# About UHN Bioinformatics and HPC Core Cluster

Our HPC cluster is a shared resource, please respect our policies and other users on the system.

* Please only request resources (cpu, memory, walltime) needed for your jobs;
* Please do not run jobs on the login node. Jobs on the login node may be terminated without notice;
* If you need to download big data sets, please contact us.
* Each user/group has limited disk quota on our system, please clean up your intermediate results as soon as possible. If you are over quota, you will not be able to write into your directory.

# Ticketing System

If you have any issues or requests on the cluster, please send email to [hpc\_support@uhnresearch.ca](mailto:hpc_support@uhnresearch.ca). A ticket will be created for the issue.

# Log into the Cluster

If you use linux or Mac machine, you need to use terminal or some ssh clients to connect to the cluster; if you use Windows machine, you need to use putty (<https://www.chiark.greenend.org.uk/~sgtatham/putty/>) or similar programs to connect to the cluster.

The cluster is on UHN network, you should log into the cluster using one of the following methods.

* If you are on UHN network, you can use “ssh -p 10022 <your username>@172.27.23.163”; For Windows users, host is 172.27.23.163, port is 10022
* If you are on mordor, you can use “ssh -p 10022 h4huhnlogin1” or “ssh -p 10022 172.27.23.163”
* If you have already logged into h4h cluster, you can "ssh mordorlogin2"
* If you are on the open network, you should use OTP authentication first, and then "ssh -p 5500 <your [username>@192.75.165.28](mailto:username>@192.75.165.28)"; For Windows users, host is 192.75.165.28, port is 5500

# Storage

Your home directory is under /cluster/home/ and you have 50GB quota. Please only store configuration files under your home and clean up often, otherwise it will be full quickly and your jobs may fail if they need to write into your home.

Your group directory will be under /cluster/projects. You should use your group directory as working directory. Different groups will have different quota. Please contact us if you need more space. It always a good idea to clean up your un-necessary intermediate files, log files. Your group directory is not mounted on the login node (h4huhnlogin2), but it is available on all the compute nodes. You can use one of the compute nodes (or the data transfer node) to access your group directory (for example, an interactive session, see below).

Our compute nodes have limited local storage, so please set up your own TMPDIR, TMP and TEMP variables. If you use system default /tmp, you may fill up the system local storage quickly and make everyone face the “no space on device” error on the same node.

# Use of the cluster

## The Scheduler

The cluster uses **Slurm** for cluster management and job scheduling. All information about Slurm can be found at its official web site, <https://slurm.schedmd.com/>. There are also lots of tutorials online.

A cluster is a set of networked computers and each computer represents one "node" of the cluster. When a user submits a job, Slurm will schedule this job on a node (or nodes) that meets the resource requested by the user. If no resources are currently available, the user’s job will wait in a queue until the resources they have requested become available for use.

Nodes are divided into distinct partitions (similar to queues in other scheduling systems) and a node may be part of multiple partitions. we have default resource limits for each partition. Please use the appropriate partition and request necessary resources based on your jobs or they may be killed by the scheduler because of exceeding the limits.

All users will have one or more Slurm usage accounts (usually their groups). Our cluster implements fair-share policy. Your job’s priority is determined by your group’s target usage, your account’s historical resource utilization, time that you job has stayed in the queue, partition that your job will be launched, and resources you request.

## Some Basic Commands

### sinfo

This command is to check the status of the cluster/partitions

sinfo

sinfo –lN # same as above, but shows per-node status

### squeue

This command is to show the status of jobs

squeue

squeue –l # long format

### scancel

This command is to kill jobs

scancel <jobID> # kill job with <jobID>.

scancel -u <username> # kill all jobs for user <username>.

scancel -t <state> # kill all jobs in state <state>. <state> can be PENDING, RUNNING, SUSPENDED

### sshare

This command is to show Slurm share information

sshare

sshare –l # long format

## Running Jobs

### sbatch

This is the command you want to use all the time. It submits your jobs to the scheduler.

sbatch <your job’s shell script>

When submitting a job to the cluster with sbatch you can specify options either within your script, or as command line options. If given within the script, they must be at the beginning of the script and preceded by #SBATCH (as shown in the following example). If given on the command line, drop the #SBATCH and just use the option as usual.

Example template:

#!/bin/bash

#SBATCH -t 1:00 <-- walltime for your job in format of days-hours:minutes:seconds

#SBATCH --mem=256M <-- amount of memory your job needs to run, in megabytes by default

#SBATCH -J test <-- name of your job

#SBATCH -p all <-- partition you job wants to run

#SBATCH -c 1 <-- number of CPUs your job needs to run

#SBATCH -N 1 <-- number of nodes your job needs to run. Please use 1 node unless you are running mpi jobs.

#SBATCH -o %x-%j.out <-- redirect job output (both stdout and stderr) to a file called “<job name>-<job id>.out”. The Slurm default output file is called “Slurm-<job id>.out”

<your commands>

**Mandatory directives: -t, --mem, -c**

### srun

Run a single command on the cluster.

srun <your command>

Example:

srun -c 1 -t 1:0 --mem 128M hostname

### salloc

Schedule an interactive job.

salloc [other slurm options here]

Example:

salloc -c 1 -t 1:0:0 --mem 1G

## Partitions (Queues)

The partitions will be changed/added/removed when necessary. You can check the up-to-date partition list using “sinfo”.

If you want to install software under your user space, you can use the “build” partition where you have internet connection.

### all

* Default=YES
* MaxNodes=1
* MaxTime=5-00:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00
* MaxMemPerNode=30720M

### short

This partition is for interactive jobs only

* Default=NO
* MaxNodes=1
* MaxTime=1-00:00:00
* DefMemPerCPU=256M
* DefaultTime=4:00:00
* MaxMemPerNode=16384M

### himem

* Default=NO
* MaxNodes=1
* MaxTime=7-00:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00
* MaxMemPerNode=61440M

### long

* Default=NO
* MaxNodes=1
* MaxTime=21-00:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00
* MaxMemPerNode=30720M

### ompi

* MaxNodes=UNLIMITED
* MaxTime=5-00:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00
* MaxMemPerNode=30720M

### gpu

* Default=NO
* MaxNodes=1
* MaxTime=3-00:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00

### build

* MaxNodes=1
* MaxTime=2:00:00
* DefMemPerCPU=256M
* DefaultTime=1-00:00:00

## Using GPU nodes

We have GPU machines. To use the GPU nodes, you need to use your “gpu” account and one of the “gpu” partitions. You also need to request the number of GPUs you need to use. For example,

sbatch -A carlsgroup\_gpu -p gpu --gres gpu:1 my\_script.sh

## Module

Our cluster uses "module" to manage software and reference data. You can use

module avai # list all the available modules

module show [<module/version>] # show the details of the module

module load <module/version> # load module for certain version of software or reference

module unload <module/version> # remove the module settings from system

Data Transfer

We have data transfer node, h4huhndata1 (172.27.23.173).

* If you are on the UHN network, you can connect to this data transfer node directly (I.e. using FileZilla)
* If you are on the open network (Internet), you can use scp on the login node to transfer file to your local machine. If the files are under your group directory, you have to copy them first to your home directory. You only have 50G under your home, so if you want to transfer huge files, please contact us. There is no internet on the compute nodes.
* If you need to transfer data between h4h and our mordor cluster
  + From mordor:

ssh xfer

scp my\_file h4huhndata1:/cluster/project/your\_group\_directory

exit

* + From h4h:

ssh h4huhndata1

scp my\_file xfer:/mnt/work1/users/my\_group\_directory

exit