Robert Jiga

4 October 2020

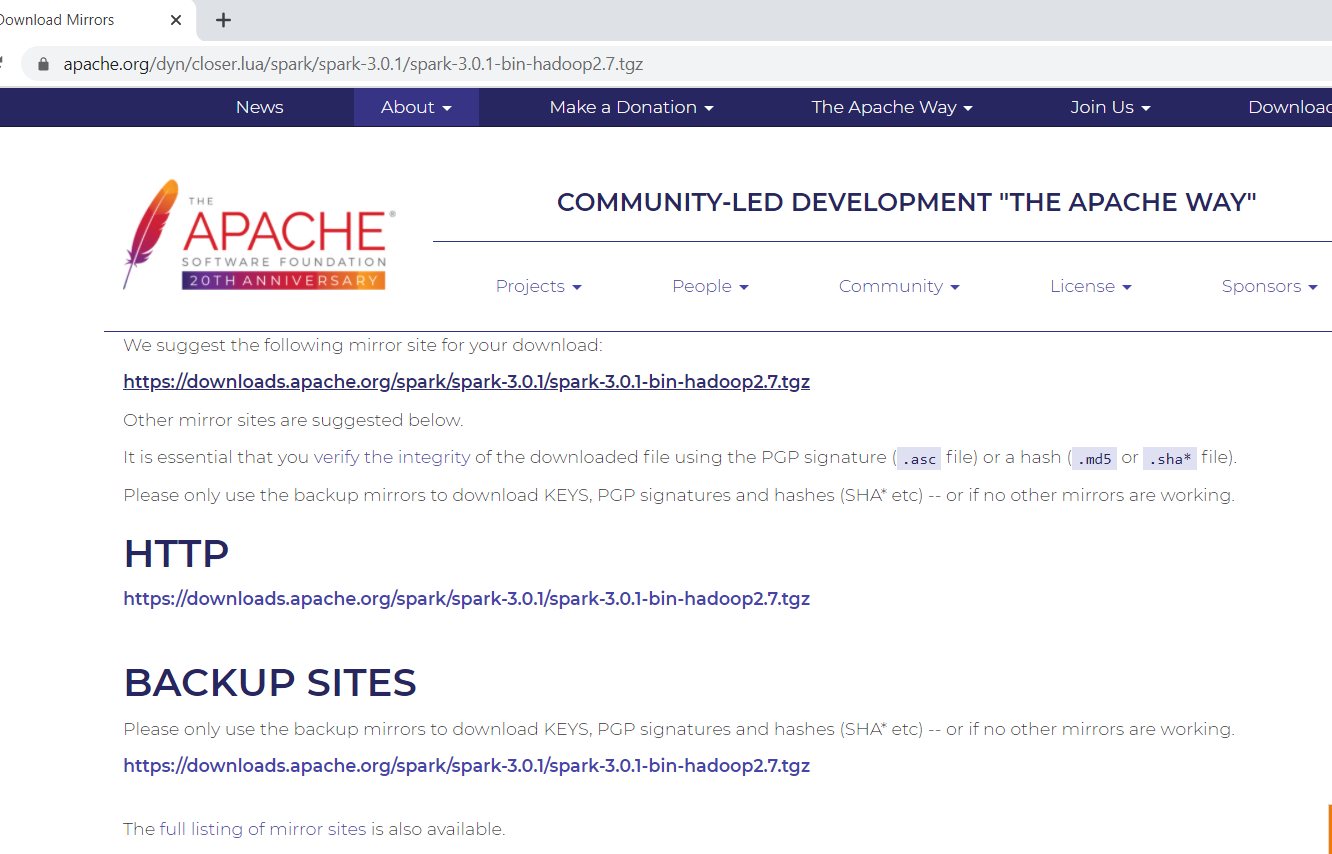
Cloud Computing

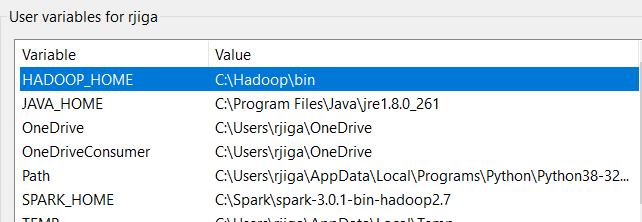
HW 1

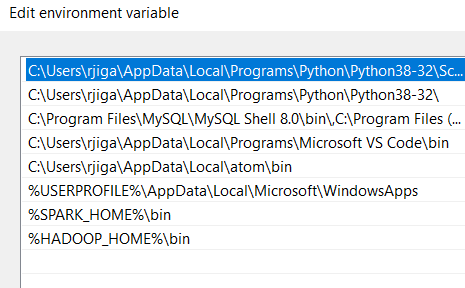
1. Make sure your computer has the following prerequisites:
   1. Java
   2. Python
   3. Memory: 8GB+
   4. CPU: 2 cores+
   5. Ubuntu/CentOS
   6. WinZip (or any decompression software to unzip .tgz files)
2. Download Apache Spark at the following link: <https://spark.apache.org/downloads.html>
   1. Ensure the following criteria below is met and click the “Download Spark” link shown below at step 3



* 1. The link will take you to the following page shown below. Click on any of the links to download the package (in my case I used the first link)

1. 
2. Download winutils.exe file at <https://github.com/cdarlint/winutils>
   1. Create Hadoop directory on the C drive and add an additional bin directory to store the winutils.exe package: C:\Hadoop\bin
3. Configure the environment variables on the local VM
   1. Set the user variables HADOOP\_HOME, JAVA\_HOME, SPARK\_HOME to their associated path locations



* 1. Edit the environment variable for “PATH” and add SPARK\_HOME and HADOOP\_HOME to the path as shown below
  2. Hit ok then close everything out

1. Pull up the command prompt and type in “spark-shell”
   1. If the paths are all aligned, this should pull up!

