|  |
| --- |
| Máster en Big Data |
| Tecnologías de Almacenamiento |
| 2. Hands-On: Uso de HDFS |

Índice

[1. Introducción 3](#_Toc96241371)

[2. Desplegar el nodo de Hadoop 3](#_Toc96241372)

[3. Uso de HDFS 4](#_Toc96241373)

[3.1. Explorar HDFS 4](#_Toc96241374)

[3.2. Insertar Archivos 4](#_Toc96241375)

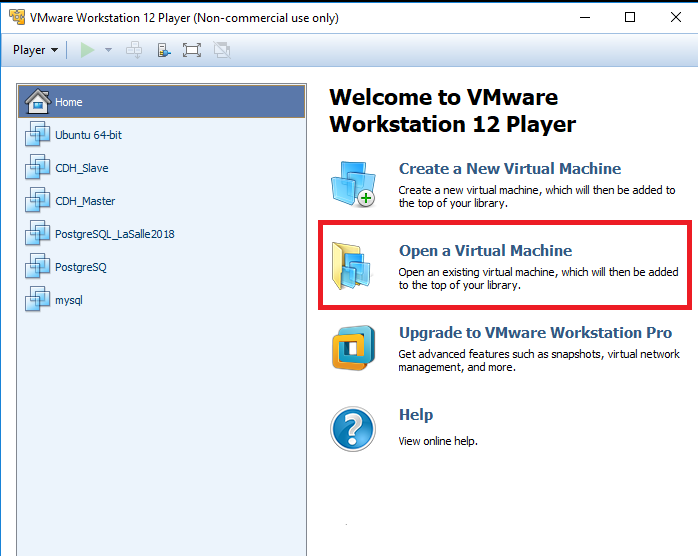
[3.3. Manipular Archivos 4](#_Toc96241376)

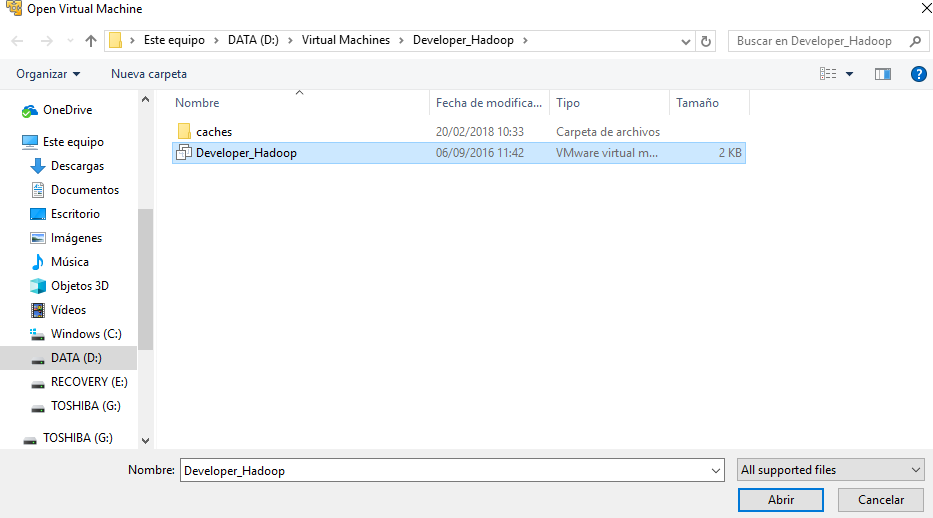
# Introducción

El objetivo de este Hands-On es ganar confianza con el uso de HDFS por línea de comandos.

# Desplegar el nodo de Hadoop

Junto con este enunciado se proporciona una instancia de clúster Hadoop pseudodistribuido. Descomprimir la carpeta e importar la máquina a VMWare como ya se ha hecho anteriormente.





# Uso de HDFS

Todas las instrucciones deben introducirse mediante Shell de Linux.

IMPORTANTE: Adjuntar el comando y captura de que los ficheros afectados por el comando se han creado/modificado/eliminado correctamente.

## Explorar HDFS

1. Muestra el comando de ayuda de Hadoop
2. Listar el contenido de la raíz
3. Listar el contenido del directorio */user*

## Insertar Archivos

1. En el directorio */home/training/training\_materials/developer/data* del filesystem local descomprimir el archivo s*hakespeare.tar.gz* e insertar en HDFS el contenido de la carpeta descomprimida en */user/training/shakespeare*
2. Listar el directorio donde se ha realizado la importación
3. Crear un directorio llamado *weblog* en HDFS en la ruta siguiente: */user/training*
4. Importar el archivo *access\_log.gz*. (gunzip –c descomprimirá el archivo y volcará todo el contenido a la salida estándar y – en hadoop leerá los datos de la salida estándar)

## Manipular Archivos

1. Listar el contenido de la carpeta *shakespeare*
2. Borra el archivo *glossary*
3. Muestra las últimas 50 líneas del archivo *histories*
4. Descarga el archivo *poems* en el filesystem local