Mathematica
Balkanica

New Series Vol. 25, 2011, Fasc. 1-2

Using Grid for Mathematical Researches

Radoslava D. Goranova[1], *Vladimir T. Dimitrov*[2]

Announced at MASSEE International Congress on Mathematics, MICOM-2009, Ohrid, Republic of Mavedonia

The Grid is a hardware and software infrastructure that coordinates access to distribute computational and data resources, shared by different institutes, computational centers and organizations. Grid is a new technology, a new way of using computers, network and devices, which can be effectively used for solving difficult mathematical, chemical or physical problems. In this article we present how Grid, can be used for mathematical researches and more specific how MATLAB can be used in g-Lite environment.

1. Introduction

The object of this article is to introduce the possibilities of the Grid infrastructure and more precisely the features, which infrastructure provides for mathematical researches. In the first part of the article we give general overview of the Grid as brief introduction and in the second part we focus on a concrete Grid environment - g-Lite and its integration with the software for mathematical calculations MATLAB.

The Grid [1] is dynamic and heterogeneous system, beyond any institutional or organizational boundaries, which provides access to nontrivial qualities of services (or resources). The resources provide services like computing power, data storage, sensors etc. The base characteristics of the Grid are:

1. Coordinates resources which are not subject of centralized control. Resources are spread in different organizations, with different security policies and access rights.

2. The user have direct access to resource, direct access to computer where services are located.

3. The user can access resource by using standard and open protocols and interfaces.

4. The Grid resources are shared over the Internet.

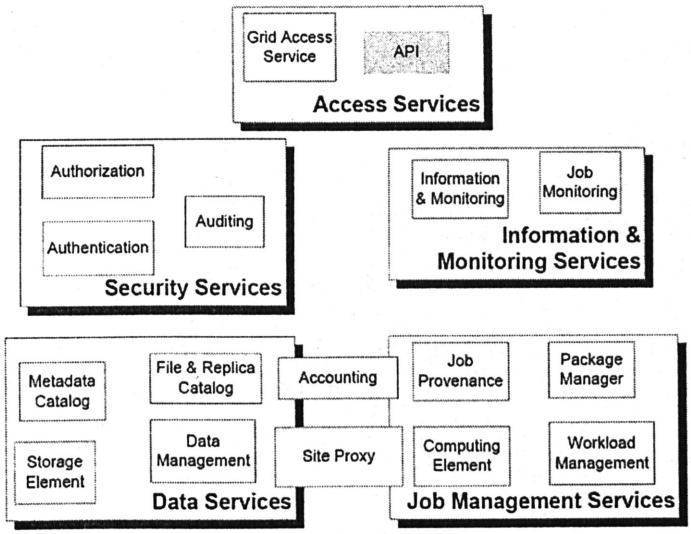
In order to organize secure access in the Grid, different security mechanisms have to be used. The most frequently used mechanism in Grid is based on X509 certificates and PKI (Public Key Infrastructure). Every Grid user needs certificate in order to access the grid. The certificate unifies user and proves user identity. On the other hand every organization which shares grid resources also has to possess certificate for each resource it shares. Separate people, institutions and organizations, which share common interests are organized into Virtual organizations (VO). VOs define access policies and users rights. More generally the Grid provides "unlimited" access to data storage and processor power and provides the user opportunity to run jobs (long running or parallel) in it infrastructure.

2. G-Lite middleware

Environment which implements Grid requirements is called Grid middleware. G-Lite [2] is a middleware designed and implemented for the needs of EGEE project [3]. The Enabling Grids for E-sciencE (EGEE) project aims to build on recent advances in grid technology and develop a service grid infrastructure which is available to scientists 24 hours-a-day. The environment provides data services, security services, services for job management, computation and brokering and information services. The storage element (SE) is the service which allows a user or an application to store data for future retrieval. It main functionality is data storage, data transfer and data replication. Usually users use these services in order to transfer data for analyze on the storage or the result from analyzed data to the storage. The environment implements the following types of storage elements:

1. CASTOR - it consists of a disk buffer frontend to a tape mass storage system. File migration between disk and tape is managed by a service called "stager"'. Uses native storage protocol and insecure RFIO protocol for access to the files.

2. dCache - it consists of a server and one or more pool nodes. The server represents the single point of access to the SE and presents files in the



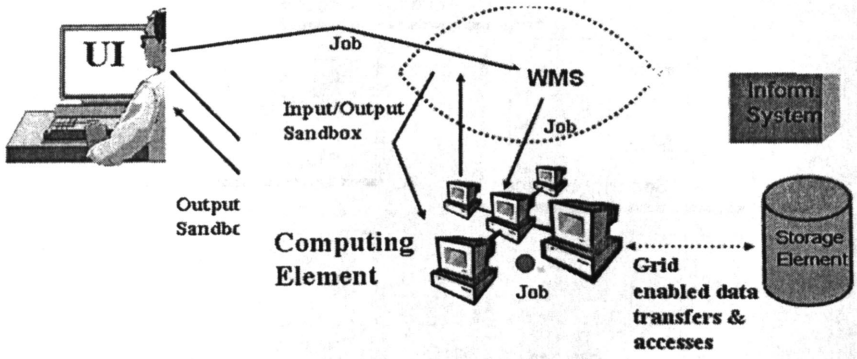
pool disks under a single virtual filesystem tree. Nodes can be dynamically added to the pool. Uses native gsidcap protocol for data access.

3. Disk pool manager (DPM) is a lightweight disk pool manager. Disks can be added dynamically to the pool. Uses secure RFIO protocol for data access.

4. Classic SE - it consists of a GridFTP server and an insecure RFIO daemon in front of a physical single disk or disk array. Currently not supported.

Job management services provide functionality for job submission, job computation and job brokering. The Workload Management System (WMS) is component that allows users to submit jobs, and performs all tasks required to execute them, without exposing the user to the complexity of the Grid. On picture bellow is shown the process of job submission in g-Lite environment. The user by user interface, submit a file described by job description language (JDL). In this file user, defines his requirements for job execution, the type of the job (parallel or not), the output path, where the end result will be stored. This JDL file is sent to WMS component of g-Lite environment. The component read the JDL file and on the base of user requirements sent it to a proper Computing Element (CE). CE is another component from job management services. It is a cluster with batch system and worker nodes, which provide computational capabilities. During the job execution data from SE can be transferred in order to

be analyzed. It is the responsibility of the user to describe his jobs and their requirements, and to retrieve the output when the jobs are finished. Information



services provide information for job monitoring, job accounting and resource discovery. They are realized by providing two types of information components BDII and R-GMA.

BDII [4] is The Berkeley Database Information Index, which is a standard LDAP database updated by an external process. The update process obtains information (LDIF files) from a number of sources (other g-Lite sites) and merges them. G-Lite middleware provides standard clients, as ldap-search, and specific environment client, as glite-sd-query, for access and interactions with the LDAP database. The environment client provides different API: C, JAVA, Python, which can be used for interactions.

R-GMA [5] is the other information component of the environment. The Relational Grid Monitoring Architecture (R-GMA) is a relational database that provides a service for information, monitoring and logging information for jobs, environment resources and sites. R-GMA consists of Producers which publish information into R-GMA, and Consumers which extracts information from the database. The environment client glite-sd-query, can be used also for searching resources and information into R-GMA provider. The client returns information for the type, name and endpoint of environment services.

Proxy certificate creation is essential part for interaction with environment services. This functionality is a part from the security services. The clients, which g-Lite environment provide for proxy creation, delegation and destroy communicate with two main components MyProxy server and VOMS Server. These components provide mechanism for users' authentication and authorization.

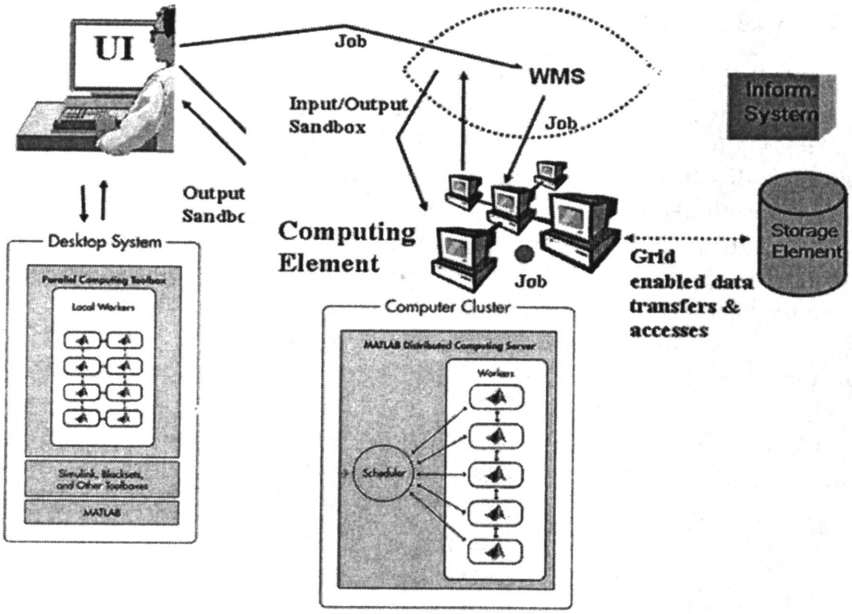
3. MATLAB Integration

MATLAB [6] is a numerical computing environment which allows matrix manipulation, plotting of functions and data, implementation of algorithms etc. Parallel Computing Toolbox and MATLAB Distributed Computing Server software provides functionality for solving computationally and data-intensive problems using multiprocessor computers. The client uses Parallel Computing Toolbox software to perform the definition of jobs (large operation, which user needs to perform). The MATLAB Distributed Computing Server product performs the execution of the job and returning the result to the client. The job manager is the part of the server software that coordinates the execution of jobs. The job manager distributes the tasks for evaluation to the server's individual MATLAB sessions called workers.

MATLAB Distributed Computing Server lets users solve computationally and data-intensive problems by executing MATLAB applications on a computer cluster. It is available for all hardware platforms and operating systems and directly supports LSF, PBS and TORQUE schedulers.

Since 2008, g-Lite and MathWorks teams, worked in collaboration in order to provide MATLAB as a grid service. On picture bellow is shown how MATLAB was integrated with the g-Lite environment. The MATLAB Distributed Computing Server (MDCS) has very similar mechanism of job submission as the Computing Element of g-Lite. Using this feature the MDCS can be installed on CE and can be configured in such a way that both of the components can use common scheduler. The thing which is worth to mention is that by Parallel Computing Toolbox software, package installed in MATLAB client, user can submit jobs to Grid and take advantage of MATLAB features. In order to use MATLAB in g-Lite, however, there are some requirements which the user have to meet. First of all the user has to be a g-Lite user, that means to have a valid X509 certificate and to be a member of VO, which to support MATLAB job submission. The user needs a MATLAB client installed on his computer, properly configured and to possess a license for MATLAB and Parallel Computing Toolbox. Additional settings have to be performed like, setting SSH key based login from the computer where MATLAB client is installed to g-Lite user interface (UI), setting LFCHOST variable (relevant with data services). Version of MATLAB is also important R2008b or R2009a with Parallel Computing Toolbox package. There is sample configuration for the MATLAB client, provided by g-lite team, where the name of the UI, user home directory and username for UI, grid username and user VO has to be specified.

The steps for submitting a MATLAB job are the same as to submit a regular job. First user has to create proxy certificate. From MATLAB user



creates jobs (by command createJob) and submit them to the Grid (by command submit). MATLAB client copies all necessary files to UI. From UI MATLAB can submit the job to the g-Lite environment and can copy files from UI to SE.

Currently, there are two g-Lite sites on which MDCE is experimentally set up and running: HG-03-AUTH located in Greece, GRIF located in France.

4. Conclusion

Usage of the grid is entirely dependable of users' needs and possibility to have access to specialize software for mathematical, chemical or physic researches. In this article we present how Grid, can be used for mathematical researches and more specific how MATLAB can be used in g-Lite environment. There are still open questions about the MATLAB Distributed Computing Server and the license for its usage in g-Lite. But g-Lite and MathWorks teams, worked in collaboration to find a solution.

References

[1] I. Foster and C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure.* Morgan Kaufmann, San Francisco (1999)

[2] g-Lite User Guide, https://edms.cern.ch/file/722398//gLite-3-UserGuide.html
[3] EGEE, http://public.eu-egee.org/
[4] g-Lite BDII, https://twiki.cern.ch/twiki/bin/view/EGEE/BDII
[5] g-Lite R-GMA, http://www.r-gma.org/index.html
[6] MATLAB, MATLAB Distributed Computing Server 4 System Administrator's Guide
[7] g-Lite and MATLAB, https://twiki.cern.ch/twiki/bin/view/EGEE/GLite AndMATLABToolkitIntegration
[8] g-Lite and MATLAB, Parallel Computing with MATLAB R2008B on EGEE cluster for Windows Client

Faculty of Mathematics and Informatics
Sofia University "St. Kliment Ohridski",
Sofia 1164, Bulgaria
E-mail: [1] *radoslava@fmi.uni-sofia.bg*
E-mail: [2] *cht@fmi.uni-sofia.bg*