HIGH PERFORMANCE RESEARCH COMPUTING

Introduction to CryoSPARC for Cryo-EM Data Processing on HPRC in Collaboration with BCBP Cryo-EM Research Center

High Performance
Research Computing
DIVISION OF RESEARCH

Spring 2023

CryoSPARC on Grace

- CryoSPARC Academic License ID
- HPRC Accounts and Service Units
- CryoSPARC on the HPRC Portal
- Service Units, Resources and Limitations
  - CPU only job
  - GPU job
- Group data directories
- CryoSPARC Live
- CryoSPARC Demo

CryoSPARC Academic License ID

- CryoSPARC provides a personal Academic License ID free of charge for TAMU staff and students
- A CryoSPARC Academic License ID is required to launch the CryoSPARC HPRC portal app
- Use your TAMU email address for the Academic License
  - https://cryosparc.com/download

Resources for Running CryoSPARC

- CryoSPARC is available as an Interactive App on the HPRC Grace Portal
  - https://portal-grace.hprc.tamu.edu
- Grace hardware overview
  - 800 x CPU only nodes
  - 100 x A100 GPU nodes (2 x A100 per node)
  - 9 x RTX 6000 GPU nodes (2 x RTX per node)
  - 8 x T4 GPU nodes (4 x T4 per node)
- SUs are charged based on type of node selected
  - CPU only = 1 SU per hour per core or per 7GB memory whichever is greater
  - GPU = same as CPU only plus additional charge for number and type of GPUs selected
- Specify enough time to allow your processing to complete
- If you launch a job for 24 hours and you finish your work in 12 hours and *Delete* the portal job, you will be reimbursed SUs for the unused 12 hours.
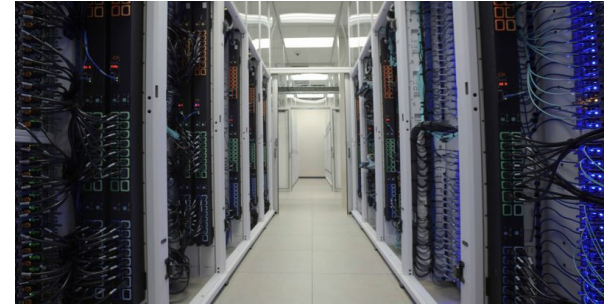
HPRC Accounts and Service Units (SUs)

- HPRC account required at no cost
  - https://hprc.tamu.edu/apply
- Types of allocations
  - **Basic** account with 5,000 SUs
  - **Startup** account of 200,000 SUs
    - PI Status; Professor, Asst Professor, Research Scientist
  - **Research** account of 1,000,000+
    - PI Status; Professor, Asst Professor, Research Scientist
  - https://hprc.tamu.edu/policies/allocations.html
- Unused SUs expire at the end of each fiscal year (Aug 31) and must be renewed
- 1 TB of initial disk space allocated per user
  - https://hprc.tamu.edu/policies/storage.html
  - Can request up to 10TB for short term (< 6 months) projects at no cost
  - Can purchase longtime (> 6 months) storage above 10TB at a cost
    - https://hprc.tamu.edu/wiki/HPRC:Storage

Grace Service Unit Calculations: CPU only

- For the Grace 384GB memory non-GPU nodes (360GB available), you are charged Service Units (SUs) based on one of the following values whichever is greater.
  - 1 SU per CPU per hour or 1 SU per 7.5 GB memory

Number of GPUs	Number of Cores	Total GB Memory per node	Hours	SUs charged
0	1	7.5	1	1
0	1	8	1	2
0	1	360	1	48
0	48	7.5	1	48
0	48	360	1	48

Grace Service Unit Calculations: GPU

- A CryoSPARC portal job reserves all the resources of one GPU node
- For the Grace 384GB memory GPU nodes (360GB available), you are charged Service Units (SUs) based on the following

Number of GPUs	Number of CPU Cores	Total GB Memory per node	Hours	Total SUs
A100 x 2	48	360	1	192
RTX 6000 x 2	48	360	1	192
T4 x 4	48	360	1	144

Each A100 GPU has 40GB memory
Each RTX 6000 GPU has 24GB memory
Each T4 GPU has 16GB memory

Resource Limitations

- Can only launch one CryoSPARC portal session at a time
  - can run multiple CryoSPARC jobs within a portal session
- CPUs
  - Some CryoSPARC jobs do not require a GPU
  - Can build a CryoSPARC job on CPU and then restart the portal app on a GPU node and run the pre-built job.
- GPUs
  - The GPU queue can get busy at times and it may take over an hour for your job to launch
  - Cancel your pending job if you will be away from your computer and it hasn't started yet

Display GPU Availability on Grace

See GPU configuration and current availability using the Grace portal shell access

- https://portal-grace.hprc.tamu.edu
- select "Clusters" -> ">_grace Shell Access"
- enter your password and authentication
- at the command line prompt, enter:

  gpuavail

```
         CONFIGURATION
NODE                NODE
TYPE                COUNT
--------------------------
gpu:a100:2          100
gpu:rtx:2           9
gpu:t4:4            8
```

There are currently 100 compute nodes on Grace that have 2 x A100 GPUs attached to each

```
         AVAILABILITY
GPU         GPU     GPU     GPU
NODE        TYPE    COUNT   AVAIL*
------------------------------
g024        a100    2       1
g090        a100    2       2
g093        a100    2       1
g094        a100    2       2
g098        a100    2       2
g099        a100    2       2
g100        a100    2       1
r001        rtx     2       2
r002        rtx     2       2
r003        rtx     2       2
r004        rtx     2       2
r005        rtx     2       2
r006        rtx     2       2
t001        t4      4       4
t002        t4      4       4
t003        t4      4       4


*  Some available GPUs may not be immediately
   immediately available due to other jobs
   using all or most of the compute node memory
```

Grace CryoSPARC Portal App

hprc.tamu.edu

You must first obtain a CryoSPARC academic license id
https://cryosparc.com/download
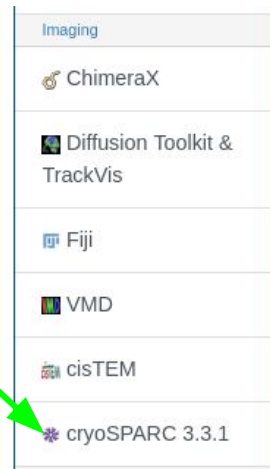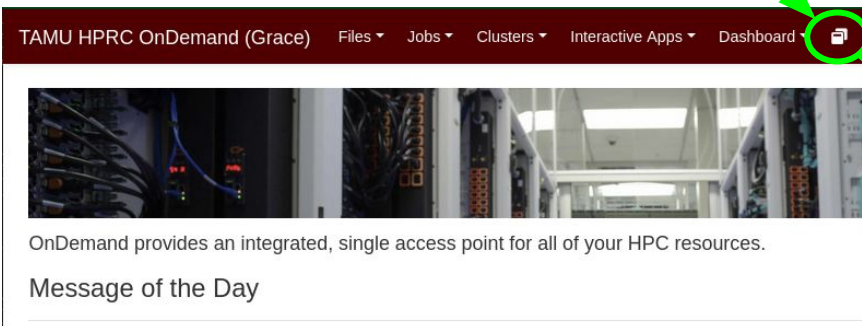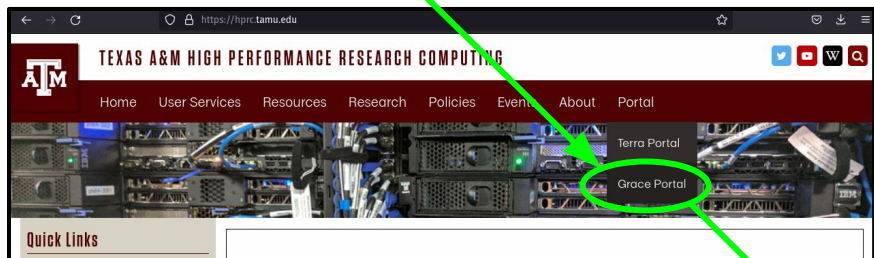
CryoSPARC 3.3.1

This app will launch CryoSPARC on one compute node

CryoSPARC license id (required)

- You must get an individual license id from cryosparc.com/download
- After you click the blue Launch button below, the instructions for connecting to CryoSPARC will be displayed on the next page once the CryoSPARC startup is complete which can take up to 5 minutes
- You can only launch one CryoSPARC portal app job but you can run multiple computational jobs within CryoSPARC

GPU type and count

A100 x 2

- SUs charged per hour
  - A100 x 2 = 192 SUs
  - RTX 6000 x 2 = 192 SUs
  - T4 x 4 = 144 SUs
  - CPU only = 48 SUs
- All multi-GPU options have 48 cores and 360GB memory for CPU processes.
- Each A100 GPU has 40GB of memory, RTX 6000 GPU 24GB and T4 GPU 16GB.
- Queue times can be long if the cluster is busy especially for GPU nodes.

Optional group directory to mount

CryoSPARC Visualization of Resource Utilization

Review CPU and GPU usage for a Job

The `jobstats` utility automatically runs for each CryoSPARC portal job and monitors CPU and GPU resource usage and creates a graph for each.
- CPU stats monitors all cores regardless of how many were configured for the job
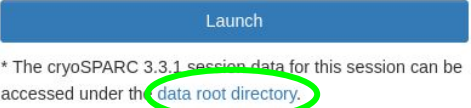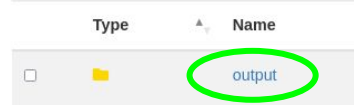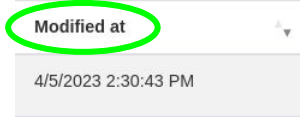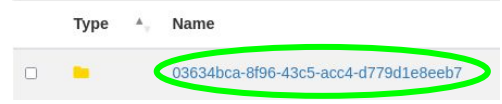- GPU stats will create a graph for each GPU that was configured for the job
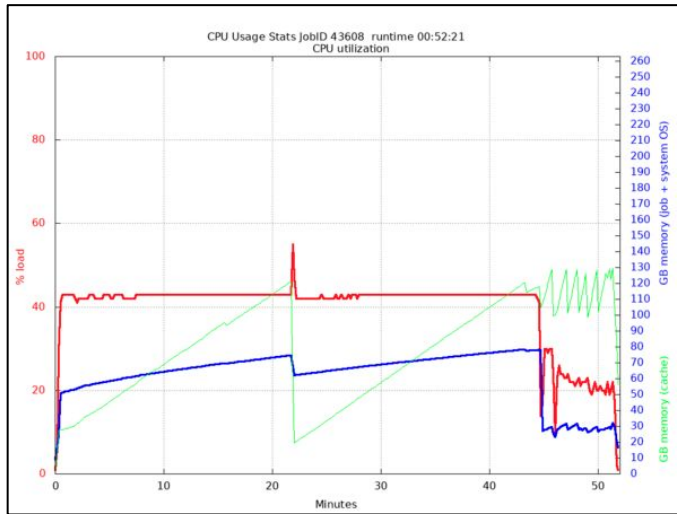
`stats_cpu.43608.png`
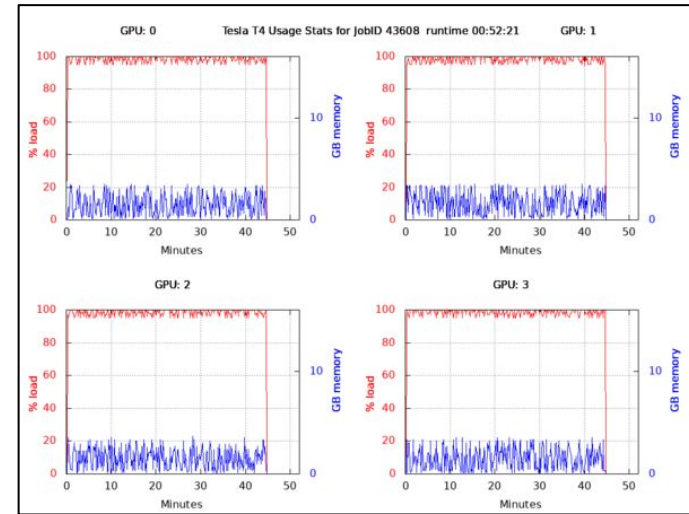
`stats_gpu.43608.png`

Review CPU and GPU usage for a CryoSPARC Job

1. At the bottom of the CryoSPARC portal app, access the CryoSPARC data root directory
2. Click 'output'
3. Click 'Modified at' to sort newest session on top
4. Click the Name of the session session you want to view based on the date
5. Click the .png image files to view the CPU and GPU usage stats

Group Data Directories

- Group data directories can be used to share input files or to have a group work space
- Send a request to the HPRC helpdesk to create a group
  - help@hprc.tamu.edu
  - provide a group name and NetIDs of members
  - group directories have their own disk quotas separate from individual users
- You will need to mount your group directory when launching the CryoSPARC portal app

Optional group directory to mount

- You must already be a member of the group
- Example values:
  - /junjiez
  - /scratch/group/davislab

CryoSPARC Live

- Allows streaming from Cryo-EM microscopes to a CryoSPARC session
- CryoSPARC Live is not yet configured on the HPRC clusters.

CryoSPARC Demo