2021 Oct Jupyter Workshop

Yasasv 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

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

HPC-Class - https://class-jupyterhub.its.iastate.edu
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

<table>
<thead>
<tr>
<th>Partition</th>
<th>Max Cores/Node</th>
<th>Max Walltime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>16</td>
<td>01:00:00</td>
</tr>
<tr>
<td>GPU</td>
<td>16</td>
<td>06:00:00</td>
</tr>
<tr>
<td>Debug</td>
<td>16</td>
<td>01:00:00</td>
</tr>
<tr>
<td>Mem</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

### Condo Cluster

<table>
<thead>
<tr>
<th>Partition</th>
<th>Max Cores/Node</th>
<th>Max Walltime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>16</td>
<td>01:00:00</td>
</tr>
<tr>
<td>GPU</td>
<td>20</td>
<td>06:00:00</td>
</tr>
<tr>
<td>Debug</td>
<td>16</td>
<td>02:00:00</td>
</tr>
<tr>
<td>Mem</td>
<td>32</td>
<td>7-00:00:00</td>
</tr>
</tbody>
</table>

UI Walkthrough
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.

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

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 %history %killbgscripts %ldir %less %lf %lk %ll %load %load_ext %loadp
%y %logoff %logon %logstart %logstop %ls %lsmagic %lx %macro %magic %man %matplotlib %mkdir %more %mv %notebook %page %pastebin %pdb %pdef %pdoc %pfile %pinfo %p
info2 %pip %pop %popd %pprint %precision %prun %psearch %psource %pushd %pwd %pycat %pylab
%qtconsole %quickref %recall %rehashx %reload_ext %rep %rerun %reset %reset_selective %rm %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
DEPRECATED: 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)
cycler (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)
wastest (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`
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
Creating/using Notebook Kernels

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

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


Aim of the demo is to visualize the dataset using seaborn
In [ ]:
```python
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 [ ]:
```python
sns.set(color_codes=True)
g = sns.pairplot(iris_df)

plt.show()
```

In [ ]:
```python
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)