



Contribution ID: 239

Type: **Oral presentation to parallel session**

## Reliability Engineering analysis of ATLAS data reprocessing campaigns

*Monday, 14 October 2013 13:52 (22 minutes)*

During three years of LHC data taking, the ATLAS collaboration completed three petascale data reprocessing campaigns on the Grid, with up to 2 PB of data being reprocessed every year. In reprocessing on the Grid, failures can occur for a variety of reasons, while Grid heterogeneity makes failures hard to diagnose and repair quickly. As a result, Big Data processing on the Grid must tolerate a continuous stream of failures, errors and faults. While ATLAS fault-tolerance mechanisms improve the reliability of Big Data processing in the Grid, their benefits come at costs and result in delays making the performance prediction difficult.

Reliability Engineering provides a framework for fundamental understanding of the Big Data processing on the Grid, which is not a desirable enhancement but a necessary requirement. In ATLAS, cost monitoring and performance prediction became critical for the success of the reprocessing campaigns conducted in preparation for the major physics conferences. In addition, our Reliability Engineering approach supported continuous improvements in data reprocessing throughput during LHC data taking. The throughput doubled in 2011 vs. 2010 reprocessing, then quadrupled in 2012 vs. 2011 reprocessing.

We present the Reliability Engineering analysis of ATLAS data reprocessing campaigns providing the foundation needed to scale up the Big Data processing technologies beyond the petascale.

### Summary

**Primary author:** Dr VANIACHINE, Alexandre (ANL)

**Co-authors:** GOLUBKOV, Dmitri (Institute for High Energy Physics (IHEP)-Unknown-Unknown); KARPENKO, Dmytro (University of Oslo (NO))

**Presenter:** KARPENKO, Dmytro (University of Oslo (NO))

**Session Classification:** Distributed Processing and Data Handling A: Infrastructure, Sites, and Virtualization

**Track Classification:** Distributed Processing and Data Handling A: Infrastructure, Sites, and Virtualization