Clusters and Storage

[ Prince ] [ Dumbo ] [ Brooklyn ]

The NYU HPC team currently maintains two clusters: The HPC cluster Prince and the Hadoop cluster Dumbo.

HPC user accounts
An HPC User account provides access to all NYU HPC and Big Data clusters. If you don't have a user account, you may apply for an HPC user account.

Old HPC clusters
NYU HPC team has retired its older clusters (Union Square, Cardiac, Bowery, Mercer). The current production HPC cluster is Prince.

- **Prince**

  *Prince* is the new HPC cluster that is currently being deployed. Prince will replace the HPC Mercer Cluster.
  - For a description of the HPC Prince cluster, see Clusters - Prince.
  - For information on how to access and use the HPC Prince cluster, see Getting started on Prince.

- **Dumbo**

  *Dumbo* is a 44 data node Hadoop cluster running Cloudera Distribution of Hadoop (CDH).
  - For a detailed description of dumbo and how to access it, please see the dumbo wiki pages.

- **Brooklyn**

  Brooklyn Research cluster is currently in its Alpha phase. It will be in its Pilot phase in Spring 2018.

  - Brooklyn is an Openstack cluster consisting of 25 compute nodes, each equipped of 4 GPUs.
  - For a detailed description of the Brooklyn Research cluster, please see Clusters - Brooklyn Research Cluster
  - Presentation Slides

The table below shows the File Systems available on the Prince Cluster.

<table>
<thead>
<tr>
<th>Mountpoint</th>
<th>Storage Capacity</th>
<th>FS Type</th>
<th>Backed up?</th>
<th>Flushed?</th>
<th>Availability</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home</td>
<td>43 TB (20 GB / user)</td>
<td>ZFS</td>
<td>Yes</td>
<td>No</td>
<td>All Prince nodes (login, compute)</td>
<td>$HOME</td>
<td>/home/$USER</td>
</tr>
<tr>
<td>/scratch</td>
<td>1.1 PB (5 TB / user)</td>
<td>Lustre</td>
<td>NO</td>
<td>YES Files unused for 60 days are deleted</td>
<td>All Prince nodes (login, compute)</td>
<td>$SCRATCH</td>
<td>/scratch/$USER</td>
</tr>
<tr>
<td>Directory</td>
<td>Size</td>
<td>Filesystem</td>
<td>Storage Policy</td>
<td>Access Locations</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>----------------</td>
<td>------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/beegfs</td>
<td>500TB</td>
<td>BeeGFS</td>
<td>NO</td>
<td>Files unused for 60 days are deleted</td>
<td>All nodes (login, compute)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2 TB / user)</td>
<td></td>
<td>YES</td>
<td></td>
<td>$BEEGFS /beegfs/$USER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/archive</td>
<td>700 TB</td>
<td>ZFS</td>
<td>Yes</td>
<td>No</td>
<td>Only on login nodes $ARCHIVE /archive/$USER</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2 TB / user)</td>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/state/partition1</td>
<td>Varies, mostly &gt;100GB</td>
<td>ext3</td>
<td>NO</td>
<td>YES at the end of each job</td>
<td>Separate <em>local filesystem</em> on each compute node $SLURM_JOBTMP /state/partition1/$SLURM_JOBID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NYU HPC is currently in the process of a major upgrade:

- **2014 Q1**: A new system with 3200 Intel Ivy Bridge (c2013) cores in 160 nodes, to be named Mercer, is being installed
- **2014 Q2**: Union Square and Cardiac, which are at the end of their working lives, will be decommissioned
- **2014 Q2**: Most of the hardware comprising Bowery will be incorporated into Mercer to form a single, heterogeneous system.
- **2014 Q2**: Hydra will be integrated into Mercer
- **2014 Q2**: Lustre will be upgraded
- **2015 Q1**: Babar will reach end-of-life and be decommissioned

Consequently the information here is in a state of flux!

TODO: mention myquota.

The diagram below shows network and storage access of the NYU clusters
Some important aspects of the cluster setup are:

- The NYU clusters cannot be directly accessed from the internet: users must first log in to the bastion host `hpc.nyu.edu` (however, Internet connections from the clusters are supported).
- Each cluster consists of login nodes and compute nodes. The login nodes are for compiling code and preparing runs, actual computation should be run on the compute nodes by submitting it as a batch job (TODO link to how to).
- The `/scratch` filesystem is available on login and compute nodes, and is on a high speed network. `/scratch` is optimized for large-block I/O - please try not to use it for frequent, small I/O transfers. For these, we recommend using the node-local filesystem `/state/partition1` (TODO: link to more info about fs usage).
- The `/archive` filesystem is only available on the login nodes.
Each cluster has three primary filesystems, described (TODO link to storage page). Stakeholder users (TODO add a link to "how to become a stakeholder") also have access to a fourth filesystem, /work.