

Ubuntu Spark installation tutorial

by Meng Cao

1. Install Java JDK

Two Options – Open Java JDK or Oracle Java JDK

Open JDK

```
>> sudo apt-get update
>> sudo apt-get install openjdk-7-jre
>> sudo apt-get install openjdk-7-jdk
```

Oracle JDK

```
>> sudo apt-get install python-software-properties
>> sudo add-apt-repository ppa:webupd8team/java
>> sudo apt-get update
```

2. Edit Environment

```
>>sudo nano ~/.bashrc
```

add following to the end

```
export SPARK_HOME=<your-spark-path
export PYTHONPATH=$SPARK_HOME/python/:$PYTHONPATH
```

notice: need to restart terminal to make it work.

3. Install python packages

if you dont have pip, install pip first

```
>> sudo apt-get install python-pip python-dev build-essential
```

then install py4j

```
>> sudo pip install py4j
```

install numpy

```
>> sudo pip install numpy
```

4. Disable spark console message

<http://stackoverflow.com/questions/25193488/how-to-turn-off-info-logging-in-pyspark>