PDSF User Meeting

- PDSF performance
- announcements
- SLURM @ PDSF
- AOB
aggregated load on PDSF interactive nodes

https://portal-auth.nersc.gov/pdsf-mon/
Jan Balewski, NERSC

CPU aggregated over month

UGE load ~½ after power dip, no Alice till today morning

SLURM is used

http://portal.nersc.gov/project/mpccc/balewski/pdsf3Load/latest/

One user job incident

Wilde fire shutdown

UGE+CHOS

SLURM+CHOS

SLURM+Shifter
queue(s) load

UGE - looks fine

SLURM - no queue limits
(no monitor vs. time)

<table>
<thead>
<tr>
<th>Partition</th>
<th>Nodes per Job</th>
<th>Physical Max Cores per Job</th>
<th>Walltime per Job</th>
</tr>
</thead>
<tbody>
<tr>
<td>shared-chos</td>
<td>1</td>
<td>1-32</td>
<td>2 days</td>
</tr>
<tr>
<td>shared</td>
<td>1</td>
<td>1-32</td>
<td>2 days</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partition</th>
<th>Total vCores [nt1]</th>
<th>Total Nodes</th>
<th>Default Mem per vCore (MB)</th>
<th>Max Mem per vCore (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>shared-chos</td>
<td>420</td>
<td>7</td>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>shared</td>
<td>176</td>
<td>4</td>
<td>1000</td>
<td>2000</td>
</tr>
</tbody>
</table>
# Aggregated utilization of UGE

## July average
Served 3.1k vCores  
Capacity 2.8k vCores  
CPU/WallT 0.87

<table>
<thead>
<tr>
<th>user \ timeRange</th>
<th>sum last 30_days</th>
</tr>
</thead>
</table>
| 0-all-users : owner | 228.3 (cpu*year) , 1300453 jobs  
wallIT frac=1.000, CPU/wallIT=0.87 |
| alice : project | 96.6 (cpu*year) , 101676 jobs  
wallIT frac=0.423, CPU/wallIT=0.84 |
| atlas : project | 23.2 (cpu*year) , 494232 jobs  
wallIT frac=0.102, CPU/wallIT=1.00 |
| dayabay : project | 20.0 (cpu*year) , 51732 jobs  
wallIT frac=0.087, CPU/wallIT=0.52 |
| dybprod : owner | 0.0 (cpu*year) , 0 jobs  
wallIT frac=0.000, CPU/wallIT=0.00 |
| hack : owner | 1.9 (cpu*year) , 16028 jobs  
wallIT frac=0.008, CPU/wallIT=0.93 |
| lux : project | 8.0 (cpu*year) , 29285 jobs  
wallIT frac=0.035, CPU/wallIT=0.91 |
| lz : project | 3.8 (cpu*year) , 4793 jobs  
wallIT frac=0.017, CPU/wallIT=1.00 |
| majorana : project | 4.1 (cpu*year) , 60960 jobs  
wallIT frac=0.018, CPU/wallIT=0.48 |
| star : project | 70.9 (cpu*year) , 551245 jobs  
wallIT frac=0.311, CPU/wallIT=0.95 |
| staremb : owner | 0.0 (cpu*year) , 0 jobs  
wallIT frac=0.000, CPU/wallIT=0.00 |

## June average
Served 3.1k vCores  
Capacity 3.4k vCores  
CPU/WallT 0.74

<table>
<thead>
<tr>
<th>user \ timeRange</th>
<th>sum last 30_days</th>
</tr>
</thead>
</table>
| 0-all-users : owner | 251.1 (cpu*year) , 1342121 jobs  
wallIT frac=1.000, CPU/wallIT=0.74 |
| alice : project | 113.8 (cpu*year) , 99351 jobs  
wallIT frac=0.453, CPU/wallIT=0.70 |
| atlas : project | 16.6 (cpu*year) , 386661 jobs  
wallIT frac=0.066, CPU/wallIT=0.93 |
| dayabay : project | 35.2 (cpu*year) , 105573 jobs  
wallIT frac=0.140, CPU/wallIT=0.36 |
| dybprod : owner | 0.0 (cpu*year) , 0 jobs  
wallIT frac=0.000, CPU/wallIT=0.00 |
| hack : owner | 3.3 (cpu*year) , 5687 jobs  
wallIT frac=0.013, CPU/wallIT=0.97 |
| lux : project | 3.5 (cpu*year) , 4195 jobs  
wallIT frac=0.014, CPU/wallIT=0.95 |
| lz : project | 2.3 (cpu*year) , 181794 jobs  
wallIT frac=0.009, CPU/wallIT=0.92 |
| majorana : project | 9.4 (cpu*year) , 141214 jobs  
wallIT frac=0.037, CPU/wallIT=0.74 |
| star : project | 69.8 (cpu*year) , 415768 jobs  
wallIT frac=0.278, CPU/wallIT=0.92 |
| staremb : owner | 0.0 (cpu*year) , 0 jobs  
wallIT frac=0.000, CPU/wallIT=0.00 |
Mendel utilization - snapshot

Load on UGE worker nodes  2017-08-08_09.01

UGE low load ~\(\frac{1}{3}\)
no Alice

SLURM nodes
Mendel RAM/task usage - snapshot

RAM/job, rack colors: mc01=r, 12=b, 13=g, 15=y  2017-08-08_09.01

<table>
<thead>
<tr>
<th>r</th>
<th>qw</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>1242</td>
<td>0</td>
<td>alice</td>
</tr>
<tr>
<td>159</td>
<td>1080</td>
<td>atlas</td>
</tr>
<tr>
<td>58</td>
<td>417</td>
<td>lux</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>majorana</td>
</tr>
<tr>
<td>1307</td>
<td>5357</td>
<td>star</td>
</tr>
<tr>
<td>2768</td>
<td>6854</td>
<td>Totals</td>
</tr>
</tbody>
</table>
### /project(a) utilization - snapshot

[http://portal.nersc.gov/project/star/jthaeder/diskUsage/overview/indexExt.html](http://portal.nersc.gov/project/star/jthaeder/diskUsage/overview/indexExt.html)

[https://my.nersc.gov/data-mgt.php](https://my.nersc.gov/data-mgt.php)

```bash
pdsf6 $ prjquota dayabay
---------- Space (GB) ------- Inode ----------
Project    Usage  Quota  InDoubt    Usage  Quota  InDoubt
----------   ---------  ---------  ---------     ----------  ----------  ----------
dayabay    794759   870400   69       135478754 150000000   558
```

```bash
pdsf6 $ prjaquota dayabay
---------- Space (GB) ------- Inode ----------
Project    Usage  Quota  InDoubt    Usage  Quota  InDoubt
----------   ---------  ---------  ---------     ----------  ----------  ----------
dayabay    706679   716800   0        2292741    10000000     0
```

```bash
pdsf6 $ prjquota majorana
---------- Space (GB) ------- Inode ----------
Project    Usage  Quota  InDoubt    Usage  Quota  InDoubt
----------   ---------  ---------  ---------     ----------  ----------  ----------
majorana   38869    40960    0        2345664    4000000     0
```

```bash
pdsf6 $ prjaquota majorana
---------- Space (GB) ------- Inode ----------
Project    Usage  Quota  InDoubt    Usage  Quota  InDoubt
----------   ---------  ---------  ---------     ----------  ----------  ----------
majorana   57901    61440    9        4445767    10000000   495
```
Set date for SLURM at PDSF

Please ask your users to test SLURM+CHOS at scale.
UNIVA vs. SLURM tutorials

Few good tutorials:

- SGE vs SLURM comparison - Uppsala Multidisciplinary Center for...
  [Link](http://www.uppmax uu.se/support/user-guides/sge-vs-slurm-comparison/)

- SGE to SLURM conversion | Stanford Research Computing Center
  [Link](https://srcr.stanford.edu/sge-slurm-conversion)

- SGE vs Slurm
  [Link](http://www.sdsc.edu/~hocks/FG/MSKCC.slurm.sge.html)

- SGE vs SLURM comparison. A guide comparing common commands in slurm and sge. Some common commands and flags in slurm and sge...

- Getting started with SLURM on the Sherlock pages.

- Univa Grid Engine development is led by CTO Fritz Ferstl, who founded the Grid Engine project and ran the business... Comparison SGE vs Slurm comparison.
SLURM @ PDSF

https://www.nersc.gov/users/computational-systems/pdsf/using-slurm-pdsf-batch-sbatch/
Alternative: https://bitbucket.org/balewski/tutornersc/src/master/2017-05-pdsf3.0/
# SLURM - outstanding high priority issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Priority</th>
<th>Imparies</th>
<th>Experiment</th>
<th>status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ganglia</td>
<td>2</td>
<td>SLURM</td>
<td>all</td>
<td>Known cause</td>
<td>Jobs running under SLURM are not reported on ganglia PDSF monitor page, ticket INC0103269</td>
</tr>
<tr>
<td>Ignored swap space</td>
<td>2</td>
<td>SLURM</td>
<td>all</td>
<td>discussed</td>
<td>SLURM is not seeing swap space on nodes, either partition (chos/shifter) INC0103391</td>
</tr>
<tr>
<td>OSG2</td>
<td>2</td>
<td>SLURM+grid jobs</td>
<td>ALICE, some STAR</td>
<td>Under eval.</td>
<td>deployment of 2nd OSG talking to SLURM on node mpdfsgrid02, will look like ...01. Jeff &amp; James</td>
</tr>
<tr>
<td>AliEn VOBX2</td>
<td>2</td>
<td>SLURM</td>
<td>ALICE</td>
<td>In progress</td>
<td>grid jobs submission, monitoring. Need 2nd system for SLURM</td>
</tr>
<tr>
<td>Bad tcsh</td>
<td>3</td>
<td>Shifter</td>
<td>STAR</td>
<td>Under eval.</td>
<td>SLURM + Shifter is not working on PDSF for users who have the default NIM shell set to tcsh, ticket INC0101446</td>
</tr>
<tr>
<td>Smooth image selection</td>
<td>3</td>
<td>shifter</td>
<td>all</td>
<td>Under eval.</td>
<td>A user logs into PDSF and their shell is run inside the preselected Shifter image without any action on their part.</td>
</tr>
</tbody>
</table>
Multi-thread jobs on PDSF

When run code be aware if it is multi-thread
documentation for the package it looks like the default is
"maximum number of threads available if not set"

Not submit as 1 core UGE tasks
- MadGraph
- Nix
- Machine Learning: BDTs, tflearn

Solution in SLURM
#submit one task running on 32 vCores
sbatch -p shared-chos -t 24:00:00 -n32 jobscript.sh
Announcements

Bi-weekly office hours  12:30 -2:30pm
Thursday, August 17, 31, 59-4016-CR

PDSF user meeting
  ●  Tuesday, September 12, 11:15am, 59-3034-CR

Outages :
Cori maintenance: today
  ●  There will be a training session on SLURM & Cori on 8/3, 1:00 - 3:00 pm in 149C. Zoom will be available for remote participants. Please sign up if you plan to attend.