Storage Resources

Overview

Today’s research generates datasets that are increasing exponentially in both complexity and size, making their 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 for storage of computational data. 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
- Store large-scale computational research data
- Request additional storage for further data analysis

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

Storage Array

A DDN SFA12KX storage array is the primary storage for all the systems. With a sustained performance of 44 GB/second raw I/O and 2PB of raw disk, expandable to 5PB.

Allocations

<table>
<thead>
<tr>
<th></th>
<th>/xdisk</th>
<th>/extra</th>
<th>/home+/pbs</th>
<th>/tmp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backed up?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Lasts as long as your HPC account</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no duration of job</td>
</tr>
<tr>
<td>Maximum space</td>
<td>200GB to 1TB limit of 45 days can renew once</td>
<td>200GB</td>
<td>14GB</td>
<td>&lt; 800GB</td>
</tr>
<tr>
<td>File count limit 600 files / GB</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

/xdisk

- /xdisk has been simplified if you have been using it. Nor more tables to guess the size or duration
- The capacity has been greatly increased - the default size is 200GB and the maximum is 1TB
- You can specify less than 45 days, but 45 days is the default and the maximum duration you can select, except that you can renew once for a total of 90 days.
- The usage is detailed on this page.

/extra

/extra is something new. When you log into Ocelote for the first time you will get an additional allocation of 200GB. It will take an hour or more to show up, but once it does, it is permanent and you can use it for anything. Remember that it is not backed up, unlike /home.

Buy-In

- Purchase disk drives to be added to DDN storage system for dedicated group storage. The end of support for the current array is at the end of 2020 so renting space will usually be a better option.
- Cost estimated at $6,600 for 43TB formatted space, with co-terminated support contract (new purchase required, market pricing determines actual cost) 43TB formatted is the only increment for adding drives to the DDN array (in sets of 10 disks)
- For groups that need less than 43TB, renting storage is probably the answer. Contact hpc-consult@list.arizona.edu
- This space is NOT backed up

Rental

- $39/TB/year - current approved rate - with annual, renewable payments, market pricing determines rate
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. Reading or writing millions of files will likely cause response time issues for other users. Please use our Consultants for ideas on efficient use of storage.

Collaboration

Research computing is implementing an iRods configuration. This resource will provide large capacity for the location of large datasets. iRods is implemented with policies for the retention of data.

Benefits of iRODS

- iRODS enables data discovery using a metadata catalog that describes every file, every directory, and every storage resource in the data grid.
- iRODS automates data workflows, with a rule engine that permits any action to be initiated by any trigger on any server or client in the grid.
- iRODS enables secure collaboration, so users only need to log in to their home grid to access data hosted on a remote grid.
- iRODS implements data virtualization, allowing access to distributed storage assets under a unified namespace, and freeing organizations from getting locked in to single-vendor storage solutions.

This section has details on how to use iRODS