

Shrija Chavan

A20381511

ITMD 521

Week 03

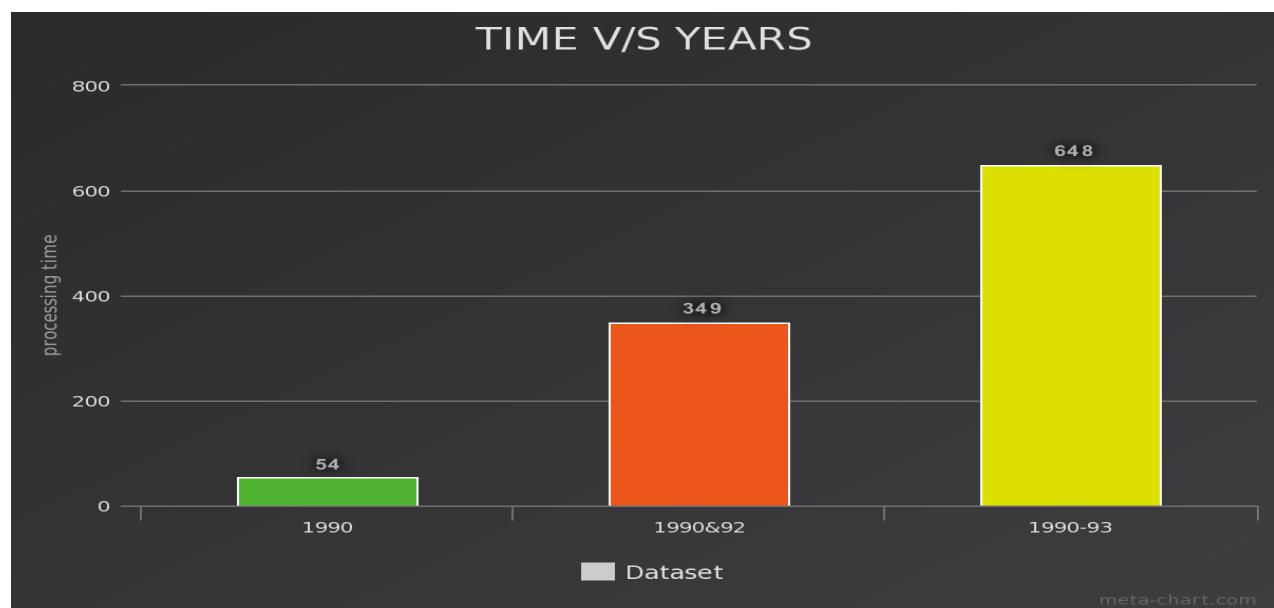
Below is a chart explaining of the amount of time required (in seconds) to run each dataset using the awk script, java, and MapReduce (with MaxTemperature and MaxTempetarureWithCombiner) with a graph combined with some screen shots to explain the process in detail and compare the time taken by each dataset.

Chart:

Datasets	awk	java	MapReduce	
			MaxTemperature	MaxTemperatureWithCombiner
1990	13.62	57	54	49
1990&1992	99.28	382	349	322
1990-1993	192.40	708	648	610

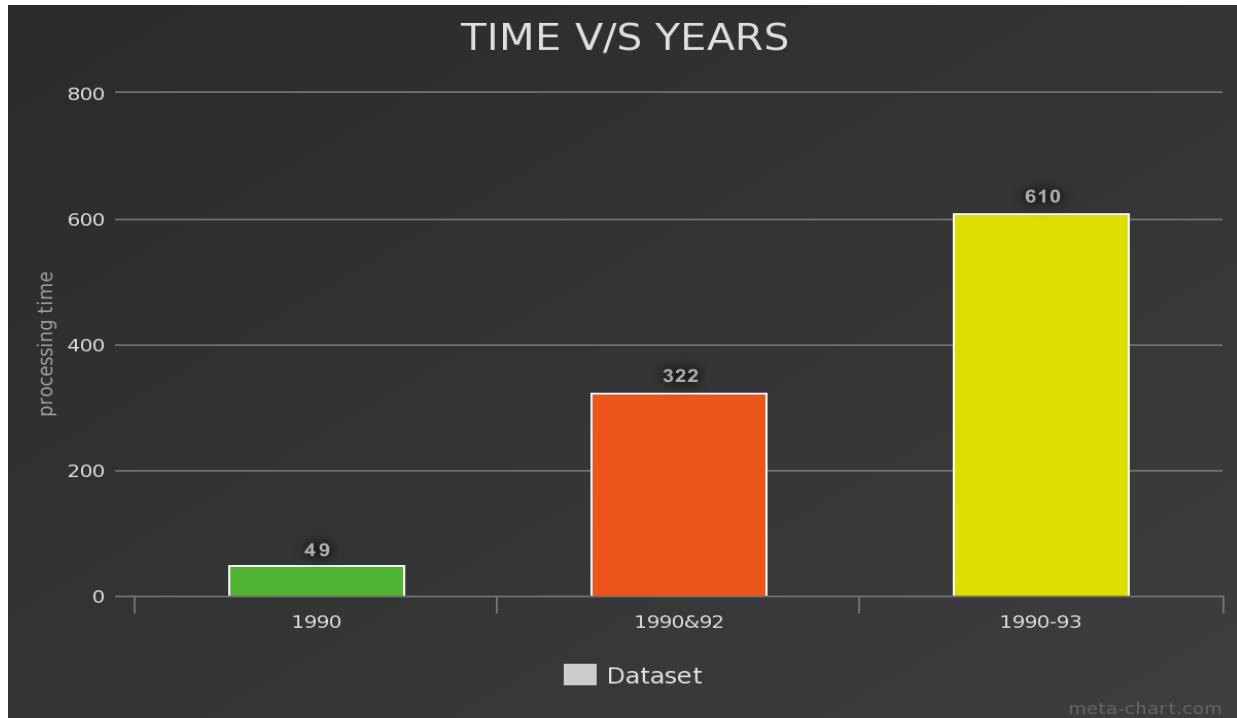
Graph:

MaxTemperature:



The above graph shows the time taken by each dataset to execute using the MaxTemperature class.

Graph: MaxTemperatureWithCombiner



The above graph shows the time taken by each dataset to execute using the MaxTemperatureWithCombiner class.

Analysis:

As you can see in the above chart, the least time taken to execute the job was by the awk script on the datasets. But, as we observe, after awk script the next minimum time taken was by the MapReduce as compared to Java.

Advantages of using MapReduce over SQL:

MapReduce is one of the efficient programming model to process a large set of data. It works on a key-value pair and can be written in any language. It can be tested on a local cluster or a distributed cluster. The job of the map is to collect the raw data as input and process it in key-value pairs and pass it as an input to the reducer. Generally, the input data will be in the HDFS file system. The reducer will then combine the key-value pair to produce the final results. As the name indicates the reduce job is always performed after the map job.

The Mapper breaks down the large dataset into several small files and distributes it to various data nodes. The data nodes will then process the data and gives the highest temperature per file back to the mapper. The mapper passes the same information to reducer and the reducer will then give you the highest temperature each year. The real preferred standpoint of MapReduce is that it is easy to scale data processing over various computing node as compared to the java which runs on a signal file at a time to give us the result.

MapReduce is also known as the Batch query processor. It can run adhoc query against the whole dataset and get the results in a reasonable time.

But, there are some scenarios where MR will not be a best choice to choose:

1. When querying small dataset, it will always be preferred to run the job on a single cluster than on a distributed cluster.
2. When you need fast response, MR is not the best choice as its not know to gets the results faster than the SQL.
3. Implementing MR to every job can be costlier, only big companies can afford the resources.
4. Works best on large datasets as compared to small datasets.

Observation:

1. Time taken by Java(SQL) and MapReduce did not differ much since the MapReduce was working on a single cluster.
2. If we would create more cluster to run the MapReduce job the results would be much faster than now.
3. The awk script was the fastest because it was executed on the compressed files as compared to Java and MR which were run on the uncompressed dataset.
4. The execution time taken by MaxTemperatureWithCombiner was noticeably less than MaxTemperature because when we use the combiner class, the data sent as input to the reducer is first collected by the combiner to sort the data and then send it to the reducer so that time taken by the reducer is less than the time taken without using the combiner class.

Screen Shots:

Using MaxTemperature 1990 Dataset

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh * vagrant ssh
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperature /user/$USER/tempdata/1990 /user/$USER/output
17/02/05 00:48:42 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/05 00:48:43 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/05 00:48:43 INFO input.FileInputFormat: Total input paths to process : 1
17/02/05 00:48:43 INFO mapreduce.JobSubmitter: number of splits:8
17/02/05 00:48:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486255348474_0001
17/02/05 00:48:44 INFO impl.YarnClientImpl: Submitted application application_1486255348474_0001
17/02/05 00:48:44 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486255348474_0001/
17/02/05 00:48:44 INFO mapreduce.Job: Running job: job_1486255348474_0001
17/02/05 00:48:51 INFO mapreduce.Job: Job job_1486255348474_0001 running in uber mode : false
17/02/05 00:48:51 INFO mapreduce.Job: map 0% reduce 0%
17/02/05 00:49:17 INFO mapreduce.Job: map 6% reduce 0%
17/02/05 00:49:20 INFO mapreduce.Job: map 18% reduce 0%
17/02/05 00:49:23 INFO mapreduce.Job: map 34% reduce 0%
17/02/05 00:49:26 INFO mapreduce.Job: map 51% reduce 0%
17/02/05 00:49:28 INFO mapreduce.Job: map 68% reduce 0%
17/02/05 00:49:29 INFO mapreduce.Job: map 75% reduce 0%
17/02/05 00:49:30 INFO mapreduce.Job: map 83% reduce 0%
17/02/05 00:49:41 INFO mapreduce.Job: map 88% reduce 0%
17/02/05 00:49:42 INFO mapreduce.Job: map 100% reduce 0%
17/02/05 00:49:44 INFO mapreduce.Job: map 100% reduce 67%
17/02/05 00:49:46 INFO mapreduce.Job: map 100% reduce 100%
17/02/05 00:49:46 INFO mapreduce.Job: Job job_1486255348474_0001 completed successfully
17/02/05 00:49:46 INFO mapreduce.Job: Counters
FILE: Number of bytes read=51062743
FILE: Number of bytes written=102995935
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=903591
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)=229745
Total time spent by all reduces in occupied slots (ms)=16000
Total time spent by all map tasks (ms)=229745

File System Counters
```

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh * vagrant ssh
Launched reduce tasks=1
Data-local map tasks=9
Total time spent by all maps in occupied slots (ms)=229745
Total time spent by all reduces in occupied slots (ms)=16000
Total time spent by all map tasks (ms)=229745
Total time spent by all reduce tasks (ms)=16000
Total core-seconds taken by all map tasks=229745
Total vcore-seconds taken by all reduce tasks=16000
Total megabyte-seconds taken by all map tasks=235258880
Total megabyte-seconds taken by all reduce tasks=16384000
Map-Reduce Framework
Map input records=5000000
Map output records=4642067
Map output bytes=41778603
Map output materialized bytes=51062785
Input split bytes=864
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=51062785
Reduce input records=4642067
Reduce output records=0
Spilled records=9284134
Shuffled Maps=8
Failed Shuffles=0
Merged Map outputs=8
GC time elapsed (ms)=3955
CPU time spent (ms)=21870
Physical memory (bytes) snapshot=1801879552
Virtual memory (bytes) snapshot=16870817792
Total committed heap usage (bytes)=1324724224
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$ hadoop fs -cat /user/$USER/output/part-r-00000
1990 607
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

MapReduce Job job_1486255348474_0001

Attempt Number	Start Time	Node	Logs
1	Sun Feb 05 00:48:47 UTC 2017	vagrant-ubuntu-trusty-64:8042	logs

1990 and 1992 Dataset

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java--ssh•vagrant ssh
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperature /user/$USER/tempdata/1990_1992 /user/$USER/output
17/02/05 00:55:30 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/05 00:55:30 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/05 00:55:31 INFO input.FileInputFormat: Total input paths to process : 1
17/02/05 00:55:31 INFO mapreduce.JobSubmitter: number of splits:60
17/02/05 00:55:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486255348474_0002
17/02/05 00:55:31 INFO impl.YarnClientImpl: Submitted application application_1486255348474_0002
17/02/05 00:55:31 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486255348474_0002/
17/02/05 00:55:31 INFO mapreduce.Job: Running job: job_1486255348474_0002
17/02/05 00:55:38 INFO mapreduce.Job: Job job_1486255348474_0002 running in uber mode : false
17/02/05 00:55:38 INFO mapreduce.Job: map 0% reduce 0%
17/02/05 00:56:02 INFO mapreduce.Job: map 1% reduce 0%
17/02/05 00:56:05 INFO mapreduce.Job: map 2% reduce 0%
17/02/05 00:56:09 INFO mapreduce.Job: map 4% reduce 0%
17/02/05 00:56:10 INFO mapreduce.Job: map 5% reduce 0%
17/02/05 00:56:12 INFO mapreduce.Job: map 6% reduce 0%
17/02/05 00:56:13 INFO mapreduce.Job: map 7% reduce 0%
17/02/05 00:56:14 INFO mapreduce.Job: map 10% reduce 0%
17/02/05 00:56:38 INFO mapreduce.Job: map 11% reduce 0%
17/02/05 00:56:39 INFO mapreduce.Job: map 12% reduce 0%
17/02/05 00:56:41 INFO mapreduce.Job: map 13% reduce 0%
17/02/05 00:56:42 INFO mapreduce.Job: map 14% reduce 0%
17/02/05 00:56:44 INFO mapreduce.Job: map 15% reduce 0%
17/02/05 00:56:45 INFO mapreduce.Job: map 16% reduce 0%
17/02/05 00:56:46 INFO mapreduce.Job: map 17% reduce 0%
17/02/05 00:56:47 INFO mapreduce.Job: map 19% reduce 0%
17/02/05 00:56:48 INFO mapreduce.Job: map 20% reduce 0%
17/02/05 00:57:10 INFO mapreduce.Job: map 21% reduce 0%
17/02/05 00:57:13 INFO mapreduce.Job: map 22% reduce 0%
17/02/05 00:57:14 INFO mapreduce.Job: map 23% reduce 0%
17/02/05 00:57:16 INFO mapreduce.Job: map 25% reduce 0%
17/02/05 00:57:19 INFO mapreduce.Job: map 26% reduce 0%
17/02/05 00:57:20 INFO mapreduce.Job: map 28% reduce 0%
17/02/05 00:57:21 INFO mapreduce.Job: map 30% reduce 0%
17/02/05 00:57:42 INFO mapreduce.Job: map 30% reduce 10%
17/02/05 00:57:43 INFO mapreduce.Job: map 31% reduce 10%
17/02/05 00:57:46 INFO mapreduce.Job: map 32% reduce 10%
17/02/05 00:57:49 INFO mapreduce.Job: map 34% reduce 10%
17/02/05 00:57:52 INFO mapreduce.Job: map 35% reduce 10%
17/02/05 00:57:53 INFO mapreduce.Job: map 38% reduce 10%
17/02/05 00:57:55 INFO mapreduce.Job: map 38% reduce 13%
17/02/05 00:58:13 INFO mapreduce.Job: map 39% reduce 13%
17/02/05 00:58:18 INFO mapreduce.Job: map 41% reduce 13%
```

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java$ ssh vagrant ssh
vagrant@vagrant-ubuntu-trusty-64: ~$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-client.jar org.apache.hadoop.mapred.lib.Collector
Data-local map tasks=60
Total time spent by all maps in occupied slots (ms)=1687213
Total time spent by all reduces in occupied slots (ms)=245547
Total time spent by all map tasks (ms)=1687213
Total time spent by all reduce tasks (ms)=245547
Total vcore-seconds taken by all map tasks=1687213
Total vcore-seconds taken by all reduce tasks=245547
Total megabyte-seconds taken by all map tasks=1727706112
Total megabyte-seconds taken by all reduce tasks=251440128
Map-Reduce Framework
  Map input records=35000000
  Map output records=30592765
  Map output bytes=275334895
  Map output materialized bytes=336520775
  Input split bytes=6780
  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)=28485
  CPU time spent (ms)=155850
  Physical memory (bytes) snapshot=12472471552
  Virtual memory (bytes) snapshot=114304675840
  Total committed heap usage (bytes)=9409646592
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=7993010283
File Output Format Counters
  Bytes Written=18
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
1990 607
1992 605
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ ]
```

The screenshot shows a Chrome browser window with the following details:

- Title Bar:** Chrome, File, Edit, View, History, Bookmarks, People, Window, Help.
- Address Bar:** localhost:19888/jobhistory/job/job_1486255348474_0002
- Page Content:**
 - Job Overview:** Shows the job name (Max temperature), user (vagrant), queue (default), state (SUCCEEDED), and submission details (Sun Feb 05 00:55:31 UTC 2017).
 - Diagnostics:** Includes average map time (28sec), average shuffle time (3mins, 53sec), average merge time (1sec), and average reduce time (10sec).
 - ApplicationMaster:** A table showing the ApplicationMaster attempt number (1), start time (Sun Feb 05 00:55:33 UTC 2017), node (vagrant-ubuntu-trusty-64:8042), and logs (logs).
 - Task Summary:** Two tables showing task types and attempt types. The first table for Task Type shows Map (60 total, 1 failed, 1 killed, 60 successful) and Reduce (0 total, 0 failed, 0 killed, 1 successful). The second table for Attempt Type shows Maps (0 total, 0 failed, 0 killed, 60 successful) and Reduces (0 total, 0 failed, 0 killed, 1 successful).

1990 to 1993 Dataset

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh • vagrant ssh
\|vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperature /user/$USER/tempdata/1990_1991_1992_1993 /user/$USER/output
17/02/05 01:12:50 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/05 01:12:50 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/05 01:12:50 INFO input.FileInputFormat: Total input paths to process : 1
17/02/05 01:12:51 INFO mapreduce.JobSubmitter: number of splits:114
17/02/05 01:12:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486255348474_0003
17/02/05 01:12:51 INFO impl.YarnClientImpl: Submitted application application_1486255348474_0003
17/02/05 01:12:51 INFO mapreduce.Job: Running job: job_1486255348474_0003
17/02/05 01:12:58 INFO mapreduce.Job: Job job_1486255348474_0003 running in uber mode : false
17/02/05 01:12:58 INFO mapreduce.Job: map 0% reduce 0%
17/02/05 01:13:25 INFO mapreduce.Job: map 1% reduce 0%
17/02/05 01:13:29 INFO mapreduce.Job: map 3% reduce 0%
17/02/05 01:13:32 INFO mapreduce.Job: map 4% reduce 0%
17/02/05 01:13:33 INFO mapreduce.Job: map 5% reduce 0%
17/02/05 01:13:53 INFO mapreduce.Job: map 6% reduce 0%
17/02/05 01:14:00 INFO mapreduce.Job: map 7% reduce 0%
17/02/05 01:14:01 INFO mapreduce.Job: map 8% reduce 0%
17/02/05 01:14:04 INFO mapreduce.Job: map 9% reduce 0%
17/02/05 01:14:06 INFO mapreduce.Job: map 11% reduce 0%
17/02/05 01:14:29 INFO mapreduce.Job: map 12% reduce 0%
17/02/05 01:14:31 INFO mapreduce.Job: map 13% reduce 0%
17/02/05 01:14:34 INFO mapreduce.Job: map 14% reduce 0%
17/02/05 01:14:35 INFO mapreduce.Job: map 15% reduce 0%
17/02/05 01:14:37 INFO mapreduce.Job: map 16% reduce 0%
17/02/05 01:15:00 INFO mapreduce.Job: map 17% reduce 0%
17/02/05 01:15:01 INFO mapreduce.Job: map 18% reduce 0%
17/02/05 01:15:04 INFO mapreduce.Job: map 19% reduce 0%
17/02/05 01:15:07 INFO mapreduce.Job: map 20% reduce 0%
17/02/05 01:15:08 INFO mapreduce.Job: map 21% reduce 0%
17/02/05 01:15:28 INFO mapreduce.Job: map 22% reduce 7%
17/02/05 01:15:31 INFO mapreduce.Job: map 23% reduce 7%
17/02/05 01:15:35 INFO mapreduce.Job: map 24% reduce 7%
17/02/05 01:15:38 INFO mapreduce.Job: map 25% reduce 8%
17/02/05 01:15:41 INFO mapreduce.Job: map 26% reduce 8%
17/02/05 01:15:54 INFO mapreduce.Job: map 27% reduce 9%
17/02/05 01:16:03 INFO mapreduce.Job: map 28% reduce 9%
17/02/05 01:16:06 INFO mapreduce.Job: map 29% reduce 9%
17/02/05 01:16:09 INFO mapreduce.Job: map 30% reduce 9%
17/02/05 01:16:11 INFO mapreduce.Job: map 30% reduce 10%
17/02/05 01:16:12 INFO mapreduce.Job: map 31% reduce 10%
17/02/05 01:16:26 INFO mapreduce.Job: map 31% reduce 10%
17/02/05 01:16:32 INFO mapreduce.Job: map 32% reduce 10%
```

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh • vagrant ssh
\|vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
Total time spent by all reduces in occupied slots (ms)=519011
Total time spent by all map tasks=3085349
Total time spent by all reduce tasks (ms)=519011
Total vcore-seconds taken by all map tasks=3085349
Total vcore-seconds taken by all reduce tasks=519011
Total megabyte-seconds taken by all map tasks=3159397376
Total megabyte-seconds taken by all reduce tasks=531467264
Map-Reduce Framework
  Map input records=65000000
  Map output records=55358499
  Map output bytes=498226491
  Map output materialized bytes=608944173
  Input split bytes=14022
  Combine input records=0
  Combine output records=0
  Reduce input groups=4
  Reduce shuffle bytes=608944173
  Reduce input records=55358499
  Reduce output records=4
  Spilled Records=110716998
  Shuffled Maps =114
  Failed Shuffles=0
  Merged Map outputs=114
  GC time elapsed (ms)=51181
  CPU time spent (ms)=287170
  Physical memory (bytes) snapshot=23587983360
  Virtual memory (bytes) snapshot=215471804416
  Total committed heap usage (bytes)=17690132480
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=15257831563
File Output Format Counters
  Bytes Written=36
\|vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
1990  607
1991  607
1992  605
1993  567
\|vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

Using MaxTemperatureWithCombiner:

1990 Dataset

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java$ ssh -v vagrant ssh
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperatureWithCombiner /user/$USER/tempdata/1990 /user/$USER/output
17/02/05 01:25:08 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/05 01:25:08 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with tool.run() if you need to override this.
17/02/05 01:25:09 INFO mapreduce.JobInputFormat: Total input paths to process : 1
17/02/05 01:25:09 INFO mapreduce.JobSubmitter: number of splits:8
17/02/05 01:25:40 INFO mapreduce.Job: map 4% reduce 0%
17/02/05 01:25:42 INFO mapreduce.Job: map 6% reduce 0%
17/02/05 01:25:43 INFO mapreduce.Job: map 8% reduce 0%
17/02/05 01:25:43 INFO mapreduce.Job: map 19% reduce 0%
17/02/05 01:25:45 INFO mapreduce.Job: map 22% reduce 0%
17/02/05 01:25:47 INFO mapreduce.Job: map 38% reduce 0%
17/02/05 01:25:50 INFO mapreduce.Job: map 58% reduce 0%
17/02/05 01:25:51 INFO mapreduce.Job: map 71% reduce 0%
17/02/05 01:25:51 INFO mapreduce.Job: map 75% reduce 0%
17/02/05 01:25:53 INFO mapreduce.Job: map 88% reduce 0%
17/02/05 01:26:04 INFO mapreduce.Job: map 100% reduce 0%
17/02/05 01:26:05 INFO mapreduce.Job: map 100% reduce 100%
17/02/05 01:26:06 INFO mapreduce.Job: Job job_1486255348474_0004 completed successfully
17/02/05 01:26:06 INFO mapreduce.Job: Counters: 50
  FILE: Number of bytes read=94
  FILE: Number of bytes written=872023
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=10309003591
  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
  Local map tasks=9
  Total time spent by all maps in occupied slots (ms)=223016
  Total time spent by all reduces in occupied slots (ms)=12291
```

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh + vagrant ssh
Launched reduce tasks=1
Data-local map tasks=9
Total time spent by all maps in occupied slots (ms)=223016
Total time spent by all reduces in occupied slots (ms)=12291
Total time spent by all map tasks (ms)=223016
Total time spent by all reduce tasks (ms)=12291
Total vcore-seconds taken by all map tasks=223016
Total vcore-seconds taken by all reduce tasks=12291
Total megabyte-seconds taken by all map tasks=228368384
Total megabyte-seconds taken by all reduce tasks=12585984
Map-Reduce Framework
  Map input records=5000000
  Map output records=4642067
  Map output bytes=41778603
  Map output materialized bytes=136
  Input split bytes=864
  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)=3451
  CPU time spent (ms)=19380
  Physical memory (bytes) snapshot=1738575872
  Virtual memory (bytes) snapshot=16883130368
  Total committed heap usage (bytes)=1277722624
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=990    607
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
1990    607
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

The screenshot shows a Chrome browser window with the following details:

- Title Bar:** Chrome, File, Edit, View, History, Bookmarks, People, Window, Help.
- Address Bar:** localhost:19888/jobhistory/job/job_1486255348474_0004
- User Information:** Logged in as: dr.who
- Content Area:**
 - Job Overview:** Shows the job name is Max temperature, user is vagrant, queue is default, state is SUCCEEDED, submitted on Sun Feb 05 01:25:09 UTC 2017, started at 01:25:14, finished at 01:26:04, and elapsed time was 49sec. Diagnostics show average map time of 27sec, average shuffle time of 11sec, and average merge/reduce times of 0sec.
 - ApplicationMaster:** Shows attempt number 1, start time Sun Feb 05 01:25:11 UTC 2017, node vagrant-ubuntu-trusty-64:8042, and logs link.
 - Task Details:** Shows task types (Map, Reduce), attempt types (Maps, Reduces), and counts of successful, failed, and killed tasks.

1990 and 1992 Dataset

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java$ ssh vagrant ssh
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperatureWithCombiner /user/$USER/tempdata/1990_1992 /user/$USER/output
17/02/05 01:28:28 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with -Dmapreduce.job.reduces=2
17/02/05 01:28:29 INFO input.FileInputFormat: Total input paths to process : 1
17/02/05 01:28:30 INFO mapreduce.JobSubmitter: number of splits:60
17/02/05 01:28:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1486255348474_0005
17/02/05 01:28:30 INFO impl.YarnClientImpl: Submitted application application_1486255348474_0005
17/02/05 01:28:30 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486255348474_0005/
17/02/05 01:28:30 INFO mapreduce.Job: Running job: job_1486255348474_0005 running in uber mode : false
17/02/05 01:28:30 INFO mapreduce.Job: map 0% reduce 0%
17/02/05 01:28:30 INFO mapreduce.Job: map 1% reduce 0%
17/02/05 01:29:03 INFO mapreduce.Job: map 2% reduce 0%
17/02/05 01:29:06 INFO mapreduce.Job: map 5% reduce 0%
17/02/05 01:29:10 INFO mapreduce.Job: map 7% reduce 0%
17/02/05 01:29:11 INFO mapreduce.Job: map 10% reduce 0%
17/02/05 01:29:32 INFO mapreduce.Job: map 11% reduce 0%
17/02/05 01:29:32 INFO mapreduce.Job: map 13% reduce 0%
17/02/05 01:29:30 INFO mapreduce.Job: map 15% reduce 0%
17/02/05 01:29:40 INFO mapreduce.Job: map 16% reduce 0%
17/02/05 01:29:42 INFO mapreduce.Job: map 18% reduce 0%
17/02/05 01:29:45 INFO mapreduce.Job: map 20% reduce 0%
17/02/05 01:30:05 INFO mapreduce.Job: map 21% reduce 0%
17/02/05 01:30:08 INFO mapreduce.Job: map 22% reduce 0%
17/02/05 01:30:09 INFO mapreduce.Job: map 23% reduce 0%
17/02/05 01:30:11 INFO mapreduce.Job: map 24% reduce 0%
17/02/05 01:30:12 INFO mapreduce.Job: map 26% reduce 0%
17/02/05 01:30:14 INFO mapreduce.Job: map 27% reduce 0%
17/02/05 01:30:15 INFO mapreduce.Job: map 28% reduce 0%
17/02/05 01:30:18 INFO mapreduce.Job: map 30% reduce 0%
17/02/05 01:30:30 INFO mapreduce.Job: map 30% reduce 10%
17/02/05 01:30:37 INFO mapreduce.Job: map 30% reduce 10%
17/02/05 01:30:40 INFO mapreduce.Job: map 33% reduce 0%
17/02/05 01:30:42 INFO mapreduce.Job: map 34% reduce 10%
17/02/05 01:30:43 INFO mapreduce.Job: map 35% reduce 10%
17/02/05 01:30:44 INFO mapreduce.Job: map 37% reduce 10%
17/02/05 01:30:45 INFO mapreduce.Job: map 38% reduce 10%
17/02/05 01:30:48 INFO mapreduce.Job: map 38% reduce 13%
17/02/05 01:31:03 INFO mapreduce.Job: map 39% reduce 13%
17/02/05 01:31:04 INFO mapreduce.Job: map 40% reduce 13%
17/02/05 01:31:07 INFO mapreduce.Job: map 42% reduce 13%
17/02/05 01:31:09 INFO mapreduce.Job: map 43% reduce 13%
17/02/05 01:31:10 INFO mapreduce.Job: map 45% reduce 13%
```

```
vagrant@vagrant-ubuntu-trusty-64: ~/hadoop-book/ch02-mr-intro/src/main/java$ ssh vagrant ssh
Data-local map tasks=61
Total time spent by all maps in occupied slots (ms)=1610447
Total time spent by all reduces in occupied slots (ms)=223445
Total time spent by all map tasks (ms)=1610447
Total time spent by all reduce tasks (ms)=223445
Total vcore-seconds taken by all map tasks=1610447
Total vcore-seconds taken by all reduce tasks=223445
Total megabyte-seconds taken by all map tasks=1649097728
Total megabyte-seconds taken by all reduce tasks=228807680
Map-Reduce Framework
  Map input records=35000000
  Map output records=30592765
  Map output bytes=275334885
  Map output materialized bytes=1031
  Input split bytes=6780
  Combine input records=30592765
  Combine output records=61
  Reduce input groups=2
  Reduce shuffle bytes=1031
  Reduce input records=61
  Reduce output records=2
  Spilled Records=122
  Shuffled Maps =60
  Failed Shuffles=0
  Merged Map outputs=60
  GC time elapsed (ms)=25265
  CPU time spent (ms)=137960
  Physical memory (bytes) snapshot=12370255872
  Virtual memory (bytes) snapshot=114370904064
  Total committed heap usage (bytes)=9265987584
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=7993010283
File Output Format Counters
  Bytes Written=18
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
1990 607
1992 605
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ ]
```

Chrome File Edit View History Bookmarks People Window Help

Illinois Institute of Technology ubunti Xenial 16.04 custom MapReduce Job job_1486255348474_0005

localhost:19888/jobhistory/job/job_1486255348474_0005

Logged in as: dr.who

hadoop MapReduce Job job_1486255348474_0005

Job Overview

Job Name:	Max temperature
User Name:	vagrant
Queue:	default
State:	SUCCEEDED
Uberized:	false
Submitted:	Sun Feb 05 01:28:30 UTC 2017
Started:	Sun Feb 05 01:28:34 UTC 2017
Finished:	Sun Feb 05 01:33:57 UTC 2017
Elapsed:	5mins, 22sec
Diagnostics:	
Average Map Time	26sec
Average Shuffle Time	3mins, 43sec
Average Merge Time	0sec
Average Reduce Time	0sec

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	Sun Feb 05 01:28:31 UTC 2017	vagrant-ubuntu-trusty-64:8042	logs

Task Type	Total	Complete
Map	60	60
Reduce	1	1

Attempt Type	Failed	Killed	Successful
Maps	0	1	60
Reduces	0	0	1

1990 to 1993 Dataset

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ ssh vagrant ssh
[vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop jar mt.jar MaxTemperatureWithCombiner /user/$USER/tempdata/1990_1991_1992_1]
993 /user/$USER/output
17/02/05 01:35:05 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/05 01:35:06 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/05 01:35:06 INFO input.FileInputFormat: Total input paths to process : 1
17/02/05 01:35:06 INFO mapreduce.JobSubmitter: number of splits:114
17/02/05 01:35:07 INFO impl.YarnClientImpl: Submitted application application_1486255348474_0006
17/02/05 01:35:07 INFO mapreduce.Job: The url to track the job: http://vagrant-ubuntu-trusty-64:8088/proxy/application_1486255348474_0006/
17/02/05 01:35:07 INFO mapreduce.Job: Running job: job_1486255348474_0006
17/02/05 01:35:13 INFO mapreduce.Job: Job job_1486255348474_0006 running in uber mode : false
17/02/05 01:35:13 INFO mapreduce.Job: map 0% reduce 0%
17/02/05 01:35:40 INFO mapreduce.Job: map 1% reduce 0%
17/02/05 01:35:43 INFO mapreduce.Job: map 2% reduce 0%
17/02/05 01:35:44 INFO mapreduce.Job: map 3% reduce 0%
17/02/05 01:35:47 INFO mapreduce.Job: map 4% reduce 0%
17/02/05 01:35:48 INFO mapreduce.Job: map 5% reduce 0%
17/02/05 01:36:10 INFO mapreduce.Job: map 6% reduce 0%
17/02/05 01:36:13 INFO mapreduce.Job: map 7% reduce 0%
17/02/05 01:36:16 INFO mapreduce.Job: map 8% reduce 0%
17/02/05 01:36:18 INFO mapreduce.Job: map 9% reduce 0%
17/02/05 01:36:19 INFO mapreduce.Job: map 10% reduce 0%
17/02/05 01:36:20 INFO mapreduce.Job: map 11% reduce 0%
17/02/05 01:36:44 INFO mapreduce.Job: map 12% reduce 0%
17/02/05 01:36:47 INFO mapreduce.Job: map 13% reduce 0%
17/02/05 01:36:48 INFO mapreduce.Job: map 14% reduce 0%
17/02/05 01:36:49 INFO mapreduce.Job: map 15% reduce 0%
17/02/05 01:36:51 INFO mapreduce.Job: map 16% reduce 0%
17/02/05 01:37:14 INFO mapreduce.Job: map 17% reduce 0%
17/02/05 01:37:17 INFO mapreduce.Job: map 18% reduce 0%
17/02/05 01:37:18 INFO mapreduce.Job: map 19% reduce 0%
17/02/05 01:37:20 INFO mapreduce.Job: map 20% reduce 0%
17/02/05 01:37:22 INFO mapreduce.Job: map 21% reduce 0%
17/02/05 01:37:44 INFO mapreduce.Job: map 21% reduce 7%
17/02/05 01:37:46 INFO mapreduce.Job: map 22% reduce 7%
17/02/05 01:37:48 INFO mapreduce.Job: map 23% reduce 7%
17/02/05 01:37:50 INFO mapreduce.Job: map 24% reduce 7%
17/02/05 01:37:52 INFO mapreduce.Job: map 25% reduce 7%
17/02/05 01:37:53 INFO mapreduce.Job: map 25% reduce 8%
17/02/05 01:38:08 INFO mapreduce.Job: map 26% reduce 8%
17/02/05 01:38:13 INFO mapreduce.Job: map 27% reduce 8%
17/02/05 01:38:14 INFO mapreduce.Job: map 27% reduce 9%
17/02/05 01:38:16 INFO mapreduce.Job: map 28% reduce 9%
17/02/05 01:38:19 INFO mapreduce.Job: map 29% reduce 9%
```

```
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ ssh vagrant@vagrant
Total time spent by all reduces in occupied slots (ms)=480833
Total time spent by all map tasks (ms)=2991451
Total time spent by all reduce tasks (ms)=480833
Total vcore-seconds taken by all map tasks=2991451
Total vcore-seconds taken by all reduce tasks=480833
Total megabyte-seconds taken by all map tasks=3063245824
Total megabyte-seconds taken by all reduce tasks=492372992
Map-Reduce Framework
Map input records=65000000
Map output records=55358499
Map output bytes=498226491
Map output materialized bytes=1971
Input split bytes=14022
Combine input records=55358499
Combine output records=117
Reduce input groups=4
Reduce shuffle bytes=1971
Reduce input records=117
Reduce output records=4
Spilled Records=234
Shuffled Maps =114
Failed Shuffles=0
Merged Map outputs=114
GC time elapsed (ms)=47990
CPU time spent (ms)=260080
Physical memory (bytes) snapshot=23415775232
Virtual memory (bytes) snapshot=215612653568
Total committed heap usage (bytes)=17561493504
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=15257831563
File Output Format Counters
Bytes Written=36
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output/part-r-00000
1990 607
1991 607
1992 605
1993 567
vagrant@vagrant-ubuntu-trusty-64:~/hadoop-book/ch02-mr-intro/src/main/java$
```

The screenshot shows a Chrome browser window with the following details:

- Title Bar:** Chrome, File, Edit, View, History, Bookmarks, People, Window, Help.
- Address Bar:** Illinois Institute of Technology > ubuntui Xenial 16.04 custom > MapReduce Job job_1486255348474_0006
- User Information:** Logged in as: dr.who
- Job Overview:**
 - Job Name:** Max temperature
 - User Name:** vagrant
 - Queue:** default
 - State:** SUCCEEDED
 - Uberized:** false
 - Submitted:** Sun Feb 05 01:35:07 UTC 2017
 - Started:** Sun Feb 05 01:35:11 UTC 2017
 - Finished:** Sun Feb 05 01:45:21 UTC 2017
 - Elapsed:** 10mins, 10sec
 - Diagnostics:**
 - Average Map Time: 26sec
 - Average Shuffle Time: 8mins, 0sec
 - Average Merge Time: 0sec
 - Average Reduce Time: 0sec
- ApplicationMaster:**

Attempt Number	Start Time	Node	Logs
1	Sun Feb 05 01:35:08 UTC 2017	vagrant-ubuntu-trusty-64:8042	logs

Task Type	Total	Complete
Map	114	114
Reduce	1	1

Attempt Type	Failed	Killed	Successful
Maps	0	0	114
Reduces	0	0	1
- System Tray:** Shows various application icons including Finder, Mail, Safari, and others.