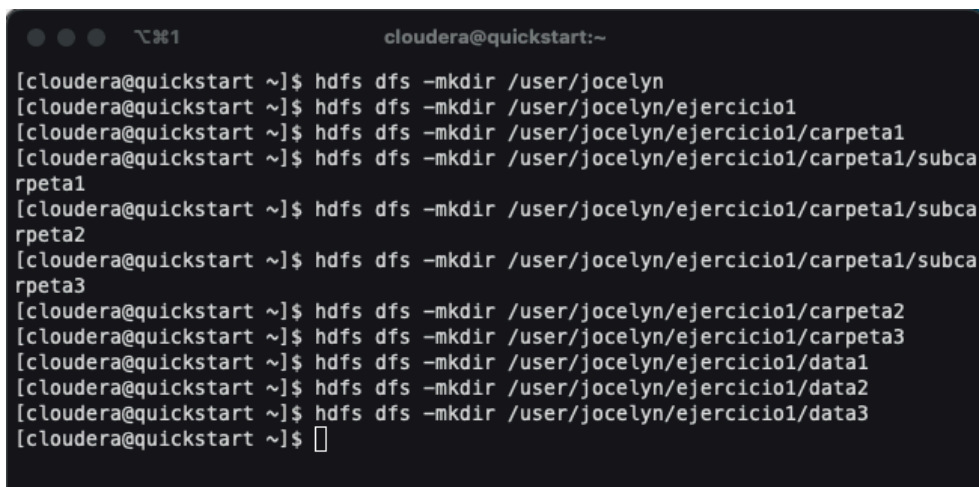


## Script Hadoop

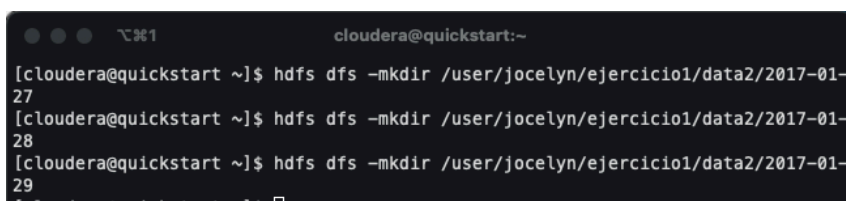
### Creación de carpetas en HDFS

```
hdfs dfs -mkdir /user/jocelyn
hdfs dfs -mkdir /user/jocelyn/ejercicio1
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subcarpetal
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subcarpetal2
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subcarpetal3
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal2
hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal3
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data1
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data3
```

A terminal window with a dark background and light text. The title bar shows three window control buttons and the text 'cloudera@quickstart:~'. The terminal displays a series of 'hdfs dfs -mkdir' commands being executed. The commands create a hierarchy of directories: /user/jocelyn, /user/jocelyn/ejercicio1, /user/jocelyn/ejercicio1/carpetal, /user/jocelyn/ejercicio1/carpetal/subcarpetal, /user/jocelyn/ejercicio1/carpetal/subcarpetal2, /user/jocelyn/ejercicio1/carpetal/subcarpetal3, /user/jocelyn/ejercicio1/carpetal2, /user/jocelyn/ejercicio1/carpetal3, /user/jocelyn/ejercicio1/data1, /user/jocelyn/ejercicio1/data2, and /user/jocelyn/ejercicio1/data3. Each command is followed by a prompt '[cloudera@quickstart ~]\$' and the next command is entered on a new line. The last line shows the prompt followed by a cursor.

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subca
rpetal
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subca
rpetal2
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal/subca
rpetal3
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal2
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/carpetal3
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data1
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data3
[cloudera@quickstart ~]$
```

```
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-27
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-28
hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-29
```

A terminal window with a dark background and light text. The title bar shows three window control buttons and the text 'cloudera@quickstart:~'. The terminal displays three 'hdfs dfs -mkdir' commands being executed. The commands create subdirectories for specific dates: /user/jocelyn/ejercicio1/data2/2017-01-27, /user/jocelyn/ejercicio1/data2/2017-01-28, and /user/jocelyn/ejercicio1/data2/2017-01-29. Each command is followed by a prompt '[cloudera@quickstart ~]\$' and the next command is entered on a new line. The last line shows the prompt followed by a cursor.

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-
27
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-
28
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/jocelyn/ejercicio1/data2/2017-01-
29
[cloudera@quickstart ~]$
```

## Cargar persona.data en HDFS

```
hdfs dfs -put /home/cloudera/datasets/persona.data
/user/jocelyn/ejercicio1/carpetal/subcarpetal
```

```
cloudera@quickstart:~/datasets
empresa.data persona.data
[cloudera@quickstart datasets]$ clear
[cloudera@quickstart datasets]$ hdfs dfs -put /home/cloudera/datasets/persona.data /user/jocelyn/ejercicio1/carpetal/subc
arpetal
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/carpetal/subcarpetal
Found 1 items
-rw-r--r-- 1 cloudera supergroup 7282 2025-02-04 18:41 /user/jocelyn/ejercicio1/carpetal/subcarpetal/persona.data
[cloudera@quickstart datasets]$
```

## Cargar empresa.data en HDFS

```
hdfs dfs -put /home/cloudera/datasets/empresa.data
/user/jocelyn/ejercicio1/carpetal2
```

```
cloudera@quickstart:~/datasets
[cloudera@quickstart datasets]$ hdfs dfs -put /home/cloudera/datasets/empresa.data /user/jocelyn/ejercicio1/carpetal2
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/carpetal2
Found 1 items
-rw-r--r-- 1 cloudera supergroup 105 2025-02-04 18:44 /user/jocelyn/ejercicio1/carpetal2/empresa.data
[cloudera@quickstart datasets]$
```

## Crear el archivo voidfile.csv desde cero en HDFS

```
hdfs dfs -touchz /user/jocelyn/ejercicio1/data1/voidfile.csv
```

```
cloudera@quickstart:~/datasets
[cloudera@quickstart datasets]$ hdfs dfs -touchz /user/jocelyn/ejercicio1/data1/voidfile.csv
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/data1
Found 1 items
-rw-r--r-- 1 cloudera supergroup 0 2025-02-04 18:48 /user/jocelyn/ejercicio1/data1/voidfile.csv
[cloudera@quickstart datasets]$
```

## Cargar el dataset file.csv en HDFS

```
hdfs dfs -put /home/cloudera/datasets/file.csv
/user/jocelyn/ejercicio1/data1
```

```
cloudera@quickstart:~/datasets
[cloudera@quickstart datasets]$ hdfs dfs -put /home/cloudera/datasets/file.csv /user/jocelyn/ejercicio1/data1
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/data1
Found 2 items
-rw-r--r-- 1 cloudera supergroup 70957403 2025-02-04 19:01 /user/jocelyn/ejercicio1/data1/file.csv
-rw-r--r-- 1 cloudera supergroup 0 2025-02-04 18:48 /user/jocelyn/ejercicio1/data1/voidfile.csv
[cloudera@quickstart datasets]$
```

## Cargar archivos al directorio data2

```
hdfs dfs -cp /user/jocelyn/ejercicio1/data1/file.csv
/user/jocelyn/ejercicio1/data2/2017-01-27/file.csv
```

```
hdfs dfs -cp /user/jocelyn/ejercicio1/data1/file.csv
/user/jocelyn/ejercicio1/data2/2017-01-28/file.csv
```

```
cloudera@quickstart:~/datasets
[cloudera@quickstart datasets]$ hdfs dfs -cp /user/jocelyn/ejercicio1/data1/file.csv /user/jocelyn/ejercicio1/data2/2017-01-27/file.csv
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/data2/2017-01-27
Found 1 items
-rw-r--r-- 1 cloudera supergroup 70957403 2025-02-04 19:06 /user/jocelyn/ejercicio1/data2/2017-01-27/file.csv
[cloudera@quickstart datasets]$ hdfs dfs -cp /user/jocelyn/ejercicio1/data1/file.csv /user/jocelyn/ejercicio1/data2/2017-01-28/file.csv
[cloudera@quickstart datasets]$ hdfs dfs -ls /user/jocelyn/ejercicio1/data2/2017-01-28
Found 1 items
-rw-r--r-- 1 cloudera supergroup 70957403 2025-02-04 19:07 /user/jocelyn/ejercicio1/data2/2017-01-28/file.csv
[cloudera@quickstart datasets]$
```

## Recursos:

Dataset de Kaggle:

<https://www.kaggle.com/datasets/thedevastator/uncovering-factors-that-affect-used-car-prices?resource=download>