OI in BigPanDA. Training data

Sergey Padolski (BNL)

Development plan

Step 1. Scope is limited by BigPanDA. Develop/test models, get a proof of the concept on small data sample, particular cases

Step 2. If Step 1 is affordable (powerful model developed) propagate OI event analysis for the whole BigPanDA scope

Step 3. Raise scope including information beyond BP: DDM, JIRA, ProdSys2, Monit,...

Data size estimation

- Number of hits to save: ~140k/week
- 1 json ~300kb
- Stream of ~42GB/week
- Could be locally stored as initial playground, then @ HDFS
- www.atlas-ml.org or Analytix - TBD
- Raw data file format: Avro
- Information retrieval management: single entry scheduler currently deploying for BP