**PRACTICA 1 HADOOP MAP REDUCE**

Dado el fichero medidas.txt que tiene el siguiente aspecto:



2017 Enero 46

2017 Enero 10

2017 Enero 6

2017 Enero 9

2017 Enero 34

2017 Enero -3

2017 Enero 48

2017 Enero 20

2017 Enero -5

2017 Enero 0

…..

**Se pide calcular a través de un map-reduce la temperatura máxima para cada año.**

En la entrega debe ir lo siguiente:

* Fichero con el mapper
* Fichero con el reducer
* Resultado del trabajo obtenido
* Log de la ejecución

El log es algo así.

packageJobJar: [] [/usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar] /tmp/streamjob9116962918006103657.jar tmpDir=null

2023-09-16 12:09:28,011 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032

2023-09-16 12:09:28,532 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032

2023-09-16 12:09:39,075 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job\_1694856642000\_0001

2023-09-16 12:09:39,830 INFO mapred.FileInputFormat: Total input files to process : 1

2023-09-16 12:09:40,417 INFO mapreduce.JobSubmitter: number of splits:2

2023-09-16 12:09:41,313 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1694856642000\_0001

2023-09-16 12:09:41,314 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-09-16 12:09:41,738 INFO conf.Configuration: resource-types.xml not found

2023-09-16 12:09:41,740 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-09-16 12:09:42,400 INFO impl.YarnClientImpl: Submitted application application\_1694856642000\_0001

2023-09-16 12:09:42,517 INFO mapreduce.Job: The url to track the job: http://LAPTOP.localdomain:8088/proxy/application\_1694856642000\_0001/

2023-09-16 12:09:42,523 INFO mapreduce.Job: Running job: job\_1694856642000\_0001

2023-09-16 12:10:25,188 INFO mapreduce.Job: Job job\_1694856642000\_0001 running in uber mode : false

2023-09-16 12:10:25,190 INFO mapreduce.Job: map 0% reduce 0%

2023-09-16 12:10:49,908 INFO mapreduce.Job: map 100% reduce 0%

2023-09-16 12:11:10,491 INFO mapreduce.Job: map 100% reduce 100%

2023-09-16 12:11:12,566 INFO mapreduce.Job: Job job\_1694856642000\_0001 completed successfully

2023-09-16 12:11:12,802 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=281

FILE: Number of bytes written=838314

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=429

HDFS: Number of bytes written=172

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=2

Launched reduce tasks=1

Data-local map tasks=2

Total time spent by all maps in occupied slots (ms)=45176

Total time spent by all reduces in occupied slots (ms)=18221

Total time spent by all map tasks (ms)=45176

Total time spent by all reduce tasks (ms)=18221

Total vcore-milliseconds taken by all map tasks=45176

Total vcore-milliseconds taken by all reduce tasks=18221

Total megabyte-milliseconds taken by all map tasks=46260224

Total megabyte-milliseconds taken by all reduce tasks=18658304

Map-Reduce Framework

Map input records=4

Map output records=27

Map output bytes=221

Map output materialized bytes=287

Input split bytes=172

Combine input records=0

Combine output records=0

Reduce input groups=20

Reduce shuffle bytes=287

Reduce input records=27

Reduce output records=20

Spilled Records=54

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=174

CPU time spent (ms)=4150

Physical memory (bytes) snapshot=760680448

Virtual memory (bytes) snapshot=2880814751744

Total committed heap usage (bytes)=761266176

Peak Map Physical memory (bytes)=292495360

Peak Map Virtual memory (bytes)=1241994481664

Peak Reduce Physical memory (bytes)=191803392

Peak Reduce Virtual memory (bytes)=884166356992

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=257

File Output Format Counters

Bytes Written=172

2023-09-16 12:11:12,879 INFO streaming.StreamJob: Output directory: hdfs:///user/salida