

2024-06-21

오전

```
(base) hadoop@producer:~$ sudo rm /etc/localtime
```

```
(base) hadoop@producer:~$ sudo ln -s /usr/share/zoneinfo/Asia/Seoul /etc/localtime
```

nohup

```
nohup jupyter notebook --ip=0.0.0.0 &
```

mysql server 설치

```
sudo apt install mysql-server
```

```
sudo systemctl status mysql
```

메모리확인하기

```
htop
```

```
free -h
```

```
df -h
```

mysql 설정파일

```
hadoop@producer:/etc/mysql/mysql.conf.d$ pwd
```

```
sudo vim mysqld.conf
```

```
31줄보면 bind -address 변경 0.0.0.0
```

그리고 재가동!!

```
sudo systemctl restart mysql
```

데이터베이스 생성

```
sudo mysql -u root
```

```
CREATE DATABASE airflow;
```

세미콜론붙이기!!!

airflow 폴더만들기

```
mkdir ~/airflow
```

지금 만든 폴더를 AIRFLOW_HOME으로 환경변수 설정

```
sudo vim ~/.bashrc
export AIRFLOW_HOME=/home/hadoop/airflow
source ~/.bashrc\
```

airflow 깔기

AIRFLOW_VERSION=2.8.1

Extract the version of Python you have installed. If you're currently using a Python version that is not supported by Airflow, you may want to set this manually.

See above for supported versions.

PYTHON_VERSION="

```
(python3 --version|cut -d' ' -f2|cut -d'.' -f1 - 2)"CONSTRAINT_URL="https://raw.githubusercontent.com/apache/airflow/constraints-
{AIRFLOW_VERSION}/constraints-
PYTHON_VERSION.txt"Forexamplethiswouldinstall2.8.1withpython3.8: https://raw.githubusercontent.com/apache/airflow/constraints-
{AIRFLOW_VERSION}" --constraint "${CONSTRAINT_URL}"
```

daga 생성

```
hadoop@producer:~/airflow$ mkdir dags
```

airflow 실행

airflow standalone

한번 실행하고 나면 비밀번호등과 같은 파일생성

```
(base) hadoop@producer:~/airflow$ ll
total 848
drwxrwxr-x  4 hadoop hadoop  4096 Jun 21 09:41 ./
drwxr-x--- 24 hadoop hadoop  4096 Jun 21 09:38 ../
-rw-----  1 hadoop hadoop  72389 Jun 21 09:40 airflow.cfg
-rw-r--r--  1 hadoop hadoop 761856 Jun 21 09:41 airflow.db
drwxrwxr-x  2 hadoop hadoop  4096 Jun 21 09:39 dags/
drwxrwxr-x  4 hadoop hadoop  4096 Jun 21 09:40 logs/
-rw-----  1 hadoop hadoop    16 Jun 21 09:40 standalone_admin_password.txt
-rw-rw-r--  1 hadoop hadoop  4761 Jun 21 09:40 webserver_config.py
```

airflow 환경파일

```
hadoop@producer:~/airflow$ vim airflow.cfg
default_timezone = Asia/Seoul 수정
```

```
cat /etc/mysql/mysql.conf.d/mysqld.cnf | grep bind
```

ese 424쳐서 이동

yy p해서 복사

yy는 복사 p가 붙여넣기

```
sql_alchemy_conn = mysql://airflow:airflow@127.0.0.1:port/airflow
```

id와 pw둘다 airflow 마지막 airflow은 데이터베이스

sql 계정만들고 권한주기

```
mysql> create user airflow@'%' identified by 'airflow';
```

```
mysql> grant all privileges on airflow.* to airflow@'%';
```

```
pip install mysqlclient
```

계정@host명: 폴더명

폴더명 간단하게 하기

```
vim ~/.bashrcd
```

```
PS1="\h@\W$ "
```

db migrate

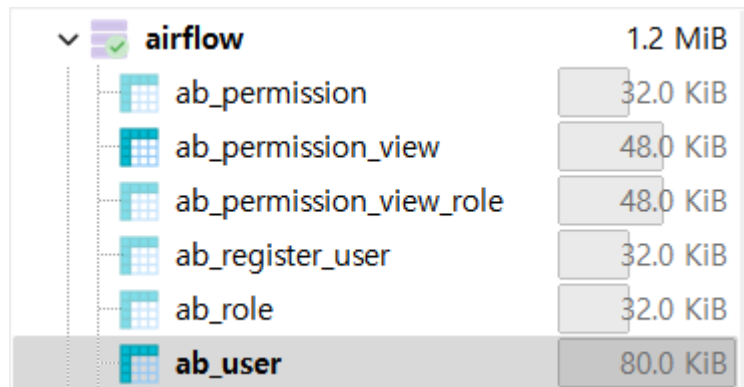
```
hadoop@producer:~/airflow$ airflow db migrate
```

heidi 들어가서 확인하기

```
airflow users create --username encore --firstname jo --lastname taesik --role Admin --email
```

[encore@encore.com](#)

비밀번호는 늘내가알던비번



airflow	1.2 MiB
ab_permission	32.0 KiB
ab_permission_view	48.0 KiB
ab_permission_view_role	48.0 KiB
ab_register_user	32.0 KiB
ab_role	32.0 KiB
ab_user	80.0 KiB

들어와있다 내가 만든 계정이

스케줄 서버 웹서버 따로따로 뛰워야한다.

웹서버 띄우기

```
hadoop@producer:~$ airflow webserver --port 8080
```

스케줄서버 띄우기

```
hadoop@producer:~$ airflow scheduler
```

airflow접속

ip:port

샘플 dags위치

에러가 뜨는데 dag

```
hadoop@producer:~$ find ./ -name "example_dags"
```

rm -rf * 로 다 지워도된다 하지만 난 안지움

결국지움

airflow가 들어간거 다 지우기

```
sudo pkill -f airflow
```

vscode로 airflow 폴더들여가기

dags 폴더 안에 mydag.py생성

vscode에서 정리하고 airflow web schedul 재실행

새로운 dag 연결해서 airflow 스케줄링

오후

airflow 크롤링 스케줄러 - 카프카연결, 카프카에서 데이터블러와서 spark