

Update on gathering of job statistical information via L&B/JP

A. Křenek, G. Romier

The procedure

- data collected from RB sites (over 30)
- stored and processed at JRA2 site

Principal drawbacks

- scalability
- direct access to L&B database
 - private, subject to change
 - different interpretation wrt. job-status output

The procedure

- use regular (daily) L&B dumps
 - private raw format (ULM)
 - present in LCG 2.x
 - should be done anyway to purge active L&B database
- processed by custom utility \Rightarrow agreed XML format
 - sustained 10 jobs per second \sim 1.7 GB per day
- digested by statistics tool

Status

- agreed on the procedure
- list of required data fields assessed
 - approx. 40 in total
 - approx. 3/4 can be filled from L&B data
- XML schema being finalised
- ready to start development of the conversion utility
 - expected for gLite release 2
 - backward compatible with dumps from current LCG
- L&B dump deployment pending

Job Provenance service

- designed to store data on jobs for long time
- permanent Primary Storage
- volatile configurable Index Servers
- JP plugin interface – any type and version of data

Foreseen usage

- bootstrap and daily queries: use JP “feed” protocol (WS)
 - client: I’m listening on host:port, want attr1, attr2, ... of yesterday jobs
 - server: starts feeding the client with reasonable-size batches
- custom-configured JP index servers to provide on-line subset
 - aggregation functions not defined in JP yet
 - plugin interface may be required for more complex ones