Parallel Computing

Overview

All SLAC users can run parallel jobs on the "bullet" shared cluster. It has 5024 cores. Each cluster node is configured as follows:

- RHEL6 64bit OS x86 nodes
- 2.2GHz Sandy Bridge CPUs
- 16 cores per node
- 64GB RAM per node
- QDR (40Gb) Infiniband for MPI comms
- 10Gb ethernet for SLAC networking

If your MPI or mulitcore job only needs <= 16 cores and the memory requirement fits on a single host, you should submit to the general queues. Use the following syntax to run an 8-core job on a single host:

```
bsub -n 8 -R "span[hosts=1]" -W <runlimit> <executable>
```

More information on the general queues: https://confluence.slac.stanford.edu/display/SCSPub/High+Performance+Computing+at+SLAC

For parallel jobs that require multiple hosts, there are 2 public queues for MPI computing: bulletmpi and bulletmpi-large. They are available to anyone with a SLAC unix account. Jobs submitted to bulletmpi and bulletmpi-large will reserve entire hosts and run on these hosts exclusively. **Please send email to unix-admin@slac.stanford.edu to request access to these queues.**

- bulletmpi for jobs between 8 and 512 cores
- bulletmpi-large for jobs between 513 and 2048 cores

Queue	min. # cores	max. # cores	default runtime	max. runtime
bulletmpi	8	512	15 mins	7 days
bulletmpi-large	513	2048	15 mins	1 day

Single slot jobs are not allowed in these queues. You should specify the wallclock runtime using the -W <minutes> or -W <hours:minutes> bsub arguments. There is also a limit on the total number of cores (slots) in use by the bulletmpi and bulletmpi-large queues. You can check the current slot usage and the slot limits by running the blimits command. The output below shows the combined slot total for bulletmpi and bulletmpi-large is limited to 3072 slots. All 3072 slots are in use:

renata@victoria $ blimits -w

INTERNAL RESOURCE LIMITS:

```
NAME                          USERS       QUEUES
HOSTS          PROJECTS      SLOTS       MEM TMP SWP JOBS
bulletmpi_total_limit             -       bulletmpi bulletmpi-large   bulletfarm/         -        3072
/3072         -   -    -    -
bulletmpi_slot_limit          hezaveh     bulletmpi                         -              -         288
/512          -   -    -    -
bulletmpi_slot_limit          lehmann     bulletmpi                         -              -         128
/512          -   -    -    -
bulletmpi_slot_limit          sforeman    bulletmpi                         -              -         256
/512          -   -    -    -
bulletmpi_slot_limit          frubio      bulletmpi                         -              -          32
/512          -   -    -    -
bulletmpi_slot_limit          weast       bulletmpi                         -              -         300
/512          -   -    -    -
bulletmpi_slot_limit          cuoco       bulletmpi                         -              -          20
/512          -   -    -    -
bulletmpi_long_slot_limit     shoeche     bulletmpi-large                   -              -        2048
/2048         -   -    -    -
```

OpenMPI environment

We recommend that you compile and run MPI jobs on the bullet cluster using the lsf-openmpi module. It is built from the RedHat OpenMPI source but compiled with support for the LSF batch job system. Login to one of the interactive bullet nodes via "ssh bullet", you will be redirected to either bullet0001 or bullet0002. Once logged in, run *which mpirun.* The command should return this path:

```
/opt/lsf-openmpi/1.8.1/bin//mpirun
```

You can also check that lsf-openmpi is in use with this command:

```
renata@bullet0002 $ module list
Currently Loaded Modulefiles:
1) lsf-openmpi_1.8.1-x86_64
```

If you are using lsf-openmpi, Make sure you do not override PATH or LD_LIBRARY_PATH with other OpenMPI directories. An example of a job submission using the lsf-openmpi module:

```
bsub -q bulletmpi -n <# cores> -W <runtime_minutes> mpirun <mpi_executable>
```

There is also an earlier version of OpenMPI, 1.5.4, available, but OpenMPI 1.8.1 has the advantage of being able to run on hosts at different infiniband speeds. OpenMPI 1.5.4 can have communication problems if it attempts to run across hosts at different speeds. However, If you find a need for the older version you can set up your environment to use it instead when running on the bullets by making the following changes to the appropriate login script; csh or tcsh users will update .cshrc and bash users will update .bash_profile or .bashrc:

```
##--FOR CSH or TCSH ----------------------------------------------------
set bulletcluster = `hostname | grep "^bullet"`
if ($bulletcluster != "") then
eval `/usr/bin/modulecmd csh unload lsf-openmpi_1.8.1-x86_64`
eval `/usr/bin/modulecmd csh load lsf-openmpi_1.5.4-x86_64`
endif
#-------------------------------------------------------------------------

##--FOR BASH -----------------------------------------------------------
bulletcluster=`hostname | grep "^bullet"`
if [ "$bulletcluster" != "" ]; then
eval `/usr/bin/modulecmd sh unload lsf-openmpi_1.8.1-x86_64`
eval `/usr/bin/modulecmd sh load lsf-openmpi_1.5.4-x86_64`
fi
#-------------------------------------------------------------------------
```

Mailing List

Please join our SLAC openmpi mailing list. You can subscribe by sending a request email to listserv@slac.stanford.edu. You can use a non-SLAC email account if you wish.

The body of the message should include:

sub openmpi <your full name>