

A Technical Overview of the GridPP Testbed

A. McNab

Schuster Laboratory, University of Manchester, Manchester, M13 9PL, United Kingdom

The future of Particle Physics is dominated by the Large Hadron Collider that is under construction at CERN. At a centre of mass energy of 14 TeV, the LHC will be the most powerful accelerator in the world for many years if not decades to come. Due to the enormously high energy, high luminosity and complexity of the detectors, the LHC experiments will produce unprecedented amounts of data, estimated to be several PetaBytes per year, for offline analysis by teams of physicists all over the world. To analyse this data and to generate the Monte Carlo simulated data necessary to understand it will require huge amounts of computing and data storage facilities. These computing resources will be distributed across the world, linked together as a Grid.

GridPP is currently building a prototype UK Grid that will enable the four LHC experiments, ATLAS, LHCb, CMS and ALICE, to generate large amounts of Monte Carlo simulated data. This is currently being tested by running experiments in the USA in which the UK is involved, BaBar at SLAC and CDF and D0 at the Tevatron, FNAL. In doing this, the largest Grid testbed in the UK has been created, consisting of more than 100 servers across 16 Institutes, incorporating a functional Grid job submission system.

This testbed is largely built upon the common core software base deployed as part of the EU-wide European DataGrid middleware development programme.

Components of this include a distributed monitoring infrastructure of resources, services and networks; data and metadata catalogues, replication and management tools; mass disk and tape storage management interfaces to the Grid; distributed authentication and authorization infrastructures for multiple virtual organisations; a job submission and brokering system, with dynamic allocation of jobs to resources using the monitoring and data location services; fabric installation, management and monitoring systems at the scale of thousands of hosts per site.

1. Introduction

The GridPP Testbed is currently distributed across 17 High Energy Physics sites in the UK: Birmingham, Bristol, Brunel, Cambridge, Edinburgh, Glasgow, Imperial College (University of London), Lancaster, Liverpool, Manchester, Oxford, Queen Mary (University of London), Rutherford Appleton Particle Physics, Royal Holloway (University of London), University College London, and a central Tier1A prototype at Rutherford Appleton Atlas Centre.

The fabric of the physical testbed is configured to form several logical testbeds, with different software versions and support commitments. The bulk of the fabric is operated as part of the EU DataGrid[2] (EDG) Application Testbed, using the systems described in the rest of this paper. The remainder of the fabric is either part of the EDG Development Testbed, the LHC Computing Grid, a dedicated testbed for R-GMA monitoring software development or development testbeds for specific High Energy Physics experiments, such as CMS.

Much of the infrastructure is built on Globus[3] software, and Globus-derived protocols such as GSI[3], and most of the middleware development effort has been to add components needed to provide automated brokering or management services. This allows us to go beyond distributed computing and begin to deploy Grids.

2. Security: Authentication and Authorization

The testbed's infrastructure of trust and permissions is built on digital certificates issued by the UK e-Science Certification Authority[4] and other EDG-certified Certification Authorities[5]. These CAs provide X.509 certificates to users and network hosts with unique names (eg /C=UK /O=eScience /OU=Manchester /L=HEP /CN=Andrew McNab), which can be used for secure protocols built on SSL or GSI, such as HTTPS, GridFTP or Globus GRAM for job submission. This authenticated identity can then be used as the basis of authorization decisions.

The primary component of the EDG authorization model is the Virtual Organisation (VO). In the testbed, this normally corresponds to one High Energy Physics experimental collaboration, usually of several hundred members at tens of sites across the world. Membership lists of VOs and their subgroups are published using LDAP servers, and local files listing authorized certificate subject names are constructed every day. Job submissions or file transfers by users are accepted if they are attempted using one of the certificates listed, and a temporary Unix account is leased to the user for the duration of their access to the site.

The Storage Element filesystems described later in this paper further restrict access to individual files by using GACL[6] access control lists, written in terms of Grid identities and VO groups.

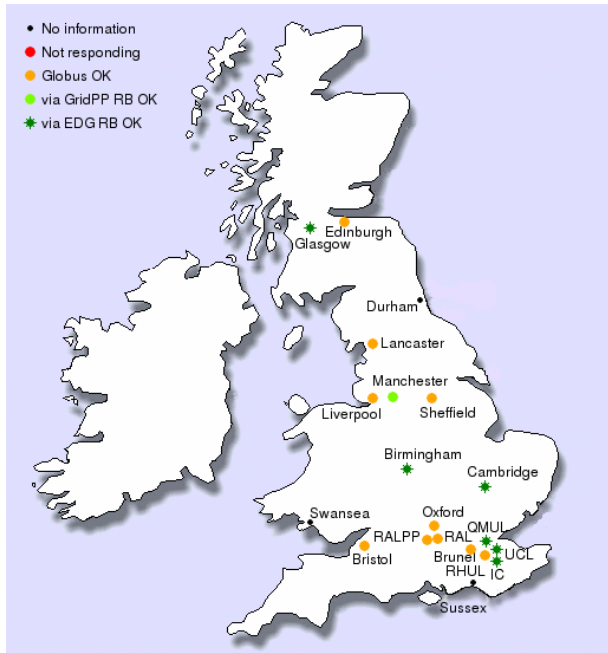


Figure 1: GridPP Testbed Status in May 2003. The symbols represent the status of sites in terms of membership of the EDG Application logical testbed. The sites with green stars were registered with the EDG and GridPP Resource Brokers' information indexes and successfully accepting test jobs; one site (Manchester), with a Green Dot, was only registered with the GridPP Resource Broker; and the remaining sites, with Amber Dots, were online and accepting direct submission of test jobs via Globus GRAM, but were not registered or successfully accepting jobs through the resource brokers.

3. Monitoring

Some Grid components need to make decisions on behalf of the user as to where best to send jobs or retrieve copies of data. For this brokering to function, it is essential that a monitoring infrastructure be in place, and desirable that the infrastructure is scalable, flexible and efficient.

GridPP has taken[7] a leading role in developing the monitoring architecture for the EU DataGrid, first by enhancements to the Globus MDS system and currently with the Relational Grid Monitoring Architecture (R-GMA)

R-GMA is based on the Global Grid Forum (GGF) Grid Monitoring Architecture, and presents information as if part of a large, distributed relational database. Multiple types of producer are supported - eg to support streaming of updates to dynamic information. The system includes a Mediator able to identify the best producer to contact and this removes some of the optimisation burden from broker. Throughout the emphasis has been on producing a scalable replacement for MDS, which takes into account the dynamic and incoherent nature of the infor-

mation in a production Grid.

4. Fabric Management

All the sites in the GridPP Testbed use fabric management software based on LCFG[8] from the University of Edinburgh. This allows farms of hundreds of hosts to be centrally installed and managed, both with unique profiles for special purpose machines such as gatekeepers or file servers and with identical, repeated profiles for the many worker nodes. We have found the predictability of automated installation especially useful in identifying problems in the distributed environment of a Grid.

The Fabric Management system also publishes information about the site status and configuration into the monitoring infrastructure, where it can be used for operational monitoring of the Grid, and by brokers needing to identify suitable sites at which to run users' jobs.

5. Resource Brokering

To go beyond the simple remote job submission system provided by Globus Gatekeeper, we need to introduce automated brokering of jobs and sites. This is done using the EDG Resource Broker[9] and Job Submission Service, which are built on the Globus Gatekeeper[3] and Condor-G[10].

The Resource Broker matches the requirements stated in the job description file, which can include sophisticated expressions involving requests for optimal characteristics (eg "the fastest site with a copy of this data") This relieves the user from having to know where data is stored and what the minute-by-minute current status of each suitable site is.

Associated with the Resource Broker are an Information Index service, which provides a locally-cached copy of relevant information from the monitoring infrastructure, and a Logging and Bookkeeping service which is notified of state transitions during job submission and execution.

6. Storage Management

GridPP has lead development of the EDG Storage Element[11] file server, which provides homogenous access to a wide variety of bulk data storage systems, including large disk arrays and tape vaults, but also single disks on modest local servers.

The Storage Element provides GridFTP, RFIO and NFS interfaces, using appropriate Grid credentials or using the correct site-specific Unix UID mapping. Fine grained control is also provided using GACL[6]

access control lists, written in terms of the Grid user credentials and groups.

7. Data Management

Due to the large volumes of data produced by experiments, most of the sites providing CPU power will not have sufficient storage to hold all of the data needed by a specific user. For this reason, the development of effective Data Management systems is essential for the exploitation of computing resources attached to the Grid by High Energy Physics applications. The key problem is to match the specific subset of the data required for a given job to a site that is able to supply it, either locally or by sufficient network capacity to a server with a copy.

The information required to perform this matching is provided by Local Replica Catalogs at each storage resource, and by Replica Location Services which aggregate this information. The EDG[12] architecture accomodates multiple Replica Location Services, and is designed to be scalable and efficient when deployed on real, production Grids.

Additional functions are provided by the Replica Metadata Service, Replica Optimization Service, Replica Subscription Service and a high-level replica management client.

8. Networking

GridPP is pursuing research into high speed networking[13] needed for moving bulk data and giving access to it from remote sites on the Grid. As preparation for this, we have deployed network monitoring software[14] at testbed sites. This records ongoing network conditions and publishes this information into the monitoring infrastructure, where it is available for optimising network-sensitive operations. In particular, this is required for the Replica Optimization Service and will enable brokers to weigh shipping jobs to data against streaming data across the network to suitable execution sites.

9. Testbed Support

GridPP has evaluated and deployed several tools for providing support to testbed sites. Our current support is largely provided by conventional email mailing-lists and telephone conferencing systems, which are the most convenient for the current number of site administrators (17); and by the GridPP website[1], which uses our GridSite software and can allow specific groups of users to authenticate using their X.509 certificate and modify subsets of the website.

We have evaluated Bugzilla[15] as a more formal bug tracking system instead of the support mailing list, but this is not yet needed for the scale at which we are operating. Some of our early meetings were held using video conferencing, but again, the reliability, convenience and ubiquity of the worldwide land-line and mobile phone systems has proved invaluable.

Acknowledgments

Several hundred people have directly contributed to the work described in this paper, as members of GridPP[1] or the EU DataGrid[2], or related projects such as Globus[3].

GridPP's development and deployment of the Testbed has been supported by the UK Particle Physics and Astronomy Research Council, the Higher Education Funding Council for England, the Scottish Higher Education Funding Council and the European Union.

References

- [1] The GridPP Project, <http://www.gridpp.ac.uk/>
- [2] The EU DataGrid Project, <http://www.eu-datagrid.org/>
- [3] The Globus Project, <http://www.globus.org/>
- [4] "UK e-Science Certification Authority", UK e-Science All Hands Conference, 2-4 September 2003
- [5] EDG Certification Authorities' Managers Group, <http://marianne.in2p3.fr/datagrid/ca/>
- [6] "EU DataGrid and GridPP Authorization and Access Control", UK e-Science All Hands Conference, 2-4 September 2003
- [7] "Relational Grid Monitoring Architecture", UK e-Science All Hands Conference, 2-4 September 2003
- [8] The LCFG Project, University of Edinburgh, <http://www.lcfg.org/>
- [9] EDG Workload Management, <http://server11.infn.it/workload-grid/>
- [10] The Condor Project, <http://www.cs.wisc.edu/condor/>
- [11] "Enabling access to mass storage", UK e-Science All Hands Conference, 2-4 September 2003
- [12] "EU DataGrid Data Management Services", UK e-Science All Hands Conference, 2-4 September 2003
- [13] "High Bandwidth High Throughput Data Transfers in the MB-NG and EU DataTAG Projects", UK e-Science All Hands Conference, 2-4 September 2003
- [14] EDG Networking Group, <http://ccwp7.in2p3.fr/>
- [15] Bugzilla, <http://www.bugzilla.org/>