

## Deliverables:

- Complete a single node cluster install as detailed in Appendix A of the Hadoop text book.
- Insert into HDFS the 1990, 1991, 1992 and 1993 temperature data sets.
- Compile and run the MaxTemperature and MaxTemperatureWithCombiner files from the text book sample code provided at <http://hadoopbook.com>
- Accurately get the execution times by enabling port forwarding in Vagrant of the specified ports listed at the end of Appendix A
- Note the execution time of the jobs and graph the differences.

**Solution:**

The screenshot below indicates the successful installation of Single Node Cluster as detailed in Appendix A.

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ jps
2625 ResourceManager
2782 NodeManager
2250 DataNode
2483 SecondaryNameNode
3093 JobHistoryServer
5292 Jps
2092 NameNode
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

## Small Set

## Insertion of 1990 data set (used in week 02 homework) into HDFS

[illegible]

Progress and Successful execution of MaxTemperature on 1990 data set is depicted below:

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperature /user/$USER/
tempdata/small /user/$USER/output_small
17/02/10 07:04:23 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/10 07:04:23 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface
and execute your application with ToolRunner to remedy this.
17/02/10 07:04:24 INFO input.FileInputFormat: Total input paths to process : 1
17/02/10 07:04:24 INFO mapreduce.JobSubmitter: number of splits:8
17/02/10 07:04:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486709182360_0001
17/02/10 07:04:25 INFO impl.YarnClientImpl: Submitted application application_1486709182360_0001
17/02/10 07:04:25 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1
486709182360_0001/
17/02/10 07:04:25 INFO mapreduce.Job: Running job: job_1486709182360_0001
17/02/10 07:04:38 INFO mapreduce.Job: Job job_1486709182360_0001 running in uber mode : false
17/02/10 07:04:38 INFO mapreduce.Job: map 0% reduce 0%
17/02/10 07:05:31 INFO mapreduce.Job: map 4% reduce 0%
17/02/10 07:05:33 INFO mapreduce.Job: map 5% reduce 0%
17/02/10 07:05:35 INFO mapreduce.Job: map 6% reduce 0%
17/02/10 07:05:37 INFO mapreduce.Job: map 13% reduce 0%
17/02/10 07:05:39 INFO mapreduce.Job: map 15% reduce 0%
17/02/10 07:05:40 INFO mapreduce.Job: map 26% reduce 0%
17/02/10 07:05:42 INFO mapreduce.Job: map 28% reduce 0%
17/02/10 07:05:43 INFO mapreduce.Job: map 35% reduce 0%
17/02/10 07:05:44 INFO mapreduce.Job: map 36% reduce 0%
17/02/10 07:05:45 INFO mapreduce.Job: map 38% reduce 0%
17/02/10 07:05:46 INFO mapreduce.Job: map 43% reduce 0%
17/02/10 07:05:47 INFO mapreduce.Job: map 46% reduce 0%
17/02/10 07:05:48 INFO mapreduce.Job: map 47% reduce 0%
17/02/10 07:05:50 INFO mapreduce.Job: map 50% reduce 0%
17/02/10 07:05:51 INFO mapreduce.Job: map 58% reduce 0%
17/02/10 07:05:52 INFO mapreduce.Job: map 75% reduce 0%
17/02/10 07:06:17 INFO mapreduce.Job: map 77% reduce 0%
17/02/10 07:06:18 INFO mapreduce.Job: map 77% reduce 25%
17/02/10 07:06:20 INFO mapreduce.Job: map 80% reduce 25%
17/02/10 07:06:21 INFO mapreduce.Job: map 86% reduce 25%
17/02/10 07:06:24 INFO mapreduce.Job: map 92% reduce 25%
17/02/10 07:06:25 INFO mapreduce.Job: map 96% reduce 25%
17/02/10 07:06:26 INFO mapreduce.Job: map 100% reduce 25%
17/02/10 07:06:28 INFO mapreduce.Job: map 100% reduce 43%
17/02/10 07:06:31 INFO mapreduce.Job: map 100% reduce 73%
17/02/10 07:06:34 INFO mapreduce.Job: map 100% reduce 100%
17/02/10 07:06:34 INFO mapreduce.Job: Job job_1486709182360_0001 completed successfully
```

Manikandan Ganesh  
A20355226

```
17/02/10 07:06:34 INFO mapreduce.Job: Job job_1486709182360_0001 completed successfully
```

```
17/02/10 07:06:34 INFO mapreduce.Job: Counters: 50
```

#### File System Counters

```
FILE: Number of bytes read=51062743
FILE: Number of bytes written=102995998
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1030903639
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=2
Launched map tasks=10
Launched reduce tasks=1
Data-local map tasks=10
Total time spent by all maps in occupied slots (ms)=521605
Total time spent by all reduces in occupied slots (ms)=38557
Total time spent by all map tasks (ms)=521605
Total time spent by all reduce tasks (ms)=38557
Total vcore-seconds taken by all map tasks=521605
Total vcore-seconds taken by all reduce tasks=38557
Total megabyte-seconds taken by all map tasks=534123520
Total megabyte-seconds taken by all reduce tasks=39482368
```

#### Map-Reduce Framework

```
Map input records=5000000
Map output records=4642067
Map output bytes=41778603
Map output materialized bytes=51062785
Input split bytes=912
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=51062785
Reduce input records=4642067
Reduce output records=1
Spilled Records=9284134
Shuffled Maps =8
Failed Shuffles=0
Merged Map outputs=8
GC time elapsed (ms)=5868
CPU time spent (ms)=38170
Physical memory (bytes) snapshot=1957867520
Virtual memory (bytes) snapshot=7099863040
Total committed heap usage (bytes)=1387724800
```

#### Shuffle Errors

```
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
```

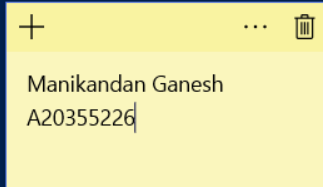
#### File Input Format Counters

```
Bytes Read=1030902727
```

#### File Output Format Counters

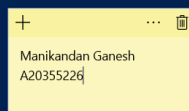
```
Bytes Written=9
```

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```



Likewise, the successful completion of MaxTemperature with Combiner on 1990 dataset is given below:

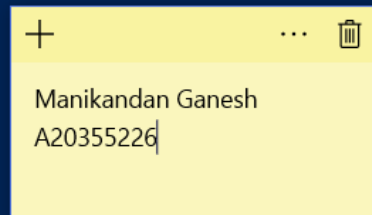
```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperatureWithCombiner /user/$USER/tempdata/small /user/$USER/output_combiner_small
17/02/10 07:11:38 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/10 07:11:39 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/10 07:11:39 INFO input.FileInputFormat: Total input paths to process : 1
17/02/10 07:11:39 INFO mapreduce.JobSubmitter: number of splits:8
17/02/10 07:11:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486709182360_0002
17/02/10 07:11:40 INFO impl.YarnClientImpl: Submitted application application_1486709182360_0002
17/02/10 07:11:40 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486709182360_0002/
17/02/10 07:11:40 INFO mapreduce.Job: Running job: job_1486709182360_0002
17/02/10 07:11:48 INFO mapreduce.Job: Job job_1486709182360_0002 running in uber mode : false
17/02/10 07:12:24 INFO mapreduce.Job: map 0% reduce 0%
17/02/10 07:12:27 INFO mapreduce.Job: map 2% reduce 0%
17/02/10 07:12:28 INFO mapreduce.Job: map 3% reduce 0%
17/02/10 07:12:31 INFO mapreduce.Job: map 16% reduce 0%
17/02/10 07:12:34 INFO mapreduce.Job: map 31% reduce 0%
17/02/10 07:12:37 INFO mapreduce.Job: map 48% reduce 0%
17/02/10 07:12:40 INFO mapreduce.Job: map 54% reduce 0%
17/02/10 07:12:41 INFO mapreduce.Job: map 71% reduce 0%
17/02/10 07:12:42 INFO mapreduce.Job: map 75% reduce 0%
17/02/10 07:13:00 INFO mapreduce.Job: map 77% reduce 0%
17/02/10 07:13:02 INFO mapreduce.Job: map 80% reduce 0%
17/02/10 07:13:03 INFO mapreduce.Job: map 86% reduce 23%
17/02/10 07:13:05 INFO mapreduce.Job: map 96% reduce 25%
17/02/10 07:13:06 INFO mapreduce.Job: map 100% reduce 29%
17/02/10 07:13:07 INFO mapreduce.Job: map 100% reduce 100%
```



```

17/02/10 07:13:08 INFO mapreduce.Job: Job job_1486709182360_0002 completed successfully
17/02/10 07:13:09 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=94
    FILE: Number of bytes written=872167
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1030903639
    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)=366687
    Total time spent by all reduces in occupied slots (ms)=24513
    Total time spent by all map tasks (ms)=366687
    Total time spent by all reduce tasks (ms)=24513
    Total vcore-seconds taken by all map tasks=366687
    Total vcore-seconds taken by all reduce tasks=24513
    Total megabyte-seconds taken by all map tasks=375487488
    Total megabyte-seconds taken by all reduce tasks=25101312
  Map-Reduce Framework
    Map input records=5000000
    Map output records=4642067
    Map output bytes=41778603
    Map output materialized bytes=136
    Input split bytes=912
    Combine input records=4642067
    Combine output records=8
    Reduce input groups=1
    Reduce shuffle bytes=136
    Reduce input records=8
    Reduce output records=1
    Spilled Records=16
    Shuffled Maps =8
    Failed Shuffles=0
    Merged Map outputs=8
    GC time elapsed (ms)=4407
    CPU time spent (ms)=28070
    Physical memory (bytes) snapshot=1864716288
    Virtual memory (bytes) snapshot=7099981824
    Total committed heap usage (bytes)=1337786368
  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=1030902727
  File Output Format Counters
    Bytes Written=9
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$

```

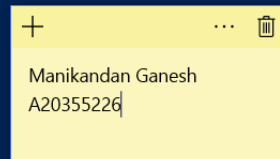


The screenshot shows the output obtained from 1990 set:

```

vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output_small/part-r-00000
1990      607
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$

```



## Insertion of medium data set (used in week 02 homework) into HDFS

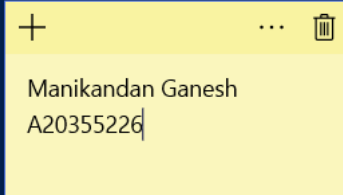
Manikandan Ganesh  
A20355226

Manikandan Ganesh  
A20355226

```

17/02/10 07:56:50 INFO mapreduce.Job: map 100% reduce 81%
17/02/10 07:56:53 INFO mapreduce.Job: map 100% reduce 87%
17/02/10 07:56:56 INFO mapreduce.Job: map 100% reduce 92%
17/02/10 07:56:59 INFO mapreduce.Job: map 100% reduce 98%
17/02/10 07:57:00 INFO mapreduce.Job: map 100% reduce 100%
17/02/10 07:57:02 INFO mapreduce.Job: Job job_1486709182360_0003 completed successfully
17/02/10 07:57:02 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=336520439
    FILE: Number of bytes written=678941739
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7993013087
    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)=2951758
    Total time spent by all reduces in occupied slots (ms)=493974
    Total time spent by all map tasks (ms)=2951758
    Total time spent by all reduce tasks (ms)=493974
    Total vcore-seconds taken by all map tasks=2951758
    Total vcore-seconds taken by all reduce tasks=493974
    Total megabyte-seconds taken by all map tasks=3022600192
    Total megabyte-seconds taken by all reduce tasks=505829376
  Map-Reduce Framework
    Map input records=35000000
    Map output records=30592765
    Map output bytes=275334885
    Map output materialized bytes=336520775
    Input split bytes=6900
    Combine input records=0
    Combine output records=0
    Reduce input groups=2
    Reduce shuffle bytes=336520775
    Reduce input records=30592765
    Reduce output records=2
    Spilled Records=61185530
    Shuffled Maps =60
    Failed Shuffles=0
    Merged Map outputs=60
    GC time elapsed (ms)=35193
    CPU time spent (ms)=247320
    Physical memory (bytes) snapshot=13463089152
    Virtual memory (bytes) snapshot=48086548480
    Total committed heap usage (bytes)=9765781504
  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=7993006187
  File Output Format Counters
    Bytes Written=18
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$

```



Execution of Max Temperature with Combiner on Medium Set – Progress and Completion



```

vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt-jar-MaxTemperatureWithCombiner -user/$USER/tempdata/medium/* -user/$USER/output_combiner_medium
17/02/10 08:06:22 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/10 08:06:22 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/10 08:06:23 INFO input.FileInputFormat: Total input paths to process : 2
17/02/10 08:06:23 INFO mapreduce.JobSubmitter: number of splits:60
17/02/10 08:06:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486709182360_0004
17/02/10 08:06:24 INFO impl.YarnClientImpl: Submitted application application_1486709182360_0004
17/02/10 08:06:24 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486709182360_0004/
17/02/10 08:06:33 INFO mapreduce.Job: Job job_1486709182360_0004 running in uber mode : false
17/02/10 08:06:33 INFO mapreduce.Job: map 0% reduce 0%
17/02/10 08:07:15 INFO mapreduce.Job: map 1% reduce 0%
17/02/10 08:07:16 INFO mapreduce.Job: map 2% reduce 0%
17/02/10 08:07:18 INFO mapreduce.Job: map 3% reduce 0%
17/02/10 08:07:19 INFO mapreduce.Job: map 4% reduce 0%
17/02/10 08:07:22 INFO mapreduce.Job: map 5% reduce 0%
17/02/10 08:07:27 INFO mapreduce.Job: map 6% reduce 0%
17/02/10 08:07:28 INFO mapreduce.Job: map 7% reduce 0%
17/02/10 08:07:38 INFO mapreduce.Job: map 8% reduce 0%
17/02/10 08:07:39 INFO mapreduce.Job: map 10% reduce 0%
17/02/10 08:08:17 INFO mapreduce.Job: map 11% reduce 0%
17/02/10 08:08:20 INFO mapreduce.Job: map 12% reduce 0%
17/02/10 08:08:21 INFO mapreduce.Job: map 13% reduce 0%
17/02/10 08:08:23 INFO mapreduce.Job: map 14% reduce 0%
17/02/10 08:08:25 INFO mapreduce.Job: map 15% reduce 0%
17/02/10 08:08:26 INFO mapreduce.Job: map 16% reduce 0%
17/02/10 08:08:27 INFO mapreduce.Job: map 17% reduce 0%
17/02/10 08:08:30 INFO mapreduce.Job: map 18% reduce 0%
17/02/10 08:08:32 INFO mapreduce.Job: map 20% reduce 0%
17/02/10 08:09:07 INFO mapreduce.Job: map 20% reduce 7%
17/02/10 08:09:08 INFO mapreduce.Job: map 21% reduce 7%

```

Manikandan Ganesh  
A20355226

```

17/02/10 08:15:43 INFO mapreduce.Job: map 95% reduce 32%
17/02/10 08:16:03 INFO mapreduce.Job: map 96% reduce 32%
17/02/10 08:16:06 INFO mapreduce.Job: map 97% reduce 32%
17/02/10 08:16:07 INFO mapreduce.Job: map 98% reduce 32%
17/02/10 08:16:09 INFO mapreduce.Job: map 100% reduce 32%
17/02/10 08:16:10 INFO mapreduce.Job: map 100% reduce 100%
17/02/10 08:16:10 INFO mapreduce.Job: Job job_1486709182360_0004 completed successfully
17/02/10 08:16:10 INFO mapreduce.Job: Counters: 50

```

#### File System Counters

```

FILE: Number of bytes read=666
FILE: Number of bytes written=5912154
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=7993013087
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)=2866063
Total time spent by all reduces in occupied slots (ms)=455543
Total time spent by all map tasks (ms)=2866063
Total time spent by all reduce tasks (ms)=455543
Total vcore-seconds taken by all map tasks=2866063
Total vcore-seconds taken by all reduce tasks=455543
Total megabyte-seconds taken by all map tasks=2934848512
Total megabyte-seconds taken by all reduce tasks=466476032

```

#### Map-Reduce Framework

```

Map input records=35000000
Map output records=30592765
Map output bytes=275334885
Map output materialized bytes=1020
Input split bytes=6900
Combine input records=30592765
Combine output records=60
Reduce input groups=2
Reduce shuffle bytes=1020
Reduce input records=60
Reduce output records=2
Spilled Records=120
Shuffled Maps =60
Failed Shuffles=0
Merged Map outputs=60
GC time elapsed (ms)=33011
CPU time spent (ms)=214350
Physical memory (bytes) snapshot=13328904192
Virtual memory (bytes) snapshot=48087384064
Total committed heap usage (bytes)=9625731072

```

#### 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=7993006187
```

#### File Output Format Counters

```
Bytes Written=18
```

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

Manikandan Ganesh  
A20355226

```

bytes in 1768=18
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output_medium/part-r-00000
1990      607
1992      605
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$

```

## LARGE SET

## Insertion of Large Set (1990 to 1993) on HDFS

[illegible]

## Progress and Completion of Max Temperature on Large Data Set

```

17/02/10 09:04:45 INFO Client.RMPProxy: Connecting to ResourceManager at localhost/main/java/hadoop.jar.mt.jar MaxTemperature /user/SUSKRY/temdata/target/ /user/SUSKRY/outputLarge
17/02/10 09:04:46 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/10 09:04:46 INFO input.FileInputFormat: Total input paths to process : 4
17/02/10 09:04:46 INFO mapreduce.JobSubmitter: number of splits:115
17/02/10 09:04:47 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486709182360_0005
17/02/10 09:04:47 INFO impl.YarnClientImpl: Submitting application application_1486709182360_0005
17/02/10 09:04:47 INFO mapreduce.Job: The url to track the job: http://vaagrant-ubuntu-trusty-64:8088/proxy/application_1486709182360_0005
17/02/10 09:04:47 INFO mapreduce.Job: Running job: job_1486709182360_0005
17/02/10 09:04:47 INFO mapreduce.Job: Job job_1486709182360_0005 running in uber mode : false
17/02/10 09:04:56 INFO mapreduce.Job: map 0% reduce 0%
17/02/10 09:05:38 INFO mapreduce.Job: map 1% reduce 0%
17/02/10 09:05:41 INFO mapreduce.Job: map 2% reduce 0%
17/02/10 09:05:49 INFO mapreduce.Job: map 3% reduce 0%
17/02/10 09:05:53 INFO mapreduce.Job: map 4% reduce 0%
17/02/10 09:05:58 INFO mapreduce.Job: map 5% reduce 0%
17/02/10 09:06:33 INFO mapreduce.Job: map 6% reduce 0%
17/02/10 09:06:39 INFO mapreduce.Job: map 7% reduce 0%
17/02/10 09:06:45 INFO mapreduce.Job: map 8% reduce 0%
17/02/10 09:06:49 INFO mapreduce.Job: map 9% reduce 0%
17/02/10 09:06:51 INFO mapreduce.Job: map 10% reduce 0%
17/02/10 09:07:27 INFO mapreduce.Job: map 11% reduce 0%
17/02/10 09:07:44 INFO mapreduce.Job: map 12% reduce 0%
17/02/10 09:07:47 INFO mapreduce.Job: map 13% reduce 0%
17/02/10 09:07:51 INFO mapreduce.Job: map 14% reduce 0%
17/02/10 09:07:54 INFO mapreduce.Job: map 15% reduce 0%
17/02/10 09:07:55 INFO mapreduce.Job: map 16% reduce 0%
17/02/10 09:08:36 INFO mapreduce.Job: map 17% reduce 0%
17/02/10 09:08:39 INFO mapreduce.Job: map 18% reduce 0%
17/02/10 09:08:42 INFO mapreduce.Job: map 19% reduce 0%
17/02/10 09:08:46 INFO mapreduce.Job: map 20% reduce 0%
17/02/10 09:09:23 INFO mapreduce.Job: map 21% reduce 0%
17/02/10 09:10:07 INFO mapreduce.Job: map 21% reduce 7%
17/02/10 09:10:14 INFO mapreduce.Job: map 22% reduce 7%
17/02/10 09:10:17 INFO mapreduce.Job: map 23% reduce 7%
17/02/10 09:10:23 INFO mapreduce.Job: map 24% reduce 7%
17/02/10 09:10:26 INFO mapreduce.Job: map 25% reduce 7%
17/02/10 09:10:27 INFO mapreduce.Job: map 25% reduce 8%
17/02/10 09:11:01 INFO mapreduce.Job: map 26% reduce 8%
17/02/10 09:11:07 INFO mapreduce.Job: map 27% reduce 8%
17/02/10 09:11:11 INFO mapreduce.Job: map 28% reduce 8%
17/02/10 09:11:16 INFO mapreduce.Job: map 30% reduce 9%
17/02/10 09:11:19 INFO mapreduce.Job: map 30% reduce 10%
17/02/10 09:12:02 INFO mapreduce.Job: map 31% reduce 10%
17/02/10 09:12:28 INFO mapreduce.Job: map 32% reduce 10%
17/02/10 09:12:33 INFO mapreduce.Job: map 33% reduce 10%
17/02/10 09:12:35 INFO mapreduce.Job: map 34% reduce 10%
17/02/10 09:12:38 INFO mapreduce.Job: map 34% reduce 11%
17/02/10 09:13:16 INFO mapreduce.Job: map 35% reduce 11%
17/02/10 09:13:20 INFO mapreduce.Job: map 36% reduce 11%
17/02/10 09:13:27 INFO mapreduce.Job: map 37% reduce 11%
17/02/10 09:13:30 INFO mapreduce.Job: map 37% reduce 12%
17/02/10 09:13:31 INFO mapreduce.Job: map 38% reduce 12%
17/02/10 09:13:33 INFO mapreduce.Job: map 38% reduce 13%
17/02/10 09:14:06 INFO mapreduce.Job: map 39% reduce 13%
17/02/10 09:14:12 INFO mapreduce.Job: map 40% reduce 13%
17/02/10 09:14:15 INFO mapreduce.Job: map 41% reduce 13%
17/02/10 09:14:18 INFO mapreduce.Job: map 42% reduce 13%
17/02/10 09:14:20 INFO mapreduce.Job: map 43% reduce 13%
17/02/10 09:14:21 INFO mapreduce.Job: map 43% reduce 14%
17/02/10 09:15:01 INFO mapreduce.Job: map 44% reduce 14%
17/02/10 09:15:04 INFO mapreduce.Job: map 45% reduce 14%
17/02/10 09:15:10 INFO mapreduce.Job: map 46% reduce 14%
17/02/10 09:15:13 INFO mapreduce.Job: map 47% reduce 15%
17/02/10 09:15:16 INFO mapreduce.Job: map 47% reduce 16%
17/02/10 09:16:00 INFO mapreduce.Job: map 48% reduce 16%

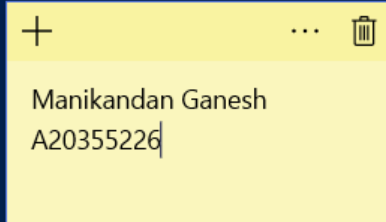
```



```

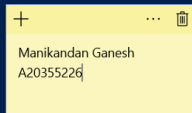
17/02/10 09:26:54 INFO mapreduce.Job: map 100% reduce 100%
17/02/10 09:26:54 INFO mapreduce.Job: Job job_1486709182360_0005 completed successfully
17/02/10 09:26:54 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=608943531
    FILE: Number of bytes written=1229108339
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=15257836481
    HDFS: Number of bytes written=36
    HDFS: Number of read operations=348
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=116
    Launched reduce tasks=1
    Data-local map tasks=116
    Total time spent by all maps in occupied slots (ms)=6412846
    Total time spent by all reduces in occupied slots (ms)=1081710
    Total time spent by all map tasks (ms)=6412846
    Total time spent by all reduce tasks (ms)=1081710
    Total vcore-seconds taken by all map tasks=6412846
    Total vcore-seconds taken by all reduce tasks=1081710
    Total megabyte-seconds taken by all map tasks=6566754304
    Total megabyte-seconds taken by all reduce tasks=1107671040
  Map-Reduce Framework
    Map input records=65000000
    Map output records=55358499
    Map output bytes=498226491
    Map output materialized bytes=608944179
    Input split bytes=13110
    Combine input records=0
    Combine output records=0
    Reduce input groups=4
    Reduce shuffle bytes=608944179
    Reduce input records=55358499
    Reduce output records=4
    Spilled Records=110716998
    Shuffled Maps =115
    Failed Shuffles=0
    Merged Map outputs=115
    GC time elapsed (ms)=61697
    CPU time spent (ms)=461670
    Physical memory (bytes) snapshot=25629028352
    Virtual memory (bytes) snapshot=91437907968
    Total committed heap usage (bytes)=18517958656
  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=15257823371
  File Output Format Counters
    Bytes Written=36
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$

```

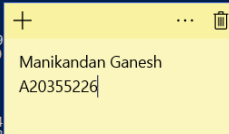


Progress and Completion of Max Temperature with Combiner on Large Data Set

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt-jar MaxTemperatureWithCombiner /user/$USER/tempdata/large/* /user/$USER/output_combiner_large
17/02/10 09:32:49 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/10 09:32:50 WARN mapreduce.jobsubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/10 09:32:50 INFO input.FileInputFormat: Total input paths to process : 4
17/02/10 09:32:51 INFO mapreduce.jobsubmitter: number of splits:115
17/02/10 09:32:51 INFO mapreduce.jobsubmitter: Submitting tokens for job: job_1486709182360_0006
17/02/10 09:32:51 INFO impl.VarClientImpl: Submitted application application_1486709182360_0006
17/02/10 09:32:51 INFO mapreduce.job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486709182360_0006/
17/02/10 09:33:01 INFO mapreduce.job: Job job_1486709182360_0006 running in uber mode : false
17/02/10 09:33:01 INFO mapreduce.job: map 0% reduce 0%
17/02/10 09:33:44 INFO mapreduce.job: map 1% reduce 0%
17/02/10 09:33:48 INFO mapreduce.job: map 2% reduce 0%
17/02/10 09:33:54 INFO mapreduce.job: map 3% reduce 0%
17/02/10 09:33:57 INFO mapreduce.job: map 4% reduce 0%
17/02/10 09:33:59 INFO mapreduce.job: map 5% reduce 0%
17/02/10 09:34:39 INFO mapreduce.job: map 6% reduce 0%
17/02/10 09:34:43 INFO mapreduce.job: map 7% reduce 0%
17/02/10 09:35:45 INFO mapreduce.job: map 8% reduce 0%
17/02/10 09:36:29 INFO mapreduce.job: map 9% reduce 0%
17/02/10 09:36:41 INFO mapreduce.job: map 10% reduce 0%
17/02/10 09:37:31 INFO mapreduce.job: map 11% reduce 0%
17/02/10 09:37:38 INFO mapreduce.job: map 12% reduce 0%
17/02/10 09:37:41 INFO mapreduce.job: map 13% reduce 0%
17/02/10 09:37:44 INFO mapreduce.job: map 14% reduce 0%
17/02/10 09:37:48 INFO mapreduce.job: map 15% reduce 0%
17/02/10 09:37:49 INFO mapreduce.job: map 16% reduce 0%
17/02/10 09:38:30 INFO mapreduce.job: map 17% reduce 0%
17/02/10 09:38:33 INFO mapreduce.job: map 18% reduce 0%
17/02/10 09:38:36 INFO mapreduce.job: map 19% reduce 0%
17/02/10 09:38:41 INFO mapreduce.job: map 20% reduce 0%
17/02/10 09:38:42 INFO mapreduce.job: map 21% reduce 0%
17/02/10 09:39:18 INFO mapreduce.job: map 21% reduce 7%
17/02/10 09:39:22 INFO mapreduce.job: map 22% reduce 7%
17/02/10 09:39:28 INFO mapreduce.job: map 23% reduce 7%
17/02/10 09:39:42 INFO mapreduce.job: map 24% reduce 7%
17/02/10 09:39:44 INFO mapreduce.job: map 25% reduce 8%
17/02/10 09:39:55 INFO mapreduce.job: map 26% reduce 8%
17/02/10 09:40:13 INFO mapreduce.job: map 26% reduce 9%
17/02/10 09:40:18 INFO mapreduce.job: map 27% reduce 9%
17/02/10 09:40:25 INFO mapreduce.job: map 28% reduce 9%
17/02/10 09:40:28 INFO mapreduce.job: map 29% reduce 9%
17/02/10 09:40:30 INFO mapreduce.job: map 30% reduce 9%
17/02/10 09:40:32 INFO mapreduce.job: map 30% reduce 10%
17/02/10 09:41:01 INFO mapreduce.job: map 31% reduce 10%
17/02/10 09:41:08 INFO mapreduce.job: map 32% reduce 10%
17/02/10 09:41:12 INFO mapreduce.job: map 33% reduce 10%
17/02/10 09:41:16 INFO mapreduce.job: map 34% reduce 10%
17/02/10 09:41:18 INFO mapreduce.job: map 34% reduce 11%
17/02/10 09:41:35 INFO mapreduce.job: map 35% reduce 11%
17/02/10 09:41:36 INFO mapreduce.job: map 35% reduce 12%
17/02/10 09:41:52 INFO mapreduce.job: map 36% reduce 12%
17/02/10 09:41:56 INFO mapreduce.job: map 37% reduce 12%
17/02/10 09:42:00 INFO mapreduce.job: map 38% reduce 12%
17/02/10 09:42:04 INFO mapreduce.job: map 38% reduce 13%
17/02/10 09:42:14 INFO mapreduce.job: map 39% reduce 13%
17/02/10 09:42:34 INFO mapreduce.job: map 40% reduce 13%
17/02/10 09:42:39 INFO mapreduce.job: map 41% reduce 13%
17/02/10 09:42:42 INFO mapreduce.job: map 42% reduce 13%
17/02/10 09:42:46 INFO mapreduce.job: map 43% reduce 13%
17/02/10 09:42:47 INFO mapreduce.job: map 43% reduce 14%
17/02/10 09:43:13 INFO mapreduce.job: map 44% reduce 14%
17/02/10 09:43:23 INFO mapreduce.job: map 45% reduce 14%
17/02/10 09:43:26 INFO mapreduce.job: map 46% reduce 14%
17/02/10 09:43:27 INFO mapreduce.job: map 46% reduce 15%
```




```
17/02/10 09:52:32 INFO mapreduce.job: map 100% reduce 99%
17/02/10 09:52:33 INFO mapreduce.job: map 100% reduce 100%
17/02/10 09:52:34 INFO mapreduce.job: Job job_1486709182360_0006 completed successfully
17/02/10 09:52:34 INFO mapreduce.job: Counters: 50
File System Counters
  FILE: Number of bytes read=1271
  FILE: Number of bytes written=11242763
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=15257836481
  HDFS: Number of bytes written=36
  HDFS: Number of read operations=348
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Killed map tasks=2
  Launched map tasks=117
  Launched reduce tasks=1
  Data-local map tasks=117
  Total time spent by all maps in occupied slots (ms)=5973810
  Total time spent by all reduces in occupied slots (ms)=829481
  Total time spent by all map tasks (ms)=5973810
  Total time spent by all reduce tasks (ms)=829481
  Total vcore-seconds taken by all map tasks=5973810
  Total vcore-seconds taken by all reduce tasks=829481
  Total megabyte-seconds taken by all map tasks=61171814
  Total megabyte-seconds taken by all reduce tasks=849381
Map-Reduce Framework
  Map input records=65000000
  Map output records=55358499
  Map output bytes=498226491
  Map output materialized bytes=1955
  Input split bytes=13110
  Combine input records=55358499
  Combine output records=115
  Reduce input groups=4
  Reduce shuffle bytes=1955
  Reduce input records=115
  Reduce output records=4
  Spilled Records=220
  Shuffled Maps =115
  Failed Shuffles=0
  Merged Map outputs=115
  GC time elapsed (ms)=62022
  CPU time spent (ms)=399430
  Physical memory (bytes) snapshot=25465253888
  Virtual memory (bytes) snapshot=91440070656
  Total committed heap usage (bytes)=18391826432
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=15257823371
File Output Format Counters
  Bytes Written=36
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```



## Output Result

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output_large/part-r-00000
1990    607
1991    607
1992    605
1993    567
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

## Job Execution history



### JobHistory

Logged in as: dr.who

Application

About Jobs

Tools

MAX TEMPERATURE WITH COMBINER → LARGE

MAX TEMPERATURE →

MAX TEMPERATURE WITH COMBINER → MEDIUM

MAX TEMPERATURE →

MAX TEMPERATURE WITH COMBINER → SMALL

MAX TEMPERATURE →

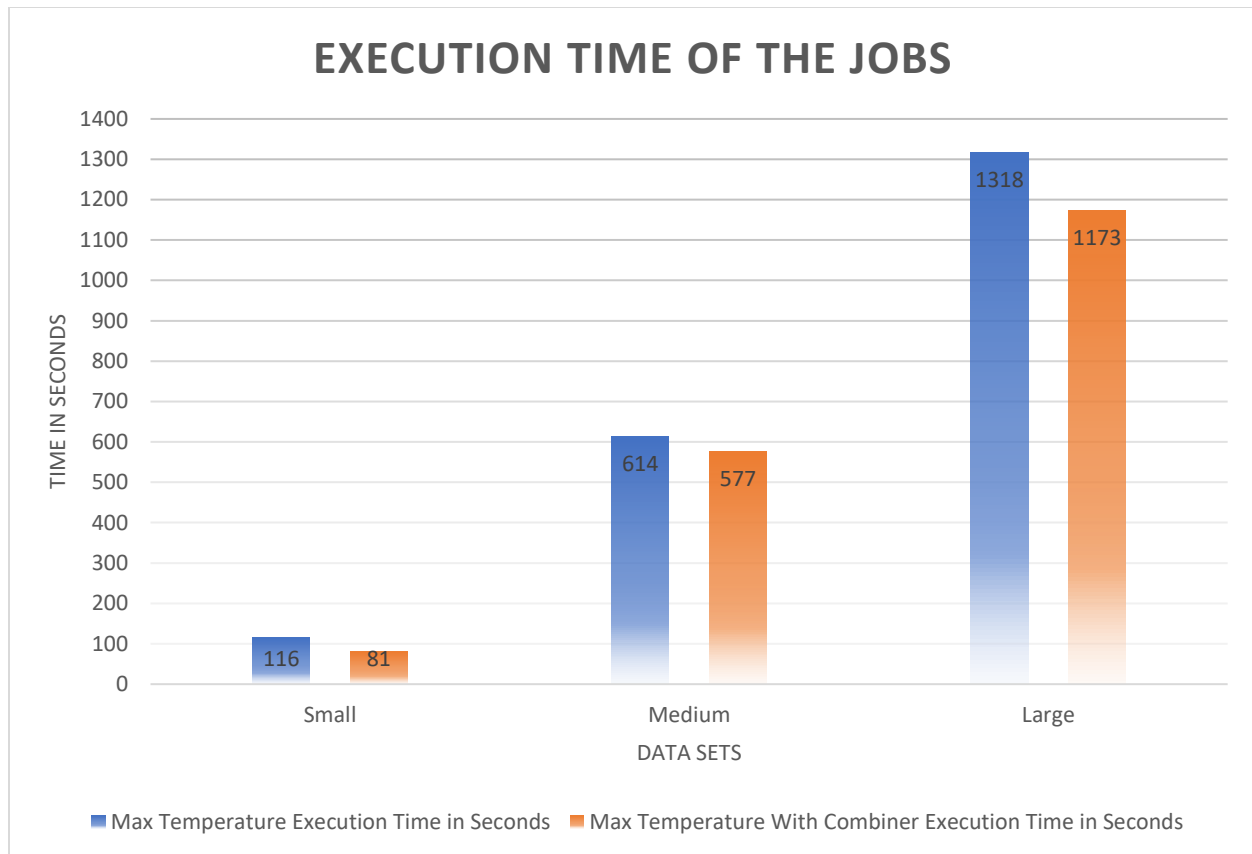
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2017.02.10 09:32:51 UTC	2017.02.10 09:32:59 UTC	2017.02.10 09:52:32 UTC	job_1486709182360_0006	Max temperature	vagrant	default	SUCCEEDED	115	115	1	1
2017.02.10 09:04:47 UTC	2017.02.10 09:04:54 UTC	2017.02.10 09:26:52 UTC	job_1486709182360_0005	Max temperature	vagrant	default	SUCCEEDED	115	115	1	1
2017.02.10 08:06:24 UTC	2017.02.10 08:06:31 UTC	2017.02.10 08:16:08 UTC	job_1486709182360_0004	Max temperature	vagrant	default	SUCCEEDED	60	60	1	1
2017.02.10 07:46:37 UTC	2017.02.10 07:46:45 UTC	2017.02.10 07:56:59 UTC	job_1486709182360_0003	Max temperature	vagrant	default	SUCCEEDED	60	60	1	1
2017.02.10 07:11:39 UTC	2017.02.10 07:11:46 UTC	2017.02.10 07:13:07 UTC	job_1486709182360_0002	Max temperature	vagrant	default	SUCCEEDED	8	8	1	1
2017.02.10 07:04:25 UTC	2017.02.10 07:04:36 UTC	2017.02.10 07:06:32 UTC	job_1486709182360_0001	Max temperature	vagrant	default	SUCCEEDED	8	8	1	1

Showing 1 to 6 of 6 entries

## ANALYSIS ON THE EXECUTION TIME OF THE JOBS

Data Set	Max Temperature Start Time	Max Temperature End Time	Max Temperature with Combiner Start Time	Max Temperature with Combiner End Time
Small	7:04:36 UTC	7:06:32 UTC	7:11:46 UTC	7:13:07 UTC
Medium	7:46:45 UTC	7:56:59 UTC	8:06:31 UTC	8:16:08 UTC
Large	9:04:54 UTC	9:26:52 UTC	9:32:59 UTC	9:52:32 UTC

Data Set	Max Temperature Execution Time (In mm:ss)	Max Temperature with Combiner Execution Time (In mm:ss)	Max Temperature Execution Time in Seconds	Max Temperature with Combiner Execution Time in Seconds
Small	01:56	01:21	116	81
Medium	10:14	09:37	614	577
Large	21:58	19:33	1318	1173



- Based on the plotted graph, the first and the foremost thing that can be observed is the quicker execution of the Max Temperature with Combiner on all the three data sets.
- The principle purpose behind the speedier execution of combiner class is that it diminishes the utilization of network bandwidth (the size of data transferred over the system per unit time) while transferring the data from mappers to reducers.
- In our case, the goal is to find the Maximum Temperature for the years 1990, 1991, 1992 and 1993 having different sizes of data sets. Having that into account, the combiner takes a shot at all the three data sets in the following way:
  - For the small data set, we have 8 splits of 128 MB data. On utilizing combiner, it acts like a local reducer to find the maximum temperature on each split, doing the sorting and rearranging process specifically from Mapper and sending the outcome to the reducer.

- We have observed 60 splits for the medium data set. Once more, the combiner chips away at each split of data to locate the maximum temperature on each split first and after that sending the outcome from the 60 splits to the reducer for pulling back the last maximum temperature.
  - Likewise, the combiner follows up on large data set having 115 splits to pull back definite maximum temperature.
- Thus, the utilization of combiner expands the throughput and enhances the execution by having better throughput and latency.
- In the instance of Max Temperature without combiner, the utilization of network bandwidth is high. Hence, the measure of time to execute is likewise high.
- The purpose behind more utilization of time would be the handling of the considerable number of splits on reducer itself in this way influencing the throughput. In this manner, all the 8 splits in the small data set will be handled on the reducer itself to identify the maximum temperature. Time taken for each split to cross from Mapper to Reducer will be high and thus the execution time will likewise be high.