**Step 1**

Without combiner:

Map input records=**163471963**

Map output records=**277762412**

Map output bytes=5648285810

Map output materialized bytes=721586150

Input split bytes=3264

Combine input records=0

Combine output records=0

Reduce input groups=2519316

Reduce shuffle bytes=721586150

Reduce input records=**277762412**

Reduce output records=**2519316**

Spilled Records=833287236

Shuffled Maps =24

Failed Shuffles=0

Merged Map outputs=24

GC time elapsed (ms)=25302

CPU time spent (ms)=1749340

Physical memory (bytes) snapshot=21563535360

Virtual memory (bytes) snapshot=84465389568

Total committed heap usage (bytes)=18353750016

With combiner:

Map input records=**163471963**

Map output records=**277762412**

Map output bytes=5648285810

Map output materialized bytes=46450635

Input split bytes=3264

Combine input records=**281972152**

Combine output records=**8046830**

Reduce input groups=2519316

Reduce shuffle bytes=46450635

Reduce input records=**3837090**

Reduce output records=**2519316**

Spilled Records=11883920

Shuffled Maps =24

Failed Shuffles=0

Merged Map outputs=24

GC time elapsed (ms)=25062

CPU time spent (ms)=1503290

Physical memory (bytes) snapshot=21325328384

Virtual memory (bytes) snapshot=84456501248

Total committed heap usage (bytes)=18223202304

**Why is the number of combiner input records more than the number of outputs of maps?** [**stackoverflow**](https://stackoverflow.com/questions/12171965/why-is-the-number-of-combiner-input-records-more-than-the-number-of-outputs-of-m)

I think it's because the Combiner can also run on the output of previous Combine steps, since your Combiner runs and produces new records which are then Combined with other records coming out of your Mappers. It may also be that Map output records is calculated after the Combiner runs, meaning that there are less records because some have been Combined.

**Step 2**

Without combiner:

Map input records=**2519316**

Map output records=**7470120**

Map output bytes=218983627

Map output materialized bytes=93612175

Input split bytes=232

Combine input records=0

Combine output records=0

Reduce input groups=2616643

Reduce shuffle bytes=93612175

Reduce input records=**7470120**

Reduce output records=**1747596**

Spilled Records=22410360

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=2042

CPU time spent (ms)=79280

Physical memory (bytes) snapshot=2709528576

Virtual memory (bytes) snapshot=11331190784

Total committed heap usage (bytes)=2346713088

With combiner:

Map input records=**2519316**

Map output records=**7470120**

Map output bytes=218983627

Map output materialized bytes=73437803

Input split bytes=210

Combine input records=**7470120**

Combine output records=**2676065**

Reduce input groups=2616643

Reduce shuffle bytes=73437803

Reduce input records=**2676065**

Reduce output records=**1747596**

Spilled Records=8028195

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=2585

CPU time spent (ms)=76890

Physical memory (bytes) snapshot=2901917696

Virtual memory (bytes) snapshot=11329560576

Total committed heap usage (bytes)=2495610880

**Step 3 – No combiner needed**

Without combiner:

Map input records=**1747596**

Map output records=**1747596**

Map output bytes=153102322

Map output materialized bytes=74891838

Input split bytes=210

Combine input records=0

Combine output records=0

Reduce input groups=**1747596**

Reduce shuffle bytes=74891838

Reduce input records=1747596

Reduce output records=**1650268**

Spilled Records=3495192

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=2013

CPU time spent (ms)=45920

Physical memory (bytes) snapshot=2859589632

Virtual memory (bytes) snapshot=11299520512

Total committed heap usage (bytes)=2466250752

**Step 4 – No combiner needed**

Without combiner:

Map input records=**1650268**

Map output records=**1650268**

Map output bytes=76728194

Map output materialized bytes=43809121

Input split bytes=210

Combine input records=0

Combine output records=0

Reduce input groups=1650268

Reduce shuffle bytes=43809121

Reduce input records=**1650268**

Reduce output records=1650268

Spilled Records=3300536

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=2901

CPU time spent (ms)=124760

Physical memory (bytes) snapshot=2794676224

Virtual memory (bytes) snapshot=11327733760

Total committed heap usage (bytes)=2412773376