**MAP/REDUCE IMPLEMENTATION**

**CSE 5331 – PROJECT 3**

**PRUTHVI RAJ MUTHYAM**

**VAMSI KRISHNA YALAMANCHILI**

**Table of contents:**

PROJECT STATEMENT …………………………………………………………………………………………………………… 3

FILE DESCRIPTIONS ………………………………………………………………………………………………………………… 3

DIVISION OF LABOUR …………………………………………………………………………………………………………….. 3

RESULT ANALYSIS …………………………………………………………………………………………………………………… 4

PERFORMANCE MEASURE ……………………………………………………………………………………………………. 12

LOGICAL ERRORS AND HANDLING …………………………………………………………………………………… 11

OUTPUT …………………………………………………………………………………………………………………………… 13

**PROJECT STATEMENT:**

In this project we implemented the Map/Reduce paradigm. We used it to analyze US Census Datasets. We used Hadoop single node cluster version 2.7.3 on windows. Average salaries of all the states of males and females for each of the 5 years are calculated. Age count of different age groups for each of the 5 years is also calculated. After that analysis of the results are done by means of charts and histograms.

For performance analysis we have reduced the max split size to be 20 MB and this resulted in input splits to be 398 instead of the 63 splits we had when dividing the inputs in size 128 MB by default. Different parameters of the map/reduce are analyzed for these two cases of split sizes. As we were working on a single node cluster there is not much scope for changing the no of mappers and reducers. As we were working on quad core processor 2 cores were dedicated for map tasks and the other two for reduce tasks.

**FILE DESCRIPTIONS:**

We created two Java files one for the first task and the other for the 2nd task and corresponding jars

**DIVISION OF WORK:**

The distribution of project work is as following:

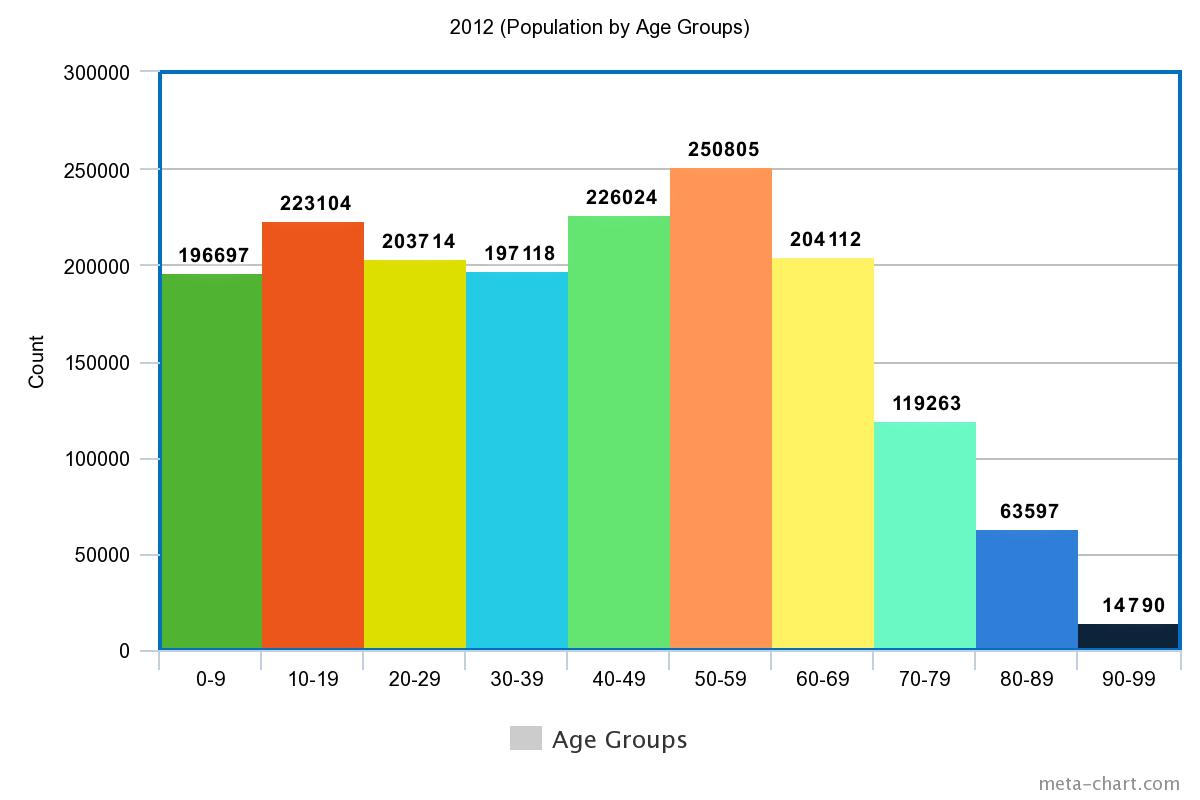
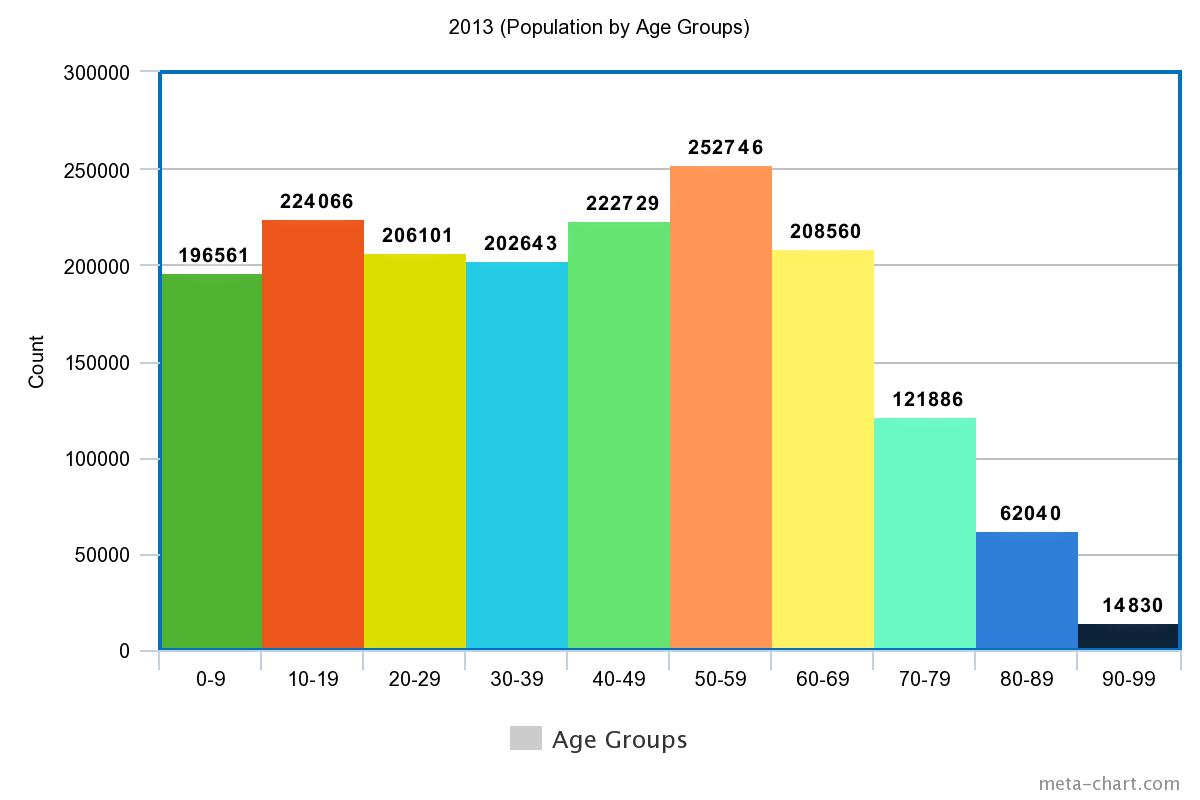
Pruthvi Raj Muthyam:

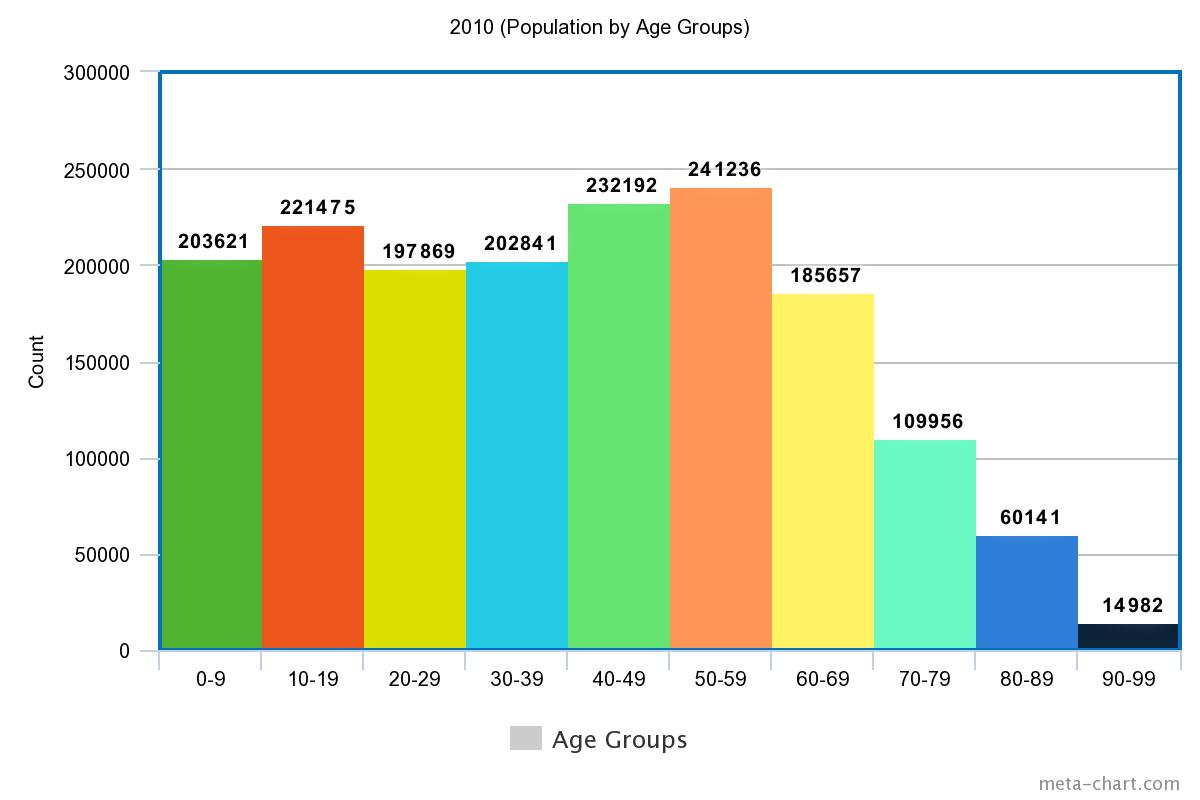
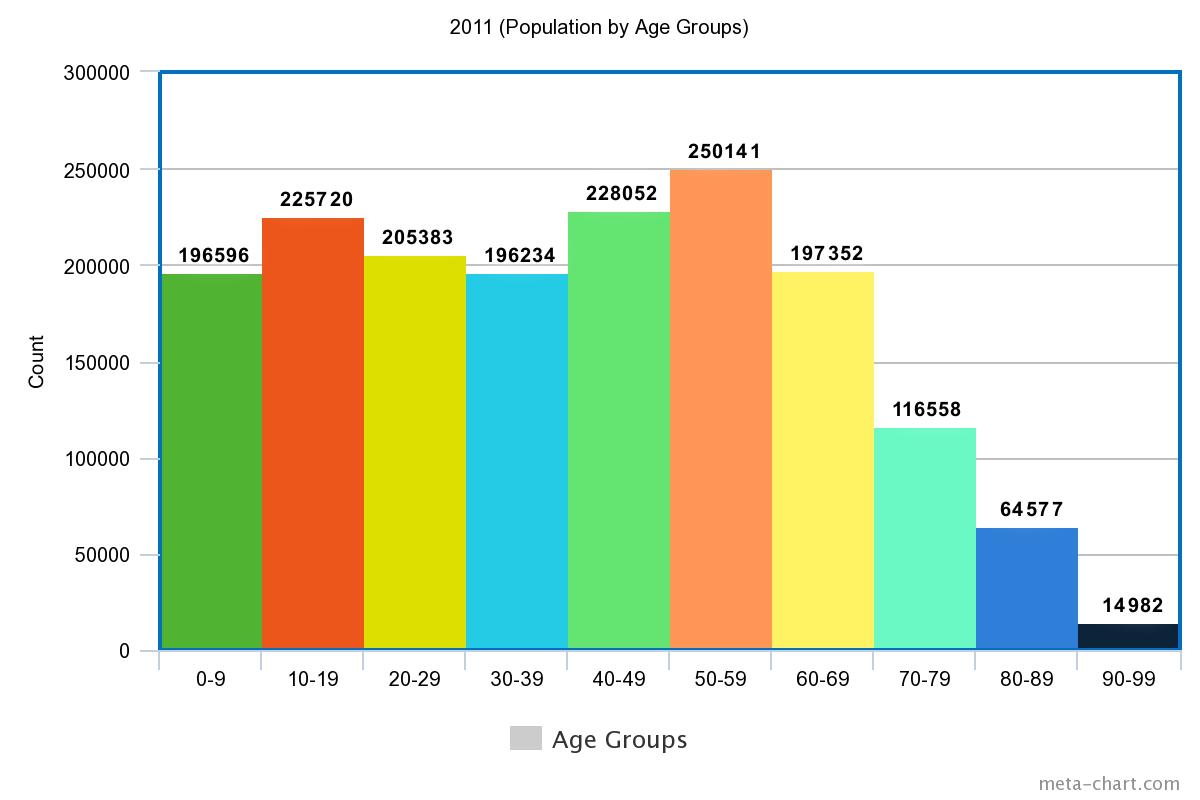
* Installed the setup of Hadoop in his system.
* Implemented the Map/Reduce for the given data set.
* Both reviewed and understood the requirements of Map/reduce and increasing the performance of the map/reduce
* Performance Analysis.

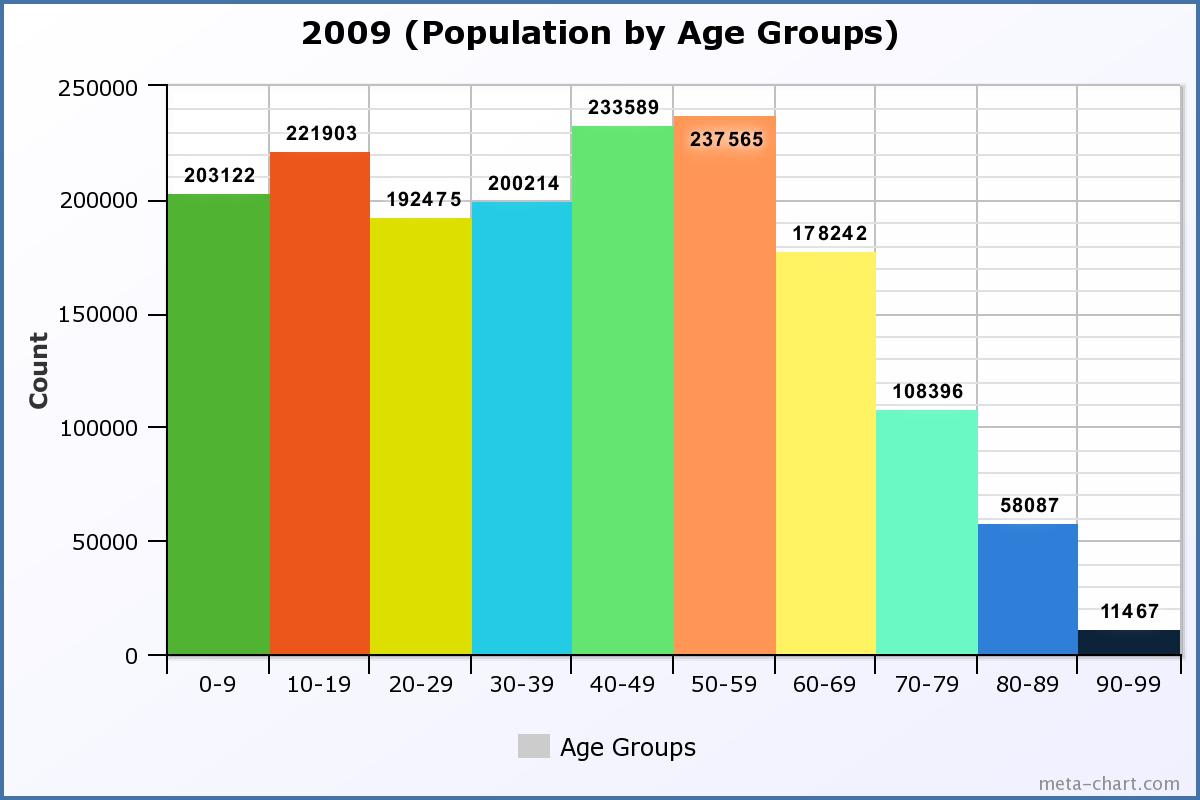
Yalamanchili Vamsi Krishna:

* Installed the setup of Hadoop in his system.
* Implemented the Map/Reduce for the given data set.
* Both reviewed and understood the requirements of Map/reduce and increasing the performance of the map/reduce
* Result Analysis.

**RESULT ANALYSIS:**

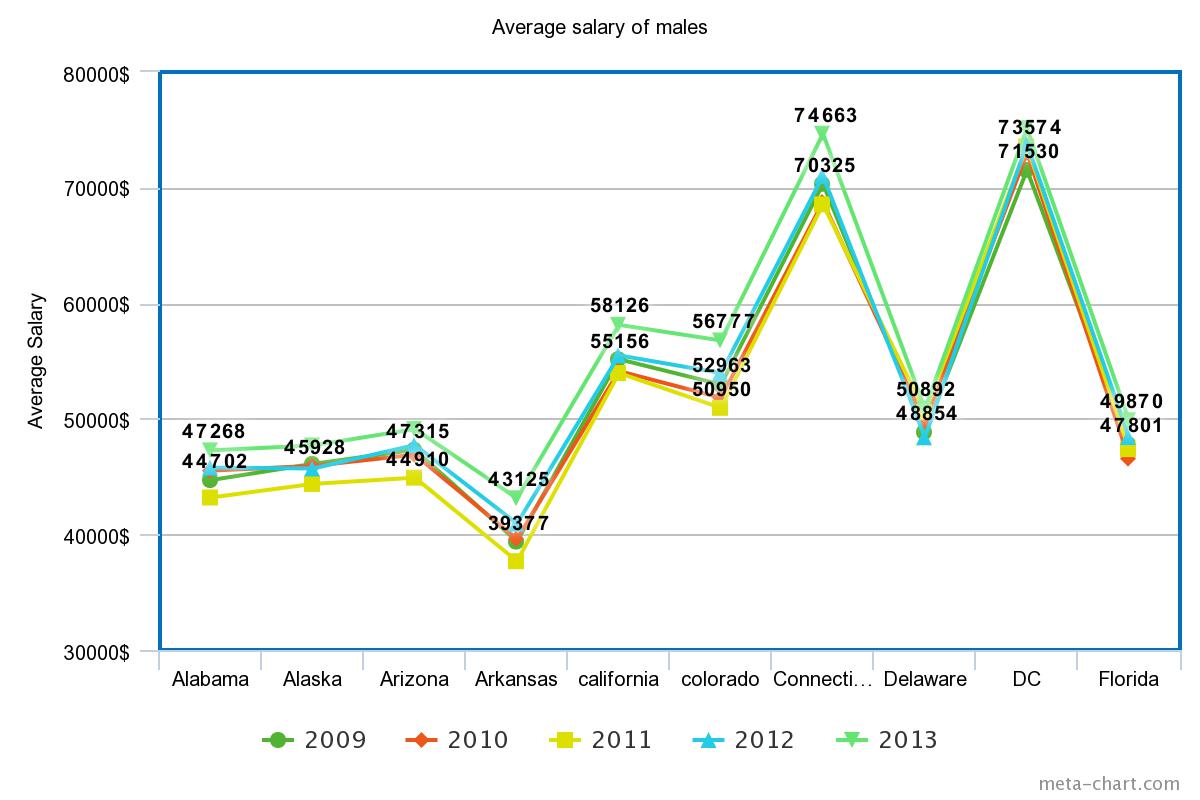
****

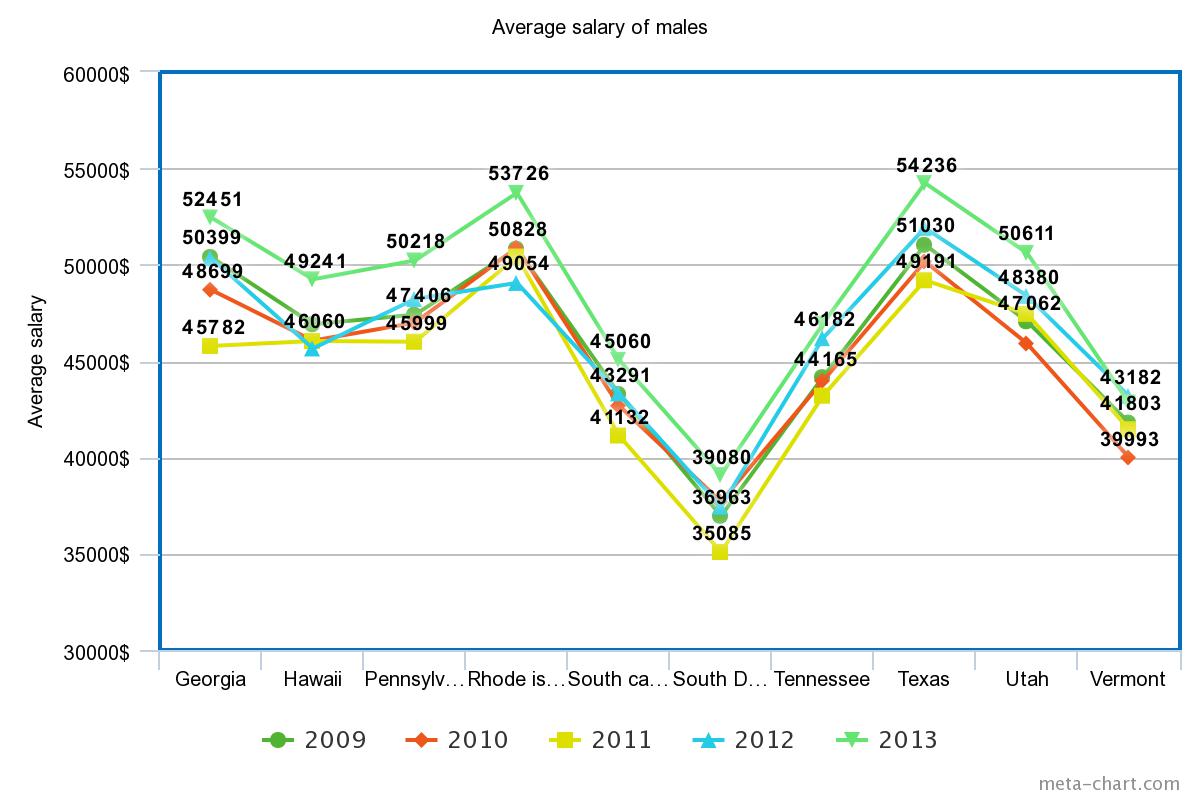
****

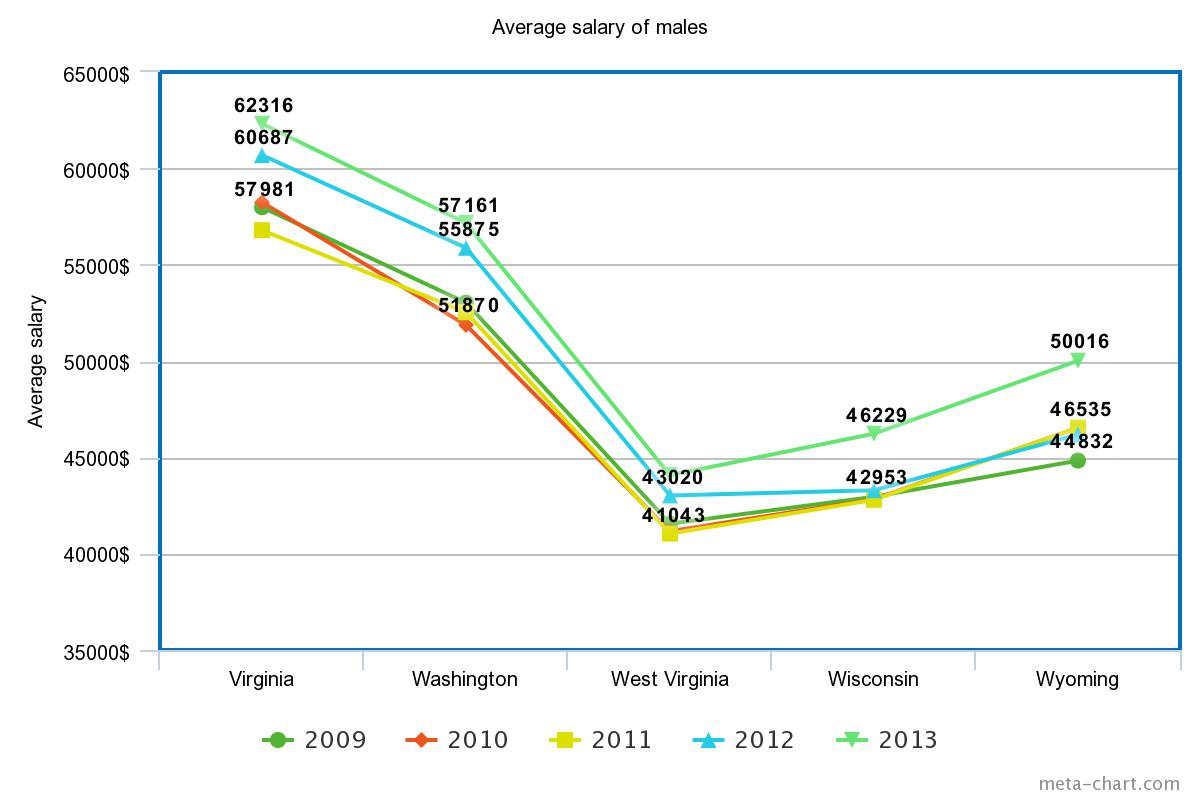
****

**2nd TASK RESULT ANALYSIS:**

* In all the years from 2009 – 2013 population in the age group 50-59 are the highest.
* In all the years from 2009 - 2013 90-99 age group had the lowest population.
* The population in the age groups 50-59, 60-69 and 70-79 increased steadily over the years from 2009 – 2013.
* The population in the age group 40-49 decreased steadily over the years from 2009-2013.
* The age group 60-69 had the highest percentage of increase in the population of 6.2% from the 2010 to 2011.
* In all the years populations in the age groups 50-59 and 40-49 stood first and second consecutively in terms of count of population.

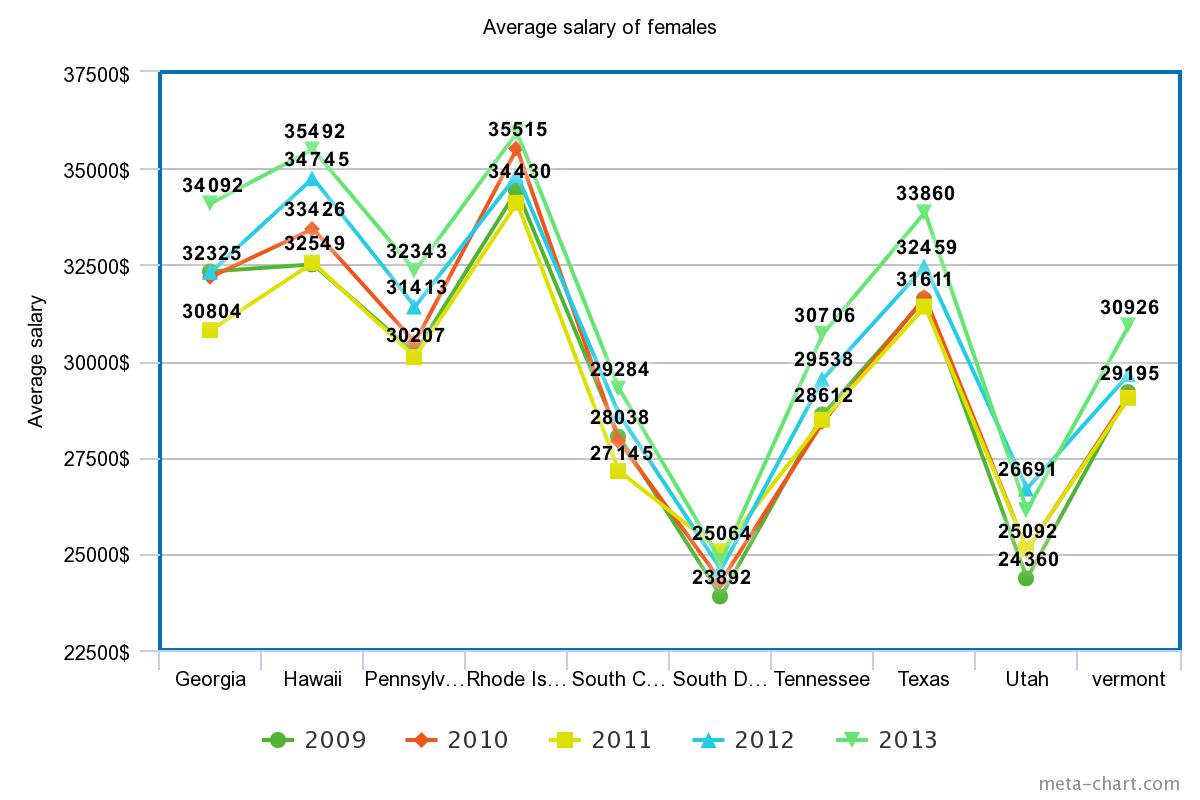


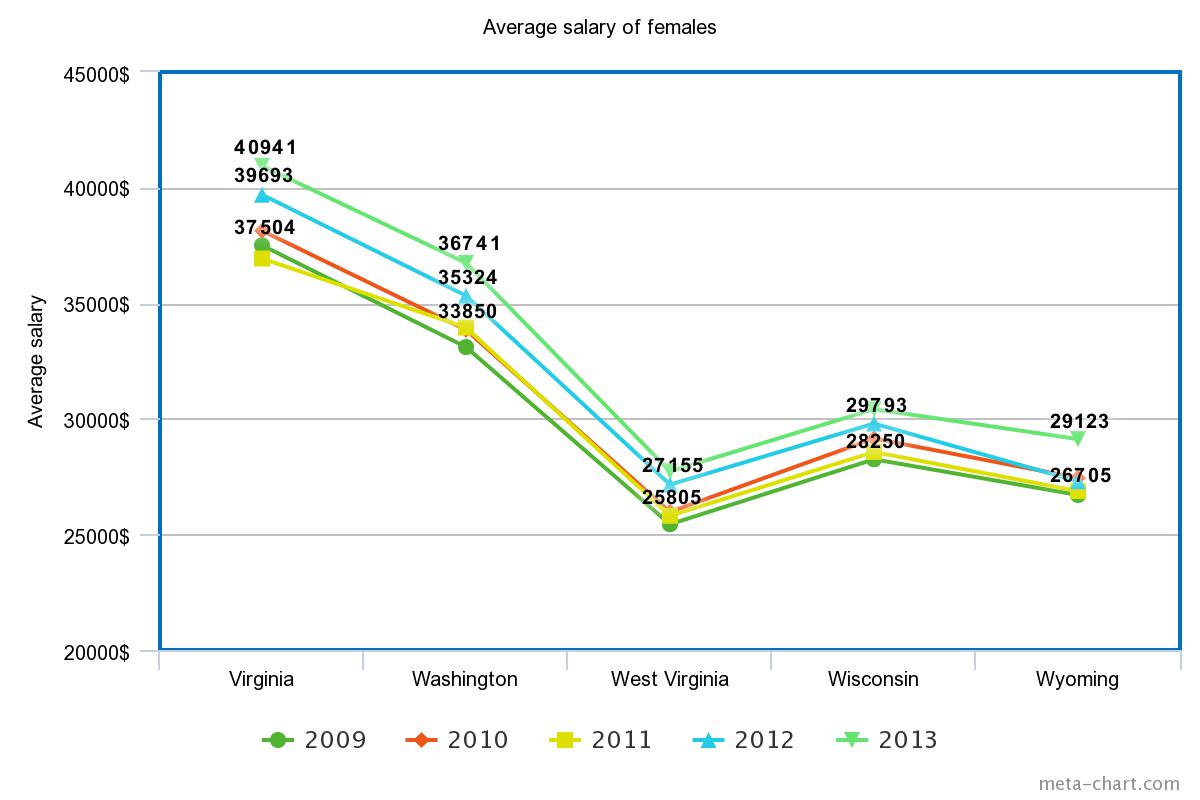


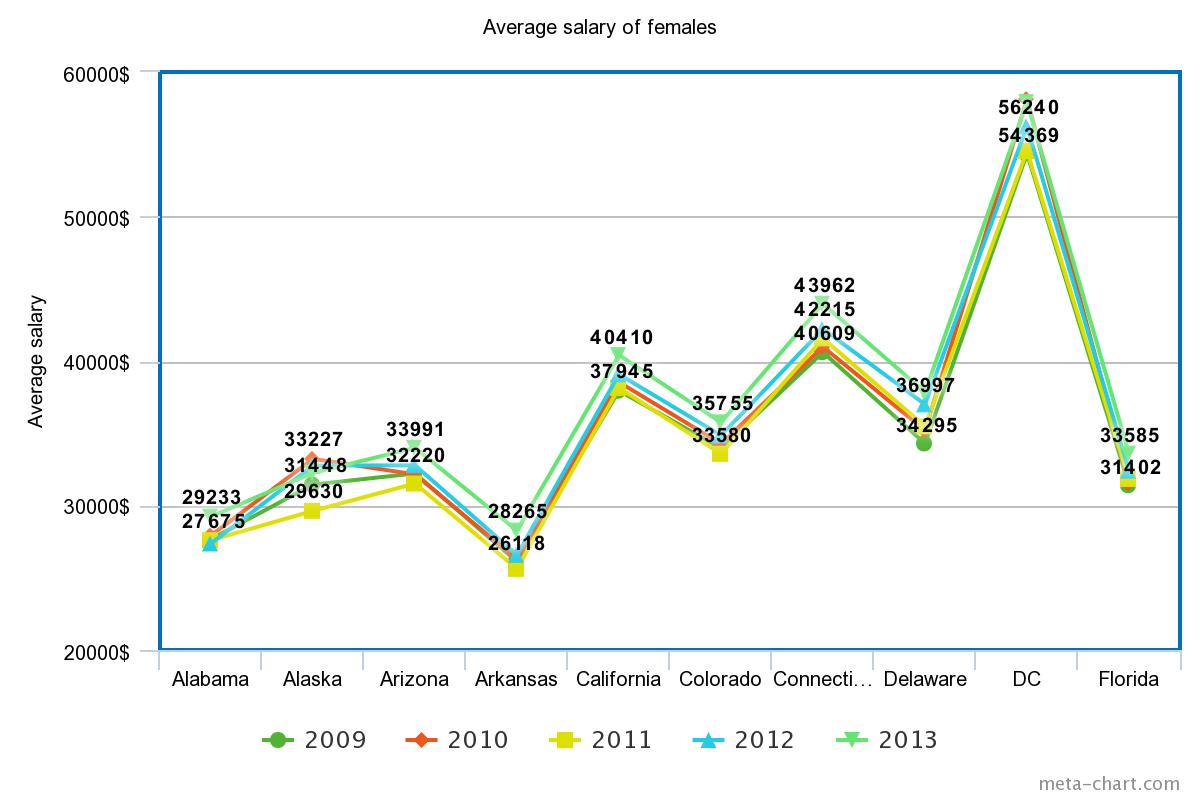


**1ST TASK RESULT ANALYSIS (MALES):**

* Males in South Dakota earned the least average salaries in all the 5 years.
* Males in District of Columbia earned the highest average salary of $75135 in the year 2013.
* The year 2011 had accounted for the least amount of average salaries earned among males in all 90% of the states.
* Males in the state of South Dakota earned the least average salary of $35085 in the year 2011. Least among all the states in all the 5 years.
* Males in the state of Georgia had the highest percentage of increase in the average salaries from 2011 to 2013 by 14.5%.







**1ST TASK RESULT ANALYSIS (FEMALES):**

* Females in DC earned the highest average salary in all the years from 2009-2013.
* Females in South Dakota earned lowest average salaries of all states in all the five years.
* Same as in males the year 2011 accounted for the least average salaries earned in females in 90% of the states in all the years.
* The year 2010 accounted for the highest average salary earned by females in all the states for all the five years by DC at $58117.
* The least average salary for females came from South Dakota in year 2013 with a value of $23892.

**Logical errors and handling:**

* **Installation issues:**

Faced numerous issues while setting up the Hadoop environment like setting up the Hadoop environment path, namenode & datanode paths and configuring various .xml configuration files. We started with setting up on Linux but eventually set up a single node cluster on Windows.

* **NumberFormatException:**

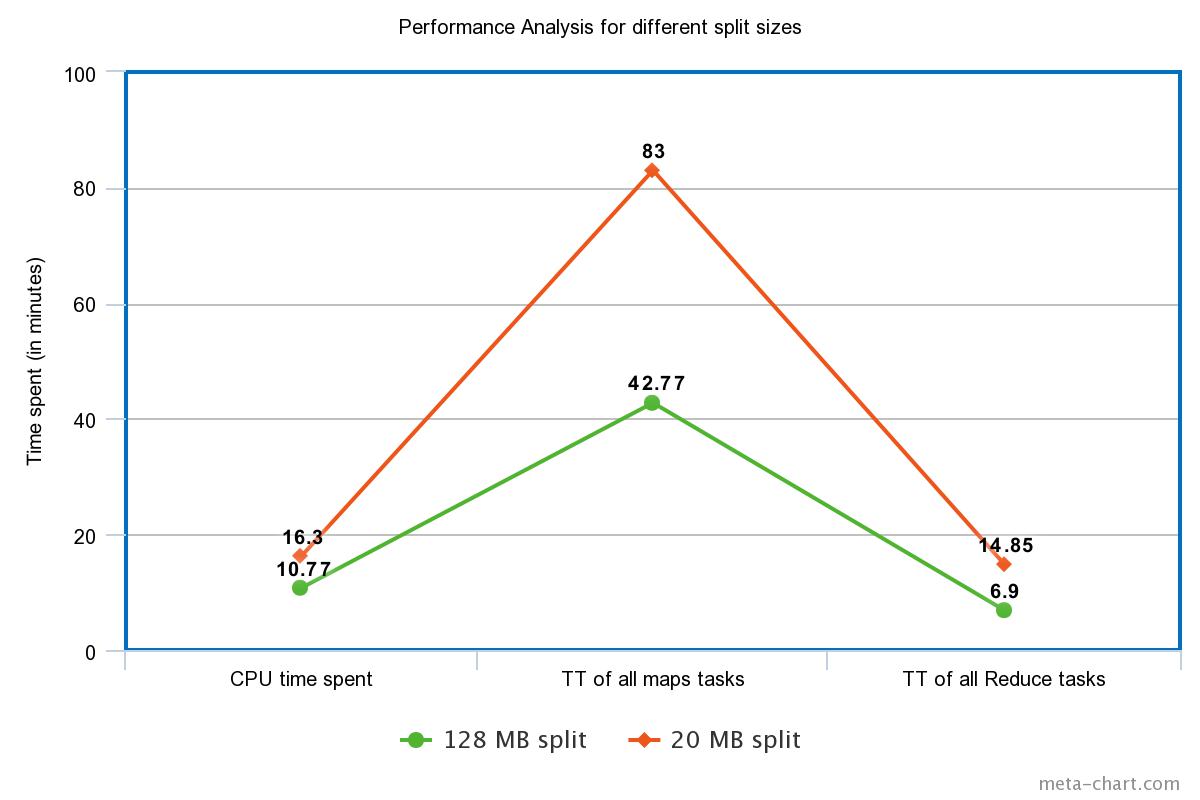
While trying to check for the the state code by converting it to an “int “datatype it was throwing a “NumberFormat” Exception. We resolved it by catching the excpetion by means of a try-catch block.

* **Negative Average Salary values:**

We were getting negative average salary values for some states. After examining the error that the sum of the salaries while calculating average might be crossing the int range. So we changed the datatype to long and corresponding LongWritables in the mapper and reducer input and output types. That solved our problem.

**PERFORMANCE ANALYSIS:**

Hadoop version 2.7.3 has a default split size of 128 MB. With this split size and the datasets given it was resulting in 63 splits which resulted in 63 map tasks and 1 reduce task. For measuring performance over different scenarios, we reduced the max split size to 20 MB. This resulted in 398 splits leading to 398 map tasks for task 1. Some important variables are analyzed for performance for the two cases below:



1. Launched map tasks: With the split size reduced the no of splits increased from 63 to 398.
2. Input split bytes: The input split size accordingly changed from 6678 bytes to 42188 bytes.
3. CPU time spent: As the no of splits increased the total time taken by CPU to process each task increased which resulted in a value of 976092(ms) as opposed to 640760 (ms).
4. Total time spent by all map tasks: As the no of map tasks increased the total time taken by all map tasks including everything from reading, processing and emitting output increased from 2566015ms to 5004102ms
5. Total time spent by all reduce tasks: The no of intermediate key value pairs emitted will increase now as the no of map tasks are increased. The time spent increased from 414155ms to 891171ms

**OUTPUT:**

**Task 1:**

C:\Users\pruthvi\Downloads\hadoop-2.7.3\sbin>start-all.cmd

C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2>javac -Xlint -classpath C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/common/hadoop-common-2.7.3.jar;C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.3.jar;C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/hdfs/hadoop-hdfs-2.7.3.jar;C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/common/lib/hadoop-annotations-2.7.3.jar;C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/common/lib/log4j-1.2.17.jar;C:/Users/pruthvi/Downloads/hadoop-2.7.3/share/hadoop/common/lib/commons-cli-1.2.jar -d Assignment\_classes Assignment.java

Assignment.java:188: warning: [deprecation] Job(Configuration,String) in Job has been deprecated

Job j2 = new Job(conf,"Assignment");

^

1 warning

C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2>jar -cvf Assignment.jar -C Assignment\_classes .

added manifest

adding: Assignment$WordMapper.class(in = 5194) (out= 2691)(deflated 48%)

adding: Assignment$WordReducer.class(in = 1673) (out= 718)(deflated 57%)

adding: Assignment.class(in = 1850) (out= 942)(deflated 49%)

**Default split size: 128 MB**

C:\Users\pruthvi\Downloads\hadoop-2.7.3\sbin>hadoop jar C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2/Assignment.jar Assignment /input2 /outputsplit -Dmapred.map.tasks=70 -Dmapred.reduce.tasks=5

17/12/06 02:40:07 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/12/06 02:40:07 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

17/12/06 02:40:08 INFO input.FileInputFormat: Total input paths to process : 2

17/12/06 02:40:09 INFO mapreduce.JobSubmitter: number of splits:63

17/12/06 02:40:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1512549199695\_0001

17/12/06 02:40:10 INFO impl.YarnClientImpl: Submitted application application\_1512549199695\_0001

17/12/06 02:40:10 INFO mapreduce.Job: The url to track the job: http://pruthvi:8088/proxy/application\_1512549199695\_0001/

17/12/06 02:40:10 INFO mapreduce.Job: Running job: job\_1512549199695\_0001

17/12/06 02:40:17 INFO mapreduce.Job: Job job\_1512549199695\_0001 running in uber mode : false

17/12/06 02:40:17 INFO mapreduce.Job: map 0% reduce 0%

17/12/06 02:40:36 INFO mapreduce.Job: map 1% reduce 0%

17/12/06 02:40:43 INFO mapreduce.Job: map 2% reduce 0%

17/12/06 02:40:52 INFO mapreduce.Job: map 3% reduce 0%

17/12/06 02:41:01 INFO mapreduce.Job: map 4% reduce 0%

17/12/06 02:41:10 INFO mapreduce.Job: map 5% reduce 0%

17/12/06 02:41:15 INFO mapreduce.Job: map 6% reduce 0%

17/12/06 02:41:18 INFO mapreduce.Job: map 7% reduce 0%

17/12/06 02:41:21 INFO mapreduce.Job: map 8% reduce 0%

17/12/06 02:41:24 INFO mapreduce.Job: map 9% reduce 0%

17/12/06 02:41:25 INFO mapreduce.Job: map 10% reduce 0%

17/12/06 02:41:33 INFO mapreduce.Job: map 11% reduce 0%

17/12/06 02:41:39 INFO mapreduce.Job: map 12% reduce 0%

17/12/06 02:41:48 INFO mapreduce.Job: map 13% reduce 0%

17/12/06 02:41:55 INFO mapreduce.Job: map 14% reduce 0%

17/12/06 02:42:03 INFO mapreduce.Job: map 15% reduce 0%

17/12/06 02:42:10 INFO mapreduce.Job: map 16% reduce 0%

17/12/06 02:42:15 INFO mapreduce.Job: map 17% reduce 0%

17/12/06 02:42:23 INFO mapreduce.Job: map 18% reduce 0%

17/12/06 02:42:24 INFO mapreduce.Job: map 19% reduce 0%

17/12/06 02:42:26 INFO mapreduce.Job: map 20% reduce 0%

17/12/06 02:42:32 INFO mapreduce.Job: map 21% reduce 0%

17/12/06 02:42:38 INFO mapreduce.Job: map 22% reduce 0%

17/12/06 02:42:48 INFO mapreduce.Job: map 23% reduce 7%

17/12/06 02:42:54 INFO mapreduce.Job: map 24% reduce 7%

17/12/06 02:42:58 INFO mapreduce.Job: map 25% reduce 7%

17/12/06 02:43:07 INFO mapreduce.Job: map 26% reduce 7%

17/12/06 02:43:11 INFO mapreduce.Job: map 27% reduce 7%

17/12/06 02:43:14 INFO mapreduce.Job: map 28% reduce 7%

17/12/06 02:43:15 INFO mapreduce.Job: map 28% reduce 8%

17/12/06 02:43:17 INFO mapreduce.Job: map 29% reduce 8%

17/12/06 02:43:18 INFO mapreduce.Job: map 29% reduce 10%

17/12/06 02:43:22 INFO mapreduce.Job: map 30% reduce 10%

17/12/06 02:43:31 INFO mapreduce.Job: map 31% reduce 10%

17/12/06 02:43:36 INFO mapreduce.Job: map 32% reduce 10%

17/12/06 02:43:46 INFO mapreduce.Job: map 33% reduce 10%

17/12/06 02:43:53 INFO mapreduce.Job: map 34% reduce 10%

17/12/06 02:43:56 INFO mapreduce.Job: map 34% reduce 11%

17/12/06 02:44:02 INFO mapreduce.Job: map 35% reduce 11%

17/12/06 02:44:06 INFO mapreduce.Job: map 36% reduce 11%

17/12/06 02:44:14 INFO mapreduce.Job: map 37% reduce 11%

17/12/06 02:44:15 INFO mapreduce.Job: map 38% reduce 12%

17/12/06 02:44:20 INFO mapreduce.Job: map 39% reduce 12%

17/12/06 02:44:21 INFO mapreduce.Job: map 39% reduce 13%

17/12/06 02:44:26 INFO mapreduce.Job: map 40% reduce 13%

17/12/06 02:44:33 INFO mapreduce.Job: map 41% reduce 13%

17/12/06 02:44:40 INFO mapreduce.Job: map 42% reduce 13%

17/12/06 02:44:47 INFO mapreduce.Job: map 43% reduce 13%

17/12/06 02:44:52 INFO mapreduce.Job: map 44% reduce 14%

17/12/06 02:45:00 INFO mapreduce.Job: map 45% reduce 14%

17/12/06 02:45:03 INFO mapreduce.Job: map 46% reduce 14%

17/12/06 02:45:04 INFO mapreduce.Job: map 47% reduce 15%

17/12/06 02:45:12 INFO mapreduce.Job: map 48% reduce 15%

17/12/06 02:45:17 INFO mapreduce.Job: map 49% reduce 16%

17/12/06 02:45:19 INFO mapreduce.Job: map 50% reduce 16%

17/12/06 02:45:26 INFO mapreduce.Job: map 51% reduce 16%

17/12/06 02:45:32 INFO mapreduce.Job: map 52% reduce 16%

17/12/06 02:45:39 INFO mapreduce.Job: map 53% reduce 16%

17/12/06 02:45:45 INFO mapreduce.Job: map 54% reduce 16%

17/12/06 02:45:47 INFO mapreduce.Job: map 55% reduce 16%

17/12/06 02:45:48 INFO mapreduce.Job: map 55% reduce 18%

17/12/06 02:45:49 INFO mapreduce.Job: map 56% reduce 18%

17/12/06 02:45:52 INFO mapreduce.Job: map 56% reduce 19%

17/12/06 02:45:53 INFO mapreduce.Job: map 57% reduce 19%

17/12/06 02:46:00 INFO mapreduce.Job: map 58% reduce 19%

17/12/06 02:46:08 INFO mapreduce.Job: map 59% reduce 19%

17/12/06 02:46:18 INFO mapreduce.Job: map 60% reduce 19%

17/12/06 02:46:25 INFO mapreduce.Job: map 61% reduce 19%

17/12/06 02:46:26 INFO mapreduce.Job: map 61% reduce 20%

17/12/06 02:46:32 INFO mapreduce.Job: map 62% reduce 20%

17/12/06 02:46:42 INFO mapreduce.Job: map 63% reduce 20%

17/12/06 02:46:46 INFO mapreduce.Job: map 64% reduce 20%

17/12/06 02:46:47 INFO mapreduce.Job: map 65% reduce 20%

17/12/06 02:46:49 INFO mapreduce.Job: map 65% reduce 21%

17/12/06 02:46:53 INFO mapreduce.Job: map 66% reduce 21%

17/12/06 02:46:55 INFO mapreduce.Job: map 66% reduce 22%

17/12/06 02:46:59 INFO mapreduce.Job: map 67% reduce 22%

17/12/06 02:47:02 INFO mapreduce.Job: map 68% reduce 22%

17/12/06 02:47:10 INFO mapreduce.Job: map 69% reduce 22%

17/12/06 02:47:15 INFO mapreduce.Job: map 70% reduce 22%

17/12/06 02:47:16 INFO mapreduce.Job: map 70% reduce 23%

17/12/06 02:47:21 INFO mapreduce.Job: map 71% reduce 23%

17/12/06 02:47:27 INFO mapreduce.Job: map 72% reduce 23%

17/12/06 02:47:33 INFO mapreduce.Job: map 73% reduce 23%

17/12/06 02:47:35 INFO mapreduce.Job: map 73% reduce 24%

17/12/06 02:47:37 INFO mapreduce.Job: map 74% reduce 24%

17/12/06 02:47:48 INFO mapreduce.Job: map 75% reduce 24%

17/12/06 02:47:57 INFO mapreduce.Job: map 76% reduce 24%

17/12/06 02:48:00 INFO mapreduce.Job: map 76% reduce 25%

17/12/06 02:48:01 INFO mapreduce.Job: map 77% reduce 25%

17/12/06 02:48:07 INFO mapreduce.Job: map 78% reduce 25%

17/12/06 02:48:15 INFO mapreduce.Job: map 79% reduce 25%

17/12/06 02:48:20 INFO mapreduce.Job: map 80% reduce 25%

17/12/06 02:48:22 INFO mapreduce.Job: map 80% reduce 26%

17/12/06 02:48:28 INFO mapreduce.Job: map 81% reduce 26%

17/12/06 02:48:36 INFO mapreduce.Job: map 82% reduce 26%

17/12/06 02:48:37 INFO mapreduce.Job: map 83% reduce 27%

17/12/06 02:48:41 INFO mapreduce.Job: map 84% reduce 27%

17/12/06 02:48:42 INFO mapreduce.Job: map 85% reduce 27%

17/12/06 02:48:43 INFO mapreduce.Job: map 85% reduce 28%

17/12/06 02:48:52 INFO mapreduce.Job: map 86% reduce 28%

17/12/06 02:49:00 INFO mapreduce.Job: map 87% reduce 28%

17/12/06 02:49:04 INFO mapreduce.Job: map 88% reduce 28%

17/12/06 02:49:05 INFO mapreduce.Job: map 88% reduce 29%

17/12/06 02:49:12 INFO mapreduce.Job: map 89% reduce 29%

17/12/06 02:49:20 INFO mapreduce.Job: map 90% reduce 29%

17/12/06 02:49:23 INFO mapreduce.Job: map 91% reduce 29%

17/12/06 02:49:24 INFO mapreduce.Job: map 91% reduce 30%

17/12/06 02:49:27 INFO mapreduce.Job: map 93% reduce 30%

17/12/06 02:49:30 INFO mapreduce.Job: map 93% reduce 31%

17/12/06 02:49:35 INFO mapreduce.Job: map 94% reduce 31%

17/12/06 02:49:38 INFO mapreduce.Job: map 95% reduce 31%

17/12/06 02:49:41 INFO mapreduce.Job: map 96% reduce 31%

17/12/06 02:49:46 INFO mapreduce.Job: map 97% reduce 31%

17/12/06 02:49:47 INFO mapreduce.Job: map 98% reduce 31%

17/12/06 02:49:49 INFO mapreduce.Job: map 98% reduce 32%

17/12/06 02:49:50 INFO mapreduce.Job: map 99% reduce 32%

17/12/06 02:49:52 INFO mapreduce.Job: map 99% reduce 33%

17/12/06 02:49:53 INFO mapreduce.Job: map 100% reduce 33%

17/12/06 02:49:55 INFO mapreduce.Job: map 100% reduce 67%

17/12/06 02:49:58 INFO mapreduce.Job: map 100% reduce 72%

17/12/06 02:50:01 INFO mapreduce.Job: map 100% reduce 78%

17/12/06 02:50:04 INFO mapreduce.Job: map 100% reduce 87%

17/12/06 02:50:07 INFO mapreduce.Job: map 100% reduce 100%

17/12/06 02:50:11 INFO mapreduce.Job: Job job\_1512549199695\_0001 completed successfully

17/12/06 02:50:11 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=120323547

FILE: Number of bytes written=248327778

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=8335795193

HDFS: Number of bytes written=6770

HDFS: Number of read operations=192

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Killed map tasks=2

Launched map tasks=65

Launched reduce tasks=1

Data-local map tasks=65

Total time spent by all maps in occupied slots (ms)=2937975

Total time spent by all reduces in occupied slots (ms)=456555

Total time spent by all map tasks (ms)=2937975

Total time spent by all reduce tasks (ms)=456555

Total vcore-milliseconds taken by all map tasks=2937975

Total vcore-milliseconds taken by all reduce tasks=456555

Total megabyte-milliseconds taken by all map tasks=3008486400

Total megabyte-milliseconds taken by all reduce tasks=467512320

Map-Reduce Framework

Map input records=8419170

Map output records=3947035

Map output bytes=112429465

Map output materialized bytes=120323913

Input split bytes=6678

Combine input records=0

Combine output records=0

Reduce input groups=250

Reduce shuffle bytes=120323913

Reduce input records=3947035

Reduce output records=250

Spilled Records=7894070

Shuffled Maps =63

Failed Shuffles=0

Merged Map outputs=63

GC time elapsed (ms)=48319

CPU time spent (ms)=642868

Physical memory (bytes) snapshot=18353029120

Virtual memory (bytes) snapshot=20331798528

Total committed heap usage (bytes)=13352042496

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=8335788515

File Output Format Counters

Bytes Written=6770

**After reducing the split size to 20 MB for task 1:**

C:\Users\pruthvi\Downloads\hadoop-2.7.3\sbin>hadoop jar C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2/Assignment.jar Assignment /input2 /outputsplittry2 -Dmapred.map.tasks=70 -Dmapred.reduce.tasks=5

17/12/06 03:06:47 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/12/06 03:06:48 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

17/12/06 03:06:49 INFO input.FileInputFormat: Total input paths to process : 2

17/12/06 03:06:49 INFO mapreduce.JobSubmitter: number of splits:398

17/12/06 03:06:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1512549199695\_0003

17/12/06 03:06:50 INFO impl.YarnClientImpl: Submitted application application\_1512549199695\_0003

17/12/06 03:06:50 INFO mapreduce.Job: The url to track the job: http://pruthvi:8088/proxy/application\_1512549199695\_0003/

17/12/06 03:06:50 INFO mapreduce.Job: Running job: job\_1512549199695\_0003

17/12/06 03:06:57 INFO mapreduce.Job: Job job\_1512549199695\_0003 running in uber mode : false

17/12/06 03:06:57 INFO mapreduce.Job: map 0% reduce 0%

17/12/06 03:07:16 INFO mapreduce.Job: map 1% reduce 0%

17/12/06 03:07:19 INFO mapreduce.Job: map 2% reduce 0%

17/12/06 03:07:37 INFO mapreduce.Job: map 3% reduce 0%

17/12/06 03:07:51 INFO mapreduce.Job: map 4% reduce 0%

17/12/06 03:07:58 INFO mapreduce.Job: map 5% reduce 0%

17/12/06 03:08:04 INFO mapreduce.Job: map 6% reduce 0%

17/12/06 03:08:20 INFO mapreduce.Job: map 7% reduce 0%

17/12/06 03:08:26 INFO mapreduce.Job: map 8% reduce 0%

17/12/06 03:08:36 INFO mapreduce.Job: map 9% reduce 0%

17/12/06 03:08:47 INFO mapreduce.Job: map 10% reduce 0%

17/12/06 03:08:58 INFO mapreduce.Job: map 11% reduce 0%

17/12/06 03:09:05 INFO mapreduce.Job: map 12% reduce 0%

17/12/06 03:09:21 INFO mapreduce.Job: map 13% reduce 0%

17/12/06 03:09:27 INFO mapreduce.Job: map 14% reduce 0%

17/12/06 03:09:38 INFO mapreduce.Job: map 15% reduce 0%

17/12/06 03:09:46 INFO mapreduce.Job: map 16% reduce 0%

17/12/06 03:09:59 INFO mapreduce.Job: map 17% reduce 0%

17/12/06 03:23:11 INFO mapreduce.Job: map 91% reduce 30%

17/12/06 03:23:21 INFO mapreduce.Job: map 92% reduce 30%

17/12/06 03:23:23 INFO mapreduce.Job: map 92% reduce 31%

17/12/06 03:23:30 INFO mapreduce.Job: map 93% reduce 31%

17/12/06 03:23:42 INFO mapreduce.Job: map 94% reduce 31%

17/12/06 03:23:50 INFO mapreduce.Job: map 95% reduce 31%

17/12/06 03:23:52 INFO mapreduce.Job: map 95% reduce 32%

17/12/06 03:24:03 INFO mapreduce.Job: map 96% reduce 32%

17/12/06 03:24:11 INFO mapreduce.Job: map 97% reduce 32%

17/12/06 03:24:22 INFO mapreduce.Job: map 98% reduce 32%

17/12/06 03:24:25 INFO mapreduce.Job: map 98% reduce 33%

17/12/06 03:24:34 INFO mapreduce.Job: map 99% reduce 33%

17/12/06 03:24:40 INFO mapreduce.Job: map 100% reduce 33%

17/12/06 03:24:44 INFO mapreduce.Job: map 100% reduce 67%

17/12/06 03:24:47 INFO mapreduce.Job: map 100% reduce 74%

17/12/06 03:24:50 INFO mapreduce.Job: map 100% reduce 87%

17/12/06 03:24:53 INFO mapreduce.Job: map 100% reduce 97%

17/12/06 03:24:54 INFO mapreduce.Job: map 100% reduce 100%

17/12/06 03:24:57 INFO mapreduce.Job: Job job\_1512549199695\_0003 completed successfully

17/12/06 03:24:58 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=120323547

FILE: Number of bytes written=288596335

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=8337202863

HDFS: Number of bytes written=6770

HDFS: Number of read operations=1197

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=398

Launched reduce tasks=1

Data-local map tasks=398

Total time spent by all maps in occupied slots (ms)=5004102

Total time spent by all reduces in occupied slots (ms)=891171

Total time spent by all map tasks (ms)=5004102

Total time spent by all reduce tasks (ms)=891171

Total vcore-milliseconds taken by all map tasks=5004102

Total vcore-milliseconds taken by all reduce tasks=891171

Total megabyte-milliseconds taken by all map tasks=5124200448

Total megabyte-milliseconds taken by all reduce tasks=912559104

Map-Reduce Framework

Map input records=8419170

Map output records=3947035

Map output bytes=112429465

Map output materialized bytes=120325923

Input split bytes=42188

Combine input records=0

Combine output records=0

Reduce input groups=250

Reduce shuffle bytes=120325923

Reduce input records=3947035

Reduce output records=250

Spilled Records=7894070

Shuffled Maps =398

Failed Shuffles=0

Merged Map outputs=398

GC time elapsed (ms)=144923

CPU time spent (ms)=976092

Physical memory (bytes) snapshot=113695031296

Virtual memory (bytes) snapshot=127850565632

Total committed heap usage (bytes)=79063678976

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=8337160675

File Output Format Counters

Bytes Written=6770

**Task 2 with default split size of 128 MB:**

C:\Users\pruthvi\Downloads\hadoop-2.7.3\sbin>hadoop jar C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2/AgeCount.jar AgeCount /input2 /outputageFinal -Dmapred.map.tasks=70 -Dmapred.reduce.tasks=5

Not a valid JAR: C:\Users\pruthvi\Downloads\hadoop-2.7.3\code2\AgeCount.jar

C:\Users\pruthvi\Downloads\hadoop-2.7.3\sbin>hadoop jar C:\Users\pruthvi\Downloads\hadoop-2.7.3\code3/AgeCount.jar AgeCount /input2 /outputageFinal -Dmapred.map.tasks=70 -Dmapred.reduce.tasks=5

17/12/06 11:52:01 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/12/06 11:52:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

17/12/06 11:52:03 INFO input.FileInputFormat: Total input paths to process : 2

17/12/06 11:52:04 INFO mapreduce.JobSubmitter: number of splits:63

17/12/06 11:52:04 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1512581008399\_0001

17/12/06 11:52:05 INFO impl.YarnClientImpl: Submitted application application\_1512581008399\_0001

17/12/06 11:52:05 INFO mapreduce.Job: The url to track the job: http://pruthvi:8088/proxy/application\_1512581008399\_0001/

17/12/06 11:52:05 INFO mapreduce.Job: Running job: job\_1512581008399\_0001

17/12/06 11:52:16 INFO mapreduce.Job: Job job\_1512581008399\_0001 running in uber mode : false

17/12/06 11:52:16 INFO mapreduce.Job: map 0% reduce 0%

17/12/06 11:52:40 INFO mapreduce.Job: map 1% reduce 0%

17/12/06 11:52:49 INFO mapreduce.Job: map 2% reduce 0%

17/12/06 11:52:56 INFO mapreduce.Job: map 3% reduce 0%

17/12/06 11:53:05 INFO mapreduce.Job: map 4% reduce 0%

17/12/06 11:53:11 INFO mapreduce.Job: map 5% reduce 0%

17/12/06 11:53:16 INFO mapreduce.Job: map 6% reduce 0%

17/12/06 11:53:24 INFO mapreduce.Job: map 7% reduce 0%

17/12/06 11:53:27 INFO mapreduce.Job: map 8% reduce 0%

17/12/06 11:53:29 INFO mapreduce.Job: map 10% reduce 0%

17/12/06 11:53:37 INFO mapreduce.Job: map 11% reduce 0%

17/12/06 11:53:45 INFO mapreduce.Job: map 12% reduce 0%

17/12/06 11:53:52 INFO mapreduce.Job: map 13% reduce 0%

17/12/06 11:54:02 INFO mapreduce.Job: map 14% reduce 0%

17/12/06 11:54:08 INFO mapreduce.Job: map 15% reduce 0%

17/12/06 11:54:16 INFO mapreduce.Job: map 16% reduce 0%

17/12/06 11:54:23 INFO mapreduce.Job: map 17% reduce 0%

17/12/06 11:54:24 INFO mapreduce.Job: map 18% reduce 0%

17/12/06 11:54:28 INFO mapreduce.Job: map 19% reduce 0%

17/12/06 11:54:31 INFO mapreduce.Job: map 20% reduce 0%

17/12/06 11:54:32 INFO mapreduce.Job: map 21% reduce 0%

17/12/06 11:54:44 INFO mapreduce.Job: map 22% reduce 0%

17/12/06 11:54:51 INFO mapreduce.Job: map 23% reduce 0%

17/12/06 11:54:59 INFO mapreduce.Job: map 24% reduce 0%

17/12/06 11:55:07 INFO mapreduce.Job: map 25% reduce 0%

17/12/06 11:55:14 INFO mapreduce.Job: map 26% reduce 0%

17/12/06 11:55:16 INFO mapreduce.Job: map 27% reduce 0%

17/12/06 11:55:18 INFO mapreduce.Job: map 27% reduce 2%

17/12/06 11:55:21 INFO mapreduce.Job: map 27% reduce 8%

17/12/06 11:55:22 INFO mapreduce.Job: map 28% reduce 8%

17/12/06 11:55:25 INFO mapreduce.Job: map 29% reduce 8%

17/12/06 11:55:26 INFO mapreduce.Job: map 30% reduce 8%

17/12/06 11:55:27 INFO mapreduce.Job: map 30% reduce 9%

17/12/06 11:55:29 INFO mapreduce.Job: map 31% reduce 9%

17/12/06 11:55:30 INFO mapreduce.Job: map 31% reduce 10%

17/12/06 11:55:37 INFO mapreduce.Job: map 32% reduce 10%

17/12/06 11:55:39 INFO mapreduce.Job: map 32% reduce 11%

17/12/06 11:55:40 INFO mapreduce.Job: map 33% reduce 11%

17/12/06 11:55:47 INFO mapreduce.Job: map 34% reduce 11%

17/12/06 11:55:57 INFO mapreduce.Job: map 35% reduce 11%

17/12/06 11:56:00 INFO mapreduce.Job: map 36% reduce 11%

17/12/06 11:56:06 INFO mapreduce.Job: map 37% reduce 11%

17/12/06 11:56:10 INFO mapreduce.Job: map 38% reduce 11%

17/12/06 11:56:14 INFO mapreduce.Job: map 39% reduce 12%

17/12/06 11:56:17 INFO mapreduce.Job: map 40% reduce 12%

17/12/06 11:56:20 INFO mapreduce.Job: map 40% reduce 13%

17/12/06 11:56:23 INFO mapreduce.Job: map 41% reduce 13%

17/12/06 11:56:31 INFO mapreduce.Job: map 42% reduce 13%

17/12/06 11:56:34 INFO mapreduce.Job: map 43% reduce 13%

17/12/06 11:56:36 INFO mapreduce.Job: map 43% reduce 14%

17/12/06 11:56:41 INFO mapreduce.Job: map 44% reduce 14%

17/12/06 11:56:48 INFO mapreduce.Job: map 45% reduce 14%

17/12/06 11:56:51 INFO mapreduce.Job: map 46% reduce 14%

17/12/06 11:56:55 INFO mapreduce.Job: map 47% reduce 14%

17/12/06 11:56:58 INFO mapreduce.Job: map 47% reduce 15%

17/12/06 11:56:59 INFO mapreduce.Job: map 48% reduce 15%

17/12/06 11:57:01 INFO mapreduce.Job: map 48% reduce 16%

17/12/06 11:57:02 INFO mapreduce.Job: map 49% reduce 16%

17/12/06 11:57:06 INFO mapreduce.Job: map 50% reduce 16%

17/12/06 11:57:12 INFO mapreduce.Job: map 51% reduce 16%

17/12/06 11:57:18 INFO mapreduce.Job: map 52% reduce 16%

17/12/06 11:57:25 INFO mapreduce.Job: map 53% reduce 16%

17/12/06 11:57:29 INFO mapreduce.Job: map 54% reduce 16%

17/12/06 11:57:30 INFO mapreduce.Job: map 54% reduce 17%

17/12/06 11:57:31 INFO mapreduce.Job: map 55% reduce 17%

17/12/06 11:57:34 INFO mapreduce.Job: map 56% reduce 17%

17/12/06 11:57:37 INFO mapreduce.Job: map 57% reduce 18%

17/12/06 11:57:40 INFO mapreduce.Job: map 58% reduce 19%

17/12/06 11:57:46 INFO mapreduce.Job: map 59% reduce 19%

17/12/06 11:57:54 INFO mapreduce.Job: map 60% reduce 19%

17/12/06 11:57:55 INFO mapreduce.Job: map 60% reduce 20%

17/12/06 11:57:59 INFO mapreduce.Job: map 61% reduce 20%

17/12/06 11:58:08 INFO mapreduce.Job: map 62% reduce 20%

17/12/06 11:58:14 INFO mapreduce.Job: map 63% reduce 20%

17/12/06 11:58:19 INFO mapreduce.Job: map 64% reduce 20%

17/12/06 11:58:21 INFO mapreduce.Job: map 65% reduce 20%

17/12/06 11:58:24 INFO mapreduce.Job: map 65% reduce 21%

17/12/06 11:58:25 INFO mapreduce.Job: map 66% reduce 21%

17/12/06 11:58:27 INFO mapreduce.Job: map 67% reduce 22%

17/12/06 11:58:42 INFO mapreduce.Job: map 68% reduce 22%

17/12/06 11:58:50 INFO mapreduce.Job: map 69% reduce 22%

17/12/06 11:58:57 INFO mapreduce.Job: map 70% reduce 22%

17/12/06 11:59:03 INFO mapreduce.Job: map 71% reduce 22%

17/12/06 11:59:10 INFO mapreduce.Job: map 72% reduce 22%

17/12/06 11:59:11 INFO mapreduce.Job: map 74% reduce 22%

17/12/06 11:59:14 INFO mapreduce.Job: map 74% reduce 24%

17/12/06 11:59:18 INFO mapreduce.Job: map 75% reduce 24%

17/12/06 11:59:21 INFO mapreduce.Job: map 75% reduce 25%

17/12/06 11:59:27 INFO mapreduce.Job: map 76% reduce 25%

17/12/06 11:59:34 INFO mapreduce.Job: map 77% reduce 25%

17/12/06 11:59:43 INFO mapreduce.Job: map 78% reduce 25%

17/12/06 11:59:47 INFO mapreduce.Job: map 79% reduce 25%

17/12/06 11:59:48 INFO mapreduce.Job: map 79% reduce 26%

17/12/06 11:59:49 INFO mapreduce.Job: map 80% reduce 26%

17/12/06 11:59:56 INFO mapreduce.Job: map 81% reduce 26%

17/12/06 12:00:02 INFO mapreduce.Job: map 82% reduce 26%

17/12/06 12:00:03 INFO mapreduce.Job: map 82% reduce 27%

17/12/06 12:00:04 INFO mapreduce.Job: map 83% reduce 27%

17/12/06 12:00:08 INFO mapreduce.Job: map 83% reduce 28%

17/12/06 12:00:09 INFO mapreduce.Job: map 84% reduce 28%

17/12/06 12:00:13 INFO mapreduce.Job: map 85% reduce 28%

17/12/06 12:00:16 INFO mapreduce.Job: map 86% reduce 28%

17/12/06 12:00:23 INFO mapreduce.Job: map 87% reduce 28%

17/12/06 12:00:24 INFO mapreduce.Job: map 87% reduce 29%

17/12/06 12:00:27 INFO mapreduce.Job: map 88% reduce 29%

17/12/06 12:00:30 INFO mapreduce.Job: map 89% reduce 29%

17/12/06 12:00:37 INFO mapreduce.Job: map 90% reduce 29%

17/12/06 12:00:47 INFO mapreduce.Job: map 91% reduce 29%

17/12/06 12:00:49 INFO mapreduce.Job: map 92% reduce 29%

17/12/06 12:00:50 INFO mapreduce.Job: map 92% reduce 30%

17/12/06 12:00:53 INFO mapreduce.Job: map 93% reduce 30%

17/12/06 12:01:00 INFO mapreduce.Job: map 94% reduce 30%

17/12/06 12:01:03 INFO mapreduce.Job: map 95% reduce 31%

17/12/06 12:01:06 INFO mapreduce.Job: map 96% reduce 31%

17/12/06 12:01:09 INFO mapreduce.Job: map 97% reduce 32%

17/12/06 12:01:12 INFO mapreduce.Job: map 98% reduce 32%

17/12/06 12:01:15 INFO mapreduce.Job: map 100% reduce 32%

17/12/06 12:01:18 INFO mapreduce.Job: map 100% reduce 67%

17/12/06 12:01:21 INFO mapreduce.Job: map 100% reduce 72%

17/12/06 12:01:24 INFO mapreduce.Job: map 100% reduce 79%

17/12/06 12:01:27 INFO mapreduce.Job: map 100% reduce 86%

17/12/06 12:01:30 INFO mapreduce.Job: map 100% reduce 94%

17/12/06 12:01:33 INFO mapreduce.Job: map 100% reduce 100%

17/12/06 12:01:38 INFO mapreduce.Job: Job job\_1512581008399\_0001 completed successfully

17/12/06 12:01:38 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=132713506

FILE: Number of bytes written=273106224

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=8335795193

HDFS: Number of bytes written=930

HDFS: Number of read operations=192

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Killed map tasks=1

Launched map tasks=64

Launched reduce tasks=1

Data-local map tasks=64

Total time spent by all maps in occupied slots (ms)=2788327

Total time spent by all reduces in occupied slots (ms)=387519

Total time spent by all map tasks (ms)=2788327

Total time spent by all reduce tasks (ms)=387519

Total vcore-milliseconds taken by all map tasks=2788327

Total vcore-milliseconds taken by all reduce tasks=387519

Total megabyte-milliseconds taken by all map tasks=2855246848

Total megabyte-milliseconds taken by all reduce tasks=396819456

Map-Reduce Framework

Map input records=8419170

Map output records=8419168

Map output bytes=115875158

Map output materialized bytes=132713872

Input split bytes=6678

Combine input records=0

Combine output records=0

Reduce input groups=50

Reduce shuffle bytes=132713872

Reduce input records=8419168

Reduce output records=50

Spilled Records=16838336

Shuffled Maps =63

Failed Shuffles=0

Merged Map outputs=63

GC time elapsed (ms)=60348

CPU time spent (ms)=604507

Physical memory (bytes) snapshot=18605985792

Virtual memory (bytes) snapshot=20569100288

Total committed heap usage (bytes)=13351518208

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=8335788515

File Output Format Counters

Bytes Written=930