. : 현재 디렉토리

.. : 상위 디렉토리

cd .. : change directory 상위 디렉토리로 이동

ls : list 현재 경로의 정보를 볼 수 있다.

clear : 화면 정리

history : 그동안 입력한 명령어 목록

.[파일명] : 숨김 속성 파일

cat : 파일 내용 확인

ifconfig : ip 확인

which : 단축아이콘 경로 확인

readlink : 실 경로 확인

1. 하둡 설치

1.1 하둡 다운로드

- wget [url] : url에서 파일을 다운받아 현재 디렉토리에 저장

1.2 하둡 설치

1.2.1 압축 풀기

- tar zxvf [파일명] : tar(명령어) zxvf(옵션)

1.2.2 파일 이동

- 이동할 경로 생성(mkdir -p [경로])

mkdir (make directory) -p는 경로 생성시 존재하지 않는 중간 디렉토리가 있으면 강제로 생성

- 압축푼 파일을 생성한 디렉토리로 이동

mv [이동할 파일] [이동될 경로]

1.2.3 파일 정리

- rm -rf [객체명] : rm(remove)은 삭제 -rf는 삭제시 이하 디렉토리 모두 물어보지 말고 무조건 삭제

sudo rm -rf /usr/local/hadoop/

ls /usr/local/

2. 하둡 관리자 등록

2.1 하둡 그룹 생성

- sudo addgroup [그룹명]

2.2 하둡 그룹에 사용자 등록

- sudo adduser --ingroup [그룹명] [사용자명]

- groups [사용자명] : 그룹 확인

2.3 사용자에게 root 권한 부여

- sudo adduser [사용자명] [부여할 권한]

2.4 폴더에 사용자 및 그룹 권한 부여

- sudo chown -R [사용자명]:[그룹명] [부여할 디렉토리]

R 옵션은 지정한 디렉토리 이하 모든 디렉토리에 같은 권한을 부여함.

3. 고정 IP 만들기

3.1 수정할 interface 열기

ifconfig

sudo vim /etc/netplan/01-network-manager-all.yaml

3.2 서비스 재시작

sudo netplan apply

4. 환경변수를 설정

4.1 사용자 변경 : 하둡 관리자로 계정 변경

su [계정명]

4.2 변경한 계정 홈 디렉토리로 이동

cd ~

4.3 환경변수 설정 : 각 계정의 홈디렉토리에 있는 .bashrc 수정

sudo vim .bashrc

4.4 설정 파일 상세 내역

#hadoop variables start

- 하둡 홈 설정

- 하둡의 실행파일 경로 설정

- 하둡의 기능별 홈 설정

#hadoop variables end

4.5 변경된 설정 적용

source .bashrc

5. 하둡 환경변수 설정

하둡의 환경변수에 자바 홈 설정

5.1 자바 경로 확인

which java

readlink -f /usr/bin/java

5.2 하둡 환경 변수에 자바 홈 등록

sudo vim /usr/local/hadoop/etc/hadoop/hadoop-env.sh

export JAVA\_HOME /usr/lib/jvm/java-11-openjdk-amd64

6. core-site 수정

6.1 임시 폴더 생성

sudo mkdir -p /usr/local/hadoop/tmp

6.2 권한 변경

sudo chown -R manager:hadoop /usr/local/hadoop/tmp

6.3 파일 열기

sudo vim /usr/local/hadoop/etc/hadoop/core-site.xml

hadoop.tmp.dir : 하둡 처리 시 발생되는 데이터를 임시 저장할 공간

fs.default.name : HDFS의 기본 이름

hdfs://localhost:62350 : 62350은 port번호로 임의의 숫자지정

7. mapred-site 수정

job tracker 설정

7.1 템블릿 복사

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml

7.2 파일 수정

sudo vim /usr/local/hadoop/etc/hadoop/mapred-site.xml

<property>

<name></name>

<value></value>

</property>

8. hdfs-site

namenode와 datanode를 설정한다.

namenode는 마스터 서버에서 데이터를 저장하기 위한 명령을 내리고 datanode는 슬레이브 서버에서 마스터 서버가 내린 명령을 처리한다.

8.1 폴더 생성(namenode, datanode)

sudo mkdir -p /usr/local/hadoop/hdfs/namenode

sudo mkdir -p /usr/local/hadoop/hdfs/datanode

sudo chown -R manager:hadoop /usr/local/hadoop

8.2 파일 수정

sudo vim /usr/local/hadoop/etc/hadoop/hdfs-site.xml

dfs.replication : 1이면 싱글, 2이상이면 완전분산모드

9. 인증키 배포

9.1 인증키 생성

ssh-keygen -t rsa

9.2 인증키 배포

ssh-copy-id -i [인증키] [배포위치]

ssh-copy-id -i .ssh/id\_rsa.pub manager@localhost

9.3 서버 재시작

sudo reboot

10. 하둡 실행

10.1 하둡 관리자 계정 이동

su manager

cd ~

10.2 하둡 초기화(최초 1회 실행)

hadoop namenode -format

10.3 하둡 실행

source .bashrc

start-dfs.sh

10.4 상태 확인

jps

2117 Jps

1622 NameNode

1787 DataNode

2014 SecondaryNameNode

http://192.168.56.101:50070/