Contribution ID: **32**   Type: **not specified**

Integration Of PanDA Workload Management System With Supercomputers

Friday 2 October 2015 10:00 (30 minutes)

Abstract. 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 over 140 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.

While PanDA currently uses more than 100,000 cores with a peak performance of 0.3 petaFLOPS, next LHC data taking runs will require more resources than Grid computing can possibly provide. To alleviate these challenges, LHC experiments are engaged in an ambitious program to expand the current computing model to include additional resources such as the opportunistic use of supercomputers.

We will describe a project aimed at integration of PanDA WMS with supercomputers in United States, Europe and Russia (in particular with Titan supercomputer at Oak Ridge Leadership Computing Facility (OLCF), Supercomputer at the National Research Center "Kurchatov Institute", IT4 in Ostrava and others).

Current approach utilizes modified PanDA pilot framework for job submission to the supercomputers batch queues and local data management, with light-weight MPI wrappers to run single threaded workloads in parallel on Titan's multi-core worker nodes.

This implementation was tested with a variety of Monte-Carlo workloads on several supercomputing platforms. We will present our current accomplishments with running PanDA WMS at 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, such as bioinformatics and astro-particle physics.

Author: Dr KLIMENTOV, Alexei (Brookhaven National Lab)

Co-authors: Mr NOVIKOV, Alexander (NRC KI); Dr POYDA, Alexei (NRC KI); Mr OLEYNIK, Danila (JINR LIT); Mr RYABINKIN, Eygene (NRC KI); Dr WELLS, Jack (ORNL); Prof. DE, Kaushik (UTA); Prof. READ, Ken (ORNL); Dr NILSSON, Paul (BNL); Dr MASHINISTOV, Ruslan (NRC KI); Dr PANITKIN, Sergey (BNL); Prof. JHA, Shantenu (Rutgers University); Dr MAENO, Tadashi (BNL); Dr WENAUS, torre (BNL)

Presenter: Dr KLIMENTOV, Alexei (Brookhaven National Lab)

Session Classification: Computing for Large Scale Accelerator Facilities (LHC, FAIR, NICA, etc.) and Big Data