



Contribution ID: 53

Type: Poster

Analyzing data flows of WLCG jobs at batch job level

Tuesday, 2 September 2014 08:00 (1 hour)

With the introduction of federated data access to the workflows of WLCG, it is becoming increasingly important for data centers to understand specific data flows regarding storage element accesses, firewall configurations, as well as the scheduling of batch jobs themselves. As existing batch system monitoring and related system monitoring tools do not support measurements at batch job level, a new tool has been developed and put into operation at the GridKa Tier1 center for monitoring continuous data streams and characteristics of WLCG jobs and pilots. Long term measurements and data collection are in progress. These measurements already have been proven to be useful analyzing misbehaviors and various issues. Therefore we aim for an automated, realtime approach for anomaly detection. As a requirement, prototypes for standard workflows have to be examined. Based on measurements of several months, different features of HEP jobs are evaluated regarding their effectiveness for data mining approaches to identify these common workflows.

The presentation will introduce the actual measurement approach and statistics as well as the general concept and first results classifying different HEP job workflows derived from the measurements at GridKa.

Primary author: KUHN, Eileen (KIT - Karlsruhe Institute of Technology (DE))

Co-authors: PETZOLD, Andreas (KIT - Karlsruhe Institute of Technology (DE)); JUNG, Christopher (KIT - Karlsruhe Institute of Technology (DE)); GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); FISCHER, Max (KIT - Karlsruhe Institute of Technology (DE))

Presenter: JUNG, Christopher (KIT - Karlsruhe Institute of Technology (DE))

Session Classification: Poster session

Track Classification: Computing Technology for Physics Research