



Contribution ID: 144

Type: Demo

Supporting Parametric Study Workflow Applications by the P-GRADE Portal

Tuesday 26 September 2006 17:00 (20 minutes)

The P-GRADE portal plays more and more important role in the life of various Grid user communities. After several successful demos at the biggest conferences and Grid user forums in Europe, Asia and the US, the representatives of several Grids and Grid based Virtual Organizations have approached us and requested to support their communities by the Portal. As a result, the P-GRADE portal is already the official portal of the VOCE (Virtual Organization Central Europe), HunGrid (Hungarian VO of EGEE) and the eGrid (Economic Grid) VOs of the EGEE Grid. It also provides service for the users of GILDA (the Grid training infrastructure of EGEE), Croatian Grid and Turkish Grid infrastructures. Moreover, the P-GRADE portal is the official portal of SEE-GRID which operates a Grid infrastructure in the South-East Europe region. Besides LCG-2 and gLite based production Grids the portal is successfully used as service for the GT2 based UK National Grid Service (NGS) and it was also successfully connected to the GT4 based Westfocus Grid (UK). Our latest achievement is that the P-GRADE Portal has been connected to the ARC middleware, thus now it is able to execute Grid applications in the Nordugrid too. After a successful demonstration at the Supercomputing'05 exhibition the representatives of the US Open Science Grid and Teragrid also expressed their interest to connect the portal to their Grid. Consequently, the P-GRADE Portal is now connected to both OSG and Teragrid, reaching the users of many large production Grid infrastructures of the World. Moreover, recently the GIN (Grid Interoperation/Interoperability Now) Grid of GGF is

supported by the portal
enabling the simultaneous access to all of its resources
coming from different
Grids.

As P-GRADE portal becomes more and more popular among users
we have

received important feedbacks asking for new features of the
portal. One of
those requests was the support of parametric studies at the
workflow level.

The idea of parametric study applications is that the same
workflow should be
executed with a large number of different input data files.
Moreover, different
jobs must be fed by different number of files and the portal
should be able to
automatically generate the cross-product of these input
files and run the
workflow for each element of this cross-product. Obviously
handling large
number of workflows and files raises many new problems. Here
we mention
only some of them just to illustrate the problems:

1. Where to place and how to organize the necessary input files?
2. Where and how to store the output of each execution of the workflow?
3. How to prevent flooding the Grid and the portal by large parameter study applications?
4. How to specify the parametric study workflows in a way that simply extends the specification of normal workflows?
5. How to manage the large number of workflows by the portal?

The main principles of supporting parametric study
application by P-GRADE
portal are as follows:

1. Any port of a PS-WF (parametric study workflow) can be used to feed many files to the WF. Such a port is called as PS-port and distinguished from the ordinary input ports both in the UI and in the inner representation of the WF. For each PS-port in a PS-WF there is a unique integer identifier (an ordering number starting from 0) generated by the portal.
2. A PS-port represents a set of input files that are stored in the same directory of a SE (storage element). It is the responsibility of the user to place these input files into the SE before submitting the PS-WF. Such a directory must not store any other files, only the input files belonging to the associated PS-port.
3. If there are several PS-ports in a PS-WF, then the portal RS (run-time system) takes care of producing the necessary cross-product of the input files of these PS-ports.
4. For each element of the cross-product the RS generates an executable WF (e-WF). The internal representation of an e-WF is the same as the normal WF.
5. Once the RS generated an e-WF it submits this e-WF in the same way as normal e-WFs are submitted (since they are the same).
6. The number (N) of e-WFs that are generated for parallel execution is the decision of the portal. When a PS-WF is submitted, the portal RS generates N e- WFs of the cross product and submit them simultaneously to the Grid. Once an e-WF is completed the portal RS generates the next element of the cross product and the related e-WF and submits it into the Grid.
7. An extra global parameter of a PS-WF is the target output directory of the workflow results. The target output directory must be in a SE.
8. Once an e-WF is completed the portal moves the zipped result into the target output directory. As a result not more than N partial WF results should be stored on the portal simultaneously for one PS-WF. Any post-processing of the results is the task of the user and not of the portal.
9. To avoid the flooding of the Grid and portal by a single user, one user can submit only one PS-WF at a time. The next one can be submitted if the previously submitted PS-WF is completed. Moreover, the portal administrator can set the maximum number of e-WFs that can be simultaneously generated from a single PS-WF as well as the maximum number of jobs that can be submitted by the portal to the Grid.

During the demo of the portal we will demonstrate how the new PS support feature of P-GRADE portal works. We hope that this new feature will significantly increase the usability of the portal and will open doors for many new Grid user communities.

Primary author: KACSUK, Peter (Prof.)

Co-authors: SIPOS, Gergely (Mr.); LOVAS, Robert (Mr.)

Presenters: KACSUK, Peter (Prof.); LOVAS, Robert (Mr.)

Session Classification: Demo session

Track Classification: Users & Applications