Next Generation Workload Management System For BigData and ATLAS
on Heterogeneous Distributed Computing Facilities

K.De, S.Jha, A.Klimentov, T.Maeno, P.Nilsson, D.Oleynik, S.Panitkin, K.Read,
T.Wenaus on behalf of ATLAS Collaboration

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 $O(10^{2})$ sites,
$O(10^{5})$ cores, $O(10^{8})$ jobs per year, $O(10^{3})$ 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. 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 Facilities (OLCF). 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.
 We will present our current accomplishments with running PanDA WMS at
OLCF and other super-computers and demonstrate our ability to use PanDA
as a portal independent of the computing facilities infrastructure for High
Energy Physics as well as other data-intensive science applications.