

Assignment 4

Part 2

Code in Package- com.hadoop.assignment4.part_2

Output Screenshot--

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/Downloads/A4.jar com.hadoop.assignment4.part_2.MinMaxDriver /nyse_merged /Assignment4
19/06/26 09:06:59 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 09:07:00 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/06/26 09:07:01 INFO Input.FileInputFormat: Total input files to process : 2
19/06/26 09:07:01 INFO mapreduce.JobSubmitter: number of splits:8
19/06/26 09:07:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0001
19/06/26 09:07:02 INFO Impl.YarnClientImpl: Submitted application application_1561554348150_0001
19/06/26 09:07:02 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8088/proxy/application_1561554348150_0001/
19/06/26 09:07:02 INFO mapreduce.Job: Running job: job_1561554348150_0001
19/06/26 09:07:10 INFO mapreduce.Job: Job job_1561554348150_0001 running in uber mode : false
19/06/26 09:07:16 INFO mapreduce.Job: map 0% reduce 0%
19/06/26 09:07:55 INFO mapreduce.Job: map 3% reduce 0%
19/06/26 09:07:57 INFO mapreduce.Job: map 6% reduce 0%
19/06/26 09:07:58 INFO mapreduce.Job: map 9% reduce 0%
19/06/26 09:08:02 INFO mapreduce.Job: map 12% reduce 0%
19/06/26 09:08:03 INFO mapreduce.Job: map 13% reduce 0%
19/06/26 09:08:04 INFO mapreduce.Job: map 17% reduce 0%
19/06/26 09:08:08 INFO mapreduce.Job: map 20% reduce 0%
```

```
19/06/26 09:09:46 INFO mapreduce.Job: map 100% reduce 71%
19/06/26 09:09:52 INFO mapreduce.Job: map 100% reduce 83%
19/06/26 09:09:58 INFO mapreduce.Job: map 100% reduce 100%
19/06/26 09:09:59 INFO mapreduce.Job: Job job_1561554348150_0001 completed successfully
19/06/26 09:09:59 INFO mapreduce.Job: Counters: 50
File System Counters
FILE: Number of bytes read=1396952508
FILE: Number of bytes written=2096853878
bytes written=270505
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part-r-00000 | head
AA dateMaxStockVolume='2009-03-19', dateMinStockVolume='1965-06-21', maxStockPriceClose=94.5
AAI dateMaxStockVolume='2009-10-08', dateMinStockVolume='1994-09-14', maxStockPriceClose=56.88
AAN dateMaxStockVolume='2009-06-08', dateMinStockVolume='1996-07-05', maxStockPriceClose=34.42
AAP dateMaxStockVolume='2006-06-29', dateMinStockVolume='2002-02-19', maxStockPriceClose=83.03
AAR dateMaxStockVolume='2003-04-15', dateMinStockVolume='2000-07-27', maxStockPriceClose=25.2
AAV dateMaxStockVolume='2006-01-24', dateMinStockVolume='2005-10-10', maxStockPriceClose=21.22
AB dateMaxStockVolume='2009-04-23', dateMinStockVolume='1989-05-26', maxStockPriceClose=93.8
ABA dateMaxStockVolume='2006-07-10', dateMinStockVolume='2007-07-05', maxStockPriceClose=27.55
ABB dateMaxStockVolume='2008-02-13', dateMinStockVolume='2002-07-09', maxStockPriceClose=32.95
ABC dateMaxStockVolume='2003-09-19', dateMinStockVolume='1996-02-12', maxStockPriceClose=83.77
cat: Unable to write to output stream.
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$
```

Part 3

Code in Package- com.hadoop.assignment4.part_3

The running time of this part was much more than running time of part_2.

Output Screenshot--

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/downloads/A4.jar com.hadoop.assignment4.part_3.StockDriver /nyse_merged /Assignment4/part_3
19/06/26 09:16:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 09:16:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/06/26 09:16:03 INFO InputFileInputFormat: Total input files to process : 2
19/06/26 09:16:03 INFO mapreduce.JobSubmitter: number of splits:9
19/06/26 09:16:04 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0002
19/06/26 09:16:04 INFO Impl.VarnClientImpl: Submitted application application_1561554348150_0002
19/06/26 09:16:04 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8088/proxy/application_1561554348150_0002/
19/06/26 09:16:04 INFO mapreduce.Job: Running Job: job_1561554348150_0002
19/06/26 09:16:11 INFO mapreduce.Job: Job job_1561554348150_0002 running in uber mode : false
19/06/26 09:16:11 INFO mapreduce.Job: map 0% reduce 0%
19/06/26 09:16:45 INFO mapreduce.Job: map 1% reduce 0%
19/06/26 09:16:56 INFO mapreduce.Job: map 2% reduce 0%
19/06/26 09:17:07 INFO mapreduce.Job: map 3% reduce 0%
19/06/26 09:17:13 INFO mapreduce.Job: map 4% reduce 0%
19/06/26 09:17:16 INFO mapreduce.Job: map 5% reduce 0%
19/06/26 09:17:17 INFO mapreduce.Job: map 6% reduce 0%
19/06/26 09:17:21 INFO mapreduce.Job: map 7% reduce 0%
19/06/26 09:17:24 INFO mapreduce.Job: map 8% reduce 0%
```

```
19/06/26 09:18:13 INFO mapreduce.Job: map 15% reduce 0%
19/06/26 09:18:14 INFO mapreduce.Job: map 75% reduce 0%
19/06/26 09:18:40 INFO mapreduce.Job: map 96% reduce 25%
19/06/26 09:18:41 INFO mapreduce.Job: map 100% reduce 25%
19/06/26 09:18:42 INFO mapreduce.Job: map 100% reduce 100%
19/06/26 09:18:44 INFO mapreduce.Job: Job job_1561554348150_0002 completed successfully
19/06/26 09:18:44 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=526214
    FILE: Number of bytes written=2215835
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1022196666
    HDFS: Number of bytes written=273358
    HDFS: Number of read operations=27
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=9
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part_3/part-r-00000 | head
AA      dateMaxStockVolume: '2009-03-19', dateMinStockVolume: '1965-06-21', maxStockPriceClose: 94.5'
AAI      dateMaxStockVolume: '2009-10-08', dateMinStockVolume: '1994-09-14', maxStockPriceClose: 56.88'
AAN      dateMaxStockVolume: '2009-06-08', dateMinStockVolume: '1999-11-26', maxStockPriceClose: 34.42'
AAP      dateMaxStockVolume: '2006-06-29', dateMinStockVolume: '2002-02-19', maxStockPriceClose: 83.03'
AAR      dateMaxStockVolume: '2003-04-15', dateMinStockVolume: '2000-07-27', maxStockPriceClose: 25.2'
AAV      dateMaxStockVolume: '2006-01-24', dateMinStockVolume: '2004-08-02', maxStockPriceClose: 21.22'
AB      dateMaxStockVolume: '2009-04-23', dateMinStockVolume: '1989-05-26', maxStockPriceClose: 93.8'
ABA      dateMaxStockVolume: '2006-07-10', dateMinStockVolume: '2007-07-05', maxStockPriceClose: 27.55'
ABB      dateMaxStockVolume: '2008-02-13', dateMinStockVolume: '2002-07-09', maxStockPriceClose: 32.95'
ABC      dateMaxStockVolume: '2003-09-19', dateMinStockVolume: '1996-02-12', maxStockPriceClose: 83.77'
cat: Unable to write to output stream
```

Part 4

Code in Package- com.hadoop.assignment4.part_4

Output Screenshot--

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/Downloads/A4.jar com.hadoop.assignment4.part_4.SecondarySortDriver /nyse_merged /Assignment4/part_4
19/06/26 10:58:56 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 10:58:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
19/06/26 10:58:57 INFO InputFileInputFormat: Total input files to process : 1
19/06/26 10:58:58 INFO mapreduce.JobSubmitter: number of splits:4
19/06/26 10:58:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0007
19/06/26 10:58:59 INFO Impl.YarnClientImpl: Submitted application application_1561554348150_0007
19/06/26 10:58:59 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8088/proxy/application_1561554348150_0007/
19/06/26 10:58:59 INFO mapreduce.Job: Running job: job_1561554348150_0007
19/06/26 10:59:06 INFO mapreduce.Job: Job Job_1561554348150_0007 running in uber mode : false
19/06/26 10:59:06 INFO mapreduce.Job:  map 0% reduce 0%
19/06/26 10:59:30 INFO mapreduce.Job:  map 67% reduce 0%
19/06/26 11:04:54 INFO mapreduce.Job:  map 75% reduce 0%
19/06/26 11:05:19 INFO mapreduce.Job:  map 75% reduce 8%
19/06/26 11:05:38 INFO mapreduce.Job:  map 83% reduce 8%
19/06/26 11:05:38 INFO mapreduce.Job:  map 83% reduce 8%
```

```
19/06/26 11:08:09 INFO mapreduce.Job:  map 100% reduce 99%
19/06/26 11:08:33 INFO mapreduce.Job:  map 100% reduce 100%
19/06/26 11:08:43 INFO mapreduce.Job: Job job_1561554348150_0007 completed successfully
19/06/26 11:08:44 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=174229098
      FILE: Number of bytes written=349248951
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=511098333
      HDFS: Number of bytes written=146595903
      HDFS: Number of read operations=15
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part_4/part-r-00000 | head
AA, 2010-02-08
AA, 2010-02-05
AA, 2010-02-04
AA, 2010-02-03
AA, 2010-02-02
AA, 2010-02-01
AA, 2010-01-29
AA, 2010-01-28
AA, 2010-01-27
AA, 2010-01-26
```

Part 5

Code in Package- com.hadoop.assignment4.part_5

Output Screenshot:

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/Downloads/A4.jar com.hadoop.assignment4.part_5.AvgStockDriver /nyse_merged /Assignment4/part_5
19/06/26 09:20:31 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 09:20:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
19/06/26 09:20:32 INFO input.FileInputFormat: Total input files to process : 2
19/06/26 09:20:33 INFO mapreduce.JobSubmitter: number of splits:8
19/06/26 09:20:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0003
19/06/26 09:20:34 INFO impl.VarnClientImpl: Submitted application application_1561554348150_0003
19/06/26 09:20:34 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8080/proxy/application_1561554348150_0003/
19/06/26 09:20:34 INFO mapreduce.Job: Running job: job_1561554348150_0003
19/06/26 09:20:42 INFO mapreduce.Job: Job job_1561554348150_0003 running in uber mode : false
19/06/26 09:20:42 INFO mapreduce.Job: map 0% reduce 0%
19/06/26 09:21:34 INFO mapreduce.Job: map 1% reduce 0%
19/06/26 09:21:40 INFO mapreduce.Job: map 6% reduce 0%
19/06/26 09:21:41 INFO mapreduce.Job: map 8% reduce 0%
19/06/26 09:21:46 INFO mapreduce.Job: map 19% reduce 0%
19/06/26 09:21:47 INFO mapreduce.Job: map 23% reduce 0%
19/06/26 09:21:52 INFO mapreduce.Job: map 37% reduce 0%
19/06/26 09:21:54 INFO mapreduce.Job: map 44% reduce 0%
19/06/26 09:21:59 INFO mapreduce.Job: map 47% reduce 0%
19/06/26 09:22:00 INFO mapreduce.Job: map 50% reduce 0%
19/06/26 09:22:01 INFO mapreduce.Job: map 63% reduce 0%
19/06/26 09:22:02 INFO mapreduce.Job: map 67% reduce 0%
19/06/26 09:22:03 INFO mapreduce.Job: map 75% reduce 0%
19/06/26 09:22:22 INFO mapreduce.Job: map 100% reduce 0%
19/06/26 09:22:23 INFO mapreduce.Job: map 100% reduce 100%
19/06/26 09:22:24 INFO mapreduce.Job: Job job_1561554348150_0003 completed successfully
19/06/26 09:22:24 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=23530
    FILE: Number of bytes written=1463344
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1022196666
    HDFS: Number of bytes written=1630
    HDFS: Number of read operations=27
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part_5/part-r-00000 | head
1962 avgStockPrice=0.8514325396825381
1963 avgStockPrice=0.9509047429254746
1964 avgStockPrice=1.1544624505928818
1965 avgStockPrice=1.3094603174603154
1966 avgStockPrice=1.4863437872171557
1967 avgStockPrice=1.7562310756972108
1968 avgStockPrice=1.9254026548672574
1969 avgStockPrice=1.9137759999999973
1970 avgStockPrice=4.814298126527329
1971 avgStockPrice=6.188950524737624
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$
```

Part 6

Code in Package- com.hadoop.assignment4.part_6

Output Screenshot:

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/Downloads/A4.jar com.hadoop.assignment4.part_6.MovieDriver /movieData/ratings.dat /Assignment4/part_6
19/06/26 09:29:44 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 09:29:45 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/06/26 09:29:45 INFO Input.FileInputFormat: Total input files to process : 2
19/06/26 09:29:45 INFO mapreduce.JobSubmitter: number of splits:2
19/06/26 09:29:46 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0005
19/06/26 09:29:46 INFO ImplVarClientImpl: Submitted application application_1561554348150_0005
19/06/26 09:29:46 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8088/proxy/application_1561554348150_0005/
19/06/26 09:29:46 INFO mapreduce.Job: Running job: job_1561554348150_0005
19/06/26 09:29:54 INFO mapreduce.Job: Job job_1561554348150_0005 running in uber mode : false
19/06/26 09:29:54 INFO mapreduce.Job: map 0% reduce 0%
19/06/26 09:30:05 INFO mapreduce.Job: map 100% reduce 0%
19/06/26 09:30:21 INFO mapreduce.Job: map 100% reduce 90%
19/06/26 09:30:27 INFO mapreduce.Job: map 100% reduce 96%
19/06/26 09:30:33 INFO mapreduce.Job: map 100% reduce 100%
19/06/26 09:30:34 INFO mapreduce.Job: Job job_1561554348150_0005 completed successfully
19/06/26 09:30:34 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=45449100
    FILE: Number of bytes written=91373255
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=49188478
    HDFS: Number of bytes written=200439
    HDFS: Number of read operations=0
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part_6/part-r-00000 | head
1      ratingMedian=4.0, ratingStdDev=0.8648081228417158
10     ratingMedian=3.0, ratingStdDev=2.421499942562664
100    ratingMedian=2.0, ratingStdDev=6.460741170375751
1000   ratingMedian=1.0, ratingStdDev=16.565332970099224
1002   ratingMedian=1.0, ratingStdDev=19.07179418233464
1003   ratingMedian=2.0, ratingStdDev=10.950662648814125
1004   ratingMedian=1.0, ratingStdDev=12.065740323421469
1005   ratingMedian=1.0, ratingStdDev=10.236082510365506
1006   ratingMedian=1.0, ratingStdDev=8.732586358927637
1007   ratingMedian=1.0, ratingStdDev=8.159320684890188
cat: Unable to write to output stream
```

Part 7

Code in Package- com.hadoop.assignment4.part_7

Output Screenshot:

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop jar /home/ankit/Downloads/A4.jar com.hadoop.assignment4.part_7.MedianSdDriver /nyse_merged /Assignment4/part_7
19/06/26 09:23:43 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/06/26 09:23:44 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/06/26 09:23:44 INFO InputFileInputFormat: Total input files to process : 2
19/06/26 09:23:44 INFO mapreduce.JobSubmitter: number of splits:0
19/06/26 09:23:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1561554348150_0004
19/06/26 09:23:45 INFO Impl.VarnClientImpl: Submitted application application_1561554348150_0004
19/06/26 09:23:45 INFO mapreduce.Job: The url to track the job: http://ankit-VirtualBox:8088/proxy/application_1561554348150_0004/
19/06/26 09:23:45 INFO mapreduce.Job: Running job: job_1561554348150_0004
19/06/26 09:23:51 INFO mapreduce.Job: Job job_1561554348150_0004 running in uber mode : false
19/06/26 09:23:51 INFO mapreduce.Job: map 0% reduce 0%
19/06/26 09:24:20 INFO mapreduce.Job: map 3% reduce 0%
19/06/26 09:24:27 INFO mapreduce.Job: map 10% reduce 0%
19/06/26 09:24:33 INFO mapreduce.Job: map 16% reduce 0%
19/06/26 09:24:39 INFO mapreduce.Job: map 22% reduce 0%
19/06/26 09:24:46 INFO mapreduce.Job: map 24% reduce 0%
19/06/26 09:24:53 INFO mapreduce.Job: map 25% reduce 0%
```

```
19/06/26 09:26:41 INFO mapreduce.Job: map 100% reduce 87%
19/06/26 09:26:42 INFO mapreduce.Job: map 100% reduce 100%
19/06/26 09:26:44 INFO mapreduce.Job: Job job_1561554348150_0004 completed successfully
19/06/26 09:26:44 INFO mapreduce.Job: Counters: 50
    File System Counters
      FILE: Number of bytes read=68353370
      FILE: Number of bytes written=95210245
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=1022196666
      HDFS: Number of bytes written=1923
      HDFS: Number of read operations=27
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
    Job Counters
```

```
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$ hadoop fs -cat /Assignment4/part_7/part-r-00000 | head
1962 median=0.57, sd=0.7639648431806878
1963 median=0.18, sd=1.1427569423191115
1964 median=0.09, sd=1.568545774728101
1965 median=0.09, sd=2.0704131319305588
1966 median=0.09, sd=2.706412785594762
1967 median=0.09, sd=3.4963260915767305
1968 median=0.09, sd=4.484096675549327
1969 median=0.09, sd=4.841988599127338
1970 median=0.17, sd=20.827274867480597
1971 median=0.1, sd=33.629667379518146
ankit@ankit-VirtualBox: /usr/local/bin/hadoop-2.8.5/bin$
```