Hadoop-Docker

Installation

- Set up Docker Engine
  - https://docs.docker.com/engine/install/ubuntu/
    - Run steps 1 and 2 from "install from apt repository".

```
# Add Docker's official GPG key:
sudo apt-get update
sudo apt-get install ca-certificates curl
sudo install -m 0755 -d /etc/apt/keyrings
sudo curl -fsSL https://download.docker.com/linux/ubuntu/gpg -o /etc/apt/keyrings/docker.
sudo chmod a+r /etc/apt/keyrings/docker.asc

# Add the repository to Apt sources:
echo \
  "deb [arch=$(dpkg --print-architecture) signed-by=/etc/apt/keyrings/docker.asc] https:/
  $(. /etc/os-release && echo "$VERSION_CODENAME") stable" | \
  sudo tee /etc/apt/sources.list.d/docker.list > /dev/null
sudo apt-get update
```

2. **Install the Docker packages.**

   Latest   Specific version

   To install the latest version, run:

   ```
   $ sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-
   ```

Installation

- Run after the previous steps

```
$ sudo systemctl enable docker.service
$ sudo systemctl enable containerd.service
```

Installation

- Clone the repository

  https://github.com/dhzdhd/hadoop-docker-compose.git

  ```
  git clone https://github.com/dhzdhd/hadoop-docker-compose.git
  ```

  Cd into the directory

  Run "sudo docker compose up –d "

  After the building of the container is complete ,

Installation

- Run "sudo docker ps –a "

```
PS D:\college\Hadoop\hadoop-docker-compose> docker ps -a
CONTAINER ID   IMAGE                          COMMAND                 CREATED       STATUS                      PORTS
                                                                      NAMES
ae58a8219a80   hadoop-docker-compose-master   "python3 -m http.ser…"  3 weeks ago   Exited (255) 43 seconds ago  0.0.0
.0:8088->8088/tcp, 0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870/tcp   master
PS D:\college\Hadoop\hadoop-docker-compose> 
```

ae58a8219a80

- This is the container id copy it (ctrl+shift+c)

Installation

- Once that is done Run "docker exec –it <container_id> /bin/bash"
- Run "init" [THIS STEP IS TO BE DONE ONLY THE FIRST TIME ,DOING IT AGAIN FORMATS HDFS]

Running docker(hadoop) after first time.

- Go to the directory , that was cloned.

- Open the directory in terminal

- Run "docker ps –a " like in the previous slide and copy the container id

- Run "docker exec –it<container_name> bin/bash"

- Run "Restart" [DO NOT RUN INIT , IT FORMATS HDFS]

Notes

- Suppose there is a problem during the building of container, that is because , the hbase website might be blocked due to manipal proxy , please stop the process(ctrl c) and run "docker compose up" after connecting to hotspot.

- For Windows do not use sudo while running the commands, install docker desktop , and clone the repository, follow the steps from slide 4-7

- You can use the "workdir" directory to copy your files in and out of the container.

- You can access the hdfs file view at "localhost:9870".

How does it look?

- It should look something like this

```
root@ae58a8219a80:/# ls
bin     derby.log    etc    lib      lib64    media          mnt    pig     root    sbin    sys    usr    workdir
boot    dev          home   lib32    libx32   metastore_db   opt    proc    run     srv     tmp    var    '~'
```

- From this point it is usual Hadoop.