

Assignment Day 4

Input Dataset:

hadoop fs -cat /dataset/television.txt

(Displays content of Input Dataset)

ScreenShot:

```
[acadgild@localhost ~]$ hadoop fs -cat /dataset/television.txt
18/07/01 09:12:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar Pradesh|232401|16200
Onida|NA|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
NA|Lucid|18|Uttar Pradesh|232401|16200
Samsung|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
[acadgild@localhost ~]$
```

Task 1:

Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Ans:

Run a Job:

hadoop jar task1na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1

(To run a MR job using created jar file.)

Output Directory:

hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1

(Listing contents of output directory)

hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/part-m-00000

(Displaying content of Output file)

hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/part-m-00000 | wc -l

(Checking count of records after removing lines that contains 'NA' in column)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment1]$ hadoop jar task1na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1
18/07/01 09:16:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/01 09:16:22 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/01 09:16:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/01 09:16:25 INFO input.FileInputFormat: Total input paths to process : 1
18/07/01 09:16:25 INFO mapreduce.JobSubmitter: number of splits:1
18/07/01 09:16:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530329599884_0021
18/07/01 09:16:26 INFO impl.YarnClientImpl: Submitted application application_1530329599884_0021
18/07/01 09:16:26 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1530329599884_0021/
18/07/01 09:16:26 INFO mapreduce.Job: Running job: job_1530329599884_0021
18/07/01 09:16:40 INFO mapreduce.Job: Job job_1530329599884_0021 running in uber mode : false
18/07/01 09:16:40 INFO mapreduce.Job: map 0% reduce 0%
18/07/01 09:16:52 INFO mapreduce.Job: map 100% reduce 0%
18/07/01 09:16:52 INFO mapreduce.Job: Job job_1530329599884_0021 completed successfully
18/07/01 09:16:52 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=107381
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=842
    HDFS: Number of bytes written=662
    HDFS: Number of read operations=5
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=7911
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=7911
    Total vcore-milliseconds taken by all map tasks=7911
    Total megabyte-milliseconds taken by all map tasks=8100864
  Map-Reduce Framework
    Map input records=18
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1
18/07/01 09:18:41 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 2018-07-01 09:16 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 662 2018-07-01 09:16 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/part-m-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/part-m-000000
18/07/01 09:18:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar Pradesh|232401|16200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Samsung|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task1/part-m-000000 | wc -l
16
18/07/01 09:20:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Task 2:

Write a Map Reduce program to calculate the total units sold for each Company.

Ans:

Run a Job:

`hadoop jar task2na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2`

(To run a MR job using created jar file.)

Output Directory:

hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2

(Listing contents of output directory)

hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2/part-r-00000

(Displaying content of Output file)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2
18/07/01 09:22:35 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 2018-07-01 09:22 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        38 2018-07-01 09:22 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2/part-r-00000
18/07/01 09:22:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai      1
Lava      3
Onida     3
Samsung   7
Zen       2
[acadgild@localhost ~]$
```

```
[acadgild@localhost Assignment1]$ hadoop jar task2na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task2
18/07/01 09:21:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/01 09:21:47 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/01 09:21:49 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/01 09:21:50 INFO input.FileInputFormat: Total input paths to process : 1
18/07/01 09:21:50 INFO mapreduce.JobSubmitter: number of splits:1
18/07/01 09:21:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530329599884_0022
18/07/01 09:21:51 INFO impl.YarnClientImpl: Submitted application application_1530329599884_0022
18/07/01 09:21:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1530329599884_0022/
18/07/01 09:21:51 INFO mapreduce.Job: Running job: job_1530329599884_0022
18/07/01 09:22:05 INFO mapreduce.Job: Job job_1530329599884_0022 running in uber mode : false
18/07/01 09:22:05 INFO mapreduce.Job:  map 0% reduce 0%
18/07/01 09:22:16 INFO mapreduce.Job:  map 100% reduce 0%
18/07/01 09:22:27 INFO mapreduce.Job:  map 100% reduce 100%
18/07/01 09:22:28 INFO mapreduce.Job: Job job_1530329599884_0022 completed successfully
18/07/01 09:22:28 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=204
  FILE: Number of bytes written=216415
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=842
  HDFS: Number of bytes written=38
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=7763
  Total time spent by all reduces in occupied slots (ms)=8520
  Total time spent by all map tasks (ms)=7763
  Total time spent by all reduce tasks (ms)=8520
  Total vcore-milliseconds taken by all map tasks=7763
```

Task 3:

Write a Map Reduce program to calculate the total units sold in each state for Onida company.

Ans:

Run a Job:

```
hadoop jar task3na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3
```

(To run a MR job using created jar file.)

Output Directory:

```
hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3
```

(Listing contents of output directory)

```
hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3/part-r-00000
```

(Displaying content of Output file)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3
18/07/01 09:51:03 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 2018-07-01 09:48 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        16 2018-07-01 09:48 /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3/part-r-00000
18/07/01 09:51:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Uttar Pradesh 3
[acadgild@localhost ~]$
```

```
[acadgild@localhost Assignment1]$ hadoop jar task3na.jar /dataset/television.txt /hadoopdata/hdfs/out/hands-on/30Jun18/Assignment1/task3
18/07/01 09:47:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built-in-java classes where applicable
18/07/01 09:47:59 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/01 09:48:01 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application
with ToolRunner to remedy this.
18/07/01 09:48:01 INFO input.FileInputFormat: Total input paths to process : 1
18/07/01 09:48:02 INFO mapreduce.JobSubmitter: number of splits:1
18/07/01 09:48:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530329599884_0023
18/07/01 09:48:02 INFO impl.YarnClientImpl: Submitted application application_1530329599884_0023
18/07/01 09:48:02 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1530329599884_0023/
18/07/01 09:48:02 INFO mapreduce.Job: Running job: job_1530329599884_0023
18/07/01 09:48:16 INFO mapreduce.Job: Job job_1530329599884_0023 running in uber mode : false
18/07/01 09:48:16 INFO mapreduce.Job: map 0% reduce 0%
18/07/01 09:48:29 INFO mapreduce.Job: map 100% reduce 0%
18/07/01 09:48:49 INFO mapreduce.Job: map 100% reduce 100%
18/07/01 09:48:50 INFO mapreduce.Job: Job job_1530329599884_0023 completed successfully
18/07/01 09:48:51 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=66
    FILE: Number of bytes written=216139
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=842
    HDFS: Number of bytes written=16
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=10269
    Total time spent by all reduces in occupied slots (ms)=16595
    Total time spent by all map tasks (ms)=10269
    Total time spent by all reduce tasks (ms)=16595
    Total vcore-milliseconds taken by all map tasks=10269
```

End
