Managing the ATLAS Grid through Harvester

Fernando Barreiro Megino, Aleksandr Alekseev, Frank Berghaus, David Cameron, Andrej Filipcic, Ivan Glushkov, Fa-Hui Lin, Tadashi Maeno, Nicolo Magini
On behalf of the ATLAS collaboration

CHEP 2019, Adelaide, Australia

Motivation

https://github.com/HSF/harvester

- Worker submitter capable of interfacing different resources (grid, cloud, HPC)
- Improve grid stability and usage efficiency
- Tighter integration with the Workload Management System (PanDA)
- Control ratio of pilot types (e.g. single vs multi core) according to ATLAS priorities/Global Shares
- Better monitoring
- This presentation focuses on the Grid usage, please see our other presentations
  - Harvester and Kubernetes
  - Harvester and HPC

Using Kubernetes as an ATLAS computing site

5 Nov 2019, 16:30
15m
Riverbank R7 (Adelaide Convention Centre)

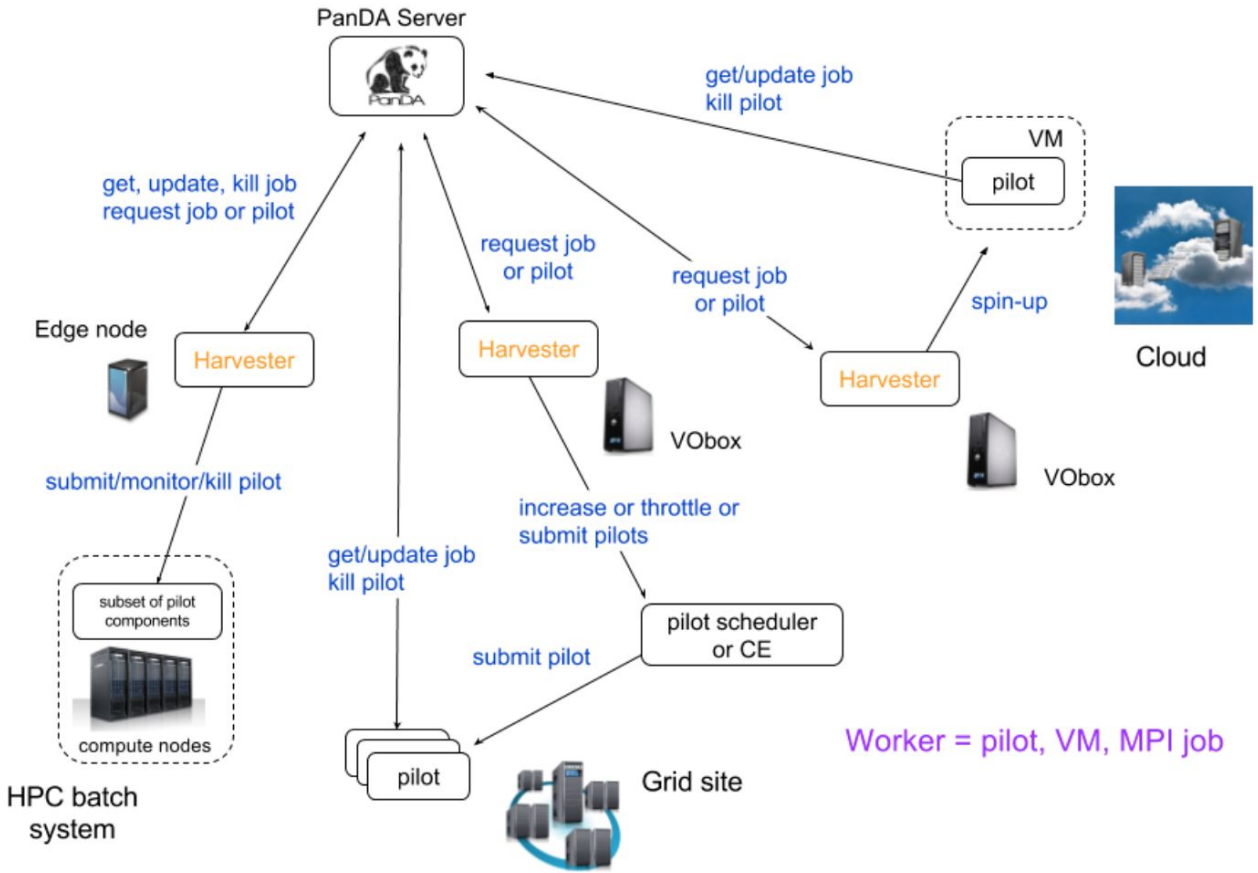
Oral | Track 7 – Facilities, Cl... | Track 7 – Facilities, Clou...

Large scale fine grain simulation workflows ("Jumbo Jobs") on HPC's by the ATLAS experiment

7 Nov 2019, 11:30
15m
Riverbank R1 (Adelaide Convention Centre)

Oral | Track 9 – Exascale Sci... | Track 9 – Exascale Scien...

Harvester global picture



Harvester: design decisions

- Flexibility to run in lightweight or high performance mode
  - Support for SQLite or MySQL/MariaDB
  - Possibility to run in multi-process mode through uWSGI framework
- Fast integration of new resources (CEs, batch, cloud API...)
  - Common core
  - Resource specific libraries (few hundred lines of code each). Typically:
    - Submitter: how to submit workers to the resource
    - Monitor: how to poll the status of the workers
    - Sweeper: how to clean up a worker
  - Plugins exist for HTCondor, ARC CE, GCE, K8s, SAGA, PBS, Slurm, Cobalt
  - Also different plugins for data movement needed in HPC environment
  - Possibility of running through RPC

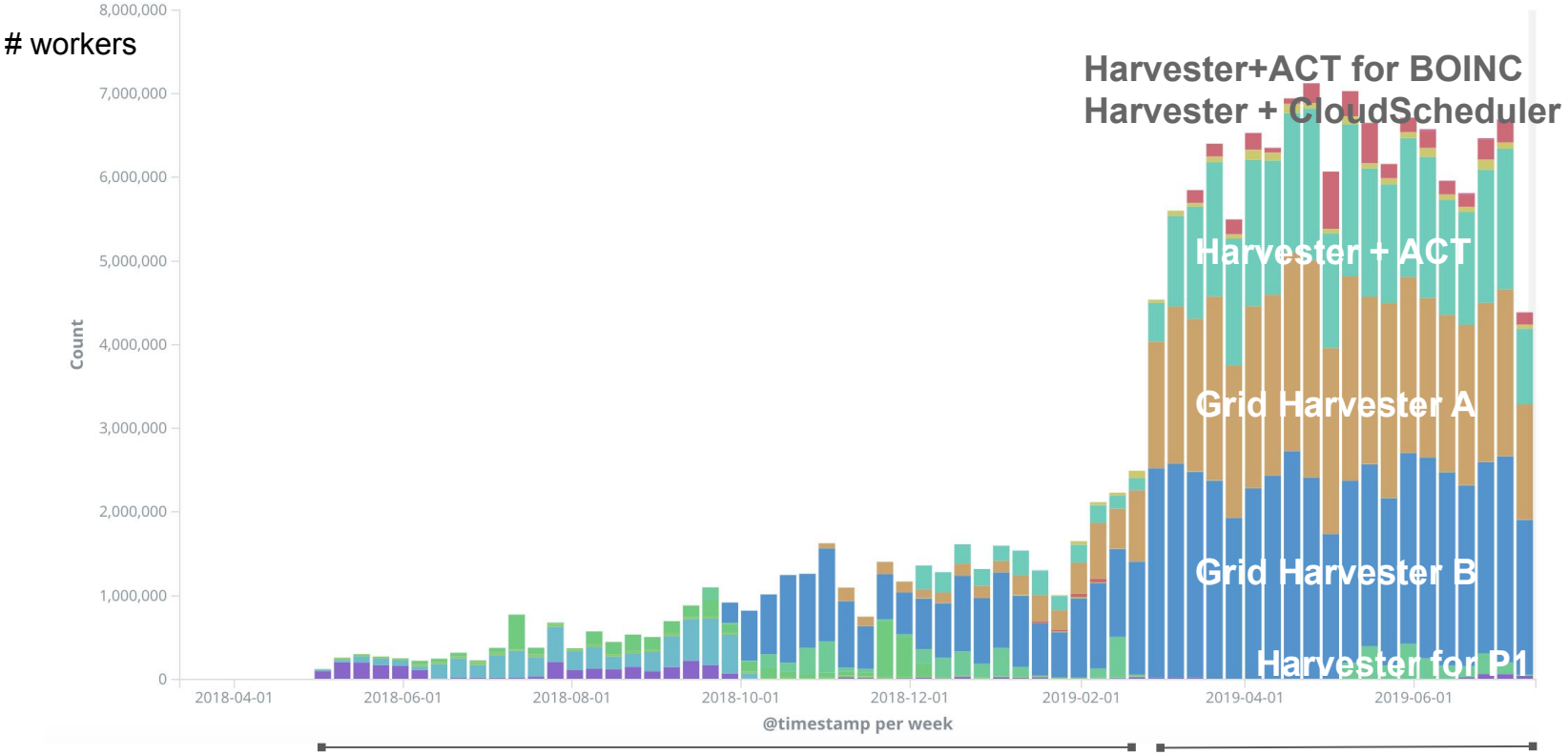
Harvester workflows

- **Push**: the queued worker is already tied to a job. Early binding
- **Pull**: the worker will retrieve a job once it starts running. Late binding. Typically several independent queues per core&memory combination at each site, leading to internal competition. Used to be most popular workflow.
- **Pull UPS**(Unified Pilot Streaming): Intended for **unified** queues, i.e. only one queue per site. The WMS[1] tells Harvester the ratio of workers out of a set of core&memory categories:
  - The worker ratio is calculated according to ATLAS shares/priorities
  - The batch system respects corecount and memory specified in the job description file
  - Wherever possible, sites were migrated to this workflow. Currently most popular workflow.

NEW

[1] *WMS: Workload Management System, in our case PanDA*

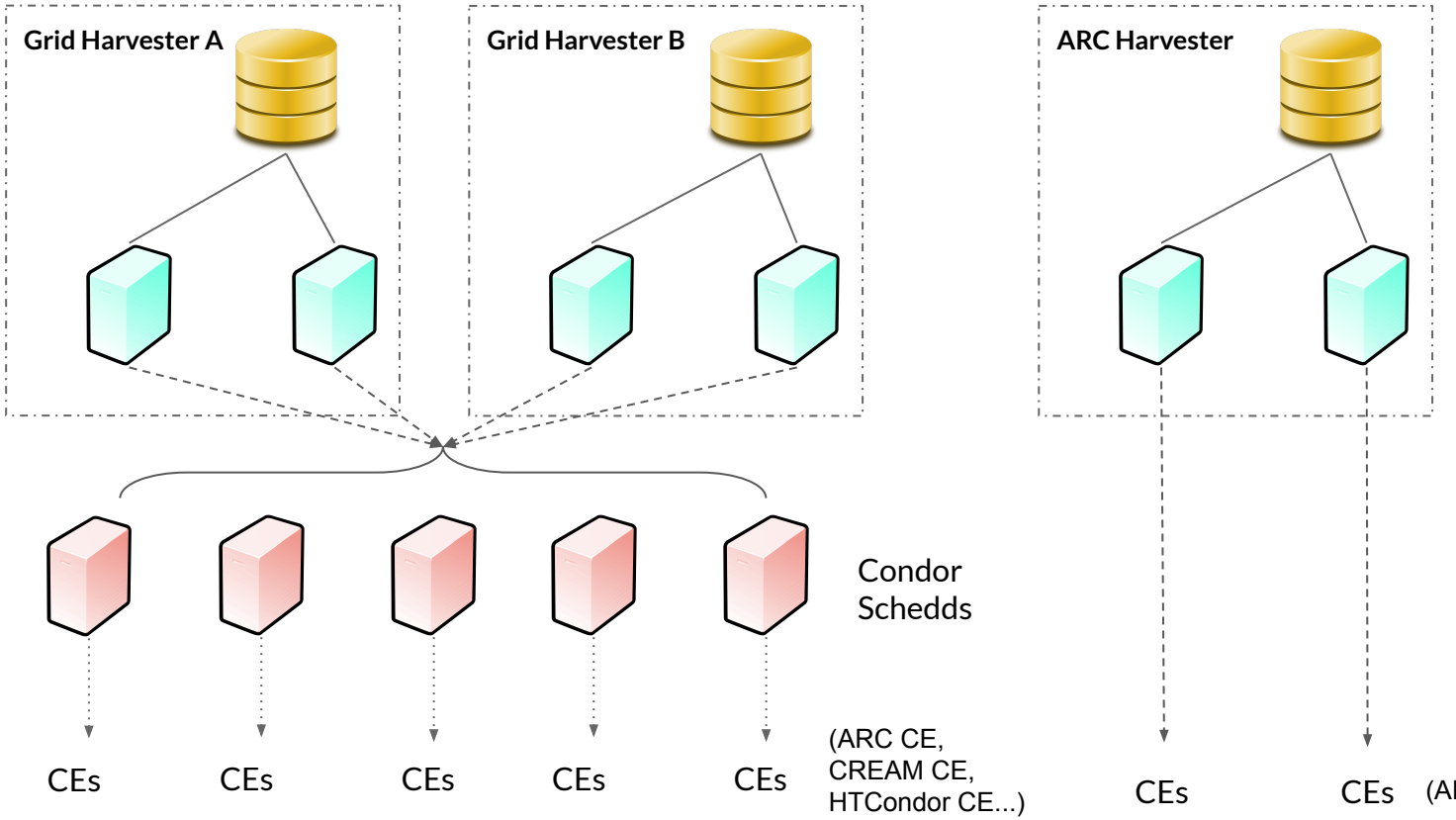
Harvester commissioning evolution



Mostly **production** migration: typically long, multi core workers

Mostly **analysis** migration: typically short, single core workers

Central Harvester infrastructure



Harvester worker monitoring: fast and flexible

Based on CERN IT provided Elastic Search, Kibana & Grafana instances

Production > Harvester

Last 24 hours

Binning 1h | Cloud All | Site All | Computingsite All | Harvester All | Status All | Resourcetype All | Computingelement All | Filter +

Filters

Harvester workers view

Search:

computingsite	status	submittime	stderr	stdout	batchlog	jdl	starttime	endti
ANALY_BEIJING	submitted	2019-07-12 10:08:40	https://aipanda183.cern.ch/condor_logs_2/19-07-12_08/grid.2185017.1.err	https://aipanda183.cern.ch/condor_logs_2/19-07-12_08/grid.2185017.1.out	https://aipanda183.cern.ch/condor_logs_2/19-07-12_08/grid.2185017.1.log	https://aipanda183.cern.ch/condor_logs_2/19-07-12_08/grid.2185017.1.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.7.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.7.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.7.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.7.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.15.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.15.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.15.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.15.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.0.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.0.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.0.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.0.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.1.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.1.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.1.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.1.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.10.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.10.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.10.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.10.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.11.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.11.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.11.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.11.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.16.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.16.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.16.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.16.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.4.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.4.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.4.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.4.jdl		
ANALY_BNL_SHORT	submitted	2019-07-12 10:08:49	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.12.err	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.12.out	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.12.log	https://aipanda156.cern.ch/condor_logs_2/19-07-12_08/grid.59328.12.jdl		

Showing 1 to 10 of 500 entries

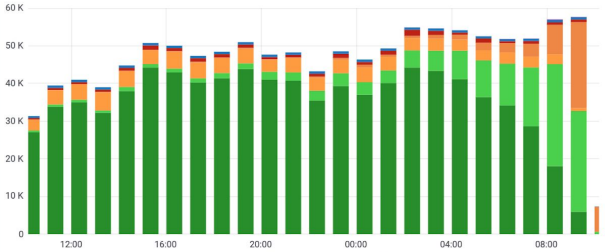
Previous 1 2 3 4 5 ... 50 Next

Worker information and links to logfiles/jdl

Workers by status

status.keyword	Count
finished	864.52 K
running	134.68 K
cancelled	83.10 K
submitted	51.75 K
failed	19.19 K
missed	15.18 K
idle	524.00

Workers by status



	min	max	avg	current	total
finished	1	44.25 K	34.57 K	1	864.22 K
running	456	27.09 K	5.39 K	596	134.69 K
cancelled	1	4.83 K	3.32 K	1	83.08 K

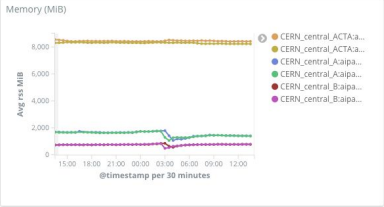
Tables and plots

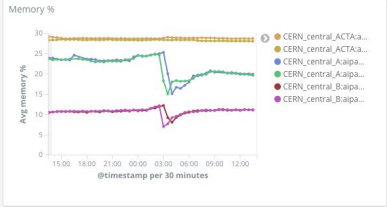
Service&infrastructure monitoring

Service monitoring for Harvester and Condor instances: memory, disk, CPU, worker submission rates. Alerting system to report issues based on most relevant metrics

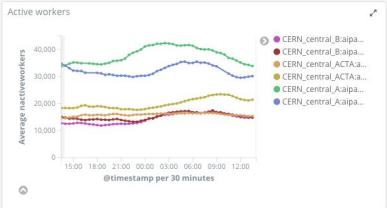
Metrics overview

harvester_id.keyword: Descending	harvester_host.keyword: Descending	Avg CPU%	Max CPU%	Avg rss(MiB)	Max rss(MiB)	Max data %
CERN_central_A	aipanda172.cern.ch	18.624	23.75	1,556.343	1,829.707	55
CERN_central_A	aipanda171.cern.ch	17.48	22	1,620.429	1,843.148	58
CERN_central_B	aipanda174.cern.ch	9.249	11.875	773.994	901.984	32
CERN_central_B	aipanda173.cern.ch	8.684	12.375	767.763	892.98	35
CERN_central_ACTA	aipanda403.cern.ch	7.337	8.369	8,303.22	8,456.766	-
CERN_central_ACTA	aipanda404.cern.ch	6.7	7.725	8,446.915	8,596.875	-





...



Grid Infrastructure monitoring: find broken or inactive sites/CEs. Future operational automation

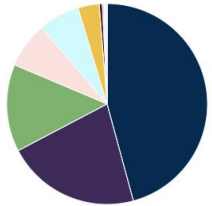
Computingsite

Computingsites error rate

computingsite	status	totalworkers	goodworkers	badworkers	error_rate
[redacted]	online	330	0	330	100.00
[redacted]	online	46	0	46	100.00
[redacted]	online	170	0	170	100.00
[redacted]	online	1	0	1	100.00
[redacted]	online	55	0	55	100.00
[redacted]	online	6507	120	6387	98.16
[redacted]	online	988	23	965	97.67
[redacted]	online	1067	34	1033	96.81
[redacted]	online	137	8	129	94.16
[redacted]	online	440	34	406	92.27
[redacted]	online	179	18	161	89.94
[redacted]	online	13047	1667	11380	87.22
[redacted]	online	1672	230	1442	86.24
[redacted]	online	934	129	805	86.19
[redacted]	online	1508	211	1297	86.01
[redacted]	online	2343	449	1894	80.84
[redacted]	online	1347	404	943	70.01
[redacted]	online	1218	383	835	68.56
[redacted]	online	1242	509	733	59.02
[redacted]	online	48	21	27	56.25
[redacted]	online	1399	622	777	55.54

1 2 3

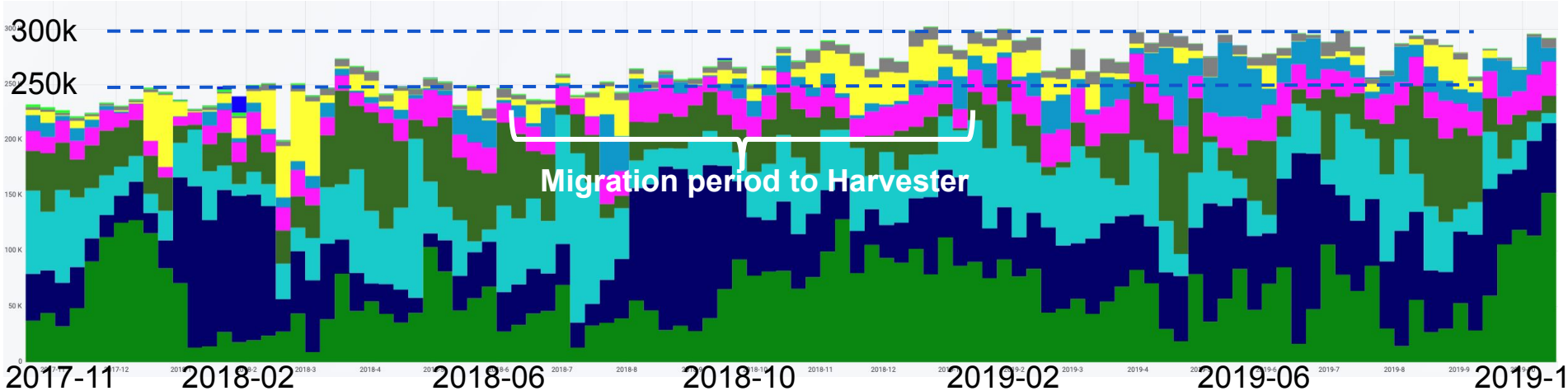
Errors count by errordesc



	current
Condor HoldReason: HTCondor-CE held job due to no matching routes, route job limit, or route failure threshold; see 'HTCondor-...	28.8 K
Condor HoldReason: Job not found ; Worker canceled by harvester due to held too long or not found	13.55 K
Condor HoldReason: CE job in status 1 put on hold by SYSTEM_PERIODIC_HOLD due to non-existent route in JOB_ROUTER_ENT...	8.85 K
Condor HoldReason: None ; Condor RemoveReason: The system macro SYSTEM_PERIODIC_REMOVE expression 'false || ((JobR...	4.65 K
Condor HoldReason: Network error talking to schedd, probably an authorization failure ; Worker canceled by harvester due to hel...	4.13 K

Results and conclusions

Running job slots last 2 years. Grid only (HPC & cloud are excluded)



Group Production	Data Processing
MC Reconstruction	Group Analysis
MC Simulation Full	Testing
MC Event Generation	MC Simulation
User Analysis	Others
MC Simulation Fast	Event Index

Conclusions

- 2 years from first discussions until project fully rolled out
- Better usage of the grid
  - More efficient and stable grid usage
  - Respecting ATLAS global shares, i.e. running the jobs that matter to the experiment
- Smarter worker submission, but also dependence on more components and their intercommunication
- Details to iron out
  - Grand unification of analysis and production queues
  - Handling single core spikes and higher submission rates
  - Automation/better handling of issues on the Grid