



# Virtualization within FermiGrid

Keith Chadwick

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



# Previous talks on FermiGrid Virtualization and High Availability

---

HEPiX 2006 at Jefferson Lab:

<https://indico.fnal.gov/conferenceDisplay.py?confId=384>

HEPiX 2007 in St. Louis:

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2513>

OSG All Hands 2008 at RENC:

<http://indico.fnal.gov/contributionDisplay.py?contribId=13&sessionId=0&confId=1037>

OSG All Hands 2009 at LIGO:

<http://indico.fnal.gov/contributionDisplay.py?contribId=52&sessionId=78&confId=2012>

Fermilab detailed documentation:

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2590>

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2539>



# FermiGrid-HA - Highly Available Grid Services

---

The majority of the services listed in the FermiGrid service catalog are deployed in high availability (HA) configuration that is collectively known as "FermiGrid-HA".

FermiGrid-HA utilizes three key technologies:

- Linux Virtual Server (LVS).
- Scientific Linux (Fermi) 5.3 + Xen Hypervisor.
- MySQL Circular Replication.



# Physical Hardware, Virtual Systems and Services

	Physical Systems	Virtual Systems	Virtualization Technology	Service Count
FermiGrid-HA Services	6	34	Xen	17
CDF, D0, GP Gatekeepers	9	28	Xen	9+6
Fermi & OSG Gratia	4	10	Xen	12
OSG ReSS	2	8	Xen	2
Integration Test Bed (ITB)	2+8	14+32	Xen	14
Grid "Access" Services	2	4	Xen	4
"FermiCloud"	8 (+16)	64 (+128)	Xen	--
"Fgtest" Systems	7	51	Xen	varies
"Cdf Sleeper Pool"	3	9	Xen	1+1
"GridWorks"	11	~20	Kvm	1



# FermiGrid – Organization of Physical Hardware, Virtual Systems and Services

---

<http://fermigrid.fnal.gov/fermigrid-systems-services.html>

- <http://fermigrid.fnal.gov/fermigrid-organization.html>
- <http://fermigrid.fnal.gov/cdfgrid-organization.html>
- <http://fermigrid.fnal.gov/d0grid-organization.html>
- <http://fermigrid.fnal.gov/gpgrid-organization.html>
- <http://fermigrid.fnal.gov/gratia-organization.html>
- <http://fermigrid.fnal.gov/fgtest-organization.html>
- <http://fermigrid.fnal.gov/fgitb-organization.html>
- <http://fermigrid.fnal.gov/ress-organization.html>

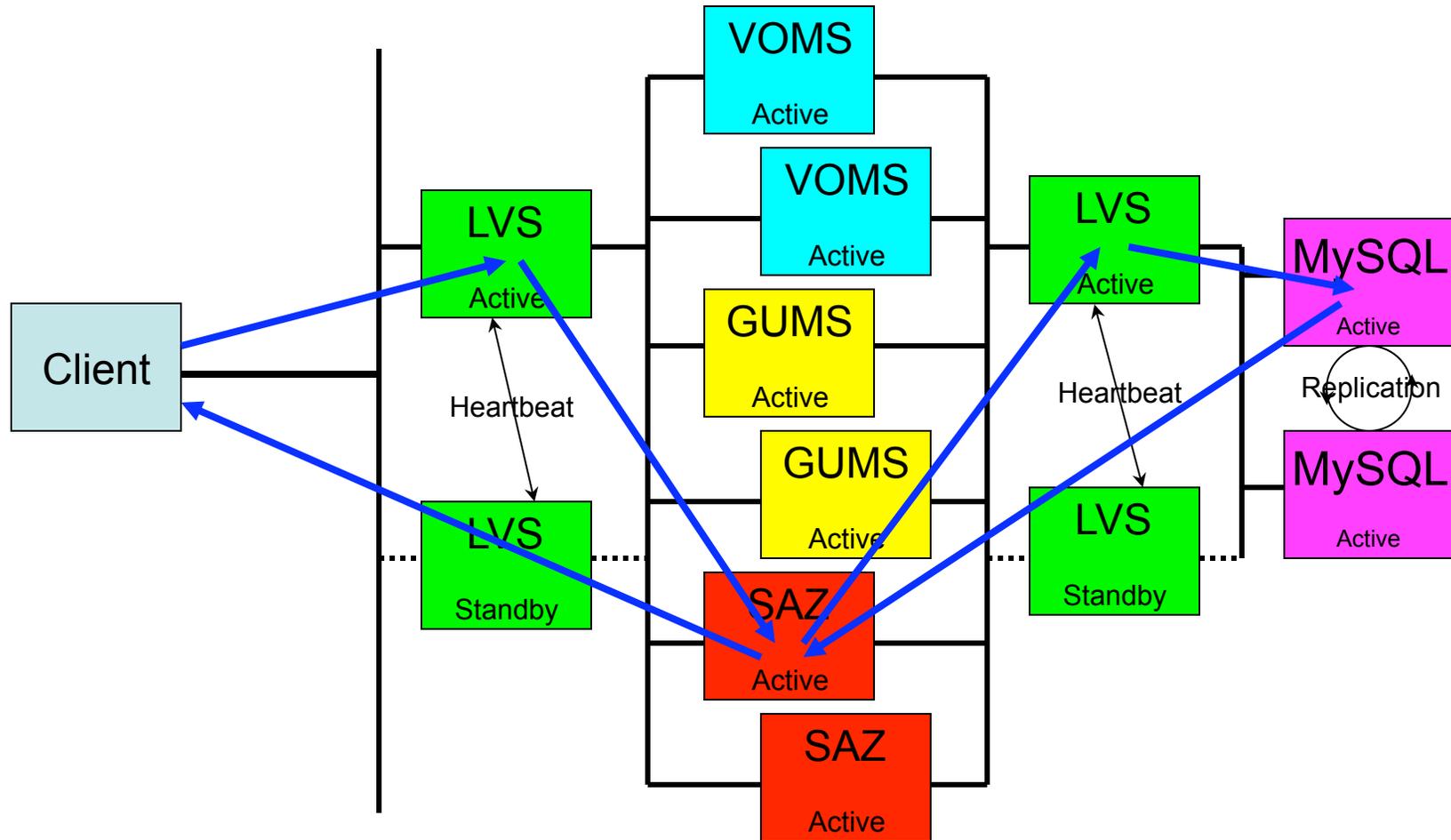


# HA Services Deployment

---

## FermiGrid employs several strategies to deploy HA services:

- Trivial monitoring or information services (examples: Ganglia and Zabbix) are deployed on two independent virtual machines.
- Services that natively support HA operation (examples: OSG ReSS, Condor Information Gatherer, FermiGrid internal ReSS deployment) are deployed in the standard service HA configuration on two independent virtual machines.
- Services that maintain intermediate routing information (example: Linux Virtual Server) are deployed in an active/standby configuration on two independent virtual machines. A periodic heartbeat process is used to perform any necessary service failover.
- Services that are pure request/response services and do not maintain intermediate context (examples: GUMS and SAZ) are deployed using a Linux Virtual Server (LVS) front end to active/active servers on two independent virtual machines.
- Services that support active-active database functions (example: circularly replicating MySQL servers) are deployed on two independent virtual machines.





# Virtualized Non-HA Services

---

The following services are virtualized, but not (yet) currently implemented as HA services:

- Globus gatekeeper services (such as the CDF and D0 experiment globus gatekeeper services) are deployed in segmented "pools".
  - Loss of any single pool will reduce the available resources by approximately 50%.
  - Expect to segment the GP Grid cluster in FY10.
- MyProxy
  - We need a secure block level replication solution to implement this in an active/standby HA configuration.
  - DRBD may be the answer, but we have not figured out how to incorporate the DRBD Kernel modifications into the Xen Kernel.
- Fermi & OSG Gratia Accounting service [Gratia]
  - Not currently implemented as an HA service.
  - If the service fails, then the service will not be available until appropriate manual intervention is performed to restart the service.
  - Equipment is on order to "HA" the Gratia services.



# Measured Service Availability

---

FermiGrid actively measures the service availability of the services in the FermiGrid service catalog:

- <http://fermigrid.fnal.gov/fermigrid-metrics.html>
- <http://fermigrid.fnal.gov/monitor/fermigrid-metrics-report.html>

The above URLs are updated on an hourly basis.

The goal for FermiGrid-HA is > 99.999% service availability.

- Not including Building or Network failures.
- These will be addressed by FermiGrid-RS (redundant services) in a future FY.

For the period 01-Dec-2007 through 30-Jun-2008, we achieved a service availability of 99.9969%.

For the last year, we have achieved a collective core service availability of 99.950%.



# FermiGrid Service Level Agreement

---

## Authentication and Authorization Services:

- The service availability goal for the critical Grid authorization and authentication services provided by the FermiGrid Services Group shall be 99.9% (measured on a weekly basis) for the periods that any supported experiment is actively involved in data collection and 99% overall.

## Incident Response:

- FermiGrid has deployed an extensive automated service monitoring and verification infrastructure that is capable of automatically restarting failed (or about to fail) services as well as performing notification to a limited pager rotation.
- It is expected that the person that receives an incident notification shall attempt to respond to the incident within 15 minutes if the notification occurs during standard business hours (Monday through Friday 8:00 through 17:00), and within 1 (one) hour for all other times, providing that this response interval does not create a hazard.

## FermiGrid SLA Document:

- <http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2903>



# Why 99.999%?

---

A service availability of 99.999% corresponds to 5m 15s of downtime in a year.

- This is a challenging availability goal.
- [http://en.wikipedia.org/wiki/High\\_availability](http://en.wikipedia.org/wiki/High_availability)

The SLA only requires 99.9% service availability = 8.76 hours.

So, really - Why target five 9's?

- Well if we try for five 9's, and miss then we are likely to hit a target that is better than the SLA.
- The hardware has shown that it is capable of supporting this goal.
- The software is also capable of meeting this goal (modulo denial of service attacks from some members of the user community...).
- The critical key is to carefully plan the service upgrades and configuration changes.



# FermiGrid Persistent ITB

---

Gatekeepers are Xen VMs.

Worker nodes are also partitioned with Xen VMs:

- Condor
- PBS (coming soon)
- Sun Grid Engine (ibid)
- + A couple of "extra" CPUs for future cloud investigation work (ibid).

<http://fermigrid.fnal.gov/fgitb-organization.html>



# Cloud Computing

---

FermiGrid is also looking at Cloud Computing.

We have a proposal in this FY, that if funded, will allow us to deploy an initial cloud computing capability:

- Dynamic provisioning of computing resources for test, development and integration efforts.
- Allow the retirement of several racks of out of warranty systems.
  - Both of the above help improve the “green-ness” of the computing facility
- Additional “peaking” capacity for the GP Grid cluster.



## Conclusions

---

### Virtualization is working well within FermiGrid.

- All services are deployed in Xen virtual machines.
- The majority of the services are also deployed in a variety of high availability configurations.

### We are actively working on:

- The configuration modifications necessary to deploy the non-HA services as HA services.
- The necessary foundation work to allow us to move forward with a cloud computing initiative (if funded).

Any questions?