The Hoffman2 Cluster is a high-performance computing cluster that provides > 7500 processor cores in three data centers.

It is a project of the Institute for Digital Research and Education (IDRE) Cluster Hosting Program at UCLA. Members of the entire UCLA research community are entitled to utilize this high-performance computing cluster.

Research groups have also purchased and contributed to the cluster to entitle them to priority access on their compute nodes.

- The Department of Biostatistics is one of those departments that has purchased compute nodes for the cluster to allow its students and research staff priority access and usage.
This resource is especially useful for students who are either currently conducting research for a GSR, masters project, or dissertation.

As students within the Biostatistics department, you get priority access and usage to the compute nodes paid for by the department.
Common Biostat projects include running jobs in R, STATA, and/or Matlab. You can run these projects on the Hoffman2 cluster, as well as other software applications.

- This results in **lots more computing power** applied to your jobs so they **complete faster**.
- It frees up your local computer so it doesn't get bogged down by heavy jobs using most of the processor.
- It allows you to let your jobs run unattended so you don't have to keep your computer on until it finishes.
- Your files are also located on a campus server (and backed up) so you aren't limited to accessing them just on your local computer.
HOFFMAN2 OFFICIAL WEBSITE

- https://idre.ucla.edu/hoffman2
  - Contains lots of easy to use useful information
  - Getting started, Access info, Software info, FAQ, etc.
Hoffman2 Cluster

About the Hoffman2 Cluster

- Real-time Usage
- Current News
- News Archive

Getting started: accounts and passwords

- Parallel Computing Classes
- Frequently Asked Questions
- High Performance Computing Consulting: hpc@ucla.edu

Using the Hoffman2 Cluster

- Access
- Computing
- Data Storage
- File Transfer
- Software
Getting an account on the Hoffman2 Cluster is a quick three-step process.

1. Fill out the “New User Registration” to gain access to UCLA Grid Portal and apply for Hoffman2 Cluster account.
2. Faculty sponsor approves your application.
3. Account is created and you’re good to go.
NEW USER REGISTRATION FORM

- This form is accessible from the Hoffman2 site under “Getting Started”

- You will be asked for your UCLA Logon ID and password to access the form.

- Make sure you choose “William Cumberland” as your sponsor and “Biostatistics” as your department.

- The password you set here is for the UCLA Grid Portal access only, not for your Hoffman2 account. That comes in the final step.
Grid Identity Manager

New User Registration

If you don’t have a cluster login id and/or Grid Username, please fill out the following form. An e-mail will be sent to your chosen sponsor for approval. If the sponsor you select is not correct, your application may be denied. You use your Grid Username and Grid Password when you access a cluster via the web, so it is important to remember them.

Select a Resource: Hoffman2 Cluster
Select a Sponsor: William Cumberland
Proposed Username: bruin

- Passwords can only contain the characters: a-z A-Z 0-9 @#/.
- Passwords must be at least 6 characters long.
- Passwords must contain at least one letter and one non-letter from: 0-9 @#/.

New Grid Password: 
New Grid Password Confirm:
First Name: Joe
Last Name: Bruin
Email Address: bruin@ucla.edu
Department, Center or Organization: Biostatistics
Preferred Shell: Bash

Create the shared secret
- You will need the following shared secret to retrieve your initial cluster login password.
- You will also need it if you forget your password and need to have it reset in future.
- The secret answer can not contain special characters. It must be alpha-numeric.

Secret Challenge Question: What’s the best school in Westwood?
Your Answer: UCLA
Project Description:

Submit
Go Back

If you have any other questions, please contact the IDRE Account Administrator at accounts@idre.ucla.edu

November 7, 2012

Any information you enter may be viewed by the system administrators. Do not use your social security number or similar identifier on this site.
Once you’ve submitted the “New User Registration” form, Dr. Cumberland receives the request and approves the account.

Upon approval, you will receive an email with your Hoffman2 Cluster account information including username and a temporary password.
IDRE supported method is with a secure shell (SSH) program. Free options available:

- Windows:
  - Putty, Cygwin
- Mac:
  - Built-in ‘terminal’

For GUI interface, requires X11 software

- NXClient, XQuartz
HOW TO LOG IN TO THE HOFFMAN2 CLUSTER

- Command line (ssh):
  - ssh login-id@hoffman2.idre.ucla.edu
  - Best way to be able to submit batch jobs

- GUI Options available:
  - NX Client
    - Requires specific key file
  - X11
### BATCH vs. INTERACTIVE

- **Batch jobs**
  - Submitted to the job queue for cluster processing
  - Runs quicker because of more processing power
  - Doesn’t require an active connection to complete
  - Notifications of job status (start/complete/terminate)
  - Can specify Biostat processors or entire Cluster
  - Time limit of 14 days
  - No user interaction while running
**Interactive**

- Interface is like that of the regular client software
- Requires “qrsh” interactive session reservation
- Meant mostly for testing before submitting a batch job
- Time limit of **24 hours maximum**
- Less processor utilization
  - Slower than Batch due to less processors
- Can’t specify just Biostat processors
- Requires active network connection to complete jobs
- Can be manually suspended and resumed later
- Lost internet connection can result in hung session that needs to be killed before being able to open a new interactive session
GENERAL RECOMMENDATION

- When to use BATCH
  - Most jobs
  - Large data sets
  - Memory intensive jobs

- When to use INTERACTIVE
  - Testing certain aspects of code
  - Small subsets of data
HOW TO SUBMIT ‘R’ BATCH JOBS
HTTPS://IDRE.UCLA.EDU/HOFFMAN2/SOFTWARE/R

- Use the following queue script to start R:
  
  R.q

- Then enter ‘build’ or just ‘b’ to start the build of your R job
- Enter the filename of your R file (filename.R – no spaces allowed)
- Enter the amount of memory (MB) to allocate to the job
- Enter the time limit for the run duration
- Choose whether or not to use just the Biostat cores or the entire cluster
- Enter any required arguments
- A filename.cmd file is created and then choose to submit now or not
- A job ID# is created
- Quit out of the R queue (q – quit)

- Type ‘myjob’ to see the status of your job
- Filename.out.jobid# contains the output of your job
R BATCH DEMO

- Demo
If you choose not to submit your batch job immediately you can do so after by typing:

```
qsub filename.cmd
```

- The .cmd file is the cluster batch file that’s created to run your job on the cluster.

- Time limits will affect when your job gets scheduled. Higher time limit might hold the job off a while until resources are available. So be realistic.

- Jobs with a time limit between 24 hours – 14 days automatically submits to your department processors. Less than 24 hours has the option but potentially will wait longer if larger jobs are set ahead on those cores.

- Jobs will automatically terminate at the time limit set with a maximum of 14 days.

- You will receive an email when job starts/ completes/ terminates.
Use the following queue script to start Stata:

```
stata.q
```

- Runs stata-mp on 1-8 processors.

Then enter ‘build’ or just ‘b’ to start the build of your Stata job.

- Enter the filename of your Stata file (filename.do – no spaces allowed)
- Enter the amount of memory (MB) to allocate to the job
- Enter the time limit for the run duration
- Choose whether or not to use just the Biostat cores or the entire cluster
- Enter any required arguments
- Choose number of processors per core
- A filename.cmd file is created and then choose to submit now or not
- A job ID# is created
- Quit out of the Stata queue (q – quit)

- Type ‘myjob’ to see the status of your job
- Filename.log contains the output of your job
STATA BATCH DEMO

- Stata Demo
STARTING AN INTERACTIVE SESSION

- DO NOT RUN PROGRAM BEFORE RESERVING AN INTERACTIVE SESSION

- `qrsh` command reserves the resources for an interactive session

- Reserve interactive session: `qrsh -l i`
  - Can specify session resource parameters. Common parameters:
    - `time`: time limit (default = 2 hrs)
    - `mem`: memory size per core (max 1GB)
    - Example: `qrsh -l i,mem=1G,time=2:00:00`
RUNNING AN INTERACTIVE SESSION

- Reserve the interactive session resources

- Load the appropriate module
  - Module is for setting environmental variables
    - Available modules can be viewed by typing: module available
    - Example: module load stata

- Run the program
  - Example: stata-se
Reserve an interactive session: qrsh -l i

At the prompt type:

module load R
R

Then run R as you normally would
- Doesn’t get the same computing power as submitting as a batch job but allows for interaction if needed
R INTERACTIVE DEMO

- Demo
Reserve an interactive session: qrsh -l i

At the prompt type:
   module load stata

To run Stata interactively, type:
   stata-se

Or, for the multi-processor version of Stata:
   stata-mp

Then run Stata as you normally would

GUI options available
STATA INTERACTIVE DEMO

- Stata Demo
GUI from Windows/Mac:
  ▪ Drag & drop SFTP/SCP software
    ▪ (Windows: WinSCP; Mac: Cyberduck, Fugu)

Command line from Linux/Mac terminal:
  ▪ ssh dtn2.hoffman2.idre.ucla.edu
  ▪ scp [-r] source:path/file target:local_path

For bigger files, use the high-speed GlobusOnline service
  ▪ Directions on Hoffman2 site under ‘File Transfer’
    http://ucla.in/U2J4qd
File Transfer Demo
Only you have access to your files by default. You can manually give permissions to others if needed.

Hoffman2 is **NOT** to be used with **Restricted Info** (SS#, PHI, etc).
- If you have data sets with Restricted Info, please talk to us so we can help you.

Home directory: cd $HOME
- 20GB quota
- backed up for 30 days

Scratch directories
- cd $SCRATCH
- 3 TB limit
- Files deleted after 7 days

Remember to check your quota in your home directory
- ATS scripts:
  - get_pan_quotas USERNAME
- Linux command:
  - cd $HOME
  - du -sh
WHERE TO GET MORE HELP IF NEEDED

- The Hoffman2 homepage contains a lot of useful information including details for certain software packages and an FAQ
  https://idre.ucla.edu/hoffman2

- High-Performance Computer Consulting email (very responsive): hpc@ucla.edu
THIS PRESENTATION


- From main Biostat homepage (http://www.biostat.ucla.edu)
  - Click “Current Students”
  - Click “Useful Links”
  - Hoffman2 quick-start guide for Biostat users: pdf