Dear Cluster User:

The following are some basic guidelines to start submitting jobs on the cluster. All our compute nodes have been upgraded to Linux Centos 7.3. In addition, the Conda package manager has been enabled to provide the flexibility application installations by the end users right in their own virtual environments. Combining this feature with our SLURM, our job scheduler, our users have the flexibility to work on single applications or complex pipelines that require a number of applications. No more module-per-application is required as before, but only a single module to activate or deactivate your **virtual environment**.

The following steps will get you up and running
Login to the cluster via SSH. If you need a VPN access to connect to the cluster outside the CUNY network, please email me to coordinate access through ICIT.

SHH To head node

```
ssh  username@bioit.hunter.cuny.edu
or
ssh username@146.95.252.94
Welcome to Bright release         7.3

                                  Based on CentOS Linux 7
                                  ID: #000002
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
---------------------------------------------------------------
```

Once you are in, load the following modules to get your profile and environment variables ready

```
module load shared
module load anaconda/default
```

The first time you login you need to create your first virtual environment; you can choose to have it in the default shared location or in your private home directory. The virtual environment will allow you to install applications or run existing applications. To create your environment, you can either do so from scratch or by cloning an existing one.
If you decide to clone the **default virtual environment**, there are over 400 Bioinformatics applications already pre-installed and you may use any one of them.
If you decide to create a new environment from scratch, this will be empty and you will need to install programs manually.

Option 1: Clone the existing virtual environment into your home folder

Before cloning an existing environment, you can selectively choose what to clone. We have prepared a number of environments pre-packaged for you.

Which Environments are available to clone?

Run the command **conda info -e**

```
[lijeron@ctbr-cluster-hn1 envs]$ conda info -e

# conda environments:
#
BiocondaBACKUP          /cm/shared/apps/anaconda/envs/BiocondaBACKUP
BiocondaMaster2         /cm/shared/apps/anaconda/envs/BiocondaMaster2
BiocondaMaster3         /cm/shared/apps/anaconda/envs/BiocondaMaster3
Py2                     /cm/shared/apps/anaconda/envs/Py2
Py3                     /cm/shared/apps/anaconda/envs/Py3
R31                     /cm/shared/apps/anaconda/envs/R31
R32                    /cm/shared/apps/anaconda/envs/R32
R33                   /cm/shared/apps/anaconda/envs/R33
root                   *  /cm/shared/apps/anaconda
```

The output above shows the available environments. Below is an explanation for each.

BiocondaBACKUP	**Backup folder - Do not use**
BiocondaMaster2	Master Environment based on Python 2.7
BiocondaMaster3	Master Environment based on Python 3
Py2	Basic environment based on Python 2.7
Py3	Basic environment based on Python 3.0
R31	Basic environment based on R 3.1
R32	Basic environment based on R 3.2
R33	Basic environment based on R 3.3
root	This is the default environment. It is about 4 Gigabytes in size and it's the most up to date. The software manifest will be listed at the end of this document.

Software Manifest

BiocondaMaster2	https://goo.gl/1Y6gPl
BiocondaMaster3	https://goo.gl/vE4M7A
Py2	https://goo.gl/Pdmi83
Py3	https://goo.gl/IV7Fxb
R31	https://goo.gl/4JMSpF
R32	https://goo.gl/IrbL3n
R33	https://goo.gl/eKVYr0
root	https://goo.gl/nw44nh

How to clone the default environment?

```
conda create --name MyVirtualEnv --clone=/cm/shared/apps/anaconda/envs/root
```

The command above will create a new virtual environment in your home folder based on the root default environment. You can clone any of the environments listed.

```
conda create --name MyVirtualEnv --clone=/cm/shared/apps/anaconda/envs/BiocondaMaster2
conda create --name MyVirtualEnv --clone=/cm/shared/apps/anaconda/envs/BiocondaMaster3
conda create --name MyVirtualEnv --clone=/cm/shared/apps/anaconda/envs/Py3
conda create --name MyVirtualEnv --clone=/cm/shared/apps/anaconda/envs/R33
```

Where is my newly cloned environment ? Go to the following directory

```
/home/YourUserName/.conda/envs/
```

How to use your environment?

In order to use your newly created environment, you must activate it. When done, you can deactivate. After the clone is complete, you need to follow the procedure outlined in the beginning: SSH to cluster, load module for anaconda, and then you can activate your environment.

```
source activate MyVirtualEnv
```

The command above will activate your environment and you will be able to use all the applications installed. You can remove or install applications as needed.

How to install applications on your virtual environment?

Because the main package manager we are using is based on conda, by setting up the **right channels** we will be able to query databases and install applications, including dependencies and binaries from multiple sources. The bioconda channel is of particular importance because that team has introduced almost every Bioinformatics-related application into this package manager. This helps to manage installation processes and requirements. For more information about bioconda, visit: https://bioconda.github.io/

Adding Channels conda config --add channels conda-forge

```
conda config --add channels defaults
conda config --add channels r
conda config --add channels bioconda
```

Search for packages

conda search PackageName
Example:

```
conda search samtools
```

```
Fetching package metadata .....
bioconductor-rsamtools      1.22.0          r3.2.2_0  bioconda
                            1.22.0          r3.2.2_1  bioconda
                            1.24.0          r3.3.1_0  bioconda
                            1.26.1          r3.3.1_0  bioconda
```

The output above gives us information about the availability of samtools. We can choose to install any version and conda/bioconda will handle the dependencies.

Command to install a specific version of an application:

```
conda install samtools=1.22.0
```

SLURM, Work Load Manager

Using SLURM, our jobs management platform along with your virtual environment. This will allow you to submit jobs to the cluster using the job management/scheduler.
Once you activate your environment and you are sure your application is available, you can simply request resources from the cluster and submit your job using srun or sbatch. Follow these steps to use SLURM

Get a compute node assigned to you:

```
[username@ctbr-cluster-hn1 envs]$ salloc
salloc: Granted job allocation 48
```

Salloc requests resources form the cluster. The number 48 in the example above is the allocation number provided to you. You can submit jobs directly to your allocation or run a batch script.
Submitting a job using the command SRUN that will make use of the resources allocated. For this example we will use Bowtie2.

```
[username@ctbr-cluster-hn1 envs]$ srun bowtie2 -x Bowtie_2/hg38 -1 RNAseq_sample_data/adrenal_1.fastq -2
RNAseq_sample_data/adrenal_2.fastq -S alignment.sam

srun = Directs the job to the compute node assigned by salloc
bowtie2 = Main binary inside the virtual environment recently activated.
Bowtie_2/hg38 = This is the reference human genome
```

RNAseq_**sam**ple_data/adrenal_1.fastq = First input strand
RNAseq_**sam**ple_data/adrenal_2.fastq = Second input strand
alignment.**sam =** The output file will be placed in the current directory

Cheat Sheet

The following PDF is a complete guide on conda. The list below are some basic and useful conda commands you may want to experiment with.
https://conda.io/docs/_downloads/conda-cheatsheet.pdf

conda list --revisions	**Shows all packages installed to allow reverting back**
conda list --revisions $Number	Allows you to revert back to any particular revision
conda remove PackageName	Allows you to remove a specific package that you don't need/want-
conda create --name MyEnvironment	Command to create your own default virtual environment. It will be installed in /home/UserName/.conda/envs/ $EnvironmentName
source activate MyEnvironment	This command activates your environment. You can have more than one environment
source deactivate MyEnvironment	This command is to deactivate your environment.
conda create --name $EnvironmentName --clone $OriginalEnvironment	This command is to clone an existing environment.
conda create --name MyClone --clone root	This clones the default environment.
conda create --name MyEnv01 python=2.7	This will create an environment in your home folder based on Python 2.7. you can then use conda install PackageName to add packages

Should you have any question, please email me @ clijeron@hunter.cuny.edu