

Monitoring HTCondor

A Marriage of Open Source and Custom

William Deck



Goals

- Help those looking to augment/roll out their own cluster Monitoring
- Possibly a different perspective on monitoring
- Get some of this work out in the open



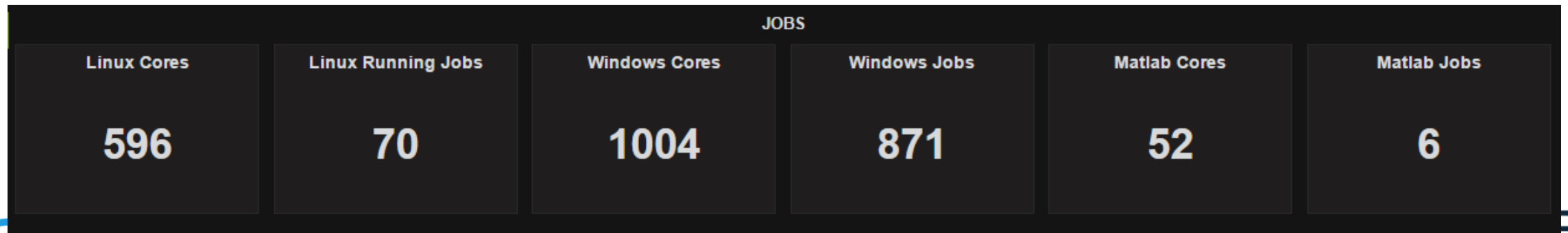
Who am I?

- Originally a mainframe developer
- Moved to SIG January of 2015
- First task: Monitor the “cluster”



Our Cluster

- Mixed Cluster (Several)
 - ~1700 cores
 - 1100 Windows
 - 600 Linux (SLES11/SLES12)
- Storage
 - Lustre: 2.8 PB
 - GPFS: 8.4 PB



Our Workload

- ~100K jobs/week
- Mixed OS Workloads (Majority Windows)
- Daily Workflows (1 – 10K)
- One Off Workflows (+100K)



What is the Problem?

- Many moving parts
- Multiple groups using condor differently
 - Single submit files
 - DAGs
 - Someone found out about **condor_run**
- Monitor 1 – 100k jobs
- Hard for users to self-support
- Correlate job failures to infrastructure problems



Monitor Evolution

- HTCondor's Job Monitor
- CycleComputing Solution
- Custom Scripts parsing condor_* commands
- Python bindings with Elastic, Grafana, and Conmon



First Solution – Tales of Condor

- Modelled after CycleComputing monitoring
- Parsing condor_status, condor_history output
- Put information into csv files
- Simple webpage displayed
 - Total jobs vs running jobs
 - Total jobs/User vs running jobs/User



First Solution: Revolutionary command **grep**

- For everything else we used grep
- Painful and tedious
- Good news!
 - Really good at grep



Enter: Monmon

- Custom webpage designed around our workflow creation scripts
- Parses nodestatus/HTCondor log files
- Specific for one group's workflows
- Useful for the user to drill down and share with others

Monmon

[Cluster Overview](#) [Condor What and Who](#) [Tales of Condor](#) [Conmon](#)

Show entries

[30 Days](#) [90 Days](#) [All Days](#)
Search:

Timestamp	Status	User	Machine	Grid	Workflow
2017-04-07 11:51:50		deck	lx00011818	view	test 4
2017-04-07 11:51:17		deck	lx00011818	view	test 3
2017-04-07 11:51:10		deck	lx00011818	view	test 2
2017-04-07 11:51:01		deck	lx00011818	view	test 1
2017-03-31 14:56:18	failed	deck	lx00011816	view	lfs migrate run small

Enter Elastic, Grafana, and Common

- Elastic (ELK) (<https://www.elastic.co/>)
 - Elasticsearch - Search/analytics engine
 - Logstash - Ingest of data
 - Kibana - Frontend
- Grafana - (<https://grafana.com/>)
 - extension of Kibana (3.0)
 - Multiple backends (graphite, ES, influxdb ...)



elastic

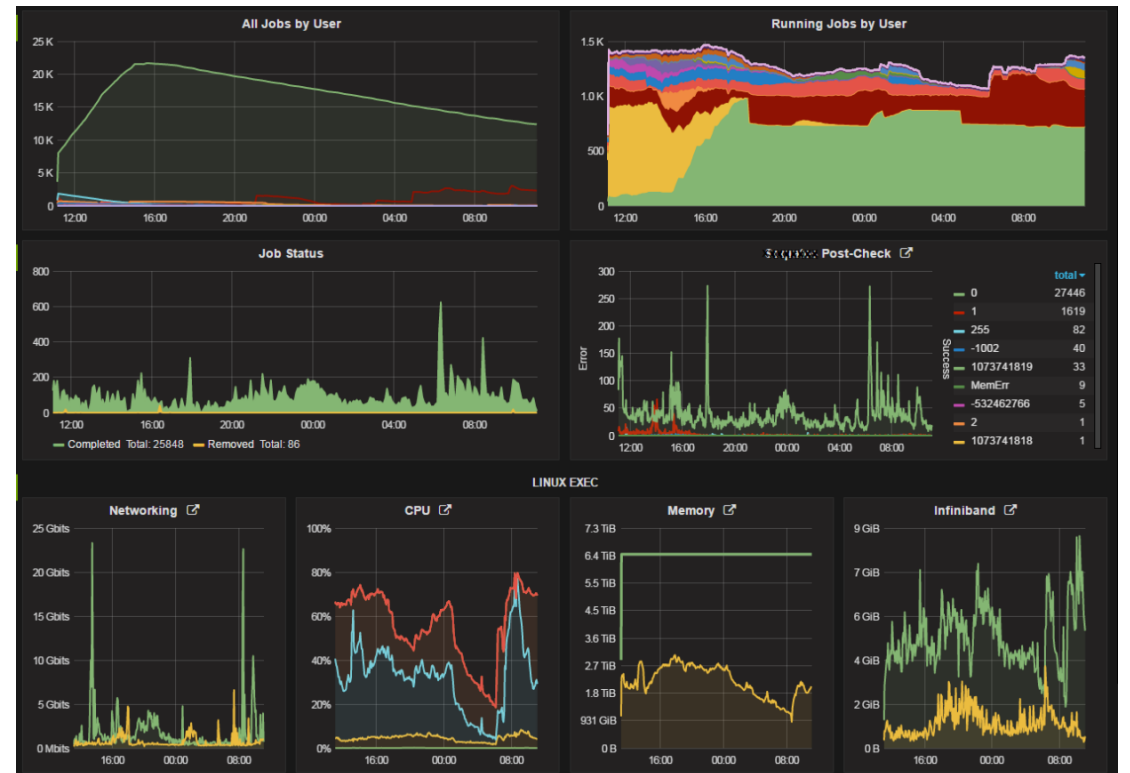


Grafana Labs



Enter Elastic, Grafana, and Common

- Poll periodically using python bindings
- Insert into ES
 - Augmented Job/Host ClassAd
 - Custom subset of ClassAd
- Use Grafana frontend
- Tales of Condor 2.0
- More Complex Dashboards
- Extended Monmon -> Common



ES Job Ad Example

- Searchable Job ClassAd
- Common Uses
 - condor_history
 - Used Resources

```
"_source": {
  "RequestMemory": 4096,
  "Requirements": "Target.Machine == \"XXXXXXXXXXXXXXXXXXXX\" && TARGET.OpSys == \"LINUX\"",
  "TotalSuspensions": 0,
  "LastJobStatus": 2,
  "BufferBlockSize": 32768,
  "OrigMaxHosts": 1,
  "JobStartDate": "2017-04-07T16:10:30",
  "WantRemoteSyscalls": false,
  "bt_action": "update",
  "ExitStatus": 0,
  "SubmitEventNotes": "DAG Node: lfs_migrate_run_meng_1.bXXXXXXXX00.Data",
  "QueueTime_Days": 0,
  "Args": "",
  "JobFinishedHookDone": 1491581935,
  "RunTime_TotalSeconds": 505,
  "JobCurrentStartDate": "2017-04-07T16:10:30",
  "RunTime_Seconds": 3,
  "JobType": "Backtest",
  "CompletionDate": "2017-04-07T16:18:55",
  "JobLeaseDuration": 2400,
  "Err": "/XXXXXXXXXXXXXXXXXXXX/XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX/XXXXXXXXXXXXXXXXXXXXX/log/Data.stderr",
  "EncryptExecuteDirectory": false,
  "MyType": "Job",
  "JobUniverse": 5,
  "RequestCpus": 1,
  "RunTime_Days": 0,
  "DAGManNodesLog": "/XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX/XXXXXXXXXXXXXXXXXXXXX/lfs_migrate_run_meng_1.dag.nodes.log",
  "StreamErr": true,
  "Rank": 0,
  "LastRejMatchTime": 1491581409,
  "AcctGroup": "group_sotbt",
  "WantRemoteIO": true,
  "LocalSysCpu": 0,
  "CondorVersion": "$CondorVersion: 8.4.3 Feb 11 2016 BuildID: UW_development $",
  "TransferIn": false,
  "MachineAttrCpus": 1,
  "CondorPlatform": "$CondorPlatform: X86_64-sles_11 $",
  "TargetType": "Machine",
  "QueueTime_Hours": 1,
  "bt_id": "XXXXXXXXXXXX5907225.0#1491577287",
  "bt_index": "backtest-2017.04.07",
  "StreamOut": true,
  "AcctGroupUser": "taskbacktest",
  "GlobalJobId": "XXXXXXXXXXXX5907225.0#1491577287",
  "Iwd": "/XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX/XXXXXXXXXXXXXXXXXXXXX/Data",
  "TransferInputSizeMB": 0,
  "LastPublicClaimId": "<10.12.253.203:9618>#1488908699#41674#...",
  "MemoryUsage": "( ( ResidentSetSize + 1023 ) / 1024 )",
  "NumSystemHolds": 0,
  "PeriodicRemove": "( JobStatus == 2 && time() - EnteredCurrentStatus > 86400 )",
  "ResidentSetSize": 4250,
  "LastRejMatchReason": "no match found ",
  "LastSuspensionTime": 0,
  "ResidentSetSize": "MFG"
```

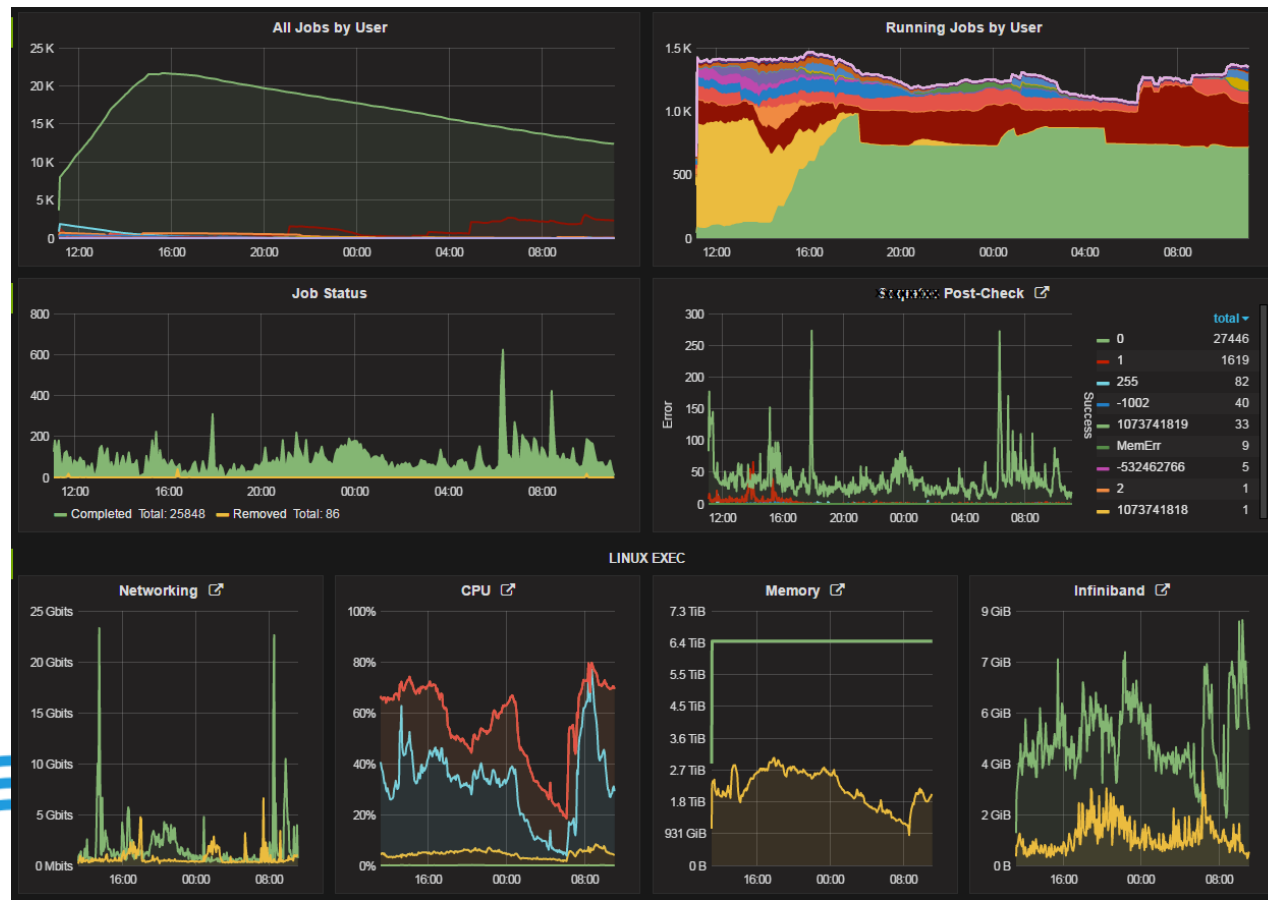


Grafana: One Stop Shop

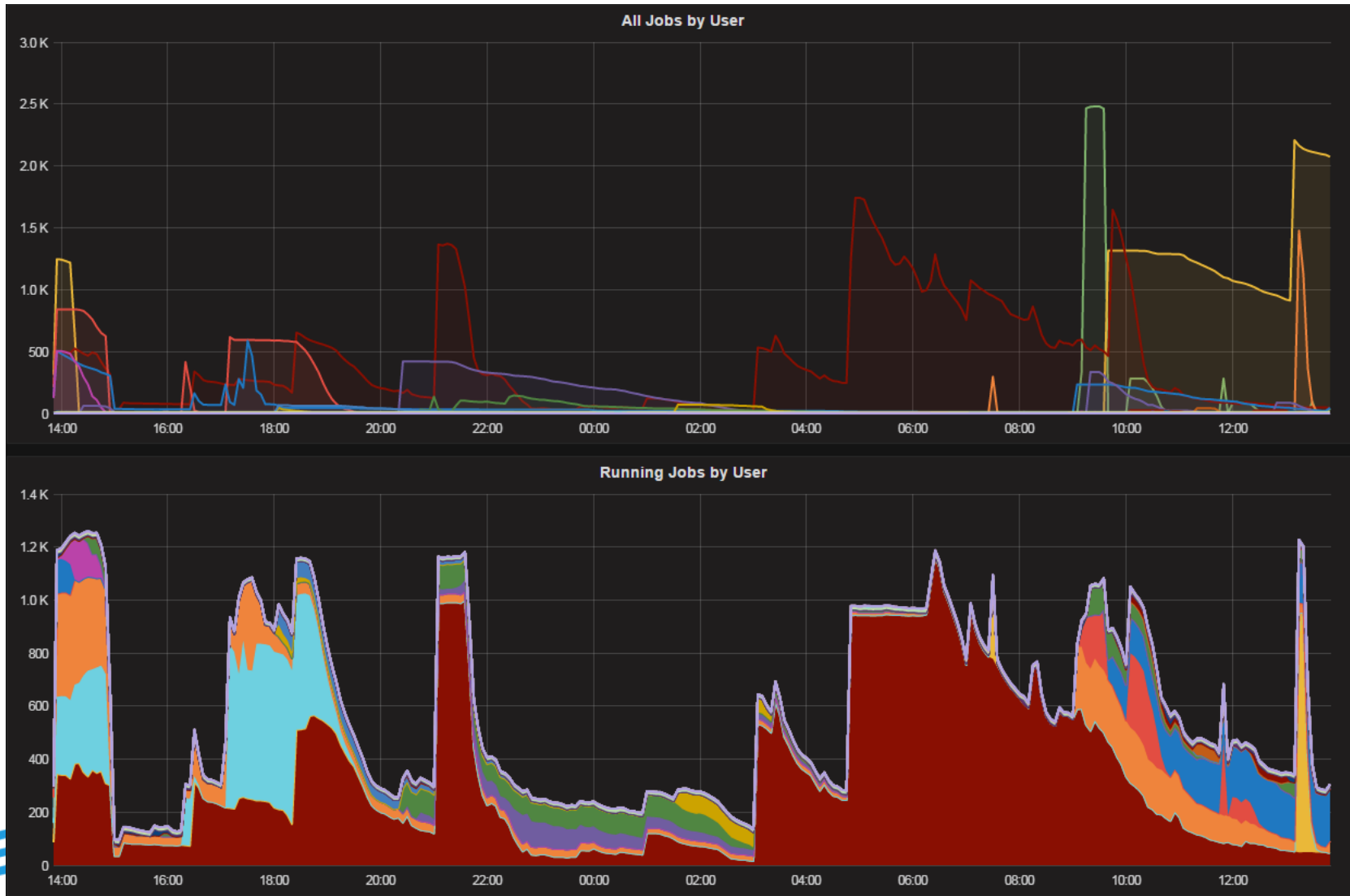
PERFORMANCE	
Condor	Host Performance
Cluster Health ☆	Cluster Performance - Aggregate ★
Condor ★	Cluster Performance - Dublin ☆
Condor - Jobs - Heat Map ☆	Cluster Performance - Linux Exec ☆
Condor - Jobs - Linux ☆	Cluster Performance - Linux Workstations ☆
Condor - Jobs - Matlab ☆	Cluster Performance - Lustre Gateways ☆
Condor - Jobs - Windows ☆	Cluster Performance - Lustre MDSs ☆
Condor - Memory Usage ☆	Cluster Performance - Lustre OSSs ☆
Condor Hosts ☆	Cluster Performance - Matlab ☆
Condor Utilization ☆	Cluster Performance - SSD Filer ☆
Condor: Burst/Overnight Cluster ☆	Cluster Performance - Windows ☆
Tales of Condor ☆	Infrastructure Performance ☆

Grafana Dashboards

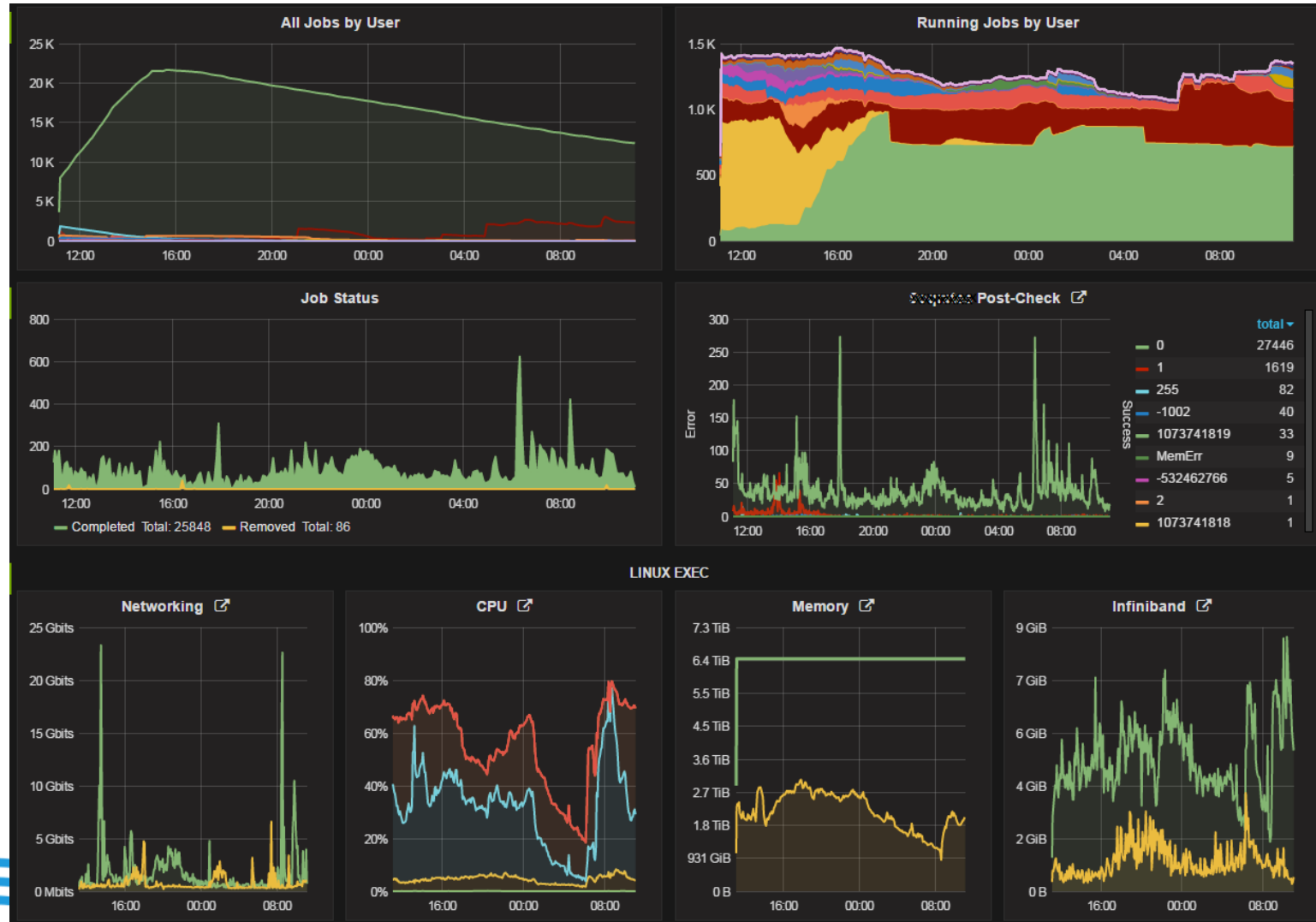
- Easily create dashboards from ES and performance metrics
- Single Pane of Glass for Condor and Infrastructure



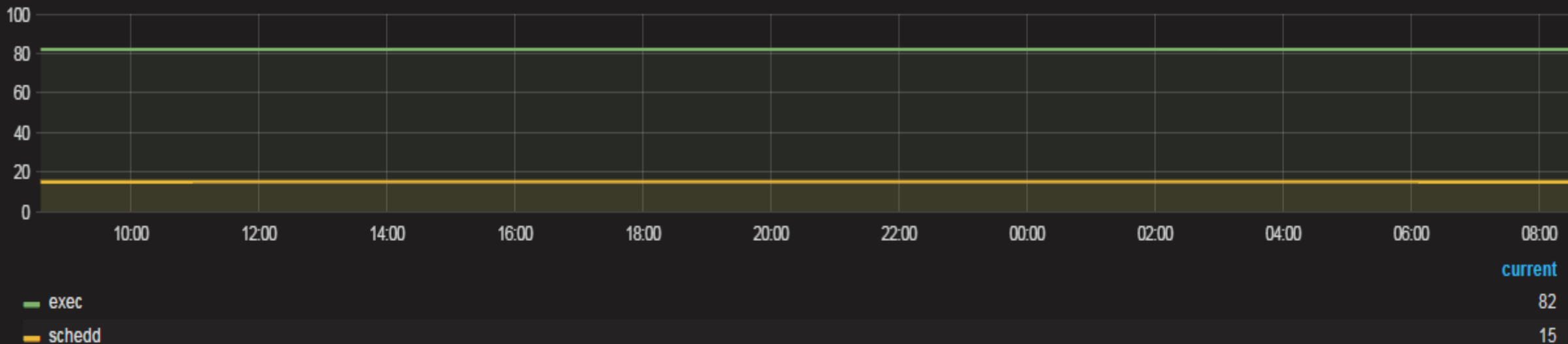
Tales of Condor 2.0



Overall Cluster Health



Hosts



Running Jobs

632

Idle Jobs

835

Running Dags

28 Dags

Held Jobs [↗](#)

0

Long Running Jobs [↗](#)

0

Long Running Dags [↗](#)

0

Long Queued Jobs [↗](#)

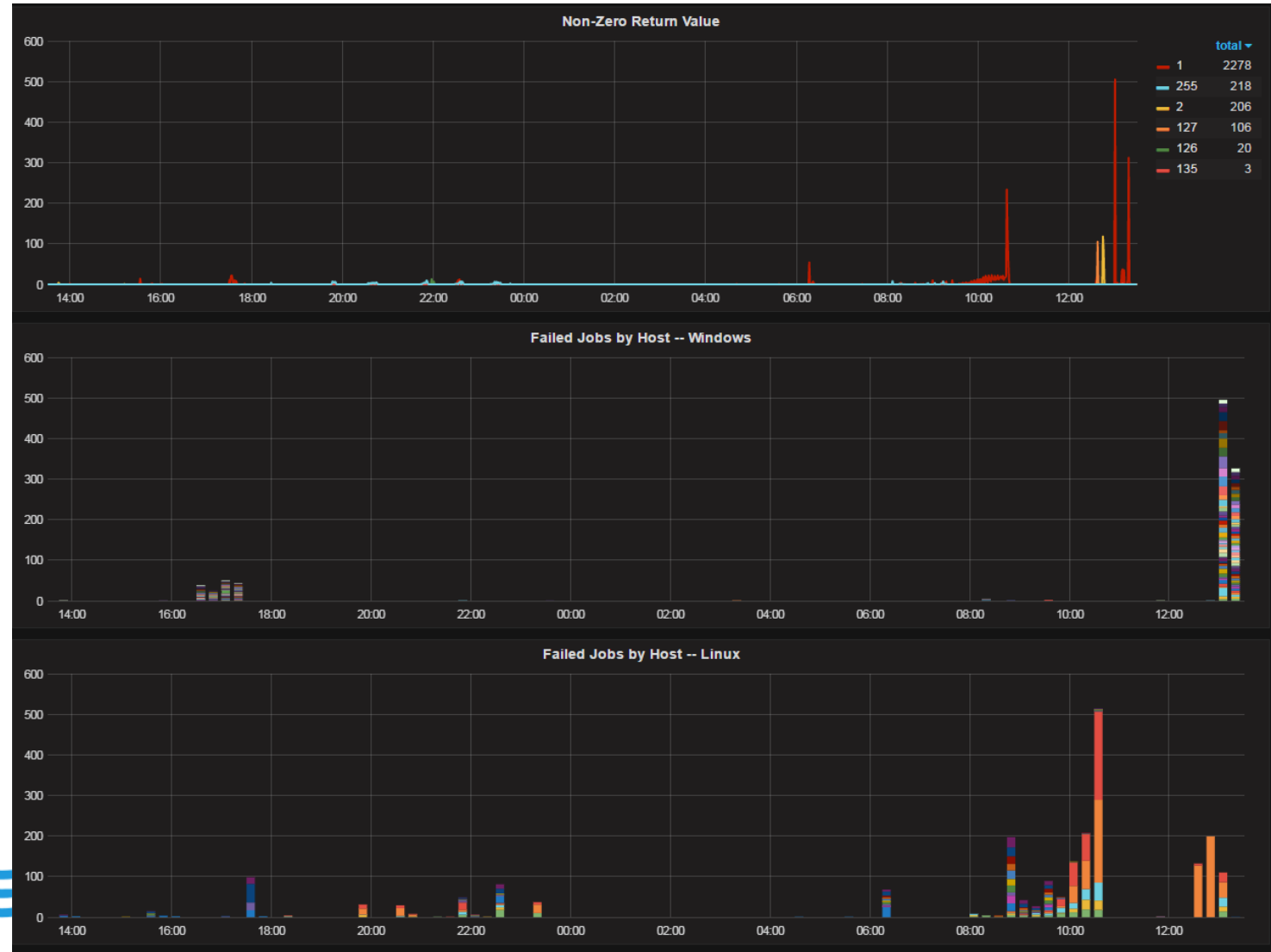
238

Multi-Start Jobs [↗](#)

0

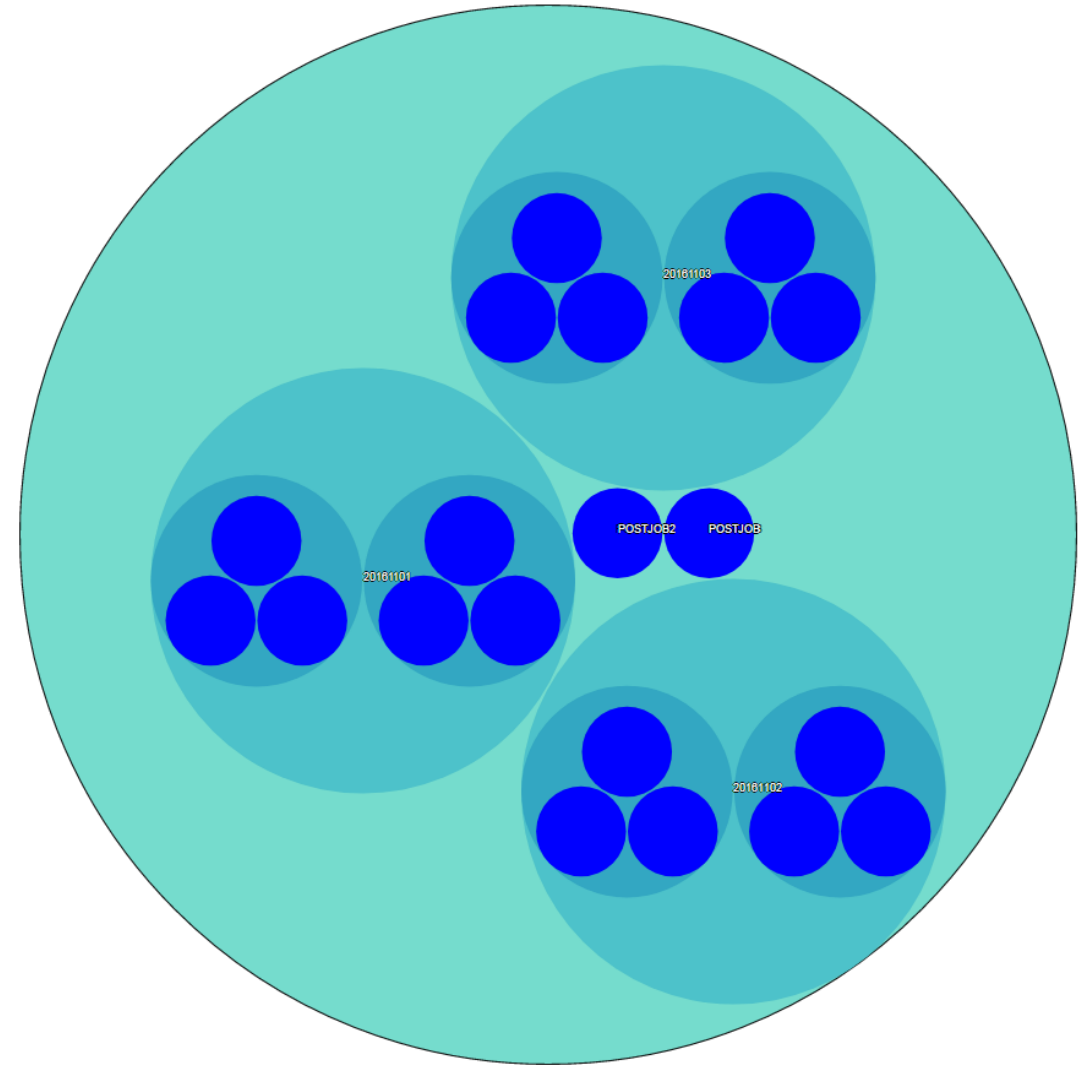
Grafana: Error Rates

- Find blackhole exec
- Postscript vs. ExitCode
- Errors by:
 - User
 - DAG
 - Is it User or infrastructure



Common

- Uses Flask
- ES Queries
- Display job classad
- Workflow overview
- Grid View
- DAG/Job Logs
- Workflow Analysis



Color Key

Description	Node Status Id	Node Count
Node has totally completed	STATUS_DONE	20
Node has failed	STATUS_ERROR	0
Node has not been seen by condor	STATUS_UNREADY	0
Node has not been seen by condor	STATUS_NOT_READY	0
The PRE script is running	STATUS_PRERUN	0
The POST script is running	STATUS_POSTRUN	0
The node is running	STATUS_SUBMITTED	0
The node is queued	STATUS_SUBMITTED_IDLE	0
Parents have completed but not queued	STATUS_READY	0
Total Number of Jobs		

Common: Home (DAGs)

Common

[Workflows](#) [Jobs](#) [Condor What/Who](#) [Error Rate](#) [Cluster Overview](#) [Condor What/Who - Old](#) [Tales of Condor](#)

Running Workflows - 42

Show entries

Search:

Start Time	Status	Owner	Schedd	Grid	Workflow	Total	Completed	Failed	Idle	Running	Not Queued
2017-04-07 12:13:43-04:00	Running	deck	XXXXXXXXXXXXXXXXXXXX	view	test_8	6	2	0	1	0	2
2017-04-07 12:13:31-04:00	Running	deck	XXXXXXXXXXXXXXXXXXXX	view	test_7	6	3	0	0	1	1
2017-04-07 12:13:24-04:00	Running	deck	XXXXXXXXXXXXXXXXXXXX	view	test_6	6	4	0	0	2	0

Showing 1 to 3 of 3 entries (filtered from 42 total entries)

[First](#) [Previous](#) [1](#) [Next](#) [Last](#)

Completed Workflows - 2153

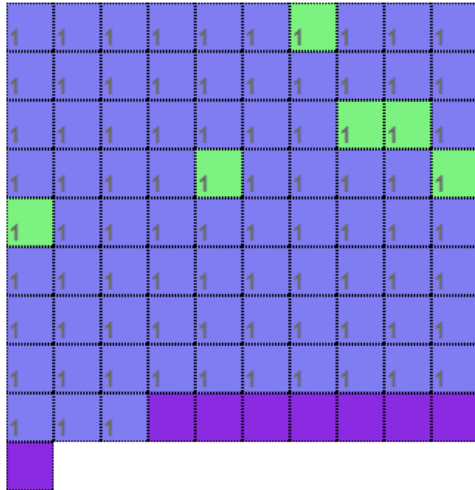
Show entries








[7 Days](#) [30 Days](#) [90 Days](#)
Search:

Start Time	Status	Owner	Schedd	Grid	Workflow	Total	Completed	Failed
2017-04-07 11:51:58-04:00	Completed	deck	XXXXXXXXXXXXXXXXXXXX	view	test_5	6	6	0
2017-04-07 11:51:49-04:00	Completed	deck	XXXXXXXXXXXXXXXXXXXX	view	test_4	6	6	0
2017-04-07 11:51:17-04:00	Completed	deck	XXXXXXXXXXXXXXXXXXXX	view	test_3	6	6	0
2017-04-07 11:51:10-04:00	Completed	deck	XXXXXXXXXXXXXXXXXXXX	view	test_2	6	6	0
2017-04-07 11:51:01-04:00	Completed	deck	XXXXXXXXXXXXXXXXXXXX	view	test_1	6	6	0

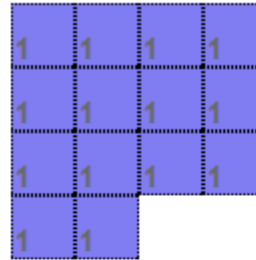
Common: Grid (DAGs)








[Link to Monmon](#)



	Description	Count
	Node has successfully completed	77
	Node has failed	0
	Node is Idle	0
	Node is Running	6
	Node is Held	0
	Node was Removed	0
	Node has not been seen by condor	8
	Total Number of Jobs	91

[Link to Monmon](#)



	Description	Count
	Node has successfully completed	14
	Node has failed	0
	Node is Idle	0
	Node is Running	0
	Node is Held	0
	Node was Removed	0
	Node has not been seen by condor	0
	Total Number of Jobs	14

Status

Completed

Dir Path

unc

XX CombineResults (Windows)

nfs

XX /CombineResults (Linux)

2017-04-07 12:13:24	CombineResults.submit	1.2 kB	[head tail raw]
2017-04-07 12:13:24	job.bat	565 Bytes	[head tail raw]
2017-04-07 12:13:24	job.sh	462 Bytes	[head tail raw]
2017-04-07 12:13:24	linux.subscript	486 Bytes	[head tail raw]
2017-04-07 12:15:23	log	4.1 kB	directory
2017-04-07 12:13:24	postscript.sh	1.4 kB	[head tail raw]
2017-04-07 12:13:24	results	4.1 kB	directory
2017-04-07 12:13:24	win.subscript	490 Bytes	[head tail raw]



Full Job Details

Show 50 entries

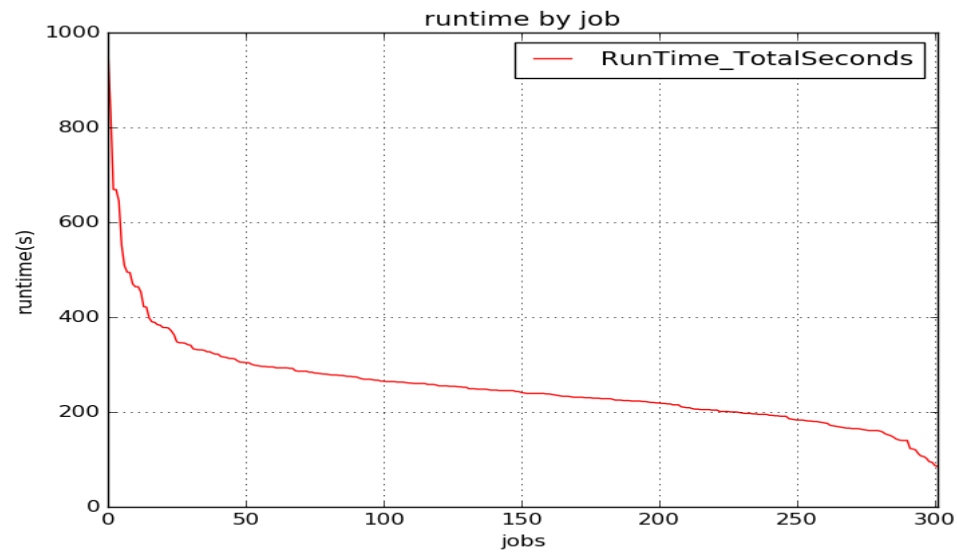
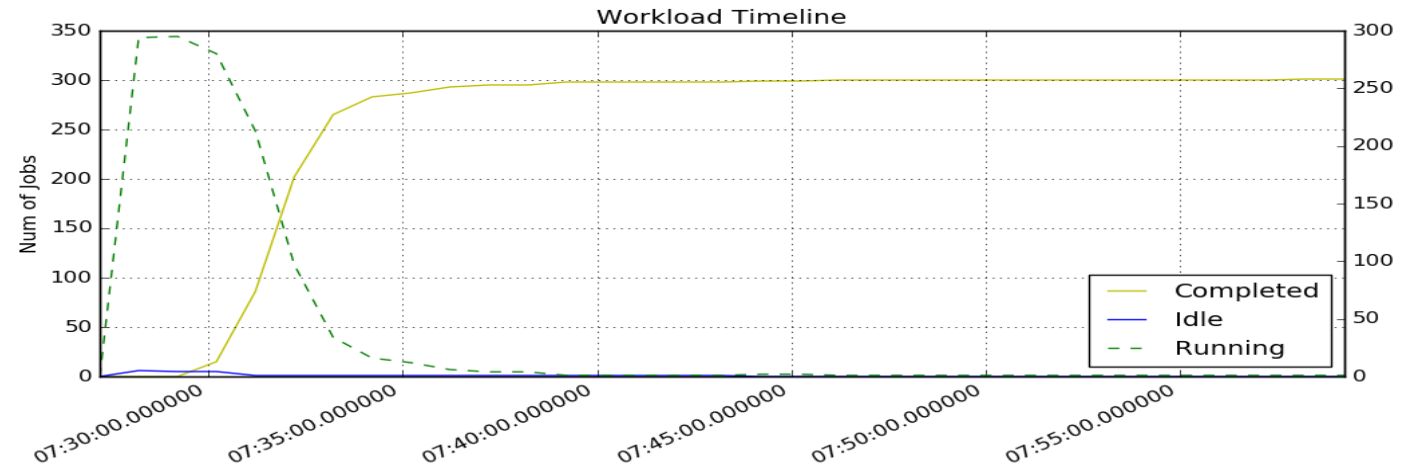
Search:

Key	Value
@timestamp:	2017-04-07T16:15:20.000Z
@version:	1
accounting_group:	group_sotbt
AccountingGroup:	group_sotbt.deck
AcctGroup:	group_sotbt
AcctGroupUser:	deck
Args:	
bt_action:	update
bt_id:	XXm#5907559.0#1491581663
bt_index:	backtest-2017.04.07
bt_type:	condor
ProcessName:	condor



Workflow Analysis

- Information from ES
- Easy to see long legs



Individual Jobs

Running Jobs - 278

Show entries

Search:

Start Time	Status	Job/Node	Cluster Id	User	Runtime (d:h:m:s)	Schedd	Exec
2017-04-12 12:39:05-04:00	Running	Node	5940564.0	root	00:00:08:59	batch	batch
2017-04-12 12:38:25-04:00	Running	Node	5940563.0	root	00:00:09:39	batch	batch
2017-04-12 12:37:04-04:00	Running	Node	84314.213	root	00:00:11:00	batch	batch
2017-04-12 12:37:04-04:00	Running	Node	84314.214	root	00:00:11:00	batch	batch
2017-04-12 12:36:54-04:00	Running	Node	740736.0	root	00:00:11:09	batch	batch
2017-04-12 12:36:50-04:00	Running	Node	740735.0	root	00:00:11:13	batch	batch
2017-04-12 12:36:50-04:00	Running	Node	740733.0	root	00:00:11:13	batch	batch
2017-04-12 12:36:50-04:00	Running	Node	740732.0	root	00:00:11:13	batch	batch
2017-04-12 12:36:44-04:00	Running	Node	740725.0	root	00:00:11:19	batch	batch
2017-04-12 12:36:44-04:00	Running	Node	740730.0	root	00:00:11:19	batch	batch
2017-04-12 12:36:44-04:00	Running	Node	740727.0	root	00:00:11:19	batch	batch
2017-04-12 12:36:24-04:00	Running	Node	84314.212	root	00:00:11:40	batch	batch
2017-04-12 12:35:03-04:00	Running	Node	84314.211	root	00:00:13:01	batch	batch
2017-04-12 12:34:02-04:00	Running	Node	84314.210	root	00:00:14:02	batch	batch
2017-04-12 12:33:21-04:00	Running	Node	84314.208	root	00:00:14:43	batch	batch
2017-04-12 12:33:21-04:00	Running	Node	84314.209	root	00:00:14:43	batch	batch
2017-04-12 12:33:01-04:00	Running	Node	84314.206	root	00:00:15:03	batch	batch
2017-04-12 12:31:40-04:00	Running	Node	84314.205	root	00:00:16:24	batch	batch
2017-04-12 12:31:39-04:00	Running	Node	5940562.0	root	00:00:16:25	batch	batch
2017-04-12 12:31:19-04:00	Running	Node	740724.0	root	00:00:16:44	batch	batch



Common: Benefits

- Loading Condor information from ES
- Can handle multiple submission/workflows types (DAGs, submit)
- User can click through jobs
- Search/Filter
- Shareable
- Consistent views

Running Jobs - 1422

Show 10 entries

Search:

Start Time	Status	Job/Node	Cluster Id	User	Runtime (d:h:m:s)	Schedd	Exec
2017-04-10 14:47:53-04:00	Running	Node	84284.202	wxxxx	00:00:00:11	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:47:35-04:00	Running	Node	84284.201	xxxxx	00:00:00:29	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.45	xxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.66	xxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.79	wxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.84	wxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.85	wxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx
2017-04-10 14:35:54-04:00	Running	Node	84285.89	xxxxx	00:00:12:10	apexqlat810xdsxsqsqxmx	apexqlat810xdsxsqsqxmx

Questions?

