

Tarea de Vacaciones

Para esta tarea se deberá responder una serie de preguntas de temas que se han abordado en el curso, el objetivo consiste entonces en reforzar los conocimientos e indagar en otros nuevos que, aunque no forman parte explícita del temario, le servirán al estudiante para incursionar en temas Big Data a plenitud.

El estudiante debe crear primero que nada un directorio dentro de su directorio local de Git llamado **TareaVacaciones** y dentro de éste crear una copia de esta tarea que lleve por nombre **TareaX y colocarla juntos con los resultados en formato PDF**, donde X es su nombre de usuario empleado en Github. Por ejemplo:

TareaYoNoFui

Con respecto de los ejercicios, a menos que se indique lo contrario, todas las respuestas constarán del código o instrucción resultante acompañada de una captura de pantalla. Ejemplo:

-1.- Indique el comando que se emplea para listar archivos en GNU/Linux de manera simple:

Respuesta: **ls**

```
aaron@aaron-trabajo-vb:~/Descargas$ ls
apache-hive-2.3.3-bin.tar.gz      R-3.3.2.tar.gz
archivo.txt                      rattle_5.0.18.tar.gz
banner-principal2.png           Respaldo Usuaría Chile
core-site.xml                   RGtk2_2.20.33.tar.gz
db-derby-10.13.1.1-bin.tar.gz   rstudio-1.0.143-amd64.deb
file01.txt                      sas2txt.py
file02.txt                      scala-2.10.4.deb
file03.txt                      scala-2.12.1.tgz
```

En este tipo de preguntas de faltar alguno de los elementos señalados se considerará como errónea la respuesta y no se obtendrá el acierto.

Es menester mencionar que hay casos donde las preguntas son de tipo abierto, entonces en esos casos lo único que se pide adjuntar es tanto la respuesta como la(s) fuente(s). Ejemplo:

0.- ¿Cuál es el significado de la vida?

Respuesta: **42**

Fuente: <https://www.independent.co.uk/life-style/history/42-the-answer-to-life-the-universe-and-everything-2205734.html>

De nueva cuenta, si no existe al menos uno de estos dos elementos, la respuesta se considerará como inválida.

La fecha límite de entrega es el **Lunes 25 de Junio a las 15:00:00**, como se había mencionado con anterioridad el flujo de archivos se mantiene única y exclusivamente por Github, para ello se dejan los comandos a emplearse:

- **git pull** (para actualizar el repositorio)
- **git add .** (para indicar todos los elementos que se desean agregar al repositorio)
- **git commit -m "TareaVacaciones nombre_usuario"** (para colocar un mensaje que distinga a esta subida de las de los demás usuarios)
- **git push origin master** (para efectuar los cambios)

Por cierto que en lo que se repara Git en Cloudera puede ocupar Git de Windows y de esta manera, ya que se sugirió la instalación de Guest Additions en VirtualBox, copiar los resultados al primer sistema.

Nuevamente, de no cumplirse al menos uno de los señalamientos anteriores la tarea se considerará no entregada.

Dicho lo anterior se les desea mucho éxito en la travesía, cualquier cosa no duden en preguntar...

SECCION 1. GNU/LINUX

1.- Del archivo **aerolineas.csv** (el archivo descomprimido que todavía debería estar en su local y no el del HDFS) use comandos de GNU/Linux para obtener las 25 primeras líneas (incluyendo encabezado) **SIN** usar el comando head.

Respuesta: sed 25q aerolinea.csv

```
[cloudera@quickstart curso2]$ sed 25q aerolinea.csv
Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin
Dest,Distance,TaxiIn,TaxiOut,Canceled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay
1987,10,14,3,741,730,912,849,PS,1451,NA,91,79,NA,23,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,15,4,729,730,903,849,PS,1451,NA,94,79,NA,14,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,17,6,741,730,918,849,PS,1451,NA,97,79,NA,29,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,18,7,729,730,847,849,PS,1451,NA,78,79,NA,-2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,19,1,749,730,922,849,PS,1451,NA,93,79,NA,33,19,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,21,3,728,730,848,849,PS,1451,NA,80,79,NA,-1,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,22,4,728,730,852,849,PS,1451,NA,84,79,NA,3,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,23,5,731,730,902,849,PS,1451,NA,91,79,NA,13,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,24,6,744,730,908,849,PS,1451,NA,84,79,NA,19,14,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,25,7,729,730,851,849,PS,1451,NA,82,79,NA,2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,26,1,735,730,904,849,PS,1451,NA,89,79,NA,15,5,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,28,3,741,725,919,855,PS,1451,NA,98,98,NA,24,16,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,29,4,742,725,906,855,PS,1451,NA,84,98,NA,11,17,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,31,6,726,725,848,855,PS,1451,NA,82,98,NA,-7,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,1,4,936,915,1035,1001,PS,1451,NA,59,46,NA,34,21,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,2,5,918,915,1017,1001,PS,1451,NA,59,46,NA,16,3,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,3,6,928,915,1037,1001,PS,1451,NA,69,46,NA,36,13,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,4,7,914,915,1003,1001,PS,1451,NA,49,46,NA,2,-1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,5,1,1042,915,1129,1001,PS,1451,NA,47,46,NA,88,87,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,6,2,934,915,1024,1001,PS,1451,NA,50,46,NA,23,19,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,7,3,946,915,1037,1001,PS,1451,NA,51,46,NA,36,31,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,8,4,932,915,1033,1001,PS,1451,NA,61,46,NA,32,17,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,9,5,947,915,1036,1001,PS,1451,NA,49,46,NA,35,32,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
1987,10,10,6,915,915,1022,1001,PS,1451,NA,67,46,NA,21,0,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA
```

2.-Como ya se ha visto, utilizar el redireccionamiento destructivo (>) implica almacenar típicamente algún contenido en un archivo (ej. **echo "contenido" > archivo**).

Pero lo cierto es que con este comando no se apreciará en pantalla lo que se desea almacenar en dicho archivo, por ello es que se necesita que, con base en el comando resultado del ejercicio 1 y

con la investigación del comando **tee**, por un lado el contenido se introduzca en el archivo **ejercicio_2.txt** y por el otro se muestre en pantalla la operación.
Caber mencionar que todo se debe registrar como una sola instrucción, es decir, no se puede ejecutar el resultado por partes, para ello tal vez quiera leer esta liga:

<http://www.linfo.org/pipes.html>

Respuesta: *sed 25q aerolinea.csv | tee ejercicio_2.txt*

```
[cloudera@quickstart curso2]$ sed 25q aerolinea.csv | tee ejercicio_2.txt
Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin
Dest,Distance,TaxiIn,TaxiOut,Canceled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay
1987,10,14,3,741,730,912,849,PS,1451,NA,91,79,NA,23,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,15,4,729,730,903,849,PS,1451,NA,94,79,NA,14,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,17,6,741,730,918,849,PS,1451,NA,97,79,NA,29,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,18,7,729,730,847,849,PS,1451,NA,78,79,NA,-2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,19,1,749,730,922,849,PS,1451,NA,93,79,NA,33,19,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,21,3,728,730,848,849,PS,1451,NA,80,79,NA,-1,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,22,4,728,730,852,849,PS,1451,NA,84,79,NA,3,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,23,5,731,730,902,849,PS,1451,NA,91,79,NA,13,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,24,6,744,730,908,849,PS,1451,NA,84,79,NA,19,14,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,25,7,729,730,851,849,PS,1451,NA,82,79,NA,2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,26,1,735,730,904,849,PS,1451,NA,89,79,NA,15,5,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,28,3,741,725,919,855,PS,1451,NA,98,98,NA,24,16,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,29,4,742,725,906,855,PS,1451,NA,84,98,NA,11,17,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,31,6,726,725,848,855,PS,1451,NA,82,98,NA,-7,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,1,4,936,915,1035,1001,PS,1451,NA,59,46,NA,34,21,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,2,5,918,915,1017,1001,PS,1451,NA,59,46,NA,16,3,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,3,6,928,915,1037,1001,PS,1451,NA,69,46,NA,36,13,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,4,7,914,915,1003,1001,PS,1451,NA,49,46,NA,2,-1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,5,1,1042,915,1129,1001,PS,1451,NA,47,46,NA,88,87,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,6,2,934,915,1024,1001,PS,1451,NA,50,46,NA,23,19,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,7,3,946,915,1037,1001,PS,1451,NA,51,46,NA,36,31,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,8,4,932,915,1033,1001,PS,1451,NA,61,46,NA,32,17,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,9,5,947,915,1036,1001,PS,1451,NA,49,46,NA,35,32,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,10,6,915,915,1022,1001,PS,1451,NA,67,46,NA,21,0,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
[cloudera@quickstart curso2]$ ls
aerolinea.csv  ejercicio_2.txt
```

3.- Cambie el nombre del archivo **ejercicio_2.txt** a **ejercicio_3.txt** SIN usar el comando **rename**

Respuesta: *mv ejercicio_2.txt ejercicio_3.txt*

```
[cloudera@quickstart curso2]$ mv ejercicio_2.txt ejercicio_3.txt
[cloudera@quickstart curso2]$ ls
aerolinea.csv  ejercicio_3.txt
[cloudera@quickstart curso2]$ █
```

4.- Con algún comando en GNU/Linux tome las 25 últimas líneas del archivo aerolínea.csv **SIN** emplear el comando tail y guárdelo como **ejercicio_4.txt**

Respuesta: `sed -e :a -e '$q;N;26,$D;ba' aerolinea.csv | tee ejercicio_4.txt`

```
[cloudera@quickstart curso2]$ sed -e :a -e '$q;N;26,$D;ba' aerolinea.csv | tee ejercicio_4.txt
2008,12,13,6,1910,1910,2017,2016,DL,1612,N927DA,67,66,38,1,0,ATL,CHS,259,5,24,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1441,1445,1604,1622,DL,1613,N973DL,83,97,65,-18,-4,IND,ATL,432,8,10,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,921,830,1112,1008,DL,1616,N907DE,111,98,82,64,51,ATL,PBI,545,8,21,0,,0,51,0,13,0,0
2008,12,13,6,1435,1440,1701,1704,DL,1618,N914DL,86,84,56,-3,-5,MSY,ATL,425,20,10,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1750,1755,2010,2015,DL,1618,N914DL,140,140,113,-5,-5,ATL,BDL,859,7,20,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,706,710,850,837,DL,1619,N949DL,104,87,49,13,-4,LEX,ATL,303,23,32,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1552,1520,1735,1718,DL,1620,N905DE,43,58,27,17,32,HSV,ATL,151,9,7,0,,0,0,0,0,0,17
2008,12,13,6,1250,1220,1617,1552,DL,1621,N938DL,147,152,120,25,30,MSP,ATL,906,9,18,0,,0,3,0,0,0,22
2008,12,13,6,1033,1041,1255,1303,DL,1622,N935DL,82,82,58,-8,-8,MSY,ATL,425,9,15,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,840,843,1025,1021,DL,1624,N3738B,105,98,53,4,-3,SLC,DEN,391,6,46,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,810,815,1504,1526,DL,1625,N3742C,234,251,210,-22,-5,LAX,CVG,1900,7,17,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,547,545,646,650,DL,1627,N621DL,59,65,38,-4,2,SAV,ATL,215,8,13,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,848,850,1024,1005,DL,1628,N920DL,156,135,108,19,-2,ATL,MCI,692,4,44,0,,0,0,0,19,0,0
2008,12,13,6,936,936,1114,1119,DL,1630,N653DL,96,86,56,6,-4,MSY,ATL,425,23,17,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,657,600,904,749,DL,1631,N3743H,127,109,78,75,57,RIC,ATL,481,15,34,0,,0,0,57,18,0,0
2008,12,13,6,1007,847,1149,1010,DL,1631,N909DA,162,143,122,99,80,ATL,IAH,689,8,32,0,,0,1,0,19,0,79
2008,12,13,6,638,640,808,753,DL,1632,N604DL,90,73,50,15,-2,JAX,ATL,270,14,26,0,,0,0,0,15,0,0
2008,12,13,6,756,800,1032,1026,DL,1633,N642DL,96,86,56,6,-4,MSY,ATL,425,23,17,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,612,615,923,907,DL,1635,N907DA,131,112,103,16,-3,GEG,SLC,546,5,23,0,,0,0,0,16,0,0
2008,12,13,6,749,750,901,859,DL,1636,N646DL,72,69,41,2,-1,SAV,ATL,215,20,11,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1002,959,1204,1150,DL,1636,N646DL,122,111,71,14,3,ATL,IAD,533,6,45,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,834,835,1021,1023,DL,1637,N908DL,167,168,139,-2,-1,ATL,SAT,874,5,23,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,655,700,856,856,DL,1638,N671DN,121,116,85,0,-5,PBI,ATL,545,24,12,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1251,1240,1446,1437,DL,1639,N646DL,115,117,89,9,11,IAD,ATL,533,13,13,0,,0,NA,NA,NA,NA,NA
2008,12,13,6,1110,1103,1413,1418,DL,1641,N908DL,123,135,104,-5,7,SAT,ATL,874,8,11,0,,0,NA,NA,NA,NA,NA
[cloudera@quickstart curso2]$ ls
aerolinea.csv  ejercicio_3.txt  ejercicio_4.txt
[cloudera@quickstart curso2]$
```

5.- Concatene los archivos **ejercicio_3.txt** y **ejercicio_4.txt** en un archivo **ejercicio_5.txt** y en esa misma pantalla resultado muestre el contenido de **ejercicio_5.txt**

Respuesta: `cat ejercicio_3.txt ejercicio_4.txt | tee ejercicio_5.txt`

```
[cloudera@quickstart curso2]$ cat ejercicio_3.txt ejercicio_4.txt | tee ejercicio_5.txt
Year,Month,DayOfMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,
Dest,Distance,TaxiIn,TaxiOut,Canceled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay
1987,10,14,3,741,730,912,849,PS,1451,NA,91,79,NA,23,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,15,4,729,730,903,849,PS,1451,NA,94,79,NA,14,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,17,6,741,730,918,849,PS,1451,NA,97,79,NA,29,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,18,7,729,730,847,849,PS,1451,NA,78,79,NA,-2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,19,1,749,730,922,849,PS,1451,NA,93,79,NA,33,19,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,21,3,728,730,848,849,PS,1451,NA,80,79,NA,-1,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,22,4,728,730,852,849,PS,1451,NA,84,79,NA,3,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,23,5,731,730,902,849,PS,1451,NA,91,79,NA,13,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,24,6,744,730,908,849,PS,1451,NA,84,79,NA,19,14,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,25,7,729,730,851,849,PS,1451,NA,82,79,NA,2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,26,1,735,730,904,849,PS,1451,NA,89,79,NA,15,5,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,28,3,741,725,919,855,PS,1451,NA,98,90,NA,24,16,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,29,4,742,725,906,855,PS,1451,NA,84,90,NA,11,17,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,31,6,726,725,848,855,PS,1451,NA,82,90,NA,-7,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,1,4,936,915,1035,1001,PS,1451,NA,59,46,NA,34,21,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,2,5,918,915,1017,1001,PS,1451,NA,59,46,NA,16,3,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,3,6,928,915,1037,1001,PS,1451,NA,69,46,NA,36,13,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,4,7,914,915,1003,1001,PS,1451,NA,49,46,NA,2,-1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,5,1,1042,915,1129,1001,PS,1451,NA,47,46,NA,88,87,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,6,2,924,915,1032,1001,PS,1451,NA,54,46,NA,77,10,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA
```

6.- Usando el comando **ls** y sus opciones, verifique el peso de **ejercicio_5.txt**, señalando en la captura de pantalla dónde se encuentra éste.

*Respuesta: **ls -lh***

```
[cloudera@quickstart curso2]$ ls -lh
total 12G
-rwxrwx--- 1 cloudera cloudera 12G Jun 13 13:37 aerolinea.csv
-rw-rw-r-- 1 cloudera cloudera 2.6K Jun 18 12:18 ejercicio_3.txt
-rw-rw-r-- 1 cloudera cloudera 2.5K Jun 18 13:38 ejercicio_4.txt
-rw-rw-r-- 1 cloudera cloudera 5.0K Jun 19 06:50 ejercicio_5.txt
```

7.- Modifique la fecha de acceso de **ejercicio_5.txt** al 25 de Agosto del 2018 y muestre en pantalla dónde se puede apreciar ese resultado.

*Respuesta: **touch -t 201808251500.00 ejercicio_5.txt***

```
[cloudera@quickstart curso2]$ touch -t 201808251500.00 ejercicio_5.txt
[cloudera@quickstart curso2]$ ls -l
total 11747296
-rwxrwx--- 1 cloudera cloudera 12029208594 Jun 13 13:37 aerolinea.csv
-rw-rw-r-- 1 cloudera cloudera 2584 Jun 18 12:18 ejercicio_3.txt
-rw-rw-r-- 1 cloudera cloudera 2462 Jun 18 13:38 ejercicio_4.txt
-rw-rw-r-- 1 cloudera cloudera 5046 Aug 25 2018 ejercicio_5.txt
[cloudera@quickstart curso2]$
```

8.- ¿Con cuál comando se puede averiguar el número de núcleos en un sistema GNU/Linux? Investigue y coloque el resultado, haciendo énfasis en el lugar donde se puede apreciar esa información.

*Respuesta: **nproc --all***

```
[cloudera@quickstart curso2]$ nproc --all
1
[cloudera@quickstart curso2]$
```

9.- Investigue en qué consiste **awk** y por medio de esa herramienta imprima en pantalla sólo la tercera y quinta columnas (de izquierda a derecha) del archivo **ejercicio_5.txt**. He aquí un ejemplo de cómo se ve el resultado con otro archivo que no tiene que ver con el curso:

```

aaron@aaron-trabajo-vb:~/Descargas$ cat muestra.txt
a,c,b,d,e,f
g,h,i,j,k,l
m,n,o,p,q,r
aaron@aaron-trabajo-vb:~/Descargas$ awk
be
ik
oq
aaron@aaron-trabajo-vb:~/Descargas$

```

Respuesta: `cat ejercicio_5.txt | awk -F "," '{ print $3 "," $5 }'`

```

[cloudera@quickstart curso2]$ cat ejercicio_5.txt | awk -F "," '{ print $3 "," $5 }'
DayOfMonth,DepTime
14,741
15,729
17,741
18,729
19,749
21,728
22,728
23,731
24,744
25,729
26,735
28,741
29,742
31,726
1,936
2,918
3,928
4,914
5,1042
6,934
7,946
8,932
9,947
10,915
13,1910
13,1441
13,921
13,1435
13,1750
13,706
13,1552

```

10.- Sin usar vim, nano o editor de texto alguno use comandos de Linux para reemplazar TODOS los elementos de la segunda columna por -1, guárdelo como **archivo_6.txt** y hágale un cat a ese mismo archivo.

Respuesta: `cat ejercicio_5.txt | awk -F "," '{ $2=-1 } { print }'`

```

[cloudera@quickstart curso2]$ cat ejercicio_5.txt | awk -F "," '{ $2=-1 } { print }'
Year -1 DayOfMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArrTime UniqueCarrier FlightNum TailNum ActualElapsedTime CRSElapsedTime AirTime ArrDelay DepDelay Origin Des
t Distance TaxiIn TaxiOut Cancelled CancellationCode Diverted CarrierDelay WeatherDelay NASDelay SecurityDelay LateAircraftDelay
1987 -1 14 3 741 730 912 849 PS 1451 NA 91 79 NA 23 11 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 15 4 729 730 903 849 PS 1451 NA 94 79 NA 14 -1 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 17 6 741 730 918 849 PS 1451 NA 97 79 NA 29 11 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 18 7 729 730 847 849 PS 1451 NA 78 79 NA -2 -1 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 19 1 749 730 922 849 PS 1451 NA 93 79 NA 33 19 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 21 3 728 730 848 849 PS 1451 NA 80 79 NA -1 -2 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 22 4 728 730 852 849 PS 1451 NA 84 79 NA 3 -2 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 23 5 731 730 902 849 PS 1451 NA 91 79 NA 13 1 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 24 6 744 730 908 849 PS 1451 NA 84 79 NA 19 14 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 25 7 729 730 851 849 PS 1451 NA 82 79 NA 2 -1 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 26 1 735 730 904 849 PS 1451 NA 89 79 NA 15 5 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 28 3 741 725 919 855 PS 1451 NA 98 90 NA 24 16 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 29 4 742 725 906 855 PS 1451 NA 84 90 NA 11 17 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 31 6 726 725 848 855 PS 1451 NA 82 90 NA -7 1 SAN SFO 447 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 1 4 936 915 1035 1001 PS 1451 NA 59 46 NA 34 21 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 2 5 918 915 1017 1001 PS 1451 NA 59 46 NA 16 3 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 3 6 928 915 1037 1001 PS 1451 NA 69 46 NA 36 13 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 4 7 914 915 1003 1001 PS 1451 NA 49 46 NA 2 -1 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 5 1 1042 915 1129 1001 PS 1451 NA 47 46 NA 88 87 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 6 2 934 915 1024 1001 PS 1451 NA 50 46 NA 23 19 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 7 3 046 015 1027 1001 PS 1451 NA 51 46 NA 26 21 SFO RNO 102 NA NA 0 NA 0 NA NA NA NA NA

```

SECCION 2. HDFS Y HIVE

11.- Se está tratando de hacer la siguiente operación:

```
hdfs dfs -head /raw/aerolínea.csv
```

Con una captura muestre qué es lo que pasa y por medio de argumentos sólidos (una captura de pantalla con la evidencia, una fuente de consulta) por qué sucede esto.

```
[cloudera@quickstart curso2]$ hdfs dfs -head /raw/aerolínea.csv
-head: Unknown command
[cloudera@quickstart curso2]$
```

Respuesta: *No existe el comando head en el HDFS, se muestra la pantalla con los comandos existentes en HDFS.*

```
[cloudera@quickstart curso2]$ hdfs dfs
Usage: hadoop fs [generic options]
    [-appendToFile <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:[GROUP]] PATH...]
    [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
    [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-count [-q] [-h] [-v] [-x] <path> ...]
    [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] [-x] <path> ...]
    [-expunge]
    [-find <path> ... <expression> ...]
    [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] <src> <localdst>]
    [-help [cmd ...]]
    [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]]
    [-mkdir [-p] <path> ...]
    [-moveFromLocal <localsrc> ... <dst>]
    [-moveToLocal <src> <localdst>]
    [-mv <src> ... <dst>]
    [-put [-f] [-p] [-l] <localsrc> ... <dst>]
    [-renameSnapshot <snapshotDir> <oldName> <newName>]
    [-rm [-f] [-r|-R] [-skipTrash] <src> ...]
    [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
    [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
    [-setfattr {-n name [-v value] | -x name} <path>]
    [-setrep [-R] [-w] <rep> <path> ...]
    [-stat [format] <path> ...]
    [-tail [-f] <file>]
```



```
[-test [-defsz] <path>]
[-text [-ignoreCrc] <src> ...]
[-touchz <path> ...]
[-usage [cmd ...]]
```

Generic options supported are

-conf <configuration file> specify an application configuration file

12.- Cuente cuántas líneas tiene el archivo **aerolínea.csv** que está en el **HDFS**. Recuerde el carácter pipe (|) empleado en ejercicios anteriores.

Respuesta: **hdfs dfs -cat /raw/aerolinea.csv | wc -l**

```
[cloudera@quickstart curso2]$ hdfs dfs -cat /raw/aerolinea.csv | wc -l
123534972
[cloudera@quickstart curso2]$ █
```

13.- Indague en la instrucción de HDFS para averiguar el factor de réplica del archivo **aerolínea.csv** y colóquelo aquí junto con captura del resultado.

Respuesta: **hdfs dfs -stat %r /raw/aerolinea.csv**

```
[cloudera@quickstart curso2]$ hdfs dfs -stat %r /raw/aerolinea.csv
1
[cloudera@quickstart curso2]$ █
```

14.- Tome como base el archivo **aerolínea.csv** del HDFS y almacene en el sistema local un archivo **ejercicio_14.txt** que contenga las primeras 15 líneas sin usar el comando **-tail** del HDFS. Muestre ese contenido también.

Respuesta: **hdfs dfs -cat /raw/aerolinea.csv | head -n 15 | tee ejercicio_14.txt**

```
[cloudera@quickstart curso2]$ hdfs dfs -cat /raw/aerolinea.csv | head -n 15 | tee ejercicio_14.txt
Year,Month,DayOfMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,
Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay
1987,10,14,3,741,730,912,849,PS,1451,NA,91,79,NA,23,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA
1987,10,15,4,729,730,903,849,PS,1451,NA,94,79,NA,14,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,17,6,741,730,918,849,PS,1451,NA,97,79,NA,29,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,18,7,729,730,847,849,PS,1451,NA,78,79,NA,-2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,19,1,749,730,922,849,PS,1451,NA,93,79,NA,33,19,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,21,3,728,730,848,849,PS,1451,NA,80,79,NA,-1,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,22,4,728,730,852,849,PS,1451,NA,84,79,NA,3,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,23,5,731,730,902,849,PS,1451,NA,91,79,NA,13,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,24,6,744,730,908,849,PS,1451,NA,84,79,NA,19,14,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,25,7,729,730,851,849,PS,1451,NA,82,79,NA,2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,26,1,735,730,904,849,PS,1451,NA,89,79,NA,15,5,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,28,3,741,725,919,855,PS,1451,NA,98,90,NA,24,16,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,29,4,742,725,906,855,PS,1451,NA,84,90,NA,11,17,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
1987,10,31,6,726,725,848,855,PS,1451,NA,82,90,NA,-7,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA
```

15.- Cree los directorios **master** y **stagin** en el directorio raíz del HDFS y además al archivo **aerolínea.csv** que está en raw cámbiele los permisos de tal manera que el propietario tenga todas las facilidades sobre él, el grupo sólo pueda leer y escribir y cualquier otro no tenga ningún permiso. Coloque las capturas de ambos ejercicios por separado.

Respuesta: **sudo -u hdfs hdfs dfs -mkdir /master /stagin**

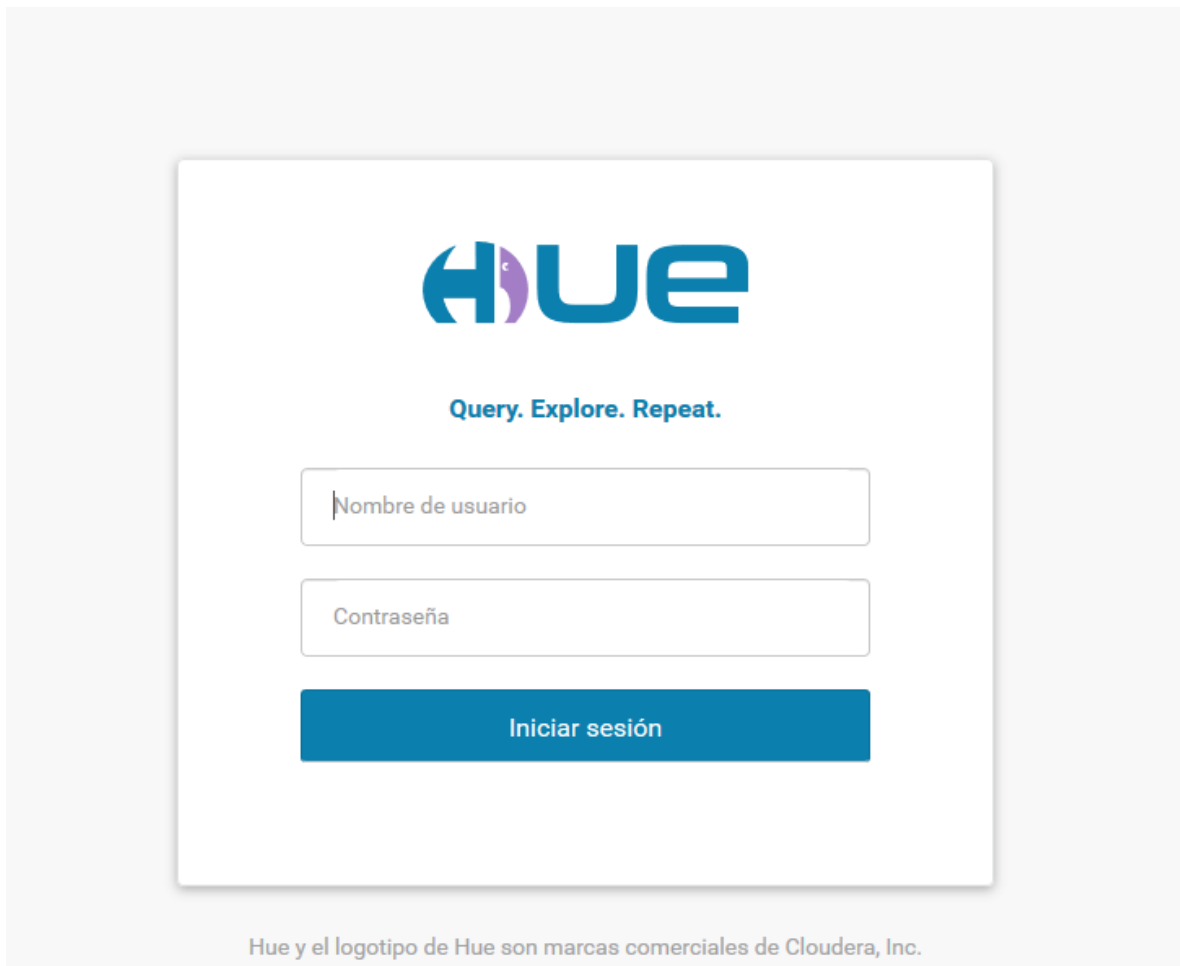

```
[cloudera@quickstart curso2]$ sudo -u hdfs hdfs dfs -mkdir /master /staging
[cloudera@quickstart curso2]$ sudo -u hdfs hdfs dfs -ls /
Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - cloudera supergroup 0 2018-03-09 09:44 /carpetal
drwxr-xr-x - hbase supergroup 0 2018-02-14 06:02 /hbase
drwxr-xr-x - hdfs supergroup 0 2018-06-20 08:11 /master
drwxr-xr-x - cloudera supergroup 0 2018-06-13 16:32 /raw
drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr
drwxr-xr-x - hdfs supergroup 0 2018-06-20 08:11 /staging
drwxrwxrwt - hdfs supergroup 0 2018-03-12 06:54 /tmp
drwxr-xr-x - hdfs supergroup 0 2018-02-14 06:58 /user
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var
```

Respuesta: `hdfs dfs -chmod 760 /raw/aerolinea.csv`

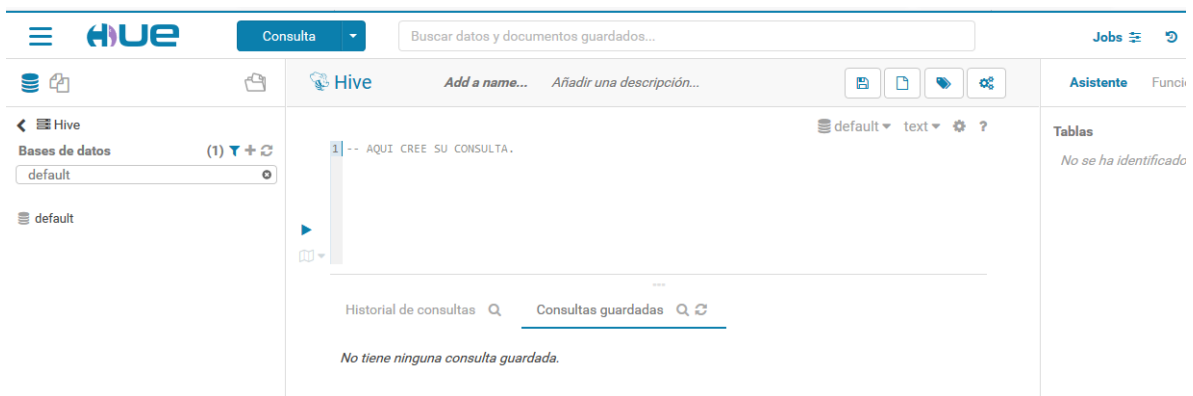
```
[cloudera@quickstart curso2]$ hdfs dfs -chmod 760 /raw/aerolinea.csv
[cloudera@quickstart curso2]$ hdfs dfs -ls /raw
Found 1 items
-rwxrw---- 1 cloudera supergroup 12029208594 2018-06-13 16:32 /raw/aerolinea.csv
[cloudera@quickstart curso2]$ █
```

16.- Para los siguientes ejercicios puede hacer uso del servicio Hue (si no ha activado los servicios en Cloudera Manager tiene que hacerlo antes, para entrar a Hue en el mismo navegador se encuentra esta opción).

Aparecerá una ventana como ésta:



Recuerde que tanto el usuario como la contraseña es **cloudera**:



Entonces tome el siguiente código y cree una tabla en Hive:

```
CREATE EXTERNAL TABLE tabla_aerolinea(
```

```

Year STRING,
Month STRING,
DayofMonth STRING,
DayOfWeek STRING,
DepTime STRING,
CRSDepTime STRING,
ArrTime STRING,
CRSArrTime STRING,
UniqueCarrier STRING,
FlightNum STRING,
TailNum STRING,
ActualElapsedTime STRING,
CRSElapsedTime STRING,
AirTime STRING,
ArrDelay STRING,
DepDelay STRING,
Origin STRING,
Dest STRING,
Distance STRING,
TaxiIn STRING,
TaxiOut STRING,
Cancelled STRING,
CancellationCode STRING,
Diverted STRING,
CarrierDelay STRING,
WeatherDelay STRING,
NASDelay STRING,
SecurityDelay STRING,
LateAircraftDelay STRING)

ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
location '/raw'

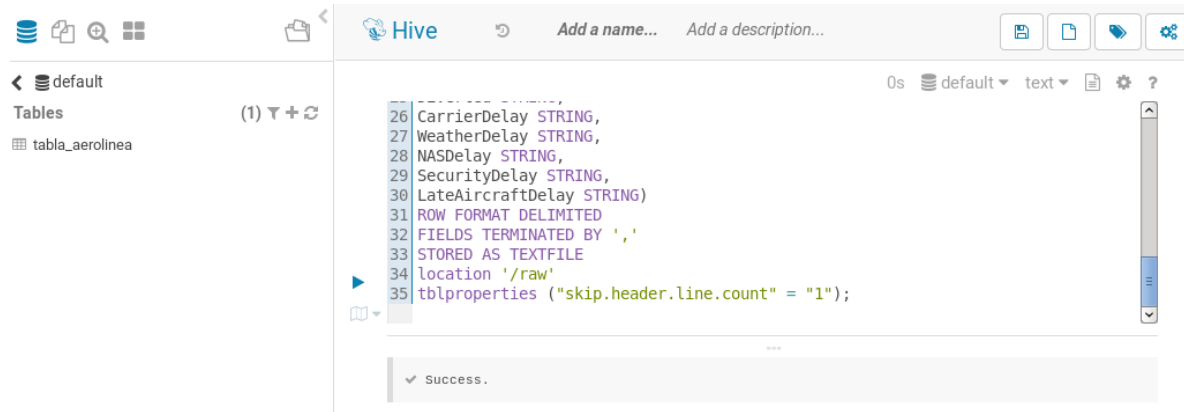
;

```

En el código anterior **NO** existe una forma de omitir los encabezados por lo que es su deber encontrar esa manera, incluirla en el código y crear la tabla.

Para acreditar el ejercicio debe mostrar la sentencia que requirió para la parte de los encabezados y hacer un SELECT de los 10 primeros elementos de la tabla.

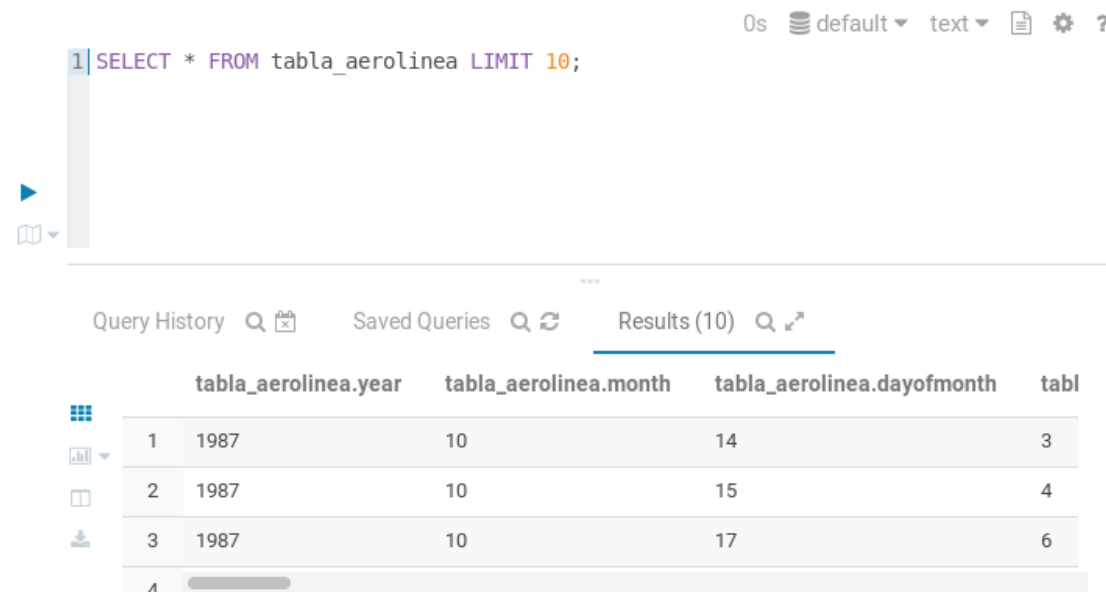
Respuesta: **tblproperties ("skip.header.line.count" = "1")**



The screenshot shows the Hive SQL editor interface. On the left, a sidebar lists the 'default' database and a table named 'tabla_aerolinea'. The main editor area contains the following SQL code:

```
26 CarrierDelay STRING,  
27 WeatherDelay STRING,  
28 NASDelay STRING,  
29 SecurityDelay STRING,  
30 LateAircraftDelay STRING)  
31 ROW FORMAT DELIMITED  
32 FIELDS TERMINATED BY ','  
33 STORED AS TEXTFILE  
34 location '/raw'  
35 tblproperties ("skip.header.line.count" = "1");
```

Below the code editor, a status bar indicates 'Success.' with a green checkmark icon.



The screenshot shows the Hive SQL editor interface with a query executed. The query is:

```
1 SELECT * FROM tabla_aerolinea LIMIT 10;
```

The results are displayed in a table with 4 columns: 'tabla_aerolinea.year', 'tabla_aerolinea.month', 'tabla_aerolinea.dayofmonth', and 'tabla_aerolinea.tabl'. The table shows the first 3 rows of data.

	tabla_aerolinea.year	tabla_aerolinea.month	tabla_aerolinea.dayofmonth	tabla_aerolinea.tabl
1	1987	10	14	3
2	1987	10	15	4
3	1987	10	17	6

17.- Borre la tabla anterior y vuélvala a crear pero ahora el tipo de dato Origin debe ser INT, entonces vuelva a ejecutar la consulta y especifique qué ha pasado y con una captura muéstrelo.

Respuesta: **Convierte todo a null ya que el origen de datos es texto y no lo puede guardar en un campo tipo INT.**

Hive

Add a name... Add a description...

0s default text

```
1 SELECT Origin A FROM tabla_aerolinea
```

Query History Saved Queries Results (100+)

	a
1	NULL
2	NULL
3	NULL
4	NULL

18.- Borre la tabla anterior, vuélvala a crear (con Origin STRING) pero ahora añada una columna después de LateAircraftDelay llamada **Adicional** con tipo de dato **STRING**, ejecute la creación, indique qué ha sucedido y coloque captura del resultado.

Respuesta: **Agrega todo NULL al no tener datos para hacer match con este campo.**

0s default text

```
1 SELECT * FROM tabla_aerolinea
```

Query History Saved Queries Results (100+)

	linea.securitydelay	tabla_aerolinea.lateaircraftdelay	tabla_aerolinea.adicional
1		NA	NULL
2		NA	NULL
3		NA	NULL

19.- En esta tabla anterior inserte un renglón a la tabla con todos los valores iguales a "NA" (tiene que investigar cómo añadir elementos a la tabla), y luego después de la inserción del elemento indague en qué parte del HDFS se ha guardado ese nuevo elemento.

Respuesta: Se genera el siguiente directorio en el HDFS.

```
drwxrwxrwx - cloudera supergroup 0 2018-06-25 10:16 /raw/.hive-staging_hive_2018-06-25_10-15-54_572_7416140232936670991-4
-rwxrwxrwx 1 cloudera supergroup 12029208594 2018-06-13 16:32 /raw/aerolinea.csv
```

SECCIÓN 3. PREGUNTAS ABIERTAS

20.- ¿Qué es el Sticky Bit? Ejemplifíquelo con el archivo **ejercicio_5.txt** adjuntando una captura de pantalla.

Respuesta:

Permite evitar que un usuario pueda borrar ficheros/directorios de otro usuario dentro de ese directorio, ya que todos tienen permiso de escritura

```
[cloudera@quickstart curso2]$ chmod 1777 ejercicio_5.txt
[cloudera@quickstart curso2]$ ls -l
total 11747316
-rwxrwx--- 1 cloudera cloudera 12029208594 Jun 13 13:37 aerolinea.csv
-rw-rw-r-- 1 cloudera cloudera          0 Jun 19 09:22 algo.txt
-rw-rw-r-- 1 cloudera cloudera      648 Jun 20 09:00 derby.log
-rw-rw-r-- 1 cloudera cloudera     1625 Jun 20 08:00 ejercicio_14.txt
-rw-rw-r-- 1 cloudera cloudera     2584 Jun 18 12:18 ejercicio_3.txt
-rw-rw-r-- 1 cloudera cloudera     2462 Jun 18 13:38 ejercicio_4.txt
-rwxrwxrwt 1 cloudera cloudera     5046 Aug 25 2018 ejercicio_5.txt
```

21.- ¿A qué se le conoce como NoSQL?, ¿considera que Hive e Impala son representantes?

Respuesta:

Las bases de datos No solo SQL son sistemas de gestión de datos que no usan SQL como lenguaje principal de consulta, con un enfoque de procesar grandes cantidades de información, Si considero que Hive e Impala son representantes de NoSQL.

22.- Investigue el uso del comando nohup en GNU/Linux y con base en esto responda: ¿cómo puede ser aplicado dicho comando en un sistema distribuido?

Respuesta:

Mantiene la ejecución de un comando pese a salir de la terminal

En un sistema distribuido nohup nos proporciona inmunidad frente a rupturas de comunicación y abandonos de la sesión en alguno o varios nodos del sistema.

23.- Se quiere averiguar la memoria RAM disponible con base en la siguiente imagen:

```
aaron@aaron-trabajo-vb:~/Descargas$ free -h
```

	total	usado	libre	compart.	buffers	almac.
Memoria:	3.9G	2.7G	1.2G	16M	66M	1.9G
-/+ buffers/cache:		700M	3.2G			
Swap:	4.0G	176K	4.0G			

Respuesta:

1.2G Total de Memoria RAM Disponible

3.2G Es la memoria Disponible tomando en cuenta los proceso que utilizaron memoria pero aun no la liberan pero que se puede considerar disponible.

24.- Se tiene el siguiente escenario: personal ajeno a su área de sistemas desea tener acceso al sistema, en particular para ver algunos datos del archivo **objetivo.txt**

Por otra parte se sabe de manera extraoficial que la meta de ellos consiste en “ensuciar” el archivo para que el área no tenga tanto repunte como la nuestra.

Por cuestiones burocráticas la creación de algún usuario nuevo no es plausible no obstante debido a asuntos políticos es prácticamente un hecho que se le tiene que dar permiso, por ello es que se optó por prestarles un usuario (**usuario_nuestro**) cuyo grupo es **grupo_nuestro**.

Con base en estas características y limitando el escenario únicamente a comandos **chmod** (y si lo desea **chown** y **chgrp**), ¿cuál sería la configuración que usted propondría para garantizar el acceso al archivo pero al mismo tiempo protegerlo de las circunstancias mencionadas y sin afectar al mismo tiempo a los demás miembros de **grupo_nuestro**?

Respuesta:

Podría usar **chown** para dar la propiedad del archivo **objetivo.txt** a **usuario_nuestro** para luego asignar los siguientes permisos **chmod 466 objetivo.txt** de esta forma el usuario propietario no tiene permisos de edición mientras que el resto del grupo no tiene afectación. ☺

(Aunque claro el usuario podría darse permisos usando el **chmod**) ☹

25.- ¿Cuál es la diferencia entre Hadoop y Cloudera?

Respuesta:

Hadoop es la distribución de Hadoop de la comunidad Apache, mientras que Cloudera Hadoop tiene su propia distribución de Hadoop, que está construida sobre Apache Hadoop. Cloudera Hadoop incluye herramientas adicionales \ tecnologías que no están presentes en otras distribuciones Hadoop como Cloudera Search, Impala, Cloudera Navigator y Cloudera

Manager. Con estas herramientas adicionales, Cloudera Hadoop resulta ser parcialmente de código abierto y parcialmente patentado.

26.- ¿Cuáles son los tipos de archivos existentes en GNU/Linux y Windows?

Respuesta:

LINUX

Archivos físicos: Estos son los archivos en el sentido más conocido, es decir, un fragmento de información con un nombre.

Directorios: Son archivos (nodos) que forman parte de la estructura jerárquica. Se conceptualizan como contenedores de archivos, aunque realmente son archivos que contienen una lista de otros archivos

Enlaces: Son archivos especiales que permiten que contienen una suerte de atajo a otros archivos y que permiten acceder a dichos archivos desde distintas rutas sin tener que tener más que una copia del archivo original.

Archivos virtuales: No son archivos reales, pues solo existen en la memoria en forma de procesos y son accesibles desde el sistema de archivos. La mayoría de estos archivos se ubican en el directorio `/proc` y como procesos se crean y se destruyen en cada sesión. Acceder a ellos nos permite obtener información del sistema (memoria, procesador, particiones, etc)

Archivos de dispositivo: Estos archivos, ubicados en el directorio `/dev` permiten acceder a los distintos dispositivos conectados al computador. Las unidades de almacenamiento y otros dispositivos, al ser tratados como archivos soportan tareas de direccionamiento y redireccionamiento.

27.- ¿Qué es el SerDe y cuál es su relación con Hive e Impala?

Respuesta:

Es la abreviatura de Serializador / Deserializador. Hive usa la interfaz SerDe para IO. La interfaz maneja tanto la serialización como la deserialización y también interpreta los resultados de la serialización como campos individuales para el procesamiento.

Un SerDe le permite a Hive leer los datos de una tabla y escribirlos nuevamente en HDFS en cualquier formato personalizado. Cualquiera puede escribir su propio SerDe para sus propios formatos de datos.

28.- ¿A qué se le conoce como Big Table y Big Query?

Respuesta:

Big Table: Sistema de gestión de base de datos creado por Google distribuido, de alta eficiencia y propietario.

BigQuery es un servicio web que permite el análisis interactivo de grandes conjuntos de datos que trabajan en conjunto con Google Storage . Es una Infraestructura como Servicio (IaaS) que puede usarse de forma complementaria con MapReduce .

29.- ¿A qué se le denomina Data Lake y Data Warehouse?

Respuesta:

Data Lake: Un data lake es un repositorio donde se almacenan todos los datos de la compañía, estructurados y sin estructurar, sin ningún tipo de preprocesamiento (raw data) y sin ningún tipo de esquema, para ser analizados posteriormente.

Data Warehouse: Es un repositorio o base de datos que se alimenta de varias fuentes que se han transformado o depurado en grupos de información de temas específicos relacionados con el mundo de los negocios

30.- ¿Existe algún otro tipo de sistemas de archivos distribuidos que NO sea HDFS? si es así, ¿de cuáles se trata?

Respuesta:

Si, AFS (Andrew File System).

SECCIÓN 4. ESPECIAL

31.- Instale Jupyter en Cloudera, para ello puede basarse en la siguiente liga:

<https://medium.com/@vando/install-jupyter-notebook-on-centos-7-1d596abf08da>

Es importante señalar que para continuar el curso es imprescindible esta herramienta y no existirán pausas para su instalación durante las sesiones, motivo por el cual es menester llevar a cabo esta operación aunque solamente valga 1 crédito. Para validar este ejercicio se requiere una captura de pantalla del menú principal, algo así:

jupyter

Logout

FilesRunningClusters

Select items to perform actions on them.

UploadNew

	Name	Last Modified
	...	seconds ago
	64bit-master	2 months ago
	_MACOSX	a year ago
	aaronMongo	3 months ago
	apache-maven-3.5.3-bin	2 months ago
	BBVAWorkbench	5 months ago
	cloudera-quickstart-vm-5.12.0-0-virtualbox	5 months ago
	codigo_completo	2 months ago
	Compilación Calculadora	5 days ago
	Curso Avanzado	a day ago
	Datio	2 months ago
	DB_VIS_201701	6 months ago
	Downloads	3 months ago

Respuesta:

cloudera-quickstart-vm-5.12.0-0-virtualbox [Corriendo] - Oracle VM VirtualBox

Archivo Máquina Ver Entradas Dispositivos Ayuda

Applications Places System

cloudera

Sin título: Bloc de notas

Archivo Edición Formato Ver Ayuda

Home - Mozilla Firefox

Cloudera Live : Welcom... x Home x

localhost:8889/tree

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

jupyter

QuitLogout

	arturo_losada.txt	3 months ago	19 B
	bashrc	3 months ago	21 B
	cleartables.py	3 months ago	812 B
	cloudera-manager	a year ago	5.39 kB
	cm_api.py	a year ago	9.96 kB
	codegen_columns.java	4 months ago	50.4 kB
	columns.java	4 months ago	50.1 kB
	customers.java	4 months ago	27.6 kB
	departments.java	4 months ago	11.6 kB