



# Jobs Buster

Sergey Padolski (BNL)

On behalf of the [operational-intelligence@cern.ch](mailto: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

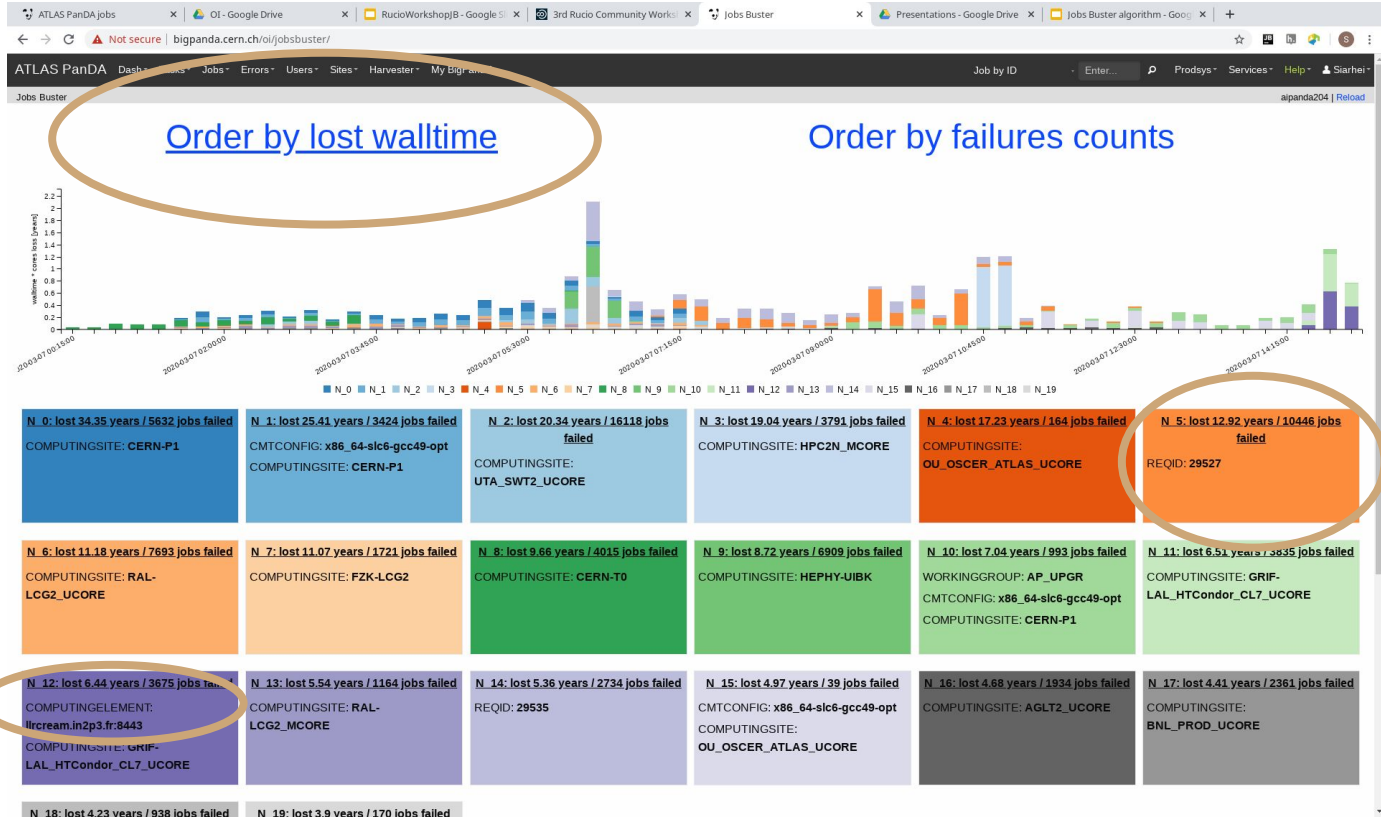
# BigPanda Monitor (in Atlas)

Job attribute summary, sort by count, alpha	
actualcorecount (8)	1 (8337) 4 (80) 6 (10) 7 (4) 8 (1637) 12 (26) 16 (11) 96 (1)
atlasrelease (85)	Atlas-19.2.3 (50) Atlas-19.2.5 (279) Atlas-20.20.14 (124) Atlas-20.20.7 (1356) Atlas-20.7.5 (55) Atlas-20.7.7 (15) Atlas-20.7.9 (98) Atlas-21.0 (19) Atlas-21.0.100 (4) Atlas-21.0.102 (5) Atlas-21.0.103 (21) Atlas-21.0.104 (8) Atlas-21.0.105 (14) Atlas-21.0.107 (771) Atlas-21.0.108 (28) ... <a href="#">more</a>
attemptnr (62)	1 (5551) 2 (1684) 3 (688) 4 (327) 5 (139) 6 (135) 7 (87) 8 (110) 9 (80) 10 (93) 11 (189) 12 (95) 13 (117) 14 (73) 15 (154) 16 (61) ... <a href="#">more</a>
cloud (13)	CA (446) CERN (1735) DE (1163) ES (72) FR (396) IT (203) ND (490) NL (161) RU (125) TW (31) UK (709) US (930) WORLD (3645)
computingsite (194)	AGLT2_UCORE (137) ANALY_AGLT2_SL7-condor (29) ANALY_ARNES (72) ANALY_ARNES_DIRECT (118) ANALY_AUSTRALIA (48) ANALY_BNL_INTEL (32) ANALY_CSOS-HPC (37) ANALY_DCSCS (23) ANALY_DESY-HH (41) ANALY_DESY-HH_TEST (1) ANALY_DESY-ZN ... <a href="#">more</a>
corecount (9)	1 (8274) 12 (26) 16 (11) 36 (1) 4 (73) 6 (10) 7 (37) 8 (1673) 96 (1)
durationmin (22)	0-0 (313) 1-314 (7430) 314-627 (841) 627-940 (426) 940-1253 (502) 1253-1566 (187) 1566-1879 (117) 1879-2192 (81) 2192-2505 (49) 2505-2818 (37) 2818-3131 (42) 3131-5009-5322 (2) 5948-6261 (1) 6261-6574 (1)
eventservice (2)	esmerge (14) ordinary (10092)
eventservicestatus (0)	
gshare (18)	Data Derivations (540) Event Index (2) Express (105) Group Analysis (4) Group Exotics (173) Group Higgs (139) MC 16 (174) MC 16 evgen (1389) MC 16 simul (160) MC Derivations (87) MC merge (163) Overlay (11) Reprocessing default (771) Special (6) Test (230) ... <a href="#">more</a>
harvesterinstance (9)	CERN-dev (46) CERN_central_0 (175) CERN_central_1 (218) CERN_central_A (4030) CERN_central_ACTA (2673) CERN_central_B (2827) CERN_central_k8s (16) cern_cloud (1) harvest ... (19)
homepackage (109)	AnalysisTransforms (150) AnalysisTransforms-AnalysisBase_2020-02-15T0348 (1) AnalysisTransforms-AnalysisBase_2020-02-19T0348 (9) AnalysisTransforms-AnalysisBase_21.2.1 (54) AnalysisTransforms-AnalysisBase_21.2.100 (43) AnalysisTransforms-AnalysisBase_21.2.102 (87) AnalysisTransforms-AnalysisBase_21.2.103 (24) AnalysisTransforms-AnalysisBase_21.2.104 (44) AnalysisTransforms-AnalysisBase_21.2.105 (5) AnalysisTransforms-AnalysisBase_21.2.106 (236) AnalysisTransforms-AnalysisBase_21.2.107 (22) AnalysisTransforms-AnalysisBase_21.2.108 (19) AnalysisTransforms-Analy ... <a href="#">more</a>
inputfileproject (19)	data15_13TeV (321) data15_hi (16) data16_13TeV (306) data17_13TeV (961) data17_5TeV (6) data18_13TeV (1114) data18_hi (76) group (48) hc_test (36) mc12_14TeV (114) mc15_13TeV (604) mc15_14TeV (1369) mc15_5TeV (14) mc15_pPb8TeV (9) mc16_13TeV (3301) ... <a href="#">more</a>
inputfiletype (46)	AOD (2853) DAOD_BPHY9 (9) DAOD_EGAM1 (85) DAOD_EGAM7 (5) DAOD_EXOT13 (28) DAOD_EXOT2 (33) DAOD_EXOT27 (359) DAOD_EXOT3 (12) DAOD_EXOT4 (295) DAOD_EXOT8 (5) DAOD_FTAG1 (2) DAOD_FTAG2 (8) DAOD_FTAG4 (339) DAOD_HIGG1D1 (1) DAOD_HIGG2D1 (1) ... <a href="#">more</a>
jedidaskid (1158)	20745846 (1356) 20748389 (221) 20738734 (192) 20708181 (190) 20742863 (167) 20361929 (166) 20604264 (145) 20624739 (130) 20734672 (114) 20660063 (99) 20624747 (94) 20604256 (92) 20604351 (90) 20604258 (89) 20604262 (87) 20747573 (87) 20624703 (84) 20604338 (80) 20624697 (80) 20744343 (75) 207501175 (75) 20041998 (73) 20604270 (73) 20744346 (70) 20604342 (69) 20748810 (64) 20042010 (60) 20736206 (57) 20713161 (56) 20742371 (56) 207501183 (56) 20604266 (55) 20738359 (53) 20749857 (53) 20553129 (52) 20738738 (52) 20749804 (50) ...

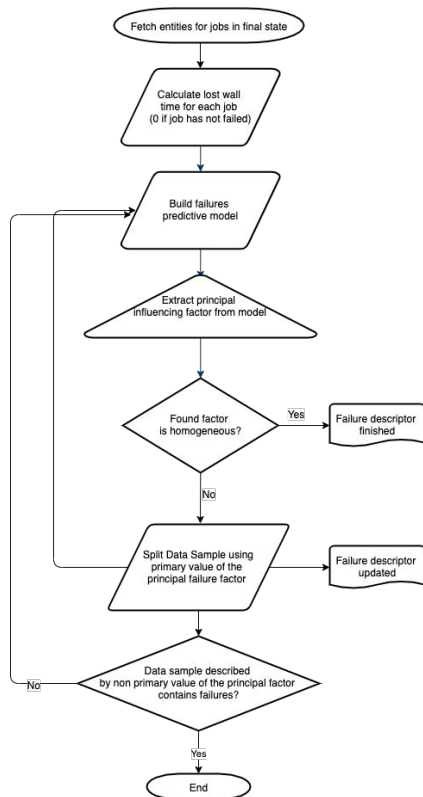
31 features

Groups for jobs failed within last 2 hours

# Jobs buster



# 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

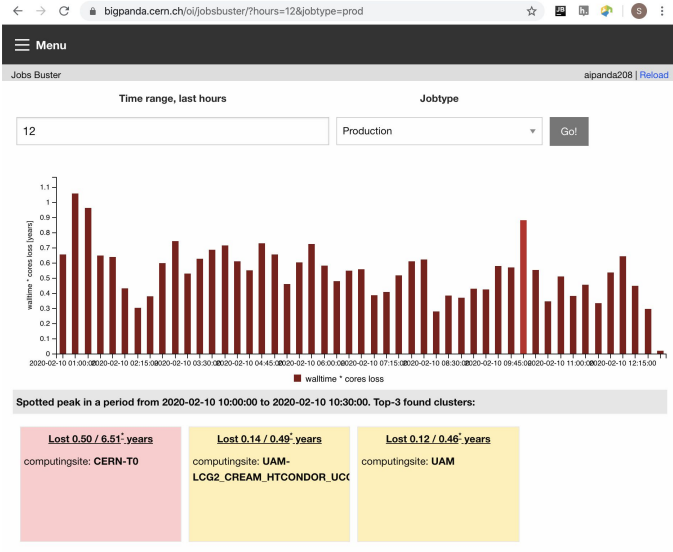
Principle factor at this step is Pilot Version.  
**One value ("Unknown") is responsible for 150 failures.** No successful jobs with this value

Repeat procedure

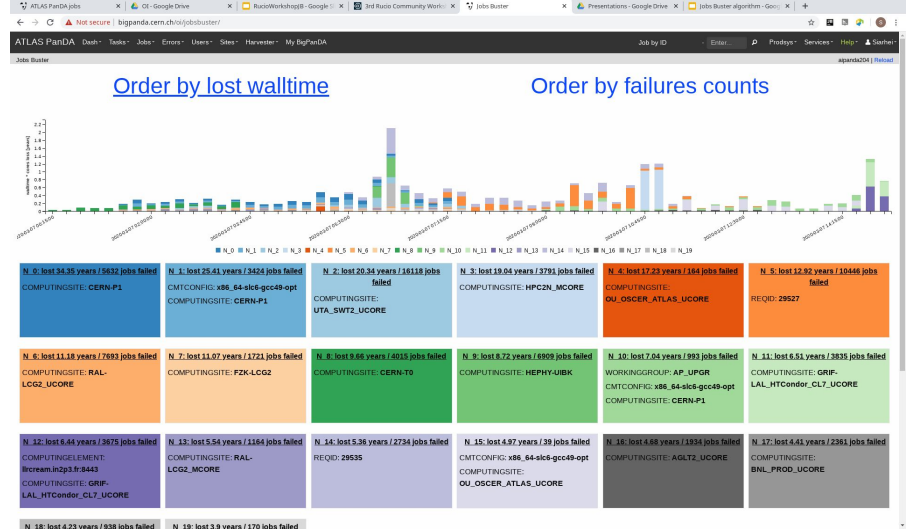
Failure spot #1 has found. We can clearly select jobs with homogeneous failure reason

Select jobs with Pilot Version  $\neq$  "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/jobbuster/>