

## Assignment 6.1

### Problem Statement

8. Write a Map Reduce program that takes the output of Task 5 (refer session 5, assignment 1) as input, and produce output which is sorted on the total units sold.

You may use a single reducer for the sorting.

Use Sequence File formats as an output for Task 5 and as input for this task.

#### Step 1: Start Hadoop Daemons:

```
$ start-all.sh
```

```
$ jps
```

```
[acadgild@localhost dataset]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
17/08/28 16:35:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 2830. Stop it first.
localhost: datanode running as process 2931. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 3086. Stop it first.
17/08/28 16:35:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
resourcemanager running as process 3263. Stop it first.
localhost: nodemanager running as process 3364. Stop it first.
[acadgild@localhost dataset]$ jps
2931 DataNode
3364 NodeManager
2830 NameNode
3086 SecondaryNameNode
3263 ResourceManager
4111 Jps
[acadgild@localhost dataset]$
```

#### Step 2: Compile the program and get JAR file into Acadgild Sandbox:

```
[acadgild@localhost dataset]$ cd assignment6/
[acadgild@localhost assignment6]$ ll
total 12
-rw-rw-r--. 1 acadgild acadgild 7180 Aug 28 16:37 mapreduce.jar
-rw-rw-r--. 1 acadgild acadgild 733 Aug 28 16:27 television.txt
[acadgild@localhost assignment6]$
```

#### Step 3: Move dataset to hadoop directory:

```
$ hadoop fs -mkdir /user/acadgild/hadoop/assignment6
```

```
$ hadoop fs -put television.txt /user/acadgild/hadoop/assignment6
```

```
$ hadoop fs -ls /user/acadgild/hadoop/assignment6
```

```
[acadgild@localhost assignment6]$ hadoop fs -mkdir /user/acadgild/hadoop/assignment6
17/08/28 16:40:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
[acadgild@localhost assignment6]$ hadoop fs -put television.txt /user/acadgild/hadoop/assignment6
17/08/28 16:40:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
[acadgild@localhost assignment6]$ hadoop fs -ls /user/acadgild/hadoop/assignment6
17/08/28 16:40:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
Found 1 items
-rw-r--r-- 1 acadgild supergroup 733 2017-08-28 16:40 /user/acadgild/hadoop/assignment6/television.txt
[acadgild@localhost assignment6]$
```

#### Step 4: run jar file with below command to get the desired result:

```
$ hadoop jar mapreduce.jar assignment61.Television /user/acadgild/hadoop/assignment6/television.txt /user/acadgild/hadoop/assignment6/output61a
```

```
[acadgild@localhost assignment6]$ hadoop jar mapreduce.jar assignment61.Television /user/acadgild/hadoop/assignment6/television.txt /user/acadgild/hadoop/assignment6/output61a
17/08/28 17:04:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/08/28 17:04:10 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/08/28 17:04:10 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/08/28 17:04:11 INFO input.FileInputFormat: Total input paths to process : 1
17/08/28 17:04:11 INFO mapreduce.JobSubmitter: number of splits:1
17/08/28 17:04:11 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1503918153486_0005
17/08/28 17:04:11 INFO impl.YarnClientImpl: Submitted application application_1503918153486_0005
17/08/28 17:04:11 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1503918153486_0005/
17/08/28 17:04:11 INFO mapreduce.Job: Running job: job_1503918153486_0005
17/08/28 17:04:17 INFO mapreduce.Job: Job job_1503918153486_0005 running in uber mode : false
17/08/28 17:04:17 INFO mapreduce.Job: map 0% reduce 0%
17/08/28 17:04:23 INFO mapreduce.Job: map 100% reduce 0%
17/08/28 17:04:28 INFO mapreduce.Job: map 100% reduce 100%
17/08/28 17:04:29 INFO mapreduce.Job: Job job_1503918153486_0005 completed successfully
17/08/28 17:04:29 INFO mapreduce.Job: Counters: 49
```

Above class file is run to get the number of units sold by each company without any order.

**Step 5: check the intermediate output file:**

We can check the intermediate output of the above step.

```
$ hadoop fs -ls /user/acadgild/hadoop/assignment6/output61a/
```

```
$ hadoop fs -text /user/acadgild/hadoop/assignment6/output61a/*
```

```
[acadgild@localhost assignment6]$ hadoop fs -ls /user/acadgild/hadoop/assignment6/output61a/
17/08/28 17:06:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-08-28 17:04 /user/acadgild/hadoop/assignment6/output61a/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 188 2017-08-28 17:04 /user/acadgild/hadoop/assignment6/output61a/part-r-00000
[acadgild@localhost assignment6]$ hadoop fs -text /user/acadgild/hadoop/assignment6/output61a/*
17/08/28 17:06:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
Lava 3
VA 1
Onida 4
Samsung 7
Zen 2
[acadgild@localhost assignment6]$
```

**Step 6: Output of the step 4 (sequence file) will be given as input to the second class file to sort companies based on the number of units sold.**

```
$ hadoop jar mapreduce.jar assignment61.Television2 /user/acadgild/hadoop/assignment6/output61a/* /user/acadgild/hadoop/assignment6/output61b
```

```
[acadgild@localhost assignment6]$ hadoop jar mapreduce.jar assignment61.Television2 /user/acadgild/hadoop/assignment6/output61a/* /user/acadgild/hadoop/assignment6/output61b
17/08/28 17:12:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/08/28 17:12:15 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/08/28 17:12:16 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/08/28 17:12:16 INFO input.FileInputFormat: Total input paths to process : 1
17/08/28 17:12:16 INFO mapreduce.JobSubmitter: number of splits:1
17/08/28 17:12:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1503918153486_0008
17/08/28 17:12:16 INFO impl.YarnClientImpl: Submitted application application_1503918153486_0008
17/08/28 17:12:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1503918153486_0008/
17/08/28 17:12:16 INFO mapreduce.Job: Running job: job_1503918153486_0008
17/08/28 17:12:22 INFO mapreduce.Job: Job job_1503918153486_0008 running in uber mode : false
17/08/28 17:12:22 INFO mapreduce.Job: map 0% reduce 0%
17/08/28 17:12:27 INFO mapreduce.Job: map 100% reduce 0%
17/08/28 17:12:34 INFO mapreduce.Job: map 100% reduce 100%
17/08/28 17:12:34 INFO mapreduce.Job: Job job_1503918153486_0008 completed successfully
17/08/28 17:12:34 INFO mapreduce.Job: Counters: 49
```

**Step 7: check the output generated by step 6:**

```
[acadgild@localhost assignment6]$ hadoop fs -ls /user/acadgild/hadoop/assignment6/output61b/
17/08/28 17:13:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2017-08-28 17:12 /user/acadgild/hadoop/assignment6/output61b/_SUCCESS
-rw-r--r--  1 acadgild supergroup       43 2017-08-28 17:12 /user/acadgild/hadoop/assignment6/output61b/part-r-00000
[acadgild@localhost assignment6]$ hadoop fs -cat /user/acadgild/hadoop/assignment6/output61b/*
17/08/28 17:13:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
NA      1
Akai    1
Zen     2
Lava    3
Onida   4
Samsung 7
[acadgild@localhost assignment6]$
```

Thus we get the final output of the problem statement.