**Communication for 10**

we have 6 steps, here is some details on how many keys+data pass from the reducers+mappers

**Record reader:**

File System Counters

**FILE: Number of bytes read=3783023959**

**FILE: Number of bytes written=5578401786**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=840**

HDFS: Number of bytes written=0

HDFS: Number of read operations=10

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=5957272885**

**S3: Number of bytes written=40037447**

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

**Launched map tasks=10**

**Launched reduce tasks=8**

Data-local map tasks=10

**Total time spent by all maps in occupied slots (ms)=508899840**

**Total time spent by all reduces in occupied slots (ms)=385332240**

**Total time spent by all map tasks (ms)=5301040**

**Total time spent by all reduce tasks (ms)=8027755**

Total vcore-milliseconds taken by all map tasks=5301040

Total vcore-milliseconds taken by all reduce tasks=8027755

Total megabyte-milliseconds taken by all map tasks=16284794880

Total megabyte-milliseconds taken by all reduce tasks=12330631680

Map-Reduce Framework

**Map input records=175910922**

**Map output records=565014321**

**Map output bytes=8303583404**

Map output materialized bytes=1818026179

Input split bytes=840

Combine input records=0

Combine output records=0

Reduce input groups=2311543

Reduce shuffle bytes=1818026179

Reduce input records=565014321

Reduce output records=2311543

Spilled Records=1695042963

Shuffled Maps =80

Failed Shuffles=0

Merged Map outputs=80

GC time elapsed (ms)=35527

CPU time spent (ms)=9288810

Physical memory (bytes) snapshot=13934362624

Virtual memory (bytes) snapshot=34470985728

Total committed heap usage (bytes)=11708923904

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

**Bytes Read=5957272885**

File Output Format Counters

Bytes Written=40037447

**indexing:**

File System Counters

**FILE: Number of bytes read=24854090**

**FILE: Number of bytes written=51766636**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=792

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=40037447**

**S3: Number of bytes written=64353310**

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=8

Launched reduce tasks=9

Data-local map tasks=8

**Total time spent by all maps in occupied slots (ms)=9144384**

**Total time spent by all reduces in occupied slots (ms)=9770304**

**Total time spent by all map tasks (ms)=95254**

**Total time spent by all reduce tasks (ms)=203548**

Total vcore-milliseconds taken by all map tasks=95254

Total vcore-milliseconds taken by all reduce tasks=203548

Total megabyte-milliseconds taken by all map tasks=292620288

Total megabyte-milliseconds taken by all reduce tasks=312649728

Map-Reduce Framework

**Map input records=2311543**

**Map output records=2311543**

**Map output bytes=44660533**

**Map output materialized bytes=24856407**

Input split bytes=792

Combine input records=0

Combine output records=0

Reduce input groups=1

Reduce shuffle bytes=24856407

Reduce input records=2311543

Reduce output records=2311543

Spilled Records=4623086

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=3776

CPU time spent (ms)=59030

Physical memory (bytes) snapshot=7823335424

Virtual memory (bytes) snapshot=30654091264

Total committed heap usage (bytes)=7192182784

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

**File Input Format Counters**

**Bytes Read=40037447**

**File Output Format Counters**

**Bytes Written=64353310**

**word vector**:

File System Counters

**FILE: Number of bytes read=561473874**

**FILE: Number of bytes written=851253931**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=840**

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

HDFS: Number of read operations=10

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

**S3: Number of bytes read=5957272885**

**S3: Number of bytes written=73498029**

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=10

Launched reduce tasks=9

Data-local map tasks=10

**Total time spent by all maps in occupied slots (ms)=130237632**

**Total time spent by all reduces in occupied slots (ms)=102321216**

**Total time spent by all map tasks (ms)=1356642**

**Total time spent by all reduce tasks (ms)=2131692**

Total vcore-milliseconds taken by all map tasks=1356642

Total vcore-milliseconds taken by all reduce tasks=2131692

Total megabyte-milliseconds taken by all map tasks=4167604224

Total megabyte-milliseconds taken by all reduce tasks=3274278912

Map-Reduce Framework

**Map input records=175910922**

**Map output records=43090238**

**Map output bytes=960790314**

**Map output materialized bytes=290073677**

Input split bytes=840

Combine input records=0

Combine output records=0

Reduce input groups=415

Reduce shuffle bytes=290073677

Reduce input records=43090238

Reduce output records=415

Spilled Records=126144579

Shuffled Maps =80

Failed Shuffles=0

Merged Map outputs=80

GC time elapsed (ms)=8264

CPU time spent (ms)=1931300

Physical memory (bytes) snapshot=11536703488

Virtual memory (bytes) snapshot=34474352640

Total committed heap usage (bytes)=10129768448

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=5957272885

File Output Format Counters

Bytes Written=73498029

**four vectors**:

File System Counters

**FILE: Number of bytes read=139824725**

**FILE: Number of bytes written=281695408**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=776**

**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=73498029**

**S3: Number of bytes written=246405224**

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=8

Launched reduce tasks=9

Data-local map tasks=8

**Total time spent by all maps in occupied slots (ms)=36826560**

**Total time spent by all reduces in occupied slots (ms)=15624768**

Total time spent by all map tasks (ms)=383610

Total time spent by all reduce tasks (ms)=325516

Total vcore-milliseconds taken by all map tasks=383610

Total vcore-milliseconds taken by all reduce tasks=325516

Total megabyte-milliseconds taken by all map tasks=1178449920

Total megabyte-milliseconds taken by all reduce tasks=499992576

Map-Reduce Framework

Map input records=415

Map output records=415

Map output bytes=246406360

Map output materialized bytes=139825589

Input split bytes=776

Combine input records=0

Combine output records=0

Reduce input groups=415

Reduce shuffle bytes=139825589

Reduce input records=415

Reduce output records=415

Spilled Records=830

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=4819

CPU time spent (ms)=327320

Physical memory (bytes) snapshot=8345595904

Virtual memory (bytes) snapshot=30706421760

Total committed heap usage (bytes)=7399276544

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=73498029

File Output Format Counters

Bytes Written=246405224

**make pair**:

File System Counters

**FILE: Number of bytes read=564742622**

**FILE: Number of bytes written=1032164864**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

**HDFS: Number of bytes read=784**

**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=246405224**

**S3: Number of bytes written=58072449**

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=8

Data-local map tasks=8

**Total time spent by all maps in occupied slots (ms)=59108928**

**Total time spent by all reduces in occupied slots (ms)=24671280**

**Total time spent by all map tasks (ms)=615718**

**Total time spent by all reduce tasks (ms)=513985**

Total vcore-milliseconds taken by all map tasks=615718

Total vcore-milliseconds taken by all reduce tasks=513985

Total megabyte-milliseconds taken by all map tasks=1891485696

Total megabyte-milliseconds taken by all reduce tasks=789480960

Map-Reduce Framework

**Map input records=415**

**Map output records=1777**

**Map output bytes=835482124**

**Map output materialized bytes=465766110**

Input split bytes=784

Combine input records=0

Combine output records=0

Reduce input groups=198

Reduce shuffle bytes=465766110

Reduce input records=1777

Reduce output records=100

Spilled Records=3959

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=6508

CPU time spent (ms)=567280

Physical memory (bytes) snapshot=8948654080

Virtual memory (bytes) snapshot=30670344192

Total committed heap usage (bytes)=7949254656

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=246405224

File Output Format Counters

Bytes Written=58072449

**similarity**:

File System Counters

**FILE: Number of bytes read=16488**

**FILE: Number of bytes written=2075543**

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=760

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=58072449**

**S3: Number of bytes written=19913**

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

Job Counters

Killed reduce tasks=1

Launched map tasks=8

Launched reduce tasks=8

Other local map tasks=2

Data-local map tasks=6

**Total time spent by all maps in occupied slots (ms)=18740160**

**Total time spent by all reduces in occupied slots (ms)=7025808**

**Total time spent by all map tasks (ms)=195210**

**Total time spent by all reduce tasks (ms)=146371**

Total vcore-milliseconds taken by all map tasks=195210

Total vcore-milliseconds taken by all reduce tasks=146371

Total megabyte-milliseconds taken by all map tasks=599685120

Total megabyte-milliseconds taken by all reduce tasks=224825856

Map-Reduce Framework

Map input records=100

Map output records=54

Map output bytes=20015

Map output materialized bytes=17352

Input split bytes=760

Combine input records=0

Combine output records=0

Reduce input groups=12

Reduce shuffle bytes=17352

Reduce input records=54

Reduce output records=54

Spilled Records=108

Shuffled Maps =64

Failed Shuffles=0

Merged Map outputs=64

GC time elapsed (ms)=3510

CPU time spent (ms)=108640

Physical memory (bytes) snapshot=7838605312

Virtual memory (bytes) snapshot=30664699904

Total committed heap usage (bytes)=7248805888

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

**File Input Format Counters**

**Bytes Read=58072449**

**File Output Format Counters**

**Bytes Written=19913**