

DR. ALVIN'S PUBLICATIONS

BUILDING AN APACHE SPARK LOCAL CLUSTER

ON WINDOWS
DR. ALVIN ANG



1 | PAGE

COPYRIGHTED BY DR ALVIN ANG
WWW.ALVINANG.SG

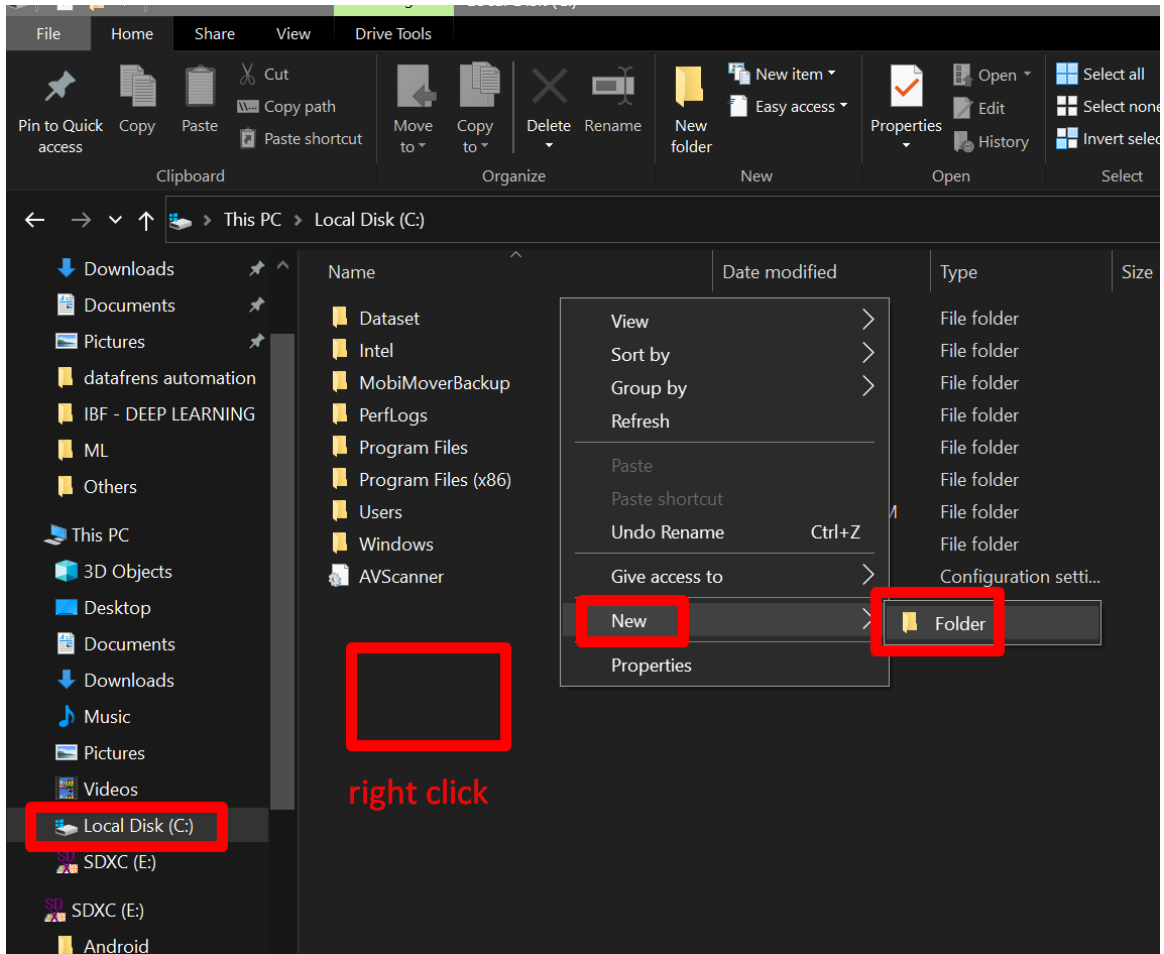
CONTENTS

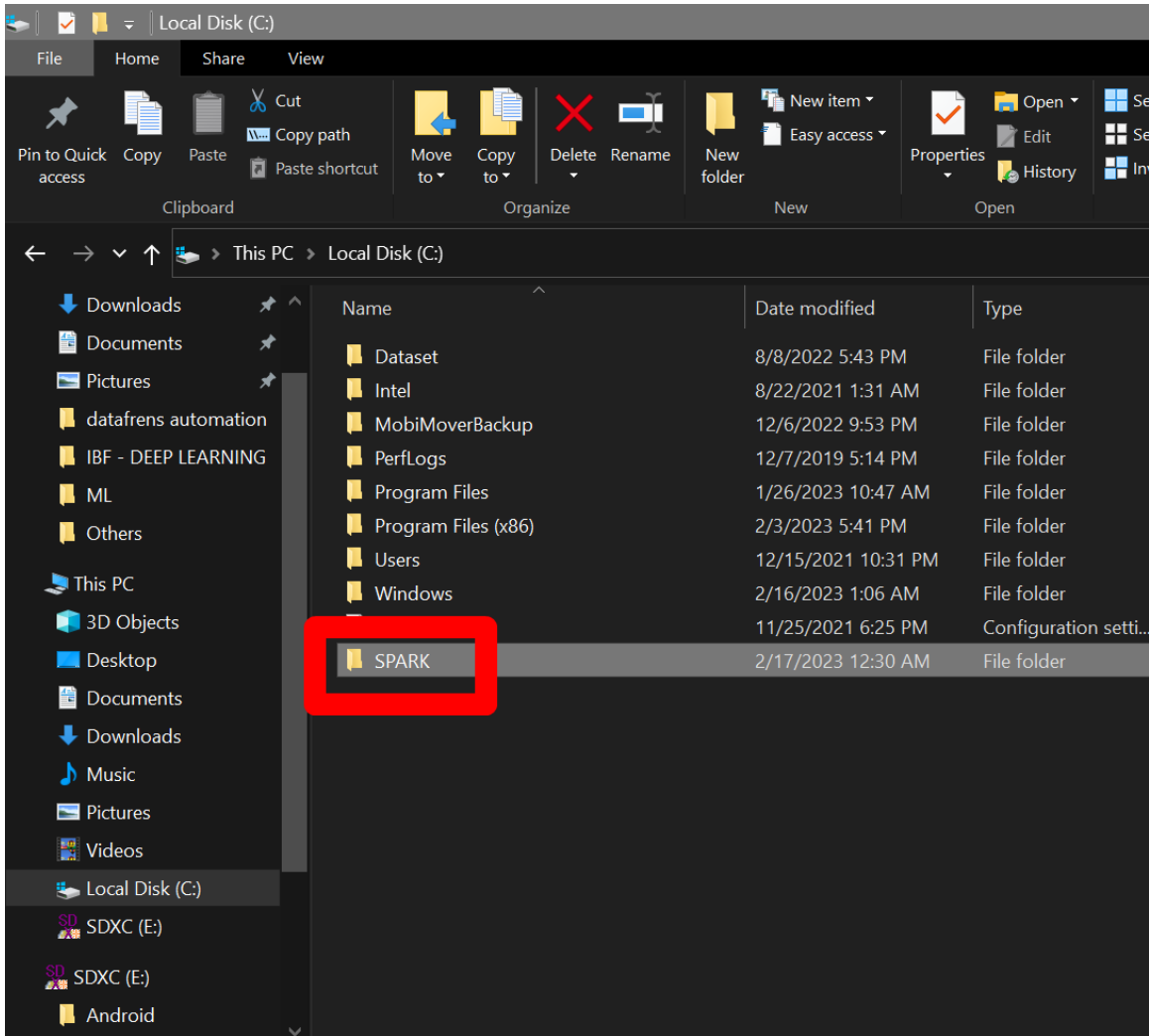
I. Installation Steps	4
A. Create a New Folder called Spark in C:\	4
B. Install Anaconda	6
C. Install Scala.....	11
D. Install Apache Spark.....	14
E. Install JDK 8	17
F. Install Hadoop	23
II. Creating Environment Variables	26
A. Creating System Variables	28
1. Create HADOOP_HOME Variable.....	29
2. Create SCALA_HOME Variable	30
3. Create SPARK_HOME Variable	31
4. Create JAVA_HOME Variable	32
5. Create PYSARK_PYTHON Variable.....	33
B. Creating PATH Variables.....	34
III. Test APACHE SPARK Now!.....	35
IV. Add Jupyter Notebook to PATH	38
V. Using Jupyter Notebook to Run Apache Spark	40
A. Create Spark Project Folder in C:\.....	40
B. TEST CODE	42
VI. Setup MASTER Computer	43
A. Step 1	43
B. Step 2	43
C. Step 3: Appoint your Master Laptop	44
1. CODE	44
D. Check to see if there are any SLAVES.....	47
VII. Setup SLAVE Computer	48
A. Check your Wifi Network (both Master and Slave)	49
B. Turn Off all your Firewalls (Master Only)	50
C. Now from SLAVE, Ping your Master Again....	52

VIII. Load / Deploy your SLAVE (or worker) into the Cluster.....	53
A. CODE	53
B. Check to see if there are any SLAVES.....	54
IX. Start Jupyter Notebook in your Master.....	55
A. CODE	56
B. Check Localhost	57
C. Open Any Ipython	58
X. Seeking Further Help	60
XI. Disclaimer	61
A. Issue: Slave App keeps on Exiting and Restarting	61
B. Possible Solutions	63
About Dr. Alvin Ang	64

I. INSTALLATION STEPS

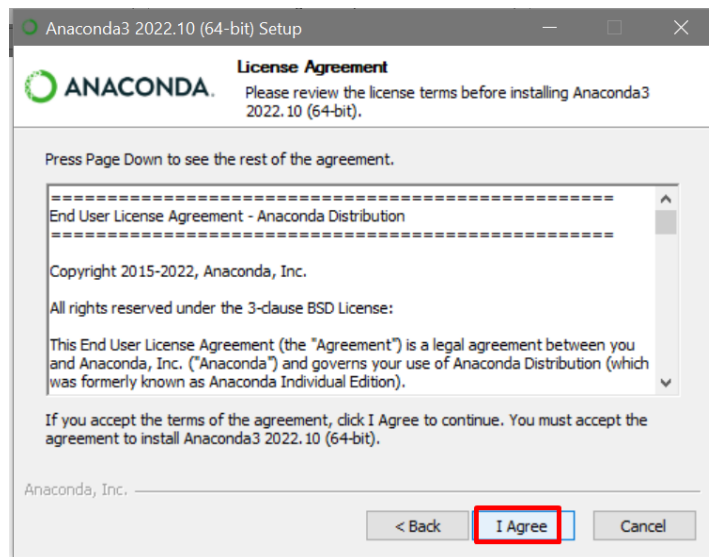
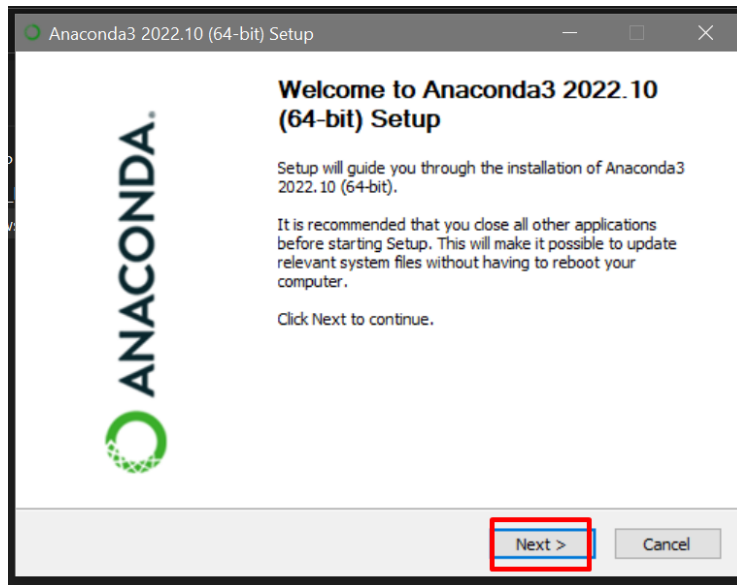
A. CREATE A NEW FOLDER CALLED SPARK IN C:\

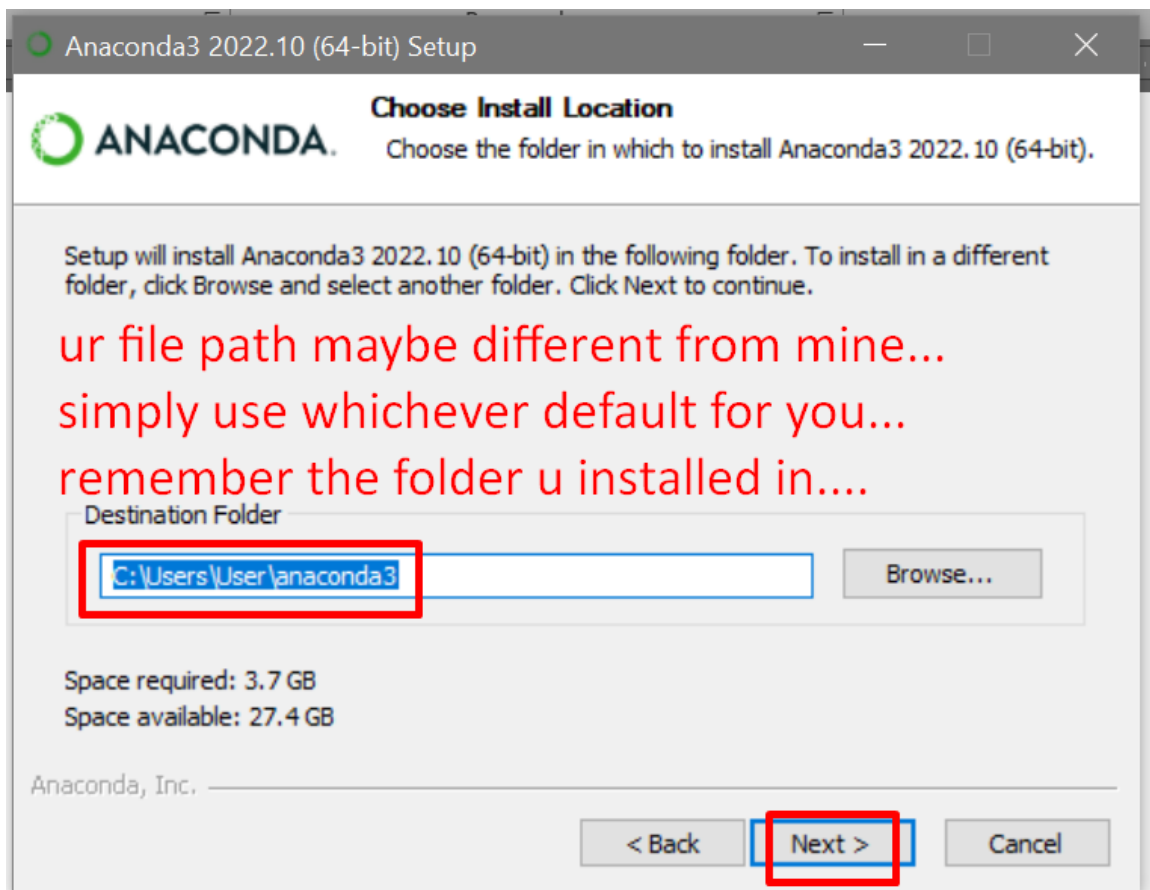
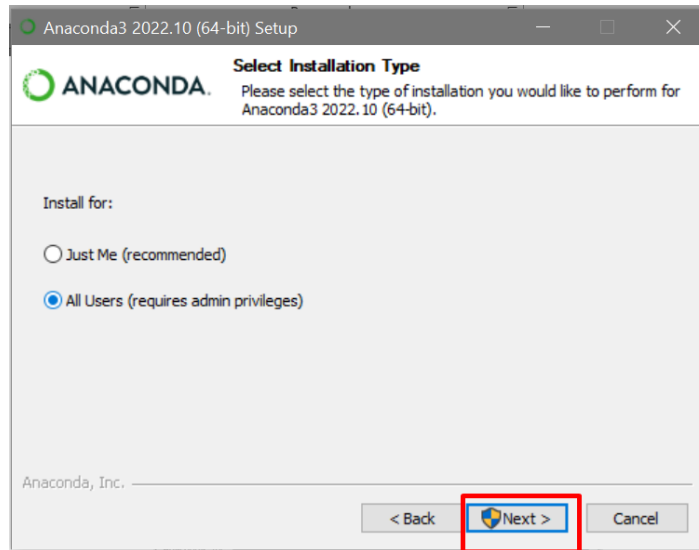


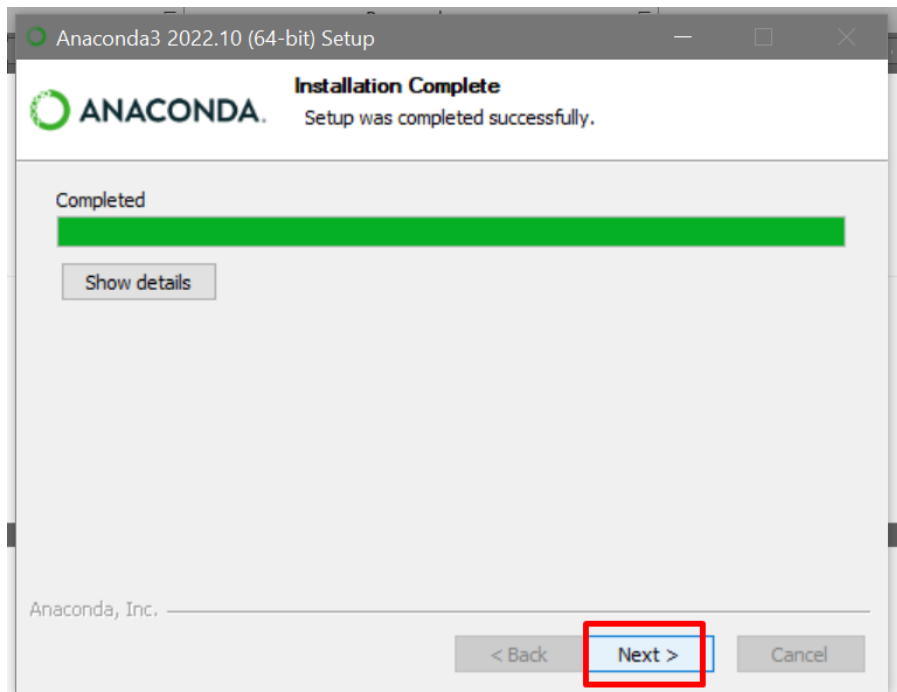
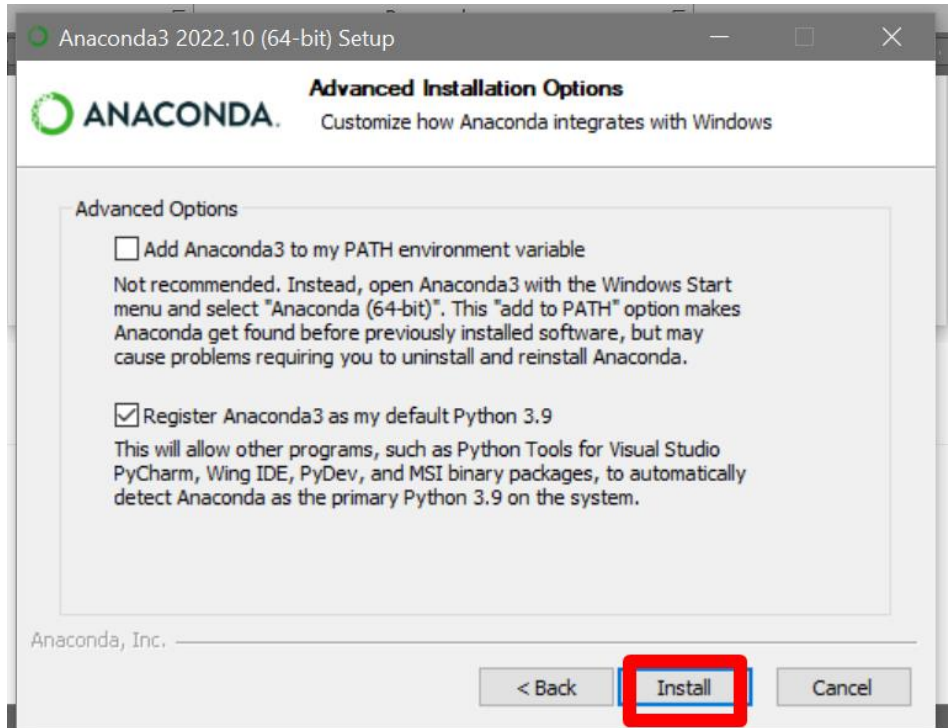


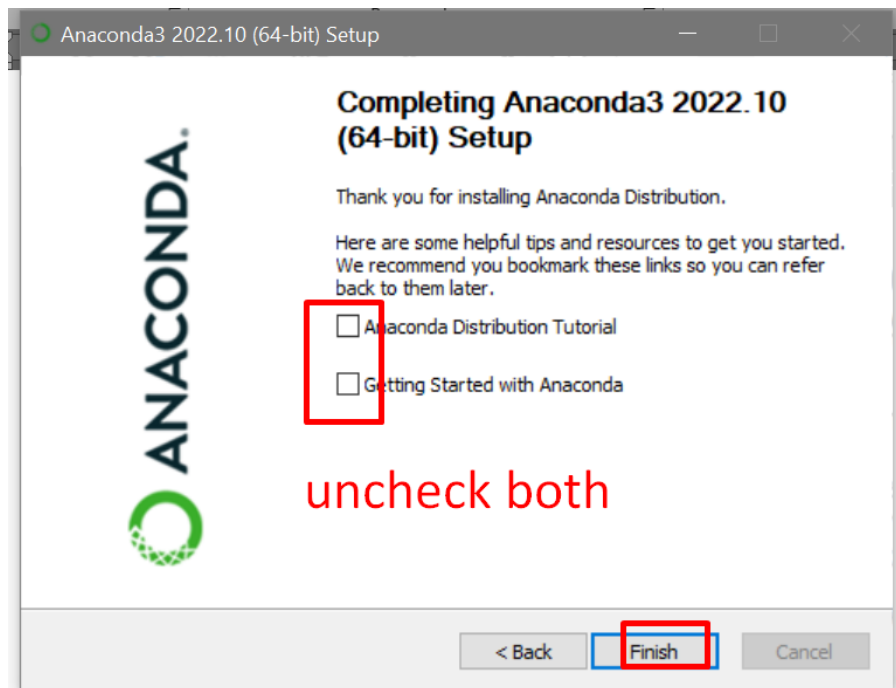
B. INSTALL ANACONDA

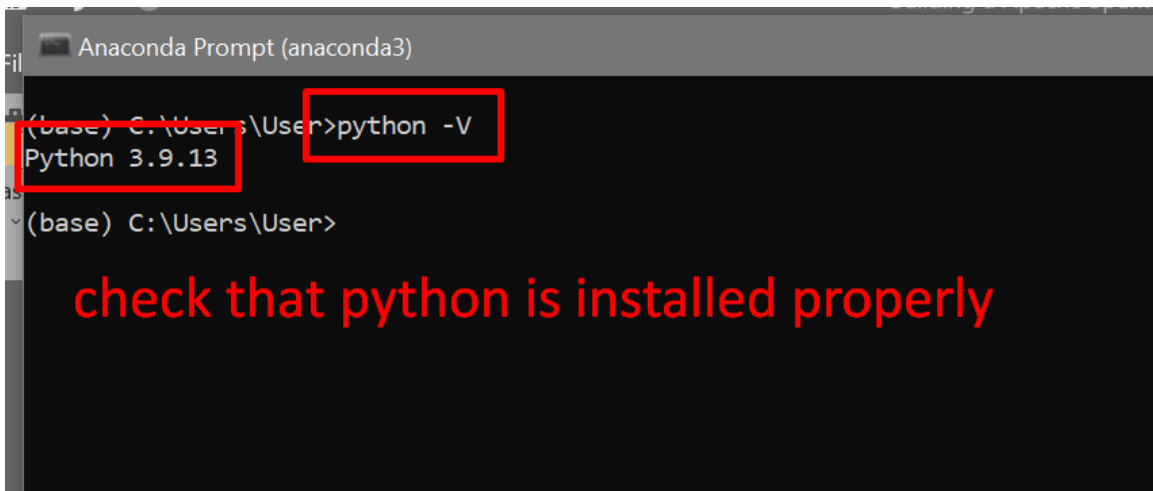
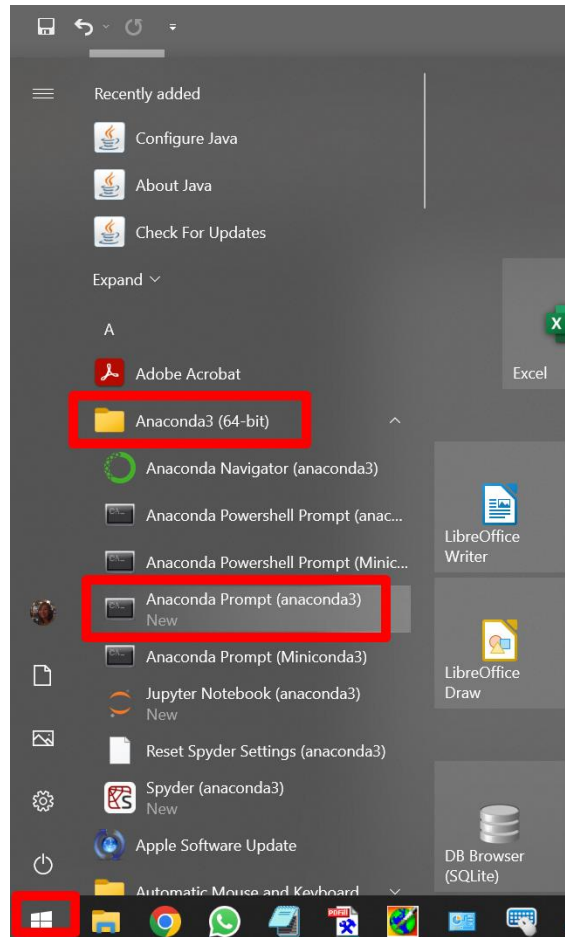
- https://repo.anaconda.com/archive/Anaconda3-2022.10-Windows-x86_64.exe





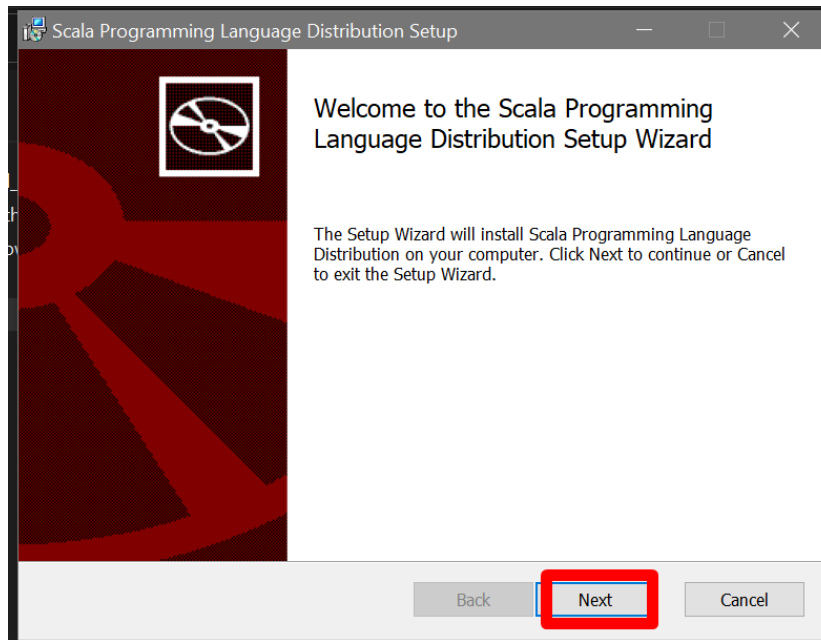


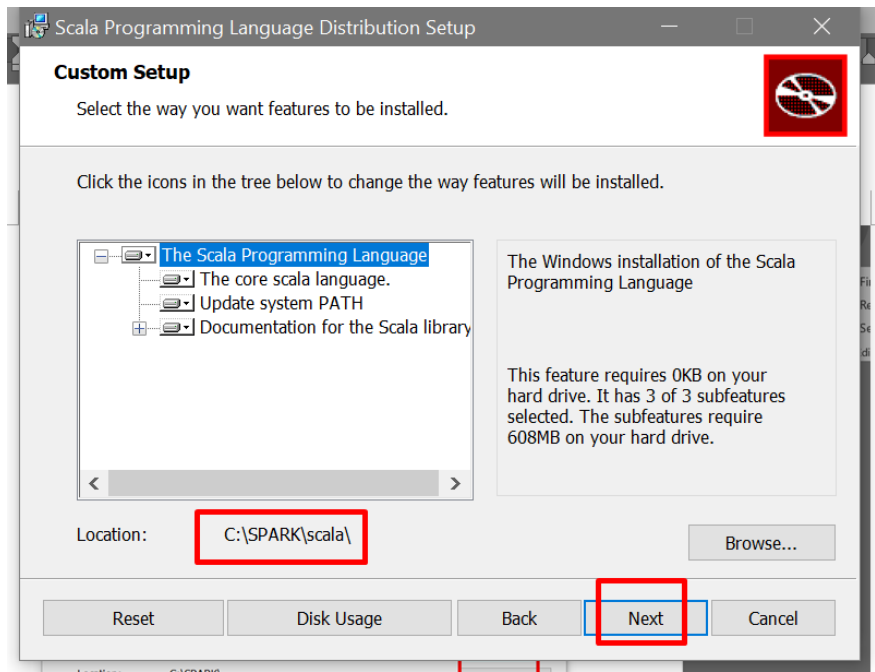
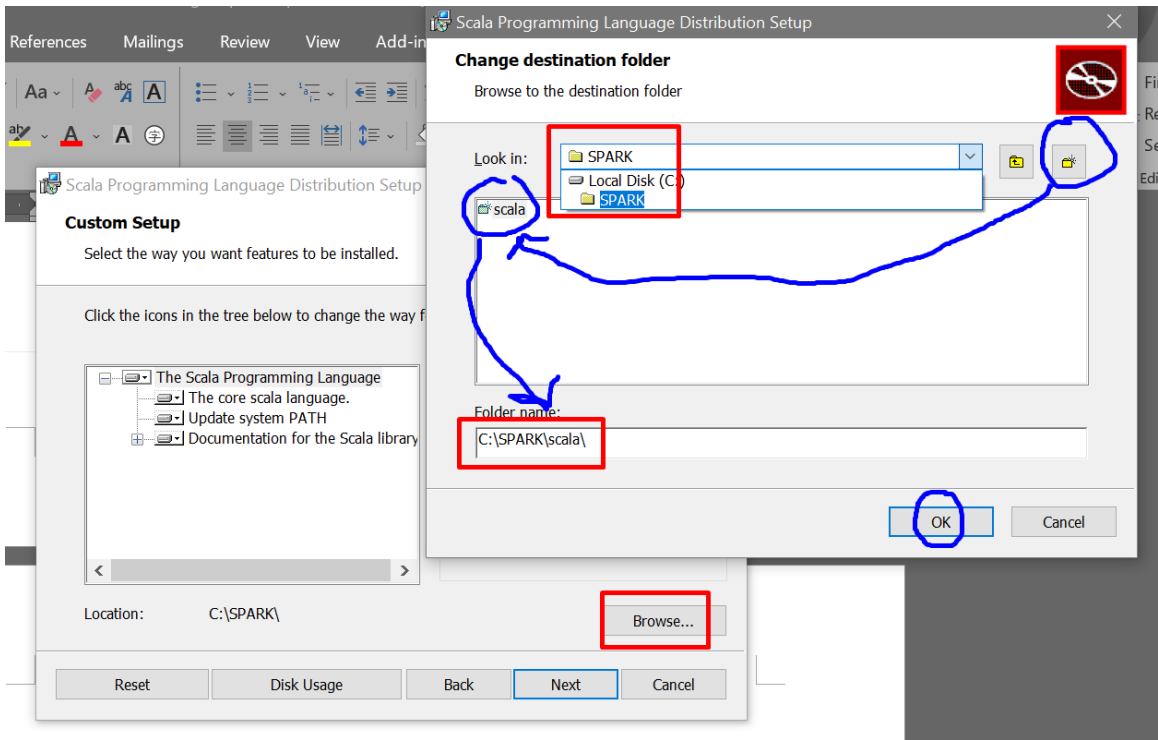


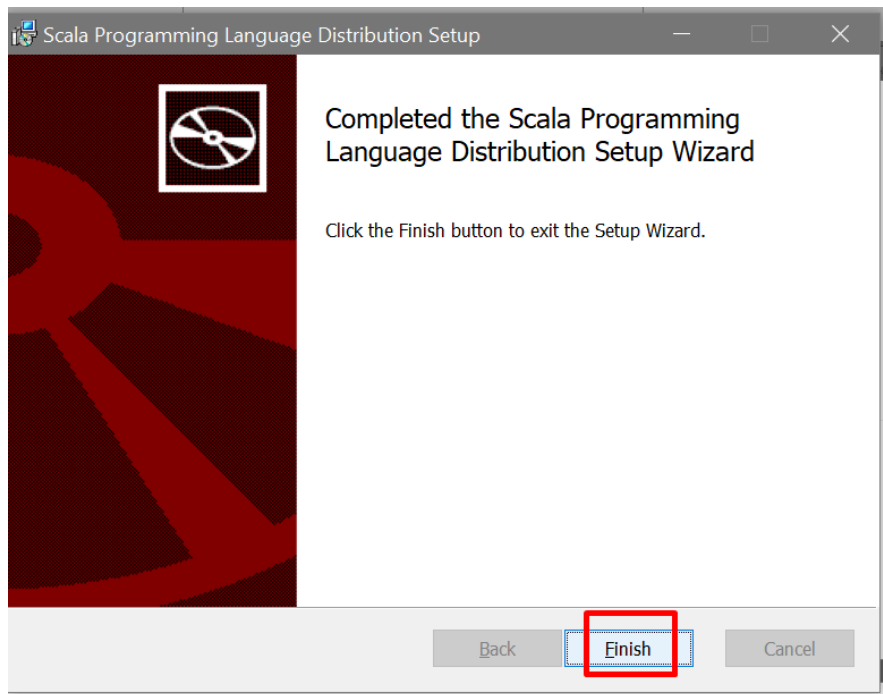
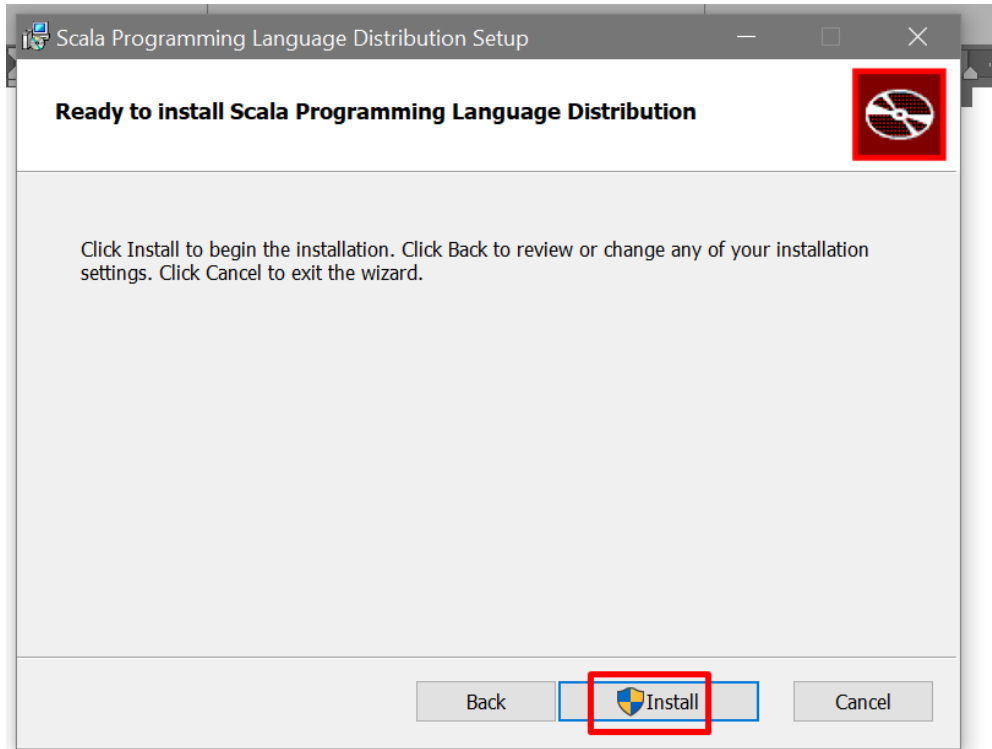


C. INSTALL SCALA

- <https://downloads.lightbend.com/scala/2.12.4/scala-2.12.4.msi>







D. INSTALL APACHE SPARK

- <https://www.apache.org/dyn/closer.lua/spark/spark-3.3.1/spark-3.3.1-bin-hadoop2.tgz>

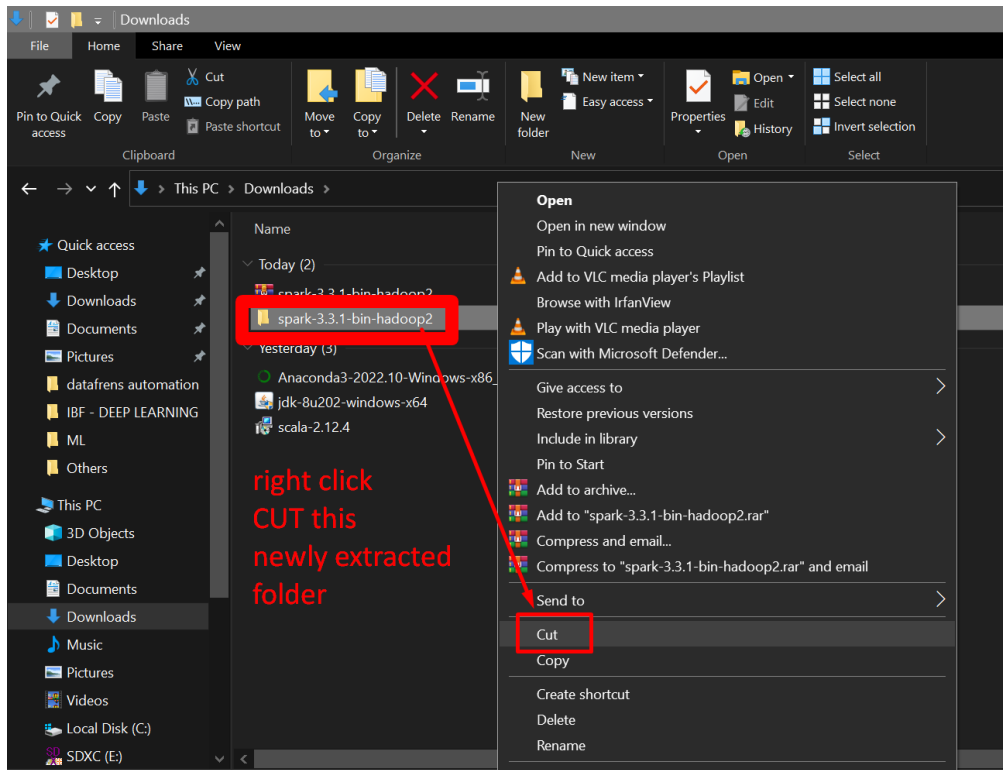
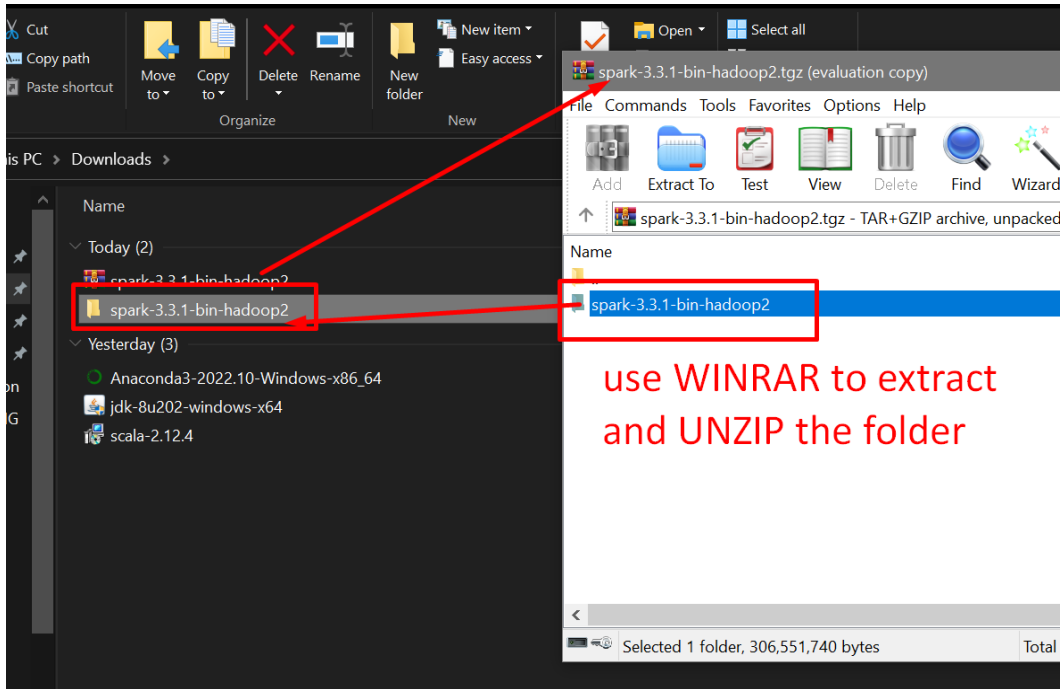
OR

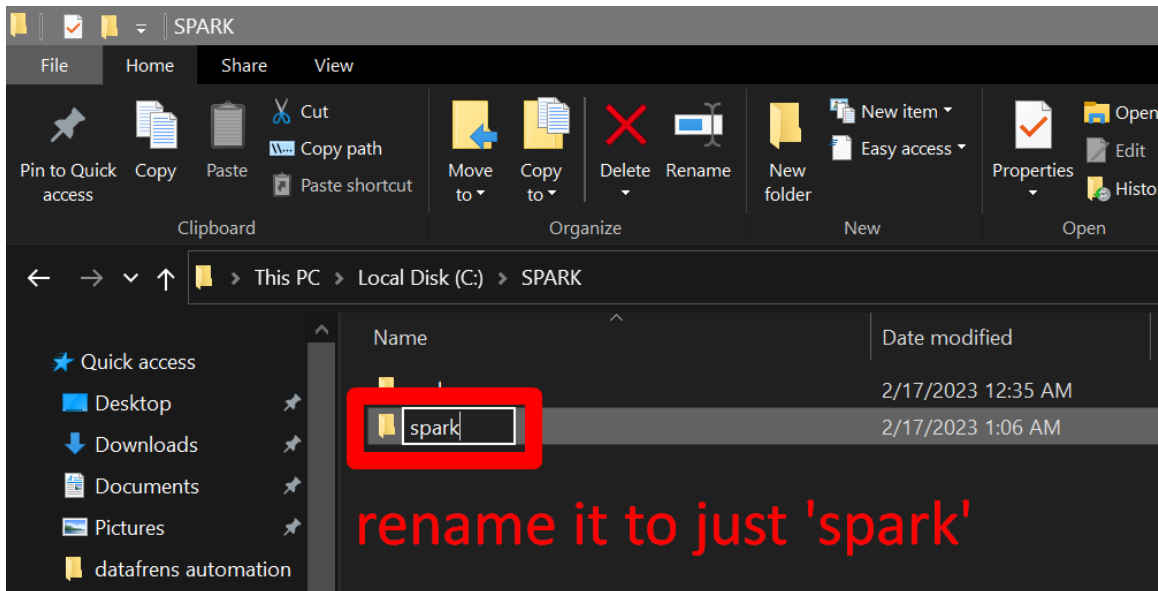
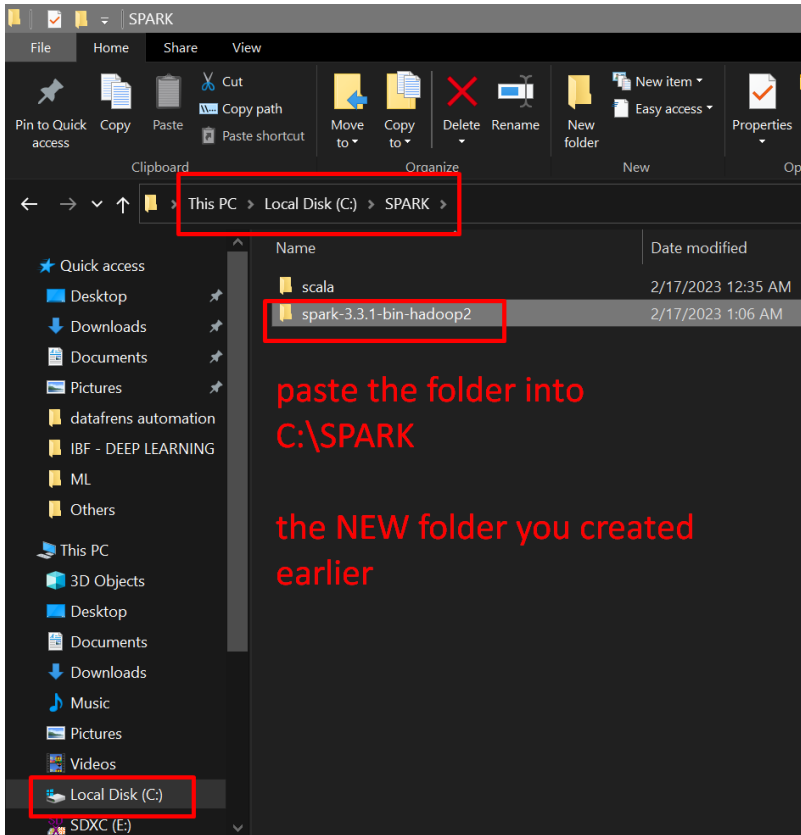
- <https://spark.apache.org/downloads.html>

The screenshot shows the Apache Spark download page. At the top, there is a navigation bar with the Apache Spark logo and links for Download, Libraries, Documentation, Examples, Community, and Developers. The main heading is "Download Apache Spark™". Below this, there are four steps:

1. Choose a Spark release: 3.3.1 (Oct 25 2022) (selected)
2. Choose a package type: Pre-built for Apache Hadoop 2.7 (selected)
3. Download Spark: spark-3.3.1-bin-hadoop2.tgz (selected)
4. Verify this release using the 3.3.1 [signatures](#), [checksums](#) and [project release KEYS](#) by following these [procedures](#).

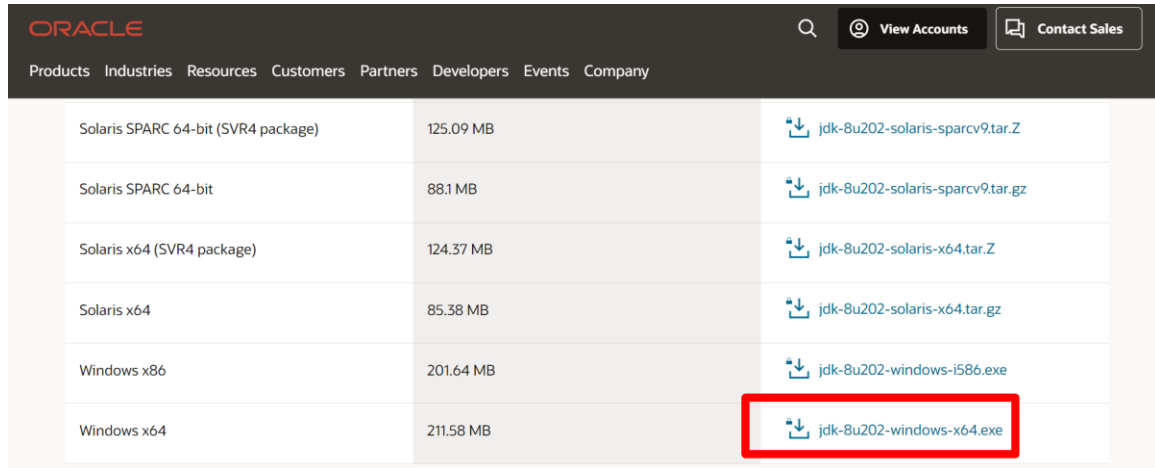
Below the steps, there is a navigation bar with links for News, About, Make a Donation, The Apache Way, Join Us, and Dow. The Apache Software Foundation logo is on the left, and the text "COMMUNITY-LED DEVELOPMENT 'THE APACHE WAY'" is on the right. Below this, there are links for Projects, People, Community, License, and Spons. The main content area says "We suggest the following site for your download:" and lists the URL <https://dlcdn.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-hadoop2.tgz>. Below this, it says "Alternate download locations are suggested below." and "It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file)." There are sections for "HTTP" and "BACKUP SITE" with the same URL. At the bottom, there is a download progress bar showing the file name "spark-3.3.1-bin-ha...tgz" and the progress "15.7/261 MB, 1 min left".



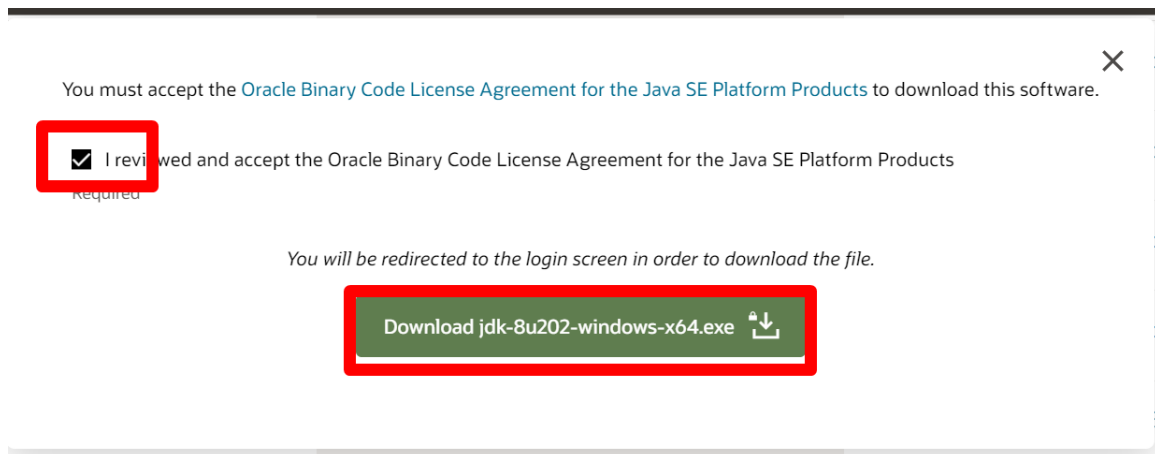


E. INSTALL JDK 8

<https://www.oracle.com/java/technologies/javase/javase8-archive-downloads.html#license-lightbox>




Operating System	Size	Download Link
Solaris SPARC 64-bit (SVR4 package)	125.09 MB	jdk-8u202-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	88.1 MB	jdk-8u202-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	124.37 MB	jdk-8u202-solaris-x64.tar.Z
Solaris x64	85.38 MB	jdk-8u202-solaris-x64.tar.gz
Windows x86	201.64 MB	jdk-8u202-windows-i586.exe
Windows x64	211.58 MB	jdk-8u202-windows-x64.exe

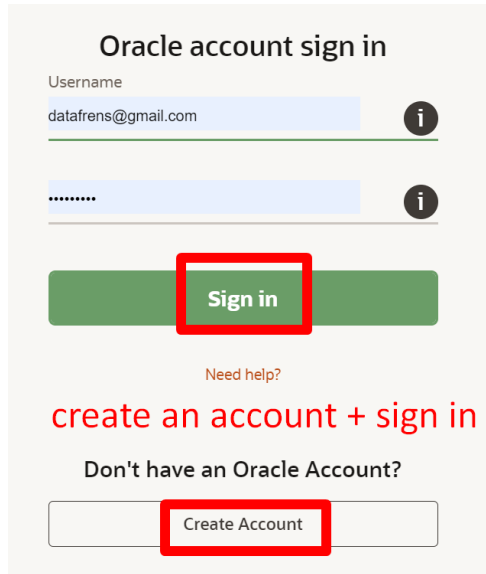


You must accept the [Oracle Binary Code License Agreement for the Java SE Platform Products](#) to download this software.

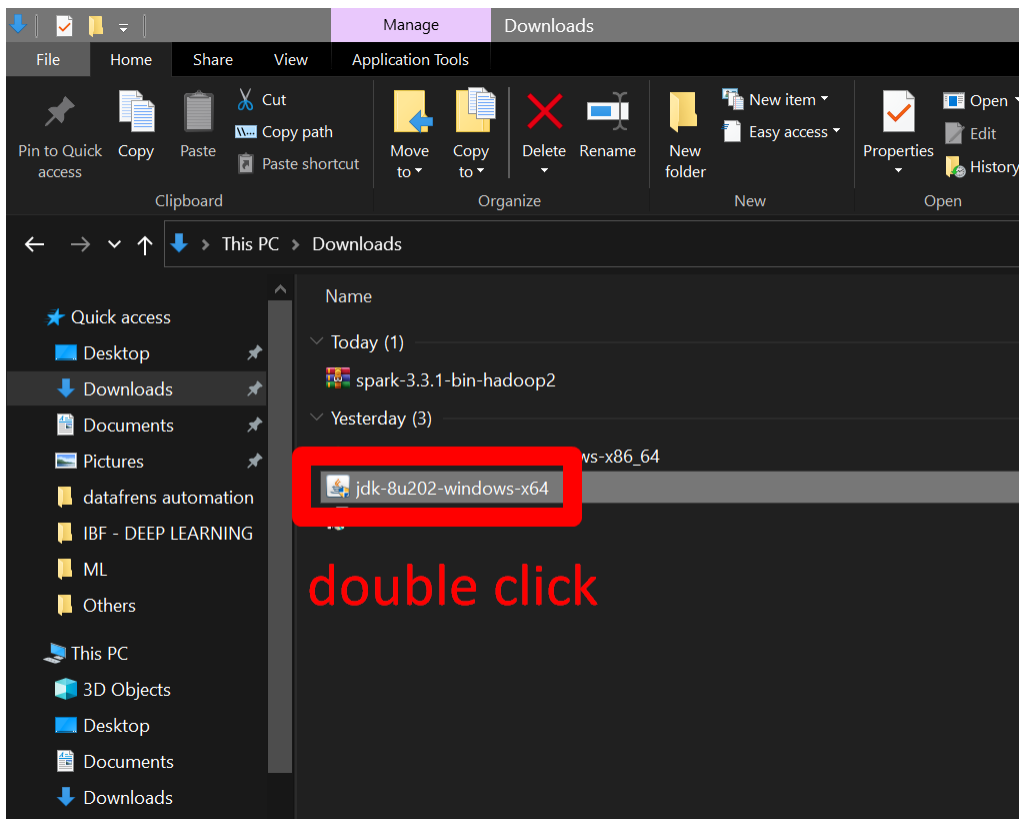
I reviewed and accept the Oracle Binary Code License Agreement for the Java SE Platform Products
Required

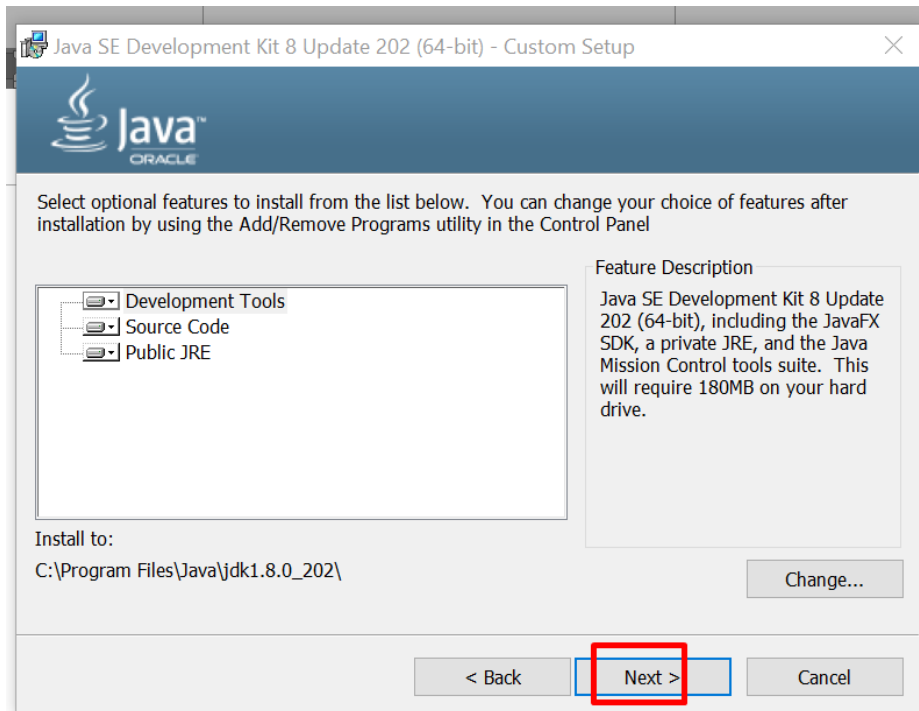
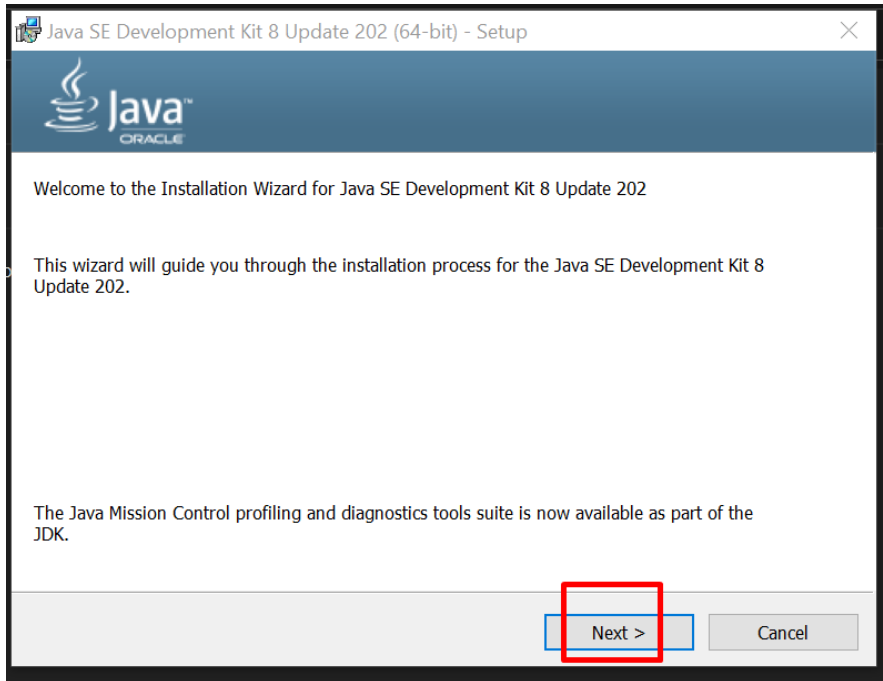
You will be redirected to the login screen in order to download the file.

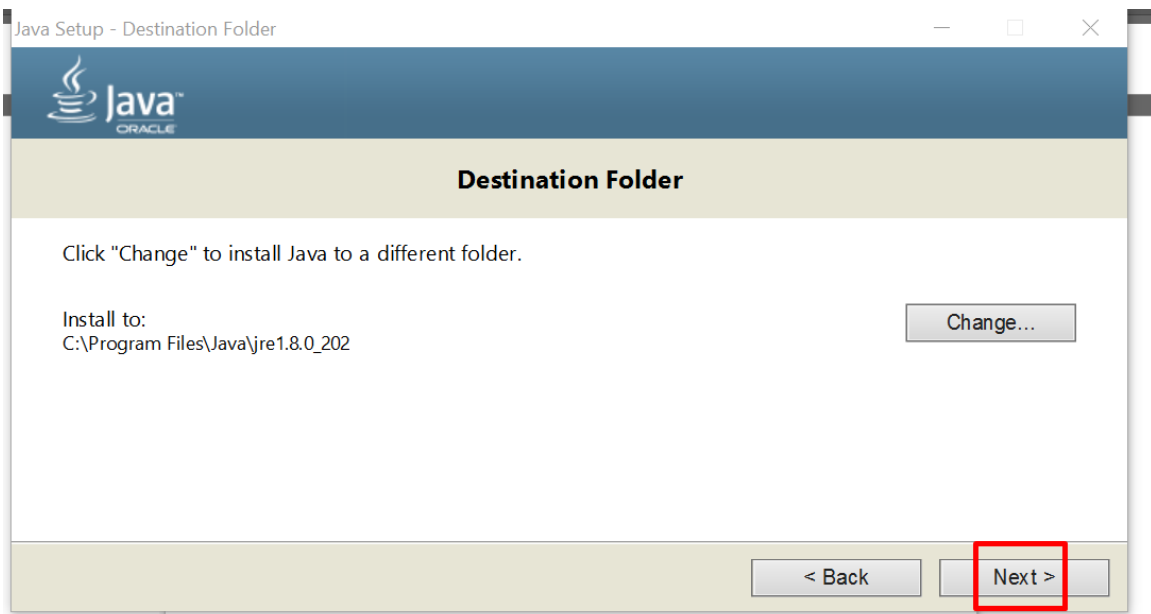
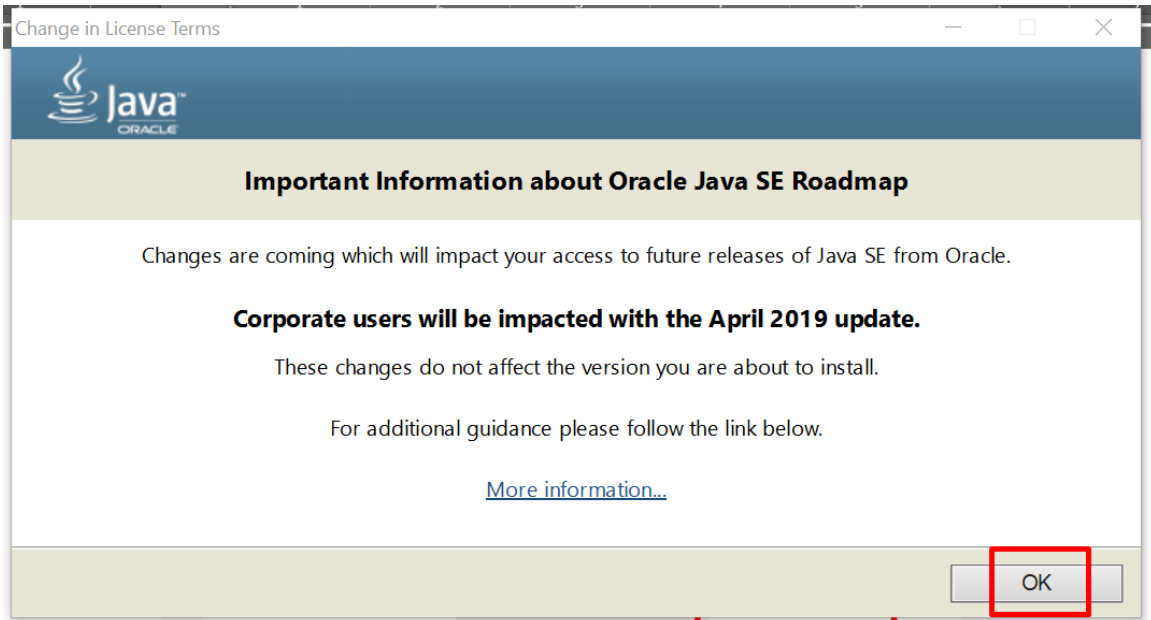
[Download jdk-8u202-windows-x64.exe](#) 

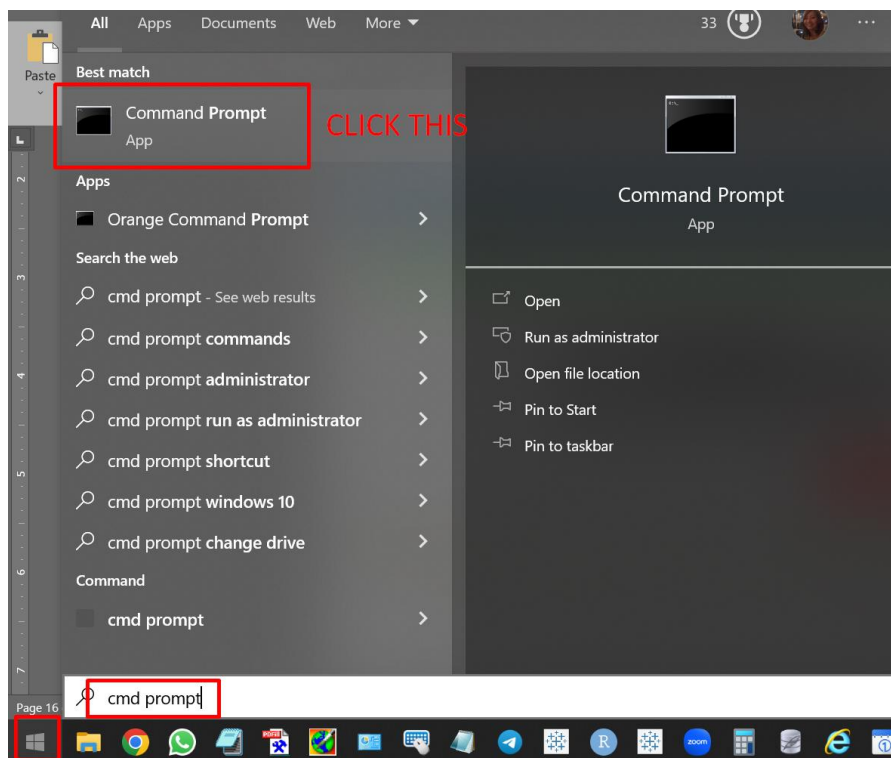


It will download a file after signing in...









```
Command Prompt
Microsoft Windows [Version 10.0.19044.2604]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User>java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\Users\User>
```

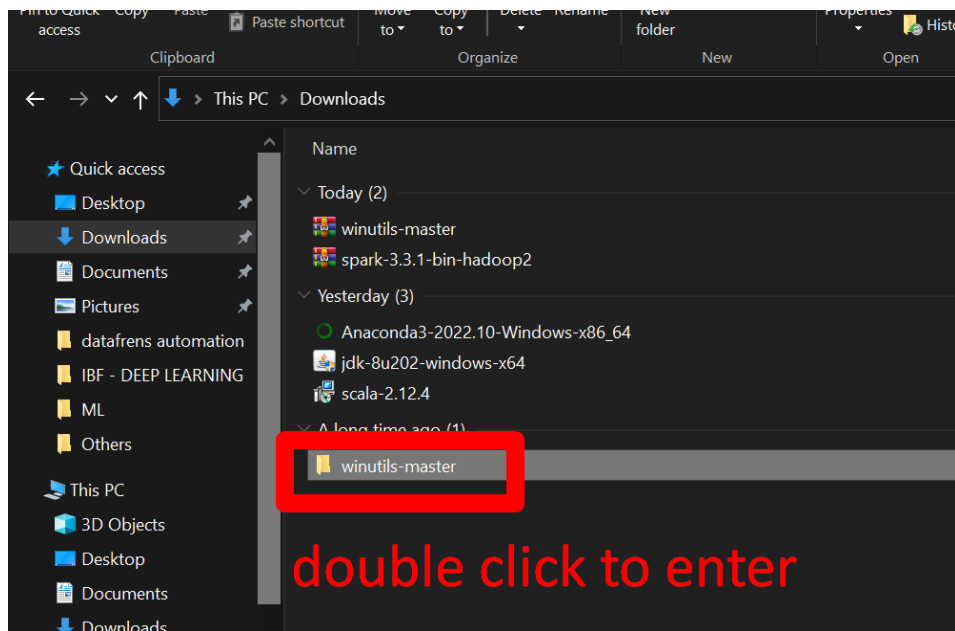
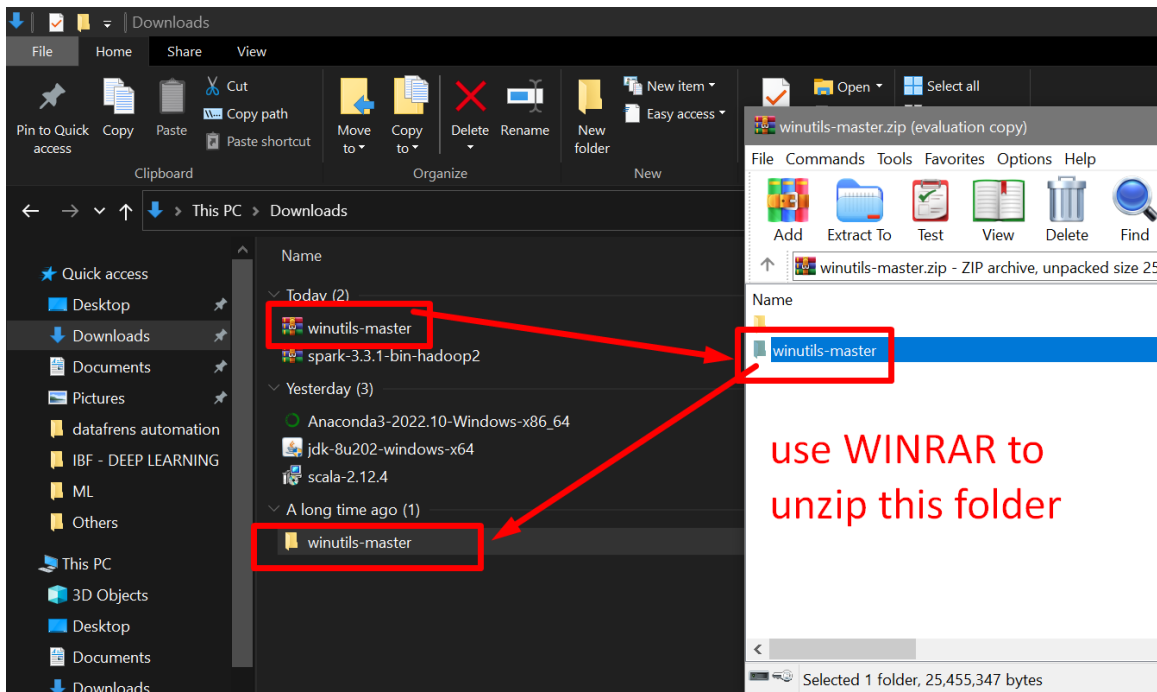
check that java is installed properly

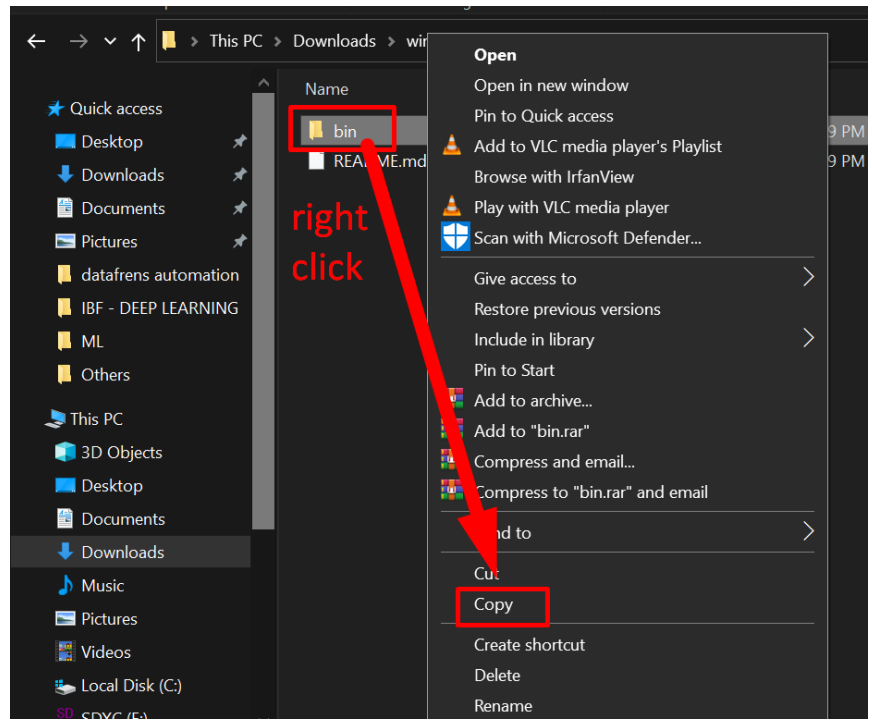
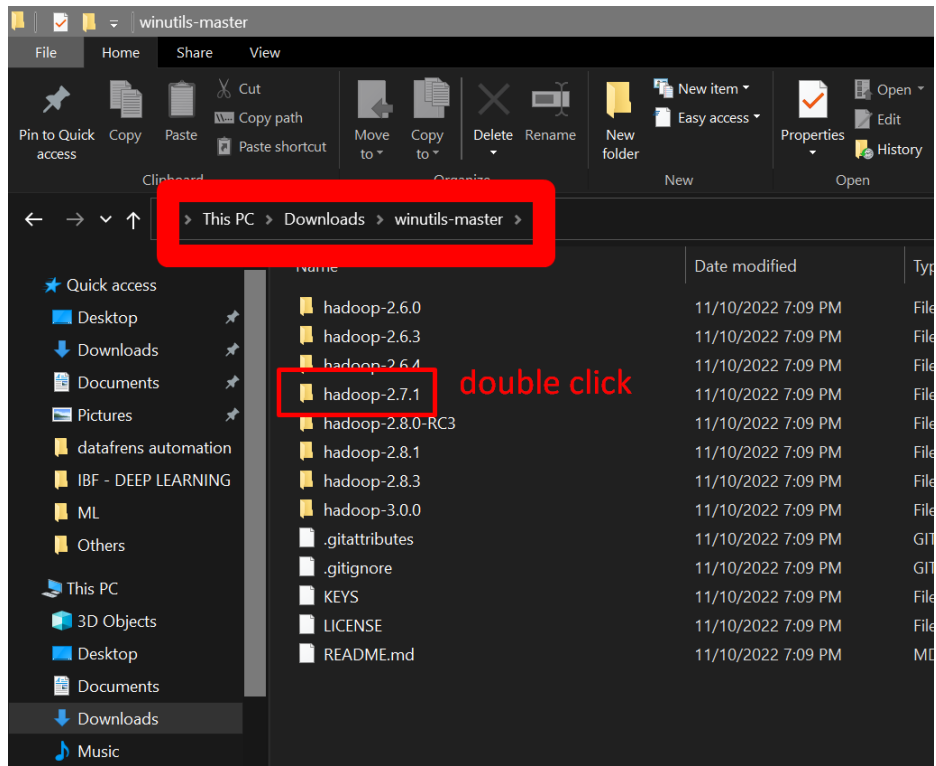
At the time of writing, I know that only JAVA 8 or 11 can work

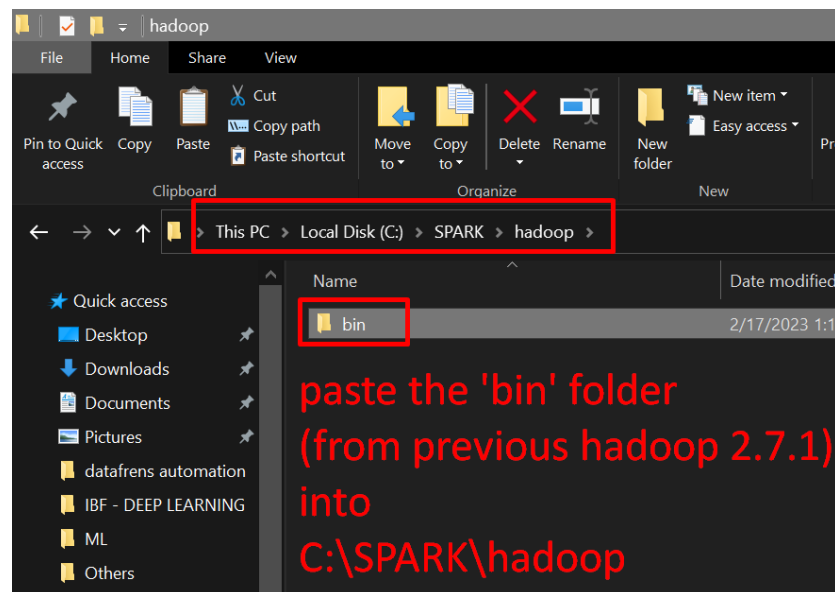
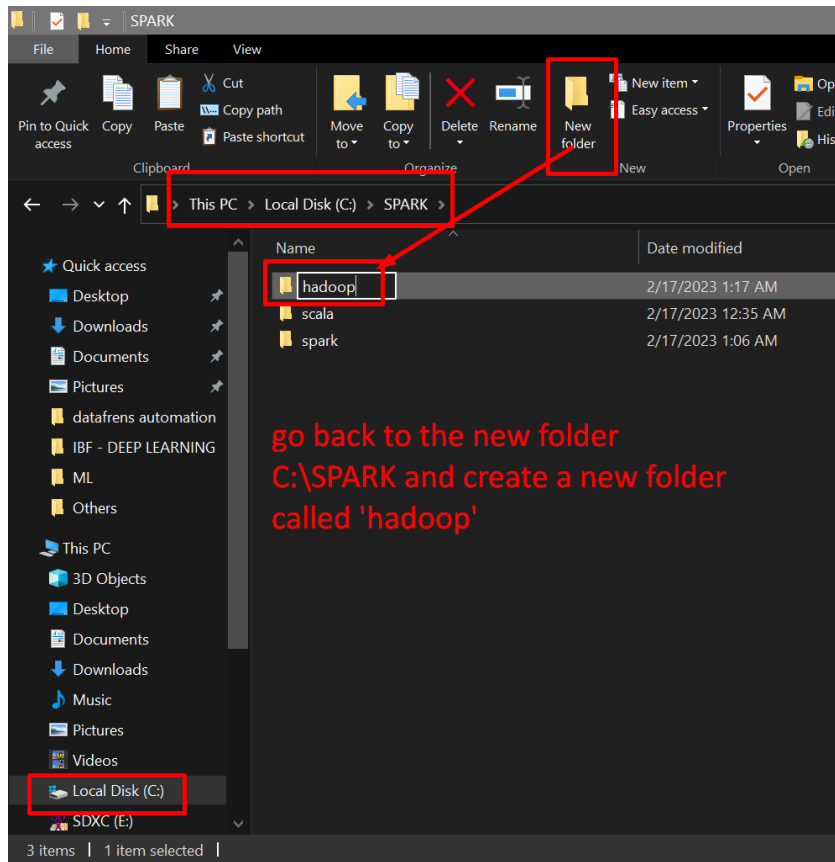
I read that JAVA 9 gives problems and should not be used.

F. INSTALL HADOOP

<https://github.com/stevloughran/winutils/archive/refs/heads/master.zip>





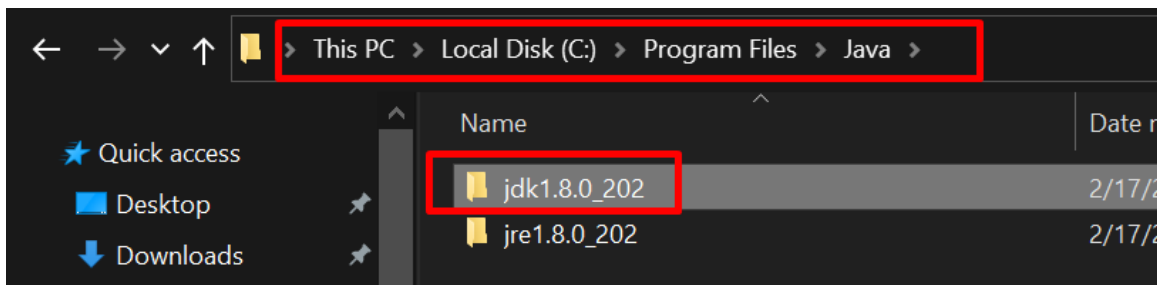
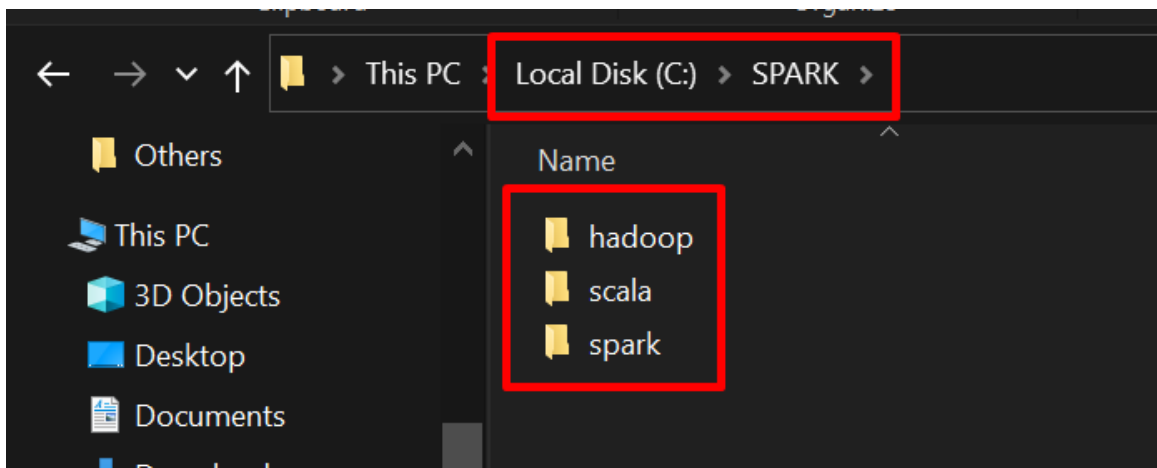


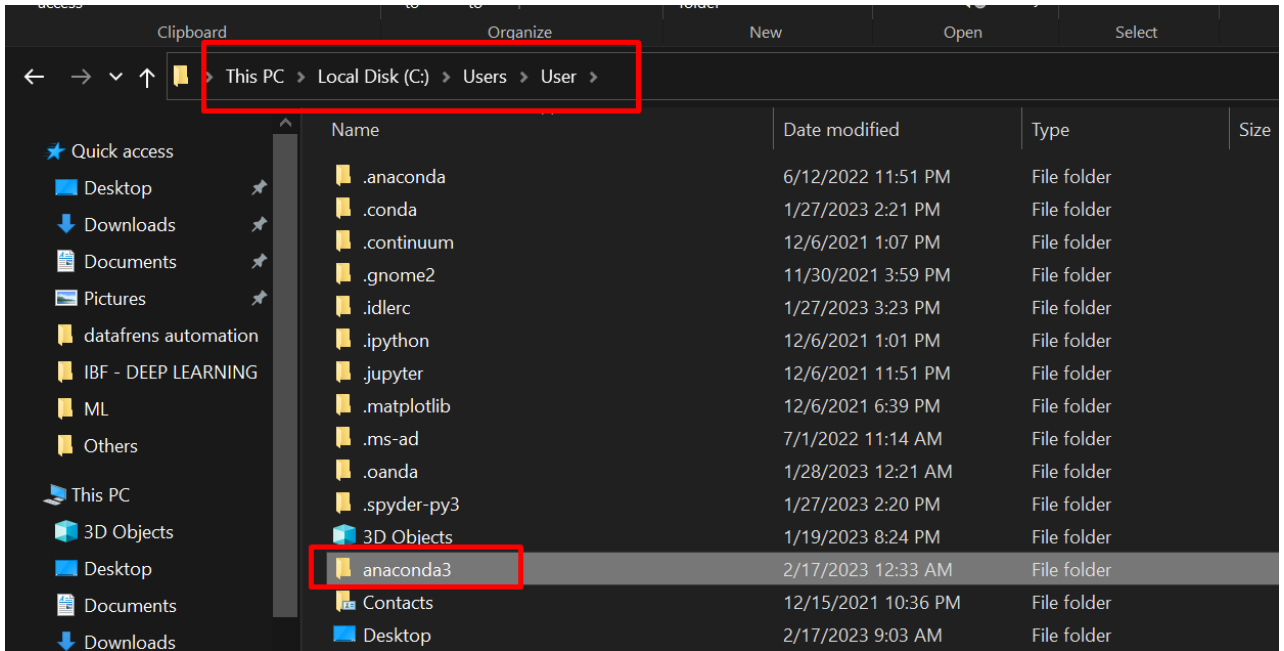
II. CREATING ENVIRONMENT VARIABLES

In a nutshell... we are trying to create all these...

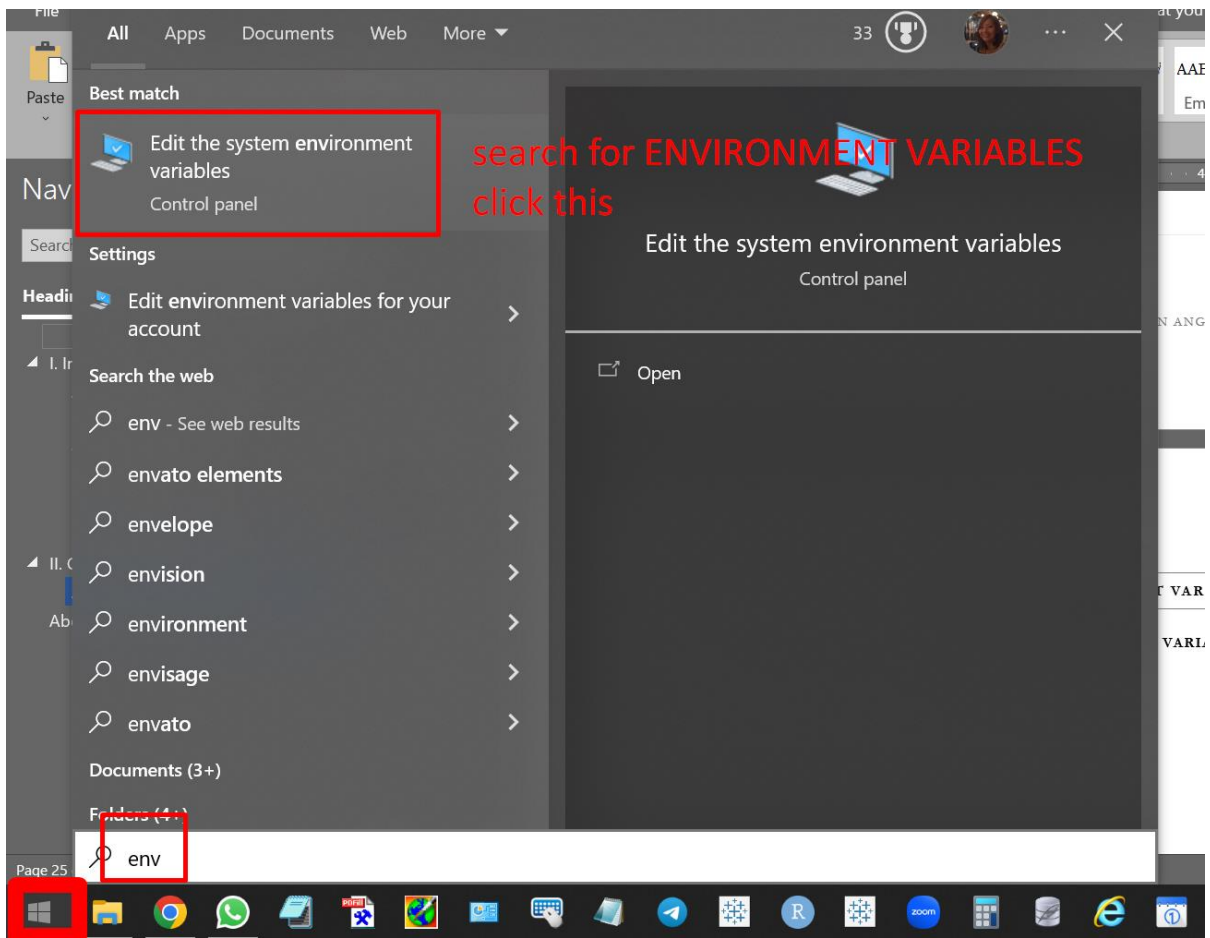
. Environment Variables (windows)

- HADOOP_HOME = C:\SPARK\hadoop
- JAVA_HOME = C:\Program Files\Java\jdk1.8.0_202
- SCALA_HOME = C:\SPARK\scala
- SPARK_HOME = C:\SPARK\spark
- PYSARK_PYTHON = C:\Users\user\anaconda3\python.exe

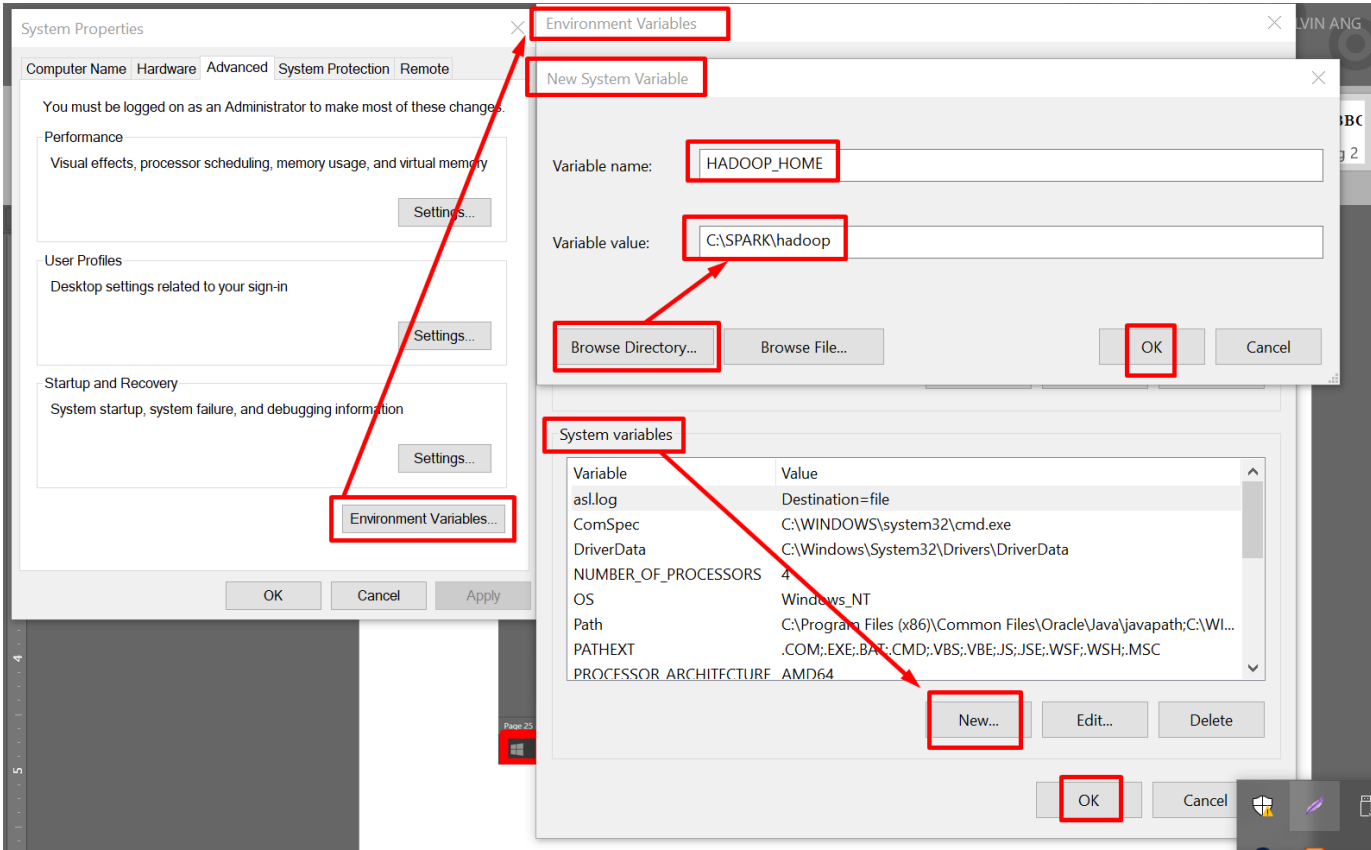




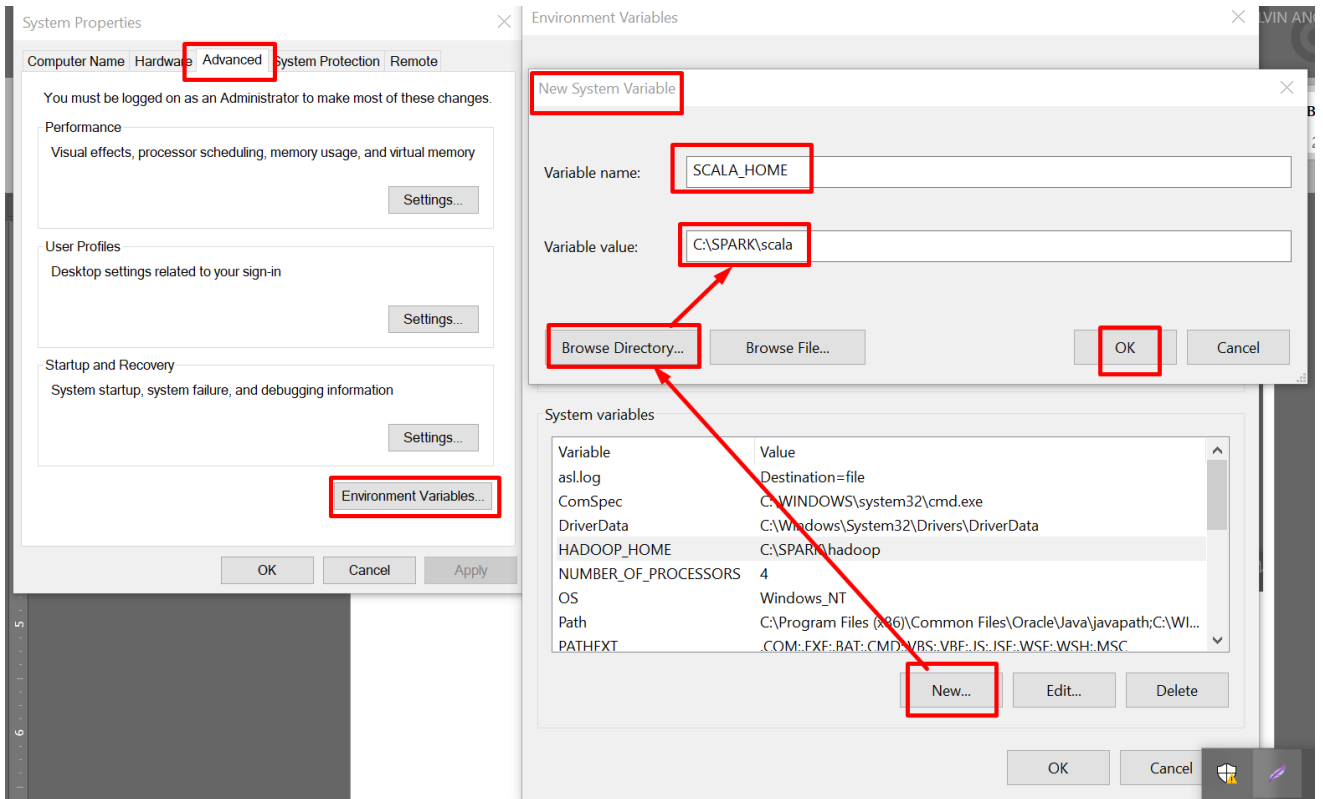
A. CREATING SYSTEM VARIABLES



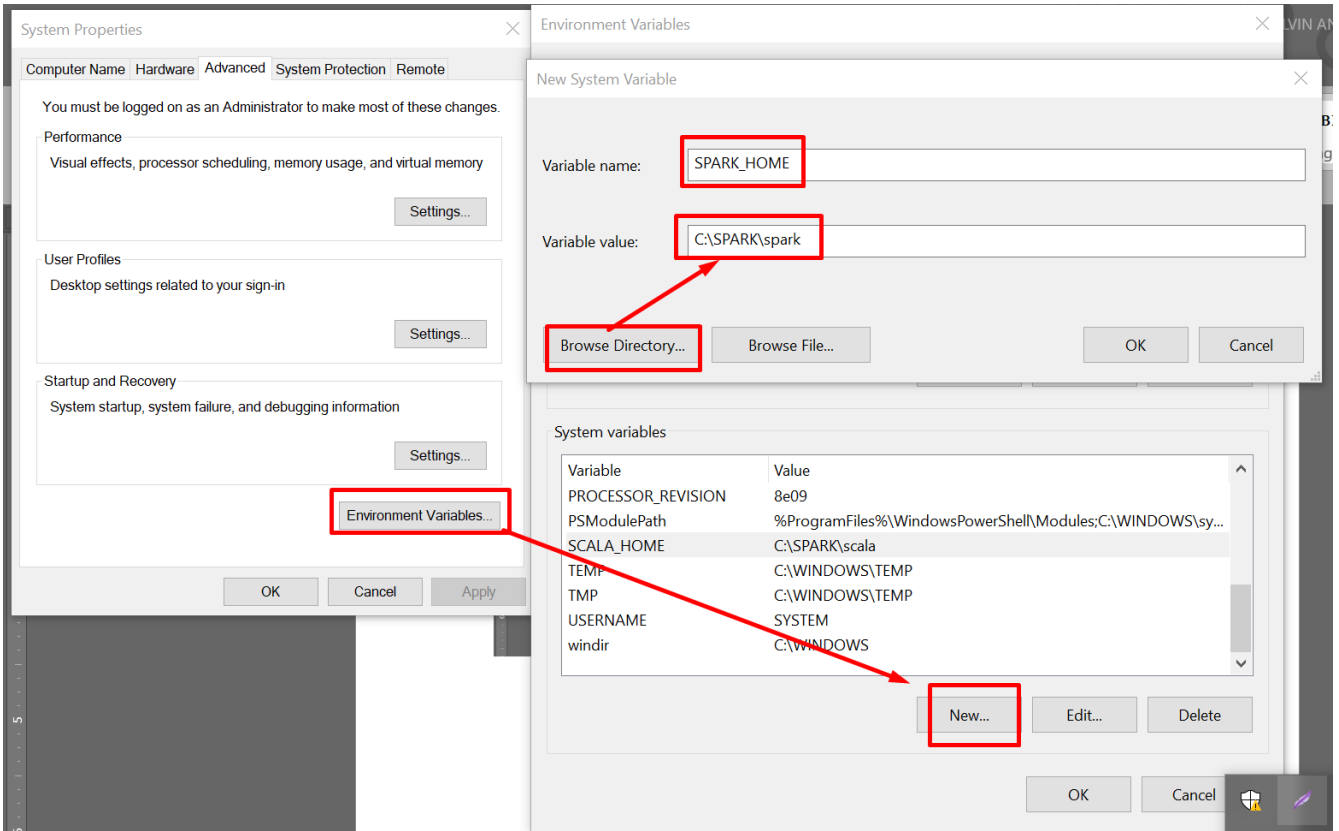
1. CREATE HADOOP_HOME VARIABLE



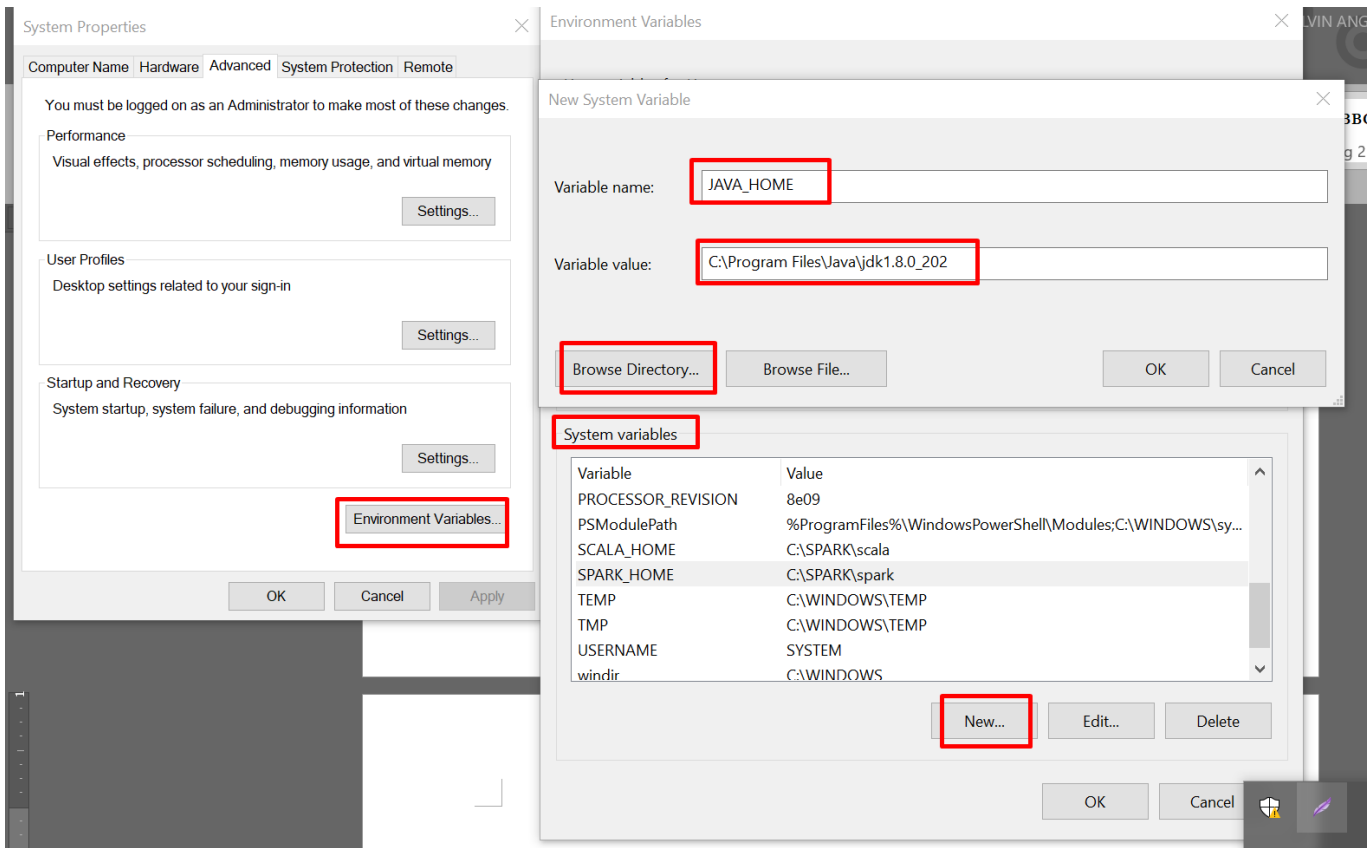
2. CREATE SCALA_HOME VARIABLE



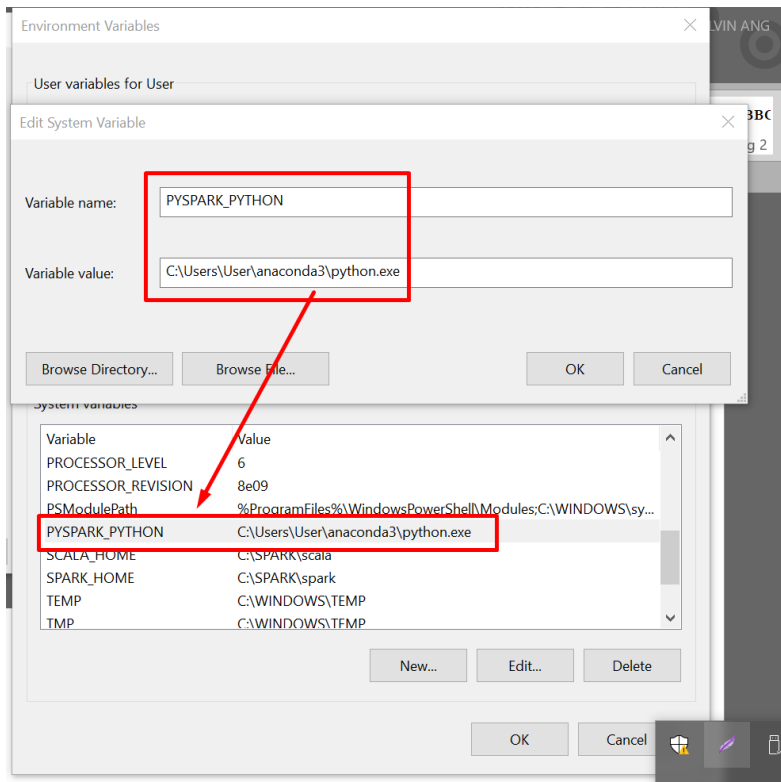
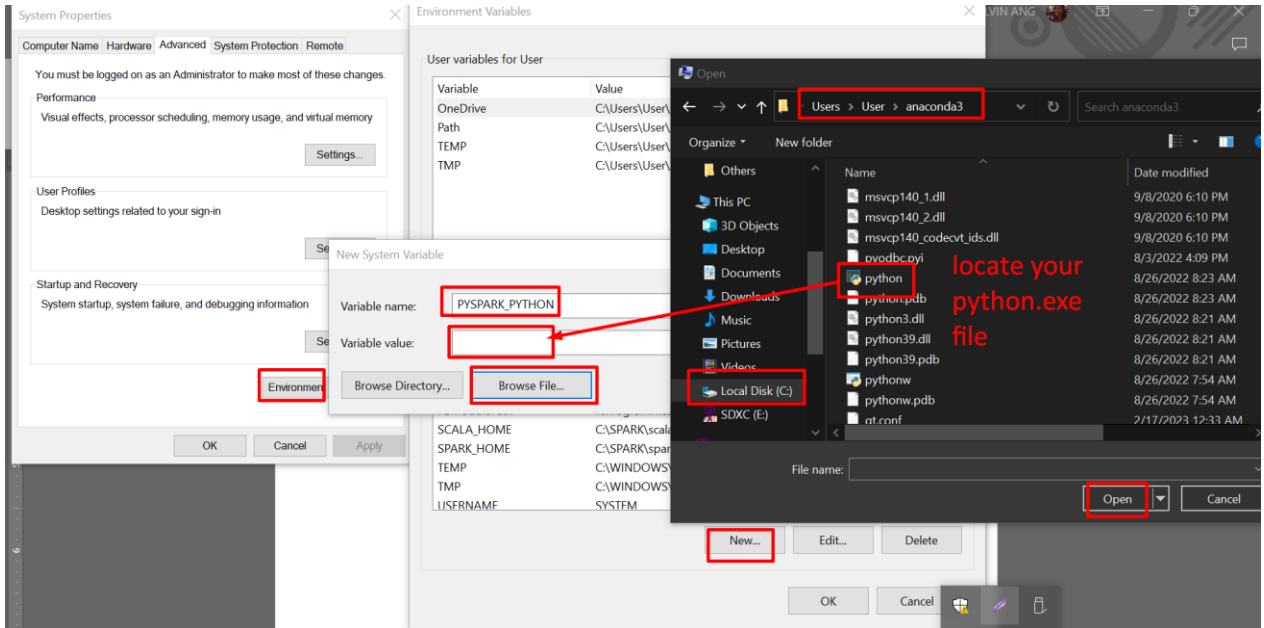
3. CREATE SPARK_HOME VARIABLE



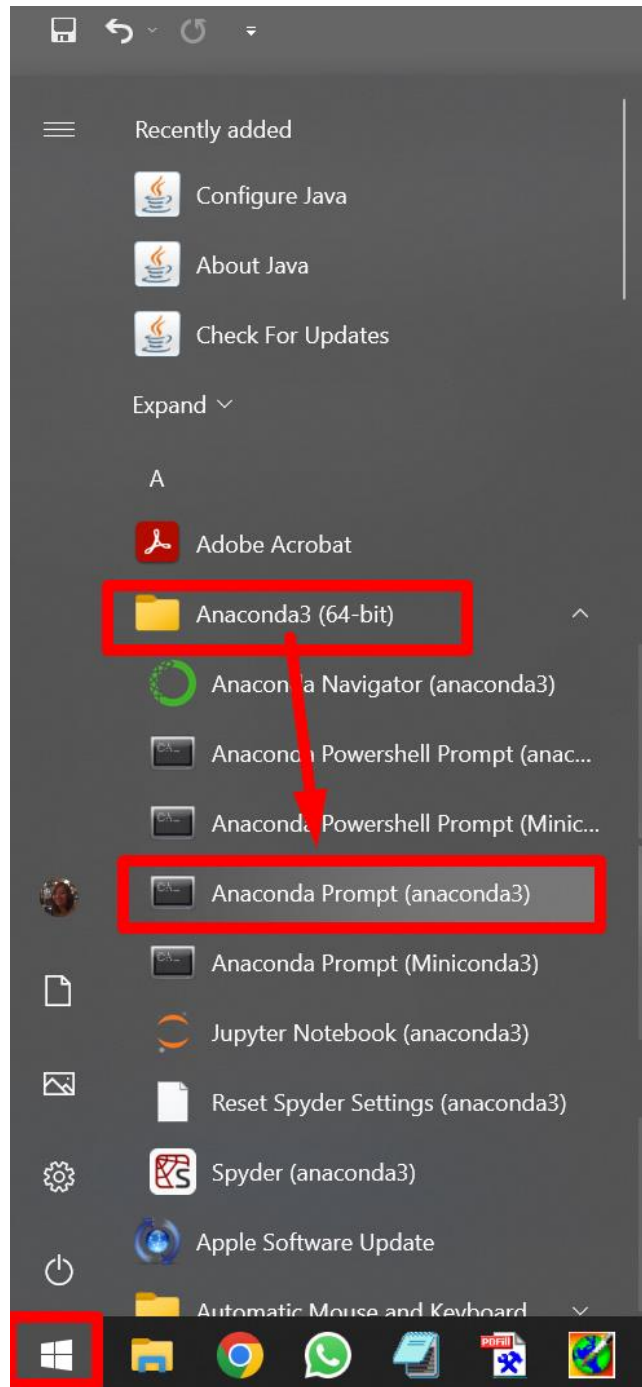
4. CREATE JAVA_HOME VARIABLE

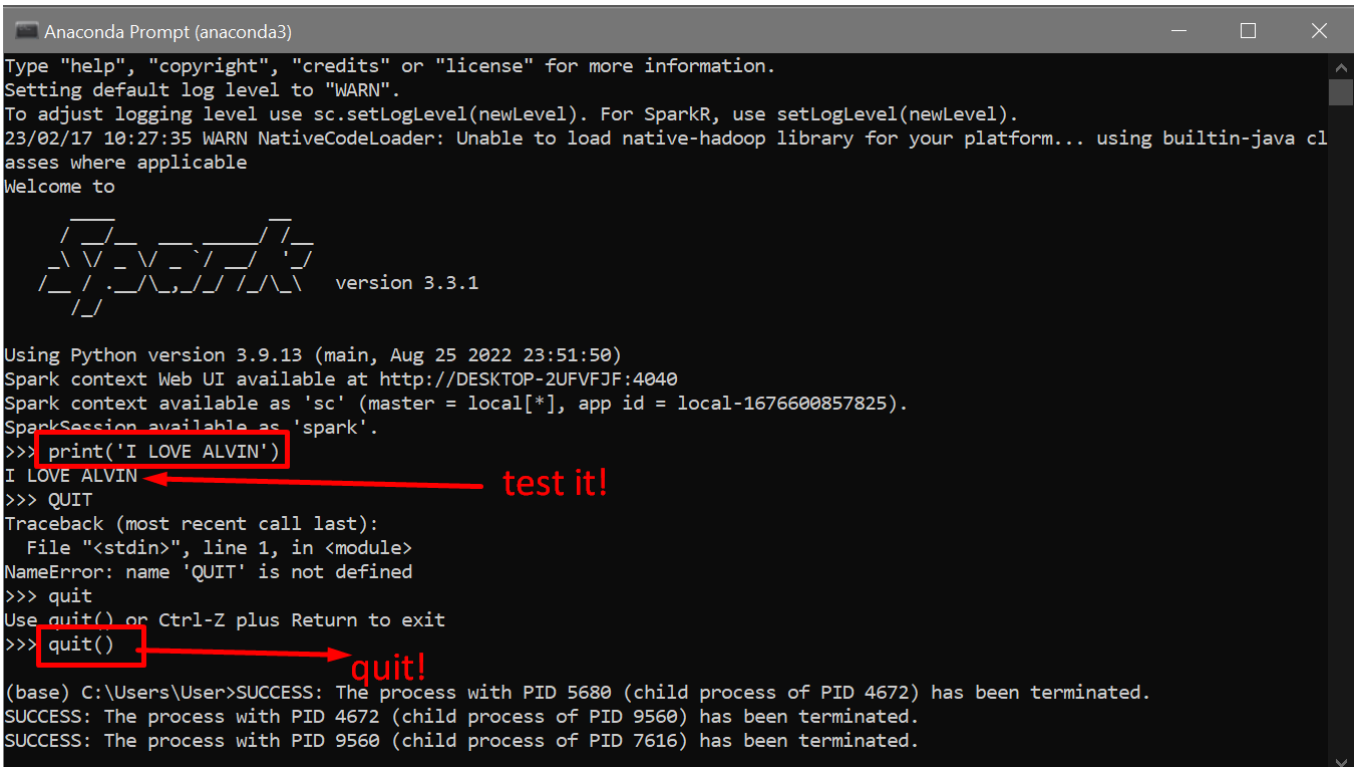
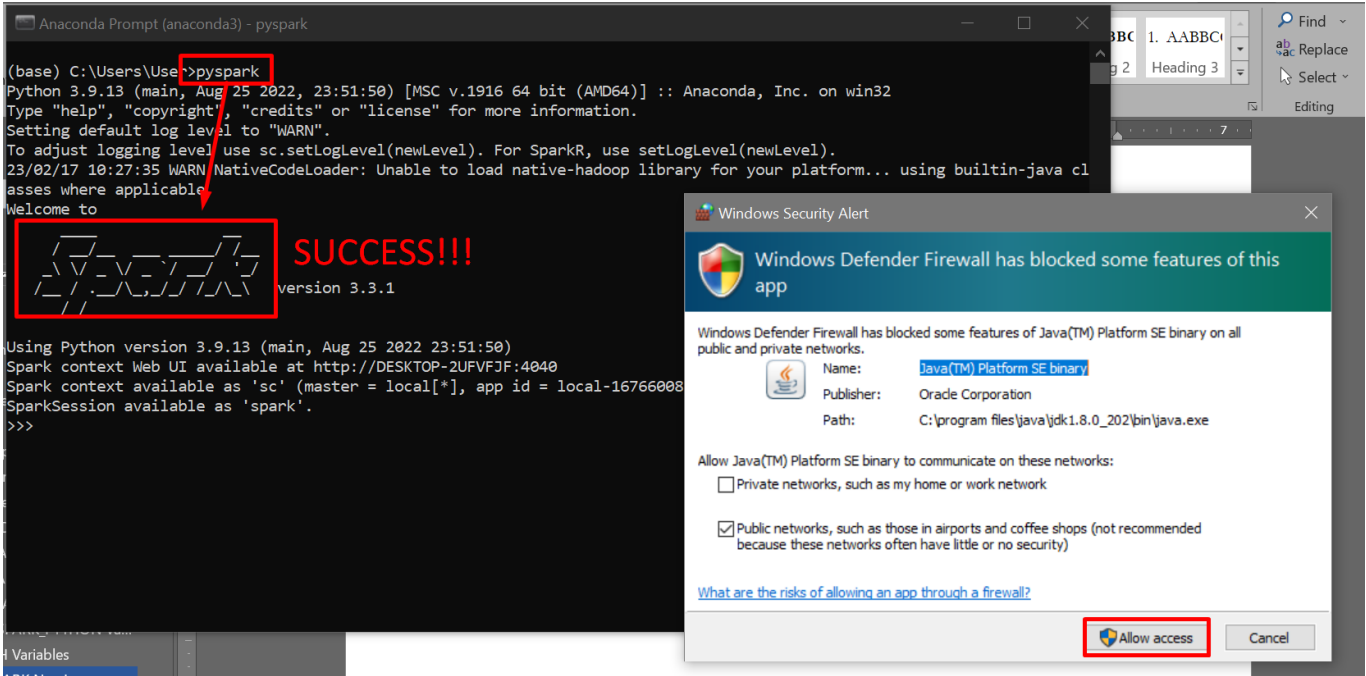


5. CREATE PYSPARK_PYTHON VARIABLE

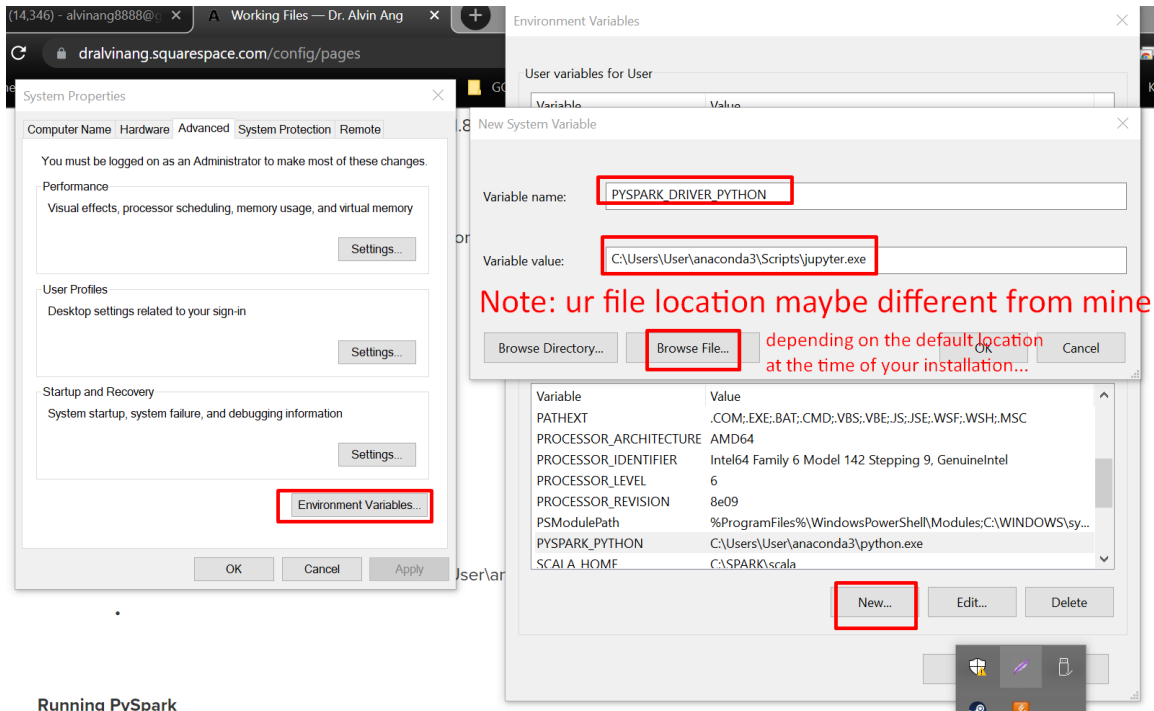
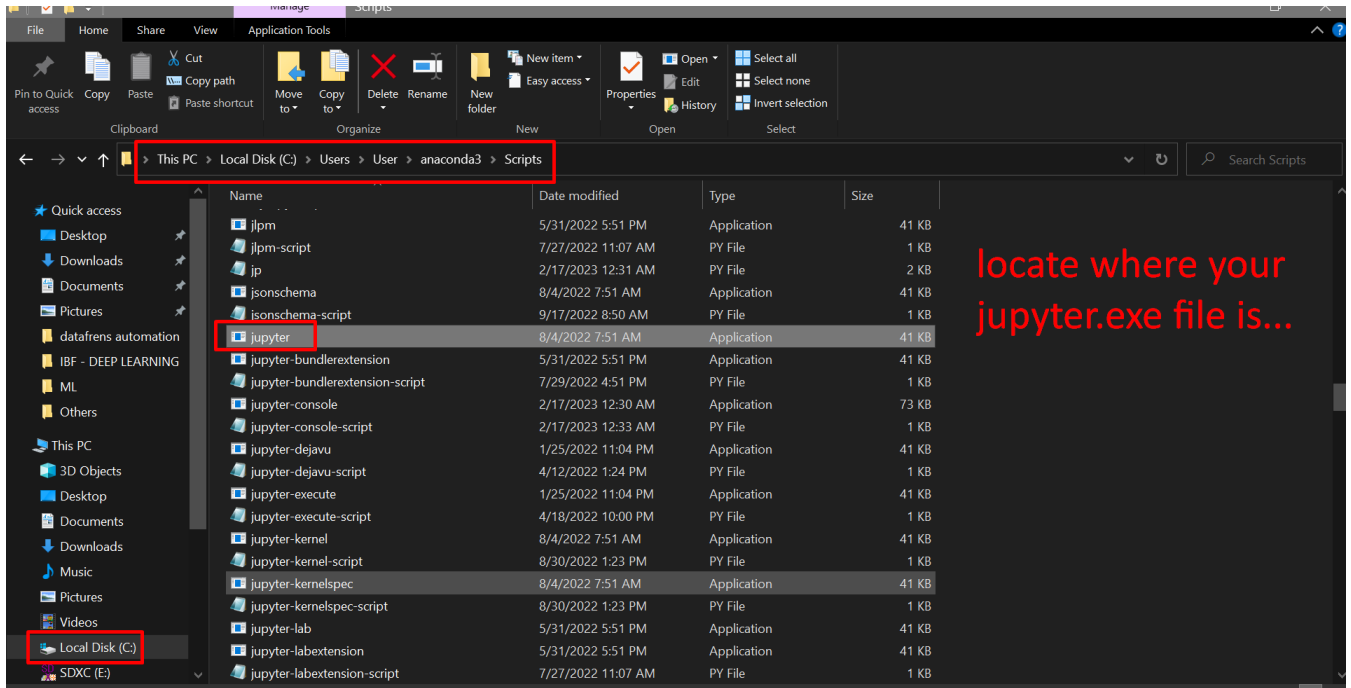


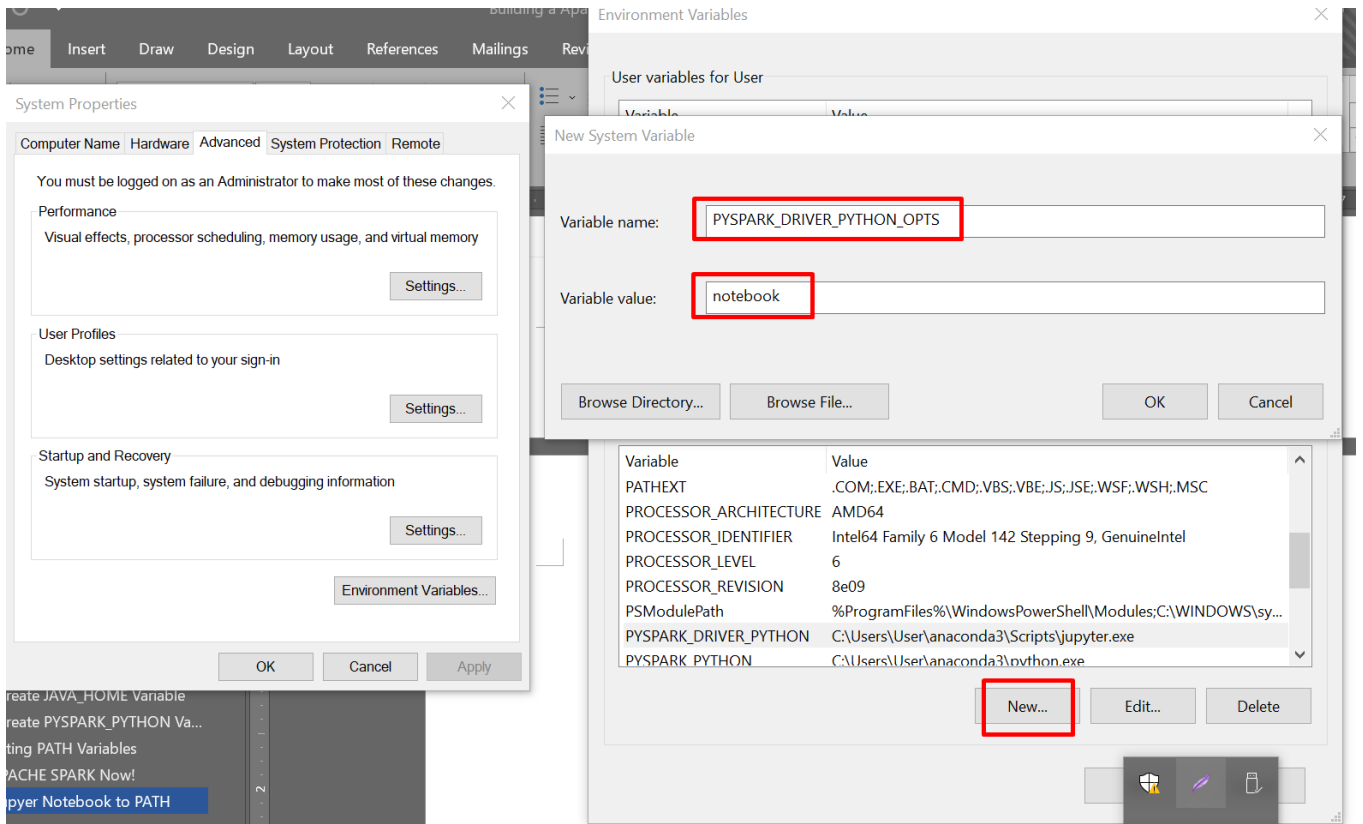
III. TEST APACHE SPARK NOW!





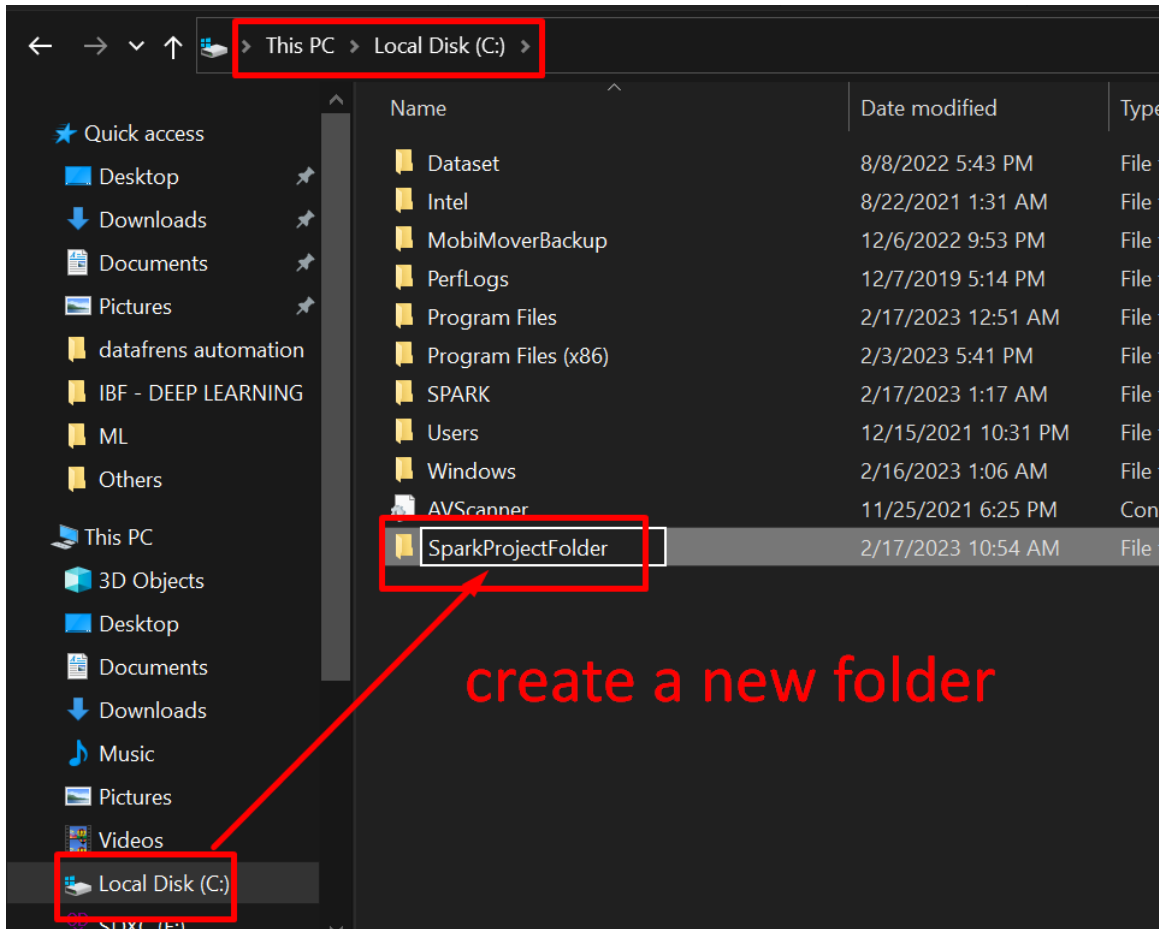
IV. ADD JUPYER NOTEBOOK TO PATH





V. USING JUPYTER NOTEBOOK TO RUN APACHE SPARK

A. CREATE SPARK PROJECT FOLDER IN C:\

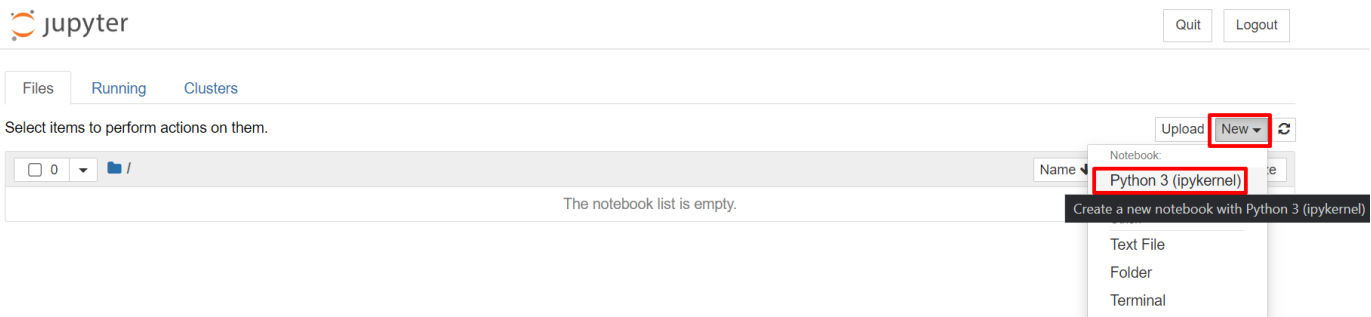



```
Anaconda Prompt (anaconda3) - pyspark
(base) C:\Users\User>cd C:\SparkProjectFolder
(base) C:\SparkProjectFolder>pyspark
[I 2023-02-17 10:56:17.879 LabApp] JupyterLab extension loaded from C:\Users\User\anaconda3\lib\site-packages\jupyterlab
[I 2023-02-17 10:56:17.879 LabApp] JupyterLab application directory is C:\Users\User\anaconda3\share\jupyter\lab
[I 10:56:17.879 NotebookApp] Serving notebooks from local directory: C:\SparkProjectFolder
[I 10:56:17.879 NotebookApp] Jupyter Notebook 6.4.12 is running at:
[I 10:56:17.879 NotebookApp] http://localhost:8888/?token=bc7e31b87109f8b052967c1b285c6c747151bc218fe9d09a
[I 10:56:17.879 NotebookApp] or http://127.0.0.1:8888/?token=bc7e31b87109f8b052967c1b285c6c747151bc218fe9d09a
[I 10:56:17.879 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 10:56:17.926 NotebookApp]

To access the notebook, open this file in a browser:
file:///C:/Users/User/AppData/Roaming/jupyter/runtime/nbserver-9332-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=bc7e31b87109f8b052967c1b285c6c747151bc218fe9d09a
or http://127.0.0.1:8888/?token=bc7e31b87109f8b052967c1b285c6c747151bc218fe9d09a
```

change the working directory to the new folder u created earlier

this command will open Jupyter NB in your browser if it does not, just copy paste this link to your browser and Jupyter will open



B. TEST CODE

```
from pyspark.sql import Row
a = Row(name = 'Anthony', age = 27, height = 168)
print("a: ", a)
```

```
#-----#
```

```
from pyspark import SparkContext, SparkConf
sc = SparkContext.getOrCreate()
sc
```

```
#-----#
```

```
from pyspark.sql import Row
df = sc.parallelize([ \
    Row(name = 'Alvin', age = 21, height = 190),\
    Row(name = 'Alvin', age = 21, height = 190),\
    Row(name = 'Alvin', age = 21, height = 190)]).toDF()

df.show()
```

The screenshot shows a Jupyter Notebook interface with the following elements:

- Header:** "jupyter FIRSTTRY" with a red box around "FIRSTTRY" and a red arrow pointing to it with the word "RENAME" written in red.
- Menu:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Toolbar:** Includes icons for file operations, a "Run" button, and a "Code" dropdown menu.
- Code Cell [2]:** Contains the first code block from the text above. The output shows: `a: Row(name='Anthony', age=27, height=168)`.
- Code Cell [4]:** Contains the second code block. The output is a `SparkContext` object with the following details in a red-bordered box:
 - Spark UI
 - Version: v3.3.1
 - Master: local[*]
 - AppName: PySparkShell
- Code Cell [5]:** Contains the third code block. The output is a table showing three rows of data: `| Alvin | 21 | 190 |`.

VI. SETUP MASTER COMPUTER

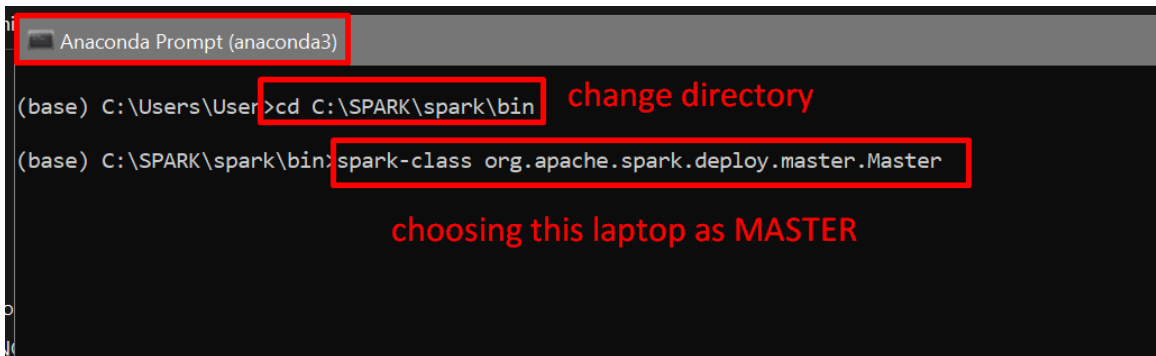
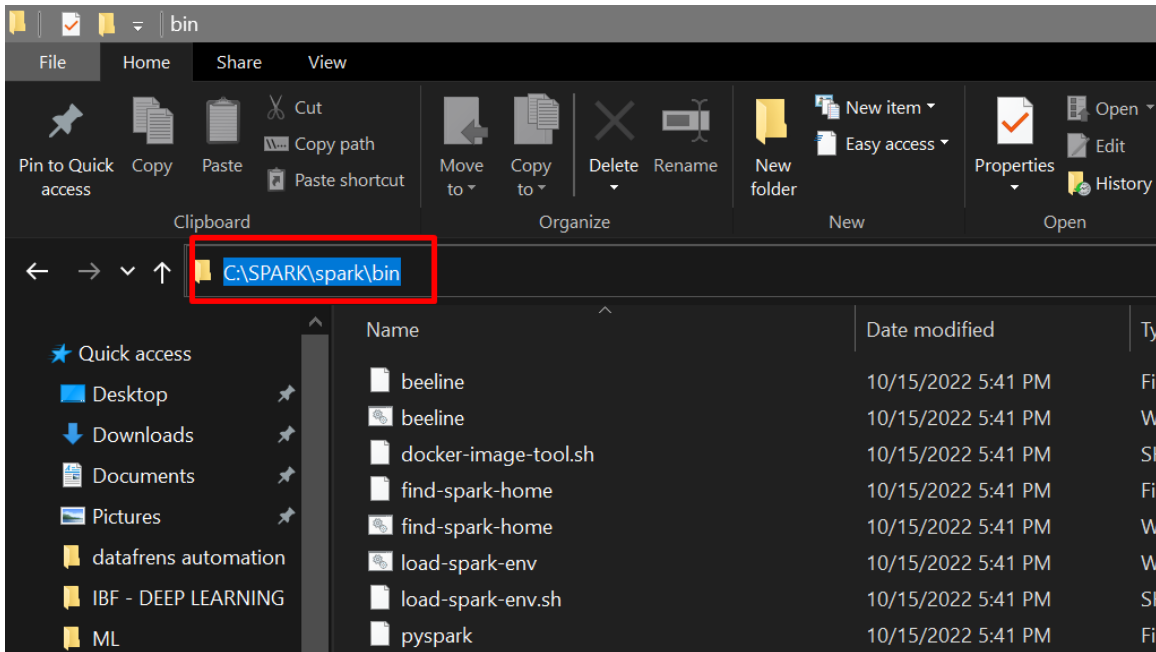
A. STEP 1

- Ensure that ALL Laptops have APACHE SPARK (and all Environment Variables) installed properly.
- In other words, ENSURE that you have followed through all previous steps in previous sections... including installing Java / Jupyter NB / Anaconda etc.... for ALL laptops.

B. STEP 2

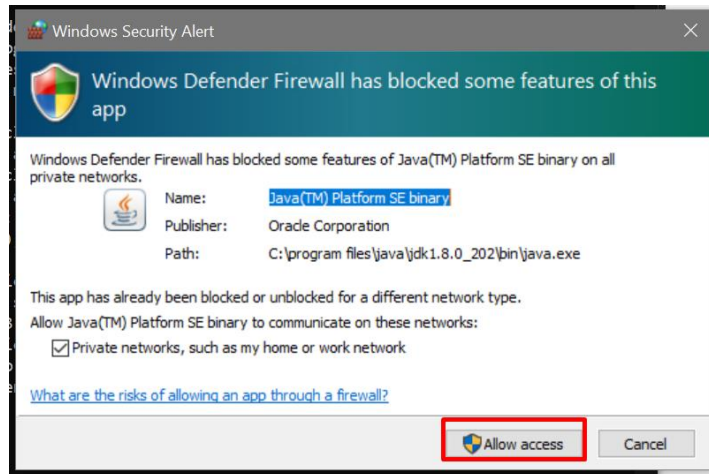
- Decide which laptops you want to be MASTER... which ones you want to be SLAVE

C. STEP 3: APPOINT YOUR MASTER LAPTOP



1. CODE

spark-class org.apache.spark.deploy.master.Master

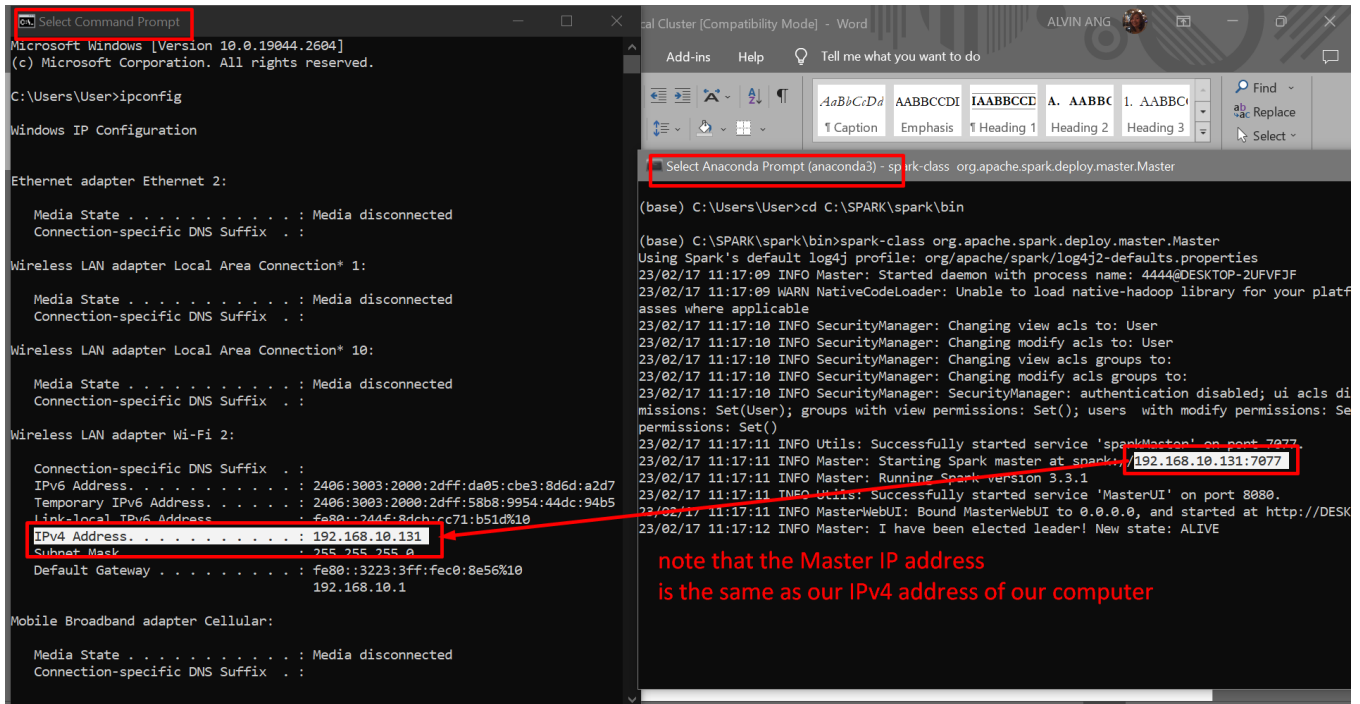
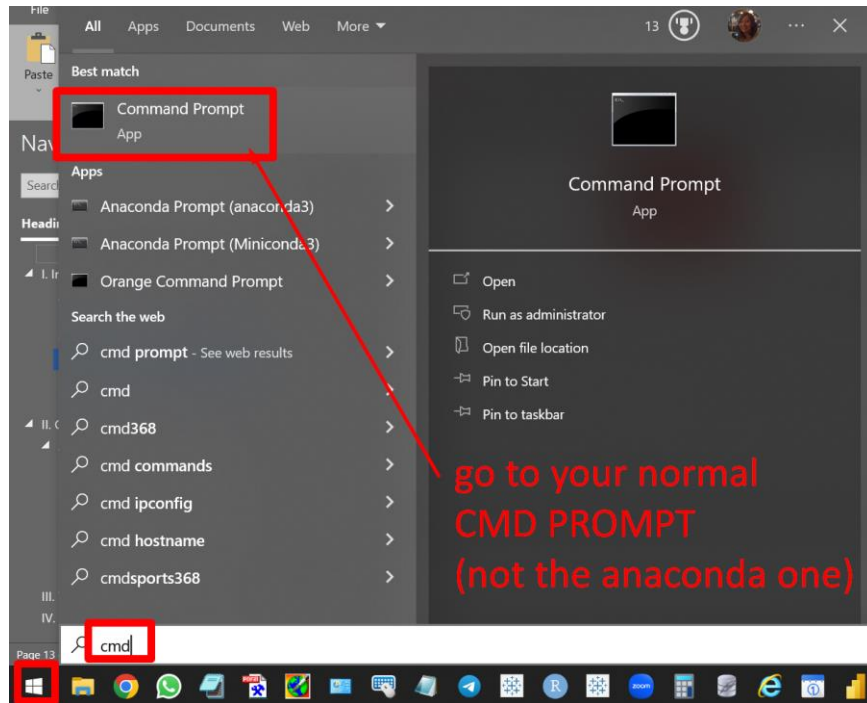


```
Anaconda Prompt (anaconda3) - spark-class org.apache.spark.deploy.master.Master
(base) C:\Users\User>cd C:\SPARK\spark\bin

(base) C:\SPARK\spark\bin>spark-class org.apache.spark.deploy.master.Master
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/02/17 11:17:09 INFO Master: Started daemon with process name: 4444@DESKTOP-2UFVFJF
23/02/17 11:17:09 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
23/02/17 11:17:10 INFO SecurityManager: Changing view acls to: User
23/02/17 11:17:10 INFO SecurityManager: Changing modify acls to: User
23/02/17 11:17:10 INFO SecurityManager: Changing view acls groups to:
23/02/17 11:17:10 INFO SecurityManager: Changing modify acls groups to:
23/02/17 11:17:10 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(User); groups with view permissions: Set(); users with modify permissions: Set(User); groups with modify permissions: Set()
23/02/17 11:17:11 INFO Utils: Successfully started service 'sparkMaster' on port 7077.
23/02/17 11:17:11 INFO Master: Starting Spark master at spark://192.168.10.131:7077
23/02/17 11:17:11 INFO Master: Running Spark version 3.3.1
23/02/17 11:17:11 INFO Utils: Successfully started service 'MasterUI' on port 8080.
23/02/17 11:17:11 INFO MasterWebUI: Bound MasterWebUI to 0.0.0.0, and started at http://DESKTOP-2UFVFJF:8080
23/02/17 11:17:12 INFO Master: I have been elected leader! New state: ALIVE
```

Master laptop ip address

Master Laptop has been appointed



D. CHECK TO SEE IF THERE ARE ANY SLAVES....

URL: spark://192.168.10.131:7077
Alive Workers: 0
Cores in use: 0 Total, 0 Used
Memory in use: 0.0 B Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (0) u can monitor and check .. there are currently no SLAVES yet!!!

Worker Id	Address	State	Cores	Memory	Resources
-----------	---------	-------	-------	--------	-----------

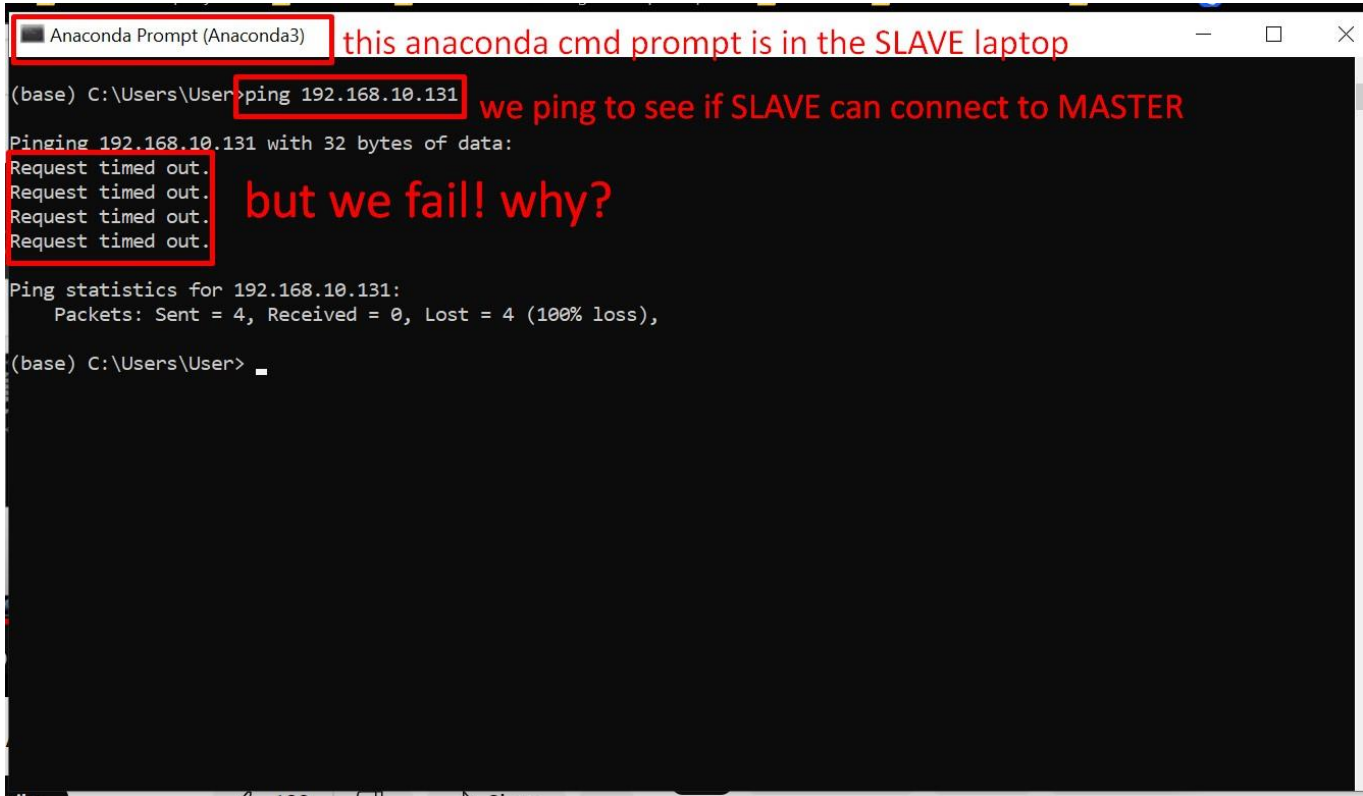
Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Now...go to your SLAVE computer....



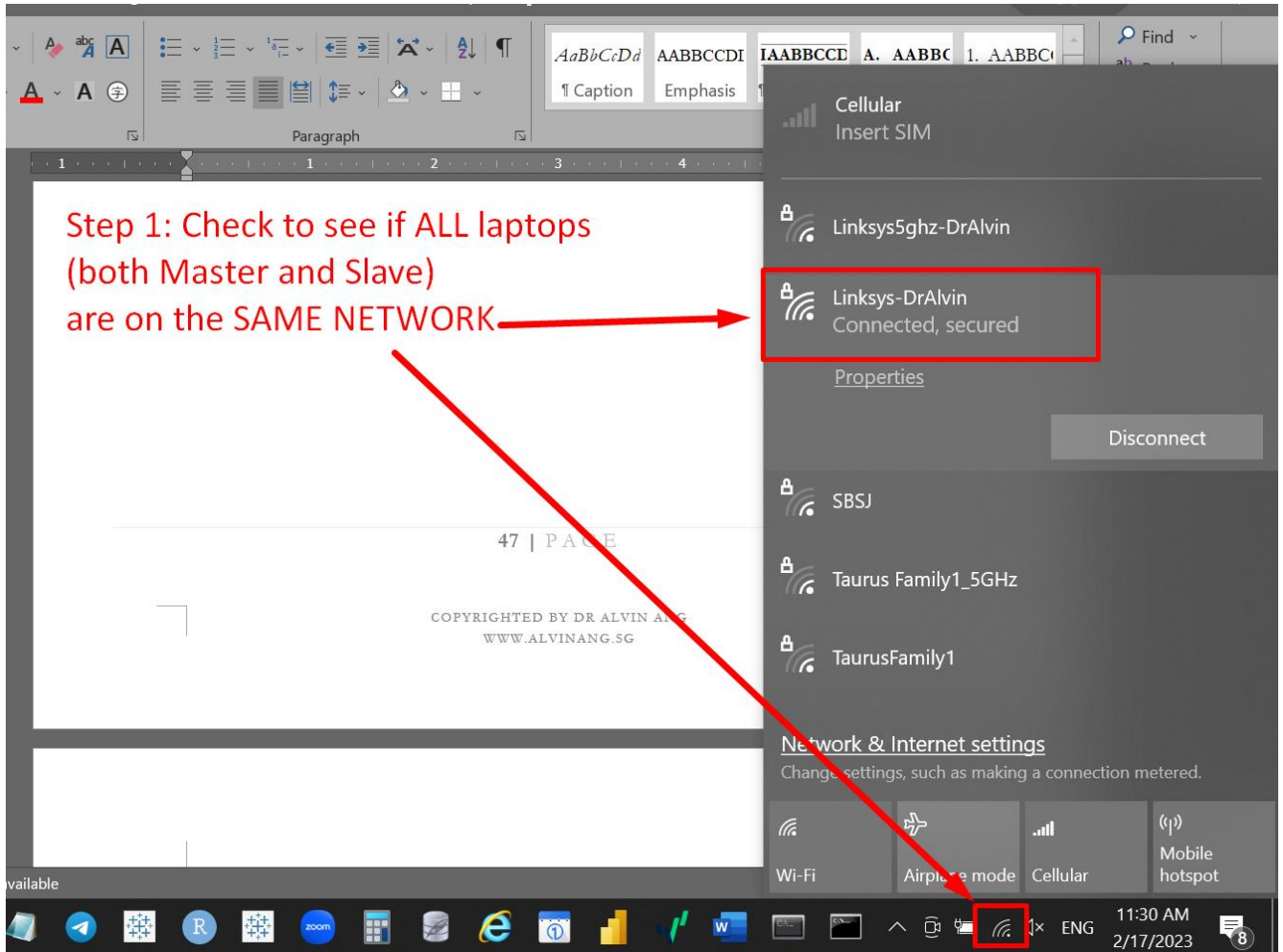
The screenshot shows an Anaconda Prompt window with the following text:

```
Anaconda Prompt (Anaconda3) this anaconda cmd prompt is in the SLAVE laptop
(base) C:\Users\User>ping 192.168.10.131 we ping to see if SLAVE can connect to MASTER
Pinging 192.168.10.131 with 32 bytes of data:
Request timed out.
Request timed out.
Request timed out.
Request timed out.
Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
(base) C:\Users\User>
```

Red annotations in the image include:

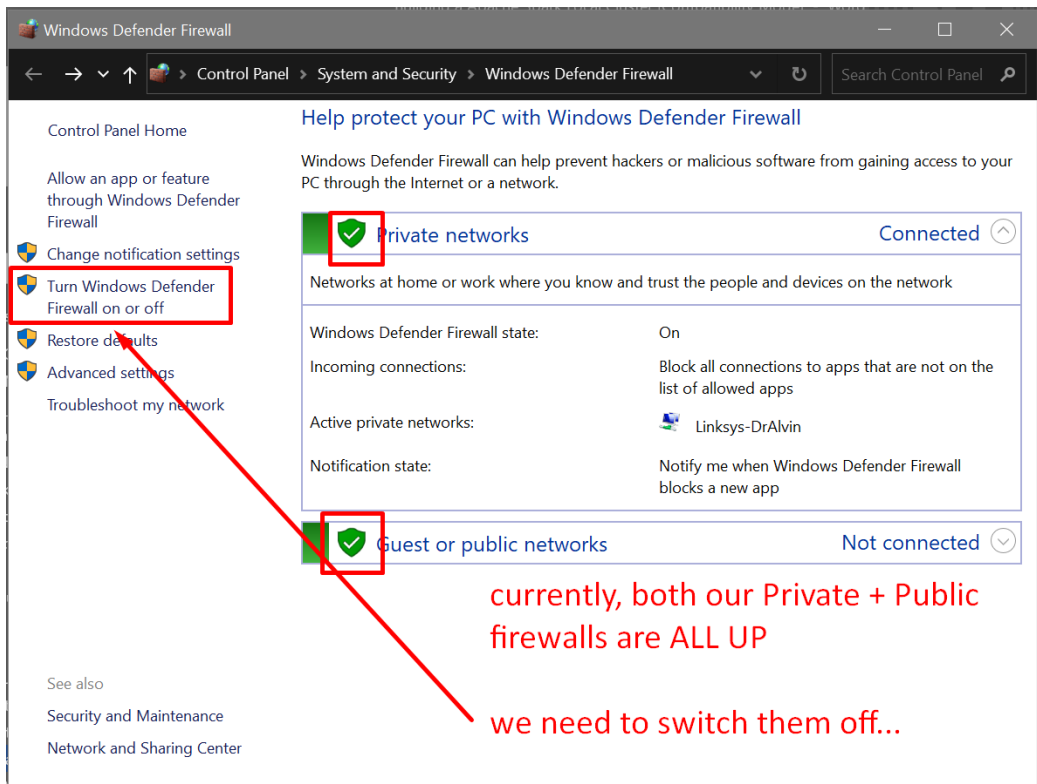
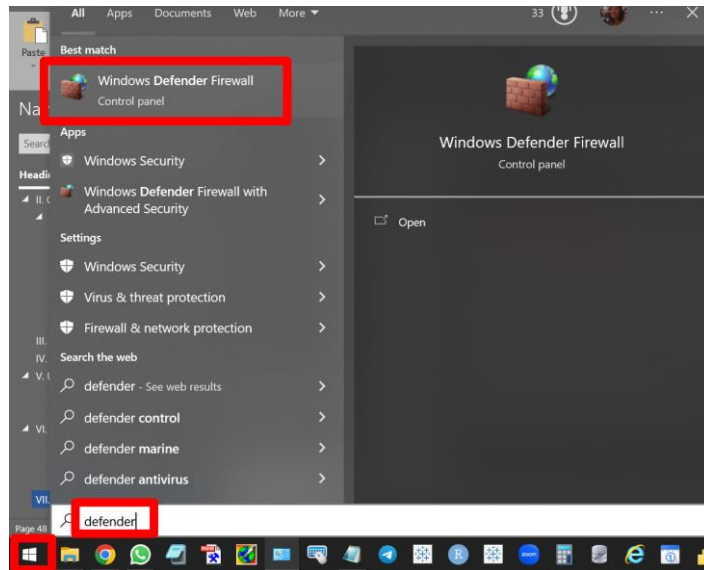
- A red box around the window title "Anaconda Prompt (Anaconda3)".
- Red text "this anaconda cmd prompt is in the SLAVE laptop" pointing to the title bar.
- A red box around the command "ping 192.168.10.131".
- Red text "we ping to see if SLAVE can connect to MASTER" pointing to the command.
- A red box around the four "Request timed out." messages.
- Red text "but we fail! why?" pointing to the "Request timed out." messages.

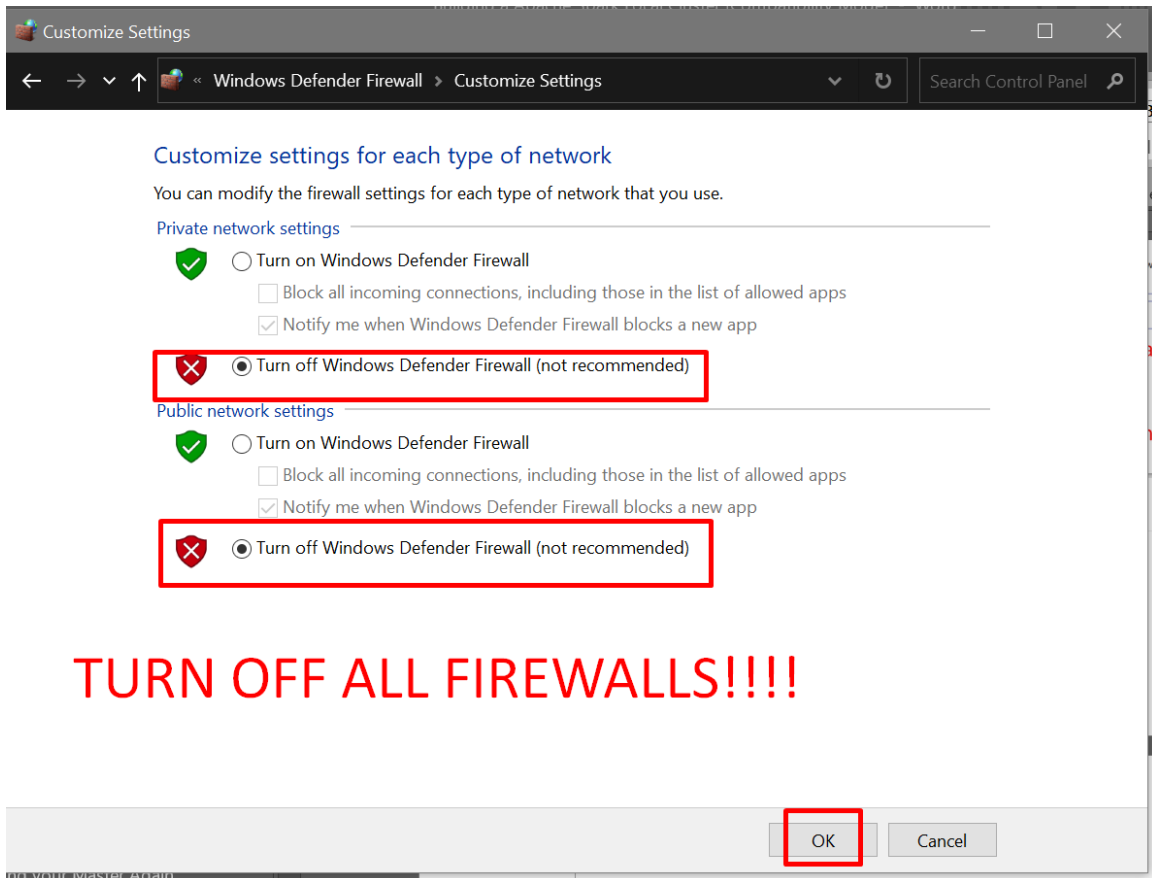
A. CHECK YOUR WIFI NETWORK (BOTH MASTER AND SLAVE)



B. TURN OFF ALL YOUR FIREWALLS (MASTER ONLY)

Now...go to your MASTER computer....





C. NOW FROM SLAVE, PING YOUR MASTER AGAIN....

Now...go to your SLAVE computer....

```

■ Anaconda Prompt (Anaconda3)
Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),

(base) C:\Users\User>ping 192.168.10.131

Pinging 192.168.10.131 with 32 bytes of data:
Request timed out.
Request timed out.
Request timed out.
Request timed out.

Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),

(base) C:\Users\User>ping 192.168.10.131

Pinging 192.168.10.131 with 32 bytes of data:
Reply from 192.168.10.131: bytes=32 time=11ms TTL=128
Reply from 192.168.10.131: bytes=32 time=4ms TTL=128
Request timed out.
Reply from 192.168.10.131: bytes=32 time=3ms TTL=128

Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 3, Lost = 1 (25% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 3ms, Maximum = 11ms, Average = 6ms

(base) C:\Users\User>ping 192.168.10.131

Pinging 192.168.10.131 with 32 bytes of data:
Reply from 192.168.10.131: bytes=32 time=2ms TTL=128
Reply from 192.168.10.131: bytes=32 time=5ms TTL=128
Reply from 192.168.10.131: bytes=32 time=3ms TTL=128
Reply from 192.168.10.131: bytes=32 time=17ms TTL=128

Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 2ms, Maximum = 17ms, Average = 6ms

(base) C:\Users\User>

```

go back to you SLAVE and try to PING again....

if all can't, try restarting both Master + Slave laptops.....

if 1 packet failed.. is still ok.... it means is already connected...

u may try pingging again.. this time all should go thru....

VIII. LOAD / DEPLOY YOUR SLAVE (OR WORKER) INTO THE CLUSTER

Now...go to your SLAVE computer....

```
Anaconda Prompt (Anaconda3) - spark-class org.apache.spark.deploy.worker.Worker spark://192.168.10.131:7077
(base) C:\Users\User>ping 192.168.10.131

Pinging 192.168.10.131 with 32 bytes of data:
Reply from 192.168.10.131: bytes=32 time=2ms TTL=128
Reply from 192.168.10.131: bytes=32 time=5ms TTL=128
Reply from 192.168.10.131: bytes=32 time=3ms TTL=128
Reply from 192.168.10.131: bytes=32 time=17ms TTL=128

Ping statistics for 192.168.10.131:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 2ms, Maximum = 17ms, Average = 6ms

(base) C:\Users\User>cd SPARK\spark\bin
The system cannot find the path specified.

(base) C:\Users\User>cd C:\SPARK\spark\bin
(base) C:\SPARK\spark\bin>spark-class org.apache.spark.deploy.worker.Worker spark://192.168.10.131:7077
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/02/17 12:14:02 INFO Worker: Started daemon with process name: 4728@DESKTOP-2UFVFJF
23/02/17 12:14:03 INFO SecurityManager: Changing view acls to: User
23/02/17 12:14:03 INFO SecurityManager: Changing modify acls to: User
23/02/17 12:14:03 INFO SecurityManager: Changing view acls groups to:
23/02/17 12:14:03 INFO SecurityManager: Changing modify acls groups to:
23/02/17 12:14:03 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(User); groups with view permissions: S
et(); users with modify permissions: Set(User); groups with modify permissions: Set()
23/02/17 12:14:04 INFO Utils: Successfully started service 'sparkworker' on port 50059.
23/02/17 12:14:04 INFO Worker: Worker decommissioning not enabled.
23/02/17 12:14:05 INFO Worker: Starting Spark worker 192.168.10.129:50059 with 4 cores, 6.9 GIB RAM
23/02/17 12:14:05 INFO Worker: Running Spark version 3.3.1
23/02/17 12:14:05 INFO Worker: Spark home: C:\SPARK\spark
23/02/17 12:14:05 INFO ResourceUtils: =====
23/02/17 12:14:05 INFO ResourceUtils: No custom resources configured for spark.worker.
23/02/17 12:14:05 INFO ResourceUtils: =====
23/02/17 12:14:05 INFO Utils: Successfully started service 'WorkerUI' on port 8081.
23/02/17 12:14:05 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://DESKTOP-2UFVFJF:8081
23/02/17 12:14:05 INFO Worker: Connecting to master 192.168.10.131:7077...
23/02/17 12:14:05 INFO TransportClientFactory: Successfully created connection to /192.168.10.131:7077 after 75 ms (0 ms spent in bootstraps)
23/02/17 12:14:06 INFO Worker: Successfully registered with master spark://192.168.10.131:7077
```

A. CODE

```
spark-class org.apache.spark.deploy.worker.Worker spark://192.168.10.131:7077
```

(your IP address will be different from mine...)

B. CHECK TO SEE IF THERE ARE ANY SLAVES....

Now...go to your MASTER computer....

Spark Master at spark://192.168.10.131:7077

URL: spark://192.168.10.131:7077
Alive Workers: 1
Cores in use: 4 Total, 0 Used
Memory in use: 6.9 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory	Resources
worker-20230217121404-192.168.10.129-50059	192.168.10.129:50059	ALIVE	4 (0 Used)	6.9 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Now...go to your SLAVE computer....

at your SLAVE laptop...
press CTRL + SHIFT + ESC to bring to Task Manager

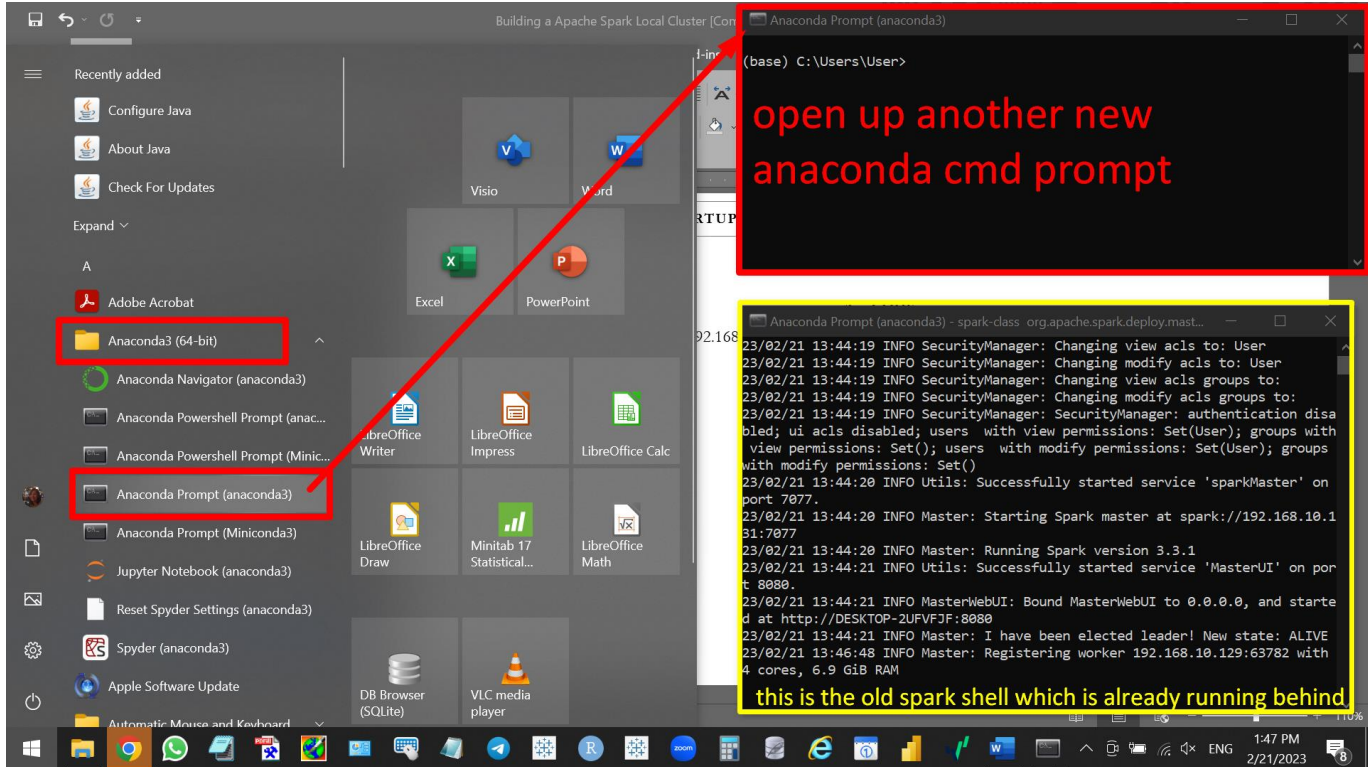
click around to see the no. of cores and Memory u have in your worker

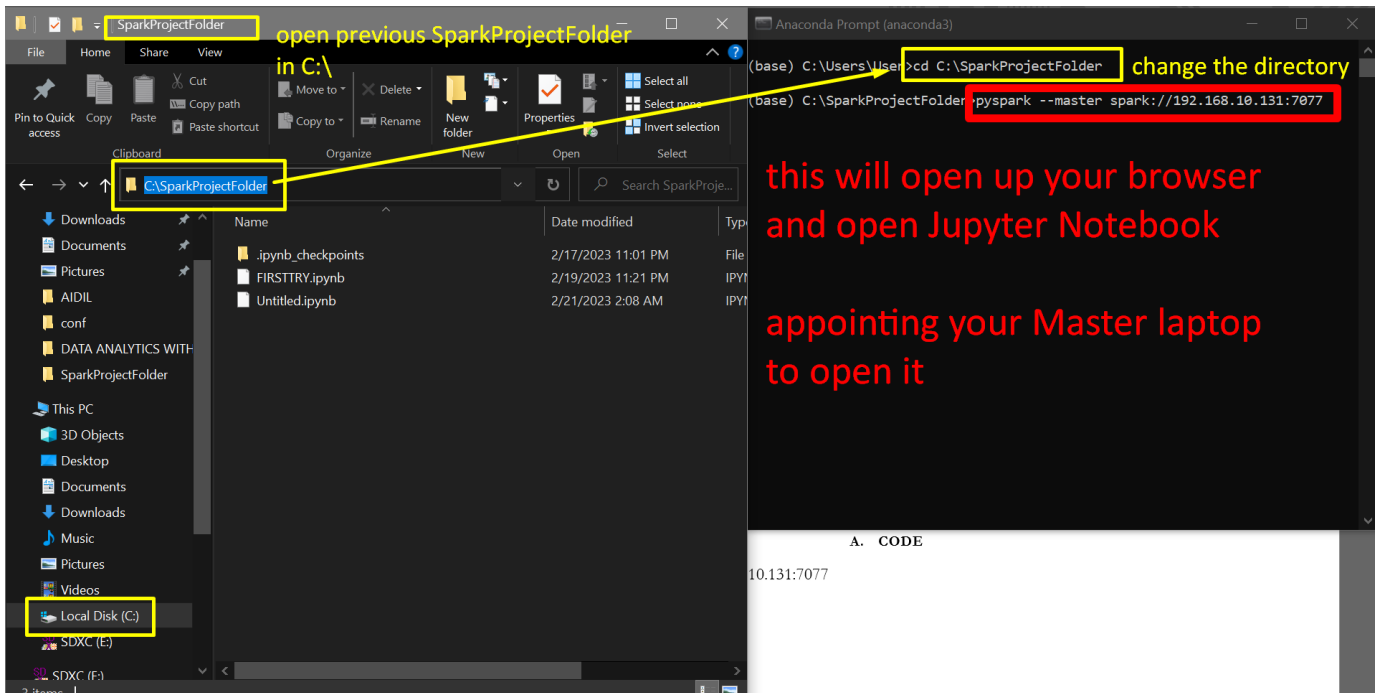
CPU 14% 2.30 GHz
Memory 4.5/7.8 GB (58%)

Utilization	Speed	Base speed:	2.71 GHz
14%	2.30 GHz	Sockets:	1
Processes	Threads	Handles	Logical processors: 4
209	2037	84580	Virtualization: Enabled
Up time	0:09:40:30		
	L1 cache:	128 KB	
	L2 cache:	512 KB	
	L3 cache:	3.0 MB	

IX. START JUPYTER NOTEBOOK IN YOUR MASTER

Now...go to your MASTER computer....





A. CODE

`pyspark --master spark://192.168.10.131:7077`

B. CHECK LOCALHOST

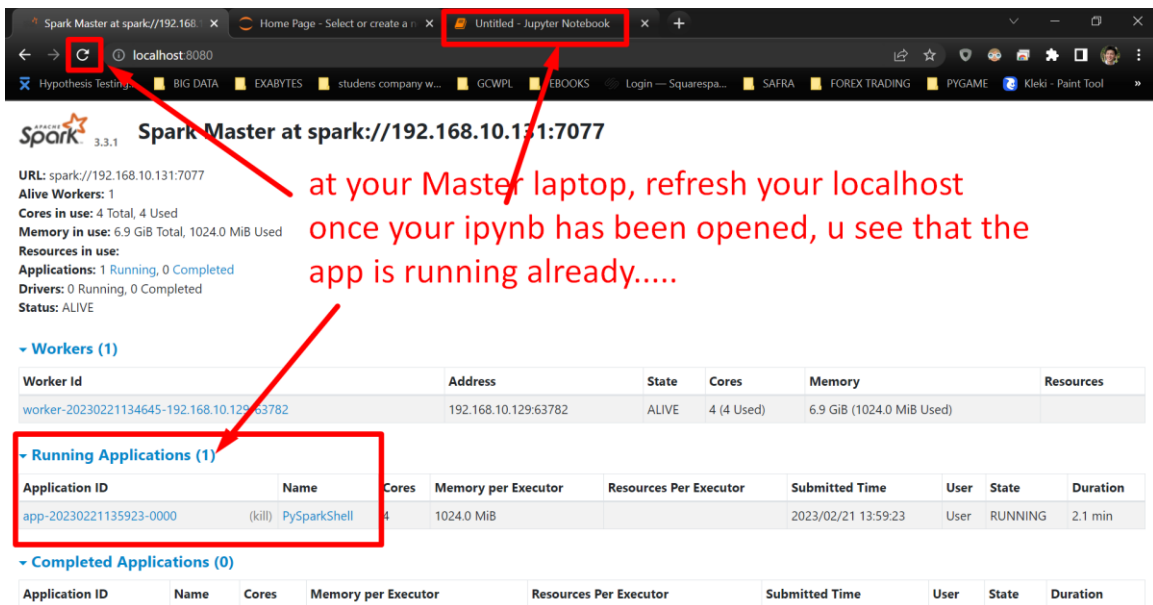
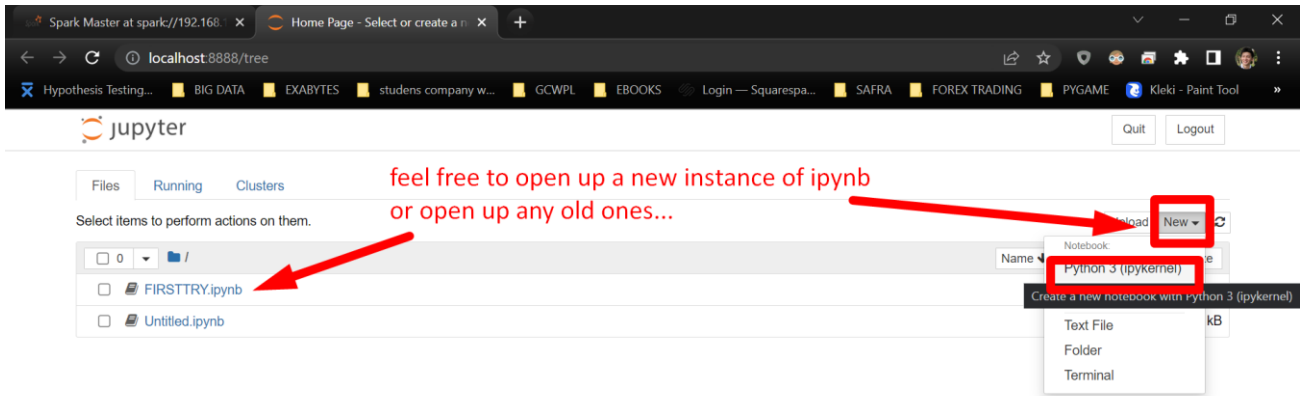
before opening up any ipynb, u may go to localhost:8080
u can see there are no apps running....
(meaning no ipynbs have been open yet...)

Worker Id	Address	State	Cores	Memory	Resources
worker-20230221134645-192.168.10.129	192.168.10.129:63782	ALIVE	4 (0 Used)	6.9 GiB (0.0 B Used)	

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

C. OPEN ANY IPYNB

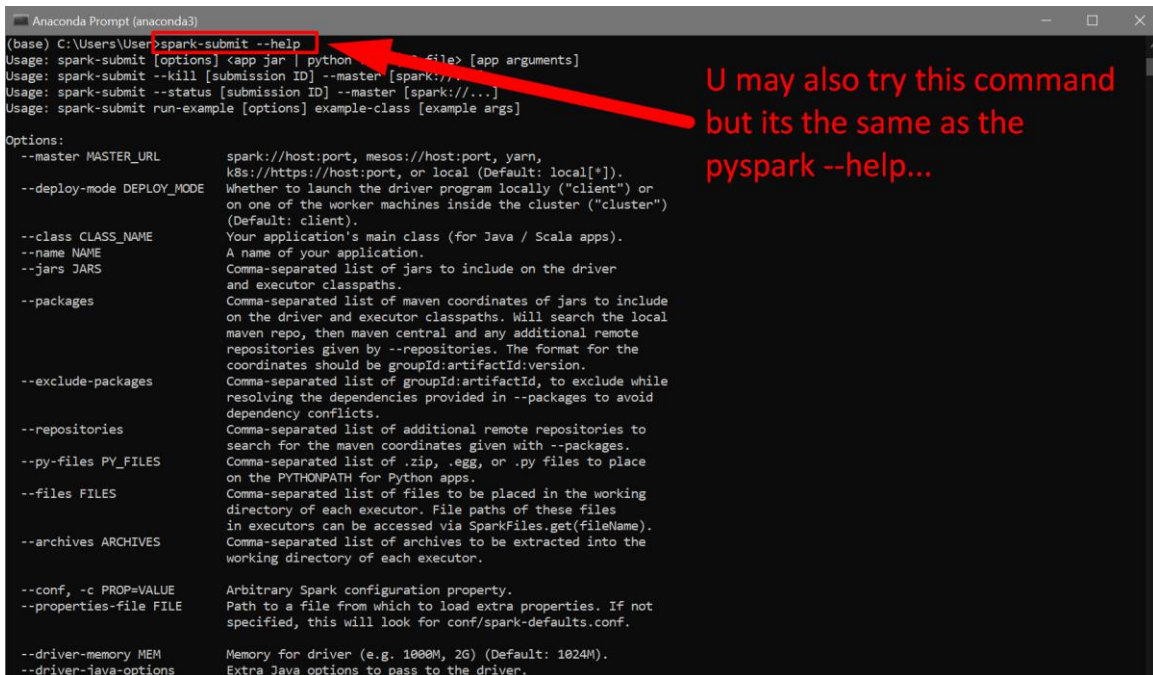
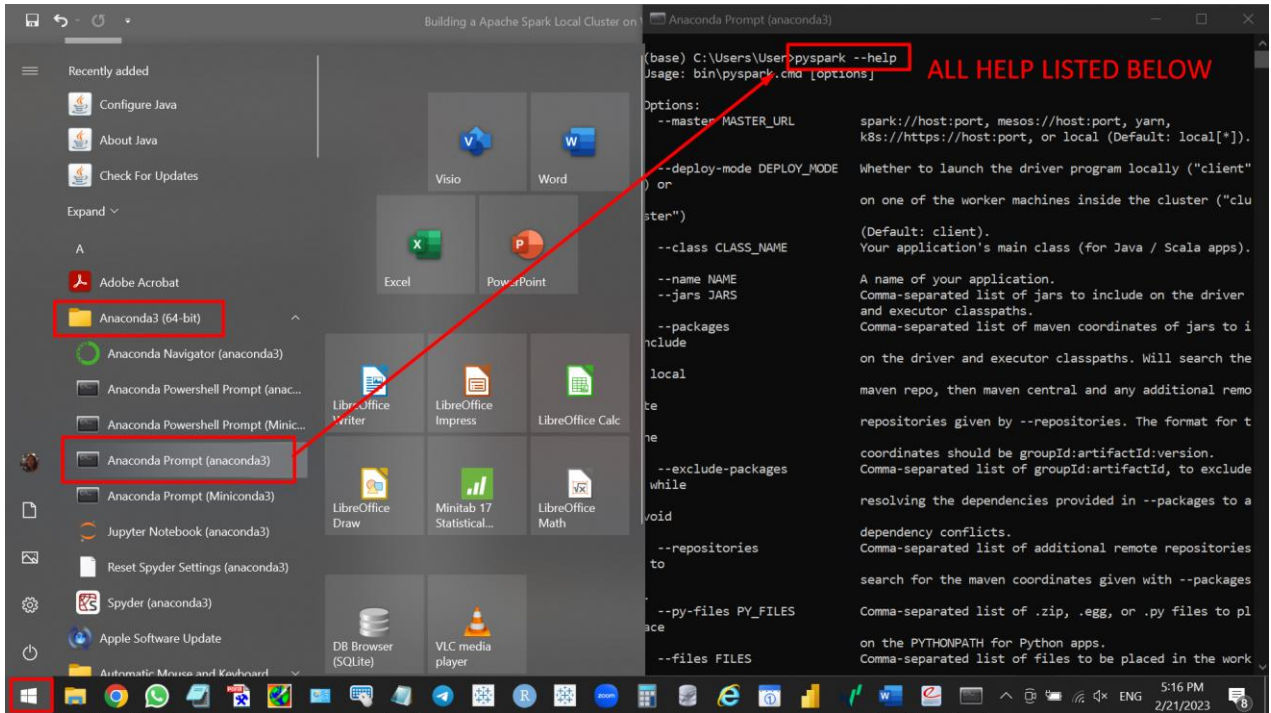


Now...go to your SLAVE computer....

```
Anaconda Prompt (Anaconda3) - spark-class org.apache.spark.deploy.worker.Worker spark://192.168.10.131:7077
et(); users with modify permissions: Set(User); groups with modify permissions: Set()
23/02/21 13:46:45 INFO Utils: Successfully started service 'sparkWorker' on port 63782.
23/02/21 13:46:45 INFO Worker: Worker decommissioning not enabled.
23/02/21 13:46:45 INFO Worker: Starting Spark worker 192.168.10.129:63782 with 4 cores, 6.9 GiB RAM
23/02/21 13:46:45 INFO Worker: Running Spark version 3.3.1
23/02/21 13:46:45 INFO Worker: Spark home: C:\SPARK\spark
23/02/21 13:46:45 INFO ResourceUtils: =====
23/02/21 13:46:45 INFO ResourceUtils: No custom resources configured for spark.worker.
23/02/21 13:46:45 INFO ResourceUtils: =====
23/02/21 13:46:46 INFO Utils: Successfully started service 'WorkerUI' on port 8081.
23/02/21 13:46:46 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://DESKTOP-2UFVFJF:8081
23/02/21 13:46:46 INFO Worker: Connecting to master 192.168.10.131:7077...
23/02/21 13:46:46 INFO TransportClientFactory: Successfully created connection to /192.168.10.131:7077 after 91 ms (0 ms spent in bootstraps)
23/02/21 13:46:46 INFO Worker: Successfully registered with master spark://192.168.10.131:7077
23/02/21 13:59:22 INFO Worker: Asked to launch executor app-20230221135923-0000/0 for PySparkShell
23/02/21 13:59:22 INFO SecurityManager: Changing view acls to: User
23/02/21 13:59:22 INFO SecurityManager: Changing modify acls to: User
23/02/21 13:59:22 INFO SecurityManager: Changing view acls groups to:
23/02/21 13:59:22 INFO SecurityManager: Changing modify acls groups to:
23/02/21 13:59:22 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(User); groups with view permissions:
et(); users with modify permissions: Set(User); groups with modify permissions: Set()
23/02/21 13:59:22 INFO ExecutorRunner: Launch command: "C:\Program Files\Java\jdk1.8.0_202\bin\java" "-cp" "C:\SPARK\spark\bin\..\conf\;C:\SPARK\spark\jars\*" "-Xmx1024M"
Dspark.driver.port=50569" "-XX:+IgnoreUnrecognizedVMOptions" "--add-opens=java.base/java.lang=ALL-UNNAMED" "--add-opens=java.base/java.lang.invoke=ALL-UNNAMED" "--add-ope
=java.base/java.lang.reflect=ALL-UNNAMED" "--add-opens=java.base/java.io=ALL-UNNAMED" "--add-opens=java.base/java.net=ALL-UNNAMED" "--add-opens=java.base/java.nio=ALL-UNN
MED" "--add-opens=java.base/java.util=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent.atomic=ALL-UNNAMED"
--add-opens=java.base/sun.nio.ch=ALL-UNNAMED" "--add-opens=java.base/sun.nio.cs=ALL-UNNAMED" "--add-opens=java.base/sun.security.action=ALL-UNNAMED" "--add-opens=java.base
sun.util.calendar=ALL-UNNAMED" "--add-opens=java.security.jgss/sun.security.krb5=ALL-UNNAMED" "org.apache.spark.executor.CoarseGrainedExecutorBackend" "--driver-url" "spar
://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569" "--executor-id" "0" "--hostname" "192.168.10.129" "--cores" "4" "--app-id" "app-20230221135923-0000" "--worker-url" "spar
//Worker@192.168.10.129:63782"
```

at your SLAVE laptop, once any ipynb has been opened... u can see that your slave starts to work already....

X. SEEKING FURTHER HELP



XI. DISCLAIMER

At the point of this writing, a key issue popped up while trying to do local Master / Slave configuration on Windows.

A. ISSUE: SLAVE APP KEEPS ON EXITING AND RESTARTING

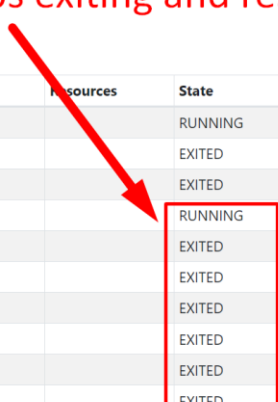
At your Master laptop....

User: User
Cores: Unlimited (8 granted)
Executor Limit: Unlimited (2 granted)
Executor Memory: 1024.0 MiB
Executor Resources:
Submit Date: 2023/02/21 14:33:55
State: RUNNING
[Application Detail UI](#)

**on Localhost: 8080,
if you click and go into the Application Details
u see that the Slave node keeps exiting and restarting**

▼ **Executor Summary (10)**

ExecutorID	Worker	Cores	Memory	Resources	State	Logs
9	worker-20230221134645-192.168.10.129-63782	4	1024		RUNNING	stdout stderr
4	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
8	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
0	worker-20230221143109-192.168.10.131-50741	4	1024		RUNNING	stdout stderr
5	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
1	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
2	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
6	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
3	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr
7	worker-20230221134645-192.168.10.129-63782	4	1024		EXITED	stdout stderr



At your Slave laptop Cmd Prompt...

```
Anaconda Prompt (Anaconda3) - spark-class org.apache.spark.deploy.worker.Worker spark://192.168.10.131:7077
"--add-opens=java.base/sun.nio.ch=ALL-UNNAMED" "--add-opens=java.base/sun.nio.cs=ALL-UNNAMED" "--add-opens=java.base/sun.security.action=ALL-UNNAMED" "--add-opens=java.base/sun.util.calendar=ALL-UNNAMED" "--add-opens=java.security.jgss/sun.security.krb5=ALL-UNNAMED" "org.apache.spark.executor.CoarseGrainedExecutorBackend" "--driver-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569" "--executor-id" "156" "--hostname" "192.168.10.129" "--cores" "4" "--app-id" "app-20230221135923-0000" "--worker-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569"

3/02/21 14:28:31 INFO Worker: Executor app-20230221135923-0000/156 finished with state EXITED message Command exited with code 1 exitStatus 1
3/02/21 14:28:31 INFO ExternalShuffleBlockResolver: Clean up non-shuffle and non-RDD files associated with the finished executor 156
3/02/21 14:28:31 INFO ExternalShuffleBlockResolver: Executor is not registered (appId=app-20230221135923-0000, execId=156)
3/02/21 14:28:31 INFO Worker: Asked to launch executor app-20230221135923-0000/157 for PySparkShell
3/02/21 14:28:31 INFO SecurityManager: Changing view acls to: User
3/02/21 14:28:31 INFO SecurityManager: Changing modify acls to: User
3/02/21 14:28:31 INFO SecurityManager: Changing view acls groups to:
3/02/21 14:28:31 INFO SecurityManager: Changing modify acls groups to:
3/02/21 14:28:31 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(User); groups with view permissions: Set(); users with modify permissions: Set(User); groups with modify permissions: Set()
3/02/21 14:28:31 INFO ExecutorRunner: Launch command: "C:\Program Files\Java\jdk1.8.0_202\bin\java" "-cp" "C:\SPARK\spark\bin\..\conf;C:\SPARK\spark\jars\*" "-Xmx1024M" "D:\spark_driver.port=50569" "-XX:+IgnoreUnrecognizedVMOptions" "--add-opens=java.base/java.lang=ALL-UNNAMED" "--add-opens=java.base/java.lang.invoke=ALL-UNNAMED" "--add-opens=java.base/java.lang.reflect=ALL-UNNAMED" "--add-opens=java.base/java.io=ALL-UNNAMED" "--add-opens=java.base/java.net=ALL-UNNAMED" "--add-opens=java.base/java.nio=ALL-UNNAMED" "--add-opens=java.base/java.util=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent.atomic=ALL-UNNAMED" "--add-opens=java.base/sun.nio.ch=ALL-UNNAMED" "--add-opens=java.base/sun.nio.cs=ALL-UNNAMED" "--add-opens=java.base/sun.security.action=ALL-UNNAMED" "--add-opens=java.base/sun.util.calendar=ALL-UNNAMED" "org.apache.spark.executor.CoarseGrainedExecutorBackend" "--driver-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569" "--executor-id" "157" "--hostname" "192.168.10.129" "--cores" "4" "--app-id" "app-20230221135923-0000" "--worker-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569"

3/02/21 14:28:42 INFO Worker: Executor app-20230221135923-0000/157 finished with state EXITED message Command exited with code 1 exitStatus 1
3/02/21 14:28:42 INFO ExternalShuffleBlockResolver: Clean up non-shuffle and non-RDD files associated with the finished executor 157
3/02/21 14:28:42 INFO ExternalShuffleBlockResolver: Executor is not registered (appId=app-20230221135923-0000, execId=157)
3/02/21 14:28:42 INFO Worker: Asked to launch executor app-20230221135923-0000/158 for PySparkShell
3/02/21 14:28:42 INFO SecurityManager: Changing view acls to: User
3/02/21 14:28:42 INFO SecurityManager: Changing modify acls to: User
3/02/21 14:28:42 INFO SecurityManager: Changing view acls groups to:
3/02/21 14:28:42 INFO SecurityManager: Changing modify acls groups to:
3/02/21 14:28:42 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(User); groups with view permissions: Set(); users with modify permissions: Set(User); groups with modify permissions: Set()
3/02/21 14:28:42 INFO ExecutorRunner: Launch command: "C:\Program Files\Java\jdk1.8.0_202\bin\java" "-cp" "C:\SPARK\spark\bin\..\conf;C:\SPARK\spark\jars\*" "-Xmx1024M" "D:\spark_driver.port=50569" "-XX:+IgnoreUnrecognizedVMOptions" "--add-opens=java.base/java.lang=ALL-UNNAMED" "--add-opens=java.base/java.lang.invoke=ALL-UNNAMED" "--add-opens=java.base/java.lang.reflect=ALL-UNNAMED" "--add-opens=java.base/java.io=ALL-UNNAMED" "--add-opens=java.base/java.net=ALL-UNNAMED" "--add-opens=java.base/java.nio=ALL-UNNAMED" "--add-opens=java.base/java.util=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent=ALL-UNNAMED" "--add-opens=java.base/java.util.concurrent.atomic=ALL-UNNAMED" "--add-opens=java.base/sun.nio.ch=ALL-UNNAMED" "--add-opens=java.base/sun.nio.cs=ALL-UNNAMED" "--add-opens=java.base/sun.security.action=ALL-UNNAMED" "--add-opens=java.base/sun.util.calendar=ALL-UNNAMED" "org.apache.spark.executor.CoarseGrainedExecutorBackend" "--driver-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569" "--executor-id" "158" "--hostname" "192.168.10.129" "--cores" "4" "--app-id" "app-20230221135923-0000" "--worker-url" "spark://CoarseGrainedScheduler@DESKTOP-2UFVFJF:50569"
```

keeps on exiting and restarting

B. POSSIBLE SOLUTIONS

1. I can't find any solution to the above problem (even after a long time searching on the internet and trying and failing).
2. I have never tested live on real big data before ...meaning, even the constant exiting and restarting could be due to "no real jobs running"....
3. You could test it on big data if you have....and maybe the problem will go away...
4. Spark was built mainly on Linux (Ubuntu).... Thus, most codes on the internet use Ubuntu...
5. <https://medium.com/codex/setup-a-spark-cluster-step-by-step-in-10-minutes-922c06f8e2b1>
6. You might want to try the link above to install only on Ubuntu systems... which may work better...(else Windows systems don't seem to work)
7. But whatever installation we have learnt so far on Windows is the exact steps as Ubuntu OS (like creating Environment Variables).
8. However, you might need to have MULTIPLE Ubuntu laptops if you really want to crunch big data.. and install all of them with Spark...which is a lot of work...
9. A better solution will be to do Spark on the cloud...

ABOUT DR. ALVIN ANG



Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.