

## **Solución tarea 2-2**

### **1. Instalamos Hadoop**

Instalamos la versión 3.4.2 de Hadoop con el comando !wget

```
✓ Instalamos Hadoop

[1] 3 min
!wget https://archive.apache.org/dist/hadoop/common/hadoop-3.4.2/hadoop-3.4.2.tar.gz

... --2026-01-28 12:06:54-- https://archive.apache.org/dist/hadoop/common/hadoop-3.4.2/hadoop-3.4.2.tar.gz
Resolving archive.apache.org (archive.apache.org)... 65.108.204.189, 2a01:4f9:1a:a084::2
Connecting to archive.apache.org (archive.apache.org)[65.108.204.189]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1065831750 (1016M) [application/x-gzip]
Saving to: 'hadoop-3.4.2.tar.gz'

hadoop-3.4.2.tar.gz 100%[=====>] 1016M 8.61MB/s in 3m 53s

2026-01-28 12:10:48 (4.36 MB/s) - 'hadoop-3.4.2.tar.gz' saved [1065831750/1065831750]
```

### **2. Descomprimos Hadoop**

Descomprimos Hadoop con el comando !tar -xzf

```
✓ Descomprimos Hadoop

[2] 29 s
!tar -xzf hadoop-3.4.2.tar.gz
```

### 3. Configuramos Java

Movemos java a la ruta que queremos !mv

```

[3]
✓ 0s
# Movemos hadoop a la ruta /usr/local
!mv hadoop-3.4.2/ /usr/local/

[4]
✓ 0s
import os

# 1. Definir la ruta de instalación
os.environ["HADOOP_HOME"] = "/usr/local/hadoop-3.4.2"

# 2. Agregar el directorio bin de Hadoop a la variable PATH
os.environ["PATH"] = os.environ["PATH"] + ":" + os.environ["HADOOP_HOME"] + "/bin"

# Buscamos la ruta de Java Home y la configuramos.
java_home = !readlink -f /usr/bin/java | sed "s:bin/java::"
os.environ["JAVA_HOME"] = java_home[0]

print("HADOOP_HOME:", os.environ.get("HADOOP_HOME"))
print("JAVA_HOME:", os.environ.get("JAVA_HOME"))

... HADOOP_HOME: /usr/local/hadoop-3.4.2
    JAVA_HOME: /usr/lib/jvm/java-17-openjdk-amd64/

[5]
✓ 1s
# Comprobamos que hadoop se ha instalado correctamente
%%bash
/usr/local/hadoop-3.4.2/bin/hadoop version

Hadoop 3.4.2
Source code repository https://github.com/apache/hadoop.git -r 84e8b89ee2ebe6923691205b9e171badde7a495c
Compiled by ahmarsu on 2025-08-20T10:30Z
Compiled on platform linux-x86_64
Compiled with protoc 3.23.4
From source with checksum fa94c67d4b4be021b9e9515c9b0f7b6
This command was run using /usr/local/hadoop-3.4.2/share/hadoop/common/hadoop-common-3.4.2.jar

```

### 4. Creamos un directorio files en HDFS

Creamos un directorio nuevo con el comando -mkdir

```

[6]
✓ 2s
%%bash
/usr/local/hadoop-3.4.2/bin/hdfs dfs -mkdir /files

```

## 5. Listamos el contenido del directorio raíz

Para listar el contenido se utiliza el comando `ls`

### ▼ Listamos el contenido del directorio raíz

[7]  
✓ 1s

```
%%bash
```

```
/usr/local/hadoop-3.4.2/bin/hdfs dfs -ls /
```

▼

Found 29 items

|            |   |      |      |        |            |       |                    |
|------------|---|------|------|--------|------------|-------|--------------------|
| -rwxr-xr-x | 1 | root | root | 0      | 2026-01-28 | 12:02 | /.dockerenv        |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-16 | 14:11 | /bin               |
| drwxr-xr-x | - | root | root | 4096   | 2022-04-18 | 10:28 | /boot              |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-28 | 12:15 | /content           |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-26 | 14:11 | /datalab           |
| drwxr-xr-x | - | root | root | 360    | 2026-01-28 | 12:02 | /dev               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-28 | 12:02 | /etc               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-28 | 12:20 | /files             |
| drwxr-xr-x | - | root | root | 4096   | 2022-04-18 | 10:28 | /home              |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-28 | 12:02 | /kaggle            |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-16 | 14:10 | /lib               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:03 | /lib32             |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:10 | /lib64             |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:03 | /libx32            |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:03 | /media             |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:03 | /mnt               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-26 | 14:14 | /opt               |
| dr-xr-xr-x | - | root | root | 0      | 2026-01-28 | 12:02 | /proc              |
| drwxrwxr-x | - | root | root | 4096   | 2026-01-16 | 14:11 | /python-apt        |
| -r-xr-xr-x | 1 | root | root | 346012 | 2000-01-01 | 08:00 | /python-apt.tar.xz |
| drwx-----  | - | root | root | 4096   | 2026-01-28 | 12:02 | /root              |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-16 | 14:07 | /run               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-28 | 12:02 | /sbin              |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-09 | 02:03 | /srv               |
| dr-xr-xr-x | - | root | root | 0      | 2026-01-28 | 12:02 | /sys               |
| drwxrwxrwt | - | root | root | 4096   | 2026-01-28 | 12:16 | /tmp               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-26 | 14:11 | /tools             |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-26 | 14:13 | /usr               |
| drwxr-xr-x | - | root | root | 4096   | 2026-01-26 | 14:11 | /var               |

## 6. Creamos el archivo hoy.txt

### Crear el archivo hoy.txt

```
%%bash
```

```
date > hoy.txt
```

```
whoami >> hoy.txt
```

## 7. Cargamos y mostramos el contenido del archivo

Con el comando put movemos el archivo hoy.txt a /

```

Cargamos y mostramos el contenido del archivo

[9] #hoy.txt está en /hoy.txt dentro de HDFS.
    %%bash
    /usr/local/hadoop-3.4.2/bin/hdfs dfs -put hoy.txt /

[10] #Mostramos el contenido del archivo
     %%bash
     /usr/local/hadoop-3.4.2/bin/hdfs dfs -cat /hoy.txt

Wed Jan 28 12:22:19 PM UTC 2026
root
```

## 8. Copiamos el archivo hoy.txt del origen al directorio files

Con el comando -cp copiamos el archivo

```

Copiamos el archivo hoy.txt del origen al directorio files

[11] %%bash
✓ 1s /usr/local/hadoop-3.4.2/bin/hdfs dfs -cp /hoy.txt /files/

[13] #Verificamos
✓ 1s %%bash
     /usr/local/hadoop-3.4.2/bin/hdfs dfs -ls /files

Found 1 items
-rw-r--r--  1 root root      37 2026-01-28 12:25 /files/hoy.txt
```

## 9. Copiamos el archivo jps.txt desde/hacia el sistema de archivos local a HDFS

```

Copiar el archivo jps.txt desde/hacia el sistema de archivos local a HDFS

[14] # Crear el archivo local
✓ 0s %%bash
     jps > jps.txt

[15] # Copiamos a HDFS
✓ 1s %%bash
     /usr/local/hadoop-3.4.2/bin/hdfs dfs -put jps.txt /
```

## 10. Mover el archivo jps.txt a files

Con el comando -mv movemos los archivos

```
✓ Mover el archivo jps.txt a files

[16] ✓ 1s % bash
/usr/local/hadoop-3.4.2/bin/hdfs dfs -mv /jps.txt /files/
```

## 11. Eliminar el archivo today.txt del directorio principal en HDFS

Eliminamos el archivo today.txt con el comando -rm

```
✓ Eliminar el archivo today.txt del directorio principal en HDFS

[17] ✓ 1s % bash
/usr/local/hadoop-3.4.2/bin/hdfs dfs -rm /hoy.txt

Deleted /hoy.txt
2026-01-28 12:31:49,667 INFO Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
```

## 12. Mostramos las últimas líneas de jps.txt

Mostramos las ultimas líneas con el comando -tail

```
✓ Mostramos las últimas líneas de jps.txt

[18] ✓ 2s % bash
/usr/local/hadoop-3.4.2/bin/hdfs dfs -tail /files/jps.txt

6292 Jps
```