HPC as a SaaS: The CloudBroker Solution

Master of Science in Engineering
Cloud Computing
HES-SO, Lausanne, April 23, 2015

Dr. Wibke Sudholt
CloudBroker GmbH
Heinrichstrasse 267, CH-8005 Zürich, Switzerland
Phone: +41 44 515 21 70
Email: info@cloudbroker.com
Web: http://www.cloudbroker.com

CloudBroker

Overview

- Cloud computing for HPC
  - Cloud computing
  - High performance computing
  - HPC in the cloud
- CloudBroker solutions
  - Company
  - CloudBroker Platform
  - CloudBroker AppCenter
- Example use cases
  - Compute-intensive applications
  - External user interfaces
  - Large collaborations
- Hands-on tutorial

Cloud Computing for HPC

Cloud Computing

NIST Definition of Cloud Computing

- From the US National Institute of Standards and Technology (2011)

- Essential characteristics:
  - On-demand self-service
  - Broad network access
  - Resource pooling
  - Rapid elasticity
  - Measured service

- Service models:
  - Software as a Service (SaaS)
  - Platform as a Service (PaaS)
  - Infrastructure as a Service (IaaS)
- Deployment models:
  - Private cloud
  - Community cloud
  - Public cloud
  - Hybrid cloud

Source: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Important Aspects

- Cloud computing primarily is not a new technology, but a new business and delivery model
- Cloud computing represents another wave in commoditization, automation and industrialization of computer technology and services
- Users can choose at which level and through which interface they would like to consume cloud services
- Different service building blocks can independently work together through their interfaces in the cloud

High Performance Computing

CloudBroker

High Performance Computing (HPC)

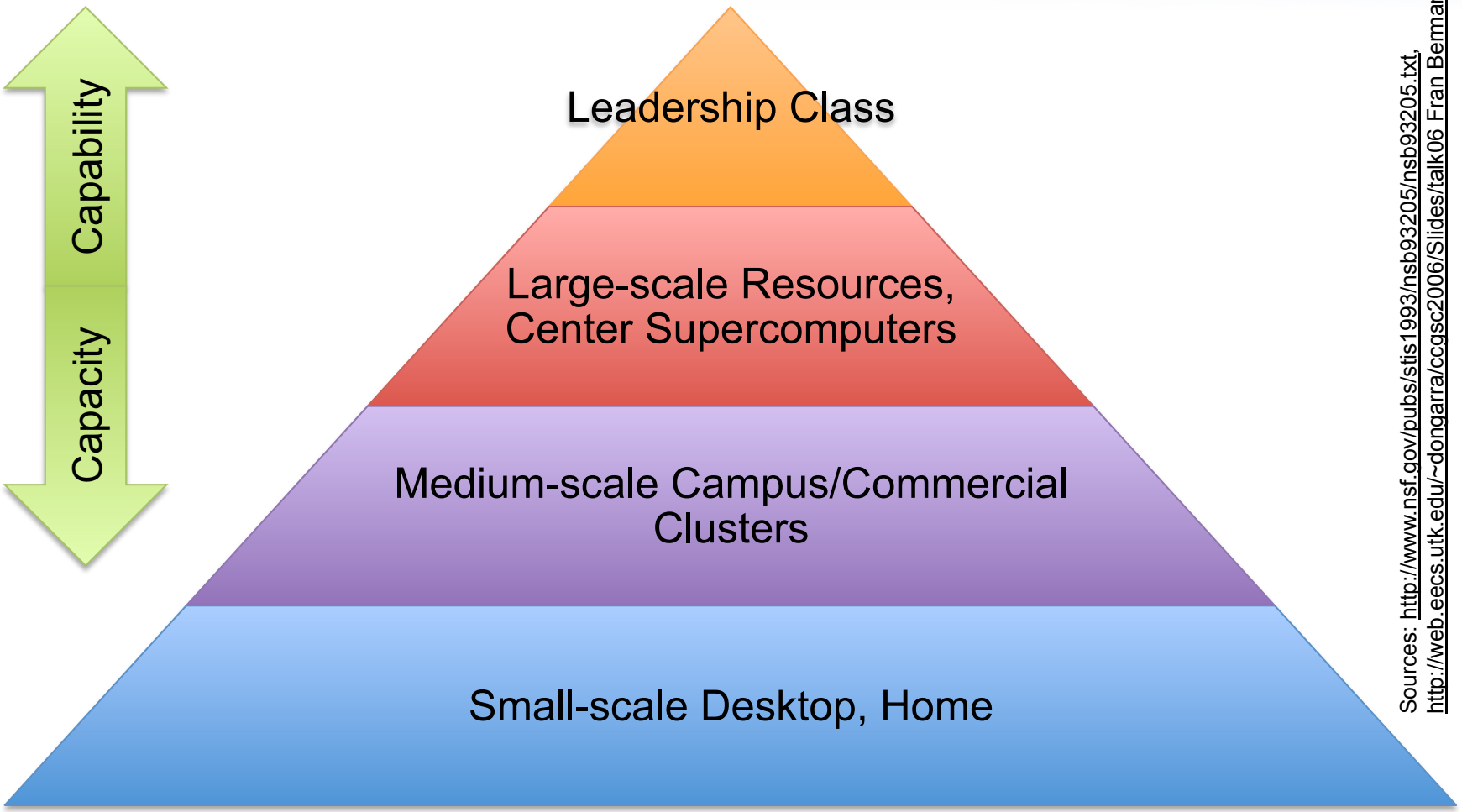
- Computing with very high demands for computing power, memory access, network speed, storage capacity, etc.
- Typically employing computer clusters and supercomputers
- Usually used to solve advanced or research problems by computer-based modeling, simulation or analysis
- Often highly specialized and optimized scientific and technical applications with complex algorithms or big data
- Compute-intensive or data-intensive, high performance or high throughput calculations
- Utilization of parallel and distributed computing and storage technologies (MPI, etc.)
- Mainly batch-oriented, non-interactive command line software running for hours, days or weeks on many cores or nodes
- Mostly project-based or periodical usage patterns
- Increasing performance and importance

Top 500 List November 2014

Rank	Site	System	Cores	Rmax (TFLOPS/s)	Rpeak (TFLOPS/s)	Power (KW)
1	National Super Computer Center in Guangzhou China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
…						
6	Swiss National Supercomputing Centre (CSCS) Switzerland	Piz Daint - Cray XC30, Xeon E5-2670 8C 2.600GHz, Aries interconnect, NVIDIA K20x Cray Inc.	115,984	6,271.0	7,788.9	2,325
…						
16	HLRS – Höchstleistungs-rechenzentrum Stuttgart Germany	Hornet - Cray XC40, Xeon E5-2680v3 12C 2.5GHz, Aries interconnect Cray Inc.	94,608	2,763.0	3,784.3	1,512
…						

Source: http://www.top500.org/list/2014/11/

Branscomb Pyramid (1993/2006)

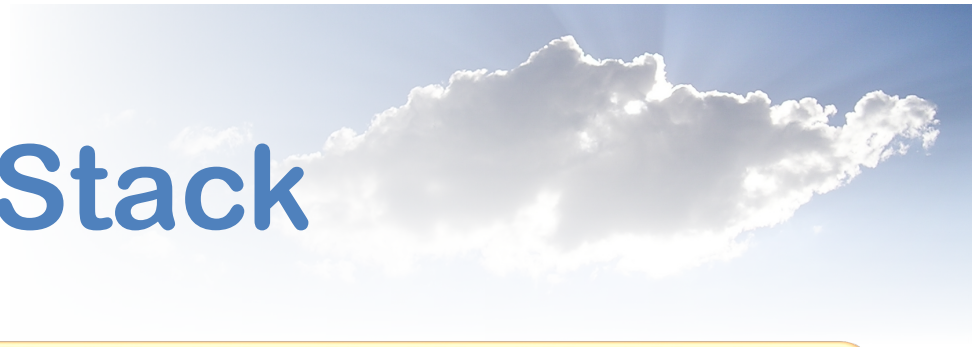


Sources: http://www.nsf.gov/pubs/stis1993/nsb93205/nsb93205.txt,
http://web.eecs.utk.edu/~dongarra/ccgsc2006/Slides/talk06 Fran Berman.ppt

HPC Application Areas

- Particle physics
- Astronomy
- Computational chemistry and molecular modeling
- Bioinformatics and drug design
- Material science
- Simulations in engineering and manufacturing
- Process optimization
- Geographical data processing
- Weather and climate simulations
- Insurance risk analysis
- Financial market calculations
- Military research
- Image processing
- Rendering
- … and others

HPC Software Stack



Application Software

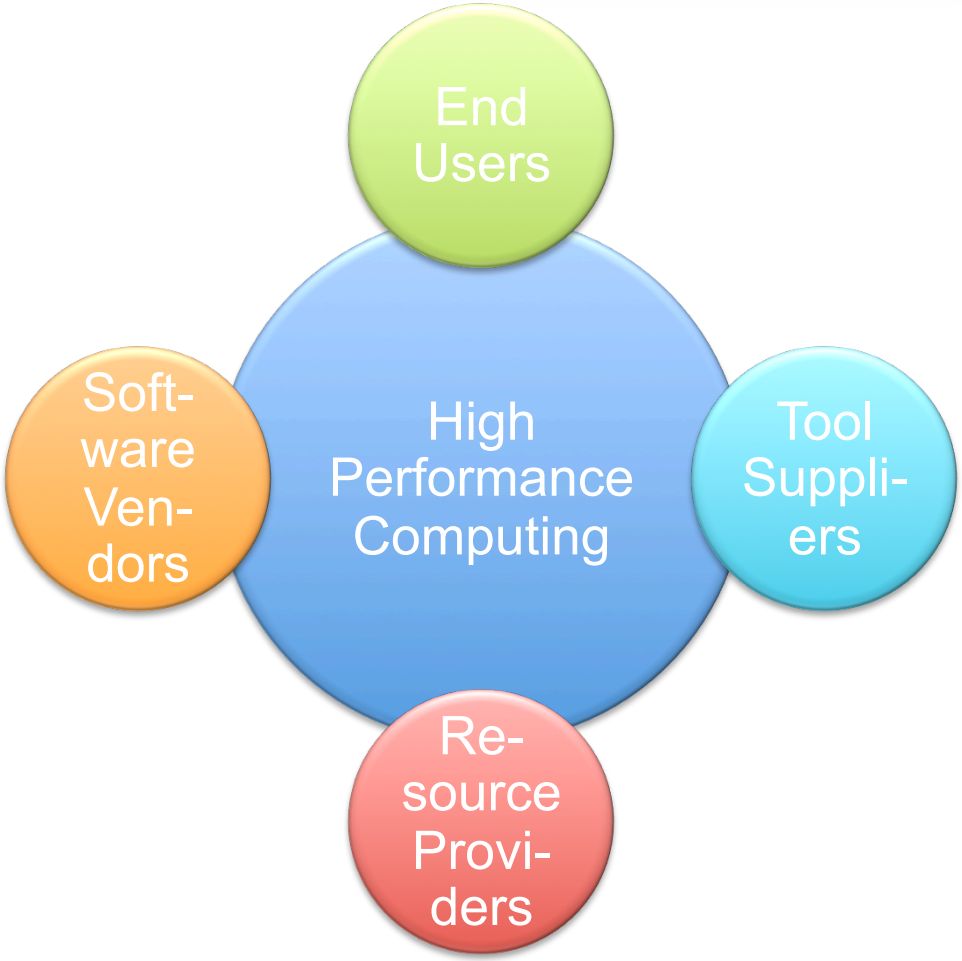
Parallelization Tools

Management Middleware

Operating System

CloudBroker

HPC Stakeholders



CloudBroker

Problems of Traditional HPC

- HPC computer infrastructure, middleware tools and application software:
  - Require expert knowledge
  - Are expensive, time-consuming and complex to buy, set up, use and maintain
  - Are hard to integrate with existing systems and processes
  - Often operate at capacity limit

⇒ Hardly accessible, affordable or marketable for:
  - Individual research groups, small institutions or SMEs
  - Special application purposes, short-term projects, etc.

Advantages of Cloud for HPC

+ Immediate access to infrastructure and/or applications on demand
+ Availability of hardware and/or software not existing in-house, no need to maintain own resources
+ High scalability, flexible adaptation to requirements
+ Temporary, non-binding utilization, no long-term contracts or licenses
+ Self service by actual application end users
+ Pay-per-use with minimal initial investment
+ New business possibilities for software vendors, tool suppliers and resource providers

CloudBroker

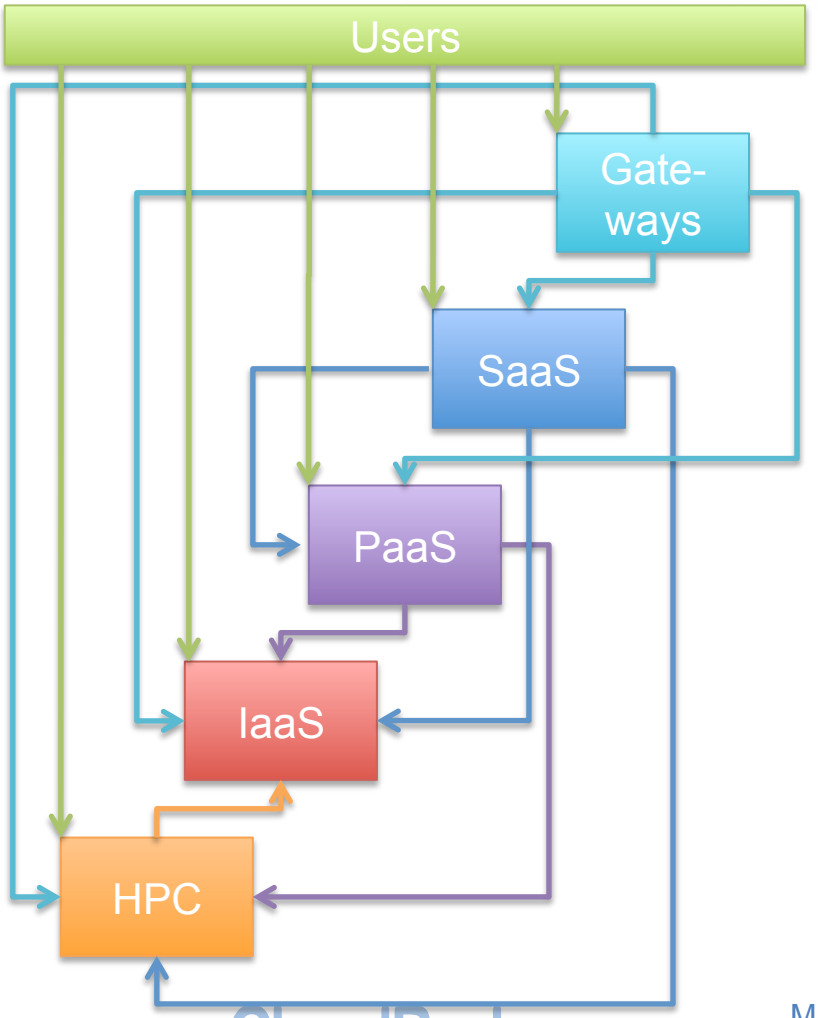
Challenges of Cloud for HPC

- Hard to get a market overview
- Infrastructure, middleware and applications in the cloud are also complex, and dynamic scaling and billing even add to the complexity
- Limitations for some applications due to internet connection, virtualization and network performance constraints, also for special requirements or steady high use the overhead might be too big
- Concerns because of security, privacy, outsourcing, internationality, legal issues, missing standardization and potential vendor lock-in
- Resource providers, tool suppliers and software vendors have to adapt to the pay-per-use and self service business model

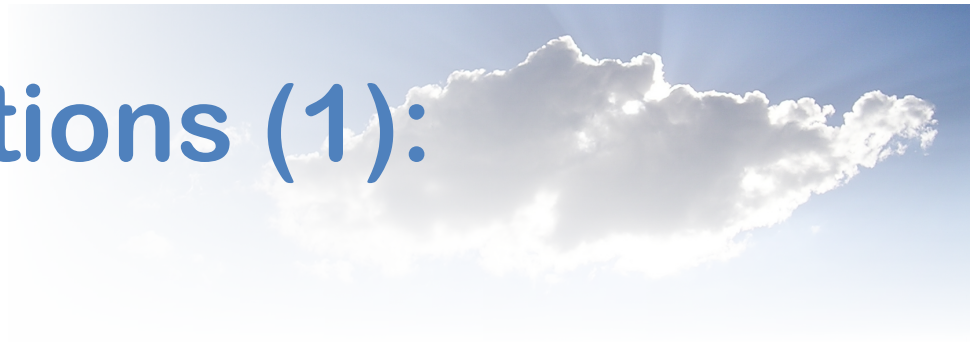
HPC in the Cloud

CloudBroker

HPC Cloud Computing Stack



HPC Cloud Solutions (1): IaaS



- On-demand access to compute and storage resources in the cloud:
  - Collection of usually virtual machines with different CPUs, memory, etc.
  - Additional storage space or system
  - Usually access through API and web interface
  - Payment often per hour usage based on machine type, etc.
- Example:
  - Amazon EC2 and S3

\+ Wide selection of machine types at different prices (from 1 to 36 CPUs, incl. GPUs etc., from cents to dollars per hour)

\+ Resources available very quickly (within minutes)

± Well suited for loosely-coupled parallel calculations, but not so suited for applications requiring high-speed network interconnects or bare-metal servers

– Need to assemble a cluster yourself

– Previous deployment of software required

HPC Cloud Solutions (2): HPC on Demand

- Pay-per-use access to clusters and supercomputers:
  - Typical HPC hardware with high-speed machines and network (e.g., Infiniband)
  - Usually access through batch queuing system
  - Additional storage space or system
  - Payment often per hour usage based on machine type, etc.
- Example:
  - HLRS

\+ Classical HPC infrastructure with corresponding high performance

\+ Cloud business model

± Users need to be familiar with HPC

– Usually not self service initially, often preparation necessary

– Previous deployment of software required

HPC Cloud Solutions (3): Application Portals

- Access to particular HPC applications as a service:
  - Individual software or group of related software that can be directly executed
  - Usually access through web interface or desktop application frontend
  - Often combined with preparation, analysis and visualization tools
  - Often special license conditions
- Example:
  - SCI-BUS gateways

\+ Easy access to specific HPC applications

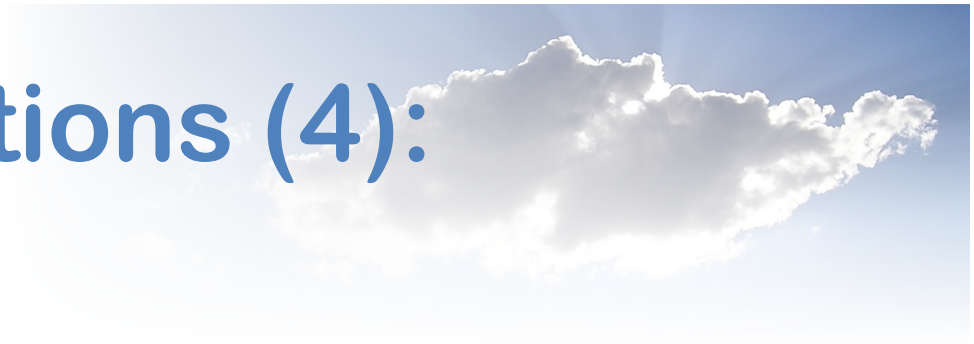
\+ Optimized software setup regarding performance and features

± Not necessarily pay-per-use

– Often intransparency or inflexibility regarding the used infrastructure resources

– Usually fixed, limited and heterogeneous application access with danger of vendor lock-in

HPC Cloud Solutions (4): One-Stop-Shop



- Marketplace for HPC applications:
  - Collection of different software that can be directly executed
  - Definition and management of underlying infrastructure resources as well as software deployment and access
  - Usually access through web interface and API
  - Payment often per hour usage based on selected infrastructure and software, etc., or subscription-based
- Example:
  - CloudBroker Platform and AppCenter

\+ Easy, uniform and pay-per-use access to multiple HPC applications

\+ Providers can offer optimized packages of infrastructure and software, and individual application interfaces can be built on top

\+ Users can compare and select from different applications and/or infrastructures

± Interfaces are different from classical HPC usage

– Might not be suitable for applications with very special requirements

CloudBroker

CloudBroker Solutions

Company

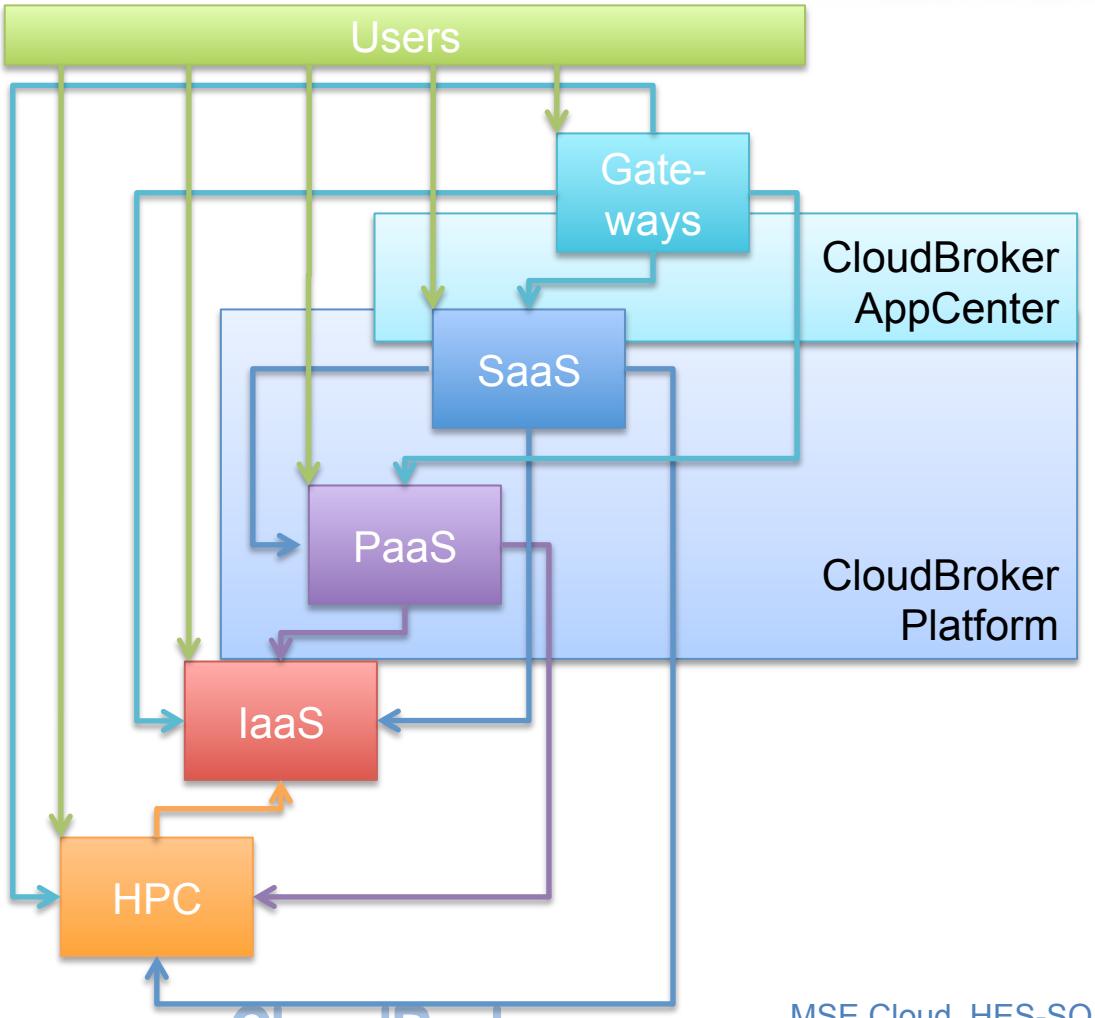
CloudBroker GmbH

- Founded in November 2008
- Office location in Zurich, Switzerland
- Originally spin-off of the ETH Zurich
- Was supported by the CTI Start-up program
- CEO: Nicola Fantini, CTO: Dr. Wibke Sudholt
- Development, support and operations in collaboration with ScaleTools AG
- http://www.cloudbroker.com

CloudBroker Solutions

- Goal:
  - Make compute-intensive scientific and technical applications easily available in the cloud by providing High Performance Computing Software as a Service (HPC SaaS)
- Products:
  - **CloudBroker Platform**: SaaS and PaaS for compute-intensive applications on different infrastructures
  - **CloudBroker AppCenter**: Web marketplace for modeling, simulation and analysis software
- Services:
  - Consulting, training and support with focus on cloud computing for scientific and technical applications

HPC Cloud Computing Stack



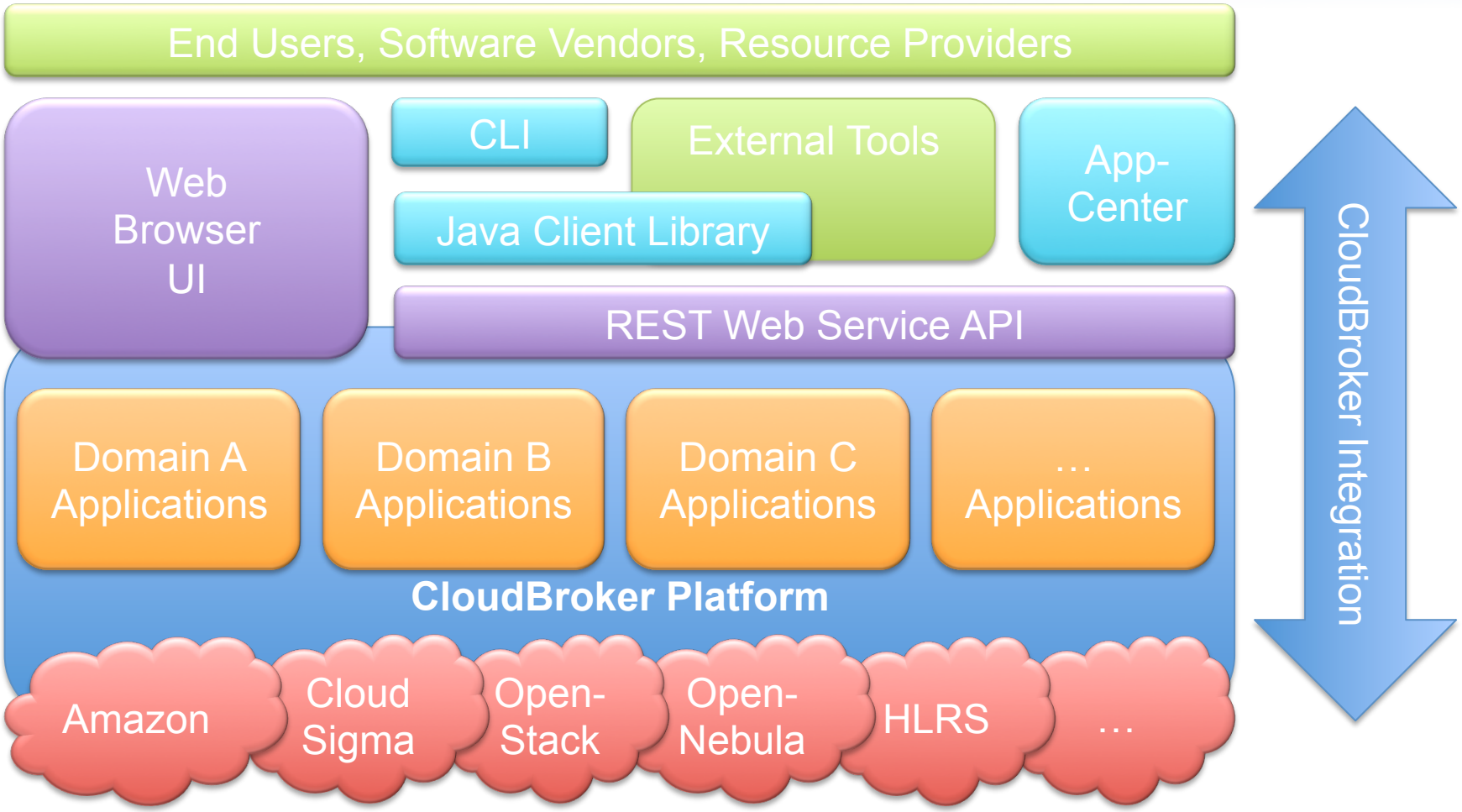
CloudBroker Platform

CloudBroker

CloudBroker Platform

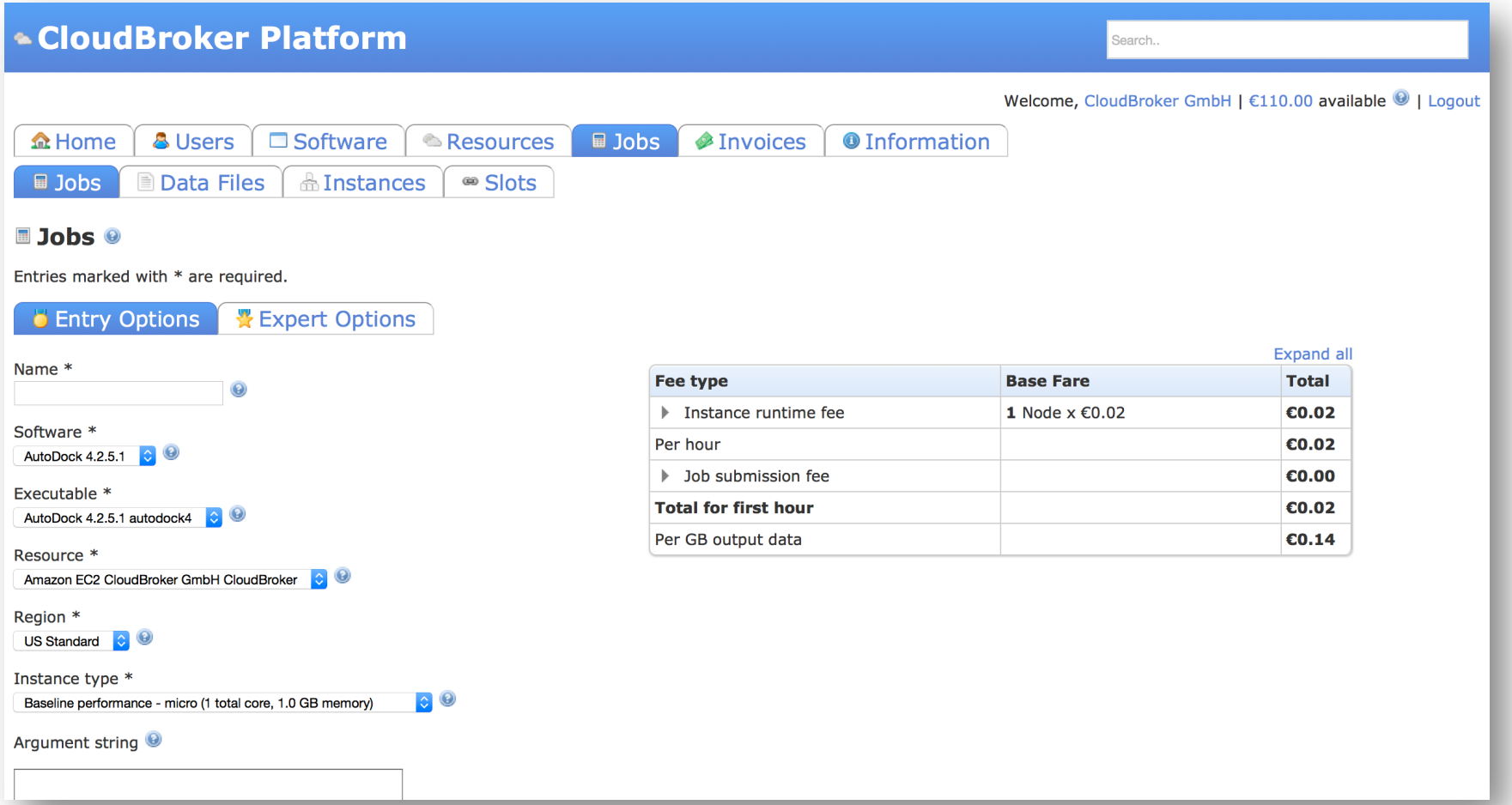
- Web-based application store and middleware for the deployment and execution of scientific and technical software on different computer infrastructures
- Uses IaaS from resource providers, offers PaaS for software vendors and SaaS for end users
- Management of users, resources, software and jobs
- Pricing, accounting, billing and payments
- Automation of software deployment, utilization of compute and storage resources, data transfer, job queuing and execution, etc.
- Users can use resources and software provided by others or register and deploy their own
- Browser, programmatic and command line access, can be used as frontend or backend, applications exposed as web service
- SSL transport layer encryption between all involved tiers
- Offered as public, hosted or in-house service or licensed software

Platform Architecture



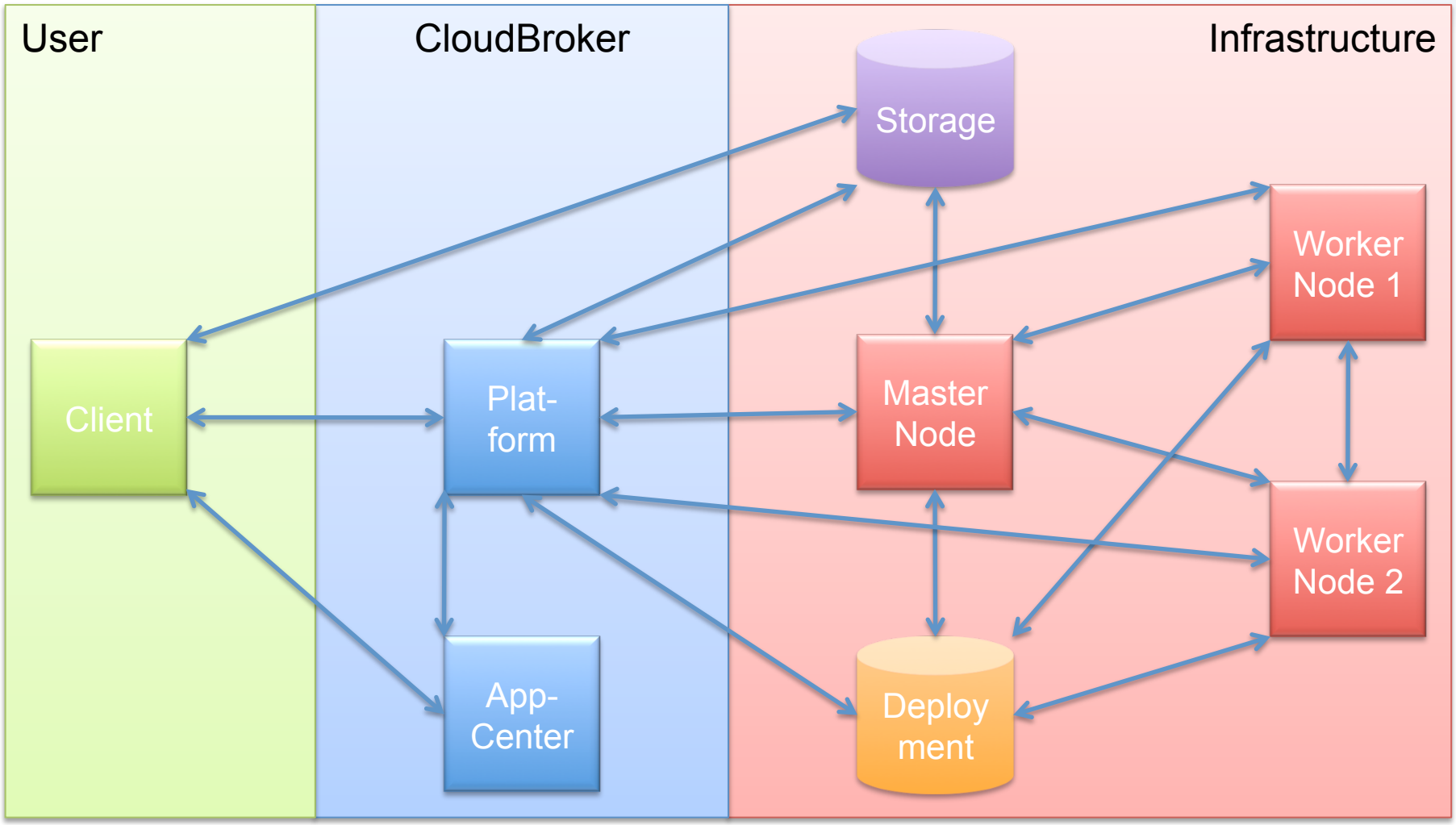
CloudBroker

Platform Screenshot



CloudBroker

Interaction of Tiers



CloudBroker

Implemented Cloud Adapters

- Compute:
  - Amazon EC2
  - CloudSigma
  - OpenStack EC2
  - OpenStack Nova
  - OpenNebula
  - Eucalyptus
  - HLRS NEC cluster
  - … more to be added
- Storage:
  - Amazon S3
  - OpenStack S3
  - OpenStack Swift
  - Ceph RADOS
  - Eucalytpus Walrus
  - SCP
  - … more to be added

Application Software

- Application requirements:
  - Scientific and technical applications from any domain, commercial or free / open source
  - Linux or Windows
  - Usually compute-intensive, batch and command line-oriented, but others are possible
  - Single-core or multi-core, single-node or multi-node (MPI, NFS, etc.)
- Application areas:
  - Biology, pharma and chemistry (Rosetta, BLAST, AutoDock, Gromacs, NWChem, etc.)
  - Engineering and manufacturing (ASCOMP TransAT, Elmer, OpenFOAM, Simul8, etc.)
  - Generic (shell scripts, R, etc.)
  - … and others

Platform Clients and APIs

Tools
- Tools from external providers
- Integration with the CloudBroker Platform via its APIs
- Examples: WS-PGRADE/gUSE, KNIME, ASCOMP TransAT

CLI
- Command Line Interface
- Job submission
- Open source at https://github.com/CloudBroker/cbp-cli

Java
- Application Programming Interface
- Java client library
- Open source at https://github.com/CloudBroker/cbp-java-api

REST
- Representational State Transfer
- Web service interface
- Documentation freely available on the CloudBroker website

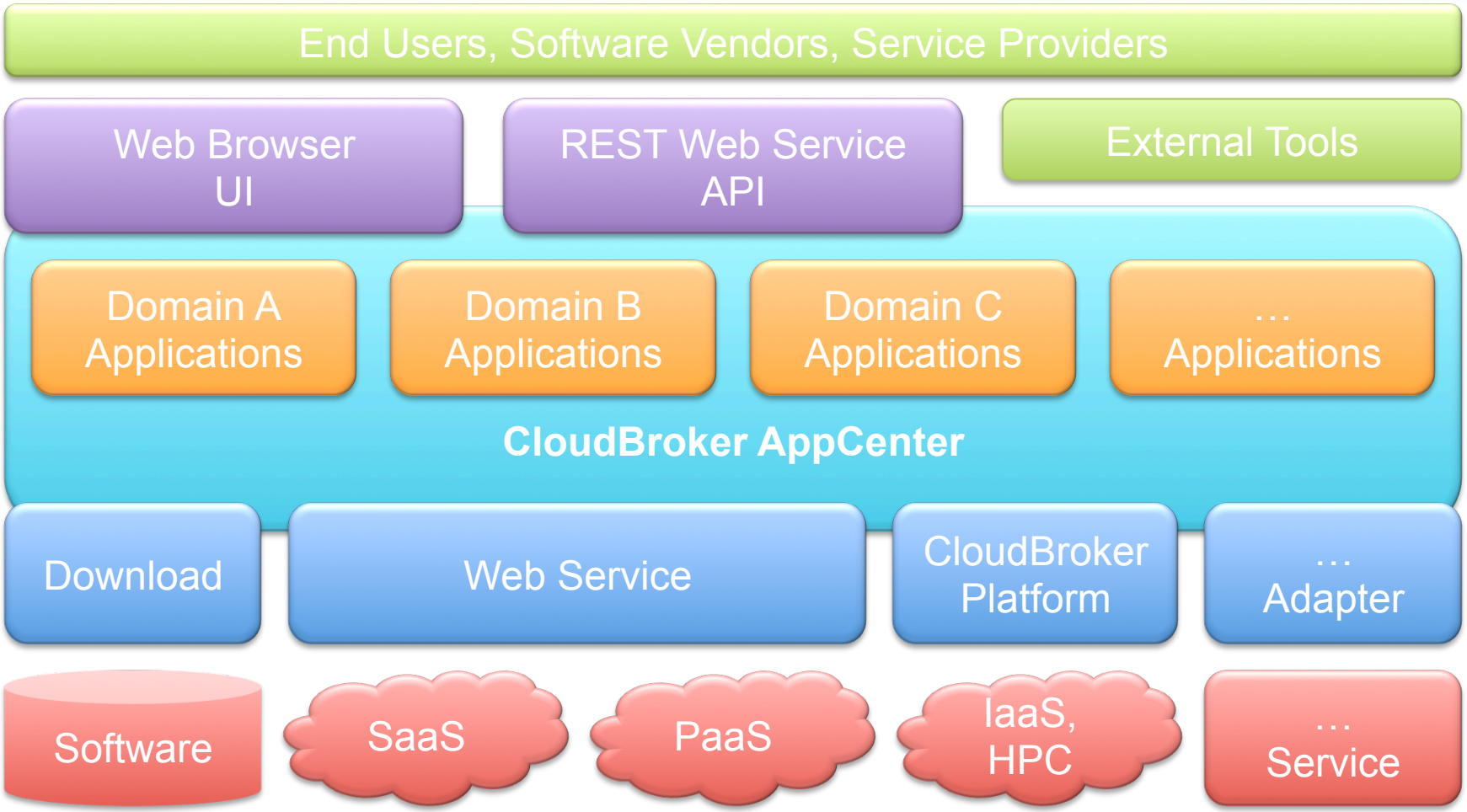
CloudBroker AppCenter

CloudBroker

CloudBroker AppCenter

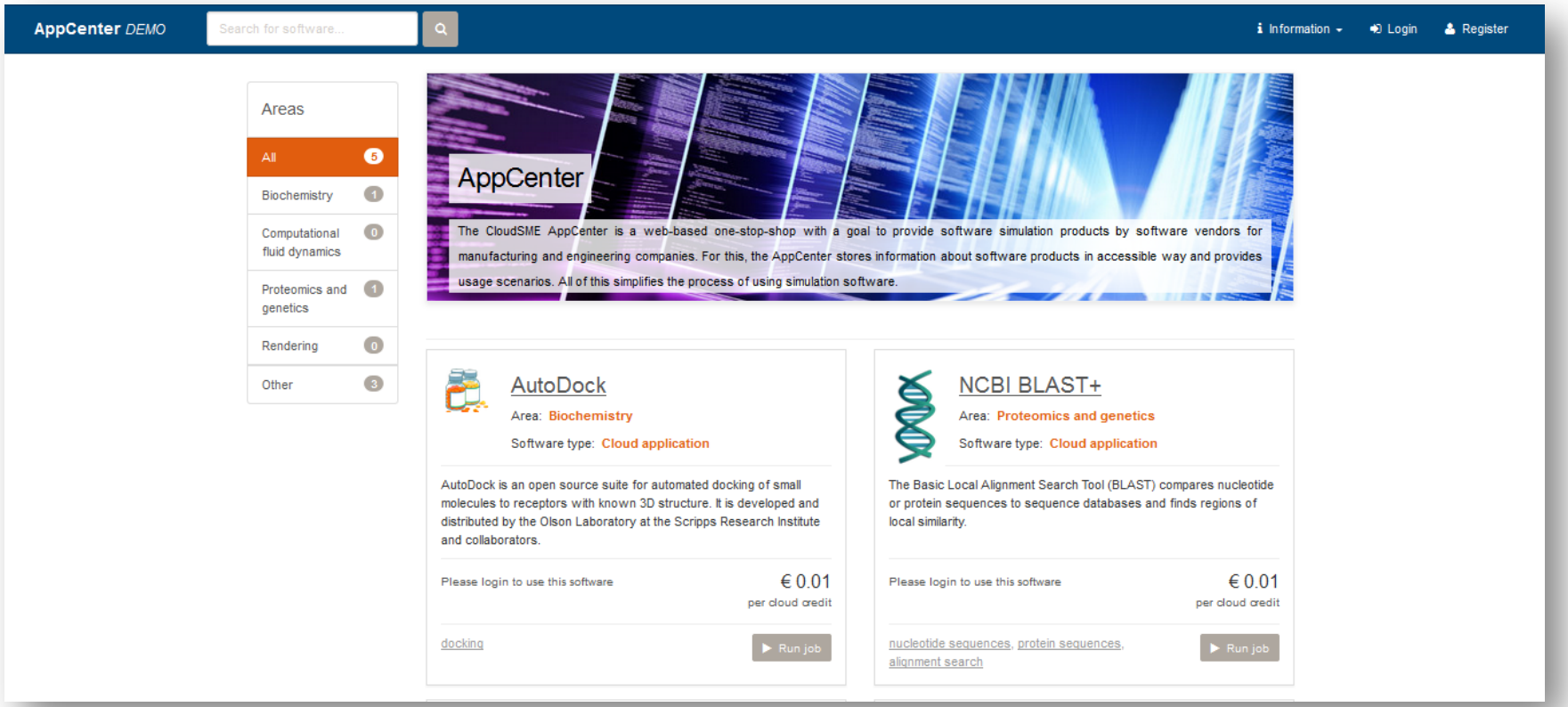
- User-friendly web one-stop-shop and marketplace to provide software products and services from software vendors to end users
- Single point of entry to different applications and technologies
- Originally developed for simulations within the CloudSME project
- Different options for end users to buy, retrieve and execute applications
- Different options for software vendors to price and provide subscriptions of applications
- Adapters to deliver software as download, web service or CloudBroker Platform deployment
- Accounting, billing and payments
- Browser and programmatic access, can be used as frontend or backend
- Offered as public, hosted or in-house service or licensed software

AppCenter Architecture



CloudBroker

AppCenter Screenshot



CloudBroker

Example Use Cases

Compute-intensive Applications

Protein Modeling

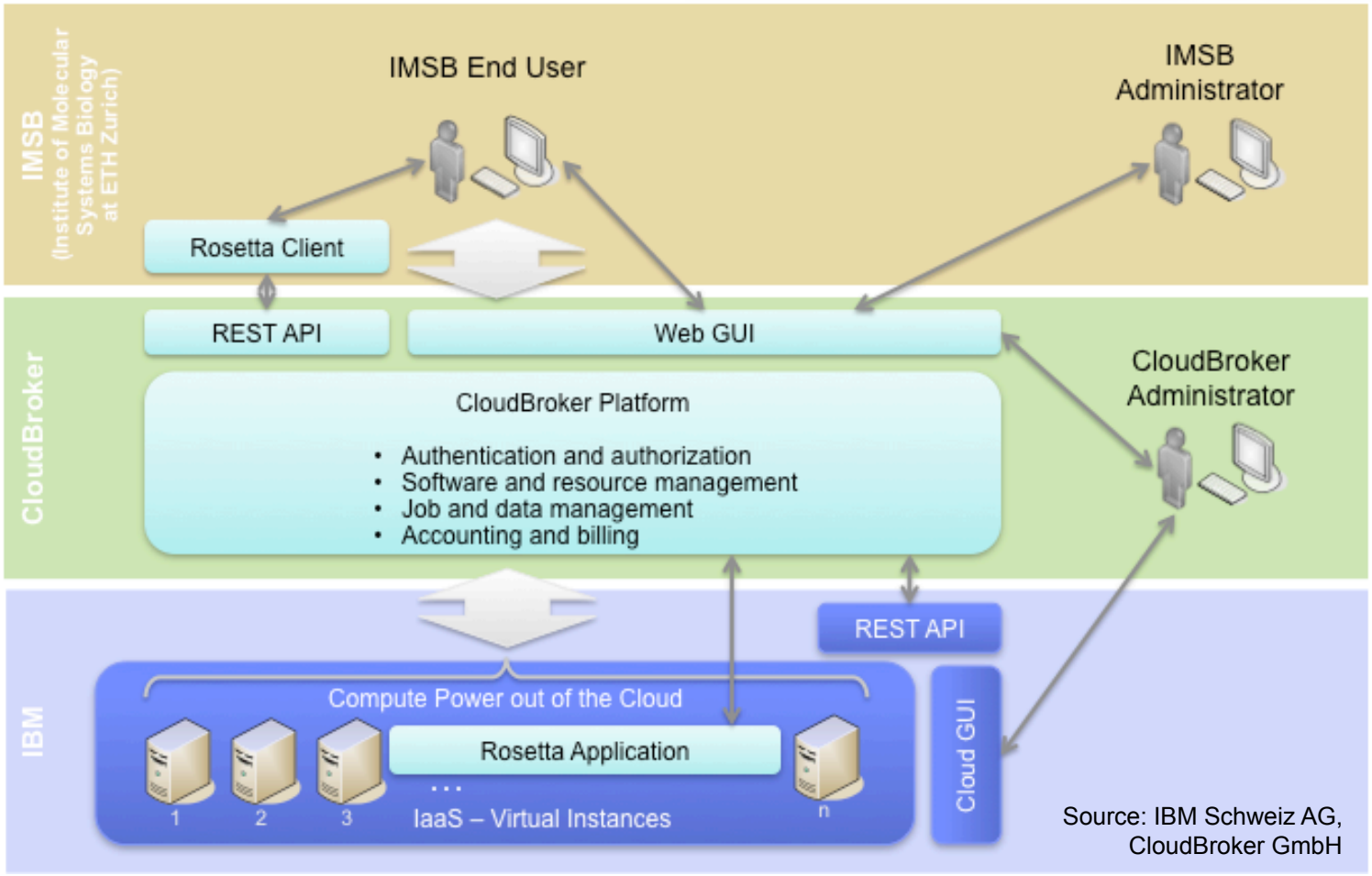
- Collaboration with the Institute of Molecular Systems Biology (IMSB) at ETH Zurich and IBM
- Goal: Better understand the mechanisms of infectious diseases to fight antibiotics resistance
- Example: Streptococcus bacterium
- Predict the 3D structures of important proteins from their 1D sequence
- Software: Rosetta
- Find the important structural differences between less and more harmful bacteria strains

⇒ Calculations expected to take several months on available small segment of ETHZ cluster

⇒ Embarrassingly parallel, compute-intensive application, perfect fit for cloud computing

⇒ Use of CloudBroker Platform to manage the calculations on top of IBM SmartCloud Enterprise compute and storage resources

Modeling Architecture



CloudBroker

Protein Modeling Results

Indicator	Amount
Modeled proteins	249
Calculation jobs	More than 36’000
Virtual machines	Up to 63 in parallel, dynamically adjusted
Virtual CPUs	Up to 1008 in parallel, dynamically adjusted
Consumed CPU hours	Almost 250’000
Real time	Less than two weeks
Generated 3D protein structures	About 2.3 Mio

⇒ Speed-up of calculations by parallelization on cloud resources leads to much shorter real time for research

⇒ Scientists can focus on project at hand and analysis of results instead of worrying about computing infrastructure

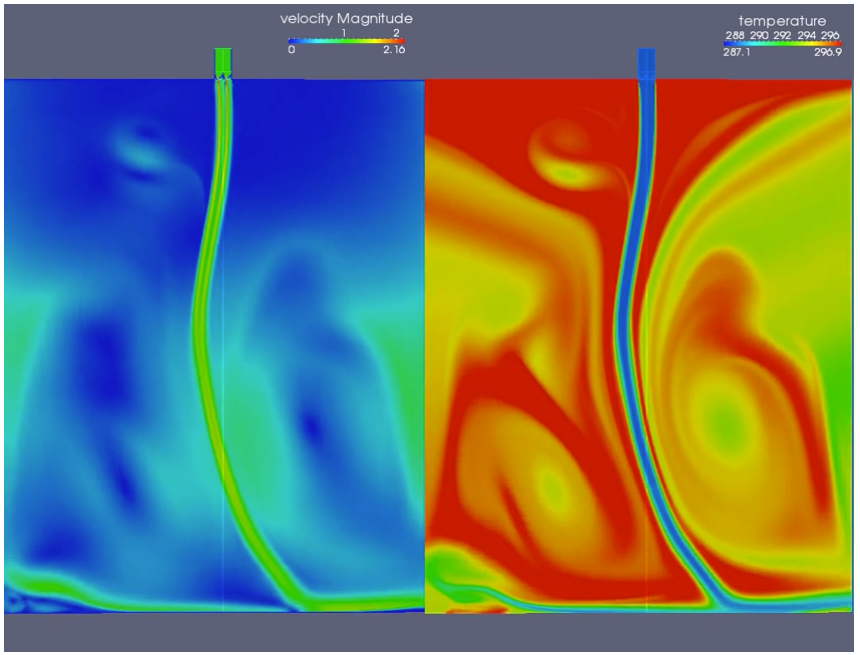
UberCloud HPC Experiment Team 30

- Heat transfer use case from Biscarri Consultoria, Spain
- Using ELMER open source finite element software for multiphysical problems from CSC, Finland
- Deployed as part of CAELinux
- Cloud resources from Amazon Web Services
- All combined by the CloudBroker Platform

⇒ Ease of use for application management and execution

⇒ Good intra-node scalability

⇒ Amazon HPC instances 10 Gbit connection not sufficient for good inter-node scalability



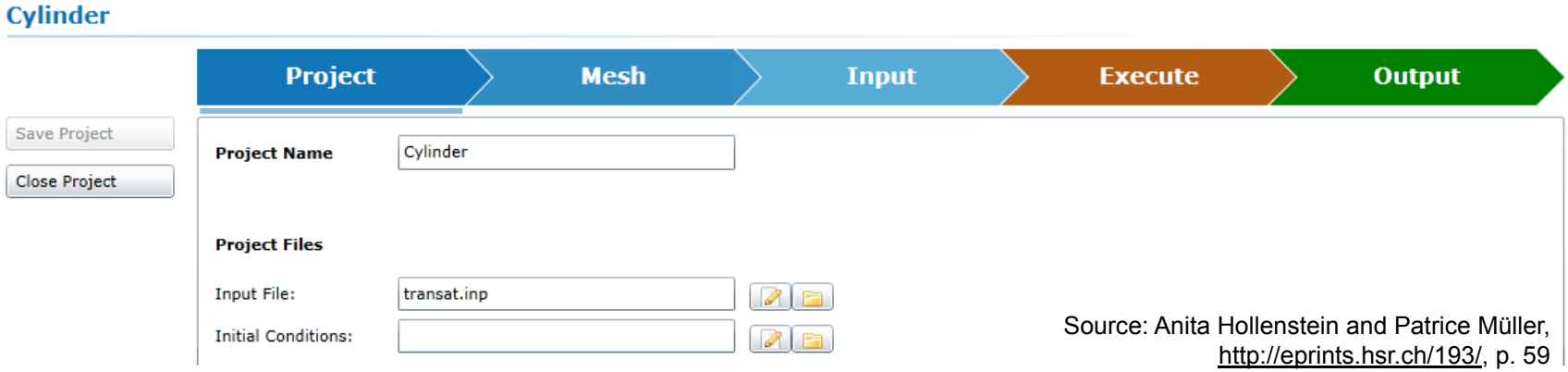
Source: http://tci.taborcommunications.com/UberCloud_HPC_Experiment

CloudBroker

External User Interfaces

Web UI for TransAT

TransAT WebUI



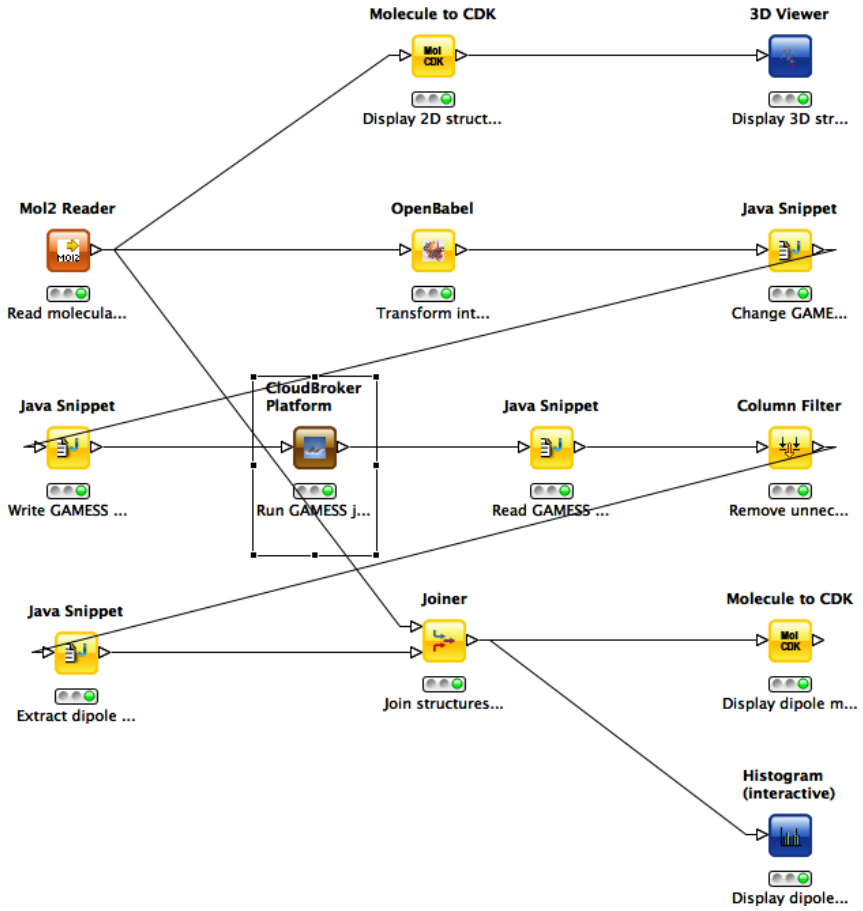
Source: Anita Hollenstein and Patrice Müller, http://eprints.hsr.ch/193/, p. 59

- TransAT = Transport phenomena Analysis Tool: Computational multiphase fluid dynamics (CMFD) software from ASCOMP GmbH
- HSR bachelor thesis “Web UI-Front-End for Fluid Dynamics Cloud”
- Microsoft Silverlight-based web UI for the CloudBroker Platform

CloudBroker

Interfacing in KNIME

- KNIME = Konstanz Information Miner: Open source data integration, processing, analysis and exploration framework by KNIME.com AG
- Development of a CloudBroker Platform KNIME node
- Computational chemistry workflow showcase: Calculation of dipole moments for a number of molecules



CloudBroker

Large Collaborations

SCI-BUS Project

- "SCIentific gateway Based User Support"
- Supported by the EU FP7 Capacities Research Infrastructures program under grant agreement no. 283481
- 10/2011-09/2014
- http://www.sci-bus.eu

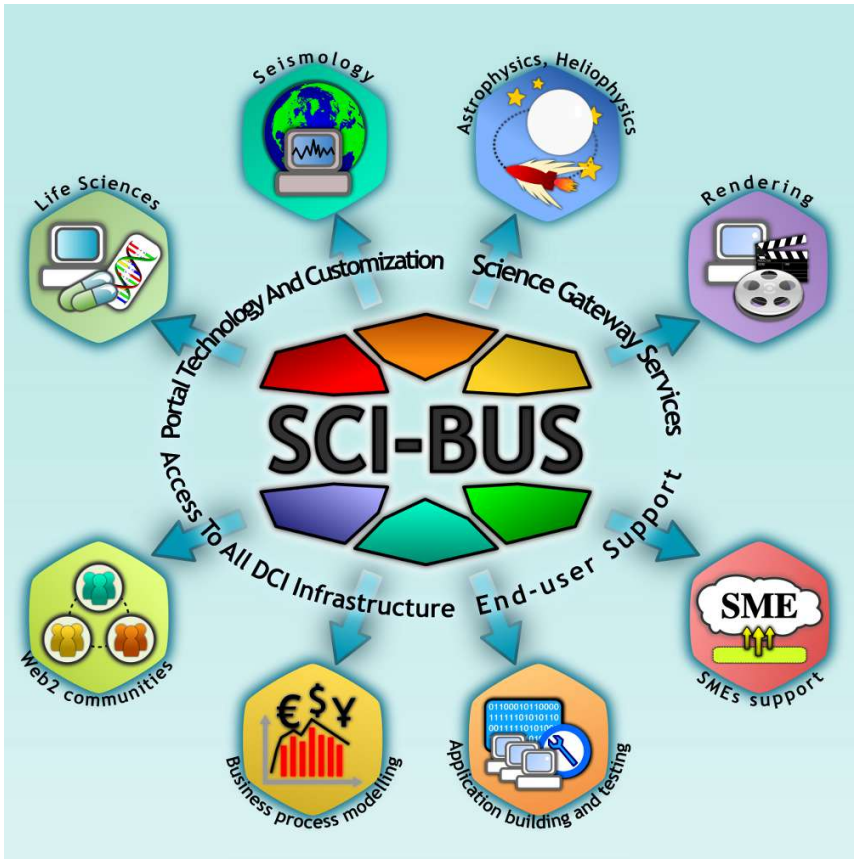
⇒ Provide researchers seamless access to major computing, data and networking infrastructures and services

⇒ Build science gateways for different domains and communities

⇒ Develop corresponding gateway technology

⇒ Focus on scientific workflows

- Base technology WS-PGRADE/gUSE
- CloudBroker Platform for connection to public and private, commercial and academic clouds



Source: SCI-BUS project, http://www.sci-bus.eu

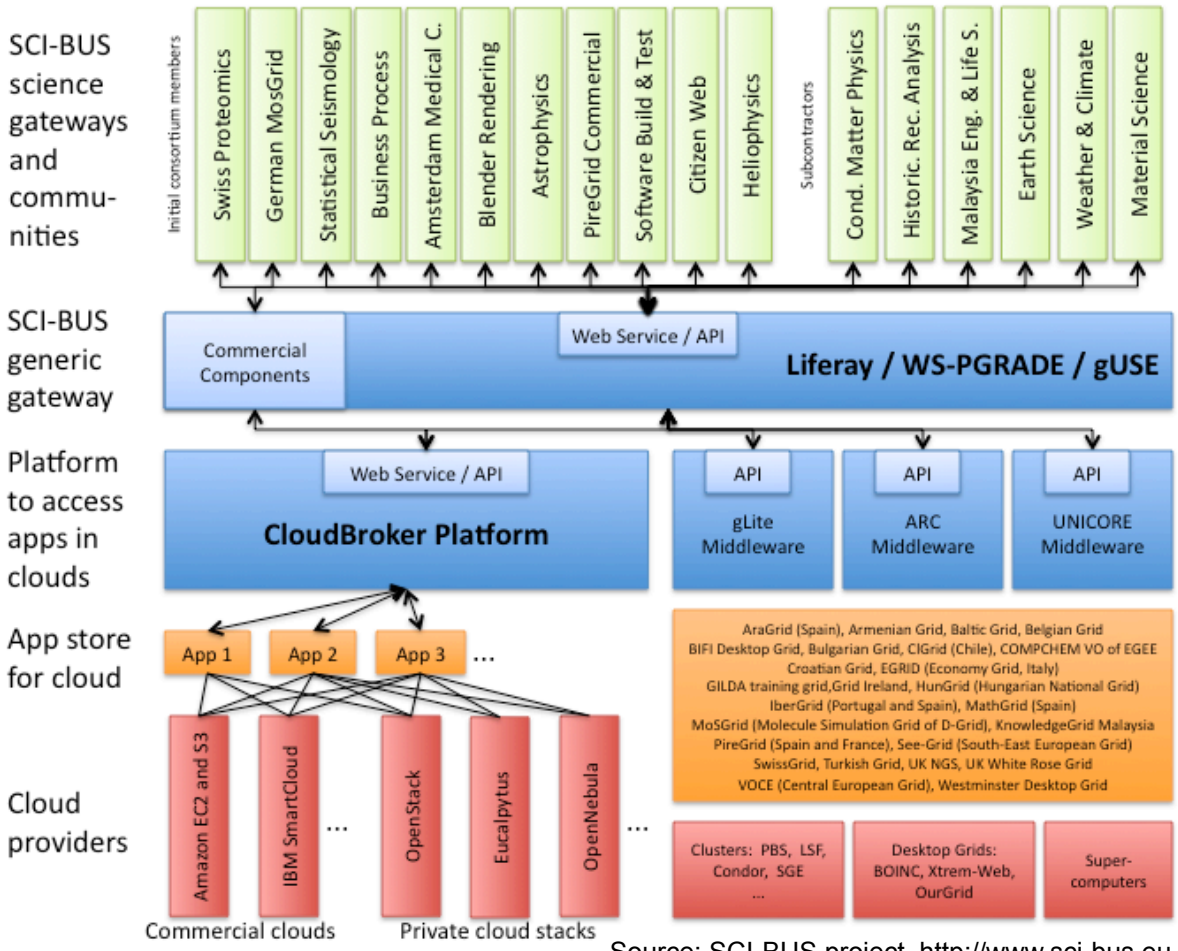
SCI-BUS Partners

+ 6 Subcontractors
+ Associated partners

Source: SCI-BUS project, http://www.sci-bus.eu

CloudBroker

SCI-BUS Architecture



Source: SCI-BUS project, http://www.sci-bus.eu

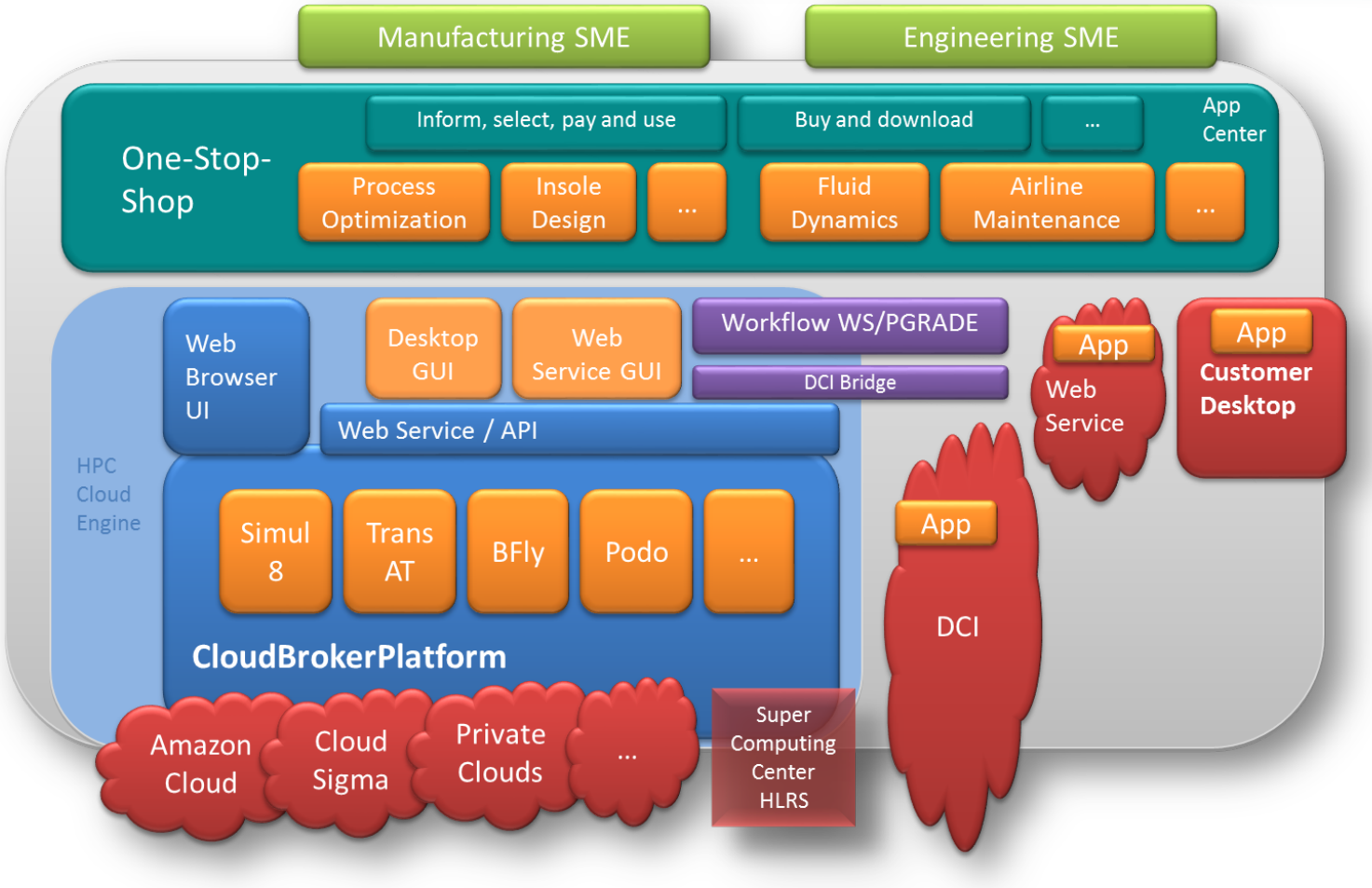
CloudSME Project

- Project:
  - “Cloud based Simulation platform for Manufacturing and Engineering”
  - Supported by the EU FP7 Public-Private Partnership Factories of the Future program under grant agreement no. 608886
  - Part of the EU I4MS (ICT Innovation for Manufacturing SMEs) initiative
  - 07/2013-12/2015
  - http://www.cloudsme.eu
- Objectives:
  - Give European SMEs access to simulations
  - Development of a SaaS and PaaS platform for manufacturing and engineering simulations
  - On top of different cloud and HPC infrastructures
  - One-stop-shop for users
- Collaboration:
  - 29 partners from 8 countries
  - 24 SMEs and 5 academic institutions

CloudSME Setup

- Platform technology:
  - CloudBroker Platform and AppCenter: CloudBroker (CH) & ScaleTools (CH)
  - WS-PGRADE/gUSE: MTA SZTAKI (HU)
- Infrastructure resources:
  - Commercial clouds: CloudSigma (CH) & Amazon (external)
  - OpenStack clouds: University of Westminster (UK) & University of Zaragoza (ES)
  - OpenNebula cloud: MTA SZTAKI (HU)
  - HPC: HLRS (external)
- Project coordination:
  - University of Westminster (UK)
- Simulation consulting:
  - Brunel University (UK) & University of Westminster (UK)
- Marketing and dissemination:
  - Sander Werbung (DE) & University of Zaragoza (ES)
- Application experiments:
  - Collaborations of software vendors and end users
  - 4 initial application experiments plus 7 additional experiments from Open Call

CloudSME Architecture



CloudBroker

CloudSME Experiments

- **Fluid dynamics: ASCOMP (CH) & Eurobios (FR)**
- *Discrete event simulation: SIMUL8 (UK) & Saker (UK)*

  Process optimization: Cutting Tools (UK)

→ *Freight transport intermodal terminals: PROYFE (ES)*

→ *Craft brewing templates: Hobsons (UK)*

- *Insole design: Ingecon (ES) & Podoactiva (ES)*

→ *Safety shoes: Base Protection (IT) & IOR (IT)*

- Aircraft maintenance: 2MoRO (FR)
- **Fluid dynamics and finite element analysis for model helicopters: DHCAE (DE) & ProcEng (CH)**
- **Emission reduction: CMCL (UK) & G-volution (UK)**
- *Business process modeling: Simsoft (TR) & Özdekan (TR) & Gökdogan (TR)*
- Inventory forecasting: Outlandish (UK) & Tidy Books (UK)

Conclusions

Summary

- Definition of cloud computing
- Definition of high performance computing
- General solutions for HPC in the cloud:
  - IaaS
  - HPC on demand
  - Application portals
  - One-stop-shop
- CloudBroker solutions for HPC in the cloud:
  - CloudBroker Platform = PaaS and SaaS
  - CloudBroker AppCenter = Marketplace
- CloudBroker use case examples:
  - Compute-intensive applications
  - External user interfaces
  - Large collaborations

Acknowledgements

- CloudBroker and ScaleTools teams, in particular:
  - Nicola
  - Andrey
  - Anna
  - … and others
- Collaborators from:
  - CloudSME
  - SCI-BUS
  - IBM
  - ETH Zurich
  - ASCOMP
  - HLRS
  - HPC Experiment
  - KNIME
  - … and others

Thanks! – Questions?

For more information:

Dr. Wibke Sudholt, wibke.sudholt@cloudbroker.com

CloudBroker

Hands-on Tutorial

CloudBroker

A Typical Use Case

- Assume you are a chemist and this is what you want to do:
  - Calculate the geometry of a water molecule
  - Using the open source computational chemistry software NWChem (http://www.nwchem-sw.org)
- What you have:
  - Expert knowledge about chemistry and how to use the NWChem software
  - A simple computer with a web browser and internet access
- What you do not have:
  - Expert knowledge about software deployment, high performance computing, clouds and the like
  - A local installation of NWChem or access to a server, cluster or supercomputer with NWChem available

Input File

```
geometry
  O 0.0 0.0 0.0
  H 1.0 0.0 0.0
  H 0.0 1.0 0.0
end
basis
  H library 6-31g**
  O library 6-31g**
end
print low
task mp2 optimize
```

Solution

- Go to the CloudBroker AppCenter (using the URL and authentication details provided)
- Register as a new user (if you do not have access to your email, please ask)
- Refill the NWChem application with 1000 test credits
- Download water.nw (from the URL provided)
- Run an NWChem job with water.nw as input file and water.nw as argument string
- Watch the CloudBroker AppCenter while the calculation is running
- Download the output file
- Unpack output.tgz and display job.out with a text editor
- Wow, you have just done a scientific calculation in the cloud!

Result in the Output File

Final and change from initial internal coordinates

Z-matrix (autoz)

Units are Angstrom for bonds and degrees for angles

	Type	Name	I	J	K	L	M	Value	Change
1	Stretch		1	2				0.96077	-0.03923
2	Stretch		1	3				0.96077	-0.03923
3	Bend		2	1	3			103.87946	13.87946

Changing Perspective

- Now you sit on the other side and are a provider of computer resources and application software.
- How do you make your computer resources (example here: Amazon cloud) and application software (example here: NWChem) easily available to users?

Solution (1)

- Go to the CloudBroker Platform (under the URL provided)
- Login with the account given to you
- Register a new Amazon resource account (use your own Amazon credentials or ask if you do not have ones)
- Download nwchem_install.sh (from the URL provided)
- Register and deploy the NWChem software on the Amazon cloud
- Run the NWChem water case as test job using your NWChem software on the Amazon cloud
- Watch what happens in the CloudBroker Platform and at Amazon

Solution (2)

- Go back to the CloudBroker AppCenter
- Login with the account given to you (NOT the account you created yourself)
- Register the NWChem software you deployed and tested in the CloudBroker Platform
- Logout
- Login with your own account (the one you initially created)
- Repeat the water calculation from the beginning with the NWChem version you deployed in the cloud yourself
- Voila, you are done!

Thanks! – Questions?

For more information:

Dr. Wibke Sudholt, wibke.sudholt@cloudbroker.com

CloudBroker