

Running jobs on Juno

Cluster resources

The Juno cluster currently has: ~148 nodes, ~7,280 CPUs, 16 GPUs, 8.1 PB usable computational storage /juno and 3.1 PB usable warm storage /warm. There are eleven different node configurations. A detailed summary of the Juno cluster hardware is available at <http://hpc.mskcc.org/compute-accounts/>

Realtime Juno cluster information is on Grafana <https://hpc-grafana.mskcc.org/d/000000005/cluster-dashboard?refresh=10s&orgId=1&var-cluster=juno&var-GPUs=All&var-gpuhost=All>

RTM has information about LSF <http://juno-rtm01.mskcc.org/cacti/index.php>

User name: guest

Getting cluster info from LSF on the command line

These are some of the commands you can use to get current information about the compute nodes and LSF configuration on the juno cluster.

- *bhosts* displays hosts and their static and dynamic resources.
- *lshost* displays hosts and their static resource information.
- *bqueues* displays information about queues.
- *bmgroup* displays the host groups.
- *bjobs -u all* will show all running and pending jobs.

Logging in

Access to the Juno cluster is by ssh only. The login node for the juno cluster is juno.mskcc.org. Please do not run compute jobs on the login node. There are compute servers available for local compute. Please use the data transfer server juno-xfer01.mskcc.org for moving large datasets.

LSF cluster defaults

We reserve ~12GB of RAM per host for the operating system and GPFS on Juno hosts.

Each LSF job slot corresponds to one CPU hyperthread. All Juno compute nodes have access to the Internet. The default LSF queue, 'cpuqueue' should be used for CPU jobs only. The gpuqueue queue should be used for GPU jobs only. It is a good idea to always specify the number of threads, memory per thread, and expected wall time for all jobs.

These are the the LSF job default parameters if you do not specify them in your bsub command:

- Queue name: cpuqueue
- Number of CPU (-n): 1
- Memory (RAM): 2GB
- span[hosts=1]
- Walltime (job runtime): 6 hours

The maximum walltime for jobs on juno is 31 days.

Local scratch space on nodes

All nodes have a local 1T /scratch drive.

Some nodes have additional local 2T NVMe /fscratch drives. Nodes with NVMe /fscratch can be requested with the bsub argument -R fscratch.

Please clean up your data on scratch drives when your job finishes. Don't use /tmp for scratch space or any job output.

The cleanup policy for scratch drives is:: files with access time > 31 days deleted daily at 6:15AM for /scratch and 6:45AM for /fscratch.

Service Level Agreements

Some subsets of compute nodes were purchased by partner PIs or individual departments. These systems are placed into the cluster under special queue configurations that enable prioritization for the contributing group. All users benefit from such systems as they allow jobs to run on them while they are idle or under low utilization. The rules for group owned nodes are defined in the LSF scheduler configuration as Service Level Agreements (SLAs) which give specific users proprietary access to subsets of the nodes and define the their loan policies.

SLA name Loan Policy Auto Attached

CMOPI (ja*,jx* hosts) 100% resources for 90 mins 75% resources for 6 hours 40% resources for <31 days No

DEVEL (ja*,jx* hosts) 100% resources for 90 mins 75% resources for 6 hours 40% resources for <31 days No

jvSC 100% resources for 6 hours Yes

jdSC 100% resources for 6 hours Yes

jbSC 100% resources for 6 hours Yes

jcSC 100% resources for 6 hours No

Auto Attached Yes: job will be attached to SLA. No request for SLA needed Auto Attached No: job has to request SLA "bsub -sla CMOPI" "bsla" shows existing SLAs "bugroup" checks the mapping UID to LSF groups

Job submission examples

To request a job with 4 CPUs and a total of 48G RAM that runs for 12 hours

```
bsub -n 4 -R "rusage[mem=16]" -W 12:00 myjob
```

New queue: gpuqueue for GPU jobs only

```
bsub -q gpuqueue -sla jcSC -n 1 -gpu "num=1" myjob
```

To check jobs which are DONE or have status EXIT, use "bhist -l JobID" or "bhist -n 0 -l JobID". bacct is also available. "bjobs -l JobID" only shows RUNNING and PEND jobs.

The bjobs command can show jobs with DONE/EXIT status for 24 hours.

Example: bsub -w "post_done('JOB_A') -J 'JOB_B'" ... if JOB_A was DONE 72 hours before JOB_B was submitted, JOB_B will never start.