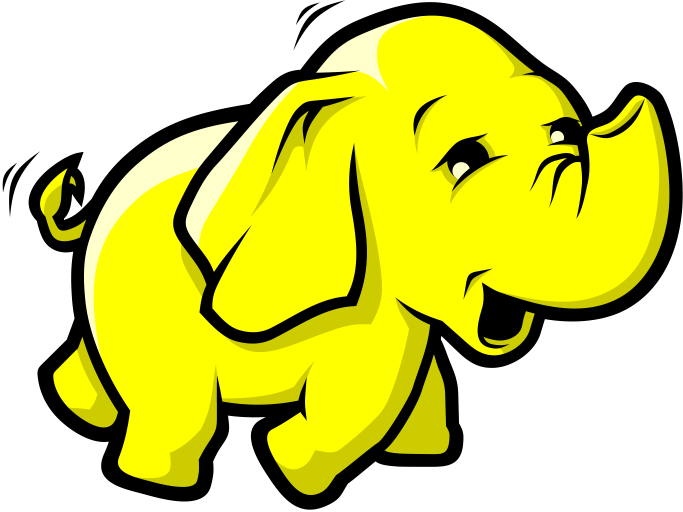
Introducción a Hadoop y HDFS



Adrián Yared Armas de la Nuez

**Contenido**

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

[**1. Instalación de Hadoop 2**](#_2c35568x6wg8)

[**1.1.1 Código 2**](#_ompw8pp0qr0w)

[**1.1.2 Prueba de ejecución 2**](#_jzosc4mhncjf)

[**1.2. Posicionar la distribución 2**](#_mptzybqzhbp3)

[**1.2.1 Código 2**](#_udqoinrqpjw5)

[**1.2.2 Prueba de ejecución 2**](#_bqphemwssy1a)

[**1.2.3 Código 2**](#_aog0dq5jsi4t)

[**1.2.4 Prueba de ejecución 3**](#_a9og31dfg1cb)

[**1.3 Configuración 3**](#_wkpuclrz112y)

[**1.3.1 Código 3**](#_4wrc5ux65mqk)

[**1.3.2 Prueba de ejecución 3**](#_ska88ygzedmf)

[**1.3.3 Código 3**](#_ihh4i9g8uqal)

[**1.3.4 Prueba de ejecución 3**](#_aijx884m2s8w)

[**1.4. Ejemplos 3**](#_qp8jzps3wog5)

[**1.4.1 Volcado de hadoop en xml 4**](#_k7yw9k8u9vm8)

[**1.4.2 Prueba de ejecución 4**](#_ai9yn33d8v3)

[**1.4.3 Ejecución de los ejemplos 4**](#_vv86julrb1iw)

[**1.4.4 Prueba de ejecución 5**](#_mh32jfu345q2)

[**1.5 Copia 5**](#_un9r209pwl5v)

[**1.5.1 Código 5**](#_6equvkrgsxcq)

[**1.5.2 Prueba de ejecución 5**](#_l40jkcuivfhv)

[**2. HDFS 5**](#_drr3con92c7s)

[**2.1 Crear el directorio prueba 5**](#_pscg04xor1ry)

[**2.1.1 Código 5**](#_9nb5ldb8tx9h)

[**2.1.2 Prueba de ejecución 5**](#_3o9x2sonno0)

[**2.2 Crear un fichero local 5**](#_4m8tefftv6sc)

[**2.2.1 Código 5**](#_ysc1yi97vipa)

[**2.2.2 Prueba de ejecución 6**](#_jq7kwxl6yj67)

[**2.2.3 Mostrar su contenido 6**](#_kx32rhuxbqo)

[**2.2.3.1 Código 6**](#_wnup88tbkfq0)

[**2.2.3.2 Prueba de ejecución 6**](#_pyyq743d6ciq)

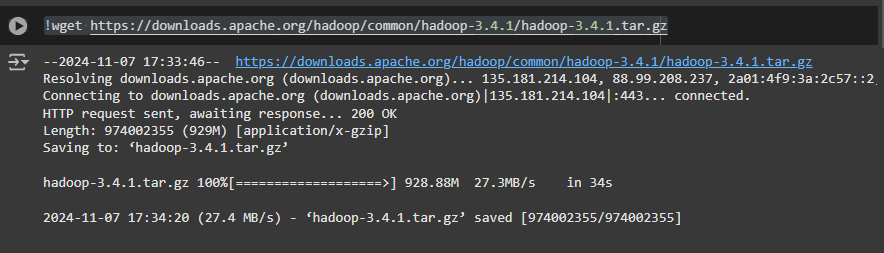
[**3. Colab 7**](#_nig0w532s51z)

# **1. Instalación de Hadoop**

### **1.1.1 Código**

!wget https://downloads.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz

### **1.1.2 Prueba de ejecución**



## **1.2. Posicionar la distribución**

Extraer la distribución descargada en el sistema de archivos de colab y moverla a /usr/local

### **1.2.1 Código**

//Extraer

!tar -xzf hadoop-3.4.1.tar.gz

### **1.2.2 Prueba de ejecución**

### 

### **1.2.3 Código**

//Mover a /usr/local/

!mv hadoop-3.4.1/ /usr/local/

### **1.2.4 Prueba de ejecución**

### 

## **1.3 Configuración**

Actualización variables de entorno (JAVA\_HOME, PATH)

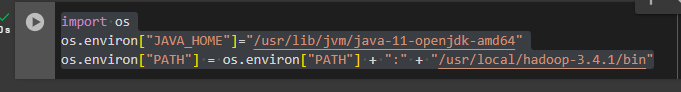
### **1.3.1 Código**

import os

os.environ["JAVA\_HOME"]="/usr/lib/jvm/java-11-openjdk-amd64"

os.environ["PATH"] = os.environ["PATH"] + ":" + "/usr/local/hadoop-3.4.1/bin"

### **1.3.2 Prueba de ejecución**

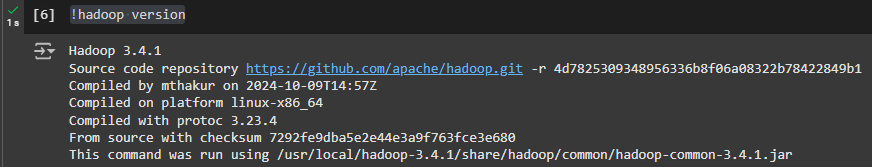


### **1.3.3 Código**

//Comprobación de la instalación

!hadoop version

### **1.3.4 Prueba de ejecución**



## **1.4. Ejemplos**

Una de las formas tradicionales de asegurarnos que un ambiente de Hadoop recién instalado funciona correctamente, es ejecutando el jar de ejemplos map-reduce incluido con toda instalación de hadoop (hadoop-mapreduce-examples.jar).

[Hadoop Map Reduce Examples](https://svn.apache.org/viewvc/hadoop/common/trunk/hadoop-mapreduce-project/hadoop-mapreduce-examples/src/main/java/org/apache/hadoop/examples/)

### **1.4.1 Volcado de hadoop en xml**

//Código:

%%bash

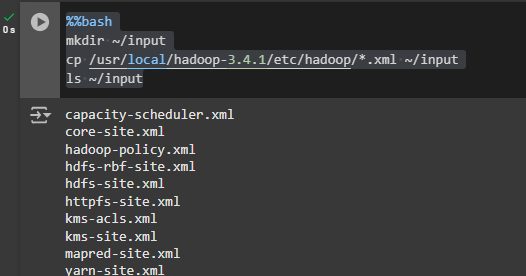
mkdir ~/input

cp /usr/local/hadoop-3.4.1/etc/hadoop/\*.xml ~/input

ls ~/input

### 

### **1.4.2 Prueba de ejecución**



### 

### **1.4.3 Ejecución de los ejemplos**

Ejecutamos hadoop jar con el fin de ejecutar uno de los ejemplos por defecto, en este caso el grep que busca expresiones regulares dentro de los ficheros que le especifiquemos.

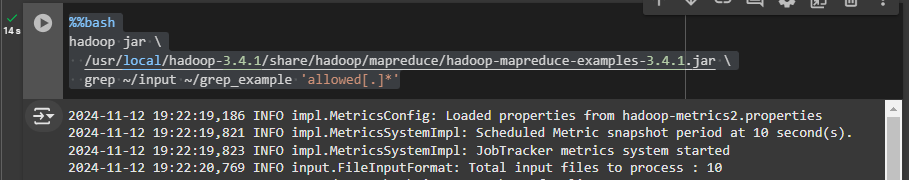
//Código:  
%%bash

hadoop jar \

/usr/local/hadoop-3.4.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.1.jar \

grep ~/input ~/grep\_example 'allowed[.]\*'

### **1.4.4 Prueba de ejecución**

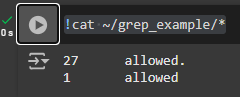


## **1.5 Copia**

### **1.5.1 Código**

!cat ~/grep\_example/\*

### **1.5.2 Prueba de ejecución**



# **2. HDFS**

## **2.1 Crear el directorio prueba**

### **2.1.1 Código**

!hdfs dfs -mkdir prueba

### **2.1.2 Prueba de ejecución**



## **2.2 Crear un fichero local**

### **2.2.1 Código**

%%bash

echo "Ejemplo de HDFS" > user.txt

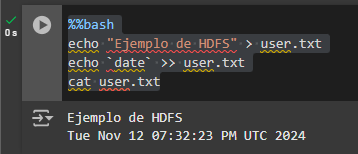
echo `date` >> user.txt

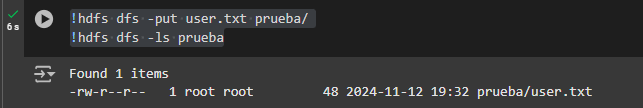
cat user.txt

!hdfs dfs -put user.txt prueba/

!hdfs dfs -ls prueba

### **2.2.2 Prueba de ejecución**





### **2.2.3 Mostrar su contenido**

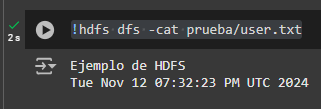
#### **2.2.3.1 Código**

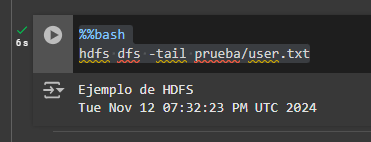
!hdfs dfs -cat prueba/user.txt

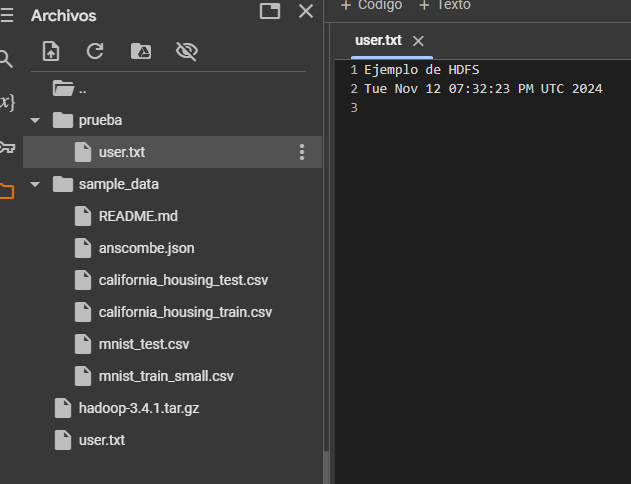
%%bash

hdfs dfs -tail prueba/user.txt

#### **2.2.3.2 Prueba de ejecución**







## **3. Colab**

[](https://colab.research.google.com/drive/18-WvUQtNLuEwfUdKoDMeus0-ggq6qMZF?usp=sharing)