# Session 3: YARN
## Assignment

**Note:** The VM available with me is having hadoop 2.6.5 and not the 2.9.0 so executing hadoop-mapreduce-examples 2.6.5 jar instead of 2.9.0 jars.

**Task 1:**

- **Execute WordMean Job**

**Command -** yarn jar **/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar** **wordmean** /**input** /**output1**

The **first highlighted** part indicates the path of the jar file to be executed.

The **second highlighted** part indicates the driver (main) class of the jar file. If main class is not defined while creating a jar file we have to specify the class name in command else we can omit the same**.**

The **third highlighted** part indicates input folder to be read.

The **forth highlighted** part indicates output folder to be written with output of job.

Below screen shot after executing the command

```
[acadgild@localhost ~]$ yarn jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /input /output1
18/03/24 12:21:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/24 12:21:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/24 12:21:49 INFO input.FileInputFormat: Total input paths to process : 1
18/03/24 12:21:50 INFO mapreduce.JobSubmitter: number of splits:1
18/03/24 12:21:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521859699712_0002
18/03/24 12:21:51 INFO impl.YarnClientImpl: Submitted application application_1521859699712_0002
18/03/24 12:21:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521859699712_0002/
18/03/24 12:21:51 INFO mapreduce.Job: Running job: job_1521859699712_0002
18/03/24 12:22:08 INFO mapreduce.Job: Job job_1521859699712_0002 running in uber mode : false
18/03/24 12:22:08 INFO mapreduce.Job:  map 0% reduce 0%
18/03/24 12:22:18 INFO mapreduce.Job:  map 100% reduce 0%
18/03/24 12:22:31 INFO mapreduce.Job:  map 100% reduce 100%
18/03/24 12:22:32 INFO mapreduce.Job: Job job_1521859699712_0002 completed successfully
18/03/24 12:22:32 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=215671
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=196
                HDFS: Number of bytes written=19
                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)=8357
                Total time spent by all reduces in occupied slots (ms)=9191
                Total time spent by all map tasks (ms)=8357
                Total time spent by all reduce tasks (ms)=9191
                Total vcore-milliseconds taken by all map tasks=8357
                Total vcore-milliseconds taken by all reduce tasks=9191
                Total megabyte-milliseconds taken by all map tasks=8557568
                Total megabyte-milliseconds taken by all reduce tasks=9411584
        Map-Reduce Framework
                Map input records=1
                Map output records=34
                Map output bytes=493
                Map output materialized bytes=39
                Input split bytes=101
                Combine input records=34
                Combine output records=2
                Reduce input groups=2
                Reduce shuffle bytes=39
                Reduce input records=2
```

```
                Spilled Records=4
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=230
                CPU time spent (ms)=2400
                Physical memory (bytes) snapshot=295473152
                Virtual memory (bytes) snapshot=4118224896
                Total committed heap usage (bytes)=170004480
        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=95
        File Output Format Counters
                Bytes Written=19
The mean is: 4.588235294117647
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Folder got created as output1

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/03/24 12:29:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 9 items
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 22:22 /Prachi
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 09:38 /hadoopdata
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 19:15 /input
drwxr-xr-x   - acadgild supergroup          0 2018-03-05 09:06 /output
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:22 /output1
drwxr-xr-x   - acadgild supergroup          0 2018-03-17 09:59 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx---   - acadgild supergroup          0 2018-03-05 08:40 /tmp
drwxr-xr-x   - acadgild supergroup          0 2018-03-18 07:10 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls /output1
18/03/24 12:30:00 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 2018-03-24 12:22 /output1/_SUCCESS
-rw-r--r--   1 acadgild supergroup         19 2018-03-24 12:22 /output1/part-r-00000
[acadgild@localhost ~]$
```

Output of word mean

```
[acadgild@localhost ~]$ hadoop fs -cat /output1/part-r-00000
18/03/24 12:30:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count   17
length  78
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

**Task 2:**

**Execute WordMedian**

**Command -** yarn jar **/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /input /output2**

Note: The color used for highlights indicates same meaning as explained above.

```
[acadgild@localhost ~]$ yarn jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /input /output2
18/03/24 12:31:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/24 12:32:00 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/24 12:32:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner
edy this.
18/03/24 12:32:03 INFO input.FileInputFormat: Total input paths to process : 1
18/03/24 12:32:04 INFO mapreduce.JobSubmitter: number of splits:1
18/03/24 12:32:04 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521859699712_0003
18/03/24 12:32:05 INFO impl.YarnClientImpl: Submitted application application_1521859699712_0003
18/03/24 12:32:05 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521859699712_0003/
18/03/24 12:32:05 INFO mapreduce.Job: Running job: job_1521859699712_0003
18/03/24 12:32:21 INFO mapreduce.Job: Job job_1521859699712_0003 running in uber mode : false
18/03/24 12:32:21 INFO mapreduce.Job:  map 0% reduce 0%
18/03/24 12:32:31 INFO mapreduce.Job:  map 100% reduce 0%
18/03/24 12:32:44 INFO mapreduce.Job:  map 100% reduce 100%
18/03/24 12:32:44 INFO mapreduce.Job: Job job_1521859699712_0003 completed successfully
18/03/24 12:32:45 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=66
                FILE: Number of bytes written=215461
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=196
                HDFS: Number of bytes written=24
                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)=8370
                Total time spent by all reduces in occupied slots (ms)=9433
                Total time spent by all map tasks (ms)=8370
                Total time spent by all reduce tasks (ms)=9433
                Total vcore-milliseconds taken by all map tasks=8370
                Total vcore-milliseconds taken by all reduce tasks=9433
                Total megabyte-milliseconds taken by all map tasks=8570880
                Total megabyte-milliseconds taken by all reduce tasks=9659392
        Map-Reduce Framework
                Map input records=1
                Map output records=17
                Map output bytes=136
                Map output materialized bytes=66
                Input split bytes=101
                Combine input records=17
                Combine output records=6
```

```
                Reduce output records=6
                Spilled Records=12
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=249
                CPU time spent (ms)=2450
                Physical memory (bytes) snapshot=295727104
                Virtual memory (bytes) snapshot=4118224896
                Total committed heap usage (bytes)=170004480
        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=95
        File Output Format Counters
                Bytes Written=24
The median is: 4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Output2 folder is created

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/03/24 12:34:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 10 items
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 22:22 /Prachi
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 09:38 /hadoopdata
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 19:15 /input
drwxr-xr-x   - acadgild supergroup          0 2018-03-05 09:06 /output
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:22 /output1
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:32 /output2
drwxr-xr-x   - acadgild supergroup          0 2018-03-17 09:59 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx---   - acadgild supergroup          0 2018-03-05 08:40 /tmp
drwxr-xr-x   - acadgild supergroup          0 2018-03-18 07:10 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls /output2
18/03/24 12:35:01 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 2018-03-24 12:32 /output2/_SUCCESS
-rw-r--r--   1 acadgild supergroup         24 2018-03-24 12:32 /output2/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /output2/part-r-00000
18/03/24 12:35:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2       4
3       1
4       4
5       3
7       4
8       1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

**Task 3:**

**Execute WordStandardDeviation**

**Command -** yarn jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /input /output3

Note: The color used for highlights indicates same meaning as explained above.

```
[acadgild@localhost ~]$ yarn jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /input /output3
18/03/24 12:36:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/24 12:36:48 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/24 12:36:51 INFO input.FileInputFormat: Total input paths to process : 1
18/03/24 12:36:52 INFO mapreduce.JobSubmitter: number of splits:1
18/03/24 12:36:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521859699712_0004
18/03/24 12:36:53 INFO impl.YarnClientImpl: Submitted application application_1521859699712_0004
18/03/24 12:36:53 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521859699712_0004/
18/03/24 12:36:53 INFO mapreduce.Job: Running job: job_1521859699712_0004
18/03/24 12:37:08 INFO mapreduce.Job: Job job_1521859699712_0004 running in uber mode : false
18/03/24 12:37:08 INFO mapreduce.Job:  map 0% reduce 0%
18/03/24 12:37:19 INFO mapreduce.Job:  map 100% reduce 0%
18/03/24 12:37:32 INFO mapreduce.Job:  map 100% reduce 100%
18/03/24 12:37:33 INFO mapreduce.Job: Job job_1521859699712_0004 completed successfully
18/03/24 12:37:33 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=215869
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=196
                HDFS: Number of bytes written=30
                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)=8547
                Total time spent by all reduces in occupied slots (ms)=9346
                Total time spent by all map tasks (ms)=8547
                Total time spent by all reduce tasks (ms)=9346
                Total vcore-milliseconds taken by all map tasks=8547
                Total vcore-milliseconds taken by all reduce tasks=9346
                Total megabyte-milliseconds taken by all map tasks=8752128
                Total megabyte-milliseconds taken by all reduce tasks=9570304
        Map-Reduce Framework
                Map input records=1
                Map output records=51
                Map output bytes=748
                Map output materialized bytes=56
                Input split bytes=101
                Combine input records=51
                Combine output records=3
                Reduce input groups=3
                Reduce shuffle bytes=56
                Reduce input records=3
```

```
                Spilled Records=6
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=222
                CPU time spent (ms)=2510
                Physical memory (bytes) snapshot=295038976
                Virtual memory (bytes) snapshot=4118224896
                Total committed heap usage (bytes)=170004480
        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=95
        File Output Format Counters
                Bytes Written=30
The standard deviation is: 1.9721240722635547
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Output3 folder is created

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/03/24 12:38:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 11 items
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 22:22 /Prachi
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 09:38 /hadoopdata
drwxr-xr-x   - acadgild supergroup          0 2018-03-04 19:15 /input
drwxr-xr-x   - acadgild supergroup          0 2018-03-05 09:06 /output
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:22 /output1
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:32 /output2
drwxr-xr-x   - acadgild supergroup          0 2018-03-24 12:37 /output3
drwxr-xr-x   - acadgild supergroup          0 2018-03-17 09:59 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx---   - acadgild supergroup          0 2018-03-05 08:40 /tmp
drwxr-xr-x   - acadgild supergroup          0 2018-03-18 07:10 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls /output3
18/03/24 12:39:06 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 2018-03-24 12:37 /output3/_SUCCESS
-rw-r--r--   1 acadgild supergroup         30 2018-03-24 12:37 /output3/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /output3/part-r-00000
18/03/24 12:39:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count   17
length  78
square  424
[acadgild@localhost ~]$
```