PDSF User Meeting

- PDSF performance/ transition to SLURM
- Announcements
- AOB
- Large scale Cori productions
  - STAR (Mustafa) : 800 nodes x 32 tasks x 2 days , completed
  - Dayabay : 300 nodes x 50 tasks x 7 days, in preparations
Modernization of PDSF

2017 Calendar

- Testing SLURM & Shifter
- Users are testing
- UGE 100%
- SLURM 10%
- UGE 90%
- SLURM 10%
- UGE 90%
- SLURM 10%

Continue until Mendel is decommissioned → Cori

Users are testing

UGE 75%
SLURM 25%

UGE 5%
SLURM 95%

SLURM 100%

? CHOS→ Shifter

Jan Balewski, NERSC
PDSF status
July 24, 2017
UGE CPU aggregated over month

9-13-2017: SLURM 30%
10-9-2017: SLURM 95%

UGE: running jobs

The End
### SLURM utilization this morning

<table>
<thead>
<tr>
<th>Rjob</th>
<th>Rcpu</th>
<th>Rcpu*h</th>
<th>PDjob</th>
<th>PDcpu</th>
<th>user:account:partition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>1933</td>
<td>9018.9</td>
<td>660</td>
<td>660</td>
<td>alicesgm alice shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>365</td>
<td>365</td>
<td>eboulton lux shared-cho</td>
</tr>
<tr>
<td>233</td>
<td>233</td>
<td>163.8</td>
<td>344</td>
<td>344</td>
<td>heshu star shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>30</td>
<td>30</td>
<td>ibles lz shared-cho</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>11.3</td>
<td>0</td>
<td>0</td>
<td>jjyj star shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1</td>
<td>jrager majoran shared-cho</td>
</tr>
<tr>
<td>378</td>
<td>378</td>
<td>47.2</td>
<td>27</td>
<td>27</td>
<td>klorley lz shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>104</td>
<td>104</td>
<td>mjd majoran shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>15016</td>
<td>15016</td>
<td>weiji lux shared-cho</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rjob</th>
<th>Rcpu</th>
<th>Rcpu*h</th>
<th>PDjob</th>
<th>PDcpu</th>
<th>account:partition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>1933</td>
<td>9018.9</td>
<td>660</td>
<td>660</td>
<td>alice shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>15381</td>
<td>15381</td>
<td>lux shared-cho</td>
</tr>
<tr>
<td>378</td>
<td>378</td>
<td>47.2</td>
<td>57</td>
<td>57</td>
<td>lz shared-cho</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>105</td>
<td>105</td>
<td>majoran shared-cho</td>
</tr>
<tr>
<td>249</td>
<td>249</td>
<td>175.1</td>
<td>344</td>
<td>344</td>
<td>star shared-cho</td>
</tr>
<tr>
<td>2560</td>
<td>2560</td>
<td>9241.2</td>
<td>16547</td>
<td>16547</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

```
psdf6 $ scontrol show partition shared-chos
PartitionName=shared-chos
 AllowGroups=ALL AllowAccounts=ALL AllowQos=ALL
 AllocNodes=ALL Default=YES QoS=N/A
 DefaultTime=1-00:00:00 DisableRootJobs=NO ExclusiveUser=NO GraceTime=0 Hidden=NO
 MaxNodes=1 MaxTime=2-00:00:00 MinNodes=1 LLN=YES MaxCPUsPerNode=UNLIMITED
 Nodes=mc01[05-46],mc12[01-19,25-34],mc13[03-18],mc15[02,03,10-25,28-34]
 PriorityJobFactor=40 PriorityTier=40 RootOnly=NO ReqResv=NO OverSubscribe=FORCE:32
 OvertimeLimit=NONE PreemptMode=OFF
 State=UP TotalCPUs=3812 TotalNodes=103 SelectTypeParameters=NONE
 DefMemPerCPU=4000 MaxMemPerNode=UNLIMITED
```
### Aggregated utilization of UGE

**end-time was set to 2017-10-09 15:49, DB collName=uge**

<table>
<thead>
<tr>
<th>user \ timeRange</th>
<th>sum last 30_days</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-all-users : owner</td>
<td>158.4 (cpu*year), 905420 jobs, wallT fract=1.000, CPU/wallT=0.89</td>
</tr>
<tr>
<td>alice : project</td>
<td>76.9 (cpu*year), 176909 jobs, wallT fract=0.485, CPU/wallT=0.84</td>
</tr>
<tr>
<td>atlas : project</td>
<td>0.3 (cpu*year), 8412 jobs, wallT fract=0.002, CPU/wallT=0.88</td>
</tr>
<tr>
<td>dayabay : project</td>
<td>22.5 (cpu*year), 90437 jobs, wallT fract=0.142, CPU/wallT=0.91</td>
</tr>
<tr>
<td>dybspade : owner</td>
<td>3.5 (cpu*year), 15439 jobs, wallT fract=0.022, CPU/wallT=0.49</td>
</tr>
<tr>
<td>lux : project</td>
<td>2.4 (cpu*year), 13581 jobs, wallT fract=0.015, CPU/wallT=0.86</td>
</tr>
<tr>
<td>lz : project</td>
<td>1.4 (cpu*year), 46345 jobs, wallT fract=0.009, CPU/wallT=0.95</td>
</tr>
<tr>
<td>majorana : project</td>
<td>2.6 (cpu*year), 22407 jobs, wallT fract=0.017, CPU/wallT=0.98</td>
</tr>
<tr>
<td>star : project</td>
<td>51.7 (cpu*year), 544732 jobs, wallT fract=0.327, CPU/wallT=0.96</td>
</tr>
<tr>
<td>staremb : owner</td>
<td>0.1 (cpu*year), 90 jobs, wallT fract=0.000, CPU/wallT=1.00</td>
</tr>
</tbody>
</table>
SLURM CPU aggregated over month

SLURM: **completed** jobs

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

Work in progress by **Ershaad Basheer**

- 17k wall-h/day
- 11k wall-h/day
- 5.6k wall-h/day

30% is SLURM

**Summary for Jul 09, 2017 (12:00) to Oct 09, 2017 (15:00)**

<table>
<thead>
<tr>
<th>Series</th>
<th>Total</th>
<th>Total (%)</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>103,059</td>
<td>33.2%</td>
<td>140.0</td>
<td>679.9</td>
</tr>
<tr>
<td>root</td>
<td>0</td>
<td>0.0%</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>nstaff</td>
<td>36,498</td>
<td>11.8%</td>
<td>49.6</td>
<td>318.9</td>
</tr>
<tr>
<td>matcomp</td>
<td>12,450</td>
<td>4.0%</td>
<td>16.9</td>
<td>139.0</td>
</tr>
<tr>
<td>majorana</td>
<td>50,256</td>
<td>16.2%</td>
<td>68.3</td>
<td>317.0</td>
</tr>
<tr>
<td>lz</td>
<td>27,246</td>
<td>8.8%</td>
<td>37.0</td>
<td>196.3</td>
</tr>
<tr>
<td>lux</td>
<td>1,806</td>
<td>0.6%</td>
<td>2.5</td>
<td>34.5</td>
</tr>
<tr>
<td>dayabay</td>
<td>5</td>
<td>0.0%</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>cuore</td>
<td>3,922</td>
<td>1.3%</td>
<td>5.3</td>
<td>44.5</td>
</tr>
<tr>
<td>atlas</td>
<td>64,454</td>
<td>20.8%</td>
<td>87.6</td>
<td>335.0</td>
</tr>
<tr>
<td>alice</td>
<td>10,486</td>
<td>3.4%</td>
<td>14.2</td>
<td>81.5</td>
</tr>
<tr>
<td>Total</td>
<td>310,193</td>
<td>100.0%</td>
<td>420.9</td>
<td>977.9</td>
</tr>
</tbody>
</table>
aggregated load on PDSF interactive nodes

https://portal-auth.nersc.gov/pdsf-mon/
/project(a) utilization - snapshot

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

```
pdsf6 $ $ prjquota dayabay
          ------- Space (GB) -------          ------- Inode -------
Project      Usage      Quota    InDoubt          Usage       Quota     InDoubt
          -------          -------     -------          -------  -------     -------
dayabay     852177     870400          0      140670048   150000000           0
pdsf6 $ prjquota dayabay
          ------- Space (GB) -------          ------- Inode -------
Project      Usage      Quota    InDoubt          Usage       Quota     InDoubt
          -------          -------     -------          -------  -------     -------
dayabay     701393     870400          0        2304462    10000000           0
pdsf6 $ prjquota majorana
          ------- Space (GB) -------          ------- Inode -------
Project      Usage      Quota    InDoubt          Usage       Quota     InDoubt
          -------          -------     -------          -------  -------     -------
majorana    37194      40960          0        2475737     4000000           0
pdsf6 $ prjquota majorana
          ------- Space (GB) -------          ------- Inode -------
Project      Usage      Quota    InDoubt          Usage       Quota     InDoubt
          -------          -------     -------          -------  -------     -------
majorana    54625      61440          0        4188227    10000000           0
```

FillStatus (Quota): PROJECT (2017-10-06 12:10)

```
star - size
       64,990/90,000 TB (72.11%)

star - inodes
       1679425/2000000 (83.97%)

starprod - size
       120,033/130,000 TB (92.74%)

starprod - inodes
       9567718/2000000 (47.53%)

alice - size
       41,088/60,076 TB (68.33%)

alice - inodes
       17744425/25000000 (70.97%)

FillStatus (Quota): PROJECTA (2017-10-06 12:10)

starprod - size
       185,950/190,050 TB (97.71%)

starprod - inodes
       5281835/5000000 (26.40%)
```
XrootD availability - snapshot

XRD file count on PDSF nodes 2017-10-06_10.14
Announcements

Bi-weekly office hours  12:30 -2:30pm  
Thursday,  October 26, November 9,  59-4016-CR 

PDSF user meeting 
  ●  Tuesday, November 11, 11:15am, 59-3034-CR 

Outages : ?
STAR: at scale data analysis on Cori
The utilization was at least 97.6%, this is a lower limit due to the frequency that our pipeline collects stats. For two days, Cori was the largest data reconstruction farm ever known to STAR experiment, 2x current farm at BNL!
Planned Dayabay production on Cori

RabbitMQ
Server
15,000 sockets

P^2 submits 750,000
NuWa tasks

300 Haswell nodes

Job 1, N=30 done
Job 2, N=30
Job 11, N=30
Job 21, N=30
Job 10, N=30
Job 20, N=30
Job 30, N=30

One week of wall time
30-node SLURM job for Dayabay analysis

One SLURM job will use N=30 nodes and run for T=36 hours