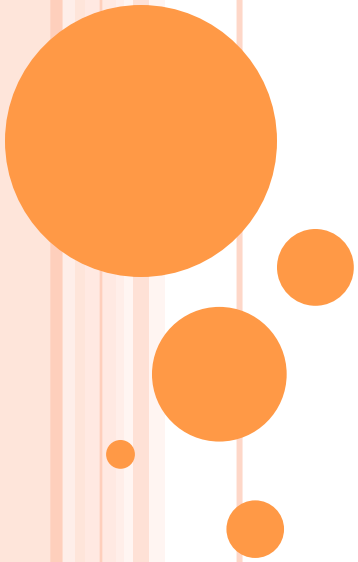


OSG Site installation and Maintenance

Suchandra Thapa
Computation Institute
University of Chicago

Introduction to OSG Terms and Operations

Introduction to OSG terms and operations
The OSG compute element
Installing an OSG site
Maintaining a site
Q&A time

Introduction to OSG

- OSG stands for Open Science Grid
- Provides high-throughput computing across US
  - Currently more than 75 sites
  - Recent stats:
    - 282,912 jobs for 433,051 hours
    - Used 75 sites
    - Jobs by ~20 different virtual organizations
    - 92% of jobs succeeded
    - Underestimate: 4 sites didn't report anything
  - Provides opportunistic computing for VOs
- Focus on high-throughput computing rather than high performance computing

Basic Terms

- CE – Compute Element
- SE – Storage Element
- VO – Virtual Organization
- WN – Worker Node
- GOC – Grid Operations Center
- VDT – Virtual Data Toolkit
- DN – Distinguished name
- VOMS – Virtual Organization Management Server
- GUMS – Grid User Management Server

OSG Organization: Virtual Organizations (VO)

- OSG is organized by Virtual Organizations
- A VO allows members of a collaboration or group to retain that same grouping on the OSG
- Each VO has different policies as to group membership and runs a VOMS server to track membership
- VOs are typically the owner of the various resources found on the OSG

Overriding principle: Autonomy

- Sites and VOs are autonomous
  - Admins are free to make decisions about site
  - OSG provides software and recommendations about configuration
  - Admins are allowed to decided when and if to upgrade
  - Admins are responsible for site but OSG provides operational support
  - VOs run and maintain own membership lists (on VOMS server)

YOUR ROLE AS AN ADMIN

- As a site admin, you should:
  - Keep in touch with OSG (downtime, security, etc.)
  - Respond to trouble tickets or inquiries from GOC
  - Plan your site's layout
  - Update software as needed (within limits)
  - Participate and be a good community member

SUPPORT PROVIDED FOR ADMINS

- OSG provides:
  - Software and ancillary information (configuration tools, documentation, recommendations)
  - Assistance in keeping site running smoothly
  - Help in troubleshooting and installing software
  - Users for your site
  - An exciting, cutting-edge, 21st-century collaborative distributed computing grid cloud buzzword-compliant environment

The OSG Compute Element

Introduction to OSG terms and operations
The OSG compute element
Installing an OSG site
Maintaining a site
Q&A time

OSG Software Stack

- Consists of:
  - VDT Software
    PLUS
  - Additional OSG Specific bits
- E.g. CE
  - VDT Subset
    - Globus
    - RSV
    - PRIMA
    - … and another dozen
  - OSG bits:
    - Information about OSG VOs
    - OSG configuration script (configure_osg.py)

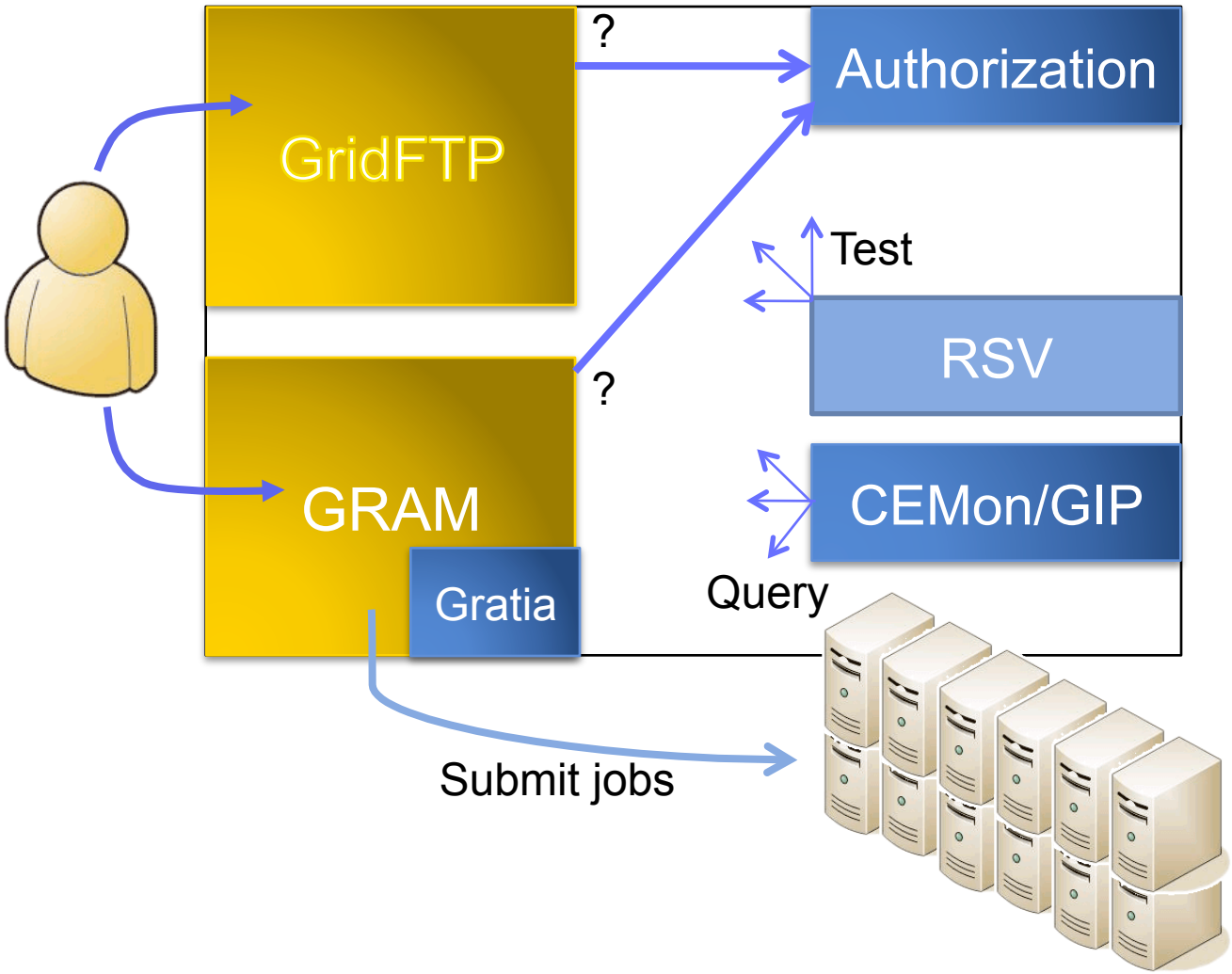
Overview of OSG components

- CE – Compute Element
  - Provides point of interface for tools attempting to run jobs or work on a cluster
  - Users submit jobs to this system
  - OSG provides a package that installs all software needed for this component
- SE – Storage Element
  - Several implementations
    - dCache
    - Bestman
  - Manages data and storage services on cluster
- WN – Worker Node
  - Software found on each compute node on grid
  - Provides software that incoming jobs may depend on (e.g. curl, srmcp, gsiftp, etc.)
- Client – Client Software
  - Provides software that users can use to submit and manage jobs and data on OSG
  - May be superseded by VO specific software
- Other tools (more specific and not necessarily used by many people)

5000 METER OVERVIEW OF CE

- GRAM : Allows job submissions and passes them on to local batch manager
- Gridftp : Provides data transfer services into and out of cluster
- CEMon / GIP : Provides information to central services
- Gratia : Sends accounting information on jobs run to central server
- RSV : Provides probes to monitor health of the CE
- User authorization : Needed to connect certificates to user accounts

Basic CE


GridFTP
?
Authorization
Test
RSV
?
GRAM
CEMon/GIP
Gratia
Query
Submit jobs

GRAM

- Two different flavors
  - OSG provides and supports both
  - Very different implementations
- GT2
  - What most users and VOs use
  - Very stable and well understood
  - On the other hand, fairly old
- GT4 (aka ws-gram)
  - Web services enabled job submission
  - Currently in transition
  - Used primarily by LIGO

GRATIA

- Collects information about what jobs have run on your site and by whom
- Hooks into GRAM and/or job manager
- Sends information to a central server
- Can connect and query central service to get reports and graphs
- Option exists for a local server

CEMon / GIP

- These work together
  - Essential for accurate information about your site
  - End-users see this information
- Generic Information Provider (GIP)
  - Scripts to scrape information about your site
  - Some information is dynamic (queue length)
  - Some is static (site name)
- CEMon
  - Reports information to OSG GOC's BDII
  - Reports to OSG Resource Selector (ReSS)

RSV

- System to run tests on various components of your site
- Presents a web page with red/green overview and links to more specific information on test results
- Optional interface to nagios
- Can be run on a server other than CE

Installing an OSG Site

Introduction to OSG terms and operations
The OSG compute element
Installing an OSG site
Maintaining a site
Q&A time

SITE PLANNING

- Bureaucratic details
- Cluster layout
- Disk layout / sharing
- Authorization

BUREAUCRACY

- Certificates (personal/host)
- VO registrations
- Registration with OSG
  - Need a site name (e.g. UC_ITB)
  - Need contacts (security, admin, etc.)
- Site policy on web

STARTING OUT

- Everyone using OSG gets a personal certificate because it is required to do any activity on an OSG resource
- Will need to know or contact someone with DOEGrids certificate in order to obtain a personal certificate

SITE REGISTRATION USING OIM

- Done using OIM at https://oim.grid.iu.edu/
- Will need to register first,
- After GOC approves registration :
  - Registrations > Resources > Add New Resource

Cluster Layout

- How is software / data being shared
  - NFS can work but gets bogged down with larger workloads
  - Where do services run?
    - Single server vs. dedicated servers
  - Worker node software?
    - Locally present on worker nodes vs. served over nfs
  - Certificates shared?

Required Directories for CE / Cluster

- **OSG_APP**: Store VO applications
  - Must be shared (usually NFS)
  - Must be writeable from CE, readable from WN
  - Must be usable by whole cluster
- **OSG_GRID**: Stores WN client software
  - May be shared or installed on each WN
  - May be read-only (no need for users to write)
  - Has a copy of CA Certs & CRLs, which must be up to date
- **OSG_WN_TMP:** temporary directory on worker node
  - May be static or dynamic
  - Must exist at start of job
  - Not guaranteed to be cleaned by batch system

Optional directories for CE

- **OSG_DATA**: Data shared between jobs
  - Must be writable from the worker nodes
  - Potentially massive performance requirements
  - Cluster file system can mitigate limitations with this file system
  - Performance & support varies widely among sites
  - 0177 permission on OSG_DATA (like /tmp)
- **Squid server**: HTTP proxy can assist many VOs and sites in reducing load
  - Reduces VO web server load
  - Efficient and reliable for site
  - Fairly low maintenance
  - Can help with CRL maintenance on worker nodes

Space Requirements

- Varies between VOs
  - Some VOs download all data & code per job (may be Squid assisted), and return data to VO per job.
  - Other VOs use hybrids of OSG_APP and/or OSG_DATA
- OSG_APP used by several VOs, not all.
  - 1 TB storage is reasonable
  - Serve from separate computer so heavy use won't affect other site services.
- OSG_DATA sees moderate usage.
  - 1 TB storage is reasonable
  - Serve it from separate computer so heavy use of OSG_DATA doesn't affect other site services.
- OSG_WN_TMP is not well managed by VOs and you should be aware of it.
  - ~100GB total local WN space
  - ~10GB per job slot.

Worker Node Storage

- Provide about 12GB per job slot
- Therefore 100GB for quad core 2 socket machine
- Not data critical, so can use RAID 0 or similar for good performance

AUTHORIZATION

- Two major setups:
  - Gridmap setup
    - File with list of mappings between DN and local account
    - Can be generated by edg-mkgridmap script
    - Doesn't handle users in mulitple VOs or with VOMS roles
  - Service with list of mappings (GUMS)
    - A little more complicated to setup
    - Centralizes mappings for entire site in single location
    - Handles complex cases better (e.g. blacklisting, roles, multiple VO membership)
    - Preferred for sites with more complex requirements
    - Ideally on dedicated system (can be VM)
    - Can add SAZ service for authorization

CE Installation Overview

- Prerequistes
  - Certificates
  - Users
- Installation
  - Pacman
- Configuration
- Getting things started

LOCAL ACCOUNTS

- You need following local accounts:
  - User for RSV
  - Daemon account used by most of VDT
  - Globus user is optional but will be used if found

PACMAN

- The OSG Software stack is installed with Pacman
  - Yes, custom installation software
- Why?
  - Mostly historical reasons
  - Makes multiple installations and non-root installations easy
- Why not?
  - It's different from what you're used to
  - It sometimes breaks in strange ways
  - Updates can be difficult (this is being improved)
- Will we always use Pacman?
  - Maybe
  - Investigating alternatives but changing existing infrastructure is hard
  - Work ongoing to support RPM/deb in the future

PACMAN CONTINUED

- Easy installation
  - Download pacman
  - Untar and source shell script
  - Start using binary
  - Look ma! No root!
- Gotcha:
  - Installs into current directory

BASIC INSTALLATION AND CONFIGURATION

- Install Pacman
  - Download
  - `http://physics.bu.edu/pacman/sample_cache/tarballs/pacman-3.26.tar.gz`
  - Untar (keep in own directory)
  - Source setup
- Make OSG directory
  - Example: /opt/osg symlink to /opt/osg-1.0
- Run pacman commands
  - Get CE (pacman –get OSG:ce)
  - Get job manager interface (pacman –get OSG:Globus-Condor-Setup)
- Configure
  - Run edg-mkgridmap or gums-host-cron
  - Configure CA certificates updater
  - Edit config.ini
  - Run configure_osg.py (configure-osg.py –c)
- Start services (vdt-control –on)

Installing CA Certificates

- The OSG installation will **not** install CA certificates by default
  - Users will not be able to access your site!
- To install CA certificates
  - Edit a configuration file to select what CA distribution you want
  - Edit `$VDT_LOCATION/vdt/etc/vdt-update-certs.conf`
  - Run a script
    - `source vdt-questions.sh`
    - `$VDT_LOCATION/vdt/sbin/vdt-setup-ca-certificates`

Site Maintenance

Introduction to OSG terms and operations
Installing an OSG site
Maintaining a site
Q&A time

UPDATING CAS

- CAs are regularly updated
  - New CAs added
  - Old CAs removed
  - Tweaks to existing CAs
- If you don't keep up to date:
  - May be unable to authenticate some user
  - May incorrectly accept some users
- Easy to keep up to date
  - vdt-update-certs
    - Runs once a day, gets latest CA certs

Monitoring Site Status

- Several tools available
- RSV
  - Part of install
  - Will present a web page with quick status update site functionality
  - Can integrate with nagios
- Nagios/Ganglia/Cacti
  - Presents information on non-grid specific details of cluster
  - Can set up alerts, pages, etc.
- Gratia
  - Provides accounting information on jobs running on your site
  - Useful to see who is using your site and how much utilization comes from various users
- Daily/Weekly email reports
  - Provides quick information on your site and osg at large at a glance

OSG RSV Status - Main
04-07-2009 15:33:07

Archived HTML pages

uct3-edge7.uchicago.edu

Probe	Metric	Last Executed	Status
osg-version-probe	org.osg.general.osg-version	2009-04-07 15:21:05 CDT	OK
ping-host-probe	org.osg.general.ping-host	2009-04-07 15:27:01 CDT	OK
cacert-crl-expiry-probe	org.osg.certificates.cacert-expiry	2009-04-07 13:48:16 CDT	OK
jobmanagers-status-probe	org.osg.batch.jobmanager-default-status	2009-04-07 14:48:09 CDT	OK
gridftp-simple-probe	org.osg.globus.gridftp-simple	2009-04-07 15:09:02 CDT	OK
vo-supported-probe	org.osg.general.vo-supported	2009-04-07 14:24:14 CDT	OK
gram-authentication-probe	org.osg.globus.gram-authentication	2009-04-07 15:31:01 CDT	OK
cacert-crl-expiry-probe	org.osg.certificates.crl-expiry	2009-04-07 14:48:20 CDT	WARNING
certificate-expiry-local-probe	org.osg.local.httpcert-expiry	2009-04-07 12:06:00 CDT	OK
certificate-expiry-local-probe	org.osg.local.containercert-expiry	2009-04-07 12:29:01 CDT	OK
certificate-expiry-local-probe	org.osg.local.hostcert-expiry	2009-04-07 12:43:00 CDT	OK
vdt-version-probe	org.osg.general.vdt-version	2009-04-07 06:47:08 CDT	OK
cacert-expiry-local-probe	org.osg.local.cacerts-expiry	2009-04-07 12:58:02 CDT	OK
jobmanagers-available-probe	org.osg.batch.jobmanagers-available	2009-04-07 04:59:05 CDT	OK
osg-directories-probe	org.osg.general.osg-directories-CE-permissions	2009-04-07 06:05:07 CDT	OK

uct3-edge6.uchicago.edu

RSV OUTPUT FOR UC_ITB RESOURCE

Web page generated by RSV showing the output of various probes. Clicking on the probe output will give history for last few invocations and error output

	Nodes status			CPUs
Farms	Idle nodes (load < 0.5)	Active nodes (load > 0.5)	Total nodes	Number of CPUs
18. OU_OCHEP_SWT2	13	52	65	-
19. OU_OSCER_ATLAS	425	104	529	-
20. OU_OSCER_CONDOR	313	17	330	-
21. OUHEP_OSG	1	39	40	-
22. PolyHub_UT	-	-	-	-
23. SBGrid-Harvard-East	8	13	21	-
24. SBGrid-Harvard-Exp	5	4	9	-
25. SPRACE	54	32	86	-
26. TTU-ANTAEUS	0	1	1	-
27. UCHC_CBG	16	0	16	-
28. UColorado_HEP	9	2	11	-
29. UCSDT2	4	93	97	-
30. UCSDT2-B	4	93	97	-
31. UmissHEP	9	13	22	-
32. UNM_HPC	-	-	-	-
Total	1093	1547	2640	0

Back to top

Running Jobs Statistics

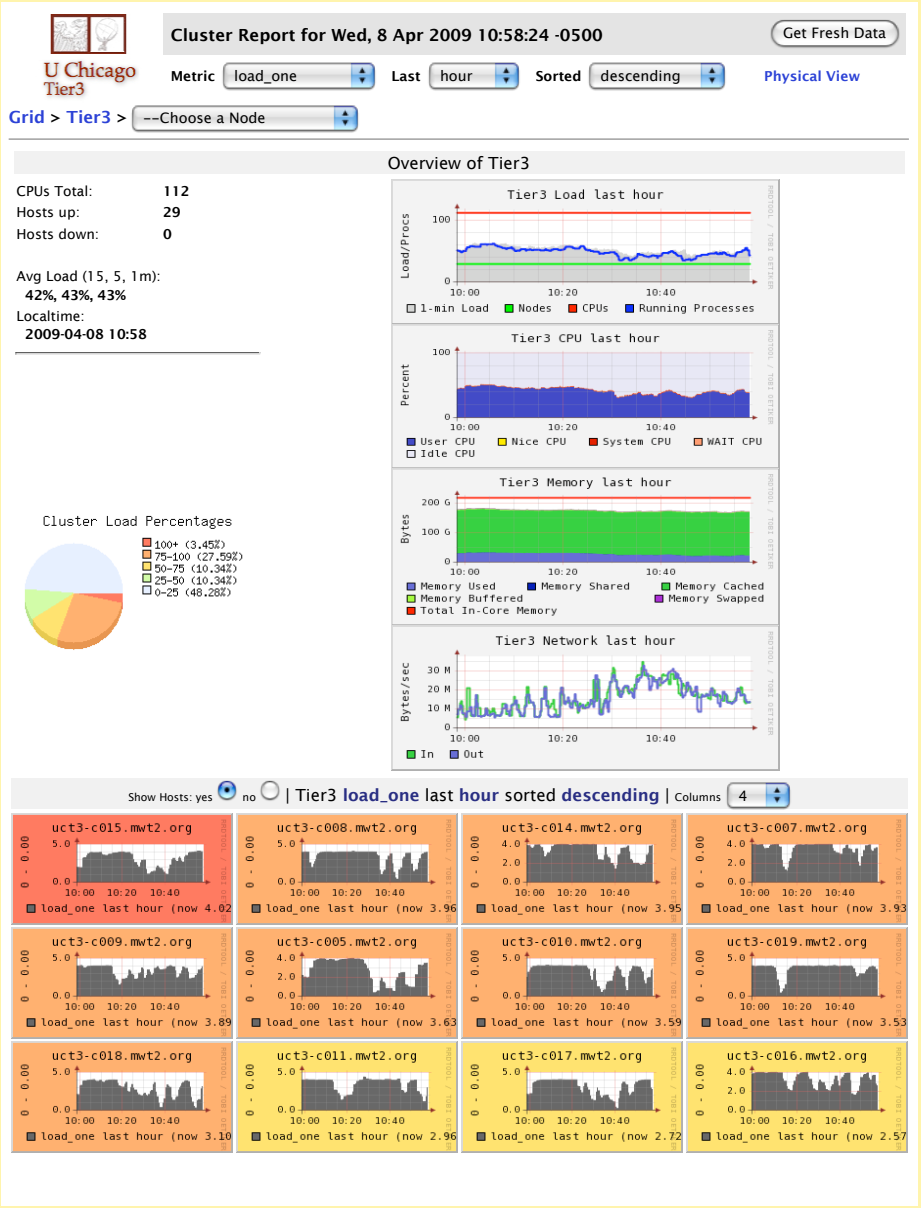
Integrated Running Jobs - last day

Running Jobs * hours															
Farms \ VOs (**)	ATLAS	CDF	CMS	DOSAR	DZERO	ENGAGE	FERMILAB	GEANT4	GLOW	LIGO	MIS	NANOHUB	RSV	SBGRID	Total
1. AGLT2	2496	-	-	-	-	-	-	-	-	-	0	-	-	-	2496
2. cinvestav	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. CIT_CMS_T2	0.8	-	10458	-	80	-	-	0.1	-	10	-	48	-	-	10597
4. FLTECH	-	-	1819	-	-	-	-	8	20	-	-	-	-	-	1847
5. FSU-HEP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. GLOW	2	2	8842	-	0	0.1	-	0	46	0	-	2	-	-	8894
7. gpnstor-MU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. GRASE-GENESEO-ROCKS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9. HEPGRID_UERJ_OSG64	4	-	305	-	-	-	-	-	-	-	-	-	-	-	309
10. Lehigh_coral	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. LONI_OSG1	-	-	-	-	1535	-	-	-	-	-	-	-	-	-	1535
12. LTU_OSG	-	-	-	-	254	-	-	-	-	-	-	-	-	-	254
13. MWT2_IU	1820	-	-	-	5850	-	-	-	-	-	-	-	-	-	7670
14. NWICG_NDCCL	-	-	-	-	-	17	-	-	-	-	-	45	-	-	62
15. NWICG_NotreDame	-	-	-	-	-	1	-	0.9	-	-	-	0.1	-	-	2
16. NYSGrid-CLUSTER04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17. NYSGRID-CUNY-GRID	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18. OU_OCHEP_SWT2	1110	-	-	-	5067	-	-	-	-	-	-	-	-	-	6177
19. OU_OSCER_ATLAS	50	-	-	7352	-	-	-	-	-	-	-	-	-	-	7402
20. OU_OSCER_CONDOR	-	-	-	-	7619	-	-	-	-	-	-	-	-	-	7619
21. OUHEP_OSG	-	-	-	-	1072	-	-	-	-	26	-	-	-	-	1098
22. PolyHub_UT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23. SBGrid-Harvard-East	69	-	-	-	-	2	-	3	-	546	-	54	-	5	679
24. SBGrid-Harvard-Exp	-	-	-	-	-	4	2	11	-	0	0.6	77	-	13	108
25. SPRACE	11	-	133	-	2921	-	0.4	-	-	1824	0.1	18	-	-	4908
26. TTU-ANTAEUS	6	-	11	-	48	13	0	3	-	-	-	20	4	-	105
27. UCHC_CBG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28. UColorado_HEP	-	-	123	-	-	-	-	-	-	-	0.1	-	-	-	123
29. UCSDT2	4	1313	7996	-	464	14	-	0.5	313	-	-	6	-	-	10111
30. UCSDT2-B	2	1251	1181	-	-	19	-	0	323	-	-	5	-	-	2781
Total	5575	2566	30868	7352	25474	70	2	27	702	2847	0.8	275	4	18	75782

Daily Email Report on Production Sites

Example of the daily email sent out to administrators with information on jobs and sites over the last day

This shows the daily usage by VOs of the Fermigrid resource over the last month as VO validations were run



Ganglia Output for Cluster at UC

Top level information on the servers and compute nodes at a small cluster at the University of Chicago, clicking on hosts will allow more detailed information on each host to be obtained

Incremental Updates

- Frequent (Every 1-4 weeks)
- Can be done within a single installation
- Either manually:
  - Turn off services
  - Backup installation directory
  - Perform update (move configuration files, pacman updates, etc.)
  - Re-enable services
- Or use vdt-updater script

Major Updates

- Irregular (Every 6-12 months)
- Must be a new installation
- Can copy configuration from old installation
- Process:
  - Point to old install
  - Perform new install
  - Turn off old services
  - Turn on new services

Questions? Thoughts? Comments?

Introduction to OSG terms and operations
Installing an OSG site
Maintaining a site
Q&A time

Acknowledgements

- Alain Roy
- Terrence Martin