



IEEE | **20**
Cluster | **14**

Madrid 22-26 September

Sergey Panitkin

Cluster 14

- IEEE Conference in Madrid, Sep. 22-26, 2014
- <http://www.cluster2014.org/>
- Main Topics
 - System Design configuration, Administration
 - System Software, Middleware, Tools
 - Resource and job management
 - Storage and File Systems
 - Algorithms, Applications and Performance
- Cluster 14 Highlight: Cluster support for Big Data

Timeline

- Abstracts submission: April 24
- Papers: May 2
- Conference: Sep 22-26

Suggested Abstract

- PanDA - workload management system for Big Data exploration
- Experiments at the Large Hadron Collider (LHC) face unprecedented computing challenges. Heterogeneous resources are distributed worldwide, thousands of physicists analyzing the data need remote access to hundreds of computing sites, the volume of processed data is beyond the exabyte scale, and data processing requires more than billions of hours of computing usage per year. The PanDA (Production and Distributed Analysis) system was developed to meet the scale and complexity of LHC distributed computing for the ATLAS experiment. In the process, the old batch job paradigm of computing in HEP was discarded in favor of a far more flexible and scalable model. The success of PanDA at the LHC is leading to widespread adoption and testing by other experiments. PanDA is the first exascale workload management system in HEP, already operating at a million computing jobs per day, and processing over an exabyte of data in 2013. We will describe the design and implementation of PanDA, present data on the performance of PanDA at the LHC, and discuss plans for future evolution of the system to meet new challenges of scale, heterogeneity and increasing user base.

Paper

- Started paper preparation
- Past ATLAS deadlines. Need to speed up
- Based on recent PanDA results and presentations
 - Structure
 - Introduction +
 - PanDA in ATLAS
 - Clouds +
 - HPC +