

- You can add any number of text files to your input folder and rerun the steps. Also change your MR program (`mapper.py` and `reducer.py`) to solve something other than wordcount and apply the steps discussed.

12. Additional tips

- If you get a connection refused error when running hadoop commands in the docker, Your name node might not have started or went down. Restart it using `service hadoop-hdfs-namenode restart` command
- `jps` can help you check the health of your hadoop environment
- Read about docker management commands:
 - i. `$ docker ps -a`
 - Lists all the running dockers on your system
 - ii. `$ docker stop containerID`
 - Stops the docker container running with the given containerID
 - iii. `$ docker stop $(docker ps -a -q)`
 - Stops all the docker containers running on your machine
 - iv. `$ docker rm $(docker ps -a -q)`
 - Removes the containers from the memory i.e. the docker can't be resumed
 - v. `$ docker images`
 - Lists all the images you have on your machine
 - vi. `$ docker rmi ImageID`
 - Removes the docker image of the given ImageID from your machine

Original Document prepared by Yesh Kumar Singh (yeshkuma@buffalo.edu), modified by Bina Ramamurthy (bina@buffalo.edu) to suit the CSE4/587 course