Welcome to the Smithsonian Institution High Performance Computing Wiki.

This Wiki holds information for the use of HPC resources at the Smithsonian.

Central to the SI HPC resources is the Smithsonian Institution High Performance Cluster (SI/HPC), named Hydra.

- High performance computing is administered by the Office of Research Computing.
- The OCIO data center in Herndon, VA houses the high performance computing cluster, Hydra.

The documentation is organized as follow:

1. A quick start guide;
2. Reference pages (technical information about the cluster);
3. Cluster Upgrade documentation (past upgrades);
4. Cookbooks pages (didactic information on how to run specific tools); (NYA: not yet available)
5. FAQs, or frequently asked questions

What's New

- **Apr 3 2020** - Hydra DOI
  - We have setup a DOI for Hydra [https://doi.org/10.25572/SIHPC](https://doi.org/10.25572/SIHPC)
  - You are now able and encouraged to cite Hydra whenever research has benefited from its use and add a DOI link to that citation or acknowledgment.

- **Mar 24 2020** - Hydra status while teleworking
  - Hydra remains up and running.
  - We will address problems that require on-site staff as fast as possible.
  - We will answer people's questions and requests as promptly as possible.
  - Access to Hydra via VPN:
    - users are asked to limit the strain on the institutional VPN resources when ever appropriate.
    - Hydra can be accessed without VPN:
      - SAO users can use login.cfa.harvard.edu to ssh to Hydra.
      - Access to the self serve password page is now working.
      - How to use Dropbox or Firefox Send ([ffsend](https://ffsend.com)) to copy files to/from Hydra is documented on the Wiki.

- The scrubbing policy has been modified as follows:
  - the scrubber will run on /pool/sao and /pool/genomics as usual, but
  - the scrubbed content will not be deleted for at least 21 days, and
  - we will accept requests to preserve what was scrubbed (beyond 21 days) as long as needed. To get your files restored, follow the usual instructions.
  - Users are asked to remain in contact via their SI email.

- **February 27, 2020** - resource and IDL 8.7.3
  - We have added a new resource, called cpu_arch, to allow users to direct jobs on nodes with CPUs of a specific (list of) architecture(s).
  - If you run jobs/codes that can only run on (a) specific type(s) of processors, look at the new section CPU Architecture under the Available Queues page.
  - IDL version 8.7.3 has been installed on Hydra, and is accessible via the idl/8.7.3 module. The idl/8.7 module is now pointing to idl/8.7.3

- **January 14, 2020** - Increased total slot limit
  - The total number of slots (CPUs) a user can grab has been increased from 512 to 640.

- **December 20, 2019** - Disk space re-tuning
  1. increased /home (doubled it) to 20T
  2. doubled each user quota on /home (100G/2M 200G/4M);
  3. resized (increased) /data/genomics and /data/sao to 45TB each;
  4. resized (increased) /pool/genomics and /pool/sao to 80TB each;
  5. increased each user quotas to 1TB/2.5M on /data/genomics (doubled);
  6. decreased /scratch/genomics and /scratch/sao from 400T 350TB each.

Reminders

- All public disks (/pool and /scratch only) are scrubbed once a week.
- Reasonable requests to restore scrubbed files must be sent no later than the following Friday, by 5pm.
- If you have any questions or encounter any problems, email:

<table>
<thead>
<tr>
<th>Email</th>
<th>For:</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:SI-HPC-Admin@si.edu">SI-HPC-Admin@si.edu</a></td>
<td>sys-admin related issues</td>
</tr>
<tr>
<td><a href="mailto:SI-HPC@si.edu">SI-HPC@si.edu</a></td>
<td>Bioinformatics/Genomics</td>
</tr>
<tr>
<td></td>
<td>questions,</td>
</tr>
<tr>
<td><a href="mailto:hpc@cfa.harvard.edu">hpc@cfa.harvard.edu</a></td>
<td>SAO users who need help.</td>
</tr>
</tbody>
</table>

- Past news with more details can be found at the What's New page.

Quick Links
Cluster monitoring page
QSub Generator (only accessible within the SI network)

FYI

- References to Hydra (publications, proposals, etc) should mention the cluster as:

  *the* Smithsonian Institution High Performance Cluster *(SI/HPC)*.

---

Recently Updated Pages

- **Conda tutorial**
  - Jul 14, 2020 • updated by Kweskin, Matthew • view change

- **Smithsonian Institution High Performance Cluster (SI/HPC) Overview**
  - Apr 03, 2020 • updated by Korzenik, Sylvain • view change

- **High Performance Computing**
  - Apr 03, 2020 • updated by Korzenik, Sylvain • view change

- **Disk Space and Disk Usage**
  - Mar 24, 2020 • updated by Korzenik, Sylvain • view change

- **Available Queues**
  - Feb 27, 2020 • updated by Korzenik, Sylvain • view change