

# 2024.06.24

## 오전

가상머신 3대 키기 - 메모리 2G로 세팅  
기존 producer 2048 mb로 바꾸고 복제

ssh 세팅되어있어야하고 비밀번호 없이 로그인 되게 하기

spark가 설치되어있어야함.

메모리 부족으로 kafka\_2 kafka\_3 producer로 진행  
spark 안갈려있으면

wget <https://dlcdn.apache.org/spark/spark-3.4.3/spark-3.4.3-bin-hadoop3.tgz>

tar -xvzf spark-3.4.3-bin-hadoop3.tgz

mv spark-3.4.3-bin-hadoop3 spark

vim ~/.bashrc

```
13 SPARK_HOME=/home/kafka/spark
14 PATH=$PATH:$SPARK_HOME/bin
```

source ~/.bashrc

잘안되어서 kafka\_1,2,3로 진행  
셋다 kafak계정

새로 환경만들거면 sudo vim /etc/netplan/00...yaml  
파일에서 ip수정

계정보려면 cd /home/ 에서 ll치면 다 나온다 encore hadoop kafka3개

sudo hostnamectl set-hostname worker2 이름 다 바꿔주기

cp spark-env.sh.template spark-env.sh  
vim spark-env.sh

```
export SPARK_MASTER_HOST=master
export SPARK_MASTER_PORT=7077
export SPARK_MASTER_WEBUI=8080
export SPARK_WORKER_CORES=2
export SPARK_WORKER_MEMORY=1g
export SPARK_WORKER_INSTANCES=1
```

```
cp workers.template workers
```

```
vim worker에서
```

```
worker1  
worker2
```

뒤늦게

```
sudo mv ./spark /opt
```

다시 ~/.bashrc 수정하기

```
kafka@worker2:/opt$ scp -r ./spark/ worker1:/home/kafka/
```

```
kafka@worker2:/opt$ ssh -t worker1 "sudo mv /home/kafka/spark/ /opt/spark/"
```

ipython이 안되시는 분들은

.bashrc에

```
alias ipython="python -m IPython"
```

위 구문 추가해주시면 됩니다.

저것도 안되면

```
alias ipython="python3 -m IPython"
```

이거로 바꿔주세요

subprocess로 shell문을 보낼수있다....

파이썬파일로 만들어서 python3 실행시키면 된다.

subprocess에서 Popen으로도 은 인자를 각각 "", 로 이루어져야한다.

```
import subprocess
```

```
for x in range(1,3):
```

```
subprocess.call(f"scp /home/hadoop/.bashrc worker{x}:/home/hadoop/", shell=True)
```

```
vim a.py
```

```
import subprocess
```

```
while(True):
```

```
subprocess.call("ping 192.168.0.160", shell=True)
```

```
nohup python a.py &
```

이름들여가는걸 모두 죽여줘!!

```
sudo kill -f "이름"
```

```
sudo chown kafka:kafka -R ./spark/
```

```
kafka@master:/opt/spark/sbin$ ./start-all.sh
```

```
sudo apt install openjdk-11-jdk -y
```

```
3156 Jps
677 QuorumPeerMain
kafka@worker1:/opt$ jps
4977 Jps
677 QuorumPeerMain
678 Kafka
```

```
kafka@master:/opt/spark/sbin$ jps
676 Kafka
3989 Jps
3925 Master
3517 Worker
```

뭔가가 이상하다....

일단 zookeeper와 kafka를 꺾다...그런데 stop-all.sh를 해도 worker가 뜬다??  
그거는 localhost에 주석을 풀면된다...

```
pyspark --master spark://master:7077 --num-executors 2
```

notebook 파일에서

```
import pyspark
from pyspark.sql import SparkSession
conf = pyspark.SparkConf().setAppName("gen").
setMaster("spark://master:7077").
set("spark.executor.instances", "2")
spark = SparkSession.builder.config(conf=conf).getOrCreate()
spark
```

주피터를 깔자...

```
6. jupyter로 이용하고 싶은 경우
1. jupyter notebook --generate-config : 환경파일 만들기
2. jupyter server password : 비밀번호 만들기
3. cat jupyter_server_config.json 에 들어가서, "hashed_password" 복사!
4. 복사 후, vim jupyter_notebook_config.py 에 들어가서, 하기와 같이 paste![[Pasted image 20240523150755.png]]
5. nohup jupyter notebook --ip=0.0.0.0 &
6. jupyter notebook 으로 연결시키는 명령어를 bashrc에 넣기!

export PYSPARK_DRIVER_PYTHON=jupyter
export PYSPARK_DRIVER_PYTHON_OPTS='notebook --ip=0.0.0.0'

7. pyspark 실행 시, jupyter 파일이 열림.
```

주피터를 깔았다

`sudo apt upgrade`

vs code에서 주피터 다운하기....

위 순서대로 하기 ipython들어가서 949번째줄

## 오후

`ssh worker2 "jps"`

ip:4040

ip:8080