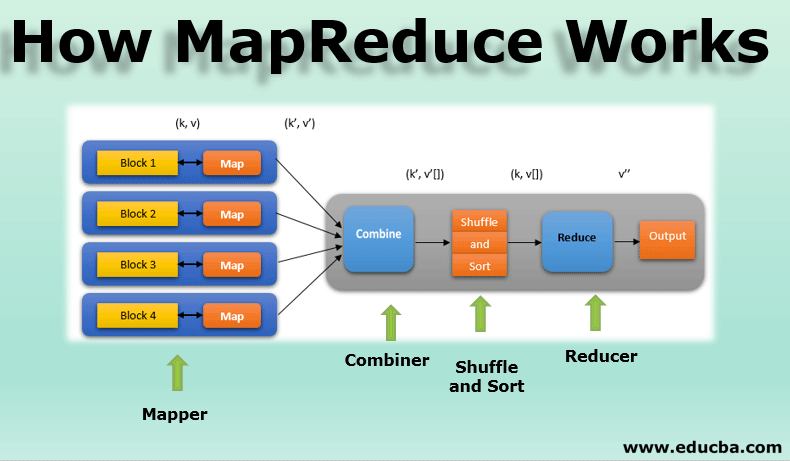
Mapreduce conteo de palabras



Adrián Yared Armas de la Nuez

**Contenido**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

[**1. Actividad 1 2**](#_2c35568x6wg8)

[**1.1 Enunciado 2**](#_rfd33ckl1d7d)

[**1.1 Resolución 2**](#_wh61e1dkm124)

[**1.1.1 Comandos 2**](#_f1a5ng9yugth)

[**2. Actividad 2 3**](#_5obt8tq7rii)

[**1.1 Enunciado 3**](#_ale2sy9kzjjw)

[**1.2 Código 3**](#_kxt8p8jimltb)

[**3. Actividad 3 4**](#_pbcedoclo0co)

[**1.1 Enunciado 4**](#_dutjksyfxypb)

[**1.2 Código 4**](#_32nntq1ko8gb)

[**4. Actividad 4 4**](#_nmfjn2pf63an)

[**4.1 Enunciado 4**](#_fbp8688rb430)

[**4.2 Código 4**](#_ec8mtotue74)

[**4.2.1 Paso 1 5**](#_5zv0qwhy33p1)

[**4.2.2 Paso 2 5**](#_y7f2okot3t28)

[**4.2.3 Paso 3 5**](#_pq0gjol8gabb)

[**4.2.4 Paso 4 5**](#_5nsta1q5qvyg)

[**4.2.5 Paso 5 6**](#_1nke48t5djph)

[**4.2.6 Paso 6 6**](#_je8edzdj9oxk)

[**4.2.7 Paso 7 7**](#_mx4ijerjvhm3)

[**4.2.8 Paso 8 7**](#_mk38nhrovht)

[**4.2.10 Paso 10 7**](#_6ybtecowhlig)

[**4.2.11 Paso 11 7**](#_10axnrrlqcgz)

[**4.2.12 Paso 12 8**](#_8xe1rxl0b1xm)

## 

## **1. Actividad 1**

### **1.1 Enunciado**

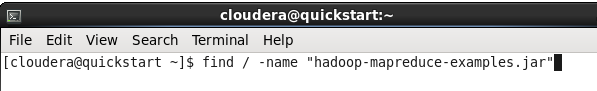
Comprueba la ruta y ejemplos de mapreduce.

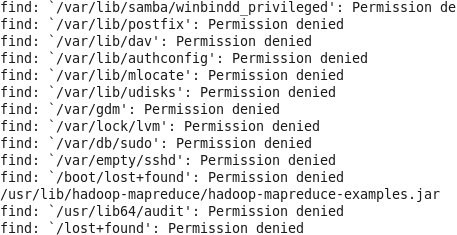
### **1.1 Resolución**

Los ejemplos se encuentran en el archivo hadoop-mapreduce-examples.jar, que generalmente se instala con Hadoop. Para verificar su ubicación: bash

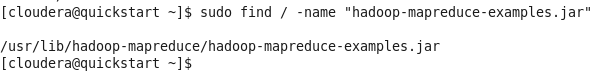
### **1.1.1 Comandos**

**Compruebo el primer comando y da error de permisos:**



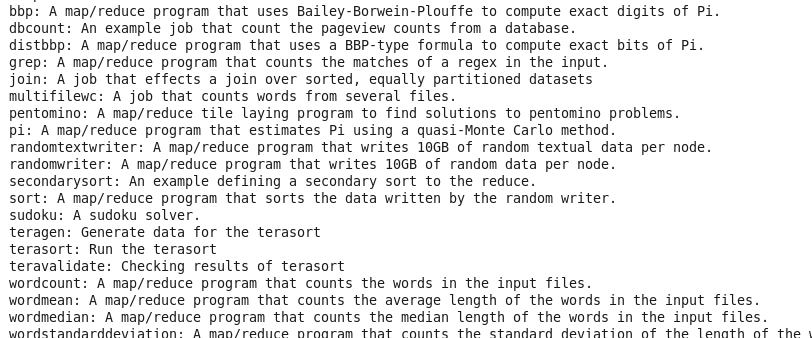


**Así que pruebo el siguiente y me da la ruta** (/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar):



**Consulta los ejemplos disponibles:**



**Me devuelve una lista gigante de ejemplos** (solo muestro algunos de ellos en la captura):  


## **2. Actividad 2**

### **1.1 Enunciado**

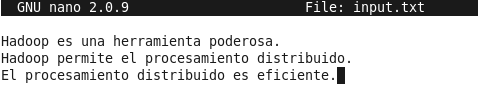
### **1.2 Código**

**Creo el directorio:**



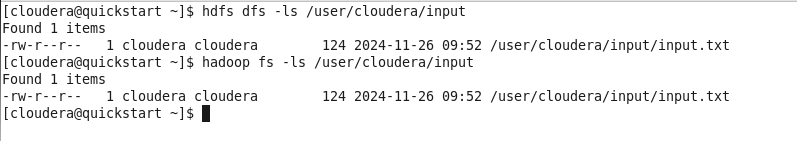
**Creo el archivo input.txt y le meto el texto que pide el enunciado:**





**Compruebo que se ha creado correctamente:**





## **3. Actividad 3**

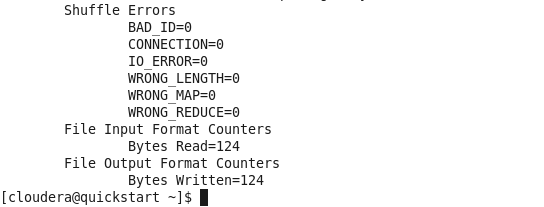
### **1.1 Enunciado**

### **1.2 Código**

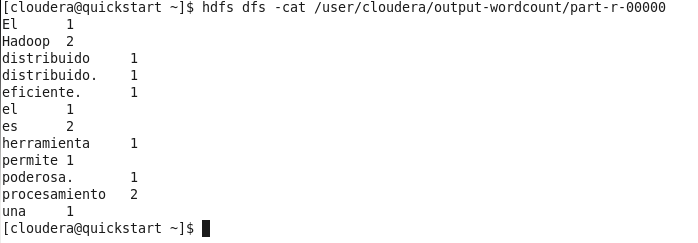
**Ejecuta el comando mapreduce**



**Final del comando:**



**Compruebo que hizo correctamente el wordcount:**



## **4. Actividad 4**

### **4.1 Enunciado**

Repite los pasos anteriores con el Quijote.

### **4.2 Código**

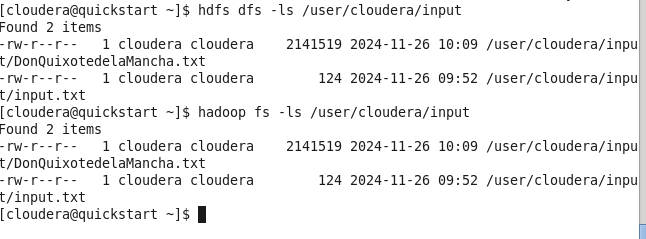
#### **4.2.1 Paso 1**

**Creo el archivo con el Quijote:**



#### **4.2.2 Paso 2**

**Compruebo que está en la ruta:**



#### **4.2.3 Paso 3**

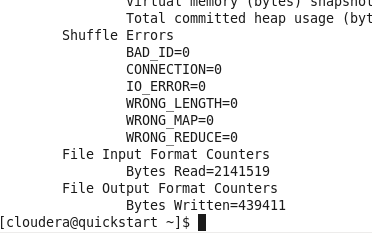
**Elimino el wordcount anterior para poder crear uno nuevo:  
**

#### **4.2.4 Paso 4**

**Ejecuto el wordcount de nuevo:**



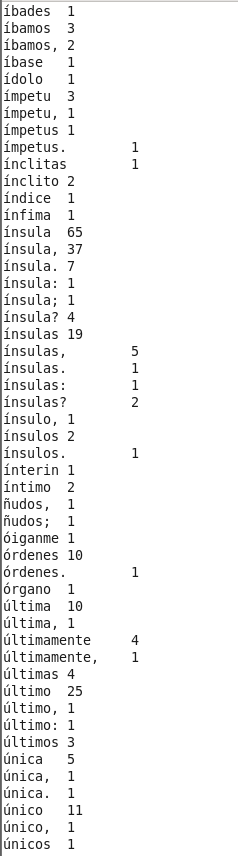
**Final del comando:**



#### **4.2.5 Paso 5**

**Leo el resultado:**  


#### **4.2.6 Paso 6**

**Resultado sin filtrar:**  


#### 

#### **4.2.7 Paso 7**

**Elimino el wordcount de nuevo:**  
ç

#### **4.2.8 Paso 8**

**Creo un filtro de palabras que no me interesan (adverbios, preposiciones y pronombres):**

echo -e

"el\nla\nde\ny\nque\nlos\nlas\nun\nuna\nunos\nunas\ncon\ndel\na\npor\npara\nen\no\nse\nsu\nal\nlo\nmuy\nmas\nmenos\nsiempre\nnunca\naqu

i\nallí\nahora\nentonces\nhoy\nmanana\ntarde\nbien\nmal\nasi\nya\ncasi\ntan\ndemasiado\ntodo\nnada\nalgo\ncomo" > stopwords.txt



**4.2.9 Paso 9**  
**Filtro sin alfanuméricos y paso todo a minúscula:**

hadoop fs -cat /user/cloudera/input/DonQuixotedelaMancha. txt | grep -o \b[A-Z][a-zA-Z]\*\b' | tr "[: upper: ]' "[: lower: ]' | grep -v -w -f stopwords.txt > filtered input.txt



#### **4.2.10 Paso 10**

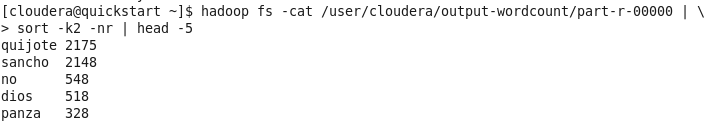
**Lo subo a HDFS:**



#### **4.2.11 Paso 11**

**Creo el wordcount:**  


#### **4.2.12 Paso 12**

**Muestro las 5 palabras más frecuentes:  
**