

# A Guide to HPC/BigPurple

Yi Li, Yuhe Xia

Associate Research Scientist  
Division of Biostatistics

# Outline

- Introduction to HPC/Bigpurple
- Getting access
- Storing & Transferring data
- SLURM commands
- Useful command lines
- Examples

# High Performance Computing (HPC)

HPC stands for **high-performance computing** and usually refers to several computers connected together in a network (forming a **HPC cluster**).

- Each different computer is called a node
- Each node has processors/cores
  - Carry out the instructions of the computer
- All these different computers talk to each other through a communications network

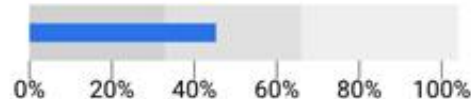
**BigPurple** is NYU Langone's distributed-memory high-performance computing cluster went live in July 2018.

- 90 compute nodes, 32 of which include graphics processing units (GPUs) for a total of 156 GPUs.
- 6 service nodes, 4 highly available login nodes, 4 data mover nodes, 4 high-memory nodes, and a 100Gb Infiniband-2 EDR interconnect.

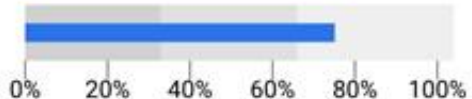
# BigPurple Status

Nodes	Cores	Ram (GB)
96	3,846.8	88,336

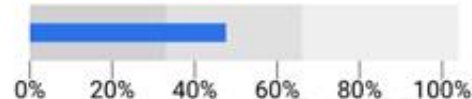
Average CPU Allocation:



Average GPU Allocation:



Average RAM Allocation:

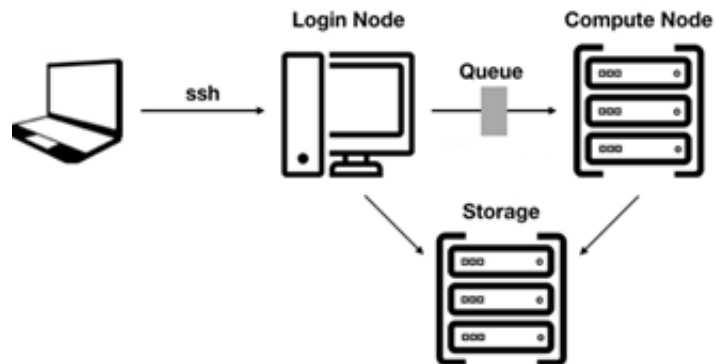


Google Data Studio



# Why/When we need HPC?

- Have many long running jobs to run
- Don't want to tie up your own machine
- Want to run in parallel to get results quicker
- Need more disk space
- Need more memory
- Want to access to the data stored on the cluster
- Want to use software installed on the cluster



# Accessing BigPurple

- Request an HPC account follow the instructions here:  
<https://hpcmed.org/guide/get-started>
- Log in using NYU Langone Kerberos ID and password
- BigPurple is only reachable through NYU on-campus network.
- Connect to VPN if you are off-site

Last login: Wed Sep 8 17:02:06 2021 from 10.127.201.54



NYU Langone Health HPC

Use the following commands to adjust your environment:

```
'module avail'          - show available modules
'module add <module>'   - adds a module to your environment for this session
'module initadd <module>' - configure module to be loaded at every login
```

2021-10-29 SAVE THE DATE: "Parallel Computing With MATLAB"  
Part I : Dec 3rd.  
Part II: Dec 9th.  
Fore registration and detailed information, Please follow the link below:

<https://www.mathworks.com/company/events/seminars/workshop-parallel-computing-with-matlab-for-nyu-2021.html>

2021-03-03 Starting March 21st, all data older than 90 days in the BigPurple and Skynet /gpfs/scratch space will be purged from the system.

2020-07-16 The HPC virtual Town Halls are held on Thursdays, 12:00-1:00.  
You may join via <https://nyumc.webex.com/meet/siavoa01>  
You may also contact us through email. We can use phone or Webex to help you.

BigPurple User Guide available at: <http://bigpurple-ws.nyumc.org/wiki>

New HPC Portal: <https://hpcmed.org/>

You may email <[hpc\\_admins@nyumc.org](mailto:hpc_admins@nyumc.org)> for any further assistance.

Home Block Quota: 11.41G/100G

Home File Quota: 6707/40000

# Accessing BigPurple

## Windows

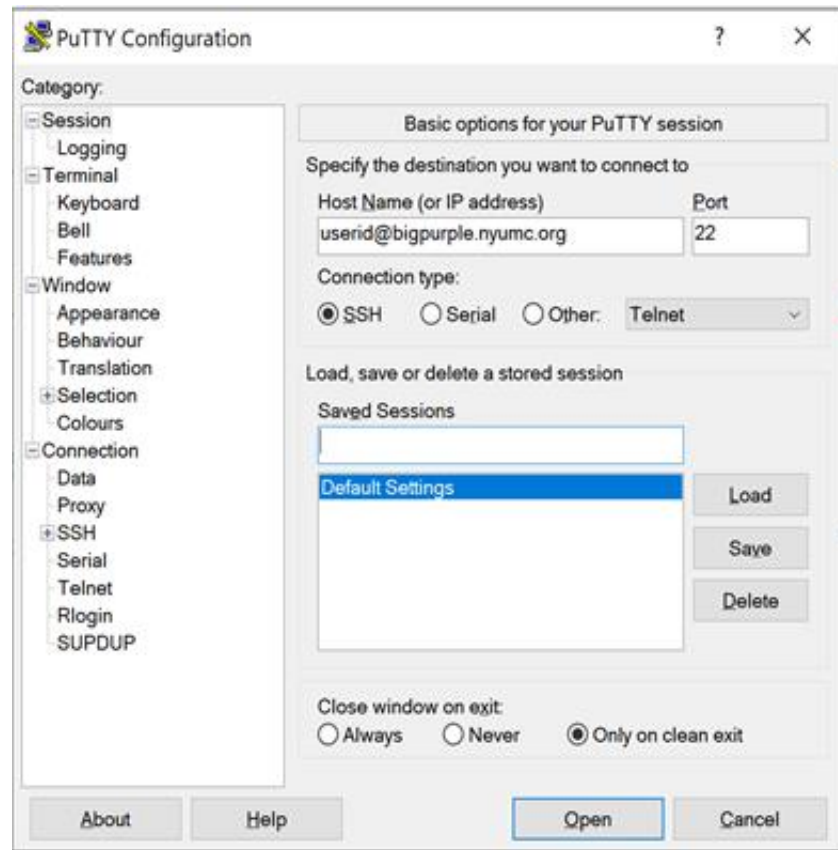
- Install SSH client *PuTTY*
- Connect *PuTTY* to *BigPurple*

## Mac OS X

- Use terminal program directly
- Type ssh [userid@bigpurple.nyumc.org](mailto:user@bigpurple.nyumc.org) to connect



```
yi -zsh - 80x24
Last login: Wed Dec 8 15:02:20 on ttys000
yi@180MADPHMLT025 ~ % ssh userid@bigpurple.nyumc.org
```



# Storage on BigPurple

The storage system of BigPurple is mainly divided up to 3 functional areas.

Name	Location	Good for	Quota	Backed up
Home	/gpfs/home	Scripts, final results	100GB	Yes
Scratch	/gpfs/scratch	Temporary files, running jobs, purged every 90 days	1PB	No
Data	/gpfs/data	Long term storage, need to request	Varies	Yes



# File transferring on BigPurple

## Copying scripts to and from the cluster

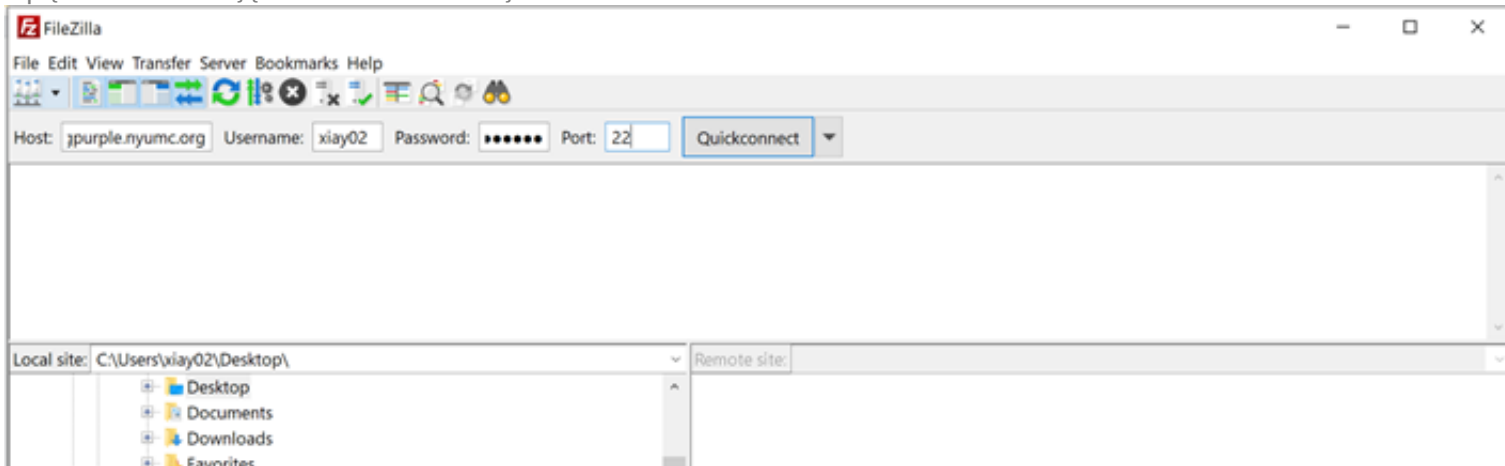
### Command Line

- Using command scp(secure copy)

scp {SOURCE PATH} {DESTINATION PATH}

### GUI based

- FileZilla
- WinSCP (Windows only)
- CyberDuck



# BigPurple Resources

Node Type	Host Names	CPU's	Sockets	Cores Per Socket	Memory GB	Threads Per Core	GPUs
Compute Nodes	cn-[0001-0054]	40	2	20	384	1	0
Fat Nodes	fn-[0001-0004]	40	2	20	1536	1	0
GPU4 Nodes	gn-[0001-0025]	40	2	20	384	1	4
GPU8 Nodes	gpu-[0001-0007]	40	2	20	768	1	8

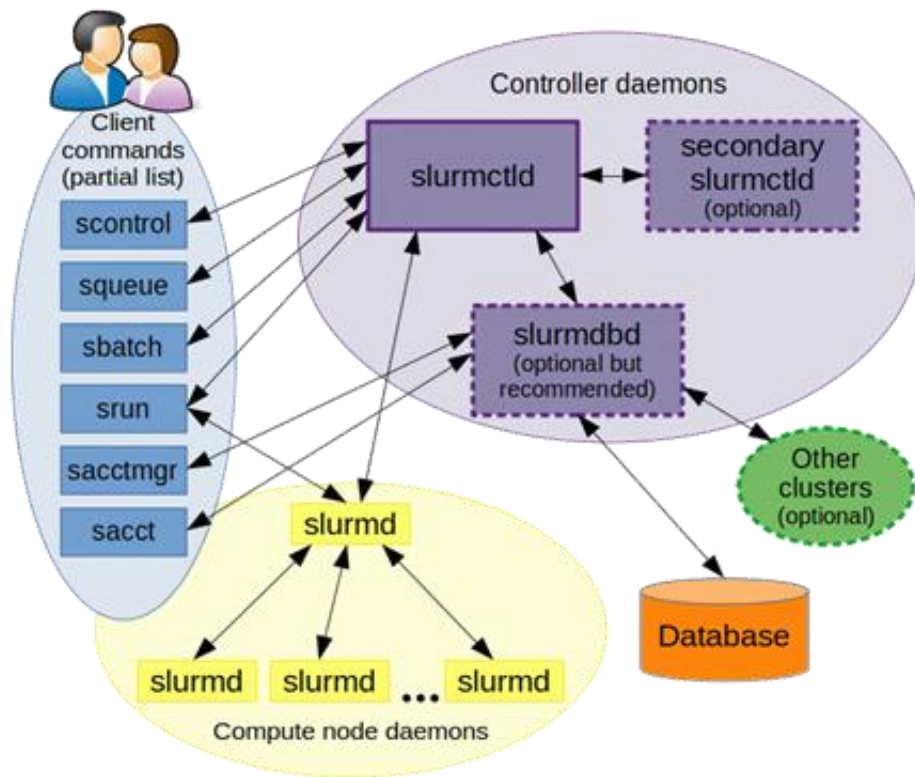
- The computing resources of the BigPurple HPC Cluster by compute node type are broken up into 4 categories (cpu,fn,gpu4,gpu8)
- Each with a different maximum time limitation (dev=4 hours, short=12 hours, medium=3 days, and long=28 days) for scheduling optimization and prediction.
- These nodes can be accessed through the Slurm job scheduler in partitions (eg. gpu4\_medium, gpu8\_long).

# SLURM

Slurm is an open source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters.

Three key functions:

- Resources allocation(node,time)
- Parallel job scheduling framework
- Contention arbitration



# Useful SLURM commands

- **srun** - run a job on terminal directly

```
$ srun --partition=gpu4_dev --nodes=1 --ntasks=2 --cpus-per-task=1 --gres=gpu:2 --x11 --pty bash
```

- **sbatch** - Used to submit a script to be queued on the cluster to run in one or more slots

```
$ sbatch ./script.sh
```

- **squeue** - Display jobs in the queueing system.

```
$ squeue -u user_name #display the list of jobs the scheduler is managing
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	ODELIST(Reason)
6474	cpu_short	serial_j	user_name	R	0:12	1	cn-0003

```
$ squeue -j 6543 #display the status of job by jobid
```

- **sacct** - Check the status of a specific job.

```
$ sacct -X -j 6289
```

JobID	JobName	Partition	Account	AllocCPUS	State	ExitCode
6289_51	cormhap_N+	cpu_medium		10	COMPLETED	0:0

- **scancel** - Stop and remove jobs in the queue.

```
$ scancel -i 6543 # cancel a job by its id
```

<https://slurm.schedmd.com/quickstart.html>

# Sbatch script example

```
#!/bin/bash

#SBATCH --job-name=ig          # your job name
#SBATCH --output=ig.out        # Standard output and error log
#SBATCH --partition=cpu_short
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --cpus-per-task=1
#SBATCH --mem-per-cpu=4G
#SBATCH --time=5:00:00         # Time limit hrs:min:sec
#SBATCH --gres=gpu:8
#SBATCH --mail-type=END,FAIL    # send email when job ends
#SBATCH --mail-user=yourname@nyulangone.org # your email address

module purge
module load r/4.0.3
module load cmdstan/2.25.0
cd /gpfs/scratch/netid/r      # specify the path where you want to run the script
Rscript --vanilla logistic_model.R
```

# Useful command lines

- **cd**: change the current working directory, eg: `cd ../../`
- **pwd**: print working directory
- **mv**: mv - move (rename) files, eg: `mv test.r ../subfolder/`
- **cp**: cp - copy files and directories, `cp Src_file Dest_file`, `cp Src_file1 Src_file2 Src_file3 Dest_directory`, `cp -R Src_directory Dest_directory`
- **scp**: scp - secure copy (remote file copy program), `scp [OPTION] [user@]SRC_HOST:]file1 [user@]DEST_HOST:]file2`

Eg: `scp id@bigpurple.nyumc.org:/gpfs/home/id/filename /Users/myname/Desktop/folder/`

- **ls**: list files or directories
- **mkdir**: create a directory, eg: `mkdir folder_name`
- **cat**: view content of a file, eg: `cat job_name.out`
- Command **—help**: list all options for each command
- ★ **vim**: text editor

# Example

## Steps

- Login to BigPurple through PuTTY
- Transfer files using FileZilla/command line
- Execute the following commands and submit sbatch script:

**cd, pwd, ls, cat, sbatch, squeue**

# Tips

- Read the manual, use google
- Test your script on local machine before running it on HPC
  - Remember to change path in your script
  - Do not abuse the shared-resources



# Resources

- <https://med.nyu.edu/research/scientific-cores-shared-resources/high-performance-computing-core>
- <https://hpcmed.org/guide>
- [http://bigpurple-  
ws.nyumc.org/wiki/index.php/BigPurple  
\\_Applications](http://bigpurple-<br/>ws.nyumc.org/wiki/index.php/BigPurple<br/>_Applications)
- <https://slurm.schedmd.com/quickstart.html>
- [https://sites.google.com/nyu.edu/nyu-hpc/training-support/tutorials/linux-tutorial?authuser=0#h.p\\_ID\\_132](https://sites.google.com/nyu.edu/nyu-hpc/training-support/tutorials/linux-tutorial?authuser=0#h.p_ID_132)
- HPC team holds weekly virtual open house and town hall style event every Thursday
- [hpc\\_admins@nyulangone.org](mailto:hpc_admins@nyulangone.org)



MERRY CHRISTMAS  
*and*  
HAPPY HOLIDAYS