

# General Purpose Cluster (GPC) and Other High Performance Computing Resources at Penn

Professor Xi Song

Department of Sociology

Methodology Working Group, Spring 2023

# Why Do You Need HPC?

# You Will Need HPC When You Need To...

- ▶ Have more computing power than a desktop can provide
- ▶ Run many models at the same time
- ▶ Run a model for a huge dataset that is bigger than your hard disk
- ▶ Use parallel computing to save time on a model that may take a very long time to run (e.g., Bayesian models)
- ▶ Run a model using software that is not installed in your own computer (e.g., Stata MP)

# Virtual Private Network (VPN)

# FortiClient

- ▶ A VPN (Virtual Private Network) Client is necessary for Remote Desktop access, when connecting to Windows Active Directory (AD) Share drives and U:'s, or other Penn resources from off-campus or on WiFi.
- ▶ You will need to install VPN in order to get access to GPC off campus
- ▶ [https://computing.sas.upenn.edu/faculty\\_staff/VPN](https://computing.sas.upenn.edu/faculty_staff/VPN)

# Install FortiClient

- ▶ Fill out the webform to request a KITE account for VPN access [https://upenn.co1.qualtrics.com/jfe/form/SV\\_d61rw3Ta4gJ0Av3](https://upenn.co1.qualtrics.com/jfe/form/SV_d61rw3Ta4gJ0Av3)
- ▶ Send an email to Alex Bond ([sschelp@sas.upenn.edu](mailto:sschelp@sas.upenn.edu)) to let him know about your application
- ▶ Download FortiClient from [https://computing.sas.upenn.edu/faculty\\_staff/VPN](https://computing.sas.upenn.edu/faculty_staff/VPN)

# Connect to FortiClient

- ▶ You will be asked to use the Two-Factor Authentication sent to your Duo Mobile app.
- ▶ You will be prompted for a password, which is your typical Pennkey password.
- ▶ you will need to do a two-factor authentication method. If you have the Duo Mobile app installed and set up on your phone, simply type **1** in the password field and hit enter to get a push notification on your phone. Otherwise, enter of the other numbers for different options.

# FortiClient



xisong

REMOTE ACCESS

Notifications

Settings

About



VPN Name

SAS VPN Neptune2

Username

xisong

Password

.....



Save Password



Auto Connect

Connect

Unlock Settings



xisong

REMOTE ACCESS

Notifications

Settings

About

Unlock Settings

VPN Connected



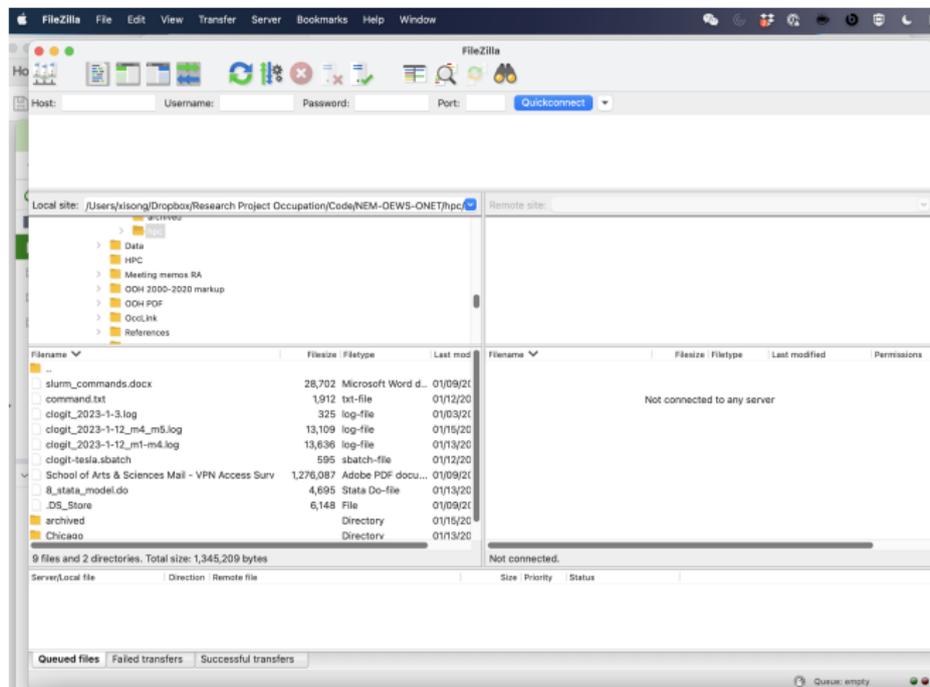
VPN Name SAS VPN Neptune2  
IP Address 10.76.0.10  
Username xisong  
Duration 00:00:24  
Bytes Received 10.53 KB  
Bytes Sent 5.37 KB

Disconnect

# Data Transfer

# How to Transfer Your Dataset to the GPC Server?

In order to use SFTP with the GPC, you will first need an SFTP client, like FileZilla: <https://filezilla-project.org/>



# Data Transfer

Once installed, open the **Site Manager** under the **File** menu and enter the following information, where “pennkey” is your own Pennkey name:

**Protocol:** SFTP

**Host:** [gpc.sas.upenn.edu](https://gpc.sas.upenn.edu)

**Port:** 22

**Logon Type:** Interactive

**User:** your Pennkey name

General

Advanced

Transfer Settings

Charset

Protocol:

SFTP - SSH File Transfer Protocol

Host:

gpc.sas.upenn.edu

Port:

22

Logon Type:

Interactive

User:

pennkey

Background color:

None

Comments:

Connect

OK

Cancel

## Once Connected to the GPC Server

- ▶ Once connected, you will see your local computer files (Local Site) on the left pane, and the GPC's files (Remote Site) on the right.
- ▶ Your own directory on the GPC will be located in `/home/pennkey` using your own Pennkey name.
- ▶ You can drag and drop files between panes, or do a manual transfer.

# Once Connected to the GPC Server

The screenshot shows the FileZilla interface with the following details:

- Title Bar:** FileZilla File Edit View Transfer Server Bookmarks Help Window
- Address Bar:** Penn-GPC - sftp://xisong@gpc.sas.upenn.edu - FileZilla
- Host:** [Empty] **Username:** [Empty] **Password:** [Empty] **Port:** [Empty] **Quickconnect** [Dropdown]
- Status Log:**
  - Status: Connecting to gpc.sas.upenn.edu...
  - Status: Using username "xisong".
  - Status: Connected to gpc.sas.upenn.edu
  - Status: Retrieving directory listing...
  - Status: Listing directory /home/xisong
  - Status: Directory listing of "/home/xisong" successful
- Local site:** /Users/xisong/Dropbox/Research Project Occupation/Code/NEM-OEWS-ONET/hpc
- Remote site:** /home/xisong
- Local File List:**

Filename	Filesize	Filetype	Last mod
..			
slurm_commands.docx	28,702	Microsoft Word d...	01/09/20
command.txt	1,912	txt-file	01/12/20
clogit_2023-1-3.log	325	log-file	01/03/20
clogit_2023-1-12_m4_m5.log	13,109	log-file	01/15/20
clogit_2023-1-12_m1-m4.log	13,636	log-file	01/13/20
clogit-tesla.sbatch	595	sbatch-file	01/12/20
School of Arts & Sciences Mail - VPN Access Surv	1,276,087	Adobe PDF docu...	01/09/20
8_stata_model.do	4,695	Stata Do-file	01/13/20
.DS_Store	6,148	File	01/09/20
archived		Directory	01/15/20

9 files and 2 directories. Total size: 1,345,209 bytes
- Remote File List:**

Filename	Filesize	Filetype	Last modified	Permissions
..				
.viminfo	735	File	01/12/2023 14:14:26	-rw-----
.kshrc	172	File	03/07/2018 08:27:43	-rw-r--r--
.emacs	334	File	09/20/2017 06:36:13	-rw-r--r--
.bashrc	264	File	10/22/2015 11:01:49	-rw-r--r--
.bash_profile	193	File	09/06/2017 12:25:27	-rw-r--r--
.bash_logout	18	File	09/06/2017 12:25:27	-rw-r--r--
.bash_history	1,502	File	01/14/2023 23:16:26	-rw-----
data		Directory	01/13/2023 19:08:44	drwxrwxr-x
code		Directory	01/03/2023 15:34:23	drwxrwxr-x
.stata17		Directory	01/03/2023 21:15:18	drwxr-xr-x

7 files and 8 directories. Total size: 3,218 bytes



# Running an Interactive Job

# In Your Mac Terminal

- ▶ Windows command shell, either command dot com or cmd dot exe, or powershell is the equivalent of terminal

```
xisong — xisong@gpc:~ — ssh xisong@gpc.sas.upenn.edu — 142x48
Last login: Tue Jan 10 15:14:49 on ttys001
xisong@setups-MacBook-Pro-3 ~ % ssh-agent Add your identification key and use the key to log into other server
SSH_AUTH_SOCK=/var/folders/cf/17brksks0m3650fw6p1gf3m0000gr/T/ssh-CMPFa16BxinI/agent.49450; export SSH_AUTH_SOCK;
SSH_AGENT_PID=49451; export SSH_AGENT_PID;
echo Agent pid 49451;
xisong@setups-MacBook-Pro-3 ~ % ssh-add Prompt the user for a private key password and add it to the list maintained by ssh-agent
xisong@setups-MacBook-Pro-3 ~ % ssh xisong@gpc.sas.upenn.edu use yourpennkey@gpc.sas.upenn.edu
(xisong@gpc.sas.upenn.edu) Password:
(xisong@gpc.sas.upenn.edu) Duo two-factor login for xisong First enter your Penkey password and then use the two-factor authentication

Enter a passcode or select one of the following options:

1. Duo Push to phone_push (iOS)
2. Phone call to XXX-XXX-0059
3. SMS passcodes to XXX-XXX-0059

Passcode or option (1-3): 1
Success. Logging you in...
Last failed login: Sun Jan 15 06:32:08 EST 2023 from 10.76.0.3 on ssh:notty
There were 3 failed login attempts since the last successful login.
Last login: Sat Jan 14 10:48:03 2023 from 10.76.0.5
Welcome to the

          GGGGGGGGGGGG  PPPPPPPPPPPPPPP  CCCCCCCCCCCC
GGG:::::::::::::G  P:::::::::::::P  CCC::::::::::::C
GG:::::::::::::G  P::::PPPPPP:::P  CC:::::::::::::C
G::::GGGGGGGG:::G  PP:::::P  P:::::P  C::::CCCCCCC:::C
G:::::G  GGGGGG  P:::P  P:::::P  C:::::C  CCCCCC
G:::::G  P:::::P  P:::::P  C:::::C
G:::::G  P::::PPPPP:::P  C:::::C
G:::::G  GGGGGGGGGG  P:::::::::::PP  C:::::C
G:::::G  G:::::G  P:::PPPPPPPP  C:::::C
G:::::G  GGGGG:::G  P:::P  C:::::C
G:::::G  G:::::G  P:::P  C:::::C
G:::::G  G:::::G  P:::::P  C:::::C  CCCCCC
G:::::GGGGGGGG:::G  PP:::::PP  C:::::CCCCCCC:::C
GG:::::::::::::G  P:::::P  CC:::::::::::::C
GGG:::::GGG:::G  P:::::P  CCG:::::::::::C
GGGGGG  GGGG  PPPPPPPPP  CCCCCCCCCCCC

          General Purpose Cluster

Use the following commands to adjust your environment:

'module avail'          - show available modules
'module load <module>' - adds a module to your environment for this session
'module unload <module>' - removes a module from your environment for this session
'module switch <m1> <m2>' - switch module version 1 with version 2
```

# Available Partitions

- ▶ **normal:** General jobs, 192GB memory available per node (many nodes available)
- ▶ **highmem:** High memory jobs; 256GB memory available per node (only three nodes available)
- ▶ **tesla:** Faster, memory unknown (only 1 node available). For students who only need a single-node server to run jobs in R or Stata, SAS provides the Tesla server, which can be run in GPC. In addition, Tesla contains one Tesla M2075 graphics processing unit (GPU) card available to faculty and students whose work lends itself to GPU processing. The server supports development in C and Fortran (through the PGI compilers) in CUDA 2.0.
- ▶ However, my jobs were killed even when I was requesting 128GB in normal or highmem nodes. If such cases happen, you'll need to email the gpc manager to increase the memory for you or help debug errors

# Available Partitions

- ▶ **normal:** General jobs, 192GB memory available per node (many nodes available)
- ▶ **highmem:** High memory jobs; 256GB memory available per node (only three nodes available)
- ▶ **tesla:** Faster, memory unknown (only 1 node available). For students who only need a single-node server to run jobs in R or Stata, SAS provides the Tesla server, which can be run in GPC. In addition, Tesla contains one Tesla M2075 graphics processing unit (GPU) card available to faculty and students whose work lends itself to GPU processing. The server supports development in C and Fortran (through the PGI compilers) in CUDA 2.0.
- ▶ However, my jobs were killed even when I was requesting 128GB in normal or highmem nodes. If such cases happen, you'll need to email the gpc manager to increase the memory for you or help debug errors

## Available Partitions

- ▶ **normal:** General jobs, 192GB memory available per node (many nodes available)
- ▶ **highmem:** High memory jobs; 256GB memory available per node (only three nodes available)
- ▶ **tesla:** Faster, memory unknown (only 1 node available). For students who only need a single-node server to run jobs in R or Stata, SAS provides the Tesla server, which can be run in GPC. In addition, Tesla contains one Tesla M2075 graphics processing unit (GPU) card available to faculty and students whose work lends itself to GPU processing. The server supports development in C and Fortran (through the PGI compilers) in CUDA 2.0.
- ▶ However, my jobs were killed even when I was requesting 128GB in normal or highmem nodes. If such cases happen, you'll need to email the gpc manager to increase the memory for you or help debug errors

## Available Partitions

- ▶ **normal:** General jobs, 192GB memory available per node (many nodes available)
- ▶ **highmem:** High memory jobs; 256GB memory available per node (only three nodes available)
- ▶ **tesla:** Faster, memory unknown (only 1 node available). For students who only need a single-node server to run jobs in R or Stata, SAS provides the Tesla server, which can be run in GPC. In addition, Tesla contains one Tesla M2075 graphics processing unit (GPU) card available to faculty and students whose work lends itself to GPU processing. The server supports development in C and Fortran (through the PGI compilers) in CUDA 2.0.
- ▶ However, my jobs were killed even when I was requesting 128GB in normal or highmem nodes. If such cases happen, you'll need to email the gpc manager to increase the memory for you or help debug errors

# Slurm Commands

- ▶ **sinfo**: Check the status of computing nodes
  - ▶ Idle means not in use. You can then request the node.
- ▶ **squeue**: View the status of all running jobs
  - ▶ Check the statuses of your jobs: `squeue | grep yourpennkey`
- ▶ **srun**: Request a partition or a node (what resources you need to run your job)
  - ▶ If you don't request anything, GPC will randomly assign you to a normal node
  - ▶ `srun --pty bash`
  - ▶ `srun --odelist=node03 --pty bash`
  - ▶ `srun --pty bash --partition=tesla`
  - ▶ `srun -p highmem --mem=160000 --pty bash`
  - ▶ Type `exit` to exit the node and switch to a new node
- ▶ **sbatch**: Submit your job in a batch file
- ▶ **scancel**: Cancel your jobs using `scancel yourjobid`
  - ▶ You can find your job id using `squeue`

# Slurm Commands

- ▶ **sinfo**: Check the status of computing nodes
  - ▶ Idle means not in use. You can then request the node.
- ▶ **squeue**: View the status of all running jobs
  - ▶ Check the statuses of your jobs: `squeue | grep yourpennkey`
- ▶ **srun**: Request a partition or a node (what resources you need to run your job)
  - ▶ If you don't request anything, GPC will randomly assign you to a normal node
  - ▶ `srun --pty bash`
  - ▶ `srun --odelist=node03 --pty bash`
  - ▶ `srun --pty bash --partition=tesla`
  - ▶ `srun -p highmem --mem=160000 --pty bash`
  - ▶ Type `exit` to exit the node and switch to a new node
- ▶ **sbatch**: Submit your job in a batch file
- ▶ **scancel**: Cancel your jobs using `scancel yourjobid`
  - ▶ You can find your job id using `squeue`

# Slurm Commands

- ▶ **sinfo**: Check the status of computing nodes
  - ▶ Idle means not in use. You can then request the node.
- ▶ **squeue**: View the status of all running jobs
  - ▶ Check the statuses of your jobs: `squeue | grep yourpennkey`
- ▶ **srun**: Request a partition or a node (what resources you need to run your job)
  - ▶ If you don't request anything, GPC will randomly assign you to a normal node
  - ▶ `srun --pty bash`
  - ▶ `srun --odelist=node03 --pty bash`
  - ▶ `srun --pty bash --partition=tesla`
  - ▶ `srun -p highmem --mem=160000 --pty bash`
  - ▶ Type `exit` to exit the node and switch to a new node
- ▶ **sbatch**: Submit your job in a batch file
- ▶ **scancel**: Cancel your jobs using `scancel yourjobid`
  - ▶ You can find your job id using `squeue`

# Slurm Commands

- ▶ **sinfo**: Check the status of computing nodes
  - ▶ Idle means not in use. You can then request the node.
- ▶ **squeue**: View the status of all running jobs
  - ▶ Check the statuses of your jobs: `squeue | grep yourpennkey`
- ▶ **srun**: Request a partition or a node (what resources you need to run your job)
  - ▶ If you don't request anything, GPC will randomly assign you to a normal node
  - ▶ `srun --pty bash`
  - ▶ `srun --odelist=node03 --pty bash`
  - ▶ `srun --pty bash --partition=tesla`
  - ▶ `srun -p highmem --mem=160000 --pty bash`
  - ▶ Type `exit` to exit the node and switch to a new node
- ▶ **sbatch**: Submit your job in a batch file
- ▶ **scancel**: Cancel your jobs using `scancel yourjobid`
  - ▶ You can find your job id using `squeue`

# Slurm Commands

- ▶ **sinfo**: Check the status of computing nodes
  - ▶ Idle means not in use. You can then request the node.
- ▶ **squeue**: View the status of all running jobs
  - ▶ Check the statuses of your jobs: `squeue | grep yourpennkey`
- ▶ **srun**: Request a partition or a node (what resources you need to run your job)
  - ▶ If you don't request anything, GPC will randomly assign you to a normal node
  - ▶ `srun --pty bash`
  - ▶ `srun --odelist=node03 --pty bash`
  - ▶ `srun --pty bash --partition=tesla`
  - ▶ `srun -p highmem --mem=160000 --pty bash`
  - ▶ Type `exit` to exit the node and switch to a new node
- ▶ **sbatch**: Submit your job in a batch file
- ▶ **scancel**: Cancel your jobs using `scancel yourjobid`
  - ▶ You can find your job id using `squeue`

For questions or help, please contact [gpc-manager@sas.upenn.edu](mailto:gpc-manager@sas.upenn.edu).

For information on GPC policy, access, installed software and how to run jobs please see: <https://computing.sas.upenn.edu/gpc>

**xisiong@gpc[~]\$ ls** list files under the current directory

code data

**xisiong@gpc[~]\$ cd /home/xisiong/data** Change the home directory

**xisiong@gpc[~/data]\$ ls**

```
8_stata_model.do      clogit_2023-1-12.log  cps_2000_2020_long_even.csv  test.419747.out  test.427798.out
8_stata_model.log    clogit-tesla.sbatch  test.419747.err              test.427798.err
```

**xisiong@gpc[~/data]\$ squeue** View the status of all running jobs

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	ODELIST(REASON)
441417	highmem	ZnCDP2_G	kmkopera	PD	0:00	1	(QOSMaxCpuPerUserLimit)
441416	highmem	ZnCDP2_G	kmkopera	PD	0:00	1	(QOSMaxCpuPerUserLimit)
437064	normal	bts_bind	aregla	PD	0:00	1	(ReqNodeNotAvail, UnavailableNodes:node01)
441415	highmem	ZnCDP2_G	kmkopera	PD	0:00	1	(QOSMaxCpuPerUserLimit)
204213_[0-69]	normal	transfer	blira	PD	0:00	1	(MaxMemPerLimit)
444429	normal	TPD0.45n	weiduow	R	20:04:51	1	node02
444428	normal	TPD0.48n	weiduow	R	21:32:30	1	node02
441414	highmem	ZnCDP2_G	kmkopera	R	1-02:53:02	1	node10
441413	highmem	ZnCDP2_G	kmkopera	R	2-02:24:32	1	node10
441403	normal	run21-15	richst	R	2-04:23:04	1	node02
440588	normal	rlax0620	weiduow	R	2-22:33:26	1	node04
437066	highmem	ZnCDP2_G	kmkopera	R	5-02:23:48	1	node10
300813	normal	run205-1	richst	R	21-11:03:14	1	node03
300812	normal	run205-1	richst	R	21-19:02:42	1	node02
300811	normal	run205-1	richst	R	21-20:49:54	1	node04
300810	normal	run205-1	richst	R	21-21:34:12	1	node02
300555	normal	run20-15	richst	R	28-17:13:25	1	node03
300554	normal	run20-15	richst	R	28-17:38:57	1	node02
300553	normal	run20-15	richst	R	28-17:44:07	1	node03
300138	normal	run21-15	richst	R	30-23:17:36	1	node02
300137	normal	run21-15	richst	R	30-23:18:00	1	node02
300136	normal	run21-15	richst	R	30-23:18:14	1	node02

**xisiong@gpc[~/data]\$ squeue | grep xisiong** View the status of all running jobs for account name xisiong

**xisiong@gpc[~/data]\$ sinfo** Check the status of the nodes. Currently, highmem on node 05 and tesla on node 09 are available

PARTITION	AVAIL	TIMELIMIT	NODES	STATE	ODELIST
normal*	up	infinite	1	drain	node01
normal*	up	infinite	2	mix	node[03-04]
normal*	up	infinite	1	alloc	node02
highmem	up	infinite	1	mix	node10
highmem	up	infinite	2	idle	node[05-06]
tesla	up	infinite	1	idle	node09
webmo	up	infinite	1	alloc	node02

xisiong@gpc[~/data]\$

# Check Available Software on the Server

## Use module avail

```
[xisong@gpc[~/data]$ srun --nodelist=node03 --pty bash  
[xisong@node03 data]$ module avail
```

```
----- /cm/local/modulefiles -----  
cluster-tools/7.2          docker/engine/1.9.1      ipmitool/1.8.15          openmpi/mlnx/gcc/64/3.0.0rc6  
cluster-tools-dell/7.2    dot                       module-git              shared  
cm-cloud-copy/7.2        flannel/0.5.4           module-info            use_owm  
cmd                       freeipmi/1.4.11         null                   version  
cmsub                     gcc/5.2.0               openldap
```

```
----- /cm/shared/modulefiles -----  
acml/gcc/64/5.3.1         intel/compiler/64/15.0/2015.5.223  nbo/7.32  
acml/gcc/fma4/5.3.1      intel/compiler/64/2017/17.0.6      ncbi-blast/2.13.0  
acml/gcc/mp/64/5.3.1     intel/mkl/64/2017/6.256            netcdf/4.8.1  
acml/gcc/mp/fma4/5.3.1   intel/mkl/mic/2017/6.256          netcdf/gcc/64/4.3.3.1
```

...

```
git/2.7.4                 mess/2021.05              SHOREmap/3.8  
GLFW/3.0.4               mpc/1.1.0                slurm/15.08.13  
globalarrays/openmpi/gcc/64/5.4  mpfr/4.0.1              slurm/20.02.4  
globalarrays/openmpi/open64/64/5.4  mpich/ge/gcc/64/3.2     smalt/0.7.6  
gmp/6.1.2                mpich/ge/open64/64/3.2  sratoolkit/2.11.2  
golang/1.13.6            mpiexec/0.84_432        STAR/2.7.10a  
gromacs/5.1.2            mvapich/gcc/64/1.2rc1   stata/13  
gsl/2.4                  mvapich/open64/64/1.2rc1 stata/14  
gurobi/8.0.0             mvapich2/gcc/64/2.2b   stata/17  
hdf5/1.12.1             mvapich2/open64/64/2.2b  tmux/3.1  
hdf5_18/1.8.16          namd/mpi/2.11           torque/6.0.0.1  
hpl/2.1                  namd/multicore/2.11     trimomatic/0.39  
hwloc/1.10.1            nbo/6                   vcftools/0.1.15  
intel/compiler/32/2017/17.0.6      nbo/7                   vovpal_wabbit/current  
[xisong@node03 data]$
```

```
[[xisiong@node02 ~]$ module load stata/17
[[xisiong@node02 ~]$ stata-mp
```

Load the module you want to use  
Open Stata MP version; now you can use it interactively



17.0

MP-Parallel Edition

Statistics and Data Science

Copyright 1985-2021 StataCorp LLC  
StataCorp  
4905 Lakeway Drive  
College Station, Texas 77845 USA  
800-STATA-PC                    <https://www.stata.com>  
979-696-4600                    [stata@stata.com](mailto:stata@stata.com)

Stata license: 2-user 8-core network perpetual  
Serial number: 501706320672  
Licensed to: Research Data Services  
University of Pennsylvania

Notes:

- 1. Unicode is supported; see help unicode\_advice.
- 2. More than 2 billion observations are allowed; see help obs\_advice.
- 3. Maximum number of variables is set to 5,000; see help set\_maxvar.

```
[. set obs 10000
Number of observations (_N) was 0, now 10,000.
```

```
[. gen x = rnormal(0,1)
```

```
[. sum x
```

Variable	Obs	Mean	Std. dev.	Min	Max
x	10,000	.0003138	.9946919	-4.468302	4.327502

```
[. exit
no; dataset in memory has changed since last saved
Save the data or specify option clear to exit anyway.
r(4);
```



# Running a Batch Job (recommend!)

# Run Your Models with Slurm Job Scheduler

If you want to run your jobs by means of wrapper scripts rather than by entering the commands line-by-line, you will need to use Slurm

- ▶ Slurm documentation  
<https://computing.sas.upenn.edu/gpc/job/slurm>
- ▶ More information about Slurm
  - ▶ Slurm Quickstart guide  
<https://slurm.schedmd.com/quickstart.html>
  - ▶ Slurm command summary  
<https://slurm.schedmd.com/pdfs/summary.pdf>
  - ▶ Slurm manual pages  
[https://slurm.schedmd.com/man\\_index.html](https://slurm.schedmd.com/man_index.html)

# A Slurm Example: clogit.batch

Save the following commands in a text file with the format .batch

```
clogit-tesla.sbatch — hpc
1  #!/bin/bash
2  #SBATCH -p highmem
3  #SBATCH -J clogit-2023-1-12
4  #SBATCH --mail-type=END,FAIL
5  #SBATCH --mail-user=xisong@upenn.edu
6  #SBATCH --nodes=1
7  #SBATCH --ntasks=1
8  #SBATCH --mem=100G
9  #SBATCH --time=36:00:00
10 #SBATCH -o test."%j".out
11 #SBATCH -e test."%j".err
12 #SBATCH --chdir=/home/xisong/data
13
14 module load stata/17
15 stata-mp -b do "8_stata_model.do"
16
```

# A Slurm Example: clogit.batch

See other slurm examples:

<https://computing.sas.upenn.edu/gpc/job/slurm>

```
clogit-tesla.sbatch — hpc
1  #!/bin/bash
2  #SBATCH -p highmem    Request the high memory partition
3  #SBATCH -J clogit-2023-1-12  Create the following error and output files using this name
4  #SBATCH --mail-type=END,FAIL  Send emails to the following address when the job ends or fails
5  #SBATCH --mail-user=xisong@upenn.edu
6  #SBATCH --nodes=1    Request 1 node
7  #SBATCH --ntasks=1  Request 1 task
8  #SBATCH --mem=100G  Request 100GB memory
9  #SBATCH --time=36:00:00  Run the model for 36 hours
10 #SBATCH -o test."%j".out
11 #SBATCH -e test."%j".err
12 #SBATCH --chdir=/home/xisong/data  Save the files in this folder
13
14 module load stata/17          Load Stata 17
15 stata-mp -b do "8_stata_model.do"  Run the do file using the multiple processor version of Stata
16
```

# Submit the sbatch file

- ▶ Submit your sbatch file by typing `clogit.sbatch` in the terminal (make sure that the file is in the current directory; otherwise, use `cd` to change the directory)

## Other Useful Commands

- ▶ use `squeue` to see the status of all running jobs on the server
- ▶ use `squeue | grep yourpennkey` to see your running jobs
  
- ▶ `tail -f filename` to see the tail of a file. `-f` means follows
- ▶ `less filename` to see the whole file
- ▶ e.g., I used `tail -f clogit.log` to view my log file from Stata. You can also use `tail -f clogit.err` to see if there is any error report.
  
- ▶ use `nano` to quickly edit your files on the server rather than upload a new file to the server using Filezilla.
- ▶ e.g., `nano clogit.sbatch`. Then use `control+C` to quit and `Y` to save.

# More Linux and Slurm Commands

- ▶ Basic Overview of Linux:

<https://www.digitalocean.com/community/tutorials/an-introduction-to-linux-basics>

- ▶ Basic File Management Commands: (e.g., ls, cat, less, tail commands)

[https://www.linuxtrainingacademy.com/linux-commands-cheat-sheet/#5\\_8211\\_FILE\\_AND\\_DIRECTORY\\_COMMANDS](https://www.linuxtrainingacademy.com/linux-commands-cheat-sheet/#5_8211_FILE_AND_DIRECTORY_COMMANDS)

- ▶ Nano Text Editor:

<https://linuxize.com/post/how-to-use-nano-text-editor/>

- ▶ Basic Slurm Commands and SBATCH configuration Options (understanding the SBATCH -J and other options, etc):

<https://hpc.nmsu.edu/discovery/slurm/slurm-commands/>

- ▶ Official Slurm Documentation Cheat Sheet:

<https://slurm.schedmd.com/pdfs/summary.pdf>

## More Information

- ▶ GPC website:  
<https://computing.sas.upenn.edu/gpc/getstarted>
- ▶ Contact GPC manager: [gpc-manager@sas.upenn.edu](mailto:gpc-manager@sas.upenn.edu)
- ▶ Be patient and keep sending emails if you don't receive a reply!

## Other HPC Resources on Campus

- ▶ <https://research-it.wharton.upenn.edu/documentation/>  
You will need to be a Wharton affiliate

The screenshot shows the Wharton Research-IT website. At the top, there is a navigation bar with the Wharton logo and links for Faculty, Youth Program, Undergrad, MBA, EMBA, PhD, ExecEd, Wharton Online, Alumni, and Knowledge @ Wharton. Below this is a dark blue header with 'RESEARCH-IT' and a secondary navigation bar with links for About, Behavioral Lab, Blog, Research Programming, Data, High Performance Computing (which is highlighted), Windows Research Computing, Publication, and Tools. The main content area features a sidebar menu on the left with items like Troubleshooting, Hardware, Getting an Account, Access, Job Management, Programming Best Practices, Sharing, Storage, Training Basics, and Cloud Bursting. The main content is titled 'High Performance Computing' and includes a paragraph about the Wharton Computing's Research and Innovation team, a link to 'Current HPC Cluster Status', and sections for 'Cloud HPC' and 'Overview'. The 'Overview' section describes the Wharton HPC Cluster (HPCC) and lists supported software like Matlab, Mathematica, R, Stata, SAS, and MySQL.

# School of Medicine HPC

- ▶ Although PSOM HPC not free, it is accessible to all Penn researchers even if they are not affiliated with the medical school. <https://www.med.upenn.edu/hpc/billing-pricing-service-descriptions.html>
- ▶ Given that HPC is a commercial software, it is not available on Penn Medicine's HPC. Alternatively, you can use Penn Medicine Limited Performance Computing (LPC) if you are a PSOM faculty, staff or affiliate
- ▶ Contact [psom-pmacshpc@penndicine.upenn.edu](mailto:psom-pmacshpc@penndicine.upenn.edu)



## High Performance Computing

Penn Medicine Academic Computing Services

Home

The Basics ▾

Account Process

Billing, Pricing and Service Descriptions

Support

Data Backup

Technical ▾

Grant

Information

System Status - ON-LINE

Service Request

Wiki

About ▾

Contact Us

Contact Us

High Performance Computing



» The Basics » Billing, Pricing and Service Descriptions

## Billing, Pricing and Service Descriptions

PMACS HPC usage costs are "pass-through" in that the PMACS HPC system cost structure seeks to cover the operating costs of the system with budget-neutrality as the goal. Costs are as follows:

Service	Cost
Compute-Hour (Virtual Core)	\$0.035
Disk storage — Per-GB/mo. RAID-6 disk (not backed up)	\$0.055
Archive Storage/GB/mo/ (mirrored tape)	\$0.015
Consulting/Customization	\$95/hr.

### Compute (\$0.035/hr.)

The computational hour can also be referred to as a "core-hour" or "virtual-core-hour." The system is configured in such a manner as to allow more computational-hours than the number of physical processors in the system. This virtualization provides a greater computational capacity than if each physical processing core was utilized individually, therefore minimizing queue wait times for jobs to complete. Each virtual core is assigned up to 8GB of RAM, however alternate RAM configurations are available as needed.

### Disk storage (\$0.055/GB/mo.)

The PMACS HPC houses 1.8 Petabytes (1,800 Terabytes) of disk storage. Some of this storage is used for system utility functions; however most of the storage (about 80%) is available for the storage and processing of HPC jobs.

Data is accessible via both NFS (Linux/Mac) and CIFS (Windows) drive shares. Shares can be mapped to biomedical equipment, such as high-throughput sequencers capable of mapping NFS and/or CIFS shares provided the equipment is also managed by PMACS staff and is on a PMACS managed network.

Once data exists on the PMACS HPC disk storage it is then accessible by the compute cores and queue management software, and can become part of computational jobs.

# Department of Economics Linux Hawk Cluster

- ▶ The cluster contains 144 compute nodes. Students can take advantage of parallel processing, which allows a program to run on more than one compute node at a time.
- ▶ Students can access up to 36 nodes at a time. The hawk cluster runs the Linux operating system and has available both Fortran (Intel and GNU) and C compilers.
- ▶ The cluster also has available a parallel optimizer called HOPSPACK.
- ▶ To get an account on hawk, contact Professor Petra Todd.

# Amazon Web Services (AWS)

← → ↻ aws.amazon.com/openservicepricing/?did=ap\_card&trk=ap\_card

aws re:invent Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Customer Enablement Events Explore More

Amazon OpenSearch Service Overview Features Pricing Getting Started Resources Migrations Partners FAQs

### On-Demand Instance pricing

Region: US East (Ohio)

General Purpose - Current Generation	vCPU	Memory (GB)	Instance Storage (GB)	Price Per hour
t3.small.search	2	2	EBS Only	\$0.056
t3.medium.search	2	4	EBS Only	\$0.075
t2.micro.search	1	1	EBS Only	\$0.018
t2.small.search	1	2	EBS Only	\$0.056
t2.medium.search	2	4	EBS Only	\$0.075
m5g.large.search	2	8	EBS Only	\$0.128
m5g.xlarge.search	4	16	EBS Only	\$0.256
m5g.2xlarge.search	8	32	EBS Only	\$0.511
m5g.4xlarge.search	16	64	EBS Only	\$1.023
m5g.8xlarge.search	32	128	EBS Only	\$2.045
<b>m5g.12xlarge.search</b>	<b>48</b>	<b>192</b>	<b>EBS Only</b>	<b>\$3.068</b>
m5.large.search	2	8	EBS Only	\$0.142
m5.xlarge.search	4	16	EBS Only	\$0.283
m5.2xlarge.search	8	32	EBS Only	\$0.566
m5.4xlarge.search	16	64	EBS Only	\$1.133
m5.12xlarge.search	48	192	EBS Only	\$3.398
m4.large.search	2	8	EBS Only	\$0.151
m4.xlarge.search	4	16	EBS Only	\$0.301
m4.2xlarge.search	8	32	EBS Only	\$0.603
m4.4xlarge.search	16	64	EBS Only	\$1.207
m4.10xlarge.search	40	160	EBS Only	\$3.017

# Acknowledgement

I would like to thank Alex Bond, Hanming Fang, Hadiya Gaiters, Michael Lachanski, Emilio Parrado, Petra Todd, Kyle Jermaine Small, Zihan Sun, and Xiuqi Yang for helpful discussions that made my work and this presentation possible.