Copy of FAQ - How do I use STATA on the HPC clusters?

<table>
<thead>
<tr>
<th>Quick Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HPC Home</strong></td>
</tr>
<tr>
<td><strong>Getting an account</strong></td>
</tr>
<tr>
<td><strong>Gentle Introduction to using HPC</strong></td>
</tr>
<tr>
<td><strong>Getting started on Prince</strong></td>
</tr>
<tr>
<td><strong>Prince How-to Articles</strong></td>
</tr>
<tr>
<td><strong>Logging in</strong></td>
</tr>
<tr>
<td><strong>Windows</strong></td>
</tr>
<tr>
<td><strong>Mac / Linux</strong></td>
</tr>
<tr>
<td><strong>Clusters and Storage</strong></td>
</tr>
<tr>
<td><strong>Prince (HPC)</strong></td>
</tr>
<tr>
<td><strong>Dumbo (Hadoop)</strong></td>
</tr>
<tr>
<td><strong>Brooklyn (OpenStack)</strong></td>
</tr>
</tbody>
</table>
Damera (NYU Abu Dhabi)

Transferring data to/from the clusters
Transferring data to/from Prince cluster using Globus

Submitting jobs with sbatch
Available software
Licensed Software Available on the HPC Cluster
Building Software
Slurm Tutorial
Tutorials
FAQs
Scratch Area Cleanup
Programming for Biologists
Something went wrong!

Why does running "ls" on /scratch take so long?
I can't login
When trying to login, I get warning s about "HOST IDENTIFICATION HAS CHANGED"

What happened to my data on /scratch?
In the library, my wireless connection keeps dropping out. How can I fix it?
I'm getting a "module: command not found" error.

Warning: no access to tty (Bad file descriptor), Thus no job control in this shell.

I get an error "Warning: no display specified." when I use -X flag with ssh.

Who killed my job, and why?

I got an email "Please do not run jobs on login nodes."

Running jobs

What resources can and should I request?

Can I make sure a job gets executed only after another one completes?

How do I log in to a specific node?
How can I make sure my job is running smoothly?

My job will take longer than 48 hours, what should I do?

My job needs (MySQL, some other service) to be running

I want to run a job at 9am every day

Using software

How do I run ... (esp, needs a license)

a STATA job?

a Gaussian job?

a Matlab job?
a parallel non-MPI job (e.g. Julia)?

I can't find (some software package)

Can you install (some software package)?

How can I view a PDF file on Prince?

Managing data

How much of my file/space quota have I used?

How do I give my colleague access to my files?

How do I get the best transfer speed to or from BuTinah?
I have a huge amount of data that I want to compress for storage or transfer.

A: To run STATA jobs, you need to:

1. Prepare a STATA do file, such as "stata-test.do" which may include,

```stata
sysuse auto, clear
summarize
graph twoway (scatter mpg weight) (lfit mpg weight)
graph export stata-test.ps, replace
```

StataMP was also installed on the USQ, which is the parallel version of Stata. To use multiple processors in StataMP, just insert "set processors X" as a line in your do file, where X is the number of processors and it should be equal to the "ppn" number in your PBS script.

2. Create a PBS script "run-stata.pbs" to run STATA jobs in the batch mode. The content in this file can be like this,

```bash
#!/bin/csh -f
#PBS -V
#PBS -S /bin/tcsh
#PBS -N stata-test
#PBS -l nodes=1:ppn=1,walltime=01:00:00
#PBS -M NetID@nyu.edu
#PBS -m abe
source /etc/profile.d/env-modules.csh
module load stata/11
cd /scratch/NetID/hpc-tutorial/stata
stata -b do stata-test.do
```

**Note**
Be sure to substitute your own "NetID" for NetID.

You need to change the paths according to your directory. Please refer to this page for more information.

3. Then submit the job by typing

```bash
$qsub run-stata.pbs
```