



University of Westminster

Centre for Parallel Computing University of Westminster

**Service Level Description
for
the UK National Grid Service Resource
University of Westminster**

Release 3.2

10 October 2008

1. Overview of the Service

1.1 Background

The UK National Grid Service (NGS) is the core UK e-Science Grid, intended for the production use of computational and data Grid resources. NGS is supported by JISC and EPSRC. More information is available at the official NGS website – <http://www.ngs.ac.uk>.

The Service Level Description defines the “Service” that the University of Westminster contributes to the NGS.

1.2 Aim of the Service

The aim of the service is to deliver the service and resources as described in this SLD document to the NGS user community.

This SLD document shall take precedence over all other documentation, except where the NGS management team explicitly declares otherwise.

1.3. The Service

The Service refers to the resources located at the Centre for Parallel Computing and supporting services provided by the Centre, as described here.

Applies to

The CARMEN cluster includes 328 AMD Opteron rev F cores of which 64 compute nodes, 18 head nodes/Grid nodes/Portal nodes, 11TB storage and Infiniband. All hardware is powered by a 60KVA three-phase UPS.

The cluster runs the Suse Linux Enterprise 10. The main head node acts as a front-end to compute nodes which are used for executing jobs submitted via Globus, the appropriate job manager, and the Maui/Torque scheduling system.

The Carmen cluster provides 11TB of useable storage. The storage is a RAID5 SAN using resilient fiber channel. There are two storage servers configured with dynamic load balancing and fail over managed using the IBM GPFS filesystem. The GPFS filesystem is accessible using RDMA (Native Infiniband protocol) for maximum performance and via the gigabit Ethernet network. Quotas are managed in LDAP and enabled using GPFS quota support.

5 head nodes and 15 compute nodes are configured with the Xen hypervisor to support virtualisation.

Cluster head nodes/Grid nodes/Portal nodes:

- 18 x IBM x3455 each featuring 2x Dual core AMD Opteron 2281 Rev F 2.6GHz processors , 2x 80Gb hard drives (mirrored) and 4GB Memory, 6 head nodes fitted with Infiniband adaptors, IPMI2

Compute nodes

- 64 x IBM x3455 each featuring 2x Dual core AMD Opteron 2218 Rev F 2.6GHz processors , 1x 80Gb hard drive and 4GB Memory, Infiniband adaptor fitted, IPMI2

Data:

- 2 x IBM System Storage DS4200, 4Gbps fiber channel
- 2 x IBM SAN16B 16-port Fibre Switch
- 2 x 16 500GB SATA2 hard drives (11TB useable)
- 2 x IBM eServer x346, 1x Intel Xeon Dual core 2.8GHz EM64T processor, 2x 36GB SCSI 10K rpm Hard Drives (Mirrored), Voltaire Infiniband
- 4 x Qlogic Single Channel Fibre HBA (4Gb), PCI-x
- Cisco 10Gbs Infiniband switches (for high performance interconnection)
- Gigabit switches (for admin)

Power:

- UPS: 60KVA / 3 phase with 30mn runtime @ 30KVA, Internal and external bypass
- Intelligent PDU for powering nodes in sequence

Cooling:

- 2 x Airdale 30KW units
- 3 x wall mounted 10KW units for backup

2. Service Provider

The Centre for Parallel Computing (CPC) at the University of Westminster operates the service on behalf of the NGS management team for the benefit of the NGS user community.

3. Service Users

The Users refers to all approved users of The Service. The Service is free for academic users. Business or commercial users should discuss their requirements directly with the Service Provider.

All users of the service are required to accept the *Regulations for Use of the NGS* stated at:

<http://www.ngs.ac.uk/NGS-tacu.html>

4. Service Inclusions

- 1) The following grid middleware and related software for the support of the UK research community is installed:
 - a) Pre-WS and WS components of the Globus Toolkit from the Virtual Data Toolkit (VDT 1.8.1) release with gLite/VOMS extensions, including GRAM2, GRAM4, GridFTP, RFT, GT4 MDS, GSIsSh and GRIS reporting to the central BDII.
 - b) Maui/Torque job manager
 - c) RUS client for reporting usage accounting statistics
 - d) Storage Resource Broker (SRB) client
 - e) Scripts supporting the Uniform Execution Environment (UEE)
 - f) Grid-mapfile populated from the central NGS VOMS server using Westminster's GridMapGen program to create one gridmapfile for Pre-WS and WS components.
 - g) GEMLCA application repository
 - h) PGRADE portal
- 2) The following software for developing and running programs is available:

- a) AMD Core Math Library (ACML)
 - b) openmpi message passing library (Category 2 support)
 - b) Java - Various versions of Sun's Java development environment
 - c) GNU compiler suite (including gcc and gfortran)
- 3) The following applications packages are currently available:
- a) Amber (licence required; category 3 support)
 - b) FFTW (Category 3 support)
 - c) R (Category 3 support)
 - d) Blender
 - e) MGLTools
 - f) AutoGrid + Autodock

The UoW NGS support staff is ready to install further application packages based on users' requests considering efforts and resources required for the installation.

- 4. Digital certificates management: accept certificates issued by the UK e-Science Certificate Authority and those CAs with which the UK e-Science Core Programme has agreements. The Certificate Revocation Lists are updated on a regular basis.
- 5. Information: *User Guide* for users of all the NGS systems is available from <http://www.ngs.ac.uk>, and standard Linux on-line HELP information is available on the system. Further information for this node is available at: <http://www.ngs.ac.uk/sites/westminster>
- 6. Provision of usernames: usernames from ngs0001 to ngs2000 will be provided.
- 7. Provision of disk space for users' files and backup of users's files.
- 8. Management of overall disk space to ensure that free space is available
- 9. Nagios is used for tracking faults and email alerts to system administrators.
- 10. Usage accounting records are uploaded to the NGS RUS service.
- 11. Batch job control: a job queuing system is operated on the cluster, and the successful operation of the system is monitored closely.
- 12. Specialist advice is available to research users who are assumed to be familiar with UNIX and general programming techniques.
- 13. Systems support - The system software products are supplied under licences specifying conditions on their use and providing a fault reporting and correction service through local support and OCF/IBM support Centres.

5. Service exclusions

- 1. Turnaround time cannot be guaranteed, as the system is heavily utilised and incoming jobs are queued waiting for existing jobs to finish.

6. Service Level

Quality

1. Ganglia (<http://ganglia.ngs.wmin.ac.uk>) results and Inca (<http://inca2.ngs.ac.uk>) results are published regularly and monitored to identify problems.
2. The Carmen cluster runs the Suse Linux Enterprise 10 operating systems and versions are normally kept as up-to-date as possible. The Novell Suse licence agreement includes patches and updates.
3. Systems software problems are dealt locally and can be reported to the OCF/IBM Support Centre; hardware maintenance contract is in effect on this equipment.

Availability

1. The Carmen cluster will be available at all times subject to: essential planned maintenance to hardware or software; the *At Risk* period; and unplanned stoppages and failures.
2. Unplanned stoppages and failures out of hours will be dealt with next working day; hardware support reflects the contract with IBM and OCF.
3. Two weeks' notice is usually given to users (via the NGS web site or the appropriate emailing list) about scheduled maintenance.

Reliability

1. Inca node monitoring results are displayed via NGS Web site at <http://inca2.ngs.ac.uk>
2. The reliability of the service is monitored by the Nagios system which alerts Westminster system administrators if the node or the service goes down.
3. The reliability of the service is monitored and the availability is presented by the Ganglia system (<http://ganglia.ngs.wmin.ac.uk/>).

Computational Capacity

1. Fair-share scheduling policy has been installed on the Carmen cluster. This scheduling policy defines job's priority considering users' historical resource utilisation information, i.e. those users who access the cluster more frequently have lower priorities than users with lower usage history. The scheduling policy decreases priority of users' jobs who used lot of resources over a given time interval and increases priority of users' jobs who have not received enough resources. The scheduling policy assigns higher priorities for Westminster users than for any other users.
2. Turnaround time cannot be guaranteed, as the systems may be heavily utilised and computations may wait in a queue until a processor becomes available.
3. The Service Provider may suspend or terminate:
 - Computations started through the batch queue system that run for longer than 24 hours (measured in wall-clock time);
 - Processes started through any means other than the batch queue system that run for longer than 24 hours (measured in wall-clock time) or longer than 15 minutes (measured in CPU time);
 - Computations that have the potential to adversely affect the system.

File Storage

1. Users will have access to a home file system on the shared disk space. They will also have temporary file space. Temporary files should also be stored in the scratch area under /scratch, which is not backed up, and must not be stored in the /tmp directory of any local server. There is no provision of permanent file space available on UoW nodes.

2. Users are allocated a soft disk quota of 1 Gb with a hard limit of 1.5 Gb and a grace period of 7 days.
3. If the User does not access his/her account for 72 hours the Service Provider may remove all files in the account. The Service Provider may also remove User files on the local disk if the User has no processes running on the server, or if the files have the potential to adversely affect the system.
4. The Service Provider will attempt to restore data in the event of the system crash.
5. Full incremental backups of the GPFS home directories occur 7 days per week with the current and previous version of each file stored in the backup repository. Once a file has been deleted it will be held in the backup repository for 90 days.
6. The backup regime is designed to protect the system from RAID failures and not as a facility to restore user files in the case of accidental deletion. However, user requests to restore individual files will be satisfied where possible as will normally be dealt with within five working days of the request.
7. Users' GASS-CACHE files that have not been accessed for 1 month will be deleted.

Network Capacity

1. Network access is provided on a best-effort basis. No guarantee is provided for bandwidth or other network metrics, either between nodes within the system, or from the system to any other system. The Service Provider may suspend or terminate computations that transfer more than 1 GB.
2. No network connectivity is provided between the execution nodes and external services. Network connectivity to external services is provided only through the head node.

7. Compliance

1. Availability statistics are published via Ganglia (see above) and Inca (see above).

8. Operational Framework

1. User problems and queries should be reported to the NGS Helpdesk (<http://www.gridsupport.ac.uk/>) either by phoning 01235 446822 or emailing support@gridsupport.ac.uk. The NGS Helpdesk will forward details of queries to the Westminster personnel for Westminster node specific issues.
2. Additional disk resources for individuals (typically up to twice the defaults) may be allocated following application.

9. Change Control

New releases of system software are generally agreed in advance with other sites at the NGS Operation Team meetings. Users are informed prior to upgrading taking place via the appropriate NGS emailing list.

10. Support Category Definitions for Software

There will be no official software support.

The UoW NGS node support staff may be contacted at the support@grid-support.ac.uk e-mail address if significant problems occur. The support staff will investigate each reported problem and will offer the required support based on their availability.

11. Definitions

1. *head node* - the front end machine for job submission and routine tasks.
2. *compute nodes* - the main cluster machines which run work through workload scheduling systems.
3. *NGS web site* - www.ngs.ac.uk.
4. *Globus* – a toolkit providing a set of services to enable grid applications and the underlying grid infrastructure to inter-operate.
5. *GridFTP* – an FTP service with extensions to meet requirements of grid environments
6. *GSIssh* – a version of ssh (secure shell) that uses the Globus Security Infrastructure (GSI) for authentication with X509 certificates.
7. *MDS – Monitoring and Discovery Service*: a way of querying a grid service about its configuration and status.
8. *PBS – Portable Batch System*: a system to organise jobs into different streams and assign different resources and priorities to these streams so as to maintain an optimum mix of jobs on the cluster.
9. *GASS – Global Access to Secondary Storage*: Libraries and utilities for file I/O to the Globus environment
10. *GASS-CACHE* – Local file cache in users' home directories used by GASS
11. *hard disk quota* - an absolute limit that cannot be exceeded
12. *soft disk quota* - a limit which can be temporarily exceeded, provided the disk holding returns below the soft limit within 7 days