

Session 4 - MR - Introduction

Assignment 1

There are 3 jar files for each of the task. Each task has a separate Package and a separate driver program.

Syntax for the hadoop command used :

Hadoop jar <location of jar in local FS> <HDFS location of input text file> <HDFS location of the output to be created.>

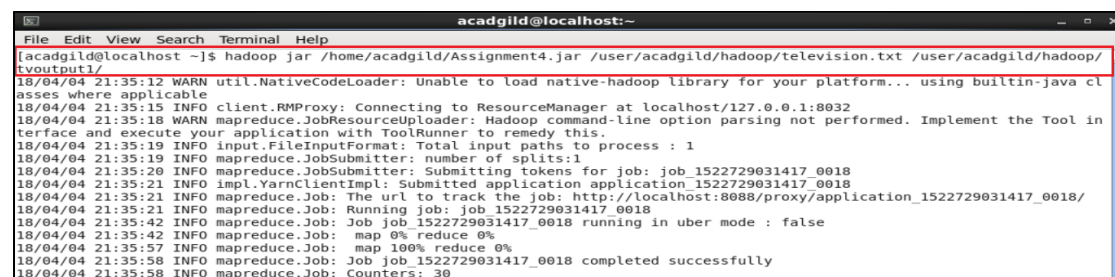
Main class for each jar file is set while exporting the jar in Eclipse.

Task1 : Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Jar name : Assignment4.jar

Hadoop command :

hadoop jar /home/acadgild/Assignment4.jar /user/acadgild/hadoop/television.txt /user/acadgild/hadoop/tvoutput1



```
acadmild@localhost:~  
File Edit View Search Terminal Help  
[acadmild@localhost ~]$ hadoop jar /home/acadmild/Assignment4.jar /user/acadmild/hadoop/television.txt /user/acadmild/hadoop/  
tvoutput1/  
18/04/04 21:35:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl  
asses where applicable  
18/04/04 21:35:15 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/04/04 21:35:18 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in  
terface and execute your application with ToolRunner to remedy this.  
18/04/04 21:35:19 INFO input.FileInputFormat: Total input paths to process : 1  
18/04/04 21:35:19 INFO mapreduce.JobSubmitter: number of splits:1  
18/04/04 21:35:20 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522729031417_0018  
18/04/04 21:35:21 INFO impl.YarnClientImpl: Submitted application application_1522729031417_0018  
18/04/04 21:35:21 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522729031417_0018/  
18/04/04 21:35:21 INFO mapreduce.Job: Running job: job_1522729031417_0018  
18/04/04 21:35:42 INFO mapreduce.Job: Job job_1522729031417_0018 running in uber mode : false  
18/04/04 21:35:42 INFO mapreduce.Job: map 0% reduce 0%  
18/04/04 21:35:57 INFO mapreduce.Job: map 100% reduce 0%  
18/04/04 21:35:58 INFO mapreduce.Job: Job job_1522729031417_0018 completed successfully  
18/04/04 21:35:58 INFO mapreduce.Job: Counters: 30
```

Output folder : tvoutput1

A '_SUCCESS' file is created indicating the successful execution of the job.

The output of the job id stored in the part file.

```
acadgild@localhost:~$ cat /var/spool/mail/acadgild
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop/tvoutput1/
-rw-r--r-- 1 acadgild supergroup 0 2018-04-04 21:35 /user/acadgild/hadoop/tvoutput1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 662 2018-04-04 21:35 /user/acadgild/hadoop/tvoutput1/part-m-000000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/tvoutput1/part-m-000000
18/04/04 21:52:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
Samsun|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
Samsun|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
Samsun|Optima|14|Madhya Pradesh|132401|14200
Samsun|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsun|Super|14|Maharashtra|619082|9200
Samsun|Super|14|Maharashtra|619082|9200
```

The output shows that the tow records containing 'NA' entries are not included.

Task2 : Write a Map Reduce program to calculate the total units sold for each Company.

Jar name : Assignment4-2.jar

Hadoop command :

`hadoop jar /home/acadgild/Assignment4-2.jar /user/acadgild/hadoop/television.txt /user/acadgild/hadoop/tvoutput2`

```
acadgild@localhost:~$ hadoop jar /home/acadgild/Assignment4-2.jar /user/acadgild/hadoop/television.txt /user/acadgild/hadoop/tvoutput2/
18/04/04 23:38:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/04/04 23:38:55 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/04 23:38:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/04/04 23:38:58 INFO input.FileInputFormat: Total input paths to process : 1
18/04/04 23:38:59 INFO mapreduce.JobSubmitter: number of splits:1
18/04/04 23:38:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522729031417_0019
18/04/04 23:39:00 INFO impl.YarnClientImpl: Submitted application application_1522729031417_0019
18/04/04 23:39:00 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522729031417_0019/
18/04/04 23:39:00 INFO mapreduce.Job: Running job: job_1522729031417_0019
18/04/04 23:39:19 INFO mapreduce.Job: Job job_1522729031417_0019 running in uber mode : false
18/04/04 23:39:20 INFO mapreduce.Job: map 0% reduce 0%
18/04/04 23:39:33 INFO mapreduce.Job: map 100% reduce 0%
18/04/04 23:39:46 INFO mapreduce.Job: map 100% reduce 100%
18/04/04 23:39:47 INFO mapreduce.Job: Job job_1522729031417_0019 completed successfully
18/04/04 23:39:47 INFO mapreduce.Job: Counters: 49
```

Output folder : tvoutput2

A '_SUCCESS' file is created indicating the successful execution of the job.

The output of the job id stored in the part file.

```
acadgild@localhost:~$ hadoop fs -ls /user/acadgild/hadoop/tvoutput2/
18/04/04 23:41:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-04-04 23:39 /user/acadgild/hadoop/tvoutput2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 43 2018-04-04 23:39 /user/acadgild/hadoop/tvoutput2/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/tvoutput2/part-r-000000
18/04/04 23:41:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Akai 1
Lava 3
NA 1
Onida 4
Samsung 7
Zen 2
```

The total units sold for each company is the number of times the record for that company occurs in the file.

The output shows the number of units sold for each company.

Task3 : Write a Map Reduce program to calculate the total units sold in each state for Onida company.

Jar name : Assignment4-3.jar

Hadoop command :

`hadoop jar /home/acadgild/Assignment4-3.jar /user/acadgild/hadoop/television.txt /user/acadgild/hadoop/tvoutput3`

```
acadgild@localhost:~$ hadoop jar /home/acadgild/Assignment4-3.jar /user/acadgild/hadoop/television.txt /user/acadgild/hadoop/tvoutput3/
18/04/04 23:43:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/04/04 23:43:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/04 23:43:34 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/04/04 23:43:35 INFO input.FileInputFormat: Total input paths to process : 1
18/04/04 23:43:35 INFO mapreduce.JobSubmitter: number of splits:1
18/04/04 23:43:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522729031417_0020
18/04/04 23:43:36 INFO impl.YarnClientImpl: Submitted application application_1522729031417_0020
18/04/04 23:43:36 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522729031417_0020/
18/04/04 23:43:36 INFO mapreduce.Job: Running job: job_1522729031417_0020
18/04/04 23:43:49 INFO mapreduce.Job: Job job_1522729031417_0020 running in uber mode : false
18/04/04 23:43:49 INFO mapreduce.Job: map 0% reduce 0%
18/04/04 23:43:58 INFO mapreduce.Job: map 100% reduce 0%
18/04/04 23:44:09 INFO mapreduce.Job: map 100% reduce 100%
18/04/04 23:44:11 INFO mapreduce.Job: Job job_1522729031417_0020 completed successfully
18/04/04 23:44:12 INFO mapreduce.Job: Counters: 49
```

Output folder : tvoutput3

A ‘_SUCCESS’ file is created indicating the successful execution of the job.

The output of the job id stored in the part file.

```
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=733
File Output Format Counters
  Bytes Written=41
I
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop/tvoutput3/
18/04/04 23:46:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-04-04 23:44 /user/acadgild/hadoop/tvoutput3/_SUCCESS
-rw-r--r-- 1 acadgild supergroup          41 2018-04-04 23:44 /user/acadgild/hadoop/tvoutput3/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/tvoutput3/part-r-000000
18/04/04 23:46:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Onida : Kerala 1
Onida : Uttar Pradesh 3
```