Proc UK e-Science All Hands Meeting 2003, © EPSRC Sept 2003, ISBN 1-904425-11-9

Building the eMinerals Minigrid

R.P. Tyer, L.J. Blanshard, K Kleese van Dam, R.J. Allan, A.J. Richards
CCLRC eScience Centre, Daresbury Laboratory, Warrington, WA4 4AD

M.T. Dove
Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ

Key words to describe the work: Grid, Portal, SRB, Globus, Metadata, Data Management, Earth Sciences

Abstract

Following a design and consultation phase, an initial eMinerals minigrid has been set up which includes compute and data resources together with the software and middleware to enable their use. The compute resources are extremely heterogeneous (comprising Beowulf machines, IBM SP's, Condor pools as well as a plethora of Linux boxes) and are distributed across the sites associated with the project. The data resources include a dedicated database server to hold the metadata catalogue for the project and multiple SRB server instances. In order to ensure full compatibility with the UK eScience Grid, the eMinerals minigrid is built on the Globus Toolkit 2. The eMinerals minigrid is accessed via a custom eMinerals portal which includes fully integrated versions of both the CCLRC Data Portal and the CCLRC HPC Portal. The DataPortal is a web based portal which allows simultaneous searches of distributed heterogeneous metadata catalogues while the HPCPortal essentially encapsulates the core services provided by the Globus Toolkit 2 technology whilst providing additional 'added value' services such a computational workflow. The Storage Resource Broker (SRB) software developed by San Diego Supercomputing, is used facilitate data management and sharing across the virtual organisation. The SRB allows files and data stored over multiple distributed heterogeneous physical resources to be accessed through a single virtual file system. This paper will describe the integration of the constituent tools / resources which comprise the eMinerals minigrid with an emphasis on the lessons learned and problems encountered (both technological and socio-political) in the process.

I. Introduction

The 'Environment from the molecular level' (eMinerals) project is a NERC funded eScience pilot project involving groups in Cambridge (Earth Sciences), Daresbury Laboratory (eScience and computational science), Bath (Chemistry), University College London (Earth Sciences, Computer Science), Reading (Computer Science), the Royal Institution, and Birkbeck College (Crystallography).

The scientific focus of the project is the use of molecular simulation techniques in order to investigate fundamental problems associated with key environmental issues such as nuclear waste storage, pollution and weathering.
The simulation of realistic systems with the desired accuracy can be intensive both in CPU time and data requirements.
Accordingly the project is extremely interested in the use of grid technology in order to facilitate both high performance computing and high throughput computing.

In addition there is the need for wider eScience technologies in order to support effective collaborative working and data management across the distributed eMinerals virtual organisation.

Accordingly the project has set-up and been using a number of eScience tools and technologies which has become known as the eMinerals minigrid. This paper will describe the different components of the minigrid, their integration and experiences gained so far in their ongoing use.

II. Components of eMinerals MiniGrid

The challenge of the project from a data management perspective is to allow effective sharing of data and metadata across the virtual organisation. Since the use of scientific codes is central to the project, the majority of the data is stored in the form of data files.

Currently these files are mostly in code specific formats although there is much effort within the project to convert to using XML based data formats namely Chemistry Markup Language (CML) and Computational Chemistry Markup Language (CMLcomp).

A. Metadata Management

CCLRC's integrated data system includes the Data Portal for high-level access to multidisciplinary data, linking to existing or new data catalogue systems. These catalogues include metadata as well as links to the data itself. The data is held in various storage resources from local disks, over databases to multi terabyte tertiary tape systems. For a number of e-Science projects, the first step is to create metadata catalogues and links to their data. The structure of these catalogues must allow storage of metadata that is common amongst the different scientific areas since the DataPortal provides common searches using a common scientific metadata format in XML. The structure must therefore allow storage of common metadata such as the name of a study and names of investigators.

CCLRC houses a database server which will hold all the metadata relating to the project. It will be stored in a relational database in a standard structure that maps directly to the CCLRC Interdisciplinary Metadata Format (XML schema) that was designed specifically to describe scientific research activities.

B. Data Management

The metadata contains links to the actual data holdings themselves, which within the context of this project are the data files produced from the simulations. In order to ensure referential integrity of the links, the Storage Resource Broker (SRB), developed by the San Diego Supercomputing Centre (SDSC), is being used.

Essentially SRB allows multiple distributed heterogeneous file systems to be treated as a single virtual file-system. This concept is extremely important within the context of the Grid where the user should not be concerned with the actual location of their files, the file system they are being stored on etc. It also allows files to be stored logically, i.e. according to their content, rather than by their location. Figure 1 shows the SRB architecture which consists of three principle components.

SRB Clients
There are a number of ways to access the SRB virtual file system. These include SRB browsers, such as InQ, SRB API and web services interfaces.

SRB Server
SRB server sits on top of the actual data storage systems translates requests from SRB clients into a format appropriate for the local systems.

MCAT
The MCAT stores the location associated with data sets, users and resources managed which is used by the SRB servers to service the requests from the SRB clients.

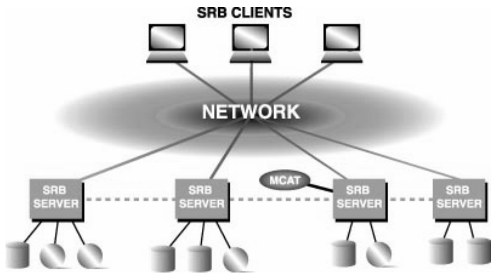

Figure 1. Architecture diagram showing the three principle components of SRB: SRB clients, SRB servers and the Metadata Catalogue (MCAT).

The use of SRB is also extremely useful for facilitating sharing of data across the VO. It is possible to define very specific access rights and conditions to data held in SRB for other members of the VO.

C. Compute Grid Infrastructure

Inspired by the development of UK eScience Grid, the project has been using the Globus Toolkit 2 to set up a compute grid. The eMinerals VO has a central Grid Index Information Service (GIIS) server to which all resources on the compute grid report.

Currently the grid contains machines from CCLRC, Cambridge and UCL. The grid is extremely heterogeneous containing parallel machines, Linux boxes and Condor pools. Some of these resources are also part of the UK eScience VO and report to two GIIS servers.

Condor
Condor is a technology developed by the University of Wisconsin in order to harness idle CPU cycles. It contains an efficient task farming

engine and is a very effective tool for the right class of computational problems, parameter space searches for example.

The eMinerals VO contains two Condor pools, one at the UCL node and the other at the Cambridge node. The UCL Condor pool is composed entirely of Windows machines from undergraduate teaching laboratories while the Cambridge pool is heterogeneous containing a variety of UNIX, Linux and Windows machines.

D. eMinerals Portal

An eMinerals web portal has been set up to provide a single point of secure access to the resources on the minigrid. The eMinerals portal is a fully integrated custom instance of two other web portals developed by CCLRC, namely the Data Portal and the HPC Portal. The eMinerals portal is also linked to a custom MDS (Monitoring and Discovery Service) service to which the resources available to users of the minigrid are setup to report.

The Data Portal (as previously outlined) is a web based portal which facilitates the discovery of scientific metadata by allowing the simultaneous searches of multiple distributed heterogeneous databases. The Data Portal acts as a broker between the user and the facilities with metadata and data holdings. The key component of the Data Portal is the XML wrapper which converts queries into a SQL statement appropriate for each specific database and then translates the search results back into the CCLRC metadata format.

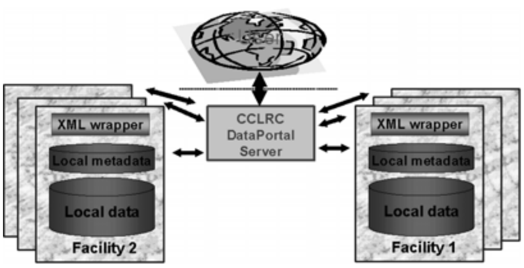

Figure 2. General Architecture of DataPortal.

The HPC Portal is another web portal developed by CCLRC but whereas the Data Portal is a broker to facilitate interactions with a data grid, the HPC portal is a broker to allow access to resources on a compute grid. Essentially the HPC Portal allows access to the core functionality of the Globus Toolkit 2 through a web portal. In addition it provides a number of 'added value' services including session management and workflow tools.

E. Collaborative Working Tools

Given the number of different groups working within the project, there is a clear need for collaborative working tools and environments. The project has made extensive use of the Access Grid technology and has been evaluating Pig / Piglet technology with the eventual aim of deploying Pigs across the VO.

In addition to AG and Pig, the Reading node of the project is developing a collaborative working suite known as the Computational Collaborative Framework (CCF) which allows the sharing of applications remotely as well as providing white board tools (Figure 3).

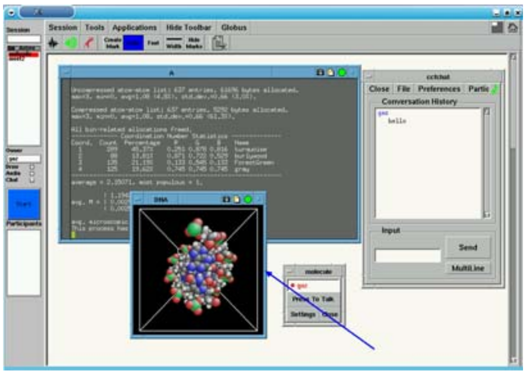

Figure 3. Example of CCF framework.

III. Integration of Minigrid Components

A. Firewall Issues

The compute resources are all running Globus Toolkit 2 middleware. By using the eMinerals portal as the central point of access for the minigrid, it has been able to avoid firewall problems as the ports necessary for Globus use need only be opened to the portal server.
The trade off for this convenience is however it that the portal server becomes a single point of failure for the minigrid.

More problems have been encountered with the use of SRB due to firewalls. SRB server middleware which are currently installed on the Cambridge and Daresbury nodes requires access to the central MCAT for the project. However as the database server which hold the eMinerals databases is heavily fire-walled as it also holds databases crucial to other eScience projects.

Hence in order to enable all nodes of the VO access to SRB it was necessary to install the SRB server software and the MCAT software on an application server outside of any site firewalls. The application server then performs port forwarding to the database server using Oracle Client software. The implementation require firewall holes only to be opened to the application server and also has the additional benefit that if the SRB software is somehow compromised then the database server is still secure.

B. Interacting with Queuing systems and Condor

Some of the compute resources on the minigrid run batch queuing systems. For example, the IBM SP machines run LoadLeveler and the DEC Beowulf cluster uses OpenPBS. In addition to the parallel / cluster machines, there are also two Condor pools in the minigrid.

Two approaches have been experimented with in order to interact with these queuing systems via Globus.

The first approach has been to use the Globus job managers for these queuing systems. The globus job-managers can be configured for different types of queuing systems, which are then interacted with through the use of the RSL syntax. This is not necessarily an all encompassing solution as the parameters passed in the RSL string are mapped to the underlying native batch system and so it is still a requirement that the sufficient parameters needed to run the job are supplied.

The second approach has been to use the default globus 'fork' job-manager, and supply a pre-written batch file, particular to the target queuing system, which after data has been transferred to the target machine is invoked via a simple globus run command. These scripts are either transferred with the data to be processed or are 'pre-installed' on the target machine and invoked in the same way as any other target executable.

B. Portal Integration Issues

As described earlier, the eMinerals portal is actually a customised integrated instance of the CCLRC Data Portal and the CCLRC HPC Portal (Figure 4).

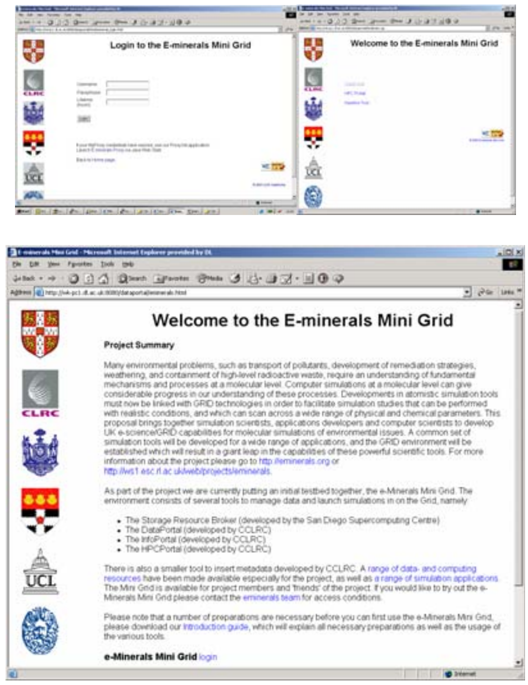

Figure 4. The customised eMinerals portal.

These portals were developed independently and in particular both had their own session management services. In order to allow a degree of flexibility for both portals to develop in the future, it was decided that both portals should retain their own individual session management service.

However within an integrated environment, such as the eMinerals portal, a typical scenario is that a user may wish to use the Data Portal services to search metadata holdings and eventually find links to specific data of interest. Having found this data, HPC Portal services may then be used to find a suitable compute resource, transfer the data and then start some computation on that data.

In order to allow a degree of flexibility for both portals to develop in the future, it was decided that both portals should retain their own individual session management service. Hence for the above scenario to be possible some mechanism of federating credentials, specifically proxy certificates and private keys, is necessary. The solution was to modify both portals so that their uses a common format for their session identities and to use the standard GSI delegation mechanism to transfer credentials between the session managers.

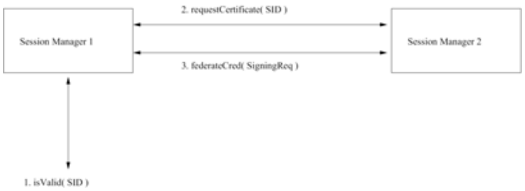

Figure 5. Session manager federation mechanism.

Figure 5 shows federation mechanism schematically which consists of the following steps:

- Initially a user has logged onto one portal and their credentials are stored in Session Manager 2 which has issued them with a session id string (SID). At some point later, the user then attempts to use a service of the other portal which sends the SID to it's own session manager (Session Manager 1) in order to verify that it is a valid SID.

- Since this SID is not recognised by Session Manager 1, it contacts Session Manager 2 asking if it recognised this SID and requesting the corresponding proxy certificate.

- Assuming the SID is valid, Session Manager 1 then generates a new public private key pair. The public key is used to form a certificate signing request which is then send to Session Manager 2 which then forms a delegate certificate by signing the certificate signing request with the private key that it holds. This certificate is then returned to Session Manager 1

Using this mechanism it is possible to transfer credentials without the private keys ever being exposed.
By implementing this functionality between the two session managers, it is possible for a user to sign on once with either portal and to then seamlessly use the services of either portal.

References

[1] Martin T Dove, Mark Calleja, Jon Wakelin, Kostya Trachenko, Guillaume Ferlat,
Peter Murray-Rust, Nora H de Leeuw, Zhimei Du, G David Price, Paul B Wilson,
John P Brodholt, Maria Alfredsson, Arnaud Marmier, Richard P Tyer, Lisa J Blanshard,
Robert J Allan, Kerstin Kleese van Dam, Ilian T Todorov, William Smith,
Vassil N Alexandrov, Gareth J Lewis, Ashish Thandavan, S Mehmood Hasan, *Environment from the molecular level: an science testbed project*, AHM 2003 (Nottingham 2-4/9/2003)

[2] CCLRC Scientific Metadata Format
http://www-dienst.rl.ac.uk/library/2002/tr/dltr-2002001.pdf

[3] CCLRC Data Portal
http://ws1.esc.rl.ac.uk/web/projects/dataportal

[4] Environment from the Molecular Level
http://ws1.esc.rl.ac.uk/web/projects/eminerals

[5] HPC Portal
http://ws1.esc.rl.ac.uk/web/projects/gridservicesportal

[6] SRB Home Page
http://www.npaci.edu/DICE/SRB/