

# Grids and Clouds

GridKa School 2009 – Forschungszentrum Karlsruhe – September 1<sup>st</sup> 2009

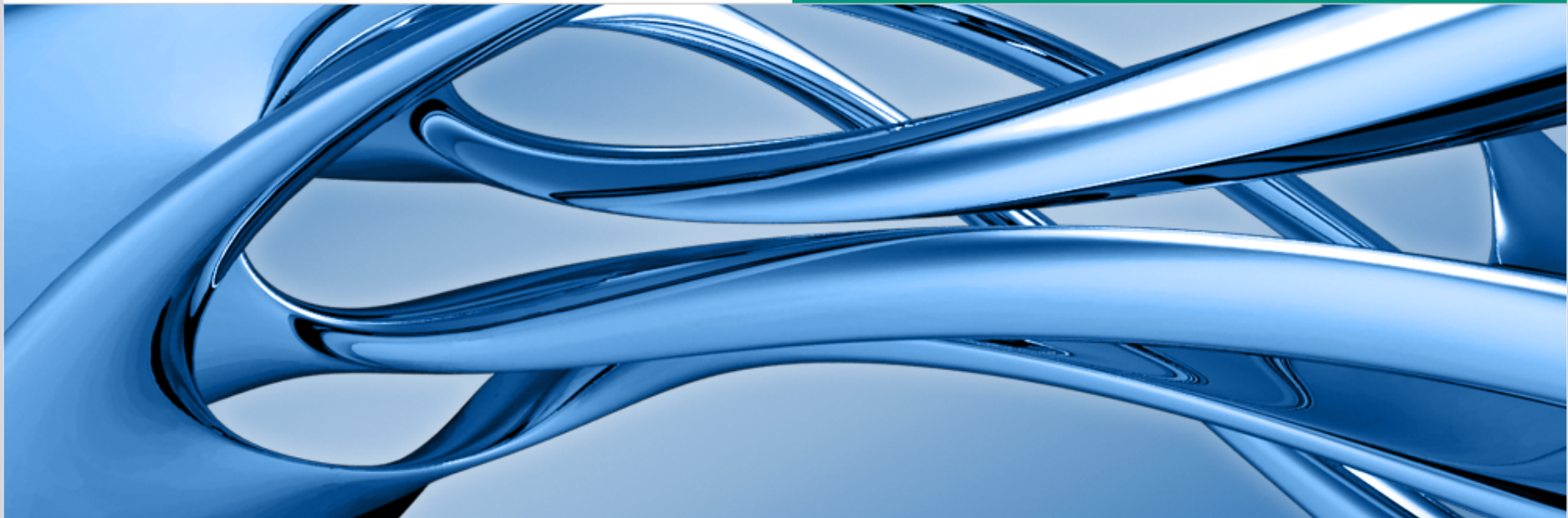
Christian Baun



Forschungszentrum Karlsruhe  
in der Helmholtz-Gemeinschaft



Universität Karlsruhe (TH)  
Forschungsuniversität • gegründet 1825



# Cloud Computing - big topic/hype

- Big topic (hype) in the press since autumn 2008

## Gartner Identifies Top Ten Disruptive Technologies for 2008 to 2012

MELBOURNE, Australia, May 28, 2008 — Social networking technologies, web mashups, multicore and hybrid processors and **cloud computing** are amongst the ten most disruptive technologies[1] that will shape the information technology (IT) landscape over the next five years, according to research and advisory firm Gartner, Inc.

QUOTABLE

## Larry Ellison on cloud computing buzzword: "Complete gibberish"

By [Jackson West](#), 1:20 PM on Fri Sep 26 2008, 1,766 views

Web 2.0: Article

## Cloud Hype at Height: Gartner

*It's supposed to be 2–5 years away from mainstream adoption*

## Cloud computing: Hot technology for 2009

Proceed with caution

By [Neal Weinberg](#), *Network World*, 01/05/2009

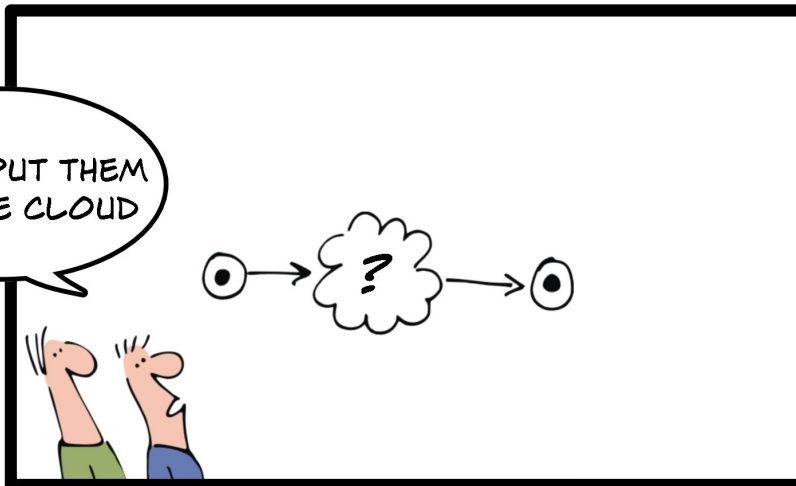
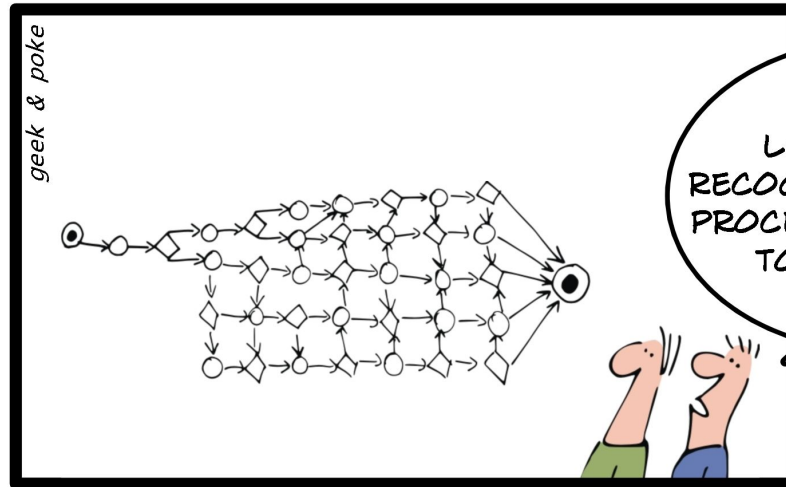
As we arrive at 2009, cloud computing is the technology creating the most buzz. Cloud technology is in its infancy, however, and enterprises would be wise to limit their efforts to small, targeted projects until the technology matures and vendors address a variety of potentially deal-breaking problems.

## Cloud computing is a trap, warns GNU founder Richard Stallman

Web-based programs like Google's Gmail will force people to buy into locked, proprietary systems that will cost more and more over time, according to the free software campaigner

**Bobbie Johnson**, technology correspondent  
[guardian.co.uk](#), Monday 29 September 2008 14.11 BST

# Why do we want/need Cloud Computing?



LET THE CLOUDS MAKE YOUR LIFE EASIER

- Cloud Computing shall:
  - revolutionize IT
  - reduce (eliminate) complexity
  - getting IT more flexible (elastic)
  - reduce cost
  - provide easy resources access
  - fulfill users demands
  - emancipate the users
  - ...

■ Remember: **GRID** !?!?!?

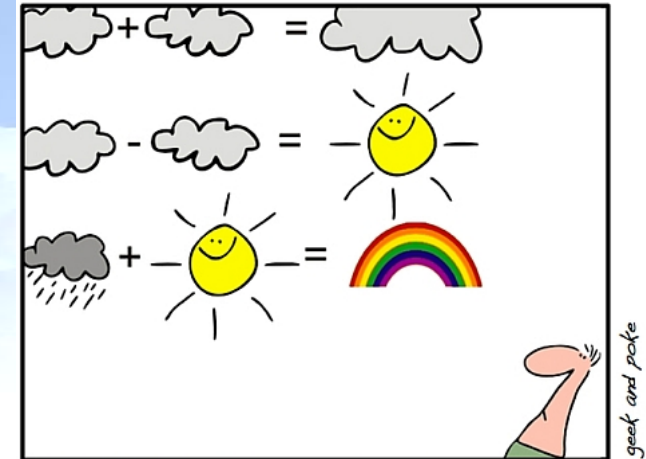
# Definitions (Cloud / Grid)

- Cloud Computing is on-demand access to virtualized IT resources that are sourced inside or outside of a data center, scalable, shared by others, simple to use, paid for via subscription or as you go and accessible over the web.

Dr. Behrend Freese (Zimory GmbH)

- A computing Cloud is a set of network enabled on demand IT services, scalable and QoS guaranteed, which could be accessed in a simple and pervasive way.

Dr. Marcel Kunze (SCC/KIT)



SIMPLY EXPLAINED - PART 17:  
CLOUD COMPUTING

- Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations.

Ian Foster (Argonne National Laboratory)

- A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.

Ian Foster & Carl Kesselman



# Cloud Computing ≠ Grid Computing

## Cloud Computing

One or few data centers containing heterogeneous or homogeneous resources under central control

Virtualized resources

Fully automated Services  
(Industrialization of IT)

Commercial business model  
(Pay-as-you-go)

Easy to use and deploy.  
No complex user interface required

Used primarily in industry and business  
(Big chance for Startups)

## Grid Computing

Geographically distributed, heterogeneous resources without central control. Follows the principle of virtual organizations

Physical resources

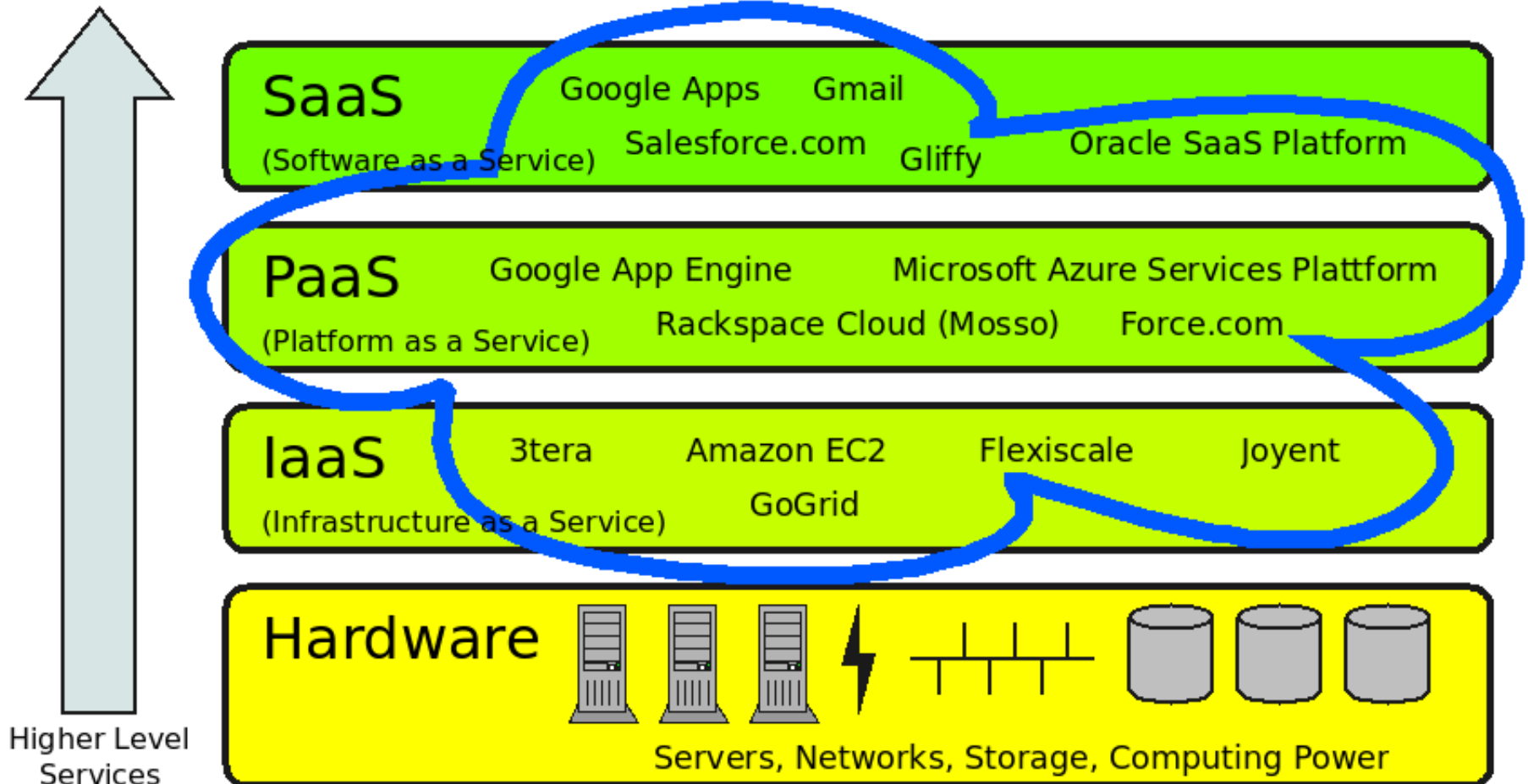
Handcrafted Services  
(Mostly Manufacture)

Publicly funded.  
Usage is for free if the resource owner approves resource access

Difficult to use and deploy

Used first and foremost by research and academic organizations  
(Large-scale scientific projects like LHC)

# Three technical Types of Cloud Services



# SaaS: Software as a Service

- Provides enterprise quality software that is run by a service provider
- Nothing has to be installed locally
- The user has no need to worry about:
  - Installation
  - Administration
  - Updates
  - Software-Licenses
- Popular Examples:
  - Clarizen: Collaborative online project management tools
  - Employease: Human resources information system
  - Gliffy: Online Diagram Software
  - Google docs: Free and collaborative Web-based office applications
  - Salesforce.com: Customer relationship management tools
  - SlideRocket: Online (collaborative) presentation tool
  - Zoho: Suite of online tools for small businesses (Apps, CRM, Project,...)



clarizen  
Connect Your Business

ADP Employease

Google docs

salesforce.com  
Success. Not Software.®

sliderocket™

ZOHO

gliffy

# PaaS: Platform as a Service

- Integrated development, runtime and/or test environment as a service
- Supports typically one or more programming languages
- Appears as one single large computer
  - Makes it simple to scale from a single server to many
- No need to worry about the operating system or other foundational software
- Developers and not users are the main target group of PaaS services.
- Popular Examples:
  - Bungee Connect: Web application development and hosting
  - Force.com: Developing and deploying customer relationship management apps
  - Google App Engine: Platform for developing and hosting Python and JAVA web applications in the Google datacenters
  - Microsoft Azure: Platform that allows developing, hosting Windows applications in Microsoft datacenters.
  - Zoho Creator: Platform for building and running online databases





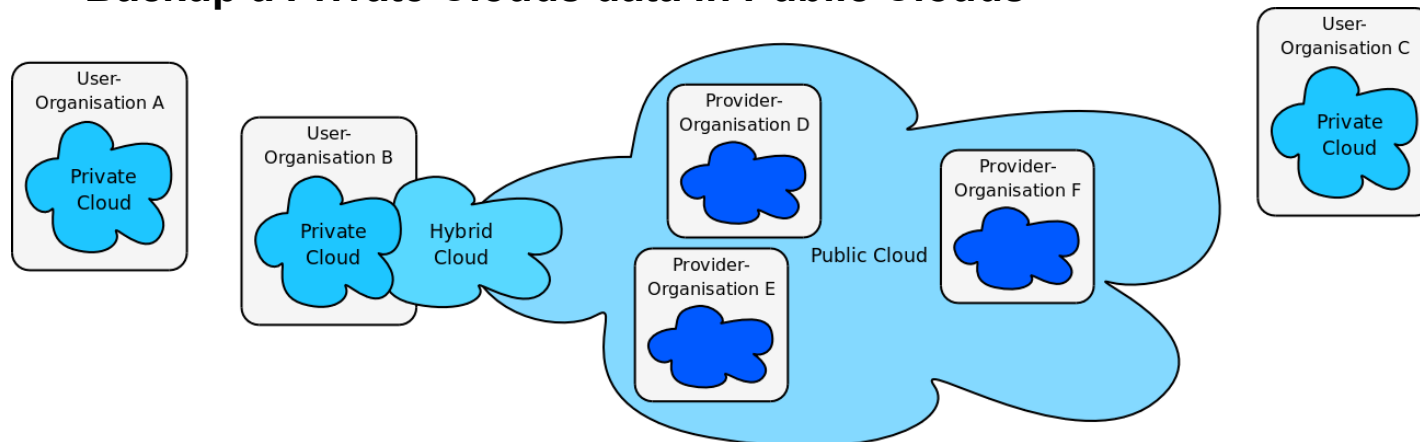
# IaaS: Infrastructure as a Service

- Abstracts away the hardware (servers, network,...)
- allows to run virtual instances of servers without ever touching the hardware
- A user interface is provided by the service provider, allowing the users...
  - create, control and terminate the virtual instances
  - change the quantity of resources, the virtual instances are equipped with
  - define and change virtual network devices
- Popular Examples:
  - Amazon EC2: Linux, Windows and Solaris VM hosting in the Amazon datacenters
  - GoGrid: Linux and Windows VM hosting
  - FlexiScale: Linux and Windows VM hosting
  - Joyent: OpenSolaris hosting
  - Rackspace cloud: Linux VM hosting
  - Todo flexIT: Linux and Windows VM hosting
- Only consumed resources are accounted following the pay-as-you-go principle



# Three organizational Types of Cloud Infrastructures

- **Public Clouds (respectively External Clouds)**
  - Customer and Cloud provider do not belong to the same organization
  - Commercial business model (Pay-as-you-go)
- **Private Clouds (respectively Internal Cloud or Intra Cloud)**
  - Cloud services are provided from the users organization
  - Main reasons for Private Clouds: Security concerns and R&D
- **Hybrid Clouds**
  - Cloud services from one or more Public and Private Clouds are used
  - Use cases:
    - Public Clouds help to manage load peaks
    - Backup a Private Clouds data in Public Clouds



# Commercial Cloud Offerings (Small Excerpt)



APPLIC | UTILITY COMPUTING | TECHNOLOGY | PARTNERS | GRID UNIVERSITY | COMPANY

Cloud Computing

Cloudware - Cloud Computing Without Compromise



- Commercial Cloud offers are usually proprietary and therefore not open for building Private Clouds and for Cloud systems research and development!
- Users are afraid of a vendor lock-in



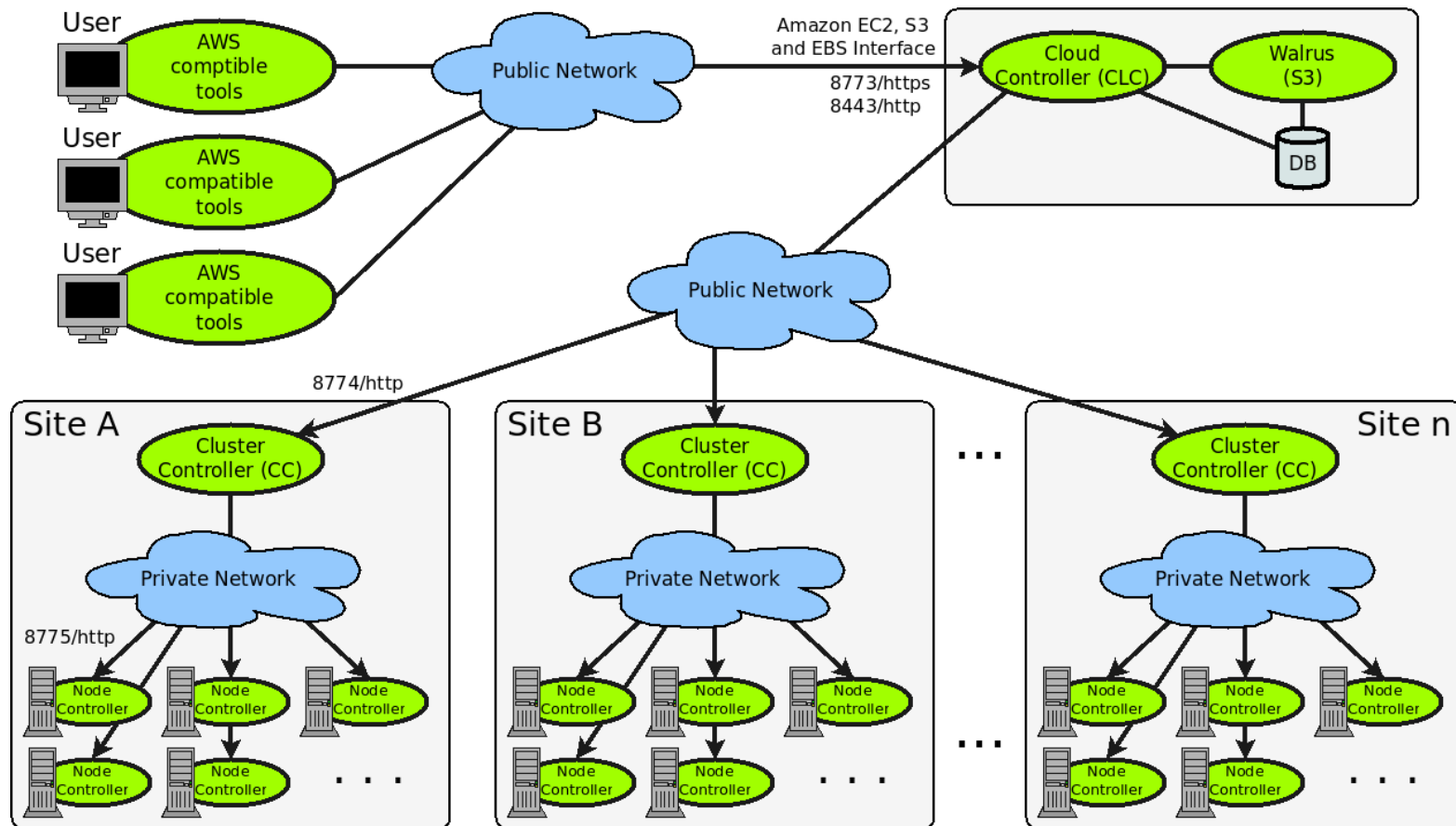
- **EUCALYPTUS - Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems**
- **Open Source software infrastructure for implementing Cloud Computing on clusters from UC Santa Barbara**
- **Developed at UC Santa Barbara**
- **Implements Infrastructure as a Service (IaaS)**
- **Gives the user the ability to run and control virtual machine instances (Xen, KVM) deployed across a variety of physical resources**
- **Interface compatible with Amazon EC2, S3 and EBS**
- **Includes „Walrus“, a storage service (provides S3 functionality)**
- **Potential to interact with the same popular tools, known to work with Amazon EC2, S3 and EBS e.g. S3 Curl, ElasticFox, s3cmd, ...**
- **Allows you to avoid vendor lock-in by enabling migrating data and services to an AWS-compatible private cloud**
- **Eucalyptus is an important step to establish an open Cloud Computing infrastructure standard**

# Eucalyptus - Components

<http://open.eucalyptus.com>



# Eucalyptus



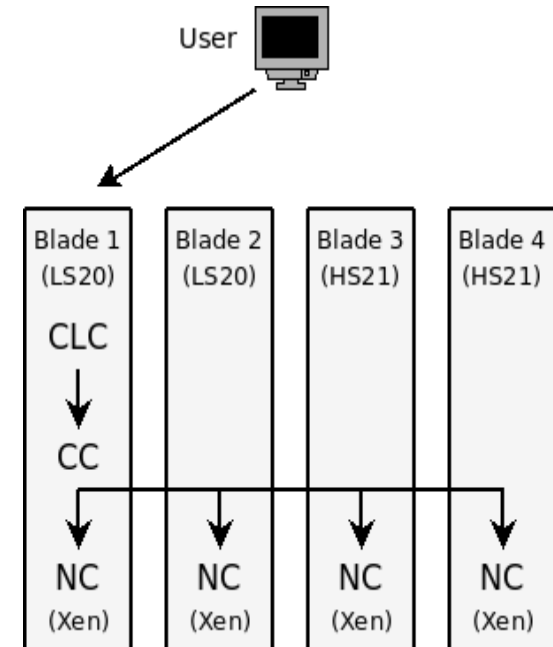
- **CLC:** Collects resource information for the CC. Operates like a meta-scheduler
- **CC:** Schedules the VM distribution to the NC. Collects (free) resource information
- **NC:** Controls Xen-Hypervisor or KVM. Provides resource information to the CC



# Eucalyptus at KIT

## ■ Private Cloud Installation I (**Eucalyptus 1.4**)

- Runs stable
- Plattform for performance testing
- 2x IBM Blade LS20
  - 2 Single Core Opteron (2,4GHz)
  - 4GB RAM
- 2x IBM Blade HS21
  - 2 Dual Core Xeon (2,33GHz)
  - 16GB RAM
- Runs stable at KIT since February 2009



## ■ Private Cloud Installation II (**Eucalyptus 1.5.2**)

- Under construction
- 5x HP Blade ProLiant BL2x220c
- 2 Server per blade:
  - 2x Intel Quad-Core Xeon (2,33GHz)
  - 16GB RAM

# Eucalyptus (installation)

- **Binary packages exist for CentOS, openSUSE, Debian and Ubuntu**  
<http://open.eucalyptus.com/downloads>  
[http://open.eucalyptus.com/wiki/EucalyptusAdministratorGuide\\_v1.5.2](http://open.eucalyptus.com/wiki/EucalyptusAdministratorGuide_v1.5.2)
- **Eucalyptus can be installed from source also**  
[http://open.eucalyptus.com/wiki/EucalyptusSourceCodeInstallation\\_v1.5.2](http://open.eucalyptus.com/wiki/EucalyptusSourceCodeInstallation_v1.5.2)
- **Requirements:  $\geq 1$  computer running with Linux and a functional Xen Hypervisor or Kernel-based Virtual Machine (KVM)**
  - **For KVM, a *modern* CPU with AMD-V (Pacifica) or Intel VT (Vanderpool) is needed**
- **Amazon EC2 command line tools to control Eucalyptus**
  - **ec2-api-tools-1.3-30349**
  - **ec2-ami-tools-1.3-26357**

# Using Eucalyptus

- Shows the cluster's front-end hostname, free resources, instance types, available NCs:

```
# ec2-describe-availability-zones verbose
AVAILABILITYZONE      iwrhpcluster      UP scchpblade03a
AVAILABILITYZONE      |- vm types      free / max      cpu      ram      disk
AVAILABILITYZONE      |- m1.small      0038 / 0048      1        128       2
AVAILABILITYZONE      |- c1.medium      0038 / 0048      1        256       5
AVAILABILITYZONE      |- m1.large      0016 / 0024      2        512      10
AVAILABILITYZONE      |- m1.xlarge     0016 / 0024      2       1024      20
AVAILABILITYZONE      |- c1.xlarge     0005 / 0012      4       2048      20
AVAILABILITYZONE      |- 141.52.167.65      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
AVAILABILITYZONE      |- 141.52.167.66      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
AVAILABILITYZONE      |- 141.52.167.67      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
AVAILABILITYZONE      |- 141.52.167.68      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
AVAILABILITYZONE      |- 141.52.167.69      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
AVAILABILITYZONE      |- 141.52.167.71      certs[cc=true,nc=true] @ Mon Aug 31
21:05:13 UTC 2009
```

## ■ Register a Filesystem-Image:

```
# ec2-bundle-image -i debian5.img  
# ec2-upload-bundle -b image-debian5 -m /tmp/debian5.img.manifest.xml  
# ec2-register image-debian5/debian5.img.manifest.xml
```

## ■ Register a Kernel-Image:

```
# ec2-bundle-image -i /boot/vmlinuz-2.6.26 --kernel true  
# ec2-upload-bundle -b kernel26 -m /tmp/vmlinuz-2.6.26.manifest.xml  
# ec2-register kernel26/vmlinuz-2.6.26.manifest.xml
```

## ■ Register a Ramdisk-Image:

```
# ec2-bundle-image -i /boot/initrd.img-2.6.26 --ramdisk true  
# ec2-upload-bundle -b ramdisk26 -m /tmp/initrd.img-2.6.26.manifest.xml  
# ec2-register ramdisk26/initrd.img-2.6.26.manifest.xml
```

# Control Registered Images

## ■ Registered images gets an unique identifier:

- Eucalyptus Machine Image: `emi-xxxxxxxx`
- Eucalyptus Kernel Image: `eki-xxxxxxxx`
- Eucalyptus Ramdisk Image: `eri-xxxxxxxx`

## ■ Get information about registered images:

```
# ec2-describe-images
```

```
IMAGE    emi-1DE4116D    image-debian5/debian5.img.manifest.xml
          admin    available    public    x86_64    machine
IMAGE    eki-791612FF    kernel26/vmlinuz-2.6.26.manifest.xml
          admin    available    public    x86_64    kernel
IMAGE    eri-CFBE1450    ramdisk26/initrd.img-2.6.26.manifest.xml
          admin    available    public    x86_64    ramdisk
```



# Keypair creation and running instances

## ■ Create and register keypair:

```
# ec2-add-keypair mykey > mykey.private
# chmod 0600 mykey.private
# ec2-describe-keypairs
KEYPAIR mykey
    33:da:6e:13:96:e6:f7:3b:b7:34:a6:28:ba:2f:64:ab:83:70:ef:70
```

## ■ Run instances:

```
# ec2-run-instances emi-1DE4116D --kernel eki-791612FF
                                --ramdisk eri-CFBE1450
                                -k mykey -n 2 -t m1.small
```

- -k <keypair>
- -n <#instances>
- -t <instance\_type>
- -z <availability\_zone>

# Control Instances and use them

## ■ Check instances:

```
# ec2-describe-instances
```

RESERVATION	r-3DDE07D9	admin	default	
INSTANCE	<b>i-4901084F</b>	emi-1DE4116D	141.52.166.160	141.52.166.160
running	mykey	0	m1.small	
2009-05-13T13:50:37+0000		iwrhpcluster	eki-791612FF	eri-CFBE1450
RESERVATION	r-42FA0732	admin	default	
INSTANCE	<b>i-463B08BE</b>	emi-1DE4116D	141.52.166.161	141.52.166.161
running	mykey	0	m1.small	
2009-05-13T13:50:10+0000		iwrhpcluster	eki-791612FF	eri-CFBE1450

## ■ Login to an instance via SSH:

```
# ssh -i mykey.private 141.52.166.160
```

## ■ Terminate instances:

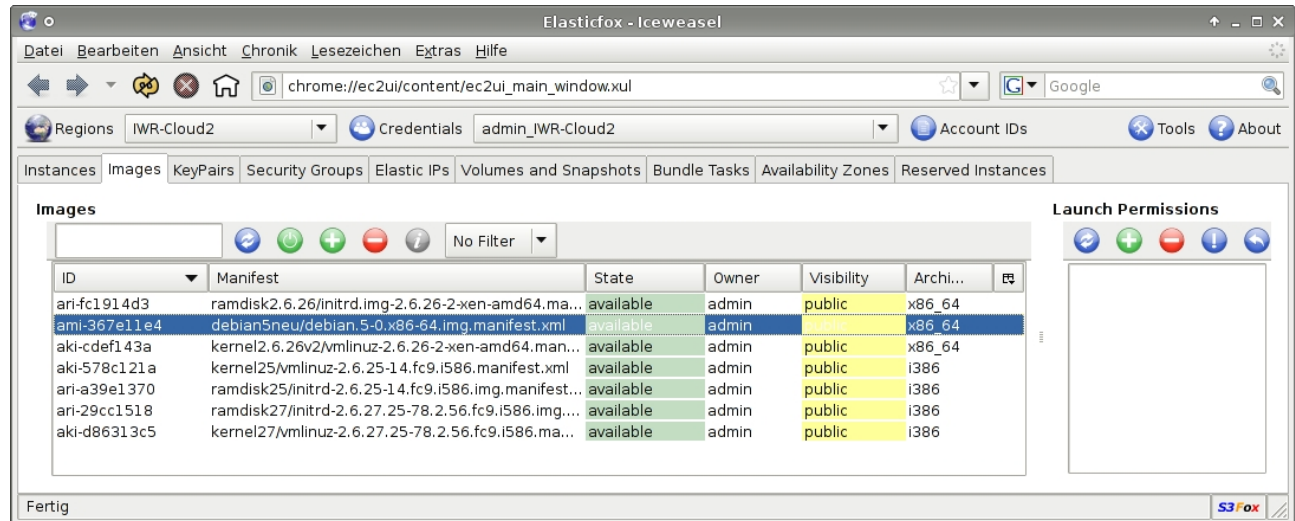
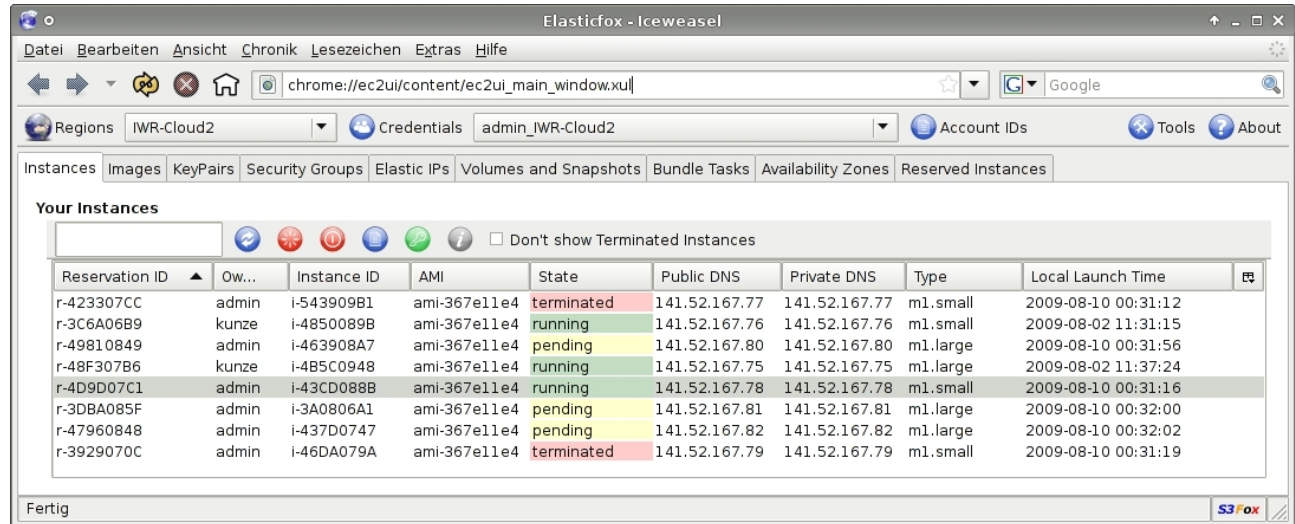
```
# ec2-terminate-instances i-4901084F i-463B08BE
```

# Handy and popular tools for Eucalyptus

- The Amazon Web-Services (EC2, S3, EBS, ...) are very popular
- Lots of handy tools supporting the AWS exist
  - Many can be utilized with Eucalyptus
  - Why not all? Because the API-Release differs!
- Some handy tools:
  - S3 Curl (command line tool)
    - <http://developer.amazonwebservices.com/connect/entry.jspa?externalID=128>
  - s3cmd (command line tool)
    - <http://s3tools.org/s3cmd>
  - s3fs (S3-Bucket can be mounted as local filesystem with FUSE)
    - <http://code.google.com/p/s3fs/>
  - ElasticFox (Firefox-Plugin)
    - <http://developer.amazonwebservices.com/connect/entry.jspa?externalID=609>

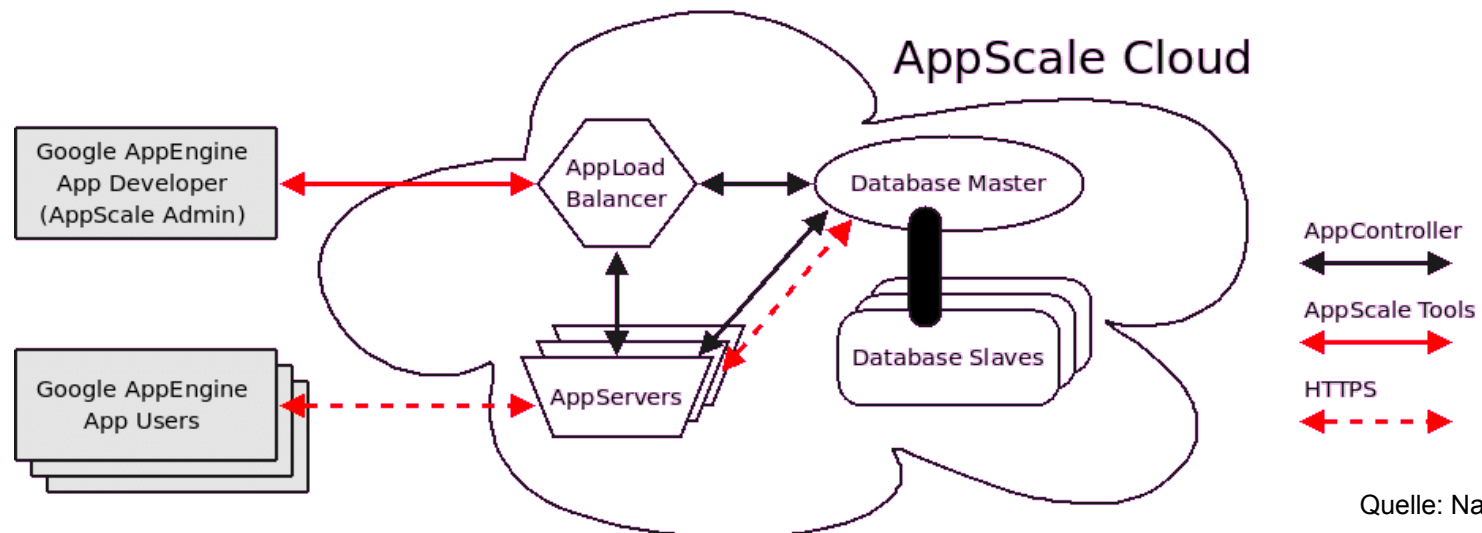
# ElasticFox

- Support to start, monitor and terminate instances in a user friendly GUI
- Much more comfortable than command line tools





- Open Source reimplementation of the Google AppEngine
  - AppEngine allows to run web applications written in Python (and JAVA) in the Google infrastructure
- Developed at UC Santa Barbara
- Implements Platform as a Service (PaaS)
- AppScale executes automatically and transparently in Eucalyptus
- With AppScale a PaaS Cloud infrastructure can be build up that allows to deploy, test, debug, measure and monitor Google AppEngine applications inside a Private Cloud

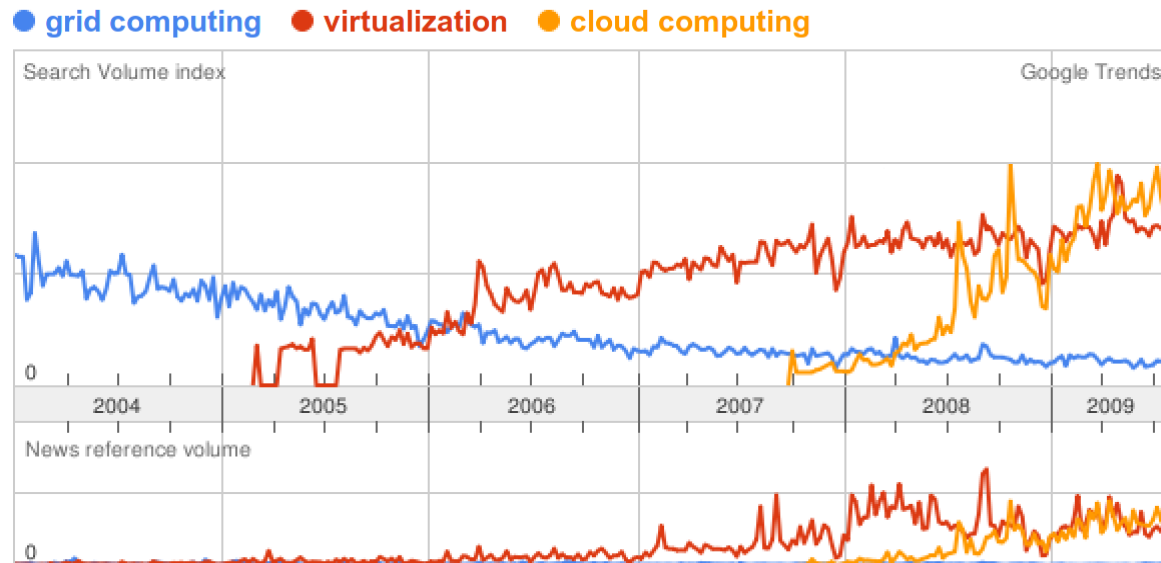


Quelle: Navraj Chohan



# Summary

- Cloud computing is the next big thing
  - Promising approach to solve some of the major challenges of IT
  - Flexible and elastic resource provisioning
  - Economy of scale makes it attractive
  - Move from manufacture towards industrialization of IT



- Eucalyptus and AppScale enable IaaS and PaaS as Open Source solutions with Linux
- With commodity hardware and Open Source software, a private cloud can be built up providing the same functionality and better performance compared to the most popular public clouds
  - ⇒ Vendor lock-in is not an issue any more

# Coming soon - this autumn

- Available in October 2009
- First cloud computing book in German language
- Covers the latest topics in cloud computing
- Only € 14,95



# Thank you for your attention

