



Contribution ID: 199

Type: **Sectional**

Applying Big Data solutions for log analytics in the PanDA infrastructure

Thursday 28 September 2017 12:30 (15 minutes)

PanDA is the workflow management system of the ATLAS experiment at the LHC and is responsible for generating, brokering and monitoring up to two million jobs per day across 150 computing centers in the Worldwide LHC Computing Grid. The PanDA core consists of several components deployed centrally on around 20 servers. The daily log volume is around 400GB per day. In certain cases, troubleshooting a particular issue on the raw log files can be compared to searching for a needle in a haystack and requires a high level of expertise. Therefore we decided to build on trending Big Data solutions and utilize the ELK infrastructure (Filebeat, Logstash, Elastic Search and Kibana) to process, index and analyze our log files. This allows to overcome troubleshooting complexity, provides a better interface to the operations team and generates advanced analytics to understand our system.

This paper will describe the features of the ELK stack, our infrastructure, optimal configuration settings and filters. We will provide examples of graphs and dashboards generated through the ELK system to demonstrate the potential of the system. Finally, we will show the current integration of Kibana with the PanDA monitoring frontend and other usage possibilities, such as proactive notification of exceptions in the system.

Author: Mr ALEKSEEV, Aleksandr (National Research Tomsk Polytechnic University)

Co-authors: Dr KLIMENTOV, Alexei (Brookhaven National Lab); Mr BARREIRO MEGINO, Fernando (University of Texas at Arlington); PADOLSKI, Siarhei (BNL); MAENO, Tadashi (BNL); KORCHUGANOVA, Tatiana (National Research Tomsk Polytechnic University)

Presenter: Mr BARREIRO MEGINO, Fernando (University of Texas at Arlington)

Session Classification: Distributed Computing. GRID & Cloud computing

Track Classification: Distributed Computing. GRID & Cloud Computing