



Contribution ID: 404

Type: **Parallel contribution**

ATLAS Analysis Data Distribution and Panda PD2P

Friday 12 August 2011 11:10 (20 minutes)

The PanDA Distributed Analysis system has been used in the ATLAS collaboration and beyond as a resilient and scalable distributed processing and analysis system. Using a central pull and distributed push (pilot job) model for task definition and job tracking, it integrates with many kinds of local batch system, data management software, and security models.

One of the principal challenges in making user jobs responsive comes from data location – since jobs go to the data, popular datasets at limited numbers of locations will attract too much user activity for the site's resources. The data are too large, however, to pre-position at all sites. PanDA has pioneered an approach to data management integration called P2DP, which automates data distribution to user analysis sites based on usage and popularity of particular datasets. By tuning the parameters that trigger these data replications, we optimize the balance between the data replication and user concentration.

The strengths and tradeoffs of both the PanDA pilot and the P2DP model will be discussed, and we will examine throughput and efficiency, security versus flexibility, and the ongoing process of tuning the system to be more responsive and intelligent.

Author: Dr STRADLING, Alden (UT Arlington)

Presenter: Dr STRADLING, Alden (UT Arlington)

Session Classification: Computing in HEP

Track Classification: Computing in HEP