Contribution ID: **34**   Type: **not specified**

Data analytics in the ATLAS Distributed Computing

Friday, 2 October 2015 09:30 (30 minutes)

The ATLAS Data analytics effort is focused on creating systems which provide the ATLAS ADC with new capabilities for understanding distributed systems and overall operational performance. These capabilities include: warehousing information from multiple systems (the production and distributed analysis system - PanDA, the distributed data management system - Rucio, the file transfer system, various monitoring services etc.); providing a platform to execute arbitrary data mining and machine learning algorithms over aggregated data; satisfy a variety of use cases for different user roles; host new third party analytics services on a scalable compute platform. We describe the implemented system where: data sources are existing RDBMS (Oracle) and Flume collectors; a Hadoop cluster is used to store the data; native Hadoop and Apache Pig scripts are used for data aggregation; and R for in-depth analytics. Part of the data is indexed in ElasticSearch so both simpler investigations and complex dashboards can be made using Kibana.

Primary author: Dr VUKOTIC, Ilija (University of Chicago)

Co-authors: Mr BRYANT, Lincoln (University of Chicago);   Dr GARDNER, Robert (University of Chicago)

Presenter: Dr VUKOTIC, Ilija (University of Chicago)

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