

Open Source Portal Framework for Job Submission

Arunachalam B., Kalasagar B., Vineeth Simon Arackal, Prahlada Rao B.B.

*Centre for development of advanced computing, CDAC Knowledge Park No1
Old Madras Road, Byappanahalli, Bangalore-560038, INDIA*

(barunachalam,kalasagarb,vineeth,prahladab)@cdac.in

Abstract: *This paper describes the Open Source Portal Framework (OSPF) for job submission to grid, which is developed based on JSR168 Portlet standard, with GridSphere Portal Framework, Globus Middleware and Gridway Meta-Scheduler. OSPF is providing web portal interface for Open Source Grid Community with the ultimate aim of accessing the distributed computing resources. This portal framework provisions user authentication and authorization using MyProxy Credentials. OSPF provides a platform for development and maintenance of their own Portlets like job Submission, Grid Status and Problem Solving Environments (PSEs) based on specific applications. This paper also captures about OSPF features and its implementation along with architecture, performance analysis followed by comparison of various open source Grid portal frameworks.*

Keywords

Portal, Portlets, Gridway Meta-Scheduler, GridSphere Portal framework, Globus Toolkit, LRMS, JSR168, OSPF, PSE.

1 Introduction

A Grid [1] is a cluster of heterogeneous workstations located at geographically distant locations, and with the ultimate aim of sharing computing resources. Grid evolved from distributed computing to focus on large scale resource sharing and high performance computing to solve huge scientific applications. Grid is a combination of one or more clusters, which makes it look like a large and powerful virtual computer. Grid supports multiple organizations to shared resources with others known as Virtual Organization. Virtual Organizations (VOs) [2] are group of researchers with similar scientific interest and requirements, who work collaboratively with other members and/or share resources (e.g. data, software, expertise, CPU, storage space), regardless of geographical location. Some of the popular Grid Computing initiatives in the world are TeraGrid, Sun Grid, NorduGrid and GARUDA Grid [26].

The most popular method of accessing the Grid is using a web portal [3]. Grid Web portal provides the users a simplified way of access to resources by avoiding underlying complexities involved in command line interface and unfamiliarity with various flavors of UNIX. Every desktop contains web browser and most of them support for dynamic content display. A web portal refers to a website or service that offers a broad array of resources and services. Grid portals build upon web portal to admittance distributed computational resources. Grid portal acts as a gateway between Grid users and resources, hides impediments involved in command line access and affords user-friendliness. There are many Grid initiatives around the world and each Grid has its own access mechanism. Many such Grid Portals have been developed for solving many domain applications including bioinformatics, e-learning etc.

Application scientists are interested to build "stove-pipe" [4] solutions because the time and manpower involved in building an advanced portal is very high.

For the development of OSPF we have used GridSphere Portal Framework 2.2.10 [5], JSR168 portlet API [6], java and Apache ant to build source. This Portal has been developed in JSR168 Portlet technology, with GridSphere Portlet Container [7] and in-house developed portlets. OSPG supports user's authentication with X509 credentials [8] with support of Myproxy server [9].

Rest of the paper is organized as follows: In section II, we examine related works in Open Source Portals, Architecture, Implementation and Features of OSPF (section III), software hierarchy (section IV), performance analysis and testing results (Section V) followed by comparison of various open source Grid Portals (section VI).

2 Related Works

Many Grid Portals have been employed for different applications. Some of them are CDAC GARUDA Access Portal [10], TeraGrid user Portal [11], P-Grade Portal [12], NEES Grid Portal [13] etc. There are only few Open Source Grid Portal frameworks has been made to build Grid Portlets with an ultimate aim to facilitate development of portlets in an easy way. Such types of frameworks are GPDK [14], Gridsphere Portal Framework [15] and P-Grade Portal.

The GridSphere Portal Framework provides Open Source Grid Portal solution with Grid Portlets based on JSR168 standard. This project is funded by EU. GPDK is also JSR168 obedient. This project supports only Globus Toolkit2 and used in project Open Grid Computing Environment (OGCE) [16]. GPDK allows construction of application specific portals by providing generic interfaces/libraries for accessing Grid resources using Globus.

P-Grade Grid Portal is a web based environment for developing, executing and monitoring workflows and workflow based parameter study applications on different type of Grid and Cluster Middleware's. This portal is developed by an open consortium coordinated by the laboratory of parallel and distributed systems at MTA-SZTAKi.

3 Open Source Portal framework for Grid

OSPF is a standardized portal framework to access distributed computing resources geographically located in distant locations through web interface. The basic software components encompasses the Grid Middleware (Globus Toolkit)[17], Grid Meta-Scheduler (Gridway)[18], Myproxy server (Credential manager) and Local Resource Management systems (LRMs) [19] like Portable Batch System (PBS)[20], Torque[21], Sun Grid Engine [22] or Load Leveler.

OSPF is providing portlets for job submission, resource monitoring, management of certificates and files in Grid. Grid is mainly UNIX-based, and it expects the user to be familiar with the UNIX operating system or the usage of command line interface. OSPF is a boon to non-UNIX based users by masking the underlying complexities to access the computing resources on the Grid environment.[23].

3.1 Architecture

OSPF is a Portlet based Portal, and consists of Portlet Container and in-house developed Portlets for job submission and resource monitoring. This Portal is installed in submit node, i.e., Grid head node. We shall use the terms Submit node and Grid Head Node interchangeably. Submit node has been installed with Grid Middleware (Globus), Grid Meta-Scheduler (Gridway) etc.

In general, Grid is a cluster of heterogeneous workstations having different types of clusters, for example Linux cluster, IBM AIX etc. All Middleware's and tools installed here like in submit node.

OSPF is deployed on cluster of four- nodes with Quad-Core Intel X3430 @ 2.40GHz Processors, UNIX based operating system with memory of 8GB, Hard disk of 1TB and 100 Mbps Ethernet network. [23]

A Web browser is a software application and first access point to access any resources in World Wide Web. A Portal is an aggregator of content and applications or a single point of entry to a set of tools and/or applications defined by the user or by the portal service provider. The GridSphere Portal framework provides an open source portlet based Web portal. GridSphere enables developers to quickly develop and package third-party portlet web applications that can be run and administered within the GridSphere portlet container. A portlet container is software running as a Java app inside a web container that allows a portlet to run. Or, in reverse order, a portlet uses a portlet container which in turn uses a web container. This framework is developed for Grid users so that we choose it as a better choice to use for development of Grid Portlets after comprehensive study of various Open Source Portal Frameworks. [23]

OSPF provides job submission to resources through Gridway Meta-Scheduler and Globus Toolkit as shown in Fig 1. To access Grid resources, user must and should authenticate with valid Grid credentials. For that, user can request certificate to Grid authority and store it in Myproxy server. By using Myproxy server user can initialize valid Grid credentials anytime.

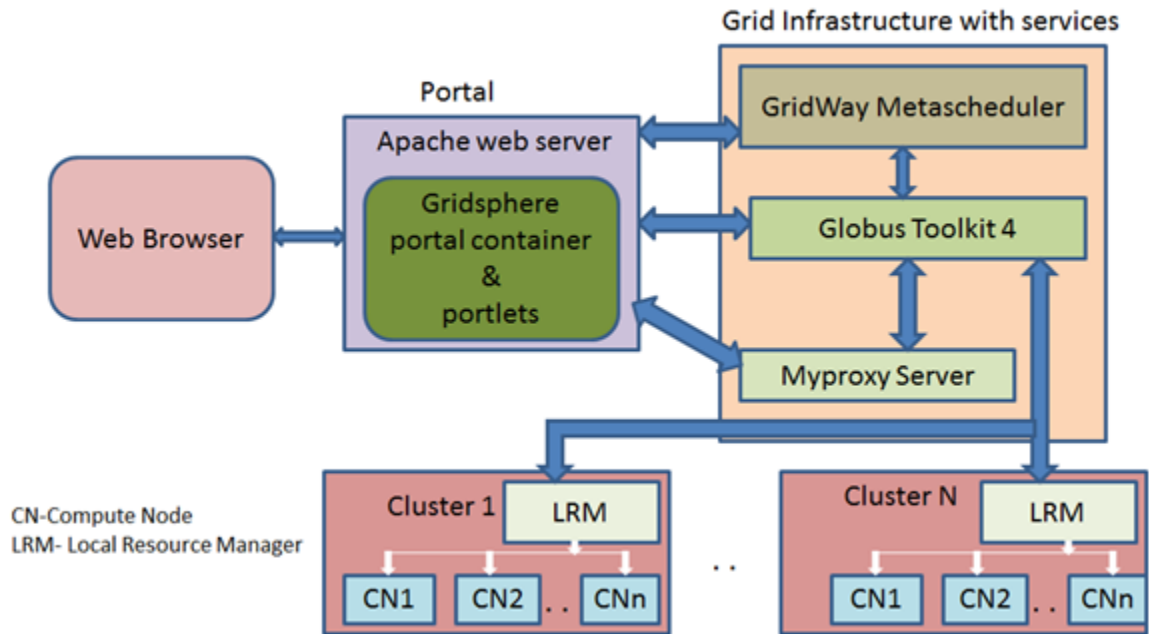


Figure 1: Component architecture of Grid with Open Source Portal framework

Meta-Scheduler plays an important role in scheduling jobs to distributed Grid resources. Most of the popular Grids use Meta-Scheduler to schedule jobs across the distributed HPC resources. Gridway is a popular Meta-Scheduler. GridWay Meta-Scheduler enables large-scale, reliable and efficient sharing of computing resources: clusters, supercomputers, stand-alone servers, etc. Gridway provides a single point of access to all resources in an organization, from in-house systems to Grid infrastructures and Cloud providers.

The Globus Toolkit is an open source software toolkit used for building Grids. It is being developed by the Globus Alliance and many others all over the world. After submission of job, job will be directed to cluster LRM to schedule across compute nodes.

3.2 Deployment Details

The OSPF is deployed on Grid Head Node, which also hosts web server and database for better job management and accounting.

Globus Toolkit Version 4 is used as a Grid Middleware. Globus is installed on Grid Head Node and on each Cluster's Head Node. Gridway Metascheduler 5.5.10 is used as a Grid Meta scheduler, for efficient sharing of resources installed in Grid Head Node. Torque is used as Local Resource Manager in Solaris and Linux clusters, and IBM Tivoli Workload Scheduler [24] Loadleveler (formerly Loadleveler) in AIX cluster. We use the terms resource manager, local job scheduler, and local resource manager interchangeably.

Database is required to store the information about job like executable, input and output data for accounting of jobs. We are using MySQL database for it.

3.3 Features

The key features of OSPF portal are as follows:

1. Credential Management to handle security issues
2. Job submission to any cluster with scheduling facility
3. Resources manager
4. File management
5. Job Status monitoring

OSPF has a pluggable architecture, which is easy to understand and deploy with minimal configuration. Each of these key features shall be explained below. Myproxy is using for credential management to handle security issues. Myproxy is for managing x509 public key infrastructure security credentials. This provides enhanced security in portal too. Please check Fig 2 for snapshot of the Myproxy interface

CDAC Open Source Grid Portal

Welcome Administration Job Management ResourcesInfo File Management **ProxyManager** Help/FAQs

MyProxyInit MyProxyLogon GRID PROXY INIT

MyProxy Init

User Name

GridProxyPassPhrase

MyProxy PassPhrase

Re enter MyProxy PassPhrase

User Cert File No file chosen

User Key File No file chosen

Proxy Life Time Hrs

[Help](#)

Tue May 14 11:33:25 GMT+05:30 2013



Figure 2: Myproxy web Interface

OSPF supports two methods of submitting job to Grid, i.e. Basic Job and Advanced Job Submission. In Basic job submission, it submits homogeneous sequential as well as parallel jobs. In basic submission, we use Gridway Meta-Scheduler for submitting job to Grid is shown in Fig 3. Using advanced job submission job will submit directly to Globus. Job status monitoring also can be done.

CDAC Open Source Grid Portal

Welcome Administration **Job Management** ResourcesInfo File Management ProxyManager Help/FAQs

Basic Job Submission Sequential/MPICH Job Submission XML Job Submission Job Info

BASIC Job Submission Portlet

Welcome osgap,

Job Name

Job Type*

Total number of processes

Operating System/Processor*

Executable/Batch File*

Input File

STDIN

Command line arguments

Environment Variables

User Specific Output File

[Help](#)

Figure 3: Job Submission web interface

The Resource Manager allows users to select the appropriate cluster for their application. It lists available clusters on the Grid with parameters like operating system, architecture, memory, total and available CPUs. The File Manager Portlet supports file operations such as copying, moving, deleting, renaming, uploading/downloading and creating directories.

4 Software Hierarchy in OSPF

At the top of the hierarchy comes the Portal (OSPF), which runs from the Grid Head Node or Submit Node. Indeed, most of the users are not even aware of the existence of the underlying software layers since OSPF hides the complexity from them.

If the user selects the Basic Submission option of OSPF, OSPF in turn submits the job to the Gridway Meta-Scheduler is shown in Fig 4. If the user has explicitly selected the cluster where he wants the job to be run, this information also will be passed to Gridway. If the user has not selected the cluster, Gridway evaluates the available clusters on the configured parameters and selects the best available cluster using efficient job scheduling algorithms.

Once Gridway has information about the cluster where the job needs to be run, the job is passed to the Globus Toolkit Version 4 (GT4) Middleware, which does the necessary permission and authorization and authentication checks, and submits the job to the Torque Local Resource Manager (LRM) that runs in the selected cluster. Torque selects the compute nodes in that cluster, and submits the job there. Note that many file transfers are involved here. For example, the executable, if present in the users' desktop, has to be first transferred to the Submit Node, and from there to the Cluster Head Node.

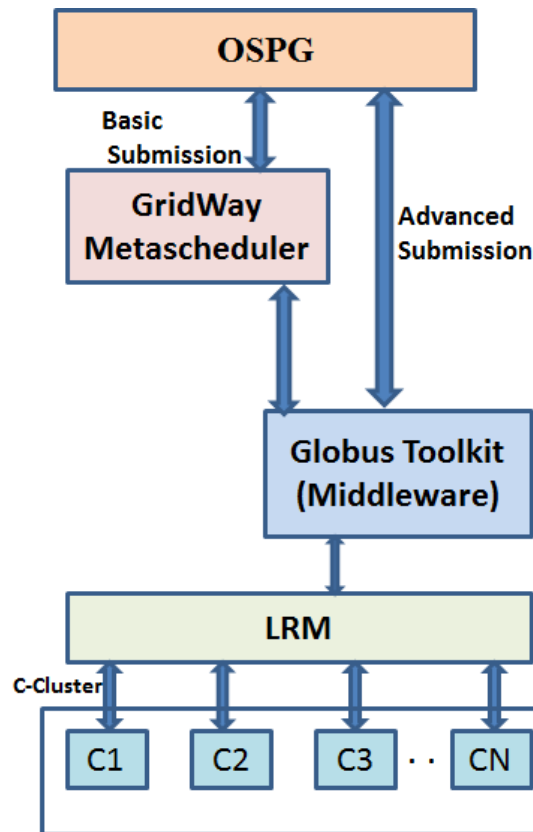


Figure 4. Software Hierarchy/Stack in OSPG for Job Submission

Once the job completes execution, the communication happens in the reverse manner. LRM notifies Globus, which in turn notifies Gridway, and OSPF obtains the status from Gridway. The output files are transferred from the Cluster Head Node to the Submit Node by Globus.

So there is a multi-level scheduling that happens if the user doesn't explicitly select the cluster where he wants to submit his jobs. First the Metascheduler selects the particular cluster where the job needs to run, and then the LRM selects the particular nodes within that particular cluster where the jobs will run. In order to use OSPF, one should follow the Hierarchy/Stack mentioned above.

5 Dashboard

In OSPF portal we are providing a dashboard with the following features.

1. Network Info
2. Grid Overview

Network Info page tabulates the various network monitored parameters namely Link Status, Bandwidth Available and Bandwidth Used , Packet Loss and Round Trip Time. Each of these data is categorized by cluster wise. Grid Overview page is a detailed graphical representation of all the major parameters relating to node, jobs and network.

6 Applications

This section presents the applications which we have run using this portal framework. We are using standard benchmark applications like MM5, GROMACS and some standard benchmarks for testing this Portal. Some other applications like FASTA, Protein Structure Prediction (PSP) also were tested.

We have identified some key features for comparison with available Open Source Grid Portals, as shown in Tab 1. We are comparing OSPF with GridSphere Grid Portlets and GPDK. GPDK Portal is in-active now.

Tab.1. Comparison with Other Open Source Frameworks

Features	OSPF	GPDK	GridSphere
User profile Manager	Yes	Yes	Yes
MPI Support	Yes	Yes	Yes
Job Submission	Yes	Yes	Yes
File Transfer Manager	Yes	Yes	Yes
Upload/Download Manager	Yes	Yes	Yes
Resource Information	Yes	Yes	Yes
Job Monitoring	Yes	Yes	Yes
Help	Yes	No	Yes
Job Submission with GridWay	Yes	No	No
Credential Manager(Upload)	Yes	No	No
Credential Manager(Download)	Yes	Yes	Yes

7 Conclusion

In this paper, we described about the Portlet based standardized OSPF Portal framework to access distributed computing resources. This Portal has a pluggable stove-pipe model for scientific communities. This Portal will attract non-Unix users to utilize Grid Resources in efficient manner. This Portal supports two types of job submissions: one is using Gridway Meta-Scheduler, which is having advanced scheduling policies and transparent job management over heterogeneous resources. Another one is using Globus Middleware. For case study, this Portal is developed on top of

GARUDA Access Portal [10] having large number of users. In future, we shall further enhance our OSPF to provide more features like supporting various Grid Middleware's and Meta-Schedulers.

References

- [1] Ian Foster, "The Grid: A New Infrastructure for 21st Century Science", Physics Today
- [2] VO www.egi.eu/community/vos/
- [3] Novotny J, Russell M, Wehrens Q Gridsphere: an advanced portal framework. Euromicro conference 2004, proceedings 30th
- [4] Gridlab portal design document <http://gridlab.org/Resources/Deliverables/D4.4.pdf>
- [5] "Introduction to (Grid) Portals/Gateways" by Anurag Shankar TeraGrid Gateways Team, Indiana University, November 2006
- [6] Liu Xianming, Li Weihua, Li Shixian Using ontology to support portlet semantic interoperability in First International Workshop on Education Technology and Computer Science-2009,ETCS 09.
- [7] Jason Novotny, Micheal Russell, Oliver Wehrens: GridSphere: An Advanced Portal Framework in Proceedings of the 30th EUROMICRO Conference (EUROMI-CRO'04) IEEE
- [8] X509 credentials <http://www.fusiongrid.org/x509certs.html>
- [9] Jason Novotny, Steven Tuecke, Von Welch An online repository for the grid: Myproxy in High performance distributed computing 2001, proceedings of tenth IEEE international symposium
- [10] Vineeth Simon Arackal, Arunachalam B, Bijoy M B, Pahlada Rao B B, Kalasagar B, Sridharan R, Subrata Chattopadhyay. An Access Mechanism for Grid GARUDA. IEEE IMSAA - 2009. Bangalore, India
- [11] Teragrid user Portal <https://portal.xsede.org/>
- [12] Galoh Rashidah Haron, Ng Kang Siong, Sea Chong Seak, Tan Fui Bee "Extending security token to directly issue proxy certificate on P-grade Portal" 2009 IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), October 4-6, 2009, Kuala Lumpur, Malaysia
- [13] Yang, X., Chohan, D., Wang, X. D., Allan, R.: A Web Portal for the National Grid Service. In UK e-Science AHM 2005, Nottingham, UK, available on CDROM,2005
- [14] GPDK <http://doesciencegrid.org/projects/GPDK/>
- [15] Micheal Russel, Jason Novotny, Oliver Wehrens, "Gridsphere Grid Portlets" Computational methods in science and technology 12(1),89-97(2006)
- [16] OGCE http://www.collab-ogce.org/ogce/index.php/Main_Page
- [17] Globus toolkit <http://www.globus.org/toolkit>
- [18] Prachi Pandey, Shamjith KV, Shikha Mehrotra, Asviya B, R Sridharan. "An ingenious approach for improving turnaround time of grid jobs with resource assurance and allocation mechanism" in High Performance Extreme Computing 2012, 10-12th September 2012 HVenue:
- [19] Local Resource Manager http://technical.bestgrid.org/index.php/Local_Resource_Managers_LRM
- [20] Portable Batch System <http://www.pbsworks.com/documentation/support/PBSProUserGuide11.3.pdf>
- [21] Torque scheduler <http://www.clusterresources.com/pages/products/torque-resource-manager.php>
- [22] Sun Grid Engine <http://www.oracle.com/us/products/tools/oracle-grid-engine-075549.html>
- [23] Arunachalam B, Kalasagar B, Vineeth S A, Pahlada rao B B: Open Source Job Submission Portal In PARCOMPTECH, CDAC, Bangalore 2013
- [24] IBM Tivoli Workload Scheduler Loadleveler <http://www.03.ibm.com/systems/clusters/software/loadleveler/index.html>
- [25] Ewing, lusk mathematics and computer science division ,Argonne National Laboratory MPI in 2002: Has it been ten years already?: in Proceedings of the IEEE International Conference on Cluster Computing (CLUSTER'02)
- [26] B B Pahlada Rao, S Ramakrishnan, M R Raja Gopalan, C.Subrata, N Mangala, R sridharan ,e-infrastructure in IT:A case study on indian national grid computing initiative-GARUDA in proceedings of computer science research and development vol 23 – issue 3-4 ,june 2009.