



Contribution ID: 171

Type: Poster

Deploying Job Provenance: First Application Experience

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

Here we demonstrate a generic gLite middleware service - Job Provenance - providing a backbone for custom application solutions requiring job catalogue capabilities. The target communities vary from small research groups trying to set up their own solution fulfilling their specific needs up to potential usage in well established communities like the high energy particle experiments.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

We demonstrate JP usage in two cases. In the first one, management of large parametric studies in computational chemistry (molecular docking) JP, together with a thin graphical front-end, was used to build the job catalogue from scratch. It allowed the researchers to easily manipulate computational jobs (input modification, jobs resubmission), to search and selected desired (finished/non-finished, aborted) jobs and finally utilize specific plugins for results presentation (e.g. visualization). In the second case, we augmented production jobs of the Atlas experiment to interact with JP, yielding functionality similar to Atlas ProdDB but with emphasis in job history. We routed part of the Atlas production traffic to JP (approx. 1100 jobs/day) as well as performed stress tests on snapshot of these jobs in order to demonstrate JP readiness for production deployment.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

gLite Job Provenance (JP) is a generic job catalogue service keeping long-term track of execution of Grid jobs. It provides a sophisticated machinery to support application and user annotations of the Grid computational jobs. Furthermore it provides data mining over the raw data and

annotations. Being a standard part of gLite middleware stack it offers continuous and guaranteed service to store all the primary information. On top of storage facilities, Job Provenance Index Servers allow for efficiently looking for expected and unexpected patterns within the stored information through user queries. While JP can be used directly by all gLite middleware users, specialized job catalogues can be built with moderate effort compared to custom solutions (custom job catalogue development) taking considerable effort.

Primary authors: Mr KRENEK, Ales (CESNET); DVORAK, Frantisek (CESNET); KMUNICEK, Jan (CESNET); FILIPOVIC, Jiri (CESNET); SITERA, Jiri (CESNET); MATYSKA, Ludek (CESNET); MULAC, Milos (CESNET); RUDA, Miroslav (CESNET); SALVET, Zdenek (CESNET); SUSTR, Zdenek (CESNET)

Presenters: Mr KRENEK, Ales (CESNET); KMUNICEK, Jan (CESNET)

Track Classification: Demo and Poster session