

FermiGrid Services Group

FermiCloud

Service Level Agreement

and Service Definitions

Steven Timm, Keith Chadwick
Last Revised November 5, 2012

Abstract:

This document is the Service Level Agreement (SLA) for the FermiCloud Services operated by the FermiGrid Services Group.

Document Revision History:

Version	Date	Author	Comments
0.1	03/28/11	Steven Timm	Initial Version
0.2	03/29/11	Steven Timm	Add service definitions per T. Currie
0.3	4/19/2011	Keith Chadwick	Minor edits
0.4	10/4/2011	Steven Timm	Adjust service definitions to economic model
0.5	11/5/2012	Steven Timm	Updates to reflect current status and discussions with Intensity Frontier

Service Level Agreement and Service Definition

Purpose:

This service level agreement document outlines the default service level roles, responsibilities, and objectives for the FermiCloud services operated by the FermiGrid Services Group. Individual FermiCloud stakeholders may negotiate higher levels of support if desired.

Scope of Services:

A number of types of cloud virtual machines are supported, as listed in the table below. A sample virtual machine and template to launch it with are provided for all classes of service. Currently all virtual machines are deployed using the command line interface of the OpenNebula virtual infrastructure manager. Use is restricted to people with an active Fermilab ID by means of requiring a Kerberos credential to log into the system where the virtual machines are launched. Fermilab developers currently are testing modifications of OpenNebula's emulation of the Amazon EC2 Query API (RESTful API) to use X.509 authentication and thus be able to deploy a restricted set of virtual machines via a web service, first on-site, and then possibly remotely pending security approval. OpenNebula also has its own web-based GUI which will be deployed when FermiCloud deploys their next release.

Service Offerings:

The following table lists the supported classes of service:

Service	Service Description	System Profile	Fulfillment
Account Provisioning	Account request for FermiCloud account provisioning (Required for standard provisioning)	Services and/or x.509 cert authenticated account with required profile.	X Days from request ticket submission
Opportunistic VM	This is for people who want to quickly try out FermiCloud. Can be pre-empted for more high-priority VM's.	<ul style="list-style-type: none">• 1 KVM Linux virtual machine with 1 CPU core• 2 GB RAM• 10 GB non-persistent disk (discard at exit)• a variable public IP	X Days from request ticket submission
Development/Integration (9x5 support)	For developers and integrators who need medium to long term virtual machines.	<ul style="list-style-type: none">• 1-5 KVM Linux virtual machines with 1-2 CPU cores and public IP's.• 2 GB RAM• 10 GB persistent disk (saved on exit and restorable).• Variable public IP standard, Static IP negotiable with FermiCloud support.	X Days from request ticket submission
Standard Production (24x7 support)	For systems which expect/require production level hosting and uptime (defined below).	<ul style="list-style-type: none">• KVM Linux virtual machine with 2 CPU cores and with fixed public IP's• 2 GB RAM	X Days from request ticket submission

		<ul style="list-style-type: none"> • 10 – 20 GB persistent disk (saved on exit and restorable). • Static IP standard. • Extra RAM available on request. 	
Custom Production (24x7 support)	Extra features including either higher density/volume of system requirements or resource on a dedicated basis that extends beyond the “Standard Production” service level mentioned above. Needs to be negotiated with the FermiCloud support staff within the FermiGrid Services Group. A task code or in-kind contributions to the project may be necessary.	<ul style="list-style-type: none"> • Xen hypervisor for high I/O loads, • X >8GB of RAM • X >2 CPU cores • Any clustered network configurations, • Any requirements of more than 20GB of persistent disk 	Negotiable
Image repository	Repository that stores saved virtual machines.		NA
EC2 query API	To be deployed on-site, date to be determined, off-site, date to be determined.		NA
Web GUI	To be deployed on-site, date to be determined.		NA

Service classes may only be added to this table with the prior agreement of the FermiCloud Project Leader and the corresponding line management (Grid and Cloud Computing Department Head).

Who Should Use FermiCloud:

- 1) People who need a virtual machine for a short amount of time to test.
- 2) Developers and integrators who need development and integration machines, both persistent and temporary.
- 3) Operators of legacy production services with modest CPU and I/O needs.
- 4) Operators of new services who need to deploy quickly without long lead time for hardware deployment.
- 5) Stakeholders who have temporary need for large amounts of virtual machines for large-scale systems testing.
- 6) Scientific experiments who have computing tasks that do not translate well to the Grid due to complexity of software stack or requirements for software that is not found on most Grid sites.

Current Users of FermiCloud:

- 1) Sysadmins doing various network and bluearc testing.
- 2) REX department staff making production SAMGrid forwarding nodes
- 3) OSG Storage group testing for Xrootd packaging

- 4) Grid and Cloud Computing Storage evaluation (Lustre, Hadoop, xrootd).
- 5) JDEM/WFIRST project, testing failure modes of distributed messaging system across a private net of several nodes.
- 6) Extenci project, testing wide-area Lustre
- 7) MCAS project, production server.
- 8) Scientist Survey Project, production web server.

Requesting FermiCloud Service

First-time FermiCloud users should submit a Service Desk Ticket of categorization Virtualization/FermiCloud. The ticket should describe which class(es) of service listed above they would like. Once an account on FermiCloud is established, the user can then launch virtual machines on demand. We will attempt to process requests for new accounts within one business day.

Service Monitoring and Service Availability Measurements:

The FermiCloud services consist of a Cluster Controller, a number of Virtual Machine Hosts. Monitoring of these physical machines and of the virtual hosts that run on them is available on the FermiCloud Ganglia server.

Service Availability Goals:

The formal service availability goal for the FermiCloud Services provided by the FermiGrid Services Group shall be 99% (measured on a weekly basis), providing that such a service availability goal is within the bounds of the reliability of the software as delivered by the software provider/developer and the operational requirements of FermiCloud.

Service Dependencies:

Any service dependencies (services operated elsewhere that depend on one or more of the Services operated by the FermiGrid Services Group) shall be explicitly documented and any such dependency shall be subject to formal authorization (signoff and agreement) by the FermiGrid Services Project Leader and the corresponding line management (Grid and Cloud Computing Department Head) prior to any such service dependency being moved to a production deployment. Should any service dependency be added without prior formal authorization, it will not be supported by the FermiGrid Services Group and may be subject to having any and all access terminated without warning to any Service operated by FermiGrid.

The current list of agreed service dependencies is:

Service	Agreed Service Dependencies
All	None

--	--

Service Constraints:

FermiCloud depends on the building power/environment, site network, DNS, NTP, and Kerberos servers. Some auxiliary files and templates are stored on Bluearc NFS server.

Service Outage Notification:

In the event of an unscheduled outage of the FermiCloud Services provided by the FermiGrid Services Group, clients shall utilize the Computing Division Service Desk procedures to notify the on-call personnel in the FermiGrid Services Group support rotation.

Test VM's may be killed at any time without notification of the user. Development and Integration persistent VM's will be suspended on reboot of VMM host and then resumed after that, saving the memory and process stacks of those VM's. Notification will be given in advance of this process. Users who need a highly available service are encouraged to have multiple VM's for this purpose. For standard and custom production servers in the cloud we will negotiate with the stakeholder to find a mutually-acceptable outage time, but also urge stakeholders to have multiple VM's for redundancy such that all of their VM's are not down simultaneously. Virtual Machine owners are responsible for maintaining their own security patches. FermiCloud staff will monitor to make sure this is being done.

Service Outage Response:

The FermiGrid support rotation that receives the incident notification shall attempt to respond to the incident within 30 minutes if the notification occurs during standard business hours (Monday through Friday 8:00 through 17:00). Off-hours incident response is not guaranteed and is subject to the availability of FermiGrid Services personnel.

Service Outage Response Escalation:

If neither the current FermiGrid primary or secondary respond within the response interval and the incident is reported within the standard business hours defined above, then the subsequent incident escalation shall be via the FermiGrid line management chain:

- FermiGrid Services Group Leader (or designee)
- Grid and Cloud Department Head (or designee)

Service Maintenance Schedules:

The following are the agreed maintenance schedules:

- Any time that the Feynman Computing Center is down.
- Any time that the Grid Computing Center Computer Rooms are down.
- Any time that is mandated by the Computing Division line management.
- Any time that automated processes run to perform routine functions (nightly log rotation, backup, database housecleaning/optimization, etc.).
- Any service that is maintained in a High-Availability (HA) configuration may have maintenance performed on the service, providing that a redundant copy of the service remains available while the maintenance is performed.
- Any time that is mutually agreeable to affected parties.

Notification of proposed maintenance periods shall be performed as far in advance of the proposed maintenance period as is possible.

Terms of Agreement:

The signatures on this document below indicate agreement to its content, that it is valid, has achievable objectives, and represents the intent of the FermiGrid Services Group to deliver the agreed set of Services.

Written notification of termination of this SLA by any party must be provided at least 60 days before expiration.

This document is controlled by Steven Timm, FermiCloud Project Manager. Any modifications to this document require the mutual consent of all parties together with notification to any affected clients.

This document will remain in effect until replaced with an updated version. It will be reviewed annually for currency, accuracy, and completeness. The next review is scheduled for 01-Apr-2012.

Signatures:

Steven Timm – FermiCloud Project Leader

Date