Production Quality e-Science Grid

Andrew Richards (a.j.richards@rl.ac.uk) CCLRC; Joanna Schmidt (j.g.schmidt@leeds.ac.uk) White Rose Grid; P M Dew (dew@comp.leeds.ac.uk) University of Leeds; Fred Youhanaie (fred.youhanaie@computing-services.oxford.ac.uk) University of Oxford; Matt Ford (matt.ford@man.ac.uk) University of Manchester; Neil Geddes (n.i.geddes@rl.ac.uk) Grid Operations Support Centre

Abstract

This paper reports on the progress made to date with the establishment of the National Grid Service, the core of which is based on the nodes recently acquired from funding allocated by the JCSR and CCLRC. Following the analysis of users’ initial requirements, the UK e-Science Grid was set up through the collaborative efforts of the four involved sites and the HPC(x) and CSAR services using the support of, and experience gained by, the Engineering Task Force and the White Rose Grid. The paper outlines the hardware and software configuration, including grid middleware, of compute and data clusters that offer a production quality grid service for use by UK academia. Furthermore our experience of selecting products for a usable and stable service is reported. Finally, it highlights some issues related to convergence interoperability with other grids, and in particular with EGEE.

1. Introduction

Over the last four years significant OST and DTI investments (approx £240m) have supported the development of e-Science in the UK. EPSRC, on behalf of all of the Research Councils, is managing the UK e-Science Core Programme which is coordinating developments in e-Science for the benefit of the full range of UK Science. The emphasis is on a broad range of e-Science projects in science and engineering as well as on the development of generic grid middleware designed to enable e-Science. New e-Science demonstrators and projects have been followed by support for many other e-Science activities such as international collaborations, and including support activities for the UK e-Science community which importantly encompass the establishment of the National Grid Service (NGS). This new facility offers a production quality grid service to UK academics. This paper reports on the substantial progress made to date with building this service, conveys the experience gained while constructing the UK Production Grid and highlights some issues pertinent to the NGS, which at present, two months into the service, already has over 70 users registered.

Initially new e-Science projects could use the Level 2 UK Grid facilities established successfully by the Engineering Task Force (ETF) with computational resources provided by the network of e-Science Centres supported by the UK Core Programme. However, with the increase in the number of e-Science projects and of concurrent major developments, there was an identifiable need to establish a production-quality service for UK researchers. Last year the Joint Information Systems Committee for Support of Research (JSCR) funded three Beowulf type clusters. These, together with a cluster funded by CCLRC and the two nodes running the national HPC service (HPCx and CSAR), have formed the basis of the new UK Production Grid, named the National Grid Service. The four sites that operate the clusters are the White Rose Grid (WRG) at Leeds, CCLRC-RAL, Oxford, and Manchester. To meet diverse users’ requirements for both types of grids: viz for a data grid offering large storage, and for a compute grid providing extensive computational power, two types of clusters with different specifications (see below) have been acquired: data clusters (at Manchester and CCLRC-RAL) and compute clusters (at Oxford and WRG at Leeds).

This paper covers issues related to the setting up of, and resources and services offered by, the four new clusters. It does not include detailed descriptions of grid aspects of CSAR and HPCx service nodes, which have their own funding arrangements and user policies. Nor does it include other Open Grid sites,
which are already negotiating to commit resources to and become part of the NGS, mainly due to the limitations of space. Furthermore the paper reports our early experience and covers some technical issues and does not set out to discuss all issues relevant to the NGS.

The paper comprises eleven sections. After the introduction, the first two sections are concerned with the objectives, delivery and growth of the service. The next section, about technical aspects of the NGS, outlines the basic hardware and software stack and reports on our selection of grid middleware installed on these four systems. This part highlights the experience gained through building the grid which is being deployed in such a way as to be largely convergent with others grids in the future. The next sub-section it about data management. It is followed by a section on the NGS monitoring and reliability, and a section describing additional services planned on the NGS. In the seventh section user management issues are summarised. Following that there is a section on gaining access to the NGS, and then a paragraph on convergence with other grids. Finally, the tenth section elaborates on the use of the NGS in these early days of the service, and then the last section offers concluding comments.

2. Objectives

The principal goal is to provide a reliable grid service to all users and not only to those engaged in current e-Science projects. Importantly the service is expected to reach beyond these groups to researchers who require significant computational resources for their projects, who wish to experiment with grid-enabling their applications, or who need to develop and support inter-institutional collaborations. The NGS is intended to be the service that allows users to effect science using grid technologies.

The objectives of the NGS are manifold, and are structured in these three main areas: the delivery of a stable grid service, the provision of computational resources, and the engagement with users as well as support teams.

The National Grid Service is operated by the Grid Operations Support Centre (GOSC), under the guidance of the Director (Neil Geddes) to achieve the following:

• building on recent advances in grid technology to offer a stable highly-available production quality grid service in accordance with its Service Level Descriptions
• to ensure where possible its convergence interoperability with other grids such as those deployed by the Enabling Grids for e-Science in Europe (EGEE) project and the LHC Computing Grid (LCG)
• to provide the broad user community with access to large scale resources through the grid interface
• to meet users’ evolving requirements for this production grid service by working with the user community, through the User Group established by the Core Programme
• to cooperate effectively with the Engineering Task Force and other teams

The ultimate vision is to meet the large-scale computing needs of the UK academic community through a deployed grid service that will integrate and grow through adding new computational assets providing flexibility and the ability to extend different technologies as they become available.

3. Delivery and Growth of the Grid Service

The NGS is a collection of nodes, each of which offers a broad set of resources, delivering jointly the grid service. This service operates in production mode, described through an individual Service Level Description (SLD) document for each node as well as a separate document stating common grid resources and services for the four clusters.

The core service of the NGS (four clusters plus the two national HPC services) is extended further by encompassing assets from those UK e-Science Centres that have offered computational resources, agreed through a separate SLD for each site, to the NGS. These resources are part of the UK e-Science Open Grid (Figure 1). In these early days the NGS will be further enhanced within the UK e-Science Open Grid with the processing capacity offered by clusters of SGI 300 servers in Cardiff (i.e. Welsh e-Science Centre) and a large number (approximately 600) of clustered PCs at the University of Bristol. These all nodes provide significant heterogeneous computational facilities for UK researchers.
The NGS’s support team, though geographically distributed across the participating sites, collaborates effectively in setting up clusters and offering the service for users. The Operational Team which has built the Grid follows up the procedures defined by the Management Team encompassing staff from all eight sites. Both teams work together in face-to-face meetings, frequently utilising Access Grid nodes and using discussion lists.

4. Basic Hardware and Software Stack

All new clusters are ClusterVision’s state-of-the-art high performance Beowulf-type systems offering significant federated computational power and large data storage. Each compute facility comprises 64 dual CPU Intel 3.06GHz nodes and 4TB storage, whilst each data facility is configured with 20 dual CPU Intel 3.06GHz nodes and 18TB Fibre SAN storage, giving a total of 44 TB of storage across all nodes available to users of the NGS.

In building two different types of services (i.e. computational and data services), it is envisaged that different types of research areas would be attracted to use the NGS. Thus, the data nodes have not been designated to act as just a data repository for the users of the compute nodes. Additionally they allow for data centric projects to take advantage of a large-scale data facility coupled with high performance computing; for example to perform tasks such as data mining of large data-sets that would normally require the large volumes of data to be first transferred to the location of the compute service.

4.1 Software Development Environments and Applications

The basic software development environment includes three sets of C/Fortran compilers (GNU, Intel, and Portland Group), Java, Intel Math Kernel Library and the Etnus Totalview parallel debugger. For those users developing parallel programs compute clusters offer MPICH with both gm (Myrinet) and P4 (Ethernet) drivers. The data clusters offer Oracle 9i RAC.

The professional edition of the PBS batch scheduler (PBSpro) is installed on the clusters. This is a flexible workload management system which offers control of computing resources. It provides many features and benefits, including time-of-day-based advanced reservation and specifying job dependencies. The scheduler comes with a graphical user interface, called xpbs, which runs under the X-Windows system. The product includes the integration scripts to Globus Toolkit (GT), though these have not been utilised. Instead PBSpro has been integrated with GT through the PBS job manager developed by Imperial College, London, because of the need to introduce several modifications to its settings.
The clusters application software base is not intended to be homogeneous across the data and compute nodes. In response to users’ requests it is being individually enhanced to meet specific requirements, where software licences are available. At present it includes GROMACS, ScalAPACK/PBLAS, MPiblacs, Gaussian, NAG libraries, BLAST, FASTA, DL_POLY and FFTW.

4.2 Grid Middleware Deployment

The choice of Grid middleware for the NGS is based around Globus version 2. This was selected as the core middleware in order to follow on from the existing work done on the Level 2 Grid project. However, with continuing changes in the development of Globus and the long term future of GT2 support it has been decided that the Virtual Data Toolkit (VDT) implementation of the Globus middleware would be adopted. The VDT offers an ensemble of grid middleware that can be easily installed and configured. This decision was taken because the VDT release provides the NGS with a more robust middleware stack to which any security updates and patches available are applied on top of the standard GT2 distribution. The use of VDT also provides for a closer working alliance in the future with other grid projects such as EGEE.

The NGS uses the Globus Monitoring and Discovery Service (through GIIS and GRIS) to provide information on the status of grid nodes. The information is propagated in a hierarchy. Individual grid nodes, which run GRIS (Grid Resource Information Service) collect and send resource information to a local site Grid Index Information Service (GIIS) from which in turn the BDII (Berkeley Database Information Index) [1] caches node status information. BDII uses standard OpenLDAP commands and strictly validates all data prior to importing. If there are any invalid data such as empty values or invalid entries, they are rejected. The central UK e-Science MDS, maintained at the GOSC, holds information from the two top-level GIISes (ginfo, used by the UK e-Science Grid, and ginfo-dev which is the top-level GIIS for the NGS).

To enable our users to carry out interactively some of the necessary developmental tasks, the GSI enabled version of the secure shell (gsiss) is being offered on the nodes. The gssih uses grid-mapfile to authenticate users.

4.3 Data Management

To transfer input and output files to and from nodes, the GridFTP has been installed on NGS systems to facilitate data transfers on the grid.

The San Diego Supercomputing Centre Storage Resource Broker (SRB) is used on clusters to provide a uniform API for connection to heterogeneous resources and to access data sets that may be replicated. SRB servers are installed on all four clusters with MCAT (Metadata Catalogue) servers run by the data clusters. The MCAT data are kept in an Oracle database. The compute clusters host a vault of 500GB each whereas the data clusters provide a vault of 2TB each. The SRB command line utilities, Scommands, are available for accessing the servers.

In addition an OGSA-DAI interface to the datastore at Manchester is also being provided for initial evaluation purposes, after which it is planned to install OGSA-DAI on both data clusters. The challenge here is that we are running GT2, chosen because of its stability, and OGSA-DAI is a GT3 component and it is not as stable as one would wish to install on a system that provides production level services.

5. NGS Monitoring and Reliability

To provide a production quality grid one of the key requirements is to be able to monitor the overall service and have procedures in place to deal with any arising issues. This involves monitoring at all levels from the hardware level through to the service level and providing a helpdesk structure to enable both external users and internal system administrators to raise and track queries.

In order to monitor the hardware, temperature, system failures, CPU load, memory usage etc, both Ganglia and Nagios are installed and used on each of the four nodes.

Ganglia is a distributed monitoring system for high-performance computing systems such as clusters and now grids. This open-source software has been ported to a broad set of operating systems and processor architectures.
Nagios is an open source host, service and network monitoring program. The monitoring daemon runs intermittent checks on individual hosts and services using external plugins which return status information to Nagios. If required the daemon can send notifications out to an administrator by different means (email, SMS, etc.). Furthermore, current status information and historical logs as well as reports may be accessed via a web browser.

Whilst Nagios is used for internal, individual site usage, Ganglia has been implemented as a federated service, whereby information is collected at each site and fed upwards to a central Ganglia server which can in turn be queried to view the overall status of the NGS service (see Figure2).

In addition, external monitoring tools provide information about the status of the service and related items such as network monitoring. For example, GridMon [2] is used to monitor the status of the network connection between the NGS sites, whilst software such as GridICE is used to monitor the information gathered from the BDII information service.

Currently Grid Integration Test Script (GITS) suite is run against all four clusters by all four sites to monitor the interoperation of grid nodes. The results of these regular point-to-point tests (e.g. see Figure 3) are published on the NGS web site. Furthermore they are delivered in an appropriate format for use with the GITS Web service. This permits the results generated by the GITS tests to be stored in a central database hosted at Manchester for inspection of historical data when required in the future.
6. Additional Services

As the service continues to grow, driven by user demands, detailed study and investigations need to be carried out to gain a full understanding of how users may develop, deploy and manage services (such as Web services and Grid services) in a container or hosting environment, and how they can publish and maintain data through DBMS on this service. Whilst the NGS itself is not aiming to be a developmental system it will continue to work closely with sister projects such as the ETF to evaluate and overcome new challenges.

7. User Management

The essential elements of this service are effective user registration and management solutions that are scalable and applicable to Virtual Organisations focused upon the NGS. These may include a very large number of grid users. The GSI (Grid Security Infrastructure) offers secure authentication through the use of a Public-Key Infrastructure (PKI) mechanism with digital certificates on the NGS. To enable users to register once with all nodes the NGS has adopted a centralised user registration model. Following web registration all users are allocated a user identity from a central pool of generic user accounts supported by all grid nodes. This user identity is mapped to a Distinguished Name (DN), used in a personal digital certificate, in a grid-map file updated on a regular basis. The use of the grid-map file distribution script (developed by M Jones) is being tested. The script downloads the grid-map file from the secure GridSite at CCLRC-RAL, and filters out any potential security breaching account mappings. Users authenticate to the NGS with their personal X.509 v3 digital certificates issued by the UK Core Programme Certification Authority (CA) or other CAs with which the UK e-Science CA has agreements. Consequently they benefit from a single sign-on to this service.

Basic user information which is collected at the point of registration may be re-used for authorisation purposes by individual nodes which may apply their own site specific authorisation policies. This information is stored on a secure web site which is accessible via digital certificates to the operational staff of each node.

At present the evaluation of grid user management tools such as the VOM software [3] developed by the London e-Science Centre, the Virtual Organisations Membership Service (VOMS) [4] from the European DataGrid project and used by LCG, as well as the Community Authorisation Service (CAS) [5] is being undertaken with the aim of selecting a reliable set of tools for user
management, including authorisation, on the NGS. The intent is to select the tools that enable central management of user registration and authorisation processes while adhering to local policies on resources access and usage.

8. Access to the NGS

NGS resources are allocated to those approved users who accept the Terms and Conditions for the Use of the UK e-Science Grid which have been developed under the auspices of the Security Task Force. Applications for NGS resources are to be reviewed and approved by a sub-panel of JCSR.

Users require a digital X.509 certificate in order to use the service, which can be obtained from the UK e-Science Certification Authority [6].

At present access to NGS resources is available through use of the Globus Toolkit and GSIssh for interactive uses such as code development and debugging. In the future it is envisaged that users will access the NGS through portals. The development of portals that will offer an easier user interface using a ubiquitous web browser from the user’s desktop to all NGS assets is seen as being of paramount importance.

9. Convergence and Federation of Grids

Throughout this paper we have emphasized our aim to set up the NGS in such a way as to be convergent with EGEE. The goal is to deploy and support EGEE releases as the basic UK grid infrastructure, and supplement them with the specific enhancements required in the UK. Building upon this common infrastructure will also facilitate closer integration with the UK Grid for Particle Physics (GridPP).

Moreover it is envisaged that further integration of the EGEE, GOSC and GridPP operational teams will take place. This synthesis (refer to Figure 4) is helped by the fact that the NGS and EGEE deployment staff are overlapping groups of people.

Further to this it is expected that ad-hoc collaborations with the TeraGrid will be solidified into a strategic relationship.

10. Working with Users

Users and potential users are the motivational force behind the implementation of the NGS. The drive for the creation of the NGS was a perceived lack of an easily accessible, substantial resource for those from all areas of research interested in using grid technologies. Consequently it is the users that are foremost in our minds whilst creating and managing the NGS.

As the service is in its relative infancy, promotion to date of the NGS has been targeted to user groups with a known interest and/or technical capability in grid technology. A small scale users’ initial requirements survey was conducted to establish potential resource allocation, environmental requirements and software needs. However, it is important to note that the NGS has been set up to provide the research community as a whole with the grid technology and as the NGS matures so will its target users’ base.

To date requests for NGS resource allocations indicate that there are 33 unique science cases (some with multiple users) ranging from large scale e-Science projects such as RealityGrid, Geodise, eMineral project, and LHCb experiment to individual application scientists (for example performing pattern induction of musical themes in polyphonic music using Java). At present the majority of requests are for both computational and data resources; however there are some “only storage” and “only computational” resource requests – justifying the strategy of providing dual resource services to attract users.

In addition there are 11 users involved in testing/evaluating the NGS for potential further use (for example, the Dept of Meteorology at the University of Reading for weather forecasting models). Moreover 2 users are using the NGS as a test bed for application development (for example the GRENADE project).
The utmost priority is to deliver an effective service to UK academics and be responsive to the ever-evolving needs of the UK e-Science community and other high performance computing users. To ensure that these goals are met a single point of contact for all NGS queries has been established through the GSC Helpdesk [6] (until the GOSC is fully in place). This forwards queries and problems related to local sites to a single point of contact, usually a Helpdesk, at the institutions involved. Furthermore the User Group’s comments (through the GOSC Director) help the NGS to deliver the service that meets users’ requirements.

A comprehensive web site with documentation covering FAQs, conditions of use, gaining access and usage reporting is seen as being an essential part of the support mechanism. The NGS web site [7] is continuously enhanced to provide user-oriented information about the service, and enables the collaborating sites to disseminate the related information in private sections using the Grid-site software accessed through X.509 certificates imported into the user’s web browser.

Whilst NGS does not at present provide dedicated training, the staff involved in supporting the NGS do provide training courses on Globus installation for the National e-Science Centre which is engaged in user training.

11. Conclusions and Further Work

This paper conveys the essence of the National Grid Service (and in particular of the four clusters within it) which is enhanced under the guidance of the GOSC Director along with helpful support from the NGS Coordinator. The service has been built using experience gained with the Level 2 Grid deployed by the ETF [8, 9] and the White Rose Grid [10]. It is anticipated that this stable though evolving service will converge with EGEE.

The important issue is to fully develop the necessary procedures for offering a resilient, reliable and robust production grid. This can only be achieved by providing a professional service, and by helping to make grid technologies easy to use and of benefit to users.

The primary task now is to establish a common acceptance of the NGS by the broader user community extending beyond the current e-Science projects. Our vision is to provide a national grid service that allows users to do research using grid technologies. The development of portals, an increase in the application packages base, and grid training are seen as vital for the widespread take-up of the National Grid Service.

This paper reports the excellent progress made with the construction of the National Grid Service and shares the experience of setting up this Production Grid with our users, with those who take an interest in it, as well as with those who may wish to build a production grid.

Acknowledgements

The authors are grateful for funding received from the JCSR in support of clusters and the hosting institutions, and technical support from the Engineering Task Force and the Grid Support Centre.

The authors wish to thank Elena Blanco formerly of the University of Oxford and now of Cranfield University, and Andrew Jones of the University of Manchester for comments on our original proposal submission.

Reference

[3] The London e-Science Centre Web pages are at: [http://www.lesc.ic.ac.uk/]
[8] R Allan, M Doherty, M Dovey, M Thompson, M McKeown, Technology Roadmap for the National Grid Service
[9] A Hardisty, J Giddy, The UK e-Science Grid: Beyond a Prototype; Project to establish a Production Quality e-Science Grid, Welsh e-Science Centre