

White Rose Grid

Evaluation of the P-Grade Portal

M.Hewitt, J.Schmidt, S.Kaushal, M.Griffiths, A.Turner
31 March 2009

1. Introduction

We have evaluated the NGS version of the P-Grade portal (based on version 2.4.1), hosted by the University of Westminster, with the purpose of assessing its suitability for adoption within the White Rose Grid. This is an enhanced version with extra capabilities such as access to the Storage Resource Broker (SRB), the OGSA-DAI interface and the Grid Execution Management for Legacy Codes (GEMLCA) software developed by the University of Westminster. It is available at: <https://grid2-portal.cpc.wmin.ac.uk:8080/gridsphere/gridsphere>. Furthermore version 2.7 has been deployed on a local system at the University of York.

2. Overview of the software product

The P-Grade portal offers a web interface for reliable job preparation, job submission and file management, including transfer of input files to and retrieval of output files from remote resources. It uses a standard web browser, thus should be easy accessible to users. The portal offers a stable environment for development of codes on the user's desktop and their submission to the back-end computational resources.

The portal is workflow oriented, and includes tools for creating, editing and managing workflows to be computed on grid resources integrated through the Globus Toolkit (GT). It offers job monitoring and may be configured to enable users to inspect grid information services output. The portal enables users to access computational resources with their proxy certificates.

3. Software provider

The core of the P-Grade portal has been developed by the Laboratory of Parallel and Distributed Systems at MTA SZTAKI in Budapest, Hungary. Further developments are undertaken by The P-Grade Portal Developer Alliance, which currently also includes the Middle East Technical University in Ankara, Turkey, and the University of Westminster in London, UK. The Alliance uses a centralised location at SourceForge.net for the P-Grade source code repository, where it coordinates joint efforts to control and manage the P-Grade portal (open source software) developments. Currently, version 2.7 can be downloaded from this web site.

4. Licence agreement and support

This is open source code, under the GNU General Public License (GPL), available from the http://sourceforge.net/projects/pgportal_web_site.

Throughout the process of the P-Grade portal's evaluation we have been offered good support by staff from the Centre for Parallel Computing at the School of Informatics of the University of Westminster, and MTA SZTAKI in Budapest. The web pages encourage referral of issues with the installation of P-GRADE via email to pgportal@lpds.sztaki.hu.

5. Platform availability

The P-Grade server software may be deployed on systems supporting Scientific Linux 3.08, and gLite User Interface if using EGEE job submission.

Users may access the P-Grade portal from any platform using common web browsers such as Mozilla, Netscape and Internet Explorer. Sheffield staff evaluated the P-Grade using Safari version 3.1.2.

6. Installation and configuration

The P-grade portal (pgportal) server software was installed on a local system at the University of York. The installation required the Globus Toolkit (GT) to be available as pgportal needs access to GT libraries to run correctly. The VDT distribution of the Globus Toolkit was installed.

The following installation steps were carried out according to instructions at <http://portal.p-grade.hu/includes/install.html>

- a) a new user, pgportal, with root rights was created
- b) the .tar.gz was unpacked
- c) the installation configuration file was changed, i.e. the setting of the DISABLE_EGEE variable to yes (DISABLE_EGEE=yes) because the inst_check.pl script complained about the lack of EGEE software.
- d) the value of GLOBUS_FLAVOR in src/c/gridftpools/Makefile was changed to solve the problem with the installation of gridftpools because the software was deployed on a 64-bit system
- e) the portal was deployed successfully and is running on <http://bifrost.wrg.york.ac.uk:8080>.

Documentation states that a number of ports need to be opened on the server in order to communicate with other resources and services; e.g. for MyProxy server, P-GRADE ports (usually 8080, 8443 for https, and 8005 for shutdown) and GLOBUS_TCP_PORT_RANGE needs to be setup (e.g. as required by the LCG setup).

To access the P-Grade portal users need to deploy Java Runtime Environment - JRE 1.4.2_x (or higher) and Java plug in, and have access to a standard web browser.

The NGS P-Grade portal was tested with resources at York and Sheffield which were added to the NGS VO. Jobs were successfully submitted and managed on these resources (globus.wrg.york.ac.uk, iceberg1.sheffield.ac.uk).

7. Functionality

The portal offers a good environment for program development and running production codes. It includes the following:

- a) a graphical interface for job preparation
 - Workflow Editor
 - Workflow Manager
- b) tools for managing user's credentials
 - Certificate Manager
- c) tool for Grid resource configuration
 - Settings (manager)
- d) information services about resources
 - Information System
- f) on-line visualization API for workflow execution monitoring
- g) OGSA-DAI DataBrowser (untested)
- h) uniform interface to heterogeneous data storage resources over a network
 - access to SRB
- g) tools for grid execution management for legacy codes (untested)
 - GEMLCA extension of Workflow Editor
 - GEMLCA extension of Workflow Manager
 - GEMLCA extension of Settings
 - GEMLCA Administration Portlet
 - GEMLCA Monitoring Tool Portlet

The NGS P-Grade version provides a convenient and easy-to-use interface for job preparation and submission on the user's desktop, within the familiar environment. Simple and more complex jobs may be run on a single resource or on a multiple-resource grid. The evaluated NGS P-Grade portal was stable and working well.

Jobs are constructed in a Java web start workflow editor application launched from within the portal. A simple task would consist of only one workflow element, with each element being either a standard or GEMMLCA job. GEMMLCA jobs are pre-defined tasks, usually wrapping up a command line application that can be selected from a list with all of the inputs and outputs for the task pre-defined. Standard jobs allow for more flexibility - the series of tasks to be performed can be uploaded as a script - but require the user to specify what input and output files there will be. The input and output files can either be local to the portal (such input files will be uploaded by the user), reside on the target resource (remote) or exist on an SRB server.

More complex tasks would contain multiple jobs with the inputs for the majority coming from outputs of previous jobs. Each output can also be an input for more than one subsequent job, leading to a tree like structure or even more complex interactions between jobs (refer to Figure 1). This can become tricky to manage if the user is writing scripts that define what each job will do, but if each job is a GEMMLCA job, this can be much easier to work with. It should also be noted that different jobs within a workflow can be configured to run on any resource, independent of where the other jobs are running and the portal will stage the script along with any output files (even if from previous jobs) to that resource.

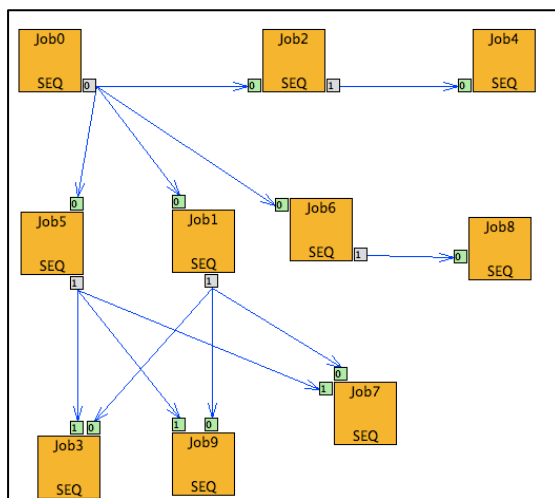


Figure 1: Complex job workflow

Our tests included single and complex jobs using Fortran source code as well as a single step scilab job and a multiprocessor matlab job using the matlab parallel computing toolbox.

Once a workflow is submitted, a graphical interface is available to monitor the progress of the workflow. If a particular part of a workflow fails, it is possible to attach to the running workflow and modify the problematic part and continue without having to restart the workflow from the beginning. The interface allows users to look at the stdout from any of the running jobs and once completed, a zip file containing all of the output files can be downloaded.

Also worth noting is that workflows can be saved, downloaded and send to other users so that they can upload them to their accounts and run from their portal environments.

Caveats:

1. The workflow editor did not update the source files if they were changed and not re-instated as input/output files or a new component of the workflow recreated.
2. Input/output (e.g. executable) filenames cannot have spaces in them. The portal uses the same names internally and it does not handle them well; it leads to file transfer issues when staging the executable.

8. Security

The portal uses a username and password mechanism for initial authentication/login. Access to resources is implemented through X.509 digital certificates and their proxies. Users may deposit their certificates in PEM format in the MyProxy server and then to download a proxy certificate (with

limited lifetime) to a portal environment in order to submit jobs and access resources. There are other options for uploading certificates to the MyProxy server (e.g. command line tools, NGS provided Java apps), all of which work in conjunction with the portal.

The portal uses the https (Hypertext Transfer Protocol over Secure Socket Layer) protocol for connection to resources available via the portal.

9. Development facilities

The P-Grade has been built using the GridSphere portal framework, which 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. Portlet API implementation is JSR 168 compliant. The portal functionality can be extended by building new services as portlets compliant to JSR168 standard e.g. SRB portlet, OGSA-DAI portlet.

Additional tabs may be very easily added to the portal's main tab.

In addition to allowing users to create their own workflows, the portal has an interface for creating GEMMLCA jobs that wrap up common tasks that can be added into subsequent workflows easily. However, neither GEMMLCA jobs nor creating GEMMLCA applications have been tested in detail.

10. Integration with other WRG software

The portal requires the Globus Toolkit (GT) middleware to be installed on its computational resources. All WRG systems should have GT installed and therefore be able to offer resources to the common P-Grade portal.

11. Ease of use

The portal offers an easy to use web interface. It uses a standard web browser, thus should be easily accessible to users.

Implementation of credentials through digital certificates brings a degree of difficulty to its use. This is a workflow-oriented portal and potential users need to understand the concept of workflows for their job preparation, submission and execution. For example, the workflow below (Figure 2) was created to illustrate that there need to be 3 steps executed in order to have a very simple parallel (MPI) program compiled, file execution permission set, and the executable run.

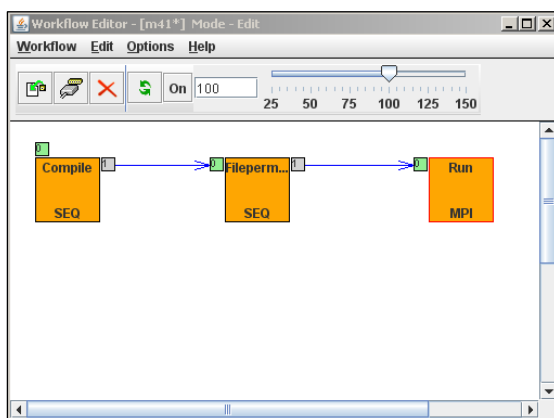


Figure 2: A simple job workflow

Once familiarity is gained with the workflow editor and its terminology, it was not too difficult to build a simple workflow. When jobs had not run as expected it was not always easy to locate the required error files in order to resolve the problem.

Development of workflows for complex jobs such as compilation and running parallel MPI jobs requires much more time, greater skills and understanding of the steps involved in the job preparation and execution.

12. Special features

The NGS P-Grade portal offers additional features such as access to files in the SRB and OGSA-DAI.

The following characteristics might be of particular interest:

- GUI and workflow based interface to Globus resources
- Workflow editing can be made much easier with the use of GEMLCA
- Integrated load information showing which resources are busy and which are not
- GEMLCA jobs can be configured to submit via the resource broker, selecting the best resource automatically.

13. Documentation for system administrators

A *Short Administrator's Manual* of the P-GRADE PORTAL is available at: <http://portal.p-grade.hu/>

The P-GRADE Portal Installation Manual is available at: <http://portal.p-grade.hu/includes/install.html>

14. Documentation for users

Comprehensive user documentation is available from the web site at <http://portal.p-grade.hu/>. It should be noted that this document introduces many grid concepts and may be difficult to read and understand for novice users not familiar with grid terminology.

The following tutorials are available:

- Parameter Study support in P-GRADE Portal 2.5, Principles & Practices
- P-GRADE Portal
- Equation Solver

15. Usage by communities (in general, and locally)

According to the portal's web page at <http://portal.p-grade.hu/?m=installations&s=0> there are a significant number of portal installations, particularly in Europe, which offer resources of multi-institutional grids. Specifically, these include the NGS user community, EGEE users, Grid-Ireland, Turkish Grid operated by ULAKBIM & METU, Baltic Grid Portal operated by Estonian Educational and Research Network (EENet), CROGRID operated by Boskovic Research Institute in Croatia, icesaga portal operated by Supercomputing Center of Galicia (CESGA) and Bulgarian Grid portal operated by Bulgarian Academy of Sciences.

16. Suitability for the WRG

The portal is suitable for deployment on the White Rose Grid. There is potentially a large user base that could find this portal useful as it requires very little knowledge of command line Linux environments. It could also be useful for more knowledgeable users as they could use it to construct large, complex workflows that run over a long time and not have to manually perform each step.

17. Conclusions and recommendations

The evaluated portal performed well and a variety of jobs (e.g. a simple serial job, complex serial jobs executing on distributed resources as well as parallel jobs) were submitted and computed on computational resources known to P-Grade. The portal offered a stable environment for development of codes on the user's desktop and their submission to the back-end computational resources.

This portal offers a robust and stable environment for the whole process of job preparation, submission and execution. Potentially it could be very useful to new WRG users who do not like to login to Linux systems and would prefer to submit their jobs to WRG HPC systems directly from their desktop. It is anticipated that training and simplified documentation will need to be developed to encourage the use of the portal.

It is recommended that a trial service be offered for selected users within the White Rose. This service will be reviewed in 6 months time.

References

User documentation <http://portal.p-grade.hu/>

Globus Toolkit www.globus.org

University of Westminster Centre for Parallel Computing
http://www.cpc.wmin.ac.uk/cpcsite/index.php/Main_Page

Laboratory of Parallel and Distributed Systems at MTA-SZTAKI, Hungary, <http://www.lpds.sztaki.hu/>