

# The 6th International Conference "Distributed Computing and Grid-technologies in Science and Education"



Contribution ID: 37

Type: **plenary reports**

## Workload Management System for Big Data on Heterogeneous Distributed Computing Resources

*Wednesday 2 July 2014 12:45 (15 minutes)*

Many research areas in the natural sciences face unprecedented computing challenges. For example, experiments at the Large Hadron Collider (LHC) use heterogeneous resources which are distributed worldwide, thousands of scientists 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 billions of hours of computing usage per year.

The PanDA (Production and Distributed Analysis) system was initially developed to meet the scale and complexity of LHC distributed computing for the ATLAS experiment. In the process, the local 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 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 every year.

In 2013 we started the project titled 'Next Generation Workload Management and Analysis System for Big Data' to expand PanDA to additional computing resources, including opportunistic use of commercial and academic clouds and Leadership Computing Facilities (LCF). Extending PanDA to clouds and LCF presents new challenges in managing heterogeneity and supporting complex workflows.

We will describe the design and implementation of PanDA, present data on the performance of PanDA, and discuss plans for future evolution of the system to meet new challenges of scale, heterogeneity and increasing user base.

**Primary author:** Mr OLEYNIK, Danila (JINR/UTA)

**Presenter:** Mr OLEYNIK, Danila (JINR/UTA)

**Session Classification:** Plenary

**Track Classification:** Section 1 - Technologies, architectures, models, methods and experiences of building distributed computing systems. Consolidation and integration of distributed resources