New York University
High Performance Computing

Shenglong Wang

hpc@nyu.edu

Topics

- Prince cluster login
- Python virtual environments setup
- Jupyter notebook on HPC cluster
- Demo

Useful links

- NYU HPC wiki
  - https://wikis.nyu.edu/display/NYUHPC/High+Performance+Computing+at+NYU
  - http://bit.ly/nyuithpc
- Prince cluster and SLURM tutorial
  - https://wikis.nyu.edu/display/NYUHPC/Slurm+Tutorial
  - http://bit.ly/nyuithpcslurm
- Python virtual environments
  - https://media.readthedocs.org/pdf/virtualenv/latest/virtualenv.pdf
- TensorFlow
  - https://www.tensorflow.org/install/
  - https://www.tensorflow.org/install/gpu#software_requirements
- PyTorch
  - https://pytorch.org
- NYU VPN
  - https://www.nyu.edu/life/information-technology/getting-started/network-and-connectivity/vpn.html

Prince cluster logins

- hostname: `prince.hpc.nyu.edu`
- open to NYU subnet
- off campus
  - setup ssh tunneling through HPC gateways
  - use NYU VPN
- Command
  - from Mac/Linux

  ```
  ssh sw77@prince.hpc.nyu.edu -Y
  ```
  - from Windows: putty

Python virtual environments setup

- to setup virtual enviorment
  - `module purge`
  - `module load python3/intel/3.6.3`
  - `mkdir ~/pyenv`
  - `cd pyenv`
  - `virtualenv --system-site-packages py3.6.3`
  - `source ~/pyenv/py3.6.3/bin/activate`
- To install Jupyter
  - `pip install jupyter`
- to install PyTorch
  - `pip3 install http://download.pytorch.org/whl/cu92/torch-0.4.1-cp36-cp36m-linux_x86_64.whl`
  - `pip3 install torchvision`
- to install TensorFlow
  - CPU version: `pip install tensorflow`
  - GPU version: `pip install tensorflow-gpu`

TensorFlow GPU version has to work with module `cudnn/9.0v7.3.0.29`

Simple interactive test

```
srun --cpus-per-task=1 --gres=gpu:1 --mem=10GB --x11 --pty /bin/bash
```

[sw77@gpu-31 ~]$ module load python3/intel/3.6.3 cudnn/9.0v7.3.0.29

[sw77@gpu-31 ~]$ source ~/pyenv/py3.6.3/bin/activate

(py3.6.3) [sw77@gpu-31 ~]$ python -c "import torch; print(torch.__version__); print(torch.__file__)"

0.4.1

/home/sw77/pyenv/py3.6.3/lib/python3.6/site-packages/torch/__init__.py

(py3.6.3) [sw77@gpu-31 ~]$ python -c "import tensorflow as tf; print(tf.__version__); print(tf.__file__)"

1.10.1

/home/sw77/pyenv/py3.6.3/lib/python3.6/site-packages/tensorflow/__init__.py

(py3.6.3) [sw77@gpu-31 ~]$ python -c "import tensorflow as tf; tf.Session()"

2018-09-27 13:44:55.268068: I tensorflow/core/platform/cpu_feature_guard.cc:141] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 FMA

2018-09-27 13:44:55.505032: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1405] Found device 0 with properties:

name: Tesla P40 major: 6 minor: 1 memoryClockRate(GHz): 1.531

pciBusID: 0000:84:00.0

totalMemory: 22.38GiB freeMemory: 22.22GiB

2018-09-27 13:44:55.505246: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1484] Adding visible gpu devices: 0

2018-09-27 13:44:56.751112: I tensorflow/core/common_runtime/gpu/gpu_device.cc:965] Device interconnect StreamExecutor with strength 1 edge matrix:

2018-09-27 13:44:56.751240: I tensorflow/core/common_runtime/gpu/gpu_device.cc:971]   0

2018-09-27 13:44:56.751278: I tensorflow/core/common_runtime/gpu/gpu_device.cc:984] 0:   N

2018-09-27 13:44:56.751805: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1097] Created TensorFlow device (/job:localhost/replica:0/task:0/device:GPU:0 with 21551 MB memory) -> physical GPU (device: 0, name: Tesla P40, pci bus id: 0000:84:00.0, compute capability: 6.1)

(py3.6.3) [sw77@gpu-31 ~]$

Jupyter notebook setup

- implement a SBATCH script, submit job to queue to run Jupyter server on a GPU node
- after job gets run, check SLURM log file for setup instructions
- setup ssh tunneling from local computer to prince cluster

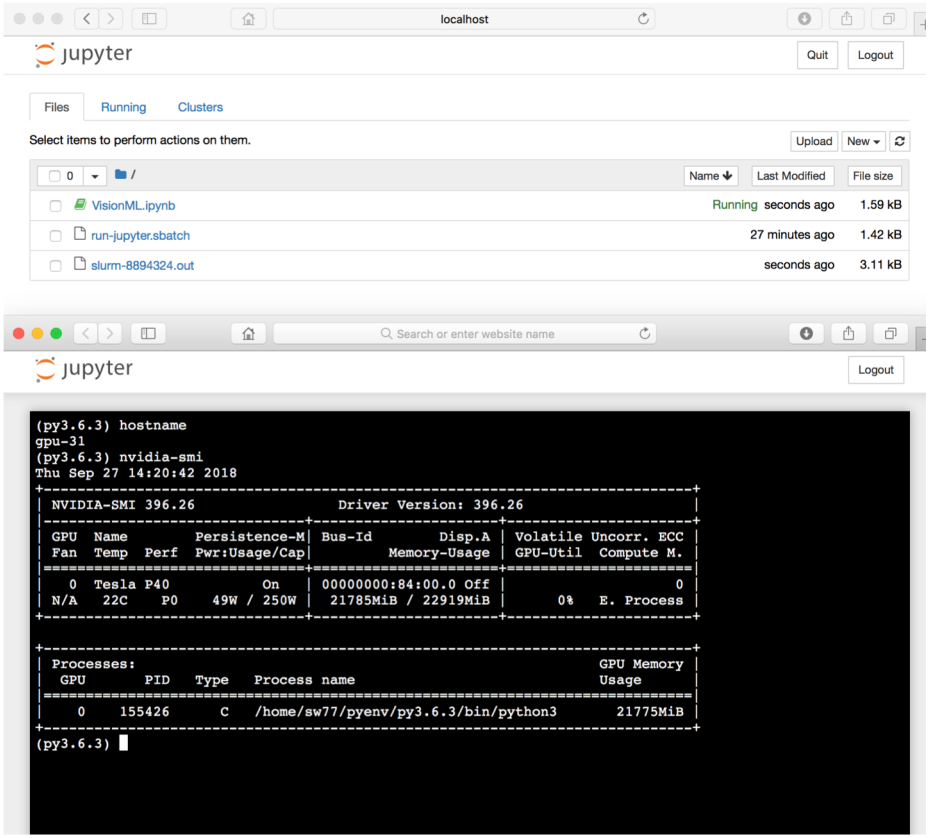
```
ssh -L 8468:localhost:8468 sw77@prince.hpc.nyu.edu
```

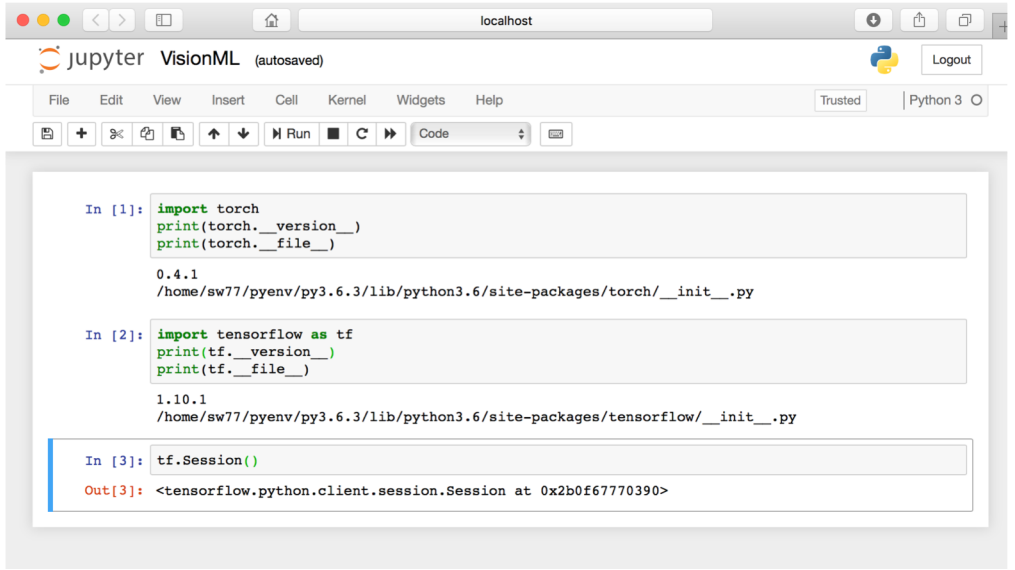
- open web browser to connect to Jupyter server with url

```
http://localhost:8468/?token=aaf8434775ac4a3429533eca2b01933961d2da82b8778a43
```

- examples are in the folder

```
/share/apps/examples/vision-ml
```


localhost
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
Name
Last Modified
File size
VisionML.ipynb
Running seconds ago
1.59 kB
run-jupyter.sbatch
27 minutes ago
1.42 kB
slurm-8894324.out
seconds ago
3.11 kB
Search or enter website name
jupyter
Logout
(py3.6.3) hostname
gpu-31
(py3.6.3) nvidia-smi
Thu Sep 27 14:20:42 2018
NVIDIA-SMI 396.26
Driver Version: 396.26
GPU Name Persistence-M Bus-Id Disp.A Volatile Uncorr. ECC
Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M.
0 Tesla P40 On 00000000:84:00.0 Off 0
N/A 22C P0 49W / 250W 21785MiB / 22919MiB 0% E. Process
Processes:
GPU PID Type Process name GPU Memory Usage
0 155426 C /home/sw77/pyenv/py3.6.3/bin/python3 21775MiB
(py3.6.3)


localhost
jupyter
VisionML
(autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code
In [1]:
import torch
print(torch.__version__)
print(torch.__file__)
0.4.1
/home/sw77/pyenv/py3.6.3/lib/python3.6/site-packages/torch/__init__.py
In [2]:
import tensorflow as tf
print(tf.__version__)
print(tf.__file__)
1.10.1
/home/sw77/pyenv/py3.6.3/lib/python3.6/site-packages/tensorflow/__init__.py
In [3]:
tf.Session()
Out[3]:
<tensorflow.python.client.session.Session at 0x2b0f67770390>

Storage allocations and shared work space

- Different file systems for different purposes
- `/home`: source code, scripts, executables, libraries, …
- `/scratch`: job running, big input/output files
- `/beegfs`: jobs running, small IOs, good for ML projects
- `/archive`: long term storage
- `/state/partition1`: local on the compute nodes for temporary files

Mount point	Quota	Backup?	Availability	Variables	Value	
`/home`	20GB	Yes	all nodes	`$HOME`	`/home/$USER`	bBackup with 30-day snapshots
`/beegfs`	2TB/3M inodes	No	all nodes	`$BEEGFS`	`/beegfs/$USER`	60 days cleanup policy
`/scratch`	5TB/1M inodes	No	all nodes	`$SCRATCH`	`/scratch/$USER`	60 days cleanup policy
`/archive`	2TB	Yes	login nodes	`$ARCHIVE`		bBackup with 30-day snapshots
`/state/partition1`	~100GB	No	compute nodes	`$SLURM_JOBTMP`		cleanup after job finishes