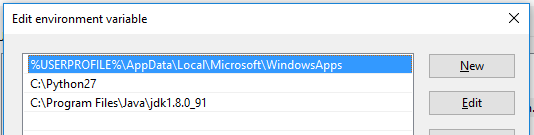
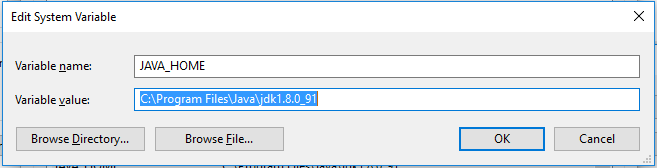
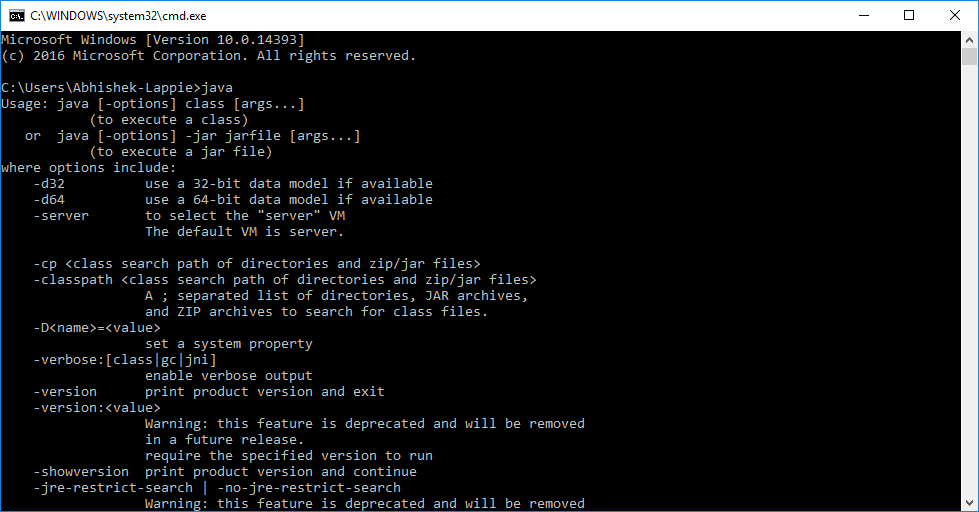
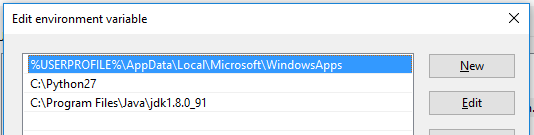
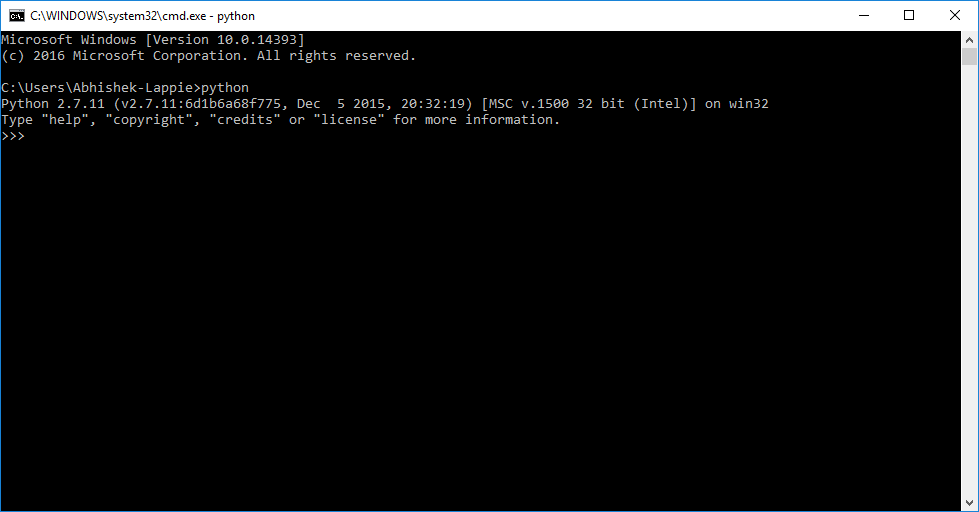
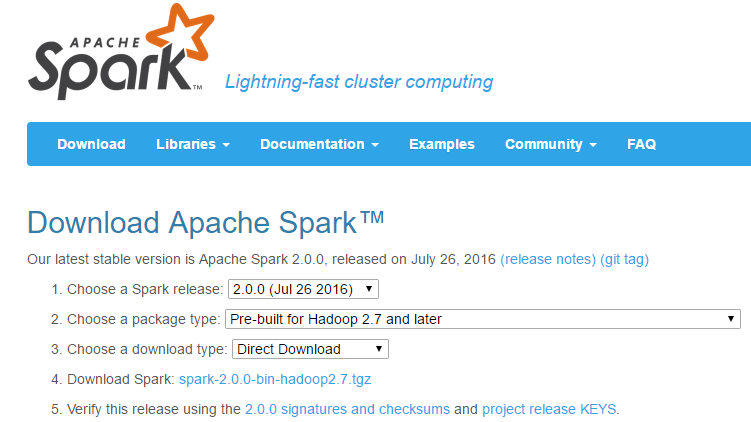
Install Java

1. Java: install Java (<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>)
2. Set Environment Variable for java
3. Set System Variable JAVA\_HOME
4. Go to command prompt and write ‘java’ to verify java is functional

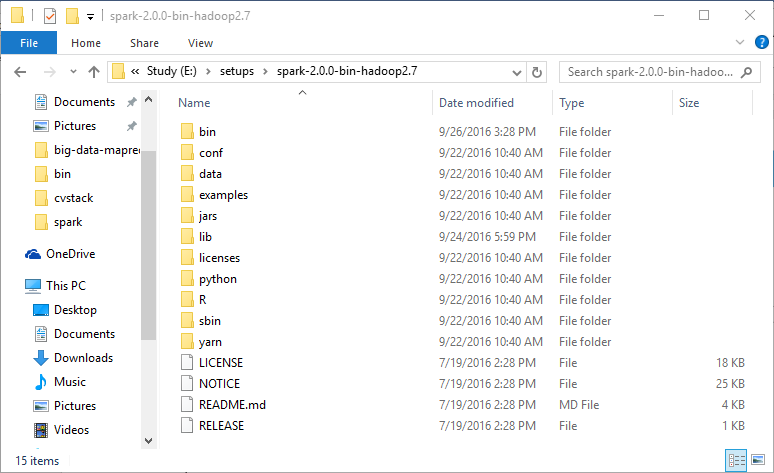
Install Python

1. Install Python(<https://www.python.org/download/releases/2.7/>)
2. Set environment variable for Python
3. Go to command prompt and write ‘python’ to verify python is functional 

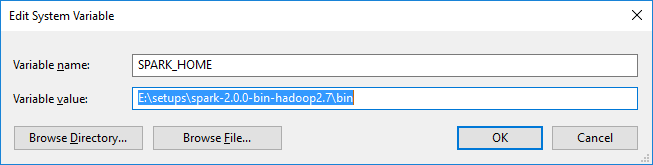
Install Spark

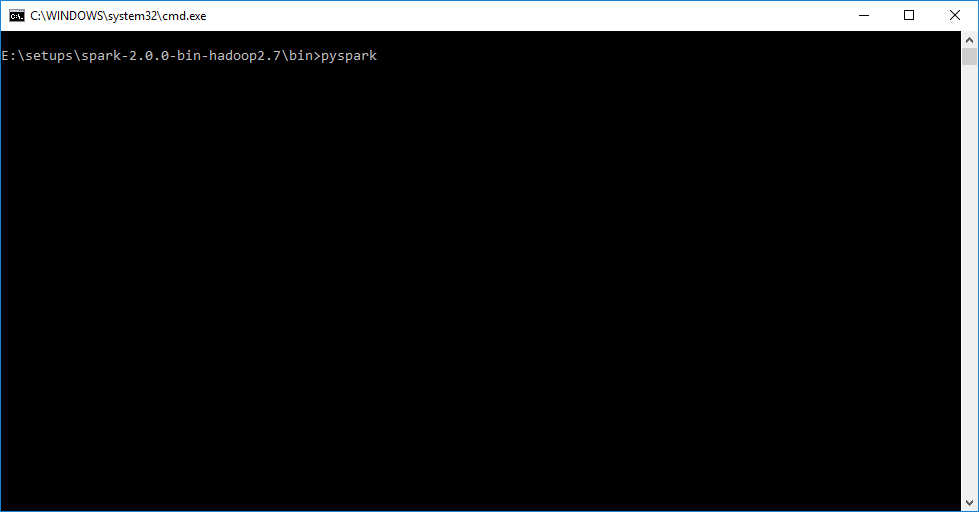
1. Download spark(http://spark.apache.org/downloads.html) 
2. Extract the contents of the *spark-2.0.0-bin-hadoop2.7.tgz* using Winrar or 7-Zip.

**IMPORTANT**: while extracting, make sure none of the folders in the path of the directory should have two words name i.e., no white spaces in the name. e.g. avoid names like ‘Program Files’. Have folder names like ‘*Program\_Files*’ or ‘*ProgramFiles*’.

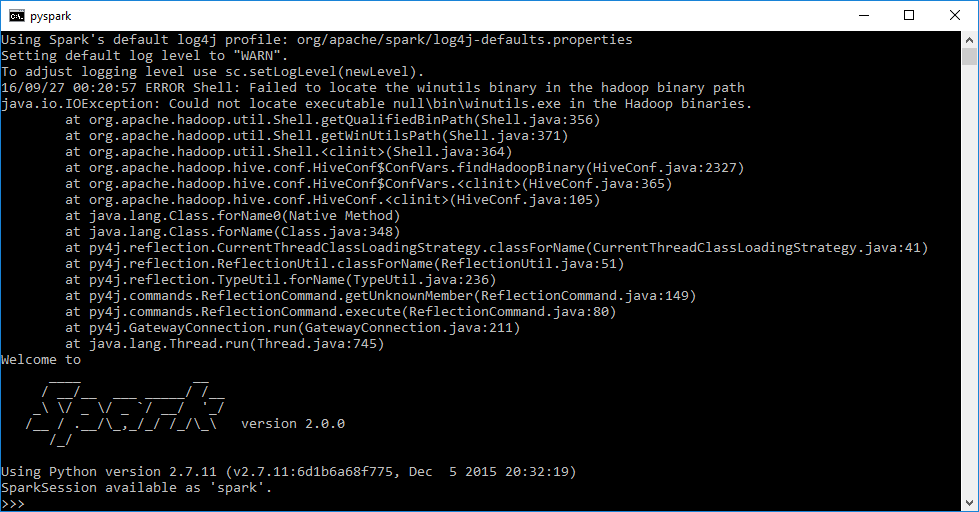
1. 

**Note**: if you find a .tar file inside the .gzip, extract the .tar file to extract the contents as shown above.

1. Set the environment variable ‘SPARK\_HOME’ for spark. 
2. Open the command prompt and go to the bin directory of the directory where you have extracted Spark. Execute the file ‘pyspark’.



You may get exceptions on the screen saying “*Failed to locate the winutils binary in the hadoop binary path java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.*” But that is ok because we are not using Hadoop.



*This document is prepared by:*

|  |
| --- |
| **Abhishek Choudhury (**achoudhury@scu.edu) |