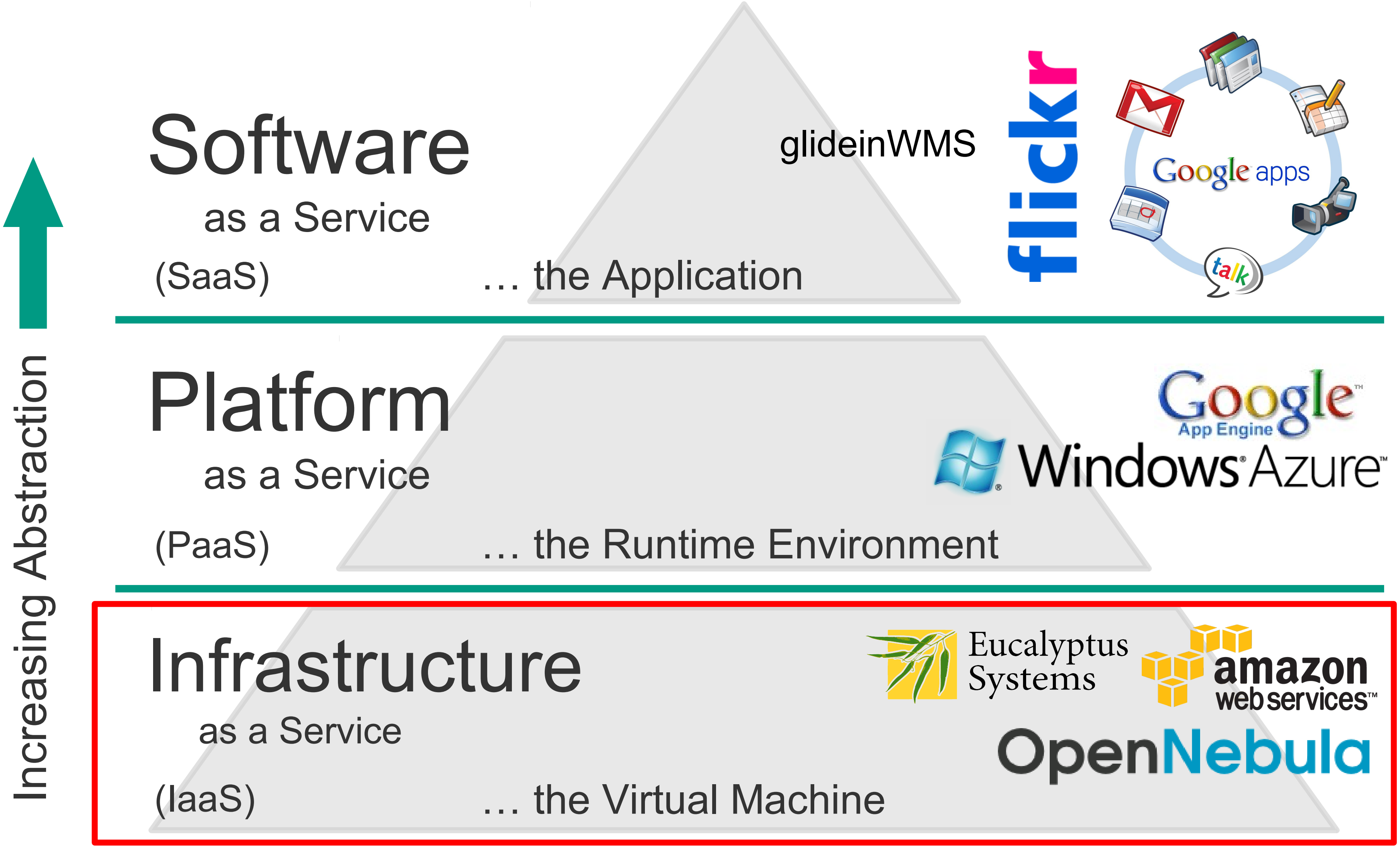
Karlsruhe Institute of Technology

Clouds in High Energy Physics

Oliver Oberst, Günter Quast, Stephan Riedel

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (EKP) – FAKULTÄT FÜR PHYSIK

What is Cloud Computing?



Increasing Abstraction

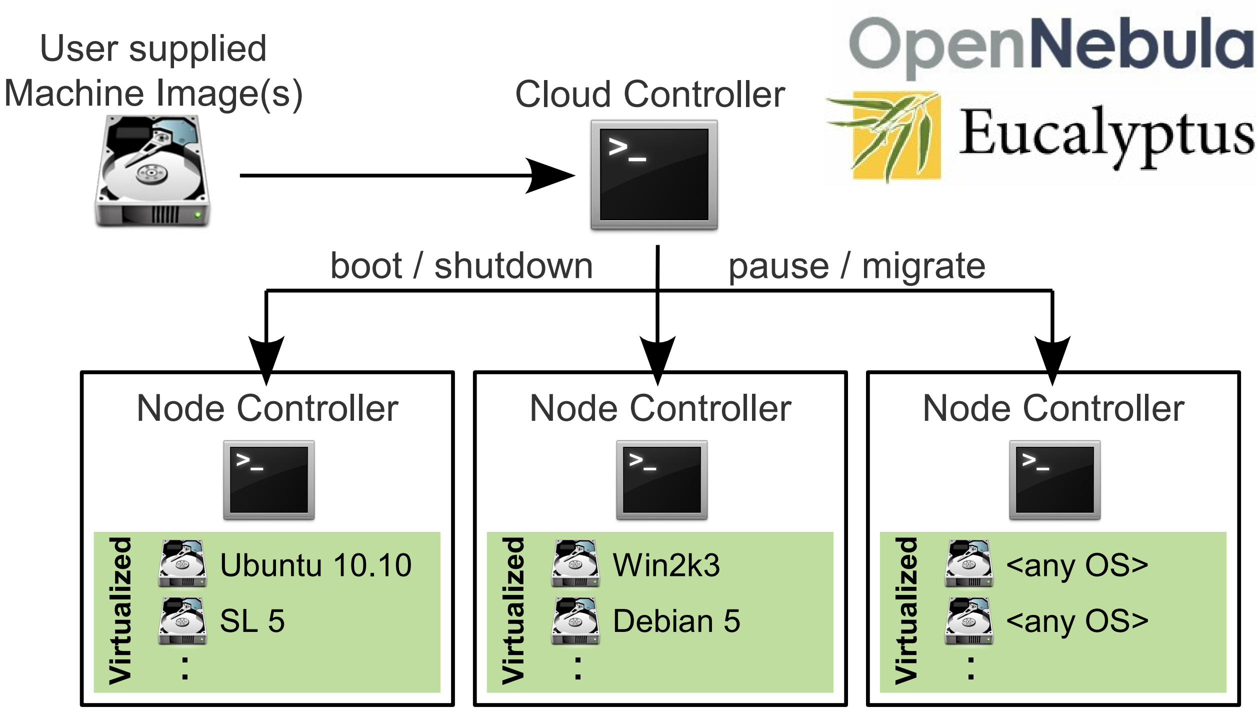
Software as a Service (SaaS) … the Application

glideinWMS

Platform as a Service (PaaS) … the Runtime Environment

Infrastructure as a Service (IaaS) … the Virtual Machine

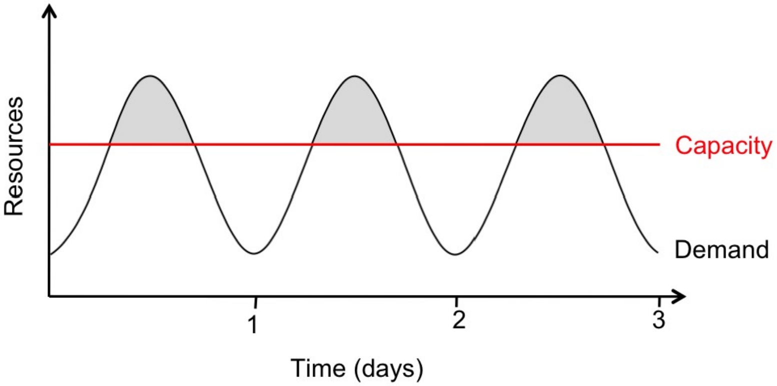
Infrastructure as a Service (IaaS)



- There are **Public Clouds** (commercial, e.g. Amazon Web Services) and **Private Clouds** (non-commercial, e.g. Cloud installations at Steinbuch Centre for Computing (SCC), KIT).
- Most **Public Cloud** Provider offer a **pay-as-you-go model** → Only pay for services actually used.
- **Network** bandwidth is still **expensive** → Public Clouds MAY currently not be an affordable solution for **I/O intense HEP data analyses**.
- A Private Cloud Interface can be set on top of already existing machines

Why do we need to extend to the Cloud?

- **Typical scenario** during HEP analyses: During **periods of high demand** (before conferences, finishing publications, etc.) local clusters are often completely occupied.
    - Buying a new Cluster is expensive and most of the time it will be underutilized.
    - Solution: **Cloud burst**. Elasticity of Cloud Providers allows to dynamically add / remove Cloud resources and thus **to handle peak loads**.
- **Constraints** in Hard- and Software setup (e.g. Scientific Linux) **prevent usage** of many shared resources by HEP community.
    - Through virtualization these resources can be made available.
- **Installation and maintenance** of HEP and Grid specific tools is a time consuming task
    - Management can be simplified by using **prepacked images** (e.g. CernVM and CernVMFS)



CernVM
SoFTware Appliance

HEP Cloud Requirements

- Toy Monte-Carlo studies:
  - Computing Power!
  - No big storage or network throughput needed
    → Easy adoptable to existing cloud solution (see later)
- Data analysis:
  - MC Productions (with detector simulation) and data reconstruction:
    - require large I/O
      → Hard/Expensive task for current available public clouds
  - User analysis:
    - May require large input (OK e.g. Amazon EC2 network input free!!)
    - Generally less data produced
      - N-tuples, Histograms,...

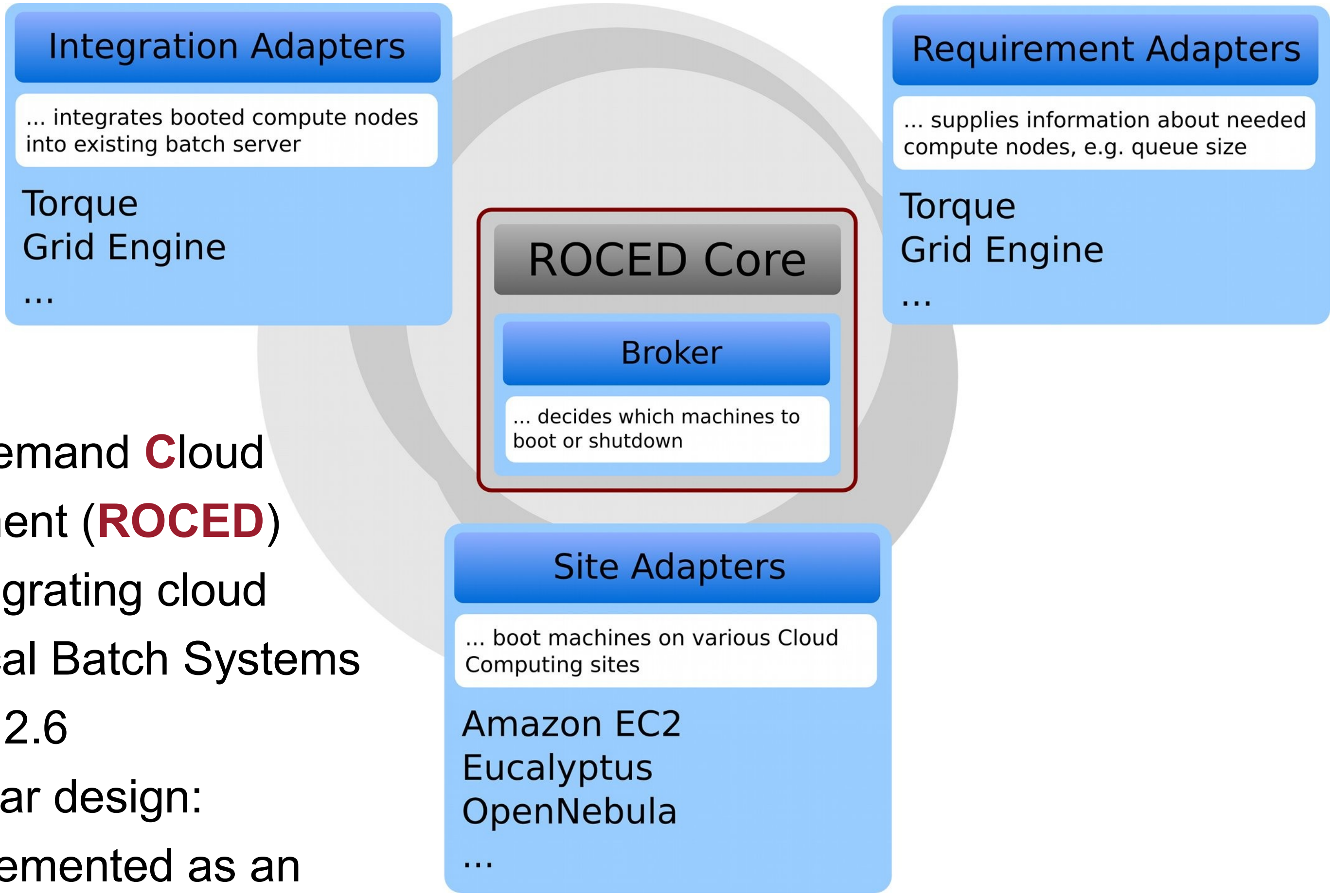
→ Current Clouds in use within HEP break “job goes where the data is”

→ I/O intensive tasks may suffer from limitations within the used Clouds

Current cloud solutions for HEP Examples

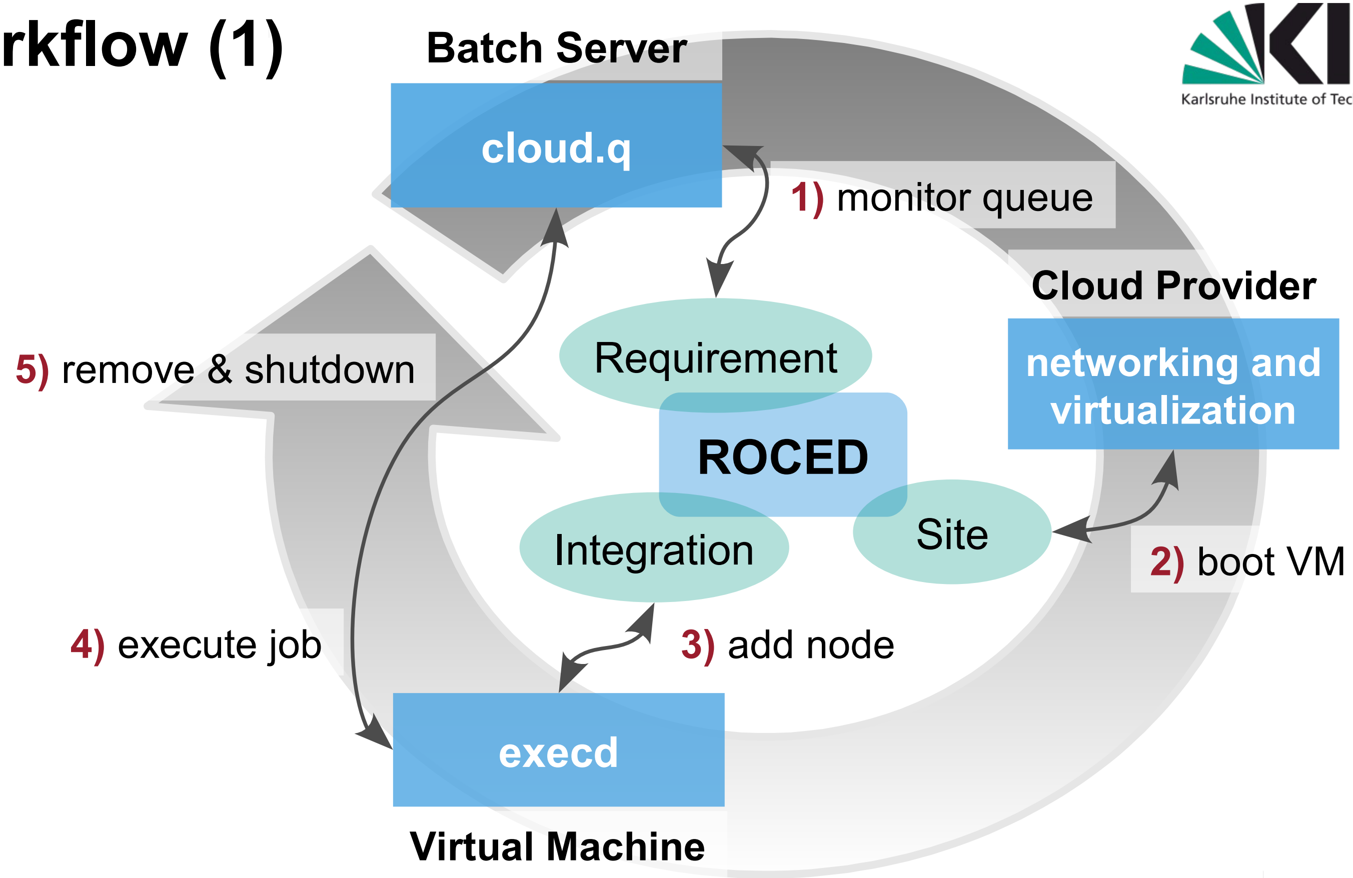
- Extension of the local batch systems
  - Two examples in the following
    - ROCED (here at KIT)
    - Nimbus
- Data storage capabilities in within a HEP Cloud?
  - Cummuls

The ROCED Concept


Integration Adapters
... integrates booted compute nodes into existing batch server
Torque
Grid Engine
...
Requirement Adapters
... supplies information about needed compute nodes, e.g. queue size
Torque
Grid Engine
...
ROCED Core
Broker
... decides which machines to boot or shutdown
Site Adapters
... boot machines on various Cloud Computing sites
Amazon EC2
Eucalyptus
OpenNebula
...

- **R**esponsive **o**n-demand **C**loud **E**nabled **D**eployment (**ROCED**)
- Scheduler for integrating cloud resources into local Batch Systems
- Written in Python 2.6
- Features a modular design: Every part is implemented as an Adapter. New Adapters can be added easily.

ROCED Workflow (1)



1) Passive monitoring of one or more batch server queues.

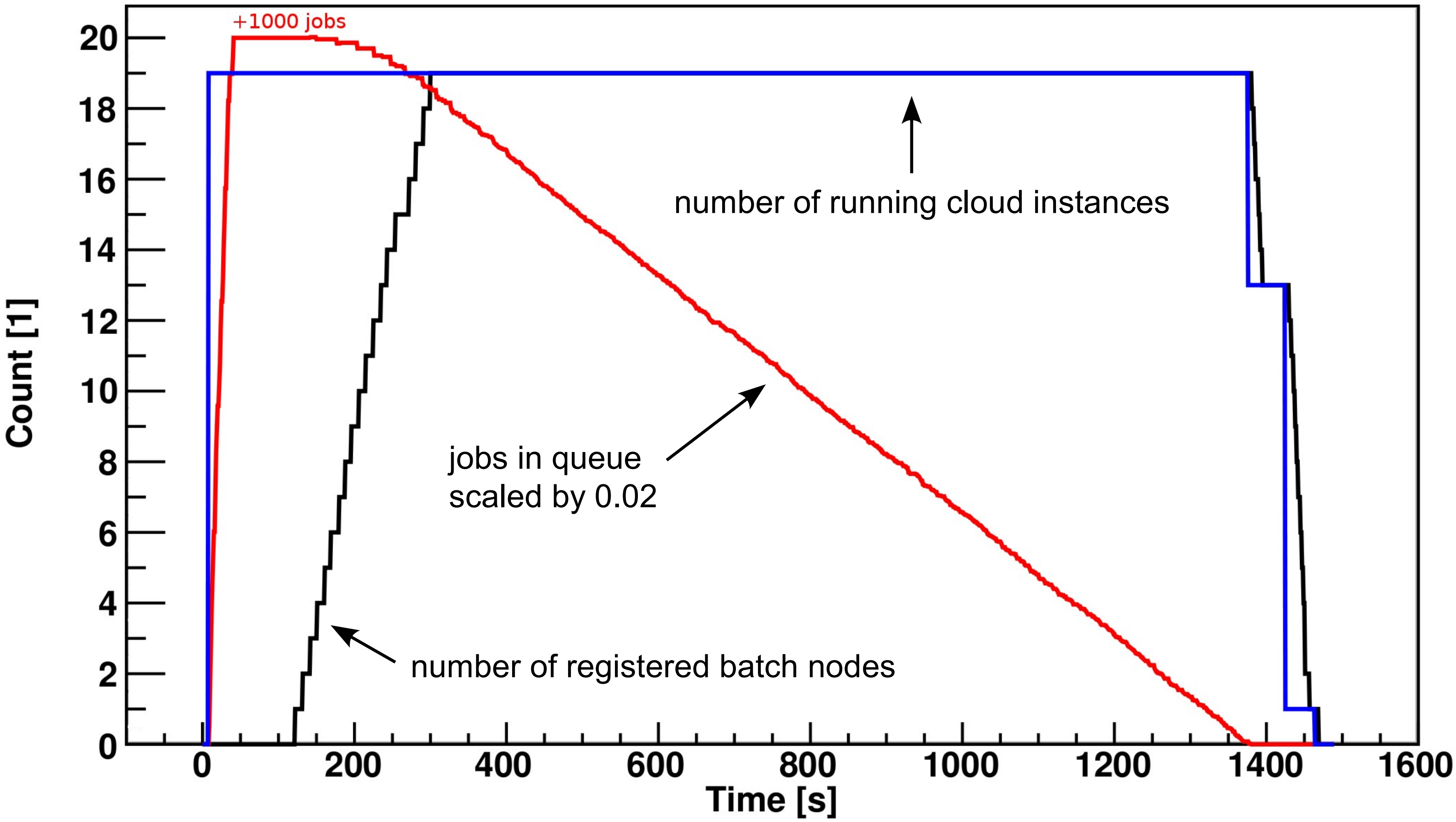
2) Depending on queue size ROCED decides how many machines are needed.

3) The needed virtual machines are started by ROCED.

4) After startup, the new hosts are added, and become available for job processing.

5) After job execution, the VMs are automaticaly removed from the batch system and shut down.

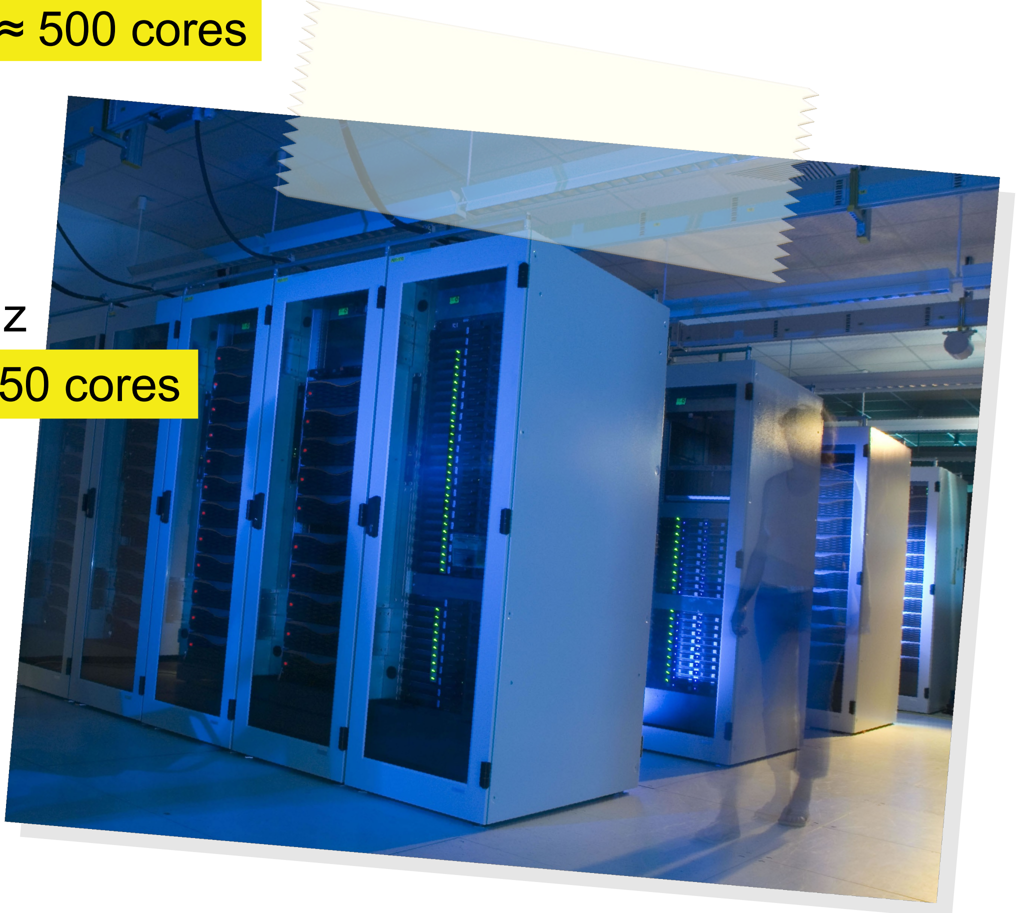
ROCED Workflow (2)



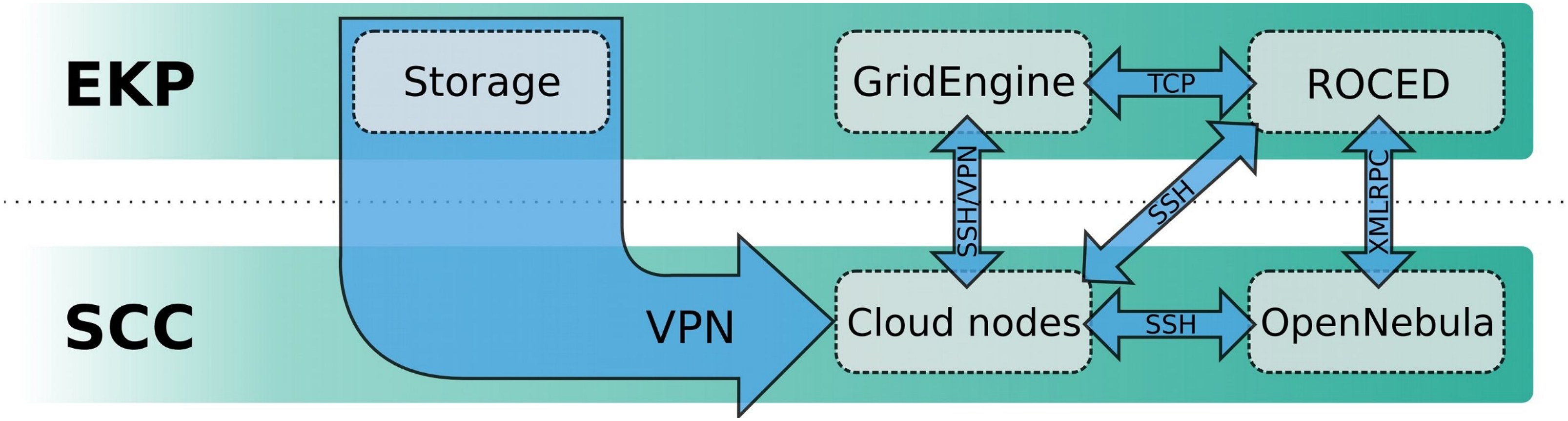
Stress test of ROCED (here with Amazon/Torque)

Cloud Resources @ KIT (SCC)

- OpenNebula Cloud:
  - 2x Xeon E5520 (Nehalem) 2.27 GHz
  - 60 nodes, each with 8 cores ≈ 500 cores
  - 36 GB RAM
  - 2 TB HD
- Eucalyptus Cloud
  - 2x Xeon 5400 series 2,33 GHz
  - 6 nodes, each with 8 cores ≈ 50 cores
  - 16 GB RAM
  - 100 GB HD
- Ran by SCC at Campus North
- In use by IEKP Campus South
  - Using ROCED



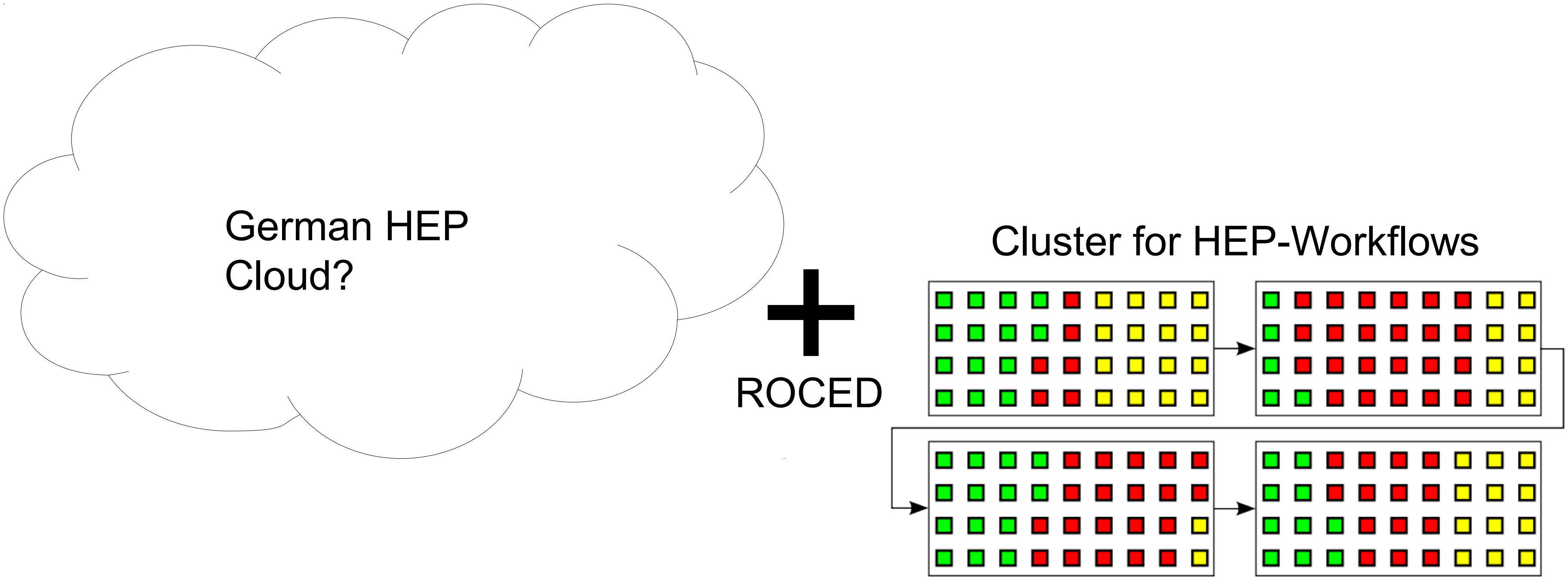
ROCED Setup



- ROCED runs on the same machine as the GridEngine.
- GridEngine communicates with its nodes and users via TCP.
- Commands to the OpenNebula host are sent via XMLRPC call.
- The Communication between the Cloud nodes, ROCED and the OpenNebula Server are done via SSH.
- No modifications to the firewall (besides VPN tunnel) needed
- When running only in a private cloud environment storage can be made available transparently (NFS/dcap...) (Expensive + I/O limited when using public clouds!)
- ROCED has still prototype status – **release/documentation ongoing**

Planned Cloud Projects at KIT

- KIT: Fusion of Dynamic Virtualization and Batch-to-Cloud Extension
  - Extension of local dynamically virtualized resources within private and public Clouds (cloudburst)
  - Extension of Grid resources into the Cloud



German HEP Cloud?

ROCED

Cluster for HEP-Workflows

Nimbus

- Open Source Science Cloud toolkit
  (formerly known as Virtual Workspace Service)
  - Provides Iaas to scientific clients with respect to their needs
    - Converts your cluster into a cloud site
    - Includes fairshairing via batch systems (PBS,SGE)
  - Supports Xen and KVM hypervisors on the virtualization layer
  - Uses “Contextualizaion” for setting up the virtual machines (“one-click” virtual cluster)
  - Supports X509 authentication/authorization
  - Also provides client software

http://nimbusproject.org

Cloud Storage

- Quasi standard: Amazon Simple Storage Service (Amazon S3)
  - Details on the underlying system not disclosed by Amazon
  - API: S3 REST, SOAP, Bittorrent
  - HDFS can run on top of S3.

ubuntu one

Dropbox

tumblr.

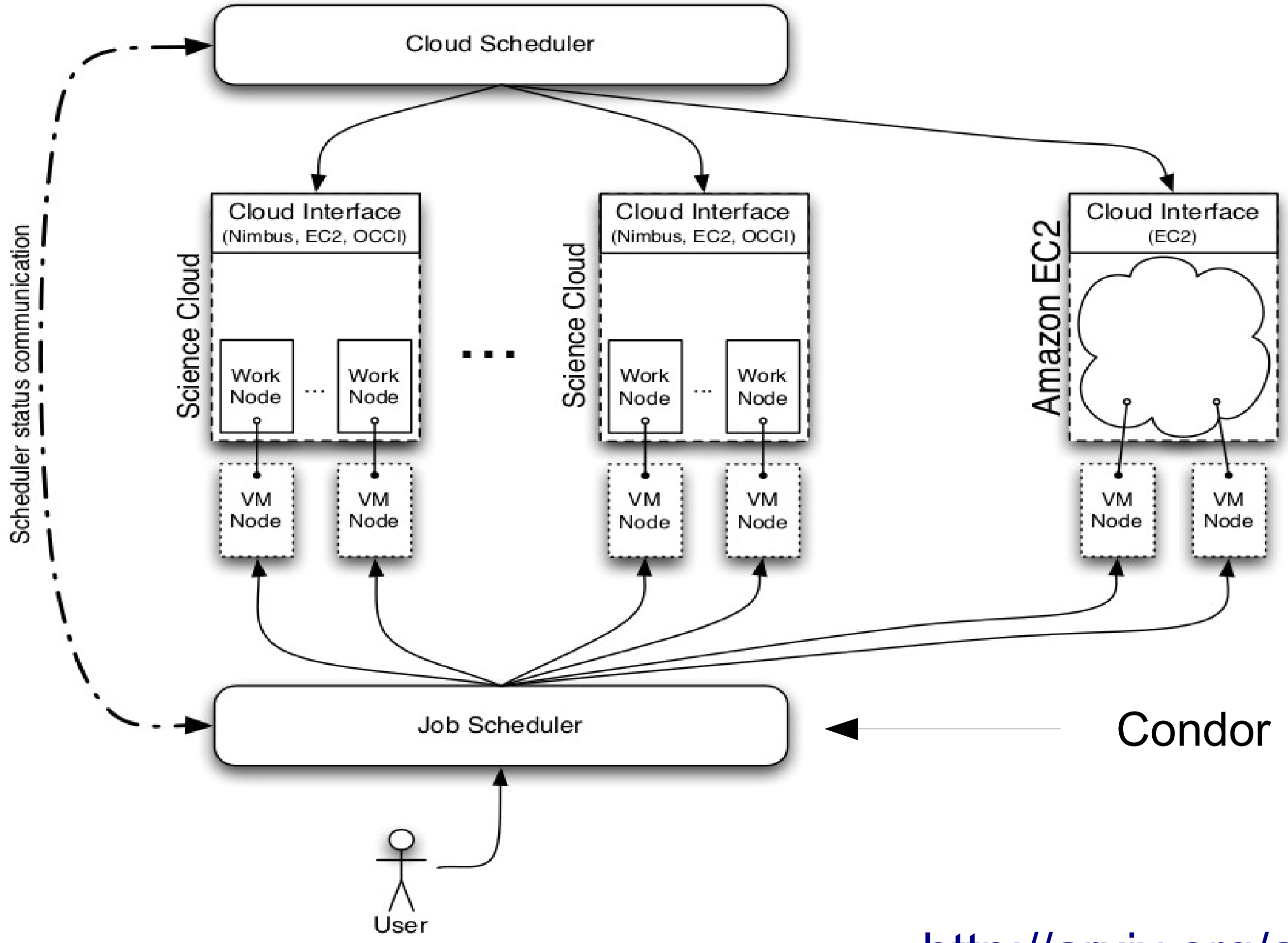
Cumulus:
An Open Source Storage Cloud for Science

- Open Source reimplementation of S3 REST API
  - Currently lacks some S3 REST functionality
- Runs on top of POSIX FS, HDFS
- Implemented in Python as a REST service (HTTP, HTTPS)
- Shipped with Nimbus or stand-alone
  - Used as VM image storage within Nimbus
- Comparable Performance to SCP or GridFTP
  (John Bresnahan et al. http://doi.acm.org/10.1145/1996109.1996115)

Clouds for HEP @ work?

- Data Intensive High Energy Physics Analysis in a Distributed Cloud http://arxiv.org/abs/1101.0357 (Jan 2011)
- Iaas Cloud using Nimbus installations at
  - National Research Council (NRC) in Ottawa
  - University of Victoria (UVIC)
- Additionally extended by Amazon EC2 resources
- Setup:
  - Xen hypervisor
  - VM Images: SLC 5.5 + BaBar Software
  - User login via X.509 Proxies
- Utilizes fast research network! (data to the job...)
- Test case accessed over a 4-5TB large test physics data sample

Clouds for HEP @ work?



Condor

http://arxiv.org/abs/1101.0357

Summary

- IaaS Clouds for HEP:
  - Field of active research!
  - First products available
  - Full HEP work-flows already tested
  - Large scale proof-of-principle missing
    (HEP cloud of many sites e.g. of a country)
- Clouds will in short term not replace the Grid in HEP, they extend the available resource space!

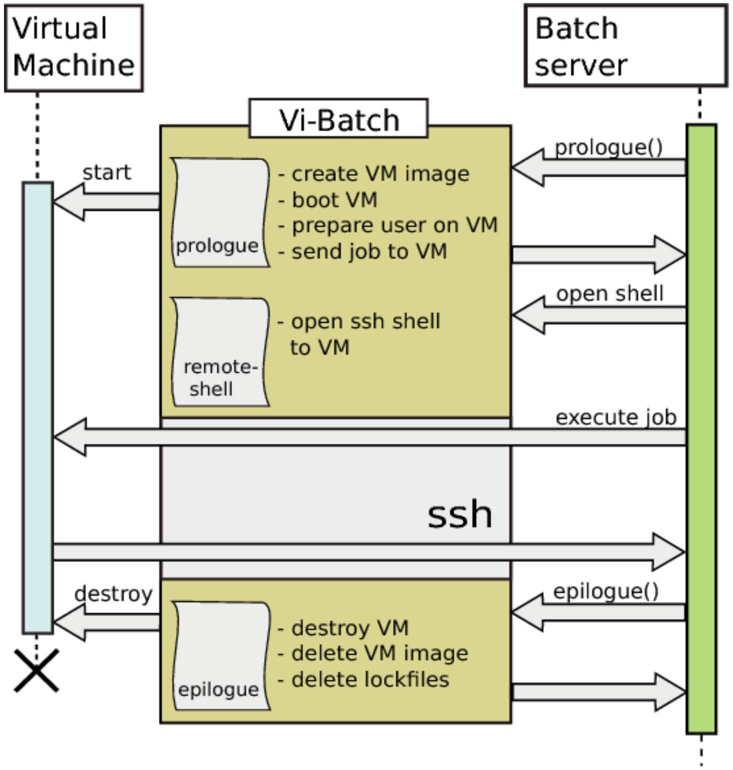
Thank you for your attention!

BACKUP

Dynamic Cluster Virtualization Project at KIT: Implementation

- ViBatch
  - Wrapper script (Prologue/Epilogue) around the actual computing job inside the batch system
  - Virtualization system hidden from the users
  - Technical Details:
    - Runs with MAUI/Torque
    - Easy portable to other batch systems (SGE tests in queue)
    - Currently uses SLC 5.1 VM images with CERNVMFS installed (CMS Software included!)
    - Runs in production for the IEKP at KIT (used now by several CMS analyses)
- Cooperation with DESY within this project (Desy implementation vmimagemanager.py)

Vibatch workflow:



https://ekptrac.physik.uni-karlsruhe.de/trac/BatchVirt