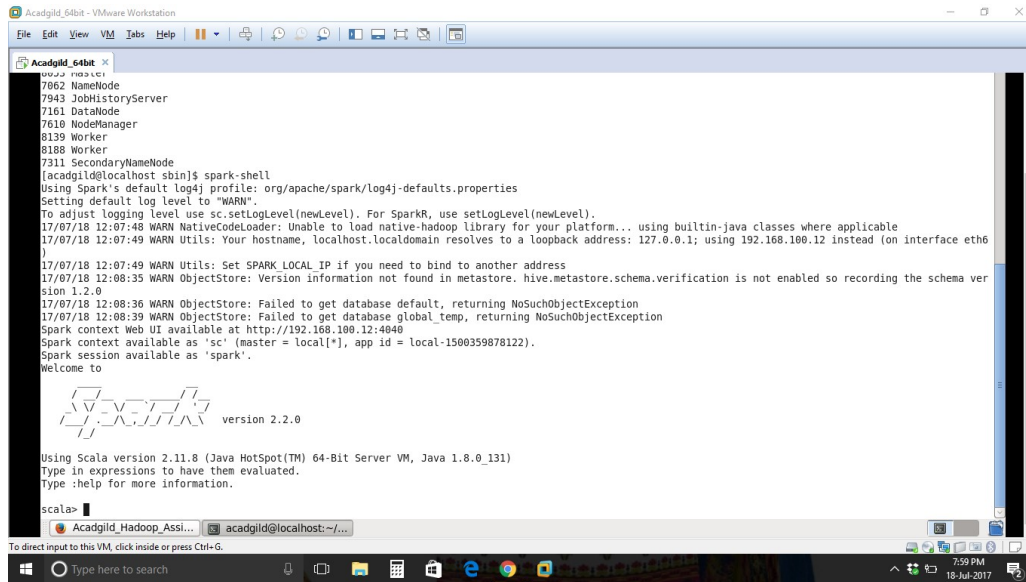


Assignment 13.2

Spark-shell

Entering spark shell



```
Acadgild_64bit
7062 NameNode
7943 JobHistoryServer
7161 DataNode
7610 NodeManager
8139 Worker
8188 Worker
7311 SecondaryNameNode
[acadgild@localhost ~]$ spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
17/07/18 12:07:48 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/07/18 12:07:49 WARN Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.168.100.12 instead (on interface eth6)
17/07/18 12:07:49 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
17/07/18 12:08:35 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema version 1.2.0
17/07/18 12:08:36 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
17/07/18 12:08:39 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Spark context Web UI available at http://192.168.100.12:4040
Spark context available as 'sc' (master = local[*], app id = local-15003598781222).
Spark session available as 'spark'.
Welcome to
  ____
 /    \        \
/_    _\        \
 \  /             \
  \/              \
   \               \
   /               /
  /_              _/
   \             /
   /_          _/
   /_        _/
   /_      _/
   /_    _/
   /_  _/
   /_/_

version 2.2.0

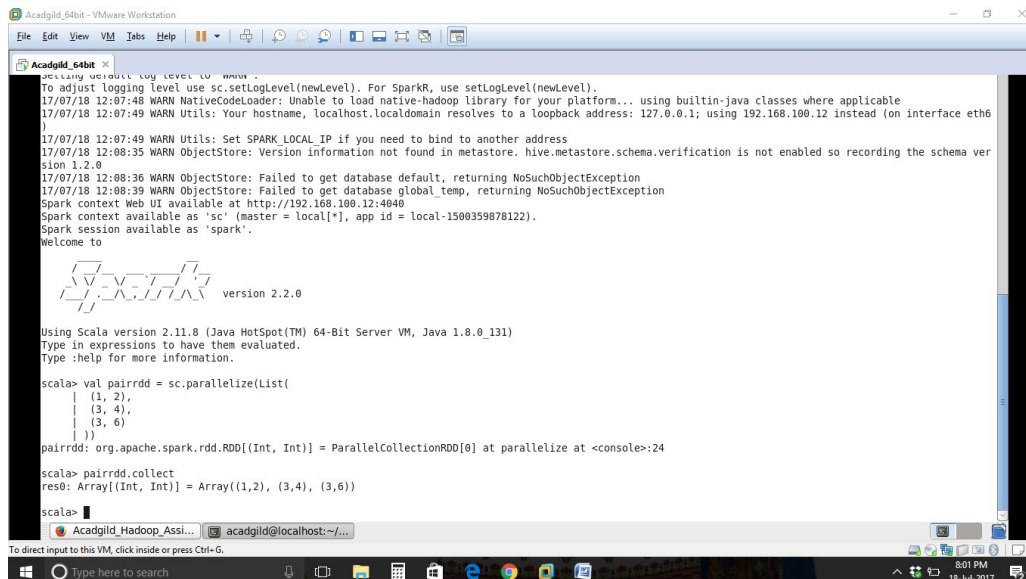
Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_131)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Task 1: Create a simple pairRDD of (1,2),(3,4),(3,6)

```
Val pairrdd=sc.parallelize(List((1,2),(3,4),(3,6)))
```

Pairrdd.collect



```
Acadgild_64bit
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
17/07/18 12:07:48 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/07/18 12:07:49 WARN Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.168.100.12 instead (on interface eth6)
17/07/18 12:07:49 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
17/07/18 12:08:35 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema version 1.2.0
17/07/18 12:08:36 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
17/07/18 12:08:39 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Spark context Web UI available at http://192.168.100.12:4040
Spark context available as 'sc' (master = local[*], app id = local-15003598781222).
Spark session available as 'spark'.
Welcome to
  ____
 /    \        \
/_    _\        \
 \  /             \
  \/              \
   \               \
   /               /
  /_              _/
   \             /
   /_          _/
   /_        _/
   /_      _/
   /_    _/
   /_  _/
   /_/_

version 2.2.0

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_131)
Type in expressions to have them evaluated.
Type :help for more information.

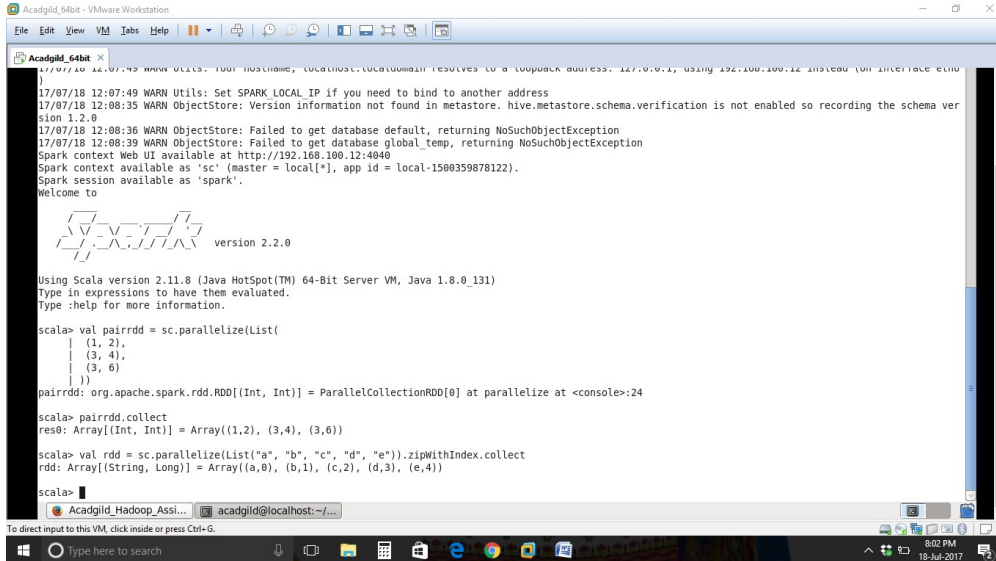
scala> val pairrdd = sc.parallelize(List(
  | (1, 2),
  | (3, 4),
  | (3, 6)
  | ))
pairrdd: org.apache.spark.rdd.RDD[(Int, Int)] = ParallelCollectionRDD[0] at parallelize at <console>:24

scala> pairrdd.collect
res0: Array[(Int, Int)] = Array((1,2), (3,4), (3,6))

scala>
```

Task 2:

Transform an RDD of ("a","b","c","d","e") to pairrdd (a,0),(b,1),(c,2),(d,3),(e,4)



```
Acadgild_64bit - VMware Workstation
File Edit View VM Help
Acadgild_64bit
17/07/18 12:07:49 WARN Utils: Your hostname, localh... resolves to a loopback address: 127.0.0.1; using 192.168.100.12 instead (UI interface enab...
17/07/18 12:07:49 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
17/07/18 12:08:35 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema ver...
17/07/18 12:08:36 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
17/07/18 12:08:39 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Spark context Web UI available at http://192.168.100.12:4040
Spark context available as 'sc' (master = local[*], app id = local-1508359878122).
Spark session available as 'spark'.
Welcome to
  ____
 /  _ \
/_/_/ \_/_/ version 2.2.0

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_131)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val pairrdd = sc.parallelize(List(
  | (1, 2),
  | (3, 4),
  | (3, 6)
  | ))
pairrdd: org.apache.spark.rdd.RDD[(Int, Int)] = ParallelCollectionRDD[0] at parallelize at <console>:24

scala> pairrdd.collect
res0: Array[(Int, Int)] = Array((1,2), (3,4), (3,6))

scala> val rdd = sc.parallelize(List("a", "b", "c", "d", "e")).zipWithIndex.collect
rdd: Array[(String, Long)] = Array((a,0), (b,1), (c,2), (d,3), (e,4))

scala>
```