PDSF User Meeting

- PDSF performance
- Announcements
- AOB
Summary of Feb 12, SC-Meeting

PDSF/Mendel retirement schedule

Instead of retiring components piecemeal, we propose a simpler model: keep all hardware DNR till agreed date, then shut down all at once.

While we considered scenarios to retire individual racks at different times based on support contracts, with service nodes distributed all over the racks move of all these nodes was going to cause NERSC staff and PDSF experts significant effort and PDSF users service interruptions.

Assume services running on Mendel end April 2019.

Contingency period of 5 months to Aug. 2019
SLURM CPU*h 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
## Existing Slurm Shifter queues

<table>
<thead>
<tr>
<th>partition</th>
<th>OS provider</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>3352</td>
<td>2 days</td>
<td></td>
<td>250</td>
<td>0</td>
<td>Share 94% of hardware</td>
</tr>
<tr>
<td>alice</td>
<td>chos</td>
<td>3224</td>
<td>2 days</td>
<td></td>
<td>1500</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>realtime-chos</td>
<td>chos</td>
<td>128</td>
<td>4 hours</td>
<td></td>
<td>50</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>debug-chos</td>
<td>chos</td>
<td>128</td>
<td>30 min</td>
<td></td>
<td>2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>long</td>
<td>shifter</td>
<td>240</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>nucori</td>
<td>shifter</td>
<td>144</td>
<td>2 days</td>
<td></td>
<td>0</td>
<td></td>
<td>Cori-like OS</td>
</tr>
<tr>
<td>realtime</td>
<td>shifter</td>
<td>128</td>
<td>4 hours</td>
<td></td>
<td>50</td>
<td>0</td>
<td>share common hardware</td>
</tr>
<tr>
<td>debug</td>
<td>shifter</td>
<td>128</td>
<td>30 min</td>
<td></td>
<td>2</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>
The following changes were enforced on Slurm all queues:

- Use true RAM per node
- compute high mem use tax: nCPU= mem/4 GB
- Num job slots per node: 2x phys cores, but …
- Each job locks 1 physical core by default but is available by using --oversubscribe
- Default mem per task: 2.5 GB (was 4, changed on June 13)
- Used cpu*h half decay time: 4 days (was 14 days)
- Cap of 250 jobs/user is enforced, pseudo-user alicesgm is an exception
- --mem → cpus conversion is 2.5 GB for all partitions, except -p long, -p short use 5 GB
### /project(a) utilization - snapshot

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

2019-01-07_20.30 project(a) usage for PDSF users, ver 1.1

<table>
<thead>
<tr>
<th>space</th>
<th>usage</th>
<th>quota</th>
<th>percent</th>
<th>usage</th>
<th>quota</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>alice</td>
<td>62161</td>
<td>62464</td>
<td>99</td>
<td>21682016</td>
<td>25000000</td>
<td>86</td>
</tr>
<tr>
<td>projecta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>69340</td>
<td>71680</td>
<td>96</td>
<td>22172104</td>
<td>25000000</td>
<td>88</td>
</tr>
<tr>
<td>projecta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dayabay</td>
<td>849694</td>
<td>870400</td>
<td>97</td>
<td>126734877</td>
<td>150000000</td>
<td>84</td>
</tr>
<tr>
<td>projecta</td>
<td>966753</td>
<td>1126400</td>
<td>85</td>
<td>6556733</td>
<td>100000000</td>
<td>65</td>
</tr>
<tr>
<td>majorana</td>
<td>48853</td>
<td>61440</td>
<td>79</td>
<td>4042829</td>
<td>60000000</td>
<td>67</td>
</tr>
<tr>
<td>projecta</td>
<td>55377</td>
<td>61440</td>
<td>90</td>
<td>8699426</td>
<td>10000000</td>
<td>87</td>
</tr>
<tr>
<td>atlas</td>
<td>80201</td>
<td>102400</td>
<td>78</td>
<td>28641579</td>
<td>400000000</td>
<td>71</td>
</tr>
<tr>
<td>projecta</td>
<td>188583</td>
<td>256000</td>
<td>73</td>
<td>28020048</td>
<td>400000000</td>
<td>70</td>
</tr>
<tr>
<td>lz</td>
<td>35038</td>
<td>40960</td>
<td>85</td>
<td>3798862</td>
<td>400000000</td>
<td>9</td>
</tr>
<tr>
<td>projecta</td>
<td>233760</td>
<td>256000</td>
<td>91</td>
<td>4759693</td>
<td>400000000</td>
<td>11</td>
</tr>
<tr>
<td>lux</td>
<td>34828</td>
<td>40960</td>
<td>85</td>
<td>6899473</td>
<td>80000000</td>
<td>86</td>
</tr>
<tr>
<td>projecta</td>
<td>217206</td>
<td>245760</td>
<td>88</td>
<td>62241056</td>
<td>700000000</td>
<td>88</td>
</tr>
<tr>
<td>cuore</td>
<td>29549</td>
<td>307200</td>
<td>96</td>
<td>2197717</td>
<td>60000000</td>
<td>36</td>
</tr>
<tr>
<td>projecta</td>
<td>36786</td>
<td>512000</td>
<td>71</td>
<td>267581</td>
<td>600000000</td>
<td>4</td>
</tr>
</tbody>
</table>
Announcements

Bi-weekly office hours  Jan 21, Feb 4, 59-4016A

PDSF user meeting: Tuesday, Feb 12

Obligatory MFA in January of 2019 (starts today)

New charge factors: Haswell=KNL=90, Edison=64
MFA help

MFA instruction:
http://www.nersc.gov/users/connecting-to-nersc/mfa/

**Virtual office hours** from 7 am to 4 pm (Pacific) on the following dates:

* Tuesday, January 8, 2019
* Wednesday, January 9, 2019

Please feel free to stop by via Zoom at https://lbnl.zoom.us/j/752299719.

Regards,
-Rebecca

**Updates**
Rebecca: must sign new User Agreement @ NIM as of today
Jan: install Google Authenticator on your smartphone before initializing MFA
Testing -p nucori

-p nucori is new partition with bare-bone OS like Cori worker nodes
You must specify and use shifter image of your choice to do meaningful computations.

Example:
runBatch.slr - defines Slurm params and starts shifter, it calls
inShifter.sh - it runs your analysis

Execute:
$sbatch runBatch.slr 101 766

$pdsf8 $ cat runBatch.slr
#!/bin/bash -l
#SBATCH --partition=nucori  --account=lz  --image=custom:pdsf-chos-sl64:v4
echo  in-base-OS
echo inShifter:`env|grep  SHIFTER_RUNTIME`
cat /etc/*release

shifter --module=cvmfs  --volume=/global/project:/project ./scripts/inShifter.sh $1 xx77 $2

$ cat inShifter.sh
#!/bin/bash
  echo  task-in-shifter
  echo inShifter:`env|grep  SHIFTER_RUNTIME`
cat /etc/*release
echo  Load environment variables and run the code
EXE=LZStats
  . ${PWD}/setup.sh
time ./${EXE} $1 $2 $3