Curso Grid y e-Ciencia 2010

Valencia, Spain, July 6-9 2010

Cloud Computing & Grid

Eduardo Huedo Cuesta

dsa-research.org

Distributed Systems Architecture Research Group
Universidad Complutense de Madrid

Universidad Complutense Madrid

La Suma de Todos
Consejería de Educación
Comunidad de Madrid
www.madrid.org

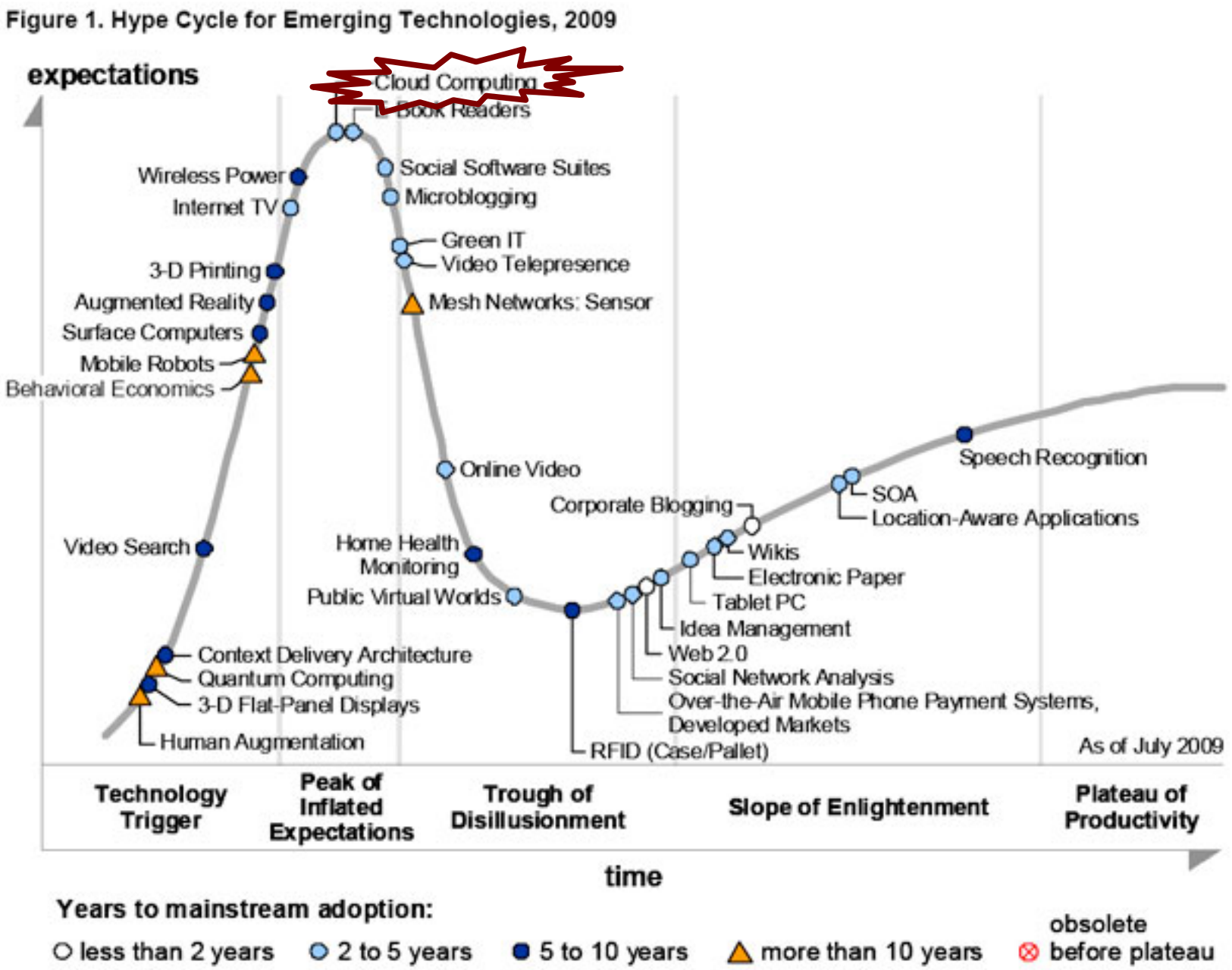
Ministerio de Educación y Ciencia

Seventh Framework Programme

Objectives

- **Provide an overview of Cloud Computing**
- Describe how Clouds can help Grids
- Discuss some experiences using Clouds and Grids

Cloud on the Hype Cycle

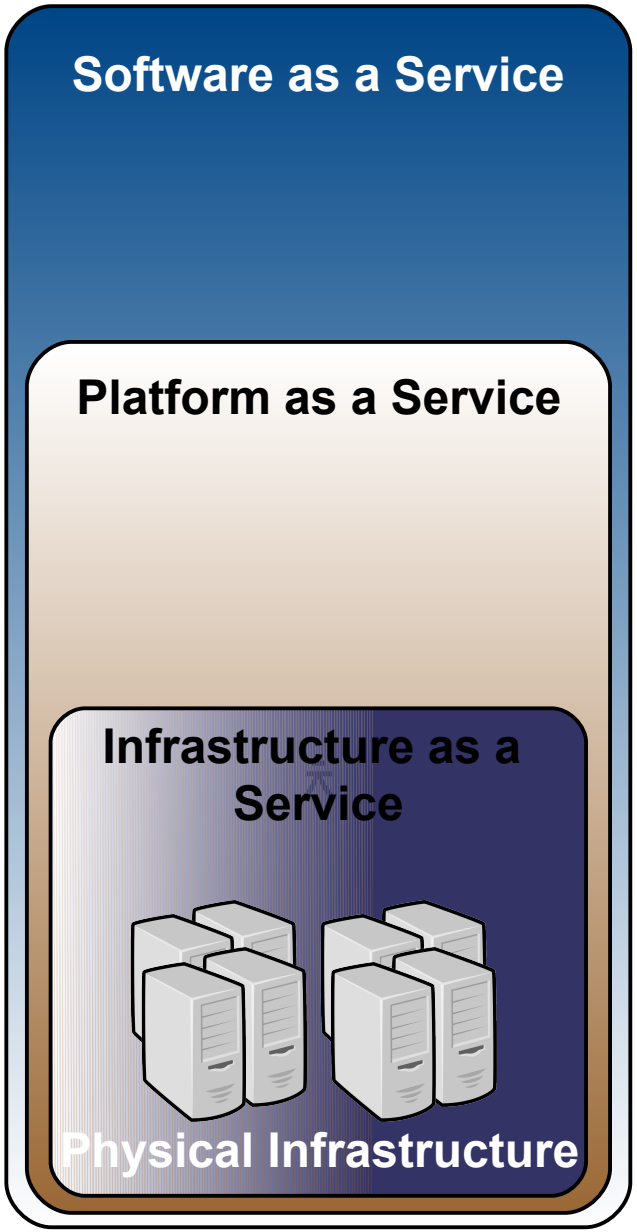


The “Cloud” Concept

- As a client...
  - You get a service from “the cloud” (* *as a service*)
  - You don't know what's behind, but it works (*transparency*)
  - You pay what you use (*utility*)
  - You immediately get what you ask for (*on demand*)
  - Later, you can get more (*elasticity*)

Most of these features have been used to describe grids, along with *sharing*, *interoperation*, *standardization*, *security*, *scalability*...

Cloud Computing in a Nutshell



What	Who
On-demand access to any application	**End-user** (does not care about hw or sw) skype facebook
Platform for building and delivering web applications	**Developer** (no managing of the underlying hw & sw layers) Windows Azure force.com platform as a service
Delivery of a *raw* computer infrastructure	**System Administrator** (complete management of the computer infrastructure) GOGRID flexiscale amazon web services

The Public IaaS Cloud

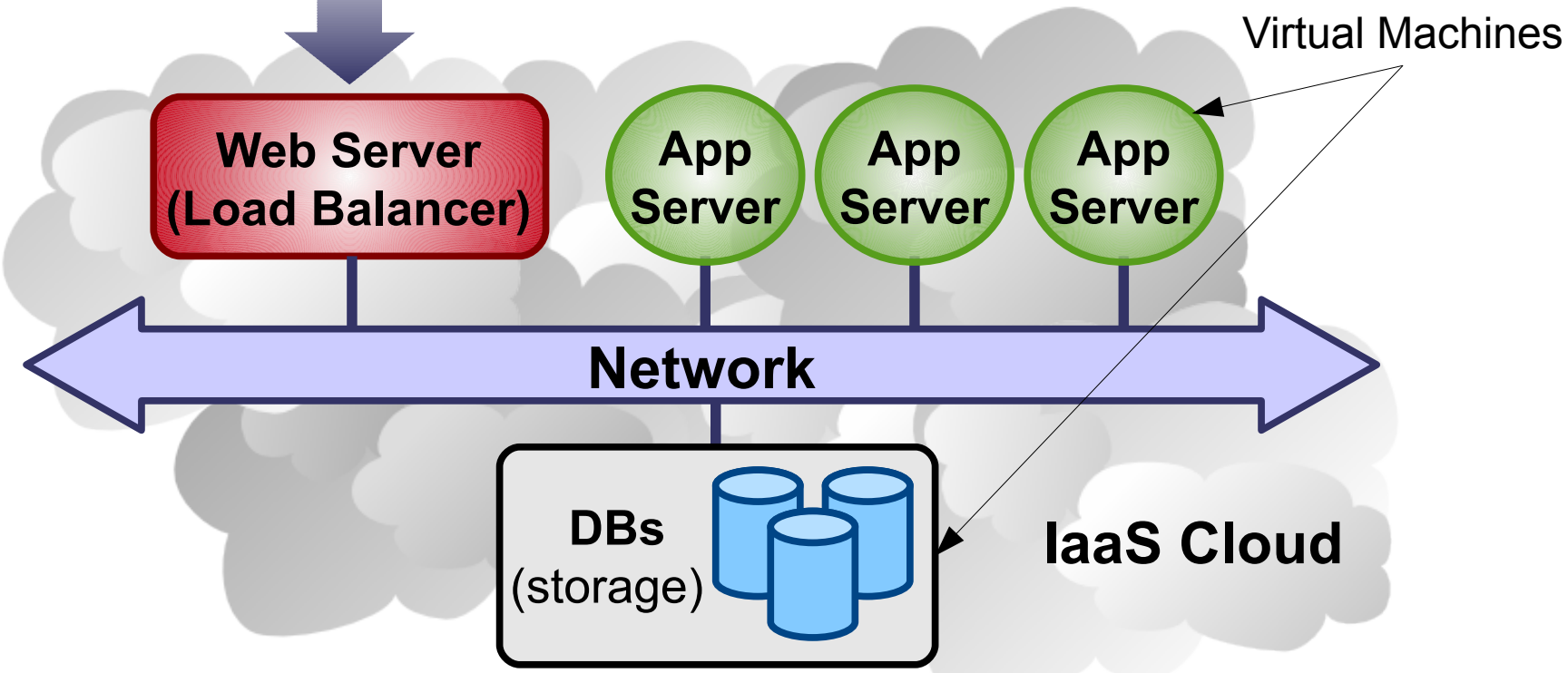
- **Simple Web interface (REST)**
- **Raw *infrastructure* resources**
  - Total control of the resources
  - Capacity leased in the form of VMs
    - Virtualization is the key enabler of IaaS clouds
    - Complete Service-HW decoupling
- **Pay-as-you-go (On-demand access)**
  - A single user can not get all the resources
  - Multi-tenancy
  - Simplified accounting
- **Elastic & *"infinite" c*apacity**

The Public IaaS Cloud

Service End-Users

Total control of service layout

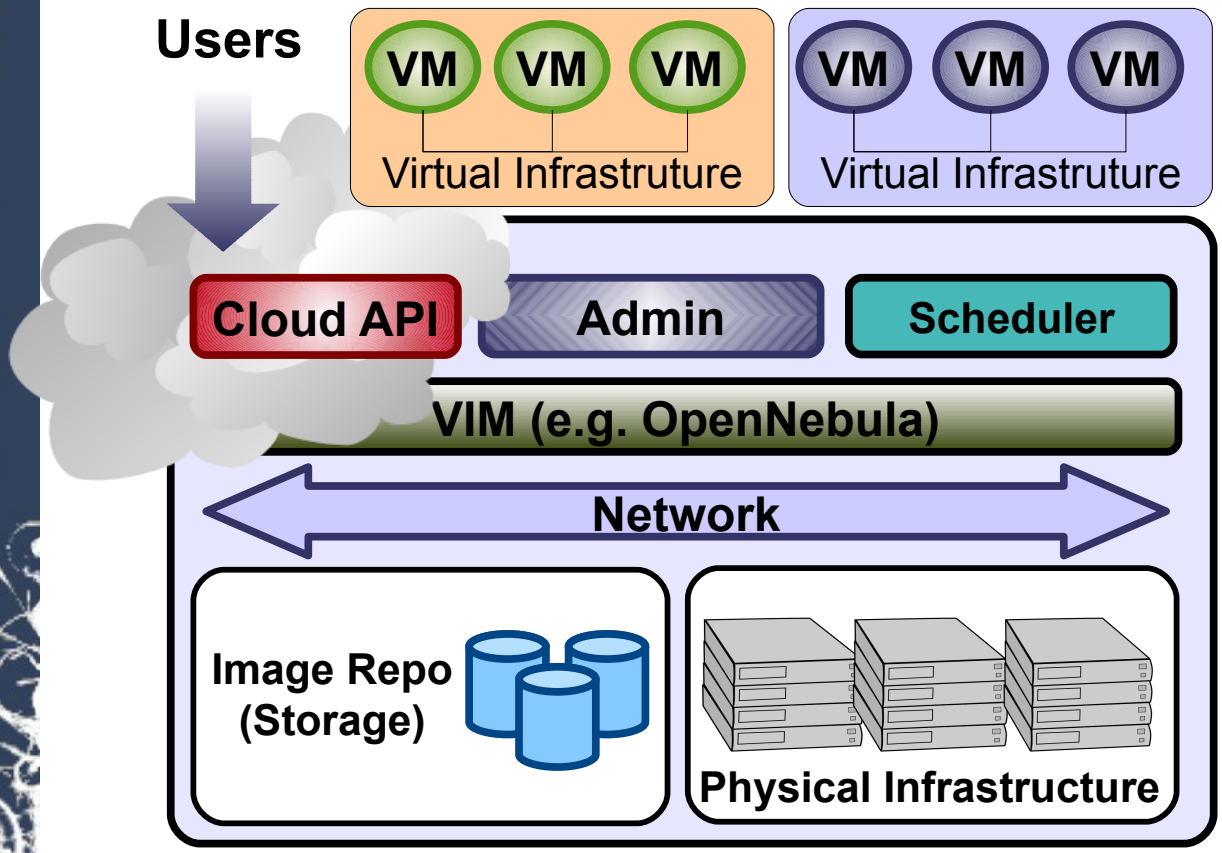
- Software stack
- Type & number of components
- Service elasticity



The Private IaaS Cloud

A "*Public Cloud behind the firewall*"

- Security concerns
- Flexible management (consolidation, adaptation, provisioning...)

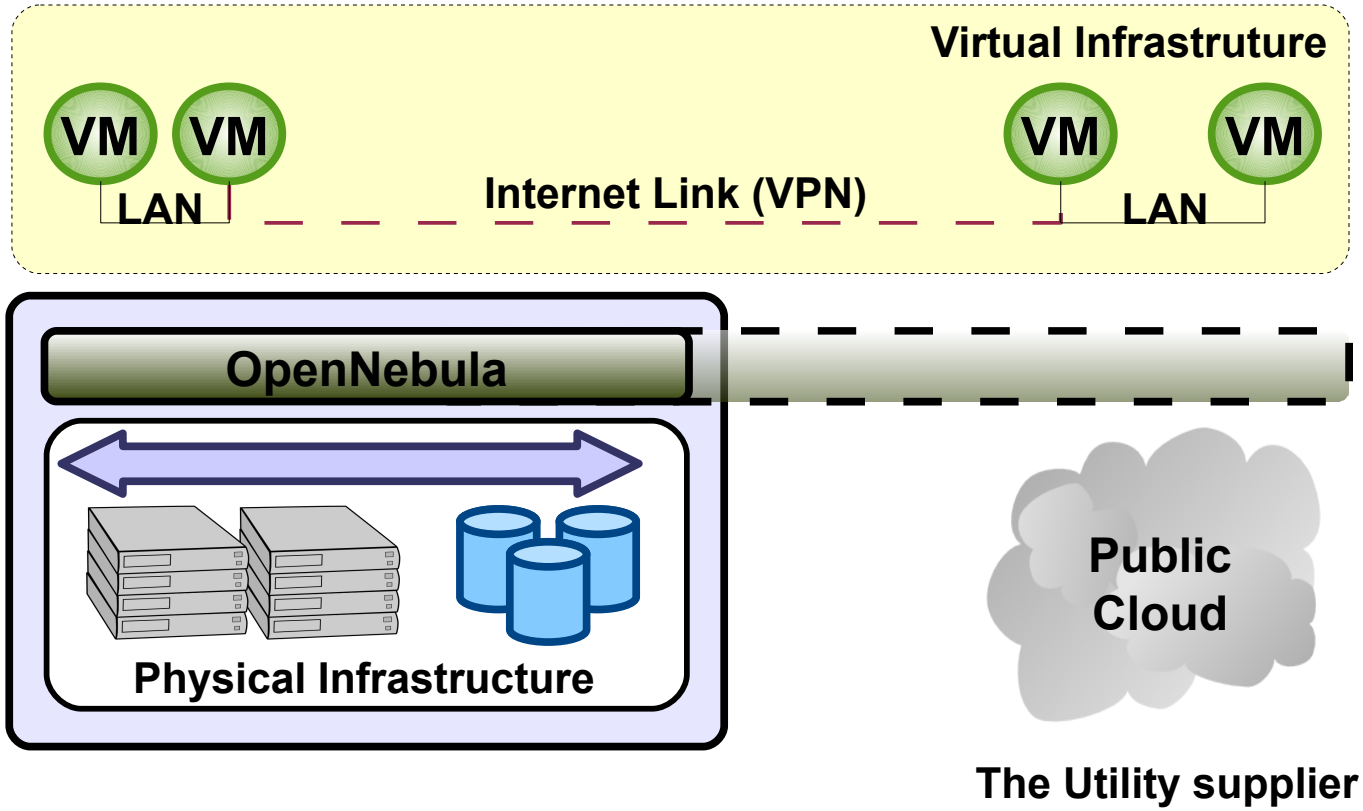


The headaches...

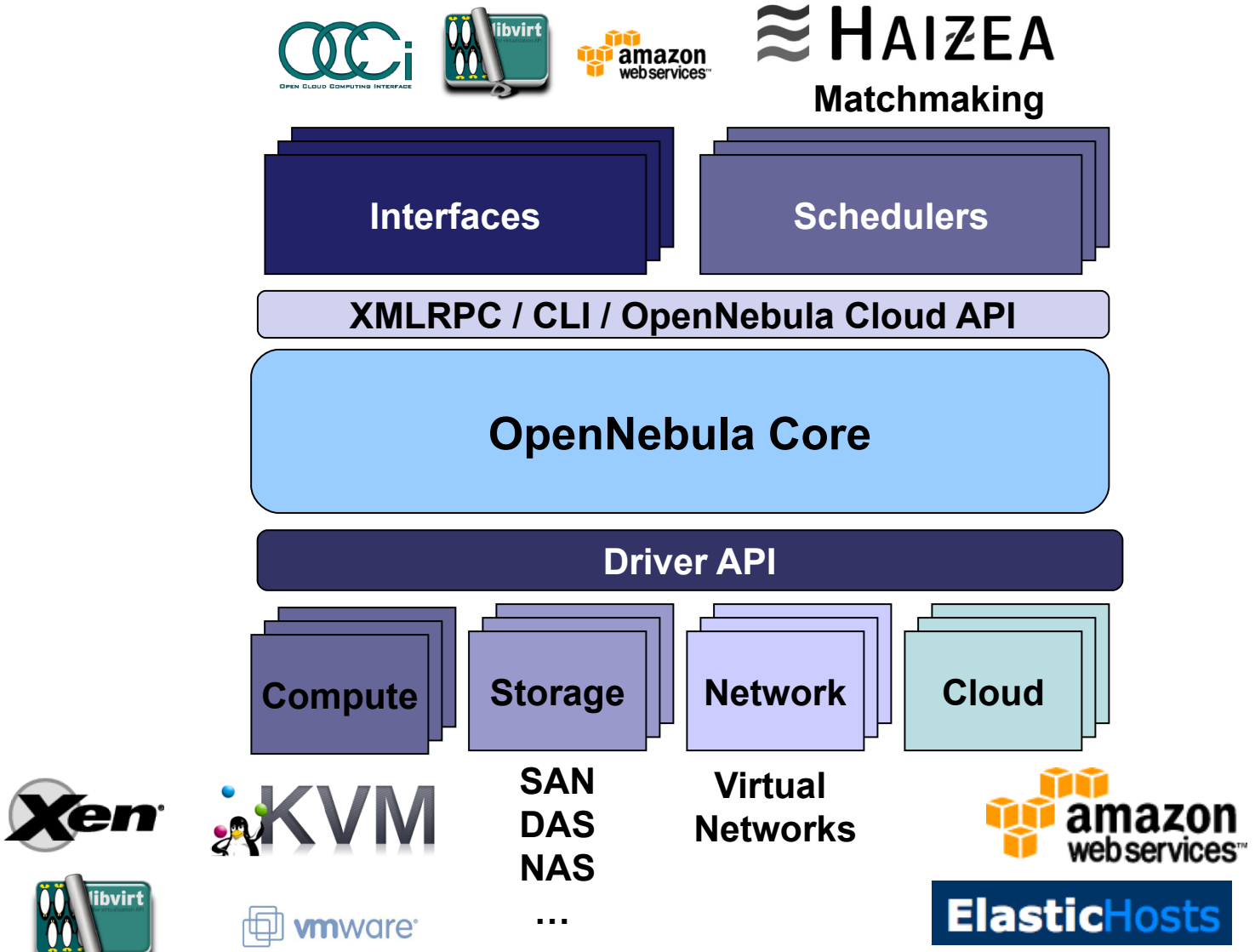
- Orchestrate:
  - Virtualization
  - Networking
  - Storage
- Admin interfaces
- VM placement

The Hybrid IaaS Cloud

- Supplement the capacity of the local infrastructure
- Transparent access to the resulting hybrid cloud
- Utility Computing dream made a reality!



OpenNebula VIM



Deployment modes

Model	Definition	Examples of Deployment
Private	Infrastructure is owned by a single organization and made available only to the organization	• Optimize and simplify **internal operation** • **SaaS**/**PaaS** support • IT consolidation within **large organizations** (Goverment Clouds, University Clouds…)
Public	Infrastructure is owned by a single organization and made available to other organizations	• **Commercial cloud providers** • **Community public clouds** by ICT service centers to enable scientific and educational projects to experiment with cloud computing • **Special purpose clouds** with dedicated capabilities (Science Clouds, HPC Clouds..) • **Regional clouds** to address regulatory or latency issues
Hybrid	Infrastructure is a composition of two or more clouds	• **Cloudbursting** to address peak demands • **Cloud Federation** to share infrastructure with partners • **Cloud Aggregation** to provide a larger resource infrastructure

Objectives

- Provide an overview of Cloud Computing
- **Describe how Clouds can help Grids**
- Discuss some experiences using Clouds and Grids

Current Grid Infrastructures...

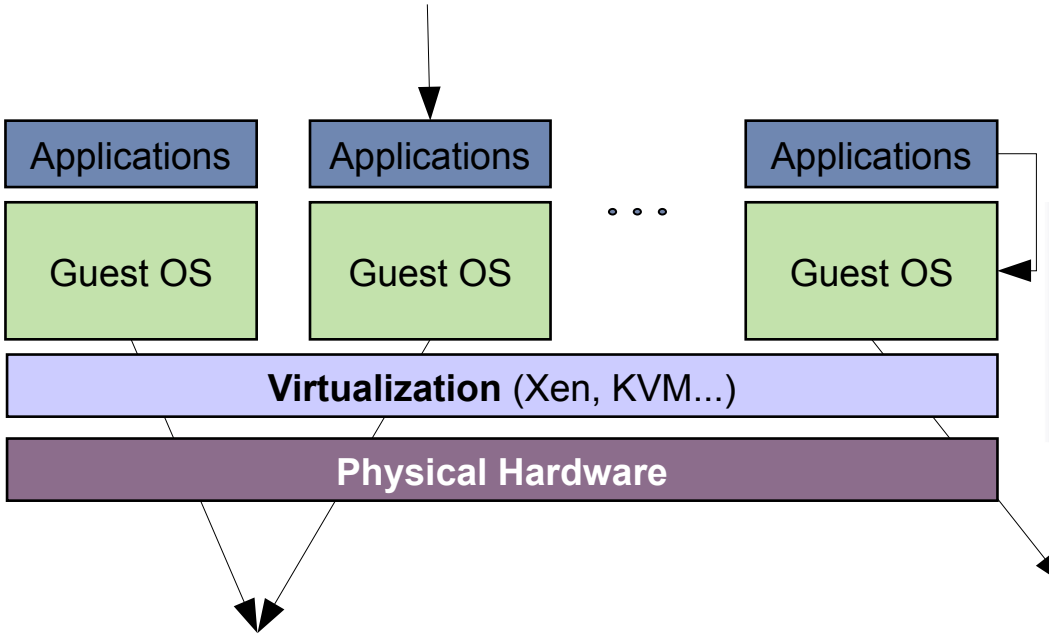
- High degree of heterogeneity (software & hardware)
- High operational costs
- Isolate and partition resources contributed to the Grid
- Specific environment requirements for different VOs
- Users simply do not feel like adopting our execution models (*pilot jobs*...)

Grids are difficult to maintain, operate and use

Grids, Clouds... and Virtual Machines

- A VM is an isolated runtime environment (guest OS and apps)
- Hypervisors: full virtualization, para-virtualization, HW-aided virtualization, kernel-based virtualization

Execution of legacy applications



Natural way to deal with the *heterogeneity*

Application / HW decoupling

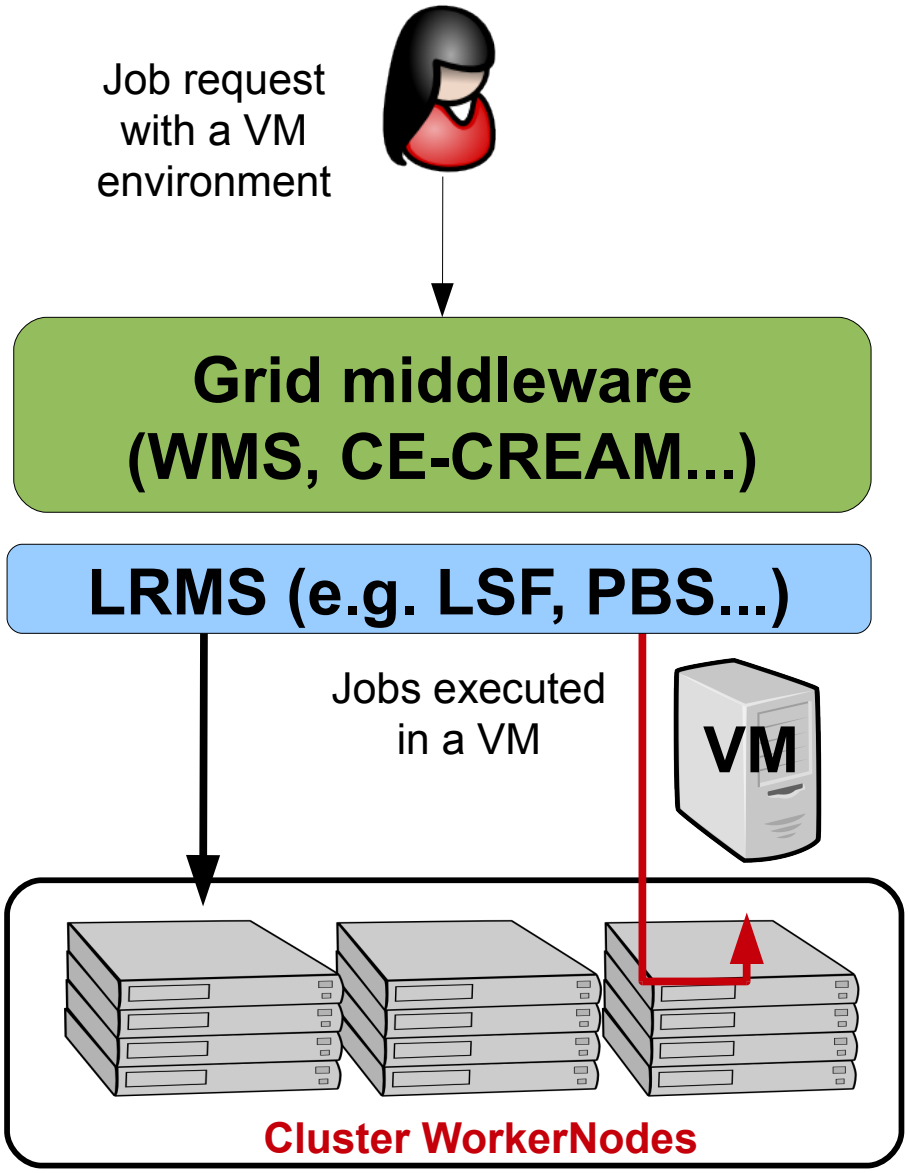
Domains are isolated

VMs can be provided with a **Cloud-like** model!

Grids, Clouds... and Virtual Machines

- Use VMs as basic building block for Grid Services
- **Current Trends**:
  - **VMs as Job Container**
  - **VMs as Grid execution service component**
    - Deal with heterogeneity
    - Simplify & improve site management
    - Give VOs control over the worker-node SW
  - **IaaS interfaces for a Grid Site**
    - Attract business users
    - Support novel execution models

Grids, Clouds: VM as a Job Container



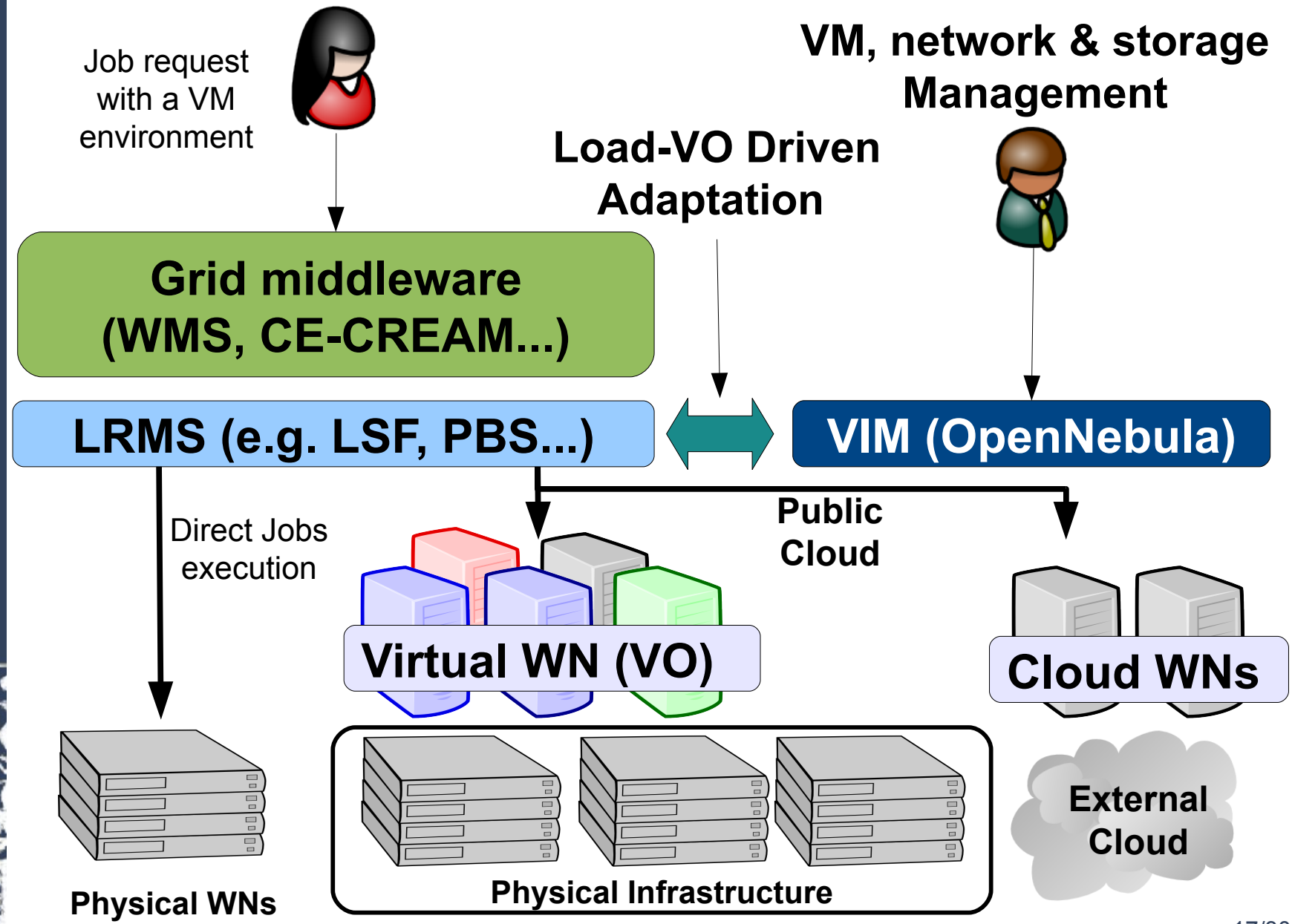
Features

- Single LRMS-based
- Integrated with Grid MW

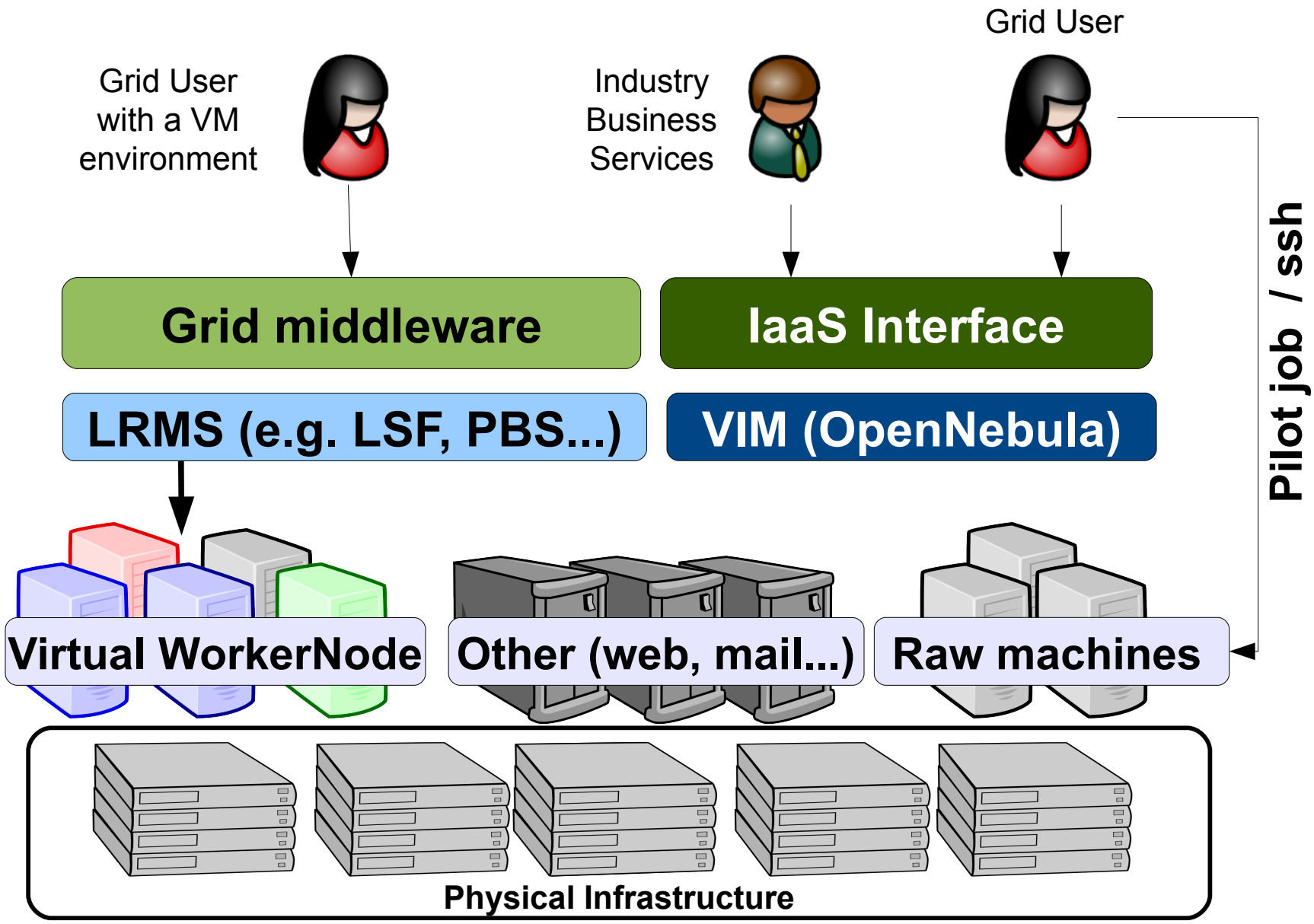
Drawbacks

- Jobs and VMs are different
- Need to integrate other resources (network, storage...)
- Do not decouple totally infrastructure from the Grid services
- Can not leverage VM features (e.g. migration...)
- Focused on job execution

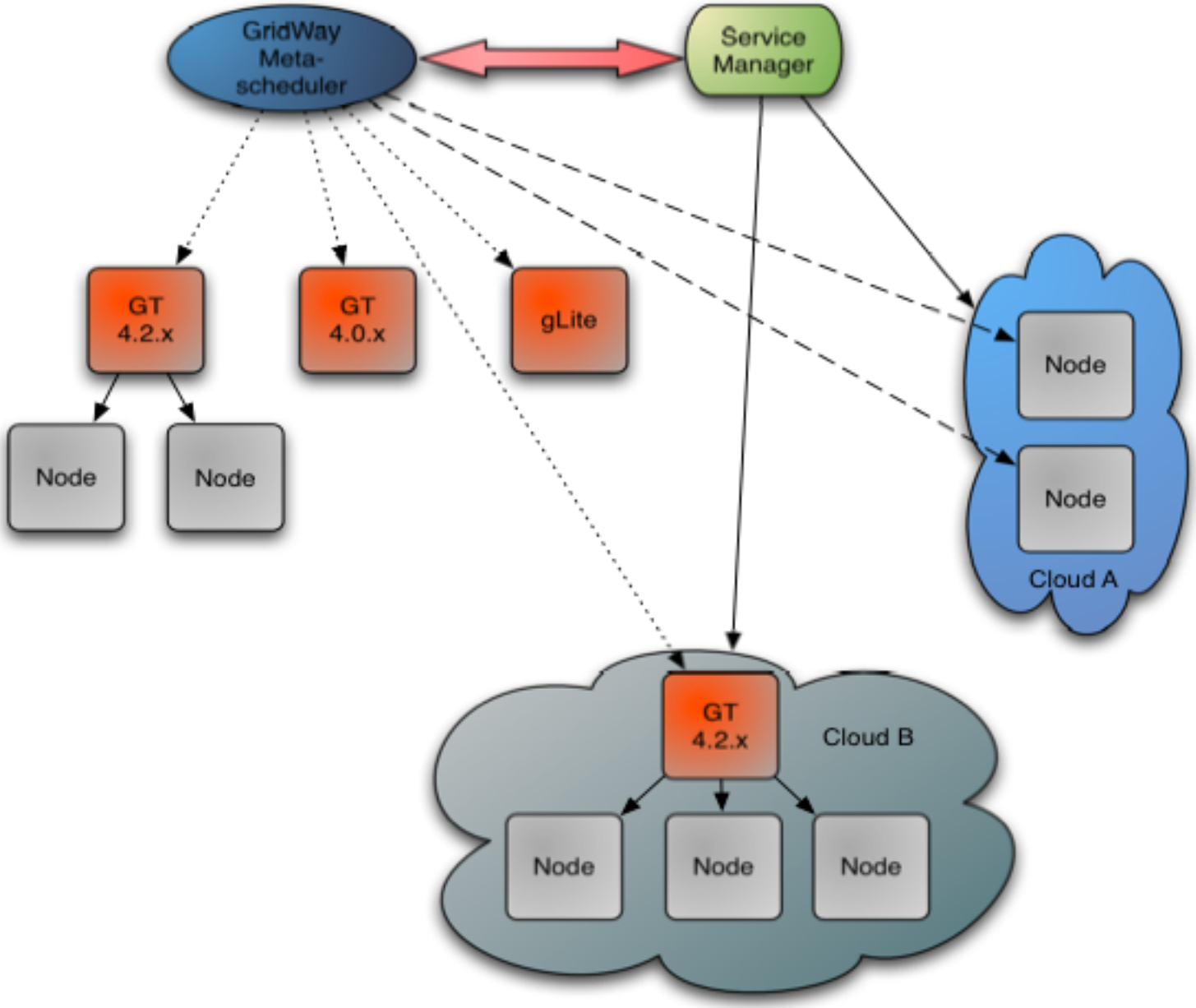
Grids, Clouds: VM as Grid Service Component



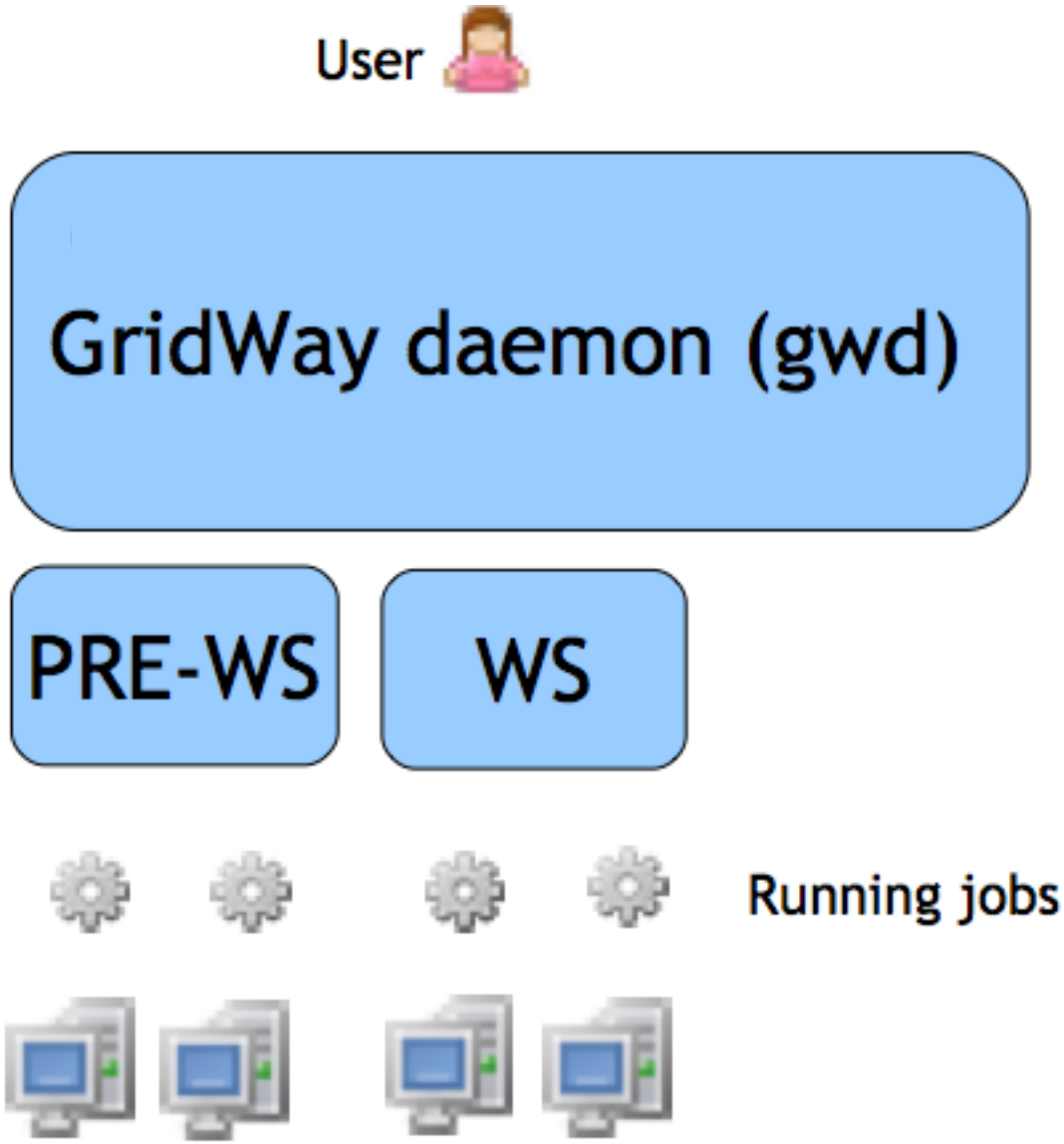
Grids, Clouds: IaaS Interface for a Site



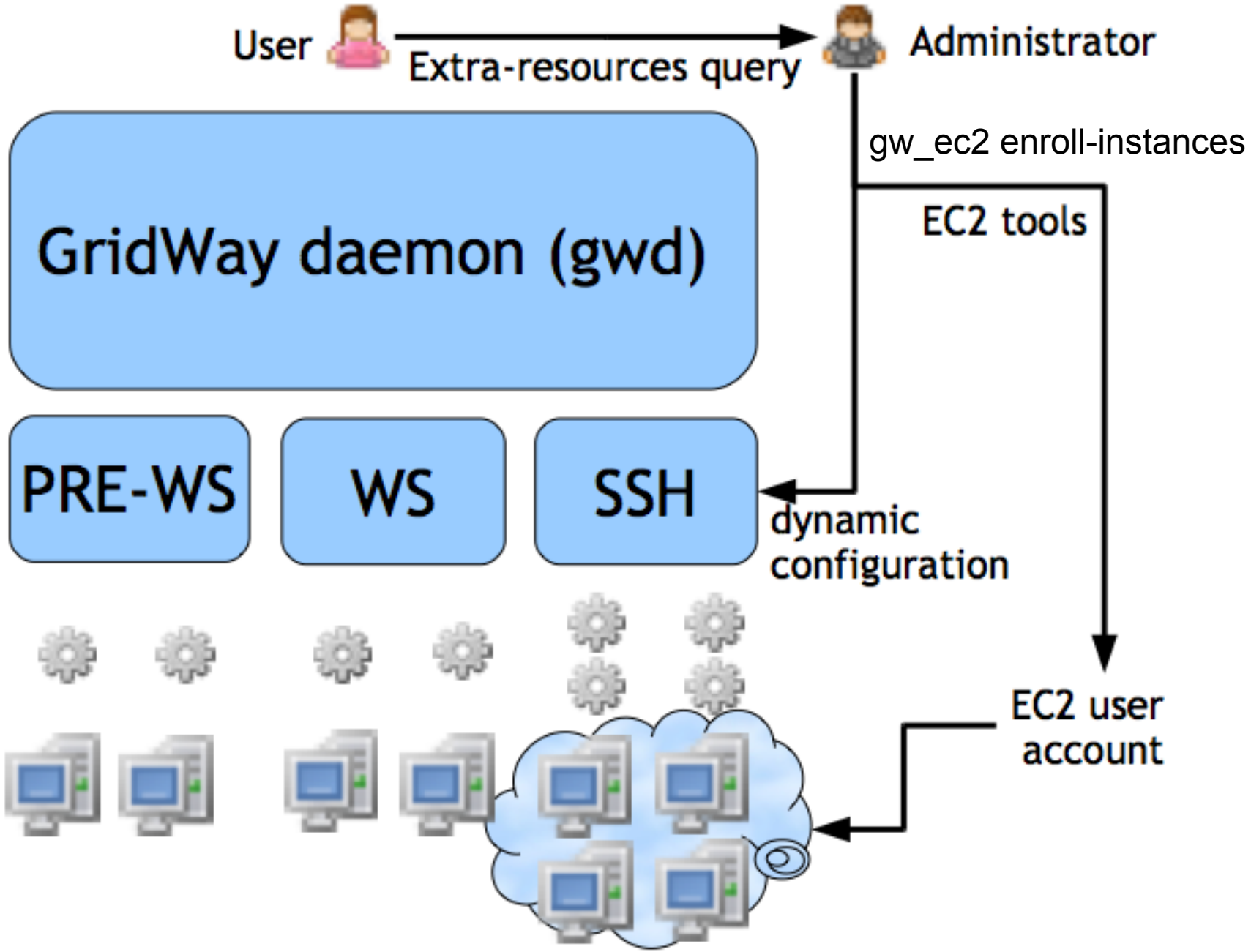
Grids, Clouds: Interoperation



Grids, Clouds: Interoperation



Grids, Clouds: Interoperation



Benefits of Cloud for Grids

- Easy support for VO-specific worker nodes
- Reduce *gridification* cycles
- Dynamic balance of resources between VO's and so maximize utility
- Fault tolerance of key infrastructure components
- Easier deployment and testing of new middleware distributions
- Distribution of pre-configured components
- Cheaper development nodes
- Simplified training machines deployment
- Simplified operation of grid sites
- Performance partitioning between local and grid services

Solve many of the obstacles to Grid adoption

Possible Roadmap for Adopting Cloud in Grid

1. Virtualization of grid sites using **private** cloud technologies
   - Provide a flexible infrastructure
2. IaaS delivery using **public** cloud technologies
   - Attract new user communities
3. Cloud scale-out using **hybrid** cloud technologies
   - Exchange resources (VMs) instead of jobs

And, in parallel...

Interoperation of grid infrastructures (physical grids) and cloud resources (virtual grids)

- Leverage existing infrastructures

Objectives

- Provide an overview of Cloud Computing
- Describe how Clouds can help Grids
- **Discuss some experiences using Clouds and Grids**

First Experiences

VMs as a Job Container

- INFN: Workernode on demand

> "Enabling Distributed Job Submission in Dynamic Virtual Execution Environments for EGEE Users", D.Salomoni, M.Cecchi, A.Ghiselli, A.Italiano, M.Orrù, D.Rebatto, V.Venturi, L.Zangrando

VMs as a Grid Service Component

- CERN: LSF + VMO/OpenNebula + Custom VM images. Tested with real life Grid experiments (ALICE)

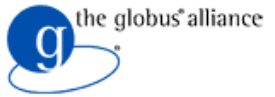
> "The batch virtualization project at CERN", Sebastien Goasguen, Ewan Roche, Tony Cass and Schwickerath Ulrich.

- NIKHEF: Torque/PBS + OpenNebula + CoW VM repository

> "Virtual Machines at a Tier-1 site." Sander Klous (Nikhef)

IaaS for Scientific Computing

- Globus Nimbus – AliEn & STAR experiments



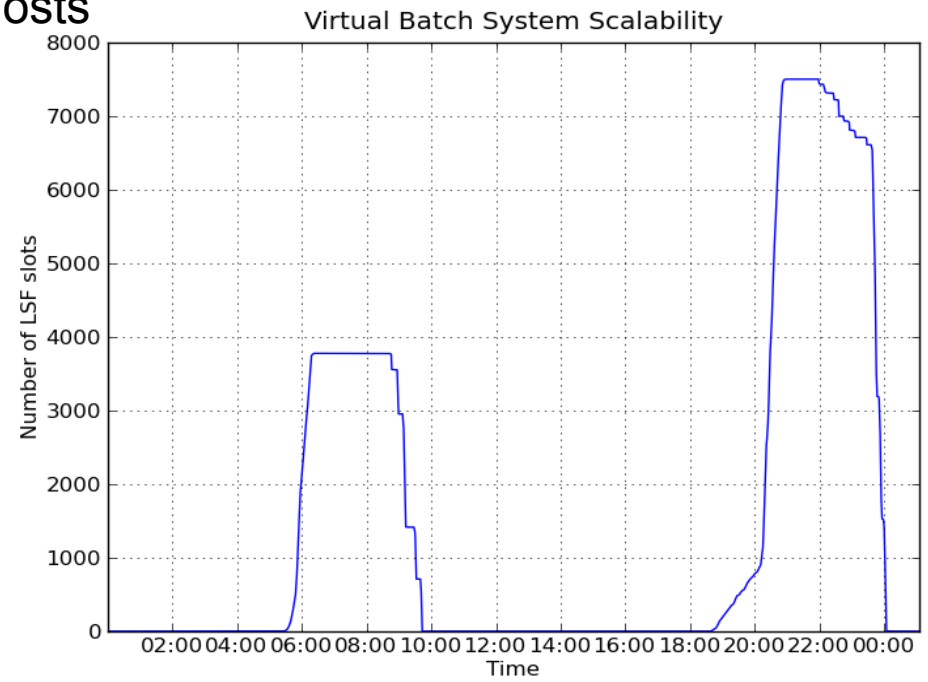
> http://workspace.globus.org/

Deployment Cases

Private Cloud to Support Batch Farm

- **IT-PES/PS Group**: Sebastien Goasguen, Ulrich Schwickerath, Ewan Roche and Belmiro Moreira
- **Configuration Management**: Quattor with lifecycle management and “self -notification” in OpenNebula
- **Network Management**: Adapted to address network infrastructure requirements regarding fixed IP/MAC leases in each box
- **Storage Management**: New LVM transfer scripts and a very fast parallel scp to push images to all the hosts

Up to 7,500 VMs on 400 hosts (3,200 cores) running Xen



Deployment Cases

Private Cloud to Support Grid Site

- The D-Grid Resource Center Ruhr (DGRZR) runs an OpenNebula private cloud on 248 blades and 1,984 cores with Xen
- OpenNebula is used to support the execution of a virtualized Grid site in D-Grid and EGEE

Public HPC Cloud

- SARA High Performance Computing Center uses OpenNebula in its new HPC Cloud service on 128 cores across 16 servers with KVM
- OpenNebula is used to support the execution of virtual clusters and HPC applications
- Authors of the OpenNebula Management Console

First Initiatives

- Study Public Clouds (Amazon EC2) to deploy an EGEE site
- EGEE site as a private cloud to deploy Grid services
- IaaS Interfaces for EGEE sites
- StratusLab will integrate, distribute and maintain a cloud turn-key toolkit for ~~EGEE~~ EGI sites

EGEE & RESERVOIR Collaboration

- Use Virtual Worker-nodes to provision Grid Services
- Explore Hybrid Cloud Computing for Grid Sites
- Virtualize a Complete EGEE site

The StratusLab EU Project

Vision

- Grid and cloud embody **complementary computing models** that will coexist and cooperate in existing and future e-infrastructures

Aim

- Incorporate **cloud innovation into existing Grid infrastructures** to:
  - **Simplify** and **optimize** its use and operation, providing a more **flexible**, **dynamic** computing environment for scientists.
  - **Enhance existing infrastructures with IaaS cloud paradigms**

Evolutionary Approach

- **Complement existing services**, being fully transparent to upper layers
- **Existing Grid middleware would continue** to provide the glue to federate the distributed resources and the services for high-level job and data management
- **Address the emerging IaaS cloud-like usage patterns**

Service Centred Project driven to support production infrastructures

- Integration, distribution, testing and maintenance the **StratusLab Toolkit**
- **Management of images** (trust, provenance…)

Conclusions

- Virtualization, cloud, and grid are complementary technologies and will coexist and cooperate at different levels of abstraction
- Virtualization can solve many obstacles for grid adoption
- Virtualization and cloud do NOT require any modification from the end-user perspectives
- Separation between service and infrastructure layers will allow the application of the utility model to Grid computing
- The use of cloud interfaces for grid sites may attract other users to e-infrastructures
- Platforms combining cloud and grid will offer their users a better service
- We'll see cloud activities in the future EGI