Open Science Grid

OSG
Area Coordinator's Report:
Workload Management

April 20th, 2011

Maxim Potekhin

BNL

631-344-3621

potekhin@bnl.gov

Summary of Workload Management: Panda

Open Science Grid

- **WBS item 2.2.1.2: Panda Pilot interface with site authentication/authorization systems**

  - Resolved a number of configuration/permission issues on various sites used by Atlas, glexec capable pilot ready for production

- **WBS item 2.2.3.1: light-weight data movement**

  - Completed as gridFTP plug-in for existing clients (Daya Bay). We'll be looking toward Globus for future implementation and integration (lots of progress made there, spoke with Steve Tuecke at OSG AHM in March 2011), further planning will depend on demand (coordinating with Atlas).

- **WBS item 2.2.4: Panda Monitoring**

  - As indicated in the previous report, the monitoring effort was reshaped to use and enhance previously existing components as opposed to a complete overhaul, and is now being handled by Atlas personnel. There may be re-use of components previously created for this work item, but for now it needs to be struck out of WBS (or marked otherwise), and the effort transferred to 2.2.6 (which is the case "de facto").

Summary of Workload Management: Panda

Open Science Grid

- ## WBS item 2.2.5.1: support of Daya Bay/LBNE
    - Currently running production of approximately 1500 jobs per week, on 2 Condor pools at BNL
    - Working on Condor issues (pilots not matching certain nodes), some tuning needed
    - Active monitoring of disk space on submission node and that of the logger web service

- ## WBS item 2.2.6: Scalability of Panda DB
    - Significant progress has been made in configuration and stress testing of a noSQL solution (Cassandra) on two different test facilities, one at CERN and another at BNL
    - Based on query patterns, the design of data and indexes has been modified
    - Deployed a 3-node cluster at BNL and performed data load of a significant and large part of Panda job data (1 year worth)
    - Collected metrics and where possible, provided comparison to similar queries done against the production instance of Oracle DB at CERN
    - Analysis still under way, the aim being to determine whether we need more horizontal scaling with potentially a large number of smaller nodes
    - Good level of support from RACF at BNL, collaboration with Atlas personnel at CERN
    - Now officially a Task Force mandate for R&D in Atlas Distributed Computing Organization with specific deliverables later in 2011 (essentially a working database)