Single Node Apache Spark 1.6.1 Setup and Configuration on Ubuntu 14.04

Data Science Lab, The Department of Computer Science, KSKV Kachchh University.

Web: http://cs.kutchuni.edu.in

1) Follow Document ‘***Apache Hadoop Single Node Setup- CSKSKV-v1.3.docx***’ and setup everything accordingly.

2) **~/.bashrc**
Edit above file and add below environment variables.

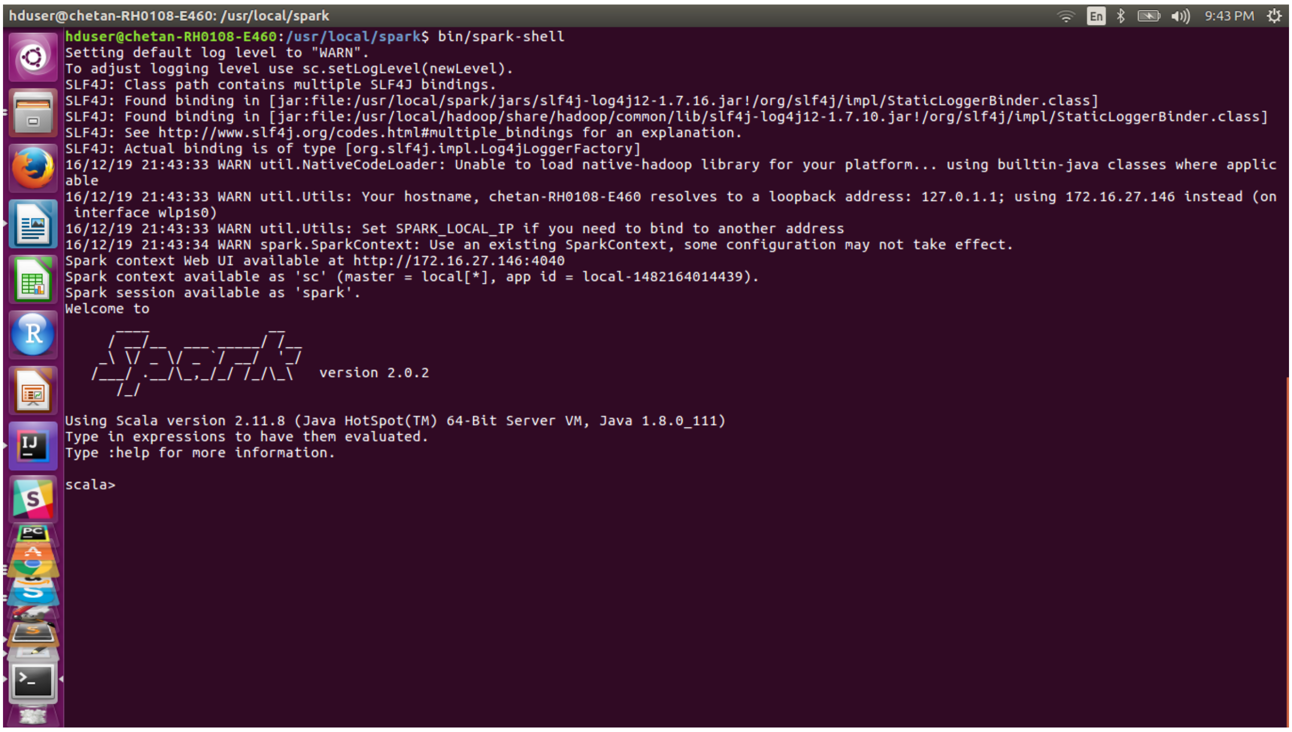
```
#SPARK VARIABLES START
export SPARK_DIST_CLASSPATH=$(hadoop --config /usr/local/hadoop/etc/hadoop classpath)
export SPARK_HOME=/usr/local/spark-1.6.1
#SPARK VARIABLES END
```

3) Download Apache Spark unzip and move to /usr/local/spark

```
wget http://d3kbcqa49mib13.cloudfront.net/spark-1.6.1-bin-hadoop2.6.tgz
$ tar xfz spark-1.6.1-bin-hadoop2.6.tgz
$ sudo mv spark-1.6.1 /usr/local/hadoop
```

4) Launching Spark-Shell

```
/usr/local/spark$ bin/spark-shell
```

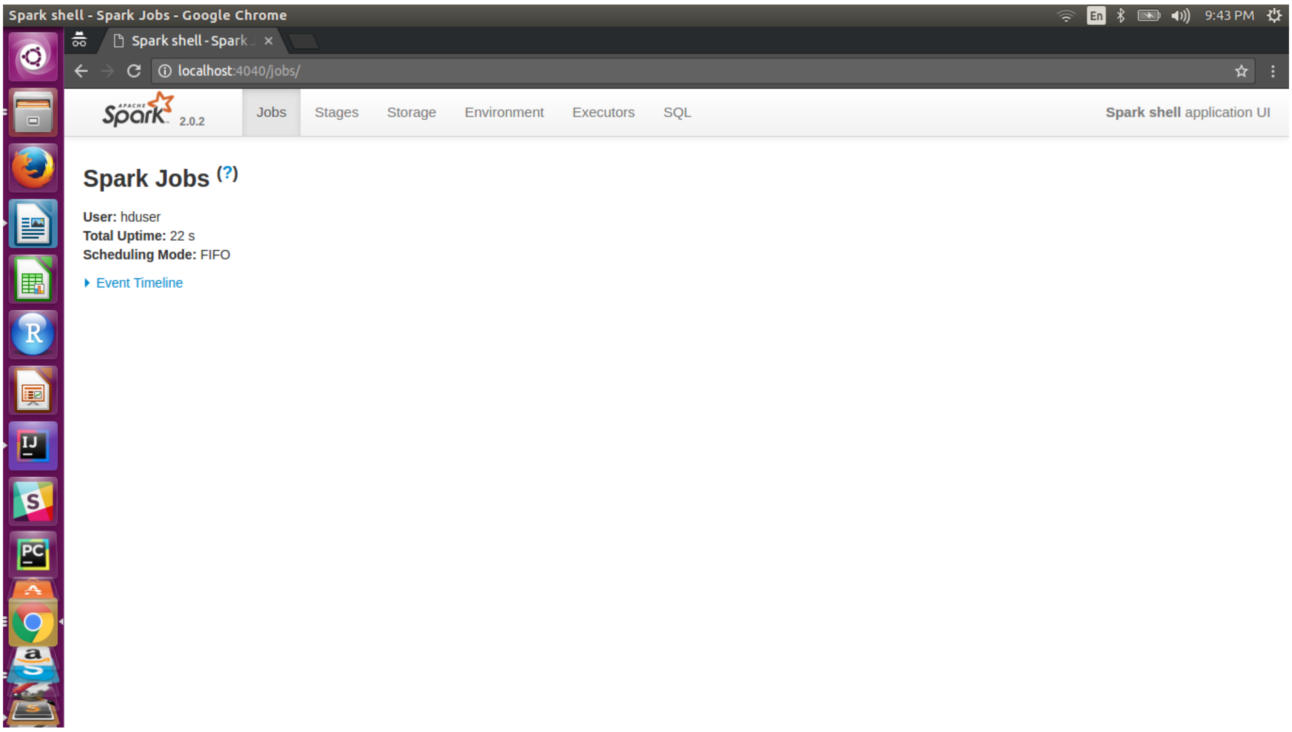


```
hduser@chetan-RH0108-E460:/usr/local/spark$ bin/spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/spark/jars/slf4j-log4j12-1.7.16.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
16/12/19 21:43:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applic
able
16/12/19 21:43:33 WARN util.Utils: Your hostname, chetan-RH0108-E460 resolves to a loopback address: 127.0.1.1; using 172.16.27.146 instead (on
 interface wlp1s0)
16/12/19 21:43:33 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
16/12/19 21:43:34 WARN spark.SparkContext: Use an existing SparkContext, some configuration may not take effect.
Spark context Web UI available at http://172.16.27.146:4040
Spark context available as 'sc' (master = local[*], app id = local-1482164014439).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.0.2
      /_/

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_111)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Figure 1: *Spark Shell Laucher Screen with Scala Interpretor*



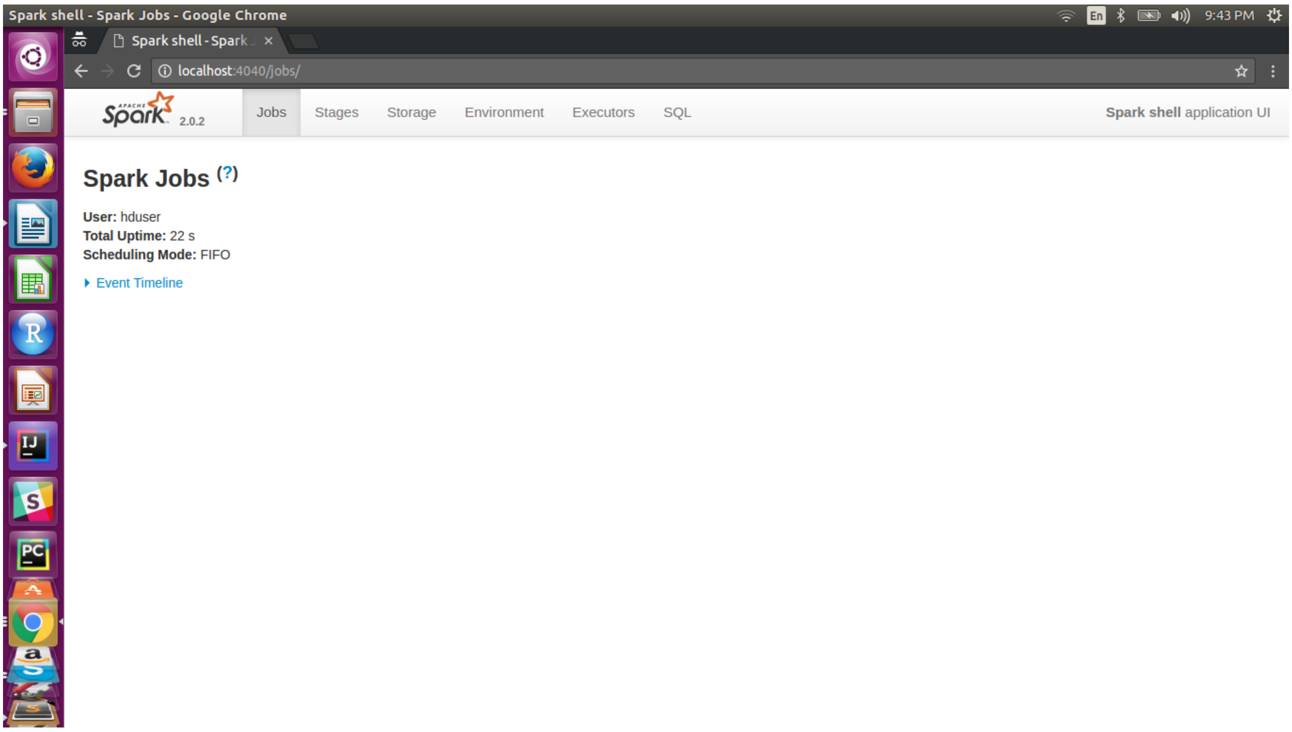
Spark Jobs (?)

User: hduser
Total Uptime: 22 s
Scheduling Mode: FIFO

Event Timeline

Spark UI Screen for Jobs Monitoring

http://localhost:4040/jobs/



Spark Jobs (?)

User: hduser
Total Uptime: 22 s
Scheduling Mode: FIFO

Event Timeline

Figure 2: *Spark UI Job Monitoring Screen*

5) Submit Spark Jobs

There are 3 ways to submit Spark Jobs
1) Spark-Shell
2) Spark Submit

3) Spark Launcher

Spark Submit Command

```
./bin/spark-submit \
  --class <main-class> \
  --master <master-url> \
  --deploy-mode <deploy-mode> \
  --conf <key>=<value> \
  ... # other options
  <application-jar> \
  [application-arguments]
```

Example,

```
bin/spark-submit --class hbase.spark.chetan.com.SparkHbaseJob /home/chetan/hbase-spark/Sparkassembly-1.0.jar
```

Where Jar must be Uber(Super) Jar.