

# Troubleshooting Jobs on Odyssey

Paul Edmon, PhD  
ITC Research Computing Associate

Bob Freeman, PhD  
Research & Education Facilitator  
XSEDE Campus Champion

- Tackle PEND, FAIL, and slow performance issues
- Highlight different approaches to troubleshooting
- Arm you with useful SLURM and Unix commands
- Enable you to “Work smarter, better, faster”



# PEND Problems

## Why are you pending?

```
queue -u USERNAME -t PD
```

```
queue -u USERNAME -t PD -o "%.18i %.9P %.8j %.8u %.2t %.10M %.6D %R %C"
```

```
[bfreeman@rclogin07 ~]$ queue -t PD | tail -n 20
 25793703_1 serial_re 1MUI-A_P ncheron PD 0:00 1 (Priority) 20
 25793729_4 serial_re 1MUI-B_P ncheron PD 0:00 1 (Priority) 20
 25793813_1 serial_re 10HR_Pro ncheron PD 0:00 1 (None) 20
 25793936_[9,17] serial_re 2AQU-B_P ncheron PD 0:00 1 (None) 20
25820727_[7,17,25] serial_re CHEMBL18 ncheron PD 0:00 1 (Resources) 16
 25793863_6 serial_re 2AQU-A_P ncheron PD 0:00 1 (Priority) 20
 25794027_18 serial_re 4HLA-A_P ncheron PD 0:00 1 (Priority) 20
25877428_[47-50] shakgpu Banjanac ncheron PD 0:00 1 (Resources) 32
 25877480_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877531_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877583_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877635_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877688_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877742_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25877794_[1-50] shakgpu Banjanac ncheron PD 0:00 1 (Priority) 32
 25885657_50 shaknovi 1HXW_Pro ncheron PD 0:00 1 (Resources) 20
25902404_[29-50] unrestric OMAx_Myr rmfernan PD 0:00 1 (Resources) 1
 25889308 vogelsber p3_30_13 ptorrey PD 0:00 1 (Resources) 64
 25890147 vogelsber p3_100_1 ptorrey PD 0:00 1 (Priority) 64
 25900499 vogelsber Aq-H-5_r kannan PD 0:00 2 (Dependency) 128
```



# PEND Problems

Why are you pending? Most common reasons are...

<i>Reason</i>	<i>Solution</i>
None	Scheduler hasn't gotten around to you yet. Hang tight...
Resources	The job is waiting for resources to become available
Priority	One or more higher priority jobs exist for this partition or reservation. Priority is influenced by age, job size, partition, QoS, and Fairshare.
Dependency	This job is waiting for a dependent job to complete (--dependency)



# PEND Problems: Resources

- What did you request?
  - What are the parameters of your SLURM submission script?
  - `scontrol show jobid -dd JOBID`

```
[bfreeman@rclogin07 ~]$ scontrol show jobid -dd 25889308
JobId=25889308 Name=p3_30_13
  UserId=ptorrey(34966) GroupId=hernquist_lab(33234)
  Priority=100000024 Nice=0 Account=hernquist_lab QOS=normal
  JobState=PENDING Reason=Resources Dependency=(null)
  Requeue=1 Restarts=0 BatchFlag=1 ExitCode=0:0
  DerivedExitCode=0:0
  RunTime=00:00:00 TimeLimit=7-00:00:00 TimeMin=N/A
  SubmitTime=2014-12-02T11:55:16 EligibleTime=2014-12-02T11:55:16
  StartTime=2014-12-03T08:47:21 EndTime=Unknown
  PreemptTime=None SuspendTime=None SecsPreSuspend=0
  Partition=vogelsberger AllocNode:Sid=holy2a24208:48180
  ReqNodeList=(null) ExcNodeList=(null)
  NodeList=(null)
  NumNodes=1-1 NumCPUs=64 CPUs/Task=1 ReqB:S:C:T=0:0:*:*
  Socks/Node=0 MinMem=0 MinSvC=0:0:*:* CoreSpec=0
  MinCPUsNode=1 MinMemoryNode=250000M MinTmpDiskNode=0
  Features=(null) Gres=(null) Reservation=(null)
  Shared=0 Contiguous=0 Licenses=(null) Network=(null)
  Command=/n/home01/ptorrey/Run/NuclearDisks/GIZMO_PSPH/10b-1e1C-1e
```

# PEND Problems: Priority

What is your priority in the scheduling queue?

- `showq-slurm -U`
- `showq-slurm -p PARTITION -o`

```
WAITING JOBS-----
JOBID      JOBNAME      USERNAME      STATE      CORE      WCLIMIT      QUEUETIME
=====
25912197   4002_b3lyp  hirzel        Waiting 8      75:00:00    Tue Dec 2 15:29:48
25911730   4027_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:26:41
25911735   4028_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:26:43
25911739   4029_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:26:44
25911743   4030_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:26:46
25911746   4031_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:26:48
25911817   4042_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:27:16
25911826   4043_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:27:18
25911828   4044_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:27:19
25911834   4045_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:27:22
25911838   4046_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 15:27:23
25912198   run_getSco  inun         Waiting 1      8:20:00     Tue Dec 2 15:29:48
25820732   CHEMBL1818 ncheron      Waiting 16     48:00:00    Tue Dec 2 05:15:18
25820733   CHEMBL1818 ncheron      Waiting 16     48:00:00    Tue Dec 2 05:18:08
25820736   CHEMBL1818 ncheron      Waiting 16     48:00:00    Tue Dec 2 05:15:21
25820737   CHEMBL1818 ncheron      Waiting 16     48:00:00    Tue Dec 2 05:17:49
25820738   CHEMBL1818 ncheron      Waiting 16     48:00:00    Tue Dec 2 05:17:48
25879921   4226_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 10:30:37
25879923   4227_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 10:30:40
25879925   4227_b3lyp  rgbombarelli  Waiting 8      75:00:00    Tue Dec 2 10:30:42
```



# PEND Problems: Fair-share

- *What is a fair-share score?*

Score assigned to *each lab* that affects the scheduling priority of jobs. A user's current and past usage is considered when determining the scheduling of job execution.

- *How is it calculated?*

Score is based on usage and shares:

$0 \leq Usage \leq 1$ , represents your proportional use of Odyssey  
*Shares*  $\sim$  slices of a pie, or the part of Odyssey that is yours

Premise: Usage == Shares, you've hit your fairshare target

$$\text{Fairshare Factor} = 2^{-Usage/Shares}$$

$$0 < FF < 1$$

When usage increases, FF decreases

When usage decreases, FF increases



# PEND Problems: Fair-share

- What is my score?

```
sshare -u USERNAME
```

```
[bfreeman@rclogin07 ~]$ sshare -u bfreeman
```

Account	User	Raw Shares	Norm Shares	Raw Usage	Effecty Usage	FairShare
root			1.000000	7537869166	1.000000	0.500000
ac275		100	0.003012	0	0.000000	1.000000
acmg		100	0.003012	0	0.000000	1.000000
adams_lab_seas		100	0.003012	2384	0.000000	0.999927
aidala_lab		100	0.003012	9645	0.000001	0.999705
airoldi_lab		100	0.003012	6173473	0.000820	0.828016
aizenberg_lab		100	0.003012	12	0.000000	1.000000
amir_lab		100	0.003012	0	0.000000	1.000000
anderson_lab		100	0.003012	188069	0.000025	0.994267
anl		100	0.003012	0	0.000000	1.000000
arlotta_lab		100	0.003012	478	0.000000	0.999985
aspuru-guzik_lab		1000	0.030120	1049842986	0.139265	0.040566
ay215		100	0.003012	0	0.000000	1.000000
aziz_lab		100	0.003012	0	0.000000	1.000000
baccarelli_lab		100	0.003012	193902	0.000026	0.994090
balskus_lab		100	0.003012	6316	0.000001	0.999807
berger_lab		100	0.003012	0	0.000000	1.000000
bortoldi_lab		100	0.003012	7748577	0.001017	0.701765

```
[bfreeman@rclogin07 ~]$ sshare -u bfreeman | grep bfreeman
```

rc_admin	bfreeman	parent	0.003012	20442	0.000014	0.996721
----------	----------	--------	----------	-------	----------	----------



# FAIL Problems

- Determining the root cause is highly dependent on your ability to trace & document the problem
- *Do you have log/error files?*  

```
#SBATCH -o job.stdout.txt  
#SBATCH -e job.stderr.txt
```
- *Did you use the -e parameter to get an error file?*  
If omitted, `STDERR` is redirected to `-o` file or `slurm-JOBID.out`
- *Is the path non-existent?*  
The path to place the output files must exist! Otherwise, SLURM won't know what to do and will do nothing but FAIL.
- *What else is going on in your SLURM submission script?*



# FAIL Problems

- Look at the fail trail in any log and/or error files.
- *Nota bene!* The last error listed may not be the root cause!

```
succeeded(260396), failed(5) 63.1007% completed. #
# There is insufficient memory for the Java Runtime Environment to continue.
# Native memory allocation (malloc) failed to allocate 1048576 bytes for AllocateHeap
# An error report file with more information is saved as:
#
/n/regal/.../BclToFastq_Lane2_Indexlength6_Run1/Project_.../
Trinity4_output/hs_err_pid16973.log
```

```
succeeded(260396), failed(6) 63.101% completed.
succeeded(260397), failed(6) 63.1012% completed. Error occurred during initialization of VM
Unable to allocate bit map for parallel garbage collection for the requested heap size.
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
[...]
```

```
Error, cmd: /n/sw/centos6/trinityrnaseq_r20140413/trinity-plugins/parafly/bin/ParaFly -c
/n/regal/.../BclToFastq_Lane2_Indexlength6_Run1/Project_.../
Trinity4_output/chrysalis/butterfly_commands -shuffle -CPU 8 -failed_cmds failed_butterfly_commands.11517.txt -v
died with ret 256 at /n/sw/centos6/trinityrnaseq_r20140413/Trinity line 1958.
```

This is an example Butterfly command:

```
java -Xmx10G -Xms1G -jar /n/sw/centos6/trinityrnaseq_r20140413/Butterfly/Butterfly.jar -N 100000 -L 200 -F 500 -C
/n/regal/.../BclToFastq_Lane2_Indexlength6_Run1/Project_...//
Trinity4_output/chrysalis/Component_bins/Cbin154/c167101.graph --path_reinforcement_distance=75 --triplet-lock
--EM_REDUCE
```

# FAIL Problems

segmentation faults

shared/static library errors

- Are the correct software packages and versions loaded?
- Are there version conflicts?
- Are too many things loaded in your `.bashrc`?

```
Loading module hpc/java-1.7.0_13.
Loading module hpc/perl-5.16.0.
Loading module hpc/intel-mkl-11.0.0.079.
Loading module centos6/tcl-8.5.14.
Loading module centos6/tk-8.5.14.
Loading module centos6/fftw-3.3_gcc-4.4.7.
Loading module centos6/gsl-1.16_gcc-4.4.7.
Loading module centos6/hdf5-1.8.11_gcc-4.4.7.
Loading module centos6/netcdf-4.3.0_gcc-4.4.7.
Loading module centos6/R-3.1.1.
Loading module centos6/python-2.7.3.
Loading module centos6/biopython-1.62b_python-2.7.3.
Loading module centos6/numpy-1.7.1_python-2.7.3.
Loading module centos6/gtk+-2.24.4.
Loading module centos6/cairo-1.12.14.
Loading module centos6/py2cairo-1.10.0_python-2.7.3.
Loading module centos6/matplotlib-1.3.1_python-2.7.3_gtk.
Loading module centos6/bx-python-3-14-2014_python-2.7.3.
Loading module hpc/glib-2.20.4.
Loading module hpc/pixman-0.17.4.
Loading module hpc/cairo-1.8.8.
Loading module hpc/pango-1.24.5.
Loading module hpc/gtk-1.20.0.
Loading module hpc/gtk+-2.16.5.
Loading module hpc/python-2.7.3_glib.
Loading module hpc/networkx-1.7_python-2.7.3.
Loading module centos6/python-2.7.3.
Loading module centos6/hdf5-1.8.11_gcc-4.4.7.
Loading module centos6/numpy-1.7.1_python-2.7.3.
Loading module centos6/scipy-0.12.0_python-2.7.3.
Loading module centos6/gtk+-2.24.4.
Loading module centos6/cairo-1.12.14.
Loading module centos6/py2cairo-1.10.0_python-2.7.3.
Loading module centos6/matplotlib-1.3.1_python-2.7.3_gtk.
Loading module centos6/cython-0.19.1_python-2.7.3.
Loading module centos6/python-2.7_modules.
Loading module centos6/pandas-0.11.0_python-2.7.3.
Loading module bio/primer3-2.2.2-beta.
```



# Poor performance

- Much harder to diagnose!
- Transient issues:
  - other jobs on the node
  - network
  - storage
  - bad node
- Software install issue
- Code compile/optimization problem

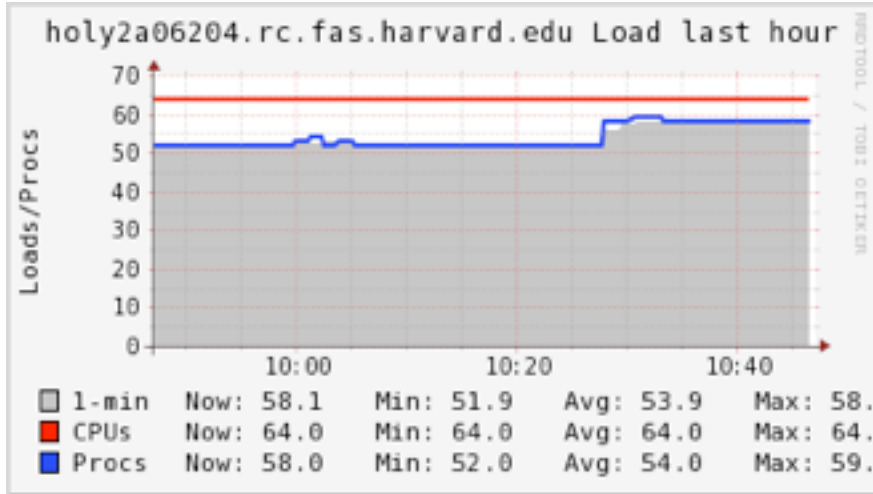
# Poor Performance

Slow run times? Most common reasons are...

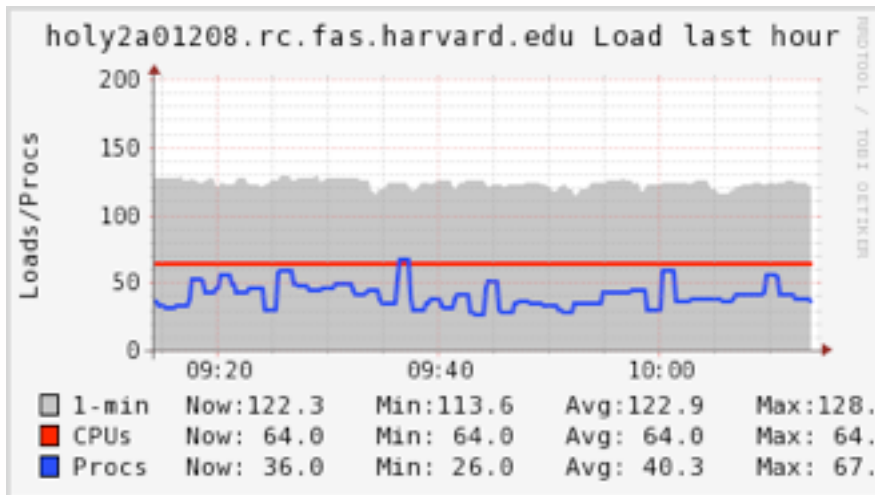
<i>Reason</i>	<i>Diagnostics &amp; Potential solutions</i>
SLURM submission errors	Check --mem/--mem-per-cpu Check cores (-n) + nodes (-N) Try running job interactively
Overloaded Storage	Is the mount point broken? Is the storage being hammered? Don't run out of home/lab directories. Is one file being accessed by hundreds of jobs? Make copies. Too many files in one directory? Check status.rc
Overloaded Node	Too much Input/Output for the given network interface Overuse of CPUs relative to SLURM request Check status.rc; use squeue and top
Sick Node	Memory errors /scratch filled up use lsload to look at node diagnostics
Sick Code	Monitor/trace program execution with strace Look at code performance with perf

# Poor performance: Nagios Compute

Check the state of the systems at <https://status.rc.fas.harvard.edu>



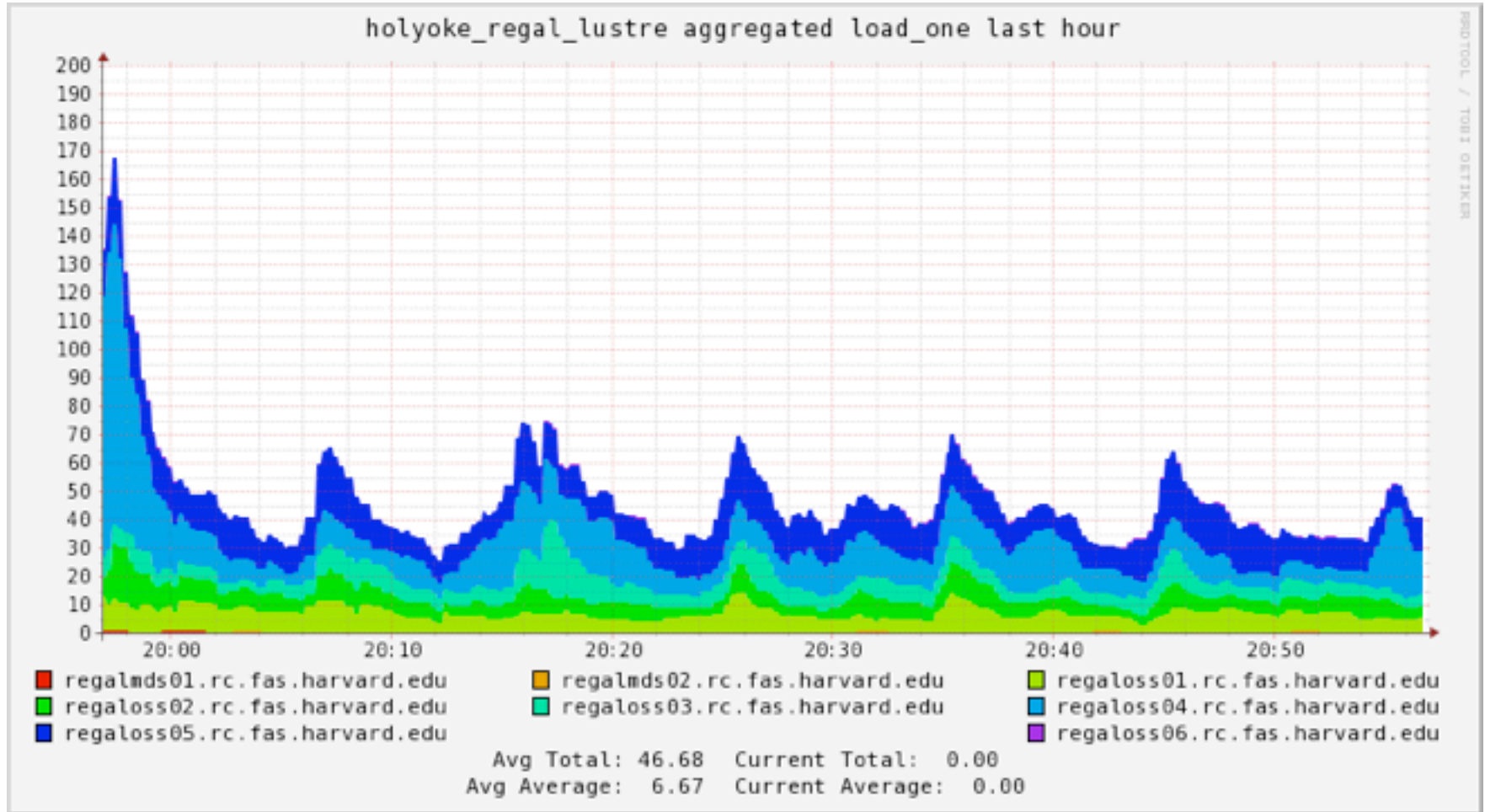
Heavy load, but normal



Heavy load, oversubscribed

# Poor performance: Nagios Storage

Includes storage like /n/regal and /n/holyscratch



# Poor performance: Node health

Check the health of the compute node (abused by other jobs?)

- need to figure out which compute node your job is on
  - queue -j JOBID
- quick check using <https://status.rc.fas.harvard.edu>
- look at load using `lsload`
- look at jobs on node using `queue`
- log in to node and profile
  - `top`
  - `strace`
  - `perftrace`





# Poor performance: Node health

Use `lsload` to examine the utilization of individual nodes

```
[bfreenan@rclogin05 ~]$ lsload
Hostname      Cores  InUse  Ratio  Load  Mem  Alloc  State
aag01         64    64   100.0  63.68  258  208   ALLOCATED
aag02         64    64   100.0  64.20  258  184   ALLOCATED
aag03         64    56   87.5   58.89  258  120   MIXED+COMPLETING
aag04         64    56   87.5   62.11  258  172   MIXED+COMPLETING

-----
holy2a01201   64    64   100.0  64.00  264  181   ALLOCATED
holy2a01202   64    64   100.0  54.34  264  250   ALLOCATED
holy2a01203   64    62   96.9   63.45  264  210   MIXED
holy2a01204   64     0    0.0    0.20  264   0   IDLE+DRAIN      /var/local/mem_error/holy2a01204_2
014-11-23_mem_error exists [root@2014-12-02T20:24:01]      ERROR Health check failed: Memory Check failed!
holy2a01205   64    64   100.0  49.60  264  132   ALLOCATED
holy2a01206   64    64   100.0  64.14  264  192   ALLOCATED
holy2a01207   64    64   100.0  64.32  264  186   ALLOCATED
holy2a01208   64    32   50.0   95.49  264  210   MIXED
holy2a02101   64    25   39.1   18.07  264   55   MIXED
holy2a02102   64    23   35.9   19.13  264   96   MIXED
holy2a02103   64    51   79.7   30.10  264  119   MIXED
holy2a02104   64    61   95.3   55.40  264  150   MIXED
holy2a02105   64    27   42.2   19.09  264   15   MIXED+DRAIN     Closed to replace DIMM - MRE 12/1/
2014 [methier@2014-12-01T12:49:20]
holy2a02106   64    47   73.4   48.92  264  124   MIXED
holy2a02107   64    45   70.3   45.33  264   71   MIXED
holy2a02108   64    56   87.5   56.90  264  201   MIXED
holy2a02201   64    64   100.0  66.14  264  114   ALLOCATED
holy2a02202   64    64   100.0  66.34  264  256   ALLOCATED
holy2a02203   64    61   95.3   56.34  264  111   MIXED
holy2a02204   64    34   53.1   33.79  264   21   MIXED
holy2a02205   64    58   90.6   50.30  264  162   MIXED
```



# Poor performance

What other jobs are running on my node?

```
queue -w NODENAME
```

```
[bfreeman@rclogin05 ~]$ queue -w holy2a18207
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)	CPUS
25901418	unrestric	tbsLNI3_	mtaylor	R	7:30:24	1	holy2a18207	8
25901380	unrestric	tbsLNI3_	mtaylor	R	7:30:53	1	holy2a18207	8
25902404_28	unrestric	OMAx_Myr	rmfernan	R	7:20:28	1	holy2a18207	1
25902404_9	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_10	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_11	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_12	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_13	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_14	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1
25902404_15	unrestric	OMAx_Myr	rmfernan	R	7:21:32	1	holy2a18207	1

e.g. Ensure that the # of CPUs requested = # CPUs used

# Poor performance

Log in to node and top

```
ssh USERNAME@NODENAME
```

```
top
```

```
top - 10:16:12 up 57 days, 22:49, 1 user, load average: 123.95, 123.25, 122.67
Tasks: 1410 total, 25 running, 1384 sleeping, 0 stopped, 1 zombie
Cpu(s): 57.7%us, 5.4%sy, 0.0%ni, 36.9%id, 0.0%wa, 0.0%hi, 0.0%si, 0.0%st
Mem: 264498560k total, 116480296k used, 148018264k free, 404116k buffers
Swap: 8388600k total, 1172k used, 8387428k free, 38103800k cached
```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
11608		20	0	71.2g	47g	1064	S	1654.2	19.0	16617:17	segemehl.x
10613		20	0	5427m	5.1g	8208	R	100.0	2.0	266:24.42	R
10639		20	0	5770m	5.4g	8192	R	100.0	2.1	267:51.29	R
10651		20	0	5866m	5.5g	8224	R	100.0	2.2	267:17.18	R
19870		20	0	389m	259m	5644	R	100.5	0.1	118:21.74	ko
19871		20	0	316m	187m	6276	R	100.5	0.1	118:26.03	ko
19872		20	0	317m	188m	6296	R	100.5	0.1	118:24.72	ko
19873		20	0	316m	186m	5532	R	100.5	0.1	118:25.13	ko
19874		20	0	316m	186m	6272	R	100.5	0.1	118:24.53	ko
19875		20	0	316m	187m	6936	R	100.5	0.1	118:25.58	ko
19876		20	0	316m	188m	6916	R	100.5	0.1	118:25.19	ko
19877		20	0	316m	187m	6092	R	100.5	0.1	118:25.35	ko
19878		20	0	316m	185m	6264	R	100.5	0.1	118:25.68	ko
19879		20	0	316m	188m	6944	R	100.5	0.1	118:25.67	ko



# Poor performance

Log in to node and monitor code execution

```
ssh USERNAME@NODENAME
```

```
ps aux | grep USERNAME
```

(obtain relevant process ID)

```
strace -p PROCESSID
```

```
Odelia:~ bfreeman$ ssh holy2a01208
Last login: Wed Dec  3 10:15:57 2014 from 10.255.12.24
[bfreeman@holy2a01208 ~]$ strace -p 29175
Process 29175 attached - interrupt to quit
select(1, [0], NULL, NULL, {2, 430607}) = 0 (Timeout)
fcntl(0, F_SETFL, O_RDWR|O_NONBLOCK|O_LARGEFILE) = 0
read(0, 0x7fffe977125c, 1)           = -1 EAGAIN (Resource temporarily unavailable)
ioctl(0, TCFLSH, 0)                  = 0
fcntl(0, F_SETFL, O_RDWR|O_LARGEFILE) = 0
open("/proc", O_RDONLY|O_NONBLOCK|O_DIRECTORY|O_CLOEXEC) = 8
getdents(8, /* 1338 entries */, 32768) = 32760
stat("/proc/1", {st_node=S_IFDIR|0555, st_size=0, ...}) = 0
open("/proc/1/stat", O_RDONLY)        = 9
read(9, "1 (init) S 0 1 1 0 -1 4202752 28"... , 1023) = 205
close(9)                               = 0
open("/proc/1/statm", O_RDONLY)       = 9
read(9, "5359 394 316 35 0 103 0\n", 1023) = 24
close(9)                               = 0
stat("/proc/2", {st_node=S_IFDIR|0555, st_size=0, ...}) = 0
open("/proc/2/stat", O_RDONLY)        = 9
read(9, "2 (kthreadd) S 0 0 0 0 -1 214961"... , 1023) = 158
close(9)                               = 0
```



# Research Computing

Please talk to your peers, and  
We wish you success in your research!

<http://rc.fas.harvard.edu>

<https://portal.rc.fas.harvard.edu>

[rchelp@fas.harvard.edu](mailto:rchelp@fas.harvard.edu)

@fasrc

Harvard Informatics

@harvardifx