Storage Resources

Overview

Today’s research generates datasets that are increasingly complex, larger, and distributed. This makes modern research analysis, archiving, and sharing ever more challenging. The support for advanced techniques to transport, store, manipulate, visualize, and interpret large datasets is critical to advancing modern science.

The University’s Research Data Center provides data storage for active analysis on the high-performance computers (HPCs). Using central computing storage services and resources, University researchers, faculty researchers, and post-doctoral researchers are able to:

- Share research data in a collaborative environment with other UA affiliates on the HPC system
- Store large-scale computational research data
- Request additional storage for further data analysis

New storage is available with the new cluster. All storage is consolidated for all compute clusters, dramatically increasing capacity.

Qumulo Storage Array

In 2020 we implemented an all-flash based Qumulo Storage array with capacity of 2.29 PB of raw disk. Our leading requirement was fast, very fast. The old array had plenty of capacity but would make for slow calculations due to contention for the slower disk drives. So the new array was no longer a bottleneck for compute, and providing faster compute increases the efficiency of the supercomputers. The trade-off is that it is very expensive and we do not currently have an option for renting storage by the TeraByte.

We are developing a plan for Tier 2 storage which will be where research data is stored that is not in active compute. Currently we recommend using Google Drive until the Tier 2 is implemented.

Check Disk Quota - Disk quotas can be checked through https://portal.hpc.arizona.edu/portal by selecting Storage.

Allocations

<table>
<thead>
<tr>
<th></th>
<th>/home</th>
<th>/groups/PI</th>
<th>/xdisk</th>
<th>On node /tmp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backed up?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Lasts as long as your HPC account</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no duration of job</td>
</tr>
<tr>
<td>Maximum space</td>
<td>50GB</td>
<td>500GB</td>
<td>Up to 20TB</td>
<td>&lt; 800GB to 1.4TB</td>
</tr>
<tr>
<td>File count limit</td>
<td>600 files / GB</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

/xdisk

- xdisk has been simplified if you have been using it. No more tables to guess the size or duration
- The capacity has been greatly increased - the maximum is now 20TB
- You can specify less than 150 days, but 150 days is the default and the maximum duration you can select, except that you can renew once for a total of 300 days.
- The usage is detailed on this page.

Buy-In

- Purchase disk drives to be added to the storage system for dedicated group storage. The end of support for the current array is at the end of 2024. That is the end of guaranteed support. Beyond that depends on when we need to buy a replacement array.
- The cost estimate for adding a unit to the Qumulo storage array is $120,000 for 133 TB. This is more expensive and less granular than our previous clusters due to the speed trade-off mentioned above.
- For groups that need less than 133TB: using the free storage allocations or /xdisk option is the best option.
- This space is NOT backed up
- Files that need to be kept for 1-3 years can be offloaded to other platforms, like the University of Arizona Google Drive.

⚠️ We strongly recommend that you do some regular housekeeping of your allocated space. Millions of files are hard to manage for both the user and systems support. Archiving or using a tool like tar will help keep our disk arrays efficient and will make data migration quicker

Collaboration
Research computing has a contract with Globus. This service provides for efficient data movement and sharing. Tier 2 in this section has the details.