int.eu.grid

http://www.interactive-grid.eu

Introduction to gLite and GridSolve

Marcus Hardt
Forschungszentrum Karlsruhe

Marcus.Hardt@iwr.fzk.de

Outline

- Different grid solutions:
  - EDG/gLite
  - Service Oriented Architectures
  - GridSolve
- GridSolve in int.eu.grid
- Demo of a solution
- Scaling considerations

Marcus.Hardt@iwr.fzk.de

Different grid solutions: EDG/gLite

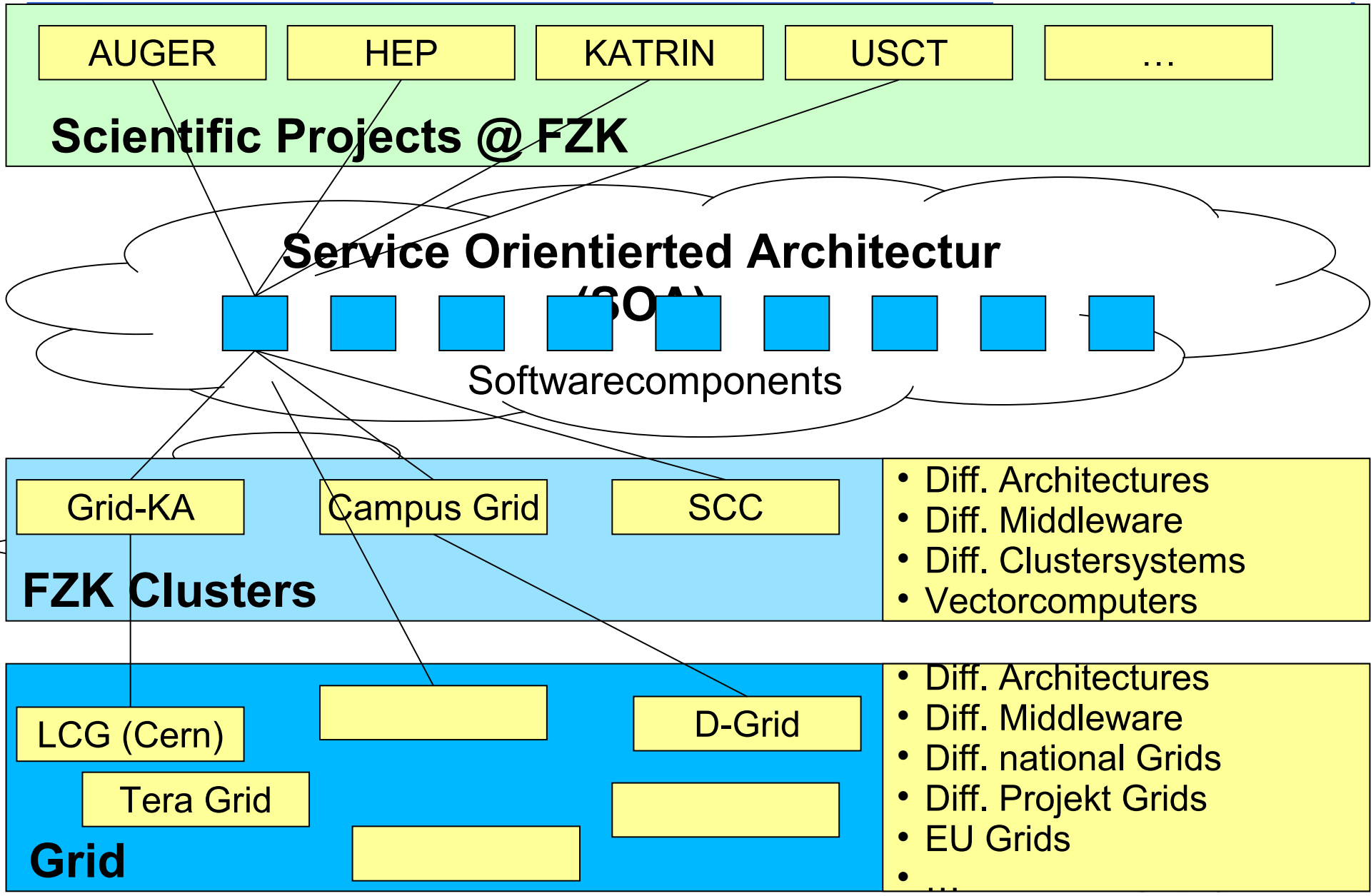
- High Energy Physics community
- Use cases
  - Data Production, Parameter sweeps
    - 10.000s of jobs, creating 100s of TB
  - Analysis on previously stored data
    - 10.000s of jobs, reading 10s of PB
- Current Infrastructure
  - 177 Centers
  - 35.000 CPUs
  - 13 PB Storage

- Interctive grid
  - Interactive extensions
  - MPI extensions
    - Inter-Job Communication across clusters
  - Resources for USCT

Current Infrastructure
- 12 Centers
- 303 CPUs
- 35 TB Disk

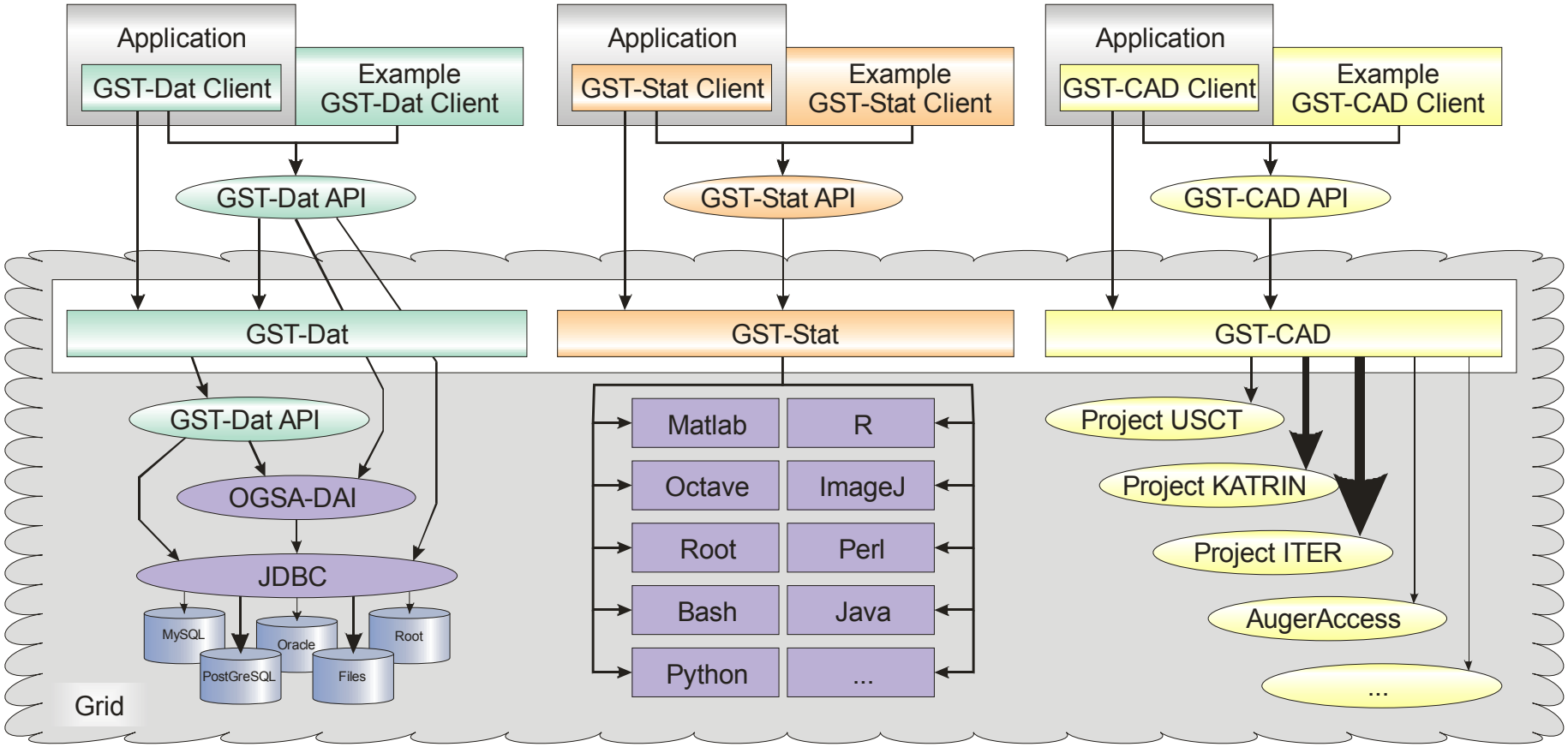
Marcus.Hardt@iwr.fzk.de

Different grid solutions: SOA



Different grid solutions: SOA

Grid Services Toolkit for Process Data Processing



Marcus.Hardt@iwr.fzk.de

Different grid solutions: GridSolve

- Self Consistent:
  - Resource Management
  - Monitoring
  - Data management
- Interface for C, Fortran, Java, Matlab, Mathamatica, ... for remote method invocation (RMI):

```
result = analysis (x, y);
result = gs_call ('analysis', x, y);
```

- "analysis" must be available in object code

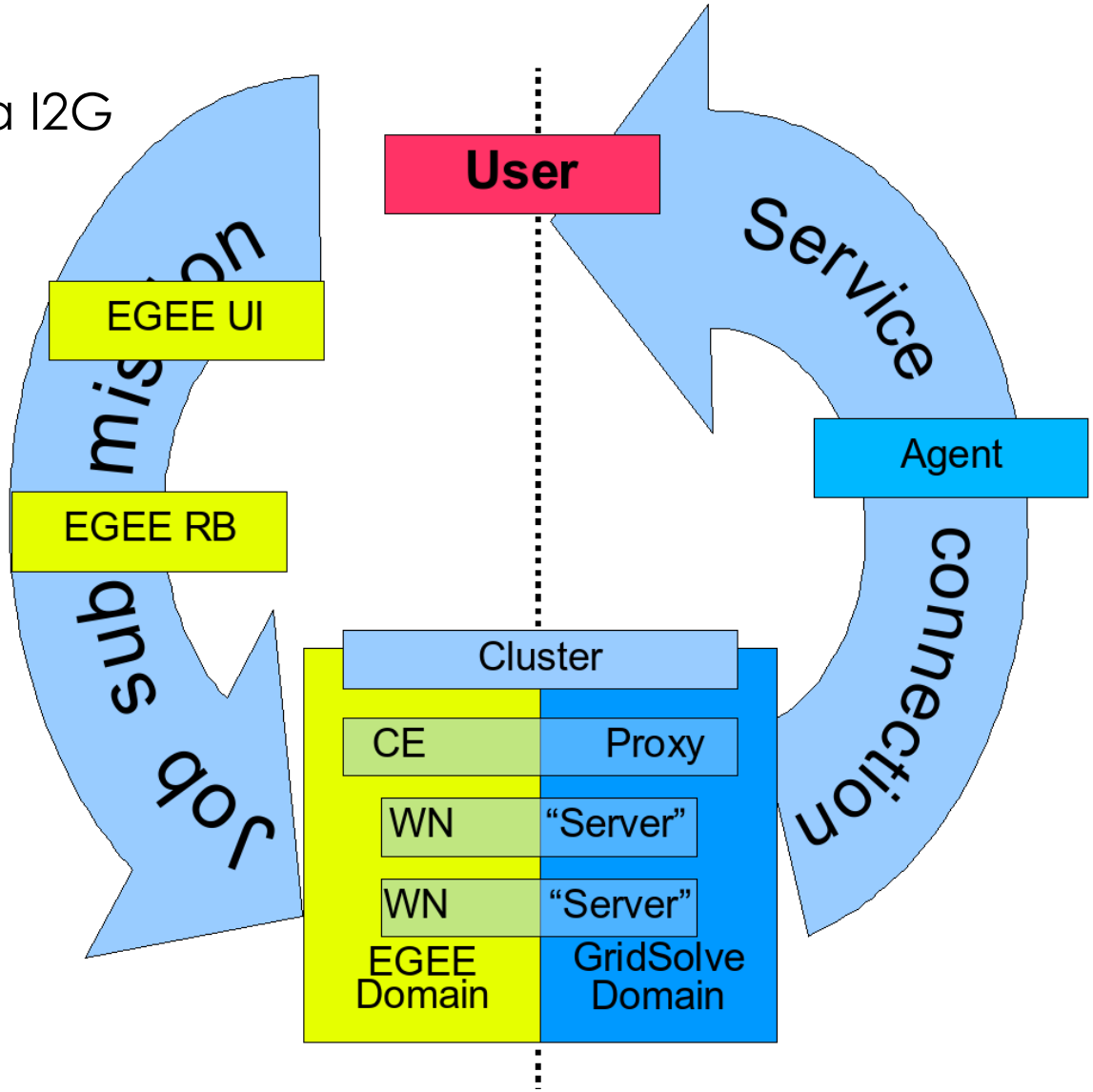
  => C, CPP, Fortran

  => Not: Java
- Asyncronous calls + "call farming" available

Marcus.Hardt@iwr.fzk.de

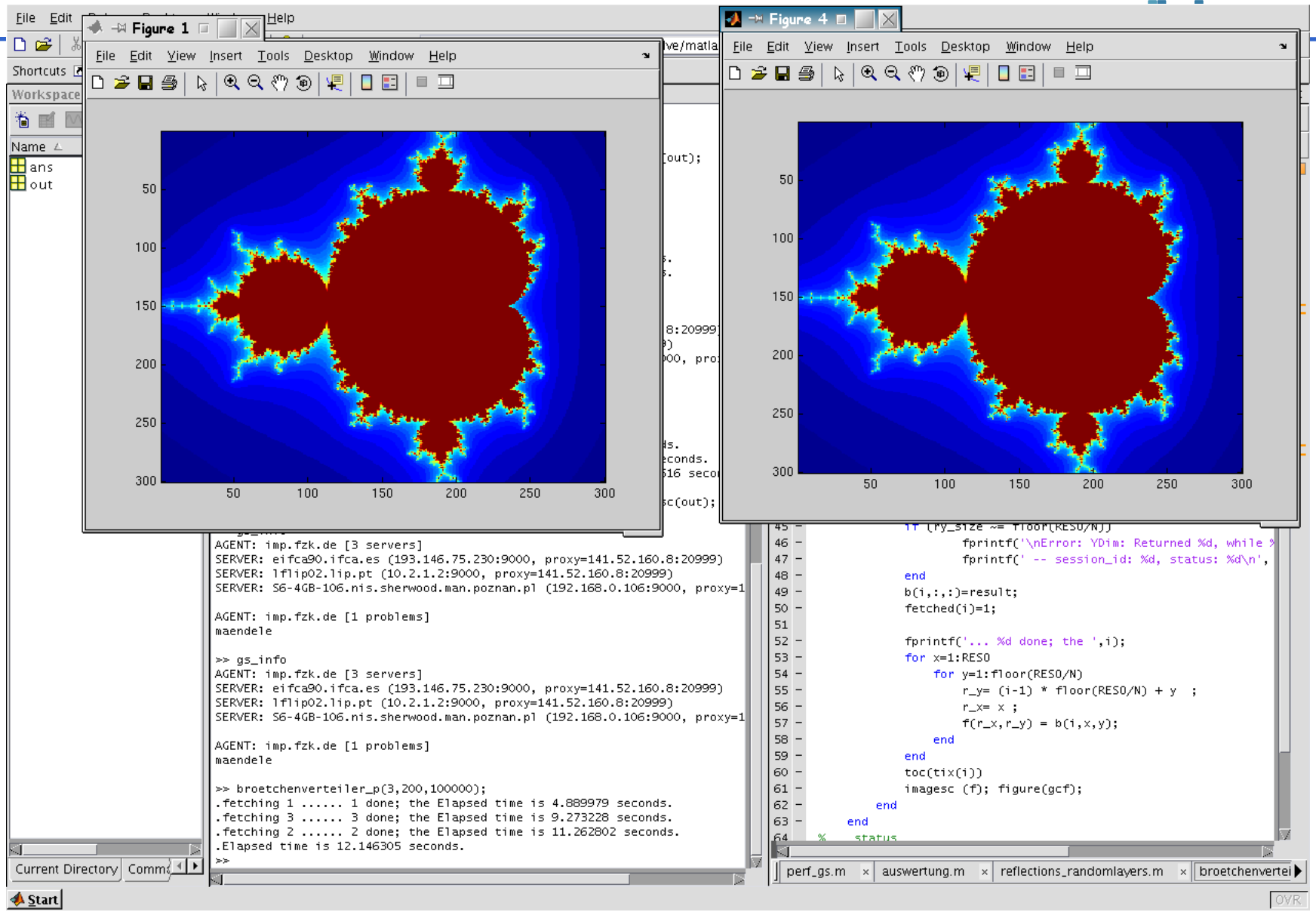
GridSolve on top of int.eu.grid

- GridSolve
  - Submission and start via I2G
    - 4-200 times
  - Agent on outside host
  - Connectivity provided by a proxy host
  - Client in Matlab / C
    - Connects to agent
    - Connects (via proxy) to resources
- Example
  - Fractal calculation
    - Resolution=> Data
    - # Iterations => CPU
    - # Resources => Distribution



Marcus.Hardt@iwr.fzk.de

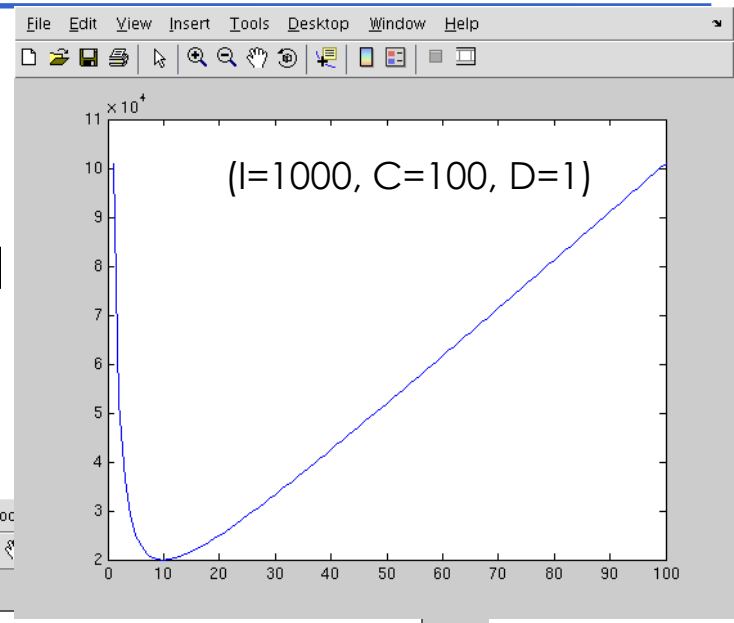
Demonstration

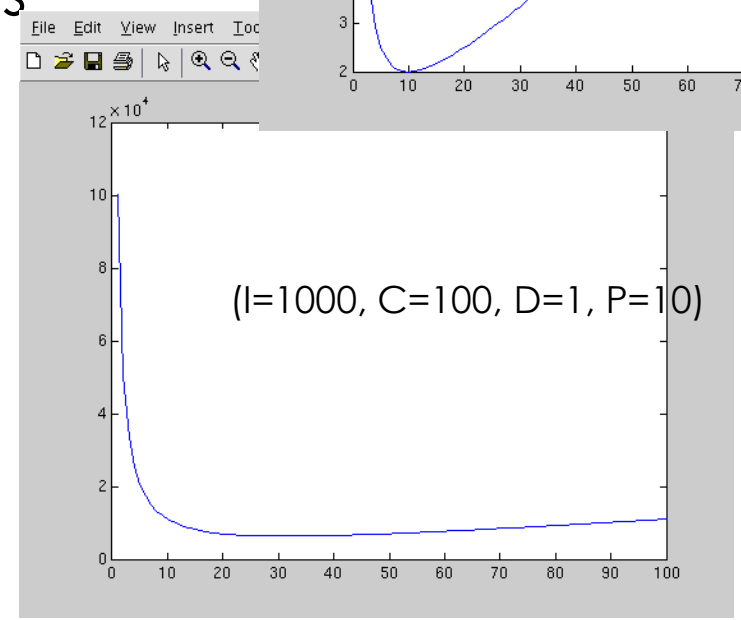


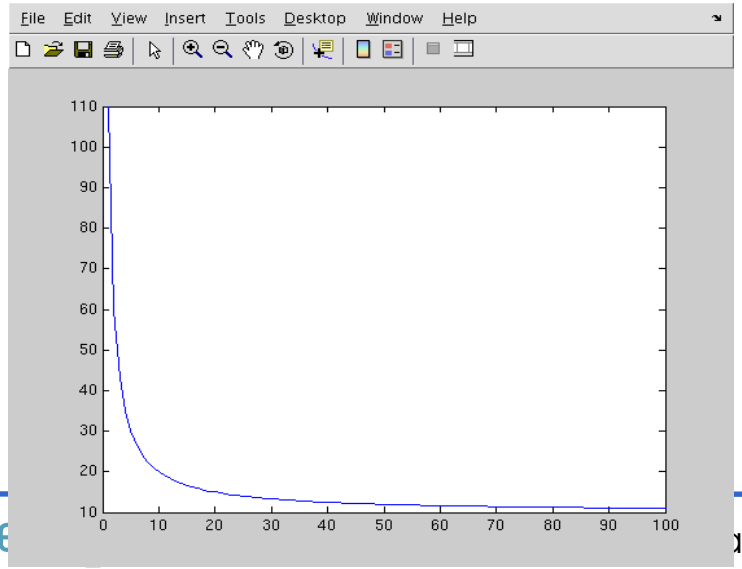
Marcus.Hardt@iwr.fzk.de

Scaling considerations

- Every Iteration can be distributed perfectly
  Update after every iteration:
  - S=I*C/N + I*N*D (I=1000, C=100, D=1
- Run P Iterations simultanously:
  - S=I*C/N + I*N*D/P
- Direct solution, return ellipsis information:
  - S=C/N + D







Backup Slides

Marcus.Hardt@iwr.fzk.de

What is a grid?

- CPU <=> Network <=> Data
  - Interconnected by Information Systems
  - Unified authentication + authorisation
  - Accounting
- Added values
  - Know where resources are available
    - Resource Brokerage
      => Send the job to where the data is
  - Data Management
    - When moving data, keep on “old” copy cached
    - Remember where which data is
    - => Replica Management

Marcus.Hardt@iwr.fzk.de

Beispiele

- **Physik** (CERN):
  - 177 Zentren
  - 34.286 CPUs
  - 13065 TB Speicherplatz
- **Google** (2006): 450.000 PCs in 3000 Centers worldwide
- Seti@home
  - 415,516 Hosts
  - 26,884 Teams
  - 258 Countries
- **Interactive European Grid (int.eu.grid):**
  - 11 Zentren
  - 190 CPUs
  - 29.6 TB Speicherplatz

int.eu.grid

- Focused on
  - Compatibility with EGEE
  - Interactivity on the grid
  - Easy access for the user (friendly user interface)
  - MPI on the grid

- My approach:
  - Use **gridsolve** to connect grid to **matlab**
  - Matlab is a well known Problem solving Environment

Marcus.Hardt@iwr.fzk.de

Basic grid middleware

- Globus-2 (old globus)
  - One of the first grid middlewares
  - Basic auth* and monitoring => Single sign on
- EDG/gLite
  - Infrastructure
  - Advanced Tools:
    - AAA
    - Data Management (Replication)
    - Configuration
- EGEE
  - = EDG/gLite
  - + Organisation

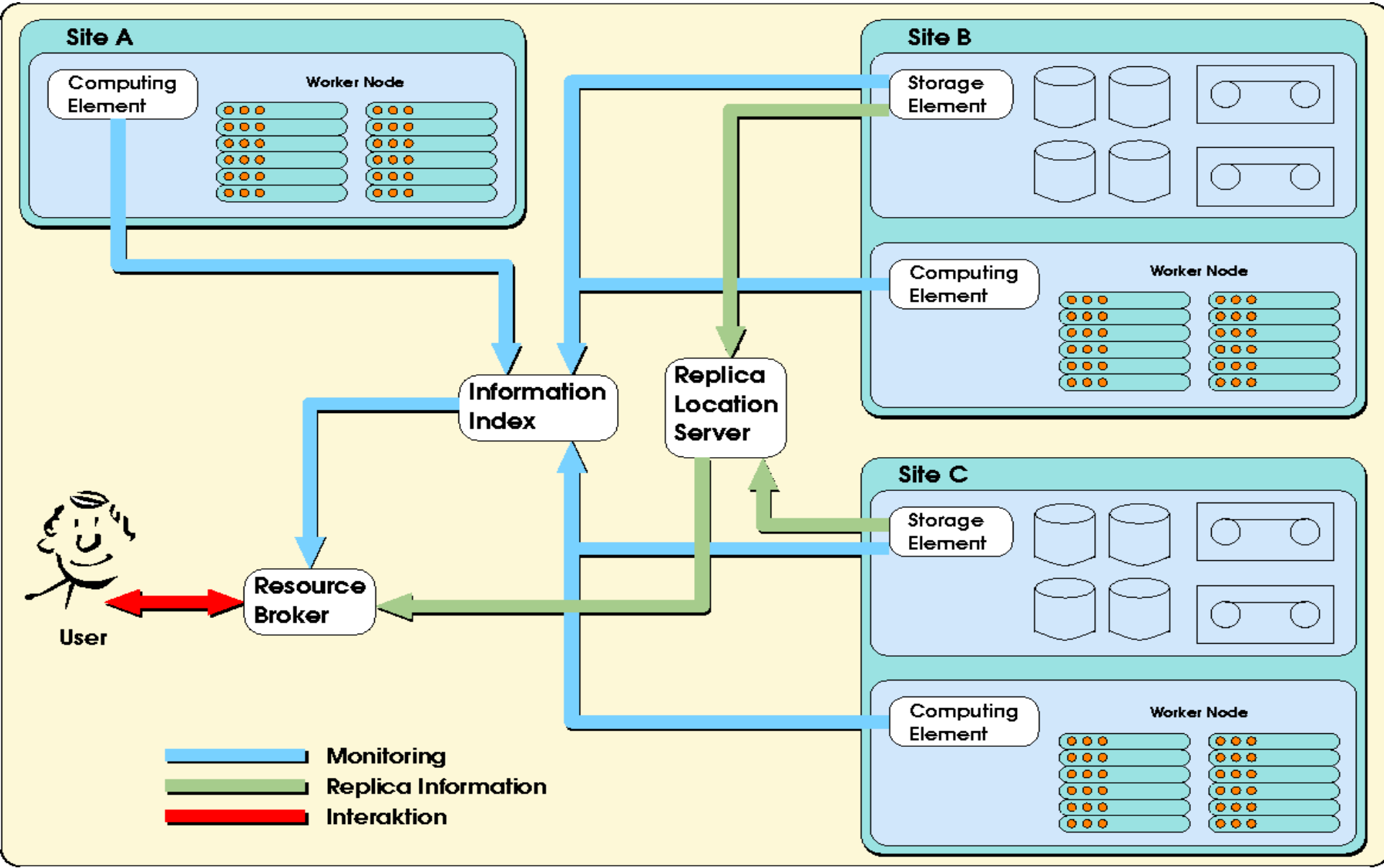
Marcus.Hardt@iwr.fzk.de

Basic grid middleware

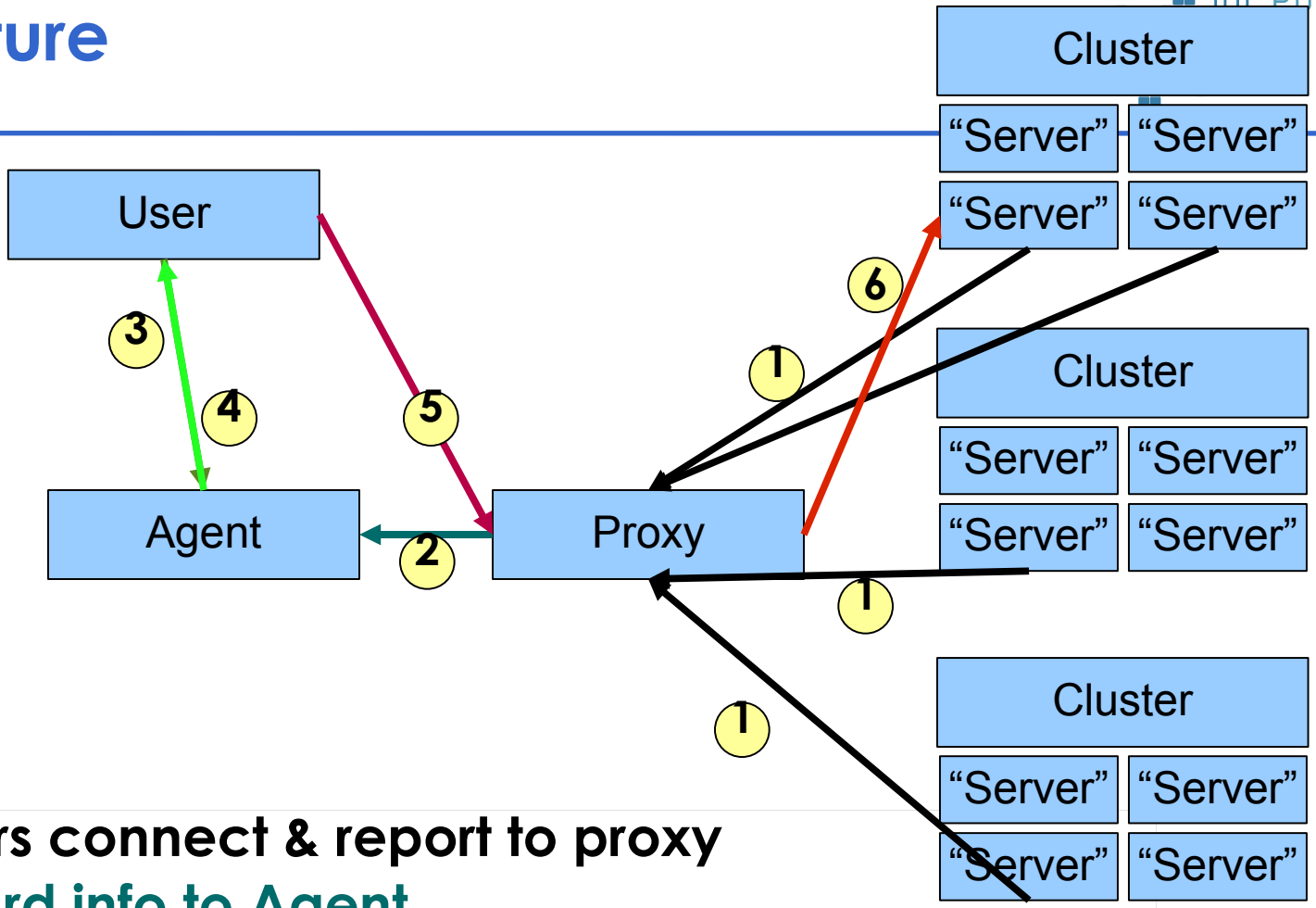
- Interactive Grid (I2G)
  - = gLite
  - + MPI support
  - + Interactive extensions
- Gew Globus-4
  - = Globus-2
  - +Webservices based

Marcus.Hardt@iwr.fzk.de

EDG-Based Grid

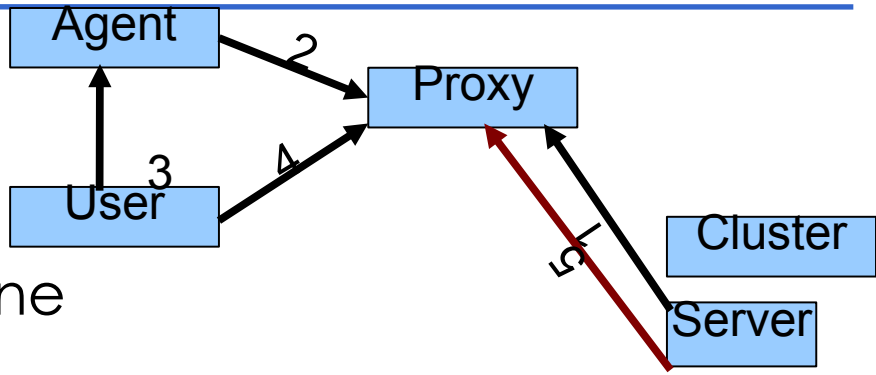


Architecture



- 1: Servers connect & report to proxy
- 2: Forward info to Agent
- 3: Query Agent for resource
- 4: Give ONE server back
- 4: Send problem via proxy
- 6: To the server6
- …

Mixing it with EGEE



- Proxy + Agent
  - Dedicated host with public IP access
  - Currently non-EGEE machine
- User:
  - My laptop with matlab installed
- Servers:
  - edg-job-submit to WNs

- Combining strengths:
  - EGEE: Resources
  - gridsolve: accessibility
  - Matlab: prototyping / development

Marcus.Hardt@iwr.fzk.de