2024-06-11

오전

하둡

여러개의 처리를 관리하는 하둡이 있고 또 그거를 대비한 보조 하둡이 있다.
namenode(처리관리), secondary namenode(보조처리관리) , datanode(처리하는곳)
==>HDFS

broker 1 2 3 켜서
hadoop 계정생성 - sudo adduser hadoop
su hadoop
sudo 권한부여 - sudo visudo 들어가 hadoop ALL=(ALL:ALL) ALL
개인 공캐키 생성 su hadoop ssh-keygen -t rsa
cat id_rsa.pub >> authorized_keys
chmod 600 authorized_keys
1,2,3 서로암호없이 로그인 가능하게하기

broker 1 2 3 전부다 hadoop으로 계정만들어야함!!

ssh broker2 mkdir /home/hadoop/.ssh
scp ~/.ssh/* broker2:/home/hadoop/.ssh
ssh broker2

ssh broker3 mkdir /home/hadoop/.ssh
scp ~/.ssh/* broker3:/home/hadoop/.ssh

broker1,2,3 전부다,producer까지도
/etc/hosts
여기들어가서
broker1 namenode datanode1
broker2 secondnode datanode2
broker3 datanode3
producer client

위에 한것처럼 producer도 sudo adduser만들어서 하기
ssh producer mkdir /home/hadoop/.ssh
scp ~/.ssh/* producer:/home/hadoop/.ssh

wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz

4대 컴퓨터에서 위 하둡 파일 압축 풀고 폴더 명을 hadoop 으로 변경
/home/hadoop/hadoop

bashrc
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoopexportHADOOP_INSTALL=HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOMEexportHADOOP_COMMON_HOME=HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOMEexportHADOOP_YARN_HOME=HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/nativeexportHADOOP_OPTS="-Djava.library.path=HADOOP_HOME/lib/native"
export JAVA_HOME=/usr/lib/jvm/java-17-openjdk-amd64/
export PATH=$PATH:HADOOP_HOME/sbin:$HADOOP_HOME/bin

tar xvzf ./hadoop-3.3.6.tar.gz
mv hadoop-3.3.6 ./hadoop

5개의 파일 파일질라를 통하여

hadoop@producer:~/hadoop/etc/hadoop$ pwd
/home/hadoop/hadoop/etc/hadoop 넣기

vim workers - 실제 데이터가 저장되는곳

hdfs-site.xml은 어디에 data가 저장되는지 나태내준다
hadoop-env는 환경설정

hadoop.tar.gz 으로 압축
tar -zcvf hadoop.tar.gz ./hadoop

tar -cvf hadoop.tar.gz ./hadoop

scp hadoop.tar.gz datanode1:/home/hadoop/
scp hadoop.tar.gz datanode2:/home/hadoop/
scp hadoop.tar.gz datanode3:/home/hadoop/

ssh datanode1 tar xzf hadoop.tar.gz
ssh datanode2 tar xzf hadoop.tar.gz

ssh datanode3 tar xzf hadoop.tar.gz

scp /home/hadoop/.bashrc datanode1:/home/hadoop/
scp /home/hadoop/.bashrc datanode2:/home/hadoop/
scp /home/hadoop/.bashrc datanode3:/home/hadoop/

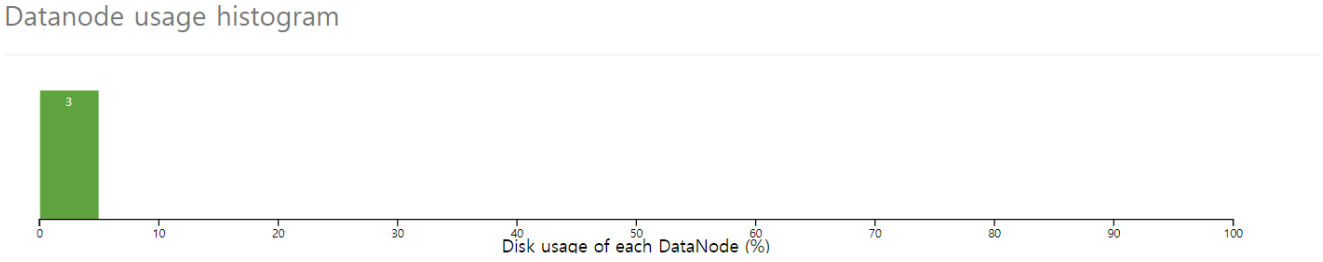
vim ~/.bashrc를 source로 안했기 때문에 다른곳으로 이동하며 읽을때 자동적으로 source실행
ssh namenode
hadoop이 실행된다
mkdir ~/data/
ssh datanode2 mkdir ~/data
ssh datanode3 mkdir ~/data
hadoop namenode -format

producer
ssh namenode start-dfs.sh
ssh secondnode start-yarn.sh
ssh namenode mr-jobhistory-daemon.sh start historyserver

jps는 자바프로세스 앞에 숫자는 pid이
ssh datanode1 jps
ssh datanode2 jps
ssh datanode3 jps

```
hadoop@producer:~$ ssh datanode1 jps
5602 Jps
4999 DataNode
4857 NameNode
```

ip:50070 으로 인터넷들어가면



3개가 살아있다.

stop-dfs.sh
stop-yarn.sh

다 끄고 재시작해보기

datanode1, node2, node3
sudo apt update
sudo apt install openjdk-11-jdk -y
vim ~/.bashrc
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/

client
vim ~/.bashrc
alias start-dfs="ssh namenode start-dfs.sh"
alias start-yarn="ssh secondnode start-yarn.sh"
alias start-mr="ssh namenode mr-jobhistory-daemon.sh start historyserver"
alias stop-dfs="ssh namenode stop-dfs.sh"
alias stop-yarn="ssh secondnode stop-yarn.sh"
alias stop-mr="ssh namenode mr-jobhistory-daemon.sh stop historyserver"

[hadoop@secondnode ~]$ jps
1700 Jps
1147 DataNode
1560 NodeManager
1235 SecondaryNameNode
1466 ResourceManager

[hadoop@namenode ~]$ jps
2017 NameNode
2557 Jps
2427 NodeManager
2139 DataNode
3021 JobHistoryServer

그리고 powertoys에
ip3개 추가하기

hdfs dfs -mkdir /encore
hdfs dfs -ls /

로컬에 있는 폴더를 enocre밑으로 넣어라
hdfs dfs -put ~/hadoop/etc/hadoop/[*].xml /encore/

wordcount 파일

hadoop jar /home/hadoop/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar grep /encore /output '[a-zA-Z]+'

hdfs에 저장이되는데 3으로 해놓았으므로 3개다 저장하고 분산시킨다 이 정보는 장부인 namenode가 갖고있다 그리고 보험으로 secondnode가 보조역할을한다. 실시간 동기x
hdfs dfs -cat /output/*