Session 3: YARN

Assignment 1

TASK 1:

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

Refer path below.
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce

EXECUTION:

We created an input file "input.txt" in local file system and moved the file in to HDFS folder "/mapreduce/input" by executing HDFS command,

hadoop fs –copyFromLocal /home/acadgild/workspace/input.txt /mapreduce/input

Then in /Hadoop/bin terminal, we call mapreduce command to calculate wordmedian, wordmean and wordstandarddeviation.

- **wordmedian**

  Mapreduce command to execute:

  <mark>hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /mapreduce/input /mapreduce/output</mark>

  ```
  hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /
  mapreduce/input /mapreduce/output
  ```

  Output:

  ```
                    CONNECTION=0
                    IO_ERROR=0
                    WRONG_LENGTH=0
                    WRONG_MAP=0
                    WRONG_REDUCE=0
          File Input Format Counters
                    Bytes Read=201
          File Output Format Counters
                    Bytes Written=34
  The median is: 5
  You have new mail in /var/spool/mail/acadgild
  [acadgild@localhost bin]$
  ```

  ```
  You have new mail in /var/spool/mail/acadgild
  [acadgild@localhost mapreduce]$ hadoop fs -ls /mapreduce/output
  19/01/13 07:10:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pl
  atform... using builtin-java classes where applicable
  Found 2 items
  -rw-r--r--   1 acadgild supergroup          0 2019-01-13 07:05 /mapreduce/output/_SUCCESS
  -rw-r--r--   1 acadgild supergroup         34 2019-01-13 07:05 /mapreduce/output/part-r-00000
  You have new mail in /var/spool/mail/acadgild
  [acadgild@localhost mapreduce]$
                               File Output Format Counters
  ```

Open the result file "part-r-00000" and the result is as below,

```
[acadgild@localhost mapreduce]$ hadoop fs -cat /mapreduce/output/part-r-00000
19/01/13 07:11:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pl
atform... using builtin-java classes where applicable
3       1
4       7
5       10
6       5
7       2
8       2
9       1
10      1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

- **Wordmean**

Mapreduce command to execute:

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

```
[acadgild@localhost bin]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/
hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /mapreduce/input /mapred
uce/wordmean/output
19/01/13 07:24:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
 your platform... using builtin-java classes where applicable
19/01/13 07:24:51 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.
0.0.1:8032
19/01/13 07:24:53 INFO input.FileInputFormat: Total input paths to process : 1
19/01/13 07:24:53 INFO mapreduce.JobSubmitter: number of splits:1
19/01/13 07:24:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15473344
68299_0003
19/01/13 07:24:53 INFO impl.YarnClientImpl: Submitted application application_15473344
68299_0003
19/01/13 07:24:54 INFO mapreduce.Job: The url to track the job: http://localhost:8088/
proxy/application_1547334468299_0003/
19/01/13 07:24:54 INFO mapreduce.Job: Running job: job_1547334468299_0003
```

Output:

```
        Total committed heap usage (bytes)=170044
        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=201
        File Output Format Counters
                Bytes Written=20
The mean is: 5.517241379310345
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$
```

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat /mapreduce/wordmean/output/part-r-00000
19/01/13 07:27:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pl
atform... using builtin-java classes where applicable
count    29
length   160
[acadgild@localhost mapreduce]$
```

- **Wordstandarddeviation**

Mapreduce command to execute:

hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /mapreduce/input /mapreduce/standarddeviation/output

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/
hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /mapreduce/
input /mapreduce/standarddeviation/output
19/01/13 07:30:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
 your platform... using builtin-java classes where applicable
19/01/13 07:30:04 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.
0.0.1:8032
19/01/13 07:30:06 INFO input.FileInputFormat: Total input paths to process : 1
19/01/13 07:30:06 INFO mapreduce.JobSubmitter: number of splits:1
19/01/13 07:30:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15473344
68299_0004
19/01/13 07:30:06 INFO impl.YarnClientImpl: Submitted application application_15473344
68299_0004
19/01/13 07:30:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/
proxy/application_1547334468299_0004/
19/01/13 07:30:06 INFO mapreduce.Job: Running job: job_1547334468299_0004
```

Output:

```
                    WRONG_LENGTH=0
                    WRONG_MAP=0
                    WRONG_REDUCE=0
          File Input Format Counters
                    Bytes Read=201
          File Output Format Counters
                    Bytes Written=31
The standard deviation is: 1.6107545812588493
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$
```

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat /mapreduce/standarddeviation/output/part-r-0000
0
19/01/13 07:32:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pl
atform... using builtin-java classes where applicable
count    29
length   160
square   958
[acadgild@localhost mapreduce]$
```