

We inserted into HDFS the same datasets used in week 02 homework and compile and run the MaxTemperature and MaxTemperatureWithCombiner files from the text book sample code provided at <http://hadoopbook.com>

DataSet-1 (1990) With MaxTemperature:

Number of bytes read= 51062743

Job counters:

Total time spent by all the maps in the slots (ms): 361317 and for reduces (ms): 27914

```
17/02/06 21:05:55 INFO mapreduce.Job: Job job_1486260480953_0005 completed successfully
17/02/06 21:05:55 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=51062743
    FILE: Number of bytes written=102996088
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1030903607
    HDFS: Number of bytes written=9
    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
    Launched reduce tasks=1
    Data-local map tasks=9
    Total time spent by all maps in occupied slots (ms)=361317
    Total time spent by all reduces in occupied slots (ms)=27913
    Total time spent by all map tasks (ms)=361317
    Total time spent by all reduce tasks (ms)=27913
    Total vcore-seconds taken by all map tasks=361317
    Total vcore-seconds taken by all reduce tasks=27913
    Total megabyte-seconds taken by all map tasks=369988608
    Total megabyte-seconds taken by all reduce tasks=28582912
  Map-Reduce Framework
```

According to the job history the time spent to complete the job of Dataset-1 is 83 secs



JobHistory

Application
 About
 Jobs
 Tools

Retired Jobs

Show 20 entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2017.02.06 21:04:23 UTC	2017.02.06 21:04:30 UTC	2017.02.06 21:05:53 UTC	job_1486260480953_0005	Max temperature	vagrant	default	SUCCEEDED	8	8

Dataset-2 (1990 & 1992) With MaxTemperature:


Number of bytes read= 336520439

Job counters:

Total time spent by all the maps in the slots (ms): 2166772 and for reduces (ms): 325268

```
17/02/06 19:09:57 INFO mapreduce.Job: Job job_1486260480953_0001 completed successfully
17/02/06 19:09:57 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=336520439
    FILE: Number of bytes written=678941434
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7993016763
    HDFS: Number of bytes written=18
    HDFS: Number of read operations=183
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=61
    Launched reduce tasks=1
    Data-local map tasks=61
    Total time spent by all maps in occupied slots (ms)=2166772
    Total time spent by all reduces in occupied slots (ms)=325268
    Total time spent by all map tasks (ms)=2166772
    Total time spent by all reduce tasks (ms)=325268
    Total vcore-seconds taken by all map tasks=2166772
    Total vcore-seconds taken by all reduce tasks=325268
    Total megabyte-seconds taken by all map tasks=2218774528
    Total megabyte-seconds taken by all reduce tasks=333074432
```

According to the job history the time spent to complete the job of Dataset-2 is 7 mins 31 secs

		JobHistory							
<div>Application</div> <div>About Jobs</div> <div>Tools</div>		Retired Jobs							
Show 20 entries									
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2017-02-06 19:02:18 UTC	2017-02-06 19:02:25 UTC	2017-02-06 19:09:56 UTC	job_1486260480953_0001	Max temperature	vagrant	default	SUCCEEDED	60	60

Dataset-3 (1990-1992) With MaxTemperature :

Number of bytes read= 608943531

Job counters:


Total time spent by all the maps in the slots (ms): 3965178 and for reduces (ms): 666145

```

17/02/06 19:35:54 INFO mapreduce.Job: Job job_1486260480953_0002 completed successfully
17/02/06 19:35:54 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=608943531
    FILE: Number of bytes written=1229011257
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=15257843875
    HDFS: Number of bytes written=36
    HDFS: Number of read operations=345
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=115
    Launched reduce tasks=1
    Data-local map tasks=115
    Total time spent by all maps in occupied slots (ms)=3965178
    Total time spent by all reduces in occupied slots (ms)=666145
    Total time spent by all map tasks (ms)=3965178
    Total time spent by all reduce tasks (ms)=666145
    Total vcore-seconds taken by all map tasks=3965178
    Total vcore-seconds taken by all reduce tasks=666145
    Total megabyte-seconds taken by all map tasks=4060342272
    Total megabyte-seconds taken by all reduce tasks=682132480

```

According to the job history the time spent to complete the job of Dataset-3 is 13 mins 49 secs



JobHistory

▼ Application

About

Jobs

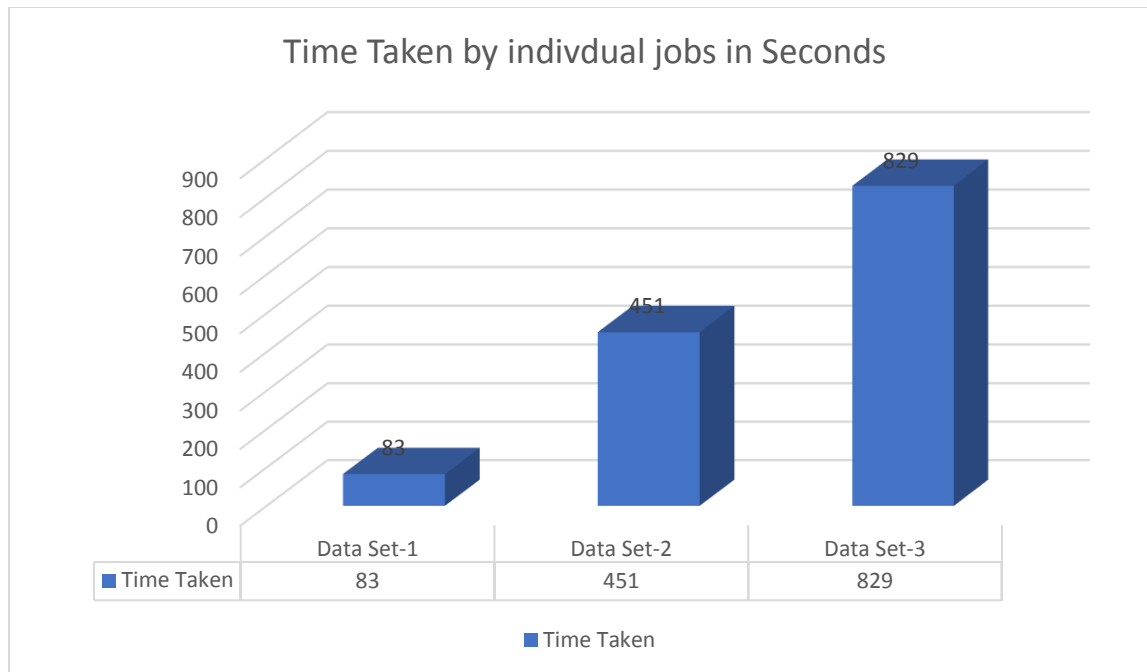
► Tools

Retired Jobs

Show 20 ▾ entries

Submit Time ↕	Start Time ↕	Finish Time ▾	Job ID ↕	Name ↕	User ↕	Queue ↕	State ↕	Maps Total ↕	Maps Completed ↕
2017.02.06 19:21:58 UTC	2017.02.06 19:22:03 UTC	2017.02.06 19:35:52 UTC	job_1486260480953_0002	Max temperature	vagrant	default	SUCCEEDED	114	114

Data Analysis with MaxTemperature.java file:

**Conclusion:**

In dataset-1 which has 1990 year data, the time taken to complete the job is 83 Seconds.

While to complete the job for 1990 & 1992-year data, it took 451 seconds which is 5.5 times the time taken to complete the dataset 1.

To complete the job for whole data of 4 years, it took 829 seconds. Which is approximately 10 times then dataset-1 and twict the time taken by dataset-2

DataSet-1 (1990) With MaxTemperatureCombiner:

Number of bytes read= 94

Job counters:

Total time spent by all the maps in the slots (ms): 344528 and for reduces (ms): 23473


```

17/02/06 21:11:12 INFO mapreduce.Job: Job job_1486260480953_0006 completed successfully
17/02/06 21:11:12 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=94
    FILE: Number of bytes written=872248
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1030903607
    HDFS: Number of bytes written=9
    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
    Launched reduce tasks=1
    Data-local map tasks=9
    Total time spent by all maps in occupied slots (ms)=344528
    Total time spent by all reduces in occupied slots (ms)=23473
    Total time spent by all map tasks (ms)=344528
    Total time spent by all reduce tasks (ms)=23473
    Total vcore-seconds taken by all map tasks=344528
    Total vcore-seconds taken by all reduce tasks=23473
    Total megabyte-seconds taken by all map tasks=352796672
    Total megabyte-seconds taken by all reduce tasks=24036352

```

According to the job history the time spent to complete the job of Dataset-1 with combiner is 1 mins 16 secs



JobHistory

Application
 About
 Jobs
 Tools

Retired Jobs

Show 20 entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2017.02.06 21:09:48 UTC	2017.02.06 21:09:54 UTC	2017.02.06 21:11:10 UTC	job_1486260480953_0006	Max temperature	vagrant	default	SUCCEEDED	8	8

DataSet-2 (1990&1992) With MaxTemperatureCombiner:

Number of bytes read= 677

Job counters:


Total time spent by all the maps in the slots (ms): 2540411 and for reduces (ms): 410745

```

17/02/06 20:55:16 INFO mapreduce.Job: Job job_1486260480953_0004 completed successfully
17/02/06 20:55:16 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=677
    FILE: Number of bytes written=5911627
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7993016763
    HDFS: Number of bytes written=18
    HDFS: Number of read operations=183
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=2
    Launched map tasks=62
    Launched reduce tasks=1
    Data-local map tasks=62
    Total time spent by all maps in occupied slots (ms)=2540411
    Total time spent by all reduces in occupied slots (ms)=410745
    Total time spent by all map tasks (ms)=2540411
    Total time spent by all reduce tasks (ms)=410745
    Total vcore-seconds taken by all map tasks=2540411
    Total vcore-seconds taken by all reduce tasks=410745
    Total megabyte-seconds taken by all map tasks=2601380864
    Total megabyte-seconds taken by all reduce tasks=420602880

```

According to the job history the time spent to complete the job of Dataset-2 with combiner is 8 mins 35 secs



JobHistory

▼ Application

[About](#)

[Jobs](#)

► Tools

Retired Jobs

Show 20 ▼ entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2017.02.06 20:46:33 UTC	2017.02.06 20:46:39 UTC	2017.02.06 20:55:14 UTC	job_1486260480953_0004	Max temperature	vagrant	default	SUCCEEDED	60	60

DataSet-3 (1990-1993) With MaxTemperatureCombiner:


Number of bytes read= 1293

Job counters:

Total time spent by all the maps in the slots (ms): 4628002 and for reduces (ms): 710298

```
17/02/06 20:41:23 INFO mapreduce.Job: Job job_1486260480953_0003 completed successfully
17/02/06 20:41:23 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=1293
    FILE: Number of bytes written=11144987
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=15257843875
    HDFS: Number of bytes written=36
    HDFS: Number of read operations=345
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=115
    Launched reduce tasks=1
    Data-local map tasks=115
    Total time spent by all maps in occupied slots (ms)=4628002
    Total time spent by all reduces in occupied slots (ms)=710698
    Total time spent by all map tasks (ms)=4628002
    Total time spent by all reduce tasks (ms)=710698
    Total vcore-seconds taken by all map tasks=4628002
    Total vcore-seconds taken by all reduce tasks=710698
    Total megabyte-seconds taken by all map tasks=4739074048
    Total megabyte-seconds taken by all reduce tasks=727754752
```

According to the job history the time spent to complete the job of Dataset-2 with combiner is 15 mins 18 secs



JobHistory

▼ Application

[About](#)

[Jobs](#)

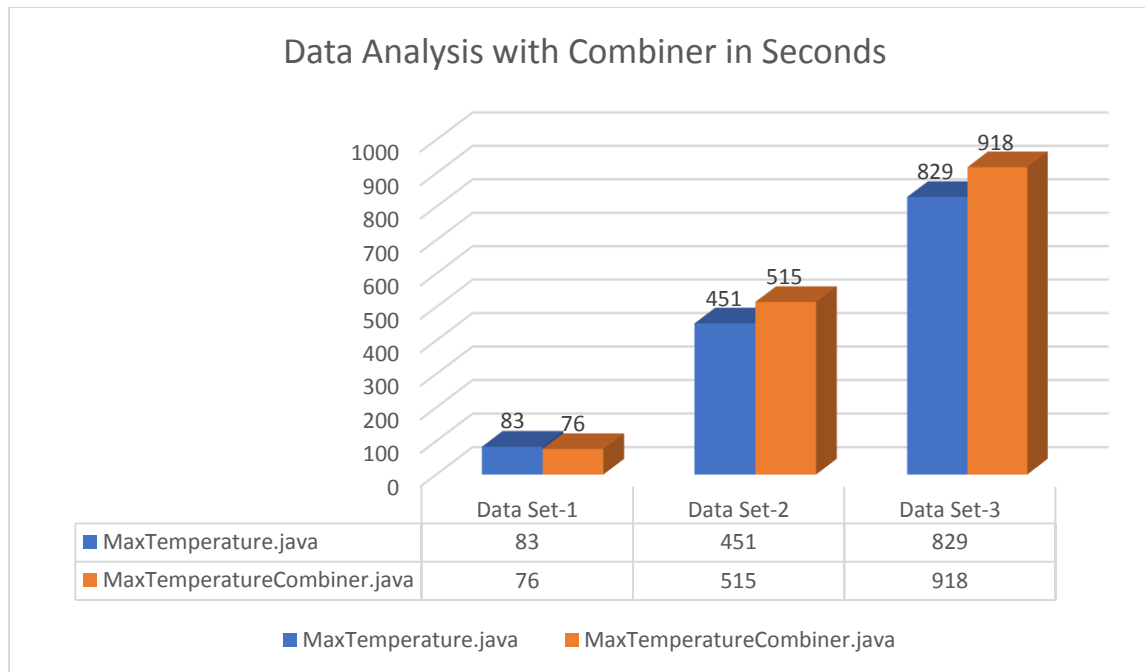
► Tools

Retired Jobs

Show 20 ▼ entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed
2017.02.06 20:25:56 UTC	2017.02.06 20:26:03 UTC	2017.02.06 20:41:21 UTC	job_1486260480953_0003	Max temperature	vagrant	default	SUCCEEDED	114	114

Data Analysis with MaxTemperature.java and MaxTemperatureCombiner.java file:

**Conclusion for using Combiner:**

As we can analysis from the given chart that the time taken to complete the job using combiner class is greater than time taken by simply using map reducer class except in data set 1.