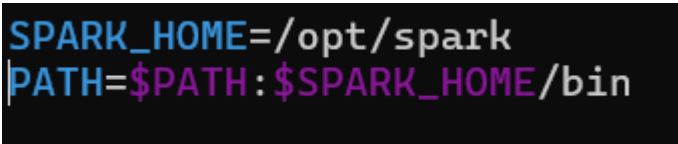


2024-06-17

오전

1. spark
2. 하둡은 저장소 음식점, spark는 배
3. airflow
4. 아파치 스파크 다운받기
5. su kafka
6. wget <https://dlcdn.apache.org/spark/spark-3.4.3/spark-3.4.3-bin-hadoop3.tgz>
7. ssh hadoop@client로 들어가서 저번에 포워딩한다고 비밀번호 바꿈..md5 hash로
8. scp hadoop@ip:/home/hadoop/spark-3.4.3-bin-hadoop3.tgz ~/
9. tar -xvzf [파일 명] 로 압축풀기
10. 파일이름 spark로 옮기기
11. /opt/spark로 옮기기
12. kafka@producer:~\$ sudo mv ./spark /opt
13. vim ~/.bashrc
14. source ~/.bashrc
15. pyspark

- 16.
- 17.
18. 하둡에다 설치해야되는줄 알고 su로 하둡들어가서 sudo chown hadoop:hadoop -R ./kafka/ 하다가 vim해서 16번 넣은다음에 source했었는데 이상한것같아서 다시 su kafka들어가서 sudo chown hadoop:hadoop -R ./kafka/ 다시 바꾸고
19. 주피터 패스워드 설정해보기!!!
20. kafka에서 진행중
21. kafka@producer:~\$ sudo apt install jupyter-core
22. pip install jupyter
23. 오류나므로 miniconda 깔기
24. wget https://repo.anaconda.com/miniconda/Miniconda3-py310_24.4.0-0-Linux-x86_64.sh
25. sh Miniconda3-py310_24.4.0-0-Linux-x86_64.sh
26. hadoop@producer:~/miniconda3/bin\$ conda init
27. source ~/.bashrc

28. (base) hadoop@producer:~/miniconda3/bin\$ jupyter notebook --generate-config
29. (base) hadoop@producer:~/miniconda3/bin\$ ipython
30. from jupyter_server.auth import passwd
31. passwd()
32. (base) hadoop@producer:~/jupyter\$ vim jupyter_notebook_config.py
33. serverApp.passwd에 넣기 949번째 줄
34. (base) hadoop@producer:~\$ vim ~/.bashrc export PYSPARK_DRIVER_PYTHON=jupyter
export PYSPARK_DRIVER_PYTHON_OPTS='notebook --ip=0.0.0.0'
35. (base) hadoop@producer:~\$ source ~/.bashrc
36. pyspark로 하면 jupyter-notebook으로 연결
37. 주피터 노트북 참조!!
38. (base) hadoop@producer:/opt/spark/conf\$ sudo cp ./spark-defaults.conf.template ./spark-defaults.conf
39. workes는 집단에 이루는 컴퓨터 이것이 분산처리 클러스터
40. spark-defaults.conf 어느집단이는 master 리더가있따
41. 환경파일은 conf에있음.
42. (base) hadoop@producer:~\$ hdfs dfs -mkdir /mort_0617 (base) hadoop@producer:~\$
hdfs dfs -put test.txt /mort_0617
43. <https://spark.apache.org/docs/latest/sql-ref-datatypes.html>