**38th INTERNATIONAL CONFERENCE
ON HIGH ENERGY PHYSICS**

AUGUST 3 - 10, 2016
CHICAGO

Contribution ID: **781**   Type: **Oral Presentation**

Getting the Most from Distributed Resources: an Analytics Platform for ATLAS Computing Services (15' + 5')

Saturday 6 August 2016 16:55 (20 minutes)

To meet a sharply increasing demand for computing resources in LHC Run 2, ATLAS distributed computing systems reach far and wide to gather CPU and storage capacity to execute an evolving ecosystem of production and analysis workflow tools. Indeed more than a hundred computing sites from the Worldwide LHC Computing Grid, plus many "opportunistic"facilities at HPC centers, universities, national laboratories, and public clouds, combine to meet these requirements. These resources have characteristics (such as local queuing availability, proximity to data sources and target destinations, network latency and bandwidth capacity, etc.) affecting the overall processing throughput. To quantitatively understand and in some instances predict behavior, we have developed a platform to aggregate, index (for user queries), and analyze the more important information streams affecting performance. These data streams come from the ATLAS production system (PanDA) and distributed data management system (Rucio), the network (throughput and latency measurements, aggregate link traffic), and from the computing facilities themselves. The platform brings new capabilities to the management of the overall system, including warehousing information, an interface to execute arbitrary data mining and machine learning algorithms over aggregated datasets, a platform to test usage scenarios, and a portal for user-designed analytics dashboards.

Author: ATLAS, Collaboration (ATLAS)

Presenter: VUKOTIC, Ilija (University of Chicago (US))

Session Classification: Computing

Track Classification: Computing and Data Handling