Docker Setup:

Step-1: open a browser and sign in (https://www.docker.com/), then download the (docker-desktop-installer.exe) file.

step-2: Search for 'turn windows features on or off' from your pc's start menu.

step-3: See the options 'Hyper-V' and 'Windows Subsystem for Linux' and check them, then click 'OK'. The pc will required restart after few minutes.

step-4: Open command prompt and type-

> wsl --status

> wsl --update

> wsl --set-default-version 2

> wsl --set-default-version 1

> wsl --set-default-version 2

step-5: install and execute (docker-desktop-installer.exe) that you have downloaded. PC will take restart again.

step-6: Open command prompt and type-

> docker version

> docker images

> docker search mysql

Docker Execution

---

step-1: open a browser and serach for 'https://docker-curriculum.com/' and see the commands.

step-2: run 'ocker-desktop-installer.exe' application and open the docker terminal and execute the following commands:

$ docker run hello-world

$ docker pull busybox

$ docker run busybox

$ docker run busybox echo "hello from busybox"

$ docker images

$ docker ps

$ docker ps -a

step-3: Run hadoop in docker terminal:

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

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

step-4: a console will open (for linux)

ls

cd home/

/home# cd hadoop/

/home/hadoop# vi student.txt

Step-5: write some thing to the student.txt file

Kabir 24

Bashar 25

Momin 26

Atik 24

Amir 25

type "esc -> : -> wq -> enter" for write and quite.

step-6: now open a local browser and check the ports are active (port:127.0.0.1:9870 and port:127.0.0.1:8088)

step-7: create a folder on hadoop ecosystem.

/home/hadoop# hdfs dfs -mkdir /samrat/

/home/hadoop# hdfs dfs -put /home/hadoop/student.txt /samrat

/home/hadoop# hive

hive> show databases;

```
hive> create database samrat-test;
hive> show databases;
hive> create database joy_test;
hive> use samrat-test;
hive> show tables;
hive> create table student(Name string, Age int)
> Row format delimited
> Fields terminated by '\t';
hive> show tables;
hive> slect * from student;
hive> load data inpath '/samrat/student.txt' into table student;
hive> slect * from student;
hive>
```

step-8: To close the the above docker window, open a new window and type

```
hive> docker stop testHadoop
```

step-9: To reopen the the closed docker window, type

```
hive> docker container start -i testHadoop
```

Create Database Statement:

hive> CREATE DATABASE IF NOT EXISTS Student;

Verify the Databases List:

hive> SHOW DATABASES;

default

Student

To use The Database:

hive> use Student;

To See the Database Location:

hadoop> hdfs dfs -ls /user/hive/warehouse;

Hive Table Types

Internal or Managed table: You can drop the table with underlying data.

External table: You can drop an external table, only table metadata from Metastore will be removed but the underlying files will not be removed and still they can be accessed via HDFS commands, Spark or any other Hadoop compatible tools.

Temporary table: For temporary purpose.

Transactional Table: For transactional data purpose.

Table Creation

```
 hive> CREATE External TABLE IF NOT EXISTS Student.Info(
Id int,
Name string,
Age int)

        PARTITIONED BY (Gender string)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
        STORED AS TEXTFILE
        LOCATION '/data/output/';
```

To See the Table Structure:

hive> DESCRIBE student.info;

Insert Data into Table:

hive> INSERT INTO student.info values(7,'Maruf',23,'M');

hive> INSERT INTO student.info values(8,'Rina',50,'F');

For Bulk Data Load:

hive> LOAD DATA INPATH '/user/data/data.txt' INTO TABLE Student.Info;

hive> LOAD DATA INPATH '/path/to/HDFS/dir/file.csv' OVERWRITE INTO TABLE Student.Info PARTITION (Gender='M');

hive> LOAD DATA INPATH '/path/to/HDFS/dir/file.csv' OVERWRITE INTO TABLE Student.Info PARTITION (Gender='F');

To Retrieves the all data:

hive> SELECT * FROM Info;

Conditional Data Retrieve:

hive> SELECT * FROM Info

WHERE age=23;

To See All Partitions:

hive> show partitions Info;

To Drop a Partition:

hive>ALTER TABLE Student.Info DROP PARTITION (gender="F");

To Repair a Partition:

hive> Msck repair table Student.Info;