Hadoop and Hive Setup using Docker

Israk Ahmed
ID: 2037820103
CSE 2019-20

Step 1: Pull and Run Hadoop Docker Container

Listing 1: Pull Hadoop Docker Image

```
docker pull macio232/hadoop-pseudo-distributed-mode
```

Explanation: Downloads the Hadoop pseudo-distributed mode image from Docker Hub.

Listing 2: Run Hadoop Container

```
docker run -p 9870:9870 -p 8088:8088 -it --name=testHadoop
    ↪ macio232/hadoop-pseudo-distributed-mode
```

Explanation: Runs the Hadoop container interactively and maps ports for HDFS (9870) and YARN (8088).

Step 2: Create and Edit a Text File

```
cd home
cd Hadoop
vi student.txt
```

Explanation: Navigate to the Hadoop directory and create or open `student.txt` using the `vi` editor.

Inside vi editor:

- Press **I** to enter insert mode (if "INSERT" not shown).
- Type your text (e.g., names and roll numbers).
- Press **ESC** to exit insert mode.
- Press **Shift + :** to open command mode.
- Type **wq** to save and quit or **q** to quit without saving.
- Press **Enter** to execute.

Step 3: Work with HDFS

```
hdfs dfs -mkdir /testDirectory
```

Explanation: Creates a new directory named `/testDirectory` in HDFS.

```
hdfs dfs -put /home/hadoop/student.txt /testDirectory
```

Explanation: Uploads the local `student.txt` file to HDFS under `/testDirectory`.

Step 4: Use Hive for Data Processing

```
hive
```

Explanation: Opens the Hive shell interface.

```
show databases;
create database mydb;
show databases;
use mydb;
show tables;
```

Explanation: Displays available databases, creates a new one named `mydb`, and switches to it.

```
create table student (name string, roll int)
row format delimited
fields terminated by '\t';
```

Explanation: Creates a table named `student` with two columns (`name`, `roll`), using tab-delimited format.

```
show tables;
load data inpath '/testDirectory/student.txt' into table student;
select * from student;
select name from student;
```

Explanation: Shows tables, loads data from HDFS, and queries the `student` table.

Step 5: Manage Docker Container

```
docker stop testHadoop
```

Explanation: Stops the running Hadoop container.

```
docker container start -i testHadoop
```

Explanation: Restarts the Hadoop container in interactive mode.