



# NCCS Brown Bag Series Introduction to ADAPT

September 2020

# Welcome!

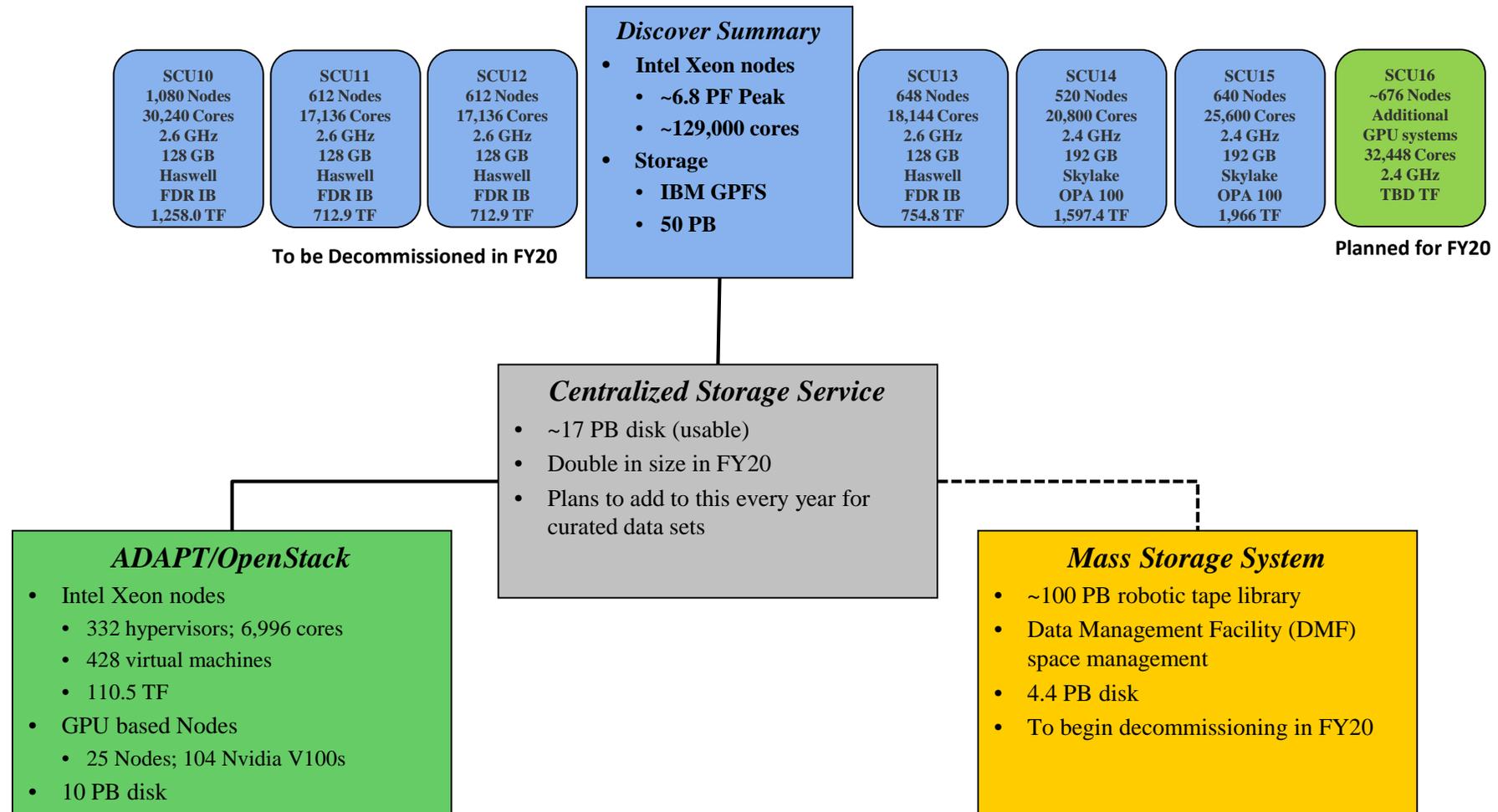
- Congratulations on your NCCS ADAPT user account!
- This presentation will guide you through all the basics of using NCCS's ADAPT computing, storage, data analysis systems, and services. Links are included for some topics where in-depth discussions are available. [New users are recommended to read through the entire package prior to login.](#)
- We are here for YOU. With an emphasis on enabling science and providing user-oriented systems and services, we encourage you to ask a lot of questions of NCCS Support!

Email to [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov)

# Introduction to ADAPT Roadmap

- **NCCS ADAPT Platform**
- **Systems and Components**
- **System Login**
- **Passwords**
- **Files and Data**
  - Discover File Systems
  - Centralized Storage System (CSS)
  - Data Management Plans
  - Managing Your Files
  - Data Sharing
  - Quota Limits
  - File Transfer
- **Compute**
  - Requesting Virtual Systems
  - Cron Jobs
  - Modules
  - Compilers
  - Running Compute Jobs via Slurm
  - Debugging and Profiling Tools
  - Licensed Application Software
  - Open Source Software Packages
- **Windows**
- **GPU Cluster**
- **JupyterHub**

# NCCS Computing and Storage Platform



# Systems and Components

- Computing
  - [ADAPT](#): A Linux managed virtual machine (VM) environment most closely resembling a platform as a service (PaaS) cloud
  - [GPU](#): A cluster of 22 GPU nodes, built for accelerating artificial intelligence (AI), machine learning (ML) and deep learning (DL) workloads
  - [Gitrepo](#): Store code in our Git repository
- Centralized Storage
  - [CSS](#): 30 PB for curated NASA and related datasets
- Data Services
  - [Dataportal](#): Provides public access to some NCCS data through various services

# System Login

- **Standard login mode**

`ssh -Y user_id@adaptlogin.nccs.nasa.gov`

**PASSCODE:** *If using a soft token (on your phone): Enter your PIN in the app and the eight-digit TOKENCODE at the PASSCODE prompt*

*If using a hard token: PIN followed by the six-digit TOKENCODE*

**Password:** *your LDAP Password provided by NCCS user services*

- **PIV login mode**

Add the following to your `.ssh/config`:

*host adaptlogin adaptlogin.nccs.nasa.gov*

*PKCS11Provider=/usr/lib/ssh-keychain.dylib*

`ssh -XY user_id@adaptlogin.nccs.nasa.gov`

**PIV login mode**

# A word on Passwords

- **RSA token passcode**

Your given PIN followed by a six-digit TOKENCODE shown in your SecurID key to access the NCCS login system.

[Call Enterprise Service Desk 1-301-286-3100 for help.](#)

- **NCCS LDAP password**

Password to access the NCCS systems.

[NCCS Support is able to help with LDAP password issues.](#)

- **NCCS Windows password**

Password to access the NCCS Windows systems.

[NCCS Support is able to help with Windows password issues.](#)

# Files and Data

- ADAPT File Systems
- Centralized Storage System (CSS)
- Data Management Plans
- Managing Your Files
- Data Sharing
- Quota Limits
- File Transfer

# ADAPT Filesystems

- Accessible from all ADAPT login and compute nodes, including the GPUs
  - `$HOME` or `/home/user_id`
    - Quota controlled. User disk space limit is 5 GB.
    - Fully backed up. Ideal for storing source code and scripts.
  - `$NOBACKUP` or `/att/nobackup/user_id`
    - Quota controlled for both the disk space and number of files (inodes)
    - User limit is 5.5 TB
    - **NOT backed up**
  - Use the command “`showquota -h`” to check usage on `$HOME` and `$NOBACKUP`. Details are available on the NCCS website.
  - `/att/pubrepo` - contains shared public datasets local to ADAPT

# Centralized Storage System (CSS)

- CSS has 15 PB of storage, 15 PB more coming online Sept 2020
- Provide access to large NASA curated data sets from our HPC, Cloud, GPU, and Dataportal environments
- Provide data discovery and usage reporting to reduce data duplication and facilitate data deletion
- Manage the data lifecycle through Data Management Plans and policies



# Centralized Storage System (CSS) Data

- Curated data products available on CSS include:
  - CMIP5: 105 TB
  - CREATE-IP: 79TB
  - GeoMIP: 14 TB
  - GOES
  - ICEBridge: 2 TB
  - ICESat: 8 TB
  - ICESat-2: 161 TB
  - Landsat: 186 TB
  - MAIAC: 107 TB
  - MERRA: 87 TB
  - MERRA2: 320 TB
  - CREATE-IP: 79TB
  - NEX GDDP: 11 TB
  - NEX DCP30: 11 TB
  - NGA: 4.5 PB
  - With more to come

# Data Management Plans

- Four types of data:
  - Input – store on Discover or, if a curated dataset, on CSS
  - Intermediate – data created during software runs, store on Discover project space:
    - Not permanent
    - Not to be shared publicly
    - Could be restart files, research results, temporary files
  - Final – used for publications, shared with the science community or collaborators, could be input to other science programs – store on CSS
  - Software – save in a Git repository for re-use

# Managing Your Files

- \$NOBACKUP is NOT backed up. It is your responsibility to copy valuable data to either \$HOME or to remote systems
- \$NOBACKUP or /att/nobackup/*user\_id* is a symlink that points to the actual disk your nobackup directories reside, e.g., /att/gpfsm/briskfs01/*user\_id*. **ALWAYS use the symlink in your scripts to specify paths**, because the actual path may be changed due to disk augmentations or system events.
- Use the “**showquota -h**” command to report your quota on ADAPT

# Data Sharing

- A common way to share files/directories with group members and others is to change permissions using *chmod* command

```
$ ls -l
drwx----- 2 cpan2 k3001 8192 2013-01-07 16:17 tmp/
$ chmod -R go+rx tmp | ls -l
drwxr-xr-x 2 cpan2 k3001 8192 2013-01-07 16:17 tmp/
$ chmod -R o-rx tmp | ls -l
drwxr-x--- 2 cpan2 k3001 8192 2013-01-07 16:17 tmp/
$ groups cpan2
cpan2 : k3001 k3002
$ chgrp -R k3002 tmp | ls -l
drwxr-x--- 2 cpan2 k3002 8192 2013-01-07 16:17 tmp
```

- **Do NOT make files/directories world-writable.** If you have a specific need to share data with group members or others, send a ticket to NCCS Support and we will help you.

# Quota Limits

- Two kinds of quotas are enforced:
  - **Limits on the total disk space occupied** by a user or a group's files on either \$HOME and \$NOBACKUP
  - **Limits on how many files (inodes)** a user can store on \$NOBACKUP, irrespective of size. For quota purposes, directories count as files
- Two types of quota limits are in place:
  - **Hard limits** – can never be exceeded. Any attempt to use more than your hard limit will be refused with an error
  - **Soft limits** – can be exceeded temporarily. When you exceed your soft limit, you can continue to work normally -- but the countdown of the grace period (**7 days**) begins and you have to bring usage back below the soft limit value. Failure to do so within the grace period will cause the soft limit to become a hard limit. See the NCCS website for “showquota” examples.

# File Transfer to and from ADAPT

- To copy data from a remote system to NCCS, a user must use the **Bastion Service Proxy Mode**
- For command line users:
  - **Initiating commands from a remote system:**
    1. Make sure to first set up for the Proxy Mode, i.e., the `$HOME/.ssh/config` file is created on the remote system (See Proxy Mode on the System Login slide)
    2. Use either **scp**, **sftp**, or **rsync** from the remote system:
      - `scp -r user_id@adaptlogin.nccs.nasa.gov:~/mydir .`
      - `sftp user_id@adaptlogin.nccs.nasa.gov`
      - `rsync -auPv ~/mydir user_id@adaptlogin:/att/nobackup/user_id/mydir`
  - **Initiating commands from ADAPT to pull/push data from a remote system is also possible**
- For WinSCP users:
  - See the NCCS website for written instructions and an instructional video
- NGA data needs special permission to be copied out, see your NGA agreement

# ADAPT Compute Systems

- Requesting Virtual Systems
- Default Shell
- Modules
- Compilers
- Running Compute Jobs via Slurm
- Monitoring Usage with Ganglia
- Debugging and Profiling Tools
- Licensed Application Software
- Open Source Software Packages

# Requesting Virtual Systems

- ADAPT provides the following sized virtual systems:

Name	Virtual CPUs	Memory
adapt.nano	1	6
adapt.small	2	12
adapt.quarter	4	48
adapt.half	10	122
adapt.full	20	245

- To request virtual systems, please submit a ticket to [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov)

# Default Shell

- “echo \$SHELL” to check your default shell, default is bash
- To change the default shell, contact NCCS Support

Shell	Startup files to edit
sh or ksh	\$HOME/.profile
bash	\$HOME/.bashrc if it exists; or \$HOME/.bash_profile if it exists; or \$HOME/.profile if it exists (in that order)
csh	\$HOME/.cshrc
tcsh	\$HOME/.tcshrc if it exists; or \$HOME/.cshrc if it exists (in that order)

# Modules

- The “module” command allows you to choose compilers, libraries, and packages to create/change your own personal environment
- When you initially log into ADAPT, **no modules** are loaded by default
- The module commands can be run in your shell startup file, your job script, or at the command line

Common Commands	Explanation
module spider	Display a complete list of available modules
module avail (av)	Display a complete list of applicable modules
module list	Display loaded modules
module load <i>module_name1</i> ...	Load new modules
module purge	Unload all loaded modules
module show <i>module_name</i>	Display the environmental variables set by the module

# Conda Environment

- Anaconda v3 is available through the modules environment
  - ADAPT – module load anaconda3
  - Legacy ML Nodes - module load anaconda3
  - GPU cluster – module load anaconda
- Users can customize their conda environments

# Compilers

- To accommodate the needs of a broad range of user groups, multiple versions of compilers from different vendors are provided
- Run `module avail` to see the versions available

Compiler	Access with:
GNU (gcc)	<code>module load gcc-version, e.g. gcc-9.1</code>
Intel	<code>module load intel-version e.g. intel-19.0.2.187</code>
PGI	<code>module load PGI-version, e.g. PGI-19.10-OpenMPI-CUDA</code>
CUDA (GPU cluster)	<code>module load CUDA-version, e.g. CUDA-10.2</code>

- While ADAPT does not have a fast interconnect for traditional HPC, both Intel MPI and OpenMPI are available through modules

# Running Compute Jobs via Slurm

- Slurm is a distributed workload management system that handles the computational workload on Discover.
- Some ADAPT communities, including the GPU cluster, utilize Slurm to access their compute nodes
- Only one QoS is configured for Slurm jobs on the GPU Cluster, no specification is required
- In order to run multi-node Slurm jobs, you have to set up a [\\$HOME/.ssh/authorized\\_keys](#) file.

# Common Slurm Commands

- Use Slurm commands to request both interactive and batch access to Discover computational resources.

QoS	Explanation
sbatch	Submit a batch job script for queueing and execution
salloc/xalloc	Submit an interactive job request
srun	Run a command within an existing job, on a subset of allocated resources
scancel	Cancel a queued or running job
squeue	query the status of your job(s) or the job queue

- View the Slurm instructional video on [nccs.nasa.gov](https://nccs.nasa.gov) for a detailed explanation of how to use Slurm on Discover

# Licensed Application Software

- IDL is installed on ADAPT systems:

- IDL:

`$ /usr/local/bin/idl`

Some groups bring their own MATLAB licenses but there is no general availability for MATLAB

# Commonly Used Open Source Software

- Python : Python distributions (2.7, 3.x) for scientific computing
- Module Environment:
  - R
- /opt software can vary from system to system but may include:
  - R
  - gdal
- User maintained:
  - Users are welcome to install their own software locally or request installs by emailing [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov)

# Windows

- ADAPT provides access to Windows VMs for applications like Esri ArcGIS Desktop and ArcGIS Pro.
- Send email to [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov) to request an account
- Access is through guacamole from your web browser:
  - <https://guac.nccs.nasa.gov>
- ADAPT and CSS filesystems are available from the Windows systems
- See the NCCS website for additional information and instructional videos

# GPU Cluster

- ADAPT provides access to 25 GPU nodes for AI/ML workflows

System	Cores	Memory	GPUs
ml4[101/102]	24	512	4 x 16 GB
ml8101	24	512	8 x 16 GB
gpu[001-0022]	20	768	4 x 32 GB

- Send email to [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov) to request access
- Access as follows:
  - ssh adaptlogin.nccs.nasa.gov
  - ssh gpulogin1
- Access to compute is through Slurm
- See the NCCS website for more information and instructional videos

# GPU Cluster Software

- All software is available through the modules and conda environments:
  - Python
  - Tensorflow
  - Pytorch
  - XGBoost
  - RapidsAI
  - Dask-cudf, dask-xgboost, dask-cuda
  - EarthML/PyViz
  - Scikit-Learn

# JupyterHub

- ADAPT provides access to two JupyterHub sites for Python notebooks and remote visualization, with more under development:
  - ICESat-2 – <https://jh-icesat.nccs.nasa.gov>
  - Legacy ML Nodes – <https://jh-ml.nccs.nasa.gov>
- Send email to [support@nccs.nasa.gov](mailto:support@nccs.nasa.gov) to request an account
- Access is through the above websites
- See the NCCS website for additional information and instructional videos

# Questions?

Help is always available by emailing  
[support@nccs.nasa.gov](mailto:support@nccs.nasa.gov)

Read our Ticket Submission Guidelines for  
information to include:

<https://www.nccs.nasa.gov/about-us/contact-us/ticket-guide>