

# Exploiting analytics techniques in CMS computing monitoring

*Monday, 10 October 2016 14:15 (15 minutes)*

The CMS experiment has collected an enormous volume of metadata about its computing operations in its monitoring systems, describing its experience in operating all of the CMS workflows on all of the Worldwide LHC Computing Grid Tiers. Data mining efforts into all these information have rarely been done, but are of crucial importance for a better understanding of how CMS did successful operations, and to reach an adequate and adaptive modelling of the CMS operations, in order to allow detailed optimizations and eventually a prediction of system behaviours. These data are now streamed into the CERN Hadoop data cluster for further analysis. Specific sets of information (e.g. data on how many replicas of datasets CMS wrote on disks at WLCG Tiers, data on which datasets were primarily requested for analysis, etc) were collected on Hadoop and processed with MapReduce applications profiting of the parallelization on the Hadoop cluster. We present the implementation of new monitoring applications on Hadoop, and discuss the new possibilities in CMS computing monitoring introduced with the ability to quickly process big data sets from multiple sources, looking forward to a predictive modeling of the system.

## Primary Keyword (Mandatory)

Monitoring

## Secondary Keyword (Optional)

## Tertiary Keyword (Optional)

**Primary authors:** BONACORSI, Daniele (University of Bologna); MAGINI, Nicolo (Fermi National Accelerator Lab. (US))

**Co-author:** VAANDERING, Eric (Fermi National Accelerator Lab. (US))

**Presenter:** VAANDERING, Eric (Fermi National Accelerator Lab. (US))

**Session Classification:** Track 7: Middleware, Monitoring and Accounting

**Track Classification:** Track 7: Middleware, Monitoring and Accounting