



UNIVERSITY OF
Nebraska®

HOLLAND COMPUTING CENTER

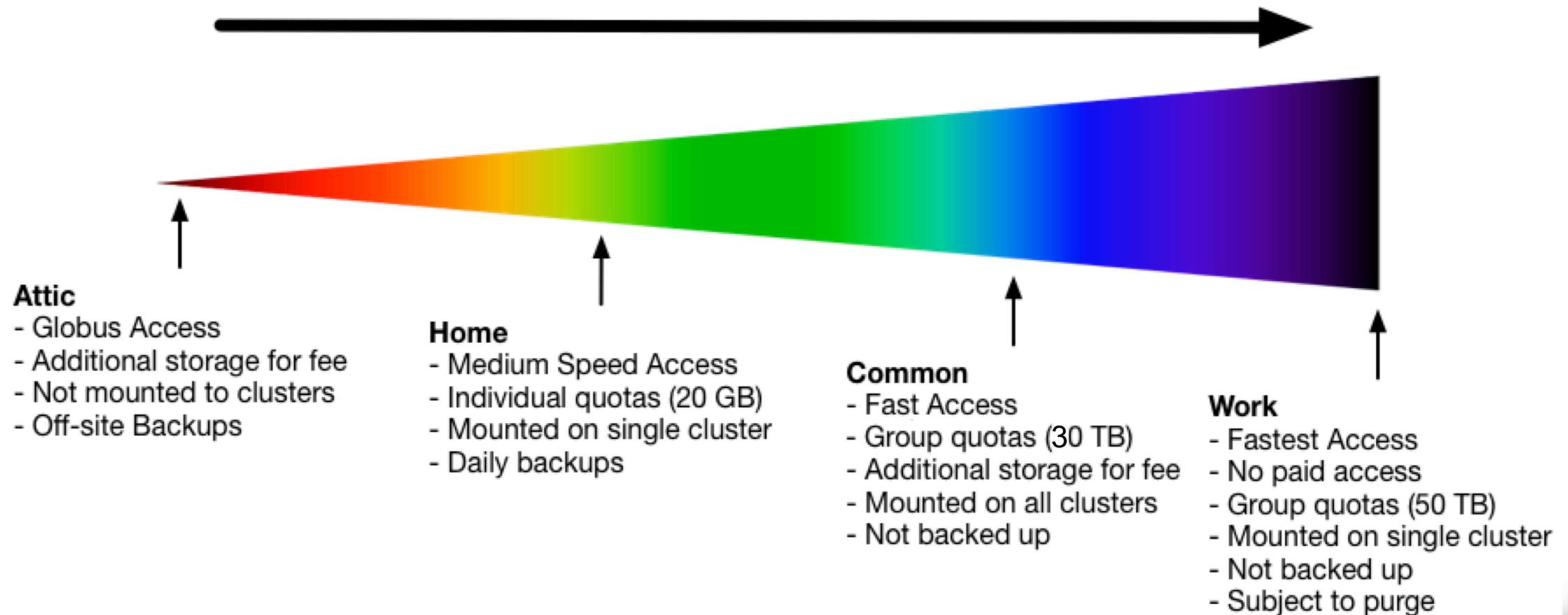
High Performance Computing Core

Data Management on HCC

- Data storage on Crane
 - Different file systems and quotas
- Transferring files to and from Crane
 - Command-line file transfers (scp)
 - GUI applications for file transfers (introduction to Globus)

Data Storage: Home vs Work vs Common vs Attic

- Faster File Access (higher bandwidth / IOPS)
- Ease of Access
- Diminishing Redundancy (backups)



Data Storage - Attic

Attic

- Data backed up in multiple locations (Lincoln and Omaha)
- Allocations for purchase
- Not mounted on the clusters
- Accessible through Globus Connect
- Intended as near-line data archive

Data Storage – Home vs. Work vs. Common

\$HOME

- /home/[group]/[username]
- 20GB / user
- Read-only on worker nodes (no active I/O for running jobs)
- NO purge policy, backed up
- Examples: source code, program binaries, configuration files

\$WORK

- /work/[group]/[username]
- 50TB / group
- Intended for fast I/O for running jobs
- Short-term scratch space
- 6-month purge policy – not backed up!
- Examples: input/output data files for running jobs

\$COMMON

- /common/[group]/[username]
- 30TB / group (larger quotas for purchase)
- Mounted on both clusters (Crane and Rhino)
- NOT intended for fast I/O for running jobs
- Accessible to worker nodes if you check out the license: `#SBATCH --licenses=common`
- NO purge policy but not backed up!
- Examples: items that need to be accessed from both clusters such as reference databases or shared data files

\$HOME/\$WORK/\$COMMON Usage Best Practices

- **Avoid large numbers of files**
- **Storage resources are finite**
 - Be judicious about the data you store.
- **Backup valuable data**

Sensitive and Protected Data

HCC currently has *no storage* that is suitable for **HIPAA** or other **PID** data sets. Users are not permitted to store such data on HCC machines.

Exercises

1. Log into Crane and find out your group's disk usage:

```
$ hcc-du
```

2. Find out what files are scheduled to be purged (dormant for 24 weeks) from your \$WORK directory: *(use space bar to page through results and `q` to quit)*

```
$ hcc-purge -l
```

3. Cd to your \$WORK directory, display your current path, and list the contents:

```
$ cd $WORK
```

```
$ pwd
```

```
$ ls
```

4. Repeat Exercise 3 above for your \$HOME and \$COMMON directories.

Transferring Files – Command Line

scp

- Mac *terminal* and Windows 10 (version 1903 and later) *Command Prompt*
- Usage:
`scp user@host:source_file user@host:target_file_or_dir`
- Example:
`scp my_file.txt demo01@crane.unl.edu:/work/demo/demo01/`
- For transferring files between your computer and the cluster, always **run the scp command on your computer** (not on the cluster)

Exercises (1&2 are optional)

1. Use scp to copy a file from your computer to your work directory on Crane:

```
$ scp my_file.txt demo01@crane.unl.edu:/work/demo/demo01/
```

2. Copy the job-examples/README.md file from your work directory on Crane to your computer:

```
$ scp demo01@crane.unl.edu:/work/demo/demo01/job-examples/README.md .
```

3. You can use the `-r` option to scp (recursively) an entire directory. Copy the job-examples/matlab directory from your work directory on Crane to your work directory on Rhino.

```
$ scp -r $WORK/job-examples/matlab demo01@rhino.unl.edu:/work/demo/demo01/
```


Transferring Files – GUI Applications

- **Transfer files using an SCP client**
 - WinSCP (<http://winscp.net>) – Windows
 - Cyberduck (<http://cyberduck.io>) – MacOS
 - Filezilla (<http://filezilla-project.org>) – Windows, MacOS and Linux
- **Globus Connect (<http://globus.org>)**
 - Fast, secure and robust transfers with user-friendly web interface
 - Uses the High-Speed transfer nodes by default
 - Can transfer directly between clusters, Attic and personal machine
 - Other features: **file sharing**, rsync transfers, command-line tool (globus-cli)
 - https://hcc.unl.edu/docs/handling_data/data_transfer/globus_connect/

Exercises

1. Log into globus.org using your *My.UNL* credentials
2. **Transfer the job-examples directory from Crane to Rhino**
 - Go to the *File Manager* tab and search for the ***hcc#crane*** and ***hcc#rhino*** endpoints.
You will have to activate each endpoint using your HCC credentials – expires every 7 days
 - Enter the source and target paths (/work/[group]/[username])
 - Select the directory to transfer (job-examples)
 - Start the transfer
3. **Download and install the *Globus Connect Personal* application. Transfer a file from your computer to Crane.**
 - In the *File Manager* tab, click in the *Collection* text box. Then click on “**Install Globus Connect Personal.**”
 - Select the image for your operating system, download and follow installation instructions.
 - In the *File Manager* tab in the Globus web portal, search for the endpoint you just created.
 - Choose a file from your computer and transfer it to your work directory on hcc#crane.

Workflow Tips

- **Test/Develop your workflow on a truncated set of data**
 - Use tutorial data if provided
 - Otherwise, make your own
- **Run commands in an interactive job first**
- **Try different combinations of SBATCH options to find what works best**
 - Run multiple jobs on your truncated data to see which run faster
- **Always check output/error files - even for apparently successful jobs**
 - Were they created?
 - What do they contain?
- **Check what resources the job actually used**
 - `sacct -j <job_number> --format Elapsed,MaxRSS`

What to do if you're stuck

- **Read the HCC Documentation**

- <http://hcc.unl.edu/docs>
 - If the documentation doesn't answer your question, leave a comment or email to let us know!
- Use **man** or **--help**
- Look at your output and error files!!

- **Consult Google**

- Useful for application specific errors or general usage

- **Contact Us**

- Open Office Hours: <http://hcc.unl.edu/OOH>
- Email hcc-support@unl.edu



**LET
ME
GOOGLE
THAT
FOR YOU**