

# Getting Started with Hive



**Estimated time needed:** 20 minutes

In this lab you will explore Apache Hive, a distributed, fault-tolerant data warehouse system that enables analytics at a massive scale. You will be creating a table and running SQL commands on it.

## Learning Objectives

At the end of this lab, you will be able to:

- Create a table in Hive
- Add data to the table using file
- Add data to the table using `insert`
- Query the data in the table using SQL commands

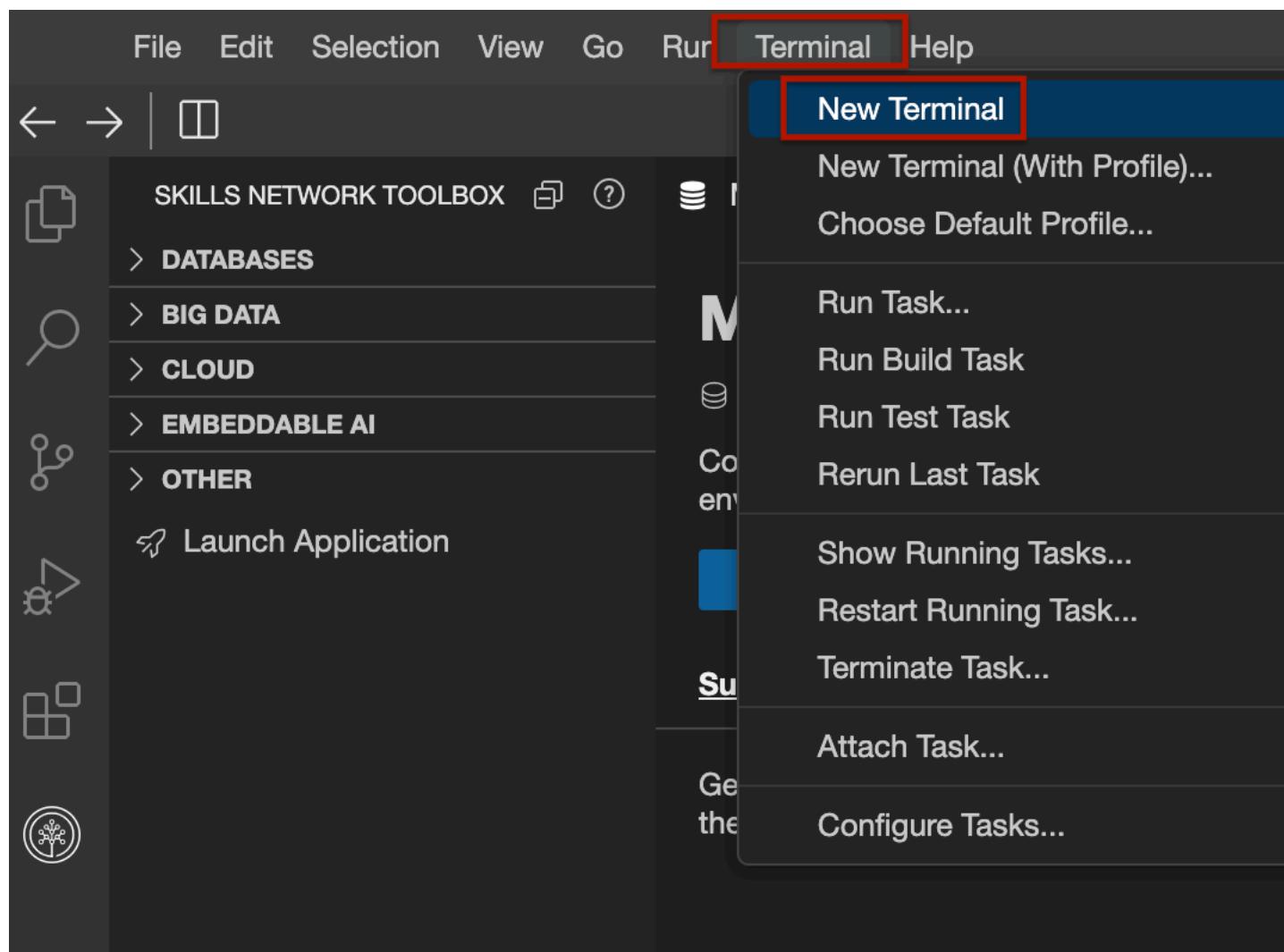
## Prerequisites

- You should be comfortable working with the Linux terminal
- Prior knowledge of SQL will be helpful

While all the terminal commands can be copied and pasted, it is highly recommended for you to type the commands for better learning.

## Step 1: Get a copy of the CSV file

1. You will run the commands in the terminal. If you don't have a terminal open, open a new terminal, by clicking on Terminal and choosing New Terminal from the submenu.



2. Create a directory named `data` under `/home/project` by running the following command.

```
mkdir /home/project/data
```

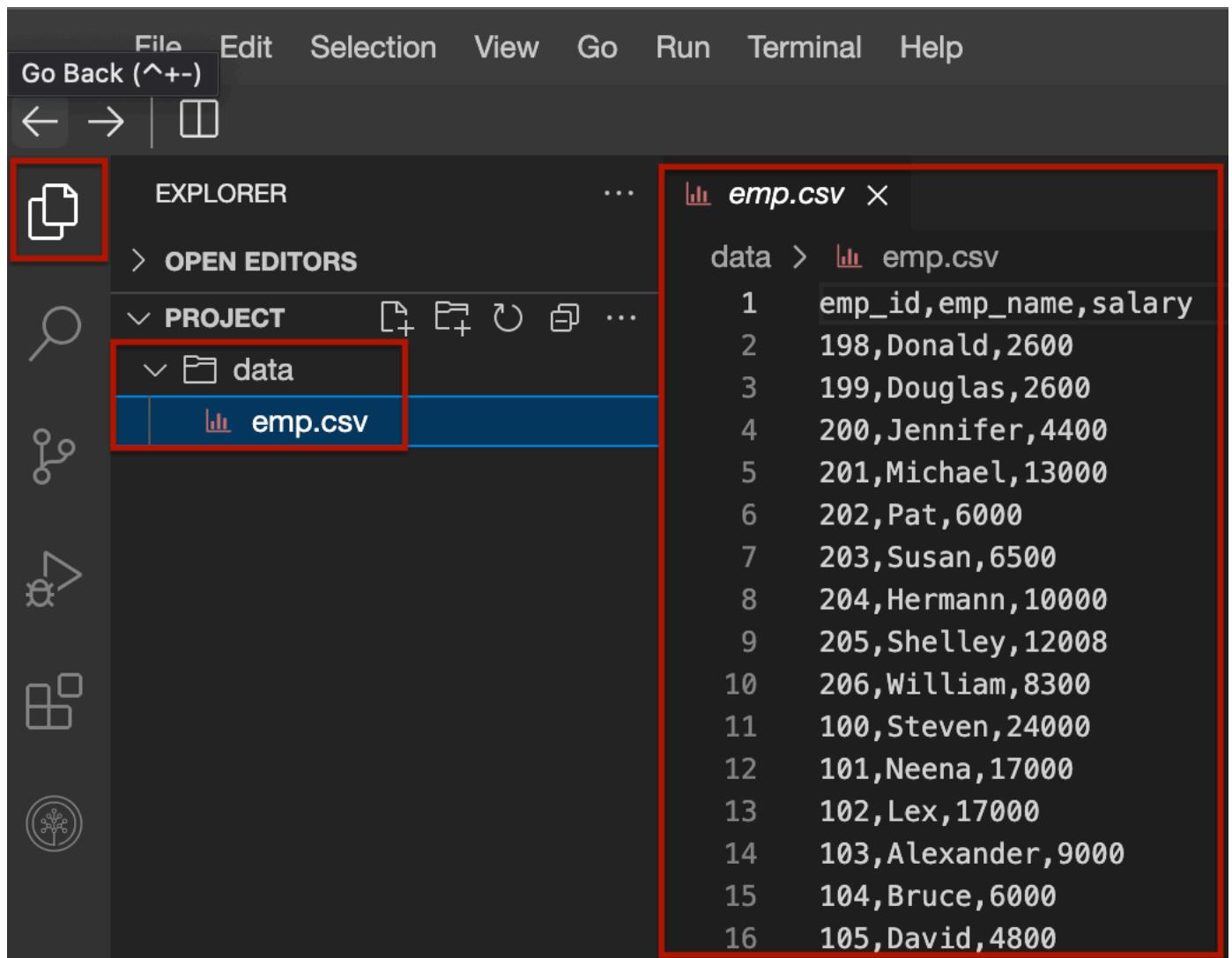
3. Change to the /home/project/data directory.

```
cd /home/project/data
```

4. Run the following command to get the emp.csv, a data file with Employee data, in a comma-separated file which you will use later to infuse data into the table you create.

```
wget https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-BD0225EN-SkillsNetwork/data/emp.csv
```

5. Open the file in editor and view the file.



## Step 2: Setup Hive and Bee

1. You will use the hive from the docker hub for this lab. Pull the hive image into your system by running the following command.

```
docker pull apache/hive:4.0.0-alpha-1
```

This will take a few seconds, depending on the speed of your internet connection.

2. Now, you will run the hive server on port **10002**. You will name the server instance **myhiveserver**. We will mount the local **data** folder in the hive server as **hive\_custom\_data**. This would mean that the whole **data** folder that you created locally, along with anything you add in the **data** folder, is copied into the container under the directory **hive\_custom\_data**.

```
docker run -d -p 10000:10000 -p 10002:10002 --env SERVICE_NAME=hiveserver2 -v /home/project/data:/hive_custom_data --name myhiveserv
```

3. You can open and take a look at the Hive server with the GUI. Click the button to open the HiveServer2 GUI.

HiveServer2 GUI

4. Now run the following command, which allows you to access **beeline**. This is a SQL cli where you can create, modify, delete table, and access data in the table.

```
docker exec -it myhiveserver beeline -u 'jdbc:hive2://localhost:10000/'
```

## Step 3: Create table, add and view data

1. To create a new table **Employee** with three columns as in the csv you downloaded - **em\_id**, **emp\_name** and **salary**, run the following command.

```
create table Employee(em_id string, emp_name string, salary int) row format delimited fields terminated by ',' ;
```

You may notice that there is an explicit mention for the fields delimited by **,** just as in the csv file.

2. Run the following command to check if the table is created.

```
show tables;
```

This should list the **Employee** table that you just created.

3. Now load the data into the table from the csv file by running the following command.

```
LOAD DATA INPATH '/hive_custom_data/emp.csv' INTO TABLE Employee;
```

3. Run the following command to list all the rows from the table to check if the data has been loaded from the CSV.

```
SELECT * FROM employee;
```

4. You can view the details of the commands and the outcome in the HiveServer2 GUI.

[HiveServer2 GUI](#)

5. To quit from the beehive prompt in the terminal, press `ctrl+D`.

Hive internally uses MapReduce to process and analyze data. When you execute a Hive query, it generates MapReduce jobs that run on the Hadoop cluster.

## Conclusion

In this lab you created a table in hive, added data to the table from csv and listed the data contained in the table.

## Next Steps

You can explore more SQL commands with table and see how it works.

## Author(s)

Lavanya T S

© IBM Corporation. All rights reserved.