Máster en Big Data

Tecnologías de Almacenamiento

3. Hands-On: Uso de HDFS

Àlex Balló Vergés

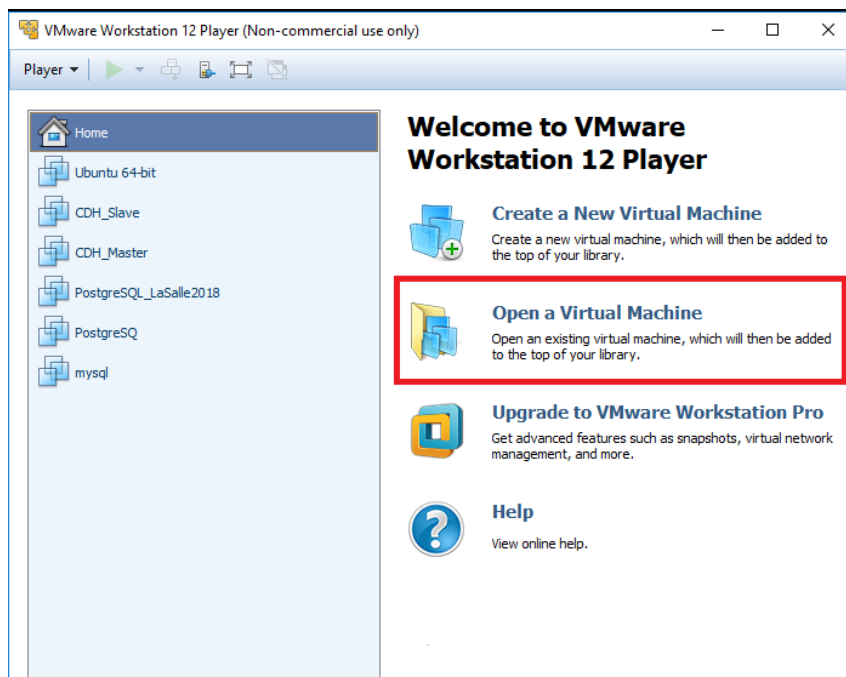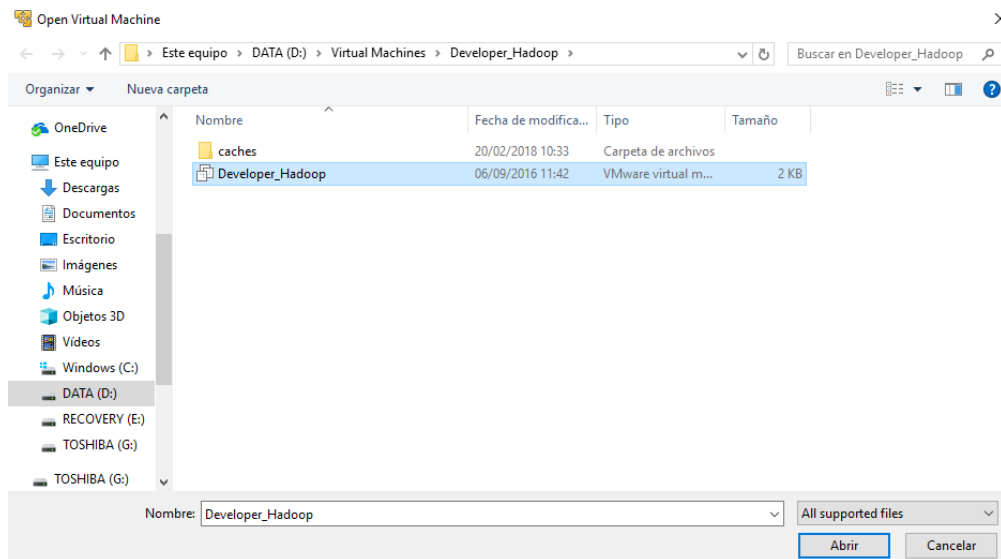2018-2019

Índice

1. Introducción

El objetivo de este Hands-On es ganar confianza con el uso de HDFS por línea de comandos. Rellena los apartados con las instrucciones pertinentes

2. 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.

3. Uso de HDFS

Todas las instrucciones deben introducirse mediante Shell de linux

3.1. Explorar HDFS

a) Muestra el comando de ayuda de hadoop

```
[training@localhost ~]$ hadoop fs -help
Usage: hadoop fs [generic options]
        [-cat [-ignoreCrc] <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal <localsrc> ... <dst>]
        [-copyToLocal [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-count [-q] <path> ...]
        [-cp <src> ... <dst>]
        [-df [-h] [<path> ...]]
        [-du [-s] [-h] <path> ...]
        [-expunge]
        [-get [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-getmerge [-nl] <src> <localdst>]
        [-help [cmd ...]]
        [-ls [-d] [-h] [-R] [<path> ...]]
        [-mkdir [-p] <path> ...]
        [-moveFromLocal <localsrc> ... <dst>]
        [-moveToLocal <src> <localdst>]
        [-mv <src> ... <dst>]
        [-put <localsrc> ... <dst>]
        [-rm [-f] [-r|-R] [-skipTrash] <src> ...]
        [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
```

b) Listar el contenido de la raíz

```
[training@localhost ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hbase supergroup          0 2019-03-11 15:50 /hbase
drwxrwxrwt   - hdfs  supergroup          0 2016-09-05 12:00 /tmp
drwxrwxrwx   - hue   supergroup          0 2016-09-05 12:01 /user
drwxr-xr-x   - hdfs  supergroup          0 2016-09-05 12:01 /var
```

c) Listar el contenido del directorio /user

```
[training@localhost ~]$ hadoop fs -ls /user
Found 3 items
drwxr-xr-x   - hue       supergroup          0 2016-09-05 12:01 /user/hive
drwxr-xr-x   - hdfs      supergroup          0 2016-09-05 12:01 /user/hue
drwxr-xr-x   - training supergroup           0 2016-09-05 12:01 /user/training
```

3.2. Insertar Archivos

a) En el directorio home/training/training_materials/developer/data del filesystem local descomprimir e insertar en HDFS el archivo Shakespeare.tar.gz

Descomprimir:

```
[training@localhost data]$ tar -xzvf shakespeare.tar.gz
shakespeare/
shakespeare/comedies
shakespeare/glossary
shakespeare/histories
shakespeare/poems
shakespeare/tragedies
[training@localhost data]$ ls
access_log.gz           movielens.readme       shakespeare
auctiondata.csv         movielens-small.sql    shakespeare-stream.tar.gz
bible.tar.gz            movielens.sql          shakespeare.tar.gz
invertedIndexInput.tgz  nameyeartestdata
```

Insertar en HDFS:

```
[training@localhost data]$ hadoop fs -put shakespeare /shakespeare
```

b) Listar el directorio donde se ha realizado la importación

```
[training@localhost data]$ hadoop fs -ls /
Found 5 items
drwxr-xr-x   - hbase     supergroup          0 2019-03-11 15:50 /hbase
drwxr-xr-x   - training supergroup           0 2019-03-11 16:04 /shakespeare
drwxrwxrwt   - hdfs      supergroup          0 2016-09-05 12:00 /tmp
drwxrwxrwx   - hue       supergroup          0 2016-09-05 12:01 /user
drwxr-xr-x   - hdfs      supergroup          0 2016-09-05 12:01 /var
```

c)
d) Crear el directorio weblog

e)
```
[training@localhost data]$ hadoop fs -mkdir /weblog
[training@localhost data]$ hadoop fs -ls /
Found 6 items
drwxr-xr-x   - hbase     supergroup          0 2019-03-11 15:50 /hbase
drwxr-xr-x   - training supergroup           0 2019-03-11 16:04 /shakespeare
drwxrwxrwt   - hdfs      supergroup          0 2016-09-05 12:00 /tmp
drwxrwxrwx   - hue       supergroup          0 2016-09-05 12:01 /user
drwxr-xr-x   - hdfs      supergroup          0 2016-09-05 12:01 /var
drwxr-xr-x   - training supergroup           0 2019-03-11 16:10 /weblog
```

f) 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)*

```
[training@localhost data]$ gzip -d access_log.gz

[training@localhost data]$ hadoop fs -put access_log /weblog
[training@localhost data]$ hadoop fs -ls /weblog
Found 1 items
-rw-r--r--   1 training supergroup  504941532 2019-03-11 16:24 /weblog/access_lo
g
```

3.3. Manipular Archivos

a) Listar el contenido de la carpeta Shakespeare
```
[training@localhost data]$ hadoop fs -ls /shakespeare
Found 5 items
-rw-r--r--   1 training supergroup    1784616 2019-03-11 16:04 /shakespeare/come
dies
-rw-r--r--   1 training supergroup      58976 2019-03-11 16:04 /shakespeare/glos
sary
-rw-r--r--   1 training supergroup    1479035 2019-03-11 16:04 /shakespeare/hist
ories
-rw-r--r--   1 training supergroup     268140 2019-03-11 16:04 /shakespeare/poem
s
-rw-r--r--   1 training supergroup    1752440 2019-03-11 16:04 /shakespeare/trag
edies
```
b) Borra el archivo glossary
```
[training@localhost data]$ hadoop fs -rm -r /shakespeare/glossary
Deleted /shakespeare/glossary
[training@localhost data]$ hadoop fs -ls /shakespeare
Found 4 items
-rw-r--r--   1 training supergroup    1784616 2019-03-11 16:04 /shakespeare/come
dies
-rw-r--r--   1 training supergroup    1479035 2019-03-11 16:04 /shakespeare/hist
ories
-rw-r--r--   1 training supergroup     268140 2019-03-11 16:04 /shakespeare/poem
s
-rw-r--r--   1 training supergroup    1752440 2019-03-11 16:04 /shakespeare/trag
edies
```

c) Muestra las últimas 50 líneas del archivo histories

```
[training@localhost data]$ hadoop fs -tail /shakespeare/histories
 rose and the red:
        Smile heaven upon this fair conjunction,
        That long have frown'd upon their enmity!
        What traitor hears me, and says not amen?
        England hath long been mad, and scarr'd herself;
        The brother blindly shed the brother's blood,
        The father rashly slaughter'd his own son,
        The son, compell'd, been butcher to the sire:
        All this divided York and Lancaster,
        Divided in their dire division,
        O, now, let Richmond and Elizabeth,
        The true succeeders of each royal house,
        By God's fair ordinance conjoin together!
        And let their heirs, God, if thy will be so.
        Enrich the time to come with smooth-faced peace,
```

d)

e) Descarga el archivo poems en el filesystem local

```
[training@localhost data]$ hadoop fs -get /shakespeare/poems
[training@localhost data]$ ls
access_log              movielens.readme        poems
auctiondata.csv         movielens-small.sql     shakespeare
bible.tar.gz            movielens.sql           shakespeare-stream.tar.gz
invertedIndexInput.tgz  nameyeartestdata        shakespeare.tar.gz
```