\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Folder P50: 50 file, total size: 1MB

----------------------------------------

w/o inverted index

-- MapReduce 1: 105s

-- MapReduce 2: 52s

----------------------------------------

w/ inverted index

-- MapReduce 1: 105s

-- MapReduce 2: 52s

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Folder P150: 150 file, total size: 4MB

----------------------------------------

w/o inverted index

-- MapReduce 1: 260s

-- MapReduce 2: 53s

----------------------------------------

w/ inverted index

-- MapReduce 1: 256s

-- MapReduce 2: 53s

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Folder P150: 50 file, total size: 50MB

----------------------------------------

w/o inverted index

-- MapReduce 1: 113s

-- MapReduce 2: 51s

----------------------------------------

w/ inverted index

-- MapReduce 1: 111s

-- MapReduce 2: 53s

Linux hadoop-cluster-1-m 5.10.0-0.deb10.16-amd64 #1 SMP Debian 5.10.127-2~bpo10+1 (2022-07-28) x86\_64

The programs included with the Debian GNU/Linux system are free software;

the exact distribution terms for each program are described in the

individual files in /usr/share/doc/\*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent

permitted by applicable law.

Last login: Sun Apr 16 05:13:26 2023 from 35.235.241.64

dannh\_cse@hadoop-cluster-1-m:~$ ./run-similarity-search.sh P5050 Q1.txt

packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-3.2.3.jar] /tmp/streamjob2753475305469249067.jar tmpDir=null

2023-04-16 13:48:59,771 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:48:59,968 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:49:00,489 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:49:00,490 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:49:01,544 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dannh\_cse/.staging/job\_1681106769753\_0054

2023-04-16 13:49:01,929 INFO mapred.FileInputFormat: Total input files to process : 51

2023-04-16 13:49:01,932 WARN concurrent.ExecutorHelper: Thread (Thread[GetFileInfo #1,5,main]) interrupted:

java.lang.InterruptedException

at com.google.common.util.concurrent.AbstractFuture.get(AbstractFuture.java:510)

at com.google.common.util.concurrent.FluentFuture$TrustedFuture.get(FluentFuture.java:88)

at org.apache.hadoop.util.concurrent.ExecutorHelper.logThrowableFromAfterExecute(ExecutorHelper.java:48)

at org.apache.hadoop.util.concurrent.HadoopThreadPoolExecutor.afterExecute(HadoopThreadPoolExecutor.java:90)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1157)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)

at java.lang.Thread.run(Thread.java:750)

2023-04-16 13:49:02,007 INFO mapreduce.JobSubmitter: number of splits:51

2023-04-16 13:49:02,180 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1681106769753\_0054

2023-04-16 13:49:02,181 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-04-16 13:49:02,379 INFO conf.Configuration: resource-types.xml not found

2023-04-16 13:49:02,379 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-04-16 13:49:02,476 INFO impl.YarnClientImpl: Submitted application application\_1681106769753\_0054

2023-04-16 13:49:02,528 INFO mapreduce.Job: The url to track the job: http://hadoop-cluster-1-m:8088/proxy/application\_1681106769753\_0054/

2023-04-16 13:49:02,530 INFO mapreduce.Job: Running job: job\_1681106769753\_0054

2023-04-16 13:49:10,618 INFO mapreduce.Job: Job job\_1681106769753\_0054 running in uber mode : false

2023-04-16 13:49:10,619 INFO mapreduce.Job: map 0% reduce 0%

2023-04-16 13:49:17,688 INFO mapreduce.Job: map 2% reduce 0%

2023-04-16 13:49:24,730 INFO mapreduce.Job: map 4% reduce 0%

2023-04-16 13:49:27,756 INFO mapreduce.Job: map 6% reduce 0%

2023-04-16 13:49:28,762 INFO mapreduce.Job: map 14% reduce 0%

2023-04-16 13:49:29,767 INFO mapreduce.Job: map 16% reduce 0%

2023-04-16 13:49:30,771 INFO mapreduce.Job: map 18% reduce 0%

2023-04-16 13:49:36,802 INFO mapreduce.Job: map 20% reduce 0%

2023-04-16 13:49:41,833 INFO mapreduce.Job: map 24% reduce 0%

2023-04-16 13:49:42,842 INFO mapreduce.Job: map 27% reduce 0%

2023-04-16 13:49:43,847 INFO mapreduce.Job: map 29% reduce 0%

2023-04-16 13:49:44,852 INFO mapreduce.Job: map 31% reduce 0%

2023-04-16 13:49:45,864 INFO mapreduce.Job: map 33% reduce 0%

2023-04-16 13:49:48,877 INFO mapreduce.Job: map 35% reduce 0%

2023-04-16 13:49:54,903 INFO mapreduce.Job: map 37% reduce 0%

2023-04-16 13:49:56,918 INFO mapreduce.Job: map 43% reduce 0%

2023-04-16 13:49:58,926 INFO mapreduce.Job: map 45% reduce 0%

2023-04-16 13:49:59,931 INFO mapreduce.Job: map 47% reduce 0%

2023-04-16 13:50:01,940 INFO mapreduce.Job: map 51% reduce 0%

2023-04-16 13:50:05,966 INFO mapreduce.Job: map 53% reduce 0%

2023-04-16 13:50:09,988 INFO mapreduce.Job: map 55% reduce 0%

2023-04-16 13:50:10,998 INFO mapreduce.Job: map 59% reduce 0%

2023-04-16 13:50:13,006 INFO mapreduce.Job: map 63% reduce 0%

2023-04-16 13:50:16,021 INFO mapreduce.Job: map 65% reduce 0%

2023-04-16 13:50:17,026 INFO mapreduce.Job: map 67% reduce 0%

2023-04-16 13:50:18,032 INFO mapreduce.Job: map 69% reduce 0%

2023-04-16 13:50:24,058 INFO mapreduce.Job: map 71% reduce 0%

2023-04-16 13:50:25,062 INFO mapreduce.Job: map 73% reduce 0%

2023-04-16 13:50:26,069 INFO mapreduce.Job: map 76% reduce 0%

2023-04-16 13:50:28,076 INFO mapreduce.Job: map 78% reduce 0%

2023-04-16 13:50:30,085 INFO mapreduce.Job: map 80% reduce 0%

2023-04-16 13:50:32,094 INFO mapreduce.Job: map 84% reduce 0%

2023-04-16 13:50:36,111 INFO mapreduce.Job: map 86% reduce 0%

2023-04-16 13:50:40,125 INFO mapreduce.Job: map 92% reduce 0%

2023-04-16 13:50:42,133 INFO mapreduce.Job: map 96% reduce 0%

2023-04-16 13:50:45,145 INFO mapreduce.Job: map 98% reduce 0%

2023-04-16 13:50:46,149 INFO mapreduce.Job: map 100% reduce 0%

2023-04-16 13:50:49,159 INFO mapreduce.Job: map 100% reduce 20%

2023-04-16 13:50:52,170 INFO mapreduce.Job: map 100% reduce 40%

2023-04-16 13:50:55,181 INFO mapreduce.Job: map 100% reduce 60%

2023-04-16 13:50:59,194 INFO mapreduce.Job: map 100% reduce 80%

2023-04-16 13:51:01,200 INFO mapreduce.Job: map 100% reduce 100%

2023-04-16 13:51:03,212 INFO mapreduce.Job: Job job\_1681106769753\_0054 completed successfully

2023-04-16 13:51:03,321 INFO mapreduce.Job: Counters: 61

File System Counters

FILE: Number of bytes read=725323

FILE: Number of bytes written=16352304

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

GS: Number of bytes read=52994444

GS: Number of bytes written=470547

GS: Number of read operations=12963

GS: Number of large read operations=0

GS: Number of write operations=3276

HDFS: Number of bytes read=6627

HDFS: Number of bytes written=0

HDFS: Number of read operations=51

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed map tasks=1

Killed reduce tasks=1

Launched map tasks=51

Launched reduce tasks=5

Rack-local map tasks=51

Total time spent by all maps in occupied slots (ms)=1189599232

Total time spent by all reduces in occupied slots (ms)=105316352

Total time spent by all map tasks (ms)=580859

Total time spent by all reduce tasks (ms)=25712

Total vcore-milliseconds taken by all map tasks=580859

Total vcore-milliseconds taken by all reduce tasks=25712

Total megabyte-milliseconds taken by all map tasks=1189599232

Total megabyte-milliseconds taken by all reduce tasks=105316352

Map-Reduce Framework

Map input records=1079588

Map output records=30550

Map output bytes=664193

Map output materialized bytes=726823

Input split bytes=6627

Combine input records=0

Combine output records=0

Reduce input groups=846

Reduce shuffle bytes=726823

Reduce input records=30550

Reduce output records=819

Spilled Records=61100

Shuffled Maps =255

Failed Shuffles=0

Merged Map outputs=255

GC time elapsed (ms)=19441

CPU time spent (ms)=102060

Physical memory (bytes) snapshot=33139421184

Virtual memory (bytes) snapshot=256205942784

Total committed heap usage (bytes)=29370613760

Peak Map Physical memory (bytes)=624570368

Peak Map Virtual memory (bytes)=4585000960

Peak Reduce Physical memory (bytes)=427917312

Peak Reduce Virtual memory (bytes)=4580597760

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=52994444

File Output Format Counters

Bytes Written=470547

2023-04-16 13:51:03,322 INFO streaming.StreamJob: Output directory: gs://dataproc-staging-us-west1-262268354580-cgi8kbph/output/P5050/output-phase1-202304161348

packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-3.2.3.jar] /tmp/streamjob2124271064825075411.jar tmpDir=null

2023-04-16 13:51:05,804 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:51:06,016 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:51:06,532 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:51:06,533 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:51:07,706 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dannh\_cse/.staging/job\_1681106769753\_0055

2023-04-16 13:51:08,251 INFO mapred.FileInputFormat: Total input files to process : 5

2023-04-16 13:51:08,324 INFO mapreduce.JobSubmitter: number of splits:16

2023-04-16 13:51:08,480 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1681106769753\_0055

2023-04-16 13:51:08,481 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-04-16 13:51:08,647 INFO conf.Configuration: resource-types.xml not found

2023-04-16 13:51:08,648 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-04-16 13:51:08,720 INFO impl.YarnClientImpl: Submitted application application\_1681106769753\_0055

2023-04-16 13:51:08,753 INFO mapreduce.Job: The url to track the job: http://hadoop-cluster-1-m:8088/proxy/application\_1681106769753\_0055/

2023-04-16 13:51:08,755 INFO mapreduce.Job: Running job: job\_1681106769753\_0055

2023-04-16 13:51:16,911 INFO mapreduce.Job: Job job\_1681106769753\_0055 running in uber mode : false

2023-04-16 13:51:16,912 INFO mapreduce.Job: map 0% reduce 0%

2023-04-16 13:51:24,008 INFO mapreduce.Job: map 6% reduce 0%

2023-04-16 13:51:27,049 INFO mapreduce.Job: map 19% reduce 0%

2023-04-16 13:51:28,057 INFO mapreduce.Job: map 31% reduce 0%

2023-04-16 13:51:30,067 INFO mapreduce.Job: map 38% reduce 0%

2023-04-16 13:51:36,102 INFO mapreduce.Job: map 50% reduce 0%

2023-04-16 13:51:37,109 INFO mapreduce.Job: map 69% reduce 0%

2023-04-16 13:51:41,132 INFO mapreduce.Job: map 75% reduce 0%

2023-04-16 13:51:45,150 INFO mapreduce.Job: map 88% reduce 0%

2023-04-16 13:51:46,155 INFO mapreduce.Job: map 100% reduce 0%

2023-04-16 13:51:53,187 INFO mapreduce.Job: map 100% reduce 20%

2023-04-16 13:51:56,200 INFO mapreduce.Job: map 100% reduce 60%

2023-04-16 13:51:57,205 INFO mapreduce.Job: map 100% reduce 80%

2023-04-16 13:51:58,209 INFO mapreduce.Job: map 100% reduce 100%

2023-04-16 13:51:59,220 INFO mapreduce.Job: Job job\_1681106769753\_0055 completed successfully

2023-04-16 13:51:59,314 INFO mapreduce.Job: Counters: 60

File System Counters

FILE: Number of bytes read=892814

FILE: Number of bytes written=7047887

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

GS: Number of bytes read=515603

GS: Number of bytes written=1858

GS: Number of read operations=132

GS: Number of large read operations=0

GS: Number of write operations=200

HDFS: Number of bytes read=2480

HDFS: Number of bytes written=0

HDFS: Number of read operations=16

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed reduce tasks=1

Launched map tasks=16

Launched reduce tasks=5

Rack-local map tasks=16

Total time spent by all maps in occupied slots (ms)=380492470

Total time spent by all reduces in occupied slots (ms)=117726225

Total time spent by all map tasks (ms)=116110

Total time spent by all reduce tasks (ms)=35925

Total vcore-milliseconds taken by all map tasks=116110

Total vcore-milliseconds taken by all reduce tasks=35925

Total megabyte-milliseconds taken by all map tasks=380492470

Total megabyte-milliseconds taken by all reduce tasks=117726225

Map-Reduce Framework

Map input records=819

Map output records=29104

Map output bytes=834576

Map output materialized bytes=893264

Input split bytes=2480

Combine input records=0

Combine output records=0

Reduce input groups=50

Reduce shuffle bytes=893264

Reduce input records=29104

Reduce output records=50

Spilled Records=58208

Shuffled Maps =80

Failed Shuffles=0

Merged Map outputs=80

GC time elapsed (ms)=6060

CPU time spent (ms)=32240

Physical memory (bytes) snapshot=11825692672

Virtual memory (bytes) snapshot=96057405440

Total committed heap usage (bytes)=10574888960

Peak Map Physical memory (bytes)=621084672

Peak Map Virtual memory (bytes)=4580618240

Peak Reduce Physical memory (bytes)=495562752

Peak Reduce Virtual memory (bytes)=4587704320

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=515603

File Output Format Counters

Bytes Written=1858

2023-04-16 13:51:59,320 INFO streaming.StreamJob: Output directory: gs://dataproc-staging-us-west1-262268354580-cgi8kbph/output/P5050/output-phase2-202304161348

dannh\_cse@hadoop-cluster-1-m:~$ ./run-similarity-search-w-prebuilt-inverted-index.sh P5050 Q1.txt

packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-3.2.3.jar] /tmp/streamjob4198390856943088428.jar tmpDir=null

2023-04-16 13:58:28,455 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:58:28,647 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:58:29,170 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 13:58:29,171 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 13:58:30,292 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dannh\_cse/.staging/job\_1681106769753\_0056

2023-04-16 13:58:30,802 INFO mapred.FileInputFormat: Total input files to process : 51

2023-04-16 13:58:30,877 INFO mapreduce.JobSubmitter: number of splits:51

2023-04-16 13:58:31,033 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1681106769753\_0056

2023-04-16 13:58:31,035 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-04-16 13:58:31,225 INFO conf.Configuration: resource-types.xml not found

2023-04-16 13:58:31,226 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-04-16 13:58:31,298 INFO impl.YarnClientImpl: Submitted application application\_1681106769753\_0056

2023-04-16 13:58:31,339 INFO mapreduce.Job: The url to track the job: http://hadoop-cluster-1-m:8088/proxy/application\_1681106769753\_0056/

2023-04-16 13:58:31,341 INFO mapreduce.Job: Running job: job\_1681106769753\_0056

2023-04-16 13:58:39,434 INFO mapreduce.Job: Job job\_1681106769753\_0056 running in uber mode : false

2023-04-16 13:58:39,435 INFO mapreduce.Job: map 0% reduce 0%

2023-04-16 13:58:46,511 INFO mapreduce.Job: map 2% reduce 0%

2023-04-16 13:58:50,536 INFO mapreduce.Job: map 6% reduce 0%

2023-04-16 13:58:51,541 INFO mapreduce.Job: map 12% reduce 0%

2023-04-16 13:58:57,573 INFO mapreduce.Job: map 14% reduce 0%

2023-04-16 13:58:59,583 INFO mapreduce.Job: map 16% reduce 0%

2023-04-16 13:59:00,589 INFO mapreduce.Job: map 20% reduce 0%

2023-04-16 13:59:01,594 INFO mapreduce.Job: map 22% reduce 0%

2023-04-16 13:59:02,599 INFO mapreduce.Job: map 24% reduce 0%

2023-04-16 13:59:07,624 INFO mapreduce.Job: map 25% reduce 0%

2023-04-16 13:59:08,629 INFO mapreduce.Job: map 27% reduce 0%

2023-04-16 13:59:09,636 INFO mapreduce.Job: map 29% reduce 0%

2023-04-16 13:59:10,642 INFO mapreduce.Job: map 31% reduce 0%

2023-04-16 13:59:11,647 INFO mapreduce.Job: map 33% reduce 0%

2023-04-16 13:59:12,652 INFO mapreduce.Job: map 35% reduce 0%

2023-04-16 13:59:17,672 INFO mapreduce.Job: map 39% reduce 0%

2023-04-16 13:59:18,676 INFO mapreduce.Job: map 41% reduce 0%

2023-04-16 13:59:20,686 INFO mapreduce.Job: map 43% reduce 0%

2023-04-16 13:59:21,694 INFO mapreduce.Job: map 45% reduce 0%

2023-04-16 13:59:22,699 INFO mapreduce.Job: map 47% reduce 0%

2023-04-16 13:59:27,720 INFO mapreduce.Job: map 51% reduce 0%

2023-04-16 13:59:28,726 INFO mapreduce.Job: map 55% reduce 0%

2023-04-16 13:59:30,737 INFO mapreduce.Job: map 57% reduce 0%

2023-04-16 13:59:32,746 INFO mapreduce.Job: map 59% reduce 0%

2023-04-16 13:59:36,769 INFO mapreduce.Job: map 61% reduce 0%

2023-04-16 13:59:37,775 INFO mapreduce.Job: map 67% reduce 0%

2023-04-16 13:59:39,788 INFO mapreduce.Job: map 69% reduce 0%

2023-04-16 13:59:42,802 INFO mapreduce.Job: map 71% reduce 0%

2023-04-16 13:59:44,810 INFO mapreduce.Job: map 73% reduce 0%

2023-04-16 13:59:46,821 INFO mapreduce.Job: map 75% reduce 0%

2023-04-16 13:59:47,825 INFO mapreduce.Job: map 80% reduce 0%

2023-04-16 13:59:52,847 INFO mapreduce.Job: map 82% reduce 0%

2023-04-16 13:59:53,851 INFO mapreduce.Job: map 84% reduce 0%

2023-04-16 13:59:55,860 INFO mapreduce.Job: map 86% reduce 0%

2023-04-16 13:59:56,866 INFO mapreduce.Job: map 90% reduce 0%

2023-04-16 13:59:57,869 INFO mapreduce.Job: map 92% reduce 0%

2023-04-16 14:00:02,889 INFO mapreduce.Job: map 94% reduce 0%

2023-04-16 14:00:04,896 INFO mapreduce.Job: map 100% reduce 0%

2023-04-16 14:00:12,928 INFO mapreduce.Job: map 100% reduce 20%

2023-04-16 14:00:15,938 INFO mapreduce.Job: map 100% reduce 60%

2023-04-16 14:00:16,942 INFO mapreduce.Job: map 100% reduce 80%

2023-04-16 14:00:17,946 INFO mapreduce.Job: map 100% reduce 100%

2023-04-16 14:00:19,961 INFO mapreduce.Job: Job job\_1681106769753\_0056 completed successfully

2023-04-16 14:00:20,048 INFO mapreduce.Job: Counters: 60

File System Counters

FILE: Number of bytes read=15883691

FILE: Number of bytes written=45806864

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

GS: Number of bytes read=52994444

GS: Number of bytes written=10888384

GS: Number of read operations=12963

GS: Number of large read operations=0

GS: Number of write operations=654408

HDFS: Number of bytes read=6627

HDFS: Number of bytes written=0

HDFS: Number of read operations=51

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed reduce tasks=1

Launched map tasks=51

Launched reduce tasks=5

Rack-local map tasks=51

Total time spent by all maps in occupied slots (ms)=1202524643

Total time spent by all reduces in occupied slots (ms)=145423429

Total time spent by all map tasks (ms)=366959

Total time spent by all reduce tasks (ms)=44377

Total vcore-milliseconds taken by all map tasks=366959

Total vcore-milliseconds taken by all reduce tasks=44377

Total megabyte-milliseconds taken by all map tasks=1202524643

Total megabyte-milliseconds taken by all reduce tasks=145423429

Map-Reduce Framework

Map input records=1079588

Map output records=605130

Map output bytes=14673401

Map output materialized bytes=15885191

Input split bytes=6627

Combine input records=0

Combine output records=0

Reduce input groups=163840

Reduce shuffle bytes=15885191

Reduce input records=605130

Reduce output records=163602

Spilled Records=1210260

Shuffled Maps =255

Failed Shuffles=0

Merged Map outputs=255

GC time elapsed (ms)=14724

CPU time spent (ms)=105790

Physical memory (bytes) snapshot=33397284864

Virtual memory (bytes) snapshot=256230453248

Total committed heap usage (bytes)=29988225024

Peak Map Physical memory (bytes)=626991104

Peak Map Virtual memory (bytes)=4580974592

Peak Reduce Physical memory (bytes)=421310464

Peak Reduce Virtual memory (bytes)=4589408256

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=52994444

File Output Format Counters

Bytes Written=10888384

2023-04-16 14:00:20,054 INFO streaming.StreamJob: Output directory: gs://dataproc-staging-us-west1-262268354580-cgi8kbph/output/P5050/output-inverted-index-202304161358

packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-3.2.3.jar] /tmp/streamjob1695291499764460100.jar tmpDir=null

2023-04-16 14:00:22,514 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 14:00:22,706 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 14:00:23,195 INFO client.RMProxy: Connecting to ResourceManager at hadoop-cluster-1-m/10.138.0.4:8032

2023-04-16 14:00:23,196 INFO client.AHSProxy: Connecting to Application History server at hadoop-cluster-1-m/10.138.0.4:10200

2023-04-16 14:00:24,322 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dannh\_cse/.staging/job\_1681106769753\_0057

2023-04-16 14:00:24,854 INFO mapred.FileInputFormat: Total input files to process : 5

2023-04-16 14:00:24,919 INFO mapreduce.JobSubmitter: number of splits:15

2023-04-16 14:00:25,093 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1681106769753\_0057

2023-04-16 14:00:25,096 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-04-16 14:00:25,278 INFO conf.Configuration: resource-types.xml not found

2023-04-16 14:00:25,279 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-04-16 14:00:25,339 INFO impl.YarnClientImpl: Submitted application application\_1681106769753\_0057

2023-04-16 14:00:25,371 INFO mapreduce.Job: The url to track the job: http://hadoop-cluster-1-m:8088/proxy/application\_1681106769753\_0057/

2023-04-16 14:00:25,373 INFO mapreduce.Job: Running job: job\_1681106769753\_0057

2023-04-16 14:00:33,491 INFO mapreduce.Job: Job job\_1681106769753\_0057 running in uber mode : false

2023-04-16 14:00:33,492 INFO mapreduce.Job: map 0% reduce 0%

2023-04-16 14:00:40,554 INFO mapreduce.Job: map 7% reduce 0%

2023-04-16 14:00:43,575 INFO mapreduce.Job: map 13% reduce 0%

2023-04-16 14:00:44,582 INFO mapreduce.Job: map 33% reduce 0%

2023-04-16 14:00:45,587 INFO mapreduce.Job: map 40% reduce 0%

2023-04-16 14:00:50,616 INFO mapreduce.Job: map 47% reduce 0%

2023-04-16 14:00:52,626 INFO mapreduce.Job: map 53% reduce 0%

2023-04-16 14:00:53,630 INFO mapreduce.Job: map 60% reduce 0%

2023-04-16 14:00:54,635 INFO mapreduce.Job: map 73% reduce 0%

2023-04-16 14:00:55,640 INFO mapreduce.Job: map 80% reduce 0%

2023-04-16 14:00:59,659 INFO mapreduce.Job: map 87% reduce 0%

2023-04-16 14:01:00,664 INFO mapreduce.Job: map 93% reduce 0%

2023-04-16 14:01:01,670 INFO mapreduce.Job: map 100% reduce 0%

2023-04-16 14:01:09,712 INFO mapreduce.Job: map 100% reduce 20%

2023-04-16 14:01:12,725 INFO mapreduce.Job: map 100% reduce 60%

2023-04-16 14:01:13,729 INFO mapreduce.Job: map 100% reduce 100%

2023-04-16 14:01:15,746 INFO mapreduce.Job: Job job\_1681106769753\_0057 completed successfully

2023-04-16 14:01:15,838 INFO mapreduce.Job: Counters: 60

File System Counters

FILE: Number of bytes read=761895

FILE: Number of bytes written=6845100

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

GS: Number of bytes read=10929344

GS: Number of bytes written=2058

GS: Number of read operations=2680

GS: Number of large read operations=0

GS: Number of write operations=204

HDFS: Number of bytes read=2445

HDFS: Number of bytes written=0

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed map tasks=1

Launched map tasks=15

Launched reduce tasks=5

Rack-local map tasks=15

Total time spent by all maps in occupied slots (ms)=341325766

Total time spent by all reduces in occupied slots (ms)=124971672

Total time spent by all map tasks (ms)=104158

Total time spent by all reduce tasks (ms)=38136

Total vcore-milliseconds taken by all map tasks=104158

Total vcore-milliseconds taken by all reduce tasks=38136

Total megabyte-milliseconds taken by all map tasks=341325766

Total megabyte-milliseconds taken by all reduce tasks=124971672

Map-Reduce Framework

Map input records=163602

Map output records=22700

Map output bytes=716465

Map output materialized bytes=762315

Input split bytes=2445

Combine input records=0

Combine output records=0

Reduce input groups=51

Reduce shuffle bytes=762315

Reduce input records=22700

Reduce output records=51

Spilled Records=45400

Shuffled Maps =75

Failed Shuffles=0

Merged Map outputs=75

GC time elapsed (ms)=5436

CPU time spent (ms)=31240

Physical memory (bytes) snapshot=11157725184

Virtual memory (bytes) snapshot=91516465152

Total committed heap usage (bytes)=10013900800

Peak Map Physical memory (bytes)=624615424

Peak Map Virtual memory (bytes)=4579741696

Peak Reduce Physical memory (bytes)=418111488

Peak Reduce Virtual memory (bytes)=4586151936

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=10929344

File Output Format Counters

Bytes Written=2058

2023-04-16 14:01:15,843 INFO streaming.StreamJob: Output directory: gs://dataproc-staging-us-west1-262268354580-cgi8kbph/output/P5050/query-w-inverted-index/Q1.txt-202304161358