

Prácticas BigData

• ZooKeeper

Instalación y configuración

- Descargamos ZooKeeper de la página zookeeper.apache.org
- Lo descomprimos en /opt/hadoop

```
tar xvf /home/hadoop/Descargas/zookeeperXXXXX
```

- Lo cambiamos de nombre para manejarlo de forma más sencilla

```
mv zookeeperXXX zoo
```

- Lo copiamos al nodo2 y nodo3

```
scp -r zoo nodo2:/opt/hadoop
scp -r zoo nodo3:/opt/hadoop
```

- Configuramos el fichero /home/hadoop/.bashrc para incluir las líneas de ZooKeeper. Y lo debemos copiar a el resto de nodos donde vamos a tener funcionando ZooKeeper

- Debería quedar algo parecido a lo siguiente:

```
export HADOOP_HOME=/opt/hadoop
export JAVA_HOME=/usr/java/jdk1.8.0_151
export HIVE_HOME=/opt/hadoop/hive
export SQOOP_HOME=/opt/hadoop/sqoop
export ZOOKEEPER_HOME=/opt/hadoop/zoo
export
PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$SQOOP_HOME/bin:$H
IVE_HOME/bin:$ZOOKEEPER_HOME/bin
```

- Si entramos de nuevo en un terminal tendremos ya cargadas las variables.
- Nos situamos en el directorio de configuración

```
cd /opt/hadoop/zoo/conf
```

- Copiamos el fichero zoo_sample.cfg como zoo.cfg

```
cp zoo_sample.cfg zoo.cfg
```

- Creamos el siguiente contenido dentro del fichero

```
# the directory where the snapshot is stored.
# do not use /tmp for storage, /tmp here is just
# example sakes.
dataDir=/datos/zoo
# the port at which the clients will connect
clientPort=2181
```

```
# the maximum number of client connections.
# increase this if you need to handle more clients
#maxClientCnxns=60
#
# Be sure to read the maintenance section of the
# administrator guide before turning on autopurge.
#
# http://zookeeper.apache.org/doc/current/zookeeperAdmin.html#sc_maintenance
#
# The number of snapshots to retain in dataDir
#autopurge.snapRetainCount=3
# Purge task interval in hours
# Set to "0" to disable auto purge feature
#autopurge.purgeInterval=1
server.1=nodo1:2888:3888
server.2=nodo2:2888:3888
server.3=nodo3:2888:3888
```

- Creamos en los 3 nodos el directorio de trabajo

```
mkdir /datos/zoo
```

- Creamos en el directorio un fichero denominado "myid" que tiene que contener el número de servidor dentro de ZooKeeper.
- Por ejemplo en el nodo1:

```
echo 1 > /datos/zoo/myid
```

- Y en el nodo2 y el nodo3 ponemos 2 y 3.
- Ejecutamos el siguiente comando en los 3 nodos

```
zkCli.sh start
```

- Comprobamos con "jps" que tenemos el proceso QuorumPeerMain funcionando

```
ps
20912 ResourceManager
17315 RunJar
20456 NameNode
16905 RunJar
22377 Jps
20699 SecondaryNameNode
26475 QuorumPeerMain
```

- Podemos preguntar el estado y si es leader o follower

zkServer.sh status

ZooKeeper JMX enabled by default

Using config: /opt/hadoop/zoo/bin/../conf/zoo.cfg

Mode: follower

-

-