

1. Verificar los Ejemplos de Hadoop Disponibles en Cloudera

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

bash

```
find / -name "hadoop-mapreduce-examples.jar"
```

si no funciona usar sudo

```
sudo find / -name "hadoop-mapreduce-examples.jar"
```

Usualmente se encuentra en:

bash

```
/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
```

Consulta los ejemplos disponibles:

bash

```
hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
```

1. Esto mostrará una lista de ejemplos como:
 - `wordcount`: Contar palabras.
 - `grep`: Buscar patrones en un archivo.
 - ...

2. Prepara los Datos de Entrada en HDFS

Crea un Directorio de Entrada en HDFS:

bash

```
hdfs dfs -mkdir -p /user/cloudera/input
```

Copia un Archivo de Texto a HDFS: Usa un archivo de texto local como entrada para los ejemplos:

```
nano input.txt
```

```
Hadoop es una herramienta poderosa.  
Hadoop permite el procesamiento distribuido.  
El procesamiento distribuido es eficiente.
```

una vez terminado de copiar el texto, pulsar ctrl+x y le damos a ok

bash

```
hdfs dfs -put /path/to/local/input.txt /user/cloudera/input/
```

Verifica que el archivo esté disponible en HDFS:

bash → los dos comandos hacen lo mismo

```
hdfs dfs -ls /user/cloudera/input
```

```
hadoop fs -ls /user/cloudera/input
```

3. Ejecuta un Ejemplo MapReduce

Ejemplo 1: Contar Palabras con `wordcount`

Ejecuta el Ejemplo:

bash

```
hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar  
wordcount /user/cloudera/input/input.txt  
/user/cloudera/output-wordcount
```

Verifica el Resultado: Después de que el trabajo se complete, los resultados estarán en `/user/cloudera/output-wordcount`.

bash

```
hdfs dfs -cat /user/cloudera/output-wordcount/part-r-00000
```

Ejemplo de Salida:

```
Hadoop      2  
distribuido  2  
el          1  
es          2  
herramienta  1  
procesamiento  2  
una         1
```

Realizar la misma operación con el siguiente archivo:

<https://babel.upm.es/~angel/teaching/pps/quijote.txt>

Presente un documento con las capturas de pantalla de todo el proceso y de los resultados finales. Comente y justifique el resultado obtenido en el segundo archivo e indique las 5 palabras más repetidas, explique qué proceso ha utilizado para obtenerlas.

Realice capturas de pantalla de los procesos MapReduce creados. Pueden acceder a esta información en el gestor de recursos de YARN.