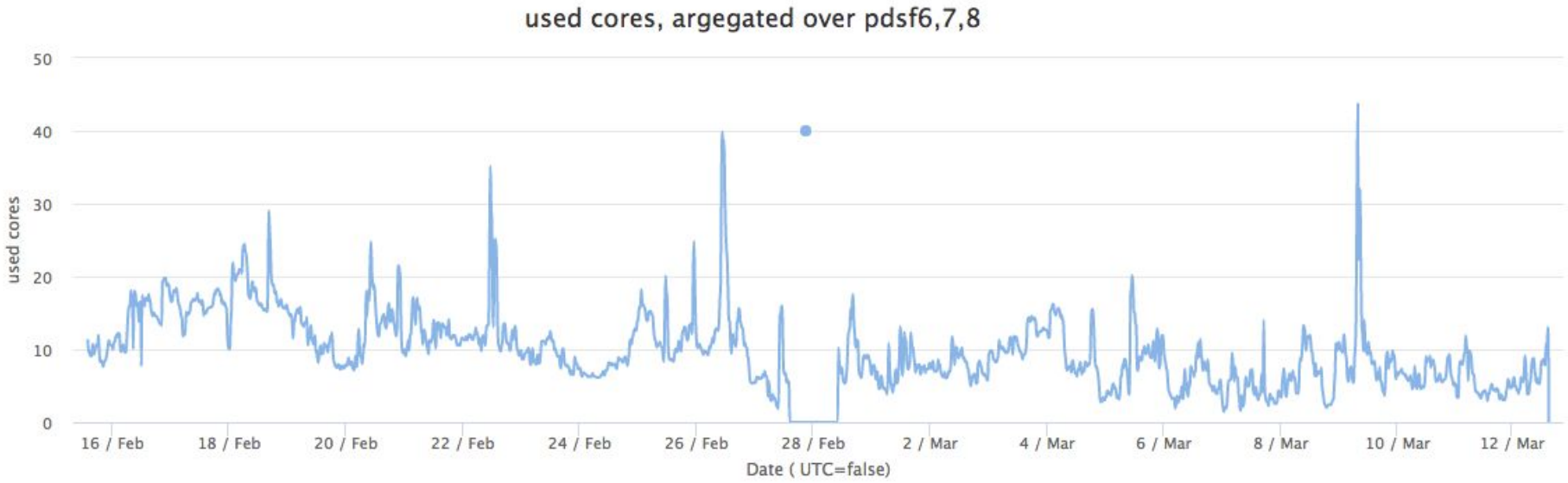
# PDSF User Meeting

- PDSF performance
- Announcements
- New PDSF queues
- New Slurm defaults
- AOB

# aggregated load on PDSF interactive nodes

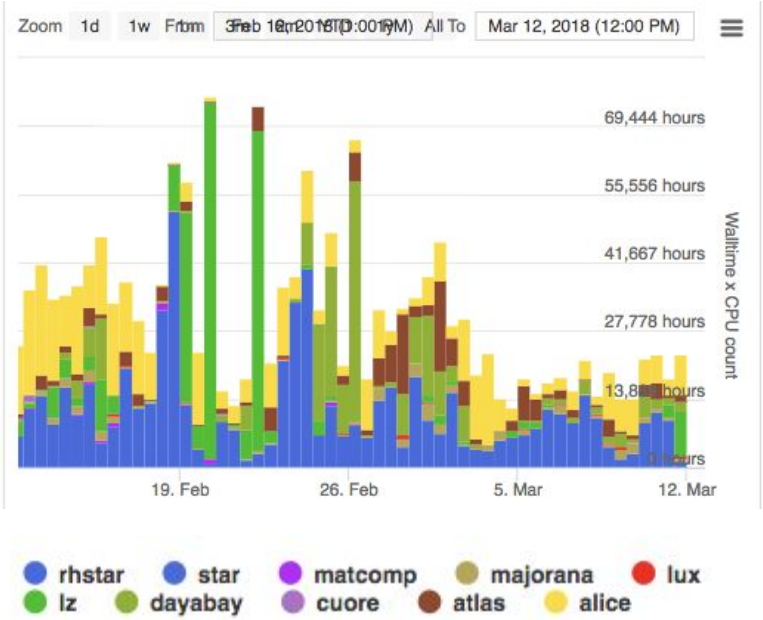
https://portal-auth.nersc.gov/pdsf-mon/



# SLURM CPU aggregated over last month

SLURM : **completed** jobs in last month

http://portal.nersc.gov/project/mpccc/ebasheer/jobbygroup.php




Summary for **Feb 12, 2018 (12:00)** to **Mar 12, 2018 (15:00)**

The average and standard deviation refer to the average and standard deviation of the points for each series across the date-time range selected. The values are all in units of the data type specified.

| Series | Total | Total (%) | Average | Standard Deviation |
|---|---|---|---|---|
| rhstar | 637,370 | 36.4% | 2,832.8 | 2,830.4 |
| star | 0 | 0.0% | 0.0 | 0.0 |
| matcomp | 6,636 | 0.4% | 29.6 | 121.4 |
| majorana | 44,288 | 2.5% | 197.7 | 318.4 |
| lux | 7,575 | 0.4% | 33.8 | 196.3 |
| lz | 257,979 | 14.8% | 1,151.7 | 5,572.7 |
| dayabay | 225,487 | 12.9% | 1,006.6 | 2,633.3 |
| cuore | 2,599 | 0.1% | 11.6 | 74.3 |
| atlas | 140,939 | 8.1% | 626.4 | 1,086.1 |
| alice | 425,945 | 24.4% | 1,893.1 | 2,065.4 |
| Total | 1,748,819 | 100.0% | 7,772.5 | 6,472.3 |

3800 jobs * 24 h = 91k cpu*h/day
→ 640k cpu*h /week
→ 2.7 M cpu*h per month

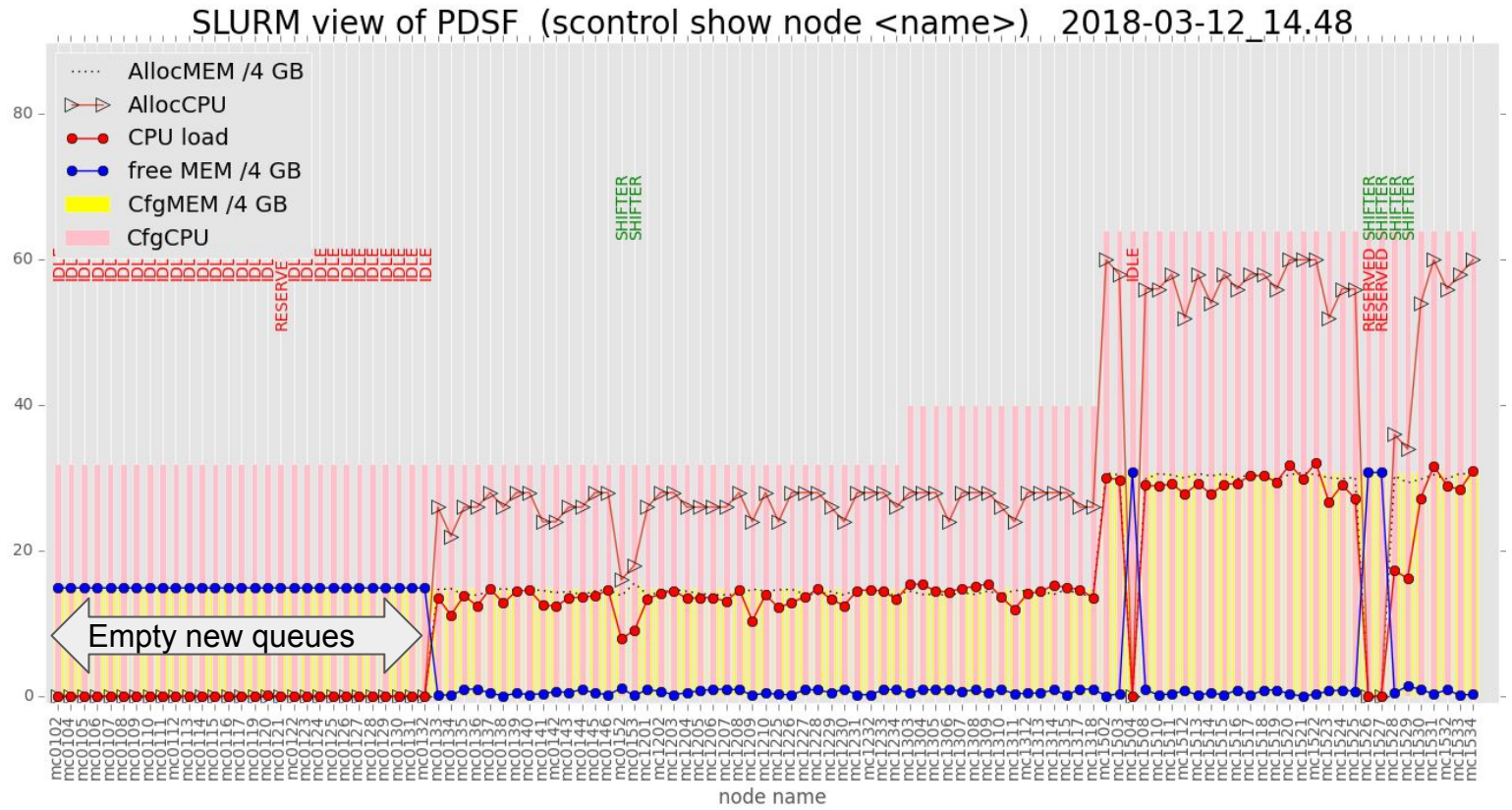
February Total: ~2.4M cpu*hours, was 28% more.
Reasons: unused new queues, changed Slurm defaults

# New Slurm Shifter queues

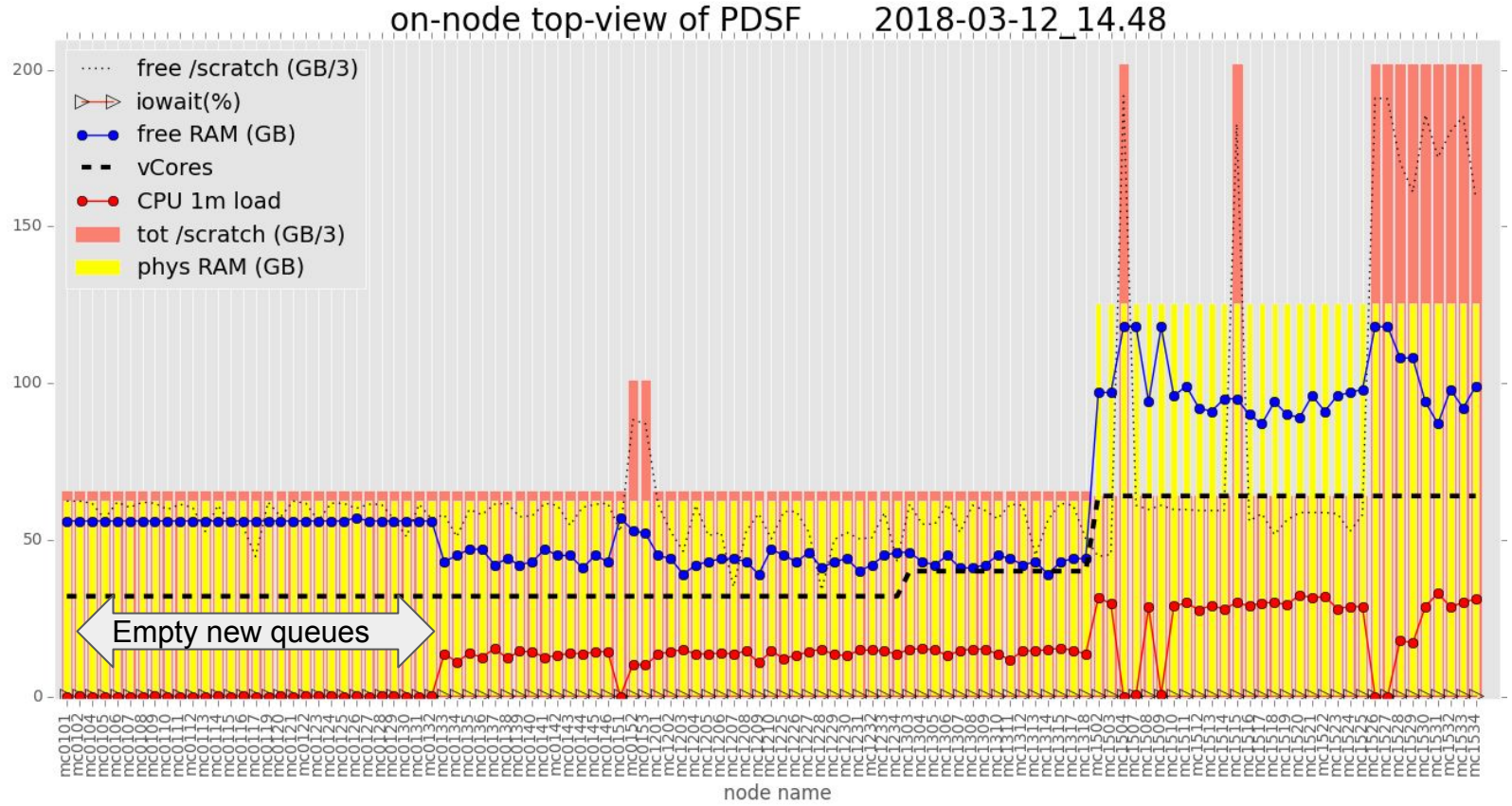
| partition | OS | TotalCPUs | Time limit | MaxJobPA (per account) | MaxJobPU (per user) | Relative priority | remarks |
|---|---|---|---|---|---|---|---|
| shared-chos | chos | 2968 | 2 days | | | 0 | |
| shared | Shifter | 296 | 2 days | | | 0 | To be cannibalized |
| long | Shifter | 384 | 2 days | | | 0 | |
| short | Shifter | 288 | 5 hours | | | 0 | |
| realtime | Shifter | 200 of 256 | 4 hours | 50 | | 0 | share common hardware |
| debug | Shifter | 60 of 256 | 30 min | | 2 | 10 | |

List QOS limits
pdsf6 $ **sacctmgr show qos format=Name,Priority,GrpTRES,MaxTRESPU,MaxTRESPA%100,MaxJobsPA,MaxJobsPU**
List timeouts: pdsf6 $ **sinfo -a ; scontrol show partition -a short**

# PDSF load SLURM snapshot



# PDSF load node snapshot



# Slurm: change of defaults

**Reason:** CSG will not tolerate failing nodes due to oversubscription of resources.

The following changes were enforced on Slurm all queues :

- Use true RAM per node (was x 2.3 oversubscribed)
- compute high mem use tax : nCPU= mem/4 GB (was mem/7.07 GB)
- Num job slots per node: 2x phys cores (was 2*(n-2))
- Default mem per task: 4 GB (no changes)
- Used cpu*h half decay time: 14 days (no change)
- Job start time reduced from ~minutes to seconds (if resources are available, e.g. salloc start immediately)

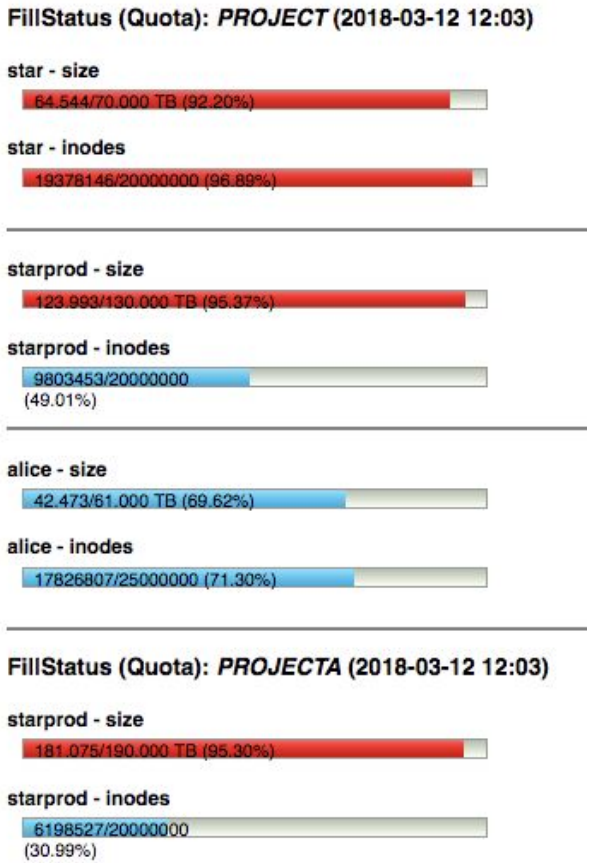
New working point set on March 7, 2018

# /project(a) utilization - snapshot

http://portal.nersc.gov/project/star/jthaeder/diskUsage/overview/indexExt.html
https://my.nersc.gov/data-mgt.php

```
cori12:~>  prjquota dayabay
           ---------- Space (GB) ---------     ------------- Inode --------------
Project          Usage     Quota    Percent        Usage      Quota     Percent
---------------  --------- --------- ---------    ---------- ---------- ----------
dayabay           853138    870400       98     100937147  150000000        67
balewski@cori06:~> prjaquota dayabay
           ---------- Space (GB) ---------     ------------- Inode --------------
Project          Usage     Quota    Percent        Usage      Quota     Percent
---------------  --------- --------- ---------    ---------- ---------- ----------
dayabay           846109   1126400       75       6245563   10000000        62


balewski@cori06:~> prjquota majorana
           ---------- Space (GB) ---------     ------------- Inode --------------
Project          Usage     Quota    Percent        Usage      Quota     Percent
---------------  --------- --------- ---------    ---------- ---------- ----------
majorana          34869     40960       85       3344039    4000000        83
balewski@cori06:~> prjaquota majorana
           ---------- Space (GB) ---------     ------------- Inode --------------
Project          Usage     Quota    Percent        Usage      Quota     Percent
---------------  --------- --------- ---------    ---------- ---------- ----------
majorana          58686     61440       95       4822902   10000000        48
```



# Announcements

Bi-weekly office hours March 15,29, 59-4016A

PDSF user meeting
- Tuesday, April 10

**Summary of Upcoming Events and Key Dates**

March 2018

| Su | Mo | Tu | We | Th | Fr | Sa | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | | |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | | |
| 11 | 12 | 13 | 14 | *15* | 16 | 17 | 15 Mar | NUG Monthly Webinar [1] |
| 18 | 19 | 20 | *21* | 22 | 23 | 24 | 21 Mar | New User Training [2] |
| | | | | | | | 21 Mar | Edison Maint [3] |
| 25 | 26 | 27 | 28 | 29 | 30 | 31 | | |

April 2018

| Su | Mo | Tu | We | Th | Fr | Sa | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 8 | 9 | *10** | *11* | 12 | 13 | 14 | 10 Apr | Quarterly Alloc Reduction [4] |
| | | | | | | | 11 Apr | Cori Monthly Maint |
| 15 | 16 | 17 | 18 | *19* | 20 | 21 | 19 Apr | NUG Monthly Webinar |
| 22 | 23 | 24 | *25* | 26 | 27 | 28 | 25 Apr | Edison Monthly Maint |
| 29 | 30 | | | | | | | |

May 2018

| Su | Mo | Tu | We | Th | Fr | Sa | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | | |
| 6 | 7 | *8* | 9 | 10 | 11 | 12 | 8 May | Quarterly Maint [5] |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 | | |

# Jeff's proposal - go back to 2014 set point

1) Reset the ½ life to the UGE value so it matches what's in our MoU.

2) Remove the memory weight on core usage. It is not included in our Share formula. Add hard limit on 8 GB RAM per job.

3) Reset the default memory to a small number – I suggest 2GB/job. It used to be 1.1GB/job. It's currently 4GB/job
   a. Implement and educate users on simple way to set all their jobs memory request
      i. Similar to the ~/.sge_request file
   b. Make sure batch kills by OOM events are relayed back to the user

4) Monitor the memory & disk I/O usage dynamically and report to PDSF user community. Not a spot check – e.g. this is CSG's responsibility. Historical data would also be useful if available. Use this information to help us set appropriate resource limits to insure stability.

5) Report on new queue usage to recommend reductions/expansions (Jan- meet users)

# **Craig's proposal:** add 2 chos queues

take the "Shared" (Shifter; 296 cores) queue, which is a replica of the Long queue, and split it into debug-chos (56 cores) and short-chos queues (240 cores).