Scientific Computing Seminar
Stanford, July 10th, 2014

SLAC NATIONAL ACCELERATOR LABORATORY

Bringing Private Cloud Computing to HPC and Science

Ignacio M. Llorente

OpenNebula Project Director

OpenNebula.org

Building Private Cloud Computing to HPC and Science

This presentation is about:

- The Private HPC Cloud Use Case
- Main Challenges for Private HPC Cloud
- Resource Provisioning Framework
- Private HPC Cloud Case Studies
- Grid and Cloud

Simple but feature-rich, production-ready, customizable solution to build clouds

SIMPLE

Easy to operate, install and upgrade, with packages for the main Linux distributions

FLEXIBLE

Really open-source and customizable to fit into any data center and policies

ROBUST

Production-ready, mature, reliable and commercially supported

POWERFUL

Innovative functionality for enterprise clouds and data center virtualization

What is OpenNebula?

From Research Project to Open-source Project for Enterprise

2005
Research Project

2008
OpenNebula.org
- Develop & innovate
- Support the community
- Collaborate

2009
SEVENTH FRAMEWORK PROGRAMME

2010
Large-scale production deployment: 16,000 VMs

2011
5,000 down/month

2013
OpenNebula Conf
24. - 26. Sept 2013 Berlin

2014
Very large-scale production deployment: 200,000 VMs (10 zones)

Tech Days
bit internet technology
TransUnion
hyve solutions

2015
Hybrid cloud partners
SOFTLAYER
Windows Azure

An Open Community Driven by Users

Use
Give us feedback
Contribute experiences
Help support new users

Develop
Open-source Apache code
Transparent process
Public roadmap

one

Communicate
User groups
Cloud technology days
OpenNebulaConf

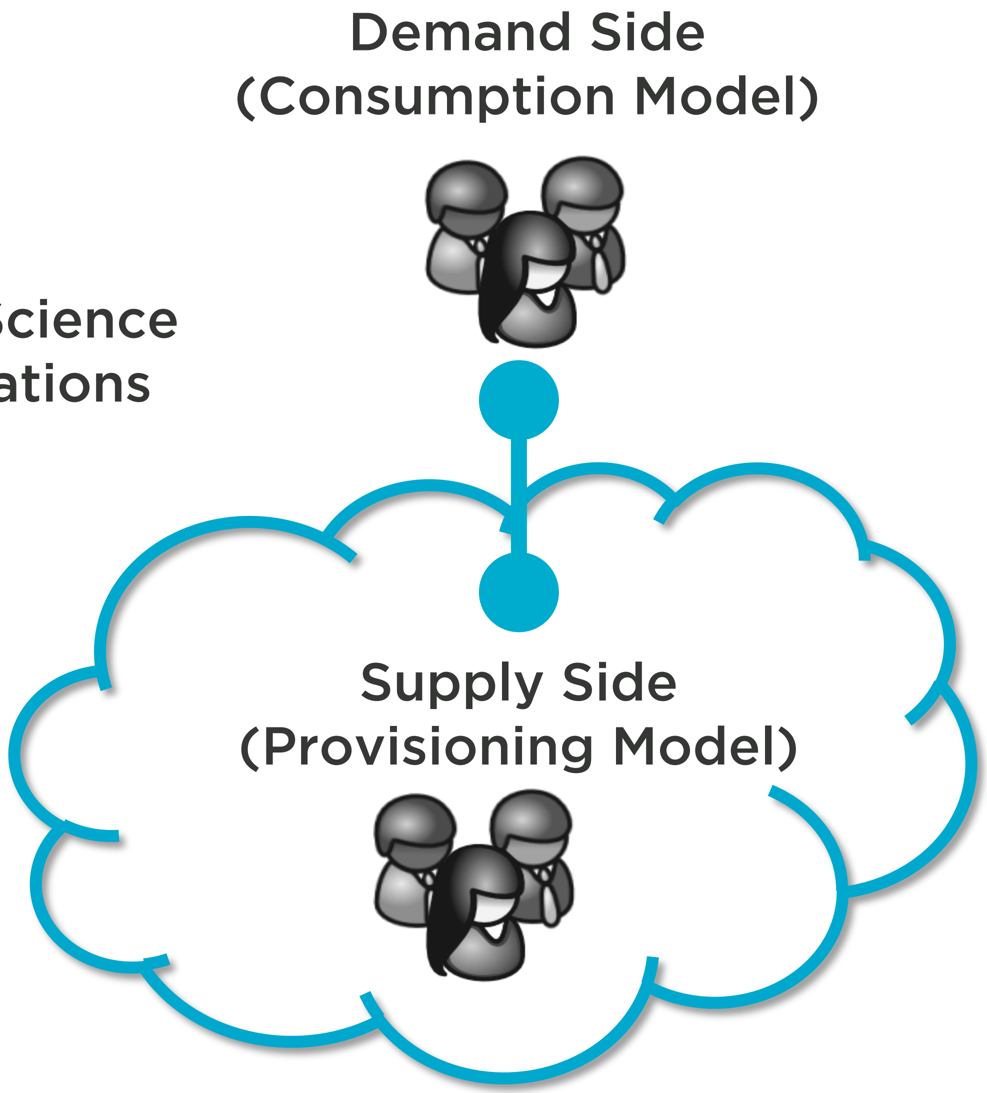
Integrate
Add-ons catalog
Ecosystem directory

What is OpenNebula?

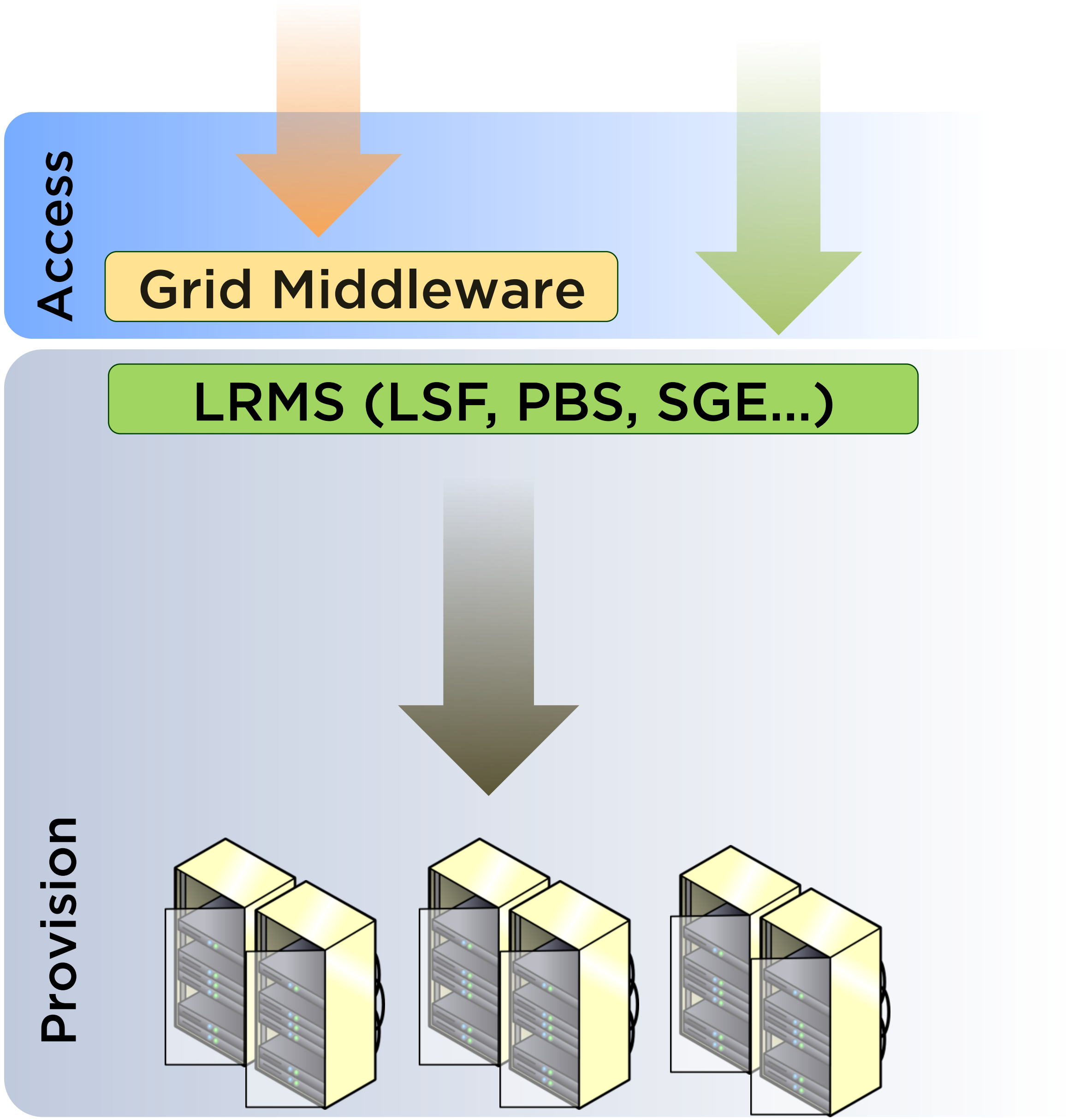
Building and Operating Enterprise Private Clouds in Medium and Large Data Centers

Different Perspectives to Present Innovations in Cloud Computing

HPC & Science Applications

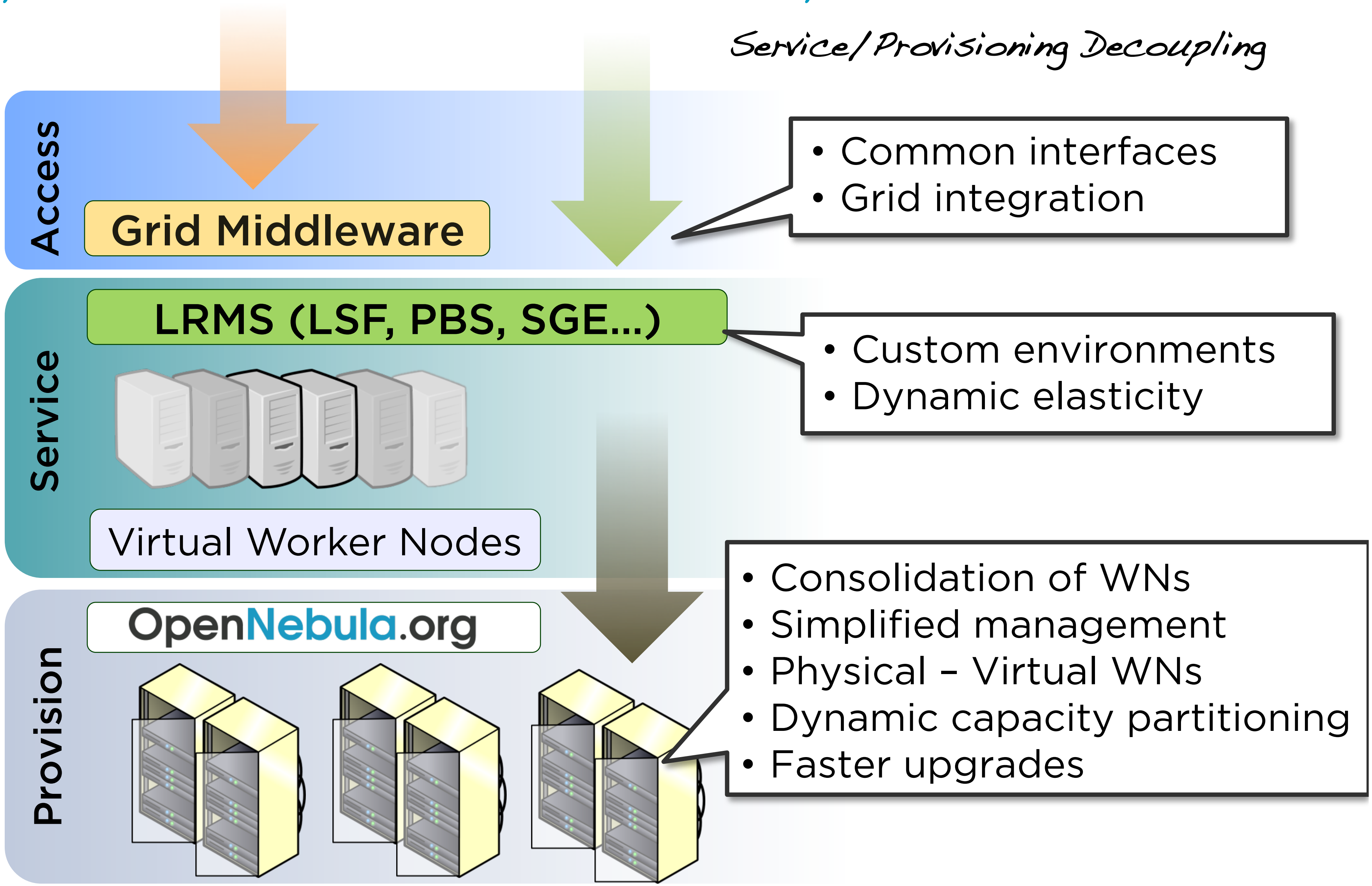


The Pre-cloud Era

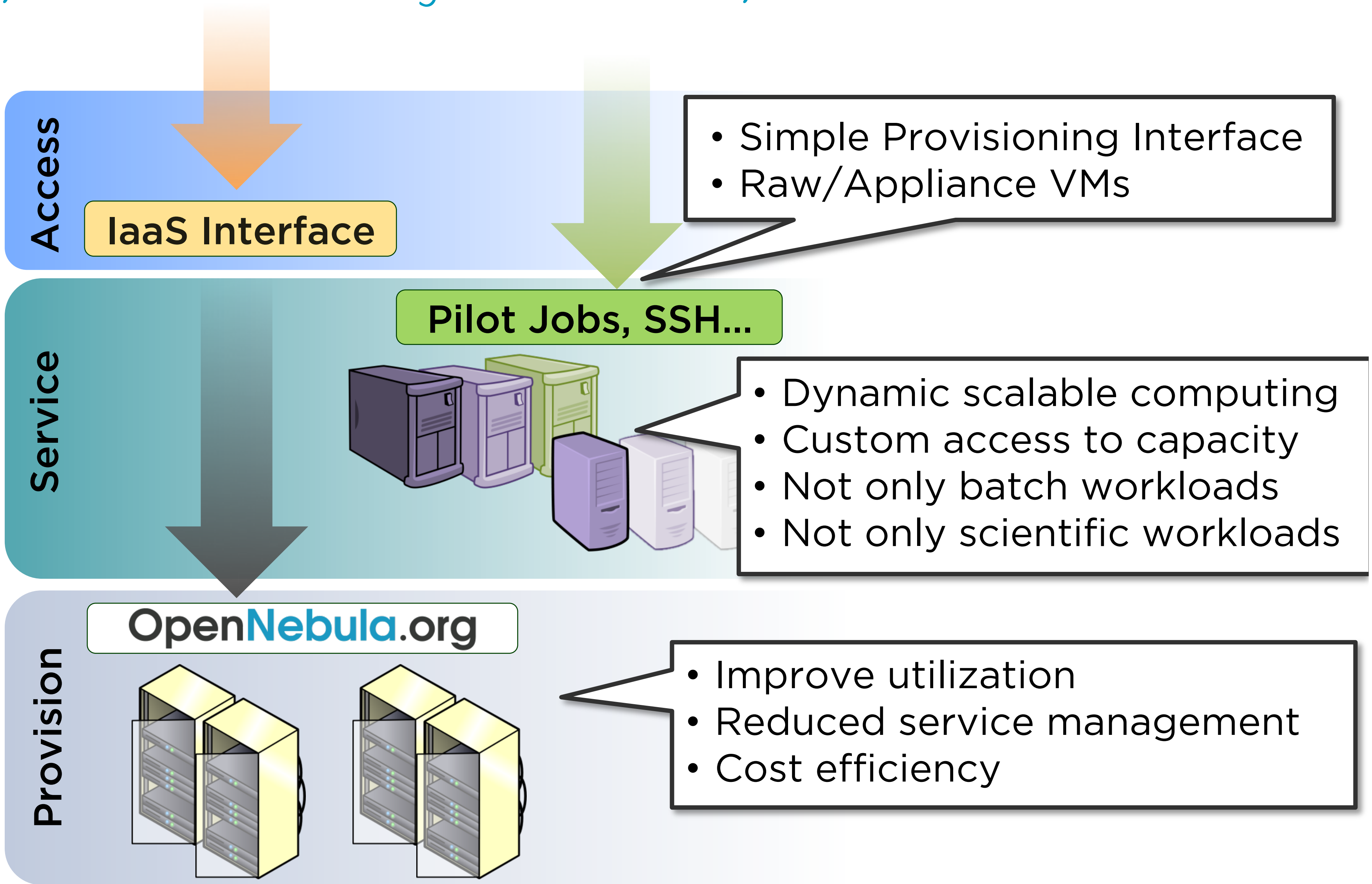


The Private HPC and Science Cloud Use Case

OpenNebula as an Infrastructure Tool – Enhanced Capabilities



OpenNebula as an Provisioning Tool – Enhanced Capabilities



Main Challenges for Private HPC Cloud

Main Demands from Engineering, Research and Supercomputing

Flexible Definition of Multi-tier Applications

Application Performance

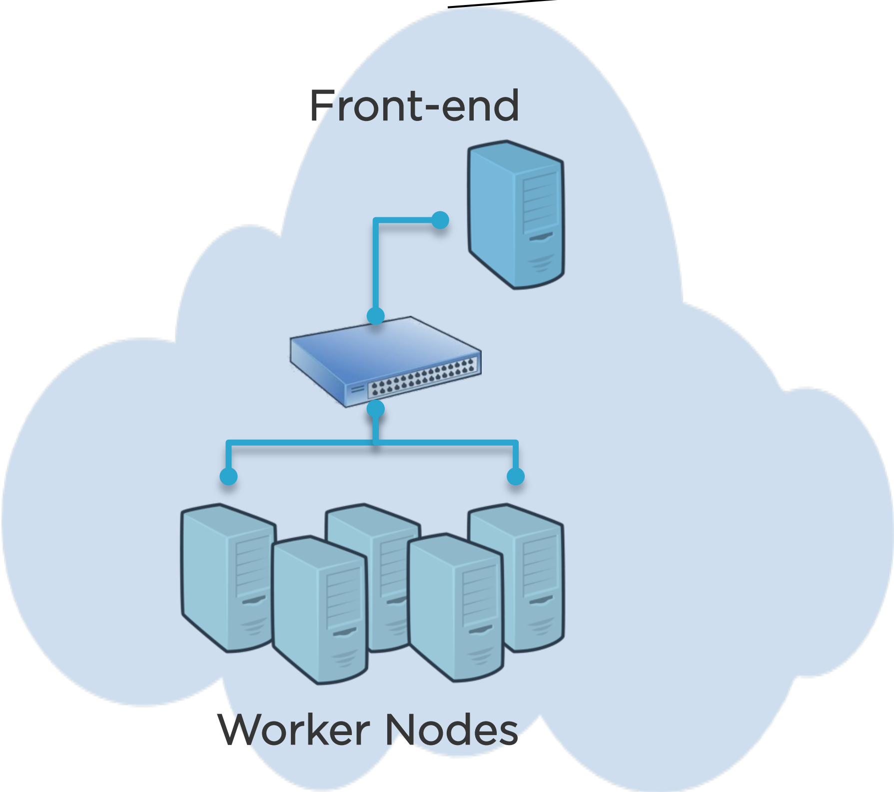
Resource Management

Provisioning Model

Execution of Multi-tiered Applications

A Comprehensive Framework to Manage Complex Applications

- Several tiers
- Deployment dependencies between components
- Each tier has its own cardinality and elasticity rules

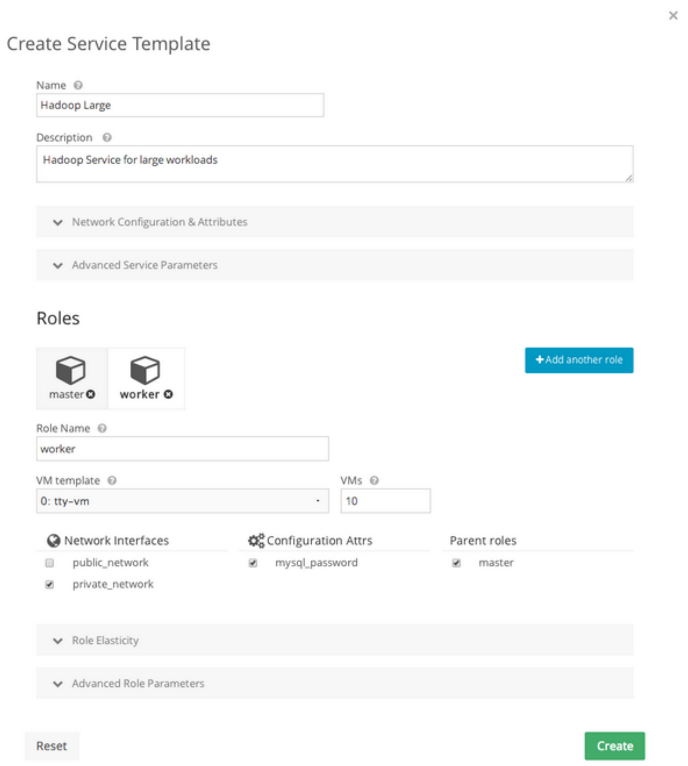


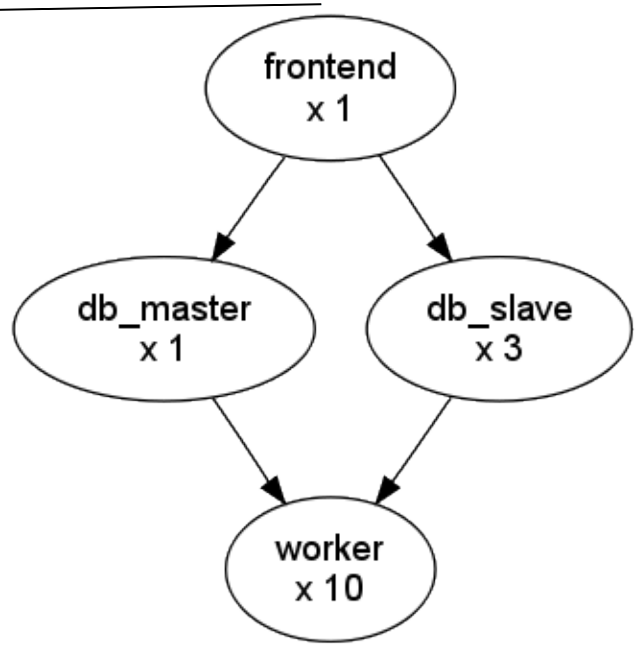

```
{ "name": ”Computing_Cluster",
  "deployment": "straight",
  "roles": [
    {
      "name": "frontend",
      "vm_template": 0
    }, {
      "name": "worker",
      "parents": frontend,
      "cardinality": 2,
      "vm_template": 3,
      "min_vms" : 1,
      "max_vms" : 5,
      "elasticity_policies" :  {
         ”expressions" : ”CPU> 90%”,
         "type" : "CHANGE",
         "adjust" : 2,
         "period_number" : 3,
         "period" : 10}, …
```

Using the Cloud - Execution of Multi-tiered Applications

Management of interconnected multi-VM applications:

- Definition of **application flows**
- **Catalog** with pre-defined applications
- **Sharing** between users and groups
- Management of **persistent scientific data**
- Automatic **elasticity**





Main Challenges for Private HPC Cloud

Performance Penalty as a Small Tax You Have to Pay

Overhead in Virtualization

- Single processor performance penalty between **1% and 5%**
- NASA has reported an overhead between **9% and 25%** (HPCC and NPB)[1]
- Growing number of users demanding containers (**OpenVZ** and **LXC**)

Overhead in Input/Output

- Growing number of **Big Data apps**
- Support for **multiple system datastores including automatic scheduling**

Need for Low-Latency High-Bandwidth Interconnection

- Lower performance, **10 GigE** typically, used in clouds has a significant negative (**x2-x10,** especially latency) impact on HPC applications[1]
- FermiCloud has reported MPI performance (HPL benchmark) on VMs and SR-IOV/**Infiniband** with **only a 4%** overhead[2]
- The Center for HPC at CSR has contributed the **KVM SR-IOV Drivers for Infiniband**[3]

(1) An Application-Based Performance Evaluation of Cloud Computing, NASA Ames, 2013
(2) FermiCloud Update, Keith Chadwick!, Fermilab, HePIX Spring Workshop 2013
(3) http://wiki.chpc.ac.za/acelab:opennebula_sr-iov_vmm_driver , 2013

Main Challenges for Private HPC Cloud

Resource Management

Optimal Placement of Virtual Machines
- Automatic placement of VM near input data
- Striping policy to maximize the resources available to VMs

Fair Share of Resources
- Resource quota management to allocate, track and limit resource utilization

Isolated Execution of Applications
- Full Isolation of performance-sensitive applications

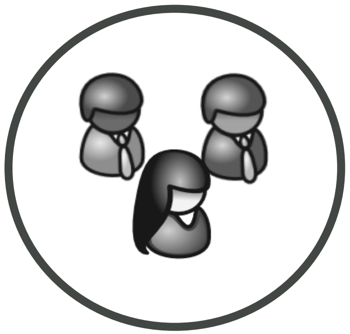
Management of Different Hardware Profiles
- Resource pools (physical clusters) with specific Hw and Sw profiles, or security levels for different workload profiles (HPC and HTC)

Hybrid Cloud Computing
- Cloudbursting to address peak or fluctuating demands for no critical and HTC workloads

Provide VOs with Isolated Cloud Environ
- Automatic provision of Virtual Data Centers

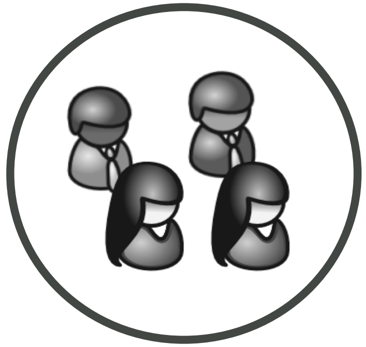
Challenges from the Organizational Perspective

Bio HTC Simulations



HPC Simulations

Big Data Analysis



Comprehensive Framework to Manage User Groups

- Several divisions, units, organizations…
- Different workloads profiles
- Different performance and security requirements
- Dynamic groups that require admin privileges

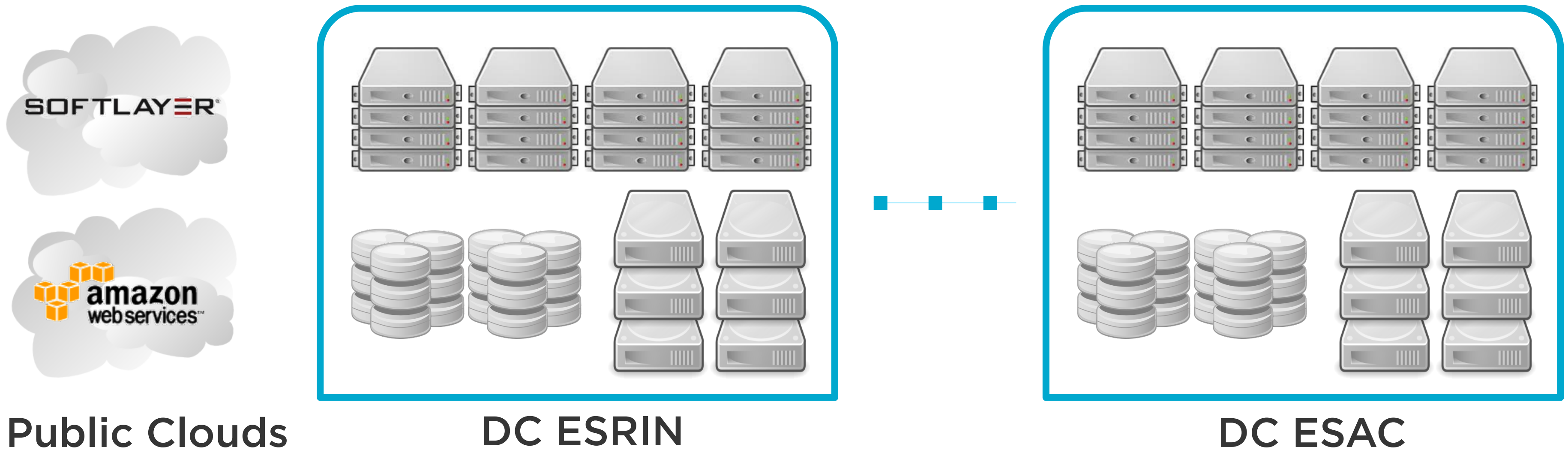
=> From many private clusters to a single consolidated environment

The Resource Provisioning Framework

Challenges from the Infrastructure Perspective

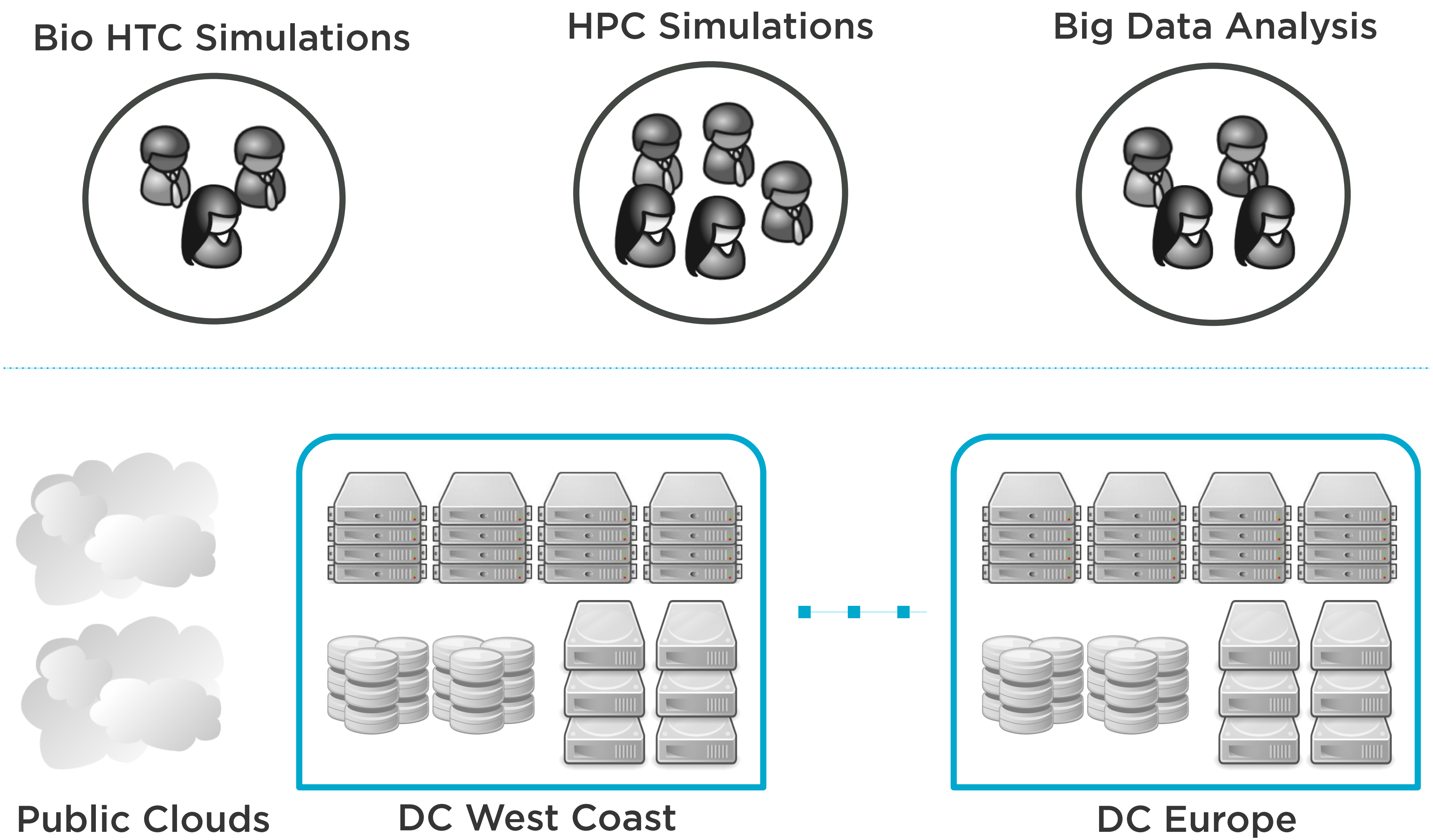
Comprehensive Framework to Manage Infrastructure Resources

- **Scalability**: Several DCs with multiple clusters
- **Outsourcing**: Access to several clouds for cloudbursting
- **Heterogeneity**: Different hardware for specific workload profiles

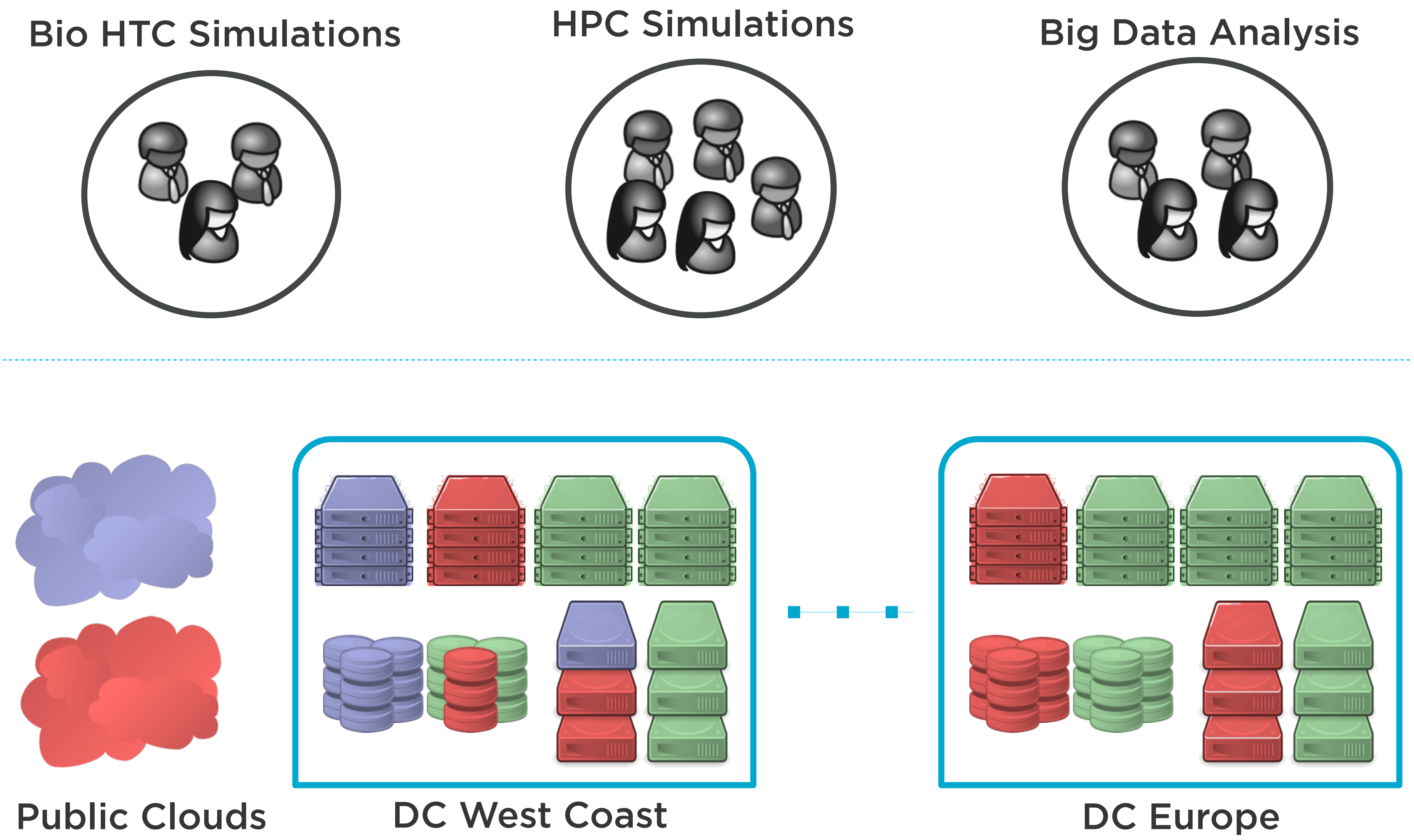


The Resource Provisioning Framework

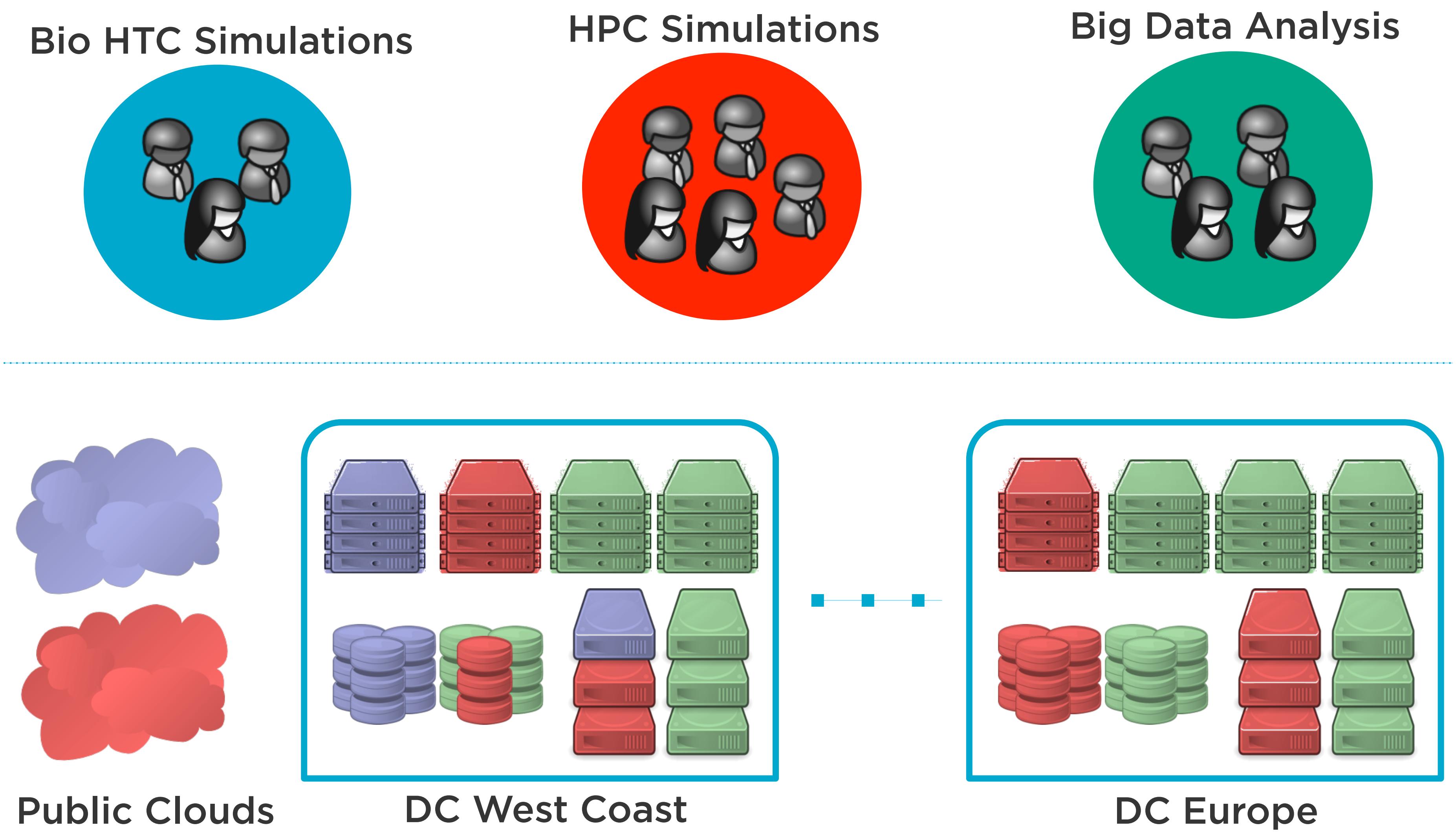
The Goal: Dynamic Allocation of Private and Public Resources to Groups of Users



Definition of Clusters (Resource Providers)

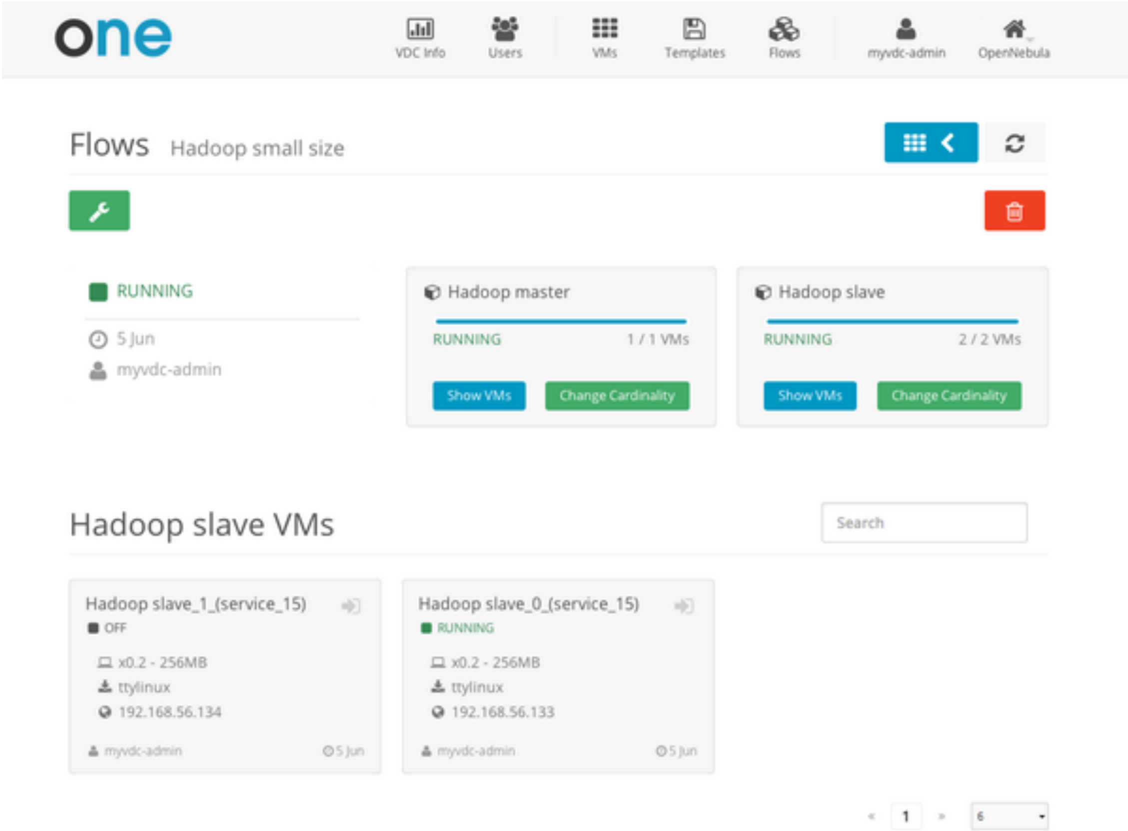


Definition of vDCs



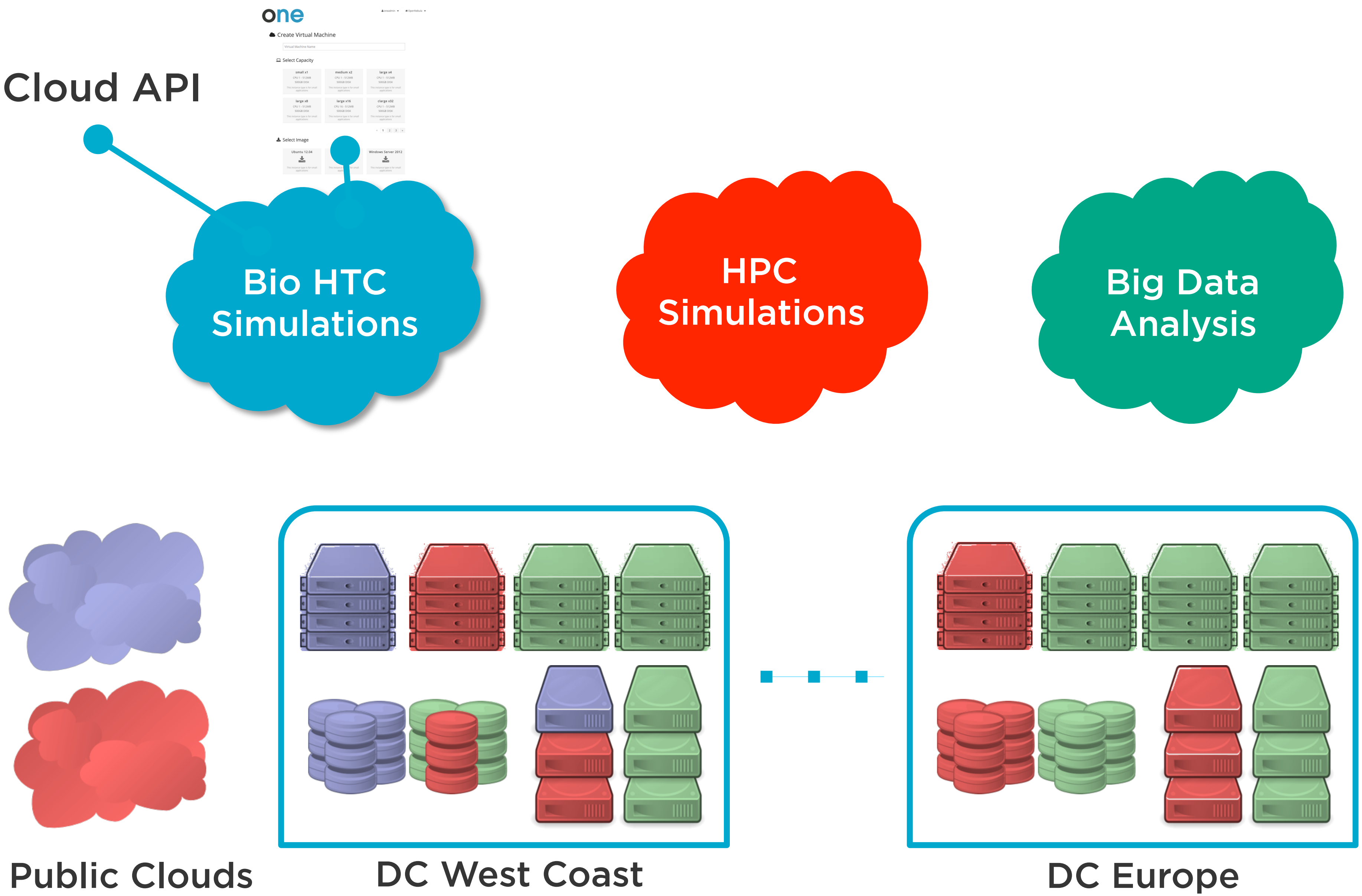
Admins in each Group/vDC Manage to its Own Virtual Private Cloud

- Each vDC has an **admin**
- **Delegation of management** in the VDC
- Only **virtual resources**, not the underlying physical infrastructure



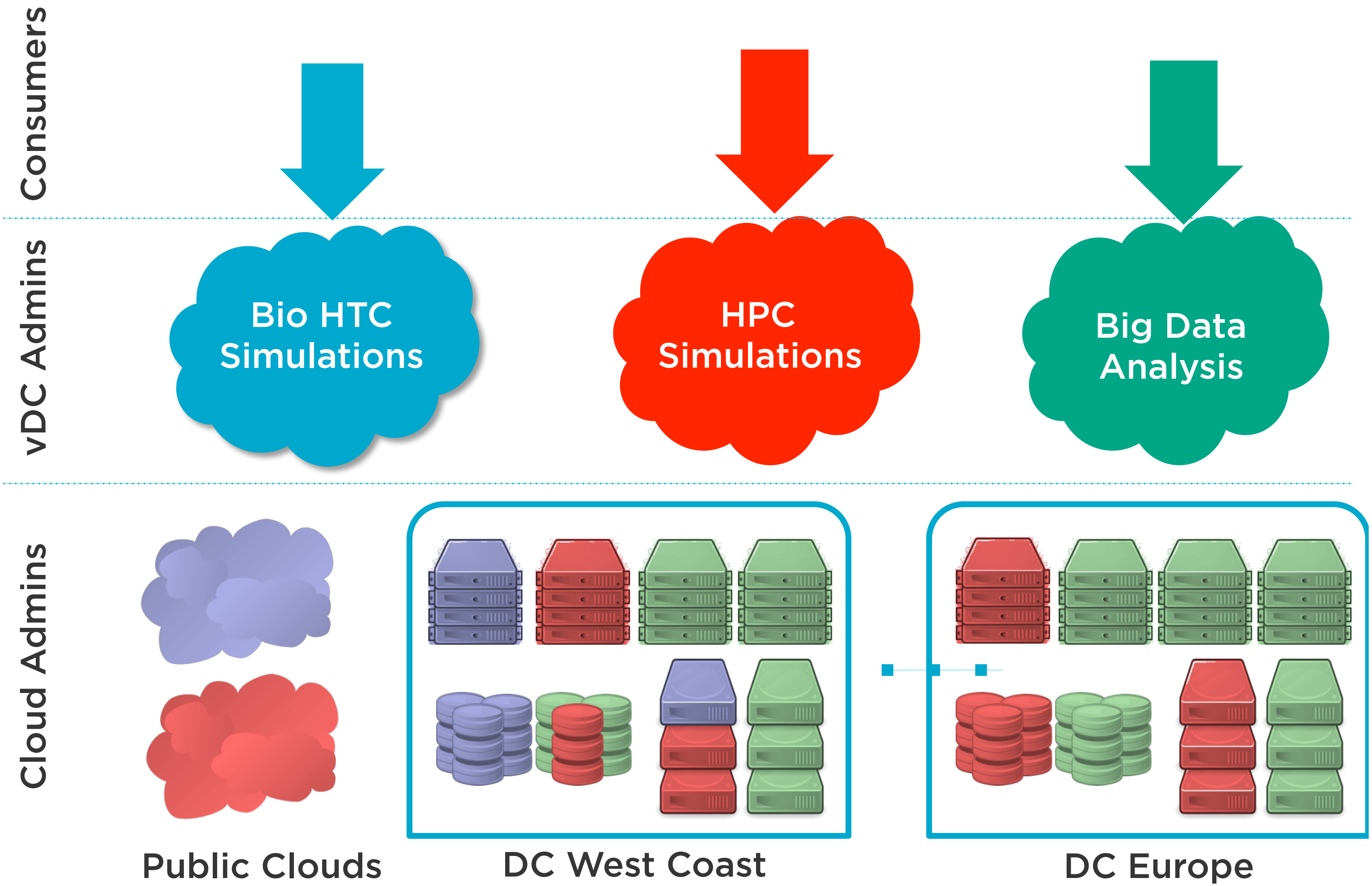
vDC Admin View

Users in each Group/vDC Access to its Own Virtual Private Cloud



The Resource Provisioning Framework

New Level of Provisioning: IaaS as a Service



Benefits

- **Partition of cloud resources**
- Complete **isolation** of users, organizations or workloads
- Allocation of Clusters with different levels of **security, performance** or **high availability** to different groups with different workload profiles
- **Containers** for the execution of virtual appliances (SDDCs)
- Way of **hiding physical resources** from Group members
- Simple **federation** and **scalability** of cloud infrastructures beyond a single cloud instance and data center

Private HPC Cloud Case Studies

One of Our Main User Communities

Supercomputing Centers

NCHC, SURF SARA, KTH, CESCA, PIC port d'informació científica, CESGA

Industry

Monte dei Paschi di Siena, Airbus, Produban, Santander

Research Centers

IPB, esa, CSIRO, MTA SZTAKI, KIT, KISTI, DESY, Fermilab, NIKHEF

Distributed Computing Infrastructures

EGI, PRAGMA, FutureGrid, BonFIRE, CHAIN-REDS, meGha

FermiCloud Fermilab http://www-fermicloud.fnal.gov/

Nodes	**KVM** on 29 nodes (2 TB RAM – 608 cores) Koi Computer
Network	Gigabit and **Infiniband**
Storage	CLVM+**GFS2** on shared 120TB NexSAN SataBeats
AuthN	**X509**
Linux	**Scientific Linux**
Interface	**Sunstone** Self-service and **EC2** API
App Profile	Legacy, HTC and **MPI HPC**

Typical Workloads

- Production VM-based batch system via the EC2 emulation => 1,000 VMs
- Scientific stakeholders get access to on-demand VMs
- Developers & integrators of new Grid applications

CESGA Cloud

http://cloud.cesga.es/

Nodes	**KVM** on 35 nodes (0.6 TB RAM – 280 cores) HP ProLiant
Network	2 x **Gigabit** (1G and 10G)
Storage	ssh from remote **EMC** storage server
AuthN	**X509** and core password
Linux	**Scientific Linux 6.4**
Interface	**Sunstone** Self-service and **OCCI**
App Profile	Individual VMs and **virtualised computing clusters**

Typical Workloads

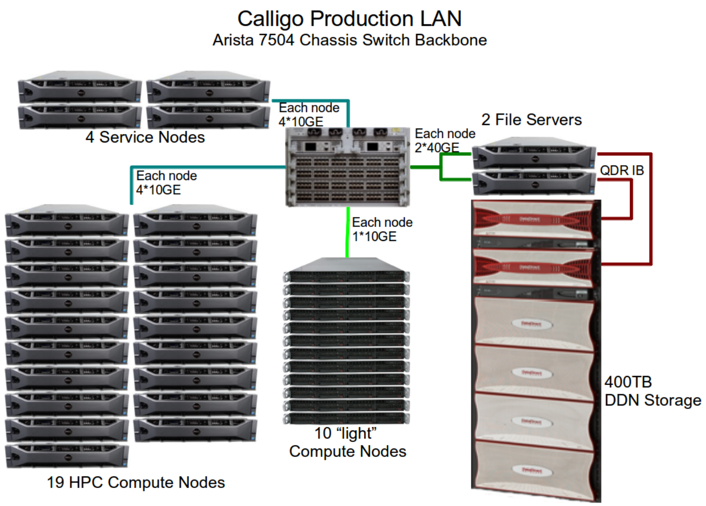
- 160 users
- Genomic, rendering…
- Grid services on production at CESGA
- Node at FedCloud project
- UMD middleware testing

SARA Cloud

SURF SARA

http://www.cloud.sara.nl

Nodes	**KVM** on 30 HPC nodes (256 GB RAM 1,300 cores + 2 TB High-memory node) Dell PowerEdge and 10 “light” nodes (64 GB RAM 80 cores) Supermicro
Network	2 x **Gigabit** (10G) with Arista switch
Storage	**NFS** on 500 TB NAS for HPC and ssh for “light”
AuthN	**Core password**
Linux	**CentOS**
Interface	**Sunstone** and **OCCI**
App Profile	MPI clusters, **windows clusters** and independent VMs



Typical Workloads

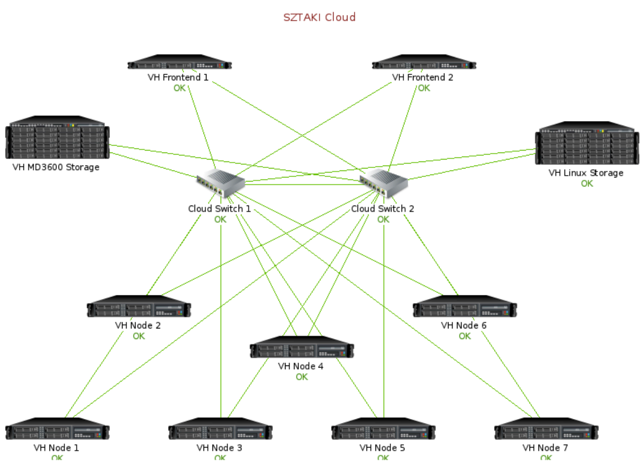
- Ad-hoc clusters with MPI and pilot jobs
- Windows clusters for Windows-bound software
- Single VMs, sometimes acting as web servers to disseminate results

SZTAKI Cloud

MTA SZTAKI

http://cloud.sztaki.hu/

Nodes	**KVM** on 8 nodes (2 TB RAM – 512 cores) DELL PowerEdge
Network	Redundant **10Gb**
Storage	Dell storage servers: **iSCSI** (36TB) and **CEPH** (288 TB)
AuthN	**X509**
Linux	**CentOS 6.5**
Interface	**Sunstone** Self-service, **EC2** and **OCCI**
App Profile	Individual VMs and **virtualised computing cluster**



Typical Workloads

- Run standard and grid services (e.g.: web servers, grid middleware...)
- Development and testing of new codes
- Research on performance and opportunistic computing

KTH Cloud

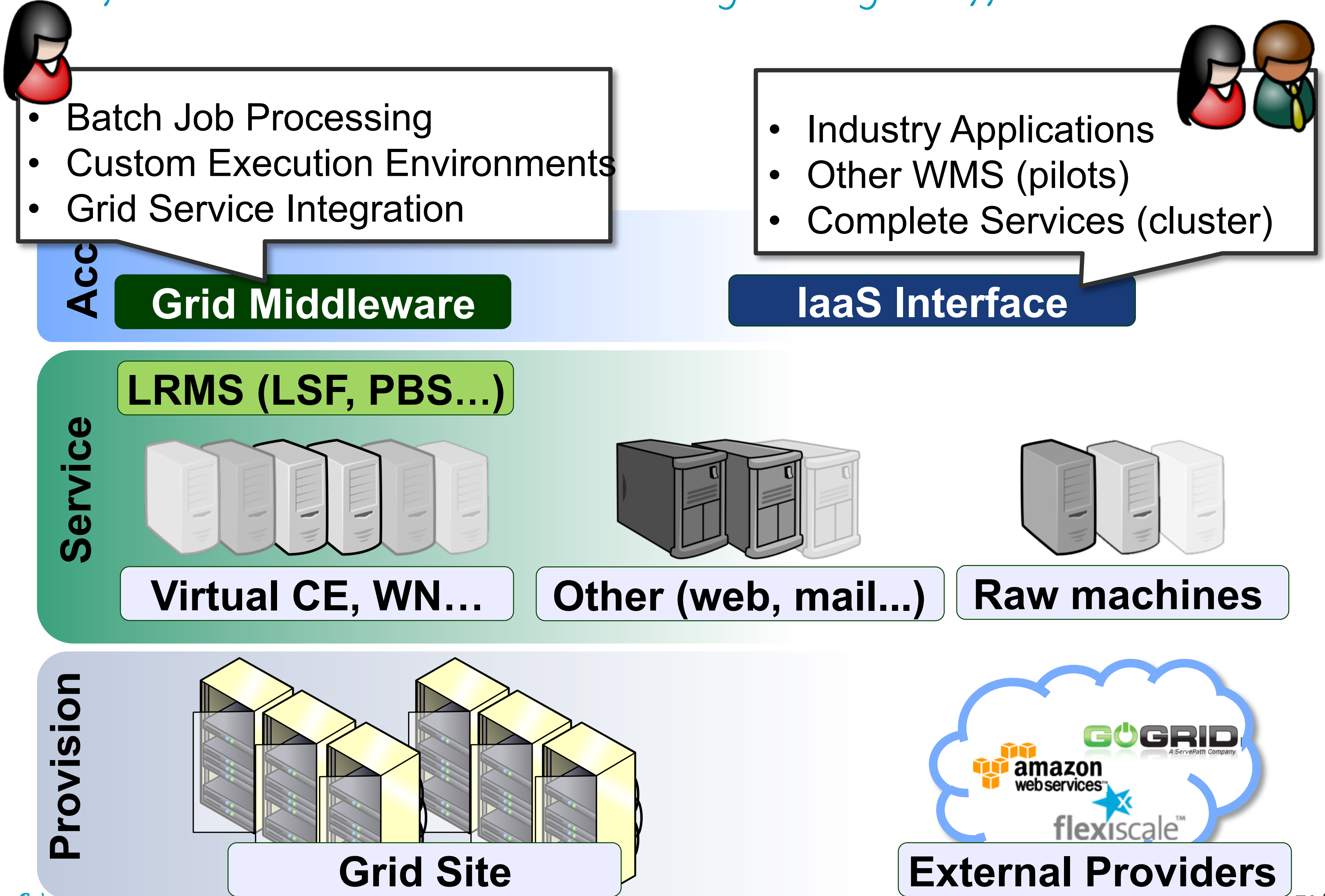
http://www.pdc.kth.se/

Nodes	**KVM** on 768 cores (768 GB RAM) HP ProLiant
Network	**Infiniband** and Gigabit
Storage	NFS and **LVM**
AuthN	**X509** and core password
Linux	**Ubuntu**
Interface	**Sunstone** self-service, **OCCI** and **EC2**
App Profile	Individual VMs and **virtualised computing cluster**

Typical Workloads

- Mainly BIO
- Hadoop, Spark, Galaxy, Cloud Bio Linux...

The OpenNebula Vision for Grid Sites: Extending the Range of Applications



The OpenNebula Vision for Grid Infrastructures

Grid Services

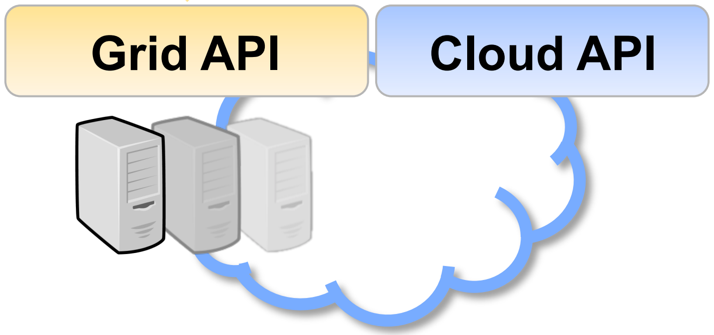
- Federation facilities
- Security
- Grid specific services

MarketPlace

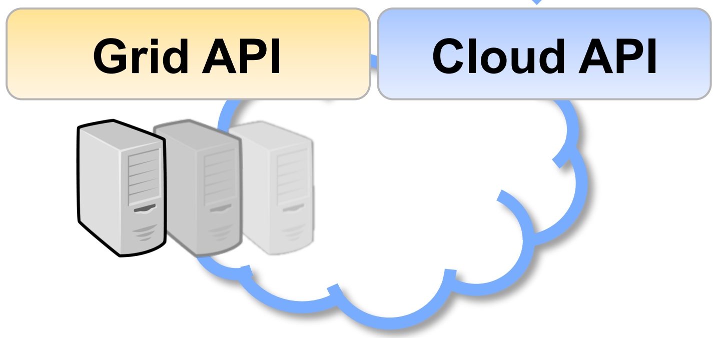
- Sharing existing VM images
- Registry of metadata
- Image are kept elsewhere
- Supports trust

Appliance Repo

- Storage VM images
- Distributed
- Multi-protocol



Cloud/Grid Site



Cloud/Grid Site

Grid and Cloud as Complementary Computing Models

Grids

Usage
- Job Processing
- Big Batch System
- File Sharing Services

Achievements
- Federation of Resources
- VO Concept

But…
- User experience
- Complexity

Clouds

Usage
- Raw infrastructure
- Elasticity & Pay-per-use
- Simple Web Interface

Achievements
- Agile Infrastructures
- IT is another Utility

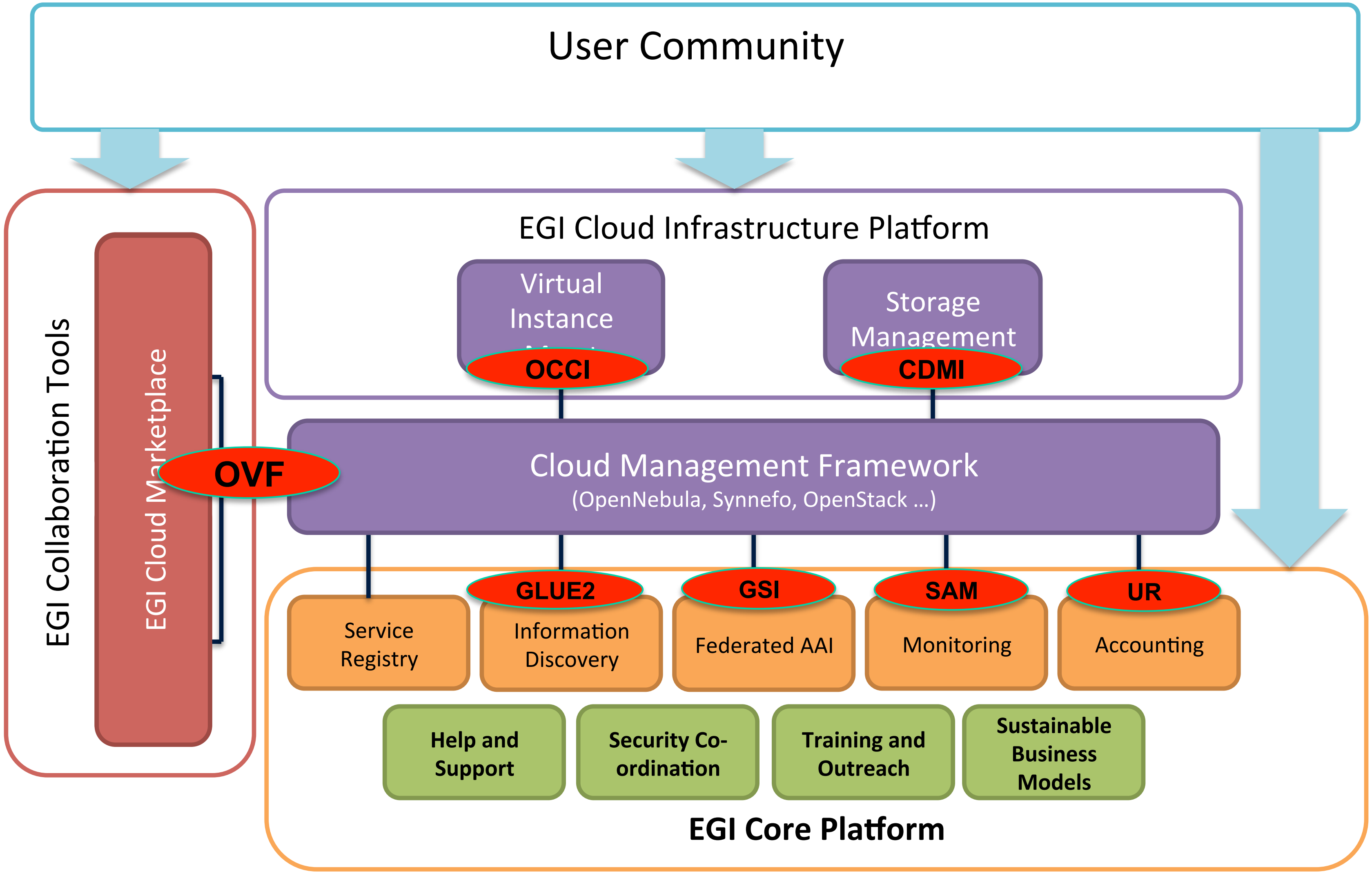
But…
- Interoperability
- Federation

Resource Sharing
Scientific Applications
Uniform Security

\+

Resource Management
Customize Environments
Flexibility & Simplicity

EGI Federated Cloud - The Architecture



EGI Federated Cloud: Use Cases and Architecture, David Wallom, July 2014

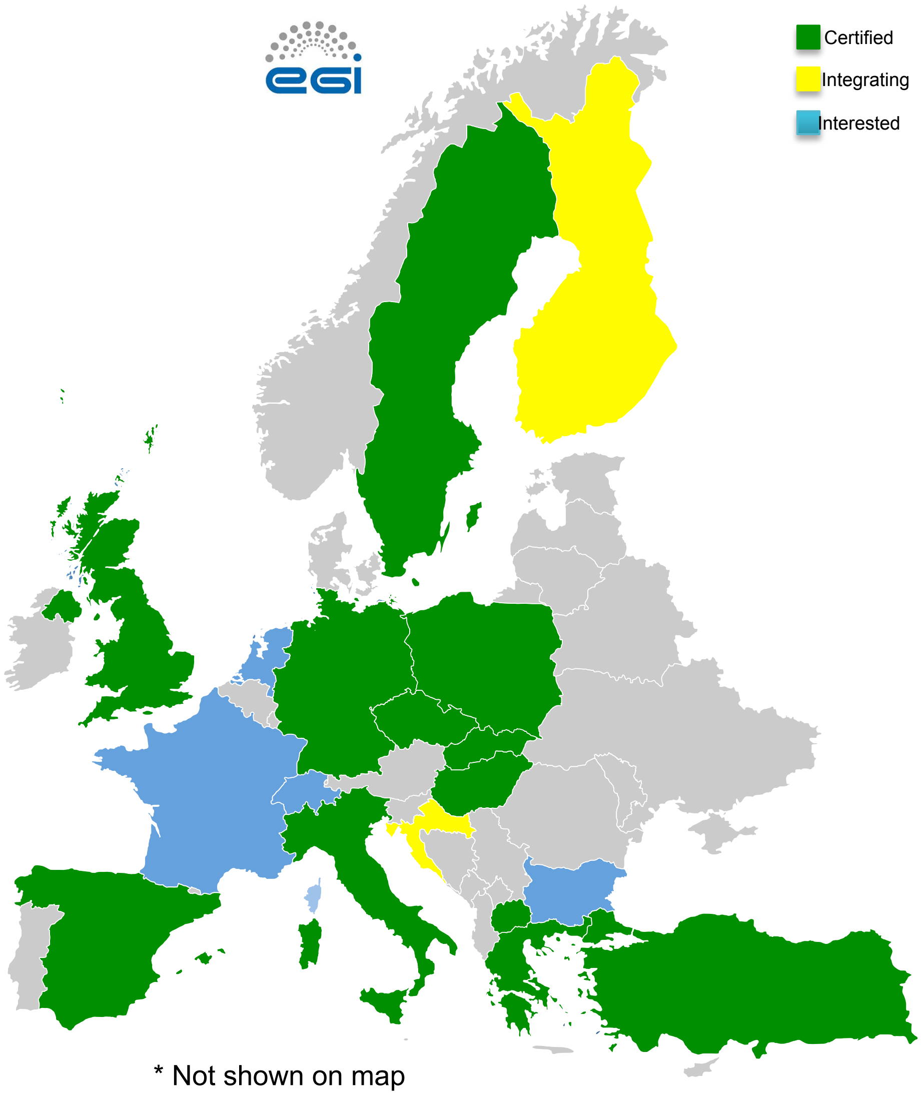
EGI Federated Cloud – The Providers

15 certified resource providers from 12 countries from the public and private sector
- Czech Republic, Germany, Greece, Hungary, Italy, Macedonia, Poland, Slovakia, Spain, Sweden, Turkey, United Kingdom

2 countries currently integrating
- Croatia, Finland

6 countries interested
- Bulgaria, France, Israel*, The Netherlands, Portugal, Switzerland

Worldwide partnership/interest
- Australia* (NECTAR)
- South Africa* (SAGrid)
- South Korea* (KISTI)
- United States* (NIST, NSF Centres)



* Not shown on map

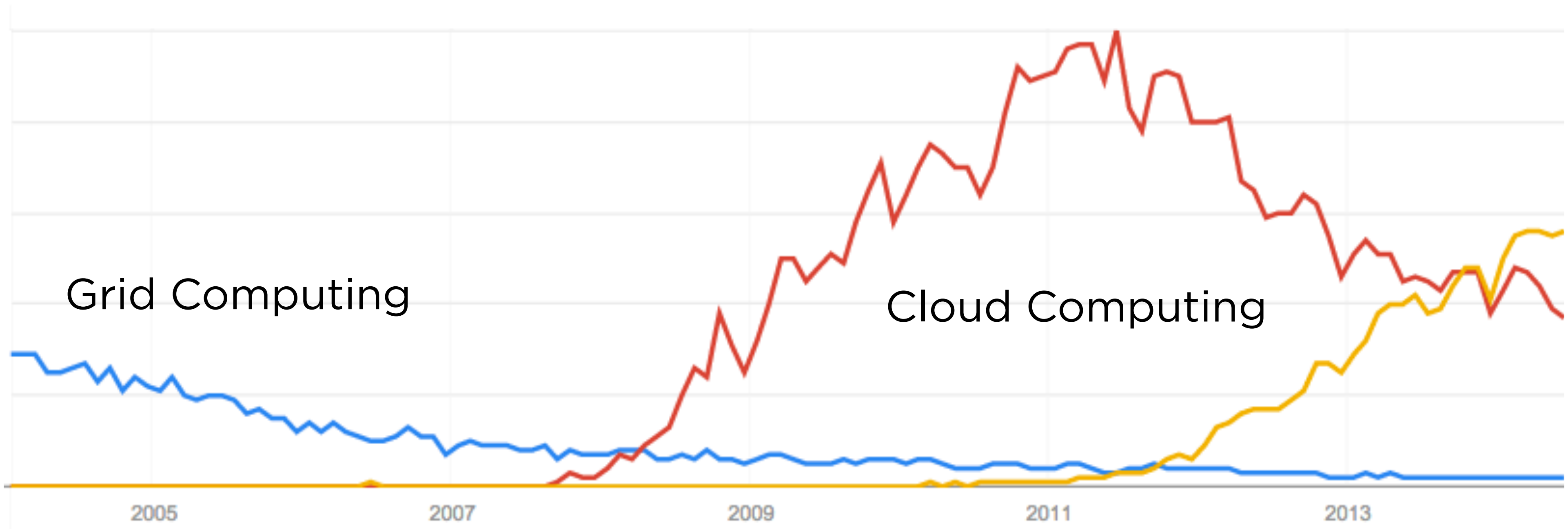
Launch capability	–	5,000 cores,	225 TB storage
Q4 2014 (planned)	–	18,000 cores,	6000 TB storage
2020 Vision	–	1,000,000 cores,	1 EB storage

EGI Federated Cloud – The Users

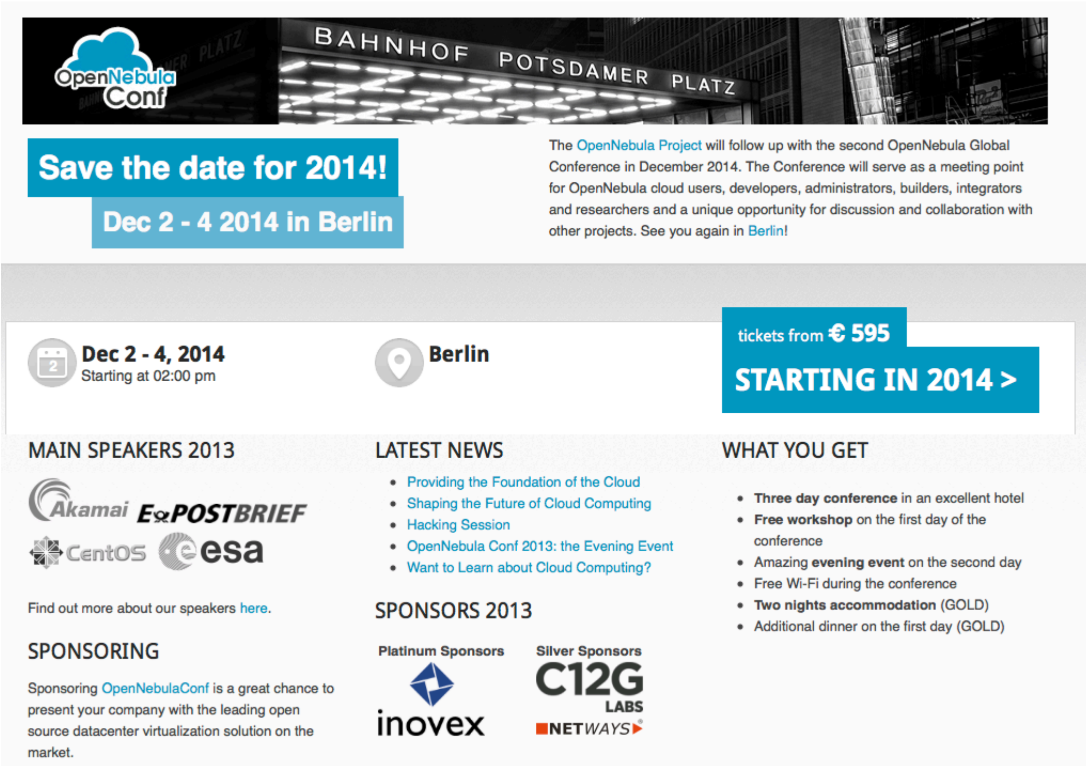
- **Ecology** – BioVeL: Biodiversity Virtual e-Laboratory
- **Structural biology** – WeNMR: a worldwide e-Infrastructure for NMR and structural biology
- **Linguistics** – CLARIN: 'British National Corpus' service (BNCWeb)
- **Earth Observation** – SSEP: European Space Agency's Supersites Exploitation Platform for volcano and earthquakes monitoring (Collaboration with Helix Nebula)
- **Software Engineering** – SCI-BUS: simulated environments for portal testing
- **Software Engineering** – DIRAC: deploying ready-to-use distributed computing systems
- **Interdisciplinary research**– Catania Science Gateway Framework
- **Musicology** – Peachnote: dynamic analysis of musical scores
- **Earth Observation** – ENVRI: Common Operations of Environmental Research infrastructures (collaboration with EISCAT3D)
- **Geology** – VERCE: Virtual Earthquake and seismology Research
- **Ecology** – LifeWatch: E-Science European Infrastructure for Biodiversity and Ecosystem Research
- **High Energy Physics** – CERN ATLAS: ATLAS processing cluster via HelixNebula

EGI Federated Cloud: Use Cases and Architecture, David Wallom, July 2014

Different Names for the Same Model? Same Challenges but Different Technologies?



Upcoming Community Events



Some Research References

Innovation in Cloud Architecture

- B. Sotomayor, R. S. Montero, I. M. Llorente and I. Foster, "Virtual Infrastructure Management in Private and Hybrid Clouds", **IEEE Internet Computing**, September/October 2009 (vol. 13 no. 5)
- Rafael Moreno-Vozmediano, Ruben S. Montero, Ignacio M. Llorente, "Multi-Cloud Deployment of Computing Clusters for Loosely-Coupled MTC Applications", **IEEE Transactions on Parallel and Distributed Systems**, 22(6): 924-930, April 2011
- Rafael Moreno-Vozmediano, Ruben S. Montero, Ignacio M. Llorente, "IaaS Cloud Architecture: From Virtualized Data Centers to Federated Cloud Infrastructures", **IEEE Computer**, 45(12):65-72, December 2012
- Rafael Moreno-Vozmediano, Ruben S. Montero, Ignacio M. Llorente, "Key Challenges in Cloud Computing to Enable the Future Internet of Services", **IEEE Internet Computing**, 17(4):18-25, 2012.

Questions?

We Will Be Happy to Answer Your Questions

OpenNebula.org   @OpenNebula