Installations Steps - Homebrew, Java, Hadoop, Hive, SQL Workbench, Spark, PySpark

Author : Asit Piri (asit.piri@gmail.com (mailto:asit.piri@gmail.com))

Date: 2nd May 2022

```
STEP 1: Go to https://brew.sh/ and copy: /bin/bash -c "$(curl -fsSL
https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)" and type it in
the terminal.

### Execute the commend in the terminal
STEP 2: echo 'eval "$(/opt/homebrew/bin/brew shellenv)"' >>
/Users/asitpiri/.zprofile
    eval "$(/opt/homebrew/bin/brew shellenv)"

### To verify the entry in .zprofile
STEP 3: code .zprofile

### To check the version Homebrew 3.4.10
STEP 4: brew --version

### To check the homebrew repository path
STEP 5: brew --repository

STEP 6: brew install hadoop

STEP 7: cd /opt/homebrew/Cellar/hadoop/3.3.2/libexec/etc/hadoop

STEP 8: Make changes to hadoop-env.sh

STEP 9: Make changes to core-site.xml

<configuration>
 <property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
 </property>
</configuration>

STEP 10: Make changes to hdfs-site.xml

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>

STEP 11: Make changes to mapred-site.xml

<configuration>
    <property>
       <name>mapreduce.framework.name</name>
```

```
       <value>yarn</value>
    </property>
    <property>
    <name>mapreduce.application.classpath</name>
  <value>
$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapre
duce/lib/*
  </value>
    </property>
</configuration>

STEP 12: Make changes to yarn-site.xml

<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
   <value>
JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DIS
TCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME
  </value>
  </property>
</configuration>

STEP 13: Download and install Arm 64 DMG Installer for Mac
        (Use the link to download
https://www.oracle.com/java/technologies/downloads/#java11)

        Java is installed in this path: /Library/Java/JavaVirtualMachines/jdk-
18.0.1.jdk!


STEP 14: Download and install Intelij Idea community Edition using Jetbeans toolbox
        (Use the link https://www.jetbrains.com/toolbox-app/)

        Download JDK 8, beacause JDK Version 18, is not fully compatible yet.

        Open Intelij Idea community Edition

        Click New Project

        Make a new project called: Hello

        Right click on src folder in the sidebar and click new -> JavaClass, and
then create a class called HelloWorld

        Then type the following code in the editor:

        public class HelloWorld {
            public static void main(String[] args) {
                System.out.println("This is a test code.");
            }
        }


        Use this command to test the java_home: /usr/libexec/java_home

        Output: /Library/Java/JavaVirtualMachines/jdk-18.0.1.jdk/Contents/Home
```

```
        To Remove the existing java_version, type:
        sudo rm -rf /Library/Java/JavaVirtualMachines/jdk-18.0.1.jdk


        sudo rm -rf /Library/Java/JavaVirtualMachines/jdk-18.0.1.jdk


STEP 15: stop-all.sh

STEP 16: hadoop namenode -format

STEP 17: start-all.sh

STEP 18: jps

STEP 19: ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa

         cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

        (Generate new keygen: See from this link:
https://stackoverflow.com/questions/48978480/hadoop-permission-denied-publickey-
password-keyboard-interactive-warning/49960886

STEP 20: http://localhost:9870/ and http://localhost:8088/cluster
        (Use the link to check Hadoop's Health)

STEP 21: hadoop fs -mkdir /user

STEP 22: touch demo.csv

STEP 23: hadoop fs -put demo.csv /user
```

Steps to Start and test Hadoop

Open Terminal

Enter: cd /opt/homebrew/Cellar/hadoop/3.3.2/libexec/etc/hadoop

Enter: stop-all.sh

Enter: hadoop namenode -format

Enter: start-all.sh

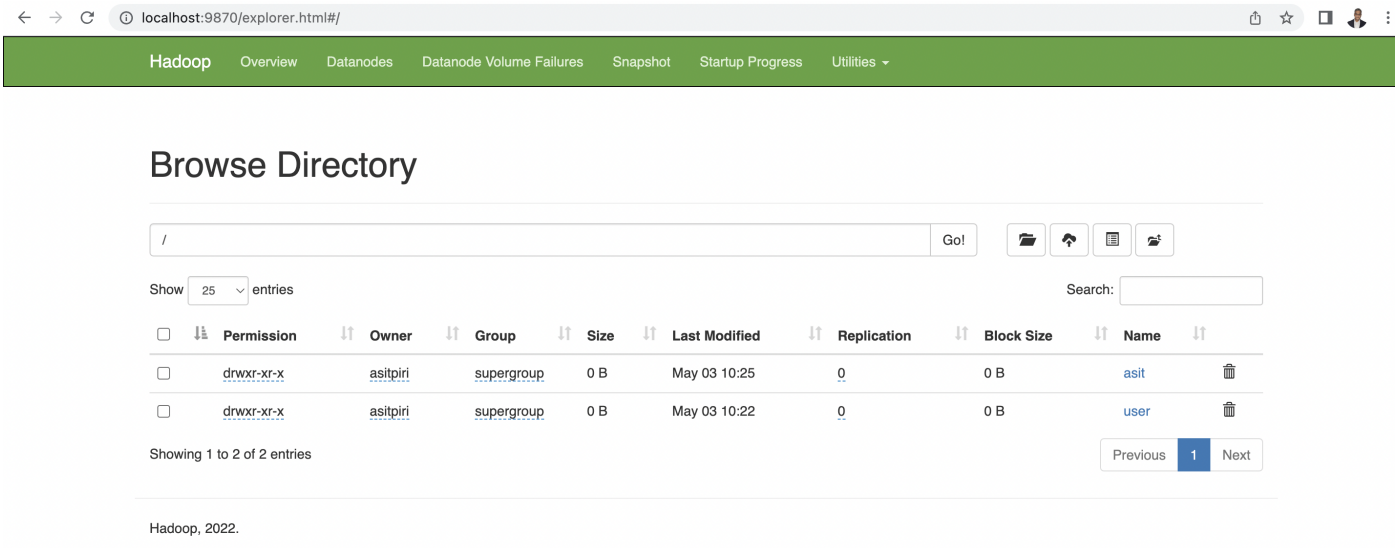
Enter in Browser: http://localhost:9870/ (http://localhost:9870/)

Enter in Browser: http://localhost:8088/cluster (http://localhost:8088/cluster)

Enter: hadoop fs -mkdir /asit

Enter: touch test.csv

Enter: hadoop fs -put demo.csv /user



Installing Spark

STEP 24: code .bash_profile

STEP 25: Type: export SPARK_HOME=/opt/homebrew/Cellar/apache-spark/3.2.1

```
export PATH=$PATH:$SPARK_HOME/bin
```

and save it

STEP 26: source .bash_profile

To run the spark-shell

STEP 27: sudo spark-shell

Steps to Start Spark-Shell

Open Terminal

Enter: spark-shell

Enter: spark

Enter: sc

Enter: val rdd = sc.parallelize(List(1,2,3,4,5))

Enter: rdd.count

asitpiri — java ◂ spark-shell — 80×24

```
<console>:23: error: not found: value exit
       exit
       ^

scala> exit()
<console>:23: error: not found: value exit
       exit()
       ^

scala> spark
res2: org.apache.spark.sql.SparkSession = org.apache.spark.sql.SparkSession@c7b2
76

scala> sc
res3: org.apache.spark.SparkContext = org.apache.spark.SparkContext@2bc41bcb

scala> val rdd = sc.parallelize(List(1,2,3,4,5))
rdd: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at
<console>:23

scala> rdd.count
res4: Long = 5
```

For Installing Pyspark

STEP 28: Open terminal and add the below configurations in the bash_profile:

export PYSPARK_DRIVER_PYTHON="jupyter" export PYSPARK_DRIVER_PYTHON_OPTS="lab"

STEP 29: source .bash_profile

STEP 30: sudo pyspark

Configuring pyspark to open in Jypyter Labs

STEP 31: conda install -c conda-forge spylon-kernel

STEP 32: sudo python -m spylon_kernel install

STEP 33: cd

STEP 34: sudo pyspark

STEP 35:

In [ ]:

Setup of Pyspark (Spylon-Kernel) Environment with Jupyter on Mac M1

STEP 1: java -version

STEP 2: brew install apache-spark

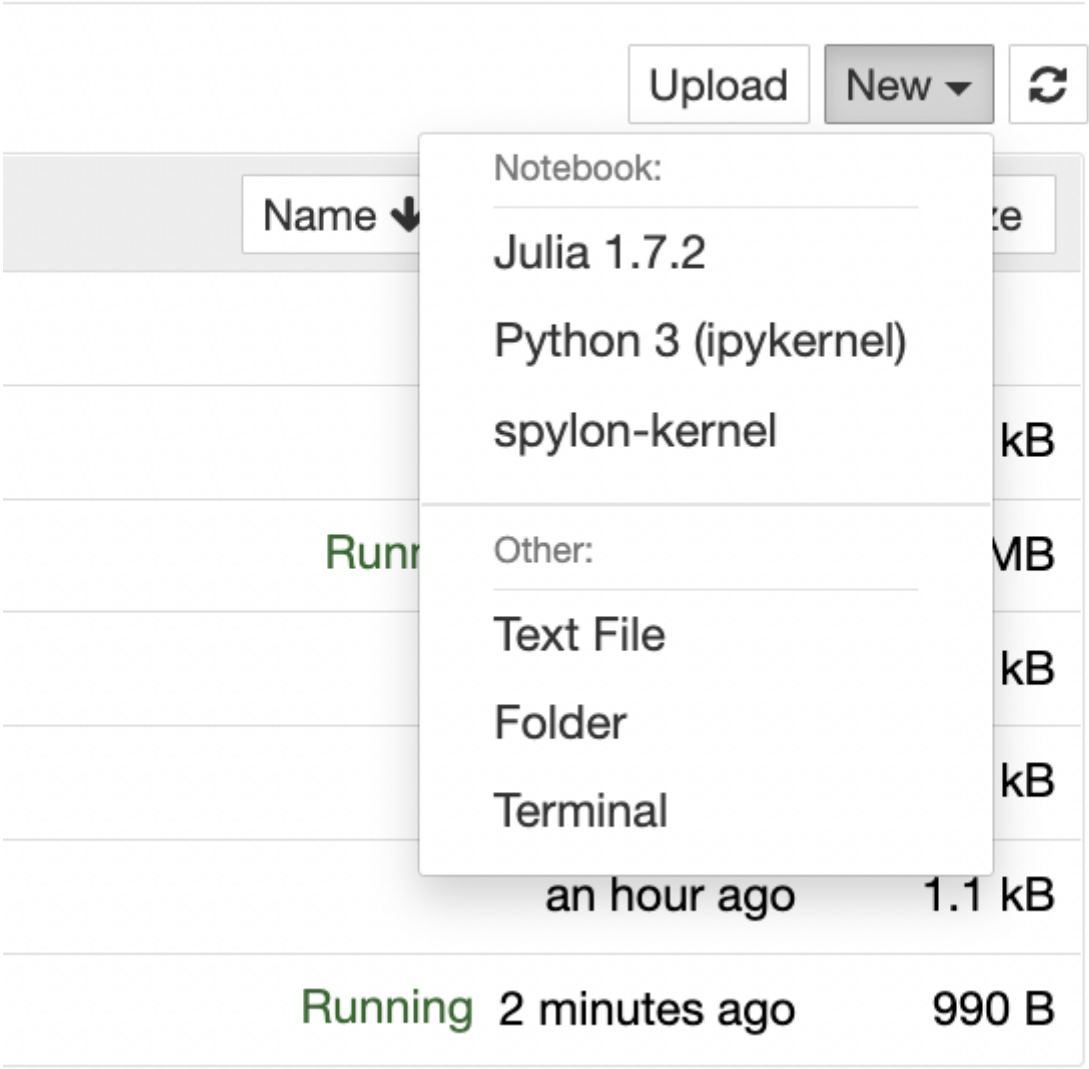
STEP 3: sudo python -m spylon_kernel install

STEP 4: sudo pyspark

STEP 5: Open Anaconda

STEP 6: Launch Jupyter Notebook

STEP 7: Open Sylon-Kernel

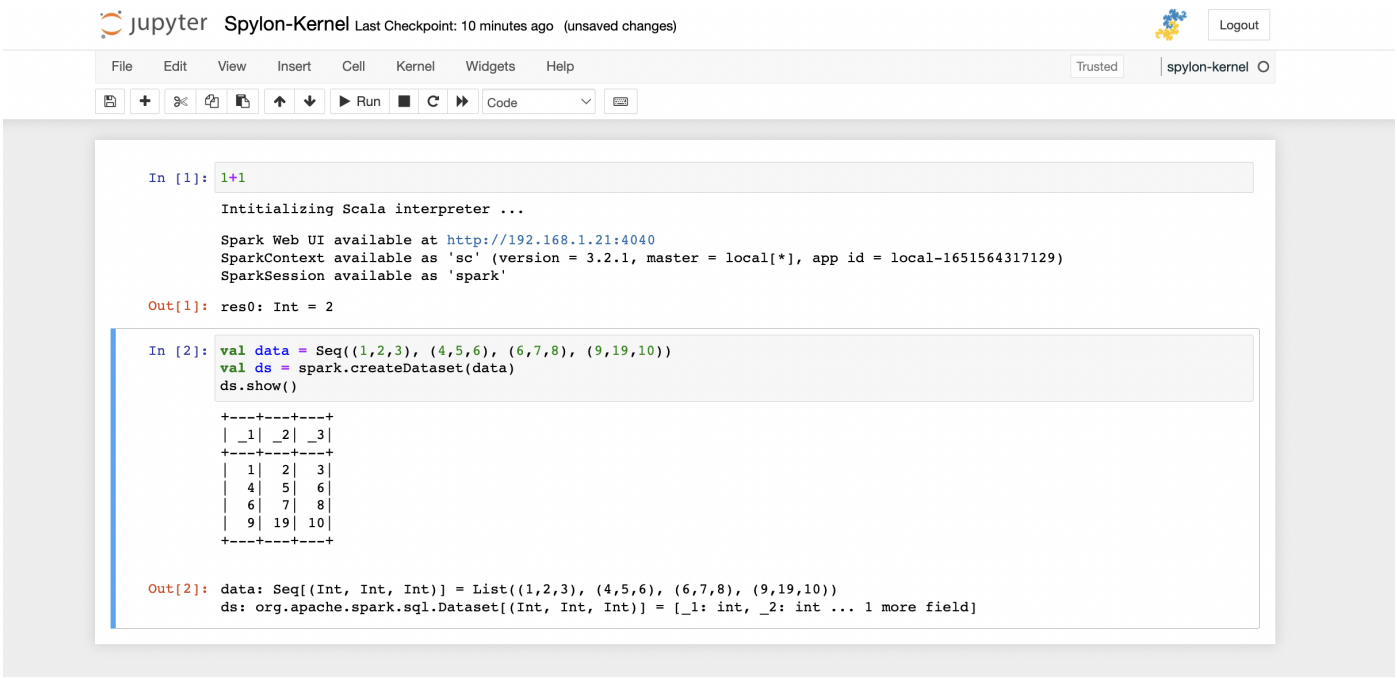


STEP 8: Write the following code in the notebook and hit run.

```
val data = Seq((1,2,3), (4,5,6), (6,7,8), (9,19,10))

val ds = spark.createDataset(data)

ds.show()
```


jupyter Spylon-Kernel Last Checkpoint: 10 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
spylon-kernel
Run
Code
In [1]: 1+1
Intitializing Scala interpreter ...
Spark Web UI available at http://192.168.1.21:4040
SparkContext available as 'sc' (version = 3.2.1, master = local[*], app id = local-1651564317129)
SparkSession available as 'spark'
Out[1]: res0: Int = 2
In [2]: val data = Seq((1,2,3), (4,5,6), (6,7,8), (9,19,10))
val ds = spark.createDataset(data)
ds.show()
+---+---+---+
| _1| _2| _3|
+---+---+---+
1	2	3
4	5	6
6	7	8
9	19	10
+---+---+---+
Out[2]: data: Seq[(Int, Int, Int)] = List((1,2,3), (4,5,6), (6,7,8), (9,19,10))
ds: org.apache.spark.sql.Dataset[(Int, Int, Int)] = [_1: int, _2: int ... 1 more field]

In [ ]: