

Zeus Users' Quick Start Training

March 14/16, 2012

Agenda

- Overview
 - Documentation
 - Zeus Architecture
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- HSMS (HPSS) Access
- Getting Help

Documentation

- Basic NESCC Web Site is

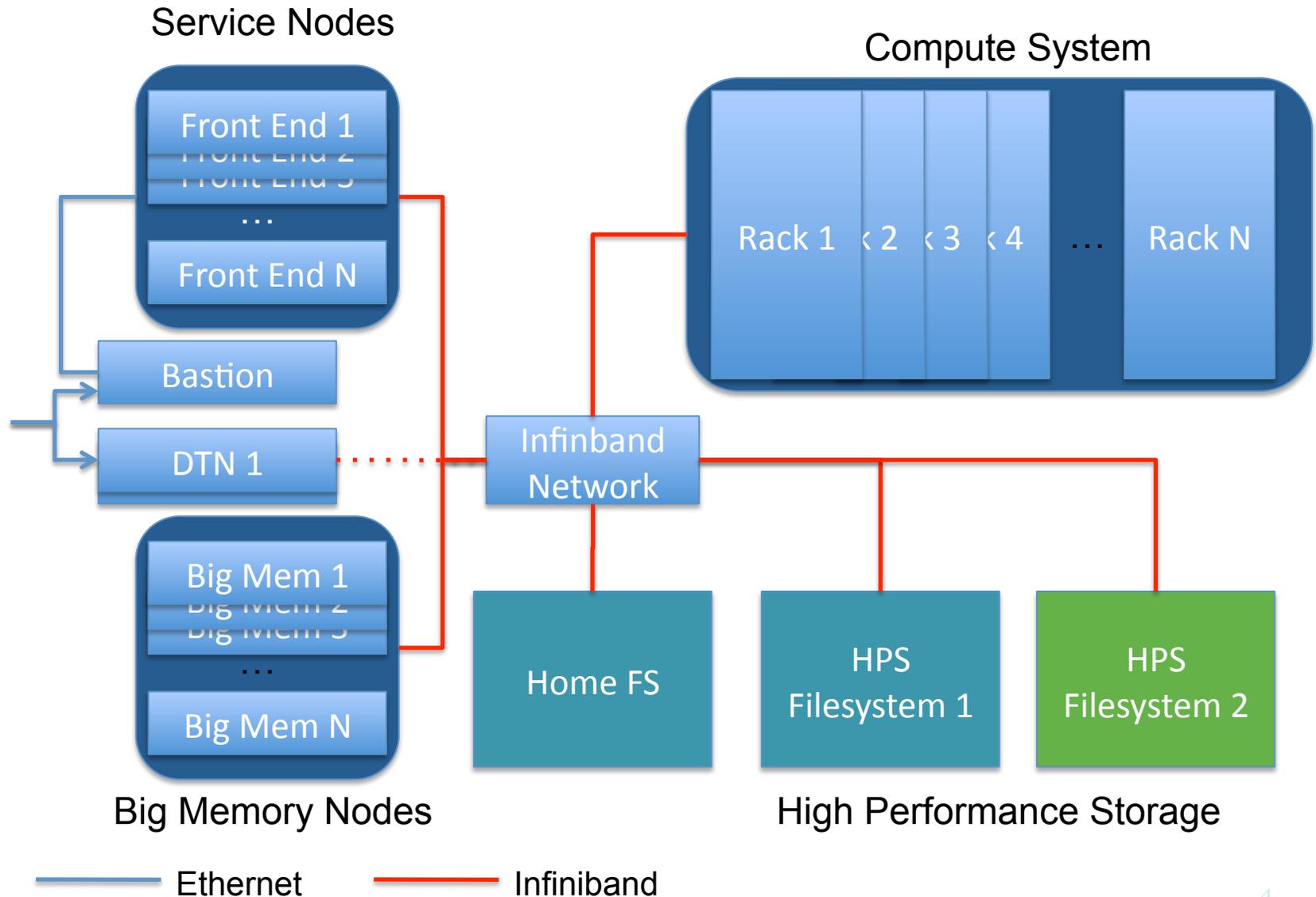
<https://rdhpcs.noaa.gov/nesc>

- On the left side is a link to FAQ

https://nescdocs.rdhpcs.noaa.gov/wiki/index.php/Main_Page

- This presentation is essentially a structured walk through of the FAQ. Web references at the bottom of the slides is the FAQ entry for that subject.

Architecture



Architecture

Subsystem	Size
Service Nodes	8 (48GBytes/node)
Big Memory Nodes	6 (96GBytes/node)
Compute Nodes	2304 (24GBytes/node)
Compute Cores	27648
Total Flops	383 TF
Total Data Transfer Nodes	2+
HPS Capacity	5.6 PB
HPS Performance	> 70 GB/s

Architecture Description

Type	Purpose
Bastion	All login sessions are routed through these hosts. They provide additional security to protect NOAA assets.
Service Nodes	Where initial login sessions are created. They are for general interactive use for operations such as editing and compiling code, managing files, initiating actions on external network including data transfer and accessing software repository.
Data Transfer Nodes (DTN)	Provide high speed Inbound and outbound data transfers. - FOR EXTERNALLY INITIATED TRANSFERS -
Compute System	Major compute resources for parallel jobs. Accessible through the batch system.
Big Memory Nodes	Large memory set for applications that require more memory than a single compute node. Accessible through the batch system.
High Performance Storage	High speed filesystems
Home Filesystem	Storage system optimized for editing and compiling code.

Policies/Conventions

- Front-ends/Service Nodes
 - File Transfers
 - Compiles
 - Editing/Organization
 - **NO** Compute Jobs
- Compute Nodes
 - Obviously, compute jobs
 - Do not have access to the data movement or HPSS
 - Do not compile

Policies/Conventions

- Disk space
 - Home
 - Backed up, User quota (initially 5GB/user)
 - Scratch
 - Not backed up, not purged, Group quota (Based on portfolio managers direction)

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- **Accessing Zeus**
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- Getting Help

Accessing Zeus

- Requesting an account on Zeus
 - This will change over time, we hope to eventually have a unified, web-based account request form. But, we aren't there yet
- For now, if you have access to Vapor, Jet, Gaea, etc. send a help request in requesting the account. Please specify the project or projects you are associated with.
 - Use: `rdhpcs.zeus.help@noaa.gov`

Accessing Zeus

- Your RSA token needs to be set up, we won't discuss this here. We are assuming you already have a valid PIN.
- Use an ssh client to login to Zeus (Linux, Mac: ssh; Windows: putty or similar)
 - Linux/Mac: `ssh -X nems.uname@zeus.rdhpcs.noaa.gov`
- If this is your first time, you will be asked to enter a three (or more) word pass phrase.
- After you enter your pass phrase, a login will be attempted. This will fail because you don't have a password. Don't worry, the process has been initiated correctly.

Accessing Zeus (continued)

- After your certificate has been signed, you can then try again logging into Zeus. (Use the same method as before.)
- You will be prompted for your pass phrase again. (This will be the last time – for a while.)
- The interaction is on the following web page:
https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Logging_in_to_Zeus
- Let's take a quick look

Selecting Specific Login Nodes

- Sometimes you may need to return to a specific Login Node. The Login Nodes are fe1-fe8, tfe1 is for Herc – the test and development system (TDS) only specific users will be enabled on the TDS
- Near the end of the login process you will see

Proxy certificate retrieved.

You will now be connected to OneNOAA RDHPCS: Zeus system

Hit ^C within 5 seconds to select another host.

After hitting Control-C you will see

Select a host. Enter the hostname, or a unique portion of a hostname:

Hostname IP Address

fe1 140.90.206.33

fe2 140.90.206.34

fe3 140.90.206.35

fe4 140.90.206.36

fe5 140.90.206.37

fe6 140.90.206.38

fe7 140.90.206.39

fe8 140.90.206.40

tfe1 140.90.206.41

Enter hostname here: fe2

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- **Moving Data to/from Zeus**
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- Getting Help

Transferring Data To/From Zeus

- Using the Data Transfer Host (Token Authentication – only works from .noaa.gov)

```
# scp bigfile John.Smith@dtm-zeus.rdhpcs.noaa.gov:/scratch2/hfip-hda/John.Smith
John.Smith@dtm-zeus.rdhpcs.noaa.gov's password:
    (This is the point where you enter your PIN+RSA Token response)
bigfile
100% 128MB 32.0MB/s 00:04
#
```

Transferring Data To/From Zeus

- Unattended Data Transfers
 - Will be set up upon special request
 - From a specific hosts within .noaa.gov
 - To a specific user name on Zeus
 - Send a request to the help system explaining your requirements and the list of IP address from which the transfers will occur
- Outbound transfers
 - Users need to request through the help system what remote hosts (IP addresses) they need to access
 - Done from the front-end nodes
 - Use the “service” nodes

See:

https://nescdocs.rdhpcs.noaa.gov/wiki/index.php/Transferring_Data

Data Transfer – Port Tunnel

```
dschorna@aphrodite-l1:~$ ssh Daniel.Schornak@zeus.rdhpcs.noaa.gov
*****
*                               WARNING!
*****
* This is a United States Government computer system. It is
* accessed and used only for official Government business by
* personnel. Unauthorized access or use of this system is
* subject violators to criminal, civil, and administrative
* penalties.
*
* All information on this computer system is confidential and
* read, copied, and disclosed by and to a third party for
* official purposes, including criminal investigations, is
* of this computer system by any person, in whole or in part,
* unauthorized, constitutes consent to the release of such
* information.
*****
Access is via First.Last username only. E
```

1. Get Your Personal Port Number:
 - a. Log in to Zeus (*zeus.rdhpcs.noaa.gov*)
`ssh John.Smith@zeus.rdhpcs.noaa.gov`
 - b. Enter RSA PIN+token code
 - c. Find port tunnel ID in Welcome Message
 - d. Log out of the session
2. On Linux station, log in with tunnel attributes:
`ssh -L21796:localhost:21796 John.Smith@zeus.rdhpcs.noaa.gov`
3. Verify tunnel is established in Welcome Message

```
Daniel.Schornak@tfe1:~$ ssh -L21796:localhost:21796 Daniel.Schornak@zeus.rdhpcs.noaa.gov
PLEASE NOTE:
=====
You will need to reconfigure your SSH client
*and* re-establish a login session
if you have never used this port tunnel before.

Configure your SSH client to local forward port 21796
to localhost:21796.
For Unix, Linux, and MacOSX users, start ssh like this:

ssh -L21796:localhost:21796 Daniel.Schornak@zeus.rdhpcs.noaa.gov.
```

Data Transfer – Port Tunnel

```
dschorna@aphrodite-l1:~$ ssh -L21796:localhost:21796 Daniel.Schornak@zeus.rdhpcs.noaa.gov
*****
*                                     WARNING!                                     *
*****
* This is a United States Government computer system. It is to be used only for official
* purposes and accessed and used only for official purposes of authorized personnel. Unauthorized access
* to this system and the information it contains is prohibited. Any unauthorized use of this system
* subject violators to criminal, civil, and administrative penalties.
*
* All information on this computer system is to be controlled, stored, read, copied, and disclosed by
* authorized personnel only. It is to be used only for official purposes, including criminal
* investigations. Any unauthorized use of this computer system by any person, whether authorized
* or not, is prohibited. Any unauthorized use of this system, whether authorized or not,
* constitutes consent to the collection, use, and disclosure of information contained
* on this system.
*****
Access is via First.Last username only. Enter RSA PASSCODE: █
```

1. ~~Get Your Personal Port Number:~~
 - a. ~~Log in through Zeus (*zeus.rdhpcs.noaa.gov*)~~
Ssh John.Smith@zeus.rdhpcs.noaa.gov
 - b. ~~Find port tunnel ID in Welcome Message~~
 - c. ~~Log out of the session~~
2. On Linux station, log in with tunnel attributes:
ssh -L21796:localhost:21796 John.Smith@zeus.rdhpcs.noaa.gov
3. Verify tunnel is established in Welcome Message

```
Daniel.Schornak@fe8:~$
Hit ^C within 5 seconds to select another host.

Attention user:

A port-tunnel has been established for SCP data transfers
on port 21796 to host zeus.rdhpcs.noaa.gov.

PLEASE NOTE:
=====
You will need to reconfigure your SSH client
```

Data Transfer – Port Tunnel

To use the port tunnel for SCP from your host:

Template:

```
scp -P 21796 /local/path/to/file John.Smith@localhost:/remote/path/to/file  
or
```

```
scp -P 21796 John.Smith@localhost:/remote/path/to/file /local/path/to/file
```

(Windows PUTTY users will use pscp instead of scp)

IMPORTANT NOTE:

=====

ALWAYS USE '**localhost**' for the SCP commands so the port tunnel is used.

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- **Modules**
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- Getting Help

Modules

- Modules allow the PATH and environment variables to be manipulated so that multiple versions of software can be supported
- As a result some things are not in your path by default including compilers and MPI
- A good set of defaults including the Intel compiler:
module load intel mpt
- A good set of defaults including the PGI compiler:
module load pgi mpt
- These commands can be in your .cshrc or .profile

Modules - Documentation

- See:
 - https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Using_Modules
- For sh, ksh, or bash see:
 - https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Frequently_Asked_Questions
 - Select: Why can't I use module commands inside of my batch script

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- **Compiling codes**
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- Getting Help

Compiling

- We currently have two compilers installed
 - Intel
 - We have more licenses and can provide more extensive support for the Intel compiler
 - PGI
- We expect to add the Lahey compiler in the future
- Many systems provide MPI compiler wrapper scripts.
 - These are provided, but are not supported nor are they guaranteed to work.

Compiling - Intel

- Compiler names:
 - Fortran: ifort
 - C, C++: icc, icpc
- Recommended options:
 - Optimize
 - `-O2 -xSSE4.2 -ip`
 - Debug
 - `-O0 -g -traceback -check all -fpe0 -ftrapuv`
- Linking MPI
 - `-lmpi`

Compiling - PGI

- Compiler names:
 - Fortran90, Fortran77: pgf90, pgf77
 - C, C++: pgcc, pgCC
- Recommended options:
 - Optimize
 - `-O2 -fastsse -tp=nehalem-64`
 - Debug
 - `-O0 -g -traceback -Mbounds -Mchkfpstk -Mchkptr -Mchkstk`
- Compiling/Linking MPI
 - Compile:
 - `-I$(MPI_ROOT)/include`
 - Link
 - `-L$(MPI_ROOT)/lib -lmpi`

Setting Compilers Within Make

- Intel

```
CC = icc
F77 = ifort
CXX = icpc
F90 = ifort
```

- PGI

```
CC = pgcc -I$(MPI_ROOT)/include
F77 = pgf77 -I$(MPI_ROOT)/include
CXX = pgCC -I$(MPI_ROOT)/include
F90 = pgf90 -I$(MPI_ROOT)/include
```

- Documentation

https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Using_the_Compilers

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- **Submitting a Job**
- Monitoring a Job
- Resource utilization inquiries
- Getting Help

Submitting a job

- qsub is the command preferred for Zeus
- qsub options
 - a <date_time> ([[MM]DD]hhmm[.SS])
 - A project (**required**)
 - l nodes=<nodes>:ppn=<proc_per_nodes> **or**
 - l procs=<number_of_cores>
 - l walltime=<hh:mm:ss> or walltime=<seconds>
 - l mem=<memsize> (default is 1.2GB/process)
 - q <queue_name> (default is batch)

Zeus Queues

Queue	Min Cores	Max Cores	Max Wallclock	Description
batch	1	4008	8:00:00	Default queue for jobs
urgent	1	4008	8:00:00	Queue for jobs that need to start ASAP. Accounts will be charge 2x the cycles used for any jobs run in this queue.
debug	1	4008	00:30:00	Highest priority queue, useful for during debugging sessions
novel	4008	27648 (full system)	8:00:00	Queue for running novel or experimental jobs where nearly the full system is required
service	1	1	24:00:00	Jobs will be run on front end nodes that have external network connectivity. Useful for data transfers or access to external resources like databases. If you have a workflow that requires pushing or pulling data to/from the HSMS, this is where they should be run.
bigmem	1	12	8:00:00	Jobs will be run on the big memory nodes.

Useful Options and Variables

- To have a job start in your current directory
-d .

or

cd \$PBS_O_WORKDIR (In your batch script)

- Some Useful Job Environment Variables

\$PBS_JOBID - The jobid of the currently running job

\$PBS_O_WORKDIR - The directory from which the batch script was submitted

\$PBS_QUEUE - The assigned queue for this job

\$PBS_NP - The number of tasks assigned to this job

Examples

- Submit a job with a 4-hour limit, 32 cores, project nescmgt, and urgent queue

```
# qsub -A nescmgt -l walltime=4:00:00,procs=32 -q urgent job.csh
```
- Submit a job with a 2-hour limit, 16 processes each with 6 threads (OpenMP/MPI hybrid), project fim, and batch (default) queue. Embed commands in script
 - qsub fimjob.sh
 - fimjob.sh:

```
#!/bin/sh -login
# Options use the PBS sentinel
#PBS -A fim
# Indicate 8 nodes with 2 MPI tasks/node (use 6 OMP threads/task)
#PBS -l nodes=8:ppn=2
#PBS -l walltime=7200
export OMP_NUM_THREADS=6
```
- See: https://nescdocs.rdhpcs.noaa.gov/wiki/index.php/Running_and_Monitoring_Jobs

Running an MPI Program

- Two MPI Implementations: SGI and Intel

- SGI

```
mpiexec_mpt -np <numMPIprocs> <executable>
```

- Intel

```
mpiexec -np <numMPIprocs> <executable>
```

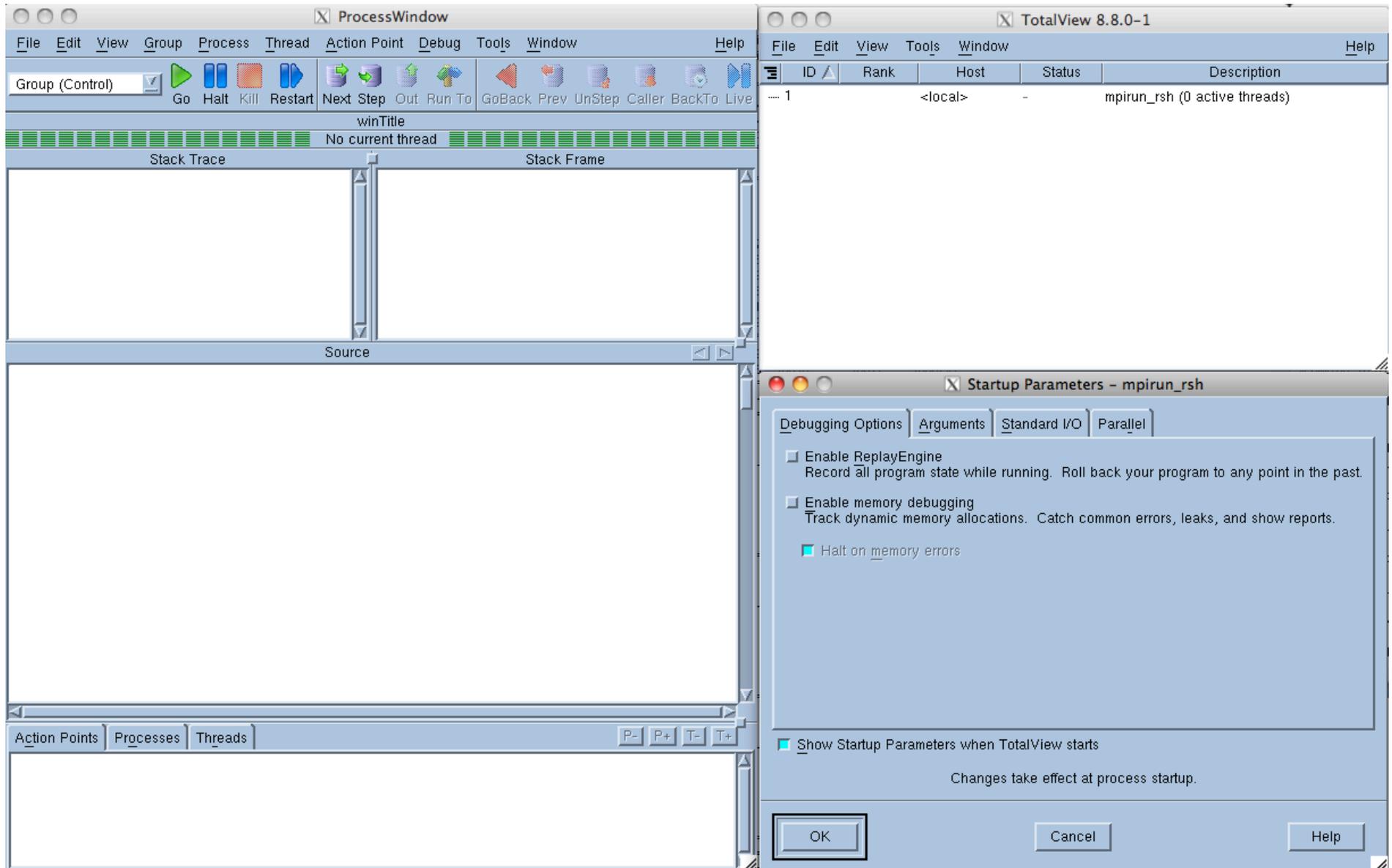
- See:

```
https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Using\_MPI
```

Submitting an Interactive Job

- To submit an interactive job use the `-I` and `-X` options
`qsub -A fim -I -X -l procs=36,walltime=1:00:00`
 - After the job works through the queue, an interactive prompt will appear
- Starting TotalView
 - At an interactive prompt enter the following (make sure you used `-X` on your `ssh` and `msub`):
`module load totalview`
`totalview` (***Give it a moment to start up, used to debug a core file***)
- Starting TotalView via `mpiexec_mpt`
`mpiexec_mpt -tv -np <numMPIprocs> <executable>`

Launching Totalview with the totalview command



Launching Totalview with mpiexec_mpt

The image shows two windows from the TotalView 8.8.0-1 IDE. The left window, titled 'mpirun_rsh', displays the process control interface. The top menu includes File, Edit, View, Group, Process, Thread, Action Point, Debug, Tools, Window, and Help. Below the menu is a toolbar with buttons for Go, Halt, Kill, Restart, Next Step, Out, Run To, GoBack, Prev, UnStep, Caller, BackTo, and Live. A status bar indicates 'Process 1 (0): mpirun_rsh (Exited or Never Created)' and 'No current thread'. The 'Stack Trace' and 'Stack Frame' panes both show 'No current thread'. The main pane displays the source code for the 'main' function in 'mpirun_rsh.c'.

```
423 }
424 }
425
426
427 int main (int argc, char *argv[])
428 {
429     int i, s, c, option_index;
430     int paramfile_on = 0;
431     #define PARAMFILE_LEN 256
432     char paramfile[PARAMFILE_LEN + 1];
433     char *param_env;
434     struct sockaddr_in sockaddr;
435     unsigned int sockaddr_len = sizeof (sockaddr);
436
437     char *env = "\0";
438     int num_of_params = 0;
439
440     char totalview_cmd[200];
441     char *tv_env;
442
443     int timeout, fastssh_threshold;
444     atexit ( remove_host_list_file );
445     atexit (free_memory);
```

The right window, titled 'TotalView 8.8.0-1', shows a table with the following data:

ID	Rank	Host	Status	Description
1	<local>	-	-	mpirun_rsh (0 active threads)

At the bottom of the left window, there are tabs for 'Action Points', 'Processes', and 'Threads', along with zoom controls (P-, P+, T-, T+).

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- **Monitoring a Job**
- Resource utilization inquiries
- Getting Help

Monitoring a job

- Show all jobs in the queue

```
# showq
```

- Show all jobs belonging to only you

```
# showq -u Christopher.W.Harrop
```

```
active jobs-----
```

JOBID	USERNAME	STATE	PROCS	REMAINING	STARTTIME
-------	----------	-------	-------	-----------	-----------

```
0 active jobs          0 of 27708 processors in use by local jobs (0.00%)
                       0 of 2310 nodes active          (0.00%)
```

```
eligible jobs-----
```

JOBID	USERNAME	STATE	PROCS	WCLIMIT	QUEUE TIME
-------	----------	-------	-------	---------	------------

5646	Christop	Idle	1	00:05:00	Thu Jan 26 16:58:52
------	----------	------	---	----------	---------------------

```
1 eligible job
```

```
blocked jobs-----
```

JOBID	USERNAME	STATE	PROCS	WCLIMIT	QUEUE TIME
-------	----------	-------	-------	---------	------------

```
0 blocked jobs
```

```
Total job: 1
```

Monitoring a Job

- Using qstat will provide current status (showq can be up to 60 seconds old)
 - Show only jobs associated with user John Smith (note mixed case)
qstat -u John.Smith
 - Show all jobs in all states
qstat
 - Many, many options (see man qstat)

Deleting a job

- `qdel <jobid>`
- `qdel -k <jobid>`
 - Please only use when `qdel` fails to terminate the job
 - Really kill it!

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- **Resource utilization inquiries**
- Getting Help

Resource Utilization Queries:

account_params

```
# account_params
```

```
Account Params -- Information regarding project associations
```

```
Processing Unix group file /etc/group
```

```
User: Leslie.B.Hart
```

```
Project: det
```

Allocation:	Id	Name	Amount	Reserved	Balance	CreditLimit	Available
Allocation:	--	-----	-----	-----	-----	-----	-----
Allocation:	40	det	194.57	0.00	194.57	0.00	194.57

```
Directory: /scratch2/portfolios/BMC/det DiskInUse=0 GB, Quota=10000 GB
```

```
Project: sepp
```

Allocation:	Id	Name	Amount	Reserved	Balance	CreditLimit	Available
Allocation:	---	-----	-----	-----	-----	-----	-----
Allocation:	106	sepp	217.94	217.94	0.00	0.00	0.00

```
Directory: /scratch2/portfolios/BMC/sepp DiskInUse=1275 GB, Quota=5000 GB
```

```
Allocation Terms:
```

```
Name -- Name of allocated project
```

```
Amount -- Total Allocation
```

```
Reserved -- Reserved requests from current jobs in queue.
```

```
Balance -- Total allocation remaining for new jobs.
```

```
Available -- Amount currently available for new jobs.
```

```
All amounts are in core-hours.
```

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- **HSMS (HPSS) Access**
- Getting Help

HPSS HSI

- HSI is a FTP-like interface to the HPSS
 - HSI is most useful for file and directory manipulation
 - HSI supports wild cards for local and HPSS pathname pattern matching, and provides recursion for many commands
 - HSI includes the ability to store, retrieve, and list entire directory trees, or change permissions on entire trees
- Copying Files to HPSS Using HSI
 - Load HPSS module

```
# module load hpss
```

- To put the file local_file into the HPSS directory /BMC/testproj/work

```
# hsi put /full_local/path/local_file : /BMC/testproj/work/local_file
```

HPSS HSI

- Retrieving a File from HPSS using HSI
 - To get the HPSS file `local_file` located in the HPSS directory `/BMC/testproj/work`
 - In your current directory (`.`)

```
# hsi get /BMC/testproj/work/local_file
```

- In the directory `/full/path` with name `new_name`

```
# hsi get /full/path/new_name : /BMC/testproj/work/local_file
```

HPSS HSI

- Listing the contents of an HPSS directory using HSI
 - To list the contents of the directory /BMC/testproj
hsi ls /BMC/testproj
- To get a listing of command line options
hsi help

For the full list of HSI supported commands

please visit: http://www.mgleicher.us/GEL/hsi/hsi_reference_manual_2/hsi_commands/

Agenda

- Overview
 - Zeus Architecture
 - Documentation
 - Policies/Conventions
- Accessing Zeus
- Moving Data to/from Zeus
- Modules
- Compiling codes
- Submitting a Job
- Monitoring a Job
- Resource utilization inquiries
- **Getting Help**

Getting Help

- Use OTRS
 - Email-based
 - Zeus Issues
`rdhpcs.zeus.help@noaa.gov`
 - HPSS Issues
`rdhpcs.hpss.help@noaa.gov`

Thank You!

now...

Open Forum