



Contribution ID: 136

Type: Plenary

## Next Generation Workload Management System for Big Data on Heterogeneous Distributed Computing

*Tuesday 2 September 2014 11:15 (35 minutes)*

The Large Hadron Collider (LHC), operating at the international CERN Laboratory in Geneva, Switzerland, is leading Big Data driven scientific explorations. Experiments at the LHC explore the fundamental nature of matter and the basic forces that shape our universe, and were recently credited for the discovery of a Higgs boson. ATLAS, one of the largest collaborations ever assembled in the sciences, is at the forefront of research at the LHC. To address an unprecedented multi-petabyte data processing challenge, the ATLAS experiment is relying on a heterogeneous distributed computational infrastructure. The ATLAS experiment uses PanDA (Production and Data Analysis) Workload Management System for managing the workflow for all data processing on hundreds of data centers. Through PanDA, ATLAS physicists see a single computing facility that enables rapid scientific breakthroughs for the experiment, even though the data centers are physically scattered all over the world. The scale is demonstrated by the following numbers: PanDA manages  $O(102)$  sites,  $O(105)$  cores,  $O(108)$  jobs per year,  $O(103)$  users and ATLAS Data Volume is  $O(10^{17})$  bytes. In 2013 we started an ambitious program to expand PanDA to all available computing resources, including opportunistic use of commercial and academic clouds and Leadership Computing Facilities (LCF). The project titled 'Next Generation Workload Management and Analysis System for Big Data' (BigPanDA) is funded by DOE ASCR and DOE HEP. Extending PanDA to clouds and LCF presents new challenges in managing heterogeneity and supporting workflow. The BigPanDA project is underway to setup and tailor PanDA at Oak Ridge Leadership Computing Facility (OLCF), and at National Research Center "Kurchatov Institute" together with ALICE Distributed Computing and ORNL computing professionals. Our approach for integration of the HPC platforms at OLCF and

elsewhere is to reuse, as much as possible, existing components of the PanDA system. The next generation of PanDA will allow many data-intensive sciences employing a variety of computing platforms to benefit from ATLAS' and ALICE experience and proven tools in highly scalable processing. We will present our current accomplishments with running PanDA WMS at OLCF and other supercomputers and demonstrate our ability to use PanDA as a portal independent of the computing facilities infrastructure for High Energy and Nuclear Physics as well as other data-intensive science applications.

**Author:** Dr KLIMENTOV, Alexei (Brookhaven National Laboratory (US))

**Co-authors:** VANIACHINE, Alexandre (ATLAS); OLEJNIK, Danila (Joint Inst. for Nuclear Research (RU)); Dr WELLS, Jack (ORNL); PORTER, Jeff (Lawrence Berkeley National Lab. (US)); DE, Kaushik (University of Texas at Arlington (US)); READ, Kenneth (Oak Ridge National Laboratory - (US)); NILSSON, Paul (University of Texas at Arlington (US)); BUNCIC, Predrag (CERN); Dr MOUNT, Richard Philip (SLAC National Accelerator Laboratory (US)); PANITKIN, Sergey (Brookhaven National Laboratory (US)); Prof. JHA, Shantenu (Rutgers University); MAENO, Tadashi (Brookhaven National Laboratory (US)); Dr WENAUS, Torre (Brookhaven National Laboratory (US))

**Presenter:** Dr KLIMENTOV, Alexei (Brookhaven National Laboratory (US))

**Session Classification:** Plenary

