

Infrastructure-as-a-Service Usage Determinants in Enterprises

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Wirtsch.-Inf. Jörg Johannes Strebel

Tag der mündlichen Prüfung: 30.07.2014

Referent: Prof. Dr. Christof Weinhardt

Korreferent: Prof. Dr. Orestis Terzidis

Karlsruhe, 2014

Abstract

Hybrid Cloud Computing, the dynamic, combined use of internal and external IT resources for business applications, is a new IT sourcing concept that was made possible by the rapid development of Internet connectivity and the broad establishment of IT standards. IaaS (Infrastructure-as-a-Service) as one type of Cloud Computing promises the flexible and scalable provisioning of IT infrastructure resources, which are offered as digital services and are billed in a pay-as-you-go fashion. Despite the advantages and the widespread press attention that these offers receive, the actual implementation and deployment rate is still small in business settings in Germany according to industry studies. Therefore, the question arises, what the determinants of Cloud Computing usage of enterprises are. The focus lies on infrastructure services, as they represent a special and interesting subset from a research point of view.

A mixed-method approach was chosen in this thesis to tackle the research question. The legal, organizational and sociological determinants are explored using case-studies and expert interviews. These determinants and the determinants gathered through a careful literature analysis result in an IaaS adoption model of enterprises, formulated as a structural equation model. This model incorporates the actual IaaS usage decision in the enterprise; it describes both enterprises that are already using IaaS and those that are not. Several established theories from social sciences and information systems research are employed, such that a broad theoretical foundation is in place. The IaaS adoption model is evaluated in a Web survey. As the economical determinants are especially important for the research question, they are separately investigated. To this end, the results of the Web survey form the basis for a cost-optimizing decision support model which yields the IaaS provider-related and software application-related determinants for an IaaS deployment. This decision support model is then applied to a potential IaaS use case of a large automobile manufacturer.

The determinants hypothesized in the IaaS adoption model are largely supported by the empirical results. Aspects of perceived usefulness (i.e. strategic, flexibility and efficiency added-values) are the strongest predictors of IaaS adoption. Small and medium enterprises are more prone to IaaS adoption and are less risk-averse regarding the uncertainties involved. The detailed analysis of quality determinants in the use case reveals that current IaaS providers tend to support the specialized IT resource quality requirements of current business applications only partially; especially strong quality-of-service guarantees are not available. Under the limitations of the use case and its quality restrictions, the decision support model shows, that taking advantage of Cloud resource elasticity yields significant cost savings. The model exhibits high predictive power in the given setting. Its suitability for supporting IaaS-related decisions is rated favorably by an industry expert.

The empirical findings of the IaaS adoption model are consistent with related studies from a Software-as-a-Service background, but extend the understanding of uncertainty, informant role and firm size on IaaS adoption. These findings help business executives determine the selling points of an IaaS adoption and the best organizational setting for an IaaS introduction. The decision support model extends similar models found in the literature and allows the identification of IaaS deployment predictors based on financial indicators. These predictors can then help business executives find software applications suitable for IaaS deployment with minimal effort. The results are helpful both for executives wishing to optimize their IaaS utilization and for researchers looking for IaaS usage determinants.

Zusammenfassung

Hybrid Cloud Computing, also die gemeinsame Nutzung von internen und externen IT-Ressourcen für Geschäftsanwendung, ist ein neues IT-Beschaffungskonzept, das durch die schnelle Entwicklung der Internetkonnektivität und durch die breite Etablierung von IT-Standards ermöglicht wurde. IaaS (Infrastructure-as-a-Service) als eine Ausprägung des Cloud Computings verspricht die flexible und skalierbare Bereitstellung von IT Infrastrukturressourcen, die als Internet-basierte Dienste angeboten und nach Nutzung abgerechnet werden. Trotz dieser Vorteile und trotz der weitverbreiteten Aufmerksamkeit der Medien, ist die tatsächliche Implementierungs- und Einsatzrate im geschäftlichen Umfeld in Deutschland nach Branchenstudien eher klein. Daher stellt sich die Frage, was die Determinanten des Einsatzes von solchen Diensten in Unternehmen sind. Der Fokus liegt in dieser Arbeit auf Infrastrukturdiensten, da diese aus Forschungssicht eine spezielle und interessante Teilmenge aller Dienste darstellen.

Zur Beantwortung der Forschungsfrage wurden mehrere Methoden kombiniert. Rechtliche, organisatorische und soziologische Determinanten werden mittels Fallstudien und Experteninterviews erkundet. Die so gewonnenen Hypothesen und die Hypothesen aus der Literatur werden zu einem IaaS-Adoptionsmodell des Unternehmens weiterentwickelt. Mehrere etablierte Theorien aus den Sozialwissenschaften und aus der Wirtschaftsinformatik werden eingesetzt, so dass eine breite theoretische Fundierung gesichert ist. Das IaaS-Adoptionsmodell wird in einer Web-basierten Umfrage evaluiert. Aufgrund der besonderen Wichtigkeit der ökonomischen Faktoren für die Forschungsfrage werden diese gesondert untersucht. Dazu werden die Ergebnisse der Umfrage als Basis für ein kostenoptimierendes Entscheidungsunterstützungsmodell genutzt, das die IaaS-Anbieter-spezifischen und die Softwareanwendungs-spezifischen Determinanten ermittelt. Dieses Modell wird anhand eines möglichen IaaS-Anwendungsszenarios eines großen deutschen Automobilherstellers evaluiert. In diesem Kontext werden auch mögliche qualitätsbezogene Determinanten diskutiert.

Die in den Hypothesen angenommenen Determinanten werden im Wesentlichen durch die empirischen Ergebnisse bestätigt. Die wahrgenommene Nützlichkeit ist der stärkste Prädiktor einer IaaS-Adoption. Kleine und mittlere Unternehmen neigen eher zur IaaS-Nutzung und sind weniger risikoavers, was die begleitenden Unsicherheiten betrifft. Aus Qualitätssicht ergibt sich, dass aktuelle IaaS Anbieter die besonderen Qualitätsanforderungen von Geschäftsanwendungen an IT Ressourcen nur teilweise abdecken; insbesondere sind keine starken Dienstgütegarantien erhältlich. Unter Berücksichtigung dieser Qualitätseinschränkung lassen sich aber durch das Entscheidungsunterstützungsmodell in der Fallstudie dann signifikante Kosteneinsparungen erzielen, wenn die Ressourcenelastizität der Cloud ausgenutzt wird. Das Modell erzielt eine hohe Vorhersagegüte und seine Brauchbarkeit zur Unterstützung von IaaS-bezogenen Entscheidung wird von einem Branchenexperten bestätigt.

Die Untersuchungsergebnisse des IaaS-Adoptionsmodells stehen im Einklang mit verwandten Studien aus dem Software-as-a-Service-Umfeld, aber sie erweitern gleichzeitig das Wissen um die Wirkung von Unsicherheit, Teilnehmerrolle im Unternehmen und Firmengröße auf die IaaS-Adoption. Die Ergebnisse helfen Führungskräften, die Alleinstellungsmerkmale einer IaaS-Adoption zu erkennen und die Verortung der IaaS-Einführung im Unternehmen zu bestimmen. Das Entscheidungsunterstützungsmodell erweitert ähnliche Modelle aus der Literatur und ermöglicht die Ableitung von kostenoptimierenden Regeln zum IaaS-Einsatz. Die Resultate sind hilfreich sowohl für Manager, die ihre IaaS-Nutzung optimieren wollen, als auch für Forscher, die nach IaaS-Nutzungsdeterminanten suchen.

Contents

List of Abbreviations	v
List of Figures	vii
List of Tables	x
I Preliminaries	1
1 Introduction	2
1.1 Research Questions	5
1.2 Thesis Outline	6
2 Foundations of IaaS Sourcing - Terms & Concepts	8
2.1 Introduction	8
2.2 Business Applications as Sourcing Demand Drivers	8
2.3 IT Resource Providers as Suppliers	10
2.3.1 Cloud Computing Nomenclature	10
2.3.2 Contrasting and Comparing Grid and Cloud Computing	15
2.3.3 IaaS Market Overview	17
2.3.4 IaaS and Outsourcing - Definitions and Comparison	19
2.4 Literature Review	23
2.5 Discussion and Summary	28
II Empirical Research	29
3 Functional Determinants for IaaS Sourcing	30
3.1 Introduction	30
3.2 BMW Case Study	31
3.2.1 Description of the BMW Group	31
3.2.2 Case Description	33
3.2.3 Case Study Interviews	38
3.3 Qualitative Model of IaaS Usage Determinants	42
3.3.1 IaaS Adoption Model	42
3.3.2 Results	46
3.4 Quantitative Model of IaaS Usage Determinants	47
3.4.1 Theoretical and Empirical Basis	48
3.4.2 Hypotheses and Causal Model	50
3.5 Discussion and Summary	53

4	Empirical Model Evaluation	55
4.1	Evaluation Approach	55
4.2	Development of the Survey Instrument	56
4.2.1	Structural Equation Modeling Concepts	56
4.2.2	Structural Model and Measurement Model	58
4.2.3	Online Questionnaire Development and Pretest	59
4.3	Data Collection and Preparation	61
4.4	Data Analysis and Results	62
4.4.1	Model Estimation Method	63
4.4.2	Descriptive Analytics	64
4.4.3	Quality Assessment of Measurement Model	73
4.4.4	Quality Assessment of Structural Model	90
4.4.5	Moderating Effects	93
4.4.6	Assessment of Hypotheses	96
4.5	Discussion of Empirical Results	97
4.5.1	Implications	97
4.5.2	Limitations of the Research Approach	99
4.5.3	Contributions	100
III	Experimental Research	103
5	Efficient Allocation Using IaaS Sourcing	104
5.1	Introduction	104
5.2	Scenario Assumptions	106
5.3	Quality Dimensions	110
5.4	IaaS Tariff Modeling	114
5.4.1	TCO Approaches	114
5.4.2	Tariff Model	115
5.5	IT Resource Requirements of Business Applications	119
5.5.1	IT Resource Model	120
5.5.2	IaaS Instance Type Selection	123
5.6	IaaS Usage Optimization Model	125
5.7	Decision Tree	127
5.7.1	Motivation for Analytical Approach	127
5.7.2	Decision Tree Set-up	128
5.7.3	Decision Tree Performance Evaluation	131
5.8	Discussion and Summary	132
6	Decision Support Model Evaluation	135
6.1	Evaluation Approach	135
6.2	Statistical Design of Experiments	136
6.3	Quality Requirements in the Case Study	137
6.3.1	Description of Relevant Quality Dimensions	138
6.3.2	Quality Model Mapping	140
6.3.3	Comparison of an In-house Storage Service to an IaaS Offer	140
6.4	Data Collection	141
6.4.1	Descriptive Workload Characteristics	145

6.5	Experimental Results	146
6.5.1	Effect of α on AWS Instance Cost	146
6.5.2	Financial Importance of Different Resource Types	148
6.5.3	Variability of Different Resource Types	150
6.5.4	Decision Tree Results	150
6.6	Discussion of Experimental Results	160
6.6.1	Implications	160
6.6.2	Limitations of the Evaluation	165
6.6.3	Contributions	166
IV	Finale	169
7	Conclusion and Outlook	170
7.1	Conclusion	170
7.1.1	Research Question R1	170
7.1.2	Research Question R2	172
7.1.3	Research Question R3	172
7.2	Outlook	174
	Declaration about the thesis	xiii
	Lebenslauf	xiv
	Appendix	xv
A	Letter of Invitation	xvi
B	The Questionnaire	xviii
C	Interview Guidelines - Case Study	xxviii
D	Interview Guidelines - Explorative Study	xxxi
E	Qualitative Survey Feedback	xxxiii
F	Amazon AWS Cost Figures	xxxv
G	Quality Model Mapping	xxxvii
H	Results of Decision Tree Parameter Optimization	xxxix
	References	xl

List of Abbreviations

ACID	<u>A</u> tomicity <u>C</u> onsistency <u>I</u> solation <u>D</u> urability
AHP	<u>A</u> nalytical <u>H</u> ierarchical <u>P</u> rocess
ANOVA	<u>A</u> nalysis of <u>V</u> ariance
API	<u>A</u> pplication <u>P</u> rogramming <u>I</u> nterface
ARIMA	<u>A</u> uto <u>R</u> egressive <u>I</u> terated <u>M</u> oving <u>A</u> verage
ASP	<u>A</u> pplication <u>S</u> ervice <u>P</u> roviding
AVE	<u>A</u> verage- <u>V</u> ariance- <u>E</u> xtracted
AWS	<u>A</u> maz <u>o</u> n <u>W</u> eb <u>S</u> ervices
BI	<u>B</u> usiness <u>I</u> ntelligence
BMW	<u>B</u> ayerische <u>M</u> otoren <u>W</u> erke
CAE	<u>C</u> omputer <u>A</u> ided <u>E</u> ngineering
CDN	<u>C</u> ontent <u>D</u> elivery <u>N</u> etwork
CFD	<u>C</u> omputation <u>F</u> luid <u>D</u> ynamics
CRM	<u>C</u> ustomer <u>R</u> elationship <u>M</u> anagement
DoE	<u>D</u> esign of <u>E</u> xperiments
EBS	<u>E</u> lastic <u>B</u> lock <u>S</u> torage
EC2	<u>E</u> lastic <u>C</u> ompute <u>C</u> loud
EOQ	<u>E</u> conomic <u>O</u> rd <u>e</u> r <u>Q</u> uantity
ERP	<u>E</u> nterprise <u>R</u> esource <u>P</u> lanning
EU	<u>E</u> uropean <u>U</u> nion
GARCH	<u>G</u> eneralized <u>A</u> utoregressive <u>C</u> onditional <u>H</u> eteroscedasticity
GUI	<u>G</u> raphical <u>U</u> ser <u>I</u> nterface
HPC	<u>H</u> igh <u>P</u> erformance <u>C</u> omputing
IaaS	<u>I</u> nfr <u>a</u> stru <u>c</u> ture- <u>a</u> s- <u>a</u> - <u>S</u> ervice
IBM	<u>I</u> nternational <u>B</u> usiness <u>M</u> achines
IEC	<u>I</u> nternational <u>E</u> lectrotechnical <u>C</u> ommission
ISO	<u>I</u> nternational <u>O</u> rgan <u>i</u> zation for <u>S</u> tandardization
ITIL	<u>I</u> T <u>I</u> nfr <u>a</u> stru <u>c</u> ture <u>L</u> ibrary
KMO	<u>K</u> aiser- <u>M</u> eyer- <u>O</u> lkin
MILP	<u>M</u> ixed <u>I</u> nteger <u>L</u> inear <u>P</u> rogramming
MIMIC	<u>M</u> ultiple <u>I</u> ndicators <u>M</u> ultiple <u>C</u> auses
MS	<u>M</u> icrosoft
MSA	<u>M</u> easure of <u>S</u> ampling <u>A</u> dequacy
NIST	<u>N</u> ational <u>I</u> nstitute of <u>S</u> tandards and <u>T</u> echnology
OGF	<u>O</u> pen <u>G</u> rid <u>F</u> orum
OTD	<u>O</u> rd <u>e</u> r- <u>t</u> o- <u>D</u> elivery
PaaS	<u>P</u> latf <u>o</u> rm- <u>a</u> s- <u>a</u> - <u>S</u> ervice

PLS	<u>P</u> artial <u>L</u> east <u>S</u> quares
QoS	<u>Q</u> uality of <u>S</u> ervice
RDBMS	<u>R</u> elational <u>D</u> atabase <u>M</u> management <u>S</u> ystem
RMSEA	<u>R</u> oot <u>M</u> ean <u>S</u> quare <u>E</u> rror of <u>A</u> pproximation
SaaS	<u>S</u> oftware- <u>a</u> s- <u>a</u> - <u>S</u> ervice
SAPS	<u>S</u> AP <u>A</u> pplication <u>P</u> erformance <u>S</u> tandard
SAS	<u>S</u> erial <u>A</u> ttached <u>S</u> CSI
SAS	<u>S</u> tatement on <u>A</u> uditing <u>S</u> tandard
SCM	<u>S</u> upply <u>C</u> hain <u>M</u> anagement
SCSI	<u>S</u> mall <u>C</u> omputer <u>S</u> ystem <u>I</u> nterface
SDLC	<u>S</u> ystem <u>D</u> evelopment <u>L</u> ife <u>C</u> ycle
SEM	<u>S</u> tructural <u>E</u> quation <u>M</u> odeling
SLA	<u>S</u> ervice <u>L</u> evel <u>A</u> greement
SME	<u>S</u> mall and <u>M</u> edium <u>E</u> nterprises
SOA	<u>S</u> ervice <u>O</u> riented <u>A</u> rchitecture
SORMA	<u>S</u> elf- <u>O</u> rganizing <u>I</u> CT <u>R</u> esource <u>M</u> anagement
SPOF	<u>S</u> ingle <u>P</u> oint of <u>F</u> ailure
SSD	<u>S</u> olid <u>S</u> tate <u>D</u> isk
SSL	<u>S</u> ecure <u>S</u> ocket <u>L</u> ayer
STARD	<u>S</u> tandardisation of <u>P</u> rocesses and <u>S</u> ystems
TAM	<u>T</u> echnology <u>A</u> cceptance <u>M</u> odel
TCO	<u>T</u> otal <u>C</u> ost of <u>O</u> wnership
TFLOPS	<u>T</u> era <u>F</u> loating <u>P</u> oint <u>O</u> perations <u>P</u> er <u>S</u> econd
VM	<u>V</u> irtual <u>M</u> achine
VO	<u>V</u> irtual <u>O</u> rganisation
VPN	<u>V</u> irtual <u>P</u> rivate <u>N</u> etwork

List of Figures

1.1	Thesis structure	7
2.1	Cloud Sourcing Options and the Enterprise Architecture Stack	12
2.2	Deployment-Delivery-Matrix	15
2.3	Typology of IaaS Cloud Computing providers	19
2.4	Outsourcing Process	21
2.5	Positioning of IaaS in the outsourcing spectrum	22
3.1	Global BMW production network	31
3.2	BMW IT Organization	32
3.3	BMW Processes	33
3.4	BMW IT Innovation Management Process	33
3.5	STARD system landscape	34
3.6	Phases of the BMW product development process	35
3.7	Plant Simulation GUI	35
3.8	Crash simulations	37
3.9	Component safety simulations	37
3.10	Rollover simulations	37
3.11	Aerodynamics simulations	37
3.12	Runtime distribution of CAE jobs (in h)	38
3.13	Research model of explorative study	46
3.14	Causal model	53
4.1	Importance of different quality criteria	66
4.2	IaaS budget shares	67
4.3	IaaS provider credentials and reputation	68
4.4	Compatibility considerations of IaaS users	68
4.5	Demonstrability properties of IaaS	69
4.6	Trialability Properties of IaaS	70
4.7	Difficulty of IaaS provider monitoring	70
4.8	Conflicting factors tested in the CFA	80
4.9	Nomological Model for TPB, TRA theories	84
4.10	Results of PLS analysis of the structural model	91
4.11	Survey methodology	97
5.1	Black box Process model	105
5.2	Scenario assumptions derived from the empirical results	107
5.3	Technical architecture of an IaaS Cloud usage scenario	110
5.4	Parametric cost model	117

5.5	Resulting total cost function	117
5.6	Exemplary IT resource demand distributions for two applications	120
5.7	Example of IaaS deployment	120
5.8	Example of in-house deployment	121
5.9	Effect of decision tree depth on classification quality	130
6.1	Experimental Process model	135
6.2	Logical data model for evaluation data	138
6.3	Workload scatterplot matrix	148
6.4	Effect of α on AWS instance cost	149
6.5	Normalized semi-variance per resource type	150
6.6	Coefficient of variation per resource type	150
6.7	Tabular Decision Tree - Unweighted Cases	152
6.8	Visualization of the Decision Tree - Unweighted Cases	153
6.9	Tabular Decision tree - Weighted Cases	154
6.10	Visualization of the Decision Tree - Weighted Cases	155

List of Tables

1.1	Enterprise Cloud Adoption Trends by Cloud Layer – Everest Survey	4
1.2	Enterprise Cloud Adoption Trends by Cloud Layer – KPMG Survey	4
1.3	Research questions	5
2.1	Types of enterprise Cloud Service Usage - IDC Survey	9
2.2	Types of enterprise Cloud Service Usage - XaaS survey	9
2.3	Types of enterprise Cloud Service Usage - Everest survey	10
2.4	Usage characteristics of Grid and Cloud Computing	16
2.5	Technical characteristics of Grid and Cloud Computing	16
2.6	Outsourcing process	21
2.7	Related work in technology adoption	23
2.8	Related work in IaaS decision support models	25
2.9	Cost-based Vendor selection approaches	27
3.1	BMW Expert Map	39
3.2	Expert interview results	42
3.3	Summarized Interview Results	46
3.4	Connection between Hypotheses and Causal Model Constructs	51
4.1	Measurement Models	59
4.2	Descriptive Statistics	65
4.3	Relationship between IaaS Quality Dimensions and Company Headcount	71
4.4	Relationship between IaaS adoption metrics and company headcount	72
4.5	Relationship between IaaS adoption metrics and IT affiliation	73
4.6	Relationship between IaaS quality metrics and IT affiliation - Descriptive Statistics	74
4.7	Relationship between IaaS quality metrics and IT affiliation - ANOVA results	74
4.8	Quality assessment criteria for reflective measurement models	75
4.9	Quality assessment criteria for formative measurement models	76
4.10	Assessment of content validity	78
4.11	Model fit measures for the CFA	81
4.12	Standardized Residual covariances	81
4.13	Assessment of PLS loadings for reflective constructs	82
4.14	PLS Reliability Scores for Reflective Constructs	83
4.15	Discriminant Validity: Fornell-Larcker Criterion	83
4.16	Global Fit Measures for the Nomological Model	85
4.17	Standardized Residual Covariances for the Nomological Model	85
4.18	Variance Inflation Factor Data for the Formative Constructs	86
4.19	Correlation Matrix for the Formative Indicators	87

4.20	Formative Indicator Prognostic Validity	89
4.21	Quality Assessment Criteria for Structural Models	90
4.22	f^2 effect size measures	92
4.23	Path-related Stone-Geisser criterion for prognostic relevance q^2	93
4.24	One-Sample Kolmogorov-Smirnov Test	94
4.25	LV scores by company size	95
4.26	Significance of company size	95
4.27	LV scores by IT affiliation	96
4.28	Significance of IT affiliation	96
4.29	Group differences in the relationship between SEU and INT	97
5.1	Objective Quality dimensions of IT services	112
5.2	Cost model for IaaS Cloud resources	115
5.3	Input variables of the decision tree	129
5.4	Base settings for the decision tree parameters	130
6.1	Mapping of BMW storage quality requirements on system-related quality dimensions	141
6.2	Mapping of BMW storage quality requirements on operations-related quality dimensions	142
6.3	Comparison of storage services	143
6.4	Contrast of storage services	144
6.5	Data Set 1 - Server Monitoring	144
6.6	Data Set 2 - Application Monitoring	145
6.7	Data Set 3 - IT application landscape	145
6.8	BMW instance types	146
6.9	Descriptive workload statistics	146
6.10	Workload correlation coefficients	147
6.11	Comparison of cost type percentages	150
6.12	Decision Tree Performance - Unweighted Cases	156
6.13	Decision Tree Performance - Weighted Cases	156
6.14	Decision Tree Performance for Random Placement - Base Settings	156
6.15	Predictive Performance of pure IaaS Deployment	157
6.16	Cost deviations per leaf caused by prediction	158
6.17	Cost deviations per leaf caused by weighted prediction	159
6.18	Reproduction of the outsourcing degree	159
6.19	Comparison of predictor variables	161
D.1	Interview guideline	xxxii
E.1	Sentiment of Feedback	xxxiii
E.2	Negative Feedback	xxxiii
E.3	Topics mentioned in the Feedback	xxxiv
F.1	Amazon AWS Compute Instances	xxxv
F.2	Amazon AWS EBS IO Tariff	xxxvi
F.3	Amazon AWS Network Tariff	xxxvi
F.4	Amazon AWS Support Tariff	xxxvi
G.1	Quality Model Mapping	xxxviii

H.1 Sensitivity analysis of decision tree parameters	xxxix
--	-------

Acknowledgments

I would like to thank Prof. Christof Weinhardt for his encouragement and his supervision throughout the creation of this thesis at his Chair. My research would not have been possible without Simon Caton PhD, Margeret Hall, Prof. Jan Krämer, Prof. Thomas Setzer, Dr. Marc Adam and Dr. Georg Wiedemann, who provided insights and encouragement in times of need. The BMW Group was also supportive in creating this work; special thanks go to Dr. Markus Greunz and Thomas Sutter.

Part I

Preliminaries

Chapter 1

Introduction

Cloud Computing gains importance as a new paradigm of using IT resources of all kind, which are provided 'as-a-service' over the Internet. Many commercial research analysts from e.g. Gartner or Forrester, consider Cloud Computing as one of the most significant trends with a great potential for changing the whole IT industry (Gartner Inc. 2009), (Hayes 2008). In fact, Cloud Computing promises from a client perspective a number of advantages compared to so-called on-premise business solutions, i.e. hardware and software systems. These systems are usually based on purchasing and licensing agreements and they are deployed on the client's site exclusively for a single corporation's users. One of the salient features of Cloud Computing is the innovative use of dynamic, pay-as-you-go pricing schemes; therefore Cloud Computing is an example of utility computing which was described in earlier research literature (e.g. by Ross and Westerman (2004)) even before the term Cloud Computing was coined. These pricing schemes eliminate the need for high one-time investments for hardware and software; billing is based on the actual consumption, i.e. the risk of capacity under-utilization is partly or completely transferred to the Cloud Computing provider. The scalability of Cloud-based IT resources ensures that unexpected peak loads (e.g. numerous users or transactions) can be processed without service interruption; these peaks are also called Cloud bursts.

These advantages are often cited by commercial analysts and software vendors, but is Cloud Computing actually a relevant real-world phenomenon that is worth analyzing scientifically? The IT industry is known for its boom-and-bust technology cycles, in which allegedly new concepts are presented every other year (e.g. SOA or ASP just to name a few examples), yet most of them fail to catch on (SOA was only a hype topic for a couple of years). The question of Cloud Computing relevance has triggered the curiosity of IT industry magazines. Their findings may not be scientifically sound, yet they paint an interesting picture of Cloud Computing adoption, which warrants a further scientific analysis.

IDC Central Europe (2009) surveyed in total 805 enterprises with more than 100 employees in Germany in March 2009 about Cloud Computing. The survey showed that 75% of those enterprises had not concerned themselves with this topic; 4% answered that they had decided against using Cloud Computing after exhaustive assessments. The main reasons for this decision lie in data security concerns and the infringement of legal regulations (compliance concerns). However, 45% of the enterprises polled assume that Cloud Computing will establish itself in the next years and that it will represent a complementary option for sourcing IT services. Moreover, many users are dubious about the added value of Cloud Computing. As a consequence, potential enterprise users tend to wait until applicable and beneficial use cases of Cloud Computing have been identified and described more clearly.

The CIO Magazine (2009) conducted a CIO Cloud Computing Survey in June 2009 with the purpose of measuring enterprise Cloud Computing adoption among IT decision-makers. The participants were CIO.com Web site visitors involved in purchasing IT-related products and services. The survey took place from June 4 – June 21, 2009; the survey findings are based on 240 responses from IT professionals in a

variety of industries including high tech, telecom & utilities, government and nonprofits including education, services, manufacturing, financial services and healthcare. Over half of the respondents are the head of IT at their company or business unit. The participating companies evenly varied in size across the whole range (from <\$100 million to >\$1 billion). The four greatest concerns surrounding Cloud Computing were security, loss of control over data, regulatory/compliance concerns and performance issues. The primary reasons for considering Cloud Computing were stated as reduced hardware infrastructure costs, reduced IT staffing/administration costs, access to skills/capabilities that the enterprise has no interest in developing in-house and the scalability on demand/flexibility to the business. For almost 68% of the enterprises polled, Cloud Computing is a technology that they are exploring or will explore within the next one to three years. "Although respondents cited cost savings most frequently as the reason for adopting Cloud Computing, many are not sure such investments will help them to reduce their IT budgets. Half of IT decision-makers expect some percent of their IT budget will be devoted to on-demand services in the next five years while slightly fewer (42%) anticipate any reduction in their IT spending as a result. Another 42% aren't sure they'll achieve any savings, while 16% say they don't expect any impact on their IT budgets." (CIO Magazine 2009, p. 1).

Avanade Inc. (2009) published in October 2009 the results of a survey among more than 500 executives and IT managers from 16 countries. One of the results was that German companies have higher security concerns than companies from other countries. Whereas only 40% of those polled world-wide mentioned security concerns, 64% of the German respondents did. In Germany, 40% of the survey participants thought, that they could save money using Cloud Computing, however 60% of the German participants considered this new IT paradigm as a tactical investment. Another important result of this global survey: the number of enterprises, that was then planning or already testing Cloud Computing, had risen sharply in the previous months. More than half of the participants had chosen a combination of Cloud-based and internal, on-premise IT systems. There is a significant trend towards the deployment of Hybrid Cloud systems, as enterprises grow familiar with the new technology. 43% of the German respondents explained that they adopt both Cloud-based and on-premise solutions. Avanade experts expect the future of Cloud-based solutions to lie definitely in a combined approach of Cloud Computing and proprietary in-house systems. At the same time, there will always be some software applications and business processes, that are not suited for the Clouds. Therefore, enterprises have to make a deliberate decision what applications to migrate and what applications to run on-premise (Avanade Inc. 2009, p. 2).

The federal association BITKOM e.V. (2010) published the results of a survey among its members in January 2010. In this study, Cloud Computing was named one of the most important IT trends in 2010 (45% of the people polled voted in favour of it). The expectations among the respondents resemble the well-known claims by Cloud-Computing proponents: this technology is supposed to make corporations leaner and more efficient, as they don't have to provision their complete IT resources beforehand, but they can access them online when needed. This federal association represents more than 1300 businesses with over 700.000 employees.

The business consulting company Everest Group conducted an online survey in the first half of 2012, where they sent out email invitations to Cloud service buyers (Everest Group 2012). The survey included 105 respondents, 68% of whom were located in North America. Rather large companies were targeted (68% of the participating companies had more than 100 million \$ in annual revenue). The respondents' roles were predominantly either senior managers or high-ranking executives. One of the survey questions was, what type of Cloud Computing was predominantly adopted by the respondents and what the plans for adoption were. Table 1.1 shows the results. Compared to the other types of Cloud Computing, IaaS shows the second-to-last smallest share of current adoption. Many participants only plan to adopt it either in the near or distant future (esp. the Hybrid Cloud variant).

Table 1.1: Enterprise Cloud Adoption Trends by Cloud Layer – Everest Survey

	Already adopted	Adopted in near future	Adopt in distant future	No plans to adopt	No. of respondents
Software-as-a-Service	57%	28%	10%	5%	81
Platform-as-a-Service	38%	25%	27%	10%	73
Infrastructure-as-a-Service (Public Cloud)	31%	26%	26%	18%	78
Infrastructure-as-a-Service (Private Cloud)	30%	36%	23%	11%	80
Infrastructure-as-a-Service (Hybrid Cloud)	17%	27%	36%	20%	70
Business-Process-as-a-Service	28%	27%	22%	23%	64

Table 1.2: Enterprise Cloud Adoption Trends by Cloud Layer – KPMG Survey (KPMG AG 2013, p.27)

	Already adopted	Adoption planned	Adoption discussed	No answer	No. of respondents
Software-as-a-Service	17%	15%	34%	34%	102
Platform-as-a-Service	13%	8%	20%	59%	102
Infrastructure-as-a-Service (Public Cloud)	14%	10%	25%	51%	102
Business-Process-as-a-Service	11%	5%	16%	68%	102

The KPMG Cloud Monitor (KPMG AG 2013) is an annual survey that analyzes the current and planned usage of different types of Cloud Computing; it takes place every year from 2011 to 2014. The 2013 survey was conducted as telephone interviews at the end of 2012. The sample includes 436 participants from German enterprises with at least 20 employees. The respondents exclusively belong to either the executive level of the information technology organization or to the management of the enterprise. The stratification of the sample ensures that enterprises of different sizes and industries are represented in sufficient number. In total, about 10% of all enterprises use Public Clouds, according to the study. When asked for the type of Cloud Computing adopted or planned, 102 participants answered. Table 1.2 summarizes these responses. Public IaaS is only deployed at 14% of the enterprises. Hybrid Cloud Computing was not part of the survey; however, two thirds of the Public Cloud users also have Private Cloud solutions in place (KPMG AG 2013, p.10). Enterprises obviously supplement their Private Cloud environments selectively with Public Cloud services.

As a conclusion, a number of similar statements become apparent in these surveys:

- Cloud Computing is expected to offer better transparency regarding cost efficiency and applicability; however, its added value remains unclear so far and its beneficial use cases still have to be identified.
- Cloud Computing has gained traction in the corporate IT world and is likely to shape the future of corporate IT service delivery. It is more than a temporary hype, but rather an emerging trend; however, enterprises are exploring it with caution and see it as a tactical, mid-term to long-term opportunity.

- The combined approach of IT service sourcing (Hybrid Cloud) is supposed to be a promising future approach to Cloud Computing, according to experts. However, its current adoption is lacking.

Hence, the market is embracing this new technology, and enterprises make their first carefully planned steps in the direction of Cloud Computing adoption (unfortunately, there are no IaaS-specific market surveys, to the best of the author's knowledge). However, enterprises also perceive the risks associated with this sourcing approach. This thesis solely focuses on the Infrastructure-as-a-Service model of Cloud Computing as the relative homogeneity of IaaS offers makes the comparison of different IaaS providers with one another and with the internal, on-premise IT service delivery possible in a rather objective fashion. This homogeneity stems from the relatively high homogeneity of IT platforms used in current IaaS Clouds and enterprise data centers (usually x86-based virtualized hardware resources). Other service models of Cloud Computing (Platform-as-a-Service, Software-as-a-Service) do not possess this level of communality.

IaaS must not be mixed up with traditional outsourcing arrangements; the distinction between those two concepts will be further explained in chapter 2.3.4. The key word in all Cloud-based sourcing models is flexibility. In this context, it describes a property of a typical IaaS offering: the duration of such a sourcing agreement and the volume of IT services purchased is not predefined at the onset of the sourcing relationship, but can be altered flexibly, as the client sees fit. This property enables the client to freely change the amount and the mixture of the IT resources that he purchases from a single IaaS provider.

1.1 Research Questions

The conclusions above directly lead to the research questions which are pursued in this thesis. Table 1.3 lists the three main research questions and the associated research approach to answer each of them.

Table 1.3: Research questions

Research questions		Approach
R1	What are the overall IaaS usage determinants of enterprises?	Explorative study: semi-structured expert interviews for hypotheses generation.
		Causal model: Empirical validation of hypotheses from the expert interviews and from literature.
R2	What determinants are relevant in an economic optimization model of hybrid IaaS sourcing?	Mathematical modeling of the outsourcing scenario using the empirically found determinants and further IaaS characteristics.
R3	What determinants in a hybrid IaaS sourcing scenario can be linked to an economically beneficial usage of public IaaS offerings?	Experimental study with a real-world use case.

Research question R1 treats the area of qualitative determinants of IaaS infrastructure sourcing within the enterprise, i.e. what are the determinants (drivers and deterrents) of Cloud-based infrastructure sourcing. Here, the focus is on the functional, technical and organizational determinants, e.g. governance, security, architectural requirements. These determinants play a major role in all stages of the decision process to use outsourcing and address the principle suitability of IaaS usage. The challenge of this research question is the wide range of possible determinants that influence an outsourcing decision, so a sensible selection of probable factors is crucial. This selection must be theoretically sound and based on previous research in the outsourcing domain and the domain of electronic services and Cloud Computing. If these determinants can indeed be identified, they will be relevant both for researchers and for practitioners: researcher benefit from

a better understanding of the fundamental theoretical constructs guiding IaaS sourcing decisions which in turn extends existing theories in this field of research; practitioners can directly apply these determinants in their sourcing decisions. This research questions requires the construction and the empirical validation of a theory of IaaS sourcing drivers. Hence, a successful answer to the research question depends on the reliability and the validity of the empirical measurement of the theoretical constructs, which can be assessed using established criteria from social science research. This line of research is further explored in chapter 3 of this thesis. The determinants found in R1 then help to answer R2 and are finally applied to a specific outsourcing situation to design a decision-support system for this situation.

Research questions R2, R3 complement R1 in that they increase the level of detail of the analysis. As economic benefits play a major role in the decision to adopt IaaS services (please see chapter 1), a deeper investigation in the cost-benefit relationship of flexible infrastructure sourcing seems warranted. To this end, a cost-based decision support model has been developed which reflects the flexible nature of IT resource usage in the Cloud; this model serves as an answer to R2. This research question is necessary as a preparation for R3, where the decision support model is experimentally evaluated. The challenges in this research question stem from the complex mathematical models required to reflect the workload characteristics of business software applications and the tariff characteristics of IaaS providers. If this decision support model is set up, it will provide practitioners with IT infrastructure cost transparency and it will help them identifying relevant determinants for their sourcing decisions. Research-wise, it will extend the state of the art as far as modeling IaaS tariffs and IT resource selection go. A successful modeling effort must be judged by the degree of applicability of the model to a real-world scenario, i.e. is the model rich enough to capture the real-world complexities? The model design is detailed in chapter 5.

In chapter 6 of this thesis, R3 is answered in the context of this model; to this end, the model is evaluated using an experimental setup. The decision support model is applied to a real-world use case; the result data is then analyzed for IaaS sourcing determinants. This approach both ensures the applicability of the decision support model in a case study and it yields IaaS sourcing determinants valid in this experimental setting. The challenges lie in the statistical analysis of the effects of various experimental factors on the outsourcing decision; these effects might be highly non-linear. The quality of a decision support model is judged by the correctness of the decisions that it recommends. Therefore, the decision quality of this model is assessed using established metrics with a statistical and machine learning background.

1.2 Thesis Outline

The thesis structure is displayed in Figure 1.1. It follows a conventional schema: Part I lays the groundwork for the whole thesis in that it introduces and motivates the research questions and defines terms and general concepts used later on. In Part II, the empirical research for R1 is described and both the expert interviews and the quantitative model are developed and evaluated. A discussion of the findings concludes this part. Part III then builds on these models and develops both a cost-based IaaS usage model and a generic quality model of IT infrastructure resources. Both models are then evaluated in an industrial setting and the findings are discussed. Part IV summarizes the answers to the research questions and contrasts and compares the results with the state of the art found in the literature. This part also gives an outlook on possible future work based on the results in this thesis.

All research is taking place as part of the Biz2Grid research project^{1,2} funded by the German D-Grid research program;³ its goal is to move business applications to the Grid. The BMW Group as an associated project partner is providing real-world use cases.

¹<http://www.im.uni-karlsruhe.de/Default.aspx?PageId=341&lang=de>, last accessed 2013-12-29

²<http://www.d-grid-gmbh.de/index.php?id=74>, last accessed 2013-12-29

³<http://www.d-grid-gmbh.de/index.php?id=1>, last accessed 2013-12-29

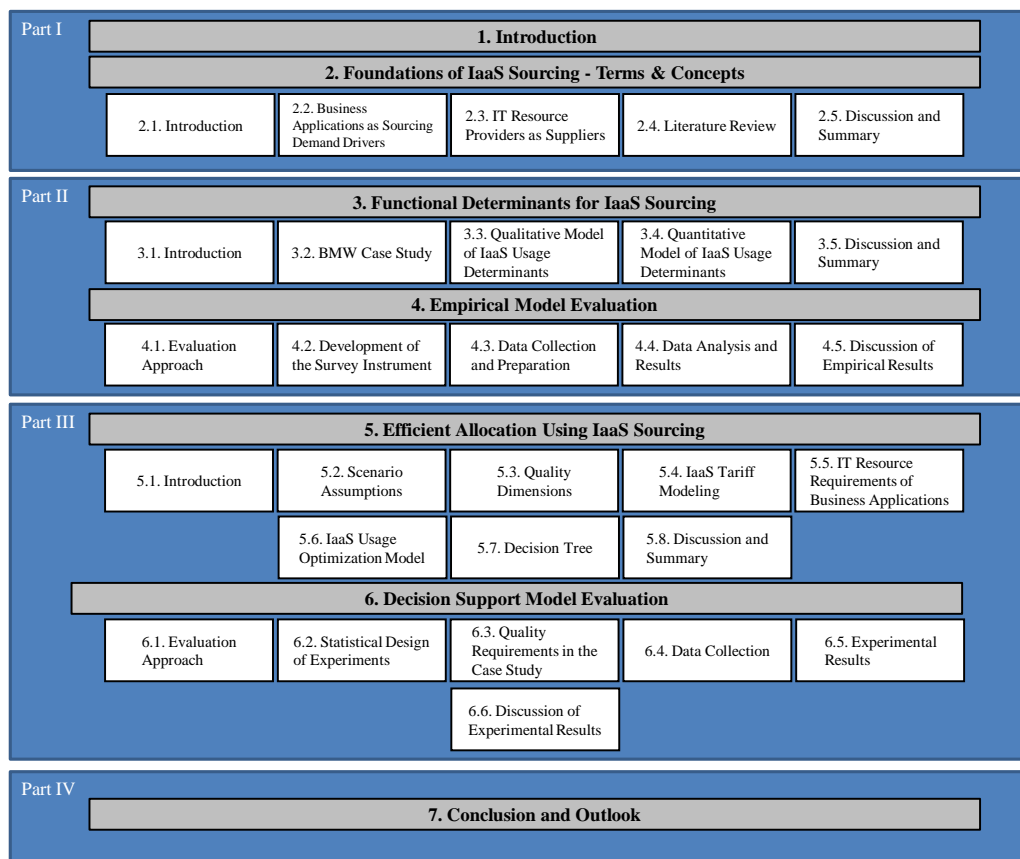


Figure 1.1: Thesis structure

Chapter 2

Foundations of IaaS Sourcing - Terms & Concepts

2.1 Introduction

This chapter gives precise definitions of key concepts and terms relevant for the further research and it lays the conceptual foundation for the remaining chapters. The Cloud Computing phenomenon is analyzed from two perspectives: the supply side and the demand side. A market-oriented approach is in order here as this research is focused on commercial Cloud offerings suitable for enterprise computing. The demand side in section 2.2 describes the current buyers' impression of business software applications that are principally fit for Cloud Computing and as such, are already sourced from the Cloud. The type of Cloud platform (IaaS, PaaS or SaaS) used for delivering the services of these business software applications is not elaborated, as only the fundamental possibility of Cloud sourcing of these applications is relevant here. The usage figures are based on commercial studies of the Cloud market and thus only show a snapshot of the situation.

The supply side in section 2.3 stresses IaaS as a special case of Cloud Computing, which is more deeply elaborated. General Cloud Computing nomenclature from recent research is introduced and it is compared to related concepts of Grid Computing and general outsourcing. These definitions are the basis for the characteristics of an IaaS provider; following these characteristics, a market analysis of IaaS providers is performed, detailing the types of offers. This section concludes the definition of terms and concepts. The investigation in this chapter is split up in both a normative and a positive research approach. The conceptual framework and the definition of the IaaS concept can be described as normative, whereas the study of the demand side and the IaaS market analysis can be described as positive, as they describe the current reality based on empirical data.

Based on these terms and concepts, a literature review regarding the three research questions follows in section 2.4. Its goal is to detail and to explain the gap in the research literature, which motivates the investigation in this thesis. The research questions are contrasted and compared to the results in the relevant literature. Several fields of research serve as a source for related work (technology adoption, decision support models, outsourcing).

2.2 Business Applications as Sourcing Demand Drivers

The business analysts from IDC (IDC Central Europe 2009) conducted a market survey among German enterprises between February and March 2009. In total, 202 enterprises were interviewed by phone using a structured questionnaire. All German industries are represented, except those that feature an especially

Table 2.1: Types of enterprise Cloud Service Usage - IDC Survey

Business applications	42%
Server (Computing capacity)	37%
E-mail Tools	30%
Office applications	25%
Storage	25%
Security tools	20%
Backup	20%
Collaborative Tools	20%
Business Community Tools	17%
System Infrastructure Software	16%
Application Development Platform	9%
Not yet decided	8%
Unified Communication Solution	7%
Other	1%

Table 2.2: Types of enterprise Cloud Service Usage - XaaS survey

Collaborative Tools	29.7%
CRM	22.5%
ERP	12.6%
external Information / Advice	12.6%
Business Process Management	9.90%
Business Intelligence	7.20%
other transaction systems	5.40%

high share of small and very small enterprises, like in the construction industry or farming. Over half of the contacts were CIOs or IT department heads and their replacements. Another 36% were IT employees in a leading position, e.g. software executives or data center managers. Business representatives like CEOs, functional managers or business executives had a share of 12% in the sample. In Table 2.1, the distribution of Cloud services is shown, that the interviewees were using or were planning to use. 161 participants contributed to this question, multiple answers were possible. Especially pronounced are business applications like CRM, ERP oder BI solutions and computing capacity.

The business analyst Wolfgang Martin and the Technical University Darmstadt set up a study in 2010 to evaluate the development of Cloud Computing in the German-speaking market (Martin, Eckert, and Repp 2010). This online-based survey took place from 25.05.2010 to 18.07.2010; the number of participants was 84, all industries were represented, but more than 50% of the participants were service providers of some sort (logistics, media, etc.) and 14% were financial service providers. Mainly large enterprises took part in the study (54% of the participating enterprises had an annual revenue of more than 100 million euros). Among other questions, the survey asked what application areas were best suited for Cloud Computing; the results of this questions are documented in Table 2.2.

The business consulting company Everest Group conducted an online survey in the first half of 2012, where they sent out email invitations to Cloud service buyers (Everest Group 2012). The survey included 105 respondents, 68% of whom are located in North America. Rather large companies were targeted (68%

Table 2.3: Types of enterprise Cloud Service Usage - Everest survey

E-mail / collaboration	87.0%
Disaster recovery / storage / data archiving	87.0%
Application development /test environment	82.0%
E-commerce and on-line tools	79.0%
Custom business applications	76.0%
BI	74.0%
ERP	62.0%

of the participating companies had more than 100 million \$ in annual revenue). The respondents' roles were predominantly either senior managers or high-ranking executives. One of the survey questions was, what type of enterprise application the participants have migrated or will migrate to the Cloud. Table 2.3 shows the percentage results.

As a conclusion, the results of all three surveys go in the same direction: a number of classical enterprise business applications are prime candidates for Cloud deployment. E-Mail, ERP and BI applications were named in all three surveys. (However, it is not indicated which type of Cloud service shall be used; it is entirely possible, that some of these applications are might be predominantly purchased from a Software-as-a-Service provider.)

2.3 IT Resource Providers as Suppliers

2.3.1 Cloud Computing Nomenclature

Cloud Computing is a service model for enabling on-demand access to a shared pool of computing resources (e.g. network, CPU, storage). Cloud Computing services can be provisioned in real-time and delivered with minimal management effort or supplier interaction. The Cloud Computing concept can partly be considered as an IT industry buzzword; a scientifically rigorous and universally accepted definition has yet to evolve. Currently, a large number of different definitions have been published (which reflects the immaturity of the whole research field). Important contributions towards a definition come from Weinhardt et al. (2009), Armbrust et al. (2009), Hayes (2008) and Vaquero et al. (2009), whose article alone features more than 20 different definitions of Cloud Computing. When the key concepts of these research papers are combined, a new suggestions for a Cloud Computing definition emerges:

Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.

The NIST (2009)(National Institute of Standards and Technology) defines Cloud Computing as follows:

Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

The two definitions are strikingly similar, so it can be assumed that they capture the essence of Cloud Computing. In the remaining thesis, the concepts of the NIST (2009) will be used as they are more comprehensive, especially when it comes to service and deployment models (which will be explained below).

Based on the literature given above, the following essential characteristics of Cloud Computing can be summarized. For Vaquero et al. (2009), these characteristics are scalability, pay-per-use utility model and virtualization. However, these properties only constitute a minimum definition, as different literature sources disagree in their perception of Cloud Computing. For the NIST, these fundamental characteristics are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.

According to Armbrust et al. (2009), the Cloud is characterized by the illusion of on-demand infinite computing resources, the elimination of an up-front commitment by Cloud users and the ability to pay for use of computing resources on a short-term basis as needed (e.g. processors by the hour and storage by the day) and release them as needed. In summary, the following criteria seem to be typical for Cloud Computing:

Scalability “Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.” (NIST 2009)

Pay-per-use utility model While NIST (2009) acknowledges that the usage of Cloud services happens in a metered fashion, their definition omits the fact, that Cloud services are metered and billed in a way resembling other utilities like gas or electricity (utility computing). However, the billing units can differ widely between different providers and can either be related to single resource units (e.g. one CPU-h, one GB per Month) or complete VM instances. The pay-per-use model also entails that there are virtually no upfront costs or large investments that have to be incurred before IT resources can be used (in contrast to conventional IT solution deployments). This property makes Clouds especially attractive for use cases that require flexibility in setting up and decommissioning IT solutions.

Virtualization Vaquero et al. (2009) considers the virtualization of hardware and software platforms as one characteristic of Cloud Computing. This view is also supported by Armbrust et al. (2009). Note that virtualization includes not only computer hardware virtualization, but also network and storage virtualization as the technological basis of Cloud Computing offers. Virtualization can also be interpreted more freely; Cloud platforms and frameworks could be considered as a form of high-level virtualization (as hinted at in (Armbrust et al. 2009) and (Vaquero et al. 2009)).

Self-Service “A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service’s provider.” (NIST 2009)

Network access “Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g. mobile phones, laptops, and PDAs)” (NIST 2009)

Resource pooling “The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g. country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.” (NIST

2009) Another aspect of resource pooling is the centralized form of control over the computing resources. These tend to be consolidated in a small number of provider-owned data centers from which customers are served globally.

Cloud Computing services can roughly classified in a rather simple ontology. For the sake of this thesis, the definitions of NIST (2009) and Vaquero et al. (2009) are applied for this ontology; (Weinhardt et al. (2009) also follow these categories, although they are referred to as business models in their research paper). A more elaborated ontology can be found in (Youseff et al. 2008), but some of the terms utilized there (like Communication-as-a-Service and Data-as-a-Service) have yet to gain wider acceptance. The relationship among the different Cloud services of this ontology is visualized in Figure 2.1.

Figure 2.1 shows how the Cloud Computing services IaaS, PaaS and SaaS can be compared to the layers of the architecture stack usually found in enterprises (Strong 2005). It becomes visible that there are both potential substitution and dependency relationships on every level of the architecture stack. Hence, the question arises where enterprises should integrate their sourcing opportunities in their IT architecture.

The application layer consists of a portfolio of business applications (as mentioned in section 2.2). These applications can principally deployed on either IaaS or PaaS Cloud services; alternatively, the corresponding applications can be migrated to a SaaS offering, where the complete functionality is provided. Usually, business applications are reliant on some sort of middleware like an application server or some other execution environment; this layer corresponds to the PaaS layer in the Cloud and could take over this function in case of a Cloud usage. The software infrastructure layer also abstracts away the peculiarities of the underlying base infrastructure. The base layer (operating system and hardware infrastructure) consists of compute resources that are utilized either in an online or a batch mode; batch computing is synonymous to the asynchronous, job-driven operation of a typical Grid middleware, so Grid infrastructure can be located at this layer. The focus of this work lies solely on the operating system / virtualization layer (see Figure 2.1). This layer corresponds to the IaaS layer in Cloud Computing.

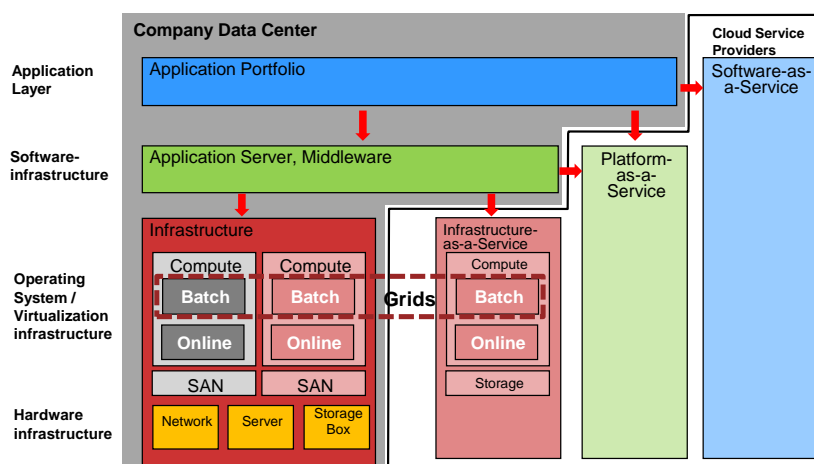


Figure 2.1: Cloud Sourcing Options and the Enterprise Architecture Stack

SaaS “The capability provided to the consumer is to use the provider’s applications running on a Cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a Web browser (e.g. Web-based email). The consumer does not manage or control the underlying Cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.” (NIST 2009) Application operations processes like software maintenance, release management, etc. are under the responsibility of the SaaS provider. The software products

offered in a SaaS manner usually support multi-tenancy, i.e. the same software product is used by a number of different users in parallel. This feature requires the separation of each user's data and each user's user-specific customizations of the software. An example of a SaaS application is the Salesforce CRM solution.¹

PaaS “The capability provided to the consumer is to deploy onto the Cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying Cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.” (NIST 2009) According to Youseff et al. (2008), one important advantage of PaaS for the developer is the availability of well-defined APIs on the platform, which support common functionality like persistent data storage, authentication or access to platform-specific functionality. An example of a PaaS offering is the Salesforce Force.com platform² or Google App Engine.³

IaaS “The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying Cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g. host firewalls).” (NIST 2009) According to Youseff et al. (2008), three different resource offerings have to be distinguished: compute services, data storage services and communication services.

Virtual machines (VMs) are the most common form for providing computing resources to the user. VMs running on the same physical system are logically isolated against each other and thus, the users can neither see other users' VMs nor the physical hardware that hosts their VMs. Within the boundaries of the VM, the user has superuser access rights and can therefore determine the software setup freely. The IaaS provider offers pre-assembled VMs that can be instantiated in the provider's Cloud as the user sees fit. Examples for IaaS providers are Amazon's Elastic Compute Cloud (EC2),⁴ part of the AWS offerings, and GoGrid's Cloud Hosting.⁵ Data storage offerings exist in two different types: one type offers stand-alone persistent data storage that can be accessed using the Internet and a Web browser (e.g. Amazon's S3). The other type is associated with a VM at the same provider and serves as a network drive for the OS (e.g. Amazon's EBS). Communication services are billed network connections sending data to or receiving data the Internet. Usually, all network connections are secured by encryption (either using a VPN or SSL connections).

Barroso and Hölzle (2009) describe the design decisions for Google's data centers; the underlying hardware hosting the VMs typically consists of commodity, x86-based servers that run a Linux distribution as a host operating system. For IaaS, a vendor-independent standard set of interfaces for Cloud resources has yet to evolve, although first steps in this direction have been taken (e.g. System Virtualization, Partitioning, and Clustering Working Group (2010)). So far, Cloud Computing providers only offer proprietary services (i.e. VMs cannot be exchanged easily between different providers, and Cloud APIs are provider-specific).

Cloud services can be deployed in several different fashions:

¹<http://www.salesforce.com>, last accessed 2013-12-29

²<http://www.salesforce.com/de/platform>, last accessed 2013-12-29

³<http://code.google.com/intl/de-DE/appengine>, last accessed 2013-12-29

⁴<http://aws.amazon.com/ec2>, last accessed 2013-12-29

⁵<http://www.gogrid.com/cloud-hosting/>, last accessed 2013-12-29

Private Cloud “The Cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.” (NIST 2009) The important difference to other deployment options is the dedicated nature of the utilized Cloud resources, which are solely used by the client organization.

Community Cloud “The Cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g. mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.” (NIST 2009)

Hybrid Cloud “The Cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g. cloud bursting for load-balancing between clouds).” (NIST 2009)

Public Cloud “The Cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling Cloud services.” (NIST 2009)

Garzotto (2010) gives a comprehensive overview of the different Cloud Computing scenarios and according examples (see Figure 2.2). The numbers in the matrix represent specific Cloud Computing sourcing options; non-sensical combinations are marked in shaded boxes. The following description explains their significance:

- No. 1** Business service delivery: complete business solutions are sourced from the Cloud. Examples include Salesforce.com CRM and Oracle On-demand.
- No. 2** This sourcing model combines in-house- and SaaS-services to deliver a complete solution. For example, MS Exchange could be deployed in-house for power users and Google Gmail could be offered for regular users.
- No. 3** In this scenario, a Cloud-based development platform is sourced, e.g. Salesforce Force.com or MS Azure.
- No. 4** A Private Cloud is used here to realize an enterprise development platform. For example, Oracle offers a corresponding middleware suite (Piech 2009).
- No. 5** This scenario is best described by the term infrastructure outsourcing; a typical example would be Amazon’s EC2 (in combination with No. 8 and 10).
- No. 6** Here, the Cloud resources are dedicated to one specific client (e.g. as an extension to his own data center); these resources can either be located in-house, and the infrastructure is operated such that it exhibits all traits of a typical Cloud service (scalability, self-service, etc.), or the resources are sourced from a third party IaaS provider. IBM’s Smart Business Storage Cloud⁶ or IBM’s Computing on Demand⁷ are examples.
- No. 7** Hybrid infrastructure resources are practical for overflow scenarios; Fraunhofer Institute’s PHAST-Grid⁸ can utilize Cloud and Grid resources transparently.
- No. 8** Storage can be sourced as a service, another type of infrastructure outsourcing; Amazon S3 is a typical example for such a service.

⁶<http://www-935.ibm.com/services/us/index.wss/offering/its/a1031610>, last accessed 2013-12-29

⁷<http://www-03.ibm.com/systems/deepcomputing/cod/index.html>, last accessed 2013-12-29

⁸<http://www.epg.fraunhofer.de/solutions/software/phastgrid/index.jsp>, last accessed 2013-12-29

No. 9 The Community Cloud concept has yet to gain wider acceptance in the literature. One conceptualization can be found in (Briscoe and Marinos 2009).

No. 10 Sourcing of Cloud communication services can become a viable scenario when additional bandwidth needs emerge, e.g. during a marketing campaign or during annual company meetings. Cloud-based CDN (Content Delivery Networks) (Buyya et al. 2009) are an active area of research.

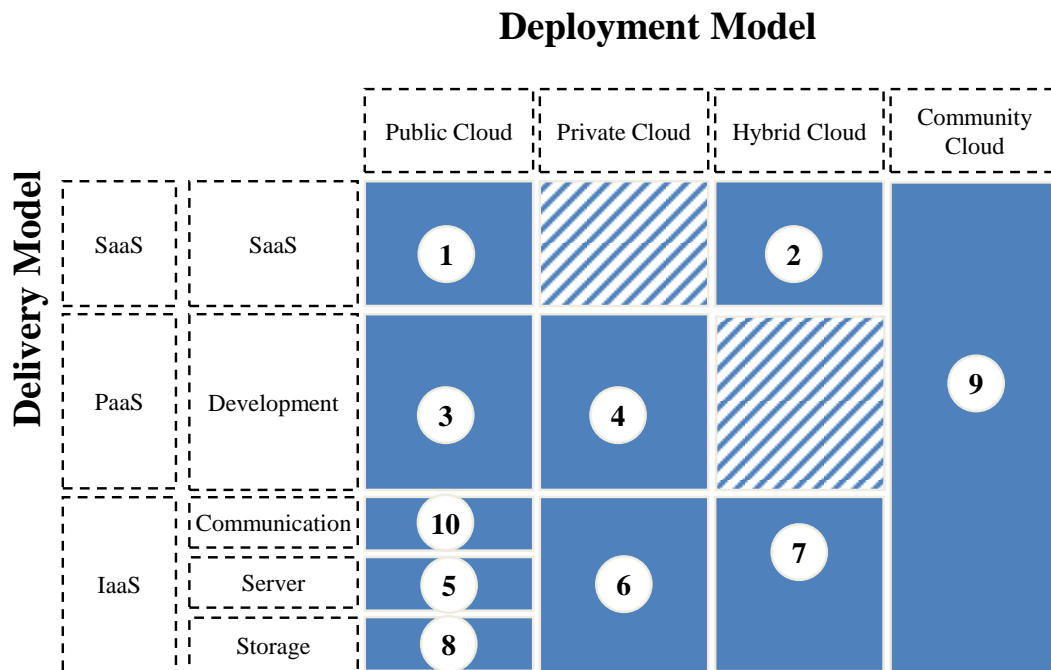


Figure 2.2: Deployment-Delivery-Matrix

Relating to the Cloud deployment models mentioned in (NIST 2009), this thesis assumes that the IaaS resources are provided by a third party which is legally independent from the IaaS client. Therefore, the above mentioned sourcing options 5, 7, 8, 10 are the focus of this thesis. Sourcing option 6 is only included if the IaaS resources are purchased as a service from a third party (private in-house Cloud is excluded, as the focus of this research lies on the outsourcing aspects of Cloud Computing).

2.3.2 Contrasting and Comparing Grid and Cloud Computing

As both concepts IaaS Cloud Computing and Grid Computing are mainly focused on offering hardware and software infrastructure services, analyzing their common features and their differences is needed to clearly separate the two concepts.

Table 2.4 shows a comparison between the usage characteristics of Grid Computing and IaaS Cloud Computing. Table 2.5 compares the main technical characteristics of Grid and IaaS Cloud Computing; it is based on the comparison in Buyya et al. (2009) and on the technical descriptions in Barroso and Hölzle (2009). Both tables mainly focus on the differences between both paradigms. As a result, both computing paradigms can be seen as separate approaches to a flexible resource usage.

According to the analysis in Franke et al. (2007) and Foster et al. (2008), current Grid technology mainly covers scientific applications and those applications in companies that resemble scientific calculations like calculation-intensive simulations (e. g. CAE calculations in the automobile development process); current Grid middleware is not adapted to accommodate general business computing needs. IaaS Cloud

Table 2.4: Usage characteristics of Grid and Cloud Computing

Characteristics	Grid Computing	Cloud Computing	References
Administrative Control of Resources	Decentralized	Centralized	
Funding	Mostly publicly funded	Privately funded by individual enterprises or consortia	e.g. European Grid initiatives (Bégin 2008, p. 6), Buyya et al. (2009)
Organizational impact	Virtual Organization	Contractual agreements among client and provider	Weinhardt et al. (2009)
Application characteristics	stateless, highly-parallelizable	stateful, serial processing	Franke et al. (2007), Foster et al. (2008)
Interactivity	little user interaction, batch processing	predominantly user interaction	Franke et al. (2007), Foster et al. (2008)
Data input and output	usually job-oriented data input and output	transaction-oriented, centralized data storage using separate RDBMS	Franke et al. (2007)
Relation to computing environment	few interfaces (e.g. license access)	numerous interfaces to existing business applications	Franke et al. (2007)
Operational characteristics	job runtime measured in hours or days	application runtime measured in years	Franke et al. (2007)
Access	Grid middleware layer (metascheduler, Grid services)	Web Services API	Weinhardt et al. (2009)
Failure handling	failure handling through job restart and checkpointing	high-availability, transactional integrity and ACID properties required; high-availability solutions installed	Buyya et al. (2009)
User-driven Monitoring	available as a part of the middleware	possible only for user application (no visibility of hardware level)	Foster et al. (2008)
User Management	Decentralized and also virtual organization (VO)-based	Centralized or can be delegated to third party	Buyya et al. (2009)
Pricing of services	Dominated by public good or privately assigned	Utility pricing, volume discounts available	Buyya et al. (2009)
Typical Use Cases	Collaborative scientific and high throughput computing applications	Dynamically provisioned legacy and Web applications, Content delivery	Buyya et al. (2009)

Table 2.5: Technical characteristics of Grid and Cloud Computing

Characteristics	Grid Computing	Cloud Computing	References
Hardware quality	High-end computers	Commodity hardware	Barroso and Hölzle (2009)
Operating Systems	Standard OS (often Linux/Unix)	Hypervisor as a host for multiple VMs	Buyya et al. (2009)
Internal Networking	Dedicated, high-end with low latency and high bandwidth	Oversubscribed commodity Ethernet	Barroso and Hölzle (2009)
External Networking	Mostly Internet	Mostly Internet	Buyya et al. (2009)
Resource Management	Distributed	Centralized, data center-oriented	Buyya et al. (2009)
Interoperability	OGF standards	provider-specific Web services	Buyya et al. (2009)
Failure Management	Limited (often failed tasks/applications are restarted)	Only on data center level; VM images can be restarted on another node.	Buyya et al. (2009)

Computing, however, is better suited for business applications due to its easy access and inherent flexibility. From an enterprise user's standpoint, the usage barriers seem to be considerably lower for Cloud than for Grid Computing.

Even as they have to be treated as technically separate approaches, it can be argued that both computing models offer similar opportunities to the enterprise that would potentially like to use or deploy Grid- or Cloud-based services. Not only are the opportunities comparable, but also the issues tend to be somewhat similar (even though details may differ):

- **User expectations:** Their users expect them to lower the cost of data processing, increase the reliability of the infrastructure layer and increase IT infrastructure flexibility by embracing an outsourcing approach to IT resource operations (IT resources are owned and operated, at least partially, by a third party) (Foster, Zhao, Raicu, and Lu 2008), (Bégin 2008, p. 4).
- **Scalability:** Both paradigms provide scalability beyond the means of even the largest in-house installations of dedicated compute clusters.

- Usage requirements: Another similarity is a common problem in Clouds and Grids. From a customer perspective, Cloud and Grid providers need to define methods by which consumers discover, request, and use resources provided by the central facilities; and to implement new software for the often highly parallel computations that execute on those resources, or to modify existing software to ensure compatibility to either one of these target platforms (according to (Foster, Zhao, Raicu, and Lu 2008)).
- Data management: Data mobility and locality will be the future challenges, that any Internet-based computing scheme has to cope with. The bottleneck in Internet computing is the network: moving data to the CPU is very expensive time-wise, especially as computing power has become ever cheaper over the last years and the amount of data has been ever increasing. Thus, data-aware scheduling and an intelligent distribution of data and applications (like Google's MapReduce (Dean and Ghemawat 2004)) are key functionalities for both Grid and Cloud Computing platforms (Foster, Zhao, Raicu, and Lu 2008).
- Sourcing decision: Internet Computing and Client Computing will coexist and evolve in parallel, however, the relationship between the two concepts is not clear-cut. There are criteria like data security, Internet availability or task-specific properties that have to be considered before a sourcing decision can be made. Those criteria are similar for both Grid and Cloud Computing (Foster, Zhao, Raicu, and Lu 2008).

As this comparison shows, Grid and Cloud Computing are two different computing paradigms, each specialized for a certain set of tasks. The uniting principles do not lie so much on the technical side, but rather on the usage side: user expectations, scalability, usage requirements, data management and sourcing decisions are all challenges that have to be faced by the potential user no matter what paradigm (Cloud or Grid) will be utilized.

2.3.3 IaaS Market Overview

When businesses decide to move infrastructure workload to an external Cloud Computing provider, it is essential to get an overview of potential market offers, especially when offers appear complex and obscure which reflects the situation in the current Cloud Computing infrastructure market. To overcome this lack in transparency, an empirical investigation of IaaS Cloud Computing offers was conducted in cooperation with Raimund Matros (University of Bayreuth) (Strebel and Matros 2010); the investigation took place as a market study and followed a combined desk-research and questionnaire approach. More details regarding the market study can also be found in Matros (2012).

First, relevant information was collected from corporate Web sites and press releases. This desk research phase took place in October 2009. If this information was incomplete or insufficient, a questionnaire was sent out to obtain the missing data from the company itself (see Matros (2012, p. 199) for the questionnaire). The questionnaires were sent and analyzed until end of November 2009; unfortunately, Matros (2012) does not give a response rate for the survey. The consciously selected target group consists of 61 companies which all offer Cloud Computing infrastructure services, according to their own publicly available information. In order to increase the comprehensiveness of the sample, it is based on earlier works of Baier, Gräfe, Jekal, Röhr, and Vörckel (2009), who also executed a market survey of IT infrastructure providers.

Surprisingly, the analyzed results reveal a great variability in Cloud offerings, although according to the Cloud Computing definition, these services are characterized by a high degree of standardization. Some providers did not meet the IaaS provider definition, although they labeled their offering as Cloud Computing, mostly because they limited their offers to certain types of IT infrastructure (e.g. Cloud storage) or

because they bundled their infrastructure service with an additional software solution (e.g. for backing up disks).

Another distinctive feature is the delivery model. When Public Cloud Computing providers make their offers available to the general public the services are delivered over the Internet and they are accessible to any customer with an Internet connection (Armbrust et al. 2009). The term private refers to internal datacenters or IT departments. 48 companies could be identified which offer their virtualized resources to the general public. The remaining 13 organizations are offering services within a private environment, e.g. LAN or dedicated WANs.

There is another compelling requirement when speaking of a Cloud: scalability (Nurmi et al. 2009; Hayes 2008). This feature is a defining property of any Cloud Computing service. Thus, it is astonishing that 41 observed organizations failed to meet this criterion (mostly because these service providers renamed their former hosting services as Cloud services without adopting distinctive Cloud features). They either offer fixed resource batches (e.g. virtual machine per month) or individual outsourcing agreements. Both groups have in common that there is no automatic resizing possible. When it comes to a situation of resource scarcity, agreements must be renegotiated in order to meet the new requirements. This is the traditional way of scaling up resources in outsourcing agreements (Wilson 1999). Real Cloud Computing providers create the illusion of infinite computing resources available on demand (Armbrust et al. 2009).

Eventually, the survey also asked for information about pricing schedules which is crucial for identifying the Cloud Computing target group. Three different pricing tariffs were observed: linear, mixed and individual. The linear tariff reflects the vision of utility computing. Consumers are only charged for their actual resource consumption. The mixed tariff is a combination of a linear and a nonlinear tariff and consists of two parts (setup fee and a linear component) or three parts (setup fee, flat rate tariff for a limited resource volume and a linear component); mixed tariffs also include fixed-fee tariffs that consist of flat rate costs and setup fees. As a result of this empirical investigation, an overview of Cloud providers' offerings was obtained. The findings are visualized in Figure 2.3. Finally, specific types of providers, who offered Cloud Computing infrastructure services at the time of the study, could be clustered and labeled with a type code.

Type A, Public Cloud provider: Customers who want to move workload to external Cloud providers need access without restrictions and with scalable resources. Two different pricing tariffs could be identified within this group: linear and mixed tariff. Public Cloud providers are the target group for the further research in this work.

Type B, Private Cloud provider: Providers who offer scalable Cloud Computing resources without providing service delivery over the public Internet are called Private Cloud providers. Normally consumers must conclude framework agreements to get access to these Cloud resources. Private Clouds offer customized services with customer-specific tariffs and reveal little of this information publicly, so a deeper analysis of these offering is not within the focus of this work.

Type C, Hosting provider: Hosting Providers normally offer dedicated or virtualized resources over the Internet. Their main handicap is the inflexible scalability of resources. Therefore they were not included in the group of Cloud providers.

Type D, Managed Services Provider: Providers who offer dedicated or virtualized resources for a private environment appear to be traditional outsourcing providers. Thus these suppliers fail to meet the criteria for IaaS Cloud providers.

Especially detailed tariff information could be gathered from Public IaaS Cloud providers. The sample included Amazon Webservices, AT&T, Elastic Hosts, Enki, FlexiScale, GoGrid, Joyent, Nirvanix,

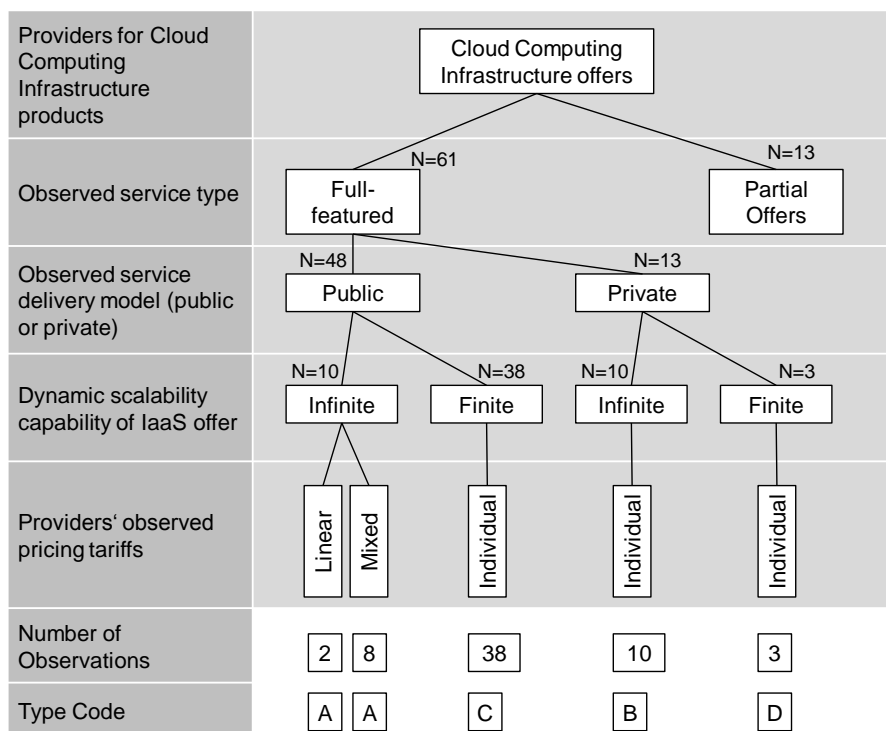


Figure 2.3: Typology of IaaS Cloud Computing providers

Rackspace and SliceHost (see (Matros 2012, p. 89) for the detailed result table). Although these IaaS providers offer a fairly standardized service, there is a large diversity in the tariff and billing conditions for those services. For example, compute services are commonly provided, but either billed per usage (CPU-h) or billed as VM instances (fixed fee depending on the instance size). Setup fees are usually not requested, but Joyent requires them. Also, minimum subscription times and the availability of volume discounts are areas, in which the provider offerings differ considerably.

As a conclusion, the IaaS Cloud Computing market exhibits a fair amount of intransparency; despite their technical similarity, provider offerings are not easily comparable, and vary both in their tariff structure and the service properties that go beyond the pure IT infrastructure level (e.g. data transfer security, monitoring options, support offerings). These findings are even more surprising, as the providers in the sample were specifically chosen for their Cloud infrastructure offerings.

2.3.4 IaaS and Outsourcing - Definitions and Comparison

Obviously, IaaS can be considered as a special type of outsourcing; therefore, the relationship between this new sourcing option and the established concept of outsourcing must be clarified. The following paragraphs will elucidate these two concepts.

Picot and Maier (1992) define IT outsourcing as follows: “the temporally limited or permanent transfer of information processing functions to external service providers.” Grover, Teng, and Cheon (1998) call IT outsourcing “an organizational decision to turn over part or all of an organization’s IS functions to external service providers in order for an organization to be able to achieve its goals.” TripleTree’s definition (TripleTree 2003) goes in the same direction:

“The transfer of operational responsibility of either business process or infrastructure management to an external service provider. The outsourced process or function is generally considered to be non-core in nature by the client, but the function can range from high volume,

repetitive processes such as electronic transaction processing to a more customized service such as technology help desk outsourcing.”

Usually, outsourcing is associated with a long-term contractual agreement, typically involving support services for IT operations. These services require the execution of tasks in the IT infrastructure or application management environment (see Kuchler (2004, p. 62)). According to the definitions given above, other simpler forms of outsourcing such as professional services (temporarily hiring specially qualified personnel e.g. for consulting tasks) or project-related contracting for work and labor (e.g. for custom software development) can also be considered valid instances of outsourcing, but the literature rather views them as border cases. In any case, outsourcing as a concept cannot be defined with scientific rigor, as an ongoing conceptual dilution has been taking place over the years through the multitude of services offered under the name of outsourcing (Bongard 1994). Hirschheim and Lacity (2000) and Willcocks, Lacity, and Cullen (2007) define types of outsourcing using the ratio of outsourced to insourced IT activities.

The following variants of IT outsourcing can be distinguished (Kuchler 2004, p. 62):

Selective Outsourcing also known as outtasking or partial outsourcing. Only a coherent subset of all IT services is outsourced. Selective outsourcing, the most common type of outsourcing, means that between 20% and 80% of the operating IT budget remains internal. If multiple outsourcing providers are utilized for different IT services, this is usually called multi-vendor outsourcing.

Full Outsourcing almost all IT-related services are covered by the outsourcing agreement. According to Willcocks, Lacity, and Cullen (2007), total outsourcing means that more than 80% of the IT budget is allocated to external providers for IT assets, leases, staff, management and delivery of services. Even at the level of full outsourcing, a minor IT function remains at the client company to manage the outsourcing provider and to retain the technical assessment competency (Bongard 1994, p. 132). The outsourcing provider, however, is usually free to subcontract parts of the services to other companies.

Full In-house Sourcing In-house sourcing means that more than 80% of the IT budget is still allocated to the internal IT department for IT assets, leases, staff, management and delivery of services (Willcocks, Lacity, and Cullen 2007).

Typical functions provided by outsourcing are Business Process outsourcing (BPO), Platform IT outsourcing, Application outsourcing and Systems and Network Infrastructure outsourcing. (TripleTree 2003, p. 30).

Business Process Outsourcing “Turning over responsibility for repetitive, well-defined processes to a third-party services provider. This definition encompasses both simple as well as complex business processes and both IT and non-IT processes, as well as varying degrees of customization.” (TripleTree 2003, p. 30)

Platform IT Outsourcing “The assumption of responsibility for managing all or part of a client’s information technology infrastructure, typically involving the transfer of IT facilities, staff, and hardware. Examples include hardware facilities management, onsite and offsite support services, server vaults, and data security services.” (TripleTree 2003, p. 30)

Application Outsourcing “providing management, maintenance, and support services for software applications. The outsourcing firm delivers application functionality via a remote hosted service and is responsible for maintaining a certain level of availability and functionality.” (TripleTree 2003, p. 30)

Systems and Network Infrastructure Outsourcing “Proactive provision and management of IT infrastructure and applications through a remote hosting environment.” (TripleTree 2003, p. 30)

Table 2.6: Outsourcing process

Phase	Activities	Key Issues
Preparation	Philosophy setting	Strategic decision
	Activity analysis	Activities identification (core/non-core)
	Outsourcing approach	big-bang, incremental or piecemeal
	Configurational arrangement	basic properties of the outsourcing relationship (No. of suppliers, duration, pricing, etc.)
Vendor selection	Creation of the RFP	Focus on objectives and results
	Determination of evaluation criteria	Development of mandatory, qualitative and cost based criteria for vendor assessment
	Evaluation and Selection of the vendor	
	Contract negotiations (and settlement)	Important contract properties: win-win orientation, flexibility, robustness
Transition	Change management	Transfer of assets, people, contracts, hardware and software, information and projects that the vendor will have responsibility for in the future
	Process integration	Rerouting of processes and integration of IT systems
Relationship Management	Performance monitoring and evaluation	effective communication
	Applying incentives (and penalties)	knowledge sharing
	Re-negotiating and managing variations	
	Handling meetings and communicating	
Reconsideration	Outsourcing performance evaluation	
	Choice of sourcing options	Decision between continuation of outsourcing relationship, change of outsourcing partner and back-sourcing

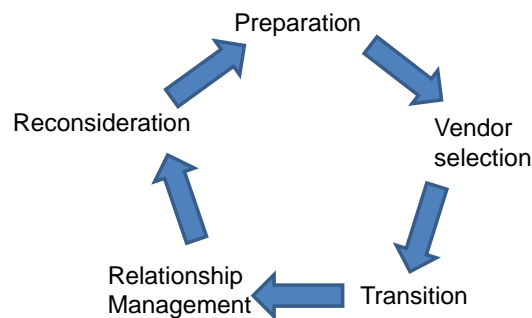


Figure 2.4: Outsourcing Process

The outsourcing process will be a major building block for all research activity in this thesis. As IaaS Cloud Computing can be viewed as a type of selective sourcing, the activities required to use IaaS have to resemble those activities usually found in selective sourcing arrangements and hence, can be captured by a generic outsourcing process definition. Moreover, using a process framework simplifies the description of dependency relationships among different outsourcing activities and guarantees a rather comprehensive analysis of the IaaS phenomenon.

The following definitions are mostly taken from Perunovic (2009), who distinguishes six defining perspective of outsourcing (Enablers, Types, Process, Theories, Outcomes, Vendors) and whose perspective of the outsourcing process shall be used in the further investigations.

Figure 2.4 shows a graphical representation of the five phases of Perunovic's aggregated outsourcing process; table 2.6 presents an overview of the corresponding activities and key issues in each of the phases (based on Perunovic (2009), Hameed (2010)). So far, the concepts displayed here are outsourcing-related, but not Cloud Computing-related. The application of the outsourcing process to Cloud-specific issues will take place in chapter 3 for qualitatively analyzable issues and chapter 5 for quantitatively analyzable issues.

In general, this thesis investigates issues related to the preparation and vendor selection phase of the outsourcing process. Later stages, that describe the operational execution of the outsourcing contract are not in the focus of this work.

Degree of outsourcing	Ownership		
	<i>Internal</i>	<i>Partial</i>	<i>External</i>
<i>Total</i>	Spin-offs (Wholly Owned Subsidiary)	Joint-Venture	Traditional Outsourcing
<i>Selective</i>			IaaS as Selective Sourcing
<i>None</i>	Insourcing / Backsourcing	Facilities Sharing among multiple clients	N/A

Figure 2.5: Positioning of IaaS in the outsourcing spectrum

Dibbern, Goles, Hirschheim, and Bandula (2004) conducted a survey and an analysis of the literature for Information Systems outsourcing in 2004. They categorized the possible types of sourcing arrangements, which can be seen in Figure 2.5. IaaS is considered to be a type of selective outsourcing activity. The term “selective outsourcing” was coined by Lacity et al. (1996), who argue against the all-or-nothing approach of outsourcing and describe situations in which differentiated outsourcing decisions are advisable. According to Dibbern, Goles, Hirschheim, and Bandula (2004), there are four fundamental parameters of any outsourcing relationship:

- degree (total, selective, none)
- mode (single vendor/client or multiple vendors/clients)
- ownership (totally owned by the company, partially owned, externally owned)
- time frame (short term, long term)

The characteristics of IaaS Cloud Computing can be mapped on those dimension. The degree is selective (always true for pure IaaS sourcing, as all other parts of the application stack remain insourced); the mode can be either single vendor/client or multiple vendors/clients, depending on the complexity of the Cloud service usage. The ownership of the IT resources is clearly externally owned for Public IaaS providers (Private IaaS providers may resort to company-owned resources, depending on the technical Cloud Computing setup; however, in this case, insourcing/backsourcing would have to be assumed). The time frame can either be short-term or long-term.

One of the main differences between the two concepts lies in the varying subject of analysis. Outsourcing contracts focus on business processes (BPO), IT management processes or other IT functions. They have a distinct business or functional orientation (see (Küchler 2004), (TripleTree 2003) for an overview of outsourcing activities). IaaS Cloud Computing, however, is focused on providing IT resources. Outsourcing contracts are settled on the level of business processes.

Service specificity is another criterion which distinguishes outsourcing services from Grid Computing services (Küchler 2004). Outsourcing contracts are highly individual to each client, even if the provided services themselves may consist of commodity tasks. Cloud Computing has a much more generic orientation according to the definition given above; its main tasks are service provisioning enabling the utilization of openly standardized services and resources.

The degree of organizational changes is another criterion that separates the two concepts. An outsourcing agreement usually foresees the transfer of assets, staff or facilities to the outsourcing provider, especially for larger outsourcing deals (Küchler 2004); the resulting financial relief is one of the reasons for outsourcing. When institutions use Cloud Computing, they usually do not transfer such resources to their computing provider.

The criterion of organizational change is connected to the criterion of the relationship between the client and the outsourcing provider. As outsourcing contracts are characterized by a high service specificity, the client enters a close relationship with the provider marked by the dependency of the client on the selected

outsourcing provider. Changing the provider would result in prohibitive switching costs. Cloud Computing suffers from this dependency dilemma to a lesser degree; initiatives like the Open Data Center Alliance⁹ are working towards a common set of standards. Higher standardization means less costs incurred for the client when switching Cloud Computing providers.

The most distinctive criterion, however, is the scope of the arrangement. Outsourcing contracts are based on exclusivity; the outsourcing provider takes over specific business activities or processes exclusively, so that the client can stop performing these tasks in-house. The outsourcing provider fully takes charge of the IT functions in question and the corporation cancels corresponding internal activities. As a result, the responsibility for the activities is transferred (Küchler 2004). IaaS Cloud Computing is not intrinsically focused on exclusivity; internal and external resources can be mixed and pooled. Every participating organization retains the responsibility for its resources and their management. As a conclusion, Cloud Computing can be subsumed as a special case of outsourcing: they share the principal inside-outside orientation and the transfer of responsibility. However, they also differ in major areas, such as the service specificity and exclusivity.

2.4 Literature Review

The related work can essentially be split in three major research areas, decision support models, technology adoption and outsourcing. Each area is split up in several topics; the related work is then clustered by research approach. For the purpose of presenting related work, Grid Computing and IaaS Cloud Computing is lumped together under the label “IT infrastructure outsourcing”. As both concepts are similar in many aspects, Grid research results are often directly relevant for similar IaaS Cloud Computing research questions as well. Table 2.7 lists the sources from the technology adoption domain; these references aim to give the scientific background for research question R1 (please see Table 1.3). Table 2.8 structures the references from the field of IaaS decision support models which are the basis of research questions R2 and R3. Table 2.9 shows the relevant references from the outsourcing domain. Vendor selection, one of the steps in the outsourcing process, lies at the intersection of both decision support approaches and outsourcing models. Hence, it is especially relevant to research question R2 and R3.

Table 2.7: Related work in technology adoption

Model Type	Topic	Research approach	References
Technology Adoption	Adoption of IT infrastructure outsourcing	Field study	Hwang and Park (2007), Baier (2008), Westhoff (2008), Heinle and Strebel (2010)
		Causal Model	Maqueira and Bruque (2006), Messerschmidt and Hinz (2013)
		Economic model	Thanos et al. (2007)
	Adoption of Application outsourcing	Causal Model	Yao (2004), Benlian and Buxmann (2009)

The following paragraphs are dedicated to the research literature on technology adoption. As mentioned before, Grid and Cloud Computing research papers were combined and form the basis for the IT infrastructure outsourcing concept. The only research paper known to the author on organizational innovativeness with respect to Grid technology adoption in businesses was written by Thanos, Courcoubetis, and Stamoulis (2007). It examines the economic factors determining the diffusion of Grid middleware and focuses on the economic forces acting upon it like network effects and market impacts. The research focus lies on the interaction among the Grid participants, each participant (e.g. an enterprise) is seen as a single unit of analysis (Rogers 2003, p. 407); the decision processes within the participating organizations were not analyzed. Thanos, Courcoubetis, and Stamoulis (2007) suggest three principal economically beneficial business scenarios viable for commercial Grid deployment:

⁹<http://www.opendatacenteralliance.org/>, last accessed 2013-12-29

1. Optimization of Processing Power in a Single Organization (inter organizational resource usage)
2. Sharing of Complementary Resources in Multi-provider Environments
3. Offering/Purchasing Utility Computing Services

It is then assumed that those economic forces influence decision to adopt one of the three business scenarios. Although this research paper discusses some of the factors influencing Grid adoption of enterprises, it does so from an overall Grid perspective; the decision process happens on an enterprise level and plays a minor role, so its findings are only marginally relevant for this thesis. However, one of the final research goals of Thanos et al. is:

... the definition of a decision model and associated methodology to be utilized by both Grid experts and business people for deciding towards the Grid adoption ... (Thanos, Courcoubetis, and Stamoulis 2007).

Maqueira and Bruque (2006) propose an adoption model which is based on TAM (Technology Acceptance Model) by Davis (1989). The Grid adoption model includes organizational and environmental factors and innovation characteristics. However, the author does not give any empirical evaluation of his model, so its explanatory power is unknown. Moreover, TAM was developed for measuring end user attitudes towards IT innovations; in the case of Grid Computing, it is not clear whether those end users are application software users, IT managers or IT operations staff. It remains unclear how Maqueira's model is supposed to be validated.

Hwang and Park (2007) identify the most important decision factors of Grid Computing solution providers for adopting Open Grid Computing technology, i. e. Grid software products for Open Grids. Their research is based on the assumption that an enterprise would want to introduce Open Grid-based solutions, and that this enterprise would then need support for the choice and the implementation of this technology. The identified decision factors seem plausible for the given scenario (e.g. usage of open standards, easiness of implementation), however, they were elicited from Korean solution providers only; the decision factors prevalent in enterprises considering Grid adoption might look different.

Baier (2008) bases her research on the technology diffusion model of Hall (2005) and distinguishes four dimensions relevant for technology diffusion (benefits, cost, network effects, information and uncertainty). After setting up her technology assessment framework, she uses Grid Computing for business information systems as a case study for evaluation. The analysis reveals two factors, observable benefit generation and data security, as the main challenges for Grid Computing. This study again approaches Grid Computing from a microeconomic angle and remains on a very high conceptual level. The six experts who gave their input for the study, all came from an academic background; this set-up limits the external validity of the results. Experts with a business background might have given more realistic input on the factors guiding Grid Computing adoption in the enterprise.

The study of Westhoff (2008) on the attitude of companies towards Grid Computing was conducted for the now-completed EU-funded SORMA research project. Its project goal was to develop an open market platform for Grid resources. Due to this background, the study tries to study the corporate requirements and influential factors towards Open Grids with market-based coordination of IT resources and with resource sharing among the participants. The data was collected using an explorative survey among experienced individuals within existing Grid research networks. Two specific issues became apparent as important results: Grid acceptance heavily depends on the management and evaluation of IT-related costs and the perceived security challenges within an open Grid environment. Grid Computing can be seen as a potential tool to reduce TCO without sacrificing QoS levels. Nevertheless, enterprises remain conservative towards

the potential of Grid Computing to reduce the TCO or increase the QoS. Moreover, security concerns hinder the deployment of open Grids tremendously. Although the survey explicitly targets the enterprise user, it also has severe limitations: only 24 participants took part in the survey, which is a rather small sample. All of the participants were actively involved in European Grid Computing research projects, so it can be assumed that there was a pro-Grid bias in the sample. Moreover, the study's Grid scenario, Open Grids with market-based coordination, does not exist in reality so far. Hence, it seems worthwhile to continue the research in this field in order to complement the results gained in this survey. First attempts of analyzing Cloud Computing enterprise usage determinants are found in the papers of Greenwood et al. (2010) and Kim et al. (2009); these conceptual works are still in their infancy and have to be considered research agendas and state-of-the-art overviews rather than full-featured research results.

Scholars have provided a sizable body of empirical research on technology adoption in related areas, such as Application Service Providing (ASP) (e.g. Yao (2004)) and SaaS (e.g. Benlian and Buxmann (2009)). General Cloud Computing adoption models are in their infancy and have little empirical support (e.g. Kim et al. (2009)). As a result, research explicitly focusing on IaaS acceptance and adoption has received little attention so far (for a detailed review see (Heinle and Strebel 2010)).

Table 2.8: Related work in IaaS decision support models

Model Type	Topic	Research approach	References
Decision Support Models	IT infrastructure outsourcing (cost focus)	Math. Optimization	Kenyon and Cheliotis (2004), Lilienthal (2013)
		Mathematical Modeling	Risch and Altmann (2008), Gray (2003)
		Mathematical Modeling / Optimization	Chaisiri et al. (2012), Chaisiri et al. (2009), Chaisiri et al. (2011), Van den Bossche et al. (2010), Zhang et al. (2009), Dastjerdi et al. (2011), Trummer et al. (2010)
	IT infrastructure outsourcing (multi-criteria)	Multi-Criteria Decision Making	Menzel et al. (2013), Khajeh-Hosseini et al. (2011), Khajeh-Hosseini et al. (2012)
	Optimized Resource usage	Linear Optimization	Rolia et al. (2003), Bichler et al. (2006), Almeida et al. (2006)
		Mathematical Modeling	Wimmer et al. (2006)
Vendor selection	Mathematical Modeling / Optimization	Weber et al. (1991), de Boer et al. (2001), Wadhwa and Ravindran (2007), Degraeve and Roodhooft (2000)	

Risch and Altmann (2008) analyzed a number of Grid Computing scenarios using a cost-based approach; they showed that Grid Computing is beneficial in scenarios, where either short and infrequent peaks have to be covered or where data backups have to be conducted or where lightly used resources have to be replaced. However, they recommend that each company performs its own cost analysis, as the benefits are depending on the cost level of the in-house resources.

Gray (2003) specifically deals with the decision when to outsource given the price ratios between the different computing resources. Generally, the business model behind Grid Computing remains case-specific; he maintains that business benefits are only realized for very CPU-intensive software applications.

Kenyon and Cheliotis (2004) addressed the area of Grid resource commercialization (also called utility computing). They conceive Grid resources as commodities and apply financial instruments for conventional commodities like gas or electricity to those Grid resources. Within the scope of their analysis, they identified the necessity for decision support, when Grid users buy or sell Grid resources on a Grid marketplace. However, the need for such elaborated decision support models will only arise, if a working Grid resource market similar to the existing markets of conventional commodities should ever exist, which is currently - despite research initiatives such as SORMA (Neumann, Stöber, Anandasivam, and Borissov 2007) and GridEcon (Altmann, Courcoubetis, Darlington, and Cohen 2007), not the case. Lilienthal (2013) also belongs to the category of mathematical optimization models for IT infrastructure with a cost focus. This paper assumes a continuous, horizontally scalable resource demand and supply for infrastructure services. Under these assumptions, the cost-optimal internal and external capacity can be determined as a function in

a mathematically closed form by deriving the underlying cost function. This model is very generic, however it requires a closed-form statistical workload distribution and is not specified for vertically scalable software solutions and for non-continuously scaled IT resource types.

Chaisiri et al. (2012) propose an OCRP (Optimal Cloud resource provisioning) algorithm, for which they assume different VM classes and consider the required number of VMs in each class to be random; also, prices can fluctuate randomly for these IT resources. The IT resources required by each application (an instance of one of the VM classes) are assumed to be fixed and certain. Their scenario seems most applicable to embarrassingly-parallel compute jobs or a pure scale-out approach. They optimize the total cost incurred for provisioning IaaS resources using a stochastic programming method. Chaisiri, Lee, and Niyato (2009) defines a similar problem, OVMP (Optimal virtual machine placement). The same assumptions hold as in the OCRP problem (see above), but the model is more carefully evaluation using numerical studies and simulation. Mark et al. (2011) solve the scenario with a solution approach based on prediction and evolutionary algorithms (EOVMP algorithm). Chaisiri, Kaewpuang, Lee, and Niyato (2011) considers the EC2 spot market and uses deterministic, stochastic programming, robust optimization and sample-average approximation (SAA) as solution approaches. Trummer et al. (2010) model a constrained-based selection of infrastructure services (constrained optimization problem). Their focus is on software components and the fulfillment of their technical requirements through Cloud services, but include cost in the constraint optimization problem. Their research concentrated on the efficiency of their solution method, as the evaluation included the running time for different number of components, different provisioning actions and different number of constraints.

Van den Bossche et al. (2010) describe a scenario for deadline-constrained jobs. Runtimes for each job on each instance type (i.e. VM class) are known beforehand. Uncertainty regarding runtimes and IT resource usage per job are excluded from the scope of the paper. Their scenario is characterized by fixed prices, no resource reservation and no discounting. Their evaluation used synthetic parameters for job-specific IT resource requirements and also concentrated on the efficiency of their model. Zhang et al. (2009) pronounce the difference between base and trespassing workload, but focus on a request-based workload model typical for Web traffic (e.g. HTTP requests, stateless applications). Typical use cases are CDNs. Dastjerdi et al. (2011) show how to model relevant QoS criteria, namely as latency, cost (data transfer cost, virtual unit, and appliance cost), and reliability for the selection of the best virtual appliances and units in a Cloud Computing environment, and present and evaluate two different selection approaches to help users in deploying a network of appliances on the different Clouds based on their QoS preferences. The model is suitable for multi-tier applications whose components are supposed to be distributed among different Cloud providers. The use case example consists of an e-Commerce application (Web Application server) and incoming user transaction requests.

As a conclusion, the related work on cost-based IT infrastructure outsourcing is only partially applicable to the research questions. Kenyon and Cheliotis (2004) assumes a market-based Grid environment, which might be relevant in the future. Risch and Altmann (2008) and Gray (2003) describe realistic outsourcing scenarios, but lack a clear mathematical model. As a contrast, the Operations Research-influenced optimization models feature an impressive array of mathematical methods, but fail to describe suitable conditions for IaaS usage.

Another set of decision support models employ multi-criteria decision making methods for arriving at reasonable outsourcing decisions Menzel, Schönherr, and Tai (2013), Khajeh-Hosseini, Greenwood, Smith, and Sommerville (2012), Khajeh-Hosseini, Sommerville, Bogaerts, and Teregowda (2011). All three research papers have in common that they elicit set of criteria (e.g. cost, security reliability, etc.) from the user and apply a multi-criteria decision making procedure (like AHP) to the resulting optimization problem. Although, this approach is more comprehensive than a cost-based decision, the added methodological complexity makes the results of these decision support models hard to evaluate objectively in reality. Therefore,

they can provide valuable methodological input for modeling the IaaS outsourcing decision (high internal validity), but they do not yield universally applicable recommendations (low external validity). Especially the results of the case study in Khajeh-Hosseini et al. (2012) support the approach for research question R2 as the case study shows “that running systems on the Cloud using a traditional ‘always on’ approach can be less cost effective, and the elastic nature of the Cloud has to be used to reduce costs. Therefore, decision makers have to model the variations in resource usage and their systems’ deployment options to obtain accurate cost estimates.” (Khajeh-Hosseini et al. 2012)

In the area of resource management, Rolia et al. (e.g. Rolia, Andrzejak, and Arlitt (2003)) suggest a resource-management framework for automatic software application placement in the data center using Grid-computing principles like resource allocation and scheduling. Their main focus lies on the optimization of in-house data-center resources, they do not address the question under which conditions to use external resources. Their optimization approach minimizes the number of CPUs and does not consider actual cost factors from an enterprise IT environment. Bichler et al. (2006), Wimmer et al. (2006) and Almeida et al. (2006) pursue the same goal and work under the same assumptions. Again, these results are only interesting from a modeling standpoint, and do not cover the complexities of the IaaS outsourcing scenario considered in this thesis.

Vendor selection is one of the steps in the outsourcing process and has attracted a fair share of research in the decision support field. For an overview of optimization approaches regarding vendor selection (e.g. multi-attribute utility theory, AHP), the review papers of Weber, Current, and Benton (1991), de Boer, Labro, and Morlacchi (2001), Wadhwa and Ravindran (2007) are a good starting point. Two authors, that specifically target the outsourcing decision are Degraeve and Roodhooft (2000). Their main focus is on the optimal decision process design, and the mathematical tools to reach it. The richest source of related work for research question R2 can be found in the purchasing and the operations research literature. Special attention in this area is justified, as the methods commonly used in vendor selection research are similarly applied in this work.

Table 2.9 gives a selection of cost-based optimization models used in the vendor selection literature that feature nonlinear price schedules. It also shows, that the current literature mainly focuses on nonlinear pricing schedules that are common for material procurement backgrounds; however, there is virtually no literature on either service procurement optimization or N-part tariffs in this area. Optimization problems resulting from bundling (e.g. in (Rosenthal, Zydiak, and Chaudhry 1995)) are not in the focus of this work.

Table 2.9: Cost-based Vendor selection approaches

Model background	Discounts used	Optimization method	References
EOQ (Economic Order Quantity) model with multiple items	incremental	Lagrange relaxation	Guder et al. (1994)
EOQ model with multiple items	all-units	Lagrange relaxation	Benton (1991), Pirkul and Aras (1985)
Cost minimization (N products, single time period)	all-units	MILP (Mixed Integer Linear Programming)	Sadrian and Yoon (1994)
Cost minimization (single product, single time period)	all-units, incremental	single-objective MILP	Chaudry et al. (1993)
TCO minimization (N products, M time periods)	all-units	single-objective MILP	Degraeve and Roodhooft (2000), Ghodsypoura and O’Brien (2001)
Cost minimization (single product, single time period)	concave cost function	Heuristic	Chauhan and Proth (2003), Burke et al. (2008)
Purchasing cost minimization (single product, single time period)	linear, incremental, all-units	Heuristics	Burke et al. (2008)
Cost, Delivery, Rejection rate optimization (single product, single time period)	all-units	multi-objective MILP	Weber and Current (1993)
Cost, Delivery, Quality optimization (N products, single time period)	incremental	various methods	Wadhwa and Ravindran (2007)

In conclusion, the existing research in decision support models is highly supportive for building outsourcing-related optimization models, but is not helpful for finding determinants of IaaS usage. The review of the related work reveals, that gaps concerning the research questions exist both in the area of decision support models and in the area of technology adoption. This thesis makes a first attempt to fill these gaps by analyzing IaaS-specific adoption determinants in an enterprise setting. IaaS is especially suitable as a research topic, as the relative technical homogeneity of the IT resources offered in Public Clouds makes it easier to compare decision factors across different enterprises. Moreover, IaaS deserves special attention, as IT infrastructure decisions are usually not strategic (Carr 2003), unlike for example decisions on software packages supporting business processes. Hence, this work argues that a dedicated study of determinants guiding an IaaS adoption decision is justified, as it is complementary to the existing SaaS- and PaaS-specific adoption research.

2.5 Discussion and Summary

This chapter lays down the conceptual framework for the further research by describing the supply side and the demand side of Cloud-based sourcing. It identifies those business software applications that act as demand drivers for Cloud resources and it introduces IT resource providers as suppliers. In order to precisely define the terms and concepts surrounding Cloud Computing, the characteristics of Cloud Computing and its types (IaaS, SaaS, PaaS) are compiled from the existing research literature. IaaS is then integrated into an outsourcing framework which clarifies the relationship between IaaS Cloud Computing and outsourcing. This normative research approach is complemented with a positive research approach (IaaS Cloud Computing market survey), which identifies four clusters of infrastructure offerings. One salient result of this market survey is, that the majority of the current IaaS providers do not adhere to the theoretical IaaS concepts and lack scalability and standardized tariffs in their offerings. However, the sampling method might contribute to this result, as meaningful data could primarily be obtained from Public IaaS providers (no random sampling). Both the sourcing aspects (supply and demand) and the outsourcing aspects then feed into the literature review. It discusses the research questions from three different perspectives; it can be concluded that both IaaS adoption determinants and the resulting decision support for IaaS deployments require additional research, but also offer chances for making a contribution to the state-of-the-art.

Part II

Empirical Research

Chapter 3

Functional Determinants for IaaS Sourcing

3.1 Introduction

The last chapter motivated the necessity of additional research concerning the adoption process of IaaS in enterprises and laid the conceptual foundation of the thesis. This chapter is primarily concerned with research question R1 as specified in section 1.1. The goal of this research question is to find the determinants of IaaS adoption in enterprises, hence the investigation needs to answer, why IaaS adoption takes place and what the determining factors are (i.e. what causes IaaS adoption?). As a first step to answer the question, a set of scientific hypotheses needs to be established which are empirically testable, generally valid, falsifiable and formulated using formally defined concepts (Bortz and Döring 2006, p. 4). In order to arrive at these hypotheses, a multi-method approach is chosen in this thesis, as documented in the following list:

- A case study in the automobile industry is supposed to give a first impression of the research problem. It follows the principles set forward by Eisenhardt (1989b), Eisenhardt and Graebner (2007) and can be classified as qualitative, explorative and inductive research. Although the context of the case study is Grid Computing, the underlying questions of external infrastructure utilization for BMW business applications are comparable to research question R1; hence, the answers should prove insightful for R1 as well. Section 3.2 describes this case study.
- The results from the case study are then challenged in a smaller field study, which methodologically follows Mayer (2008). Again, the research approach can be classified as qualitative, explorative and inductive. A set of hypotheses explaining IaaS adoption is derived. In section 3.3, both the methodological approach and the results of the expert interviews can be found.
- The results from the expert interviews and established information science research theories are then combined in a causal model, in which the known hypotheses describe causal relationships. The research approach here is quantitative, explanatory and deductive. This step also helps to validate the preceding qualitative work. The model is established in section 3.4.

The case study and the expert interviews are inductive in nature, as their goal is to generate testable hypotheses as a basis for further research. The causal model requires a deductive approach, as it tries to test the validity of the established and allegedly true hypotheses. The data collection of all three approaches happens in the field, which allows a much better and easier access to the needed informants for organizational research.

3.2 BMW Case Study

3.2.1 Description of the BMW Group

The BMW Group is a German automobile manufacturer, which also commands a sizeable market share in motorcycles; additionally, it offers financial services for both car-related and non-car-related transactions. The company was founded in 1916 as a maker of airplane engines, hence the spinning propeller as its emblem. Meanwhile, it has developed into one of the top premium automobile manufacturers in the world (BMW Group 2013, p. 18); its headquarters are located in Munich, Germany. The following figures and charts were taken from BMW's 2013 annual report (BMW Group 2013). The number of employees worldwide rose from around 96230 in 2009 to around 110350 in 2013 due to the recovery from the global economic crisis in 2008 and 2009. BMW's 2013 revenues were ca. 76 billion €, resulting from around 1.96 Mio. cars sold (BMW, MINI and Rolls-Royce brands). Figure 3.1 shows the number of plants and their location on the globe, forming a veritable global production network with full-featured production sites and specialized assembly plants, where semi-knocked down and completely knocked-down car kits are reassembled. The company focuses on the premium market sector, with its vehicle brands BMW, MINI and Rolls-Royce; apart from the automobile business, the motorcycle business also sold more than 115000 units in 2013. The predominant car model in sales is the BMW 3 series which alone accounts for ca. 30% of total sales. BMW's largest markets are China and USA which contribute almost 20% resp. more than 19% to the total sales volume in 2013. Despite the already strong sales volume in China, the Asian market still shows very strong growth rates (a 17.3% increase in 2013 alone).

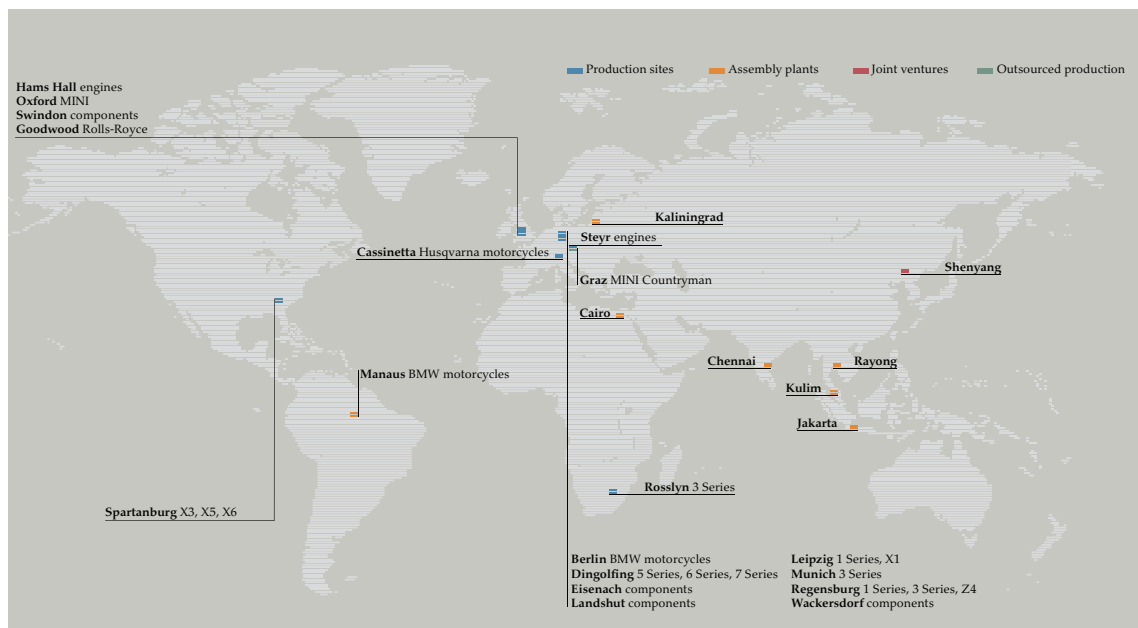


Figure 3.1: Global BMW production network (BMW Group 2011)

Figure 3.2 shows the setup of the central BMW IT organization at the start of 2014. The central IT function has a clear mission of delivering cost-efficient and standardized IT services and solutions for the whole BMW Group; it can be seen as an in-house IT service provider. Essentially, the BMW IT function is organized along the IT system development life cycle (SDLC) (Elliott 2004, p. 87): the business relationship management (BRM) discipline takes care of a preliminary analysis of the end user requirements which typically emerge in the business departments or in the BMW plants. This function is responsible for setting up and controlling the multi-project management framework necessary to conduct the following system development activities. The BRM function is split up in the following regions: the Americas, APAC (Asia-

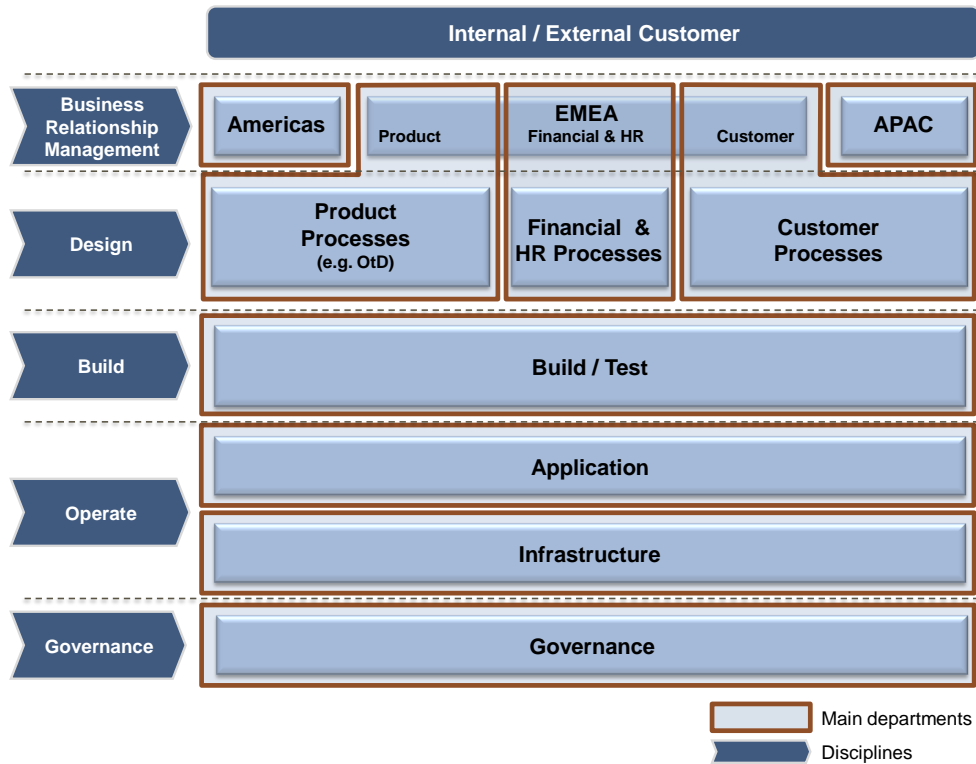


Figure 3.2: BMW IT Organization (BMW Group 2014b)

Pacific region) and EMEA (Europe, Middle East and Africa). The EMEA BRM function is stretched out over three main departments, one for product-related IT projects, one for customer-related IT projects and one for finance- and human resource-related IT projects.

The design discipline refines the analysis of the end user requirements and documents these requirements into corresponding system design specifications. This function is also usually responsible for IT project management and is organized along the main business processes (e.g. the Order-to-Delivery (OtD) process in product-related IT projects). Figure 3.3 shows a chart of the high-level BMW business processes. On the top level, the six main business processes are given. Additionally, the main business sub-processes of the Order-to-Delivery process are detailed. The Order-to-Delivery process encompasses all business activities from the reception of a customer order to the hand-over of the finished car to the physical distribution (activities include for example production control and planning, external parts ordering, paint shop, engine integration, final assembly, quality inspection); as the BMW Group predominantly follows a built-to-order philosophy, the customer order is the starting point of the manufacturing process. The OtD process and its associated business applications will be in the focus of the following investigations.

After receiving the system specifications from the design discipline, the build discipline launches the actual software development phase, where code generation, system integration and testing activities take place. The operate discipline then installs the IT system for production use and operates both the application software components and the hardware / infrastructure components of the IT system. The operate function is also responsible for the evaluation and the eventual disposal of the system, which concludes the SDLC.

The governance function is responsible for setting overall IT project quality and IT security standards and works in parallel to the original SDLC.

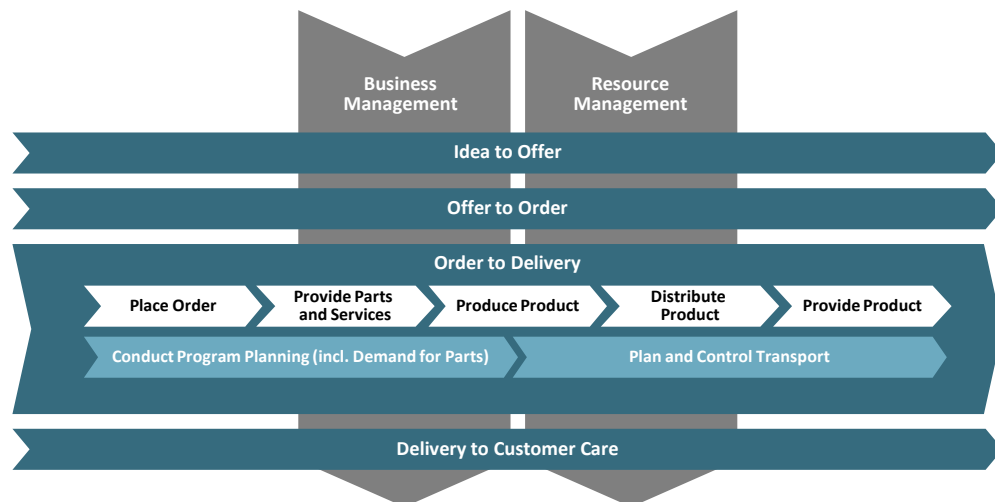


Figure 3.3: BMW Processes (BMW Group 2014a)

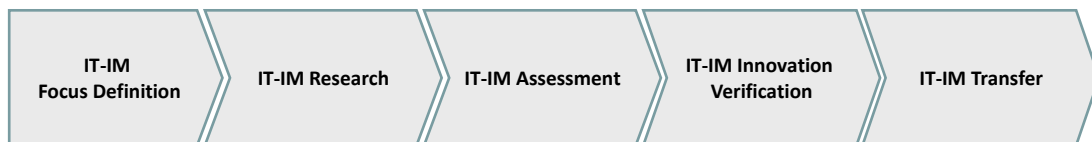


Figure 3.4: BMW IT Innovation Management Process (BMW Group 2008a)

3.2.2 Case Description

The assessment of possible Grid /IaaS sourcing options took place in the BMW IT innovation management context shown in Figure 3.4; this Figure shows the schematic steps of the IT Innovation management process in the IT department in place at the time of the case study (ca. 2008). Generally, it is supposed to be a lightweight process with short cycle time (a couple of months at most). During the focus definition step, the innovation topics are chosen in alignment with strategic business topics. In the research phase, the topic is detailed and compared to similar topics already in use within BMW and other automobile manufacturers; also, the topic's innovation potential is identified. In the assessment phase, the topic's benefits potential and its feasibility is assessed in cooperation with the business. If this assessment delivers a promising outcome, the topic is verified in the context of a feasibility study called 'Proof-of-Concept'. If this study confirms the assumed benefits, the topic is turned into a full-featured project, which is then added to BMW's project portfolio and prioritized in the context of the division-wide multi-project management. The IT Innovation management function provided the organizational context of the complete research work for this thesis. The above explanations and figures act as the background for the following case study on Grid /IaaS sourcing in enterprises. In 2007, BMW's IT department had identified Grid Computing as an IT innovation management topic. There was hope that this technology could make IT service delivery more resource-efficient.

One of the innovation management research projects was Biz2Grid,¹ which was federally-funded and was supposed to analyze how Grid middleware like the Globus Toolkit² can be used to take commercial applications to the Grid. Biz2Grid was part of a larger program, D-Grid, which was aimed at fostering Grid usage in Germany. BMW Group became an associated partner with the project.

¹<http://www.d-grid-gmbh.de/index.php?id=74>, last accessed 2013-12-29

²<http://toolkit.globus.org/toolkit/>, last accessed 2013-12-29

BMW chose three initial project scenarios, in which efficiency gains from Grid Computing seemed likely: a SAP ERP system, a material flow simulation called Plant Simulation sold by Siemens Tecnomatix and a suite of scenarios with a product development background (CAE tools). These three scenarios were BMW's input for the project and the assessment of those scenarios took place in the context of the Biz2Grid project.

From a research perspective, this setup offers a fascinating possibility: technology adoption can be studied in an enterprise environment; earlier Grid Computing approaches targeted only scientific environments (Buyya et al. 2009).

3.2.2.1 SAP ERP Scenario

The specific SAP ERP system under analysis is a SAP logistics solution based on SAP R\3 called STARD. This solution standardizes a number of processes in the areas of materials provision, plant maintenance, finance and processes for implementing the internal customer-supplier-relationships between car assembly plants and component-manufacturing plants in the BMW production network. The STARD system is deployed in six plants in the production network and consists of six SAP R\3 production systems, six quality testing systems and two development system. Additionally, there are two SAP SCM production systems, two SAP SCM testing systems and one SAP SCM development system. In total, there are eight SAP systems consisting of four to eight application servers each, which brings the number of servers to ca. 40-60. The complete landscape is shown in Figure 3.5 (as of 2009). As a first step in determining whether these SAP systems were suited for hybrid sourcing, a requirements analysis of those systems was conducted, i.e. the necessary preconditions for a Grid middleware were investigated, so that it can support the STARD landscape.

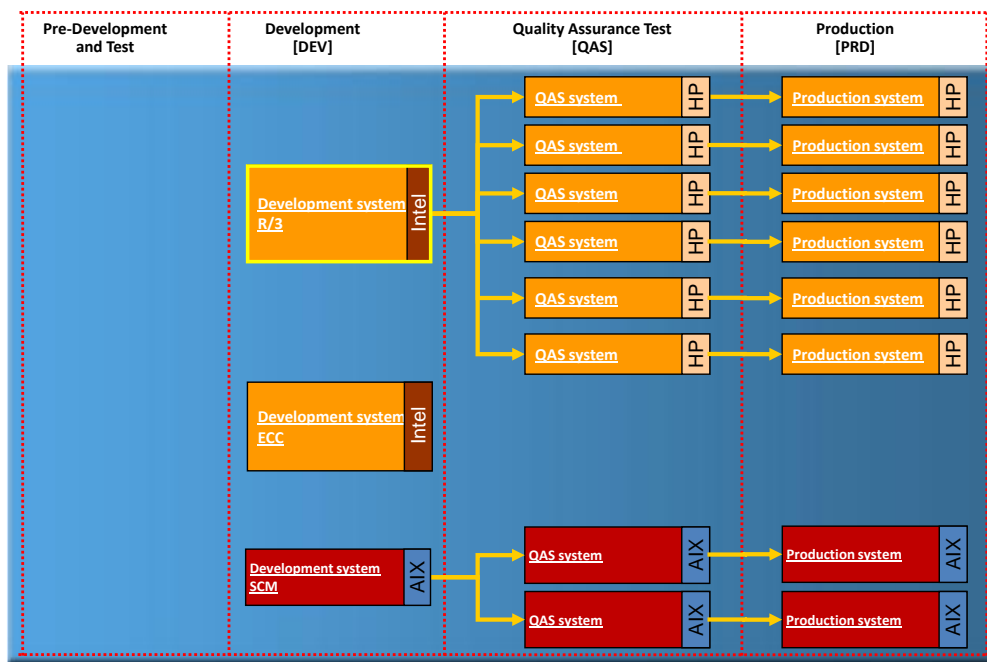


Figure 3.5: STARD system landscape (BMW Group 2008b)

3.2.2.2 Plant Simulation Scenario

Siemens Tecnomatix is the vendor of an event-based simulation software called Plant Simulation;³ this software tool is used throughout the whole automobile product development process for a wide variety of purposes, like intra-plant and inter-plant logistics, value flow analyses, production process simulations and improvements (like bottleneck identification, throughput / inventory / cycle time analyses). The BMW product development process is structured in several phases which are presented in Figure 3.6. Figure 3.7 shows the Plant Simulation GUI; the tool runs on the Microsoft Windows platform and is the BMW standard solution for logistics simulations. Simulation experts use it to graphically create object-oriented, hierarchical simulation models, that can later on be executed and analyzed; the simulation results, for example throughput or resource utilization data, can be analyzed using built-in diagrams and other visualization tools. It also includes an experiment manager module in which series of experiments can be defined and can automatically be executed.



Figure 3.6: Phases of the BMW product development process (BMW Group 2007)

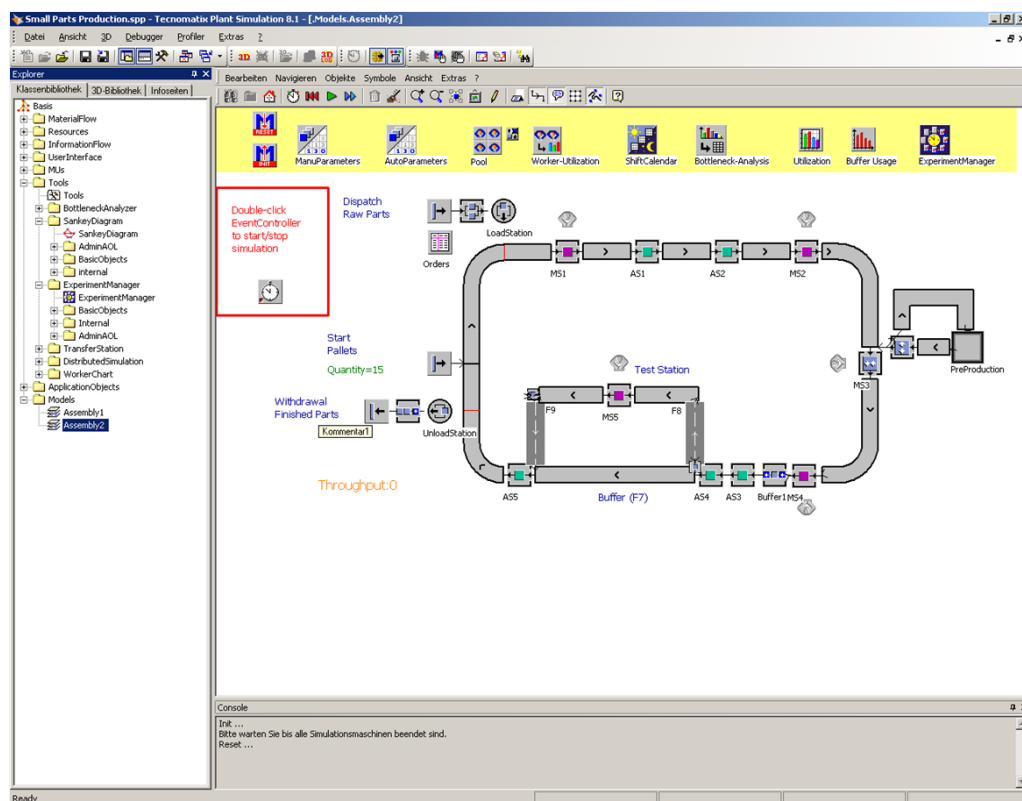


Figure 3.7: Plant Simulation GUI

In each of the product development process phases, specific development activities and tasks need to take place. In the initial phase, the fundamental task is the goal identification - the formulation of all requirements that the newly developed vehicle has to fulfill. During the concept phase, different possible technical solutions are developed, assessed, compared and selected. In the end, a coherent concept for the complete

³http://www.plm.automation.siemens.com/en_us/products/tecnomatix/plant_design/plant_simulation.shtml, last accessed 2013-12-29

vehicle is ready (+ prototyping). In the serial development phase, system, component and complete vehicle testing is executed with the goal of a complete virtual and real testing coverage. The potential use of simulations ranges from the preparation phase to the run phase. A simulation study (also called simulation project) contains a number of experiments, each of which consists of several simulation runs leading to one observation. The runs executed in the context of the experiments are independent from each other (neither data nor control dependencies exist); consequently, those runs are perfectly parallelizable.

A simulation study requires the actual simulation model and a number of different input parameter types. Some typical examples of those production process parameters are machine performance data (throughput), machine availability data, production programs, takt times, failure statistics, etc. Typical results are total system throughputs and work-in-progress, locations of bottlenecks and buffer sizings. Using this information, a material flow planner can eliminate redundant buffers, optimize inventory and reduce cycle times.

Some of the tasks during simulation projects are especially computation-intensive:

Statistical validation a stochastic simulation model will be fed with identical input parameter values but with differing seed parameters for the random number generators; this approach yields higher-quality simulation results with higher statistical support. During the time of the research project, a low number of simulation runs are executed for statistical validation due to computational and time constraints.

Input parameter variation The simulation model will be fed with differing input parameter values; each configuration of input parameters constitutes a simulation experiment. The computing time increases rapidly with the number of input parameters and the number of tested values for each parameter. Like in the statistical validation use case, only a limited number of simulation experiments can be executed.

The simulation project's results support the design decisions taken in the product development process. However, the simulation projects do not lie on the critical path of the product development process; the unavailability of simulation results does not hinder the decision-making in this process.

The Plant Simulation scenario has been chosen as a business setting because it exhibits several properties advantageous to Grid usage. First, each single simulation run is computationally intensive (up to several hours per run); second, the parallelizability of the runs makes them amenable to their simultaneous execution in a Grid Computing environment. Therefore, it was assumed that the positive impacts of Grid Computing would be especially visible and would lead to a seizable business case.

3.2.2.3 CAE Applications Scenario

CAE software applications play a vital role in the BMW product development process. Several exemplary challenges in the development process can be addressed by running numerical simulations on virtual car bodies:

- Complete Vehicle Safety: Figure 3.8 shows the visualization of an exemplary crash simulation load case.
- Component Safety: Figure 3.9 shows an exemplary load case of a pole impact.
- Rollover Safety: Figure 3.10 shows the simulation of an over roll scenario. Those additional load cases have emerged in the last couple of years due to legal obligations in the area of passive car safety.
- Aerodynamics: Figure 3.11 shows the air pressure profile for a car in motion. Those load cases become increasingly important for the design of future, fuel-efficient vehicles.

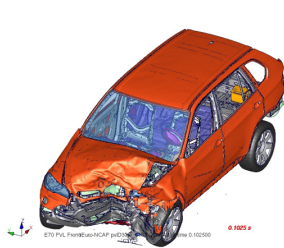


Figure 3.8: Crash simulations

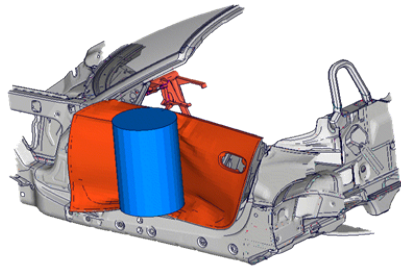


Figure 3.9: Component safety simulations

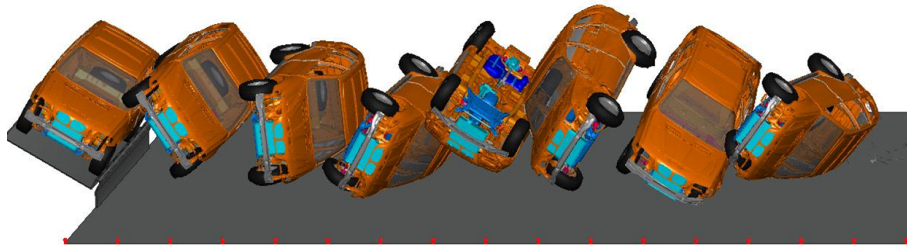


Figure 3.10: Rollover simulations

The figures and technical details were obtained from internal BMW documentation, except Figure 3.12 which has been created by the author.

The CAE load cases presented here are very compute-intensive. Figure 3.12 shows the running time distribution (in hours) of CAE jobs during one month. The data was collected in February 2009 using the scheduling protocols of BMW's HPC cluster. Typical CAE jobs run in parallel across several compute nodes; the calculations on each node require data from the other nodes of the same job, so the software processes on those compute nodes are tightly coupled and depend on a low-latency, high-bandwidth network connection for the ongoing intensive data exchange. It is clear that these jobs are also extremely CPU-intensive. As Figure 3.12 reveals, typical CAE jobs tend to have significant run times. Jobs running up to 8h only account for 10% of the utilized capacity, but for 72% of the job count. 45% of the capacity is utilized by jobs that run for more than 24h. This runtime distribution makes it evident that these jobs require dedicated hardware resources; solutions like Desktop Grids or shared server resources cannot accommodate these requirements.

It could be argued that the HPC cluster might not have been powerful enough, as suggested by the running time distribution, but the underlying hardware consisted of then state-of-the-art Intel x86 servers

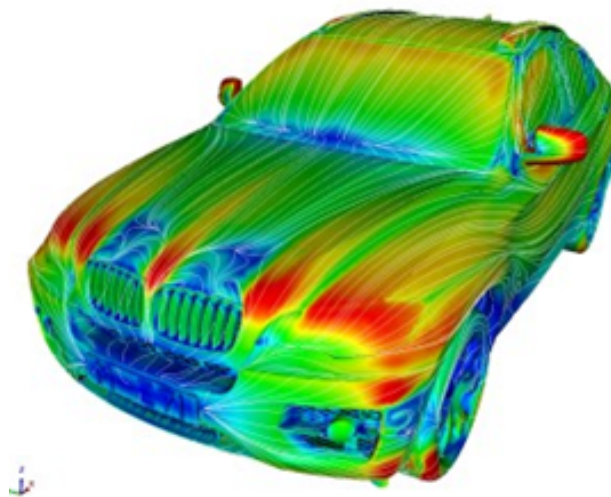


Figure 3.11: Aerodynamics simulations

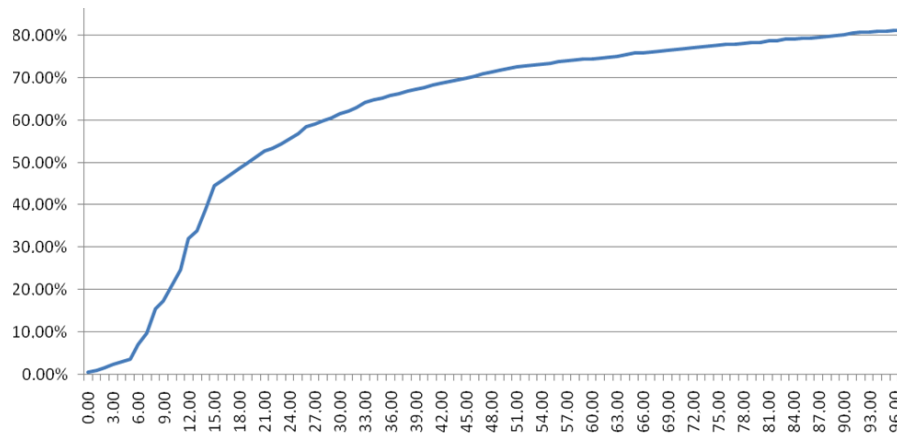


Figure 3.12: Runtime distribution of CAE jobs (in h)

connected through a Gigabit-Ethernet connection (47 TFLOPS Peak). Altogether, more than 15 CAE applications ran on 850 servers / 4 000 cores; the overall utilization target was 80%. 20 000-30 000 simulations were run per month in miscellaneous use cases and disciplines (as of February 2009).

3.2.3 Case Study Interviews

As the literature review in section 2.4 revealed, there are only a few relevant literature sources available for understanding flexible resource usage in enterprises. As the Biz2Grid research project was explicitly designed to facilitate the use of Grid Computing in businesses, the project setting was used to explore the determinants of flexible resource usage. A first approach was the diploma thesis of Kai Hachenberg in cooperation with the BMW Group and IBM Germany (see Hachenberg (2009) for further details). The results are presented in the following sections.

3.2.3.1 Research Methodology

The diploma thesis was designed to employ a qualitative and explorative/inductive research approach, as the topic at hand had only been superficially studied in the literature and hence, there were no specific, established IT infrastructure-related theories that could serve as a starting point for this research (except for general-purpose technology adoption-related theories). The goal of this study was to establish a first understanding of the phenomenon.

The data collection consisted of semi-structured expert interviews with potential Grid Computing users. From a Grid value chain perspective, the experts chosen for the interviews belong to the group of Business Users, a type of consumer (Altmann, Ion, and Mohammed 2007); the participating organizations do not offer any Grid services or provide consulting. As Grid Computing is still in its infancy in enterprises, the data gathered mainly reflects corporate attitudes and opinions on IT infrastructure outsourcing, rather than first-hand implementation experiences. In total, 12 participants took part in the survey.

The interviews were structured using an interview guideline consisting of a broad spectrum of questions that were supposed to cover the relevant aspects of Grid usage in enterprises. The input for the guideline was extracted from corresponding literature sources Fernández and Martrat (2008), Forge and Blackman (2006), Stockinger (2006), Schikuta et al. (2005), Jiménez-Peris et al. (2007). Other inputs were research papers and documents mentioned in the related work section 2.4. The complete interview guideline can be found in Appendix C. The interviews and the following analyses are based on methods and principles found in (Mayer 2008). The sample of expert interviewees was deliberately chosen for this case study; the resulting limitation in external validity is acceptable for a case study approach.

The experts were chosen because of their functional coverage of the areas affected by a potential Grid Computing introduction. Table 3.1 gives an overview of the available BMW experts. The expert selection was geared towards a full coverage of each scenario with each role. This goal could be realized for the CAE and the Plant Simulation scenario. There were no user experts questioned for the SAP scenario, as it was assumed, that the use of Grid Computing would be transparent for the user and that the system location did not matter for the user, if the system availability and the system responsiveness remained unaffected by Grid Computing. This assumption was later confirmed during the interviews.

Table 3.1: BMW Expert Map

	SAP scenario	CAE scenario	Plant Sim. scenario
User	n.a.	BMW4b	BMW4a
System Architect	XXX	BMW3	BMW7
Innovation Management	BMW5	BMW5	BMW5
IT Operations	BMW2a, BMW2b	BMW6	BMW1

As the BMW Group had little experience with Grid Computing, the research team turned to IBM Research GmbH, one of the other Biz2Grid project partners, for experts to be interviewed; these experts were supposed to provide additional knowledge from their Grid Computing implementations. The desired profile of the experts included experience in the deployment of Grid technology or comparable experience in related research areas like Cloud Computing. They were supposed to provide the researchers with their insights to the Grid Computing challenges outlined in the interview (the IBM experts are not listed in Table 3.1). The interviews were conducted from June 16th 2009 to June 22nd 2009 at the BMW IT headquarters in Munich by K. Hachenberg and the author. The interviews were face-to-face meetings, usually lasting about 45 minutes each; they were recorded on tape for later transcription. The expert BMW2a was interviewed together with BMW expert BMW2b. Their statements were summarized in one interview recording. Also, experts BMW4a and BMW4b were interviewed together and their statements were summarized.

3.2.3.2 Expert Profiles

The contents of this section follows Hachenberg (2009); the BMW expert profiles were prepared by the author, the IBM expert profiles were prepared by K. Hachenberg.

The expert IBM1 used to be a member of IBM Europa's Grid Team, but at the time of the interview, he was active in the field of Cloud Computing. His importance as an expert stems from his deployment to the Grid Team in which he acquired comprehensive experience in the area of Grid system development. During this time, solutions for industrial customers were thought up and developed almost up to marketability. Especially his experiences in introducing Grid Computing in enterprises are of high relevance in this interview. He was classified as a system architect.

The expert IBM2 is professionally occupied with Cloud Computing at IBM. Currently, he is responsible for the operations of an internal (private) Cloud solution. His importance as an expert stems from his function as an operations manager, as he could gather experience in Cloud requirements engineering in this role. This knowledge is relevant when analyzing Cloud Computing obstacles. He was classified as member of IT operations.

The expert IBM3 is mainly occupied with IBM data center management, application infrastructure deployment and its automation. Although this expert had had little exposure to Grid Computing-related problems, the researchers were confident that he could contribute significant insights into the challenges of Grid

Computing introduction especially from an organizational perspective due to his application infrastructure setup knowledge. He was classified as a system architect.

The expert BMW1 is mainly responsible for the application operations of the Siemens Tecnomatix Plant Simulation tool used within BMW (see also section 3.2.2.2). His relevance as an expert results from his professional responsibilities, which consist of planning and managing application deployment and the set-up of measures to ensure smooth application operations. In his cross-functional role in the central IT department, he also coordinates operations specialists and is responsible for designing and checking the application SLAs for compliance. As this tool is one of the candidates for Grid deployment, the expert's opinion and expectation on this technology are relevant for identifying possible integration obstacles. This expert was classified as a member of IT operations.

The expert BMW2a is responsible for the application operations of the large BMW SAP logistics system STARD (please see section 3.2.2.1). Similar to expert BMW1, BMW2a is also employed in a cross-functional role in the central IT department. He also has to ensure the smooth applications operations of the STARD systems. The STARD systems support the material management of a number of BMW plants, so their availability and performance is of vital importance to the company. SAP systems are also candidates for Grid deployment, but as their requirements are fundamentally different from those of the other candidates, BMW2a was a very relevant expert to the survey because of his expectations on Grid Computing and his opinion on the appropriateness of such a technology for ERP software. He was classified as a member of IT operations.

The expert BMW2b is a team leader for the specific infrastructure operations of the SAP platform. This expert was chosen, as he is responsible mainly for system administration, output management, user administration and interface configuration and monitoring in STARD. As these functions are vital to the operations of a SAP system, this expert's views on the technical requirements necessary for the integration of SAP in a Grid Computing environment are highly relevant for this study. He was classified as a member of IT operations.

The expert BMW3 is a team leader in the HPC architecture group, which is a part of the central IT solution design department. His responsibilities and his long-lasting track record of designing and operating high performance cluster systems make BMW3 an important interviewee. As an architect, he gathered experience in the operations and the billing mechanisms of shared IT resources; therefore, he possesses the necessary knowledge to realistically assess the requirements and problems of HPC cluster operations. The software applications run on these clusters are prime candidates for future Grid deployment, so his views are vital for this study. Moreover, the expert worked in the IT sector for years before joining BMW. He was classified as a systems architect.

The experts BMW4a and BMW4b are simulation tool users; BMW4a works with Siemens Tecnomatix Plant Simulation, BMW4b uses Exa PowerFLOW⁴ as a CFD solution. As long-term users of their respective software packages, both experts are highly relevant for capturing the user requirements of potential Grid-based software applications. Earlier research showed that user acceptance plays a major role in technology adoption (e.g. Rogers (2003)), so these experts' views need to be included in the study. Both users have been working with their tools for years and know the functional and non-functional requirements very well. These experts were classified as users.

The expert BMW5 is employed in the central, cross-functional IT pre-development.⁵ Due to this role, he was able to provide information on how BMW innovation assessment processes look like and how the deployment of new technologies is organized. This perspective is especially relevant for Grid Computing, as Open and even Enterprise Grids were completely new to BMW at the time of writing and hence consid-

⁴Exa produces simulation software for fluids engineering. Please see <http://www.exa.com>, last accessed 2013-12-29

⁵In the automobile industry, the pre-development stage serves as a preparation stage for the product development process of series vehicles. Pre-development is supposed to reduce the technological risk associated with technical innovations. IT pre-development serves the same purpose for innovative IT solutions that are destined for company-wide deployment in IT projects.

ered an innovation. Moreover, the expert has received his doctorate on fault-tolerant algorithms for Grid Computing, so he should therefore have a good understanding of the research area. He was classified as an innovation manager.

The expert BMW6 is responsible for the evolution of a number of CAE tools in use at BMW; he bundles functional change requests into software releases, assesses user-driven functional change requests and acts as a contact for the functional departments when new functional requirements become apparent. In this role, he possesses several years of professional experience; therefore, he should be able to judge whether Grid Computing is able to cover typical user requirements and expectations. From a functional perspective, he can tell the possible pain points of Grid Computing integration. He was classified as a member of IT operations.

The expert BMW7 is an IT architect in the central production IT department. Thanks to this background, he is able to formulate application development requirements as they relate to Grid Computing. As an architect, he already gained experience in implementing a Desktop Grid prototype at BMW together with IBM. He was classified as an architect.

3.2.3.3 Results

The contents of this section follows Hachenberg (2009), who analyzed the contents of the expert interviews and categorized similar opinions from the different interviews into general statements. Table 3.2 lists these statements and the supporting experts for each one.

Statement 1 questions the assumption that Grid users rationally choose the level of their resource usage. The economic theory predicts that they would act rationally if they had to balance usage against incurred cost. However, in enterprise settings, Grid users and budget owners are rarely identical; for example, a simulation expert might not even know the effective costs that the enterprise incurs for running his compute jobs. He mainly focuses on solving functional problems using simulations; the Grid allows to do it in a very effective and flexible way, as there are always sufficient IT resources available. In this situation, the budget owner, be it a first-line manager or a department head, has to have tools and processes that allow him to track and control the expenses incurred by Grid Computing in such a way that the maximum value is created for the enterprise.

Statement 2 questions the assumption that Grid resource usage can be planned using traditional capacity management schemes. These schemes usually determine the capacity of the following period by aggregating the capacity requirement forecasts of the individual departments; the central IT department then purchases those resources. Their quality and their price is known exactly beforehand. If the central IT department has to purchase Grid resources at some point in the future, neither their quality nor their price may be known beforehand.

Statement 3 suggests another obstacle for Grid Computing adoption. The reasons for the claimed cost intransparency are unknown initial setup cost, demand variations and additional effort for regular and ongoing cost accounting activities because of the utility pricing scheme of Grid resources.

Statement 4 confirms what Gray (2003) also shows in his research paper: the relevance of networking costs.

Statement 5 tackles the well-known security issues of Grid Computing. Corresponding standard processes will likely have to resemble vendor selection processes used in conventional outsourcing. In this context, it must be emphasized that most experts agreed on the fact that compliance with security standards like ISO/IEC 27001 or Sarbanes-Oxley Act/SAS70 can be valuable indicators for Grid provider security, but are not sufficient.

Statement 9 summarizes the experts' opinion on the organizational effects of Grid Computing on the IT organization. One of the concerns results from the similarity of Grid Computing and outsourcing, which

Table 3.2: Expert interview results

No.	Statement	Supporting experts
1	An improved model for demand management is needed to plan the resource demand and resource utilization of system users with limited rationality and competing goals.	BMW1, BMW2, BMW3, BMW4, BMW6
2	Current capacity management processes in enterprises are not suited for Grid / IaaS sourcing, where both resource quality and resource prices can vary widely.	BMW3, BMW5, IBM1
3	The commercial adoption of Grid Computing is hindered by lacking cost transparency during the planning and usage phase of Grid Computing services.	IBM1, BMW5
4	The current network communication fees reduce the number of potential use cases, that can adequately be addressed with Grid Computing.	BMW3
5	Clearly defined standard processes need to be established for assessing the risks of and the trust in potential Grid Computing providers.	BMW3, BMW5, BMW7
6	There is a need for improved server monitoring tools supporting distributed and virtualized IT resources. Resource monitoring is essential for ensuring the necessary availability and responsiveness of Grid-based business applications and for detecting SLA violations.	BMW2
7	Enterprises lack sufficient experience to quantify the business value of Grid Computing usage in financial terms. Non-monetary benefits are generally agreed on.	IBM1, IBM2, IBM3, BMW4, BMW5, BMW6
8	The preparation of current IT software applications and their interfaces for Grid deployment causes substantial effort (due to the interconnectedness of the systems).	BMW4, BMW5, BMW7
9	The adoption of Grid Computing is negatively affected by the uncertain effects of Grid Computing on the IT organization (e.g. role, budget changes, headcount, etc.)	IBM1, BMW5
10	Limiting license agreements of used commercial software obstruct the Grid deployment of such applications.	IBM1, IBM3, BMW1, BMW3, BMW4
11	Grid Computing can only be motivated through functional requirements of the business departments and has to be backed by management decisions.	BMW3, BMW4, BMW5, BMW6, IBM3

might lead to the transfer of former IT department activities to Grid providers; also, IT management processes (e.g. resource provision and acquisition) will likely have to be adapted when a Grid provider is used. The current IT organization might likely suffer from a loss of power in the organization.

Statement 11 emphasizes that the business functions play an essential role in ensuring Grid adoption, as they have to support the decision to put “their” applications on the Grid. In any case, management support is crucial, as the considerable risks created by Grid usage have to be accepted by the management.

When looking at the results in Table 3.2, it is striking that enterprise Grid adoption is mainly affected by organizational issues like IT governance, demand and capacity management, and the assessment of business value. In addition to the organizational issues, there are also legal, technical and cost-based problems, which makes the questions when to outsource a formidable one.

3.3 Qualitative Model of IaaS Usage Determinants

The uptake of infrastructure outsourcing for business applications in the enterprise environment is not without its challenges; among the most important ones are organizational issues, as it became clear in the preceding BMW case study (see section 3.2.3.3). This section presents an IaaS acceptance model based on multiple theoretical dimensions, which focuses on organizational drivers and barriers to IaaS deployment. It was challenged in a series of expert interviews and the resulting hypotheses give new insights into IaaS adoption drivers and represent a solid foundation for future empirical studies. The study and its results have already been published in a research paper (Heinle and Strebel 2010).

3.3.1 IaaS Adoption Model

3.3.1.1 Research Method

A qualitative research approach was chosen from the diverse range of available methods; qualitative interviews seem to be particularly suitable due to the explorative nature of the research question. In the context of this research, the guided expert interview was selected; this type of semi-structured interview is regarded as an important basic approach to collecting qualitative data (Bortz and Döring 2006, p. 308) and allows

for open-ended types of questions, as well as for intensive research using small sample counts (Bortz and Döring 2006, p. 381).

An interview guideline was prepared for the expert interviews; the contents of this guideline reflects the hypotheses derived in section 3.3.1.2. The interview guideline is supposed to serve as an orientation for the interviewer, such that no important detail is forgotten and such that a certain comparability among the collected data is ensured. The research literature considers guideline tests and test interviews as essential (Bortz and Döring 2006, p. 248); therefore, three test interviews were conducted and the interview guideline was reviewed through several external specialists. The resulting feedback was incorporated into the guideline. The final interview guideline can be found in Appendix D.

Expert interviews are not suited for a large number of survey participants, as the focus lies on the quality and the expressive power of each individual interview (Bortz and Döring 2006, p. 297). Research literature recommends sample sizes of 20 to 30 interviews (Meuser and Nagel 2009, p. 441). It was decided to focus on German enterprises or enterprises with German subsidiaries. German IT executives tend to be conservative towards Cloud Computing (e.g. (IDC Central Europe 2009)), so German experts should be especially aware of the potential obstacles to IaaS adoption. The actual experts were searched using the social business network XING,⁶ which lists the functional skills and the hierarchical position of each participant. Among the chosen experts were IT executives, IT consultants, lawyers specialized in outsourcing as well as research-oriented enterprises with an IT focus. Further search parameters included experience in the areas of Cloud Computing, Grid Computing, Virtualization, IT outsourcing and Data Center operations. In total, 215 potential interviewees were identified and invited, out of which ca. 50 experts were willing to participate in an interview. From those, 20 were selected for the actual interviews. They were invited electronically and they were informed that the telephone interviews would take about 20 minutes (as recommended by Bortz and Döring (2006, p. 242)); this relatively short time interval was chosen due to the assumed time pressure that these experts are facing. The actual minimum interview length was 15 minutes and the maximum length was 75 minutes; the average interview length was 35 minutes. This rather high variation can be explained by the use of open questions, the target of a free-flowing interview and the specific time pressure of some of the participants. The interviews were conducted during three weeks from April to May 2010.

In research literature, the recording of telephone interviews for later interpretation is perceived as indispensable. After the interviews, the written notes and the recorded talks were available for analysis (as recommended by Bortz and Döring (2006, p. 311)). Every interview was transcribed word by word using a transcription software for greater clarity in the later interpretation; the transcriptions were then anonymized to hide the experts' identity (see Bortz and Döring (2006, p. 313) for methodological questions).

The further analysis follows Meuser and Nagel's method (Meuser and Nagel 2009). In a first step, the interviews were paraphrased and then topically ordered. Each passage was associated with a specific headline. In the next step, passages from different interviews, that match topically, were selected and compiled; the associated headlines were harmonized. Then, a conceptualization step showed similarities and differences in the data. The final step, the theoretical generalization, consisted of the inclusion of theories and the arrangement of topics. This step is detailed in section 3.3.2, as it exhibits the actual results of the interviews.

3.3.1.2 Model design

Despite the strong media presence of general Cloud Computing, the IaaS concept is far from being clearly defined. Several sources give varying definitions (e.g. Vaquero et al. (2009), NIST (2009)). According to Hall, information is a determining factor for the diffusion of new technology; the choice of implementing

⁶<http://www.xing.de>, last accessed 2013-12-29

the technology requires knowledge about its existence and its applicability in an enterprise context (Hall 2005, p. 19). Therefore, hypothesis 1 is put forward: the unclear definition of IaaS negatively affects IaaS adoption propensity.

The impetus for new technology deployment (e.g. IaaS), mostly builds up in the IT-related areas of any enterprise, as these employees tend to have the technical background knowledge. Many business executives cannot position the current IaaS offerings correctly in respect to the conventional IT sourcing options due to the recency of this technology. Hence, innovation champions from the IT department are key success factors of IaaS adoption (see Rogers (2003, p. 414)). Hypothesis 2 states: IaaS Innovation champions in the IT departments positively affect IaaS adoption propensity.

As in every sourcing scenario, vendor selection is also an issue for enterprises planning to use IaaS; the challenge here is to assess and select suitable IaaS providers. Among the different provider attributes, absolute size (e.g. in terms of employees, turnover) acts as a signal for trustworthiness. It demonstrates that numerous clients can be served and is hence an expression of provider performance (Bensaou and Anderson 1999, p. 466). Relative size (in terms of market share) suggests the superiority of the services and resources offered by a specific provider (Doney and Cannon 1997, p. 38). Provider reputation is also considered a positive attribute (Doney and Cannon 1997, p. 38). Hypothesis 3 therefore assumes: Provider characteristics (relative and absolute size, positive reputation) positively affect IaaS adoption propensity.

Although provider reputation is an important concept for vendor selection, determining the reputation of a certain provider might be difficult and would require the usage of reputation measurement processes and methods. Hypothesis 4 states: Lacking processes for assessing provider risk and reputation negatively affect IaaS adoption propensity.

The processing of personal data of EU citizens on IT resources located out of EU territory is only permitted if complicated data protection regulations are followed (e.g. the Safe-Harbor treaty between the EU and the USA). Moreover, German data protection laws stipulate that a client has to have control over his data at any time during a commissioned data processing job (Meents 2010). This principle collides with the basic premise of IaaS, that the location of the data remains unknown to the client. Parrilli (2009) points out that enterprises enjoy almost no legal protection when using Cloud Computing; the business relationship between an enterprise user and an IaaS provider is solely depending on the contract that those two parties agree on (which might not be fair due to the information asymmetry at play here). Those legal challenges lead to hypothesis 5: the unfavorable legal situation and binding compliance requirements negatively affect IaaS adoption propensity.

Outsourcing of business data to an IaaS provider generally leads to a certain loss of control over data security and data protection. This problem was first investigated in the context of conventional IT outsourcing, and was identified as one of the biggest risks of this sourcing option (Barthélemy and Adsit 1993, p. 92); it is directly applicable to IaaS. Hypothesis 6 states: Data security and data protection issues surrounding IaaS negatively affect IaaS adoption propensity.

According to Armbrust et al. (2009) and Kim et al. (2009), clients' worries about the availability of externally purchased services are the biggest challenges to IaaS providers. Hence, hypothesis 7 states that concerns about service availability negatively affect IaaS adoption propensity.

Hypothesis 8 suggests that lacking IaaS monitoring and reporting solutions negatively affect IaaS adoption propensity. This assumption is informed by the results in (Hachenberg 2009) and is also based on the past situation in IaaS monitoring, which was in its infancy at the time of the study. There are public monitoring services (e.g. CloudSleuth⁷ or CloudClimate⁸). At the time of the study, those offerings were not yet comparable to or integrated into existing in-house monitoring tools.

⁷<https://cloudsleuth.net/>, last accessed 2013-12-29

⁸<http://www.cloudclimate.com/ec2-eu/>, last accessed 2013-12-29

Missing Application Programming Interfaces (APIs) and incompatible standards are still more the rule than the exception for IaaS offerings. Many providers use proprietary standards for their virtual machine containers and their APIs (Kim, Kim, Lee, and Lee 2009, p. 68). This situation leads to low interoperability among the IaaS providers and hence, current users can become locked-in with one provider. Buyya et al. perceive standardized interfaces of service offerings as indispensable for the success of IaaS Cloud Computing (Buyya, Yeo, and Venugopal 2008, p. 7). Hypothesis 9 therefore assumes that incompatible standards among IaaS providers negatively affect IaaS adoption propensity.

Price transparency exists if customers can acquire a clear, comprehensive and easily understandable overview of the service tariffs of a provider; especially the comparability of tariffs and benefits across different providers increases customer satisfaction (Diller and Herrmann 2003, p. 309). When looking at current IaaS offers, customers can hardly compare the individual offerings (e.g. Amazon's ECU (EC2 Compute Unit) vs. Rackspace's concept of guaranteed CPU core percentages). Although tariffs are known, these quality differences lead to price intransparency (Durkee 2010; Hachenberg 2009). Moreover, according to Gartner analysts, the Cloud Computing market is on its way from the past pioneer phase to a consolidated market (Driver 2008). This leads to hypothesis 10: the difficult cost-benefit evaluation of current IaaS offerings due to price intransparency and market immaturity negatively affects IaaS adoption propensity.

Egle et al. (2008) showed in an empirical study that the systematic management of IT costs hitherto only happens for hardware and software expenditures. Communication cost, personnel and operating expenses are often neglected (Egle et al. 2008, p. 12). As a result, enterprises can state the expenses incurred for certain IT services only very roughly. When sourcing external services, enterprises receive exact information on the costs. Thus, hypothesis 11 claims that increased internal cost transparency through IaaS usage positively affects IaaS adoption propensity.

Every new technology has to be adapted to the needs of the enterprise, but also, the structure of the enterprise has to be adapted to the new technology (Rogers 2003, p. 424). This adoption process, triggered by innovations, can cause uncertainty and resistance in the organization; the BMW case study exemplarily revealed some of the possible uncertainties that an organization might face (Hachenberg 2009). In the IaaS context, those actions especially affect employees in IT departments (Yanosky 2008, p. 134). Thus, hypothesis 12 suggests that the unknown organizational impact of IaaS introduction negatively affects IaaS adoption propensity.

The hypotheses detailed above can be firmly grounded in existing theoretical frameworks. Hypotheses 3, 4, 7, 8, 9 can be derived from agency theory (Logan 2000), especially from principal-agent-dynamics. Hypotheses 1, 2, 10, 11, 12 are grounded in the concepts of diffusion of innovation theory (Rogers 2003), especially as far as diffusion in organizations is concerned. IT governance theory (Weill and Ross 2004) is applicable to hypotheses 5 and 6, as they both address accountability and decision rights.

Other recognized theories, which are frequently used to explain SaaS adoption (e.g. in (Benlian and Buxmann 2009) or (Xin and Levina 2008)) are not applicable here for the following reasons. First, the transaction cost theory (Williamson 1985) is not particularly helpful, as the resource specificity of IaaS resources is so low, that only a market-based approach seems reasonable here. Second, the same line of reasoning can be applied to the resource-based view of the enterprise (Barney 1991). As argued earlier, for most enterprises, IT infrastructure resources can hardly be considered unique capabilities which offer a strategic advantage.

Figure 3.13 summarizes the hypotheses given above and acts as a research model for the following expert interviews.

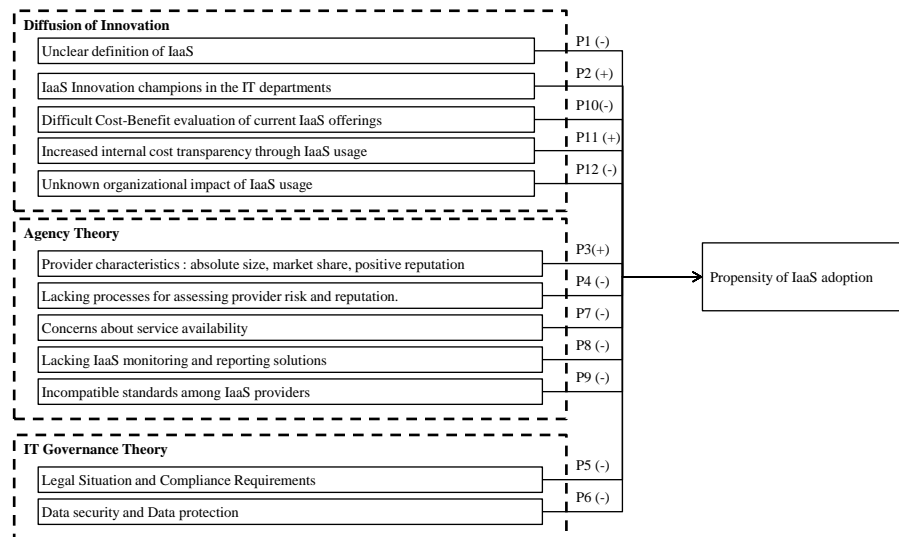


Figure 3.13: Research model of explorative study

3.3.2 Results

In this section, the summarized results of the expert interviews are given in Table 3.3; the interview transcripts were used to derive hypotheses which will be compared to the hypotheses from section 3.3.1.2 also in Table 3.3; column two shows the original hypothesis, column three shows the matching hypothesis derived from the expert interviews and the rightmost column shows the percentage of interviewees that supported this hypothesis.

Table 3.3: Summarized Interview Results

No.	Hypothesis	Matching Statements	Support of Hypothesis
1	Unclear definition of IaaS	Hypothesis supported	60%
2	IaaS innovation champions in the IT departments	The integration decision of IaaS has to be made according to the requirements of the functional departments and in accordance with management.	40%
3	Provider characteristics: size, market share, positive reputation, client references	Prov. characteristics: absolute size, positive reputation, references, further trust-building measures (certifications, data center tours)	55%
4	Lacking processes for assessing provider risk and reputation	Hypothesis supported	15%
5	Legal situation and compliance requirements	Hypothesis supported	40%
6	Data security and Data protection	Hypothesis supported	50%
7	Concerns about service availability	Hypothesis supported	25%
8	Lacking IaaS monitoring and reporting solutions	Hypothesis supported	25%
9	Incompatible standards among IaaS providers	Hypothesis supported	25%
10	Difficult cost-benefit evaluation of current IaaS offerings	Hypothesis supported	65%
11	Increased internal cost transparency through IaaS usage	Hypothesis supported	20%
12	Unknown organizational impact of IaaS usage	Hypothesis supported	40%

The interview results yielded another general statement in addition to the twelve ones described in Table 3.3. This statement can be summarized as follows: “Insufficient Service Level Agreements and Policies in case of lacking service availability negatively affect the introduction and the usage of IaaS”. The support of this statement is ca. 20%. This result specifically addresses the practically non-existent penalty payments in case of service failures.

The results presented in the last section give some interesting insights into IaaS adoption drivers. It becomes clear that the hypotheses developed in section 3.3.1.2 were on the whole comprehensive and generally supported by the experts. It is also a sign of external validity, that the general Cloud Computing

issues usually noted in surveys like data security and legal issues were also mentioned frequently by the interviewees of this IaaS-centered study. However, the usual potential IaaS benefits like infrastructure agility were not a common topic in the experts' answers, although they were asked for the possible rewards of Cloud Computing.

The most notable discrepancy between the hypotheses and the interview results becomes apparent in the area of IT-based innovation champions supposedly driving IaaS adoption. The experts interviewed tend to put the responsibility for the successful IaaS adoption in the hands of management executives, as they are the only ones that can decide to bear the risk of an IaaS implementation. IT departments tend to prepare IaaS adoption decisions according to functional business requirements, but they are not the IaaS innovation champions (perhaps they are more able than the business units to anticipate the possible organizational consequences).

The top three issues of IaaS adoption are difficult cost-benefit evaluations, the unclear definition of the IaaS concept and the importance of provider characteristics as decisive factors. Every new technology will have to prove its value sooner or later and IaaS makes no exception here (in accordance to Rogers' Diffusion of Innovation theory). More surprising is the fact, that conceptual difficulties surrounding IaaS prevent its success; this result has not been visible in other recent surveys on Cloud Computing (e.g. those mentioned in section 1). It seems that IaaS adoption depends on the dissemination of a clearer notion of the concept in organizations, especially to a non-technical audience. The third major issue is the provider-customer relationship, especially the creation of trust among the parties involved. Conventional and well understood trust-inducing signals as provider size and reputation are important, but further trust-building measures (like certifications e.g. ISO27001 or data center tours) are essential for IaaS adoption.

This research approach generated a number of interesting hypotheses, that warrant further investigation, especially the role of trust in the provider-client relationship and the role of the IT department in IaaS adoption. Those hypotheses form an ideal basis for a quantitative empirical research study as a next step.

The research method utilized here shows some inherent weaknesses. Telephone interviews are perceived as unfavorable in the research literature, because of their impersonal character and possible, uncontrollable circumstances (e.g. distractions) during the call (Bortz and Döring 2006, p. 242). Moreover, expert selection using a social network entails the risk of making poor choices, as qualifications of the social network participants are self-reported and cannot be verified. Another inherent weakness arises through the non-standardized IaaS nomenclature; therefore, a certain conceptual fuzziness is introduced into the investigation, as every expert probably had a slightly different notion of the IaaS concept.

This section aimed at investigating the organizational factors of IaaS adoption, as this issues is of high practical relevance and as the related scientific literature failed to answer this relevant question in a sufficient way so far. Towards this end, an IaaS adoption model containing adoption drivers and deterrents was proposed, based on an appropriate theoretical foundation. The model served as an input to rigorously planned expert interviews following scientific best-practices. The hypotheses of the adoption model were generally supported by the experts, however, the role of trust and executive involvement were underestimated in the initial hypotheses. As a result, trust between provider and client, transparency in IaaS offerings and conceptual clarity of IaaS can be assumed to be decisive issues for IaaS adoption.

3.4 Quantitative Model of IaaS Usage Determinants

This section presents and tests a set of hypotheses explaining corporate IaaS (Infrastructure-as-a-Service) acceptance and adoption (Wiedemann and Strebel 2011). With the market survey background given in section 1 in mind, it is surprising to find relatively few academic studies directed at the service model's basic pre-condition for success, i.e. IaaS acceptance in organizations. This part of the thesis tries to answer research question R1; to this end, a theoretical model is developed which is based on the Technology

Acceptance Model (TAM) (Davis 1989), the Theory of Reasoned Action (TRA) (Fishbein and Ajzen 1975), the Theory of Planned Behavior (TPB) (Ajzen 1991), the Transaction Cost Theory (TCT) (Dibbern, Goles, Hirschheim, and Bandula 2004), and the Principal Agent Theory (PAT) (Eisenhardt 1989a).

The research design can be described as quantitative, confirmatory research, as its aim is the empirical validation of a theory-based causal model. The results of Christian Heinle's research (Heinle 2010) and the literature review were used to formulate the underlying hypotheses. Those hypotheses are then tested using a quantitative approach (see Chapter 4 for the results).

3.4.1 Theoretical and Empirical Basis

Following the methodological recommendations in (Cheon, Grover, and Teng 1995; Bortz and Döring 2006), a multi-theoretical research approach for explaining corporate IaaS usage has been chosen. This section provides an overview of the theories derived from IS research and economics and motivate why these theories are appropriate for explaining IaaS acceptance and adoption. Moreover, the concepts used in the theoretical framework are defined.

The first theory is the "Technology Acceptance Model" (TAM) which is a multi-attribute model that predicts technology acceptance based on perceptions of user-friendliness and usefulness (Davis 1989). The model was chosen as an appropriate model in this study for three reasons. First, as IaaS is an information technology, the intentions to use IaaS and actual usage of IaaS should be explained in part by the TAM. Second, numerous empirical studies have shown that TAM is a robust model of technology acceptance behaviors in a wide variety of technologies and users (e.g. (Venkatesh, Morris, Davis, and Davis 2003)). (Legris, Ingham, and Colletette 2003) (Venkatesh, Morris, Davis, and Davis 2003) Third, past outsourcing research has considered the influence of decision maker attitude toward outsourcing on their decision (Benamati and Rajkumar 2002). Fourth, IaaS services are functional services adopted for utilitarian reasons (e.g. (Heinle and Strebel 2010)).

The TAM includes five concepts. Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" and perceived ease of use is defined as "the degree to which a person believes that using a particular system would be free of effort" (Davis 1989, p. 320). Both concepts influence one's attitude toward system usage, which influences one's behavioral intention to use a system, which, in turn, determines actual system usage. Attitude toward use is referred to as "an individual's positive or negative feelings (evaluative affect) about performing the target behavior" (Fishbein and Ajzen 1975, p. 216). Intention to use is based on Fishbein's and Ajzen's definition of behavioral intention: "the strength of one's intention to perform a specified behavior" (Fishbein and Ajzen 1975, p. 288).

The second theory is the "Theory of Reasoned Action" (TRA). Several studies have emphasized the importance of social influences on technology usage in general (e.g. (Venkatesh et al. 2003), (Venkatesh and Davis 2000); (Venkatesh, Morris, Davis, and Davis 2003)). According to (Dibbern 2003), IT sourcing and adoption are management decisions made by individuals rather than by organizations. According to (Benlian and Buxmann 2009) and (Xin and Levina 2008), IT executives are influenced by their social environment. Thus, the inclusion of individual's thoughts and feelings affected by other people represents an important addition to the model at hand to fill the gap related to the effects of social influence. The TRA (Fishbein and Ajzen 1975) captures the effects of normative pressure from outside and inside the company. The TRA model includes four general concepts: subjective norm, attitude toward the behavior, usage intention, and actual use. The subjective norm is defined as "the person's perception that most people who are important to him think he should or should not perform the behavior in question" (Fishbein and Ajzen 1975, p. 302).

The third theory is the “Theory of Planned Behaviour” (TPB). IaaS has some notable differences compared to traditional in-house infrastructure. First, despite the strong media presence, the IaaS concept is far from being clearly defined and understood (Heinle and Strebel 2010). However, information is a determining factor for the diffusion of new technology. Hall (2005) argues that the choice of implementing the technology requires knowledge about its existence and its applicability in enterprises. Second, IaaS providers deliver services through the Internet, which increases uncertainty about availability, reliability, response time, data security, and data protection (Heinle and Strebel 2010). Third, as IaaS sourcing causes a new legal situation (as compared to traditional in-house infrastructure sourcing), other binding compliance requirements have to be met (Parrilli 2009), (Heinle and Strebel 2010). These differences reduce decision makers’ perception of control, confidence, and effortlessness over outsourcing activities in terms of IaaS, creating a barrier to IaaS acceptance and adoption. Therefore, perceived behavioral control, as described in the TPB (Ajzen 1991) (an extension of the TRA), is likely to play a critical role in IaaS.

Perceived behavioral control is defined as “people’s perception of the ease or difficulty of performing the behavior of interest” (Ajzen 1991, p. 183). The well-researched TPB has been one of the most influential theories in explaining and predicting behavior across a variety of settings (Sheppard, Hartwick, and Warshaw 1988). Thus, it is reasonable to expect that a model integrating TPB could explain decision-making behavior towards IaaS.

The fourth theory is the “Transaction Cost Theory” (TCT); it was developed originally by (Williamson 1985) and is heavily used in the outsourcing research field for explaining the boundaries of firms and the dynamics of exchange relationships among business partners (Dibbern 2003). Therefore, TCT has been used for studying ASP- and SaaS-related IT sourcing models (e.g. (Benlian and Buxmann 2009), (Yao 2004), (Susarla, Barua, and Whinston 2009)). There are two reasons why TCT is a suitable theory in the IaaS context. First, IaaS is centered on commodity IT resources like storage and servers; as a result, questions whether asset specificity can be assumed for those resources and whether the concept is relevant in IaaS decision making, need to be evaluated. Second, IaaS shares some of the traits of outsourcing (e.g. external resource usage); as a result, it seems reasonable to utilize TCT in the IaaS context. Two of TCT’s main concepts are uncertainty and asset specificity (David and Han 2004). “Uncertainty refers to the volatility of the environment that cannot be anticipated.” (Nam, Rajagopalan, Rao, and Chaudhury 1996). Uncertainty is a multi-faceted concept and, as a result, is measured in variety of ways (David and Han 2004). This thesis follows (Benlian and Buxmann 2009) and (Susarla, Barua, and Whinston 2009) who consider business, technological, legal and transactional uncertainty. Business uncertainty reflects the concerns that business processes may be negatively affected by the outsourcing relationship. Technology-driven uncertainty captures the degree to which the required technical functions or features of the outsourced infrastructure change over time. (Benlian and Buxmann 2009). Legal uncertainty describes the effects of lacking or country-specific laws governing Cloud services (e.g. jurisdiction (Parrilli 2009)). According to (Susarla, Barua, and Whinston 2009), uncertainty also results from the transactional environment (e.g. uncertainty about service cost or the IaaS implementation period). This thesis combines these uncertainty measures, similar to (Benlian and Buxmann 2009) and (Susarla, Barua, and Whinston 2009).

Another important concept of TCT is asset specificity; “asset specificity is a key issue in the transaction-cost-based view of business relationships, referring to the degree to which an asset can be redeployed to alternative uses and by alternative users without sacrificing productive value.” (Nam, Rajagopalan, Rao, and Chaudhury 1996) Regarding outsourcing decisions, the TCT assumes that a low level of asset specificity results in market-based sourcing of assets (i.e. outsourcing), whereas a high level of asset specificity results in in-house production of these assets. In the context of IaaS, asset specificity is considered as “infrastructure specificity” and is defined as the degree the corporate infrastructure solutions are proprietary. This specificity can be estimated by analyzing how the IaaS offerings would have to be altered to fit the organizational requirements and how much organization-specific technical and business knowledge would

have to be established at a potential IaaS provider for him to successfully support business processes with IT resources (Benlian and Buxmann 2009), (Yao 2004).

The Principal Agent Theory (PAT) considers enterprises as a collection of contractual relationships between principals and agents. As stated by (Eisenhardt 1989a) “one party (the principal) delegates work to another (the agent), who performs that work”. One major theoretical assumption of PAT is the existence of asymmetric information between the principal and the agents. As the agents are supposed to be better able to judge the risk involved with the work (Dibbern, Goles, Hirschheim, and Bandula 2004) opportunistic behavior may occur on the part of the agents. The PAT is applicable to IaaS sourcing decisions, as corporate decision makers (the principals) face IaaS providers (the agents) with superior knowledge of their operations and the associated quality of service offered in terms of availability and data security (see (Durkee 2010) for examples of provider’s shirking). Under these circumstances and in the absence of effective control mechanisms, IaaS providers may be tempted to act opportunistically to pursue their own self-interests. Thus, it is assumed that the corporate decision makers’ fear of provider opportunism influences their behavior towards IaaS adoption. Moreover, Pavlou et al. (2007) showed that this concept is helpful in explaining electronic commerce adoption.

One of the biggest and mostly cited concerns about IaaS and Cloud Computing in general are data security issues (e.g. (CIO Magazine 2009)). Pavlou et al. (2007) introduced information security concerns as an important concept for studying buyer-supplier relationships. “Information security concerns are defined as the buyer’s beliefs about a seller’s inability and unwillingness to safeguard their monetary information from security breaches during transmission and storage” (Pavlou, Liang, and Xue 2007). In the context of IaaS, the concept of information security includes monetary information as well as business data, e.g. from human resources or customer relationship management applications. In the IaaS provider-client relationship, the potential client cannot accurately judge the quality of the provider’s data protection schemes beforehand due to the information asymmetry (Cheon, Grover, and Teng 1995), (Eisenhardt 1989a) between the IaaS provider and the client and due to the black-box nature of IaaS.

Table 3.4 shows the mapping of the hypotheses gathered from the expert interviews and the causal model constructs as explained above. Each bullet point indicates the relevance of an expert’s hypothesis for the associated causal model construct.

As a sixth theory, Modheji (2010) developed in his bachelor’s thesis a quality model for public IaaS offers. The six quality dimensions of this model were the basis for six e-Service quality-related questions, which have the goal of determining the importance of each of the quality dimensions. The six dimensions are appropriate service level agreements, support of corporate IT operations processes by the IaaS provider, adherence to security and compliance regulations, transparency and practicability of provider tariffs, the possibility of drafting individual contracts (e.g. duration, type, penalties...) and the quality of customer service with the IaaS provider. These concepts are evaluated along with the other hypotheses, but they are not part of the causal model.

3.4.2 Hypotheses and Causal Model

Following the structure put forward by Dibbern, Goles, Hirschheim, and Bandula (2004), the individual level is addressed by analyzing the motivations, preferences or attitudes of individuals and their impact on the IaaS outsourcing decision. Thus, this paragraph shows the effects on the decision maker’s attitude towards use IaaS based on the theories TAM, TRA, and TPB. A main point in multi-attribute models like the TRA or TPB is that the evaluation of salient beliefs about a product or service directly affects the overall attitude toward the product or service usage (Fishbein and Ajzen 1975). Decision makers tend to have a positive attitude toward services associated with characteristics that they perceive to be good, and vice versa (Nysveen, Pedersen, and Thorbjørnsen 2005). Thus, decision makers’ beliefs about an infrastructure

Table 3.4: Connection between Hypotheses and Causal Model Constructs

	Causal model constructs								
	Infrastructure Specificity	Perceived Uncertainty	Attitude towards IaaS Usage	Subjective Norm	Perceived Beh. Control	Information Sec. Concerns	Fear of Provider Opportunism	Perceived Usefulness	Ease of Use
Hypotheses from Expert Interviews									
Unclear definition of IaaS			•						
The integration decision of IaaS has to be made according to the requirements of the functional departments in accordance with management.					•				
Provider characteristics: absolute size, positive reputation, references, further trust-building measures (certifications, data center tours)		•	•						
Lacking processes for assessing provider risk and reputation						•	•		
Legal situation and compliance requirements						•			
Data security and data protection						•			
Concerns about service availability		•					•		
Lacking IaaS monitoring and reporting solutions		•							
Incompatible standards among IaaS providers	•	•							•
Difficult cost-benefit evaluation of current IaaS offerings							•	•	
Increased internal cost transparency through IaaS usage								•	
Unknown organizational impact of IaaS usage		•		•					

services' usefulness and ease of use should positively influence their attitude toward using IaaS. Although the direct effect of beliefs (i.e. perceived usefulness and perceived ease of use) on behavioral intention is not included in the TRA and TPB, such effects are theoretically justified in the TAM and other intention models (Bagozzi 1982) and empirically confirmed in several studies (e.g. (Venkatesh and Davis 2000)).

According to Bazijanec, Pousttchi, and Turowski (2004), the usefulness of an electronic solution may be based on efficiency added values (benefits through an increase of operating efficiency and cost effectiveness), effectiveness added values (benefits through an augmentation in output quality), flexible added values (benefits through creation of a higher level of flexibility), organizational added values (benefits through new forms of organization), innovative added values (benefits through entirely new products or services), and strategic added values (benefits through significant competitive advantage).⁹ Outsourcing success is usually viewed as the outcome of a mix among economic, technological, or business-related benefits. Literature shows that cost savings being a dominant factor for organizations when deciding to outsource and evaluating the outcomes afterwards (Dibbern, Goles, Hirschheim, and Bandula 2004). Thus, Hypothesis 1 (H1) can be formulated as follows: the decision maker will have a more positive attitude toward using IaaS, if the perceived usefulness and the perceived ease of use of IaaS are rated more positively by him.

This paragraph hypothesizes effects on perceived uncertainty based on the theories explained above. As the IaaS sourcing relationship entails the potential of moral hazard (Eisenhardt 1989a) for the IaaS provider, fears of seller opportunism are proposed to increase uncertainty since buyers are normally unable to post-contractually control or enforce provider behavior (Pavlou, Liang, and Xue 2007). Moral hazard considerations are also applicable when assessing information security. Information security concerns lead to uncertainty, which stems from the buyers' difficulty in assessing a provider's ability to safeguard vital business information. Moreover, buyers also have to bear the uncertain consequences of a potential security breach at the provider's facilities, which may result in financial problems in the future (Pavlou, Liang, and Xue 2007). Also, it can be argued that IT infrastructure specificity may lead to a higher perceived uncertainty as the potential IaaS users may be in doubt whether generic IaaS resources are able to fulfill their infrastructure requirements. The continually evolving Cloud market and its technical progress could contribute to this notion. Based on this discussion, Hypothesis 2 (H2) is formulated as follows: the decision maker will experience more perceived uncertainty, if his or her fear of provider opportunism, information security concerns and infrastructure specificity are more elevated.

This paragraph hypothesizes effects on intention to use. According to the TAM, the effect of ease of use on behavioral intention is proposed to be mediated by attitude toward behavior (e.g. (Taylor and Todd 1995)). In general, people want to behave in ways that are in accordance with their attitude (Fishbein and Ajzen 1975). Therefore, the assumption is justified that the attitude mediates the effect on behavioral intention towards IaaS usage. Moreover, positive effects of attitude towards a behavior on behavioral intention have been empirically shown in several IS studies (Legris, Ingham, and Collerette 2003). According to the TRA and TPB, the subjective norm is postulated to have a positive effect on intention to use IaaS. Similar to (Nysveen, Pedersen, and Thorbjørnsen 2005), it can be argued that decision makers and the workforces can use a technology based on social pressure alone, although their attitude toward using the technology can be neutral or negative. According to the TPB, a positive effect of perceived behavioral control on intention to use IaaS is suggested. This control concept covers organizational factors like the availability of resources, knowledge and capabilities of effective IaaS usage. The organization may have a low intention to use IaaS due to the lack of skills or high costs related to the use of the infrastructure service. From a TCT background, it can be argued that an increased perception of uncertainty results in increased transaction costs (e.g. data search costs) which would reduce the benefits of outside infrastructure resource usage. Moreover, previous studies in the e-commerce field have shown that a heightened uncertainty perception is detrimental to technology adoption (Pavlou, Liang, and Xue 2007), (Pavlou 2003). Hence, if IaaS users are

⁹Macroeconomic and aesthetic-emotional added values may occur but were not focus of the study.

worried about using IaaS due to the numerous possibilities of negative effects, they are less likely to engage in such an online exchange relationship. Based on this discussion, Hypothesis 3 (H3) reads as follows: The decision maker's intention to use will be greater, if his or her attitude towards use, the subjective norm and the perceived behavioral control is rated more positively and the perceived uncertainty is rated lower.

This paragraph describes determinants related directly to the actual behavior of the decision maker (i.e. infrastructure outsourcing activities). The role of intention as a predictor of behavior is critical (Venkatesh, Morris, Davis, and Davis 2003) and has been well established in information systems research and the reference disciplines (see (Ajzen 1991; Sheppard, Hartwick, and Warshaw 1988)). When technology use is considered as the dependent variable, TPB and other theories described in (Venkatesh, Morris, Davis, and Davis 2003) employ perceived behavioral control as another predictor, in addition to intention as a key predictor. Therefore, perceived behavioral control and intention to use were used to predict behavior. Thus, Hypothesis 4 (H4) follows: The actual usage of IaaS by a decision maker will be higher, if his or her perceived behavioral control and intention to use are rated higher. Figure 3.14 gives an overview of the complete causal model, which will be evaluated in chapter 4.

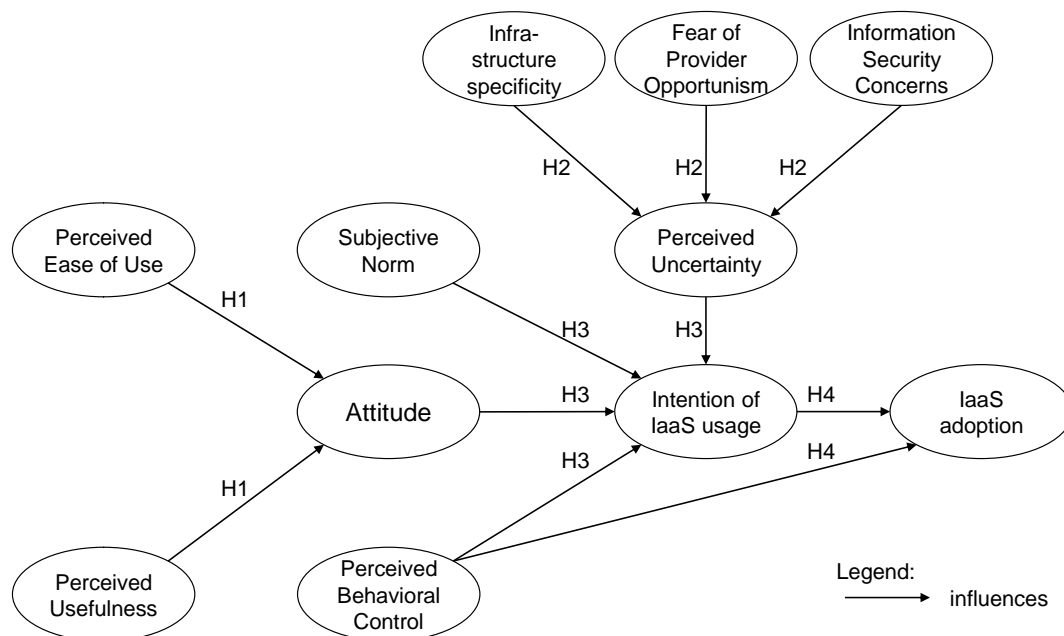


Figure 3.14: Causal model (based on (Wiedemann and Strebel 2011))

3.5 Discussion and Summary

The goal of this chapter was to establish testable hypotheses through a multi-method research approach. As a summary, it can be concluded that the results of all three research approaches converge on a set of testable hypotheses with a common theme. The case study results in Table 3.2 are surprising, as enterprise Grid adoption is mainly affected by organizational issues like IT governance, demand and capacity management, and the assessment of business value. The expert interviews continue this theme, as the hypotheses developed in the case study were on the whole comprehensive and generally supported by the results of the expert interviews. It is also a sign of external validity, that the general Cloud Computing issues usually noted in surveys like data security and legal issues were also mentioned frequently by the interviewees of this IaaS-centered study. For example, the issue of cost-benefit evaluations in the expert interviews is directly related

to the assessment of business value reported in the case study. The expert interviews generate a number of interesting hypotheses, that warrant further investigation, especially the role of trust in the provider-client relationship and the role of the IT department in IaaS adoption.

The hypotheses formed in the case study and the expert interviews are then aligned with the hypotheses gathered in literature research. For example, the topic of cost-benefit evaluation is directly addressed in the concept of perceived usefulness; other hypotheses from the expert interviews are similarly mapped onto the theoretical concepts. The result is a unified set of causal hypotheses, that aim to explain IaaS adoption in enterprises, and that is based both on exploratory and explanative modes of research.

One of the assumptions made in section 3.4.2 is the causality among the hypotheses (causal hypotheses according to Bortz and Döring (2006, p. 517)). This assumption is necessary for showing that the influencing factors on IaaS adoption are in fact responsible for its adoption (i.e. that they cause the adoption decision to some degree). The hypotheses have to be understood statistically; in mathematical terms, the network of hypotheses (causal model) is tested using correlation measures; however, a high correlation among two concepts (variables) alone is insufficient to prove a causal relationship. The statistical dependence between two variables is only a necessary condition for causality (Weiber and Mühlhaus 2010, p. 9); however, the absence of any statistical dependence can be used to falsify causal hypotheses. In order to be able to infer a causal relationship from a statistical dependency, a careful theoretical and fact-based argument for the assumed causality and its direction has to be established (Weiber and Mühlhaus 2010, p. 8). Therefore, the causal model in this thesis relies on a multitude of established theories (see section 3.4.1); the causal relationships among the concepts of these theories are documented in the research literature.

Chapter 4

Empirical Model Evaluation

4.1 Evaluation Approach

Following the methodological recommendations in (Cheon, Grover, and Teng 1995) and (Bortz and Döring 2006, p. 12), a multi-theoretical research approach for explaining corporate IaaS usage has been chosen. The preceding chapter developed a causal model explaining the IaaS adoption of enterprise in terms of several hypotheses, which were derived from two studies with expert interviews and from established concepts in information science research theories. These hypotheses form a complex cause-and-effect network reflecting the various determinants for IaaS adoption and their direct and indirect effects on each other.

The empirical assessment of hypotheses follows the deductive-nomological model by Hempel and Oppenheim (1948). In this approach, the truth of the explanandum, i.e. the propensity of IaaS adoption, follows from the truth of universally accepted laws (i.e. a number of well-published theories) and the truth of the antecedent (i.e. the required necessary conditions of the applied theories). Thus, the research approach here is quantitative, and the mode of reasoning is deductive. The relationship between the antecedent and the explanandum has to be tested empirically (Bortz and Döring 2006, p. 17), thereby verifying the truth of the hypotheses. The value of a deductive-nomological explanation depends on how well the underlying theories are empirically supported (Bortz and Döring 2006, p. 17). Empirically verifying the truth of hypotheses is logically impossible, as the empirical data is usually only a sample of the population. However, if the sample data does not falsify the theory, the truth of the theory is assumed and it is thought to be generally valid for the population (inductive inference following Popper's critical rationalism (Popper 1934)).

The hypotheses in the model need to be transformed into statistical hypotheses, i.e. it has to be determined how and in which context the variables that make up the hypotheses can be measured. As scientific hypotheses are probabilistic statements (Bortz and Döring 2006, p. 10), their truth is not absolute, but has to be assessed using statistical significance. The results of the significance calculations plus the definition of a significance level allows the statistical falsification of the aforementioned theories. Hence, in the best case, the hypotheses cannot be rejected.

The type of investigation can be deduced from research question R1. Its goal is to find the determinants of IaaS adoption in enterprises, hence the investigation needs to answer, why IaaS adoption takes place and what the determining factors are (i.e. what causes IaaS adoption?). As the state of the art offers a sufficiently large body of theories and related work, and as hypotheses based on the state of the art can be substantiated, an explanatory investigation can be chosen in accordance with Bortz and Döring (2006, p. 52). The hypotheses are formulated as causal hypotheses in correspondence to the goal of research question R1. Each hypothesis has its independent and dependent variables; the goal of this research is the determination of the relative explanatory value of the various independent variables of this causal model.

Investigations concerning the assessment of causal relationships are called interdependence analyses, as the relationships between the variables are based on the measurement of the degree of common variations among these variables Bortz and Döring (2006, p. 506). Ensuring that these variations are due to causality requires contextual and factual considerations, which might rule out alternative causal models (e.g. because of temporal relationships). These considerations are already reflected in the causal model designed in the preceding chapter, as it is based on existing, empirically well-supported causal structures (esp. the theory of planned behaviour). The interdependence analysis in this thesis is a one-shot cross-sectional design, which samples several attributes from a representative set of enterprises at one point in time in the field. This matches well with the desired experimental design.

The experimental design is one of the key points influencing the validity of the investigation. A quasi-experimental design is chosen here, as randomization is not possible. A randomization regime would require experimental control over the organizational and environmental factors acting upon an enterprise (with IaaS adoption being the outcome). However, IaaS adoption is considered as the logical consequence of factors beyond the experimenter's control. Possible confounding variables like company size are included in the design and are analyzed. As IaaS adoption in enterprises is a natural phenomenon which cannot be experimentally controlled, the selection of a quasi-experimental design is justified. A true experimental design is hardly imaginable for this research question, as the various organizational and environmental factors are hard to replicate in a lab setting. So a field study is chosen for conducting research, as the informants are only available in their organizational setting. An experimental field study would be impossible for this research question as this design would entail putting whole organizations in an experimental setting (e.g. through randomization). Quasi-experimental field designs have a lower internal validity than experimental field designs, yet their external validity is usually on par with experimental field designs Bortz and Döring (2006, p. 58).

The practical implementation of these epistemological assumptions follows the schema put forward by Bortz and Döring (2006, p. 22). The state of the art already supplies ample preceding research for formulating sensible hypotheses (see 3.4.1). These hypotheses have to be transformed into statistical hypotheses which entails their operationalization and the definition of the measurement process, which is detailed in section 4.2.2. The required steps for checking the validity of the measurement process (e.g. the selection of the significance test, the calculation of p-values, etc.) are described in section 4.4.3. Section 4.2 defines the concepts of the data analysis technique Structural Equation Modeling (SEM) used in the further course of the investigation; this section also contains the details of the pretest. In section 4.3, details of the participant acquisition and the sampling strategy are highlighted. Based on the survey data, the descriptive analysis, the quality assessment of both the structural and the measurement model and the assessment of the research hypotheses take place in section 4.4.

4.2 Development of the Survey Instrument

4.2.1 Structural Equation Modeling Concepts

The following sections will rely heavily on SEM-specific terms and concepts, hence a short introduction to the most important ones shall be given here.

Weiber and Mühlhaus (Weiber and Mühlhaus 2010, p. 73) understand SEM as "... the complete process ranging from the theoretical and factually logical formulation of the structural model and its measurement models to the assessment of the empirically found results using structural equation analysis". Structural equation analysis (SEA) includes statistical techniques for analyzing complex interdependency structures among manifest and/or latent variables and enables the quantitative estimation of cause-effect relationships. The goal of the analysis is the best possible reproduction of the input data through the structural equation

model (Weiber and Mühlhaus 2010, p. 17). To this end, structural equation models representing complex interdependencies among variables in a linear equation system, are formed and are utilized to estimate the coefficients among the observed variables as well as the measurement errors (Weiber and Mühlhaus 2010, p. 6).

As structural equation analysis represents an extension of the classical multivariate regression analysis, the common notion of independent (predictor) and dependent variables (target variables) also has to be extended. Structural equation models feature three types of variables (Weiber and Mühlhaus 2010, p. 18):

endogenous variables These variables are always target variables, whose values are explained by the influence of other variables in the structural model.

exogenous variables These variables are always predictor variables, which are externally given and which serve to explain the values of the endogenous variables in a structural model. They are not explained by the model.

intervening variables These variables are both target and predictor variables, who serve as an input to other predictor variables in a structural model.

Another difference between multivariate regression analysis and structural equation analysis becomes visible when looking at the values of the variables. While regression analysis deals with manifest variables (i.e. directly empirically measurable variable values), SEA can handle both manifest and latent variables (i.e. not directly empirically measurable variable values). Latent variables are also called (theoretical) constructs; examples for constructs are usually related to social science theories and could be concepts like reputation, trust or competency (Weiber and Mühlhaus 2010, p. 19). Latent variables require suitable measurement models that contain instructions on how a latent variable, i.e. a hypothetic construct, can be assigned to an observable fact (operationalization) and can be numerically captured (measurement). The measurement result is mapped to a measurement variable which is directly empirically observable and hence constitutes a manifest variable (Weiber and Mühlhaus 2010, p. 35).

According to Weiber and Mühlhaus (2010, p. 31), structural equation models with latent variables consist of three partial models:

1. The structural model represents the theoretically assumed relationships among the latent variables. In the model, the endogenous variables are explained by the assumed causal dependencies, the exogenous variables serve as explanatory values, which themselves are not explained by the causal model.
2. The measurement model of the latent exogenous variables contains the empirically measured values from the exogenous variables' operationalization and reflects the assumed relationships between the measured values and the exogenous quantities.
3. The measurement model of the latent endogenous variables contains the empirically measured values from the endogenous variables' operationalization and reflects the assumed relationships between the measured values and the endogenous quantities.

It was already mentioned that the regression analysis is one of the predecessors of SEA; one of the other major predecessors is path analysis, which aims at analyzing interdependencies among (strictly manifest) variables in the path model. It also assumes an a-priori formulation of the causal relationship among the variables, which can then be assessed by the path analysis. For a complete list of SEA assumptions, the assumptions associated with path analysis also have to be included. The complete list is as follows (Weiber and Mühlhaus 2010, p. 30):

- One variable has to precede another variable in a causal fashion

- There exist linear and additive relationships among the variables
- Metrically scaled and standardized variables are examined, whose (manifest) values can be gathered without measurement error.
- The residuals are normally distributed
- No multicollinearity among exogenous and intervening variables
- The residual paths of the endogenous variables are uncorrelated and do not correlate with the exogenous variables.
- The measurement error variables assumed in the measurement model for latent variables are neither correlated with the latent variables nor with other measurement error variables.

4.2.2 Structural Model and Measurement Model

The model includes constructs, most of which are well founded in IS research literature. Table 4.1 summarizes the literature for the formative and reflective items in this survey. The second to last column in this table shows the associated code name for the hypothetical constructs; it will be used throughout the further quality assessment of the measurement models. A number attached to the code name signifies the number of the indicator (e.g. REL2 for the second indicator of the REL construct). The last column in this table lists the number of the question in the questionnaire that is associated with the construct in the same line. The questionnaire can be found in Appendix B.

There are two principle types of measurement models to operationalize a given theoretical construct, formative and reflective measurements (Weiber and Mühlhaus 2010, p. 35). The definition of a reflective measurement model is given by Weiber and Mühlhaus (2010, p. 90): in reflective measurement models, a hypothetical construct represents the causes of the changes in the measurement indicators collected on an observational level. Ideally, changes in the construct values are reflected simultaneously by all measurement indicators; it is assumed that the construct acts as an independent variable which affects the indicators.

In formative measurement models, the hypothetical construct is understood as the consequence of the measurement indicators in effect on the observational level. Hence, a construct constitutes a linear combination of measurement indicators, which corresponds to a linear regression-type of approach (Weiber and Mühlhaus 2010, p. 202). The construct is the dependent variable in the corresponding regression formula. A single indicator represents a factual facet of the hypothetical construct, therefore, the construct is extensionally defined by the combination of its indicators.

When constructs are operationalized, care has to be taken to correctly distinguish and apply these two types of measurement models. The key questions is: “do the changes in the measurement indicator values cause value changes in the latent variables (formative) or do the changes in the latent variable values cause changes in the measurement indicator values (reflective)?” (Weiber and Mühlhaus 2010, p. 36). Also, Jarvis, MacKenzie, and Podsakoff (2003) give a comprehensive list of decision criteria to further clarify the identification of formative and reflective construct measurement models.

The measurement indicators listed in Table 4.1 were taken from literature sources also given in the table, so their validity had already been established before they were used in this survey. However, care was taken to ensure the correct type of measurement model by reviewing the measurement indicators under the criteria given above. (Eggert and Fassott (2005, p. 44) shows in a review of 25 articles, that 109 out of 125 reflective latent variables were operationalized in such a way that a formative measurement model would have been more appropriate.) The use of standardized measurement models is well established and recommended in the social science literature (Bortz and Döring 2006, p. 191), (Weiber and Mühlhaus 2010, p. 86). However, the existing operationalization of the constructs had to be adapted to the IaaS research

Table 4.1: Measurement Models

Theory	Construct Name	Source	Type	Code	Question No.
TRA	IaaS Adoption	Benlian and Buxmann (2009), Cheon et al. (1995)	reflective	ADO	14
	Attitude towards IaaS	Benlian and Buxmann (2009)	reflective	ATT	6
	Intention of IaaS Usage	Ajzen (2010)	reflective	INT	7
	Subjective Norm	Eckhardt et al. (2009)	formative	SNO	8
TPB	Perceived Behavioral Control	Taylor and Todd (1995)	reflective	PBC	9
TCT	Infrastructure Specificity	Yao (2004)	reflective	INS	10
	Perceived Uncertainty	Benlian and Buxmann (2009), Susarla et al. (2009)	formative	SEU	11
PAT	Fear of Provider Opportunism	Pavlou et al. (2007)	reflective	SEO	5
	Information Security Concerns (reverse coded)	Pavlou et al. (2007)	reflective	ISC	5
TAM	Perceived Usefulness	Davis (1989), Bazijanec et al. (2004)	formative	REL	12
	Ease of Use	Davis (1989), Bradford and Florin (2003)	reflective	EOU	13

topic, which made a slight reformulation necessary. For example, some constructs had originally been developed for ASP investigations, so the term “ASP” had to be replaced by the term “IaaS”.

The next step in the development of the survey instrument is the construction of the measuring approach. This process is also called scaling. Scaling generally denotes the construction of a measuring approach which assigns numbers to qualitative real-world properties and thus helps to gather those properties in a quantitative fashion (Weiber and Mühlhaus 2010, p. 95). The research literature knows various scaling approaches; for the above defined constructs, a bipolar six-point Likert-type rating scale was used. Using this scale, the participants can express their consent with a measurement indicator on a continuum from total rejection to complete agreement. The number of six points is universally recommended in the research literature (Weiber and Mühlhaus 2010, p. 97). Moreover, the even number of points forces the participant to gravitate towards one of the two sides of the question, as no neutral middle element exists. (The final questionnaire also contained a small number of questions with a different scale, but only for the demographical and socio-economic status of the participant; the constructs were all measured as described above.)

To reduce the number of missing values, forced ratings were used, i.e. each question had to be answered without the option of a default answer option (e.g. “don’t know”). This design decision is sensible, as the target audience of this survey consists of IT experts who can be expected to be knowledgeable about the subject IaaS. The rating scale for the construct “Information Security Concerns” was reverse coded: this change in direction is explicitly recommended in the literature (Weiber and Mühlhaus 2010, p. 99), as it allows the detection of inattentive participants or participants that answer the survey in patterns. The application of a rating scale also implies the application of closed-ended questions. These questions should preferably be deployed in questionnaires (Bortz and Döring 2006, p. 254), as the answers can be analyzed more easily and are less prone to missing values.

4.2.3 Online Questionnaire Development and Pretest

The choice of an online survey as a medium is an important survey design decision. Online surveys have become over the last years an important alternative to established questionnaires sent by mail. In 2008,

online surveys had a share of 31% of all surveys with quantitative survey types (among German market and social science research institutes) (Thielsch and Weltzin 2009, p. 69). The decision for an online survey will be further motivated in the following paragraphs.

Thielsch and Weltzin (2009, p. 70) gives a list of advantages and disadvantages of online surveys. Especially advantages like time efficiency, lower expenses and the possibility of automation are convincing arguments. Moreover, a higher return data quality can be expected as the questionnaire is designed to check for the completeness of the data and as time stamps on submitted questionnaires allow the easy identification of inconsistent entries.

Some of the disadvantages shall also be discussed: the members of the target group might not be available online (Thielsch and Weltzin 2009, p. 70), but this is unlikely in the case of a target group consisting of IT experts. As this survey is designed as a Web survey, there is no need for the participants to invest any effort in learning a specific survey software. Also, an online survey runs the principle risk of multiple entries per participant (Thielsch and Weltzin 2009, p. 70). The standard software used in this survey uses browser cookies to mark participants who already submitted their entry, so this risk can be minimized. As a conclusion, the use of an online survey has significant advantages, whereas the disadvantages can be reduced by technical measures, so this medium is perfectly suitable for this survey. The design decisions of the online questionnaire shall be discussed in the following paragraphs. The complete questionnaire as it appeared online can be found in Appendix B.

As recommended by Thielsch and Weltzin (2009, p. 71), the initial page of the survey features the contact data and the organization of the persons in charge of the survey, describes the goals and the contents of the survey and gives a honest estimate of the average duration of the survey. Also, a note regarding the confidentiality and the scientific usage of the collected data was visible. As an incentive for potential participants, the first page also contained a description of the three IT-related books that were raffled off among the participants that provided their e-mail address at the end of the survey.¹

The questionnaire itself starts with an “ice-breaker” question, which is simple to answer, and which helps to relax the participant and makes him feel comfortable with the survey tool and the general navigation. The later questions were group by related topics, such that participant knows the context of the questions. The questions (indicators) belonging to one construct were shuffled randomly in order to avoid sequence-related effects in the responses.

The questionnaire design aims to give visual cues to the reader for easy orientation; the rating scales are both numerically (1...6) and textually labeled, and radio buttons were used for choosing the answers which works in favor of lower item non-response rate (Vicente and Reis 2010). Additionally, the design was screen-oriented rather than scrolling-oriented; this way, the item non-response rate can be reduced (Vicente and Reis 2010). Also, both a graphic and a numeric progress indicator was visible on every page of the questionnaire; this visual aid generally decreases drop-outs (Vicente and Reis 2010).

The overall design goal was to have a short questionnaire, which is better for the completion rate (Vicente and Reis 2010), as the target audience cannot be expected and is probably not willing to spend extended stretches of time in front of the questionnaire. The possibility to leave comments at the end of the questionnaire was offered and this channel resulted in valuable insights both in the pretest and the production run.

Atteslander (2008, p. 277) recommends a pretest after the questionnaire design to test the capability of the survey instrument. Four main aspects deserve special attention: reliability and validity, comprehensibility of questions, uniqueness of categories and specific data gathering problems (Atteslander 2008, p. 278).

For the pretest, 22 invitations to experts in the field of IS research or to business executives were sent out; the pretest followed the same procedure and used the same tools as the main study. 12 completed ques-

¹The books were sponsored by Proventa AG, where Dr. Wiedemann, one of the survey organizers, is employed

tionnaires were retrieved. Some of the participants also provided written feedback with detailed suggestions for improvements. Additionally, two sessions with an expert panel of IS researchers from the Information & Market Engineering Chair at KIT (Prof. Weinhardt) were held which also resulted in numerous suggestions. All pretest activities took place in December 2010. The reliability and the validity of the constructs were not checked during this pretest, as the construct operationalization was based on literature sources and hence assumed to be valid. Also, the number of pretest participants was much too low to assess the reliability and validity of a rather complex structural model.

The linguistic and content-related comprehensibility of the questions was thoroughly checked in the pretest. The following improvements consisted of a changed wording of the question items and the ice-breaker question, the clarification of question contexts, etc. The changes in wording often helped to reduce the possibility of ambiguous formulations. The uniqueness of the items in categorical questions is less of a problem in this survey, as only few categorical questions were asked (besides the regular rating scales).

Data gathering problems were also tested; one major issue in Web surveys is the expected duration of the survey, as a long-running survey tends to have higher drop-out rates. The pretest was able to correctly estimate the duration to be on average ca. 12 minutes (which was also found later in the main survey). Moreover, the pretest ensured the correct functioning of the survey software infrastructure and its data export facilities.

4.3 Data Collection and Preparation

After several rounds of pretests and revisions of the questionnaire, a Web-based survey was used for data collection. The regular, officially communicated data gathering period was between January and March 2011. The final sample also contains 12 laggards, which participated between April and July 2011.

The software infrastructure was provided by the Forschungszentrum Informatik (FZI) in Karlsruhe; they operate a Web server with an installation of the Open Source survey tool LimeSurvey.² The questionnaire was available as a Web site on the WWW.³ As a service for the participants and for better usability, an alternative URL was set up for the survey.⁴

A key-informant, single-respondent approach was used, as the necessary data can only be obtained from knowledgeable specialists within an enterprise, who are willing to pass on their insights (Kumar, Stern, and Anderson 1993). Informants are not necessarily representative members of an organization, but are chosen nonetheless, as they can generalize about the observed or expected organizational relations (Kumar, Stern, and Anderson 1993). Unfortunately, this methodology introduces both informant bias and random error in the collected data (Kumar, Stern, and Anderson 1993); the alternative of selecting and surveying multiple informants from the same enterprise is methodologically preferable, but also not without challenges, as those matching informants have to be identified and their potentially dissimilar responses have to be combined. The literature offers little methodological support for either problem (Kumar, Stern, and Anderson 1993), so the predominant organizational research approach remains the key-informant one.

With focus on German-speaking IT management executives, participants were recruited by direct invitations using the XING social business network.⁵ The personal profiles on the Xing network made the identification of suitable participants possible; obviously, these profiles are self-reported, which limits their trustworthiness. However, care was taken to inspect the complete expert profile and assess its soundness; moreover, the candidates were part of the social network and largely organized themselves in special user groups, thematic exchanges and forums. Thus, a familiarity of the candidates with the topic can be assumed.

²<http://www.limesurvey.org/>, last accessed 2013-12-29

³<http://amazonas.fzi.de/limesurvey/index.php?sid=31459&lang=de>, now defunct

⁴<http://www.iaas-studie.de>, now defunct

⁵<http://www.xing.de/>, last accessed 2013-12-29

The target profiles of the informants were either IT executives (e.g. CTO, IT lead) or IT experts (e.g. IT architects, IT consultants). Table 4.2 shows the roles of the participants.

The letter of invitation was emailed once to each candidate (the letter of invitation can be found in Appendix A). This was the only invitation attempt, a second round of invitations was not sent out due to the large pool of suitable candidates. In total, 1441 quasi-randomly selected candidates were invited to participate (additional invitations were sent to a small number of the researchers' personal contacts and some advertising of the questionnaire was done in several Cloud-related blog entries covering the topics of the study).

After checking the plausibility, integrity and completeness of the 452 received questionnaires, 276 could be utilized for further analysis, which equates to a response rate of 19.15%. This low response rate reflects the challenges in obtaining responses from top management informants (a common problem in the IS area (Benlian and Buxmann 2009)).

The following steps were taken to prepare the raw survey data for further analysis:

1. All incomplete responses were inspected and treated as follows: if 10% or more of overall missing values per question were detected in the complete data set, the responses with the highest number of missing values were deleted. In the end, all questions used for further analysis had at less than 10% of overall missing values. This procedure follows the available case analysis for missing values (Göthlich 2007, p. 123); according to Hair, Anderson, Tatham, and Black (2006), a rate of 10% missing values is permissible without a deeper analysis of the effects of missing values on the survey results. 175 unfinished questionnaires had thus to be deleted, 5 unfinished questionnaires remain.
2. The two indicators for the construct "complexity" were reverse-coded in the questionnaire. Their values had to be inverted to match the formulation of the initial hypotheses.
3. The time for answering each question was calculated. In the final data set, all participants needed at least four minutes to complete the questionnaire. Responses with shorter timings were incomplete and thus were eliminated.
4. Two complete answers had to be removed, because the participant chose to answer every question with the same value from the scale and because they finished the survey extraordinarily quick (less than five minutes). As an assumption, those informants probably wanted to finish as quick as possible, but still be able to participate in the raffle and receive the final survey result document which was distributed among all participants.
5. One adoption indicator (ADO1) was removed because of its high percentage of missing values. This correction turns the adoption construct into a single-item construct. The remaining item also had around 8% missing values, but according to (Hair, Anderson, Tatham, and Black 2006), the item data can be used without assessing the missing value pattern. The missing values were replaced with the sample mean (Hair, Anderson, Tatham, and Black 2006). Although literature heavily debates the use of single-item constructs, (Fuchs and Diamantopoulos 2009) claim that such a measurement model may be used for constructs with a high level of concreteness, which is the case here. Although, ADO1 was eliminated from the structural/measurement model, it was used for descriptive analytics, even though its validity is limited.

4.4 Data Analysis and Results

The following sections give an overview of the survey results. The collected data is analyzed in multiple ways: first, a descriptive statistical analysis is executed which directly evaluates indicator data and tests statistically several interesting hypotheses that extend the SEM analysis (section 4.4.2). Second, two sections

show the results of the quality checks of both the measurement and the structural model. As there is no single overall quality criterion for PLS (Partial Least Squares) models, the quality assessment of the SEM model involves a large number of individual checks both on the constructs and the structural model (section 4.4.3 and 4.4.4). Third, the initial hypotheses of the structural model are assessed based on the results of the PLS estimation of the model variables (section 4.4.6). The reasons for using a variance analytical approach like PLS are given in the following section.

4.4.1 Model Estimation Method

In principle, there are two approaches suitable for the numerical treatment of SEA, PLS path analysis (implemented in software packages like SmartPLS (Ringle, Wende, and Will 2005)) and covariance structure analysis (implemented in software packages like AMOS and LISREL) (Bliemel et al. 2005, p. 9). For a better understanding of factors that influence attitude, intention, and behavior in IaaS usage decisions, the PLS method was applied in this research setting. In order to motivate this decision, the favorable properties of PLS for the research setting will be detailed, so that the better fit of PLS path analysis becomes apparent.

Covariance structure analysis, generally covariance-based methods, are based on a factor analytical approach, in which the interdependencies among all parameters in the model are estimated simultaneously. The goal of this analysis is the most exact reproduction possible of the empirical variance-covariance matrix. The latent variables represent factors in the sense of a classical factor analysis. The construct values remain latent throughout the estimation process (but can be explicitly estimated after the model estimation). There are several estimation methods, with Maximum Likelihood being the most popular; under its multi-normality assumptions, an array of statistical inferences can be tested (Weiber and Mühlhaus 2010, p. 57). The mathematical formulation can be found in (Weiber and Mühlhaus 2010, pp. 47).

PLS is a powerful method of analysis with comparatively low demands on sample size, measurement scales, and residual distributions (Chin 1998). PLS path analysis is characterized by a regression analytical approach, in which the interdependencies among all parameters are estimated successively. The goal of this approach is the most exact reproduction of the empirical input data matrix while minimizing the measurement error in the model. In a first step, a specific construct value is calculated as the weighted linear combination of the measurement variables (indicators) associated with the construct. This step involves an iterative estimation algorithm, which estimates the construct value both based on the structural model and the measurement model. As soon as both construct value estimates converge up to a difference of 0.000001, the algorithm stops. The construct values are then utilized in a second step to estimate the path analysis parameters in the structural model via linear regression. The detailed mathematical description of the steps can be found in (Weiber and Mühlhaus 2010, p. 59).

PLS does not make any assumptions about the statistical distribution of the sample values; this precludes the application of tests for statistical inferences. However, the estimation of standard errors for the model parameters is possible using resampling methods like bootstrapping, if the distribution of the sample data is assumed to be statistically representative of the population distribution (Weiber and Mühlhaus 2010, p. 63). PLS is generally an approach with low statistical preconditions (no assumption of normal distribution, no parametrical assumptions for resampling methods, no requirement of sample independence, no identical distributions for residues (Huber et al. 2007, pp. 10)). A detailed comparison of both approaches can be found in (Huber et al. 2007, pp. 9), (Weiber and Mühlhaus 2010, pp. 66) and (Bliemel et al. 2005, p. 11) which shall not be fully reproduced here. Chin and Newsted (1999) recommends to apply the PLS approach under the following circumstances (list also in (Weiber and Mühlhaus 2010, p. 69)):

- PLS is preferable if the investigated phenomenon is relatively new and no established measurement and construct theories are available.

- The models show a high number of measurement variables and are structurally complex.
- Prediction-making is the focus of the research.
- Only relatively small samples are available.

Additionally, PLS is able to integrate formative constructs seamlessly (a feat that covariance-based methods can only indirectly emulate) (Weiber and Mühlhaus 2010, pp. 66). The direct availability of construct values with PLS is highly beneficial for the intended evaluations, as the construct values have to be put in relation to socio-economic factors also polled in the survey (this usage also stresses the importance of PLS as an approach that emphasizes the predictive power of the resulting model). Moreover, the PLS approach is more appropriate for the further evaluation due to the innovative character of this study and the limited availability of sample data. Out of the available PLS software packages, the software SmartPLS (Ringle, Wende, and Will 2005) was used for executing the quality checks and the model estimation.

4.4.2 Descriptive Analytics

Table 4.2 gives an overview of the fundamental metrics of the surveyed enterprises and the roles of the participants in their respective enterprise. The identifiers PER1, PER2, PER5, PER6 in brackets behind each question in the “Question” column act as labels for further references to this questions. The descriptive statistical analytics were created using SPSS Statistics 19.0.⁶

The survey findings are based on the responses by 61% IT executives, 15% business executives, 18% functional/technical specialists (6% missing values) from a variety of industries including IT and Telecommunications (42%), Production (15%) and Services (Law, Counseling, Real Estate) (13%). 49% of all participants were employed in executive functions. The participating companies evenly varied in size across the whole range (from <9 to >5000 employees) with a slightly greater number of large companies (also visible in the gross revenue of the enterprises, where a large portion of the participating companies reported an annual revenue of >100 Mio. Euros).

The categories for the informant’s role PER1 and the industry sectors PER2 were taken from CIO magazine’s survey (CIO Magazine 2009). The lower ranks of PER5 and PER6 are based on the EU definition of small and medium enterprises (SMEs) (European Commission 2003), as it was assumed that IaaS Cloud Computing would be especially relevant for SMEs (Marston, Li, Bandyopadhyay, Zhang, and Ghalsasi 2011).

The following paragraph gives the results of the e-Service quality model developed by Modheji (2010). The informants could rate each of the six dimension separately on the above described scale from 1 to 6. The underlying data set had 16 missing values. Figure 4.1 shows a box-and-whisker plot of the results. The adherence to security and compliance regulations was the most important quality feature (average score 5.73), followed by appropriate SLA (average score 5.52). Quality of customer service (average score 5.48) and transparency and practicability of provider tariffs (average score 5.28) are grouped in the center and ranked third and fourth. The two lowest-ranked dimensions “Possibility of drafting individual contracts (e.g. duration, type, penalties...) ” (average score 4.87) and “Support of corporate IT operations processes by the provider” (average score 4.72) are ranked inconsistently, as indicated by the variations in extreme values (in the whiskers), so there is no clear consensus among the informants about these two quality dimensions (which cannot be said about the other four dimensions).

⁶<http://www-01.ibm.com/software/de/analytics/spss/products/statistics/>, last accessed 2013-12-29

Table 4.2: Descriptive Statistics

Question	Category	Count	Share
How would you describe your role in the enterprise? (PER1)	CIO/CTO	50.0	0.18
	CSO/CISO	0.0	0.00
	Superior IT Management Function	45.0	0.16
	IT Manager / Team lead / Project lead	75.0	0.27
	IT-Specialist Infrastructure	16.0	0.06
	IT-Specialist Applications	13.0	0.05
	CEO/President/Owner/Principal/COO	29.0	0.11
	CFO/Finance lead/ Superior Finance Management Function	0.0	0.00
	Business Executive (Manager, Team lead)	13.0	0.05
	Specialist in Functional domain (Purchasing, Production, Sales, etc.)	19.0	0.07
	No answer	16.0	0.06
In what industry sector does your enterprise operate? (PER2)	Information Technology and Telecommunications	113.0	0.41
	Utilities	7.0	0.03
	Welfare Organizations	1.0	0.00
	Public Sector (including education)	11.0	0.04
	Services (Law, Counseling, Real Estate)	31.0	0.11
	Production (including Automobile, Chemicals, Construction, Mechanical Engineering)	45.0	0.16
	Financial Services (Banks, Insurance companies)	18.0	0.07
	Health Sector (Facilities and Pharmaceutical industry)	8.0	0.03
	Retailing, Wholesaling and Distribution	18.0	0.07
	Transportation (Airlines, railways, shipping, logistics)	12.0	0.04
	Building industry	1.0	0.00
	No answer	11.0	0.04
How many employees does your enterprise have? (PER5)	0 - 9 MA	15.0	0.05
	10 - 49 MA	36.0	0.13
	50 - 249 MA	50.0	0.18
	250 - 999 MA	40.0	0.14
	1000 - 5000 MA	35.0	0.13
	>5000 MA	75.0	0.27
	No answer	25.0	0.09
What was the gross revenue of your enterprise in 2009? (PER6)	< 0.5 Mio. €	11.0	0.04
	0.5 - 1 Mio. €	8.0	0.03
	1 - 2 Mio. €	5.0	0.02
	2 - 10 Mio. €	16.0	0.06
	10 - 50 Mio. €	16.0	0.06
	50 - 100 Mio. €	18.0	0.07
	> 100 Mio. €	95.0	0.34
	No answer	107.0	0.39

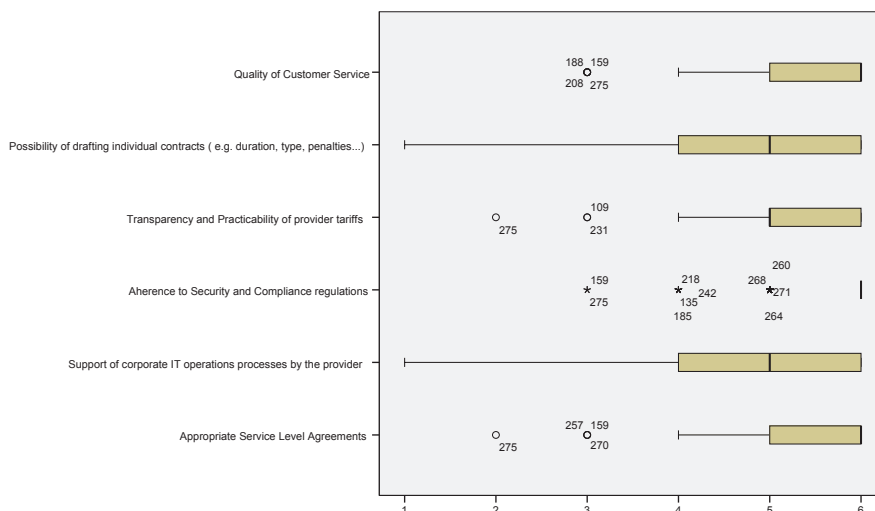


Figure 4.1: Importance of different quality criteria

The boxes in this plot are bounded by the 25%(Q1) and 75%(Q3) quantile; the extreme values are located at data points in the vicinity of $Q1-1.5*IQR^7$ and $Q3+1.5*IQR$. Data points that are larger or smaller are marked with an asterisk on the plot and the ID of the data point.

The questionnaire also measured the estimated percentage of the IT budget allocated to IaaS compared to the total IT budget in your enterprise in 2009. The answer was voluntary as this number is a rather sensitive piece of data. The item could not be used in the PLS measurement model due to the high number of missing values, however it still gives an interesting insight in overall IaaS adoption. Figure 4.2 shows the distribution of the IaaS budget shares. There were 146 answers (130 missing values, ca. 47.1%). The majority of respondents (48 respondents) shows no IaaS adoption, their IaaS budget is at 0%. But an almost as large group (40 respondents) seems to have experimented with this technology and invested a small portion (up to 5%) of their IT budget. Another portion of the respondents (45 respondents) integrates IaaS in their IT landscape and devotes from 10%-30% of their IT budget for this cause. There were few respondents that invested more than 30% of their IT budget in IaaS (13 respondents).

Heinle and Strebel (2010) identified several preferential IaaS provider characteristics that foster IaaS adoption: absolute size, positive reputation, references, further trust-building measures (certifications, data center tours). The survey follows this line of research and tries to identify factors that may influence the decision for a specific IaaS provider. Bensaou and Anderson (1999) researched the conditions under which buyers make idiosyncratic investment decision in suppliers; they identified the provider credentials number of employees and market share as important antecedents. Doney and Cannon (1997) investigated the nature of trust in buyer-seller relationships. Among other factors, the size of the supplier firm and its reputation influenced the level of trust that the buying firm had towards the supplier firm. In this context, reputation was measured using the indicators honesty, concern for the customer and estimated reputation in the market.

Figure 4.3 shows the result scores for these five indicators; in total, this question only had 9 missing values. Having a reputation for being honest scored highest (avg. score 5.62), followed by the other indicators of Doney and Cannon (1997). The number of employees and the market share showed significantly lower scores as compared to the indicators of Doney and Cannon (1997) (at an error level of $\alpha = 0.05$). The

⁷Interquartile range, i.e. the difference between the 25% and 75% quantile

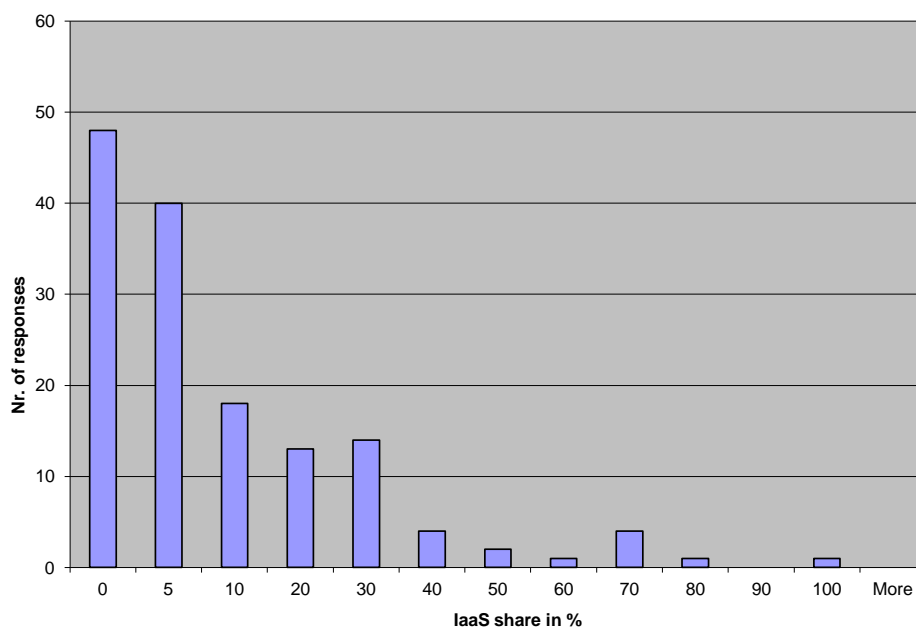


Figure 4.2: IaaS budget shares

number of employees had an average score of 3.51 out of 6, so few informants consider this indicator to be very meaningful or they have limited experience with different provider sizes and thus give an indifferent answer.

As diffusion research shows (Rogers 2003), the relative advantage and the compatibility are two of the strongest predictors of an innovation's rate of adoption. "Relative advantage is defined as the ratio of the expected benefits and the costs of adoption of an innovation" (Rogers 2003, 233); "Compatibility is the degree to which an innovation is perceived as consistent with the existing values, past experiences and the needs of potential adopters" (Rogers 2003, 243). Even though Rogers (2003) established the aforementioned relationships, he does not provide an instrument to measure these concepts. Here, compatibility shall be explored further using the measurement indicators defined by Bradford and Florin (2003) and Moore and Benbasat (1991).

Figure 4.4 shows the three indicators of compatibility, legacy system software, existing hardware and existing IT processes. There were no missing values for these three questions. The three questions were very similarly scored (avg. scores 3.83; 3.88; 3.72) and show a small tendency towards IaaS being compatible with the firm's systems, hardware and IT processes. Note that the box-and-whisker plot in Figure 4.4 shows the median as a line in the middle of the box; comparing the median with the average scores can reveal the skewness of the underlying distribution.

Diffusion research discovered another important requirement for adoption success, and that is demonstrability (Rogers 2003). According to Moore and Benbasat (1991), demonstrability consists of the two concepts observability and communicability; the indicators for demonstrability were derived from these two concepts and can be seen in Figure 4.5 along with survey scores. The indicators were suggested and evaluated by Moore and Benbasat (1991), but had to be adapted to the IaaS context by replacing the original subject with "IaaS". No missing values were found for these three questions.

Similar to compatibility, the scores were rather evenly distributed (avg. scores 4.01; 3.54; 4.05). The small affirmative tendency seems to indicate that the informants felt vaguely optimistic about their ability to talk about their IaaS adoption results. This weak result is arguably linked with the finding that an unclear IaaS definition hinders enterprise IaaS adoption (Heinle and Strebel 2010), as the communicability of IaaS adoption results depends on the common understanding of the term IaaS.

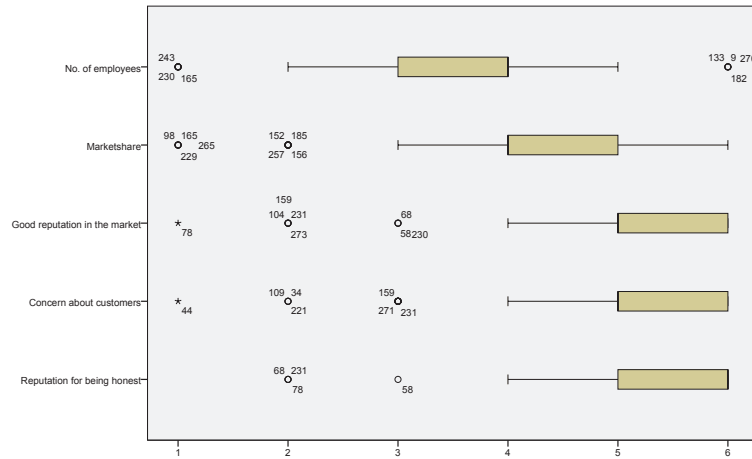


Figure 4.3: IaaS provider credentials and reputation

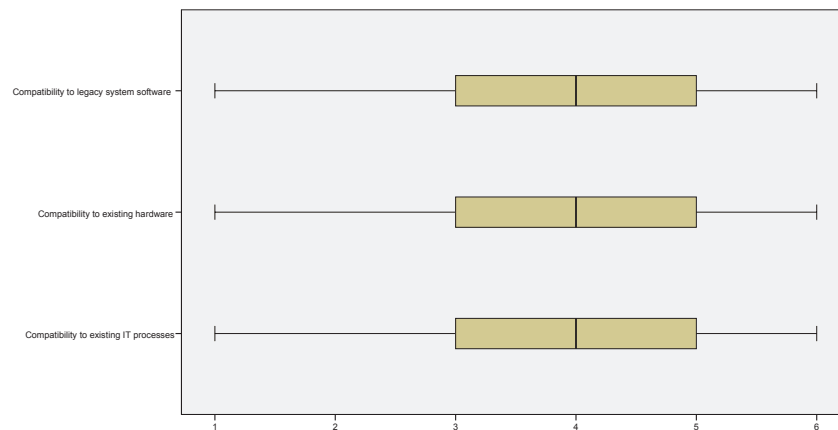


Figure 4.4: Compatibility considerations of IaaS users

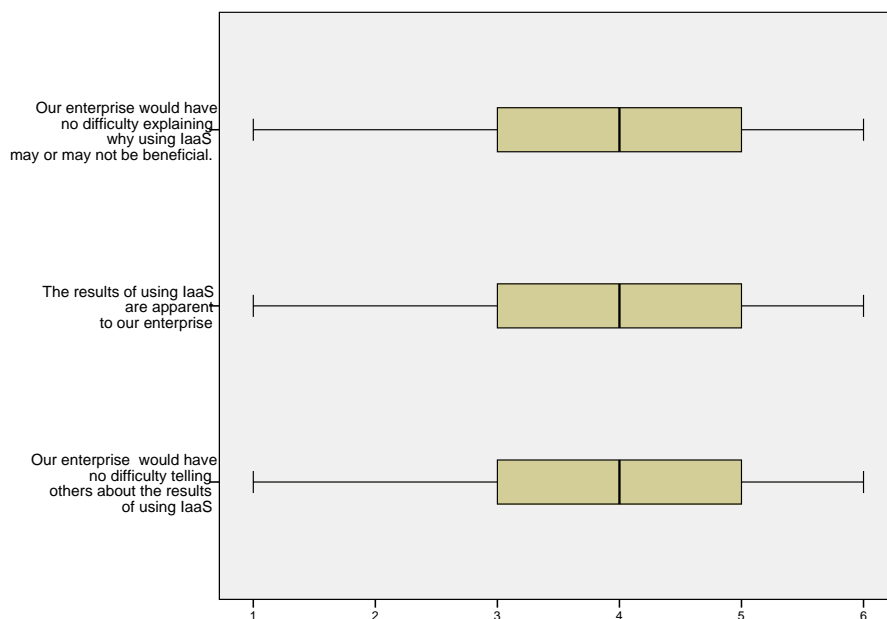


Figure 4.5: Demonstrability properties of IaaS

The importance of trialability for the diffusion of an innovation was also shown by Rogers (2003), who defined it as “the degree to which an innovation may be experimented with before adoption” (Rogers 2003, p. 224). Moore and Benbasat (1991) also provide indicators for this concept (which were altered as described for the demonstrability concept). Figure 4.6 displays the scores for the two questions (avg. scores 4.12; 3.87). There were no missing values for these questions. The slightly positive answers might indicate that the informants principally know about the possibilities of testing IaaS, but have insufficient testing experience to answer more affirmatively.

The following results come from questions that have their theoretical background in the principal-agent theory (Eisenhardt 1989a). This theory explained the efficient contract-based cooperation between an agent and a principal who have partially conflicting goals, different risk preferences and different access to information. In the case of IaaS Cloud Computing, the principal is assumed to be the enterprise user or executive that has to decide among several IT infrastructure sourcing options. The agent is then a potential IaaS provider. The contractual agreement between these two parties might range from maximal cooperation (complete outsourcing) to minimal cooperation (mainly insourcing), depending on the risk aversion of the principal and the measurability of the outcome of the business relationship.

The survey therefore contained three questions relating to the efforts required to monitor the outcome of the outsourcing relationship. The corresponding measurement indicators were developed by Loh (1994) in a study about information technology outsourcing; they were slightly adapted for the IaaS outsourcing scenario. Figure 4.7 displays the results; the data contains only one missing value. The questions were worded such that the complexity of the monitoring task had to be assessed by the informants; a higher value thus signals a higher estimated difficulty of the monitoring task.

The monitoring of IaaS providers’ investments in staff development were rated as most difficult (avg. score 4.71), followed by the monitoring of investments in technological innovation (avg. score 4.21). The monitoring of the operations performance of the IaaS provider was judged as neither easy nor difficult on average (avg. score 3.46); however, this questions has the highest variability (visible in the standard error and the interquartile range in the plot). Thus, it has to be assumed that the informants rather disagreed on the difficulty of this monitoring task.

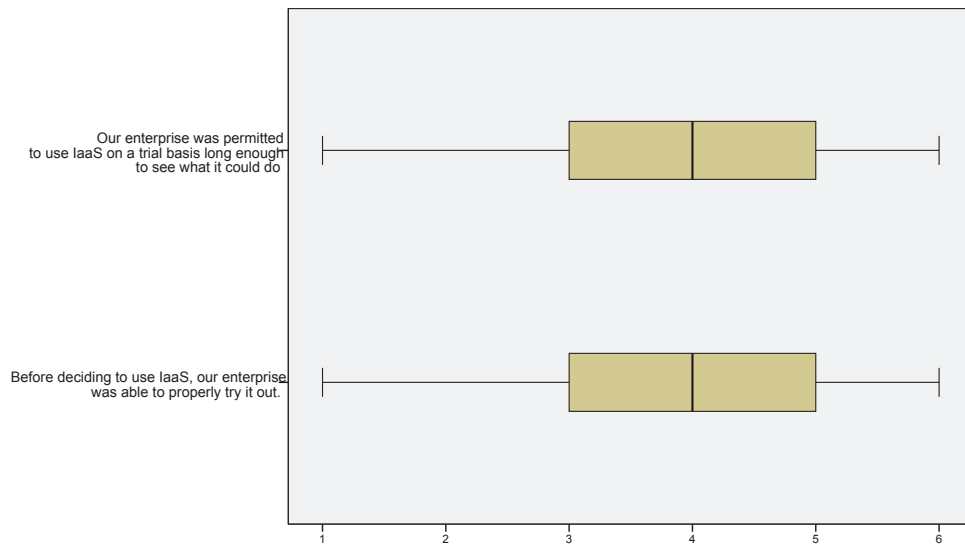


Figure 4.6: Trialability Properties of IaaS

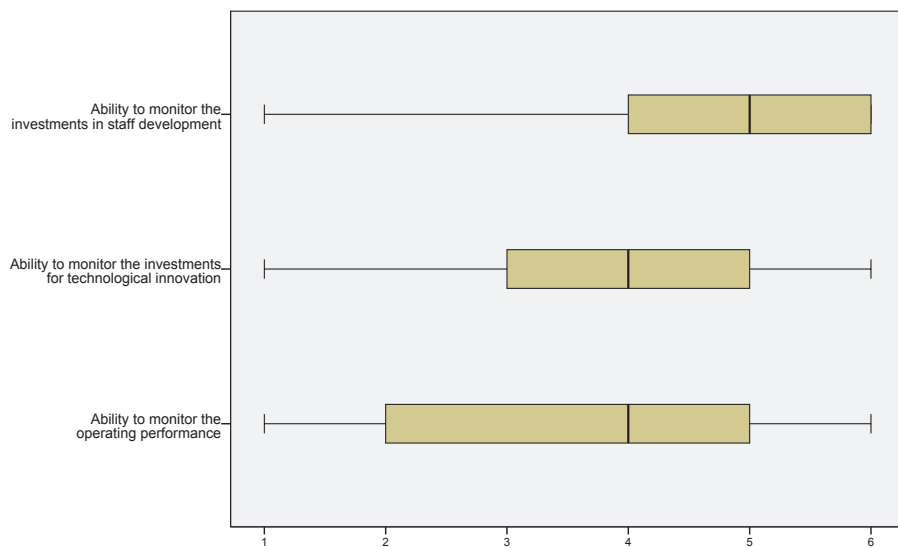


Figure 4.7: Difficulty of IaaS provider monitoring

Table 4.3: Relationship between IaaS Quality Dimensions and Company Headcount

Code	Quality criterion	F-Value	Sig. (p-value)	Mean SME score	Mean Large Ent. score
QUA1	Appropriate service level agreements	6.49	0.011*	5.38	5.61
QUA2	Support of corporate IT operations processes by the provider	7.32	0.007**	4.49	4.90
QUA3	Adherence to security and compliance regulations	3.24	0.073 ⁺	5.66	5.79
QUA4	Transparency and practicability of provider tariffs	0.79	0.374	5.24	5.33
QUA5	Possibility of drafting individual contracts (e.g. duration, type, penalties...)	2.25	0.135	4.70	4.91
QUA6	Quality of customer service	0.23	0.632	5.49	5.45

Table 4.3 shows the results of an analysis that was performed to answer the question whether SME (small and medium enterprises) have different quality priorities than large enterprises. The survey contained ca. 100 answers from SMEs and ca. 150 answers from larger enterprises (the rest were missing values in the company size question). According to the European Commission (2003), an SME is an enterprise with less than 250 employees, which was also the criterion for splitting up the two groups in this analysis.

The statistical procedure applied here is a simple one-way ANOVA (Analysis of Variance); the mathematical background for this analysis can be found in Backhaus et al. (2006, p. 154) and will not be discussed here. The quality dimensions are analyzed independently; each quality dimension is seen as a dependent variable in the ANOVA model. The two groups (SME, large company) are numerically coded in a factor which acts as the independent variable. Table 4.3 shows the test statistics for the resulting six ANOVA calculations. The results indicate that QUA1 is significant on the $\alpha = 0.05$ level, QUA2 is significant on the $\alpha = 0.01$ level and QUA3 is still significant on the $\alpha = 0.1$ level. QUA4, QUA5 and QUA6 are not sufficiently differing in regards to the factor company size; the differences might be attributable to chance.

It can be concluded that appropriate service level agreements are more important for larger enterprises than for SMEs; the absolute score (measured as agreement) is comparatively high in both cases, which indicates that SLAs are generally an important topic. Also, the support of corporate IT operations processes by the IaaS provider is a topic more relevant for large enterprises than for SMEs, but overall less relevant in absolute terms (as indicated by the lower mean score).

The adherence of the IaaS provider to security and compliance regulations is the most important topic for both SMEs and large enterprises, but there is evidence that large enterprises take this quality dimension more seriously than their smaller counterparts. Where a difference is discernible among the two groups, the large enterprises seem to be more quality-sensitive than the SMEs.

The survey data also helps to answer the question whether there is a relationship between IaaS adoption metrics and company headcount. Table 4.4 lists the results of the corresponding one-way ANOVA. ADO1 is measured as a percentage of the total enterprise IT budget and follows the measurement indicator for SaaS adoption in (Benlian and Buxmann 2009). ADO2 is a score value on the standard survey scale and is modeled after an adoption indicator in (Teng, Cheon, and Grover 1995); higher ADO2 values indicate a

Table 4.4: Relationship between IaaS adoption metrics and company headcount

Code	Adoption Metric	F-Value	Sig. (p-value)	Mean SME data	Mean Large Ent. data
ADO2	IaaS usage for business applications has strongly increased in the last 3 years in your enterprise	4.568	0.034*	3.63	3.15
ADO1	The estimated percentage of the IT budget allocated to IaaS compared to the total IT budget in your enterprise in 2009	2.983	0.086 ⁺	14.80%	9.54%

higher increase of IaaS usage. The ADO1 question yielded only 140 answers usable in this analysis (SME 69, Large Ent. 71). The ADO2 question supplied 232 usable answers (SME 93, Large Ent. 139). ADO1 and ADO2 point in the same direction as far as IaaS usage of SMEs is concerned; both indicators show higher absolute values for SMEs than for large enterprises. The difference for ADO1 is significant on an $\alpha = 0.1$ error level, while the difference for ADO2 is significant on an $\alpha = 0.05$ error level. Thus, it can be argued that SMEs show a statistically significant higher IaaS adoption propensity than large enterprises.

Another question investigated in the survey was the relationship between IaaS adoption metrics and IT affiliation. The results of the corresponding one-way ANOVA with IT affiliation as the independent variable are listed in Table 4.5.

The IT affiliation of each informant is characterized by his role in the enterprise. Table 4.2 lists all role options that were given in the survey. The roles “CIO/CTO”, “CSO/CISO”, “Superior IT Management Function”, “IT Manager / Team lead / Project lead”, “IT-Specialist Infrastructure”, “IT-Specialist Applications” were categorized as IT roles; the roles “CEO/President/Owner/Principal/COO”, “CFO/Finance lead/ Superior Finance Management Function”, “Business Executive (Manager, Team lead)”, “Specialist in Functional domain (Purchasing, Production, Sales, etc.)” were labeled as business roles. Hence, the IT affiliation is a two-level factor in the ANOVA. The ADO1 question yielded only 140 answers usable in this analysis (32 business informants, 108 IT informants). The ADO2 question supplied 241 usable answers (53 business informants, 188 IT informants).

The analysis reveals an overall greater propensity of business informants to adopt IaaS than IT informants for both ADO1 and ADO2 adoption measures. The differences between IT and business informants are highly statistically significant for the increase of IaaS usage over the last three years ($\alpha \leq 0.01$ error level) and moderately statistically significant for the budget IaaS share (close to the $\alpha \leq 0.1$ error level).

Thus, it can be postulated that business executives are generally more open towards an IaaS usage in their enterprise than IT executives. However, this results also raises some concerns about the effectiveness of IT governance in enterprises. If the IaaS usage perception clearly differs between IT and business informants, the existence of a certain shadow IT function located in the business departments of the enterprise can be assumed; it probably deploys IaaS solutions without the knowledge of the official IT functions (please see Rentrop and Zimmermann (2012) for a description of the shadow IT concept; they define it “as a collection of systems developed by business departments without support of the official IT department”).

The final question in the descriptive analytics section is concerned with the relationship between IaaS quality metrics and IT affiliation (both concepts were introduced above). Table 4.6 shows the number of informants and missing values for each quality dimension and gives the mean values and the standard deviation for each type of IT affiliation, separated by quality dimension. On average, there were around 20

Table 4.5: Relationship between IaaS adoption metrics and IT affiliation

Code	Adoption Metric	F-Value	Sig.	Mean IT roles	Mean Business roles
ADO2	IaaS usage for business applications has strongly increased in the last 3 years in your enterprise	8.842	0.003**	3.09	3.87
ADO1	The estimated percentage of the IT budget allocated to IaaS compared to the total IT budget in your enterprise in 2009	2.625	0.107	10.77%	16.66%

missing values per question. Table 4.7 lists the results of the one-way ANOVA, that was performed with the IT affiliation as a two-level independent.

The absolute score values reveal that the IT informants generally have slightly higher quality requirements across almost all quality dimensions (i.e. when they have to select an IaaS provider, they rate the importance of these quality dimensions higher than their business counterparts). The sole exception is the quality of the customer service, which is more important to business informants. However, few of these differences in importance are statistically significant, as Table 4.7 proves. Only appropriate service level agreements (QUA1) and the transparency and practicability of provider tariffs (QUA4) turn out to be statistically significant on an $\alpha = 0.1$ error level.

As the last descriptive statistical analysis, the evaluation of the qualitative feedback gathered on the questionnaire has to be mentioned. The numbers can be found in Appendix E. They paint an interesting picture of the informants' perception of the survey and their feelings towards IaaS adoption.

4.4.3 Quality Assessment of Measurement Model

4.4.3.1 Definition of quality criteria for SEM constructs

The following paragraphs summarize the quality criteria for the constructs in SEM models; these criteria are listed as the relevant quality criteria in the standard references (e.g. (Weiber and Mühlhaus 2010), (Krafft, Götz, and Liehr-Gobbers 2005), (Huber et al. 2007)).

One of the first criteria for checking the applicability of any SEM approach is the required sample size. According to the rules of thumb by Chin (1998), the minimal sample size would be 50. However, Weiber and Mühlhaus (2010, p. 259) and Huber et al. (2007, p. 2) reference earlier research that requires a sample size of at least 100 for moderately complex PLS models. In any case, the survey in this work fulfills these requirements. Table 4.8 displays the quality checklist for reflective constructs. Each criterion will be explained in more detail in the following paragraphs.

Generally, the assessment of the quality of a structural model is of paramount importance before any conclusions can be drawn from its structure. Two main concepts serve this purpose: reliability measures and validity measures. In this context, reliability is defined as the degree to which repeated measurements of the facts using the same instrument yield the same results (Weiber and Mühlhaus 2010, p. 109). Validity is characterized by the degree to which an instrument measures what it is supposed to measure and represents the conceptual correctness of an instrument (Weiber and Mühlhaus 2010, p. 127).

The literature knows several measures of validity. One of them is construct validity, which is assumed if the measurement of a specific construct is not falsified by other constructs or by a systematic error (Weiber and Mühlhaus 2010, p. 131). Construct validity can be claimed by confirming the three related validity measures, convergent validity, discriminant validity and nomological validity. The test for convergent

Table 4.6: Relationship between IaaS quality metrics and IT affiliation - Descriptive Statistics

		Number	Mean	Std. Dev.
QUA1	IT	196	5.57	0.608
	Business	60	5.33	0.914
	Missing	20		
QUA2	IT	189	4.79	1.108
	Business	60	4.53	1.282
	Missing	27		
QUA3	IT	194	5.76	0.537
	Business	61	5.69	0.593
	Missing	21		
QUA4	IT	197	5.31	0.737
	Business	60	5.12	0.922
	Missing	19		
QUA5	IT	197	4.87	0.974
	Business	61	4.80	1.209
	Missing	18		
QUA6	IT	196	5.48	0.705
	Business	61	5.51	0.674
	Missing	19		

Table 4.7: Relationship between IaaS quality metrics and IT affiliation - ANOVA results

Code	Quality criterion	F value	Sig. (p-value)
QUA1	Appropriate Service Level Agreements	5.45	0.020*
QUA2	Support of corporate IT operations processes by the provider	2.32	0.129
QUA3	Adherence to Security and Compliance regulations	0.73	0.393
QUA4	Transparency and Practicability of provider tariffs	2.94	0.088 ⁺
QUA5	Possibility of drafting individual contracts (e.g. duration, type, penalties...)	0.21	0.645
QUA6	Quality of Customer Service	0.05	0.819

Table 4.8: Quality assessment criteria for reflective measurement models

Quality criterion	Metric	Thresholds	References
Uni-dimensionality of the indicator set	Loadings of exploratory factor analysis	Acceptable Loading $\lambda > 0.7$, Indicator elimination at $\lambda < 0.4$	Krafft et al. (2005, p. 73), Backhaus et al. (2006, p. 334)
	Explained Variance	Communalities $h^2 > 0.5$	Weiber and Mühlhaus (2010, p. 107)
	Kaiser-Meyer-Olkin criterion (KMO)	Overall KMO value ≥ 0.6	Kaiser and Rice (1974)
	Measure of Sampling Adequacy (MSA)	Indicator elimination for indicator MSA values < 0.5	Kaiser and Rice (1974)
	Simple Structure of the Loadings Matrix	Display of loadings $\lambda > 0.3$ for easier visibility.	Gerbing and Anderson (1988)
Construct reliability	Composite Reliability (Factor reliability)	$Rel \geq 0.6$ per construct	Bagozzi and Yi (1988, p. 82)
	Cronbach alpha	$\alpha \geq 0.7$ per construct	Nunnally and Bernstein (1994, p. 252)
	Average Variance Extracted (AVE)	$AVE \geq 0.5$ per construct	Fornell and Larcker (1981, p. 45)
Content validity	The construct indicators represent the factual-semantic domain of the construct and the indicators cover all defined facets of meaning in a construct.	Assessment criteria (e.g. suitable indicator selection, expert judgment, pretest) fulfilled	Weiber and Mühlhaus (2010, p. 128)
Construct validity	Nomological validity	Interdependencies among the constructs in the nomological network correspond to the theoretically expected relationships	Weiber and Mühlhaus (2010, p. 132)
	Discriminant validity	Fornell-Larcker criterion: AVE of each latent variable greater than the squared correlation of that latent variable with all other latent variables in the model.	Fornell and Larcker (1981, p. 45)

validity requires that the measurements of a specific construct using two maximally different methods coincide (Weiber and Mühlhaus 2010, p. 132). This validity measure was not applied in this study as it is relatively complex and rarely found in the research literature (Weiber and Mühlhaus 2010, p. 132).

Another validity measure found in the literature is criterion validity, which is assumed if the construct values and the values of a valid outside criterion are highly correlated (Weiber and Mühlhaus 2010, p. 129). An outside criterion is also a hypothetical construct which is conceptually closely related to the target criterion. Finding and operationalizing such an outside criterion is not trivial and was omitted in this survey (also for reasons of brevity of the overall questionnaire).

Similar to the aforementioned checklist for reflective constructs, Table 4.9 displays a quality checklist for formative constructs. This comprehensive list is compiled from several standard reference source. There are no reliability measures for formative constructs, so the reliability assessment is not possible in this case (Weiber and Mühlhaus 2010, p. 208). Thanks to the formative operationalization, the indicators within one construct indicator set ideally show no or little redundancy (low collinearity), hence the calculation of a composite reliability score (similar to reflective constructs) would be pointless. Another way to prove reliability consists in conducting the same test twice, with some delay between the tests. The test-retest reliability can then be calculated based on the comparison of the two tests. This approach was omitted due to the research setting; it is unlikely, that the target group would be willing to fill in the same questionnaire twice.

Table 4.9: Quality assessment criteria for formative measurement models Weiber and Mühlhaus (2010, p. 210)

Quality criterion	Metric	Thresholds	References
Collinearity check	Variance Inflation Factor (VIF) per indicator of each formative construct	Conceptual assessment required for VIF values >3; indicator elimination for VIF values >5 and non-significant regression coefficient β	Diamantopoulos and Riefler (2008, p. 1193)
	Correlation matrix (for pair-wise indicator dependencies)	Matrix values around 0; indicator combination for high correlation values.	(Krafft et al. 2005, p. 79)
Indicator validity	indicator regression coefficients and their statistical significance	regression coefficient $\beta > 0.1$ and significant	Seltin and Keeves (1994, p. 4356)
Construct validity (nomological validity)	r^2 of formative constructs	r^2 of each construct should be sufficiently large ($r^2 \geq 0.3$)	Chin (1998, p. 325)
	Path coefficients of the formative constructs to other constructs in the nomological network	Path coefficients have to be significant and show the theoretically expected sign for the relationship.	Diamantopoulos and Winklhofer (2001, p. 273)

Krafft, Götz, and Liehr-Gobbers (2005, p. 76) and Anderson and Gerbing (1991) suggest to test substantive validity, where two indices (the proportion of substantive agreement and the substantive-validity coefficient) are calculated during a pre-test to assess the validity of the indicator sets for each construct. High values indicate a correct indicator mapping. Perceived usefulness, perceived uncertainty and the subjective norm were the only formative hypothetical constructs in this SEM. As these constructs are based on indicators from already validated sources, the test for substantive validity was not conducted.

An additional possibility to show the reliability and the validity of a formative construct can be seen in the MIMIC models (Multiple Indicators, Multiple Causes). It uses both reflective and formative indicators to measure the latent variable, hence the quality criteria for reflective constructs can be applied as well and they can be used to check whether the formative operationalization explains the construct in a similar fashion like the reflective operationalization (Krafft et al. 2005, p. 80). MIMIC models were not applied in this survey for two reasons: first, the additional redundant reflective items would have added to the overall length of the questionnaire and second, the SmartPLS software used for calculating the structural model cannot process MIMIC models directly.

4.4.3.2 Evaluation of quality criteria for SEM constructs

As the quality criteria for both reflective and formative constructs have been defined, the following paragraphs describe the evaluation results of these criteria. First, the reflective constructs will be evaluated. The unidimensionality of the indicator sets for each construct is assessed using an exploratory factor analysis (EFA). The requirement of unidimensionality follows from the concept of reflective constructs, which assumes that the hypothetical construct causes the observations in the measurement indicators, hence measurement indicators of one construct should be similarly affected. (Krafft, Götz, and Liehr-Gobbers 2005, p. 73). Unidimensionality is a precondition for later reliability assessments (Weiber and Mühlhaus 2010, p. 106).

An EFA can only be applied if the indicator data possesses certain statistical properties, which have to be checked beforehand. Bortz and Döring (2006, p. 383) and Hair, Anderson, Tatham, and Black (2006, p. 113) give slightly differing criteria for the necessary sample size of a factor analysis. In any case, metric variables are required, the absolute number of observations should be above 50 and the sample must have more observations than variables. All three requirements are fulfilled in this survey, as there are only 17 reflective variables in the survey, but 276 observations. Additionally, a minimum of five observations per variable is required as a general rule (Hair, Anderson, Tatham, and Black 2006, p. 113), which is also easily fulfilled in this sample.

Also, the data matrix has to have sufficient correlations among the variables for a factor analysis. Bartlett's test of sphericity checks whether the sample comes from a population in which the variables are uncorrelated, which is also the null hypothesis for this test. The test statistics shows a Chi-Square value of 2247.104 and a p-value of <0.0005 , so the null hypothesis can be rejected and it can be concluded that there are significant correlations present (Weiber and Mühlhaus 2010, p. 107).

One of the best criterion for the applicability of a factor analysis is the KMO criterion (Backhaus et al. 2006, p. 336). Its is based on the measure of sampling adequacy (MSA) of the whole data matrix which has a value of 0.810; according to the KMO criterion, this is a "meritorious" value according to Hair, Anderson, Tatham, and Black (2006, p. 114). Hence, it can be concluded that the survey data is suited for a factor analysis.

The extraction of the factors is a process, which is based on certain assumptions about the measurement of the factors. It is assumed that the measurement process is not free of errors and that the total variance of a measurement variable (indicator) cannot be fully explained by the factor (hypothetical construct). The principal axis analysis works under this assumption and splits up the total variance of a variable in its communality and its residual variance (Backhaus et al. 2006, p. 350); goal of this procedure is to explain the variance of a variable up to its communality (the sum of the explained variance across all factors). The factors in the solution are interpreted as the common causes of the correlations among the measurement variables (Backhaus et al. 2006, p. 351).

The number of extracted factors is guided by the Kaiser criterion, which is widely used for this purpose (Weber and Mühlhaus 2010, p. 107). This criterion demands that only factors should be extracted whose share of explained variance across all measurement variables has to be greater than 1.

The factors extracted are symmetrical under rotation, so a rotated set of factors is also a valid solution for replicating the correlation matrix. Rotation is helpful in aligning the factors with the assumed hypothetical constructs and may improve the unidimensionality of the loadings matrix (Backhaus et al. 2006, p. 356). It is assumed that the extracted factors are not perfectly perpendicular, which means they can be correlated themselves. Hence, an oblique rotation is chosen to correct for this effect and to arrive at uncorrelated constructs. The direct oblimin method is suited for this task; its only parameter, Delta, is set to -1 to decorrelate medium-sized factor correlations (Heck 1998).

The EFA was calculated using SPSS Statistics 19. The variables, the extracted factors, the factor loadings, the MSA and the communalities for each variable can be found in Table 4.10.

Table 4.10: Assessment of content validity

Factor	Variable	Factor loadings					MSA	Communalities
		1	2	3	4	5		
Fear of Provider Opportunism	SEO1			0.69			0.62	0.49
	SEO2			0.74			0.66	0.58
Information Security Concerns	ISC1					0.67	0.82	0.58
	ISC2					0.82	0.76	0.68
Attitude	ATT1	0.70					0.90	0.67
	ATT2	0.75					0.90	0.67
Intention	INT1	0.80					0.84	0.81
	INT2	0.76					0.85	0.74
Perceived Behavioral Control	PBC1			0.77			0.81	0.72
	PBC3			0.76			0.77	0.61
Infrastructure Specificity	INS1		0.71				0.75	0.52
	INS2		0.75				0.75	0.63
	INS3		0.78				0.74	0.61
	INS4		0.74				0.75	0.55
Ease of Use	EOU1			0.67			0.83	0.54
	EOU2			0.59			0.84	0.56

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Factor loadings <0.3 are hidden

Using five extracted factors, 62.2% of the cumulative variance can be explained. Factor loadings smaller than 0.3 are suppressed in the output. All variable-specific MSA values are well above the recommended value of 0.5 (Weber and Mühlhaus 2010). All communalities are very close to or above the critical value of 0.5, so the 16 variables shown were retained for further analysis. During the first round of factor analysis,

it became clear that PBC2 has to be removed due to its small communality (<0.5) (Weiber and Mühlhaus 2010). Regarding the reflective indicators, the results of the explorative factor analysis indicate item reliability. All indicator loadings are well above 0.4 (Krafft, Götz, and Liehr-Gobbers 2005), otherwise they would have to be removed.

Uni-dimensionality for the theoretical constructs can directly be shown for the infrastructure specificity construct (which loads on factor 2), fear of provider opportunism (which loads on factor 4) and information security concerns (which loads on factor 5). Convergent validity was demonstrated by the indicator's high loadings on its own dimension in the model, in which all are close to the acceptable value of 0.7 (Krafft, Götz, and Liehr-Gobbers 2005). The indicators of intention and attitude charge on factor 1 and the indicators of perceived behavioral control and ease of use charge on factor 3. Therefore, uni-dimensionality could not be proven for these constructs using an explorative factor analysis.

The literature (Hildebrandt and Temme 2006, p. 12) recommends a confirmatory factor analysis to demonstrate the correct specification of the constructs. If the global fit measures of the model are satisfactory and the residual covariances of the indicators are small, then the uni-dimensionality can nonetheless be assumed, according to Hildebrandt and Temme (2006). This approach rests on the assumption that the constructs and their measurements models under analysis are theoretically well established, otherwise the usage of a structure-checking method like CFA would be pointless (no structure to confirm). Alternatively, a new construct plus its measurement model would have to be developed; a CFA would then be necessary to prove its validity. However, this assumption can be justified in this case, as the constructs are well-researched (section 3.4.1). The conflicting four constructs were tested together using the model shown in Figure 4.8.

The model is a fully connected graph, as the goal is construct assessment and not the validation of specific construct interdependencies. Hence, the weights of the relationships among the constructs play no role in the further investigation. The construct-specific measurement indicators (rectangular boxes) are attached to the corresponding constructs (ovals), each indicator possesses its own error term. A maximum-likelihood estimator was chosen, as it is one of the most commonly used methods in social science research for causal analysis (Weiber and Mühlhaus 2010, p. 155). The model fit was assessed using the inferential statistics standard criteria Chi-Square, RMSEA (Root Mean Square Error of Approximation) and Hoelter (Weiber and Mühlhaus 2010, pp. 160). The calculations were executed using SPSS AMOS 19.⁸

Table 4.11 shows the results of the test statistics. CMIN is the value of the Chi-Square test function, P is the probability level which describes the probability that rejecting the null hypothesis is a wrong decision. The null hypothesis of the Chi-Square test assumes the equality of the empirical and the model-based covariance matrices. In this case, the probability level is 0.20 which leads to the acceptance of the null hypothesis and the conclusion that the model fit is satisfactory (according to (Weiber and Mühlhaus 2010, p. 161), the null hypothesis is rejected for $P < 0.1$). The CMIN value corrected by the degrees of freedom (CMIN/DF) also shows a satisfactory level according to (Weiber and Mühlhaus 2010, p. 162). The other test criteria also point in the same direction: the RMSEA value is 0.03 which has to be interpreted as a close model fit (Weiber and Mühlhaus 2010, p. 162). The two margins for the RMSEA confidence interval support this conclusion. The p-value of $RMSEA \leq 0.05$ (PCLOSE) is 0.73, hence the null hypothesis $RMSEA \leq 0.05$ cannot be rejected. The Hoelter test computes the minimum sample size under which the Chi-Square test would not be significant for an error level of $\alpha = 0.05$. Hoelter (1983) gives a rule of thumb for the value of the critical sample size N, which should exceed 200 in order to arrive at adequately fitting models. This value is reached here. Table 4.12 lists the residual covariances, which are small.

Hence, it can be concluded that the conflicting constructs intention, attitude, perceived behavioral control and ease of use charge are indeed uni-dimensional, which is the basis for all further reliability assessments. This conclusion is further supported by the factor loadings calculated during the PLS iterative

⁸<http://www-142.ibm.com/software/products/us/en/spss-amos>, last accessed 2013-12-29

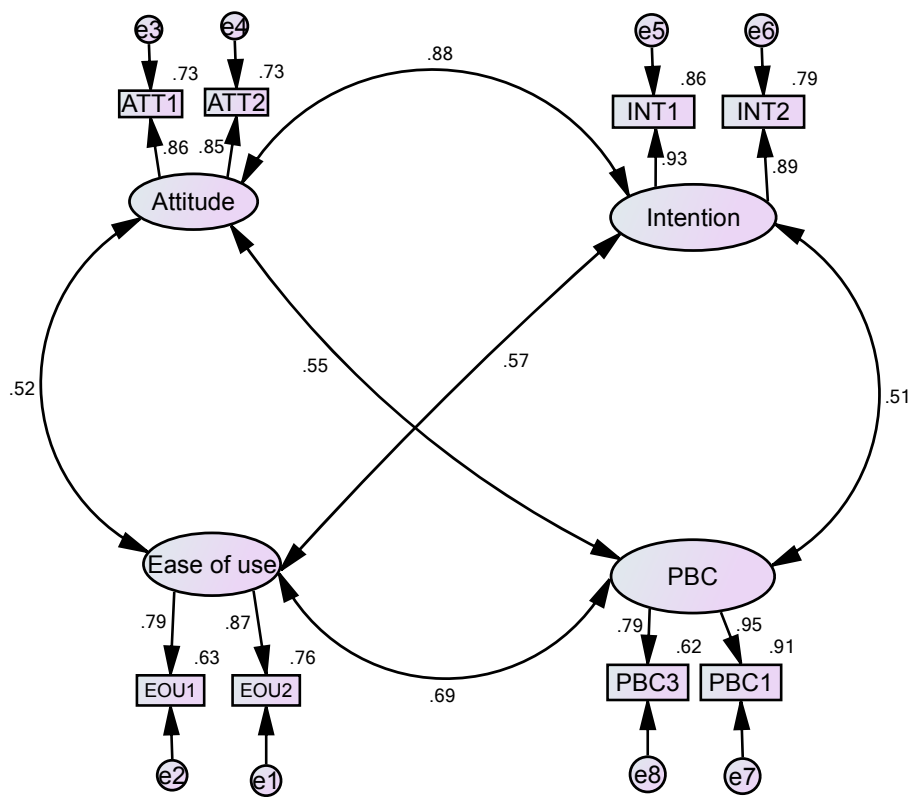


Figure 4.8: Conflicting factors tested in the CFA

Table 4.11: Model fit measures for the CFA

Test	Criterion	Value
Chi-Square	CMIN	18.11
	p	0.20
	CMIN/DF	1.29
RMSEA	RMSEA	0.03
	Low margin 90% confidence interval	0.00
	High margin 90% confidence interval	0.07
	α for RMSEA \leq 0.05	0.73
Hoelter	HOELTER ($\alpha = 0.05$)	360.00

Table 4.12: Standardized Residual covariances

	EOU1	EOU2	PBC3	PBC1	INT2	INT1	ATT2	ATT1
EOU1	0							
EOU2	0	0						
PBC3	1.179	-0.261	0					
PBC1	0.2	-0.195	0	0				
INT2	-0.502	0.563	-0.238	0.473	0			
INT1	-0.727	0.251	-0.401	-0.189	0	0		
ATT2	-0.652	0.478	-0.627	-0.05	-0.106	0.119	0	
ATT1	-0.563	0.229	-0.322	0.228	-0.002	-0.051	0	0

estimation algorithm; these loadings are listed in Table 4.13 and show a simple structure of the loadings matrix and hence prove the uni-dimensionality of the indicators.

The following paragraphs proceed with the quality assessment of indicator and construct reliability for reflective constructs. As table 4.14 shows, the reliability criteria are fulfilled for all reflective constructs. Composite reliability is >0.6 for all constructs, Cronbach's alpha is >0.7 for all constructs except SEO. As SEO still score 0.69, it remains in the analysis. AVE is consistently higher than 0.5 for all constructs. All values of composite reliability and average variance extracted were considered satisfactory. The values of composite reliability at 0.81 or above exceeded the recommended level of 0.7 (Chin 1998). Hence, indicator and construct reliability can be assumed for the reflective constructs in this model.

Table 4.15 shows the correlations of the reflective latent variables. The numbers printed in bold above the main diagonal are the squared correlations. The Fornell-Larcker criterion grants discriminant validity if all squared correlations of a certain variable with all other variables are smaller than the AVE of this certain variable (Krafft, Götz, and Liehr-Gobbers 2005). This criterion is true for all reflective variables used in this model. Therefore, discriminant validity is successfully verified.

The nomological validity of the reflective measurement model is tested using the constructs and the predicted relationships of the two theories TPB and TRA. The resulting structural model can be seen in Figure 4.9. The model fit characteristics are listed in Table 4.16 and show a close model fit. Table 4.17 lists

Table 4.13: Assessment of PLS loadings for reflective constructs

	Attitude towards IaaS usage [ATT]	Ease of use [EOU]	Fear of Provider Opportunism [SEO]	Information Security Concerns [ISC]	Infrastructure Specificity [INS]	Intention of IaaS Usage [INT]	Perceived Control [PBC]
ATT1	0.93						
ATT2	0.93						
EOU1		0.90					
EOU2		0.94					
INS1					0.84		
INS2					0.86		
INS3					0.80		
INS4					0.78		
INT1						0.96	
INT2						0.96	
ISC1				0.94			
ISC2				0.83			
PBC1							0.95
PBC3							0.91
SEO1			0.87				
SEO2			0.88				

Table 4.14: PLS Reliability Scores for Reflective Constructs

	AVE	Composite Reliability	Cronbachs Alpha	Communality
Attitude towards IaaS usage [ATT]	0.87	0.93	0.85	0.87
Ease of use [EOU]	0.84	0.91	0.82	0.84
Fear of Provider Opportunism [SEO]	0.76	0.86	0.69	0.76
Information Security Concerns [ISC]	0.79	0.88	0.75	0.79
Infrastructure Specificity [INS]	0.67	0.89	0.84	0.67
Intention of IaaS Usage [INT]	0.91	0.95	0.91	0.91
Perceived Control [PBC]	0.88	0.93	0.86	0.88

Table 4.15: Discriminant Validity: Fornell-Larcker Criterion

	Attitude towards IaaS usage [ATT]	Ease of use [EOU]	Fear of Provider Opportunism [SEO]	Information Security Concerns [ISC]	Infrastructure Specificity [INS]	Intention of IaaS Usage [INT]	Perceived Control [PBC]	AVE
Attitude towards IaaS usage [ATT]	1.000	0.183	0.034	0.197	0.001	0.587	0.217	0.866
Ease of use [EOU]	0.427	1.000	0.029	0.132	0.000	0.243	0.359	0.842
Fear of Provider Opportunism [SEO]	-0.184	-0.169	1.000	0.033	0.042	0.020	0.011	0.762
Information Security Concerns [ISC]	0.444	0.363	-0.183	1.000	0.002	0.208	0.087	0.790
Infrastructure Specificity [INS]	-0.023	-0.004	0.205	-0.039	1.000	0.006	0.009	0.671
Intention of IaaS Usage [INT]	0.766	0.493	-0.143	0.456	-0.077	1.000	0.202	0.914
Perceived Control [PBC]	0.466	0.600	-0.106	0.295	-0.097	0.450	1.000	0.880

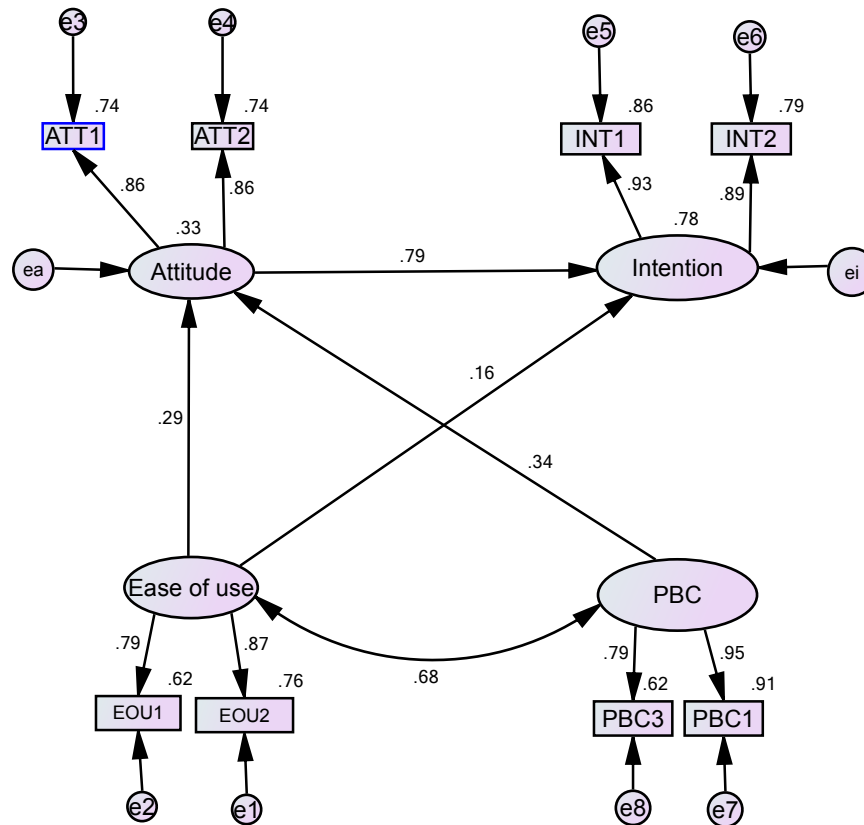


Figure 4.9: Nomological Model for TPB, TRA theories

the standardized residual covariances. No extreme values can be observed, which adds to the conclusion of a close model fit. Hence, the nomological validity of the TPB and TRA constructs can be assumed. The test for nomological validity concludes the quality assessment of the reflective constructs. All test criteria (unidimensionality, construct reliability, content validity, construct validity) given in Table 4.9 were successfully executed and the results unanimously show the high reliability and validity of the measuring models.

The following paragraphs are dedicated to the quality assessment of the formative measurement models. The required tests are described in Table 4.9 above. Regarding the formative indicators, experts in IS research checked the indicators and verified the measurement models (as recommended by Krafft et al. (2005)). During the pretest, the questionnaire containing the indicators was discussed twice with an academic expert panel consisting of members of the Chair Prof. Weinhardt; also, two external reviewers (Ludwig-Maximilians-University Munich and University of Augsburg) submitted their input via e-mail. Each feedback item was recorded and the questionnaire was reworked accordingly. Moreover, the appropriate instrument construction (theory-based for perceived usefulness, literature-base for perceived uncertainty and subjective norm, please see 4.1) also helps to ensure that the construct indicators represent the factual-semantic domain of the construct and that the indicators cover all defined facets of meaning in a construct. Therefore, content validity of the formative measurement models is assumed.

As a first collinearity check, the variance inflation factor (VIF) is calculated per indicator. Table 4.18 lists the VIFs for the formative constructs used in the model (the VIFs are computed within each construct). According to Krafft et al. (2005) and Weiber and Mühlhaus (2010), VIF values below 3 are satisfactory,

Table 4.16: Global Fit Measures for the Nomological Model

Test	Criterion	Value
Chi-Square	CMIN	19.82
	P	0.18
	CMIN/DF	1.32
RMSEA	RMSEA	0.03
	Low margin 90% confidence interval	0.00
	High margin 90% confidence interval	0.07
	α for RMSEA ≤ 0.05	0.72
Hoelter	HOELTER ($\alpha = 0.05$)	347.00

Table 4.17: Standardized Residual Covariances for the Nomological Model

	EOU1	EOU2	PBC3	PBC1	INT2	INT1	ATT2	ATT1
EOU1	0							
EOU2	-0.022	0						
PBC3	1.235	-0.237	0					
PBC1	0.259	-0.172	0	0				
INT2	-0.447	0.592	-0.536	0.112	0			
INT1	-0.65	0.302	-0.692	-0.536	0	0		
ATT2	-0.673	0.425	-0.507	0.089	-0.11	0.144	0	
ATT1	-0.574	0.188	-0.189	0.382	0.012	-0.008	-0.036	0

Table 4.18: Variance Inflation Factor Data for the Formative Constructs

Construct	Indicator	Adj. r^2	VIF
Subjective Norm	SNO1	0.420	1.724
	SNO2	0.427	1.745
	SNO3	0.499	1.996
	SNO4	0.441	1.789
	SNO5	0.380	1.613
	SNO6	0.408	1.689
	SNO7	0.572	2.336
Perceived Uncertainty	SEU1	0.239	1.314
	SEU2	0.264	1.359
	SEU3	0.336	1.506
	SEU4	0.146	1.171
	SEU5	0.184	1.225
	SEU6	0.310	1.449
	SEU7	0.257	1.346
	SEU8	0.325	1.481
Perceived Usefulness	REL1	0.426	1.742
	REL2	0.505	2.020
	REL3	0.416	1.712
	REL4	0.579	2.375
	REL5	0.592	2.451
	REL6	0.417	1.715

so the correlations among the indicators for each formative construct do not threaten further analysis steps. All VIFs remain below this threshold.

The second collinearity check consists of the indicator correlation matrix (please see Table 4.19). The correlation values are mostly small; correlations >0.5 are printed in bold. They appear mostly in the constructs “Perceived Usefulness” and “Subjective Norm”. As most of the problematic indicators will be removed in the next step anyway, no indicators have to be combined here.

Determining the indicator validity is the next necessary step; it can be demonstrated using estimates for the prognostic validity of each indicator; significant and important regression coefficients in formative constructs are a sign of prognostic validity. As the PLS approach makes no assumption about the statistical distribution of the underlying data, a parametrical test cannot be used for checking the significance of the path coefficients of each indicator. However, resampling methods allow the non-parametric estimation of regression parameters and their confidence intervals and hence open up the possibility to check statistical significance (Huber et al. 2007, p. 10). This method replaces the missing theoretical distribution function

Table 4.19: Correlation Matrix for the Formative Indicators

	SNO1	SNO2	SNO3	SNO4	SNO5	SNO6	SNO7	SEU1	SEU2	SEU3	SEU4	SEU5	SEU6	SEU7	SEU8	REL1	REL2	REL3	REL4	REL5	REL6	
SNO1	1.00																					
SNO2	0.26	1.00																				
SNO3	0.15	0.54	1.00																			
SNO4	0.26	0.53	0.53	1.00																		
SNO5	0.60	0.06	0.11	0.20	1.00																	
SNO6	0.33	0.47	0.51	0.49	0.26	1.00																
SNO7	0.28	0.57	0.66	0.61	0.15	0.56	1.00															
SEU1	0.05	0.10	-0.03	0.06	0.02	0.06	0.02	1.00														
SEU2	0.07	0.06	-0.05	-0.02	0.08	0.01	-0.03	0.23	1.00													
SEU3	0.02	0.02	-0.07	-0.07	0.07	-0.02	-0.06	0.34	0.38	1.00												
SEU4	-0.01	-0.09	-0.09	-0.17	0.04	-0.11	-0.16	0.22	0.27	0.32	1.00											
SEU5	0.02	0.02	-0.06	-0.01	0.01	-0.01	-0.03	0.30	0.28	0.28	0.23	1.00										
SEU6	0.08	0.01	-0.04	-0.03	0.13	-0.03	-0.04	0.29	0.29	0.48	0.22	0.32	1.00									
SEU7	0.06	0.01	-0.12	-0.03	0.01	-0.04	-0.01	0.42	0.29	0.32	0.14	0.27	0.34	1.00								
SEU8	0.09	0.02	-0.08	0.00	0.11	-0.01	-0.09	0.31	0.45	0.33	0.30	0.33	0.40	0.36	1.00							
REL1	0.28	0.29	0.32	0.34	0.19	0.38	0.35	0.08	-0.06	0.00	-0.08	-0.12	-0.03	-0.01	-0.02	1.00						
REL2	0.25	0.30	0.27	0.28	0.21	0.35	0.28	0.02	-0.05	-0.09	-0.02	-0.05	-0.10	-0.08	-0.09	0.59	1.00					
REL3	0.23	0.24	0.31	0.27	0.09	0.21	0.35	-0.05	-0.16	-0.13	-0.05	-0.06	-0.12	-0.16	-0.22	0.48	0.48	1.00				
REL4	0.21	0.35	0.33	0.32	0.06	0.35	0.41	-0.08	-0.26	-0.14	-0.17	-0.19	-0.16	-0.12	-0.34	0.48	0.56	0.60	1.00			
REL5	0.22	0.32	0.27	0.30	0.08	0.21	0.35	-0.09	-0.18	-0.11	-0.14	-0.12	-0.05	-0.06	-0.22	0.45	0.59	0.55	0.71	1.00		
REL6	0.21	0.27	0.24	0.29	0.14	0.26	0.35	0.02	0.00	0.02	-0.08	0.01	0.05	-0.02	-0.07	0.53	0.57	0.45	0.52	0.59	1.00	

of the data with the empirically calculated distribution function (under the assumption that the survey data is statistically representative of the population) (Efron 1979).

For this analysis, the Bootstrapping method was performed on the measurement models to test the statistical significance of regression coefficients using t-tests (Krafft, Götz, and Liehr-Gobbers 2005). Bootstrapping consists of drawing repeated samples with replacement from the empirical data set and using them to calculate test statistics (Weiber and Mühlhaus 2010, p. 256). These single test statistics are then averaged to arrive at an overall test statistic.

Table 4.20 shows the results of the bootstrapping method applied to the formative measurement models. Each bootstrap sample has the size of the original data set; 800 of these samples were taken for these calculations. These parameters are in line with the recommendations in the literature (Weiber and Mühlhaus 2010, p. 256). The last column in the table lists the status of each indicator; all indicators but REL1, SEU7, SEU8 are modestly or highly significant. All path coefficients (except REL1's one) are sizable (>0.1). REL1 is a weak indicator but should not be eliminated, as its path coefficient is not negligible and contributes to the overall explanatory value of the construct (corresponding tests with and without the indicator were executed on the full model). This reasoning is in line with Weiber and Mühlhaus (2010, p. 256), who point out that an indicator must not be removed only for statistical reasons; the deletion has to be backed up by theoretical arguments as well.

All other indicators were removed from the formative measurement models for both statistical and theoretical reasons. This indicator deletion poses a serious problem, as each indicator in a formative construct defines a semantical facet of the hypothetical construct; if an indicator is removed, the semantic content of the construct is changed. This may come as a threat to the content validity of the affected construct. Therefore, the removed indicators warrant a closer inspection of their semantic contribution to the construct.

In the case of the construct "Subjective Norm", the survey asks the informant to estimate whether certain institutions/persons recommend or reject the IaaS usage for his/her enterprise. The indicator for the IT trade press (SNO5) was removed; this is plausible as the influence of the IT trade press on IT decision makers is supposedly limited. The indicator for colleagues of the informant (SNO7) was removed as it exhibited strong correlations with several other indicators of the same construct (please see Table 4.19). This can be explained by the ambiguity of the term "colleagues": other indicators already ask for potential specific colleagues of the informant (e.g. employees of the IT department), such that the question for general colleagues is redundant.

In the case of the construct "Perceived Uncertainty", the informant had to estimate whether the indicators were easy or hard to ascertain when using IaaS resources in the enterprise. The indicator for technical difficulties when integrating IaaS resources in the enterprise IT landscape (SEU1) had to be removed; this seems plausible, as IaaS resources are technically standardized and their specifications are publicly available, hence the technical uncertainty induced by them should be minimal. The indicator for measuring the appearance of a lock-in effect with the IaaS provider (SEU5) was also deleted; the same line of reasoning shown for SEU1 can be applied accordingly here.

The construct "Perceived Usefulness" measures the comparative benefits of an IaaS enterprise usage. The indicator for an increased effectiveness (REL5) had to be removed (effectiveness means better attainment of a given goal). It can be argued that IaaS itself does directly contribute to business goals due to its commodity nature (Carr 2003). (However, it can make businesses more efficient). The indicator for organizational added-value (REL6) was also deleted; a direct relationship of enterprise IaaS usage and enterprise organization can be questioned due to the same reasons given for REL5. (An organizational added-value comprises the possibilities to apply new organizational structures through the deployment of Cloud infrastructure (Bazijanec, Pousttchi, and Turowski 2004).) The aforementioned results clearly define the formative constructs and demonstrate the high validity of their indicators.

Table 4.20: Formative Indicator Prognostic Validity

Construct	Indicator	Regr. Coeff.	t-value	t-Test result
Subjective Norm	SNO1	0.149	1.782	+
	SNO2	0.260	2.588	**
	SNO3	0.174	1.688	+
	SNO4	0.414	4.369	**
	SNO6	0.327	3.199	**
Perceived Uncertainty	SEU2	0.266	2.258	*
	SEU3	0.312	2.350	*
	SEU4	0.365	2.372	*
	SEU6	0.253	1.829	+
	SEU7	0.151	1.433	
	SEU8	0.153	1.229	
Perceived Usefulness	REL1	0.075	0.688	
	REL2	0.342	3.388	**
	REL3	0.244	2.646	**
	REL4	0.538	5.426	**
+ significant for $\alpha = 0.1$				
* significant for $\alpha = 0.05$				
** significant for $\alpha = 0.01$				

The last step in the validity assessment of the formative measurement models consists of the construct validity assessment. Construct validity is analyzed by checking the interdependencies of the formative constructs to each other and to other constructs in the full structural model. According to Diamantopoulos and Winklhofer (2001, pp. 272), if the full structural model is estimated using PLS and these interdependencies (path coefficients) turn out to be statistically significant and exhibit the theoretically predicted direction, then the nomological validity can be assumed. If furthermore the coefficient of determination R^2 of the formative constructs is sufficiently large (>0.3) (Chin 1998), then construct validity is assumed.

Figure 4.10 reveals the coefficient of determination, the significance level and the path coefficients of the formative constructs (among others) and demonstrates that the requirements of construct validity are fulfilled for all formative constructs.

4.4.4 Quality Assessment of Structural Model

Following Hair et al. (2006), it is essential to analyze the structural model after the validity of the measurement models has been established. Table 4.21 lists the criteria that need to be checked to ensure the quality of the structural model. These single, independent tests are necessary, as there is no global criterion to judge model validity. If all tests return a satisfactory result, the model as a whole is deemed reliable (Weiber and Mühlhaus 2010, p. 259).

Table 4.21: Quality Assessment Criteria for Structural Models (Weiber and Mühlhaus 2010, p. 259)

Quality criterion	Metric	Thresholds	References
Construct assessment	Coefficient of determination R^2 of the endogenous constructs	Interpretation of R^2 value: >0.67 substantial [0.33; 0.67[average [0.19; 0.33[weak	Chin (1998, p. 323)
	Stone-Geisser test criterion Q^2 for endogenous reflective constructs	If $Q^2 > 0$, the model possesses predictive power.	(Fornell and Bookstein 1982, p. 440)
Assessment of Path coefficients	Standardized β values	$\beta > 0.1$	Lohmöller (1989, pp. 60)
	t-test statistics	t-values according to a two-sided t-test (e.g. 1.65 for $\alpha = 0.1$) for null hypothesis $\beta = 0$	(Weiber and Mühlhaus 2010, p. 259)
	Effect size f^2 of an exogenous variable on an endogenous variable	Interpretation of f^2 value: >0.35 large [0.15; 0.35[medium [0.02; 0.15[small	Chin (1998, p. 316)

Figure 4.10 presents the results of the analysis with overall coefficients of determination, estimated path coefficients β , and associated significance test results. According to the interpretation of Chin (1998, p. 323), all endogenous constructs but “Perceived Uncertainty” reach an average explanatory power; all endogenous constructs are clearly above the threshold of $R^2 \geq 0.19$, as requested by Chin (1998, p. 325).

All path coefficients except one are greater than 0.1 and are significant at least on an $\alpha = 0.10$ error level. “Perceived Uncertainty” fails to exert any meaningful influence on the IaaS usage intention. The path remains in the structural model for two reasons: first, “Perceived Uncertainty” and “Intention of IaaS usage” are both important construct and their construct reliability and validity metrics were sufficient so

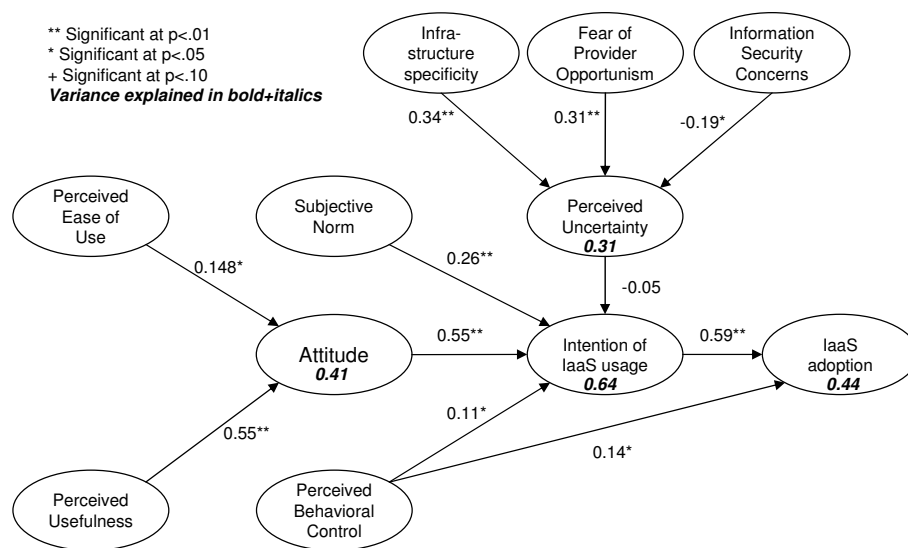


Figure 4.10: Results of PLS analysis of the structural model

far; hence, it can be assumed that they are correctly measured. Second, the relationship between “Perceived Uncertainty” and “Intention of IaaS usage” has been established theoretically (please see section 3.4.2); therefore, this negative result warrants a closer inspection and explanation. In section 4.4.5, the role of “Perceived Uncertainty” is analyzed further using moderating variables; the corresponding results partly explain why no direct effect between “Perceived Uncertainty” and “Intention of IaaS usage” was measured.

The effect size f^2 allows the assessment whether an exogenous latent variable exerts a measurable effect on an endogenous latent variable (Weiber and Mühlhaus 2010, p. 257) (in addition to the path coefficients). Cohen (1988, p. 410) defined this metric for multiple regression and correlation analysis scenarios. Equation 4.1 shows the formula. R^2_{incl} is the coefficient of determination of the endogenous variable j if all exogenous latent variables are present. If the exogenous latent variable i is removed, then R^2_{excl} is the resulting coefficient of determination.

$$f_{ij}^2 = \frac{R^2_{incl} - R^2_{excl}}{1 - R^2_{incl}}. \quad (4.1)$$

The effect size measures f^2 are listed in the last column in table 4.22. Their interpretation follows Chin (1998, p. 316). It is remarkable that mainly the TPB constructs show a large effect size, and that the “Perceived Usefulness” construct (REL) also seems to be very important. All other effect sizes are rather small, but existent (>0). Krafft, Götz, and Liehr-Gobbers (2005, p. 85) consider existent effect sizes as one of the validity criteria for structural models. Only “Perceived Behavioral Control” entirely fails to influence its endogenous latent variable “Intention of IaaS usage” (despite its significant path coefficient). This observation does not lead to the removal of “Perceived Behavioral Control” from the model, as the construct exerts a non-negligible effect on IaaS adoption (ADO). The removal of the path between “Perceived Behavioral Control” and “Intention of IaaS usage” would be an option, but as the other validity criteria are fulfilled, the path remains in the model.

The Stone-Geisser Criterion Q^2 (Geisser 1974), (Stone 1974) can be used to judge the prognostic relevance of reflective, endogenous latent variables (Fornell and Bookstein 1982, p. 450). Its application to PLS and its detailed derivation are described by (Fornell and Cha 1994); they also give the mathematical

Table 4.22: f^2 effect size measures

Path from...	...to	R^2_{incl}	R^2_{excl}	f^2	Interpretation
INT	ADO	0.445	0.168	0.499	large
ATT	INT	0.643	0.475	0.471	large
REL	ATT	0.407	0.183	0.378	large
INS	SEU	0.314	0.220	0.137	small
SEO	SEU	0.314	0.226	0.128	small
SNO	INT	0.643	0.600	0.120	small
PBC	ADO	0.445	0.428	0.031	small
ISC	SEU	0.314	0.293	0.031	small
EOU	ATT	0.407	0.390	0.029	small
SEU	INT	0.643	0.641	0.006	small
PBC	INT	0.643	0.643	0.000	small

formulation. In essence, the criterion compares the differences between the latent variable sample data and the latent variable estimates from the structural model with the differences between the latent variable sample data and a latent variable estimate based on its indicator averages. The criterion states how well a latent variable can be reconstructed by its estimated indicators (Huber et al. 2007, p. 37).

During the calculation, it is assumed that parts of the sample data are missing and have to be estimated based on the parameters of the structural model that were computed earlier. This procedure is called blindfolding (Weiber and Mühlhaus 2010, p. 258), whose foundations are described in (Tenenhaus, Vinzi, Chatelin, and Lauro 2005, pp. 174). SmartPLS implements a blindfolding approach and returns a cross-validated redundancy score, that is equivalent to the Stone-Geisser criterion. The SmartPLS blindfolding approach possesses one parameter, omission distance. It signifies the number of partitions that are used during the blindfolding procedure; the sample values of each partition are deleted and are estimated using the remaining values. This is done for every partition and the results of each partition are combined in the cross-validated redundancy score (Tenenhaus, Vinzi, Chatelin, and Lauro 2005, p. 176). The omission distance is set to seven, as recommended by (Wold 1982). For mathematical reasons, the number of observations must not be a multiple of the omission distance, which is not the case in this study.

The construct-related Stone-Geisser criterion $Q^2 > 0$ is defined in (Fornell and Bookstein 1982, p. 440); if this condition holds, the model is granted prognostic relevance. The only reflective, endogenous constructs are Attitude (ATT), Intention of IaaS usage (INT) and IaaS adoption (ADO). In Table 4.23, the construct-related Stone-Geisser criterion is listed in column 3 (Q^2_{incl}); the values indicate a good prognostic relevance for these constructs.

The Stone-Geisser criterion $q^2 > 0$ can also be calculated for paths among reflective constructs (Chin 1998, p. 318). There, the Q^2 values for the endogenous latent variables j are determined twice: once including all exogenous latent variables and once excluding the specific exogenous latent variable i forming the path. These two Q^2 values are the input for the equation 4.2. If $q^2_{ij} > 0$, then a high prognostic relevance of the excluded exogenous latent variable can be assumed (Chin 1998). In Table 4.23, all exogenous latent variables in column 1 fulfill this condition.

$$q^2_{ij} = \frac{Q^2_{incl} - Q^2_{excl}}{1 - Q^2_{incl}}. \quad (4.2)$$

Table 4.23: Path-related Stone-Geisser criterion for prognostic relevance q^2

Path from...	...to	Q_{incl}^2	Q_{excl}^2	q^2
REL	ATT	0.349	0.157	0.295
EOU	ATT	0.349	0.336	0.020
SNO	INT	0.587	0.547	0.097
ATT	INT	0.587	0.430	0.380
PBC	INT	0.587	0.577	0.024
SEU	INT	0.587	0.584	0.007
PBC	ADO	0.435	0.422	0.023
INT	ADO	0.435	0.162	0.483

As a conclusion, the comprehensive validity criteria given in Table 4.21 are generally fulfilled by the presented structural model. Nevertheless, there are several exceptions (e.g. the missing uni-dimensionality for four constructs, the necessary removals of indicators of formative measurement models or the missing effect size on PBC), but these do not lead to the rejection of the model, as they can either be remedied with additional statistical analyses or they are compensated by other validity measures or theoretical considerations based on existing research. Overall, the structural model can be considered valid and delivers a good fit for the survey data.

4.4.5 Moderating Effects

As mentioned above, the role that the construct “Perceived Uncertainty” plays in IaaS adoption has to be inspected more closely. The last section established the validity of the structural model, hence the latent variable scores estimated using the PLS approach can be utilized to further analyze the perceived uncertainty of different informant groups. The calculations are based on unstandardized scores, as Henseler and Fassott (2010, p. 728) argue that unstandardized latent variable scores should be used if the researcher is interested in interpreting the outcomes in terms of the original scales. It is assumed that the attributes of each informant group (e.g. company size, IT affiliation) act as a moderating effect on the relationship between perceived uncertainty and the intention to use IaaS in the enterprise. According to the definition found in the research literature, a moderator is a “variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable” (Baron and Kenny 1986, p. 1174).

Moderating effects belong to a type of interaction effects, which include more general causal relationships in structural models. The importance of analyzing these interaction effects is emphasized by Chin, Marcolin, and Newsted (2003), who see them as a natural step towards more complex IS theories. The importance stems from the observation that direct causal relationships are often well-known or trivial, whereas interaction effects elucidate the situational effects on a causal relationship. However, Chin, Marcolin, and Newsted (2003) also illustrate using a literature review the small share of researcher that actually investigate moderating relationships in their research papers.

As Henseler and Fassott (2010) point out, there are principally two approaches to estimating moderating effects in a PLS setting: the product-term approach and the group comparison approach. In the case of at least one formative construct, the product-term approach is calculated in two stages, hence it is usually referenced in the research literature as the two-stage approach (Henseler and Fassott 2010, p. 725). They recommend using the product-term approach for its better result quality (Henseler and Fassott 2010, p. 721), although Henseler and Fassott (2010, p. 721) argue that the group comparison approach can be considered, “if the moderator variable is categorical or if the researcher wants a quick overview of a possible moderator

effect” (Henseler and Fassott 2010, p. 721). Other researchers (e.g. Reinecke (1999)) are not as dismissive of the group comparison approach and see it as a first identification step for moderating effects. As the moderator variables in this study are categorical, manifest variables, the group comparison approach is chosen for all further analyses.

The analysis of moderating effects is split up in two parts: first, an ANOVA is executed that investigates the effects of company size and IT affiliation on a subset of the hypothetical constructs (section 4.4.2 also contains ANOVA calculations, however these are related to measurement indicators that were not used in the causal model). Second, two group comparisons are conducted to estimate the moderating effects of company size and IT affiliation on the relationship between perceived uncertainty and intention to use IaaS.

The selection of the six constructs for further analysis is mainly motivated by the need to understand the properties of perceived uncertainty SEU, hence the three antecedents ISC, SEO, INS and the endogenous variable INT are included. The perceived usefulness REL is one of the strongest influences in the whole structural model and a deeper understanding should prove worthwhile.

The application of ANOVA is tied to the fulfillment of a number of statistical preconditions (Backhaus et al. 2006, p. 177); for the one-way ANOVA, the dependent variable has to be metric and normally distributed and the subgroups have to exhibit homogeneity of variances. Table 4.24 shows the results of the test for normality (Kolmogorov-Smirnov Test). All variables pass the test (the null hypothesis cannot be rejected on an $\alpha \leq 0.1$ error level).

Table 4.24: One-Sample Kolmogorov-Smirnov Test

	ISC	SEO	SEU	REL	INT	INS
N	276.00	276.00	276.00	276.00	276.00	276.00
Mean	2.89	3.48	4.23	4.16	3.76	3.69
Std. Deviation	1.15	1.06	0.82	1.07	1.39	1.13
Kolmogorov-Smirnov Z	1.59	1.76	1.29	1.35	2.13	1.27
Asymp. Sig. (2-tailed)	0.01**	0.00***	0.07 ⁺	0.05*	0.00***	0.08 ⁺

After the completion of the ANOVA, the moderating effect of the company size on the different constructs is analyzed using a group comparison. The company size is coded as a two-level factor (SME, large entity); the exact definition of this factor can be found in section 4.4.2. Table 4.25 shows the descriptive statistics of the latent variable scores grouped by company size and the test result of the mean difference significance test ($+$: $\alpha \leq 0.1$; $*$: $\alpha \leq 0.05$; $**$: $\alpha \leq 0.01$). Table 4.26 lists an overview of the test statistics. In case the Levene test led to the rejection of the null hypothesis (e.g. for REL), the test results of the Mann-Whitney-U (Mann and Whitney 1947) test are reported; mostly, these are identical to the ANOVA test results. The Mann-Whitney-U test is a non-parametric test, where the dependent variable does not have to be normally distributed.

The interpretation of the results is straightforward: SME exhibit a greater trust in using IaaS Cloud providers (higher value of ISC) and also perceive less uncertainty than larger enterprises. This might be due to the fact that their infrastructural needs are less elaborated than those of larger companies (lower value of INS) and that they are better able to extract value from their IaaS deployments (higher value of REL). Both groups have comparable opinions on IaaS provider opportunism (SEO) and their intention to use IaaS (INT).

Analog to the company size, the IT affiliation of the informant might also influence the perception of IaaS usage of the enterprise. His/her role is coded as a two-level factor (IT, Business); the exact definition of this factor can be found in section 4.4.2. The results are prepared and presented similarly to the tests for

Table 4.25: LV scores by company size

		ISC	SEO	SEU	REL	INT	INS
SME	Mean	3.20	3.45	4.10	4.48	3.97	3.50
	N	101.00	101.00	101.00	101.00	101.00	101.00
	Std. Deviation	1.17	0.99	0.85	0.91	1.30	1.14
Large Entity	Mean	2.79	3.50	4.32	4.02	3.68	3.86
	N	150.00	150.00	150.00	150.00	150.00	150.00
	Std. Deviation	1.13	1.08	0.76	1.11	1.44	1.07
Total	Mean	2.95	3.48	4.23	4.20	3.80	3.71
	N	251.00	251.00	251.00	251.00	251.00	251.00
	Std. Deviation	1.16	1.04	0.80	1.06	1.39	1.11
Significance		**		*	**		*

Table 4.26: Significance of company size

	Levene		Mann-Whitney-U		ANOVA	
	Test Statistic	Sig.	Test Statistic (Z)	Sig.	Test Statistic (F)	Sig.
ISC	0.07	0.79	-2.77	0.01**	7.56	0.01**
SEO	1.14	0.29	-0.24	0.81	0.13	0.71
SEU	1.73	0.19	-1.75	0.08 ⁺	4.67	0.03*
REL	6.01	0.01**	-3.33	0.00***	11.85	0.00***
INT	4.01	0.05*	-1.58	0.11	2.62	0.11
INS	0.60	0.44	-2.34	0.02*	6.60	0.01**

company size in the last paragraphs. Table 4.27 summarizes the findings; detail values for the test statistics can be found in Table 4.28.

The interpretation yields a slightly different picture compared to the company size analysis. In general, the effects of IT affiliation are less pronounced than the effects of company size across all constructs. The intention to use IaaS is significantly higher on the business side than on the IT side, which fits the fact that business informants perceive IaaS to be less insecure (higher ISC) and more valuable to the company (higher REL). These observations perfectly match the findings from section 4.4.2, where it was shown that business executives are more likely to adopt IaaS than IT leaders, and extend them by providing the underlying causes for this behaviour.

The first part containing ANOVA calculations is now complete. In the second part, two group comparisons are conducted to estimate the moderating effects of company size and IT affiliation on the relationship between perceived uncertainty and intention to use IaaS.

Table 4.29 shows the results of the group comparisons. The β values, the t-Test statistics and the standard errors in the group comparisons were estimated using the same PLS and bootstrapping parameters like the ones used in the full model.

Nitzl (2010, p. 46) provides a formula for a t-test statistic that can be used to test whether the differences found in the PLS group comparison approach are significant. The last column significance in Table 4.29 shows these calculated t-values; they have to be interpreted with 259 (IT affiliation) and 249 (Company size) degrees of freedom. For this magnitude of degrees of freedom, the Student's t distribution can be approximated with the normal distribution.

Table 4.27: LV scores by IT affiliation

		ISC	SEO	SEU	REL	INT	INS
IT	Mean	2.82	3.45	4.27	4.08	3.63	3.71
	N	199.00	199.00	199.00	199.00	199.00	199.00
	Std. Deviation	1.10	1.02	0.78	1.08	1.38	1.11
Business	Mean	3.13	3.52	4.15	4.37	4.12	3.63
	N	61.00	61.00	61.00	61.00	61.00	61.00
	Std. Deviation	1.28	1.09	0.87	1.10	1.41	1.13
Total	Mean	2.89	3.47	4.24	4.15	3.75	3.69
	N	260.00	260.00	260.00	260.00	260.00	260.00
	Std. Deviation	1.15	1.04	0.80	1.09	1.40	1.12
Significance		+			+	*	

Table 4.28: Significance of IT affiliation

	Levene		Mann-Whitney-U		ANOVA	
	Test Statistic	Sig.	Test Statistic (Z)	Sig.	Test Statistic (F)	Sig.
ISC	1.37	0.24	-1.62	0.11	3.32	0.07 ⁺
SEO	0.12	0.73	-0.51	0.61	0.21	0.65
SEU	1.41	0.24	-0.82	0.41	0.99	0.32
REL	0.14	0.70	-1.92	0.06 ⁺	3.16	0.08 ⁺
INT	0.00	0.98	-2.41	0.02 [*]	5.74	0.02 [*]
INS	0.02	0.88	-0.23	0.81	0.28	0.60

As a result, the company size can be considered a moderating effect on the relationship between perceived uncertainty and intention to use IaaS ($\alpha \leq 0.1$). Only larger companies let their perception of uncertainty negatively influence their IaaS usage intentions, whereas smaller companies (SME) are not intimidated by their perceived uncertainty; the effect on usage intention is insignificantly different from 0.

The influence of IT affiliation remains dubious, mainly because the number of business informants is rather small compared to the number of IT informants. If the sample sizes of the two groups had been comparable, a more powerful test for the significance of the moderating effect would have been possible. For the IT group, the perceived uncertainty already has a significantly negative influence on their IaaS usage intention, whereas the business group seems unaffected by their uncertainty perception.

4.4.6 Assessment of Hypotheses

The structural model identifies perceived usefulness and ease of use as drivers of attitude towards IaaS usage; it also reveals attitude, perceived usefulness, and subjective norm as strongest drivers of intention to use. The hypothesis, that Intention to use drives IaaS adoption, could not be rejected (Wiedemann and Strebel 2011). Although the hypotheses regarding the influencing factors of perceived uncertainty could not be rejected, the latter construct seems to have no direct influence on intention to use (Wiedemann and Strebel 2011). However, moderating factors can partly explain this counter-intuitive result.

Table 4.29: Group differences in the relationship between SEU and INT

		Sample size	β values	t-Test statistic (Bootstrapping)	Standard Error	Sig. (t value)
IT affiliation	IT	199	-0.08	1.715	0.05	-0.94
	Business	62	0.02	0.166	0.12	
Company size	SME	101	0.05	0.767	0.07	1.91 ⁺
	Large Entity	150	-0.11	2.144	0.05	

4.5 Discussion of Empirical Results

The sections above presented the evaluation results of the quantitative model evaluation; these findings are summarized and discussed further in light of the research questions. Also, the limitations of both research approaches are detailed, followed by the contributions of the research for both researchers and practitioners. Figure 4.11 summarizes the evaluation steps of the methodological process followed in this thesis.

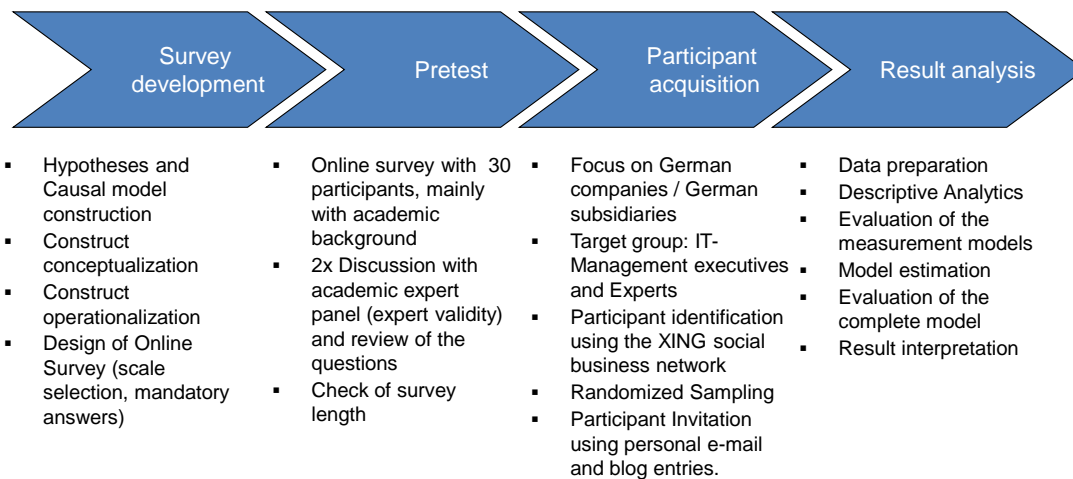


Figure 4.11: Survey methodology (based on General SEM process (Weiber and Mühlhaus 2010, p. 254))

The hypotheses in the causal model (established in section 3.4) are empirically tested by means of an online survey (N=276); the outcome is a framework explaining organization's acceptance and adoption of infrastructure resources over the Internet (IaaS). The steps "Survey development", "Pretest" and "Participant acquisition" were executed collaboratively with Carsten Frietsch and Dr. Dieter Wiedemann. C. Frietsch also documented the steps in his diploma thesis (Frietsch 2011).⁹ The step "Result analysis" consists entirely of results created solely by the author.

4.5.1 Implications

This section will give an interpretation of the empirical findings and clarify their implications for research questions R1 and R2 put forward for this thesis.

⁹The diploma thesis also contains an evaluation section which tests an alternative set of hypotheses using an early, incomplete version of the data set.

4.5.1.1 Implications for research question R1

Research question R1 was termed “What are the overall IaaS usage determinants of enterprises?”. To this end, an empirical study was thoroughly prepared and executed. The primary objective of the study was to examine organizational acceptance and adoption of IaaS in the light of a framework of various well-researched theories from multiple backgrounds (e.g. theory of planned behaviour, transaction cost theory and principal-agent theory). This framework was formulated as a structural equation model with appropriate hypothetical constructs for the usage determinants.

The evaluation of the tested model allows for the confirmation of all hypotheses developed in Section 3.4.2, except the hypothesis related to the direct influence of uncertainty on intention to use. The implications will be discussed in detail in the following paragraphs.

The structural model identifies perceived usefulness and ease of use as drivers of attitude towards IaaS usage (hypothesis H1). Ease of use measures the mental effort required to learn the usage of IaaS services and the effort required to use IaaS as intended, so usability of IaaS services is a major driver for the attitude that employees have towards this technology. Perceived usefulness, i.e. the benefits of IaaS usage, also massively impacts employees’ attitude. This construct had an overall “large” effect size, the third largest in the whole model. Especially interesting are the constituents of this construct, its significant indicators. The first one, the strategic added-value, measures advantages exceeding the operational and tactical level and influences the positioning of the enterprise in its market segment. The second one, the flexibility added-value, enables the enterprise to react elastically to varying demands for IT infrastructure. The third one, the efficiency added-value, improves the speed and/or the cost efficiency of existing processes or systems. Thus, IaaS positively influences the strategic positioning, the flexibility and the efficiency of enterprises, which in turn indirectly drives its adoption. Notable is also the fact that the efficiency added-value is the strongest of the three indicators. Thus, traditional antecedents based on the TAM explain IaaS adoption either directly or indirectly (via attitude and intention). However, the level of perceived usefulness is dependent on the company size and the IT affiliation. SME are generally better able to extract value from their IaaS deployments than larger enterprises and business executives perceive IaaS to be more valuable to the company than IT executives. Nevertheless, it is clear that a great deal of additional research is necessary to obtain a more satisfactory understanding of IaaS adoption processes, especially concerning the role of uncertainty in IaaS decision making (Wiedemann and Strebel 2011).

The hypothesis H2 regarding the influencing factors of perceived uncertainty could not be rejected; infrastructure specificity, fear of provider opportunism and information security concerns were all statistically significant drivers and partly explain the construct of perceived uncertainty. However, the level of the individual drivers again depends on the company size and the informant’s background. SME exhibit less data security concerns when using IaaS Cloud providers. This observation matches the overall lower perceived uncertainty of SMEs than of larger enterprises. This reduced uncertainty perception might also be due to the fact that their IT infrastructural demands are less specific than those of larger companies, and hence, the risk of lock-in situations and other provider dependencies is lower. In general, business informants perceive IaaS to be less insecure than IT informants. This fact might be attributable to the deeper expertise of IT informants, which allows them to more realistically judge the security risks associated with an IaaS usage.

Hypothesis H3 could not be rejected for the most part: attitude, subjective norm and perceived behavioral control have a statistically significant effect on the organization’s intention to use IaaS. Attitude possesses a “large” effect size, the second largest in the model. The hypothesis for perceived uncertainty, also a hypothesized antecedent of intention to use, needs to be rejected. Regarding the intention to use, the findings highlight that subjective norms have an important influence on behavioral intention. The formative measurement model shows that significant impacts comes from external consulting agencies, comparable

enterprises, employees in the business departments, and superiors. These persons or institutions actively exert influence and recommend the IaaS usage. However, it is interesting to see that the business department has the strongest effect and the IT department has one of the lowest effects. This observation fits the previous findings; as the IT informants generally have a more critical perception of IaaS usage, it is comprehensible that they do not actively promote the deployment of IaaS services in the enterprise (and vice versa for business executives, that are eager to deploy those services).

Regarding the perceived uncertainty construct, the findings indicate that this construct seems to have no direct influence on intention to use. However, this counterintuitive result has to be put in perspective, as the company size can be considered a moderating variable for the relationship between perceived uncertainty and intention to use IaaS. The intention to use for SMEs is not negatively affected by their level of perceived uncertainty, whereas the intention to use IaaS for larger companies is strongly negatively affected by their level of perceived uncertainty.

Perceived uncertainty is supposed to measure the difficulties in predicting environmental risks. These comprehensive risks were aggregated in different risk classes. The most important difficulties are business-related difficulties caused by price changes of IaaS services, business-related difficulties caused by changes in process-critical/ operational performance indicators (e.g. unplanned downtime), legal difficulties caused by external data storage (e.g. because of unclear legal situation regarding compliance regulations) and technical difficulties (compatibility of IaaS providers with IT standards in the enterprise). From the informants' perspective, legal difficulties were the hardest ones to predict and contributed most strongly to the perceived uncertainty. The intention to use IaaS is significantly higher on the business side than on the IT side. The company size as an independent variable does not seem to play any role in the intention to use IaaS (but plays a role as a moderating variable).

Hypothesis H4 regarding IaaS adoption could also not be rejected. Intention to use was found to be a very strong predictor of actual usage; perceived behavioral control affects the adoption only marginally. This result can be most likely attributed due to the difficulties surrounding the single-item measurement model for IaaS adoption (perceived behavioral control is a significant predictor when using both items for the adoption construct). The intention to use has a "large" effect size; the largest in the model. SMEs show a higher IaaS adoption propensity than large enterprises for both adoption measures. The analysis also reveals an overall greater propensity of business informants to adopt IaaS than IT informants for both adoption measures. These observations match the aforementioned findings regarding the role of SMEs, and can be explained by them, as they provide the underlying causes for this behaviour (the differences in attitude and intention to use between SMEs and large companies).

4.5.1.2 Implications for Research Question R2

Research question R2 is relevant for enterprises that are principally interested in IaaS, but have to yet figure out the most economical way of doing so (the importance of economics is motivated by the survey results). In the light of the previous finding, an SME would be the most likely candidate for such a decision problem (also because it would probably be unfazed by the perceived uncertainty and the commodity IT resources offered by IaaS providers).

4.5.2 Limitations of the Research Approach

4.5.2.1 Limitations of the survey

Although the survey results can be considered statistically significant in most parts, the study has several limitations. First, this study uses a more general perspective on the drivers of IaaS adoption and neglects use case types. The questionnaire does not differentiate between the provision of single IT resources, e.g.

CPU cores, storage, networks, etc., and it does not include typical Cloud Computing use cases (e.g. as identified by (Cloud Computing Use Case Discussion Group 2010)). Second, although the sample size was quite large, it consisted of German-speaking decision makers only. Thus, a replication of the study in other international markets would be desirable. Third, the survey is hampered by a rather low response rate, but this situation is typical for studies targeted at business executives (e.g. (Benlian and Buxmann 2009)). Fourth, the number of measurement items for some constructs could have been higher, especially for the adoption construct and for those reflective constructs with two measurement indicators (e.g. SEO or INT). The adoption construct especially suffered from a high number of missing values.

4.5.3 Contributions

The theoretical and practical contributions of the empirical model are discussed in the following paragraphs. Their aim is to clearly expose the knowledge increase over the state-of-the-art found in the research literature and to demonstrate the practical relevance of the findings. The contributions relating to the survey results were already published in (Wiedemann and Strebel 2011).

4.5.3.1 Contributions to Researchers

The starting point for this work's consideration was the need to understand the motivations, attitudes, and adoption behaviors of those corporate executives deciding on infrastructure service usage "out of the Cloud". In this thesis, IaaS adoption was explored on two complementary levels: first, using a multi-theoretical empirical model and second, using an outsourcing-process based analytical model. The two approaches complement each other, as the empirical approach analyzes the general overall factors and the analytical approach models the specific outsourcing situation. Therefore, the outcomes of both approaches are mentioned here side by side.

The study investigates an unexplored area in Cloud Computing research, the determinants of IaaS usage. Previously existing studies either focused on generic IT outsourcing determinants, or they investigated the drivers in SaaS / ASP use cases. This situation calls for a special approach from researchers: the results in outsourcing research are not specific enough to understand the Cloud Computing phenomenon and it can be argued that Cloud Computing is an outsourcing special case. SaaS/ASP use cases are not directly comparable to IaaS ones, as IT infrastructure is seen as a commodity in the research literature, whereas business application software is defined by its process-oriented functionality and must usually be customized.

The proposed theory developed in section 3.4.2 provides researchers with a useful first step to better understand the decision makers' behavior in potential IaaS sourcing scenarios (Wiedemann and Strebel 2011). From a theoretical standpoint, the results contribute to the existing literature in a number of ways.

First, the thesis contributes to Cloud Computing literature by providing insights on the drivers of IaaS acceptance and adoption based on an integrated model. In summary, this study has proved the validity of the TAM, TRA, and TPB for research in the area of IaaS. These findings are consistent with studies on SaaS (Benlian and Buxmann 2009) and outsourcing literature (Dibbern, Goles, Hirschheim, and Bandula 2004). In extension to (Benlian and Buxmann 2009), the study also found that subjective norm and perceived behavioral control have a significant impact on intention to use IaaS. Moreover, the thesis model allows a deeper understanding of the impact of company size and informant role on the motivations, attitudes, and behaviors of decision makers. The informant role is shown to be a relevant added value of this model over the model by Benlian and Buxmann (2009). Unlike the results by Benlian and Buxmann (2009), this study can also prove the effect of company size on the IaaS adoption decision.

Second, the findings indicate that perceived uncertainty had no direct overall influence on intention to use. This is in contradiction to many outsourcing studies (e.g. (Nam, Rajagopalan, Rao, and Chaudhury

1996)) and not consistent with the results reported in the electronic commerce literature (e.g. (Pavlou, Liang, and Xue 2007)). One possible explanation of this discrepancy may be found in the IaaS usage scenarios; the immature sourcing option IaaS may have been used up to then mainly for non-critical and commodity IT applications for which the uncertainty associated with IaaS plays no role. Further research is definitely necessary to clarify this issue. Another possible explanation may be found in Herzberg's two-factor theory (Herzberg, Mausner, and Snyderman 1959). A main assumption of the theory is that the presence of hygiene factors is necessary, but not sufficient enough to lead to job satisfaction. Consistent with this theory, uncertainty can be identified as a hygiene factor (and not as a motivation factor). Therefore, uncertainty may negatively influence an unfavorable evaluation of intention to use, but the lack of uncertainty, e.g. through mitigating with internal IT governance processes, does not positively influence a favorable evaluation of intention to use IaaS. Thus, a decision maker may or may not decide on IaaS when he feels that technological, business and legal certainty exist; however, he/she will definitely not decide on IaaS when a high level of uncertainty exists. Clearly, this two-factor hypothesis requires further theoretical analysis and empirical research, but a progress in this area might shed more light on the internal structure of the construct of perceived uncertainty.

A part of the explanation of the discrepancy is provided by the company size, which can be shown to act as a moderating variable between the perceived uncertainty and the intention to use. This result is an extension to existing studies. Benlian and Buxmann (2009) shows that the application adoption uncertainty is also statistically insignificant in the overall model, but depends on the application type. Yao (2004, p. 138) also examines the role of uncertainty, but does not execute appropriate statistical tests to further establish any connection between uncertainty and descriptive survey items (Yao 2004, p. 138).

4.5.3.2 Contributions to Practitioners

By providing practitioners with some insight into the decision makers' perception of IaaS, the survey research framework serves as a basis for the management of the so far poorly understood IaaS acceptance process and for IaaS-related outsourcing decisions. This knowledge is especially relevant for IaaS providers which aim at maximizing the effectiveness of their sales activities. But also IT executives can profit from these results by better understanding the mindset of their business departments and by reacting accordingly.

The results suggest that IaaS providers should consider the whole process of IaaS acceptance including attitude, intention, and adoption. The study helps IaaS providers better understand the critical elements of IaaS acceptance and adoption and allows better marketing of Cloud infrastructure services.

Regarding the attitude, the findings explicate a substantial influence of perceived usefulness and ease of use. The findings based on TAM are hardly groundbreaking; nonetheless, they tell IaaS providers that the basic requirements for acceptance of information technology are also essential for IaaS. The results from the formative measurement model of perceived usefulness imply that IaaS providers should pay close attention to aspects of strategic added value, flexibility added value, and/or efficiency added value. If Cloud infrastructure services are designed to provide one or more of these added values, decision makers will develop a positive attitude towards IaaS leading to the behavioral intention to use IaaS. Perceived ease of use also showed a direct effect on attitude. As a result, when promoting IaaS offers, providers should particularly highlight aspects of user-friendliness, e.g. fast deployment, easy configuration or access to online trainings and support.

The overall findings, that the peer group influences the intention to use IaaS, tells providers to use testimonials from their customers and case studies to promote IaaS offers. Moreover, IaaS providers should educate the prospective customers' business departments, which might then make a better case for IaaS sourcing and exert pressure on the internal IT department. (Usually, the client-internal IT departments are responsible for IT landscape planning and solution design.)

Regarding actual IaaS usage behavior, the results show that intention to use was a significant factor and perceived behavioral control is probably a contributing factor. A possible marketing strategy for increasing organizational adoption of IaaS through effects of perceived behavioral control could be to offer free use of the services for a trial period. This would enable potential users to learn the infrastructure services, thus increasing their perceived control of the service (Nysveen, Pedersen, and Thorbjørnsen 2005). An example of such a strategy can already be observed at the Cloud provider Amazon; it offers 750h per month of free EC2 micro VM instances for new customers over the course of a year.¹⁰

The results suggest, that IaaS services are especially suited for start-ups and small businesses, which have more favorable perceptions of data security, efficiency and uncertainty associated with IaaS services than larger companies. Therefore, these SMEs would be prime candidates for IaaS providers' marketing efforts. However, as IaaS clients try to mitigate these risks by carefully selecting their future IaaS provider, those IaaS providers, that have a reputation for being honest and that show their concern about customers, are better positioned to convince IaaS clients. Official certifications (e.g. ISO 27001) or data-center audits by external experts might signal the provider's honesty to potential customers. But the IaaS adoption model is also helpful for IT executives in enterprises that are potential IaaS clients. The outcomes indicate, that business executives and IT executives have differing perceptions of using IaaS. The IT department as a service function to the business must take this fact seriously and must proactively work with the business department to address their IT infrastructure needs. Otherwise, the business departments may use the outside option, that IaaS provides; this might hurt the long-term relevance of the in-house IT department.

¹⁰<http://aws.amazon.com/de/free/>, last accessed 2013-12-29. Micro instances are the smallest VM instance type on offer at Amazon (as of 2012-12-31).

Part III

Experimental Research

Chapter 5

Efficient Allocation Using IaaS Sourcing

5.1 Introduction

This chapter tackles research question R2 which aims at identifying the relevant determinants in an economic optimization model of hybrid IaaS sourcing. The motivation for focusing on the economic aspects stems from the results of the empirical survey and are detailed in the following section 5.2 along with the fundamental modeling assumptions which are derived from research literature. The relevant determinants are identified by mathematically modeling the outsourcing decision process. The goal of this chapter is to arrive at a mathematical model that satisfies the assumptions, that models the outsourcing process and that can be experimentally evaluated.

Research questions R2 and R3 use an experimental research approach, i.e. numerical experiments are conducted and the results are analyzed to arrive at answers to the research questions. Generally, in an experiment, one or more process variables (or factors) are deliberately and systematically changed in order to observe an effect on one or more response variables (NIST/SEMATECH 2014). These experiments are planned according to the design of experiments (DoE) methodology (NIST/SEMATECH 2014). The process model underlying this methodology is displayed in Figure 5.1. It assumes the decision process to be of a black box type; this process receives a number of controlled inputs (factors), which can be deliberately set to certain levels by the experimenter, and a number of uncontrolled inputs (co-factors), which are beyond the experimenter's control and may fluctuate independently from the factors. Experiments consist of setting the factors to certain levels and let the decision process react on these inputs and produce an output (a response). This combination of factors, co-factors and responses can then be analyzed, and an empirical approximation function can be derived. This empirical function then links the inputs and the output in a mathematical fashion and thus reveals the determining factors.

Choosing the DoE methodology is motivated by the effectiveness of this methodology for the given research questions. According to (NIST/SEMATECH 2014), experimental designs are suitable for different purposes, one of which is selecting the key factors affecting a response among a multitude of given inputs (screening experiments). These are efficient at determining the important factors with a minimal number of experiments; therefore, this approach matches the goal of research question R3. This experimental approach can be classified as an exploratory type of research, as its aim is to empirically discover the determining factors and not to fit a predefined function. The experimental unit is the single application; the experiments are designed to test the effects of the controlled inputs on the applications.

The DoE methodology and this Black box process model serve as the structure which frames the research activities in chapter 5 and 6. In chapter 5, the concepts of this experimental approach are designed. The underlying decision process is the outsourcing process, which was already introduced in the last chapter (see Table 2.6 in section 2.3.4); the output of the outsourcing process is a placement decision for each

software application. The generation of the output (response) is modeled in section 5.6, which builds on the tariff and resource models, and devises a cost-based linear optimization model. It calculates a cost-efficient allocation of software applications to two different IT resource providers; thus, the output consists of a placement decision for each software application as it would have been made by a solely cost-driven business executive. The goal of this experimental setup is to derive an empirical approximation function for this decision process; the function is the result of a machine learning approach which is developed in section 5.7, along with performance criteria that allow a quality assessment of the approximation. The controlled inputs (factors) are presented in section 5.2; it introduces fundamental parameters and assumptions regarding an IaaS outsourcing decision; these definitions will be used throughout the following sections. It also motivates the design choices by applying the results from the empirical part. The uncontrolled inputs (co-factors) are divided into the following three groups: software application workload, IaaS provider tariffs and quality requirements. The software application workload (e.g. CPU or RAM requirements) are modeled in section 5.5 in a generic IT resource model. The provider tariff models are elaborated in section 5.4.1 and 5.4.2 for both in-house and IaaS offerings. The quality requirements are modeled in section 5.3 where a generic quality model for IaaS resources is suggested.

A complementary, multi-method approach is chosen in this part of the thesis: the case study follows the principles set forward by Eisenhardt (1989b), Eisenhardt and Graebner (2007) and can be classified as qualitative, explorative and inductive research. The experimental methodology is characterized as quantitative, explanatory and inductive. The results of both methods yield a unified set of hypotheses which describe the conditions under which IaaS usage is beneficial for an enterprise, thereby providing answers for research questions R2 and R3. The philosophical perspective of both strands of research is positivist; “positivist studies are premised on the existence of a-priori fixed relationships within phenomena which are typically investigated with structured instrumentation” (Orlikowski and Baroudi 1991). “Positivists generally assume that reality is objectively given and can be described by measurable properties which are independent of the observer (researcher) and his or her instruments” (Myers 1997).

As an outlook, the experimental setup detailed above is evaluated in chapter 6, using the DoE methodology on real-world business software applications. In order to estimate the possible effect of the experimentally uncontrolled quality requirements, the generic quality model is also evaluated in chapter 6 using a real-world case study. Both evaluations take place in the same company in which the functional determinants of IaaS usage were investigated (see section 3.2), i.e. they share a common organizational environment.

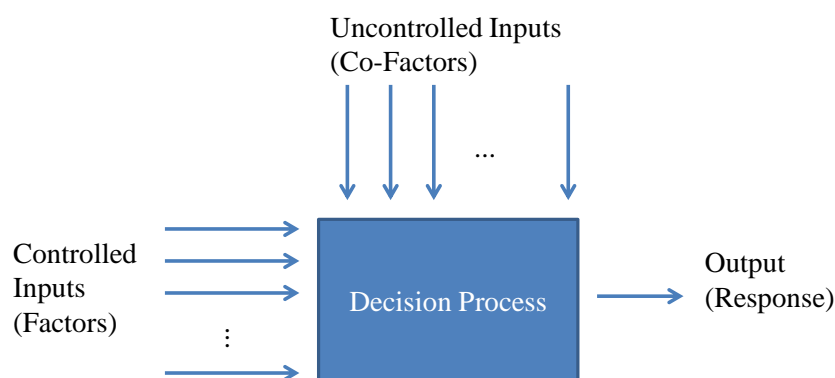


Figure 5.1: Black box Process model (based on (NIST/SEMATECH 2014))

5.2 Scenario Assumptions

In order to arrive at a mathematical model for the IaaS-based outsourcing process, the informal descriptions of the outsourcing process given in section 2.3.4 need to be mathematically defined. The term “outsourcing scenario” is the main concept in such a definition. An outsourcing scenario encompasses a number of relevant contextual aspects, that uniquely define and characterize an outsourcing decision situation for a business executive. The mathematical notation for such an outsourcing scenario and its properties are established in the following paragraphs. The complete set of possible outsourcing scenarios is defined as the relation

$S = (I, W, \eta, \alpha, \beta, a_\Delta, p_\Delta, t_\Delta)$ with the following input sets. S is supposed to contain the Cartesian product of all the input sets.

I : I is the index set of software applications that are in principle suitable for IaaS deployment. $i \in I = \{1 \dots N_s\}$ is used as an index to refer to a specific software application i . $|I| = N_s$ is the total number of suitable software applications at a potential IaaS client company. The “suitability” of a software application is more thoroughly defined in the following compilation of scenario assumptions about IT resource quality and software applications. Suitable quality dimensions and a discussion of the suitability of current IaaS offerings can be found in section 5.3.

W : a collection of resource demand observations \vec{w}_i of each software application i . \vec{w}_i is considered a multivariate random variable, where the different IT infrastructure resources act as dimensions (IT infrastructure resources could be CPU cores or storage for example). A formal definition is given in equation 5.9 below. The statistical distribution of the elements in \vec{w}_i is generally unknown; the sample size for each \vec{w}_i is supposed to be identical and of size n . If the samples are gathered from real-world applications, this assumption might not hold; however, missing values do not pose a problem for the optimization model presented here, as it relies only on the empirical distribution and not on single values.

α : it is generally assumed that the software application workload can be divided into a base workload and a peak workload; in the case of IaaS resource usage, the base workload is placed on IT resources whose tariffs favor a long and steady resource usage, whereas the peak workload can be covered dynamically using on-demand tariffs. This parameter describes the level of continuous base resource utilization. It regulates the available base resources. To this end, the α -quantile of \vec{w}_i is calculated (e.g. for $\alpha = 0.5$, the base load would be the median of the IT infrastructure resource demands in \vec{w}_i).

β : this value signifies the maximum IT resource usage of a software application and is modeled as a high-percentage quantile of the IT resource requirements to capture the tail values of the IT resource distribution (peak resource usage). The β -quantile is calculated on the distribution of \vec{w}_i .

η : The outsourcing degree $\eta \in [0 \dots 1]$ denotes the share of software applications that are actually deployed in an IaaS Cloud in this scenario. This value represents a target figure and can be thought of as a strategic goal of a CIO in terms of application outsourcing.

p_Δ : this value is a sensitivity parameter that models the internal price structure differences of IT resources in potential IaaS clients. It is assumed that the price level of IT resources varies systematically among the enterprises considered in the outsourcing scenario due to enterprise-dependent factors like IT resource utilization, IT management efficiency, economies of scope or organizational purchasing power. To reflect the differences, p_Δ for enterprises may vary considerably (better-run or larger enterprises will most likely have lower infrastructure price levels).

a_{Δ} : this value is a sensitivity parameter that models the differences in price structures among the providers of IaaS resources. It is assumed that the price level of IT resources varies among the IaaS providers considered in the outsourcing scenario; it might also vary for a single provider over time. To reflect the differences between the IaaS tariffs, a_{Δ} for IaaS providers may vary considerably.

t_{Δ} : A single time interval t_{Δ} is under analysis; iterated purchasing decisions are excluded from the analysis. The time period is assumed to be discrete (i.e. no continuous time line) and without pauses or interruptions. A long-term allocation (months, years) of IT resources is presupposed.

Whereas the last chapter dealt with the principle, high-level determinants influencing an IaaS outsourcing decision, this chapter looks in more detail at the quantitative key factors in such an outsourcing decision, i.e. quality metrics, business software workload characteristics and the cost associated with an outsourcing decision. These factors become especially relevant once the principle readiness for the usage of IaaS resources has been established in the company. The empirical research conducted in the previous part of the thesis yielded insightful results about the determinants of IaaS usage in enterprises. These results are now harnessed in the experimental part of the thesis; they motivate a number of outsourcing scenario assumptions for the mathematical model of the outsourcing decision process. Figure 5.2 integrates the theoretical constructs, their empirical results and the derived scenario assumptions. These assumptions need to be supplemented by some more technical aspects; all assumptions are listed in more detail below.

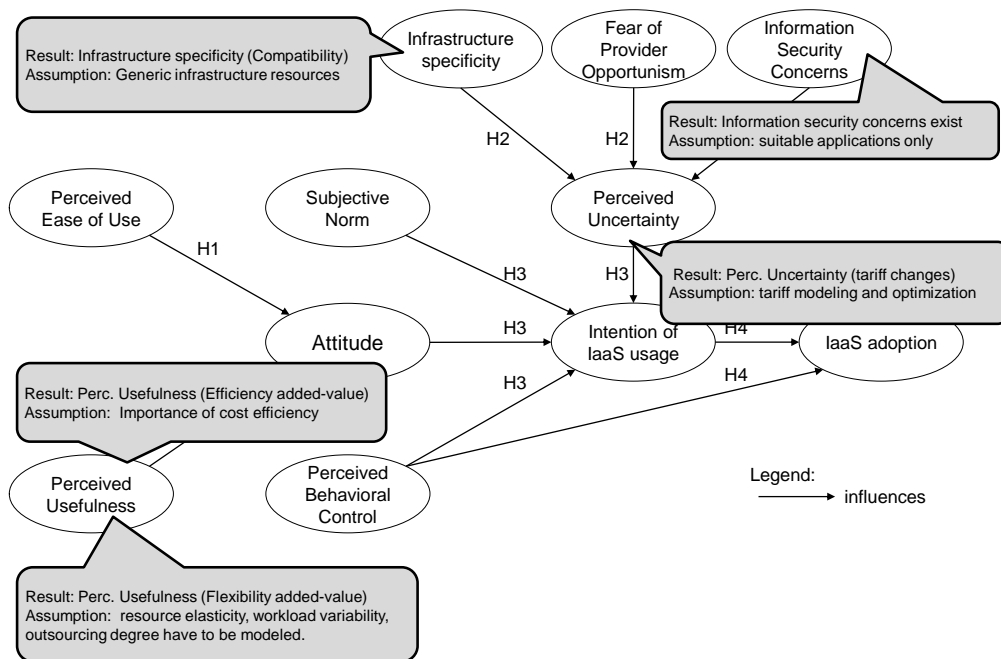


Figure 5.2: Scenario assumptions derived from the empirical results

The following list explains the outsourcing-related assumptions:

- As literature research and the survey results in section 4.4 suggest, IaaS' perceived usefulness is one of the major adoption drivers and efficiency considerations are the most important part of the usefulness concept (REL4 regression coefficient in Table 4.20). Hence, the following analysis will focus on the cost-benefit evaluation of IaaS offerings in the enterprise IT context.
- The selection of the business services considered for outsourcing has already taken place. The selection process is assumed to follow the general outsourcing process (see Table 2.6, step "Activity

Analysis”). This premise also caters to the information security concerns identified in the empirical research.

- The selected business services are actually amenable to outsourcing in terms of non-monetary criteria like security, availability, organizational matters, etc. Such non-functional properties either have to be modeled economically or they have to be checked during earlier phases of the outsourcing process.
- The selection of business services/applications and their sourcing model is independent from the available IaaS providers. In principal, each business service can be delivered by any available provider. Once a business service has been selected for outsourcing, it can be outsourced to any provider.
- Vendor selection is a structured, separate task within the outsourcing process in its own right with its own research background.
- IaaS providers are primarily distinguished by their pricing schedules. When different providers are compared based on empirical results, the analysis is based on comparable offers (e.g. pricing schedules, additional services, etc.) Care is taken, that only similar offers are compared. It will be discussed in section 6.4 how far the differing qualitative criteria of Amazon’s and BMW’s tariffs can be reproduced using a TCO-based approach.
- Sufficient market transparency: preferred providers and their offers /tariffs are universally known by all market participants.

Two technical assumptions are required for simpler modeling:

- The statistical distribution of the random variables representing the workload on servers running business applications is available for the time period t_{Δ} under analysis. A possible source for this workload distribution might be historical server monitoring data.
- Business application workload is created by the execution of specific business transactions on the system under analysis. Those can be data entry tasks, Web page views or any other functionality that the system offers. It is assumed that a specific time series of transactions uses the same amount of underlying hardware resources, no matter what Cloud provider hosts the virtual machine in which this service is run. This assumption equates to the necessary condition that the resource offerings of Cloud providers need to be comparable in performance among each other.

The quality aspects are not explicitly integrated into the mathematical model of the outsourcing decision process; however, two premises need to be established to increase the external validity of the results.

- It is assumed that the selection of both suitable applications and suitable IaaS providers has already taken place. As a part of this selection process, a number of service quality requirements will have to be matched between applications and providers (e.g. using a quality model as described in Sec 5.3). This premise relates to the information security concerns identified in the empirical research.
- The infrastructure resource requirements are generic, i.e. no special-purpose hardware is demanded. This also relates to the empirically supported compatibility demands of enterprise users.

A simplified notion of business software applications serves as the basis for the mathematical model. The following statements summarize this notion:

- Software applications follow a monolithic application model. Each application is self-contained and runs in exactly one virtual machine. There are no interdependencies among the applications.

- Stateful services: Applications are assumed to be complex software services running continuously. They exhibit significant changes in resource requirements over time. Applications run in an online environment, not a batch environment. Job-based workloads are not in the focus of this research; the application workload at hand is not easily parallelizable (e.g. unlike HTTP requests to a Web server, which are easily parallelizable), as complex, stateful transactions are involved.
- The software applications cannot be executed in parallel (embarrassingly parallel). Thus, horizontal scaling (scale-out) is ruled out, vertical scaling (scale-up) is preferable for complexity reasons. (Horizontal scaling would require a special application architecture, whereas vertical scaling is usually supported out-of-the-box by many currently popular software platforms like Java or SAP). The model assumes perfect vertical scalability of the business applications involved, i.e. if a business service consumes 30% of CPU time on a two-core system, it would roughly consume 60% of CPU time on a single-core system. Vertical scalability also entails that a software application can autonomously and automatically take advantage of additional computing resources if these are provided by the operating system.
- Software licences are supposed to be mobile (e.g. SmartLM¹) and are migrated along with the application, when the application is moved from the corporate data center to a new Cloud environment.

The following provider assumptions are required for the mathematical treatment:

- Tariffs: Two basic cost charging schemes are assumed: virtual machine instance-based and workload-based. CPUs and RAM are not charged separately, but as part of the virtual machine instance usage over time. All other resources (e.g. network traffic, storage) are charged based on the IT resource amount consumed over time. IaaS provider tariffs and the granularity of IT resource offerings (e.g. types of VMs available) are assumed to be externally given and fixed; their variation is not in the focus of this research. Multiple tariff types per VM instance are assumed to exist (e.g. an on-demand tariff and a non-linear tariff). Tariffs differ for different VM instance types; the tariff model is further elaborated in section 5.4.2. The empirical research supports the need for an accurate tariff model, as perceived uncertainty is directly influenced by provider tariff changes.
- Elasticity: There are IaaS instances that can be scaled to a sufficient size for every application. The resource allotment of any virtual machine can be adapted or the virtual machine can be migrated among servers. These changes happen with very little time delay (e.g. VMware Elastic Memory for Java²). Elasticity is a defining property of IaaS providers; client enterprises are assumed not to have this capability. The importance of this elasticity premise is supported by the empirical research as perceived usefulness is mainly determined by IaaS resource flexibility.
- A unique migration path exists for all VM instances of one provider. If a software application needs to be migrated from one type of VM instance to a smaller or larger one, a unique migration path ensures that there is exactly one unique smaller or larger type of VM instance. This implies mathematically that a total order can be established on the types of the various available VM instances, such that a unique, size-dependent migration path exists. As a VM instance is characterized by both its compute core count and its RAM size, the comparison of two VM instances might be ambiguous. A definite comparison is only possible, if the size of the larger VM instance dominates the size of the smaller VM instance (i.e. the comparisons for both core count and RAM point in the same direction). The necessity for this assumptions is explained in section 5.5.1.

¹<http://www.smartlm.eu/>, last accessed 2013-12-29

²<http://www.vmware.com/support/pubs/vfabric-em4j.html>, last accessed 2013-12-29

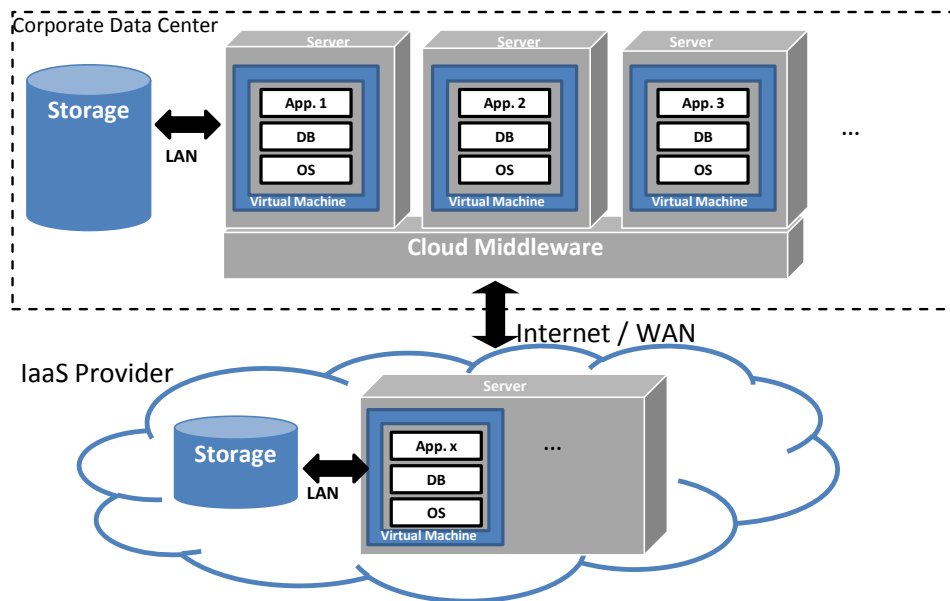


Figure 5.3: Technical architecture of an IaaS Cloud usage scenario

In general, the technical architecture of an IaaS Cloud usage scenario is assumed to look comparable to the architecture shown in Figure 5.3. The corporate data center is composed of a dedicated mass storage solution and of physical servers each of which runs a virtual machine. Each virtual machine acts as a host for the complete software stack (operating system, database and application software) of a single business solution. The software applications access the mass storage using the local area network (LAN). Each IaaS provider offers a similar setup in his data center, such that the virtual machines can be migrated from the corporate data center to the Cloud effortlessly. The migration process is managed by a Cloud middleware layer, which automates the transfer from and to the Cloud and monitors the status of the virtual machines. The main difference between the corporate data center and the Cloud data center lies in the fact, that the IaaS client cannot choose the physical server for each of his virtual machines; the IaaS provider is free to place the virtual machines as he sees fit.

5.3 Quality Dimensions

In section 3.4.1, the e-Service quality model of Modheji (2010) was already briefly introduced. It had been specifically developed for the survey and thus only covers the most important aspects in a high-level fashion. It is insufficient to represent the more fine-grained quality concepts needed in the quality analysis of IaaS services. Thus, the model is now expanded to enable a more elaborated and detailed view of IaaS service quality. Especially the aspects relevant to SLAs were refined. Please see Appendix G for the mapping table. For this work, the following quality definition DIN EN ISO 9000:2005 (Quality Management, Statistics and Certification Standards Committee 2005) will be used; according to this standard, quality is the “degree to which a set of inherent characteristics fulfills requirements”. The standard makes it clear that it only covers permanent characteristics. A requirement is a “need or expectation that is stated, generally implied or obligatory” (Quality Management, Statistics and Certification Standards Committee 2005).

The quality assessment of each product or service starts with the definition and separation of quality criteria, each of which is associated with one or more requirements that need to be fulfilled. The degree of fulfillment of these requirements eventually determines the quality (Berger 2005, p. 41). The quality

assessment of current IaaS market offers necessitates the application of the definition of the quality concept. A service can usually be judged by numerous quality criteria; for a comprehensive quality model, these different quality criteria need to be aggregated and combined to a reasonable number of quality dimensions. Such an abstraction of these criteria is necessary to ensure the comparability of different service offers along the aforementioned quality dimensions (Modheji 2010). A multi-faceted quality view is definitely necessary; e.g. Bruhn (2008) argues against the isolated consideration of single criteria.

A wide range of e-Service-quality models exist (see (Barrutia and Gilsanz 2009) for an overview). A well-published example of an e-Service-quality model is “E-S-Qual” (Parasuraman, Zeithaml, and Malhotra 2005). Here, the customer’s impression of the service quality is argued to be the comparison of the actual with the expected service performance. The ultimate goal is customer satisfaction. However, these e-Service-quality models are focused on applications in business-to-consumer eCommerce scenarios, especially for Web-based eCommerce offers, hence they are not directly applicable for IaaS quality assessments. The primary target metric of those models is customer satisfaction; this metric plays only an indirect role for IaaS services, as the service consumer is usually a software application running on the IaaS infrastructure. The primary concern in this work is rather the technical service quality defined by QoS targets (the question how the IaaS quality of service affects the satisfaction of the buyer is omitted in this work; it is assumed that the fulfillment of the IaaS QoS targets ensures a sufficient customer satisfaction).

This work rather follows (Berger 2005, p. 53) and (Beims 2012, p. 227), which maintain that service quality is the degree of accordance of the requirements for certain service properties defined between the customer and the service provider with the actual characteristics of these properties during the service rendition. Berger (2005, p. 111) maintains that for the area of IT service operations, there exists no common catalogue of relevant quality dimensions in the research literature up to the time of writing. The goal of the following paragraphs is to establish such a catalogue for IaaS offerings.

Ma, Pearson, and Tadisina (2005) explores quality dimensions for ASP providers, but is still centered around customer perceptions, not agreed-upon QoS targets or defined product features. There are however objectively quantifiable dimensions of the ASP service quality model that deserve to be considered. Garvin (1987) suggests quality dimensions for (physical) products; a subset of these quality dimensions is also relevant for IT services. However, “durability” as one of Garvin’s dimensions is omitted because of its inapplicability to digital services; “aesthetics” is also omitted, because digital services exhibit no sensory properties like regular products (visual, tactile or audible clues as to its quality). “Perceived quality” i.e. customer perceptions are also omitted because first they are difficult to measure objectively and second, because these perceptions are partly based on the reputation and the experiences that customers gained in the past. Those perceptions might not be substantial or wide-spread enough for a new technology like IaaS services.

Berger (2005, p. 160) proposes the degree of technological innovation as another quality dimension. The degree of technological innovation characterizes the up-to-dateness of an IT system (hardware and software) related to the current state-of-the-art. The literature does not define any practical metric to operationalize this notion, thus this dimension is omitted as well. However, the single components of an IT system (especially software, but also for hardware) tend to have version numbers to signify the release level of this component. This level could act as a base for comparison. The remaining factual quality dimensions of IT infrastructure service providers are listed in Table 5.1, which is based on Berger (2005, p. 112) and Ma et al. (2005).

System-related quality dimensions are quality dimensions measuring the properties of the underlying infrastructure system. Operations-related quality dimensions are quality dimensions measuring properties of the support processes surrounding the infrastructure system. These processes are usually labour-intensive services whose quality depends on the functional expertise and the personal skills of the personnel involved. This personnel is usually grouped in distinct operations units within an organization (or outside an organiza-

Table 5.1: Objective Quality dimensions of IT services

	Quality dimension	Source
System-related	Operational time	Berger (2005, p. 119)
	Performance	Berger (2005, p. 120), Garvin (1987)
	Availability	Berger (2005, p. 134), Ma et al. (2005)
	Stability	Berger (2005, p. 152), Garvin (1987)
	Recoverability	Berger (2005, p. 153), Garvin (1987)
	Security	Berger (2005, p. 157), Ma et al. (2005)
	Standardization	Berger (2005, p. 159), Ma et al. (2005), Garvin (1987)
	Product Features	Ma et al. (2005), Garvin (1987)
Operations-related	Service time	Berger (2005, p. 161)
	Performance	Berger (2005, p. 165)
	Reliability (on-time delivery)	Berger (2005, p. 167), Ma et al. (2005)
	Empathy	Ma et al. (2005), Parasuraman et al. (1988)
	Assurance	Ma et al. (2005), Parasuraman et al. (2005)

tion in case these services have been outsourced). The operations processes are assumed to be coordinated using a ticket-based workflow system, which assigns a ticket number to each service request.

In this work, an IT-system is seen as an abstract entity which is defined by technical properties. An IT-system is thus comprised of multiple technical components that exhibit a certain functionality (e.g. storage space or compute instance). Those technical properties are independent of the user's view of the system, but more related to its technical architecture. The technical components are also called the IT infrastructure of an IT system.

The following list defines the concepts introduced in Table 5.1. The level of description remains abstract; the inclusion of specific KPIs for measuring these quality criteria is omitted as those KPIs and their measuring process are highly context- and provider-specific.

Operational Time The operational time of an IT system is the time interval in which the system is technically available as planned and can be used (Berger 2005, p. 119).

System Performance The performance of an IT system determines the time that an IT system takes to execute specific actions under defined constraints (Berger 2005, p. 120). An example for a performance metric would be the response time of the system for certain computing operations or the input-output (I/O) operations per second of a storage subsystem.

Availability The availability of an IT system is defined as the error-free usage of the system functionality under pre-defined conditions within a specific time frame (Berger 2005, p. 134), (DIN 1990). However, the availability of an IT system does not allow any conclusions as to its quality of service.

Stability The stability of an IT system describes the proneness to failures (frequency of break-downs) within a specific time frame. (Berger 2005, p. 152) A typical stability metric would be MTBF (Mean Time Between Failures).

Recoverability The recoverability of an IT system assesses both the chance of principally putting a broken-down system back to work and the speed with which this restore action can take place (Berger 2005,

p. 153) A typical recoverability metric would be MTTR (Mean Time To Repair). A metric typically found in technical SLAs is the maximal permissible resolution time for a certain incident (e.g. 1d).

Security Security characterizes the state of an IT system in which external thread-related risks, that exist during IT system deployment, are limited by appropriate security measures to a manageable size (Berger 2005, p. 157), (BSI 1992).

Degree of Standardization The degree to which single IT systems in an IT infrastructure pool resemble each other. (Berger 2005, p. 159) This definition is rather abstract and Berger (2005) does not specify any actual metrics for this quality dimension. A more practical definition would have to include the different configuration options that exist for any IT system. Regarding IT infrastructure, it could be distinguished between the type of operating system or the type of processor architecture for example.

Product Features Features are those characteristics that supplement the basic functioning of products and services (Garvin 1987). They can be thought of as a secondary aspect of performance and give the customer the perception of choice among different options. Features are an especially relevant quality dimension of software products, which are defined by their functionality features.

Service time The service time is the time interval in which the service unit can render a certain service or in which a certain service can be rendered by this unit upon request (Berger 2005, p. 161). This quality dimension delineates the time interval in which the service unit is available and executes regular activities (like server monitoring) or occasional or unexpected activities (like a server restart after a break-down).

Operations performance The operations performance assesses the time, which a service unit requires for the execution of specific actions under defined constraints (Berger 2005, p. 165). In case of a problem ticket, the performance is both related to the responsiveness of the service unit and the time until a solution is reached. The definition of responsiveness is understood here that the client service request is addressed within a specified time interval. The client is then notified that the service request is being processed. In the case of a problem ticket, this notification also means the start of the problem solution process and can be followed by a stage of inquiry where the client provides further detail as to the specifics of his request. There are several metrics that measure the time for each execution of a specific action. As an example, the solution time and the recoverability of a system fall in the category of metrics for problem-related actions. The provisioning time of an IT system is a metric for the performance of a provisioning action (e.g. when setting up new hard/software components) (Berger 2005, p. 166).

Reliability The reliability of a service unit is understood as the adherence to agreed-upon deadlines or actions under defined constraints (Berger 2005, p. 167). Suitable metrics could be the degree of actions delivered on-time in relation to all actions or the error rate of service operations.

Empathy “Caring, individualized attention the firm provides its customers” (Parasuraman, Zeithaml, and Berry 1988). This broad definition encompasses also the communication aspect of service delivery. Berger (2005, p. 169) sees the communication behaviour of a service as one of its quality features; this behaviour is characterized by the communication frequency and fashion of this unit with its environment, i.e. the IT users, for whom the service unit is responsible. Exemplary metrics include the message count and the message frequency of support messages related to some support issue within a certain time period. One precondition of successful communication is the accessibility of the service provider, hence Berger (2005, p. 162) includes this dimension in his quality model. Accessibility denotes the service unit’s property that a successful contact can be established with this unit within

a certain time interval using a defined means of communication. In this work, both communication behaviour and accessibility are seen as facets of the broader concept of empathy. Parasuraman, Zeithaml, and Malhotra (2005) also include this communication aspect in their e-Service-Quality model as a “contact” dimension (especially for recovery operations in the case of service failures). The empathy quality dimension is also valid for an IaaS service quality model, as the customer is directly integrated in the service process; in the German research literature, the customer is referred to as the “external factor” (Bruhn 2008, p. 22). This direct contact is needed for instance, when the customers order IT infrastructure or when they have to interact with the provider to solve operational issues.

Assurance According to (Parasuraman, Zeithaml, and Malhotra 2005), it is the confidence the customer feels in dealing with a Web site and is due to the reputation of the site and the products or services it sells, as well as clear and truthful information presented. For (Ma, Pearson, and Tadisina 2005), this concept is operationalized differently and is related to the trustworthiness of the ASP provider; this perception of trust is fostered for example by the availability of customer support, quality assurance systems and tools, and the availability of a secured physical environment for the ASP’s data center. This work does not intend to measure constructs like trust or confidence of customers; rather, the presence or availability of trust-inducing processes or properties of an IaaS provider indicates the fulfillment of this quality dimension. According to (Ma, Pearson, and Tadisina 2005), one of those factors is the technical expertise of the support staff: it is understood as the property of a service unit to be able to fulfill certain tasks under given constraints because of the technical know-how found with the persons in this unit (Berger 2005, p. 168). Exemplary metrics for this quality dimension are the first call resolution rate and the overall resolution rate (in the case of call center operations e.g. for user help desks).

The goal of this section is to derive a comprehensive, abstract list of IaaS service quality dimensions from the existing body of research in order to compare the quality aspects of the offerings of multiple providers in different sourcing variants. As the quality dimensions outlined above are rather abstract, the categorization based on these categories will be rather coarse-grained as well, but it should give a basic conceptual framework for comparing IT infrastructure service offerings. The entries in the list presented above are not weighted; if these quality dimensions were to be applied in a real outsourcing situation, a situation-specific weighting scheme would have to be adopted (weights would likely be depending on the non-functional requirements of the software applications that are supposed to run on the target IT infrastructure). Weighting of multiple dimensions can be achieved with different weighting schemes (e.g. see (Hwang and Yoon 1981)), each of which has its own characteristics and has to fit the specific situation.

5.4 IaaS Tariff Modeling

5.4.1 TCO Approaches

This work focuses on a cost-oriented decision model only, as the nonfinancial benefits resulting from the usage of IaaS resources are generally hard to quantify. One of the most important cost-oriented models both used in research and in real-life settings is the Total-Cost-of-Ownership (TCO) model (Silver 2007). One of the latest varieties of TCO models applies the approach to the field of IaaS (Leong 2009; Li, Li, Liu, Qiu, and Wang 2009), however the suggested model does not capture the effects of nonlinear provider tariffs.

Another drawback in current literature is the treatment of Clouds as a black box. As shown in section 2.3, the provider side of IaaS Cloud Computing cannot be generalized in this fashion, as the quality and pricing variations in IaaS offerings are considerable. As a conclusion, TCO models for Cloud Computing exist, however, they are not powerful enough to capture the complexities of the current IaaS offerings.

Table 5.2: Cost model for IaaS Cloud resources

Recurrence	Cost type	Cost factor	Explanation
One-time expenses	Set-up cost	Fee per VM instance	These costs can vary among the providers It includes the effort for making the software application Cloud-ready and the set-up fee from the provider.
Recurring expenses	Cloud Infra-structure	VM CPU hour	Cost for running a virtual machine in a time contingent under the hourly tariff. It is assumed that the IaaS Cloud provider offers the OS licenses along with the server and that support and management service charges are already included in the virtual machine charges.
		Contingent fee	Cost of buying a contingent which represent a minimum charge.
		Cloud storage	Cost for using a unit of storage during a time interval.
	Network	Internet bandwidth	The variable cost of Internet bandwidth is considered in this model, as the Internet connection will most likely be used by other IT systems in the enterprise.
	Data transfer	outbound	Cost of serving data from the Cloud.
		inbound	Cost of uploading data to the Cloud.

Table 5.2 describes the cost model used throughout this thesis; it is based on the work of (Leong 2009) and (Barroso and Hölzle 2009). The cost model for enterprise resources does not feature the cost of enterprise-internal LAN infrastructure; LAN hardware costs are neglected as LAN transports typically cost 10000 times less than Internet transports (Gray 2003). Moreover, they are not helpful in distinguishing the cost between enterprise and Cloud resources, as all data has to pass through the LAN, even if the software application is placed on an IaaS Cloud. Internet transports are not required for the evaluated business applications in this model, if the application runs within the enterprise data center.

5.4.2 Tariff Model

Now that the cost types are defined, the question remains how these cost types are charged. A general type of charging scheme is the non-linear tariff; non-linear tariff schedules are commonly used in the telecommunication sector or in logistics, but they are also making inroads into Cloud Computing. Generally speaking, any price schedule in which the total cost is not directly proportional to the amount bought, is called nonlinear (see Wilson (1999) for a definition). Nonlinear schedules can be designed using several concepts; four of the most common are N-part tariffs, incremental quantity discounts, all-units quantity discounts. When only considering the total cost incurred by the customer, all three types of discounts can be modeled using an N-part cost function (Benton 1991).

Nonlinear pricing has mostly been researched in the context of revenue management for the IaaS provider; numerous authors analyze the optimal pricing schedules for utility computing service providers (e.g. Paleologo (2004), Huang and Sundararajan (2005)). Based on the literature review, the value propo-

sition of the single IaaS user paying nonlinear tariffs has not been analyzed in the literature so far; the aforementioned research remains on a (macro-)economic level. Consequentially, an IaaS tariff model for the single IaaS user shall be developed in this section. The model will be explained with one specific provider and one specific resource type in mind without loss of generality. Different provider-resource type combinations can be modeled accordingly.

A business software application requires certain IT resources; those resource demands are matched with the products from the IaaS suppliers. A product represents a specific IT resource type like CPU cores or data storage. For one IaaS provider, each product can be associated with a number of different tariffs (a client can have the choice between multiple tariffs for the same resource type, although he ultimately chooses only one).

Figure 5.4 depicts graphs for the total cost (y-axis) depending on the amount purchased (x-axis) and shows the defining parameters of a tariff and different tariff options T_v, T_1, T_2, T_3 (for example, the IaaS provider GoGrid³ uses such a pricing scheme for its computing resources). C_s represents the fixed setup costs that may be charged for new clients. C_{vol_1} is the cost of the volume discount in the 3-part tariff T_1 (the index number “1” signifies that C_{vol_1} belongs to tariff T_1); the client receives Q_{T_1} units of the product in exchange (the resulting effective variable cost per unit is $\frac{C_{vol_1}}{Q_{T_1}}$). a_1 is the variable cost per unit after the discount contingent has been fully utilized in tariff T_1 (the same notation is used parallel For example, the total cost C for a given consumption x of a specific IT resource in a given period is calculated using equation 5.1. Amazon AWS⁴ uses a similar discount model for their reserved instances pricing (the adjective “reserved” indicates, that the VM instances are dedicated to one client over an extended period of time and have lengthy minimum subscription times of over a year).

The main observation is, that those multiple independent tariffs for one resource type can be combined into one combined total cost function consisting of N parts. An exemplary result of such a combination is shown in Figure 5.5; it parallels Figure 5.4, which is proven by the faded lines of the original tariffs.

$$C(x) = \min_{y \in \{1 \dots N; v\}} (C^{T_y}(x)). \quad (5.1)$$

$$C^{T_v}(x) = a_v x. \quad (5.2)$$

$$C^{T_n}(x) = \begin{cases} C_s + C_{vol_n} & \text{if } 0 \leq x \leq Q_{T_n}. \\ C_s + C_{vol_n} + a_n(x - Q_{T_n}) & \text{if } x > Q_{T_n}. \end{cases} \quad (5.3)$$

The following paragraphs describe, how the resulting total cost function can be calculated, i.e. it answers how the individual tariffs can be combined to yield the correct total cost and it gives a computational complexity estimation.

Such an aggregated total cost function has the nice property of easily being usable in a linear optimization model. The original formulation of equation 5.3 (corresponding to Figure 5.4) is not directly suitable for a linear optimization approach; it would need to be part of the objective function, but it cannot be guaranteed to be linear, smooth and either convex or concave. These missing properties preclude the utilization of efficient, exact optimization methods.

A closer look at the mathematical properties of the tariffs $T_1 \dots T_n$ reveals that these functions are piecewise-linear. If the lower envelope of the tariff functions as defined by equation 5.1 could be computed, this lower envelope would also be a piecewise-linear function and hence, it could be modeled as a convex combination of its linear components. This convex combination could be inserted as a part of an objective

³<http://www.gogrid.com>, last accessed 2013-12-29

⁴<http://aws.amazon.com>, last accessed 2013-12-29

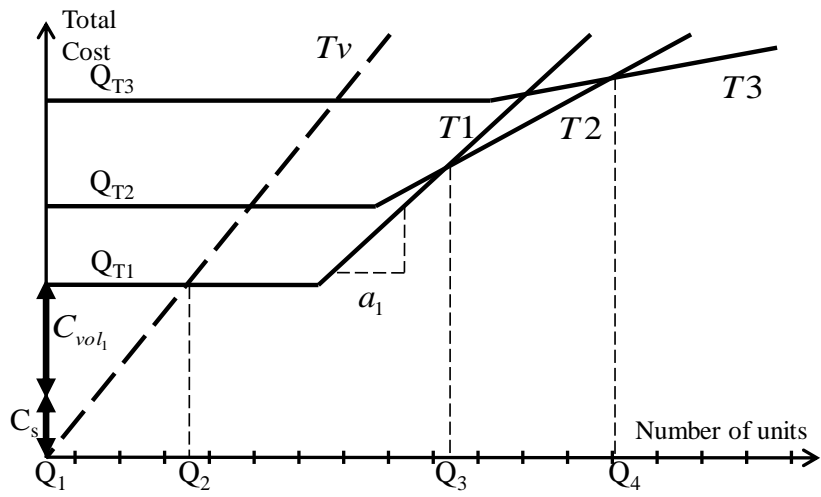


Figure 5.4: Parametric cost model

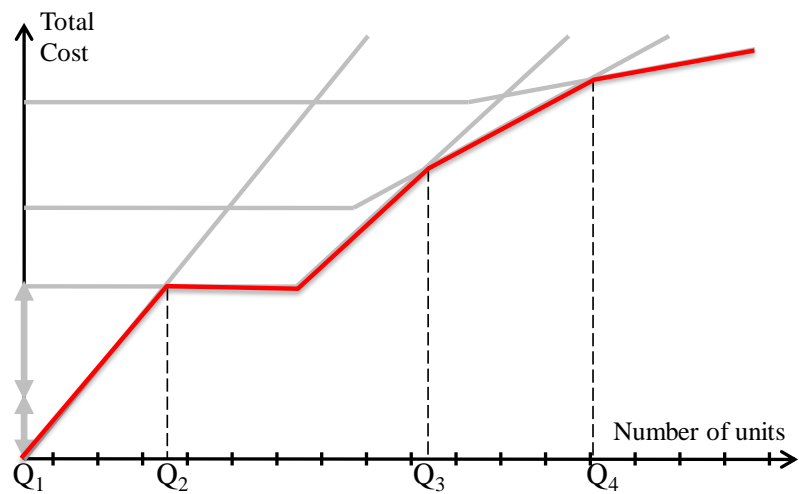


Figure 5.5: Resulting total cost function

function in a linear model and hence, the formulation would be amenable to exact optimization methods like IP (Integer Programming), for which efficient solution methods exist. (An example for the lower envelope of a set of piecewise-linear tariff functions is the function drawn in bold in Figure 5.5. In general, the lower envelope is defined as the minimum over the functions defined on the same domain.)

The problem to compute the lower envelope of equation 5.1 can be solved using the so called sweep-line paradigm (see Klein (1997, p. 51)), a result from computational geometry. Originally, Klein showed how this paradigm can be used to find the lower envelope of a set of linear function defined on an interval (Klein 1997, p. 79). The piecewise-linear functions discussed in this thesis are defined on \mathfrak{R}_0^+ in theory, but as the number of IT infrastructure resource units purchased will inevitably be finite for any given finite time interval, a finite upper bound x_{max} on x in equation 5.1 can be assumed. So the algorithms presented in (Klein 1997, p. 71) can be directly applied to this problem. The only preparation step necessary is the extraction of line segments from equations 5.2 and 5.3, which is simple given their piecewise-linear nature and the assumed upper bound x_{max} . The runtime complexity of the basic sweep-line algorithm for finding the lower envelope of line segments was shown to be $\Theta(n \log n)$; the storage complexity is $\Theta(n)$ (with n as the number of line segments) (Klein 1997, p. 86). The alterations required for preparing the line segments from the cost functions do not change this result.

The sweep-line algorithm returns C^{lev} , the piecewise-linear cost function representing the lower envelope of the input tariffs; each piece j of this function will henceforth be called $C^{lev,j}$. Equation 5.1 can now be formulated according to equation 5.4. The index of pieces j in $C^{lev,j}$ and the index n of input tariff functions are mutually independent, i.e. $C^{lev,j}$ is not necessarily associated with C^{Tn} .

$$C^{lev}(x) = \begin{cases} C^{lev,1}(x) = a^1x + b^1 = a^1(x-0) + b^1 & \text{if } 0 \leq x \leq Q_1. \\ C^{lev,2}(x) = a^2x + b^2 = a^2(x-Q_1) + b^2 & \text{if } Q_1 < x \leq Q_2. \\ \dots \\ C^{lev,j}(x) = a^jx + b^j = a^j(x-Q_j) + b^j & \text{if } Q_j < x < \infty. \end{cases} \quad (5.4)$$

Equation 5.4 signifies that each part of the piecewise linear function is a linear function, as its gradient is constant per piece (a^j). Equation 5.5 expresses that the cost function is continuous. This property follows from the fact, that each single tariff function C^{Tn} is continuous and that the pieces that the sweep-line algorithm returns, are either from one of these tariff functions, or they result from an intersection of two tariff functions, in which the continuity of the lower envelope is also preserved.

$$\forall j \quad \forall x \geq 0 : \lim_{x \rightarrow x_0^-} C^{lev,j}(x) = \lim_{x \rightarrow x_0^+} C^{lev,j}(x) = C^{lev,j}(x_0). \quad (5.5)$$

As a result, it was shown that an N-part cost function is general enough to model the most important discounts and that it can accommodate the combination of several multi-part tariffs by calculating the lower envelope of the tariff functions involved. This combination is necessary for the applicability of efficient optimization approaches that need to use this total cost function.

For the following considerations, a simplified version of the N-part cost functions will be used for complexity reasons, and it is assumed that set-up cost and contingent fees (Table 5.2) are zero and thus can be neglected. However, the models described below can easily be extended to include N-part pricing schedules, if the decision scenario requires it (see section 7.2).

As an application of the above introduced N-part tariff model, this model will now be applied to the tariff system of Amazon AWS in the context of the IaaS outsourcing scenario. In the case of IaaS sourcing from Amazon AWS, the IaaS Cloud TCO model 5.2 consists of four major cost types: VM instances, WAN/LAN transfer cost, storage cost. At the time of writing, two tariffs T_V^{VM}, T_I^{VM} are being offered

when renting one of their VM instances: on-demand pricing (Tv) and reserved instance pricing ($T1$). Let the IaaS instance types be indexed by $e \in E = \{1, 2, \dots\}$ with E being the set of available IaaS instance types.

Consequently, each VM instance type e is associated with two tariffs $Tv^{VM,e}, T1^{VM,e}$. The cost function C for each of these tariffs follows the corresponding equation defined above (equation 5.2, resp. equation 5.3). In this case, the sweep-line algorithm would be required to calculate the lower envelope C_e^{lenv} of the two tariffs per VM instance e . For later calculations, the price vector \vec{p}^{lenv} is now defined in equation 5.6, in order to shorten the following notation. As this model assumes a one-shot outsourcing scenario with a fixed length, the time interval t_Δ is entered into the cost formulas which then give the constant total cost factor for this time interval.

$$\vec{p}^{lenv} = \begin{pmatrix} \min(C_1^{Tv^{VM}}(t_\Delta), C_1^{T1^{VM}}(t_\Delta)) \\ \vdots \\ \min(C_{|E|}^{Tv^{VM}}(t_\Delta), C_{|E|}^{T1^{VM}}(t_\Delta)) \end{pmatrix} = \begin{pmatrix} C_1^{lenv}(t_\Delta) \\ \vdots \\ C_{|E|}^{lenv}(t_\Delta) \end{pmatrix} = \begin{pmatrix} p_1^{lenv} \\ \vdots \\ p_{|E|}^{lenv} \end{pmatrix} \in (\mathfrak{R}_0^+)^{|E|}. \quad (5.6)$$

If only the on-demand VM tariffs are supposed to be used, the price vector can be simplified to \vec{p}^{ondem} . (The same logic applies for the exclusive use of reserved tariffs, yielding \vec{p}^{res}). Please note that each p_e^{ondem} is a constant.

$$\vec{p}^{ondem} = \begin{pmatrix} C_1^{Tv^{VM}}(t_\Delta) \\ \vdots \\ C_{|E|}^{Tv^{VM}}(t_\Delta) \end{pmatrix} = \begin{pmatrix} p_1^{ondem} \\ \vdots \\ p_{|E|}^{ondem} \end{pmatrix}. \quad (5.7)$$

The other IT IaaS resources (WAN/LAN transfer cost, storage cost) are billed using a variable tariff each ($Tv^{InboundLAN}, Tv^{OutboundLAN}, Tv^{Store}$). The cost function C for each of these tariffs follows the corresponding equation 5.3 defined above. The corresponding price vector is \vec{p}^{vol} (equation 5.8). These definitions are necessary, because compute power is charged on a VM instance basis, whereas WAN/LAN transfer cost and storage cost are charged on a volume basis (yet all cost types are charged in a pay-as-you-go manner per software application). The tariff model for in-house IT resources can be defined accordingly.

$$\vec{p}^{vol} = \begin{pmatrix} C^{Tv^{InboundLAN}}(t_\Delta) \\ C^{Tv^{OutboundLAN}}(t_\Delta) \\ C^{Tv^{Store}}(t_\Delta) \end{pmatrix} = \begin{pmatrix} p^{InboundLAN} \\ p^{OutboundLAN} \\ p^{Store} \end{pmatrix}. \quad (5.8)$$

This section demonstrated how both the reserved and the variable tariffs can be fitted into the N-part tariff model. This successful application of the N-part tariff model is a first step to validate its real-world relevance.

5.5 IT Resource Requirements of Business Applications

As the provider-side models for quality and resource tariffs were discussed in the last section, it now becomes necessary to look at the client-side resource requirements and also develop corresponding models. The resource requirements can be classified in several resource types; for the sake of this model, four fundamental IT resource types $k \in \{CPU, RAM, LAN_in, LAN_out, Storage\}$ are distinguished.

The following example outlines how the scenario structure defined in section 5.2 affects the concept of IT resources. In Figure 5.6, an idealized empirical probability density function of the IT resource requirements of two exemplary software applications is shown. Application 1 (App1) has a more variable

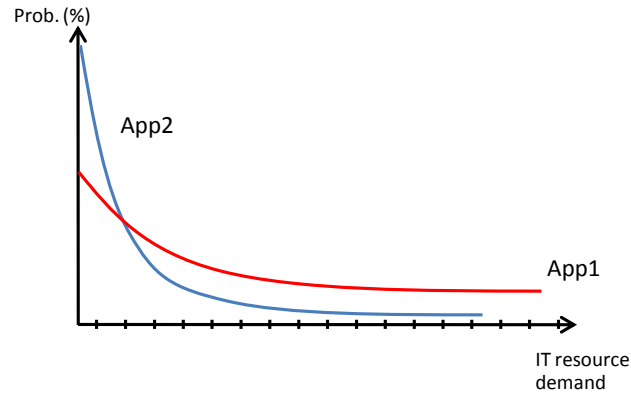


Figure 5.6: Exemplary IT resource demand distributions for two applications

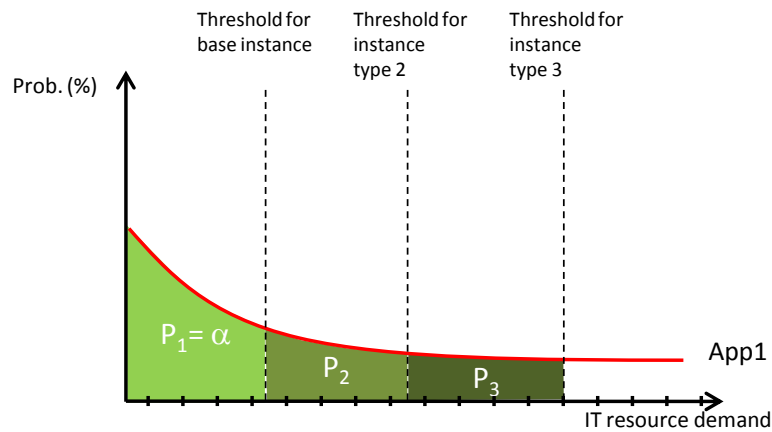


Figure 5.7: Example of IaaS deployment

IT resource demand as compared to Application 2 (App2). During the outsourcing process, it is decided that Application 1 is deployed in an IaaS Cloud (to leverage the resource elasticity of the IaaS resource), whereas Application 2 is deployed in the in-house data center.

Figure 5.7 shows the elastic resource usage for IaaS resources. Let the IT resource in question be the VM instance size for the sake of this example. With probability $P_1 = \alpha$, the resource requirements can be satisfied with a suitably sized base VM instance. For a certain probability P_2 , Application 1 will exhibit a greater resource demand, that is satisfied by vertically scaling the base instance to the next larger VM instance size. With probabilities P_3, P_4, \dots , even greater VM instances will be used. It is assumed, that the provider can accommodate the maximum application load with a suitable VM instance.

Figure 5.8 shows the fixed resource usage for in-house resources. Any application running on these IT resources has to reserve the maximum required amount of IT resources beforehand, no matter how infrequently these resources are fully utilized. $P = \beta$ signifies that the sizing of those IT resources is oriented towards the maximum case.

5.5.1 IT Resource Model

The IT resource model further formalizes the notion of \vec{w}_i . As already mentioned, the IT resource demand W , the software application workload, is modeled as a multivariate random variable. Equation 5.9 shows

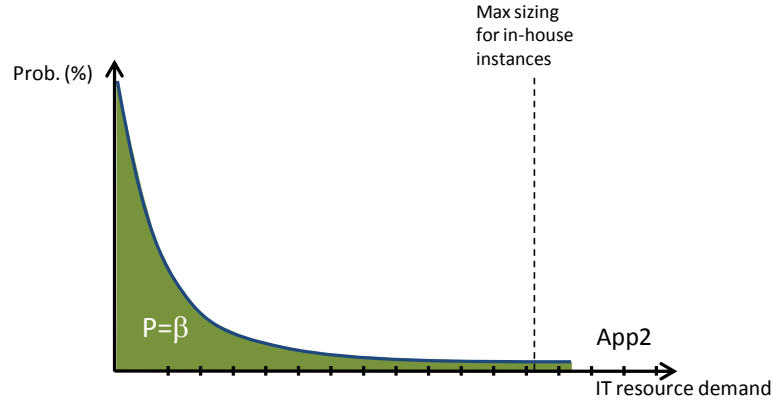


Figure 5.8: Example of in-house deployment

the random vector for application i . Please note that w_i^k is a proper random variable, defined by its empirical distribution.

$$\vec{w}_i = (w_i^k) = \begin{pmatrix} w_i^{CPU} \\ w_i^{RAM} \\ w_i^{InboundLAN} \\ w_i^{OutboundLAN} \\ w_i^{Storage} \end{pmatrix}. \quad (5.9)$$

Realizations of \vec{w}_i can be gained from historical workload traces of application i ; usually, those samples are only available at discrete points in time. It is assumed that n samples exist for each distribution w_i^k . Let $(w_n)_i^k$ be the n -th sample in this set.

It has to be pointed out that the above definition does not imply that the application workload is modeled as a time series or analyzed using time-series-related methods (using e.g. ARIMA, GARCH models). The time index is omitted as only the empirical distribution gained from the different observations is relevant for the model. This approach alleviates the necessity to have high-quality, continuous and complete time series sample data; missing values are less of an issue, as only the statistical distribution of the workload is required. Moreover, general time series models require significant knowledge about the statistical properties of the underlying random process creating the series. However, this process is unknown in the case of IT resource demands, so the usage of these models would necessitate the introduction of additional assumptions.

The α -quantile $Q_\alpha(w_i^k)$ for each component k of vector \vec{w}_i is defined as the value that satisfies the condition in equation 5.10. This definition directly corresponds to the standard statistical definition of a quantile (Rinne 2008, p. 33). F_n is the cumulative distribution function for w_i^k ; n denotes the number of observations in this empirical distribution function.

$$Q_\alpha(w_i^k) = F_n^{-1}(\alpha) := \inf\{x | F_n(x) \geq \alpha\}. \quad (5.10)$$

The definition in equation 5.10 can now serve as a basis to define the random vector $Q_\alpha(w_i) = (Q_\alpha(w_i^k))$. The resource demand distribution \vec{w}_i is the basis for the anticipated amount of IT resources that have to be bought at the IaaS provider or provisioned at the in-house data center. So far, \vec{w}_i denotes the actual IT resource demand distribution. When these resources are sourced from an IaaS provider, the question of

how much to purchase from this provider becomes relevant. The following paragraphs show how to arrive at these estimates in the context of the model.

The purchased amount of IT resources d_i of each application i differs depending on the deployment of the application. If the application is deployed in the IaaS Cloud then equation 5.11 shows the assumed purchased amount, where \bar{w} denotes the arithmetic mean of the corresponding random variable. The mean is chosen here as an estimator of the expected value of the random variable. If the application is deployed in-house, then equation 5.12 shows the assumed purchased amount. Q_β is the β quantile function. Hence, it is assumed that the worst-case storage demand has to be specified upfront when an application is supposed to be deployed in-house, as the in-house data center does not offer the same level of resource elasticity as the IaaS environment. The comparatively large size of Cloud Computing data centers and their advanced virtualization techniques make the rapid provisioning and deprovisioning of IT resources possible; in-house data centers usually do not share these advantages and have to manage their IT resources more rigidly.

$$d_i^{IaaS} = \begin{pmatrix} \bar{w}_i^{InboundLAN} \\ \bar{w}_i^{OutboundLAN} \\ \bar{w}_i^{Storage} \end{pmatrix}. \quad (5.11)$$

$$d_i^{In-house} = \begin{pmatrix} \bar{w}_i^{InboundLAN} \\ \bar{w}_i^{OutboundLAN} \\ Q_\beta(w_i^{Storage}) \end{pmatrix}. \quad (5.12)$$

Calculating the purchased amount for CPU and RAM resources work differently, as it is tied to the size of the virtual machine. Section 5.5.2 describes the VM instance selection and the calculation of the associated purchased amount of VM resources.

For the later analysis, a statistical property of the workload will be defined here, its variability. The measure used here for calculating the variability is the semi-variance of the workload. The semi-variance was suggested by (Markowitz 1959, p. 188) for the computation of efficient financial portfolios and is defined in equation 5.13 for the discrete random variable x , where \bar{x} is its mean, x_n the n -th observation and $\mathbb{E}()$ is the expected value function. (Markowitz considered x to be the returns from some sort of financial portfolio).

$$S_{\mathbb{E}} = \mathbb{E}(\min(x_n - \bar{x}, 0)^2). \quad (5.13)$$

For the sake of this research, the semi-variance is defined on the application workload and is therefore slightly adapted. Equation 5.14 shows the definition; for the purpose of this research, two modifications to the original concept have to be introduced. First, the semi-variance now measures the excess value above a threshold, and not the shortfall as in the original concept. For further investigation, the semi-variance then measures how much variability is left when a certain part of the workload has already been covered by the deployment of a certain VM instance. Second, the α quantile is used as a threshold, not the mean as in the original concept. This change is motivated by the fact that the α quantile of the workload also corresponds to the size of the base VM instance in the elastic tariff. Therefore, the semi-variance sv_i gives an estimate of how much variability is not covered by the base instance. sv_i is defined for application i in the workload data. The semi-variance is only defined on the level of a single application, as the outsourcing decision is also made for each individual application. Furthermore, it is assumed, that the discrete workload function has n sample points and that all sample points contribute equally to the variability. Equation 5.14 is applicable to all resource types.

As the different IT resource types have vastly different scales (e.g. CPU usage is measured in percent, whereas main memory is measured in bytes), the absolute value of the semi-variance will differ considerably across the different resource types. In order to make the semi-variance comparable across different resource types, it is normalized with the sample mean (equation 5.15). This step parallels the computation of the coefficient of variation (CoV), a measure of dispersion, which is also normalized with the mean (Rinne 2008, p. 45); the equation of this measure can be found in equation 5.16. The CoV is utilized in this study as a benchmark for determining the relevance of the semi-variance in this research context. As a result, both measures deliver a dimensionless number that can be compared across different dimensions.

In both the SV and CoV formula, \overline{w}_i^k represents the arithmetic mean of the random variable w_i^k .

$$sv_i^k = \frac{1}{n-1} \sum_1^n \max\left((w_n)_i^k - Q_\alpha(w_i^k), 0\right)^2. \quad (5.14)$$

$$v_i^k = \frac{\sqrt{sv_i^k}}{\overline{w}_i^k}. \quad (5.15)$$

$$cv_i^k = \frac{\sqrt{\frac{1}{n-1} \sum_1^n ((w_n)_i^k - \overline{w}_i^k)^2}}{\overline{w}_i^k}. \quad (5.16)$$

When the normalized semi-variance is applied component-wise to the multivariate workload vector \vec{w}_i , a multivariate semi-variance vector (sv_i^k) is received. In order to make the semi-variance uniquely comparable between two software applications, this semi-variance vector has to be mapped to a scalar; to this end, a vector norm is applied. The single resource type variabilities are aggregated to an application workload variability. The same procedure is applied to the CoV, which yields an aggregated CoV value.

$$\|\vec{v}_i\|_l = \sqrt[l]{\sum_k (v_i^k)^l} = \sqrt[l]{(v_i^{CPU})^l + (v_i^{RAM})^l + \dots} \quad (5.17)$$

As a default, the Euclidean distance ($l = 2$) is used; the City-Block distance ($l = 1$), the Chebyshev distance ($l = \infty$) and the L10 norm ($l = 10$) are evaluated as part of the sensitivity analysis. For a definition of these concepts, please see Rinne (2008, p. 678).

5.5.2 IaaS Instance Type Selection

The IaaS instance type selection is a necessary step to derive the type of VM from the IT resource demands. It serves to determine the purchased amount of CPU and RAM resources. IaaS providers are assumed to offer several types of VM which are equipped with a certain amount of RAM and a number of CPU cores. For example, AWS offers small, medium and large instances as a part of their EC2 service. These IaaS instance types are indexed by $e \in E = \{1, 2, \dots\}$ with E being the set of available IaaS instance types. The in-house instance types are indexed using $g \in G = \{1, 2, \dots\}$ with G being the set of available in-house instance types.

The VM instance size, or the capacity of an instance, is modeled in a capacity matrix for both in-house and IaaS resources. Equation 5.18 and 5.19 define the capacity matrices $H^{IaaS}, H^{In-house}$ for both deployment cases.

$$H^{IaaS} = \begin{pmatrix} h_1^{CPU} & \dots & h_{|E|}^{CPU} \\ h_1^{RAM} & \dots & h_{|E|}^{RAM} \end{pmatrix}. \quad (5.18)$$

$$H^{In-house} = \begin{pmatrix} h_1^{CPU} & \dots & h_{|G|}^{CPU} \\ h_1^{RAM} & \dots & h_{|G|}^{RAM} \end{pmatrix}. \quad (5.19)$$

It is assumed that a total order can be established among the different VM instances, i.e. there exists a binary relation \leq on the column vectors of each of the capacity matrices H^{IaaS} and $H^{In-house}$, which satisfies the total order conditions. As a consequence, there exists one and only one migration path among the VM instances; each VM instance has exactly one successor to which it can be vertically scaled, if the capacity has to be expanded by a given value (vertically scaling may imply the need to migrate the VM instance to another physical server. In any case, scaling up also means adjusting the currently billed rate to the new VM instance type). The case, in which there is a partial order defined on the VM capacities, is not discussed in detail here (a partial order would lead to various simultaneously possible migration paths; finding out the cost-optimal one would in turn be an optimization problem on its own).

One of the decisions, that an IT manager faces when pondering the use of IaaS, is the selection of a sufficiently large VM type for the software application that is supposed to be deployed on that VM type. H^{IaaS} and $H^{In-house}$ define the capacity of the available resource bundles for both deployment cases. Usually, one important activity in capacity management is the sizing of future systems to accommodate the application workload. This activity can be supported by elaborate tooling or experience, but for this research, it is assumed that an independent variable α exists that describes the degree to which the application workload percentage is covered by the VM instance. The size of the VM instance is directly positively correlated to α . Hence, the selection of the appropriate VM instance can happen on the basis of α .

Equation 5.20 defines a function $bf()$ that returns “1”, when called with the right combination of VM instance and software application; it maps the smallest suitable VM instance in the capacity matrix to the application. The first parameter x indexes the VM instance in the capacity matrix, hence x has to be between 1 and $|G|$ for the in-house case, and between 1 and $|E|$ for the IaaS case. The second parameter indexes the software application ($y \in I$). Q_α is the α -quantile of the random variable. Function $bf()$ makes it possible to define a vector \vec{b} that selects the suitable VM instance in the capacity matrix by calculating the matrix-vector product $H \cdot \vec{b}$. Equation 5.21 shows the case for the IaaS VM instances.

$$bf(x, y, \alpha) = \begin{cases} 0 & \text{if otherwise.} \\ 1 & \text{if } Q_\alpha(w_y^{CPU}) \leq h_1^{CPU} \wedge Q_\alpha(w_y^{RAM}) \leq h_1^{RAM} \wedge x = 1. \\ 1 & \text{if } h_{x-1}^{CPU} \leq Q_\alpha(w_y^{CPU}) \leq h_x^{CPU} \wedge h_{x-1}^{RAM} \leq Q_\alpha(w_y^{RAM}) \leq h_x^{RAM} \wedge x > 1. \end{cases} \quad (5.20)$$

$$\vec{b}_i(\alpha) = \begin{pmatrix} bf(1, i, \alpha) \\ \vdots \\ bf(|E|, i, \alpha) \end{pmatrix} \in \{0, 1\}^{|E|}. \quad (5.21)$$

Possible IaaS provider tariffs were already discussed in section 5.4.2. The elastic tariff does not require a dedicated VM instance selection step, as this tariff considers all available instances for the workload at hand (as opposed to the regular tariff which has to pick a single VM instance of suitable size). The elastic tariff therefore has to establish VM type usage frequencies. Equation 5.22 calculates the share of workload $\gamma(x, y)$, which a given VM instance x takes over for a specified application y . If the application resource demand exceeds the capacity of a certain VM type, γ will be 0.

Equation 5.23 defines a function $of()$, which considers the effect of α on $\gamma()$. $\gamma()$ is only relevant, if the IT resource demands surpass the base load VM type, as calculated in equation 5.20. In this case,

$of()$ calculates the usage frequencies of the VM types beyond the base VM type. Please note that the VM instance types $h_1 \dots h_{|E|}$ are supposed to be ascendingly ordered by size, which excludes the case of $\gamma(x, y) \leq \alpha \wedge \gamma(x-1, y) > \alpha$.

$$\gamma(x, y) = P(w_y^{CPU} \leq h_x^{CPU} \wedge w_y^{RAM} \leq h_x^{RAM}). \quad (5.22)$$

$$of(x, y, \alpha) = \begin{cases} 0 & \text{if } \gamma(x, y) \leq \alpha \wedge \gamma(x-1, y) \leq \alpha. \\ 0 & \text{if } \gamma(x, y) > \alpha \wedge \gamma(x-1, y) \leq \alpha. \\ \gamma(x, y) - \gamma(x-1, y) & \text{if } \gamma(x, y) > \alpha \wedge \gamma(x-1, y) > \alpha. \end{cases} \quad (5.23)$$

Function $of()$ makes it possible to define a vector \vec{o} that selects the utilized VM instances of the capacity matrix by calculating the matrix-vector product $H \cdot \vec{o}$. Equation 5.24 shows the case for the IaaS VM instances. Please note that \vec{o} is a vector with real-valued components, whereas \vec{b} is a vector with binary components.

$$\vec{o}_i(\alpha) = \begin{pmatrix} of(1, i, \alpha) \\ \vdots \\ of(|E|, i, \alpha) \end{pmatrix} \in [0 \dots 1]^{|E|}. \quad (5.24)$$

5.6 IaaS Usage Optimization Model

The optimization model for the above defined outsourcing scenario is presented in this section; it is used throughout the rest of the thesis. In the preceding section, the tariff model, the resource demands and the VM instance selection logic were introduced. These elements are combined in the IaaS usage optimization model, in which the cost-optimal deployment for a set of software applications is calculated. This work focuses on a cost-oriented decision model only, as benefits resulting from the usage of IaaS resources are generally hard to quantify. Moreover, the literature review of Gonzalez, Gasco, and Llopis (2006) and the review of Dibbern, Goles, Hirschheim, and Bandula (2004) show that the question of what to outsource has mostly been analyzed conceptually or in a positivist fashion so far, but not through mathematical modeling, even though cost is universally recognized as a dominant criterion for outsourcing (Liker and Choi 2004), (Gottfredson, Puryear, and Phillips 2005).

As a first step, the cost functions per application i for each deployment option are derived. As a second step, the optimization model is formulated using these cost functions. In principle, there is one N-part cost function for every IT resource. The cost functions for the elastic and regular IaaS instances can be seen in equation 5.25 and 5.26. The cost function for the in-house deployment option is displayed in equation 5.27.

$$m_i^{\text{Elastic}} = \vec{p}^{\text{IaaS}} \left(\vec{b}_i(\alpha) + \vec{o}_i(\alpha) \right) + \vec{p}^{\text{IaaS}} \vec{d}_i^{\text{IaaS}} \quad (5.25)$$

$$m_i^{\text{reg}} = \vec{p}^{\text{IaaS}} \vec{b}_i(\beta) + \vec{p}^{\text{IaaS}} \vec{d}_i^{\text{IaaS}} \quad (5.26)$$

$$m_i^{\text{In-house}} = \vec{p}^{\text{In-house}} \vec{b}_i(\beta) + \vec{p}^{\text{In-house}} \vec{d}_i^{\text{In-house}} \quad (5.27)$$

The tariff based on the cost function 5.25 is modeling a hypothetical tariff, because AWS does not offer the functionality assumed in this scenario. To make it more realistic, it has to be assumed that no physical migration takes place, but only an up/downgrade of the current VM to an enhanced set of resources.

Following the past AWS tariff logic (as of 2013-12-01), the base instances would have to be paid in full, but the additional resources would be billed like on-demand instances (but they also enjoyed the same flexibility). This is also the reason why α plays a role in the elastic cost function; it describes the size of the base instances (which are likely reserved instances).

In the optimization model, the total cost of the regular and the elastic tariff, m_i^{reg} and m_i^{Elastic} have to be compared, because it cannot be guaranteed that the elastic tariff is always the cheaper one. Thus, the cost of both tariffs has to be calculated and the cheaper tariff option is chosen. The current model does not include discounts, but they could be easily integrated with the total cost functions defined above. Using the above definitions, the optimization model can be defined in equation 5.28. The decision variables y_i are binary; a value of “1” is equal to the placement of software application i on IaaS servers. However, only the number of ηN_s software applications can be placed in total there.

$$\min m_{\text{total}} = \sum_{i \in I} (\min(m_i^{\text{reg}}, m_i^{\text{Elastic}}) y_i + p_{\Delta} m_i^{\text{In-house}} (1 - y_i)) \quad \text{subject to} \quad (5.28)$$

$$\sum_{i \in I} y_i = \eta N_s \quad (5.29)$$

$$y_i \in \{0, 1\} \quad \forall i \in I \quad (5.30)$$

Although the optimization model can technically be solved using an existing Integer Programming solver software, the simplicity of the model makes it amenable to a much more efficient solution algorithm. The following transformation of equation 5.28 reveals an insight helpful for finding another solution approach.

$$\min \sum_{i \in I} (\min(m_i^{\text{reg}}, m_i^{\text{Elastic}}) y_i + p_{\Delta} m_i^{\text{In-house}} (1 - y_i)) \quad (5.31)$$

$$= \min \sum_{i \in I} (m' y_i + p_{\Delta} m_i^{\text{In-house}} (1 - y_i)) \quad (5.32)$$

$$= \min \sum_{i \in I} (m' y_i - p_{\Delta} m_i^{\text{In-house}} y_i + c') \quad (5.33)$$

m' is an arbitrary cost figure which can be precalculated, c' is essentially a constant, so the goal is to minimize a sum of differences. The minimum of this sum is obviously the sum of the top ηN_s smallest differences. To find those top elements, two principle approaches are possible: the PICK algorithm (Blum, Floyd, Pratt, Rivest, and Tarjan 1973) and a sorting procedure of the complete list (e.g. using QuickSort (Musser 1997)). For Quicksort, “the average computing time on uniformly distributed inputs is $\Theta(N \log N)$ ” (Musser 1997); for PICK, the worst-case computing time for picking the i^{th} -largest element is $\Theta(N)$ (Blum, Floyd, Pratt, Rivest, and Tarjan 1973). No matter what approach is chosen, the computational complexity is negligible. Hence, this optimization problem can be solved exactly and efficiently even for a very large number of software applications. The availability of an exact solution is a preferable property of an optimization model; if an optimal solution exists, it will be found. The solution algorithm is not going to get stuck in local optima, it always returns the global optimum.

5.7 Decision Tree

The last sections introduced models for the controlled inputs, the uncontrolled inputs and the calculation of the placement decisions (as a proxy for the actual decisions made during outsourcing). This section explains the rationale behind the selection of the decision tree algorithm for finding an approximation function for the calculated placement decision. First, the methodology is motivated in section 5.7.1, then the set-up of the tree is discussed in section 5.7.2 and finally, the quality metrics for assessing the tree's predictive performance are listed in section 5.7.3 based on recommendations from research literature.

5.7.1 Motivation for Analytical Approach

First, an explorative approach to understanding the drivers of the application placement decision is required. The relevant independent variables are not known beforehand, hence all variables of the outsourcing scenario (41 variables in total, see Table 5.3) must be included in the analysis. Second, the functional relationship between the dependent and the independent variables is unclear, but it is likely that the function is highly discontinuous due to the non-linearities in the provider tariffs. Therefore, the analysis of the placement decision cannot be accomplished by fitting a pre-specified function.

For exploring the interdependencies in this high-dimensional space, a decision tree is chosen according to the criteria put forward by Kotsiantis (2007). He compared a wide range of different machine learning algorithms, but only Decision Trees show the combination of favorable features needed in this classification task. Especially their explanation ability sets them apart from the other approaches, i.e. the resulting decision tree is very easy to interpret and the relevant decision drivers are directly visible. Moreover, the independent variable "Deployment Target" (abbrev. TARGET) in this learning task is binary scaled (only two values "iaas" and "in-house"), which further limits the possible approaches. From a statistical standpoint, logistic regression and discriminant analysis can handle nominally scaled independent variables (Backhaus et al. 2006, p. 12); however, they are reliant on metrically scaled independent variables and they assume a linear model (please see (Backhaus et al. 2006, p. 186) for discriminant analysis and (Backhaus et al. 2006, p. 249) for logistic regression), which is unlikely in this case, as explained above.

Other statistical learning algorithms include Bayesian ones like Naive Bayes classifiers and Bayesian Networks. Naive Bayes classifiers rely on the statistical independence of the input variables (Kotsiantis 2007, p. 257), which is clearly not the case of this data set. Furthermore, neither type of the two Bayesian classifiers can handle metric input variables; these would have to be discretized (Kotsiantis 2007, p. 259). However, this input variable transformation step brings up the problem of finding adequate bin sizes. Generally, Bayesian classifiers seem to be a poor fit for the classification task at hand.

Neural Networks are pretty flexible in terms of inputs and outputs and can be considered as another option. They are not suitable for this classification task, as their structure (no matter how the neural network itself is grown), is not amenable to interpretation. The weights and nodes of the network do not bear any direct semantic relationship to the input variables. Therefore, the neural network structure does not reveal the leverage of the variables and hence it cannot answer the research question in this study. Network-type algorithms (neural networks and support vector machines) generally produce network models that are poorly interpretable (Kotsiantis 2007, p. 263).

Kotsiantis (2007) additionally includes kNN (k-Nearest Neighbors) as a classification algorithm, but also mentions the drawbacks of this approach. One of the most important ones for this study is again the interpretability of the algorithm's output, which consists in this case of 2 subsets of maximally similar training cases. The algorithm itself does not indicate which input variables were especially relevant for making the clustering decision; this judgment would be left to the experimenter. Consequently, this algorithm is not suitable for this study.

The principle workings of the algorithm for decision trees in the RapidMiner data mining software is described in the online documentation: “This decision tree learner works similar to Quinlan’s C4.5 or CART. Roughly speaking, the tree induction algorithm works as follows: whenever a new node is created at a certain stage, an attribute is picked to maximize the discriminative power of that node with respect to the cases assigned to the particular subtree. This discriminative power is measured by a criterion which can be selected by the user (information gain, gain ratio, Gini index, etc.)” (Rapid-I 2010). A node in this context is associated with an independent variable and it splits the domain of this variable in two regions, such that the cases in each of the regions differ maximally in terms of the independent variable. A description of the general concepts of decision trees and their application in classification tasks can be found in (Maimon and Rokach 2010) and (Tan, Steinbach, and Kumar 2006). Details on CART are available in (Breiman et al. 1993) and Quinlan’s C4.5 tree is documented in (Quinlan 1993).

5.7.2 Decision Tree Set-up

A decision tree is applied to the classification task of finding decisive factors that should guide a cost-based outsourcing decision. The independent input variables are listed in Table 5.3 and encompass all aspects of the decision problem (outsourcing scenario parameters, resource-specific cost figures, software application workload figures and workload variability measures). VARIAB and AGGCOEFVAR are calculated according to equation 5.17 using the controlled input parameter L_{norm} . Some independent variables of the optimization model are excluded as input variables. APP, the label for the individual application, is excluded as the focus of the analysis rests on the general cost and workload properties and not on individual application names. NWI_IH and NWO_IH are excluded, as the in-house cost model does not assign a separate price to network traffic, hence the value would always be zero and no predictive value can be expected. The scenario parameter β is excluded, as it is effectively a constant in this evaluation and therefore is also not helpful in a classification task. The complete decision tree and the input variable values are specified at the single application level.

The decision tree applied here is a complex algorithmic approach to a learning task, and has several parameters that determine its learning behaviour. The settings of these parameters can be found in Table 5.4; these settings will henceforth be referred to as base settings.

The base settings given in Table 5.4 are used for a couple of reasons. First, they are the default values in the data mining suite used.⁵ They can be expected to work reasonably well for a wide range of learning tasks. Second, the results of a sensitivity analysis prove the validity of these parameter setting. The values for the parameters “Minimal size for split”, “Minimal leaf size”, “Minimal gain” and “Confidence” were changed systematically and the resulting performance (accuracy) were recorded. All in all, 374 parameter configurations were tried; Appendix H shows the results. For the performance measure “accuracy”, the above base settings were among the top performance settings in the sensitivity analysis. Third, the maximal depth of the tree has to be constrained to a sensible (albeit subjective) limit. A high maximal depth value produces a more elaborate, complex tree with a higher classification performance, but in turn this complexity decreases the interpretability of the resulting tree because of the numerous cases. A low maximal depth value diminishes the classification performance, but creates a smaller tree which is less susceptible to overfitting (Tan, Steinbach, and Kumar 2006, p. 181). Therefore, the value in the base settings is considered to be a good compromise. The decision made in the base settings for the maximal tree depth can be evaluated as well using decision tree performance metrics (the exact definition of the three metrics accuracy, precision and recall is given below in section 5.7.3). Figure 5.9 shows the development of the three evaluation metrics based on the permissible layers of the unweighted decision tree (on the x-axis). As expected, the tree depth positively affects all three metrics, although precision and recall tend to swing wildly especially for smaller

⁵RapidMiner 5.0 <http://www.rapid-i.com>, last accessed 2013-12-29

Table 5.3: Input variables of the decision tree

Variable Type	Input Variable Name	Explanation	Unit
Parameters	OSDEGR (η), ADELTA (a_{Δ}), PDELTA (p_{Δ}), QALPHA (α)	outsourcing scenario parameters	n.a.
Total Cost Ratios	DYNRATIO	Total cost for elastic IaaS tariff / Total cost for in-house tariff	n.a.
	RESRATIO	Total cost for reserved IaaS tariff / Total cost for in-house tariff	n.a.
Resource-specific Cost Ratios	AINSTDYNRATIO	Base part of the AWS instance cost (elastic IaaS tariff) / Total cost for elastic IaaS tariff	n.a.
	AINSTRESRATIO	AWS reserved instance cost / Total cost for reserved IaaS tariff	n.a.
	DYNINSTRATIO	Elastic part of the AWS instance cost (elastic IaaS tariff) / In-house instance cost	n.a.
	OINSTRATIO	AWS reserved instance cost / In-house instance cost	n.a.
	OINSTDYNRATIO	AWS reserved instance cost / Total cost for elastic IaaS tariff	n.a.
	OINSTRESRATIO	AWS reserved instance cost / Total cost for reserved IaaS tariff	n.a.
	INSTIHRATIO	in-house instance cost / Total cost for in-house tariff	n.a.
	NWIDYNRATIO	Inbound IaaS Networking cost / Total cost for elastic IaaS tariff	n.a.
	NWIRESRATIO	Inbound IaaS Networking cost / Total cost for reserved IaaS tariff	n.a.
	NWODYNRATIO	Outbound IaaS Networking cost / Total cost for elastic IaaS tariff	n.a.
	NWORES RATIO	Outbound IaaS Networking cost / Total cost for reserved IaaS tariff	n.a.
	STOIHRATIO	In-house Storage cost / Total in-house Cost	n.a.
	STODYNRATIO	IaaS Storage cost / Total cost for elastic IaaS tariff	n.a.
	STORESRATIO	IaaS Storage cost / Total cost for reserved IaaS tariff	n.a.
STORATIO	IaaS Storage cost / In-house Storage Cost	n.a.	
Workload Levels	CPU_EXP, CPU_QA, CPU_QO	mean, α, β quantiles of CPU core count	CPU cores
	RAM_EXP, RAM_QA, RAM_QO	mean, α, β quantiles of RAM consumption	Gbytes
	NWI_EXP, NWI_QA, NWI_QO	mean, α, β -quantiles of inbound network consumption	Gbytes
	NWO_EXP, NWO_QA, NWO_QO	mean, α, β -quantiles of outbound network consumption	Gbytes
	STO_EXP, STO_QA, STO_QO	mean, α, β -quantiles of storage utilization	Gbytes
Workload Variability	CPU_SV, RAM_SV, STO_SV, NWI_SV, NWO_SV	Resource-specific semi-variances	n.a.
	CPU_CV, RAM_CV, STO_CV, NWI_CV, NWO_CV	Resource-specific coefficients of variation	n.a.
	VARIAB, AGGCOEF-VAR	Aggregated variability, aggregated coefficient of variation	n.a.
Case Weighting Factor	CASEWEIGHT	(COSTIAAS+COSTIAASRES+COSTIH)/3 Only used for weighted decision trees	Dollars

Table 5.4: Base settings for the decision tree parameters (based on (Rapid-I 2010))

Parameter	Value	Description
Criterion	Gini Index	
Minimal size for split	5	The minimal number of cases in a node in order to allow a split
Minimal leaf size	3	The minimal number of cases in each of the leaves of the tree
Minimal gain	0.1	Threshold for producing a split
Maximal depth	6	Size limitation of tree depth
Confidence	0.3	The confidence level used for the pessimistic error calculation of pruning (Maimon and Rokach 2010, p. 176)
Number of prepruning alternatives	3	The number of alternative nodes tried when prepruning would prevent a split.

layer counts. As more layers are added to the decision tree, the three evaluation metrics seem to converge and the variations for layers 4 to 7 become relatively small.

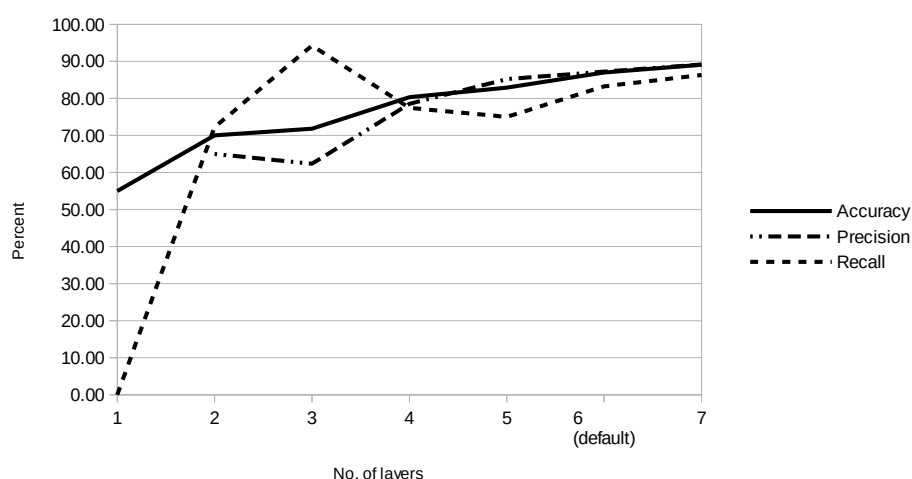


Figure 5.9: Effect of decision tree depth on classification quality

The criterion “Gini index” is a measure for the discriminative power of a node in the tree. Several other measures would be possible here, especially information gain and gain ratio (Wu et al. 2008). However, in a series of decision tree test runs using the base settings, neither the information gain nor the gain ratio produced superior results compared to the Gini index, so the latter is chosen (Raileanu and Stoffel (2004) come to a similar conclusion on theoretical grounds. The Gini index and the information gain criteria produce almost identical split criteria and hence do not differ in discriminative power). Pre-pruning and post-pruning steps are executed during the training phase in order to mitigate the risk of overfitting (Tan et al. 2006, p. 184). These steps stunt the tree growth before the tree perfectly fits the training data (i.e. before overfitting takes place).

A definition of the cost measure used for weighting needs to be developed. Table 5.3 includes the variable CASEWEIGHT which is applied as a weighting factor. It is defined as the average total cost across all three tariffs (elastic, reserved and in-house). This definition ensures the relevance of the weighting factor for all cases and it also guarantees that more expensive cases in terms of total cost receive a higher weight. The choice of this weighting function is guided by simplicity (a linear function), but many other,

more complicated weighting functions would be imaginable. In any case, the weighting function must not include any reference to the cost-optimal placement value calculated by the optimization model and has to rely solely on the variables in Table 5.3 (or combinations thereof). If the target variable and any input variable were conflated in any way, the learning algorithm would pick this input variable for its special predictive qualities. Hence, both the tree and its performance metrics would be flawed.

5.7.3 Decision Tree Performance Evaluation

Both the unweighted and the weighted decision trees are evaluated using a 10-fold cross-validation procedure with stratified sampling; cross-validation is a widely-used method for evaluating the performance of a classifier on previously unseen data (Tan, Steinbach, and Kumar 2006, p. 187). Cross-validation is a “method for estimating the accuracy (or error) of a prediction algorithm by dividing the data into k mutually exclusive subsets (the ‘folds’) of approximately equal size. The algorithm is trained and tested k times. Each time it is trained on the data set minus a fold and tested on that fold. The accuracy estimate is the average accuracy for the k folds” (Kohavi and Provost 1998, p. 272).

A confusion matrix (Kohavi and Provost 1998, p. 272) is a contingency table (cross tab), which contrasts the predicted and the actual classification results of a machine learning algorithm. The decision support model is concerned with the independent variable “Deployment Target” in this learning task, which is binary scaled (only two values “iaas” and “in-house”); hence, the confusion matrix is of size 2×2 . The following confusion matrix shows an example.

	pred. in-house (cases)	pred. iaas (cases)
actual in-house (cases)	a	b
actual iaas (cases)	c	d

The confusion matrix is a useful concept for evaluating the results of classification algorithms, as a number of key quality metrics can be derived from it. The decision trees are evaluated based on the three metrics accuracy, precision and recall, which are commonly applied in the assessment of machine learning algorithms (Kohavi and Provost 1998, p. 271) and which are calculated for each of the 10 iterations in the cross-validation procedure. Accuracy is defined as “the rate of correct predictions made by the model over a data set” (Kohavi and Provost 1998, p. 271); correct predictions are related to both IaaS and in-house cases. Precision is defined as the ratio between actual IaaS deployments and the total number of predicted IaaS deployments (correct and false ones) (Kohavi and Provost 1998, p. 272). Recall, the true positive rate, is defined as the ratio between actual IaaS deployments and the total number of predicted deployments (IaaS and in-house ones) (Kohavi and Provost 1998, p. 272).

$$\text{Accuracy} = \frac{a + d}{a + b + c + d}. \quad (5.34)$$

$$\text{Precision} = \frac{d}{b + d}. \quad (5.35)$$

$$\text{Recall} = \frac{d}{c + d}. \quad (5.36)$$

In order to be able to interpret and to put in context the classification performance of the aforementioned decision tree, there must be a base line classification performance to compare against. This baseline in this case is the classification performance on the results of the random outsourcing process (see section 5.7). Again, the decision tree is trained using the base settings, but the values of the independent variable are chosen randomly from the two deployment options. The random tree is again evaluated using a 10-fold cross-validation procedure with stratified sampling. The decision tree performance can not only be

compared to random placement decisions but to completely biased decisions (all “iaas” or “in-house”), favoring exclusively IaaS or in-house placements. Both baseline metrics are evaluated in section 6.5.4.

Another way of evaluating the performance of a classification algorithm is the Matthews correlation coefficient (MCC), which is well-known in the Machine Learning community (Baldi, Brunak, Chauvin, Andersen, and Nielsen 2000), (Powers 2011). It measures the correlation coefficient between the binary classification variable, the predicted placement, and the binary input variable, the actual placement. An MCC value of 1.0 signifies a perfect classification; a value of 0.0 indicates independence, hence completely random predictions.

For a deeper evaluation, the question whether there are circumstances under which each tree exhibits a heightened error rate, is interesting. These circumstances are described in terms of the partitions that the tree creates through its leaf structure, as the set of the cases in the tree leaves is a partition of the set of all cases (i.e. each individual case ends up in exactly one of the tree leaves). The error rate in this study is defined as the cost deviation between the total cost associated with the placement prediction and the optimized total cost (as calculated by the optimization model). The predicted total cost and the optimized total cost is available for every case, hence all the cases in one leaf can be aggregated and the resulting cost deviation can be calculated. If the prediction matches the optimized deployment decision for some case, then the predicted total cost and the optimized total cost are equal and no cost deviation occurs for this case. If the prediction does not match the optimized deployment decision for another case, then the predicted total cost and the optimized total cost are usually different and a cost deviation may occur for this case. Please note, that such a cost deviation does not always have to be positive (i.e. more expensive). Under certain circumstances, the predicted total cost for a case can be lower than the optimized total cost, so a wrong prediction can actually save money. When this logic is applied to the above mentioned decision trees, the cost deviations per tree leaf can be computed.

5.8 Discussion and Summary

Two strands of modeling are described in this chapter, one of which is quantitative and the other is qualitative. The quantitative model is developed as part of an experimental research methodology that targets the IaaS outsourcing process and that focuses exclusively on economic criteria. The context of these experiments is defined in terms of outsourcing scenario assumptions. Quality is considered as a largely uncontrolled input in the above quantitative model; however, a few general conditions on quality attributes of IaaS resources are needed for the comparability of different resource offerings. The decision process (based on the aforementioned IaaS tariff model, resource model and cost optimization model) is established and a decision tree is used as the decision process’ approximation model. In combination, these elements form a complete quantitative decision support tool for the IaaS outsourcing process. The qualitative model is developed using a case study-based research methodology; it complements the economic considerations of the quantitative model. In this chapter, dimensions relevant for assessing IT infrastructure quality are identified from the research literature and are compiled into an IaaS quality model.

The following discussion reflects on the most important scenario assumptions for the quantitative model. One of the basic principles is the summary calculation of the economic value of the different decision options. The economic value is seen as the product of the summary resource demands and the applicable tariffs. As a consequence, time series (e.g. for IT resource demand), which are the temporal sequence of resource requirements, are not needed for this model. Within the time slot under investigation, it plays no role, when exactly a particular IT resource is needed. The cost is always the same. This model property corresponds with other cost-based models (e.g. (Lilienthal 2013; Chaisiri et al. 2011)). The irrelevance of the temporal structure can only be upheld, as long as tariff attributes do not interfere with it. Tariff options like bundling or billing cycles impose a temporal structure on the IT resource demand; this structure will

have to be respected by the cost-based approaches, if they aim to compute the multi-period economic value. Both multi-period and sophisticated tariff options are currently not included in the models in this chapter, but they are briefly outlined in section 7.2 as an outlook.

In the outsourcing model, η , the outsourcing degree, is assumed to be exogenous and to be set by corporate IT management, i.e. only a certain share of the suitable business software applications are deployed externally. As an alternative, the outsourcing degree could also be determined as an endogenous variable, which means, the value of η could be discovered or calculated within the model itself. This approach would result in placing software applications solely based on cost efficiency. Both alternatives are valid modeling options, but they differ in the way they support a business executive in its decision-making process. Using a model with an exogenous variable η , the decision maker first selects the remaining model parameter values according to his or her outsourcing scenario. Then the cost-optimal deployment options are calculated for each level of η . The executive can then pick the level of η that promises the lowest cost (but he or she also sees the cost values of other levels of η). Using a model with an endogenous variable η , the decision maker is only presented with one cost-optimal deployment and the corresponding value of η . Either way, a cost-optimal solution is reached, but the model with an exogenous variable η offers greater transparency. This feature comes at a price: the η values need to be discrete and would most likely be chosen such, that they are evenly distributed in the $[0 \dots 1]$ interval, whereas the values of an endogenous variable η would be continuously distributed; hence, the optimal value of η could be determined more exactly. If the goal is to provide the decision-maker with actionable options, then the model with an exogenous variable η is preferable, as the outsourcing degree might not be solely depending on the total cost, but also on organizational or environmental considerations (whose influence was shown in chapter 4); a conscious deviation from the cost-optimal η might then be called for.

Similarly, the level of base VM instance utilization α from the IT resource model in section 5.5.1 is an exogenous variable. The reasoning for this modeling choice resembles the one for the outsourcing degree η , as an explicit assignment of α makes the economic effect of this variable more clearly visible. However, a decision-maker might not be interested in the level of α per se, as it is conceptually on a much more technical level than the outsourcing degree. From a decision support standpoint, making α an exogenous variable does not preclude decision-makers from identifying and selecting the cost-optimal level of α themselves, so the explicit modeling of α does not limit the model's utility or informative value. As an endogenous determination of α results in a substantially more complex optimization model (see section 7.2), the aforementioned approach was chosen.

The role of risk, decisions under uncertainty and their relations to the presented model shall be discussed in more detail. Possible risks in the outsourcing scenario were already identified in the empirical research, e.g. through constructs like "Perceived Uncertainty", "Fear of Provider Opportunism" and "Information Security Concerns" defined in section 4.2.2. Examples of such risks are provider-related SLA violations (Michalk 2011) or uncertainty about the actual IT resource demand and the cost associated with it, especially in the face of constantly-changing IaaS provider tariffs. In general, the current quantitative model does not evaluate these risks; they are reflected in the quality model (see section 5.3), where quality dimensions for measuring and for evaluating IaaS service quality are devised. The decisions supported by the quantitative model are decisions under certainty; the model is designed to mitigate any possible IT resource availability risk by purchasing sufficient IT resources. The explicit incorporation of uncertainty (e.g. relating workload) would be a possible extension of the current decision support model; it is described in more detail in the outlook section 7.2.

The application of the variability measure semi-variance, which has a risk management background, does not conflict with the deterministic nature of the decision support model. As the model aims at identifying IaaS usage determinants, it has to include workload variability as one potential determinant. Variability is calculated in this model on an application level only, as the single instance is in focus. The literature

offers two principle domains from which to choose variability measures: variability measures with a statistical background (e.g. variance, coefficient of variation) and variability measures with a risk management background (e.g. lower partial moments (LPM) (Unser 2000) with the semi-variance as a special case). Among the many possible variability measures, semi-variance and the coefficient of variation are selected as the two variability measures. The coefficient of variation measures variability symmetrically around the mean, whereas the semi-variance only measures the variability above the α -quantile. As the two variability measures are part of the input variable set of the decision tree, the experimental evaluation is able to reveal whether symmetric or asymmetric variability measures are more effective determinants of IaaS usage.

Section 5.5.2 describes a VM instance type selection scheme for IaaS providers, which is an integral part of the overall cost calculation. Other researchers (e.g. (Lilienthal 2013; Chaisiri et al. 2011)) use a different approach; they simply assume a given base VM instance type and linearly scale this instance type to the required size. The compute costs are calculated accordingly. For example, a base VM instance might have one CPU core and one GB of RAM; all subsequently used VM instance types and their costs are small multiples of this size. The underlying assumption is that VM instances and their associated tariffs are continuously scalable. But this assumption does not hold up in reality (see Table F.1 in the appendix for an exemplary list of Amazon AWS VM instance types). VM instance type offered by real-world IaaS providers might very well be specially equipped with disproportionately more CPU cores or more main memory than an arbitrary base VM instance type. Therefore, the VM instance type selection scheme developed in this chapter relaxes this assumption and merely requires that a total order can be established on the set of different VM instance types, i.e. all VM instance types can be unambiguously ranked. The VM instance types themselves can feature any combination of CPU cores and RAM resources.

This chapter defined the notion of elasticity for IaaS resources; however, the IT resource model in section 5.5.1 also introduced the notion of limited flexibility for in-house IT resources (e.g. reservations needed for storage and reserved tariff only for computing resources). These flexibility differences are based on an assumed gap in technological capabilities between the IaaS provider and the client enterprise. The IaaS provider's technological superiority leads to the advanced flexibility of its data centers, and hence to the elasticity of IaaS resources. The assumption of this superiority can be explained by two factors, economies of scale and IT architecture maturity. IaaS providers can take advantage of effects of scale and scope in their IT infrastructure operations as compared to regular enterprises (Hamilton 2008) and exhibit IT infrastructure growth rates unlike any regular enterprise unrelated to the Cloud (Hamilton 2014). Hence, they develop specialized capabilities in building and running large data centers effectively and efficiently, which are not reproducible by regular enterprises. Moreover, the flexibility and the cost efficiency of a data center depend on the architectural maturity of the IT function. Enterprise IT architecture refers to "the organizing logic for applications, data and infrastructure technologies, as captured in a set of policies and technical choices, intended to enable the firm's business strategy" (Ross 2003). According to Ross and Beath (2006), IT architecture maturity can be distinguished in four levels, which build on top of each other. The IT resource model in this chapter places the IT architecture maturity of a regular enterprise below the maturity of an IaaS provider for the same reasons as above, which also contributes to the flexibility differences. Models for evaluating and measuring IT architecture maturity have been developed by Perko (2008, pp. 228) and Engels (2007, p. 48). The consequences of these flexibility differences eventually manifest themselves in higher prices for in-house IT resources (partly due to reservation costs); nonetheless, IT outsourcing arrangements can be used to help an enterprise move from one architectural stage to the next. If an enterprise manages to develop its internal technical capabilities to a higher level of architecture maturity, the necessity of IaaS service sourcing might have to be questioned and investigated critically, as the technological gap between the in-house resources and the IaaS resources closes.

Chapter 6

Decision Support Model Evaluation

The following chapter evaluates the approximation function put forward in section 5.7. Thereby, it aims to provide answers to research question R3 by identifying factors in an outsourcing scenario that can be linked to an economically beneficial usage of IaaS offerings. First, IaaS' potential to fulfill corporate IT quality requirements is discussed in section 6.3; this will highlight the potential benefits and the shortcomings of IaaS resource usage. Second, the cost aspects are covered in section 6.5.4, where the IaaS placement model is evaluated experimentally as part of the vendor selection step in the outsourcing process. Both the quality discussion and the placement analysis are based on the BMW case study.

6.1 Evaluation Approach

The outsourcing scenario definitions $S_{1...n}$ (please see section 5.5) serve as an input for the two vendor selection processes. The results of each vendor selection process is a deployment decision for each software application in each scenario. These deployment decisions act as an input to the IaaS deployments per scenario (as defined by the outsourcing degree). Both the outsourcing scenarios $S_{1...n}$ and the IaaS deployments per scenario provide input parameters for the outsourcing driver deduction step (section 5.7.2).

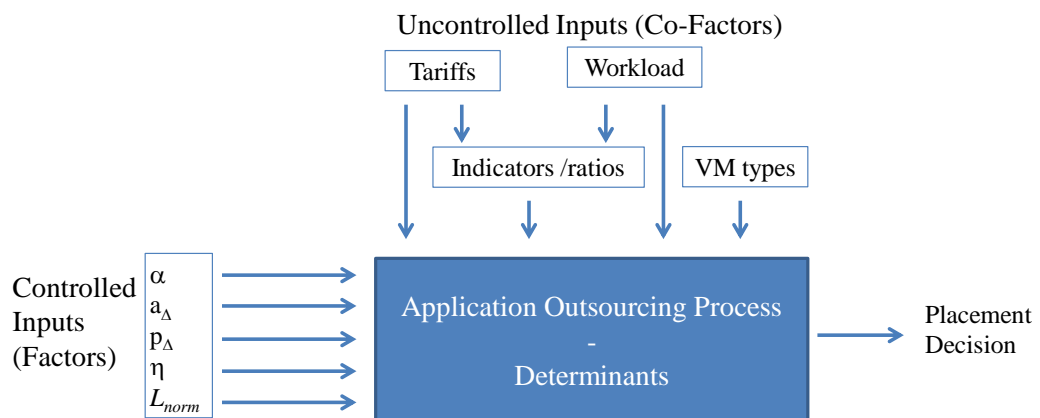


Figure 6.1: Experimental Process model

6.2 Statistical Design of Experiments

In section 5.5, the set of outsourcing scenarios S and its defining parameters were introduced. In the context of the scenario, the basic cost optimization model was derived in equation 5.28; in it, the total cost made up the objective function. An evaluation function F_s for such an outsourcing scenario is established in equation 6.1. This function maps the parameter set of one specific outsourcing scenario on the total cost of this scenario as defined by the basic optimization model target function.

$$F_s : s \in S \rightarrow m_{\text{total}}. \quad (6.1)$$

The optimization of scenario s follows a cost-minimizing approach. It attempts to find the optimal selection of ηN_s applications for IaaS placement by minimizing the financial outcome m_{total} per scenario.

The experimental approach uses the following treatments:

- Level α for base instance resource usage
- Levels of a_Δ
- Levels of p_Δ
- Degree of IaaS usage η
- L_{norm}

These treatments are modeled as factors with varying number of factor levels. The following sets show the factor levels of each factor. The levels were chosen to both cover a reasonable subset of the total input space and to limit the combinatorial explosion in the experimental design which resembles a full factorial one (when speaking in terms of statistical design of experiments (Rinne 2008, p. 799)). The domain of a_Δ assumes that the IT cost structure of most enterprises is at most 30% more expensive or 30% cheaper than the BMW cost structure. As BMW IT resource prices are comparable to other enterprises in the marketplace according to BMW-internal benchmark studies, -30% as a lower bound on internal IT resource prices seem reasonable. The +30% as an upper bound are chosen as a sensible value for the markup on commodity IT resources (theoretically, the IT cost structure of an enterprise might be arbitrarily bad). The domain of p_Δ assumes that the tariff structure of most IaaS providers is at most 50% more expensive or 50% cheaper than the Amazon AWS cost structure. As AWS is one of the biggest public IaaS providers, the economies of scale gained from that position should make it hard for contenders to sustainably undercut AWS. L_{norm} , the norm for calculating the application workload variability, looks at two obvious choices, the City-Block distance and the Euclidean norm. For comparison purposes, the Chebyshev distance is also included.

$$\begin{aligned} \alpha &\in \{0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1\} \\ a_\Delta &\in \{0.7; 0.8; 0.9; 1.0; 1.1; 1.2; 1.3\} \\ p_\Delta &\in \{0.5; 0.6; 0.7; 0.8; 0.9; 1.0; 1.1; 1.2; 1.3; 1.4; 1.5\} \\ L_{\text{norm}} &\in \{1; 2; \infty\} \\ \eta &\in \{0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8\} \end{aligned}$$

The complete decision tree and the input variable values are specified at the single application level. Hence, the total number of training cases for a decision tree at a certain L_{norm} level are 338800 ($= 6776 \frac{\text{scenarios}}{\text{application}} * 50 \frac{\text{applications}}{\text{scenario}}$). Also, all cost figures and cost ratios are application- and scenario-specific.

The number of suitable software applications N_s is not used as a treatment; every scenario is launched with the maximum number of possible applications. As a consequence, the workload distributions W associated with these software applications are also reused for each scenario. Moreover, the value of $\beta = 0.9999$ is considered a constant throughout the study (β is a percentage for a workload percentile that reflects the almost certain maximum level of resource usage).

The outsourcing scenario includes one IaaS provider “Amazon AWS” and considers an outsourcing time interval t_Δ of one calendar year (365 days) as a time-frame for all calculations. This time interval corresponds to the minimum subscription time of the reserved AWS server instances. The above experimental approach leads to $|S| = 11 * 11 * 3 * 8 * 7 = 20328$ potential outsourcing scenarios for the single IaaS client “BMW” examined in this evaluation.

The algorithm which was followed during the evaluation is described in algorithm 1. Its output consists of the values for the optimization model decision variables y_i and the cost statistics per scenario in the variable m_{total} .

Algorithm 1: Scenario evaluation steps
Input: S, W, I
Output: set of $(y_i, m_{total})_{1... S }$
foreach $s \in S$ do
Set parameters $(\eta^s, \alpha^s, p_\Delta^s, a_\Delta^s, L_{norm}^s)$;
Workload variability calculation per application from W ;
Calculation of resource demand distributions per application from W ;
Evaluate F_s ;
Store $y_{1...N_s}^s$ and m_{total}^s ;
end

Storing the results is the last step in algorithm 1. The logical data model used for this purpose during the evaluation can be seen in Figure 6.2. It is an entity-relationship model, in which the fundamental objects in the domain of discourse are modeled as entities (such as instance or applications) and the interdependencies among those objects are modeled as relationships; both entities and m:n relation structures are stored as database tables in the physical data model. An Oracle 11g Express Edition database facilitated the storing.¹

The table OPTAPPRES records the results of the scenario evaluations. The table SZENARIO stores the scenario parameters. The scenario master data can be found in the tables CTYPE (cost type), APPS, TARIFF, INSTANCE, PROVIDER. The LOAD table hosts the resource demand distributions per application.

6.3 Quality Requirements in the Case Study

The discussion of IaaS’ potential to fulfill the IT quality requirements in the case study is a three-step process. First, the quality dimensions that are relevant to the BMW storage infrastructure are described in detail. The second step involves the mapping of these quality dimensions on the generic quality model (c.f Table 5.1) to ensure coverage and representativeness. Finally, the quality attributes of BMW’s quality dimensions are compared and contrasted to the ones of a typical IaaS provider, Amazon AWS, and the outcomes are discussed.

¹<http://www.oracle.com/technetwork/products/express-edition/downloads/index.html>, last accessed 2013-02-21.

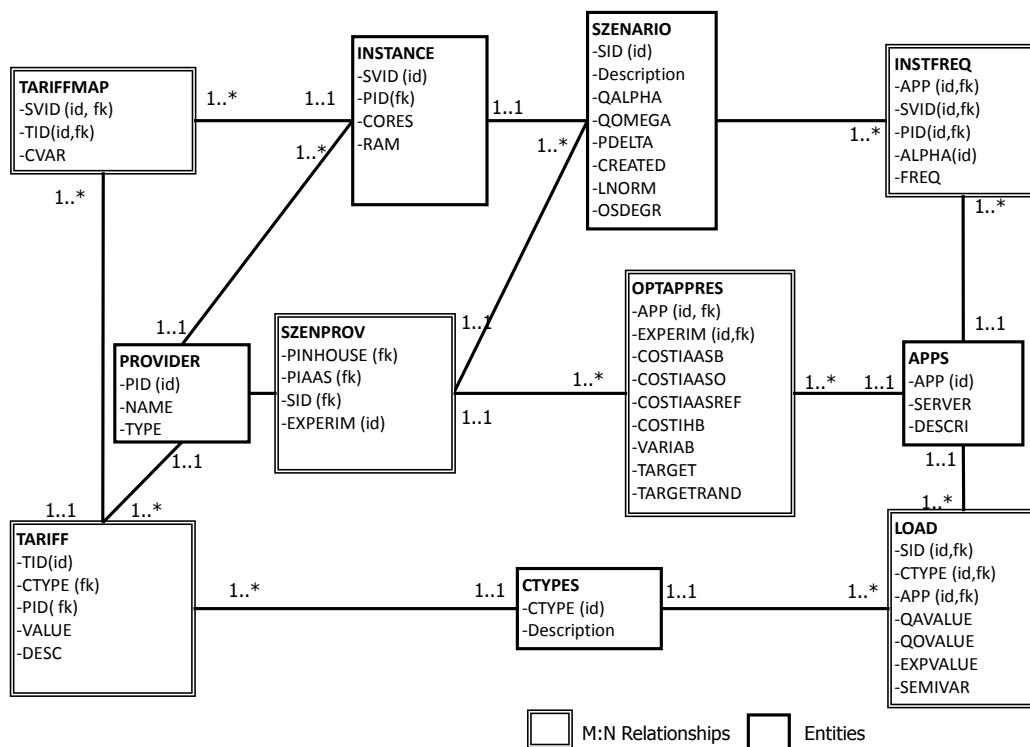


Figure 6.2: Logical data model for evaluation data

6.3.1 Description of Relevant Quality Dimensions

The quality dimensions for the storage infrastructure are defined in several internal documents (BMW Group 2012b), (BMW Group 2012c). The following paragraphs summarize the relevant parts of these documents. The storage system-related quality requirements will be described first, followed by the operations-related requirements.

- The first system-related requirement is the existence of defined storage infrastructure operations hours, i.e. hours in which the system is nominally available. The storage performance measures the technical parameters of the storage system. Relevant attributes of this quality dimension include the type of disk drive (SSD, SAS), the Read Cache Hit Rate, the Read-Write Cache to capacity ratio, the I/O response time as measured on the server and the I/O response time as measured at the storage system connector. Each of the above mentioned performance attributes is associated with performance guarantees, i.e. specific target values for metrics that are monitored during operations.
- All storage systems are monitored by the operations service unit. This monitoring comprises a wide range of attributes of the storage system itself, e.g. performance, thresholds, interruptions or break-downs of hard/software, processor and cache utilization, throughput, and the surrounding infrastructure e.g. network switches and connectivity. These monitoring informations are centrally logged; in critical situations, an alerts is created and a ticket is dispatched to the service team.
- Reporting as a quality aspect serves to inform the business executives about the storage system status. It is based on aggregated monitoring data and gives an overview of filling levels, free capacity, key performance metrics, etc.
- Disaster protection is a stability feature, as it guarantees the uninterrupted system usage even in the case of a catastrophic failure at the data center. The storage system automatically copies all data to

two separate locations and switches transparently for the user between these locations in case in case of a disaster.

- A Single Point Of Failure (SPOF) in a technical system exists, if the failure of one of its components disables the complete system. If storage systems need to be highly available, there must not be a single point of failure in the system (usually guaranteed by the automatic deployment of redundantly available components such as multiple network cards, power units or disk drives).
- File system snapshots are an important feature of a storage infrastructure and are defined as follows: “Snapshot is a common industry term denoting the ability to record the state of a storage device at any given moment and preserve that snapshot as a guide for restoring the storage device in the event that it fails. A snapshot primarily creates a point-in-time copy of the data.” (Garimella 2006) File system snapshots are transparent to the user, they are executed without service interruption or degradation.
- Backup (Tape), as another means of recovering data in case of a disaster, is executed periodically. It uses magnetic tape as its storage medium. The recovery time, i.e. the time needed to restore the data from backup, is one important aspect of this quality requirement.
- Rights and Access Administration describes a software infrastructure that defines user roles and their permissions in terms of storage access and manipulation. It also offers the functionality to grant rights to other users. It is based on the user concept of the underlying operating system (i.e. Microsoft Windows or Unix/Linux).
- Software licenses and software maintenance includes software components and their licencing fees required for the operations of the storage infrastructure like transfer protocol services (CIFS, NFS, sFTP, etc.), management, monitoring and alerting components and backup/snapshot, restore components and virus scanners as the file storage infrastructure might need to be systematically scanned for computer viruses.
- The subscription time is the minimal time period in which a customer is obliged to keep (and pay for) the ordered storage. At the end of the subscription time, the storage order can be canceled, and the cancellation period defines, when this cancellation goes into effect (until then the customer is still contractually bound). The storage provider might impose a minimum order quantity on its clients such that any order has to exceed a certain size (either in terms of storage units or monetary value).

This concludes the description of the storage system-related quality requirements; the next paragraphs name the operations-related requirements.

- Serviced operations ensure that the storage system is available and running within the specified parameters; serviced operations are offered during pre-defined service hours. Non-serviced operations characterize the time in which a service operator is on stand-by, but not actually available online.
- The provisioning time is defined as the required time for setting up and making available the storage that was ordered by a customer.
- The availability of IT support is another quality criterion. The IT support function is usually executed by the above mentioned service unit and is usually triggered by an incident. According to the ITIL standard (Office of Government Commerce 2005), an incident is “an event, which is not part of the standard service operations and which causes or threatens to cause an interruption of the service or a degradation of its quality”. The support function acts on these incident reports and manages the incident resolution process. One of the key metrics of this incident management process is the reaction time, which is the time between the incident ticket submission and the reception message to

the submitter of the incident ticket. The reaction time can be a contractually defined as part of the SLA. Not only is the service unit obliged to react in a certain time interval, it may also be required to solve the root cause of the incident within a similar time constraint. This solution time is another performance metric of the service unit.

- When entering a service contract, the customer may have to fulfill certain collaboration duties. For an IT service, exemplary duties could include naming a contact person for the service provider, adhering to certain security policies, submitting a quarterly capacity planning, negotiating and enforcing maintenance windows in the organization, etc. These duties limit the flexibility of the service customer, so that the service provider can use this reduction in variability for more efficient service operations. From a service customer's point of view, the reduction of these duties is a quality criterion.
- In case of a storage service, data security is a major issue. One important activity in this area is the irreversible deletion of the contents of old storage equipment on decommissioning (disk wipe), such that the data cannot be restored. The adherence to this policy and to the relevant standards for this activity is another quality criterion.
- SLAs for storage services can contain penalties for the breach of quality-of-service clauses. The existence and the conditions of these penalties also define the quality of a service.

6.3.2 Quality Model Mapping

In Table 6.1 and Table 6.2, the above mentioned quality criteria are mapped to the generic IaaS quality model developed in section 5.3. The mapping of the quality criteria to the quality dimensions of the IaaS quality model is exemplary for this use case; some quality criteria were assigned to two different quality dimensions, as they exhibit traits of both dimensions. The mapping shows an even distribution of quality criteria, as every quality dimension has a corresponding quality criterion. Therefore, a broad coverage of quality aspects can be assumed.

6.3.3 Comparison of an In-house Storage Service to an IaaS Offer

As the quality criteria have been defined, they can be utilized for a comparison between different storage service providers. In this case study, one of BMW's file-based storage services and Amazon's EC2 EBS are compared and contrasted with each other. Table 6.3 contains the quality criteria for which both storage services show a comparable evaluation. Table 6.4 lists the quality criteria for which both services differ considerably. This comparison shown here is mostly qualitative. Although the exact metrics of the BMW service are known to the author, these cannot be given here due to confidentiality reasons.

The analysis of the results gives a mixed picture: neither option offers any performance guarantees (although BMW claims to offer a better performance), and most of the technical criteria are comparable across both options. The biggest differences become visible when contrasting the properties associated with the flexibility of the service usage. Provisioning time, cancellation period and the minimum order quantity are considerably different. AWS only places minor duties on the service customer (e.g. adherence to intellectual property rights), whereas the BMW service customer is deeply integrated in the IT processes of the enterprise. Moreover, guaranteed solution times with penalties for quality-of-service violations offer the business a greater level of assurance than the best-effort policies of AWS.

As a conclusion, the two storage services look similar at a first, superficial glance; after all, file storage is widely considered a commodity. A closer look reveals, that the two services cannot realistically be compared in terms of quality. Even though the aforementioned BMW file storage service offers very limited quality-of-service levels compared to the other available file storage services in the BMW Group, it is

Table 6.1: Mapping of BMW storage quality requirements on system-related quality dimensions

Generic system-related quality dimension	Case study quality requirement
Operational Time	Storage infrastructure operations
Performance	Provisioning time Storage performance Performance guarantees
Availability	Monitoring Reporting
Stability	Disaster protection Single point of failure (SPOF)
Recoverability	File system snapshots Backup (Tape)
Security	Rights and access administration
Standardization	Software licenses and maintenance
Product feature	Virus scan Software licenses and maintenance Subscription time Cancellation period Minimum order quantity

characterized by a sizable number of criteria (and so is the AWS storage service). Hence, the notion of file storage being a commodity is not justified. Moreover, BMW's storage service is tailored towards specific business needs, whereas Amazon's storage service does not have such a specific use case. These two reasons explain the differences in both file storage services, and it can be argued that these reasons are not limited to this case study, but are applicable in other IaaS usage settings as well.

Only if the business needs of a potential IaaS client are generic enough to fit both the IaaS service quality profile and the company-specific profile, the client will realistically consider IaaS an outsourcing option and thus will have to decide, what applications to outsource. If only one of the storage services offers the required quality attributes, this hybrid Cloud approach is unlikely to be pursued. A similar line of reasoning can be applied to other IT resources as well (e.g. compute services).

All cost optimization model evaluations in the following sections are based on the assumption, that the client's quality requirements are fulfilled by both the in-house IT provider and the outside IaaS provider and that the quality requirements are on a basic level (generic infrastructure) (as mentioned in the model assumptions in section 5.2). The evaluation of possible quality-cost trade-offs in the outsourcing process are beyond the scope of this work.

6.4 Data Collection

The evaluation of the IaaS decision support model presented in chapter 5 is executed in the context of the BMW SAP case study (see section 3.2), as SAP ERP systems are typical representatives of business software applications, especially for larger enterprises. As the decision support model requires a broad range of input data, a comprehensive data collection approach had to take place. The description of the data source used for the evaluation are displayed in Table 6.5, Table 6.6 and Table 6.7. The totally available

Table 6.2: Mapping of BMW storage quality requirements on operations-related quality dimensions

Generic operations-related quality dimension	Case study quality requirement
Service time	Serviced operations Non-serviced operations
Operations Performance	Reaction time
Reliability	Solution time
Empathy	Stand-by Collaboration duties of the service client Support
Assurance	Support Penalties Disk wipe on decommissioning

number of SAP application instances was determined as the intersecting set of SAP instances commonly available in each data set; eventually, each available SAP application instance had the full range of data (system monitoring, storage, IT landscape master data). If only one of the data items was missing for a certain SAP instance, this instance was not used for further calculations. This procedure resulted in 50 SAP application instances being available for the evaluation. The required data cleansing activities are also listed in each table; these cleansing operations further limited the available data.

- Data set 1 in Table 6.5 was recorded by the server monitoring software in the BMW data center and consists of samples taken as 5-minute averages for CPU, RAM and LAN utilization of each server.
- Data set 2 in Table 6.6 consists of samples of 1-day average application storage requirements; the numbers represent the actual storage usage, not storage reservations. They were extracted from the output of a proprietary SAP application monitoring service (Early Watch Alert), that continually logs vital system statistics. Please note that the gathering period for this data set is different from data set 1; therefore, the sample size for storage data is larger than the one for system monitoring data, but this is not a problem, as there is no direct temporal matching of the samples in the two different data sets. This temporal matching is not required as the evaluation approach only uses statistical properties of the data sets like the mean or specific quantiles. In essence, only the empirical distribution is relevant, not the time series underlying the data set.
- Data set 3 in Table 6.7 was extracted from the IT application landscape administration software SAP Solution Manager.² The set is a snapshot of a part of the SAP application landscape at BMW and contains system master data like sizing information, deployed IT software infrastructure and system status (e.g. production system, development system, integration system).

Data set 1 and 2 contained missing values to a varying degree. These missing values can be explained with the fact that nor the system monitoring software neither the target SAP instance were always available. As the subsequent calculations only rely on the empirical distribution of the data, there is no need to patch the gaps in the time series, as the time series is never analyzed and therefore does not have to be continuous. It is assumed that the system outages happened randomly, such that the empirical distribution is not systematically affected.

²<http://www.sap.com/germany/plattform/netweaver/components/solutionmanager/index.epx>, last accessed 2013-02-09

Table 6.3: Comparison of storage services

Quality criterion	AWS EC2 EBS	BMW File storage	Comparison
Virus scan	no	no	comparable
Storage infrastructure operations	yes	yes	comparable
Software Licenses and Maintenance	yes (AWS-proprietary software)	yes (BMW-proprietary software)	comparable
Support	yes, AWS Premium Support ¹	yes	comparable
File system Snapshots	no	no	comparable
Backup (Tape)	no	no	comparable
Rights and Access Administration	yes, AWS Identity and Access Management ²	yes	comparable
Monitoring	yes (AWS-proprietary)	yes (BMW-proprietary)	comparable
Reporting	yes	yes	comparable
Disaster protection	no	no	comparable
Single Point Of Failure (SPOF)	no (data replication in one data center)	no (data replication in one data center)	comparable
Subscription time	none	none	comparable
Performance guarantees	no	no	comparable
Disk Wipe on decommissioning	yes	yes	comparable

¹ <http://aws.amazon.com/premiumsupport/pricing/>, last accessed 2013-12-29

² <http://aws.amazon.com/iam/>, last accessed 2013-12-29

Another important type of input data consists of tariff informations. The Amazon AWS tariff informations were copied from the Amazon AWS Web site³ in the beginning of Aug. 2012. The actual cost figures and the server instance types used in the evaluation can be found in Appendix F. It is assumed that an Amazon EC2 Compute Unit roughly corresponds to a virtualized core of an Intel x86 multicore processor used in the BMW setting. The evaluation time interval t_{Δ} is one year, as this is the minimal subscription time of an AWS reserved instance. In order to guarantee comparability with BMW hardware, the AWS servers are specially configured:

- The server instances run SLES (SUSE Linux Enterprise Server) as an operating system, as those come equipped with commercial software licenses and license support (which matches BMW instances).
- The utilization of the virtual servers is assumed to be 10% on average (medium utilization according to AWS terms). This also matches observations in the BMW data center.
- Backup, virus scans and file system snapshots were excluded.

BMW infrastructure tariffs are based on the 2012 figures (BMW Group 2012a); the exact numbers are confidential, but they also follow the established cost types (compute, storage, network). Table 6.8 lists the BMW virtual server instance types utilized in the evaluation. The types marked with an asterisk cannot regularly be ordered, but are assumed to be technically feasible. The Euro-Dollar exchange rate is set to 1.27\$ per 1.00€. The evaluation exclusively focuses on the infrastructure cost factors, but of course there will be a number of other potential cost factors if an external IaaS provider is an outsourcing partner. These factors shall be discussed in the following paragraphs.

As this study perceives software applications to be monolithic and stand-alone (section 5.2 for the outsourcing scenario assumptions), no costs are incurred for application interfaces between the IaaS Cloud

³<http://aws.amazon.com/de/suse/>, last accessed 2013-12-29

Table 6.4: Contrast of storage services

Quality criterion	AWS EC2 EBS	BMW File storage	Comparison
Provisioning time	< 1h (≤ 1 TB)	Days - Weeks	different
Cancellation period	none	One Month by the end of the month	different
Serviced operations	24x7x365	limited time interval	different
Non-serviced operations	operations always serviced	no standby	different
Reaction time	1h (Level: Business)	longer	different
Solution time	no guarantee	guarantee	different
Storage Performance	ca. 100 IOPS on average (regular instance)	more	different
Minimum order quantity	1 GB	more	different
Collaboration duties of the service client	minor duties	intensive collaboration	different
Penalties	no	yes	different

and the client's data center. The expenses for server operations are excluded for both parties, but are assumed to be comparable in both scenarios (Cloud, In-house), as the same virtualized server should lead to the same operations expenses, no matter where the system is run. Currently, only data center operations are included in the compute cost but these are factored automatically in the tariff by the IaaS provider. No one-time transaction cost is incurred when choosing an external resource provider. In this instance, the effort required by the vendor selection process is used as a proxy for the transaction cost. In the industry, an RFP (request-for-proposal) is launched whenever there is outsourcing work to be done and whenever there are several potential vendors. The incurred expenses are extremely hard to estimate and vary greatly depending on the complexity of the outsourcing deal. The enterprise requires personnel to control the outsourcing provider. This periodic transaction cost level is similarly hard to estimate as the one-time transaction cost level. It is neglected as most BMW IT services are already externally procured, so the effort of controlling an IaaS provider should be comparable to the effort of controlling a substitute conventional infrastructure provider.

Table 6.5: Data Set 1 - Server Monitoring

Attributes	Values	Modifications
Source	Linux System Monitoring	
Dimensions	CPU/LAN/Memory	CPU utilization was calculated as CPU user and system load (no wait I/O); CPU count was calculated as the product of CPU workload and the number of CPUs in the original server
SAP instance count	65	50 instances remained after data cleansing
Gathering period	17.03.2009 - 16.05.2009	
Samples per instance per dimension	ca. 8520	Removed outliers and instances with different data recording format
Temporal granularity	5min	

Table 6.6: Data Set 2 - Application Monitoring

Attributes	Values	Modifications
Source	SAP Early Watch Alert	
Dimensions	Storage size in GB	
SAP instance count	516	Only instances were used with more than 30 data points which can be mapped to the instances from data set 1
Gathering period	24.01.2008 - 02.05.2012	
Samples per instance	37 - 1019	many missing values (1560 days across all instances)
Temporal granularity	1d	

Table 6.7: Data Set 3 - IT application landscape

Attributes	Values	Modifications
Source	SAP Solution Manager	
Dimensions	SAP instance master data (sizing, system status)	
SAP instance count	73	Only instances which can be mapped to the instances from data set 1
Gathering period	as of April 2009	
Samples per instance per dimension	1	
Temporal granularity	none (snapshot)	

6.4.1 Descriptive Workload Characteristics

The following paragraphs give some examples of the statistical properties of the empirical workload distributions used for this research. The focus here lies on descriptive statistics, such that a rough understanding of the fundamental characteristics is established. The tables and the graphics were created using the R statistical software package (R Core Team 2012).

Table 6.9 shows the mean, the median, the standard deviation, the standard error and the inter-quartile range for the five resource types used in the evaluation. (The dimensions ‘‘SAP application’’ and time are not considered here.) The metrics reveal that the workload data is long-tailed and positively skewed. It is obvious that a normal distribution is unlikely to be a good fit for this kind of data.

The correlation coefficients of the four server-related resource dimensions are detailed in Table 6.10. As explained above, the application storage data was collected on another time scale, hence it cannot be incorporated into this matrix. In general, the correlation coefficients are rather low, so the linear relationships among the variable pairs are very weak.

Figure 6.3 displays a scatterplot matrix of the aforementioned four variables. The panels show no clear functional relationship or clustering among the variables. The large skewness becomes visible again.

As the workload data exhibits a great deal of randomness, a parametric approach to modeling the workload data has been abandoned. The utilization of workload data in the following evaluation is purely non-parametric and completely based on the joint empirical distribution. This seems justified also because of the large sample size of the data set.

Table 6.8: BMW instance types

BMW Instance type	Virtual Cores	RAM (GB)
B3 virt. Server 1x1	1	1
B3 virt. Server 1x2	1	2
B3 virt. Server 2x2	2	2
B3 virt. Server 2x4	2	4
B3 virt. Server 4x4	4	4
B3 virt. Server 4x8	4	8
B3 virt. Server 6x6	6	6
B3 virt. Server 6x12	6	12
B3 virt. Server 8x16	8	16
B3 virt. Server 8x32	8	32
B3 virt. Server 16x64*	16	64
B3 virt. Server 20x136*	20	136

Table 6.9: Descriptive workload statistics

	LAN_in (MB/h)	LAN_out (MB/h)	CPU Count	RAM (MB)	Storage (GB)
Mean	2.65	1.78	0.70	7820.28	908.90
Median	0.42	0.18	0.19	4559.00	828.19
Std. Dev.	8.76	10.82	1.35	11305.45	583.13
Std. Err.	0.01	0.02	0.00	17.51	2.16
IQR	2.39	0.74	0.79	9996.70	663.82
Min	0.00	0.00	0.00	4.72	13.00
Max	228.24	424.95	19.68	95982.66	3771.68

6.5 Experimental Results

The evaluation results allow the assessment of the effects of parameter variations on the IaaS placement decisions and on the total cost. First, the effect of the α base resource level on the IaaS instance cost for the elastic IaaS tariff is analyzed, then the financial importance of the different resource types is highlighted and last, the variability of different IT resources is investigated. All these observations are descriptive and can be evaluated independently of the application outsourcing decision (which will act as the dependent variable later on).

Finally, the outsourcing decision is analyzed: to this end, the complete parameter set is used as independent variables in a machine learning algorithm in order to identify direct and higher-order interaction effects on the placement decision.

6.5.1 Effect of α on AWS Instance Cost

The effect of the base instance α -level on the AWS instance cost under the elastic tariff can be seen in Figure 6.4. This figure is a box-whisker-plot; the small circles above the boxes represent outliers in the data. All elements of this plot refer to the y-axis on the left-hand side of the figure. The locations of the median in the boxes reveals that the distributions per α -level are positively skewed and show some outliers

Table 6.10: Workload correlation coefficients

	CPU Count	RAM (MB)	LAN_in (MB/h)	LAN_out (MB/h)
CPU Count	1.00			
RAM (MB)	-0.02	1.00		
LAN_in (MB/h)	0.17	-0.04	1.00	
LAN_out (MB/h)	0.17	0.10	0.21	1.00

in the long tail of the distribution. The other scenario parameters L_{norm} , d , β , a_{Δ} and p_{Δ} do not influence the AWS instance cost level, hence their setting is irrelevant for this purpose.

For $\alpha = 0.3$, the median of instance costs reaches a minimum, i.e. this level is the most efficient level at which reserved instances should be rented, if the elastic tariff described in section 5.5.2 was available. The median was selected here as a characteristic measure of the cost distribution, because it is not sensitive to outliers (unlike the mean). The following table shows the median cost per α -level.

α	Median VM Instance Cost p.a. (\$)
0.0	4356.76
0.1	4060.20
0.2	4060.20
0.3	4003.82
0.4	4098.16
0.5	4098.16
0.6	4216.96
0.7	4638.09
0.8	5255.24
0.9	5534.86
1.0	8256.00

In order to prove the significance of the effect of α on the cost distribution, a Friedman test is performed (Hollander and Wolfe 1973, pp. 139-146). This non-parametric test is used for within-subject experimental designs and does not depend on the normal distribution of the input data (unlike an ANOVA). The AWS instance cost under the elastic tariff is the dependent variable, the α -level is the independent variable and the applications act as blocks. The value of Friedman's chi-squared statistic is 177.60 with 10 degrees of freedom. The p-value is smaller than $2.2e-16$, hence the null hypothesis of equal instance costs per level of α can be rejected. As a conclusion, the effect of α on the instance cost level in an elastic IaaS tariff cannot be neglected. As the most efficient level of α has been established, it is possible to compare the reserved tariff with the elastic tariff at that level and determine, if the elastic tariff offers any added value.

The evaluation results include the total IaaS cost per application for each of the two tariffs (elastic and reserved). To compare the total cost per application, a paired Wilcoxon signed rank test (Rinne 2008, p. 551) with continuity correction is used, which is available in the R statistics package (Crawley 2007, p. 297). This non-parametric test does not depend on the two samples being normally distributed, which is the case here. The test has to be paired, as the total cost of the same application under different tariffs is compared. The test results allow the rejection of the null hypothesis of equal means for $\alpha \in \{0.1, 0.6, 0.9\}$ with extremely low p-values ($<1.942e-05$). Hence, it can be concluded that the elastic tariff guarantees significantly lower total costs than the reserved tariff under the same conditions.

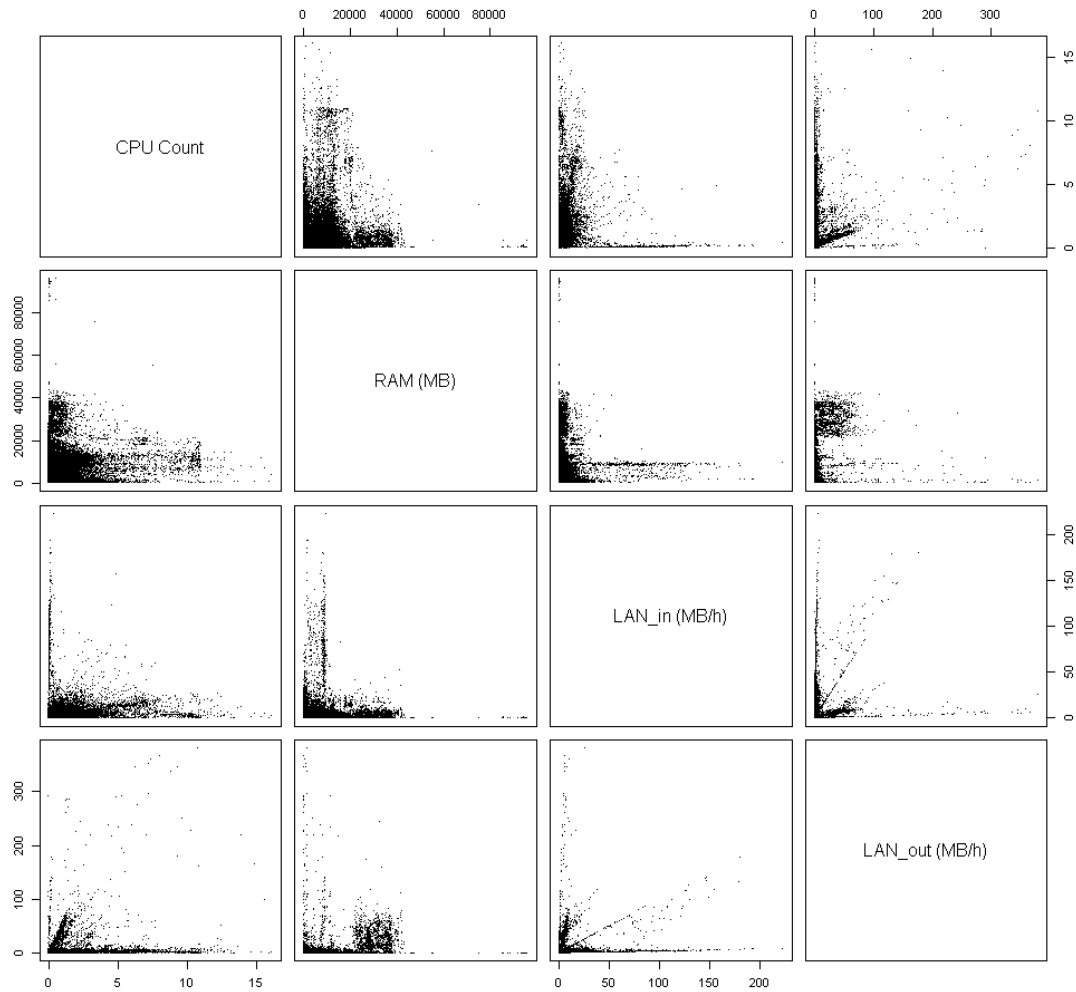


Figure 6.3: Workload scatterplot matrix

α	Median Elastic Total Cost p.a. (\$)	Median Reserved Total Cost p.a. (\$)
0.1	5407.02	10165.43
0.6	5738.85	10165.43
0.9	7056.20	10165.43

6.5.2 Financial Importance of Different Resource Types

The financial importance of the different resource types need to be highlighted, as the cost share of each resource type determines the relevance of the resource type. Table 6.11 displays the average cost shares per resource type and tariff. Each row adds up to 100%; the percentages are based on the total cost sum per tariff. These numbers are aggregated across all applications. The financial figures are standardized per row; the underlying absolute financial values differ widely among the different tariffs and cannot be derived from this table. It becomes clear that instance-related costs claim between ca. 77% and 84% of the total cost (depending on the α level). Storage-related costs represent the second largest share; networking costs are negligible. These ratios suggest that storage costs are a significant factor, that has to be accounted for when optimizing IaaS Cloud usage. Therefore, it is surprising if recent research papers on IaaS cost optimization leave out this cost type and solely analyze VM instance cost (e.g. Lilienthal (2013)).

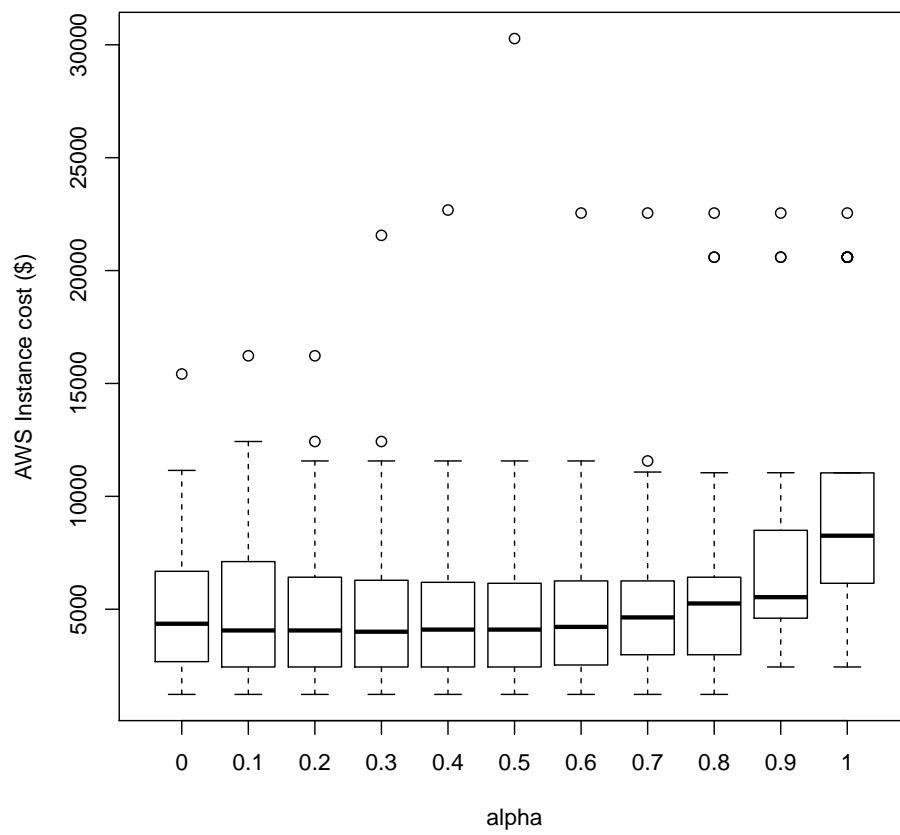
Figure 6.4: Effect of α on AWS instance cost

Table 6.11: Comparison of cost type percentages

Tariff	α	Instance cost (%)	LAN out (%)	LAN in (%)	Storage (%)
IaaS (elastic)	0.1	78.5%	0.6%	0.5%	20.3%
	0.6	77.7%	0.7%	0.6%	21.1%
	0.9	83.7%	0.5%	0.4%	15.4%
IaaS (reserved)	n.a.	87.8%	0.4%	0.3%	11.5%
In-house	n.a.	73.6%	0.0%	0.0%	26.4%

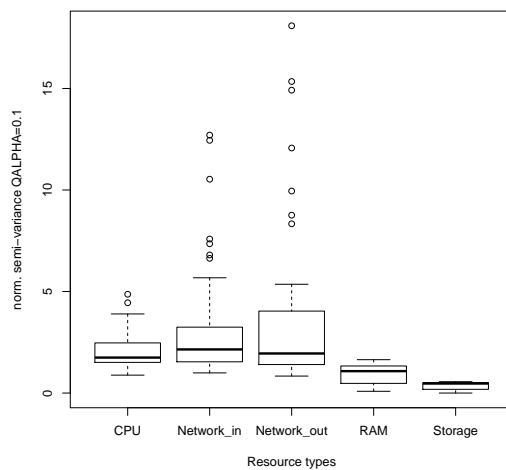


Figure 6.5: Normalized semi-variance per resource type

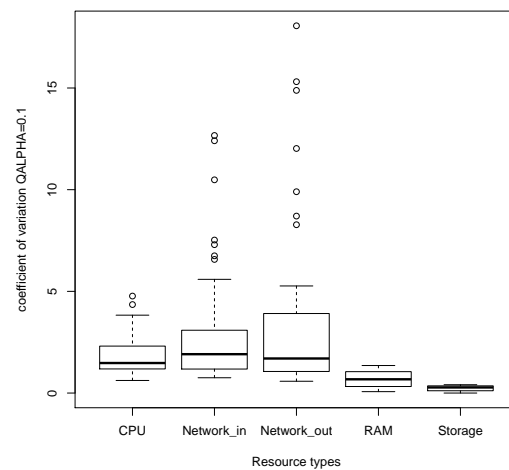


Figure 6.6: Coefficient of variation per resource type

6.5.3 Variability of Different Resource Types

Before the placement decision will be analyzed, the variability of different IT resources is investigated. Figure 6.5 shows a box-and-whisker plot of the different semi-variances per resource type evaluated in this work. The $\alpha = 0.1$ base resource level is displayed as it minimizes the amount of reserved infrastructure and hence should allow potential fluctuations in the workload to fully manifest themselves. The plot aggregates the semi-variances across the applications; the variations in semi-variance in each resource factor hence are attributable to the variability in the application workloads. As the application semi-variance was normalized with the application mean (equation 5.15), the factor distributions are comparable. It becomes clear that storage demands fluctuate least and are almost stagnant, whereas networking resources exhibit massive variations in demand. Instance resources (CPU, RAM) are somewhere in between. The picture looks practically identical for the coefficient of variation in Figure 6.6; the resource types exhibit the same order of variability as is the case for semi-variance (e.g. the networking resources are still ranked in the first positions). When these findings are applied to the IaaS placement decision, they seem to suggest that semi-variance may not have a strong impact on the placement decision, as the three most cost-intensive resource types are the ones that exhibit the least fluctuation.

6.5.4 Decision Tree Results

The resulting decision tree with its splits and leaf nodes is visible in Figure 6.7. Figure 6.7 contains line numbers at the beginning of each line, e.g. L05 for line number 5 in the tree listing. After the colon at each leaf node, the predicted placement and the case count for each placement option is printed (e.g. in

L05 “iaas” and in-house=6, iaas=410). In this leaf, 6 cases of in-house placement as calculated by the optimization algorithm would be misclassified by the decision tree, as all the cases in this leaf would be classified as “iaas” cases. A graphical representation of the tree is shown in Figure 6.8. The rectangular boxes signify the tree leaves and contain the target variable value distribution in a separate bar; its binary color coding uses red for the share of IaaS cases and blue for the share of in-house cases. The box labels in the tree leaves are chosen according to the value of the target variable TARGET predicted by the algorithm for this subset of cases. The figure also shows the share of misclassified cases per leaf (whose class does not match the predicted class).

But in fact, there are good reasons to assume that not all training cases can be considered equal. Each training case represents the placement of a certain software application in a specific outsourcing scenario. This placement decision is associated with a total cost sum which is different with each training case. This associated cost sum represents a good candidate for a weighting factor. If the overall goal is cost optimization, then more expensive training cases should influence the tree learning algorithm more than less expensive cases, because the financial consequences of mispredictions of expensive cases are more severe than of cheap cases. (The financial impact of mispredictions in an equal-weighting scheme was already derived in Table 6.16 and its implications were discussed.) The question arises whether a cost-based weighting scheme might reduce this financial impact of mispredictions. To address this question, a weighting scheme is applied to the machine learning approach described in the last section and the effects on the structure of the decision tree and on the cost deviations are derived and compared to the results of the equal-weighting approach.

For the decision tree in Figure 6.7, every training case has the same weight, i.e. it contributes equally to the impurity metric on which the tree is based. As a next step, the decision tree learning algorithm uses the case-specific weighting factor CASEWEIGHT during its learning phase. The resulting tree nodes and branching criteria are shown in Figure 6.9; a graphic representation of the decision tree can be found in Figure 6.10. The tree structure also allows to identify the most important variables in an outsourcing decision, as the learning algorithm behaves greedily (Quinlan 1993, p. 20): when it starts assembling the tree, it uses a statistical criterion (in this case the Gini criterion) on each variable to determine the one that produces the best split of the test cases. So the root node is by definition the node that best separates the cases into in-house and IaaS cases. For the tree in Figure 6.9, DYNRATIO, the ratio between the total cost for the elastic IaaS tariff and the total cost for the in-house tariff, is the single most important variable (L01, L24), which is not surprising for a cost-based optimization model with cost-based case weights. However, the second layer in the tree is deserves more attention, as the value of DYNRATIO determines the importance of the variables in this layer. A value of DYNRATIO greater than 0.917 means that the outsourcing (OSDEGR) gains a large importance in this layer (L02, L13). Its importance is obvious, as it limits the outsourcing activities, hence it can be expected to play a major role in the tree. A value of DYNRATIO smaller than 0.917 leads to a totally different situation: in this part of the tree, CPU_QO is the second most important variable (L25, L32). This variable among others is responsible for the size of the required VM instance, which is in turn a determining cost factor. This relationship comes as no surprise, but the importance of CPU_QO is nonetheless interesting. Especially noteworthy is the fact, that CPU_QO only appears in the bottom part of the tree (L25 and larger), where the cost ratio for IaaS resources is more favorable (smaller DYNRATIO) than in the upper part of the tree (larger DYNRATIO).

The performance of the unweighted decision tree according to the cross-validation procedure is displayed in Table 6.12, a confusion matrix (Kohavi and Provost 1998, p. 272). The overall average accuracy was 86.97% with a range of 0.32%. The overall average precision was 87.22% with a range of 0.66% (positive class: iaas). The overall average recall was 83.25% with a range of 1.48% (positive class: iaas).

L01: OSDEGR > 0.450
 L02: CPU_QO > 8.235
 L03: DYNRATIO > 1.096
 L04: CPU_QO > 13.007
 L05: OINSTDYNRATIO > 2.544: iaas (Cases: in-house=6, iaas=410)
 L06: OINSTDYNRATIO ≤ 2.544: in-house (Cases: in-house=4279, iaas=301)
 L07: CPU_QO ≤ 13.007
 L08: AGGCOEFVAR > 2.149: iaas (Cases: in-house=82, iaas=1926)
 L09: AGGCOEFVAR ≤ 2.149: in-house (Cases: in-house=2779, iaas=2545)
 L10: DYNRATIO ≤ 1.096
 L11: VARIAB > 19.108
 L12: DYNRATIO > 0.890: in-house (Cases: in-house=430, iaas=62)
 L13: DYNRATIO ≤ 0.890: iaas (Cases: in-house=63, iaas=617)
 L14: VARIAB ≤ 19.108: iaas (Cases: in-house=689, iaas=46795)
 L15: CPU_QO ≤ 8.235
 L16: STODYNRATIO > 0.266
 L17: NWIRESRATIO > 0.005
 L18: OSDEGR > 0.650: iaas (Cases: in-house=3655, iaas=6509)
 L19: OSDEGR ≤ 0.650: in-house (Cases: in-house=8000, iaas=2164)
 L20: NWIRESRATIO ≤ 0.005
 L21: STO_EXP > 1558.609: in-house (Cases: in-house=2749, iaas=2179)
 L22: STO_EXP ≤ 1558.609: iaas (Cases: in-house=4380, iaas=29808)
 L23: STODYNRATIO ≤ 0.266
 L24: RAM_SV > 0.009: in-house (Cases: in-house=28091, iaas=9793)
 L25: RAM_SV ≤ 0.009
 L26: NWL_CV > 1.005: iaas (Cases: in-house=3271, iaas=6893)
 L27: NWL_CV ≤ 1.005: in-house (Cases: in-house=816, iaas=108)
 L28: OSDEGR ≤ 0.450
 L29: CPU_QO > 8.235
 L30: DYNRATIO > 0.925
 L31: OINSTRESRATIO > 0.872
 L32: OINSTDYNRATIO > 2.544: iaas (Cases: in-house=354, iaas=614)
 L33: OINSTDYNRATIO ≤ 2.544: in-house (Cases: in-house=13098, iaas=1014)
 L34: OINSTRESRATIO ≤ 0.872
 L35: OSDEGR > 0.150: iaas (Cases: in-house=693, iaas=1611)
 L36: OSDEGR ≤ 0.150: in-house (Cases: in-house=679, iaas=89)
 L37: DYNRATIO ≤ 0.925
 L38: OSDEGR > 0.250: iaas (Cases: in-house=1748, iaas=19668)
 L39: OSDEGR ≤ 0.250
 L40: STO_CV > 0.320: in-house (Cases: in-house=8345, iaas=2049)
 L41: STO_CV ≤ 0.320: iaas (Cases: in-house=2482, iaas=8540)
 L42: CPU_QO ≤ 8.235
 L43: STODYNRATIO > 0.382
 L44: OSDEGR > 0.350
 L45: OINSTDYNRATIO > 1.780: iaas (Cases: in-house=664, iaas=1954)
 L46: OINSTDYNRATIO ≤ 1.780: in-house (Cases: in-house=756, iaas=245)
 L47: OSDEGR ≤ 0.350: in-house (Cases: in-house=9103, iaas=1754)
 L48: STODYNRATIO ≤ 0.382
 L49: STORATIO > 1.323
 L50: OINSTDYNRATIO > 1.897: iaas (Cases: in-house=502, iaas=850)
 L51: OINSTDYNRATIO ≤ 1.897: in-house (Cases: in-house=4961, iaas=783)
 L52: STORATIO ≤ 1.323: in-house (Cases: in-house=83665, iaas=3179)

Figure 6.7: Tabular Decision Tree - Unweighted Cases

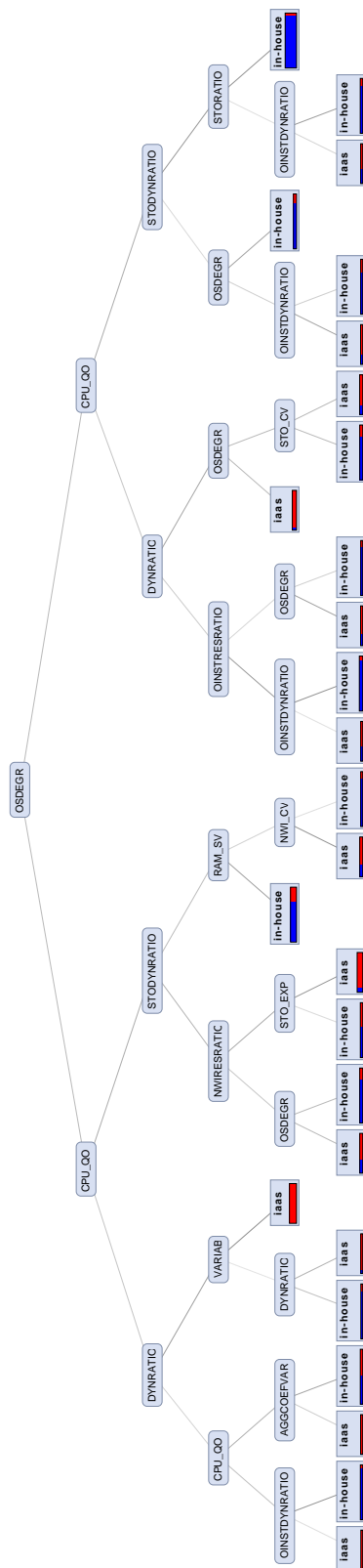


Figure 6.8: Visualization of the Decision Tree - Unweighted Cases

L01: DYNRATIO > 0.917
 L02: OSDEGR > 0.550
 L03: RAM_QO > 35.650 GB
 L04: RESRATIO > 0.939: in-house (Cases: in-house=5779, iaas=53)
 L05: RESRATIO ≤ 0.939
 L06: DYNINSTRATIO > 0.433: in-house (Cases: in-house=18, iaas=3)
 L07: DYNINSTRATIO ≤ 0.433: iaas (Cases: in-house=17, iaas=34)
 L08: RAM_QO ≤ 35.650 GB
 L09: OINSTDYNRATIO > 1.319: iaas (Cases: in-house=4087, iaas=15599)
 L10: OINSTDYNRATIO ≤ 1.319
 L11: AGGCOEFVAR > 3.681: in-house (Cases: in-house=9821, iaas=4315)
 L12: AGGCOEFVAR ≤ 3.681: iaas (Cases: in-house=5625, iaas=8625)
 L13: OSDEGR ≤ 0.550
 L14: OINSTDYNRATIO > 2.522
 L15: OSDEGR > 0.250: iaas (Cases: in-house=234, iaas=1446)
 L16: OSDEGR ≤ 0.250
 L17: OINSTDYNRATIO > 3.477: iaas (Cases: in-house=108, iaas=296)
 L18: OINSTDYNRATIO ≤ 3.477: in-house (Cases: in-house=605, iaas=111)
 L19: OINSTDYNRATIO ≤ 2.522
 L20: STODYNRATIO > 0.382
 L21: OSDEGR > 0.250: iaas (Cases: in-house=448, iaas=1763)
 L22: OSDEGR ≤ 0.250: in-house (Cases: in-house=1217, iaas=257)
 L23: STODYNRATIO ≤ 0.382: in-house (Cases: in-house=73636, iaas=9839)
 L24: DYNRATIO ≤ 0.917
 L25: CPU_QO > 8.235
 L26: OSDEGR > 0.250: iaas (Cases: in-house=1933, iaas=62027)
 L27: OSDEGR ≤ 0.250
 L28: STO_CV > 0.320: in-house (Cases: in-house=8300, iaas=2044)
 L29: STO_CV ≤ 0.320
 L30: STO_CV > 0.172: iaas (Cases: in-house=872, iaas=7620)
 L31: STO_CV ≤ 0.172: in-house (Cases: in-house=1595, iaas=889)
 L32: CPU_QO ≤ 8.235
 L33: OSDEGR > 0.450
 L34: STO_EXP > 1095.781: iaas (Cases: in-house=4123, iaas=23389)
 L35: STO_EXP ≤ 1095.781
 L36: OINSTDYNRATIO > 2.136: iaas (Cases: in-house=1753, iaas=5603)
 L37: OINSTDYNRATIO ≤ 2.136: in-house (Cases: in-house=15437, iaas=4487)
 L38: OSDEGR ≤ 0.450
 L39: STODYNRATIO > 0.382
 L40: OSDEGR > 0.350: iaas (Cases: in-house=1299, iaas=1601)
 L41: OSDEGR ≤ 0.350: in-house (Cases: in-house=7633, iaas=1067)
 L42: STODYNRATIO ≤ 0.382: in-house (Cases: in-house=41800, iaas=1392)

Figure 6.9: Tabular Decision tree - Weighted Cases

Table 6.12: Decision Tree Performance - Unweighted Cases

	true in-house (cases)	true iaas (cases)	class precision
pred. in-house (cases)	167717	25530	86.79%
pred. iaas (cases)	18623	126930	87.21%
class recall	90.01%	83.25%	

When a decision tree is supposed to be applied as a decision support tool, then a simpler tree is a better tree, as it offers a higher level of generalization. However, tree simplicity has to be balanced with predictive performance, hence it needs to be clarified whether the simpler weighted tree performs as well as the unweighted tree. Tab 6.13 shows the confusion matrix of the weighted tree, calculated using 10-fold cross-validation (analogue to the one of the unweighted tree). The cells of the matrix contain weighted cases hence they feature Dollars (\$) as a unit. The overall performance metrics are as follows: accuracy is at 86.71% with a range of 0.15%, precision is at 86.43% with a range of 0.64% and recall is at 83.62% with a range of 0.55%. When these overall values are compared to the ones of the unweighted tree in Table 6.12, accuracy, precision and recall are on comparable levels (the differences between the unweighted and the weighted variants are smaller than 1%). The MCC for the weighted tree is 0.731 (unweighted tree 0.736). Both trees can be judged comparable in terms of predictive performance with slight advantages of the unweighted tree, but as the weighted tree is structurally simpler (lower line count), it lends itself more to decision support uses.

Table 6.13: Decision Tree Performance - Weighted Cases

	true in-house (cases)	true iaas (cases)	class precision
pred. in-house (cases)	166309	24979	86.94%
pred. iaas (cases)	20031	127481	86.42%
class recall	89.25%	83.62%	

As explained in section 5.7.3, a decision tree based on random placement decisions is used as a performance baseline for the decision tree based on cost-optimal placement decisions. The performance of the random decision tree according to the cross-validation procedure is displayed in Table 6.14. The overall average accuracy was 69.96% with a range of 0.21%. The overall average precision was 68.53% with a range of 1.56% and the overall average recall was 61.87% with a range of 4.88%. It must be noted that these numbers only represent one possible outcome of the random outsourcing process, so the external validity of these figures must be questioned. A bootstrapping approach would be required to arrive at more reliable validity estimates for the random tree. This task could be part of the future work.

Table 6.14: Decision Tree Performance for Random Placement - Base Settings

	true in-house (cases)	true iaas (cases)	class precision
pred. in-house (cases)	142700	58134	71.05%
pred. iaas (cases)	43640	94326	68.37%
class recall	76.58%	61.87%	

When Table 6.12 and Table 6.14 are compared, it can be deducted that the original decision tree has a significantly higher predictive power than the random decision tree. This result is not surprising in itself, however it helps to put the predictive power of the original decision tree in perspective. The random decision tree measures, how many placements can be explained sensibly, even if they were chosen randomly. Hence,

it exposes the added value of the optimization step in the outsourcing process, as it increases the sensibility of the placements.

The decision tree performance can not only be compared to random placement decisions, but also to completely biased decisions, favoring exclusively IaaS or in-house placements. Table 6.15 shows the confusion matrix for an IaaS-only software application deployment (the positive class in terms of the confusion matrix definition is IaaS). Per definition, all true IaaS cases were correctly predicted, which explains the recall rate of 100%, but none of the in-house cases was predicted, which leads to an accuracy of 45% and to a precision of 45% (this is exactly the base ratio of IaaS cases to all cases). The class precision for predicted in-house cases could not be calculated, as zero predictions for in-house placements were made. In essence, a purely IaaS-oriented prediction algorithm would only achieve about half the accuracy of the original decision tree. This comparison demonstrates the improvement of predictive power gained through the decision tree.

Table 6.15: Predictive Performance of pure IaaS Deployment

	true in-house (cases)	true iaas (cases)	class precision
pred. in-house (cases)	0	0	n.a.
pred. iaas (cases)	186340	152460	45.00%
class recall	0.00%	100.00%	

The MCC value for the baseline decision tree in Figure 6.7 is 0.736; as a comparison, the MCC value for the random placement decision tree is 0.389. The MCC value for a pure IaaS deployment cannot be calculated, as this indicator is a fraction and its denominator becomes zero in this case. Hence, the baseline decision tree shows an overall satisfactory predictive performance, when all outsourcing scenarios and all software applications are considered collectively.

Table 6.16 shows the corresponding results for the unweighted tree. Column 1 contains the line number of the decision tree (as defined by Figure 6.7). Column 2 lists the additional costs generated by wrong predictions and column 3 lists the costs saved by wrong predictions. The “Cost Delta” column adds up these two values; this sum is then compared to the optimal cost sum, leading to the percentage deviation in the last column. The last two table rows show aggregated values (column sum, resp. the column average). The table nicely shows that although there are leaves with high percentage deviations ranging from 21.73% cost savings to 20.34% cost increases, these deviations almost cancel out in the end, leading to a cost saving of 0.39%, which seems a low price to pay for using predicted values.

Table 6.17 lists the cost deviation values in Dollars and percent for the weighted tree (similar to Table 6.16). When these deviation numbers are compared to the ones of the unweighted tree, two notable observations can be made. First, the percentage deviations per leaf exhibit a different range (from 29.69% savings to 15.32% increases) than the ones of the unweighted tree. Second, the total percentage deviation is lower in absolute terms than the one of the unweighted tree (-0.03% vs. -0.39%).

Across all outsourcing scenarios and across all software applications, an average outsourcing degree of 45% is predetermined by the experimental setup (selection of levels for η). This ratio is also strictly enforced during the optimization step. It would be interesting to know, how well the average outsourcing degree is reproduced by the predicted decisions of the two decision trees. The outsourcing degree could be considered as a management decision which needs to be complied with during outsourcing activities. Hence, the decision tree should aim to reproduce this ratio.

Table 6.18 shows the average outsourcing degree across all outsourcing scenarios. Both decision trees closely preserve the outsourcing degree (ca. 43%); this fits the observation, that OSDEGR (η) is one of the most decisive input variables in both trees.

Table 6.16: Cost deviations per leaf caused by prediction

Tree leaf (line no.)	Cost added by prediction errors (\$)	Cost saved by prediction errors (\$)	Cost Delta (\$)	Optimal Sum (\$)	Cost	Percentage Deviation
L05	16364.39	0.00	16364.39	3043842.45		0.54%
L06	15759.19	-771497.45	-755738.26	58286634.68		-1.30%
L08	276950.45	0.00	276950.45	18910243.79		1.46%
L09	0.00	-7453296.81	-7453296.81	49735790.46		-14.99%
L12	194179.23	0.00	194179.23	10538472.87		1.84%
L13	0.00	-230740.48	-230740.48	12128051.55		-1.90%
L14	226997.76	-332332.55	-105334.79	346011504.76		-0.03%
L18	534337.64	-5887317.59	-5352979.95	42212591.67		-12.68%
L19	694658.07	-1860693.18	-1166035.11	46500141.42		-2.51%
L21	7588424.36	-205241.81	7383182.55	36292500.10		20.34%
L22	2477722.03	-9194979.51	-6717257.49	171143293.30		-3.92%
L24	11412197.53	-9031512.28	2380685.25	223329159.30		1.07%
L26	13150782.38	-635527.48	12515254.90	72962381.61		17.15%
L27	4163.72	-637564.97	-633401.25	7565575.23		-8.37%
L32	464817.15	-11947.07	452870.08	6336702.63		7.15%
L33	237590.44	-869655.73	-632065.28	148956531.82		-0.42%
L34						
L35	1564922.48	-40396.84	1524525.64	19097463.74		7.98%
L36	10157.63	-56047.13	-45889.50	6660823.04		-0.69%
L38	0.00	-7443668.76	-7443668.76	161013507.27		-4.62%
L40	13804352.91	0.00	13804352.91	135790225.85		10.17%
L41	0.00	-14772564.51	-14772564.51	77851553.93		-18.98%
L45	9650.50	-2861857.08	-2852206.58	13126647.75		-21.73%
L46	154859.07	-136710.34	18148.74	5112983.63		0.35%
L47	2042999.78	-494131.30	1548868.48	66968161.13		2.31%
L50	553756.30	-23065.60	530690.70	3543838.30		14.98%
L51	0.00	-2084366.94	-2084366.94	18460029.32		-11.29%
L52	3666965.27	-3001085.00	665880.27	555717487.92		0.12%
Total Sum	59102608.29	-68036200.40	-8933592.11	2317296139.53		
Total Average						-0.39%

As a conclusion to the decision tree evaluation, it can be argued that the presented tree models possess a high predictive performance and generalize well (as demonstrated by the cross-validated evaluation metrics).

The following paragraphs contrast and compare the weighted and the unweighted decision tree and try to answer the question, what the differences and the similarities between the two trees are. The tree complexity is clearly one differentiating factor; the weighted tree is structurally simpler than the unweighted tree. There are fewer leaves and nodes overall in the weighted tree (42), as compared to the unweighted tree (52). It also features only 14 leaves on the lowest tree level as opposed to 22 leaves for the unweighted tree. What this means in terms of predictive performance is discussed above; what this means for applying the tree in

Table 6.17: Cost deviations per leaf caused by weighted prediction

Tree leaf (line no.)	Cost added by prediction errors (\$)	Cost saved by prediction errors (\$)	Cost Delta (\$)	Optimal Cost Sum (\$)	Percentage Deviation
L04	28184.35	-34715.28	-6530.93	73268055.89	-0.01%
L06	9475.34	0.00	9475.34	473494.00	2.00%
L07	25730.21	-13151.47	12578.74	1024189.46	1.23%
L09	5583743.98	-113138.74	5470605.24	105336220.06	5.19%
L11	435906.75	-8123396.07	-7687489.33	89221586.08	-8.62%
L12	15281476.70	-171965.28	15109511.42	113044415.19	13.37%
L15	195105.98	-9733.08	185372.89	7933219.94	2.34%
L17	60488.64	-6846.27	53642.36	1801024.14	2.98%
L18	10302.79	-84094.24	-73791.46	3154244.95	-2.34%
L21	218210.52	-28898.72	189311.80	10254095.67	1.85%
L22	11462.78	-198572.01	-187109.22	6220407.52	-3.01%
L23	3015248.88	-13537545.12	-10522296.24	554297380.02	-1.90%
L26	0.00	-7948605.56	-7948605.56	465965272.54	-1.71%
L28	13798535.71	0.00	13798535.71	135145635.87	10.21%
L30	0.00	-5296624.36	-5296624.36	53594718.14	-9.88%
L31	3665474.08	0.00	3665474.08	23927562.27	15.32%
L34	0.00	-13067674.46	-13067674.46	163174078.40	-8.01%
L36	0.00	-4875927.71	-4875927.71	22817341.43	-21.37%
L37	9156118.97	0.00	9156118.97	104034261.66	8.80%
L40	0.00	-4471897.18	-4471897.18	15059814.20	-29.69%
L41	2021866.30	0.00	2021866.30	58175366.79	3.48%
L42	3675259.14	0.00	3675259.14	309373755.30	1.19%
Total sum	57192591.11	-57982785.55	-790194.44	2317296139.53	
Total average					-0.03%

Table 6.18: Reproduction of the outsourcing degree

Placement method	IaaS placement (%)	In-house placement (%)	Sum
Optimization	45.00%	55.00%	100.0%
Prediction (unweighted)	42.73%	57.27%	100.0%
Prediction (weighted)	43.83%	56.17%	100.0%

a decision support setting is discussed in section 6.6.1. Another difference between the two trees are the importance of the input variables. The unweighted tree has OSDEGR as the root node, then CPU_QO as the two nodes on the first level. The weighted tree ranks DYNRATIO as the variable with supreme importance; on the first level, OSDEGR and CPU_QO follow. The weighting obviously reduces the importance of OSDEGR and increases the importance of DYNRATIO, the ratio between the elastic IaaS tariff's total cost per application and the in-house tariff's total cost per application.

Table 6.19 lists the input variables used in each tree type. Using a weighting factor produces a more compact tree; the weighted tree requires 11 different variables, whereas the unweighted tree needs 14 different variables. The application of a weighting factor seems to help the tree learning algorithm focus

on the more relevant variables in the input set and it seems to decrease redundancy among variables. For example, the weighted tree only uses AGGCOEFVAR as an overall measures of workload variability, but the unweighted tree requires both VARIAB and AGGCOEFVAR, the two of which feature highly similar values. The unused variables are similar in both tree types: network-related variables are hardly used as predictors. This observation matches the fact, that network-related costs are an unimportant overall cost factor.

6.6 Discussion of Experimental Results

In the last sections, the evaluation results of the analytical model are presented. This chapter summarizes and further discusses these findings in light of the current research. Also, the limitations of the research approach are detailed, followed by the contributions of the work for both researchers and practitioners.

6.6.1 Implications

After the quality of the decision tree has been established, it is fit for actually providing decision support in the outsourcing scenarios described above. Quinlan (1993, p. 47) points out the structural similarities between the tree structure and production rules. These production rules capture all the factors necessary to arrive at a placement decision and can be derived from the tree by starting at the root node and follow the tree up to the leaves. Each tree branch connecting two nodes is another condition for the rule. As an example, these rules will be derived for the weighted decision tree in Figure 6.9. In order to focus on the most relevant rules, only the top 5 rules for each placement target are elaborated (according to the data, it makes no difference whether the top 5 based on total cost per leaf or the top 5 based on total number of cases per leaf are chosen - they are identical). The following paragraphs describe these rules (i.e. the placement decisions and the conditions under which these decisions were chosen). The line numbers in these rules refer to Figure 6.9.

An in-house placement is the optimal choice for business applications in the investigated outsourcing scenarios under the following conditions:

- Line L28: $DYNRATIO \leq 0.917$ and $CPU_QO > 8.235$ and $OSDEGR \leq 0.25$ and $STO_CV > 0.32$. A low outsourcing degree is unsurprisingly favorable for in-house sourcing; an elevated storage variability is understandable as well, as in-house storage is purchased in total upfront, hence variability does not alter the cost.
- Line L11: $DYNRATIO > 0.917$ and $OSDEGR > 0.55$ and $RAM_QO \leq 35.650$ GB and $OINST-DYNRATIO \leq 1.319$ and $AGGCOEFVAR > 3.681$. In this situation, the overall variability metric AGGCOEFVAR is the key, as it determines the placement. If this value lies below 3.681 (see L12), then IaaS is the preferred placement option. If this value lies above 3.681 (L11), then the decision changes. This behaviour indicates, that there is an in-house niche for software applications with higher IT resource variability.
- Line L37: $DYNRATIO \leq 0.917$ and $CPU_QO \leq 8.235$ and $OSDEGR > 0.45$ and $STO_EXP \leq 1095.781$ GB and $OINST-DYNRATIO \leq 2.136$. In this situation, the resource-specific cost ratio OINST-DYNRATIO is the key, as it determines the placement. If this value lies above ca. 2.136 (see L36), then IaaS is the preferred placement option. If this value lies below or at 2.136 (L37) then the decision changes.
- Line L42: $DYNRATIO \leq 0.917$ and $CPU_QO \leq 8.235$ and $OSDEGR \leq 0.45$ and $STODYNRATIO \leq 0.382$. If the storage cost only make up a certain part of the total IaaS cost ($STODYNRATIO \leq$

Table 6.19: Comparison of predictor variables

Variable Type	Input Variable Name	Variable in un-weighted tree	Variable in weighted tree
Parameters	OSDEGR (η), ADELTA (a_Δ), PDELTA (p_Δ), QALPHA (α)	OSDEGR	OSDEGR
Total Cost Ratios	DYNRATIO	DYNRATIO	DYNRATIO
	RESRATIO		RESRATIO
Resource-specific Cost Ratios	AINSTDYNRATIO		
	AINSTRESRATIO		
	DYNINSTRATIO		DYNINSTRATIO
	OINSTRATIO		
	OINSTDYNRATIO	OINSTDYNRATIO	OINSTDYNRATIO
	OINSTRESRATIO	OINSTRESRATIO	
	INSTIHRATIO		
	NWIDYNRATIO		
	NWIRESRATIO	NWIRESRATIO	
	NWODYNRATIO		
	NWORES RATIO		
	STOIHRATIO		
	STODYNRATIO	STODYNRATIO	STODYNRATIO
	STORESRATIO		
STORATIO	STORATIO		
Workload Levels	CPU_EXP, CPU_QA, CPU_QO	CPU_QO	CPU_QO
	RAM_EXP, RAM_QA, RAM_QO		RAM_QO
	NWI_EXP, NWI_QA, NWI_QO		
	NWO_EXP, NWO_QA, NWO_QO		
	STO_EXP, STO_QA, STO_QO	STO_EXP	STO_EXP
Workload Variability	CPU_SV, RAM_SV, STO_SV, NWI_SV, NWO_SV	RAM_SV	
	CPU_CV, RAM_CV, STO_CV, NWI_CV, NWO_CV	STO_CV, NWI_CV	STO_CV
	VARIAB, AGGCOEFVAR	VARIAB, AGGCOEFVAR	AGGCOEFVAR

0.382) and the outsourcing degree is rather low and if only applications with rather limited CPU core count are in focus, then in-house placement is the best option.

- Line L23: $DYNRATIO > 0.917$ and $OSDEGR \leq 0.55$ and $OINSTDYNRATIO \leq 2.522$ and $STO_DYNRATIO \leq 0.382$. This is the most important rule for in-house placement. It resembles L42 in certain ways: if the storage costs only make up a certain part of the total IaaS costs ($STODYNRATIO \leq 0.382$) and the outsourcing degree is rather low and if $DYNRATIO$ crosses a certain threshold (i.e. IaaS resources get comparatively more expensive than in-house resources), then in-house placement is recommended.

An IaaS placement is the optimal choice for applications in the investigated outsourcing scenarios under the following conditions:

- Line L30: $DYNRATIO \leq 0.917$ and $CPU_QO > 8.235$ and $OSDEGR \leq 0.25$ and $STO_CV \leq 0.320$ and $STO_CV > 0.172$. The storage coefficient of variation (STO_CV) determines the placement decision. As long as it remains between 0.172 and 0.32 and if $DYNRATIO$ lies below a certain threshold (i.e. IaaS resources are comparatively cheaper than in-house resource), then an IaaS placement is better. If STO_CV drops below 0.172, an in-house placement is preferred (L31), which seems the logical choice given the low outsourcing degree.
- Line L12: $DYNRATIO > 0.917$ and $OSDEGR > 0.55$ and $RAM_QO \leq 35.650$ GB and $OINSTDYNRATIO \leq 1.319$ and $AGGCOEFVAR \leq 3.681$. This is the mirror case to L11 explained above.
- Line L09: $DYNRATIO > 0.917$ and $OSDEGR > 0.55$ and $RAM_QO \leq 35.650$ GB and $OINSTDYNRATIO > 1.319$. In this rule, the outsourcing degree is rather high, so although $DYNRATIO$ is unfavorable, a high number of applications have to be placed in the Cloud. Therefore, it makes sense to pick those applications with low IT resource requirements ($RAM_QO \leq 35.650$ GB), which are cheap in absolute terms.
- Line L34: $DYNRATIO \leq 0.917$ and $CPU_QO \leq 8.2348$ and $OSDEGR > 0.45$ and $STO_EXP > 1095.78$ GB. In this rule, the outsourcing degree is rather high, but $DYNRATIO$ is favorable for IaaS placement. It makes sense to pick those applications with high IT storage requirements ($STO_EXP > 1095.78$ GB), as Amazon AWS storage prices are very competitive.
- Line L26: $DYNRATIO \leq 0.917$ and $CPU_QO > 8.235$ and $OSDEGR > 0.25$. This is the most important rule for IaaS placement. A large portion of applications with CPU core requirements greater than 8 cores end up in this leaf of the tree. The favorable range of $DYNRATIO$ only supports this conclusion. It is instructive to look at the cases not ending up in this leaf (L27-L31); the rules L28 and L30 are discussed above. It can be assumed that an outsourcing degree lower than 0.25 is strongly correlated to in-house placement decisions.

It can be concluded that the cost-related outsourcing decision drivers themselves are largely as expected, e.g. $DYNRATIO$ (the quality-related outsourcing decision drivers were already discussed in section 6.3.3). However, the number and the effects of the input variable interactions are unexpected and are at times surprising, but they can be understood in the context of the outsourcing scenario. These interaction effects are of major importance for explaining cost-based placement decisions; they also justify the application of a decision tree as a machine learning algorithm, as it is able to capture and to visualize these interaction effects. The most important conditions under which a clear cost-based decision can be recommended are presented above.

The evaluation results allow the determination of the most important decision variables from the tree models in the case study outsourcing scenario (see Table 6.19). For the weighted decision tree, these

are OSDEGR, DYNRATIO, RESRATIO, OINSTDYNRATIO, STODYNRATIO, CPU_QO, RAM_QO, STO_EXP, STO_CV and AGGCOEFVAR. As part of the discussion of the experimental results, these decision variables shall be compared to those decision variables that are featured in the research results of other decision support tools with an IaaS background. Gray (2003) gives a rule of thumb for outsourcing decisions relating to Grid Computing. His main decision variable is the ratio between compute cost and Internet/WAN network cost; for a beneficial use of outsourcing, this ratio has to be at least 100,000 instructions per byte transferred over Internet/WAN network or around 8 CPU-h per Gigabyte of Internet/WAN network traffic. So, his ideal use case for outsourcing are compute-intensive, stateless jobs with little data transfer. The one-time job model is fundamentally different from the continuous-service model in this thesis; for example, the one-time data transfer costs are neglected in the above continuous-service model, as the cost are spread out over a longer period of IaaS usage time. On-going network charges are insignificant compared to storage and compute costs (see Table 6.11). Unlike continuous services, one-time jobs are thought to recover their one-time overhead costs (like network transfer) in one session, whereas services can spread out these overhead costs over the course of their lifetime. Therefore, the economies of these two outsourcing modes are hardly comparable, thus resulting in different decision variables. In summary, the research in this thesis complements and extends the results obtained by Gray (2003).

A classical decision support systems research approach was taken by Khajeh-Hosseini, Greenwood, Smith, and Sommerville from the University of St. Andrews. In a series of research articles, they develop a so-called Cloud Adoption Toolkit, implement it in an IT system and evaluate it using industrial case studies. As a first step, Khajeh-Hosseini et al. (2010) executes a first industrial case study, which features a cost comparison for infrastructure, support and maintenance costs and which includes a stakeholder impact analysis which lists chances and risks of IaaS adoption. The conceptual level is comparable to the expert interviews in this thesis. As a next step, Greenwood et al. (2010) introduces the Cloud Adoption Toolkit, a collection of models for technology suitability analysis, cost modeling, stakeholder impact analysis and responsibility modeling. The models are highly conceptual, so the results have more of a framework character. In Khajeh-Hosseini et al. (2011), the Cloud Adoption Toolkit is applied to two more case studies. The cost modeling is more elaborated, but the degree of decision support is low: the model provides factual, consolidated information, but no automatic recommendations for an optimized IaaS placement. Software application workload elasticity is reflected in the cost model, but not based on statistical properties of historical workloads, but by manual estimates of some predefined workload variation patterns. The outcome is a collection of useful information that increases transparency and awareness for the decision maker in a given outsourcing situation, but does not generate general decision variables or determinants for IaaS usage. In their summary paper (Khajeh-Hosseini et al. 2012), the Cloud Adoption Toolkit is described in terms of five useful tools/techniques (technology suitability analysis, energy consumption analysis, stakeholder impact analysis, responsibility modeling, and cost modeling). The article focuses on the tool for IaaS cost modeling, as it is the most mature. The tool includes a simple language to describe variable workload patterns for computing resources, so software application workload patterns can be entered manually. The deployment decisions need to be entered manually, which enables a what-if analysis. The IaaS cost types are comparable to the one presented in this thesis, but the cost model features multiple providers. However, only a simple tariff model is supported (per-unit price, no non-linear tariffs). The focus rests on the system implementation and the evaluation case study. In summary, these models/tools provide cost transparency to the decision maker, but no higher forms of decision support (like suggesting or even optimizing a decision). Hence, the results in this thesis go beyond the scope of the cost-modeling-related parts of their research.

Similar to Khajeh-Hosseini et al. (2012), Menzel et al. (2013) present a very generic and comprehensive decision-making framework for Cloud Computing scenarios. It combines well-established decision support methods (e.g. AHP for multi-criteria decision making, satisficing for criteria importance rating) and structures the decision-making process by defining key activities (scenario definition, criteria definition, re-

quirements definition, etc.), but mostly on a conceptual level. They present generic, literature-based criteria, requirements and cost-category catalogs which support IT infrastructure decisions qualitatively. The complete process is implemented as a Web-based tool. The result is a ranking of the alternatives that have to be defined earlier in the course of the decision-making process. The goal of this framework is to provide improved information and transparency to the decision maker and to structure the decision-making process. As such, its scope is more conceptual than the scope of this thesis, whose results (quality dimensions, IaaS usage determinants, decision tree as a decision support tool) could be integrated into their framework.

Chaisiri et al. (2011) and Chaisiri et al. (2012) focus on modeling the IaaS VM resource provisioning problem for multiple IaaS providers. The authors solve the resulting cost optimization problem using stochastic optimization with multi-stage recourse. Their comprehensive model is based on the AWS tariff system, and includes staged sourcing phases (reservation phase, expanding phase, on-demand phase). They approach the topic with a strong operations research background; large parts of the paper describe the optimization problem, the solution algorithm and various algorithmic improvements for solving multi-stage, integer stochastic models. The experimental evaluation of the resource provisioning algorithm is conducted with synthetic and real-world case studies; a comparison of the proposed algorithm with other provisioning algorithms shows the superiority of their approach. Their optimization model is similar to the model designed in this thesis (see section 5.6), but could not be used as an extension or replacement, as their model optimizes the bulk sourcing of VMs (i.e. it answers the question, how many VMs of a certain type are needed). They use a flat, job-based, summary resource model, which assumes horizontal scalability of the software applications to be placed in an IaaS Cloud. So, their research is complementary to this thesis, as they analyze a different software application model: their planning is on the level of application type, but not application-specific. Also, their research is extended by this thesis, as it establishes a method to link outsourcing scenario characteristics and workload characteristics to the deployment decision.

Lilienthal (2013) suggests an elegant economic decision support model for determining the optimal amount of internal IT infrastructure capacity. His application model aims at job-based, horizontally scalable workloads and continuously scalable IaaS infrastructure resources; he does not distinguish specific applications, but uses the summary workload of all applications. The optimization model features as decision variable the amount of internal IT infrastructure capacity (as opposed to the amount of external IT infrastructure capacity sourced from an IaaS provider). The three-part cost model in his work can be considered as a special case of the N-part cost model developed in this thesis. The model evaluation uses Grid Computing and compute cluster workloads. As use cases, the author presents Cloud bursting and resource pooling, which are obvious scenarios for this type of workload. In terms of IaaS usage determinants, only the pay-per-use tariff is discussed; the analytic solution for the 3-part tariff with fixed costs and volume discounts is not explicitly analyzed. Nonetheless, the comparison of the paper's results for the pay-per-use tariff with the thesis' results obtained from the decision tree is insightful. The paper's single determinant of the optimal amount of internal capacity is the ratio of internal fix cost and external variable cost. In comparison, DYNRATIO (i.e. the ratio of total cost for the elastic IaaS tariff and the total cost for the in-house tariff) is the most important predictor in the decision tree (see Figure 6.9). The two models identify a similar cost ratio, but the results are not directly comparable, as the paper's cost ratio is not tracked as an input variable in this thesis and thus cannot appear in the decision tree. The reason why the paper's model yields a much simpler determinant structure lies in its use of a continuous workload and continuously scalable IaaS VM resources, whereas this thesis' model breaks down the workload for specific software applications and assumes discrete and unevenly-sized IaaS VM resources. These discontinuities most likely require a more complex structure of IaaS usage determinants.

6.6.2 Limitations of the Evaluation

Vendor selection is generally a multi-objective problem (Wadhwa and Ravindran 2007), and the selection of an infrastructure service provider makes no exception to this rule. Therefore, the cost-based objective function can be criticized for being one-dimensional. However, many quality-related decision factors can be modeled economically (e.g. availability could be addressed by additional costs for safety capacity) or are covered in earlier stages of the outsourcing process and hence are treated using different methods (e.g. workload selection) (see section 2.3.4 on the outsourcing process). Thus, a single-objective, cost-focused approach might still be reasonable, especially given the high practical relevance of cost-efficient IT solutions (which became also apparent in the survey results). Moreover, Degraeve, Labro, and Roodhooft (2000) showed in a case study, that a TCO-based approach delivers comparable or even superior decision quality compared to approaches only focusing on certain cost aspects or to approaches with a different methodology altogether (e.g. rating models). Also, multi-criteria models suffer from difficulties in objectively interpreting the resulting multiple solutions lying on the efficient frontier.

The fundamental problems of any TCO approach are discussed in detail in Treber, Teipel, and Schwickert (2004, p. 39). Especially the missing standardization of the TCO models from different consulting agencies (e.g. Gartner and Meta Group) hinders the comparability of TCO values from different TCO schemes (Treber, Teipel, and Schwickert 2004, p. 41). Each consulting agency applies its own set of assumptions in the course of the TCO calculation, so that the TCO values show a great variability even for standardized IT services like Desktop PCs. Also, the weak scientific foundation of the TCO concept leads to a superficial application in practice (Treber, Teipel, and Schwickert 2004, p. 40). TCO calculations are valuable for increasing cost transparency, but have to be critically interpreted. Therefore, the IaaS tariff model tries to include the relevant infrastructure cost factors that are comparable across different providers.

The scenario introduced in section 5.5 concentrates on a generic IaaS use case, but it rests on numerous assumptions (please see section 5.2). Its practical relevance must be considered high, as there are existing Cloud providers, that offer dynamically scaled SAP ERP systems and as there are enterprise clients that actively pursue the outsourcing of ERP business applications.⁴ However, as the public IaaS Cloud services only offer very limited QoS guarantees, the pool of suitable business applications, that can be supported on those guarantees, might be small. Additionally, the outsourcing of business applications entails a substantial up-front effort (e.g. for provider negotiations, modifications of the business software, set-up of interface connections to the outsourcing partner, etc.), such that a short-term migration to and from the Cloud is rather unlikely (“short term” means here days or a few weeks). A medium- to long-term perspective on software migrations in a hybrid Cloud seems more realistic, even for SAP ERP applications.

The analytical model assumes comparable performance ratings for both in-house and IaaS IT resources. This assumption must be considered an approximation because of service provider shirking (Durkee 2010) and fundamental problem of incomplete contracts in outsourcing relationships for non-trivial services (Nam, Rajagopalan, Rao, and Chaudhury 1996). However, the continuing success of commercial IaaS offers has also created an eco-system of IaaS monitoring and benchmarking solutions, so that performance deviations can be tracked (e.g. using a service like CloudClimate⁵).

For the evaluation of the decision tree, a random tree was used, which tried to learn from random placement decisions. Currently, only a single trial was executed, i.e. only one random placement decision was randomly chosen for each training example, although the solution space of possible random placements is vast. A bootstrapping approach over the random placement decisions would be required to arrive at more reliable validity estimates for the random tree.

⁴<http://www.t-systems.de/ueber-t-systems/cloud-computing-neue-perspektiven-in-der-cloud/753500>, last accessed 2013-12-29

⁵<http://www.cloudclimate.com/ec2-eu/>, last accessed 2013-12-29

The experimental decision support model evaluation is based on a specific real-world data set (see the data collection section 6.4) and on a specific set of assumptions regarding the outsourcing scenario. Hence, the results produced in this chapter (e.g. the structure of the decision trees, the identification of the most important decision variables and the IT resource quality considerations) have to be interpreted in this context. So, the external validity of the results in this chapter can be critically questioned. Although measures were taken to preserve external validity (e.g. generic scenario variables, cost input variables are defined as ratios, a_{Δ} and p_{Δ} are included to model different cost structures among IaaS providers and business clients), the tariff data and the workload input data are irrevocably tied to a specific client-provider combination. So, the actual results might vary, when the decision support model is applied to a different outsourcing scenario. Nonetheless, the evaluation demonstrates the external validity of the tariff, workload and optimization model by successfully applying them to a real-world use case. Moreover, the hypothesis, that a decision tree is an apt instrument for the identification of IaaS decision variables, is also supported by the evaluation results, so this method can be generally transferred to other outsourcing scenarios.

6.6.3 Contributions

The theoretical and practical contributions of the optimization model are discussed in the following paragraphs. Their aim is to clearly expose the knowledge increase over the state-of-the-art found in the research literature and to demonstrate the practical relevance of the findings.

6.6.3.1 Contributions to Researchers

The starting point for this work's consideration was the need to understand the motivations, attitudes, and adoption behaviors of those corporate executives deciding on infrastructure service usage "out of the Cloud". In this thesis, IaaS adoption was explored on two complementary levels: first, using a multi-theoretical empirical model and second, using an outsourcing-process based analytical model. The two approaches complement each other, as the empirical approach analyzes the general overall factors and the analytical approach models the specific outsourcing situation.

One of the main findings of the infrastructure quality model is the relevance of IT infrastructure specificity. The research literature (e.g. (Klems, Nimis, and Tai 2008), (Gray 2003), (Leong 2009), (Carr 2003)) generally considers IT infrastructure as an interchangeable commodity, without paying attention to the technical and operational differences in IT infrastructure services. However, this view is not substantiated by the findings of this thesis. The differences in commodity IT infrastructure are motivated by business needs, at least for large companies, as these companies feature high ratings for infrastructure specificity in the survey. The results in the BMW storage quality comparison further support this claim. This higher specificity of large enterprises may be rooted in a more complex business model or more resources to implement elaborated IT solutions. Hence, although IT infrastructure can be seen as a commodity, its specific application becomes strategic for larger companies. A similar view, yet on a more conceptual level, was voiced by opponents of Carr (2003) (e.g. Vandenbosch and Lyytinen (2004) or Brown and Hagel (2003)). Consequently, future research needs to adopt a differentiated view towards IT infrastructure for large enterprises.

As another research contribution, the analytical model represents the outsourcing situation more systematically and comprehensively than the case study approaches taken, e.g. by (Khajeh-Hosseini, Greenwood, and Sommerville 2010), (Khajeh-Hosseini, Greenwood, Smith, and Sommerville 2012) or (Risch and Altmann 2008), which focus on the individual company and the contingencies of the situation. Moreover, the suggested model identifies specific drivers as opposed to general cost-calculation models like (Leong 2009) or (Khajeh-Hosseini, Sommerville, Bogaerts, and Teregowda 2011). It also extends the Cloud adoption toolkit proposed by Khajeh-Hosseini, Greenwood, Smith, and Sommerville (2012). Its authors demand that "decision makers have to model the variations in resource usage and their systems' deployment options to

obtain accurate cost estimates” (Khajeh-Hosseini, Greenwood, Smith, and Sommerville 2012); the analytical model in this thesis follows up on this suggestion and implements an elastic tariff for variable software application workloads. The analytical model also extends simple VM instance-based cost optimization approaches as in (Chaisiri, Kaewpuang, Lee, and Niyato 2011), as it considers outsourcing process-related situational factors and includes storage and networking costs. As this work shows, storage costs are non-negligible and the storage and RAM workload variations can be a determining factor for the application placement decision.

6.6.3.2 Contributions to Practitioners

By providing practitioners with some insight into the decision makers’ perception of IaaS, the survey research framework serves as a basis for the management of the so far poorly understood IaaS acceptance process and for IaaS-related outsourcing decisions. This knowledge is especially relevant for IaaS providers which aim at maximizing the effectiveness of their sales activities. But also IT executives can profit from these results by better understanding the mindset of their business departments and by reacting accordingly.

The experimental part features two elaborated decision support models: one for the IaaS quality dimensions (see section 5.3) and one for the financial dimension (see 5.6). Both models are evaluated using a BMW use case and the results were presented in section 6.3 and in section 6.5.4. The practical contributions of these models and the relevance of the research questions R2 and R3 can be judged best by IT experts in the field. Therefore, these evaluation results were presented to BMW IT experts and discussed in expert interviews, which took place in the first quarter of 2013. Their outcomes are recorded in writing and the experts then reviewed these notes for accuracy. The following two paragraphs summarize these expert interviews.

The BMW expert works as an IT infrastructure architect in the storage systems group within the BMW Data Centers IT department. Its tasks are various: the centralized, uninterrupted 7*24 operations of the data centers and the networks, the provisioning of hardware servers following up on business requirements, cost-optimizations by increased virtualization of servers and the standardization of operations tools, processes and methods across different sites. This expert has clear expectations as far as technical properties of Cloud Storage services are concerned, some of which are certain metrics that are important for the performance of a storage system (e.g. storage system cache utilization or the number of cached, but pending write transactions). Public Cloud providers do not publicly disclose the measurements for these metrics, nor do they give guarantees as to their minimum thresholds. In general, performance SLAs for Cloud storage services are hard to find, according to the expert. Current BMW storage systems are engineered to correspond to the specific technical requirements dictated by the business processes which those systems are supposed to support. Consequently, the commodity Cloud storage services may only be suitable for a subset of BMW applications, which happen to have matching QoS requirement. So, the current utility and the architectural potential of current Public Cloud storage services is limited, from this expert’s point of view. Nonetheless, the instantiation of the IaaS tariff model using the AWS tariff data (see sections 5.4.2 and 6.4) and the comparison against an instantiation of the BMW tariff model were considered interesting and relevant.

Another expert works as an IT infrastructure architect in the Data Centers IT department. His areas of specialty are IT Unix/Linux servers and he is responsible for Linux as a software product within BMW. Linux is an enabler for further IT infrastructure topics like HA (high availability) concepts and server virtualization. He is the lead for the BMW-internal “Infrastructure as a Service” project. The research approach, the data set and the evaluation results of the optimization model were discussed with this expert. According to the expert, the enterprise is in contact with IaaS providers, however a sustainable outsourcing relationship has not yet been established for various reasons. These include financial details and the problem of downtime constraints imposed by the provider (i.e. time windows for scheduled maintenance cannot be freely

chosen). A future hybrid sourcing model with externally-provided IaaS services is still being considered however. The strategy can be very selective, as BMW operates a critical mass of IT infrastructure, which can be run efficiently in-house. A recent Gartner IT benchmark analysis confirmed the competitiveness of the in-house tariffs, according to the expert. The long-term goal is a hybrid Private Cloud using an automation abstraction layer and advanced virtualization techniques; in this Cloud, the central IT department is going to be the Cloud broker, coordinating the service provisioning and consumption. In fact, virtualization is seen as a mature technique, which can also support highly critical business functions like production control. As the majority of BMW applications follows a scale-up capacity management approach and as downtimes have to be minimized, live migration capabilities enabled by virtualization are seen as a necessity by this expert. These statements support the principle interest in and the relevance of research questions R2 and R3.

In a second interview, the same expert offered his feedback on the experimental validation results of the decision support model, which are judged generally interesting; the decision tree would lead to simplified decisions. The inclusion of dependencies among software applications was considered a necessity, but it would result in a more complicated model with clusters of dependent software applications. As a possible use case, the optimization of the BMW data center landscape would be imaginable with a similar model (e.g. placement decisions based on network latency instead of financial cost). Hence, the results of research question R3 were rated favorably by the expert.

As a summary to the expert discussion, the realization of a truly hybrid Cloud incorporating BMW in-house and external resources will only be achieved in the future. The (cost-)efficiency of IT solutions is always of interest; the optimization of the BMW IT landscape and the adaptation of the IT building blocks to the business requirements are a constant challenge for the internal IT department. The specificity of BMW business requirements necessitates the deployment of highly specific IT infrastructure and corresponding operations processes, both of which are hard to find at current IaaS Cloud providers. However, current public IaaS offerings are already a topic of great interest for the experts, but primarily as a benchmark, against which BMW IT services are compared (by both business and IT executives) and which motivates the development of comparable IT services for a Private BMW Cloud. Hence, the development of tariff and cost-optimization models allows the comparison of in-house and external resources and thus provides relevant input for these benchmarking activities.

Part IV

Finale

Chapter 7

Conclusion and Outlook

7.1 Conclusion

The conclusion picks up the problem motivation and the research questions exposed in the introduction chapter and gives a short summary of the relevant findings for each research question.

7.1.1 Research Question R1

Research question R1 was termed “What are the overall IaaS usage determinants of enterprises?” To this end, an empirical study was thoroughly prepared and executed. The primary objective of the study was to examine organizational acceptance and adoption of IaaS in the light of a framework of various well-researched theories from multiple backgrounds (e.g. theory of planned behaviour, transaction cost theory and principal-agent theory). Research question R1 treats the area of qualitative determinants of flexible infrastructure sourcing within the enterprise, i.e. what are the determinants (drivers and deterrents) of flexible infrastructure sourcing. Here, the focus is on the functional, technical and organizational determinants, e.g. governance, security, architectural requirements. These determinants play a major role in all stages of the decision process to use outsourcing and address the principle suitability of an IaaS usage.

The empirical evaluation of the pre-tested causal model allows for the confirmation of all hypotheses developed in Section 3.4.2, except the hypotheses related to the direct influence of uncertainty on intention to use. The determinants hypothesized in the IaaS adoption model are largely supported by the empirical results. Perceived usefulness (as defined by strategic, flexibility and efficiency added-values) and ease of use are the two strongest predictors of IaaS adoption. Small and medium enterprises are more prone to IaaS adoption and less risk-averse regarding the uncertainties involved. The empirical findings of the IaaS adoption model are consistent with related studies from a Software-as-a-Service background, but extend the understanding of uncertainty, informant role and firm size on IaaS adoption. These findings help business executives determine the selling points of an IaaS adoption and the best organizational setting for an IaaS introduction.

Perceived usefulness has an overall “large” effect size, the third largest in the whole model. The first indicator in this construct, the strategic added-value, measures advantages exceeding the operational and tactical level and influences the positioning of the enterprise in its market segment. The second one, the flexibility added-value, enables the enterprise to react elastically to varying demands for IT infrastructure. The third one, the efficiency added-value, improves the speed and/or the cost efficiency of existing processes or systems. Thus, IaaS positively influences the strategic positioning, the flexibility and the efficiency of enterprises, which in turn indirectly drives its adoption. Notable is also the fact that the efficiency added-value is the strongest of the three indicators. The level of perceived usefulness is dependent on the company

size and the IT affiliation of the respondent. SMEs are generally better able to extract value from their IaaS deployments than larger enterprises and business executives perceive IaaS to be more valuable to the company than IT executives.

As far as perceived uncertainty is concerned, infrastructure specificity, fear of provider opportunism and information security concerns were all statistically significant drivers and partly explain the construct of perceived uncertainty. However, the level of the individual drivers again depends on the company size and the informant's background. SME exhibit less data security concerns when using IaaS Cloud providers. This observation matches the overall lower perceived uncertainty of SMEs than of larger enterprises. This reduced uncertainty perception might also be due to the fact that their IT infrastructural demands are less specific than those of larger companies, and hence, the risk of lock-in situations and other provider dependencies is lower. In general, business informants perceive IaaS to be less insecure than IT informants. This fact might be attributable to the deeper expertise of IT informants, which allows them to more realistically judge the security risks associated with an IaaS usage.

The intention to use IaaS resources could also be explained satisfactorily: attitude, subjective norm and perceived behavioral control have a statistically significant effect on the organization's intention to use IaaS. Attitude possesses a "large" effect size, the second largest in the model. The hypothesis for perceived uncertainty, also a hypothesized antecedent of intention to use, needs to be rejected. Regarding the intention to use, the findings highlight that subjective norms have an important influence on behavioral intention. Normative pressure is especially felt from external consulting agencies, comparable enterprises, employees in the business departments, and superiors. These persons or institutions actively exert influence and recommend the IaaS usage. However, it is interesting to see that the business department has the strongest effect and the IT department has one of the lowest effects. This observation fits the previous findings; as the IT informants generally have a more critical perception of IaaS usage, it is comprehensible that they do not actively promote the deployment of IaaS services in the enterprise (and vice versa for business executives, that are eager to deploy those services).

Perceived uncertainty is supposed to measure the difficulties in predicting environmental risks. Regarding the perceived uncertainty construct, the findings indicate that this construct seems to have no direct influence on the intention to use IaaS. However, this counterintuitive result has to be put in perspective. First, a part of the explanation is provided by the company size, which can be shown to act as a moderating variable between the perceived uncertainty and the intention to use IaaS. Second, another possible explanation may be found in the IaaS usage scenarios at that time; the immature sourcing option IaaS may have been used then mainly for non-critical and commodity IT applications for which the uncertainty associated with IaaS plays no role.

The most important uncertainty factors are business-related difficulties caused by price changes of IaaS services, business-related difficulties caused by changes in process-critical/ operational performance indicators (e.g. unplanned downtime), legal difficulties caused by external data storage (e.g. because of unclear legal situation regarding compliance regulations) and technical difficulties (compatibility of IaaS providers with IT standards in the enterprise). From the informants' perspective, legal difficulties were the hardest ones to predict and contributed most strongly to the perceived uncertainty. The intention to use IaaS is significantly higher on the business side than on the IT side. The company size as an independent variable does not seem to play any role in the intention to use IaaS (but plays a role as a moderating variable).

Intention to use IaaS was found to be a very strong predictor of actual usage; perceived behavioral control affects the adoption only marginally. The intention to use has a "large" effect size; the largest in the model. SMEs show a higher IaaS adoption propensity than large enterprises for both adoption measures. The analysis also reveals an overall greater propensity of business informants to adopt IaaS than IT informants for both adoption measures. These observations match the aforementioned findings regarding the role of SMEs, and can be explained by them, as they provide the underlying causes for this behaviour (the

differences in attitude and intention to use between SMEs and large companies). The determinants found in R1 from the basis for the modeling activities needed for R2; also, these determinants are finally applied to a specific outsourcing situation to design a decision-support system for this situation.

7.1.2 Research Question R2

Research question R2 is worded “What determinants are relevant in an economic optimization model of hybrid IaaS sourcing?” Research questions R2, R3 complement R1 in that they increase the level of detail of the analysis. As economic benefits play a major role in the decision to adopt IaaS services, a deeper investigation in the cost-benefit relationship of flexible infrastructure sourcing is warranted. To this end, a cost-based decision support model has been developed using the results of R1. This model reflects the flexible nature of IT resource usage in the Cloud and it serves as an answer to R2.

The relevant determinants in an economic optimization model of hybrid IaaS sourcing are the input variables of the decision tree in Table 5.3. These input variables themselves are based on the scenario assumptions, and the economic optimization model, which can be split up in a tariff model, an IT resource model and an optimization model, which builds on the tariff and resource models, and offers a cost-based linear optimization approach which calculates a cost-efficient allocation of software applications to two different IT resource providers. The optimization model is only dependent on the empirical distributions of IT resource demands; no parametric assumptions regarding these distributions were taken and no time-series calculations are involved. Hence, assumptions often taken for time-series like stationarity or periodicity are not needed. The optimization model tolerates missing values in the historical recordings of the IT resource workloads and can be instantiated with a minimal amount of workload data to be collected. Elasticity is built-in, i.e. the automatic selection of the most appropriate VM instance size. The model is shown to be computationally simple (essentially as complex as a standard sorting algorithm).

The economic model must be complemented with a quality model for IaaS infrastructure offerings; quality dimensions are needed for assessing the comparability of QoS criteria across different providers. The quality model is based on a comprehensive list of quality dimensions from the e-Service quality research background; these dimensions are split up in system-related quality dimensions (operational time, performance, availability, stability, recoverability, security, standardization, product features) and operations-related quality dimensions (service time, service performance, reliability (on-time delivery), empathy, assurance) (see Table 5.1 for a complete description).

7.1.3 Research Question R3

Research question R3 is called “What determinants in a hybrid IaaS sourcing scenario can be linked to an economically beneficial usage of public IaaS offerings?” It is a follow-up question to R2 with the goal to evaluate the theoretical models developed for R2. The analysis of the IaaS usage drivers is grounded in an abstract model of the outsourcing process taken from the research literature. The vendor selection step of this process is the most likely place for such an analysis, as it entails the comparison of potential IaaS providers. The notion of an outsourcing scenario is defined; it is described by a number of key parameters relevant to the IT decision makers in an enterprise. A unique data set could be obtained for the purpose of the evaluation: real-world application system workload traces and realistic historic BMW and AWS cost figures. Care was taken to ensure the comparability of the two infrastructure providers, especially in terms of tariff models and the underlying cost structures of both enterprises.

From a descriptive analysis of the statistical workload characteristics, it can be concluded that the workload data is long-tailed and positively skewed in every IT resource dimension. Hence the non-parametric nature of the optimization model is well-suited to handle this data set. The descriptive evaluation results reveal the differences between the elastic and the reserved IaaS tariff. To minimize VM instance costs in

the elastic tariff, the base VM instance size should be set at 30% of the maximum application workload (in terms of CPU and RAM utilization). The significance of this effect was statistically shown. In general, the IaaS elastic tariff guarantees significantly lower costs than the IaaS reserved tariff under the same conditions; hence, the relevance of this artificially created tariff becomes obvious. The analysis of the total IaaS costs shows that VM instance-related costs claim between ca. 78% and 84% of the total cost (depending on the α level); networking costs are negligible. The remaining 15-21% are incurred for storage infrastructure.

An algorithmic machine learning approach (decision tree) was applied to identify the IaaS usage drivers in the aforementioned outsourcing scenario. As the algorithm relies on certain configuration parameter settings, a systematic search for the optimal parameter settings was successfully conducted. State-of-the-art procedures and metrics ensure the validity of the resulting decision tree. The tree was controlled for overfitting and for the stability of the prediction quality. These preparations lead to high values for all measures of predictive performance (accuracy, precision, recall and MCC). In an extended evaluation step, the training cases were weighted with the averaged costs across the three tariffs as devised by the pre-calculated optimized placement decision. Cost-related weighting of the training cases lead to a simplified, hence a better understandable decision tree, with negligible changes in the predictive power and in the cost deviations caused by wrong placement predictions.

The execution of the machine learning approach yields the relevant and the irrelevant factors of the application placement decision in the outsourcing scenario. The expectation, that the semi-variance measures would be strongly linked to the placement decision, had to be revised, as they were not found in the weighted tree (and only played a marginal role in the unweighted tree); variability measures based on the coefficient of variation obviously have higher predictive power.

The ratio between the total cost for the elastic IaaS tariff and the total cost for the in-house tariff (DYNRATIO) is the single most important explanatory variable in the weighted decision tree. The placement decisions seem to be fundamentally different for software applications having a DYNRATIO either below or above 0.92. A higher DYNRATIO lends more weight to the outsourcing degree, which is the most important variable in this subtree. The other subtree with a lower DYNRATIO is structured differently, as the maximum number of CPU cores CPU_QO is the most important variable there (even before the outsourcing degree).

The outsourcing degree is also of major importance. However, it was shown, that the obvious direct interdependency between OSDEGR and the number of IaaS placements also has exceptions and depends on additional application workload properties.

The cost-based IaaS placement decisions have to be discussed in the context of the enterprise quality requirements for IT infrastructure. Therefore, a short synopsis of the quality feedback from the survey and from the quality model shall be summarized here. The survey revealed that “Adherence to Security and Compliance regulations” (i.e. Assurance) was highly rated in absolute terms and the most important quality factor in the eyes of the informants, closely followed by the also highly rated “appropriate Service Level Agreements” (grouping all types of system-related quality dimensions). Moreover, large enterprises are significantly more concerned with the contents of these SLAs than small or medium-size enterprises; IT departments are more concerned with SLAs than business departments.

These findings partly explain the situation that could be observed, when the quality attributes of BMW storage services were compared to the quality attributes of AWS. The BMW IT department features storage services with elaborated quality attributes matching specific business requirements, as suggested by the survey results for SLAs in large enterprises. Although appropriate Service Level Agreements are generally highly rated, a large portion of the BMW quality attributes for storage services cannot be satisfied in the Amazon Cloud yet. Therefore, only applications that are generic enough to fit the IaaS service quality profile are potential IaaS outsourcing candidates. Thus, the preceding cost-based outsourcing model is limited to these candidates, and can only give recommendations for this subset of software applications.

7.2 Outlook

The following paragraphs serve to illustrate potential future research topics and potential extensions to the models in this work.

The empirical model could be complemented with a hard-modeling approach like LISREL. This approach might be justified in the future, as the enterprise IaaS adoption is constantly rising and maturing, and so are the experiences of company users with this new sourcing option. The IaaS market expansion would also make available a larger sample size with more business informants. An established phenomenon, solid theoretical foundations and a large sample size are the preconditions for the aforementioned hard-modeling approaches.

Another extension of the empirical model consists in additional theoretical constructs explaining attitude, intention, adoption and perceived uncertainty. One of these constructs could be infrastructure quality with its associated dimensions; these could be the basis for a formative measurement model of IT infrastructure quality. A deeper inspection of IaaS quality dimensions, their perception and their effect on the adoption decision is certainly worthwhile, especially in light of the demonstrated relevance of infrastructure specificity and appropriate SLAs.

Another construct could be power and politics in business organizations, in accordance with Eisenhardt and Bourgeois III (1988), who investigated the effects of power and politics in strategic decision making. An application of the power and politics construct to IT infrastructure decision making could be rewarding, as the survey in this work already shows, that social factors play a significant role in the adoption decision.

A final additional construct is represented by the maturity of enterprise IT architecture, which is also part of the adoption model of Xin and Levina (2008). Enterprise IT architecture refers to “the organizing logic for applications, data and infrastructure technologies, as captured in a set of policies and technical choices, intended to enable the firm’s business strategy.” (Ross 2003) The aforementioned article also hypothesizes the maturity stages of enterprise IT architecture. Ross and Beath (2006) showed, that IT outsourcing arrangements can be used to help an enterprise move from one architectural stage to the next. Hence, a future study could investigate the role that IaaS might play for developing an enterprise IT architecture. As different architectural stages lead to different IT capabilities, the questions of IaaS added-value and of IaaS service sourcing possibilities at the different stages might also prove worthwhile. Instruments for measuring IT architecture maturity were developed by Perko (2008, pp. 228) and Engels (2007, p. 48).

Rogers (2003, p. 281) emphasizes the time dimension of technology adoption; according to him, adopters can be grouped in five categories describing the innovativeness of the adopter. The speed of adoption is directly correlated with the level of innovativeness of the adopter, hence there is a significant share of adopters that introduce new technologies later than average or even never. This temporal distribution of adoption could be analyzed for IaaS as well, using a longitudinal research approach to track the adoption in a predefined panel over time, especially from informants that rejected IaaS so far. Additionally, this approach would make it possible to include new decision situations (e.g. if the renewal of IaaS contracts is an option).

The decision support model lends itself to several possible extensions as follows: a first extension could incorporate inter-application dependencies into the resource model and subsequently into the placement algorithm. One option is the clustering of dependent applications into a combined VM node/new application. This extension relaxes the current model assumption of software applications being individual, independent and monolithic and it would enable the inclusion of functional dependencies and placement constraints in the placement algorithm. Such an extension would be helpful, when distributed applications (commonly found e.g. in multi-tiered server applications) have to be optimized. However, such an extension must not be confused with already existing server consolidation approaches (using techniques described in e.g. (Rolia,

Andrzejak, and Arlitt 2003)), which are more suited for IaaS providers trying to maximize the utilization of their data centers.

Currently, the resource variability is modeled as the semi-variance and as the coefficient of variation, and the workload variability is modeled as a vector norm aggregating the different resource variabilities. To arrive at a more general concept of workload variance, a multivariate measure of variability is required. One interesting candidate for this purpose is the generalized variance (Wilks 1932), a multivariate extension of the variance defined on univariate random variables. The generalized variance is calculated as the determinant of the variance/covariance matrix of a multivariate random variable. Applying the generalized variance to the multivariate application workload vector would yield an elegant metric for application workload variability. This metric could be seamlessly integrated in the placement optimization algorithm described above.

IT resource bundling and optimization without an exogenous variable α are two extensions that need to be discussed in the context of the optimization model in section 5.6. The optimization model defined in equation 5.28 was created under the assumption that α is an exogenous parameter, i.e. the level of base load is given. As an extension to the above model, this assumption can be dropped and the model can be formulated with α being one of the decision variables, so that the optimization model determines the optimal level of α . The principle set-up is similar to the basic model, however the objective function and the constraints need to be altered. Equation 7.1 defines the extended optimization model.

$$\min \sum_{i \in I} (\min (m_i^{\text{reg}}(\alpha_i), m_i^{\text{Elastic}}(\alpha_i)) y_i + p_{\Delta} m_i^{\text{In-house}}(\alpha_i) * (1 - y_i)) \quad \text{subject to} \quad (7.1)$$

$$\sum_{i \in I} y_i = \eta N_s \quad (7.2)$$

$$0 \leq \alpha_i \leq 1 \quad \forall i \in I \quad (7.3)$$

$$y_i \in \{0, 1\} \quad \forall i \in I \quad (7.4)$$

Unfortunately, this extended optimization problem is non-linear, as the value of $m_i(\alpha_i)$ is non-linearly dependent on α_i (equation 5.25, 5.26, 5.27). α determines the calculation of the quantiles (equation 5.10), and those quantile values can only be numerically determined if the underlying workload is of unknown statistical distribution, as in this work. Hence, $m_i(\alpha_i)$ is only numerically solvable by meta-heuristics (e.g. Tabu search, Simulated Annealing), for which only asymptotic performance guarantees exist. Thus, the extended optimization problem can be considered numerically hard.

The definition of bundling follows (Rosenthal, Zydiak, and Chaudhry 1995). Generally, bundling can be part of a providers pricing schedule and must therefore be analyzed in conjunction with discounts to obtain the complete pricing schedule of a provider. Whereas discounts are granted based on the purchased amount of any specific resource and hence only affect the cost of that specific resource, bundles are defined as a distinct combination of different resources, so the total cost can no longer be defined as the sum of the individual resource costs, but only as a combined cost function. In a way, both concepts, discounts and bundles are therefore conceptually orthogonal. An illustrative example of bundling is given in the following paragraphs:

Let there be two resource types r_1, r_2 with their respective prices c_1, c_2 . x_1, x_2 denote the purchased quantities of each resource. As bundles are defined in a specific provider price schedule, the example does not include a multi-provider analysis. $q_1, q_2 > 0$ are the minimum quantities for which bundling becomes possible and k are the free units of resource r_2 when bundling is used. Those parameters are defined by the provider tariff. The resulting cost function can be stated as follows:

$$C(x_1, x_2) = \begin{cases} x_1 c_1 + x_2 c_2. & \text{if bundling is not chosen} \\ x_1 c_1 + x_2 c_2. & \text{if } x_1 < q_1 \vee x_2 < q_2 \\ x_1 c_1 + x_2 c_2 (1 - \frac{k}{x_2}). & \text{if } x_1 = q_1 \wedge x_2 = q_2 \\ q_1 c_1 + q_2 c_2 (1 - \frac{k}{q_2}) + (x_1 - q_1) c_1 + (x_2 - q_2) c_2. & \text{if } x_1 > q_1 \wedge x_2 > q_2 \end{cases} \quad (7.5)$$

$c_2(1 - \frac{k}{x_2})$ is the effective price of r_2 . Proof: if $x_1 = 0$, the total cost is $c_2 x_2 - c_2 k$. If c' is assumed to be the effective price of r_2 , then the equations $c' x_2 = c_2 x_2 - c_2 k$ holds. Solving this equation for c' yields the result in equation 7.5. For example, a popular bundle is “buy one, get one free”. This bundle can be modeled using $q_1 = q_2 = 1$ and $k = 1$. Another possible bundling strategy could be described as follows: “if you buy q_1 of r_1 , you get each of the q_2 units of r_2 for a reduced price p'_2 .” This bundle can be modeled using $k = q_2(1 - \frac{c'_2}{c_2})$.

Bundling has not been analyzed deeply in this thesis, as this instrument is not widely used among current IaaS Cloud Computing providers (as it was shown in section 2.3). From an optimization standpoint, bundling poses a number of challenges in the formulation and in the solution of TCO-based optimization models. Those can be summarized as follows:

- As shown above, bundling leads to multi-dimensional cost functions. Those functions are piecewise linear, and hence have to be linearized using the same techniques as in the one-dimensional case. However, the linearized results in the multi-dimensional case can only be an approximation to the true cost function values, as the surface of the function usually cannot be linearized exactly.
- The formulation of optimization models for such bundling cost functions becomes even more complex, with an exponentially rising number of decision variables as a consequence of the higher number of dimensions.
- It is not clear, how bundling and discounts must be combined; from a mathematical standpoint, the two concepts are not commutative: the total costs depend on the sequence in which bundling and discounts are calculated (i.e. how the billing process of the purchased quantities is implemented).

The current optimization model is a deterministic model, that operates on measures of location (expected values of mean, quantile, etc.) for the random vectors involved. Alternatively, the random vectors could be directly incorporated into the optimization model, which would turn it into a single-level stochastic optimization model (Kall and Mayer 2011). Depending on certain statistical properties of the underlying distributions and depending on additionally required modeling assumptions, a number of possible solution procedures exists for these types of optimization problems (it could be argued that the current approach is already one possible solution procedure, at least according to Kall and Mayer (2011, pp. 140)). Hence, such an extension of the original problem would be a major endeavor.

Declaration about the thesis

Ich versichere hiermit wahrheitsgemäß, die Arbeit bis auf die dem Aufgabensteller bereits bekannte Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderung entnommen wurde.

Karlsruhe, 31.07.2014

Jörg Strebel

Lebenslauf

Name: Jörg Johannes Strebel
geboren am: 19.02.1978
Staatsangehörigkeit: deutsch
Heimatanschrift: Aachener Straße 2
80804 München
Familienstand: ledig
Telefon: +49 89 382 49768
Mobil: +49 151 601 49768
E-Mail: joerg.strebel@kit.edu

Schulischer Werdegang:

September 1988 - Juni 1997 Besuch des Friedrich-Alexander-Gymnasiums in Neustadt/Aisch mit Abschluss des bayerischen Abiturs (Note 1,1)
November 1998 - Juni 2003 Studium der Wirtschaftsinformatik an der Universität Erlangen-Nürnberg; (10 Semester; Note 1,4, "sehr gut")
August 2003 - Dezember 2004 Auslandsstudium an der Clemson University (South Carolina, USA); Abschluss als "Master of Science in Industrial Management" (GPA: 4.0)

Bisherige Arbeitserfahrung:

August - Oktober 2001 Forschungsaufenthalt im europäischen IBM-Forschungszentrum in Zürich
August 2003 - Dezember 2004 Teaching Assistant und Tutor an der Clemson University
Januar 2005 - Oktober 2007 Consultant bei der Accenture GmbH in München
Oktober 2007 - August 2010 Teilnahme am Doktorandenprogramm der BMW AG in München
seit Januar 2008 Start der Promotion am Lehrstuhl Prof. Weinhardt am KIT
seit August 2010 Tätigkeit als IT-Spezialist bei der BMW Group

Besondere Kenntnisse:

Sommer 1997 Hochbegabtenstipendiums des Freistaates Bayern
Oktober 2002 Ablegung des GMAT (710/800 Pkt., 95% Quantil) und des TOEFL (290/300 Pkt.)
März 2003 Förderung des Auslandsaufenthaltes durch die Fulbright-Kommission
August 2003 Förderung des USA-Studiums durch ein Graduate Assistantship der Clemson University
Februar 2010 eFellows-Stipendium

Appendix

Appendix A

Letter of Invitation

The invitation was sent out to the enterprises in form of an e-mail. The text is included in its original language German, as the invitation and all subsequent communication was also in German:

Subject: Wissenschaftliche Cloud-Computing Studie: Mitmachen, gewinnen und wertvolle Studien-Ergebnisse erhalten!

E-Mail body:

Sehr geehrter Herr / Sehr geehrte Frau ...,

der Lehrstuhl Information & Market Engineering am Karlsruher Institut für Technologie führt im Rahmen des staatlich geförderten Forschungsprojektes Biz2Grid (<http://www.biz2grid.de>) zurzeit eine Studie zur Infrastructure-as-a-Service(IaaS)-Nutzung von Unternehmen durch.

Ich möchte Sie als IT-Entscheider gerne exklusiv zur Teilnahme per Online-Fragebogen (<http://www.iaas-studie.de>) einladen. Selbst wenn Sie Cloud Computing momentan noch nicht nutzen, ist Ihre Meinung dennoch sehr wertvoll für mich.

Die Fragestellungen der Studie sind die folgenden:

- Was halten Unternehmen von Infrastrukturdiensten aus der Wolke?
- Welchen Einfluss haben technische, rechtliche, ökonomische und organisatorische Faktoren auf die Nutzungsentscheidung?
- Welche IaaS-Qualitätsfaktoren sind für Unternehmen wichtig?

Als Teilnehmer erhalten Sie als Dankeschön bei Angabe einer E-Mailadresse die Studienergebnisse (bisher mehr als 220 Teilnehmer). Außerdem verlosen wir unter allen Teilnehmern mehrere Bücher zum Thema "Cloud Computing".

Bitte nehmen Sie sich die Zeit und füllen Sie den Fragebogen vollständig aus. Zur Umfrage gelangen Sie unter <http://www.iaas-studie.de>.

Die Dauer dieser Umfrage beträgt ca. 12min (ermittelt durch einen Vorab-Test). Ihre Antworten werden selbstverständlich anonym erfasst und werden nur für wissenschaftliche Zwecke verwendet. Es ist technisch dafür gesorgt, dass keine Verknüpfung Ihrer Antworten mit Ihrer E-Mailadresse möglich ist.

Für Fragen stehe ich Ihnen natürlich gerne per Mail unter strebel@iism.uni-karlsruhe.de zur Verfügung. Weitere Informationen zur Studie finden Sie unter dem Link <http://www.im.uni-karlsruhe.de/biz2grid/umfrage/>.

Ich bedanke mich herzlichst für Ihr Interesse und Ihre Unterstützung!

Mit freundlichen Grüßen

Jörg Strebel

Institut für Informationswirtschaft und -management (IISM)

Forschungsgruppe Information & Market Engineering

Karlsruher Institut für Technologie

Englerstr. 14

D- 76131 Karlsruhe

Email strebel@iism.uni-karlsruhe.de

Web: <http://www.im.uni-karlsruhe.de/Default.aspx?PageId=379>

Appendix B

The Questionnaire

The Web-based questionnaire for the survey is presented here. All questions marked with an asterisk were mandatory (only a subset of questions were mandatory). Each question was presented on a separate screen; the participant had to click a button on the Web site to proceed to the next question.

IaaS-Studie

Die Umfrage zur **Infrastructure-as-a-Service-Nutzung** wird im Rahmen des Projektes **“Biz2Grid”** des Instituts für Informationswirtschaft und -management am **Karlsruhe Institut für Technologie (KIT)** durchgeführt. Ziel der Umfrage ist es, Treiber und Hemmnisse der IaaS-Nutzung zu untersuchen und die Verbreitung der Technologie zu analysieren.

Die folgenden Fragen sind speziell an **IT-Entscheider** und **IT-Manager** gerichtet! Bei den nachfolgenden Fragen geht es um Ihre persönlichen Einschätzungen, “falsche” oder “richtige” Antworten gibt es daher nicht.

Unter allen Teilnehmern, die den Fragebogen vollständig ausfüllen, werden am Ende der Studie Buchpreise im Wert von insgesamt 150€ verlost (siehe unten). Außerdem haben alle Teilnehmer die Möglichkeit, eine umfassende Auswertung der Studienergebnisse zu erhalten. Bitte geben Sie in beiden Fällen eine gültige E-Mailadresse an.

Da wir wissen, wie wertvoll Ihre Zeit ist, haben wir diesen Fragebogen so entwickelt, dass er in ca. 12min zu beantworten ist.

Ihre Antworten werden selbstverständlich anonymisiert abgespeichert und nur für wissenschaftliche Zwecke verwendet. Alle abgeleiteten Ergebnisse werden nur in aggregierter Form weitergegeben.

Jörg Strebel

Institut für Informationswirtschaft und -management (IISM)
Forschungsgruppe Information & Market Engineering
Karlsruher Institut für Technologie
Englerstr. 14
D- 76131 Karlsruhe

E-Mail: strebel@iism.uni-karlsruhe.de
Web: www.im.uni-karlsruhe.de/

Folgende Buchpreise werden verlost:



2x Cloud Computing Explained: Implementation Handbook for Enterprises von John Rhoton
 2x Cloud Computing: A Practical Approach von Toby Velte, Anthony Velte
 3x Cloud Computing: Web-basierte dynamische IT-Services (Informatik Im Fokus) von Christian Baun, Marcel Kunze, Jens Nimis, und Stefan Tai

Diese Umfrage enthält 21 Fragen.

Frage 1/15

1. Verfolgen Sie aktuelle Entwicklungen in der IT und hat das Thema Cloud Computing bei Ihnen das Interesse geweckt? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- O Ja
- O Nein

Definition von Schlüsselbegriffen: Infrastructure-as-a-Service (IaaS), eine Form des Cloud Computing, umfasst die Bereitstellung von Verarbeitungs-, Speicher- und Netzwerkkapazitäten und anderer grundlegender Rechenressourcen über das Internet durch einen externen Provider. Typische aktuelle Angebote sind Amazon's Elastic Compute Cloud (EC2), GoGrid's Cloud Service oder RackSpace's Cloud Hosting.

Frage 2/15

2. Stellen Sie sich vor, Sie planen den Einsatz von IaaS in Ihrem Unternehmen und müssten sich für einen IaaS-Anbieter entscheiden. Ihr bevorzugter Anbieter... *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne etwas ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
...hat eine sehr hohe Anzahl an Mitarbeitern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...hat einen sehr hohen Marktanteil.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...hat einen Ruf dafür, ein vertrauenswürdiger Partner zu sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ist bekannt dafür, um die Kunden bemüht zu sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...genießt ein hohes Ansehen im Markt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 3/15

3. Stimmen Sie zu oder lehnen Sie ab, dass die folgenden Qualitätsfaktoren Ihnen bei der Wahl Ihres IaaS-Providers sehr wichtig wären?

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne et- was ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Passende Service Level Agreements (SLA)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Unterstützung der Unternehmens-IT-Betriebsprozesse durch den Anbieter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Einhaltung von Sicherheits- und Compliance-Vorgaben	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Transparenz und Zweckmäßigkeit der angebotenen Preismodelle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Möglichkeit der individuellen Vertragsgestaltung (z.B. Vertragslaufzeiten, Vertragsart, Strafen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Qualität des Kundenservice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Service Level Agreements = Dienstgütereinbarungen mit dem IaaS-Anbieter über zugesicherte Leistungseigenschaften (bspw. Leistungsumfang, Reaktionszeit, Schnelligkeit der Bearbeitung)

Frage 4/15

4. Wie einfach oder schwierig wäre es Ihrer Meinung nach, einen IaaS-Anbieter gemäß folgender Kriterien zu überwachen? *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne et- was ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Die Überwachung der Leistungen des IaaS-Betriebs ist sehr schwer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Überwachung der Investitionen des IaaS-Anbieters in technische Innovationen ist sehr schwer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Überwachung der Investitionen des IaaS-Anbieters in seine eigene Mitarbeiterentwicklung ist sehr schwer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 5/15

5. Schätzen Sie das mögliche Verhalten von IaaS-Anbietern ein! *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne etwas ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Ich denke, dass der Anbieter möglicherweise vereinbarte und/oder informelle Übereinkünfte zu seinen Gunsten übertreten würde, wenn er die Chance hätte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, dass mir der Anbieter möglicherweise virtualisierte Server mit CPUs mit wechselnder Leistungsfähigkeit / älteren Festplatten / schlechterer Netzanbindung zur Verfügung stellen würde, wenn er die Chance hätte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mein Unternehmen fühlt sich sicher, wenn es den IaaS-Anbieter nutzt, um sensible Informationen zu verarbeiten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generell ist ein IaaS-Anbieter ein sicherer Ort, um vertrauliche Informationen hin zu senden und zu speichern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 6/15

6. Für unser Unternehmen ist die Nutzung von IaaS für Geschäftsanwendungen momentan und in naher Zukunft insgesamt eher: *

Bitte wählen Sie die zutreffende Antwort aus

	1	2	3	4	5	6	
Sehr unvorteilhaft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sehr vorteilhaft
Sehr unwichtig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sehr wichtig

Frage 7/15

7. Bitte bewerten Sie die Nutzungsabsicht von IaaS in Ihrem Unternehmen. *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne etwas ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Unser Unternehmen wird sich bemühen, IaaS regelmäßig für Geschäftsanwendungen zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen plant, innerhalb der nächsten zwei Jahre IaaS für Geschäftsanwendungen zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 8/15

8. Im Folgenden werden Institutionen/Personen aufgelistet, deren Meinung Ihnen oder Ihrem Unternehmen sehr wichtig sein könnte. Bitte bewerten Sie, ob diese Institutionen/Personen die Nutzung von IaaS für Ihr Unternehmen empfehlen oder ablehnen. *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne et- was ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Externe Beratungshäuser empfehlen die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vergleichbare Unternehmen empfehlen die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitarbeiter unserer IT-Abteilung empfehlen die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitarbeiter unserer Fachabteilungen empfehlen die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die IT-Fach-Presse empfiehlt die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ihr Vorgesetzter empfiehlt die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ihre Kollegen empfehlen die Nutzung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 9/15

9. Bitte bewerten Sie, ob Ihr Unternehmen fähig wäre, IaaS zu nutzen! *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne et- was ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Unser Unternehmen wäre fähig, IaaS zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Nutzung von IaaS ist vollständig unter der Kontrolle unseres Unternehmens.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen hat die Ressourcen, das Wissen und die Fähigkeiten, um IaaS zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 10/15

10 Um unsere IT-Infrastruktur zu beherrschen, verlangt unser Unternehmen, dass ein IaaS-Anbieter ... *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab 2	Lehne et- was ab 3	Stimme etwas zu 4	Stimme zu 5	Stimme voll und ganz zu 6
... eine beträchtliche Investition in Anlagen tätigen sollte, die auf unsere Bedürfnisse zugeschnitten sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... sich stark anstrengen sollte, seine Infrastruktur auf unsere Geschäftsanwendungen anzupassen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... auf unser Unternehmen spezialisiertes technisches Wissen besitzt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... auf unser Unternehmen spezialisiertes Geschäftswissen besitzt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 11/15

11. Stimmen Sie zu oder lehnen Sie ab, dass folgende Punkte schwer einzuschätzen sind bei einer Nutzung von IaaS-Ressourcen in Ihrem Unternehmen? *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab 2	Lehne et- was ab 3	Stimme etwas zu 4	Stimme zu 5	Stimme voll und ganz zu 6
Technischen Schwierigkeiten bei der Integration von IaaS-Ressourcen in die Unternehmenslandschaft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geschäftsbezogene Schwierigkeiten durch Preisänderungen von IaaS-Leistungen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geschäftsbezogene Schwierigkeiten durch Änderung prozesskritischer / operativer Leistungsindikatoren (z.B. unvorhergesehene Downtime bei kritischen Prozessen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rechtliche Schwierigkeiten durch externe Datenspeicherung (z.B. wegen der unsicheren rechtlichen Situation bezüglich Compliance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Auftreten eines Lock-In Effektes beim IaaS-Anbieter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kompatibilität von IaaS-Anbietern mit IT-Standards im Unternehmen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer des IaaS-Einführungsprojekts in unserem Unternehmen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kostenumfang der IaaS-Nutzung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Lock-In Effekt: hohe entstehende Kosten beim Wechsel des IaaS-Anbieters durch mangelnde Standards

Frage 12/15

12 Im Folgenden werden Sie gebeten, die relativen Vor- und Nachteile von Cloud-Infrastrukturdiensten gegenüber bisherigen Lösungen zu bewerten. Die Verwendung von IaaS ... *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne etwas ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
...führt zu mehr Innovation in unserem Unternehmen (z.B. die Erstellung von komplett neuen Produkten oder Diensten).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...schafft einen strategischen Mehrwert in unserem Unternehmen. (Der strategische Mehrwert bemisst Vorteile, die über die operationelle und taktische Ebene hinausgehen, indem er die Stellung einer Firma in einem Marktsegment beeinflusst.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...schafft einen Flexibilitätsmehrwert (in dem er unserem Unternehmen ermöglicht, flexibel auf unterschiedlichen Bedarf nach IT-Infrastruktur zu reagieren).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... schafft einen Effizienz Mehrwert in unserem Unternehmen (die Geschwindigkeit oder die Kosteneffizienz von Prozessen wird verbessert).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...führt zu einer erhöhten Effektivität in unserem Unternehmen bei der Leistungserstellung. (Effektivität bezieht sich auf die Verbesserung der Ergebnisqualität. Dies drückt sich durch die bessere Erreichung eines gegebenen Ziels oder die Ermöglichung von vorher unerreichbaren Zielen aus.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...führt zu einem organisatorischen Mehrwert in unserem Unternehmen. Ein organisatorischer Mehrwert umfasst die Möglichkeit, neue Organisationsformen durch den Einsatz von Cloud-Infrastrukturdiensten zu nutzen.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ist technisch kompatibel zur bestehenden IT-Infrastruktur unseres Unternehmens.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ist technisch kompatibel zu bestehenden Softwareapplikationen unseres Unternehmens.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ist kompatibel zu unseren bestehenden IT-Prozessen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 13/15

13 Bitte bewerten Sie die folgenden Aussagen zu Ihren bisherigen Erfahrungen mit der IaaS-Nutzung. *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab 2	Lehne et- was ab 3	Stimme etwas zu 4	Stimme zu 5	Stimme voll und ganz zu 6
Unserem Unternehmen fällt das Erlernen des Umgangs mit IaaS leicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unserem Unternehmen fällt es leicht, IaaS wie beabsichtigt einzusetzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vor der Entscheidung, IaaS überhaupt zu nutzen, könnte unser Unternehmen es richtig ausprobieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen könnte IaaS versuchsweise lange genug nutzen, um beurteilen zu können, was man damit machen kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die IaaS-Nutzung ist nicht sehr sichtbar in unserem Unternehmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In unserem Unternehmen sieht man die IaaS-Nutzung in vielen Bereichen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen hätte keine Schwierigkeit, anderen von den Ergebnissen der IaaS-Nutzung zu erzählen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Ergebnisse der IaaS-Nutzung erscheinen unserem Unternehmen offenkundig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen hätte keine Schwierigkeiten zu erklären, warum IaaS vorteilhaft sein könnte oder nicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Frage 14/15

Gleich haben Sie es geschafft. Bitte beantworten Sie nur noch wenige Fragen zu Ihnen und Ihrem Unternehmen. Wie würden Sie Ihre Rolle im Unternehmen beschreiben?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- CIO/CTO CIO/CTO
- CSO/CISO CSO/CISO
- Übergeordnete IT-Leitungsfunktion Übergeordnete IT-Leitungsfunktion
- IT-Manager / Teamleiter / Projektleiter IT-Manager / Teamleiter / Projektleiter
- IT-Spezialist Infrastruktur IT-Spezialist Infrastruktur
- IT-Spezialist Anwendungen IT-Spezialist Anwendungen
- CEO/Präsident/Eigner/Partner/COO CEO/Präsident/Eigner/Partner/COO
- CFO/Finanzleiter/ Übergeordnete Finanz-Leitungsfunktion CFO/Finanzleiter/ Übergeordnete Finanz-Leitungsfunktion
- Geschäftliche Leitungsfunktion (Manager, Teamleiter) Geschäftliche Leitungsfunktion (Manager, Teamleiter)
- Spezialist in Funktionalabteilung (Einkauf, Fertigung, Vertrieb etc.) Spezialist in Funktionalabteilung (Einkauf, Fertigung, Vertrieb etc.)

In welcher Branche ist Ihr Unternehmen tätig?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- Informationstechnologie und Telekommunikation Informationstechnologie und Telekommunikation
- Versorgungsunternehmen Versorgungsunternehmen
- Gemeinnützige Organisationen Gemeinnützige Organisationen
- Institutionen und öffentlicher Sektor (einschließlich Bildungssektor) Institutionen und öffentlicher Sektor (einschließlich Bildungssektor)
- Dienstleistungen (Recht, Beratung, Immobilien) Dienstleistungen (Recht, Beratung, Immobilien)
- Fertigung (inkl. Automobilbau, Chemie, Bauwesen, Maschinenbau, usw.) Fertigung (inkl. Automobilbau, Chemie, Bauwesen, Maschinenbau, usw.)
- Finanzdienstleistungen (Banken, Versicherungen) Finanzdienstleistungen (Banken, Versicherungen)
- Gesundheitswesen (Einrichtungen und Pharmahersteller) Gesundheitswesen (Einrichtungen und Pharmahersteller)
- Einzelhandel, Großhandel und Vertrieb Einzelhandel, Großhandel und Vertrieb
- Transportwesen (Fluglinien, Eisenbahnen, Schiffsverkehr, Logistik) Transportwesen (Fluglinien, Eisenbahnen, Schiffsverkehr, Logistik)
- Baugewerbe Baugewerbe

Die Nutzung von IaaS für Geschäftsanwendungen hat in den letzten drei Jahren in Ihrem Unternehmen stark zugenommen?

Bitte wählen Sie die zutreffende Antwort aus:

	1	2	3	4	5	6	
Lehne voll und ganz ab	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Stimme voll und ganz zu

Der geschätzte Prozentsatz des für IaaS zugeteilten IT-Budgets am Gesamt-IT-Budget - in Ihrem Unternehmen in 2009 liegt bei (Angaben freiwillig):

Bitte schreiben Sie Ihre Antwort hier

Wie viele Mitarbeiter hat Ihr Unternehmen? (Angaben freiwillig)

Bitte wählen Sie nur eine der folgenden Antworten aus:

- 0 - 9
- 10 - 49
- 50 - 249
- 250 - 999
- 1000 - 5000
- >5000

Wie hoch war der Umsatz Ihres Unternehmens im Jahr 2009? (Angaben freiwillig)

Bitte wählen Sie nur eine der folgenden Antworten aus:

- < 0.5 Mio.
- 0.5 - 1 Mio.
- 1 - 2 Mio.
- 2 - 10 Mio.
- 10 - 50 Mio.
- 50 - 100 Mio.
- > 100 Mio.

Frage 15/15

Wie schätzen Sie die Risikoneigung der Firmen-IT-Verantwortlichen Ihres eigenen Unternehmens ein? *

Bitte wählen Sie die zutreffende Antwort aus

	Lehne voll und ganz ab	Lehne ab	Lehne et- was ab	Stimme etwas zu	Stimme zu	Stimme voll und ganz zu
	1	2	3	4	5	6
Unser Unternehmen scheint eher e- ne-konservative Haltung bei wichti- gen Entscheidungen einzunehmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unser Unternehmen unterstützt Pro- jekte eher, wenn der erwartete Ge- winn oder Return on Investment si- cher ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Der Geschäftsbetrieb folgte bisher im Allgemeinen getesteten und für gut befundenen Wegen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Hier können Sie noch Ihre Kommentare zur Umfrage eintragen.

Bitte beachten Sie: Erst im nächsten Schritt werden Sie zu einem separaten System weitergeleitet, in das Sie Ihre E-Mail-Adresse für die Auswertung und die Verlosung eingeben können.

Appendix C

Interview Guidelines - Case Study

As the interviews were conducted in German, the following interview guidelines were also formulated in German:

Leitfaden Experten-Interview Organisationelle Anforderungen zur Integration von Grid-Computing

1. Einführung

- a. Haben Sie bereits Erfahrungen mit dem Einsatz von Grid-Technologien gemacht? Wenn ja, welche?
- b. Welche (anderen) Anwendungsszenarien für Grid Computing sind Ihnen bekannt?
ERP-Software (z.B. BMW: emPlant - Simulation von Produktionsabläufen)
CAE-Anwendungen (z.B. Batch Meshing)
SAP-Szenario
Data Mining/Storage

2. Organisation

- a. Wie wird über den Einsatz von Grid Computing entschieden? Wo liegt die Verantwortlichkeit für die Entscheidung zum Einsatz von Grid Computing (auch: Wer entscheidet)? Wie sieht der Prozess für das Technologie Assessment aus? Wer ist alles beteiligt? Was sind die Kriterien, die darüber entscheiden, ob Grid Computing eingesetzt wird?
- b. Wer ist verantwortlich/bestimmt über Zugriff, Accounting und Verrechnung? Steuert eine einzelne Abteilung den Zugriff auf alle Knoten? Wer misst/erfasst/verarbeitet/verteilt die Jobs? Wer bestimmt über die Preise in einem internen Verrechnungsmodell? Durch wen und wie werden etwaige Rechenbudgets festgelegt?
- c. Welche organisatorischen Veränderungen sind notwendig, um erfolgreiches Ressourcen-Sharing zu etablieren? Wie können Betriebsverantwortliche für die Business-Anwendungen bzw. IT-Projekte dazu gebracht werden, "ihre" unausgelasteten Server zu teilen? Wie könnten Anreize dafür aussehen? Was muss man verändern, um Ressourcenteilung zu begünstigen?
- d. Entstehen organisatorische Konsequenzen aus der Dezentralisierung der Dienste? Führt Grid Computing zu geänderten Kostenstrukturen in der IT? Findet eine Umschichtung der Etats aus Invest-Bereichen (Infrastruktur) in Budget-Bereiche (Dienstleistungen) statt? Wie werden diese wahrgenommen (z.B. "Kontroll"-Verlust o.Ä.)? Entstehen andere Konsequenzen aus der Entscheidung?

- e. Wurden bereits Erfahrungen mit Outsourcing gemacht (welche und würde man Dienste von "bekannten Anbietern" bevorzugen)? Wenn ja, was waren dabei die organisatorischen Hemmnisse? Wie wurden diese beseitigt? Wie wichtig ist es im Bezug auf einen Einsatz von Grid Computing, dass bereits Erfahrungen mit einem potentiellen Anbieter gemacht wurden?

3. Sicherheit

- a. Würde man Daten auf einem Knoten verarbeiten, den man nicht 100% unter Kontrolle hat? Gibt es Maßnahmen, die die Daten vor einem unbefugten Zugriff anderer Nutzer des Knoten schützen? Wie sehen diese aus? Was sind typische Methoden der Zugriffskontrolle? Wie kann man einen unbefugten Zugriff verhindern? (Verschlüsselung, VMs in separaten VLANs pro Kunde) Gibt es Daten die man selbst auf einem innerbetrieblichen Knoten, den man nicht 100%ig unter Kontrolle hat, nicht rechnen würde?
- b. Wie kann der Missbrauch von Daten (insbesondere bei Bezug der Leistungen durch einen externen Anbieter) wirksam ausgeschlossen werden? Wie sicherheitskritisch/vertraulich sind Daten? Wie sehen die Sicherheitsanforderungen an Daten darüber hinaus aus?
Für welche Arten von Daten ist man bereit, das latente Risiko zu akzeptieren?
- c. Existiert eine Gefährdung interner Systeme? (können Schwachstellen ausgenutzt werden? Ermöglicht bidirektionale Kommunikation im Grid den Zugriff auf Rechner die eigentlich geschützt sein müssten/sollten?
Kann sichergestellt werden, dass interne Systeme nur auf den spezifizierten Schnittstellen erreichbar sind?
Für welche Form von Systemen ist bereit, eine solche Gefährdung in Kauf zu nehmen?
- d. Wie wichtig ist Vertrauen zu anderen Partnern/Abteilungen? Kann man es messen, erfassen und wo möglich zu anderen Nutzern kommunizieren?
Was sind Indikatoren aus, die Vertrauen schaffen (Sicherheits-Audit, Zertifikate, Referenzen etc.)?
Wo kann man solche Indikatoren auffinden?
Wie könnte ein Mechanismus zur Bewertung aussehen?

4. Wissen

- a. Umfasst der Begriff Grid-Computing für die meisten Nutzer/Entscheidungsträger nur die Bereitstellung einer großen Menge an Rechenleistung?
Wie wird Grid von den Nutzern verstanden? Wie von Leuten, die über den Einsatz oder Verzicht auf Grid entscheiden?
Welche Vorteile kann man kommunizieren um das Grid schmackhafter zu machen?
Welche anderen Einsatzszenarien existieren und sind diese außerhalb der IT-Abteilung/Experten bekannt?
- b. Können die Anforderungen der Nutzer an die Anwendungen im Grid formalisiert und umgesetzt werden?
Was sind die typischen Anforderungen von Nutzern an die IT?
Sind diese auf einem Grid überhaupt möglich?
- c. Sind Anwendungsfälle komplex genug, damit es sich überhaupt lohnt sie auf dem Grid zu verarbeiten?
Wie sieht ein typischer Job für die Verarbeitung auf dem Grid aus?
Kann es sich auch lohnen, Jobs die permanenter Steuerung bedürfen, auf dem Grid zu verarbeiten?

- d. Würden sie Rechenjobs in ein offenes Grid vergeben, wenn es über ausreichende Sicherheitsmechanismen verfügt? Wie müssen diese Mechanismen aussehen? Ist eine Virtualisierungslösung ausreichend? Ist der generierte Nutzen ausreichend, um das Restrisiko akzeptieren zu können?

5. Kosten/Nutzen

- a. Führt der Einsatz von Grid-Technologien zu einer Kostenreduktion?
Führt der Einsatz von Grid-Technologien zu einer Kostenreduktion bei den IT-Kosten?
Sind diese abschätzbar?
Wo sehen Sie Sparpotentiale?
Sind die Kosten, die durch die Einführung von Grid Computing entstehen planbar?
- b. Was für einen Nutzen hat der Einsatz von Grid Technologien?
Gibt es einen Nutzen für das Unternehmen über eine Kostenreduktion hinaus?
Wie groß ist der Nutzen aus einem flexibleren Einsatz der Ressourcen?
Wenn ja, welchen Nutzen kann man zusätzlich aus dem Einsatz generieren?
- c. Wie kann man den Nutzen messen? Ist der Nutzen einfach vorab kalkulierbar oder später messbar?
Wie sieht generell das Innovations-Assessment aus? Welche Faktoren werden zur Messung des Nutzens herangezogen?
Inwiefern kann es hemmend sein, dass man den Nutzen nicht abschätzen kann?

6. Anwendung

- a. Kann eine Grid-Infrastruktur überhaupt ausreichend genutzt werden um Anwendung/Einführung zu rechtfertigen?
Besteht aufgrund der ständigen Zunahme von Rechenleistung überhaupt noch ein Bedarf an der Rechenleistung eines Grids (insbesondere emPlant und CAE)?
Gibt es genug Anwendungen, die auf das Grid portiert werden könnten?
- b. Wie sehen Sie die Chancen, bestehende Lizenzvereinbarungen dahingehend abzuändern, so dass diese eine flexible/dynamische Verwendung im Grid ermöglichen?
Sind die aktuellen Lizenzmodelle auf einen verteilten Einsatz im Grid anwendbar?
Lassen sich die Lizenzmodelle leicht abändern, um eine Verwendung von jedem Knoten für jede mögliche Task zu ermöglichen?
Welche typischen Einschränkungen der Lizenznutzung existieren, die den Einsatz bestimmter Grid-Provider verhindern (z.B. Gültigkeit der Lizenz auf bestimmte Ressourcen oder räumliche Beschränkungen, z.B. für Corporate Lizenz, die nur auf dem Grundstück des Unternehmens gilt)?

Appendix D

Interview Guidelines - Explorative Study

Tabelle D.1: Interview guideline (Heinle 2010)

Offengelegt Frage	Nachfragen	Erwartungen
1. Wie verstehen Sie „Cloud Computing“ / „Infrastructure as a Service“?	Schichtenmodell	schnell skalierende virtuelle Infrastruktur (IaaS: Amazon EC2)
	Kennen Sie Szenarien für die Anwendung von Cloud Computing?	
2. Wie und durch wen wird über Cloud Computing informiert?	Informieren Anbieter umfassend?	Information hauptsächlich für IT Abteilungen
	Information von Geschäftspartner / Consulting?	
3. Würden bereits Erfahrungen mit dem Outsourcing von Infrastruktur gemacht?	Wie wird über den Einsatz von Cloud Computing entschieden?	Outsourcing von Telefonanlage, Drucker, Desktop,...
	Wer entscheidet über den Einsatz neuer Technologien? Kriterien?	
	Konnten die Erwartungen erfüllt werden?	
	Was waren eventuelle Probleme? Lösungen?	
4. Würden Sie zum aktuellen Zeitpunkt Infrastructure as a Service Angebote nutzen?	Einsatz (relativ) neuer / junger Technologien?	
	Entscheidende Faktoren?	
	Instabiler Marktlage	
	Geringe Marktreife	
	Könnte Cloud Computing zum aktuellen Zeitpunkt eingesetzt werden?	
5. Welche Eigenschaften müsste ein potentieller Cloudanbieter aufweisen?	Welches ist das bevorzugte Abrechnungsmodell?	Grösse, Reputation, Referenzen, zusätzliche Angebote, Support
	Wie findet die Risikobewertung statt?	
	Wie wird Vertrauen in Anbieter gemessen?	
6. Welche kritischen Punkte sehen Sie vor dem Bezug von Cloud Computing Diensten?	Preistransparenz?	
	Bedarfsplanung?	
	Migration von Anwendungen?	
	Legacy Dienste?	
	Abrechnung von Cloud Diensten?	
	Wer darf Cloud Dienste einkaufen?	
7. Welche kritischen Punkte sehen Sie während des Bezugs von Cloud Computing Diensten?	Sind die angebotenen Leistungen im Bereich Monitoring und Reporting ausreichend?	Mangelndes Reporting und Monitoring begünstigt durch fehlende Schnittstellen, Gewährleistung der Einhaltung von SLAs nicht ausreichend gegeben, Daten nicht sicher, mangelhafte Standardisierung
	Überwachung durch fehlende Schnittstellen?	
	Sehen Sie Probleme bei der Verfügbarkeit der Services?	
	Einhaltung der SLAs?	
	Entschädigung bei Ausfällen?	
	Ist die Sicherheit der Daten in der Cloud ausreichend gewährleistet?	
	Ist der Support von Anbieterseite ausreichend?	
	Wie wichtig ist Standardisierung?	
Sehen Sie die Gefahr des Lock-In?		
8. Welche Auswirkungen hat der Einsatz von Cloud Computing auf Unternehmen?	Veränderung	
	Erhöhte Kostentransparenz	
9. Welche Chancen und Risiken sehen Sie beim Einsatz von Cloud Computing?	Wettbewerbsvorteile durch den Einsatz von Cloud Computing?	Kostensenkung, Flexibilitätssteigerung, Rechtliche Probleme, Compliance, Sicherheit der Daten, Standort der Daten, Zukunft der Technologie
	Risiken für Daten in der Cloud?	
	Compliance?	
	Integration in bestehende Organisation / Abläufe?	

Appendix E

Qualitative Survey Feedback

The questionnaire described in Appendix B contained a field, where the informants could leave a comment. These comments were also analyzed and paint an interesting picture of the participant's perception of this survey. As the comments field was at the end of the questionnaire, the comments must come from informants that finished the questionnaire up to that point. Table E.1 gives a subjective evaluation of the comments' sentiment. Negative remarks make up the greatest share, which is understandable, as the comments field gives informants a way to vent their frustration. However, there are some explicitly positive remarks as well.

Table E.2 lists the most common topics in these negative remarks. Poor wording of questions and missing aspects (both formal and content-related) were mentioned most. In general, the comments contained IaaS-related key words, that were indexed in Table E.3. The aggregation of the actual wording to the topics in the table required some interpretation of the comment itself, so this summary is necessarily subjective. Even though, client-side risks and data security are two topics that stand out among all comments; these two seem to concern participants most, when they think about IaaS adoption.

Table E.1: Sentiment of Feedback

Type of Feedback	Count	%
Positive Feedback	6	17%
Negative Feedback	18	50%
General Feedback	12	33%
Number of comments	36	100%

Table E.2: Negative Feedback

Negative Feedback	Count	%	Examples
Question concerning Data from 2009	2	5.6%	
Misunderstandings	2	5.6%	Contact data, Private Cloud
Wording of questions	7	19.4%	crude translations
Missing aspect	7	19.4%	missing neutral answering option, distinction between IaaS adopters and non-adopters, public sector

Table E.3: Topics mentioned in the Feedback

Topics mentioned	Count	%
Client-side risk (e.g. incomplete contracts, trust in provider, reliability, continuity, lock-in effects)	7	21.9%
Data security	5	15.6%
SLA-related topics	3	9.4%
Suitable business processes (also necessary adaptation)	3	9.4%
IaaS in banks	2	6.3%
IaaS drawbacks	2	6.3%
Standardization (application, infrastructure)	2	6.3%
Peak load coverage	1	3.1%
Size of the provider	1	3.1%
Individual service selection	1	3.1%
Compatibility	1	3.1%
SaaS/PaaS	1	3.1%
Visibility	1	3.1%

Appendix F

Amazon AWS Cost Figures

This section documents the cost figures used in the decision support model evaluation; the data was copied from the Amazon AWS Elastic Compute Cloud (EC2) Web site <http://aws.amazon.com/de/ec2/> in the beginning of August 2012. The calculation of the yearly cost figures for the cost optimization model was done using the Amazon Monthly Calculator.¹ All Amazon prices are excluding discounts for first-time customers.

Table F.1 lists the prices of Amazon AWS EC2 compute instances (in US Dollars). These are the prices of the most general instance types, which are used in the evaluation; specialized instance types like micro instances were not considered. All instances are online 100% of the time, and are charged accordingly.

Table F.1: Amazon AWS Compute Instances

VM type	Instance	RAM in GB	EC2 Cores	On-demand instance (\$ per month)	On-demand instance (\$ per year)	Reserved instance (\$ per year)
m1.small		1.7	1.0	87.84	1054.08	642.52
m1.medium		3.8	2.0	168.36	2020.32	1179.56
m1.large		7.5	4.0	336.72	4040.64	2044.24
m1.xlarge		15.0	8.0	600.24	7202.88	3782.24
m2.xlarge		17.1	6.5	443.59	5323.08	2781.28
m2.2xlarge		34.2	13.0	814.72	9776.64	5256.32
m2.4xlarge		68.4	26.0	1550.50	18606.00	10206.40
NEWM2.8xlarge		136.8	52.0	3000.00	36000.00	20412.80

As storage, Amazon Elastic Block Storage (EBS) is added to the total cost. The tariff per GB is 0.11\$ per month and 1.43\$ per year. In the Amazon Cloud, storage costs are a combination of the actual storage used over time and the number of input/output operations on this storage. As the BMW server monitoring data does not include IO operations, a realistic number for IO operations has to be estimated. To this end, the SAPS per EC2 Compute Unit are estimated to 531.91, based on the BMW infrastructure documentation (BMW Group 2012b) and on the definition of SAPS², which is a hardware-independent benchmark for SAP systems. According to Engelbart (2011), the IO operations per second (IOPS) depend on the SAPS of a system; for each unit of SAPS, 0.4 IOPS are needed. Combining these numbers with the fact that Amazon AWS charges 0.11\$ per 1 Million IO requests, the cost figures for IO operations can be derived (Table F.2). The IO costs are depending on the system size, as it is assumed that a fast system can put out more IO

¹<http://calculator.s3.amazonaws.com/calc5.html>, last accessed 2013-02-10

²http://www.sap.com/campaigns/benchmark/bob_glossary.epx, last accessed 2013-12-29

requests than a slow one. However, Amazon acknowledges, that standard EBS volumes only deliver around 100 IOPS (Amazon 2012), hence the IOPS in the table are capped.

Table F.2: Amazon AWS EBS IO Tariff

UCL	IOPS	IO Count per month	IO cost (\$ per month)
1.00	21.28	40442553.19	4.45
2.00	42.55	80885106.38	8.90
4.00	85.11	161770212.77	17.79
8.00	170.21	323540425.53	35.59
6.50	170.21	323540425.53	35.59
13.00	170.21	323540425.53	35.59
26.00	170.21	323540425.53	35.59
52.00	170.21	323540425.53	35.59

Network costs must be distinguished between LAN and WAN costs. LAN tariffs are those charges that Amazon AWS as an IaaS provider bills for data traffic that flow in (inbound) and out (outbound) of its Cloud. Table F.3 lists these two tariffs. WAN charges have to added on top, as the data traffic has to first reach the Amazon Cloud using regular Internet connections. An exemplary offer for a medium WAN uplink at a German provider was used for estimating the WAN charges.³ They turnout to be 0.15\$ per month per GB per direction, which is comparable to the figures given in (Gray 2003). Amazon AWS also charges for a dedicated VPN connection to one of their data centers, but this cost factors is negligibly small.

Table F.3: Amazon AWS Network Tariff

Network connection	Cost (\$ per month per GB)	Cost + WAN charge	Total Cost (\$ per year per GB)
Inbound	0.00	0.15	1.95
Outbound	0.12	0.27	3.51

The tariffs for all infrastructure cost types (compute, storage, network), that are used in this evaluation, are known. In order to make the Amazon AWS infrastructure services comparable to traditional industry IT in-house services, the surcharge for support operations has to be included. In the case of AWS, these surcharges depend on the client revenue.⁴ For this study, pure AWS usage would cost approx. 300 000\$ per year, or ca. 25 000\$ per month. Table F.4 shows how to calculate the actual support charges. For this study, the surcharge turns out to be 8.2% of the total revenue, hence all AWS prices were increased by this rate.

Table F.4: Amazon AWS Support Tariff

Tariff breaks	Usage cost (\$ per month)	Support rate	Support cost (\$ per month)
10% of monthly AWS usage for the first \$0-\$10K	10000	10%	1000
7% of monthly AWS usage from \$10K-\$80K	15000	7%	1050
5% of monthly AWS usage from \$80K-\$250K			
3% of monthly AWS usage from \$250K+			
Total	25000		2050

³<http://www.m-net.de/geschaeftskunden/internet/sdsl/tarife.html>, last accessed 2012-08-01

⁴<http://aws.amazon.com/de/premiumsupport/#pricing>, last accessed 2013-12-29

Appendix G

Quality Model Mapping

The columns represent the e-Service quality model by Modheji (2010). The rows are show the quality dimensions of the extended IaaS quality model. The quality dimension “Transparency and Practicability of provider tariffs” has only limited support in the research literature and is therefore not continued in the new model.

Table G.1: Quality Model Mapping

	Appropriate Service Level Agreements	Support of corporate IT operations processes by the provider	Adherence to Security and Compliance regulations	Transparency and Practicability of provider tariffs	Possibility of drafting individual contracts (e.g. duration, type, penalties...)	Quality of Customer Service
System-related	Operational time	•				
	Performance	•				
	Availability	•				
	Stability	•				
	Recoverability	•				
	Security	•				
	Standardization	•				
	Product features					•
	Service time	•				
	Performance	•				
Operations-related	Reliability (on-time delivery)		•			
	Empathy					•
	Assurance			•		

Appendix H

Results of Decision Tree Parameter Optimization

Table H.1 lists the 71 (out of 374) parameter combinations that yielded the best performance on the IaaS placement classification problem. The table shows the full combinatorics of the factor levels for the parameters “Minimal size for split”, “Min. Leaf size” and “Confidence”. It is obvious that the decision tree performance is not sensitive to variations of these parameters under the given factor levels. However, “Minimal gain” is a decisive factor; the best performance is only reached for a value of 0.1. All other 303 parameter combinations had other values for “Minimal gain” and fared worse in terms of performance.

Table H.1: Sensitivity analysis of decision tree parameters

<u>Minimal size for split</u>	<u>Min Leaf size</u>	<u>Minimal gain</u>	<u>Confidence</u>	<u>Performance</u>
9.0	2.0	0.1	0.2	0.888
12.0	2.0	0.1	0.2	0.888
15.0	2.0	0.1	0.2	0.888
3.0	4.0	0.1	0.2	0.888
6.0	4.0	0.1	0.2	0.888
9.0	4.0	0.1	0.2	0.888
12.0	4.0	0.1	0.2	0.888
15.0	4.0	0.1	0.2	0.888
3.0	6.0	0.1	0.2	0.888
6.0	6.0	0.1	0.2	0.888
9.0	6.0	0.1	0.2	0.888
12.0	6.0	0.1	0.2	0.888
15.0	6.0	0.1	0.2	0.888
3.0	8.0	0.1	0.2	0.888
6.0	8.0	0.1	0.2	0.888
9.0	8.0	0.1	0.2	0.888
12.0	8.0	0.1	0.2	0.888
15.0	8.0	0.1	0.2	0.888
3.0	10.0	0.1	0.2	0.888
6.0	10.0	0.1	0.2	0.888
9.0	10.0	0.1	0.2	0.888
12.0	10.0	0.1	0.2	0.888
15.0	10.0	0.1	0.2	0.888
9.0	2.0	0.1	0.3	0.888
12.0	2.0	0.1	0.3	0.888
15.0	2.0	0.1	0.3	0.888
3.0	4.0	0.1	0.3	0.888
6.0	4.0	0.1	0.3	0.888

9.0	4.0	0.1	0.3	0.888
12.0	4.0	0.1	0.3	0.888
15.0	4.0	0.1	0.3	0.888
3.0	6.0	0.1	0.3	0.888
6.0	6.0	0.1	0.3	0.888
9.0	6.0	0.1	0.3	0.888
12.0	6.0	0.1	0.3	0.888
15.0	6.0	0.1	0.3	0.888
3.0	8.0	0.1	0.3	0.888
6.0	8.0	0.1	0.3	0.888
9.0	8.0	0.1	0.3	0.888
12.0	8.0	0.1	0.3	0.888
15.0	8.0	0.1	0.3	0.888
3.0	10.0	0.1	0.3	0.888
6.0	10.0	0.1	0.3	0.888
9.0	10.0	0.1	0.3	0.888
12.0	10.0	0.1	0.3	0.888
15.0	10.0	0.1	0.3	0.888
3.0	2.0	0.1	0.4	0.888
6.0	2.0	0.1	0.4	0.888
9.0	2.0	0.1	0.4	0.888
12.0	2.0	0.1	0.4	0.888
15.0	2.0	0.1	0.4	0.888
3.0	4.0	0.1	0.4	0.888
6.0	4.0	0.1	0.4	0.888
9.0	4.0	0.1	0.4	0.888
12.0	4.0	0.1	0.4	0.888
15.0	4.0	0.1	0.4	0.888
3.0	6.0	0.1	0.4	0.888
6.0	6.0	0.1	0.4	0.888
9.0	6.0	0.1	0.4	0.888
12.0	6.0	0.1	0.4	0.888
15.0	6.0	0.1	0.4	0.888
3.0	8.0	0.1	0.4	0.888
6.0	8.0	0.1	0.4	0.888
9.0	8.0	0.1	0.4	0.888
12.0	8.0	0.1	0.4	0.888
15.0	8.0	0.1	0.4	0.888
3.0	10.0	0.1	0.4	0.888
6.0	10.0	0.1	0.4	0.888
9.0	10.0	0.1	0.4	0.888
12.0	10.0	0.1	0.4	0.888
15.0	10.0	0.1	0.4	0.888

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2), 179 – 211. Theories of Cognitive Self-Regulation.
- Ajzen, I. (2010). Constructing a theory of planned behavior questionnaire: Conceptual and methodological considerations. Online [last accessed 2010-12-15]. <http://www.people.umass.edu/ajzen/pdf/tpb.measurement.pdf>.
- Almeida, J., V. Almeida, D. Ardagna, C. Francalanci, and M. Trubian (2006, June). Resource management in the autonomic service-oriented architecture. In *IEEE International Conference on Autonomic Computing, 2006. ICAC '06.*, pp. 84–92.
- Altmann, J., C. Courcoubetis, J. Darlington, and J. Cohen (2007). GridEcon - The Economic-Enhanced Next-Generation Internet. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Berlin, pp. 188–193. Springer-Verlag.
- Altmann, J., M. Ion, and A. A. B. Mohammed (2007). Taxonomy of Grid Business Models. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Berlin, pp. 29 – 43. Springer-Verlag.
- Amazon (2012, Dec.). Amazon Elastic Compute Cloud - User Guide. Online [last accessed 2013-02-11]. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AmazonEBS.html>.
- Anderson, J. C. and D. W. Gerbing (1991, Oct.). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology* 76(5), 732 – 740.
- Armbrust, M., A. Fox, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia (2009, Feb.). Above the Clouds: A Berkeley View of Cloud Computing. Technical Report UCB/EECS-2009-28, University of California at Berkeley.
- Attelander, P. (2008). *Methoden der empirischen Sozialforschung* (12th ed.). Berlin: Erich Schmidt Verlag.
- Avanade Inc. (2009, Oct). Cloud Computing Studie. Online. http://www.avanade.com/de/_uploaded/pdf/pressrelease/200910cloudcomputingstudieii854656.pdf [last accessed 2010-05-13].
- Backhaus, K., B. Erichson, W. Plinke, and R. Weiber (2006). *Multivariate Analysemethoden* (11th ed.). Berlin: Springer Verlag.
- Bagozzi, R. and Y. Yi (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science* 16, 74–94.
- Bagozzi, R. P. (1982, Nov.). A field investigation of causal relations among cognitions, affect, intentions, and behavior. *Journal of Marketing Research* 19(4), 562 – 583.

- Baier, M. (2008). Die Diffusion von Innovationen im Markt managen: Fallstudie zur Nutzung von Grid-Technologien für betriebliche Informationssysteme (BIS). In M. Bichler, T. Hess, H. Krcmar, U. Lechner, F. Matthes, A. Picot, B. Speitkamp, and P. Wolf (Eds.), *Multikonferenz Wirtschaftsinformatik 2008*, Berlin, pp. 104 – 114. GITO-Verlag.
- Baier, M., G. Gräfe, M. Jekal, F. Röhr, and T. Vörckel (2009, Jul). BIS-Grid Marktanalyse für kommerzielles Grid-Providing - Marktstudie Stand Juli 2009. Online. https://bi.offis.de/bisgrid/tiki-download_file.php?fileId=299 [last accessed 2010-04-19].
- Baldi, P., S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics Review* 16(5), 412 – 424.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management* 17(1), 99 – 120.
- Baron, R. M. and D. A. Kenny (1986, Dec.). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51(6), 1173–1182.
- Barroso, L. A. and U. Hölzle (2009). The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture* 4(1), 1–108.
- Barrutia, J. M. and A. Gilsanz (2009). e-service quality: overview and research agenda. *International Journal of Quality and Service Sciences* 1(1), 29 –50.
- Barthélemy, J. and D. Adsit (1993, May). The seven deadly sins of outsourcing. *The Academy of Management Executive* 17(2), 87 – 100.
- Bazijanec, B., K. Pousttchi, and K. Turowski (2004). An approach for assessment of electronic offers. In M. Núñez, Z. Maamar, F. Pelayo, K. Pousttchi, and F. Rubio (Eds.), *Applying Formal Methods: Testing, Performance, and M/E-Commerce*, Volume 3236 of *Lecture Notes in Computer Science*, pp. 44–57. Springer Berlin / Heidelberg.
- Bégin, M.-E. (2008, May). An EGEE comparative study: Grids and Clouds - Evolution or Revolution? Technical Report 925013, EGEE-II collaboration. <https://edms.cern.ch/file/925013/3/EGEE-Grid-Cloud.pdf>, [last accessed 2008-10-14].
- Beims, M. (2012). *IT-Service Management in der Praxis mit ITIL* (3rd ed.). München: Carl Hanser Verlag.
- Benamati, J. and T. M. Rajkumar (2002). The application development outsourcing decision: an application of the technology acceptance model. *Journal of Computer Information Systems* 42(4), 35–43.
- Benlian, A. and P. Buxmann (2009). Treiber der Adoption SaaS-basierter Anwendungen - Eine empirische Untersuchung auf Basis verschiedener Applikationstypen. *Wirtschaftsinformatik* 5, 414 – 428.
- Bensaou, M. and E. Anderson (1999). Buyer-supplier relations in industrial markets: When do buyers risk making idiosyncratic investments? *Organization Science* 10(4), 460 – 481.
- Benton, W. C. (1991). Quantity discount decisions und conditions of multiple items, multiple suppliers and resource limitations. *International Journal of Production Research* 29(10), 1953–1961.
- Berger, T. (2005). *Konzeption und Management für Service-Level Agreements für IT-Dienstleistungen*. Ph. D. thesis, Technische Universität Darmstadt, Darmstadt.
- Bichler, M., T. Setzer, and B. Speitkamp (2006). Capacity planning for virtualized servers. In *Workshop on Information Technologies and Systems (WITS)*, Milwaukee, Wisconsin, USA.

- BITKOM e.V. (2010, Jan). IT- und Telekommunikations-Trends 2010. Online. http://www.bitkom.org/files/documents/BITKOM-Presseinfo_IT-Trends_2010_-_13_01_2010.pdf [last accessed 2010-05-05].
- Bliemel, F., A. Eggert, G. Fassott, and J. Henseler (2005). Die PLS-Pfadmodellierung: Mehr als eine Alternative zur Kovarianzstrukturanalyse. In F. Bliemel, A. Eggert, G. Fassott, and J. Henseler (Eds.), *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele*, pp. 9 – 16. Stuttgart: Schäffer-Poeschel.
- Blum, M., R. W. Floyd, V. Pratt, R. L. Rivest, and R. E. Tarjan (1973). Time bounds for selection. *Journal of Computer and System Sciences* 7(4), 448 – 461.
- BMW Group (2007, Sep.). Von der Idee zum Kunden - Vortrag Entwicklung und Einkauf. Internal document.
- BMW Group (2008a, Jan.). IT-Innovationsmanagement (IT-IM). Internal document.
- BMW Group (2008b, Jan.). STARD LS3. Internal document.
- BMW Group (2011). *A Company in Its Time*. Munich, Germany: BMW Group. http://www.bmwgroup.com/e/0_0_www_bmwgroup_com/unternehmen/publikationen/_ebook/EBook_Stationen_en/index.html [last accessed 2014-08-15].
- BMW Group (2012a, Jan.). Infrastructure price table 2012. Internal document.
- BMW Group (2012b, Jan.). Leistungskatalog Applikationsbetrieb und IT-Infrastruktur 2012. Internal document.
- BMW Group (2012c). Statement of Work - Los NAS. Internal document.
- BMW Group (2013). *Annual Report 2013*. Munich, Germany: BMW Group. http://www.bmwgroup.com/e/0_0_www_bmwgroup_com/investor_relations/finanzberichte/geschaeftsberichte/2013/_pdf/report2013.pdf [last accessed 2014-08-15].
- BMW Group (2014a, Mar.). BMW Group Prozessportal. Internal document.
- BMW Group (2014b, Apr.). IT Functional Strategy 2014. Internal document.
- Bongard, S. (1994). *Outsourcing - Entscheidungen in der Informationsverarbeitung*. Unternehmensführung und Controlling. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH.
- Bortz, J. and N. Döring (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler [Research methods and Evaluation for Social scientists]* (4. ed.). Berlin: Springer.
- Bradford, M. and J. Florin (2003). Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning systems. *International Journal of Accounting Information Systems* 4(3), 205 – 225.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1993). *Classification and Regression Trees*. New York, USA: Chapman & Hall.
- Briscoe, G. and A. Marinos (2009, Dec). Community Cloud Computing. In *First International Conference on Cloud Computing*, Beijing, China.
- Brown, J. S. and J. Hagel (2003). Does IT matter? An HBR debate. *Harvard Business Review* June, pp. 1 – 17.
- Bruhn, M. (2008). *Qualitätsmanagement für Dienstleistungen : Grundlagen, Konzepte, Methoden* (7th ed.). Berlin: Springer.

- BSI (1992). Handbuch für die sichere Anwendung der Informationstechnik (IT) IT - Sicherheitshandbuch. Technical Report BSI 7105, Bundesamt für Sicherheit in der Informationstechnik (BSI).
- Burke, G. J., J. Carrillo, and A. J. Vakharia (2008, Apr). Heuristics for sourcing from multiple suppliers with alternative quantity discounts. *European Journal of Operational Research* 186(1), 317–329.
- Burke, G. J., J. Geunes, H. E. Romeijn, and A. Vakharia (2008). Allocating procurement to capacitated suppliers with concave quantity discounts. *Operations Research Letters* 36(1), 103–109.
- Buyya, R., C. S. Yeo, and S. Venugopal (2008). Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. In *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications*, Dalian, China.
- Buyya, R., C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic (2009). Cloud Computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems* 25(6), 599 – 616.
- Carr, N. G. (2003, May). IT doesn't matter. *Harvard Business Review* 81(5), 41–49.
- Chaisiri, S., R. Kaewpuang, B.-S. Lee, and D. Niyato (2011, july). Cost Minimization for Provisioning Virtual Servers in Amazon Elastic Compute Cloud. In *Modeling, Analysis Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on*, pp. 85–95.
- Chaisiri, S., B.-S. Lee, and D. Niyato (2009, dec.). Optimal virtual machine placement across multiple Cloud providers. In *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*, pp. 103–110.
- Chaisiri, S., B.-S. Lee, and D. Niyato (2012). Optimization of Resource Provisioning Cost in Cloud Computing. *Services Computing, IEEE Transactions on* 5(2), 164–177.
- Chaudry, S. S., F. G. Forst, and J. L. Zydiak (1993). Vendor selection with price breaks. *European Journal of Operational Research* 70, 52–66.
- Chauhan, S. S. and J.-M. Proth (2003). The concave cost supply problem. *European Journal of Operational Research* 148, 374–383.
- Cheon, M., V. Grover, and J. Teng (1995). Theoretical perspectives on the outsourcing of information systems. *Journal of Information Technology* 10(4), 209–219.
- Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (8 ed.), pp. 295–336. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Chin, W. W., B. L. Marcolin, and P. R. Newsted (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a monte carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research* 14(2), 189–217.
- Chin, W. W. and P. R. Newsted (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research*, Chapter Structural Equation Modeling Analysis with small samples using Partial Least Squares, pp. 307–342. Thousand Oaks, CA: Sage Publications.
- CIO Magazine (2009, Jun). Cloud Computing Survey. Online. <http://www.cio.com/documents/whitepapers/CIOCloudComputingSurveyJune2009V3.pdf> [last accessed 2010-05-13].

- Cloud Computing Use Case Discussion Group (2010, Jul). Cloud computing use cases. Online. http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf [last accessed 2012-04-05].
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Crawley, M. J. (2007). *The R Book*. The Atrium, England: John Wiley & Sons.
- Dastjerdi, A., S. Garg, and R. Buyya (2011, dec.). QoS-aware Deployment of Network of Virtual Appliances Across Multiple Clouds. In *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, pp. 415–423.
- David, R. J. and S.-K. Han (2004). A systematic assessment of the empirical support for transaction cost economics. *Strategic Management Journal* 25(1), 39–58.
- Davis, F. D. (1989, Sep). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13(3), 319–340.
- de Boer, L., E. Labro, and P. Morlacchi (2001). A review of methods for supporting supplier selection. *European Journal of Purchasing & Supply Management* 7(2), 75–89.
- Dean, J. and S. Ghemawat (2004, Dec). Mapreduce: Simplified data processing on large clusters. In *Proceedings of the 6th Symposium on Operating Systems Design & Implementation*, San Francisco. USENIX.
- Degraeve, Z., E. Labro, and F. Roodhooft (2000). An evaluation of vendor selection models from a total cost of ownership perspective. *European Journal of Operational Research* 125, 34–58.
- Degraeve, Z. and F. Roodhooft (2000). A mathematical programming approach for procurement using activity based costing. *Journal of Business Finance & Accounting* 27(1 & 2), 69–98.
- Diamantopoulos, A. and P. Riefler (2008). Formative Indikatoren: Einige Anmerkungen zu ihrer Art, Validität und Multikollinearität. *Zeitschrift für Betriebswirtschaft* 78, 1183–1196.
- Diamantopoulos, A. and H. M. Winklhofer (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research* 38(2), 269–277.
- Dibbern, J. (2003). *The Sourcing of Application Software Services*. Heidelberg: Physica-Verlag.
- Dibbern, J., T. Goles, R. Hirschheim, and J. Bandula (2004). Information systems outsourcing: A survey and analysis of the literature. *Database for Advances in Information Systems* 35(4), 6–102.
- Diller, H. and A. Herrmann (2003). *Handbuch Preispolitik [Handbook on Pricing]* (1 ed.). Wiesbaden: Gabler.
- DIN (1990). DIN 40041 - Zuverlässigkeit - Begriffe.
- Doney, P. M. and J. P. Cannon (1997, Apr.). An examination of the nature of trust in buyer-seller relationships. *The Journal of Marketing* 61(2), 35 – 51.
- Driver, M. (2008). Cloud Application Infrastructure Technologies Need Seven Years to Mature. Research report G00162990, Gartner Inc., Stamford, USA.
- Durkee, D. (2010). Why Cloud Computing Will Never Be Free. *Queue* 8(4), 20 – 29.
- Eckhardt, A., S. Laumer, and T. Weitzel (2009). Who influences whom? analyzing workplace referents' social influence on it adoption and non-adoption. *Journal of Information Technology* 24(1), 11–24.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7(1), 1 – 26.

- Eggert, A. and G. Fassott (2005). Zur Verwendung formativer und reflektiver Indikatoren in Strukturgleichungsmodellen: Bestandsaufnahme und Anwendungsempfehlung. In F. Bliemel, A. Eggert, G. Fassott, and J. Henseler (Eds.), *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele*, pp. 31 – 47. Stuttgart: Schäffer-Poeschel.
- Egle, U., D. Weibel, and T. Myrach (2008). Ziele und erfasste Kosten im IT-Kostenmanagement: Eine empirische Untersuchung.[Targets and recorded costs in the IT cost management]. In M. Bichler, T. Hess, H. Krcmar, U. Lechner, F. Matthes, A. Picot, B. Speitkamp, and P. Wolf (Eds.), *Multikonferenz Wirtschaftsinformatik 2008*. GITO-Verlag.
- Eisenhardt, K. M. (1989a, Jan.). Agency theory: An assessment and review. *The Academy of Management Review* 14(1), 57 – 74.
- Eisenhardt, K. M. (1989b). Building theories form case study research. *Academy of Managment Review* 14(4), 532–550.
- Eisenhardt, K. M. and L. J. Bourgeois III (1988, Dec.). Politics of strategic decision making in high-velocity environments: Toward a midrange theory. *Academy of management journal* 31(4), 737–770.
- Eisenhardt, K. M. and M. S. Graebner (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal* 50(1), 25–32.
- Elliott, G. (2004). *Global Business Information Technology* (1st. ed.). Harlow, England: Pearson Education Ltd.
- Engelbart, M. (2011, Mar.). SAP Sizing Daumenwerte. Online [last accessed 2013-02-11]. <http://www.common-d.de/pdf11/sap-240311/SAP%20Sizing%20RoT%202011common1.pdf>.
- Engels, R. (2007, Nov.). Workplace technologies, enterprise architecture and dimensions of work: Empirical research at rabobank nederland. Master's thesis, RSM Erasmus University Rotterdam, Rotterdam, Netherlands.
- European Commission (2003, May). Recommendation 2003/361/EC. Online. http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index_en.htm [last accessed 2011-08-08].
- Everest Group (2012, Aug). Enterprise Cloud Adoption Survey - Results. Online [last accessed 2012-12-04]. <http://www.everestgrp.com/wp-content/uploads/2012/08/CloudConnect-Everest-Group-Enterprise-Cloud-Adoption-Survey-2012-FINAL.pdf>.
- Fernández, M. and J. Martrat (2008, Jan). Draft report on the impact of grids in business and life improvement. Online. ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/ssai/challengers-research-agenda-roadmap-extended_en.pdf, [last accessed 2010-04-07].
- Fishbein, M. and I. Ajzen (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, Mass.: Addison-Wesley Pub. Co.
- Forge, S. and C. Blackman (2006, November). Commercial exploitation of grid technologies and services - drivers and barriers, business models and impacts of using free and open source licensing schemes. Online. ftp://ftp.cordis.europa.eu/pub/ist/docs/grids/study-report-commercial-exploitation-of-grid-technologies-services-2006-11-25_en.pdf, [last accessed 2008-10-05].
- Fornell, C. and F. L. Bookstein (1982, Nov.). Two structural equation models: Lisrel and pls applied to consumer exit-voice theory. *Journal of Marketing Research* 19(4), 440 – 452.

- Fornell, C. and J. Cha (1994). Partial least squares. In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research*, Chapter Partial least squares, pp. 52 – 78. Cambridge: Blackwell Publishers.
- Fornell, C. and D. F. Larcker (1981, Feb.). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18(1), 39 – 50.
- Foster, I., Y. Zhao, I. Raicu, and S. Lu (2008). Cloud Computing and Grid Computing 360-Degree Compared. In *2008 Grid Computing Environments Workshop*, pp. 1 – 10.
- Franke, C., A. Hohl, P. Robinson, and B. Scheuermann (2007, Aug). On business grid demands and approaches. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Rennes, France, pp. 124–134. Springer-Verlag Berlin Heidelberg.
- Frietsch, C. (2011, Mar.). Determinanten der betrieblichen IaaS-Nutzung. Diplomarbeit, Karlsruher Institut für Technologie, Karlsruhe. Supervising assistant J. Strebel.
- Fuchs, C. and A. Diamantopoulos (2009). Using single-item measures for construct measurement in management research. *Die Betriebswirtschaft* 69(2), 195 – 210.
- Garimella, N. (2006, Apr.). Understanding and exploiting snapshot technology for data protection, part 1: Snapshot technology overview. Online [last accessed 2012-11-25]. <http://www.ibm.com/developerworks/tivoli/library/t-snaptsml/index.html>.
- Gartner Inc. (2009, Jul). Hype Cycle for Cloud Computing. Research Report G00168780, Gartner Inc., Stamford, USA.
- Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review* 65(6), 101 – 109.
- Garzotto, F. (2010). Ein Entscheidungsmodell für den Einsatz von Cloud Computing. Diplomarbeit, Universität Karlsruhe (TU), Karlsruhe. Advisor: Jörg Strebel.
- Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika* 61(1), 101–107.
- Gerbing, D. W. and J. C. Anderson (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research* 25(2), pp. 186–192.
- Ghodsypoura, S. H. and C. O'Brien (2001). The total cost of logistics in supplier selection, under conditions of multiple sourcing, multiple criteria and capacity constraint. *International Journal of Production Economics* 73(1), 15–27.
- Gonzalez, R., J. Gasco, and J. Llopis (2006, Oct.). Information systems outsourcing: A literature analysis. *Information & Management* 43(7), 821 – 834.
- Göthlich, S. E. (2007). Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen. In S. Albers, D. Klapper, U. Konradt, A. Walter, and J. Wolf (Eds.), *Methodik der empirischen Forschung* (2nd ed.), pp. 120 – 134. Wiesbaden: Gabler Verlag.
- Gottfredson, M., R. Puryear, and S. Phillips (2005). Strategic sourcing: From periphery to the core. *Harvard Business Review* 83(2), 132–139.
- Gray, J. (2003, Mar). Distributed Computing Economics. Online. http://research.microsoft.com/research/pubs/view.aspx?tr_id=655, [last accessed 2008-10-05].
- Greenwood, D., A. Khajeh-Hosseini, J. Smith, and I. Sommerville (2010). The Cloud Adoption Toolkit: Addressing the Challenges of Cloud Adoption in Enterprise. *Computing Research Repository abs/1003.3866*, 10.

- Grover, V., J. T. Teng, and M. J. Cheon (1998). Strategic source of information systems: Perspective and practices. In L. P. Willcocks and M. Lacity (Eds.), *Towards a Theoretically-based Contingency Model of Information Systems Outsourcing*. New York: Wiley.
- Guder, F., J. Zydiak, and S. Chaudhry (1994). Capacitated multiple item ordering with incremental quantity discounts. *The Journal of the Operational Research Society* 45(10), 1197–1205.
- Hachenberg, K. (2009, Aug). Organisationsanalyse zur Integration von Grid Computing im Unternehmenskontext. Diploma thesis, Universität Karlsruhe (TU), Institute of Information Systems and Management (IISM), Karlsruhe. Thesis advisors Jochen Stößer, Steffen Haak, Jörg Strebel.
- Hair, Jr., J. F., R. E. Anderson, R. L. Tatham, and W. C. Black (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Hall, B. H. (2005). Ch. 17 innovation and diffusion. In J. Fagerberg, D. Mowery, and R. R. Nelson (Eds.), *The Oxford Handbook of Innovation*, pp. 459–484. Oxford: Oxford University Press.
- Hameed, I. (2010, Feb). Outsourcing-based Decision Framework for Cloud Computing Usage in Enterprises. Diploma thesis, Universität Karlsruhe (TU), Institute of Information Systems and Management (IISM), Karlsruhe. Thesis advisor Jörg Strebel, Thomas Meinl.
- Hamilton, J. (2008). Internet-scale service efficiency. In *Large-Scale Distributed Systems and Middleware (LADIS 2008)*. URL: http://www.cs.cornell.edu/projects/ladis2008/materials/JamesRH_Ladis2008.pdf, [last accessed March 2 2010].
- Hamilton, J. (2014, Nov.). Why Scale Matters and how the Cloud Really is Different. In *AWS re:Invent*, Las Vegas, USA. Amazon Web Services. http://mvdirona.com/jrh/TalksAndPapers/JamesHamilton_Reinvent20131115.pdf, [last accessed 2014-04-13].
- Hayes, B. (2008). Cloud Computing. *Commun. ACM* 51(7), 9–11.
- Heck, R. H. (1998). Factor analysis: Exploratory and confirmatory approaches. In G. Marcoulides (Ed.), *Modern Methods for Business Research*, Chapter Factor Analysis: Exploratory and Confirmatory Approaches, pp. 177 – 216. Mahwah: Lawrence Erlbaum Associates Publishers.
- Heinle, C. (2010, May). Organisatorisch Faktoren der Cloud Computing Nutzung in Unternehmen. Diploma thesis, Karlsruher Institut für Technologie, Karlsruhe. Supervising Assistant J. Strebel.
- Heinle, C. and J. Strebel (2010). IaaS Adoption Determinants in Enterprises. In J. Altmann and O. Rana (Eds.), *Economics of Grids, Clouds, Systems, and Services*, Volume 6296 of *Lecture Notes in Computer Science*, pp. 93–104. Springer Berlin / Heidelberg.
- Hempel, C. G. and P. Oppenheim (1948, Apr). Studies in the logic of explanation. *Philosophy of Science* 15(2), 135 – 175.
- Henseler, J. and G. Fassott (2010). Testing moderating effects in pls path models: An illustration of available procedures. In V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (Eds.), *Handbook of Partial Least Squares*, Springer Handbooks of Computational Statistics, pp. 713–735. Springer Berlin Heidelberg.
- Herzberg, F., B. Mausner, and B. Snyderman (1959). *The motivation to work*. New York: Wiley.
- Hildebrandt, L. and D. Temme (2006, Dec). Probleme der validierung mit strukturgleichungsmodellen. Sfb 649 discussion paper 2006-082, Humboldt-Universität Berlin, Institute of Marketing, Berlin.
- Hirschheim, R. and M. Lacity (2000). The myths and realities of information technology insourcing. *Commun. ACM* 43(2), 99–107.
- Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research* 11(3), 325–344.

- Hollander, M. and D. A. Wolfe (1973). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.
- Huang, K.-W. and A. Sundararajan (2005, Nov.). Pricing Models for On-Demand Computing. CeDER Working Paper No. 05-26. Available at SSRN: <http://ssrn.com/abstract=859504>.
- Huber, F., A. Herrmann, F. Meyer, J. Vogel, and K. Vollhardt (2007). *Kausalmodellierung mit Partial Least Squares* (1st ed.). Wiesbaden: Gabler Verlag.
- Hwang, C.-L. and K. Yoon (1981). *Multiple Attribute Decision Making*. Lecture Notes in Economical and Mathematical Systems. Berlin: Springer Verlag.
- Hwang, J. and J. Park (2007, Aug). Decision factors of enterprises for adopting grid computing. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Rennes, France, pp. 16–28. Springer-Verlag Berlin Heidelberg.
- IDC Central Europe (2009, June). IDC-Studie: Cloud Computing in Deutschland ist noch nicht angekommen. Online. http://www.idc.com/germany/press/presse_cloudcomp.jsp, [last accessed 2009-10-07].
- Jarvis, C. B., S. B. MacKenzie, and P. M. Podsakoff (2003, Sep.). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30(2), 199 – 218.
- Jiménez-Peris, R., M. Patiño-Martínez, and B. Kemme (2007, Sept). Enterprise grids: Challenges ahead. *Journal of Grid Computing* 5(3), 283–294.
- Kaiser, H. F. and J. Rice (1974). Little jiffy, mark iv. *Educational and Psychological Measurement* 34(1), 111–117.
- Kall, P. and J. Mayer (2011). *Stochastic Linear Programming - Models, Theory and Computation*. (2nd ed.), Volume vol. 156 of *International Series in Operations Research & Management Science*. New York: Springer.
- Kenyon, C. and G. Cheliotis (2004). Grid resource commercialization: Economic engineering and delivery scenarios. In J. Nabrzyski, J. M. Schopf, and J. Weglarz (Eds.), *Grid Resource Management: State of the Art and Future Trends* (1st edition ed.), International Series in Operations Research & Management Science, Chapter 28, pp. 465–478. Kluwer Academic Publishers.
- Khajeh-Hosseini, A., D. Greenwood, J. W. Smith, and I. Sommerville (2012). The Cloud Adoption Toolkit: supporting cloud adoption decisions in the enterprise. *Software: Practice and Experience* 42(4), 447–465.
- Khajeh-Hosseini, A., D. Greenwood, and I. Sommerville (2010). Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pp. 450–457.
- Khajeh-Hosseini, A., I. Sommerville, J. Bogaerts, and P. Teregowda (2011, july). Decision Support Tools for Cloud Migration in the Enterprise. In *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, pp. 541 –548.
- Kim, W., S. D. Kim, E. Lee, and S. Lee (2009, Dec). Adoption Issues for Cloud Computing. In *The 11th International Conference on Information Integration and Web-based Applications & Services(iiWAS2009)*, Kuala Lumpur. URL: http://uclab.khu.ac.kr/resources/publication/C_196.pdf [last accessed 2010-04-08].
- Klein, R. (1997). *Algorithmische Geometrie* (1st ed. ed.). Bonn: Addison-Wesley-Longman.

- Klems, M., J. Nimis, and S. Tai (2008). Do Clouds Compute? A Framework for Estimating the Value of Cloud Computing. In *Proceedings of Web2008*.
- Kohavi, R. and F. Provost (1998, Feb.). Glossary of terms. *Machine Learning* 30(2-3), 271–274.
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatika* 31, 249 – 268.
- KPMG AG (2013, Feb.). Cloud-Monitor 2013. Online. <http://www.kpmg.com/de/de/bibliothek/2013/seiten/cloud-monitor-2013.aspx>, [last accessed 2014-03-10].
- Krafft, M., O. Götz, and K. Liehr-Gobbers (2005). Die Validierung von Strukturgleichungsmodellen mit Hilfe des Partial-Least-Squares (PLS)-Ansatzes. In F. Bliemel, A. Eggert, G. Fassott, and J. Henseler (Eds.), *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele*, pp. 71 – 86. Stuttgart: Schäffer-Poeschel.
- Küchler, P. (2004). Technische und wirtschaftliche Grundlagen. In P. Bräutigam (Ed.), *IT-Outsourcing*, pp. 51–159. Berlin: Erich Schmidt Verlag.
- Kumar, N., L. W. Stern, and J. C. Anderson (1993, Dec.). Conducting interorganizational research using key informants. *The Academy of Management Journal* 36(6), 1633 – 1651.
- Lacity, M. C., L. P. Willcocks, and D. F. Feeny (1996, Apr). The Value of Selective IT Sourcing. *Sloan Management Review* 37(3), 13 – 25.
- Legris, P., J. Ingham, and P. Colletette (2003). Why do people use information technology? a critical review of the technology acceptance model. *Information & Management* 40(3), 191 – 204.
- Leong, L. (2009, Mar). Toolkit: Estimating the Cost of Cloud Infrastructure. Technical Report G00165397, Gartner Inc., Stamford, CT.
- Li, X., Y. Li, T. Liu, J. Qiu, and F. Wang (2009, sept.). The Method and Tool of Cost Analysis for Cloud Computing. In *Cloud Computing, 2009. CLOUD '09. IEEE International Conference on*, pp. 93 –100.
- Liker, J. K. and T. Y. Choi (2004). Building deep supplier relationships. *Harvard Business Review* 82(12), 104 – 114.
- Lilienthal, M. (2013). Ein Entscheidungsmodell für Cloud-Bursting. *Wirtschaftsinformatik* 55(2), 69–81.
- Logan, M. S. (2000). Using agency theory to design successful outsourcing relationships. *The International Journal of Logistics Management* 11(2), 21 – 32.
- Loh, L. (1994). An organizational-economic blueprint for information technology outsourcing: Concepts and evidence. In *ICIS 1994 Proceedings*, Volume Paper 8.
- Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Springer.
- Ma, Q., M. Pearson, and S. Tadisina (2005). An exploratory study into factors of service quality for application service providers. *Information & Management* 42(8), 1067 – 1080.
- Maimon, O. and L. Rokach (2010). *Data Mining and Knowledge Discovery Handbook* (2nd ed.), Chapter Ch.9 Decision Trees, pp. 149 – 174. New York: Springer.
- Mann, H. B. and D. R. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18(1), 50–60.
- Maqueira, J. and S. Bruque (2006). Towards an adoption model of grid information technology in the organisational arena. *2006 7th IEEE/ACM International Conference on Grid Computing* 0, 325–326.

- Mark, C., D. Niyato, and T. Chen-Khong (2011, march). Evolutionary Optimal Virtual Machine Placement and Demand Forecaster for Cloud Computing. In *Advanced Information Networking and Applications (AINA), 2011 IEEE International Conference on*, pp. 348–355.
- Markowitz, H. M. (1959). *Portfolio Selection*. New York: John Wiley & Sons.
- Marston, S., Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi (2011, Apr.). Cloud Computing - The business perspective. *Decision Support Systems* 51(1), 176 – 189.
- Martin, W., J. Eckert, and N. Repp (2010, Sep.). XaaS check 2010. Online. <http://www.wolfgang-martin-team.net/pdf/2010-XaaS-Check-09-30-Final.pdf>, [last accessed 2013-02-07].
- Matros, R. (2012). *Der Einfluss von Cloud Computing auf IT-Dienstleister*. Dissertation, Universität Bayreuth, Bayreuth.
- Mayer, H. O. (2008). *Interview und schriftliche Befragung - Entwicklung, Durchführung, Auswertung* (4th. ed.). München: Oldenbourg Wissenschaftsverlag.
- Meents, J. G. (2010, May). Cloud Computing: Rechtlich Aspekte [Legal Aspects]. Online. URL: <http://www.haufe.de/recht/newsDetails?newsID=1272627923.17> [last accessed 2010-05-25].
- Menzel, M., M. Schönherr, and S. Tai (2013). (MC2)2: criteria, requirements and a software prototype for cloud infrastructure decisions. *Software: Practice and Experience* 43(11), 1283–1297.
- Messerschmidt, C. M. and O. Hinz (2013). Explaining the adoption of grid computing: An integrated institutional theory and organizational capability approach. *The Journal of Strategic Information Systems* 22(2), 137 – 156.
- Meuser, M. and U. Nagel (2009). Das Experteninterview: Konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth, and D. Jahn (Eds.), *Methoden der vergleichenden Politik- und Sozialwissenschaft*, pp. 465 – 479. Wiesbaden: VS Verlag.
- Michalk, W. A. (2011). *SLA Establishment Decisions: Minimizing the Risk of SLA Violations*. Ph. D. thesis, Karlsruhe Service Research Institute (KSRI), Karlsruhe.
- Modheji, S. (2010). Dimensionen eines Qualitätsmodells zur Bewertung von Public Cloud Infrastruktur. Bachelor thesis, Karlsruher Institut für Technologie, Karlsruhe.
- Moore, G. C. and I. Benbasat (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research* 2(3), 192–222.
- Musser, D. (1997). Introspective sorting and selection algorithms. *Software Practice and Experience* 27, 983–993.
- Myers, M. D. (1997, Jun). Qualitative research in information systems. *MIS Quarterly* 21(2), 241–242. updated version, last modified: September 2, 2008, last accessed on Oct. 13 2008.
- Nam, K., S. Rajagopalan, H. R. Rao, and A. Chaudhury (1996, July). A two-level investigation of information systems outsourcing. *Commun. ACM* 39, 36–44.
- Neumann, D., J. Stöber, A. Anandasivam, and N. Borissov (2007). SORMA - Building an Open Grid Market for Grid Resource Allocation. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Berlin, pp. 194 – 200. Springer-Verlag.
- NIST (2009, Aug). Draft NIST Working Definition of Cloud Computing. Online. URL: <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>, [last accessed 2009-11-20].

- NIST/SEMATECH (2014, Jan.). e-handbook of statistical methods - process improvement. Online. URL: <http://www.itl.nist.gov/div898/handbook/pri/pri.htm>, [last accessed 2014-02-12].
- Nitzl, C. (2010, Jun.). *Eine anwenderorientierte Einführung in die Partial Least Square (PLS)-Methode*. Hamburg: Universität Hamburg. <http://ssrn.com/abstract=2097324>.
- Nunnally, J. C. and I. H. Bernstein (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill Book Company.
- Nurmi, D., R. Wolski, C. Grzegorzczak, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov (2009). The Eucalyptus Open-Source Cloud-Computing System. In *CCGRID '09: Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, Washington, DC, USA, pp. 124–131. IEEE Computer Society.
- Nysveen, H., P. Pedersen, and H. Thorbjørnsen (2005). Intentions to use mobile services: Antecedents and cross-service comparisons. *Journal of the Academy of Marketing Science* 33(3), 330–346. 10.1177/0092070305276149.
- Office of Government Commerce (2005). *Service Support*. Norwich: The Stationary Office.
- Orlikowski, W. J. and J. J. Baroudi (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research* 2(1), 1 – 28.
- Paleologo, G. A. (2004). Price-at-risk: A methodology for pricing utility computing services. *IBM Syst. J.* 43(1), 20–31.
- Parasuraman, A., V. A. Zeithaml, and L. L. Berry (1988). Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing* 64(1), 12 – 40.
- Parasuraman, A., V. A. Zeithaml, and A. Malhotra (2005). E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality. *Journal of Service Research* 7(3), 213–233.
- Parrilli, D. (2009). The Determination of Jurisdiction in Grid and Cloud Service Level Agreements. In J. Altmann, R. Buyya, and O. Rana (Eds.), *Grid Economics and Business Models*, Volume 5745 of *Lecture Notes in Computer Science*, pp. 128–139. Springer Berlin / Heidelberg. 10.1007/978-3-642-03864-8_10.
- Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce* 7(3), 101 – 134.
- Pavlou, P. A., H. Liang, and Y. Xue (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly* 31(1), 105 – 136.
- Perko, J. (2008, Dec.). *IT Governance and Enterprise Architecture as Prerequisites for Assimilation of Service-Oriented Architecture*. Ph. D. thesis, Tampere University of Technology, Tampere, Finland.
- Perunovic, Z. (2009, Feb.). *The Utilisation of Information and Communication Technology across the Outsourcing Process : The Vendor's Perspective*. Dissertation, Technical University of Denmark (DTU), Lyngby, Denmark.
- Picot, A. and M. Maier (1992). Analyse und Gestaltungskonzepte für das Outsourcing [Analysis and Design concepts for Outsourcing]. *Information Management* 4, 14–27.
- Piech, M. (2009, Oct). Platform-as-a-Service Private Cloud with Oracle Fusion Middleware. Whitepaper. <http://www.oracle.com/ocom/groups/public/documents/webcontent/036500.pdf> [last accessed 2010-03-29].
- Pirkul, H. and O. A. Aras (1985, Sep). Capacitated multiple item ordering problem with quantity discounts. *IIE Transactions* 17(3), 206–211.

- Popper, K. R. (1934). *Logik der Forschung*. Tübingen: Mohr Siebeck.
- Powers, D. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2(1), 37–63.
- Quality Management, Statistics and Certification Standards Committee (2005). Quality management systems - Fundamentals and vocabulary (ISO 9000:2005).
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, USA: Morgan Kaufmann.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Raileanu, L. E. and K. Stoffel (2004, May). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence* 41(1), 77–93.
- Rapid-I (2010, Nov.). Rapid-i wiki - decision tree. Online. http://rapid-i.com/wiki/index.php?title=Decision_Tree [last accessed on 2013-02-26].
- Reinecke, J. (1999). Interaktionseffekte in Strukturgleichungsmodellen mit der Theorie des geplanten Verhaltens : multiple Gruppenvergleiche und Produktterme mit latenten Variablen. *ZUMA Nachrichten* 23(45), 88–114.
- Rentrop, C. and S. Zimmermann (2012). Shadow IT - Management and Control of Unofficial IT. In *ICDS 2012, The Sixth International Conference on Digital Society*, Valencia, pp. 98–102. IARIA.
- Ringle, C. M., S. Wende, and S. Will (2005). SmartPLS 2.0 (M3) Beta. Online. <http://www.smartpls.de> [last accessed 2012-12-23].
- Rinne, H. (2008). *Taschenbuch der Statistik* (4th. ed.). Frankfurt am Main: Verlag Harri Deutsch.
- Risch, M. and J. Altmann (2008). Cost analysis of current grids and its implications for future grid markets. In J. Altmann, D. Neumann, and T. Fahringer (Eds.), *Grid Economics and Business Models 5th International Workshop, GECON 2008*, Volume 5206 of *LNCS*, Berlin Heidelberg, pp. 13–27. Springer-Verlag.
- Rogers, E. M. (2003, Aug). *Diffusion of Innovations* (5th edition ed.). New York: Free Press.
- Rolia, J., A. Andrzejak, and M. Arlitt (2003). Automating enterprise application placement in resource utilities. In M. Brunner and A. Keller (Eds.), *Self-Managing Distributed Systems - 14th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management*, Volume 2867 of *Lecture Notes in Computer Science*, Berlin, pp. 118–129. Springer.
- Rosenthal, E. C., J. L. Zydiak, and S. S. Chaudhry (1995). Vendor selection with bundling. *Decision Sciences* 26(1), 35–48.
- Ross, J. W. (2003). Creating a strategic it architecture competency: Learning in stages. *MIS Quarterly Executive* 2(1), 31–43.
- Ross, J. W. and C. M. Beath (2006). Sustainable it outsourcing success: Let enterprise architecture be your guide. *MIS Quarterly Executive* 5(4), 181–192.
- Ross, J. W. and G. Westerman (2004). Preparing for utility computing: The role of it architecture and relationship management. *IBM Systems Journal* 43(1), 5–19.
- Sadrian, A. A. and Y. S. Yoon (1994). A procurement decision support system in business volume discount environments. *Operations Research* 42(1), 14–23.
- Schikuta, E., F. Donno, H. Stockinger, E. Vinek, H. Wanek, T. Weishäupl, and C. Witzany (2005). Business in the grid: Project results. Online. URL: http://www.pri.univie.ac.at/Publications/2005/Schikuta_austriangrid_bigresults.pdf [last accessed 2008-10-05].

- Seltin, N. and J. P. Keeves (1994). Path analysis with latent variables. In T. Postlethwaite and T. Husen (Eds.), *The International Encyclopedia of Education* (2nd ed.), pp. 4352 – 4359. Oxford: Pergamon.
- Sheppard, B. H., J. Hartwick, and P. R. Warshaw (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research* 15(3), 325 – 343.
- Silver, M. A. (2007, Dec). Gartner PC TCO: The Next Generation. Research report G00153432, Gartner Inc., Stamford, CT, USA.
- Stockinger, H. (2006, Jun). Grid computing: A critical discussion on business applicability. IEEE Distributed Systems Online. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1644718&isnumber=34467>, [last accessed 2013-04-03].
- Stone, M. (1974). Cross-validators choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2), 111 – 147.
- Strebel, J. and R. Matros (2010, Mar.). IaaS Vendor selection in hybrid Cloud environments. Unpublished (planned submission for Electronic Markets journal).
- Strong, P. (2005). Enterprise grid computing. *ACM Queue* 3(6), 50–59.
- Susarla, A., A. Barua, and A. B. Whinston (2009). A transaction cost perspective of the "software as a service" business model. *Journal of Management Information Systems* 26(2), 205 – 240.
- System Virtualization, Partitioning, and Clustering Working Group (2010, Jan). Open virtualization format specification. Online. URL: http://www.dmtf.org/standards/published_documents/DSP0243_1.1.0.pdf, [last accessed 2010-03-02].
- Tan, P.-N., M. Steinbach, and V. Kumar (2006). *Introduction to Data Mining*, Chapter Ch. 4 Classification: Basic Concepts, Decision Trees and Model evaluation, pp. 145 – 205. Boston: Addison-Wesley Pub. Co.
- Taylor, S. and P. Todd (1995). Understanding information technology usage: A test of competing models. *Information systems research* 6(2), 144–176.
- Tenenhaus, M., V. E. Vinzi, Y.-M. Chatelin, and C. Lauro (2005). Pls path modeling. *Computational Statistics & Data Analysis* 48(1), 159 – 205.
- Teng, J. T. C., M. J. Cheon, and V. Grover (1995). Decisions to outsource information systems functions: Testing a strategy-theoretic discrepancy model. *Decision Sciences* 26(1), 75 – 103.
- Thanos, G. A., C. Courcoubetis, and G. D. Stamoulis (2007, Aug). Adopting the grid for business purposes: The main objectives and the associated economic issues. In D. J. Veit and J. Altmann (Eds.), *Grid Economics and Business Models: 4th international workshop; proceedings / GECON 2007*, Number 4685 in LNCS, Rennes, France, pp. 1–15. Springer-Verlag Berlin Heidelberg.
- Thielsch, M. T. and S. Weltzin (2009). Online-Befragungen in der Praxis. In T. Brandenburg and M. T. Thielsch (Eds.), *Praxis der Wirtschaftspsychologie*, Chapter Online-Befragungen in der Praxis, pp. 69 – 85. Münster: MV-Verlag.
- Treber, U., P. Teipel, and A. C. Schwickert (2004). Total Cost of Ownership - Stand und Entwicklungstendenzen 2003 (Total Cost of Ownership - State of the Art and Research directions 2003). In Professur BWL - Wirtschaftsinformatik (Ed.), *Arbeitspapiere WI*, Number 1. Justus-Liebig-Universität Gießen.
- TripleTree (2003, Mar). 2003 outsourcing update. Spotlight report vol 6 nr. 1, Triple-Tree LLC, Minneapolis, USA. http://www.triple-tree.com/research/business/outsourcing_apr_03.pdf, [last accessed on 2009-02-02].

- Trummer, I., F. Leymann, R. Mietzner, and W. Binder (2010, dec.). Cost-Optimal Outsourcing of Applications into the Clouds. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pp. 135–142.
- Unser, M. (2000). Lower partial moments as measures of perceived risk: An experimental study. *Journal of Economic Psychology* 21(3), 253–280.
- Van den Bossche, R., K. Vanmechelen, and J. Broeckhove (2010, July). Cost-Optimal Scheduling in Hybrid IaaS Clouds for Deadline Constrained Workloads. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pp. 228–235.
- Vandenbosch, B. and K. Lyytinen (2004). Much ado about IT: a response to “the corrosion of IT advantage” by Nicholas G. Carr. *Journal of Business Strategy* 25(6), 10–12.
- Vaquero, L. M., L. Rodero-Merino, J. Caceres, and M. Lindner (2009). A break in the Clouds: towards a Cloud definition. *SIGCOMM Comput. Commun. Rev.* 39(1), 50–55.
- Venkatesh, V. and F. Davis (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* 6(2), 186–204.
- Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis (2003, Sep). User acceptance of information technology: Towards a unified view. *MIS Quarterly* 27(3), 425–478.
- Vicente, P. and E. Reis (2010, May). Using questionnaire design to fight nonresponse bias in web surveys. *Social Science Computer Review* 28(2), 251–267.
- Wadhwa, V. and A. R. Ravindran (2007). Vendor selection in outsourcing. *Computers & Operations Research* 34, 3725–3737.
- Weber, C. A. and J. R. Current (1993). A multiobjective approach to vendor selection. *European Journal of Operational Research* 68, 173–184.
- Weber, C. A., J. R. Current, and W. C. Benton (1991). Vendor selection criteria and methods. *European Journal of Operational Research* 50, 2–18.
- Weiber, R. and D. Mühlhaus (2010). *Strukturgleichungsmodellierung*. Berlin: Springer.
- Weill, P. and J. Ross (2004). *IT Governance*. Boston, USA: Harvard Business School Press.
- Weinhardt, C., A. Anandasivam, B. Blau, N. Borissov, T. Meinl, W. Michalk, and J. Stößer (2009). Cloud-Computing. *Wirtschaftsinformatik* 51(5), 453–462.
- Westhoff, J. (2008, Apr). Grid computing in small and medium-sized enterprises: An exploratory study of corporate attitudes towards economic and security-related issues. Bayreuth Reports on Information Systems Management 32, Universität Bayreuth, Bayreuth.
- Wiedemann, D. and J. Strebel (2011, sept.). Organizational Determinants of Corporate IaaS Usage. In *Commerce and Enterprise Computing (CEC), 2011 IEEE 13th Conference on*, pp. 191–196.
- Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* 24(3/4), pp. 471–494.
- Willcocks, L., M. Lacity, and S. Cullen (2007). Ch. 10. In R. Mansell, C. Avgerou, D. Quah, and R. Silverstone (Eds.), *The Oxford Handbook of Information and Communication Technologies*, pp. 244–272. Oxford: Oxford Univ. Press.
- Williamson, O. E. (1985). *The economic institutions of capitalism: firms, markets, relational contracting*. New York: The Free Press.
- Wilson, R. (1999, Oct). Short course on nonlinear pricing. Online. URL: <http://faculty-gsb.stanford.edu/wilson/PDF/Mechanism%20Design/Short%20course%20on%20nonlinear%20pricing.pdf>, [last accessed 2009-11-30] Course notes.

- Wimmer, M., V. Nicolescu, D. Gmach, M. Mohr, A. Kemper, and H. Krcmar (2006). Evaluation of Adaptive Computing Concepts for Classical ERP Systems and Enterprise Services. In *E-Commerce Technology, 2006. The 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, The 3rd IEEE International Conference on*, San Francisco, CA, USA, pp. 48–51. IEEE Computer Society.
- Wold, H. (1982). Soft modeling: the basic design and some extensions. In K. G. Joreskog (Ed.), *Systems under Indirect Observation, Part 2*, pp. 1 – 54. Amsterdam: Elsevier Science Ltd.
- Wu, X., V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems 14*(1), 1 – 37.
- Xin, M. and N. Levina (2008). Software-as-a service model: Elaborating client-side adoption factors. In *ICIS 2008 Proceedings*. Paper 86.
- Yanosky, R. (2008). From Users to Choosers: The Cloud and the Changing Shape of Enterprise Authority. In R. N. Katz (Ed.), *The Tower and the Cloud*, Chapter From Users to Choosers: The Cloud and the Changing Shape of Enterprise Authority, pp. 126 – 136. Educause.
- Yao, Y. (2004). *An integrative model of clients' decision to adopt an Application Service Provider*. Ph. D. thesis, Louisiana State University, Baton Rouge, LA 70803.
- Youseff, L., M. Butrico, and D. D. Silva (2008). Toward a Unified Ontology of Cloud Computing. In *Grid Computing Environments Workshop 2008 (GCE '08)*, Austin, USA, pp. 1–10.
- Zhang, H., G. Jiang, K. Yoshihira, H. Chen, and A. Saxena (2009, july). Intelligent Workload Factoring for a Hybrid Cloud Computing Model. In *Services - I, 2009 World Conference on*, pp. 701 –708.