Hummingbird Computational Cluster
Overview and Introduction

Hummingbird Support

Team hummingbird@ucsc.edu
Hummingbird Etiquette

The DOs and DON'Ts of using the Hummingbird Cluster
Software - the module system
Using the cluster - Scheduling
Template SLURM scripts

Premade templates for your convenience. Just make a copy to your working directory and edit to meet your needs.

If you have a template that is useful to more than just yourself, let us know and we can include it here.
Anatomy of a basic SLURM Script

- Name of Partition to use
- Name to give your run
- Name for output/error logs (use for troubleshooting)
- Number of nodes to request
- Number of tasks your program will require
- Amount of RAM requested for your program
- Time limit for the run (!!!)
- Get status emails about your job

```
#!/bin/bash
SBATCH -p 128x24
SBATCH -j example_job
SBATCH -o job.%j.out
SBATCH -e job.%j.err
SBATCH -N 1
SBATCH -n 24
SBATCH --mem=600mb
SBATCH --time=00:05:00
SBATCH --mail-type=ALL
SBATCH --mail-user=rkparson@ucsc.edu

module load python-3.6.5
export EXMP_VAR="foo"
python my_python_script.py
```
Any modules your program requires in order to run

Any system variables your program might require

The program to run
Submitting your job

```bash
sbatch your_job_script.slurm
```

Verifying your job's settings

```bash
scontrol show job <jobid>
```

That's a lot of details!
A view into what's running - **squeue**

```
[rkparson@hb ~]$ squeue -l
Wed Dec 2 13:15:07 2020

<table>
<thead>
<tr>
<th>JOBID</th>
<th>PARTITION</th>
<th>NAME</th>
<th>USER</th>
<th>STATE</th>
<th>TIME</th>
<th>TIME_LIMI</th>
<th>NODES</th>
<th>NODELIST(REASON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>88174</td>
<td>128x24</td>
<td>6_first_</td>
<td>blufox</td>
<td>RUNNING</td>
<td>16-20:12:26</td>
<td>62-00:00:00</td>
<td>1</td>
<td>hbcomp-024</td>
</tr>
<tr>
<td>88297</td>
<td>128x24</td>
<td>mapping_</td>
<td>jharenc</td>
<td>RUNNING</td>
<td>12-02:44:22</td>
<td>UNLIMITED</td>
<td>3</td>
<td>[012-014]</td>
</tr>
<tr>
<td>88370</td>
<td>128x24</td>
<td>7_first_</td>
<td>blufox</td>
<td>RUNNING</td>
<td>7-21:54:56</td>
<td>62-00:00:00</td>
<td>1</td>
<td>hbcomp-020</td>
</tr>
<tr>
<td>88493</td>
<td>128x24</td>
<td>8_first_</td>
<td>blufox</td>
<td>RUNNING</td>
<td>1-00:48:10</td>
<td>62-00:00:00</td>
<td>1</td>
<td>hbcomp-025</td>
</tr>
<tr>
<td>88498</td>
<td>128x24</td>
<td>PMZ7bi2</td>
<td>vroger</td>
<td>RUNNING</td>
<td>21:38:24</td>
<td>20-20:00:00</td>
<td>1</td>
<td>hbcomp-006</td>
</tr>
<tr>
<td>88499</td>
<td>128x24</td>
<td>P7unfre</td>
<td>vroger</td>
<td>RUNNING</td>
<td>21:35:09</td>
<td>20-20:00:00</td>
<td>1</td>
<td>hbcomp-007</td>
</tr>
<tr>
<td>88501</td>
<td>128x24</td>
<td>P7cunfre</td>
<td>vroger</td>
<td>RUNNING</td>
<td>21:28:56</td>
<td>20-20:00:00</td>
<td>1</td>
<td>hbcomp-009</td>
</tr>
<tr>
<td>88507</td>
<td>128x24</td>
<td>75898_3p</td>
<td>amstahl</td>
<td>RUNNING</td>
<td>20:07:49</td>
<td>INVALID</td>
<td>3</td>
<td>[010-011,018]</td>
</tr>
<tr>
<td>88511</td>
<td>128x24</td>
<td>extsemim</td>
<td>oross</td>
<td>RUNNING</td>
<td>5:01:56</td>
<td>20-10:00:00</td>
<td>3</td>
<td>[015-017]</td>
</tr>
</tbody>
</table>
```

Much easier to parse!
Hummingbird is NOT a SUPERCOMPUTER!

- Hummingbird does NOT use a high speed backplane, thus it is not suitable for highly parallelized jobs requiring many cores/nodes acting synchronously (FEM, FDTD, etc...)
- Humming IS good for serialized style processing (Monte Carlo, batch processing, etc...)
- Users are limited to using three nodes for a single job (a maximum of 72 cores)
- Most nodes only have 128 GB or RAM
- Hummingbird works on First-In, First-Out with Backfilling
- We don't *enforce* a MaxTime, so you *should*
A quick way to check the cluster - `sinfo`

Some queues are restricted and are not for general use

We have four NVIDIA GPUs, and are setup for CUDA and TensorFlow!
Knowing your Hummingbird variants

hb.ucsc.edu
- Cluster login node
- Used for:
  - Compiling your code
  - Checking the status on your submitted jobs

hbfeeder.ucsc.edu
- Cluster storage node
- Used for:
  - Accessing your data when you don't need the cluster
  - Transferring your data to or from the cluster
Everyone is limited to 1TB of storage in their home folders.
If you exceed this, your jobs won't run correctly.
How to check:

- `du -sh ~` total used space summary
- `du -d1 -h ~` used space per-folder (one-layer deep)
Key points to take away

- Do Check the Message of the Day for useful updates
- Do Check in the modules before you ask for software
- Do Use the scheduler when submitting jobs
- Don't Use the cluster for heavily parallel style workflows
- Do Set time limits on your jobs
- Don't Overprovision when submitting a job
- Do regularly check your disk space usage
- Don't Leave data in Scratch long-term
- Do Reach out for help!
Reaching out for help!

Open a ticket!
https://ucsc.service-now.com/ess/

Come to our weekly open-hours sessions!
Find the link in our Message of the Day on the Hummingbird login node:

```
Do you have questions? Need help? Want to speak to an expert?
Join the Hummingbird Zoom-in Help Clinic Thursdays at 1pm
https://ucsc.zoom.us/j/98567158305?pwd=emZhYzBlN0tqeS9UbVpYdllpcHAYdze9 (UCSC log-in required)
rkparson@hb ~]$
```

Check out the website for more information!
https://www.hb.ucsc.edu/
Please take a moment and fill out our post-session survey

https://docs.google.com/forms/d/e/1FAIpQLSeaLMqv3yPcgxy3-sfwZsnCLu5PHvyXg4i60rYnSRxJVVEZog/viewform?usp=sf_link