**Communication for 100**

we run 6 steps in this project.

In each step you can see the data that entered /exit from mappers+ reducers.

**Record reader:**

File System Counters

**FILE: Number of bytes read=40773418336**

FILE: Number of bytes written=60115714191

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=8316**

HDFS: Number of bytes written=0

**HDFS: Number of read operations=99**

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=59059468351**

**S3: Number of bytes written=141388695**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Failed reduce tasks=1

Killed map tasks=1

Killed reduce tasks=1

Launched map tasks=99

Launched reduce tasks=10

Data-local map tasks=99

**Total time spent by all maps in occupied slots (ms)=5325434016**

**Total time spent by all reduces in occupied slots (ms)=4087413360**

**Total time spent by all map tasks (ms)=55473271**

**Total time spent by all reduce tasks (ms)=85154445**

Total vcore-milliseconds taken by all map tasks=55473271

Total vcore-milliseconds taken by all reduce tasks=85154445

Total megabyte-milliseconds taken by all map tasks=170413888512

Total megabyte-milliseconds taken by all reduce tasks=130797227520

Map-Reduce Framework

**Map input records=1783018535**

**Map output records=6136060879**

**Map output bytes=89760355728**

**Map output materialized bytes=19596727595**

Input split bytes=8316

Combine input records=0

Combine output records=0

Reduce input groups=8250505

Reduce shuffle bytes=19596727595

Reduce input records=6136060879

Reduce output records=8250505

Spilled Records=18408182637

Shuffled Maps =792

Failed Shuffles=0

Merged Map outputs=792

GC time elapsed (ms)=158127

CPU time spent (ms)=98634650

Physical memory (bytes) snapshot=76505862144

Virtual memory (bytes) snapshot=204382560256

Total committed heap usage (bytes)=61932044288

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=59059468351

File Output Format Counters

Bytes Written=141388695

**indexing:**

File System Counters

**FILE: Number of bytes read=88835151**

**FILE: Number of bytes written=179730933**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=800**

HDFS: Number of bytes written=0

HDFS: Number of read operations=8

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=141388695**

**S3: Number of bytes written=231033140**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Killed map tasks=1

Killed reduce tasks=2

Launched map tasks=8

Launched reduce tasks=9

Data-local map tasks=8

**Total time spent by all maps in occupied slots (ms)=10238496**

**Total time spent by all reduces in occupied slots (ms)=15531072**

Total time spent by all map tasks (ms)=106651

Total time spent by all reduce tasks (ms)=323564

Total vcore-milliseconds taken by all map tasks=106651

Total vcore-milliseconds taken by all reduce tasks=323564

Total megabyte-milliseconds taken by all map tasks=327631872

Total megabyte-milliseconds taken by all reduce tasks=496994304

Map-Reduce Framework

**Map input records=8250505**

**Map output records=8250505**

**Map output bytes=157889718**

**Map output materialized bytes=88839378**

Input split bytes=800

Combine input records=0

Combine output records=0

Reduce input groups=1

Reduce shuffle bytes=88839378

Reduce input records=8250505

Reduce output records=8250505

Spilled Records=16501010

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=3700

CPU time spent (ms)=139800

Physical memory (bytes) snapshot=8093073408

Virtual memory (bytes) snapshot=30710046720

Total committed heap usage (bytes)=7357333504

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=141388695

File Output Format Counters

Bytes Written=231033140

**word vector**:

File System Counters

**FILE: Number of bytes read=88835151**

**FILE: Number of bytes written=179730933**

**FILE: Number of read operations=0**

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=800**

**HDFS: Number of bytes written=0**

HDFS: Number of read operations=8

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=141388695**

**S3: Number of bytes written=231033140**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Killed map tasks=1

Killed reduce tasks=2

Launched map tasks=8

Launched reduce tasks=9

Data-local map tasks=8

Total time spent by all maps in occupied slots (ms)=10238496

Total time spent by all reduces in occupied slots (ms)=15531072

Total time spent by all map tasks (ms)=106651

Total time spent by all reduce tasks (ms)=323564

Total vcore-milliseconds taken by all map tasks=106651

Total vcore-milliseconds taken by all reduce tasks=323564

Total megabyte-milliseconds taken by all map tasks=327631872

Total megabyte-milliseconds taken by all reduce tasks=496994304

Map-Reduce Framework

**Map input records=8250505**

**Map output records=8250505**

**Map output bytes=157889718**

Map output materialized bytes=88839378

Input split bytes=800

Combine input records=0

Combine output records=0

Reduce input groups=1

Reduce shuffle bytes=88839378

Reduce input records=8250505

Reduce output records=8250505

Spilled Records=16501010

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=3700

CPU time spent (ms)=139800

Physical memory (bytes) snapshot=8093073408

Virtual memory (bytes) snapshot=30710046720

Total committed heap usage (bytes)=7357333504

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=141388695

File Output Format Counters

Bytes Written=231033140

**four vectors**:

File System Counters

FILE: Number of bytes read=4222229573

FILE: Number of bytes written=6413309959

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=1764

HDFS: Number of bytes written=0

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

S3: Number of bytes read=1103913231

S3: Number of bytes written=3758088724

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Killed map tasks=1

Killed reduce tasks=1

Launched map tasks=18

Launched reduce tasks=9

Data-local map tasks=18

Total time spent by all maps in occupied slots (ms)=421887744

Total time spent by all reduces in occupied slots (ms)=187992384

Total time spent by all map tasks (ms)=4394664

Total time spent by all reduce tasks (ms)=3916508

Total vcore-milliseconds taken by all map tasks=4394664

Total vcore-milliseconds taken by all reduce tasks=3916508

Total megabyte-milliseconds taken by all map tasks=13500407808

Total megabyte-milliseconds taken by all reduce tasks=6015756288

Map-Reduce Framework

**Map input records=5256**

**Map output records=5256**

**Map output bytes=3758103195**

**Map output materialized bytes=2194181856**

Input split bytes=1764

Combine input records=0

Combine output records=0

Reduce input groups=5256

Reduce shuffle bytes=2194181856

Reduce input records=5256

Reduce output records=5256

Spilled Records=15325

Shuffled Maps =144

Failed Shuffles=0

Merged Map outputs=144

GC time elapsed (ms)=18946

CPU time spent (ms)=4477000

Physical memory (bytes) snapshot=16976371712

Virtual memory (bytes) snapshot=49833304064

Total committed heap usage (bytes)=15187574784

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=1103913231

File Output Format Counters

Bytes Written=3758088724

**make pair**:

File System Counters

**FILE: Number of bytes read=21033197735**

**FILE: Number of bytes written=31821067292**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=5742**

**HDFS: Number of bytes written=0**

HDFS: Number of read operations=58

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=3816853366**

**S3: Number of bytes written=18571815958**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Failed reduce tasks=1

Killed map tasks=1

Killed reduce tasks=3

Launched map tasks=58

Launched reduce tasks=12

Data-local map tasks=58

Total time spent by all maps in occupied slots (ms)=1178134016

Total time spent by all reduces in occupied slots (ms)=499679520

Total time spent by all map tasks (ms)=9204172

Total time spent by all reduce tasks (ms)=10409990

Total vcore-milliseconds taken by all map tasks=9204172

Total vcore-milliseconds taken by all reduce tasks=10409990

Total megabyte-milliseconds taken by all map tasks=37700288512

Total megabyte-milliseconds taken by all reduce tasks=15989744640

Map-Reduce Framework

**Map input records=5256**

**Map output records=28580**

**Map output bytes=18832454054**

**Map output materialized bytes=10812763738**

Input split bytes=5742

Combine input records=0

Combine output records=0

Reduce input groups=200

Reduce shuffle bytes=10812763738

Reduce input records=28580

Reduce output records=27892

Spilled Records=83651

Shuffled Maps =464

Failed Shuffles=0

Merged Map outputs=464

GC time elapsed (ms)=51613

CPU time spent (ms)=9900730

Physical memory (bytes) snapshot=47889620992

Virtual memory (bytes) snapshot=126272872448

Total committed heap usage (bytes)=43517476864

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=3816853366

File Output Format Counters

Bytes Written=18571815958

**similarity**:

File System Counters

**FILE: Number of bytes read=4230074**

**FILE: Number of bytes written=45207630**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=26784**

HDFS: Number of bytes written=0

**HDFS: Number of read operations=279**

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=19089054144**

**S3: Number of bytes written=5409320**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Killed map tasks=1

Killed reduce tasks=1

**Launched map tasks=279**

**Launched reduce tasks=9**

**Data-local map tasks=279**

**Total time spent by all maps in occupied slots (ms)=4634933760**

**Total time spent by all reduces in occupied slots (ms)=1462534992**

**Total time spent by all map tasks (ms)=36210420**

**Total time spent by all reduce tasks (ms)=30469479**

**Total vcore-milliseconds taken by all map tasks=36210420**

**Total vcore-milliseconds taken by all reduce tasks=30469479**

**Total megabyte-milliseconds taken by all map tasks=148317880320**

**Total megabyte-milliseconds taken by all reduce tasks=46801119744**

Map-Reduce Framework

**Map input records=27892**

**Map output records=14047**

**Map output bytes=5437383**

**Map output materialized bytes=4322787**

**Input split bytes=26784**

Combine input records=0

Combine output records=0

**Reduce input groups=200**

**Reduce shuffle bytes=4322787**

**Reduce input records=14047**

**Reduce output records=14047**

**Spilled Records=28094**

**Shuffled Maps =2232**

Failed Shuffles=0

**Merged Map outputs=2232**

**GC time elapsed (ms)=50858**

**CPU time spent (ms)=32926480**

**Physical memory (bytes) snapshot=177041129472**

**Virtual memory (bytes) snapshot=548063649792**

**Total committed heap usage (bytes)=158338646016**

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

**File Input Format Counters**

**Bytes Read=19089054144**

**File Output Format Counters**

**Bytes Written=5409320**