

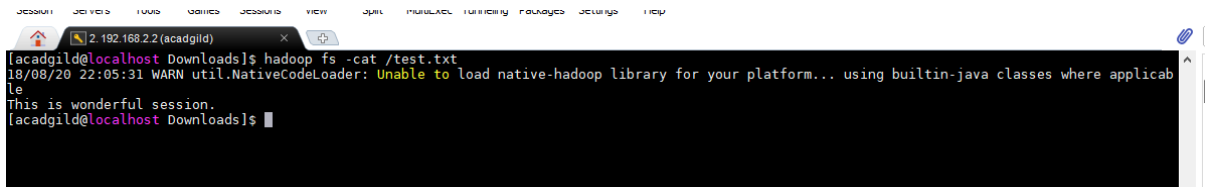
Problem Statement

Task 1:

Execute **WordMedian** , **WordMean** , **WordStandardDeviation** programs using **hadoop-mapreduce-examples-2.9.0.jar** file present in your AcadGild VM.

Solution

Command to show the contents of the input file used to test the above mentioned programs present in - /home/acadgild/



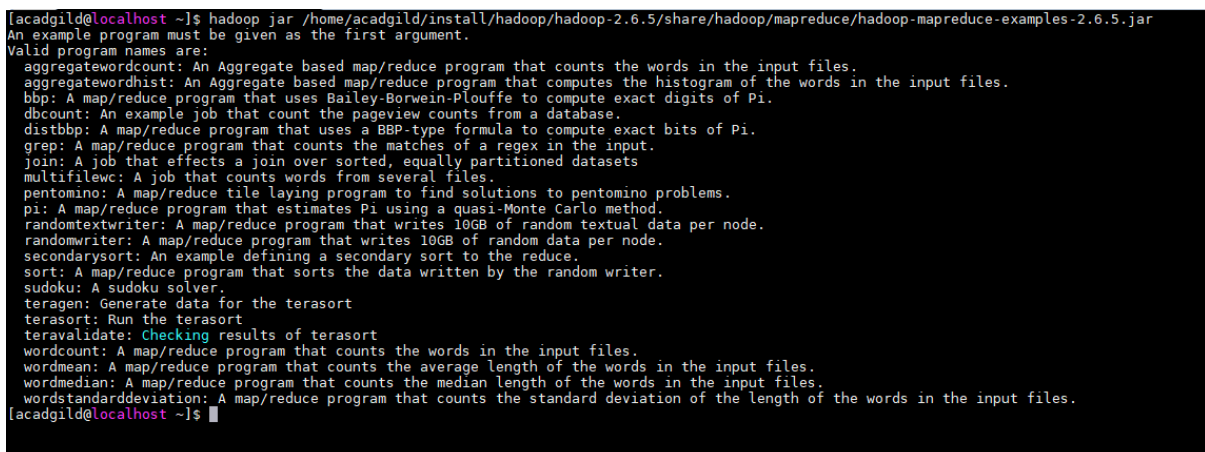
```
acadgild@localhost: Downloads | 2.192.168.2.2 (acadgild)
[acadgild@localhost Downloads]$ hadoop fs -cat /test.txt
18/08/20 22:05:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is wonderful session.
[acadgild@localhost Downloads]$
```

JAR file : **hadoop-mapreduce-examples-2.6.5.jar** (present in Acadgild VM)

JAR file location:

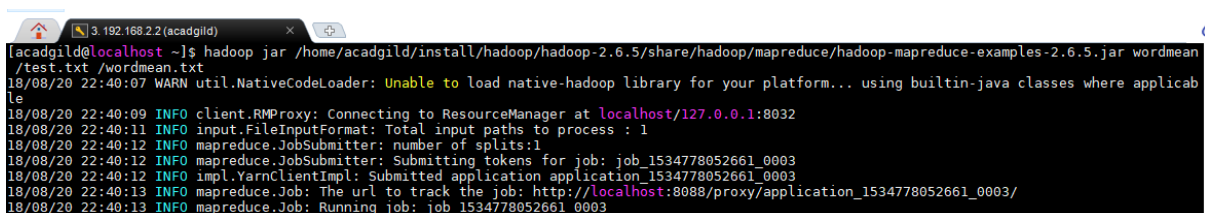
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar

Execute the below command to get the description of the different programs present in the jar. This contains the **WordMedian** , **WordMean** , **WordStandardDeviation** programs shown in the below screenshot



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar
An example program must be given as the first argument.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
dbcount: An example job that count the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifilewc: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
terasgen: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
wordmean: A map/reduce program that counts the average length of the words in the input files.
wordmedian: A map/reduce program that counts the median length of the words in the input files.
wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[acadgild@localhost ~]$
```

- To test the **wordmean** program run the below command



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean
/test.txt /wordmean.txt
18/08/20 22:40:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/20 22:40:09 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/20 22:40:11 INFO input.FileInputFormat: Total input paths to process : 1
18/08/20 22:40:12 INFO mapreduce.JobSubmitter: number of splits:1
18/08/20 22:40:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1534778052661_0003
18/08/20 22:40:12 INFO impl.YarnClientImpl: Submitted application application_1534778052661_0003
18/08/20 22:40:13 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1534778052661_0003/
18/08/20 22:40:13 INFO mapreduce.Job: Running job: job_1534778052661_0003
```

Output of the program is displayed in the terminal

```
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=27
File Output Format Counters
  Bytes Written=18
The mean is: 5.75
[acadgild@localhost ~]$
```

Command to check the output files (Success and number of reducers = 1)

```
[acadgild@localhost ~]$ hadoop fs -ls /wordmean.txt
18/08/20 22:44:59 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 2018-08-20 22:40 /wordmean.txt/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        18 2018-08-20 22:40 /wordmean.txt/part-r-00000
[acadgild@localhost ~]$
```

Command to show the content of the output reducer file. The output file shows the number of words and the length of the input file

```
[acadgild@localhost ~]$ hadoop fs -cat /wordmean.txt/part-r-00000
18/08/20 22:47:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 4
length 23
[acadgild@localhost ~]$
```

- To test the **wordmedian** program run the below command

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /wordmedian.txt
18/08/20 22:50:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/20 22:50:30 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/20 22:50:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
```

Output of the program is displayed in the terminal

```
Total committed heap usage (bytes)=14003
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=27
File Output Format Counters
  Bytes Written=16
The median is: 4
[acadgild@localhost ~]$
```

Command to check the output file(Success and number of reducers = 1)

```
3.192.168.2.2 (acadgild)
[acadgild@localhost ~]$ hadoop fs -ls /wordmedian.txt
18/08/20 22:52:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicab
le
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-08-20 22:51 /wordmedian.txt/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        16 2018-08-20 22:51 /wordmedian.txt/part-r-00000
[acadgild@localhost ~]$
```

Command to show the content of the output reducer file. The output file shows the length of the words and its occurrence

```
3.192.168.2.2 (acadgild)
[acadgild@localhost ~]$ hadoop fs -cat /wordmedian.txt/part-r-00000
le
2      1
4      1
8      1
9      1
[acadgild@localhost ~]$
```

- To test the **wordstandarddeviation** program run the below command

```
3.192.168.2.2 (acadgild)
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstan
darddeviation /test.txt /wordstddeviation.txt
18/08/20 22:55:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicab
le
18/08/20 22:55:10 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/20 22:55:12 INFO input.FileInputFormat: Total input paths to process : 1
18/08/20 22:55:13 INFO mapreduce.JobSubmitter: number of splits:1
18/08/20 22:55:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1534778052661_0005
18/08/20 22:55:14 INFO impl.YarnClientImpl: Submitted application application_1534778052661_0005
```

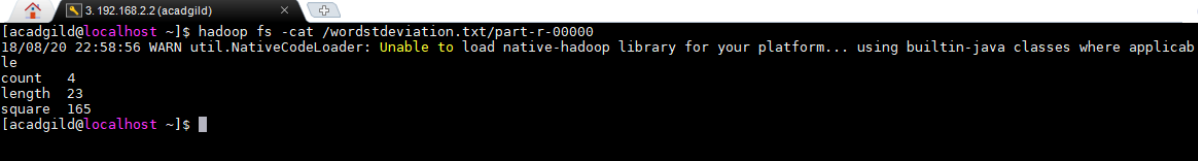
Output of the program is displayed in the terminal

```
total committed heap usage (bytes)=1
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=27
File Output Format Counters
Bytes Written=29
The standard deviation is: 2.8613807855648994
[acadgild@localhost ~]$
```

Command to check the output files(Success and number of reducers = 1)

```
3.192.168.2.2 (acadgild)
[acadgild@localhost ~]$ hadoop fs -ls /wordstddeviation.txt
18/08/20 22:58:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicab
le
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-08-20 22:55 /wordstddeviation.txt/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        29 2018-08-20 22:55 /wordstddeviation.txt/part-r-00000
[acadgild@localhost ~]$
```

Command to show the content of the output reducer file.



```
[acadgild@localhost ~]$ hadoop fs -cat /wordstdeviation.txt/part-r-00000
18/08/20 22:58:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicab
le
count    4
length   23
square   165
[acadgild@localhost ~]$
```

The image shows a terminal window with a dark background. The prompt is [acadgild@localhost ~]\$. The command hadoop fs -cat /wordstdeviation.txt/part-r-00000 is entered. The output shows a warning message from WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicab, followed by the content of the file: le, count 4, length 23, and square 165. The prompt returns to [acadgild@localhost ~]\$.