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

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

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

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

<table>
<thead>
<tr>
<th>Series</th>
<th>Total</th>
<th>Total (%)</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rhstar</td>
<td>637,370</td>
<td>36.4%</td>
<td>2,832.8</td>
<td>2,830.4</td>
</tr>
<tr>
<td>star</td>
<td>0</td>
<td>0.0%</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>matcomp</td>
<td>6,636</td>
<td>0.4%</td>
<td>29.6</td>
<td>121.4</td>
</tr>
<tr>
<td>majorana</td>
<td>44,286</td>
<td>2.5%</td>
<td>197.7</td>
<td>316.4</td>
</tr>
<tr>
<td>lux</td>
<td>7,575</td>
<td>0.4%</td>
<td>33.6</td>
<td>196.3</td>
</tr>
<tr>
<td>lz</td>
<td>257,979</td>
<td>14.8%</td>
<td>1,151.7</td>
<td>5,572.7</td>
</tr>
<tr>
<td>dayabay</td>
<td>225,487</td>
<td>12.9%</td>
<td>1,006.6</td>
<td>2,833.3</td>
</tr>
<tr>
<td>cuore</td>
<td>2,598</td>
<td>0.1%</td>
<td>11.6</td>
<td>74.3</td>
</tr>
<tr>
<td>atlas</td>
<td>140,939</td>
<td>8.1%</td>
<td>626.4</td>
<td>1,096.1</td>
</tr>
<tr>
<td>alice</td>
<td>425,945</td>
<td>24.4%</td>
<td>1,893.1</td>
<td>2,005.4</td>
</tr>
<tr>
<td>Total</td>
<td>1,748,819</td>
<td>100.0%</td>
<td>7,772.5</td>
<td>6,472.3</td>
</tr>
</tbody>
</table>

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

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

List QOS limits
```bash
pdsf6 $ sacctmgr show qos format=Name,Priority,GrpTRES,MaxTRESPU,MaxTRESPA%100,MaxJobsPA,MaxJobsPU
```

List timeouts:
```bash
pdsf6 $ sinfo -a ; scontrol show partition -a short
```

Jan Balewski, NERSC

PDSF User Meeting
PDSF load SLURM snapshot

SLURM view of PDSF (scontrol show node <name>)  2018-03-12_14.48

Empty new queues
PDSF load node snapshot

on-node top-view of PDSF  2018-03-12_14.48

Empty new queues
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
Jan Balewski, NERSC
PDSF User Meeting

/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:~> prjquota 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:~> prjquota majorana
---------- Space (GB) ---------     ------------- Inode -------------
Project    Usage    Quota    Percent      Usage    Quota    Percent
---------------   ---------  ---------  ---------     ----------  ----------  ----------
majorana     58686     61440     95        4822902    10000000          48

FillStatus (Quota): PROJECT (2018-03-12 12:03)

star - size
44.544/70.000 TB (63.20%)

star - inodes
19378148/20000000 (96.89%)

starprod - size
123.000/130.000 TB (95.37%)

starprod - inodes
9803453/20000000 (49.31%)

alice - size
42.473/61.000 TB (69.82%)

alice - inodes
17828807/25000000 (71.30%)

FillStatus (Quota): PROJECTA (2018-03-12 12:03)

starprod - size
181.075/190.000 TB (95.30%)

starprod - inodes
6198527/20000000 (30.99%)
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]
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
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).