

Step 1: Spark latest version can be downloaded from the below site,

**File name: spark-2.2.0-bin-hadoop2.7 (Installed version)**

<http://spark.apache.org/downloads.html> You may select the last build. As of today, it is 2.2.0 with prebuild Hadoop 2.7 Just click the download link to get the package

Pre-requisite:

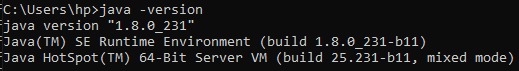
Prior installation of Spark, it is recommended that you have installed one of the java versions

Java can be downloaded from

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

### Setting Permanent Path

* Go to My Computer ---> Right Click on it ---> Advanced System Settings ---> Advanced Tab ---> Click on Environment Variables
* Click on New tab of User variables, assign value JAVA\_HOME to Variable Name
* java\jdk\_version\bin path (copied path) to Variable Value and click on OK Button
* Finally, click on OK button.



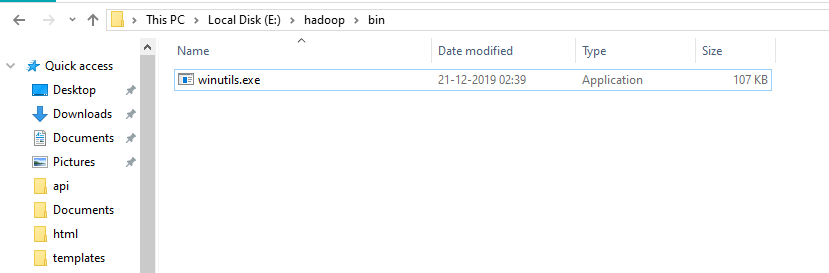
Step 2: Unzip and extract your download into a local folder.

Due to the installation is packaged by gzip then tar. So, you will need to unpack it by any zip tools to get a "spark-2.2.0-bin-hadoop2.7". Rename it to spark-2.2.0-bin-hadoop2.7.zip and unzip it again to get all material.

Step 3: Download / Install Hadoop 2.7.1 binary for windows 10

Unzip the winutils.exe and copy to local folder E:\hadoop\bin (It may be C|D|E drive of your choice)





Step 4: Configure your environment.

From windows logo=>search to launch: "Search advanced system settings" program-> click the button of "Environment Variables”, Change below environment variables according to your software version and location. This is an example.

JAVA\_HOME=C:\Program Files\Java\jre1.8.0\_121

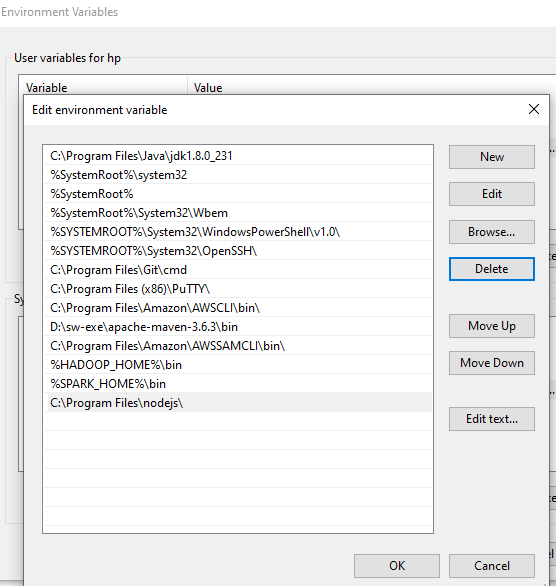
SPARK\_HOME=C:\somewhere\spark-2.2.0-bin-hadoop2.7

HADOOP\_HOME=C:\somewhere\hadoop-2.7.3

Append below variable into "**Path**"

%SPARK\_HOME%\bin

Screenshot of variables assigned in Path,



Step 5. Grant permission to temp folder

Create a temp folder under c:\tmp\hive (C/D/E of your choice)

Execute below command as administrator.

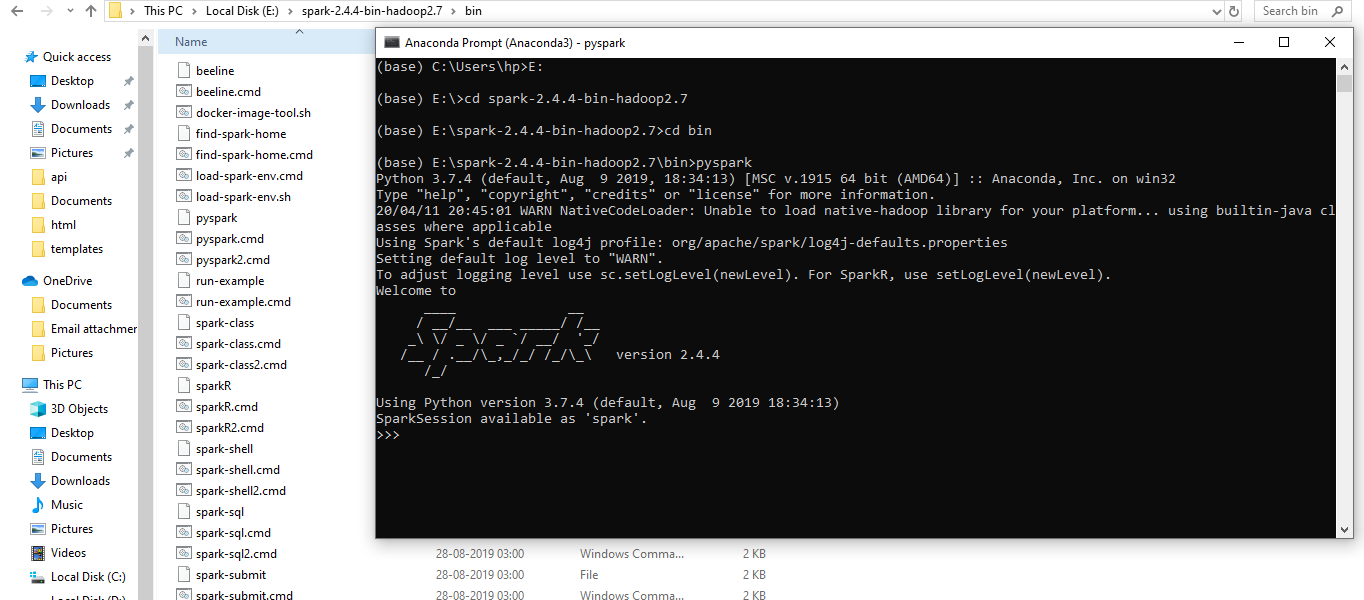
winutils.exe chmod 777 C:\tmp\hive

Step 6:

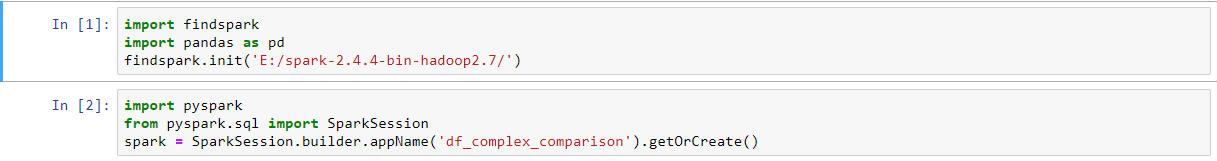
Go to the c:\somewhere\spark-2.2.0-bin-hadoop2.7\bin\

execute "spark-shell", "pyspark" or "spark-submit <app\_name>" for your program.

Hope everything good. If you found some error messages related to HiveSessionState, you may try to execute command as administrator to avoid it.



To Run from Jupyter we need to run the below command



Import findspark

findspark. init (“Spark installed directory”)