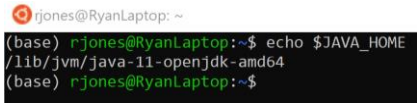
Tutorial – Install Spark on Windows Machine

Ryan Jones – COP 6526 – Dr. Wang – Fall 2020 MSDA Program

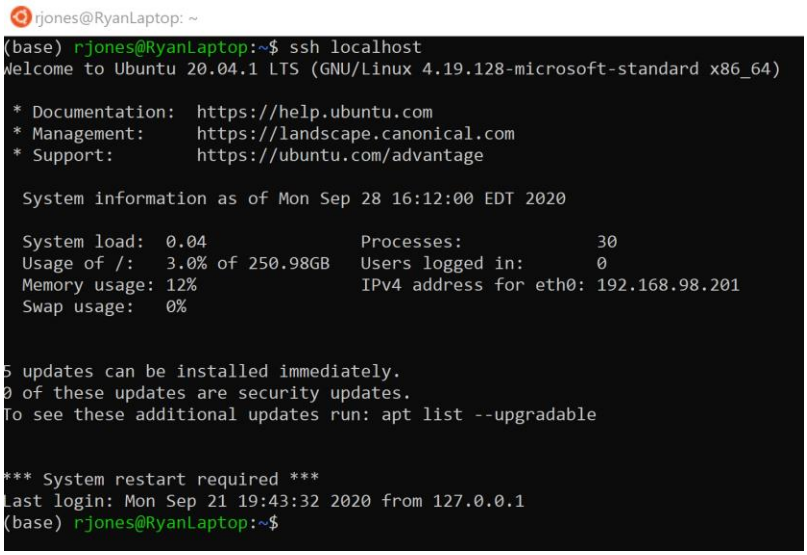
1. I have a laptop running Windows 10. I will be installing Spark using the WSL2 interface and Ubuntu.
2. Install Java and Set Path.

 # sudo apt update
 # sudo apt install default-jre default-jdk
 # java -version
 # export JAVA_HOME= /lib/jvm/java-11-openjdk-amd64
 # echo $JAVA_HOME

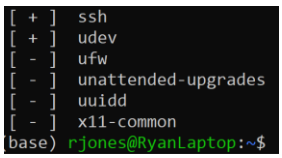


3. Setup SSH Without Password.

 # ssh-keygen -t rsa
 # cd ~/.ssh/
 # cat id_rsa.pub >> authorized_keys
 # ssh localhost



 # service –status-all

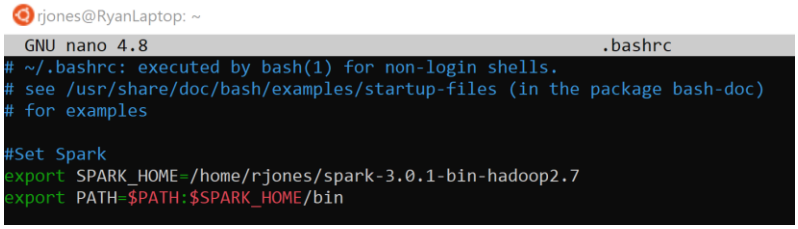


4. Install Apache Spark and Configure

 # wget https://downloads.apache.org/spark/spark-3.0.1/spark-3.0.1-bin-hadoop2.7.tgz
 # tar xf spark-3.0.1-bin-hadoop2.7.tgz

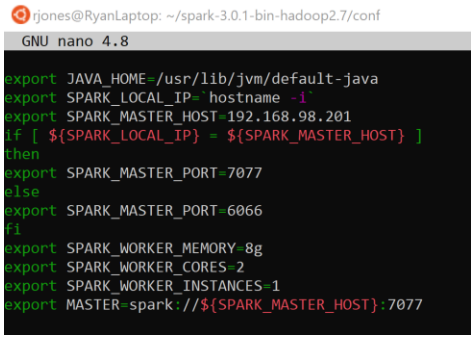
 # cd ~
 # vi .bashrc

Enter the below information into the .bashrc file and save it.



 # cd $SPARK_HOME/conf
 # cp spark-env.sh.template spark-env.sh
 # nano spark-env.sh

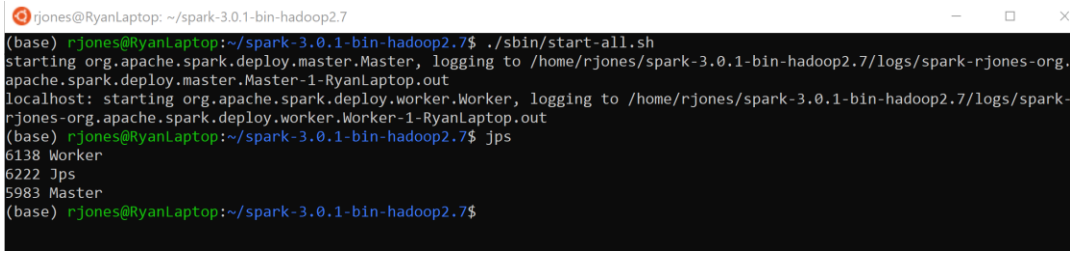
Enter the below information into the spark-env.sh file and save it.



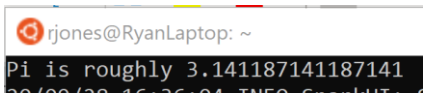
 # cp slaves.template slaves

5. Run a Program on Spark

 # cd $SPARK_HOME
 # ./sbin/start-all.sh
 # jps



 # spark-submit --class org.apache.spark.examples.SparkPi --master spark://192.168.98.201:6066 --executor-memory 6G /home/rjones/spark-3.0.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.0.1.jar 10



 # cd $SPARK_HOME

 # ./sbin/stop-all.sh

 # jps