COASTAL CAROLINA UNIVERSITY

CLEMSON SCHOOL OF COMPUTING

PENNSTATE

Using Virtual Machines for OSG Virtual Organizations

Dr. Sebastien Goasguen
Lance Stout

Outline

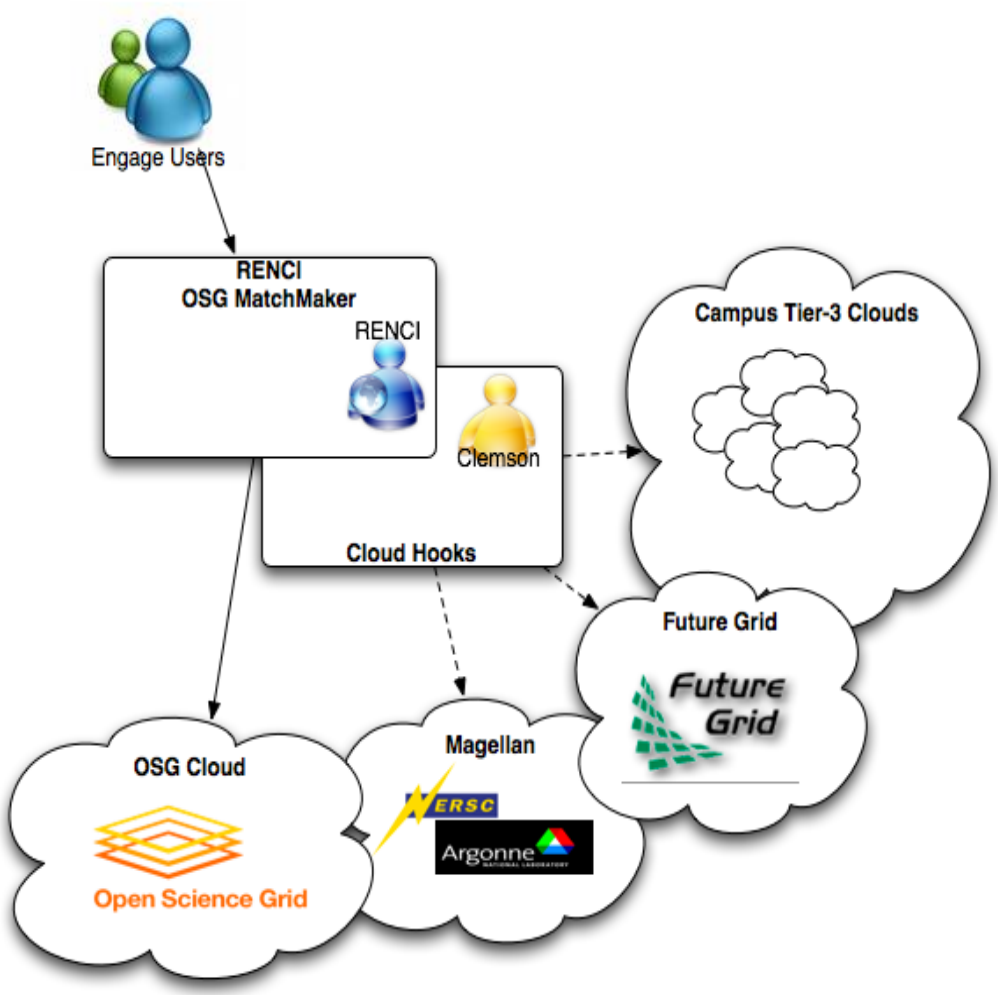
- Motivation and background
- Adapting Grids for use in Clouds
- Clemson's Contributions to OSG
- VM Enabled Clusters
- Networking Issues
- Kestrel and XMPP
- Submitting VMs using Engage
- Virtual Machine Image Catalogue

Motivation and background

Can we let users customize their computing environment and start this environment on grid sites?

Thereby providing an elastic, on-demand "cluster" over the OSG infrastructure, a.k.a a Cloud

Paradigm shift from running an application as a job to running an environment as a job.
Virtualization seen as a key enabler.

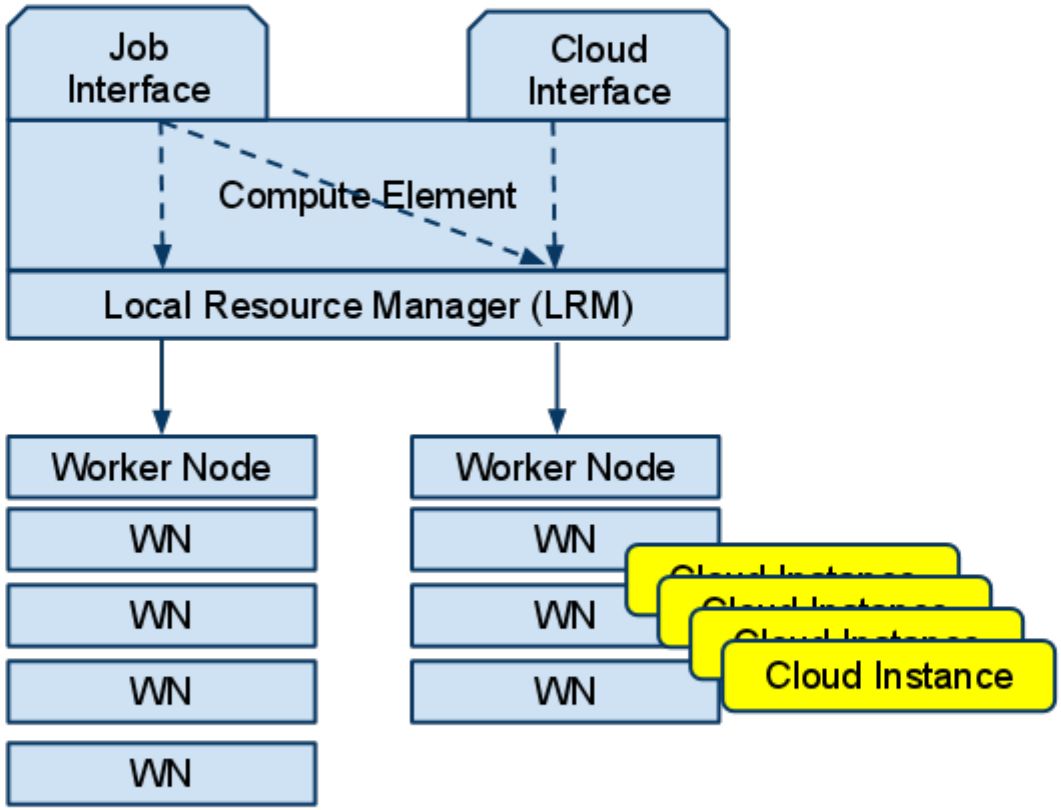


Adapting / Changing OSG sites for the Cloud

Adapting the Compute Elements to receive requests for servers ?

Or creating a new Cloud Element aka Cloud interface service ala EC2 ?

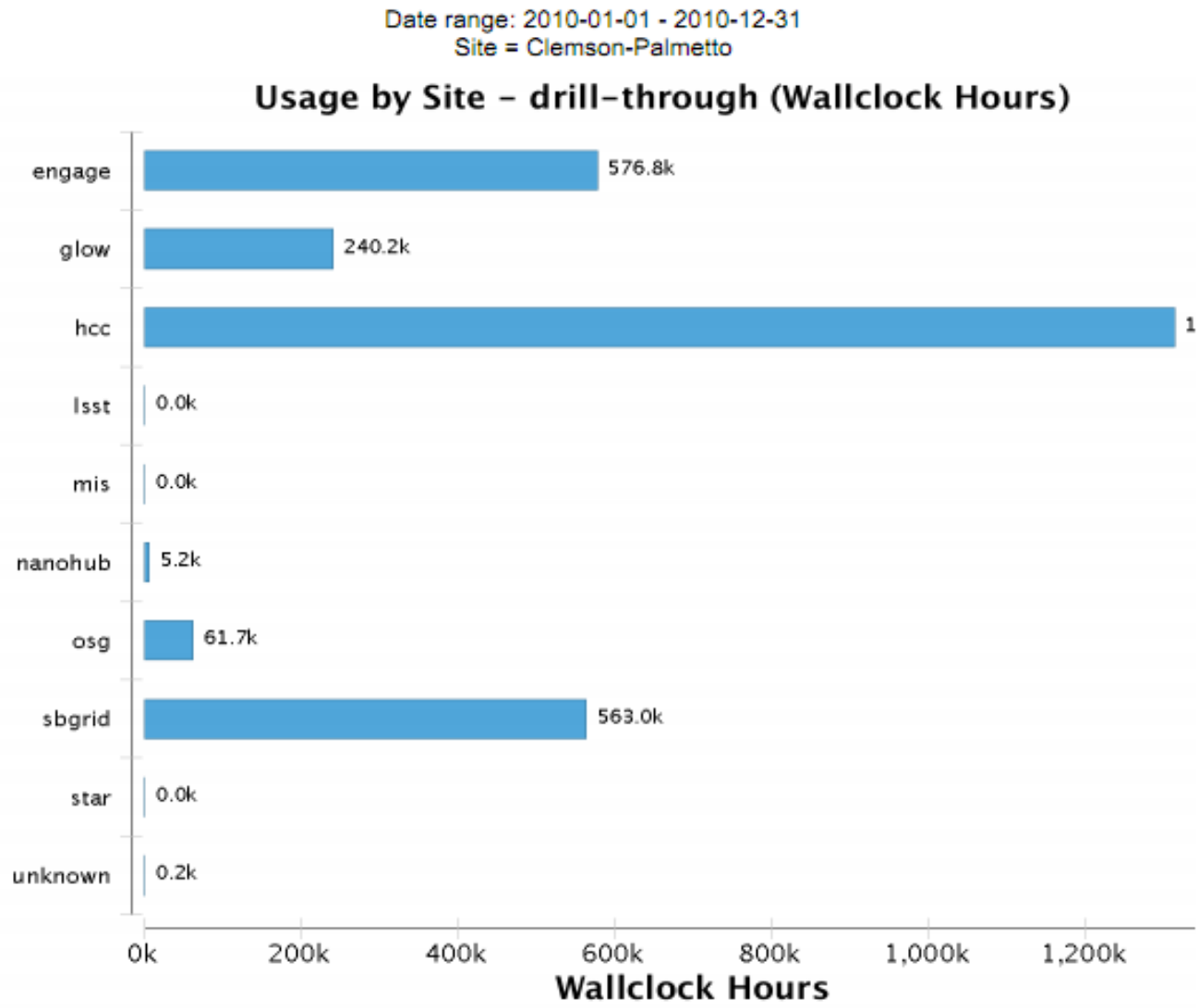
Both are possible, the Cloud interface enables on-demand characteristics while using the standard CE does not change much of the infrastructure.



Clemson contribution to OSG in 2010

~2.8M hours
Not only Engage but also:
GLOW
HCC
SBGRID

Also #8 in total credits to Einstein@home via our Windows Condor pool.



VM enabled Cluster at Clemson

KVM based
CentOS 5.x, ~12000 cores

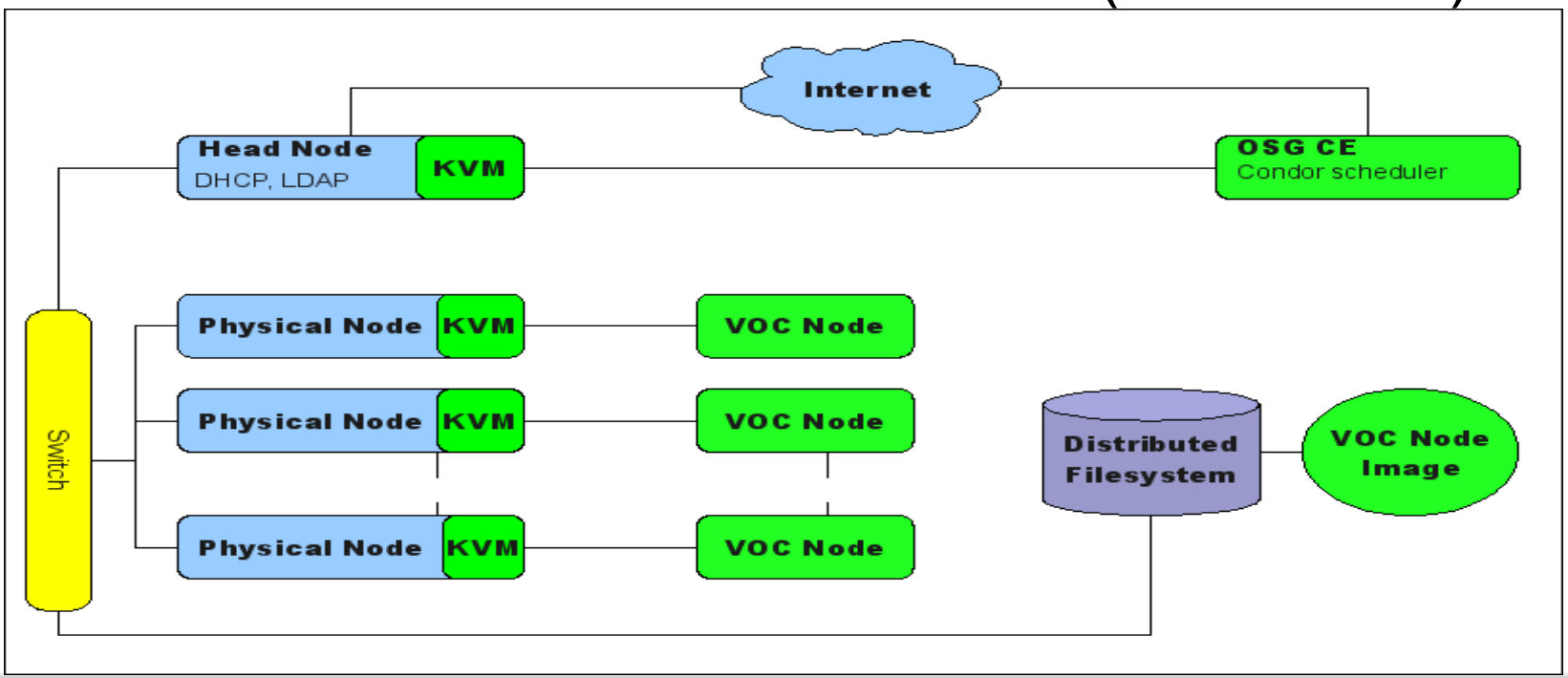
VMs started locally via either PBS or Condor

VMs started remotely via Condor-G

Images stored on shared file system using OrangeFS (formerly PVFS)

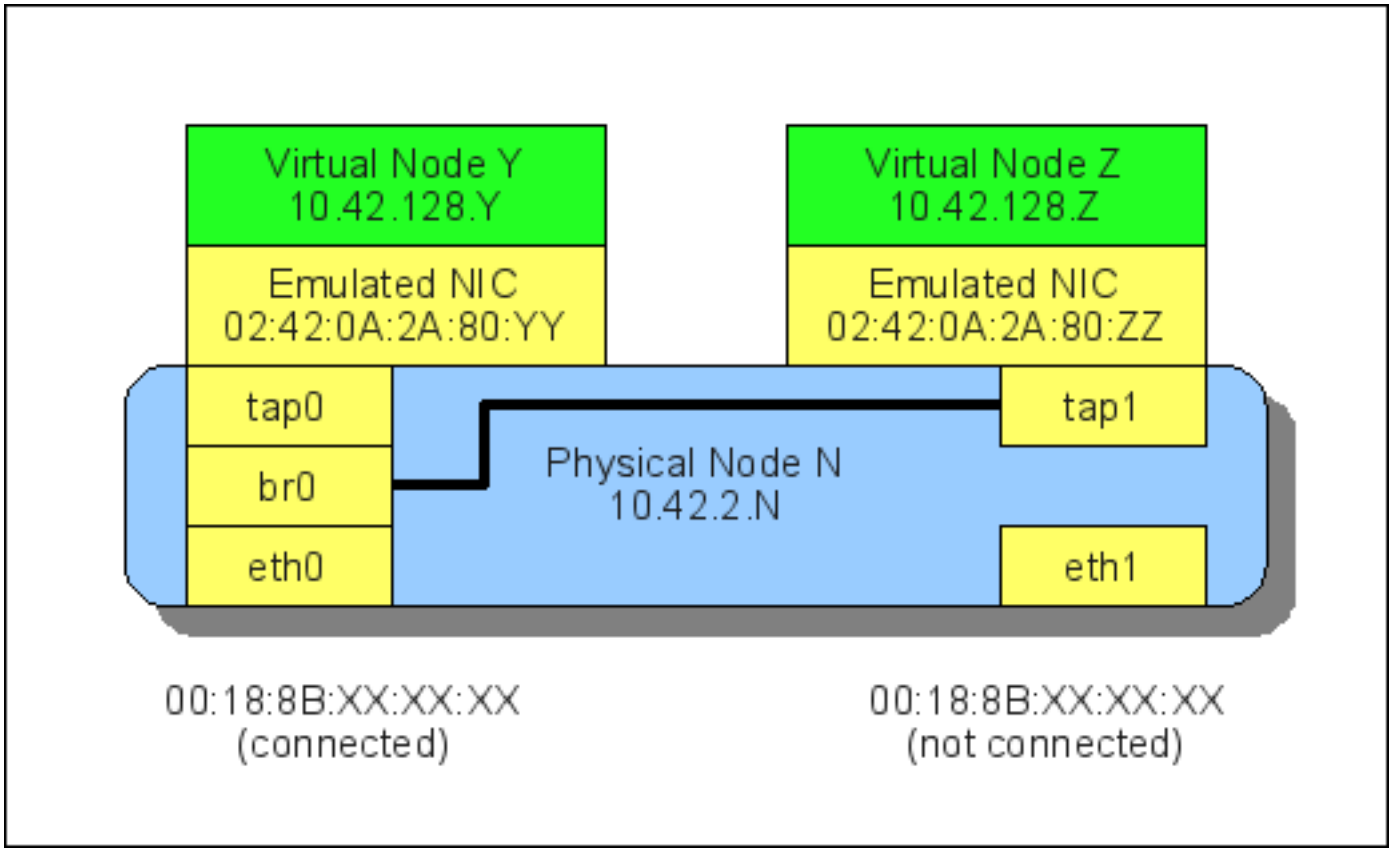
KVM snapshot used to keep base image intact.

Tested scale up to 4,000 VMs (1VM / core)



Preferred Networking Layout

Bridged networking is preferred, but carries security considerations for VM isolation. Can a VM monitor or alter the networking traffic of another VM?



Networking issues

Local administrative policy for VMs requires the use of KVM's userspace networking. The result is each VM is effectively behind its own NAT layer, blocking inbound access.

The cluster itself has a NAT boundary, creating two layers between each VM and outside entities.

Condor GCB could be used but each VM appears as having the same IP (the connection broker's).

An alternative XMPP approach has been investigated for creating an overlay network.

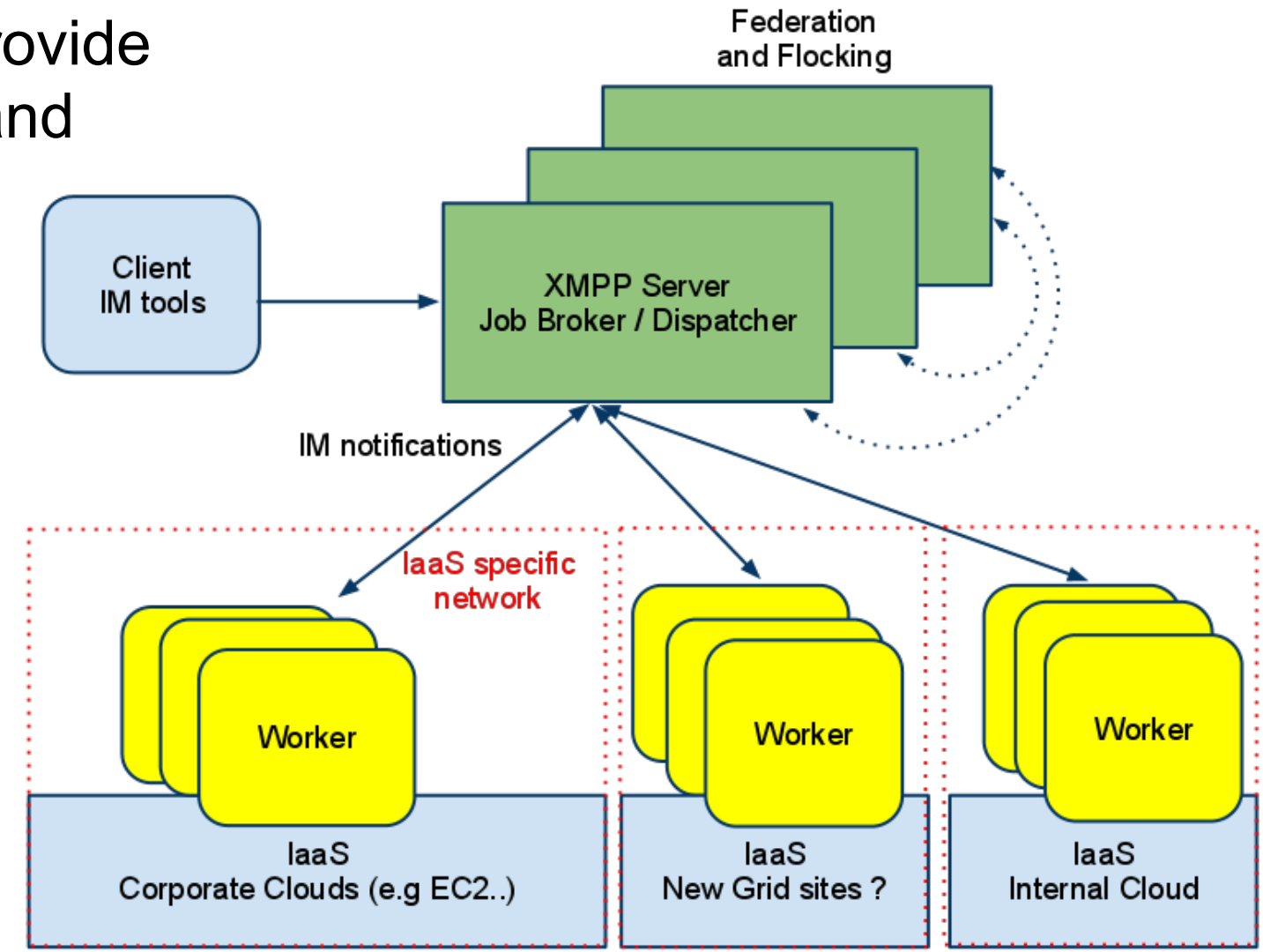
Kestrel basics

- An XMPP-based system for scheduling and dispatching jobs.
- Primarily suited for bag-of-task applications, not MPI.
- XMPP's presence notifications allows for quick response to changes in the worker pool.
- Each worker is a VM with an XMPP client agent that can be tagged with capabilities.
- Workers report to a manager component. Presence notifications update the manager about resource outages or utilization.

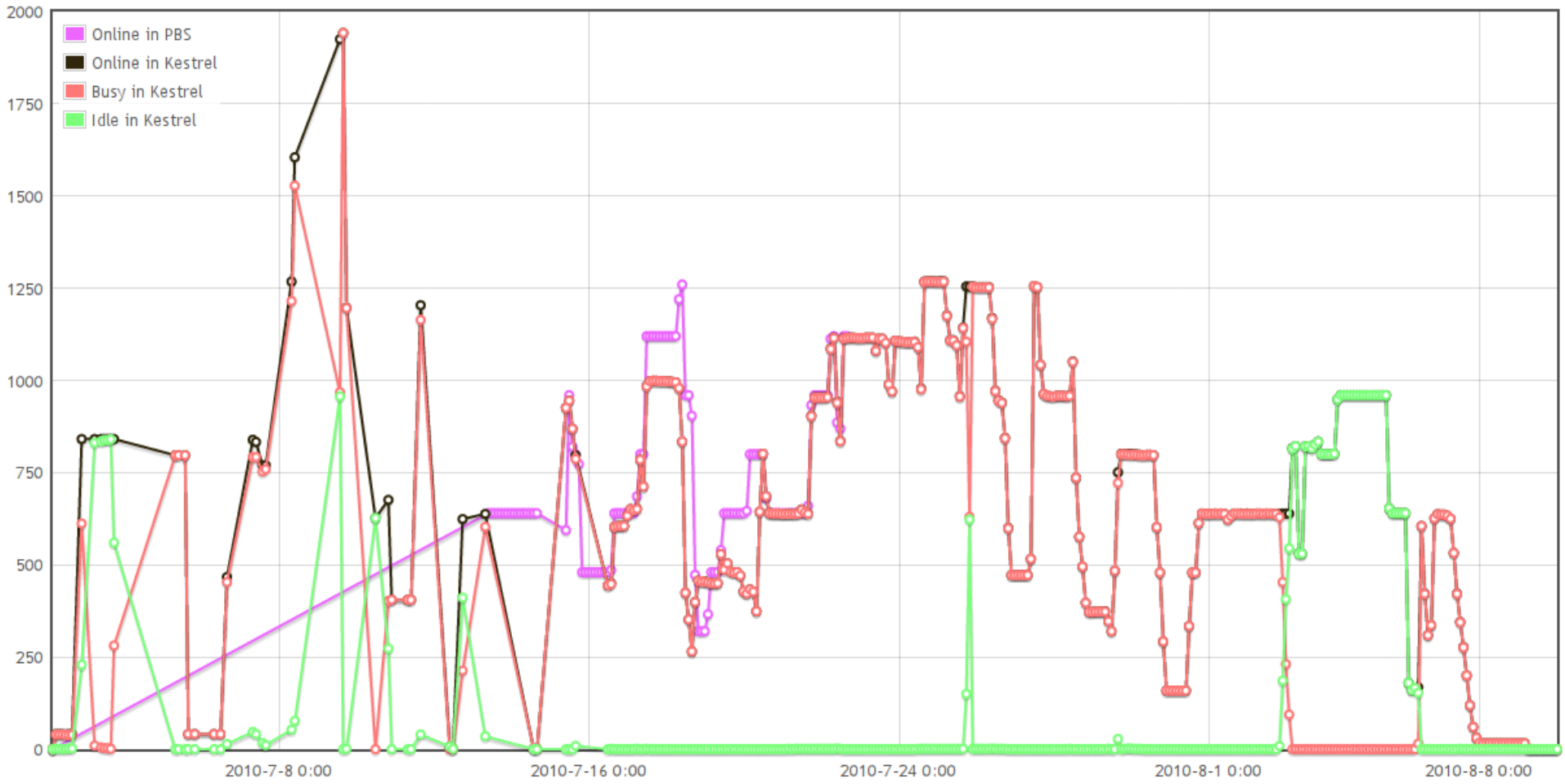
Kestrel Architecture

XMPP and its server implementations provide built-in federation and clustering support.

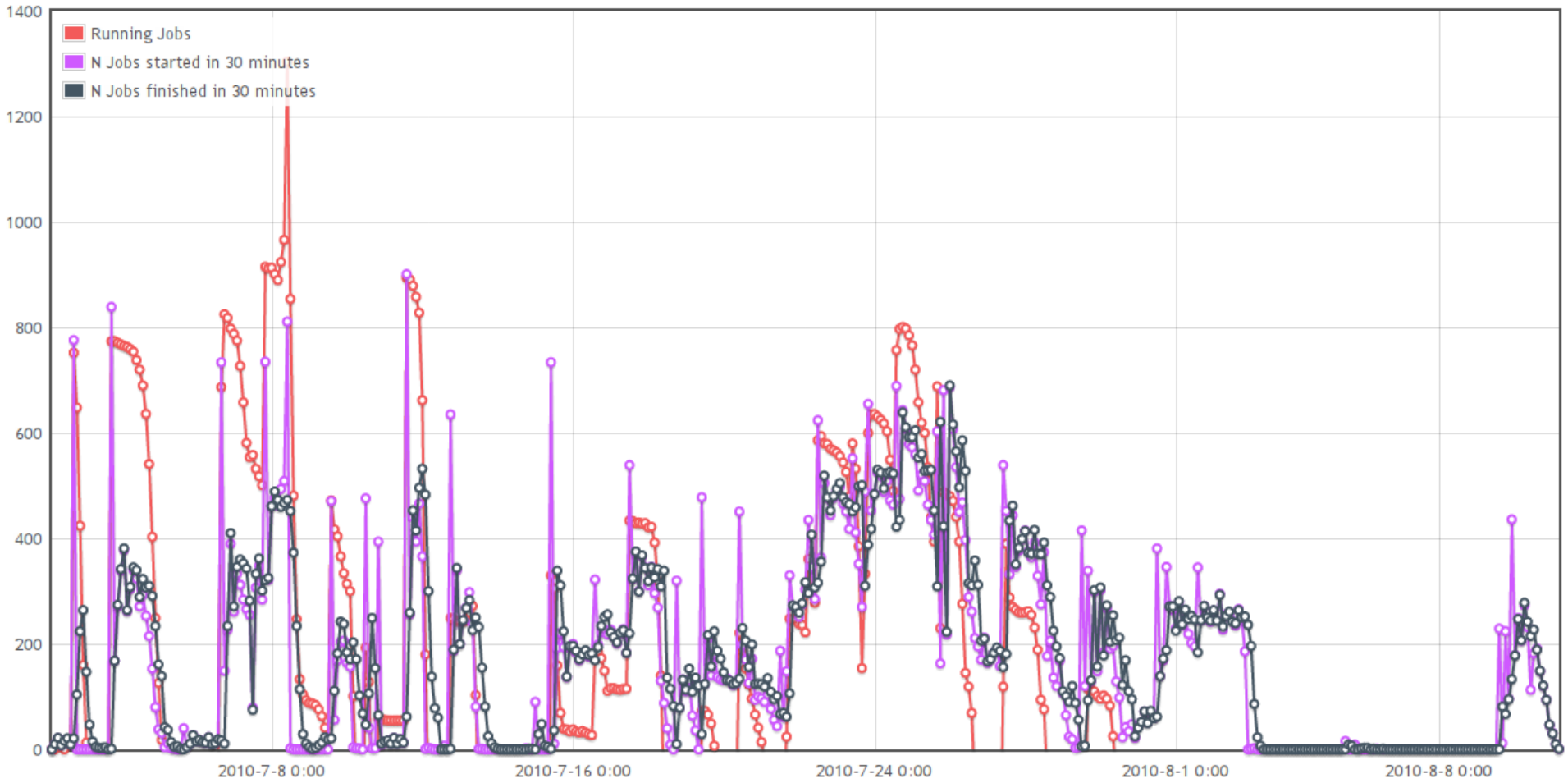
Allows for Kestrel pools to span grid sites.



Use of Kestrel by STAR



Use of Kestrel by STAR



80,000 tasks
400,000 CPU Hours
7TB of generated data

Use of Kestrel by RENCI Bio group

Last year, we have used the VMs on OSG to run the PolyPhen-2 tool for annotating the full human genomic sequence. PolyPhen-2 (Polymorphism Phenotyping v2) is a bioinformatics tool that predicts the possible impact of amino acid substitutions on human proteins. ... we have built a VM, where PolyPhen-2 was installed. In addition, the VM has also install kestrel, which is used to manage the job queening.

In the past year, we have been using this VM several times and our experience has been very satisfying. Our jobs, typically consisting of 150,000 queries that will take us about months, if not years, to run on our workstation, were able to complete within a few days. Needless to say, such a reduction in analysis time is a tremendous help to the researchers.

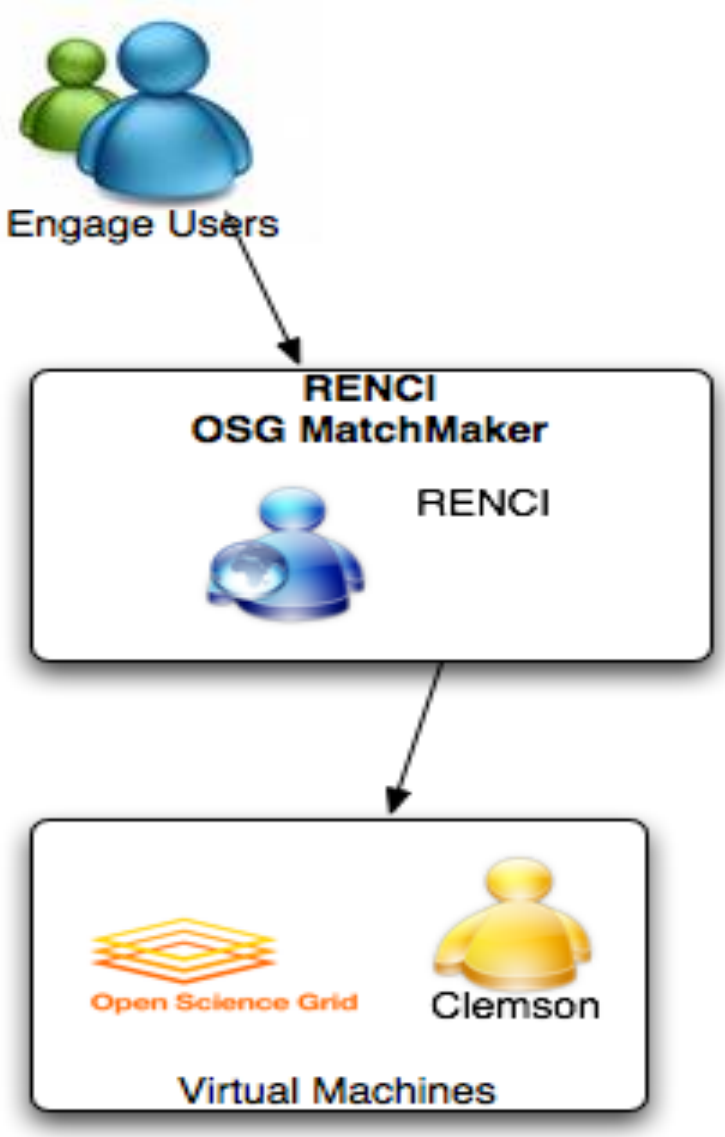
Submitting VMs on OSG using Engage job submission framework

Write bash script that runs the VM

Upload image to Engage RENCI submission node

Submit jobs to VM enabled OSG site using the Engage OSGMM.

VM and script get bundled, staged and executed as any regular job and data dependency



Virtual Machine Image Catalogue (VMIC)

Developed at CERN and DESY within the context of the HEPiX virtualization working group

Define processes to produced and endorse Virtual Machine Images and share these images between sites

Internal VMIC deployed at CERN, Clemson, DESY
Alternative developed at University of Victoria (https://github.com/hep-gc/repoman)

Metadata for virtual machine images being developed by Stratuslab project in EU and by HEPiX.

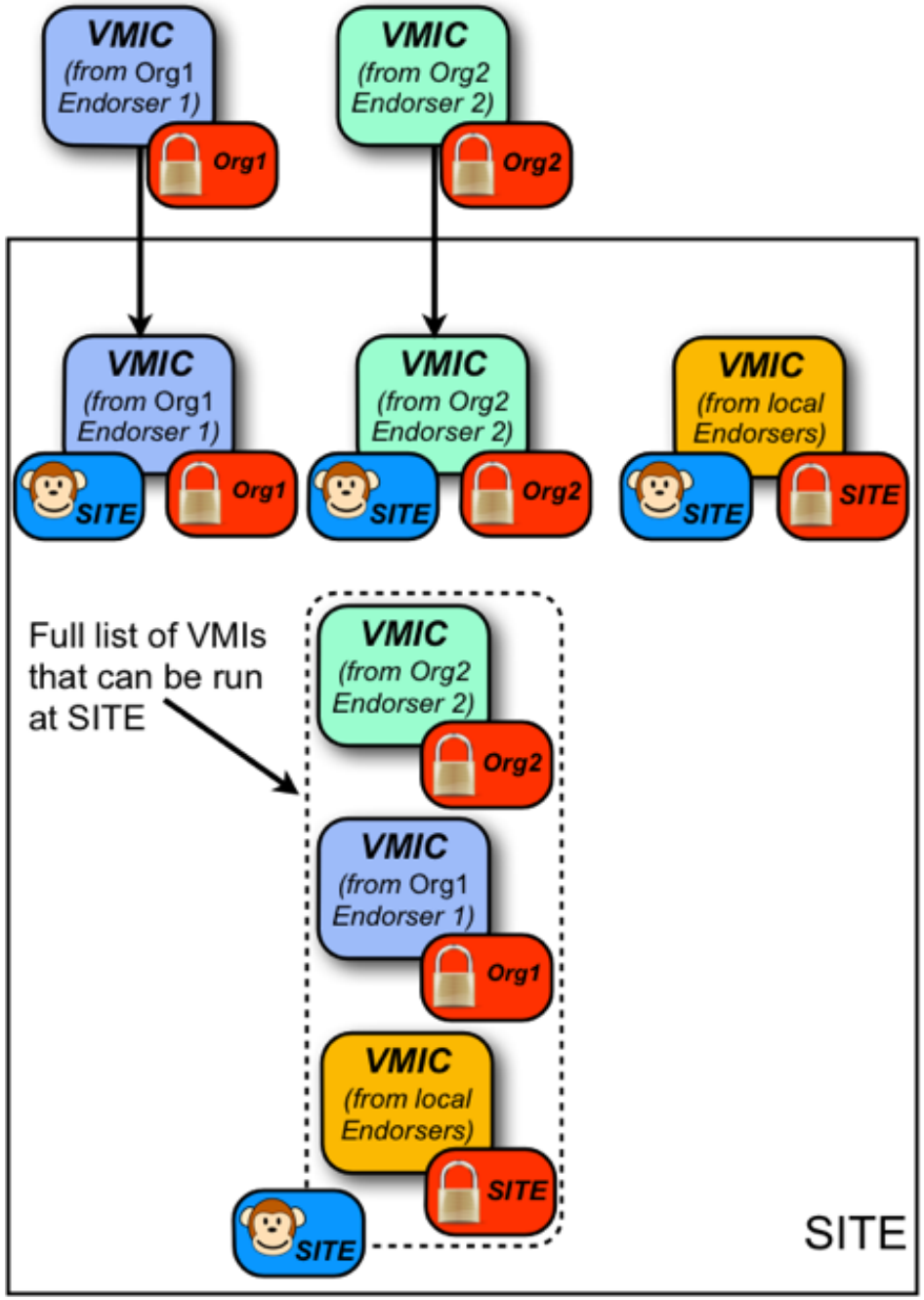
Virtual Machine Image Catalogue

CERN development primarily by Romain Wartel (WLCG).

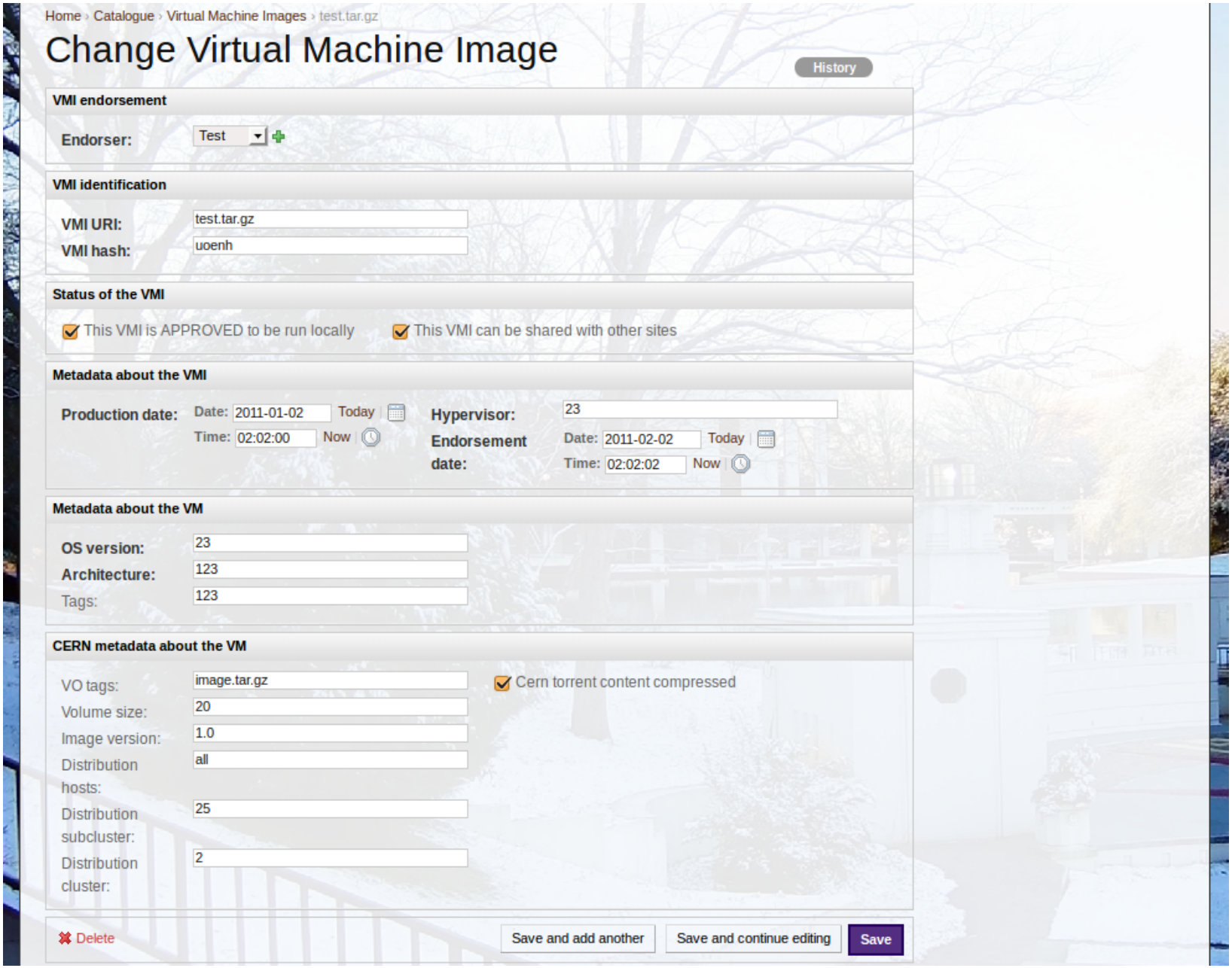
Local catalogue can be built from list of trusted endorsers who publish their own VMIC.

Site keep the control of which VM gets instantiated by approving images.

Catalogue and images digitally signed and contain metadata.



VMIC


Home › Catalogue › Virtual Machine Images › test.tar.gz
Change Virtual Machine Image
History
VMI endorsement
Endorser: Test
VMI identification
VMI URI: test.tar.gz
VMI hash: uoenh
Status of the VMI
This VMI is APPROVED to be run locally
This VMI can be shared with other sites
Metadata about the VMI
Production date: Date: 2011-01-02 Today
Time: 02:02:00 Now
Hypervisor: 23
Endorsement date: Date: 2011-02-02 Today
Time: 02:02:02 Now
Metadata about the VM
OS version: 23
Architecture: 123
Tags: 123
CERN metadata about the VM
VO tags: image.tar.gz
Cern torrent content compressed
Volume size: 20
Image version: 1.0
Distribution hosts: all
Distribution subcluster: 25
Distribution cluster: 2
Delete
Save and add another
Save and continue editing
Save

VMIC


CLEMSON UNIVERSITY
About Photo | Background Off
CLEMSON UNIVERSITY
SEARCH:LEARN
Manage
Authentication
Endorse
Sponsors
Welcome, Loren. Change password / Log out
Home › Endorse
Endorse
The following images will be endorsed on behalf of Loren Klingman :
Name
Volume
Produced
test.tar.gz
23_123_23_123
2011-01-02 02:02:00
The endorser confirms all images meet the criteria defined in the Policy Trusted Virtual Machines.
Actions:
Do not change anything, and go back
Endorse all images!
About
VMIC 1.1
Image Distribution: Wave
Cloud: Palmetto Cluster
Contact
lorenk@clemson.edu
Clemson University
104 McAdams Hall Clemson, SC 29634
Developed By:
Clemson University, School of Computing
CERN
Policy/Terms of Use
Terms of Use
Released Under: ____ License
Copyright © 2011 Clemson University, Clemson, SC. All Rights Reserved.

Overall expected VM workflow on OSG

- VO users upload image to catalogue
- VM images get endorsed by site
- VM get provisioned at sites via Condor-G
- Usage gets reported to Gratia like regular jobs do
- VO either use standard job submission system (shared scheduler at the site) or use overlay scheduler like glideins, Kestrel and the like

Conclusions

- VMs allow for elastic on-demand clusters on top of OSG infrastructure.
- New paradigm of submitting an environment as a job instead of an application.
- Overlay networks can resolve issues caused by limited networking installations.
- Implementing a VM image catalog will increase confidence in preventing malicious images through web of trust mechanisms.

Work was supported by NSF/DOE via CI-TEAM and ExtenCI projects.

Thanks to Mike Murphy, Michael Fenn, Linton Abraham and all the other students…

Thanks to Michael Moore, Eddy Duffy and Barr Von Oehsen