

Install and Configure Ipython Notebook with Apache Spark in HDP

Required Components

- HDP 2.5 (containing Apache Spark) on Hortonworks sandbox
 - Go to <http://hortonworks.com/downloads/> and download for Virtualbox
 - Install HDP in Virtualbox
- Python 2.7X

Configure the Dependencies

At Centos prompt, **type** the following commands:

- `yum install nano centos-release-SCL zlib-devel`
- `yum install bzip2-devel openssl-devel ncurses-devel`
- `yum install sqlite-devel readline-devel tk-devel`
- `yum install gdbm-devel db4-devel libpcap-devel xz-devel`
- `yum install libpng-devel libjpeg-devel atlas-devel`

Install the “Development tools” Dependency for Python 2.7

lpython requires Python 2.7 or higher (the version came with the HDP installation is Python 2.6).

Execute:

```
yum groupinstall "Development tools"
```

Install Python 2.7

- Run the following command to get Python 2.7:
yum install python27
(if it does not work, first run **yum install centos-release-scl**)
- Get the Python Anaconda distribution to upgrade Python to current version and to obtain the most useful Python packages for scientific computation.
 - `yum install wget`
 - `wget https://repo.continuum.io/archive/Anaconda2-4.3.1-Linux-x86_64.sh`
 - `bash Anaconda2-4.3.1-Linux-x86_64.sh`
(accept the license and type 'yes')

Install or Upgrade Python Analytics Packages

By installing the latest version of Anaconda, you should have all the following Python packages for data science: numpy, scipy, matplotlib, pandas, scikit-learn, tornado, pyzmq, pygments, jsonschema and jinja2. Use pip to check installation or to upgrade them if already installed:

```
pip install numpy scipy pandas matplotlib scikit-learn
```

```
pip install tornado pyzmq pygments jsonschema
```

```
pip install jinja2 --upgrade
```

```
pip install matplotlib --upgrade
```

```
pip install scikit-learn --upgrade
```

```
pip install numpy --upgrade
```

```
...
```

Install and Configure IPython Notebook

- Run the following command to install ipython notebook if not already installed:
`pip install "ipython[notebook]"`
- Configure a Python interpreter built with Apache Spark by creating an Ipython profile for pyspark:
`ipython profile create pyspark`
- Generate a Jupyter configuration file:
`jupyter notebook --generate-config`

(You should see the following output:

Writing default config to:

`/root/.jupyter/jupyter_notebook_config.py`)

Create a Shell Script to Start IPython

In a text editor (vi), in the home directory (~) create a shell script called start_ipython_notebook.sh.

```
#!/bin/bash  
source /opt/rh/python27/enable  
IPYTHON_OPTS="notebook --port 8889  
--notebook-dir='/usr/hdp/2.5.0.0-1245/spark/'  
--ip='*' --no-browser" pyspark
```

Make the above shell script executable:

```
chmod +x start_ipython_notebook.sh
```


Port Forwarding

In order for IPython notebook to be accessible from a browser on your host machine, if not already completed, you must forward the port 8889 from the virtual machine to the host machine:

- Open the VirtualBox
- Click the **Setting** button
- Select the **Network** tab

Port Forwarding (cont'd)

Hortonworks Sandbox with HDP 2.4 - Network

General System Display Storage Audio **Network** Ports Shared Folders User Interface

Adapter 1 Adapter 2 Adapter 3 Adapter 4

☒ Enable Network Adapter

Attached to: NAT

Name:

▼ Advanced

Adapter Type: PCnet-FAST III (Am79C973)

Promiscuous Mode: Deny

MAC Address: 080027383F58


☒ Cable Connected


Click → **Port Forwarding**


? Cancel OK


Port Forwarding (cont'd)


Hortonworks Sandbox with HDP 2.4 - Network


General


System


Display


Storage

Audio


Network

Ports

Shared Folders

User Interface

Click



Name	Protocol	Host IP	Host Port	Guest IP	Guest Port
Accumulo	TCP	127.0.0.1	50095		50095
AmbariShell	TCP	127.0.0.1	4200		4200
Atlas	TCP	127.0.0.1	21000		21000
Datanode	TCP	127.0.0.1	50075		50075
Falcon	TCP	127.0.0.1	15000		15000
HBaseMaster	TCP	127.0.0.1	16010		16010
HBaseRegion	TCP	127.0.0.1	16030		16030
HS2	TCP	127.0.0.1	10000		10000
HS2Http	TCP	127.0.0.1	10001		10001
HST	TCP	127.0.0.1	9000		9000
JobHistory	TCP	127.0.0.1	19888		19888

Port Forwarding

Cancel

OK

Port Forwarding (cont'd)

Hortonworks Sandbox with HDP 2.4 - Network

General System Display Storage Audio Network Ports Shared Folders User Interface

Name	Protocol	Host IP	Host Port	Guest IP	Guest Port
WebHcat	TCP	127.0.0.1	50111		50111
WebHdfs	TCP	127.0.0.1	50070		50070
XASecure	TCP	127.0.0.1	6080		6080
YarnATS	TCP	127.0.0.1	8188		8188
YarnRM	TCP	127.0.0.1	8088		8088
Zeppelin1	TCP	127.0.0.1	9995		9995
Zeppelin2	TCP	127.0.0.1	9996		9996
ambari	TCP	127.0.0.1	8080		8080
apache	TCP	127.0.0.1	42080		80
hdfs	TCP	127.0.0.1	8020		8020
ipython	TCP	127.0.0.1	8889		8889

Port Forwarding

Cancel OK

Add a new port (ipython) as indicated and then click OK

Running IPython Notebook

- Execute the shell script from CentOS (home director ~) :

[./start_ipython_notebook.sh](#)

Copy the URL created by the script or record the token for login in next step.

```
[root@sandbox ~]# ./start_ipython_notebook.sh
Multiple versions of Spark are installed but SPARK_MAJOR_VERSION
Spark1 will be picked by default
[TerminalIPythonApp] WARNING | Subcommand `ipython notebook` is d
[TerminalIPythonApp] WARNING | You likely want to use `jupyter no
[W 06:01:57.086 NotebookApp] WARNING: The notebook server is list
recommended.
[I 06:01:57.091 NotebookApp] Serving notebooks from local directo
[I 06:01:57.091 NotebookApp] 0 active kernels
[I 06:01:57.091 NotebookApp] The Jupyter Notebook is running at:
3dfb6bce269232429c857bd64373e133214de
[I 06:01:57.091 NotebookApp] Use Control-C to stop this server an
[C 06:01:57.092 NotebookApp]
```

Copy/paste this URL into your browser when you connect for th
to login with a token:

<http://localhost:8889/?token=b96c93385a63dfb6bce269232429>

Running IPython Notebook

- Open a browser on your host machine, paste in the URL from last step to enter, or enter <http://127.0.0.1:8889> to the URL window and you should see a login window. Enter the token to enter.

127.0.0.1:8889/login?next=%2Ftree%3F

jupyter

Password or token: Log in

Invalid password

Token authentication is enabled. You need to open the notebook server with its first-time login token in the URL, or enable a password in order to gain access. The command:

```
jupyter notebook list
```

will show you the URLs of running servers with their tokens, which you can copy and paste into your browser. For example:

```
Currently running servers:  
http://localhost:8888/?token=c8de56fa... :: /Users/you/notebooks
```


Or you can paste just the token value into the password field on this page.

Cookies are required for authenticated access to notebooks.

127.0.0.1:8889/login?next=%2Ftree%3F

Search

Most VisitedGetting StartedDicLong Beach, Calif...LinuxShellA Basic Introducti...AmbariYT



Password or token:

.....

Log in

Invalid password

Token authentication is enabled. You need to open the notebook server with its first-time login token in the URL, or enable a password in order to gain access. The command:

```
jupyter notebook list
```

will show you the URLs of running servers with their tokens, which you can copy and paste into your browser. For example:

```
Currently running servers:
http://localhost:8888/?token=c8de56fa... :: /Users/you/notebooks
```

Or you can paste just the token value into the password field on this page.

Cookies are required for authenticated access to notebooks.

Jupyter (IPython) Notebook

