

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.- ¿Cuáles 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 -n 1,25p D:/INDRA/Cursos/AVANZADO/aerolinea.csv

```
MINGW64/c/Users/W7/aquinoahaynada/deruiz
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz (master)
$ sed -n 1,25p D:/INDRA/Cursos/AVANZADO/aerolinea.csv
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
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
1987,10,6,2,934,915,1024,1001,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

[3]+ Stopped sed -n 1,25p D:/INDRA/Cursos/AVANZADO/aerolinea.csv
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz (master)
```

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 -n 1,25p D:/INDRA/Cursos/AVANZADO/aerolinea.csv | tee ejercicio_2.txt

```
MINGW64:/c/Users/W7/aquinohaynada/deruiz
$ AC
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz (master)
$ sed -n 1,25p D:/INDRA/Cursos/AVANZADO/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,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
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,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,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,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,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,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,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,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,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,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,NA
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz (master)
$ |
```

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**

```
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ ls
'~$readeruiz.docx'  ejercicio_2.txt  Tareaderuiz.docx
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ mv ejercicio_2.txt ejercicio_3.txt
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ ls
'~$readeruiz.docx'  ejercicio_3.txt  Tareaderuiz.docx
W7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ |
```

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' D:/INDRA/Cursos/AVANZADO/aerolinea.csv | tee ejercicio_4.txt**

```
MINGW64/c/Users/W7/aquinoHaynada/deruiz/TareaVacaciones
w7@XMY2655-INDRA MINGW64 ~/aquinoHaynada/deruiz/TareaVacaciones (master)
$ sed -e :a -e '$q;N;26,$D;ba' D:/INDRA/Cursos/AVANZADO/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,98,103,70,-5,0,ATL,RSW,515,4,24,0,,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

w7@XMY2655-INDRA MINGW64 ~/aquinoHaynada/deruiz/TareaVacaciones (master)
$ ls
'~$readeruiz.docx'  ejercicio_3.txt  ejercicio_4.txt  Tareaderuiz.docx

w7@XMY2655-INDRA MINGW64 ~/aquinoHaynada/deruiz/TareaVacaciones (master)
$ |
```

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`


```
MINGW64/c:/Users/W7/aquinoahaynada/deruiz/TareaVacaciones
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz/TareaVacaciones (master)
$ 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,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
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
1987,10,6,2,934,915,1024,1001,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
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,98,103,70,-5,0,ATL,RSW,515,4,24,0,0,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
2008,12,13,6,612,615,923,907,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
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz/TareaVacaciones (master)
$ wc ejercicio_5.txt
 50   50 5046 ejercicio_5.txt
```

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 ejercicio_5.txt`

```
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz/TareaVacaciones (master)
$ ls -lh ejercicio_5.txt
-rw-r--r-- 1 w7 197121 5,0K jun 20 13:06 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 201808251330 ejercicio_5.txt`

```
w7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ touch -t 201808251330 ejercicio_5.txt

w7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ stat ejercicio_5.txt
  File: ejercicio_5.txt
  Size: 5046          Blocks: 8          IO Block: 65536  regular file
Device: fc452effh/4232392447d Inode: 5629499534628869 Links: 1
Access: (0644/-rw-r--r--)  Uid: (197608/   w7)   Gid: (197121/ UNKNOWN)
Access: 2018-08-25 13:30:00.000000000 -0500
Modify: 2018-08-25 13:30:00.000000000 -0500
Change: 2018-06-20 15:11:38.202255400 -0500
 Birth: 2018-06-20 13:06:19.150214000 -0500
```

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: <http://www.daniloaz.com/es/como-saber-cuantos-procesadores-y-nucleos-tiene-una-maquina-linux/>

```
w7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ grep processor /proc/cpuinfo
processor       : 0
processor       : 1
processor       : 2
processor       : 3

w7@XMY2655-INDRA MINGW64 ~/aquinohaynada/deruiz/TareaVacaciones (master)
$ cat /proc/cpuinfo | grep "model name"
model name     : Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz
model name     : Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz
model name     : Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz
model name     : Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz
```

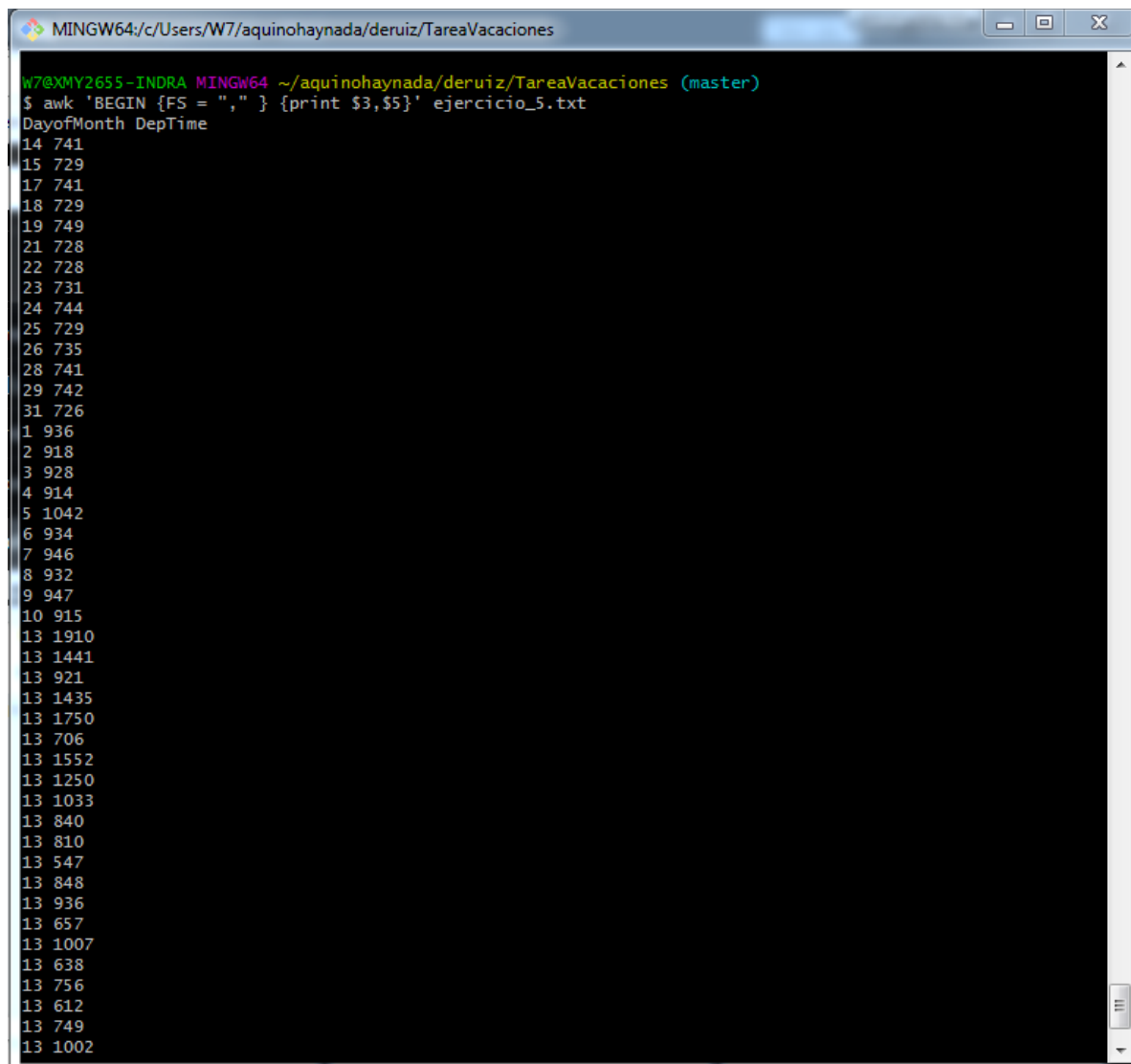
9.- Investigue en qué consiste **awk** y por medio de esa herramienta imprima en pantalla sólo la tercera y quinta columnas (de izquierda a derecha) del archivo **ejercicio_5.txt**. He aquí un ejemplo de cómo se ve el resultado con otro archivo que no tiene que ver con el curso:

```
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: `awk 'BEGIN {FS = "," } {print $3,$5}' ejercicio_5.txt`

<http://francisconi.org/linux/comandos/awk>

http://www.sromero.org/wiki/linux/aplicaciones/uso_de_awk

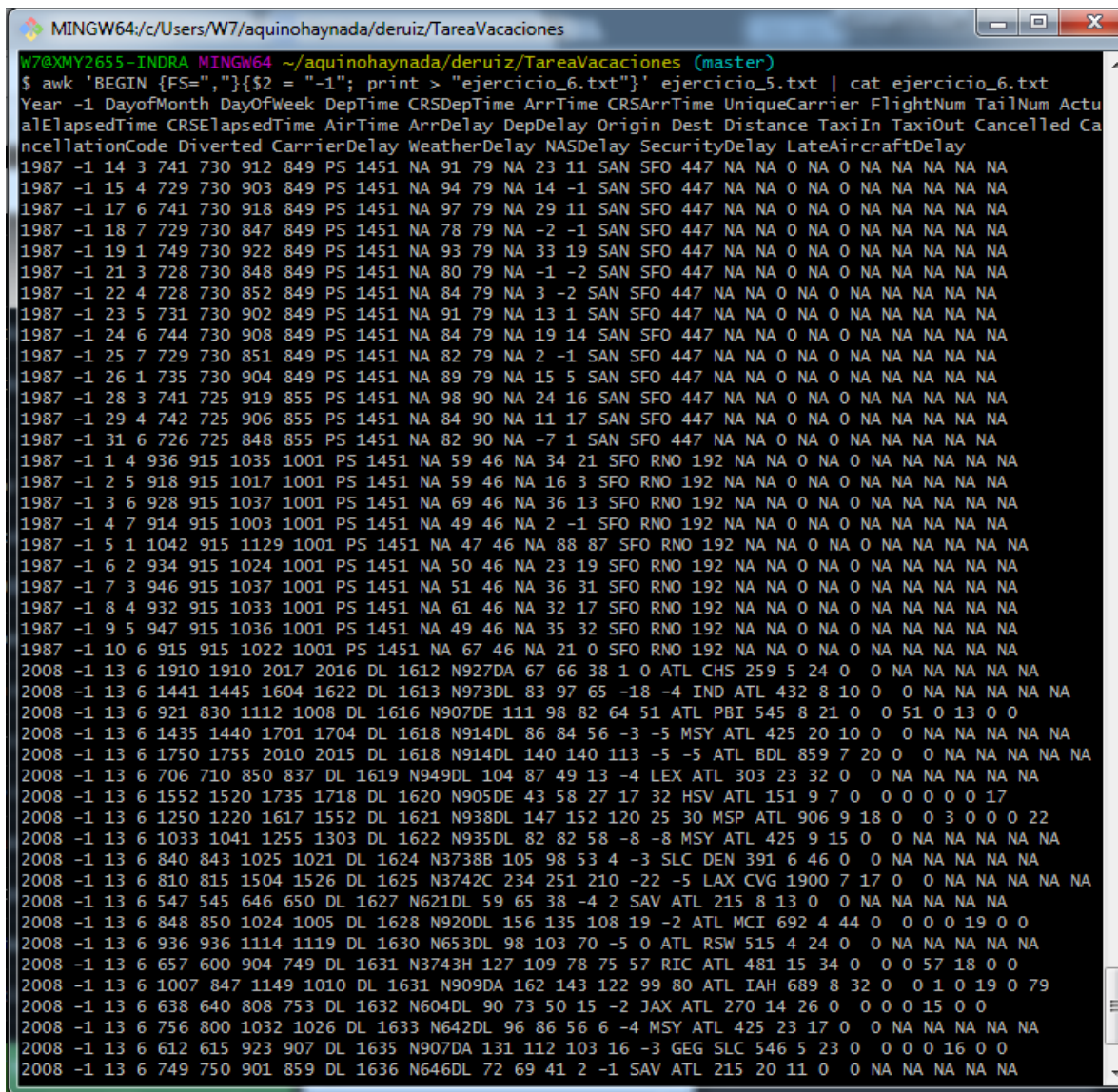


The screenshot shows a Windows command prompt window titled "MINGW64:/c/Users/W7/aquinoHaynada/deruiz/TareaVacaciones". The prompt is "W7@XMY2655-INDRA MINGW64 ~/aquinoHaynada/deruiz/TareaVacaciones (master)". The user has entered the command: `$ awk 'BEGIN {FS = "," } {print $3,$5}' ejercicio_5.txt`. The output is a list of pairs of values, separated by a space, representing the third and fifth fields of each line in the file "ejercicio_5.txt". The output is as follows:

```
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
13 1250
13 1033
13 840
13 810
13 547
13 848
13 936
13 657
13 1007
13 638
13 756
13 612
13 749
13 1002
```


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: `awk 'BEGIN {FS=","}{ $2 = "-1"; print > "ejercicio_6.txt"}' ejercicio_5.txt | cat ejercicio_6.txt`



```
MINGW64/c:/Users/W7/aquinoahaynada/deruiz/TareaVacaciones
w7@XMY2655-INDRA MINGW64 ~/aquinoahaynada/deruiz/TareaVacaciones (master)
$ awk 'BEGIN {FS=","}{ $2 = "-1"; print > "ejercicio_6.txt"}' ejercicio_5.txt | cat ejercicio_6.txt
Year -1 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 -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 3 946 915 1037 1001 PS 1451 NA 51 46 NA 36 31 SFO RNO 192 NA NA 0 NA 0 NA NA NA NA NA
1987 -1 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 -1 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 -1 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
2008 -1 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 -1 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 -1 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 -1 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 -1 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 -1 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 -1 13 6 1552 1520 1735 1718 DL 1620 N905DE 43 58 27 17 32 HSV ATL 151 9 7 0 0 0 0 0 17
2008 -1 13 6 1250 1220 1617 1552 DL 1621 N938DL 147 152 120 25 30 MSP ATL 906 9 18 0 0 3 0 0 22
2008 -1 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 -1 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 -1 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 -1 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 -1 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 -1 13 6 936 936 1114 1119 DL 1630 N653DL 98 103 70 -5 0 ATL RSW 515 4 24 0 0 NA NA NA NA NA
2008 -1 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 -1 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 -1 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 -1 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 -1 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 -1 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
```

SECCION 2. HDFS Y HIVE

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

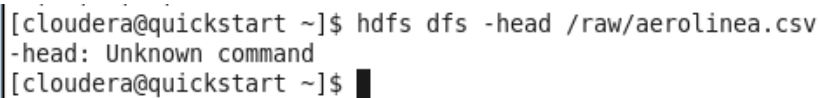
hdfs dfs -head /raw/aerolínea.csv

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.

RESPUESTA: En las paginas que investigué, el comando Head no está disponible para HDFS

<https://hadoop.apache.org/docs/r2.7.1/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html>

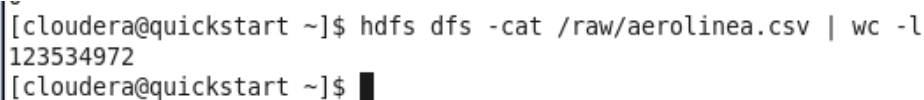
https://www.tutorialspoint.com/es/hadoop/hadoop_command_reference.htm



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

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/aerolíneas.csv | wc -ls

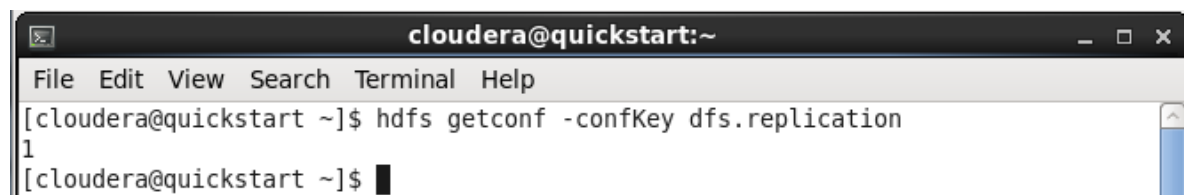


```
[cloudera@quickstart ~]$ hdfs dfs -cat /raw/aerolinea.csv | wc -l
123534972
[cloudera@quickstart ~]$
```

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 getconf -confKey dfs.replication

<https://community.toadworld.com/platforms/nosql/b/weblog/archive/2016/06/14/pri-meros-pasos-con-hdfs-el-sistema-de-archivos-distribuido-de-hadoop>



```
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs getconf -confKey dfs.replication
1
[cloudera@quickstart ~]$
```

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/aerolínea.csv |sed 15q |tee ejercicio_14.txt`

```
[cloudera@quickstart ~]$ hdfs dfs -cat /raw/aerolinea.csv |sed 15q |tee ejercicio_14.txt
cat: Unable to write to output stream.
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
[cloudera@quickstart ~]$
```

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 1:


```
[cloudera@quickstart ~]$ sudo -u hdfs hdfs dfs -mkdir /master
[cloudera@quickstart ~]$ sudo -u hdfs hdfs dfs -mkdir /stagin
[cloudera@quickstart ~]$ sudo -u hdfs hdfs dfs -chown cloudera /master
[cloudera@quickstart ~]$ sudo -u hdfs hdfs dfs -chown cloudera /stagin
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxrwxrwx - hdfs supergroup 0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup 0 2018-06-20 15:59 /hbase
drwxr-xr-x - cloudera supergroup 0 2018-06-22 15:11 /master
drwxr-xr-x - cloudera supergroup 0 2018-06-13 16:35 /raw
drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr
drwxr-xr-x - cloudera supergroup 0 2018-06-22 15:12 /stagin
drwxrwxrwt - hdfs supergroup 0 2018-06-12 16:42 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var
[cloudera@quickstart ~]$
```

RESPUESTA 2: `hdfs dfs -chmod 760 /raw/aerolínea.csv`

```
[cloudera@quickstart ~]$ hdfs dfs -chmod 760 /raw/aerolinea.csv
[cloudera@quickstart ~]$ hdfs dfs -ls /raw
Found 1 items
-rwxrw---- 1 cloudera supergroup 12029208594 2018-06-13 16:35 /raw/aerolinea.csv
[cloudera@quickstart ~]$
```

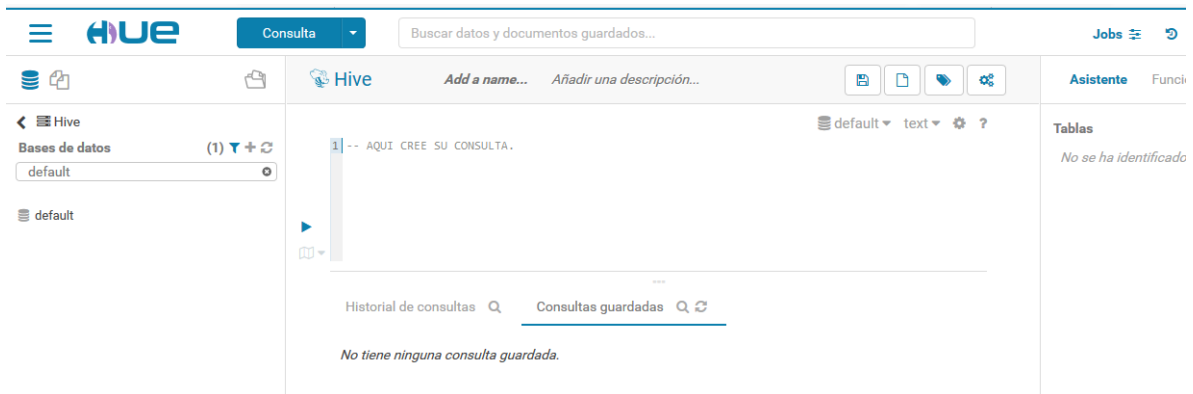
16.- Para los siguientes ejercicios puede hacer uso del servicio Hue (si no ha activado los servicios en Cloudera Manager tiene que hacerlo antes, para entrar a Hue en el mismo navegador se encuentra esta opción).

Aparecerá una ventana como ésta:



The image shows the Hue login interface. At the top is the Hue logo, which consists of a stylized 'H' in blue and purple followed by 'ue' in blue. Below the logo is the tagline 'Query. Explore. Repeat.' in blue. There are two input fields: the first is labeled 'Nombre de usuario' and the second is labeled 'Contraseña'. Below these fields is a blue button labeled 'Iniciar sesión'. At the bottom of the interface, there is a small disclaimer: 'Hue y el logotipo de Hue son marcas comerciales de Cloudera, Inc.'

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 a SQL editor interface. The top toolbar includes a timer at '0s', a database icon, a dropdown menu set to 'default', a 'text' dropdown, and icons for document, settings, and help. The code editor contains the following SQL statement:

```
28 ...  
29 NASDelay STRING,  
30 SecurityDelay STRING,  
31 LateAircraftDelay STRING)  
32 ROW FORMAT DELIMITED  
33 FIELDS TERMINATED BY ','  
34 STORED AS TEXTFILE  
35 location '/raw'  
36 TBLPROPERTIES ("skip.header.line.count"="1")  
37 ;
```

Below the editor, a status bar shows a green checkmark and the text 'Success.'.

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

Query History

Saved Queries

Results (10)

	tabla_aerolinea.year	tabla_aerolinea.month	tabla_aerolinea.dayofmonth	tabla_aerolinea.dayofweek
1	1987	10	14	3
2	1987	10	15	4
3	1987	10	17	6
4	1987	10	18	7
5	1987	10	19	1
6	1987	10	21	3

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 muéstrela.

```
1| DROP TABLE tabla_aerolinea;
```

✓ Success.

```

38
39 SELECT * FROM tabla_aerolinea LIMIT 10
40 ;

```

	tabla_aerolinea.depdelay	tabla_aerolinea.origin	tabla_aerolinea.dest	tabla_aerolinea.distance	tabla_aerolinea.taxiin
1	11	NULL	SFO	447	NA
2	-1	NULL	SFO	447	NA
3	11	NULL	SFO	447	NA
4	-1	NULL	SFO	447	NA
5	19	NULL	SFO	447	NA
6	-2	NULL	SFO	447	NA

RESPUESTA: Crea la tabla, pero no inserta datos en el campo Origin ya que contiene valores de tipo String.

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.

```

24 CancellationCode STRING,
25 Diverted STRING,
26 CarrierDelay STRING,
27 WeatherDelay STRING,
28 NASDelay STRING,
29 SecurityDelay STRING,
30 LateAircraftDelay STRING,
31 ADICIONAL STRING
32 )
33 ROW FORMAT DELIMITED
34 FIELDS TERMINATED BY ','
35 STORED AS TEXTFILE
36 location '/raw'
37 TBLPROPERTIES ("skip.header.line.count"="1")

```

```

41 SELECT * FROM tabla_aerolinea LIMIT 10
42 ;

```

erdelay	tabla_aerolinea.nasdelay	tabla_aerolinea.securitydelay	tabla_aerolinea.lateaircraftdelay	tabla_aerolinea.adicional
NA	NA	NA	NA	NULL
NA	NA	NA	NA	NULL
NA	NA	NA	NA	NULL

RESPUESTA: Crea la tabla y el campo Adicional no se guarda información ya que no tiene ningún campo para llenarlo.

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.

```
1 INSERT INTO tabla_aerolinea VALUES('NA','NA','NA','NA','NA','NA','NA','NA','NA','NA','NA',
2                                     'NA','NA','NA','NA','NA','NA','NA','NA','NA','NA','NA',
3                                     'NA','NA','NA','NA','NA','NA','NA','NA','NA','NA','NA')
✓ Success.
```

SECCIÓN 3. PREGUNTAS ABIERTAS

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

RESPUESTA: Es un tipo de permiso especial que suele asignarse a directorios o ficheros y ayuda a que los usuarios puedan borrar el fichero o directorio

<http://rm-rf.es/permisos-especiales-setuid-setgid-sticky-bit/>

```
[cloudera@quickstart TareaVacaciones]$ chmod o+t ejercicio_5.txt
[cloudera@quickstart TareaVacaciones]$ ls -l
total 1100
-rwxrwxrwx 1 cloudera cloudera 2584 Jun 20 09:20 ejercicio_3.txt
-rwxrwxrwx 1 cloudera cloudera 2462 Jun 20 10:56 ejercicio_4.txt
-rwxrwxrwt 1 cloudera cloudera 5046 Aug 25 2018 ejercicio_5.txt
-rwxrwxrwx 1 cloudera cloudera 5043 Jun 22 09:16 ejercicio_6.txt
-rwxrwxrwx 1 cloudera cloudera 162 Jun 25 08:31 ~$readeruiz.docx
-rwxrwxrwx 1 cloudera cloudera 1094548 Jun 25 08:22 Tareaderuiz.docx
[cloudera@quickstart TareaVacaciones]$
```

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

RESPUESTA. Lenguajes no estructurados, no tiene esquemas entidad-relación. Pueden manejar gran cantidad de datos, y tiene escalabilidad horizontal.

<https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>

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: Este comando nos ayuda a que, cuando se ejecute un proceso o comando, aunque se cierre la terminal, siga en ejecución dicho proceso o comando.

En base a lo anterior, podemos ejecutar el comando para la carga de información que sabemos que tarda mucho o para ejecutar algún programa cuando realicemos el análisis de información.

<http://rm-rf.es/nohup-mantiene-ejecucion-comando-pese-salir-terminal/>

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.   búffers   almac.
Memoria:      3.9G         2.7G         1.2G         16M         66M        1.9G
-/+ buffers/cache:  700M         3.2G
Swap:         4.0G         176K         4.0G
```

Indique el o los valores adecuados y por qué.

RESPUESTA: La memoria libre es 1,2G y 4G para Swap (cuando se está acabando el espacio en disco, utilizan esta que es espacio en disco, pero es lenta)

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: **chmod 700 archivo.txt >> solo le daríamos todos los permisos al archivo al propietario usuario_nuestro, de lectura y escritura a grupo_nuestro.**

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

RESPUESTA:

CLOUDERA	HADOOP
----------	--------

Es una firma especializada en Big Data, que permite añadir funciones a la arquitectura Hadoop de seguridad, control y gestión necesarios para establecer una solución empresarial robusta y fiable.	Hadoop es un ecosistema de componentes de código abierto que cambia fundamentalmente la manera en que las empresas almacenan, procesan y analizan datos.
	Hadoop permite que múltiples tipos de cargas de trabajo y analíticas se ejecuten en los mismos datos al mismo tiempo a gran escala.

<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/387604/m-s-all-de-cloudera-soluciones-de-hadoop-y-big-data>

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

RESPUESTA:

LINUX: http://mural.uv.es/oshuso/824_tipos_de_archivos_en_linux.html

- ordinarios
- Enlaces físicos o duros
- Enlaces simbólicos
- Directorios
- Archivos especiales

WINDOWS: <http://www.windowsfacil.com/manuales/extensiones/extensiones.htm>

- EXTENSION	- TIPO DE ARCHIVO	- PROGRAMA PARA CREARLO O VISIONARLO
.ace	Archivo comprimido	WinAce, Winrar, Izarc
.bmp	Archivo de imagen	Visor de imágenes, Xnview, Acdsee
.doc	Documento de texto	Word, Wordpad, Openoffice.
.exe	Aplicación	Autoejecutable
.gif	Archivo de imagen	Visor de imágenes, Xnview, Acdsee
.jpg	Archivo de imagen	Visor de imágenes, Xnview, Acdsee
.log	Archivo de texto	Bloc de notas, Wordpad
.mp3	Archivo de audio	Reproductores de audio
.ogg	Archivo de audio	Reproductores de audio
.pdf	Documento de texto	Adobe Acrobat Reader
.ppt	Presentación	Power point
.pps	Presentación	Power point
.ra	Archivo de audio	Reproductor Real Player
.rar	Archivo comprimido	Winrar, Izarc
.rtf	Documento de texto	Word, Wordpad, Openoffice
.tif	Archivo de imagen	Visor de imágenes, Xnview, Acdsee
.wav	Archivo de audio	Reproductores de audio
.wma	Archivo de audio	Reproductores de audio, Windows media player
.xls	Hoja de cálculo	Excell, Openoffice
.zip	Archivo comprimido	Winzip, Winrar, Izarc

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

RESPUESTA: SerDe permite indicarle a Hive como debe procesar un registro. SerDe es una combinación de Serializer y Deserializer.

- Deserializer **toma una representación string o binaria y lo convierte a un objeto Java que Hive puede manipular**
- Serializer: **toma un objeto Java y lo convierte en algo que Hive puede escribir a HDFS**

<https://unpocodejava.com/2013/01/24/apache-hive-y-serde/>

No encontré información de SerDe con Impala

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

BIG TABLE: es el servicio de base de datos de Big Data NoSQL de Google, está diseñado para administrar cargas de trabajo colosales con baja latencia y alto rendimiento uniformes

<https://cloud.google.com/bigtable/>

BIG QUERY: Almacén de datos empresariales rápido, económico, escalable y totalmente administrado para analizar datos a cualquier escala.

<https://cloud.google.com/bigquery/>

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

Data Lake: es un repositorio de almacenamiento que contienen una gran cantidad de datos en bruto y que se mantienen allí hasta que sea necesario. A diferencia de un data warehouse jerárquico que almacena datos en ficheros o carpetas, un data lake utiliza una arquitectura plana para almacenar los datos.

Data Warehouse: es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

<https://www.powerdata.es/data-lake>

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

RESPUESTA: SI, <https://www.linuxlinks.com/filesystems/>

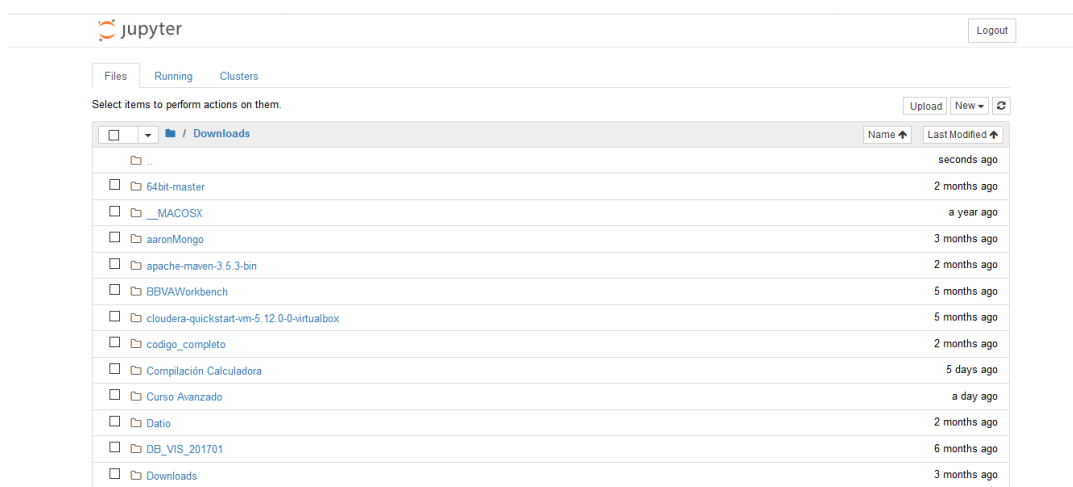
File Systems	
Quantcast File System	High-performance, fault-tolerant, distributed file system
HDFS	Distributed file system providing high-throughput access
Ceph	Unified, distributed storage system
Lustre	File system for computer clusters
GlusterFS	Scale-out NAS file system
PVFS	Designed to scale to petabytes of storage

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:



RESULTADO:

