Running wordcount program with Local Hadoop

To run wordcount sample program provided by Hadoop,

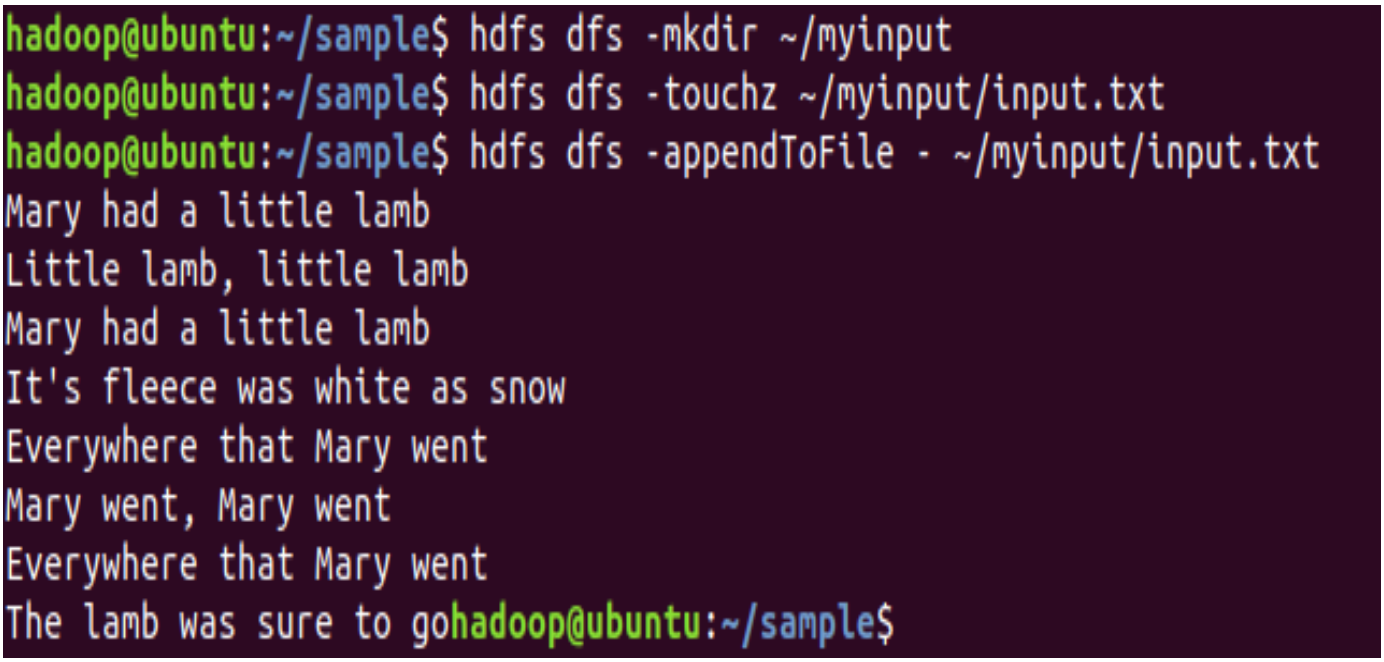
Hadoop can run program using mapreduce and can be run using below command,

1. Specify jar filename
2. Input folder having input file to run
3. Output folder where output will be created

$hadoop jar jarname –classpath input output

Map reduce can only run jar files.

Step 1: Create input file with some content,



Step 2: Run wordcount jar sample program provided by Hadoop

```
hadoop@ubuntu:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.1.4.jar wordcount ~/myinput ~/myoutput
2020-12-08 19:35:25,579 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2020-12-08 19:35:26,191 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging
2020-12-08 19:35:26,398 INFO input.FileInputFormat: Total input files to process : 1
2020-12-08 19:35:26,514 INFO mapreduce.JobSubmitter: number of splits:1
2020-12-08 19:35:26,701 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1607229528034_0001
2020-12-08 19:35:26,702 INFO mapreduce.JobSubmitter: Executing with tokens: []
2020-12-08 19:35:26,892 INFO conf.Configuration: resource-types.xml not found
2020-12-08 19:35:26,892 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2020-12-08 19:35:27,522 INFO impl.YarnClientImpl: Submitted application application_1607229528034_0001
2020-12-08 19:35:27,579 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1607229528034_0001/
2020-12-08 19:35:27,580 INFO mapreduce.Job: Running job: job_1607229528034_0001
2020-12-08 19:35:35,730 INFO mapreduce.Job: Job job_1607229528034_0001 running in uber mode : false
2020-12-08 19:35:35,733 INFO mapreduce.Job:  map 0% reduce 0%
```

Step 3: Get the output from output folder,

Output is saved in file part-r-00000

```
hadoop@ubuntu:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -ls ~/myoutput
Found 2 items
-rw-r--r--   1 hadoop supergroup          0 2020-12-08 19:35 /home/hadoop/myoutput/_SUCCESS
-rw-r--r--   1 hadoop supergroup        150 2020-12-08 19:35 /home/hadoop/myoutput/part-r-00000
hadoop@ubuntu:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -cat ~/myoutput/part*
Everywhere      2
It's    1
Little  1
Mary    6
The     1
a       2
as      1
fleece  1
go      1
had     2
lamb    4
lamb,   1
little  3
snow    1
sure    1
that    2
to      1
was     2
went    3
went,   1
white   1
```