Contribution ID: **448**   Type: **poster**

Experimental Evaluation of Job Provenance in ATLAS environment

Wednesday 5 September 2007 08:00 (20 minutes)

Grid middleware stacks, including gLite, matured into the state of being able to process upto millions of jobs per day. Logging and Bookkeeping, the gLite job-tracking service keeps pace with this rate, however it is not designed to provide a long-term archive of executed jobs.

ATLAS—representative of large user community— addresses this issue with its own job catalogue (prodDB). Development of such a customized service took considerable effort which is not easily affordable by smaller communities and is not easily reused.

On the contrary, Job Provenance (JP) is a generic gLite service designed for long-term archive of information on executed jobs. Its design priorities are:
(i) scalability – store data on billions of jobs;
(ii) extensibility – virtually any data format can be uploaded and handled by plugins;
(iii) uniform data view – all data are logically transformed into RDF-like data model, using appropriate namespaces to avoid ambiguities;
(iv) configurability – highly customizable components maintaining pre-cooked queries provide efficient query interface.

We present first results of experimental JP deployment for the ATLAS production infrastructure. JP installation was fed with a part of ATLAS production jobs (thousands of jobs per day). We provide a functional comparison of JP and ATLAS prodDB, discuss reliability, performance and scalability issues, and focus on the application level functionality as opposed to pure Grid middleware functions.

The main outcome of this work is a demonstration that JP can complement large-scale application-specific job catalogue services, as well as serve similar purpose where these are not available.

Authors: KRENEK, Ales (CESNET); MATYSKA, Ludek (CESNET)

Co-authors: DVORAK, Frantisek (CESNET); CHUDOBA, Jiri (CESNET); FILIPOVIC, Jiri (CESNET); SITERA, Jiri (CESNET); PERINI, Laura (INFN); MULAC, Milos (CESNET); RUDA, Miroslav (CESNET); CAMPANA, Simone (CERN); SALVET, Zdenek (CESNET); SUSTR, Zdenek (CESNET)

Presenter: MATYSKA, Ludek (CESNET)

Session Classification: Poster 2

Track Classification: Grid middleware and tools