ATLAS  BigPanDA monitoring
BelleII  DDM

S. Padolski (NPPS)
BigPanda Monitoring
Introduction

Reference talks:

- Alexei (https://indico.bnl.gov/event/6290/)
- Tadashi (https://indico.bnl.gov/event/6333/)

Instances: Atlas (https://bigpanda.cern.ch), Compass, EC2 (LSST, LQCD)
http://pandamonitor.org
Current content

- A window into the PanDA system
- > 100 different views
- From production dashboards to logs
- Covers scope in range $1 \ldots 10^{11}$ events
Current usage

17000 json requests a day

- 6.5k (+3%) user queries a day
- From 1 to 626 pages a day per user
- 1110 unique monthly active users
- 342 unique daily users

Is a one of primary tool ATLAS wide for shifters, experts and ADC in general
Architecture

MONIT

DB Cache

Rucio

DB

plots

Django applications

load balancer

firewall

logs

cache

JSON

elastic
Data-flow diagram of the BigPanDA monitor
Self monitoring
Operation Intelligence

- A lot of routine work is required for exposing and digging into problems
- We develop a mechanization such of work
  - BigPanDA attempts to provide information in the user friendly, quickly accessible way
  - OI attempts to process this information as a user would do
Operation Intelligence

A prototype of spotting failed jobs

- Finds time windows with concentration of failures
- Build a failure model within each time window
- Extract from models most influencing factors (pilot version, computing site, task,...)
- Builds clusters of failed jobs within defined space
- Provides links to the monitoring

Spot #0
Time window 2019-10-08 22:30:00 to 2019-10-08 23:00:00
https://biopanda.cern.ch/jobs?endpointrange=2019-10-09T22:30:00/2019-10-09T23:00:00&computingsite=AMNE_MORE&pilotw

Spot #1
Time window 2019-10-09 01:45:00 to 2019-10-09 02:15:00
https://biopanda.cern.ch/jobs?endpointrange=2019-10-09T01:45:00/2019-10-09T02:15:00&computingsite=UKT-SOUTHGRID-CNC;

Spot #2
Time window 2019-10-09 04:30:00 to 2019-10-09 05:00:00
https://biopanda.cern.ch/jobs?endpointrange=2019-10-09T04:30:00/2019-10-09T05:00:00&editaskid=19338038&destination;
https://biopanda.cern.ch/jobs?endpointrange=2019-10-09T04:30:00/2019-10-09T05:00:00&editaskid=19339914&destination;
Forecasting

- GRID computing is a big queue
  - When is my order to get served?
  - When I do receive needed result?
  - How much can I order in principle to get result in reasonable time?

- Campaign is the ProdSys2 object which unites large computing activity ($10^{10}$ events) and involves different parties (physicists, managers, shifters)

- A “Hot” model was developed

- Currently it is getting to production

---

Remaining events: 26,006,909

Estimation of completion time for remaining events: 0.89 d

Current production rate: 29,068,813 ev/day
Usability R&D

Aims:

- Raise UP level of BigPanDA monitoring usability
- Build the whole system at one presentation technology stack
- Make development easy, from well defined bricks
- Technology assessment is finishing (Tatiana is leading this effort)
Bellell DDM
Introduction

Reference talks: Paul [https://indico.bnl.gov/event/6124/]

Pictures taken from Ueda I talks on B2 meetings
Distributed computing
Distributed Data Management

Initial picture taken from Paul introductory talk
Distributed Data Management Scale

Daily transfers
BNL time Developments

- Data mover scripts (PNNL->BNL)
- Functional tests subsystem
- Tider integration with DIRAC Resource system
- Advanced transfers load balancer. Developed to prevent possible data transfers stucks. Implements accurate submissions to the FTS served respecting:
  - SE current performance
  - Links quality
  - Recent experience of completed operations
  - Current queue state
  - Operations importance
- Deletion at scales (bulk operations, parallel processing)
- SQL queries analysis and optimization
- Simple DDM monitoring
- Operational support

Operational experience: 2 DDM (and whole DIRAC) down due to server disk spaces exhausting and one due to RMS system failure, one failed deployment. Successful raw files distribution during data taking.