	Hadoop	2 cores m5.xlarge	5 cores m5.xlarge	7 cores m5.xlarge
FILE: Number of bytes read	2700772	361829	362113	362329
FILE: Number of bytes written	6117970	3887167	8631477	11176373
FILE: Number of read operations	0	0	0	0
FILE: Number of large read operations	0	0	0	0
FILE: Number of write operations	0	0	0	0
HDFS: Number of bytes read	2784184	3952656	4052656	4567247
HDFS: Number of bytes written	4698	2700532	2700532	2700532
HDFS: Number of read operations	11	35	48	70
HDFS: Number of large read operations	0	0	0	0
HDFS: Number of write operations	4698	6	18	
HDFS: Number of bytes read erasure-coded	0	0	0	0
S3: Number of bytes read		15257243	15357618	15469079
S3: Number of bytes written		4698	4698	4698
S3: Number of read operations		0	0	0
S3: Number of large read operations		0	0	0
S3: Number of write operations		0	0	0
Job:Killed map tasks	1	1	1	1
Job:Launched map tasks	2	10	22	35
Job:Launched reduce tasks	1	3	9	13
Job:Data-local map tasks	2	10	22	32
Job:Total time spent by all maps in occupied slots (ms)	43313	10058592	28694592	37227072
Job:Total time spent by all reduces in occupied slots (ms)	15732	2233344	7260672	16563072
Job:Total time spent by all map tasks (ms)	43313	104777	298902	387782
Job:Total time spent by all reduce tasks (ms)	15732	11632	37816	86266
Job:Total vcore-milliseconds taken by all map tasks	43313	104777	298902	387782
Job:Total vcore-milliseconds taken by all reduce tasks	15732	11632	37816	86266
Job:Total megabyte-milliseconds taken by all map tasks	44352512	321874944	918226944	1191266304
Job:Total megabyte-milliseconds taken by all reduce tasks	16109568	71467008	232341504	530018304
MR: Map input records	39	30000	30000	30000
MR:Map output records	39	72000	72000	72000
MR:Map output bytes	2700610	3365810	3365810	3365810
MR:Map output materialized bytes	2700778	388703	412672	431233

MR:Input split bytes	318	960	2112	2880
MR:Combine input records	0	0	0	0
MR: Combine output records	0	0	0	0
MR:Reduce input groups	39	39	39	39
MR:Reduce shuffle bytes	2700778	388703	412672	431233
MR:Reduce input records	39	72000	72000	72000
MR:Reduce output records	39	39	39	39
MR:Spilled Records	78	144000	144000	144000
MR:Shuffled Maps	2	30	198	390
MR:Failed Shuffles	0	0	0	0
MR:Merged Map outputs	2	30	198	390
MR:GC time elapsed (ms)	1011	2815	7326	11258
MR:CPU time spent (ms)	3780	62860	138170	188710
MR:Physical memory (bytes) snapshot	554360832	6672097280	15757266944	21893304320
MR:Virtual memory (bytes) snapshot	7450382336	65699921920	1,6174E+11	2,26022E+11
MR:Total committed heap usage (bytes)	349904896	6655311872	15335424000	21317025792
MR:Peak Map Physical memory (bytes)	218980352	623874048	719966208	944103424
MR:Peak Map Virtual memory (bytes)	2480910336	4486156288	4497489920	4511358976
MR:Peak Reduce Physical memory (bytes)	117026816	232091648	239951872	242769920
MR:Peak Reduce Virtual memory (bytes)	2488561664	7089987584	7096573952	7095021568
File Input: Bytes Read	2783866	15257243	15357618	15469079
File Output: Bytes Read	4698	2700532	2700532	2700532

Escogí utilizar instancias de tipo m5.xlarge para hacer las pruebas en EMR ya que los clústers pedían instancias grandes para poder llevar a cabo las operaciones con Hadoop. Utilicé 2 5 y 7 núcleos ya que eran números equilibrados, además, a partir de 8 núcleos el clúster se volvía más inestable y empezaba a fallar.

En los datos de Hadoop no están los datos de S3 ya que evidentemente en local no se interactúa con S3. En el caso de Hadoop hay un único nodo pseudodistribuido, mientras que en EMR el número más bajo es de 2. Se puede observar como a medida que el número de núcleos aumenta, las estadísticas relacionadas con bytes y los tiempos se mantienen iguales o aumentan, nunca disminuyen. Hay otras estadísticas directamente relacionadas con el número de nodos, como el *Launched reduce tasks* que puede obtenerse como el número de nodos multiplicado por 2 y menos 1. En general las estadísticas presentan una tendencia común.