

Everest Cluster User Guide

CS 441/541 – Summer 2004

1. Login to the cluster

Your username is your blazer id, and your initial password is your student id (for example 999554444). You would need to login to moat first, if not in the CIS domain

```
shell$ ssh username@everest00.cis.uab.edu
```

You will then be asked *3 questions shown below*. Press “**Enter**” key for all the questions (**entering no other input**) and your ssh keys will generated:

```
It doesn't appear that you have set up your ssh key.
This process will make the files:
    /home/username/.ssh/identity.pub
    /home/username/.ssh/identity
    /home/username/.ssh/authorized_keys

Generating public/private rsa1 key pair.
Enter file in which to save the key (/home/username/.ssh/identity):
Created directory '/home/username/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/username/.ssh/identity.
Your public key has been saved in /home/username/.ssh/identity.pub.
The key fingerprint is:
several 2 digit hex numbers separated by :>
username@everest00.cis.uab.edu
```

2. Changing Password

It is strongly advised that you change your initial password

```
[user@everest00 user]$ passwd
```

You would be asked to enter your old password and your new password.

3. Compiling MPI program

```
[user@everest00 user]$ mpicc -o executable program.c
```

Use the mpicc compiler to compile programs at the shell prompt.

4. Sun Grid Engine

Sun Grid Engine has a large set of programs that let the user submit/delete jobs, check job status, and have information about available queues and environments. For the normal user the knowledge of the following basic commands should be sufficient to get started with Grid Engine and have full control of his jobs:

- `qconf`: Shows (-s) the user the configurations and access permissions only.
For example `qconf -sql` will give you a list of all available queues.
- `qdel`: Gives the user the ability to delete his own jobs only.
- `qhost`: Displays status information about Sun Grid Engine execution hosts.
- `qmod`: Modify the status of your jobs (like suspend/resume).
- `qstat`: Provides a status listing of all jobs and queues associated with the cluster.
- `qsub`: Is the user interface for submitting a job to Grid Engine.

For further information, see the **SGE User's Guide**

<http://www.sun.com/products-n-solutions/hardware/docs/pdf/816-2077-12.pdf> ([PDF](#))

<http://docs.sun.com/source/816-4739-11/enterpri.htm> ([HTML](#))

4.1 Writing and submitting batch jobs

To run a job with grid engine you have to submit it from the command line.

But first, you have to write a batch script file that contains all the commands and environment requests that you want for this job. If, for example, `serial.sh` is the name of the script file then use the command `qsub` to submit the job:

```
#!/bin/bash
#
#$ -cwd
#$ -j y
#$ -S /bin/bash
#
#$ -M myemail
#$ -e error_file
#$ -o output_file
date
sleep 10
date
```

And, if the submission of the job is successful, you will see this message:

```
[user@everest00 user]$ qsub serial.sh
your job 1 ("serial.sh") has been submitted.
```

After that, you can monitor the status of your job with the command `qstat`

When the job is finished you will have two output files called "output_file" and "error_file" (if there were any output/error messages).

```
[user@everest00 user]$ qstat
job-ID prior name      user      state submit/start at   queue      master  ja-task-ID
-----
      4     0 serial.sh  user      qw      06/15/2004 21:40:49
```

In Grid Engine, it is a batch script that contains additionally to normal UNIX command special comments lines defined by the leading prefix ``#\$".

The first line of the batch file starts with

#\$ -S /bin/bash

which is default shell interpreter for Grid Engine. to tell GE to run the job from the current working directory add this script line

#\$ -cwd

if you want to pass some environment variable VAR (or a list of variables separated by commas) use the **-v** option like this

#\$ -v VAR (**#\$ -V** passes all variables listed in env).

Insert the full path name of the files to which you want to redirect the standard **output/error** respectively.

#\$ -o <path_name>

#\$ -e <path_name>

The prefix **#\$** has many options and is used the same way you use **qsub**, so check **qsub** man pages to take a look at those options.

Insert you email-address after the "**#\$ -M**", and also insert the full path name of the files to which you want to redirect the standard output/error. after the "**#\$ -o**" (the "**#\$ -e**") statement, respectively.

Note that that **qsub** accepts shell scripts only, not executable files, and also that shell scripts need to be executable, if it's not the case run the command

```
chmod u+rx serial.sh
```

And after that, to submit the job you simply type

```
qsub serial.sh
```

An example of parallel (MPI) job (parallel.sh) that requests 4 processors:

```
#!/bin/bash
#
## -cwd
## -j y
## -S /bin/bash
#
## -M user@uab.edu
## -e /home/user/error_file
## -o /home/user/output_file
## -pe mpi 4
MPI_DIR=/opt/mpich/gnu/bin
EXECUTABLE=/home/user/hello

$MPI_DIR/mpirun -nolocal -np $NSLOTS -machinefile $TMPDIR/machines $EXECUTABLE
```

And after that, to submit the job you simply type

```
qsub parallel.sh
```

[A sample snapshot of job submission, and monitoring is provided towards the end of this document.](#)

Due to the tight integration of MPI with SGE (via the qsub command), SGE automatically configures a number of environment variables containing values required by mpirun.

The first is \$NSLOTS, the number of slots (or processors) granted by SGE for this MPI job, which corresponds to the (range) value given by the user as the *second argument* to the -pe option. The second variable is \$TMPDIR, a temporary directory which will contain a file titled machines, itself containing an automatically-generated list of nodes on which the MPI job will be run. The temporary directory and its contents will be automatically removed upon completion of the MPI job.

Note: \$MPI_DIR and \$EXECUTABLE are provided for clarity and may be dispensed with, substituting them with their actual values. However, \$NSLOTS and \$TMPDIR are mandatory.

Both these values are passed to mpirun via the specified arguments. The next argument should be the name of the MPI program itself, followed by any optional arguments to be sent to that program. The above script should suffice to run any simple MPI job by changing the name of the program (myprogram) in the mpirun line.

4.2 Monitoring and Controlling Jobs

After submitting your job to Grid Engine you may track its status by using either the qstat command, or by email.

a. Monitoring with qstat

The `qstat` command provides the status of all jobs and queues in the cluster. The most useful options are:

- `qstat`: Displays list of all jobs with no queue status information.
- `qstat -u hpc1***`: Displays list of all jobs belonging to user `hpc1***`
- `qstat -f`: gives full information about jobs and queues.
- `qstat -j [job_id]`: Gives the reason why the pending job (if any) is not being scheduled.

You can refer to the man pages for a complete description of all the options of the `qstat` command.

b. Monitoring Jobs by Electronic Mail

Another way to monitor your jobs is to make Grid Engine notify you by email on status of the job.

In your batch script or from the command line use the `-m` option to request that an email should be send and `-M` option to precise the email address where this should be sent. This will look like:

```
#$ -M myaddress@work
#$ -m beas
```

Where the (`-m`) option can select after which events you want to receive your email. In particular you can select to be notified at the **beginning/end** of the job, or when the job is **aborted/suspended** (see the sample script lines above).

And from the command line you can use the same options (for example): [1]

```
qsub -M myaddress@work -m be job.sh
```

5. Controlling jobs

Based on the status of the job displayed, you can control the job by the following actions:

- **Modify a job:** As a user, you have certain rights that apply exclusively to your jobs. The Grid Engine command line used is `qmod`. Check the man pages for the options that you are allowed to use.
- **Suspend/(or Resume) a job:** This uses the UNIX `kill` command, and applies only to running jobs, in practice you type

```
qmod -s/(or -r) job_id (where job_id is given by qstat or qsub).
```

- **Delete a job:** You can delete a job that is running or spooled in the queue by using the `qdel` command like this

`qdel job_id` (where `job_id` is given by `qstat` or `qsub`).

6. Sample snapshot of MPI job submission along with monitoring

```
[user@everest00 user]$ qsub parallel.sh
your job 15 ("parallel.sh") has been submitted
[user@everest00 user]$ qstat
job-ID prior name      user      state submit/start at      queue      master  ja-task-ID
-----
      15     0 parallel.s user      qw   06/16/2004 14:51:07
[user@everest00 user]$ qstat
job-ID prior name      user      state submit/start at      queue      master  ja-task-ID
-----
      15     0 parallel.s user      t    06/16/2004 14:51:07 everest-0- SLAVE
           0 parallel.s user      t    06/16/2004 14:51:07 everest-0- SLAVE
      15     0 parallel.s user      t    06/16/2004 14:51:07 everest-0- MASTER
           0 parallel.s user      t    06/16/2004 14:51:07 everest-0- SLAVE
[user@everest00 user]$ qstat
[user@everest00 user]$ more output_file
/opt/gridengine/default/spool/everest-0-6/active_jobs/16.1/pe_hostfile
everest-0-6
everest-0-6
everest-0-14
everest-0-14
Warning: Permanently added 'everest-0-6' (RSA1) to the list of known hosts.
Warning: Permanently added 'everest-0-14' (RSA1) to the list of known hosts.
[1]: Hello World, 1 of 4 alive
[3]: Hello World, 3 of 4 alive
[2]: Hello World, 2 of 4 alive
[0]: Hello World, 0 of 4 alive
rm: cannot remove `~/tmp/16.1.everest-0-6.q/rsh': No such file or directory
```

Note that the first time `qstat` was executed, the job was in the queue in submitted state. The job was in execution state when `qstat` was executed the second time, and the third time it had completed.

For further information, see the **SGE User's Guide**

<http://www.sun.com/products-n-solutions/hardware/docs/pdf/816-2077-12.pdf> ([PDF](#))

<http://docs.sun.com/source/816-4739-11/enterpri.htm> ([HTML](#))