**Spark 2 docker**

<https://github.com/gettyimages/docker-spark> - working

Download codes

Docker build -t sparkimage .

docker run -d \  
 --name devtest \  
 --mount type=volume,source=my-vol-1,target=/home/bl \

-it -p 4040:4040 sparkimage

docker run --rm -d \  
 --name devtest \  
 --mount type=volume,source=my-vol-1,target=/home/bl \

-it -p 4040:4040 sparkimage

Stop & remove the container if already running

Sudo docker stop <container id>

Sudo docker stop 09178ab3ec59c00776504fa47676d404f64570f986e2e1e71180738f4c32aabf

Sudo docker rm <container id>

Sudo docker rm 09178ab3ec59c00776504fa47676d404f64570f986e2e1e71180738f4c32aabf

docker exec -it devtest /bin/bash

root@09178ab3ec59:/# ls

bin dev home lib64 mnt proc run srv tmp var

boot etc lib media opt root sbin sys usr

Mkdir home

Cd home

Mkdir bl

host machine CMD

Edit

Nano mife-copy5.py

lines = sc.textFile("/home/bl/carbon.log")

Copy files from host machine to docker

sudo docker cp /home/devops/Desktop/spark2/mife-copy5.py devtest:/home/bl/

sudo docker cp /home/devops/Desktop/spark2/carbon.log devtest:/home/bl/

Run the pyspark file

root@09178ab3ec59:/# /usr/spark-2.4.0/bin/spark-submit /home/bl/mife-copy5.py