

## Enunciat:

Les Administracions Públiques, des de l'àmbit municipal a l'internacional, publiquen gran quantitat de dades mitjançant els seus portals de dades obertes. Entre d'altres, tenim a la nostra disposició el portal de dades obertes del Govern de les Illes Balears (<https://dadesobertes.caib.es/>), el del Govern d'Espanya (<https://datos.gob.es>) i el de la Comissió Europea (<https://data.europa.eu/es>). Es tracta d'un sistema jeràrquic, de manera que per exemple en el portal espanyol podem trobar també les dades publicades per les administracions autonòmiques i municipals del país. Al llarg del curs veurem altres repositoris on podem trobar conjunts de dades emprats per la comunitat de ciència de dades i intel·ligència artificial.

Una vegada configurat el clúster Hadoop seguint les instruccions que trobaràs en l'[apartat 6 dels apunts](#), has de descarregar des del node mestre el conjunt de dades de qualitat de l'aire de la comunitat autònoma de Castella i Lleó, que trobaràs al portal espanyol de dades obertes. La descripció del conjunt de dades la tens a <https://datos.gob.es/es/catalogo/a07002862-calidad-del-aire-por-horas>. I l'enllaç directe per descarregar el fitxer CSV és <https://datosabiertos.jcyl.es/web/jcyl/risp/es/medio-ambiente/calidad-aire-historico-horario/1284808467480.csv>

A continuació, has d'executar algunes operacions fent servir les comandes de FS Shell sobre el teu clúster de 3 nodes:

1. Crea un nou directori en HDFS anomenat "bigdataaplicat"
2. Crea dos directoris dins "bigdataaplicat": "lliurament1" i "castellailleo"
3. Puja el fitxer CSV de les dades al directori "bigdataaplicat"
4. Mostra el contingut del directori "bigdataaplicat"
5. Mostra el contingut del fitxer que has pujat
6. Copia el fitxer al directori "castellailleo"
7. Esborra el fitxer que queda al directori "bigdataaplicat"
8. Mou el fitxer des del directori "castellailleo" al directori "lliurament1"

## Solució

és d'haver descarregat el fitxer (1284808467480.csv), les ordres corresponents a cada artat serien les següents.

```
hdfs dfs -mkdir bigdataaplicat
```

```
hdfs dfs -mkdir bigdataaplicat/liurament1  
hdfs dfs -mkdir bigdataaplicat/castellailleo
```

```
hdfs dfs -put 1284808467480.csv bigdataaplicat
```

```
hdfs dfs -ls bigdataaplicat
```

```
hdfs dfs -head bigdataaplicat/1284808467480.csv  
per veure només les darreres files: hdfs dfs -tail bigdataaplicat/1284808467480.csv
```

```
hdfs dfs -cp  
bigdataaplicat/1284808467480.csv bigdataaplicat/castellailleo/1284808467480.csv
```

```
hdfs dfs -rm bigdataaplicat/1284808467480.csv
```

```
hdfs dfs -mv  
bigdataaplicat/castellailleo/1284808467480.csv bigdataaplicat/liurament1/1284808467480.csv
```

una solució única, també són correctes altres variants.