Karlsruher Institut für Technologie

Private Clouds with Open Source

GridKa School 2010 – KIT – September 7th 2010

Christian Baun

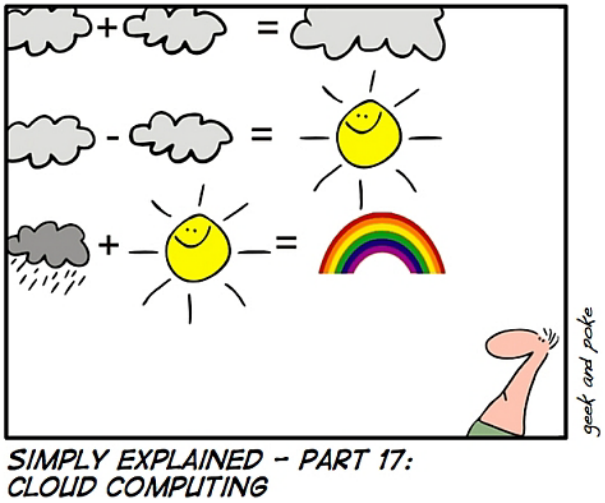
baun@kit.edu

Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

Cloud-Computing ?

- **Building on compute and storage virtualization, and leveraging the modern Web, Cloud Computing provides scalable, network-centric, abstracted IT infrastructure, platforms, and applications as on-demand services that are billed by consumption**
- Organizational Types
  - Public Cloud
  - Private Cloud
  - Hybrid Cloud
- Categories of Cloud services
  - IaaS
  - PaaS
  - SaaS
  - HuaaS



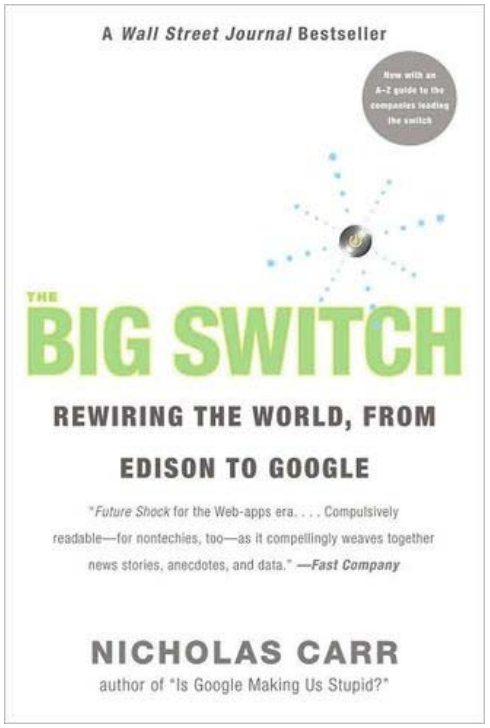
SIMPLY EXPLAINED – PART 17:
CLOUD COMPUTING

Monday, September 6

12:00-14:00	[illegible]
14:00 - 14:30	[illegible]
14:30 - 15:30	**Introduction to Grid and Cloud Computing** (T. Cass, CERN) abstract
15:30-16:00	[illegible]
16:00-17:00	[illegible]

Public Cloud vs. Private Cloud

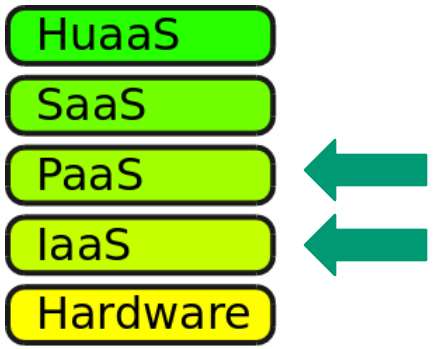
- Reasons for using Public Cloud services
  - Users have no costs for purchase, operation and maintenance of own server hardware
  - Fully automated services
  - Virtualized resources (no driver issues)
  - Pay-as-you-go principle
  - Services are elastic
  - **N. Carr: Transition of IT into the era of industrialization**



- Reasons for <u>not</u> using Public Cloud services
  - Fear for a lock-in situation
  - Security or privacy reasons
  - Local server hardware already exists

→ Where can I get a private cloud?

Focus: IaaS and PaaS Private Clouds



- Platform as a Service (PaaS)
  - Scalable runtime environment and (sometimes) development environment for 1 or 2 programming languages
  - No administrative effort for the users concerning the operation environment
- Infrastructure as a Service (IaaS)
  - Users run virtual server instances with unmodified applications
  - No direct contact to physical hardware for the users
  - Administrative user rights
  - Users can define the firewall rules independently

Requirements for a Private Cloud PaaS and IaaS

- Easy to install and use (for administrators and users) and Secure
- Open Source
  - No purchase costs
  - Easy to adopt (flexible)
- API compatible to popular Public Cloud services
  - Amazon Elastic Compute Cloud (EC2) is the most popular Public Cloud IaaS
    - EC2 is a part of the Amazon Web Services (AWS), a collection of different Cloud services
    - Billing according to consumption
    - Dynamic development
    - Popular services within the AWS are EC2, S3, EBS, ELB…
  - Google App Engine (GAE) is the most popular Public Cloud PaaS
    - Allows to run Python and Java web applications
- Integration of Public Cloud resources inside a Private Cloud (=> Hybrid Cloud)
  - Useful Scenarios for a Hybrid Cloud
    - Processing of load peaks with EC2 instances while running fewer servers locally
    - Outsourcing of data copies inside S3 to increase availability

Private Cloud PaaS Frameworks – An Overview

- Only few Private Cloud PaaS solutions available are Open Source
- Number of available solutions is shorter than it appears at first view

10gen	`http://www.10gen.com`
Reasonably Smart	`http://reasonablysmart.com`
AppScale	`http://appscale.cs.ucsb.edu`
typhoonAE	`http://code.google.com/p/typhoonae/`

10gen, Reasonably Smart

- 10gen
  - First Private Cloud PaaS available that was Open Source
  - Platform with a Java-based application server named “Hedley”
  - Support for JavaScript and Ruby applications
  - Web application framework Django10 available
  - Object-oriented database Mongo used to store data
  - Today, 10gen does only development and support for the Mongo database
  - Source code is still available (**`http://github.com/10gen`**)

- Reasonably Smart
  - Uses the free version control system Git
  - Support for JavaScript applications
  - Reasonably Smart was acquired in January 2009 from Joyent
  - Source code appears to be not available any longer

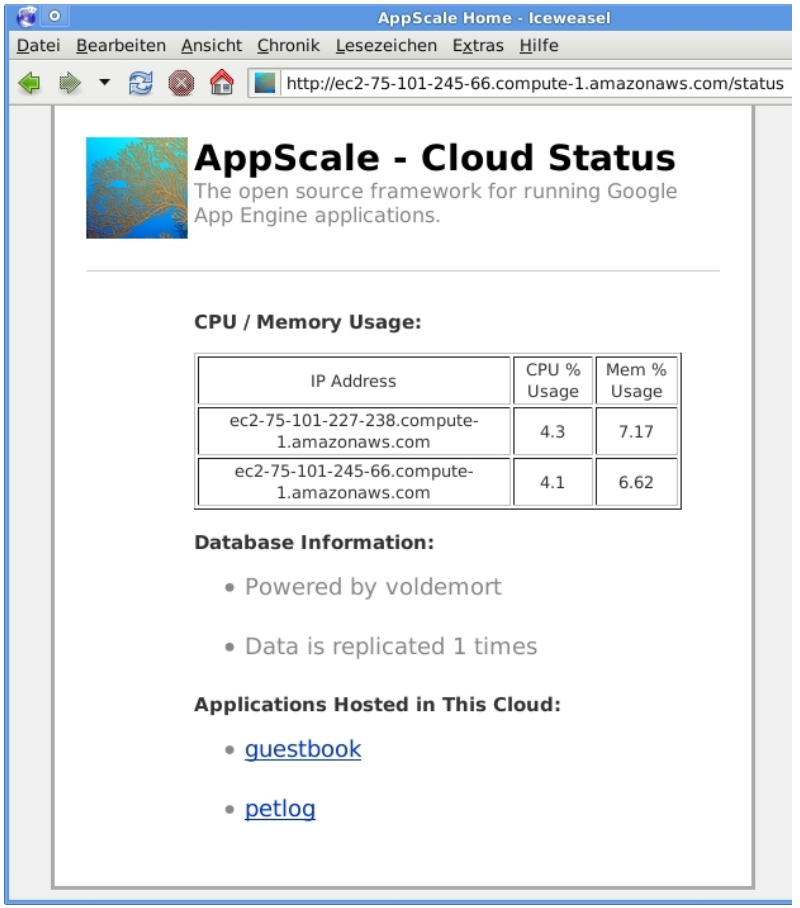
    “Yes, We are Open Source” (**`http://code.reasonablysmart.com`**)

    “Our repositories are offline at the moment, but we'll be back shortly “

AppScale

http://appscale.cs.ucsb.edu

- Open Source re-implementation of the Google App Engine (GAE)
- GAE compatible applications can be developed, deployed and run inside
  - Private Clouds IaaS (Eucalyptus)
  - Public Cloud IaaS (EC2)
  - Virtualized systems (Xen and KVM)
- Supports Python and Java applications
- Emulates Google's infrastructure services Datastore, XMPP, Memcache, Mail, authentication …
- AppScale 1.0 (March 2009)
- AppScale 1.3 (December 2009)
- Supported GAE version: 1.2.7
- Supported by Google and IBM Research

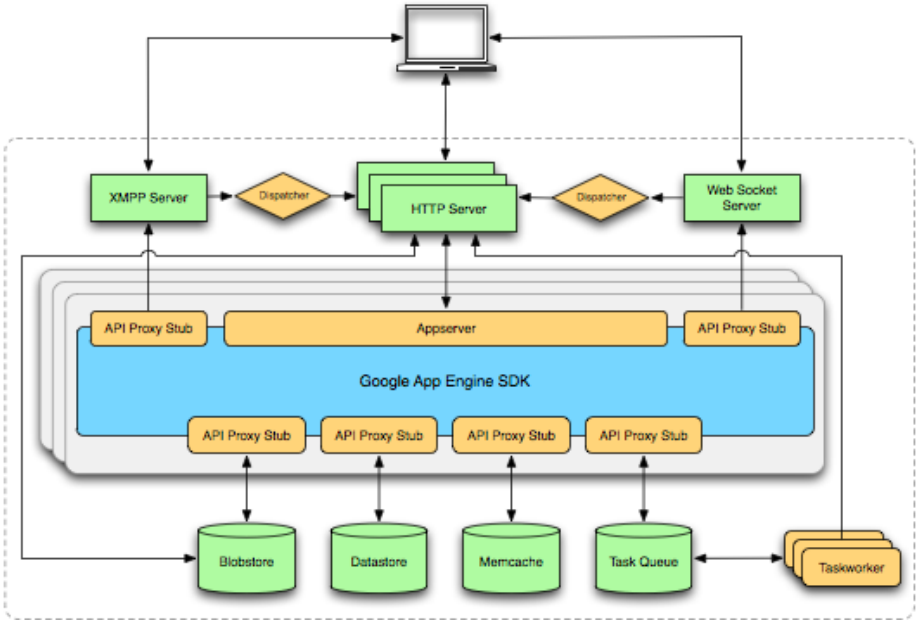


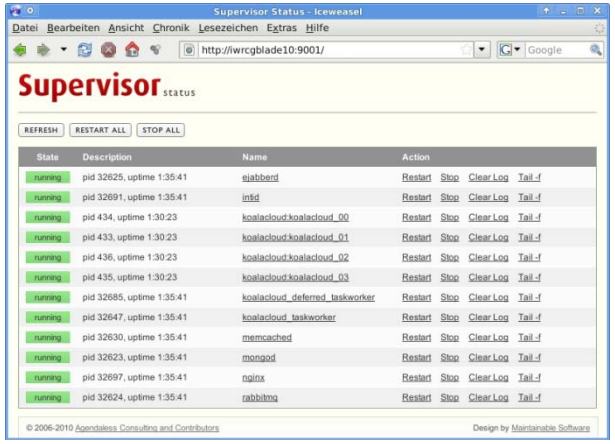
typhoonAE

http://code.google.com/p/typhoonae

typhoonae
Typhoon App Engine

- Open Source re-implementation of the Google App Engine (GAE)
- GAE compatible applications can be developed, deployed and run
  - Locally (Linux or Mac OS X)
  - inside a Private Cloud IaaS
  - Inside a Public Cloud IaaS
- Supports Python applications
- Uses the development server from the App Engine SDK and popular open source packages like MySQL and memcached to emulate Google's infrastructure services
- Supported GAE version: 1.3.5





Private Cloud IaaS Frameworks – An Overview

- Lots of Private Cloud IaaS solutions available that are Open Source
- Some are already used in science projects
  - CERN builds an Cloud Environment with OpenNebula with the goal to manage up to 45,000 Virtual Machine instances

Cloud.com CloudStack	`http://cloud.com`
Abiquo (AbiCloud)	`http://www.abicloud.org`
OpenNebula	`http://www.opennebula.org`
Nimbus	`http://www.nimbusproject.org`
Tashi	`http://incubator.apache.org/tashi/`
Enomaly ECP	`http://src.enomaly.com`
OpenECP	`http://www.openecp.org`
Eucalyptus	`http://open.eucalyptus.com`

Enomaly Elastic Compute Cloud, OpenECP

- Enomaly Elastic Compute Cloud
  - Also known as Enomalism Elastic computing platform (ECP)
  - First Private Cloud IaaS solution (since 2005) available that was Open Source
  - No support for EC2 API implemented
  - Max 10 nodes
  - Only few documentation
  - No storage service included
  - Since autumn 2009, Enomaly ECP is not officially available any longer
  - The company left it's Open Source strategy behind

    "Current customer demands require that we focus on expanding our commercial offerings via the Service Provider Edition and the High Assurance Edition" (**`http://src.enomaly.com`**)

- OpenECP
  - Fork of Enomaly ECP
  - Project start: February 2010
  - No support for EC2 API implemented
  - No storage service included
  - Popularity of the project is weak and therefore the future is unclear

Abiquo, Cloud.com

- Abiquo
  - Also known as AbiCloud
  - Version 1.0.0 since February 2010
  - No support for EC2 API implemented
  - No storage service included
  - Community is small

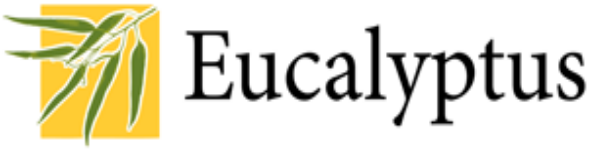
- Cloud.com CloudStack
  - Developed by a start-up company
  - Only a small part of the EC2 API implemented
  - Difficult to set up
  - Available as Community, Enterprise and Service Provider edition
  - No storage service included but an S3 compatible service is in development
  - Software is still buggy
  - Community is small
  - Small startup company => Future is unclear

Nimbus, Tashi

- Nimbus
  - Build on top of the Grid middleware Globus 4
  - Only a small part of the EC2 API implemented
    - describe images
    - describe, run, reboot und terminate instances
    - add und delete keypair
  - EC2 compatible resources can be used via remote (=> Hybrid Cloud)
  - Includes „Cumulus“, a storage service that is compatible with S3 REST API
  - Schedulers like PBS (Portable Batch System) or SGE (Sun Grid Engine) can be used to schedule virtual machines
  - Community is small

- Tashi
  - Development is done from the Intel Labs Pittsburgh
  - Focus is data organization in cluster systems
  - No support for EC2 API implemented
  - No storage service included
  - Only few installations world wide
  - Part of the OpenCirrus Cloud Computing Research Testbed

Eucalyptus

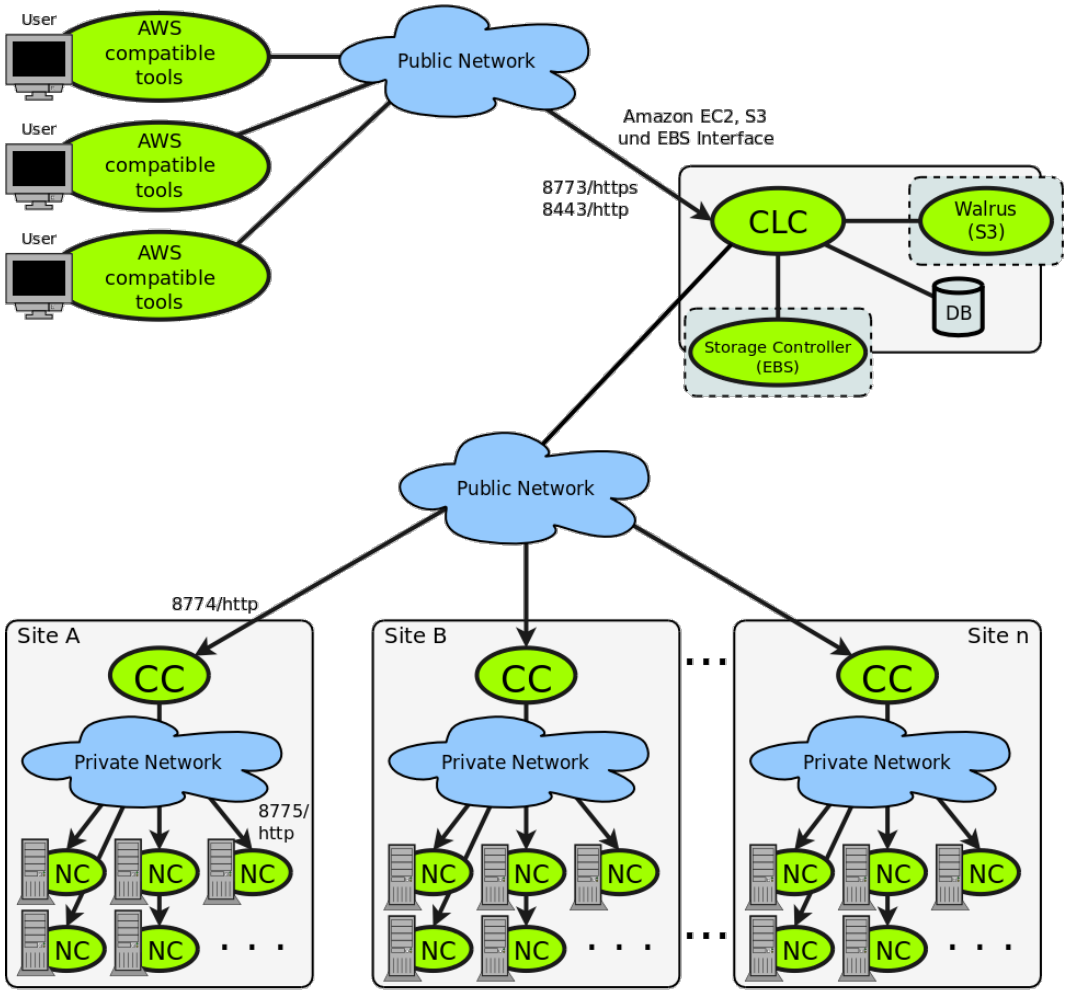


- **E**lastic **U**tility **C**omputing **Ar**chitecture for **L**inking **Y**our **P**rograms **T**o **U**seful **S**ystems
- One of the most popular Private Cloud IaaS solutions
- May 2008: Version 1.0
- August 2010: Version 2.0
- Emulates the most popular AWS services
  - Fully API compatible to Amazon EC2
  - Includes „*Walrus*“, a S3 compatible storage service
  - Includes „*Storage Controller*“, an EBS compatible storage service
- Easy to install via Ubuntu 10.04 („*Lcid Lynx*“) LTS Server
- Difficult to realize a high availability level
  - Source code looks a bit *obscure*
  - Still some bugs
  - Implementation of additional features is difficult
  - Lots of different log files distributed over the nodes
- Open Source version lacks some features of the Enterprise Edition
  - e.g. support for VMware ESX(i)
- Supports Windows instances since version 2.0

Eucalyptus – Components

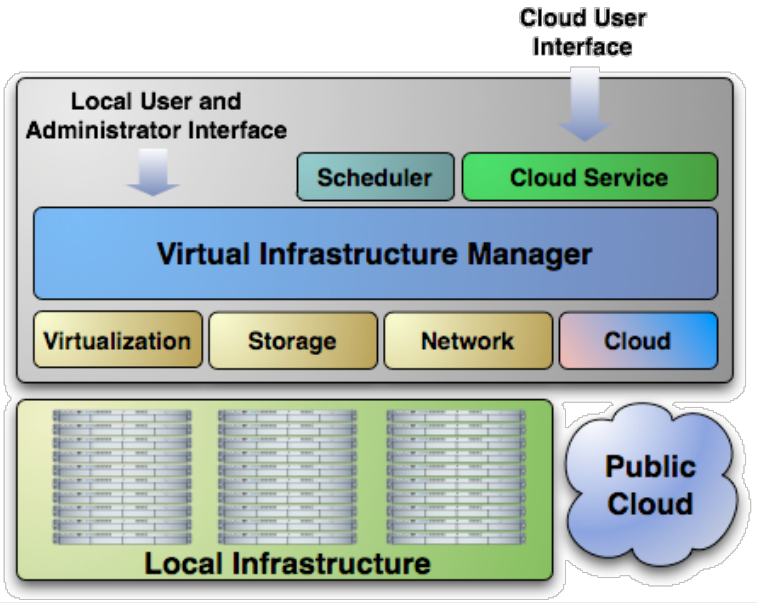
http://open.eucalyptus.com

- Cloud Controller (CLC)
  - Operates like a meta scheduler
  - Collects resource information from the CCs
- Cluster Controller (CC)
  - Schedules the distribution of virtual machines to the NCs
  - Collects free resource information from the NCs
- Node Controller (NC)
  - Runs on every worker node in the cloud
  - Xen hypervisor or KVM running
  - Provides resource information to the CC
- Walrus
  - S3 compatible storage service
- Storage Controller
  - EBS compatible storage service



OpenNebula

- EC2 compatible resources and resources from ElasticHosts can be used via remote (=> Hybrid Cloud)
- Only a small part of the EC2 API implemented since OpenNebula 2.0 Beta1
  - describe images
  - describe, run, reboot und terminate instances
- Nodes can be grouped
  - Important for HPCaaS and network latency (e.g. MPI)
- Trivial architecture
  - Easy to implement additional features
  - Easy to debug because of central log data
- No storage service included
- Supports Windows instances
- Software is still buggy



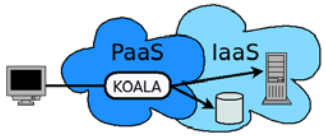
Ways to work with Public and Private Clouds

- All existing tools for cloud services face several advantages and drawbacks
- **Online tools**
  - **AWS Management Console** is in line with just Amazon's cloud servics. It is impossible to configure it in a way to work with Private Cloud services
  - **Ylastic** offers support for most AWS services and Eucalyptus infrastructures but not e.g. Nimbus
  - As the access keys are stored with the provider, the customer needs to trust the provider of the tool regarding privacy and availability
- **Browser-Plugins**
  - **ElasticFox** and **Hybridfox** only work with the Firefox browser
  - Require a local installation, a fact that does not reflect the cloud paradigm very well
- **Command-line tools**
  - **AWS tools** offered by Amazon only support the AWS public cloud offerings
  - **Euca2ools** from the Eucalyptus project support both, public and private cloud services
  - Require a local installation
  - Lack ease of use as they implement no graphical user interface (GUI)

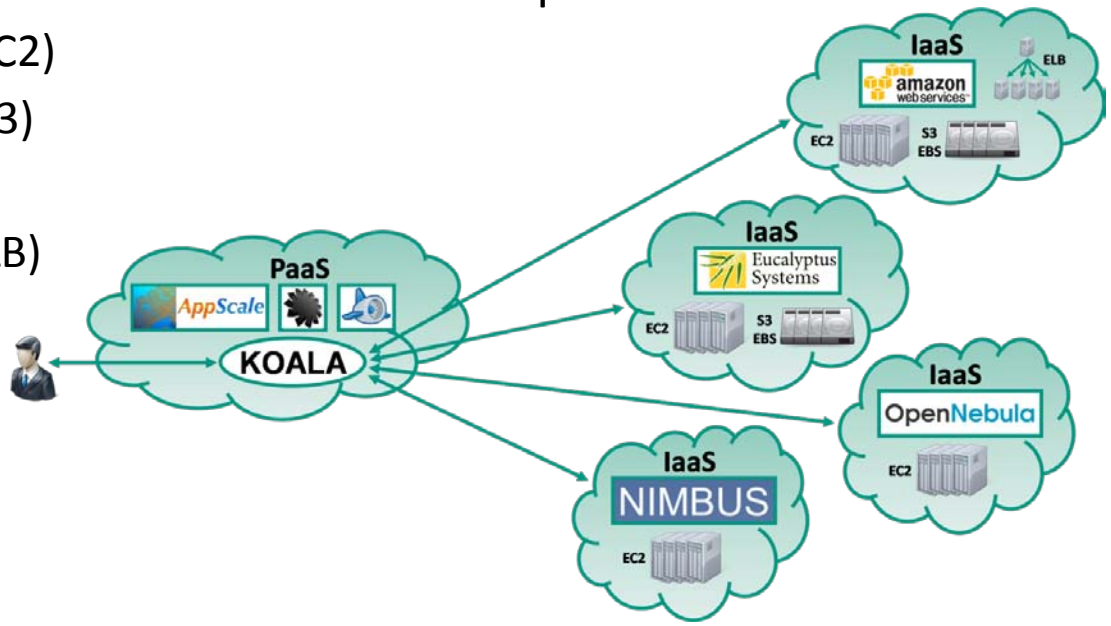
We need a tool that integrates public and private cloud services from different providers!

KOALA

http://koalacloud.appspot.com



- Software service, designed to assist working with Public and Private Cloud services that are compatible to the AWS
- KOALA helps interacting with cloud services that implement the APIs of
  - Elastic Compute Cloud (EC2)
  - Simple Storage Service (S3)
  - Elastic Block Store (EBS)
  - Elastic Load Balancing (ELB)
- Support for
  - Amazon AWS
  - Eucalyptus
  - Nimbus
  - OpenNebula



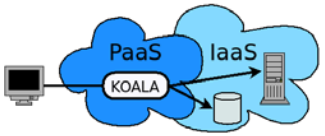
- KOALA itself is able to run inside the Public Cloud platform (PaaS) Google App Engine and inside Private Cloud platforms with AppScale or typhoonAE
- KOALA is Open Source (Apache License 2.0)
- Project site with source code and documentation:
  `http://code.google.com/p/koalacloud/`

Remember psDooM? (2000)

- psDooM is based on id Software's Doom
  - **`http://psdoom.sourceforge.net`**
- It is a process monitor and manager for Unix/Linux systems
- It can be considered a graphical interface to **ps**, **renice**, and **kill**
- The monsters represent processes currently running on your machine
- Killing a monster sends a **kill -9** to the associated process
- What happens to the game when init is killed?
  - The same can be done with KOALA!

KOALA

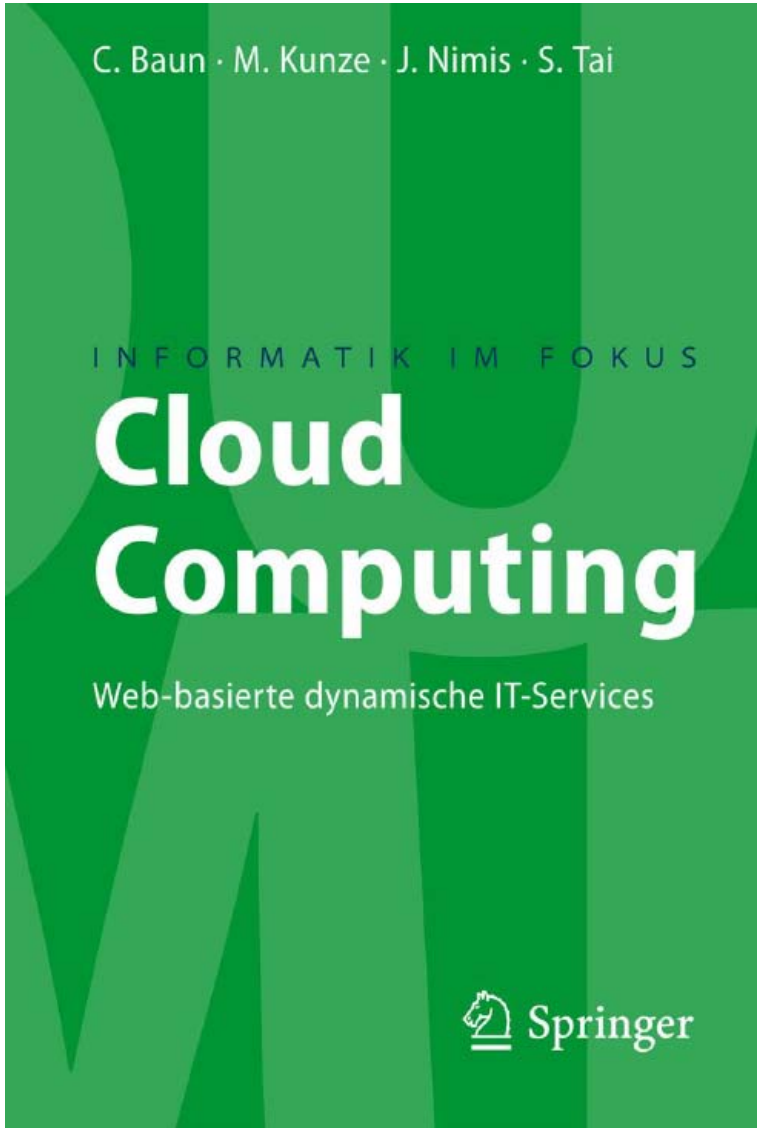
http://koalacloud.appspot.com



- Interesting scenario
  - When KOALA runs inside a Private Cloud PaaS (AppScale or typhoonAE) …
  - … and this PaaS runs inside a Public or Private Cloud IaaS (EC2 or Eucalyptus), …
  - … it is possible to work with the cloud services from inside

Advertisement

- First cloud computing book in German language
- Only € 14,95
- English version is work in process



Vielen Dank für Ihre Aufmerksamkeit
Thank you for your attention