How to run Apache Spark on Windows7 in standalone mode

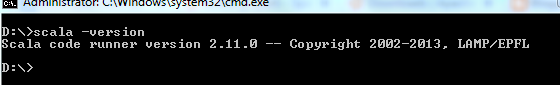
So far, we might have done setup of Spark with Hadoop, EC2 or mesos on Linux machine.  But what if we don’t want with Hadoop/EC2, we just want to run it in standalone mode on windows.

Here we’ll see how we can run Spark on Windows machine.

**Prerequisites:**

* Java6+
* Scala 2.10.x
* Python 2.6 +
* Spark 1.2.x
* sbt ( In case of building Spark Source code)
* GIT( If you use sbt tool)

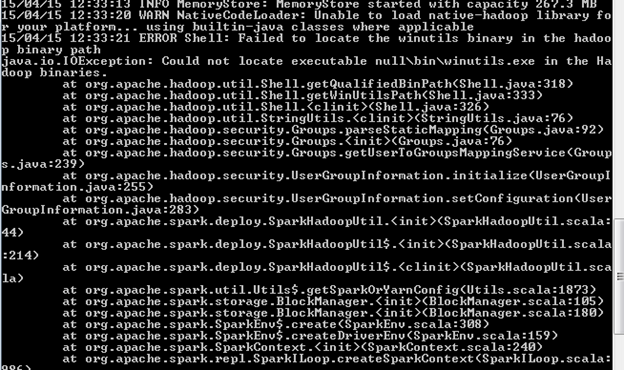
Now we’ll see the installation steps :

* Install Java 7 or later. Set JAVA\_HOME and PATH variable as environment variables.
* Download [Scala 2.10](http://www.scala-lang.org/download/2.10.5.html) or Scala 2.11 and install. Set **SCALA\_HOME**  and add **%SCALA\_HOME%\bin** in   PATH variable in environment variables. To test whether Scala is installed or not, run following command. [](http://4.bp.blogspot.com/-cgSmScJ771A/VS6VzSdLp8I/AAAAAAAAEQ0/ipkZWy-CZeo/s1600/scalaversion.png)
* Next thing is Spark. Spark can be installed in two ways.
  + Building Spark using SBT
  + Use Prebuilt Spark package

**Building Spark with SBT :**

* Download [SBT](http://www.scala-sbt.org/download.html) and install. Set SBT\_HOME and PATH variable in environment variables.
* Download [source code](http://spark.apache.org/downloads.html) from Spark website against any of the Hadoop version.
* Run **sbt assembly**command to build the [Spark package](http://spark.apache.org/docs/1.2.1/building-spark.html#building-with-sbt)
* You need to set Hadoop version also while building as follows :
* **sbt –Pyarn –pHadoop 2.3 assembly**

**Using Spark Prebuilt Package:**

* Choose a [Spark](http://spark.apache.org/downloads.html) prebuilt package for Hadoop i.e.  **Prebuilt for Hadoop 2.3/2.4 or later**. Download and extract it to any drive i.e. D:\spark-1.2.1-bin-hadoop2.3
* Set **SPARK\_HOME** and add **%SPARK\_HOME%\bin**in PATH in environment variables
* Run following command on command line.                              [https://1.bp.blogspot.com/-qMav-Ju9l6E/VS6Vzlyp7-I/AAAAAAAAEQ4/NK_d5kCvpXQ/s1600/spark-shell.png](http://1.bp.blogspot.com/-qMav-Ju9l6E/VS6Vzlyp7-I/AAAAAAAAEQ4/NK_d5kCvpXQ/s1600/spark-shell.png)
* You’ll get and error for winutils.exe:[](http://4.bp.blogspot.com/-NBYnbjvIiPQ/VS6VzMNvlQI/AAAAAAAAEQw/zSce7PMz6dA/s1600/error.png)

    Though we aren’t using Hadoop with Spark, but somewhere it checks for HADOOP\_HOME variable in configuration. So to overcome this error, download [winutils.exe](https://github.com/steveloughran/winutils/blob/master/hadoop-2.6.0/bin/winutils.exe) and place it in any location (i.e. D:\winutils\bin\winutils.exe).  
  
**P.S.** As per the Operating system version, this winutils.exe may vary. So in case, if it doesn't support to your OS, please find another one and use. You can refer this [Problems running Hadoop on Windows](https://wiki.apache.org/hadoop/WindowsProblems) link for winutils.exe.

* Set **HADOOP\_HOME = D:\winutils** in environment variable
* Now, Re run the command “spark-shell’ , you’ll see the scala shell. For latest spark releases, if you get the permission error for /tmp/hive directory as given below:

The root scratch dir: /tmp/hive on HDFS should be writable. Current permissions are: rw-rw-rw-

 You need to run following command :

**D:\spark>D:\winutils\bin\winutils.exe chmod 777 D:\tmp\hive**

* For Spark UI : open <http://localhost:4040/> in browser
* For testing the successful setup you can run the example :[https://2.bp.blogspot.com/-tVd2T9ofDb8/VS6VyNnxFXI/AAAAAAAAEQk/15qYpC5_VI8/s1600/runexample.png](http://2.bp.blogspot.com/-tVd2T9ofDb8/VS6VyNnxFXI/AAAAAAAAEQk/15qYpC5_VI8/s1600/runexample.png)
* It will execute the program and return the result :[https://2.bp.blogspot.com/-ZXQ-UR8GIZo/VS6Vx0tZeQI/AAAAAAAAEQg/p1dTLryP_5U/s1600/result.png](http://2.bp.blogspot.com/-ZXQ-UR8GIZo/VS6Vx0tZeQI/AAAAAAAAEQg/p1dTLryP_5U/s1600/result.png)

Now your cluster is successfully launched, start writing your Java/Python/Scala programs…!!!! 

