

# 2021 Oct Jupyter Workshop

Yasasvy Nanyam

## Table of Contents

1. [Introduction \(Introduction\)](#)
2. [Spawning a session \(Spawning-a-session\)](#)
3. [UI walkthrough \(UI-walkthrough\)](#)
4. [Magic commands \(Magic-commands\)](#)
5. [Creating/using Notebook Kernels \(Creating/using-Conda-envs\)](#)
6. [Python notebooks and HPC \(python-notebooks-and-HPC\)](#)
7. [Demo \(Demo\)](#)
8. [Future considerations \(future-considerations\)](#)

## Introduction

## JupyterHub

- Used to serve Jupyter Notebook/Lab for multiple users
- Okta integration for user authentication
- Integrates with SLURM to allocate resources

## Jupyter Notebook

- Combination of a web application and IDE
- Develop, execute, debug code
- Share your work (with caveats!)
- Future development uncertain - see <https://github.com/jupyter/notebook/issues/6210> (<https://github.com/jupyter/notebook/issues/6210>)

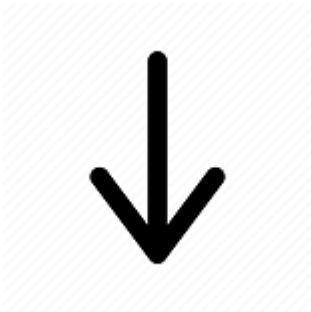
This notebook and the input data files are located in `/ptmp/jupyterhub-workshop-oct-2021`

## Spawning a session

Condo - <https://hpc-jupyterhub.its.iastate.edu> (<https://hpc-jupyterhub.its.iastate.edu>)

HPC-Class - <https://class-jupyterhub.its.iastate.edu> (<https://class-jupyterhub.its.iastate.edu>)

Sign in with Okta@ISU



# Spawner Options

Below are the commands to launch an instance of Jupyter Notebook on the ISU Condo cluster.

----

**Node Type**

short ▾

**Number of Cores (required)**

**Job duration(HH:MM:SS, required)**

**Working Directory (optional, defaults to \$HOME)**

**Slurm Sbatch Args (optional)**

Specify --gres here if selecting GPU partition

----

Spawn

## Class Cluster

Partition	Max Cores/Node	Max Walltime
Short	16	01:00:00
GPU	16	06:00:00
Debug	16	01:00:00
Mem	NA	NA

## Condo Cluster

Partition	Max Cores/Node	Max Walltime
Short	16	01:00:00
GPU	20	06:00:00
Debug	16	02:00:00
Mem	32	7-00:00:00

**UI Walkthrough**

Select items to perform actions on them.

Upload New ↕


0 /		Name ↓	Last Modified	File size
<input type="checkbox"/>	bitreverse		3 years ago	
<input type="checkbox"/>	ccp4_tmp		5 years ago	
<input type="checkbox"/>	cluster-config		2 months ago	
<input type="checkbox"/>	CUDA		4 years ago	
<input type="checkbox"/>	ondemand		21 days ago	
<input type="checkbox"/>	project_opencv		3 years ago	
<input type="checkbox"/>	R		2 years ago	
<input type="checkbox"/>	R_test		4 years ago	
<input type="checkbox"/>	rhel7-cluster		4 years ago	
<input type="checkbox"/>	spark-warehouse		2 years ago	
<input type="checkbox"/>	sruntest		4 years ago	
<input type="checkbox"/>	tensorflow		3 years ago	
<input type="checkbox"/>	Untitled Folder		2 years ago	
<input type="checkbox"/>	2021-OCT-Jupyter-Workshop.ipynb	Running	40 minutes ago	224 kB
<input type="checkbox"/>	pytorch_logistic_regression.ipynb		a year ago	202 kB
<input type="checkbox"/>	aft_survival_regression.py		2 years ago	2.15 kB
<input type="checkbox"/>	aft_survival_regression.py.1		2 years ago	2.15 kB
<input type="checkbox"/>	app_list_hpc-class		4 years ago	2.38 kB
<input type="checkbox"/>	bitreverse.zip		11 years ago	1.69 kB
<input type="checkbox"/>	cnvator2VCF.pl		3 years ago	5.37 kB
<input type="checkbox"/>	example.c		4 years ago	200 B

localhost

jupyter Welcome to Python (unsaved changes) Python 3

File Edit View Insert Cell Kernel Help Menubar

Markdown CellToolbar Toolbar Cell Mode Indicator Kernel Indicator

jupyter 

**Welcome to the Temporary Notebook (tmpnb) service!**

This Notebook Server was **launched just for you**. It's a temporary way for you to try out a recent development version of the IPython/Jupyter notebook.

**WARNING**  
Don't rely on this server for anything you want to last - your server will be *deleted after 10 minutes of inactivity*.

Your server is hosted thanks to [Rackspace](#), on their on-demand bare metal servers, [OnMetal](#).

**Cell In Command Mode**

**Run some Python code!**

To run the code below:

1. Click on the cell to select it.
2. Press **SHIFT+ENTER** on your keyboard or press the play button (▶) in the toolbar above.

A full tutorial for using the notebook interface is available [here](#).

```
In [ ]: %matplotlib inline
import pandas as pd
import numpy as np
import matplotlib
```

Source - [https://jupyter-notebook.readthedocs.io/en/stable/ui\\_components.html](https://jupyter-notebook.readthedocs.io/en/stable/ui_components.html) ([https://jupyter-notebook.readthedocs.io/en/stable/ui\\_components.html](https://jupyter-notebook.readthedocs.io/en/stable/ui_components.html))

For an interactive tour select *Help* and then select *User Interface Tour*

## Magic Commands

Magic commands are provided by IPython kernel

```
In [1]: %lsmagic
```

```
Out[1]: Available line magics:
```

```
%alias %alias_magic %autoawait %autocall %automagic %autosave %bookmark %cat %cd %clear  
%colors %conda %config %connect_info %cp %debug %dhist %dirs %doctest_mode %ed %edit %  
env %gui %hist %history %killbgscripts %ldir %less %lf %lk %ll %load %load_ext %loadp  
y %logoff %logon %logstart %logstate %logstop %ls %lsmagic %lx %macro %magic %man %ma  
tplotlib %mkdir %more %mv %notebook %page %pastebin %pdb %pdef %pdoc %pfile %pinfo %p  
info2 %pip %popd %pprint %precision %prun %psearch %psource %pushd %pwd %pycat %pylab  
%qtconsole %quickref %recall %rehashx %reload_ext %rep %rerun %reset %reset_selective %r  
m %rmdir %run %save %sc %set_env %store %sx %system %tb %time %timeit %unalias %unlo  
ad_ext %who %who_ls %whos %xdel %xmode
```

```
Available cell magics:
```

```
%%! %%HTML %%SVG %%bash %%capture %%debug %%file %%html %%javascript %%js %%latex %%ma  
rkdown %%perl %%prun %%pypy %%python %%python2 %%python3 %%ruby %%script %%sh %%svg %%  
sx %%system %%time %%timeit %%writefile
```

```
Automagic is ON, % prefix IS NOT needed for line magics.
```

```
In [2]: %ls /ptmp/jupyterhub-workshop-oct-2021
```

```
2021-OCT-Jupyter-Workshop.ipynb  iris.csv  
hello.py                        kernel-install-instructions
```



```
In [3]: %pip list
```

DEPRECATION: The default format will switch to columns in the future. You can use --format=(legacy|columns) (or define a format=(legacy|columns) in your pip.conf under the [list] section) to disable this warning.

backcall (0.2.0)  
cyclcr (0.10.0)  
decorator (5.1.0)  
entrypoints (0.3)  
ipykernel (5.5.6)  
ipython (7.16.1)  
ipython-genutils (0.2.0)  
jedi (0.18.0)  
joblib (1.1.0)  
jupyter-client (7.0.6)  
jupyter-core (4.9.1)  
kiwisolver (1.3.1)  
matplotlib (3.3.4)  
nest-asyncio (1.5.1)  
numpy (1.19.5)  
pandas (1.1.5)  
parso (0.8.2)  
pexpect (4.8.0)  
pickleshare (0.7.5)  
Pillow (8.4.0)  
pip (9.0.1)  
prompt-toolkit (3.0.21)  
ptyprocess (0.7.0)  
Pygments (2.10.0)  
pyparsing (3.0.3)  
python-dateutil (2.8.2)  
pytz (2021.3)  
pyzmq (22.3.0)  
scikit-learn (0.24.2)  
scipy (1.5.4)  
seaborn (0.11.2)  
setuptools (28.8.0)  
six (1.16.0)  
threadpoolctl (3.0.0)  
tornado (6.1)  
traitlets (4.3.3)  
wcwidth (0.2.5)

You are using pip version 9.0.1, however version 21.3.1 is available.

You should consider upgrading via the 'pip install --upgrade pip' command.

Note: you may need to restart the kernel to use updated packages.

In [4]: `%env`

```
Out[4]: {'SLURM_NODELIST': 'hpc-class05',
'MANPATH': '/usr/share/lmod/lmod/share/man:',
'SLURM_JOB_NAME': 'jupyterhub',
'XDG_SESSION_ID': 'c4',
'SLURM_NTASKS_PER_NODE': '2',
'HOSTNAME': 'hpc-class05',
'SLURM_TOPOLOGY_ADDR': 'hpc-class05',
'SLURMD_NODENAME': 'hpc-class05',
'SLURM_PRIO_PROCESS': '0',
'SLURM_NODE_ALIASES': '(null)',
'__LMOD_REF_COUNT_MODULEPATH': '/opt/rit/singularity/modules:1;/opt/rit/spack-modules/lmod/linux-rhel7-x86_64/Core:1;/shared/hpc/modulefiles:1;/etc/modulefiles:1;/usr/share/modulefiles:1;/usr/share/modulefiles/Linux:1;/usr/share/modulefiles/Core:1;/usr/share/lmod/lmod/modulefiles/Core:1',
'SHELL': '/bin/bash',
'SLURM_EXPORT_ENV': 'PATH,LANG,CONDA_PREFIX,JUPYTERHUB_API_TOKEN,JPY_API_TOKEN,JUPYTERHUB_CLIENT_ID,JUPYTERHUB_HOST,JUPYTERHUB_OAUTH_CALLBACK_URL,JUPYTERHUB_USER,JUPYTERHUB_SERVER_NAME,JUPYTERHUB_API_URL,JUPYTERHUB_ACTIVITY_URL,JUPYTERHUB_BASE_URL,JUPYTERHUB_SERVICE_PREFIX,USER,HOME,SHELL',
'LMOD_ROOT': '/usr/share/lmod',
'HISTSIZE': '1000',
'SLURM_JOB_QOS': 'normal',
'MODULEPATH_ROOT': '/usr/share/modulefiles',
'TMPDIR': '/scratch/ynanyam/101445',
'JPY_API_TOKEN': '<hidden>',
'SLURM_TOPOLOGY_ADDR_PATTERN': 'node',
'LMOD_PKG': '/usr/share/lmod/lmod',
'QTDIR': '/usr/lib64/qt-3.3',
'LMOD_VERSION': '8.2.7',
'QTINC': '/usr/lib64/qt-3.3/include',
'QT_GRAPHICSSYSTEM_CHECKED': '1',
'SLURM_MEM_PER_CPU': '7770',
'USER': 'ynanyam',
'SLURM_NNODES': '1',
'LMOD_sys': 'Linux',
'JUPYTERHUB_CLIENT_ID': 'jupyterhub-user-ynanyam%40iastate.edu',
'JUPYTERHUB_BASE_URL': '/',
'SLURM_JOBID': '101445',
'JUPYTERHUB_API_TOKEN': '<hidden>',
'SLURM_NTASKS': '2',
'SINGULARITY_TMPDIR': '',
'_ModuleTable001_': 'X01vZHVszVRhYmxlXz17WyJNVHZlcnNpb24iXT0zLFsiY19yZWJlaWxkVGltZSJdpWZhbHhN1LFs
```

```
iY19zaG9ydFRpbWUiXT1mYWxzZSxkZXB0aFQ9e30sZmFtaWx5Pxt9LG1UPxt9LG1wYXR0QT17Ii9vcHQvcml0L3Npbmd1bGFy
aXR5L21vZHVvsZXMiLCIvb3B0L3JpdC9zcGFjay1tb2R1bGVzL2xtb2QvbGludXgtcmhlbDcteDg2XzY0L0Nvcml0L3Npbmd1bGFy
mVkl2hwYy9tb2R1bGVmaWxlcYIsIi9ldGMvbW9kdWxlZmlsZXMiLCIvdXNyL3NoYXJlL21vZHVvsZWZpbGVzIiwil3Vzci9zaG
FyZS9tb2R1bGVmaWxlcY9MaW5leCIsIi9lc3Ivc2hhcmUvbW9kdWxlZmlsZXMvQ29yZSIsIi9lc3Ivc2hhcmUvbG1vZC9sbW9
kL21vZHVvsZWZpbGVzL0Nvcml0L3Npbmd1bGFyJzeXN0ZW1CYXNlTVBB',
  'MAIL': '/var/spool/mail/ynanyam',
  'PATH': '/sbin:/bin:/usr/sbin:/usr/bin:/shared/hpc/jupyterhub/1.1.0/bin:/shared/hpc/miniconda/4.
8.2/bin',
  'SLURM_TASKS_PER_NODE': '2',
  'SLURM_CONF': '/etc/slurm/slurm.conf',
  'SLURM_WORKING_CLUSTER': 'hpc-class:172.22.100.1:6817:9216:101',
  '_': '/shared/hpc/jupyterhub/1.1.0/bin/jupyterhub-singleuser',
  'SLURM_JOB_ID': '101445',
  'CONDA_PREFIX': '/shared/hpc/miniconda/4.8.2',
  'PWD': '/home/ynanyam',
  'SLURM_JOB_USER': 'ynanyam',
  'LANG': 'en_US.UTF-8',
  'JUPYTERHUB_SERVER_NAME': '',
  'MODULEPATH': '/opt/rit/singularity/modules:/opt/rit/spack-modules/lmod/linux-rhel7-x86_64/Cor
e:/shared/hpc/modulefiles:/etc/modulefiles:/usr/share/modulefiles:/usr/share/modulefiles/Linux:/u
sr/share/modulefiles/Core:/usr/share/lmod/lmod/modulefiles/Core',
  '_ModuleTable_Sz_': '2',
  'SLURM_JOB_UID': '366067',
  'SLURM_NODEID': '0',
  'SLURM_SUBMIT_DIR': '/',
  'LMOD_CMD': '/usr/share/lmod/lmod/libexec/lmod',
  'SLURM_NPROCS': '2',
  'SLURM_TASK_PID': '8218',
  'SLURM_CPUS_ON_NODE': '2',
  'HISTCONTROL': 'ignoredups',
  'JUPYTERHUB_API_URL': 'http://172.22.100.7:8081/hub/api',
  'ENVIRONMENT': 'BATCH',
  'SLURM_PROCID': '0',
  'SLURM_JOB_NODELIST': 'hpc-class05',
  'SHLVL': '2',
  'HOME': '/home/ynanyam',
  'JUPYTERHUB_USER': 'ynanyam@iastate.edu',
  'SLURM_GET_USER_ENV': '1',
  'SLURM_LOCALID': '0',
  '_ModuleTable002_': 'VEgiXT0iL2V0Yy9tb2R1bGVmaWxlczoVdXNyL3NoYXJlL21vZHVvsZWZpbGVzOi9lc3Ivc2hhcmU
vbW9kdWxlZmlsZXMvTGluZG6L3Vzci9zaGFyZS9tb2R1bGVmaWxlcY9Db3JlOi9lc3Ivc2hhcmUvbG1vZC9sbW9kdWxlZHVvs
ZWZpbGVzL0Nvcml0L3Npbmd1bGFy=',
```

```
'SLURM_CLUSTER_NAME': 'hpc-class',
'SLURM_JOB_CPUS_PER_NODE': '2',
'SLURM_JOB_GID': '101',
'SLURM_SUBMIT_HOST': 'class-jupyterhub',
'SLURM_GTIDS': '0',
'BASH_ENV': '/usr/share/lmod/lmod/init/bash',
'SLURM_JOB_PARTITION': 'medium',
'LOGNAME': 'ynanyam',
'RSTUDIO_PASSWORD': 'test123',
'QTLIB': '/usr/lib64/qt-3.3/lib',
'SLURM_JOB_ACCOUNT': 'ccresearch',
'MODULESHOME': '/usr/share/lmod/lmod',
'JUPYTERHUB_ACTIVITY_URL': 'http://172.22.100.7:8081/hub/api/users/ynanyam@iastate.edu/activity
',
'SLURM_JOB_NUM_NODES': '1',
'LMOD_SETTARG_FULL_SUPPORT': 'no',
'LESSOPEN': '|/usr/bin/lesspipe.sh %s',
'JUPYTERHUB_OAUTH_CALLBACK_URL': '/user/ynanyam@iastate.edu/oauth_callback',
'XDG_RUNTIME_DIR': '/run/user/366067',
'LMOD_DIR': '/usr/share/lmod/lmod/libexec',
'JUPYTERHUB_HOST': '',
'PBS_NODEFILE': '/scratch/ynanyam/101445/0nTTGYADaH',
'JUPYTERHUB_SERVICE_PREFIX': '/user/ynanyam@iastate.edu/',
'BASH_FUNC_module()': '() { eval $(($LMOD_CMD bash "$@") && eval ${${LMOD_SETTARG_CMD:-} -s s
h)\n}',
'BASH_FUNC_ml()': '() { eval $(($LMOD_DIR/ml_cmd "$@")\n}',
'JPY_PARENT_PID': '8226',
'TERM': 'xterm-color',
'CLICOLOR': '1',
'PAGER': 'cat',
'GIT_PAGER': 'cat',
```

```
In [5]: %time sum(range(10000)) # for a single line
```

```
CPU times: user 159 µs, sys: 37 µs, total: 196 µs
Wall time: 200 µs
```

```
Out[5]: 49995000
```

In [2]:

```
%%time  
  
y = sum(range(10000))  
print(y)
```

```
49995000  
CPU times: user 478 µs, sys: 155 µs, total: 633 µs  
Wall time: 493 µs
```

In [ ]:

```
%run /ptmp/jupyterhub-workshop-oct-2021/hello.py
```

In [ ]:

```
%cat /ptmp/jupyterhub-workshop-oct-2021/hello.py
```

## Creating/using Notebook Kernels

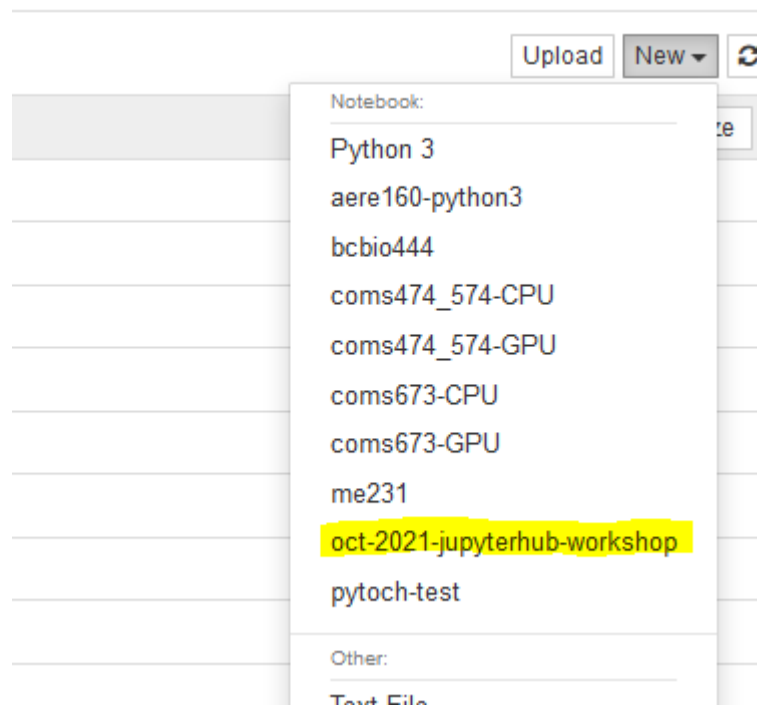
Follow the instructions from the email to create a Python virtual env and then add it to Jupyter Notebook

In [ ]:

```
%cat /ptmp/jupyterhub-workshop-oct-2021/kernel-install-instructions
```



After succesful install, JupyterHub session should have a new kernel *oct-2021-jupyterhub-workshop*



## Python Notebooks and HPC

- Python Notebooks in HPC is considered inefficient
- Batch processing is preferred
- Not guaranteed to get an allocation when the cluster is busy. Timeout of 240 sec.

But there are certain cases where it is useful

- Debugging code prior to batch processing
- For class instructions
- Visualizing data

Note on HOME directory

- By default, python and conda use `$HOME` to save config and packages
- Only 5G of available space so using `/ptmp/$USER` and `$TMPDIR` is suggested to make sure `$HOME` isn't over the limit.
- JupyterHub fails to spawn a session if `$HOME` is full

## Demo

Using Iris flower dataset - [https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set) ([https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set))

Dataset details:

- 50 samples each from three species of Iris - *Iris setosa*, *Iris virginica* and *Iris versicolor*
- Four features measured - length and width of sepals and petals in centimeters

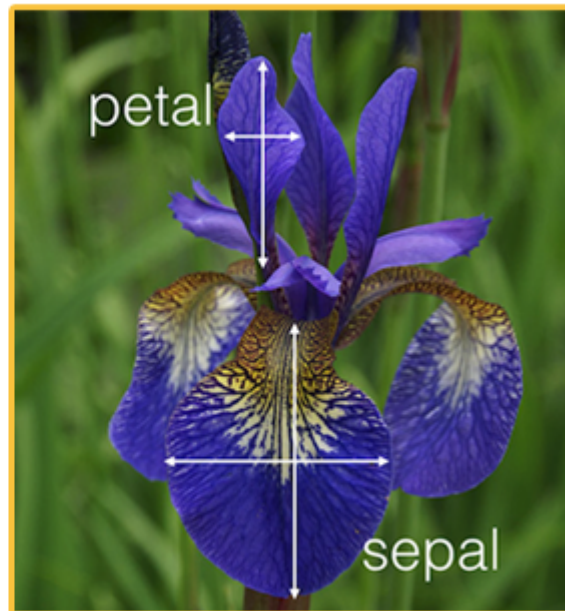
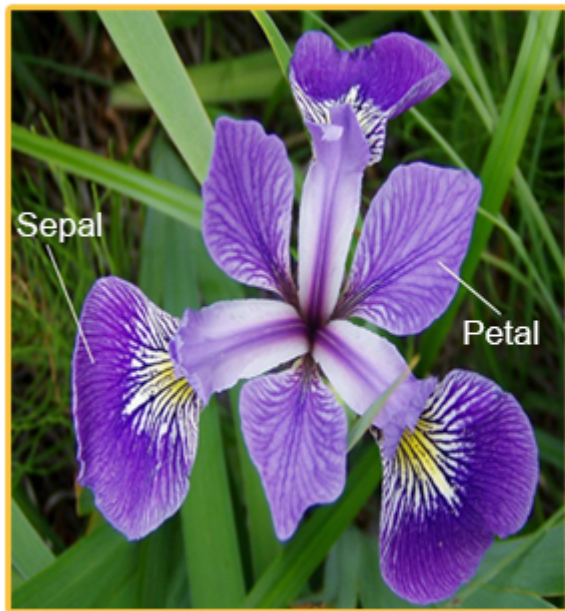


Image Source - <https://www.w3resource.com/machine-learning/scikit-learn/iris/index.php> (<https://www.w3resource.com/machine-learning/scikit-learn/iris/index.php>)

Aim of the demo is to visualize the dataset using seaborn

```
In [ ]: import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

iris_df = pd.read_csv('/ptmp/jupyterhub-workshop-oct-2021/iris.csv')

display(iris_df.head())

display(iris_df.describe())
```

```
In [ ]: sns.set(color_codes=True)

g = sns.pairplot(iris_df)

plt.show()
```

```
In [ ]: sns.pairplot(iris_df, hue='species')
```

## Future Considerations

- Jupyter Lab will be the default soon(Notebook will be available as an option)
- Introduce nbgrader for instructors - <https://github.com/jupyter/nbgrader> (<https://github.com/jupyter/nbgrader>)

```
In [ ]:
```