

Installation (Linux or OS X)

I highly recommend installing Brew if you do not have it in your machine, it is an easy install utility similar to apt-get on Linux that will make installations very easy.

1) In a new terminal window install Brew:

```
/usr/bin/ruby -e "$(curl -fsSL  
https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

2) Once installed install “wget” (Will take 15-30 min to download, 250 mb)

```
brew install wget
```

3) Change directory to a location you would like to save Spark and then execute:

```
wget http://d3kbcqa49mib13.cloudfront.net/spark-2.0.1-bin-hadoop2.7.tgz
```

4) Unpackage Spark using the tar command

```
tar -zxvf spark-2.0.1-bin-hadoop2.7.tgz
```

5) Launch either pySpark or sparkR by navigating to ./bin folder and executing:

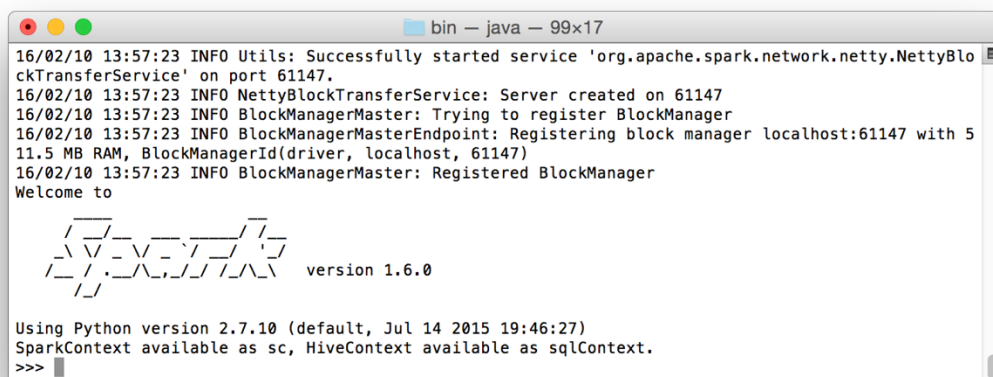
```
cd ./spark-2.0.1-bin-hadoop2.7/bin
```

pySpark - ./pyspark

sparkR - ./sparkR

You should see the following console open:

pySpark Console:



```
bin — java — 99x17
16/02/10 13:57:23 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 61147.
16/02/10 13:57:23 INFO NettyBlockTransferService: Server created on 61147
16/02/10 13:57:23 INFO BlockManagerMaster: Trying to register BlockManager
16/02/10 13:57:23 INFO BlockManagerMasterEndpoint: Registering block manager localhost:61147 with 5
11.5 MB RAM, BlockManagerId(driver, localhost, 61147)
16/02/10 13:57:23 INFO BlockManagerMaster: Registered BlockManager
Welcome to

  ____      _
 / ___|    / \
| |   |  _/  \
| |___| / ___\
|_____|/_   \_\
version 1.6.0

Using Python version 2.7.10 (default, Jul 14 2015 19:46:27)
SparkContext available as sc, HiveContext available as sqlContext.
>>>
```

```
bin — R — 99x17
```

```
ckTransferService' on port 61161.  
16/02/10 13:58:30 INFO NettyBlockTransferService: Server created on 61161  
16/02/10 13:58:30 INFO BlockManagerMaster: Trying to register BlockManager  
16/02/10 13:58:30 INFO BlockManagerMasterEndpoint: Registering block manager localhost:61161 with 5  
11.5 MB RAM, BlockManagerId(driver, localhost, 61161)  
16/02/10 13:58:30 INFO BlockManagerMaster: Registered BlockManager  
  
Welcome to  
  
    / \_/_/_/_/_/_/_/_/_/_/_/_/_/_/_\  
   / \_/_/_/_/_/_/_\_/_/_/_\_/_\_/_\_\  
  / \_/_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/  
 / \_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/_/_/  
/ \_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/_\_/_/_/  
 version 1.6.0  
  
Spark context is available as sc, SQL context is available as sqlContext  
>
```

Ex: Install pandas from terminal (very useful for manipulating data structures) and then call the package before use in pySpark.

```
import pandas
```

```
library(RCurl)
```

FYI: You can easily install things on Linux use **apt-get** or **yum** (depends on version of linux). You can similarly easily install python packages using **easy_install** or **pip**.