

Popularity of CMS datasets

Presenter: Justinas Rumševičius
Supervisors: Valentin Kuznetsov
Carl Vuosalo
Jean-Roch Vlimant

CERN
2017

Procedure outline

- Sequence of procedures:
 - Lookup activity logs
 - Extract specific attributes from logs
 - Export data in CSV format
 - Upload output to MONIT system
 - Visualize using Kibana/Grafana

Current progress

- Data from AAA, EOS, CMSSW, CRAB streams
- Most of the attributes are extracted
- Data is exported in CSV
- Each day is in separate directory
- Successfully uploaded data to MONIT
- Successfully plotted basic pie chart and histogram in Kibana

Next steps

- Code refinement
- Error handling
- Replace missing values with default values if needed
- Documentation, wiki page
- Setup cron job (automatic periodic script run)
- Discuss what plots will be needed
- Start making plots

My github repositories

- Github account:
 - <https://github.com/justinasr>
- Forked CMSSpark (“development” branch):
 - <https://github.com/justinasr/CMSSpark>
 - <https://github.com/justinasr/CMSSpark/tree/development>
- Weekly reports:
 - <https://github.com/justinasr/CERNReports>