



Contribution ID: 274

Type: **Parallel**

## PD2P : PanDA Dynamic Data Placement for ATLAS

*Thursday, May 24, 2012 5:50 PM (25 minutes)*

The PanDA Production and Distributed Analysis System is the ATLAS workload management system for processing user analysis, group analysis and production jobs.

In 2011 more than 1400 users have submitted jobs through PanDA to the ATLAS grid infrastructure. The system processes more than 2 million analysis jobs per week. Analysis jobs are routed to sites based on the availability of relevant data and processing resources, taking account of the nonuniform distribution of CPU and storage resources in the ATLAS grid. The data distribution has to be optimized to fit the resource distribution, and also has to be dynamically changed to meet rapidly evolving requirements for analysis use cases. The PanDA Dynamic Data Placement (PD2P) system has been developed to cope with difficulties of data placement for ATLAS. PD2P is an intelligent subsystem of PanDA to distribute data by taking the following factors into account: popularity, locality, the usage pattern of the data, the distribution of CPU and storage resources, network topology between sites, site operation downtime and reliability, and so on. We will describe the design of the new system, its performance during the past year of data taking, dramatic improvements it has brought about in the efficient use of storage and processing resources, associated reductions in average wait time for user analysis jobs, and plans for the future.

**Primary author:** ATLAS, Collaboration (Atlas)

**Co-authors:** DE, Kaushik (University of Texas at Arlington (US)); PANITKIN, Sergey (Brookhaven National Laboratory (US)); MAENO, Tadashi (Brookhaven National Laboratory (US))

**Presenter:** MAENO, Tadashi (Brookhaven National Laboratory (US))

**Session Classification:** Distributed Processing and Analysis on Grids and Clouds

**Track Classification:** Distributed Processing and Analysis on Grids and Clouds (track 3)