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 UArizona affiliates on the HPC system
- Store large-scale computational research data
- Request additional storage for further data analysis

Your storage is mounted as a filesystem and all the clusters have access to the same filesystems.

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 bottle-neck 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 have developed a plan for Tier 2 storage which will be where research data is stored that is not in active compute.

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

⚠️ 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.

<table>
<thead>
<tr>
<th></th>
<th>/home</th>
<th>/groups/PI</th>
<th>/xdisk/PI</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>Maximum duration</td>
<td>Persists as long as your HPC account</td>
<td>Persists as long as your faculty sponsor's HPC account</td>
<td>300 days. A new allocation may be created once the previous one expires.</td>
<td>Only accessible during for the duration of your 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>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

Buy-In

- Disk drives may be purchased 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. 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 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.

Rental Storage

Two new storage rental options have been rolled out as of 2022. More information can be found under:

- Tier 2 Storage
- Rental Storage

Collaboration

Research computing has a contract with Globus which facilitates efficient data movement and sharing. Information on Globus and other data transfer method can be found in our Transferring Data page.