



Memorial Sloan Kettering  
Cancer Center

# HPC user group Luna and Juno clusters

Sep 27 2018

# Agenda

- The Juno cluster (successor to the Luna cluster)
- Why is Juno replacing Luna?
- Overview of LSF configuration changes, particularly differences from Luna
- How to execute jobs on Juno (LSF)
- Software available on Juno
- How to get help on Luna/Juno?
- July network upgrade
- Documentation wiki
- Q&A



# Luna computational resources as of September 27, 2018

name	#	model	CPU	cores	RAM	Interc	NVMe
s01-24	24	HPE DL160G8	2Xeon(R)2.20GHz	16	384	10GB	
t01-02	2	HPE DL580G8	4Xeon(R)2.00GHz	32	1536	10GB	
u01-36	36	HPE DL160G9	2Xeon(R)2.60GHz	16	256	10GB	
w01	1	HPE DL160G9	2Xeon(R)2.40GHz	16	256	10GB	
x11-24	24	Supermicro	2Xeon(R)2.40GHz	20	256	25Gb	2x2TB
y01-03	3	Supermicro	2Xeon(R)2.40GHz	20	512	25Gb	2x2TB
Total	90			1,580	29,440		



# Juno computational resources as of September 27, 2018

name	#	model	CPU	OS	cores	RAM	Interc	NVMe
jx01-10	10	Supermicro	2Xeon®2.40GHz	CentOS7	20	256	25Gb	2x2TB
ju14	1	HPE DL160G9	2Xeon®2.60GHz	CentOS6	16	256	10Gb	
Total	11				216	2,816		



# Why is Juno replacing Luna?

- New 2.6PB GPFS 5.1 storage /juno available only on CentOS7 nodes
- New LSF 10.1 FP6 servers and configuration
- LSF can run jobs on CentOS6 and CentOS7 nodes



# Overview of LSF configuration changes

- RAM in GB is per task(slot), not per job!
- All jobs must have -W (Walltime) and LSF will terminate the job which exceeds it.
- Cgroups: Memory enforced by cgroups. No /swap.
- To check FINISHED job use “bhist” or “bacct”. “bjobs” will not have information on FINISHED jobs.
- Two loan policies: Short (90 minutes, 100% of resources) and Medium (4 hours, 75% of resources).
- Users can request nodes with NVMe.
- HealthCheck: Jobs are only dispatched to “healthy” hosts.
- GPFS and OS reserve ~12GB of RAM per host.
- No iounits, no custom esub to overwrite bsub parameters



# Job default settings on Juno (LSF)

- Job default parameters
  - Queue name: general
  - Operating System: CentOS7
  - Number of slots (-n): 1
  - Walltime (max job runtime): 6 hours
  - Memory (RAM): 2GB
- “bsub” will overwrite the default parameters
- To check the queue configuration:  
`bqueues -l general`
- To check the default application configuration:  
`bapp -l defaultOS7`



# How to execute jobs on Juno (LSF)

- To submit a job to CentOS7 nodes:

```
bsub -n 1 -W 1:00 -R "rusage[mem=2]"
```

```
bsub -n 1 -W 1:00 -app anyOS -R "select[type==CentOS7] rusage[mem=2]"
```

- To submit a job to CentOS6 nodes:

```
bsub -n 1 -W 1:00 -app anyOS -R "select[type==CentOS6] rusage[mem=2]"
```

- To submit a job to any nodes, either CentOS6 or CentOS7:

```
bsub -n 1 -W 1:00 -app anyOS -R "rusage[mem=2]"
```

- To submit a job to nodes with NVMe /fscratch or /pic:

```
bsub -n 1 -W 1:00 -R fscratch
```

```
bsub -n 1 -W 1:00 -R pic
```





# LSF job monitoring

- Check all my jobs  
`bjobs`
- Check my job's stdout and stderr while the job is running  
`bpeek JID`
- Check status of my job using JobID (JID)  
`#1: bjobs -l JID`
  - Why can't my job run now? Check "PENDING REASONS" in the output #1
  - When will my job start to run? Check "ESTIMATION" in the output #1
- Why did my job exit abnormally?  
`bhist -l JID`  
`bhist -n 0 -l JID`
- To kill my job  
`bkill -l JID`
- To kill all my jobs  
`bkill 0`



# Useful LSF commands

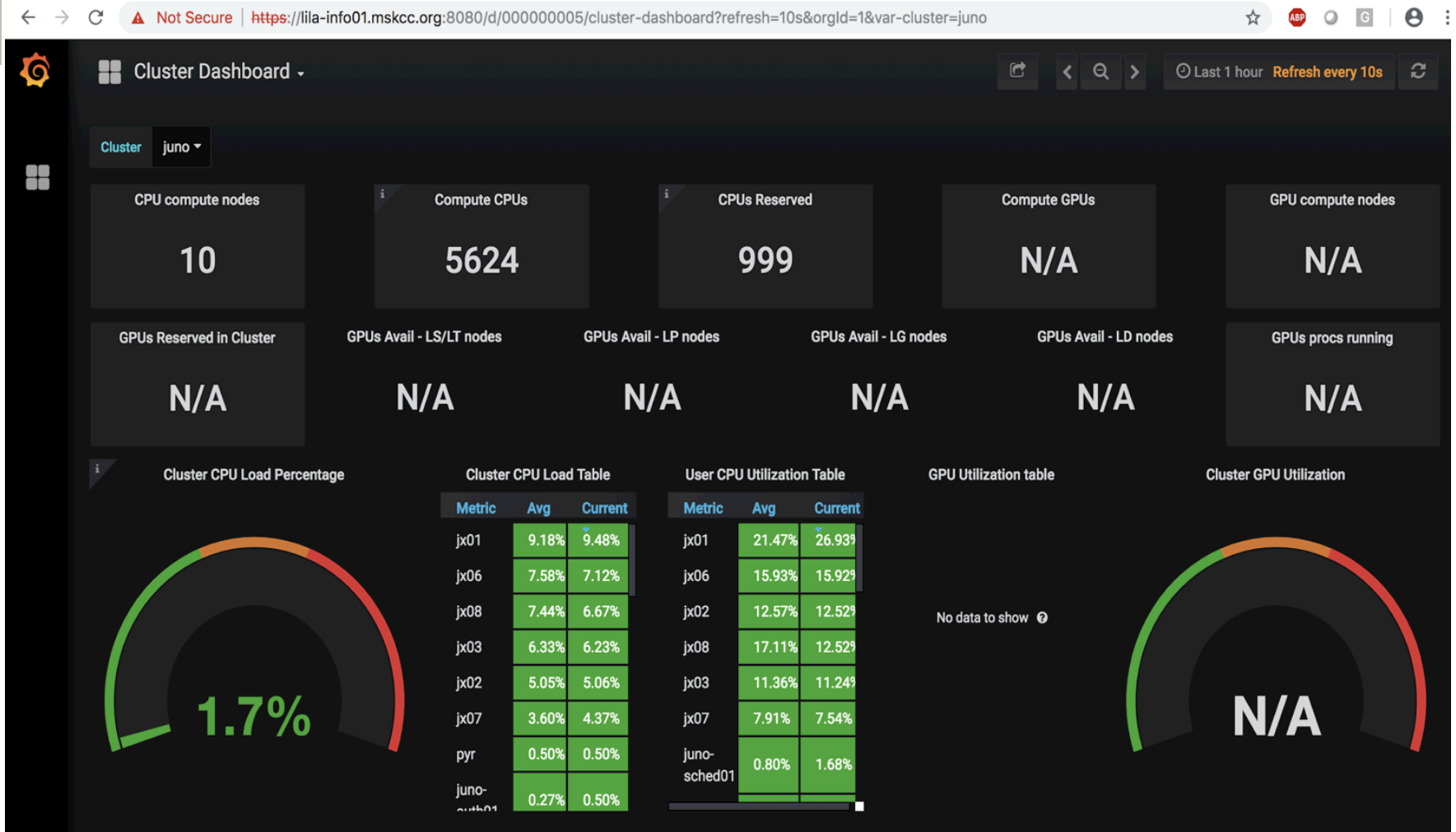
Please use: `man "command"`; or `"command" -h` to check all flags

- `bsub`
- `bhosts ; bhosts -l node_name`
- `bqueues; bqueues -l queue_name`
- `bmggroup`
- `lsload; lsload -l host_name`
- `lshosts`
- `bsla`
- `bjobs -uall -m host_name; bjobs -p`
- `bkill`
- `bhist -l JID; bhist -n 0 -l JID`
- `lsload -l healthy | grep 0.0`



# Grafana Juno cluster Dashboard

<https://hpc-grafana.mskcc.org/>



# Software available on Juno

- /opt/common/CentOS\_7
- Working with modules:

```
[sveta@juno:~]#module list
Currently Loaded Modulefiles:
  1) singularity/2.6.0
[sveta@juno:~]#module avail
```

```
----- /usr/share/Modules/modulefiles -----
dot          module-git  module-info modules      null          use.own
```

```
----- /opt/common/modulefiles -----
singularity/2.6.0
[sveta@juno:~]#module load singularity
[sveta@juno:~]#module list
Currently Loaded Modulefiles:
  1) singularity/2.6.0
[sveta@juno:~]#module purge
[sveta@juno:~]#module list
No Modulefiles Currently Loaded.
[sveta@juno:~]#
```



# How to get help on Luna/Juno

- Please, send email to: [hpc-request@cbio.mskcc.org](mailto:hpc-request@cbio.mskcc.org)
- All information on how to contact us:  
<http://hpc.mskcc.org/contact-us/>



# July Network Upgrade, 1/3

Our July work included three main network changes.

We consolidated the Lilac and Luna/Juno clusters onto a single network switch.

- This will enable us to loosen restrictions between the Lilac and Juno clusters in the future, and make it easier for you to leverage both clusters.



## July Network Upgrade, 2/3

The new switch is designed to run at 100gbps, while the old switch was 10gbps based.

- The switch currently has 224 \* 100 gigabit Ethernet ports.
- Each port can run at either 100gbps or 40gbps.
- Each port can also be broken out into 4 sub-ports, running at either 25gbps or 10gbps — 896 connections if we broke them all out.
- Most of our equipment still connects to our private network at 10gbps (compute nodes) or 20gbps (named servers), but recent equipment and new purchases use 25gbps.



# July Network Upgrade, 3/3

We upgraded our newest equipment to 25gbps.

- Luna x## nodes have 25gbps connections.
- Lilac It09..22 nodes have 25gbps connections.
- Juno and Lilac GPFS storage servers have been upgraded to 100gbps. We see throughput on sequential 16M block size: for writes=15GBps and reads=27GBps
- All new equipment will connect to the private network at 25gbps or better.
- We are working to make Juno/Lilac more similar, and exploring future changes to remove some barriers between them.





# Documentation wiki

- <http://mskcchpc.org/display/CLUS/Juno+Cluster+Guide>
- <http://hpc.mskcc.org/compute-accounts/>





# Questions/Answers

