Với 4 máy, chúng ta sẽ tạo ra 1 máy chủ(master) với 3 máy con(slave) để phục vụ nhu cầu tính toán phân tán.

Yêu cầu:

HĐH: Linux

Cài đặt: Java 8, SSH

Các cài đặt thêm: Hadoop, Zookeeper, Kafka, Hbase

Cách cấu hình:

**1, Hadoop**

1. Lấy ip các máy tính trong hệ thống
2. Tắt firewall
3. Mở file host và thêm các ip tương ứng với máy con, máy chủ
4. Khởi động lại dịch vụ sshd
5. Tạo khóa ssh ở máy chủ
6. Sao chép khóa được tạo ra để làm khóa ủy quyền cho máy chủ
7. Sao chép khóa được tạo ra để làm khóa ủy quyền cho máy con
8. Thêm đường dẫn của Hadoop và Java 8 vào file .bashrc ở tất cả các máy
9. Tạo và file masters ở tất cả các máy, nội dung “masters”
10. Với máy chủ(master), tạo file slaves với nội dung “master slave”
11. Với máy con(slave), tạo file slaves với nội dung “ slave”
12. Sửa file core-site.xml ở tất cả các máy, với nội dung:

<?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>

1. Sửa file hdfs-site ở máy chủ như sau:

<?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>/home/edureka/hadoop-2.7.3/namenode</value>

</property>

<property>

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

<value>/home/edureka/hadoop-2.7.3/datanode</value>

</property>

</configuration>

1. Sửa file hdfs-site ở máy con như sau:

<?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>/home/edureka/hadoop-2.7.3/datanode</value>

</property>

</configuration>

1. Sao chép file mapred-site.xml ở tất cả các máy từ mẫu của thư mục cấu hình và sửa với nội dung như sau:

<?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>

1. Sửa file yarn-site.xml ở tất cả các máy như sau:

<?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>

</configuration>

Cấu hình thành công.

**2, Kafka và Zookeeper**

* Chỉnh sửa file “zookeeper.properties” trên tất cả các máy(với 1 máy chủ và 3 máy con) như sau:

vi ~/kafka/config/zookeeper.properties

clientPort=2080 #Changing the Port from default "2181" to "2080"

server.1=master:2888:3888

server.2=slave1:2888:3888

server.3=slave2:2888:3888

server.4=slave3:2888:3888

#add here more servers if you want

initLimit=5

syncLimit=2

* Chỉnh sửa file "server.properties" trên tất cả các máy và thêm vào như sau:

vi ~/kafka/config/server.properties

broker.id=1

port=9092

host.name=master

zookeeper.connect=master:2080,slave1:2080,slave2:2080,slave3:2080

#Add all the Instances here for each Instance

* Tạo thư mục zookeeper trong /tmp

Các máy đã được cấu hình xong

**3, Hbase**

* Chỉnh file hbase-site.xml như sau:
* Ở máy chủ

<configuration>

<property>

<name>hbase.rootdir</name>

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

</property>

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

</property>

<property>

<name>hbase.zookeeper.property.dataDir</name>

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

</property>

<property>

<name>hbase.zookeeper.quorum</name>

<value>master,slave1,slave2</value>

</property>

<property>

<name>hbase.zookeeper.property.clientPort</name>

<value>2181</value>

</property>

</configuration>

* Ở các máy con:

<configuration>

<property>

<name>hbase.rootdir</name>

<value>hdfs://master/hbase</value>

</property>

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

</property>

</configuration>

Như vậy chúng ta đã cấu hình xong hbase.