

## [One time setup] Install IPython

From the top left menu, Open a terminal: Applications => System Tools => Terminal

Type:

```
sudo easy_install ipython==1.2.1
```

Hit enter, administrator password is **cloudera**.

## Launch pyspark with IPython

Every time you need to open the pyspark shell, open a terminal and type:

```
PYSPARK_DRIVER_PYTHON=ipython pyspark
```

Hit enter, after the startup logs, you should see the pyspark console:

A screenshot of a terminal window showing the PySpark console output. The text is as follows:

```
Welcome to
      _ _ _ _ _
     / /   \ \
    / /     \ \
   / /       \ \
  / /         \ \
 / /           \ \
/_/            \_\
version 1.3.0

Using Python version 2.6.6 (r266:84292, Feb 22 2013 00:00:18)
SparkContext available as sc, HiveContext available as sqlCtx.
In [1]:
```

## Check version

To make sure that PySpark started correctly, print out the version by typing in the PySpark IPython terminal:

```
sc.version
```

Verify that the output is:

```
u'1.3.0'
```

## Python resources

This module requires a basic knowledge of Python to perform the programming assignments.

I recommend to take the interactive tutorial at: <http://learnpython.org>, the "Learn the basics" section should be enough.

Then if you want to learn more about Python, even if not necessary for this class, you can check the Coursera class:

<https://www.coursera.org/learn/python>