py

|  |
| --- |
| **Submitted By: Ahmed Dider Rahat**  Matriculation Number: 916146 |

Assignment E1

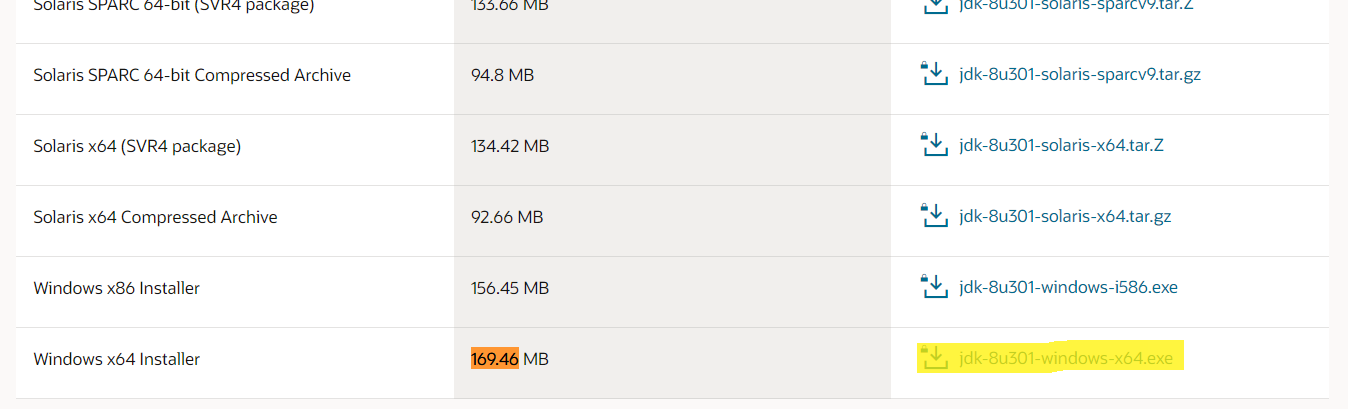
Spark Example for CS4B

**Assignment Code:** After implementing all the assignment I pushed my code on my [Github](https://github.com/AhmedDiderRahat/csfb-wise2122/tree/main/assignment_9) link.

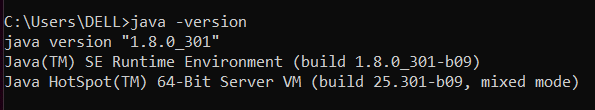
**Answer to the question no. 1**

I have completed the installation process by using some steps. They are:

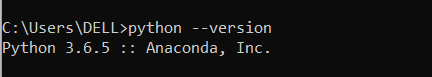
1. **Java Installation:** I had already installed java 17 in my system. So, at first I remove my version of java and then install java 8 from <https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html> . There are several version of jdk uploaded. I choose the marked one.



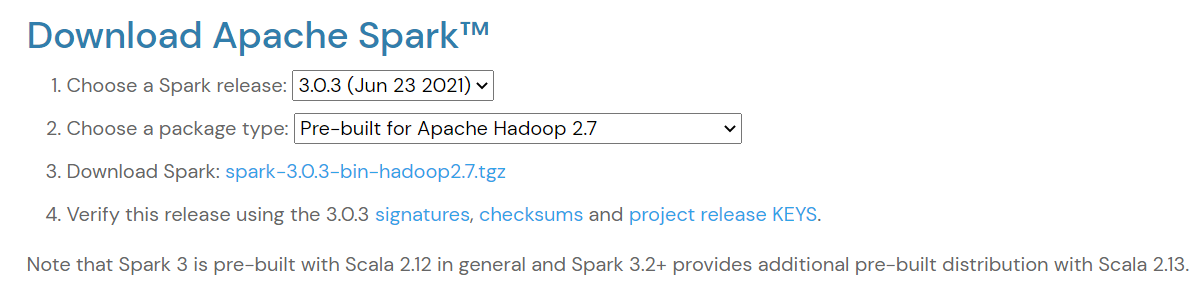
So, the current java version become:



1. **Python Installation:** Python 3 already installed before. So, just check the version:

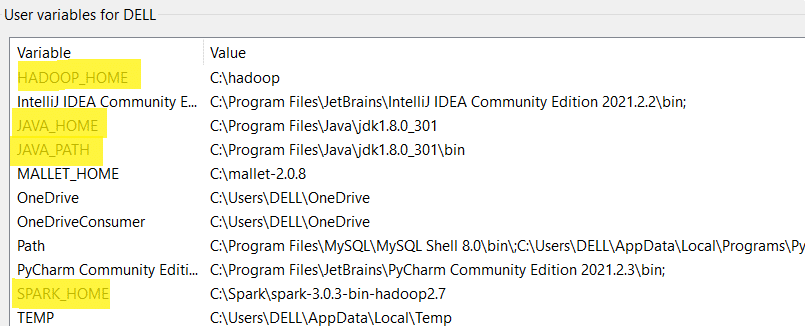
****

1. **Folder Creation:** Create two empty folder in C drive. One is Spark and another one is Hadoop/bin.
2. **Download Spark:** Download spark from <https://spark.apache.org/downloads.html> and download the spark zip file.

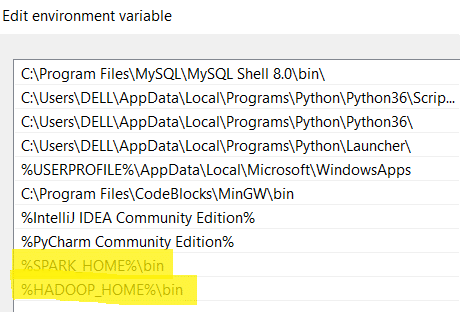
****

After download the zip file, I unzip it and move it to the C:\Spark location.

1. **Download winutils:** I download the winutils file from github. As I used hadoop 2.7 my link was <https://github.com/cdarlint/winutils/tree/master/hadoop-2.7.7/bin>. Then I move it to C:\hadoop\bin.
2. Setup environment variables: For the setting I need to setup the environment variables. So, I added them in the environment variable.

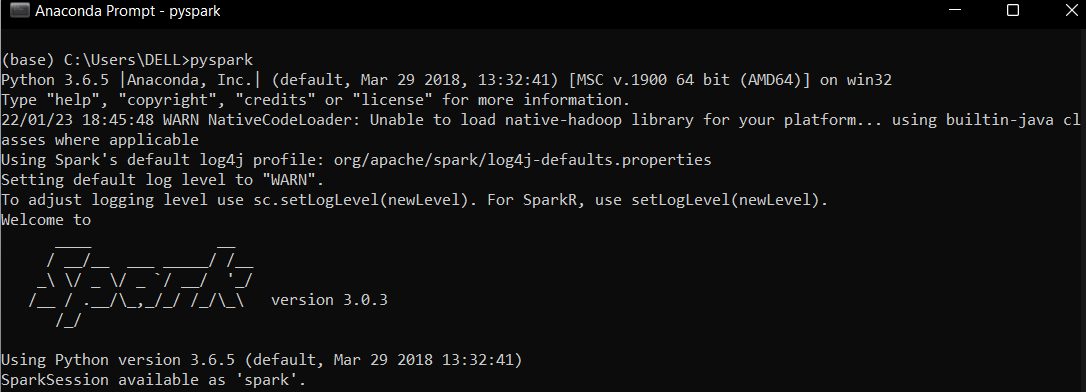
****

Then add the path of spark and hadoop in the path section.

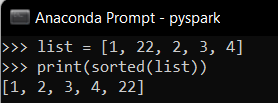
****

**Answer to the question no. 2**

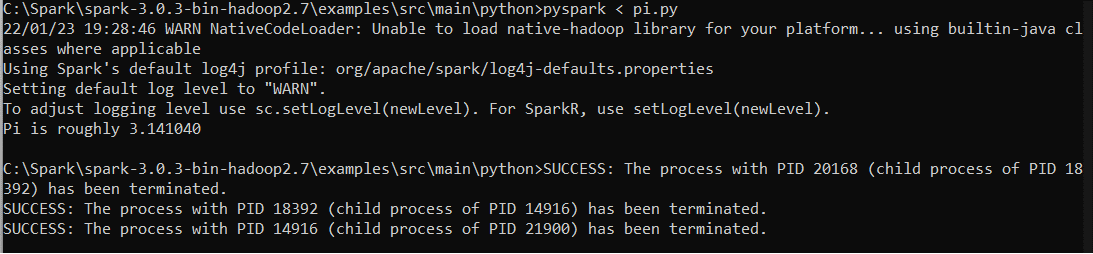
After successfully installation my Anaconda promt become:



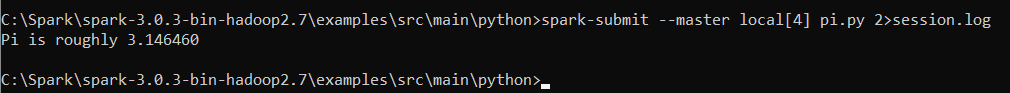
i) **Run simple python command and see the output:**



ii) **Run pi.py:**



iii) **spark-submit ‐‐master local[4] pi.py 2>session.log:**



**Answer to the question no. 3**

