py

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

Assignment E1

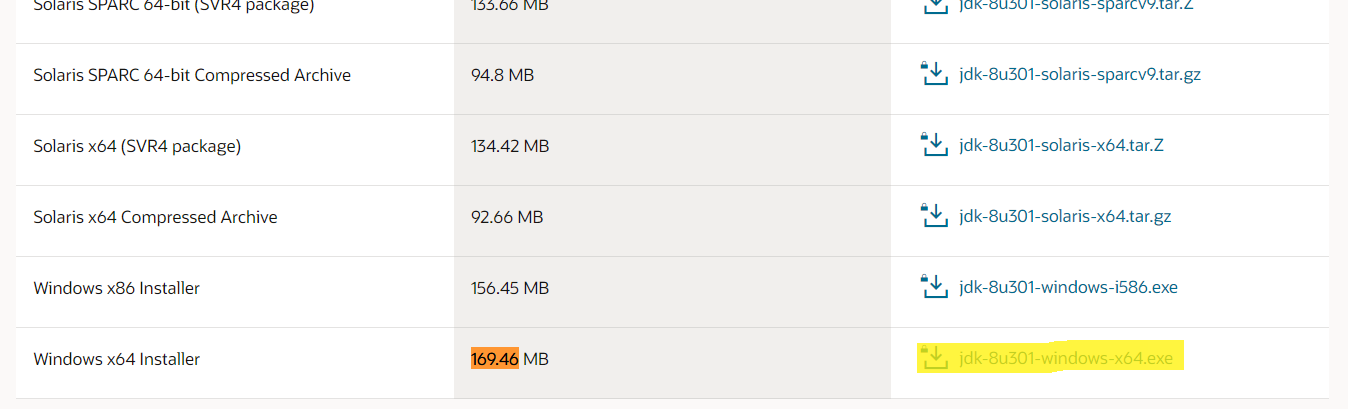
Spark Example for CS4B

**Assignment Code:** After implementing all the assignment I pushed my code on my [Github](https://github.com/AhmedDiderRahat/csfb-wise2122/tree/main/assignment_7) link.

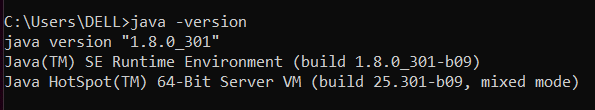
**Answer to the question no. 1**

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

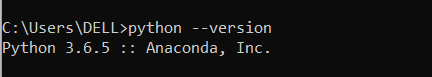
1. **Java Installation:** I had already installed java 17 in my system. So, at first I remove my version of java and then install java 8 from <https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html> . There are several version of jdk uploaded. I choose the marked one.



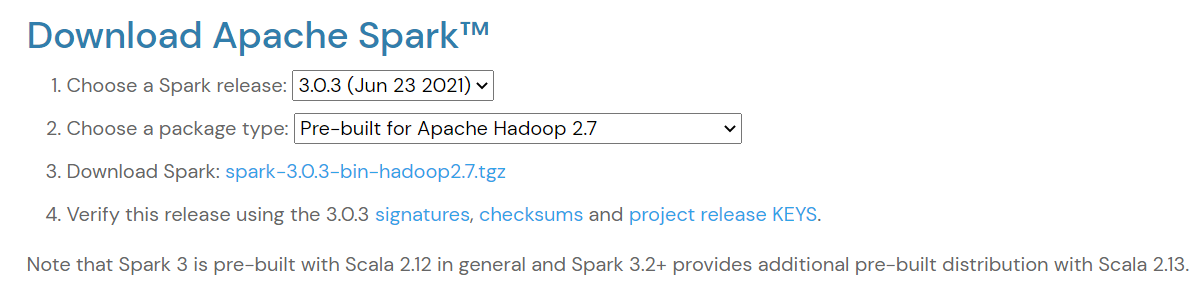
So, the current java version become:



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

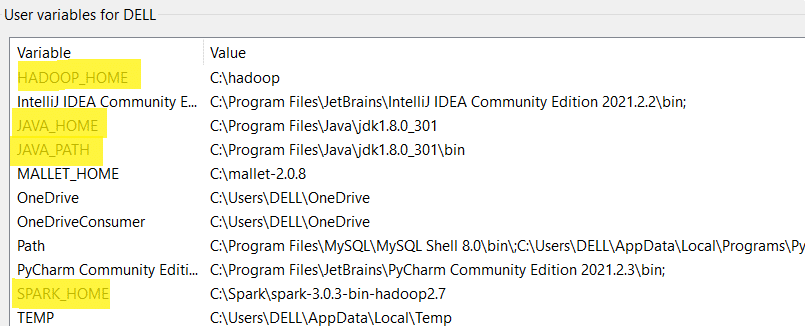
****

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.

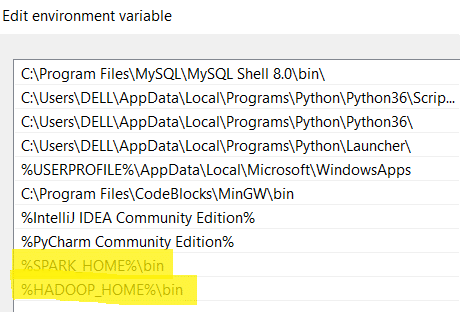
****

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.

****

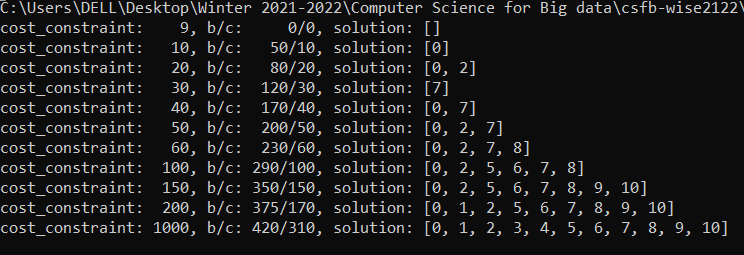
Then add the path of spark and hadoop in the path section.

****

The implementation is done in Qustion1.py file.

**Answer to the question no. 2**

After running the code, the following output is found:



**Answer to the question no. 3**

**Optimized total beefit:** Choose a sequence which has maximum benefits under the given cost costraint.

**Optimal total cost:** Get the sequence which has maximum benefits and also minimum cost within a certain cost constraint.

**Answer to the question no. 4**

The code of this question is implemented in Question\_5.py file. The optimal solution under cost constaint of 200 is given bellow:

**a) Total benefit:** 580

**b) Total Cost:** 200

**c) Task arrangement:** [0, 2, 5, 6, 7, 8, 13, 14, 15, 16, 17]

**Answer to the question no. 5**

When uncommented the last two task and run the code, the execution took much time then before. Because of,

1. Now there are 20 item so all possible combination become 2^20 = 1,048,576 > 1 M.
2. For 18 items, 2^18 = 262,144. So, total number of operation increse almost 4 times.

**Answer to the question no. 6**

For 20 tasks total solution space become, 2^20 = 1,048,576.

For 11 tasks total solution space become, 2^11 = 2048.

So, the expansion become: (1,048,576 - 2048) = 1046528 > 1M.