SPARK INSTALLATION:

After having Java and Hadoop installed in the VM the next step is to install Spark service on to the VM.

The procedure begins by checking the Hadoop HDFS and YARN service and the respective environment configuration

* jps and
* echo $Hadoop\_CONF\_DIR

After the single node cluster is seen to successfully working the next step is to download and extract the spark package with the following command:

* wget <http://www-eu.apache.org/dist/spark/spark-2.1.0/spark-2.1.0-bin-hadoop2.7.tgz>
* tar zxvf spark-2.1.0-bin-hadoop2.7.tgz

Then after this step, the Spark installation is checked in the local mode

* ./bin/run-example SparkPi 10

Once the value of Pi is successfully retrieved then the next step is to Run Spark Shell with Yarn,

* ./bin/spark-shell –master yarn –deploy-mode client

However, since the VM does not have sufficient memory to run the yarn and so it was to be asked to test in the local machine using the following commands,

* ./bin/spark-submit –class org.apache.spark.examples.SparkPi \
* --master local \
* --deploy-mode cluster \
* --driver-memory 512m \
* --executor memory 512m \
* --executor-cores 1 \
* --queue default \
* Examples/jars/spark-examples\*.jar \