Cloud Computing

(and virtualization at CERN)

Ulrich Schwickerath et al

With special thanks to the many contributors to this presentation!

GridKa School 2011, Karlsruhe

Disclaimer: largely personal view of things

CERN IT Departmen
CH-1211 Genève 2
Switzerlan
www.cern.ch/i

Outline

CERN IT Department

What is this all about ?

- Definition, Key features and manifestations
- Is it useful ? For what ? For whom ?
- Is it relevant for HEP community

An IaaS cook book

- Ingredients and recipe
- Caveats and work to do

The practice

- Results from a prototype setup at CERN

Conclusions

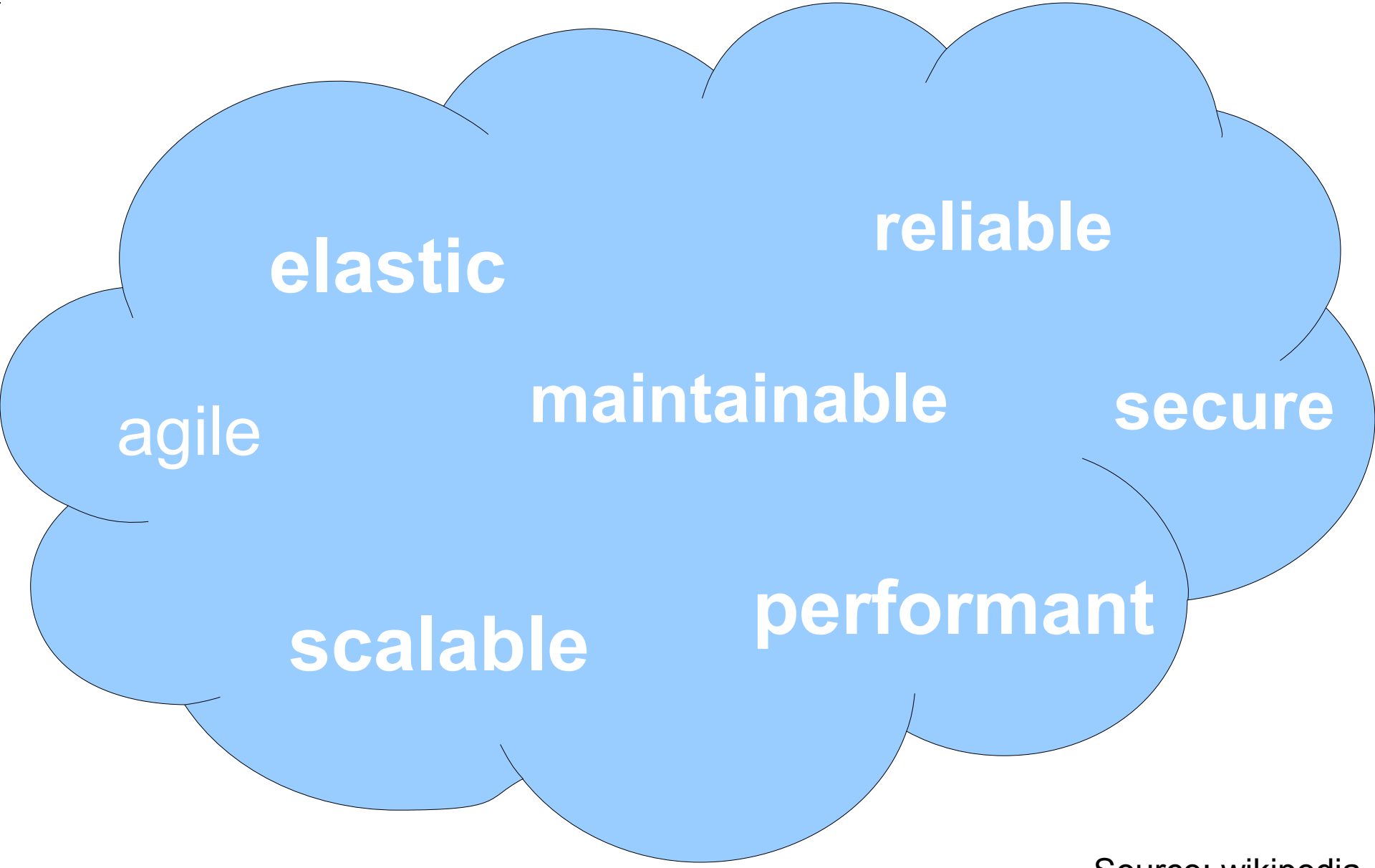
What is Cloud Computing ?

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet).

Source: Wikipedia

Key features

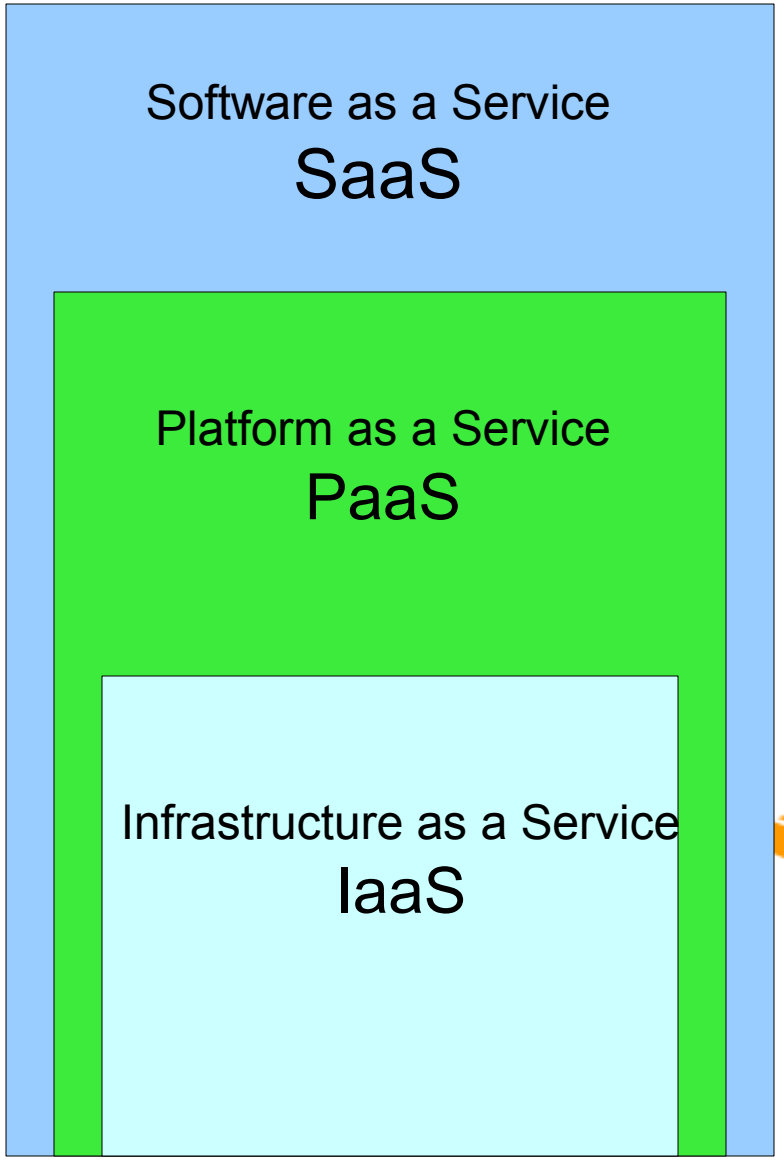
CERN IT Department



Source: wikipedia

Manifestation: XaaS

CERN IT Department



On demand access to applications

skype Gmail by Google Google Docs

Service-now.com
On Demand IT Service Management

Platform for building & delivering Web applications

force.com platform as a service

Utility computing

amazon web services

ElasticHosts
Flexible servers in the cloud

flexiscale

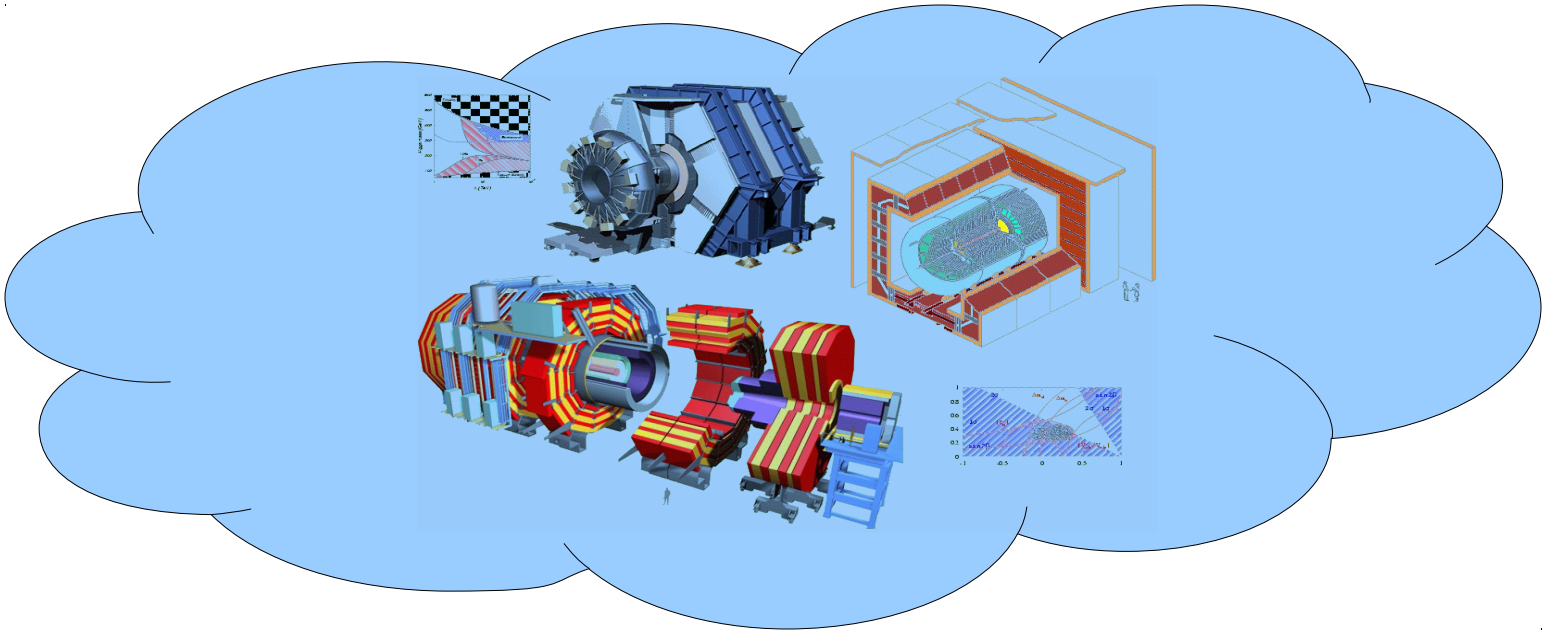
GOGRID

greenqloud

Thanks to Tony Cass

The key question is ...

What is the relevance for HEP community ?



Disclaimer: personal and site biased perspective, with a focus on IaaS

What are the challenges?

Resource provider challenges

- **Increasing** demand for computing and storage resources
- More (new) **users** to support
- More **communities** to support (VOs), with different requirements
- **Conflicting** software **requirements** for applications and new hardware

What are the challenges?

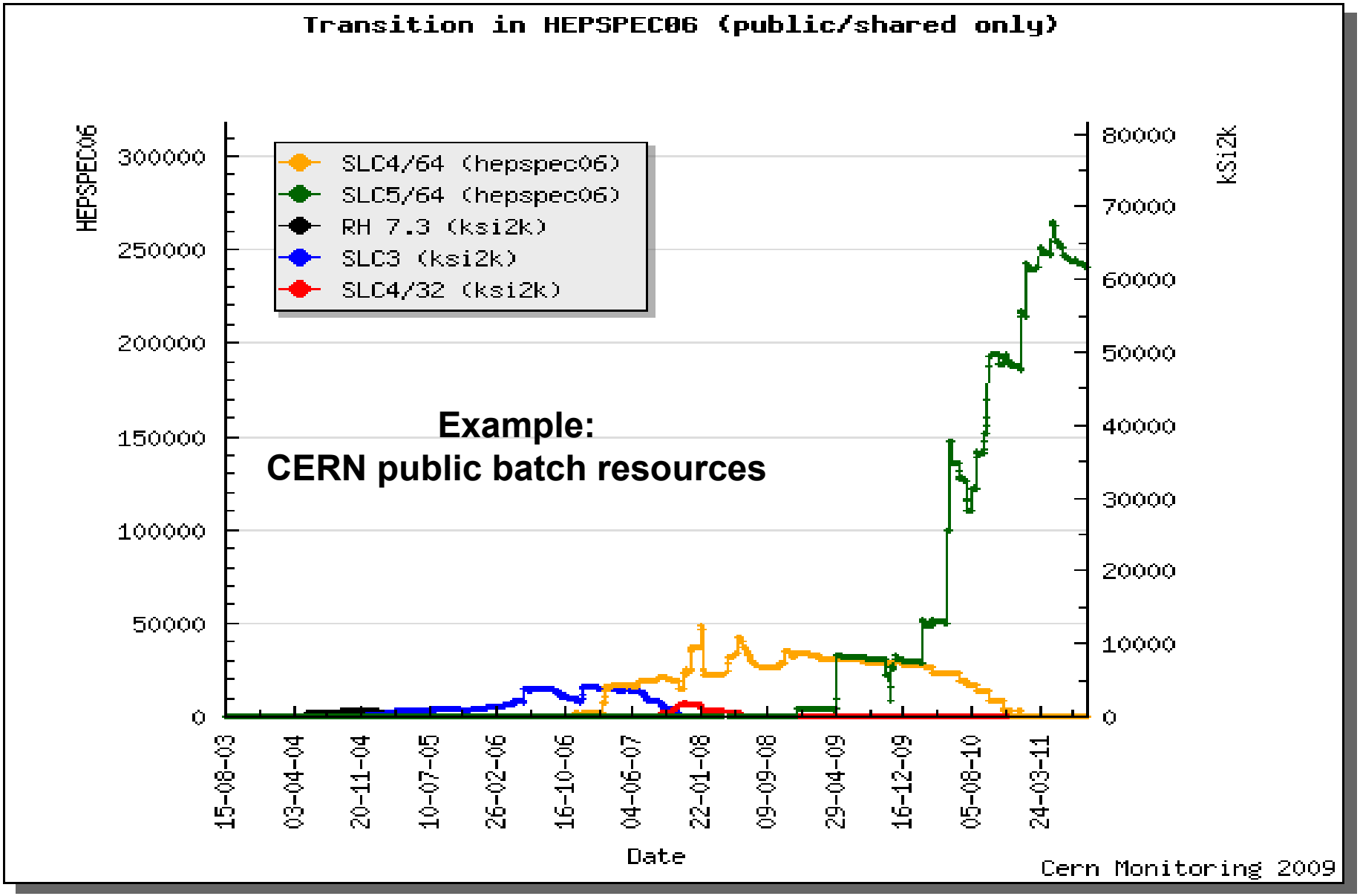
Resource provider boundary conditions

- Decreasing resources for developments and maintenance
- Constant or decreasing number of people to provide computing services

Necessity to optimize the use of existing resources

What are the challenges ?

CERN IT Department



Example:
CERN public batch resources

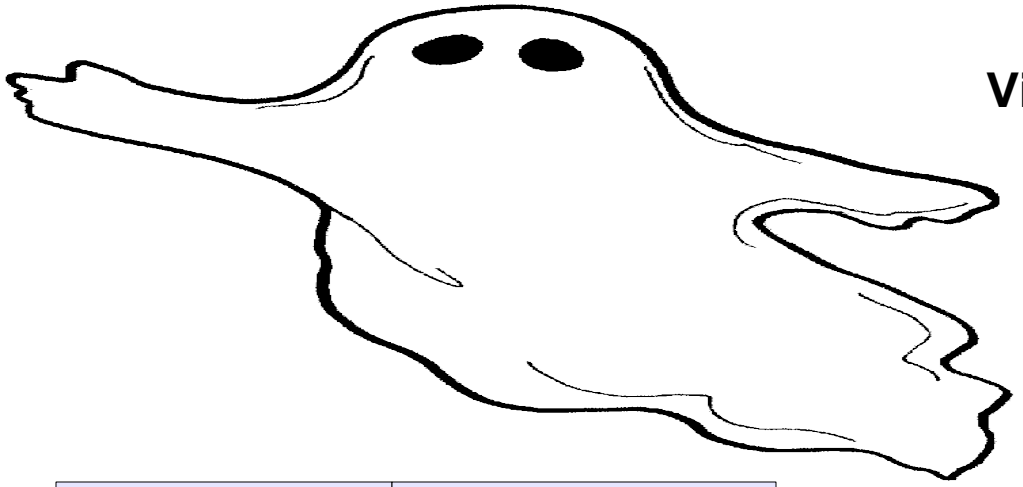
What are the challenges ?

- VOs often **know their users** better than the sites do
- VOs know best where their **active data** is
- VOs often unhappy with sites scheduling decisions

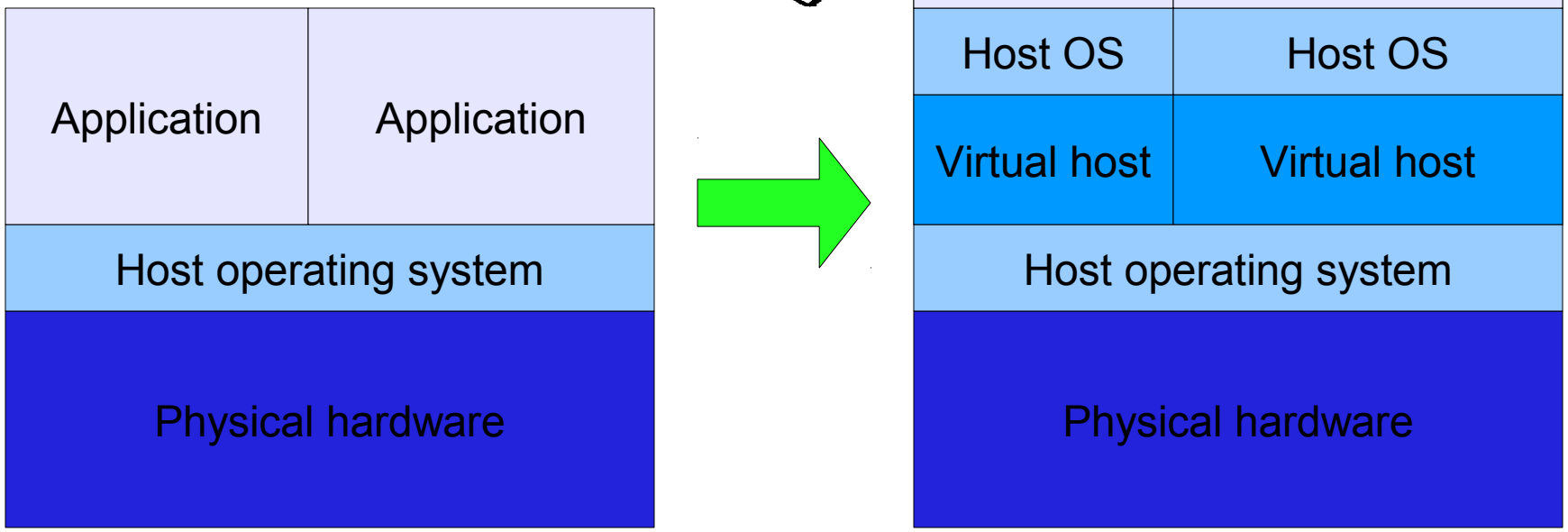
Example: Pilot job frameworks:

- Essentially a work-around for VOs to do the job scheduling themselves
- Causes some overhead for the sites who have to maintain new services
  - glExec, SCAS/Argus
  - UID switching within a single user job

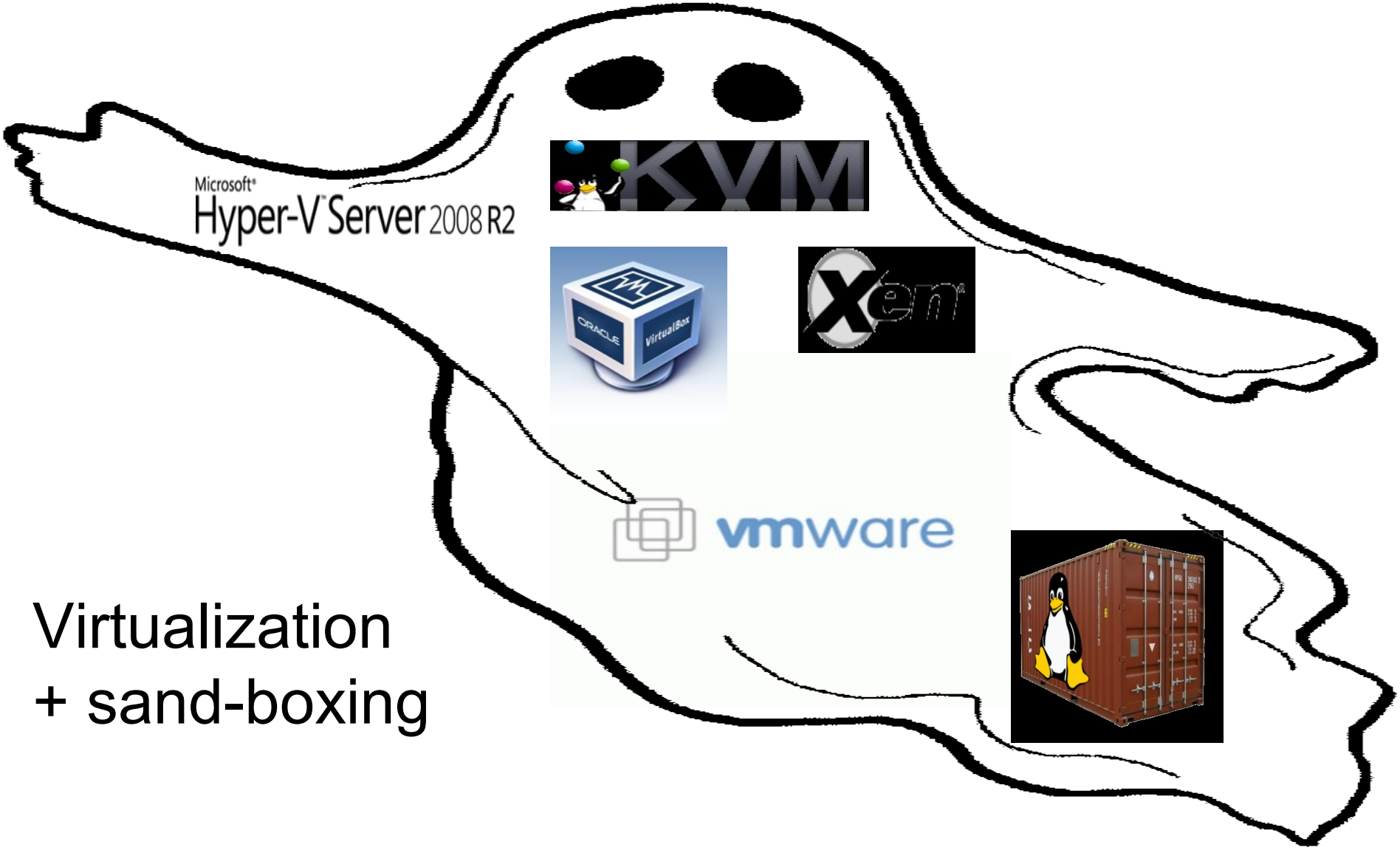
… and the new technologies



Virtualization as part of the solution



… and the new technologies



Virtualization
+ sand-boxing

Virtual machine management systems

Central place to manage your VMs in the computer center
Some complemented by services required for Clouds

openstack™

EUCALYPTUS

OpenNebula.org

vmware

Platform Computing

ISF

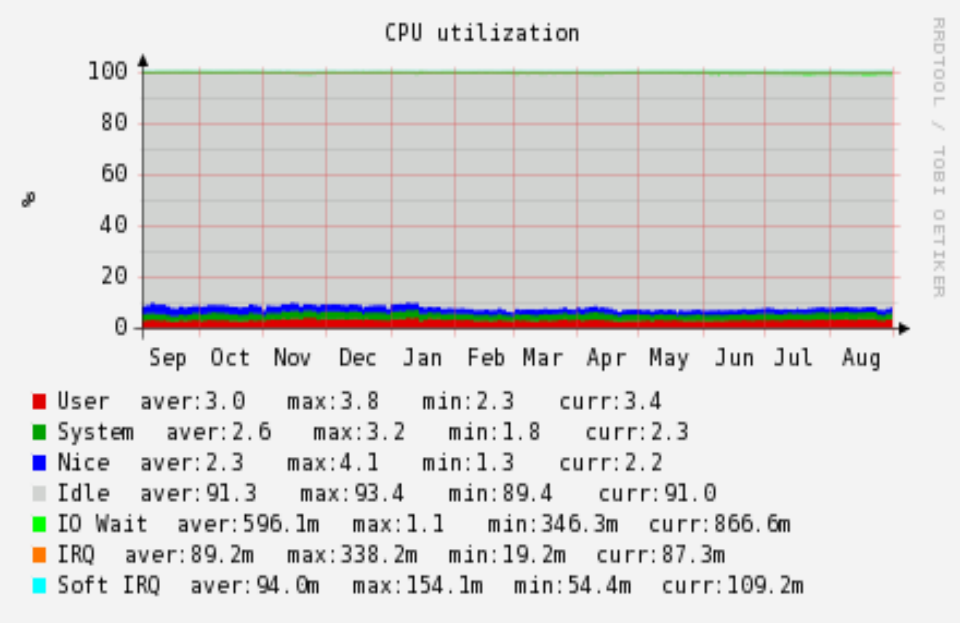
Microsoft System Center Virtual Machine Manager

How can virtualization help ?

CERN IT Department

Resource usage optimization

- Many dedicated machines
- Mainly managed by VOs
- Specific applications
- Low CPU and I/O usage
- Large reliability required



Current CERN solution for consolidation using CERN Virtual Infrastructure (CVI)

- Specialized fully redundant hardware with shared dedicated storage via iSCSI
- Support for live migration for services
- Selected solution: Microsoft Hyper-V + SCVMM as orchestrator
- Completed by Self-Service on cheap hardware
- Different configurations: enclosures+shared storage, disk servers, worker nodes

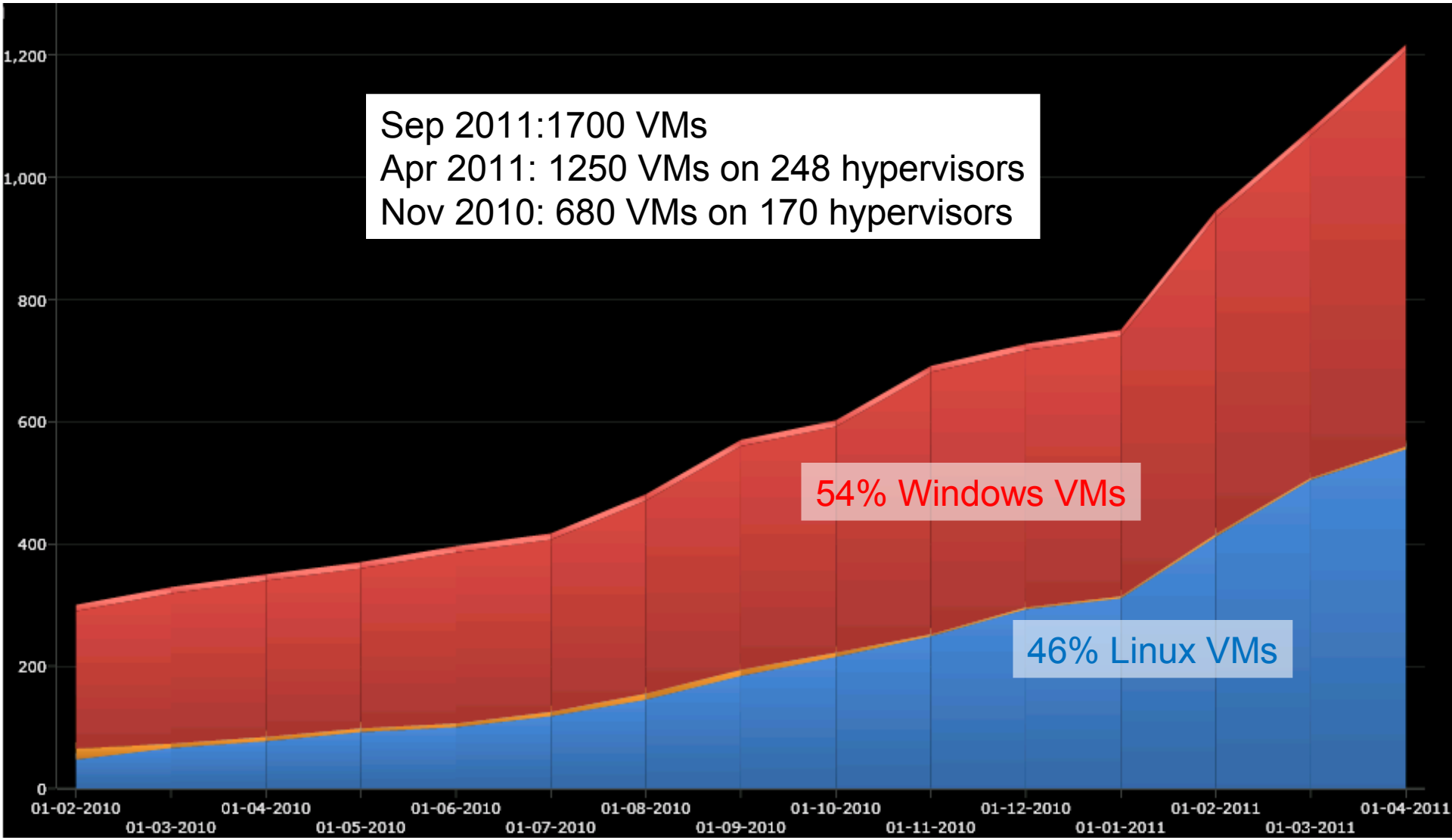
How can virtualization help ?

Virtualization for service consolidation is reality, wide spread and routine

Example: CERN CVI

CVI status (April 2011)

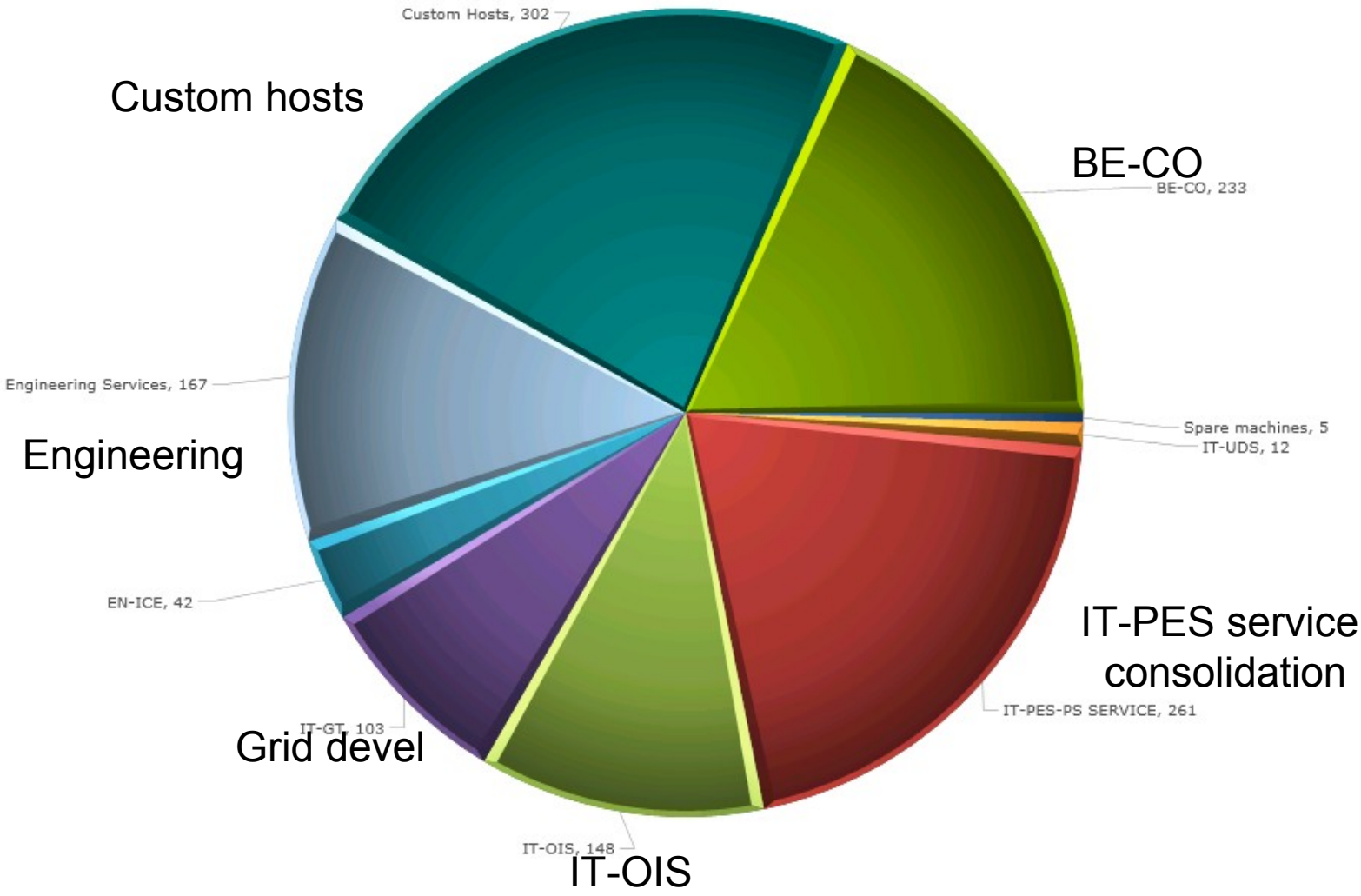
CERN IT Department



Number of Virtual Machines per Operating System

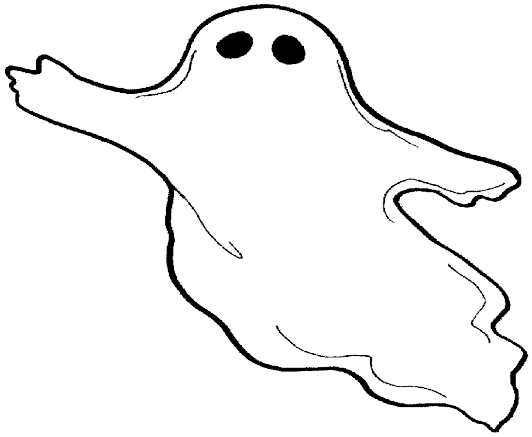
Thanks to Jan van Eldik et al

April 2011: CVI 1250 VMs, 8 Customer groups



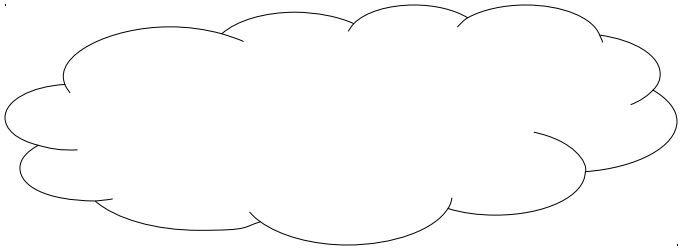
Thanks to Jan van Eldik et al

How can virtualization help?

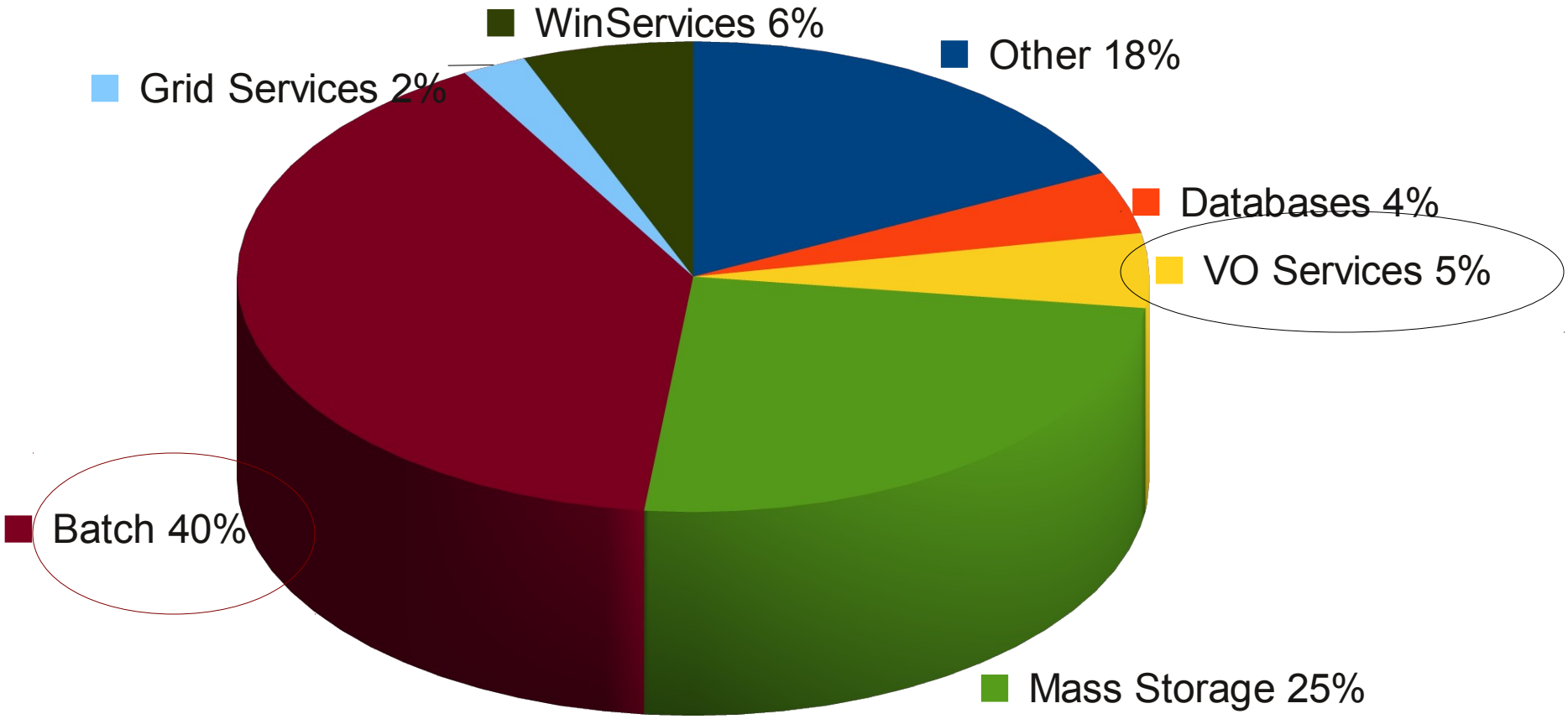


Note:
virtualization is NOT cloud computing

≠

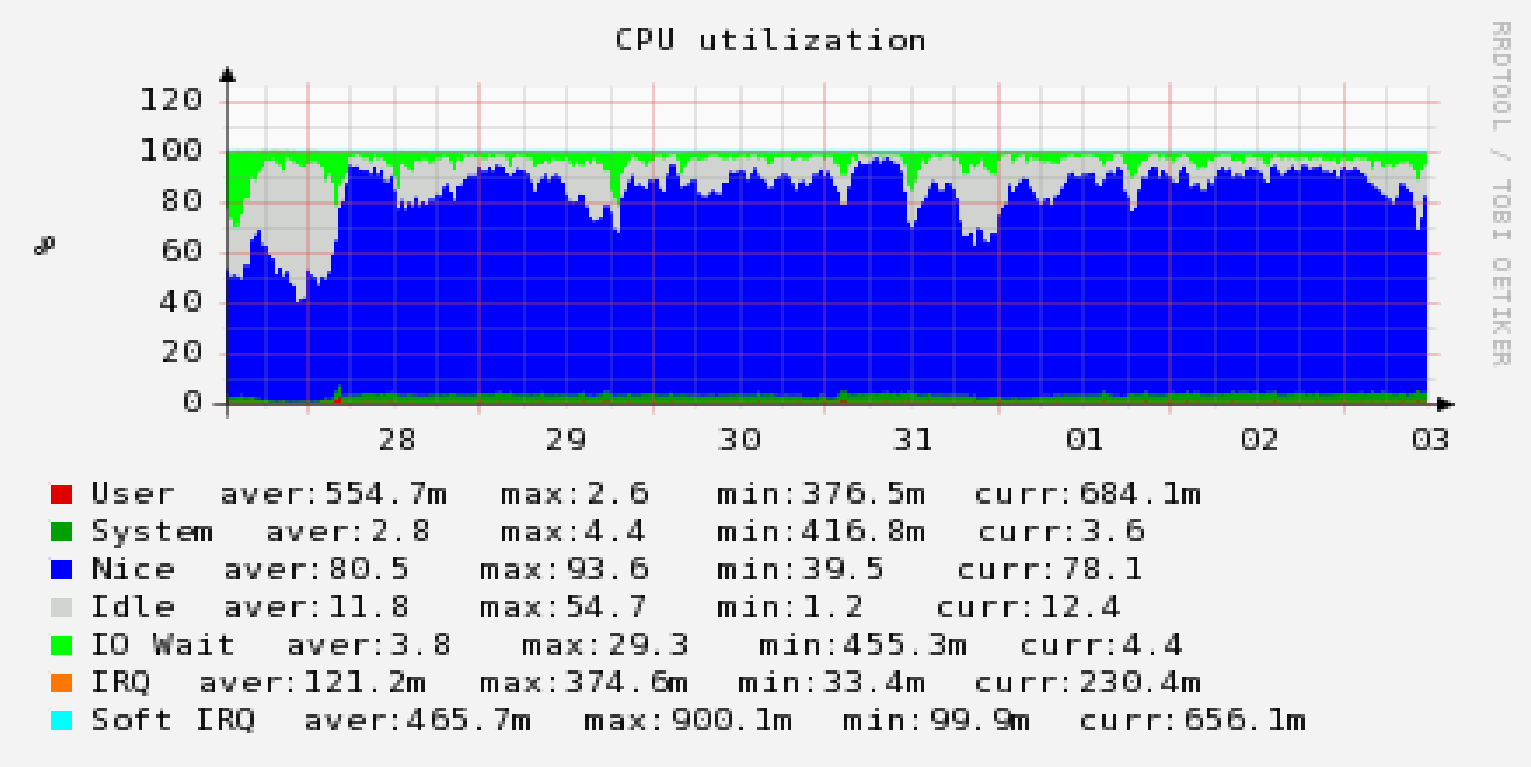


Can we gain elsewhere ?



Example: CERN Computer Center in September 2011

Batch processing resources



- Managed by the site
- Grid Worker node setup
- Lots of identical machines

- CPU, I/O and network demanding
- Cheap hardware
- individual failures OK

Challenging use case due to scale and performance requirements!

Batch virtualization: principles

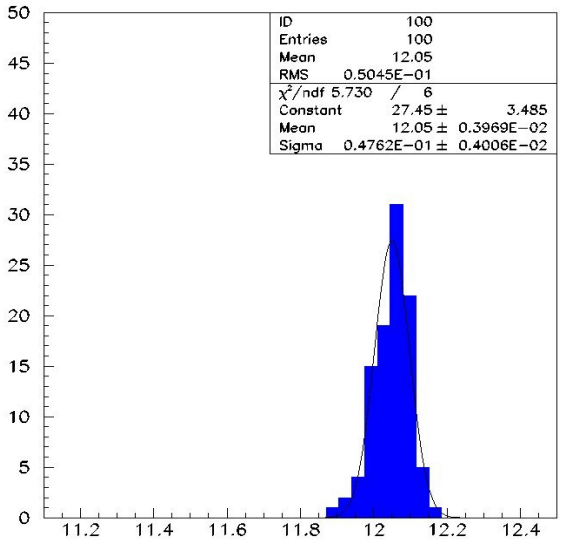
What is(are) the problem(s) to solve ?

- Optimize
  - Job throughput
  - Job success rate
- Minimize
  - Operational overhead
  - Downtime for updates
- Require
  - Performance and speed

The price to pay for virtualization

CERN IT Department

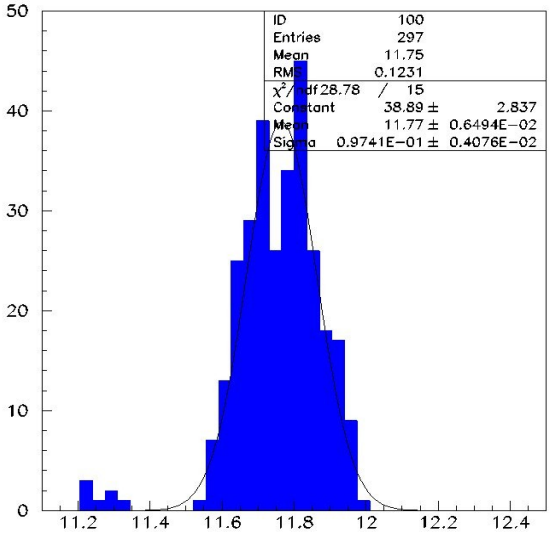
HS06 CPU benchmark test



Bare metal:

- SLC5
- 2x L5520 Intel Xeon
- 2.27GHz

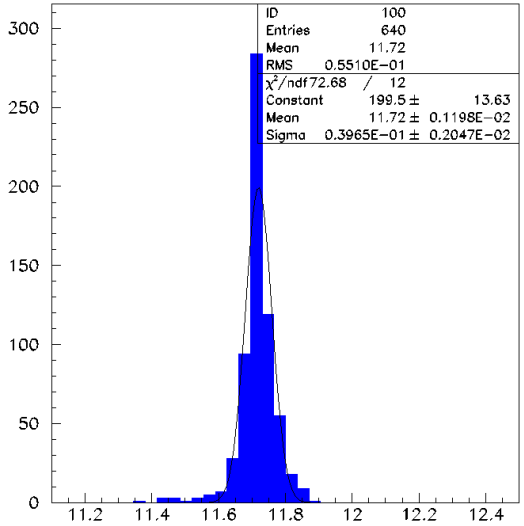
HS06=12.05/core



KVM

- HW as before
- SLC5/6 hypervisor
- 8 SLC5 guests
- No KSM,ept off (SLC5)
- Pinned VMs

HS06=11.4/core (SLC5)
HS06=11.8/core (SLC6)



HyperV

- HW as before
- 8 SLC5 guests

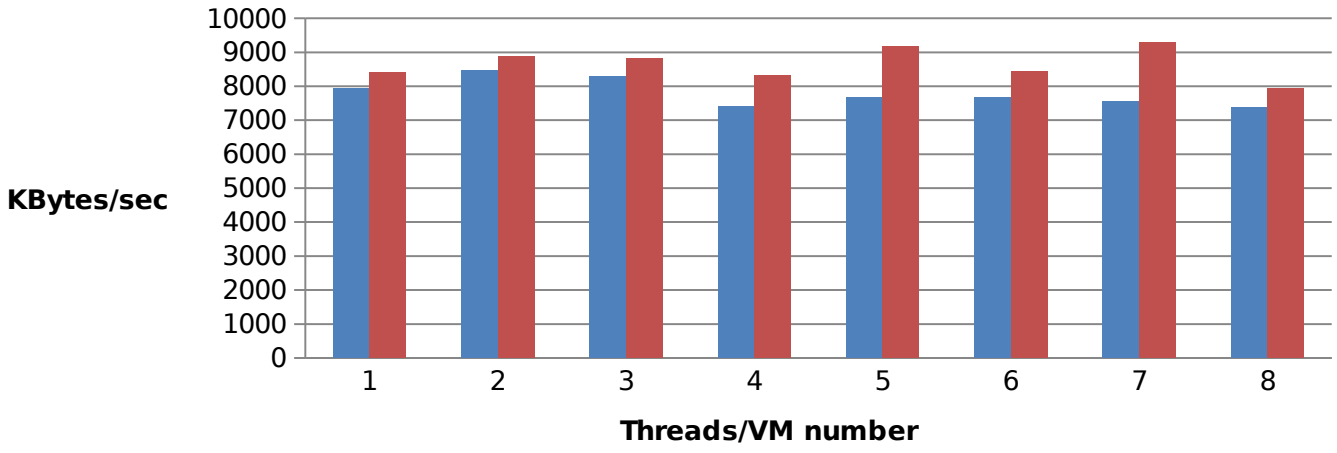
HS06=11.7/core

I/O Benchmarking

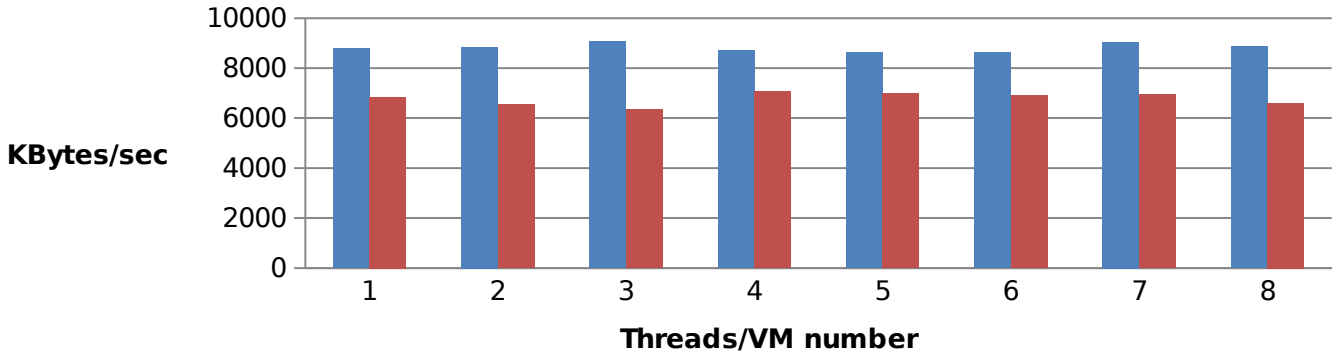
CERN IT Department

Worst case scenario: 1VM/physical core running IOzone, 8 threads on bare metal

preliminary (random) **Read**



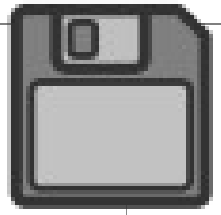
preliminary (random) **Write**



Analysis by Qiulan Huang (Chinese academy of science), December 2010

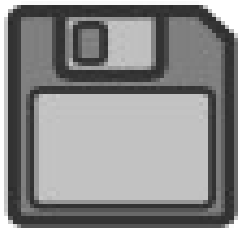
Notes:

- Caching off
- SLC5 Hypervisor
- **20% penalty**
- write worse
- block device disk on LV, exported to VM



I/O Benchmarking

- **Analysis by Qiulan Huang** (Chinese academy of science), December 2010
  - Caching off, SLC5 based KVM hypervisors
  - LV raw device, imported into the VM
  - 1-8 VM per hypervisor with one IOZone benchmark each
  - Bare metal test with 1-8 concurrent IOZone threads
  - 20% penalty, write performance penatly is worse
- **New analysis ongoing by Belmiro Moreira (CERN)**
  - Trying qcow2 and compare to LV
  - Based on SLC6.1 KVM
  - No final results yet

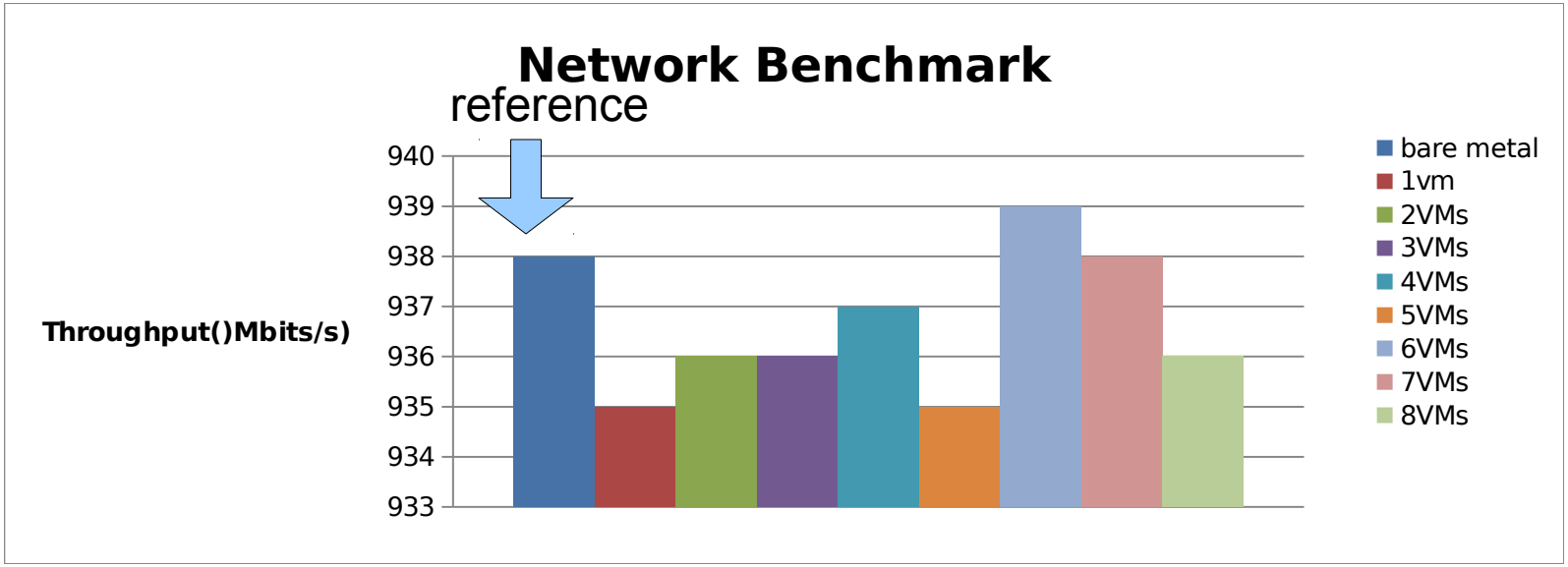


Network Benchmarking

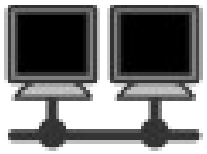
CERN IT Department

Analysis by
Qiulan Huang
(Chinese academy of science),
December 2010, CERN

Penalty <=1%



Iperf with TCP window size of 256k and 60s test time



Virtualizing batch resources

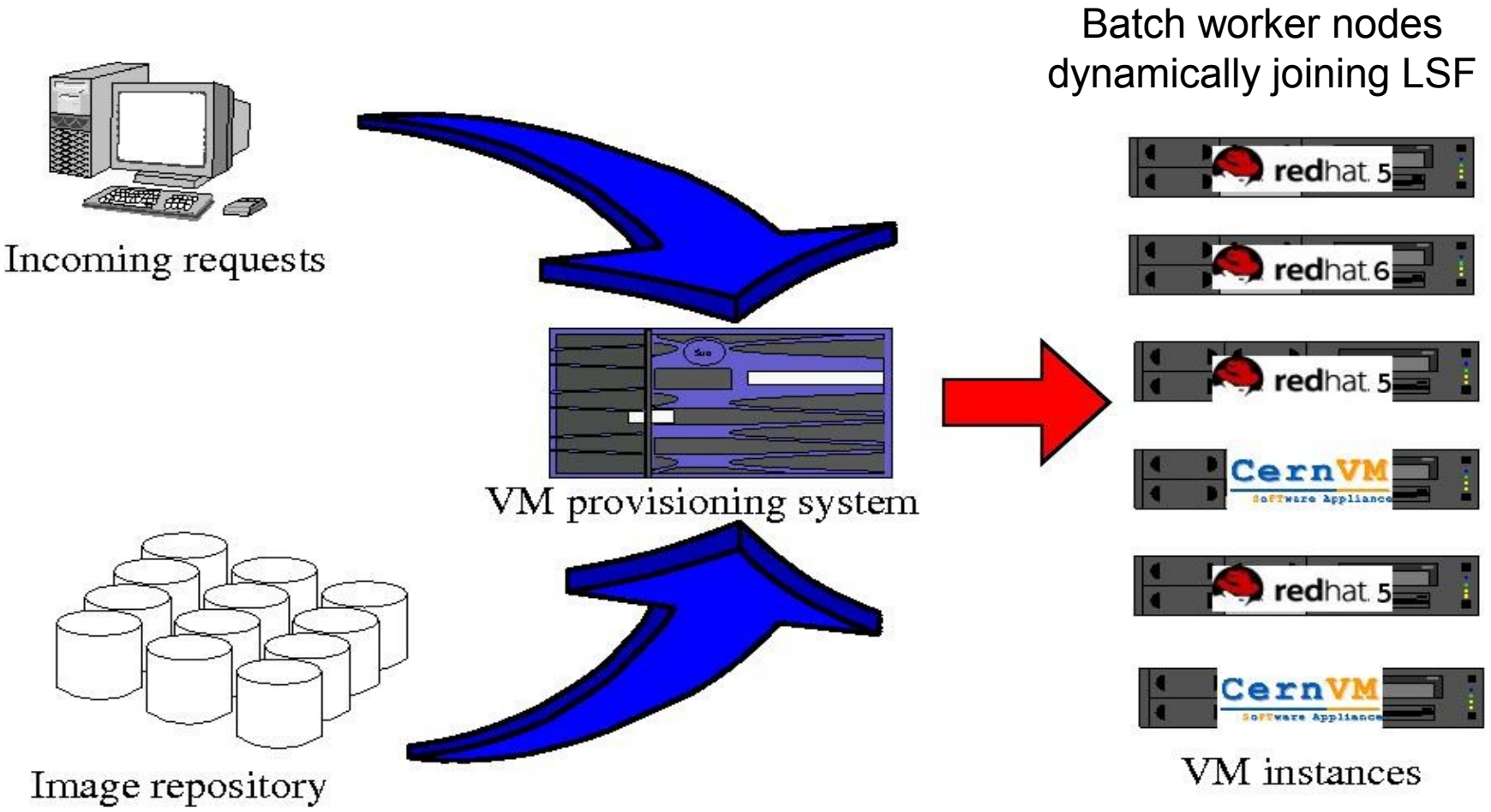
Key ideas for prototype at CERN

- **Isolation:** one job per virtual machine only
- **Limited Live-Time** for each worker node
- **Be agile:** always start from the latest image with the newest software
- **Be demand driven:** if possible, adjust running VMs to current or expected demand

Virtualization of batch resources

CERN IT Department

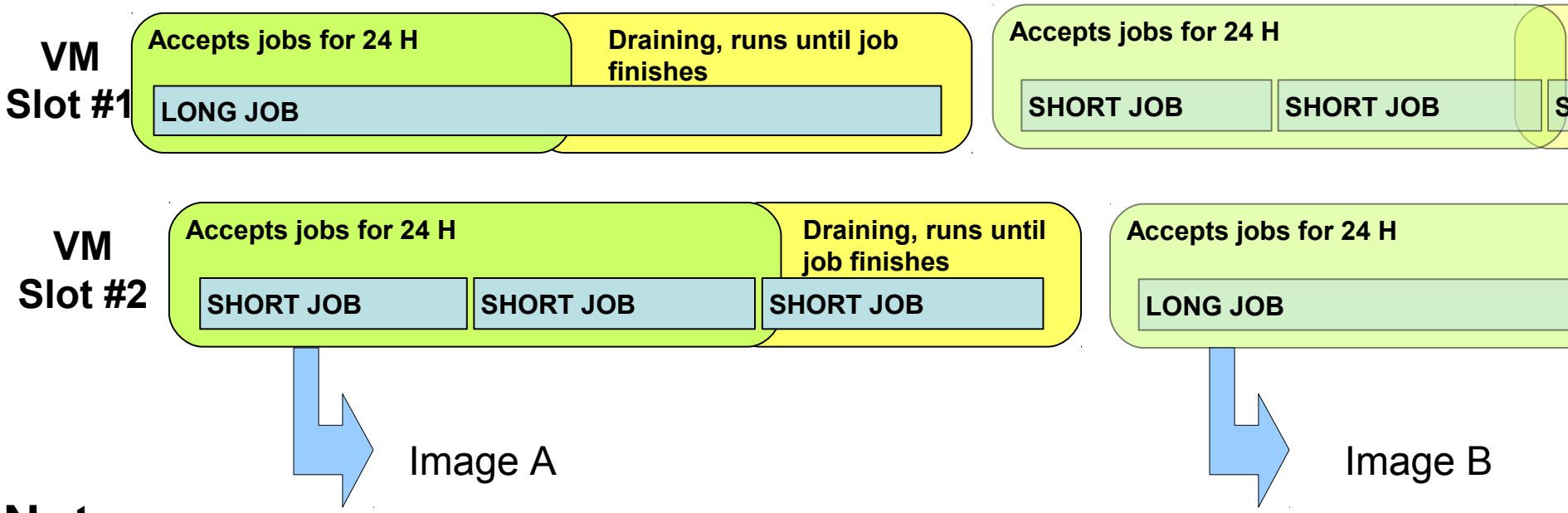
How to provide the **right mix** of environments matching needs ?



Virtualization of batch resources

CERN IT Department

Automate intrusive interventions: kernel, afs, glibc updates



Notes:

- NEW virtual machines always start with the latest image
- Image A and Image B can correspond to different OS versions

Thanks to Ricardo Silva

Cooking up the infrastructure

CERN IT Department

Administrator

Image creation

Image repository

Provisioning system
VM Orchestrator

Image distribution

Application
(batch)

Computing resources

Ups ... looks like an IaaS infrastructure, doesn't it ?

CERN's “lxcloud” system choices

CERN IT Department

Image creation:

Software setup can be derived from centrally managed computers

Image distribution:

currently Bit-torrent in use

VM orchestrator:

tested **OpenNebula** and **Platform ISF**, interested in **OpenStack**

CERN's “lxcloud” system choices

CERN IT Department

Virtualization technology

Started with XEN, moved to KVM in 2010

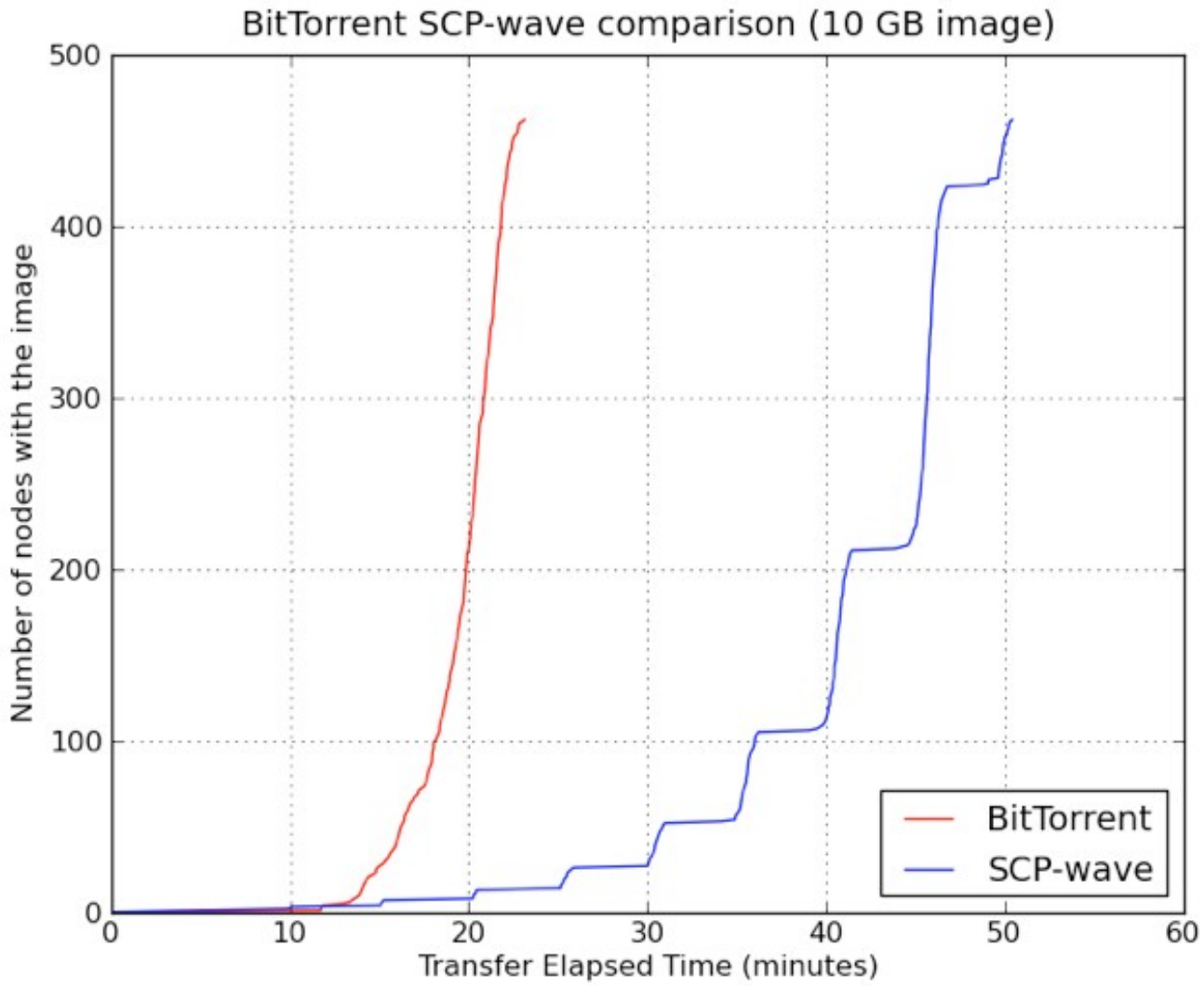
Image treatment:

Prestaged images, using LVM snapshot

Public cloud interface(s)

Tests with ONE EC2 access using ATLAS hammer-cloud tests (selected users only)

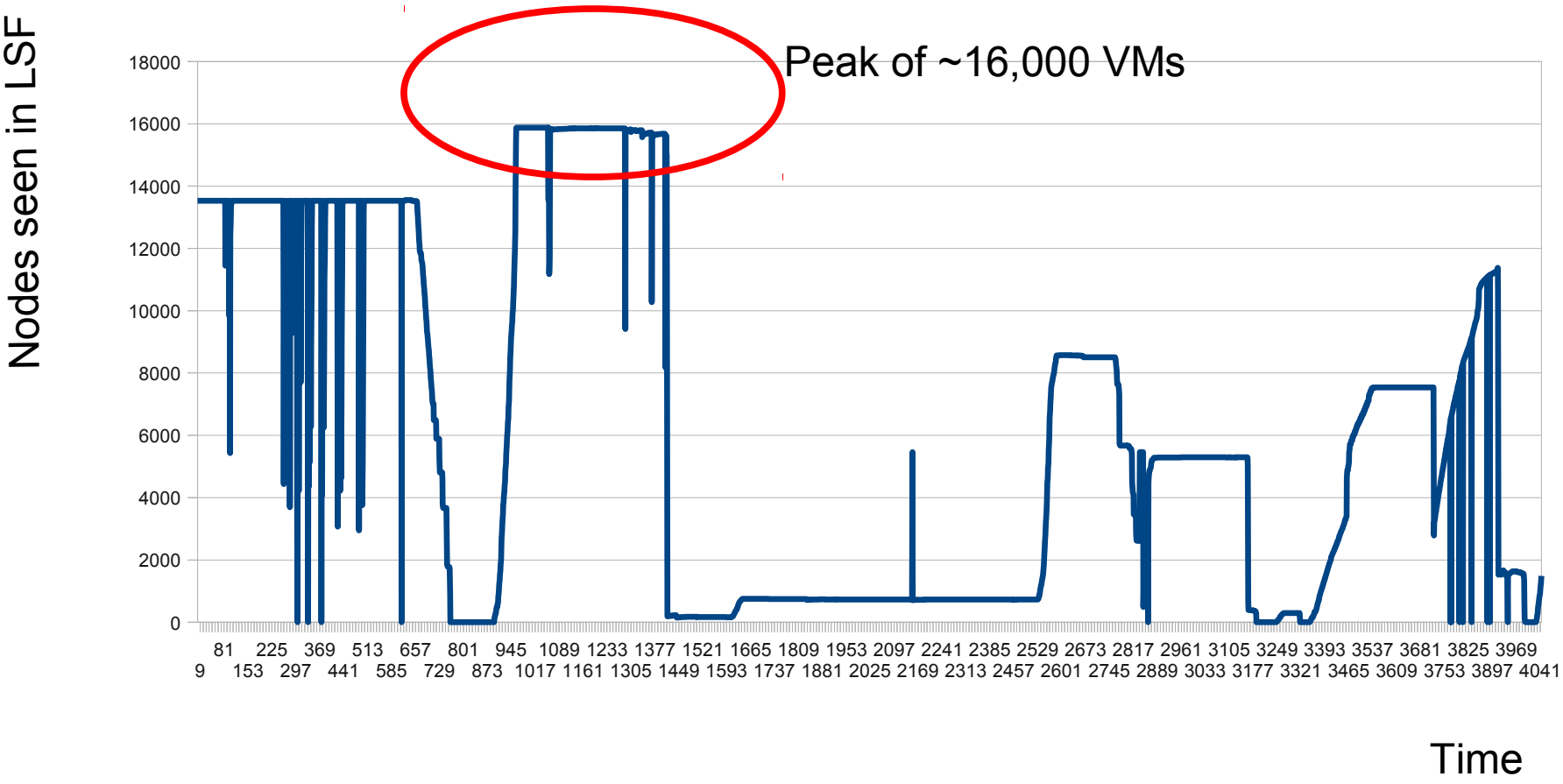
Image distribution performance



VM provisioning (ONE)

CERN IT Department

Virtual machine provisioning system tests



Is this a cloud infrastructure

Is this a cloud infrastructure

CERN IT Department

Does it matter ?

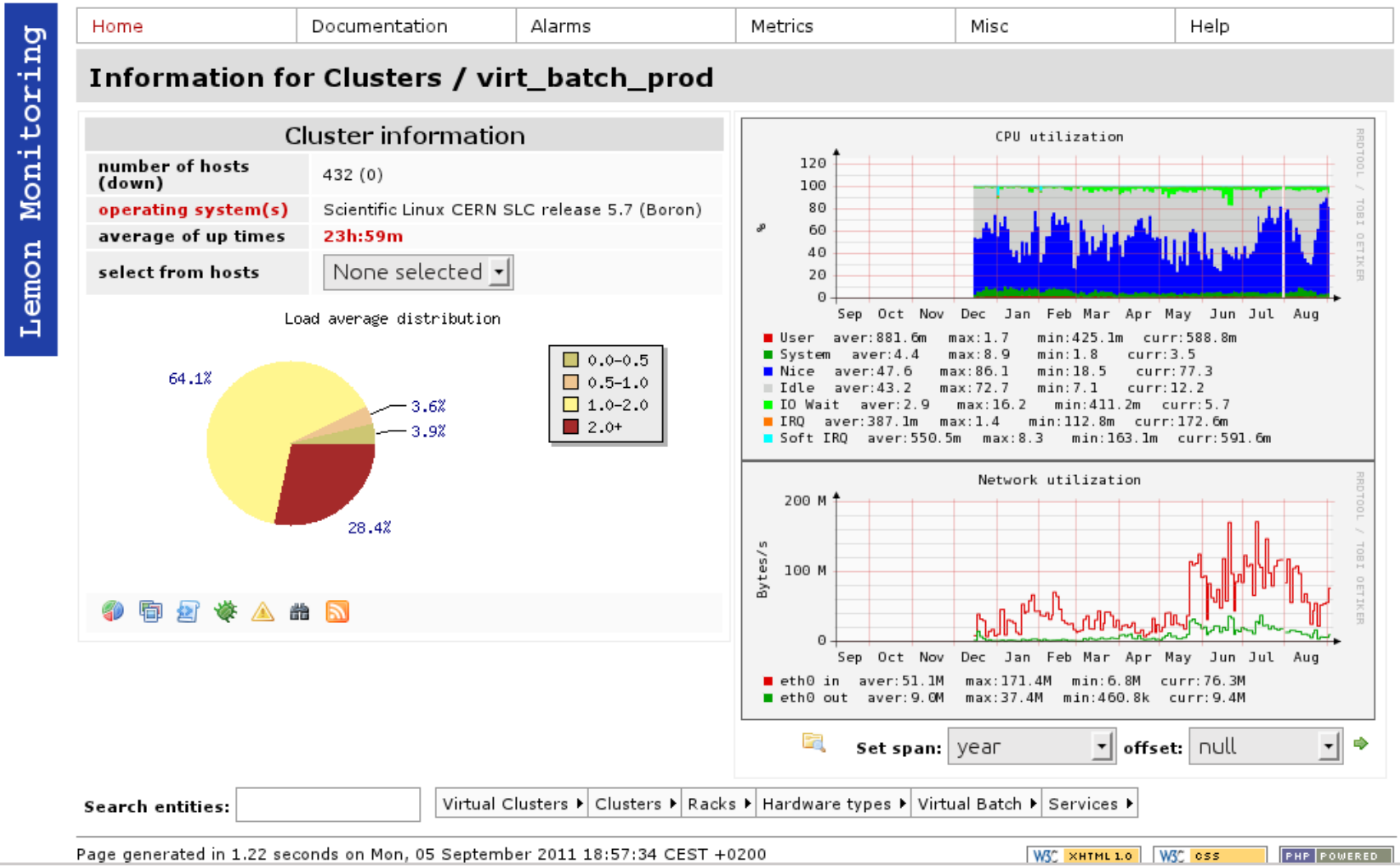
… as long as it does what we need

It has shares some features with the Cloud definition.

Virtual batch at CERN

CERN IT Department

In production since December 2010 at CERN



Batch virtualization: caveat

The scalability challenge

- The bulk of applications are single threaded.
- Therefore, a full virtualization of current CERN batch resources translates into

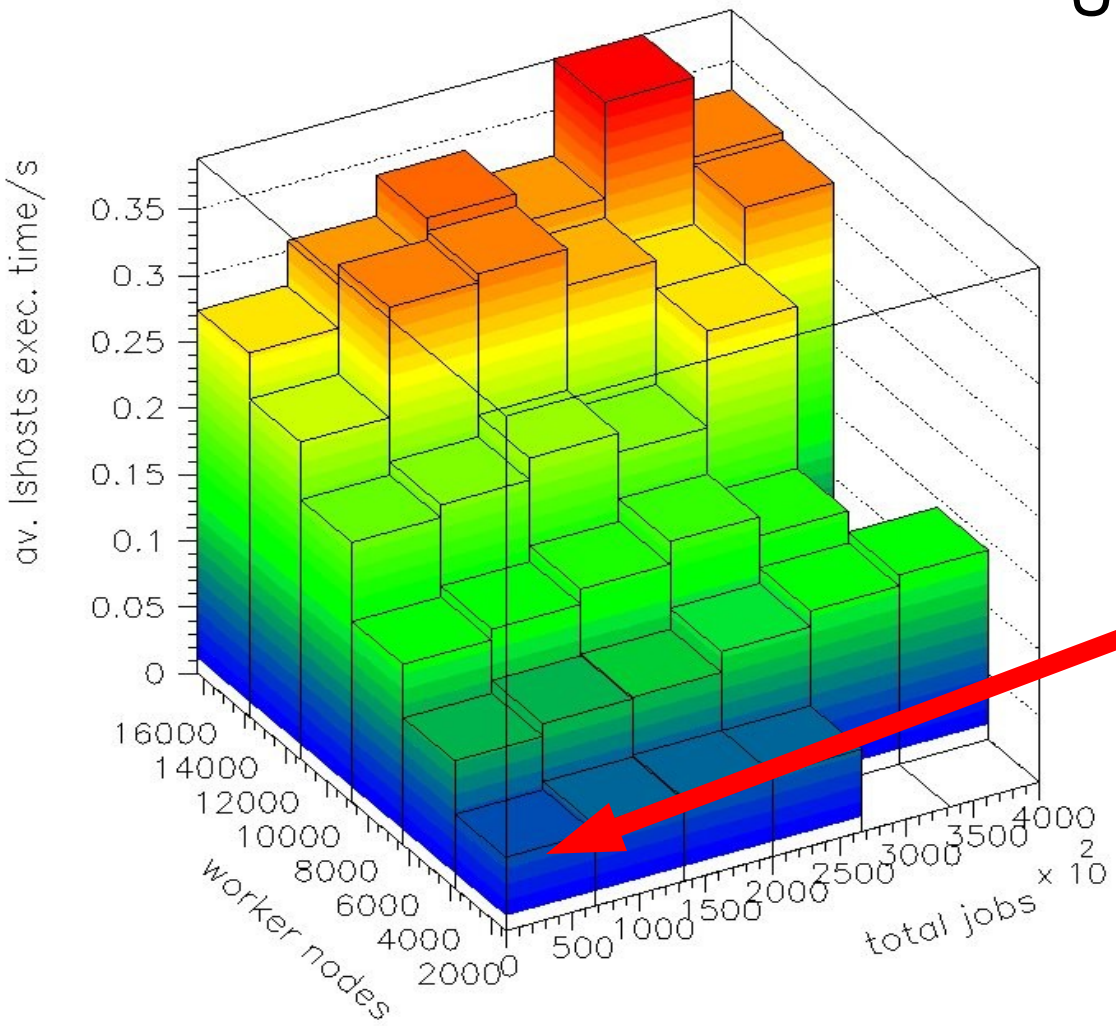
  > 31,000 VMs in a single batch instance

Feasibility studies

CERN IT Department

Batch system tests: resource layer

Up to 15,000 nodes
Up to 400,000 jobs



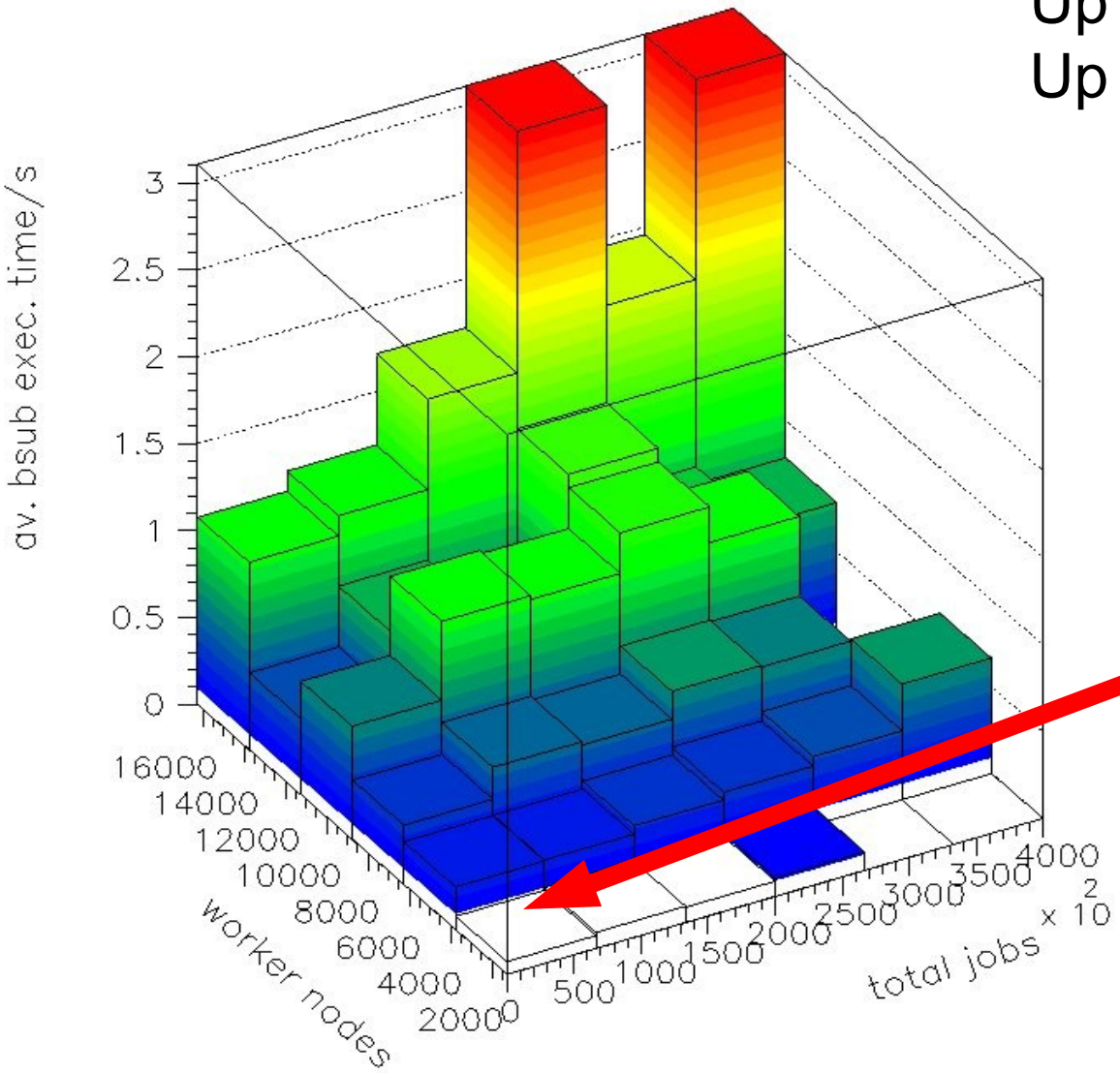
More than **3x** of what is officially supported

Current production system

Feasibility studies

CERN IT Department

Batch system tests: job submission

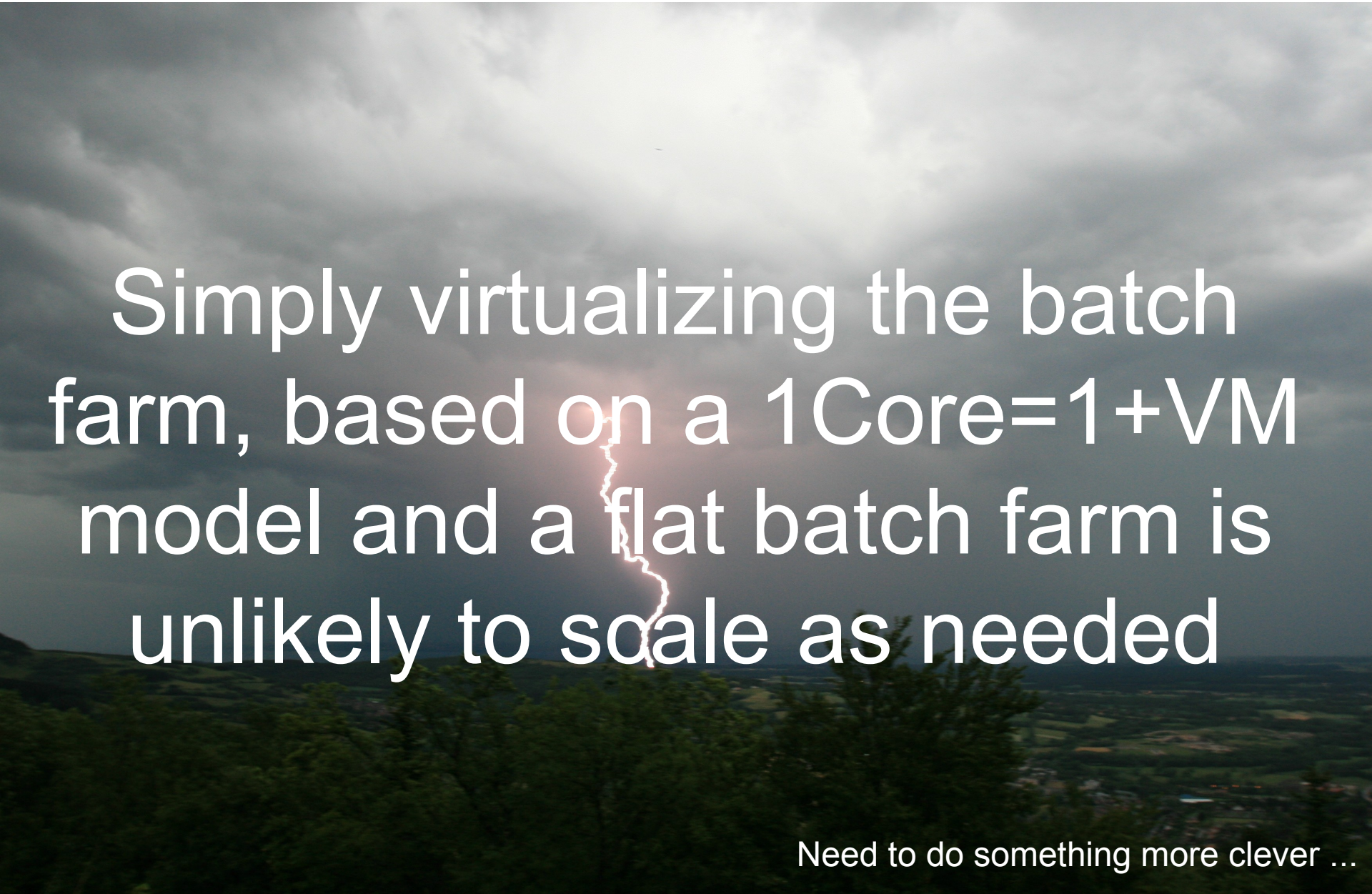


Up to 15,000 nodes
Up to 400,000 jobs

More than **3x** of what is officially supported

Current production system

Batch virtualization scalability



Simply virtualizing the batch farm, based on a 1Core=1+VM model and a flat batch farm is unlikely to scale as needed

Need to do something more clever ...

Going cloud ...

CERN IT Department

Enduser
VO

Quattor

CERNVM

“Golden Nodes”

Application
Manager
(lxbatch)

ONE EC2 interface

Image creation

ONE
test master

CERN VMIC

ONE
production master

VM provisioning

Lxcloud - Test

Lxcloud production

Physical Resources

Public cloud access tests

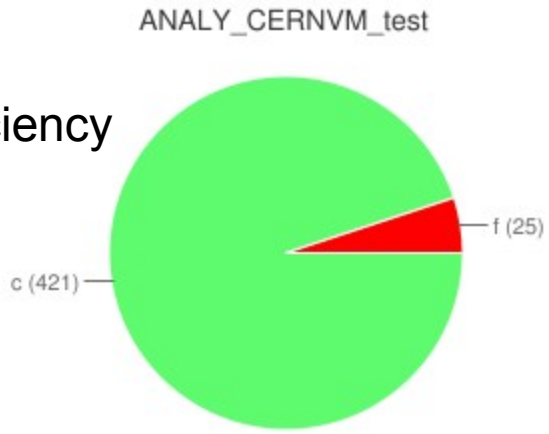
CERN IT Department

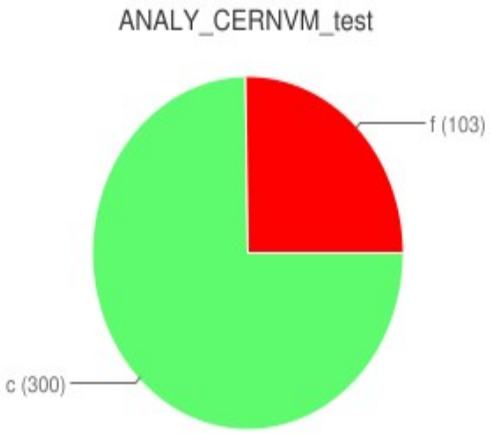
EC2 lxcloud

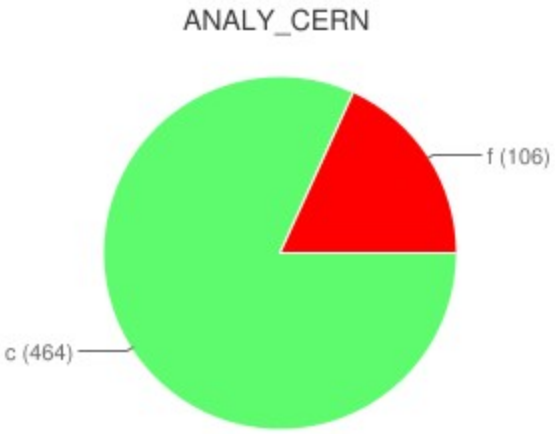
Direct submission

Regular GRID jobs

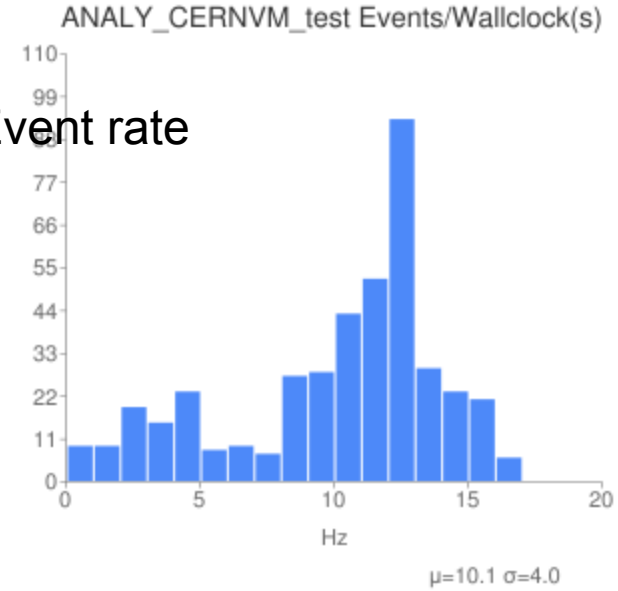
Efficiency

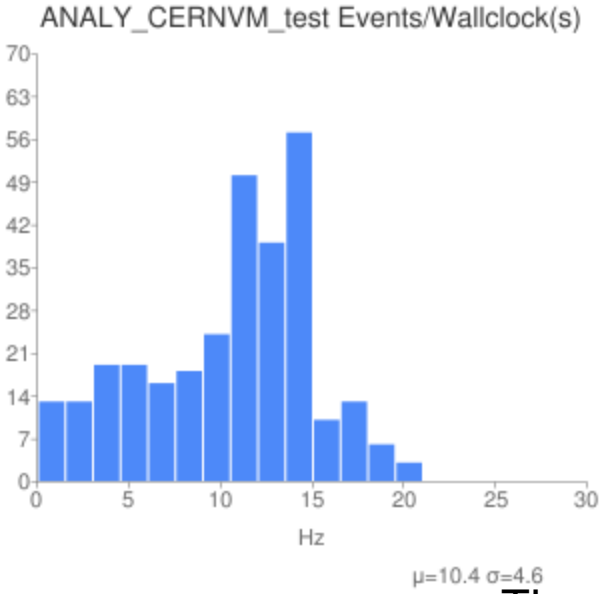


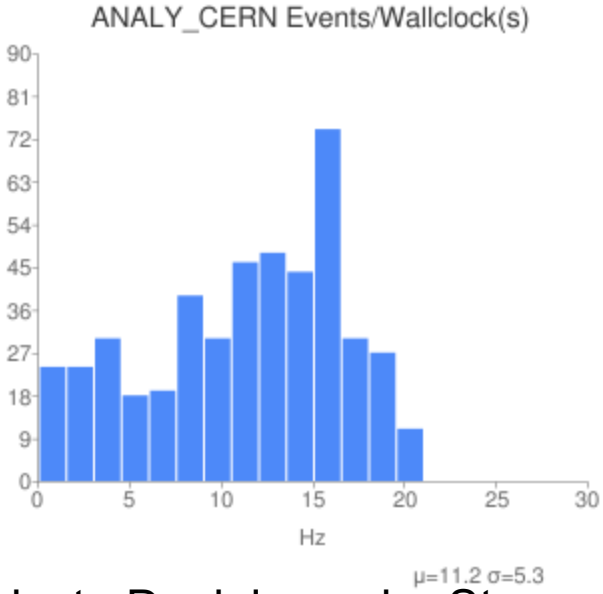




Event rate

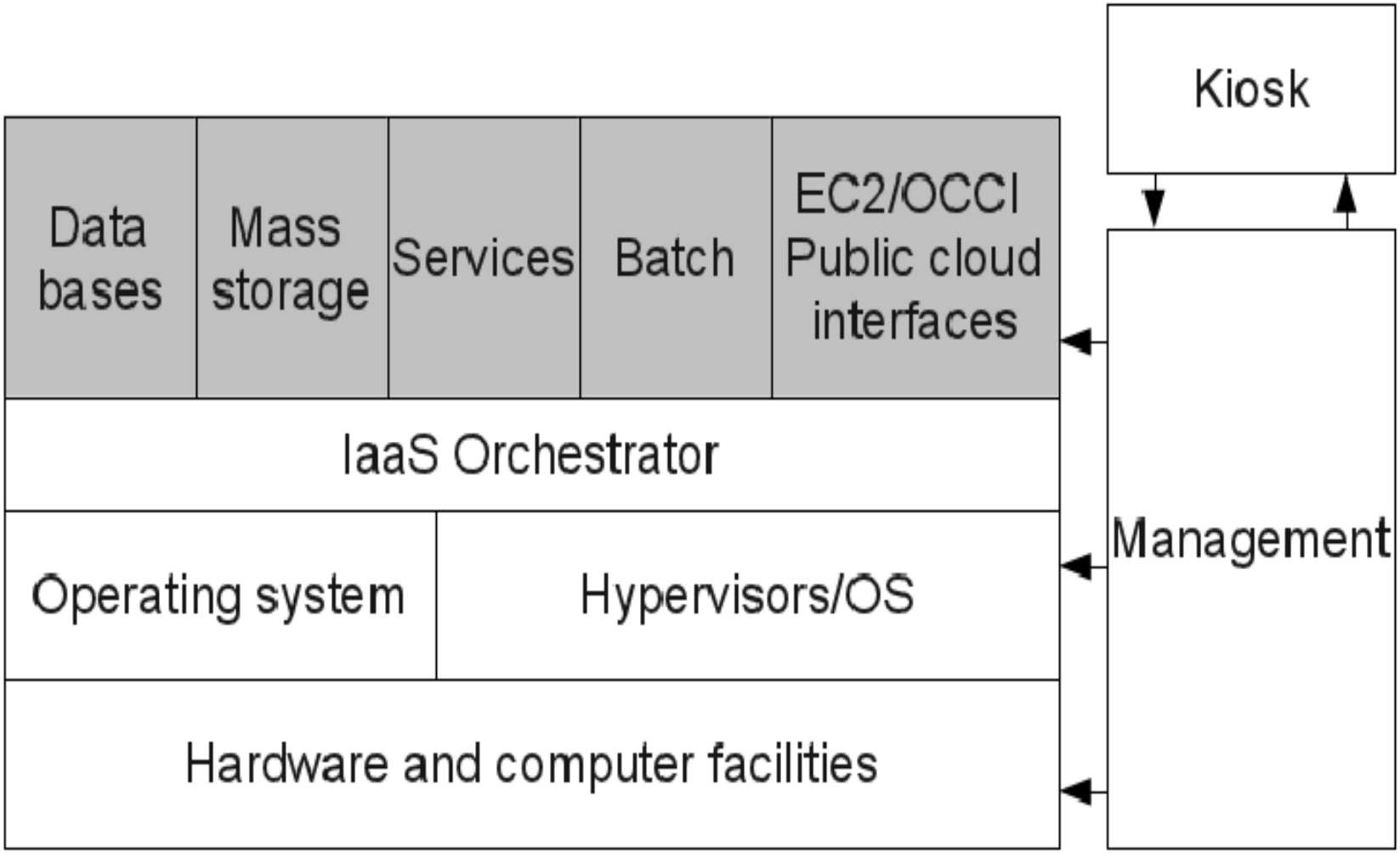






Thanks to Daniel van der Ster

A more general IaaS model ?



Conclusions

- Cloud computing is certainly a nice new technology, and technology is maturing
- First attempts to apply it are in progress at several sites, to solve operational issues rather than to please users
- Virtualization plays a vital role in the applied models, and is already reality
- Application scalability is an issue for larger scale deployments
- Looking forward to an interesting future


Thank you!
Any questions ?