Turing HPC Technical Policies

The overarching goal of these policies is to promote equity in the distribution of the Turing HPC resources while at the same time ensuring that the HPC resources are used effectively and efficiently.

SLURM Scheduling Policy – Fair Share

The SLURM scheduler on the Turing HPC uses a fair share policy in which the priority of a submitted job in the queue is based on a combination of the resources being requested in the user's pending jobs and the resources consumed in the user's previously completed jobs. For example, a user who submits small jobs requesting half a node of resources that run for 2 days will get a higher placing priority in the wait queue as compared to someone who submits larger jobs requesting 2 nodes of resources that run for 4 days. The goal of the fair share policy is to promote equity among the users of the Turing HPC cluster.

SLURM Partition to Quality of Service (QOS)

A set of nodes are assigned to a SLURM partition (queue) based on the type of users and the set of consistent resources (cores, memory, GPU cards) on node types. Table 1 presents the QOS partitions for Turing. All users are expected to be familiar with the maximum number of cores per node, average memory per core, and number of GPU cards per node within a partition. This is important so that they correctly set the SBATCH parameters in the job script. If the job script requests more than the maximum or average amount of resources, SLURM will not allow the job to run.

Quality of Service (QOS) – Fair & Equitable Access

Each user is assigned to an account type that has a default QOS and a certain set of QOS's. These QOS's assign limits on number of cores, amount of memory, number of running jobs, and walltime. These limits (rail guards) ensure that all users anytime will have fair and equitable access to the HPC. The Accounts and their assigned QOS's are shown in Table 2. Each QOS and its limits are shown in Table 3.

Requesting Access to GPU Nodes

The GPU nodes require applications that have been designed to work with GPU hardware. As such, new users are not given access to these nodes by default.

In order to obtain access to these nodes, the Principal Investigator must email a request to HPC Administrator that includes:

- The specific applications that will be used (including documentation and/or links to online documentation)
- A brief description of the proposed usage of the GPU nodes (which can include code development)

The HPC Administrator has the authority to either grant the request or bring the request to the CCAM Computation Committee.

Table 1: QOS Partitions on the Turing HPC Cluster

Partition Name / Purpose	Total Nodes per Part	Available Cores per Node	Available Mem(GB) per Node	Average Mem(GB) per Core	Associated Quality of Service (QOS)
p_nsm_compute_short Used for learning the job environment and for debugging small jobs that require limited resources. Accessible to all users.	1	46	182	3.95 =4044MB	q_nsm_short
p_nsm_compute Used for general computational workloads. Accessible to all users.	31	46	182	3.95 =4044MB	q_nsm_faculty_w,q_nsm_faculty_j, q_nsm_student_w,q_nsm_student_j, q_nsm_long, q_nsm_default
p_nsm_gpu Used by advance users for high performance computing; graphical computing (CUDA and Tensor cores) and visualization. Each node contains 2 GPU cards and can support up to 2 simultaneous GPU jobs (1 per card).	4	46	364	7.9 =8090MB	q_gpu_faculty_w, q_gpu_faculty_j, q_gpu_student_w, q_gpu_student_j,

p_groves_compute_priority Computational nodes purchased by Michael Groves. Members of the GrovesGroup have priority access. All other users have guest access.	8	48	118	2.45 =2508MB	q_groves_faculty_w, q_groves_student_w, q_groves_all
p_groves_compute_guest This partition points to the same Groves nodes as the priority partition. Used by all other users to access Groves nodes.	8	48	118	2.45 =2508MB	q_groves_guest_w, q_groves_guest_j

Table 2: QOS by Account

Account	Default QOS	QOS	Optional
nsmfaculty	q_nsm_faculty_j	q_nsm_faculty_w, q_nsm_faculty_j, q_groves_guest_j, q_nsm_short	q_gpu_faculty_w, q_gpu_faculty_j
nsmstudent	q_nsm_student_j	q_nsm_student_w, q_nsm_student_j, q_groves_guest_j, q_nsm_short	q_gpu_student_w, q_gpu_student_j
chemfaculty	q_nsm_faculty_w	q_nsm_faculty_w, q_nsm_faculty_j, q_groves_guest_w,q_groves_guest_j, q_nsm_short	q_gpu_faculty_w, q_gpu_faculty_j
chemstudent	q_nsm_student_w	q_nsm_student_w, q_nsm_student_j, q_groves_guest_w,q_groves_guest_j, q_nsm_short	q_gpu_student_w, q_gpu_student_j
grovesfaculty	q_groves_faculty_w	q_groves_faculty_w, q_nsm_faculty_w, q_nsm_short	q_gpu_faculty_w, q_gpu_faculty_j
grovesstudent	q_groves_student_w	q_groves_student_w, q_nsm_student_w, q_nsm_short	q_gpu_faculty_w, q_gpu_faculty_j

Table 3: QOS Limits

Quality of Service (QOS)	Max Cores per User	Max Mem (GB) per User	Max Run Jobs per User	Max Walltime per Job (hours)	Max GPU Cards per User	Comments
q_nsm_faculty_w	368	1488	12	100		368cores=8nodes, 1488GB=8nodes
q_nsm_faculty_j	368	1488	24	50		
q_nsm_student_w	184	744	12	100		184cores=4nodes, 744GB=4nodes
q_nsm_student_j	184	744	24	50		
q_nsm_short	23	90	1	6		20cores=0.4node, 80GB=0.4node
q_gpu_faculty_w	46	372	6	100	2	46cores=1node, 372GB=1node

q_gpu_faculty_j	46	372	12	50	2	
q_gpu_student_w	23	186	6	100	1	23cores=1/2node, 186GB=1/2node
q_gpu_student_j	23	186	12	50	1	
q_groves_faculty_w	384		12	100		384cores=8nodes
q_groves_student_w	192		12	100		192cores=4nodes
q_groves_all	384					
q_groves_guest_w	384		12	100		384cores=8nodes
q_groves_guest_j	384		24	50		

NSM Short Queue

An important aspect of ensuring equity in the use of the Turing HPC cluster is the permanent reservation of a single compute node for "short" jobs – jobs that require at most 23 cores, 90 GB of memory, and run for up to 6 hours. Each user will be limited to 1 such calculation at a time. The purpose of this is to facilitate code debugging, testing, and jobs that require limited HPC resources with limited wait-times.

Reviewing and Modifying the SLURM Scheduling Policies

At least once annually, the CCAM Computation Committee will solicit feedback from users about their experiences with Turing. This information, in conjunction with usage data collected by the HPC Administrator, will inform any potential changes to the SLURM scheduling policies.

Review of user feedback and usage data will typically occur during the fall semester with any changes to the SLURM scheduler implemented during a system downtime over winter break. Any changes to the SLURM scheduler that will impact users beyond a single research group will require majority approval form the CCAM Computation Committee.

User Home Directory

Each user account on Turing will get a home directory with fixed size quota. Principal Investigators will receive 5GB and students will receive 2GB for their home directories. The user home directory is used for storing jobs scripts, used as the base for submitting jobs, and storing small computational (result) files. Large files containing raw or processed data will be stored in the Principal Investigator's data space. The user home directory will be backed up on a weekly basis and up to two recent weeks kept on Tier 2 Storage (HDD).

The HPC Administrator will be responsible for creating the original home directory for each user of the Turing HPC cluster.

Data Directory

Principle Investigators will have a data directory that will be used by themselves and members of their research group to store raw and processed data. Applications will read in large data sets from the data directory and output data will be written to the data directory. This organization of storage equips Principal Investigators to develop data management policies that are appropriate for their research group and/or class. The use of the data directory will not negatively impact the performance of the Turing HPC cluster.

The initial size of the data directory will depend on the type of data as well as type of research; the HPC Administrator will collaborate with a new Principal Investigator to determine an appropriate initial allocation of storage space. Typical initial allocations will range from 50GB to 500GB and be managed on Tier 1 Storage (NVME). If beyond 500GB, the data directory will have to be on Tier 2 Storage (HDD).

The HPC Administrator will be responsible for creating the original data directory for each Principal Investigator. The Principal Investigator will then be responsible for managing this directory and creating subdirectories for their research group. Information about doing this will be provided by CCAM.

A Principal Investigator can request additional storage space for their data directory by contacting the HPC Administrator and providing justification for the need for additional storage space. The HPC Administrator has the authority to either grant the request or bring the request to the CCAM Computation Committee.

The Principal Investigator is responsible for managing and backing up the data stored in their data directory. The Principal Investigator is also responsible for housecleaning the data generated by their students.

User Apps Directory

Researchers can request their own software install directory. The default size is 5GB. 10GB or more can be requested depending on type of software. One example could be Anaconda software. Another example could be any beta application that needs to be vetted before installing on the system's directory, /share/apps. Small application databases can be installed like genome classification databases. Please do not install a software package called an RPM because it will try to install on the system directories. Install a source code tar or zip file that you can specify your user apps directory as the install location.

In order to obtain a user apps directory, the Principal Investigator will need to submit a request to the HPC Administrator that includes

- The specific applications that will be installed along with documentation (if appropriate)
- The amount of storage space required
- A justification for the need for a user apps directory

The HPC Administrator has the authority to either approve the request or bring the request to the CCAM Computation Committee.

Software Install Request

If there is Linux based scientific software that is not on the Turing HPC, submit a request to the HPC Administrator that includes the following information:

- Software name and version
- Link to the website containing documentation about the software
- Description of who will use the software
- If possible, the prerequisites for the software and an estimate of the data space required
- Date when the software is needed

The HPC Administrator will review the request within 2 weeks and may have follow-up questions before providing an estimate of the timeline for the software install. The HPC Administrator has the authority to deny software install requests that will negatively impact other users on the Turing cluster. In such situations, the Principal Investigator may bring the matter to the CCAM Computation Committee.