

# ORION & NIAGARA QUICK START

Jan, 31 2020

- ❑ Orion System basics
  - ❑ Overview
  - ❑ Accounts
  - ❑ Allocations
  - ❑ Modules
  - ❑ Compiling and running MPI codes
  - ❑ Using the batch system
  - ❑ File systems
  - ❑ Documentation
  - ❑ Reporting problems
- ❑ Niagara System basics
  - ❑ Overview
  - ❑ File systems
  - ❑ Data Migration
  - ❑ Documentation
  - ❑ Reporting problems

# Quickstart Overview

- ❑ **Orion** *was a giant huntsman whom Zeus placed among the stars as the constellation of Orion. Also one of NOAA's specially equipped fleet of aircraft which plays an integral role in hurricane forecasting. Data collected during hurricanes by these high-flying meteorological stations help forecasters make accurate predictions during a hurricane and help hurricane researchers achieve a better understanding of storm processes, improving their forecast models.*
- ❑ **Based on the Intel Skylake (same as Hera)**
  - ❑ 20 cores/socket, 2.4 GHz, 2 sockets/node
  - ❑ 1,792 Compute nodes (192GB)
  - ❑ 8 Big memory (384GB) nodes
  - ❑ 4 Login nodes / Front-Ends (192GB)
  - ❑ 4 Data Transfer Nodes
  - ❑ 2 Development Nodes
- ❑ Mellanox HDR-100 Infiniband (a partially blocking fat-tree similar to Hera)
- ❑ File systems: /work
  - ❑ Lustre, 9PB Capacity (about .5x of Hera's /scratch1 + /scratch2), >53GB/s reads and writes
- ❑ File systems: /home
  - ❑ NFS, approximately 37TB, 10GB per user quota

## System Overview

- ❑ MSU's HPC Account Management System and onboarding process will be used for users working on NOAA's R&D HPC projects.
  - ❑ Custom web forms have been provided to support NOAA's projects and project members.
  - ❑ MSU accounts will be managed completely independent of NOAA RDHPCS Accounts.
  - ❑ The web based account management system has been created so that each Principal Investigator (PI) can manage the membership of their project(s).
  - ❑ In the event that the PI is not available, then the Portfolio Manager (PM) may manage the membership of any project under their portfolio.
  - ❑ The system also allows for PIs and PMs to request accounts for new users, by filling out the web based Computer Access Request form.
- ❑ For the initial creation of accounts, all Jet and Hera users have had accounts requested. Users will have until **February 29, 2020** to complete their onboarding. Afterward, new account requests will have to be submitted.

## User Accounts

- ❑ The NOAA RDHPCS Allocation committee has not yet defined which Portfolios nor what their allocations will be on Orion .
- ❑ To get started, Projects within Portfolios which have Hurricane Supplemental work and which have had recent activity on Hera, were added to Orion.
- ❑ Emails have been sent to Portfolio Managers with Project/Allocation/Quota information so they can begin to manage (add/remove/change) their Projects/Allocations/Quotas on Orion.
- ❑ NOAA RDHPCS Allocation committee decisions in the future could affect Orion allocations.
- ❑ To see a list of the initial Portfolios and Projects, please see:

[https://oriondocs.rdhpcs.noaa.gov/wiki/index.php/Managing\\_Allocations#Current\\_Allocations\\_on\\_Orion](https://oriondocs.rdhpcs.noaa.gov/wiki/index.php/Managing_Allocations#Current_Allocations_on_Orion)

- ❑ There are 4 Login nodes, orion-login-[1-4]
- ❑ SSH dual-factor login via DNS round-robin
  - ❑ SSH <username>@orion-login.hpc.msstate.edu
  - ❑ To authenticate you will need to type in your password and Duo pin number

## Orion Logging in

- ❑ We are assuming that users are already familiar with using Jet and/or Hera.
- ❑ Orion is set up to be as close as possible to Hera and Jet.
- ❑ So it will be a very familiar environment.
- ❑ The key differences are being emphasized in this presentation.

## Transitioning to Orion

- ❑ Nodes are named differently than you are used to.
- ❑ However, all node names start with “Orion-”
  - ❑ Orion-login-[1-4]  
*Interactive user login nodes and cron services*
  - ❑ Orion-dtn-[1-4]  
*Data transfer tools: scp, sftp, rsync (via ssh), and Globus*
  - ❑ Orion-devel-[1-4]  
For managing “contrib” software directories and modules
  - ❑ Orion-[01-25]-[01-72]  
*25 compute racks with 72 compute nodes per rack*

## Node Name Convention



- ❑ Based on LMOD
  - ❑ Similar to Hera and Jet
- ❑ Designed to prevent loading incompatible modules
- ❑ Not all modules appear with *module avail*
  - ❑ Use *module spider* to see everything that is available
- ❑ NOAA specific tools, workflow management software, and reporting scripts will be managed under “/apps/contrib”
  - ❑ Users will need to load the “contrib” module to see these modules
  - ❑ Once this module is load other contrib modules may be loaded
    - ❑ noaatools, rocoto, nceplibs, gfdl, etc...

## Modules

- ❑ Intel compilers, Intel MPI, PGI compilers, OpenMPI are available
  - ❑ Compilers: **intel 2018.4, intel/2019.5, intel/2020**
  - ❑ MPI: ***impi/2018.4, impi/2019.6, impi/2020***
  - ❑ PGI: ***pgi/2019(19.9) is available***
  - ❑ OpenMPI: ***openmpi/4.0.2***
- ❑ ***Allinea tools***
  - ❑ ***“Forge” and “performance reports” for profiling***
  - ❑ ***“DDT” for debugging***

## Compiling and Running MPI Codes and Tools

- ❑ Skylake processor has 2 AVX512 units, each capable of 8 DP operations
  - ❑ Also has mult-add capability
  - ❑ If your code can use these instructions, it will perform very well
- ❑ Intel recommended options
  - ❑ `-g -O3 -ftz -traceback -fpe0 -xHost`
  - ❑ `-g -O3 -ftz -traceback -fpe0 -xHOST -axcore-avx512`
  - ❑ `-g -O3 -ftz -traceback -fpe0 -xHOST -axcore-avx512 -qno-opt-dynamic-align`
- ❑ PGI recommended options
  - ❑ `-g -O2 -tp skylake`

## Compiling and Running MPI Codes

- ❑ The batch system chosen by MSU is Slurm. Much of their configurations are similar to NOAA's.
- ❑ We are assuming you are already familiar with Slurm from your NOAA experience.
- ❑ We have installed a number of NOAA's tools on Orion under noaatools in contrib (module load contrib noaatools).

## Using the Batch System

- ❑ QOSes Available:
  - ❑ Batch: priority=20, max node count 500, max walltime of 8 hours.
  - ❑ Urgent: priority=40, max node count 500, max of one job per project, max walltime of 8 hours.
  - ❑ Debug: priority=30, max node count 20 (will be 500), max of 2 jobs per user, max walltime of 30 minutes.
  - ❑ Novel: priority=50, min node count of 501, max walltime of 8 hours. (Runs after maintenance)
  - ❑ Windfall: priority=1, max node count 500, max walltime of 8 hours.

## Using the Batch System: QOSes

- ❑ Partitions Available (subject to change):
  - ❑ Orion (max 500 nodes)
    - ❑ QOS: batch,urgent,windfall,novel
  - ❑ debug (default, max 20 nodes)
    - ❑ QOS: debug,batch,windfall
  - ❑ bigmem (max 8 nodes)
    - ❑ QOS: debug,batch,urgent,windfall
  - ❑ Service
    - ❑ QOS: batch,urgent,windfall

- ❑ Default Partition is debug
  - ❑ Max wall time is 30 minutes
  - ❑ Max nodes is 20
  - ❑ Specify orion partition for more nodes and longer wall time
- ❑ Orion partition has higher limits
  - ❑ Max wall time is 8 hours
  - ❑ Max nodes is 500
- ❑ Salloc leaves you on the submitting node
  - ❑ Use srun to get an interactive job
    - ❑ `srun -A noaatest -p service -q batch [--x11] --pty bash`
- ❑ Thread layout is still being worked on
  - ❑ Hyperthreading is not enabled
- ❑ [t]csh currently is not configured correctly
- ❑ module load contrib noaatools for:
  - ❑ `saccount_params`, `shpcrpt`, `layout.pl`, `arbitrary.pl`, etc

- ❑ Lustre
  - ❑ Quotas are tracked by “Project ID” (The same as group ID and directory name)
    - ❑ The Project ID is assigned to top-level dirs and will be inherited for all new subdirs.
  - ❑ Tracking and enforcement is based on maximum capacity.
  - ❑ Please DO NOT run “find”, “du” (or anything) similar to produce reports
  - ❑ Your initial disk allocation is based on your project’s core-hour allocation



- ❑ `saccount_params` will give you information about your projects, partitions, QOSes and project directory (see alternative below)
  - ❑ Must module load `contrib noatools`

- ❑ Finding your new Project directory:

```
/bin/ls -ld /work/noaa/<project>
```

- ❑ For example:

```
/bin/ls -ld /work/noaa/noaatest  
drwxrwsr-x 16 root noaatest 4096 Jan 14 09:49 /work/noaa/noaatest/
```

## Where is my Project Space?

## ❑ Known Issues

- ❑ There is currently limited RDHPCS staffing to support Orion. This may cause delays in responding to Help Desk support tickets.
- ❑ Currently no restricted data (rstprod) is allowed to be copied to Orion. We're still working through NCEP's rstprod data from both a technical and policy standpoint
- ❑ CRON services are only currently available on orion-login-1
- ❑ Still need to deploy reporting tools for reporting system and disk usage (shpcrpt, Lustre quota report, etc...) under the noatools contrib directory.
- ❑ Still working through some Batch and Fairshare configuration issues.
- ❑ Shared user accounts (i.e. Role account), for managing automated model workflows, are not currently allowed on Orion.
- ❑ No remote desktop support (i.e. X2GO) is currently available on Orion.
- ❑ No externally initiated unattended data transfers (via scp & rsync) are currently allowed on Orion

- ❑ Finding your quota and usage
- ❑ The `saccount_params` command is ready but you must first load the `contrib` and `noaatools` modules
- ❑ Quotas are “project” based
  - ❑ Run the “`id`” command to get your **project ID number** (not the name)
  - ❑ `lfs quota -p <project ID number> /work`
- ❑ User and Group usage (capacity and file count) is tracked but not limited
- ❑ You can also find your usage across all projects:
  - ❑ `lfs quota -u <username> /work`
  - ❑ `lfs quota -g <groupname> /work`

## What is my Quota and Usage

## □ Simple rules of thumb

- Space is limited, so please actively manage your data. Only keep what is critical to your project's mission.
- Keep source code and critical configuration files on /home
- Move data back to Niagara and then back up critical data to NOAA's HPSS. (Orion's /work file system is *not backed up!*)
- Tar up old small files (or delete them) to free up space and stay under your file count quota
- Large files are still optimal for HPC batch job performance.

# File Systems

❑ Primary documentation at:

❑ <https://intranet.hpc.msstate.edu/helpdesk/resource-docs/>

Note: Active MSU account required for accessing documentation

❑ Supplementary documentation at:

❑ [https://oriondocs.rdhpcs.noaa.gov/wiki/index.php/Main\\_Page](https://oriondocs.rdhpcs.noaa.gov/wiki/index.php/Main_Page)

Note: Active NOAA email account required for accessing documentation

❑ Help: [rdhpcs.orion.help@noaa.gov](mailto:rdhpcs.orion.help@noaa.gov)

Documentation and Reporting Problems

# Orion Questions?

- ❑ **Niagara** *falls sits on the border between the U.S. and Canada. NOAA's Niagara system sits on the border between NOAA's R&D HPCS program and the rest of the world. The Niagara system is intended to be a collaborative location where data can be securely copied to and from any location, by any authorized user. It can also be used as a service to post process, analyze, and disseminate R&D data to NOAA's collaborators around the globe.*
- ❑ **Based on the Intel Skylake (similar to Hera)**
  - ❑ 20 cores/socket, 2.5 GHz, 2 sockets/node
  - ❑ 12 Interactive / Login nodes
  - ❑ 2 Data Transfer Nodes (DTNs) available from trusted (pre-approved) NOAA and non-NOAA sites
  - ❑ 2 Untrusted Data Transfer Nodes (UDTNs) available from anywhere on the internet
  - ❑ 2 Web servers
- ❑ Mellanox EDR Infiniband (a partially blocking fat-tree similar to Hera)
- ❑ File systems: /collab1
  - ❑ Lustre, ~2PB Capacity, >27GB/s reads and writes
  - ❑ Includes: apps, home, data, and data\_untrstuted directories

- ❑ 12 Interactive nodes, nfe-[01-12]
- ❑ CAC log in
  - ❑ sshg3 bastion-niagara.princeton.rdhpcs.noaa.gov
  - ❑ sshg3 bastion-niagara.boulder.rdhpcs.noaa.gov
- ❑ RSA log in
  - ❑ ssh niagara-rsa.princeton.rdhpcs.noaa.gov
  - ❑ ssh niagara-rsa.boulder.rdhpcs.noaa.gov

Note: Before being able to transfer data via the DTNs, a user must first login to Niagara.

## Niagara Logging in



- ❑ Niagara is intended to be used as a data collection, post processing, and data dissemination point.
- ❑ All data written to Niagara is transient.
- ❑ Data intended for long term storage should be either written to HPSS or moved off of R&D HPC systems to another NOAA storage location.
- ❑ Niagara's user environment and application software is set up to be identical to Hera and Jet. So it will be a very familiar environment.
- ❑ Any user with an active RDHPCS user account may login and use Niagara.
- ❑ CRON services are available on all Interactive / Login nodes for automating workflows.

## General Niagara information

- ❑ Lustre
  - ❑ There are no “project” based directories. All directories are individual user based.
  - ❑ Upon initial login a home directory is created. This directory has a 10GB storage limit and should not be used for data.
  - ❑ Upon initial login two data directories are created under “/collab1/data/\$USER” and “/collab1/data\_untrusted/\$USER”
  - ❑ The File system’s maximum capacity is managed via file purgers
    - ❑ There is a 5 day purger on “/collab1/data\_untrusted/\$USER”
    - ❑ There is a 90 day purger on “/collab1/data/\$USER”
  - ❑ All file system directories are reachable via the Interactive nodes

- ❑ Data transfer services:
  - ❑ scp, sftp, rsync (via ssh), and Globus
- ❑ Accessible Directories
  - ❑ “/collab1/data/\$USER” (**90 day purge policy**)
- ❑ Accessible From
  - ❑ Pre-authorized “trusted” NOAA and non-NOAA sites. A user help ticket and security approval is required.
- ❑ Externally initiated unattended scp and rsync (via ssh) allowed on a per user basis. User help ticket and justification required.
- ❑ With the exception of unattended data transfers, RSA token authentication required for all transfers.

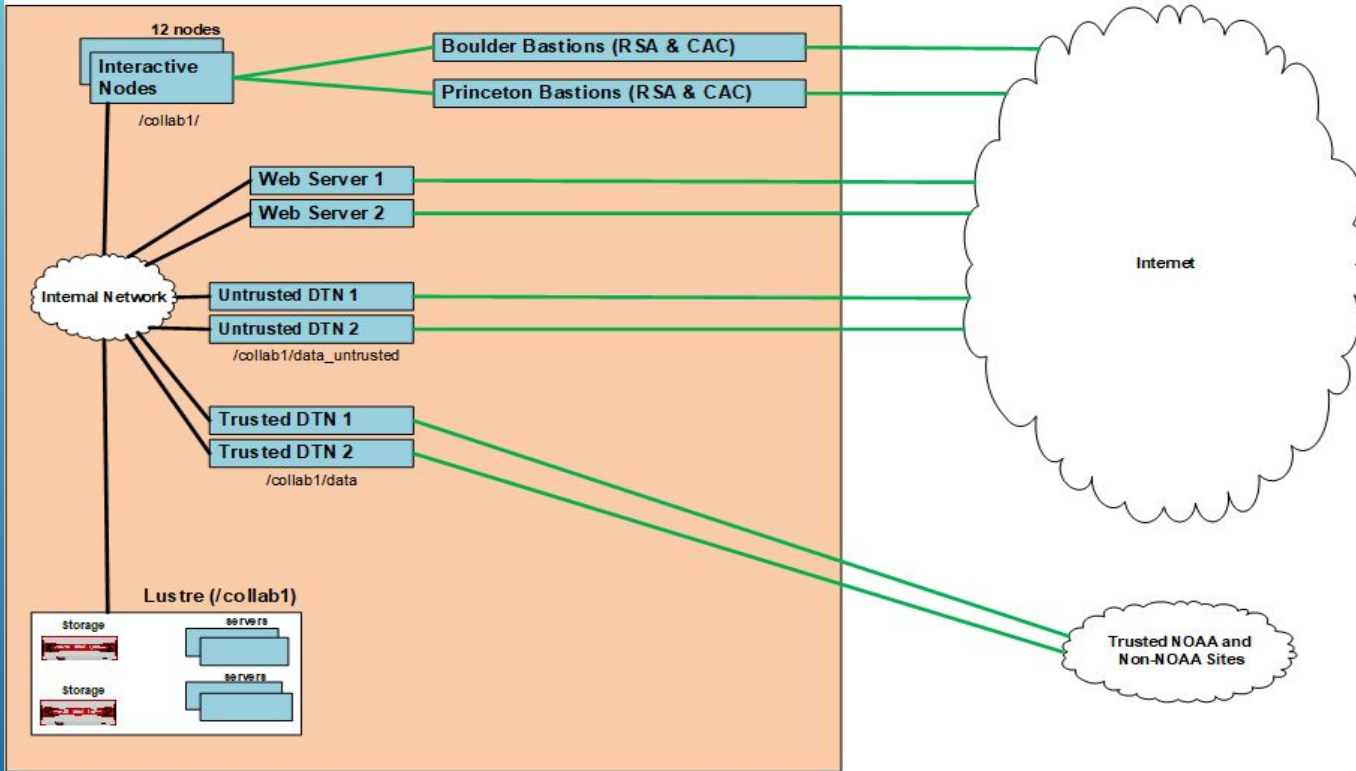
- ❑ Data transfer services:
  - ❑ scp, sftp, rsync (via ssh), and Globus
- ❑ Accessible Directories
  - ❑ “/collab1/data\_untrusted/\$USER” (**5 day purge policy**)
- ❑ Accessible From
  - ❑ Allow connections from any inbound location
- ❑ Externally initiated unattended scp and rsync (via ssh) are not allowed
- ❑ RSA token authentication required for all transfers

- ❑ The RDHPCS program is in the process of deploying the Globus service on all of its DTNs.
- ❑ This service will allow users to use a web based interface for moving data to and from other Globus supported sites and systems (i.e. Orion's DTN Endpoints).
- ❑ The native parallel transfer capability of Globus allows for large amounts of data to be transferred at very high speeds.
- ❑ This service is currently only available on the RDHPCS Niagara system and is the recommended data transfer tool.
- ❑ For information on how to use Globus, please see:  
[https://rdhpcs-common-docs.rdhpcs.noaa.gov/wiki/index.php/Transferring\\_Data\\_Globus](https://rdhpcs-common-docs.rdhpcs.noaa.gov/wiki/index.php/Transferring_Data_Globus)

- ❑ Only transfer important data!
- ❑ Use Data Transfer Nodes for all transfers (and not login nodes)
- ❑ Tar up directories and small file
- ❑ Make sure the person doing the transfers has adequate permissions on both systems.
- ❑ Use Globus to transfer data.

## Data Migration from Orion to Niagara: Guidelines

# NOAA's RDHPCS Niagara System



Niagara System Diagram

❑ Documentation at:

❑ [https://niagaradocs.rdhpcs.noaa.gov/wiki/index.php/Main\\_Page](https://niagaradocs.rdhpcs.noaa.gov/wiki/index.php/Main_Page)

❑ Help: [rdhpcs.niagara.help@noaa.gov](mailto:rdhpcs.niagara.help@noaa.gov)

Documentation and Reporting Problems



Thanks!

Further Questions?