

```

bfezo001@lena:~$ mahout seqdirectory \
> -i dsm010/question_2/british-fiction-corpus \
> -o dsm010/question_2/british-fiction-corpus-seqfiles \
> -c UTF-8 \
> -chunk 5
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 13:39:14 INFO AbstractJob: Command line arguments: [--charset=[UTF-8], --
chunkSize=[5], --endPhase=[2147483647], --
fileFilterClass=[org.apache.mahout.text.PrefixAdditionFilter],
--input=[dsm010/question_2/british-fiction-corpus], --keyPrefix=[], --
method=[mapreduce], --output=[dsm010/question_2/british-fiction-corpus-seqfiles],
--startPhase=[0], --tempDir=[temp]]
23/01/04 13:39:14 INFO deprecation: mapred.input.dir is deprecated. Instead, use
mapreduce.input.fileinputformat.inputdir
23/01/04 13:39:14 INFO deprecation: mapred.compress.map.output is deprecated.
Instead, use mapreduce.map.output.compress
23/01/04 13:39:14 INFO deprecation: mapred.output.dir is deprecated. Instead, use
mapreduce.output.fileoutputformat.outputdir
23/01/04 13:39:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException:
Output directory dsm010/question_2/british-fiction-corpus-seqfiles already exists
    at
    org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutput
Format.java:164)
    at
    org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
    at
    org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1576)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1573)
    at java.base/java.security.AccessController.doPrivileged(Native Method)
    at java.base/javax.security.auth.Subject.doAs(Subject.java:423)
    at
    org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1845
)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1573)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1594)
    at
    org.apache.mahout.text.SequenceFilesFromDirectory.runMapReduce(SequenceFilesFromDir
ectory.java:177)
    at
    org.apache.mahout.text.SequenceFilesFromDirectory.run(SequenceFilesFromDirectory.ja
va:90)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:90)
    at
    org.apache.mahout.text.SequenceFilesFromDirectory.main(SequenceFilesFromDirectory.j
ava:64)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native
Method)
    at
    java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessor
Impl.java:62)
    at
    java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethod

```

```

AccessorImpl.java:43)
    at java.base/java.lang.reflect.Method.invoke(Method.java:566)
    at
org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:7
1)
    at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
    at org.apache.hadoop.util.ProgramDriver.driver(ProgramDriver.java:152)
    at org.apache.mahout.driver.MahoutDriver.main(MahoutDriver.java:195)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native
Method)
    at
java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessor
Impl.java:62)
    at
java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethod
AccessorImpl.java:43)
    at java.base/java.lang.reflect.Method.invoke(Method.java:566)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
bfezo001@lena:~$ mahout seq2sparse \
> -nv \> -i dsm010/question_2/british-fiction-corpus-seqfiles \
> -o dsm010/question_2/british-fiction-corpus-vectors \
>
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 13:41:48 INFO SparseVectorsFromSequenceFiles: Maximum n-gram size is: 1
23/01/04 13:41:48 INFO SparseVectorsFromSequenceFiles: Minimum LLR value: 1.0
23/01/04 13:41:48 INFO SparseVectorsFromSequenceFiles: Number of reduce tasks: 1
23/01/04 13:41:48 INFO SparseVectorsFromSequenceFiles: Tokenizing documents in
dsm010/question_2/british-fiction-corpus-seqfiles
23/01/04 13:41:48 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-vectors/tokenized-documents
23/01/04 13:41:48 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:41:48 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0056
23/01/04 13:41:49 INFO FileInputFormat: Total input files to process : 6
23/01/04 13:41:49 INFO JobSubmitter: number of splits:6
23/01/04 13:41:49 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0056
23/01/04 13:41:49 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:41:49 INFO Configuration: resource-types.xml not found
23/01/04 13:41:49 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 13:41:49 INFO YarnClientImpl: Submitted application
application_1672817468245_0056
23/01/04 13:41:49 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0056/
23/01/04 13:41:49 INFO Job: Running job: job_1672817468245_0056
23/01/04 13:41:55 INFO Job: Job job_1672817468245_0056 running in uber mode : false
23/01/04 13:41:55 INFO Job: map 0% reduce 0%
23/01/04 13:42:01 INFO Job: map 67% reduce 0%
23/01/04 13:42:02 INFO Job: map 100% reduce 0%
23/01/04 13:42:02 INFO Job: Job job_1672817468245_0056 completed successfully
23/01/04 13:42:02 INFO Job: Counters: 34
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=1590750

```

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=13528567  
HDFS: Number of bytes written=26388306  
HDFS: Number of read operations=48  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=12  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=6  
Data-local map tasks=3  
Rack-local map tasks=3  
Total time spent by all maps in occupied slots (ms)=116730  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=23346  
Total vcore-milliseconds taken by all map tasks=23346  
Total megabyte-milliseconds taken by all map tasks=119531520

#### Map-Reduce Framework

Map input records=28  
Map output records=28  
Input split bytes=960  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=322  
CPU time spent (ms)=16820  
Physical memory (bytes) snapshot=2302377984  
Virtual memory (bytes) snapshot=38256054272  
Total committed heap usage (bytes)=6316621824  
Peak Map Physical memory (bytes)=409391104  
Peak Map Virtual memory (bytes)=6383161344

#### File Input Format Counters

Bytes Read=13527607

#### File Output Format Counters

Bytes Written=26388306

23/01/04 13:42:02 INFO SparseVectorsFromSequenceFiles: Creating Term Frequency Vectors

23/01/04 13:42:02 INFO DictionaryVectorizer: Creating dictionary from dsm010/question\_2/british-fiction-corpus-vectors/tokenized-documents and saving at dsm010/question\_2/british-fiction-corpus-vectors/wordcount

23/01/04 13:42:02 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/wordcount

23/01/04 13:42:02 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:42:02 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0057

23/01/04 13:42:03 INFO FileInputFormat: Total input files to process : 6

23/01/04 13:42:03 INFO JobSubmitter: number of splits:6

23/01/04 13:42:03 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0057

23/01/04 13:42:03 INFO JobSubmitter: Executing with tokens: []

23/01/04 13:42:03 INFO YarnClientImpl: Submitted application application\_1672817468245\_0057

23/01/04 13:42:03 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0057/

23/01/04 13:42:03 INFO Job: Running job: job\_1672817468245\_0057

23/01/04 13:42:09 INFO Job: Job job\_1672817468245\_0057 running in uber mode : false

23/01/04 13:42:09 INFO Job: map 0% reduce 0%

23/01/04 13:42:15 INFO Job: map 100% reduce 0%  
23/01/04 13:42:19 INFO Job: map 100% reduce 100%  
23/01/04 13:42:20 INFO Job: Job job\_1672817468245\_0057 completed successfully  
23/01/04 13:42:20 INFO Job: Counters: 55

#### File System Counters

FILE: Number of bytes read=2783707  
FILE: Number of bytes written=7427107  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=26389380  
HDFS: Number of bytes written=955129  
HDFS: Number of read operations=29  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=6  
Launched reduce tasks=1  
Data-local map tasks=3  
Rack-local map tasks=3  
Total time spent by all maps in occupied slots (ms)=103960  
Total time spent by all reduces in occupied slots (ms)=9715  
Total time spent by all map tasks (ms)=20792  
Total time spent by all reduce tasks (ms)=1943  
Total vcore-milliseconds taken by all map tasks=20792  
Total vcore-milliseconds taken by all reduce tasks=1943  
Total megabyte-milliseconds taken by all map tasks=106455040  
Total megabyte-milliseconds taken by all reduce tasks=9948160

#### Map-Reduce Framework

Map input records=28  
Map output records=318679  
Map output bytes=5204855  
Map output materialized bytes=2783737  
Input split bytes=1074  
Combine input records=318679  
Combine output records=149762  
Reduce input groups=57492  
Reduce shuffle bytes=2783737  
Reduce input records=149762  
Reduce output records=38756  
Spilled Records=299524  
Shuffled Maps =6  
Failed Shuffles=0  
Merged Map outputs=6  
GC time elapsed (ms)=238  
CPU time spent (ms)=20230  
Physical memory (bytes) snapshot=3195654144  
Virtual memory (bytes) snapshot=44722348032  
Total committed heap usage (bytes)=7369392128  
Peak Map Physical memory (bytes)=474165248  
Peak Map Virtual memory (bytes)=6391619584  
Peak Reduce Physical memory (bytes)=394817536  
Peak Reduce Virtual memory (bytes)=6412750848

#### 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=26388306  
File Output Format Counters  
  Bytes Written=955129  
23/01/04 13:42:20 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 13:42:20 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0058  
23/01/04 13:42:20 INFO FileInputFormat: Total input files to process : 6  
23/01/04 13:42:20 INFO JobSubmitter: number of splits:6  
23/01/04 13:42:20 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0058  
23/01/04 13:42:20 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:42:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0058  
23/01/04 13:42:20 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0058/](http://lena-master:8088/proxy/application_1672817468245_0058/)  
23/01/04 13:42:20 INFO Job: Running job: job\_1672817468245\_0058  
23/01/04 13:42:27 INFO Job: Job job\_1672817468245\_0058 running in uber mode : false  
23/01/04 13:42:27 INFO Job: map 0% reduce 0%  
23/01/04 13:42:32 INFO Job: map 100% reduce 0%  
23/01/04 13:42:40 INFO Job: map 100% reduce 100%  
23/01/04 13:42:40 INFO Job: Job job\_1672817468245\_0058 completed successfully  
23/01/04 13:42:40 INFO Job: Counters: 55

#### File System Counters

FILE: Number of bytes read=26387254  
FILE: Number of bytes written=54646367  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=27189444  
HDFS: Number of bytes written=3028543  
HDFS: Number of read operations=31  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=6  
Launched reduce tasks=1  
Data-local map tasks=3  
Rack-local map tasks=3  
Total time spent by all maps in occupied slots (ms)=102140  
Total time spent by all reduces in occupied slots (ms)=23525  
Total time spent by all map tasks (ms)=20428  
Total time spent by all reduce tasks (ms)=4705  
Total vcore-milliseconds taken by all map tasks=20428  
Total vcore-milliseconds taken by all reduce tasks=4705  
Total megabyte-milliseconds taken by all map tasks=104591360  
Total megabyte-milliseconds taken by all reduce tasks=24089600

#### Map-Reduce Framework

Map input records=28  
Map output records=28  
Map output bytes=26387108  
Map output materialized bytes=26387284  
Input split bytes=1074  
Combine input records=0  
Combine output records=0

Reduce input groups=27  
Reduce shuffle bytes=26387284  
Reduce input records=28  
Reduce output records=27  
Spilled Records=56  
Shuffled Maps =6  
Failed Shuffles=0  
Merged Map outputs=6  
GC time elapsed (ms)=619  
CPU time spent (ms)=21360  
Physical memory (bytes) snapshot=3719184384  
Virtual memory (bytes) snapshot=44627128320  
Total committed heap usage (bytes)=7909408768  
Peak Map Physical memory (bytes)=514498560  
Peak Map Virtual memory (bytes)=6379859968  
Peak Reduce Physical memory (bytes)=797908992  
Peak Reduce Virtual memory (bytes)=6387535872

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=26388306

File Output Format Counters

Bytes Written=3028543

23/01/04 13:42:40 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 13:42:40 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0059  
23/01/04 13:42:40 INFO FileInputFormat: Total input files to process : 1  
23/01/04 13:42:40 INFO JobSubmitter: number of splits:1  
23/01/04 13:42:40 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0059  
23/01/04 13:42:40 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:42:40 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0059  
23/01/04 13:42:40 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0059/](http://lena-master:8088/proxy/application_1672817468245_0059/)  
23/01/04 13:42:40 INFO Job: Running job: job\_1672817468245\_0059  
23/01/04 13:42:46 INFO Job: Job job\_1672817468245\_0059 running in uber mode : false  
23/01/04 13:42:46 INFO Job: map 0% reduce 0%  
23/01/04 13:42:52 INFO Job: map 100% reduce 0%  
23/01/04 13:42:58 INFO Job: map 100% reduce 100%  
23/01/04 13:42:58 INFO Job: Job job\_1672817468245\_0059 completed successfully  
23/01/04 13:42:58 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3028037  
FILE: Number of bytes written=6588041  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3028720  
HDFS: Number of bytes written=3028543  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16570  
Total time spent by all reduces in occupied slots (ms)=16550  
Total time spent by all map tasks (ms)=3314  
Total time spent by all reduce tasks (ms)=3310  
Total vcore-milliseconds taken by all map tasks=3314  
Total vcore-milliseconds taken by all reduce tasks=3310  
Total megabyte-milliseconds taken by all map tasks=16967680  
Total megabyte-milliseconds taken by all reduce tasks=16947200

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=3027897  
Map output materialized bytes=3028037  
Input split bytes=177  
Combine input records=0  
Combine output records=0  
Reduce input groups=27  
Reduce shuffle bytes=3028037  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=2700  
Physical memory (bytes) snapshot=676319232  
Virtual memory (bytes) snapshot=12761415680  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=382726144  
Peak Map Virtual memory (bytes)=6385655808  
Peak Reduce Physical memory (bytes)=297013248  
Peak Reduce Virtual memory (bytes)=6379667456

#### 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=3028543

#### File Output Format Counters

Bytes Written=3028543

23/01/04 13:42:58 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/partial-vectors-0

23/01/04 13:42:58 INFO SparseVectorsFromSequenceFiles: Calculating IDF

23/01/04 13:42:58 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/df-count

23/01/04 13:42:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:42:58 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0060

23/01/04 13:42:58 INFO FileInputFormat: Total input files to process : 1

23/01/04 13:42:59 INFO JobSubmitter: number of splits:1  
23/01/04 13:42:59 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0060  
23/01/04 13:42:59 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:42:59 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0060  
23/01/04 13:42:59 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0060/  
23/01/04 13:42:59 INFO Job: Running job: job\_1672817468245\_0060  
23/01/04 13:43:05 INFO Job: Job job\_1672817468245\_0060 running in uber mode : false  
23/01/04 13:43:05 INFO Job: map 0% reduce 0%  
23/01/04 13:43:12 INFO Job: map 100% reduce 0%  
23/01/04 13:43:19 INFO Job: map 100% reduce 100%  
23/01/04 13:43:19 INFO Job: Job job\_1672817468245\_0060 completed successfully  
23/01/04 13:43:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=542604  
FILE: Number of bytes written=1616297  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3028721  
HDFS: Number of bytes written=775373  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17825  
Total time spent by all reduces in occupied slots (ms)=16785  
Total time spent by all map tasks (ms)=3565  
Total time spent by all reduce tasks (ms)=3357  
Total vcore-milliseconds taken by all map tasks=3565  
Total vcore-milliseconds taken by all reduce tasks=3357  
Total megabyte-milliseconds taken by all map tasks=18252800  
Total megabyte-milliseconds taken by all reduce tasks=17187840

#### Map-Reduce Framework

Map input records=27  
Map output records=286077  
Map output bytes=3432924  
Map output materialized bytes=542604  
Input split bytes=178  
Combine input records=286077  
Combine output records=38757  
Reduce input groups=38757  
Reduce shuffle bytes=542604  
Reduce input records=38757  
Reduce output records=38757  
Spilled Records=77514  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=4250  
Physical memory (bytes) snapshot=721539072  
Virtual memory (bytes) snapshot=12773052416



Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396406784  
Peak Map Virtual memory (bytes)=6376562688  
Peak Reduce Physical memory (bytes)=325132288  
Peak Reduce Virtual memory (bytes)=6396489728

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=3028543

File Output Format Counters

Bytes Written=775373

23/01/04 13:43:19 INFO SparseVectorsFromSequenceFiles: Pruning  
23/01/04 13:43:19 INFO deprecation: mapred.input.dir is deprecated. Instead, use  
mapreduce.input.fileinputformat.inputdir  
23/01/04 13:43:19 INFO deprecation: mapred.compress.map.output is deprecated.  
Instead, use mapreduce.map.output.compress  
23/01/04 13:43:19 INFO deprecation: mapred.output.dir is deprecated. Instead, use  
mapreduce.output.fileoutputformat.outputdir  
23/01/04 13:43:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 13:43:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0061  
23/01/04 13:43:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 13:43:19 INFO JobSubmitter: number of splits:1  
23/01/04 13:43:20 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0061  
23/01/04 13:43:20 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:43:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0061  
23/01/04 13:43:20 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0061/](http://lena-master:8088/proxy/application_1672817468245_0061/)  
23/01/04 13:43:20 INFO Job: Running job: job\_1672817468245\_0061  
23/01/04 13:43:26 INFO Job: Job job\_1672817468245\_0061 running in uber mode : false  
23/01/04 13:43:26 INFO Job: map 0% reduce 0%  
23/01/04 13:43:31 INFO Job: map 100% reduce 0%  
23/01/04 13:43:37 INFO Job: map 100% reduce 100%  
23/01/04 13:43:38 INFO Job: Job job\_1672817468245\_0061 completed successfully  
23/01/04 13:43:38 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=1799950  
FILE: Number of bytes written=2582533  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3028721  
HDFS: Number of bytes written=2680027  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=16355  
Total time spent by all reduces in occupied slots (ms)=17590  
Total time spent by all map tasks (ms)=3271  
Total time spent by all reduce tasks (ms)=3518  
Total vcore-milliseconds taken by all map tasks=3271  
Total vcore-milliseconds taken by all reduce tasks=3518  
Total megabyte-milliseconds taken by all map tasks=16747520  
Total megabyte-milliseconds taken by all reduce tasks=18012160

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=3027897  
Map output materialized bytes=1024589  
Input split bytes=178  
Combine input records=0  
Combine output records=0  
Reduce input groups=27  
Reduce shuffle bytes=1024589  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=5070  
Physical memory (bytes) snapshot=864075776  
Virtual memory (bytes) snapshot=12903006208  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=384282624  
Peak Map Virtual memory (bytes)=6375542784  
Peak Reduce Physical memory (bytes)=479793152  
Peak Reduce Virtual memory (bytes)=6527463424

#### 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=3028543

#### File Output Format Counters

Bytes Written=2680027

23/01/04 13:43:38 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors

23/01/04 13:43:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:43:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0062

23/01/04 13:43:38 INFO FileInputFormat: Total input files to process : 1

23/01/04 13:43:38 INFO JobSubmitter: number of splits:1

23/01/04 13:43:38 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0062

23/01/04 13:43:38 INFO JobSubmitter: Executing with tokens: []

23/01/04 13:43:38 INFO YarnClientImpl: Submitted application application\_1672817468245\_0062

23/01/04 13:43:38 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0062/](http://lena-master:8088/proxy/application_1672817468245_0062/)

23/01/04 13:43:38 INFO Job: Running job: job\_1672817468245\_0062  
23/01/04 13:43:44 INFO Job: Job job\_1672817468245\_0062 running in uber mode : false  
23/01/04 13:43:44 INFO Job: map 0% reduce 0%  
23/01/04 13:43:50 INFO Job: map 100% reduce 0%  
23/01/04 13:43:54 INFO Job: map 100% reduce 100%  
23/01/04 13:43:54 INFO Job: Job job\_1672817468245\_0062 completed successfully  
23/01/04 13:43:55 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2679517  
FILE: Number of bytes written=5890131  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2680215  
HDFS: Number of bytes written=2680027  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16290  
Total time spent by all reduces in occupied slots (ms)=9170  
Total time spent by all map tasks (ms)=3258  
Total time spent by all reduce tasks (ms)=1834  
Total vcore-milliseconds taken by all map tasks=3258  
Total vcore-milliseconds taken by all reduce tasks=1834  
Total megabyte-milliseconds taken by all map tasks=16680960  
Total megabyte-milliseconds taken by all reduce tasks=9390080

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2679381  
Map output materialized bytes=2679517  
Input split bytes=188  
Combine input records=0  
Combine output records=0  
Reduce input groups=27  
Reduce shuffle bytes=2679517  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=2670  
Physical memory (bytes) snapshot=691691520  
Virtual memory (bytes) snapshot=12777390080  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=381575168  
Peak Map Virtual memory (bytes)=6382010368  
Peak Reduce Physical memory (bytes)=310116352  
Peak Reduce Virtual memory (bytes)=6395379712

#### 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=2680027
File Output Format Counters
    Bytes Written=2680027
23/01/04 13:43:55 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors-partial
23/01/04 13:43:55 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors-toprune
23/01/04 13:43:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:43:55 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0063
23/01/04 13:43:55 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:43:55 INFO JobSubmitter: number of splits:1
23/01/04 13:43:55 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0063
23/01/04 13:43:55 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:43:55 INFO YarnClientImpl: Submitted application
application_1672817468245_0063
23/01/04 13:43:55 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0063/
23/01/04 13:43:55 INFO Job: Running job: job_1672817468245_0063
23/01/04 13:44:01 INFO Job: Job job_1672817468245_0063 running in uber mode : false
23/01/04 13:44:01 INFO Job:  map 0% reduce 0%
23/01/04 13:44:06 INFO Job:  map 100% reduce 0%
23/01/04 13:44:12 INFO Job:  map 100% reduce 100%
23/01/04 13:44:12 INFO Job: Job job_1672817468245_0063 completed successfully
23/01/04 13:44:12 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=2679517
    FILE: Number of bytes written=5893509
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=3455550
    HDFS: Number of bytes written=2680027
    HDFS: Number of read operations=11
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=16340
    Total time spent by all reduces in occupied slots (ms)=17485
    Total time spent by all map tasks (ms)=3268
    Total time spent by all reduce tasks (ms)=3497
    Total vcore-milliseconds taken by all map tasks=3268
    Total vcore-milliseconds taken by all reduce tasks=3497
    Total megabyte-milliseconds taken by all map tasks=16732160
    Total megabyte-milliseconds taken by all reduce tasks=17904640
Map-Reduce Framework
    Map input records=27
    Map output records=27

```

Map output bytes=2679381  
Map output materialized bytes=2679517  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=27  
Reduce shuffle bytes=2679517  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=3390  
Physical memory (bytes) snapshot=717561856  
Virtual memory (bytes) snapshot=12781117440  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=385015808  
Peak Map Virtual memory (bytes)=6382927872  
Peak Reduce Physical memory (bytes)=332546048  
Peak Reduce Virtual memory (bytes)=6402097152

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=2680027

File Output Format Counters

Bytes Written=2680027

23/01/04 13:44:12 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/tfidf-vectors

23/01/04 13:44:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:44:12 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0064

23/01/04 13:44:12 INFO FileInputFormat: Total input files to process : 1

23/01/04 13:44:12 INFO JobSubmitter: number of splits:1

23/01/04 13:44:13 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0064

23/01/04 13:44:13 INFO JobSubmitter: Executing with tokens: []

23/01/04 13:44:13 INFO YarnClientImpl: Submitted application application\_1672817468245\_0064

23/01/04 13:44:13 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0064/

23/01/04 13:44:13 INFO Job: Running job: job\_1672817468245\_0064

23/01/04 13:44:19 INFO Job: Job job\_1672817468245\_0064 running in uber mode : false

23/01/04 13:44:19 INFO Job: map 0% reduce 0%

23/01/04 13:44:24 INFO Job: map 100% reduce 0%

23/01/04 13:44:30 INFO Job: map 100% reduce 100%

23/01/04 13:44:31 INFO Job: Job job\_1672817468245\_0064 completed successfully

23/01/04 13:44:31 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2679517  
FILE: Number of bytes written=5890991  
FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2680204  
HDFS: Number of bytes written=2680027  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16000  
Total time spent by all reduces in occupied slots (ms)=16625  
Total time spent by all map tasks (ms)=3200  
Total time spent by all reduce tasks (ms)=3325  
Total vcore-milliseconds taken by all map tasks=3200  
Total vcore-milliseconds taken by all reduce tasks=3325  
Total megabyte-milliseconds taken by all map tasks=16384000  
Total megabyte-milliseconds taken by all reduce tasks=17024000

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2679381  
Map output materialized bytes=2679517  
Input split bytes=177  
Combine input records=0  
Combine output records=0  
Reduce input groups=27  
Reduce shuffle bytes=2679517  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=35  
CPU time spent (ms)=2840  
Physical memory (bytes) snapshot=685490176  
Virtual memory (bytes) snapshot=12761649152  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=374620160  
Peak Map Virtual memory (bytes)=6382657536  
Peak Reduce Physical memory (bytes)=310870016  
Peak Reduce Virtual memory (bytes)=6380949504

#### 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=2680027

#### File Output Format Counters

Bytes Written=2680027

23/01/04 13:44:31 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-vectors/partial-vectors-0

23/01/04 13:44:31 INFO MahoutDriver: Program took 163153 ms (Minutes:

```

2.7192166666666666)
bfezo001@lena:~$ # list the generated output in HDFS
bfezo001@lena:~$ hadoop fs -ls dsm010/question_2/british-fiction-corpus-vectors/tf-vectors
Found 2 items
-rw-r--r--    3 bfezo001 users          0 2023-01-04 13:43
dsm010/question_2/british-fiction-corpus-vectors/tf-vectors/_SUCCESS
-rw-r--r--    3 bfezo001 users    2680027 2023-01-04 13:43
dsm010/question_2/british-fiction-corpus-vectors/tf-vectors/part-r-00000
bfezo001@lena:~$ hadoop fs -ls
dsm010/question_2/british-fiction-corpus-vectors/tfidf-vectors
Found 2 items
-rw-r--r--    3 bfezo001 users          0 2023-01-04 13:44
dsm010/question_2/british-fiction-corpus-vectors/tfidf-vectors/_SUCCESS
-rw-r--r--    3 bfezo001 users    2680027 2023-01-04 13:44
dsm010/question_2/british-fiction-corpus-vectors/tfidf-vectors/part-r-00000
bfezo001@lena:~$
bfezo001@lena:~$ # we can generate "canopies", which will help approximate
centroids for K-Means
bfezo001@lena:~$ mahout canopy \
> -i dsm010/question_2/british-fiction-corpus-vectors/tf-vectors \
> -ow \
> -o dsm010/question_2/british-fiction-corpus-canopy-centroids \
> -dm org.apache.mahout.common.distance.CosineDistanceMeasure \
> -t1 0.5 \
> -t2 0.3
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT_JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 13:44:38 INFO AbstractJob: Command line arguments: {--
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
method=[mapreduce], --output=[dsm010/question_2/british-fiction-corpus-canopy-
centroids], --overwrite=null, --startPhase=[0], --t1=[0.5], --t2=[0.3], --
tempDir=[temp]}
23/01/04 13:44:39 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids
23/01/04 13:44:39 INFO CanopyDriver: Build Clusters Input:
dsm010/question_2/british-fiction-corpus-vectors/tf-vectors Out:
dsm010/question_2/british-fiction-corpus-canopy-centroids Measure:
org.apache.mahout.common.distance.CosineDistanceMeasure@4e877ad5 t1: 0.5 t2: 0.3
23/01/04 13:44:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:44:39 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0065
23/01/04 13:44:39 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:44:39 INFO JobSubmitter: number of splits:1
23/01/04 13:44:40 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0065
23/01/04 13:44:40 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:44:40 INFO Configuration: resource-types.xml not found
23/01/04 13:44:40 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 13:44:40 INFO YarnClientImpl: Submitted application
application_1672817468245_0065
23/01/04 13:44:40 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0065/
23/01/04 13:44:40 INFO Job: Running job: job_1672817468245_0065

```

23/01/04 13:44:46 INFO Job: Job job\_1672817468245\_0065 running in uber mode : false  
23/01/04 13:44:46 INFO Job: map 0% reduce 0%  
23/01/04 13:44:52 INFO Job: map 100% reduce 0%  
23/01/04 13:44:58 INFO Job: map 100% reduce 100%  
23/01/04 13:44:58 INFO Job: Job job\_1672817468245\_0065 completed successfully  
23/01/04 13:44:58 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2749672  
FILE: Number of bytes written=6033059  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2680197  
HDFS: Number of bytes written=2763394  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17760  
Total time spent by all reduces in occupied slots (ms)=18200  
Total time spent by all map tasks (ms)=3552  
Total time spent by all reduce tasks (ms)=3640  
Total vcore-milliseconds taken by all map tasks=3552  
Total vcore-milliseconds taken by all reduce tasks=3640  
Total megabyte-milliseconds taken by all map tasks=18186240  
Total megabyte-milliseconds taken by all reduce tasks=18636800

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2749536  
Map output materialized bytes=2749672  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=2749672  
Reduce input records=27  
Reduce output records=26  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=759566336  
Virtual memory (bytes) snapshot=12759130112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=439107584  
Peak Map Virtual memory (bytes)=6373928960  
Peak Reduce Physical memory (bytes)=320458752  
Peak Reduce Virtual memory (bytes)=6385201152

#### 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=2680027
    File Output Format Counters
        Bytes Written=2763394
23/01/04 13:44:58 INFO MahoutDriver: Program took 19992 ms (Minutes: 0.3332)
bfezo001@lena:~$
bfezo001@lena:~$ #t1 is known as loose distance and t2 is known as tight distance
and t1 > t2. t1 and t2 values must be compatible with the selected distance
measure. Let's have a look at what was generated:
bfezo001@lena:~$ hadoop fs -ls dsm010/question_2/british-fiction-corpus-canopy-
centroids/
Found 1 items
drwxr-xr-x   - bfezo001 users          0 2023-01-04 13:44
dsm010/question_2/british-fiction-corpus-canopy-centroids/clusters-0-final
bfezo001@lena:~$ mahout kmeans \
>   -i dsm010/question_2/british-fiction-corpus-vectors/tf-vectors \
>   -c dsm010/question_2/british-fiction-corpus-canopy-centroids \
>   -o hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans-
clusters \
>   -dm org.apache.mahout.common.distance.CosineDistanceMeasure \
>   -cl -cd 0.1 -ow -x 20 \
>   -k 10
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 13:49:50 INFO AbstractJob: Command line arguments: [--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids], --
convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[10],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans-clusters],--overwrite=null, --startPhase=[0], --tempDir=[temp]]
23/01/04 13:49:50 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans-clusters
23/01/04 13:49:50 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids
23/01/04 13:49:50 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 13:49:50 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 13:49:51 INFO RandomSeedGenerator: Wrote 10 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/part-randomSeed
23/01/04 13:49:51 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans-clusters
23/01/04 13:49:51 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 13:49:51 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 13:49:51 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:803223/01/04 13:49:51 INFO
JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0066
23/01/04 13:49:51 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:49:52 INFO JobSubmitter: number of splits:123/01/04 13:49:52 INFO

```

JobSubmitter: Submitting tokens for job: job\_1672817468245\_0066  
23/01/04 13:49:52 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:49:52 INFO Configuration: resource-types.xml not found  
23/01/04 13:49:52 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 13:49:52 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0066  
23/01/04 13:49:52 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0066/  
23/01/04 13:49:52 INFO Job: Running job: job\_1672817468245\_0066  
23/01/04 13:49:58 INFO Job: Job job\_1672817468245\_0066 running in uber mode : false  
23/01/04 13:49:58 INFO Job: map 0% reduce 0%  
23/01/04 13:50:03 INFO Job: map 100% reduce 0%  
23/01/04 13:50:09 INFO Job: map 100% reduce 100%  
23/01/04 13:50:09 INFO Job: Job job\_1672817468245\_0066 completed successfully  
23/01/04 13:50:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3889653  
FILE: Number of bytes written=8311483  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8498915  
HDFS: Number of bytes written=2435188  
HDFS: Number of read operations=55  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18305  
Total time spent by all reduces in occupied slots (ms)=18200  
Total time spent by all map tasks (ms)=3661  
Total time spent by all reduce tasks (ms)=3640  
Total vcore-milliseconds taken by all map tasks=3661  
Total vcore-milliseconds taken by all reduce tasks=3640  
Total megabyte-milliseconds taken by all map tasks=18744320  
Total megabyte-milliseconds taken by all reduce tasks=18636800

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=3889597  
Map output materialized bytes=3889653  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=3889653  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=74  
CPU time spent (ms)=4610  
Physical memory (bytes) snapshot=714534912  
Virtual memory (bytes) snapshot=12767817728

```

Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=411193344
Peak Map Virtual memory (bytes)=6381215744
Peak Reduce Physical memory (bytes)=303341568
Peak Reduce Virtual memory (bytes)=6386601984
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=2680027
File Output Format Counters
  Bytes Written=2435188
23/01/04 13:50:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:50:09 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0067
23/01/04 13:50:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:50:09 INFO JobSubmitter: number of splits:1
23/01/04 13:50:10 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0067
23/01/04 13:50:10 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:50:10 INFO YarnClientImpl: Submitted application
application_1672817468245_0067
23/01/04 13:50:10 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0067/
23/01/04 13:50:10 INFO Job: Running job: job_1672817468245_0067
23/01/04 13:50:16 INFO Job: Job job_1672817468245_0067 running in uber mode : false
23/01/04 13:50:16 INFO Job:  map 0% reduce 0%
23/01/04 13:50:22 INFO Job:  map 100% reduce 0%
23/01/04 13:50:27 INFO Job:  map 100% reduce 100%
23/01/04 13:50:28 INFO Job: Job job_1672817468245_0067 completed successfully
23/01/04 13:50:28 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=5353575
  FILE: Number of bytes written=11239327
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=7550961
  HDFS: Number of bytes written=2435188
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18400
  Total time spent by all reduces in occupied slots (ms)=17665
  Total time spent by all map tasks (ms)=3680
  Total time spent by all reduce tasks (ms)=3533
  Total vcore-milliseconds taken by all map tasks=3680
  Total vcore-milliseconds taken by all reduce tasks=3533
  Total megabyte-milliseconds taken by all map tasks=18841600

```

```

        Total megabyte-milliseconds taken by all reduce tasks=18088960
Map-Reduce Framework
    Map input records=27
    Map output records=10
    Map output bytes=5353519
    Map output materialized bytes=5353575
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=10
    Reduce shuffle bytes=5353575
    Reduce input records=10
    Reduce output records=10
    Spilled Records=20
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=97
    CPU time spent (ms)=4240
    Physical memory (bytes) snapshot=728494080
    Virtual memory (bytes) snapshot=12766670848
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=418775040
    Peak Map Virtual memory (bytes)=6377070592
    Peak Reduce Physical memory (bytes)=309719040
    Peak Reduce Virtual memory (bytes)=6389600256
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=2680027
File Output Format Counters
    Bytes Written=2435188
23/01/04 13:50:28 INFO KMeansDriver: Clustering data
23/01/04 13:50:28 INFO KMeansDriver: Running Clustering
23/01/04 13:50:28 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans-clusters
Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans-
clusters
23/01/04 13:50:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:50:28 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0068
23/01/04 13:50:28 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:50:28 INFO JobSubmitter: number of splits:1
23/01/04 13:50:28 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0068
23/01/04 13:50:28 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:50:28 INFO YarnClientImpl: Submitted application
application_1672817468245_0068
23/01/04 13:50:28 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0068/
23/01/04 13:50:28 INFO Job: Running job: job_1672817468245_0068
23/01/04 13:50:34 INFO Job: Job job_1672817468245_0068 running in uber mode : false

```

23/01/04 13:50:34 INFO Job: map 0% reduce 0%  
23/01/04 13:50:39 INFO Job: map 100% reduce 0%  
23/01/04 13:50:40 INFO Job: Job job\_1672817468245\_0068 completed successfully  
23/01/04 13:50:41 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265840  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5115579  
HDFS: Number of bytes written=2680688  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17765  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3553  
Total vcore-milliseconds taken by all map tasks=3553  
Total megabyte-milliseconds taken by all map tasks=18191360

Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=39  
CPU time spent (ms)=2110  
Physical memory (bytes) snapshot=294510592  
Virtual memory (bytes) snapshot=6372814848  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=294510592  
Peak Map Virtual memory (bytes)=6372814848

File Input Format Counters

Bytes Read=2680027

File Output Format Counters

Bytes Written=2680688

23/01/04 13:50:41 INFO MahoutDriver: Program took 50935 ms (Minutes:  
0.8489166666666667)

bfezo001@lena:~\$ hadoop fs -ls dsm010/question\_2/british-fiction-corpus-kmeans-  
clusters

Found 5 items

```
-rw-r--r--  3 bfezo001 users          194 2023-01-04 13:50
dsm010/question_2/british-fiction-corpus-kmeans-clusters/_policy
drwxr-xr-x  - bfezo001 users           0 2023-01-04 13:50
dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusteredPoints
drwxr-xr-x  - bfezo001 users           0 2023-01-04 13:49
dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusters-0
drwxr-xr-x  - bfezo001 users           0 2023-01-04 13:50
dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusters-1
drwxr-xr-x  - bfezo001 users           0 2023-01-04 13:50
dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusters-2-final
bfezo001@lena:~$ mahout clusterdump \
> -dt sequencefile \
```

```

> -d dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-* \
> -i dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusters-2-final
\> -o ~/dsm010/question_2/output/british-fiction-corpus-clusters.txt -b 100 \
> -p dsm010/question_2/british-fiction-corpus-kmeans-clusters/clusteredPoints \
> -n 20 --evaluate
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.Running on hadoop,
using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 13:54:27 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.SquaredEuclideanDistanceMeasure]
, --endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-
fiction-corpus-kmeans-clusters/clusters-2-final], --numWords=[20],
--output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-
clusters.txt], --outputFormat=[TEXT], --pointsDir=[dsm010/question_2/british-
fiction-corpus-kmeans-clusters/clusteredPoints], --startPhase=[0], --
substring=[100], --tempDir=[temp]}
23/01/04 13:54:28 INFO HadoopUtil: Deleting tmp/representative
23/01/04 13:54:28 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans-clusters/
clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.SquaredEuclideanDistanceMeasure]
, --endPhase=[2147483647], --input=[dsm010/question_2/british-fiction-corpus-
kmeans-clusters/clusters-2-final], --maxIter=[5], --method=[mapreduce], --
output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 13:54:28 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 13:54:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 13:54:28 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0072
23/01/04 13:54:29 INFO FileInputFormat: Total input files to process : 1
23/01/04 13:54:29 INFO JobSubmitter: number of splits:1
23/01/04 13:54:29 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0072
23/01/04 13:54:29 INFO JobSubmitter: Executing with tokens: []
23/01/04 13:54:29 INFO Configuration: resource-types.xml not found
23/01/04 13:54:29 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 13:54:29 INFO YarnClientImpl: Submitted application
application_1672817468245_0072
23/01/04 13:54:29 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0072/
23/01/04 13:54:29 INFO Job: Running job: job_1672817468245_0072
23/01/04 13:54:35 INFO Job: Job job_1672817468245_0072 running in uber mode : false
23/01/04 13:54:35 INFO Job: map 0% reduce 0%
23/01/04 13:54:41 INFO Job: map 100% reduce 0%
23/01/04 13:54:46 INFO Job: map 100% reduce 100%
23/01/04 13:54:46 INFO Job: Job job_1672817468245_0072 completed successfully
23/01/04 13:54:46 INFO Job: Counters: 54
File System Counters
FILE: Number of bytes read=1038972
FILE: Number of bytes written=2610467
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5600349
HDFS: Number of bytes written=2498776

```

HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16705  
Total time spent by all reduces in occupied slots (ms)=16160  
Total time spent by all map tasks (ms)=3341  
Total time spent by all reduce tasks (ms)=3232  
Total vcore-milliseconds taken by all map tasks=3341  
Total vcore-milliseconds taken by all reduce tasks=3232  
Total megabyte-milliseconds taken by all map tasks=17105920  
Total megabyte-milliseconds taken by all reduce tasks=16547840

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1038917  
Map output materialized bytes=1038972  
Input split bytes=183  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1038972  
Reduce input records=10  
Reduce output records=20  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=3040  
Physical memory (bytes) snapshot=705720320  
Virtual memory (bytes) snapshot=12752510976  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404406272  
Peak Map Virtual memory (bytes)=6375079936  
Peak Reduce Physical memory (bytes)=301314048  
Peak Reduce Virtual memory (bytes)=6381338624

#### 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=2680688

#### File Output Format Counters

Bytes Written=2498776

23/01/04 13:54:46 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 13:54:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:54:46 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0073

23/01/04 13:54:47 INFO FileInputFormat: Total input files to process : 1

23/01/04 13:54:47 INFO JobSubmitter: number of splits:1  
23/01/04 13:54:47 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0073  
23/01/04 13:54:47 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:54:47 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0073  
23/01/04 13:54:47 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0073/  
23/01/04 13:54:47 INFO Job: Running job: job\_1672817468245\_0073  
23/01/04 13:54:53 INFO Job: Job job\_1672817468245\_0073 running in uber mode : false  
23/01/04 13:54:53 INFO Job: map 0% reduce 0%  
23/01/04 13:54:59 INFO Job: map 100% reduce 0%  
23/01/04 13:55:05 INFO Job: map 100% reduce 100%  
23/01/04 13:55:05 INFO Job: Job job\_1672817468245\_0073 completed successfully  
23/01/04 13:55:05 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=955339  
FILE: Number of bytes written=2443201  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7678423  
HDFS: Number of bytes written=3454201  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16960  
Total time spent by all reduces in occupied slots (ms)=16945  
Total time spent by all map tasks (ms)=3392  
Total time spent by all reduce tasks (ms)=3389  
Total vcore-milliseconds taken by all map tasks=3392  
Total vcore-milliseconds taken by all reduce tasks=3389  
Total megabyte-milliseconds taken by all map tasks=17367040  
Total megabyte-milliseconds taken by all reduce tasks=17351680

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=955285  
Map output materialized bytes=955339  
Input split bytes=183  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=955339  
Reduce input records=10  
Reduce output records=30  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=3320  
Physical memory (bytes) snapshot=692494336  
Virtual memory (bytes) snapshot=12756623360



Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397578240  
Peak Map Virtual memory (bytes)=6378643456  
Peak Reduce Physical memory (bytes)=294916096  
Peak Reduce Virtual memory (bytes)=6377979904

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=2680688

File Output Format Counters

Bytes Written=3454201

23/01/04 13:55:05 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 13:55:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 13:55:05 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0074

23/01/04 13:55:05 INFO FileInputFormat: Total input files to process : 1

23/01/04 13:55:05 INFO JobSubmitter: number of splits:1

23/01/04 13:55:05 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0074

23/01/04 13:55:05 INFO JobSubmitter: Executing with tokens: []

23/01/04 13:55:05 INFO YarnClientImpl: Submitted application application\_1672817468245\_0074

23/01/04 13:55:05 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0074/

23/01/04 13:55:05 INFO Job: Running job: job\_1672817468245\_0074

23/01/04 13:55:11 INFO Job: Job job\_1672817468245\_0074 running in uber mode : false

23/01/04 13:55:11 INFO Job: map 0% reduce 0%

23/01/04 13:55:17 INFO Job: map 100% reduce 0%

23/01/04 13:55:22 INFO Job: map 100% reduce 100%

23/01/04 13:55:23 INFO Job: Job job\_1672817468245\_0074 completed successfully

23/01/04 13:55:23 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=1027942  
FILE: Number of bytes written=2588407  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9589273  
HDFS: Number of bytes written=4482228  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17120  
Total time spent by all reduces in occupied slots (ms)=16865  
Total time spent by all map tasks (ms)=3424  
Total time spent by all reduce tasks (ms)=3373  
Total vcore-milliseconds taken by all map tasks=3424

Total vcore-milliseconds taken by all reduce tasks=3373  
Total megabyte-milliseconds taken by all map tasks=17530880  
Total megabyte-milliseconds taken by all reduce tasks=17269760  
Map-Reduce Framework  
Map input records=27  
Map output records=10  
Map output bytes=1027887  
Map output materialized bytes=1027942  
Input split bytes=183  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1027942  
Reduce input records=10  
Reduce output records=40  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3300  
Physical memory (bytes) snapshot=705605632  
Virtual memory (bytes) snapshot=12757508096  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412065792  
Peak Map Virtual memory (bytes)=6368460800  
Peak Reduce Physical memory (bytes)=297746432  
Peak Reduce Virtual memory (bytes)=6389047296  
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=2680688  
File Output Format Counters  
Bytes Written=4482228

23/01/04 13:55:23 INFO RepresentativePointsDriver: Representative Points Iteration 3  
23/01/04 13:55:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 13:55:23 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0075  
23/01/04 13:55:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 13:55:24 INFO JobSubmitter: number of splits:1  
23/01/04 13:55:24 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0075  
23/01/04 13:55:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:55:24 INFO YarnClientImpl: Submitted application application\_1672817468245\_0075  
23/01/04 13:55:24 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0075/  
23/01/04 13:55:24 INFO Job: Running job: job\_1672817468245\_0075  
23/01/04 13:55:30 INFO Job: Job job\_1672817468245\_0075 running in uber mode : false  
23/01/04 13:55:30 INFO Job: map 0% reduce 0%  
23/01/04 13:55:36 INFO Job: map 100% reduce 0%  
23/01/04 13:55:42 INFO Job: map 100% reduce 100%

23/01/04 13:55:42 INFO Job: Job job\_1672817468245\_0075 completed successfully

23/01/04 13:55:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=987529  
FILE: Number of bytes written=2507581  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11645327  
HDFS: Number of bytes written=5469843  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17435  
Total time spent by all reduces in occupied slots (ms)=16735  
Total time spent by all map tasks (ms)=3487  
Total time spent by all reduce tasks (ms)=3347  
Total vcore-milliseconds taken by all map tasks=3487  
Total vcore-milliseconds taken by all reduce tasks=3347  
Total megabyte-milliseconds taken by all map tasks=17853440  
Total megabyte-milliseconds taken by all reduce tasks=17136640

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=987475  
Map output materialized bytes=987529  
Input split bytes=183  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=987529  
Reduce input records=10  
Reduce output records=50  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3580  
Physical memory (bytes) snapshot=713781248  
Virtual memory (bytes) snapshot=12758589440  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=416813056  
Peak Map Virtual memory (bytes)=6382936064  
Peak Reduce Physical memory (bytes)=296968192  
Peak Reduce Virtual memory (bytes)=6375653376

#### 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=2680688  
File Output Format Counters  
Bytes Written=5469843  
23/01/04 13:55:42 INFO RepresentativePointsDriver: Representative Points Iteration  
4  
23/01/04 13:55:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 13:55:42 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0076  
23/01/04 13:55:42 INFO FileInputFormat: Total input files to process : 1  
23/01/04 13:55:42 INFO JobSubmitter: number of splits:1  
23/01/04 13:55:42 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0076  
23/01/04 13:55:42 INFO JobSubmitter: Executing with tokens: []  
23/01/04 13:55:42 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0076  
23/01/04 13:55:42 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0076/  
23/01/04 13:55:42 INFO Job: Running job: job\_1672817468245\_0076  
23/01/04 13:55:48 INFO Job: Job job\_1672817468245\_0076 running in uber mode : false  
23/01/04 13:55:48 INFO Job: map 0% reduce 0%  
23/01/04 13:55:54 INFO Job: map 100% reduce 0%  
23/01/04 13:55:59 INFO Job: map 100% reduce 100%  
23/01/04 13:56:00 INFO Job: Job job\_1672817468245\_0076 completed successfully  
23/01/04 13:56:00 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=919123  
FILE: Number of bytes written=2370769  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=13620557  
HDFS: Number of bytes written=6389052  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17655  
Total time spent by all reduces in occupied slots (ms)=17100  
Total time spent by all map tasks (ms)=3531  
Total time spent by all reduce tasks (ms)=3420  
Total vcore-milliseconds taken by all map tasks=3531  
Total vcore-milliseconds taken by all reduce tasks=3420  
Total megabyte-milliseconds taken by all map tasks=18078720  
Total megabyte-milliseconds taken by all reduce tasks=17510400

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=919069  
Map output materialized bytes=919123  
Input split bytes=183  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=919123

Reduce input records=10  
Reduce output records=60  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=3810  
Physical memory (bytes) snapshot=722026496  
Virtual memory (bytes) snapshot=12760588288  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413999104  
Peak Map Virtual memory (bytes)=6375100416  
Peak Reduce Physical memory (bytes)=308027392  
Peak Reduce Virtual memory (bytes)=6385487872

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=2680688

File Output Format Counters

Bytes Written=6389052

23/01/04 13:56:01 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.33727847188331084

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[23] =  
0.6481058078537116

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[11] = NaN

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[3] = NaN

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[10] =  
0.48333333333333334

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[6] =  
0.5004702682331529

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[26] =  
0.48333333333333334

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[18] =  
0.5712065174716331

23/01/04 13:56:01 INFO ClusterEvaluator: Intra-Cluster Density[9] = NaN

23/01/04 13:56:01 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.5372898520450329

23/01/04 13:56:02 INFO ClusterDumper: Wrote 10 clusters

23/01/04 13:56:02 INFO MahoutDriver: Program took 95230 ms (Minutes:  
1.5871666666666666)

bfezo001@lana:~\$

bfezo001@lana:~\$ # The final lines give a statistical evaluation of the clustering  
solution.

bfezo001@lana:~\$ tail -n 5 ~/dsm010/question\_2/output/british-fiction-corpus-  
clusters.txt

Inter-Cluster Density: 0.33727847188331084

Intra-Cluster Density: 0.5372898520450329

CDBw Inter-Cluster Density: 0.0

CDBw Intra-Cluster Density: 0.9600622344095872

CDBw Separation: 2.772156712372451E8

bfezo001@lana:~\$ k\_start=1

```

bfezo001@lana:~$ k_end=10
bfezo001@lana:~$
bfezo001@lana:~$ # vectors and dictionary path
bfezo001@lana:~$ path_vectors=dsm010/question_2/british-fiction-corpus-vectors/tf-
vectors
bfezo001@lana:~$
path_dictionary=dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*
bfezo001@lana:~$
bfezo001@lana:~$ # clusters output base path and result's local output path
bfezo001@lana:~$ path_hdfs_base=hdfs://lana/user/bfezo001/
bfezo001@lana:~$ path_clusters_base=dsm010/question_2/british-fiction-corpus-kmeans
bfezo001@lana:~$ path_results_base=~ /dsm010/question_2/output/british-fiction-
corpus-clusters
bfezo001@lana:~$
bfezo001@lana:~$ # distance metrics used Cosine, Euclidean, Manhattan
bfezo001@lana:~$
distance_metrics=("org.apache.mahout.common.distance.CosineDistanceMeasure"
"org.apache.mahout.common.distance.EuclideanDistanceMeasure"
"org.apache.mahout.common.distance.ManhattanDistanceMeasure")
bfezo001@lana:~$
bfezo001@lana:~$ for distance_metric in "${distance_metrics[@]}"
> do
>   echo "--- Distance Metric: $distance_metric"
>
>   # perform the canopy clustering
>   distance_name=${distance_metric##*.}
>   path_centroids=dsm010/question_2/british-fiction-corpus-canopy-centroids/$
{distance_name}
>
>   mahout canopy \
>     -i $path_vectors \
>     -ow \
>     -o $path_centroids \
>     -dm $distance_metric \
>     -t1 0.5 \
>     -t2 0.3
>
>   for ((k = $k_start; k <= $k_end; k++))
>   do
>
>     k_padded=$(printf %02d $k)
>
>     # set the output path for the clusters
>     path_clusters="${path_clusters_base}/${distance_name}/${k_padded}"
>
>     echo "---- K: $k_padded -- $path_clusters"
>
>     # run k-means clustering
>     mahout kmeans \
>       -i $path_vectors \
>       -c $path_centroids \
>       -o "${path_hdfs_base}${path_clusters}" \
>       -ow \
>       -dm $distance_metric \
>       -cl -cd 0.1 -ow -x 20 \
>       -k $k
>
>     # set the path for output
>     path_final_clusters=`hadoop fs -ls -d -C "${path_clusters}/clusters-* -final"`

```

```

> path_clusterpoints="${path_clusters}/clusteredPoints"
> path_results="${path_results_base}/${distance_name}/${k_padded}.txt"
>
> # results output
> mahout clusterdump -dt sequencefile \
>   -d $path_dictionary \
>   -i $path_final_clusters \
>   -o $path_results \
>   -of TEXT \
>   -b 100 \
>   -p $path_clusterpoints \
>   -dm $distance_metric \
>   -n 20 --evaluate
> done
> done
--- Distance Metric: org.apache.mahout.common.distance.CosineDistanceMeasure
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:28:34 INFO AbstractJob: Command line arguments: {--
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
method=[mapreduce], --output=[dsm010/question_2/british-fiction-corpus-canopy-
centroids/CosineDistanceMeasure], --overwrite=null, --startPhase=[0], --t1=[0.5],
--t2=[0.3], --tempDir=[temp]}
23/01/04 14:28:35 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:28:35 INFO CanopyDriver: Build Clusters Input:
dsm010/question_2/british-fiction-corpus-vectors/tf-vectors Out:
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure
Measure: org.apache.mahout.common.distance.CosineDistanceMeasure@4e877ad5 t1: 0.5
t2: 0.3
23/01/04 14:28:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:28:35 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0136
23/01/04 14:28:36 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:28:36 INFO JobSubmitter: number of splits:1
23/01/04 14:28:36 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0136
23/01/04 14:28:36 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:28:36 INFO Configuration: resource-types.xml not found
23/01/04 14:28:36 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:28:36 INFO YarnClientImpl: Submitted application
application_1672817468245_0136
23/01/04 14:28:36 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0136/
23/01/04 14:28:36 INFO Job: Running job: job_1672817468245_0136
23/01/04 14:28:42 INFO Job: Job job_1672817468245_0136 running in uber mode : false
23/01/04 14:28:42 INFO Job:  map 0% reduce 0%
23/01/04 14:28:48 INFO Job:  map 100% reduce 0%
23/01/04 14:28:54 INFO Job:  map 100% reduce 100%
23/01/04 14:28:54 INFO Job: Job job_1672817468245_0136 completed successfully
23/01/04 14:28:54 INFO Job: Counters: 54
File System Counters
FILE: Number of bytes read=2749672
FILE: Number of bytes written=6033103

```

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2680197  
HDFS: Number of bytes written=2763394  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18180  
Total time spent by all reduces in occupied slots (ms)=18465  
Total time spent by all map tasks (ms)=3636  
Total time spent by all reduce tasks (ms)=3693  
Total vcore-milliseconds taken by all map tasks=3636  
Total vcore-milliseconds taken by all reduce tasks=3693  
Total megabyte-milliseconds taken by all map tasks=18616320  
Total megabyte-milliseconds taken by all reduce tasks=18908160

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2749536  
Map output materialized bytes=2749672  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=2749672  
Reduce input records=27  
Reduce output records=26  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=55  
CPU time spent (ms)=4390  
Physical memory (bytes) snapshot=727330816  
Virtual memory (bytes) snapshot=12766138368  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=414490624  
Peak Map Virtual memory (bytes)=6377193472  
Peak Reduce Physical memory (bytes)=312840192  
Peak Reduce Virtual memory (bytes)=6388944896

#### 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=2680027

#### File Output Format Counters

Bytes Written=2763394

23/01/04 14:28:54 INFO MahoutDriver: Program took 19855 ms (Minutes:  
0.3309166666666667)



```

---- K: 01 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/01
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:28:59 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[1],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/01], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:28:59 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/01
23/01/04 14:28:59 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:28:59 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:28:59 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:28:59 INFO RandomSeedGenerator: Wrote 1 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:28:59 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/01
23/01/04 14:28:59 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:28:59 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:29:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:29:00 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0138
23/01/04 14:29:00 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:29:00 INFO JobSubmitter: number of splits:1
23/01/04 14:29:00 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0138
23/01/04 14:29:00 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:29:00 INFO Configuration: resource-types.xml not found
23/01/04 14:29:00 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:29:00 INFO YarnClientImpl: Submitted application
application_1672817468245_0138
23/01/04 14:29:00 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0138/
23/01/04 14:29:00 INFO Job: Running job: job_1672817468245_0138
23/01/04 14:29:06 INFO Job: Job job_1672817468245_0138 running in uber mode : false
23/01/04 14:29:06 INFO Job: map 0% reduce 0%
23/01/04 14:29:13 INFO Job: map 100% reduce 0%
23/01/04 14:29:19 INFO Job: map 100% reduce 100%
23/01/04 14:29:19 INFO Job: Job job_1672817468245_0138 completed successfully
23/01/04 14:29:19 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=887008
        FILE: Number of bytes written=2306289

```

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3239077  
HDFS: Number of bytes written=794072  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17295  
Total time spent by all reduces in occupied slots (ms)=17400  
Total time spent by all map tasks (ms)=3459  
Total time spent by all reduce tasks (ms)=3480  
Total vcore-milliseconds taken by all map tasks=3459  
Total vcore-milliseconds taken by all reduce tasks=3480  
Total megabyte-milliseconds taken by all map tasks=17710080  
Total megabyte-milliseconds taken by all reduce tasks=17817600

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=886997  
Map output materialized bytes=887008  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=887008  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=50  
CPU time spent (ms)=3550  
Physical memory (bytes) snapshot=727683072  
Virtual memory (bytes) snapshot=12770000896  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408698880  
Peak Map Virtual memory (bytes)=6375268352  
Peak Reduce Physical memory (bytes)=318984192  
Peak Reduce Virtual memory (bytes)=6394732544

#### 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=2680027

#### File Output Format Counters

Bytes Written=794072

23/01/04 14:29:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:29:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0139  
23/01/04 14:29:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:29:19 INFO JobSubmitter: number of splits:1  
23/01/04 14:29:19 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0139  
23/01/04 14:29:19 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:29:19 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0139  
23/01/04 14:29:19 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0139/  
23/01/04 14:29:19 INFO Job: Running job: job\_1672817468245\_0139  
23/01/04 14:29:25 INFO Job: Job job\_1672817468245\_0139 running in uber mode : false  
23/01/04 14:29:25 INFO Job: map 0% reduce 0%  
23/01/04 14:29:31 INFO Job: map 100% reduce 0%  
23/01/04 14:29:37 INFO Job: map 100% reduce 100%  
23/01/04 14:29:37 INFO Job: Job job\_1672817468245\_0139 completed successfully  
23/01/04 14:29:37 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1587758  
FILE: Number of bytes written=3707789  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4268729  
HDFS: Number of bytes written=794072  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17290  
Total time spent by all reduces in occupied slots (ms)=17475  
Total time spent by all map tasks (ms)=3458  
Total time spent by all reduce tasks (ms)=3495  
Total vcore-milliseconds taken by all map tasks=3458  
Total vcore-milliseconds taken by all reduce tasks=3495  
Total megabyte-milliseconds taken by all map tasks=17704960  
Total megabyte-milliseconds taken by all reduce tasks=17894400

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=1587747  
Map output materialized bytes=1587758  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=1587758  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=46

CPU time spent (ms)=3810  
Physical memory (bytes) snapshot=755056640  
Virtual memory (bytes) snapshot=12765265920  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419901440  
Peak Map Virtual memory (bytes)=6377746432  
Peak Reduce Physical memory (bytes)=335155200  
Peak Reduce Virtual memory (bytes)=6387519488

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=2680027

File Output Format Counters

Bytes Written=794072

23/01/04 14:29:37 INFO KMeansDriver: Clustering data  
23/01/04 14:29:37 INFO KMeansDriver: Running Clustering  
23/01/04 14:29:37 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/01 Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/01  
23/01/04 14:29:37 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:29:37 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0140  
23/01/04 14:29:37 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:29:38 INFO JobSubmitter: number of splits:1  
23/01/04 14:29:38 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0140  
23/01/04 14:29:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:29:38 INFO YarnClientImpl: Submitted application application\_1672817468245\_0140  
23/01/04 14:29:38 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0140/  
23/01/04 14:29:38 INFO Job: Running job: job\_1672817468245\_0140  
23/01/04 14:29:44 INFO Job: Job job\_1672817468245\_0140 running in uber mode : false  
23/01/04 14:29:44 INFO Job: map 0% reduce 0%  
23/01/04 14:29:49 INFO Job: map 100% reduce 0%  
23/01/04 14:29:50 INFO Job: Job job\_1672817468245\_0140 completed successfully  
23/01/04 14:29:50 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265888  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3474463  
HDFS: Number of bytes written=2680755  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17460  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3492  
Total vcore-milliseconds taken by all map tasks=3492  
Total megabyte-milliseconds taken by all map tasks=17879040

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=31  
CPU time spent (ms)=1880  
Physical memory (bytes) snapshot=297791488  
Virtual memory (bytes) snapshot=6375034880  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=297791488  
Peak Map Virtual memory (bytes)=6375034880

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680755

23/01/04 14:29:50 INFO MahoutDriver: Program took 51284 ms (Minutes: 0.85475)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:29:56 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --

endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/CosineDistanceMeasure/01/clusters-2-final], --numWords=[20], --

output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
CosineDistanceMeasure/01.txt], --outputFormat=[TEXT],

--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/  
01/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:29:57 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:29:57 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/01/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --

endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/01/  
clusters-2-final], --maxIter=[5], --method=[mapreduce],

--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 14:29:57 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 14:29:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:29:58 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0141

23/01/04 14:29:58 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:29:58 INFO JobSubmitter: number of splits:1

23/01/04 14:29:58 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0141

23/01/04 14:29:58 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:29:58 INFO Configuration: resource-types.xml not found  
23/01/04 14:29:58 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:29:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0141  
23/01/04 14:29:58 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0141/  
23/01/04 14:29:58 INFO Job: Running job: job\_1672817468245\_0141  
23/01/04 14:30:04 INFO Job: Job job\_1672817468245\_0141 running in uber mode : false  
23/01/04 14:30:04 INFO Job: map 0% reduce 0%  
23/01/04 14:30:09 INFO Job: map 100% reduce 0%  
23/01/04 14:30:15 INFO Job: map 100% reduce 100%  
23/01/04 14:30:15 INFO Job: Job job\_1672817468245\_0141 completed successfully  
23/01/04 14:30:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=36531  
FILE: Number of bytes written=605633  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3474979  
HDFS: Number of bytes written=433533  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16325  
Total time spent by all reduces in occupied slots (ms)=16560  
Total time spent by all map tasks (ms)=3265  
Total time spent by all reduce tasks (ms)=3312  
Total vcore-milliseconds taken by all map tasks=3265  
Total vcore-milliseconds taken by all reduce tasks=3312  
Total megabyte-milliseconds taken by all map tasks=16716800  
Total megabyte-milliseconds taken by all reduce tasks=16957440

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=36521  
Map output materialized bytes=36531  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=36531  
Reduce input records=1  
Reduce output records=2  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=2510  
Physical memory (bytes) snapshot=675991552  
Virtual memory (bytes) snapshot=12761190400  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=389017600

```

Peak Map Virtual memory (bytes)=6370635776
Peak Reduce Physical memory (bytes)=286973952
Peak Reduce Virtual memory (bytes)=6390763520
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=2680755
File Output Format Counters
  Bytes Written=433533
23/01/04 14:30:16 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 14:30:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:30:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0142
23/01/04 14:30:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:30:16 INFO JobSubmitter: number of splits:1
23/01/04 14:30:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0142
23/01/04 14:30:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:30:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0142
23/01/04 14:30:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0142/
23/01/04 14:30:16 INFO Job: Running job: job_1672817468245_0142
23/01/04 14:30:22 INFO Job: Job job_1672817468245_0142 running in uber mode : false
23/01/04 14:30:22 INFO Job:  map 0% reduce 0%
23/01/04 14:30:27 INFO Job:  map 100% reduce 0%
23/01/04 14:30:33 INFO Job:  map 100% reduce 100%
23/01/04 14:30:33 INFO Job: Job job_1672817468245_0142 completed successfully
23/01/04 14:30:33 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=85326
  FILE: Number of bytes written=703223
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=3548021
  HDFS: Number of bytes written=518868
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=16135
  Total time spent by all reduces in occupied slots (ms)=16480
  Total time spent by all map tasks (ms)=3227
  Total time spent by all reduce tasks (ms)=3296
  Total vcore-milliseconds taken by all map tasks=3227
  Total vcore-milliseconds taken by all reduce tasks=3296
  Total megabyte-milliseconds taken by all map tasks=16522240

```

```

        Total megabyte-milliseconds taken by all reduce tasks=16875520
Map-Reduce Framework
    Map input records=27
    Map output records=1
    Map output bytes=85315
    Map output materialized bytes=85326
    Input split bytes=200
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=85326
    Reduce input records=1
    Reduce output records=3
    Spilled Records=2
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=59
    CPU time spent (ms)=2330
    Physical memory (bytes) snapshot=673042432
    Virtual memory (bytes) snapshot=12752252928
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=386252800
    Peak Map Virtual memory (bytes)=6372777984
    Peak Reduce Physical memory (bytes)=286789632
    Peak Reduce Virtual memory (bytes)=6383382528
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=2680755
File Output Format Counters
    Bytes Written=518868
23/01/04 14:30:33 INFO RepresentativePointsDriver: Representative Points Iteration
2
23/01/04 14:30:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:30:33 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0143
23/01/04 14:30:33 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:30:33 INFO JobSubmitter: number of splits:1
23/01/04 14:30:33 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0143
23/01/04 14:30:33 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:30:33 INFO YarnClientImpl: Submitted application
application_1672817468245_0143
23/01/04 14:30:33 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0143/
23/01/04 14:30:33 INFO Job: Running job: job_1672817468245_0143
23/01/04 14:30:39 INFO Job: Job job_1672817468245_0143 running in uber mode : false
23/01/04 14:30:39 INFO Job:  map 0% reduce 0%
23/01/04 14:30:44 INFO Job:  map 100% reduce 0%
23/01/04 14:30:50 INFO Job:  map 100% reduce 100%
23/01/04 14:30:50 INFO Job: Job job_1672817468245_0143 completed successfully
23/01/04 14:30:50 INFO Job: Counters: 54

```



#### File System Counters

FILE: Number of bytes read=81234  
FILE: Number of bytes written=695039  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3718691  
HDFS: Number of bytes written=600091  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16700  
Total time spent by all reduces in occupied slots (ms)=16435  
Total time spent by all map tasks (ms)=3340  
Total time spent by all reduce tasks (ms)=3287  
Total vcore-milliseconds taken by all map tasks=3340  
Total vcore-milliseconds taken by all reduce tasks=3287  
Total megabyte-milliseconds taken by all map tasks=17100800  
Total megabyte-milliseconds taken by all reduce tasks=16829440

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=81223  
Map output materialized bytes=81234  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=81234  
Reduce input records=1  
Reduce output records=4  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=2450  
Physical memory (bytes) snapshot=672669696  
Virtual memory (bytes) snapshot=12758757376  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=388546560  
Peak Map Virtual memory (bytes)=6381621248  
Peak Reduce Physical memory (bytes)=284123136  
Peak Reduce Virtual memory (bytes)=6379089920

#### 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=2680755

#### File Output Format Counters

Bytes Written=600091  
23/01/04 14:30:50 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 14:30:51 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:30:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0144  
23/01/04 14:30:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:30:51 INFO JobSubmitter: number of splits:1  
23/01/04 14:30:51 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0144  
23/01/04 14:30:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:30:51 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0144  
23/01/04 14:30:51 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0144/  
23/01/04 14:30:51 INFO Job: Running job: job\_1672817468245\_0144  
23/01/04 14:30:57 INFO Job: Job job\_1672817468245\_0144 running in uber mode : false  
23/01/04 14:30:57 INFO Job: map 0% reduce 0%  
23/01/04 14:31:03 INFO Job: map 100% reduce 0%  
23/01/04 14:31:08 INFO Job: map 100% reduce 100%  
23/01/04 14:31:09 INFO Job: Job job\_1672817468245\_0144 completed successfully  
23/01/04 14:31:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=109331  
FILE: Number of bytes written=751233  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3881137  
HDFS: Number of bytes written=709431  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16525  
Total time spent by all reduces in occupied slots (ms)=16630  
Total time spent by all map tasks (ms)=3305  
Total time spent by all reduce tasks (ms)=3326  
Total vcore-milliseconds taken by all map tasks=3305  
Total vcore-milliseconds taken by all reduce tasks=3326  
Total megabyte-milliseconds taken by all map tasks=16921600  
Total megabyte-milliseconds taken by all reduce tasks=17029120

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=109320  
Map output materialized bytes=109331  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=109331  
Reduce input records=1  
Reduce output records=5

Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=29  
CPU time spent (ms)=2540  
Physical memory (bytes) snapshot=674242560  
Virtual memory (bytes) snapshot=12760039424  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=386506752  
Peak Map Virtual memory (bytes)=6370992128  
Peak Reduce Physical memory (bytes)=287735808  
Peak Reduce Virtual memory (bytes)=6389202944

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=2680755

File Output Format Counters

Bytes Written=709431

23/01/04 14:31:09 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:31:09 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:31:09 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0146

23/01/04 14:31:09 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:31:09 INFO JobSubmitter: number of splits:1

23/01/04 14:31:09 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0146

23/01/04 14:31:09 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:31:10 INFO YarnClientImpl: Submitted application application\_1672817468245\_0146

23/01/04 14:31:10 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0146/](http://lena-master:8088/proxy/application_1672817468245_0146/)

23/01/04 14:31:10 INFO Job: Running job: job\_1672817468245\_0146

23/01/04 14:31:16 INFO Job: Job job\_1672817468245\_0146 running in uber mode : false

23/01/04 14:31:16 INFO Job: map 0% reduce 0%

23/01/04 14:31:21 INFO Job: map 100% reduce 0%

23/01/04 14:31:27 INFO Job: map 100% reduce 100%

23/01/04 14:31:28 INFO Job: Job job\_1672817468245\_0146 completed successfully

23/01/04 14:31:28 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=111620  
FILE: Number of bytes written=755811  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4099817  
HDFS: Number of bytes written=821060  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

```

Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16555
Total time spent by all reduces in occupied slots (ms)=16500
Total time spent by all map tasks (ms)=3311
Total time spent by all reduce tasks (ms)=3300
Total vcore-milliseconds taken by all map tasks=3311
Total vcore-milliseconds taken by all reduce tasks=3300
Total megabyte-milliseconds taken by all map tasks=16952320
Total megabyte-milliseconds taken by all reduce tasks=16896000
Map-Reduce Framework
  Map input records=27
  Map output records=1
  Map output bytes=111609
  Map output materialized bytes=111620
  Input split bytes=200
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=111620
  Reduce input records=1
  Reduce output records=6
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=44
  CPU time spent (ms)=2600
  Physical memory (bytes) snapshot=687697920
  Virtual memory (bytes) snapshot=12749819904
  Total committed heap usage (bytes)=2105540608
  Peak Map Physical memory (bytes)=394850304
  Peak Map Virtual memory (bytes)=6369517568
  Peak Reduce Physical memory (bytes)=292847616
  Peak Reduce Virtual memory (bytes)=6384209920
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=2680755
File Output Format Counters
  Bytes Written=821060
23/01/04 14:31:28 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN
23/01/04 14:31:28 INFO ClusterEvaluator: Intra-Cluster Density[1] =
0.656657647846596
23/01/04 14:31:28 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.656657647846596
23/01/04 14:31:28 INFO ClusterDumper: Wrote 1 clusters
23/01/04 14:31:28 INFO MahoutDriver: Program took 92610 ms (Minutes: 1.5435)
---- K: 02 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/02
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop

```

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar  
23/01/04 14:31:33 INFO AbstractJob: Command line arguments: {--clustering=null, --clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure], --convergenceDelta=[0.1], --distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --endPhase=[2147483647], --input=[dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors], --maxIter=[20], --method=[mapreduce], --numClusters=[2], --output=[hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/02], --overwrite=null, --startPhase=[0], --tempDir=[temp]}

23/01/04 14:31:34 INFO HadoopUtil: Deleting  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/02

23/01/04 14:31:34 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure

23/01/04 14:31:34 INFO ZlibFactory: Successfully loaded & initialized native-zlib library

23/01/04 14:31:34 INFO CodecPool: Got brand-new compressor [.deflate]

23/01/04 14:31:34 INFO RandomSeedGenerator: Wrote 2 Klusters to  
dsm010/question\_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/part-randomSeed

23/01/04 14:31:34 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In: dsm010/question\_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/02

23/01/04 14:31:34 INFO KMeansDriver: convergence: 0.1 max Iterations: 20

23/01/04 14:31:34 INFO CodecPool: Got brand-new decompressor [.deflate]

23/01/04 14:31:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:31:35 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0147

23/01/04 14:31:35 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:31:35 INFO JobSubmitter: number of splits:1

23/01/04 14:31:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0147

23/01/04 14:31:35 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:31:35 INFO Configuration: resource-types.xml not found

23/01/04 14:31:35 INFO ResourceUtils: Unable to find 'resource-types.xml'.

23/01/04 14:31:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0147

23/01/04 14:31:35 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0147/

23/01/04 14:31:35 INFO Job: Running job: job\_1672817468245\_0147

23/01/04 14:31:41 INFO Job: Job job\_1672817468245\_0147 running in uber mode : false

23/01/04 14:31:41 INFO Job: map 0% reduce 0%

23/01/04 14:31:46 INFO Job: map 100% reduce 0%

23/01/04 14:31:52 INFO Job: map 100% reduce 100%

23/01/04 14:31:52 INFO Job: Job job\_1672817468245\_0147 completed successfully

23/01/04 14:31:53 INFO Job: Counters: 54

File System Counters

- FILE: Number of bytes read=1526974
- FILE: Number of bytes written=3586221
- FILE: Number of read operations=0
- FILE: Number of large read operations=0
- FILE: Number of write operations=0
- HDFS: Number of bytes read=4042753
- HDFS: Number of bytes written=1300151

HDFS: Number of read operations=23  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17120  
Total time spent by all reduces in occupied slots (ms)=17575  
Total time spent by all map tasks (ms)=3424  
Total time spent by all reduce tasks (ms)=3515  
Total vcore-milliseconds taken by all map tasks=3424  
Total vcore-milliseconds taken by all reduce tasks=3515  
Total megabyte-milliseconds taken by all map tasks=17530880  
Total megabyte-milliseconds taken by all reduce tasks=17996800

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=1526958  
Map output materialized bytes=1526974  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=1526974  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=3830  
Physical memory (bytes) snapshot=735551488  
Virtual memory (bytes) snapshot=12765396992  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399605760  
Peak Map Virtual memory (bytes)=6380978176  
Peak Reduce Physical memory (bytes)=335945728  
Peak Reduce Virtual memory (bytes)=6384418816

#### 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=2680027

#### File Output Format Counters

Bytes Written=1300151

23/01/04 14:31:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:31:53 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0148

23/01/04 14:31:53 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:31:53 INFO JobSubmitter: number of splits:1

23/01/04 14:31:53 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0148  
23/01/04 14:31:53 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:31:53 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0148  
23/01/04 14:31:53 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0148/  
23/01/04 14:31:53 INFO Job: Running job: job\_1672817468245\_0148  
23/01/04 14:31:59 INFO Job: Job job\_1672817468245\_0148 running in uber mode : false  
23/01/04 14:31:59 INFO Job: map 0% reduce 0%  
23/01/04 14:32:05 INFO Job: map 100% reduce 0%  
23/01/04 14:32:11 INFO Job: map 100% reduce 100%  
23/01/04 14:32:11 INFO Job: Job job\_1672817468245\_0148 completed successfully  
23/01/04 14:32:11 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2599710  
FILE: Number of bytes written=5731693  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5280887  
HDFS: Number of bytes written=1300151  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17535  
Total time spent by all reduces in occupied slots (ms)=17440  
Total time spent by all map tasks (ms)=3507  
Total time spent by all reduce tasks (ms)=3488  
Total vcore-milliseconds taken by all map tasks=3507  
Total vcore-milliseconds taken by all reduce tasks=3488  
Total megabyte-milliseconds taken by all map tasks=17955840  
Total megabyte-milliseconds taken by all reduce tasks=17858560

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=2599694  
Map output materialized bytes=2599710  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=2599710  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=3950  
Physical memory (bytes) snapshot=733011968  
Virtual memory (bytes) snapshot=12753895424  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403419136

```

Peak Map Virtual memory (bytes)=6371901440
Peak Reduce Physical memory (bytes)=329592832
Peak Reduce Virtual memory (bytes)=6381993984
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=2680027
File Output Format Counters
  Bytes Written=1300151
23/01/04 14:32:11 INFO KMeansDriver: Clustering data
23/01/04 14:32:11 INFO KMeansDriver: Running Clustering
23/01/04 14:32:11 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/02 Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-
fiction-corpus-kmeans/CosineDistanceMeasure/02
23/01/04 14:32:11 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:32:11 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0149
23/01/04 14:32:11 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:32:11 INFO JobSubmitter: number of splits:1
23/01/04 14:32:11 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0149
23/01/04 14:32:11 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:32:11 INFO YarnClientImpl: Submitted application
application_1672817468245_0149
23/01/04 14:32:11 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0149/
23/01/04 14:32:11 INFO Job: Running job: job_1672817468245_0149
23/01/04 14:32:18 INFO Job: Job job_1672817468245_0149 running in uber mode : false
23/01/04 14:32:18 INFO Job:  map 0% reduce 0%
23/01/04 14:32:24 INFO Job:  map 100% reduce 0%
23/01/04 14:32:24 INFO Job: Job job_1672817468245_0149 completed successfully
23/01/04 14:32:24 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265888
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=3980542
  HDFS: Number of bytes written=2680752
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17245
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=3449
  Total vcore-milliseconds taken by all map tasks=3449

```



Total megabyte-milliseconds taken by all map tasks=17658880  
Map-Reduce Framework  
Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=19  
CPU time spent (ms)=1940  
Physical memory (bytes) snapshot=311480320  
Virtual memory (bytes) snapshot=6379782144  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=311480320  
Peak Map Virtual memory (bytes)=6379782144  
File Input Format Counters  
Bytes Read=2680027  
File Output Format Counters  
Bytes Written=2680752

23/01/04 14:32:24 INFO MahoutDriver: Program took 50420 ms (Minutes:  
0.84033333333333334)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:32:30 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --

endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/CosineDistanceMeasure/02/clusters-2-final], --numWords=[20], --

output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
CosineDistanceMeasure/02.txt], --outputFormat=[TEXT],

--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/  
02/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:32:32 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:32:32 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/02/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/02/  
clusters-2-final], --maxIter=[5], --method=[mapreduce],

--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 14:32:32 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 14:32:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:32:32 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0150

23/01/04 14:32:32 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:32:32 INFO JobSubmitter: number of splits:1

23/01/04 14:32:33 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0150

23/01/04 14:32:33 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:32:33 INFO Configuration: resource-types.xml not found

23/01/04 14:32:33 INFO ResourceUtils: Unable to find 'resource-types.xml'.

23/01/04 14:32:33 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0150

23/01/04 14:32:33 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0150/  
23/01/04 14:32:33 INFO Job: Running job: job\_1672817468245\_0150  
23/01/04 14:32:39 INFO Job: Job job\_1672817468245\_0150 running in uber mode : false  
23/01/04 14:32:39 INFO Job: map 0% reduce 0%  
23/01/04 14:32:44 INFO Job: map 100% reduce 0%  
23/01/04 14:32:50 INFO Job: map 100% reduce 100%  
23/01/04 14:32:50 INFO Job: Job job\_1672817468245\_0150 completed successfully  
23/01/04 14:32:50 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=121851  
FILE: Number of bytes written=776273  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3980946  
HDFS: Number of bytes written=771853  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16150  
Total time spent by all reduces in occupied slots (ms)=16495  
Total time spent by all map tasks (ms)=3230  
Total time spent by all reduce tasks (ms)=3299  
Total vcore-milliseconds taken by all map tasks=3230  
Total vcore-milliseconds taken by all reduce tasks=3299  
Total megabyte-milliseconds taken by all map tasks=16537600  
Total megabyte-milliseconds taken by all reduce tasks=16890880

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=121836  
Map output materialized bytes=121851  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=121851  
Reduce input records=2  
Reduce output records=4  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=2410  
Physical memory (bytes) snapshot=683347968  
Virtual memory (bytes) snapshot=12766371840  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398286848  
Peak Map Virtual memory (bytes)=6376349696  
Peak Reduce Physical memory (bytes)=285061120  
Peak Reduce Virtual memory (bytes)=6391975936

#### 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=2680752
    File Output Format Counters
        Bytes Written=771853
23/01/04 14:32:50 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 14:32:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:32:50 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0151
23/01/04 14:32:50 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:32:50 INFO JobSubmitter: number of splits:1
23/01/04 14:32:50 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0151
23/01/04 14:32:50 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:32:50 INFO YarnClientImpl: Submitted application
application_1672817468245_0151
23/01/04 14:32:50 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0151/
23/01/04 14:32:50 INFO Job: Running job: job_1672817468245_0151
23/01/04 14:32:57 INFO Job: Job job_1672817468245_0151 running in uber mode : false
23/01/04 14:32:57 INFO Job:  map 0% reduce 0%
23/01/04 14:33:03 INFO Job:  map 100% reduce 0%
23/01/04 14:33:08 INFO Job:  map 100% reduce 100%
23/01/04 14:33:09 INFO Job: Job job_1672817468245_0151 completed successfully
23/01/04 14:33:09 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=183809
        FILE: Number of bytes written=900189
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4224658
        HDFS: Number of bytes written=955666
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17145
        Total time spent by all reduces in occupied slots (ms)=16515
        Total time spent by all map tasks (ms)=3429
        Total time spent by all reduce tasks (ms)=3303
        Total vcore-milliseconds taken by all map tasks=3429
        Total vcore-milliseconds taken by all reduce tasks=3303
        Total megabyte-milliseconds taken by all map tasks=17556480
        Total megabyte-milliseconds taken by all reduce tasks=16911360
    Map-Reduce Framework
        Map input records=27
        Map output records=2

```

Map output bytes=183793  
Map output materialized bytes=183809  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=183809  
Reduce input records=2  
Reduce output records=6  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=2660  
Physical memory (bytes) snapshot=690577408  
Virtual memory (bytes) snapshot=12753641472  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396935168  
Peak Map Virtual memory (bytes)=6372118528  
Peak Reduce Physical memory (bytes)=293642240  
Peak Reduce Virtual memory (bytes)=6383476736

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=2680752

File Output Format Counters

Bytes Written=955666

23/01/04 14:33:09 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 14:33:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:33:09 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0152

23/01/04 14:33:09 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:33:09 INFO JobSubmitter: number of splits:1

23/01/04 14:33:09 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0152

23/01/04 14:33:09 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:33:09 INFO YarnClientImpl: Submitted application application\_1672817468245\_0152

23/01/04 14:33:09 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0152/

23/01/04 14:33:09 INFO Job: Running job: job\_1672817468245\_0152

23/01/04 14:33:15 INFO Job: Job job\_1672817468245\_0152 running in uber mode : false

23/01/04 14:33:15 INFO Job: map 0% reduce 0%

23/01/04 14:33:21 INFO Job: map 100% reduce 0%

23/01/04 14:33:27 INFO Job: map 100% reduce 100%

23/01/04 14:33:27 INFO Job: Job job\_1672817468245\_0152 completed successfully

23/01/04 14:33:27 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=139195

FILE: Number of bytes written=810961

FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4592284  
HDFS: Number of bytes written=1094866  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16665  
Total time spent by all reduces in occupied slots (ms)=16620  
Total time spent by all map tasks (ms)=3333  
Total time spent by all reduce tasks (ms)=3324  
Total vcore-milliseconds taken by all map tasks=3333  
Total vcore-milliseconds taken by all reduce tasks=3324  
Total megabyte-milliseconds taken by all map tasks=17064960  
Total megabyte-milliseconds taken by all reduce tasks=17018880

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=139180  
Map output materialized bytes=139195  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=139195  
Reduce input records=2  
Reduce output records=8  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=2750  
Physical memory (bytes) snapshot=681140224  
Virtual memory (bytes) snapshot=12740022272  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392302592  
Peak Map Virtual memory (bytes)=6365097984  
Peak Reduce Physical memory (bytes)=288837632  
Peak Reduce Virtual memory (bytes)=6382010368

#### 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=2680752

#### File Output Format Counters

Bytes Written=1094866

23/01/04 14:33:27 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 14:33:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:33:27 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0153  
23/01/04 14:33:27 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:33:27 INFO JobSubmitter: number of splits:1  
23/01/04 14:33:27 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0153  
23/01/04 14:33:27 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:33:27 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0153  
23/01/04 14:33:27 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0153/  
23/01/04 14:33:27 INFO Job: Running job: job\_1672817468245\_0153  
23/01/04 14:33:33 INFO Job: Job job\_1672817468245\_0153 running in uber mode : false  
23/01/04 14:33:34 INFO Job: map 0% reduce 0%  
23/01/04 14:33:39 INFO Job: map 100% reduce 0%  
23/01/04 14:33:45 INFO Job: map 100% reduce 100%  
23/01/04 14:33:45 INFO Job: Job job\_1672817468245\_0153 completed successfully  
23/01/04 14:33:45 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=193705  
FILE: Number of bytes written=919981  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4870684  
HDFS: Number of bytes written=1288555  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16680  
Total time spent by all reduces in occupied slots (ms)=16285  
Total time spent by all map tasks (ms)=3336  
Total time spent by all reduce tasks (ms)=3257  
Total vcore-milliseconds taken by all map tasks=3336  
Total vcore-milliseconds taken by all reduce tasks=3257  
Total megabyte-milliseconds taken by all map tasks=17080320  
Total megabyte-milliseconds taken by all reduce tasks=16675840

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=193689  
Map output materialized bytes=193705  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=193705  
Reduce input records=2  
Reduce output records=10  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=69
CPU time spent (ms)=2690
Physical memory (bytes) snapshot=683241472
Virtual memory (bytes) snapshot=12756017152
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=391385088
Peak Map Virtual memory (bytes)=6369030144
Peak Reduce Physical memory (bytes)=291856384
Peak Reduce Virtual memory (bytes)=6388940800
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=2680752
File Output Format Counters
  Bytes Written=1288555
23/01/04 14:33:45 INFO RepresentativePointsDriver: Representative Points Iteration
4
23/01/04 14:33:45 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:33:45 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0154
23/01/04 14:33:45 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:33:45 INFO JobSubmitter: number of splits:1
23/01/04 14:33:45 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0154
23/01/04 14:33:45 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:33:45 INFO YarnClientImpl: Submitted application
application_1672817468245_0154
23/01/04 14:33:45 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0154/
23/01/04 14:33:45 INFO Job: Running job: job_1672817468245_0154
23/01/04 14:33:51 INFO Job: Job job_1672817468245_0154 running in uber mode : false
23/01/04 14:33:51 INFO Job:  map 0% reduce 0%
23/01/04 14:33:57 INFO Job:  map 100% reduce 0%
23/01/04 14:34:03 INFO Job:  map 100% reduce 100%
23/01/04 14:34:03 INFO Job: Job job_1672817468245_0154 completed successfully
23/01/04 14:34:03 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=248428
  FILE: Number of bytes written=1029427
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5258062
  HDFS: Number of bytes written=1537007
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=16545

```

Total time spent by all reduces in occupied slots (ms)=16725  
Total time spent by all map tasks (ms)=3309  
Total time spent by all reduce tasks (ms)=3345  
Total vcore-milliseconds taken by all map tasks=3309  
Total vcore-milliseconds taken by all reduce tasks=3345  
Total megabyte-milliseconds taken by all map tasks=16942080  
Total megabyte-milliseconds taken by all reduce tasks=17126400

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=248412  
Map output materialized bytes=248428  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=248428  
Reduce input records=2  
Reduce output records=12  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=2870  
Physical memory (bytes) snapshot=692051968  
Virtual memory (bytes) snapshot=12759179264  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396181504  
Peak Map Virtual memory (bytes)=6375981056  
Peak Reduce Physical memory (bytes)=295870464  
Peak Reduce Virtual memory (bytes)=6385152000

#### 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=2680752

#### File Output Format Counters

Bytes Written=1537007

23/01/04 14:34:03 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN

23/01/04 14:34:03 INFO ClusterEvaluator: Intra-Cluster Density[8] =

0.6090761597660228

23/01/04 14:34:03 INFO ClusterEvaluator: Intra-Cluster Density[21] =

0.7250167378006033

23/01/04 14:34:03 INFO ClusterEvaluator: Average Intra-Cluster Density =

0.667046448783313

23/01/04 14:34:03 INFO ClusterDumper: Wrote 2 clusters

23/01/04 14:34:03 INFO MahoutDriver: Program took 93162 ms (Minutes: 1.5527)

---- K: 03 --

dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/03

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:34:08 INFO AbstractJob: Command line arguments: {--clustering=null, --



```

clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[3],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/03], --overwrite=null, --startPhase=[0], --
tempDir=[temp]]
23/01/04 14:34:09 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/03
23/01/04 14:34:09 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:34:09 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:34:09 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:34:09 INFO RandomSeedGenerator: Wrote 3 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:34:09 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/03
23/01/04 14:34:09 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:34:09 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:34:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:34:10 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0155
23/01/04 14:34:10 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:34:10 INFO JobSubmitter: number of splits:1
23/01/04 14:34:10 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0155
23/01/04 14:34:10 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:34:10 INFO Configuration: resource-types.xml not found
23/01/04 14:34:10 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:34:10 INFO YarnClientImpl: Submitted application
application_1672817468245_0155
23/01/04 14:34:10 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0155/
23/01/04 14:34:10 INFO Job: Running job: job_1672817468245_0155
23/01/04 14:34:16 INFO Job: Job job_1672817468245_0155 running in uber mode : false
23/01/04 14:34:16 INFO Job:  map 0% reduce 0%
23/01/04 14:34:23 INFO Job:  map 100% reduce 0%
23/01/04 14:34:28 INFO Job:  map 100% reduce 100%
23/01/04 14:34:29 INFO Job: Job job_1672817468245_0155 completed successfully
23/01/04 14:34:29 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=1942588
        FILE: Number of bytes written=4417449
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4192421
        HDFS: Number of bytes written=1690814
        HDFS: Number of read operations=27
        HDFS: Number of large read operations=0

```

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17545  
Total time spent by all reduces in occupied slots (ms)=17570  
Total time spent by all map tasks (ms)=3509  
Total time spent by all reduce tasks (ms)=3514  
Total vcore-milliseconds taken by all map tasks=3509  
Total vcore-milliseconds taken by all reduce tasks=3514  
Total megabyte-milliseconds taken by all map tasks=17966080  
Total megabyte-milliseconds taken by all reduce tasks=17991680

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=1942567  
Map output materialized bytes=1942588  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=1942588  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=55  
CPU time spent (ms)=3920  
Physical memory (bytes) snapshot=734302208  
Virtual memory (bytes) snapshot=12762460160  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419950592  
Peak Map Virtual memory (bytes)=6378467328  
Peak Reduce Physical memory (bytes)=314351616  
Peak Reduce Virtual memory (bytes)=6383992832

#### 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=2680027

#### File Output Format Counters

Bytes Written=1690814

23/01/04 14:34:29 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:34:29 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0156

23/01/04 14:34:29 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:34:29 INFO JobSubmitter: number of splits:1

23/01/04 14:34:29 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0156

23/01/04 14:34:29 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:34:29 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0156  
23/01/04 14:34:29 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0156/  
23/01/04 14:34:29 INFO Job: Running job: job\_1672817468245\_0156  
23/01/04 14:34:35 INFO Job: Job job\_1672817468245\_0156 running in uber mode : false  
23/01/04 14:34:35 INFO Job: map 0% reduce 0%  
23/01/04 14:34:41 INFO Job: map 100% reduce 0%  
23/01/04 14:34:47 INFO Job: map 100% reduce 100%  
23/01/04 14:34:47 INFO Job: Job job\_1672817468245\_0156 completed successfully  
23/01/04 14:34:47 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3380882  
FILE: Number of bytes written=7294037  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6062213  
HDFS: Number of bytes written=1690814  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17810  
Total time spent by all reduces in occupied slots (ms)=17500  
Total time spent by all map tasks (ms)=3562  
Total time spent by all reduce tasks (ms)=3500  
Total vcore-milliseconds taken by all map tasks=3562  
Total vcore-milliseconds taken by all reduce tasks=3500  
Total megabyte-milliseconds taken by all map tasks=18237440  
Total megabyte-milliseconds taken by all reduce tasks=17920000

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=3380861  
Map output materialized bytes=3380882  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=3380882  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=3960  
Physical memory (bytes) snapshot=755560448  
Virtual memory (bytes) snapshot=12764581888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403968000  
Peak Map Virtual memory (bytes)=6373298176  
Peak Reduce Physical memory (bytes)=351592448

```

        Peak Reduce Virtual memory (bytes)=6391283712
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=2680027
File Output Format Counters
    Bytes Written=1690814
23/01/04 14:34:47 INFO KMeansDriver: Clustering data
23/01/04 14:34:47 INFO KMeansDriver: Running Clustering
23/01/04 14:34:47 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/03 Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-
fiction-corpus-kmeans/CosineDistanceMeasure/03
23/01/04 14:34:47 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:34:47 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0157
23/01/04 14:34:48 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:34:48 INFO JobSubmitter: number of splits:1
23/01/04 14:34:48 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0157
23/01/04 14:34:48 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:34:48 INFO YarnClientImpl: Submitted application
application_1672817468245_0157
23/01/04 14:34:48 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0157/
23/01/04 14:34:48 INFO Job: Running job: job_1672817468245_0157
23/01/04 14:34:54 INFO Job: Job job_1672817468245_0157 running in uber mode : false
23/01/04 14:34:54 INFO Job:  map 0% reduce 0%
23/01/04 14:35:00 INFO Job:  map 100% reduce 0%
23/01/04 14:35:00 INFO Job: Job job_1672817468245_0157 completed successfully
23/01/04 14:35:00 INFO Job: Counters: 33
File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=265888
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4371205
    HDFS: Number of bytes written=2680758
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17580
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=3516
    Total vcore-milliseconds taken by all map tasks=3516
    Total megabyte-milliseconds taken by all map tasks=18001920
Map-Reduce Framework

```

```

Map input records=27
Map output records=27
Input split bytes=170
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=42
CPU time spent (ms)=2050
Physical memory (bytes) snapshot=301518848
Virtual memory (bytes) snapshot=6378082304
Total committed heap usage (bytes)=1052770304
Peak Map Physical memory (bytes)=301518848
Peak Map Virtual memory (bytes)=6378082304
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680758
23/01/04 14:35:00 INFO MahoutDriver: Program took 51745 ms (Minutes:
0.8624166666666667)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:35:06 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/CosineDistanceMeasure/03/clusters-2-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
CosineDistanceMeasure/03.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/
03/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:35:08 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:35:08 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/03/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/03/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:35:08 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:35:08 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:35:08 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0158
23/01/04 14:35:08 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:35:08 INFO JobSubmitter: number of splits:1
23/01/04 14:35:09 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0158
23/01/04 14:35:09 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:35:09 INFO Configuration: resource-types.xml not found
23/01/04 14:35:09 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:35:09 INFO YarnClientImpl: Submitted application
application_1672817468245_0158
23/01/04 14:35:09 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0158/

```

23/01/04 14:35:09 INFO Job: Running job: job\_1672817468245\_0158  
23/01/04 14:35:15 INFO Job: Job job\_1672817468245\_0158 running in uber mode : false  
23/01/04 14:35:15 INFO Job: map 0% reduce 0%  
23/01/04 14:35:20 INFO Job: map 100% reduce 0%  
23/01/04 14:35:26 INFO Job: map 100% reduce 100%  
23/01/04 14:35:26 INFO Job: Job job\_1672817468245\_0158 completed successfully  
23/01/04 14:35:26 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=193757  
FILE: Number of bytes written=920085  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4371558  
HDFS: Number of bytes written=1039057  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16470  
Total time spent by all reduces in occupied slots (ms)=16425  
Total time spent by all map tasks (ms)=3294  
Total time spent by all reduce tasks (ms)=3285  
Total vcore-milliseconds taken by all map tasks=3294  
Total vcore-milliseconds taken by all reduce tasks=3285  
Total megabyte-milliseconds taken by all map tasks=16865280  
Total megabyte-milliseconds taken by all reduce tasks=16819200

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=193737  
Map output materialized bytes=193757  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=193757  
Reduce input records=3  
Reduce output records=6  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=35  
CPU time spent (ms)=2540  
Physical memory (bytes) snapshot=694108160  
Virtual memory (bytes) snapshot=12752064512  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406876160  
Peak Map Virtual memory (bytes)=6376181760  
Peak Reduce Physical memory (bytes)=287232000  
Peak Reduce Virtual memory (bytes)=6377836544

#### 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=2680758
    File Output Format Counters
        Bytes Written=1039057
23/01/04 14:35:26 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 14:35:26 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:35:26 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0159
23/01/04 14:35:26 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:35:26 INFO JobSubmitter: number of splits:1
23/01/04 14:35:26 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0159
23/01/04 14:35:26 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:35:26 INFO YarnClientImpl: Submitted application
application_1672817468245_0159
23/01/04 14:35:26 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0159/
23/01/04 14:35:26 INFO Job: Running job: job_1672817468245_0159
23/01/04 14:35:32 INFO Job: Job job_1672817468245_0159 running in uber mode : false
23/01/04 14:35:32 INFO Job:  map 0% reduce 0%
23/01/04 14:35:38 INFO Job:  map 100% reduce 0%
23/01/04 14:35:43 INFO Job:  map 100% reduce 100%
23/01/04 14:35:44 INFO Job: Job job_1672817468245_0159 completed successfully
23/01/04 14:35:44 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=275377
        FILE: Number of bytes written=1083325
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4759072
        HDFS: Number of bytes written=1314433
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=16850
        Total time spent by all reduces in occupied slots (ms)=16490
        Total time spent by all map tasks (ms)=3370
        Total time spent by all reduce tasks (ms)=3298
        Total vcore-milliseconds taken by all map tasks=3370
        Total vcore-milliseconds taken by all reduce tasks=3298
        Total megabyte-milliseconds taken by all map tasks=17254400
        Total megabyte-milliseconds taken by all reduce tasks=16885760
    Map-Reduce Framework
        Map input records=27
        Map output records=3
        Map output bytes=275356
        Map output materialized bytes=275377

```

Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=275377  
Reduce input records=3  
Reduce output records=9  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=2740  
Physical memory (bytes) snapshot=688254976  
Virtual memory (bytes) snapshot=12762349568  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=393744384  
Peak Map Virtual memory (bytes)=6379536384  
Peak Reduce Physical memory (bytes)=294510592  
Peak Reduce Virtual memory (bytes)=6388551680

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=2680758

File Output Format Counters

Bytes Written=1314433

23/01/04 14:35:44 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 14:35:45 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:35:45 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0160

23/01/04 14:35:45 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:35:45 INFO JobSubmitter: number of splits:1

23/01/04 14:35:45 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0160

23/01/04 14:35:45 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:35:45 INFO YarnClientImpl: Submitted application application\_1672817468245\_0160

23/01/04 14:35:45 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0160/

23/01/04 14:35:45 INFO Job: Running job: job\_1672817468245\_0160

23/01/04 14:35:51 INFO Job: Job job\_1672817468245\_0160 running in uber mode : false

23/01/04 14:35:51 INFO Job: map 0% reduce 0%

23/01/04 14:35:56 INFO Job: map 100% reduce 0%

23/01/04 14:36:02 INFO Job: map 100% reduce 100%

23/01/04 14:36:03 INFO Job: Job job\_1672817468245\_0160 completed successfully

23/01/04 14:36:03 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=280608

FILE: Number of bytes written=1093787

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0



HDFS: Number of bytes read=5309824  
HDFS: Number of bytes written=1595080  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17055  
Total time spent by all reduces in occupied slots (ms)=17220  
Total time spent by all map tasks (ms)=3411  
Total time spent by all reduce tasks (ms)=3444  
Total vcore-milliseconds taken by all map tasks=3411  
Total vcore-milliseconds taken by all reduce tasks=3444  
Total megabyte-milliseconds taken by all map tasks=17464320  
Total megabyte-milliseconds taken by all reduce tasks=17633280

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=280587  
Map output materialized bytes=280608  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=280608  
Reduce input records=3  
Reduce output records=12  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=50  
CPU time spent (ms)=2640  
Physical memory (bytes) snapshot=684589056  
Virtual memory (bytes) snapshot=12757057536  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396599296  
Peak Map Virtual memory (bytes)=6374051840  
Peak Reduce Physical memory (bytes)=287989760  
Peak Reduce Virtual memory (bytes)=6391988224

#### 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=2680758

#### File Output Format Counters

Bytes Written=1595080

23/01/04 14:36:03 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 14:36:03 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:36:03 INFO JobResourceUploader: Disabling Erasure Coding for path:

/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0162  
23/01/04 14:36:03 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:36:03 INFO JobSubmitter: number of splits:1  
23/01/04 14:36:03 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0162  
23/01/04 14:36:03 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:36:03 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0162  
23/01/04 14:36:03 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0162/  
23/01/04 14:36:03 INFO Job: Running job: job\_1672817468245\_0162  
23/01/04 14:36:09 INFO Job: Job job\_1672817468245\_0162 running in uber mode : false  
23/01/04 14:36:09 INFO Job: map 0% reduce 0%  
23/01/04 14:36:15 INFO Job: map 100% reduce 0%  
23/01/04 14:36:20 INFO Job: map 100% reduce 100%  
23/01/04 14:36:21 INFO Job: Job job\_1672817468245\_0162 completed successfully  
23/01/04 14:36:21 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=297659  
FILE: Number of bytes written=1127889  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5871118  
HDFS: Number of bytes written=1892739  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17275  
Total time spent by all reduces in occupied slots (ms)=16780  
Total time spent by all map tasks (ms)=3455  
Total time spent by all reduce tasks (ms)=3356  
Total vcore-milliseconds taken by all map tasks=3455  
Total vcore-milliseconds taken by all reduce tasks=3356  
Total megabyte-milliseconds taken by all map tasks=17689600  
Total megabyte-milliseconds taken by all reduce tasks=17182720

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=297639  
Map output materialized bytes=297659  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=297659  
Reduce input records=3  
Reduce output records=15  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=77  
CPU time spent (ms)=2840

Physical memory (bytes) snapshot=694886400  
Virtual memory (bytes) snapshot=12766380032  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398610432  
Peak Map Virtual memory (bytes)=6378029056  
Peak Reduce Physical memory (bytes)=296275968  
Peak Reduce Virtual memory (bytes)=6390304768

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=2680758

File Output Format Counters

Bytes Written=1892739

23/01/04 14:36:21 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:36:21 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:36:21 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0163

23/01/04 14:36:22 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:36:22 INFO JobSubmitter: number of splits:1

23/01/04 14:36:22 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0163

23/01/04 14:36:22 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:36:22 INFO YarnClientImpl: Submitted application application\_1672817468245\_0163

23/01/04 14:36:22 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0163/

23/01/04 14:36:22 INFO Job: Running job: job\_1672817468245\_0163

23/01/04 14:36:28 INFO Job: Job job\_1672817468245\_0163 running in uber mode : false

23/01/04 14:36:28 INFO Job: map 0% reduce 0%

23/01/04 14:36:34 INFO Job: map 100% reduce 0%

23/01/04 14:36:40 INFO Job: map 100% reduce 100%

23/01/04 14:36:40 INFO Job: Job job\_1672817468245\_0163 completed successfully

23/01/04 14:36:40 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=339395  
FILE: Number of bytes written=1211361  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6466436  
HDFS: Number of bytes written=2232173  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16985  
Total time spent by all reduces in occupied slots (ms)=16690  
Total time spent by all map tasks (ms)=3397

Total time spent by all reduce tasks (ms)=3338  
Total vcore-milliseconds taken by all map tasks=3397  
Total vcore-milliseconds taken by all reduce tasks=3338  
Total megabyte-milliseconds taken by all map tasks=17392640  
Total megabyte-milliseconds taken by all reduce tasks=17090560

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=339374  
Map output materialized bytes=339395  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=339395  
Reduce input records=3  
Reduce output records=18  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=2750  
Physical memory (bytes) snapshot=693161984  
Virtual memory (bytes) snapshot=12747898880  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398987264  
Peak Map Virtual memory (bytes)=6369009664  
Peak Reduce Physical memory (bytes)=294174720  
Peak Reduce Virtual memory (bytes)=6382796800

#### 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=2680758

#### File Output Format Counters

Bytes Written=2232173

23/01/04 14:36:40 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.5936570599357017

23/01/04 14:36:40 INFO ClusterEvaluator: Intra-Cluster Density[13] =  
0.6020073078755733

23/01/04 14:36:40 INFO ClusterEvaluator: Intra-Cluster Density[10] =  
0.5921044637222637

23/01/04 14:36:40 INFO ClusterEvaluator: Intra-Cluster Density[19] =  
0.6077541437052906

23/01/04 14:36:40 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.6006219717677093

23/01/04 14:36:40 INFO ClusterDumper: Wrote 3 clusters

23/01/04 14:36:40 INFO MahoutDriver: Program took 94072 ms (Minutes:  
1.5678666666666667)

---- K: 04 --

dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/04

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

```

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:36:45 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[4],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/04], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:36:45 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/04
23/01/04 14:36:45 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:36:46 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:36:46 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:36:46 INFO RandomSeedGenerator: Wrote 4 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:36:46 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/04
23/01/04 14:36:46 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:36:46 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:36:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:36:46 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0164
23/01/04 14:36:46 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:36:46 INFO JobSubmitter: number of splits:1
23/01/04 14:36:47 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0164
23/01/04 14:36:47 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:36:47 INFO Configuration: resource-types.xml not found
23/01/04 14:36:47 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:36:47 INFO YarnClientImpl: Submitted application
application_1672817468245_0164
23/01/04 14:36:47 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0164/
23/01/04 14:36:47 INFO Job: Running job: job_1672817468245_0164
23/01/04 14:36:53 INFO Job: Job job_1672817468245_0164 running in uber mode : false
23/01/04 14:36:53 INFO Job: map 0% reduce 0%
23/01/04 14:36:59 INFO Job: map 100% reduce 0%
23/01/04 14:37:05 INFO Job: map 100% reduce 100%
23/01/04 14:37:05 INFO Job: Job job_1672817468245_0164 completed successfully
23/01/04 14:37:05 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2095990
        FILE: Number of bytes written=4724253
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5212777
        HDFS: Number of bytes written=1561705

```

HDFS: Number of read operations=31  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17695  
Total time spent by all reduces in occupied slots (ms)=17585  
Total time spent by all map tasks (ms)=3539  
Total time spent by all reduce tasks (ms)=3517  
Total vcore-milliseconds taken by all map tasks=3539  
Total vcore-milliseconds taken by all reduce tasks=3517  
Total megabyte-milliseconds taken by all map tasks=18119680  
Total megabyte-milliseconds taken by all reduce tasks=18007040

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2095964  
Map output materialized bytes=2095990  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2095990  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=76  
CPU time spent (ms)=3900  
Physical memory (bytes) snapshot=717373440  
Virtual memory (bytes) snapshot=12757098496  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398983168  
Peak Map Virtual memory (bytes)=6380892160  
Peak Reduce Physical memory (bytes)=318390272  
Peak Reduce Virtual memory (bytes)=6376206336

#### 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=2680027

#### File Output Format Counters

Bytes Written=1561705

23/01/04 14:37:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:37:05 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0165

23/01/04 14:37:05 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:37:05 INFO JobSubmitter: number of splits:1

23/01/04 14:37:05 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0165  
23/01/04 14:37:05 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:37:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0165  
23/01/04 14:37:05 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0165/  
23/01/04 14:37:05 INFO Job: Running job: job\_1672817468245\_0165  
23/01/04 14:37:12 INFO Job: Job job\_1672817468245\_0165 running in uber mode : false  
23/01/04 14:37:12 INFO Job: map 0% reduce 0%  
23/01/04 14:37:17 INFO Job: map 100% reduce 0%  
23/01/04 14:37:23 INFO Job: map 100% reduce 100%  
23/01/04 14:37:23 INFO Job: Job job\_1672817468245\_0165 completed successfully  
23/01/04 14:37:23 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3234958  
FILE: Number of bytes written=7002189  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5803995  
HDFS: Number of bytes written=1561705  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17730  
Total time spent by all reduces in occupied slots (ms)=17690  
Total time spent by all map tasks (ms)=3546  
Total time spent by all reduce tasks (ms)=3538  
Total vcore-milliseconds taken by all map tasks=3546  
Total vcore-milliseconds taken by all reduce tasks=3538  
Total megabyte-milliseconds taken by all map tasks=18155520  
Total megabyte-milliseconds taken by all reduce tasks=18114560

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3234932  
Map output materialized bytes=3234958  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3234958  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3840  
Physical memory (bytes) snapshot=738205696  
Virtual memory (bytes) snapshot=12748197888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413646848

```

Peak Map Virtual memory (bytes)=6376636416
Peak Reduce Physical memory (bytes)=324558848
Peak Reduce Virtual memory (bytes)=6371561472
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=2680027
File Output Format Counters
  Bytes Written=1561705
23/01/04 14:37:23 INFO KMeansDriver: Clustering data
23/01/04 14:37:23 INFO KMeansDriver: Running Clustering
23/01/04 14:37:23 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/04 Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-
fiction-corpus-kmeans/CosineDistanceMeasure/04
23/01/04 14:37:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:37:23 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0166
23/01/04 14:37:23 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:37:23 INFO JobSubmitter: number of splits:1
23/01/04 14:37:23 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0166
23/01/04 14:37:23 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:37:23 INFO YarnClientImpl: Submitted application
application_1672817468245_0166
23/01/04 14:37:23 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0166/
23/01/04 14:37:23 INFO Job: Running job: job_1672817468245_0166
23/01/04 14:37:29 INFO Job: Job job_1672817468245_0166 running in uber mode : false
23/01/04 14:37:29 INFO Job:  map 0% reduce 0%
23/01/04 14:37:35 INFO Job:  map 100% reduce 0%
23/01/04 14:37:35 INFO Job: Job job_1672817468245_0166 completed successfully
23/01/04 14:37:35 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265888
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4242096
  HDFS: Number of bytes written=2680740
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17920
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=3584
  Total vcore-milliseconds taken by all map tasks=3584

```



```

        Total megabyte-milliseconds taken by all map tasks=18350080
Map-Reduce Framework
    Map input records=27
    Map output records=27
    Input split bytes=170
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=31
    CPU time spent (ms)=2170
    Physical memory (bytes) snapshot=316309504
    Virtual memory (bytes) snapshot=6379040768
    Total committed heap usage (bytes)=1052770304
    Peak Map Physical memory (bytes)=316309504
    Peak Map Virtual memory (bytes)=6379040768
File Input Format Counters
    Bytes Read=2680027
File Output Format Counters
    Bytes Written=2680740
23/01/04 14:37:35 INFO MahoutDriver: Program took 50799 ms (Minutes: 0.84665)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:37:41 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/CosineDistanceMeasure/04/clusters-2-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
CosineDistanceMeasure/04.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/
04/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:37:43 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:37:43 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/04/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/04/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:37:43 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:37:43 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:37:43 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0167
23/01/04 14:37:44 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:37:44 INFO JobSubmitter: number of splits:1
23/01/04 14:37:44 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0167
23/01/04 14:37:44 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:37:44 INFO Configuration: resource-types.xml not found
23/01/04 14:37:44 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:37:44 INFO YarnClientImpl: Submitted application
application_1672817468245_0167
23/01/04 14:37:44 INFO Job: The url to track the job:

```

http://lena-master:8088/proxy/application\_1672817468245\_0167/  
23/01/04 14:37:44 INFO Job: Running job: job\_1672817468245\_0167  
23/01/04 14:37:50 INFO Job: Job job\_1672817468245\_0167 running in uber mode : false  
23/01/04 14:37:50 INFO Job: map 0% reduce 0%  
23/01/04 14:37:55 INFO Job: map 100% reduce 0%  
23/01/04 14:38:01 INFO Job: map 100% reduce 100%  
23/01/04 14:38:01 INFO Job: Job job\_1672817468245\_0167 completed successfully  
23/01/04 14:38:01 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=314693  
FILE: Number of bytes written=1161957  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4354822  
HDFS: Number of bytes written=1151649  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16620  
Total time spent by all reduces in occupied slots (ms)=16570  
Total time spent by all map tasks (ms)=3324  
Total time spent by all reduce tasks (ms)=3314  
Total vcore-milliseconds taken by all map tasks=3324  
Total vcore-milliseconds taken by all reduce tasks=3314  
Total megabyte-milliseconds taken by all map tasks=17018880  
Total megabyte-milliseconds taken by all reduce tasks=16967680

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=314668  
Map output materialized bytes=314693  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=314693  
Reduce input records=4  
Reduce output records=8  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=35  
CPU time spent (ms)=2680  
Physical memory (bytes) snapshot=698200064  
Virtual memory (bytes) snapshot=12758622208  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408150016  
Peak Map Virtual memory (bytes)=6376054784  
Peak Reduce Physical memory (bytes)=290050048  
Peak Reduce Virtual memory (bytes)=6384521216

#### 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=2680740
File Output Format Counters
    Bytes Written=1151649
23/01/04 14:38:01 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 14:38:01 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:38:01 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0168
23/01/04 14:38:02 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:38:02 INFO JobSubmitter: number of splits:1
23/01/04 14:38:02 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0168
23/01/04 14:38:02 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:38:02 INFO YarnClientImpl: Submitted application
application_1672817468245_0168
23/01/04 14:38:02 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0168/
23/01/04 14:38:02 INFO Job: Running job: job_1672817468245_0168
23/01/04 14:38:08 INFO Job: Job job_1672817468245_0168 running in uber mode : false
23/01/04 14:38:08 INFO Job:  map 0% reduce 0%
23/01/04 14:38:14 INFO Job:  map 100% reduce 0%
23/01/04 14:38:19 INFO Job:  map 100% reduce 100%
23/01/04 14:38:20 INFO Job: Job job_1672817468245_0168 completed successfully
23/01/04 14:38:20 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=412712
    FILE: Number of bytes written=1357995
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4984238
    HDFS: Number of bytes written=1564375
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=16905
    Total time spent by all reduces in occupied slots (ms)=16460
    Total time spent by all map tasks (ms)=3381
    Total time spent by all reduce tasks (ms)=3292
    Total vcore-milliseconds taken by all map tasks=3381
    Total vcore-milliseconds taken by all reduce tasks=3292
    Total megabyte-milliseconds taken by all map tasks=17310720
    Total megabyte-milliseconds taken by all reduce tasks=16855040
Map-Reduce Framework
    Map input records=27
    Map output records=4
    Map output bytes=412686

```

Map output materialized bytes=412712  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=412712  
Reduce input records=4  
Reduce output records=12  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=56  
CPU time spent (ms)=2690  
Physical memory (bytes) snapshot=681132032  
Virtual memory (bytes) snapshot=12754239488  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390406144  
Peak Map Virtual memory (bytes)=6367686656  
Peak Reduce Physical memory (bytes)=290725888  
Peak Reduce Virtual memory (bytes)=6392803328

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=2680740

File Output Format Counters

Bytes Written=1564375

23/01/04 14:38:20 INFO RepresentativePointsDriver: Representative Points Iteration  
2

23/01/04 14:38:20 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:38:20 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0169

23/01/04 14:38:20 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:38:20 INFO JobSubmitter: number of splits:1

23/01/04 14:38:20 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0169

23/01/04 14:38:20 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:38:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0169

23/01/04 14:38:20 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0169/](http://lena-master:8088/proxy/application_1672817468245_0169/)

23/01/04 14:38:20 INFO Job: Running job: job\_1672817468245\_0169

23/01/04 14:38:26 INFO Job: Job job\_1672817468245\_0169 running in uber mode : false

23/01/04 14:38:26 INFO Job: map 0% reduce 0%

23/01/04 14:38:32 INFO Job: map 100% reduce 0%

23/01/04 14:38:38 INFO Job: map 100% reduce 100%

23/01/04 14:38:38 INFO Job: Job job\_1672817468245\_0169 completed successfully

23/01/04 14:38:38 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=431717  
FILE: Number of bytes written=1396005  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=5809690  
HDFS: Number of bytes written=1996147  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16775  
Total time spent by all reduces in occupied slots (ms)=16450  
Total time spent by all map tasks (ms)=3355  
Total time spent by all reduce tasks (ms)=3290  
Total vcore-milliseconds taken by all map tasks=3355  
Total vcore-milliseconds taken by all reduce tasks=3290  
Total megabyte-milliseconds taken by all map tasks=17177600  
Total megabyte-milliseconds taken by all reduce tasks=16844800

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=431692  
Map output materialized bytes=431717  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=431717  
Reduce input records=4  
Reduce output records=16  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=35  
CPU time spent (ms)=2900  
Physical memory (bytes) snapshot=687292416  
Virtual memory (bytes) snapshot=12762423296  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394854400  
Peak Map Virtual memory (bytes)=6376431616  
Peak Reduce Physical memory (bytes)=292438016  
Peak Reduce Virtual memory (bytes)=6387945472

#### 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=2680740

#### File Output Format Counters

Bytes Written=1996147

23/01/04 14:38:38 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 14:38:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:38:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0170  
23/01/04 14:38:39 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:38:39 INFO JobSubmitter: number of splits:1  
23/01/04 14:38:39 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0170  
23/01/04 14:38:39 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:38:39 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0170  
23/01/04 14:38:39 INFO Job: The url to track the job:  
http://lana-master:8088/proxy/application\_1672817468245\_0170/  
23/01/04 14:38:39 INFO Job: Running job: job\_1672817468245\_0170  
23/01/04 14:38:45 INFO Job: Job job\_1672817468245\_0170 running in uber mode : false  
23/01/04 14:38:45 INFO Job: map 0% reduce 0%  
23/01/04 14:38:51 INFO Job: map 100% reduce 0%  
23/01/04 14:38:56 INFO Job: map 100% reduce 100%  
23/01/04 14:38:57 INFO Job: Job job\_1672817468245\_0170 completed successfully  
23/01/04 14:38:57 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=346913  
FILE: Number of bytes written=1226397  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6673234  
HDFS: Number of bytes written=2343075  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17030  
Total time spent by all reduces in occupied slots (ms)=16475  
Total time spent by all map tasks (ms)=3406  
Total time spent by all reduce tasks (ms)=3295  
Total vcore-milliseconds taken by all map tasks=3406  
Total vcore-milliseconds taken by all reduce tasks=3295  
Total megabyte-milliseconds taken by all map tasks=17438720  
Total megabyte-milliseconds taken by all reduce tasks=16870400

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=346888  
Map output materialized bytes=346913  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=346913  
Reduce input records=4  
Reduce output records=20  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52

CPU time spent (ms)=2680  
Physical memory (bytes) snapshot=692686848  
Virtual memory (bytes) snapshot=12754501632  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397078528  
Peak Map Virtual memory (bytes)=6372749312  
Peak Reduce Physical memory (bytes)=295608320  
Peak Reduce Virtual memory (bytes)=6385659904

#### 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=2680740

#### File Output Format Counters

Bytes Written=2343075

23/01/04 14:38:57 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:38:57 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:38:57 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0171

23/01/04 14:38:57 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:38:57 INFO JobSubmitter: number of splits:1

23/01/04 14:38:57 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0171

23/01/04 14:38:57 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:38:57 INFO YarnClientImpl: Submitted application application\_1672817468245\_0171

23/01/04 14:38:57 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0171/

23/01/04 14:38:57 INFO Job: Running job: job\_1672817468245\_0171

23/01/04 14:39:03 INFO Job: Job job\_1672817468245\_0171 running in uber mode : false

23/01/04 14:39:03 INFO Job: map 0% reduce 0%

23/01/04 14:39:09 INFO Job: map 100% reduce 0%

23/01/04 14:39:15 INFO Job: map 100% reduce 100%

23/01/04 14:39:15 INFO Job: Job job\_1672817468245\_0171 completed successfully

23/01/04 14:39:15 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=354192  
FILE: Number of bytes written=1240955  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7367090  
HDFS: Number of bytes written=2697301  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17225  
Total time spent by all reduces in occupied slots (ms)=16440

Total time spent by all map tasks (ms)=3445  
Total time spent by all reduce tasks (ms)=3288  
Total vcore-milliseconds taken by all map tasks=3445  
Total vcore-milliseconds taken by all reduce tasks=3288  
Total megabyte-milliseconds taken by all map tasks=17638400  
Total megabyte-milliseconds taken by all reduce tasks=16834560

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=354166  
Map output materialized bytes=354192  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=354192  
Reduce input records=4  
Reduce output records=24  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=3010  
Physical memory (bytes) snapshot=689733632  
Virtual memory (bytes) snapshot=12751429632  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392830976  
Peak Map Virtual memory (bytes)=6370258944  
Peak Reduce Physical memory (bytes)=296902656  
Peak Reduce Virtual memory (bytes)=6385078272

#### 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=2680740

#### File Output Format Counters

Bytes Written=2697301

23/01/04 14:39:16 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.6232749870842145

23/01/04 14:39:16 INFO ClusterEvaluator: Intra-Cluster Density[17] =  
0.7058771257825506

23/01/04 14:39:16 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN

23/01/04 14:39:16 INFO ClusterEvaluator: Intra-Cluster Density[8] =  
0.5957545946315335

23/01/04 14:39:16 INFO ClusterEvaluator: Intra-Cluster Density[19] =  
0.6375025448658466

23/01/04 14:39:16 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.6463780884266436

23/01/04 14:39:16 INFO ClusterDumper: Wrote 4 clusters

23/01/04 14:39:16 INFO MahoutDriver: Program took 94555 ms (Minutes:  
1.5759166666666666)

---- K: 05 --

dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/05  
MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.



```

Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:39:20 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[5],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/05], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:39:21 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/05
23/01/04 14:39:21 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:39:21 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:39:21 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:39:21 INFO RandomSeedGenerator: Wrote 5 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:39:21 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/05
23/01/04 14:39:21 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:39:21 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:39:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:39:22 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0172
23/01/04 14:39:22 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:39:22 INFO JobSubmitter: number of splits:1
23/01/04 14:39:22 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0172
23/01/04 14:39:22 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:39:22 INFO Configuration: resource-types.xml not found
23/01/04 14:39:22 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:39:22 INFO YarnClientImpl: Submitted application
application_1672817468245_0172
23/01/04 14:39:22 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0172/
23/01/04 14:39:22 INFO Job: Running job: job_1672817468245_0172
23/01/04 14:39:29 INFO Job: Job job_1672817468245_0172 running in uber mode : false
23/01/04 14:39:29 INFO Job: map 0% reduce 0%
23/01/04 14:39:34 INFO Job: map 100% reduce 0%
23/01/04 14:39:40 INFO Job: map 100% reduce 100%
23/01/04 14:39:40 INFO Job: Job job_1672817468245_0172 completed successfully
23/01/04 14:39:40 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=1992297
        FILE: Number of bytes written=4516867
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0

```

HDFS: Number of bytes read=5481795  
HDFS: Number of bytes written=1335501  
HDFS: Number of read operations=35  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17865  
Total time spent by all reduces in occupied slots (ms)=17170  
Total time spent by all map tasks (ms)=3573  
Total time spent by all reduce tasks (ms)=3434  
Total vcore-milliseconds taken by all map tasks=3573  
Total vcore-milliseconds taken by all reduce tasks=3434  
Total megabyte-milliseconds taken by all map tasks=18293760  
Total megabyte-milliseconds taken by all reduce tasks=17582080

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=1992266  
Map output materialized bytes=1992297  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=1992297  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=4070  
Physical memory (bytes) snapshot=716972032  
Virtual memory (bytes) snapshot=12755861504  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396337152  
Peak Map Virtual memory (bytes)=6371749888  
Peak Reduce Physical memory (bytes)=320634880  
Peak Reduce Virtual memory (bytes)=6384111616

#### 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=2680027

#### File Output Format Counters

Bytes Written=1335501

23/01/04 14:39:40 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:39:40 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0173  
23/01/04 14:39:40 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:39:40 INFO JobSubmitter: number of splits:1  
23/01/04 14:39:40 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0173  
23/01/04 14:39:40 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:39:40 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0173  
23/01/04 14:39:40 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0173/  
23/01/04 14:39:40 INFO Job: Running job: job\_1672817468245\_0173  
23/01/04 14:39:46 INFO Job: Job job\_1672817468245\_0173 running in uber mode : false  
23/01/04 14:39:46 INFO Job: map 0% reduce 0%  
23/01/04 14:39:52 INFO Job: map 100% reduce 0%  
23/01/04 14:39:56 INFO Job: map 100% reduce 100%  
23/01/04 14:39:56 INFO Job: Job job\_1672817468245\_0173 completed successfully  
23/01/04 14:39:56 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2860147  
FILE: Number of bytes written=6252567  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5351587  
HDFS: Number of bytes written=1335501  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17685  
Total time spent by all reduces in occupied slots (ms)=9855  
Total time spent by all map tasks (ms)=3537  
Total time spent by all reduce tasks (ms)=1971  
Total vcore-milliseconds taken by all map tasks=3537  
Total vcore-milliseconds taken by all reduce tasks=1971  
Total megabyte-milliseconds taken by all map tasks=18109440  
Total megabyte-milliseconds taken by all reduce tasks=10091520

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=2860116  
Map output materialized bytes=2860147  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=2860147  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=4010  
Physical memory (bytes) snapshot=764502016  
Virtual memory (bytes) snapshot=12758507520

Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=416579584  
Peak Map Virtual memory (bytes)=6374797312  
Peak Reduce Physical memory (bytes)=347922432  
Peak Reduce Virtual memory (bytes)=6383710208

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=2680027

File Output Format Counters

Bytes Written=1335501

23/01/04 14:39:56 INFO KMeansDriver: Clustering data  
23/01/04 14:39:56 INFO KMeansDriver: Running Clustering  
23/01/04 14:39:56 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/05 Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/05  
23/01/04 14:39:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:39:56 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0174  
23/01/04 14:39:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:39:57 INFO JobSubmitter: number of splits:1  
23/01/04 14:39:57 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0174  
23/01/04 14:39:57 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:39:57 INFO YarnClientImpl: Submitted application application\_1672817468245\_0174  
23/01/04 14:39:57 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0174/  
23/01/04 14:39:57 INFO Job: Running job: job\_1672817468245\_0174  
23/01/04 14:40:03 INFO Job: Job job\_1672817468245\_0174 running in uber mode : false  
23/01/04 14:40:03 INFO Job: map 0% reduce 0%  
23/01/04 14:40:08 INFO Job: map 100% reduce 0%  
23/01/04 14:40:09 INFO Job: Job job\_1672817468245\_0174 completed successfully  
23/01/04 14:40:09 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265888  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4015892  
HDFS: Number of bytes written=2680743  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17425  
Total time spent by all reduces in occupied slots (ms)=0

```

Total time spent by all map tasks (ms)=3485
Total vcore-milliseconds taken by all map tasks=3485
Total megabyte-milliseconds taken by all map tasks=17843200
Map-Reduce Framework
  Map input records=27
  Map output records=27
  Input split bytes=170
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=23
  CPU time spent (ms)=2030
  Physical memory (bytes) snapshot=301531136
  Virtual memory (bytes) snapshot=6380027904
  Total committed heap usage (bytes)=1052770304
  Peak Map Physical memory (bytes)=301531136
  Peak Map Virtual memory (bytes)=6380027904
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680743
23/01/04 14:40:09 INFO MahoutDriver: Program took 48611 ms (Minutes:
0.81018333333333334)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:40:15 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/CosineDistanceMeasure/05/clusters-2-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
CosineDistanceMeasure/05.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/
05/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:40:16 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:40:16 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/05/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/05/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:40:16 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:40:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:40:17 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0175
23/01/04 14:40:17 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:40:17 INFO JobSubmitter: number of splits:1
23/01/04 14:40:17 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0175
23/01/04 14:40:17 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:40:17 INFO Configuration: resource-types.xml not found
23/01/04 14:40:17 INFO ResourceUtils: Unable to find 'resource-types.xml'.

```

23/01/04 14:40:17 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0175  
23/01/04 14:40:17 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0175/](http://lena-master:8088/proxy/application_1672817468245_0175/)  
23/01/04 14:40:17 INFO Job: Running job: job\_1672817468245\_0175  
23/01/04 14:40:27 INFO Job: Job job\_1672817468245\_0175 running in uber mode : false  
23/01/04 14:40:27 INFO Job: map 0% reduce 0%  
23/01/04 14:40:33 INFO Job: map 100% reduce 0%  
23/01/04 14:40:39 INFO Job: map 100% reduce 100%  
23/01/04 14:40:39 INFO Job: Job job\_1672817468245\_0175 completed successfully  
23/01/04 14:40:39 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=352102  
FILE: Number of bytes written=1236775  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4206195  
HDFS: Number of bytes written=1114759  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16250  
Total time spent by all reduces in occupied slots (ms)=16190  
Total time spent by all map tasks (ms)=3250  
Total time spent by all reduce tasks (ms)=3238  
Total vcore-milliseconds taken by all map tasks=3250  
Total vcore-milliseconds taken by all reduce tasks=3238  
Total megabyte-milliseconds taken by all map tasks=16640000  
Total megabyte-milliseconds taken by all reduce tasks=16578560

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=352073  
Map output materialized bytes=352102  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=352102  
Reduce input records=5  
Reduce output records=10  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=37  
CPU time spent (ms)=2600  
Physical memory (bytes) snapshot=674746368  
Virtual memory (bytes) snapshot=12754595840  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=389267456  
Peak Map Virtual memory (bytes)=6377193472  
Peak Reduce Physical memory (bytes)=285478912

Peak Reduce Virtual memory (bytes)=6379356160  
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=2680743

File Output Format Counters

Bytes Written=1114759

23/01/04 14:40:39 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 14:40:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:40:39 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0176

23/01/04 14:40:39 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:40:39 INFO JobSubmitter: number of splits:1

23/01/04 14:40:39 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0176

23/01/04 14:40:39 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:40:39 INFO YarnClientImpl: Submitted application application\_1672817468245\_0176

23/01/04 14:40:39 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0176/

23/01/04 14:40:39 INFO Job: Running job: job\_1672817468245\_0176

23/01/04 14:40:45 INFO Job: Job job\_1672817468245\_0176 running in uber mode : false

23/01/04 14:40:45 INFO Job: map 0% reduce 0%

23/01/04 14:40:50 INFO Job: map 100% reduce 0%

23/01/04 14:40:56 INFO Job: map 100% reduce 100%

23/01/04 14:40:56 INFO Job: Job job\_1672817468245\_0176 completed successfully

23/01/04 14:40:56 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=431417

FILE: Number of bytes written=1395405

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4910461

HDFS: Number of bytes written=1546186

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=16665

Total time spent by all reduces in occupied slots (ms)=16475

Total time spent by all map tasks (ms)=3333

Total time spent by all reduce tasks (ms)=3295

Total vcore-milliseconds taken by all map tasks=3333

Total vcore-milliseconds taken by all reduce tasks=3295

Total megabyte-milliseconds taken by all map tasks=17064960

Total megabyte-milliseconds taken by all reduce tasks=16870400

Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=431387  
Map output materialized bytes=431417  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=431417  
Reduce input records=5  
Reduce output records=15  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=2750  
Physical memory (bytes) snapshot=676454400  
Virtual memory (bytes) snapshot=12754247680  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394498048  
Peak Map Virtual memory (bytes)=6373183488  
Peak Reduce Physical memory (bytes)=281956352  
Peak Reduce Virtual memory (bytes)=6390534144

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=2680743

File Output Format Counters

Bytes Written=1546186

23/01/04 14:40:56 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 14:40:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:40:56 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0177  
23/01/04 14:40:56 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:40:57 INFO JobSubmitter: number of splits:1  
23/01/04 14:40:57 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0177  
23/01/04 14:40:57 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:40:57 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0177  
23/01/04 14:40:57 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0177/  
23/01/04 14:40:57 INFO Job: Running job: job\_1672817468245\_0177  
23/01/04 14:41:03 INFO Job: Job job\_1672817468245\_0177 running in uber mode : false  
23/01/04 14:41:03 INFO Job: map 0% reduce 0%  
23/01/04 14:41:09 INFO Job: map 100% reduce 0%  
23/01/04 14:41:14 INFO Job: map 100% reduce 100%  
23/01/04 14:41:15 INFO Job: Job job\_1672817468245\_0177 completed successfully  
23/01/04 14:41:15 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=376081



FILE: Number of bytes written=1284733  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5773315  
HDFS: Number of bytes written=1922298  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16775  
Total time spent by all reduces in occupied slots (ms)=16420  
Total time spent by all map tasks (ms)=3355  
Total time spent by all reduce tasks (ms)=3284  
Total vcore-milliseconds taken by all map tasks=3355  
Total vcore-milliseconds taken by all reduce tasks=3284  
Total megabyte-milliseconds taken by all map tasks=17177600  
Total megabyte-milliseconds taken by all reduce tasks=16814080

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=376052  
Map output materialized bytes=376081  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=376081  
Reduce input records=5  
Reduce output records=20  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=2940  
Physical memory (bytes) snapshot=683266048  
Virtual memory (bytes) snapshot=12753358848  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=389074944  
Peak Map Virtual memory (bytes)=6373089280  
Peak Reduce Physical memory (bytes)=294191104  
Peak Reduce Virtual memory (bytes)=6384242688

#### 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=2680743

#### File Output Format Counters

Bytes Written=1922298

3

```
23/01/04 14:41:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:41:15 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0178
23/01/04 14:41:15 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:41:15 INFO JobSubmitter: number of splits:1
23/01/04 14:41:15 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0178
23/01/04 14:41:15 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:41:15 INFO YarnClientImpl: Submitted application
application_1672817468245_0178
23/01/04 14:41:15 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0178/
23/01/04 14:41:15 INFO Job: Running job: job_1672817468245_0178
23/01/04 14:41:21 INFO Job: Job job_1672817468245_0178 running in uber mode : false
23/01/04 14:41:21 INFO Job:  map 0% reduce 0%
23/01/04 14:41:26 INFO Job:  map 100% reduce 0%
23/01/04 14:41:32 INFO Job:  map 100% reduce 100%
23/01/04 14:41:32 INFO Job: Job job_1672817468245_0178 completed successfully
23/01/04 14:41:32 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=404033
FILE: Number of bytes written=1340637
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=6525539
HDFS: Number of bytes written=2326362
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16700
Total time spent by all reduces in occupied slots (ms)=16510
Total time spent by all map tasks (ms)=3340
Total time spent by all reduce tasks (ms)=3302
Total vcore-milliseconds taken by all map tasks=3340
Total vcore-milliseconds taken by all reduce tasks=3302
Total megabyte-milliseconds taken by all map tasks=17100800
Total megabyte-milliseconds taken by all reduce tasks=16906240
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=5
Map output bytes=404004
Map output materialized bytes=404033
Input split bytes=200
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=404033
Reduce input records=5
Reduce output records=25
Spilled Records=10
Shuffled Maps =1
```

Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=30  
CPU time spent (ms)=2900  
Physical memory (bytes) snapshot=688975872  
Virtual memory (bytes) snapshot=12747780096  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396238848  
Peak Map Virtual memory (bytes)=6368419840  
Peak Reduce Physical memory (bytes)=292737024  
Peak Reduce Virtual memory (bytes)=6383267840

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=2680743

File Output Format Counters

Bytes Written=2326362

23/01/04 14:41:32 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:41:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:41:32 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0179

23/01/04 14:41:33 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:41:33 INFO JobSubmitter: number of splits:1

23/01/04 14:41:33 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0179

23/01/04 14:41:33 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:41:33 INFO YarnClientImpl: Submitted application application\_1672817468245\_0179

23/01/04 14:41:33 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0179/

23/01/04 14:41:33 INFO Job: Running job: job\_1672817468245\_0179

23/01/04 14:41:39 INFO Job: Job job\_1672817468245\_0179 running in uber mode : false

23/01/04 14:41:39 INFO Job: map 0% reduce 0%

23/01/04 14:41:44 INFO Job: map 100% reduce 0%

23/01/04 14:41:50 INFO Job: map 100% reduce 100%

23/01/04 14:41:50 INFO Job: Job job\_1672817468245\_0179 completed successfully

23/01/04 14:41:50 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=369331  
FILE: Number of bytes written=1271233  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7333667  
HDFS: Number of bytes written=2695704  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1

Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16980  
Total time spent by all reduces in occupied slots (ms)=16640  
Total time spent by all map tasks (ms)=3396  
Total time spent by all reduce tasks (ms)=3328  
Total vcore-milliseconds taken by all map tasks=3396  
Total vcore-milliseconds taken by all reduce tasks=3328  
Total megabyte-milliseconds taken by all map tasks=17387520  
Total megabyte-milliseconds taken by all reduce tasks=17039360

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=369302  
Map output materialized bytes=369331  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=369331  
Reduce input records=5  
Reduce output records=30  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=33  
CPU time spent (ms)=3120  
Physical memory (bytes) snapshot=689418240  
Virtual memory (bytes) snapshot=12744916992  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396726272  
Peak Map Virtual memory (bytes)=6371577856  
Peak Reduce Physical memory (bytes)=292691968  
Peak Reduce Virtual memory (bytes)=6375292928

#### 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=2680743

#### File Output Format Counters

Bytes Written=2695704

23/01/04 14:41:50 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.6452505922259761

23/01/04 14:41:50 INFO ClusterEvaluator: Intra-Cluster Density[5] =  
0.6247324265428018

23/01/04 14:41:50 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN

23/01/04 14:41:50 INFO ClusterEvaluator: Intra-Cluster Density[2] =  
0.5000243138814683

23/01/04 14:41:50 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN

23/01/04 14:41:50 INFO ClusterEvaluator: Intra-Cluster Density[20] =  
0.5623327154284357

23/01/04 14:41:50 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.5623631519509019

23/01/04 14:41:51 INFO ClusterDumper: Wrote 5 clusters

23/01/04 14:41:51 INFO MahoutDriver: Program took 96146 ms (Minutes:

```

1.60243333333333334)
---- K: 06 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/06
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:41:55 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[6],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/06], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:41:56 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/06
23/01/04 14:41:56 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:41:56 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:41:56 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:41:56 INFO RandomSeedGenerator: Wrote 6 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:41:56 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/06
23/01/04 14:41:56 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:41:56 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:41:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:41:57 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0180
23/01/04 14:41:57 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:41:57 INFO JobSubmitter: number of splits:1
23/01/04 14:41:57 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0180
23/01/04 14:41:57 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:41:57 INFO Configuration: resource-types.xml not found
23/01/04 14:41:57 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:41:57 INFO YarnClientImpl: Submitted application
application_1672817468245_0180
23/01/04 14:41:58 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0180/
23/01/04 14:41:58 INFO Job: Running job: job_1672817468245_0180
23/01/04 14:42:04 INFO Job: Job job_1672817468245_0180 running in uber mode : false
23/01/04 14:42:04 INFO Job:  map 0% reduce 0%
23/01/04 14:42:10 INFO Job:  map 100% reduce 0%
23/01/04 14:42:16 INFO Job:  map 100% reduce 100%
23/01/04 14:42:16 INFO Job: Job job_1672817468245_0180 completed successfully
23/01/04 14:42:16 INFO Job: Counters: 54
      File System Counters
          FILE: Number of bytes read=2762865

```

FILE: Number of bytes written=6058003  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5537355  
HDFS: Number of bytes written=2108984  
HDFS: Number of read operations=39  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17985  
Total time spent by all reduces in occupied slots (ms)=17795  
Total time spent by all map tasks (ms)=3597  
Total time spent by all reduce tasks (ms)=3559  
Total vcore-milliseconds taken by all map tasks=3597  
Total vcore-milliseconds taken by all reduce tasks=3559  
Total megabyte-milliseconds taken by all map tasks=18416640  
Total megabyte-milliseconds taken by all reduce tasks=18222080

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=2762829  
Map output materialized bytes=2762865  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=2762865  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=4420  
Physical memory (bytes) snapshot=716533760  
Virtual memory (bytes) snapshot=12756299776  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=407289856  
Peak Map Virtual memory (bytes)=6377013248  
Peak Reduce Physical memory (bytes)=309243904  
Peak Reduce Virtual memory (bytes)=6379286528

#### 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=2680027

#### File Output Format Counters

Bytes Written=2108984

23/01/04 14:42:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:42:16 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0181  
23/01/04 14:42:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:42:16 INFO JobSubmitter: number of splits:1  
23/01/04 14:42:16 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0181  
23/01/04 14:42:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:42:16 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0181  
23/01/04 14:42:16 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0181/  
23/01/04 14:42:16 INFO Job: Running job: job\_1672817468245\_0181  
23/01/04 14:42:22 INFO Job: Job job\_1672817468245\_0181 running in uber mode : false  
23/01/04 14:42:22 INFO Job: map 0% reduce 0%  
23/01/04 14:42:27 INFO Job: map 100% reduce 0%  
23/01/04 14:42:33 INFO Job: map 100% reduce 100%  
23/01/04 14:42:34 INFO Job: Job job\_1672817468245\_0181 completed successfully  
23/01/04 14:42:34 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4367371  
FILE: Number of bytes written=9267015  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6898553  
HDFS: Number of bytes written=2081323  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17780  
Total time spent by all reduces in occupied slots (ms)=17765  
Total time spent by all map tasks (ms)=3556  
Total time spent by all reduce tasks (ms)=3553  
Total vcore-milliseconds taken by all map tasks=3556  
Total vcore-milliseconds taken by all reduce tasks=3553  
Total megabyte-milliseconds taken by all map tasks=18206720  
Total megabyte-milliseconds taken by all reduce tasks=18191360

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4367335  
Map output materialized bytes=4367371  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=4367371  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=52
CPU time spent (ms)=4100
Physical memory (bytes) snapshot=759996416
Virtual memory (bytes) snapshot=12751757312
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=412430336
Peak Map Virtual memory (bytes)=6365167616
Peak Reduce Physical memory (bytes)=347566080
Peak Reduce Virtual memory (bytes)=6386589696
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=2680027
File Output Format Counters
  Bytes Written=2081323
23/01/04 14:42:34 INFO KMeansDriver: Clustering data
23/01/04 14:42:34 INFO KMeansDriver: Running Clustering
23/01/04 14:42:34 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/06 Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-
fiction-corpus-kmeans/CosineDistanceMeasure/06
23/01/04 14:42:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:42:35 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0182
23/01/04 14:42:35 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:42:35 INFO JobSubmitter: number of splits:1
23/01/04 14:42:35 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0182
23/01/04 14:42:35 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:42:35 INFO YarnClientImpl: Submitted application
application_1672817468245_0182
23/01/04 14:42:35 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0182/
23/01/04 14:42:35 INFO Job: Running job: job_1672817468245_0182
23/01/04 14:42:41 INFO Job: Job job_1672817468245_0182 running in uber mode : false
23/01/04 14:42:41 INFO Job:  map 0% reduce 0%
23/01/04 14:42:47 INFO Job:  map 100% reduce 0%
23/01/04 14:42:47 INFO Job: Job job_1672817468245_0182 completed successfully
23/01/04 14:42:47 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265888
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4761714
  HDFS: Number of bytes written=2680752
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters

```



Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17755  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3551  
Total vcore-milliseconds taken by all map tasks=3551  
Total megabyte-milliseconds taken by all map tasks=18181120

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=34  
CPU time spent (ms)=2110  
Physical memory (bytes) snapshot=291033088  
Virtual memory (bytes) snapshot=6370639872  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=291033088  
Peak Map Virtual memory (bytes)=6370639872

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680752

23/01/04 14:42:47 INFO MahoutDriver: Program took 51517 ms (Minutes:  
0.8586166666666667)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:42:53 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/CosineDistanceMeasure/06/clusters-2-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
CosineDistanceMeasure/06.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/  
06/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:42:54 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:42:54 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/06/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647],  
--input=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/06/  
clusters-2-final], --maxIter=[5], --method=[mapreduce],  
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 14:42:54 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 14:42:54 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:42:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0183

23/01/04 14:42:55 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:42:55 INFO JobSubmitter: number of splits:1

23/01/04 14:42:55 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0183

23/01/04 14:42:55 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:42:55 INFO Configuration: resource-types.xml not found

23/01/04 14:42:55 INFO ResourceUtils: Unable to find 'resource-types.xml'.

23/01/04 14:42:55 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0183

23/01/04 14:42:55 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0183/](http://lena-master:8088/proxy/application_1672817468245_0183/)

23/01/04 14:42:55 INFO Job: Running job: job\_1672817468245\_0183

23/01/04 14:43:04 INFO Job: Job job\_1672817468245\_0183 running in uber mode : false

23/01/04 14:43:04 INFO Job: map 0% reduce 0%

23/01/04 14:43:09 INFO Job: map 100% reduce 0%

23/01/04 14:43:15 INFO Job: map 100% reduce 100%

23/01/04 14:43:15 INFO Job: Job job\_1672817468245\_0183 completed successfully

23/01/04 14:43:15 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=430069

FILE: Number of bytes written=1392709

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4940042

HDFS: Number of bytes written=1559660

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=16505

Total time spent by all reduces in occupied slots (ms)=16535

Total time spent by all map tasks (ms)=3301

Total time spent by all reduce tasks (ms)=3307

Total vcore-milliseconds taken by all map tasks=3301

Total vcore-milliseconds taken by all reduce tasks=3307

Total megabyte-milliseconds taken by all map tasks=16901120

Total megabyte-milliseconds taken by all reduce tasks=16931840

#### Map-Reduce Framework

Map input records=27

Map output records=6

Map output bytes=430035

Map output materialized bytes=430069

Input split bytes=200

Combine input records=0

Combine output records=0

Reduce input groups=6

Reduce shuffle bytes=430069

Reduce input records=6

Reduce output records=12

Spilled Records=12

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=27

CPU time spent (ms)=2690

Physical memory (bytes) snapshot=693092352

Virtual memory (bytes) snapshot=12752445440

Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397979648  
Peak Map Virtual memory (bytes)=6372339712  
Peak Reduce Physical memory (bytes)=295112704  
Peak Reduce Virtual memory (bytes)=6384021504

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=2680752

File Output Format Counters

Bytes Written=1559660

23/01/04 14:43:15 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 14:43:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:43:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0184

23/01/04 14:43:16 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:43:16 INFO JobSubmitter: number of splits:1

23/01/04 14:43:16 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0184

23/01/04 14:43:16 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:43:16 INFO YarnClientImpl: Submitted application application\_1672817468245\_0184

23/01/04 14:43:16 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0184/

23/01/04 14:43:16 INFO Job: Running job: job\_1672817468245\_0184

23/01/04 14:43:22 INFO Job: Job job\_1672817468245\_0184 running in uber mode : false

23/01/04 14:43:22 INFO Job: map 0% reduce 0%

23/01/04 14:43:28 INFO Job: map 100% reduce 0%

23/01/04 14:43:33 INFO Job: map 100% reduce 100%

23/01/04 14:43:34 INFO Job: Job job\_1672817468245\_0184 completed successfully

23/01/04 14:43:34 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=551027  
FILE: Number of bytes written=1634625  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5800272  
HDFS: Number of bytes written=2110691  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16875  
Total time spent by all reduces in occupied slots (ms)=16705  
Total time spent by all map tasks (ms)=3375  
Total time spent by all reduce tasks (ms)=3341  
Total vcore-milliseconds taken by all map tasks=3375

Total vcore-milliseconds taken by all reduce tasks=3341  
Total megabyte-milliseconds taken by all map tasks=17280000  
Total megabyte-milliseconds taken by all reduce tasks=17105920  
Map-Reduce Framework  
Map input records=27  
Map output records=6  
Map output bytes=550991  
Map output materialized bytes=551027  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=551027  
Reduce input records=6  
Reduce output records=18  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=2580  
Physical memory (bytes) snapshot=684924928  
Virtual memory (bytes) snapshot=12760043520  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395857920  
Peak Map Virtual memory (bytes)=6379671552  
Peak Reduce Physical memory (bytes)=289067008  
Peak Reduce Virtual memory (bytes)=6384279552  
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=2680752  
File Output Format Counters  
Bytes Written=2110691

23/01/04 14:43:34 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 14:43:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:43:34 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0185  
23/01/04 14:43:34 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:43:34 INFO JobSubmitter: number of splits:1  
23/01/04 14:43:34 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0185  
23/01/04 14:43:34 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:43:34 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0185  
23/01/04 14:43:34 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0185/  
23/01/04 14:43:34 INFO Job: Running job: job\_1672817468245\_0185  
23/01/04 14:43:40 INFO Job: Job job\_1672817468245\_0185 running in uber mode : false  
23/01/04 14:43:40 INFO Job: map 0% reduce 0%  
23/01/04 14:43:45 INFO Job: map 100% reduce 0%  
23/01/04 14:43:51 INFO Job: map 100% reduce 100%

23/01/04 14:43:51 INFO Job: Job job\_1672817468245\_0185 completed successfully

23/01/04 14:43:51 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=653953  
FILE: Number of bytes written=1840477  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6902334  
HDFS: Number of bytes written=2764708  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16770  
Total time spent by all reduces in occupied slots (ms)=16730  
Total time spent by all map tasks (ms)=3354  
Total time spent by all reduce tasks (ms)=3346  
Total vcore-milliseconds taken by all map tasks=3354  
Total vcore-milliseconds taken by all reduce tasks=3346  
Total megabyte-milliseconds taken by all map tasks=17172480  
Total megabyte-milliseconds taken by all reduce tasks=17131520

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=653917  
Map output materialized bytes=653953  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=653953  
Reduce input records=6  
Reduce output records=24  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=2950  
Physical memory (bytes) snapshot=687321088  
Virtual memory (bytes) snapshot=12751859712  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394248192  
Peak Map Virtual memory (bytes)=6376185856  
Peak Reduce Physical memory (bytes)=293072896  
Peak Reduce Virtual memory (bytes)=6375673856

#### 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=2680752  
File Output Format Counters  
Bytes Written=2764708  
23/01/04 14:43:51 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 14:43:52 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:43:52 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0186  
23/01/04 14:43:52 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:43:52 INFO JobSubmitter: number of splits:1  
23/01/04 14:43:52 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0186  
23/01/04 14:43:52 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:43:52 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0186  
23/01/04 14:43:52 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0186/  
23/01/04 14:43:52 INFO Job: Running job: job\_1672817468245\_0186  
23/01/04 14:43:58 INFO Job: Job job\_1672817468245\_0186 running in uber mode : false  
23/01/04 14:43:58 INFO Job: map 0% reduce 0%  
23/01/04 14:44:03 INFO Job: map 100% reduce 0%  
23/01/04 14:44:08 INFO Job: map 100% reduce 100%  
23/01/04 14:44:08 INFO Job: Job job\_1672817468245\_0186 completed successfully  
23/01/04 14:44:08 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=532399  
FILE: Number of bytes written=1597369  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8210368  
HDFS: Number of bytes written=3297132  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16795  
Total time spent by all reduces in occupied slots (ms)=8930  
Total time spent by all map tasks (ms)=3359  
Total time spent by all reduce tasks (ms)=1786  
Total vcore-milliseconds taken by all map tasks=3359  
Total vcore-milliseconds taken by all reduce tasks=1786  
Total megabyte-milliseconds taken by all map tasks=17198080  
Total megabyte-milliseconds taken by all reduce tasks=9144320

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=532364  
Map output materialized bytes=532399  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=532399

Reduce input records=6  
Reduce output records=30  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=2910  
Physical memory (bytes) snapshot=686362624  
Virtual memory (bytes) snapshot=12751560704  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390541312  
Peak Map Virtual memory (bytes)=6374862848  
Peak Reduce Physical memory (bytes)=295821312  
Peak Reduce Virtual memory (bytes)=6380605440

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=2680752

File Output Format Counters

Bytes Written=3297132

23/01/04 14:44:08 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:44:08 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:44:08 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0187

23/01/04 14:44:08 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:44:08 INFO JobSubmitter: number of splits:1

23/01/04 14:44:08 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0187

23/01/04 14:44:08 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:44:09 INFO YarnClientImpl: Submitted application application\_1672817468245\_0187

23/01/04 14:44:09 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0187/](http://lena-master:8088/proxy/application_1672817468245_0187/)

23/01/04 14:44:09 INFO Job: Running job: job\_1672817468245\_0187

23/01/04 14:44:15 INFO Job: Job job\_1672817468245\_0187 running in uber mode : false

23/01/04 14:44:15 INFO Job: map 0% reduce 0%

23/01/04 14:44:20 INFO Job: map 100% reduce 0%

23/01/04 14:44:26 INFO Job: map 100% reduce 100%

23/01/04 14:44:26 INFO Job: Job job\_1672817468245\_0187 completed successfully

23/01/04 14:44:26 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=545852  
FILE: Number of bytes written=1624275  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9275216  
HDFS: Number of bytes written=3843029  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16960  
Total time spent by all reduces in occupied slots (ms)=16545  
Total time spent by all map tasks (ms)=3392  
Total time spent by all reduce tasks (ms)=3309  
Total vcore-milliseconds taken by all map tasks=3392  
Total vcore-milliseconds taken by all reduce tasks=3309  
Total megabyte-milliseconds taken by all map tasks=17367040  
Total megabyte-milliseconds taken by all reduce tasks=16942080

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=545817  
Map output materialized bytes=545852  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=545852  
Reduce input records=6  
Reduce output records=36  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=3220  
Physical memory (bytes) snapshot=710348800  
Virtual memory (bytes) snapshot=12747849728  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=407072768  
Peak Map Virtual memory (bytes)=6368108544  
Peak Reduce Physical memory (bytes)=303276032  
Peak Reduce Virtual memory (bytes)=6383656960

#### 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=2680752

#### File Output Format Counters

Bytes Written=3843029

23/01/04 14:44:26 INFO ClusterEvaluator: Scaled Inter-Cluster Density = 0.591802327898766

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[19] = 0.6077541437052906

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[2] = 0.6189308846031941

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[24] = 0.6564650507099812



```

23/01/04 14:44:26 INFO ClusterEvaluator: Intra-Cluster Density[12] =
0.7223631084115724
23/01/04 14:44:26 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6513782968575096
23/01/04 14:44:26 INFO ClusterDumper: Wrote 6 clusters
23/01/04 14:44:26 INFO MahoutDriver: Program took 93864 ms (Minutes: 1.5644)
---- K: 07 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/07
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:44:31 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[7],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/07], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:44:32 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/07
23/01/04 14:44:32 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:44:32 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:44:32 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:44:32 INFO RandomSeedGenerator: Wrote 7 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:44:32 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/07
23/01/04 14:44:32 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:44:32 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:44:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:44:33 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0188
23/01/04 14:44:33 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:44:33 INFO JobSubmitter: number of splits:1
23/01/04 14:44:33 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0188
23/01/04 14:44:33 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:44:33 INFO Configuration: resource-types.xml not found
23/01/04 14:44:33 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:44:33 INFO YarnClientImpl: Submitted application
application_1672817468245_0188
23/01/04 14:44:33 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0188/
23/01/04 14:44:33 INFO Job: Running job: job_1672817468245_0188
23/01/04 14:44:39 INFO Job: Job job_1672817468245_0188 running in uber mode : false
23/01/04 14:44:39 INFO Job: map 0% reduce 0%
23/01/04 14:44:44 INFO Job: map 100% reduce 0%

```

23/01/04 14:44:51 INFO Job: map 100% reduce 100%  
23/01/04 14:44:51 INFO Job: Job job\_1672817468245\_0188 completed successfully  
23/01/04 14:44:51 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3370926  
FILE: Number of bytes written=7274125  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6866517  
HDFS: Number of bytes written=2549135  
HDFS: Number of read operations=43  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17780  
Total time spent by all reduces in occupied slots (ms)=17660  
Total time spent by all map tasks (ms)=3556  
Total time spent by all reduce tasks (ms)=3532  
Total vcore-milliseconds taken by all map tasks=3556  
Total vcore-milliseconds taken by all reduce tasks=3532  
Total megabyte-milliseconds taken by all map tasks=18206720  
Total megabyte-milliseconds taken by all reduce tasks=18083840

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3370885  
Map output materialized bytes=3370926  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3370926  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=33  
CPU time spent (ms)=4420  
Physical memory (bytes) snapshot=722358272  
Virtual memory (bytes) snapshot=12750483456  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397037568  
Peak Map Virtual memory (bytes)=6367719424  
Peak Reduce Physical memory (bytes)=325320704  
Peak Reduce Virtual memory (bytes)=6382764032

#### 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=2680027

#### File Output Format Counters

Bytes Written=2549135

23/01/04 14:44:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:44:51 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0189  
23/01/04 14:44:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:44:51 INFO JobSubmitter: number of splits:1  
23/01/04 14:44:51 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0189  
23/01/04 14:44:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:44:51 INFO YarnClientImpl: Submitted application application\_1672817468245\_0189  
23/01/04 14:44:51 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0189/  
23/01/04 14:44:51 INFO Job: Running job: job\_1672817468245\_0189  
23/01/04 14:44:57 INFO Job: Job job\_1672817468245\_0189 running in uber mode : false  
23/01/04 14:44:57 INFO Job: map 0% reduce 0%  
23/01/04 14:45:03 INFO Job: map 100% reduce 0%  
23/01/04 14:45:09 INFO Job: map 100% reduce 100%  
23/01/04 14:45:09 INFO Job: Job job\_1672817468245\_0189 completed successfully  
23/01/04 14:45:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5213907  
FILE: Number of bytes written=10960087  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7778855  
HDFS: Number of bytes written=2541737  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17740  
Total time spent by all reduces in occupied slots (ms)=17590  
Total time spent by all map tasks (ms)=3548  
Total time spent by all reduce tasks (ms)=3518  
Total vcore-milliseconds taken by all map tasks=3548  
Total vcore-milliseconds taken by all reduce tasks=3518  
Total megabyte-milliseconds taken by all map tasks=18165760  
Total megabyte-milliseconds taken by all reduce tasks=18012160

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=5213866  
Map output materialized bytes=5213907  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=5213907  
Reduce input records=7

```

Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=42
CPU time spent (ms)=4350
Physical memory (bytes) snapshot=742084608
Virtual memory (bytes) snapshot=12754370560
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=421629952
Peak Map Virtual memory (bytes)=6373171200
Peak Reduce Physical memory (bytes)=320454656
Peak Reduce Virtual memory (bytes)=6381199360

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=2680027
File Output Format Counters
  Bytes Written=2541737
23/01/04 14:45:09 INFO KMeansDriver: Clustering data
23/01/04 14:45:09 INFO KMeansDriver: Running Clustering
23/01/04 14:45:09 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/07 Out: hdfs://lena/user/bfezo001/dsm010/question_2/british-
fiction-corpus-kmeans/CosineDistanceMeasure/07
23/01/04 14:45:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:45:09 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0190
23/01/04 14:45:10 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:45:10 INFO JobSubmitter: number of splits:1
23/01/04 14:45:10 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0190
23/01/04 14:45:10 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:45:10 INFO YarnClientImpl: Submitted application
application_1672817468245_0190
23/01/04 14:45:10 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0190/
23/01/04 14:45:10 INFO Job: Running job: job_1672817468245_0190
23/01/04 14:45:16 INFO Job: Job job_1672817468245_0190 running in uber mode : false
23/01/04 14:45:16 INFO Job:  map 0% reduce 0%
23/01/04 14:45:22 INFO Job:  map 100% reduce 0%
23/01/04 14:45:22 INFO Job: Job job_1672817468245_0190 completed successfully
23/01/04 14:45:22 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265888
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5222128
  HDFS: Number of bytes written=2680763

```

HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17320  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3464  
Total vcore-milliseconds taken by all map tasks=3464  
Total megabyte-milliseconds taken by all map tasks=17735680

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=18  
CPU time spent (ms)=2240  
Physical memory (bytes) snapshot=298987520  
Virtual memory (bytes) snapshot=6376628224  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=298987520  
Peak Map Virtual memory (bytes)=6376628224

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680763

23/01/04 14:45:22 INFO MahoutDriver: Program took 50804 ms (Minutes:  
0.8467333333333333)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:45:27 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --  
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/CosineDistanceMeasure/07/clusters-2-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
CosineDistanceMeasure/07.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/  
07/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:45:29 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:45:29 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/07/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/07/  
clusters-2-final], --maxIter=[5], --method=[mapreduce],  
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 14:45:29 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 14:45:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:45:29 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0191  
23/01/04 14:45:30 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:45:30 INFO JobSubmitter: number of splits:1  
23/01/04 14:45:30 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0191  
23/01/04 14:45:30 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:45:30 INFO Configuration: resource-types.xml not found  
23/01/04 14:45:30 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:45:30 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0191  
23/01/04 14:45:30 INFO Job: The url to track the job:  
http://lana-master:8088/proxy/application\_1672817468245\_0191/  
23/01/04 14:45:30 INFO Job: Running job: job\_1672817468245\_0191  
23/01/04 14:45:36 INFO Job: Job job\_1672817468245\_0191 running in uber mode : false  
23/01/04 14:45:36 INFO Job: map 0% reduce 0%  
23/01/04 14:45:41 INFO Job: map 100% reduce 0%  
23/01/04 14:45:47 INFO Job: map 100% reduce 100%  
23/01/04 14:45:47 INFO Job: Job job\_1672817468245\_0191 completed successfully  
23/01/04 14:45:47 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=603161  
FILE: Number of bytes written=1738893  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5346655  
HDFS: Number of bytes written=1936047  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16625  
Total time spent by all reduces in occupied slots (ms)=16650  
Total time spent by all map tasks (ms)=3325  
Total time spent by all reduce tasks (ms)=3330  
Total vcore-milliseconds taken by all map tasks=3325  
Total vcore-milliseconds taken by all reduce tasks=3330  
Total megabyte-milliseconds taken by all map tasks=17024000  
Total megabyte-milliseconds taken by all reduce tasks=17049600

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=603121  
Map output materialized bytes=603161  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=603161  
Reduce input records=7  
Reduce output records=14  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0

Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=2520  
Physical memory (bytes) snapshot=682557440  
Virtual memory (bytes) snapshot=12755091456  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395771904  
Peak Map Virtual memory (bytes)=6372544512  
Peak Reduce Physical memory (bytes)=286785536  
Peak Reduce Virtual memory (bytes)=6391681024

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=2680763

File Output Format Counters

Bytes Written=1936047

23/01/04 14:45:47 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 14:45:47 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:45:47 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0192

23/01/04 14:45:48 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:45:48 INFO JobSubmitter: number of splits:1

23/01/04 14:45:48 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0192

23/01/04 14:45:48 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:45:48 INFO YarnClientImpl: Submitted application application\_1672817468245\_0192

23/01/04 14:45:48 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0192/

23/01/04 14:45:48 INFO Job: Running job: job\_1672817468245\_0192

23/01/04 14:45:54 INFO Job: Job job\_1672817468245\_0192 running in uber mode : false

23/01/04 14:45:54 INFO Job: map 0% reduce 0%

23/01/04 14:46:00 INFO Job: map 100% reduce 0%

23/01/04 14:46:04 INFO Job: map 100% reduce 100%

23/01/04 14:46:04 INFO Job: Job job\_1672817468245\_0192 completed successfully

23/01/04 14:46:04 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=702145  
FILE: Number of bytes written=1936861  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6553057  
HDFS: Number of bytes written=2638212  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=16410  
Total time spent by all reduces in occupied slots (ms)=9380  
Total time spent by all map tasks (ms)=3282  
Total time spent by all reduce tasks (ms)=1876  
Total vcore-milliseconds taken by all map tasks=3282  
Total vcore-milliseconds taken by all reduce tasks=1876  
Total megabyte-milliseconds taken by all map tasks=16803840  
Total megabyte-milliseconds taken by all reduce tasks=9605120

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=702105  
Map output materialized bytes=702145  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=702145  
Reduce input records=7  
Reduce output records=21  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=3150  
Physical memory (bytes) snapshot=704548864  
Virtual memory (bytes) snapshot=12749451264  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395304960  
Peak Map Virtual memory (bytes)=6367342592  
Peak Reduce Physical memory (bytes)=309243904  
Peak Reduce Virtual memory (bytes)=6386016256

#### 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=2680763

#### File Output Format Counters

Bytes Written=2638212

23/01/04 14:46:04 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 14:46:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:46:04 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0193

23/01/04 14:46:04 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:46:04 INFO JobSubmitter: number of splits:1

23/01/04 14:46:04 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0193

23/01/04 14:46:04 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:46:04 INFO YarnClientImpl: Submitted application application\_1672817468245\_0193

23/01/04 14:46:04 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0193/](http://lena-master:8088/proxy/application_1672817468245_0193/)



23/01/04 14:46:04 INFO Job: Running job: job\_1672817468245\_0193  
23/01/04 14:46:10 INFO Job: Job job\_1672817468245\_0193 running in uber mode : false  
23/01/04 14:46:10 INFO Job: map 0% reduce 0%  
23/01/04 14:46:16 INFO Job: map 100% reduce 0%  
23/01/04 14:46:21 INFO Job: map 100% reduce 100%  
23/01/04 14:46:22 INFO Job: Job job\_1672817468245\_0193 completed successfully  
23/01/04 14:46:22 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=692351  
FILE: Number of bytes written=1917273  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7957387  
HDFS: Number of bytes written=3330662  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16780  
Total time spent by all reduces in occupied slots (ms)=16590  
Total time spent by all map tasks (ms)=3356  
Total time spent by all reduce tasks (ms)=3318  
Total vcore-milliseconds taken by all map tasks=3356  
Total vcore-milliseconds taken by all reduce tasks=3318  
Total megabyte-milliseconds taken by all map tasks=17182720  
Total megabyte-milliseconds taken by all reduce tasks=16988160

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=692310  
Map output materialized bytes=692351  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=692351  
Reduce input records=7  
Reduce output records=28  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=35  
CPU time spent (ms)=2940  
Physical memory (bytes) snapshot=698839040  
Virtual memory (bytes) snapshot=12761653248  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=389464064  
Peak Map Virtual memory (bytes)=6375976960  
Peak Reduce Physical memory (bytes)=309374976  
Peak Reduce Virtual memory (bytes)=6385676288

#### 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=2680763
File Output Format Counters
    Bytes Written=3330662
23/01/04 14:46:22 INFO RepresentativePointsDriver: Representative Points Iteration
3
23/01/04 14:46:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:46:22 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0194
23/01/04 14:46:23 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:46:23 INFO JobSubmitter: number of splits:1
23/01/04 14:46:23 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0194
23/01/04 14:46:23 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:46:23 INFO YarnClientImpl: Submitted application
application_1672817468245_0194
23/01/04 14:46:23 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0194/
23/01/04 14:46:23 INFO Job: Running job: job_1672817468245_0194
23/01/04 14:46:29 INFO Job: Job job_1672817468245_0194 running in uber mode : false
23/01/04 14:46:29 INFO Job:  map 0% reduce 0%
23/01/04 14:46:36 INFO Job:  map 100% reduce 0%
23/01/04 14:46:40 INFO Job:  map 100% reduce 100%
23/01/04 14:46:40 INFO Job: Job job_1672817468245_0194 completed successfully
23/01/04 14:46:40 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=692132
    FILE: Number of bytes written=1916835
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=9342287
    HDFS: Number of bytes written=4022835
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17005
    Total time spent by all reduces in occupied slots (ms)=9140
    Total time spent by all map tasks (ms)=3401
    Total time spent by all reduce tasks (ms)=1828
    Total vcore-milliseconds taken by all map tasks=3401
    Total vcore-milliseconds taken by all reduce tasks=1828
    Total megabyte-milliseconds taken by all map tasks=17413120
    Total megabyte-milliseconds taken by all reduce tasks=9359360
Map-Reduce Framework
    Map input records=27
    Map output records=7
    Map output bytes=692093
    Map output materialized bytes=692132

```

Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=692132  
Reduce input records=7  
Reduce output records=35  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=37  
CPU time spent (ms)=3110  
Physical memory (bytes) snapshot=694706176  
Virtual memory (bytes) snapshot=12761870336  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397643776  
Peak Map Virtual memory (bytes)=6381203456  
Peak Reduce Physical memory (bytes)=297062400  
Peak Reduce Virtual memory (bytes)=6380666880

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=2680763

File Output Format Counters

Bytes Written=4022835

23/01/04 14:46:40 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:46:40 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:46:40 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0195

23/01/04 14:46:40 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:46:40 INFO JobSubmitter: number of splits:1

23/01/04 14:46:40 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0195

23/01/04 14:46:40 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:46:40 INFO YarnClientImpl: Submitted application application\_1672817468245\_0195

23/01/04 14:46:40 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0195/

23/01/04 14:46:40 INFO Job: Running job: job\_1672817468245\_0195

23/01/04 14:46:46 INFO Job: Job job\_1672817468245\_0195 running in uber mode : false

23/01/04 14:46:46 INFO Job: map 0% reduce 0%

23/01/04 14:46:52 INFO Job: map 100% reduce 0%

23/01/04 14:46:57 INFO Job: map 100% reduce 100%

23/01/04 14:46:57 INFO Job: Job job\_1672817468245\_0195 completed successfully

23/01/04 14:46:58 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=681023  
FILE: Number of bytes written=1894617  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0

HDFS: Number of bytes read=10726633  
HDFS: Number of bytes written=4703918  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17000  
Total time spent by all reduces in occupied slots (ms)=16625  
Total time spent by all map tasks (ms)=3400  
Total time spent by all reduce tasks (ms)=3325  
Total vcore-milliseconds taken by all map tasks=3400  
Total vcore-milliseconds taken by all reduce tasks=3325  
Total megabyte-milliseconds taken by all map tasks=17408000  
Total megabyte-milliseconds taken by all reduce tasks=17024000

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=680983  
Map output materialized bytes=681023  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=681023  
Reduce input records=7  
Reduce output records=42  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3180  
Physical memory (bytes) snapshot=706531328  
Virtual memory (bytes) snapshot=12759154688  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413962240  
Peak Map Virtual memory (bytes)=6379200512  
Peak Reduce Physical memory (bytes)=292569088  
Peak Reduce Virtual memory (bytes)=6379954176

#### 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=2680763

#### File Output Format Counters

Bytes Written=4703918

23/01/04 14:46:58 INFO ClusterEvaluator: Scaled Inter-Cluster Density = 0.5969094940624287

23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[0] = 0.6859311010321779

23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN

```

23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[21] =
0.6904405695987956
23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[10] =
0.521851636455987
23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[16] =
0.5263580925425024
23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.629902413107075
23/01/04 14:46:58 INFO ClusterEvaluator: Intra-Cluster Density[23] =
0.6886952834671087
23/01/04 14:46:58 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6238631827006077
23/01/04 14:46:58 INFO ClusterDumper: Wrote 7 clusters
23/01/04 14:46:58 INFO MahoutDriver: Program took 90908 ms (Minutes:
1.5151333333333334)
---- K: 08 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/08
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:47:03 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[8],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/08], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:47:03 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/08
23/01/04 14:47:03 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:47:04 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:47:04 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:47:04 INFO RandomSeedGenerator: Wrote 8 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:47:04 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/08
23/01/04 14:47:04 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:47:04 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:47:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:47:05 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0196
23/01/04 14:47:05 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:47:05 INFO JobSubmitter: number of splits:1
23/01/04 14:47:05 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0196
23/01/04 14:47:05 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:47:05 INFO Configuration: resource-types.xml not found

```

23/01/04 14:47:05 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:47:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0196  
23/01/04 14:47:05 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0196/  
23/01/04 14:47:05 INFO Job: Running job: job\_1672817468245\_0196  
23/01/04 14:47:11 INFO Job: Job job\_1672817468245\_0196 running in uber mode : false  
23/01/04 14:47:11 INFO Job: map 0% reduce 0%  
23/01/04 14:47:17 INFO Job: map 100% reduce 0%  
23/01/04 14:47:21 INFO Job: map 100% reduce 100%  
23/01/04 14:47:21 INFO Job: Job job\_1672817468245\_0196 completed successfully  
23/01/04 14:47:21 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3400285  
FILE: Number of bytes written=7332843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7458351  
HDFS: Number of bytes written=2416399  
HDFS: Number of read operations=47  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18045  
Total time spent by all reduces in occupied slots (ms)=10665  
Total time spent by all map tasks (ms)=3609  
Total time spent by all reduce tasks (ms)=2133  
Total vcore-milliseconds taken by all map tasks=3609  
Total vcore-milliseconds taken by all reduce tasks=2133  
Total megabyte-milliseconds taken by all map tasks=18478080  
Total megabyte-milliseconds taken by all reduce tasks=10920960

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=3400239  
Map output materialized bytes=3400285  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=3400285  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=4580  
Physical memory (bytes) snapshot=746332160  
Virtual memory (bytes) snapshot=12763803648  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=421396480  
Peak Map Virtual memory (bytes)=6380609536

```

        Peak Reduce Physical memory (bytes)=324935680
        Peak Reduce Virtual memory (bytes)=6383194112
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=2680027
File Output Format Counters
    Bytes Written=2416399
23/01/04 14:47:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:47:22 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0197
23/01/04 14:47:22 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:47:22 INFO JobSubmitter: number of splits:1
23/01/04 14:47:22 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0197
23/01/04 14:47:22 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:47:22 INFO YarnClientImpl: Submitted application
application_1672817468245_0197
23/01/04 14:47:22 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0197/
23/01/04 14:47:22 INFO Job: Running job: job_1672817468245_0197
23/01/04 14:47:28 INFO Job: Job job_1672817468245_0197 running in uber mode : false
23/01/04 14:47:28 INFO Job:  map 0% reduce 0%
23/01/04 14:47:34 INFO Job:  map 100% reduce 0%
23/01/04 14:47:38 INFO Job:  map 100% reduce 100%
23/01/04 14:47:38 INFO Job: Job job_1672817468245_0197 completed successfully
23/01/04 14:47:38 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=5019106
    FILE: Number of bytes written=10570485
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7513383
    HDFS: Number of bytes written=2416399
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=18070
    Total time spent by all reduces in occupied slots (ms)=10175
    Total time spent by all map tasks (ms)=3614
    Total time spent by all reduce tasks (ms)=2035
    Total vcore-milliseconds taken by all map tasks=3614
    Total vcore-milliseconds taken by all reduce tasks=2035
    Total megabyte-milliseconds taken by all map tasks=18503680
    Total megabyte-milliseconds taken by all reduce tasks=10419200
Map-Reduce Framework
    Map input records=27

```

Map output records=8  
Map output bytes=5019060  
Map output materialized bytes=5019106  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=5019106  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=4310  
Physical memory (bytes) snapshot=747753472  
Virtual memory (bytes) snapshot=12760662016  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419033088  
Peak Map Virtual memory (bytes)=6374907904  
Peak Reduce Physical memory (bytes)=328720384  
Peak Reduce Virtual memory (bytes)=6385754112

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=2680027

File Output Format Counters

Bytes Written=2416399

23/01/04 14:47:38 INFO KMeansDriver: Clustering data  
23/01/04 14:47:38 INFO KMeansDriver: Running Clustering  
23/01/04 14:47:38 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/08 Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/08  
23/01/04 14:47:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:47:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0199  
23/01/04 14:47:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:47:38 INFO JobSubmitter: number of splits:1  
23/01/04 14:47:38 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0199  
23/01/04 14:47:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:47:38 INFO YarnClientImpl: Submitted application application\_1672817468245\_0199  
23/01/04 14:47:38 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0199/  
23/01/04 14:47:38 INFO Job: Running job: job\_1672817468245\_0199  
23/01/04 14:47:45 INFO Job: Job job\_1672817468245\_0199 running in uber mode : false  
23/01/04 14:47:45 INFO Job: map 0% reduce 0%  
23/01/04 14:47:51 INFO Job: map 100% reduce 0%  
23/01/04 14:47:51 INFO Job: Job job\_1672817468245\_0199 completed successfully



23/01/04 14:47:51 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265888  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5096790  
HDFS: Number of bytes written=2680738  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17680  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3536  
Total vcore-milliseconds taken by all map tasks=3536  
Total megabyte-milliseconds taken by all map tasks=18104320

Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=21  
CPU time spent (ms)=1990  
Physical memory (bytes) snapshot=296804352  
Virtual memory (bytes) snapshot=6382710784  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=296804352  
Peak Map Virtual memory (bytes)=6382710784

File Input Format Counters

Bytes Read=2680027

File Output Format Counters

Bytes Written=2680738

23/01/04 14:47:51 INFO MahoutDriver: Program took 47802 ms (Minutes: 0.7967)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:47:56 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/CosineDistanceMeasure/08/clusters-2-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
CosineDistanceMeasure/08.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/  
08/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:47:58 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:47:58 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/08/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --

```
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/08/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:47:58 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:47:58 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:47:58 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0200
23/01/04 14:47:59 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:47:59 INFO JobSubmitter: number of splits:1
23/01/04 14:47:59 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0200
23/01/04 14:47:59 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:47:59 INFO Configuration: resource-types.xml not found
23/01/04 14:47:59 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:47:59 INFO YarnClientImpl: Submitted application
application_1672817468245_0200
23/01/04 14:47:59 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0200/
23/01/04 14:47:59 INFO Job: Running job: job_1672817468245_0200
23/01/04 14:48:05 INFO Job: Job job_1672817468245_0200 running in uber mode : false
23/01/04 14:48:05 INFO Job:  map 0% reduce 0%
23/01/04 14:48:10 INFO Job:  map 100% reduce 0%
23/01/04 14:48:16 INFO Job:  map 100% reduce 100%
23/01/04 14:48:16 INFO Job: Job job_1672817468245_0200 completed successfully
23/01/04 14:48:16 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=709510
FILE: Number of bytes written=1951591
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5284602
HDFS: Number of bytes written=2011397
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16385
Total time spent by all reduces in occupied slots (ms)=16465
Total time spent by all map tasks (ms)=3277
Total time spent by all reduce tasks (ms)=3293
Total vcore-milliseconds taken by all map tasks=3277
Total vcore-milliseconds taken by all reduce tasks=3293
Total megabyte-milliseconds taken by all map tasks=16778240
Total megabyte-milliseconds taken by all reduce tasks=16860160
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=8
Map output bytes=709465
Map output materialized bytes=709510
Input split bytes=200
Combine input records=0
```

Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=709510  
Reduce input records=8  
Reduce output records=16  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=50  
CPU time spent (ms)=2610  
Physical memory (bytes) snapshot=691273728  
Virtual memory (bytes) snapshot=12753711104  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398344192  
Peak Map Virtual memory (bytes)=6366613504  
Peak Reduce Physical memory (bytes)=292929536  
Peak Reduce Virtual memory (bytes)=6391005184  
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=2680738  
File Output Format Counters  
Bytes Written=2011397

23/01/04 14:48:16 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 14:48:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:48:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0201  
23/01/04 14:48:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:48:17 INFO JobSubmitter: number of splits:1  
23/01/04 14:48:17 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0201  
23/01/04 14:48:17 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:48:17 INFO YarnClientImpl: Submitted application application\_1672817468245\_0201  
23/01/04 14:48:17 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0201/  
23/01/04 14:48:17 INFO Job: Running job: job\_1672817468245\_0201  
23/01/04 14:48:23 INFO Job: Job job\_1672817468245\_0201 running in uber mode : false  
23/01/04 14:48:23 INFO Job: map 0% reduce 0%  
23/01/04 14:48:29 INFO Job: map 100% reduce 0%  
23/01/04 14:48:34 INFO Job: map 100% reduce 100%  
23/01/04 14:48:35 INFO Job: Job job\_1672817468245\_0201 completed successfully  
23/01/04 14:48:35 INFO Job: Counters: 54  
File System Counters  
FILE: Number of bytes read=753610  
FILE: Number of bytes written=2039791  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6703732  
HDFS: Number of bytes written=2765003

HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16600  
Total time spent by all reduces in occupied slots (ms)=16245  
Total time spent by all map tasks (ms)=3320  
Total time spent by all reduce tasks (ms)=3249  
Total vcore-milliseconds taken by all map tasks=3320  
Total vcore-milliseconds taken by all reduce tasks=3249  
Total megabyte-milliseconds taken by all map tasks=16998400  
Total megabyte-milliseconds taken by all reduce tasks=16634880

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=753566  
Map output materialized bytes=753610  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=753610  
Reduce input records=8  
Reduce output records=24  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=24  
CPU time spent (ms)=2890  
Physical memory (bytes) snapshot=704860160  
Virtual memory (bytes) snapshot=12745834496  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411172864  
Peak Map Virtual memory (bytes)=6369796096  
Peak Reduce Physical memory (bytes)=293687296  
Peak Reduce Virtual memory (bytes)=6376038400

#### 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=2680738

#### File Output Format Counters

Bytes Written=2765003

23/01/04 14:48:35 INFO RepresentativePointsDriver: Representative Points Iteration  
2

23/01/04 14:48:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:48:35 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0202

23/01/04 14:48:35 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:48:35 INFO JobSubmitter: number of splits:1  
23/01/04 14:48:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0202  
23/01/04 14:48:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:48:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0202  
23/01/04 14:48:35 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0202/  
23/01/04 14:48:35 INFO Job: Running job: job\_1672817468245\_0202  
23/01/04 14:48:41 INFO Job: Job job\_1672817468245\_0202 running in uber mode : false  
23/01/04 14:48:41 INFO Job: map 0% reduce 0%  
23/01/04 14:48:47 INFO Job: map 100% reduce 0%  
23/01/04 14:48:52 INFO Job: map 100% reduce 100%  
23/01/04 14:48:53 INFO Job: Job job\_1672817468245\_0202 completed successfully  
23/01/04 14:48:53 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=729095  
FILE: Number of bytes written=1990761  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8210944  
HDFS: Number of bytes written=3494213  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17190  
Total time spent by all reduces in occupied slots (ms)=16635  
Total time spent by all map tasks (ms)=3438  
Total time spent by all reduce tasks (ms)=3327  
Total vcore-milliseconds taken by all map tasks=3438  
Total vcore-milliseconds taken by all reduce tasks=3327  
Total megabyte-milliseconds taken by all map tasks=17602560  
Total megabyte-milliseconds taken by all reduce tasks=17034240

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=729050  
Map output materialized bytes=729095  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=729095  
Reduce input records=8  
Reduce output records=32  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=3010  
Physical memory (bytes) snapshot=671793152  
Virtual memory (bytes) snapshot=12759343104

Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394244096  
Peak Map Virtual memory (bytes)=6379855872  
Peak Reduce Physical memory (bytes)=281309184  
Peak Reduce Virtual memory (bytes)=6383394816

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=2680738

File Output Format Counters

Bytes Written=3494213

23/01/04 14:48:53 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 14:48:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:48:53 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0203

23/01/04 14:48:54 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:48:54 INFO JobSubmitter: number of splits:1

23/01/04 14:48:54 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0203

23/01/04 14:48:54 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:48:54 INFO YarnClientImpl: Submitted application application\_1672817468245\_0203

23/01/04 14:48:54 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0203/

23/01/04 14:48:54 INFO Job: Running job: job\_1672817468245\_0203

23/01/04 14:49:00 INFO Job: Job job\_1672817468245\_0203 running in uber mode : false

23/01/04 14:49:00 INFO Job: map 0% reduce 0%

23/01/04 14:49:06 INFO Job: map 100% reduce 0%

23/01/04 14:49:11 INFO Job: map 100% reduce 100%

23/01/04 14:49:12 INFO Job: Job job\_1672817468245\_0203 completed successfully

23/01/04 14:49:12 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=809194  
FILE: Number of bytes written=2150959  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9669364  
HDFS: Number of bytes written=4303442  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16935  
Total time spent by all reduces in occupied slots (ms)=16385  
Total time spent by all map tasks (ms)=3387  
Total time spent by all reduce tasks (ms)=3277  
Total vcore-milliseconds taken by all map tasks=3387

```

Total vcore-milliseconds taken by all reduce tasks=3277
Total megabyte-milliseconds taken by all map tasks=17341440
Total megabyte-milliseconds taken by all reduce tasks=16778240
Map-Reduce Framework
  Map input records=27
  Map output records=8
  Map output bytes=809149
  Map output materialized bytes=809194
  Input split bytes=200
  Combine input records=0
  Combine output records=0
  Reduce input groups=8
  Reduce shuffle bytes=809194
  Reduce input records=8
  Reduce output records=40
  Spilled Records=16
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=41
  CPU time spent (ms)=3010
  Physical memory (bytes) snapshot=703508480
  Virtual memory (bytes) snapshot=12754894848
  Total committed heap usage (bytes)=2105540608
  Peak Map Physical memory (bytes)=402710528
  Peak Map Virtual memory (bytes)=6378991616
  Peak Reduce Physical memory (bytes)=300797952
  Peak Reduce Virtual memory (bytes)=6375903232
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=2680738
File Output Format Counters
  Bytes Written=4303442

```

```

23/01/04 14:49:12 INFO RepresentativePointsDriver: Representative Points Iteration
4
23/01/04 14:49:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:49:12 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0204
23/01/04 14:49:12 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:49:12 INFO JobSubmitter: number of splits:1
23/01/04 14:49:12 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0204
23/01/04 14:49:12 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:49:12 INFO YarnClientImpl: Submitted application
application_1672817468245_0204
23/01/04 14:49:12 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0204/
23/01/04 14:49:12 INFO Job: Running job: job_1672817468245_0204
23/01/04 14:49:19 INFO Job: Job job_1672817468245_0204 running in uber mode : false
23/01/04 14:49:19 INFO Job: map 0% reduce 0%
23/01/04 14:49:25 INFO Job: map 100% reduce 0%
23/01/04 14:49:30 INFO Job: map 100% reduce 100%

```

23/01/04 14:49:31 INFO Job: Job job\_1672817468245\_0204 completed successfully

23/01/04 14:49:31 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=675892  
FILE: Number of bytes written=1884355  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11287822  
HDFS: Number of bytes written=4979390  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17150  
Total time spent by all reduces in occupied slots (ms)=17075  
Total time spent by all map tasks (ms)=3430  
Total time spent by all reduce tasks (ms)=3415  
Total vcore-milliseconds taken by all map tasks=3430  
Total vcore-milliseconds taken by all reduce tasks=3415  
Total megabyte-milliseconds taken by all map tasks=17561600  
Total megabyte-milliseconds taken by all reduce tasks=17484800

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=675848  
Map output materialized bytes=675892  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=675892  
Reduce input records=8  
Reduce output records=48  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=3140  
Physical memory (bytes) snapshot=692383744  
Virtual memory (bytes) snapshot=12748500992  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=393056256  
Peak Map Virtual memory (bytes)=6376607744  
Peak Reduce Physical memory (bytes)=299327488  
Peak Reduce Virtual memory (bytes)=6371893248

#### 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=2680738
File Output Format Counters
Bytes Written=4979390
23/01/04 14:49:31 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.6701767568794718
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[23] =
0.6434028707487564
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[11] =
0.5199904417006127
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[19] = NaN
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[21] =
0.6763847446383956
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.5648179782292129
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[25] =
0.49599999285035035
23/01/04 14:49:31 INFO ClusterEvaluator: Intra-Cluster Density[7] =
0.514332289282225
23/01/04 14:49:31 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.5691547195749255
23/01/04 14:49:31 INFO ClusterDumper: Wrote 8 clusters
23/01/04 14:49:31 INFO MahoutDriver: Program took 95083 ms (Minutes:
1.5847166666666668)
---- K: 09 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/09
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:49:36 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[9],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/09], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:49:36 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/09
23/01/04 14:49:36 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:49:37 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:49:37 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:49:37 INFO RandomSeedGenerator: Wrote 9 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/
part-randomSeed
23/01/04 14:49:37 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/09
23/01/04 14:49:37 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:49:37 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:49:38 INFO DefaultNoHARMFailoverProxyProvider: Connecting to

```

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:49:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0205  
23/01/04 14:49:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:49:38 INFO JobSubmitter: number of splits:1  
23/01/04 14:49:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0205  
23/01/04 14:49:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:49:38 INFO Configuration: resource-types.xml not found  
23/01/04 14:49:38 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:49:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0205  
23/01/04 14:49:38 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0205/  
23/01/04 14:49:38 INFO Job: Running job: job\_1672817468245\_0205  
23/01/04 14:49:45 INFO Job: Job job\_1672817468245\_0205 running in uber mode : false  
23/01/04 14:49:45 INFO Job: map 0% reduce 0%  
23/01/04 14:49:51 INFO Job: map 100% reduce 0%  
23/01/04 14:49:57 INFO Job: map 100% reduce 100%  
23/01/04 14:49:57 INFO Job: Job job\_1672817468245\_0205 completed successfully  
23/01/04 14:49:57 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3957730  
FILE: Number of bytes written=8447733  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8393453  
HDFS: Number of bytes written=2765205  
HDFS: Number of read operations=51  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18060  
Total time spent by all reduces in occupied slots (ms)=18210  
Total time spent by all map tasks (ms)=3612  
Total time spent by all reduce tasks (ms)=3642  
Total vcore-milliseconds taken by all map tasks=3612  
Total vcore-milliseconds taken by all reduce tasks=3642  
Total megabyte-milliseconds taken by all map tasks=18493440  
Total megabyte-milliseconds taken by all reduce tasks=18647040

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=3957679  
Map output materialized bytes=3957730  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=3957730  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1

```

Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=49
CPU time spent (ms)=4710
Physical memory (bytes) snapshot=747401216
Virtual memory (bytes) snapshot=12753424384
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=431468544
Peak Map Virtual memory (bytes)=6376804352
Peak Reduce Physical memory (bytes)=315932672
Peak Reduce Virtual memory (bytes)=6376620032
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=2680027
File Output Format Counters
  Bytes Written=2765205
23/01/04 14:49:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:49:57 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0206
23/01/04 14:49:57 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:49:57 INFO JobSubmitter: number of splits:1
23/01/04 14:49:57 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0206
23/01/04 14:49:57 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:49:57 INFO YarnClientImpl: Submitted application
application_1672817468245_0206
23/01/04 14:49:57 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0206/
23/01/04 14:49:57 INFO Job: Running job: job_1672817468245_0206
23/01/04 14:50:03 INFO Job: Job job_1672817468245_0206 running in uber mode : false
23/01/04 14:50:03 INFO Job:  map 0% reduce 0%
23/01/04 14:50:09 INFO Job:  map 100% reduce 0%
23/01/04 14:50:15 INFO Job:  map 100% reduce 100%
23/01/04 14:50:15 INFO Job: Job job_1672817468245_0206 completed successfully
23/01/04 14:50:15 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=5769373
  FILE: Number of bytes written=12071019
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=8210995
  HDFS: Number of bytes written=2765205
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18390

```

Total time spent by all reduces in occupied slots (ms)=17530  
Total time spent by all map tasks (ms)=3678  
Total time spent by all reduce tasks (ms)=3506  
Total vcore-milliseconds taken by all map tasks=3678  
Total vcore-milliseconds taken by all reduce tasks=3506  
Total megabyte-milliseconds taken by all map tasks=18831360  
Total megabyte-milliseconds taken by all reduce tasks=17950720

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5769322  
Map output materialized bytes=5769373  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5769373  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=4480  
Physical memory (bytes) snapshot=730198016  
Virtual memory (bytes) snapshot=12756320256  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406151168  
Peak Map Virtual memory (bytes)=6379868160  
Peak Reduce Physical memory (bytes)=324046848  
Peak Reduce Virtual memory (bytes)=6376452096

#### 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=2680027

#### File Output Format Counters

Bytes Written=2765205

23/01/04 14:50:15 INFO KMeansDriver: Clustering data  
23/01/04 14:50:15 INFO KMeansDriver: Running Clustering  
23/01/04 14:50:15 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/09 Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/09  
23/01/04 14:50:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:50:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0207  
23/01/04 14:50:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:50:16 INFO JobSubmitter: number of splits:1  
23/01/04 14:50:16 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0207  
23/01/04 14:50:16 INFO JobSubmitter: Executing with tokens: []

```

23/01/04 14:50:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0207
23/01/04 14:50:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0207/
23/01/04 14:50:16 INFO Job: Running job: job_1672817468245_0207
23/01/04 14:50:22 INFO Job: Job job_1672817468245_0207 running in uber mode : false
23/01/04 14:50:22 INFO Job: map 0% reduce 0%
23/01/04 14:50:28 INFO Job: map 100% reduce 0%
23/01/04 14:50:28 INFO Job: Job job_1672817468245_0207 completed successfully
23/01/04 14:50:28 INFO Job: Counters: 33
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=265888
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5445596
        HDFS: Number of bytes written=2680770
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17625
        Total time spent by all reduces in occupied slots (ms)=0
        Total time spent by all map tasks (ms)=3525
        Total vcore-milliseconds taken by all map tasks=3525
        Total megabyte-milliseconds taken by all map tasks=18048000
    Map-Reduce Framework
        Map input records=27
        Map output records=27
        Input split bytes=170
        Spilled Records=0
        Failed Shuffles=0
        Merged Map outputs=0
        GC time elapsed (ms)=25
        CPU time spent (ms)=2300
        Physical memory (bytes) snapshot=314839040
        Virtual memory (bytes) snapshot=6379626496
        Total committed heap usage (bytes)=1052770304
        Peak Map Physical memory (bytes)=314839040
        Peak Map Virtual memory (bytes)=6379626496
    File Input Format Counters
        Bytes Read=2680027
    File Output Format Counters
        Bytes Written=2680770
23/01/04 14:50:28 INFO MahoutDriver: Program took 52153 ms (Minutes:
0.8692333333333333)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:50:34 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-

```

```
corpus-kmeans/CosineDistanceMeasure/09/clusters-2-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
CosineDistanceMeasure/09.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/
09/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:50:35 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:50:35 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/09/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/09/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:50:35 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:50:36 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:50:36 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0208
23/01/04 14:50:36 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:50:36 INFO JobSubmitter: number of splits:1
23/01/04 14:50:36 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0208
23/01/04 14:50:36 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:50:36 INFO Configuration: resource-types.xml not found
23/01/04 14:50:36 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:50:36 INFO YarnClientImpl: Submitted application
application_1672817468245_0208
23/01/04 14:50:36 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0208/
23/01/04 14:50:36 INFO Job: Running job: job_1672817468245_0208
23/01/04 14:50:42 INFO Job: Job job_1672817468245_0208 running in uber mode : false
23/01/04 14:50:42 INFO Job:  map 0% reduce 0%
23/01/04 14:50:49 INFO Job:  map 100% reduce 0%
23/01/04 14:50:54 INFO Job:  map 100% reduce 100%
23/01/04 14:50:55 INFO Job: Job job_1672817468245_0208 completed successfully
23/01/04 14:50:55 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=766563
FILE: Number of bytes written=2065697
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5686192
HDFS: Number of bytes written=2269224
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16545
Total time spent by all reduces in occupied slots (ms)=16305
Total time spent by all map tasks (ms)=3309
Total time spent by all reduce tasks (ms)=3261
Total vcore-milliseconds taken by all map tasks=3309
```

Total vcore-milliseconds taken by all reduce tasks=3261  
 Total megabyte-milliseconds taken by all map tasks=16942080  
 Total megabyte-milliseconds taken by all reduce tasks=16696320  
 Map-Reduce Framework  
   Map input records=27  
   Map output records=9  
   Map output bytes=766513  
   Map output materialized bytes=766563  
   Input split bytes=200  
   Combine input records=0  
   Combine output records=0  
   Reduce input groups=9  
   Reduce shuffle bytes=766563  
   Reduce input records=9  
   Reduce output records=18  
   Spilled Records=18  
   Shuffled Maps =1  
   Failed Shuffles=0  
   Merged Map outputs=1  
   GC time elapsed (ms)=43  
   CPU time spent (ms)=2800  
   Physical memory (bytes) snapshot=688611328  
   Virtual memory (bytes) snapshot=12761493504  
   Total committed heap usage (bytes)=2105540608  
   Peak Map Physical memory (bytes)=387985408  
   Peak Map Virtual memory (bytes)=6374838272  
   Peak Reduce Physical memory (bytes)=300625920  
   Peak Reduce Virtual memory (bytes)=6390562816  
 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=2680770  
 File Output Format Counters  
   Bytes Written=2269224

23/01/04 14:50:55 INFO RepresentativePointsDriver: Representative Points Iteration 1  
 23/01/04 14:50:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
 23/01/04 14:50:55 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0209  
 23/01/04 14:50:55 INFO FileInputFormat: Total input files to process : 1  
 23/01/04 14:50:55 INFO JobSubmitter: number of splits:1  
 23/01/04 14:50:55 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0209  
 23/01/04 14:50:55 INFO JobSubmitter: Executing with tokens: []  
 23/01/04 14:50:55 INFO YarnClientImpl: Submitted application application\_1672817468245\_0209  
 23/01/04 14:50:55 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0209/  
 23/01/04 14:50:55 INFO Job: Running job: job\_1672817468245\_0209  
 23/01/04 14:51:01 INFO Job: Job job\_1672817468245\_0209 running in uber mode : false  
 23/01/04 14:51:01 INFO Job: map 0% reduce 0%  
 23/01/04 14:51:08 INFO Job: map 100% reduce 0%  
 23/01/04 14:51:12 INFO Job: map 100% reduce 100%

23/01/04 14:51:12 INFO Job: Job job\_1672817468245\_0209 completed successfully

23/01/04 14:51:12 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=935321  
FILE: Number of bytes written=2403213  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7219418  
HDFS: Number of bytes written=3204575  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16960  
Total time spent by all reduces in occupied slots (ms)=9380  
Total time spent by all map tasks (ms)=3392  
Total time spent by all reduce tasks (ms)=1876  
Total vcore-milliseconds taken by all map tasks=3392  
Total vcore-milliseconds taken by all reduce tasks=1876  
Total megabyte-milliseconds taken by all map tasks=17367040  
Total megabyte-milliseconds taken by all reduce tasks=9605120

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=935271  
Map output materialized bytes=935321  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=935321  
Reduce input records=9  
Reduce output records=27  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=3130  
Physical memory (bytes) snapshot=687570944  
Virtual memory (bytes) snapshot=12765749248  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=391168000  
Peak Map Virtual memory (bytes)=6375182336  
Peak Reduce Physical memory (bytes)=296402944  
Peak Reduce Virtual memory (bytes)=6394474496

#### 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=2680770  
File Output Format Counters  
Bytes Written=3204575  
23/01/04 14:51:12 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 14:51:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:51:12 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0210  
23/01/04 14:51:12 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:51:12 INFO JobSubmitter: number of splits:1  
23/01/04 14:51:12 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0210  
23/01/04 14:51:12 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:51:12 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0210  
23/01/04 14:51:12 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0210/  
23/01/04 14:51:12 INFO Job: Running job: job\_1672817468245\_0210  
23/01/04 14:51:18 INFO Job: Job job\_1672817468245\_0210 running in uber mode : false  
23/01/04 14:51:18 INFO Job: map 0% reduce 0%  
23/01/04 14:51:25 INFO Job: map 100% reduce 0%  
23/01/04 14:51:31 INFO Job: map 100% reduce 100%  
23/01/04 14:51:31 INFO Job: Job job\_1672817468245\_0210 completed successfully  
23/01/04 14:51:31 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=977409  
FILE: Number of bytes written=2487389  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9090120  
HDFS: Number of bytes written=4182093  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16805  
Total time spent by all reduces in occupied slots (ms)=16705  
Total time spent by all map tasks (ms)=3361  
Total time spent by all reduce tasks (ms)=3341  
Total vcore-milliseconds taken by all map tasks=3361  
Total vcore-milliseconds taken by all reduce tasks=3341  
Total megabyte-milliseconds taken by all map tasks=17208320  
Total megabyte-milliseconds taken by all reduce tasks=17105920

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=977358  
Map output materialized bytes=977409  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=977409

Reduce input records=9  
Reduce output records=36  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3270  
Physical memory (bytes) snapshot=687673344  
Virtual memory (bytes) snapshot=12753252352  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399888384  
Peak Map Virtual memory (bytes)=6368870400  
Peak Reduce Physical memory (bytes)=287784960  
Peak Reduce Virtual memory (bytes)=6384381952

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=2680770

File Output Format Counters

Bytes Written=4182093

23/01/04 14:51:31 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 14:51:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:51:31 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0211

23/01/04 14:51:31 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:51:31 INFO JobSubmitter: number of splits:1

23/01/04 14:51:31 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0211

23/01/04 14:51:31 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:51:31 INFO YarnClientImpl: Submitted application application\_1672817468245\_0211

23/01/04 14:51:31 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0211/](http://lena-master:8088/proxy/application_1672817468245_0211/)

23/01/04 14:51:31 INFO Job: Running job: job\_1672817468245\_0211

23/01/04 14:51:37 INFO Job: Job job\_1672817468245\_0211 running in uber mode : false

23/01/04 14:51:37 INFO Job: map 0% reduce 0%

23/01/04 14:51:43 INFO Job: map 100% reduce 0%

23/01/04 14:51:49 INFO Job: map 100% reduce 100%

23/01/04 14:51:49 INFO Job: Job job\_1672817468245\_0211 completed successfully

23/01/04 14:51:49 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=827514

FILE: Number of bytes written=2187599

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=11045156

HDFS: Number of bytes written=5009659

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=16945  
  Total time spent by all reduces in occupied slots (ms)=16920  
  Total time spent by all map tasks (ms)=3389  
  Total time spent by all reduce tasks (ms)=3384  
  Total vcore-milliseconds taken by all map tasks=3389  
  Total vcore-milliseconds taken by all reduce tasks=3384  
  Total megabyte-milliseconds taken by all map tasks=17351680  
  Total megabyte-milliseconds taken by all reduce tasks=17326080

Map-Reduce Framework  
  Map input records=27  
  Map output records=9  
  Map output bytes=827466  
  Map output materialized bytes=827514  
  Input split bytes=200  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=9  
  Reduce shuffle bytes=827514  
  Reduce input records=9  
  Reduce output records=45  
  Spilled Records=18  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=46  
  CPU time spent (ms)=3090  
  Physical memory (bytes) snapshot=705368064  
  Virtual memory (bytes) snapshot=12755169280  
  Total committed heap usage (bytes)=2105540608  
  Peak Map Physical memory (bytes)=409882624  
  Peak Map Virtual memory (bytes)=6375907328  
  Peak Reduce Physical memory (bytes)=295485440  
  Peak Reduce Virtual memory (bytes)=6379261952

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=2680770

File Output Format Counters

  Bytes Written=5009659

23/01/04 14:51:49 INFO RepresentativePointsDriver: Representative Points Iteration  
4  
23/01/04 14:51:49 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:51:49 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0213  
23/01/04 14:51:49 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:51:49 INFO JobSubmitter: number of splits:1  
23/01/04 14:51:49 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0213

23/01/04 14:51:49 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:51:50 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0213  
23/01/04 14:51:50 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0213/  
23/01/04 14:51:50 INFO Job: Running job: job\_1672817468245\_0213  
23/01/04 14:51:56 INFO Job: Job job\_1672817468245\_0213 running in uber mode : false  
23/01/04 14:51:56 INFO Job: map 0% reduce 0%  
23/01/04 14:52:01 INFO Job: map 100% reduce 0%  
23/01/04 14:52:07 INFO Job: map 100% reduce 100%  
23/01/04 14:52:08 INFO Job: Job job\_1672817468245\_0213 completed successfully  
23/01/04 14:52:08 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=897438  
FILE: Number of bytes written=2327447  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=12700288  
HDFS: Number of bytes written=5907166  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17040  
Total time spent by all reduces in occupied slots (ms)=16955  
Total time spent by all map tasks (ms)=3408  
Total time spent by all reduce tasks (ms)=3391  
Total vcore-milliseconds taken by all map tasks=3408  
Total vcore-milliseconds taken by all reduce tasks=3391  
Total megabyte-milliseconds taken by all map tasks=17448960  
Total megabyte-milliseconds taken by all reduce tasks=17361920

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=897387  
Map output materialized bytes=897438  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=897438  
Reduce input records=9  
Reduce output records=54  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=83  
CPU time spent (ms)=3640  
Physical memory (bytes) snapshot=706879488  
Virtual memory (bytes) snapshot=12747997184  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406413312  
Peak Map Virtual memory (bytes)=6370365440

```

        Peak Reduce Physical memory (bytes)=300466176
        Peak Reduce Virtual memory (bytes)=6377631744
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=2680770
File Output Format Counters
    Bytes Written=5907166
23/01/04 14:52:08 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.6853061757770863
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[26] =
0.49599999285035035
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[14] =
0.521851636455987
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[24] = NaN
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[17] =
0.5360213631254261
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[9] =
0.5997318058999354
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[6] =
0.6375025448658466
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[18] =
0.6456853989902614
23/01/04 14:52:08 INFO ClusterEvaluator: Intra-Cluster Density[21] =
0.5487362020127529
23/01/04 14:52:08 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.5693612777429371
23/01/04 14:52:09 INFO ClusterDumper: Wrote 9 clusters
23/01/04 14:52:09 INFO MahoutDriver: Program took 95056 ms (Minutes:
1.5842666666666667)
---- K: 10 --
dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/10
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:52:14 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
CosineDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[10],
--output=[hdfs://lana/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/CosineDistanceMeasure/10], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:52:15 INFO HadoopUtil: Deleting
hdfs://lana/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/10
23/01/04 14:52:15 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/CosineDistanceMeasure
23/01/04 14:52:15 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library

```

23/01/04 14:52:15 INFO CodecPool: Got brand-new compressor [.deflate]  
23/01/04 14:52:15 INFO RandomSeedGenerator: Wrote 10 Klusters to  
dsm010/question\_2/british-fiction-corpus-canopy-centroids/CosineDistanceMeasure/  
part-randomSeed  
23/01/04 14:52:15 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-  
corpus-vectors/tf-vectors Clusters In: dsm010/question\_2/british-fiction-corpus-  
canopy-centroids/CosineDistanceMeasure/part-randomSeed Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/10  
23/01/04 14:52:15 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 14:52:15 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 14:52:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:52:16 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0217  
23/01/04 14:52:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:52:16 INFO JobSubmitter: number of splits:1  
23/01/04 14:52:16 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0217  
23/01/04 14:52:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:52:17 INFO Configuration: resource-types.xml not found  
23/01/04 14:52:17 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:52:17 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0217  
23/01/04 14:52:17 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0217/  
23/01/04 14:52:17 INFO Job: Running job: job\_1672817468245\_0217  
23/01/04 14:52:23 INFO Job: Job job\_1672817468245\_0217 running in uber mode : false  
23/01/04 14:52:23 INFO Job: map 0% reduce 0%  
23/01/04 14:52:29 INFO Job: map 100% reduce 0%  
23/01/04 14:52:33 INFO Job: map 100% reduce 100%  
23/01/04 14:52:33 INFO Job: Job job\_1672817468245\_0217 completed successfully  
23/01/04 14:52:33 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3947542  
FILE: Number of bytes written=8427357  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8309649  
HDFS: Number of bytes written=2682608  
HDFS: Number of read operations=55  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18285  
Total time spent by all reduces in occupied slots (ms)=10550  
Total time spent by all map tasks (ms)=3657  
Total time spent by all reduce tasks (ms)=2110  
Total vcore-milliseconds taken by all map tasks=3657  
Total vcore-milliseconds taken by all reduce tasks=2110  
Total megabyte-milliseconds taken by all map tasks=18723840  
Total megabyte-milliseconds taken by all reduce tasks=10803200

#### Map-Reduce Framework

Map input records=27

Map output records=10  
Map output bytes=3947486  
Map output materialized bytes=3947542  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=3947542  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65  
CPU time spent (ms)=4500  
Physical memory (bytes) snapshot=718884864  
Virtual memory (bytes) snapshot=12757098496  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406757376  
Peak Map Virtual memory (bytes)=6376714240  
Peak Reduce Physical memory (bytes)=312127488  
Peak Reduce Virtual memory (bytes)=6380384256

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=2680027

File Output Format Counters

Bytes Written=2682608

23/01/04 14:52:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:52:33 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0220  
23/01/04 14:52:33 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:52:33 INFO JobSubmitter: number of splits:1  
23/01/04 14:52:33 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0220  
23/01/04 14:52:33 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:52:33 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0220  
23/01/04 14:52:33 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0220/](http://lena-master:8088/proxy/application_1672817468245_0220/)  
23/01/04 14:52:33 INFO Job: Running job: job\_1672817468245\_0220  
23/01/04 14:52:39 INFO Job: Job job\_1672817468245\_0220 running in uber mode : false  
23/01/04 14:52:39 INFO Job: map 0% reduce 0%  
23/01/04 14:52:45 INFO Job: map 100% reduce 0%  
23/01/04 14:52:49 INFO Job: map 100% reduce 100%  
23/01/04 14:52:49 INFO Job: Job job\_1672817468245\_0220 completed successfully  
23/01/04 14:52:49 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=5690403  
FILE: Number of bytes written=11913079  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=8045801  
HDFS: Number of bytes written=2682608  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18325  
Total time spent by all reduces in occupied slots (ms)=10195  
Total time spent by all map tasks (ms)=3665  
Total time spent by all reduce tasks (ms)=2039  
Total vcore-milliseconds taken by all map tasks=3665  
Total vcore-milliseconds taken by all reduce tasks=2039  
Total megabyte-milliseconds taken by all map tasks=18764800  
Total megabyte-milliseconds taken by all reduce tasks=10439680

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5690347  
Map output materialized bytes=5690403  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5690403  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=55  
CPU time spent (ms)=4230  
Physical memory (bytes) snapshot=762011648  
Virtual memory (bytes) snapshot=12763283456  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408686592  
Peak Map Virtual memory (bytes)=6373949440  
Peak Reduce Physical memory (bytes)=353325056  
Peak Reduce Virtual memory (bytes)=6389334016

#### 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=2680027

#### File Output Format Counters

Bytes Written=2682608

23/01/04 14:52:50 INFO KMeansDriver: Clustering data  
23/01/04 14:52:50 INFO KMeansDriver: Running Clustering  
23/01/04 14:52:50 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-  
corpus-vectors/tf-vectors Clusters In:



hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
CosineDistanceMeasure/10 Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-  
fiction-corpus-kmeans/CosineDistanceMeasure/10  
23/01/04 14:52:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:52:50 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0223  
23/01/04 14:52:50 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:52:50 INFO JobSubmitter: number of splits:1  
23/01/04 14:52:50 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0223  
23/01/04 14:52:50 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:52:50 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0223  
23/01/04 14:52:50 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0223/  
23/01/04 14:52:50 INFO Job: Running job: job\_1672817468245\_0223  
23/01/04 14:52:56 INFO Job: Job job\_1672817468245\_0223 running in uber mode : false  
23/01/04 14:52:56 INFO Job: map 0% reduce 0%  
23/01/04 14:53:02 INFO Job: map 100% reduce 0%  
23/01/04 14:53:02 INFO Job: Job job\_1672817468245\_0223 completed successfully  
23/01/04 14:53:02 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265888  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5362999  
HDFS: Number of bytes written=2680718  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18095  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3619  
Total vcore-milliseconds taken by all map tasks=3619  
Total megabyte-milliseconds taken by all map tasks=18529280

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=43  
CPU time spent (ms)=2130  
Physical memory (bytes) snapshot=289333248  
Virtual memory (bytes) snapshot=6370758656  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=289333248  
Peak Map Virtual memory (bytes)=6370758656

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

```

        Bytes Written=2680718
23/01/04 14:53:02 INFO MahoutDriver: Program took 47753 ms (Minutes: 0.7959)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:53:08 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/CosineDistanceMeasure/10/clusters-2-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
CosineDistanceMeasure/10.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/
10/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:53:10 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:53:10 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
CosineDistanceMeasure/10/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.CosineDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/CosineDistanceMeasure/10/
clusters-2-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:53:10 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:53:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:53:10 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0227
23/01/04 14:53:10 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:53:11 INFO JobSubmitter: number of splits:1
23/01/04 14:53:11 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0227
23/01/04 14:53:11 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:53:11 INFO Configuration: resource-types.xml not found
23/01/04 14:53:11 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:53:11 INFO YarnClientImpl: Submitted application
application_1672817468245_0227
23/01/04 14:53:11 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0227/
23/01/04 14:53:11 INFO Job: Running job: job_1672817468245_0227
23/01/04 14:53:17 INFO Job: Job job_1672817468245_0227 running in uber mode : false
23/01/04 14:53:17 INFO Job:  map 0% reduce 0%
23/01/04 14:53:23 INFO Job:  map 100% reduce 0%
23/01/04 14:53:28 INFO Job:  map 100% reduce 100%
23/01/04 14:53:29 INFO Job: Job job_1672817468245_0227 completed successfully
23/01/04 14:53:29 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=769810
        FILE: Number of bytes written=2072191
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5689864
        HDFS: Number of bytes written=2274311
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0

```

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16725  
Total time spent by all reduces in occupied slots (ms)=16420  
Total time spent by all map tasks (ms)=3345  
Total time spent by all reduce tasks (ms)=3284  
Total vcore-milliseconds taken by all map tasks=3345  
Total vcore-milliseconds taken by all reduce tasks=3284  
Total megabyte-milliseconds taken by all map tasks=17126400  
Total megabyte-milliseconds taken by all reduce tasks=16814080

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=769758  
Map output materialized bytes=769810  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=769810  
Reduce input records=10  
Reduce output records=20  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=2750  
Physical memory (bytes) snapshot=684060672  
Virtual memory (bytes) snapshot=12755697664  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390754304  
Peak Map Virtual memory (bytes)=6372823040  
Peak Reduce Physical memory (bytes)=293306368  
Peak Reduce Virtual memory (bytes)=6386782208

#### 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=2680718

#### File Output Format Counters

Bytes Written=2274311

23/01/04 14:53:29 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 14:53:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:53:29 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0230

23/01/04 14:53:29 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:53:29 INFO JobSubmitter: number of splits:1

23/01/04 14:53:29 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0230

23/01/04 14:53:29 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:53:29 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0230

23/01/04 14:53:29 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0230/](http://lena-master:8088/proxy/application_1672817468245_0230/)

23/01/04 14:53:29 INFO Job: Running job: job\_1672817468245\_0230

23/01/04 14:53:36 INFO Job: Job job\_1672817468245\_0230 running in uber mode : false

23/01/04 14:53:36 INFO Job: map 0% reduce 0%

23/01/04 14:53:42 INFO Job: map 100% reduce 0%

23/01/04 14:53:46 INFO Job: map 100% reduce 100%

23/01/04 14:53:46 INFO Job: Job job\_1672817468245\_0230 completed successfully

23/01/04 14:53:46 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=953592

FILE: Number of bytes written=2439755

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=7229540

HDFS: Number of bytes written=3227949

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=16445

Total time spent by all reduces in occupied slots (ms)=9460

Total time spent by all map tasks (ms)=3289

Total time spent by all reduce tasks (ms)=1892

Total vcore-milliseconds taken by all map tasks=3289

Total vcore-milliseconds taken by all reduce tasks=1892

Total megabyte-milliseconds taken by all map tasks=16839680

Total megabyte-milliseconds taken by all reduce tasks=9687040

#### Map-Reduce Framework

Map input records=27

Map output records=10

Map output bytes=953538

Map output materialized bytes=953592

Input split bytes=200

Combine input records=0

Combine output records=0

Reduce input groups=10

Reduce shuffle bytes=953592

Reduce input records=10

Reduce output records=30

Spilled Records=20

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=41

CPU time spent (ms)=2990

Physical memory (bytes) snapshot=683548672

Virtual memory (bytes) snapshot=12760506368

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=391446528

```

Peak Map Virtual memory (bytes)=6377156608
Peak Reduce Physical memory (bytes)=292102144
Peak Reduce Virtual memory (bytes)=6387257344
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=2680718
File Output Format Counters
  Bytes Written=3227949
23/01/04 14:53:46 INFO RepresentativePointsDriver: Representative Points Iteration
2
23/01/04 14:53:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:53:46 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0233
23/01/04 14:53:46 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:53:46 INFO JobSubmitter: number of splits:1
23/01/04 14:53:46 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0233
23/01/04 14:53:46 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:53:46 INFO YarnClientImpl: Submitted application
application_1672817468245_0233
23/01/04 14:53:46 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0233/
23/01/04 14:53:46 INFO Job: Running job: job_1672817468245_0233
23/01/04 14:53:52 INFO Job: Job job_1672817468245_0233 running in uber mode : false
23/01/04 14:53:52 INFO Job:  map 0% reduce 0%
23/01/04 14:53:57 INFO Job:  map 100% reduce 0%
23/01/04 14:54:03 INFO Job:  map 100% reduce 100%
23/01/04 14:54:04 INFO Job: Job job_1672817468245_0233 completed successfully
23/01/04 14:54:04 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=967928
  FILE: Number of bytes written=2468427
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=9136816
  HDFS: Number of bytes written=4195983
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=16800
  Total time spent by all reduces in occupied slots (ms)=16790
  Total time spent by all map tasks (ms)=3360
  Total time spent by all reduce tasks (ms)=3358
  Total vcore-milliseconds taken by all map tasks=3360
  Total vcore-milliseconds taken by all reduce tasks=3358
  Total megabyte-milliseconds taken by all map tasks=17203200

```

Total megabyte-milliseconds taken by all reduce tasks=17192960  
Map-Reduce Framework  
Map input records=27  
Map output records=10  
Map output bytes=967874  
Map output materialized bytes=967928  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=967928  
Reduce input records=10  
Reduce output records=40  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=3130  
Physical memory (bytes) snapshot=695107584  
Virtual memory (bytes) snapshot=12751433728  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394596352  
Peak Map Virtual memory (bytes)=6372995072  
Peak Reduce Physical memory (bytes)=300511232  
Peak Reduce Virtual memory (bytes)=6378438656

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=2680718

File Output Format Counters

Bytes Written=4195983

23/01/04 14:54:04 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 14:54:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:54:04 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0237  
23/01/04 14:54:04 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:54:04 INFO JobSubmitter: number of splits:1  
23/01/04 14:54:04 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0237  
23/01/04 14:54:04 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:54:04 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0237  
23/01/04 14:54:04 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0237/  
23/01/04 14:54:04 INFO Job: Running job: job\_1672817468245\_0237  
23/01/04 14:54:10 INFO Job: Job job\_1672817468245\_0237 running in uber mode : false  
23/01/04 14:54:10 INFO Job: map 0% reduce 0%  
23/01/04 14:54:17 INFO Job: map 100% reduce 0%  
23/01/04 14:54:23 INFO Job: map 100% reduce 100%  
23/01/04 14:54:23 INFO Job: Job job\_1672817468245\_0237 completed successfully  
23/01/04 14:54:23 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=935844  
FILE: Number of bytes written=2404259  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11072884  
HDFS: Number of bytes written=5131874  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16715  
Total time spent by all reduces in occupied slots (ms)=17055  
Total time spent by all map tasks (ms)=3343  
Total time spent by all reduce tasks (ms)=3411  
Total vcore-milliseconds taken by all map tasks=3343  
Total vcore-milliseconds taken by all reduce tasks=3411  
Total megabyte-milliseconds taken by all map tasks=17116160  
Total megabyte-milliseconds taken by all reduce tasks=17464320

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=935791  
Map output materialized bytes=935844  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=935844  
Reduce input records=10  
Reduce output records=50  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3340  
Physical memory (bytes) snapshot=683819008  
Virtual memory (bytes) snapshot=12759609344  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394240000  
Peak Map Virtual memory (bytes)=6379905024  
Peak Reduce Physical memory (bytes)=289579008  
Peak Reduce Virtual memory (bytes)=6379704320

#### 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=2680718

#### File Output Format Counters

Bytes Written=5131874

23/01/04 14:54:23 INFO RepresentativePointsDriver: Representative Points Iteration 4  
23/01/04 14:54:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:54:23 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0240  
23/01/04 14:54:23 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:54:23 INFO JobSubmitter: number of splits:1  
23/01/04 14:54:23 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0240  
23/01/04 14:54:23 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:54:23 INFO YarnClientImpl: Submitted application application\_1672817468245\_0240  
23/01/04 14:54:23 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0240/  
23/01/04 14:54:23 INFO Job: Running job: job\_1672817468245\_0240  
23/01/04 14:54:29 INFO Job: Job job\_1672817468245\_0240 running in uber mode : false  
23/01/04 14:54:29 INFO Job: map 0% reduce 0%  
23/01/04 14:54:35 INFO Job: map 100% reduce 0%  
23/01/04 14:54:41 INFO Job: map 100% reduce 100%  
23/01/04 14:54:41 INFO Job: Job job\_1672817468245\_0240 completed successfully  
23/01/04 14:54:41 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=962730  
FILE: Number of bytes written=2458031  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=12944666  
HDFS: Number of bytes written=6094690  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17630  
Total time spent by all reduces in occupied slots (ms)=17010  
Total time spent by all map tasks (ms)=3526  
Total time spent by all reduce tasks (ms)=3402  
Total vcore-milliseconds taken by all map tasks=3526  
Total vcore-milliseconds taken by all reduce tasks=3402  
Total megabyte-milliseconds taken by all map tasks=18053120  
Total megabyte-milliseconds taken by all reduce tasks=17418240

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=962676  
Map output materialized bytes=962730  
Input split bytes=200  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=962730  
Reduce input records=10  
Reduce output records=60



Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=82  
CPU time spent (ms)=3150  
Physical memory (bytes) snapshot=705208320  
Virtual memory (bytes) snapshot=12759461888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403668992  
Peak Map Virtual memory (bytes)=6377607168  
Peak Reduce Physical memory (bytes)=301539328  
Peak Reduce Virtual memory (bytes)=6381854720

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=2680718

File Output Format Counters

Bytes Written=6094690

23/01/04 14:54:41 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.69038947281991  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[22] = NaN  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[21] =  
0.5304732605365603  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[2] =  
0.5000243138814683  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[3] = NaN  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[23] =  
0.6867516163406496  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[10] =  
0.521851636455987  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[6] =  
0.6375025448658466  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[25] =  
0.49599999285035035  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[18] =  
0.6701921205115067  
23/01/04 14:54:41 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN  
23/01/04 14:54:41 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.5775422122060527  
23/01/04 14:54:42 INFO ClusterDumper: Wrote 10 clusters  
23/01/04 14:54:42 INFO MahoutDriver: Program took 93957 ms (Minutes: 1.56595)  
--- Distance Metric: org.apache.mahout.common.distance.EuclideanDistanceMeasure  
MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.  
Running on hadoop, using /opt/hadoop/current/bin/hadoop and  
HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop  
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar  
23/01/04 14:54:47 INFO AbstractJob: Command line arguments: {--  
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --  
endPhase=[2147483647],  
--input=[dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors], --  
method=[mapreduce], --output=[dsm010/question\_2/british-fiction-corpus-canopy-  
centroids/EuclideanDistanceMeasure], --overwrite=null, --startPhase=[0], --  
t1=[0.5], --t2=[0.3], --tempDir=[temp]}

23/01/04 14:54:48 INFO CanopyDriver: Build Clusters Input:  
dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Out:  
dsm010/question\_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure  
Measure: org.apache.mahout.common.distance.EuclideanDistanceMeasure@6accbb21 t1:  
0.5 t2: 0.3  
23/01/04 14:54:48 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:54:48 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0244  
23/01/04 14:54:49 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:54:49 INFO JobSubmitter: number of splits:1  
23/01/04 14:54:49 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0244  
23/01/04 14:54:49 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:54:49 INFO Configuration: resource-types.xml not found  
23/01/04 14:54:49 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:54:49 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0244  
23/01/04 14:54:49 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0244/  
23/01/04 14:54:49 INFO Job: Running job: job\_1672817468245\_0244  
23/01/04 14:54:55 INFO Job: Job job\_1672817468245\_0244 running in uber mode : false  
23/01/04 14:54:55 INFO Job: map 0% reduce 0%  
23/01/04 14:55:01 INFO Job: map 100% reduce 0%  
23/01/04 14:55:06 INFO Job: map 100% reduce 100%  
23/01/04 14:55:07 INFO Job: Job job\_1672817468245\_0244 completed successfully  
23/01/04 14:55:07 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2678621  
FILE: Number of bytes written=5891013  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2680197  
HDFS: Number of bytes written=2682998  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17665  
Total time spent by all reduces in occupied slots (ms)=17825  
Total time spent by all map tasks (ms)=3533  
Total time spent by all reduce tasks (ms)=3565  
Total vcore-milliseconds taken by all map tasks=3533  
Total vcore-milliseconds taken by all reduce tasks=3565  
Total megabyte-milliseconds taken by all map tasks=18088960  
Total megabyte-milliseconds taken by all reduce tasks=18252800

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2678485  
Map output materialized bytes=2678621  
Input split bytes=170  
Combine input records=0  
Combine output records=0

```
Reduce input groups=1
Reduce shuffle bytes=2678621
Reduce input records=27
Reduce output records=27
Spilled Records=54
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=75
CPU time spent (ms)=4080
Physical memory (bytes) snapshot=726245376
Virtual memory (bytes) snapshot=12749905920
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=409047040
Peak Map Virtual memory (bytes)=6369501184
Peak Reduce Physical memory (bytes)=317198336
Peak Reduce Virtual memory (bytes)=6380404736
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=2680027
File Output Format Counters
  Bytes Written=2682998
```

```
23/01/04 14:55:07 INFO MahoutDriver: Program took 19928 ms (Minutes:
0.33213333333333334)
```

```
---- K: 01 --
```

```
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/01
```

```
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
```

```
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
```

```
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
```

```
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
```

```
23/01/04 14:55:13 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[1],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/01], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
```

```
23/01/04 14:55:13 INFO HadoopUtil: Deleting
```

```
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/01
```

```
23/01/04 14:55:13 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
```

```
23/01/04 14:55:14 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
```

```
23/01/04 14:55:14 INFO CodecPool: Got brand-new compressor [.deflate]
```

```
23/01/04 14:55:14 INFO RandomSeedGenerator: Wrote 1 Klusters to
```

```
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
```

```
23/01/04 14:55:14 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
```

canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/01  
23/01/04 14:55:14 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 14:55:14 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 14:55:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:55:14 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0248  
23/01/04 14:55:14 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:55:14 INFO JobSubmitter: number of splits:1  
23/01/04 14:55:14 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0248  
23/01/04 14:55:14 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:55:15 INFO Configuration: resource-types.xml not found  
23/01/04 14:55:15 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 14:55:15 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0248  
23/01/04 14:55:15 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0248/  
23/01/04 14:55:15 INFO Job: Running job: job\_1672817468245\_0248  
23/01/04 14:55:21 INFO Job: Job job\_1672817468245\_0248 running in uber mode : false  
23/01/04 14:55:21 INFO Job: map 0% reduce 0%  
23/01/04 14:55:27 INFO Job: map 100% reduce 0%  
23/01/04 14:55:33 INFO Job: map 100% reduce 100%  
23/01/04 14:55:33 INFO Job: Job job\_1672817468245\_0248 completed successfully  
23/01/04 14:55:33 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=938295  
FILE: Number of bytes written=2408881  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3546755  
HDFS: Number of bytes written=794075  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17570  
Total time spent by all reduces in occupied slots (ms)=17630  
Total time spent by all map tasks (ms)=3514  
Total time spent by all reduce tasks (ms)=3526  
Total vcore-milliseconds taken by all map tasks=3514  
Total vcore-milliseconds taken by all reduce tasks=3526  
Total megabyte-milliseconds taken by all map tasks=17991680  
Total megabyte-milliseconds taken by all reduce tasks=18053120

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=938284  
Map output materialized bytes=938295  
Input split bytes=170  
Combine input records=0  
Combine output records=0

Reduce input groups=1  
Reduce shuffle bytes=938295  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65  
CPU time spent (ms)=3430  
Physical memory (bytes) snapshot=715849728  
Virtual memory (bytes) snapshot=12762419200  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=387506176  
Peak Map Virtual memory (bytes)=6375333888  
Peak Reduce Physical memory (bytes)=328343552  
Peak Reduce Virtual memory (bytes)=6387085312

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=2680027

File Output Format Counters

Bytes Written=794075

23/01/04 14:55:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:55:33 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0252  
23/01/04 14:55:33 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:55:33 INFO JobSubmitter: number of splits:1  
23/01/04 14:55:33 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0252  
23/01/04 14:55:33 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:55:33 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0252  
23/01/04 14:55:33 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0252/  
23/01/04 14:55:33 INFO Job: Running job: job\_1672817468245\_0252  
23/01/04 14:55:39 INFO Job: Job job\_1672817468245\_0252 running in uber mode : false  
23/01/04 14:55:39 INFO Job: map 0% reduce 0%  
23/01/04 14:55:44 INFO Job: map 100% reduce 0%  
23/01/04 14:55:50 INFO Job: map 100% reduce 100%  
23/01/04 14:55:51 INFO Job: Job job\_1672817468245\_0252 completed successfully  
23/01/04 14:55:51 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=1587761  
FILE: Number of bytes written=3707813  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4268735  
HDFS: Number of bytes written=794075  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17565  
  Total time spent by all reduces in occupied slots (ms)=17175  
  Total time spent by all map tasks (ms)=3513  
  Total time spent by all reduce tasks (ms)=3435  
  Total vcore-milliseconds taken by all map tasks=3513  
  Total vcore-milliseconds taken by all reduce tasks=3435  
  Total megabyte-milliseconds taken by all map tasks=17986560  
  Total megabyte-milliseconds taken by all reduce tasks=17587200

Map-Reduce Framework  
  Map input records=27  
  Map output records=1  
  Map output bytes=1587750  
  Map output materialized bytes=1587761  
  Input split bytes=170  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=1  
  Reduce shuffle bytes=1587761  
  Reduce input records=1  
  Reduce output records=1  
  Spilled Records=2  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=49  
  CPU time spent (ms)=3390  
  Physical memory (bytes) snapshot=725270528  
  Virtual memory (bytes) snapshot=12749783040  
  Total committed heap usage (bytes)=2105540608  
  Peak Map Physical memory (bytes)=401162240  
  Peak Map Virtual memory (bytes)=6372544512  
  Peak Reduce Physical memory (bytes)=324108288  
  Peak Reduce Virtual memory (bytes)=6377238528

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=2680027

File Output Format Counters

  Bytes Written=794075

23/01/04 14:55:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:55:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0255  
23/01/04 14:55:52 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:55:52 INFO JobSubmitter: number of splits:1  
23/01/04 14:55:52 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0255  
23/01/04 14:55:52 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:55:52 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0255  
23/01/04 14:55:52 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0255/  
23/01/04 14:55:52 INFO Job: Running job: job\_1672817468245\_0255  
23/01/04 14:55:58 INFO Job: Job job\_1672817468245\_0255 running in uber mode : false  
23/01/04 14:55:58 INFO Job: map 0% reduce 0%  
23/01/04 14:56:04 INFO Job: map 100% reduce 0%  
23/01/04 14:56:10 INFO Job: map 100% reduce 100%  
23/01/04 14:56:10 INFO Job: Job job\_1672817468245\_0255 completed successfully  
23/01/04 14:56:10 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1587761  
FILE: Number of bytes written=3707813  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4268735  
HDFS: Number of bytes written=794075  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17625  
Total time spent by all reduces in occupied slots (ms)=17690  
Total time spent by all map tasks (ms)=3525  
Total time spent by all reduce tasks (ms)=3538  
Total vcore-milliseconds taken by all map tasks=3525  
Total vcore-milliseconds taken by all reduce tasks=3538  
Total megabyte-milliseconds taken by all map tasks=18048000  
Total megabyte-milliseconds taken by all reduce tasks=18114560

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=1587750  
Map output materialized bytes=1587761  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=1587761  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=3850  
Physical memory (bytes) snapshot=742092800  
Virtual memory (bytes) snapshot=12764909568  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418504704  
Peak Map Virtual memory (bytes)=6373388288  
Peak Reduce Physical memory (bytes)=323588096  
Peak Reduce Virtual memory (bytes)=6391521280

#### 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=2680027

#### File Output Format Counters

Bytes Written=794075

23/01/04 14:56:10 INFO KMeansDriver: Clustering data  
23/01/04 14:56:10 INFO KMeansDriver: Running Clustering  
23/01/04 14:56:10 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/01 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/01  
23/01/04 14:56:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:56:10 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0259  
23/01/04 14:56:10 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:56:10 INFO JobSubmitter: number of splits:1  
23/01/04 14:56:10 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0259  
23/01/04 14:56:10 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:56:10 INFO YarnClientImpl: Submitted application application\_1672817468245\_0259  
23/01/04 14:56:10 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0259/  
23/01/04 14:56:10 INFO Job: Running job: job\_1672817468245\_0259  
23/01/04 14:56:16 INFO Job: Job job\_1672817468245\_0259 running in uber mode : false  
23/01/04 14:56:16 INFO Job: map 0% reduce 0%  
23/01/04 14:56:21 INFO Job: map 100% reduce 0%  
23/01/04 14:56:21 INFO Job: Job job\_1672817468245\_0259 completed successfully  
23/01/04 14:56:22 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3474466  
HDFS: Number of bytes written=2680734  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17055  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3411  
Total vcore-milliseconds taken by all map tasks=3411  
Total megabyte-milliseconds taken by all map tasks=17464320

#### Map-Reduce Framework



```

Map input records=27
Map output records=27
Input split bytes=170
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=16
CPU time spent (ms)=1860
Physical memory (bytes) snapshot=287490048
Virtual memory (bytes) snapshot=6379646976
Total committed heap usage (bytes)=1052770304
Peak Map Physical memory (bytes)=287490048
Peak Map Virtual memory (bytes)=6379646976
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680734
23/01/04 14:56:22 INFO MahoutDriver: Program took 68899 ms (Minutes:
1.1483166666666667)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 14:56:27 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/EuclideanDistanceMeasure/01/clusters-3-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
EuclideanDistanceMeasure/01.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu
re/01/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 14:56:29 INFO HadoopUtil: Deleting tmp/representative
23/01/04 14:56:29 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/01/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
01/clusters-3-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:56:29 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:56:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:56:29 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0262
23/01/04 14:56:29 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:56:29 INFO JobSubmitter: number of splits:1
23/01/04 14:56:30 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0262
23/01/04 14:56:30 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:56:30 INFO Configuration: resource-types.xml not found
23/01/04 14:56:30 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:56:30 INFO YarnClientImpl: Submitted application
application_1672817468245_0262
23/01/04 14:56:30 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0262/

```

23/01/04 14:56:30 INFO Job: Running job: job\_1672817468245\_0262  
23/01/04 14:56:36 INFO Job: Job job\_1672817468245\_0262 running in uber mode : false  
23/01/04 14:56:36 INFO Job: map 0% reduce 0%  
23/01/04 14:56:42 INFO Job: map 100% reduce 0%  
23/01/04 14:56:46 INFO Job: map 100% reduce 100%  
23/01/04 14:56:46 INFO Job: Job job\_1672817468245\_0262 completed successfully  
23/01/04 14:56:46 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=134189  
FILE: Number of bytes written=800971  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3474961  
HDFS: Number of bytes written=531210  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16580  
Total time spent by all reduces in occupied slots (ms)=8730  
Total time spent by all map tasks (ms)=3316  
Total time spent by all reduce tasks (ms)=1746  
Total vcore-milliseconds taken by all map tasks=3316  
Total vcore-milliseconds taken by all reduce tasks=1746  
Total megabyte-milliseconds taken by all map tasks=16977920  
Total megabyte-milliseconds taken by all reduce tasks=8939520

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=134178  
Map output materialized bytes=134189  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=134189  
Reduce input records=1  
Reduce output records=2  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=46  
CPU time spent (ms)=2320  
Physical memory (bytes) snapshot=683556864  
Virtual memory (bytes) snapshot=12763729920  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=393281536  
Peak Map Virtual memory (bytes)=6374436864  
Peak Reduce Physical memory (bytes)=290275328  
Peak Reduce Virtual memory (bytes)=6392987648

#### 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=2680734
File Output Format Counters
    Bytes Written=531210
23/01/04 14:56:46 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 14:56:46 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:56:46 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0265
23/01/04 14:56:46 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:56:46 INFO JobSubmitter: number of splits:1
23/01/04 14:56:46 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0265
23/01/04 14:56:46 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:56:46 INFO YarnClientImpl: Submitted application
application_1672817468245_0265
23/01/04 14:56:46 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0265/
23/01/04 14:56:46 INFO Job: Running job: job_1672817468245_0265
23/01/04 14:56:52 INFO Job: Job job_1672817468245_0265 running in uber mode : false
23/01/04 14:56:52 INFO Job:  map 0% reduce 0%
23/01/04 14:56:57 INFO Job:  map 100% reduce 0%
23/01/04 14:57:03 INFO Job:  map 100% reduce 100%
23/01/04 14:57:04 INFO Job: Job job_1672817468245_0265 completed successfully
23/01/04 14:57:04 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=150434
    FILE: Number of bytes written=833461
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=3743357
    HDFS: Number of bytes written=681653
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=16390
    Total time spent by all reduces in occupied slots (ms)=16385
    Total time spent by all map tasks (ms)=3278
    Total time spent by all reduce tasks (ms)=3277
    Total vcore-milliseconds taken by all map tasks=3278
    Total vcore-milliseconds taken by all reduce tasks=3277
    Total megabyte-milliseconds taken by all map tasks=16783360
    Total megabyte-milliseconds taken by all reduce tasks=16778240
Map-Reduce Framework
    Map input records=27
    Map output records=1
    Map output bytes=150423
    Map output materialized bytes=150434

```

Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=150434  
Reduce input records=1  
Reduce output records=3  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=30  
CPU time spent (ms)=2510  
Physical memory (bytes) snapshot=691339264  
Virtual memory (bytes) snapshot=12760875008  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400711680  
Peak Map Virtual memory (bytes)=6385250304  
Peak Reduce Physical memory (bytes)=290627584  
Peak Reduce Virtual memory (bytes)=6379524096

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=2680734

File Output Format Counters

Bytes Written=681653

23/01/04 14:57:04 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 14:57:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:57:04 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0269  
23/01/04 14:57:04 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:57:04 INFO JobSubmitter: number of splits:1  
23/01/04 14:57:04 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0269  
23/01/04 14:57:04 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:57:04 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0269  
23/01/04 14:57:04 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0269/](http://lena-master:8088/proxy/application_1672817468245_0269/)  
23/01/04 14:57:04 INFO Job: Running job: job\_1672817468245\_0269  
23/01/04 14:57:10 INFO Job: Job job\_1672817468245\_0269 running in uber mode : false  
23/01/04 14:57:10 INFO Job: map 0% reduce 0%  
23/01/04 14:57:15 INFO Job: map 100% reduce 0%  
23/01/04 14:57:21 INFO Job: map 100% reduce 100%  
23/01/04 14:57:21 INFO Job: Job job\_1672817468245\_0269 completed successfully  
23/01/04 14:57:21 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=113177  
FILE: Number of bytes written=758947  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0

HDFS: Number of bytes read=4044243  
HDFS: Number of bytes written=794839  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16520  
Total time spent by all reduces in occupied slots (ms)=16175  
Total time spent by all map tasks (ms)=3304  
Total time spent by all reduce tasks (ms)=3235  
Total vcore-milliseconds taken by all map tasks=3304  
Total vcore-milliseconds taken by all reduce tasks=3235  
Total megabyte-milliseconds taken by all map tasks=16916480  
Total megabyte-milliseconds taken by all reduce tasks=16563200

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=113166  
Map output materialized bytes=113177  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=113177  
Reduce input records=1  
Reduce output records=4  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=2640  
Physical memory (bytes) snapshot=685817856  
Virtual memory (bytes) snapshot=12749152256  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=393826304  
Peak Map Virtual memory (bytes)=6367924224  
Peak Reduce Physical memory (bytes)=291991552  
Peak Reduce Virtual memory (bytes)=6385135616

#### 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=2680734

#### File Output Format Counters

Bytes Written=794839

23/01/04 14:57:21 INFO RepresentativePointsDriver: Representative Points Iteration  
3

23/01/04 14:57:21 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:57:21 INFO JobResourceUploader: Disabling Erasure Coding for path:

/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0272  
23/01/04 14:57:21 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:57:21 INFO JobSubmitter: number of splits:1  
23/01/04 14:57:21 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0272  
23/01/04 14:57:21 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:57:21 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0272  
23/01/04 14:57:21 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0272/  
23/01/04 14:57:21 INFO Job: Running job: job\_1672817468245\_0272  
23/01/04 14:57:28 INFO Job: Job job\_1672817468245\_0272 running in uber mode : false  
23/01/04 14:57:28 INFO Job: map 0% reduce 0%  
23/01/04 14:57:33 INFO Job: map 100% reduce 0%  
23/01/04 14:57:39 INFO Job: map 100% reduce 100%  
23/01/04 14:57:39 INFO Job: Job job\_1672817468245\_0272 completed successfully  
23/01/04 14:57:39 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=134189  
FILE: Number of bytes written=800971  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4270615  
HDFS: Number of bytes written=929037  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16920  
Total time spent by all reduces in occupied slots (ms)=16540  
Total time spent by all map tasks (ms)=3384  
Total time spent by all reduce tasks (ms)=3308  
Total vcore-milliseconds taken by all map tasks=3384  
Total vcore-milliseconds taken by all reduce tasks=3308  
Total megabyte-milliseconds taken by all map tasks=17326080  
Total megabyte-milliseconds taken by all reduce tasks=16936960

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=134178  
Map output materialized bytes=134189  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=134189  
Reduce input records=1  
Reduce output records=5  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=2710

Physical memory (bytes) snapshot=700874752  
Virtual memory (bytes) snapshot=12756733952  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404361216  
Peak Map Virtual memory (bytes)=6375100416  
Peak Reduce Physical memory (bytes)=296513536  
Peak Reduce Virtual memory (bytes)=6385549312

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=2680734

File Output Format Counters

Bytes Written=929037

23/01/04 14:57:39 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 14:57:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:57:39 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0275

23/01/04 14:57:39 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:57:39 INFO JobSubmitter: number of splits:1

23/01/04 14:57:39 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0275

23/01/04 14:57:39 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:57:39 INFO YarnClientImpl: Submitted application application\_1672817468245\_0275

23/01/04 14:57:39 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0275/

23/01/04 14:57:39 INFO Job: Running job: job\_1672817468245\_0275

23/01/04 14:57:45 INFO Job: Job job\_1672817468245\_0275 running in uber mode : false

23/01/04 14:57:45 INFO Job: map 0% reduce 0%

23/01/04 14:57:51 INFO Job: map 100% reduce 0%

23/01/04 14:57:56 INFO Job: map 100% reduce 100%

23/01/04 14:57:57 INFO Job: Job job\_1672817468245\_0275 completed successfully

23/01/04 14:57:57 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=85326  
FILE: Number of bytes written=703245  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4539011  
HDFS: Number of bytes written=1014352  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16905  
Total time spent by all reduces in occupied slots (ms)=16395  
Total time spent by all map tasks (ms)=3381

Total time spent by all reduce tasks (ms)=3279  
Total vcore-milliseconds taken by all map tasks=3381  
Total vcore-milliseconds taken by all reduce tasks=3279  
Total megabyte-milliseconds taken by all map tasks=17310720  
Total megabyte-milliseconds taken by all reduce tasks=16788480

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=85315  
Map output materialized bytes=85326  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=85326  
Reduce input records=1  
Reduce output records=6  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=2620  
Physical memory (bytes) snapshot=670396416  
Virtual memory (bytes) snapshot=12747067392  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=387469312  
Peak Map Virtual memory (bytes)=6375501824  
Peak Reduce Physical memory (bytes)=282927104  
Peak Reduce Virtual memory (bytes)=6375473152

#### 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=2680734

#### File Output Format Counters

Bytes Written=1014352

23/01/04 14:57:57 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN

23/01/04 14:57:57 INFO ClusterEvaluator: Intra-Cluster Density[23] =

0.7151365780668429

23/01/04 14:57:57 INFO ClusterEvaluator: Average Intra-Cluster Density =

0.7151365780668429

23/01/04 14:57:57 INFO ClusterDumper: Wrote 1 clusters

23/01/04 14:57:57 INFO MahoutDriver: Program took 89988 ms (Minutes: 1.4998)

---- K: 02 --

dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/02

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:58:02 INFO AbstractJob: Command line arguments: {--clustering=null, --

clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/

EuclideanDistanceMeasure], --convergenceDelta=[0.1], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --

endPhase=[2147483647],



```
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[2],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/02], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 14:58:02 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/02
23/01/04 14:58:02 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 14:58:03 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 14:58:03 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 14:58:03 INFO RandomSeedGenerator: Wrote 2 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 14:58:03 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/02
23/01/04 14:58:03 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 14:58:03 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 14:58:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:58:03 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0279
23/01/04 14:58:04 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:58:04 INFO JobSubmitter: number of splits:1
23/01/04 14:58:04 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0279
23/01/04 14:58:04 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:58:04 INFO Configuration: resource-types.xml not found
23/01/04 14:58:04 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:58:04 INFO YarnClientImpl: Submitted application
application_1672817468245_0279
23/01/04 14:58:04 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0279/
23/01/04 14:58:04 INFO Job: Running job: job_1672817468245_0279
23/01/04 14:58:10 INFO Job: Job job_1672817468245_0279 running in uber mode : false
23/01/04 14:58:10 INFO Job:  map 0% reduce 0%
23/01/04 14:58:16 INFO Job:  map 100% reduce 0%
23/01/04 14:58:22 INFO Job:  map 100% reduce 100%
23/01/04 14:58:22 INFO Job: Job job_1672817468245_0279 completed successfully
23/01/04 14:58:22 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=1071155
FILE: Number of bytes written=2674601
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=3702383
HDFS: Number of bytes written=847249
HDFS: Number of read operations=23
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
```

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17300  
Total time spent by all reduces in occupied slots (ms)=17145  
Total time spent by all map tasks (ms)=3460  
Total time spent by all reduce tasks (ms)=3429  
Total vcore-milliseconds taken by all map tasks=3460  
Total vcore-milliseconds taken by all reduce tasks=3429  
Total megabyte-milliseconds taken by all map tasks=17715200  
Total megabyte-milliseconds taken by all reduce tasks=17556480

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=1071139  
Map output materialized bytes=1071155  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=1071155  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=3570  
Physical memory (bytes) snapshot=732602368  
Virtual memory (bytes) snapshot=12759064576  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403148800  
Peak Map Virtual memory (bytes)=6374002688  
Peak Reduce Physical memory (bytes)=329453568  
Peak Reduce Virtual memory (bytes)=6385061888

#### 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=2680027

#### File Output Format Counters

Bytes Written=847249

23/01/04 14:58:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:58:22 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0280  
23/01/04 14:58:22 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:58:23 INFO JobSubmitter: number of splits:1  
23/01/04 14:58:23 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0280  
23/01/04 14:58:23 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:58:23 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0280  
23/01/04 14:58:23 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0280/](http://lena-master:8088/proxy/application_1672817468245_0280/)

23/01/04 14:58:23 INFO Job: Running job: job\_1672817468245\_0280  
23/01/04 14:58:29 INFO Job: Job job\_1672817468245\_0280 running in uber mode : false  
23/01/04 14:58:29 INFO Job: map 0% reduce 0%  
23/01/04 14:58:36 INFO Job: map 100% reduce 0%  
23/01/04 14:58:40 INFO Job: map 100% reduce 100%  
23/01/04 14:58:41 INFO Job: Job job\_1672817468245\_0280 completed successfully  
23/01/04 14:58:41 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1747744  
FILE: Number of bytes written=4027779  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4375083  
HDFS: Number of bytes written=847249  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17235  
Total time spent by all reduces in occupied slots (ms)=10160  
Total time spent by all map tasks (ms)=3447  
Total time spent by all reduce tasks (ms)=2032  
Total vcore-milliseconds taken by all map tasks=3447  
Total vcore-milliseconds taken by all reduce tasks=2032  
Total megabyte-milliseconds taken by all map tasks=17648640  
Total megabyte-milliseconds taken by all reduce tasks=10403840

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=1747728  
Map output materialized bytes=1747744  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=1747744  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=69  
CPU time spent (ms)=3710  
Physical memory (bytes) snapshot=758833152  
Virtual memory (bytes) snapshot=12758507520  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418172928  
Peak Map Virtual memory (bytes)=6377213952  
Peak Reduce Physical memory (bytes)=340660224  
Peak Reduce Virtual memory (bytes)=6381293568

#### 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=2680027
File Output Format Counters
    Bytes Written=847249
23/01/04 14:58:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:58:41 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0281
23/01/04 14:58:41 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:58:41 INFO JobSubmitter: number of splits:1
23/01/04 14:58:41 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0281
23/01/04 14:58:41 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:58:41 INFO YarnClientImpl: Submitted application
application_1672817468245_0281
23/01/04 14:58:41 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0281/
23/01/04 14:58:41 INFO Job: Running job: job_1672817468245_0281
23/01/04 14:58:47 INFO Job: Job job_1672817468245_0281 running in uber mode : false
23/01/04 14:58:47 INFO Job:  map 0% reduce 0%
23/01/04 14:58:53 INFO Job:  map 100% reduce 0%
23/01/04 14:58:59 INFO Job:  map 100% reduce 100%
23/01/04 14:58:59 INFO Job: Job job_1672817468245_0281 completed successfully
23/01/04 14:58:59 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=1747744
    FILE: Number of bytes written=4027779
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4375083
    HDFS: Number of bytes written=847249
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17650
    Total time spent by all reduces in occupied slots (ms)=17385
    Total time spent by all map tasks (ms)=3530
    Total time spent by all reduce tasks (ms)=3477
    Total vcore-milliseconds taken by all map tasks=3530
    Total vcore-milliseconds taken by all reduce tasks=3477
    Total megabyte-milliseconds taken by all map tasks=18073600
    Total megabyte-milliseconds taken by all reduce tasks=17802240
Map-Reduce Framework
    Map input records=27
    Map output records=2
    Map output bytes=1747728
    Map output materialized bytes=1747744
    Input split bytes=170
    Combine input records=0

```

```

Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=1747744
Reduce input records=2
Reduce output records=2
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=49
CPU time spent (ms)=3560
Physical memory (bytes) snapshot=719859712
Virtual memory (bytes) snapshot=12757704704
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=405577728
Peak Map Virtual memory (bytes)=6375415808
Peak Reduce Physical memory (bytes)=314281984
Peak Reduce Virtual memory (bytes)=6382288896
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=2680027
File Output Format Counters
  Bytes Written=847249
23/01/04 14:58:59 INFO KMeansDriver: Clustering data
23/01/04 14:58:59 INFO KMeansDriver: Running Clustering
23/01/04 14:58:59 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/02 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/02
23/01/04 14:58:59 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:58:59 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0282
23/01/04 14:59:00 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:59:00 INFO JobSubmitter: number of splits:1
23/01/04 14:59:00 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0282
23/01/04 14:59:00 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:59:00 INFO YarnClientImpl: Submitted application
application_1672817468245_0282
23/01/04 14:59:00 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0282/
23/01/04 14:59:00 INFO Job: Running job: job_1672817468245_0282
23/01/04 14:59:06 INFO Job: Job job_1672817468245_0282 running in uber mode : false
23/01/04 14:59:06 INFO Job:  map 0% reduce 0%
23/01/04 14:59:12 INFO Job:  map 100% reduce 0%
23/01/04 14:59:12 INFO Job: Job job_1672817468245_0282 completed successfully
23/01/04 14:59:12 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265897

```

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3527640  
HDFS: Number of bytes written=2680724  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17475  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3495  
Total vcore-milliseconds taken by all map tasks=3495  
Total megabyte-milliseconds taken by all map tasks=17894400

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=12  
CPU time spent (ms)=1970  
Physical memory (bytes) snapshot=307646464  
Virtual memory (bytes) snapshot=6370529280  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=307646464  
Peak Map Virtual memory (bytes)=6370529280

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680724

23/01/04 14:59:12 INFO MahoutDriver: Program took 70244 ms (Minutes:

1.1707333333333334)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 14:59:18 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],

--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --

endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-

corpus-kmeans/EuclideanDistanceMeasure/02/clusters-3-final], --numWords=[20], --

output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/

EuclideanDistanceMeasure/02.txt], --outputFormat=[TEXT],

--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu

re/02/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 14:59:19 INFO HadoopUtil: Deleting tmp/representative

23/01/04 14:59:19 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/

EuclideanDistanceMeasure/02/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --

endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/

02/clusters-3-final], --maxIter=[5], --method=[mapreduce],

```
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 14:59:19 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 14:59:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 14:59:19 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0283
23/01/04 14:59:19 INFO FileInputFormat: Total input files to process : 1
23/01/04 14:59:20 INFO JobSubmitter: number of splits:1
23/01/04 14:59:20 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0283
23/01/04 14:59:20 INFO JobSubmitter: Executing with tokens: []
23/01/04 14:59:20 INFO Configuration: resource-types.xml not found
23/01/04 14:59:20 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 14:59:20 INFO YarnClientImpl: Submitted application
application_1672817468245_0283
23/01/04 14:59:20 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0283/
23/01/04 14:59:20 INFO Job: Running job: job_1672817468245_0283
23/01/04 14:59:26 INFO Job: Job job_1672817468245_0283 running in uber mode : false
23/01/04 14:59:26 INFO Job: map 0% reduce 0%
23/01/04 14:59:31 INFO Job: map 100% reduce 0%
23/01/04 14:59:37 INFO Job: map 100% reduce 100%
23/01/04 14:59:37 INFO Job: Job job_1672817468245_0283 completed successfully
23/01/04 14:59:37 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=188033
FILE: Number of bytes written=908659
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=3581797
HDFS: Number of bytes written=638473
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16735
Total time spent by all reduces in occupied slots (ms)=16645
Total time spent by all map tasks (ms)=3347
Total time spent by all reduce tasks (ms)=3329
Total vcore-milliseconds taken by all map tasks=3347
Total vcore-milliseconds taken by all reduce tasks=3329
Total megabyte-milliseconds taken by all map tasks=17136640
Total megabyte-milliseconds taken by all reduce tasks=17044480
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=2
Map output bytes=188018
Map output materialized bytes=188033
Input split bytes=203
Combine input records=0
Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=188033
```

Reduce input records=2  
Reduce output records=4  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=2500  
Physical memory (bytes) snapshot=678260736  
Virtual memory (bytes) snapshot=12761370624  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399060992  
Peak Map Virtual memory (bytes)=6377693184  
Peak Reduce Physical memory (bytes)=279199744  
Peak Reduce Virtual memory (bytes)=6387585024

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=2680724

File Output Format Counters

Bytes Written=638473

23/01/04 14:59:37 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 14:59:37 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 14:59:37 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0284

23/01/04 14:59:37 INFO FileInputFormat: Total input files to process : 1

23/01/04 14:59:37 INFO JobSubmitter: number of splits:1

23/01/04 14:59:37 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0284

23/01/04 14:59:37 INFO JobSubmitter: Executing with tokens: []

23/01/04 14:59:38 INFO YarnClientImpl: Submitted application application\_1672817468245\_0284

23/01/04 14:59:38 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0284/](http://lena-master:8088/proxy/application_1672817468245_0284/)

23/01/04 14:59:38 INFO Job: Running job: job\_1672817468245\_0284

23/01/04 14:59:44 INFO Job: Job job\_1672817468245\_0284 running in uber mode : false

23/01/04 14:59:44 INFO Job: map 0% reduce 0%

23/01/04 14:59:49 INFO Job: map 100% reduce 0%

23/01/04 14:59:55 INFO Job: map 100% reduce 100%

23/01/04 14:59:55 INFO Job: Job job\_1672817468245\_0284 completed successfully

23/01/04 14:59:55 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=204278

FILE: Number of bytes written=941149

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=3957873

HDFS: Number of bytes written=842776

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2



HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17015  
Total time spent by all reduces in occupied slots (ms)=16630  
Total time spent by all map tasks (ms)=3403  
Total time spent by all reduce tasks (ms)=3326  
Total vcore-milliseconds taken by all map tasks=3403  
Total vcore-milliseconds taken by all reduce tasks=3326  
Total megabyte-milliseconds taken by all map tasks=17423360  
Total megabyte-milliseconds taken by all reduce tasks=17029120

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=204263  
Map output materialized bytes=204278  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=204278  
Reduce input records=2  
Reduce output records=6  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=2650  
Physical memory (bytes) snapshot=675512320  
Virtual memory (bytes) snapshot=12747411456  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390840320  
Peak Map Virtual memory (bytes)=6375419904  
Peak Reduce Physical memory (bytes)=284672000  
Peak Reduce Virtual memory (bytes)=6375899136

#### 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=2680724

#### File Output Format Counters

Bytes Written=842776

23/01/04 14:59:55 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 14:59:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 14:59:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0285  
23/01/04 14:59:55 INFO FileInputFormat: Total input files to process : 1  
23/01/04 14:59:55 INFO JobSubmitter: number of splits:1  
23/01/04 14:59:55 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0285

23/01/04 14:59:55 INFO JobSubmitter: Executing with tokens: []  
23/01/04 14:59:55 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0285  
23/01/04 14:59:55 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0285/  
23/01/04 14:59:55 INFO Job: Running job: job\_1672817468245\_0285  
23/01/04 15:00:01 INFO Job: Job job\_1672817468245\_0285 running in uber mode : false  
23/01/04 15:00:01 INFO Job: map 0% reduce 0%  
23/01/04 15:00:07 INFO Job: map 100% reduce 0%  
23/01/04 15:00:12 INFO Job: map 100% reduce 100%  
23/01/04 15:00:13 INFO Job: Job job\_1672817468245\_0285 completed successfully  
23/01/04 15:00:13 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=167021  
FILE: Number of bytes written=866635  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4366479  
HDFS: Number of bytes written=1009802  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16835  
Total time spent by all reduces in occupied slots (ms)=16500  
Total time spent by all map tasks (ms)=3367  
Total time spent by all reduce tasks (ms)=3300  
Total vcore-milliseconds taken by all map tasks=3367  
Total vcore-milliseconds taken by all reduce tasks=3300  
Total megabyte-milliseconds taken by all map tasks=17239040  
Total megabyte-milliseconds taken by all reduce tasks=16896000

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=167006  
Map output materialized bytes=167021  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=167021  
Reduce input records=2  
Reduce output records=8  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=2830  
Physical memory (bytes) snapshot=681353216  
Virtual memory (bytes) snapshot=12752912384  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390918144  
Peak Map Virtual memory (bytes)=6375579648

Peak Reduce Physical memory (bytes)=290435072  
Peak Reduce Virtual memory (bytes)=6381240320  
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=2680724

File Output Format Counters

Bytes Written=1009802

23/01/04 15:00:13 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:00:13 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:00:13 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0286

23/01/04 15:00:13 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:00:14 INFO JobSubmitter: number of splits:1

23/01/04 15:00:14 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0286

23/01/04 15:00:14 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:00:14 INFO YarnClientImpl: Submitted application application\_1672817468245\_0286

23/01/04 15:00:14 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0286/

23/01/04 15:00:14 INFO Job: Running job: job\_1672817468245\_0286

23/01/04 15:00:20 INFO Job: Job job\_1672817468245\_0286 running in uber mode : false

23/01/04 15:00:20 INFO Job: map 0% reduce 0%

23/01/04 15:00:26 INFO Job: map 100% reduce 0%

23/01/04 15:00:32 INFO Job: map 100% reduce 100%

23/01/04 15:00:32 INFO Job: Job job\_1672817468245\_0286 completed successfully

23/01/04 15:00:32 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=188033

FILE: Number of bytes written=908659

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4700531

HDFS: Number of bytes written=1197860

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=17105

Total time spent by all reduces in occupied slots (ms)=16290

Total time spent by all map tasks (ms)=3421

Total time spent by all reduce tasks (ms)=3258

Total vcore-milliseconds taken by all map tasks=3421

Total vcore-milliseconds taken by all reduce tasks=3258

Total megabyte-milliseconds taken by all map tasks=17515520

Total megabyte-milliseconds taken by all reduce tasks=16680960

## Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=188018  
Map output materialized bytes=188033  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=188033  
Reduce input records=2  
Reduce output records=10  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=72  
CPU time spent (ms)=2810  
Physical memory (bytes) snapshot=684134400  
Virtual memory (bytes) snapshot=12748894208  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392794112  
Peak Map Virtual memory (bytes)=6374109184  
Peak Reduce Physical memory (bytes)=291340288  
Peak Reduce Virtual memory (bytes)=6378692608

## 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=2680724

## File Output Format Counters

Bytes Written=1197860

23/01/04 15:00:32 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:00:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:00:32 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0287

23/01/04 15:00:32 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:00:32 INFO JobSubmitter: number of splits:1

23/01/04 15:00:32 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0287

23/01/04 15:00:32 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:00:32 INFO YarnClientImpl: Submitted application application\_1672817468245\_0287

23/01/04 15:00:32 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0287/

23/01/04 15:00:32 INFO Job: Running job: job\_1672817468245\_0287

23/01/04 15:00:38 INFO Job: Job job\_1672817468245\_0287 running in uber mode : false

23/01/04 15:00:38 INFO Job: map 0% reduce 0%

23/01/04 15:00:44 INFO Job: map 100% reduce 0%

23/01/04 15:00:49 INFO Job: map 100% reduce 100%

23/01/04 15:00:50 INFO Job: Job job\_1672817468245\_0287 completed successfully

23/01/04 15:00:50 INFO Job: Counters: 54

## File System Counters

FILE: Number of bytes read=139170  
FILE: Number of bytes written=810933  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5076647  
HDFS: Number of bytes written=1337035  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17225  
Total time spent by all reduces in occupied slots (ms)=16200  
Total time spent by all map tasks (ms)=3445  
Total time spent by all reduce tasks (ms)=3240  
Total vcore-milliseconds taken by all map tasks=3445  
Total vcore-milliseconds taken by all reduce tasks=3240  
Total megabyte-milliseconds taken by all map tasks=17638400  
Total megabyte-milliseconds taken by all reduce tasks=16588800

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=139155  
Map output materialized bytes=139170  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=139170  
Reduce input records=2  
Reduce output records=12  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=2920  
Physical memory (bytes) snapshot=682811392  
Virtual memory (bytes) snapshot=12753092608  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=393687040  
Peak Map Virtual memory (bytes)=6369292288  
Peak Reduce Physical memory (bytes)=289124352  
Peak Reduce Virtual memory (bytes)=6387707904

#### 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=2680724

#### File Output Format Counters

Bytes Written=1337035

```

23/01/04 15:00:50 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN
23/01/04 15:00:51 INFO ClusterEvaluator: Intra-Cluster Density[3] = NaN
23/01/04 15:00:51 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.7147719780730082
23/01/04 15:00:51 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.7147719780730082
23/01/04 15:00:51 INFO ClusterDumper: Wrote 2 clusters
23/01/04 15:00:51 INFO MahoutDriver: Program took 93215 ms (Minutes:
1.5535833333333333)
---- K: 03 --
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/03
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:00:55 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[3],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/03], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:00:56 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/03
23/01/04 15:00:56 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:00:56 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:00:56 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:00:56 INFO RandomSeedGenerator: Wrote 3 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:00:56 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/03
23/01/04 15:00:56 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:00:56 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:00:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:00:57 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0288
23/01/04 15:00:57 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:00:57 INFO JobSubmitter: number of splits:1
23/01/04 15:00:57 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0288
23/01/04 15:00:57 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:00:57 INFO Configuration: resource-types.xml not found
23/01/04 15:00:57 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:00:57 INFO YarnClientImpl: Submitted application
application_1672817468245_0288
23/01/04 15:00:57 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0288/
23/01/04 15:00:57 INFO Job: Running job: job_1672817468245_0288

```

23/01/04 15:01:03 INFO Job: Job job\_1672817468245\_0288 running in uber mode : false  
23/01/04 15:01:03 INFO Job: map 0% reduce 0%  
23/01/04 15:01:09 INFO Job: map 100% reduce 0%  
23/01/04 15:01:15 INFO Job: map 100% reduce 100%  
23/01/04 15:01:16 INFO Job: Job job\_1672817468245\_0288 completed successfully  
23/01/04 15:01:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1710861  
FILE: Number of bytes written=3954013  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4727531  
HDFS: Number of bytes written=1220795  
HDFS: Number of read operations=27  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17890  
Total time spent by all reduces in occupied slots (ms)=17195  
Total time spent by all map tasks (ms)=3578  
Total time spent by all reduce tasks (ms)=3439  
Total vcore-milliseconds taken by all map tasks=3578  
Total vcore-milliseconds taken by all reduce tasks=3439  
Total megabyte-milliseconds taken by all map tasks=18319360  
Total megabyte-milliseconds taken by all reduce tasks=17607680

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=1710840  
Map output materialized bytes=1710861  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=1710861  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=81  
CPU time spent (ms)=3690  
Physical memory (bytes) snapshot=710029312  
Virtual memory (bytes) snapshot=12762869760  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398086144  
Peak Map Virtual memory (bytes)=6375919616  
Peak Reduce Physical memory (bytes)=311943168  
Peak Reduce Virtual memory (bytes)=6386950144

#### 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=2680027
    File Output Format Counters
        Bytes Written=1220795
23/01/04 15:01:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:01:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0289
23/01/04 15:01:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:01:16 INFO JobSubmitter: number of splits:1
23/01/04 15:01:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0289
23/01/04 15:01:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:01:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0289
23/01/04 15:01:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0289/
23/01/04 15:01:16 INFO Job: Running job: job_1672817468245_0289
23/01/04 15:01:22 INFO Job: Job job_1672817468245_0289 running in uber mode : false
23/01/04 15:01:22 INFO Job:  map 0% reduce 0%
23/01/04 15:01:29 INFO Job:  map 100% reduce 0%
23/01/04 15:01:36 INFO Job:  map 100% reduce 100%
23/01/04 15:01:36 INFO Job: Job job_1672817468245_0289 completed successfully
23/01/04 15:01:36 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2589937
        FILE: Number of bytes written=5712165
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5122175
        HDFS: Number of bytes written=1220795
        HDFS: Number of read operations=19
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17965
        Total time spent by all reduces in occupied slots (ms)=17525
        Total time spent by all map tasks (ms)=3593
        Total time spent by all reduce tasks (ms)=3505
        Total vcore-milliseconds taken by all map tasks=3593
        Total vcore-milliseconds taken by all reduce tasks=3505
        Total megabyte-milliseconds taken by all map tasks=18396160
        Total megabyte-milliseconds taken by all reduce tasks=17945600
    Map-Reduce Framework
        Map input records=27
        Map output records=3
        Map output bytes=2589916
        Map output materialized bytes=2589937
        Input split bytes=170
        Combine input records=0
        Combine output records=0

```



Reduce input groups=3  
Reduce shuffle bytes=2589937  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=3760  
Physical memory (bytes) snapshot=738893824  
Virtual memory (bytes) snapshot=12767686656  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395431936  
Peak Map Virtual memory (bytes)=6380666880  
Peak Reduce Physical memory (bytes)=343461888  
Peak Reduce Virtual memory (bytes)=6387019776

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=2680027

File Output Format Counters

Bytes Written=1220795

23/01/04 15:01:36 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:01:36 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0290  
23/01/04 15:01:36 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:01:36 INFO JobSubmitter: number of splits:1  
23/01/04 15:01:36 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0290  
23/01/04 15:01:36 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:01:37 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0290  
23/01/04 15:01:37 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0290/](http://lena-master:8088/proxy/application_1672817468245_0290/)  
23/01/04 15:01:37 INFO Job: Running job: job\_1672817468245\_0290  
23/01/04 15:01:43 INFO Job: Job job\_1672817468245\_0290 running in uber mode : false  
23/01/04 15:01:43 INFO Job: map 0% reduce 0%  
23/01/04 15:01:49 INFO Job: map 100% reduce 0%  
23/01/04 15:01:55 INFO Job: map 100% reduce 100%  
23/01/04 15:01:55 INFO Job: Job job\_1672817468245\_0290 completed successfully  
23/01/04 15:01:55 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2528113  
FILE: Number of bytes written=5588517  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5122175  
HDFS: Number of bytes written=1196023  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17600  
  Total time spent by all reduces in occupied slots (ms)=17525  
  Total time spent by all map tasks (ms)=3520  
  Total time spent by all reduce tasks (ms)=3505  
  Total vcore-milliseconds taken by all map tasks=3520  
  Total vcore-milliseconds taken by all reduce tasks=3505  
  Total megabyte-milliseconds taken by all map tasks=18022400  
  Total megabyte-milliseconds taken by all reduce tasks=17945600

Map-Reduce Framework  
  Map input records=27  
  Map output records=3  
  Map output bytes=2528092  
  Map output materialized bytes=2528113  
  Input split bytes=170  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=3  
  Reduce shuffle bytes=2528113  
  Reduce input records=3  
  Reduce output records=3  
  Spilled Records=6  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=56  
  CPU time spent (ms)=3790  
  Physical memory (bytes) snapshot=734695424  
  Virtual memory (bytes) snapshot=12758056960  
  Total committed heap usage (bytes)=2105540608  
  Peak Map Physical memory (bytes)=393420800  
  Peak Map Virtual memory (bytes)=6373744640  
  Peak Reduce Physical memory (bytes)=341274624  
  Peak Reduce Virtual memory (bytes)=6384312320

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=2680027

File Output Format Counters

  Bytes Written=1196023

23/01/04 15:01:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:01:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0291  
23/01/04 15:01:55 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:01:55 INFO JobSubmitter: number of splits:1  
23/01/04 15:01:55 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0291  
23/01/04 15:01:55 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:01:55 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0291  
23/01/04 15:01:55 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0291/  
23/01/04 15:01:55 INFO Job: Running job: job\_1672817468245\_0291  
23/01/04 15:02:01 INFO Job: Job job\_1672817468245\_0291 running in uber mode : false  
23/01/04 15:02:01 INFO Job: map 0% reduce 0%  
23/01/04 15:02:07 INFO Job: map 100% reduce 0%  
23/01/04 15:02:11 INFO Job: map 100% reduce 100%  
23/01/04 15:02:11 INFO Job: Job job\_1672817468245\_0291 completed successfully  
23/01/04 15:02:11 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2503341  
FILE: Number of bytes written=5538973  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5072631  
HDFS: Number of bytes written=1196023  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17545  
Total time spent by all reduces in occupied slots (ms)=9780  
Total time spent by all map tasks (ms)=3509  
Total time spent by all reduce tasks (ms)=1956  
Total vcore-milliseconds taken by all map tasks=3509  
Total vcore-milliseconds taken by all reduce tasks=1956  
Total megabyte-milliseconds taken by all map tasks=17966080  
Total megabyte-milliseconds taken by all reduce tasks=10014720

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=2503320  
Map output materialized bytes=2503341  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=2503341  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65  
CPU time spent (ms)=3770  
Physical memory (bytes) snapshot=733519872  
Virtual memory (bytes) snapshot=12754309120  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403075072  
Peak Map Virtual memory (bytes)=6373777408  
Peak Reduce Physical memory (bytes)=330444800  
Peak Reduce Virtual memory (bytes)=6380531712

#### 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=2680027

#### File Output Format Counters

Bytes Written=1196023

23/01/04 15:02:11 INFO KMeansDriver: Clustering data  
23/01/04 15:02:11 INFO KMeansDriver: Running Clustering  
23/01/04 15:02:11 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/03 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/03  
23/01/04 15:02:11 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:02:11 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0292  
23/01/04 15:02:12 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:02:12 INFO JobSubmitter: number of splits:1  
23/01/04 15:02:12 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0292  
23/01/04 15:02:12 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:02:12 INFO YarnClientImpl: Submitted application application\_1672817468245\_0292  
23/01/04 15:02:12 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0292/  
23/01/04 15:02:12 INFO Job: Running job: job\_1672817468245\_0292  
23/01/04 15:02:18 INFO Job: Job job\_1672817468245\_0292 running in uber mode : false  
23/01/04 15:02:18 INFO Job: map 0% reduce 0%  
23/01/04 15:02:23 INFO Job: map 100% reduce 0%  
23/01/04 15:02:24 INFO Job: Job job\_1672817468245\_0292 completed successfully  
23/01/04 15:02:24 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3876414  
HDFS: Number of bytes written=2680702  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17490  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3498  
Total vcore-milliseconds taken by all map tasks=3498  
Total megabyte-milliseconds taken by all map tasks=17909760

#### Map-Reduce Framework

```

Map input records=27
Map output records=27
Input split bytes=170
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=33
CPU time spent (ms)=2190
Physical memory (bytes) snapshot=291446784
Virtual memory (bytes) snapshot=6376882176
Total committed heap usage (bytes)=1052770304
Peak Map Physical memory (bytes)=291446784
Peak Map Virtual memory (bytes)=6376882176
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680702
23/01/04 15:02:24 INFO MahoutDriver: Program took 88612 ms (Minutes:
1.4768666666666668)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:02:30 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/EuclideanDistanceMeasure/03/clusters-4-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
EuclideanDistanceMeasure/03.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu
re/03/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:02:31 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:02:31 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/03/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
03/clusters-4-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:02:31 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:02:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:02:32 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0293
23/01/04 15:02:32 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:02:32 INFO JobSubmitter: number of splits:1
23/01/04 15:02:32 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0293
23/01/04 15:02:32 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:02:32 INFO Configuration: resource-types.xml not found
23/01/04 15:02:32 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:02:32 INFO YarnClientImpl: Submitted application
application_1672817468245_0293
23/01/04 15:02:32 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0293/

```

23/01/04 15:02:32 INFO Job: Running job: job\_1672817468245\_0293  
23/01/04 15:02:38 INFO Job: Job job\_1672817468245\_0293 running in uber mode : false  
23/01/04 15:02:38 INFO Job: map 0% reduce 0%  
23/01/04 15:02:44 INFO Job: map 100% reduce 0%  
23/01/04 15:02:51 INFO Job: map 100% reduce 100%  
23/01/04 15:02:51 INFO Job: Job job\_1672817468245\_0293 completed successfully  
23/01/04 15:02:51 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=428900  
FILE: Number of bytes written=1390393  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3988755  
HDFS: Number of bytes written=1082864  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16735  
Total time spent by all reduces in occupied slots (ms)=16615  
Total time spent by all map tasks (ms)=3347  
Total time spent by all reduce tasks (ms)=3323  
Total vcore-milliseconds taken by all map tasks=3347  
Total vcore-milliseconds taken by all reduce tasks=3323  
Total megabyte-milliseconds taken by all map tasks=17136640  
Total megabyte-milliseconds taken by all reduce tasks=17013760

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=428879  
Map output materialized bytes=428900  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=428900  
Reduce input records=3  
Reduce output records=6  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=2670  
Physical memory (bytes) snapshot=699441152  
Virtual memory (bytes) snapshot=12765802496  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396881920  
Peak Map Virtual memory (bytes)=6382698496  
Peak Reduce Physical memory (bytes)=302559232  
Peak Reduce Virtual memory (bytes)=6387019776

#### 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=2680702  
File Output Format Counters  
  Bytes Written=1082864  
23/01/04 15:02:51 INFO RepresentativePointsDriver: Representative Points Iteration  
1  
23/01/04 15:02:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:02:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0294  
23/01/04 15:02:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:02:51 INFO JobSubmitter: number of splits:1  
23/01/04 15:02:51 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0294  
23/01/04 15:02:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:02:51 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0294  
23/01/04 15:02:51 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0294/  
23/01/04 15:02:51 INFO Job: Running job: job\_1672817468245\_0294  
23/01/04 15:02:57 INFO Job: Job job\_1672817468245\_0294 running in uber mode : false  
23/01/04 15:02:57 INFO Job: map 0% reduce 0%  
23/01/04 15:03:02 INFO Job: map 100% reduce 0%  
23/01/04 15:03:08 INFO Job: map 100% reduce 100%  
23/01/04 15:03:08 INFO Job: Job job\_1672817468245\_0294 completed successfully  
23/01/04 15:03:08 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=391643  
  FILE: Number of bytes written=1315879  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=4846633  
  HDFS: Number of bytes written=1474546  
  HDFS: Number of read operations=15  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17090  
  Total time spent by all reduces in occupied slots (ms)=16835  
  Total time spent by all map tasks (ms)=3418  
  Total time spent by all reduce tasks (ms)=3367  
  Total vcore-milliseconds taken by all map tasks=3418  
  Total vcore-milliseconds taken by all reduce tasks=3367  
  Total megabyte-milliseconds taken by all map tasks=17500160  
  Total megabyte-milliseconds taken by all reduce tasks=17239040  
Map-Reduce Framework  
  Map input records=27  
  Map output records=3  
  Map output bytes=391622  
  Map output materialized bytes=391643

Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=391643  
Reduce input records=3  
Reduce output records=9  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=3250  
Physical memory (bytes) snapshot=699596800  
Virtual memory (bytes) snapshot=12762980352  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402071552  
Peak Map Virtual memory (bytes)=6380331008  
Peak Reduce Physical memory (bytes)=297525248  
Peak Reduce Virtual memory (bytes)=6384607232

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=2680702

File Output Format Counters

Bytes Written=1474546

23/01/04 15:03:08 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 15:03:08 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:03:08 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0295  
23/01/04 15:03:08 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:03:08 INFO JobSubmitter: number of splits:1  
23/01/04 15:03:08 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0295  
23/01/04 15:03:08 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:03:08 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0295  
23/01/04 15:03:08 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0295/](http://lena-master:8088/proxy/application_1672817468245_0295/)  
23/01/04 15:03:08 INFO Job: Running job: job\_1672817468245\_0295  
23/01/04 15:03:15 INFO Job: Job job\_1672817468245\_0295 running in uber mode : false  
23/01/04 15:03:15 INFO Job: map 0% reduce 0%  
23/01/04 15:03:21 INFO Job: map 100% reduce 0%  
23/01/04 15:03:27 INFO Job: map 100% reduce 100%  
23/01/04 15:03:27 INFO Job: Job job\_1672817468245\_0295 completed successfully  
23/01/04 15:03:27 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=363792  
FILE: Number of bytes written=1260177  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0



HDFS: Number of bytes read=5629997  
HDFS: Number of bytes written=1838357  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16490  
Total time spent by all reduces in occupied slots (ms)=16665  
Total time spent by all map tasks (ms)=3298  
Total time spent by all reduce tasks (ms)=3333  
Total vcore-milliseconds taken by all map tasks=3298  
Total vcore-milliseconds taken by all reduce tasks=3333  
Total megabyte-milliseconds taken by all map tasks=16885760  
Total megabyte-milliseconds taken by all reduce tasks=17064960

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=363771  
Map output materialized bytes=363792  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=363792  
Reduce input records=3  
Reduce output records=12  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=2800  
Physical memory (bytes) snapshot=693972992  
Virtual memory (bytes) snapshot=12755742720  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394563584  
Peak Map Virtual memory (bytes)=6378156032  
Peak Reduce Physical memory (bytes)=299409408  
Peak Reduce Virtual memory (bytes)=6379544576

#### 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=2680702

#### File Output Format Counters

Bytes Written=1838357

23/01/04 15:03:27 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:03:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:03:27 INFO JobResourceUploader: Disabling Erasure Coding for path:

/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0296  
23/01/04 15:03:27 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:03:27 INFO JobSubmitter: number of splits:1  
23/01/04 15:03:27 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0296  
23/01/04 15:03:27 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:03:27 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0296  
23/01/04 15:03:27 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0296/  
23/01/04 15:03:27 INFO Job: Running job: job\_1672817468245\_0296  
23/01/04 15:03:33 INFO Job: Job job\_1672817468245\_0296 running in uber mode : false  
23/01/04 15:03:33 INFO Job: map 0% reduce 0%  
23/01/04 15:03:38 INFO Job: map 100% reduce 0%  
23/01/04 15:03:44 INFO Job: map 100% reduce 100%  
23/01/04 15:03:44 INFO Job: Job job\_1672817468245\_0296 completed successfully  
23/01/04 15:03:44 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=389764  
FILE: Number of bytes written=1312121  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6357619  
HDFS: Number of bytes written=2228160  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17095  
Total time spent by all reduces in occupied slots (ms)=16250  
Total time spent by all map tasks (ms)=3419  
Total time spent by all reduce tasks (ms)=3250  
Total vcore-milliseconds taken by all map tasks=3419  
Total vcore-milliseconds taken by all reduce tasks=3250  
Total megabyte-milliseconds taken by all map tasks=17505280  
Total megabyte-milliseconds taken by all reduce tasks=16640000

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=389743  
Map output materialized bytes=389764  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=389764  
Reduce input records=3  
Reduce output records=15  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=3160

Physical memory (bytes) snapshot=707530752  
Virtual memory (bytes) snapshot=12763947008  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409853952  
Peak Map Virtual memory (bytes)=6373150720  
Peak Reduce Physical memory (bytes)=297676800  
Peak Reduce Virtual memory (bytes)=6394703872

#### 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=2680702

#### File Output Format Counters

Bytes Written=2228160

23/01/04 15:03:44 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:03:44 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:03:44 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0297

23/01/04 15:03:44 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:03:44 INFO JobSubmitter: number of splits:1

23/01/04 15:03:44 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0297

23/01/04 15:03:44 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:03:44 INFO YarnClientImpl: Submitted application application\_1672817468245\_0297

23/01/04 15:03:44 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0297/

23/01/04 15:03:44 INFO Job: Running job: job\_1672817468245\_0297

23/01/04 15:03:51 INFO Job: Job job\_1672817468245\_0297 running in uber mode : false

23/01/04 15:03:51 INFO Job: map 0% reduce 0%

23/01/04 15:03:57 INFO Job: map 100% reduce 0%

23/01/04 15:04:02 INFO Job: map 100% reduce 100%

23/01/04 15:04:03 INFO Job: Job job\_1672817468245\_0297 completed successfully

23/01/04 15:04:03 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=387797  
FILE: Number of bytes written=1308187  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7137225  
HDFS: Number of bytes written=2615996  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16945  
Total time spent by all reduces in occupied slots (ms)=16915  
Total time spent by all map tasks (ms)=3389

Total time spent by all reduce tasks (ms)=3383  
Total vcore-milliseconds taken by all map tasks=3389  
Total vcore-milliseconds taken by all reduce tasks=3383  
Total megabyte-milliseconds taken by all map tasks=17351680  
Total megabyte-milliseconds taken by all reduce tasks=17320960

Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=387776  
Map output materialized bytes=387797  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=387797  
Reduce input records=3  
Reduce output records=18  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=29  
CPU time spent (ms)=3250  
Physical memory (bytes) snapshot=699797504  
Virtual memory (bytes) snapshot=12760817664  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411226112  
Peak Map Virtual memory (bytes)=6376914944  
Peak Reduce Physical memory (bytes)=288571392  
Peak Reduce Virtual memory (bytes)=6383902720

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=2680702

File Output Format Counters

Bytes Written=2615996

23/01/04 15:04:03 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.6538686552003133

23/01/04 15:04:03 INFO ClusterEvaluator: Intra-Cluster Density[18] = NaN

23/01/04 15:04:03 INFO ClusterEvaluator: Intra-Cluster Density[23] = NaN

23/01/04 15:04:03 INFO ClusterEvaluator: Intra-Cluster Density[6] =

0.5686641540304515

23/01/04 15:04:03 INFO ClusterEvaluator: Average Intra-Cluster Density =

0.5686641540304515

23/01/04 15:04:03 INFO ClusterDumper: Wrote 3 clusters

23/01/04 15:04:03 INFO MahoutDriver: Program took 93461 ms (Minutes:

1.5576833333333333)

---- K: 04 --

dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/04

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:04:08 INFO AbstractJob: Command line arguments: {--clustering=null, --

```

clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[4],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/04], --overwrite=null, --startPhase=[0], --
tempDir=[temp]]
23/01/04 15:04:08 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/04
23/01/04 15:04:08 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:04:09 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:04:09 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:04:09 INFO RandomSeedGenerator: Wrote 4 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:04:09 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/04
23/01/04 15:04:09 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:04:09 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:04:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:04:09 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0298
23/01/04 15:04:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:04:09 INFO JobSubmitter: number of splits:1
23/01/04 15:04:10 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0298
23/01/04 15:04:10 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:04:10 INFO Configuration: resource-types.xml not found
23/01/04 15:04:10 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:04:10 INFO YarnClientImpl: Submitted application
application_1672817468245_0298
23/01/04 15:04:10 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0298/
23/01/04 15:04:10 INFO Job: Running job: job_1672817468245_0298
23/01/04 15:04:16 INFO Job: Job job_1672817468245_0298 running in uber mode : false
23/01/04 15:04:16 INFO Job:  map 0% reduce 0%
23/01/04 15:04:22 INFO Job:  map 100% reduce 0%
23/01/04 15:04:27 INFO Job:  map 100% reduce 100%
23/01/04 15:04:27 INFO Job: Job job_1672817468245_0298 completed successfully
23/01/04 15:04:27 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=1852735
        FILE: Number of bytes written=4237761
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4643875
        HDFS: Number of bytes written=1309991
        HDFS: Number of read operations=31
        HDFS: Number of large read operations=0

```

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17745  
Total time spent by all reduces in occupied slots (ms)=17205  
Total time spent by all map tasks (ms)=3549  
Total time spent by all reduce tasks (ms)=3441  
Total vcore-milliseconds taken by all map tasks=3549  
Total vcore-milliseconds taken by all reduce tasks=3441  
Total megabyte-milliseconds taken by all map tasks=18170880  
Total megabyte-milliseconds taken by all reduce tasks=17617920

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=1852709  
Map output materialized bytes=1852735  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=1852735  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=4180  
Physical memory (bytes) snapshot=730382336  
Virtual memory (bytes) snapshot=12750389248  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402108416  
Peak Map Virtual memory (bytes)=6370631680  
Peak Reduce Physical memory (bytes)=328273920  
Peak Reduce Virtual memory (bytes)=6379757568

#### 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=2680027

#### File Output Format Counters

Bytes Written=1309991

23/01/04 15:04:27 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:04:27 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0299

23/01/04 15:04:27 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:04:27 INFO JobSubmitter: number of splits:1

23/01/04 15:04:28 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0299

23/01/04 15:04:28 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:04:28 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0299  
23/01/04 15:04:28 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0299/  
23/01/04 15:04:28 INFO Job: Running job: job\_1672817468245\_0299  
23/01/04 15:04:34 INFO Job: Job job\_1672817468245\_0299 running in uber mode : false  
23/01/04 15:04:34 INFO Job: map 0% reduce 0%  
23/01/04 15:04:40 INFO Job: map 100% reduce 0%  
23/01/04 15:04:46 INFO Job: map 100% reduce 100%  
23/01/04 15:04:46 INFO Job: Job job\_1672817468245\_0299 completed successfully  
23/01/04 15:04:46 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2816930  
FILE: Number of bytes written=6166151  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5300567  
HDFS: Number of bytes written=1287373  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17880  
Total time spent by all reduces in occupied slots (ms)=17510  
Total time spent by all map tasks (ms)=3576  
Total time spent by all reduce tasks (ms)=3502  
Total vcore-milliseconds taken by all map tasks=3576  
Total vcore-milliseconds taken by all reduce tasks=3502  
Total megabyte-milliseconds taken by all map tasks=18309120  
Total megabyte-milliseconds taken by all reduce tasks=17930240

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2816904  
Map output materialized bytes=2816930  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2816930  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=3810  
Physical memory (bytes) snapshot=742141952  
Virtual memory (bytes) snapshot=12764381184  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412250112  
Peak Map Virtual memory (bytes)=6381887488  
Peak Reduce Physical memory (bytes)=329891840

```

        Peak Reduce Virtual memory (bytes)=6382493696
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=2680027
File Output Format Counters
    Bytes Written=1287373
23/01/04 15:04:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:04:46 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0300
23/01/04 15:04:46 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:04:46 INFO JobSubmitter: number of splits:1
23/01/04 15:04:46 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0300
23/01/04 15:04:46 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:04:46 INFO YarnClientImpl: Submitted application
application_1672817468245_0300
23/01/04 15:04:46 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0300/
23/01/04 15:04:46 INFO Job: Running job: job_1672817468245_0300
23/01/04 15:04:52 INFO Job: Job job_1672817468245_0300 running in uber mode : false
23/01/04 15:04:52 INFO Job:  map 0% reduce 0%
23/01/04 15:04:58 INFO Job:  map 100% reduce 0%
23/01/04 15:05:03 INFO Job:  map 100% reduce 100%
23/01/04 15:05:04 INFO Job: Job job_1672817468245_0300 completed successfully
23/01/04 15:05:04 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=2734174
    FILE: Number of bytes written=6000639
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=5255331
    HDFS: Number of bytes written=1264287
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17630
    Total time spent by all reduces in occupied slots (ms)=17870
    Total time spent by all map tasks (ms)=3526
    Total time spent by all reduce tasks (ms)=3574
    Total vcore-milliseconds taken by all map tasks=3526
    Total vcore-milliseconds taken by all reduce tasks=3574
    Total megabyte-milliseconds taken by all map tasks=18053120
    Total megabyte-milliseconds taken by all reduce tasks=18298880
Map-Reduce Framework
    Map input records=27
    Map output records=4

```



Map output bytes=2734148  
Map output materialized bytes=2734174  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2734174  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=61  
CPU time spent (ms)=3760  
Physical memory (bytes) snapshot=721018880  
Virtual memory (bytes) snapshot=12751912960  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400703488  
Peak Map Virtual memory (bytes)=6371491840  
Peak Reduce Physical memory (bytes)=320315392  
Peak Reduce Virtual memory (bytes)=6380421120

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=2680027

File Output Format Counters

Bytes Written=1264287

23/01/04 15:05:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:05:04 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0301  
23/01/04 15:05:05 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:05:05 INFO JobSubmitter: number of splits:1  
23/01/04 15:05:05 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0301  
23/01/04 15:05:05 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:05:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0301  
23/01/04 15:05:05 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0301/](http://lena-master:8088/proxy/application_1672817468245_0301/)  
23/01/04 15:05:05 INFO Job: Running job: job\_1672817468245\_0301  
23/01/04 15:05:11 INFO Job: Job job\_1672817468245\_0301 running in uber mode : false  
23/01/04 15:05:11 INFO Job: map 0% reduce 0%  
23/01/04 15:05:17 INFO Job: map 100% reduce 0%  
23/01/04 15:05:23 INFO Job: map 100% reduce 100%  
23/01/04 15:05:23 INFO Job: Job job\_1672817468245\_0301 completed successfully  
23/01/04 15:05:23 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954467  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0

HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17735  
Total time spent by all reduces in occupied slots (ms)=17200  
Total time spent by all map tasks (ms)=3547  
Total time spent by all reduce tasks (ms)=3440  
Total vcore-milliseconds taken by all map tasks=3547  
Total vcore-milliseconds taken by all reduce tasks=3440  
Total megabyte-milliseconds taken by all map tasks=18160640  
Total megabyte-milliseconds taken by all reduce tasks=17612800

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3900  
Physical memory (bytes) snapshot=725336064  
Virtual memory (bytes) snapshot=12752822272  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404107264  
Peak Map Virtual memory (bytes)=6367313920  
Peak Reduce Physical memory (bytes)=321228800  
Peak Reduce Virtual memory (bytes)=6385508352

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:05:23 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:05:23 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0302

23/01/04 15:05:23 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:05:23 INFO JobSubmitter: number of splits:1  
23/01/04 15:05:23 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0302  
23/01/04 15:05:23 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:05:23 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0302  
23/01/04 15:05:23 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0302/](http://lena-master:8088/proxy/application_1672817468245_0302/)  
23/01/04 15:05:23 INFO Job: Running job: job\_1672817468245\_0302  
23/01/04 15:05:29 INFO Job: Job job\_1672817468245\_0302 running in uber mode : false  
23/01/04 15:05:29 INFO Job: map 0% reduce 0%  
23/01/04 15:05:35 INFO Job: map 100% reduce 0%  
23/01/04 15:05:42 INFO Job: map 100% reduce 100%  
23/01/04 15:05:42 INFO Job: Job job\_1672817468245\_0302 completed successfully  
23/01/04 15:05:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954467  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17665  
Total time spent by all reduces in occupied slots (ms)=17155  
Total time spent by all map tasks (ms)=3533  
Total time spent by all reduce tasks (ms)=3431  
Total vcore-milliseconds taken by all map tasks=3533  
Total vcore-milliseconds taken by all reduce tasks=3431  
Total megabyte-milliseconds taken by all map tasks=18088960  
Total megabyte-milliseconds taken by all reduce tasks=17566720

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=3950  
Physical memory (bytes) snapshot=745570304  
Virtual memory (bytes) snapshot=12762980352

```

    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=418983936
    Peak Map Virtual memory (bytes)=6375047168
    Peak Reduce Physical memory (bytes)=326586368
    Peak Reduce Virtual memory (bytes)=6387933184
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=2680027
File Output Format Counters
    Bytes Written=1264287
23/01/04 15:05:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:05:42 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0303
23/01/04 15:05:42 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:05:42 INFO JobSubmitter: number of splits:1
23/01/04 15:05:42 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0303
23/01/04 15:05:42 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:05:42 INFO YarnClientImpl: Submitted application
application_1672817468245_0303
23/01/04 15:05:42 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0303/
23/01/04 15:05:42 INFO Job: Running job: job_1672817468245_0303
23/01/04 15:05:48 INFO Job: Job job_1672817468245_0303 running in uber mode : false
23/01/04 15:05:48 INFO Job:  map 0% reduce 0%
23/01/04 15:05:54 INFO Job:  map 100% reduce 0%
23/01/04 15:05:59 INFO Job:  map 100% reduce 100%
23/01/04 15:05:59 INFO Job: Job job_1672817468245_0303 completed successfully
23/01/04 15:05:59 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=2711088
    FILE: Number of bytes written=5954467
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=5209159
    HDFS: Number of bytes written=1264287
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17840
    Total time spent by all reduces in occupied slots (ms)=17265
    Total time spent by all map tasks (ms)=3568
    Total time spent by all reduce tasks (ms)=3453
    Total vcore-milliseconds taken by all map tasks=3568
    Total vcore-milliseconds taken by all reduce tasks=3453
    Total megabyte-milliseconds taken by all map tasks=18268160

```

```

        Total megabyte-milliseconds taken by all reduce tasks=17679360
Map-Reduce Framework
    Map input records=27
    Map output records=4
    Map output bytes=2711062
    Map output materialized bytes=2711088
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=4
    Reduce shuffle bytes=2711088
    Reduce input records=4
    Reduce output records=4
    Spilled Records=8
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=50
    CPU time spent (ms)=3790
    Physical memory (bytes) snapshot=722989056
    Virtual memory (bytes) snapshot=12762038272
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=406196224
    Peak Map Virtual memory (bytes)=6381109248
    Peak Reduce Physical memory (bytes)=316792832
    Peak Reduce Virtual memory (bytes)=6380929024
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=2680027
File Output Format Counters
    Bytes Written=1264287
23/01/04 15:05:59 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:05:59 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0304
23/01/04 15:06:00 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:06:00 INFO JobSubmitter: number of splits:1
23/01/04 15:06:00 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0304
23/01/04 15:06:00 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:06:00 INFO YarnClientImpl: Submitted application
application_1672817468245_0304
23/01/04 15:06:00 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0304/
23/01/04 15:06:00 INFO Job: Running job: job_1672817468245_0304
23/01/04 15:06:06 INFO Job: Job job_1672817468245_0304 running in uber mode : false
23/01/04 15:06:06 INFO Job:  map 0% reduce 0%
23/01/04 15:06:12 INFO Job:  map 100% reduce 0%
23/01/04 15:06:18 INFO Job:  map 100% reduce 100%
23/01/04 15:06:18 INFO Job: Job job_1672817468245_0304 completed successfully
23/01/04 15:06:18 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=2711088

```

FILE: Number of bytes written=5954467  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17655  
Total time spent by all reduces in occupied slots (ms)=17415  
Total time spent by all map tasks (ms)=3531  
Total time spent by all reduce tasks (ms)=3483  
Total vcore-milliseconds taken by all map tasks=3531  
Total vcore-milliseconds taken by all reduce tasks=3483  
Total megabyte-milliseconds taken by all map tasks=18078720  
Total megabyte-milliseconds taken by all reduce tasks=17832960

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=4100  
Physical memory (bytes) snapshot=756928512  
Virtual memory (bytes) snapshot=12759027712  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=414572544  
Peak Map Virtual memory (bytes)=6376136704  
Peak Reduce Physical memory (bytes)=342355968  
Peak Reduce Virtual memory (bytes)=6382891008

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:06:18 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:06:18 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0305  
23/01/04 15:06:18 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:06:18 INFO JobSubmitter: number of splits:1  
23/01/04 15:06:18 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0305  
23/01/04 15:06:18 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:06:18 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0305  
23/01/04 15:06:18 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0305/  
23/01/04 15:06:18 INFO Job: Running job: job\_1672817468245\_0305  
23/01/04 15:06:24 INFO Job: Job job\_1672817468245\_0305 running in uber mode : false  
23/01/04 15:06:24 INFO Job: map 0% reduce 0%  
23/01/04 15:06:30 INFO Job: map 100% reduce 0%  
23/01/04 15:06:36 INFO Job: map 100% reduce 100%  
23/01/04 15:06:36 INFO Job: Job job\_1672817468245\_0305 completed successfully  
23/01/04 15:06:36 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954467  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17820  
Total time spent by all reduces in occupied slots (ms)=17035  
Total time spent by all map tasks (ms)=3564  
Total time spent by all reduce tasks (ms)=3407  
Total vcore-milliseconds taken by all map tasks=3564  
Total vcore-milliseconds taken by all reduce tasks=3407  
Total megabyte-milliseconds taken by all map tasks=18247680  
Total megabyte-milliseconds taken by all reduce tasks=17443840

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=39
CPU time spent (ms)=4140
Physical memory (bytes) snapshot=748269568
Virtual memory (bytes) snapshot=12754935808
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=412528640
Peak Map Virtual memory (bytes)=6376558592
Peak Reduce Physical memory (bytes)=335740928
Peak Reduce Virtual memory (bytes)=6378377216
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=2680027
File Output Format Counters
  Bytes Written=1264287
23/01/04 15:06:36 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:06:36 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0306
23/01/04 15:06:36 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:06:36 INFO JobSubmitter: number of splits:1
23/01/04 15:06:36 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0306
23/01/04 15:06:36 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:06:36 INFO YarnClientImpl: Submitted application
application_1672817468245_0306
23/01/04 15:06:36 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0306/
23/01/04 15:06:36 INFO Job: Running job: job_1672817468245_0306
23/01/04 15:06:42 INFO Job: Job job_1672817468245_0306 running in uber mode : false
23/01/04 15:06:42 INFO Job:  map 0% reduce 0%
23/01/04 15:06:47 INFO Job:  map 100% reduce 0%
23/01/04 15:06:53 INFO Job:  map 100% reduce 100%
23/01/04 15:06:53 INFO Job: Job job_1672817468245_0306 completed successfully
23/01/04 15:06:53 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=2711088
  FILE: Number of bytes written=5954467
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5209159
  HDFS: Number of bytes written=1264287
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17445
  Total time spent by all reduces in occupied slots (ms)=17305
  Total time spent by all map tasks (ms)=3489

```



Total time spent by all reduce tasks (ms)=3461  
Total vcore-milliseconds taken by all map tasks=3489  
Total vcore-milliseconds taken by all reduce tasks=3461  
Total megabyte-milliseconds taken by all map tasks=17863680  
Total megabyte-milliseconds taken by all reduce tasks=17720320

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=4060  
Physical memory (bytes) snapshot=723226624  
Virtual memory (bytes) snapshot=12768342016  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402632704  
Peak Map Virtual memory (bytes)=6384361472  
Peak Reduce Physical memory (bytes)=320593920  
Peak Reduce Virtual memory (bytes)=6383980544

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:06:53 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:06:53 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0307  
23/01/04 15:06:54 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:06:54 INFO JobSubmitter: number of splits:1  
23/01/04 15:06:54 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0307  
23/01/04 15:06:54 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:06:54 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0307  
23/01/04 15:06:54 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0307/  
23/01/04 15:06:54 INFO Job: Running job: job\_1672817468245\_0307  
23/01/04 15:07:00 INFO Job: Job job\_1672817468245\_0307 running in uber mode : false  
23/01/04 15:07:00 INFO Job: map 0% reduce 0%  
23/01/04 15:07:06 INFO Job: map 100% reduce 0%  
23/01/04 15:07:12 INFO Job: map 100% reduce 100%

23/01/04 15:07:12 INFO Job: Job job\_1672817468245\_0307 completed successfully

23/01/04 15:07:12 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954471  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17510  
Total time spent by all reduces in occupied slots (ms)=17300  
Total time spent by all map tasks (ms)=3502  
Total time spent by all reduce tasks (ms)=3460  
Total vcore-milliseconds taken by all map tasks=3502  
Total vcore-milliseconds taken by all reduce tasks=3460  
Total megabyte-milliseconds taken by all map tasks=17930240  
Total megabyte-milliseconds taken by all reduce tasks=17715200

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=37  
CPU time spent (ms)=4020  
Physical memory (bytes) snapshot=733822976  
Virtual memory (bytes) snapshot=12753346560  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415576064  
Peak Map Virtual memory (bytes)=6369173504  
Peak Reduce Physical memory (bytes)=318246912  
Peak Reduce Virtual memory (bytes)=6384173056

#### 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=2680027  
File Output Format Counters  
Bytes Written=1264287  
23/01/04 15:07:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:07:12 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0308  
23/01/04 15:07:12 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:07:12 INFO JobSubmitter: number of splits:1  
23/01/04 15:07:12 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0308  
23/01/04 15:07:12 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:07:12 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0308  
23/01/04 15:07:12 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0308/  
23/01/04 15:07:12 INFO Job: Running job: job\_1672817468245\_0308  
23/01/04 15:07:18 INFO Job: Job job\_1672817468245\_0308 running in uber mode : false  
23/01/04 15:07:18 INFO Job: map 0% reduce 0%  
23/01/04 15:07:24 INFO Job: map 100% reduce 0%  
23/01/04 15:07:28 INFO Job: map 100% reduce 100%  
23/01/04 15:07:28 INFO Job: Job job\_1672817468245\_0308 completed successfully  
23/01/04 15:07:28 INFO Job: Counters: 54  
File System Counters  
FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0  
Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17610  
Total time spent by all reduces in occupied slots (ms)=9975  
Total time spent by all map tasks (ms)=3522  
Total time spent by all reduce tasks (ms)=1995  
Total vcore-milliseconds taken by all map tasks=3522  
Total vcore-milliseconds taken by all reduce tasks=1995  
Total megabyte-milliseconds taken by all map tasks=18032640  
Total megabyte-milliseconds taken by all reduce tasks=10214400  
Map-Reduce Framework  
Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4

Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=3910  
Physical memory (bytes) snapshot=744026112  
Virtual memory (bytes) snapshot=12758102016  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409133056  
Peak Map Virtual memory (bytes)=6370463744  
Peak Reduce Physical memory (bytes)=334893056  
Peak Reduce Virtual memory (bytes)=6387638272

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=2680027

File Output Format Counters

Bytes Written=1264287

23/01/04 15:07:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:07:28 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0309  
23/01/04 15:07:29 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:07:29 INFO JobSubmitter: number of splits:1  
23/01/04 15:07:29 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0309  
23/01/04 15:07:29 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:07:29 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0309  
23/01/04 15:07:29 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0309/](http://lena-master:8088/proxy/application_1672817468245_0309/)  
23/01/04 15:07:29 INFO Job: Running job: job\_1672817468245\_0309  
23/01/04 15:07:35 INFO Job: Job job\_1672817468245\_0309 running in uber mode : false  
23/01/04 15:07:35 INFO Job: map 0% reduce 0%  
23/01/04 15:07:41 INFO Job: map 100% reduce 0%  
23/01/04 15:07:46 INFO Job: map 100% reduce 100%  
23/01/04 15:07:46 INFO Job: Job job\_1672817468245\_0309 completed successfully  
23/01/04 15:07:46 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1

Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17525  
Total time spent by all reduces in occupied slots (ms)=10260  
Total time spent by all map tasks (ms)=3505  
Total time spent by all reduce tasks (ms)=2052  
Total vcore-milliseconds taken by all map tasks=3505  
Total vcore-milliseconds taken by all reduce tasks=2052  
Total megabyte-milliseconds taken by all map tasks=17945600  
Total megabyte-milliseconds taken by all reduce tasks=10506240

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=3970  
Physical memory (bytes) snapshot=764436480  
Virtual memory (bytes) snapshot=12768264192  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=427352064  
Peak Map Virtual memory (bytes)=6380810240  
Peak Reduce Physical memory (bytes)=337084416  
Peak Reduce Virtual memory (bytes)=6387453952

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:07:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:07:46 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0310  
23/01/04 15:07:46 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:07:46 INFO JobSubmitter: number of splits:1  
23/01/04 15:07:46 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0310  
23/01/04 15:07:46 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:07:46 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0310  
23/01/04 15:07:46 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0310/](http://lena-master:8088/proxy/application_1672817468245_0310/)  
23/01/04 15:07:46 INFO Job: Running job: job\_1672817468245\_0310

23/01/04 15:07:52 INFO Job: Job job\_1672817468245\_0310 running in uber mode : false  
23/01/04 15:07:52 INFO Job: map 0% reduce 0%  
23/01/04 15:07:58 INFO Job: map 100% reduce 0%  
23/01/04 15:08:03 INFO Job: map 100% reduce 100%  
23/01/04 15:08:03 INFO Job: Job job\_1672817468245\_0310 completed successfully  
23/01/04 15:08:03 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17930  
Total time spent by all reduces in occupied slots (ms)=17185  
Total time spent by all map tasks (ms)=3586  
Total time spent by all reduce tasks (ms)=3437  
Total vcore-milliseconds taken by all map tasks=3586  
Total vcore-milliseconds taken by all reduce tasks=3437  
Total megabyte-milliseconds taken by all map tasks=18360320  
Total megabyte-milliseconds taken by all reduce tasks=17597440

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=3820  
Physical memory (bytes) snapshot=754561024  
Virtual memory (bytes) snapshot=12763926528  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406257664  
Peak Map Virtual memory (bytes)=6378475520  
Peak Reduce Physical memory (bytes)=348303360  
Peak Reduce Virtual memory (bytes)=6385451008

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1264287  
23/01/04 15:08:04 INFO DefaultNoHARMAiloverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:08:04 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0311  
23/01/04 15:08:04 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:08:04 INFO JobSubmitter: number of splits:1  
23/01/04 15:08:04 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0311  
23/01/04 15:08:04 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:08:04 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0311  
23/01/04 15:08:04 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0311/  
23/01/04 15:08:04 INFO Job: Running job: job\_1672817468245\_0311  
23/01/04 15:08:10 INFO Job: Job job\_1672817468245\_0311 running in uber mode : false  
23/01/04 15:08:10 INFO Job:  map 0% reduce 0%  
23/01/04 15:08:17 INFO Job:  map 100% reduce 0%  
23/01/04 15:08:24 INFO Job:  map 100% reduce 100%  
23/01/04 15:08:24 INFO Job: Job job\_1672817468245\_0311 completed successfully  
23/01/04 15:08:24 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=2711088  
  FILE: Number of bytes written=5954475  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=5209159  
  HDFS: Number of bytes written=1264287  
  HDFS: Number of read operations=19  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=18360  
  Total time spent by all reduces in occupied slots (ms)=17280  
  Total time spent by all map tasks (ms)=3672  
  Total time spent by all reduce tasks (ms)=3456  
  Total vcore-milliseconds taken by all map tasks=3672  
  Total vcore-milliseconds taken by all reduce tasks=3456  
  Total megabyte-milliseconds taken by all map tasks=18800640  
  Total megabyte-milliseconds taken by all reduce tasks=17694720  
Map-Reduce Framework  
  Map input records=27  
  Map output records=4  
  Map output bytes=2711062  
  Map output materialized bytes=2711088  
  Input split bytes=170  
  Combine input records=0  
  Combine output records=0

Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=54  
CPU time spent (ms)=4030  
Physical memory (bytes) snapshot=748138496  
Virtual memory (bytes) snapshot=12754010112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=420106240  
Peak Map Virtual memory (bytes)=6373203968  
Peak Reduce Physical memory (bytes)=328032256  
Peak Reduce Virtual memory (bytes)=6380806144

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=2680027

File Output Format Counters

Bytes Written=1264287

23/01/04 15:08:24 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:08:24 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0312  
23/01/04 15:08:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:08:24 INFO JobSubmitter: number of splits:1  
23/01/04 15:08:24 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0312  
23/01/04 15:08:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:08:24 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0312  
23/01/04 15:08:24 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0312/  
23/01/04 15:08:24 INFO Job: Running job: job\_1672817468245\_0312  
23/01/04 15:08:30 INFO Job: Job job\_1672817468245\_0312 running in uber mode : false  
23/01/04 15:08:30 INFO Job: map 0% reduce 0%  
23/01/04 15:08:36 INFO Job: map 100% reduce 0%  
23/01/04 15:08:41 INFO Job: map 100% reduce 100%  
23/01/04 15:08:42 INFO Job: Job job\_1672817468245\_0312 completed successfully  
23/01/04 15:08:42 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2



HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17815  
  Total time spent by all reduces in occupied slots (ms)=17305  
  Total time spent by all map tasks (ms)=3563  
  Total time spent by all reduce tasks (ms)=3461  
  Total vcore-milliseconds taken by all map tasks=3563  
  Total vcore-milliseconds taken by all reduce tasks=3461  
  Total megabyte-milliseconds taken by all map tasks=18242560  
  Total megabyte-milliseconds taken by all reduce tasks=17720320

Map-Reduce Framework  
  Map input records=27  
  Map output records=4  
  Map output bytes=2711062  
  Map output materialized bytes=2711088  
  Input split bytes=170  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=4  
  Reduce shuffle bytes=2711088  
  Reduce input records=4  
  Reduce output records=4  
  Spilled Records=8  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=44  
  CPU time spent (ms)=3810  
  Physical memory (bytes) snapshot=758407168  
  Virtual memory (bytes) snapshot=12746792960  
  Total committed heap usage (bytes)=2105540608  
  Peak Map Physical memory (bytes)=407416832  
  Peak Map Virtual memory (bytes)=6369005568  
  Peak Reduce Physical memory (bytes)=350990336  
  Peak Reduce Virtual memory (bytes)=6377787392

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=2680027

File Output Format Counters

  Bytes Written=1264287

23/01/04 15:08:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:08:43 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0313  
23/01/04 15:08:43 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:08:43 INFO JobSubmitter: number of splits:1  
23/01/04 15:08:43 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0313  
23/01/04 15:08:43 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:08:43 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0313

23/01/04 15:08:43 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0313/](http://lena-master:8088/proxy/application_1672817468245_0313/)

23/01/04 15:08:43 INFO Job: Running job: job\_1672817468245\_0313

23/01/04 15:08:49 INFO Job: Job job\_1672817468245\_0313 running in uber mode : false

23/01/04 15:08:49 INFO Job: map 0% reduce 0%

23/01/04 15:08:55 INFO Job: map 100% reduce 0%

23/01/04 15:09:01 INFO Job: map 100% reduce 100%

23/01/04 15:09:01 INFO Job: Job job\_1672817468245\_0313 completed successfully

23/01/04 15:09:01 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088

FILE: Number of bytes written=5954475

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=5209159

HDFS: Number of bytes written=1264287

HDFS: Number of read operations=19

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=18055

Total time spent by all reduces in occupied slots (ms)=17075

Total time spent by all map tasks (ms)=3611

Total time spent by all reduce tasks (ms)=3415

Total vcore-milliseconds taken by all map tasks=3611

Total vcore-milliseconds taken by all reduce tasks=3415

Total megabyte-milliseconds taken by all map tasks=18488320

Total megabyte-milliseconds taken by all reduce tasks=17484800

#### Map-Reduce Framework

Map input records=27

Map output records=4

Map output bytes=2711062

Map output materialized bytes=2711088

Input split bytes=170

Combine input records=0

Combine output records=0

Reduce input groups=4

Reduce shuffle bytes=2711088

Reduce input records=4

Reduce output records=4

Spilled Records=8

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=65

CPU time spent (ms)=3890

Physical memory (bytes) snapshot=742428672

Virtual memory (bytes) snapshot=12765491200

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=417394688

Peak Map Virtual memory (bytes)=6381658112

Peak Reduce Physical memory (bytes)=325033984

Peak Reduce Virtual memory (bytes)=6383833088

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:09:01 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:09:01 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0315  
23/01/04 15:09:01 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:09:01 INFO JobSubmitter: number of splits:1  
23/01/04 15:09:01 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0315  
23/01/04 15:09:01 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:09:01 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0315  
23/01/04 15:09:01 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0315/](http://lena-master:8088/proxy/application_1672817468245_0315/)  
23/01/04 15:09:01 INFO Job: Running job: job\_1672817468245\_0315  
23/01/04 15:09:07 INFO Job: Job job\_1672817468245\_0315 running in uber mode : false  
23/01/04 15:09:07 INFO Job: map 0% reduce 0%  
23/01/04 15:09:13 INFO Job: map 100% reduce 0%  
23/01/04 15:09:19 INFO Job: map 100% reduce 100%  
23/01/04 15:09:19 INFO Job: Job job\_1672817468245\_0315 completed successfully  
23/01/04 15:09:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17390  
Total time spent by all reduces in occupied slots (ms)=18030  
Total time spent by all map tasks (ms)=3478  
Total time spent by all reduce tasks (ms)=3606  
Total vcore-milliseconds taken by all map tasks=3478  
Total vcore-milliseconds taken by all reduce tasks=3606  
Total megabyte-milliseconds taken by all map tasks=17807360  
Total megabyte-milliseconds taken by all reduce tasks=18462720

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062

Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=94  
CPU time spent (ms)=3930  
Physical memory (bytes) snapshot=731111424  
Virtual memory (bytes) snapshot=12755693568  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411205632  
Peak Map Virtual memory (bytes)=6372323328  
Peak Reduce Physical memory (bytes)=319905792  
Peak Reduce Virtual memory (bytes)=6383370240

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=2680027

File Output Format Counters

Bytes Written=1264287

23/01/04 15:09:20 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:09:20 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0317  
23/01/04 15:09:20 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:09:20 INFO JobSubmitter: number of splits:1  
23/01/04 15:09:20 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0317  
23/01/04 15:09:20 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:09:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0317  
23/01/04 15:09:20 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0317/  
23/01/04 15:09:20 INFO Job: Running job: job\_1672817468245\_0317  
23/01/04 15:09:26 INFO Job: Job job\_1672817468245\_0317 running in uber mode : false  
23/01/04 15:09:26 INFO Job: map 0% reduce 0%  
23/01/04 15:09:32 INFO Job: map 100% reduce 0%  
23/01/04 15:09:38 INFO Job: map 100% reduce 100%  
23/01/04 15:09:38 INFO Job: Job job\_1672817468245\_0317 completed successfully  
23/01/04 15:09:38 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159

HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17695  
Total time spent by all reduces in occupied slots (ms)=17155  
Total time spent by all map tasks (ms)=3539  
Total time spent by all reduce tasks (ms)=3431  
Total vcore-milliseconds taken by all map tasks=3539  
Total vcore-milliseconds taken by all reduce tasks=3431  
Total megabyte-milliseconds taken by all map tasks=18119680  
Total megabyte-milliseconds taken by all reduce tasks=17566720

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=4010  
Physical memory (bytes) snapshot=736747520  
Virtual memory (bytes) snapshot=12766244864  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401461248  
Peak Map Virtual memory (bytes)=6375264256  
Peak Reduce Physical memory (bytes)=335286272  
Peak Reduce Virtual memory (bytes)=6390980608

#### 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=2680027

#### File Output Format Counters

Bytes Written=1264287

23/01/04 15:09:38 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:09:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0319

23/01/04 15:09:38 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:09:38 INFO JobSubmitter: number of splits:1

23/01/04 15:09:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0319  
23/01/04 15:09:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:09:39 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0319  
23/01/04 15:09:39 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0319/  
23/01/04 15:09:39 INFO Job: Running job: job\_1672817468245\_0319  
23/01/04 15:09:45 INFO Job: Job job\_1672817468245\_0319 running in uber mode : false  
23/01/04 15:09:45 INFO Job: map 0% reduce 0%  
23/01/04 15:09:50 INFO Job: map 100% reduce 0%  
23/01/04 15:09:56 INFO Job: map 100% reduce 100%  
23/01/04 15:09:56 INFO Job: Job job\_1672817468245\_0319 completed successfully  
23/01/04 15:09:56 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17670  
Total time spent by all reduces in occupied slots (ms)=17225  
Total time spent by all map tasks (ms)=3534  
Total time spent by all reduce tasks (ms)=3445  
Total vcore-milliseconds taken by all map tasks=3534  
Total vcore-milliseconds taken by all reduce tasks=3445  
Total megabyte-milliseconds taken by all map tasks=18094080  
Total megabyte-milliseconds taken by all reduce tasks=17638400

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3890  
Physical memory (bytes) snapshot=738189312  
Virtual memory (bytes) snapshot=12767965184  
Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=410537984  
Peak Map Virtual memory (bytes)=6381297664  
Peak Reduce Physical memory (bytes)=327651328  
Peak Reduce Virtual memory (bytes)=6386667520

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=2680027

File Output Format Counters

Bytes Written=1264287

23/01/04 15:09:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:09:56 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0321  
23/01/04 15:09:56 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:09:56 INFO JobSubmitter: number of splits:1  
23/01/04 15:09:56 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0321  
23/01/04 15:09:56 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:09:56 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0321  
23/01/04 15:09:56 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0321/](http://lena-master:8088/proxy/application_1672817468245_0321/)  
23/01/04 15:09:56 INFO Job: Running job: job\_1672817468245\_0321  
23/01/04 15:10:02 INFO Job: Job job\_1672817468245\_0321 running in uber mode : false  
23/01/04 15:10:02 INFO Job: map 0% reduce 0%  
23/01/04 15:10:08 INFO Job: map 100% reduce 0%  
23/01/04 15:10:14 INFO Job: map 100% reduce 100%  
23/01/04 15:10:14 INFO Job: Job job\_1672817468245\_0321 completed successfully  
23/01/04 15:10:14 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2711088  
FILE: Number of bytes written=5954475  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5209159  
HDFS: Number of bytes written=1264287  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18005  
Total time spent by all reduces in occupied slots (ms)=17550  
Total time spent by all map tasks (ms)=3601  
Total time spent by all reduce tasks (ms)=3510  
Total vcore-milliseconds taken by all map tasks=3601  
Total vcore-milliseconds taken by all reduce tasks=3510  
Total megabyte-milliseconds taken by all map tasks=18437120  
Total megabyte-milliseconds taken by all reduce tasks=17971200

## Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2711062  
Map output materialized bytes=2711088  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2711088  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=4200  
Physical memory (bytes) snapshot=755101696  
Virtual memory (bytes) snapshot=12750458880  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418172928  
Peak Map Virtual memory (bytes)=6374019072  
Peak Reduce Physical memory (bytes)=336928768  
Peak Reduce Virtual memory (bytes)=6376439808

## 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=2680027

## File Output Format Counters

Bytes Written=1264287

23/01/04 15:10:14 INFO KMeansDriver: Clustering data  
23/01/04 15:10:14 INFO KMeansDriver: Running Clustering  
23/01/04 15:10:14 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/04 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/04  
23/01/04 15:10:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:10:14 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0323  
23/01/04 15:10:15 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:10:15 INFO JobSubmitter: number of splits:1  
23/01/04 15:10:15 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0323  
23/01/04 15:10:15 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:10:15 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0323  
23/01/04 15:10:15 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0323/  
23/01/04 15:10:15 INFO Job: Running job: job\_1672817468245\_0323  
23/01/04 15:10:21 INFO Job: Job job\_1672817468245\_0323 running in uber mode : false



23/01/04 15:10:21 INFO Job: map 0% reduce 0%  
23/01/04 15:10:27 INFO Job: map 100% reduce 0%  
23/01/04 15:10:27 INFO Job: Job job\_1672817468245\_0323 completed successfully  
23/01/04 15:10:27 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265898  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3944678  
HDFS: Number of bytes written=2680693  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17560  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3512  
Total vcore-milliseconds taken by all map tasks=3512  
Total megabyte-milliseconds taken by all map tasks=17981440

Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=19  
CPU time spent (ms)=2150  
Physical memory (bytes) snapshot=290738176  
Virtual memory (bytes) snapshot=6369898496  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=290738176  
Peak Map Virtual memory (bytes)=6369898496

File Input Format Counters

Bytes Read=2680027

File Output Format Counters

Bytes Written=2680693

23/01/04 15:10:27 INFO MahoutDriver: Program took 379213 ms (Minutes: 6.320216666666667)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:10:33 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --  
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/EuclideanDistanceMeasure/04/clusters-20-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
EuclideanDistanceMeasure/04.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu  
re/04/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 15:10:34 INFO HadoopUtil: Deleting tmp/representative

```

23/01/04 15:10:34 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/04/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
04/clusters-20-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:10:35 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:10:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:10:35 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0325
23/01/04 15:10:35 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:10:35 INFO JobSubmitter: number of splits:1
23/01/04 15:10:35 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0325
23/01/04 15:10:35 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:10:35 INFO Configuration: resource-types.xml not found
23/01/04 15:10:35 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:10:35 INFO YarnClientImpl: Submitted application
application_1672817468245_0325
23/01/04 15:10:35 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0325/
23/01/04 15:10:35 INFO Job: Running job: job_1672817468245_0325
23/01/04 15:10:42 INFO Job: Job job_1672817468245_0325 running in uber mode : false
23/01/04 15:10:42 INFO Job: map 0% reduce 0%
23/01/04 15:10:47 INFO Job: map 100% reduce 0%
23/01/04 15:10:53 INFO Job: map 100% reduce 100%
23/01/04 15:10:53 INFO Job: Job job_1672817468245_0325 completed successfully
23/01/04 15:10:53 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=500363
        FILE: Number of bytes written=1533319
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4128326
        HDFS: Number of bytes written=1224132
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=16760
        Total time spent by all reduces in occupied slots (ms)=16590
        Total time spent by all map tasks (ms)=3352
        Total time spent by all reduce tasks (ms)=3318
        Total vcore-milliseconds taken by all map tasks=3352
        Total vcore-milliseconds taken by all reduce tasks=3318
        Total megabyte-milliseconds taken by all map tasks=17162240
        Total megabyte-milliseconds taken by all reduce tasks=16988160
    Map-Reduce Framework
        Map input records=27
        Map output records=4

```

Map output bytes=500337  
Map output materialized bytes=500363  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=500363  
Reduce input records=4  
Reduce output records=8  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=2810  
Physical memory (bytes) snapshot=687869952  
Virtual memory (bytes) snapshot=12747735040  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395870208  
Peak Map Virtual memory (bytes)=6370971648  
Peak Reduce Physical memory (bytes)=291999744  
Peak Reduce Virtual memory (bytes)=6380679168

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=2680693

File Output Format Counters

Bytes Written=1224132

23/01/04 15:10:53 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 15:10:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:10:53 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0327

23/01/04 15:10:53 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:10:53 INFO JobSubmitter: number of splits:1

23/01/04 15:10:53 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0327

23/01/04 15:10:53 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:10:53 INFO YarnClientImpl: Submitted application application\_1672817468245\_0327

23/01/04 15:10:53 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0327/

23/01/04 15:10:53 INFO Job: Running job: job\_1672817468245\_0327

23/01/04 15:10:59 INFO Job: Job job\_1672817468245\_0327 running in uber mode : false

23/01/04 15:10:59 INFO Job: map 0% reduce 0%

23/01/04 15:11:05 INFO Job: map 100% reduce 0%

23/01/04 15:11:10 INFO Job: map 100% reduce 100%

23/01/04 15:11:11 INFO Job: Job job\_1672817468245\_0327 completed successfully

23/01/04 15:11:11 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=461227

FILE: Number of bytes written=1455047

FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5129160  
HDFS: Number of bytes written=1685413  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16800  
Total time spent by all reduces in occupied slots (ms)=16350  
Total time spent by all map tasks (ms)=3360  
Total time spent by all reduce tasks (ms)=3270  
Total vcore-milliseconds taken by all map tasks=3360  
Total vcore-milliseconds taken by all reduce tasks=3270  
Total megabyte-milliseconds taken by all map tasks=17203200  
Total megabyte-milliseconds taken by all reduce tasks=16742400

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=461201  
Map output materialized bytes=461227  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=461227  
Reduce input records=4  
Reduce output records=12  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=50  
CPU time spent (ms)=2700  
Physical memory (bytes) snapshot=681869312  
Virtual memory (bytes) snapshot=12759027712  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394313728  
Peak Map Virtual memory (bytes)=6379020288  
Peak Reduce Physical memory (bytes)=287555584  
Peak Reduce Virtual memory (bytes)=6383915008

#### 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=2680693

#### File Output Format Counters

Bytes Written=1685413

23/01/04 15:11:11 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 15:11:11 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:11:11 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0329  
23/01/04 15:11:11 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:11:11 INFO JobSubmitter: number of splits:1  
23/01/04 15:11:12 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0329  
23/01/04 15:11:12 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:11:12 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0329  
23/01/04 15:11:12 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0329/  
23/01/04 15:11:12 INFO Job: Running job: job\_1672817468245\_0329  
23/01/04 15:11:18 INFO Job: Job job\_1672817468245\_0329 running in uber mode : false  
23/01/04 15:11:18 INFO Job: map 0% reduce 0%  
23/01/04 15:11:23 INFO Job: map 100% reduce 0%  
23/01/04 15:11:29 INFO Job: map 100% reduce 100%  
23/01/04 15:11:29 INFO Job: Job job\_1672817468245\_0329 completed successfully  
23/01/04 15:11:29 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=435255  
FILE: Number of bytes written=1403103  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6051722  
HDFS: Number of bytes written=2120702  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17220  
Total time spent by all reduces in occupied slots (ms)=16210  
Total time spent by all map tasks (ms)=3444  
Total time spent by all reduce tasks (ms)=3242  
Total vcore-milliseconds taken by all map tasks=3444  
Total vcore-milliseconds taken by all reduce tasks=3242  
Total megabyte-milliseconds taken by all map tasks=17633280  
Total megabyte-milliseconds taken by all reduce tasks=16599040

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=435229  
Map output materialized bytes=435255  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=435255  
Reduce input records=4  
Reduce output records=16  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=48
CPU time spent (ms)=3060
Physical memory (bytes) snapshot=694718464
Virtual memory (bytes) snapshot=12757315584
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=399872000
Peak Map Virtual memory (bytes)=6382587904
Peak Reduce Physical memory (bytes)=294846464
Peak Reduce Virtual memory (bytes)=6376681472
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=2680693
File Output Format Counters
  Bytes Written=2120702
23/01/04 15:11:29 INFO RepresentativePointsDriver: Representative Points Iteration
3
23/01/04 15:11:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:11:29 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0331
23/01/04 15:11:29 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:11:29 INFO JobSubmitter: number of splits:1
23/01/04 15:11:29 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0331
23/01/04 15:11:29 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:11:29 INFO YarnClientImpl: Submitted application
application_1672817468245_0331
23/01/04 15:11:29 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0331/
23/01/04 15:11:29 INFO Job: Running job: job_1672817468245_0331
23/01/04 15:11:35 INFO Job: Job job_1672817468245_0331 running in uber mode : false
23/01/04 15:11:35 INFO Job:  map 0% reduce 0%
23/01/04 15:11:42 INFO Job:  map 100% reduce 0%
23/01/04 15:11:46 INFO Job:  map 100% reduce 100%
23/01/04 15:11:46 INFO Job: Job job_1672817468245_0331 completed successfully
23/01/04 15:11:46 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=459260
  FILE: Number of bytes written=1451113
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6922300
  HDFS: Number of bytes written=2580016
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17320

```

Total time spent by all reduces in occupied slots (ms)=9180  
Total time spent by all map tasks (ms)=3464  
Total time spent by all reduce tasks (ms)=1836  
Total vcore-milliseconds taken by all map tasks=3464  
Total vcore-milliseconds taken by all reduce tasks=1836  
Total megabyte-milliseconds taken by all map tasks=17735680  
Total megabyte-milliseconds taken by all reduce tasks=9400320

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=459234  
Map output materialized bytes=459260  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=459260  
Reduce input records=4  
Reduce output records=20  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3080  
Physical memory (bytes) snapshot=692596736  
Virtual memory (bytes) snapshot=12766064640  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400293888  
Peak Map Virtual memory (bytes)=6381068288  
Peak Reduce Physical memory (bytes)=292302848  
Peak Reduce Virtual memory (bytes)=6386950144

#### 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=2680693

#### File Output Format Counters

Bytes Written=2580016

23/01/04 15:11:46 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:11:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:11:46 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0333

23/01/04 15:11:46 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:11:46 INFO JobSubmitter: number of splits:1

23/01/04 15:11:47 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0333

23/01/04 15:11:47 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:11:47 INFO YarnClientImpl: Submitted application application\_1672817468245\_0333

23/01/04 15:11:47 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0333/](http://lena-master:8088/proxy/application_1672817468245_0333/)

23/01/04 15:11:47 INFO Job: Running job: job\_1672817468245\_0333

23/01/04 15:11:53 INFO Job: Job job\_1672817468245\_0333 running in uber mode : false  
23/01/04 15:11:53 INFO Job: map 0% reduce 0%  
23/01/04 15:11:58 INFO Job: map 100% reduce 0%  
23/01/04 15:12:04 INFO Job: map 100% reduce 100%  
23/01/04 15:12:05 INFO Job: Job job\_1672817468245\_0333 completed successfully  
23/01/04 15:12:05 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=494218  
FILE: Number of bytes written=1521029  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7840928  
HDFS: Number of bytes written=3074288  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16810  
Total time spent by all reduces in occupied slots (ms)=16810  
Total time spent by all map tasks (ms)=3362  
Total time spent by all reduce tasks (ms)=3362  
Total vcore-milliseconds taken by all map tasks=3362  
Total vcore-milliseconds taken by all reduce tasks=3362  
Total megabyte-milliseconds taken by all map tasks=17213440  
Total megabyte-milliseconds taken by all reduce tasks=17213440

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=494192  
Map output materialized bytes=494218  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=494218  
Reduce input records=4  
Reduce output records=24  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=3240  
Physical memory (bytes) snapshot=689106944  
Virtual memory (bytes) snapshot=12762882048  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397492224  
Peak Map Virtual memory (bytes)=6381211648  
Peak Reduce Physical memory (bytes)=291614720  
Peak Reduce Virtual memory (bytes)=6385586176

#### 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=2680693
    File Output Format Counters
        Bytes Written=3074288
23/01/04 15:12:05 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.5683087556196019
23/01/04 15:12:05 INFO ClusterEvaluator: Intra-Cluster Density[0] =
0.6171908322634607
23/01/04 15:12:05 INFO ClusterEvaluator: Intra-Cluster Density[16] = NaN
23/01/04 15:12:05 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN
23/01/04 15:12:05 INFO ClusterEvaluator: Intra-Cluster Density[4] = NaN
23/01/04 15:12:05 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6171908322634607
23/01/04 15:12:05 INFO ClusterDumper: Wrote 4 clusters
23/01/04 15:12:05 INFO MahoutDriver: Program took 92448 ms (Minutes: 1.5408)
---- K: 05 --
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/05
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:12:11 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[5],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/05], --overwrite=null, --startPhase=[0], --
tempDir=[tmp]}
23/01/04 15:12:11 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/05
23/01/04 15:12:11 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:12:12 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:12:12 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:12:12 INFO RandomSeedGenerator: Wrote 5 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:12:12 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/05
23/01/04 15:12:12 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:12:12 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:12:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:12:12 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0334
23/01/04 15:12:12 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:12:12 INFO JobSubmitter: number of splits:1
23/01/04 15:12:13 INFO JobSubmitter: Submitting tokens for job:

```

job\_1672817468245\_0334  
23/01/04 15:12:13 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:12:13 INFO Configuration: resource-types.xml not found  
23/01/04 15:12:13 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 15:12:13 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0334  
23/01/04 15:12:13 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0334/  
23/01/04 15:12:13 INFO Job: Running job: job\_1672817468245\_0334  
23/01/04 15:12:19 INFO Job: Job job\_1672817468245\_0334 running in uber mode : false  
23/01/04 15:12:19 INFO Job: map 0% reduce 0%  
23/01/04 15:12:25 INFO Job: map 100% reduce 0%  
23/01/04 15:12:30 INFO Job: map 100% reduce 100%  
23/01/04 15:12:31 INFO Job: Job job\_1672817468245\_0334 completed successfully  
23/01/04 15:12:31 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2443248  
FILE: Number of bytes written=5418787  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5892989  
HDFS: Number of bytes written=1582665  
HDFS: Number of read operations=35  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18320  
Total time spent by all reduces in occupied slots (ms)=17020  
Total time spent by all map tasks (ms)=3664  
Total time spent by all reduce tasks (ms)=3404  
Total vcore-milliseconds taken by all map tasks=3664  
Total vcore-milliseconds taken by all reduce tasks=3404  
Total megabyte-milliseconds taken by all map tasks=18759680  
Total megabyte-milliseconds taken by all reduce tasks=17428480

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=2443217  
Map output materialized bytes=2443248  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=2443248  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=4350  
Physical memory (bytes) snapshot=720318464  
Virtual memory (bytes) snapshot=12748103680

```

Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=400666624
Peak Map Virtual memory (bytes)=6370795520
Peak Reduce Physical memory (bytes)=319651840
Peak Reduce Virtual memory (bytes)=6377308160
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=2680027
File Output Format Counters
  Bytes Written=1582665
23/01/04 15:12:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:12:31 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0336
23/01/04 15:12:31 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:12:31 INFO JobSubmitter: number of splits:1
23/01/04 15:12:31 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0336
23/01/04 15:12:31 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:12:31 INFO YarnClientImpl: Submitted application
application_1672817468245_0336
23/01/04 15:12:31 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0336/
23/01/04 15:12:31 INFO Job: Running job: job_1672817468245_0336
23/01/04 15:12:37 INFO Job: Job job_1672817468245_0336 running in uber mode : false
23/01/04 15:12:37 INFO Job:  map 0% reduce 0%
23/01/04 15:12:44 INFO Job:  map 100% reduce 0%
23/01/04 15:12:48 INFO Job:  map 100% reduce 100%
23/01/04 15:12:48 INFO Job: Job job_1672817468245_0336 completed successfully
23/01/04 15:12:48 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=3465861
  FILE: Number of bytes written=7464013
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5845915
  HDFS: Number of bytes written=1558789
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17600
  Total time spent by all reduces in occupied slots (ms)=10285
  Total time spent by all map tasks (ms)=3520
  Total time spent by all reduce tasks (ms)=2057
  Total vcore-milliseconds taken by all map tasks=3520
  Total vcore-milliseconds taken by all reduce tasks=2057
  Total megabyte-milliseconds taken by all map tasks=18022400

```

```

        Total megabyte-milliseconds taken by all reduce tasks=10531840
Map-Reduce Framework
    Map input records=27
    Map output records=5
    Map output bytes=3465830
    Map output materialized bytes=3465861
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=5
    Reduce shuffle bytes=3465861
    Reduce input records=5
    Reduce output records=5
    Spilled Records=10
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=88
    CPU time spent (ms)=3940
    Physical memory (bytes) snapshot=742420480
    Virtual memory (bytes) snapshot=12755746816
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=410697728
    Peak Map Virtual memory (bytes)=6370611200
    Peak Reduce Physical memory (bytes)=331722752
    Peak Reduce Virtual memory (bytes)=6385135616
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=2680027
File Output Format Counters
    Bytes Written=1558789
23/01/04 15:12:48 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:12:48 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0338
23/01/04 15:12:48 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:12:48 INFO JobSubmitter: number of splits:1
23/01/04 15:12:48 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0338
23/01/04 15:12:48 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:12:48 INFO YarnClientImpl: Submitted application
application_1672817468245_0338
23/01/04 15:12:48 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0338/
23/01/04 15:12:48 INFO Job: Running job: job_1672817468245_0338
23/01/04 15:12:54 INFO Job: Job job_1672817468245_0338 running in uber mode : false
23/01/04 15:12:54 INFO Job:  map 0% reduce 0%
23/01/04 15:13:00 INFO Job:  map 100% reduce 0%
23/01/04 15:13:06 INFO Job:  map 100% reduce 100%
23/01/04 15:13:06 INFO Job: Job job_1672817468245_0338 completed successfully
23/01/04 15:13:06 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=3392355

```

FILE: Number of bytes written=7317001  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5798163  
HDFS: Number of bytes written=1508820  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18095  
Total time spent by all reduces in occupied slots (ms)=17105  
Total time spent by all map tasks (ms)=3619  
Total time spent by all reduce tasks (ms)=3421  
Total vcore-milliseconds taken by all map tasks=3619  
Total vcore-milliseconds taken by all reduce tasks=3421  
Total megabyte-milliseconds taken by all map tasks=18529280  
Total megabyte-milliseconds taken by all reduce tasks=17515520

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3392324  
Map output materialized bytes=3392355  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3392355  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3990  
Physical memory (bytes) snapshot=741089280  
Virtual memory (bytes) snapshot=12749074432  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409780224  
Peak Map Virtual memory (bytes)=6369943552  
Peak Reduce Physical memory (bytes)=331309056  
Peak Reduce Virtual memory (bytes)=6379130880

#### 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=2680027

#### File Output Format Counters

Bytes Written=1508820

23/01/04 15:13:06 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:13:06 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0340  
23/01/04 15:13:06 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:13:06 INFO JobSubmitter: number of splits:1  
23/01/04 15:13:06 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0340  
23/01/04 15:13:06 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:13:06 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0340  
23/01/04 15:13:06 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0340/  
23/01/04 15:13:06 INFO Job: Running job: job\_1672817468245\_0340  
23/01/04 15:13:13 INFO Job: Job job\_1672817468245\_0340 running in uber mode : false  
23/01/04 15:13:13 INFO Job: map 0% reduce 0%  
23/01/04 15:13:19 INFO Job: map 100% reduce 0%  
23/01/04 15:13:25 INFO Job: map 100% reduce 100%  
23/01/04 15:13:25 INFO Job: Job job\_1672817468245\_0340 completed successfully  
23/01/04 15:13:25 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3231684  
FILE: Number of bytes written=6995659  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5698225  
HDFS: Number of bytes written=1393593  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17755  
Total time spent by all reduces in occupied slots (ms)=17595  
Total time spent by all map tasks (ms)=3551  
Total time spent by all reduce tasks (ms)=3519  
Total vcore-milliseconds taken by all map tasks=3551  
Total vcore-milliseconds taken by all reduce tasks=3519  
Total megabyte-milliseconds taken by all map tasks=18181120  
Total megabyte-milliseconds taken by all reduce tasks=18017280

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3231653  
Map output materialized bytes=3231684  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3231684  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=57
CPU time spent (ms)=4210
Physical memory (bytes) snapshot=764846080
Virtual memory (bytes) snapshot=12752879616
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=425353216
Peak Map Virtual memory (bytes)=6374285312
Peak Reduce Physical memory (bytes)=339492864
Peak Reduce Virtual memory (bytes)=6378594304
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=2680027
File Output Format Counters
  Bytes Written=1393593
23/01/04 15:13:25 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:13:25 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0342
23/01/04 15:13:25 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:13:25 INFO JobSubmitter: number of splits:1
23/01/04 15:13:25 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0342
23/01/04 15:13:25 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:13:25 INFO YarnClientImpl: Submitted application
application_1672817468245_0342
23/01/04 15:13:25 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0342/
23/01/04 15:13:25 INFO Job: Running job: job_1672817468245_0342
23/01/04 15:13:31 INFO Job: Job job_1672817468245_0342 running in uber mode : false
23/01/04 15:13:31 INFO Job:  map 0% reduce 0%
23/01/04 15:13:36 INFO Job:  map 100% reduce 0%
23/01/04 15:13:42 INFO Job:  map 100% reduce 100%
23/01/04 15:13:43 INFO Job: Job job_1672817468245_0342 completed successfully
23/01/04 15:13:43 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=3055823
  FILE: Number of bytes written=6643937
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5467771
  HDFS: Number of bytes written=1370011
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17520
  Total time spent by all reduces in occupied slots (ms)=17190
  Total time spent by all map tasks (ms)=3504

```

Total time spent by all reduce tasks (ms)=3438  
Total vcore-milliseconds taken by all map tasks=3504  
Total vcore-milliseconds taken by all reduce tasks=3438  
Total megabyte-milliseconds taken by all map tasks=17940480  
Total megabyte-milliseconds taken by all reduce tasks=17602560

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3055792  
Map output materialized bytes=3055823  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3055823  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=3920  
Physical memory (bytes) snapshot=728162304  
Virtual memory (bytes) snapshot=12762877952  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406454272  
Peak Map Virtual memory (bytes)=6377377792  
Peak Reduce Physical memory (bytes)=321708032  
Peak Reduce Virtual memory (bytes)=6385500160

#### 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=2680027

#### File Output Format Counters

Bytes Written=1370011

23/01/04 15:13:43 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:13:43 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0344  
23/01/04 15:13:43 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:13:44 INFO JobSubmitter: number of splits:1  
23/01/04 15:13:44 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0344  
23/01/04 15:13:44 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:13:44 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0344  
23/01/04 15:13:44 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0344/  
23/01/04 15:13:44 INFO Job: Running job: job\_1672817468245\_0344  
23/01/04 15:13:50 INFO Job: Job job\_1672817468245\_0344 running in uber mode : false  
23/01/04 15:13:50 INFO Job: map 0% reduce 0%  
23/01/04 15:13:56 INFO Job: map 100% reduce 0%  
23/01/04 15:14:02 INFO Job: map 100% reduce 100%



23/01/04 15:14:02 INFO Job: Job job\_1672817468245\_0344 completed successfully

23/01/04 15:14:02 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3032241  
FILE: Number of bytes written=6596773  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5420607  
HDFS: Number of bytes written=1370011  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17825  
Total time spent by all reduces in occupied slots (ms)=17465  
Total time spent by all map tasks (ms)=3565  
Total time spent by all reduce tasks (ms)=3493  
Total vcore-milliseconds taken by all map tasks=3565  
Total vcore-milliseconds taken by all reduce tasks=3493  
Total megabyte-milliseconds taken by all map tasks=18252800  
Total megabyte-milliseconds taken by all reduce tasks=17884160

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3032210  
Map output materialized bytes=3032241  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3032241  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65  
CPU time spent (ms)=3700  
Physical memory (bytes) snapshot=736665600  
Virtual memory (bytes) snapshot=12756979712  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410873856  
Peak Map Virtual memory (bytes)=6373900288  
Peak Reduce Physical memory (bytes)=325791744  
Peak Reduce Virtual memory (bytes)=6383079424

#### 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=2680027
    File Output Format Counters
        Bytes Written=1370011
23/01/04 15:14:02 INFO KMeansDriver: Clustering data
23/01/04 15:14:02 INFO KMeansDriver: Running Clustering
23/01/04 15:14:02 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/05 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/05
23/01/04 15:14:02 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:14:02 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0345
23/01/04 15:14:02 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:14:02 INFO JobSubmitter: number of splits:1
23/01/04 15:14:02 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0345
23/01/04 15:14:02 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:14:02 INFO YarnClientImpl: Submitted application
application_1672817468245_0345
23/01/04 15:14:02 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0345/
23/01/04 15:14:02 INFO Job: Running job: job_1672817468245_0345
23/01/04 15:14:08 INFO Job: Job job_1672817468245_0345 running in uber mode : false
23/01/04 15:14:08 INFO Job:  map 0% reduce 0%
23/01/04 15:14:13 INFO Job:  map 100% reduce 0%
23/01/04 15:14:14 INFO Job: Job job_1672817468245_0345 completed successfully
23/01/04 15:14:14 INFO Job: Counters: 33
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=265897
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4050402
        HDFS: Number of bytes written=2680676
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17540
        Total time spent by all reduces in occupied slots (ms)=0
        Total time spent by all map tasks (ms)=3508
        Total vcore-milliseconds taken by all map tasks=3508
        Total megabyte-milliseconds taken by all map tasks=17960960
    Map-Reduce Framework
        Map input records=27
        Map output records=27
        Input split bytes=170
        Spilled Records=0
        Failed Shuffles=0
        Merged Map outputs=0
        GC time elapsed (ms)=33
        CPU time spent (ms)=2230

```

```

Physical memory (bytes) snapshot=297578496
Virtual memory (bytes) snapshot=6378905600
Total committed heap usage (bytes)=1052770304
Peak Map Physical memory (bytes)=297578496
Peak Map Virtual memory (bytes)=6378905600
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680676
23/01/04 15:14:14 INFO MahoutDriver: Program took 123575 ms (Minutes:
2.0595833333333333)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:14:20 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/EuclideanDistanceMeasure/05/clusters-6-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
EuclideanDistanceMeasure/05.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu
re/05/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:14:21 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:14:21 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/05/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
05/clusters-6-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:14:22 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:14:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:14:22 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0346
23/01/04 15:14:22 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:14:22 INFO JobSubmitter: number of splits:1
23/01/04 15:14:22 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0346
23/01/04 15:14:22 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:14:22 INFO Configuration: resource-types.xml not found
23/01/04 15:14:22 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:14:22 INFO YarnClientImpl: Submitted application
application_1672817468245_0346
23/01/04 15:14:22 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0346/
23/01/04 15:14:22 INFO Job: Running job: job_1672817468245_0346
23/01/04 15:14:29 INFO Job: Job job_1672817468245_0346 running in uber mode : false
23/01/04 15:14:29 INFO Job:  map 0% reduce 0%
23/01/04 15:14:35 INFO Job:  map 100% reduce 0%
23/01/04 15:14:41 INFO Job:  map 100% reduce 100%
23/01/04 15:14:41 INFO Job: Job job_1672817468245_0346 completed successfully
23/01/04 15:14:41 INFO Job: Counters: 54
File System Counters

```

FILE: Number of bytes read=610324  
FILE: Number of bytes written=1753241  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4343835  
HDFS: Number of bytes written=1441831  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16655  
Total time spent by all reduces in occupied slots (ms)=16720  
Total time spent by all map tasks (ms)=3331  
Total time spent by all reduce tasks (ms)=3344  
Total vcore-milliseconds taken by all map tasks=3331  
Total vcore-milliseconds taken by all reduce tasks=3344  
Total megabyte-milliseconds taken by all map tasks=17054720  
Total megabyte-milliseconds taken by all reduce tasks=17121280

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=610293  
Map output materialized bytes=610324  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=610324  
Reduce input records=5  
Reduce output records=10  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=2660  
Physical memory (bytes) snapshot=670375936  
Virtual memory (bytes) snapshot=12764094464  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=379039744  
Peak Map Virtual memory (bytes)=6376951808  
Peak Reduce Physical memory (bytes)=291336192  
Peak Reduce Virtual memory (bytes)=6388645888

#### 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=2680676

#### File Output Format Counters

Bytes Written=1441831

23/01/04 15:14:41 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 15:14:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:14:41 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0347  
23/01/04 15:14:41 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:14:41 INFO JobSubmitter: number of splits:1  
23/01/04 15:14:41 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0347  
23/01/04 15:14:41 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:14:41 INFO YarnClientImpl: Submitted application application\_1672817468245\_0347  
23/01/04 15:14:41 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0347/  
23/01/04 15:14:41 INFO Job: Running job: job\_1672817468245\_0347  
23/01/04 15:14:47 INFO Job: Job job\_1672817468245\_0347 running in uber mode : false  
23/01/04 15:14:47 INFO Job: map 0% reduce 0%  
23/01/04 15:14:52 INFO Job: map 100% reduce 0%  
23/01/04 15:14:58 INFO Job: map 100% reduce 100%  
23/01/04 15:14:58 INFO Job: Job job\_1672817468245\_0347 completed successfully  
23/01/04 15:14:58 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=573067  
FILE: Number of bytes written=1678727  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5564541  
HDFS: Number of bytes written=2014967  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16875  
Total time spent by all reduces in occupied slots (ms)=16820  
Total time spent by all map tasks (ms)=3375  
Total time spent by all reduce tasks (ms)=3364  
Total vcore-milliseconds taken by all map tasks=3375  
Total vcore-milliseconds taken by all reduce tasks=3364  
Total megabyte-milliseconds taken by all map tasks=17280000  
Total megabyte-milliseconds taken by all reduce tasks=17223680

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=573036  
Map output materialized bytes=573067  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=573067  
Reduce input records=5  
Reduce output records=15  
Spilled Records=10

Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=2900  
Physical memory (bytes) snapshot=694308864  
Virtual memory (bytes) snapshot=12757200896  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403939328  
Peak Map Virtual memory (bytes)=6384435200  
Peak Reduce Physical memory (bytes)=290369536  
Peak Reduce Virtual memory (bytes)=6374719488

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=2680676

File Output Format Counters

Bytes Written=2014967

23/01/04 15:14:58 INFO RepresentativePointsDriver: Representative Points Iteration 2  
23/01/04 15:14:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:14:58 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0348  
23/01/04 15:14:58 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:14:59 INFO JobSubmitter: number of splits:1  
23/01/04 15:14:59 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0348  
23/01/04 15:14:59 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:14:59 INFO YarnClientImpl: Submitted application application\_1672817468245\_0348  
23/01/04 15:14:59 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0348/  
23/01/04 15:14:59 INFO Job: Running job: job\_1672817468245\_0348  
23/01/04 15:15:05 INFO Job: Job job\_1672817468245\_0348 running in uber mode : false  
23/01/04 15:15:05 INFO Job: map 0% reduce 0%  
23/01/04 15:15:11 INFO Job: map 100% reduce 0%  
23/01/04 15:15:17 INFO Job: map 100% reduce 100%  
23/01/04 15:15:17 INFO Job: Job job\_1672817468245\_0348 completed successfully  
23/01/04 15:15:17 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=545216  
FILE: Number of bytes written=1623025  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6710813  
HDFS: Number of bytes written=2560212  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17030  
Total time spent by all reduces in occupied slots (ms)=16810  
Total time spent by all map tasks (ms)=3406  
Total time spent by all reduce tasks (ms)=3362  
Total vcore-milliseconds taken by all map tasks=3406  
Total vcore-milliseconds taken by all reduce tasks=3362  
Total megabyte-milliseconds taken by all map tasks=17438720  
Total megabyte-milliseconds taken by all reduce tasks=17213440

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=545185  
Map output materialized bytes=545216  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=545216  
Reduce input records=5  
Reduce output records=20  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=46  
CPU time spent (ms)=3220  
Physical memory (bytes) snapshot=707739648  
Virtual memory (bytes) snapshot=12744335360  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405151744  
Peak Map Virtual memory (bytes)=6369771520  
Peak Reduce Physical memory (bytes)=302587904  
Peak Reduce Virtual memory (bytes)=6374563840

#### 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=2680676

#### File Output Format Counters

Bytes Written=2560212

23/01/04 15:15:17 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:15:17 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:15:17 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0349

23/01/04 15:15:17 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:15:17 INFO JobSubmitter: number of splits:1

23/01/04 15:15:17 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0349

23/01/04 15:15:17 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:15:17 INFO YarnClientImpl: Submitted application application\_1672817468245\_0349

23/01/04 15:15:17 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0349/  
23/01/04 15:15:17 INFO Job: Running job: job\_1672817468245\_0349  
23/01/04 15:15:23 INFO Job: Job job\_1672817468245\_0349 running in uber mode : false  
23/01/04 15:15:23 INFO Job: map 0% reduce 0%  
23/01/04 15:15:28 INFO Job: map 100% reduce 0%  
23/01/04 15:15:34 INFO Job: map 100% reduce 100%  
23/01/04 15:15:34 INFO Job: Job job\_1672817468245\_0349 completed successfully  
23/01/04 15:15:34 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=571188  
FILE: Number of bytes written=1674969  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7801303  
HDFS: Number of bytes written=3131489  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16925  
Total time spent by all reduces in occupied slots (ms)=17175  
Total time spent by all map tasks (ms)=3385  
Total time spent by all reduce tasks (ms)=3435  
Total vcore-milliseconds taken by all map tasks=3385  
Total vcore-milliseconds taken by all reduce tasks=3435  
Total megabyte-milliseconds taken by all map tasks=17331200  
Total megabyte-milliseconds taken by all reduce tasks=17587200

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=571157  
Map output materialized bytes=571188  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=571188  
Reduce input records=5  
Reduce output records=25  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=26  
CPU time spent (ms)=3100  
Physical memory (bytes) snapshot=686628864  
Virtual memory (bytes) snapshot=12748718080  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=388706304  
Peak Map Virtual memory (bytes)=6368067584  
Peak Reduce Physical memory (bytes)=297922560  
Peak Reduce Virtual memory (bytes)=6384566272

#### 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=2680676
    File Output Format Counters
        Bytes Written=3131489
23/01/04 15:15:34 INFO RepresentativePointsDriver: Representative Points Iteration
4
23/01/04 15:15:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:15:34 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0351
23/01/04 15:15:34 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:15:35 INFO JobSubmitter: number of splits:1
23/01/04 15:15:35 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0351
23/01/04 15:15:35 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:15:35 INFO YarnClientImpl: Submitted application
application_1672817468245_0351
23/01/04 15:15:35 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0351/
23/01/04 15:15:35 INFO Job: Running job: job_1672817468245_0351
23/01/04 15:15:41 INFO Job: Job job_1672817468245_0351 running in uber mode : false
23/01/04 15:15:41 INFO Job:  map 0% reduce 0%
23/01/04 15:15:47 INFO Job:  map 100% reduce 0%
23/01/04 15:15:52 INFO Job:  map 100% reduce 100%
23/01/04 15:15:53 INFO Job: Job job_1672817468245_0351 completed successfully
23/01/04 15:15:53 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=569221
        FILE: Number of bytes written=1671035
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=8943857
        HDFS: Number of bytes written=3700739
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17555
        Total time spent by all reduces in occupied slots (ms)=16345
        Total time spent by all map tasks (ms)=3511
        Total time spent by all reduce tasks (ms)=3269
        Total vcore-milliseconds taken by all map tasks=3511
        Total vcore-milliseconds taken by all reduce tasks=3269
        Total megabyte-milliseconds taken by all map tasks=17976320
        Total megabyte-milliseconds taken by all reduce tasks=16737280
    Map-Reduce Framework
        Map input records=27
        Map output records=5

```

Map output bytes=569190  
Map output materialized bytes=569221  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=569221  
Reduce input records=5  
Reduce output records=30  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=80  
CPU time spent (ms)=3330  
Physical memory (bytes) snapshot=697921536  
Virtual memory (bytes) snapshot=12771930112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396460032  
Peak Map Virtual memory (bytes)=6375231488  
Peak Reduce Physical memory (bytes)=303063040  
Peak Reduce Virtual memory (bytes)=6398201856

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=2680676

File Output Format Counters

Bytes Written=3700739

23/01/04 15:15:53 INFO ClusterEvaluator: Scaled Inter-Cluster Density = 0.505781262060261

23/01/04 15:15:53 INFO ClusterEvaluator: Intra-Cluster Density[5] = NaN

23/01/04 15:15:53 INFO ClusterEvaluator: Intra-Cluster Density[1] = 0.5652519810480305

23/01/04 15:15:53 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN

23/01/04 15:15:53 INFO ClusterEvaluator: Intra-Cluster Density[7] = NaN

23/01/04 15:15:53 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN

23/01/04 15:15:53 INFO ClusterEvaluator: Average Intra-Cluster Density = 0.5652519810480305

23/01/04 15:15:54 INFO ClusterDumper: Wrote 5 clusters

23/01/04 15:15:54 INFO MahoutDriver: Program took 93559 ms (Minutes: 1.5593166666666667)

---- K: 06 --

dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/06

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:15:58 INFO AbstractJob: Command line arguments: {--clustering=null, --clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure], --convergenceDelta=[0.1], --distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors], --maxIter=[20], --method=[mapreduce], --numClusters=[6],

```

--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/06], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:15:59 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/06
23/01/04 15:15:59 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:15:59 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:15:59 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:15:59 INFO RandomSeedGenerator: Wrote 6 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:15:59 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/06
23/01/04 15:15:59 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:15:59 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:16:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:16:00 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0353
23/01/04 15:16:00 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:16:00 INFO JobSubmitter: number of splits:1
23/01/04 15:16:00 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0353
23/01/04 15:16:00 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:16:00 INFO Configuration: resource-types.xml not found
23/01/04 15:16:00 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:16:00 INFO YarnClientImpl: Submitted application
application_1672817468245_0353
23/01/04 15:16:00 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0353/
23/01/04 15:16:00 INFO Job: Running job: job_1672817468245_0353
23/01/04 15:16:08 INFO Job: Job job_1672817468245_0353 running in uber mode : false
23/01/04 15:16:08 INFO Job: map 0% reduce 0%
23/01/04 15:16:13 INFO Job: map 100% reduce 0%
23/01/04 15:16:19 INFO Job: map 100% reduce 100%
23/01/04 15:16:19 INFO Job: Job job_1672817468245_0353 completed successfully
23/01/04 15:16:19 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2561009
        FILE: Number of bytes written=5654309
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5885311
        HDFS: Number of bytes written=1765892
        HDFS: Number of read operations=39
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1

```

Total time spent by all maps in occupied slots (ms)=18020  
Total time spent by all reduces in occupied slots (ms)=17690  
Total time spent by all map tasks (ms)=3604  
Total time spent by all reduce tasks (ms)=3538  
Total vcore-milliseconds taken by all map tasks=3604  
Total vcore-milliseconds taken by all reduce tasks=3538  
Total megabyte-milliseconds taken by all map tasks=18452480  
Total megabyte-milliseconds taken by all reduce tasks=18114560

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=2560973  
Map output materialized bytes=2561009  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=2561009  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=56  
CPU time spent (ms)=4260  
Physical memory (bytes) snapshot=708702208  
Virtual memory (bytes) snapshot=12748537856  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399880192  
Peak Map Virtual memory (bytes)=6370304000  
Peak Reduce Physical memory (bytes)=308822016  
Peak Reduce Virtual memory (bytes)=6378233856

#### 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=2680027

#### File Output Format Counters

Bytes Written=1765892

23/01/04 15:16:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:16:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0355  
23/01/04 15:16:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:16:19 INFO JobSubmitter: number of splits:1  
23/01/04 15:16:19 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0355  
23/01/04 15:16:19 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:16:19 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0355  
23/01/04 15:16:19 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0355/](http://lena-master:8088/proxy/application_1672817468245_0355/)  
23/01/04 15:16:19 INFO Job: Running job: job\_1672817468245\_0355  
23/01/04 15:16:25 INFO Job: Job job\_1672817468245\_0355 running in uber mode : false

23/01/04 15:16:25 INFO Job: map 0% reduce 0%  
23/01/04 15:16:31 INFO Job: map 100% reduce 0%  
23/01/04 15:16:36 INFO Job: map 100% reduce 100%  
23/01/04 15:16:36 INFO Job: Job job\_1672817468245\_0355 completed successfully  
23/01/04 15:16:36 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3791363  
FILE: Number of bytes written=8115017  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6212369  
HDFS: Number of bytes written=1765450  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18420  
Total time spent by all reduces in occupied slots (ms)=17520  
Total time spent by all map tasks (ms)=3684  
Total time spent by all reduce tasks (ms)=3504  
Total vcore-milliseconds taken by all map tasks=3684  
Total vcore-milliseconds taken by all reduce tasks=3504  
Total megabyte-milliseconds taken by all map tasks=18862080  
Total megabyte-milliseconds taken by all reduce tasks=17940480

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3791327  
Map output materialized bytes=3791363  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3791363  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=74  
CPU time spent (ms)=4020  
Physical memory (bytes) snapshot=723759104  
Virtual memory (bytes) snapshot=12753772544  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409272320  
Peak Map Virtual memory (bytes)=6370189312  
Peak Reduce Physical memory (bytes)=314486784  
Peak Reduce Virtual memory (bytes)=6383583232

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1765450  
23/01/04 15:16:36 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:16:36 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0357  
23/01/04 15:16:37 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:16:37 INFO JobSubmitter: number of splits:1  
23/01/04 15:16:37 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0357  
23/01/04 15:16:37 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:16:37 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0357  
23/01/04 15:16:37 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0357/](http://lena-master:8088/proxy/application_1672817468245_0357/)  
23/01/04 15:16:37 INFO Job: Running job: job\_1672817468245\_0357  
23/01/04 15:16:43 INFO Job: Job job\_1672817468245\_0357 running in uber mode : false  
23/01/04 15:16:43 INFO Job:  map 0% reduce 0%  
23/01/04 15:16:49 INFO Job:  map 100% reduce 0%  
23/01/04 15:16:55 INFO Job:  map 100% reduce 100%  
23/01/04 15:16:55 INFO Job: Job job\_1672817468245\_0357 completed successfully  
23/01/04 15:16:55 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3762995  
FILE: Number of bytes written=8058281  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6211485  
HDFS: Number of bytes written=1760150  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18110  
Total time spent by all reduces in occupied slots (ms)=17640  
Total time spent by all map tasks (ms)=3622  
Total time spent by all reduce tasks (ms)=3528  
Total vcore-milliseconds taken by all map tasks=3622  
Total vcore-milliseconds taken by all reduce tasks=3528  
Total megabyte-milliseconds taken by all map tasks=18544640  
Total megabyte-milliseconds taken by all reduce tasks=18063360

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3762959  
Map output materialized bytes=3762995  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6

Reduce shuffle bytes=3762995  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=722173952  
Virtual memory (bytes) snapshot=12755292160  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419135488  
Peak Map Virtual memory (bytes)=6373580800  
Peak Reduce Physical memory (bytes)=303038464  
Peak Reduce Virtual memory (bytes)=6381711360

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=2680027

File Output Format Counters

Bytes Written=1760150

23/01/04 15:16:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:16:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0359  
23/01/04 15:16:55 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:16:55 INFO JobSubmitter: number of splits:1  
23/01/04 15:16:55 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0359  
23/01/04 15:16:55 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:16:55 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0359  
23/01/04 15:16:55 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0359/](http://lena-master:8088/proxy/application_1672817468245_0359/)  
23/01/04 15:16:55 INFO Job: Running job: job\_1672817468245\_0359  
23/01/04 15:17:01 INFO Job: Job job\_1672817468245\_0359 running in uber mode : false  
23/01/04 15:17:01 INFO Job: map 0% reduce 0%  
23/01/04 15:17:07 INFO Job: map 100% reduce 0%  
23/01/04 15:17:12 INFO Job: map 100% reduce 100%  
23/01/04 15:17:12 INFO Job: Job job\_1672817468245\_0359 completed successfully  
23/01/04 15:17:12 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3742449  
FILE: Number of bytes written=8017189  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6200885  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17955  
Total time spent by all reduces in occupied slots (ms)=17250  
Total time spent by all map tasks (ms)=3591  
Total time spent by all reduce tasks (ms)=3450  
Total vcore-milliseconds taken by all map tasks=3591  
Total vcore-milliseconds taken by all reduce tasks=3450  
Total megabyte-milliseconds taken by all map tasks=18385920  
Total megabyte-milliseconds taken by all reduce tasks=17664000

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3742413  
Map output materialized bytes=3742449  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3742449  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=4110  
Physical memory (bytes) snapshot=758816768  
Virtual memory (bytes) snapshot=12765216768  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411897856  
Peak Map Virtual memory (bytes)=6372028416  
Peak Reduce Physical memory (bytes)=346918912  
Peak Reduce Virtual memory (bytes)=6393188352

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904

23/01/04 15:17:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:17:13 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0361  
23/01/04 15:17:13 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:17:13 INFO JobSubmitter: number of splits:1  
23/01/04 15:17:13 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0361  
23/01/04 15:17:13 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:17:13 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0361



23/01/04 15:17:13 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0361/  
23/01/04 15:17:13 INFO Job: Running job: job\_1672817468245\_0361  
23/01/04 15:17:19 INFO Job: Job job\_1672817468245\_0361 running in uber mode : false  
23/01/04 15:17:19 INFO Job: map 0% reduce 0%  
23/01/04 15:17:25 INFO Job: map 100% reduce 0%  
23/01/04 15:17:31 INFO Job: map 100% reduce 100%  
23/01/04 15:17:31 INFO Job: Job job\_1672817468245\_0361 completed successfully  
23/01/04 15:17:31 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986697  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17910  
Total time spent by all reduces in occupied slots (ms)=17360  
Total time spent by all map tasks (ms)=3582  
Total time spent by all reduce tasks (ms)=3472  
Total vcore-milliseconds taken by all map tasks=3582  
Total vcore-milliseconds taken by all reduce tasks=3472  
Total megabyte-milliseconds taken by all map tasks=18339840  
Total megabyte-milliseconds taken by all reduce tasks=17776640

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=3940  
Physical memory (bytes) snapshot=730484736  
Virtual memory (bytes) snapshot=12754128896  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406151168  
Peak Map Virtual memory (bytes)=6378283008  
Peak Reduce Physical memory (bytes)=324333568  
Peak Reduce Virtual memory (bytes)=6375845888

#### 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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:17:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:17:31 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0362  
23/01/04 15:17:31 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:17:31 INFO JobSubmitter: number of splits:1  
23/01/04 15:17:31 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0362  
23/01/04 15:17:31 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:17:31 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0362  
23/01/04 15:17:31 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0362/  
23/01/04 15:17:31 INFO Job: Running job: job\_1672817468245\_0362  
23/01/04 15:17:37 INFO Job: Job job\_1672817468245\_0362 running in uber mode : false  
23/01/04 15:17:37 INFO Job: map 0% reduce 0%  
23/01/04 15:17:44 INFO Job: map 100% reduce 0%  
23/01/04 15:17:48 INFO Job: map 100% reduce 100%  
23/01/04 15:17:49 INFO Job: Job job\_1672817468245\_0362 completed successfully  
23/01/04 15:17:49 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986697  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17870  
Total time spent by all reduces in occupied slots (ms)=10270  
Total time spent by all map tasks (ms)=3574  
Total time spent by all reduce tasks (ms)=2054  
Total vcore-milliseconds taken by all map tasks=3574  
Total vcore-milliseconds taken by all reduce tasks=2054  
Total megabyte-milliseconds taken by all map tasks=18298880  
Total megabyte-milliseconds taken by all reduce tasks=10516480

Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203

Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=4100  
Physical memory (bytes) snapshot=713105408  
Virtual memory (bytes) snapshot=12755116032  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404885504  
Peak Map Virtual memory (bytes)=6376636416  
Peak Reduce Physical memory (bytes)=308219904  
Peak Reduce Virtual memory (bytes)=6378479616

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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:17:49 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:17:49 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0365  
23/01/04 15:17:49 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:17:49 INFO JobSubmitter: number of splits:1  
23/01/04 15:17:49 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0365  
23/01/04 15:17:49 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:17:49 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0365  
23/01/04 15:17:49 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0365/  
23/01/04 15:17:49 INFO Job: Running job: job\_1672817468245\_0365  
23/01/04 15:17:55 INFO Job: Job job\_1672817468245\_0365 running in uber mode : false  
23/01/04 15:17:55 INFO Job: map 0% reduce 0%  
23/01/04 15:18:01 INFO Job: map 100% reduce 0%  
23/01/04 15:18:07 INFO Job: map 100% reduce 100%  
23/01/04 15:18:07 INFO Job: Job job\_1672817468245\_0365 completed successfully  
23/01/04 15:18:07 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986697  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904

HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17995  
Total time spent by all reduces in occupied slots (ms)=17380  
Total time spent by all map tasks (ms)=3599  
Total time spent by all reduce tasks (ms)=3476  
Total vcore-milliseconds taken by all map tasks=3599  
Total vcore-milliseconds taken by all reduce tasks=3476  
Total megabyte-milliseconds taken by all map tasks=18426880  
Total megabyte-milliseconds taken by all reduce tasks=17797120

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=4120  
Physical memory (bytes) snapshot=727429120  
Virtual memory (bytes) snapshot=12769017856  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=416055296  
Peak Map Virtual memory (bytes)=6378958848  
Peak Reduce Physical memory (bytes)=311373824  
Peak Reduce Virtual memory (bytes)=6390059008

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904

23/01/04 15:18:07 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:18:07 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0367

23/01/04 15:18:07 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:18:07 INFO JobSubmitter: number of splits:1

23/01/04 15:18:07 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0367

23/01/04 15:18:07 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:18:07 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0367

23/01/04 15:18:07 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0367/](http://lena-master:8088/proxy/application_1672817468245_0367/)

23/01/04 15:18:07 INFO Job: Running job: job\_1672817468245\_0367

23/01/04 15:18:15 INFO Job: Job job\_1672817468245\_0367 running in uber mode : false

23/01/04 15:18:15 INFO Job: map 0% reduce 0%

23/01/04 15:18:21 INFO Job: map 100% reduce 0%

23/01/04 15:18:26 INFO Job: map 100% reduce 100%

23/01/04 15:18:27 INFO Job: Job job\_1672817468245\_0367 completed successfully

23/01/04 15:18:27 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203

FILE: Number of bytes written=7986697

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=6170393

HDFS: Number of bytes written=1744904

HDFS: Number of read operations=19

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=18115

Total time spent by all reduces in occupied slots (ms)=17615

Total time spent by all map tasks (ms)=3623

Total time spent by all reduce tasks (ms)=3523

Total vcore-milliseconds taken by all map tasks=3623

Total vcore-milliseconds taken by all reduce tasks=3523

Total megabyte-milliseconds taken by all map tasks=18549760

Total megabyte-milliseconds taken by all reduce tasks=18037760

#### Map-Reduce Framework

Map input records=27

Map output records=6

Map output bytes=3727167

Map output materialized bytes=3727203

Input split bytes=170

Combine input records=0

Combine output records=0

Reduce input groups=6

Reduce shuffle bytes=3727203

Reduce input records=6

Reduce output records=6

Spilled Records=12

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=54

CPU time spent (ms)=4100

Physical memory (bytes) snapshot=735227904

Virtual memory (bytes) snapshot=12765249536

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=423088128

```

Peak Map Virtual memory (bytes)=6370627584
Peak Reduce Physical memory (bytes)=312139776
Peak Reduce Virtual memory (bytes)=6394621952
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=2680027
File Output Format Counters
  Bytes Written=1744904
23/01/04 15:18:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:18:27 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0369
23/01/04 15:18:27 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:18:27 INFO JobSubmitter: number of splits:1
23/01/04 15:18:27 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0369
23/01/04 15:18:27 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:18:27 INFO YarnClientImpl: Submitted application
application_1672817468245_0369
23/01/04 15:18:27 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0369/
23/01/04 15:18:27 INFO Job: Running job: job_1672817468245_0369
23/01/04 15:18:33 INFO Job: Job job_1672817468245_0369 running in uber mode : false
23/01/04 15:18:33 INFO Job:  map 0% reduce 0%
23/01/04 15:18:39 INFO Job:  map 100% reduce 0%
23/01/04 15:18:45 INFO Job:  map 100% reduce 100%
23/01/04 15:18:45 INFO Job: Job job_1672817468245_0369 completed successfully
23/01/04 15:18:45 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=3727203
  FILE: Number of bytes written=7986697
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6170393
  HDFS: Number of bytes written=1744904
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17795
  Total time spent by all reduces in occupied slots (ms)=17465
  Total time spent by all map tasks (ms)=3559
  Total time spent by all reduce tasks (ms)=3493
  Total vcore-milliseconds taken by all map tasks=3559
  Total vcore-milliseconds taken by all reduce tasks=3493
  Total megabyte-milliseconds taken by all map tasks=18222080
  Total megabyte-milliseconds taken by all reduce tasks=17884160
Map-Reduce Framework

```

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=4040  
Physical memory (bytes) snapshot=714182656  
Virtual memory (bytes) snapshot=12750450688  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404619264  
Peak Map Virtual memory (bytes)=6370095104  
Peak Reduce Physical memory (bytes)=309563392  
Peak Reduce Virtual memory (bytes)=6380355584

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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:18:45 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:18:45 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0370  
23/01/04 15:18:45 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:18:45 INFO JobSubmitter: number of splits:1  
23/01/04 15:18:45 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0370  
23/01/04 15:18:45 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:18:45 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0370  
23/01/04 15:18:45 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0370/  
23/01/04 15:18:45 INFO Job: Running job: job\_1672817468245\_0370  
23/01/04 15:18:52 INFO Job: Job job\_1672817468245\_0370 running in uber mode : false  
23/01/04 15:18:52 INFO Job: map 0% reduce 0%  
23/01/04 15:18:57 INFO Job: map 100% reduce 0%  
23/01/04 15:19:03 INFO Job: map 100% reduce 100%  
23/01/04 15:19:03 INFO Job: Job job\_1672817468245\_0370 completed successfully  
23/01/04 15:19:03 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986701  
FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18615  
Total time spent by all reduces in occupied slots (ms)=17035  
Total time spent by all map tasks (ms)=3723  
Total time spent by all reduce tasks (ms)=3407  
Total vcore-milliseconds taken by all map tasks=3723  
Total vcore-milliseconds taken by all reduce tasks=3407  
Total megabyte-milliseconds taken by all map tasks=19061760  
Total megabyte-milliseconds taken by all reduce tasks=17443840

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=4100  
Physical memory (bytes) snapshot=739270656  
Virtual memory (bytes) snapshot=12752138240  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=407977984  
Peak Map Virtual memory (bytes)=6374916096  
Peak Reduce Physical memory (bytes)=331292672  
Peak Reduce Virtual memory (bytes)=6377222144

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904

23/01/04 15:19:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:19:03 INFO JobResourceUploader: Disabling Erasure Coding for path:



/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0372  
23/01/04 15:19:03 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:19:03 INFO JobSubmitter: number of splits:1  
23/01/04 15:19:03 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0372  
23/01/04 15:19:03 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:19:03 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0372  
23/01/04 15:19:03 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0372/  
23/01/04 15:19:03 INFO Job: Running job: job\_1672817468245\_0372  
23/01/04 15:19:09 INFO Job: Job job\_1672817468245\_0372 running in uber mode : false  
23/01/04 15:19:09 INFO Job: map 0% reduce 0%  
23/01/04 15:19:15 INFO Job: map 100% reduce 0%  
23/01/04 15:19:19 INFO Job: map 100% reduce 100%  
23/01/04 15:19:19 INFO Job: Job job\_1672817468245\_0372 completed successfully  
23/01/04 15:19:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18370  
Total time spent by all reduces in occupied slots (ms)=10230  
Total time spent by all map tasks (ms)=3674  
Total time spent by all reduce tasks (ms)=2046  
Total vcore-milliseconds taken by all map tasks=3674  
Total vcore-milliseconds taken by all reduce tasks=2046  
Total megabyte-milliseconds taken by all map tasks=18810880  
Total megabyte-milliseconds taken by all reduce tasks=10475520

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=3960

Physical memory (bytes) snapshot=725819392  
Virtual memory (bytes) snapshot=12752277504  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413499392  
Peak Map Virtual memory (bytes)=6373732352  
Peak Reduce Physical memory (bytes)=312320000  
Peak Reduce Virtual memory (bytes)=6378545152

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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:19:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:19:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0374  
23/01/04 15:19:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:19:19 INFO JobSubmitter: number of splits:1  
23/01/04 15:19:19 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0374  
23/01/04 15:19:19 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:19:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0374  
23/01/04 15:19:20 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0374/](http://lena-master:8088/proxy/application_1672817468245_0374/)  
23/01/04 15:19:20 INFO Job: Running job: job\_1672817468245\_0374  
23/01/04 15:19:26 INFO Job: Job job\_1672817468245\_0374 running in uber mode : false  
23/01/04 15:19:26 INFO Job: map 0% reduce 0%  
23/01/04 15:19:32 INFO Job: map 100% reduce 0%  
23/01/04 15:19:39 INFO Job: map 100% reduce 100%  
23/01/04 15:19:39 INFO Job: Job job\_1672817468245\_0374 completed successfully  
23/01/04 15:19:39 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17980  
Total time spent by all reduces in occupied slots (ms)=17085  
Total time spent by all map tasks (ms)=3596  
Total time spent by all reduce tasks (ms)=3417  
Total vcore-milliseconds taken by all map tasks=3596

Total vcore-milliseconds taken by all reduce tasks=3417  
Total megabyte-milliseconds taken by all map tasks=18411520  
Total megabyte-milliseconds taken by all reduce tasks=17495040  
Map-Reduce Framework  
Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=54  
CPU time spent (ms)=3910  
Physical memory (bytes) snapshot=730755072  
Virtual memory (bytes) snapshot=12761493504  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402714624  
Peak Map Virtual memory (bytes)=6370533376  
Peak Reduce Physical memory (bytes)=328040448  
Peak Reduce Virtual memory (bytes)=6390960128

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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:19:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:19:39 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0376  
23/01/04 15:19:39 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:19:39 INFO JobSubmitter: number of splits:1  
23/01/04 15:19:39 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0376  
23/01/04 15:19:39 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:19:39 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0376  
23/01/04 15:19:39 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0376/  
23/01/04 15:19:39 INFO Job: Running job: job\_1672817468245\_0376  
23/01/04 15:19:45 INFO Job: Job job\_1672817468245\_0376 running in uber mode : false  
23/01/04 15:19:45 INFO Job: map 0% reduce 0%  
23/01/04 15:19:51 INFO Job: map 100% reduce 0%  
23/01/04 15:19:57 INFO Job: map 100% reduce 100%  
23/01/04 15:19:57 INFO Job: Job job\_1672817468245\_0376 completed successfully  
23/01/04 15:19:57 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17615  
Total time spent by all reduces in occupied slots (ms)=17390  
Total time spent by all map tasks (ms)=3523  
Total time spent by all reduce tasks (ms)=3478  
Total vcore-milliseconds taken by all map tasks=3523  
Total vcore-milliseconds taken by all reduce tasks=3478  
Total megabyte-milliseconds taken by all map tasks=18037760  
Total megabyte-milliseconds taken by all reduce tasks=17807360

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3920  
Physical memory (bytes) snapshot=712331264  
Virtual memory (bytes) snapshot=12750901248  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408363008  
Peak Map Virtual memory (bytes)=6370516992  
Peak Reduce Physical memory (bytes)=303968256  
Peak Reduce Virtual memory (bytes)=6380384256

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904  
23/01/04 15:19:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:19:57 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0378  
23/01/04 15:19:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:19:58 INFO JobSubmitter: number of splits:1  
23/01/04 15:19:58 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0378  
23/01/04 15:19:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:19:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0378  
23/01/04 15:19:58 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0378/  
23/01/04 15:19:58 INFO Job: Running job: job\_1672817468245\_0378  
23/01/04 15:20:05 INFO Job: Job job\_1672817468245\_0378 running in uber mode : false  
23/01/04 15:20:05 INFO Job: map 0% reduce 0%  
23/01/04 15:20:10 INFO Job: map 100% reduce 0%  
23/01/04 15:20:16 INFO Job: map 100% reduce 100%  
23/01/04 15:20:16 INFO Job: Job job\_1672817468245\_0378 completed successfully  
23/01/04 15:20:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17870  
Total time spent by all reduces in occupied slots (ms)=17415  
Total time spent by all map tasks (ms)=3574  
Total time spent by all reduce tasks (ms)=3483  
Total vcore-milliseconds taken by all map tasks=3574  
Total vcore-milliseconds taken by all reduce tasks=3483  
Total megabyte-milliseconds taken by all map tasks=18298880  
Total megabyte-milliseconds taken by all reduce tasks=17832960

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1

```

Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=72
CPU time spent (ms)=4050
Physical memory (bytes) snapshot=709324800
Virtual memory (bytes) snapshot=12750786560
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=410857472
Peak Map Virtual memory (bytes)=6374907904
Peak Reduce Physical memory (bytes)=298467328
Peak Reduce Virtual memory (bytes)=6375878656
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=2680027
File Output Format Counters
  Bytes Written=1744904
23/01/04 15:20:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:20:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0380
23/01/04 15:20:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:20:16 INFO JobSubmitter: number of splits:1
23/01/04 15:20:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0380
23/01/04 15:20:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:20:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0380
23/01/04 15:20:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0380/
23/01/04 15:20:16 INFO Job: Running job: job_1672817468245_0380
23/01/04 15:20:22 INFO Job: Job job_1672817468245_0380 running in uber mode : false
23/01/04 15:20:22 INFO Job:  map 0% reduce 0%
23/01/04 15:20:28 INFO Job:  map 100% reduce 0%
23/01/04 15:20:34 INFO Job:  map 100% reduce 100%
23/01/04 15:20:34 INFO Job: Job job_1672817468245_0380 completed successfully
23/01/04 15:20:34 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=3727203
  FILE: Number of bytes written=7986705
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6170393
  HDFS: Number of bytes written=1744904
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17810

```

Total time spent by all reduces in occupied slots (ms)=17710  
Total time spent by all map tasks (ms)=3562  
Total time spent by all reduce tasks (ms)=3542  
Total vcore-milliseconds taken by all map tasks=3562  
Total vcore-milliseconds taken by all reduce tasks=3542  
Total megabyte-milliseconds taken by all map tasks=18237440  
Total megabyte-milliseconds taken by all reduce tasks=18135040

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=4390  
Physical memory (bytes) snapshot=739287040  
Virtual memory (bytes) snapshot=12757250048  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408694784  
Peak Map Virtual memory (bytes)=6383210496  
Peak Reduce Physical memory (bytes)=330592256  
Peak Reduce Virtual memory (bytes)=6374039552

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904

23/01/04 15:20:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:20:34 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0382  
23/01/04 15:20:34 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:20:35 INFO JobSubmitter: number of splits:1  
23/01/04 15:20:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0382  
23/01/04 15:20:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:20:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0382  
23/01/04 15:20:35 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0382/  
23/01/04 15:20:35 INFO Job: Running job: job\_1672817468245\_0382  
23/01/04 15:20:41 INFO Job: Job job\_1672817468245\_0382 running in uber mode : false  
23/01/04 15:20:41 INFO Job: map 0% reduce 0%

23/01/04 15:20:47 INFO Job: map 100% reduce 0%  
23/01/04 15:20:53 INFO Job: map 100% reduce 100%  
23/01/04 15:20:53 INFO Job: Job job\_1672817468245\_0382 completed successfully  
23/01/04 15:20:53 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17565  
Total time spent by all reduces in occupied slots (ms)=9975  
Total time spent by all map tasks (ms)=3513  
Total time spent by all reduce tasks (ms)=1995  
Total vcore-milliseconds taken by all map tasks=3513  
Total vcore-milliseconds taken by all reduce tasks=1995  
Total megabyte-milliseconds taken by all map tasks=17986560  
Total megabyte-milliseconds taken by all reduce tasks=10214400

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=3940  
Physical memory (bytes) snapshot=761249792  
Virtual memory (bytes) snapshot=12756865024  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=424366080  
Peak Map Virtual memory (bytes)=6370885632  
Peak Reduce Physical memory (bytes)=336883712  
Peak Reduce Virtual memory (bytes)=6385979392

#### 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=2680027
    File Output Format Counters
        Bytes Written=1744904
23/01/04 15:20:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:20:53 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0384
23/01/04 15:20:53 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:20:53 INFO JobSubmitter: number of splits:1
23/01/04 15:20:53 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0384
23/01/04 15:20:53 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:20:53 INFO YarnClientImpl: Submitted application
application_1672817468245_0384
23/01/04 15:20:53 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0384/
23/01/04 15:20:53 INFO Job: Running job: job_1672817468245_0384
23/01/04 15:20:59 INFO Job: Job job_1672817468245_0384 running in uber mode : false
23/01/04 15:20:59 INFO Job:  map 0% reduce 0%
23/01/04 15:21:05 INFO Job:  map 100% reduce 0%
23/01/04 15:21:09 INFO Job:  map 100% reduce 100%
23/01/04 15:21:10 INFO Job: Job job_1672817468245_0384 completed successfully
23/01/04 15:21:10 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=3727203
        FILE: Number of bytes written=7986705
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=6170393
        HDFS: Number of bytes written=1744904
        HDFS: Number of read operations=19
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17950
        Total time spent by all reduces in occupied slots (ms)=10320
        Total time spent by all map tasks (ms)=3590
        Total time spent by all reduce tasks (ms)=2064
        Total vcore-milliseconds taken by all map tasks=3590
        Total vcore-milliseconds taken by all reduce tasks=2064
        Total megabyte-milliseconds taken by all map tasks=18380800
        Total megabyte-milliseconds taken by all reduce tasks=10567680
    Map-Reduce Framework
        Map input records=27
        Map output records=6
        Map output bytes=3727167
        Map output materialized bytes=3727203
        Input split bytes=170
        Combine input records=0
        Combine output records=0
        Reduce input groups=6
        Reduce shuffle bytes=3727203

```

Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=90  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=727756800  
Virtual memory (bytes) snapshot=12769751040  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412069888  
Peak Map Virtual memory (bytes)=6377357312  
Peak Reduce Physical memory (bytes)=315686912  
Peak Reduce Virtual memory (bytes)=6392393728

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=2680027

File Output Format Counters

Bytes Written=1744904

23/01/04 15:21:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:21:10 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0386  
23/01/04 15:21:10 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:21:11 INFO JobSubmitter: number of splits:1  
23/01/04 15:21:11 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0386  
23/01/04 15:21:11 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:21:11 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0386  
23/01/04 15:21:11 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0386/](http://lena-master:8088/proxy/application_1672817468245_0386/)  
23/01/04 15:21:11 INFO Job: Running job: job\_1672817468245\_0386  
23/01/04 15:21:17 INFO Job: Job job\_1672817468245\_0386 running in uber mode : false  
23/01/04 15:21:17 INFO Job: map 0% reduce 0%  
23/01/04 15:21:23 INFO Job: map 100% reduce 0%  
23/01/04 15:21:27 INFO Job: map 100% reduce 100%  
23/01/04 15:21:27 INFO Job: Job job\_1672817468245\_0386 completed successfully  
23/01/04 15:21:27 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17740  
Total time spent by all reduces in occupied slots (ms)=9955  
Total time spent by all map tasks (ms)=3548  
Total time spent by all reduce tasks (ms)=1991  
Total vcore-milliseconds taken by all map tasks=3548  
Total vcore-milliseconds taken by all reduce tasks=1991  
Total megabyte-milliseconds taken by all map tasks=18165760  
Total megabyte-milliseconds taken by all reduce tasks=10193920

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=81  
CPU time spent (ms)=4290  
Physical memory (bytes) snapshot=727699456  
Virtual memory (bytes) snapshot=12756885504  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401948672  
Peak Map Virtual memory (bytes)=6372311040  
Peak Reduce Physical memory (bytes)=325750784  
Peak Reduce Virtual memory (bytes)=6384574464

#### 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=2680027

#### File Output Format Counters

Bytes Written=1744904

23/01/04 15:21:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:21:27 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0388  
23/01/04 15:21:27 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:21:27 INFO JobSubmitter: number of splits:1  
23/01/04 15:21:27 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0388  
23/01/04 15:21:27 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:21:27 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0388  
23/01/04 15:21:27 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0388/  
23/01/04 15:21:27 INFO Job: Running job: job\_1672817468245\_0388  
23/01/04 15:21:33 INFO Job: Job job\_1672817468245\_0388 running in uber mode : false  
23/01/04 15:21:33 INFO Job: map 0% reduce 0%  
23/01/04 15:21:38 INFO Job: map 100% reduce 0%  
23/01/04 15:21:43 INFO Job: map 100% reduce 100%  
23/01/04 15:21:43 INFO Job: Job job\_1672817468245\_0388 completed successfully  
23/01/04 15:21:43 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3727203  
FILE: Number of bytes written=7986705  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6170393  
HDFS: Number of bytes written=1744904  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18365  
Total time spent by all reduces in occupied slots (ms)=10035  
Total time spent by all map tasks (ms)=3673  
Total time spent by all reduce tasks (ms)=2007  
Total vcore-milliseconds taken by all map tasks=3673  
Total vcore-milliseconds taken by all reduce tasks=2007  
Total megabyte-milliseconds taken by all map tasks=18805760  
Total megabyte-milliseconds taken by all reduce tasks=10275840

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=3727167  
Map output materialized bytes=3727203  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=3727203  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=3930  
Physical memory (bytes) snapshot=724361216  
Virtual memory (bytes) snapshot=12759719936  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395636736  
Peak Map Virtual memory (bytes)=6380593152  
Peak Reduce Physical memory (bytes)=328724480  
Peak Reduce Virtual memory (bytes)=6379126784

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1744904  
23/01/04 15:21:43 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:21:43 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0391  
23/01/04 15:21:44 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:21:44 INFO JobSubmitter: number of splits:1  
23/01/04 15:21:44 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0391  
23/01/04 15:21:44 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:21:44 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0391  
23/01/04 15:21:44 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0391/  
23/01/04 15:21:44 INFO Job: Running job: job\_1672817468245\_0391  
23/01/04 15:21:50 INFO Job: Job job\_1672817468245\_0391 running in uber mode : false  
23/01/04 15:21:50 INFO Job: map 0% reduce 0%  
23/01/04 15:21:56 INFO Job: map 100% reduce 0%  
23/01/04 15:22:02 INFO Job: map 100% reduce 100%  
23/01/04 15:22:02 INFO Job: Job job\_1672817468245\_0391 completed successfully  
23/01/04 15:22:02 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=3727203  
  FILE: Number of bytes written=7986705  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=6170393  
  HDFS: Number of bytes written=1744904  
  HDFS: Number of read operations=19  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17890  
  Total time spent by all reduces in occupied slots (ms)=17755  
  Total time spent by all map tasks (ms)=3578  
  Total time spent by all reduce tasks (ms)=3551  
  Total vcore-milliseconds taken by all map tasks=3578  
  Total vcore-milliseconds taken by all reduce tasks=3551  
  Total megabyte-milliseconds taken by all map tasks=18319360  
  Total megabyte-milliseconds taken by all reduce tasks=18181120  
Map-Reduce Framework  
  Map input records=27  
  Map output records=6  
  Map output bytes=3727167  
  Map output materialized bytes=3727203  
  Input split bytes=170

```

Combine input records=0
Combine output records=0
Reduce input groups=6
Reduce shuffle bytes=3727203
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=89
CPU time spent (ms)=4410
Physical memory (bytes) snapshot=731828224
Virtual memory (bytes) snapshot=12761268224
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=409575424
Peak Map Virtual memory (bytes)=6375604224
Peak Reduce Physical memory (bytes)=322252800
Peak Reduce Virtual memory (bytes)=6385664000
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=2680027
File Output Format Counters
  Bytes Written=1744904
23/01/04 15:22:02 INFO KMeansDriver: Clustering data
23/01/04 15:22:02 INFO KMeansDriver: Running Clustering
23/01/04 15:22:02 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/06 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/06
23/01/04 15:22:02 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:22:02 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0394
23/01/04 15:22:02 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:22:02 INFO JobSubmitter: number of splits:1
23/01/04 15:22:02 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0394
23/01/04 15:22:02 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:22:02 INFO YarnClientImpl: Submitted application
application_1672817468245_0394
23/01/04 15:22:02 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0394/
23/01/04 15:22:02 INFO Job: Running job: job_1672817468245_0394
23/01/04 15:22:08 INFO Job: Job job_1672817468245_0394 running in uber mode : false
23/01/04 15:22:08 INFO Job:  map 0% reduce 0%
23/01/04 15:22:14 INFO Job:  map 100% reduce 0%
23/01/04 15:22:14 INFO Job: Job job_1672817468245_0394 completed successfully
23/01/04 15:22:14 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0

```

FILE: Number of bytes written=265898  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4425295  
HDFS: Number of bytes written=2680676  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17460  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3492  
Total vcore-milliseconds taken by all map tasks=3492  
Total megabyte-milliseconds taken by all map tasks=17879040

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=21  
CPU time spent (ms)=2140  
Physical memory (bytes) snapshot=290557952  
Virtual memory (bytes) snapshot=6372904960  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=290557952  
Peak Map Virtual memory (bytes)=6372904960

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680676

23/01/04 15:22:14 INFO MahoutDriver: Program took 376205 ms (Minutes:  
6.270083333333333)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:22:20 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/EuclideanDistanceMeasure/06/clusters-20-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
EuclideanDistanceMeasure/06.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu  
re/06/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 15:22:22 INFO HadoopUtil: Deleting tmp/representative

23/01/04 15:22:22 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/06/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --  
endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/

```
06/clusters-20-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:22:22 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:22:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:22:22 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0398
23/01/04 15:22:22 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:22:22 INFO JobSubmitter: number of splits:1
23/01/04 15:22:23 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0398
23/01/04 15:22:23 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:22:23 INFO Configuration: resource-types.xml not found
23/01/04 15:22:23 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:22:23 INFO YarnClientImpl: Submitted application
application_1672817468245_0398
23/01/04 15:22:23 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0398/
23/01/04 15:22:23 INFO Job: Running job: job_1672817468245_0398
23/01/04 15:22:29 INFO Job: Job job_1672817468245_0398 running in uber mode : false
23/01/04 15:22:29 INFO Job:  map 0% reduce 0%
23/01/04 15:22:35 INFO Job:  map 100% reduce 0%
23/01/04 15:22:40 INFO Job:  map 100% reduce 100%
23/01/04 15:22:41 INFO Job: Job job_1672817468245_0398 completed successfully
23/01/04 15:22:41 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=607462
FILE: Number of bytes written=1747517
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=4663941
HDFS: Number of bytes written=1599018
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16565
Total time spent by all reduces in occupied slots (ms)=16855
Total time spent by all map tasks (ms)=3313
Total time spent by all reduce tasks (ms)=3371
Total vcore-milliseconds taken by all map tasks=3313
Total vcore-milliseconds taken by all reduce tasks=3371
Total megabyte-milliseconds taken by all map tasks=16962560
Total megabyte-milliseconds taken by all reduce tasks=17259520
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=6
Map output bytes=607427
Map output materialized bytes=607462
Input split bytes=203
Combine input records=0
Combine output records=0
Reduce input groups=6
```



Reduce shuffle bytes=607462  
Reduce input records=6  
Reduce output records=12  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=2800  
Physical memory (bytes) snapshot=676691968  
Virtual memory (bytes) snapshot=12750860288  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=378564608  
Peak Map Virtual memory (bytes)=6372839424  
Peak Reduce Physical memory (bytes)=298127360  
Peak Reduce Virtual memory (bytes)=6381936640

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=2680676

File Output Format Counters

Bytes Written=1599018

23/01/04 15:22:41 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 15:22:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:22:41 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0402

23/01/04 15:22:41 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:22:41 INFO JobSubmitter: number of splits:1

23/01/04 15:22:41 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0402

23/01/04 15:22:41 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:22:41 INFO YarnClientImpl: Submitted application application\_1672817468245\_0402

23/01/04 15:22:41 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0402/](http://lena-master:8088/proxy/application_1672817468245_0402/)

23/01/04 15:22:41 INFO Job: Running job: job\_1672817468245\_0402

23/01/04 15:22:47 INFO Job: Job job\_1672817468245\_0402 running in uber mode : false

23/01/04 15:22:47 INFO Job: map 0% reduce 0%

23/01/04 15:22:53 INFO Job: map 100% reduce 0%

23/01/04 15:22:57 INFO Job: map 100% reduce 100%

23/01/04 15:22:57 INFO Job: Job job\_1672817468245\_0402 completed successfully

23/01/04 15:22:58 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=653598  
FILE: Number of bytes written=1839789  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5878915  
HDFS: Number of bytes written=2252701  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16830  
Total time spent by all reduces in occupied slots (ms)=9255  
Total time spent by all map tasks (ms)=3366  
Total time spent by all reduce tasks (ms)=1851  
Total vcore-milliseconds taken by all map tasks=3366  
Total vcore-milliseconds taken by all reduce tasks=1851  
Total megabyte-milliseconds taken by all map tasks=17233920  
Total megabyte-milliseconds taken by all reduce tasks=9477120

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=653563  
Map output materialized bytes=653598  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=653598  
Reduce input records=6  
Reduce output records=18  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=2900  
Physical memory (bytes) snapshot=689414144  
Virtual memory (bytes) snapshot=12758913024  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=391245824  
Peak Map Virtual memory (bytes)=6373789696  
Peak Reduce Physical memory (bytes)=298168320  
Peak Reduce Virtual memory (bytes)=6389313536

#### 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=2680676

#### File Output Format Counters

Bytes Written=2252701

23/01/04 15:22:58 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 15:22:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:22:58 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0406

23/01/04 15:22:58 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:22:58 INFO JobSubmitter: number of splits:1

23/01/04 15:22:58 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0406  
23/01/04 15:22:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:22:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0406  
23/01/04 15:22:58 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0406/  
23/01/04 15:22:58 INFO Job: Running job: job\_1672817468245\_0406  
23/01/04 15:23:04 INFO Job: Job job\_1672817468245\_0406 running in uber mode : false  
23/01/04 15:23:04 INFO Job: map 0% reduce 0%  
23/01/04 15:23:10 INFO Job: map 100% reduce 0%  
23/01/04 15:23:15 INFO Job: map 100% reduce 100%  
23/01/04 15:23:16 INFO Job: Job job\_1672817468245\_0406 completed successfully  
23/01/04 15:23:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=598725  
FILE: Number of bytes written=1730043  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7186281  
HDFS: Number of bytes written=2851491  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17035  
Total time spent by all reduces in occupied slots (ms)=16600  
Total time spent by all map tasks (ms)=3407  
Total time spent by all reduce tasks (ms)=3320  
Total vcore-milliseconds taken by all map tasks=3407  
Total vcore-milliseconds taken by all reduce tasks=3320  
Total megabyte-milliseconds taken by all map tasks=17443840  
Total megabyte-milliseconds taken by all reduce tasks=16998400

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=598690  
Map output materialized bytes=598725  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=598725  
Reduce input records=6  
Reduce output records=24  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=50  
CPU time spent (ms)=3150  
Physical memory (bytes) snapshot=691544064  
Virtual memory (bytes) snapshot=12748361728  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390172672

```

Peak Map Virtual memory (bytes)=6373560320
Peak Reduce Physical memory (bytes)=301371392
Peak Reduce Virtual memory (bytes)=6378717184
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=2680676
File Output Format Counters
  Bytes Written=2851491
23/01/04 15:23:16 INFO RepresentativePointsDriver: Representative Points Iteration
3
23/01/04 15:23:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:23:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0410
23/01/04 15:23:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:23:16 INFO JobSubmitter: number of splits:1
23/01/04 15:23:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0410
23/01/04 15:23:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:23:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0410
23/01/04 15:23:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0410/
23/01/04 15:23:16 INFO Job: Running job: job_1672817468245_0410
23/01/04 15:23:22 INFO Job: Job job_1672817468245_0410 running in uber mode : false
23/01/04 15:23:22 INFO Job:  map 0% reduce 0%
23/01/04 15:23:28 INFO Job:  map 100% reduce 0%
23/01/04 15:23:33 INFO Job:  map 100% reduce 100%
23/01/04 15:23:34 INFO Job: Job job_1672817468245_0410 completed successfully
23/01/04 15:23:34 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=589199
  FILE: Number of bytes written=1710991
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=8383861
  HDFS: Number of bytes written=3440735
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17370
  Total time spent by all reduces in occupied slots (ms)=16690
  Total time spent by all map tasks (ms)=3474
  Total time spent by all reduce tasks (ms)=3338
  Total vcore-milliseconds taken by all map tasks=3474
  Total vcore-milliseconds taken by all reduce tasks=3338
  Total megabyte-milliseconds taken by all map tasks=17786880

```

Total megabyte-milliseconds taken by all reduce tasks=17090560  
Map-Reduce Framework  
Map input records=27  
Map output records=6  
Map output bytes=589164  
Map output materialized bytes=589199  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=589199  
Reduce input records=6  
Reduce output records=30  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=66  
CPU time spent (ms)=3370  
Physical memory (bytes) snapshot=684756992  
Virtual memory (bytes) snapshot=12758503424  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=385556480  
Peak Map Virtual memory (bytes)=6374076416  
Peak Reduce Physical memory (bytes)=299200512  
Peak Reduce Virtual memory (bytes)=6384427008

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=2680676

File Output Format Counters

Bytes Written=3440735

23/01/04 15:23:34 INFO RepresentativePointsDriver: Representative Points Iteration  
4  
23/01/04 15:23:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:23:34 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0414  
23/01/04 15:23:35 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:23:35 INFO JobSubmitter: number of splits:1  
23/01/04 15:23:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0414  
23/01/04 15:23:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:23:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0414  
23/01/04 15:23:35 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0414/  
23/01/04 15:23:35 INFO Job: Running job: job\_1672817468245\_0414  
23/01/04 15:23:41 INFO Job: Job job\_1672817468245\_0414 running in uber mode : false  
23/01/04 15:23:41 INFO Job: map 0% reduce 0%  
23/01/04 15:23:47 INFO Job: map 100% reduce 0%  
23/01/04 15:23:53 INFO Job: map 100% reduce 100%  
23/01/04 15:23:53 INFO Job: Job job\_1672817468245\_0414 completed successfully  
23/01/04 15:23:53 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=606231  
FILE: Number of bytes written=1745055  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9562349  
HDFS: Number of bytes written=4047011  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17100  
Total time spent by all reduces in occupied slots (ms)=16865  
Total time spent by all map tasks (ms)=3420  
Total time spent by all reduce tasks (ms)=3373  
Total vcore-milliseconds taken by all map tasks=3420  
Total vcore-milliseconds taken by all reduce tasks=3373  
Total megabyte-milliseconds taken by all map tasks=17510400  
Total megabyte-milliseconds taken by all reduce tasks=17269760

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=606196  
Map output materialized bytes=606231  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=606231  
Reduce input records=6  
Reduce output records=36  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=3300  
Physical memory (bytes) snapshot=717307904  
Virtual memory (bytes) snapshot=12751450112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419160064  
Peak Map Virtual memory (bytes)=6376091648  
Peak Reduce Physical memory (bytes)=298147840  
Peak Reduce Virtual memory (bytes)=6375358464

#### 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=2680676

#### File Output Format Counters

```

Bytes Written=4047011
23/01/04 15:23:53 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.5685677893382413
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[6] =
0.6116744625863928
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[4] = NaN
23/01/04 15:23:53 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.6709502844786016
23/01/04 15:23:53 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6413123735324973
23/01/04 15:23:54 INFO ClusterDumper: Wrote 6 clusters
23/01/04 15:23:54 INFO MahoutDriver: Program took 93515 ms (Minutes:
1.5585833333333334)
---- K: 07 --
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/07
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:23:58 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[7],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/07], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:23:59 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/07
23/01/04 15:23:59 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:23:59 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:23:59 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:23:59 INFO RandomSeedGenerator: Wrote 7 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:23:59 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/07
23/01/04 15:23:59 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:23:59 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:24:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:24:00 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0420
23/01/04 15:24:00 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:24:00 INFO JobSubmitter: number of splits:1
23/01/04 15:24:00 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0420
23/01/04 15:24:00 INFO JobSubmitter: Executing with tokens: []

```

23/01/04 15:24:00 INFO Configuration: resource-types.xml not found  
23/01/04 15:24:00 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 15:24:00 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0420  
23/01/04 15:24:00 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0420/  
23/01/04 15:24:00 INFO Job: Running job: job\_1672817468245\_0420  
23/01/04 15:24:07 INFO Job: Job job\_1672817468245\_0420 running in uber mode : false  
23/01/04 15:24:07 INFO Job: map 0% reduce 0%  
23/01/04 15:24:13 INFO Job: map 100% reduce 0%  
23/01/04 15:24:17 INFO Job: map 100% reduce 100%  
23/01/04 15:24:17 INFO Job: Job job\_1672817468245\_0420 completed successfully  
23/01/04 15:24:17 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2714316  
FILE: Number of bytes written=5960923  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6243969  
HDFS: Number of bytes written=1692940  
HDFS: Number of read operations=43  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18250  
Total time spent by all reduces in occupied slots (ms)=10545  
Total time spent by all map tasks (ms)=3650  
Total time spent by all reduce tasks (ms)=2109  
Total vcore-milliseconds taken by all map tasks=3650  
Total vcore-milliseconds taken by all reduce tasks=2109  
Total megabyte-milliseconds taken by all map tasks=18688000  
Total megabyte-milliseconds taken by all reduce tasks=10798080

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=2714275  
Map output materialized bytes=2714316  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=2714316  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=4890  
Physical memory (bytes) snapshot=780054528  
Virtual memory (bytes) snapshot=12760027136  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=424505344



```

Peak Map Virtual memory (bytes)=6372372480
Peak Reduce Physical memory (bytes)=355549184
Peak Reduce Virtual memory (bytes)=6387654656
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=2680027
File Output Format Counters
  Bytes Written=1692940
23/01/04 15:24:17 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:24:17 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0423
23/01/04 15:24:17 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:24:17 INFO JobSubmitter: number of splits:1
23/01/04 15:24:17 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0423
23/01/04 15:24:17 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:24:17 INFO YarnClientImpl: Submitted application
application_1672817468245_0423
23/01/04 15:24:17 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0423/
23/01/04 15:24:17 INFO Job: Running job: job_1672817468245_0423
23/01/04 15:24:23 INFO Job: Job job_1672817468245_0423 running in uber mode : false
23/01/04 15:24:23 INFO Job:  map 0% reduce 0%
23/01/04 15:24:29 INFO Job:  map 100% reduce 0%
23/01/04 15:24:35 INFO Job:  map 100% reduce 100%
23/01/04 15:24:35 INFO Job: Job job_1672817468245_0423 completed successfully
23/01/04 15:24:35 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=3776924
  FILE: Number of bytes written=8086139
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6066465
  HDFS: Number of bytes written=1676885
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18055
  Total time spent by all reduces in occupied slots (ms)=17350
  Total time spent by all map tasks (ms)=3611
  Total time spent by all reduce tasks (ms)=3470
  Total vcore-milliseconds taken by all map tasks=3611
  Total vcore-milliseconds taken by all reduce tasks=3470
  Total megabyte-milliseconds taken by all map tasks=18488320
  Total megabyte-milliseconds taken by all reduce tasks=17766400
Map-Reduce Framework

```

Map input records=27  
Map output records=7  
Map output bytes=3776883  
Map output materialized bytes=3776924  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3776924  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=72  
CPU time spent (ms)=3980  
Physical memory (bytes) snapshot=731344896  
Virtual memory (bytes) snapshot=12754948096  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411717632  
Peak Map Virtual memory (bytes)=6372712448  
Peak Reduce Physical memory (bytes)=319627264  
Peak Reduce Virtual memory (bytes)=6382235648

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=2680027

File Output Format Counters

Bytes Written=1676885

23/01/04 15:24:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:24:35 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0426  
23/01/04 15:24:36 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:24:36 INFO JobSubmitter: number of splits:1  
23/01/04 15:24:36 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0426  
23/01/04 15:24:36 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:24:36 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0426  
23/01/04 15:24:36 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0426/  
23/01/04 15:24:36 INFO Job: Running job: job\_1672817468245\_0426  
23/01/04 15:24:42 INFO Job: Job job\_1672817468245\_0426 running in uber mode : false  
23/01/04 15:24:42 INFO Job: map 0% reduce 0%  
23/01/04 15:24:48 INFO Job: map 100% reduce 0%  
23/01/04 15:24:53 INFO Job: map 100% reduce 100%  
23/01/04 15:24:53 INFO Job: Job job\_1672817468245\_0426 completed successfully  
23/01/04 15:24:53 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3729977  
FILE: Number of bytes written=7992245  
FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6034355  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17865  
Total time spent by all reduces in occupied slots (ms)=17020  
Total time spent by all map tasks (ms)=3573  
Total time spent by all reduce tasks (ms)=3404  
Total vcore-milliseconds taken by all map tasks=3573  
Total vcore-milliseconds taken by all reduce tasks=3404  
Total megabyte-milliseconds taken by all map tasks=18293760  
Total megabyte-milliseconds taken by all reduce tasks=17428480

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3729936  
Map output materialized bytes=3729977  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3729977  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=770359296  
Virtual memory (bytes) snapshot=12757008384  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=431579136  
Peak Map Virtual memory (bytes)=6378266624  
Peak Reduce Physical memory (bytes)=338780160  
Peak Reduce Virtual memory (bytes)=6378741760

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:24:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:24:53 INFO JobResourceUploader: Disabling Erasure Coding for path:

/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0428  
23/01/04 15:24:53 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:24:53 INFO JobSubmitter: number of splits:1  
23/01/04 15:24:53 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0428  
23/01/04 15:24:53 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:24:53 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0428  
23/01/04 15:24:53 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0428/  
23/01/04 15:24:53 INFO Job: Running job: job\_1672817468245\_0428  
23/01/04 15:24:59 INFO Job: Job job\_1672817468245\_0428 running in uber mode : false  
23/01/04 15:24:59 INFO Job: map 0% reduce 0%  
23/01/04 15:25:05 INFO Job: map 100% reduce 0%  
23/01/04 15:25:11 INFO Job: map 100% reduce 100%  
23/01/04 15:25:11 INFO Job: Job job\_1672817468245\_0428 completed successfully  
23/01/04 15:25:11 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18065  
Total time spent by all reduces in occupied slots (ms)=17600  
Total time spent by all map tasks (ms)=3613  
Total time spent by all reduce tasks (ms)=3520  
Total vcore-milliseconds taken by all map tasks=3613  
Total vcore-milliseconds taken by all reduce tasks=3520  
Total megabyte-milliseconds taken by all map tasks=18498560  
Total megabyte-milliseconds taken by all reduce tasks=18022400

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=4160

Physical memory (bytes) snapshot=741052416  
Virtual memory (bytes) snapshot=12763103232  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409088000  
Peak Map Virtual memory (bytes)=6381518848  
Peak Reduce Physical memory (bytes)=331964416  
Peak Reduce Virtual memory (bytes)=6381584384

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=2680027

File Output Format Counters

Bytes Written=1662184

23/01/04 15:25:11 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:25:11 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0431  
23/01/04 15:25:11 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:25:11 INFO JobSubmitter: number of splits:1  
23/01/04 15:25:11 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0431  
23/01/04 15:25:11 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:25:11 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0431  
23/01/04 15:25:11 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0431/  
23/01/04 15:25:11 INFO Job: Running job: job\_1672817468245\_0431  
23/01/04 15:25:17 INFO Job: Job job\_1672817468245\_0431 running in uber mode : false  
23/01/04 15:25:17 INFO Job: map 0% reduce 0%  
23/01/04 15:25:23 INFO Job: map 100% reduce 0%  
23/01/04 15:25:28 INFO Job: map 100% reduce 100%  
23/01/04 15:25:28 INFO Job: Job job\_1672817468245\_0431 completed successfully  
23/01/04 15:25:28 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18160  
Total time spent by all reduces in occupied slots (ms)=17090  
Total time spent by all map tasks (ms)=3632  
Total time spent by all reduce tasks (ms)=3418  
Total vcore-milliseconds taken by all map tasks=3632

Total vcore-milliseconds taken by all reduce tasks=3418  
Total megabyte-milliseconds taken by all map tasks=18595840  
Total megabyte-milliseconds taken by all reduce tasks=17500160  
Map-Reduce Framework  
Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=4090  
Physical memory (bytes) snapshot=728006656  
Virtual memory (bytes) snapshot=12753489920  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404447232  
Peak Map Virtual memory (bytes)=6371987456  
Peak Reduce Physical memory (bytes)=323559424  
Peak Reduce Virtual memory (bytes)=6381502464

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=2680027

File Output Format Counters

Bytes Written=1662184

23/01/04 15:25:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:25:28 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0432  
23/01/04 15:25:29 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:25:29 INFO JobSubmitter: number of splits:1  
23/01/04 15:25:29 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0432  
23/01/04 15:25:29 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:25:29 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0432  
23/01/04 15:25:29 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0432/  
23/01/04 15:25:29 INFO Job: Running job: job\_1672817468245\_0432  
23/01/04 15:25:35 INFO Job: Job job\_1672817468245\_0432 running in uber mode : false  
23/01/04 15:25:35 INFO Job: map 0% reduce 0%  
23/01/04 15:25:40 INFO Job: map 100% reduce 0%  
23/01/04 15:25:46 INFO Job: map 100% reduce 100%  
23/01/04 15:25:46 INFO Job: Job job\_1672817468245\_0432 completed successfully  
23/01/04 15:25:46 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17985  
Total time spent by all reduces in occupied slots (ms)=17275  
Total time spent by all map tasks (ms)=3597  
Total time spent by all reduce tasks (ms)=3455  
Total vcore-milliseconds taken by all map tasks=3597  
Total vcore-milliseconds taken by all reduce tasks=3455  
Total megabyte-milliseconds taken by all map tasks=18416640  
Total megabyte-milliseconds taken by all reduce tasks=17689600

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=70  
CPU time spent (ms)=4130  
Physical memory (bytes) snapshot=724750336  
Virtual memory (bytes) snapshot=12758171648  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409624576  
Peak Map Virtual memory (bytes)=6378938368  
Peak Reduce Physical memory (bytes)=315125760  
Peak Reduce Virtual memory (bytes)=6379233280

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184  
23/01/04 15:25:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:25:46 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0434  
23/01/04 15:25:46 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:25:46 INFO JobSubmitter: number of splits:1  
23/01/04 15:25:46 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0434  
23/01/04 15:25:46 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:25:47 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0434  
23/01/04 15:25:47 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0434/  
23/01/04 15:25:47 INFO Job: Running job: job\_1672817468245\_0434  
23/01/04 15:25:53 INFO Job: Job job\_1672817468245\_0434 running in uber mode : false  
23/01/04 15:25:53 INFO Job: map 0% reduce 0%  
23/01/04 15:25:59 INFO Job: map 100% reduce 0%  
23/01/04 15:26:05 INFO Job: map 100% reduce 100%  
23/01/04 15:26:05 INFO Job: Job job\_1672817468245\_0434 completed successfully  
23/01/04 15:26:05 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18105  
Total time spent by all reduces in occupied slots (ms)=17395  
Total time spent by all map tasks (ms)=3621  
Total time spent by all reduce tasks (ms)=3479  
Total vcore-milliseconds taken by all map tasks=3621  
Total vcore-milliseconds taken by all reduce tasks=3479  
Total megabyte-milliseconds taken by all map tasks=18539520  
Total megabyte-milliseconds taken by all reduce tasks=17812480

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1



Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=77  
CPU time spent (ms)=4220  
Physical memory (bytes) snapshot=758726656  
Virtual memory (bytes) snapshot=12766019584  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410365952  
Peak Map Virtual memory (bytes)=6379642880  
Peak Reduce Physical memory (bytes)=348360704  
Peak Reduce Virtual memory (bytes)=6386376704  
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=2680027  
File Output Format Counters  
Bytes Written=1662184  
23/01/04 15:26:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:26:05 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0436  
23/01/04 15:26:05 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:26:05 INFO JobSubmitter: number of splits:1  
23/01/04 15:26:05 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0436  
23/01/04 15:26:05 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:26:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0436  
23/01/04 15:26:05 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0436/  
23/01/04 15:26:05 INFO Job: Running job: job\_1672817468245\_0436  
23/01/04 15:26:11 INFO Job: Job job\_1672817468245\_0436 running in uber mode : false  
23/01/04 15:26:11 INFO Job: map 0% reduce 0%  
23/01/04 15:26:17 INFO Job: map 100% reduce 0%  
23/01/04 15:26:23 INFO Job: map 100% reduce 100%  
23/01/04 15:26:23 INFO Job: Job job\_1672817468245\_0436 completed successfully  
23/01/04 15:26:23 INFO Job: Counters: 54  
File System Counters  
FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0  
Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18285

Total time spent by all reduces in occupied slots (ms)=17435  
Total time spent by all map tasks (ms)=3657  
Total time spent by all reduce tasks (ms)=3487  
Total vcore-milliseconds taken by all map tasks=3657  
Total vcore-milliseconds taken by all reduce tasks=3487  
Total megabyte-milliseconds taken by all map tasks=18723840  
Total megabyte-milliseconds taken by all reduce tasks=17853440

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=80  
CPU time spent (ms)=4300  
Physical memory (bytes) snapshot=738177024  
Virtual memory (bytes) snapshot=12761051136  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=417386496  
Peak Map Virtual memory (bytes)=6376452096  
Peak Reduce Physical memory (bytes)=320790528  
Peak Reduce Virtual memory (bytes)=6384599040

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:26:23 INFO DefaultNoHARMAiloverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:26:23 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0438  
23/01/04 15:26:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:26:24 INFO JobSubmitter: number of splits:1  
23/01/04 15:26:24 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0438  
23/01/04 15:26:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:26:24 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0438  
23/01/04 15:26:24 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0438/  
23/01/04 15:26:24 INFO Job: Running job: job\_1672817468245\_0438  
23/01/04 15:26:30 INFO Job: Job job\_1672817468245\_0438 running in uber mode : false  
23/01/04 15:26:30 INFO Job: map 0% reduce 0%

23/01/04 15:26:36 INFO Job: map 100% reduce 0%  
23/01/04 15:26:40 INFO Job: map 100% reduce 100%  
23/01/04 15:26:40 INFO Job: Job job\_1672817468245\_0438 completed successfully  
23/01/04 15:26:40 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18395  
Total time spent by all reduces in occupied slots (ms)=10075  
Total time spent by all map tasks (ms)=3679  
Total time spent by all reduce tasks (ms)=2015  
Total vcore-milliseconds taken by all map tasks=3679  
Total vcore-milliseconds taken by all reduce tasks=2015  
Total megabyte-milliseconds taken by all map tasks=18836480  
Total megabyte-milliseconds taken by all reduce tasks=10316800

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=91  
CPU time spent (ms)=4130  
Physical memory (bytes) snapshot=753807360  
Virtual memory (bytes) snapshot=12752879616  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418967552  
Peak Map Virtual memory (bytes)=6367924224  
Peak Reduce Physical memory (bytes)=334839808  
Peak Reduce Virtual memory (bytes)=6384955392

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1662184  
23/01/04 15:26:40 INFO DefaultNoHARMAiloverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:26:40 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0440  
23/01/04 15:26:40 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:26:40 INFO JobSubmitter: number of splits:1  
23/01/04 15:26:40 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0440  
23/01/04 15:26:40 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:26:40 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0440  
23/01/04 15:26:40 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0440/  
23/01/04 15:26:40 INFO Job: Running job: job\_1672817468245\_0440  
23/01/04 15:26:46 INFO Job: Job job\_1672817468245\_0440 running in uber mode : false  
23/01/04 15:26:46 INFO Job: map 0% reduce 0%  
23/01/04 15:26:52 INFO Job: map 100% reduce 0%  
23/01/04 15:26:57 INFO Job: map 100% reduce 100%  
23/01/04 15:26:57 INFO Job: Job job\_1672817468245\_0440 completed successfully  
23/01/04 15:26:57 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=3715276  
  FILE: Number of bytes written=7962847  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=6004953  
  HDFS: Number of bytes written=1662184  
  HDFS: Number of read operations=19  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=18170  
  Total time spent by all reduces in occupied slots (ms)=17430  
  Total time spent by all map tasks (ms)=3634  
  Total time spent by all reduce tasks (ms)=3486  
  Total vcore-milliseconds taken by all map tasks=3634  
  Total vcore-milliseconds taken by all reduce tasks=3486  
  Total megabyte-milliseconds taken by all map tasks=18606080  
  Total megabyte-milliseconds taken by all reduce tasks=17848320  
Map-Reduce Framework  
  Map input records=27  
  Map output records=7  
  Map output bytes=3715235  
  Map output materialized bytes=3715276  
  Input split bytes=170  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=7  
  Reduce shuffle bytes=3715276

Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=4270  
Physical memory (bytes) snapshot=778760192  
Virtual memory (bytes) snapshot=12764217344  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=431828992  
Peak Map Virtual memory (bytes)=6374481920  
Peak Reduce Physical memory (bytes)=346931200  
Peak Reduce Virtual memory (bytes)=6389735424

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=2680027

File Output Format Counters

Bytes Written=1662184

23/01/04 15:26:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:26:58 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0442  
23/01/04 15:26:58 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:26:58 INFO JobSubmitter: number of splits:1  
23/01/04 15:26:58 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0442  
23/01/04 15:26:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:26:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0442  
23/01/04 15:26:58 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0442/](http://lena-master:8088/proxy/application_1672817468245_0442/)  
23/01/04 15:26:58 INFO Job: Running job: job\_1672817468245\_0442  
23/01/04 15:27:04 INFO Job: Job job\_1672817468245\_0442 running in uber mode : false  
23/01/04 15:27:04 INFO Job: map 0% reduce 0%  
23/01/04 15:27:10 INFO Job: map 100% reduce 0%  
23/01/04 15:27:14 INFO Job: map 100% reduce 100%  
23/01/04 15:27:14 INFO Job: Job job\_1672817468245\_0442 completed successfully  
23/01/04 15:27:14 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

```

    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=18050
    Total time spent by all reduces in occupied slots (ms)=10255
    Total time spent by all map tasks (ms)=3610
    Total time spent by all reduce tasks (ms)=2051
    Total vcore-milliseconds taken by all map tasks=3610
    Total vcore-milliseconds taken by all reduce tasks=2051
    Total megabyte-milliseconds taken by all map tasks=18483200
    Total megabyte-milliseconds taken by all reduce tasks=10501120
Map-Reduce Framework
    Map input records=27
    Map output records=7
    Map output bytes=3715235
    Map output materialized bytes=3715276
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=7
    Reduce shuffle bytes=3715276
    Reduce input records=7
    Reduce output records=7
    Spilled Records=14
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=66
    CPU time spent (ms)=4090
    Physical memory (bytes) snapshot=737824768
    Virtual memory (bytes) snapshot=12770291712
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=403480576
    Peak Map Virtual memory (bytes)=6377361408
    Peak Reduce Physical memory (bytes)=334344192
    Peak Reduce Virtual memory (bytes)=6392930304
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=2680027
File Output Format Counters
    Bytes Written=1662184
23/01/04 15:27:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:27:14 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0443
23/01/04 15:27:14 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:27:14 INFO JobSubmitter: number of splits:1
23/01/04 15:27:14 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0443
23/01/04 15:27:14 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:27:14 INFO YarnClientImpl: Submitted application
application_1672817468245_0443
23/01/04 15:27:14 INFO Job: The url to track the job:

```

http://lena-master:8088/proxy/application\_1672817468245\_0443/  
23/01/04 15:27:14 INFO Job: Running job: job\_1672817468245\_0443  
23/01/04 15:27:20 INFO Job: Job job\_1672817468245\_0443 running in uber mode : false  
23/01/04 15:27:20 INFO Job: map 0% reduce 0%  
23/01/04 15:27:26 INFO Job: map 100% reduce 0%  
23/01/04 15:27:30 INFO Job: map 100% reduce 100%  
23/01/04 15:27:30 INFO Job: Job job\_1672817468245\_0443 completed successfully  
23/01/04 15:27:30 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18245  
Total time spent by all reduces in occupied slots (ms)=10525  
Total time spent by all map tasks (ms)=3649  
Total time spent by all reduce tasks (ms)=2105  
Total vcore-milliseconds taken by all map tasks=3649  
Total vcore-milliseconds taken by all reduce tasks=2105  
Total megabyte-milliseconds taken by all map tasks=18682880  
Total megabyte-milliseconds taken by all reduce tasks=10777600

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=55  
CPU time spent (ms)=4450  
Physical memory (bytes) snapshot=742416384  
Virtual memory (bytes) snapshot=12762812416  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398888960  
Peak Map Virtual memory (bytes)=6383472640  
Peak Reduce Physical memory (bytes)=343527424  
Peak Reduce Virtual memory (bytes)=6379339776

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1662184  
23/01/04 15:27:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:27:31 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0445  
23/01/04 15:27:31 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:27:31 INFO JobSubmitter: number of splits:1  
23/01/04 15:27:31 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0445  
23/01/04 15:27:31 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:27:31 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0445  
23/01/04 15:27:31 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0445/  
23/01/04 15:27:31 INFO Job: Running job: job\_1672817468245\_0445  
23/01/04 15:27:37 INFO Job: Job job\_1672817468245\_0445 running in uber mode : false  
23/01/04 15:27:37 INFO Job: map 0% reduce 0%  
23/01/04 15:27:43 INFO Job: map 100% reduce 0%  
23/01/04 15:27:49 INFO Job: map 100% reduce 100%  
23/01/04 15:27:49 INFO Job: Job job\_1672817468245\_0445 completed successfully  
23/01/04 15:27:49 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=3715276  
  FILE: Number of bytes written=7962851  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=6004953  
  HDFS: Number of bytes written=1662184  
  HDFS: Number of read operations=19  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Data-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17785  
  Total time spent by all reduces in occupied slots (ms)=17260  
  Total time spent by all map tasks (ms)=3557  
  Total time spent by all reduce tasks (ms)=3452  
  Total vcore-milliseconds taken by all map tasks=3557  
  Total vcore-milliseconds taken by all reduce tasks=3452  
  Total megabyte-milliseconds taken by all map tasks=18211840  
  Total megabyte-milliseconds taken by all reduce tasks=17674240  
Map-Reduce Framework  
  Map input records=27  
  Map output records=7  
  Map output bytes=3715235  
  Map output materialized bytes=3715276  
  Input split bytes=170



Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=4010  
Physical memory (bytes) snapshot=737636352  
Virtual memory (bytes) snapshot=12763623424  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412254208  
Peak Map Virtual memory (bytes)=6375874560  
Peak Reduce Physical memory (bytes)=325382144  
Peak Reduce Virtual memory (bytes)=6387748864

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:27:49 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:27:49 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0449  
23/01/04 15:27:49 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:27:49 INFO JobSubmitter: number of splits:1  
23/01/04 15:27:49 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0449  
23/01/04 15:27:49 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:27:49 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0449  
23/01/04 15:27:49 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0449/  
23/01/04 15:27:49 INFO Job: Running job: job\_1672817468245\_0449  
23/01/04 15:27:55 INFO Job: Job job\_1672817468245\_0449 running in uber mode : false  
23/01/04 15:27:55 INFO Job: map 0% reduce 0%  
23/01/04 15:28:01 INFO Job: map 100% reduce 0%  
23/01/04 15:28:06 INFO Job: map 100% reduce 100%  
23/01/04 15:28:07 INFO Job: Job job\_1672817468245\_0449 completed successfully  
23/01/04 15:28:07 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19

HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18210  
Total time spent by all reduces in occupied slots (ms)=17440  
Total time spent by all map tasks (ms)=3642  
Total time spent by all reduce tasks (ms)=3488  
Total vcore-milliseconds taken by all map tasks=3642  
Total vcore-milliseconds taken by all reduce tasks=3488  
Total megabyte-milliseconds taken by all map tasks=18647040  
Total megabyte-milliseconds taken by all reduce tasks=17858560

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3910  
Physical memory (bytes) snapshot=743424000  
Virtual memory (bytes) snapshot=12768190464  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412971008  
Peak Map Virtual memory (bytes)=6379450368  
Peak Reduce Physical memory (bytes)=330452992  
Peak Reduce Virtual memory (bytes)=6388740096

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:28:08 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:28:08 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0452

23/01/04 15:28:08 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:28:08 INFO JobSubmitter: number of splits:1

23/01/04 15:28:08 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0452

23/01/04 15:28:08 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:28:08 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0452  
23/01/04 15:28:08 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0452/  
23/01/04 15:28:08 INFO Job: Running job: job\_1672817468245\_0452  
23/01/04 15:28:14 INFO Job: Job job\_1672817468245\_0452 running in uber mode : false  
23/01/04 15:28:14 INFO Job: map 0% reduce 0%  
23/01/04 15:28:20 INFO Job: map 100% reduce 0%  
23/01/04 15:28:25 INFO Job: map 100% reduce 100%  
23/01/04 15:28:25 INFO Job: Job job\_1672817468245\_0452 completed successfully  
23/01/04 15:28:25 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17670  
Total time spent by all reduces in occupied slots (ms)=17020  
Total time spent by all map tasks (ms)=3534  
Total time spent by all reduce tasks (ms)=3404  
Total vcore-milliseconds taken by all map tasks=3534  
Total vcore-milliseconds taken by all reduce tasks=3404  
Total megabyte-milliseconds taken by all map tasks=18094080  
Total megabyte-milliseconds taken by all reduce tasks=17428480

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=4030  
Physical memory (bytes) snapshot=733712384  
Virtual memory (bytes) snapshot=12755496960  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400760832  
Peak Map Virtual memory (bytes)=6377123840

```

        Peak Reduce Physical memory (bytes)=332951552
        Peak Reduce Virtual memory (bytes)=6378373120
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=2680027
File Output Format Counters
    Bytes Written=1662184
23/01/04 15:28:25 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:28:25 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0455
23/01/04 15:28:25 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:28:25 INFO JobSubmitter: number of splits:1
23/01/04 15:28:25 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0455
23/01/04 15:28:25 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:28:26 INFO YarnClientImpl: Submitted application
application_1672817468245_0455
23/01/04 15:28:26 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0455/
23/01/04 15:28:26 INFO Job: Running job: job_1672817468245_0455
23/01/04 15:28:32 INFO Job: Job job_1672817468245_0455 running in uber mode : false
23/01/04 15:28:32 INFO Job:  map 0% reduce 0%
23/01/04 15:28:37 INFO Job:  map 100% reduce 0%
23/01/04 15:28:43 INFO Job:  map 100% reduce 100%
23/01/04 15:28:43 INFO Job: Job job_1672817468245_0455 completed successfully
23/01/04 15:28:43 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=3715276
    FILE: Number of bytes written=7962851
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6004953
    HDFS: Number of bytes written=1662184
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=18345
    Total time spent by all reduces in occupied slots (ms)=17795
    Total time spent by all map tasks (ms)=3669
    Total time spent by all reduce tasks (ms)=3559
    Total vcore-milliseconds taken by all map tasks=3669
    Total vcore-milliseconds taken by all reduce tasks=3559
    Total megabyte-milliseconds taken by all map tasks=18785280
    Total megabyte-milliseconds taken by all reduce tasks=18222080
Map-Reduce Framework
    Map input records=27

```

Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=3910  
Physical memory (bytes) snapshot=735588352  
Virtual memory (bytes) snapshot=12763090944  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402808832  
Peak Map Virtual memory (bytes)=6376005632  
Peak Reduce Physical memory (bytes)=332779520  
Peak Reduce Virtual memory (bytes)=6387085312

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=2680027

File Output Format Counters

Bytes Written=1662184

23/01/04 15:28:43 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:28:43 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0459  
23/01/04 15:28:43 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:28:43 INFO JobSubmitter: number of splits:1  
23/01/04 15:28:43 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0459  
23/01/04 15:28:43 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:28:43 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0459  
23/01/04 15:28:43 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0459/](http://lena-master:8088/proxy/application_1672817468245_0459/)  
23/01/04 15:28:43 INFO Job: Running job: job\_1672817468245\_0459  
23/01/04 15:28:49 INFO Job: Job job\_1672817468245\_0459 running in uber mode : false  
23/01/04 15:28:49 INFO Job: map 0% reduce 0%  
23/01/04 15:28:55 INFO Job: map 100% reduce 0%  
23/01/04 15:28:59 INFO Job: map 100% reduce 100%  
23/01/04 15:29:00 INFO Job: Job job\_1672817468245\_0459 completed successfully  
23/01/04 15:29:00 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18110  
Total time spent by all reduces in occupied slots (ms)=10160  
Total time spent by all map tasks (ms)=3622  
Total time spent by all reduce tasks (ms)=2032  
Total vcore-milliseconds taken by all map tasks=3622  
Total vcore-milliseconds taken by all reduce tasks=2032  
Total megabyte-milliseconds taken by all map tasks=18544640  
Total megabyte-milliseconds taken by all reduce tasks=10403840

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=4160  
Physical memory (bytes) snapshot=746491904  
Virtual memory (bytes) snapshot=12763545600  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412016640  
Peak Map Virtual memory (bytes)=6372610048  
Peak Reduce Physical memory (bytes)=334475264  
Peak Reduce Virtual memory (bytes)=6390935552

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:29:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:29:00 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0462

23/01/04 15:29:00 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:29:00 INFO JobSubmitter: number of splits:1  
23/01/04 15:29:01 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0462  
23/01/04 15:29:01 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:29:01 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0462  
23/01/04 15:29:01 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0462/](http://lena-master:8088/proxy/application_1672817468245_0462/)  
23/01/04 15:29:01 INFO Job: Running job: job\_1672817468245\_0462  
23/01/04 15:29:07 INFO Job: Job job\_1672817468245\_0462 running in uber mode : false  
23/01/04 15:29:07 INFO Job: map 0% reduce 0%  
23/01/04 15:29:13 INFO Job: map 100% reduce 0%  
23/01/04 15:29:18 INFO Job: map 100% reduce 100%  
23/01/04 15:29:19 INFO Job: Job job\_1672817468245\_0462 completed successfully  
23/01/04 15:29:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18100  
Total time spent by all reduces in occupied slots (ms)=17285  
Total time spent by all map tasks (ms)=3620  
Total time spent by all reduce tasks (ms)=3457  
Total vcore-milliseconds taken by all map tasks=3620  
Total vcore-milliseconds taken by all reduce tasks=3457  
Total megabyte-milliseconds taken by all map tasks=18534400  
Total megabyte-milliseconds taken by all reduce tasks=17699840

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=4210  
Physical memory (bytes) snapshot=755613696

Virtual memory (bytes) snapshot=12762521600  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=429355008  
Peak Map Virtual memory (bytes)=6370971648  
Peak Reduce Physical memory (bytes)=326258688  
Peak Reduce Virtual memory (bytes)=6391549952

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=2680027

File Output Format Counters

Bytes Written=1662184

23/01/04 15:29:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:29:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0465  
23/01/04 15:29:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:29:19 INFO JobSubmitter: number of splits:1  
23/01/04 15:29:19 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0465  
23/01/04 15:29:19 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:29:19 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0465  
23/01/04 15:29:19 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0465/  
23/01/04 15:29:19 INFO Job: Running job: job\_1672817468245\_0465  
23/01/04 15:29:25 INFO Job: Job job\_1672817468245\_0465 running in uber mode : false  
23/01/04 15:29:25 INFO Job: map 0% reduce 0%  
23/01/04 15:29:31 INFO Job: map 100% reduce 0%  
23/01/04 15:29:36 INFO Job: map 100% reduce 100%  
23/01/04 15:29:37 INFO Job: Job job\_1672817468245\_0465 completed successfully  
23/01/04 15:29:37 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18090  
Total time spent by all reduces in occupied slots (ms)=17130  
Total time spent by all map tasks (ms)=3618  
Total time spent by all reduce tasks (ms)=3426  
Total vcore-milliseconds taken by all map tasks=3618  
Total vcore-milliseconds taken by all reduce tasks=3426



```

        Total megabyte-milliseconds taken by all map tasks=18524160
        Total megabyte-milliseconds taken by all reduce tasks=17541120
Map-Reduce Framework
    Map input records=27
    Map output records=7
    Map output bytes=3715235
    Map output materialized bytes=3715276
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=7
    Reduce shuffle bytes=3715276
    Reduce input records=7
    Reduce output records=7
    Spilled Records=14
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=52
    CPU time spent (ms)=4140
    Physical memory (bytes) snapshot=780480512
    Virtual memory (bytes) snapshot=12761141248
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=423116800
    Peak Map Virtual memory (bytes)=6373724160
    Peak Reduce Physical memory (bytes)=357363712
    Peak Reduce Virtual memory (bytes)=6387417088
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=2680027
File Output Format Counters
    Bytes Written=1662184
23/01/04 15:29:37 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:29:37 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0471
23/01/04 15:29:37 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:29:38 INFO JobSubmitter: number of splits:1
23/01/04 15:29:38 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0471
23/01/04 15:29:38 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:29:38 INFO YarnClientImpl: Submitted application
application_1672817468245_0471
23/01/04 15:29:38 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0471/
23/01/04 15:29:38 INFO Job: Running job: job_1672817468245_0471
23/01/04 15:29:44 INFO Job: Job job_1672817468245_0471 running in uber mode : false
23/01/04 15:29:44 INFO Job:  map 0% reduce 0%
23/01/04 15:29:50 INFO Job:  map 100% reduce 0%
23/01/04 15:29:56 INFO Job:  map 100% reduce 100%
23/01/04 15:29:56 INFO Job: Job job_1672817468245_0471 completed successfully
23/01/04 15:29:56 INFO Job: Counters: 54
File System Counters

```

FILE: Number of bytes read=3715276  
FILE: Number of bytes written=7962851  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6004953  
HDFS: Number of bytes written=1662184  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18155  
Total time spent by all reduces in occupied slots (ms)=17575  
Total time spent by all map tasks (ms)=3631  
Total time spent by all reduce tasks (ms)=3515  
Total vcore-milliseconds taken by all map tasks=3631  
Total vcore-milliseconds taken by all reduce tasks=3515  
Total megabyte-milliseconds taken by all map tasks=18590720  
Total megabyte-milliseconds taken by all reduce tasks=17996800

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=3715235  
Map output materialized bytes=3715276  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=3715276  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=69  
CPU time spent (ms)=4050  
Physical memory (bytes) snapshot=748519424  
Virtual memory (bytes) snapshot=12760825856  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412246016  
Peak Map Virtual memory (bytes)=6369464320  
Peak Reduce Physical memory (bytes)=336273408  
Peak Reduce Virtual memory (bytes)=6391361536

#### 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=2680027

#### File Output Format Counters

Bytes Written=1662184

23/01/04 15:29:56 INFO KMeansDriver: Clustering data  
23/01/04 15:29:56 INFO KMeansDriver: Running Clustering  
23/01/04 15:29:56 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/07 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/07  
23/01/04 15:29:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:29:56 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0475  
23/01/04 15:29:56 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:29:56 INFO JobSubmitter: number of splits:1  
23/01/04 15:29:56 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0475  
23/01/04 15:29:56 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:29:56 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0475  
23/01/04 15:29:56 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0475/  
23/01/04 15:29:56 INFO Job: Running job: job\_1672817468245\_0475  
23/01/04 15:30:02 INFO Job: Job job\_1672817468245\_0475 running in uber mode : false  
23/01/04 15:30:02 INFO Job: map 0% reduce 0%  
23/01/04 15:30:08 INFO Job: map 100% reduce 0%  
23/01/04 15:30:08 INFO Job: Job job\_1672817468245\_0475 completed successfully  
23/01/04 15:30:08 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265898  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4342575  
HDFS: Number of bytes written=2680650  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17705  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3541  
Total vcore-milliseconds taken by all map tasks=3541  
Total megabyte-milliseconds taken by all map tasks=18129920

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=54  
CPU time spent (ms)=2490  
Physical memory (bytes) snapshot=292675584  
Virtual memory (bytes) snapshot=6380445696  
Total committed heap usage (bytes)=1052770304

```

        Peak Map Physical memory (bytes)=292675584
        Peak Map Virtual memory (bytes)=6380445696
    File Input Format Counters
        Bytes Read=2680027
    File Output Format Counters
        Bytes Written=2680650
23/01/04 15:30:08 INFO MahoutDriver: Program took 370158 ms (Minutes:
6.169316666666667)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:30:14 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/EuclideanDistanceMeasure/07/clusters-20-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
EuclideanDistanceMeasure/07.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu
re/07/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:30:16 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:30:16 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/07/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
07/clusters-20-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:30:16 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:30:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:30:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0479
23/01/04 15:30:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:30:16 INFO JobSubmitter: number of splits:1
23/01/04 15:30:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0479
23/01/04 15:30:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:30:16 INFO Configuration: resource-types.xml not found
23/01/04 15:30:16 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:30:17 INFO YarnClientImpl: Submitted application
application_1672817468245_0479
23/01/04 15:30:17 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0479/
23/01/04 15:30:17 INFO Job: Running job: job_1672817468245_0479
23/01/04 15:30:23 INFO Job: Job job_1672817468245_0479 running in uber mode : false
23/01/04 15:30:23 INFO Job:  map 0% reduce 0%
23/01/04 15:30:28 INFO Job:  map 100% reduce 0%
23/01/04 15:30:34 INFO Job:  map 100% reduce 100%
23/01/04 15:30:34 INFO Job: Job job_1672817468245_0479 completed successfully
23/01/04 15:30:34 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=815914
        FILE: Number of bytes written=2164421
        FILE: Number of read operations=0

```

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4734745  
HDFS: Number of bytes written=1842939  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16690  
Total time spent by all reduces in occupied slots (ms)=16780  
Total time spent by all map tasks (ms)=3338  
Total time spent by all reduce tasks (ms)=3356  
Total vcore-milliseconds taken by all map tasks=3338  
Total vcore-milliseconds taken by all reduce tasks=3356  
Total megabyte-milliseconds taken by all map tasks=17090560  
Total megabyte-milliseconds taken by all reduce tasks=17182720

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=815873  
Map output materialized bytes=815914  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=815914  
Reduce input records=7  
Reduce output records=14  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=26  
CPU time spent (ms)=2750  
Physical memory (bytes) snapshot=690900992  
Virtual memory (bytes) snapshot=12764553216  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399011840  
Peak Map Virtual memory (bytes)=6379347968  
Peak Reduce Physical memory (bytes)=291889152  
Peak Reduce Virtual memory (bytes)=6385213440

#### 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=2680650

#### File Output Format Counters

Bytes Written=1842939

23/01/04 15:30:34 INFO RepresentativePointsDriver: Representative Points Iteration  
1  
23/01/04 15:30:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:30:34 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0481  
23/01/04 15:30:34 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:30:34 INFO JobSubmitter: number of splits:1  
23/01/04 15:30:34 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0481  
23/01/04 15:30:34 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:30:34 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0481  
23/01/04 15:30:34 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0481/  
23/01/04 15:30:34 INFO Job: Running job: job\_1672817468245\_0481  
23/01/04 15:30:40 INFO Job: Job job\_1672817468245\_0481 running in uber mode : false  
23/01/04 15:30:40 INFO Job: map 0% reduce 0%  
23/01/04 15:30:45 INFO Job: map 100% reduce 0%  
23/01/04 15:30:51 INFO Job: map 100% reduce 100%  
23/01/04 15:30:51 INFO Job: Job job\_1672817468245\_0481 completed successfully  
23/01/04 15:30:51 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=778657  
FILE: Number of bytes written=2089907  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6366731  
HDFS: Number of bytes written=2621675  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16830  
Total time spent by all reduces in occupied slots (ms)=16735  
Total time spent by all map tasks (ms)=3366  
Total time spent by all reduce tasks (ms)=3347  
Total vcore-milliseconds taken by all map tasks=3366  
Total vcore-milliseconds taken by all reduce tasks=3347  
Total megabyte-milliseconds taken by all map tasks=17233920  
Total megabyte-milliseconds taken by all reduce tasks=17136640

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=778616  
Map output materialized bytes=778657  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=778657  
Reduce input records=7  
Reduce output records=21  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=65
CPU time spent (ms)=2990
Physical memory (bytes) snapshot=690946048
Virtual memory (bytes) snapshot=12754788352
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=392560640
Peak Map Virtual memory (bytes)=6379786240
Peak Reduce Physical memory (bytes)=298385408
Peak Reduce Virtual memory (bytes)=6378909696
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=2680650
File Output Format Counters
  Bytes Written=2621675
23/01/04 15:30:51 INFO RepresentativePointsDriver: Representative Points Iteration
2
23/01/04 15:30:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:30:51 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0484
23/01/04 15:30:51 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:30:52 INFO JobSubmitter: number of splits:1
23/01/04 15:30:52 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0484
23/01/04 15:30:52 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:30:52 INFO YarnClientImpl: Submitted application
application_1672817468245_0484
23/01/04 15:30:52 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0484/
23/01/04 15:30:52 INFO Job: Running job: job_1672817468245_0484
23/01/04 15:30:58 INFO Job: Job job_1672817468245_0484 running in uber mode : false
23/01/04 15:30:58 INFO Job:  map 0% reduce 0%
23/01/04 15:31:04 INFO Job:  map 100% reduce 0%
23/01/04 15:31:10 INFO Job:  map 100% reduce 100%
23/01/04 15:31:10 INFO Job: Job job_1672817468245_0484 completed successfully
23/01/04 15:31:10 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=750806
  FILE: Number of bytes written=2034205
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=7924203
  HDFS: Number of bytes written=3372540
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17290

```

Total time spent by all reduces in occupied slots (ms)=16605  
Total time spent by all map tasks (ms)=3458  
Total time spent by all reduce tasks (ms)=3321  
Total vcore-milliseconds taken by all map tasks=3458  
Total vcore-milliseconds taken by all reduce tasks=3321  
Total megabyte-milliseconds taken by all map tasks=17704960  
Total megabyte-milliseconds taken by all reduce tasks=17003520

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=750765  
Map output materialized bytes=750806  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=750806  
Reduce input records=7  
Reduce output records=28  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=3220  
Physical memory (bytes) snapshot=687566848  
Virtual memory (bytes) snapshot=12756086784  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396742656  
Peak Map Virtual memory (bytes)=6374096896  
Peak Reduce Physical memory (bytes)=292691968  
Peak Reduce Virtual memory (bytes)=6383943680

#### 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=2680650

#### File Output Format Counters

Bytes Written=3372540

23/01/04 15:31:10 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:31:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:31:10 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0487

23/01/04 15:31:10 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:31:10 INFO JobSubmitter: number of splits:1

23/01/04 15:31:10 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0487

23/01/04 15:31:10 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:31:10 INFO YarnClientImpl: Submitted application application\_1672817468245\_0487

23/01/04 15:31:10 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0487/](http://lena-master:8088/proxy/application_1672817468245_0487/)

23/01/04 15:31:10 INFO Job: Running job: job\_1672817468245\_0487



23/01/04 15:31:17 INFO Job: Job job\_1672817468245\_0487 running in uber mode : false  
23/01/04 15:31:17 INFO Job: map 0% reduce 0%  
23/01/04 15:31:23 INFO Job: map 100% reduce 0%  
23/01/04 15:31:29 INFO Job: map 100% reduce 100%  
23/01/04 15:31:29 INFO Job: Job job\_1672817468245\_0487 completed successfully  
23/01/04 15:31:29 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=774811  
FILE: Number of bytes written=2082215  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9425933  
HDFS: Number of bytes written=4147450  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16975  
Total time spent by all reduces in occupied slots (ms)=16570  
Total time spent by all map tasks (ms)=3395  
Total time spent by all reduce tasks (ms)=3314  
Total vcore-milliseconds taken by all map tasks=3395  
Total vcore-milliseconds taken by all reduce tasks=3314  
Total megabyte-milliseconds taken by all map tasks=17382400  
Total megabyte-milliseconds taken by all reduce tasks=16967680

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=774770  
Map output materialized bytes=774811  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=774811  
Reduce input records=7  
Reduce output records=35  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3400  
Physical memory (bytes) snapshot=684040192  
Virtual memory (bytes) snapshot=12747997184  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=390529024  
Peak Map Virtual memory (bytes)=6367096832  
Peak Reduce Physical memory (bytes)=293511168  
Peak Reduce Virtual memory (bytes)=6380900352

#### 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=2680650  
File Output Format Counters  
  Bytes Written=4147450  
23/01/04 15:31:29 INFO RepresentativePointsDriver: Representative Points Iteration  
4  
23/01/04 15:31:29 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:31:30 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0489  
23/01/04 15:31:30 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:31:30 INFO JobSubmitter: number of splits:1  
23/01/04 15:31:30 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0489  
23/01/04 15:31:30 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:31:30 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0489  
23/01/04 15:31:30 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0489/  
23/01/04 15:31:30 INFO Job: Running job: job\_1672817468245\_0489  
23/01/04 15:31:36 INFO Job: Job job\_1672817468245\_0489 running in uber mode : false  
23/01/04 15:31:36 INFO Job: map 0% reduce 0%  
23/01/04 15:31:42 INFO Job: map 100% reduce 0%  
23/01/04 15:31:48 INFO Job: map 100% reduce 100%  
23/01/04 15:31:48 INFO Job: Job job\_1672817468245\_0489 completed successfully  
23/01/04 15:31:48 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=804702  
  FILE: Number of bytes written=2141997  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=10975753  
  HDFS: Number of bytes written=4952211  
  HDFS: Number of read operations=15  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Data-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17435  
  Total time spent by all reduces in occupied slots (ms)=16860  
  Total time spent by all map tasks (ms)=3487  
  Total time spent by all reduce tasks (ms)=3372  
  Total vcore-milliseconds taken by all map tasks=3487  
  Total vcore-milliseconds taken by all reduce tasks=3372  
  Total megabyte-milliseconds taken by all map tasks=17853440  
  Total megabyte-milliseconds taken by all reduce tasks=17264640  
Map-Reduce Framework  
  Map input records=27  
  Map output records=7  
  Map output bytes=804661  
  Map output materialized bytes=804702  
  Input split bytes=203

```

Combine input records=0
Combine output records=0
Reduce input groups=7
Reduce shuffle bytes=804702
Reduce input records=7
Reduce output records=42
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=51
CPU time spent (ms)=3370
Physical memory (bytes) snapshot=711634944
Virtual memory (bytes) snapshot=12753702912
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=419037184
Peak Map Virtual memory (bytes)=6374428672
Peak Reduce Physical memory (bytes)=292597760
Peak Reduce Virtual memory (bytes)=6379274240
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=2680650
File Output Format Counters
  Bytes Written=4952211
23/01/04 15:31:48 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.48850396603925395
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[0] =
0.5620277815838468
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[11] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[2] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[4] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Intra-Cluster Density[23] = NaN
23/01/04 15:31:48 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.5620277815838468
23/01/04 15:31:49 INFO ClusterDumper: Wrote 7 clusters
23/01/04 15:31:49 INFO MahoutDriver: Program took 94856 ms (Minutes:
1.5809333333333333)
---- K: 08 --
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/08
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:31:54 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[8],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-

```

```
kmeans/EuclideanDistanceMeasure/08], --overwrite=null, --startPhase=[0], --
tempDir=[temp]]
23/01/04 15:31:54 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/08
23/01/04 15:31:54 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:31:54 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:31:54 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:31:55 INFO RandomSeedGenerator: Wrote 8 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:31:55 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/08
23/01/04 15:31:55 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:31:55 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:31:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:31:55 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0491
23/01/04 15:31:55 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:31:55 INFO JobSubmitter: number of splits:1
23/01/04 15:31:56 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0491
23/01/04 15:31:56 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:31:56 INFO Configuration: resource-types.xml not found
23/01/04 15:31:56 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:31:56 INFO YarnClientImpl: Submitted application
application_1672817468245_0491
23/01/04 15:31:56 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0491/
23/01/04 15:31:56 INFO Job: Running job: job_1672817468245_0491
23/01/04 15:32:02 INFO Job: Job job_1672817468245_0491 running in uber mode : false
23/01/04 15:32:02 INFO Job: map 0% reduce 0%
    Map-Reduce Framework
        Map input records=27
        Map output records=8
        Map output bytes=3067198
        Map output materialized bytes=3067244
        Input split bytes=170
        Combine input records=0
        Combine output records=0
        Reduce input groups=8
        Reduce shuffle bytes=3067244
        Reduce input records=8
        Reduce output records=8
        Spilled Records=16
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=65
        CPU time spent (ms)=4810
        Physical memory (bytes) snapshot=753762304
        Virtual memory (bytes) snapshot=12758728704
        Total committed heap usage (bytes)=2105540608
```

```

Peak Map Physical memory (bytes)=433094656
Peak Map Virtual memory (bytes)=6375981056
Peak Reduce Physical memory (bytes)=320667648
Peak Reduce Virtual memory (bytes)=6382747648
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=2680027
File Output Format Counters
  Bytes Written=1732864
23/01/04 15:32:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:32:14 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0493
23/01/04 15:32:14 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:32:14 INFO JobSubmitter: number of splits:1
23/01/04 15:32:14 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0493
23/01/04 15:32:14 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:32:14 INFO YarnClientImpl: Submitted application
application_1672817468245_0493
23/01/04 15:32:14 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0493/
23/01/04 15:32:14 INFO Job: Running job: job_1672817468245_0493
23/01/04 15:32:21 INFO Job: Job job_1672817468245_0493 running in uber mode : false
23/01/04 15:32:21 INFO Job:  map 0% reduce 0%
23/01/04 15:32:27 INFO Job:  map 100% reduce 0%
23/01/04 15:32:33 INFO Job:  map 100% reduce 100%
23/01/04 15:32:33 INFO Job: Job job_1672817468245_0493 completed successfully
23/01/04 15:32:33 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4120779
  FILE: Number of bytes written=8773849
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6146313
  HDFS: Number of bytes written=1812933
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17820
  Total time spent by all reduces in occupied slots (ms)=18050
  Total time spent by all map tasks (ms)=3564
  Total time spent by all reduce tasks (ms)=3610
  Total vcore-milliseconds taken by all map tasks=3564
  Total vcore-milliseconds taken by all reduce tasks=3610
  Total megabyte-milliseconds taken by all map tasks=18247680
  Total megabyte-milliseconds taken by all reduce tasks=18483200

```

## Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4120733  
Map output materialized bytes=4120779  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4120779  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=4090  
Physical memory (bytes) snapshot=749838336  
Virtual memory (bytes) snapshot=12748812288  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412737536  
Peak Map Virtual memory (bytes)=6370258944  
Peak Reduce Physical memory (bytes)=337100800  
Peak Reduce Virtual memory (bytes)=6378553344

## 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=2680027

## File Output Format Counters

Bytes Written=1812933

23/01/04 15:32:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:32:33 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0495  
23/01/04 15:32:33 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:32:33 INFO JobSubmitter: number of splits:1  
23/01/04 15:32:33 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0495  
23/01/04 15:32:33 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:32:33 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0495  
23/01/04 15:32:33 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0495/](http://lena-master:8088/proxy/application_1672817468245_0495/)  
23/01/04 15:32:33 INFO Job: Running job: job\_1672817468245\_0495  
23/01/04 15:32:39 INFO Job: Job job\_1672817468245\_0495 running in uber mode : false  
23/01/04 15:32:39 INFO Job: map 0% reduce 0%  
23/01/04 15:32:45 INFO Job: map 100% reduce 0%  
23/01/04 15:32:51 INFO Job: map 100% reduce 100%  
23/01/04 15:32:51 INFO Job: Job job\_1672817468245\_0495 completed successfully  
23/01/04 15:32:51 INFO Job: Counters: 54

## File System Counters

FILE: Number of bytes read=4183818  
FILE: Number of bytes written=8899927

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6306451  
HDFS: Number of bytes written=1795903  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18240  
Total time spent by all reduces in occupied slots (ms)=17415  
Total time spent by all map tasks (ms)=3648  
Total time spent by all reduce tasks (ms)=3483  
Total vcore-milliseconds taken by all map tasks=3648  
Total vcore-milliseconds taken by all reduce tasks=3483  
Total megabyte-milliseconds taken by all map tasks=18677760  
Total megabyte-milliseconds taken by all reduce tasks=17832960

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4183772  
Map output materialized bytes=4183818  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4183818  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=4320  
Physical memory (bytes) snapshot=742793216  
Virtual memory (bytes) snapshot=12754432000  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400117760  
Peak Map Virtual memory (bytes)=6370660352  
Peak Reduce Physical memory (bytes)=342675456  
Peak Reduce Virtual memory (bytes)=6383771648

#### 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=2680027

#### File Output Format Counters

Bytes Written=1795903

23/01/04 15:32:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:32:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0497  
23/01/04 15:32:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:32:52 INFO JobSubmitter: number of splits:1  
23/01/04 15:32:52 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0497  
23/01/04 15:32:52 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:32:52 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0497  
23/01/04 15:32:52 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0497/  
23/01/04 15:32:52 INFO Job: Running job: job\_1672817468245\_0497  
23/01/04 15:32:58 INFO Job: Job job\_1672817468245\_0497 running in uber mode : false  
23/01/04 15:32:58 INFO Job: map 0% reduce 0%  
23/01/04 15:33:04 INFO Job: map 100% reduce 0%  
23/01/04 15:33:09 INFO Job: map 100% reduce 100%  
23/01/04 15:33:10 INFO Job: Job job\_1672817468245\_0497 completed successfully  
23/01/04 15:33:10 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4181306  
FILE: Number of bytes written=8894903  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6272391  
HDFS: Number of bytes written=1810421  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17925  
Total time spent by all reduces in occupied slots (ms)=17715  
Total time spent by all map tasks (ms)=3585  
Total time spent by all reduce tasks (ms)=3543  
Total vcore-milliseconds taken by all map tasks=3585  
Total vcore-milliseconds taken by all reduce tasks=3543  
Total megabyte-milliseconds taken by all map tasks=18355200  
Total megabyte-milliseconds taken by all reduce tasks=18140160

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4181260  
Map output materialized bytes=4181306  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4181306  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60



```

CPU time spent (ms)=4100
Physical memory (bytes) snapshot=730050560
Virtual memory (bytes) snapshot=12758261760
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=400670720
Peak Map Virtual memory (bytes)=6379368448
Peak Reduce Physical memory (bytes)=329379840
Peak Reduce Virtual memory (bytes)=6378893312
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=2680027
File Output Format Counters
  Bytes Written=1810421
23/01/04 15:33:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:33:10 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0499
23/01/04 15:33:10 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:33:10 INFO JobSubmitter: number of splits:1
23/01/04 15:33:10 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0499
23/01/04 15:33:10 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:33:10 INFO YarnClientImpl: Submitted application
application_1672817468245_0499
23/01/04 15:33:10 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0499/
23/01/04 15:33:10 INFO Job: Running job: job_1672817468245_0499
23/01/04 15:33:16 INFO Job: Job job_1672817468245_0499 running in uber mode : false
23/01/04 15:33:16 INFO Job:  map 0% reduce 0%
23/01/04 15:33:22 INFO Job:  map 100% reduce 0%
23/01/04 15:33:28 INFO Job:  map 100% reduce 100%
23/01/04 15:33:28 INFO Job: Job job_1672817468245_0499 completed successfully
23/01/04 15:33:28 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4200882
  FILE: Number of bytes written=8934055
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6301427
  HDFS: Number of bytes written=1815479
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18095
  Total time spent by all reduces in occupied slots (ms)=17965
  Total time spent by all map tasks (ms)=3619
  Total time spent by all reduce tasks (ms)=3593

```

Total vcore-milliseconds taken by all map tasks=3619  
Total vcore-milliseconds taken by all reduce tasks=3593  
Total megabyte-milliseconds taken by all map tasks=18529280  
Total megabyte-milliseconds taken by all reduce tasks=18396160

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4200836  
Map output materialized bytes=4200882  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4200882  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=4200  
Physical memory (bytes) snapshot=730583040  
Virtual memory (bytes) snapshot=12749094912  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=425312256  
Peak Map Virtual memory (bytes)=6374871040  
Peak Reduce Physical memory (bytes)=305270784  
Peak Reduce Virtual memory (bytes)=6374223872

#### 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=2680027

#### File Output Format Counters

Bytes Written=1815479

23/01/04 15:33:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:33:28 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0501  
23/01/04 15:33:29 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:33:29 INFO JobSubmitter: number of splits:1  
23/01/04 15:33:29 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0501  
23/01/04 15:33:29 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:33:29 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0501  
23/01/04 15:33:29 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0501/  
23/01/04 15:33:29 INFO Job: Running job: job\_1672817468245\_0501  
23/01/04 15:33:35 INFO Job: Job job\_1672817468245\_0501 running in uber mode : false  
23/01/04 15:33:35 INFO Job: map 0% reduce 0%  
23/01/04 15:33:40 INFO Job: map 100% reduce 0%  
23/01/04 15:33:46 INFO Job: map 100% reduce 100%  
23/01/04 15:33:47 INFO Job: Job job\_1672817468245\_0501 completed successfully

23/01/04 15:33:47 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4199664  
FILE: Number of bytes written=8931619  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6311543  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17700  
Total time spent by all reduces in occupied slots (ms)=17755  
Total time spent by all map tasks (ms)=3540  
Total time spent by all reduce tasks (ms)=3551  
Total vcore-milliseconds taken by all map tasks=3540  
Total vcore-milliseconds taken by all reduce tasks=3551  
Total megabyte-milliseconds taken by all map tasks=18124800  
Total megabyte-milliseconds taken by all reduce tasks=18181120

Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4199618  
Map output materialized bytes=4199664  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4199664  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=4170  
Physical memory (bytes) snapshot=726818816  
Virtual memory (bytes) snapshot=12755320832  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410394624  
Peak Map Virtual memory (bytes)=6374748160  
Peak Reduce Physical memory (bytes)=316424192  
Peak Reduce Virtual memory (bytes)=6380572672

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=2680027

#### File Output Format Counters

Bytes Written=1809203

23/01/04 15:33:47 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:33:47 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0503  
23/01/04 15:33:47 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:33:47 INFO JobSubmitter: number of splits:1  
23/01/04 15:33:47 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0503  
23/01/04 15:33:47 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:33:47 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0503  
23/01/04 15:33:47 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0503/  
23/01/04 15:33:47 INFO Job: Running job: job\_1672817468245\_0503  
23/01/04 15:33:53 INFO Job: Job job\_1672817468245\_0503 running in uber mode : false  
23/01/04 15:33:53 INFO Job: map 0% reduce 0%  
23/01/04 15:34:00 INFO Job: map 100% reduce 0%  
23/01/04 15:34:05 INFO Job: map 100% reduce 100%  
23/01/04 15:34:06 INFO Job: Job job\_1672817468245\_0503 completed successfully  
23/01/04 15:34:06 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18175  
Total time spent by all reduces in occupied slots (ms)=17175  
Total time spent by all map tasks (ms)=3635  
Total time spent by all reduce tasks (ms)=3435  
Total vcore-milliseconds taken by all map tasks=3635  
Total vcore-milliseconds taken by all reduce tasks=3435  
Total megabyte-milliseconds taken by all map tasks=18611200  
Total megabyte-milliseconds taken by all reduce tasks=17587200

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16

```

Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=60
CPU time spent (ms)=4210
Physical memory (bytes) snapshot=717365248
Virtual memory (bytes) snapshot=12759863296
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=406994944
Peak Map Virtual memory (bytes)=6380511232
Peak Reduce Physical memory (bytes)=310370304
Peak Reduce Virtual memory (bytes)=6379352064

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=2680027
File Output Format Counters
  Bytes Written=1809203
23/01/04 15:34:06 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:34:06 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0504
23/01/04 15:34:06 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:34:06 INFO JobSubmitter: number of splits:1
23/01/04 15:34:06 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0504
23/01/04 15:34:06 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:34:06 INFO YarnClientImpl: Submitted application
application_1672817468245_0504
23/01/04 15:34:06 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0504/
23/01/04 15:34:06 INFO Job: Running job: job_1672817468245_0504
23/01/04 15:34:12 INFO Job: Job job_1672817468245_0504 running in uber mode : false
23/01/04 15:34:12 INFO Job:  map 0% reduce 0%
23/01/04 15:34:18 INFO Job:  map 100% reduce 0%
23/01/04 15:34:24 INFO Job:  map 100% reduce 100%
23/01/04 15:34:24 INFO Job: Job job_1672817468245_0504 completed successfully
23/01/04 15:34:24 INFO Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=4193388
    FILE: Number of bytes written=8919067
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6298991
    HDFS: Number of bytes written=1809203
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1

```

Total time spent by all maps in occupied slots (ms)=18360  
Total time spent by all reduces in occupied slots (ms)=17390  
Total time spent by all map tasks (ms)=3672  
Total time spent by all reduce tasks (ms)=3478  
Total vcore-milliseconds taken by all map tasks=3672  
Total vcore-milliseconds taken by all reduce tasks=3478  
Total megabyte-milliseconds taken by all map tasks=18800640  
Total megabyte-milliseconds taken by all reduce tasks=17807360

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=4070  
Physical memory (bytes) snapshot=724049920  
Virtual memory (bytes) snapshot=12758327296  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402612224  
Peak Map Virtual memory (bytes)=6380675072  
Peak Reduce Physical memory (bytes)=321437696  
Peak Reduce Virtual memory (bytes)=6377652224

#### 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=2680027

#### File Output Format Counters

Bytes Written=1809203

23/01/04 15:34:24 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:34:24 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0506  
23/01/04 15:34:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:34:24 INFO JobSubmitter: number of splits:1  
23/01/04 15:34:24 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0506  
23/01/04 15:34:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:34:24 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0506  
23/01/04 15:34:24 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0506/](http://lena-master:8088/proxy/application_1672817468245_0506/)  
23/01/04 15:34:24 INFO Job: Running job: job\_1672817468245\_0506  
23/01/04 15:34:30 INFO Job: Job job\_1672817468245\_0506 running in uber mode : false

23/01/04 15:34:30 INFO Job: map 0% reduce 0%  
23/01/04 15:34:37 INFO Job: map 100% reduce 0%  
23/01/04 15:34:43 INFO Job: map 100% reduce 100%  
23/01/04 15:34:43 INFO Job: Job job\_1672817468245\_0506 completed successfully  
23/01/04 15:34:43 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18415  
Total time spent by all reduces in occupied slots (ms)=17275  
Total time spent by all map tasks (ms)=3683  
Total time spent by all reduce tasks (ms)=3455  
Total vcore-milliseconds taken by all map tasks=3683  
Total vcore-milliseconds taken by all reduce tasks=3455  
Total megabyte-milliseconds taken by all map tasks=18856960  
Total megabyte-milliseconds taken by all reduce tasks=17689600

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=4320  
Physical memory (bytes) snapshot=720158720  
Virtual memory (bytes) snapshot=12761169920  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411754496  
Peak Map Virtual memory (bytes)=6374834176  
Peak Reduce Physical memory (bytes)=308404224  
Peak Reduce Virtual memory (bytes)=6386335744

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1809203  
23/01/04 15:34:43 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:34:43 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0508  
23/01/04 15:34:43 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:34:43 INFO JobSubmitter: number of splits:1  
23/01/04 15:34:43 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0508  
23/01/04 15:34:43 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:34:43 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0508  
23/01/04 15:34:43 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0508/](http://lena-master:8088/proxy/application_1672817468245_0508/)  
23/01/04 15:34:43 INFO Job: Running job: job\_1672817468245\_0508  
23/01/04 15:34:50 INFO Job: Job job\_1672817468245\_0508 running in uber mode : false  
23/01/04 15:34:50 INFO Job:  map 0% reduce 0%  
23/01/04 15:34:55 INFO Job:  map 100% reduce 0%  
23/01/04 15:35:01 INFO Job:  map 100% reduce 100%  
23/01/04 15:35:01 INFO Job: Job job\_1672817468245\_0508 completed successfully  
23/01/04 15:35:01 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919071  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17975  
Total time spent by all reduces in occupied slots (ms)=17335  
Total time spent by all map tasks (ms)=3595  
Total time spent by all reduce tasks (ms)=3467  
Total vcore-milliseconds taken by all map tasks=3595  
Total vcore-milliseconds taken by all reduce tasks=3467  
Total megabyte-milliseconds taken by all map tasks=18406400  
Total megabyte-milliseconds taken by all reduce tasks=17751040

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8



Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=4330  
Physical memory (bytes) snapshot=728285184  
Virtual memory (bytes) snapshot=12760518656  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405020672  
Peak Map Virtual memory (bytes)=6378565632  
Peak Reduce Physical memory (bytes)=323264512  
Peak Reduce Virtual memory (bytes)=6381953024

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=2680027

File Output Format Counters

Bytes Written=1809203

23/01/04 15:35:01 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:35:01 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0511  
23/01/04 15:35:01 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:35:01 INFO JobSubmitter: number of splits:1  
23/01/04 15:35:01 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0511  
23/01/04 15:35:01 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:35:02 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0511  
23/01/04 15:35:02 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0511/  
23/01/04 15:35:02 INFO Job: Running job: job\_1672817468245\_0511  
23/01/04 15:35:08 INFO Job: Job job\_1672817468245\_0511 running in uber mode : false  
23/01/04 15:35:08 INFO Job: map 0% reduce 0%  
23/01/04 15:35:14 INFO Job: map 100% reduce 0%  
23/01/04 15:35:19 INFO Job: map 100% reduce 100%  
23/01/04 15:35:20 INFO Job: Job job\_1672817468245\_0511 completed successfully  
23/01/04 15:35:20 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919075  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18170  
Total time spent by all reduces in occupied slots (ms)=17515  
Total time spent by all map tasks (ms)=3634  
Total time spent by all reduce tasks (ms)=3503  
Total vcore-milliseconds taken by all map tasks=3634  
Total vcore-milliseconds taken by all reduce tasks=3503  
Total megabyte-milliseconds taken by all map tasks=18606080  
Total megabyte-milliseconds taken by all reduce tasks=17935360

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=4400  
Physical memory (bytes) snapshot=775249920  
Virtual memory (bytes) snapshot=12768837632  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=427507712  
Peak Map Virtual memory (bytes)=6382628864  
Peak Reduce Physical memory (bytes)=347742208  
Peak Reduce Virtual memory (bytes)=6386208768

#### 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=2680027

#### File Output Format Counters

Bytes Written=1809203

23/01/04 15:35:20 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:35:20 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0513

23/01/04 15:35:20 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:35:20 INFO JobSubmitter: number of splits:1

23/01/04 15:35:20 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0513

23/01/04 15:35:20 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:35:20 INFO YarnClientImpl: Submitted application application\_1672817468245\_0513

23/01/04 15:35:20 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0513/  
23/01/04 15:35:20 INFO Job: Running job: job\_1672817468245\_0513  
23/01/04 15:35:26 INFO Job: Job job\_1672817468245\_0513 running in uber mode : false  
23/01/04 15:35:26 INFO Job: map 0% reduce 0%  
23/01/04 15:35:32 INFO Job: map 100% reduce 0%  
23/01/04 15:35:38 INFO Job: map 100% reduce 100%  
23/01/04 15:35:38 INFO Job: Job job\_1672817468245\_0513 completed successfully  
23/01/04 15:35:38 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919075  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17850  
Total time spent by all reduces in occupied slots (ms)=17340  
Total time spent by all map tasks (ms)=3570  
Total time spent by all reduce tasks (ms)=3468  
Total vcore-milliseconds taken by all map tasks=3570  
Total vcore-milliseconds taken by all reduce tasks=3468  
Total megabyte-milliseconds taken by all map tasks=18278400  
Total megabyte-milliseconds taken by all reduce tasks=17756160

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=37  
CPU time spent (ms)=4310  
Physical memory (bytes) snapshot=749191168  
Virtual memory (bytes) snapshot=12750966784  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=422694912  
Peak Map Virtual memory (bytes)=6377132032  
Peak Reduce Physical memory (bytes)=326496256  
Peak Reduce Virtual memory (bytes)=6373834752

#### 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=2680027

File Output Format Counters

Bytes Written=1809203

23/01/04 15:35:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:35:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0515

23/01/04 15:35:39 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:35:39 INFO JobSubmitter: number of splits:1

23/01/04 15:35:39 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0515

23/01/04 15:35:39 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:35:39 INFO YarnClientImpl: Submitted application application\_1672817468245\_0515

23/01/04 15:35:39 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0515/](http://lena-master:8088/proxy/application_1672817468245_0515/)

23/01/04 15:35:39 INFO Job: Running job: job\_1672817468245\_0515

23/01/04 15:35:45 INFO Job: Job job\_1672817468245\_0515 running in uber mode : false

23/01/04 15:35:45 INFO Job: map 0% reduce 0%

23/01/04 15:35:51 INFO Job: map 100% reduce 0%

23/01/04 15:35:56 INFO Job: map 100% reduce 100%

23/01/04 15:35:57 INFO Job: Job job\_1672817468245\_0515 completed successfully

23/01/04 15:35:57 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4193388

FILE: Number of bytes written=8919075

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=6298991

HDFS: Number of bytes written=1809203

HDFS: Number of read operations=19

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=18090

Total time spent by all reduces in occupied slots (ms)=17000

Total time spent by all map tasks (ms)=3618

Total time spent by all reduce tasks (ms)=3400

Total vcore-milliseconds taken by all map tasks=3618

Total vcore-milliseconds taken by all reduce tasks=3400

Total megabyte-milliseconds taken by all map tasks=18524160

Total megabyte-milliseconds taken by all reduce tasks=17408000

Map-Reduce Framework

Map input records=27

Map output records=8

Map output bytes=4193342

Map output materialized bytes=4193388

Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=33  
CPU time spent (ms)=4100  
Physical memory (bytes) snapshot=728469504  
Virtual memory (bytes) snapshot=12760682496  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=414031872  
Peak Map Virtual memory (bytes)=6373781504  
Peak Reduce Physical memory (bytes)=314437632  
Peak Reduce Virtual memory (bytes)=6386900992

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=2680027

File Output Format Counters

Bytes Written=1809203

23/01/04 15:35:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:35:57 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0517  
23/01/04 15:35:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:35:57 INFO JobSubmitter: number of splits:1  
23/01/04 15:35:57 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0517  
23/01/04 15:35:57 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:35:57 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0517  
23/01/04 15:35:57 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0517/  
23/01/04 15:35:57 INFO Job: Running job: job\_1672817468245\_0517  
23/01/04 15:36:03 INFO Job: Job job\_1672817468245\_0517 running in uber mode : false  
23/01/04 15:36:03 INFO Job: map 0% reduce 0%  
23/01/04 15:36:09 INFO Job: map 100% reduce 0%  
23/01/04 15:36:15 INFO Job: map 100% reduce 100%  
23/01/04 15:36:15 INFO Job: Job job\_1672817468245\_0517 completed successfully  
23/01/04 15:36:15 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4193388  
FILE: Number of bytes written=8919075  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6298991  
HDFS: Number of bytes written=1809203

HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17625  
Total time spent by all reduces in occupied slots (ms)=18565  
Total time spent by all map tasks (ms)=3525  
Total time spent by all reduce tasks (ms)=3713  
Total vcore-milliseconds taken by all map tasks=3525  
Total vcore-milliseconds taken by all reduce tasks=3713  
Total megabyte-milliseconds taken by all map tasks=18048000  
Total megabyte-milliseconds taken by all reduce tasks=19010560

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4193342  
Map output materialized bytes=4193388  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4193388  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=4270  
Physical memory (bytes) snapshot=743616512  
Virtual memory (bytes) snapshot=12755165184  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411525120  
Peak Map Virtual memory (bytes)=6372995072  
Peak Reduce Physical memory (bytes)=332091392  
Peak Reduce Virtual memory (bytes)=6382170112

#### 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=2680027

#### File Output Format Counters

Bytes Written=1809203

23/01/04 15:36:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:36:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0519

23/01/04 15:36:16 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:36:16 INFO JobSubmitter: number of splits:1

23/01/04 15:36:16 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0519  
23/01/04 15:36:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:36:16 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0519  
23/01/04 15:36:16 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0519/  
23/01/04 15:36:16 INFO Job: Running job: job\_1672817468245\_0519  
1672817468245\_0541  
23/01/04 15:38:45 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:38:45 INFO JobSubmitter: number of splits:1  
23/01/04 15:38:45 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0541  
23/01/04 15:38:45 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:38:45 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0541  
23/01/04 15:38:45 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0541/  
23/01/04 15:38:45 INFO Job: Running job: job\_1672817468245\_0541  
23/01/04 15:38:51 INFO Job: Job job\_1672817468245\_0541 running in uber mode : false  
23/01/04 15:38:51 INFO Job: map 0% reduce 0%  
23/01/04 15:38:57 INFO Job: map 100% reduce 0%  
23/01/04 15:39:03 INFO Job: map 100% reduce 100%  
23/01/04 15:39:03 INFO Job: Job job\_1672817468245\_0541 completed successfully  
23/01/04 15:39:03 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=866659  
FILE: Number of bytes written=2265911  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6659079  
HDFS: Number of bytes written=2855868  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16695  
Total time spent by all reduces in occupied slots (ms)=16790  
Total time spent by all map tasks (ms)=3339  
Total time spent by all reduce tasks (ms)=3358  
Total vcore-milliseconds taken by all map tasks=3339  
Total vcore-milliseconds taken by all reduce tasks=3358  
Total megabyte-milliseconds taken by all map tasks=17095680  
Total megabyte-milliseconds taken by all reduce tasks=17192960

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=866614  
Map output materialized bytes=866659  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=866659  
Reduce input records=8

Reduce output records=24  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=3130  
Physical memory (bytes) snapshot=699035648  
Virtual memory (bytes) snapshot=12761169920  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402042880  
Peak Map Virtual memory (bytes)=6379646976  
Peak Reduce Physical memory (bytes)=296992768  
Peak Reduce Virtual memory (bytes)=6383480832

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=2680648

File Output Format Counters

Bytes Written=2855868

23/01/04 15:39:03 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 15:39:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:39:03 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0544

23/01/04 15:39:03 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:39:03 INFO JobSubmitter: number of splits:1

23/01/04 15:39:03 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0544

23/01/04 15:39:03 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:39:03 INFO YarnClientImpl: Submitted application application\_1672817468245\_0544

23/01/04 15:39:03 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0544/

23/01/04 15:39:03 INFO Job: Running job: job\_1672817468245\_0544

23/01/04 15:39:09 INFO Job: Job job\_1672817468245\_0544 running in uber mode : false

23/01/04 15:39:09 INFO Job: map 0% reduce 0%

23/01/04 15:39:15 INFO Job: map 100% reduce 0%

23/01/04 15:39:21 INFO Job: map 100% reduce 100%

23/01/04 15:39:21 INFO Job: Job job\_1672817468245\_0544 completed successfully

23/01/04 15:39:21 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=811786  
FILE: Number of bytes written=2156165  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8392587  
HDFS: Number of bytes written=3667729  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0



#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17035  
Total time spent by all reduces in occupied slots (ms)=16510  
Total time spent by all map tasks (ms)=3407  
Total time spent by all reduce tasks (ms)=3302  
Total vcore-milliseconds taken by all map tasks=3407  
Total vcore-milliseconds taken by all reduce tasks=3302  
Total megabyte-milliseconds taken by all map tasks=17443840  
Total megabyte-milliseconds taken by all reduce tasks=16906240

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=811741  
Map output materialized bytes=811786  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=811786  
Reduce input records=8  
Reduce output records=32  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=54  
CPU time spent (ms)=3290  
Physical memory (bytes) snapshot=706588672  
Virtual memory (bytes) snapshot=12751577088  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412200960  
Peak Map Virtual memory (bytes)=6369701888  
Peak Reduce Physical memory (bytes)=294387712  
Peak Reduce Virtual memory (bytes)=6383828992

#### 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=2680648

#### File Output Format Counters

Bytes Written=3667729

23/01/04 15:39:21 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:39:21 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:39:21 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0547

23/01/04 15:39:21 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:39:21 INFO JobSubmitter: number of splits:1

23/01/04 15:39:21 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0547

23/01/04 15:39:21 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:39:21 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0547  
23/01/04 15:39:21 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0547/  
23/01/04 15:39:21 INFO Job: Running job: job\_1672817468245\_0547  
23/01/04 15:39:27 INFO Job: Job job\_1672817468245\_0547 running in uber mode : false  
23/01/04 15:39:27 INFO Job: map 0% reduce 0%  
23/01/04 15:39:32 INFO Job: map 100% reduce 0%  
23/01/04 15:39:38 INFO Job: map 100% reduce 100%  
23/01/04 15:39:38 INFO Job: Job job\_1672817468245\_0547 completed successfully  
23/01/04 15:39:38 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=773385  
FILE: Number of bytes written=2079363  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=10016309  
HDFS: Number of bytes written=4441190  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18005  
Total time spent by all reduces in occupied slots (ms)=16695  
Total time spent by all map tasks (ms)=3601  
Total time spent by all reduce tasks (ms)=3339  
Total vcore-milliseconds taken by all map tasks=3601  
Total vcore-milliseconds taken by all reduce tasks=3339  
Total megabyte-milliseconds taken by all map tasks=18437120  
Total megabyte-milliseconds taken by all reduce tasks=17095680

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=773341  
Map output materialized bytes=773385  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=773385  
Reduce input records=8  
Reduce output records=40  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=99  
CPU time spent (ms)=3520  
Physical memory (bytes) snapshot=691077120  
Virtual memory (bytes) snapshot=12748943360  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401920000  
Peak Map Virtual memory (bytes)=6373203968  
Peak Reduce Physical memory (bytes)=293580800

Peak Reduce Virtual memory (bytes)=6375739392  
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=2680648  
File Output Format Counters  
Bytes Written=4441190

23/01/04 15:39:38 INFO RepresentativePointsDriver: Representative Points Iteration 4  
23/01/04 15:39:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:39:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0550  
23/01/04 15:39:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:39:39 INFO JobSubmitter: number of splits:1  
23/01/04 15:39:39 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0550  
23/01/04 15:39:39 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:39:39 INFO YarnClientImpl: Submitted application application\_1672817468245\_0550  
23/01/04 15:39:39 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0550/  
23/01/04 15:39:39 INFO Job: Running job: job\_1672817468245\_0550  
23/01/04 15:39:45 INFO Job: Job job\_1672817468245\_0550 running in uber mode : false  
23/01/04 15:39:45 INFO Job: map 0% reduce 0%  
23/01/04 15:39:50 INFO Job: map 100% reduce 0%  
23/01/04 15:39:56 INFO Job: map 100% reduce 100%  
23/01/04 15:39:56 INFO Job: Job job\_1672817468245\_0550 completed successfully  
23/01/04 15:39:56 INFO Job: Counters: 54

File System Counters  
FILE: Number of bytes read=806618  
FILE: Number of bytes written=2145829  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11563231  
HDFS: Number of bytes written=5247863  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17860  
Total time spent by all reduces in occupied slots (ms)=16770  
Total time spent by all map tasks (ms)=3572  
Total time spent by all reduce tasks (ms)=3354  
Total vcore-milliseconds taken by all map tasks=3572  
Total vcore-milliseconds taken by all reduce tasks=3354  
Total megabyte-milliseconds taken by all map tasks=18288640  
Total megabyte-milliseconds taken by all reduce tasks=17172480  
Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=806573  
Map output materialized bytes=806618  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=806618  
Reduce input records=8  
Reduce output records=48  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3450  
Physical memory (bytes) snapshot=703107072  
Virtual memory (bytes) snapshot=12754010112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404942848  
Peak Map Virtual memory (bytes)=6376640512  
Peak Reduce Physical memory (bytes)=298164224  
Peak Reduce Virtual memory (bytes)=6377369600

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=2680648

File Output Format Counters

Bytes Written=5247863

23/01/04 15:39:56 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.5886736324975431  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[24] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[1] =  
0.6709502844786016  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[21] =  
0.7696631966848229  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[16] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[4] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[14] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN  
23/01/04 15:39:56 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.7203067405817123  
23/01/04 15:39:57 INFO ClusterDumper: Wrote 8 clusters  
23/01/04 15:39:57 INFO MahoutDriver: Program took 92184 ms (Minutes: 1.5364)  
---- K: 09 --  
dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/09  
MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.  
Running on hadoop, using /opt/hadoop/current/bin/hadoop and  
HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop  
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar  
23/01/04 15:40:02 INFO AbstractJob: Command line arguments: {--clustering=null, --  
clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/

```

EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[9],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/09], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:40:03 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/09
23/01/04 15:40:03 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:40:03 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:40:03 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:40:03 INFO RandomSeedGenerator: Wrote 9 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:40:03 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/09
23/01/04 15:40:03 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:40:03 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:40:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:40:04 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0553
23/01/04 15:40:04 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:40:04 INFO JobSubmitter: number of splits:1
23/01/04 15:40:04 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0553
23/01/04 15:40:04 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:40:04 INFO Configuration: resource-types.xml not found
23/01/04 15:40:04 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:40:04 INFO YarnClientImpl: Submitted application
application_1672817468245_0553
23/01/04 15:40:04 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0553/
23/01/04 15:40:04 INFO Job: Running job: job_1672817468245_0553
23/01/04 15:40:10 INFO Job: Job job_1672817468245_0553 running in uber mode : false
23/01/04 15:40:10 INFO Job: map 0% reduce 0%
23/01/04 15:40:16 INFO Job: map 100% reduce 0%
23/01/04 15:40:23 INFO Job: map 100% reduce 100%
23/01/04 15:40:23 INFO Job: Job job_1672817468245_0553 completed successfully
23/01/04 15:40:23 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=3520040
        FILE: Number of bytes written=7572371
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=8047449
        HDFS: Number of bytes written=1934315
        HDFS: Number of read operations=51
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2

```

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18305  
Total time spent by all reduces in occupied slots (ms)=17230  
Total time spent by all map tasks (ms)=3661  
Total time spent by all reduce tasks (ms)=3446  
Total vcore-milliseconds taken by all map tasks=3661  
Total vcore-milliseconds taken by all reduce tasks=3446  
Total megabyte-milliseconds taken by all map tasks=18744320  
Total megabyte-milliseconds taken by all reduce tasks=17643520

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=3519989  
Map output materialized bytes=3520040  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=3520040  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=4670  
Physical memory (bytes) snapshot=734842880  
Virtual memory (bytes) snapshot=12746604544  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412127232  
Peak Map Virtual memory (bytes)=6371483648  
Peak Reduce Physical memory (bytes)=322715648  
Peak Reduce Virtual memory (bytes)=6375120896

#### 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=2680027

#### File Output Format Counters

Bytes Written=1934315

23/01/04 15:40:23 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:40:23 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0557

23/01/04 15:40:23 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:40:23 INFO JobSubmitter: number of splits:1

23/01/04 15:40:23 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0557

23/01/04 15:40:23 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:40:23 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0557

23/01/04 15:40:23 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0557/](http://lena-master:8088/proxy/application_1672817468245_0557/)

23/01/04 15:40:23 INFO Job: Running job: job\_1672817468245\_0557

23/01/04 15:40:29 INFO Job: Job job\_1672817468245\_0557 running in uber mode : false

23/01/04 15:40:29 INFO Job: map 0% reduce 0%

23/01/04 15:40:35 INFO Job: map 100% reduce 0%

23/01/04 15:40:41 INFO Job: map 100% reduce 100%

23/01/04 15:40:41 INFO Job: Job job\_1672817468245\_0557 completed successfully

23/01/04 15:40:41 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4552241

FILE: Number of bytes written=9636773

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=6549215

HDFS: Number of bytes written=1928083

HDFS: Number of read operations=19

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=17715

Total time spent by all reduces in occupied slots (ms)=17415

Total time spent by all map tasks (ms)=3543

Total time spent by all reduce tasks (ms)=3483

Total vcore-milliseconds taken by all map tasks=3543

Total vcore-milliseconds taken by all reduce tasks=3483

Total megabyte-milliseconds taken by all map tasks=18140160

Total megabyte-milliseconds taken by all reduce tasks=17832960

#### Map-Reduce Framework

Map input records=27

Map output records=9

Map output bytes=4552190

Map output materialized bytes=4552241

Input split bytes=170

Combine input records=0

Combine output records=0

Reduce input groups=9

Reduce shuffle bytes=4552241

Reduce input records=9

Reduce output records=9

Spilled Records=18

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=61

CPU time spent (ms)=4060

Physical memory (bytes) snapshot=729821184

Virtual memory (bytes) snapshot=12760592384

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=415531008

Peak Map Virtual memory (bytes)=6378549248

Peak Reduce Physical memory (bytes)=314290176

Peak Reduce Virtual memory (bytes)=6382043136

#### 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=2680027

#### File Output Format Counters

Bytes Written=1928083

23/01/04 15:40:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:40:41 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0560

23/01/04 15:40:41 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:40:41 INFO JobSubmitter: number of splits:1

23/01/04 15:40:41 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0560

23/01/04 15:40:41 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:40:41 INFO YarnClientImpl: Submitted application application\_1672817468245\_0560

23/01/04 15:40:41 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0560/](http://lena-master:8088/proxy/application_1672817468245_0560/)

23/01/04 15:40:41 INFO Job: Running job: job\_1672817468245\_0560

23/01/04 15:40:48 INFO Job: Job job\_1672817468245\_0560 running in uber mode : false

23/01/04 15:40:48 INFO Job: map 0% reduce 0%

23/01/04 15:40:54 INFO Job: map 100% reduce 0%

23/01/04 15:41:00 INFO Job: map 100% reduce 100%

23/01/04 15:41:00 INFO Job: Job job\_1672817468245\_0560 completed successfully

23/01/04 15:41:00 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4538111  
FILE: Number of bytes written=9608513  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6536751  
HDFS: Number of bytes written=1920175  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17910  
Total time spent by all reduces in occupied slots (ms)=17530  
Total time spent by all map tasks (ms)=3582  
Total time spent by all reduce tasks (ms)=3506  
Total vcore-milliseconds taken by all map tasks=3582  
Total vcore-milliseconds taken by all reduce tasks=3506  
Total megabyte-milliseconds taken by all map tasks=18339840  
Total megabyte-milliseconds taken by all reduce tasks=17950720

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4538060



Map output materialized bytes=4538111  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4538111  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=4370  
Physical memory (bytes) snapshot=720183296  
Virtual memory (bytes) snapshot=12758716416  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411930624  
Peak Map Virtual memory (bytes)=6382112768  
Peak Reduce Physical memory (bytes)=308252672  
Peak Reduce Virtual memory (bytes)=6376603648

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=2680027

File Output Format Counters

Bytes Written=1920175

23/01/04 15:41:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:41:00 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0563  
23/01/04 15:41:00 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:41:00 INFO JobSubmitter: number of splits:1  
23/01/04 15:41:00 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0563  
23/01/04 15:41:00 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:41:00 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0563  
23/01/04 15:41:00 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0563/  
23/01/04 15:41:00 INFO Job: Running job: job\_1672817468245\_0563  
23/01/04 15:41:06 INFO Job: Job job\_1672817468245\_0563 running in uber mode : false  
23/01/04 15:41:06 INFO Job: map 0% reduce 0%  
23/01/04 15:41:12 INFO Job: map 100% reduce 0%  
23/01/04 15:41:16 INFO Job: map 100% reduce 100%  
23/01/04 15:41:16 INFO Job: Job job\_1672817468245\_0563 completed successfully  
23/01/04 15:41:16 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4453771  
FILE: Number of bytes written=9439833  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6520935

HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18120  
Total time spent by all reduces in occupied slots (ms)=10095  
Total time spent by all map tasks (ms)=3624  
Total time spent by all reduce tasks (ms)=2019  
Total vcore-milliseconds taken by all map tasks=3624  
Total vcore-milliseconds taken by all reduce tasks=2019  
Total megabyte-milliseconds taken by all map tasks=18554880  
Total megabyte-milliseconds taken by all reduce tasks=10337280

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4453720  
Map output materialized bytes=4453771  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4453771  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=4340  
Physical memory (bytes) snapshot=726843392  
Virtual memory (bytes) snapshot=12751101952  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396947456  
Peak Map Virtual memory (bytes)=6372872192  
Peak Reduce Physical memory (bytes)=329895936  
Peak Reduce Virtual memory (bytes)=6378229760

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:41:16 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:41:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0566

23/01/04 15:41:16 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:41:16 INFO JobSubmitter: number of splits:1

23/01/04 15:41:17 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0566  
23/01/04 15:41:17 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:41:17 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0566  
23/01/04 15:41:17 INFO Job: The url to track the job:  
http://lana-master:8088/proxy/application\_1672817468245\_0566/  
23/01/04 15:41:17 INFO Job: Running job: job\_1672817468245\_0566  
23/01/04 15:41:23 INFO Job: Job job\_1672817468245\_0566 running in uber mode : false  
23/01/04 15:41:23 INFO Job: map 0% reduce 0%  
23/01/04 15:41:29 INFO Job: map 100% reduce 0%  
23/01/04 15:41:34 INFO Job: map 100% reduce 100%  
23/01/04 15:41:35 INFO Job: Job job\_1672817468245\_0566 completed successfully  
23/01/04 15:41:35 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286059  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17900  
Total time spent by all reduces in occupied slots (ms)=17715  
Total time spent by all map tasks (ms)=3580  
Total time spent by all reduce tasks (ms)=3543  
Total vcore-milliseconds taken by all map tasks=3580  
Total vcore-milliseconds taken by all reduce tasks=3543  
Total megabyte-milliseconds taken by all map tasks=18329600  
Total megabyte-milliseconds taken by all reduce tasks=18140160

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=4130  
Physical memory (bytes) snapshot=708046848  
Virtual memory (bytes) snapshot=12762054656  
Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=398077952  
Peak Map Virtual memory (bytes)=6373203968  
Peak Reduce Physical memory (bytes)=309968896  
Peak Reduce Virtual memory (bytes)=6388850688

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=2680027

File Output Format Counters

Bytes Written=1843288

23/01/04 15:41:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:41:35 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0569  
23/01/04 15:41:35 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:41:35 INFO JobSubmitter: number of splits:1  
23/01/04 15:41:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0569  
23/01/04 15:41:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:41:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0569  
23/01/04 15:41:35 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0569/](http://lena-master:8088/proxy/application_1672817468245_0569/)  
23/01/04 15:41:35 INFO Job: Running job: job\_1672817468245\_0569  
23/01/04 15:41:41 INFO Job: Job job\_1672817468245\_0569 running in uber mode : false  
23/01/04 15:41:41 INFO Job: map 0% reduce 0%  
23/01/04 15:41:47 INFO Job: map 100% reduce 0%  
23/01/04 15:41:51 INFO Job: map 100% reduce 100%  
23/01/04 15:41:51 INFO Job: Job job\_1672817468245\_0569 completed successfully  
23/01/04 15:41:51 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286059  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17875  
Total time spent by all reduces in occupied slots (ms)=9865  
Total time spent by all map tasks (ms)=3575  
Total time spent by all reduce tasks (ms)=1973  
Total vcore-milliseconds taken by all map tasks=3575  
Total vcore-milliseconds taken by all reduce tasks=1973  
Total megabyte-milliseconds taken by all map tasks=18304000  
Total megabyte-milliseconds taken by all reduce tasks=10101760

## Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=4110  
Physical memory (bytes) snapshot=720613376  
Virtual memory (bytes) snapshot=12760190976  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408563712  
Peak Map Virtual memory (bytes)=6374354944  
Peak Reduce Physical memory (bytes)=312049664  
Peak Reduce Virtual memory (bytes)=6385836032

## 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=2680027

## File Output Format Counters

Bytes Written=1843288

23/01/04 15:41:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:41:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0572  
23/01/04 15:41:52 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:41:52 INFO JobSubmitter: number of splits:1  
23/01/04 15:41:52 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0572  
23/01/04 15:41:52 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:41:52 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0572  
23/01/04 15:41:52 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0572/](http://lena-master:8088/proxy/application_1672817468245_0572/)  
23/01/04 15:41:52 INFO Job: Running job: job\_1672817468245\_0572  
23/01/04 15:41:58 INFO Job: Job job\_1672817468245\_0572 running in uber mode : false  
23/01/04 15:41:58 INFO Job: map 0% reduce 0%  
23/01/04 15:42:04 INFO Job: map 100% reduce 0%  
23/01/04 15:42:10 INFO Job: map 100% reduce 100%  
23/01/04 15:42:10 INFO Job: Job job\_1672817468245\_0572 completed successfully  
23/01/04 15:42:10 INFO Job: Counters: 54

## File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286059

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17955  
Total time spent by all reduces in occupied slots (ms)=17530  
Total time spent by all map tasks (ms)=3591  
Total time spent by all reduce tasks (ms)=3506  
Total vcore-milliseconds taken by all map tasks=3591  
Total vcore-milliseconds taken by all reduce tasks=3506  
Total megabyte-milliseconds taken by all map tasks=18385920  
Total megabyte-milliseconds taken by all reduce tasks=17950720

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=4330  
Physical memory (bytes) snapshot=721448960  
Virtual memory (bytes) snapshot=12754415616  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405639168  
Peak Map Virtual memory (bytes)=6371827712  
Peak Reduce Physical memory (bytes)=315809792  
Peak Reduce Virtual memory (bytes)=6382587904

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:42:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:42:10 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0575  
23/01/04 15:42:10 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:42:10 INFO JobSubmitter: number of splits:1  
23/01/04 15:42:10 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0575  
23/01/04 15:42:10 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:42:10 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0575  
23/01/04 15:42:10 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0575/  
23/01/04 15:42:10 INFO Job: Running job: job\_1672817468245\_0575  
23/01/04 15:42:16 INFO Job: Job job\_1672817468245\_0575 running in uber mode : false  
23/01/04 15:42:16 INFO Job: map 0% reduce 0%  
23/01/04 15:42:22 INFO Job: map 100% reduce 0%  
23/01/04 15:42:27 INFO Job: map 100% reduce 100%  
23/01/04 15:42:27 INFO Job: Job job\_1672817468245\_0575 completed successfully  
23/01/04 15:42:27 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286059  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18090  
Total time spent by all reduces in occupied slots (ms)=17575  
Total time spent by all map tasks (ms)=3618  
Total time spent by all reduce tasks (ms)=3515  
Total vcore-milliseconds taken by all map tasks=3618  
Total vcore-milliseconds taken by all reduce tasks=3515  
Total megabyte-milliseconds taken by all map tasks=18524160  
Total megabyte-milliseconds taken by all reduce tasks=17996800

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68

```

CPU time spent (ms)=4100
Physical memory (bytes) snapshot=707280896
Virtual memory (bytes) snapshot=12750471168
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=401227776
Peak Map Virtual memory (bytes)=6371811328
Peak Reduce Physical memory (bytes)=306053120
Peak Reduce Virtual memory (bytes)=6378659840
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=2680027
File Output Format Counters
  Bytes Written=1843288
23/01/04 15:42:28 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:42:28 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0577
23/01/04 15:42:28 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:42:28 INFO JobSubmitter: number of splits:1
23/01/04 15:42:28 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0577
23/01/04 15:42:28 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:42:28 INFO YarnClientImpl: Submitted application
application_1672817468245_0577
23/01/04 15:42:28 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0577/
23/01/04 15:42:28 INFO Job: Running job: job_1672817468245_0577
23/01/04 15:42:34 INFO Job: Job job_1672817468245_0577 running in uber mode : false
23/01/04 15:42:34 INFO Job:  map 0% reduce 0%
23/01/04 15:42:40 INFO Job:  map 100% reduce 0%
23/01/04 15:42:45 INFO Job:  map 100% reduce 100%
23/01/04 15:42:46 INFO Job: Job job_1672817468245_0577 completed successfully
23/01/04 15:42:46 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4376884
  FILE: Number of bytes written=9286059
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6367161
  HDFS: Number of bytes written=1843288
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18120
  Total time spent by all reduces in occupied slots (ms)=17675
  Total time spent by all map tasks (ms)=3624
  Total time spent by all reduce tasks (ms)=3535

```



Total vcore-milliseconds taken by all map tasks=3624  
Total vcore-milliseconds taken by all reduce tasks=3535  
Total megabyte-milliseconds taken by all map tasks=18554880  
Total megabyte-milliseconds taken by all reduce tasks=18099200

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=75  
CPU time spent (ms)=4180  
Physical memory (bytes) snapshot=743133184  
Virtual memory (bytes) snapshot=12765057024  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=407420928  
Peak Map Virtual memory (bytes)=6379937792  
Peak Reduce Physical memory (bytes)=335712256  
Peak Reduce Virtual memory (bytes)=6385119232

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:42:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:42:46 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0580  
23/01/04 15:42:46 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:42:47 INFO JobSubmitter: number of splits:1  
23/01/04 15:42:47 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0580  
23/01/04 15:42:47 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:42:47 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0580  
23/01/04 15:42:47 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0580/  
23/01/04 15:42:47 INFO Job: Running job: job\_1672817468245\_0580  
23/01/04 15:42:53 INFO Job: Job job\_1672817468245\_0580 running in uber mode : false  
23/01/04 15:42:53 INFO Job: map 0% reduce 0%  
23/01/04 15:42:59 INFO Job: map 100% reduce 0%  
23/01/04 15:43:05 INFO Job: map 100% reduce 100%  
23/01/04 15:43:05 INFO Job: Job job\_1672817468245\_0580 completed successfully

23/01/04 15:43:05 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286063  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18275  
Total time spent by all reduces in occupied slots (ms)=17680  
Total time spent by all map tasks (ms)=3655  
Total time spent by all reduce tasks (ms)=3536  
Total vcore-milliseconds taken by all map tasks=3655  
Total vcore-milliseconds taken by all reduce tasks=3536  
Total megabyte-milliseconds taken by all map tasks=18713600  
Total megabyte-milliseconds taken by all reduce tasks=18104320

Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=37  
CPU time spent (ms)=4040  
Physical memory (bytes) snapshot=734494720  
Virtual memory (bytes) snapshot=12758220800  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401133568  
Peak Map Virtual memory (bytes)=6371966976  
Peak Reduce Physical memory (bytes)=333361152  
Peak Reduce Virtual memory (bytes)=6386253824

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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:43:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:43:05 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0583  
23/01/04 15:43:05 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:43:05 INFO JobSubmitter: number of splits:1  
23/01/04 15:43:05 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0583  
23/01/04 15:43:05 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:43:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0583  
23/01/04 15:43:05 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0583/  
23/01/04 15:43:05 INFO Job: Running job: job\_1672817468245\_0583  
23/01/04 15:43:11 INFO Job: Job job\_1672817468245\_0583 running in uber mode : false  
23/01/04 15:43:11 INFO Job: map 0% reduce 0%  
23/01/04 15:43:17 INFO Job: map 100% reduce 0%  
23/01/04 15:43:23 INFO Job: map 100% reduce 100%  
23/01/04 15:43:23 INFO Job: Job job\_1672817468245\_0583 completed successfully  
23/01/04 15:43:23 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17950  
Total time spent by all reduces in occupied slots (ms)=17995  
Total time spent by all map tasks (ms)=3590  
Total time spent by all reduce tasks (ms)=3599  
Total vcore-milliseconds taken by all map tasks=3590  
Total vcore-milliseconds taken by all reduce tasks=3599  
Total megabyte-milliseconds taken by all map tasks=18380800  
Total megabyte-milliseconds taken by all reduce tasks=18426880

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18

Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=4080  
Physical memory (bytes) snapshot=724475904  
Virtual memory (bytes) snapshot=12754231296  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408903680  
Peak Map Virtual memory (bytes)=6370684928  
Peak Reduce Physical memory (bytes)=315572224  
Peak Reduce Virtual memory (bytes)=6383546368

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=2680027

File Output Format Counters

Bytes Written=1843288

23/01/04 15:43:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:43:23 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0586  
23/01/04 15:43:23 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:43:24 INFO JobSubmitter: number of splits:1  
23/01/04 15:43:24 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0586  
23/01/04 15:43:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:43:24 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0586  
23/01/04 15:43:24 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0586/](http://lena-master:8088/proxy/application_1672817468245_0586/)  
23/01/04 15:43:24 INFO Job: Running job: job\_1672817468245\_0586  
23/01/04 15:43:30 INFO Job: Job job\_1672817468245\_0586 running in uber mode : false  
23/01/04 15:43:30 INFO Job: map 0% reduce 0%  
23/01/04 15:43:36 INFO Job: map 100% reduce 0%  
23/01/04 15:43:42 INFO Job: map 100% reduce 100%  
23/01/04 15:43:42 INFO Job: Job job\_1672817468245\_0586 completed successfully  
23/01/04 15:43:42 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=17685  
Total time spent by all reduces in occupied slots (ms)=17525  
Total time spent by all map tasks (ms)=3537  
Total time spent by all reduce tasks (ms)=3505  
Total vcore-milliseconds taken by all map tasks=3537  
Total vcore-milliseconds taken by all reduce tasks=3505  
Total megabyte-milliseconds taken by all map tasks=18109440  
Total megabyte-milliseconds taken by all reduce tasks=17945600

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=66  
CPU time spent (ms)=3950  
Physical memory (bytes) snapshot=731746304  
Virtual memory (bytes) snapshot=12766183424  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=420683776  
Peak Map Virtual memory (bytes)=6380924928  
Peak Reduce Physical memory (bytes)=311062528  
Peak Reduce Virtual memory (bytes)=6385258496

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:43:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:43:42 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0589  
23/01/04 15:43:42 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:43:42 INFO JobSubmitter: number of splits:1  
23/01/04 15:43:42 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0589  
23/01/04 15:43:42 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:43:42 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0589  
23/01/04 15:43:42 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0589/](http://lena-master:8088/proxy/application_1672817468245_0589/)  
23/01/04 15:43:42 INFO Job: Running job: job\_1672817468245\_0589  
23/01/04 15:43:48 INFO Job: Job job\_1672817468245\_0589 running in uber mode : false

23/01/04 15:43:48 INFO Job: map 0% reduce 0%  
23/01/04 15:43:54 INFO Job: map 100% reduce 0%  
23/01/04 15:43:59 INFO Job: map 100% reduce 100%  
23/01/04 15:44:00 INFO Job: Job job\_1672817468245\_0589 completed successfully  
23/01/04 15:44:00 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18010  
Total time spent by all reduces in occupied slots (ms)=17775  
Total time spent by all map tasks (ms)=3602  
Total time spent by all reduce tasks (ms)=3555  
Total vcore-milliseconds taken by all map tasks=3602  
Total vcore-milliseconds taken by all reduce tasks=3555  
Total megabyte-milliseconds taken by all map tasks=18442240  
Total megabyte-milliseconds taken by all reduce tasks=18201600

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=102  
CPU time spent (ms)=4140  
Physical memory (bytes) snapshot=700424192  
Virtual memory (bytes) snapshot=12755656704  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402800640  
Peak Map Virtual memory (bytes)=6374264832  
Peak Reduce Physical memory (bytes)=297623552  
Peak Reduce Virtual memory (bytes)=6381391872

#### 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=2680027  
File Output Format Counters  
Bytes Written=1843288  
23/01/04 15:44:01 INFO DefaultNoHARMAiloverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:44:01 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0592  
23/01/04 15:44:01 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:44:01 INFO JobSubmitter: number of splits:1  
23/01/04 15:44:01 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0592  
23/01/04 15:44:01 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:44:01 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0592  
23/01/04 15:44:01 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0592/](http://lena-master:8088/proxy/application_1672817468245_0592/)  
23/01/04 15:44:01 INFO Job: Running job: job\_1672817468245\_0592  
23/01/04 15:44:07 INFO Job: Job job\_1672817468245\_0592 running in uber mode : false  
23/01/04 15:44:07 INFO Job: map 0% reduce 0%  
23/01/04 15:44:13 INFO Job: map 100% reduce 0%  
23/01/04 15:44:17 INFO Job: map 100% reduce 100%  
23/01/04 15:44:17 INFO Job: Job job\_1672817468245\_0592 completed successfully  
23/01/04 15:44:17 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18630  
Total time spent by all reduces in occupied slots (ms)=10635  
Total time spent by all map tasks (ms)=3726  
Total time spent by all reduce tasks (ms)=2127  
Total vcore-milliseconds taken by all map tasks=3726  
Total vcore-milliseconds taken by all reduce tasks=2127  
Total megabyte-milliseconds taken by all map tasks=19077120  
Total megabyte-milliseconds taken by all reduce tasks=10890240

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9

Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=87  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=729927680  
Virtual memory (bytes) snapshot=12757942272  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=412635136  
Peak Map Virtual memory (bytes)=6370902016  
Peak Reduce Physical memory (bytes)=317292544  
Peak Reduce Virtual memory (bytes)=6387040256

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=2680027

File Output Format Counters

Bytes Written=1843288

23/01/04 15:44:17 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:44:17 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0594  
23/01/04 15:44:17 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:44:17 INFO JobSubmitter: number of splits:1  
23/01/04 15:44:17 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0594  
23/01/04 15:44:17 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:44:17 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0594  
23/01/04 15:44:17 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0594/  
23/01/04 15:44:17 INFO Job: Running job: job\_1672817468245\_0594  
23/01/04 15:44:23 INFO Job: Job job\_1672817468245\_0594 running in uber mode : false  
23/01/04 15:44:23 INFO Job: map 0% reduce 0%  
23/01/04 15:44:29 INFO Job: map 100% reduce 0%  
23/01/04 15:44:33 INFO Job: map 100% reduce 100%  
23/01/04 15:44:33 INFO Job: Job job\_1672817468245\_0594 completed successfully  
23/01/04 15:44:34 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0



#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18120  
Total time spent by all reduces in occupied slots (ms)=10200  
Total time spent by all map tasks (ms)=3624  
Total time spent by all reduce tasks (ms)=2040  
Total vcore-milliseconds taken by all map tasks=3624  
Total vcore-milliseconds taken by all reduce tasks=2040  
Total megabyte-milliseconds taken by all map tasks=18554880  
Total megabyte-milliseconds taken by all reduce tasks=10444800

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=4250  
Physical memory (bytes) snapshot=728264704  
Virtual memory (bytes) snapshot=12752437248  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413065216  
Peak Map Virtual memory (bytes)=6377205760  
Peak Reduce Physical memory (bytes)=315199488  
Peak Reduce Virtual memory (bytes)=6375231488

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:44:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:44:34 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0597

23/01/04 15:44:34 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:44:34 INFO JobSubmitter: number of splits:1

23/01/04 15:44:34 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0597

23/01/04 15:44:34 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:44:34 INFO YarnClientImpl: Submitted application application\_1672817468245\_0597

23/01/04 15:44:34 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0597/  
23/01/04 15:44:34 INFO Job: Running job: job\_1672817468245\_0597  
23/01/04 15:44:40 INFO Job: Job job\_1672817468245\_0597 running in uber mode : false  
23/01/04 15:44:40 INFO Job: map 0% reduce 0%  
23/01/04 15:44:51 INFO JobSubmitter: number of splits:1  
23/01/04 15:44:51 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0600  
23/01/04 15:44:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:44:52 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0600  
23/01/04 15:44:52 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0600/  
23/01/04 15:44:52 INFO Job: Running job: job\_1672817468245\_0600  
23/01/04 15:44:58 INFO Job: Job job\_1672817468245\_0600 running in uber mode : false  
23/01/04 15:44:58 INFO Job: map 0% reduce 0%  
23/01/04 15:45:03 INFO Job: map 100% reduce 0%  
23/01/04 15:45:09 INFO Job: map 100% reduce 100%  
23/01/04 15:45:09 INFO Job: Job job\_1672817468245\_0600 completed successfully  
23/01/04 15:45:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17785  
Total time spent by all reduces in occupied slots (ms)=17495  
Total time spent by all map tasks (ms)=3557  
Total time spent by all reduce tasks (ms)=3499  
Total vcore-milliseconds taken by all map tasks=3557  
Total vcore-milliseconds taken by all reduce tasks=3499  
Total megabyte-milliseconds taken by all map tasks=18211840  
Total megabyte-milliseconds taken by all reduce tasks=17914880

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0

```

Merged Map outputs=1
GC time elapsed (ms)=58
CPU time spent (ms)=4200
Physical memory (bytes) snapshot=728543232
Virtual memory (bytes) snapshot=12770328576
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=421433344
Peak Map Virtual memory (bytes)=6382399488
Peak Reduce Physical memory (bytes)=307109888
Peak Reduce Virtual memory (bytes)=6387929088
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=2680027
File Output Format Counters
  Bytes Written=1843288
23/01/04 15:45:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:45:09 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0602
23/01/04 15:45:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:45:09 INFO JobSubmitter: number of splits:1
23/01/04 15:45:09 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0602
23/01/04 15:45:09 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:45:09 INFO YarnClientImpl: Submitted application
application_1672817468245_0602
23/01/04 15:45:09 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0602/
23/01/04 15:45:09 INFO Job: Running job: job_1672817468245_0602
23/01/04 15:45:16 INFO Job: Job job_1672817468245_0602 running in uber mode : false
23/01/04 15:45:16 INFO Job:  map 0% reduce 0%
23/01/04 15:45:21 INFO Job:  map 100% reduce 0%
23/01/04 15:45:26 INFO Job:  map 100% reduce 100%
23/01/04 15:45:26 INFO Job: Job job_1672817468245_0602 completed successfully
23/01/04 15:45:26 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4376884
  FILE: Number of bytes written=9286067
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6367161
  HDFS: Number of bytes written=1843288
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18405
  Total time spent by all reduces in occupied slots (ms)=11535

```

Total time spent by all map tasks (ms)=3681  
Total time spent by all reduce tasks (ms)=2307  
Total vcore-milliseconds taken by all map tasks=3681  
Total vcore-milliseconds taken by all reduce tasks=2307  
Total megabyte-milliseconds taken by all map tasks=18846720  
Total megabyte-milliseconds taken by all reduce tasks=11811840

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=4240  
Physical memory (bytes) snapshot=723214336  
Virtual memory (bytes) snapshot=12761088000  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403521536  
Peak Map Virtual memory (bytes)=6373576704  
Peak Reduce Physical memory (bytes)=319692800  
Peak Reduce Virtual memory (bytes)=6387511296

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:45:26 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:45:26 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0605  
23/01/04 15:45:26 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:45:26 INFO JobSubmitter: number of splits:1  
23/01/04 15:45:26 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0605  
23/01/04 15:45:26 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:45:26 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0605  
23/01/04 15:45:26 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0605/  
23/01/04 15:45:26 INFO Job: Running job: job\_1672817468245\_0605  
23/01/04 15:45:34 INFO Job: Job job\_1672817468245\_0605 running in uber mode : false  
23/01/04 15:45:34 INFO Job: map 0% reduce 0%  
23/01/04 15:45:39 INFO Job: map 100% reduce 0%

23/01/04 15:45:45 INFO Job: map 100% reduce 100%  
23/01/04 15:45:45 INFO Job: Job job\_1672817468245\_0605 completed successfully  
23/01/04 15:45:45 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17990  
Total time spent by all reduces in occupied slots (ms)=17370  
Total time spent by all map tasks (ms)=3598  
Total time spent by all reduce tasks (ms)=3474  
Total vcore-milliseconds taken by all map tasks=3598  
Total vcore-milliseconds taken by all reduce tasks=3474  
Total megabyte-milliseconds taken by all map tasks=18421760  
Total megabyte-milliseconds taken by all reduce tasks=17786880

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=3990  
Physical memory (bytes) snapshot=726306816  
Virtual memory (bytes) snapshot=12754436096  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413519872  
Peak Map Virtual memory (bytes)=6374711296  
Peak Reduce Physical memory (bytes)=312786944  
Peak Reduce Virtual memory (bytes)=6379724800

#### 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=2680027

#### File Output Format Counters

Bytes Written=1843288

23/01/04 15:45:45 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:45:45 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0608  
23/01/04 15:45:45 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:45:45 INFO JobSubmitter: number of splits:1  
23/01/04 15:45:45 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0608  
23/01/04 15:45:45 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:45:45 INFO YarnClientImpl: Submitted application application\_1672817468245\_0608  
23/01/04 15:45:45 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0608/  
23/01/04 15:45:45 INFO Job: Running job: job\_1672817468245\_0608  
23/01/04 15:45:51 INFO Job: Job job\_1672817468245\_0608 running in uber mode : false  
23/01/04 15:45:51 INFO Job: map 0% reduce 0%  
23/01/04 15:45:57 INFO Job: map 100% reduce 0%  
23/01/04 15:46:02 INFO Job: map 100% reduce 100%  
23/01/04 15:46:03 INFO Job: Job job\_1672817468245\_0608 completed successfully  
23/01/04 15:46:03 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4376884  
FILE: Number of bytes written=9286067  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6367161  
HDFS: Number of bytes written=1843288  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17985  
Total time spent by all reduces in occupied slots (ms)=17515  
Total time spent by all map tasks (ms)=3597  
Total time spent by all reduce tasks (ms)=3503  
Total vcore-milliseconds taken by all map tasks=3597  
Total vcore-milliseconds taken by all reduce tasks=3503  
Total megabyte-milliseconds taken by all map tasks=18416640  
Total megabyte-milliseconds taken by all reduce tasks=17935360

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4376833  
Map output materialized bytes=4376884  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4376884  
Reduce input records=9

```

Reduce output records=9
Spilled Records=18
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=61
CPU time spent (ms)=4250
Physical memory (bytes) snapshot=726179840
Virtual memory (bytes) snapshot=12763369472
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=397451264
Peak Map Virtual memory (bytes)=6369906688
Peak Reduce Physical memory (bytes)=328728576
Peak Reduce Virtual memory (bytes)=6393462784

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=2680027
File Output Format Counters
  Bytes Written=1843288
23/01/04 15:46:03 INFO KMeansDriver: Clustering data
23/01/04 15:46:03 INFO KMeansDriver: Running Clustering
23/01/04 15:46:03 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/09 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/09
23/01/04 15:46:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:46:03 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0611
23/01/04 15:46:03 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:46:03 INFO JobSubmitter: number of splits:1
23/01/04 15:46:04 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0611
23/01/04 15:46:04 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:46:04 INFO YarnClientImpl: Submitted application
application_1672817468245_0611
23/01/04 15:46:04 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0611/
23/01/04 15:46:04 INFO Job: Running job: job_1672817468245_0611
23/01/04 15:46:10 INFO Job: Job job_1672817468245_0611 running in uber mode : false
23/01/04 15:46:10 INFO Job:  map 0% reduce 0%
23/01/04 15:46:16 INFO Job:  map 100% reduce 0%
23/01/04 15:46:16 INFO Job: Job job_1672817468245_0611 completed successfully
23/01/04 15:46:16 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265898
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4523679

```

HDFS: Number of bytes written=2680628  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17990  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3598  
Total vcore-milliseconds taken by all map tasks=3598  
Total megabyte-milliseconds taken by all map tasks=18421760

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=48  
CPU time spent (ms)=2290  
Physical memory (bytes) snapshot=289918976  
Virtual memory (bytes) snapshot=6382297088  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=289918976  
Peak Map Virtual memory (bytes)=6382297088

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680628

23/01/04 15:46:16 INFO MahoutDriver: Program took 373702 ms (Minutes: 6.228366666666667)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:46:22 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --  
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/EuclideanDistanceMeasure/09/clusters-20-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
EuclideanDistanceMeasure/09.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu  
re/09/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 15:46:24 INFO HadoopUtil: Deleting tmp/representative

23/01/04 15:46:24 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/

EuclideanDistanceMeasure/09/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --

endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/  
09/clusters-20-final], --maxIter=[5], --method=[mapreduce],

--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 15:46:24 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 15:46:24 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to



ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:46:24 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0615  
23/01/04 15:46:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:46:25 INFO JobSubmitter: number of splits:1  
23/01/04 15:46:25 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0615  
23/01/04 15:46:25 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:46:25 INFO Configuration: resource-types.xml not found  
23/01/04 15:46:25 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 15:46:25 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0615  
23/01/04 15:46:25 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0615/  
23/01/04 15:46:25 INFO Job: Running job: job\_1672817468245\_0615  
23/01/04 15:46:31 INFO Job: Job job\_1672817468245\_0615 running in uber mode : false  
23/01/04 15:46:31 INFO Job: map 0% reduce 0%  
23/01/04 15:46:36 INFO Job: map 100% reduce 0%  
23/01/04 15:46:42 INFO Job: map 100% reduce 100%  
23/01/04 15:46:42 INFO Job: Job job\_1672817468245\_0615 completed successfully  
23/01/04 15:46:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=935641  
FILE: Number of bytes written=2403875  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5215421  
HDFS: Number of bytes written=2203006  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16835  
Total time spent by all reduces in occupied slots (ms)=16525  
Total time spent by all map tasks (ms)=3367  
Total time spent by all reduce tasks (ms)=3305  
Total vcore-milliseconds taken by all map tasks=3367  
Total vcore-milliseconds taken by all reduce tasks=3305  
Total megabyte-milliseconds taken by all map tasks=17239040  
Total megabyte-milliseconds taken by all reduce tasks=16921600

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=935591  
Map output materialized bytes=935641  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=935641  
Reduce input records=9  
Reduce output records=18  
Spilled Records=18  
Shuffled Maps =1

Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=2810  
Physical memory (bytes) snapshot=687693824  
Virtual memory (bytes) snapshot=12756697088  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394293248  
Peak Map Virtual memory (bytes)=6377680896  
Peak Reduce Physical memory (bytes)=293400576  
Peak Reduce Virtual memory (bytes)=6380969984

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=2680628

File Output Format Counters

Bytes Written=2203006

23/01/04 15:46:42 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 15:46:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:46:42 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0618

23/01/04 15:46:42 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:46:42 INFO JobSubmitter: number of splits:1

23/01/04 15:46:42 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0618

23/01/04 15:46:42 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:46:43 INFO YarnClientImpl: Submitted application application\_1672817468245\_0618

23/01/04 15:46:43 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0618/](http://lena-master:8088/proxy/application_1672817468245_0618/)

23/01/04 15:46:43 INFO Job: Running job: job\_1672817468245\_0618

23/01/04 15:46:49 INFO Job: Job job\_1672817468245\_0618 running in uber mode : false

23/01/04 15:46:49 INFO Job: map 0% reduce 0%

23/01/04 15:46:54 INFO Job: map 100% reduce 0%

23/01/04 15:47:00 INFO Job: map 100% reduce 100%

23/01/04 15:47:00 INFO Job: Job job\_1672817468245\_0618 completed successfully

23/01/04 15:47:00 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=981777

FILE: Number of bytes written=2496147

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=7086843

HDFS: Number of bytes written=3184873

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17205  
Total time spent by all reduces in occupied slots (ms)=16545  
Total time spent by all map tasks (ms)=3441  
Total time spent by all reduce tasks (ms)=3309  
Total vcore-milliseconds taken by all map tasks=3441  
Total vcore-milliseconds taken by all reduce tasks=3309  
Total megabyte-milliseconds taken by all map tasks=17617920  
Total megabyte-milliseconds taken by all reduce tasks=16942080

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=981727  
Map output materialized bytes=981777  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=981777  
Reduce input records=9  
Reduce output records=27  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3310  
Physical memory (bytes) snapshot=687681536  
Virtual memory (bytes) snapshot=12750471168  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405479424  
Peak Map Virtual memory (bytes)=6370185216  
Peak Reduce Physical memory (bytes)=285773824  
Peak Reduce Virtual memory (bytes)=6384197632

#### 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=2680628

#### File Output Format Counters

Bytes Written=3184873

23/01/04 15:47:00 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 15:47:00 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:47:00 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0620

23/01/04 15:47:00 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:47:00 INFO JobSubmitter: number of splits:1

23/01/04 15:47:00 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0620

23/01/04 15:47:00 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:47:00 INFO YarnClientImpl: Submitted application application\_1672817468245\_0620

23/01/04 15:47:00 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0620/  
23/01/04 15:47:00 INFO Job: Running job: job\_1672817468245\_0620  
23/01/04 15:47:06 INFO Job: Job job\_1672817468245\_0620 running in uber mode : false  
23/01/04 15:47:06 INFO Job: map 0% reduce 0%  
23/01/04 15:47:12 INFO Job: map 100% reduce 0%  
23/01/04 15:47:18 INFO Job: map 100% reduce 100%  
23/01/04 15:47:18 INFO Job: Job job\_1672817468245\_0620 completed successfully  
23/01/04 15:47:18 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=915039  
FILE: Number of bytes written=2362671  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9050577  
HDFS: Number of bytes written=4100002  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16970  
Total time spent by all reduces in occupied slots (ms)=16585  
Total time spent by all map tasks (ms)=3394  
Total time spent by all reduce tasks (ms)=3317  
Total vcore-milliseconds taken by all map tasks=3394  
Total vcore-milliseconds taken by all reduce tasks=3317  
Total megabyte-milliseconds taken by all map tasks=17377280  
Total megabyte-milliseconds taken by all reduce tasks=16983040

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=914989  
Map output materialized bytes=915039  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=915039  
Reduce input records=9  
Reduce output records=36  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=32  
CPU time spent (ms)=3340  
Physical memory (bytes) snapshot=687362048  
Virtual memory (bytes) snapshot=12756230144  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392904704  
Peak Map Virtual memory (bytes)=6376574976  
Peak Reduce Physical memory (bytes)=294457344  
Peak Reduce Virtual memory (bytes)=6379655168

#### 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=2680628
File Output Format Counters
    Bytes Written=4100002
23/01/04 15:47:18 INFO RepresentativePointsDriver: Representative Points Iteration
3
23/01/04 15:47:18 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:47:18 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0624
23/01/04 15:47:19 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:47:19 INFO JobSubmitter: number of splits:1
23/01/04 15:47:19 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0624
23/01/04 15:47:19 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:47:19 INFO YarnClientImpl: Submitted application
application_1672817468245_0624
23/01/04 15:47:19 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0624/
23/01/04 15:47:19 INFO Job: Running job: job_1672817468245_0624
23/01/04 15:47:25 INFO Job: Job job_1672817468245_0624 running in uber mode : false
23/01/04 15:47:25 INFO Job:  map 0% reduce 0%
23/01/04 15:47:31 INFO Job:  map 100% reduce 0%
23/01/04 15:47:37 INFO Job:  map 100% reduce 100%
23/01/04 15:47:37 INFO Job: Job job_1672817468245_0624 completed successfully
23/01/04 15:47:37 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=965714
    FILE: Number of bytes written=2464021
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=10880835
    HDFS: Number of bytes written=5065806
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17320
    Total time spent by all reduces in occupied slots (ms)=16560
    Total time spent by all map tasks (ms)=3464
    Total time spent by all reduce tasks (ms)=3312
    Total vcore-milliseconds taken by all map tasks=3464
    Total vcore-milliseconds taken by all reduce tasks=3312
    Total megabyte-milliseconds taken by all map tasks=17735680
    Total megabyte-milliseconds taken by all reduce tasks=16957440
Map-Reduce Framework
    Map input records=27
    Map output records=9
    Map output bytes=965664

```

Map output materialized bytes=965714  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=965714  
Reduce input records=9  
Reduce output records=45  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=3630  
Physical memory (bytes) snapshot=696913920  
Virtual memory (bytes) snapshot=12754923520  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400334848  
Peak Map Virtual memory (bytes)=6376599552  
Peak Reduce Physical memory (bytes)=296579072  
Peak Reduce Virtual memory (bytes)=6378323968

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=2680628

File Output Format Counters

Bytes Written=5065806

23/01/04 15:47:37 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:47:37 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:47:37 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0625

23/01/04 15:47:37 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:47:37 INFO JobSubmitter: number of splits:1

23/01/04 15:47:37 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0625

23/01/04 15:47:37 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:47:37 INFO YarnClientImpl: Submitted application application\_1672817468245\_0625

23/01/04 15:47:37 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0625/

23/01/04 15:47:37 INFO Job: Running job: job\_1672817468245\_0625

23/01/04 15:47:43 INFO Job: Job job\_1672817468245\_0625 running in uber mode : false

23/01/04 15:47:43 INFO Job: map 0% reduce 0%

23/01/04 15:47:49 INFO Job: map 100% reduce 0%

23/01/04 15:47:55 INFO Job: map 100% reduce 100%

23/01/04 15:47:55 INFO Job: Job job\_1672817468245\_0625 completed successfully

23/01/04 15:47:55 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=945136  
FILE: Number of bytes written=2422865  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=12812443  
HDFS: Number of bytes written=6011032  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17475  
Total time spent by all reduces in occupied slots (ms)=16785  
Total time spent by all map tasks (ms)=3495  
Total time spent by all reduce tasks (ms)=3357  
Total vcore-milliseconds taken by all map tasks=3495  
Total vcore-milliseconds taken by all reduce tasks=3357  
Total megabyte-milliseconds taken by all map tasks=17894400  
Total megabyte-milliseconds taken by all reduce tasks=17187840

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=945086  
Map output materialized bytes=945136  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=945136  
Reduce input records=9  
Reduce output records=54  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=72  
CPU time spent (ms)=3660  
Physical memory (bytes) snapshot=692457472  
Virtual memory (bytes) snapshot=12749242368  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408543232  
Peak Map Virtual memory (bytes)=6368755712  
Peak Reduce Physical memory (bytes)=287055872  
Peak Reduce Virtual memory (bytes)=6380486656

#### 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=2680628

#### File Output Format Counters

Bytes Written=6011032

23/01/04 15:47:56 INFO ClusterEvaluator: Scaled Inter-Cluster Density = 0.61522289489521

23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[11] = NaN

23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[9] = NaN

```

23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[16] = NaN
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[3] = NaN
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.6294363421102389
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[6] =
0.6222699678562017
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[23] = NaN
23/01/04 15:47:56 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN
23/01/04 15:47:56 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6258531549832203
23/01/04 15:47:57 INFO ClusterDumper: Wrote 9 clusters
23/01/04 15:47:57 INFO MahoutDriver: Program took 94170 ms (Minutes: 1.5695)
---- K: 10 --
dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/10
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:48:01 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
EuclideanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[10],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/EuclideanDistanceMeasure/10], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:48:02 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/10
23/01/04 15:48:02 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/EuclideanDistanceMeasure
23/01/04 15:48:02 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:48:02 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:48:02 INFO RandomSeedGenerator: Wrote 10 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/EuclideanDistanceMeasure/
part-randomSeed
23/01/04 15:48:02 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/EuclideanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/10
23/01/04 15:48:02 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:48:02 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:48:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:48:03 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0630
23/01/04 15:48:03 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:48:03 INFO JobSubmitter: number of splits:1
23/01/04 15:48:03 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0630
23/01/04 15:48:03 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:48:03 INFO Configuration: resource-types.xml not found
23/01/04 15:48:03 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:48:03 INFO YarnClientImpl: Submitted application

```



application\_1672817468245\_0630

23/01/04 15:48:03 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0630/](http://lena-master:8088/proxy/application_1672817468245_0630/)

23/01/04 15:48:03 INFO Job: Running job: job\_1672817468245\_0630

23/01/04 15:48:09 INFO Job: Job job\_1672817468245\_0630 running in uber mode : false

23/01/04 15:48:09 INFO Job: map 0% reduce 0%

23/01/04 15:48:15 INFO Job: map 100% reduce 0%

23/01/04 15:48:21 INFO Job: map 100% reduce 100%

23/01/04 15:48:22 INFO Job: Job job\_1672817468245\_0630 completed successfully

23/01/04 15:48:22 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3649112

FILE: Number of bytes written=7830515

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=7766097

HDFS: Number of bytes written=2256628

HDFS: Number of read operations=55

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=18605

Total time spent by all reduces in occupied slots (ms)=17980

Total time spent by all map tasks (ms)=3721

Total time spent by all reduce tasks (ms)=3596

Total vcore-milliseconds taken by all map tasks=3721

Total vcore-milliseconds taken by all reduce tasks=3596

Total megabyte-milliseconds taken by all map tasks=19051520

Total megabyte-milliseconds taken by all reduce tasks=18411520

#### Map-Reduce Framework

Map input records=27

Map output records=10

Map output bytes=3649056

Map output materialized bytes=3649112

Input split bytes=170

Combine input records=0

Combine output records=0

Reduce input groups=10

Reduce shuffle bytes=3649112

Reduce input records=10

Reduce output records=10

Spilled Records=20

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=65

CPU time spent (ms)=4590

Physical memory (bytes) snapshot=719773696

Virtual memory (bytes) snapshot=12751486976

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=410390528

Peak Map Virtual memory (bytes)=6380318720

Peak Reduce Physical memory (bytes)=309383168

Peak Reduce Virtual memory (bytes)=6371168256

#### 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=2680027

#### File Output Format Counters

Bytes Written=2256628

23/01/04 15:48:22 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:48:22 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0633  
23/01/04 15:48:22 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:48:22 INFO JobSubmitter: number of splits:1  
23/01/04 15:48:22 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0633  
23/01/04 15:48:22 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:48:22 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0633  
23/01/04 15:48:22 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0633/](http://lena-master:8088/proxy/application_1672817468245_0633/)  
23/01/04 15:48:22 INFO Job: Running job: job\_1672817468245\_0633  
23/01/04 15:48:28 INFO Job: Job job\_1672817468245\_0633 running in uber mode : false  
23/01/04 15:48:28 INFO Job: map 0% reduce 0%  
23/01/04 15:48:34 INFO Job: map 100% reduce 0%  
23/01/04 15:48:41 INFO Job: map 100% reduce 100%  
23/01/04 15:48:41 INFO Job: Job job\_1672817468245\_0633 completed successfully  
23/01/04 15:48:41 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5023133  
FILE: Number of bytes written=10578557  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7193841  
HDFS: Number of bytes written=2242465  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18185  
Total time spent by all reduces in occupied slots (ms)=17720  
Total time spent by all map tasks (ms)=3637  
Total time spent by all reduce tasks (ms)=3544  
Total vcore-milliseconds taken by all map tasks=3637  
Total vcore-milliseconds taken by all reduce tasks=3544  
Total megabyte-milliseconds taken by all map tasks=18621440  
Total megabyte-milliseconds taken by all reduce tasks=18145280

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5023077

Map output materialized bytes=5023133  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5023133  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=4280  
Physical memory (bytes) snapshot=719437824  
Virtual memory (bytes) snapshot=12762378240  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=388907008  
Peak Map Virtual memory (bytes)=6379499520  
Peak Reduce Physical memory (bytes)=330530816  
Peak Reduce Virtual memory (bytes)=6382878720

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=2680027

File Output Format Counters

Bytes Written=2242465

23/01/04 15:48:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:48:41 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0636  
23/01/04 15:48:42 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:48:42 INFO JobSubmitter: number of splits:1  
23/01/04 15:48:42 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0636  
23/01/04 15:48:42 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:48:42 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0636  
23/01/04 15:48:42 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0636/](http://lena-master:8088/proxy/application_1672817468245_0636/)  
23/01/04 15:48:42 INFO Job: Running job: job\_1672817468245\_0636  
23/01/04 15:48:48 INFO Job: Job job\_1672817468245\_0636 running in uber mode : false  
23/01/04 15:48:48 INFO Job: map 0% reduce 0%  
23/01/04 15:48:54 INFO Job: map 100% reduce 0%  
23/01/04 15:49:00 INFO Job: map 100% reduce 100%  
23/01/04 15:49:00 INFO Job: Job job\_1672817468245\_0636 completed successfully  
23/01/04 15:49:00 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4889332  
FILE: Number of bytes written=10310955  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7165515

HDFS: Number of bytes written=2192571  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18150  
Total time spent by all reduces in occupied slots (ms)=17935  
Total time spent by all map tasks (ms)=3630  
Total time spent by all reduce tasks (ms)=3587  
Total vcore-milliseconds taken by all map tasks=3630  
Total vcore-milliseconds taken by all reduce tasks=3587  
Total megabyte-milliseconds taken by all map tasks=18585600  
Total megabyte-milliseconds taken by all reduce tasks=18365440

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=4889276  
Map output materialized bytes=4889332  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=4889332  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=4160  
Physical memory (bytes) snapshot=773603328  
Virtual memory (bytes) snapshot=12769267712  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=434946048  
Peak Map Virtual memory (bytes)=6373343232  
Peak Reduce Physical memory (bytes)=338657280  
Peak Reduce Virtual memory (bytes)=6395924480

#### 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=2680027

#### File Output Format Counters

Bytes Written=2192571

23/01/04 15:49:00 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:49:00 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0639  
23/01/04 15:49:00 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:49:00 INFO JobSubmitter: number of splits:1

23/01/04 15:49:00 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0639  
23/01/04 15:49:00 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:49:00 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0639  
23/01/04 15:49:00 INFO Job: The url to track the job:  
http://lana-master:8088/proxy/application\_1672817468245\_0639/  
23/01/04 15:49:00 INFO Job: Running job: job\_1672817468245\_0639  
23/01/04 15:49:06 INFO Job: Job job\_1672817468245\_0639 running in uber mode : false  
23/01/04 15:49:06 INFO Job: map 0% reduce 0%  
23/01/04 15:49:12 INFO Job: map 100% reduce 0%  
23/01/04 15:49:16 INFO Job: map 100% reduce 100%  
23/01/04 15:49:16 INFO Job: Job job\_1672817468245\_0639 completed successfully  
23/01/04 15:49:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4839438  
FILE: Number of bytes written=10211167  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7065727  
HDFS: Number of bytes written=2192571  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18350  
Total time spent by all reduces in occupied slots (ms)=10405  
Total time spent by all map tasks (ms)=3670  
Total time spent by all reduce tasks (ms)=2081  
Total vcore-milliseconds taken by all map tasks=3670  
Total vcore-milliseconds taken by all reduce tasks=2081  
Total megabyte-milliseconds taken by all map tasks=18790400  
Total megabyte-milliseconds taken by all reduce tasks=10654720

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=4839382  
Map output materialized bytes=4839438  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=4839438  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=4420  
Physical memory (bytes) snapshot=735801344  
Virtual memory (bytes) snapshot=12759449600  
Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=429142016  
Peak Map Virtual memory (bytes)=6375186432  
Peak Reduce Physical memory (bytes)=306659328  
Peak Reduce Virtual memory (bytes)=6384263168

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=2680027

File Output Format Counters

Bytes Written=2192571

23/01/04 15:49:16 INFO KMeansDriver: Clustering data

23/01/04 15:49:16 INFO KMeansDriver: Running Clustering

23/01/04 15:49:16 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:

hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/10 Out:

hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
EuclideanDistanceMeasure/10

23/01/04 15:49:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:49:16 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0641

23/01/04 15:49:17 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:49:17 INFO JobSubmitter: number of splits:1

23/01/04 15:49:17 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0641

23/01/04 15:49:17 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:49:17 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0641

23/01/04 15:49:17 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0641/

23/01/04 15:49:17 INFO Job: Running job: job\_1672817468245\_0641

23/01/04 15:49:23 INFO Job: Job job\_1672817468245\_0641 running in uber mode : false

23/01/04 15:49:23 INFO Job: map 0% reduce 0%

23/01/04 15:49:29 INFO Job: map 100% reduce 0%

23/01/04 15:49:29 INFO Job: Job job\_1672817468245\_0641 completed successfully

23/01/04 15:49:29 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4872962  
HDFS: Number of bytes written=2680604  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18035  
Total time spent by all reduces in occupied slots (ms)=0

```

    Total time spent by all map tasks (ms)=3607
    Total vcore-milliseconds taken by all map tasks=3607
    Total megabyte-milliseconds taken by all map tasks=18467840
Map-Reduce Framework
    Map input records=27
    Map output records=27
    Input split bytes=170
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=23
    CPU time spent (ms)=2520
    Physical memory (bytes) snapshot=316407808
    Virtual memory (bytes) snapshot=6373548032
    Total committed heap usage (bytes)=1052770304
    Peak Map Physical memory (bytes)=316407808
    Peak Map Virtual memory (bytes)=6373548032
File Input Format Counters
    Bytes Read=2680027
File Output Format Counters
    Bytes Written=2680604
23/01/04 15:49:29 INFO MahoutDriver: Program took 87991 ms (Minutes:
1.4665166666666667)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:49:35 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/EuclideanDistanceMeasure/10/clusters-4-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
EuclideanDistanceMeasure/10.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasu
re/10/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:49:36 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:49:36 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
EuclideanDistanceMeasure/10/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.EuclideanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/EuclideanDistanceMeasure/
10/clusters-4-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:49:36 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:49:36 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:49:37 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0644
23/01/04 15:49:37 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:49:37 INFO JobSubmitter: number of splits:1
23/01/04 15:49:37 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0644
23/01/04 15:49:37 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:49:37 INFO Configuration: resource-types.xml not found
23/01/04 15:49:37 INFO ResourceUtils: Unable to find 'resource-types.xml'.

```

23/01/04 15:49:37 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0644  
23/01/04 15:49:37 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0644/  
23/01/04 15:49:37 INFO Job: Running job: job\_1672817468245\_0644  
23/01/04 15:49:43 INFO Job: Job job\_1672817468245\_0644 running in uber mode : false  
23/01/04 15:49:43 INFO Job: map 0% reduce 0%  
23/01/04 15:49:49 INFO Job: map 100% reduce 0%  
23/01/04 15:49:54 INFO Job: map 100% reduce 100%  
23/01/04 15:49:55 INFO Job: Job job\_1672817468245\_0644 completed successfully  
23/01/04 15:49:56 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1134634  
FILE: Number of bytes written=2801861  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5328765  
HDFS: Number of bytes written=2458717  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16610  
Total time spent by all reduces in occupied slots (ms)=16570  
Total time spent by all map tasks (ms)=3322  
Total time spent by all reduce tasks (ms)=3314  
Total vcore-milliseconds taken by all map tasks=3322  
Total vcore-milliseconds taken by all reduce tasks=3314  
Total megabyte-milliseconds taken by all map tasks=17008640  
Total megabyte-milliseconds taken by all reduce tasks=16967680

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1134578  
Map output materialized bytes=1134634  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1134634  
Reduce input records=10  
Reduce output records=20  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=2970  
Physical memory (bytes) snapshot=689782784  
Virtual memory (bytes) snapshot=12763385856  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=395042816  
Peak Map Virtual memory (bytes)=6373822464  
Peak Reduce Physical memory (bytes)=294739968



Peak Reduce Virtual memory (bytes)=6391517184  
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=2680604  
File Output Format Counters  
Bytes Written=2458717

23/01/04 15:49:56 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 15:49:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:49:56 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0647  
23/01/04 15:49:56 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:49:56 INFO JobSubmitter: number of splits:1  
23/01/04 15:49:56 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0647  
23/01/04 15:49:56 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:49:56 INFO YarnClientImpl: Submitted application application\_1672817468245\_0647  
23/01/04 15:49:56 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0647/  
23/01/04 15:49:56 INFO Job: Running job: job\_1672817468245\_0647  
23/01/04 15:50:02 INFO Job: Job job\_1672817468245\_0647 running in uber mode : false  
23/01/04 15:50:02 INFO Job: map 0% reduce 0%  
23/01/04 15:50:08 INFO Job: map 100% reduce 0%  
23/01/04 15:50:13 INFO Job: map 100% reduce 100%  
23/01/04 15:50:13 INFO Job: Job job\_1672817468245\_0647 completed successfully  
23/01/04 15:50:13 INFO Job: Counters: 54

File System Counters  
FILE: Number of bytes read=1128489  
FILE: Number of bytes written=2789571  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7598241  
HDFS: Number of bytes written=3587330  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16765  
Total time spent by all reduces in occupied slots (ms)=16185  
Total time spent by all map tasks (ms)=3353  
Total time spent by all reduce tasks (ms)=3237  
Total vcore-milliseconds taken by all map tasks=3353  
Total vcore-milliseconds taken by all reduce tasks=3237  
Total megabyte-milliseconds taken by all map tasks=17167360  
Total megabyte-milliseconds taken by all reduce tasks=16573440  
Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1128433  
Map output materialized bytes=1128489  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1128489  
Reduce input records=10  
Reduce output records=30  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=3230  
Physical memory (bytes) snapshot=708595712  
Virtual memory (bytes) snapshot=12753764352  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408498176  
Peak Map Virtual memory (bytes)=6377861120  
Peak Reduce Physical memory (bytes)=300097536  
Peak Reduce Virtual memory (bytes)=6379810816

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=2680604

File Output Format Counters

Bytes Written=3587330

23/01/04 15:50:13 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 15:50:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:50:13 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0650  
23/01/04 15:50:13 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:50:13 INFO JobSubmitter: number of splits:1  
23/01/04 15:50:13 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0650  
23/01/04 15:50:13 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:50:13 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0650  
23/01/04 15:50:13 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0650/  
23/01/04 15:50:13 INFO Job: Running job: job\_1672817468245\_0650  
23/01/04 15:50:19 INFO Job: Job job\_1672817468245\_0650 running in uber mode : false  
23/01/04 15:50:19 INFO Job: map 0% reduce 0%  
23/01/04 15:50:25 INFO Job: map 100% reduce 0%  
23/01/04 15:50:30 INFO Job: map 100% reduce 100%  
23/01/04 15:50:31 INFO Job: Job job\_1672817468245\_0650 completed successfully  
23/01/04 15:50:31 INFO Job: Counters: 54  
File System Counters  
FILE: Number of bytes read=1086781

FILE: Number of bytes written=2706155  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9855467  
HDFS: Number of bytes written=4674215  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17175  
Total time spent by all reduces in occupied slots (ms)=16445  
Total time spent by all map tasks (ms)=3435  
Total time spent by all reduce tasks (ms)=3289  
Total vcore-milliseconds taken by all map tasks=3435  
Total vcore-milliseconds taken by all reduce tasks=3289  
Total megabyte-milliseconds taken by all map tasks=17587200  
Total megabyte-milliseconds taken by all reduce tasks=16839680

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1086725  
Map output materialized bytes=1086781  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1086781  
Reduce input records=10  
Reduce output records=40  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=3380  
Physical memory (bytes) snapshot=691249152  
Virtual memory (bytes) snapshot=12759654400  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=409276416  
Peak Map Virtual memory (bytes)=6383857664  
Peak Reduce Physical memory (bytes)=281972736  
Peak Reduce Virtual memory (bytes)=6375796736

#### 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=2680604

#### File Output Format Counters

Bytes Written=4674215

23/01/04 15:50:31 INFO RepresentativePointsDriver: Representative Points Iteration

3

```
23/01/04 15:50:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:50:32 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0654
23/01/04 15:50:32 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:50:32 INFO JobSubmitter: number of splits:1
23/01/04 15:50:32 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0654
23/01/04 15:50:32 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:50:32 INFO YarnClientImpl: Submitted application
application_1672817468245_0654
23/01/04 15:50:32 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0654/
23/01/04 15:50:32 INFO Job: Running job: job_1672817468245_0654
23/01/04 15:50:38 INFO Job: Job job_1672817468245_0654 running in uber mode : false
23/01/04 15:50:38 INFO Job:  map 0% reduce 0%
23/01/04 15:50:44 INFO Job:  map 100% reduce 0%
23/01/04 15:50:50 INFO Job:  map 100% reduce 100%
23/01/04 15:50:50 INFO Job: Job job_1672817468245_0654 completed successfully
23/01/04 15:50:50 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=1116402
FILE: Number of bytes written=2765397
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=12029237
HDFS: Number of bytes written=5790741
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=17025
Total time spent by all reduces in occupied slots (ms)=18040
Total time spent by all map tasks (ms)=3405
Total time spent by all reduce tasks (ms)=3608
Total vcore-milliseconds taken by all map tasks=3405
Total vcore-milliseconds taken by all reduce tasks=3608
Total megabyte-milliseconds taken by all map tasks=17433600
Total megabyte-milliseconds taken by all reduce tasks=18472960
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=10
Map output bytes=1116346
Map output materialized bytes=1116402
Input split bytes=203
Combine input records=0
Combine output records=0
Reduce input groups=10
Reduce shuffle bytes=1116402
Reduce input records=10
Reduce output records=50
Spilled Records=20
Shuffled Maps =1
```

Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=61  
CPU time spent (ms)=3630  
Physical memory (bytes) snapshot=714170368  
Virtual memory (bytes) snapshot=12771856384  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415842304  
Peak Map Virtual memory (bytes)=6383042560  
Peak Reduce Physical memory (bytes)=298328064  
Peak Reduce Virtual memory (bytes)=6388813824

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=2680604

File Output Format Counters

Bytes Written=5790741

23/01/04 15:50:50 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:50:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:50:50 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0656

23/01/04 15:50:50 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:50:50 INFO JobSubmitter: number of splits:1

23/01/04 15:50:50 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0656

23/01/04 15:50:50 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:50:50 INFO YarnClientImpl: Submitted application application\_1672817468245\_0656

23/01/04 15:50:50 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0656/](http://lena-master:8088/proxy/application_1672817468245_0656/)

23/01/04 15:50:50 INFO Job: Running job: job\_1672817468245\_0656

23/01/04 15:50:56 INFO Job: Job job\_1672817468245\_0656 running in uber mode : false

23/01/04 15:50:56 INFO Job: map 0% reduce 0%

23/01/04 15:51:02 INFO Job: map 100% reduce 0%

23/01/04 15:51:06 INFO Job: map 100% reduce 100%

23/01/04 15:51:06 INFO Job: Job job\_1672817468245\_0656 completed successfully

23/01/04 15:51:06 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=1075268  
FILE: Number of bytes written=2683129  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=14262289  
HDFS: Number of bytes written=6866093  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1

Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17805  
Total time spent by all reduces in occupied slots (ms)=9635  
Total time spent by all map tasks (ms)=3561  
Total time spent by all reduce tasks (ms)=1927  
Total vcore-milliseconds taken by all map tasks=3561  
Total vcore-milliseconds taken by all reduce tasks=1927  
Total megabyte-milliseconds taken by all map tasks=18232320  
Total megabyte-milliseconds taken by all reduce tasks=9866240

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1075212  
Map output materialized bytes=1075268  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1075268  
Reduce input records=10  
Reduce output records=60  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=3800  
Physical memory (bytes) snapshot=760164352  
Virtual memory (bytes) snapshot=12755427328  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=431583232  
Peak Map Virtual memory (bytes)=6373822464  
Peak Reduce Physical memory (bytes)=332328960  
Peak Reduce Virtual memory (bytes)=6381604864

#### 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=2680604

#### File Output Format Counters

Bytes Written=6866093

23/01/04 15:51:07 INFO ClusterEvaluator: Scaled Inter-Cluster Density = 0.46670822782180127

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[0] = 0.5980394278108104

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[2] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[10] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[4] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[11] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[24] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN

23/01/04 15:51:07 INFO ClusterEvaluator: Average Intra-Cluster Density =

```

0.5980394278108104
23/01/04 15:51:08 INFO ClusterDumper: Wrote 10 clusters
23/01/04 15:51:08 INFO MahoutDriver: Program took 93118 ms (Minutes:
1.5519666666666667)
--- Distance Metric: org.apache.mahout.common.distance.ManhattanDistanceMeasure
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:51:13 INFO AbstractJob: Command line arguments: {--
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
method=[mapreduce], --output=[dsm010/question_2/british-fiction-corpus-canopy-
centroids/ManhattanDistanceMeasure], --overwrite=null, --startPhase=[0], --
t1=[0.5], --t2=[0.3], --tempDir=[temp]}
23/01/04 15:51:13 INFO CanopyDriver: Build Clusters Input:
dsm010/question_2/british-fiction-corpus-vectors/tf-vectors Out:
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure
Measure: org.apache.mahout.common.distance.ManhattanDistanceMeasure@6accbb21 t1:
0.5 t2: 0.3
23/01/04 15:51:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:51:13 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0659
23/01/04 15:51:14 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:51:14 INFO JobSubmitter: number of splits:1
23/01/04 15:51:14 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0659
23/01/04 15:51:14 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:51:14 INFO Configuration: resource-types.xml not found
23/01/04 15:51:14 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:51:14 INFO YarnClientImpl: Submitted application
application_1672817468245_0659
23/01/04 15:51:14 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0659/
23/01/04 15:51:14 INFO Job: Running job: job_1672817468245_0659
23/01/04 15:51:20 INFO Job: Job job_1672817468245_0659 running in uber mode : false
23/01/04 15:51:20 INFO Job: map 0% reduce 0%
23/01/04 15:51:26 INFO Job: map 100% reduce 0%
23/01/04 15:51:33 INFO Job: map 100% reduce 100%
23/01/04 15:51:33 INFO Job: Job job_1672817468245_0659 completed successfully
23/01/04 15:51:34 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2678621
        FILE: Number of bytes written=5891013
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=2680197
        HDFS: Number of bytes written=2682998
        HDFS: Number of read operations=9
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1

```

Total time spent by all maps in occupied slots (ms)=19815  
Total time spent by all reduces in occupied slots (ms)=20085  
Total time spent by all map tasks (ms)=3963  
Total time spent by all reduce tasks (ms)=4017  
Total vcore-milliseconds taken by all map tasks=3963  
Total vcore-milliseconds taken by all reduce tasks=4017  
Total megabyte-milliseconds taken by all map tasks=20290560  
Total megabyte-milliseconds taken by all reduce tasks=20567040

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Map output bytes=2678485  
Map output materialized bytes=2678621  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=2678621  
Reduce input records=27  
Reduce output records=27  
Spilled Records=54  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=66  
CPU time spent (ms)=5690  
Physical memory (bytes) snapshot=856977408  
Virtual memory (bytes) snapshot=12760276992  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=463273984  
Peak Map Virtual memory (bytes)=6375157760  
Peak Reduce Physical memory (bytes)=393703424  
Peak Reduce Virtual memory (bytes)=6385119232

#### 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=2680027

#### File Output Format Counters

Bytes Written=2682998

23/01/04 15:51:34 INFO MahoutDriver: Program took 21091 ms (Minutes:  
0.35151666666666664)

---- K: 01 --

dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/01

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:51:38 INFO AbstractJob: Command line arguments: {--clustering=null, --  
clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/  
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --  
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647],  
--input=[dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors], --  
maxIter=[20], --method=[mapreduce], --numClusters=[1],



```

--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/01], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:51:38 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/01
23/01/04 15:51:38 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 15:51:39 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:51:39 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:51:39 INFO RandomSeedGenerator: Wrote 1 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 15:51:39 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/01
23/01/04 15:51:39 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 15:51:39 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 15:51:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:51:39 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0663
23/01/04 15:51:39 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:51:40 INFO JobSubmitter: number of splits:1
23/01/04 15:51:40 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0663
23/01/04 15:51:40 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:51:40 INFO Configuration: resource-types.xml not found
23/01/04 15:51:40 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:51:40 INFO YarnClientImpl: Submitted application
application_1672817468245_0663
23/01/04 15:51:40 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0663/
23/01/04 15:51:40 INFO Job: Running job: job_1672817468245_0663
23/01/04 15:51:46 INFO Job: Job job_1672817468245_0663 running in uber mode : false
23/01/04 15:51:46 INFO Job: map 0% reduce 0%
23/01/04 15:51:52 INFO Job: map 100% reduce 0%
23/01/04 15:51:58 INFO Job: map 100% reduce 100%
23/01/04 15:51:58 INFO Job: Job job_1672817468245_0663 completed successfully
23/01/04 15:51:58 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=928204
        FILE: Number of bytes written=2388699
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=3486161
        HDFS: Number of bytes written=794075
        HDFS: Number of read operations=19
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1

```

Total time spent by all maps in occupied slots (ms)=17865  
Total time spent by all reduces in occupied slots (ms)=17390  
Total time spent by all map tasks (ms)=3573  
Total time spent by all reduce tasks (ms)=3478  
Total vcore-milliseconds taken by all map tasks=3573  
Total vcore-milliseconds taken by all reduce tasks=3478  
Total megabyte-milliseconds taken by all map tasks=18293760  
Total megabyte-milliseconds taken by all reduce tasks=17807360

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=928193  
Map output materialized bytes=928204  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=928204  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=3430  
Physical memory (bytes) snapshot=719228928  
Virtual memory (bytes) snapshot=12753047552  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401190912  
Peak Map Virtual memory (bytes)=6377463808  
Peak Reduce Physical memory (bytes)=318038016  
Peak Reduce Virtual memory (bytes)=6375583744

#### 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=2680027

#### File Output Format Counters

Bytes Written=794075

23/01/04 15:51:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:51:58 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0666  
23/01/04 15:51:58 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:51:58 INFO JobSubmitter: number of splits:1  
23/01/04 15:51:58 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0666  
23/01/04 15:51:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:51:59 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0666  
23/01/04 15:51:59 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0666/](http://lena-master:8088/proxy/application_1672817468245_0666/)  
23/01/04 15:51:59 INFO Job: Running job: job\_1672817468245\_0666  
23/01/04 15:52:05 INFO Job: Job job\_1672817468245\_0666 running in uber mode : false

23/01/04 15:52:05 INFO Job: map 0% reduce 0%  
23/01/04 15:52:10 INFO Job: map 100% reduce 0%  
23/01/04 15:52:16 INFO Job: map 100% reduce 100%  
23/01/04 15:52:16 INFO Job: Job job\_1672817468245\_0666 completed successfully  
23/01/04 15:52:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1587761  
FILE: Number of bytes written=3707813  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4268735  
HDFS: Number of bytes written=794075  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17565  
Total time spent by all reduces in occupied slots (ms)=17205  
Total time spent by all map tasks (ms)=3513  
Total time spent by all reduce tasks (ms)=3441  
Total vcore-milliseconds taken by all map tasks=3513  
Total vcore-milliseconds taken by all reduce tasks=3441  
Total megabyte-milliseconds taken by all map tasks=17986560  
Total megabyte-milliseconds taken by all reduce tasks=17617920

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=1587750  
Map output materialized bytes=1587761  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=1587761  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=84  
CPU time spent (ms)=3590  
Physical memory (bytes) snapshot=729112576  
Virtual memory (bytes) snapshot=12759433216  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405127168  
Peak Map Virtual memory (bytes)=6377988096  
Peak Reduce Physical memory (bytes)=323985408  
Peak Reduce Virtual memory (bytes)=6381445120

#### 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=2680027
    File Output Format Counters
        Bytes Written=794075
23/01/04 15:52:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:52:16 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0669
23/01/04 15:52:16 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:52:16 INFO JobSubmitter: number of splits:1
23/01/04 15:52:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0669
23/01/04 15:52:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:52:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0669
23/01/04 15:52:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0669/
23/01/04 15:52:16 INFO Job: Running job: job_1672817468245_0669
23/01/04 15:52:22 INFO Job: Job job_1672817468245_0669 running in uber mode : false
23/01/04 15:52:22 INFO Job:  map 0% reduce 0%
23/01/04 15:52:28 INFO Job:  map 100% reduce 0%
23/01/04 15:52:34 INFO Job:  map 100% reduce 100%
23/01/04 15:52:34 INFO Job: Job job_1672817468245_0669 completed successfully
23/01/04 15:52:34 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=1587761
        FILE: Number of bytes written=3707813
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4268735
        HDFS: Number of bytes written=794075
        HDFS: Number of read operations=19
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17520
        Total time spent by all reduces in occupied slots (ms)=17485
        Total time spent by all map tasks (ms)=3504
        Total time spent by all reduce tasks (ms)=3497
        Total vcore-milliseconds taken by all map tasks=3504
        Total vcore-milliseconds taken by all reduce tasks=3497
        Total megabyte-milliseconds taken by all map tasks=17940480
        Total megabyte-milliseconds taken by all reduce tasks=17904640
    Map-Reduce Framework
        Map input records=27
        Map output records=1
        Map output bytes=1587750
        Map output materialized bytes=1587761
        Input split bytes=170
        Combine input records=0
        Combine output records=0
        Reduce input groups=1

```

Reduce shuffle bytes=1587761  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=66  
CPU time spent (ms)=3490  
Physical memory (bytes) snapshot=715096064  
Virtual memory (bytes) snapshot=12765130752  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396107776  
Peak Map Virtual memory (bytes)=6380425216  
Peak Reduce Physical memory (bytes)=318988288  
Peak Reduce Virtual memory (bytes)=6384705536

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=2680027

File Output Format Counters

Bytes Written=794075

23/01/04 15:52:34 INFO KMeansDriver: Clustering data  
23/01/04 15:52:34 INFO KMeansDriver: Running Clustering  
23/01/04 15:52:34 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/01 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/01  
23/01/04 15:52:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:52:34 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0671  
23/01/04 15:52:35 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:52:35 INFO JobSubmitter: number of splits:1  
23/01/04 15:52:35 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0671  
23/01/04 15:52:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:52:35 INFO YarnClientImpl: Submitted application application\_1672817468245\_0671  
23/01/04 15:52:35 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0671/  
23/01/04 15:52:35 INFO Job: Running job: job\_1672817468245\_0671  
23/01/04 15:52:41 INFO Job: Job job\_1672817468245\_0671 running in uber mode : false  
23/01/04 15:52:41 INFO Job: map 0% reduce 0%  
23/01/04 15:52:47 INFO Job: map 100% reduce 0%  
23/01/04 15:52:47 INFO Job: Job job\_1672817468245\_0671 completed successfully  
23/01/04 15:52:47 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=3474466  
HDFS: Number of bytes written=2680727  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17475  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3495  
Total vcore-milliseconds taken by all map tasks=3495  
Total megabyte-milliseconds taken by all map tasks=17894400

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=45  
CPU time spent (ms)=1890  
Physical memory (bytes) snapshot=297627648  
Virtual memory (bytes) snapshot=6383157248  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=297627648  
Peak Map Virtual memory (bytes)=6383157248

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680727

23/01/04 15:52:47 INFO MahoutDriver: Program took 69068 ms (Minutes: 1.1511333333333333)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:52:53 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/ManhattanDistanceMeasure/01/clusters-3-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
ManhattanDistanceMeasure/01.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu  
re/01/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 15:52:54 INFO HadoopUtil: Deleting tmp/representative

23/01/04 15:52:54 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/01/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/  
01/clusters-3-final], --maxIter=[5], --method=[mapreduce],  
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 15:52:55 INFO RepresentativePointsDriver: Representative Points Iteration

0

```
23/01/04 15:52:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:52:55 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0675
23/01/04 15:52:55 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:52:55 INFO JobSubmitter: number of splits:1
23/01/04 15:52:55 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0675
23/01/04 15:52:55 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:52:55 INFO Configuration: resource-types.xml not found
23/01/04 15:52:55 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:52:55 INFO YarnClientImpl: Submitted application
application_1672817468245_0675
23/01/04 15:52:55 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0675/
23/01/04 15:52:55 INFO Job: Running job: job_1672817468245_0675
23/01/04 15:53:01 INFO Job: Job job_1672817468245_0675 running in uber mode : false
23/01/04 15:53:01 INFO Job: map 0% reduce 0%
23/01/04 15:53:08 INFO Job: map 100% reduce 0%
23/01/04 15:53:13 INFO Job: map 100% reduce 100%
23/01/04 15:53:14 INFO Job: Job job_1672817468245_0675 completed successfully
23/01/04 15:53:14 INFO Job: Counters: 54
```

#### File System Counters

```
FILE: Number of bytes read=134189
FILE: Number of bytes written=800971
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=3474954
HDFS: Number of bytes written=531210
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
```

#### Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16600
Total time spent by all reduces in occupied slots (ms)=16230
Total time spent by all map tasks (ms)=3320
Total time spent by all reduce tasks (ms)=3246
Total vcore-milliseconds taken by all map tasks=3320
Total vcore-milliseconds taken by all reduce tasks=3246
Total megabyte-milliseconds taken by all map tasks=16998400
Total megabyte-milliseconds taken by all reduce tasks=16619520
```

#### Map-Reduce Framework

```
Map input records=27
Map output records=1
Map output bytes=134178
Map output materialized bytes=134189
Input split bytes=203
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=134189
Reduce input records=1
Reduce output records=2
```

Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=2300  
Physical memory (bytes) snapshot=690438144  
Virtual memory (bytes) snapshot=12751241216  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392708096  
Peak Map Virtual memory (bytes)=6371442688  
Peak Reduce Physical memory (bytes)=297730048  
Peak Reduce Virtual memory (bytes)=6379798528

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=2680727

File Output Format Counters

Bytes Written=531210

23/01/04 15:53:14 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 15:53:14 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:53:14 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0677

23/01/04 15:53:14 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:53:14 INFO JobSubmitter: number of splits:1

23/01/04 15:53:14 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0677

23/01/04 15:53:14 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:53:14 INFO YarnClientImpl: Submitted application application\_1672817468245\_0677

23/01/04 15:53:14 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0677/](http://lena-master:8088/proxy/application_1672817468245_0677/)

23/01/04 15:53:14 INFO Job: Running job: job\_1672817468245\_0677

23/01/04 15:53:20 INFO Job: Job job\_1672817468245\_0677 running in uber mode : false

23/01/04 15:53:20 INFO Job: map 0% reduce 0%

23/01/04 15:53:26 INFO Job: map 100% reduce 0%

23/01/04 15:53:31 INFO Job: map 100% reduce 100%

23/01/04 15:53:32 INFO Job: Job job\_1672817468245\_0677 completed successfully

23/01/04 15:53:32 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=150434  
FILE: Number of bytes written=833461  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3743350  
HDFS: Number of bytes written=681653  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters



Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16820  
Total time spent by all reduces in occupied slots (ms)=16545  
Total time spent by all map tasks (ms)=3364  
Total time spent by all reduce tasks (ms)=3309  
Total vcore-milliseconds taken by all map tasks=3364  
Total vcore-milliseconds taken by all reduce tasks=3309  
Total megabyte-milliseconds taken by all map tasks=17223680  
Total megabyte-milliseconds taken by all reduce tasks=16942080

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=150423  
Map output materialized bytes=150434  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=150434  
Reduce input records=1  
Reduce output records=3  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=2520  
Physical memory (bytes) snapshot=687796224  
Virtual memory (bytes) snapshot=12751110144  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396304384  
Peak Map Virtual memory (bytes)=6374449152  
Peak Reduce Physical memory (bytes)=291491840  
Peak Reduce Virtual memory (bytes)=6376660992

#### 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=2680727

#### File Output Format Counters

Bytes Written=681653

23/01/04 15:53:32 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 15:53:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:53:32 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0680

23/01/04 15:53:32 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:53:32 INFO JobSubmitter: number of splits:1

23/01/04 15:53:33 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0680

23/01/04 15:53:33 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:53:33 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0680

23/01/04 15:53:33 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0680/](http://lena-master:8088/proxy/application_1672817468245_0680/)

23/01/04 15:53:33 INFO Job: Running job: job\_1672817468245\_0680

23/01/04 15:53:39 INFO Job: Job job\_1672817468245\_0680 running in uber mode : false

23/01/04 15:53:39 INFO Job: map 0% reduce 0%

23/01/04 15:53:45 INFO Job: map 100% reduce 0%

23/01/04 15:53:51 INFO Job: map 100% reduce 100%

23/01/04 15:53:51 INFO Job: Job job\_1672817468245\_0680 completed successfully

23/01/04 15:53:51 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=134189

FILE: Number of bytes written=800971

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4044236

HDFS: Number of bytes written=815851

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=17155

Total time spent by all reduces in occupied slots (ms)=16665

Total time spent by all map tasks (ms)=3431

Total time spent by all reduce tasks (ms)=3333

Total vcore-milliseconds taken by all map tasks=3431

Total vcore-milliseconds taken by all reduce tasks=3333

Total megabyte-milliseconds taken by all map tasks=17566720

Total megabyte-milliseconds taken by all reduce tasks=17064960

#### Map-Reduce Framework

Map input records=27

Map output records=1

Map output bytes=134178

Map output materialized bytes=134189

Input split bytes=203

Combine input records=0

Combine output records=0

Reduce input groups=1

Reduce shuffle bytes=134189

Reduce input records=1

Reduce output records=4

Spilled Records=2

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=72

CPU time spent (ms)=2730

Physical memory (bytes) snapshot=691691520

Virtual memory (bytes) snapshot=12747702272

Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=407040000

Peak Map Virtual memory (bytes)=6370951168

Peak Reduce Physical memory (bytes)=284651520

Peak Reduce Virtual memory (bytes)=6378704896

#### 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=2680727

#### File Output Format Counters

Bytes Written=815851

23/01/04 15:53:51 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:53:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:53:51 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0683

23/01/04 15:53:51 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:53:51 INFO JobSubmitter: number of splits:1

23/01/04 15:53:51 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0683

23/01/04 15:53:51 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:53:51 INFO YarnClientImpl: Submitted application application\_1672817468245\_0683

23/01/04 15:53:51 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0683/](http://lena-master:8088/proxy/application_1672817468245_0683/)

23/01/04 15:53:51 INFO Job: Running job: job\_1672817468245\_0683

23/01/04 15:53:57 INFO Job: Job job\_1672817468245\_0683 running in uber mode : false

23/01/04 15:53:57 INFO Job: map 0% reduce 0%

23/01/04 15:54:03 INFO Job: map 100% reduce 0%

23/01/04 15:54:08 INFO Job: map 100% reduce 100%

23/01/04 15:54:09 INFO Job: Job job\_1672817468245\_0683 completed successfully

23/01/04 15:54:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=139404  
FILE: Number of bytes written=811401  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4312632  
HDFS: Number of bytes written=955264  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17365  
Total time spent by all reduces in occupied slots (ms)=16465  
Total time spent by all map tasks (ms)=3473  
Total time spent by all reduce tasks (ms)=3293  
Total vcore-milliseconds taken by all map tasks=3473  
Total vcore-milliseconds taken by all reduce tasks=3293  
Total megabyte-milliseconds taken by all map tasks=17781760  
Total megabyte-milliseconds taken by all reduce tasks=16860160

#### Map-Reduce Framework

Map input records=27

Map output records=1  
Map output bytes=139393  
Map output materialized bytes=139404  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=139404  
Reduce input records=1  
Reduce output records=5  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=30  
CPU time spent (ms)=3110  
Physical memory (bytes) snapshot=693133312  
Virtual memory (bytes) snapshot=12765065216  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404398080  
Peak Map Virtual memory (bytes)=6384668672  
Peak Reduce Physical memory (bytes)=288735232  
Peak Reduce Virtual memory (bytes)=6384304128

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=2680727

File Output Format Counters

Bytes Written=955264

23/01/04 15:54:09 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:54:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:54:09 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0686

23/01/04 15:54:09 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:54:09 INFO JobSubmitter: number of splits:1

23/01/04 15:54:10 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0686

23/01/04 15:54:10 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:54:10 INFO YarnClientImpl: Submitted application application\_1672817468245\_0686

23/01/04 15:54:10 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0686/

23/01/04 15:54:10 INFO Job: Running job: job\_1672817468245\_0686

23/01/04 15:54:16 INFO Job: Job job\_1672817468245\_0686 running in uber mode : false

23/01/04 15:54:16 INFO Job: map 0% reduce 0%

23/01/04 15:54:22 INFO Job: map 100% reduce 0%

23/01/04 15:54:27 INFO Job: map 100% reduce 100%

23/01/04 15:54:28 INFO Job: Job job\_1672817468245\_0686 completed successfully

23/01/04 15:54:28 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=139222

FILE: Number of bytes written=811037

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4591458  
HDFS: Number of bytes written=1094495  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17810  
Total time spent by all reduces in occupied slots (ms)=16400  
Total time spent by all map tasks (ms)=3562  
Total time spent by all reduce tasks (ms)=3280  
Total vcore-milliseconds taken by all map tasks=3562  
Total vcore-milliseconds taken by all reduce tasks=3280  
Total megabyte-milliseconds taken by all map tasks=18237440  
Total megabyte-milliseconds taken by all reduce tasks=16793600

#### Map-Reduce Framework

Map input records=27  
Map output records=1  
Map output bytes=139211  
Map output materialized bytes=139222  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=139222  
Reduce input records=1  
Reduce output records=6  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=38  
CPU time spent (ms)=3200  
Physical memory (bytes) snapshot=708591616  
Virtual memory (bytes) snapshot=12758462464  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=427618304  
Peak Map Virtual memory (bytes)=6375239680  
Peak Reduce Physical memory (bytes)=280973312  
Peak Reduce Virtual memory (bytes)=6386192384

#### 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=2680727

#### File Output Format Counters

Bytes Written=1094495

23/01/04 15:54:28 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN  
23/01/04 15:54:28 INFO ClusterEvaluator: Intra-Cluster Density[17] =

0.6616020772324018  
23/01/04 15:54:28 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.6616020772324018  
23/01/04 15:54:28 INFO ClusterDumper: Wrote 1 clusters  
23/01/04 15:54:28 INFO MahoutDriver: Program took 94909 ms (Minutes:  
1.5818166666666666)  
---- K: 02 --  
dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/02  
MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.  
Running on hadoop, using /opt/hadoop/current/bin/hadoop and  
HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop  
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar  
23/01/04 15:54:33 INFO AbstractJob: Command line arguments: [--clustering=null, --  
clusters=[dsm010/question\_2/british-fiction-corpus-canopy-centroids/  
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --  
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647],  
--input=[dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors], --  
maxIter=[20], --method=[mapreduce], --numClusters=[2],  
--output=[hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-  
kmeans/ManhattanDistanceMeasure/02], --overwrite=null, --startPhase=[0], --  
tempDir=[temp]]  
23/01/04 15:54:34 INFO HadoopUtil: Deleting  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/02  
23/01/04 15:54:34 INFO HadoopUtil: Deleting dsm010/question\_2/british-fiction-  
corpus-canopy-centroids/ManhattanDistanceMeasure  
23/01/04 15:54:34 INFO ZlibFactory: Successfully loaded & initialized native-zlib  
library  
23/01/04 15:54:34 INFO CodecPool: Got brand-new compressor [.deflate]  
23/01/04 15:54:34 INFO RandomSeedGenerator: Wrote 2 Klusters to  
dsm010/question\_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/  
part-randomSeed  
23/01/04 15:54:34 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-  
corpus-vectors/tf-vectors Clusters In: dsm010/question\_2/british-fiction-corpus-  
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/02  
23/01/04 15:54:34 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 15:54:34 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 15:54:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:54:35 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0690  
23/01/04 15:54:35 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:54:35 INFO JobSubmitter: number of splits:1  
23/01/04 15:54:35 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0690  
23/01/04 15:54:35 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:54:35 INFO Configuration: resource-types.xml not found  
23/01/04 15:54:35 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 15:54:35 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0690  
23/01/04 15:54:35 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0690/  
23/01/04 15:54:35 INFO Job: Running job: job\_1672817468245\_0690  
23/01/04 15:54:42 INFO Job: Job job\_1672817468245\_0690 running in uber mode : false  
23/01/04 15:54:42 INFO Job: map 0% reduce 0%  
23/01/04 15:54:48 INFO Job: map 100% reduce 0%

23/01/04 15:54:53 INFO Job: map 100% reduce 100%  
23/01/04 15:54:54 INFO Job: Job job\_1672817468245\_0690 completed successfully  
23/01/04 15:54:54 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1210809  
FILE: Number of bytes written=2953909  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3751475  
HDFS: Number of bytes written=1019494  
HDFS: Number of read operations=23  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17870  
Total time spent by all reduces in occupied slots (ms)=17410  
Total time spent by all map tasks (ms)=3574  
Total time spent by all reduce tasks (ms)=3482  
Total vcore-milliseconds taken by all map tasks=3574  
Total vcore-milliseconds taken by all reduce tasks=3482  
Total megabyte-milliseconds taken by all map tasks=18298880  
Total megabyte-milliseconds taken by all reduce tasks=17827840

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=1210793  
Map output materialized bytes=1210809  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=1210809  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=75  
CPU time spent (ms)=3810  
Physical memory (bytes) snapshot=723914752  
Virtual memory (bytes) snapshot=12760391680  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398495744  
Peak Map Virtual memory (bytes)=6376906752  
Peak Reduce Physical memory (bytes)=325419008  
Peak Reduce Virtual memory (bytes)=6383484928

#### 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=2680027

#### File Output Format Counters

Bytes Written=1019494

23/01/04 15:54:54 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:54:54 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0693  
23/01/04 15:54:54 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:54:54 INFO JobSubmitter: number of splits:1  
23/01/04 15:54:54 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0693  
23/01/04 15:54:54 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:54:54 INFO YarnClientImpl: Submitted application application\_1672817468245\_0693  
23/01/04 15:54:54 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0693/  
23/01/04 15:54:54 INFO Job: Running job: job\_1672817468245\_0693  
23/01/04 15:55:00 INFO Job: Job job\_1672817468245\_0693 running in uber mode : false  
23/01/04 15:55:00 INFO Job: map 0% reduce 0%  
23/01/04 15:55:06 INFO Job: map 100% reduce 0%  
23/01/04 15:55:11 INFO Job: map 100% reduce 100%  
23/01/04 15:55:12 INFO Job: Job job\_1672817468245\_0693 completed successfully  
23/01/04 15:55:12 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2051425  
FILE: Number of bytes written=4635141  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4719573  
HDFS: Number of bytes written=1019494  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17605  
Total time spent by all reduces in occupied slots (ms)=17235  
Total time spent by all map tasks (ms)=3521  
Total time spent by all reduce tasks (ms)=3447  
Total vcore-milliseconds taken by all map tasks=3521  
Total vcore-milliseconds taken by all reduce tasks=3447  
Total megabyte-milliseconds taken by all map tasks=18027520  
Total megabyte-milliseconds taken by all reduce tasks=17648640

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=2051409  
Map output materialized bytes=2051425  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=2051425  
Reduce input records=2



```

Reduce output records=2
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=43
CPU time spent (ms)=3750
Physical memory (bytes) snapshot=738267136
Virtual memory (bytes) snapshot=12752609280
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=407126016
Peak Map Virtual memory (bytes)=6374199296
Peak Reduce Physical memory (bytes)=331141120
Peak Reduce Virtual memory (bytes)=6378409984

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=2680027

File Output Format Counters
  Bytes Written=1019494

23/01/04 15:55:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:55:13 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0696
23/01/04 15:55:13 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:55:13 INFO JobSubmitter: number of splits:1
23/01/04 15:55:13 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0696
23/01/04 15:55:13 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:55:13 INFO YarnClientImpl: Submitted application
application_1672817468245_0696
23/01/04 15:55:13 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0696/
23/01/04 15:55:13 INFO Job: Running job: job_1672817468245_0696
23/01/04 15:55:19 INFO Job: Job job_1672817468245_0696 running in uber mode : false
23/01/04 15:55:19 INFO Job:  map 0% reduce 0%
23/01/04 15:55:25 INFO Job:  map 100% reduce 0%
23/01/04 15:55:31 INFO Job:  map 100% reduce 100%
23/01/04 15:55:31 INFO Job: Job job_1672817468245_0696 completed successfully
23/01/04 15:55:31 INFO Job: Counters: 54

File System Counters
  FILE: Number of bytes read=2051425
  FILE: Number of bytes written=4635141
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4719573
  HDFS: Number of bytes written=1019494
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0

Job Counters
  Launched map tasks=1

```

Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18160  
Total time spent by all reduces in occupied slots (ms)=17150  
Total time spent by all map tasks (ms)=3632  
Total time spent by all reduce tasks (ms)=3430  
Total vcore-milliseconds taken by all map tasks=3632  
Total vcore-milliseconds taken by all reduce tasks=3430  
Total megabyte-milliseconds taken by all map tasks=18595840  
Total megabyte-milliseconds taken by all reduce tasks=17561600

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=2051409  
Map output materialized bytes=2051425  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=2051425  
Reduce input records=2  
Reduce output records=2  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=118  
CPU time spent (ms)=4050  
Physical memory (bytes) snapshot=740487168  
Virtual memory (bytes) snapshot=12760043520  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=403943424  
Peak Map Virtual memory (bytes)=6378713088  
Peak Reduce Physical memory (bytes)=336543744  
Peak Reduce Virtual memory (bytes)=6381330432

#### 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=2680027

#### File Output Format Counters

Bytes Written=1019494

23/01/04 15:55:31 INFO KMeansDriver: Clustering data  
23/01/04 15:55:31 INFO KMeansDriver: Running Clustering  
23/01/04 15:55:31 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/02 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/02  
23/01/04 15:55:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:55:31 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0699  
23/01/04 15:55:31 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:55:31 INFO JobSubmitter: number of splits:1  
23/01/04 15:55:31 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0699  
23/01/04 15:55:31 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:55:31 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0699  
23/01/04 15:55:31 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0699/  
23/01/04 15:55:31 INFO Job: Running job: job\_1672817468245\_0699  
23/01/04 15:55:37 INFO Job: Job job\_1672817468245\_0699 running in uber mode : false  
23/01/04 15:55:37 INFO Job: map 0% reduce 0%  
23/01/04 15:55:43 INFO Job: map 100% reduce 0%  
23/01/04 15:55:43 INFO Job: Job job\_1672817468245\_0699 completed successfully  
23/01/04 15:55:43 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=3699885  
HDFS: Number of bytes written=2680707  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17650  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3530  
Total vcore-milliseconds taken by all map tasks=3530  
Total megabyte-milliseconds taken by all map tasks=18073600

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=19  
CPU time spent (ms)=2210  
Physical memory (bytes) snapshot=324898816  
Virtual memory (bytes) snapshot=6384279552  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=324898816  
Peak Map Virtual memory (bytes)=6384279552

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680707

23/01/04 15:55:43 INFO MahoutDriver: Program took 69484 ms (Minutes:  
1.1580666666666666)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 15:55:50 INFO AbstractJob: Command line arguments:

```

{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/02/clusters-3-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/02.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu
re/02/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:55:51 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:55:51 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/02/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
02/clusters-3-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:55:51 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:55:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:55:51 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0702
23/01/04 15:55:51 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:55:51 INFO JobSubmitter: number of splits:1
23/01/04 15:55:52 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0702
23/01/04 15:55:52 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:55:52 INFO Configuration: resource-types.xml not found
23/01/04 15:55:52 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:55:52 INFO YarnClientImpl: Submitted application
application_1672817468245_0702
23/01/04 15:55:52 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0702/
23/01/04 15:55:52 INFO Job: Running job: job_1672817468245_0702
23/01/04 15:55:58 INFO Job: Job job_1672817468245_0702 running in uber mode : false
23/01/04 15:55:58 INFO Job: map 0% reduce 0%
23/01/04 15:56:03 INFO Job: map 100% reduce 0%
23/01/04 15:56:09 INFO Job: map 100% reduce 100%
23/01/04 15:56:09 INFO Job: Job job_1672817468245_0702 completed successfully
23/01/04 15:56:09 INFO Job: Counters: 54

```

#### File System Counters

```

FILE: Number of bytes read=219509
FILE: Number of bytes written=971611
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=3713276
HDFS: Number of bytes written=735696
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0

```

#### Job Counters

```

Launched map tasks=1
Launched reduce tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16815

```

Total time spent by all reduces in occupied slots (ms)=16400  
Total time spent by all map tasks (ms)=3363  
Total time spent by all reduce tasks (ms)=3280  
Total vcore-milliseconds taken by all map tasks=3363  
Total vcore-milliseconds taken by all reduce tasks=3280  
Total megabyte-milliseconds taken by all map tasks=17218560  
Total megabyte-milliseconds taken by all reduce tasks=16793600

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=219493  
Map output materialized bytes=219509  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=219509  
Reduce input records=2  
Reduce output records=4  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=2630  
Physical memory (bytes) snapshot=681332736  
Virtual memory (bytes) snapshot=12770832384  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=385867776  
Peak Map Virtual memory (bytes)=6374330368  
Peak Reduce Physical memory (bytes)=295464960  
Peak Reduce Virtual memory (bytes)=6398234624

#### 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=2680707

#### File Output Format Counters

Bytes Written=735696

23/01/04 15:56:09 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 15:56:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:56:09 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0704

23/01/04 15:56:09 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:56:09 INFO JobSubmitter: number of splits:1

23/01/04 15:56:09 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0704

23/01/04 15:56:09 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:56:09 INFO YarnClientImpl: Submitted application application\_1672817468245\_0704

23/01/04 15:56:09 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0704/](http://lena-master:8088/proxy/application_1672817468245_0704/)

23/01/04 15:56:09 INFO Job: Running job: job\_1672817468245\_0704

23/01/04 15:56:16 INFO Job: Job job\_1672817468245\_0704 running in uber mode : false  
23/01/04 15:56:16 INFO Job: map 0% reduce 0%  
23/01/04 15:56:21 INFO Job: map 100% reduce 0%  
23/01/04 15:56:27 INFO Job: map 100% reduce 100%  
23/01/04 15:56:27 INFO Job: Job job\_1672817468245\_0704 completed successfully  
23/01/04 15:56:27 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=241496  
FILE: Number of bytes written=1015585  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4152302  
HDFS: Number of bytes written=977196  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16580  
Total time spent by all reduces in occupied slots (ms)=16385  
Total time spent by all map tasks (ms)=3316  
Total time spent by all reduce tasks (ms)=3277  
Total vcore-milliseconds taken by all map tasks=3316  
Total vcore-milliseconds taken by all reduce tasks=3277  
Total megabyte-milliseconds taken by all map tasks=16977920  
Total megabyte-milliseconds taken by all reduce tasks=16778240

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=241480  
Map output materialized bytes=241496  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=241496  
Reduce input records=2  
Reduce output records=6  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=2460  
Physical memory (bytes) snapshot=695312384  
Virtual memory (bytes) snapshot=12761108480  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408072192  
Peak Map Virtual memory (bytes)=6377033728  
Peak Reduce Physical memory (bytes)=287240192  
Peak Reduce Virtual memory (bytes)=6387982336

#### 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=2680707  
File Output Format Counters  
  Bytes Written=977196  
23/01/04 15:56:27 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 15:56:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:56:27 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0707  
23/01/04 15:56:27 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:56:27 INFO JobSubmitter: number of splits:1  
23/01/04 15:56:27 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0707  
23/01/04 15:56:27 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:56:27 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0707  
23/01/04 15:56:27 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0707/  
23/01/04 15:56:27 INFO Job: Running job: job\_1672817468245\_0707  
23/01/04 15:56:33 INFO Job: Job job\_1672817468245\_0707 running in uber mode : false  
23/01/04 15:56:33 INFO Job: map 0% reduce 0%  
23/01/04 15:56:38 INFO Job: map 100% reduce 0%  
23/01/04 15:56:44 INFO Job: map 100% reduce 100%  
23/01/04 15:56:44 INFO Job: Job job\_1672817468245\_0707 completed successfully  
23/01/04 15:56:44 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=219509  
  FILE: Number of bytes written=971611  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=4635302  
  HDFS: Number of bytes written=1196729  
  HDFS: Number of read operations=15  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17320  
  Total time spent by all reduces in occupied slots (ms)=16660  
  Total time spent by all map tasks (ms)=3464  
  Total time spent by all reduce tasks (ms)=3332  
  Total vcore-milliseconds taken by all map tasks=3464  
  Total vcore-milliseconds taken by all reduce tasks=3332  
  Total megabyte-milliseconds taken by all map tasks=17735680  
  Total megabyte-milliseconds taken by all reduce tasks=17059840  
Map-Reduce Framework  
  Map input records=27  
  Map output records=2  
  Map output bytes=219493  
  Map output materialized bytes=219509  
  Input split bytes=203

Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=219509  
Reduce input records=2  
Reduce output records=8  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=70  
CPU time spent (ms)=2960  
Physical memory (bytes) snapshot=678977536  
Virtual memory (bytes) snapshot=12750508032  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=399261696  
Peak Map Virtual memory (bytes)=6374256640  
Peak Reduce Physical memory (bytes)=279715840  
Peak Reduce Virtual memory (bytes)=6379835392

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=2680707

File Output Format Counters

Bytes Written=1196729

23/01/04 15:56:44 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 15:56:44 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:56:44 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0710  
23/01/04 15:56:45 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:56:45 INFO JobSubmitter: number of splits:1  
23/01/04 15:56:45 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0710  
23/01/04 15:56:45 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:56:45 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0710  
23/01/04 15:56:45 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0710/  
23/01/04 15:56:45 INFO Job: Running job: job\_1672817468245\_0710  
23/01/04 15:56:51 INFO Job: Job job\_1672817468245\_0710 running in uber mode : false  
23/01/04 15:56:51 INFO Job: map 0% reduce 0%  
23/01/04 15:56:57 INFO Job: map 100% reduce 0%  
23/01/04 15:57:03 INFO Job: map 100% reduce 100%  
23/01/04 15:57:03 INFO Job: Job job\_1672817468245\_0710 completed successfully  
23/01/04 15:57:03 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=230466  
FILE: Number of bytes written=993525  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5074368



HDFS: Number of bytes written=1427199  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17350  
Total time spent by all reduces in occupied slots (ms)=16285  
Total time spent by all map tasks (ms)=3470  
Total time spent by all reduce tasks (ms)=3257  
Total vcore-milliseconds taken by all map tasks=3470  
Total vcore-milliseconds taken by all reduce tasks=3257  
Total megabyte-milliseconds taken by all map tasks=17766400  
Total megabyte-milliseconds taken by all reduce tasks=16675840

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=230450  
Map output materialized bytes=230466  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=230466  
Reduce input records=2  
Reduce output records=10  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=3280  
Physical memory (bytes) snapshot=708452352  
Virtual memory (bytes) snapshot=12749119488  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415617024  
Peak Map Virtual memory (bytes)=6372147200  
Peak Reduce Physical memory (bytes)=292835328  
Peak Reduce Virtual memory (bytes)=6380888064

#### 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=2680707

#### File Output Format Counters

Bytes Written=1427199

23/01/04 15:57:03 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:57:03 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:57:03 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0712

23/01/04 15:57:03 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:57:03 INFO JobSubmitter: number of splits:1  
23/01/04 15:57:03 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0712  
23/01/04 15:57:03 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:57:03 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0712  
23/01/04 15:57:03 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0712/  
23/01/04 15:57:03 INFO Job: Running job: job\_1672817468245\_0712  
23/01/04 15:57:09 INFO Job: Job job\_1672817468245\_0712 running in uber mode : false  
23/01/04 15:57:09 INFO Job: map 0% reduce 0%  
23/01/04 15:57:15 INFO Job: map 100% reduce 0%  
23/01/04 15:57:19 INFO Job: map 100% reduce 100%  
23/01/04 15:57:19 INFO Job: Job job\_1672817468245\_0712 completed successfully  
23/01/04 15:57:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=224542  
FILE: Number of bytes written=981677  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5535308  
HDFS: Number of bytes written=1651765  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17570  
Total time spent by all reduces in occupied slots (ms)=9440  
Total time spent by all map tasks (ms)=3514  
Total time spent by all reduce tasks (ms)=1888  
Total vcore-milliseconds taken by all map tasks=3514  
Total vcore-milliseconds taken by all reduce tasks=1888  
Total megabyte-milliseconds taken by all map tasks=17991680  
Total megabyte-milliseconds taken by all reduce tasks=9666560

#### Map-Reduce Framework

Map input records=27  
Map output records=2  
Map output bytes=224526  
Map output materialized bytes=224542  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=224542  
Reduce input records=2  
Reduce output records=12  
Spilled Records=4  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=3230  
Physical memory (bytes) snapshot=723656704

```

Virtual memory (bytes) snapshot=12753272832
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=424284160
Peak Map Virtual memory (bytes)=6370668544
Peak Reduce Physical memory (bytes)=299372544
Peak Reduce Virtual memory (bytes)=6382604288
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=2680707
File Output Format Counters
  Bytes Written=1651765
23/01/04 15:57:19 INFO ClusterEvaluator: Scaled Inter-Cluster Density = NaN
23/01/04 15:57:20 INFO ClusterEvaluator: Intra-Cluster Density[25] =
0.5666666666666667
23/01/04 15:57:20 INFO ClusterEvaluator: Intra-Cluster Density[1] =
0.6609247936383614
23/01/04 15:57:20 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6137957301525141
23/01/04 15:57:20 INFO ClusterDumper: Wrote 2 clusters
23/01/04 15:57:20 INFO MahoutDriver: Program took 90331 ms (Minutes:
1.5055166666666666)
---- K: 03 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/03
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:57:24 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[3],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/03], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 15:57:25 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/03
23/01/04 15:57:25 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 15:57:25 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 15:57:25 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 15:57:25 INFO RandomSeedGenerator: Wrote 3 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 15:57:25 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/03

```

23/01/04 15:57:25 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 15:57:25 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 15:57:26 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:57:26 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0716  
23/01/04 15:57:26 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:57:26 INFO JobSubmitter: number of splits:1  
23/01/04 15:57:26 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0716  
23/01/04 15:57:26 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:57:26 INFO Configuration: resource-types.xml not found  
23/01/04 15:57:26 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 15:57:26 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0716  
23/01/04 15:57:26 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0716/  
23/01/04 15:57:26 INFO Job: Running job: job\_1672817468245\_0716  
23/01/04 15:57:32 INFO Job: Job job\_1672817468245\_0716 running in uber mode : false  
23/01/04 15:57:32 INFO Job: map 0% reduce 0%  
23/01/04 15:57:38 INFO Job: map 100% reduce 0%  
23/01/04 15:57:44 INFO Job: map 100% reduce 100%  
23/01/04 15:57:45 INFO Job: Job job\_1672817468245\_0716 completed successfully  
23/01/04 15:57:45 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1928602  
FILE: Number of bytes written=4389495  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4897689  
HDFS: Number of bytes written=1420123  
HDFS: Number of read operations=27  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18190  
Total time spent by all reduces in occupied slots (ms)=17460  
Total time spent by all map tasks (ms)=3638  
Total time spent by all reduce tasks (ms)=3492  
Total vcore-milliseconds taken by all map tasks=3638  
Total vcore-milliseconds taken by all reduce tasks=3492  
Total megabyte-milliseconds taken by all map tasks=18626560  
Total megabyte-milliseconds taken by all reduce tasks=17879040

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=1928581  
Map output materialized bytes=1928602  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=1928602  
Reduce input records=3

Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=4030  
Physical memory (bytes) snapshot=736145408  
Virtual memory (bytes) snapshot=12752293888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415440896  
Peak Map Virtual memory (bytes)=6373535744  
Peak Reduce Physical memory (bytes)=320704512  
Peak Reduce Virtual memory (bytes)=6378758144

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=2680027

File Output Format Counters

Bytes Written=1420123

23/01/04 15:57:45 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:57:45 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0719  
23/01/04 15:57:45 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:57:45 INFO JobSubmitter: number of splits:1  
23/01/04 15:57:45 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0719  
23/01/04 15:57:45 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:57:45 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0719  
23/01/04 15:57:45 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0719/](http://lena-master:8088/proxy/application_1672817468245_0719/)  
23/01/04 15:57:45 INFO Job: Running job: job\_1672817468245\_0719  
23/01/04 15:57:51 INFO Job: Job job\_1672817468245\_0719 running in uber mode : false  
23/01/04 15:57:51 INFO Job: map 0% reduce 0%  
23/01/04 15:57:57 INFO Job: map 100% reduce 0%  
23/01/04 15:58:04 INFO Job: map 100% reduce 100%  
23/01/04 15:58:04 INFO Job: Job job\_1672817468245\_0719 completed successfully  
23/01/04 15:58:04 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=2975737  
FILE: Number of bytes written=6483765  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5520831  
HDFS: Number of bytes written=1417111  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18525  
Total time spent by all reduces in occupied slots (ms)=17780  
Total time spent by all map tasks (ms)=3705  
Total time spent by all reduce tasks (ms)=3556  
Total vcore-milliseconds taken by all map tasks=3705  
Total vcore-milliseconds taken by all reduce tasks=3556  
Total megabyte-milliseconds taken by all map tasks=18969600  
Total megabyte-milliseconds taken by all reduce tasks=18206720

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=2975716  
Map output materialized bytes=2975737  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=2975737  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=4150  
Physical memory (bytes) snapshot=739000320  
Virtual memory (bytes) snapshot=12761001984  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=417415168  
Peak Map Virtual memory (bytes)=6372970496  
Peak Reduce Physical memory (bytes)=321585152  
Peak Reduce Virtual memory (bytes)=6388031488

#### 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=2680027

#### File Output Format Counters

Bytes Written=1417111

23/01/04 15:58:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 15:58:04 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0722  
23/01/04 15:58:04 INFO FileInputFormat: Total input files to process : 1  
23/01/04 15:58:04 INFO JobSubmitter: number of splits:1  
23/01/04 15:58:04 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0722  
23/01/04 15:58:04 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:58:04 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0722  
23/01/04 15:58:04 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0722/](http://lena-master:8088/proxy/application_1672817468245_0722/)

23/01/04 15:58:04 INFO Job: Running job: job\_1672817468245\_0722  
23/01/04 15:58:11 INFO Job: Job job\_1672817468245\_0722 running in uber mode : false  
23/01/04 15:58:11 INFO Job: map 0% reduce 0%  
23/01/04 15:58:17 INFO Job: map 100% reduce 0%  
23/01/04 15:58:23 INFO Job: map 100% reduce 100%  
23/01/04 15:58:23 INFO Job: Job job\_1672817468245\_0722 completed successfully  
23/01/04 15:58:23 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2972725  
FILE: Number of bytes written=6477741  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5514807  
HDFS: Number of bytes written=1417111  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18325  
Total time spent by all reduces in occupied slots (ms)=17565  
Total time spent by all map tasks (ms)=3665  
Total time spent by all reduce tasks (ms)=3513  
Total vcore-milliseconds taken by all map tasks=3665  
Total vcore-milliseconds taken by all reduce tasks=3513  
Total megabyte-milliseconds taken by all map tasks=18764800  
Total megabyte-milliseconds taken by all reduce tasks=17986560

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=2972704  
Map output materialized bytes=2972725  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=2972725  
Reduce input records=3  
Reduce output records=3  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=83  
CPU time spent (ms)=4180  
Physical memory (bytes) snapshot=744509440  
Virtual memory (bytes) snapshot=12754948096  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406306816  
Peak Map Virtual memory (bytes)=6376976384  
Peak Reduce Physical memory (bytes)=338202624  
Peak Reduce Virtual memory (bytes)=6377971712

#### 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=2680027
File Output Format Counters
    Bytes Written=1417111
23/01/04 15:58:23 INFO KMeansDriver: Clustering data
23/01/04 15:58:23 INFO KMeansDriver: Running Clustering
23/01/04 15:58:23 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/03 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/03
23/01/04 15:58:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:58:23 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0725
23/01/04 15:58:23 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:58:23 INFO JobSubmitter: number of splits:1
23/01/04 15:58:23 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0725
23/01/04 15:58:23 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:58:23 INFO YarnClientImpl: Submitted application
application_1672817468245_0725
23/01/04 15:58:23 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0725/
23/01/04 15:58:23 INFO Job: Running job: job_1672817468245_0725
23/01/04 15:58:29 INFO Job: Job job_1672817468245_0725 running in uber mode : false
23/01/04 15:58:29 INFO Job:  map 0% reduce 0%
23/01/04 15:58:35 INFO Job:  map 100% reduce 0%
23/01/04 15:58:35 INFO Job: Job job_1672817468245_0725 completed successfully
23/01/04 15:58:35 INFO Job: Counters: 33
File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=265897
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4097502
    HDFS: Number of bytes written=2680719
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=18345
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=3669
    Total vcore-milliseconds taken by all map tasks=3669
    Total megabyte-milliseconds taken by all map tasks=18785280
Map-Reduce Framework
    Map input records=27
    Map output records=27
    Input split bytes=170

```



```

    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=22
    CPU time spent (ms)=2500
    Physical memory (bytes) snapshot=327192576
    Virtual memory (bytes) snapshot=6383362048
    Total committed heap usage (bytes)=1052770304
    Peak Map Physical memory (bytes)=327192576
    Peak Map Virtual memory (bytes)=6383362048
File Input Format Counters
    Bytes Read=2680027
File Output Format Counters
    Bytes Written=2680719
23/01/04 15:58:35 INFO MahoutDriver: Program took 70871 ms (Minutes:
1.1811833333333333)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 15:58:41 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/03/clusters-3-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/03.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu
re/03/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 15:58:43 INFO HadoopUtil: Deleting tmp/representative
23/01/04 15:58:43 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/03/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
03/clusters-3-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 15:58:43 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 15:58:43 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 15:58:43 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0728
23/01/04 15:58:43 INFO FileInputFormat: Total input files to process : 1
23/01/04 15:58:43 INFO JobSubmitter: number of splits:1
23/01/04 15:58:43 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0728
23/01/04 15:58:43 INFO JobSubmitter: Executing with tokens: []
23/01/04 15:58:44 INFO Configuration: resource-types.xml not found
23/01/04 15:58:44 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 15:58:44 INFO YarnClientImpl: Submitted application
application_1672817468245_0728
23/01/04 15:58:44 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0728/
23/01/04 15:58:44 INFO Job: Running job: job_1672817468245_0728
23/01/04 15:58:50 INFO Job: Job job_1672817468245_0728 running in uber mode : false
23/01/04 15:58:50 INFO Job: map 0% reduce 0%

```

23/01/04 15:58:55 INFO Job: map 100% reduce 0%  
23/01/04 15:59:01 INFO Job: map 100% reduce 100%  
23/01/04 15:59:01 INFO Job: Job job\_1672817468245\_0728 completed successfully  
23/01/04 15:59:01 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=423833  
FILE: Number of bytes written=1380259  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4237068  
HDFS: Number of bytes written=1201945  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16270  
Total time spent by all reduces in occupied slots (ms)=16610  
Total time spent by all map tasks (ms)=3254  
Total time spent by all reduce tasks (ms)=3322  
Total vcore-milliseconds taken by all map tasks=3254  
Total vcore-milliseconds taken by all reduce tasks=3322  
Total megabyte-milliseconds taken by all map tasks=16660480  
Total megabyte-milliseconds taken by all reduce tasks=17008640

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=423812  
Map output materialized bytes=423833  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=423833  
Reduce input records=3  
Reduce output records=6  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=2650  
Physical memory (bytes) snapshot=676745216  
Virtual memory (bytes) snapshot=12745310208  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=392577024  
Peak Map Virtual memory (bytes)=6367711232  
Peak Reduce Physical memory (bytes)=284168192  
Peak Reduce Virtual memory (bytes)=6381506560

#### 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=2680719  
 File Output Format Counters  
   Bytes Written=1201945  
 23/01/04 15:59:01 INFO RepresentativePointsDriver: Representative Points Iteration 1  
 23/01/04 15:59:01 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
 23/01/04 15:59:01 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0731  
 23/01/04 15:59:01 INFO FileInputFormat: Total input files to process : 1  
 23/01/04 15:59:01 INFO JobSubmitter: number of splits:1  
 23/01/04 15:59:01 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0731  
 23/01/04 15:59:01 INFO JobSubmitter: Executing with tokens: []  
 23/01/04 15:59:01 INFO YarnClientImpl: Submitted application application\_1672817468245\_0731  
 23/01/04 15:59:01 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0731/  
 23/01/04 15:59:01 INFO Job: Running job: job\_1672817468245\_0731  
 23/01/04 15:59:07 INFO Job: Job job\_1672817468245\_0731 running in uber mode : false  
 23/01/04 15:59:07 INFO Job: map 0% reduce 0%  
 23/01/04 15:59:13 INFO Job: map 100% reduce 0%  
 23/01/04 15:59:19 INFO Job: map 100% reduce 100%  
 23/01/04 15:59:19 INFO Job: Job job\_1672817468245\_0731 completed successfully  
 23/01/04 15:59:20 INFO Job: Counters: 54  
   File System Counters  
     FILE: Number of bytes read=422903  
     FILE: Number of bytes written=1378399  
     FILE: Number of read operations=0  
     FILE: Number of large read operations=0  
     FILE: Number of write operations=0  
     HDFS: Number of bytes read=5084812  
     HDFS: Number of bytes written=1624887  
     HDFS: Number of read operations=15  
     HDFS: Number of large read operations=0  
     HDFS: Number of write operations=2  
     HDFS: Number of bytes read erasure-coded=0  
   Job Counters  
     Launched map tasks=1  
     Launched reduce tasks=1  
     Rack-local map tasks=1  
     Total time spent by all maps in occupied slots (ms)=17000  
     Total time spent by all reduces in occupied slots (ms)=16710  
     Total time spent by all map tasks (ms)=3400  
     Total time spent by all reduce tasks (ms)=3342  
     Total vcore-milliseconds taken by all map tasks=3400  
     Total vcore-milliseconds taken by all reduce tasks=3342  
     Total megabyte-milliseconds taken by all map tasks=17408000  
     Total megabyte-milliseconds taken by all reduce tasks=17111040  
   Map-Reduce Framework  
     Map input records=27  
     Map output records=3  
     Map output bytes=422882  
     Map output materialized bytes=422903  
     Input split bytes=203  
     Combine input records=0  
     Combine output records=0

Reduce input groups=3  
Reduce shuffle bytes=422903  
Reduce input records=3  
Reduce output records=9  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47  
CPU time spent (ms)=3020  
Physical memory (bytes) snapshot=698466304  
Virtual memory (bytes) snapshot=12750065664  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411697152  
Peak Map Virtual memory (bytes)=6371917824  
Peak Reduce Physical memory (bytes)=286769152  
Peak Reduce Virtual memory (bytes)=6384103424

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=2680719

File Output Format Counters

Bytes Written=1624887

23/01/04 15:59:20 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 15:59:20 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:59:20 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0734

23/01/04 15:59:20 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:59:20 INFO JobSubmitter: number of splits:1

23/01/04 15:59:20 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0734

23/01/04 15:59:20 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:59:20 INFO YarnClientImpl: Submitted application application\_1672817468245\_0734

23/01/04 15:59:20 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0734/](http://lena-master:8088/proxy/application_1672817468245_0734/)

23/01/04 15:59:20 INFO Job: Running job: job\_1672817468245\_0734

23/01/04 15:59:26 INFO Job: Job job\_1672817468245\_0734 running in uber mode : false

23/01/04 15:59:26 INFO Job: map 0% reduce 0%

23/01/04 15:59:32 INFO Job: map 100% reduce 0%

23/01/04 15:59:38 INFO Job: map 100% reduce 100%

23/01/04 15:59:38 INFO Job: Job job\_1672817468245\_0734 completed successfully

23/01/04 15:59:38 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=405475

FILE: Number of bytes written=1343543

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=5930696

HDFS: Number of bytes written=2030401

HDFS: Number of read operations=15

HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17245  
Total time spent by all reduces in occupied slots (ms)=16565  
Total time spent by all map tasks (ms)=3449  
Total time spent by all reduce tasks (ms)=3313  
Total vcore-milliseconds taken by all map tasks=3449  
Total vcore-milliseconds taken by all reduce tasks=3313  
Total megabyte-milliseconds taken by all map tasks=17658880  
Total megabyte-milliseconds taken by all reduce tasks=16962560

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=405454  
Map output materialized bytes=405475  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=405475  
Reduce input records=3  
Reduce output records=12  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=61  
CPU time spent (ms)=3110  
Physical memory (bytes) snapshot=714854400  
Virtual memory (bytes) snapshot=12749750272  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415907840  
Peak Map Virtual memory (bytes)=6369349632  
Peak Reduce Physical memory (bytes)=298946560  
Peak Reduce Virtual memory (bytes)=6382354432

#### 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=2680719

#### File Output Format Counters

Bytes Written=2030401

23/01/04 15:59:38 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 15:59:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:59:38 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0736

23/01/04 15:59:38 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:59:38 INFO JobSubmitter: number of splits:1

23/01/04 15:59:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0736  
23/01/04 15:59:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 15:59:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0736  
23/01/04 15:59:38 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0736/  
23/01/04 15:59:38 INFO Job: Running job: job\_1672817468245\_0736  
23/01/04 15:59:44 INFO Job: Job job\_1672817468245\_0736 running in uber mode : false  
23/01/04 15:59:44 INFO Job: map 0% reduce 0%  
23/01/04 15:59:50 INFO Job: map 100% reduce 0%  
23/01/04 15:59:55 INFO Job: map 100% reduce 100%  
23/01/04 15:59:56 INFO Job: Job job\_1672817468245\_0736 completed successfully  
23/01/04 15:59:56 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=373993  
FILE: Number of bytes written=1280579  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6741724  
HDFS: Number of bytes written=2404433  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17530  
Total time spent by all reduces in occupied slots (ms)=16465  
Total time spent by all map tasks (ms)=3506  
Total time spent by all reduce tasks (ms)=3293  
Total vcore-milliseconds taken by all map tasks=3506  
Total vcore-milliseconds taken by all reduce tasks=3293  
Total megabyte-milliseconds taken by all map tasks=17950720  
Total megabyte-milliseconds taken by all reduce tasks=16860160

#### Map-Reduce Framework

Map input records=27  
Map output records=3  
Map output bytes=373972  
Map output materialized bytes=373993  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=3  
Reduce shuffle bytes=373993  
Reduce input records=3  
Reduce output records=15  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3340  
Physical memory (bytes) snapshot=695017472  
Virtual memory (bytes) snapshot=12748767232  
Total committed heap usage (bytes)=2105540608

Peak Map Physical memory (bytes)=402759680  
Peak Map Virtual memory (bytes)=6370312192  
Peak Reduce Physical memory (bytes)=292257792  
Peak Reduce Virtual memory (bytes)=6378455040

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=2680719

File Output Format Counters

Bytes Written=2404433

23/01/04 15:59:56 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 15:59:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 15:59:56 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0738

23/01/04 15:59:57 INFO FileInputFormat: Total input files to process : 1

23/01/04 15:59:57 INFO JobSubmitter: number of splits:1

23/01/04 15:59:57 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0738

23/01/04 15:59:57 INFO JobSubmitter: Executing with tokens: []

23/01/04 15:59:57 INFO YarnClientImpl: Submitted application application\_1672817468245\_0738

23/01/04 15:59:57 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0738/

23/01/04 15:59:57 INFO Job: Running job: job\_1672817468245\_0738

23/01/04 16:00:03 INFO Job: Job job\_1672817468245\_0738 running in uber mode : false

23/01/04 16:00:03 INFO Job: map 0% reduce 0%

23/01/04 16:00:09 INFO Job: map 100% reduce 0%

23/01/04 16:00:15 INFO Job: map 100% reduce 100%

23/01/04 16:00:15 INFO Job: Job job\_1672817468245\_0738 completed successfully

23/01/04 16:00:15 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=402821  
FILE: Number of bytes written=1338235  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7489788  
HDFS: Number of bytes written=2807293  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18300  
Total time spent by all reduces in occupied slots (ms)=16840  
Total time spent by all map tasks (ms)=3660  
Total time spent by all reduce tasks (ms)=3368  
Total vcore-milliseconds taken by all map tasks=3660  
Total vcore-milliseconds taken by all reduce tasks=3368

```

        Total megabyte-milliseconds taken by all map tasks=18739200
        Total megabyte-milliseconds taken by all reduce tasks=17244160
Map-Reduce Framework
    Map input records=27
    Map output records=3
    Map output bytes=402800
    Map output materialized bytes=402821
    Input split bytes=203
    Combine input records=0
    Combine output records=0
    Reduce input groups=3
    Reduce shuffle bytes=402821
    Reduce input records=3
    Reduce output records=18
    Spilled Records=6
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=49
    CPU time spent (ms)=3550
    Physical memory (bytes) snapshot=721514496
    Virtual memory (bytes) snapshot=12756103168
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=435781632
    Peak Map Virtual memory (bytes)=6378717184
    Peak Reduce Physical memory (bytes)=285732864
    Peak Reduce Virtual memory (bytes)=6381301760
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=2680719
File Output Format Counters
    Bytes Written=2807293

```

```

23/01/04 16:00:15 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.5677011769728706
23/01/04 16:00:15 INFO ClusterEvaluator: Intra-Cluster Density[24] =
0.6538715010834588
23/01/04 16:00:15 INFO ClusterEvaluator: Intra-Cluster Density[7] =
0.7370081153461865
23/01/04 16:00:15 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN
23/01/04 16:00:15 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6954398082148227
23/01/04 16:00:16 INFO ClusterDumper: Wrote 3 clusters
23/01/04 16:00:16 INFO MahoutDriver: Program took 94719 ms (Minutes: 1.57865)
---- K: 04 --

```

```

dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/04
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

```

```

23/01/04 16:00:20 INFO AbstractJob: Command line arguments: [--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --

```



```

endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[4],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/04], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 16:00:21 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/04
23/01/04 16:00:21 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 16:00:21 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 16:00:21 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 16:00:21 INFO RandomSeedGenerator: Wrote 4 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 16:00:21 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/04
23/01/04 16:00:21 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 16:00:21 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 16:00:21 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:00:21 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0743
23/01/04 16:00:22 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:00:22 INFO JobSubmitter: number of splits:1
23/01/04 16:00:22 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0743
23/01/04 16:00:22 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:00:22 INFO Configuration: resource-types.xml not found
23/01/04 16:00:22 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:00:22 INFO YarnClientImpl: Submitted application
application_1672817468245_0743
23/01/04 16:00:22 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0743/
23/01/04 16:00:22 INFO Job: Running job: job_1672817468245_0743
23/01/04 16:00:28 INFO Job: Job job_1672817468245_0743 running in uber mode : false
23/01/04 16:00:28 INFO Job:  map 0% reduce 0%
23/01/04 16:00:34 INFO Job:  map 100% reduce 0%
23/01/04 16:00:38 INFO Job:  map 100% reduce 100%
23/01/04 16:00:38 INFO Job: Job job_1672817468245_0743 completed successfully
23/01/04 16:00:38 INFO Job: Counters: 54

```

#### File System Counters

```

FILE: Number of bytes read=2254457
FILE: Number of bytes written=5041205
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=4927235
HDFS: Number of bytes written=1722715
HDFS: Number of read operations=31
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0

```

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18300  
Total time spent by all reduces in occupied slots (ms)=10255  
Total time spent by all map tasks (ms)=3660  
Total time spent by all reduce tasks (ms)=2051  
Total vcore-milliseconds taken by all map tasks=3660  
Total vcore-milliseconds taken by all reduce tasks=2051  
Total megabyte-milliseconds taken by all map tasks=18739200  
Total megabyte-milliseconds taken by all reduce tasks=10501120

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=2254431  
Map output materialized bytes=2254457  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=2254457  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=4190  
Physical memory (bytes) snapshot=719781888  
Virtual memory (bytes) snapshot=12744589312  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406171648  
Peak Map Virtual memory (bytes)=6368649216  
Peak Reduce Physical memory (bytes)=313610240  
Peak Reduce Virtual memory (bytes)=6375940096

#### 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=2680027

#### File Output Format Counters

Bytes Written=1722715

23/01/04 16:00:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:00:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0746  
23/01/04 16:00:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:00:39 INFO JobSubmitter: number of splits:1  
23/01/04 16:00:39 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0746  
23/01/04 16:00:39 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:00:39 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0746  
23/01/04 16:00:39 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0746/  
23/01/04 16:00:39 INFO Job: Running job: job\_1672817468245\_0746  
23/01/04 16:00:45 INFO Job: Job job\_1672817468245\_0746 running in uber mode : false  
23/01/04 16:00:45 INFO Job: map 0% reduce 0%  
23/01/04 16:00:51 INFO Job: map 100% reduce 0%  
23/01/04 16:00:57 INFO Job: map 100% reduce 100%  
23/01/04 16:00:57 INFO Job: Job job\_1672817468245\_0746 completed successfully  
23/01/04 16:00:57 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3594752  
FILE: Number of bytes written=7721795  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6126015  
HDFS: Number of bytes written=1715447  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18310  
Total time spent by all reduces in occupied slots (ms)=17460  
Total time spent by all map tasks (ms)=3662  
Total time spent by all reduce tasks (ms)=3492  
Total vcore-milliseconds taken by all map tasks=3662  
Total vcore-milliseconds taken by all reduce tasks=3492  
Total megabyte-milliseconds taken by all map tasks=18749440  
Total megabyte-milliseconds taken by all reduce tasks=17879040

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3594726  
Map output materialized bytes=3594752  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3594752  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=4070  
Physical memory (bytes) snapshot=742580224  
Virtual memory (bytes) snapshot=12768096256  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=428208128  
Peak Map Virtual memory (bytes)=6381780992  
Peak Reduce Physical memory (bytes)=314372096  
Peak Reduce Virtual memory (bytes)=6386315264

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1715447  
23/01/04 16:00:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:00:57 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0749  
23/01/04 16:00:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:00:57 INFO JobSubmitter: number of splits:1  
23/01/04 16:00:57 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0749  
23/01/04 16:00:57 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:00:57 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0749  
23/01/04 16:00:57 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0749/  
23/01/04 16:00:57 INFO Job: Running job: job\_1672817468245\_0749  
23/01/04 16:01:03 INFO Job: Job job\_1672817468245\_0749 running in uber mode : false  
23/01/04 16:01:03 INFO Job:  map 0% reduce 0%  
23/01/04 16:01:09 INFO Job:  map 100% reduce 0%  
23/01/04 16:01:13 INFO Job:  map 100% reduce 100%  
23/01/04 16:01:13 INFO Job: Job job\_1672817468245\_0749 completed successfully  
23/01/04 16:01:13 INFO Job: Counters: 54  
File System Counters  
  FILE: Number of bytes read=3343420  
  FILE: Number of bytes written=7219131  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=6111479  
  HDFS: Number of bytes written=1629224  
  HDFS: Number of read operations=19  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=18410  
  Total time spent by all reduces in occupied slots (ms)=10170  
  Total time spent by all map tasks (ms)=3682  
  Total time spent by all reduce tasks (ms)=2034  
  Total vcore-milliseconds taken by all map tasks=3682  
  Total vcore-milliseconds taken by all reduce tasks=2034  
  Total megabyte-milliseconds taken by all map tasks=18851840  
  Total megabyte-milliseconds taken by all reduce tasks=10414080  
Map-Reduce Framework  
  Map input records=27  
  Map output records=4  
  Map output bytes=3343394  
  Map output materialized bytes=3343420  
  Input split bytes=170

Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3343420  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=4120  
Physical memory (bytes) snapshot=733904896  
Virtual memory (bytes) snapshot=12758929408  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=416071680  
Peak Map Virtual memory (bytes)=6379753472  
Peak Reduce Physical memory (bytes)=317833216  
Peak Reduce Virtual memory (bytes)=6379175936

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=2680027

File Output Format Counters

Bytes Written=1629224

23/01/04 16:01:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:01:13 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0750  
23/01/04 16:01:14 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:01:14 INFO JobSubmitter: number of splits:1  
23/01/04 16:01:14 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0750  
23/01/04 16:01:14 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:01:14 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0750  
23/01/04 16:01:14 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0750/  
23/01/04 16:01:14 INFO Job: Running job: job\_1672817468245\_0750  
23/01/04 16:01:20 INFO Job: Job job\_1672817468245\_0750 running in uber mode : false  
23/01/04 16:01:20 INFO Job: map 0% reduce 0%  
23/01/04 16:01:26 INFO Job: map 100% reduce 0%  
23/01/04 16:01:32 INFO Job: map 100% reduce 100%  
23/01/04 16:01:32 INFO Job: Job job\_1672817468245\_0750 completed successfully  
23/01/04 16:01:32 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3280385  
FILE: Number of bytes written=7093061  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5939033  
HDFS: Number of bytes written=1667716  
HDFS: Number of read operations=19

HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17980  
Total time spent by all reduces in occupied slots (ms)=17115  
Total time spent by all map tasks (ms)=3596  
Total time spent by all reduce tasks (ms)=3423  
Total vcore-milliseconds taken by all map tasks=3596  
Total vcore-milliseconds taken by all reduce tasks=3423  
Total megabyte-milliseconds taken by all map tasks=18411520  
Total megabyte-milliseconds taken by all reduce tasks=17525760

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3280359  
Map output materialized bytes=3280385  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3280385  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=94  
CPU time spent (ms)=4450  
Physical memory (bytes) snapshot=752615424  
Virtual memory (bytes) snapshot=12766896128  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=427225088  
Peak Map Virtual memory (bytes)=6380797952  
Peak Reduce Physical memory (bytes)=325390336  
Peak Reduce Virtual memory (bytes)=6386098176

#### 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=2680027

#### File Output Format Counters

Bytes Written=1667716

23/01/04 16:01:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:01:32 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0753

23/01/04 16:01:32 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:01:32 INFO JobSubmitter: number of splits:1

23/01/04 16:01:32 INFO JobSubmitter: Submitting tokens for job:

job\_1672817468245\_0753

23/01/04 16:01:32 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:01:32 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0753  
23/01/04 16:01:32 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0753/  
23/01/04 16:01:32 INFO Job: Running job: job\_1672817468245\_0753  
23/01/04 16:01:38 INFO Job: Job job\_1672817468245\_0753 running in uber mode : false  
23/01/04 16:01:38 INFO Job: map 0% reduce 0%  
23/01/04 16:01:44 INFO Job: map 100% reduce 0%  
23/01/04 16:01:49 INFO Job: map 100% reduce 100%  
23/01/04 16:01:50 INFO Job: Job job\_1672817468245\_0753 completed successfully  
23/01/04 16:01:50 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3381631  
FILE: Number of bytes written=7295553  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6016017  
HDFS: Number of bytes written=1733602  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18360  
Total time spent by all reduces in occupied slots (ms)=17190  
Total time spent by all map tasks (ms)=3672  
Total time spent by all reduce tasks (ms)=3438  
Total vcore-milliseconds taken by all map tasks=3672  
Total vcore-milliseconds taken by all reduce tasks=3438  
Total megabyte-milliseconds taken by all map tasks=18800640  
Total megabyte-milliseconds taken by all reduce tasks=17602560

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3381605  
Map output materialized bytes=3381631  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3381631  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=4090  
Physical memory (bytes) snapshot=745365504  
Virtual memory (bytes) snapshot=12766789632  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406790144  
Peak Map Virtual memory (bytes)=6376538112

```

        Peak Reduce Physical memory (bytes)=338575360
        Peak Reduce Virtual memory (bytes)=6390251520
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=2680027
File Output Format Counters
    Bytes Written=1733602
23/01/04 16:01:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:01:50 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0756
23/01/04 16:01:51 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:01:51 INFO JobSubmitter: number of splits:1
23/01/04 16:01:51 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0756
23/01/04 16:01:51 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:01:51 INFO YarnClientImpl: Submitted application
application_1672817468245_0756
23/01/04 16:01:51 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0756/
23/01/04 16:01:51 INFO Job: Running job: job_1672817468245_0756
23/01/04 16:01:57 INFO Job: Job job_1672817468245_0756 running in uber mode : false
23/01/04 16:01:57 INFO Job:  map 0% reduce 0%
23/01/04 16:02:03 INFO Job:  map 100% reduce 0%
23/01/04 16:02:09 INFO Job:  map 100% reduce 100%
23/01/04 16:02:09 INFO Job: Job job_1672817468245_0756 completed successfully
23/01/04 16:02:09 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=3528017
    FILE: Number of bytes written=7588325
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6147789
    HDFS: Number of bytes written=1814048
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=18710
    Total time spent by all reduces in occupied slots (ms)=17720
    Total time spent by all map tasks (ms)=3742
    Total time spent by all reduce tasks (ms)=3544
    Total vcore-milliseconds taken by all map tasks=3742
    Total vcore-milliseconds taken by all reduce tasks=3544
    Total megabyte-milliseconds taken by all map tasks=19159040
    Total megabyte-milliseconds taken by all reduce tasks=18145280
Map-Reduce Framework
    Map input records=27

```



Map output records=4  
Map output bytes=3527991  
Map output materialized bytes=3528017  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3528017  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=58  
CPU time spent (ms)=4040  
Physical memory (bytes) snapshot=726675456  
Virtual memory (bytes) snapshot=12751273984  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401354752  
Peak Map Virtual memory (bytes)=6371864576  
Peak Reduce Physical memory (bytes)=325320704  
Peak Reduce Virtual memory (bytes)=6379409408

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=2680027

File Output Format Counters

Bytes Written=1814048

23/01/04 16:02:09 INFO DefaultNoHARMFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:02:09 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0759  
23/01/04 16:02:09 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:02:09 INFO JobSubmitter: number of splits:1  
23/01/04 16:02:09 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0759  
23/01/04 16:02:09 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:02:09 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0759  
23/01/04 16:02:09 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0759/](http://lena-master:8088/proxy/application_1672817468245_0759/)  
23/01/04 16:02:09 INFO Job: Running job: job\_1672817468245\_0759  
23/01/04 16:02:15 INFO Job: Job job\_1672817468245\_0759 running in uber mode : false  
23/01/04 16:02:15 INFO Job: map 0% reduce 0%  
23/01/04 16:02:21 INFO Job: map 100% reduce 0%  
23/01/04 16:02:25 INFO Job: map 100% reduce 100%  
23/01/04 16:02:25 INFO Job: Job job\_1672817468245\_0759 completed successfully  
23/01/04 16:02:25 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3566333  
FILE: Number of bytes written=7664957  
FILE: Number of read operations=0  
FILE: Number of large read operations=0

FILE: Number of write operations=0  
HDFS: Number of bytes read=6308681  
HDFS: Number of bytes written=1771236  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18415  
Total time spent by all reduces in occupied slots (ms)=10460  
Total time spent by all map tasks (ms)=3683  
Total time spent by all reduce tasks (ms)=2092  
Total vcore-milliseconds taken by all map tasks=3683  
Total vcore-milliseconds taken by all reduce tasks=2092  
Total megabyte-milliseconds taken by all map tasks=18856960  
Total megabyte-milliseconds taken by all reduce tasks=10711040

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3566307  
Map output materialized bytes=3566333  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3566333  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=68  
CPU time spent (ms)=4400  
Physical memory (bytes) snapshot=734650368  
Virtual memory (bytes) snapshot=12764868608  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=417792000  
Peak Map Virtual memory (bytes)=6373232640  
Peak Reduce Physical memory (bytes)=316858368  
Peak Reduce Virtual memory (bytes)=6391635968

#### 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=2680027

#### File Output Format Counters

Bytes Written=1771236

23/01/04 16:02:25 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:02:25 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0762

23/01/04 16:02:25 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:02:26 INFO JobSubmitter: number of splits:1  
23/01/04 16:02:26 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0762  
23/01/04 16:02:26 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:02:26 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0762  
23/01/04 16:02:26 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0762/](http://lena-master:8088/proxy/application_1672817468245_0762/)  
23/01/04 16:02:26 INFO Job: Running job: job\_1672817468245\_0762  
23/01/04 16:02:32 INFO Job: Job job\_1672817468245\_0762 running in uber mode : false  
23/01/04 16:02:32 INFO Job: map 0% reduce 0%  
23/01/04 16:02:38 INFO Job: map 100% reduce 0%  
23/01/04 16:02:42 INFO Job: map 100% reduce 100%  
23/01/04 16:02:42 INFO Job: Job job\_1672817468245\_0762 completed successfully  
23/01/04 16:02:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3523521  
FILE: Number of bytes written=7579333  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6223057  
HDFS: Number of bytes written=1771236  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18575  
Total time spent by all reduces in occupied slots (ms)=9770  
Total time spent by all map tasks (ms)=3715  
Total time spent by all reduce tasks (ms)=1954  
Total vcore-milliseconds taken by all map tasks=3715  
Total vcore-milliseconds taken by all reduce tasks=1954  
Total megabyte-milliseconds taken by all map tasks=19020800  
Total megabyte-milliseconds taken by all reduce tasks=10004480

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=3523495  
Map output materialized bytes=3523521  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3523521  
Reduce input records=4  
Reduce output records=4  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=4300  
Physical memory (bytes) snapshot=739168256

Virtual memory (bytes) snapshot=12754399232  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410767360  
Peak Map Virtual memory (bytes)=6371938304  
Peak Reduce Physical memory (bytes)=328400896  
Peak Reduce Virtual memory (bytes)=6382460928

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=2680027

File Output Format Counters

Bytes Written=1771236

23/01/04 16:02:42 INFO KMeansDriver: Clustering data  
23/01/04 16:02:42 INFO KMeansDriver: Running Clustering  
23/01/04 16:02:42 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/04 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/04  
23/01/04 16:02:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:02:42 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0765  
23/01/04 16:02:42 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:02:42 INFO JobSubmitter: number of splits:1  
23/01/04 16:02:42 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0765  
23/01/04 16:02:42 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:02:42 INFO YarnClientImpl: Submitted application application\_1672817468245\_0765  
23/01/04 16:02:42 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0765/  
23/01/04 16:02:42 INFO Job: Running job: job\_1672817468245\_0765  
23/01/04 16:02:48 INFO Job: Job job\_1672817468245\_0765 running in uber mode : false  
23/01/04 16:02:48 INFO Job: map 0% reduce 0%  
23/01/04 16:02:54 INFO Job: map 100% reduce 0%  
23/01/04 16:02:54 INFO Job: Job job\_1672817468245\_0765 completed successfully  
23/01/04 16:02:54 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4451627  
HDFS: Number of bytes written=2680669  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1

```

Total time spent by all maps in occupied slots (ms)=18855
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=3771
Total vcore-milliseconds taken by all map tasks=3771
Total megabyte-milliseconds taken by all map tasks=19307520
Map-Reduce Framework
  Map input records=27
  Map output records=27
  Input split bytes=170
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=20
  CPU time spent (ms)=2610
  Physical memory (bytes) snapshot=314372096
  Virtual memory (bytes) snapshot=6375702528
  Total committed heap usage (bytes)=1052770304
  Peak Map Physical memory (bytes)=314372096
  Peak Map Virtual memory (bytes)=6375702528
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680669
23/01/04 16:02:54 INFO MahoutDriver: Program took 154282 ms (Minutes:
2.5713833333333334)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:03:00 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/04/clusters-8-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/04.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu
re/04/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 16:03:02 INFO HadoopUtil: Deleting tmp/representative
23/01/04 16:03:02 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/04/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
04/clusters-8-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 16:03:02 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 16:03:02 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:03:02 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0767
23/01/04 16:03:02 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:03:02 INFO JobSubmitter: number of splits:1
23/01/04 16:03:02 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0767
23/01/04 16:03:02 INFO JobSubmitter: Executing with tokens: []

```

23/01/04 16:03:03 INFO Configuration: resource-types.xml not found  
23/01/04 16:03:03 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:03:03 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0767  
23/01/04 16:03:03 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0767/  
23/01/04 16:03:03 INFO Job: Running job: job\_1672817468245\_0767  
23/01/04 16:03:09 INFO Job: Job job\_1672817468245\_0767 running in uber mode : false  
23/01/04 16:03:09 INFO Job: map 0% reduce 0%  
23/01/04 16:03:15 INFO Job: map 100% reduce 0%  
23/01/04 16:03:20 INFO Job: map 100% reduce 100%  
23/01/04 16:03:21 INFO Job: Job job\_1672817468245\_0767 completed successfully  
23/01/04 16:03:21 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=535125  
FILE: Number of bytes written=1602843  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4433786  
HDFS: Number of bytes written=1411636  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16650  
Total time spent by all reduces in occupied slots (ms)=16495  
Total time spent by all map tasks (ms)=3330  
Total time spent by all reduce tasks (ms)=3299  
Total vcore-milliseconds taken by all map tasks=3330  
Total vcore-milliseconds taken by all reduce tasks=3299  
Total megabyte-milliseconds taken by all map tasks=17049600  
Total megabyte-milliseconds taken by all reduce tasks=16890880

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=535099  
Map output materialized bytes=535125  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=535125  
Reduce input records=4  
Reduce output records=8  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=2620  
Physical memory (bytes) snapshot=692350976  
Virtual memory (bytes) snapshot=12752457728  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=400625664

```

Peak Map Virtual memory (bytes)=6371622912
Peak Reduce Physical memory (bytes)=291725312
Peak Reduce Virtual memory (bytes)=6382788608
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=2680669
File Output Format Counters
  Bytes Written=1411636
23/01/04 16:03:21 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 16:03:21 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:03:21 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0770
23/01/04 16:03:21 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:03:21 INFO JobSubmitter: number of splits:1
23/01/04 16:03:21 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0770
23/01/04 16:03:21 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:03:21 INFO YarnClientImpl: Submitted application
application_1672817468245_0770
23/01/04 16:03:21 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0770/
23/01/04 16:03:21 INFO Job: Running job: job_1672817468245_0770
23/01/04 16:03:27 INFO Job: Job job_1672817468245_0770 running in uber mode : false
23/01/04 16:03:27 INFO Job:  map 0% reduce 0%
23/01/04 16:03:33 INFO Job:  map 100% reduce 0%
23/01/04 16:03:38 INFO Job:  map 100% reduce 100%
23/01/04 16:03:39 INFO Job: Job job_1672817468245_0770 completed successfully
23/01/04 16:03:39 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=525974
  FILE: Number of bytes written=1584541
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5504144
  HDFS: Number of bytes written=1937664
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=16950
  Total time spent by all reduces in occupied slots (ms)=16960
  Total time spent by all map tasks (ms)=3390
  Total time spent by all reduce tasks (ms)=3392
  Total vcore-milliseconds taken by all map tasks=3390
  Total vcore-milliseconds taken by all reduce tasks=3392
  Total megabyte-milliseconds taken by all map tasks=17356800

```

Total megabyte-milliseconds taken by all reduce tasks=17367040  
Map-Reduce Framework  
Map input records=27  
Map output records=4  
Map output bytes=525948  
Map output materialized bytes=525974  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=525974  
Reduce input records=4  
Reduce output records=12  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=2790  
Physical memory (bytes) snapshot=683823104  
Virtual memory (bytes) snapshot=12749615104  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=388100096  
Peak Map Virtual memory (bytes)=6369533952  
Peak Reduce Physical memory (bytes)=295723008  
Peak Reduce Virtual memory (bytes)=6383996928

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=2680669

File Output Format Counters

Bytes Written=1937664

23/01/04 16:03:39 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 16:03:39 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:03:39 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0773  
23/01/04 16:03:40 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:03:40 INFO JobSubmitter: number of splits:1  
23/01/04 16:03:40 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0773  
23/01/04 16:03:40 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:03:40 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0773  
23/01/04 16:03:40 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0773/  
23/01/04 16:03:40 INFO Job: Running job: job\_1672817468245\_0773  
23/01/04 16:03:46 INFO Job: Job job\_1672817468245\_0773 running in uber mode : false  
23/01/04 16:03:46 INFO Job: map 0% reduce 0%  
23/01/04 16:03:52 INFO Job: map 100% reduce 0%  
23/01/04 16:03:58 INFO Job: map 100% reduce 100%  
23/01/04 16:03:58 INFO Job: Job job\_1672817468245\_0773 completed successfully  
23/01/04 16:03:58 INFO Job: Counters: 54



#### File System Counters

FILE: Number of bytes read=520263  
FILE: Number of bytes written=1573119  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6556200  
HDFS: Number of bytes written=2457981  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17065  
Total time spent by all reduces in occupied slots (ms)=16650  
Total time spent by all map tasks (ms)=3413  
Total time spent by all reduce tasks (ms)=3330  
Total vcore-milliseconds taken by all map tasks=3413  
Total vcore-milliseconds taken by all reduce tasks=3330  
Total megabyte-milliseconds taken by all map tasks=17474560  
Total megabyte-milliseconds taken by all reduce tasks=17049600

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=520237  
Map output materialized bytes=520263  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=520263  
Reduce input records=4  
Reduce output records=16  
Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=53  
CPU time spent (ms)=2930  
Physical memory (bytes) snapshot=694882304  
Virtual memory (bytes) snapshot=12756389888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=398565376  
Peak Map Virtual memory (bytes)=6374739968  
Peak Reduce Physical memory (bytes)=296316928  
Peak Reduce Virtual memory (bytes)=6385565696

#### 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=2680669

#### File Output Format Counters

Bytes Written=2457981

23/01/04 16:03:58 INFO RepresentativePointsDriver: Representative Points Iteration 3  
23/01/04 16:03:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:03:58 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0776  
23/01/04 16:03:58 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:03:58 INFO JobSubmitter: number of splits:1  
23/01/04 16:03:58 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0776  
23/01/04 16:03:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:03:58 INFO YarnClientImpl: Submitted application application\_1672817468245\_0776  
23/01/04 16:03:58 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0776/  
23/01/04 16:03:58 INFO Job: Running job: job\_1672817468245\_0776  
23/01/04 16:04:04 INFO Job: Job job\_1672817468245\_0776 running in uber mode : false  
23/01/04 16:04:04 INFO Job: map 0% reduce 0%  
23/01/04 16:04:10 INFO Job: map 100% reduce 0%  
23/01/04 16:04:15 INFO Job: map 100% reduce 100%  
23/01/04 16:04:16 INFO Job: Job job\_1672817468245\_0776 completed successfully  
23/01/04 16:04:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=490921  
FILE: Number of bytes written=1514435  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7596834  
HDFS: Number of bytes written=2948936  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17380  
Total time spent by all reduces in occupied slots (ms)=17170  
Total time spent by all map tasks (ms)=3476  
Total time spent by all reduce tasks (ms)=3434  
Total vcore-milliseconds taken by all map tasks=3476  
Total vcore-milliseconds taken by all reduce tasks=3434  
Total megabyte-milliseconds taken by all map tasks=17797120  
Total megabyte-milliseconds taken by all reduce tasks=17582080

#### Map-Reduce Framework

Map input records=27  
Map output records=4  
Map output bytes=490895  
Map output materialized bytes=490921  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=490921  
Reduce input records=4  
Reduce output records=20

Spilled Records=8  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=64  
CPU time spent (ms)=3470  
Physical memory (bytes) snapshot=710602752  
Virtual memory (bytes) snapshot=12753776640  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410931200  
Peak Map Virtual memory (bytes)=6374912000  
Peak Reduce Physical memory (bytes)=299671552  
Peak Reduce Virtual memory (bytes)=6382772224

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=2680669

File Output Format Counters

Bytes Written=2948936

23/01/04 16:04:16 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 16:04:16 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:04:16 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0779

23/01/04 16:04:17 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:04:17 INFO JobSubmitter: number of splits:1

23/01/04 16:04:17 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0779

23/01/04 16:04:17 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:04:17 INFO YarnClientImpl: Submitted application

application\_1672817468245\_0779

23/01/04 16:04:17 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0779/](http://lena-master:8088/proxy/application_1672817468245_0779/)

23/01/04 16:04:17 INFO Job: Running job: job\_1672817468245\_0779

23/01/04 16:04:23 INFO Job: map 0% reduce 0%

23/01/04 16:04:29 INFO Job: map 100% reduce 0%

23/01/04 16:04:34 INFO Job: map 100% reduce 100%

23/01/04 16:04:34 INFO Job: Job job\_1672817468245\_0779 completed successfully

23/01/04 16:04:34 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=555903  
FILE: Number of bytes written=1644399  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8578744  
HDFS: Number of bytes written=3504893  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

```

    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17775
    Total time spent by all reduces in occupied slots (ms)=16690
    Total time spent by all map tasks (ms)=3555
    Total time spent by all reduce tasks (ms)=3338
    Total vcore-milliseconds taken by all map tasks=3555
    Total vcore-milliseconds taken by all reduce tasks=3338
    Total megabyte-milliseconds taken by all map tasks=18201600
    Total megabyte-milliseconds taken by all reduce tasks=17090560
Map-Reduce Framework
    Map input records=27
    Map output records=4
    Map output bytes=555877
    Map output materialized bytes=555903
    Input split bytes=203
    Combine input records=0
    Combine output records=0
    Reduce input groups=4
    Reduce shuffle bytes=555903
    Reduce input records=4
    Reduce output records=24
    Spilled Records=8
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=55
    CPU time spent (ms)=3370
    Physical memory (bytes) snapshot=710594560
    Virtual memory (bytes) snapshot=12748079104
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=406908928
    Peak Map Virtual memory (bytes)=6369746944
    Peak Reduce Physical memory (bytes)=303685632
    Peak Reduce Virtual memory (bytes)=6378332160
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=2680669
File Output Format Counters
    Bytes Written=3504893
23/01/04 16:04:34 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.5145907538668104
23/01/04 16:04:34 INFO ClusterEvaluator: Intra-Cluster Density[19] =
0.7327838989762803
23/01/04 16:04:34 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN
23/01/04 16:04:34 INFO ClusterEvaluator: Intra-Cluster Density[11] = NaN
23/01/04 16:04:34 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.5986653264418702
23/01/04 16:04:34 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6657246127090752
23/01/04 16:04:35 INFO ClusterDumper: Wrote 4 clusters
23/01/04 16:04:35 INFO MahoutDriver: Program took 94852 ms (Minutes:
1.5808666666666666)

```

```

---- K: 05 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/05
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:04:39 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[5],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/05], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 16:04:40 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/05
23/01/04 16:04:40 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 16:04:40 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 16:04:40 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 16:04:40 INFO RandomSeedGenerator: Wrote 5 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 16:04:40 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/05
23/01/04 16:04:40 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 16:04:40 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 16:04:41 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:04:41 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0782
23/01/04 16:04:41 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:04:41 INFO JobSubmitter: number of splits:1
23/01/04 16:04:41 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0782
23/01/04 16:04:41 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:04:41 INFO Configuration: resource-types.xml not found
23/01/04 16:04:41 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:04:42 INFO YarnClientImpl: Submitted application
application_1672817468245_0782
23/01/04 16:04:42 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0782/
23/01/04 16:04:42 INFO Job: Running job: job_1672817468245_0782
23/01/04 16:04:48 INFO Job: Job job_1672817468245_0782 running in uber mode : false
23/01/04 16:04:48 INFO Job:  map 0% reduce 0%
23/01/04 16:04:53 INFO Job:  map 100% reduce 0%
23/01/04 16:04:59 INFO Job:  map 100% reduce 100%
23/01/04 16:04:59 INFO Job: Job job_1672817468245_0782 completed successfully
23/01/04 16:04:59 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2242105
        FILE: Number of bytes written=5016501

```

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5537423  
HDFS: Number of bytes written=1579502  
HDFS: Number of read operations=35  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18700  
Total time spent by all reduces in occupied slots (ms)=18335  
Total time spent by all map tasks (ms)=3740  
Total time spent by all reduce tasks (ms)=3667  
Total vcore-milliseconds taken by all map tasks=3740  
Total vcore-milliseconds taken by all reduce tasks=3667  
Total megabyte-milliseconds taken by all map tasks=19148800  
Total megabyte-milliseconds taken by all reduce tasks=18775040

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=2242074  
Map output materialized bytes=2242105  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=2242105  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=86  
CPU time spent (ms)=4410  
Physical memory (bytes) snapshot=735686656  
Virtual memory (bytes) snapshot=12764708864  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418652160  
Peak Map Virtual memory (bytes)=6380191744  
Peak Reduce Physical memory (bytes)=317034496  
Peak Reduce Virtual memory (bytes)=6384517120

#### 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=2680027

#### File Output Format Counters

Bytes Written=1579502

23/01/04 16:04:59 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:04:59 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0785  
23/01/04 16:04:59 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:04:59 INFO JobSubmitter: number of splits:1  
23/01/04 16:04:59 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0785  
23/01/04 16:04:59 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:04:59 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0785  
23/01/04 16:04:59 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0785/  
23/01/04 16:04:59 INFO Job: Running job: job\_1672817468245\_0785  
23/01/04 16:05:05 INFO Job: Job job\_1672817468245\_0785 running in uber mode : false  
23/01/04 16:05:05 INFO Job: map 0% reduce 0%  
23/01/04 16:05:11 INFO Job: map 100% reduce 0%  
23/01/04 16:05:15 INFO Job: map 100% reduce 100%  
23/01/04 16:05:15 INFO Job: Job job\_1672817468245\_0785 completed successfully  
23/01/04 16:05:15 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3474706  
FILE: Number of bytes written=7481703  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5839589  
HDFS: Number of bytes written=1771727  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18155  
Total time spent by all reduces in occupied slots (ms)=10380  
Total time spent by all map tasks (ms)=3631  
Total time spent by all reduce tasks (ms)=2076  
Total vcore-milliseconds taken by all map tasks=3631  
Total vcore-milliseconds taken by all reduce tasks=2076  
Total megabyte-milliseconds taken by all map tasks=18590720  
Total megabyte-milliseconds taken by all reduce tasks=10629120

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3474675  
Map output materialized bytes=3474706  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3474706  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65

CPU time spent (ms)=4320  
Physical memory (bytes) snapshot=723988480  
Virtual memory (bytes) snapshot=12749508608  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402399232  
Peak Map Virtual memory (bytes)=6371131392  
Peak Reduce Physical memory (bytes)=321589248  
Peak Reduce Virtual memory (bytes)=6378377216

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=2680027

File Output Format Counters  
Bytes Written=1771727

23/01/04 16:05:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:05:15 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0788  
23/01/04 16:05:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:05:16 INFO JobSubmitter: number of splits:1  
23/01/04 16:05:16 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0788  
23/01/04 16:05:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:05:16 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0788  
23/01/04 16:05:16 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0788/](http://lena-master:8088/proxy/application_1672817468245_0788/)  
23/01/04 16:05:16 INFO Job: Running job: job\_1672817468245\_0788  
23/01/04 16:05:22 INFO Job: Job job\_1672817468245\_0788 running in uber mode : false  
23/01/04 16:05:22 INFO Job: map 0% reduce 0%  
23/01/04 16:05:28 INFO Job: map 100% reduce 0%  
23/01/04 16:05:32 INFO Job: map 100% reduce 100%  
23/01/04 16:05:32 INFO Job: Job job\_1672817468245\_0788 completed successfully  
23/01/04 16:05:32 INFO Job: Counters: 54

File System Counters  
FILE: Number of bytes read=3832723  
FILE: Number of bytes written=8197737  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6224039  
HDFS: Number of bytes written=1974612  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18390  
Total time spent by all reduces in occupied slots (ms)=10020  
Total time spent by all map tasks (ms)=3678  
Total time spent by all reduce tasks (ms)=2004



Total vcore-milliseconds taken by all map tasks=3678  
Total vcore-milliseconds taken by all reduce tasks=2004  
Total megabyte-milliseconds taken by all map tasks=18831360  
Total megabyte-milliseconds taken by all reduce tasks=10260480

Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3832692  
Map output materialized bytes=3832723  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3832723  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=4360  
Physical memory (bytes) snapshot=733696000  
Virtual memory (bytes) snapshot=12756414464  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=404123648  
Peak Map Virtual memory (bytes)=6374891520  
Peak Reduce Physical memory (bytes)=329572352  
Peak Reduce Virtual memory (bytes)=6381522944

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=2680027

File Output Format Counters

Bytes Written=1974612

23/01/04 16:05:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:05:32 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0791  
23/01/04 16:05:32 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:05:32 INFO JobSubmitter: number of splits:1  
23/01/04 16:05:32 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0791  
23/01/04 16:05:32 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:05:32 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0791  
23/01/04 16:05:32 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0791/  
23/01/04 16:05:32 INFO Job: Running job: job\_1672817468245\_0791  
23/01/04 16:05:38 INFO Job: Job job\_1672817468245\_0791 running in uber mode : false  
23/01/04 16:05:38 INFO Job: map 0% reduce 0%  
23/01/04 16:05:44 INFO Job: map 100% reduce 0%  
23/01/04 16:05:49 INFO Job: map 100% reduce 100%  
23/01/04 16:05:50 INFO Job: Job job\_1672817468245\_0791 completed successfully

23/01/04 16:05:50 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=3966188  
FILE: Number of bytes written=8464667  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6629809  
HDFS: Number of bytes written=1901346  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18465  
Total time spent by all reduces in occupied slots (ms)=17505  
Total time spent by all map tasks (ms)=3693  
Total time spent by all reduce tasks (ms)=3501  
Total vcore-milliseconds taken by all map tasks=3693  
Total vcore-milliseconds taken by all reduce tasks=3501  
Total megabyte-milliseconds taken by all map tasks=18908160  
Total megabyte-milliseconds taken by all reduce tasks=17925120

Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3966157  
Map output materialized bytes=3966188  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3966188  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=4140  
Physical memory (bytes) snapshot=710189056  
Virtual memory (bytes) snapshot=12753465344  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402468864  
Peak Map Virtual memory (bytes)=6372753408  
Peak Reduce Physical memory (bytes)=307720192  
Peak Reduce Virtual memory (bytes)=6380711936

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=2680027

#### File Output Format Counters

Bytes Written=1901346

23/01/04 16:05:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:05:51 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0794  
23/01/04 16:05:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:05:51 INFO JobSubmitter: number of splits:1  
23/01/04 16:05:51 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0794  
23/01/04 16:05:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:05:51 INFO YarnClientImpl: Submitted application application\_1672817468245\_0794  
23/01/04 16:05:51 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0794/  
23/01/04 16:05:51 INFO Job: Running job: job\_1672817468245\_0794  
23/01/04 16:05:57 INFO Job: Job job\_1672817468245\_0794 running in uber mode : false  
23/01/04 16:05:57 INFO Job: map 0% reduce 0%  
23/01/04 16:06:03 INFO Job: map 100% reduce 0%  
23/01/04 16:06:09 INFO Job: map 100% reduce 100%  
23/01/04 16:06:09 INFO Job: Job job\_1672817468245\_0794 completed successfully  
23/01/04 16:06:09 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3892922  
FILE: Number of bytes written=8318135  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6483277  
HDFS: Number of bytes written=1901346  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18515  
Total time spent by all reduces in occupied slots (ms)=17560  
Total time spent by all map tasks (ms)=3703  
Total time spent by all reduce tasks (ms)=3512  
Total vcore-milliseconds taken by all map tasks=3703  
Total vcore-milliseconds taken by all reduce tasks=3512  
Total megabyte-milliseconds taken by all map tasks=18959360  
Total megabyte-milliseconds taken by all reduce tasks=17981440

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=3892891  
Map output materialized bytes=3892922  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=3892922  
Reduce input records=5  
Reduce output records=5  
Spilled Records=10

```

Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=60
CPU time spent (ms)=4490
Physical memory (bytes) snapshot=727580672
Virtual memory (bytes) snapshot=12758396928
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=408109056
Peak Map Virtual memory (bytes)=6378336256
Peak Reduce Physical memory (bytes)=319471616
Peak Reduce Virtual memory (bytes)=6380060672

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=2680027
File Output Format Counters
  Bytes Written=1901346
23/01/04 16:06:09 INFO KMeansDriver: Clustering data
23/01/04 16:06:09 INFO KMeansDriver: Running Clustering
23/01/04 16:06:09 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/05 Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/05
23/01/04 16:06:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:06:09 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0797
23/01/04 16:06:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:06:09 INFO JobSubmitter: number of splits:1
23/01/04 16:06:09 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0797
23/01/04 16:06:09 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:06:09 INFO YarnClientImpl: Submitted application
application_1672817468245_0797
23/01/04 16:06:09 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0797/
23/01/04 16:06:09 INFO Job: Running job: job_1672817468245_0797
23/01/04 16:06:15 INFO Job: Job job_1672817468245_0797 running in uber mode : false
23/01/04 16:06:15 INFO Job:  map 0% reduce 0%
23/01/04 16:06:21 INFO Job:  map 100% reduce 0%
23/01/04 16:06:21 INFO Job: Job job_1672817468245_0797 completed successfully
23/01/04 16:06:22 INFO Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=265897
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4581737
  HDFS: Number of bytes written=2680683
  HDFS: Number of read operations=15

```

HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18440  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3688  
Total vcore-milliseconds taken by all map tasks=3688  
Total megabyte-milliseconds taken by all map tasks=18882560

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=16  
CPU time spent (ms)=2410  
Physical memory (bytes) snapshot=300134400  
Virtual memory (bytes) snapshot=6378688512  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=300134400  
Peak Map Virtual memory (bytes)=6378688512

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680683

23/01/04 16:06:22 INFO MahoutDriver: Program took 102196 ms (Minutes: 1.7032666666666667)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 16:06:28 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --

distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/ManhattanDistanceMeasure/05/clusters-5-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
ManhattanDistanceMeasure/05.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu  
re/05/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 16:06:29 INFO HadoopUtil: Deleting tmp/representative

23/01/04 16:06:29 INFO AbstractJob: Command line arguments: {--

clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/05/clusteredPoints], --

distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647],

--input=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/  
05/clusters-5-final], --maxIter=[5], --method=[mapreduce],

--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}

23/01/04 16:06:29 INFO RepresentativePointsDriver: Representative Points Iteration  
0

23/01/04 16:06:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:06:30 INFO JobResourceUploader: Disabling Erasure Coding for path:

/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0799  
23/01/04 16:06:30 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:06:30 INFO JobSubmitter: number of splits:1  
23/01/04 16:06:30 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0799  
23/01/04 16:06:30 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:06:30 INFO Configuration: resource-types.xml not found  
23/01/04 16:06:30 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:06:30 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0799  
23/01/04 16:06:30 INFO Job: The url to track the job:  
[http://lana-master:8088/proxy/application\\_1672817468245\\_0799/](http://lana-master:8088/proxy/application_1672817468245_0799/)  
23/01/04 16:06:30 INFO Job: Running job: job\_1672817468245\_0799  
23/01/04 16:06:36 INFO Job: Job job\_1672817468245\_0799 running in uber mode : false  
23/01/04 16:06:36 INFO Job: map 0% reduce 0%  
23/01/04 16:06:42 INFO Job: map 100% reduce 0%  
23/01/04 16:06:48 INFO Job: map 100% reduce 100%  
23/01/04 16:06:48 INFO Job: Job job\_1672817468245\_0799 completed successfully  
23/01/04 16:06:48 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=633119  
FILE: Number of bytes written=1798831  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4673188  
HDFS: Number of bytes written=1629319  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16915  
Total time spent by all reduces in occupied slots (ms)=16625  
Total time spent by all map tasks (ms)=3383  
Total time spent by all reduce tasks (ms)=3325  
Total vcore-milliseconds taken by all map tasks=3383  
Total vcore-milliseconds taken by all reduce tasks=3325  
Total megabyte-milliseconds taken by all map tasks=17320960  
Total megabyte-milliseconds taken by all reduce tasks=17024000

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=633088  
Map output materialized bytes=633119  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=633119  
Reduce input records=5  
Reduce output records=10  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=55
CPU time spent (ms)=2870
Physical memory (bytes) snapshot=673374208
Virtual memory (bytes) snapshot=12760055808
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=391217152
Peak Map Virtual memory (bytes)=6373289984
Peak Reduce Physical memory (bytes)=282157056
Peak Reduce Virtual memory (bytes)=6390673408
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=2680683
File Output Format Counters
  Bytes Written=1629319
23/01/04 16:06:48 INFO RepresentativePointsDriver: Representative Points Iteration
1
23/01/04 16:06:48 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:06:48 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0802
23/01/04 16:06:48 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:06:48 INFO JobSubmitter: number of splits:1
23/01/04 16:06:48 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0802
23/01/04 16:06:48 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:06:48 INFO YarnClientImpl: Submitted application
application_1672817468245_0802
23/01/04 16:06:48 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0802/
23/01/04 16:06:48 INFO Job: Running job: job_1672817468245_0802
23/01/04 16:06:54 INFO Job: Job job_1672817468245_0802 running in uber mode : false
23/01/04 16:06:54 INFO Job:  map 0% reduce 0%
23/01/04 16:07:00 INFO Job:  map 100% reduce 0%
23/01/04 16:07:04 INFO Job:  map 100% reduce 100%
23/01/04 16:07:04 INFO Job: Job job_1672817468245_0802 completed successfully
23/01/04 16:07:04 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=629035
  FILE: Number of bytes written=1790663
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=5939524
  HDFS: Number of bytes written=2258403
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=16770

```

Total time spent by all reduces in occupied slots (ms)=9350  
Total time spent by all map tasks (ms)=3354  
Total time spent by all reduce tasks (ms)=1870  
Total vcore-milliseconds taken by all map tasks=3354  
Total vcore-milliseconds taken by all reduce tasks=1870  
Total megabyte-milliseconds taken by all map tasks=17172480  
Total megabyte-milliseconds taken by all reduce tasks=9574400

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=629004  
Map output materialized bytes=629035  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=629035  
Reduce input records=5  
Reduce output records=15  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=2930  
Physical memory (bytes) snapshot=714190848  
Virtual memory (bytes) snapshot=12762030080  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419643392  
Peak Map Virtual memory (bytes)=6376648704  
Peak Reduce Physical memory (bytes)=294547456  
Peak Reduce Virtual memory (bytes)=6385381376

#### 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=2680683

#### File Output Format Counters

Bytes Written=2258403

23/01/04 16:07:04 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 16:07:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:07:04 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0805

23/01/04 16:07:04 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:07:04 INFO JobSubmitter: number of splits:1

23/01/04 16:07:04 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0805

23/01/04 16:07:04 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:07:04 INFO YarnClientImpl: Submitted application application\_1672817468245\_0805

23/01/04 16:07:04 INFO Job: The url to track the job:

[http://lena-master:8088/proxy/application\\_1672817468245\\_0805/](http://lena-master:8088/proxy/application_1672817468245_0805/)

23/01/04 16:07:04 INFO Job: Running job: job\_1672817468245\_0805



23/01/04 16:07:10 INFO Job: Job job\_1672817468245\_0805 running in uber mode : false  
23/01/04 16:07:10 INFO Job: map 0% reduce 0%  
23/01/04 16:07:17 INFO Job: map 100% reduce 0%  
23/01/04 16:07:22 INFO Job: map 100% reduce 100%  
23/01/04 16:07:23 INFO Job: Job job\_1672817468245\_0805 completed successfully  
23/01/04 16:07:23 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=617382  
FILE: Number of bytes written=1767357  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7197692  
HDFS: Number of bytes written=2875854  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17300  
Total time spent by all reduces in occupied slots (ms)=16630  
Total time spent by all map tasks (ms)=3460  
Total time spent by all reduce tasks (ms)=3326  
Total vcore-milliseconds taken by all map tasks=3460  
Total vcore-milliseconds taken by all reduce tasks=3326  
Total megabyte-milliseconds taken by all map tasks=17715200  
Total megabyte-milliseconds taken by all reduce tasks=17029120

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=617351  
Map output materialized bytes=617382  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=617382  
Reduce input records=5  
Reduce output records=20  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=3180  
Physical memory (bytes) snapshot=685010944  
Virtual memory (bytes) snapshot=12755374080  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397340672  
Peak Map Virtual memory (bytes)=6371897344  
Peak Reduce Physical memory (bytes)=287670272  
Peak Reduce Virtual memory (bytes)=6387384320

#### 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=2680683
    File Output Format Counters
        Bytes Written=2875854
23/01/04 16:07:23 INFO RepresentativePointsDriver: Representative Points Iteration
3
23/01/04 16:07:23 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:07:23 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0808
23/01/04 16:07:23 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:07:23 INFO JobSubmitter: number of splits:1
23/01/04 16:07:23 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0808
23/01/04 16:07:23 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:07:23 INFO YarnClientImpl: Submitted application
application_1672817468245_0808
23/01/04 16:07:23 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0808/
23/01/04 16:07:23 INFO Job: Running job: job_1672817468245_0808
23/01/04 16:07:29 INFO Job: Job job_1672817468245_0808 running in uber mode : false
23/01/04 16:07:29 INFO Job:  map 0% reduce 0%
23/01/04 16:07:35 INFO Job:  map 100% reduce 0%
23/01/04 16:07:40 INFO Job:  map 100% reduce 100%
23/01/04 16:07:40 INFO Job: Job job_1672817468245_0808 completed successfully
23/01/04 16:07:40 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=601002
        FILE: Number of bytes written=1734597
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=8432594
        HDFS: Number of bytes written=3476885
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17155
        Total time spent by all reduces in occupied slots (ms)=16435
        Total time spent by all map tasks (ms)=3431
        Total time spent by all reduce tasks (ms)=3287
        Total vcore-milliseconds taken by all map tasks=3431
        Total vcore-milliseconds taken by all reduce tasks=3287
        Total megabyte-milliseconds taken by all map tasks=17566720
        Total megabyte-milliseconds taken by all reduce tasks=16829440
    Map-Reduce Framework
        Map input records=27
        Map output records=5
        Map output bytes=600971
        Map output materialized bytes=601002
        Input split bytes=203

```

Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=601002  
Reduce input records=5  
Reduce output records=25  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=3310  
Physical memory (bytes) snapshot=696324096  
Virtual memory (bytes) snapshot=12764704768  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405622784  
Peak Map Virtual memory (bytes)=6380654592  
Peak Reduce Physical memory (bytes)=292569088  
Peak Reduce Virtual memory (bytes)=6386003968

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=2680683

File Output Format Counters

Bytes Written=3476885

23/01/04 16:07:40 INFO RepresentativePointsDriver: Representative Points Iteration  
4  
23/01/04 16:07:40 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:07:40 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0810  
23/01/04 16:07:40 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:07:41 INFO JobSubmitter: number of splits:1  
23/01/04 16:07:41 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0810  
23/01/04 16:07:41 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:07:41 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0810  
23/01/04 16:07:41 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0810/  
23/01/04 16:07:41 INFO Job: Running job: job\_1672817468245\_0810  
23/01/04 16:07:47 INFO Job: Job job\_1672817468245\_0810 running in uber mode : false  
23/01/04 16:07:47 INFO Job: map 0% reduce 0%  
23/01/04 16:07:52 INFO Job: map 100% reduce 0%  
23/01/04 16:07:58 INFO Job: map 100% reduce 100%  
23/01/04 16:07:58 INFO Job: Job job\_1672817468245\_0810 completed successfully  
23/01/04 16:07:58 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=654079  
FILE: Number of bytes written=1840751  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9634656

HDFS: Number of bytes written=4131053  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17500  
Total time spent by all reduces in occupied slots (ms)=16580  
Total time spent by all map tasks (ms)=3500  
Total time spent by all reduce tasks (ms)=3316  
Total vcore-milliseconds taken by all map tasks=3500  
Total vcore-milliseconds taken by all reduce tasks=3316  
Total megabyte-milliseconds taken by all map tasks=17920000  
Total megabyte-milliseconds taken by all reduce tasks=16977920

#### Map-Reduce Framework

Map input records=27  
Map output records=5  
Map output bytes=654048  
Map output materialized bytes=654079  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=654079  
Reduce input records=5  
Reduce output records=30  
Spilled Records=10  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=3720  
Physical memory (bytes) snapshot=735031296  
Virtual memory (bytes) snapshot=12757327872  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=423907328  
Peak Map Virtual memory (bytes)=6378692608  
Peak Reduce Physical memory (bytes)=311123968  
Peak Reduce Virtual memory (bytes)=6378635264

#### 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=2680683

#### File Output Format Counters

Bytes Written=4131053

23/01/04 16:07:58 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.4683995886303535

23/01/04 16:07:58 INFO ClusterEvaluator: Intra-Cluster Density[22] =  
0.5666666666666667

23/01/04 16:07:58 INFO ClusterEvaluator: Intra-Cluster Density[18] = NaN

23/01/04 16:07:58 INFO ClusterEvaluator: Intra-Cluster Density[2] =

```
0.5991208405202104
23/01/04 16:07:58 INFO ClusterEvaluator: Intra-Cluster Density[3] =
0.6931841341254931
23/01/04 16:07:58 INFO ClusterEvaluator: Intra-Cluster Density[24] = NaN
23/01/04 16:07:58 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.61965721377079
23/01/04 16:07:59 INFO ClusterDumper: Wrote 5 clusters
23/01/04 16:07:59 INFO MahoutDriver: Program took 91782 ms (Minutes: 1.5297)
---- K: 06 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/06
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:08:04 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[6],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/06], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 16:08:04 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/06
23/01/04 16:08:04 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 16:08:05 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 16:08:05 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 16:08:05 INFO RandomSeedGenerator: Wrote 6 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 16:08:05 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/06
23/01/04 16:08:05 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 16:08:05 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 16:08:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:08:06 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0814
23/01/04 16:08:06 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:08:06 INFO JobSubmitter: number of splits:1
23/01/04 16:08:06 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0814
23/01/04 16:08:06 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:08:06 INFO Configuration: resource-types.xml not found
23/01/04 16:08:06 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:08:06 INFO YarnClientImpl: Submitted application
application_1672817468245_0814
23/01/04 16:08:06 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0814/
23/01/04 16:08:06 INFO Job: Running job: job_1672817468245_0814
23/01/04 16:08:12 INFO Job: Job job_1672817468245_0814 running in uber mode : false
```

23/01/04 16:08:12 INFO Job: map 0% reduce 0%  
23/01/04 16:08:17 INFO Job: map 100% reduce 0%  
23/01/04 16:08:23 INFO Job: map 100% reduce 100%  
23/01/04 16:08:23 INFO Job: Job job\_1672817468245\_0814 completed successfully  
23/01/04 16:08:24 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=2942865  
FILE: Number of bytes written=6418021  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6765895  
HDFS: Number of bytes written=1854169  
HDFS: Number of read operations=39  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18330  
Total time spent by all reduces in occupied slots (ms)=18000  
Total time spent by all map tasks (ms)=3666  
Total time spent by all reduce tasks (ms)=3600  
Total vcore-milliseconds taken by all map tasks=3666  
Total vcore-milliseconds taken by all reduce tasks=3600  
Total megabyte-milliseconds taken by all map tasks=18769920  
Total megabyte-milliseconds taken by all reduce tasks=18432000

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=2942829  
Map output materialized bytes=2942865  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=2942865  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=56  
CPU time spent (ms)=4470  
Physical memory (bytes) snapshot=730193920  
Virtual memory (bytes) snapshot=12763938816  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=420868096  
Peak Map Virtual memory (bytes)=6384754688  
Peak Reduce Physical memory (bytes)=309325824  
Peak Reduce Virtual memory (bytes)=6379184128

#### 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=2680027  
File Output Format Counters  
  Bytes Written=1854169  
23/01/04 16:08:24 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:08:24 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0816  
23/01/04 16:08:24 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:08:24 INFO JobSubmitter: number of splits:1  
23/01/04 16:08:24 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0816  
23/01/04 16:08:24 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:08:24 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0816  
23/01/04 16:08:24 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0816/  
23/01/04 16:08:24 INFO Job: Running job: job\_1672817468245\_0816  
23/01/04 16:08:30 INFO Job: Job job\_1672817468245\_0816 running in uber mode : false  
23/01/04 16:08:30 INFO Job: map 0% reduce 0%  
23/01/04 16:08:36 INFO Job: map 100% reduce 0%  
23/01/04 16:08:42 INFO Job: map 100% reduce 100%  
23/01/04 16:08:42 INFO Job: Job job\_1672817468245\_0816 completed successfully  
23/01/04 16:08:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4126058  
FILE: Number of bytes written=8784407  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6388923  
HDFS: Number of bytes written=1865089  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19005  
Total time spent by all reduces in occupied slots (ms)=17565  
Total time spent by all map tasks (ms)=3801  
Total time spent by all reduce tasks (ms)=3513  
Total vcore-milliseconds taken by all map tasks=3801  
Total vcore-milliseconds taken by all reduce tasks=3513  
Total megabyte-milliseconds taken by all map tasks=19461120  
Total megabyte-milliseconds taken by all reduce tasks=17986560

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4126022  
Map output materialized bytes=4126058  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6

Reduce shuffle bytes=4126058  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=71  
CPU time spent (ms)=4570  
Physical memory (bytes) snapshot=730288128  
Virtual memory (bytes) snapshot=12760047616  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405794816  
Peak Map Virtual memory (bytes)=6372761600  
Peak Reduce Physical memory (bytes)=324493312  
Peak Reduce Virtual memory (bytes)=6387286016

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=2680027

File Output Format Counters

Bytes Written=1865089

23/01/04 16:08:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:08:42 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0819  
23/01/04 16:08:42 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:08:42 INFO JobSubmitter: number of splits:1  
23/01/04 16:08:42 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0819  
23/01/04 16:08:42 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:08:42 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0819  
23/01/04 16:08:42 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0819/  
23/01/04 16:08:42 INFO Job: Running job: job\_1672817468245\_0819  
23/01/04 16:08:49 INFO Job: Job job\_1672817468245\_0819 running in uber mode : false  
23/01/04 16:08:49 INFO Job: map 0% reduce 0%  
23/01/04 16:08:55 INFO Job: map 100% reduce 0%  
23/01/04 16:09:01 INFO Job: map 100% reduce 100%  
23/01/04 16:09:01 INFO Job: Job job\_1672817468245\_0819 completed successfully  
23/01/04 16:09:01 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4177812  
FILE: Number of bytes written=8887915  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6410763  
HDFS: Number of bytes written=1905923  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0



#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18880  
Total time spent by all reduces in occupied slots (ms)=17715  
Total time spent by all map tasks (ms)=3776  
Total time spent by all reduce tasks (ms)=3543  
Total vcore-milliseconds taken by all map tasks=3776  
Total vcore-milliseconds taken by all reduce tasks=3543  
Total megabyte-milliseconds taken by all map tasks=19333120  
Total megabyte-milliseconds taken by all reduce tasks=18140160

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4177776  
Map output materialized bytes=4177812  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=4177812  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=4250  
Physical memory (bytes) snapshot=756699136  
Virtual memory (bytes) snapshot=12753625088  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=411303936  
Peak Map Virtual memory (bytes)=6368116736  
Peak Reduce Physical memory (bytes)=345395200  
Peak Reduce Virtual memory (bytes)=6385508352

#### 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=2680027

#### File Output Format Counters

Bytes Written=1905923

23/01/04 16:09:01 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:09:01 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0821

23/01/04 16:09:01 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:09:01 INFO JobSubmitter: number of splits:1

23/01/04 16:09:01 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0821

23/01/04 16:09:01 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:09:01 INFO YarnClientImpl: Submitted application application\_1672817468245\_0821

23/01/04 16:09:01 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0821/  
23/01/04 16:09:01 INFO Job: Running job: job\_1672817468245\_0821  
23/01/04 16:09:07 INFO Job: Job job\_1672817468245\_0821 running in uber mode : false  
23/01/04 16:09:07 INFO Job: map 0% reduce 0%  
23/01/04 16:09:13 INFO Job: map 100% reduce 0%  
23/01/04 16:09:19 INFO Job: map 100% reduce 100%  
23/01/04 16:09:19 INFO Job: Job job\_1672817468245\_0821 completed successfully  
23/01/04 16:09:19 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4198282  
FILE: Number of bytes written=8928855  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6492431  
HDFS: Number of bytes written=1885485  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19000  
Total time spent by all reduces in occupied slots (ms)=17410  
Total time spent by all map tasks (ms)=3800  
Total time spent by all reduce tasks (ms)=3482  
Total vcore-milliseconds taken by all map tasks=3800  
Total vcore-milliseconds taken by all reduce tasks=3482  
Total megabyte-milliseconds taken by all map tasks=19456000  
Total megabyte-milliseconds taken by all reduce tasks=17827840

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4198246  
Map output materialized bytes=4198282  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=4198282  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=4410  
Physical memory (bytes) snapshot=754970624  
Virtual memory (bytes) snapshot=12745826304  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410849280  
Peak Map Virtual memory (bytes)=6373273600  
Peak Reduce Physical memory (bytes)=344121344  
Peak Reduce Virtual memory (bytes)=6372552704

#### 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=2680027

File Output Format Counters

Bytes Written=1885485

23/01/04 16:09:19 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:09:19 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0824  
23/01/04 16:09:19 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:09:20 INFO JobSubmitter: number of splits:1  
23/01/04 16:09:20 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0824  
23/01/04 16:09:20 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:09:20 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0824  
23/01/04 16:09:20 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0824/  
23/01/04 16:09:20 INFO Job: Running job: job\_1672817468245\_0824  
23/01/04 16:09:26 INFO Job: Job job\_1672817468245\_0824 running in uber mode : false  
23/01/04 16:09:26 INFO Job: map 0% reduce 0%  
23/01/04 16:09:32 INFO Job: map 100% reduce 0%  
23/01/04 16:09:38 INFO Job: map 100% reduce 100%  
23/01/04 16:09:38 INFO Job: Job job\_1672817468245\_0824 completed successfully  
23/01/04 16:09:38 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4156840  
FILE: Number of bytes written=8845971  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6451555  
HDFS: Number of bytes written=1863824  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19120  
Total time spent by all reduces in occupied slots (ms)=17115  
Total time spent by all map tasks (ms)=3824  
Total time spent by all reduce tasks (ms)=3423  
Total vcore-milliseconds taken by all map tasks=3824  
Total vcore-milliseconds taken by all reduce tasks=3423  
Total megabyte-milliseconds taken by all map tasks=19578880  
Total megabyte-milliseconds taken by all reduce tasks=17525760

Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4156804  
Map output materialized bytes=4156840

Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=4156840  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=4840  
Physical memory (bytes) snapshot=769134592  
Virtual memory (bytes) snapshot=12763066368  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=420573184  
Peak Map Virtual memory (bytes)=6375505920  
Peak Reduce Physical memory (bytes)=348561408  
Peak Reduce Virtual memory (bytes)=6387560448

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=2680027

File Output Format Counters

Bytes Written=1863824

23/01/04 16:09:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:09:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0827  
23/01/04 16:09:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:09:38 INFO JobSubmitter: number of splits:1  
23/01/04 16:09:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0827  
23/01/04 16:09:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:09:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0827  
23/01/04 16:09:38 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0827/  
23/01/04 16:09:38 INFO Job: Running job: job\_1672817468245\_0827  
23/01/04 16:09:44 INFO Job: Job job\_1672817468245\_0827 running in uber mode : false  
23/01/04 16:09:44 INFO Job: map 0% reduce 0%  
23/01/04 16:09:50 INFO Job: map 100% reduce 0%  
23/01/04 16:09:54 INFO Job: map 100% reduce 100%  
23/01/04 16:09:55 INFO Job: Job job\_1672817468245\_0827 completed successfully  
23/01/04 16:09:55 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4135179  
FILE: Number of bytes written=8802649  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6408233  
HDFS: Number of bytes written=1863824

HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18765  
Total time spent by all reduces in occupied slots (ms)=10185  
Total time spent by all map tasks (ms)=3753  
Total time spent by all reduce tasks (ms)=2037  
Total vcore-milliseconds taken by all map tasks=3753  
Total vcore-milliseconds taken by all reduce tasks=2037  
Total megabyte-milliseconds taken by all map tasks=19215360  
Total megabyte-milliseconds taken by all reduce tasks=10429440

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=4135143  
Map output materialized bytes=4135179  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=4135179  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=77  
CPU time spent (ms)=4310  
Physical memory (bytes) snapshot=734150656  
Virtual memory (bytes) snapshot=12764016640  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=426196992  
Peak Map Virtual memory (bytes)=6382280704  
Peak Reduce Physical memory (bytes)=307953664  
Peak Reduce Virtual memory (bytes)=6381735936

#### 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=2680027

#### File Output Format Counters

Bytes Written=1863824

23/01/04 16:09:55 INFO KMeansDriver: Clustering data  
23/01/04 16:09:55 INFO KMeansDriver: Running Clustering  
23/01/04 16:09:55 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/06 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/

ManhattanDistanceMeasure/06

23/01/04 16:09:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:09:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0830  
23/01/04 16:09:56 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:09:56 INFO JobSubmitter: number of splits:1  
23/01/04 16:09:56 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0830  
23/01/04 16:09:56 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:09:56 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0830  
23/01/04 16:09:56 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0830/  
23/01/04 16:09:56 INFO Job: Running job: job\_1672817468245\_0830  
23/01/04 16:10:02 INFO Job: Job job\_1672817468245\_0830 running in uber mode : false  
23/01/04 16:10:02 INFO Job: map 0% reduce 0%  
23/01/04 16:10:08 INFO Job: map 100% reduce 0%  
23/01/04 16:10:08 INFO Job: Job job\_1672817468245\_0830 completed successfully  
23/01/04 16:10:08 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4544215  
HDFS: Number of bytes written=2680630  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18715  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3743  
Total vcore-milliseconds taken by all map tasks=3743  
Total megabyte-milliseconds taken by all map tasks=19164160

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=28  
CPU time spent (ms)=2710  
Physical memory (bytes) snapshot=313069568  
Virtual memory (bytes) snapshot=6378037248  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=313069568  
Peak Map Virtual memory (bytes)=6378037248

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680630

23/01/04 16:10:08 INFO MahoutDriver: Program took 124145 ms (Minutes:

```

2.0690833333333334)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:10:13 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/06/clusters-6-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/06.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu
re/06/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 16:10:15 INFO HadoopUtil: Deleting tmp/representative
23/01/04 16:10:15 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/06/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
06/clusters-6-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 16:10:15 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 16:10:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:10:15 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0833
23/01/04 16:10:15 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:10:16 INFO JobSubmitter: number of splits:1
23/01/04 16:10:16 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0833
23/01/04 16:10:16 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:10:16 INFO Configuration: resource-types.xml not found
23/01/04 16:10:16 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:10:16 INFO YarnClientImpl: Submitted application
application_1672817468245_0833
23/01/04 16:10:16 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0833/
23/01/04 16:10:16 INFO Job: Running job: job_1672817468245_0833
23/01/04 16:10:22 INFO Job: Job job_1672817468245_0833 running in uber mode : false
23/01/04 16:10:22 INFO Job: map 0% reduce 0%
23/01/04 16:10:28 INFO Job: map 100% reduce 0%
23/01/04 16:10:34 INFO Job: map 100% reduce 100%
23/01/04 16:10:34 INFO Job: Job job_1672817468245_0833 completed successfully
23/01/04 16:10:34 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=787694
        FILE: Number of bytes written=2107981
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4953011
        HDFS: Number of bytes written=1923867
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2

```

HDFS: Number of bytes read erasure-coded=0  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Rack-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=17090  
  Total time spent by all reduces in occupied slots (ms)=16540  
  Total time spent by all map tasks (ms)=3418  
  Total time spent by all reduce tasks (ms)=3308  
  Total vcore-milliseconds taken by all map tasks=3418  
  Total vcore-milliseconds taken by all reduce tasks=3308  
  Total megabyte-milliseconds taken by all map tasks=17500160  
  Total megabyte-milliseconds taken by all reduce tasks=16936960

Map-Reduce Framework  
  Map input records=27  
  Map output records=6  
  Map output bytes=787658  
  Map output materialized bytes=787694  
  Input split bytes=203  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=6  
  Reduce shuffle bytes=787694  
  Reduce input records=6  
  Reduce output records=12  
  Spilled Records=12  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=45  
  CPU time spent (ms)=2960  
  Physical memory (bytes) snapshot=705294336  
  Virtual memory (bytes) snapshot=12748926976  
  Total committed heap usage (bytes)=2105540608  
  Peak Map Physical memory (bytes)=414138368  
  Peak Map Virtual memory (bytes)=6373339136  
  Peak Reduce Physical memory (bytes)=291155968  
  Peak Reduce Virtual memory (bytes)=6375587840

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=2680630

File Output Format Counters

  Bytes Written=1923867

23/01/04 16:10:34 INFO RepresentativePointsDriver: Representative Points Iteration  
1  
23/01/04 16:10:34 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:10:34 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0836  
23/01/04 16:10:34 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:10:34 INFO JobSubmitter: number of splits:1  
23/01/04 16:10:34 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0836



23/01/04 16:10:34 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:10:34 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0836  
23/01/04 16:10:34 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0836/  
23/01/04 16:10:34 INFO Job: Running job: job\_1672817468245\_0836  
23/01/04 16:10:41 INFO Job: Job job\_1672817468245\_0836 running in uber mode : false  
23/01/04 16:10:41 INFO Job: map 0% reduce 0%  
23/01/04 16:10:47 INFO Job: map 100% reduce 0%  
23/01/04 16:10:52 INFO Job: map 100% reduce 100%  
23/01/04 16:10:53 INFO Job: Job job\_1672817468245\_0836 completed successfully  
23/01/04 16:10:53 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=724496  
FILE: Number of bytes written=1981585  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6528567  
HDFS: Number of bytes written=2648427  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17065  
Total time spent by all reduces in occupied slots (ms)=16855  
Total time spent by all map tasks (ms)=3413  
Total time spent by all reduce tasks (ms)=3371  
Total vcore-milliseconds taken by all map tasks=3413  
Total vcore-milliseconds taken by all reduce tasks=3371  
Total megabyte-milliseconds taken by all map tasks=17474560  
Total megabyte-milliseconds taken by all reduce tasks=17259520

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=724460  
Map output materialized bytes=724496  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=724496  
Reduce input records=6  
Reduce output records=18  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=65  
CPU time spent (ms)=3020  
Physical memory (bytes) snapshot=702631936  
Virtual memory (bytes) snapshot=12747685888  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408481792  
Peak Map Virtual memory (bytes)=6370435072

Peak Reduce Physical memory (bytes)=294150144  
Peak Reduce Virtual memory (bytes)=6377250816  
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=2680630  
File Output Format Counters  
Bytes Written=2648427

23/01/04 16:10:53 INFO RepresentativePointsDriver: Representative Points Iteration  
2  
23/01/04 16:10:53 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:10:53 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0839  
23/01/04 16:10:53 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:10:53 INFO JobSubmitter: number of splits:1  
23/01/04 16:10:53 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0839  
23/01/04 16:10:53 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:10:53 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0839  
23/01/04 16:10:53 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0839/](http://lena-master:8088/proxy/application_1672817468245_0839/)  
23/01/04 16:10:53 INFO Job: Running job: job\_1672817468245\_0839  
23/01/04 16:10:59 INFO Job: Job job\_1672817468245\_0839 running in uber mode : false  
23/01/04 16:10:59 INFO Job: map 0% reduce 0%  
23/01/04 16:11:06 INFO Job: map 100% reduce 0%  
23/01/04 16:11:12 INFO Job: map 100% reduce 100%  
23/01/04 16:11:12 INFO Job: Job job\_1672817468245\_0839 completed successfully  
23/01/04 16:11:12 INFO Job: Counters: 54

File System Counters  
FILE: Number of bytes read=780814  
FILE: Number of bytes written=2094221  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7977687  
HDFS: Number of bytes written=3429325  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17340  
Total time spent by all reduces in occupied slots (ms)=16360  
Total time spent by all map tasks (ms)=3468  
Total time spent by all reduce tasks (ms)=3272  
Total vcore-milliseconds taken by all map tasks=3468  
Total vcore-milliseconds taken by all reduce tasks=3272  
Total megabyte-milliseconds taken by all map tasks=17756160  
Total megabyte-milliseconds taken by all reduce tasks=16752640

## Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=780778  
Map output materialized bytes=780814  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=780814  
Reduce input records=6  
Reduce output records=24  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=36  
CPU time spent (ms)=3210  
Physical memory (bytes) snapshot=715968512  
Virtual memory (bytes) snapshot=12761759744  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413548544  
Peak Map Virtual memory (bytes)=6375071744  
Peak Reduce Physical memory (bytes)=302419968  
Peak Reduce Virtual memory (bytes)=6386688000

## 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=2680630

## File Output Format Counters

Bytes Written=3429325

23/01/04 16:11:12 INFO RepresentativePointsDriver: Representative Points Iteration 3  
23/01/04 16:11:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:11:12 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0842  
23/01/04 16:11:12 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:11:13 INFO JobSubmitter: number of splits:1  
23/01/04 16:11:13 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0842  
23/01/04 16:11:13 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:11:13 INFO YarnClientImpl: Submitted application application\_1672817468245\_0842  
23/01/04 16:11:13 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0842/  
23/01/04 16:11:13 INFO Job: Running job: job\_1672817468245\_0842  
23/01/04 16:11:19 INFO Job: Job job\_1672817468245\_0842 running in uber mode : false  
23/01/04 16:11:19 INFO Job: map 0% reduce 0%  
23/01/04 16:11:25 INFO Job: map 100% reduce 0%  
23/01/04 16:11:31 INFO Job: map 100% reduce 100%  
23/01/04 16:11:31 INFO Job: Job job\_1672817468245\_0842 completed successfully  
23/01/04 16:11:31 INFO Job: Counters: 54  
File System Counters

FILE: Number of bytes read=772254  
FILE: Number of bytes written=2077101  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9539483  
HDFS: Number of bytes written=4201663  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17645  
Total time spent by all reduces in occupied slots (ms)=16740  
Total time spent by all map tasks (ms)=3529  
Total time spent by all reduce tasks (ms)=3348  
Total vcore-milliseconds taken by all map tasks=3529  
Total vcore-milliseconds taken by all reduce tasks=3348  
Total megabyte-milliseconds taken by all map tasks=18068480  
Total megabyte-milliseconds taken by all reduce tasks=17141760

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=772218  
Map output materialized bytes=772254  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=772254  
Reduce input records=6  
Reduce output records=30  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=3300  
Physical memory (bytes) snapshot=710201344  
Virtual memory (bytes) snapshot=12753604608  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410439680  
Peak Map Virtual memory (bytes)=6372302848  
Peak Reduce Physical memory (bytes)=299761664  
Peak Reduce Virtual memory (bytes)=6381301760

#### 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=2680630

#### File Output Format Counters

Bytes Written=4201663

23/01/04 16:11:31 INFO RepresentativePointsDriver: Representative Points Iteration 4  
23/01/04 16:11:31 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:11:31 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0845  
23/01/04 16:11:31 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:11:31 INFO JobSubmitter: number of splits:1  
23/01/04 16:11:31 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0845  
23/01/04 16:11:31 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:11:31 INFO YarnClientImpl: Submitted application application\_1672817468245\_0845  
23/01/04 16:11:31 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0845/  
23/01/04 16:11:31 INFO Job: Running job: job\_1672817468245\_0845  
23/01/04 16:11:37 INFO Job: Job job\_1672817468245\_0845 running in uber mode : false  
23/01/04 16:11:37 INFO Job: map 0% reduce 0%  
23/01/04 16:11:43 INFO Job: map 100% reduce 0%  
23/01/04 16:11:48 INFO Job: map 100% reduce 100%  
23/01/04 16:11:48 INFO Job: Job job\_1672817468245\_0845 completed successfully  
23/01/04 16:11:48 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=787694  
FILE: Number of bytes written=2107981  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11084159  
HDFS: Number of bytes written=4989441  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17895  
Total time spent by all reduces in occupied slots (ms)=16515  
Total time spent by all map tasks (ms)=3579  
Total time spent by all reduce tasks (ms)=3303  
Total vcore-milliseconds taken by all map tasks=3579  
Total vcore-milliseconds taken by all reduce tasks=3303  
Total megabyte-milliseconds taken by all map tasks=18324480  
Total megabyte-milliseconds taken by all reduce tasks=16911360

#### Map-Reduce Framework

Map input records=27  
Map output records=6  
Map output bytes=787658  
Map output materialized bytes=787694  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=6  
Reduce shuffle bytes=787694  
Reduce input records=6  
Reduce output records=36  
Spilled Records=12

```

Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=50
CPU time spent (ms)=3560
Physical memory (bytes) snapshot=725880832
Virtual memory (bytes) snapshot=12743131136
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=422490112
Peak Map Virtual memory (bytes)=6367039488
Peak Reduce Physical memory (bytes)=303390720
Peak Reduce Virtual memory (bytes)=6376091648

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=2680630
File Output Format Counters
  Bytes Written=4989441
23/01/04 16:11:48 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.4345443114932381
23/01/04 16:11:48 INFO ClusterEvaluator: Intra-Cluster Density[22] =
0.5666666666666667
23/01/04 16:11:49 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN
23/01/04 16:11:49 INFO ClusterEvaluator: Intra-Cluster Density[2] =
0.7567517705573877
23/01/04 16:11:49 INFO ClusterEvaluator: Intra-Cluster Density[15] =
0.7595394108042938
23/01/04 16:11:49 INFO ClusterEvaluator: Intra-Cluster Density[17] = NaN
23/01/04 16:11:49 INFO ClusterEvaluator: Intra-Cluster Density[16] = NaN
23/01/04 16:11:49 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6943192826761161
23/01/04 16:11:50 INFO ClusterDumper: Wrote 6 clusters
23/01/04 16:11:50 INFO MahoutDriver: Program took 96884 ms (Minutes:
1.6147333333333334)
---- K: 07 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/07
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:11:55 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[7],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/07], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 16:11:56 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/07
23/01/04 16:11:56 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-

```

corpus-canopy-centroids/ManhattanDistanceMeasure  
23/01/04 16:11:56 INFO ZlibFactory: Successfully loaded & initialized native-zlib library  
23/01/04 16:11:56 INFO CodecPool: Got brand-new compressor [.deflate]  
23/01/04 16:11:56 INFO RandomSeedGenerator: Wrote 7 Klusters to  
dsm010/question\_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/  
part-randomSeed  
23/01/04 16:11:57 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-  
corpus-vectors/tf-vectors Clusters In: dsm010/question\_2/british-fiction-corpus-  
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/07  
23/01/04 16:11:57 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 16:11:57 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 16:11:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:11:57 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0848  
23/01/04 16:11:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:11:57 INFO JobSubmitter: number of splits:1  
23/01/04 16:11:58 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0848  
23/01/04 16:11:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:11:58 INFO Configuration: resource-types.xml not found  
23/01/04 16:11:58 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:11:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0848  
23/01/04 16:11:58 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0848/  
23/01/04 16:11:58 INFO Job: Running job: job\_1672817468245\_0848  
23/01/04 16:12:04 INFO Job: Job job\_1672817468245\_0848 running in uber mode : false  
23/01/04 16:12:04 INFO Job: map 0% reduce 0%  
23/01/04 16:12:10 INFO Job: map 100% reduce 0%  
23/01/04 16:12:14 INFO Job: map 100% reduce 100%  
23/01/04 16:12:15 INFO Job: Job job\_1672817468245\_0848 completed successfully  
23/01/04 16:12:15 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3290821  
FILE: Number of bytes written=7113933  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7815199  
HDFS: Number of bytes written=1866809  
HDFS: Number of read operations=43  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19170  
Total time spent by all reduces in occupied slots (ms)=10760  
Total time spent by all map tasks (ms)=3834  
Total time spent by all reduce tasks (ms)=2152  
Total vcore-milliseconds taken by all map tasks=3834  
Total vcore-milliseconds taken by all reduce tasks=2152  
Total megabyte-milliseconds taken by all map tasks=19630080

```

        Total megabyte-milliseconds taken by all reduce tasks=11018240
Map-Reduce Framework
    Map input records=27
    Map output records=7
    Map output bytes=3290780
    Map output materialized bytes=3290821
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=7
    Reduce shuffle bytes=3290821
    Reduce input records=7
    Reduce output records=7
    Spilled Records=14
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=56
    CPU time spent (ms)=4830
    Physical memory (bytes) snapshot=751149056
    Virtual memory (bytes) snapshot=12752306176
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=418607104
    Peak Map Virtual memory (bytes)=6366695424
    Peak Reduce Physical memory (bytes)=332541952
    Peak Reduce Virtual memory (bytes)=6385610752
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=2680027
File Output Format Counters
    Bytes Written=1866809
23/01/04 16:12:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:12:15 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0851
23/01/04 16:12:15 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:12:15 INFO JobSubmitter: number of splits:1
23/01/04 16:12:15 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0851
23/01/04 16:12:15 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:12:15 INFO YarnClientImpl: Submitted application
application_1672817468245_0851
23/01/04 16:12:15 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0851/
23/01/04 16:12:15 INFO Job: Running job: job_1672817468245_0851
23/01/04 16:12:21 INFO Job: Job job_1672817468245_0851 running in uber mode : false
23/01/04 16:12:21 INFO Job:  map 0% reduce 0%
23/01/04 16:12:27 INFO Job:  map 100% reduce 0%
23/01/04 16:12:38 INFO Job:  map 100% reduce 100%
23/01/04 16:12:38 INFO Job: Job job_1672817468245_0851 completed successfully
23/01/04 16:12:38 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=4300685

```



FILE: Number of bytes written=9133661  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6414203  
HDFS: Number of bytes written=1866809  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19170  
Total time spent by all reduces in occupied slots (ms)=39380  
Total time spent by all map tasks (ms)=3834  
Total time spent by all reduce tasks (ms)=7876  
Total vcore-milliseconds taken by all map tasks=3834  
Total vcore-milliseconds taken by all reduce tasks=7876  
Total megabyte-milliseconds taken by all map tasks=19630080  
Total megabyte-milliseconds taken by all reduce tasks=40325120

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=4300644  
Map output materialized bytes=4300685  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=4300685  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=4590  
Physical memory (bytes) snapshot=722169856  
Virtual memory (bytes) snapshot=12747214848  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=407298048  
Peak Map Virtual memory (bytes)=6371000320  
Peak Reduce Physical memory (bytes)=314871808  
Peak Reduce Virtual memory (bytes)=6376214528

#### 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=2680027

#### File Output Format Counters

Bytes Written=1866809

23/01/04 16:12:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:12:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0854  
23/01/04 16:12:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:12:38 INFO JobSubmitter: number of splits:1  
23/01/04 16:12:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0854  
23/01/04 16:12:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:12:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0854  
23/01/04 16:12:38 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0854/  
23/01/04 16:12:38 INFO Job: Running job: job\_1672817468245\_0854  
23/01/04 16:12:44 INFO Job: Job job\_1672817468245\_0854 running in uber mode : false  
23/01/04 16:12:44 INFO Job: map 0% reduce 0%  
23/01/04 16:12:49 INFO Job: map 100% reduce 0%  
23/01/04 16:12:55 INFO Job: map 100% reduce 100%  
23/01/04 16:12:55 INFO Job: Job job\_1672817468245\_0854 completed successfully  
23/01/04 16:12:55 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4245269  
FILE: Number of bytes written=9022829  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6414203  
HDFS: Number of bytes written=1848445  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19065  
Total time spent by all reduces in occupied slots (ms)=17180  
Total time spent by all map tasks (ms)=3813  
Total time spent by all reduce tasks (ms)=3436  
Total vcore-milliseconds taken by all map tasks=3813  
Total vcore-milliseconds taken by all reduce tasks=3436  
Total megabyte-milliseconds taken by all map tasks=19522560  
Total megabyte-milliseconds taken by all reduce tasks=17592320

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=4245228  
Map output materialized bytes=4245269  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=4245269  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=52
CPU time spent (ms)=4390
Physical memory (bytes) snapshot=720744448
Virtual memory (bytes) snapshot=12761284608
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=408051712
Peak Map Virtual memory (bytes)=6373687296
Peak Reduce Physical memory (bytes)=312692736
Peak Reduce Virtual memory (bytes)=6387597312
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=2680027
File Output Format Counters
  Bytes Written=1848445
23/01/04 16:12:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:12:55 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0856
23/01/04 16:12:56 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:12:56 INFO JobSubmitter: number of splits:1
23/01/04 16:12:56 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0856
23/01/04 16:12:56 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:12:56 INFO YarnClientImpl: Submitted application
application_1672817468245_0856
23/01/04 16:12:56 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0856/
23/01/04 16:12:56 INFO Job: Running job: job_1672817468245_0856
23/01/04 16:13:02 INFO Job: Job job_1672817468245_0856 running in uber mode : false
23/01/04 16:13:02 INFO Job:  map 0% reduce 0%
23/01/04 16:13:08 INFO Job:  map 100% reduce 0%
23/01/04 16:13:12 INFO Job:  map 100% reduce 100%
23/01/04 16:13:12 INFO Job: Job job_1672817468245_0856 completed successfully
23/01/04 16:13:12 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4226905
  FILE: Number of bytes written=8986101
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=6377475
  HDFS: Number of bytes written=1848445
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=18990
  Total time spent by all reduces in occupied slots (ms)=10075
  Total time spent by all map tasks (ms)=3798

```

Total time spent by all reduce tasks (ms)=2015  
Total vcore-milliseconds taken by all map tasks=3798  
Total vcore-milliseconds taken by all reduce tasks=2015  
Total megabyte-milliseconds taken by all map tasks=19445760  
Total megabyte-milliseconds taken by all reduce tasks=10316800

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=4226864  
Map output materialized bytes=4226905  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=4226905  
Reduce input records=7  
Reduce output records=7  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=76  
CPU time spent (ms)=4470  
Physical memory (bytes) snapshot=723492864  
Virtual memory (bytes) snapshot=12753969152  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408952832  
Peak Map Virtual memory (bytes)=6374928384  
Peak Reduce Physical memory (bytes)=314540032  
Peak Reduce Virtual memory (bytes)=6379040768

#### 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=2680027

#### File Output Format Counters

Bytes Written=1848445

23/01/04 16:13:12 INFO KMeansDriver: Clustering data  
23/01/04 16:13:12 INFO KMeansDriver: Running Clustering  
23/01/04 16:13:12 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/07 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/07  
23/01/04 16:13:12 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:13:12 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0859  
23/01/04 16:13:12 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:13:12 INFO JobSubmitter: number of splits:1  
23/01/04 16:13:12 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0859  
23/01/04 16:13:12 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:13:12 INFO YarnClientImpl: Submitted application

```

application_1672817468245_0859
23/01/04 16:13:12 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0859/
23/01/04 16:13:12 INFO Job: Running job: job_1672817468245_0859
23/01/04 16:13:18 INFO Job: Job job_1672817468245_0859 running in uber mode : false
23/01/04 16:13:18 INFO Job:  map 0% reduce 0%
23/01/04 16:13:24 INFO Job:  map 100% reduce 0%
23/01/04 16:13:24 INFO Job: Job job_1672817468245_0859 completed successfully
23/01/04 16:13:24 INFO Job: Counters: 33
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=265897
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=4528836
        HDFS: Number of bytes written=2680640
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=18850
        Total time spent by all reduces in occupied slots (ms)=0
        Total time spent by all map tasks (ms)=3770
        Total vcore-milliseconds taken by all map tasks=3770
        Total megabyte-milliseconds taken by all map tasks=19302400
    Map-Reduce Framework
        Map input records=27
        Map output records=27
        Input split bytes=170
        Spilled Records=0
        Failed Shuffles=0
        Merged Map outputs=0
        GC time elapsed (ms)=32
        CPU time spent (ms)=2720
        Physical memory (bytes) snapshot=310697984
        Virtual memory (bytes) snapshot=6372466688
        Total committed heap usage (bytes)=1052770304
        Peak Map Physical memory (bytes)=310697984
        Peak Map Virtual memory (bytes)=6372466688
    File Input Format Counters
        Bytes Read=2680027
    File Output Format Counters
        Bytes Written=2680640
23/01/04 16:13:24 INFO MahoutDriver: Program took 89092 ms (Minutes:
1.4848833333333333)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:13:30 INFO AbstractJob: Command line arguments:
[--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/07/clusters-4-final], --numWords=[20], --

```

```

output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/07.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/07/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 16:13:32 INFO HadoopUtil: Deleting tmp/representative
23/01/04 16:13:32 INFO AbstractJob: Command line arguments: {--
clusteredPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/07/clusteredPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
07/clusters-4-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 16:13:32 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 16:13:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:13:32 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0862
23/01/04 16:13:33 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:13:33 INFO JobSubmitter: number of splits:1
23/01/04 16:13:33 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0862
23/01/04 16:13:33 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:13:33 INFO Configuration: resource-types.xml not found
23/01/04 16:13:33 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:13:33 INFO YarnClientImpl: Submitted application
application_1672817468245_0862
23/01/04 16:13:33 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0862/
23/01/04 16:13:33 INFO Job: Running job: job_1672817468245_0862
23/01/04 16:13:39 INFO Job: Job job_1672817468245_0862 running in uber mode : false
23/01/04 16:13:39 INFO Job:  map 0% reduce 0%
23/01/04 16:13:44 INFO Job:  map 100% reduce 0%
23/01/04 16:13:50 INFO Job:  map 100% reduce 100%
23/01/04 16:13:50 INFO Job: Job job_1672817468245_0862 completed successfully
23/01/04 16:13:50 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=907322
        FILE: Number of bytes written=2347237
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5060223
        HDFS: Number of bytes written=2097111
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17240
        Total time spent by all reduces in occupied slots (ms)=16290
        Total time spent by all map tasks (ms)=3448
        Total time spent by all reduce tasks (ms)=3258
        Total vcore-milliseconds taken by all map tasks=3448
        Total vcore-milliseconds taken by all reduce tasks=3258

```

Total megabyte-milliseconds taken by all map tasks=17653760  
Total megabyte-milliseconds taken by all reduce tasks=16680960  
Map-Reduce Framework  
Map input records=27  
Map output records=7  
Map output bytes=907281  
Map output materialized bytes=907322  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=907322  
Reduce input records=7  
Reduce output records=14  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=34  
CPU time spent (ms)=2760  
Physical memory (bytes) snapshot=694116352  
Virtual memory (bytes) snapshot=12755853312  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=394932224  
Peak Map Virtual memory (bytes)=6369923072  
Peak Reduce Physical memory (bytes)=299184128  
Peak Reduce Virtual memory (bytes)=6387884032  
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=2680640  
File Output Format Counters  
Bytes Written=2097111

23/01/04 16:13:50 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 16:13:50 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:13:50 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0865  
23/01/04 16:13:50 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:13:51 INFO JobSubmitter: number of splits:1  
23/01/04 16:13:51 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0865  
23/01/04 16:13:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:13:51 INFO YarnClientImpl: Submitted application application\_1672817468245\_0865  
23/01/04 16:13:51 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0865/  
23/01/04 16:13:51 INFO Job: Running job: job\_1672817468245\_0865  
23/01/04 16:13:57 INFO Job: Job job\_1672817468245\_0865 running in uber mode : false  
23/01/04 16:13:57 INFO Job: map 0% reduce 0%  
23/01/04 16:14:03 INFO Job: map 100% reduce 0%  
23/01/04 16:14:08 INFO Job: map 100% reduce 100%  
23/01/04 16:14:09 INFO Job: Job job\_1672817468245\_0865 completed successfully

23/01/04 16:14:09 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=951198  
FILE: Number of bytes written=2434989  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6875065  
HDFS: Number of bytes written=3048408  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17040  
Total time spent by all reduces in occupied slots (ms)=16265  
Total time spent by all map tasks (ms)=3408  
Total time spent by all reduce tasks (ms)=3253  
Total vcore-milliseconds taken by all map tasks=3408  
Total vcore-milliseconds taken by all reduce tasks=3253  
Total megabyte-milliseconds taken by all map tasks=17448960  
Total megabyte-milliseconds taken by all reduce tasks=16655360

Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=951157  
Map output materialized bytes=951198  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=951198  
Reduce input records=7  
Reduce output records=21  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=56  
CPU time spent (ms)=3140  
Physical memory (bytes) snapshot=686002176  
Virtual memory (bytes) snapshot=12757082112  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=387018752  
Peak Map Virtual memory (bytes)=6378835968  
Peak Reduce Physical memory (bytes)=298983424  
Peak Reduce Virtual memory (bytes)=6386204672

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=2680640



#### File Output Format Counters

Bytes Written=3048408

23/01/04 16:14:09 INFO RepresentativePointsDriver: Representative Points Iteration 2  
23/01/04 16:14:09 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:14:09 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0868  
23/01/04 16:14:09 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:14:09 INFO JobSubmitter: number of splits:1  
23/01/04 16:14:09 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0868  
23/01/04 16:14:09 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:14:09 INFO YarnClientImpl: Submitted application application\_1672817468245\_0868  
23/01/04 16:14:09 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0868/  
23/01/04 16:14:09 INFO Job: Running job: job\_1672817468245\_0868  
23/01/04 16:14:15 INFO Job: Job job\_1672817468245\_0868 running in uber mode : false  
23/01/04 16:14:15 INFO Job: map 0% reduce 0%  
23/01/04 16:14:21 INFO Job: map 100% reduce 0%  
23/01/04 16:14:27 INFO Job: map 100% reduce 100%  
23/01/04 16:14:27 INFO Job: Job job\_1672817468245\_0868 completed successfully  
23/01/04 16:14:27 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=909201  
FILE: Number of bytes written=2350995  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8777659  
HDFS: Number of bytes written=3957708  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17275  
Total time spent by all reduces in occupied slots (ms)=16455  
Total time spent by all map tasks (ms)=3455  
Total time spent by all reduce tasks (ms)=3291  
Total vcore-milliseconds taken by all map tasks=3455  
Total vcore-milliseconds taken by all reduce tasks=3291  
Total megabyte-milliseconds taken by all map tasks=17689600  
Total megabyte-milliseconds taken by all reduce tasks=16849920

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=909160  
Map output materialized bytes=909201  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=909201  
Reduce input records=7

Reduce output records=28  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=3330  
Physical memory (bytes) snapshot=721817600  
Virtual memory (bytes) snapshot=12758294528  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=417583104  
Peak Map Virtual memory (bytes)=6375723008  
Peak Reduce Physical memory (bytes)=304234496  
Peak Reduce Virtual memory (bytes)=6382571520

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=2680640

File Output Format Counters

Bytes Written=3957708

23/01/04 16:14:27 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 16:14:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:14:27 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0871

23/01/04 16:14:28 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:14:28 INFO JobSubmitter: number of splits:1

23/01/04 16:14:28 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0871

23/01/04 16:14:28 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:14:28 INFO YarnClientImpl: Submitted application application\_1672817468245\_0871

23/01/04 16:14:28 INFO Job: The url to track the job:

http://lena-master:8088/proxy/application\_1672817468245\_0871/

23/01/04 16:14:28 INFO Job: Running job: job\_1672817468245\_0871

23/01/04 16:14:34 INFO Job: Job job\_1672817468245\_0871 running in uber mode : false

23/01/04 16:14:34 INFO Job: map 0% reduce 0%

23/01/04 16:14:40 INFO Job: map 100% reduce 0%

23/01/04 16:14:46 INFO Job: map 100% reduce 100%

23/01/04 16:14:46 INFO Job: Job job\_1672817468245\_0871 completed successfully

23/01/04 16:14:46 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=943996

FILE: Number of bytes written=2420585

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=10596259

HDFS: Number of bytes written=4901803

HDFS: Number of read operations=15

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17590  
Total time spent by all reduces in occupied slots (ms)=16510  
Total time spent by all map tasks (ms)=3518  
Total time spent by all reduce tasks (ms)=3302  
Total vcore-milliseconds taken by all map tasks=3518  
Total vcore-milliseconds taken by all reduce tasks=3302  
Total megabyte-milliseconds taken by all map tasks=18012160  
Total megabyte-milliseconds taken by all reduce tasks=16906240

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=943955  
Map output materialized bytes=943996  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=943996  
Reduce input records=7  
Reduce output records=35  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=40  
CPU time spent (ms)=3500  
Physical memory (bytes) snapshot=722423808  
Virtual memory (bytes) snapshot=12756393984  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=417677312  
Peak Map Virtual memory (bytes)=6378971136  
Peak Reduce Physical memory (bytes)=304746496  
Peak Reduce Virtual memory (bytes)=6377422848

#### 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=2680640

#### File Output Format Counters

Bytes Written=4901803

23/01/04 16:14:46 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 16:14:46 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:14:46 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0874

23/01/04 16:14:46 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:14:46 INFO JobSubmitter: number of splits:1

23/01/04 16:14:46 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0874

23/01/04 16:14:46 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:14:46 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0874  
23/01/04 16:14:46 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0874/  
23/01/04 16:14:46 INFO Job: Running job: job\_1672817468245\_0874  
23/01/04 16:14:52 INFO Job: Job job\_1672817468245\_0874 running in uber mode : false  
23/01/04 16:14:52 INFO Job: map 0% reduce 0%  
23/01/04 16:14:58 INFO Job: map 100% reduce 0%  
23/01/04 16:15:03 INFO Job: map 100% reduce 100%  
23/01/04 16:15:04 INFO Job: Job job\_1672817468245\_0874 completed successfully  
23/01/04 16:15:04 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=881350  
FILE: Number of bytes written=2295293  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=12484449  
HDFS: Number of bytes written=5783232  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17920  
Total time spent by all reduces in occupied slots (ms)=16385  
Total time spent by all map tasks (ms)=3584  
Total time spent by all reduce tasks (ms)=3277  
Total vcore-milliseconds taken by all map tasks=3584  
Total vcore-milliseconds taken by all reduce tasks=3277  
Total megabyte-milliseconds taken by all map tasks=18350080  
Total megabyte-milliseconds taken by all reduce tasks=16778240

#### Map-Reduce Framework

Map input records=27  
Map output records=7  
Map output bytes=881309  
Map output materialized bytes=881350  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=7  
Reduce shuffle bytes=881350  
Reduce input records=7  
Reduce output records=42  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=62  
CPU time spent (ms)=3520  
Physical memory (bytes) snapshot=691879936  
Virtual memory (bytes) snapshot=12741533696  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401604608  
Peak Map Virtual memory (bytes)=6369349632  
Peak Reduce Physical memory (bytes)=290275328

```

        Peak Reduce Virtual memory (bytes)=6372184064
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=2680640
File Output Format Counters
    Bytes Written=5783232
23/01/04 16:15:05 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.41485922032023387
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[11] =
0.5976114661652916
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[23] = NaN
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[22] = NaN
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[5] =
0.5666666666666667
23/01/04 16:15:05 INFO ClusterEvaluator: Intra-Cluster Density[18] = NaN
23/01/04 16:15:05 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.5821390664159791
23/01/04 16:15:07 INFO ClusterDumper: Wrote 7 clusters
23/01/04 16:15:07 INFO MahoutDriver: Program took 96600 ms (Minutes: 1.61)
---- K: 08 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/08
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:15:11 INFO AbstractJob: Command line arguments: {--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[8],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/08], --overwrite=null, --startPhase=[0], --
tempDir=[temp]}
23/01/04 16:15:12 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/08
23/01/04 16:15:12 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 16:15:12 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 16:15:12 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 16:15:13 INFO RandomSeedGenerator: Wrote 8 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 16:15:13 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/08

```

23/01/04 16:15:13 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 16:15:13 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 16:15:13 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:15:13 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0877  
23/01/04 16:15:13 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:15:13 INFO JobSubmitter: number of splits:1  
23/01/04 16:15:14 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0877  
23/01/04 16:15:14 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:15:14 INFO Configuration: resource-types.xml not found  
23/01/04 16:15:14 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:15:14 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0877  
23/01/04 16:15:14 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0877/  
23/01/04 16:15:14 INFO Job: Running job: job\_1672817468245\_0877  
23/01/04 16:15:20 INFO Job: Job job\_1672817468245\_0877 running in uber mode : false  
23/01/04 16:15:20 INFO Job: map 0% reduce 0%  
23/01/04 16:15:26 INFO Job: map 100% reduce 0%  
23/01/04 16:15:31 INFO Job: map 100% reduce 100%  
23/01/04 16:15:32 INFO Job: Job job\_1672817468245\_0877 completed successfully  
23/01/04 16:15:32 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3355430  
FILE: Number of bytes written=7243151  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7618369  
HDFS: Number of bytes written=2069687  
HDFS: Number of read operations=47  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19290  
Total time spent by all reduces in occupied slots (ms)=17535  
Total time spent by all map tasks (ms)=3858  
Total time spent by all reduce tasks (ms)=3507  
Total vcore-milliseconds taken by all map tasks=3858  
Total vcore-milliseconds taken by all reduce tasks=3507  
Total megabyte-milliseconds taken by all map tasks=19752960  
Total megabyte-milliseconds taken by all reduce tasks=17955840

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=3355384  
Map output materialized bytes=3355430  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=3355430  
Reduce input records=8

Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=4890  
Physical memory (bytes) snapshot=713736192  
Virtual memory (bytes) snapshot=12755816448  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=401072128  
Peak Map Virtual memory (bytes)=6375845888  
Peak Reduce Physical memory (bytes)=312664064  
Peak Reduce Virtual memory (bytes)=6379970560

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=2680027

File Output Format Counters

Bytes Written=2069687

23/01/04 16:15:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:15:32 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0880  
23/01/04 16:15:32 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:15:32 INFO JobSubmitter: number of splits:1  
23/01/04 16:15:32 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0880  
23/01/04 16:15:32 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:15:33 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0880  
23/01/04 16:15:33 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0880/](http://lena-master:8088/proxy/application_1672817468245_0880/)  
23/01/04 16:15:33 INFO Job: Running job: job\_1672817468245\_0880  
23/01/04 16:15:39 INFO Job: Job job\_1672817468245\_0880 running in uber mode : false  
23/01/04 16:15:39 INFO Job: map 0% reduce 0%  
23/01/04 16:15:45 INFO Job: map 100% reduce 0%  
23/01/04 16:15:51 INFO Job: map 100% reduce 100%  
23/01/04 16:15:51 INFO Job: Job job\_1672817468245\_0880 completed successfully  
23/01/04 16:15:51 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=4754334  
FILE: Number of bytes written=10040959  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=6819959  
HDFS: Number of bytes written=2315639  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=18885  
Total time spent by all reduces in occupied slots (ms)=17870  
Total time spent by all map tasks (ms)=3777  
Total time spent by all reduce tasks (ms)=3574  
Total vcore-milliseconds taken by all map tasks=3777  
Total vcore-milliseconds taken by all reduce tasks=3574  
Total megabyte-milliseconds taken by all map tasks=19338240  
Total megabyte-milliseconds taken by all reduce tasks=18298880

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4754288  
Map output materialized bytes=4754334  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4754334  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=69  
CPU time spent (ms)=4840  
Physical memory (bytes) snapshot=721838080  
Virtual memory (bytes) snapshot=12760989696  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402120704  
Peak Map Virtual memory (bytes)=6376792064  
Peak Reduce Physical memory (bytes)=319717376  
Peak Reduce Virtual memory (bytes)=6384197632

#### 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=2680027

#### File Output Format Counters

Bytes Written=2315639

23/01/04 16:15:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:15:51 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0883  
23/01/04 16:15:51 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:15:51 INFO JobSubmitter: number of splits:1  
23/01/04 16:15:51 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0883  
23/01/04 16:15:51 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:15:51 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0883  
23/01/04 16:15:51 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0883/](http://lena-master:8088/proxy/application_1672817468245_0883/)



23/01/04 16:15:51 INFO Job: Running job: job\_1672817468245\_0883  
23/01/04 16:15:57 INFO Job: Job job\_1672817468245\_0883 running in uber mode : false  
23/01/04 16:15:57 INFO Job: map 0% reduce 0%  
23/01/04 16:16:03 INFO Job: map 100% reduce 0%  
23/01/04 16:16:08 INFO Job: map 100% reduce 100%  
23/01/04 16:16:08 INFO Job: Job job\_1672817468245\_0883 completed successfully  
23/01/04 16:16:08 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4972262  
FILE: Number of bytes written=10476815  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7311863  
HDFS: Number of bytes written=2330732  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19365  
Total time spent by all reduces in occupied slots (ms)=10455  
Total time spent by all map tasks (ms)=3873  
Total time spent by all reduce tasks (ms)=2091  
Total vcore-milliseconds taken by all map tasks=3873  
Total vcore-milliseconds taken by all reduce tasks=2091  
Total megabyte-milliseconds taken by all map tasks=19829760  
Total megabyte-milliseconds taken by all reduce tasks=10705920

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=4972216  
Map output materialized bytes=4972262  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=4972262  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=4670  
Physical memory (bytes) snapshot=760619008  
Virtual memory (bytes) snapshot=12755193856  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=419491840  
Peak Map Virtual memory (bytes)=6376620032  
Peak Reduce Physical memory (bytes)=341127168  
Peak Reduce Virtual memory (bytes)=6378573824

#### 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=2680027
File Output Format Counters
    Bytes Written=2330732
23/01/04 16:16:08 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:16:08 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0886
23/01/04 16:16:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:16:09 INFO JobSubmitter: number of splits:1
23/01/04 16:16:09 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0886
23/01/04 16:16:09 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:16:09 INFO YarnClientImpl: Submitted application
application_1672817468245_0886
23/01/04 16:16:09 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0886/
23/01/04 16:16:09 INFO Job: Running job: job_1672817468245_0886
23/01/04 16:16:15 INFO Job: Job job_1672817468245_0886 running in uber mode : false
23/01/04 16:16:15 INFO Job:  map 0% reduce 0%
23/01/04 16:16:21 INFO Job:  map 100% reduce 0%
23/01/04 16:16:26 INFO Job:  map 100% reduce 100%
23/01/04 16:16:27 INFO Job: Job job_1672817468245_0886 completed successfully
23/01/04 16:16:27 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=5015349
    FILE: Number of bytes written=10562989
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7342049
    HDFS: Number of bytes written=2441328
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=19150
    Total time spent by all reduces in occupied slots (ms)=17500
    Total time spent by all map tasks (ms)=3830
    Total time spent by all reduce tasks (ms)=3500
    Total vcore-milliseconds taken by all map tasks=3830
    Total vcore-milliseconds taken by all reduce tasks=3500
    Total megabyte-milliseconds taken by all map tasks=19609600
    Total megabyte-milliseconds taken by all reduce tasks=17920000
Map-Reduce Framework
    Map input records=27
    Map output records=8
    Map output bytes=5015303
    Map output materialized bytes=5015349
    Input split bytes=170
    Combine input records=0

```

Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=5015349  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=51  
CPU time spent (ms)=4550  
Physical memory (bytes) snapshot=741363712  
Virtual memory (bytes) snapshot=12763271168  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=416817152  
Peak Map Virtual memory (bytes)=6381002752  
Peak Reduce Physical memory (bytes)=324546560  
Peak Reduce Virtual memory (bytes)=6382268416

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=2680027

File Output Format Counters

Bytes Written=2441328

23/01/04 16:16:27 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:16:27 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0889  
23/01/04 16:16:27 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:16:27 INFO JobSubmitter: number of splits:1  
23/01/04 16:16:27 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0889  
23/01/04 16:16:27 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:16:27 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0889  
23/01/04 16:16:27 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0889/  
23/01/04 16:16:27 INFO Job: Running job: job\_1672817468245\_0889  
23/01/04 16:16:33 INFO Job: Job job\_1672817468245\_0889 running in uber mode : false  
23/01/04 16:16:33 INFO Job: map 0% reduce 0%  
23/01/04 16:16:39 INFO Job: map 100% reduce 0%  
23/01/04 16:16:46 INFO Job: map 100% reduce 100%  
23/01/04 16:16:46 INFO Job: Job job\_1672817468245\_0889 completed successfully  
23/01/04 16:16:46 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=5282163  
FILE: Number of bytes written=11096617  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7563241  
HDFS: Number of bytes written=2602416  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19560  
Total time spent by all reduces in occupied slots (ms)=17785  
Total time spent by all map tasks (ms)=3912  
Total time spent by all reduce tasks (ms)=3557  
Total vcore-milliseconds taken by all map tasks=3912  
Total vcore-milliseconds taken by all reduce tasks=3557  
Total megabyte-milliseconds taken by all map tasks=20029440  
Total megabyte-milliseconds taken by all reduce tasks=18211840

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=5282117  
Map output materialized bytes=5282163  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=5282163  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=4740  
Physical memory (bytes) snapshot=753754112  
Virtual memory (bytes) snapshot=12757573632  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413933568  
Peak Map Virtual memory (bytes)=6376488960  
Peak Reduce Physical memory (bytes)=339820544  
Peak Reduce Virtual memory (bytes)=6381084672

#### 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=2680027

#### File Output Format Counters

Bytes Written=2602416

23/01/04 16:16:46 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:16:46 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0892

23/01/04 16:16:47 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:16:47 INFO JobSubmitter: number of splits:1

23/01/04 16:16:47 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0892

23/01/04 16:16:47 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:16:47 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0892  
23/01/04 16:16:47 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0892/  
23/01/04 16:16:47 INFO Job: Running job: job\_1672817468245\_0892  
23/01/04 16:16:54 INFO Job: Job job\_1672817468245\_0892 running in uber mode : false  
23/01/04 16:16:54 INFO Job: map 0% reduce 0%  
23/01/04 16:16:59 INFO Job: map 100% reduce 0%  
23/01/04 16:17:05 INFO Job: map 100% reduce 100%  
23/01/04 16:17:05 INFO Job: Job job\_1672817468245\_0892 completed successfully  
23/01/04 16:17:05 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5420267  
FILE: Number of bytes written=11372825  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7885417  
HDFS: Number of bytes written=2577983  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19300  
Total time spent by all reduces in occupied slots (ms)=17620  
Total time spent by all map tasks (ms)=3860  
Total time spent by all reduce tasks (ms)=3524  
Total vcore-milliseconds taken by all map tasks=3860  
Total vcore-milliseconds taken by all reduce tasks=3524  
Total megabyte-milliseconds taken by all map tasks=19763200  
Total megabyte-milliseconds taken by all reduce tasks=18042880

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=5420221  
Map output materialized bytes=5420267  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=5420267  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=45  
CPU time spent (ms)=4620  
Physical memory (bytes) snapshot=755171328  
Virtual memory (bytes) snapshot=12764303360  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=433057792  
Peak Map Virtual memory (bytes)=6384787456  
Peak Reduce Physical memory (bytes)=322113536

Peak Reduce Virtual memory (bytes)=6379515904  
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=2680027

File Output Format Counters

Bytes Written=2577983

23/01/04 16:17:05 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:17:05 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0894  
23/01/04 16:17:05 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:17:05 INFO JobSubmitter: number of splits:1  
23/01/04 16:17:05 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0894  
23/01/04 16:17:05 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:17:05 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0894  
23/01/04 16:17:05 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0894/](http://lena-master:8088/proxy/application_1672817468245_0894/)  
23/01/04 16:17:05 INFO Job: Running job: job\_1672817468245\_0894  
23/01/04 16:17:11 INFO Job: Job job\_1672817468245\_0894 running in uber mode : false  
23/01/04 16:17:11 INFO Job: map 0% reduce 0%  
23/01/04 16:17:17 INFO Job: map 100% reduce 0%  
23/01/04 16:17:24 INFO Job: map 100% reduce 100%  
23/01/04 16:17:24 INFO Job: Job job\_1672817468245\_0894 completed successfully  
23/01/04 16:17:24 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=5395834  
FILE: Number of bytes written=11323959  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7836551  
HDFS: Number of bytes written=2577983  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19410  
Total time spent by all reduces in occupied slots (ms)=18025  
Total time spent by all map tasks (ms)=3882  
Total time spent by all reduce tasks (ms)=3605  
Total vcore-milliseconds taken by all map tasks=3882  
Total vcore-milliseconds taken by all reduce tasks=3605  
Total megabyte-milliseconds taken by all map tasks=19875840  
Total megabyte-milliseconds taken by all reduce tasks=18457600

Map-Reduce Framework

Map input records=27  
Map output records=8

Map output bytes=5395788  
Map output materialized bytes=5395834  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=5395834  
Reduce input records=8  
Reduce output records=8  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=94  
CPU time spent (ms)=4590  
Physical memory (bytes) snapshot=733949952  
Virtual memory (bytes) snapshot=12761583616  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415735808  
Peak Map Virtual memory (bytes)=6377627648  
Peak Reduce Physical memory (bytes)=318214144  
Peak Reduce Virtual memory (bytes)=6383955968

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=2680027

File Output Format Counters

Bytes Written=2577983

23/01/04 16:17:24 INFO KMeansDriver: Clustering data  
23/01/04 16:17:24 INFO KMeansDriver: Running Clustering  
23/01/04 16:17:24 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/08 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/08  
23/01/04 16:17:25 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:17:25 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0897  
23/01/04 16:17:25 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:17:25 INFO JobSubmitter: number of splits:1  
23/01/04 16:17:25 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0897  
23/01/04 16:17:25 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:17:25 INFO YarnClientImpl: Submitted application application\_1672817468245\_0897  
23/01/04 16:17:25 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0897/  
23/01/04 16:17:25 INFO Job: Running job: job\_1672817468245\_0897  
23/01/04 16:17:31 INFO Job: Job job\_1672817468245\_0897 running in uber mode : false  
23/01/04 16:17:31 INFO Job: map 0% reduce 0%  
23/01/04 16:17:37 INFO Job: map 100% reduce 0%  
23/01/04 16:17:37 INFO Job: Job job\_1672817468245\_0897 completed successfully

23/01/04 16:17:37 INFO Job: Counters: 33

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5258374  
HDFS: Number of bytes written=2680613  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19110  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3822  
Total vcore-milliseconds taken by all map tasks=3822  
Total megabyte-milliseconds taken by all map tasks=19568640

Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=19  
CPU time spent (ms)=2440  
Physical memory (bytes) snapshot=320557056  
Virtual memory (bytes) snapshot=6378819584  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=320557056  
Peak Map Virtual memory (bytes)=6378819584

File Input Format Counters

Bytes Read=2680027

File Output Format Counters

Bytes Written=2680613

23/01/04 16:17:37 INFO MahoutDriver: Program took 145511 ms (Minutes: 2.4251833333333335)

MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.

Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop

MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar

23/01/04 16:17:43 INFO AbstractJob: Command line arguments:

{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --  
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/ManhattanDistanceMeasure/08/clusters-7-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
ManhattanDistanceMeasure/08.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu  
re/08/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}

23/01/04 16:17:44 INFO HadoopUtil: Deleting tmp/representative

23/01/04 16:17:44 INFO AbstractJob: Command line arguments: {--  
clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/08/clusteredPoints], --



```
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
08/clusters-7-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 16:17:45 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 16:17:45 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:17:45 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0900
23/01/04 16:17:45 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:17:45 INFO JobSubmitter: number of splits:1
23/01/04 16:17:45 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0900
23/01/04 16:17:45 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:17:45 INFO Configuration: resource-types.xml not found
23/01/04 16:17:45 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:17:45 INFO YarnClientImpl: Submitted application
application_1672817468245_0900
23/01/04 16:17:46 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0900/
23/01/04 16:17:46 INFO Job: Running job: job_1672817468245_0900
23/01/04 16:17:52 INFO Job: Job job_1672817468245_0900 running in uber mode : false
23/01/04 16:17:52 INFO Job:  map 0% reduce 0%
23/01/04 16:17:58 INFO Job:  map 100% reduce 0%
23/01/04 16:18:03 INFO Job:  map 100% reduce 100%
23/01/04 16:18:04 INFO Job: Job job_1672817468245_0900 completed successfully
23/01/04 16:18:04 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=936078
        FILE: Number of bytes written=2404749
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5499684
        HDFS: Number of bytes written=2345586
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=16810
        Total time spent by all reduces in occupied slots (ms)=16705
        Total time spent by all map tasks (ms)=3362
        Total time spent by all reduce tasks (ms)=3341
        Total vcore-milliseconds taken by all map tasks=3362
        Total vcore-milliseconds taken by all reduce tasks=3341
        Total megabyte-milliseconds taken by all map tasks=17213440
        Total megabyte-milliseconds taken by all reduce tasks=17105920
    Map-Reduce Framework
        Map input records=27
        Map output records=8
        Map output bytes=936032
        Map output materialized bytes=936078
        Input split bytes=203
```

Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=936078  
Reduce input records=8  
Reduce output records=16  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=2860  
Physical memory (bytes) snapshot=698241024  
Virtual memory (bytes) snapshot=12759007232  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=396787712  
Peak Map Virtual memory (bytes)=6370512896  
Peak Reduce Physical memory (bytes)=301453312  
Peak Reduce Virtual memory (bytes)=6390226944

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=2680613

File Output Format Counters

Bytes Written=2345586

23/01/04 16:18:04 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 16:18:04 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:18:04 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0903  
23/01/04 16:18:04 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:18:04 INFO JobSubmitter: number of splits:1  
23/01/04 16:18:04 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0903  
23/01/04 16:18:04 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:18:04 INFO YarnClientImpl: Submitted application application\_1672817468245\_0903  
23/01/04 16:18:04 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0903/  
23/01/04 16:18:04 INFO Job: Running job: job\_1672817468245\_0903  
23/01/04 16:18:10 INFO Job: Job job\_1672817468245\_0903 running in uber mode : false  
23/01/04 16:18:10 INFO Job: map 0% reduce 0%  
23/01/04 16:18:15 INFO Job: map 100% reduce 0%  
23/01/04 16:18:21 INFO Job: map 100% reduce 100%  
23/01/04 16:18:21 INFO Job: Job job\_1672817468245\_0903 completed successfully  
23/01/04 16:18:21 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=971309  
FILE: Number of bytes written=2475211  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7371988

HDFS: Number of bytes written=3316989  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17370  
Total time spent by all reduces in occupied slots (ms)=16730  
Total time spent by all map tasks (ms)=3474  
Total time spent by all reduce tasks (ms)=3346  
Total vcore-milliseconds taken by all map tasks=3474  
Total vcore-milliseconds taken by all reduce tasks=3346  
Total megabyte-milliseconds taken by all map tasks=17786880  
Total megabyte-milliseconds taken by all reduce tasks=17131520

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=971263  
Map output materialized bytes=971309  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=971309  
Reduce input records=8  
Reduce output records=24  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=78  
CPU time spent (ms)=3310  
Physical memory (bytes) snapshot=689504256  
Virtual memory (bytes) snapshot=12751204352  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=389840896  
Peak Map Virtual memory (bytes)=6373261312  
Peak Reduce Physical memory (bytes)=299663360  
Peak Reduce Virtual memory (bytes)=6381858816

#### 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=2680613

#### File Output Format Counters

Bytes Written=3316989

23/01/04 16:18:21 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 16:18:21 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:18:21 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0906

23/01/04 16:18:22 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:18:22 INFO JobSubmitter: number of splits:1  
23/01/04 16:18:22 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0906  
23/01/04 16:18:22 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:18:22 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0906  
23/01/04 16:18:22 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0906/](http://lena-master:8088/proxy/application_1672817468245_0906/)  
23/01/04 16:18:22 INFO Job: Running job: job\_1672817468245\_0906  
23/01/04 16:18:28 INFO Job: Job job\_1672817468245\_0906 running in uber mode : false  
23/01/04 16:18:28 INFO Job: map 0% reduce 0%  
23/01/04 16:18:34 INFO Job: map 100% reduce 0%  
23/01/04 16:18:39 INFO Job: map 100% reduce 100%  
23/01/04 16:18:40 INFO Job: Job job\_1672817468245\_0906 completed successfully  
23/01/04 16:18:40 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=943572  
FILE: Number of bytes written=2419737  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9314794  
HDFS: Number of bytes written=4260655  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17225  
Total time spent by all reduces in occupied slots (ms)=16965  
Total time spent by all map tasks (ms)=3445  
Total time spent by all reduce tasks (ms)=3393  
Total vcore-milliseconds taken by all map tasks=3445  
Total vcore-milliseconds taken by all reduce tasks=3393  
Total megabyte-milliseconds taken by all map tasks=17638400  
Total megabyte-milliseconds taken by all reduce tasks=17372160

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=943526  
Map output materialized bytes=943572  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=943572  
Reduce input records=8  
Reduce output records=32  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=3220  
Physical memory (bytes) snapshot=702164992

Virtual memory (bytes) snapshot=12763914240  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=405499904  
Peak Map Virtual memory (bytes)=6378041344  
Peak Reduce Physical memory (bytes)=296665088  
Peak Reduce Virtual memory (bytes)=6385872896

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=2680613

File Output Format Counters

Bytes Written=4260655

23/01/04 16:18:40 INFO RepresentativePointsDriver: Representative Points Iteration  
3

23/01/04 16:18:40 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:18:40 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0909

23/01/04 16:18:40 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:18:40 INFO JobSubmitter: number of splits:1

23/01/04 16:18:40 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0909

23/01/04 16:18:40 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:18:40 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0909

23/01/04 16:18:40 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0909/](http://lena-master:8088/proxy/application_1672817468245_0909/)

23/01/04 16:18:40 INFO Job: Running job: job\_1672817468245\_0909

23/01/04 16:18:46 INFO Job: Job job\_1672817468245\_0909 running in uber mode : false

23/01/04 16:18:46 INFO Job: map 0% reduce 0%

23/01/04 16:18:52 INFO Job: map 100% reduce 0%

23/01/04 16:18:57 INFO Job: map 100% reduce 100%

23/01/04 16:18:57 INFO Job: Job job\_1672817468245\_0909 completed successfully

23/01/04 16:18:57 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=946061  
FILE: Number of bytes written=2424715  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11202126  
HDFS: Number of bytes written=5206790  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17620  
Total time spent by all reduces in occupied slots (ms)=16840  
Total time spent by all map tasks (ms)=3524  
Total time spent by all reduce tasks (ms)=3368

Total vcore-milliseconds taken by all map tasks=3524  
Total vcore-milliseconds taken by all reduce tasks=3368  
Total megabyte-milliseconds taken by all map tasks=18042880  
Total megabyte-milliseconds taken by all reduce tasks=17244160

Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=946015  
Map output materialized bytes=946061  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=946061  
Reduce input records=8  
Reduce output records=40  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=54  
CPU time spent (ms)=3360  
Physical memory (bytes) snapshot=728350720  
Virtual memory (bytes) snapshot=12746416128  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=427483136  
Peak Map Virtual memory (bytes)=6372503552  
Peak Reduce Physical memory (bytes)=300867584  
Peak Reduce Virtual memory (bytes)=6373912576

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=2680613

File Output Format Counters

Bytes Written=5206790

23/01/04 16:18:57 INFO RepresentativePointsDriver: Representative Points Iteration 4

23/01/04 16:18:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:18:57 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0910

23/01/04 16:18:57 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:18:58 INFO JobSubmitter: number of splits:1

23/01/04 16:18:58 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0910

23/01/04 16:18:58 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:18:58 INFO YarnClientImpl: Submitted application application\_1672817468245\_0910

23/01/04 16:18:58 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0910/

23/01/04 16:18:58 INFO Job: Running job: job\_1672817468245\_0910

23/01/04 16:19:04 INFO Job: Job job\_1672817468245\_0910 running in uber mode : false

23/01/04 16:19:04 INFO Job: map 0% reduce 0%

23/01/04 16:19:10 INFO Job: map 100% reduce 0%

23/01/04 16:19:16 INFO Job: map 100% reduce 100%  
23/01/04 16:19:16 INFO Job: Job job\_1672817468245\_0910 completed successfully  
23/01/04 16:19:16 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=889675  
FILE: Number of bytes written=2311943  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=13094396  
HDFS: Number of bytes written=6096580  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17730  
Total time spent by all reduces in occupied slots (ms)=16955  
Total time spent by all map tasks (ms)=3546  
Total time spent by all reduce tasks (ms)=3391  
Total vcore-milliseconds taken by all map tasks=3546  
Total vcore-milliseconds taken by all reduce tasks=3391  
Total megabyte-milliseconds taken by all map tasks=18155520  
Total megabyte-milliseconds taken by all reduce tasks=17361920

#### Map-Reduce Framework

Map input records=27  
Map output records=8  
Map output bytes=889630  
Map output materialized bytes=889675  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=8  
Reduce shuffle bytes=889675  
Reduce input records=8  
Reduce output records=48  
Spilled Records=16  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=63  
CPU time spent (ms)=3500  
Physical memory (bytes) snapshot=702349312  
Virtual memory (bytes) snapshot=12744818688  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408530944  
Peak Map Virtual memory (bytes)=6365720576  
Peak Reduce Physical memory (bytes)=293818368  
Peak Reduce Virtual memory (bytes)=6379098112

#### 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=2680613
File Output Format Counters
  Bytes Written=6096580
23/01/04 16:19:16 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.3427141634209242
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[0] =
0.757986995838954
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[26] =
0.5666666666666667
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[21] =
0.5159728440331924
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[23] =
0.5666666666666667
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[16] = NaN
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[6] =
0.6580053441178733
23/01/04 16:19:16 INFO ClusterEvaluator: Intra-Cluster Density[7] = NaN
23/01/04 16:19:16 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6130597034646705
23/01/04 16:19:19 INFO ClusterDumper: Wrote 8 clusters
23/01/04 16:19:19 INFO MahoutDriver: Program took 96321 ms (Minutes: 1.60535)
---- K: 09 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/09
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:19:24 INFO AbstractJob: Command line arguments: [--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[9],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/09], --overwrite=null, --startPhase=[0], --
tempDir=[temp]]
23/01/04 16:19:24 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/09
23/01/04 16:19:24 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-
corpus-canopy-centroids/ManhattanDistanceMeasure
23/01/04 16:19:25 INFO ZlibFactory: Successfully loaded & initialized native-zlib
library
23/01/04 16:19:25 INFO CodecPool: Got brand-new compressor [.deflate]
23/01/04 16:19:25 INFO RandomSeedGenerator: Wrote 9 Klusters to
dsm010/question_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/
part-randomSeed
23/01/04 16:19:25 INFO KMeansDriver: Input: dsm010/question_2/british-fiction-
corpus-vectors/tf-vectors Clusters In: dsm010/question_2/british-fiction-corpus-
canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out:
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/09
23/01/04 16:19:25 INFO KMeansDriver: convergence: 0.1 max Iterations: 20
23/01/04 16:19:25 INFO CodecPool: Got brand-new decompressor [.deflate]
23/01/04 16:19:25 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032

```



23/01/04 16:19:25 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0915  
23/01/04 16:19:26 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:19:26 INFO JobSubmitter: number of splits:1  
23/01/04 16:19:26 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0915  
23/01/04 16:19:26 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:19:26 INFO Configuration: resource-types.xml not found  
23/01/04 16:19:26 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:19:26 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0915  
23/01/04 16:19:26 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0915/  
23/01/04 16:19:26 INFO Job: Running job: job\_1672817468245\_0915  
23/01/04 16:19:32 INFO Job: Job job\_1672817468245\_0915 running in uber mode : false  
23/01/04 16:19:32 INFO Job: map 0% reduce 0%  
23/01/04 16:19:38 INFO Job: map 100% reduce 0%  
23/01/04 16:19:42 INFO Job: map 100% reduce 100%  
23/01/04 16:19:42 INFO Job: Job job\_1672817468245\_0915 completed successfully  
23/01/04 16:19:42 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=3349952  
FILE: Number of bytes written=7232195  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7150117  
HDFS: Number of bytes written=2166968  
HDFS: Number of read operations=51  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19320  
Total time spent by all reduces in occupied slots (ms)=9950  
Total time spent by all map tasks (ms)=3864  
Total time spent by all reduce tasks (ms)=1990  
Total vcore-milliseconds taken by all map tasks=3864  
Total vcore-milliseconds taken by all reduce tasks=1990  
Total megabyte-milliseconds taken by all map tasks=19783680  
Total megabyte-milliseconds taken by all reduce tasks=10188800

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=3349901  
Map output materialized bytes=3349952  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=3349952  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0

```

Merged Map outputs=1
GC time elapsed (ms)=38
CPU time spent (ms)=4780
Physical memory (bytes) snapshot=741593088
Virtual memory (bytes) snapshot=12750725120
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=407416832
Peak Map Virtual memory (bytes)=6371131392
Peak Reduce Physical memory (bytes)=334176256
Peak Reduce Virtual memory (bytes)=6379593728
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=2680027
File Output Format Counters
  Bytes Written=2166968
23/01/04 16:19:42 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:19:42 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0918
23/01/04 16:19:42 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:19:43 INFO JobSubmitter: number of splits:1
23/01/04 16:19:43 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0918
23/01/04 16:19:43 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:19:43 INFO YarnClientImpl: Submitted application
application_1672817468245_0918
23/01/04 16:19:43 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0918/
23/01/04 16:19:43 INFO Job: Running job: job_1672817468245_0918
23/01/04 16:19:49 INFO Job: Job job_1672817468245_0918 running in uber mode : false
23/01/04 16:19:49 INFO Job:  map 0% reduce 0%
23/01/04 16:19:55 INFO Job:  map 100% reduce 0%
23/01/04 16:20:02 INFO Job:  map 100% reduce 100%
23/01/04 16:20:02 INFO Job: Job job_1672817468245_0918 completed successfully
23/01/04 16:20:02 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=4902052
  FILE: Number of bytes written=10336395
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=7014521
  HDFS: Number of bytes written=2351572
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=19705
  Total time spent by all reduces in occupied slots (ms)=18260

```

Total time spent by all map tasks (ms)=3941  
Total time spent by all reduce tasks (ms)=3652  
Total vcore-milliseconds taken by all map tasks=3941  
Total vcore-milliseconds taken by all reduce tasks=3652  
Total megabyte-milliseconds taken by all map tasks=20177920  
Total megabyte-milliseconds taken by all reduce tasks=18698240

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=4902001  
Map output materialized bytes=4902052  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=4902052  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=66  
CPU time spent (ms)=4410  
Physical memory (bytes) snapshot=720781312  
Virtual memory (bytes) snapshot=12746117120  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=413155328  
Peak Map Virtual memory (bytes)=6370291712  
Peak Reduce Physical memory (bytes)=307625984  
Peak Reduce Virtual memory (bytes)=6375825408

#### 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=2680027

#### File Output Format Counters

Bytes Written=2351572

23/01/04 16:20:02 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:20:02 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0921  
23/01/04 16:20:02 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:20:02 INFO JobSubmitter: number of splits:1  
23/01/04 16:20:02 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0921  
23/01/04 16:20:02 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:20:02 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0921  
23/01/04 16:20:02 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0921/  
23/01/04 16:20:02 INFO Job: Running job: job\_1672817468245\_0921  
23/01/04 16:20:08 INFO Job: Job job\_1672817468245\_0921 running in uber mode : false  
23/01/04 16:20:08 INFO Job: map 0% reduce 0%  
23/01/04 16:20:14 INFO Job: map 100% reduce 0%

23/01/04 16:20:21 INFO Job: map 100% reduce 100%  
23/01/04 16:20:21 INFO Job: Job job\_1672817468245\_0921 completed successfully  
23/01/04 16:20:21 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5141254  
FILE: Number of bytes written=10814799  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7383729  
HDFS: Number of bytes written=2400589  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19595  
Total time spent by all reduces in occupied slots (ms)=17970  
Total time spent by all map tasks (ms)=3919  
Total time spent by all reduce tasks (ms)=3594  
Total vcore-milliseconds taken by all map tasks=3919  
Total vcore-milliseconds taken by all reduce tasks=3594  
Total megabyte-milliseconds taken by all map tasks=20065280  
Total megabyte-milliseconds taken by all reduce tasks=18401280

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5141203  
Map output materialized bytes=5141254  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5141254  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=57  
CPU time spent (ms)=4440  
Physical memory (bytes) snapshot=739106816  
Virtual memory (bytes) snapshot=12755165184  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=423866368  
Peak Map Virtual memory (bytes)=6371258368  
Peak Reduce Physical memory (bytes)=315240448  
Peak Reduce Virtual memory (bytes)=6383906816

#### 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=2680027

#### File Output Format Counters

Bytes Written=2400589

23/01/04 16:20:21 INFO DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:20:21 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0924  
23/01/04 16:20:22 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:20:22 INFO JobSubmitter: number of splits:1  
23/01/04 16:20:22 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0924  
23/01/04 16:20:22 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:20:22 INFO YarnClientImpl: Submitted application application\_1672817468245\_0924  
23/01/04 16:20:22 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0924/  
23/01/04 16:20:22 INFO Job: Running job: job\_1672817468245\_0924  
23/01/04 16:20:28 INFO Job: Job job\_1672817468245\_0924 running in uber mode : false  
23/01/04 16:20:28 INFO Job: map 0% reduce 0%  
23/01/04 16:20:34 INFO Job: map 100% reduce 0%  
23/01/04 16:20:38 INFO Job: map 100% reduce 100%  
23/01/04 16:20:38 INFO Job: Job job\_1672817468245\_0924 completed successfully  
23/01/04 16:20:38 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5166929  
FILE: Number of bytes written=10866149  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7481763  
HDFS: Number of bytes written=2420572  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19135  
Total time spent by all reduces in occupied slots (ms)=10540  
Total time spent by all map tasks (ms)=3827  
Total time spent by all reduce tasks (ms)=2108  
Total vcore-milliseconds taken by all map tasks=3827  
Total vcore-milliseconds taken by all reduce tasks=2108  
Total megabyte-milliseconds taken by all map tasks=19594240  
Total megabyte-milliseconds taken by all reduce tasks=10792960

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5166878  
Map output materialized bytes=5166929  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5166929  
Reduce input records=9

Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=59  
CPU time spent (ms)=4540  
Physical memory (bytes) snapshot=741302272  
Virtual memory (bytes) snapshot=12767334400  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418045952  
Peak Map Virtual memory (bytes)=6379106304  
Peak Reduce Physical memory (bytes)=323256320  
Peak Reduce Virtual memory (bytes)=6388228096

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=2680027

File Output Format Counters

Bytes Written=2420572

23/01/04 16:20:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:20:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0925  
23/01/04 16:20:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:20:38 INFO JobSubmitter: number of splits:1  
23/01/04 16:20:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0925  
23/01/04 16:20:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:20:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0925  
23/01/04 16:20:38 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0925/](http://lena-master:8088/proxy/application_1672817468245_0925/)  
23/01/04 16:20:38 INFO Job: Running job: job\_1672817468245\_0925  
23/01/04 16:20:44 INFO Job: Job job\_1672817468245\_0925 running in uber mode : false  
23/01/04 16:20:44 INFO Job: map 0% reduce 0%  
23/01/04 16:20:51 INFO Job: map 100% reduce 0%  
23/01/04 16:20:58 INFO Job: map 100% reduce 100%  
23/01/04 16:20:58 INFO Job: Job job\_1672817468245\_0925 completed successfully  
23/01/04 16:20:58 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=5282250  
FILE: Number of bytes written=11096791  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7521729  
HDFS: Number of bytes written=2517604  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=20255  
Total time spent by all reduces in occupied slots (ms)=18080  
Total time spent by all map tasks (ms)=4051  
Total time spent by all reduce tasks (ms)=3616  
Total vcore-milliseconds taken by all map tasks=4051  
Total vcore-milliseconds taken by all reduce tasks=3616  
Total megabyte-milliseconds taken by all map tasks=20741120  
Total megabyte-milliseconds taken by all reduce tasks=18513920

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5282199  
Map output materialized bytes=5282250  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5282250  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=70  
CPU time spent (ms)=4680  
Physical memory (bytes) snapshot=735784960  
Virtual memory (bytes) snapshot=12773818368  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418222080  
Peak Map Virtual memory (bytes)=6389207040  
Peak Reduce Physical memory (bytes)=317562880  
Peak Reduce Virtual memory (bytes)=6384611328

#### 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=2680027

#### File Output Format Counters

Bytes Written=2517604

23/01/04 16:20:58 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:20:58 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0928  
23/01/04 16:20:58 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:20:58 INFO JobSubmitter: number of splits:1  
23/01/04 16:20:58 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0928  
23/01/04 16:20:58 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:20:58 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0928  
23/01/04 16:20:58 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0928/](http://lena-master:8088/proxy/application_1672817468245_0928/)

23/01/04 16:20:58 INFO Job: Running job: job\_1672817468245\_0928  
23/01/04 16:21:04 INFO Job: Job job\_1672817468245\_0928 running in uber mode : false  
23/01/04 16:21:04 INFO Job: map 0% reduce 0%  
23/01/04 16:21:10 INFO Job: map 100% reduce 0%  
23/01/04 16:21:17 INFO Job: map 100% reduce 100%  
23/01/04 16:21:17 INFO Job: Job job\_1672817468245\_0928 completed successfully  
23/01/04 16:21:17 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5404800  
FILE: Number of bytes written=11341891  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7715793  
HDFS: Number of bytes written=2541451  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19450  
Total time spent by all reduces in occupied slots (ms)=17870  
Total time spent by all map tasks (ms)=3890  
Total time spent by all reduce tasks (ms)=3574  
Total vcore-milliseconds taken by all map tasks=3890  
Total vcore-milliseconds taken by all reduce tasks=3574  
Total megabyte-milliseconds taken by all map tasks=19916800  
Total megabyte-milliseconds taken by all reduce tasks=18298880

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5404749  
Map output materialized bytes=5404800  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5404800  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=4730  
Physical memory (bytes) snapshot=774660096  
Virtual memory (bytes) snapshot=12756574208  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=443310080  
Peak Map Virtual memory (bytes)=6369894400  
Peak Reduce Physical memory (bytes)=331350016  
Peak Reduce Virtual memory (bytes)=6386679808

#### 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=2680027
File Output Format Counters
    Bytes Written=2541451
23/01/04 16:21:17 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:21:17 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0931
23/01/04 16:21:17 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:21:17 INFO JobSubmitter: number of splits:1
23/01/04 16:21:17 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0931
23/01/04 16:21:17 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:21:17 INFO YarnClientImpl: Submitted application
application_1672817468245_0931
23/01/04 16:21:17 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0931/
23/01/04 16:21:17 INFO Job: Running job: job_1672817468245_0931
23/01/04 16:21:23 INFO Job: Job job_1672817468245_0931 running in uber mode : false
23/01/04 16:21:23 INFO Job:  map 0% reduce 0%
23/01/04 16:21:30 INFO Job:  map 100% reduce 0%
23/01/04 16:21:37 INFO Job:  map 100% reduce 100%
23/01/04 16:21:37 INFO Job: Job job_1672817468245_0931 completed successfully
23/01/04 16:21:37 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=5398503
    FILE: Number of bytes written=11329297
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7763487
    HDFS: Number of bytes written=2501920
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=19735
    Total time spent by all reduces in occupied slots (ms)=17755
    Total time spent by all map tasks (ms)=3947
    Total time spent by all reduce tasks (ms)=3551
    Total vcore-milliseconds taken by all map tasks=3947
    Total vcore-milliseconds taken by all reduce tasks=3551
    Total megabyte-milliseconds taken by all map tasks=20208640
    Total megabyte-milliseconds taken by all reduce tasks=18181120
Map-Reduce Framework
    Map input records=27
    Map output records=9
    Map output bytes=5398452
    Map output materialized bytes=5398503
    Input split bytes=170
    Combine input records=0

```

Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5398503  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=60  
CPU time spent (ms)=4660  
Physical memory (bytes) snapshot=749957120  
Virtual memory (bytes) snapshot=12755935232  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=430006272  
Peak Map Virtual memory (bytes)=6369468416  
Peak Reduce Physical memory (bytes)=319950848  
Peak Reduce Virtual memory (bytes)=6386466816

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=2680027

File Output Format Counters

Bytes Written=2501920

23/01/04 16:21:37 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:21:37 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0934  
23/01/04 16:21:37 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:21:37 INFO JobSubmitter: number of splits:1  
23/01/04 16:21:37 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0934  
23/01/04 16:21:37 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:21:37 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0934  
23/01/04 16:21:37 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0934/](http://lena-master:8088/proxy/application_1672817468245_0934/)  
23/01/04 16:21:37 INFO Job: Running job: job\_1672817468245\_0934  
23/01/04 16:21:43 INFO Job: Job job\_1672817468245\_0934 running in uber mode : false  
23/01/04 16:21:43 INFO Job: map 0% reduce 0%  
23/01/04 16:21:49 INFO Job: map 100% reduce 0%  
23/01/04 16:21:55 INFO Job: map 100% reduce 100%  
23/01/04 16:21:55 INFO Job: Job job\_1672817468245\_0934 completed successfully  
23/01/04 16:21:55 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=5358972  
FILE: Number of bytes written=11250235  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7684425  
HDFS: Number of bytes written=2501920  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0

HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19540  
Total time spent by all reduces in occupied slots (ms)=17470  
Total time spent by all map tasks (ms)=3908  
Total time spent by all reduce tasks (ms)=3494  
Total vcore-milliseconds taken by all map tasks=3908  
Total vcore-milliseconds taken by all reduce tasks=3494  
Total megabyte-milliseconds taken by all map tasks=20008960  
Total megabyte-milliseconds taken by all reduce tasks=17889280

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=5358921  
Map output materialized bytes=5358972  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=5358972  
Reduce input records=9  
Reduce output records=9  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=49  
CPU time spent (ms)=4740  
Physical memory (bytes) snapshot=763789312  
Virtual memory (bytes) snapshot=12753887232  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=432304128  
Peak Map Virtual memory (bytes)=6373548032  
Peak Reduce Physical memory (bytes)=331485184  
Peak Reduce Virtual memory (bytes)=6380339200

#### 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=2680027

#### File Output Format Counters

Bytes Written=2501920

23/01/04 16:21:55 INFO KMeansDriver: Clustering data  
23/01/04 16:21:55 INFO KMeansDriver: Running Clustering  
23/01/04 16:21:55 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/09 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/09  
23/01/04 16:21:55 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:21:55 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0937  
23/01/04 16:21:55 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:21:55 INFO JobSubmitter: number of splits:1  
23/01/04 16:21:55 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0937  
23/01/04 16:21:55 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:21:55 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0937  
23/01/04 16:21:55 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0937/  
23/01/04 16:21:55 INFO Job: Running job: job\_1672817468245\_0937  
23/01/04 16:22:01 INFO Job: Job job\_1672817468245\_0937 running in uber mode : false  
23/01/04 16:22:01 INFO Job: map 0% reduce 0%  
23/01/04 16:22:08 INFO Job: map 100% reduce 0%  
23/01/04 16:22:08 INFO Job: Job job\_1672817468245\_0937 completed successfully  
23/01/04 16:22:08 INFO Job: Counters: 33

#### File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5182311  
HDFS: Number of bytes written=2680570  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19360  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3872  
Total vcore-milliseconds taken by all map tasks=3872  
Total megabyte-milliseconds taken by all map tasks=19824640

#### Map-Reduce Framework

Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=35  
CPU time spent (ms)=2580  
Physical memory (bytes) snapshot=309673984  
Virtual memory (bytes) snapshot=6375997440  
Total committed heap usage (bytes)=1052770304  
Peak Map Physical memory (bytes)=309673984  
Peak Map Virtual memory (bytes)=6375997440

#### File Input Format Counters

Bytes Read=2680027

#### File Output Format Counters

Bytes Written=2680570

23/01/04 16:22:08 INFO MahoutDriver: Program took 164055 ms (Minutes: 2.73425)  
MAHOUT\_LOCAL is not set; adding HADOOP\_CONF\_DIR to classpath.  
Running on hadoop, using /opt/hadoop/current/bin/hadoop and

HADOOP\_CONF\_DIR=/opt/hadoop/current/etc/hadoop  
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar  
23/01/04 16:22:13 INFO AbstractJob: Command line arguments:  
{--dictionary=[dsm010/question\_2/british-fiction-corpus-vectors/dictionary.file-\*],  
--dictionaryType=[sequencefile], --  
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question\_2/british-fiction-  
corpus-kmeans/ManhattanDistanceMeasure/09/clusters-8-final], --numWords=[20], --  
output=[/home/bfezo001/dsm010/question\_2/output/british-fiction-corpus-clusters/  
ManhattanDistanceMeasure/09.txt], --outputFormat=[TEXT],  
--pointsDir=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu  
re/09/clusteredPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}  
23/01/04 16:22:15 INFO HadoopUtil: Deleting tmp/representative  
23/01/04 16:22:15 INFO AbstractJob: Command line arguments: {--  
clusteredPoints=[dsm010/question\_2/british-fiction-corpus-kmeans/  
ManhattanDistanceMeasure/09/clusteredPoints], --  
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --  
endPhase=[2147483647],  
--input=[dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/  
09/clusters-8-final], --maxIter=[5], --method=[mapreduce],  
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}  
23/01/04 16:22:15 INFO RepresentativePointsDriver: Representative Points Iteration  
0  
23/01/04 16:22:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:22:15 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0940  
23/01/04 16:22:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:22:16 INFO JobSubmitter: number of splits:1  
23/01/04 16:22:16 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0940  
23/01/04 16:22:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:22:16 INFO Configuration: resource-types.xml not found  
23/01/04 16:22:16 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:22:16 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0940  
23/01/04 16:22:16 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0940/  
23/01/04 16:22:16 INFO Job: Running job: job\_1672817468245\_0940  
23/01/04 16:22:22 INFO Job: Job job\_1672817468245\_0940 running in uber mode : false  
23/01/04 16:22:22 INFO Job: map 0% reduce 0%  
23/01/04 16:22:27 INFO Job: map 100% reduce 0%  
23/01/04 16:22:33 INFO Job: map 100% reduce 100%  
23/01/04 16:22:33 INFO Job: Job job\_1672817468245\_0940 completed successfully  
23/01/04 16:22:33 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1025569  
FILE: Number of bytes written=2583731  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5538819  
HDFS: Number of bytes written=2454681  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1

Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17305  
Total time spent by all reduces in occupied slots (ms)=17175  
Total time spent by all map tasks (ms)=3461  
Total time spent by all reduce tasks (ms)=3435  
Total vcore-milliseconds taken by all map tasks=3461  
Total vcore-milliseconds taken by all reduce tasks=3435  
Total megabyte-milliseconds taken by all map tasks=17720320  
Total megabyte-milliseconds taken by all reduce tasks=17587200

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=1025518  
Map output materialized bytes=1025569  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=1025569  
Reduce input records=9  
Reduce output records=18  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=41  
CPU time spent (ms)=3070  
Physical memory (bytes) snapshot=695713792  
Virtual memory (bytes) snapshot=12753354752  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=397078528  
Peak Map Virtual memory (bytes)=6375084032  
Peak Reduce Physical memory (bytes)=298635264  
Peak Reduce Virtual memory (bytes)=6382178304

#### 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=2680570

#### File Output Format Counters

Bytes Written=2454681

23/01/04 16:22:33 INFO RepresentativePointsDriver: Representative Points Iteration 1

23/01/04 16:22:33 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:22:33 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0943

23/01/04 16:22:33 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:22:33 INFO JobSubmitter: number of splits:1

23/01/04 16:22:34 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0943

23/01/04 16:22:34 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:22:34 INFO YarnClientImpl: Submitted application application\_1672817468245\_0943

23/01/04 16:22:34 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0943/  
23/01/04 16:22:34 INFO Job: Running job: job\_1672817468245\_0943  
23/01/04 16:22:40 INFO Job: Job job\_1672817468245\_0943 running in uber mode : false  
23/01/04 16:22:40 INFO Job: map 0% reduce 0%  
23/01/04 16:22:46 INFO Job: map 100% reduce 0%  
23/01/04 16:22:51 INFO Job: map 100% reduce 100%  
23/01/04 16:22:52 INFO Job: Job job\_1672817468245\_0943 completed successfully  
23/01/04 16:22:52 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=956087  
FILE: Number of bytes written=2444767  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7590135  
HDFS: Number of bytes written=3410858  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17625  
Total time spent by all reduces in occupied slots (ms)=17180  
Total time spent by all map tasks (ms)=3525  
Total time spent by all reduce tasks (ms)=3436  
Total vcore-milliseconds taken by all map tasks=3525  
Total vcore-milliseconds taken by all reduce tasks=3436  
Total megabyte-milliseconds taken by all map tasks=18048000  
Total megabyte-milliseconds taken by all reduce tasks=17592320

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=956037  
Map output materialized bytes=956087  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=956087  
Reduce input records=9  
Reduce output records=27  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=42  
CPU time spent (ms)=3120  
Physical memory (bytes) snapshot=719605760  
Virtual memory (bytes) snapshot=12750508032  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=418017280  
Peak Map Virtual memory (bytes)=6368182272  
Peak Reduce Physical memory (bytes)=301588480  
Peak Reduce Virtual memory (bytes)=6384279552

#### 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=2680570
    File Output Format Counters
        Bytes Written=3410858
23/01/04 16:22:52 INFO RepresentativePointsDriver: Representative Points Iteration
2
23/01/04 16:22:52 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:22:52 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0946
23/01/04 16:22:52 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:22:52 INFO JobSubmitter: number of splits:1
23/01/04 16:22:52 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0946
23/01/04 16:22:52 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:22:52 INFO YarnClientImpl: Submitted application
application_1672817468245_0946
23/01/04 16:22:52 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0946/
23/01/04 16:22:52 INFO Job: Running job: job_1672817468245_0946
23/01/04 16:22:58 INFO Job: Job job_1672817468245_0946 running in uber mode : false
23/01/04 16:22:58 INFO Job:  map 0% reduce 0%
23/01/04 16:23:04 INFO Job:  map 100% reduce 0%
23/01/04 16:23:10 INFO Job:  map 100% reduce 100%
23/01/04 16:23:10 INFO Job: Job job_1672817468245_0946 completed successfully
23/01/04 16:23:10 INFO Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=965395
        FILE: Number of bytes written=2463383
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=9502489
        HDFS: Number of bytes written=4376322
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=17405
        Total time spent by all reduces in occupied slots (ms)=17030
        Total time spent by all map tasks (ms)=3481
        Total time spent by all reduce tasks (ms)=3406
        Total vcore-milliseconds taken by all map tasks=3481
        Total vcore-milliseconds taken by all reduce tasks=3406
        Total megabyte-milliseconds taken by all map tasks=17822720
        Total megabyte-milliseconds taken by all reduce tasks=17438720
    Map-Reduce Framework
        Map input records=27
        Map output records=9

```



Map output bytes=965344  
Map output materialized bytes=965395  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=965395  
Reduce input records=9  
Reduce output records=36  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=54  
CPU time spent (ms)=3190  
Physical memory (bytes) snapshot=693342208  
Virtual memory (bytes) snapshot=12749295616  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408584192  
Peak Map Virtual memory (bytes)=6373347328  
Peak Reduce Physical memory (bytes)=291979264  
Peak Reduce Virtual memory (bytes)=6377902080

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=2680570

File Output Format Counters

Bytes Written=4376322

23/01/04 16:23:10 INFO RepresentativePointsDriver: Representative Points Iteration  
3  
23/01/04 16:23:10 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:23:10 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0949  
23/01/04 16:23:11 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:23:11 INFO JobSubmitter: number of splits:1  
23/01/04 16:23:11 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0949  
23/01/04 16:23:11 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:23:11 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0949  
23/01/04 16:23:11 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0949/  
23/01/04 16:23:11 INFO Job: Running job: job\_1672817468245\_0949  
23/01/04 16:23:17 INFO Job: Job job\_1672817468245\_0949 running in uber mode : false  
23/01/04 16:23:17 INFO Job: map 0% reduce 0%  
23/01/04 16:23:23 INFO Job: map 100% reduce 0%  
23/01/04 16:23:28 INFO Job: map 100% reduce 100%  
23/01/04 16:23:29 INFO Job: Job job\_1672817468245\_0949 completed successfully  
23/01/04 16:23:29 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=963933  
FILE: Number of bytes written=2460459  
FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=11433417  
HDFS: Number of bytes written=5340364  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17175  
Total time spent by all reduces in occupied slots (ms)=17010  
Total time spent by all map tasks (ms)=3435  
Total time spent by all reduce tasks (ms)=3402  
Total vcore-milliseconds taken by all map tasks=3435  
Total vcore-milliseconds taken by all reduce tasks=3402  
Total megabyte-milliseconds taken by all map tasks=17587200  
Total megabyte-milliseconds taken by all reduce tasks=17418240

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=963882  
Map output materialized bytes=963933  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=963933  
Reduce input records=9  
Reduce output records=45  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=48  
CPU time spent (ms)=3440  
Physical memory (bytes) snapshot=725909504  
Virtual memory (bytes) snapshot=12751339520  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=410345472  
Peak Map Virtual memory (bytes)=6372499456  
Peak Reduce Physical memory (bytes)=315564032  
Peak Reduce Virtual memory (bytes)=6378840064

#### 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=2680570

#### File Output Format Counters

Bytes Written=5340364

23/01/04 16:23:29 INFO RepresentativePointsDriver: Representative Points Iteration  
4

23/01/04 16:23:29 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:23:29 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0952  
23/01/04 16:23:29 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:23:29 INFO JobSubmitter: number of splits:1  
23/01/04 16:23:29 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0952  
23/01/04 16:23:29 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:23:29 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0952  
23/01/04 16:23:29 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0952/  
23/01/04 16:23:29 INFO Job: Running job: job\_1672817468245\_0952  
23/01/04 16:23:35 INFO Job: Job job\_1672817468245\_0952 running in uber mode : false  
23/01/04 16:23:35 INFO Job: map 0% reduce 0%  
23/01/04 16:23:41 INFO Job: map 100% reduce 0%  
23/01/04 16:23:47 INFO Job: map 100% reduce 100%  
23/01/04 16:23:47 INFO Job: Job job\_1672817468245\_0952 completed successfully  
23/01/04 16:23:47 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1025569  
FILE: Number of bytes written=2583731  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=13361501  
HDFS: Number of bytes written=6366022  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17210  
Total time spent by all reduces in occupied slots (ms)=16985  
Total time spent by all map tasks (ms)=3442  
Total time spent by all reduce tasks (ms)=3397  
Total vcore-milliseconds taken by all map tasks=3442  
Total vcore-milliseconds taken by all reduce tasks=3397  
Total megabyte-milliseconds taken by all map tasks=17623040  
Total megabyte-milliseconds taken by all reduce tasks=17392640

#### Map-Reduce Framework

Map input records=27  
Map output records=9  
Map output bytes=1025518  
Map output materialized bytes=1025569  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=9  
Reduce shuffle bytes=1025569  
Reduce input records=9  
Reduce output records=54  
Spilled Records=18  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

```

GC time elapsed (ms)=82
CPU time spent (ms)=3390
Physical memory (bytes) snapshot=724242432
Virtual memory (bytes) snapshot=12766146560
Total committed heap usage (bytes)=2105540608
Peak Map Physical memory (bytes)=412295168
Peak Map Virtual memory (bytes)=6378745856
Peak Reduce Physical memory (bytes)=311947264
Peak Reduce Virtual memory (bytes)=6387400704
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=2680570
File Output Format Counters
  Bytes Written=6366022
23/01/04 16:23:48 INFO ClusterEvaluator: Scaled Inter-Cluster Density =
0.33197862332237604
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[25] = NaN
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[1] =
0.7327838989762803
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[3] =
0.7424407538719928
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[4] =
0.5666666666666667
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[24] = NaN
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[6] =
0.7225625638590281
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[15] = NaN
23/01/04 16:23:48 INFO ClusterEvaluator: Intra-Cluster Density[26] = NaN
23/01/04 16:23:48 INFO ClusterEvaluator: Average Intra-Cluster Density =
0.6911134708434918
23/01/04 16:23:52 INFO ClusterDumper: Wrote 9 clusters
23/01/04 16:23:52 INFO MahoutDriver: Program took 98853 ms (Minutes: 1.64755)
---- K: 10 --
dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/10
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:23:58 INFO AbstractJob: Command line arguments: [--clustering=null, --
clusters=[dsm010/question_2/british-fiction-corpus-canopy-centroids/
ManhattanDistanceMeasure], --convergenceDelta=[0.1], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-vectors/tf-vectors], --
maxIter=[20], --method=[mapreduce], --numClusters=[10],
--output=[hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-
kmeans/ManhattanDistanceMeasure/10], --overwrite=null, --startPhase=[0], --
tempDir=[temp]]
23/01/04 16:23:59 INFO HadoopUtil: Deleting
hdfs://lena/user/bfezo001/dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/10
23/01/04 16:23:59 INFO HadoopUtil: Deleting dsm010/question_2/british-fiction-

```

corpus-canopy-centroids/ManhattanDistanceMeasure  
23/01/04 16:23:59 INFO ZlibFactory: Successfully loaded & initialized native-zlib library  
23/01/04 16:23:59 INFO CodecPool: Got brand-new compressor [.deflate]  
23/01/04 16:23:59 INFO RandomSeedGenerator: Wrote 10 Klusters to dsm010/question\_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/part-randomSeed  
23/01/04 16:23:59 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In: dsm010/question\_2/british-fiction-corpus-canopy-centroids/ManhattanDistanceMeasure/part-randomSeed Out: hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/10  
23/01/04 16:23:59 INFO KMeansDriver: convergence: 0.1 max Iterations: 20  
23/01/04 16:23:59 INFO CodecPool: Got brand-new decompressor [.deflate]  
23/01/04 16:24:00 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:24:00 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0955  
23/01/04 16:24:00 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:24:00 INFO JobSubmitter: number of splits:1  
23/01/04 16:24:00 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0955  
23/01/04 16:24:00 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:24:01 INFO Configuration: resource-types.xml not found  
23/01/04 16:24:01 INFO ResourceUtils: Unable to find 'resource-types.xml'.  
23/01/04 16:24:01 INFO YarnClientImpl: Submitted application application\_1672817468245\_0955  
23/01/04 16:24:01 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0955/  
23/01/04 16:24:01 INFO Job: Running job: job\_1672817468245\_0955  
23/01/04 16:24:07 INFO Job: Job job\_1672817468245\_0955 running in uber mode : false  
23/01/04 16:24:07 INFO Job: map 0% reduce 0%  
23/01/04 16:24:13 INFO Job: map 100% reduce 0%  
23/01/04 16:24:18 INFO Job: map 100% reduce 100%  
23/01/04 16:24:18 INFO Job: Job job\_1672817468245\_0955 completed successfully  
23/01/04 16:24:18 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=4080120  
FILE: Number of bytes written=8692531  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8886829  
HDFS: Number of bytes written=2342635  
HDFS: Number of read operations=55  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19785  
Total time spent by all reduces in occupied slots (ms)=10565  
Total time spent by all map tasks (ms)=3957  
Total time spent by all reduce tasks (ms)=2113  
Total vcore-milliseconds taken by all map tasks=3957  
Total vcore-milliseconds taken by all reduce tasks=2113  
Total megabyte-milliseconds taken by all map tasks=20259840

```

        Total megabyte-milliseconds taken by all reduce tasks=10818560
Map-Reduce Framework
    Map input records=27
    Map output records=10
    Map output bytes=4080064
    Map output materialized bytes=4080120
    Input split bytes=170
    Combine input records=0
    Combine output records=0
    Reduce input groups=10
    Reduce shuffle bytes=4080120
    Reduce input records=10
    Reduce output records=10
    Spilled Records=20
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=52
    CPU time spent (ms)=4900
    Physical memory (bytes) snapshot=735350784
    Virtual memory (bytes) snapshot=12753162240
    Total committed heap usage (bytes)=2105540608
    Peak Map Physical memory (bytes)=420511744
    Peak Map Virtual memory (bytes)=6368645120
    Peak Reduce Physical memory (bytes)=314839040
    Peak Reduce Virtual memory (bytes)=6384517120
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=2680027
File Output Format Counters
    Bytes Written=2342635
23/01/04 16:24:18 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:24:18 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0958
23/01/04 16:24:18 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:24:18 INFO JobSubmitter: number of splits:1
23/01/04 16:24:18 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0958
23/01/04 16:24:18 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:24:18 INFO YarnClientImpl: Submitted application
application_1672817468245_0958
23/01/04 16:24:18 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0958/
23/01/04 16:24:18 INFO Job: Running job: job_1672817468245_0958
23/01/04 16:24:24 INFO Job: Job job_1672817468245_0958 running in uber mode : false
23/01/04 16:24:24 INFO Job:  map 0% reduce 0%
23/01/04 16:24:31 INFO Job:  map 100% reduce 0%
23/01/04 16:24:37 INFO Job:  map 100% reduce 100%
23/01/04 16:24:38 INFO Job: Job job_1672817468245_0958 completed successfully
23/01/04 16:24:38 INFO Job: Counters: 54
File System Counters
    FILE: Number of bytes read=5470220

```

FILE: Number of bytes written=11472731  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7365855  
HDFS: Number of bytes written=2507534  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19375  
Total time spent by all reduces in occupied slots (ms)=17490  
Total time spent by all map tasks (ms)=3875  
Total time spent by all reduce tasks (ms)=3498  
Total vcore-milliseconds taken by all map tasks=3875  
Total vcore-milliseconds taken by all reduce tasks=3498  
Total megabyte-milliseconds taken by all map tasks=19840000  
Total megabyte-milliseconds taken by all reduce tasks=17909760

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5470164  
Map output materialized bytes=5470220  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5470220  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=44  
CPU time spent (ms)=4570  
Physical memory (bytes) snapshot=742445056  
Virtual memory (bytes) snapshot=12748156928  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=426799104  
Peak Map Virtual memory (bytes)=6373298176  
Peak Reduce Physical memory (bytes)=315645952  
Peak Reduce Virtual memory (bytes)=6374858752

#### 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=2680027

#### File Output Format Counters

Bytes Written=2507534

23/01/04 16:24:38 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to

ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:24:38 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0961  
23/01/04 16:24:38 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:24:38 INFO JobSubmitter: number of splits:1  
23/01/04 16:24:38 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0961  
23/01/04 16:24:38 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:24:38 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0961  
23/01/04 16:24:38 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0961/  
23/01/04 16:24:38 INFO Job: Running job: job\_1672817468245\_0961  
23/01/04 16:24:44 INFO Job: Job job\_1672817468245\_0961 running in uber mode : false  
23/01/04 16:24:44 INFO Job: map 0% reduce 0%  
23/01/04 16:24:50 INFO Job: map 100% reduce 0%  
23/01/04 16:24:57 INFO Job: map 100% reduce 100%  
23/01/04 16:24:57 INFO Job: Job job\_1672817468245\_0961 completed successfully  
23/01/04 16:24:57 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5720671  
FILE: Number of bytes written=11973633  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7695653  
HDFS: Number of bytes written=2575304  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19605  
Total time spent by all reduces in occupied slots (ms)=18495  
Total time spent by all map tasks (ms)=3921  
Total time spent by all reduce tasks (ms)=3699  
Total vcore-milliseconds taken by all map tasks=3921  
Total vcore-milliseconds taken by all reduce tasks=3699  
Total megabyte-milliseconds taken by all map tasks=20075520  
Total megabyte-milliseconds taken by all reduce tasks=18938880

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5720615  
Map output materialized bytes=5720671  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5720671  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1



GC time elapsed (ms)=41  
CPU time spent (ms)=4860  
Physical memory (bytes) snapshot=782569472  
Virtual memory (bytes) snapshot=12752080896  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=448724992  
Peak Map Virtual memory (bytes)=6375907328  
Peak Reduce Physical memory (bytes)=333844480  
Peak Reduce Virtual memory (bytes)=6376173568

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=2680027

File Output Format Counters  
Bytes Written=2575304

23/01/04 16:24:57 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:24:57 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0964  
23/01/04 16:24:57 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:24:57 INFO JobSubmitter: number of splits:1  
23/01/04 16:24:57 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0964  
23/01/04 16:24:57 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:24:57 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0964  
23/01/04 16:24:57 INFO Job: The url to track the job:  
[http://lena-master:8088/proxy/application\\_1672817468245\\_0964/](http://lena-master:8088/proxy/application_1672817468245_0964/)  
23/01/04 16:24:57 INFO Job: Running job: job\_1672817468245\_0964  
23/01/04 16:25:03 INFO Job: Job job\_1672817468245\_0964 running in uber mode : false  
23/01/04 16:25:03 INFO Job: map 0% reduce 0%  
23/01/04 16:25:10 INFO Job: map 100% reduce 0%  
23/01/04 16:25:15 INFO Job: map 100% reduce 100%  
23/01/04 16:25:16 INFO Job: Job job\_1672817468245\_0964 completed successfully  
23/01/04 16:25:16 INFO Job: Counters: 54

File System Counters  
FILE: Number of bytes read=5703573  
FILE: Number of bytes written=11939437  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7831193  
HDFS: Number of bytes written=2477498  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters  
Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19755  
Total time spent by all reduces in occupied slots (ms)=10385  
Total time spent by all map tasks (ms)=3951

Total time spent by all reduce tasks (ms)=2077  
Total vcore-milliseconds taken by all map tasks=3951  
Total vcore-milliseconds taken by all reduce tasks=2077  
Total megabyte-milliseconds taken by all map tasks=20229120  
Total megabyte-milliseconds taken by all reduce tasks=10634240

Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5703517  
Map output materialized bytes=5703573  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5703573  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=4680  
Physical memory (bytes) snapshot=746299392  
Virtual memory (bytes) snapshot=12754915328  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=430649344  
Peak Map Virtual memory (bytes)=6374711296  
Peak Reduce Physical memory (bytes)=315650048  
Peak Reduce Virtual memory (bytes)=6380204032

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=2680027

File Output Format Counters

Bytes Written=2477498

23/01/04 16:25:16 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to  
ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:25:16 INFO JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0967  
23/01/04 16:25:16 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:25:16 INFO JobSubmitter: number of splits:1  
23/01/04 16:25:16 INFO JobSubmitter: Submitting tokens for job:  
job\_1672817468245\_0967  
23/01/04 16:25:16 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:25:16 INFO YarnClientImpl: Submitted application  
application\_1672817468245\_0967  
23/01/04 16:25:16 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0967/  
23/01/04 16:25:16 INFO Job: Running job: job\_1672817468245\_0967  
23/01/04 16:25:22 INFO Job: Job job\_1672817468245\_0967 running in uber mode : false  
23/01/04 16:25:22 INFO Job: map 0% reduce 0%  
23/01/04 16:25:28 INFO Job: map 100% reduce 0%  
23/01/04 16:25:34 INFO Job: map 100% reduce 100%

23/01/04 16:25:35 INFO Job: Job job\_1672817468245\_0967 completed successfully

23/01/04 16:25:35 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=5605767  
FILE: Number of bytes written=11743825  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=7635581  
HDFS: Number of bytes written=2477498  
HDFS: Number of read operations=19  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=20065  
Total time spent by all reduces in occupied slots (ms)=17605  
Total time spent by all map tasks (ms)=4013  
Total time spent by all reduce tasks (ms)=3521  
Total vcore-milliseconds taken by all map tasks=4013  
Total vcore-milliseconds taken by all reduce tasks=3521  
Total megabyte-milliseconds taken by all map tasks=20546560  
Total megabyte-milliseconds taken by all reduce tasks=18027520

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=5605711  
Map output materialized bytes=5605767  
Input split bytes=170  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=5605767  
Reduce input records=10  
Reduce output records=10  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=61  
CPU time spent (ms)=4490  
Physical memory (bytes) snapshot=742555648  
Virtual memory (bytes) snapshot=12753825792  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=422244352  
Peak Map Virtual memory (bytes)=6374252544  
Peak Reduce Physical memory (bytes)=320311296  
Peak Reduce Virtual memory (bytes)=6379573248

#### 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=2680027  
File Output Format Counters  
Bytes Written=2477498  
23/01/04 16:25:35 INFO KMeansDriver: Clustering data  
23/01/04 16:25:35 INFO KMeansDriver: Running Clustering  
23/01/04 16:25:35 INFO KMeansDriver: Input: dsm010/question\_2/british-fiction-corpus-vectors/tf-vectors Clusters In:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/10 Out:  
hdfs://lena/user/bfezo001/dsm010/question\_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/10  
23/01/04 16:25:35 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:25:35 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0970  
23/01/04 16:25:36 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:25:36 INFO JobSubmitter: number of splits:1  
23/01/04 16:25:36 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0970  
23/01/04 16:25:36 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:25:36 INFO YarnClientImpl: Submitted application application\_1672817468245\_0970  
23/01/04 16:25:36 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0970/  
23/01/04 16:25:36 INFO Job: Running job: job\_1672817468245\_0970  
23/01/04 16:25:42 INFO Job: Job job\_1672817468245\_0970 running in uber mode : false  
23/01/04 16:25:42 INFO Job: map 0% reduce 0%  
23/01/04 16:25:48 INFO Job: map 100% reduce 0%  
23/01/04 16:25:48 INFO Job: Job job\_1672817468245\_0970 completed successfully  
23/01/04 16:25:48 INFO Job: Counters: 33  
File System Counters  
FILE: Number of bytes read=0  
FILE: Number of bytes written=265897  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5157889  
HDFS: Number of bytes written=2680606  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0  
Job Counters  
Launched map tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=19630  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=3926  
Total vcore-milliseconds taken by all map tasks=3926  
Total megabyte-milliseconds taken by all map tasks=20101120  
Map-Reduce Framework  
Map input records=27  
Map output records=27  
Input split bytes=170  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=26  
CPU time spent (ms)=2540

```

Physical memory (bytes) snapshot=316280832
Virtual memory (bytes) snapshot=6375456768
Total committed heap usage (bytes)=1052770304
Peak Map Physical memory (bytes)=316280832
Peak Map Virtual memory (bytes)=6375456768
File Input Format Counters
  Bytes Read=2680027
File Output Format Counters
  Bytes Written=2680606
23/01/04 16:25:48 INFO MahoutDriver: Program took 109625 ms (Minutes:
1.8270833333333334)
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /opt/hadoop/current/bin/hadoop and
HADOOP_CONF_DIR=/opt/hadoop/current/etc/hadoop
MAHOUT-JOB: /opt/mahout/current/mahout-examples-0.13.0-job.jar
23/01/04 16:25:54 INFO AbstractJob: Command line arguments:
{--dictionary=[dsm010/question_2/british-fiction-corpus-vectors/dictionary.file-*],
--dictionaryType=[sequencefile], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647], --evaluate=null, --input=[dsm010/question_2/british-fiction-
corpus-kmeans/ManhattanDistanceMeasure/10/clusters-5-final], --numWords=[20], --
output=[/home/bfezo001/dsm010/question_2/output/british-fiction-corpus-clusters/
ManhattanDistanceMeasure/10.txt], --outputFormat=[TEXT],
--pointsDir=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasu
re/10/clusteringPoints], --startPhase=[0], --substring=[100], --tempDir=[temp]}
23/01/04 16:25:55 INFO HadoopUtil: Deleting tmp/representative
23/01/04 16:25:55 INFO AbstractJob: Command line arguments: {--
clusteringPoints=[dsm010/question_2/british-fiction-corpus-kmeans/
ManhattanDistanceMeasure/10/clusteringPoints], --
distanceMeasure=[org.apache.mahout.common.distance.ManhattanDistanceMeasure], --
endPhase=[2147483647],
--input=[dsm010/question_2/british-fiction-corpus-kmeans/ManhattanDistanceMeasure/
10/clusters-5-final], --maxIter=[5], --method=[mapreduce],
--output=[tmp/representative], --startPhase=[0], --tempDir=[temp]}
23/01/04 16:25:56 INFO RepresentativePointsDriver: Representative Points Iteration
0
23/01/04 16:25:56 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:25:56 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0972
23/01/04 16:25:56 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:25:56 INFO JobSubmitter: number of splits:1
23/01/04 16:25:56 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0972
23/01/04 16:25:56 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:25:56 INFO Configuration: resource-types.xml not found
23/01/04 16:25:56 INFO ResourceUtils: Unable to find 'resource-types.xml'.
23/01/04 16:25:56 INFO YarnClientImpl: Submitted application
application_1672817468245_0972
23/01/04 16:25:57 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0972/
23/01/04 16:25:57 INFO Job: Running job: job_1672817468245_0972
23/01/04 16:26:03 INFO Job: Job job_1672817468245_0972 running in uber mode : false
23/01/04 16:26:03 INFO Job:  map 0% reduce 0%
23/01/04 16:26:09 INFO Job:  map 100% reduce 0%
23/01/04 16:26:15 INFO Job:  map 100% reduce 100%
23/01/04 16:26:15 INFO Job: Job job_1672817468245_0972 completed successfully
23/01/04 16:26:15 INFO Job: Counters: 54
File System Counters

```

FILE: Number of bytes read=1229980  
FILE: Number of bytes written=2992553  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=5810289  
HDFS: Number of bytes written=2794844  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=16545  
Total time spent by all reduces in occupied slots (ms)=16520  
Total time spent by all map tasks (ms)=3309  
Total time spent by all reduce tasks (ms)=3304  
Total vcore-milliseconds taken by all map tasks=3309  
Total vcore-milliseconds taken by all reduce tasks=3304  
Total megabyte-milliseconds taken by all map tasks=16942080  
Total megabyte-milliseconds taken by all reduce tasks=16916480

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1229924  
Map output materialized bytes=1229980  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1229980  
Reduce input records=10  
Reduce output records=20  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=43  
CPU time spent (ms)=2890  
Physical memory (bytes) snapshot=708329472  
Virtual memory (bytes) snapshot=12756152320  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=406966272  
Peak Map Virtual memory (bytes)=6375723008  
Peak Reduce Physical memory (bytes)=301363200  
Peak Reduce Virtual memory (bytes)=6384336896

#### 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=2680606

#### File Output Format Counters

Bytes Written=2794844

23/01/04 16:26:15 INFO RepresentativePointsDriver: Representative Points Iteration 1  
23/01/04 16:26:15 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032  
23/01/04 16:26:15 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0975  
23/01/04 16:26:15 INFO FileInputFormat: Total input files to process : 1  
23/01/04 16:26:15 INFO JobSubmitter: number of splits:1  
23/01/04 16:26:15 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0975  
23/01/04 16:26:15 INFO JobSubmitter: Executing with tokens: []  
23/01/04 16:26:15 INFO YarnClientImpl: Submitted application application\_1672817468245\_0975  
23/01/04 16:26:15 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0975/  
23/01/04 16:26:15 INFO Job: Running job: job\_1672817468245\_0975  
23/01/04 16:26:21 INFO Job: Job job\_1672817468245\_0975 running in uber mode : false  
23/01/04 16:26:21 INFO Job: map 0% reduce 0%  
23/01/04 16:26:26 INFO Job: map 100% reduce 0%  
23/01/04 16:26:32 INFO Job: map 100% reduce 100%  
23/01/04 16:26:32 INFO Job: Job job\_1672817468245\_0975 completed successfully  
23/01/04 16:26:32 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1151238  
FILE: Number of bytes written=2835069  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=8270497  
HDFS: Number of bytes written=3946186  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17480  
Total time spent by all reduces in occupied slots (ms)=16455  
Total time spent by all map tasks (ms)=3496  
Total time spent by all reduce tasks (ms)=3291  
Total vcore-milliseconds taken by all map tasks=3496  
Total vcore-milliseconds taken by all reduce tasks=3291  
Total megabyte-milliseconds taken by all map tasks=17899520  
Total megabyte-milliseconds taken by all reduce tasks=16849920

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1151182  
Map output materialized bytes=1151238  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1151238  
Reduce input records=10  
Reduce output records=30  
Spilled Records=20

Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=39  
CPU time spent (ms)=3010  
Physical memory (bytes) snapshot=697774080  
Virtual memory (bytes) snapshot=12741427200  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=402276352  
Peak Map Virtual memory (bytes)=6366179328  
Peak Reduce Physical memory (bytes)=295497728  
Peak Reduce Virtual memory (bytes)=6379020288

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=2680606

File Output Format Counters

Bytes Written=3946186

23/01/04 16:26:32 INFO RepresentativePointsDriver: Representative Points Iteration 2

23/01/04 16:26:32 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:26:32 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0978

23/01/04 16:26:32 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:26:33 INFO JobSubmitter: number of splits:1

23/01/04 16:26:33 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0978

23/01/04 16:26:33 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:26:33 INFO YarnClientImpl: Submitted application application\_1672817468245\_0978

23/01/04 16:26:33 INFO Job: The url to track the job: http://lena-master:8088/proxy/application\_1672817468245\_0978/

23/01/04 16:26:33 INFO Job: Running job: job\_1672817468245\_0978

23/01/04 16:26:39 INFO Job: Job job\_1672817468245\_0978 running in uber mode : false

23/01/04 16:26:39 INFO Job: map 0% reduce 0%

23/01/04 16:26:45 INFO Job: map 100% reduce 0%

23/01/04 16:26:50 INFO Job: map 100% reduce 100%

23/01/04 16:26:51 INFO Job: Job job\_1672817468245\_0978 completed successfully

23/01/04 16:26:51 INFO Job: Counters: 54

File System Counters

FILE: Number of bytes read=1153165  
FILE: Number of bytes written=2838923  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=10573181  
HDFS: Number of bytes written=5099476  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=1



Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17120  
Total time spent by all reduces in occupied slots (ms)=16855  
Total time spent by all map tasks (ms)=3424  
Total time spent by all reduce tasks (ms)=3371  
Total vcore-milliseconds taken by all map tasks=3424  
Total vcore-milliseconds taken by all reduce tasks=3371  
Total megabyte-milliseconds taken by all map tasks=17530880  
Total megabyte-milliseconds taken by all reduce tasks=17259520

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1153110  
Map output materialized bytes=1153165  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1153165  
Reduce input records=10  
Reduce output records=40  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=52  
CPU time spent (ms)=3240  
Physical memory (bytes) snapshot=717385728  
Virtual memory (bytes) snapshot=12752027648  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=425598976  
Peak Map Virtual memory (bytes)=6368333824  
Peak Reduce Physical memory (bytes)=295727104  
Peak Reduce Virtual memory (bytes)=6387904512

#### 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=2680606

#### File Output Format Counters

Bytes Written=5099476

23/01/04 16:26:51 INFO RepresentativePointsDriver: Representative Points Iteration 3

23/01/04 16:26:51 INFO DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at lena-master/128.86.245.64:8032

23/01/04 16:26:51 INFO JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job\_1672817468245\_0981

23/01/04 16:26:51 INFO FileInputFormat: Total input files to process : 1

23/01/04 16:26:51 INFO JobSubmitter: number of splits:1

23/01/04 16:26:51 INFO JobSubmitter: Submitting tokens for job: job\_1672817468245\_0981

23/01/04 16:26:51 INFO JobSubmitter: Executing with tokens: []

23/01/04 16:26:51 INFO YarnClientImpl: Submitted application application\_1672817468245\_0981

23/01/04 16:26:51 INFO Job: The url to track the job:  
http://lena-master:8088/proxy/application\_1672817468245\_0981/  
23/01/04 16:26:51 INFO Job: Running job: job\_1672817468245\_0981  
23/01/04 16:26:57 INFO Job: Job job\_1672817468245\_0981 running in uber mode : false  
23/01/04 16:26:57 INFO Job: map 0% reduce 0%  
23/01/04 16:27:02 INFO Job: map 100% reduce 0%  
23/01/04 16:27:08 INFO Job: map 100% reduce 100%  
23/01/04 16:27:08 INFO Job: Job job\_1672817468245\_0981 completed successfully  
23/01/04 16:27:08 INFO Job: Counters: 54

#### File System Counters

FILE: Number of bytes read=1198984  
FILE: Number of bytes written=2930561  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=12879761  
HDFS: Number of bytes written=6298604  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Rack-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=17665  
Total time spent by all reduces in occupied slots (ms)=16735  
Total time spent by all map tasks (ms)=3533  
Total time spent by all reduce tasks (ms)=3347  
Total vcore-milliseconds taken by all map tasks=3533  
Total vcore-milliseconds taken by all reduce tasks=3347  
Total megabyte-milliseconds taken by all map tasks=18088960  
Total megabyte-milliseconds taken by all reduce tasks=17136640

#### Map-Reduce Framework

Map input records=27  
Map output records=10  
Map output bytes=1198928  
Map output materialized bytes=1198984  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1198984  
Reduce input records=10  
Reduce output records=50  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=76  
CPU time spent (ms)=3080  
Physical memory (bytes) snapshot=721457152  
Virtual memory (bytes) snapshot=12753768448  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=415686656  
Peak Map Virtual memory (bytes)=6372233216  
Peak Reduce Physical memory (bytes)=309661696  
Peak Reduce Virtual memory (bytes)=6385586176

#### 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=2680606
File Output Format Counters
  Bytes Written=6298604
23/01/04 16:27:08 INFO RepresentativePointsDriver: Representative Points Iteration
4
23/01/04 16:27:08 INFO DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at lena-master/128.86.245.64:8032
23/01/04 16:27:08 INFO JobResourceUploader: Disabling Erasure Coding for path:
/tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0984
23/01/04 16:27:09 INFO FileInputFormat: Total input files to process : 1
23/01/04 16:27:09 INFO JobSubmitter: number of splits:1
23/01/04 16:27:09 INFO JobSubmitter: Submitting tokens for job:
job_1672817468245_0984
23/01/04 16:27:09 INFO JobSubmitter: Executing with tokens: []
23/01/04 16:27:09 INFO YarnClientImpl: Submitted application
application_1672817468245_0984
23/01/04 16:27:09 INFO Job: The url to track the job:
http://lena-master:8088/proxy/application_1672817468245_0984/
23/01/04 16:27:09 INFO Job: Running job: job_1672817468245_0984
23/01/04 16:27:15 INFO Job: Job job_1672817468245_0984 running in uber mode : false
23/01/04 16:27:15 INFO Job:  map 0% reduce 0%
23/01/04 16:27:21 INFO Job:  map 100% reduce 0%
23/01/04 16:27:27 INFO Job:  map 100% reduce 100%
23/01/04 16:27:27 INFO Job: Job job_1672817468245_0984 completed successfully
23/01/04 16:27:27 INFO Job: Counters: 54
File System Counters
  FILE: Number of bytes read=1178454
  FILE: Number of bytes written=2889501
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=15278017
  HDFS: Number of bytes written=7477182
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=17925
  Total time spent by all reduces in occupied slots (ms)=16550
  Total time spent by all map tasks (ms)=3585
  Total time spent by all reduce tasks (ms)=3310
  Total vcore-milliseconds taken by all map tasks=3585
  Total vcore-milliseconds taken by all reduce tasks=3310
  Total megabyte-milliseconds taken by all map tasks=18355200
  Total megabyte-milliseconds taken by all reduce tasks=16947200
Map-Reduce Framework
  Map input records=27
  Map output records=10
```

Map output bytes=1178398  
Map output materialized bytes=1178454  
Input split bytes=203  
Combine input records=0  
Combine output records=0  
Reduce input groups=10  
Reduce shuffle bytes=1178454  
Reduce input records=10  
Reduce output records=60  
Spilled Records=20  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=80  
CPU time spent (ms)=3490  
Physical memory (bytes) snapshot=708071424  
Virtual memory (bytes) snapshot=12749455360  
Total committed heap usage (bytes)=2105540608  
Peak Map Physical memory (bytes)=408535040  
Peak Map Virtual memory (bytes)=6371090432  
Peak Reduce Physical memory (bytes)=299536384  
Peak Reduce Virtual memory (bytes)=6378364928

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=2680606

File Output Format Counters

Bytes Written=7477182

23/01/04 16:27:27 INFO ClusterEvaluator: Scaled Inter-Cluster Density =  
0.35303095125361245  
23/01/04 16:27:27 INFO ClusterEvaluator: Intra-Cluster Density[0] =  
0.7139889475666085  
23/01/04 16:27:27 INFO ClusterEvaluator: Intra-Cluster Density[1] =  
0.5666666666666667  
23/01/04 16:27:27 INFO ClusterEvaluator: Intra-Cluster Density[13] = NaN  
23/01/04 16:27:27 INFO ClusterEvaluator: Intra-Cluster Density[3] =  
0.5666666666666667  
23/01/04 16:27:27 INFO ClusterEvaluator: Intra-Cluster Density[22] =  
0.5666666666666667  
23/01/04 16:27:28 INFO ClusterEvaluator: Intra-Cluster Density[17] = NaN  
23/01/04 16:27:28 INFO ClusterEvaluator: Intra-Cluster Density[12] = NaN  
23/01/04 16:27:28 INFO ClusterEvaluator: Intra-Cluster Density[11] =  
0.478705230126814  
23/01/04 16:27:28 INFO ClusterEvaluator: Intra-Cluster Density[14] = NaN  
23/01/04 16:27:28 INFO ClusterEvaluator: Intra-Cluster Density[20] = NaN  
23/01/04 16:27:28 INFO ClusterEvaluator: Average Intra-Cluster Density =  
0.5785388355386845  
23/01/04 16:27:32 INFO ClusterDumper: Wrote 10 clusters  
23/01/04 16:27:32 INFO MahoutDriver: Program took 97994 ms (Minutes:  
1.6332333333333333)