A Slurm Odyssey: Slurm at Harvard FAS Research Computing

Paul Edmon, PhD
Harvard - Research Computing

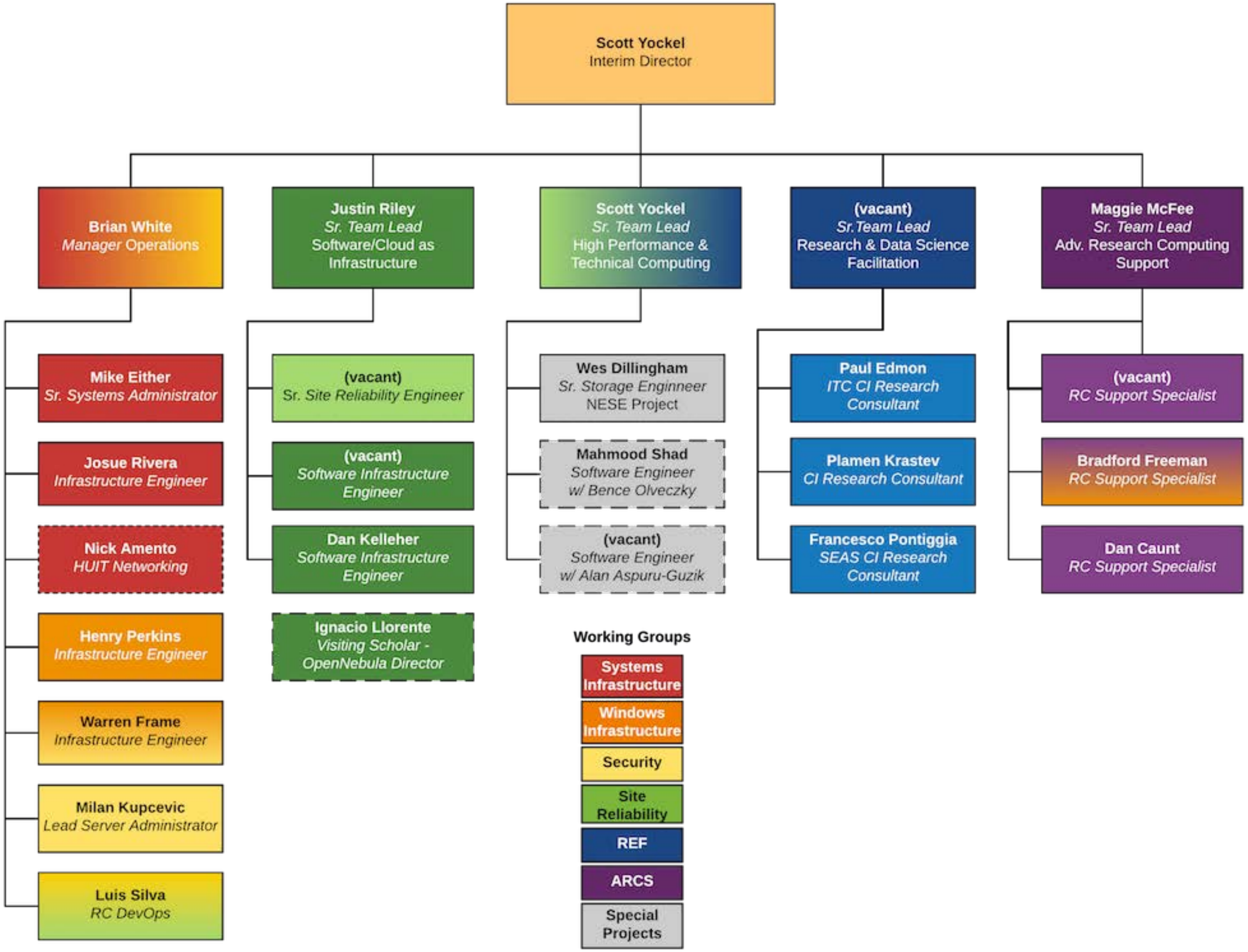
Slurm User Group 2017

Berkeley, CA

Overview

- Research Computing at Harvard
- Odyssey
- Slurm on Odyssey
- Monitoring and Optimizing Slurm
- Future Work

Research Computing



Scott Yockel
Interim Director

Brian White
Manager Operations

Justin Riley
Sr. Team Lead
Software/Cloud as Infrastructure

Scott Yockel
Sr. Team Lead
High Performance & Technical Computing

(vacant)
Sr.Team Lead
Research & Data Science Facilitation

Maggie McFee
Sr. Team Lead
Adv. Research Computing Support

Mike Either
Sr. Systems Administrator

Josue Rivera
Infrastructure Engineer

Nick Amento
HUIT Networking

Henry Perkins
Infrastructure Engineer

Warren Frame
Infrastructure Engineer

Milan Kupcevic
Lead Server Administrator

Luis Silva
RC DevOps

(vacant)
Sr. Site Reliability Engineer

(vacant)
Software Infrastructure Engineer

Dan Kelleher
Software Infrastructure Engineer

Ignacio Llorente
Visiting Scholar - OpenNebula Director

Wes Dillingham
Sr. Storage Enginneer
NESE Project

Mahmood Shad
Software Engineer w/ Bence Olveczky

(vacant)
Software Engineer w/ Alan Aspuru-Guzik

Paul Edmon
ITC CI Research Consultant

Plamen Krastev
CI Research Consultant

Francesco Pontiggia
SEAS CI Research Consultant

(vacant)
RC Support Specialist

Bradford Freeman
RC Support Specialist

Dan Caunt
RC Support Specialist

Working Groups

- Systems Infrastructure
- Windows Infrastructure
- Security
- Site Reliability
- REF
- ARCS
- Special Projects

Job Openings

- **Area Lead, Data Science & Research Facilitation**
- **Sr. Site Reliability Engineer**
- **Software Infrastructure Engineer**
- **Research Computing Support Specialist**
- **Visit: https://www.rc.fas.harvard.edu/about/employment/**

ODYSSEY

Harvard's Largest Cluster

60K+ Cores

35PB+ Storage

Nodes 2100

RAM 260 TB

CUDA 1million

24.7M CPU Hours/month

2.8M Jobs Run/month

8 Miles Network Cabling

Network Ports 5000+

Config Lines 300,000

Racks 190

InfiniBand Switching 56 Gb/s

600+ Active Lab Groups

3 Data Centers

Holyoke, MA
Boston, MA
Cambridge, MA

Total Square Footage

10,000+

Odyssey 3

- 29 Dell M1000 Chassis with 16 M630 Blades: 464 nodes
- Each Blade has 128 GB and two 16 Core Intel Broadwell Chips: 14,848 cores
- Connected by Mellanox FDR Interconnect
- Slated for community release November 2017

Slurm on Odyssey

- Odyssey 1.0 ran LSF, Odyssey 2.0+ has been using Slurm since 2013 beginning with version 2.6.5

- Current Slurm version 17.02.7
  - Upgrade to keep pace with minor releases
  - For major releases we wait till first .1 release to ensure stability
  - Built with:
    - Lua
    - MariaDB
    - HWLoc
    - PMI2

- Slurm Master CentOS 7, Compute Nodes CentOS 6
  - Planning on full CentOS 7 by December 2017

Slurm on Odyssey

	general	serial_requeue	Interact	Bigmem	Unrestricted
Time Limit	7 days	7 days	3 days	No limit	No limit
# Nodes	134	1289	8	7	8
# Cores/Node	64	Varies	64	64	64
Memory/Node (GB)	256	Varies	256	512	256

- 116 Partitions, mostly for various PI owned hardware
- TRES Billing: Memory: 0.25G and AMD: 1CPU, Sandy Bridge: 2CPU and Haswell/Broadwell: 4CPU, and 0.5 for serial_requeue

Slurm on Odyssey

Scheduler Parameters

- Primary Loop
  - default_queue_depth: PartitionNumber*10
  - partition_job_depth: 10

- Backfill Loop
  - bf_continue
  - bf_window: 11520 min
  - bf_resolution: 600 s
  - bf_max_job_test: 10000
  - bf_max_job_start: 1000
  - bf_max_job_user: 10
  - bf_min_prio_reserve: 10000000 (equivalent of Fairshare 0.5)

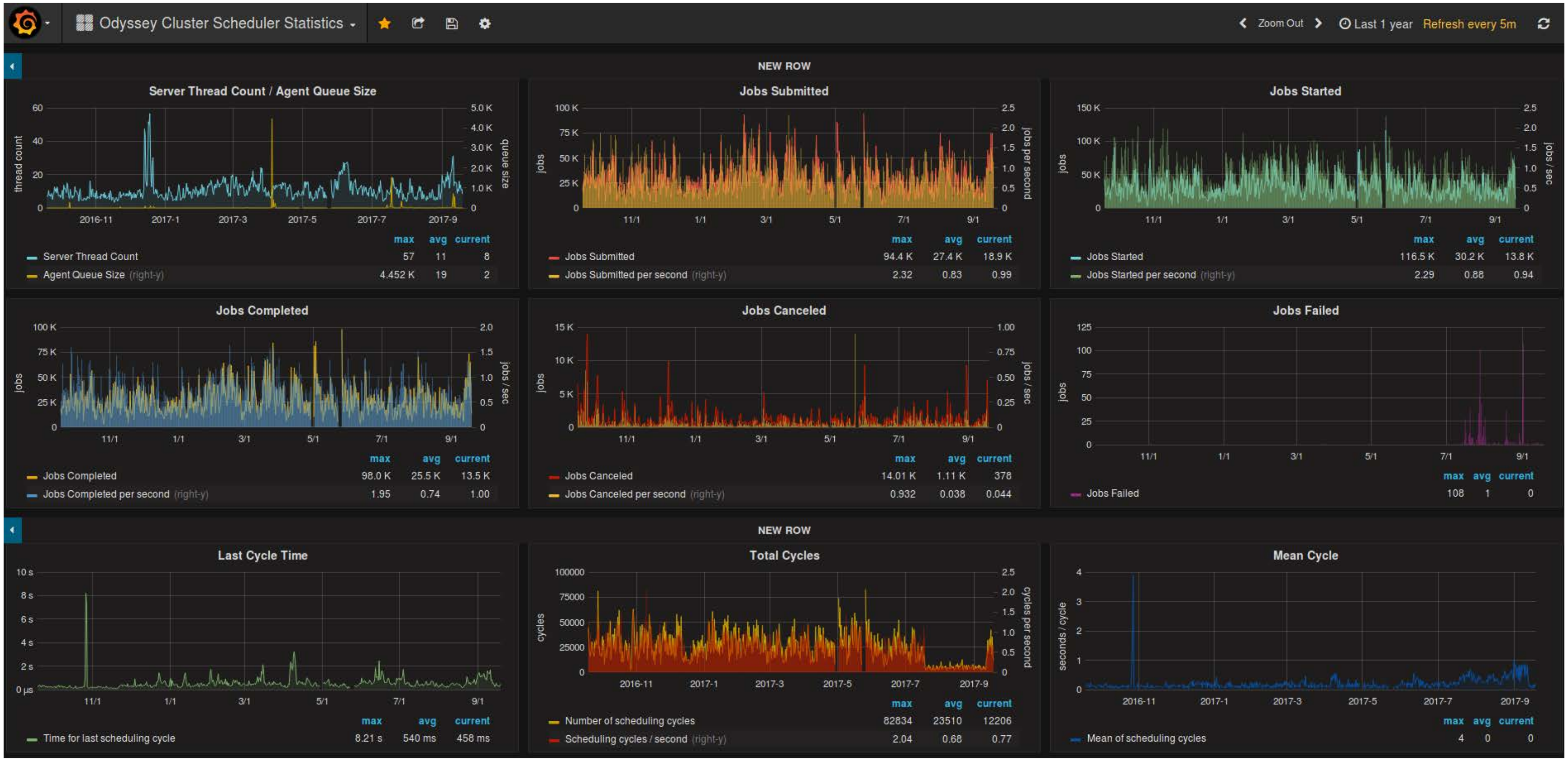
- Other Parameters
  - preempt_youngest_first
  - max_rpc_cnt: 8

Slurm on Odyssey

- Scheduling Priority:
  - Fairshare
    - Each Group gets 100
    - Groups that purchase hardware get Fairshare:
      - AMD: 1
      - Ivy Bridge: 2
      - Broadwell: 4
    - Fairshare HalfLife: 2 days
  - Job Age
    - Maximum score is equivalent to Fairshare 0.5
    - 7 day maximum score

Monitoring

- https://github.com/fasrc/slurm-diamond-collector
- Giovanni Torres' Sdiag Monitor



Odyssey Cluster Scheduler Statistics

Server Thread Count / Agent Queue Size

	max	avg	current
Server Thread Count	57	11	8
Agent Queue Size (right-y)	4.452 K	19	2

Jobs Submitted

	max	avg	current
Jobs Submitted	94.4 K	27.4 K	18.9 K
Jobs Submitted per second (right-y)	2.32	0.83	0.99

Jobs Started

	max	avg	current
Jobs Started	116.5 K	30.2 K	13.8 K
Jobs Started per second (right-y)	2.29	0.88	0.94

Jobs Completed

	max	avg	current
Jobs Completed	98.0 K	25.5 K	13.5 K
Jobs Completed per second (right-y)	1.95	0.74	1.00

Jobs Canceled

	max	avg	current
Jobs Canceled	14.01 K	1.11 K	378
Jobs Canceled per second (right-y)	0.932	0.038	0.044

Jobs Failed

	max	avg	current
Jobs Failed	108	1	0

Last Cycle Time

	max	avg	current
Time for last scheduling cycle	8.21 s	540 ms	458 ms

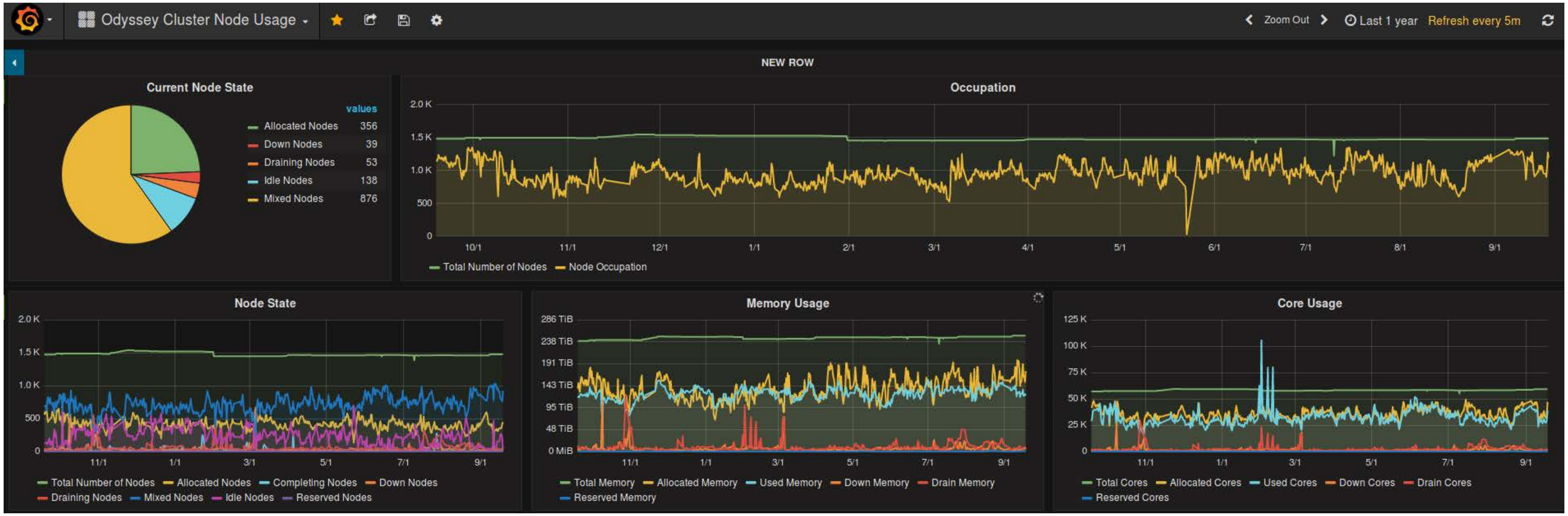
Total Cycles

	max	avg	current
Number of scheduling cycles	82834	23510	12206
Scheduling cycles / second (right-y)	2.04	0.68	0.77

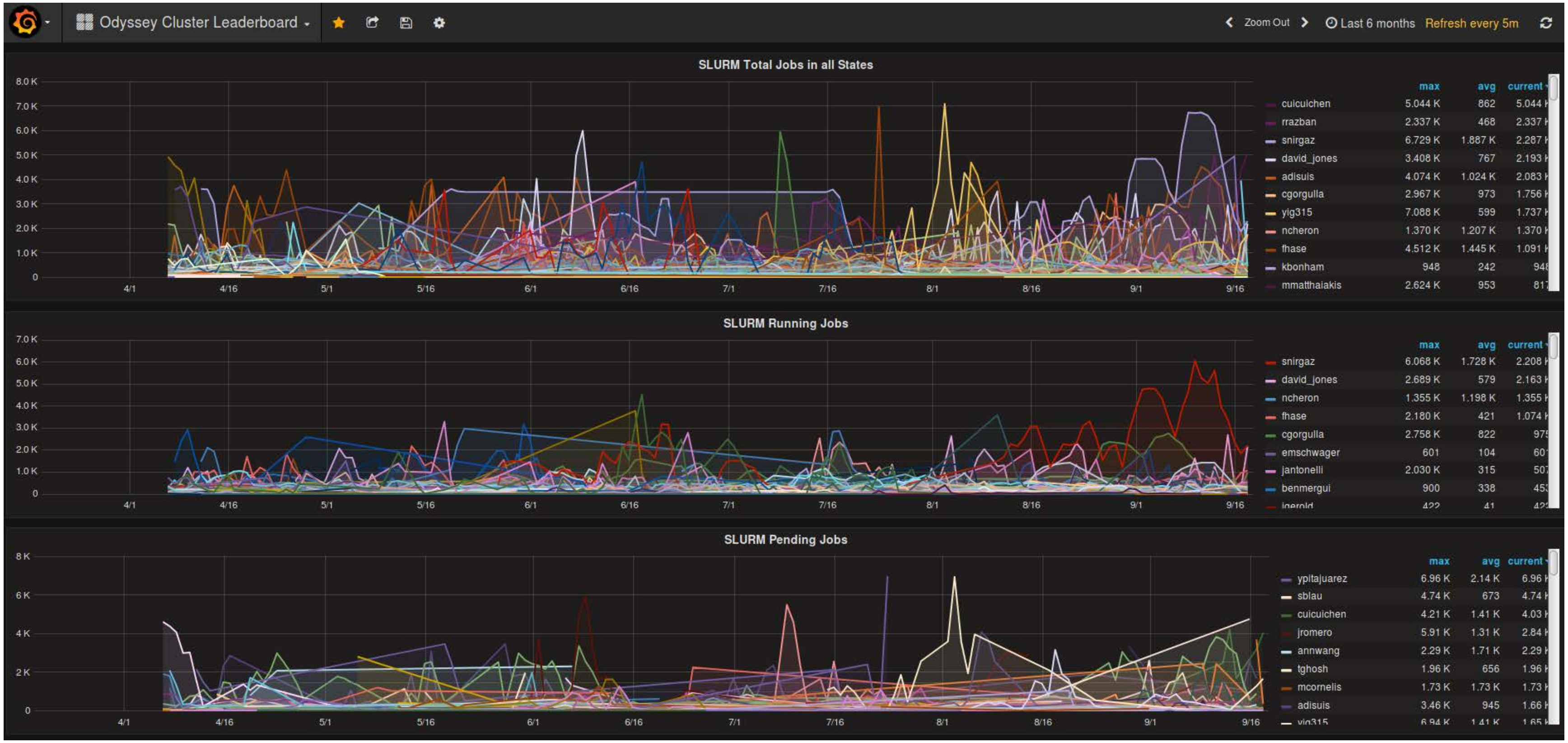
Mean Cycle

	max	avg	current
Mean of scheduling cycles	4	0	0

Monitoring


Odyssey Cluster Node Usage
Zoom Out
Last 1 year
Refresh every 5m
NEW ROW
Current Node State
values
Allocated Nodes 356
Down Nodes 39
Draining Nodes 53
Idle Nodes 138
Mixed Nodes 876
Occupation
2.0 K
1.5 K
1.0 K
500
0
10/1
11/1
12/1
1/1
2/1
3/1
4/1
5/1
6/1
7/1
8/1
9/1
Total Number of Nodes
Node Occupation
Node State
2.0 K
1.5 K
1.0 K
500
0
11/1
1/1
3/1
5/1
7/1
9/1
Total Number of Nodes
Allocated Nodes
Completing Nodes
Down Nodes
Draining Nodes
Mixed Nodes
Idle Nodes
Reserved Nodes
Memory Usage
286 TiB
238 TiB
191 TiB
143 TiB
95 TiB
48 TiB
0 MiB
11/1
1/1
3/1
5/1
7/1
9/1
Total Memory
Allocated Memory
Used Memory
Down Memory
Drain Memory
Reserved Memory
Core Usage
125 K
100 K
75 K
50 K
25 K
0
11/1
1/1
3/1
5/1
7/1
9/1
Total Cores
Allocated Cores
Used Cores
Down Cores
Drain Cores
Reserved Cores

Monitoring



SLURM Total Jobs in all States

	max	avg	current
cuicuichen	5.044 K	862	5.044 K
rrazban	2.337 K	468	2.337 K
snirgaz	6.729 K	1.887 K	2.287 K
david_jones	3.408 K	767	2.193 K
adisuis	4.074 K	1.024 K	2.083 K
cgorgulla	2.967 K	973	1.756 K
yig315	7.088 K	599	1.737 K
ncheron	1.370 K	1.207 K	1.370 K
fhase	4.512 K	1.445 K	1.091 K
kbonham	948	242	948
mmatthaiakis	2.624 K	953	817

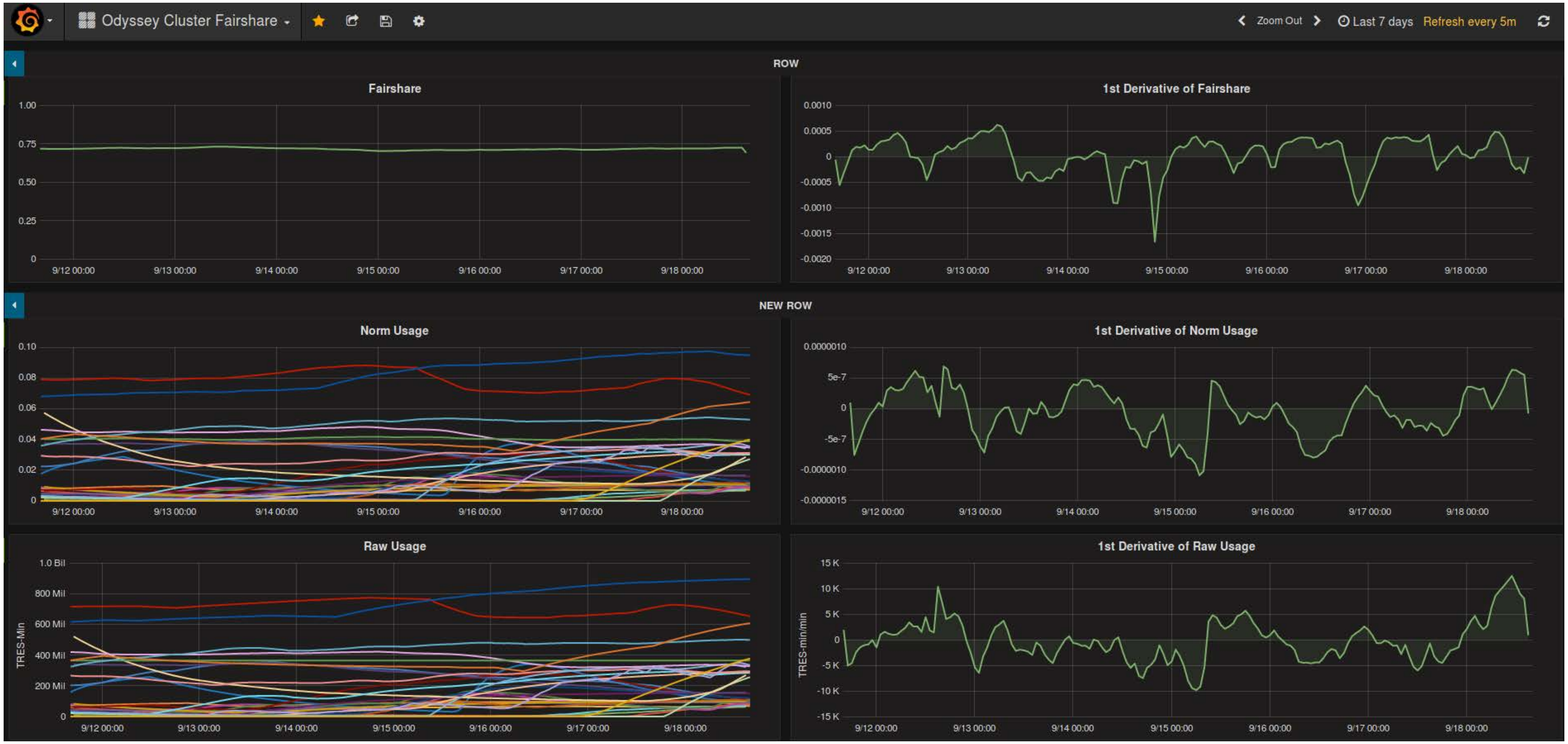
SLURM Running Jobs

	max	avg	current
snirgaz	6.068 K	1.728 K	2.208 K
david_jones	2.689 K	579	2.163 K
ncheron	1.355 K	1.198 K	1.355 K
fhase	2.180 K	421	1.074 K
cgorgulla	2.758 K	822	975
emschwager	601	104	601
jantonelli	2.030 K	315	507
benmergui	900	338	453
igerold	422	41	422

SLURM Pending Jobs

	max	avg	current
ypitajuarez	6.96 K	2.14 K	6.96 K
sblau	4.74 K	673	4.74 K
cuicuichen	4.21 K	1.41 K	4.03 K
jromero	5.91 K	1.31 K	2.84 K
annwang	2.29 K	1.71 K	2.29 K
tghosh	1.96 K	656	1.96 K
mcornelis	1.73 K	1.73 K	1.73 K
adisuis	3.46 K	945	1.66 K
yiq315	6.94 K	1.41 K	1.65 K

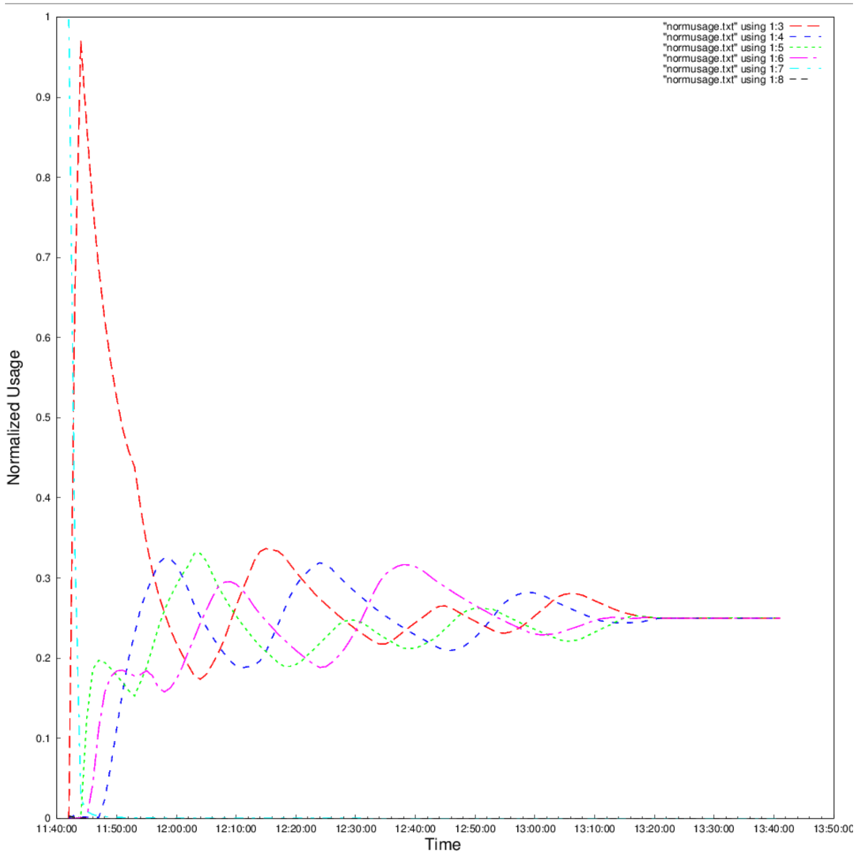
Monitoring

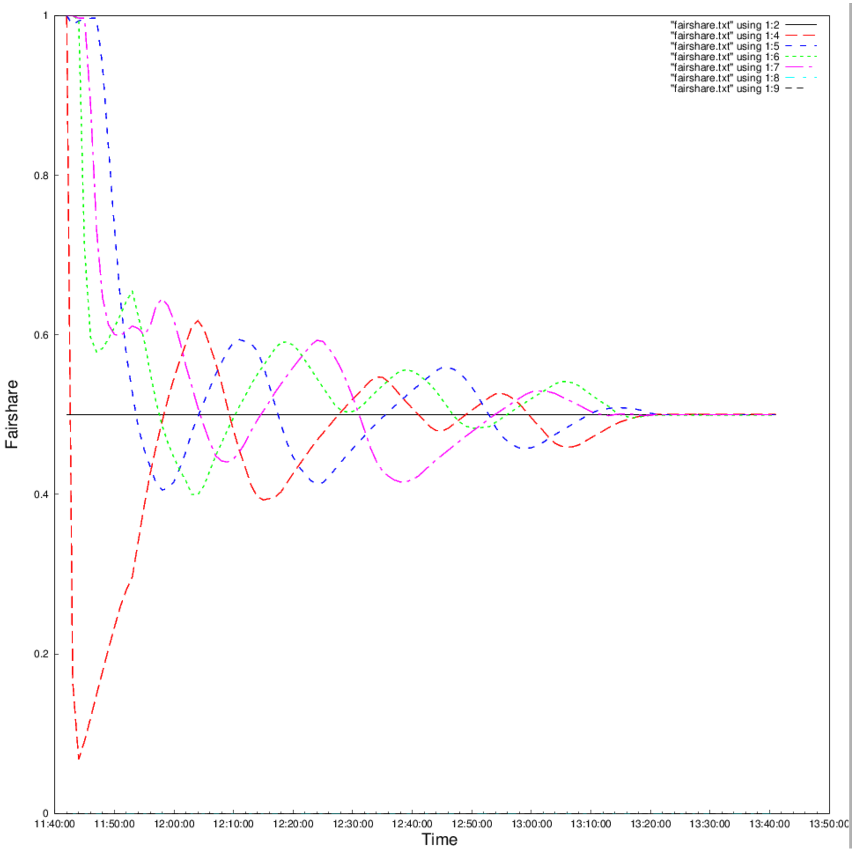

Odyssey Cluster Fairshare
Zoom Out
Last 7 days
Refresh every 5m
ROW
Fairshare
1.00
0.75
0.50
0.25
0
9/12 00:00
9/13 00:00
9/14 00:00
9/15 00:00
9/16 00:00
9/17 00:00
9/18 00:00
1st Derivative of Fairshare
0.0010
0.0005
0
-0.0005
-0.0010
-0.0015
-0.0020
NEW ROW
Norm Usage
0.10
0.08
0.06
0.04
0.02
0
1st Derivative of Norm Usage
0.0000010
5e-7
0
-5e-7
-0.0000010
-0.0000015
Raw Usage
TRES-Min
1.0 Bil
800 Mil
600 Mil
400 Mil
200 Mil
0
1st Derivative of Raw Usage
TRES-min/min
15 K
10 K
5 K
0
-5 K
-10 K
-15 K

Testing And Optimization

- https://github.com/fasrc/stdg
- Simple bash scripts that generate test decks based on a configuration file
- Performance Testing
  - Odyssey 2 grew from 28,000 cores to 60,000 cores. User count and number of jobs also increased.
  - Engaged in a study to optimize scheduler to maximize throughput and lower latency
  - Changes
    - AMD Abu Dhabi -> Intel Haswell
    - HDD -> SSD
    - Mysql 5.1 -> MariaDB 10.1.18
    - CentOS 6 -> CentOS 7
  - Result: 10x speed up in scheduler performance

Fairshare Testing





Future Work

- Consolidate Partitions
  - Make High Priority Partition with a Fairshare Gate
  - Move From PI Owned Partitions to Fairshare Resource Allocation
  - Make GPU Specific Requeue Queue
  - Use Constraints for Serial Queues
- Integrate Fairshare Calculation and Assignment with User Portal
- Feature Requests:
  - Have Pending Jobs Count Against Fairshare
  - Reservation Charge Back
  - Negative Match Constraint
  - slurm.conf Syntax Checker