Contribution ID: **347**   Type: **oral presentation**

Panda: Production and Distributed Analysis System for ATLAS

Wednesday 15 February 2006 15:00 (20 minutes)

A new offline processing system for production and analysis, Panda, has been developed for the ATLAS experiment and deployed in OSG. ATLAS will accrue tens of petabytes of data per year, and the Panda design is accordingly optimized for data intensive processing. Its development followed three years of production experience, the lessons from which drove a markedly different design for the new system.

Key design features include
- data-driven workflow with managed placement of datasets (file collections) at processing sites; job-placement that assures jobs are dispatched to sites holding their input data
- tight integration with the ATLAS DDM system Don Quijote 2 which provides all data management services
- late binding of jobs to worker nodes for dynamic and flexible prioritization and scheduling, and to isolate workloads from latencies and failure modes in acquiring processing resources
- a service oriented architecture with a fast and lightweight communication layer based on REST-style web services
- queue management, data placement, job dispatching, and processor acquisition operate as asynchronous services, with support for multiple instances, for fast throughput and maximum scalability.

In this paper we motivate and describe the design and implementation of the system, the current state of its deployment for production and analysis operations, and the work remaining to achieve readiness for ATLAS datataking.

Primary authors: Prof. DE, Kaushik (UNIVERSITY OF TEXAS AT ARLINGTON); Dr WENAUS, Torre (Brookhaven National Laboratory)

Co-authors: Dr KLIMENTOV, Alexei (Brookhaven National Laboratory); Dr ADAMS, David (Brookhaven National Laboratory); Dr SEVERINI, Horst (University of Oklahoma); Mr ARUNACHALAM, Karthik (University of Oklahoma); Dr MAMBELLI, Marco (University of Chicago); DR SOSEBEE, Mark (UNIVERSITY OF TEXAS AT ARLINGTON); Dr OZTURK, Nurcan (UNIVERSITY OF TEXAS AT ARLINGTON); Dr NEVSKI, Pavel (Brookhaven National Laboratory); Mr MAENO, Tadashi (Brookhaven National Laboratory); Dr DENG, Wensheng (Brookhaven National Laboratory); Dr ZHAO, Xin (Brookhaven National Laboratory); Dr SMIRNOV, Yuri (Brookhaven National Laboratory)

Presenter: Prof. DE, Kaushik (UNIVERSITY OF TEXAS AT ARLINGTON)

Session Classification: Distributed Data Analysis

Track Classification: Distributed Data Analysis