Laboratorio 24/01/2020

1. Crear red virtual en VBOX

2. Cambiar la configuracion de red para que se conecte a la red virtual

3. Cambiar los recursos del server a 1 cpu y 2 gb de RAM

4. Clonar la VM

5. Arrancar las 2 VMs en modo Headless

6. Revisar las IP asignadas a cada hostname

$ ip addr show

master -> 10.0.2.15

slave -> 10.0.2.4

Laboratorio 2: HDFS Master Slave

2.1. Agregar los nombres al archivo de hosts

$ vim /etc/hosts

10.0.2.15 master

10.0.2.4 slave

2.2. Reiniciar los nodos

$ reboot

2.3. Revisar que el Firewall este deshabilitado

$ systemctl status firewalld

EJ.

● firewalld.service - firewalld - dynamic firewall daemon

Loaded: loaded (/usr/lib/systemd/system/firewalld.service; disabled; vendor preset: enabled)

Active: inactive (dead)

Docs: man:firewalld(1)

2.4. Crear las nuevas llaves para cada server y registrarlas

$ cd .ssh

$ rm authorized\_keys

$ rm id\_rsa

$ rm id\_rsa.pub

$ ssh-keygen -t rsa

$ systemctl restart sshd

2.5. Enviar las llaves a los otros hosts

$ ssh-copy-id root@master

$ ssh-copy-id root@slave

2.6. Verificar la configuracion de Java y Hadoop

$ java -version

$ bin/hadoop version

2.7. Crear los archivos de configuracion

$ vim etc/hadoop/masters (master / slave)

master

$ vim etc/hadoop/workers (master)

master

slave

$ vim etc/hadoop/workers (slave)

slave

2.8. Actualizar la configuracion del CORE (master / slave)

$ vim etc/hadoop/core-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://master:9000</value>

</property>

</configuration>

2.9. Actualizar la configuracion del HDFS (Master)

$ vim etc/hadoop/hdfs-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>/root/hadoop-3.2.1/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>/root/hadoop-3.2.1/datanode</value>

</property>

</configuration>

2.10. Actualizar la configuracion del HDFS (Slave)

$ vim etc/hadoop/hdfs-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>/root/hadoop-3.2.1/datanode</value>

</property>

</configuration>

2.11. Actualizar la configuracion del MapReduce (master / slave)

$ vim etc/hadoop/marped-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

2.12. Actualizar la configuracion del YARN (master / slave)

$ vim etc/hadoop/yarn-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>master</value>

</property>

</configuration>

2.13. Formatear el namenode (master)

$ bin/hadoop namenode -format

2.14. Iniciar los servicios

$ sbin/start-all.sh (master)

$ sbin/start-dfs.sh (slave)

2.15. Verificar que se hayan levantado los servicios

$ ps -fea | grep java

2.16. Verificar el estado de los nodos desde un navegador

master:9870

master:8088