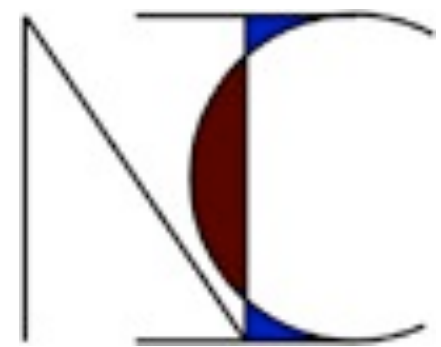




UNIVERSITY OF OREGON

Building a Scientific Cloud Computer with OpenStack



Chris Hoge
Neuroinformatics Center
hoge@uoregon.edu





ACISS Grant - May 1, 2010 through April 30, 2013

\$1.97 million proposal to National Science Foundation (NSF).

Major Research Instrumentation (MRI) grant.

Awarded through the Office of Cyber Infrastructure (OCI).

Funded by the American Recovery and Reinvestment Act (ARRI).





Applied Computational Instrument for Scientific Synthesis (ACISS)

ACISS Objectives:

Improve the productivity of HPC research and administration.

Increase computing access to the research community.

Initiate a new approach to computing at the University of Oregon.



Applied Computational Instrument for Scientific Synthesis (ACISS)

Project Strategy:

Heterogeneous cluster to address a wide range of needs.

Offer a traditional HPC cluster to get off the ground quickly.

Adopt a private cloud to maximize flexibility.



----- /usr/local/packages/Modules/modulefiles -----				
FastQC/v0.10.0	eclipse/eclipse-juno	infernai/1.0.2	mpi/openmpi-1.5.3_gcc-4.4.6	python/2.7.2(default)
FastQ_Screen/v0.3.1	emboss/6.4.0	intel/12.0.4(default)	mpi/openmpi-1.7-alpha_intel-12.1.4	qiime/1.3.0
Mathematica/8.0(default)	fasttree/2.1.3	intel/12.1.0-mpi	mpi-tor/openmpi-1.4.5_gcc-4.4.6	qiime/1.4.0
PANDAsq	fasttree/2.1.4	intel/12.1.4	mpi-tor/openmpi-1.5.4_gcc-4.5.3	qiime/1.5.0(default)
R/2.13.2	fastx_toolkit/0.0.13(default)	iprscan/4.8(default)	mpi-tor/openmpi-1.5.4_intel-12.0.4	raxml/7.2.8
R/2.14.2(default)	fftw/2.1.5	jacket/32-bit	mpi-tor/openmpi-1.5.4_pgi-11.10	rmblast/1.2-ncbi-blast-2.2.23+
SPECFEM/2D-6.1.5	fftw/2.1.5-omp	jacket/64-bit	mpi-tor/openmpi-1.5.5_intel-12.1.4	rnnotator/2.4.12(default)
SPECFEM/3D_V2.0.1	fftw/3.2.2	java/1.5.0_22	mpi-tor/openmpi-1.5.5_pgi-12.5	ruby/1.9.3(default)
allpathslg/41370(default)	fftw/3.2.2-omp	java/1.6	mpi-tor/openmpi-1.6.0_gcc-4.6.3	samtools/0.1.18(default)
bedtools/2.14.3	flash/1.0.3	java/1.6.0_31	mrconvert/2.0-r235	stacks/0.998
binutils/2.22	fossil/1.20	java/1.7	mrconvert/20111111	stacks/0.999
blast/2.2.25	fsl/fsl-4.1.9(default)	java/1.7.0	muscle/3.8.31	stacks/0.9993(default)
blast/2.2.25+	gcc/4.3.6	java/1.7.0_5	nagware/5.3.836	swift/0.93
blast/2.2.26+(default)	gcc/4.5.3	jruby/1.6.7.2(default)	nagware/5.3.840	tagdust/1.12(default)
blat/130	gcc/4.6.3	lammps/120210(default)	nagware/5.3.842(default)	tau/2.21.2
boost/1.48(default)	gmap-gsnap/2011-12-28	lastz/1.02.00	netcdf/netcdf-4.2_gcc-4.5.3	tophat/1.4.1(default)
bowtie/0.12.7	gmap-gsnap/2012-04-05(default)	matlab/r2011b(default)	netcdf/netcdf-4.2_intel-12.1.4	tophat/2.0.3
bowtie/0.12.8(default)	gmp/5.0.2	matlab/r2011b_gcc	null	tophat/2.0.4
bowtie/2.0.0-beta5	gmp/5.0.5	mercurial/2.2.2(default)	numpy/1.6.1(default)	trf/4.04
bowtie/2.0.0-beta6	grace/5.1.22(default)	mkl/12.0	nwchem/6.0(default)	uclust/1.2.22
candis/candis-5.10_gcc-4.5.3	graphviz/2.28.0-1	mkl/12.1	oases/0.2.07(default)	ucsc_utils/130
clearcut/1.0.9	gromacs/4.5.5(default)	module-cvs	papi/4.1.3	use.own
clustalw/2.1(default)	gromacs/4.5.5_sngl	module-info	papi/4.2.1	usearch/5.2.32
cmake/2.8.6	hdf5/hdf5-1.8.9_gcc-4.5.3	modules	paraview/3.14.1	velvet/1.1.06-93mer-omp
comsol/4.3	hdf5/hdf5-1.8.9_intel-12.1.4	mothur/1.22.2(default)	perl/5.14.2(default)	velvet/1.2.05
cuda/4.0(default)	hmmer/3.0(default)	mpc/0.9	perl/5.16.0	velvet/1.2.07(default)
cuda/4.1	hmpp/2.4.0	mpc/0.9-2	pgi/11(default)	wpp/2.1.5
cufflinks/1.3.0(default)	hmpp/2.5.2	mpfr/3.0.1	pgi/11.10	
cufflinks/2.0.0	hmpp/3.0.5	mpfr/3.1.0	pgi/12	
cytoscape/2.8.2(default)	hmpp/3.1.0	mpi/intel-4.0.3	pgi/12.5	
dot	hpupc/3.5	mpi/openmpi-1.4.5_gcc-4.4.6	pvcogent/1.5.1	

Our software support list
(as of a month ago)

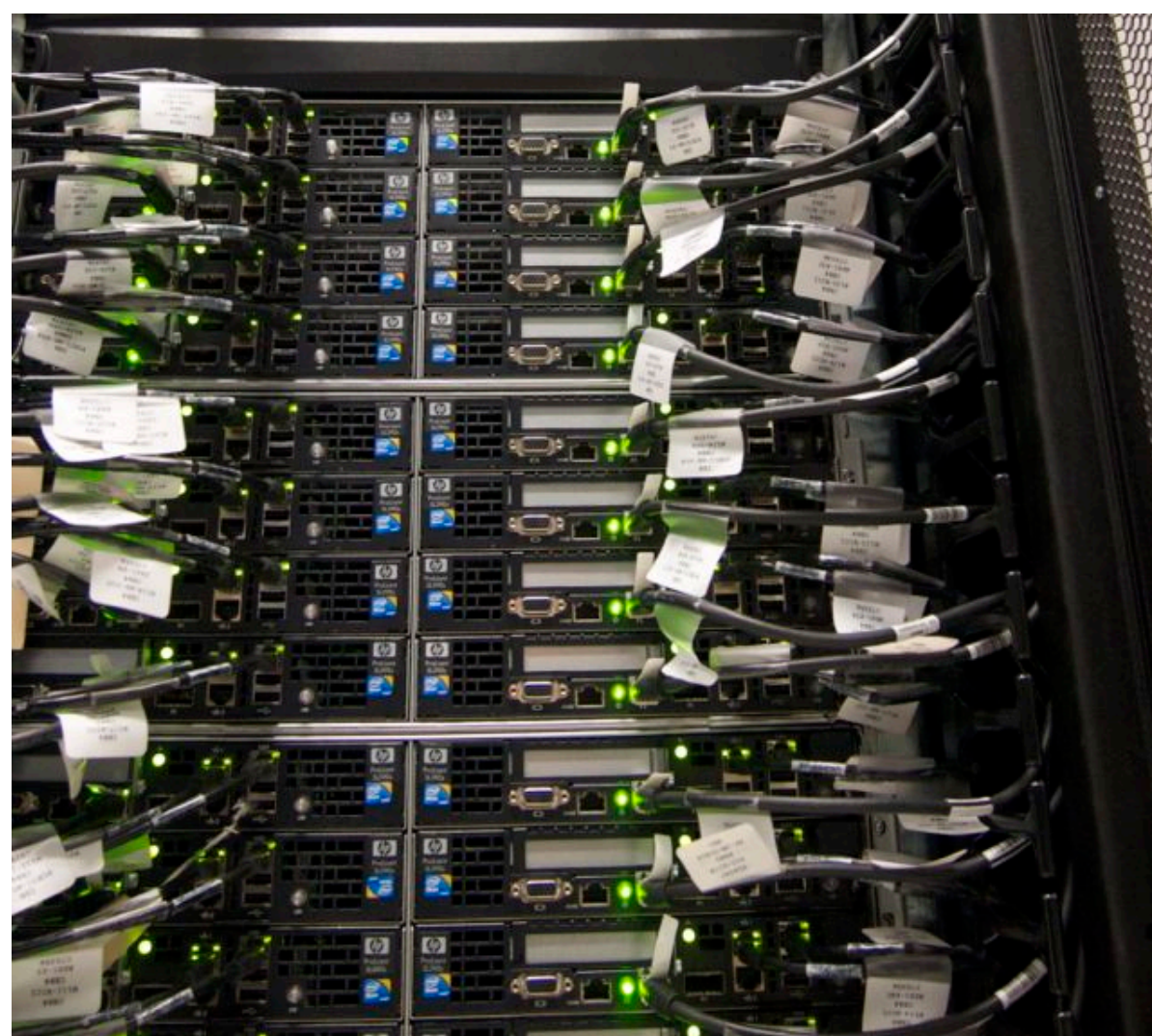
Basic Nodes

128 HP ProLiant SL390 G7.

Two Intel X5650 2.66 GHz 6-core CPUs per node.

1,536 total cores.

72GB DDR3 RAM per basic node.



Fat Nodes

16 HP ProLiant DL 580 G7.

Four Intel X7560 2.266 GHz 8-core CPUs per node.

512 total cores.

384GB DDR3 RAM per fat node.



GPU Nodes

52 HP ProLiant SL390 G7.

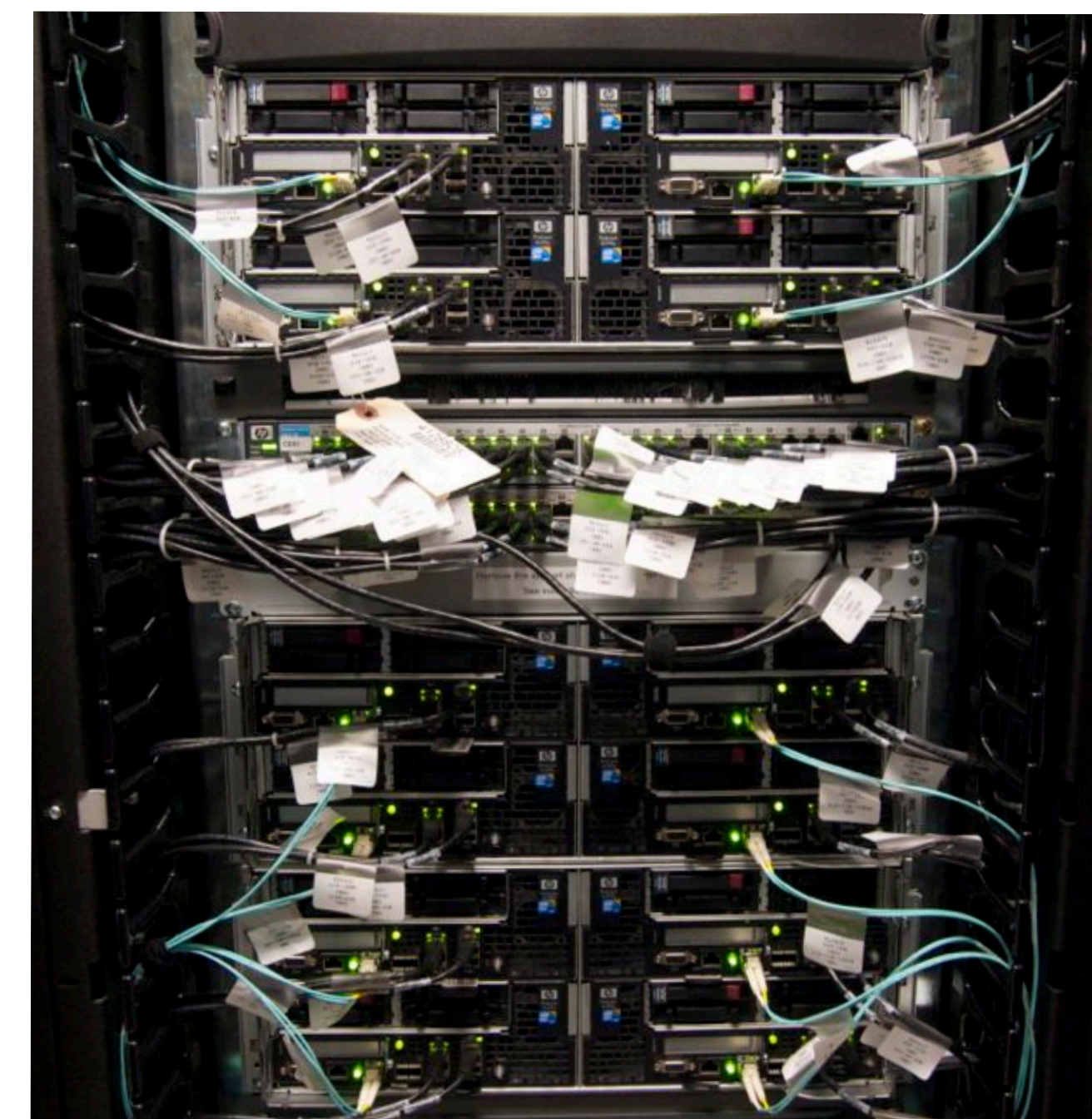
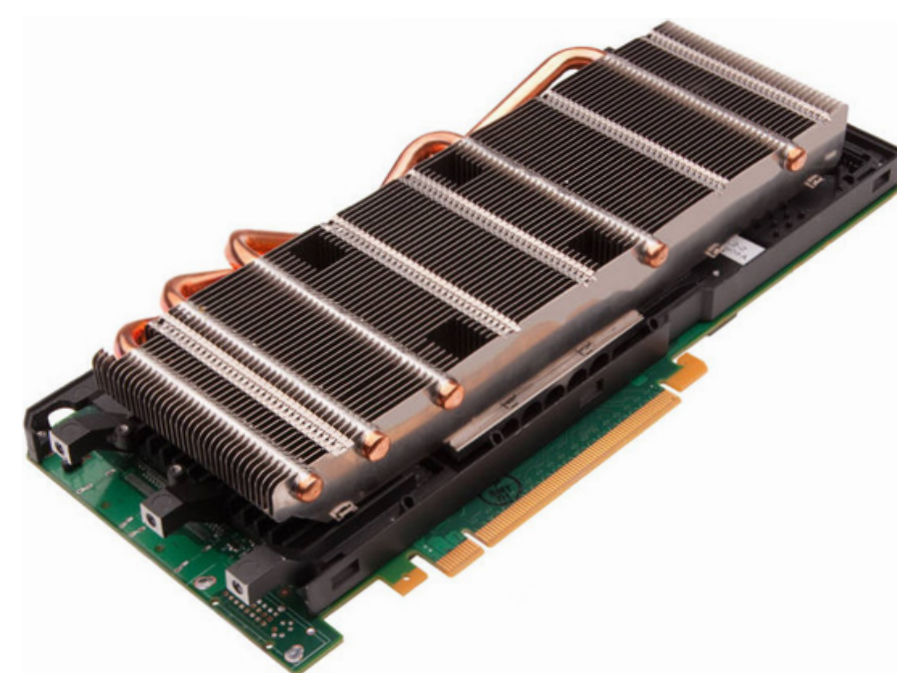
Two Intel X5650 2.66 GHz 6-core CPUs per node.

624 total cores.

72GB DDR3 per GPU node.

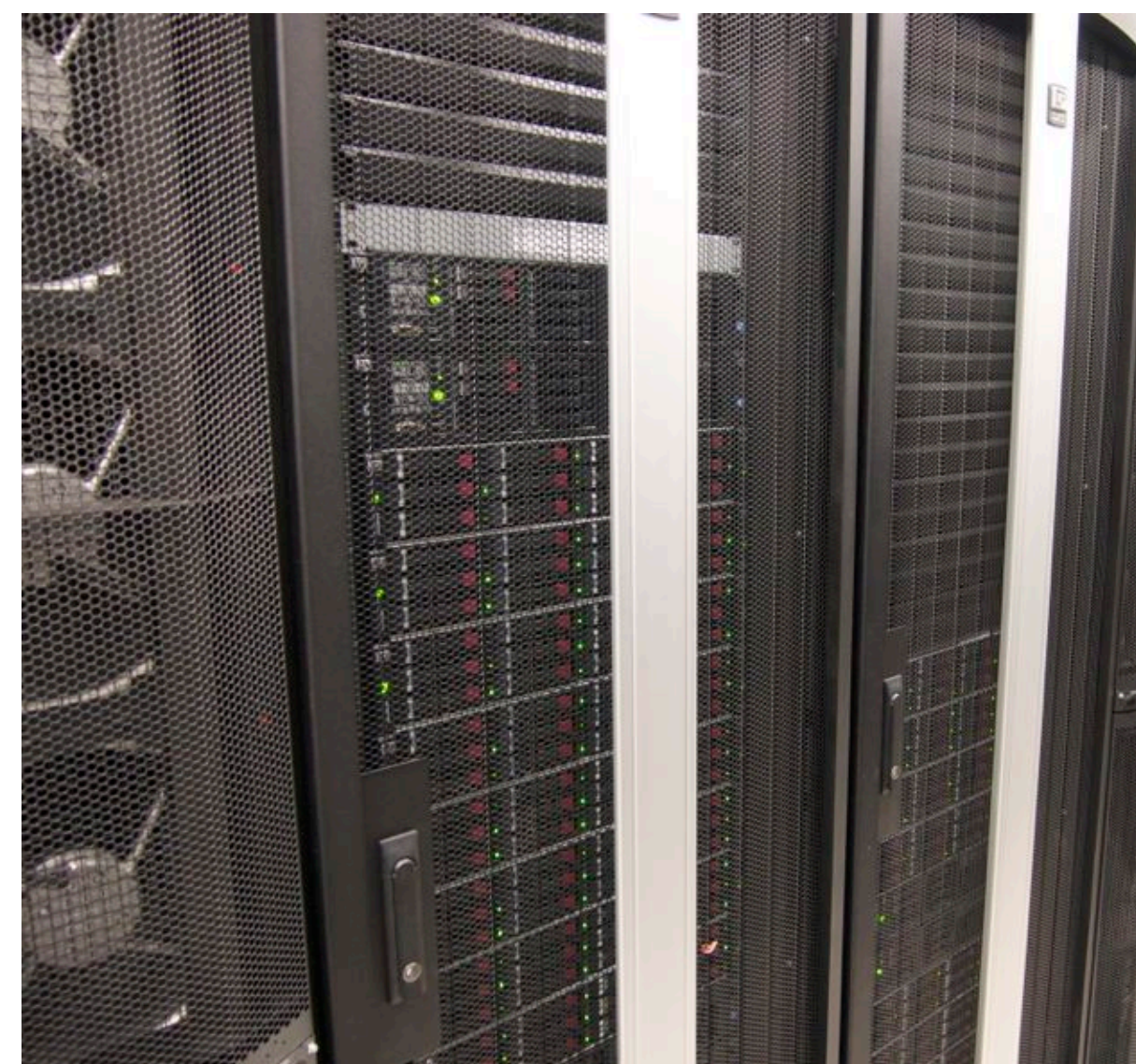
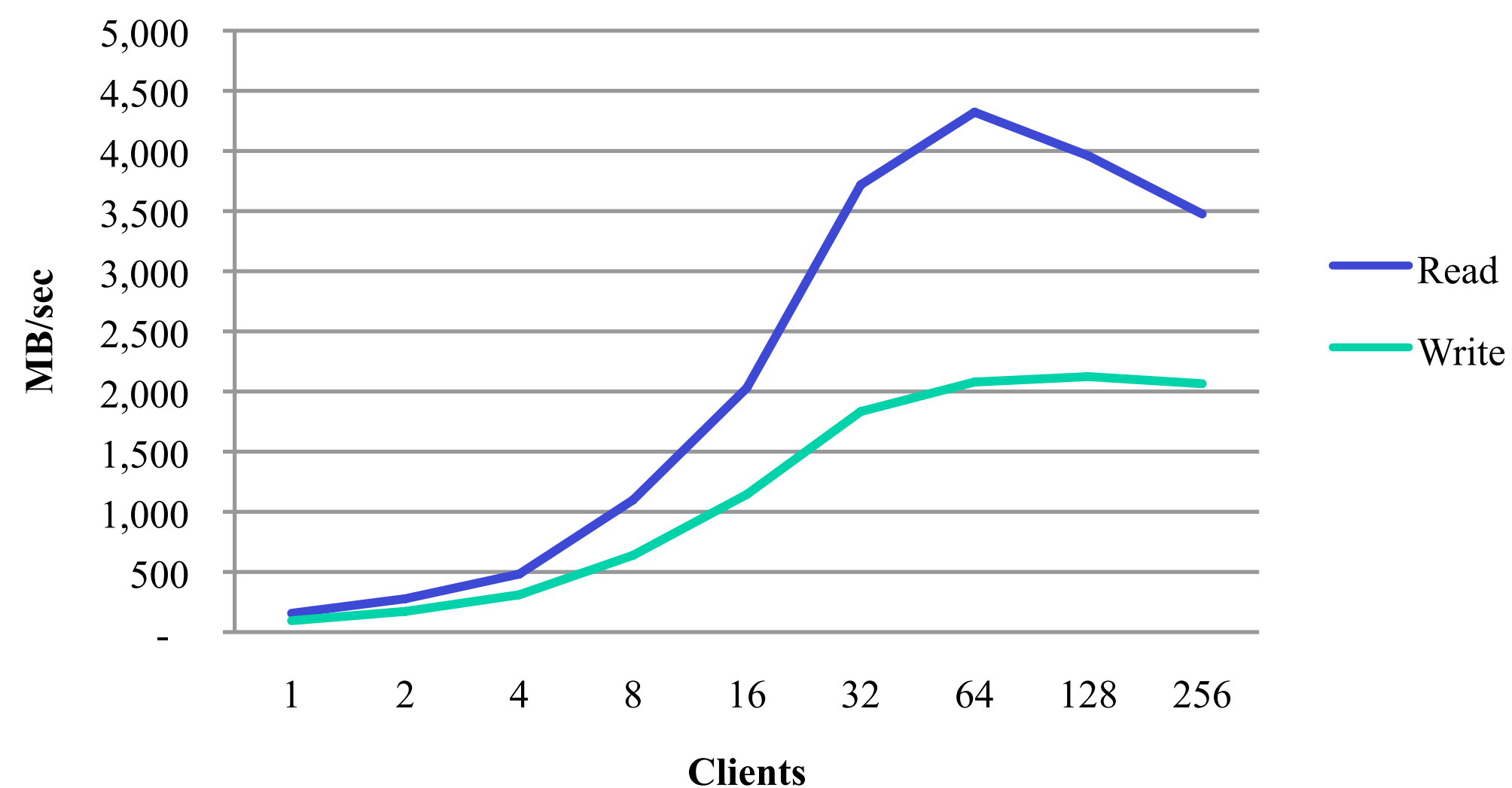
3 NVidia M2070 GPUs per node.

156 total GPUS.



Parallel Network Attached Storage
HP Ibrix X9320 Storage Array.
4 couplets connected to the 10 GigE network.
400 TB useable storage.
Auto mounted as home directory to cluster users.

4x X9320 7.2K 2TB ML-SAS RAID6



Networking

Voltaire Vantage 8500 10 GigE ethernet switch.

1:1 non-blocking 10 GigE networking.

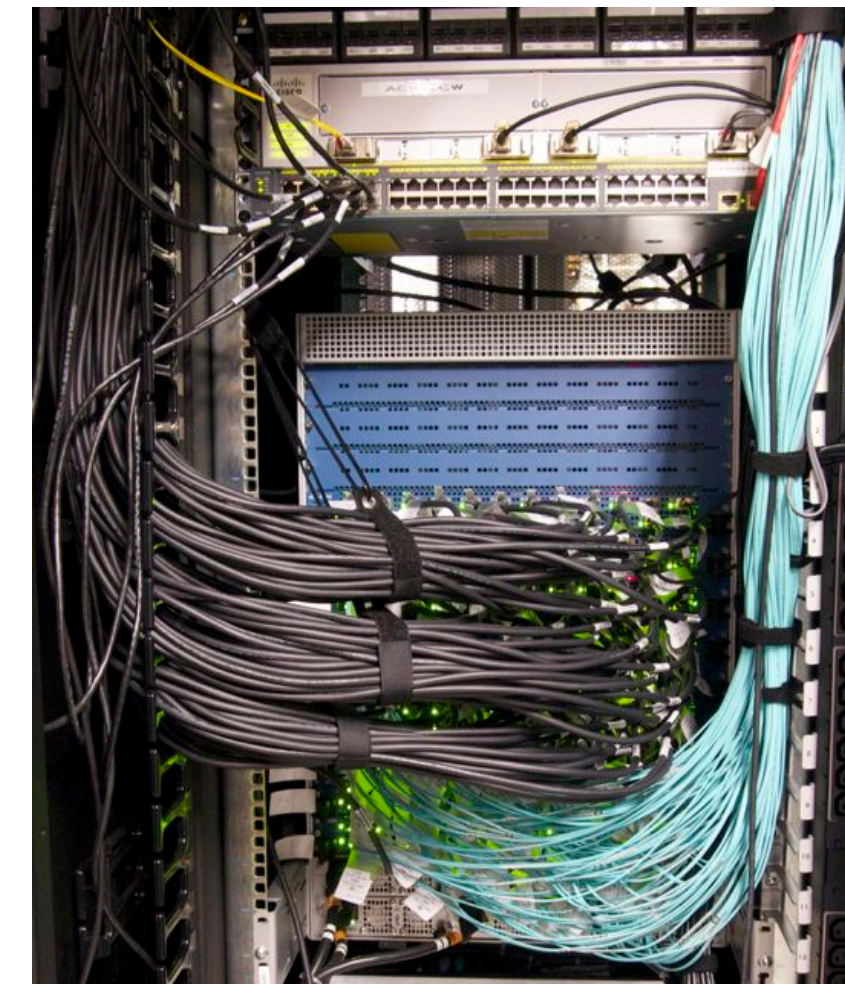
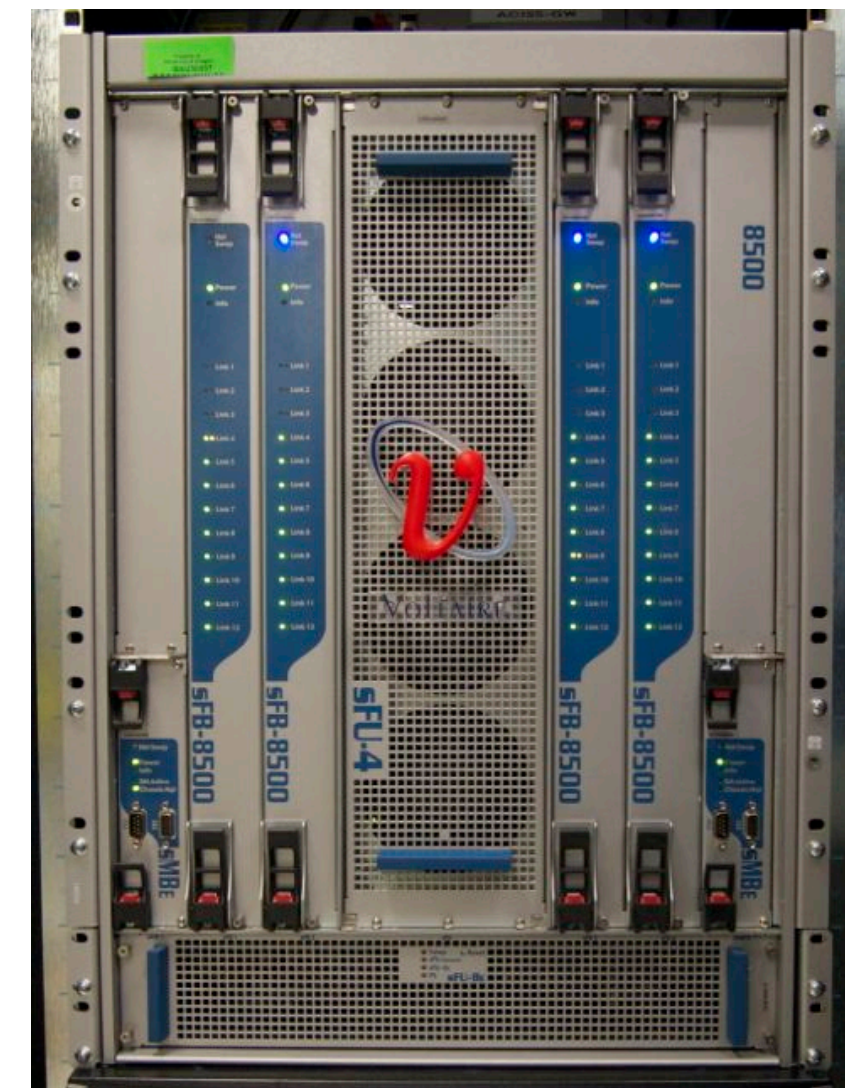
Redundant 10 GigE trunk to UONET.

Communication and storage, and public networks segmented by VLANs.

Tiered 1 GigE administration networks for CMU and ILO.

Mixed copper/fiber connections to nodes.

Hardware option to build out InfiniBand network.



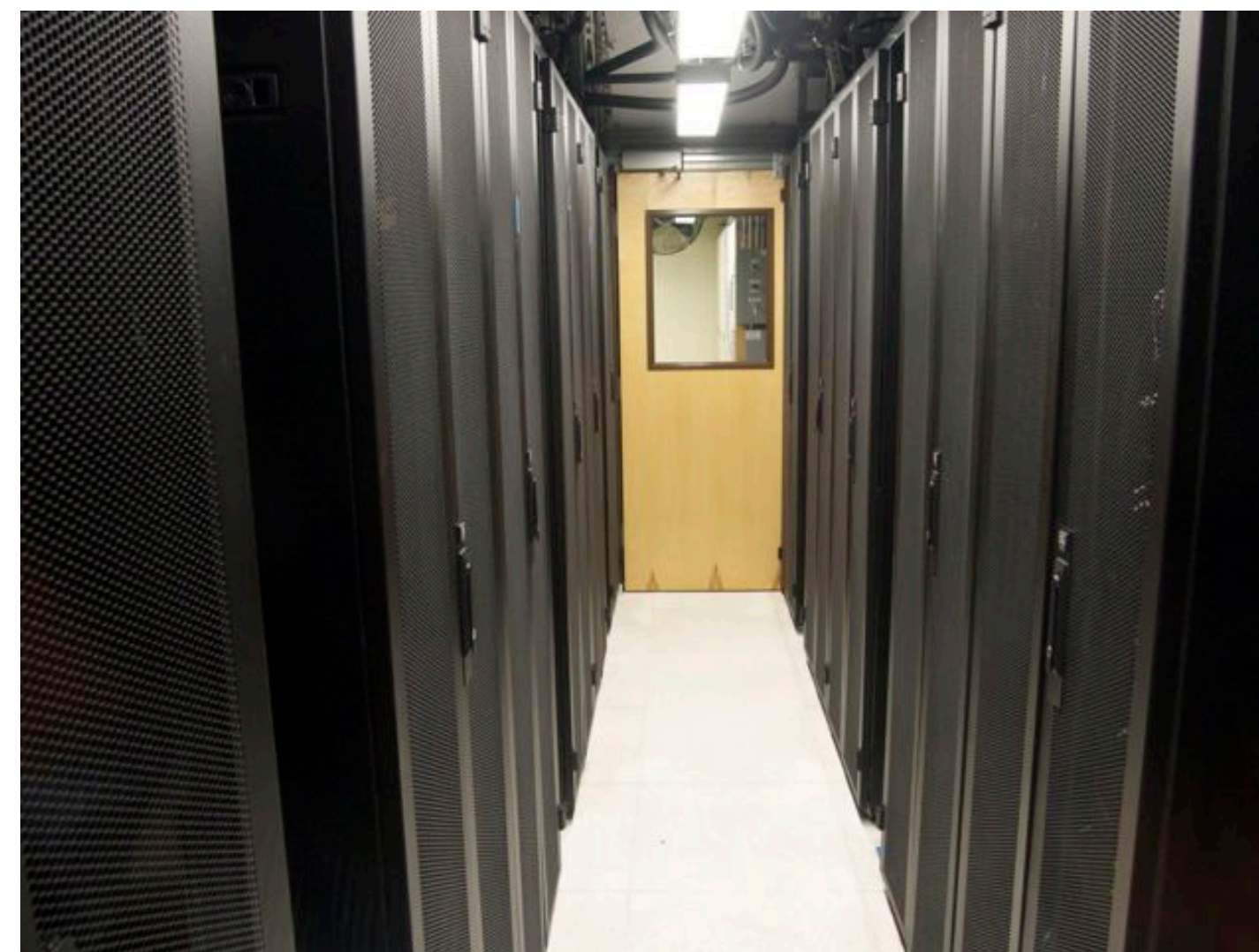
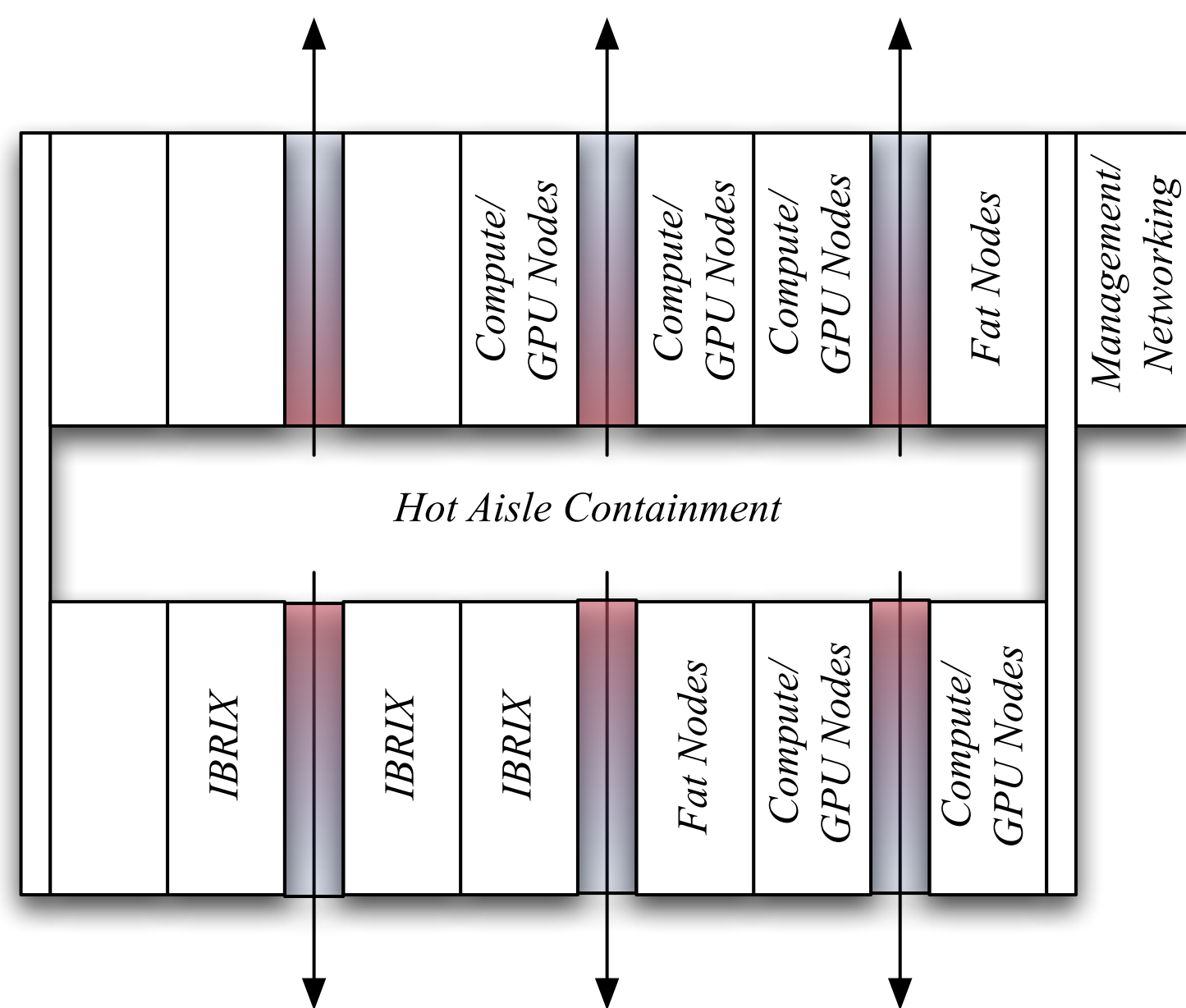
Computing Center Installation

Ten racks with hot aisle containment.

Network and management rack outside of containment.

Redundant power from campus power plant.

Chilled water supply used for AC heat exchange.



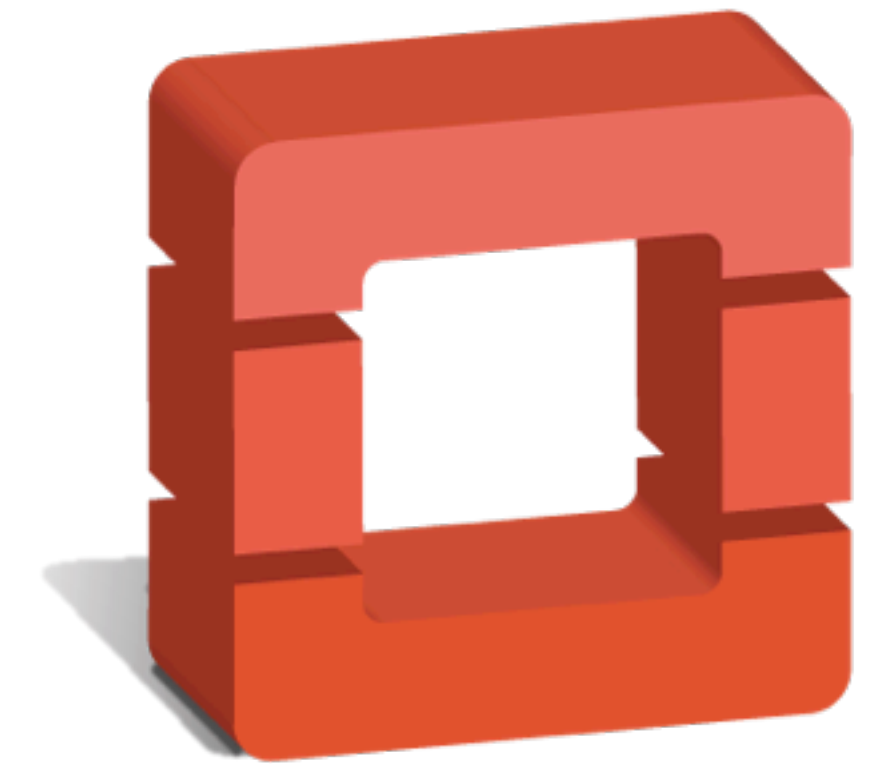
Why OpenStack?

Full cloud stack: compute, storage, identity,
and interface

Large community.

100% open source.

Well planned road map with predictable
releases.



openstack™
CLOUD SOFTWARE



How we deployed OpenStack:

RedHat with KVM as Hypervisor.

Installed from stable github source.

Why?

We have many patches in place, and many more planned.



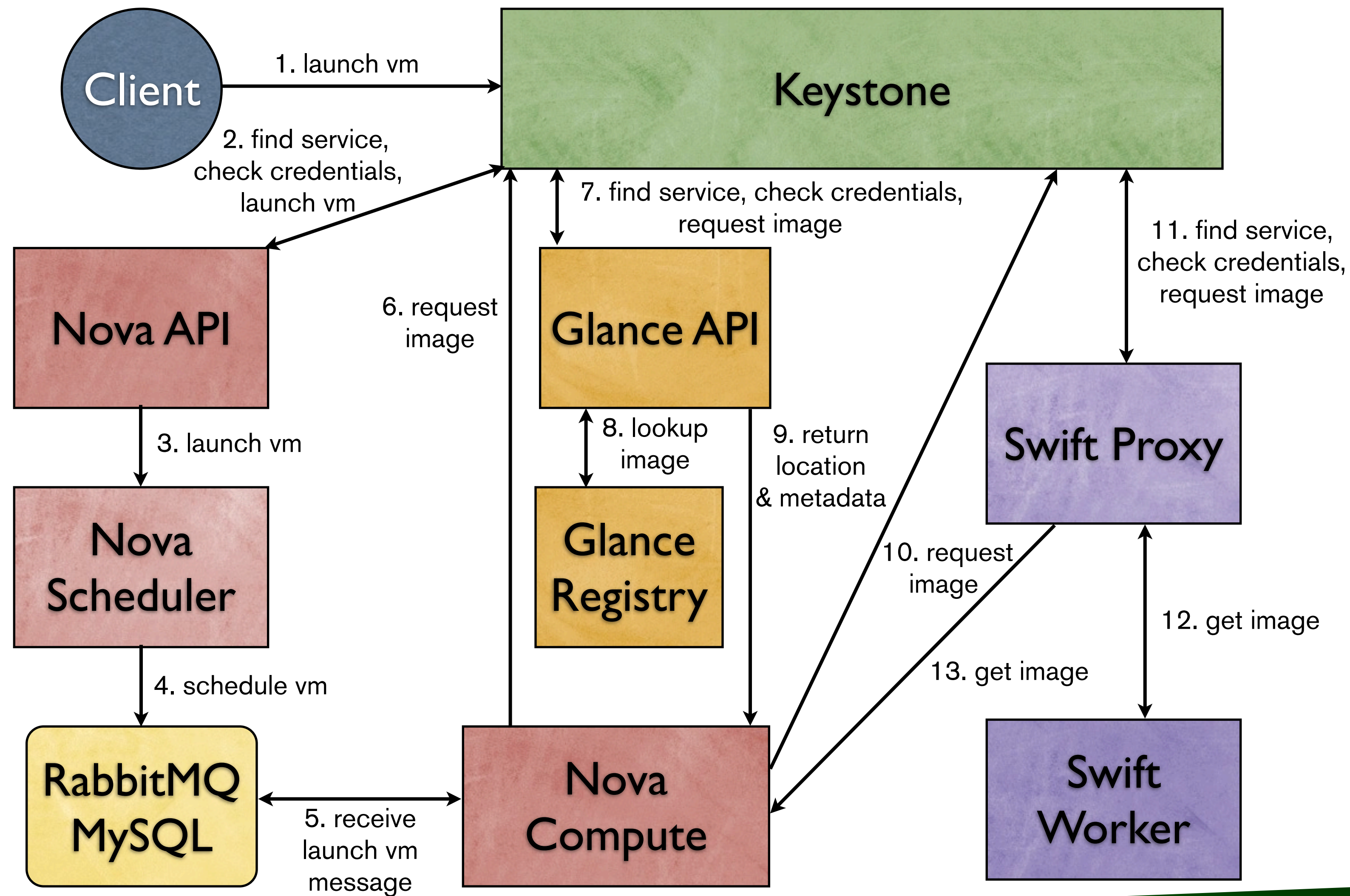
Things that made our life easier:

Puppet for node management and software deployment.

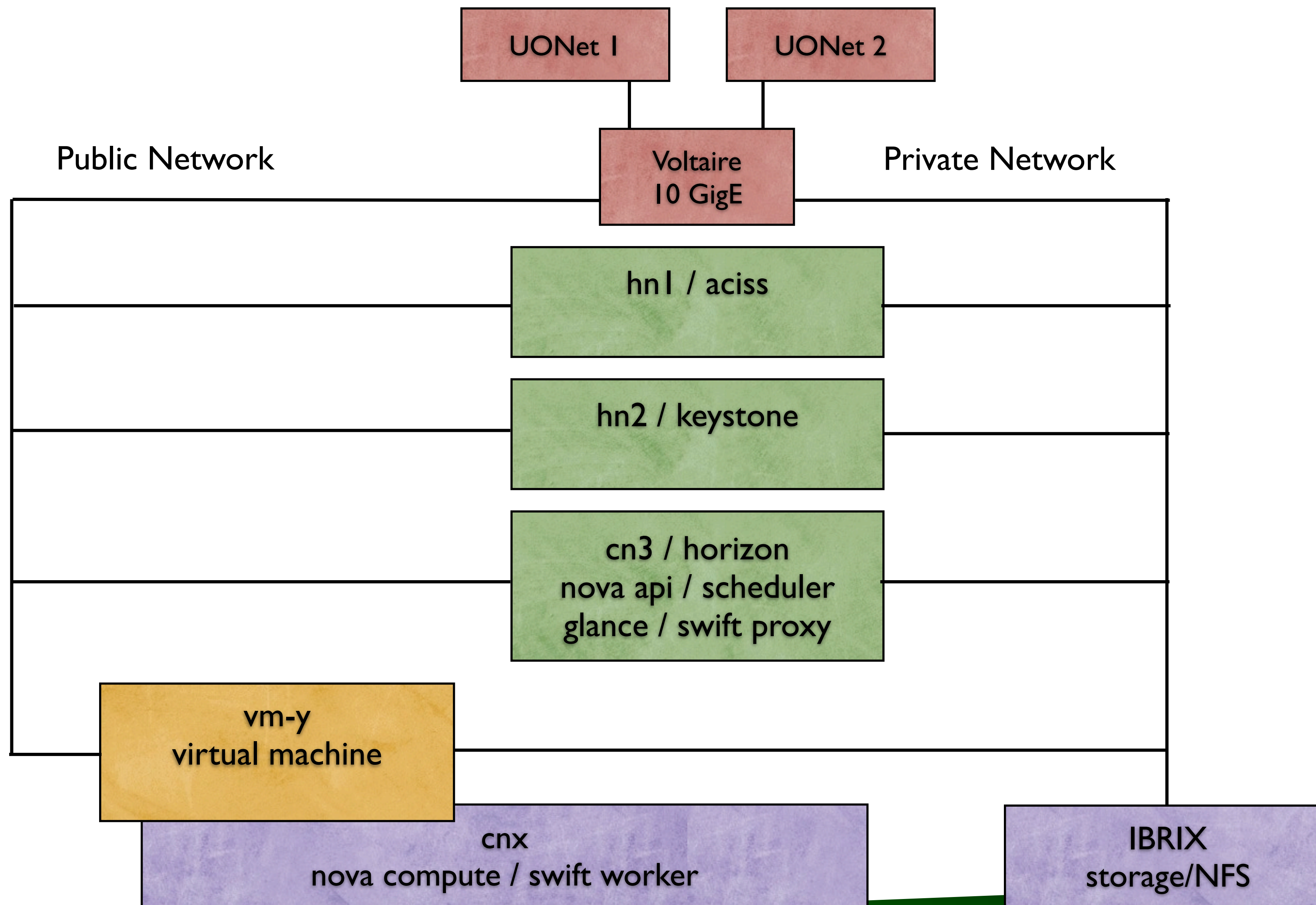
Python **virtualenv** for dependencies.

EPEL packages to helped with setup.

The **OpenStack community** for support and debugging.



14. Launch VM





Hacking Keystone

Integration with campus authentication (DuckID) essential.

One condition for using DuckID was security

But we still we needed to support off-campus users.

Hacking Keystone: Connecting to UO Directory Services

```
@@ -144,7 +145,8 @@ class Identity(sql.Base, identity.Driver):
    """
    user_ref = self._get_user(user_id)
    if (not user_ref
-       or not utils.check_password(password, user_ref.get('password'))):
+       or not (ro_ldap.ro_ldap_check_password(user_ref.get('name'), password)
+           or utils.check_password(password, user_ref.get('password')))):
        raise AssertionError('Invalid user / password')
```


Hacking Keystone: Connecting to UO Directory Services

```
@@ -144,7 +145,8 @@ class Identity(sql.Base, identity.Driver):
    """
    user_ref = self._get_user(user_id)
    if (not user_ref
-       or not utils.check_password(password, user_ref.get('password'))):
+       or not (ro_ldap.ro_ldap_check_password(user_ref.get('name'), password)
+           or utils.check_password(password, user_ref.get('password')))):
        raise AssertionError('Invalid user / password')
```



Hacking Keystone: Connecting to UO Directory Services

```
+++ b/keystone/config.py
+#ro_ldap
+register_str('use_ro_ldap', default=False, group='ro_ldap')
+register_str('anonymous_dn', group='ro_ldap')
+register_str('anonymous_pw', group='ro_ldap')
+register_str('auth_dn', group='ro_ldap')
+register_str('ldap_conn_str', group='ro_ldap')
```


Hacking Keystone: Securing the Server

OpenStack Identity (Keystone)

[Overview](#)
[Code](#)
[Bugs](#)
[Blueprints](#)
[Translations](#)
[Answers](#)

Keystone Essex does not support TLS over HTTPS

Keystone » Bugs » Bug #980864

Reported by Chris Hoge on 2012-04-13

This bug affects 2 people 12

Affects	Status	Importance	Assigned to	Milestone
▶ Keystone	Fix Committed	Wishlist	Unassigned	

Also affects project
 Also affects distribution Nominate for series

Bug Description

The most recent stable release of Keystone does not support TLS security over HTTPS. This functionality was available in the Diablo release, but was removed in Essex.

TLS should be enabled by default. Credentials should never be sent to an authentication server as plain text. If OpenStack APIs are made publicly available, the Keystone server must also be on a public interface to be accessible.



Hacking Keystone: Securing the Server

Dolph Mathews (dolph) on 2012-04-23

Changed in keystone:

`status:New → Confirmed`

`importance:Undecided → Wishlist`

Alan Pevec (apevec) wrote 8 hours ago:

#4

This is fixed on master <https://github.com/openstack/keystone/commit/8de61f8af43563b1d93291c868634810d9e42902>

but as a new feature isn't appropriate for stable/essex (see http://wiki.openstack.org/StableBranch#Appropriate_Fixes)

Alternative for stable/essex deployments is to use SSL termination in front of services as suggested in previous comments.

Alan Pevec (apevec) 7 hours ago

`tags:removed: essex`

Thierry Carrez (ttx) 5 hours ago

Changed in keystone:



Installing Swift: The Wrong Hardware

Swift wants dedicated JBODs.

It's pathologically worst case for RAID 6.
(we are using RAID 6 on our parallel filesystem)

To use Swift we carved off disk space on compute servers.
Then increased timeouts to account for additional server load.

The Swift-Glance-Nova timeout bug.



Database Inconsistencies (Horizon Example)

```
@@ -205,9 +205,21 @@ def get_ips(instance):
def get_size(instance):
    if hasattr(instance, "full_flavor"):
        size_string = _("%(RAM)s RAM | %(VCPU)s VCPU | %(disk)s Disk")
-       vals = {'RAM': sizeformat.mbformat(instance.full_flavor.ram),
-               'VCPU': instance.full_flavor.vcpus,
-               'disk': sizeformat.diskgbformat(instance.full_flavor.disk)}
+       try:
+           ram = sizeformat.mbformat(instance.full_flavor.ram)
+       except Exception, e:
+           ram = 0
+       try:
+           vcpu = instance.full_flavor.vcpus
+       except Exception, e:
+           vcpu = 0
+       try:
+           disk = sizeformat.diskgbformat(instance.full_flavor.disk)
+       except Exception, e:
```

Database Inconsistencies (Horizon Example)

```
@@ -205,9 +205,21 @@ def get_ips(instance):
def get_size(instance):
    if hasattr(instance, "full_flavor"):
        size_string = _("%(RAM)s RAM | %(VCPU)s VCPU | %(disk)s Disk")
-        vals = {'RAM': sizeformat.mbformat(instance.full_flavor.ram),
-               'VCPU': instance.full_flavor.vcpus,
-               'disk': sizeformat.diskgbformat(instance.full_flavor.disk)}
+
+        try:
+            ram = sizeformat.mbformat(instance.full_flavor.ram)
+        except Exception, e:
+            ram = 0
+        try:
+            vcpu = instance.full_flavor.vcpus
+        except Exception, e:
+            vcpu = 0
+        try:
+            disk = sizeformat.diskgbformat(instance.full_flavor.disk)
+        except Exception, e:
+            disk = 0
```




Planned Extensions

Service monitoring and auto-recovery.

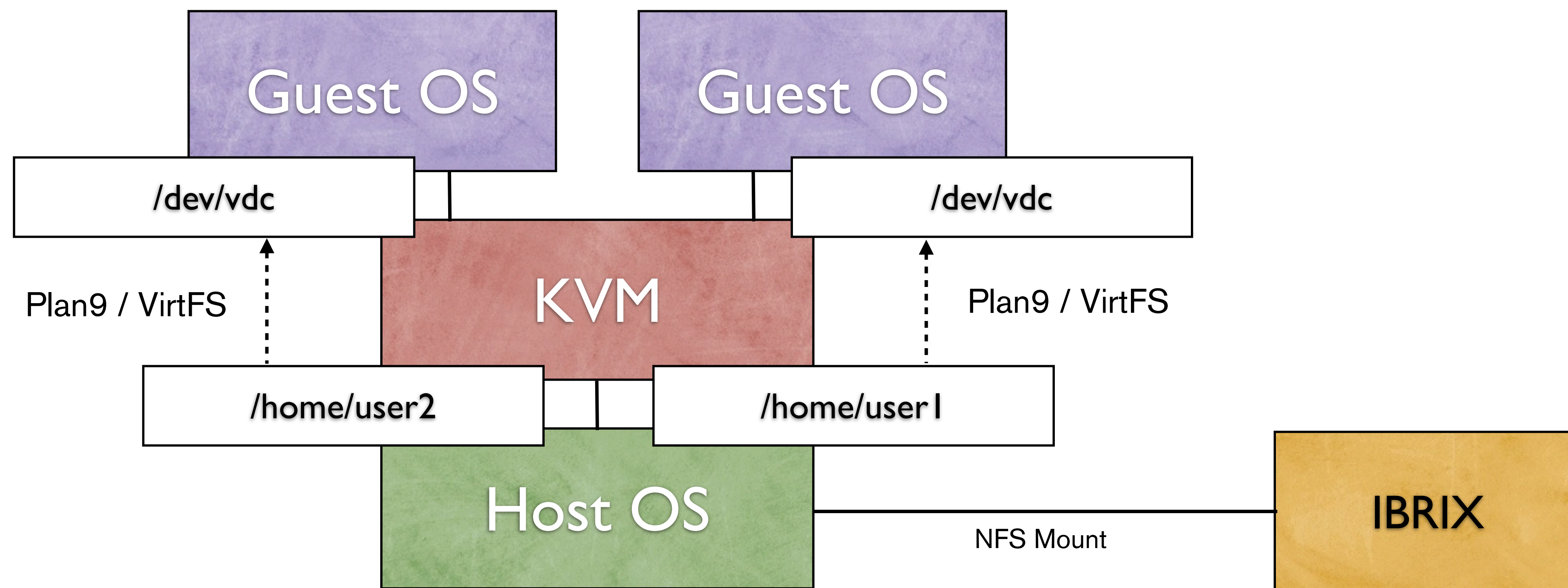
Porting existing scientific cloud computing projects.

Galaxy project: <http://galaxy.psu.edu/>

Building “applications in a browser” using VNCs.

Securely attaching network storage to VMs.

Attaching Nova to our Parallel NAS: Plan9 filesystem on KVM to the rescue





Usage to date:

Graduate seminar in Scientific Cloud Computing.
<http://prodigal.nic.uoregon.edu/~hoge/cis607>

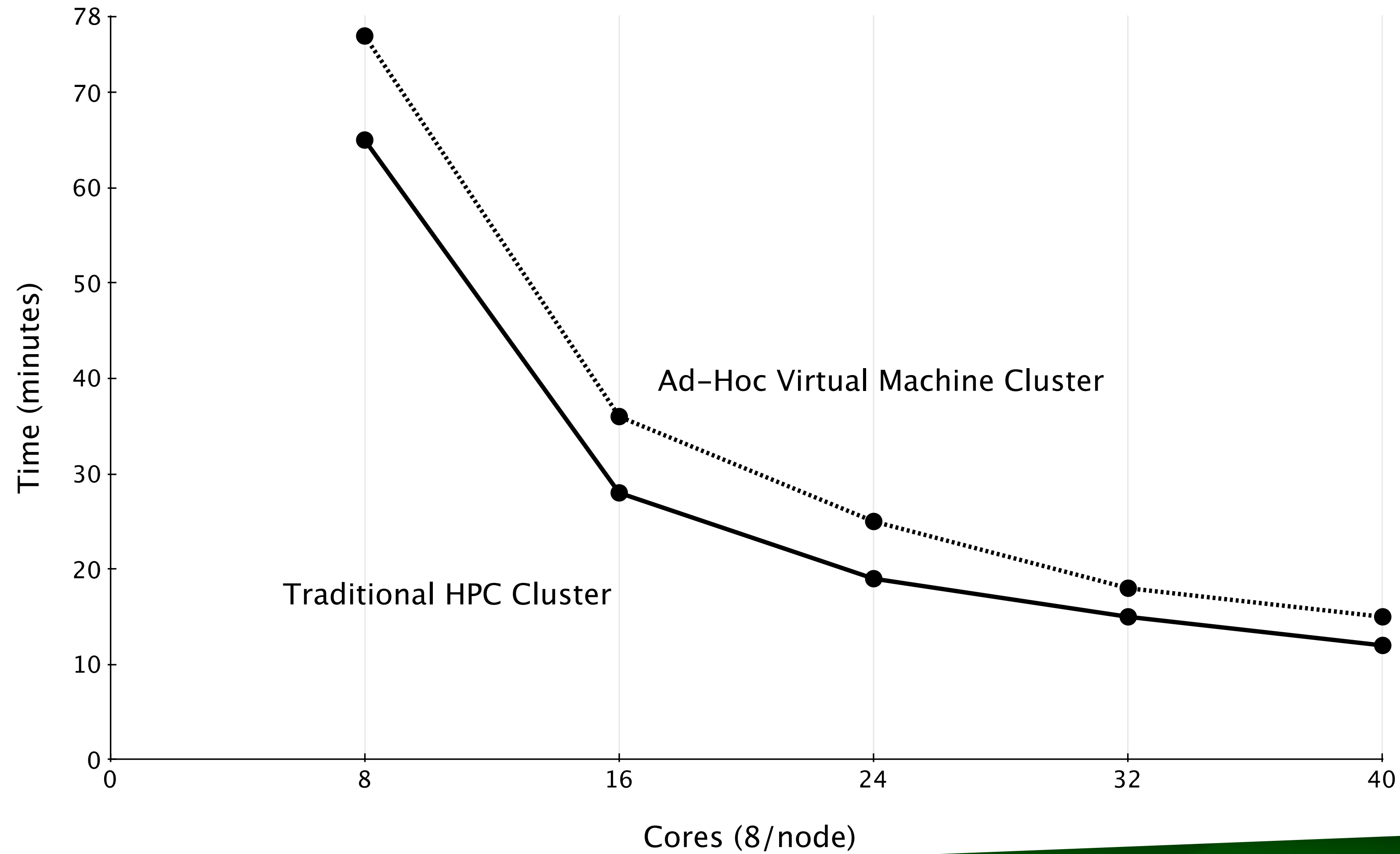
Summer workshop in Molecular Biology

VM security research.

Scientific data storage on Swift.



Comparison of OpenMP/MPI computation on Cloud to Bare Metal





General Impressions:

Amazing community.

Stability and features are growing rapidly.

For a production environment, you need to hack.

Fortunately, OpenStack is easy to hack.

OpenStack has a very bright future.



Thanks to...



CASIT





And especially to...

Dr. Allen Malony, who had the vision for the grant.

Robert Yelle, for his awesome system administration.

Micah Sardell and the Computing Center, for hosting.

José Dominguez, for networking support.

Rob Chevalier, for getting us plugged into DuckID.



Interested in getting involved?

Opportunities for:

Visualization experts

System administrators

Graduate students

contact me: hoge@uoregon.edu



Questions?

Chris Hoge
Scientific and Cloud Computing Technical Lead
Neuroinformatics Center
hoge@uoregon.edu