

The PanDA Distributed Production and Analysis System

Torre Wenaus

BNL, US

wenaus@gmail.com

The PanDA workload management system for distributed production and analysis has been developed since fall 2005 to meet the requirements of Petabyte scale data processing for the ATLAS experiment at the LHC. Principal architectural features of PanDA include a service architecture based on the LAMP stack, a late binding or 'pilot job' approach to job management and submission, a central job queue offering flexible brokerage and a unified system view, tight integration with data management, and a heavy emphasis on automation, error recovery and monitoring to maximize efficiency and operational scalability. In this talk we will describe the PanDA system, operational experience with it in ATLAS production, and current priorities and directions for its development, including security issues. The talk will also touch on PanDA's role in the workload management activities of the Open Science Grid.