Jobs Buster

Sergey Padolski (BNL)

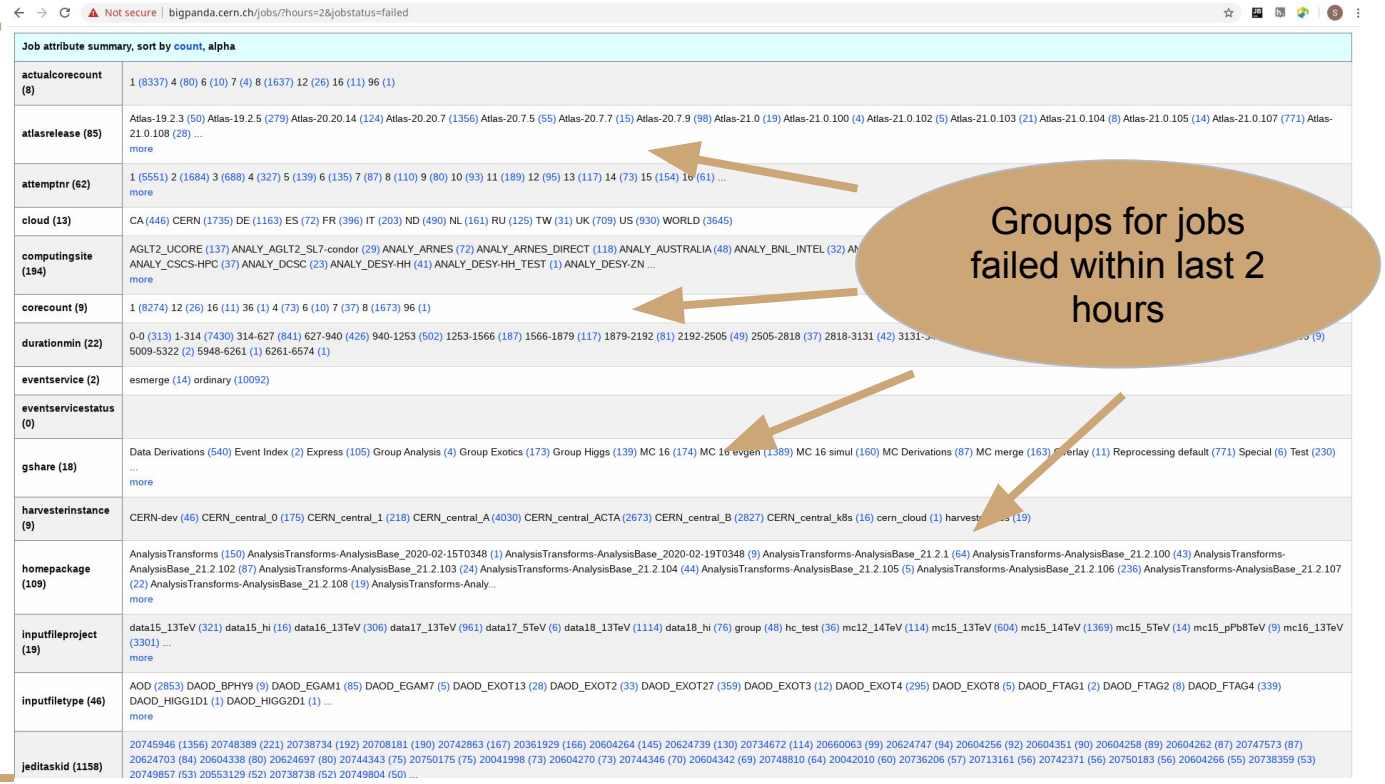
On behalf of the operational-intelligence@cern.ch community

Motivation

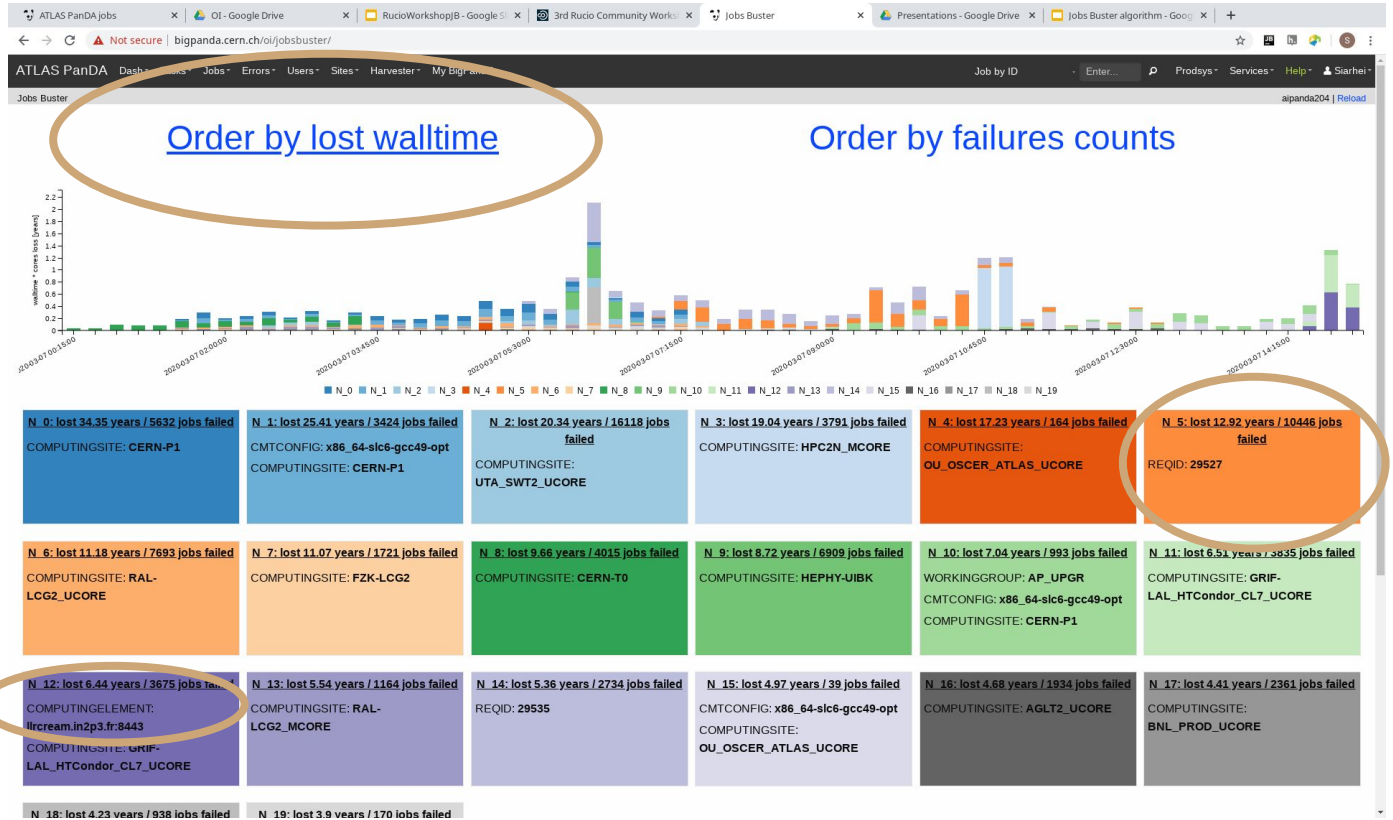
- Impressive numbers and facts about scale and FTEs of the current operational efforts are in the Federica Legger talk
- How can we spot problems automatically, using if-else, ML, OI,...?
- Automatically is the point

BigPanda Monitor (in Atlas)

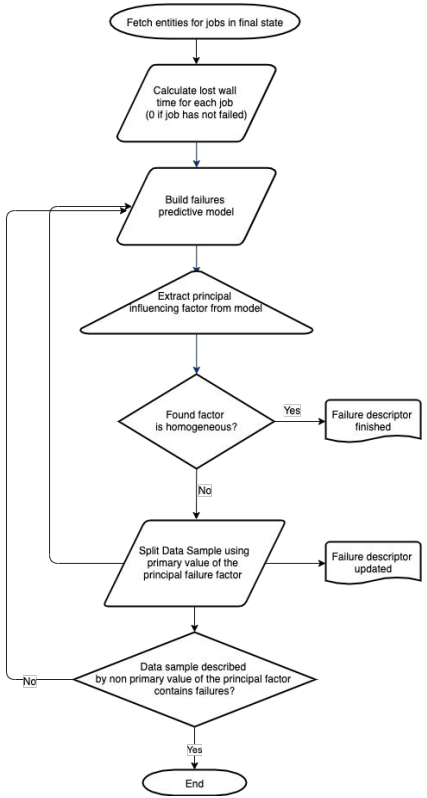
31 features



Jobs buster


ATLAS PanDA jobs
OI - Google Drive
RucioWorkshopJB - Google S
3rd Rucio Community Works
Jobs Buster
Presentations - Google Drive
Jobs Buster algorithm - Goog
Not secure
bigpanda.cern.ch/oi/jobsbuster/
ATLAS PanDA
Dash
Jobs
Errors
Users
Sites
Harvester
Job by ID
Enter...
Prodsys
Services
Help
Siarhei
Jobs Buster
aipanda204 | Reload
Order by lost walltime
Order by failures counts
N_0 N_1 N_2 N_3 N_4 N_5 N_6 N_7 N_8 N_9 N_10 N_11 N_12 N_13 N_14 N_15 N_16 N_17 N_18 N_19
N_0: lost 34.35 years / 5632 jobs failed
COMPUTINGSITE: CERN-P1
N_1: lost 25.41 years / 3424 jobs failed
CMTCONFIG: x86_64-slc6-gcc49-opt
COMPUTINGSITE: CERN-P1
N_2: lost 20.34 years / 16118 jobs failed
COMPUTINGSITE: UTA_SWT2_UCORE
N_3: lost 19.04 years / 3791 jobs failed
COMPUTINGSITE: HPC2N_MCORE
N_4: lost 17.23 years / 164 jobs failed
COMPUTINGSITE: OU_OSCER_ATLAS_UCORE
N_5: lost 12.92 years / 10446 jobs failed
REQID: 29527
N_6: lost 11.18 years / 7693 jobs failed
COMPUTINGSITE: RAL-LCG2_UCORE
N_7: lost 11.07 years / 1721 jobs failed
COMPUTINGSITE: FZK-LCG2
N_8: lost 9.66 years / 4015 jobs failed
COMPUTINGSITE: CERN-T0
N_9: lost 8.72 years / 6909 jobs failed
COMPUTINGSITE: HEPHY-UIBK
N_10: lost 7.04 years / 993 jobs failed
WORKINGGROUP: AP_UPGR
CMTCONFIG: x86_64-slc6-gcc49-opt
COMPUTINGSITE: CERN-P1
N_11: lost 6.51 years / 3835 jobs failed
COMPUTINGSITE: GRIF-LAL_HTCondor_CL7_UCORE
N_12: lost 6.44 years / 3675 jobs failed
COMPUTINGELEMENT: llrcream.in2p3.fr:8443
COMPUTINGSITE: GRIF-LAL_HTCondor_CL7_UCORE
N_13: lost 5.54 years / 1164 jobs failed
COMPUTINGSITE: RAL-LCG2_MCORE
N_14: lost 5.36 years / 2734 jobs failed
REQID: 29535
N_15: lost 4.97 years / 39 jobs failed
CMTCONFIG: x86_64-slc6-gcc49-opt
COMPUTINGSITE: OU_OSCER_ATLAS_UCORE
N_16: lost 4.68 years / 1934 jobs failed
COMPUTINGSITE: AGLT2_UCORE
N_17: lost 4.41 years / 2361 jobs failed
COMPUTINGSITE: BNL_PROD_UCORE
N_18: lost 4.23 years / 938 jobs failed
N_19: lost 3.9 years / 170 jobs failed

How it works



Step 1: Fetched 20424 failed and 145546 finished jobs

Build failure (wall time loss) predictive model using both successful and failed statistics

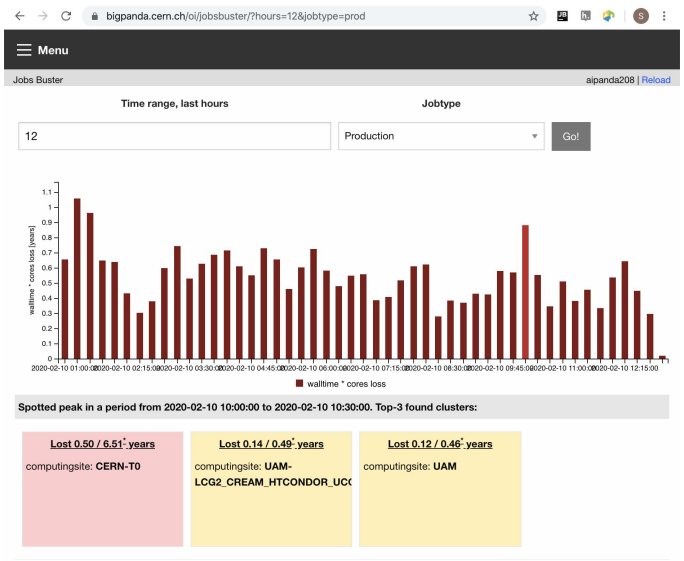
Principle factor at this step is Pilot Version.
One value (“Unknown”) is responsible for 150 failures. No successful jobs with this value

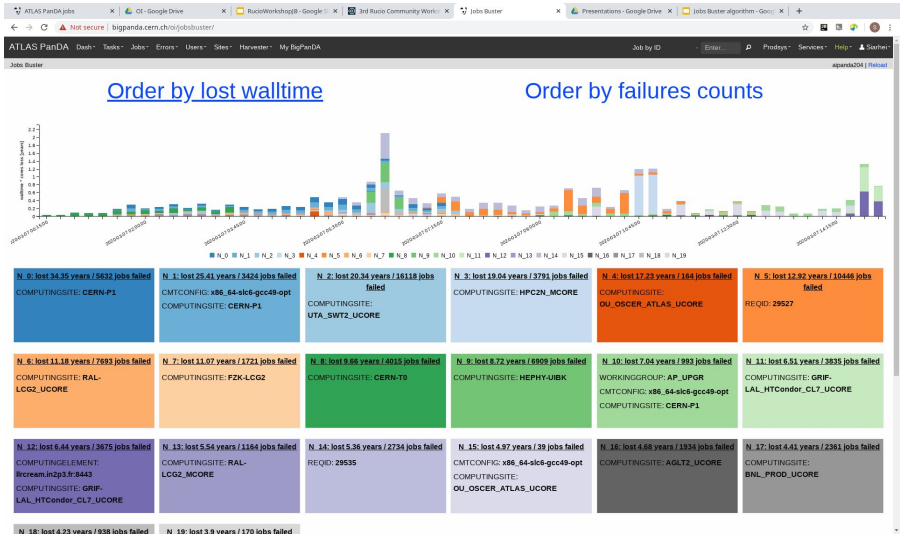
Repeat procedure

Failure spot #1 has found. We can clearly select jobs with homogeneous failure reason

Select jobs with Pilot Version ≠ “Unknown”

Current Status





- Problems identification within tiny windows around peaks
- No persistency / problems merge
- Stand alone prototype

- Problems identification at whole timeline
- Persistency
- A use case pushing the OpInt framework development

Plans

- Assessment of the developed prototype. Use operation shifts reports as a source of information for findings comparison
- Tune up algorithms
- Add rules formalized by humans: task is a part of request, computing element is a part of site, etc.
- Enrich feature selection with log based semantic analysis

https://bigpanda.cern.ch/oi/jobsbuster/