HPC Cluster Queues

Queues and Scheduling

Fawkes Cluster Queue Structure

Fawkes has a four queue structure:

default:
- this is the default queue that re-routes jobs to other queues
- the re-routing policy is based on requested resources
- check queue resources below

albus:
- Re-routes jobs to the default queue for Albus cluster
- check below for queue resource details

express:
- all nodes available
- up to 8 hours of run time available

serial:
- all nodes available
- from 24 to 72 hours of run time available

short:
- all nodes available
- from 8 to 24 hours of run time available

medium:
- All nodes available
- from 24 to 72 hours of run-time available

long:
- all nodes available
- from 72 to 168 hours of run time available

verylong:
- not all nodes available
- lowest priority

HPC Cluster Queues

- jobs on this queue can be paused to allow higher priority jobs to run and finish
- from 168 to 400 hours of run time available

Notes:

- the job queuing software is aware of available gpus therefore to use a gpu simply add “gpus=1” to your resource request line and your job will be allocated to a node with gpus.
- the job queuing software is aware of nodes with large memory therefore to use simply request one node with resources of greater than 128Gb of memory and your job will be allocated to a node with large amount of memory.

	express	serial	short	medium	long	verylong
Priority	160	140	160	160	160	50
Min Walltime (hr)	0	24	8	24	72	168
Max Walltime (hr)	8	72	24	72	168	400
Default CPU	1	1	1	1	1	1
Default Memory (Gb)	2	2	2	2	2	2
Max Running Jobs	500	450	450	400	200	40
Soft limit of running jobs	N/A	200	N/A	N/A	N/A	N/A
Max CPUs/job	500	1	200	100	40	10
Max Queued Jobs	20000	10000	20000	5000	2000	2000
Notes		Jobs on this queue can be paused for higher priority jobs				Jobs on this queue can be paused for higher priority jobs

HPC Cluster Queues

Albus Cluster Queue Structure

Albus has a three queue structure:

default:

- this is the default queue that re-routes jobs to other queues
- the re-routing policy is based on requested resources
- check queue resources below

fawkes:

- default queue for Fawkes cluster
- see above for queue resource details

short:

- all nodes available
- up to 24 hours of run time available

medium:

- all nodes available
- from 24 to 72 hours of run time available

long:

- all nodes available
- from 72 to 168 hours of run time available

	short	medium	long
Priority	14	12	10
Default CPU	1	1	1
Default Memory (Gb)	1	2	2
Min Walltime (hr)	0	24	72
Max Walltime (hr)	24	72	168
Max CPUs/job	200	100	40
Max Running Jobs	200	160	96
Max Queued Jobs	20000	5000	1000

HPC Cluster Queues

Queue Design Criteria

The scheduling algorithm used on both clusters aims to:

- promote large scale parallel use of the HPC
- allow equal access to resources for all users
- provide good turnaround for all users
- minimize the impact of jobs on one another

Some of the scheduler features to achieve these aims are:

- resources are strictly allocated so jobs will not start unless there are sufficient free resources (e.g. cpus and memory).
- queued jobs are shuffled so that jobs from different users are "interleaved". This means your first job should appear near the top of the queue even if there are many jobs in the queue.

From a user's perspective, it is very important that you minimize your requests for resources (e.g. walltime, memory and cpus) without restricting your job(s). Otherwise your job may be queued or suspended longer than necessary. Of course, make sure you ask for sufficient resources - a little experimentation might help.