

December 2019 Central Computing Outage (Fermi)

First Edition: 6 Dec 2019

Version 1.1 (10:50 PT 6 Dec 2019)



On 8 Dec 2019 this outage was postponed until July 2020

Power outage time-line.

[Tentative proposal] Not many details are currently known, but this power outage will affect substations #7 (next to bldg 50) and #8 (located on the 4th floor of bldg 50). All of bldg 50 will be without normal power. The facilities (F&O) group plan to do their maintenance during the 4-day period starting 26 Dec 2019. However, the outage will start earlier due to lack of staff during the holiday shutdown. Minimally, it is expected that all H.A. (High Availability) and experiment-critical equipment will be powered throughout the 16+ days of the holiday shutdown. This page captures what Fermi will need to maintain a minimal data processing effort running during the outage.



Note that the ability to perform general science analysis at SLAC by the LAT collaboration will be seriously hindered by this outage due to the fact that much of the batch farm will be unavailable.

Date	Time	Action
A day or two prior to 20 Dec 2019	TBA	Test of power source switching (i.e., normal line power to generator)
Fri 20 Dec 2019	TBA	switch to generator power (this could happen earlier) This will require a several-hour outage
Mon 6 Jan 2020	TBA	return to normal power. This will require a several-hour outage

Outage preparation task list

- ☒ Define needed xrootd resources (Wilko Kroeger)
- ☒ Confirm sufficient xrootd space to handle 16+ day HPSS outage (Wilko Kroeger)
- ☒ Define needed Level 0 and half-pipe resources (Steve Tether)
- ☒ Define needed Level 1 resources (Tom Stephens)
- ☐ Update service VM hypervisor host mapping (Brian Van Klaveren)
- ☒ Suggest/Move non-HA VMs to HA (Brian Van Klaveren)
- ☒ Define needed ASP resources (Jim Chiang)
- ☒ Define needed resources for critical Fermi Science pipelines (various)
 - ☒ Gravitational wave analysis (Nicola Omodei)
 - ☒ Flare Advocates (Gulli Johannesson, Stefano Ciprini)
 - ☒ Burst Analysis (Dan Kocevski)

LISOC Operations Functions

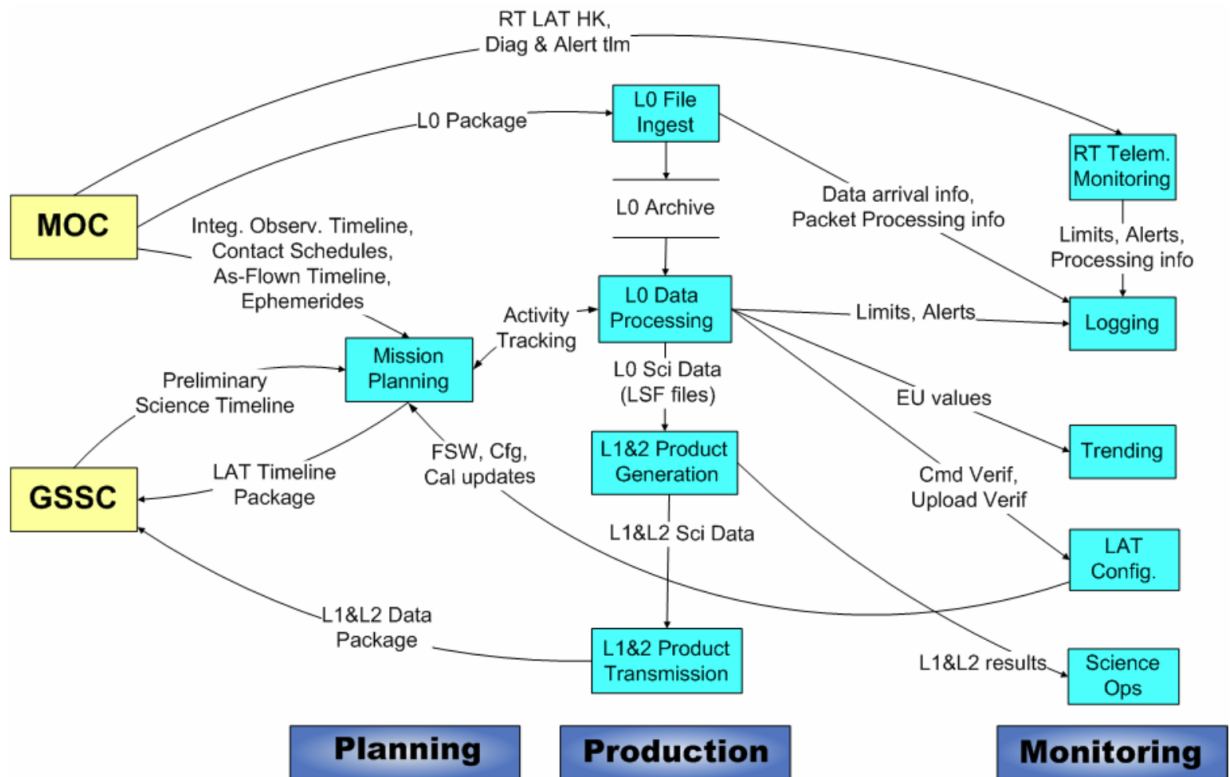


Table of LISOC Tasks and Services

Function/Service	Sub-Functions	Needed Servers	Needed Databases	Needed File Systems	Other Needs	Needed During Shutdown?	Available During Shutdown?
Mission Planning, LAT Configurations	FastCopy	fermilnx01 and fermilnx02	TCDB	AFS	Fermi LAT Portal: Timeline Webview; Confluence, JIRA, Mission Planning s/w, FastCopy Monitoring Sharepoint (reference for PROCs and Narrative Procedures for commanding in case of anomalies)	yes	
Real Time Telemetry Monitoring		fermilnx01 and fermilnx02			spread Fermi LAT Portal: Real Time Telemetry, Telemetry Monitor	during anomalies	
Logging		fermilnx01 and fermilnx02	TCDB		Fermi LAT Portal: Log Watcher	yes	
Trending			TCDB		Fermi LAT Portal: Telemetry Trending	yes	
L0 File Ingest and Archive	FastCopy		L0 Archive			yes	
Data Gap Checking and Reporting	FastCopy	fermilnx01 and fermilnx02	L0 Archive			yes, continuously	
L1 processing	pipeline	SLAC Farm	Data Catalog		Fermi LAT Portal: Pipeline, Data Processing	yes	
L1 Data Quality Monitoring					Fermi LAT Portal, Telemetry Trending		
L1 delivery	FastCopy	fermilnx01 and fermilnx02	Data Catalog			yes	
L2 processing (ASP) and Delivery	FastCopy	fermilnx01 and fermilnx02	Data Catalog		Fermi LAT Portal: Pipeline, Data Processing	daily, weekly	

The following table of servers must remain powered up and operational for Fermi Level 1 and critical Science Pipelines to function.

Fermi has requested that **all** VMs be relocated (at least temporarily) to the two H.A. hypervisor machines, thus some of the tasks listed below are no longer relevant.

- ☒ Confirm current H.A. rack occupants. [spreadsheet from Christian Pama](#)
Old (2017) spreadsheet [here](#)
- ☐ Confirm the VM-master for a given VM. Use the 'node' command, e.g., \$ node -whereis fermilnx-v12 (obsolete)
- ☐ Confirm the tomcat <-> service associations. Table [here](#).
- ☐ Confirm the tomcat-VM associations in this table. Use the 'node' command, e.g., \$ node -whereis glast-tomcat01



NOTE: Fermi has four VMware hypervisors, each of which contain some number of VMs running Fermi services. Two of these hypervisor machines are in the H.A. racks (fermi-vmclust03/04), while the others (fermi-vmclust01/02) are not. At this writing there are no user-level tools to allow one to discover which VMs are running on which hypervisor machines.

Category†	server	VM/service	function
XC	fermi-gpfs01	xrootd	xrootd server and storage
	fermi-gpfs02		
	fermi-gpfs05		
	fermi-gpfs06		
	fermi-gpfs07		
	fermi-gpfs08		
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v02	xrootd redirector
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v12	xrootd redirector
XC	fermi-gpfs03	GPFS	Fermi NFS/GPFS storage
	fermi-gpfs04		
XC	fermi-cnfs01	GPFS/NFS bridge	Fermi NFS storage access
	fermi-cnfs02		
HA	staas-gpfs50		Critical ISOC NFS storage
	staas-gpfs51		
HA	fermilnx01		LAT config, fastcopy and real-time telemetry
HA	fermilnx02		LAT config, fastcopy and real-time telemetry
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v03	archiver
HA	fermi-oracle03		oracle primary
XC	fermi-oracle04		oracle secondary
HA	mysql05	mysql-node03	calibration, etc. DB
	mysql06		
XC	400 cores		(50 "hequ" equivalents) batch hosts for LISOC queues={express,short,medium,long,glstdataq} users={glst,lsstsim,lsstprod,glstmc,glstrow}
XC	200 cores		(25 "hequ" equivalents) batch hosts for Science Pipelines
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v07 /tomcat01	Commons, Group manager
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v16 /tomcat06	rm2
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v05 /tomcat08	dataCatalog
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v17 /tomcat09	Pipeline-II
XC/HA	fermi-vmclust01/02 /03/04	fermilnx-v15 /pipeline-mail01	Pipeline-II email server

XC/HA	fermi-vmclust01/02/03/04	fermilnx-v18/tomcat10	FCWebView, ISOCLogging, MPWebView TelemetryMonitor, TelemetryTableWebUI
XC/HA	fermi-vmclust01/02/03/04	fermilnx-v10/tomcat11	DataProcessing
XC/HA	fermi-vmclust01/02/03/04	fermilnx-v11/tomcat12	TelemetryTrending
NC	(non-Fermi server)	astore-new (HPSS)	FastCopy data archive **We have been granted a temporary quota increase of 1 TB on /nfs/farm/g/glast/u23, which has allowed this item to become "NC"***
HA	(non-Fermi server)	trscron	tokenized cron
HA	(non-Fermi server)	lnxcron	cron
XC	(non-Fermi server)	(farm manager, etc.)	LSF management
HA	yfs01/NN (non-Fermi)		basically all of AFS
HA	(non-Fermi server)	JIRA	issue tracking (HA as of 10/20/2017)
XC	rhel6-64		public login nodes (a small number is needed for interactive access)

† Equipment categories

Category	Machine status
NC	non-critical for entire 16-day shutdown period
XC	experiment critical but not in H.A. rack, only a few, short outages acceptable
HA	high-availability (continuous operation)

Total non-HA machines to receive emergency power:

Machine Type	Total	Notes
GPFS servers	8	
NFS/GPFS bridge	2	
VMware hypervisors	2	Not needed if all Fermi services can be moved to the two H.A. hypervisors
batch nodes ("hequ" equivalents)	75	Depending on which batch nodes are selected, some may already be in H.A. power
Oracle servers	1	There is rumor that this machine may already be on H.A. power – to be confirmed
Public login nodes	N	(where "N" is a small integer)
TOTAL	88+N	

★ Note that HPSS is NOT required by Fermi.

The services for L1:

oracle

- pipeline
- data catalog
- group manager

mysql

- calibrations

tomcats

- pipeline
- data catalog
- data processing

isoc servers
xroot

The following servers are needed to allow processing of new data (older data on fermi-xrd or HPSS will not be available):

- fermi-gpfs01/02, fermi-gpfs05/06 and fermi-gpfs07/08
These are the servers that make up the fermi xrootd gpfs space. Each pair servers a part of the total gpfs space.
 - fermi-gpfs05 runs the xrootd server for the gpfs space
- fermilnx-v02 (redirector)
- fermilnx-v12 (redirector)

nfs

- Pretty much everything that's currently on staas-gpfs50/51
- Parts of the non-HA Fermi NFS file system

LSF

- 50 hosts should let us keep up (including ASP)

Here's what ISOC tasks need:

FASTCopy chain

staas-gpfs50/51
fermilnx01
fermilnx02
trscron
fermilnx-v03 (Archiver)
Whatever the pipeline server runs on.
xroot servers
astore-new system (HPSS)

Web servers

tomcat01 Commons
tomcat06 rm2
tomcat09 Pipeline-II
tomcat10 FCWebView, ISOCLogging, MPWebView
TelemetryMonitor, TelemetryTableWebUI
tomcat11 DataProcessing
tomcat12 TelemetryTrending

Science Pipelines

Gravitational Wave analysis (Nicola)

- Runs once per GW event reported from Global GW detectors
- Large variability in CPU requirement due to varying size of GW localization in sky
- Estimate 300 core-hours per day per GW event (e.g., 10 hours on 30 cores)
This would be 4 hequ hosts for about 10 hours per GW event

Flare Advocate analysis

- Batch jobs submitted to follow up on flare alert, typically once/day at most
- Batch job runs ~30 minutes on ~100 cores?
About 12 hequ-class nodes are needed for this analysis

Burst Analysis

- Batch job(s) submitted to follow up on gamma-ray burst detection
- Six jobs/burst, medium queue, rhel6
- Recent 7-day week had 11 triggers, so >1/day
Six hequ batch nodes should cover this need

FAVA (Fermi All-Sky Variability Analysis)

- Runs weekly
- Can postpone routine FAVA analysis until after the outage

High availability racks

For general information about the High-availability racks, Shirley provided this pointer to the latest [list](#):

"Service Now, Knowledge Base, search for "High Availability" , following link for current servers"

And here is the current statement about high-availability functionality:

```
Current Services in HA Racks
•CATER application
•Confluence application
•Data center management tool
•Drupal web
•Email lists
•Email transport infrastructure
•ERP application
•Exchange email
•EXO application
•Facilities monitoring
•Fermi application
•IT Ticketing system
•Network infrastructure
•Site Security infrastructure
•Unix authentication infrastructure
•Unix AFS infrastructure
•Unix mailboxes
•Unix monitoring
•VPN
•Windows authentication infrastructure
•Windows file servers and SAN
•Windows monitoring
•Windows web
```

Supporting documentation

Email from Steve Tether with some storage-related information:

Change "fermilnx01 or fermilnx02" to "fermilnx01 and fermilnx02". While services can all be shifted to one of those machines, frankly it's a pain.

The partition staas-cnfs501b:/gpfs/slac/ha/fsl/g/fermi/u23 currently has 554 GB free. This is where we store:

- Incoming FASTCopy packages (L0 data, HSK data).
- Outgoing FASTCopy packages (L1 data, mission planning).
- Unpacked LAT raw data (L0, HSK, etc.)

FASTCopy packages for both L0 and L1 data are archived daily to "astore-new" and are then deleted within 24 hours. "astore-new" is a POSIX-compliant filesystem interface to HPSS that replaced the old "astore" interface. This is driven by the old GLAST Disk Archiver service. The packages are also archived to xrootd daily. Unpacked raw data is also archived to xrootd but is retained for 60 days on u23. The unpacked raw data on xrootd is a "live" backup in the sense that it can be accessed by ISOC tools and L1 reconstruction if needed, though that option is not normally enabled.

We get something like 16 GB of L0 data daily. If archiving to astore-new is turned off then we would have to retain the original incoming L0 FC packages, the unpacked L0 data and the L1 FC packages. Naively assuming that all of these to be about the same size that means retaining 48GB or more per day so u23 would fill up in 11.5 days or less. And we'd probably start experiencing problems as it approached being 100% full.

If the astore-new archiving were kept going but the xrootd archiving were suspended, then we would retain only the 16 GB of unpacked L0 data per day which would fill up u23 in 30 days or so.

So I would recommend changing the classification of "astore (non-Fermi server)" from NC to XC for this long of an outage. And rename "astore" to "astore-new (HPSS)". I see that the Archiver server fermilnx-v03 is already classified as XC, so that's good.

The partition staas-cnfs501b:/gpfs/slac/ha/fsl/g/fermi/u41 is used by the halfpipe to store events extracted from LAT raw data. The events would take up 16 GB daily times some modest expansion factor. That partition needs to be kept going for normal processing. I don't know how long the event data is retained but the partition currently has 4.4 TB free so it shouldn't be a problem in any event.

All the rest of the page seems OK.

Wilko's statement regarding space currently available in xrootd:

there are currently about ~290TB free in the xrootd gpfs space, which is plenty . Also, if needed we can always purge old recon files from disk.

Nicola's estimate of batch power needed for GW follow up pipeline:

I am trying to figure out the right numbers looking at the resource plots....

Not sure how to read the plots. I think they are running on 300 cores for about an hour. So my estimation was 30 cores for 10 hours...

Stefano's comment on Flare Advocates:

for the FA shifts
in case of no connection, tokens decaying,
servers down or lack of processed ASP drp/pgwave data
and DB tables, simply the FA shifters can postpone the daily FA analysis and the run of FA-script for each day when the system will be working again
(after 2, 3, 6 of january?).
Even if with an about 10-days delay there will not be losses of checked daily sky and daily confluence reports.

Sara's response to Flare advocate script question:

They are usually submitted to the "medium" queue, in general relatively fast to finish (~30min) and not that much demanding resources wise, roughly a hundred cores I'd say. If you need more detailed info I do not know them on top of my head, there were also some changes lately to the code (I believe). I or some of the FA coord. managing this would need to take a look into the code. E.g. you were asking about the batch node, what other info may help?

Teddy adds: "The FA scripts tend to be run each day at most."

Dan's statement on various Pipelines (including FAVA):

We have a few analysis pipelines that currently use the batch system. These include the burst advocate analysis, the gravitational wave followup, and FAVA. The gravitational wave analysis typically requires thousands of jobs to be launched to analyze a large portion of the sky, so I think it's probably hopeless to keep that up during the outage. FAVA runs on weekly timescales, so we can probably safely catch up that analysis once the batch farm comes back to full strength. The burst advocate analysis gets launched a little more than once a day. Counting up the past week, we had 11 triggers in 7 days. Each trigger launches 6 jobs and each job goes to the medium queue using rhel6.

I can take the appropriate steps to deactivate the gravitational wave followup analysis and FAVA leading up to the outage. Let me know if you think we'd be able to keep the burst advocate analysis running and I'll take the appropriate actions.

Brian's proposal to move all VMs to H.A.:

I think we can move all fermilnx VMs to HA without oversubscribing memory or disk. Can we verify this?

* I think each fermilnx VM, except for fermilnx01 and fermilnx02, has 384GB memory.

* I think we have two VMWare Hypervisors in in HA.
I'd suggest distributing the VMs such that:
* fermilnx01 is on one hypervisor
* fermilnx02 is on another hypervisor (I think this is currently the case)
All other fermilnx-v* VMs are distributed between the other two hypervisors (live migration if possible)

Juliyana's report on VMs in H.A.:

Currently Fermi has these hypervisors and VMS in HA and Non-HA

HA:
Fermilnx-v06/v07/08
Fermilnx01/02/04

NON-HA:
Fermi-ci-test01
Fermilnx-v02/v03/v04/v05/v10/v11/v12/v13/v14/v15/v16/v17/v18/v19

These are fermi hypervisors. There are four of them. Two in HA and two in Non-HA.

HA: fermi-vmclust03 and fermivmclust04
Non-HA: fermi-vmclust01 and fermi-vmclust02

Excerpt from Christian Pama's spreadsheet on H.A. rack contents:

fermi-vmclust03
fermi-vmclust04
fermilnx03-vmm
fermi-oracle03
wain031
staas-gpfs50
staas-gpfs51

sca-oracle01
scalnx02-vmm
scalnx10-vmm01
scalnx11-vmm02
scalnx12-vmm01
scalnx13-vmm02

Current tomcat server VMs:

As of 11:30 on 5 Dec 2019

tomcat	VM	Hypervisor (obsolete)	Notes
01	FERMILNX-V07	FERMILNX03-VMM	H.A.
02	FERMILNX-V08	FERMILNX03-VMM	H.A.
03	FERMILNX-V12	FERMILNX07-VMM	
04	FERMILNX-V19	FERMILNX10-VMM	
05	FERMILNX-V13	FERMILNX08-VMM	
06	FERMILNX-V16	FERMILNX09-VMM	
07	GLASTLNX09		defunct
08	FERMILNX-V05	FERMILNX06-VMM	
09	FERMILNX-V17	FERMILNX09-VMM	
10	FERMILNX-V18	FERMILNX09-VMM	
11	FERMILNX-V10	FERMILNX07-VMM	
12	FERMILNX-V11	FERMILNX07-VMM	

Note that old 'kvm' hypervisors have been replaced by two VMware hypervisors: fermi-vmclust01/02 (non-HA) + fermi-vmclust03/04 (HA)

Other VM's:

VM	Hypervisor	Notes
fermilnx01		

fermilnx02		
fermilnx04		

Gotchas from the Dec 2017 outage

1. We did not specify the "medium" LSF queue in our requirements, but ended up needing it
 2. There was a delay in getting all 50 hequ's operating (some were/are on H.A., but others are not)
 3. The xrootd redirector had a problem and needed a restart
-