

FermiCloud: Enabling Scientific Workflows with Federation and Interoperability

Steven C. Timm

FermiCloud Project Lead

Grid & Cloud Computing Department

Fermilab

Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359

FermiCloud Background

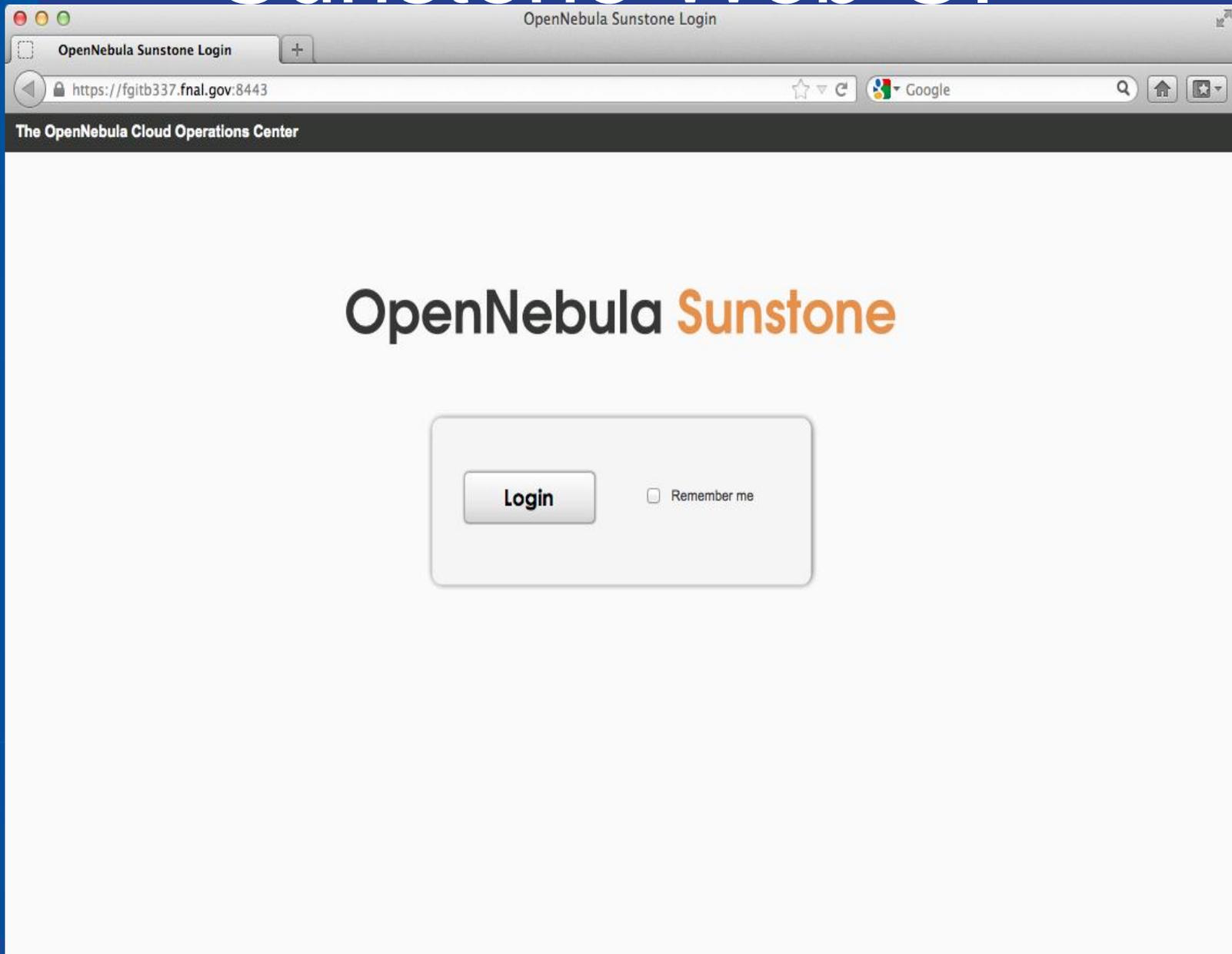
Infrastructure-as-a-service facility for Fermilab employees, users, and collaborators

- Project started in 2010.
- OpenNebula 2.0 cloud available to users since fall 2010.
- Condensed 7 racks of junk machines to 1.5 racks of good machines
- Provider of integration and test machines to the OSG Software team.
- OpenNebula 3.2 cloud up since June 2012
- This talk will focus mostly on current user experience and future directions
- More technical details available on request

Who can use FermiCloud

- Any employee, user, or contractor of Fermilab with a current ID.
- Most OSG staff have been able to get Fermilab “Offsite Only” ID’s.
- With Fermilab ID in hand, request FermiCloud login via Service Desk form.
- Instructions on our new web page at <http://fclweb.fnal.gov>
- Note new web UI at <https://fermicloud.fnal.gov:8443/>
- **Doesn't work with Internet Explorer yet**

Sunstone Web UI



Selecting a template

OpenNebula Sunstone Documentation | Support | Community Welcome timm2 | Sign out

+ New Update properties Instantiate Previous action Delete

Show 10 entries Search:

<input type="checkbox"/> All	ID	Owner	Group	Name	Registration time
<input type="checkbox"/>	50	oneadmin	oneadmin	SLF6 New Base	16:16:25 11/26/2012
<input type="checkbox"/>	53	oneadmin	oneadmin	SCT-slf58	22:31:08 11/28/2012

Showing 1 to 2 of 2 entries First Previous 1 Next Last

Copyright 2002-2012 © OpenNebula Project Leads (OpenNebula.org). All Rights Reserved. OpenNebula 3.2.1

Launching the Virtual Machine

The screenshot displays the OpenNebula Sunstone Cloud Operations Center interface. The browser address bar shows the URL <https://fgitb337.fnal.gov:8443>. The page title is "OpenNebula Sunstone: Cloud Operations Center". The navigation bar includes "OpenNebula Sunstone", "Documentation | Support | Community", and "Welcome hyunwoo | Sign out".

The left sidebar contains the following menu items: Dashboard, Virtual Machines (highlighted), Templates, Virtual Networks, Images, and Configuration.

The main content area shows a table of virtual machines. The table has the following columns: All, ID, Owner, Group, Name, Status, CPU, Memory, Hostname, Start Time, and VNC Access. A single entry is visible:

All	ID	Owner	Group	Name	Status	CPU	Memory	Hostname	Start Time	VNC Access
<input type="checkbox"/>	119	hyunwoo	users	one-119	PROLOG	0	0K	fgitb338	14:04:03 10/16/2012	

Below the table, it says "Showing 1 to 1 of 1 entries". Navigation buttons for "First", "Previous", "1", "Next", and "Last" are present.

At the top right of the main content area, there are buttons for "+ New", "Update properties", "Shutdown", "Previous action", and "Delete".

Copyright 2002-2012 © OpenNebula Project Leads (OpenNebula.org). All Rights Reserved. OpenNebula 3.2.1

Monitoring VM's

OpenNebula Sunstone Documentation | Support | Community Welcome timm2 | Sign out

Dashboard

- Virtual Machines
- Templates
- Virtual Networks
- Images
- Configuration

Summary of resources

VM Templates (total/public)	3 / 0
VM Instances (total/running/failed)	0 / 0 / 0
Virtual Networks (total/public)	1 / 0
Images (total/public)	4 / 0

Quickstart

New:

- VM Template
- VM Instance
- Virtual Network

Image

Historical monitoring information

Total VM count

total active error

Time	total	active	error
13:10:00	28	0	0
13:13:20	28	0	0
13:16:40	28	0	0
13:20:00	28	0	0
13:23:20	28	0	0
13:26:40	28	0	0

Total VM CPU

cpu_usage

Time	cpu_usage
13:10:00	273,000
13:13:20	273,000
13:16:40	273,000
13:20:00	273,000
13:23:20	273,000
13:26:40	273,000

Total VM Memory

mem_usage

Time	mem_usage
13:10:00	64.7G
13:13:20	64.7G
13:16:40	64.7G
13:20:00	64.8G
13:23:20	64.8G
13:26:40	64.8G

VM Network stats

net_tx net_rx

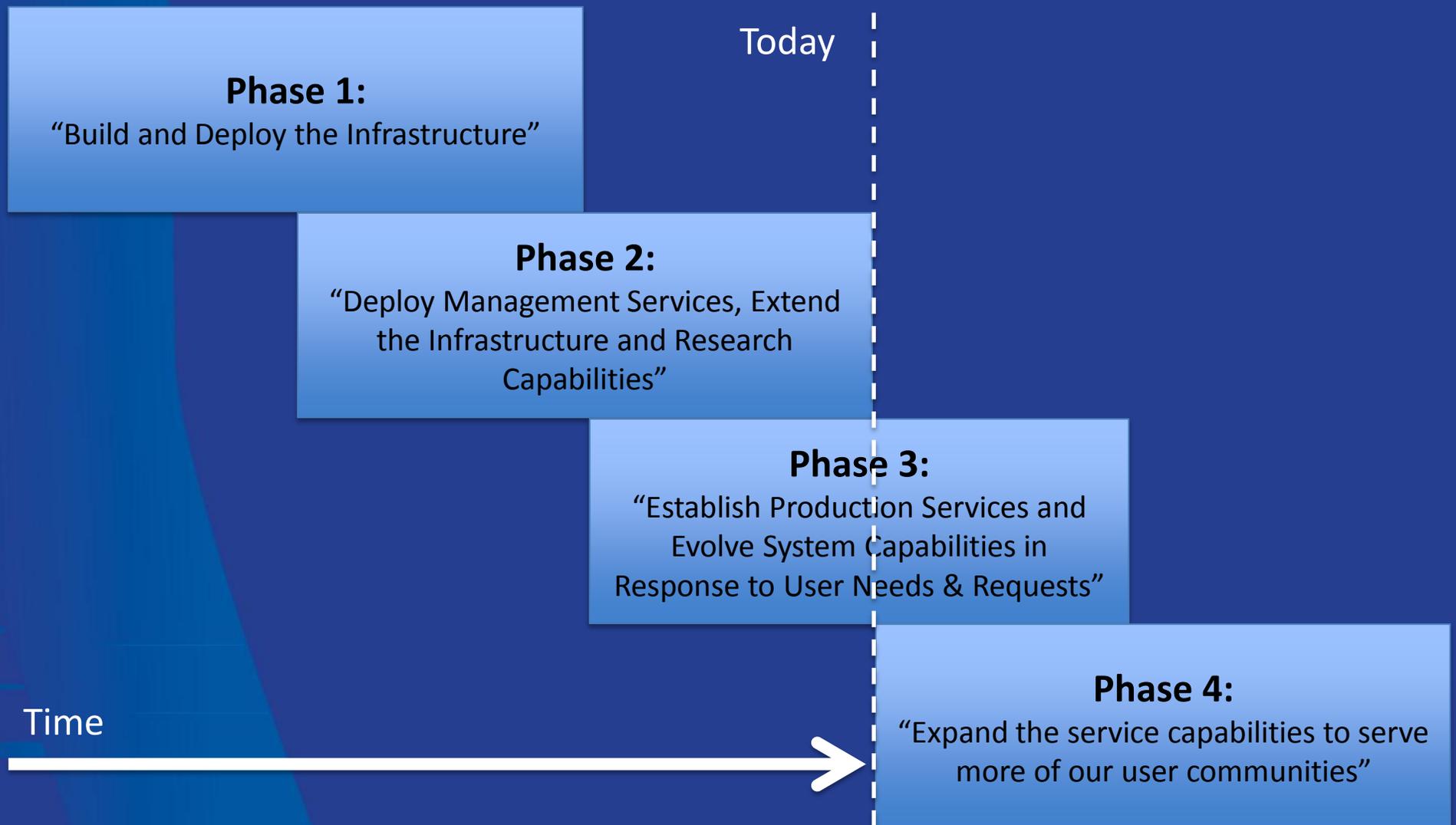
Time	net_tx	net_rx
13:10:00	23.3T	23.3T
13:13:20	23.3T	23.3T
13:16:40	23.3T	23.3T
13:20:00	23.3T	23.3T
13:23:20	23.3T	23.3T
13:26:40	23.3T	23.3T

FermiCloud Development Goals

Goal: Make virtual machine-based workflows practical for scientific users:

- **Cloud bursting:** Send virtual machines from private cloud to commercial cloud if needed
- **Grid bursting:** Expand grid clusters to the cloud based on demand for batch jobs in the queue.
- **Federation:** Let a set of users operate between different clouds
- **Portability:** How to get virtual machines from desktop → FermiCloud → commercial cloud and back.
- **Fabric Studies:** enable access to hardware capabilities via virtualization (100G, Infiniband, ...)

Overlapping Phases



Virtual Machines as Jobs

OpenNebula (and all other open-source IaaS stacks) provide an emulation of Amazon EC2.

Condor team has added code to their “Amazon EC2” universe to support the X.509-authenticated protocol.

Planned use case for GlideinWMS to run Monte Carlo on clouds public and private.

Feature already exists,

- **this is a testing/integration task only.**

Grid Bursting

Seo-Young Noh, KISTI visitor @ FNAL, showed proof-of-principle of “vCluster” in summer 2011:

- Look ahead at Condor batch queue,
- Submit worker node virtual machines of various VO's to FermiCloud or Amazon EC2 based on user demand,
- Machines join grid cluster and run grid jobs from the matching VO.

Need to strengthen proof-of-principle, then make cloud slots available to FermiGrid.

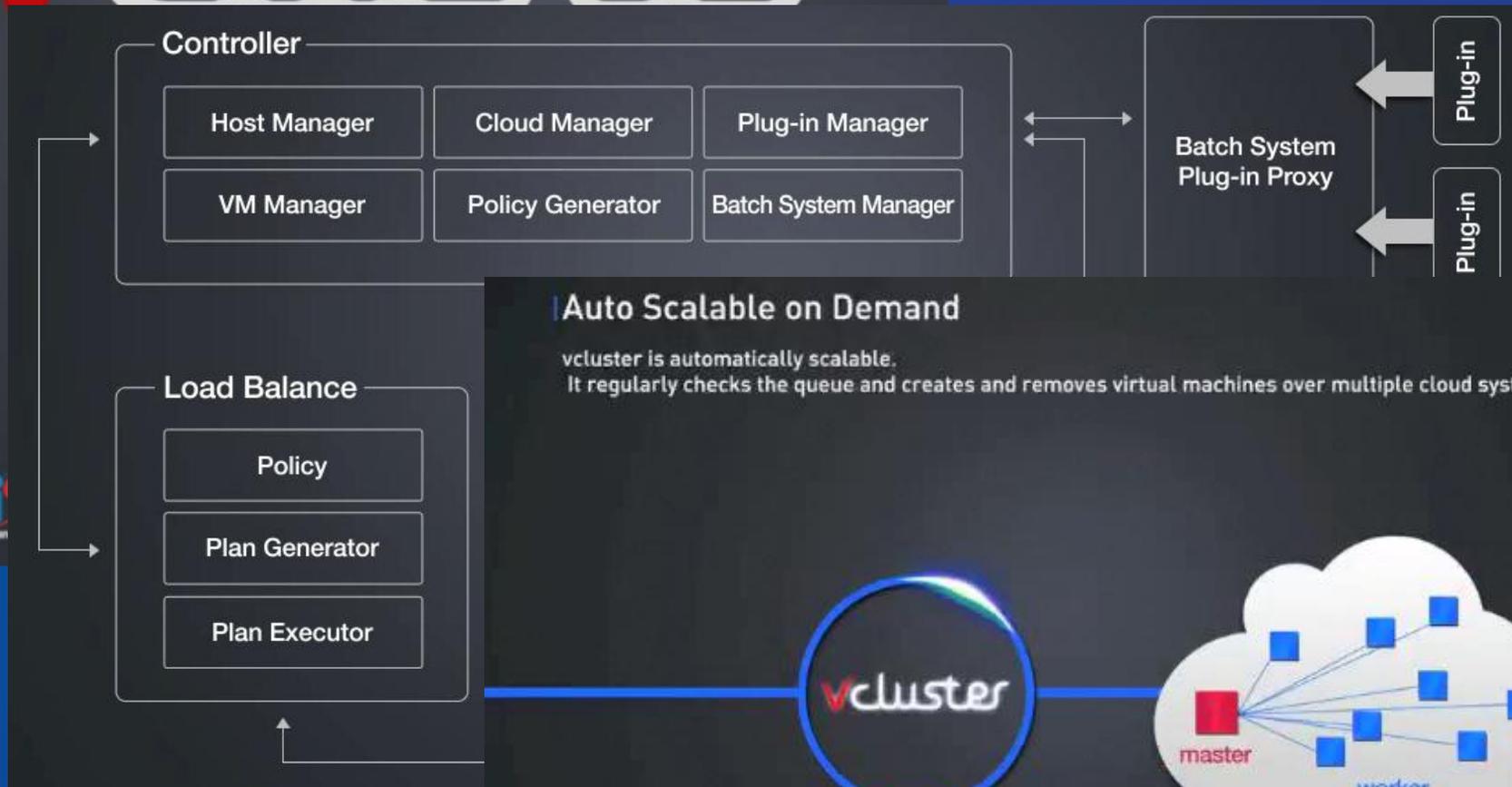
Several other institutions have expressed interest in extending vCluster to other batch systems such as Grid Engine.

KISTI staff has a program of work for the development of vCluster.

GlideinWMS project has significant experience submitting worker node virtual machines to cloud. In discussions to collaborate.

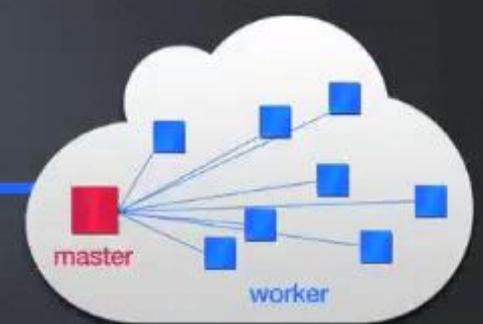
vCluster at SC2012

vcluster



Auto Scalable on Demand

vcluster is automatically scalable. It regularly checks the queue and creates and removes virtual machines over multiple cloud systems.



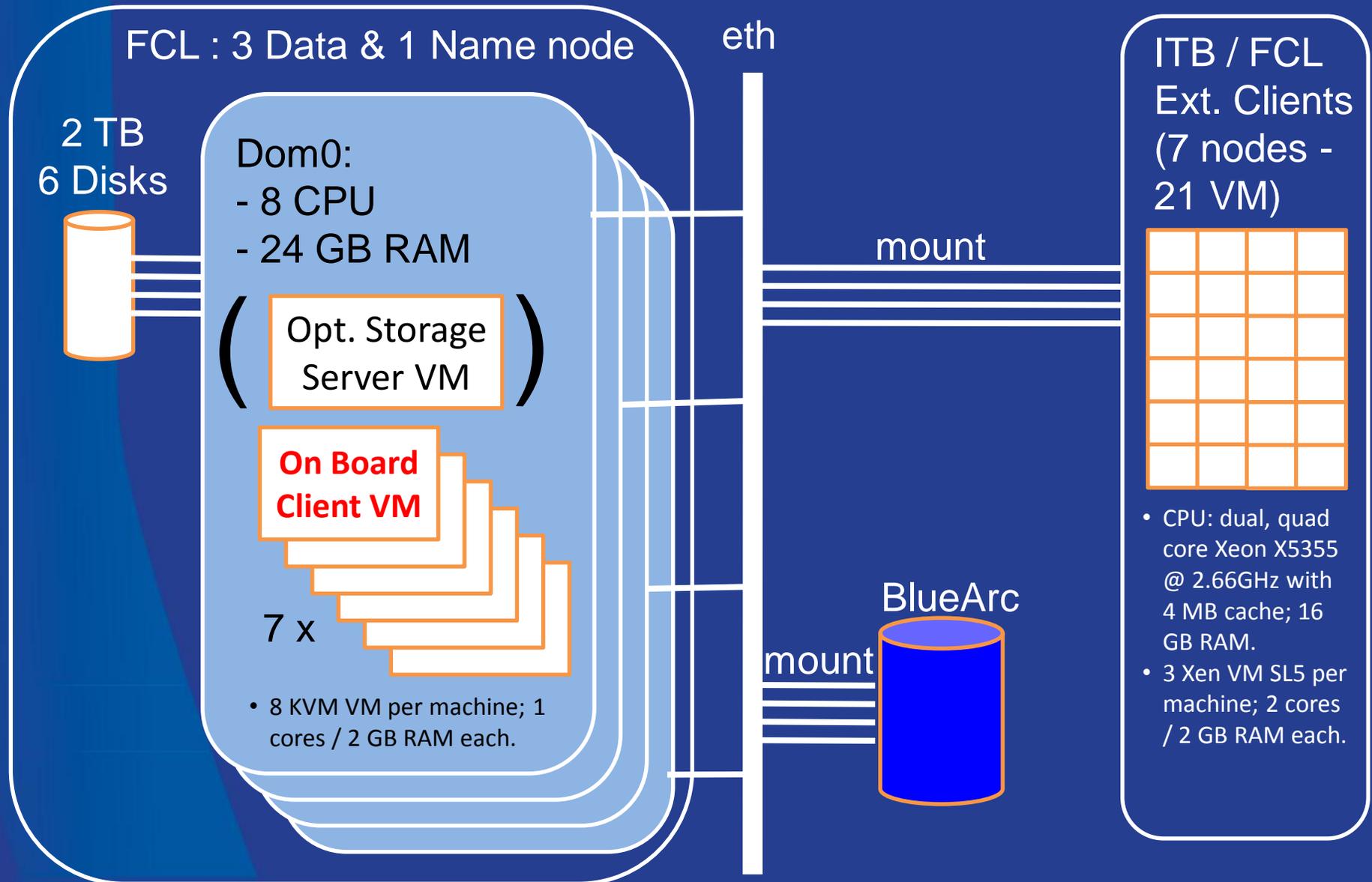
Cloud Bursting

OpenNebula already has built-in “Cloud Bursting” feature to send machines to Amazon EC2 if the OpenNebula private cloud is full.

Need to evaluate/test it, see if it meets our technical and business requirements, or if something else is necessary.

Need to test interoperability against other stacks.

FermiCloud Test Bed - Virtualized Server



True Idle VM Detection

In times of resource need, we want the ability to suspend or “shelve” idle VMs in order to free up resources for higher priority usage.

- This is especially important in the event of constrained resources (e.g. during building or network failure).

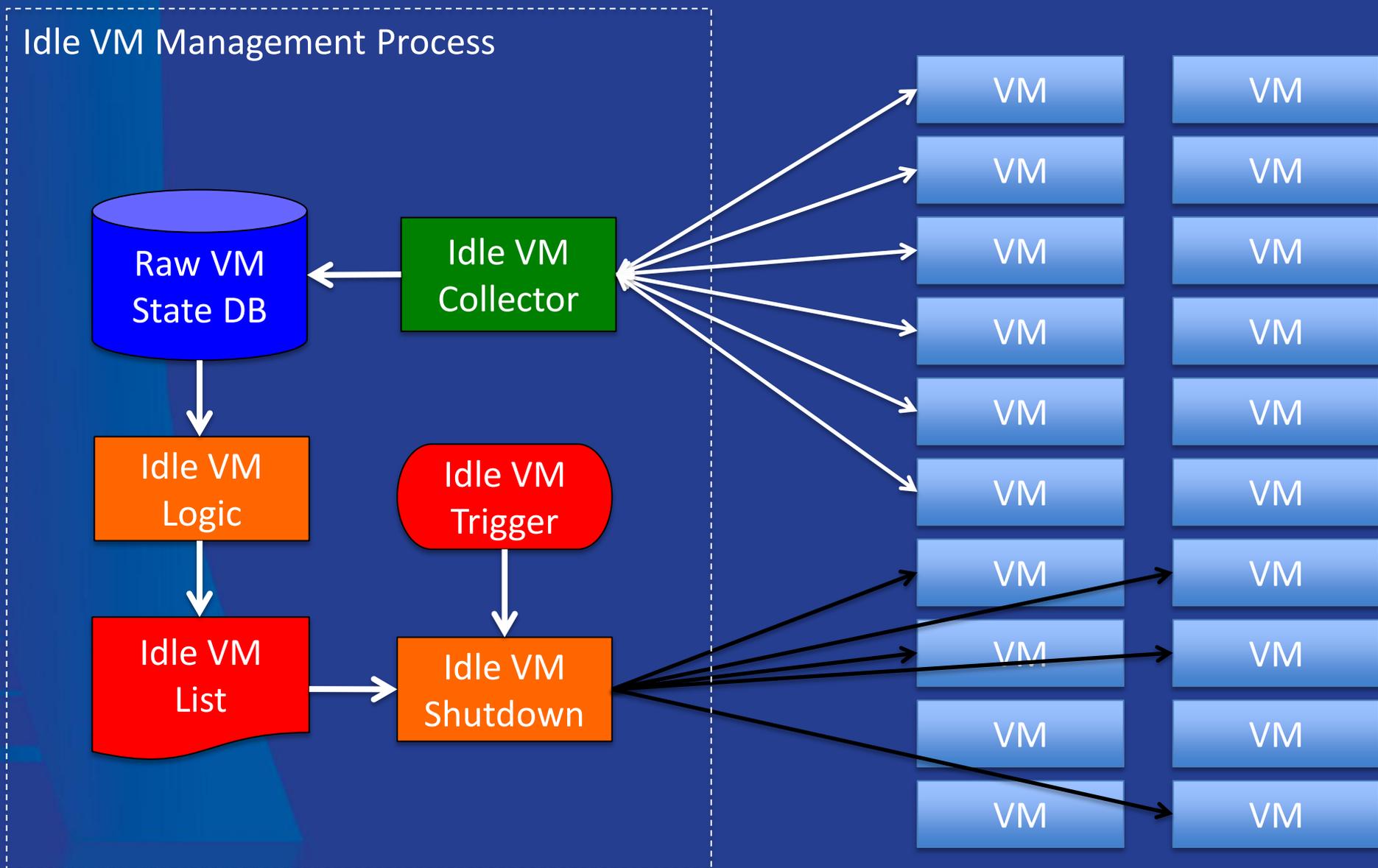
Shelving of “9x5” and “opportunistic” VMs allows us to use FermiCloud resources for Grid worker node VMs during nights and weekends

- This is part of the draft economic model.

Giovanni Franzini (an Italian co-op student) has written (extensible) code for an “Idle VM Probe” that can be used to detect idle virtual machines based on CPU, disk I/O and network I/O.

- This is the biggest pure coding task left in the FermiCloud project,
- If KISTI joint project approved—good candidate for 3-month consultant.

Idle VM Information Flow



Federation

Driver:

- Global scientific collaborations such as LHC experiments will have to interoperate across facilities with heterogeneous cloud infrastructure.

European efforts:

- EGI Cloud Federation Task Force – several institutional clouds (OpenNebula, OpenStack, StratusLab).
- HelixNebula—Federation of commercial cloud providers

Our goals:

- Show proof of principle—Federation including FermiCloud + KISTI “G Cloud” + one or more commercial cloud providers + other research institution community clouds if possible.
- Participate in existing federations if possible.

Core Competency:

- FermiCloud project can contribute to these cloud federations given our expertise in X.509 Authentication and Authorization, and our long experience in grid federation

Virtual Image Formats and Distribution

Clouds have different VM image formats:

- **FS, Partition table, LVM , Kernel**

Identify differences, find conversion tools

Investigate image marketplaces (Hepix, UVic)

Do we need S3 or Gridftp image upload facility?

- **OpenNebula doesn't have one now.**

Develop auto security scan for VM images

- **Scan them like laptop coming onto site.**

Interoperability/Compatibility of API's

Amazon EC2 API is not open source, it is a moving target that changes frequently.

Open-source emulations have various feature levels and accuracy of implementation:

- Compare and contrast OpenNebula, OpenStack, and commercial clouds,
- Identify lowest common denominator(s) that work on all.
- Contribute bug reports and fixes where possible.

High-Throughput Fabric Virtualization

Follow up earlier virtualized MPI work:

- Use it in real scientific workflows
- Example – simulation of data acquisition systems (the existing FermiCloud Infiniband fabric has already been used for such).

Will also use FermiCloud machines on 100Gbit Ethernet test bed

- Evaluate / optimize virtualization of 10G NIC for the use case of HEP data management applications
- Compare and contrast against Infiniband

FermiCloud X.509 Authentication

OpenNebula came with “pluggable” authentication, but few plugins initially available.

OpenNebula 2.0 Web services by default used “access key” / “secret key” mechanism similar to Amazon EC2. No https available.

Four ways to access OpenNebula:

- Command line tools,
- Sunstone Web GUI,
- “ECONE” web service emulation of Amazon Restful (Query) API,
- OCCl web service.

FermiCloud project wrote X.509-based authentication plugins:

- Available in OpenNebula 3.2 and onward.
- X.509 plugins available for command line and for web services authentication

FermiCloud X.509 Authorization

- OpenNebula authorization plugins written in Ruby
- As shipped: One DN -> one userid—no VO support
 - For now, must use Fermilab KCA short-lived cert.
- Use existing Grid AuthZ routines:
 - Use Ruby-Java bridge to call Java-based routines from VO Privilege project
 - GUMS returns uid/gid, SAZ returns yes/no.
 - Works with OpenNebula command line and non-interactive web services
- Almost impossible to send VOMS proxy through web browser
 - For OpenNebula's Sunstone Web GUI develop VO selection pull-down (similar to Gratia Admin Web UI)
 - Get desired vo/role from user and poll GUMS/VOMS on server side

FermiCloud User Comments1:

- Bothered by no-root-squash NFS (Mat)
 - FermiCloud stores all host/http certs in volatile RAM disk for just that reason.
- Clock problem with pause/resume (Mat)
 - FC admins solve that by launching a process to restart ntpd on a loop just before we pause. Will try to make this available to users.
- Way for user to intervene on down VM (several)
 - We as admins can get serial console or access via virt-viewer
- Can't save VM snapshot (several)
 - `Onevm saveas <vmid> 0 <name_of_image>` ; `onevm shutdown <vmid>`
- IP finding tool needed (several)
 - `source /cloud/images/OpenNebula/scripts/one3.2/hostname.sh`
 - Other OSG software team have written scripts too

FermiCloud User Comments 2

- Change name of VM (several)
 - Have asked OpenNebula for this feature.
- Community repo for users to share scripts(Marco)
 - OSG SW team already added
- Better docs on best practices (shutdown) (Tim)
 - Next version of OpenNebula expected to beat a lot of race conditions
 - Until then—we will modify onevm delete so it will not delete a VM still in “shut” or “epil” state
 - And add an “are you sure?” prompt for VM in “run” state.
- Make docs available as HTML pages, not word docs (Tim)
 - See <http://fclweb.fnal.gov/fermicloud-dummies.html>
 - And <http://fclweb.fnal.gov/fermicloud-geeks.html>

Acknowledgements

None of this work could have been accomplished without:

- FermiCloud Project Team: Currently Steven Timm (lead), Keith Chadwick, Gerard Bernabeu, Karen Shepelak, Hyunwoo Kim, Parag Mhashilkar, Tanya Levshina, Gabriele Garzoglio
- The excellent support from other departments of the Fermilab Computing Sector – including Computing Facilities, Site Networking, and Logistics.
- The excellent collaboration with the open source communities – especially Scientific Linux and OpenNebula,
- As well as the excellent collaboration and contributions from KISTI.

FermiCloud Summary

FermiCloud Development Collaboration:

- Leveraging external work as much as possible,
- Contribution of our work back to external collaborations.
- Using (and if necessary extending) existing standards:
- AuthZ, OGF UR, Gratia, etc.

FermiCloud Facility

- Deploying 24by7 capabilities, redundancy and HA.
- Delivering support for science collaborations at Fermilab
- Making new types of computing work possible

The future is mostly cloudy.