

2024-06-12

오전

hadoop 공부 pptx

하둡의 에코시스템

vscode키기

hadoop@producer:~\$ mkdir hadoop_test1

second노드에서 pip install mysql

아이디 비번 생성 계정 비번만들기

mysql 을 쳤는데 notfound면 sudo apt install mysql-client-core-8.0

sudo apt-get install mysql-server

cd /etc/mysql/mysql.conf.d

sudo vim mysqld.cnf

bind 0.0.0.0으로 바꾸기

sudo mysql -u root

mysql로 들어와서 계정생성하고 권한부여

extendsion 자바깔구 create java project

maven 선택 no 누르고 다 엔터

porm.xml파일의 dependencies은 import나 pip와 비슷하다

dependencies 추가하기 필요한 기능 가져와야하기에

encore.java 공부

vscode에서 java를 통하여 hdfs와 연결확인

미니콘다설치

wget https://repo.anaconda.com/miniconda/Miniconda3-py310_24.4.0-0-Linux-x86_64.sh

잘모르겠으면 q누르고 엔터

sh ./Miniconda3-py310_24.4.0-0-Linux-x86_64.sh

source ~/.bashrc

변화가 생겼으므로 bashrc한다.

/(base) hadoop@producer:~\$ conda install anaconda =>하지말기

대신에 pip install pyarrow
sudo apt install htop =>
htop으로 보는데 swp이 없어야 좋은것이다

pip install jupyter
토큰이 있는값으로 들어가기

주피터하기

병렬처리 파이썬 dask
파케이 파일형식으로 저장 parquet - 하둡방식의 데이터 저장 포맷방식

오후

ubuntu 22.04 설치

앞에 노트북은 xxx.xxx.0.201~205
F2 F12 DELETE키 누르기
BIOS에서
Secure Boot Control OFF

F10으로 UEFI신청하고
UBUNTU 설정
재부팅하고 와이파이설정 WLP2S0 WLAN 2번째꺼
와이파이설정하고 without으로 넘기기 다시

UEFI and CSM OS 설정
mirror.kakao.com
용량을 최대한으로 하는구나 max로 잡고
openssh 체크

hostnamectl set-hostname 이용해서 hostname바꾸고
sudo apt upgrade
sudo apt install openjdk-11-jdk -y

cd /etc/netplan/ wifi.yaml

오타없는지 꼭 확인하기!!

ssh-keygen -t rsa해서 auth에 저장

wget <https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz> 하둡
다운하고 tar -zcvf로 압출풀기
그리고 mv를 통해서 이름 바꾸기

mkdir data 만들기
sudo vim /etc/hosts

namenode ,secondnode datanode1~7 전부다 ssh 파일 동일시하기 (진짜 사기적인 기술)
sudo vim /etc/hosts에 ip와 hostname넣으면 ssh로 연결할때 비번없어도 된다

방화벽끄기
sudo systemctl disable --now ufw //부팅할때 끄는거
sudo systemctl stop ufw
sudo service ufw status

노트북 닫아도 서버 안꺼지는 법
[Login]
HandleLidSwitch=ignore
HandleLidSwitchDocked=ignore

sudo vim /etc/systemd/logind.conf
sudo systemctl restart systemd-logind

powertoys에 등록하기
하지만 일일이 하기 귀찮으므로
C:\Windows\System32\drivers\etc
여기에 hosts 파일 붙여넣기