Copiamos la traza pcap y el fichero hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar al cluster

```
merve@onur-ideacenter:~/Desktop/Infraestructura_para_big_data/3Evaluación_de_prestaciones$ scp traza_serde2.pcap
uambd20@150.244.65.34:.
uambd20@150.244.65.34's password:
traza_serde2.pcap                                          100%  146MB  10.1MB/s    00:14
merve@onur-ideacenter:~/Desktop/Infraestructura_para_big_data/3Evaluación_de_prestaciones$ scp hadoop-pcap-serde-
1.2-SNAPSHOT-jar-with-dependencies.jar uambd20@150.244.65.34:.
uambd20@150.244.65.34's password:
hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar    100% 3468KB   8.6MB/s    00:00
```

```
merve@onur-ideacenter:~/Desktop/Infraestructura_para_big_data/3Evaluación_de_prestaciones$ ssh uambd20@150.244.65.34
uambd20@150.244.65.34's password:
Last login: Wed Jan 31 00:14:25 2024 from 172.30.130.89
[uambd20@master ~]$ ls
hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar  traza_serde2.pcap  veryBig.txt
TestDFSIO_results.log                                      veryBig2.txt
```

Creamos directorios libs y pcaps en el HDFS y movemos la traza pcap y el fichero jar

```
                                    uambd20@master:~                              _  □  ✕

File  Edit  View  Search  Terminal  Help
mkdir: `/home/uambd20/pcaps': No such file or directory
[uambd20@master ~]$ hadoop fs -mkdir /user/uambd20/pcaps
[uambd20@master ~]$ hadoop fs -mkdir /user/uambd20/libs
[uambd20@master ~]$ hadoop fs -put traza_serde2.pcap /user/uambd20/pcaps
[uambd20@master ~]$ hadoop fs -put hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar /user/
uambd20/libs
[uambd20@master ~]$ hadoop fs -ls /user/uambd20/pcaps
Found 1 items
-rw-r--r--   3 uambd20 uambd20  153403944 2024-01-31 00:44 /user/uambd20/pcaps/traza_serde2.pcap
[uambd20@master ~]$ hadoop fs -ls /user/uambd20/libs
Found 1 items
-rw-r--r--   3 uambd20 uambd20    3551149 2024-01-31 00:45 /user/uambd20/libs/hadoop-pcap-serde-1.
2-SNAPSHOT-jar-with-dependencies.jar
[uambd20@master ~]$
```

También creamos el directorio hive_script para guardar los scripts de Hive

```
[uambd20@master ~]$ mkdir /home/uambd20/hive_scripts/tables
[uambd20@master ~]$ mkdir /home/uambd20/hive_scripts/otros
[uambd20@master ~]$ mkdir /home/uambd20/hive_scripts/ejercicios
[uambd20@master ~]$ ls
hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar  traza_serde2.pcap
hive_scripts                                              veryBig2.txt
TestDFSIO_results.log                                     veryBig.txt
[uambd20@master ~]$ cd hive_scripts/
[uambd20@master hive_scripts]$ ls
ejercicios  otros  tables
```

Por otro lado creamos los directorios en hive_scripts para los scripts de tablas y ejercicios

Creamos una tabla pcaps para SerDe y adjunto a esta práctica el ficheros "create_table_serde.hql" correspondiente al que he llamado.

```
[uambd20@master hive_scripts]$ cd tables/
[uambd20@master tables]$ ls
[uambd20@master tables]$ nano create_table_serde.hql
[uambd20@master tables]$ hive -f create_table_serde.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/j
ars/hive-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/10e463e3-3a52-496f-a63d-9f19d9b5810a_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-depe
ndencies.jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.575 seconds
OK
Time taken: 0.111 seconds
OK
Time taken: 0.077 seconds
OK
Time taken: 0.139 seconds
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master tables]$
```

Creamos un script para visualizar dirección IP origen(src) y número de bytes con protocol="UDP", también adjunto a esta práctica el ficheros "select_udp.hql" correspondiente al que he llamado.

```
[uambd20@master otros]$ cd ..
[uambd20@master hive_scripts]$ cd tables/
[uambd20@master tables]$ cd ..
[uambd20@master hive_scripts]$ cd otros/
[uambd20@master otros]$ nano select_udp.hql
[uambd20@master otros]$ hive -f select_udp.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/hive-commo
n-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/ff31a7f3-3aaa-47cd-9b7d-78a714e24db2_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.499 seconds
Query ID = uambd20_20240131135454_21d74027-1828-4ae4-9de2-3d54ecc4241b
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1706529435081_0084, Tracking URL = http://master:8088/proxy/application_1706529435081_0084/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1706529435081_0
084
```

```
Ended Job = job_1706529435081_0085
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 16.67 sec   HDFS Read: 306827971 HDFS Write: 569 SUCCESS
Stage-Stage-2: Map: 3  Reduce: 1   Cumulative CPU: 5.6 sec   HDFS Read: 9567 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 270 msec
OK
91.189.91.157   96
91.189.89.198   96
fe80:0:0:0:250:56ff:fec0:8      354
192.168.158.254 1200
0.0.0.0 1200
192.168.158.1   11860
192.168.158.128 260352
192.168.158.2   970018
Time taken: 42.76 seconds, Fetched: 8 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master otros]$
```

Creamos una tabla http y adjunto a esta práctica el ficheros "create_http_table.hql" correspondiente al que he llamado.


```
  GNU nano 2.3.1                    File: create_http_table.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.HttpPcapReader;
CREATE DATABASE IF NOT EXISTS uambd20;
USE uambd20;
DROP TABLE http;
CREATE EXTERNAL TABLE http(ts bigint,protocol string,src string,src_port string,dst string, dst_port string,le$
ROW FORMAT SERDE 'net.ripe.hadoop.pcap.serde.PcapDeserializer'
STORED AS INPUTFORMAT 'net.ripe.hadoop.pcap.mr1.io.PcapInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 'hdfs:/user/uambd20/pcaps/';
```



```
                            uambd20@master:~/hive_scripts/tables                    _  □  ✕

File  Edit  View  Search  Terminal  Help
[uambd20@master tables]$ nano create_http_table.hql
[uambd20@master tables]$ hive -f create_http_table.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/hive
-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/8e842628-e2d9-4a57-a0db-d2066d771455_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies
.jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.564 seconds
OK
Time taken: 0.028 seconds
OK
Time taken: 0.008 seconds
OK
Time taken: 0.067 seconds
OK
Time taken: 0.143 seconds
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master tables]$
```

Creamos un script para visualizar líneas de respuesta HTTP con "UserAgent" cambiando el reader a la clase HTTP y adjunto a esta práctica el ficheros "http_request.hql" correspondiente al que he llamado.


```
  GNU nano 2.3.1                    File: http_request.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.HttpPcapReader;
SELECT http_request,`header_user-agent` FROM http where http_request is not NULL;
```

```
uambd20@master:~/hive_scripts/otros

File  Edit  View  Search  Terminal  Help

GET /css/rtve.2015/rtve.sections/rtve.rtve.css HTTP/1.1 Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Gecko/2010
0101 Firefox/55.0
GET /pages/pfPortada/1.8.3/js/userFavor/UserFavor.js HTTP/1.1   Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Ge
cko/20100101 Firefox/55.0
GET /vpaid/vPlayer/player/v8.2.235/assets/video-js/video.js HTTP/1.1    Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:
55.0) Gecko/20100101 Firefox/55.0
GET /vpaid/units/17_2_1/assets/css/cmOsUnit.css HTTP/1.1        Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Ge
cko/20100101 Firefox/55.0
GET /css?family=Roboto+Slab:300,400,500,100,100italic,300,300italic,400italic,500italic,700,700italic,900,900italic
 HTTP/1.1       Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
POST /gsalphasha2g2 HTTP/1.1    Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
GET /piwik/piwik.php?action_name=Nivel%201&idsite=1&rec=1&r=944933&h=3&m=35&s=58&url=http%3A%2F%2Fwww.madrid.org%2F
cs%2FSatellite%3Fpagename%3DComunidadMadrid%252FHome&urlref=http%3A%2F%2Fwww.madrid.org%2Fcs%2FSatellite%3Fcid%3D12
73078206797%26language%3Des%26pagename%3DComunidadMadrid%252FEstructura&_id=2ef494f588f51635&_idts=1511782494&_idvc
=1&_idn=0&_refts=0&_viewts=1511782494&send_image=0&cookie=1&res=1916x958&gt_ms=26 HTTP/1.1     Mozilla/5.0 (X11; Ub
untu; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
GET /vpaid/units/14_12_0/creatives/creative_js.js HTTP/1.1        Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Ge
cko/20100101 Firefox/55.0
GET /vpaid/vPlayer/player/v8.2.235/OvaMediaPlayer.js HTTP/1.1   Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Ge
cko/20100101 Firefox/55.0
Time taken: 23.833 seconds, Fetched: 2414 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master otros]$
```

Creamos una tabla dns y adjunto a esta práctica el ficheros "create_table_dns.hql" correspondiente al que he llamado.


```
  GNU nano 2.3.1                    File: create_table_dns.hql                                            Modified

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.DnsPcapReader;
CREATE DATABASE IF NOT EXISTS uambd20;
USE uambd20;
DROP TABLE dns;
CREATE EXTERNAL TABLE dns(ts bigint, src string,dst string,dns_queryid int,dns_qr boolean,dns_opcode string,dns_rcode string, dns_qname string,$
ROW FORMAT SERDE 'net.ripe.hadoop.pcap.serde.PcapDeserializer'
STORED AS INPUTFORMAT 'net.ripe.hadoop.pcap.mr1.io.PcapInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 'hdfs:/user/uambd20/pcaps/';
```

EJERCICIOS:

Ejercicio1 (0,75 puntos): Creamos un script HIVE que obtiene la serie temporal en bits/s con granularidad 1 segundo. El script selecciona los tiempos y los bits, que se han convertido a bits multiplicando los bytes por 8.

```
                          uambd20@master:~/hive_scripts/ejercicios                    _  □  ⊗

 File  Edit  View  Search  Terminal  Help
   GNU nano 2.3.1                    File: ejercicio1.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SELECT ts,sum(len)*8 as bits FROM pcaps group by ts order by ts asc;




^G Get Help     ^O WriteOut     ^R Read File     ^Y Prev Page     ^K Cut Text      ^C Cur Pos
^X Exit         ^J Justify      ^W Where Is      ^V Next Page     ^U UnCut Text    ^T To Spell
```

```
1511782531     577744
1511782532     0
1511782533     20624
1511782534     496
1511782535     181728
1511782536     1472
1511782537     0
1511782538     2538880
1511782539     544512
1511782540     2247504
1511782541     531744
1511782542     396912
1511782543     569504
1511782544     272336
1511782545     40352
1511782546     47440
1511782547     5632
1511782548     139392
1511782549     2135392
1511782550     189280
1511782551     1570176
1511782552     1849536
1511782553     4371680
1511782554     1374656
1511782555     3127440
1511782556     156480
1511782557     65312
1511782558     26352
1511782559     57632
1511782560     17376
1511782561     992
1511782562     2944
1511782563     22224
1511782564     36880
1511782565     992
1511782566     0
1511782567     9824
1511782568     9040
1511782569     12672
1511782570     12944
1511782571     4497168
1511782572     18320
1511782573     2944
Time taken: 42.174 seconds, Fetched: 894 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master ejercicios]$
```

Ejercicio2 (0,75 puntos): Creamos un script HIVE que obtiene registros de flujos junto con el número de bytes y paquetes. Los registros de flujos incluyen información sobre tiempo mínimo, tiempo máximo, dirección IP origen (src) y dirección IP destino (dst), puerto origen (src_port) y puerto destino (dst_port), protocolo (TCP o UDP), número de bytes (sum(len)), número de paquetes.

```
  GNU nano 2.3.1                     File: ejercicio2.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SELECT min(ts) as first_packet, max(ts) as last_packet, src, src_port, dst, dst_port, protocol, sum(le$
FROM pcaps
WHERE src is not NULL and dst is NOT NULL
GROUP BY src, src_port, dst, dst_port, protocol
ORDER BY bytes DESC;
```

```
[uambd20@master ejercicios]$ nano ejercicio2.hql
[uambd20@master ejercicios]$ time hive -f ejercicio2.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/j
ars/hive-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/c216b23f-4fb1-4d8f-92c7-594e2423e336_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-depe
ndencies.jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.578 seconds
Query ID = uambd20_20240131145555_7c6752df-97cb-403c-8cb3-0ca3c7637cb9
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1706529435081_0094, Tracking URL = http://master:8088/proxy/application_170652943508
1_0094/
```

```
1511782000    1511782005    192.168.158.128 46962    130.206.192.10    80      TCP    0    8
1511781874    1511781880    192.168.158.128 52670    216.58.210.129    80      TCP    0    8
1511781711    1511781711    93.184.220.70    443     192.168.158.128 45768    TCP    0    6
1511782220    1511782225    192.168.158.128 45858    192.0.76.3        80      TCP    0    8
1511782556    1511782562    192.168.158.128 45066    8.254.94.254      80      TCP    0    8
1511782361    1511782366    192.168.158.128 45812    62.73.184.139     80      TCP    0    8
1511781892    1511781897    192.168.158.128 53674    67.226.210.13     80      TCP    0    8
1511781869    1511781869    192.168.158.128 52424    69.172.216.111    80      TCP    0    4
1511782355    1511782360    192.168.158.128 45758    62.73.184.139     80      TCP    0    8
1511782556    1511782556    192.168.158.128 45722    69.172.216.111    443     TCP    0    4
1511782516    1511782525    93.184.221.133   80      192.168.158.128 46894    TCP    0    6
1511781896    1511781896    192.168.158.128 45622    62.67.193.31      80      TCP    0    4
1511781892    1511781897    192.168.158.128 52192    2.18.115.117      80      TCP    0    8
1511781893    1511781898    192.168.158.128 47070    52.51.73.34       80      TCP    0    8
1511781896    1511781896    192.168.158.128 45624    62.67.193.31      80      TCP    0    4
1511781879    1511781885    192.168.158.128 52734    34.204.227.165    80      TCP    0    8
1511781996    1511781996    104.121.156.80   80      192.168.158.128 47160    TCP    0    2
1511781699    1511781704    192.168.158.128 52712    150.244.214.206 80      TCP    0    8
1511782183    1511782188    192.168.158.128 45462    216.58.214.174    80      TCP    0    8
1511782427    1511782432    4.23.62.126      80      192.168.158.128 47348    TCP    0    6
1511781890    1511781896    192.168.158.128 51334    62.67.193.21      80      TCP    0    8
1511781701    1511781706    150.244.214.21   80      192.168.158.128 42738    TCP    0    6
1511782303    1511782308    192.168.158.128 51242    216.58.201.130    80      TCP    0    8
1511782171    1511782176    192.168.158.128 52586    54.231.114.187    80      TCP    0    8
1511781879    1511781884    46.228.164.11    80      192.168.158.128 53942    TCP    0    6
1511782549    1511782554    216.58.211.194   80      192.168.158.128 53536    TCP    0    6
1511781855    1511781855    192.168.158.128 45154    52.52.228.19      443     TCP    0    2
1511781854    1511781854    192.168.158.128 45124    52.52.228.19      443     TCP    0    2
1511782419    1511782482    192.168.158.128 51148    217.148.69.24     443     TCP    0    14
Time taken: 41.161 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.587s
user    0m21.728s
sys     0m1.936s
[uambd20@master ejercicios]$
```

Ejercicio3 (0,75 puntos): Creamos un script HIVE que obtiene los (`header_host`) 'Host' (a nivel HTTP) más populares (en número de peticiones).

```
  GNU nano 2.3.1                          File: ejercicio3.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.HttpPcapReader;
SELECT `header_host`, count(*) as npet FROM http where http_request is not NULL group by `header_host` sort by npet asc;
```

```
[uambd20@master hive_scripts]$ cd ejercicios/
[uambd20@master ejercicios]$ ls
ejercicio1.hql  ejercicio2.hql
[uambd20@master ejercicios]$ nano ejercicio3.hql
[uambd20@master ejercicios]$ nano ejercicio3.hql
[uambd20@master ejercicios]$ hive -f ejercicio3.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.
1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/86cd9117-6ace-474c-9419-40029da8b1a4_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.542 seconds
```

```
adserver-us.adtech.advertising.com        12
bilder.bild.de   12
www.ucm.es       12
dle.rae.es       12
jadserve.postrelease.com          12
img.irtve.es     12
adserver.adtechus.com     12
gn.symcd.com     12
hackersonlineclub.com     12
aax-eu.amazon-adsystem.com        14
r.nexac.com      14
cdn.cnn.com      14
www.uam.es       14
active.cache.el-mundo.net         14
tags.bluekai.com          14
dt.adsafeprotected.com    14
dx.steelhousemedia.com    14
www.latimes.com 14
as.casalemedia.com        14
www.eluniversohoy.net     16
clients1.google.com       16
pixel.adsafeprotected.com         16
www.hpcn-uam.es 16
www.trbimg.com   16
cdn02.nintendo-europe.com         16
edition.i.cdn.cnn.com     18
tags.tiqcdn.com 18
ministeriodeagricultura.solution.weborama.fr     18
e00-elmundo.uecdn.es      18
ocsp.usertrust.com        20
aa.agkn.com      20
e00-marca.uecdn.es        20
ocsp.pki.goog    20
aax.amazon-adsystem.com 24
www.irtve.es     24
img.rtve.es      24
tribune-d.openx.net       28
ams1-ib.adnxs.com         30
beacon.krxd.net 30
secure.adnxs.com          46
ocsp.godaddy.com          48
ib.adnxs.com     56
ocsp.comodoca.com         72
fastlane.rubiconproject.com       100
Time taken: 49.822 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master ejercicios]$
```

Ejercicio4 (0,75 puntos): Creamos un script HIVE que obtiene la cantidad de 'UserAgent' (`header_user-agent`) distintos junto con el número de peticiones observadas.

```
  GNU nano 2.3.1                          File: ejercicio4.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.HttpPcapReader;
SELECT `header_user-agent`, count(*) as npet FROM http where http_request is not NULL group by `header_user-agent` sort by npet asc;
```

```
[uambd20@master ejercicios]$ ls
ejercicio1.hql  ejercicio2.hql  ejercicio3.hql
[uambd20@master ejercicios]$ nano ejercicio4.hql
[uambd20@master ejercicios]$ hive -f ejercicio4.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/hive-common-1.
1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/98492c04-d7af-4ce1-a0b3-fd96d7c35d5e_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar] to c
lass path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.581 seconds
Query ID = uambd20_20240131162626_62570f37-ffbb-496b-abb9-5e2ede5447fe
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
```

```
                        uambd20@master:~/hive_scripts/ejercicios                    _  □  ✕

 File  Edit  View  Search  Terminal  Help
2024-01-31 16:27:18,797 Stage-2 map = 0%,  reduce = 0%
2024-01-31 16:27:24,034 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.36 sec
2024-01-31 16:27:30,317 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.14 sec
MapReduce Total cumulative CPU time: 3 seconds 140 msec
Ended Job = job_1706529435081_0114
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 36.03 sec   HDFS Read: 306828779 HDFS Write: 385 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.14 sec   HDFS Read: 5743 HDFS Write: 82 SUCCESS
Total MapReduce CPU Time Spent: 39 seconds 170 msec
OK
Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0    2414
Time taken: 50.73 seconds, Fetched: 1 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master ejercicios]$
```

Solo se observa Mozilla/5.0 con un total de 2414 peticiones.

Ejercicio 5 (0,75 puntos): Creamos un script HIVE que obtiene las diferentes líneas de respuesta HTTP (http_response) junto con el número de veces que se han observado.

```
  GNU nano 2.3.1                        File: ejercicio5.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.HttpPcapReader;
SELECT http_response, count(*) as nresp FROM http where http_response is not NULL group by http_response sort by nresp asc;
```

```
[uambd20@master ejercicios]$ nano ejercicio5.hql
[uambd20@master ejercicios]$ hive -f ejercicio5.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/hive-
common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/d92cca07-17e7-4504-9e5f-e668b7eeb051_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.
jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.56 seconds
Query ID = uambd20_20240131163333_38001660-0722-4f74-bc7d-3cd98a1b9818
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job 1706529435081 0115, Tracking URL = http://master:8088/proxy/application 1706529435081 0115/
```

```
                          uambd20@master:~/hive_scripts/ejercicios              _  □  ✕

 File  Edit  View  Search  Terminal  Help
Ended Job = job_1706529435081_0116
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 3   Cumulative CPU: 36.09 sec   HDFS Read: 306826605 HDFS Write: 1133 SUCCESS
Stage-Stage-2: Map: 3  Reduce: 1   Cumulative CPU: 5.52 sec    HDFS Read: 10143 HDFS Write: 532 SUCCESS
Total MapReduce CPU Time Spent: 41 seconds 610 msec
OK
HTTP/1.1 400 Bad Request         2
HTTP/1.1 307 Temporary Redirect  2
HTTP/1.1 204 No Response          2
HTTP/1.1 302 nginx/1.11.3         2
HTTP/1.1 206 Partial Content     2
HTTP/1.1 204     2
HTTP/1.1 101 Switching Protocols        2
HTTP/1.0 301 Moved Permanently   2
HTTP/1.0 200 OK 2
HTTP/1.1 303 See Other   4
HTTP/1.1 404 Not Found   4
HTTP/1.1 302 Moved Permanently  8
HTTP/1.0 400 Bad request        8
HTTP/1.1 302     12
HTTP/1.1 200     14
HTTP/1.1 301 Moved Permanently  32
HTTP/1.1 204 No Content 60
HTTP/1.1 302 Moved Temporarily  96
HTTP/1.1 302 Found       250
HTTP/1.1 200 OK 1792
Time taken: 50.452 seconds, Fetched: 20 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master ejercicios]$
```

Ejercicio 6 (0,75 puntos): Creamos un script HIVE que obtiene la popularidad de los nombre de dominio (dns_qname) por número de peticiones.

```
 GNU nano 2.3.1                    File: ejercicio6.hql

ADD JAR hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar;
USE uambd20;
SET net.ripe.hadoop.pcap.io.reader.class=net.ripe.hadoop.pcap.DnsPcapReader;
SELECT dns_qname, count(*) as requests
FROM dns
WHERE dns_qr=false
GROUP BY dns_qname
ORDER BY requests DESC;
```

```
[uambd20@master hive_scripts]$ cd ejercicios/
[uambd20@master ejercicios]$ ls
ejercicio1.hql  ejercicio2.hql  ejercicio3.hql  ejercicio4.hql  ejercicio5.hql
[uambd20@master ejercicios]$ nano ejercicio6.hql
[uambd20@master ejercicios]$ hive -f ejercicio6.hql

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/h
ive-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
converting to local hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar
Added [/tmp/5923b8d9-429a-4c55-91a7-1832e5e0d1ac_resources/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependenc
ies.jar] to class path
Added resources: [hdfs:///user/uambd20/libs/hadoop-pcap-serde-1.2-SNAPSHOT-jar-with-dependencies.jar]
OK
Time taken: 1.609 seconds
Query ID = uambd20_20240131182020_634c9589-faba-4e4b-8f3d-006612482b0e
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 3
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1706529435081_0117, Tracking URL = http://master:8088/proxy/application_1706529435081_011
7/
```

```
Total MapReduce CPU Time Spent: 46 seconds 560 msec
OK
bidder.criteo.com.        56
pix04.revsci.net.         48
www.facebook.com.         48
www.youtube.com.          46
www.google.com. 40
ib.adnxs.com.    38
securepubads.g.doubleclick.net. 38
connect.facebook.net.    36
versioncheck-bg.addons.mozilla.org.       36
api.permutive.com.        36
www.google-analytics.com.         36
adservice.google.com.    34
pagead2.googlesyndication.com.   34
syndication.twitter.com.         34
d2tjil5tz17cbz.cloudfront.net.   32
d2tjil5tz17cbz.cloudfront.net.localdomain.       32
ad.doubleclick.net.       32
www.googletagservices.com.        32
rb.adnxs.com.    32
b.scorecardresearch.com.          32
platform.twitter.com.    30
as.casalemedia.com.       28
fonts.googleapis.com.    28
tpc.googlesyndication.com.        28
www.google.es.  28
ajax.googleapis.com.      26
staticxx.facebook.com.   24
ping.chartbeat.net.       24
pix04.revsci.net.localdomain.    24
cm.g.doubleclick.net.    24
secure.adnxs.com.         24
googleads.g.doubleclick.net.     24
beacon.krxd.net.          24
bs.serving-sys.com.       24
acdn.adnxs.com. 24
pbs.twimg.com.  24
fonts.gstatic.com.        22
```

```
                    uambd20@master:~/hive_scripts/ejercicios

 File   Edit   View   Search   Terminal   Help
asset.flite.com.            4
apps.uam.es.       4
api.pinterest.com.          4
api.elconfidencial.com.  4
api.bounceexchange.com.  4
analytics.brandcrumb.com.        4
an.pxl.ace.advertising.com.       4
amplify.outbrain.com.      4
amazon.es.          4
alumni.uam.es.    4
ak1.abmr.net.       4
ak.sail-horizon.com.        4
ak-ads-ns.prisasd.com.  4
advertising.amazon.com.  4
adserver.adtechus.com.   4
adserver-us.adtech.advertising.com.       4
ads.undertone.com.          4
ad.yieldlab.net.            4
a.visualrevenue.com.        4
a.rfihub.com.       4
a.amz.mshcdn.com.           4
a-ssl.ligatus.com.          4
71i.nuggad.net.  4
5808301.fls.doubleclick.net.     4
58011369.log.optimizely.com.     4
5365628.fls.doubleclick.net.     4
4711.lineadirecta.com.  4
4.bp.blogspot.com.          4
33across-d.openx.net.    4
1to1.bbva.com.   4
as.es.   2
quit.   2
local.   2
quit.localdomain.        2
ww.google.es.    2
as.com. 2
google.es.         2
Time taken: 52.571 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[uambd20@master ejercicios]$
```

Ejercicio 7 (1 punto): Obtenemos los valores por defecto de los siguientes parámetros (desde HIVE) usando los comandos:

set mapreduce.map.memory.mb;
(La memoria (en MB) dedicada a las tareas del mapper es 0)

set mapreduce.map.java.opts;
(Las opciones de la JVM que se usarán para las tareas del mapper es TRUE)

set mapreduce.reduce.memory.mb;
(La memoria (en MB) dedicada a las tareas del reducer es 0)

set mapreduce.reduce.java.opts;
(Las opciones de la JVM que se usarán para las tareas del mapper es TRUE)

set mapreduce.task.io.sort.mb;
(La cantidad de memoria (en MB) usada para realizar ordenación de ficheros es 256)

set yarn.scheduler.minimum-allocation-mb;
(El valor mínimo de memoria (en MB) usado para un container es 1024)

set yarn.scheduler.maximum-allocation-mb;
(El valor máximo de memoria (en MB) usado para un container es 65536)

set yarn.scheduler.minimum-allocation-vcores;
(El número mínimo de vCores que puede usar un container es 1)

set yarn.scheduler.maximum-allocation-vcores;
(El número máximo de vCores que puede usar un container es 32)

```
merve@onur-ideacenter:~/Desktop/Infraestructura_para_big_data/3Evaluación_de_pre
staciones$ ssh uambd20@150.244.65.34
uambd20@150.244.65.34's password:
Last login: Wed Jan 31 18:02:28 2024 from 172.30.130.89
[uambd20@master ~]$ hive

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.
1-1.cdh5.15.1.p0.4/jars/hive-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> set mapreduce.map.memory.mb;
mapreduce.map.memory.mb=0
hive> set mapreduce.map.java.opts;
mapreduce.map.java.opts=-Djava.net.preferIPv4Stack=true
hive> set mapreduce.reduce.memory.mb;
mapreduce.reduce.memory.mb=0
hive> set mapreduce.reduce.java.opts;
mapreduce.reduce.java.opts=-Djava.net.preferIPv4Stack=true
hive> set mapreduce.task.io.sort.mb;
mapreduce.task.io.sort.mb=256
hive> set yarn.scheduler.minimum-allocation-mb;
yarn.scheduler.minimum-allocation-mb=1024
hive> set yarn.scheduler.maximum-allocation-mb;
yarn.scheduler.maximum-allocation-mb=65536
hive> set yarn.scheduler.minimum-allocation-vcores;
yarn.scheduler.minimum-allocation-vcores=1
hive> set yarn.scheduler.maximum-allocation-vcores;
yarn.scheduler.maximum-allocation-vcores=32
hive>
```

Ejercicio 8 (1,5 puntos): Utilizando el script desarrollado en ejercicio 2, observar la influencia del parámetro "mapreduce.map.cpu.vcores". Variar su valor entre 1 y 3. Realizar 4 ejecuciones con cada valor y obtener la media y al desviación estándar. Representar los datos de forma gráfica o mediante una tabla.

```
hive> set mapreduce.map.cpu.vcores;
mapreduce.map.cpu.vcores=1
```

El valor por defecto del parámetro mapreduce.map.cpu.vcores es 1.

Realizamos 4 ejecuciones del script de Hive cuando el parámetro mapreduce.map.cpu.vcores está configurado en 1.

```
[uambd20@master ejercicios]$ time hive -f ejercicio2.hql --hiveconf mapreduce.map.cpu.vcores=1
```

```
1511782355      1511782360      192.168.158.128 55190   172.227.120.189 80      TCP     0       8
1511781896      1511781896      192.168.158.128 44540   185.33.223.83   80      TCP     0       4
Time taken: 42.247 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m51.606s
user    0m21.511s
sys     0m1.947s
[uambd20@master ejercicios]$
```

```
1511781854      1511781854      192.168.158.128 45124   52.52.228.19    443     TCP     0       2
1511782419      1511782482      192.168.158.128 51148   217.148.69.24   443     TCP     0       14
Time taken: 42.0 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.935s
user    0m21.005s
sys     0m1.931s
[uambd20@master ejercicios]$
```

```
1511781902      1511781907      192.168.158.128 37794   54.245.233.125  80      TCP     0       8
1511782127      1511782132      192.168.158.128 42674   216.58.210.131  80      TCP     0       8
Time taken: 40.883 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m49.743s
user    0m20.324s
sys     0m1.762s
[uambd20@master ejercicios]$
```

```
1511781902      1511781907      192.168.158.128 37794   54.245.233.125  80      TCP     0       8
1511782127      1511782132      192.168.158.128 42674   216.58.210.131  80      TCP     0       8
Time taken: 44.439 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m53.686s
user    0m21.403s
sys     0m2.110s
```

Realizamos 4 ejecuciones del script de Hive cuando el parámetro mapreduce.map.cpu.vcores está configurado en 2.

```
[uambd20@master ejercicios]$ time hive -f ejercicio2.hql --hiveconf mapreduce.map.cpu.vcores=2
```

```
1511781902      1511781907      192.168.158.128 37794   54.245.233.125  80      TCP     0       8
1511782127      1511782132      192.168.158.128 42674   216.58.210.131  80      TCP     0       8
Time taken: 42.022 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m51.591s
user    0m22.069s
sys     0m1.959s
```

```
1511781902      1511781907      192.168.158.128 37794   54.245.233.125  80      TCP     0       8
1511782127      1511782132      192.168.158.128 42674   216.58.210.131  80      TCP     0       8
Time taken: 41.35 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.770s
user    0m21.998s
sys     0m1.846s
```

```
1511781902      1511781907      192.168.158.128 37794   54.245.233.125  80      TCP     0       8
1511782127      1511782132      192.168.158.128 42674   216.58.210.131  80      TCP     0       8
Time taken: 41.042 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.199s
user    0m21.202s
sys     0m1.929s
```

```
1511782355      1511782360      192.168.158.128 55190   172.227.120.189 80      TCP     0       8
1511781896      1511781896      192.168.158.128 44540   185.33.223.83   80      TCP     0       4
Time taken: 41.369 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.659s
user    0m22.060s
sys     0m1.952s
```

Realizamos 4 ejecuciones del script de Hive cuando el parámetro mapreduce.map.cpu.vcores está configurado en 3.

```
[uambd20@master ejercicios]$ time hive -f ejercicio2.hql --hiveconf mapreduce.map.cpu.vcores=3
```

```
1511781902        1511781907        192.168.158.128 37794    54.245.233.125  80        TCP      0        8
1511782127        1511782132        192.168.158.128 42674    216.58.210.131  80        TCP      0        8
Time taken: 41.613 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.795s
user    0m21.094s
sys     0m2.010s
```

```
1511781902        1511781907        192.168.158.128 37794    54.245.233.125  80        TCP      0        8
1511782127        1511782132        192.168.158.128 42674    216.58.210.131  80        TCP      0        8
Time taken: 41.212 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.515s
user    0m22.039s
sys     0m1.872s
```

```
1511781902        1511781907        192.168.158.128 37794    54.245.233.125  80        TCP      0        8
1511782127        1511782132        192.168.158.128 42674    216.58.210.131  80        TCP      0        8
Time taken: 41.346 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.291s
user    0m20.934s
sys     0m1.871s
```

```
1511781902        1511781907        192.168.158.128 37794    54.245.233.125  80        TCP      0        8
1511782127        1511782132        192.168.158.128 42674    216.58.210.131  80        TCP      0        8
Time taken: 40.935 seconds, Fetched: 8681 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m50.221s
user    0m21.329s
sys     0m1.853s
```

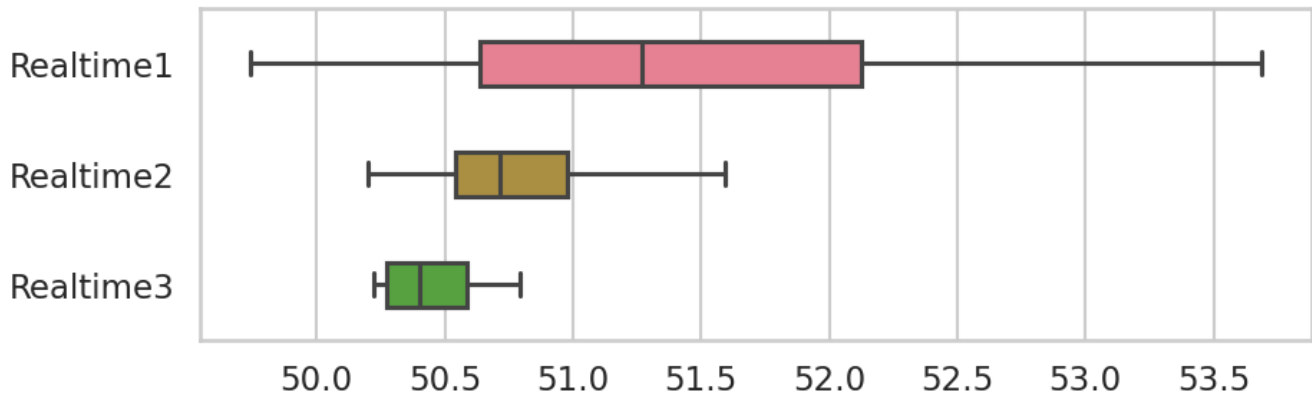| Question 8	Mean of Execution Time	Standard Deviation of Execution Time
Real Time 1 mapreduce.map.cpu.vcore=1	51.492	1.431
Real Time 2 mapreduce.map.cpu.vcore=2	50.805	0.502
Real Time 3 mapreduce.map.cpu.vcore=3	50.456	0.224

Figure 1: Resultados de 4 ejecuciones con valores diferentes para un parámetro específico.

Realtime1,2 y 3 indica que mapreduce.map.cpu.vcore=1,2 y 3

Observamos en la gráfica que el valor óptimo para el parámetro mapreduce.map.cpu.vcores es 3. Porque se reduce el tiempo de ejecución medio y la desviación estándar del tiempo de ejecución.

En BoxPlot:
Línea dentro de la caja: Esta línea representa la mediana (Q2) del conjunto de datos. Da una idea de la tendencia central de los datos.

Bigotes: Los bigotes se extienden desde los bordes de la caja hasta los valores mínimo y máximo dentro de un rango determinado. El rango puede basarse en una distancia especificada desde el IQR o puede extenderse a los valores mínimo y máximo dentro de los datos que no se consideran valores atípicos.

Ejercicio 9 (1,5 puntos): Utilizando el script desarrollado en ejercicio 3, observar la influencia del parámetro "mapreduce.map.memory.mb". Probar con los siguientes valores [2048,4096,8192] . Realizar 4 ejecuciones con cada valor y obtener la media y al desviación estándar. Representar los datos de forma gráfica o mediante una tabla. Al realizar las ejecuciones se deben tener en cuenta la influencia de otros parámetros del clúster como por ejemplo:
"mapreduce.map.java.opts"
"mapreduce.task.io.sort.mb"

Tomando en consideración las siguientes condiciones:
IO Sort Mem << Mapper Maximum JVM Heap
Mapper Maximum JVM Heap - IO Sort Mem > 768MB

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb=2048 y los valores de mapreduce.map.java.opts, mapreduce.task.io.sort.mb(=256) son los valores por defecto.

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=2048
ib.adnxs.com      56
ocsp.comodoca.com         72
fastlane.rubiconproject.com      100
Time taken: 49.445 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m57.631s
user    0m20.862s
sys     0m1.747s
```

```
ib.adnxs.com      56
ocsp.comodoca.com         72
fastlane.rubiconproject.com      100
Time taken: 48.446 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m56.782s
user    0m20.490s
sys     0m1.586s
```

```
ib.adnxs.com      56
ocsp.comodoca.com         72
fastlane.rubiconproject.com      100
Time taken: 51.447 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m59.814s
user    0m20.284s
sys     0m1.738s
```

```
ib.adnxs.com      56
ocsp.comodoca.com         72
fastlane.rubiconproject.com      100
Time taken: 53.711 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    1m1.975s
user    0m21.585s
sys     0m1.603s
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb=2048 y mapreduce.map.java.opts se establece en -Xmx1000m (Establece el tamaño máximo del montón en 1 GB) y mapreduce.task.io.sort.mb se establece en 100 (Asigna 100 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=2048
--hiveconf mapreduce.map.java.opts=-Xmx1000m --hiveconf mapreduce.task.io.sort.mb=100
```

```
beacon.krxd.net 30
ams1-ib.adnxs.com       30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 49.422 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#r
elease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanat
ion.

real    0m57.815s
user    0m21.169s
sys     0m1.499s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 53.351 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#r
elease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanat
ion.

real    1m1.798s
user    0m21.143s
sys     0m1.710s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 49.859 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#r
elease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanat
ion.

real    0m58.385s
user    0m21.434s
sys     0m1.700s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 51.733 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#r
elease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanat
ion.

real    0m59.820s
user    0m20.634s
sys     0m1.718s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb=2048 y mapreduce.map.java.opts se establece en -Xmx2000m (Establece el tamaño máximo del montón en 2 GB) y mapreduce.task.io.sort.mb se establece en 300 (Asigna 300 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=2048
 --hiveconf mapreduce.map.java.opts=-Xmx2000m --hiveconf mapreduce.task.io.sort.mb=300
```

```
beacon.krxd.net 30
ams1-ib.adnxs.com       30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 50.092 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rel
ease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanatio
n.

real    0m57.996s
user    0m20.771s
sys     0m1.530s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 53.119 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rel
ease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanatio
n.

real    1m1.413s
user    0m20.846s
sys     0m1.642s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 50.561 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rel
ease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanatio
n.

real    0m59.026s
user    0m21.881s
sys     0m1.743s
[uambd20@master ejercicios]$
```

```
beacon.krxd.net 30
ams1-ib.adnxs.com       30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 47.963 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rel
ease() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanatio
n.

real    0m56.469s
user    0m21.663s
sys     0m1.576s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb=4096 y los valores de mapreduce.map.java.opts, mapreduce.task.io.sort.mb(=256) son los valores por defecto.

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=4096
```

```
tribune-d.openx.net     28
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 50.718 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#relea
) was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m58.851s
user    0m21.012s
sys     0m1.671s
[uambd20@master ejercicios]$
```

```
tribune-d.openx.net     28
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 51.894 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release(
) was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    1m0.426s
user    0m21.996s
sys     0m1.724s
[uambd20@master ejercicios]$
```

```
tribune-d.openx.net     28
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 49.435 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release(
) was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m57.813s
user    0m20.568s
sys     0m1.877s
[uambd20@master ejercicios]$
```

```
tribune-d.openx.net     28
beacon.krxd.net 30
ams1-ib.adnxs.com       30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 48.257 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release(
) was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m56.757s
user    0m20.950s
sys     0m1.747s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb= 4096 y mapreduce.map.java.opts se establece en -Xmx1000m (Establece el tamaño máximo del montón en 1 GB) y mapreduce.task.io.sort.mb se establece en 100 (Asigna 100 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=4096
--hiveconf mapreduce.map.java.opts=-Xmx1000m --hiveconf mapreduce.task.io.sort.mb=100
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 46.043 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m54.332s
user    0m20.865s
sys     0m1.764s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 53.075 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    1m1.658s
user    0m21.537s
sys     0m1.597s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 54.515 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    1m2.798s
user    0m20.216s
sys     0m1.791s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 51.249 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m59.778s
user    0m21.974s
sys     0m1.819s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb= 4096 y mapreduce.map.java.opts se establece en -Xmx2000m (Establece el tamaño máximo del montón en 2 GB) y mapreduce.task.io.sort.mb se establece en 300 (Asigna 300 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=4096
 --hiveconf mapreduce.map.java.opts=-Xmx2000m --hiveconf mapreduce.task.io.sort.mb=300
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com     56
ocsp.comodoca.com        72
fastlane.rubiconproject.com      100
Time taken: 52.027 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactor
y#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expla
nation.

real    1m0.415s
user    0m21.080s
sys     0m1.770s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com     56
ocsp.comodoca.com        72
fastlane.rubiconproject.com      100
Time taken: 57.793 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactor
y#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expla
nation.

real    1m6.530s
user    0m20.926s
sys     0m1.691s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com     56
ocsp.comodoca.com        72
fastlane.rubiconproject.com      100
Time taken: 50.589 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactor
y#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expla
nation.

real    0m58.813s
user    0m20.696s
sys     0m1.603s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com     56
ocsp.comodoca.com        72
fastlane.rubiconproject.com      100
Time taken: 55.639 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactor
y#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expla
nation.

real    1m4.208s
user    0m22.436s
sys     0m1.702s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb=8192 y los valores de mapreduce.map.java.opts, mapreduce.task.io.sort.mb(=256) son los valores por defecto.

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=8192
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 46.137 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rele
ase() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation
.

real    0m54.497s
user    0m20.688s
sys     0m1.726s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 50.501 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rele
ase() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation
.

real    0m58.731s
user    0m20.822s
sys     0m1.822s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 48.197 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rele
ase() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation
.

real    0m56.650s
user    0m21.762s
sys     0m1.599s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 49.555 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#rele
ase() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation
.

real    0m58.009s
user    0m21.471s
sys     0m1.676s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb= 8192 y mapreduce.map.java.opts se establece en -Xmx1000m (Establece el tamaño máximo del montón en 1 GB) y mapreduce.task.io.sort.mb se establece en 100 (Asigna 100 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=8192
 --hiveconf mapreduce.map.java.opts=-Xmx1000m --hiveconf mapreduce.task.io.sort.mb=100
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 45.558 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m54.061s
user    0m20.042s
sys     0m1.623s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 46.7 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m55.396s
user    0m21.062s
sys     0m1.706s
[uambd20@master ejercicios]$
```

```
beacon.krxd.net 30
ams1-ib.adnxs.com       30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 48.433 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m56.854s
user    0m20.797s
sys     0m1.669s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com       30
beacon.krxd.net 30
secure.adnxs.com        46
ocsp.godaddy.com        48
ib.adnxs.com    56
ocsp.comodoca.com       72
fastlane.rubiconproject.com     100
Time taken: 45.132 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#
release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explana
tion.

real    0m53.482s
user    0m19.841s
sys     0m1.755s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones con la configuración cuando mapreduce.map.memory.mb= 8192 y mapreduce.map.java.opts se establece en -Xmx2000m (Establece el tamaño máximo del montón en 2 GB) y mapreduce.task.io.sort.mb se establece en 300 (Asigna 300 MB de memoria para realizar ordenación de ficheros).

```
[uambd20@master ejercicios]$ time hive -f ejercicio3.hql --hiveconf mapreduce.map.memory.mb=8192
 --hiveconf mapreduce.map.java.opts=-Xmx2000m --hiveconf mapreduce.task.io.sort.mb=300
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com    56
ocsp.comodoca.com        72
fastlane.rubiconproject.com     100
Time taken: 46.354 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory
#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explan
ation.

real    0m54.779s
user    0m20.321s
sys     0m1.749s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com    56
ocsp.comodoca.com        72
fastlane.rubiconproject.com     100
Time taken: 49.466 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory
#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explan
ation.

real    0m58.308s
user    0m21.296s
sys     0m1.648s
[uambd20@master ejercicios]$
```

```
ams1-ib.adnxs.com        30
beacon.krxd.net 30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com    56
ocsp.comodoca.com        72
fastlane.rubiconproject.com     100
Time taken: 46.349 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory
#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explan
ation.

real    0m54.736s
user    0m20.902s
sys     0m1.823s
[uambd20@master ejercicios]$
```

```
beacon.krxd.net 30
ams1-ib.adnxs.com        30
secure.adnxs.com         46
ocsp.godaddy.com         48
ib.adnxs.com    56
ocsp.comodoca.com        72
fastlane.rubiconproject.com     100
Time taken: 49.393 seconds, Fetched: 468 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory
#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explan
ation.

real    0m57.585s
user    0m20.883s
sys     0m1.602s
[uambd20@master ejercicios]$
```

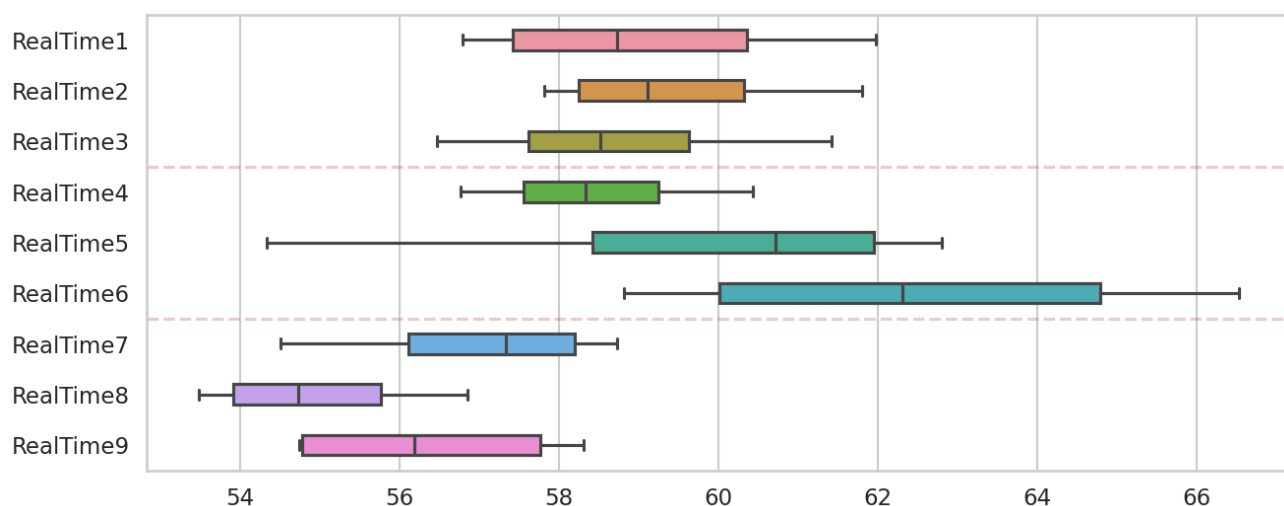| Question 9	Mean	Standard Deviation
Real Time 1 mapreduce.map.memory.mb=2048 mapreduce.map.java.opts=por defecto mapreduce.task.io.sort.mb=por defecto(256)	59.050	2.018
Real Time 2 mapreduce.map.memory.mb=2048 mapreduce.map.java.opts=-Xmx1000m mapreduce.task.io.sort.mb=100	59.454	1.538
Real Time 3 mapreduce.map.memory.mb=2048 mapreduce.map.java.opts=-Xmx2000m mapreduce.task.io.sort.mb=300	58.726	1.798
Real Time 4 mapreduce.map.memory.mb=4096 mapreduce.map.java.opts=por defecto mapreduce.task.io.sort.mb=por defecto(256)	58.462	1.354
Real Time 5 mapreduce.map.memory.mb=4096 mapreduce.map.java.opts=-Xmx1000m mapreduce.task.io.sort.mb=100	59.642	3.250
Real Time 6 mapreduce.map.memory.mb=4096 mapreduce.map.java.opts=-Xmx2000m mapreduce.task.io.sort.mb=300	62.491	3.045
Real Time 7 mapreduce.map.memory.mb=8192 mapreduce.map.java.opts=por defecto mapreduce.task.io.sort.mb=por defecto(256)	56.972	1.612
Real Time 8 mapreduce.map.memory.mb=8192 mapreduce.map.java.opts=-Xmx1000m mapreduce.task.io.sort.mb=100	54.948	1.301
Real Time 9 mapreduce.map.memory.mb=8192 mapreduce.map.java.opts=-Xmx2000m mapreduce.task.io.sort.mb=300	56.352	1.615

Figure 2: Resultados de 4 ejecuciones con valores diferentes para 3 parámetros distintos.

RealTime1 indica que mapreduce.map.memory.mb=2048 mapreduce.map.java.opts=por defecto, mapreduce.task.io.sort.mb=por defecto(256)
RealTime2 indica que mapreduce.map.memory.mb=2048, mapreduce.map.java.opts=-Xmx1000m, mapreduce.task.io.sort.mb=100
RealTime3 indica que mapreduce.map.memory.mb=2048, mapreduce.map.java.opts=-Xmx2000m, mapreduce.task.io.sort.mb=300

RealTime4 indica que mapreduce.map.memory.mb=4096 mapreduce.map.java.opts=por defecto, mapreduce.task.io.sort.mb=por defecto(256)
RealTime5 indica que mapreduce.map.memory.mb=4096, mapreduce.map.java.opts=-Xmx1000m, mapreduce.task.io.sort.mb=100
RealTime6 indica que mapreduce.map.memory.mb=4096, mapreduce.map.java.opts=-Xmx2000m, mapreduce.task.io.sort.mb=300

RealTime7 indica que mapreduce.map.memory.mb=8192 mapreduce.map.java.opts=por defecto, mapreduce.task.io.sort.mb=por defecto(256)
RealTime8 indica que mapreduce.map.memory.mb=8192, mapreduce.map.java.opts=-Xmx1000m, mapreduce.task.io.sort.mb=100
RealTime9 indica que mapreduce.map.memory.mb=8192, mapreduce.map.java.opts=-Xmx2000m, mapreduce.task.io.sort.mb=300

Observamos en la gráfica que para reducir tanto el tiempo de ejecución medio como la desviación estándar del tiempo de ejecución, los valores óptimos de los parámetros son:

mapreduce.map.memory.mb=8192
mapreduce.map.java.opts=-Xmx1000m
mapreduce.task.io.sort.mb=100.

Ejercicio 10 (1,5 puntos): Utilizando el script desarrollado en ejercicio 6, observar la influencia de la aplicación de la vectorización de HIVE (activar todos los parámetros de vectorización) . Realizar 4 ejecuciones con cada una de las dos opciones y obtener la media y al desviación estándar. Representar los datos de forma gráfica o mediante una tabla.

Los parámetros de vectorización de HIVE:

```
hive> set hive.vectorized.execution.reduce.groupby.enabled;
hive.vectorized.execution.reduce.groupby.enabled=true
hive> set hive.vectorized.execution.reduce.enabled;
hive.vectorized.execution.reduce.enabled=false
hive> set hive.vectorized.execution.enabled;
hive.vectorized.execution.enabled=true
hive>
```

Realizamos 4 ejecuciones activando todos los parámetros de vectorización de Hive:

```
[uambd20@master ejercicios]$ time hive -f ejercicio6.hql --hiveconf hive.vectorized.execution.enabled=true --hiveconf
 hive.vectorized.execution.reduce.enabled=true --hiveconf hive.vectorized.execution.reduce.groupby.enabled=true
```

```
20674359p.rfihub.com.   4
2.bp.blogspot.com.      4
ww.google.es.   2
as.com. 2
google.es.      2
quit.localdomain.       2
local.  2
as.es.  2
quit.   2
Time taken: 48.167 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release
() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.

real    0m56.751s
user    0m22.850s
sys     0m1.760s
[uambd20@master ejercicios]$
```

```
2.bp.blogspot.com.      4
ww.google.es.   2
as.com. 2
google.es.      2
quit.localdomain.       2
local.  2
as.es.  2
quit.   2
Time taken: 48.49 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFacto
ry#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expl
anation.

real    0m57.161s
user    0m22.141s
sys     0m1.800s
[uambd20@master ejercicios]$
```

```
1to1.bbva.com.  4
as.es.  2
quit.   2
local.  2
quit.localdomain.       2
ww.google.es.   2
as.com. 2
google.es.      2
Time taken: 50.263 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFacto
ry#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expl
anation.

real    0m58.769s
user    0m21.314s
sys     0m1.820s
[uambd20@master ejercicios]$
```

```
1to1.bbva.com.  4
as.es.  2
quit.   2
local.  2
quit.localdomain.       2
ww.google.es.   2
as.com. 2
google.es.      2
Time taken: 51.502 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFacto
ry#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an expl
anation.

real    1m0.039s
user    0m21.357s
sys     0m1.747s
[uambd20@master ejercicios]$
```

Realizamos 4 ejecuciones desactivando todos los parámetros de vectorización de Hive:

```
[uambd20@master ejercicios]$ time hive -f ejercicio6.hql --hiveconf hive.vectorized.execution.enabled=false --hiveconf
 hive.vectorized.execution.reduce.enabled=false --hiveconf hive.vectorized.execution.reduce.groupby.enabled=false
```

```
1to1.bbva.com.  4
as.es.  2
quit.   2
local.  2
quit.localdomain.       2
ww.google.es.   2
as.com. 2
google.es.      2
Time taken: 51.606 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLo
gFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for a
n explanation.

real    1m0.026s
user    0m21.510s
sys     0m1.679s
[uambd20@master ejercicios]$
```

```
1to1.bbva.com.  4
as.es.  2
quit.   2
local.  2
quit.localdomain.       2
ww.google.es.   2
as.com. 2
google.es.      2
Time taken: 61.243 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLo
gFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for a
n explanation.

real    1m10.104s
user    0m22.532s
sys     0m1.637s
[uambd20@master ejercicios]$
```

```
1to1.bbva.com.  4
as.es.  2
quit.   2
local.  2
quit.localdomain.       2
ww.google.es.   2
as.com. 2
google.es.      2
Time taken: 58.123 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLo
gFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for a
n explanation.

real    1m6.712s
user    0m21.762s
sys     0m1.824s
[uambd20@master ejercicios]$
```

```
2.bp.blogspot.com.      4
ww.google.es.   2
as.com. 2
google.es.      2
quit.localdomain.       2
local.  2
as.es.  2
quit.   2
Time taken: 50.769 seconds, Fetched: 1024 row(s)
WARN: The method class org.apache.commons.logging.impl.SLF4JLo
gFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for a
n explanation.

real    0m59.210s
user    0m20.928s
sys     0m1.452s
[uambd20@master ejercicios]$
```

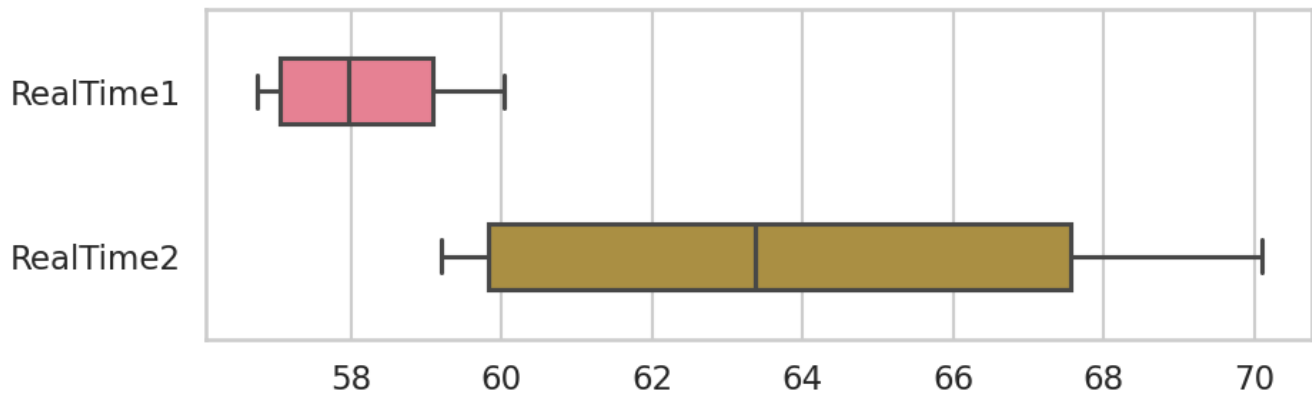| Question 10	Mean	Standard Deviation
hive.vectorized.execution.enabled=true hive.vectorized.execution.reduce.enabled=true hive.vectorized.execution.reduce.groupby.enabled=true	58.180	1.312
hive.vectorized.execution.enabled=false hive.vectorized.execution.reduce.enabled=false hive.vectorized.execution.reduce.groupby.enabled=false	64.013	4.565

Figura 3: Resultados de 4 ejecuciones con 3 parámetros diferentes y dos opciones true y false.

RealTime1 indica que hive.vectorized.execution.enabled=true,
hive.vectorized.execution.reduce.enabled=true,
hive.vectorized.execution.reduce.groupby.enabled=true.

RealTime2 indica que hive.vectorized.execution.enabled=false,
hive.vectorized.execution.reduce.enabled=false,
hive.vectorized.execution.reduce.groupby.enabled=false.

Observamos en la gráfica que al activar todos los parámetros de vectorización de Hive, conseguimos reducir tanto el tiempo de ejecución medio como la desviación estándar del tiempo de ejecución.