Pyspark installation guide:

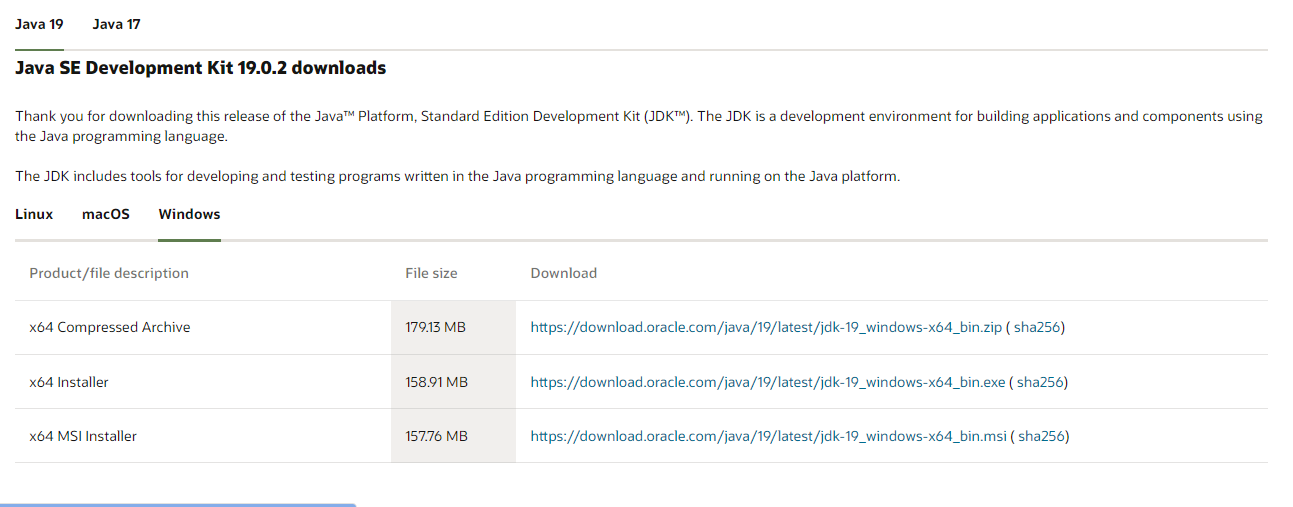
Installing PySpark involves the following steps:

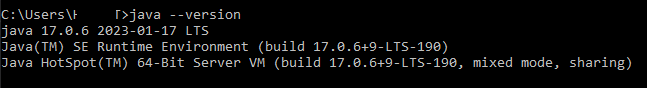
1. Install Java Development Kit (JDK) 8 or higher
2. Download and install Apache Spark
3. Configure PySpark on your system

Here is a step-by-step guide to install PySpark on your windows machine:

**Step 1: Install Java Development Kit (JDK) 8 or higher** PySpark requires Java Development Kit (JDK) version 8 or higher. If you do not have it already installed on your machine, follow the steps below to download and install it:

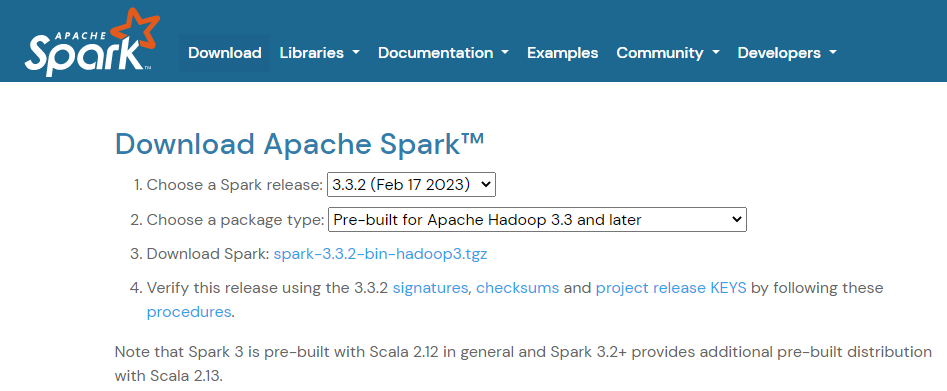
* Go to the official Oracle JDK download page: <https://www.oracle.com/java/technologies/downloads/>
* Select the version of JDK that corresponds to your operating system



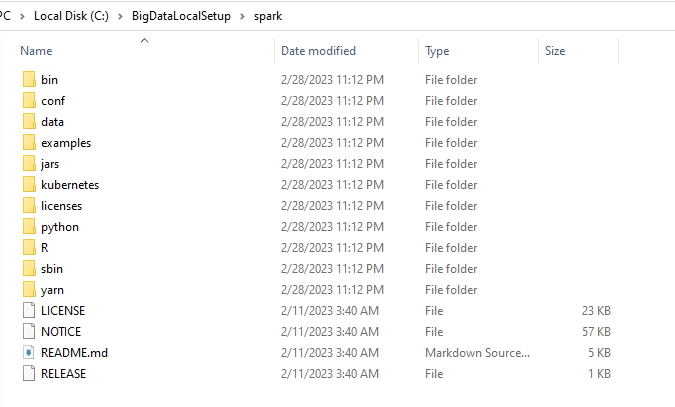
* Accept the license agreement and click on the download link
* Once the download is complete, double-click on the downloaded file to start the installation process
* Follow the prompts to complete the installation
* Verify your installation:  
  

**Step 2: Download and install Apache Spark** The next step is to download and install Apache Spark. Follow these steps:

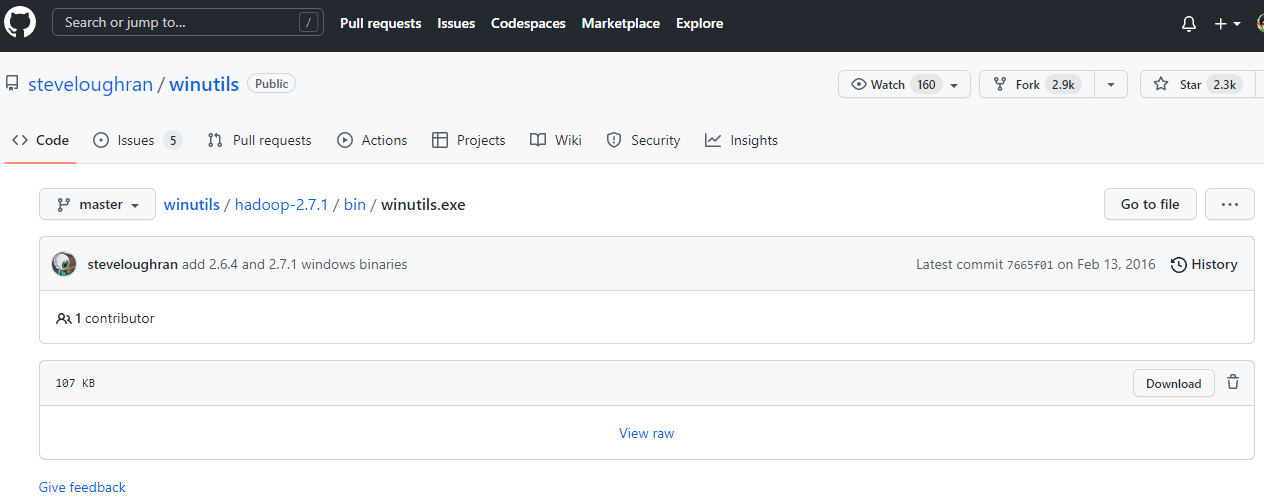
* Go to the official Apache Spark download page: <https://spark.apache.org/downloads.html>
* Select the latest version of Spark



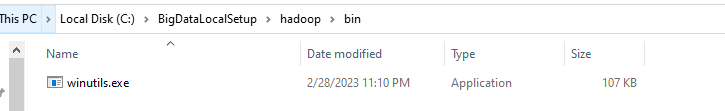
* Choose the package type that corresponds to your operating system and the version of Hadoop you have installed (if applicable)
* Click on the download link to begin the download
* Once the download is complete, extract the contents of the archive to a directory of your choice



* Download Hadoop winutils.exe from this git repo: <https://github.com/steveloughran/winutils/blob/master/hadoop-2.6.0/bin/winutils.exe>

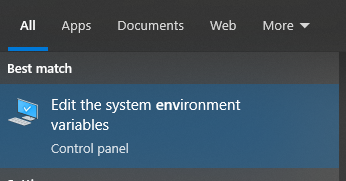


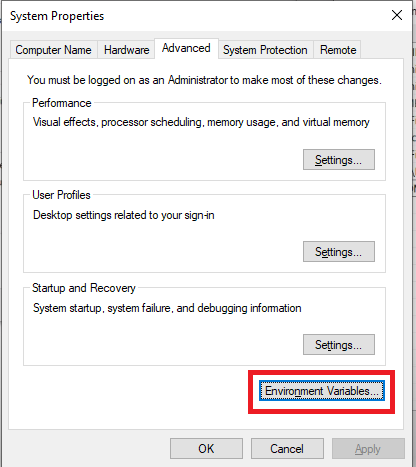
* Once the download is complete, put the file to the directory of your choice



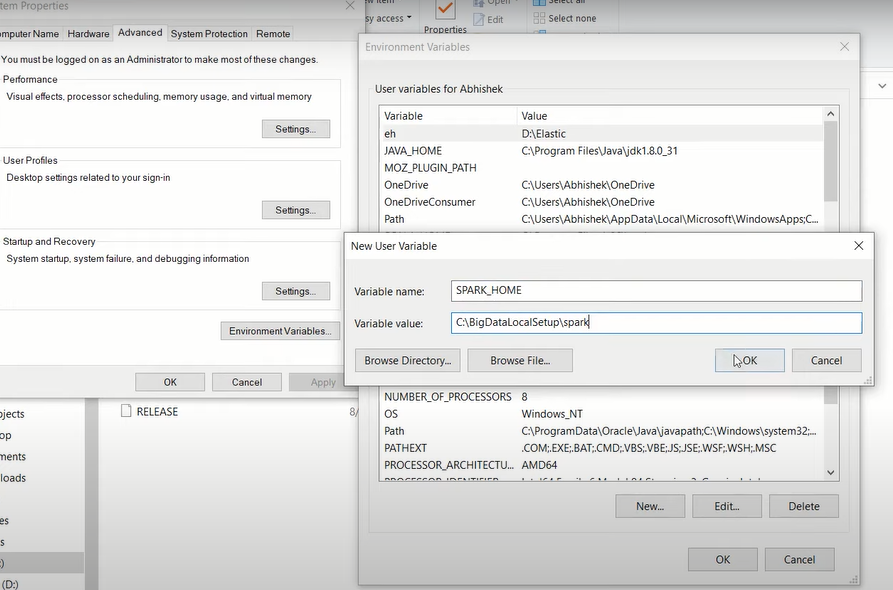
**Step 3: Configure PySpark on your system** The final step is to configure PySpark on your machine. Here are the steps:

* Open your system environment variables

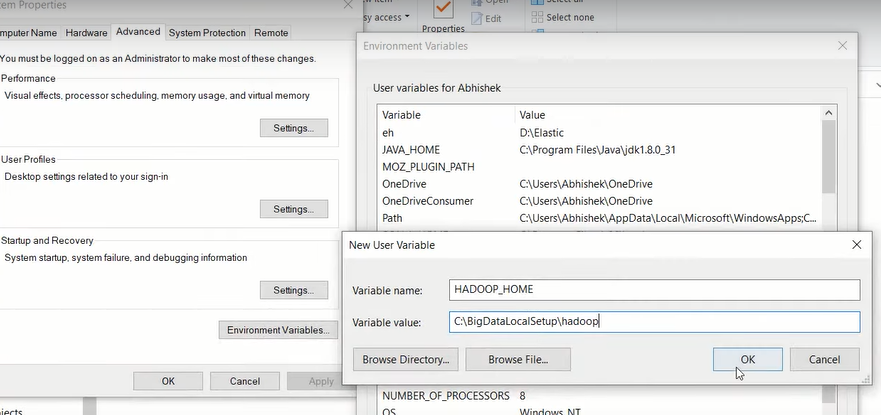




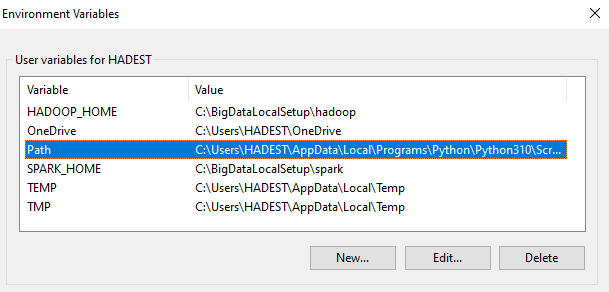
* Setup environment variable SPARK\_HOME with your spark file directory “C:\BigDataLocalSetup\spark”



* Set the variable HADOOP\_HOME for directory “C:\BigDataLocalSetup\hadoop” too and click ok.



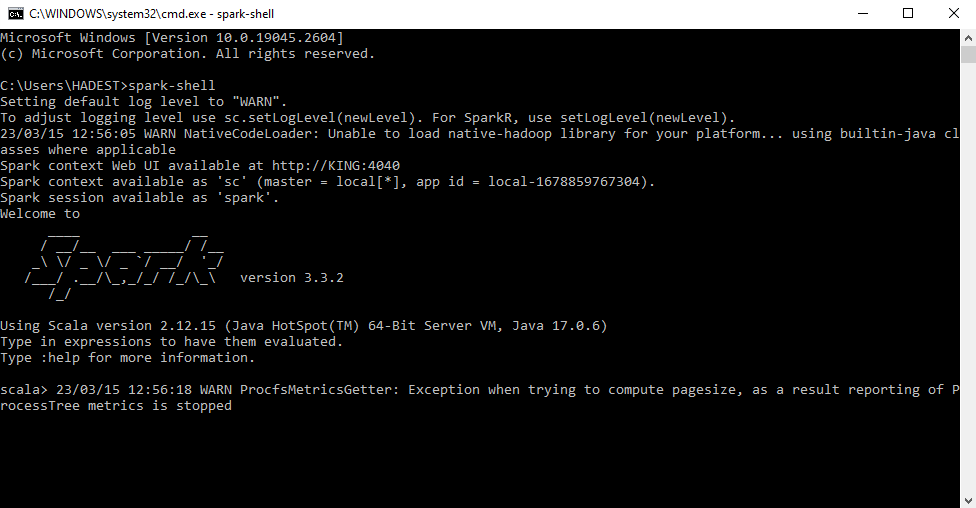
* Next, open the Path variable



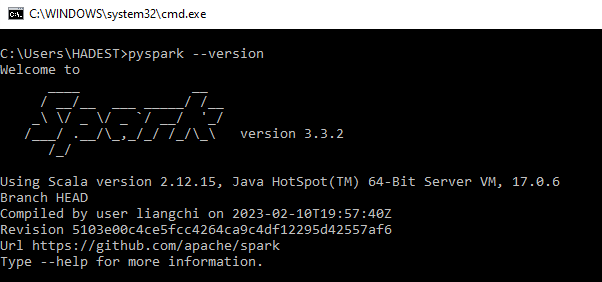
* Add two variable below “%SPARK\_HOME%\bin” and “%HADOOP\_HOME%\bin” then click ok.



* Then open cmd promt and run prompt : “spark --shell”



* Or simply check the version:



* This will print the version of PySpark you have installed.

END