

## Dimensional Modeling in Hive

The screenshot shows a Hive query editor interface. At the top, a text area contains the SQL query: `1 Create database DW_project;|`. Below the text area are buttons for **Execute**, **Save as...**, **Explain**, **Format**, and **New query**. Below these buttons is a tabbed interface with tabs for **Recent queries**, **Query**, **Log**, **Columns**, **Results**, and **Chart**. The **Recent queries** tab is active, displaying a table of recent queries:

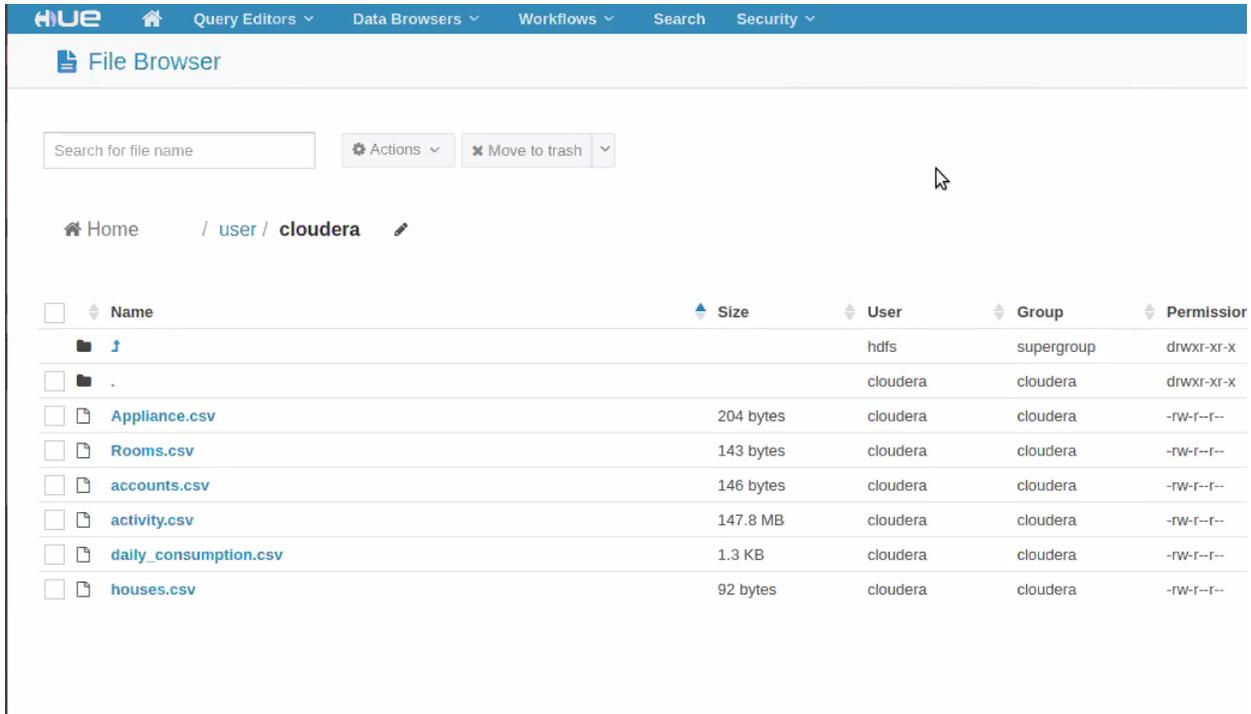
Time	Query	Result
01/07/2021 10:59:27 PM	<code>create table student( id bigint,name string,remarks string );</code>	<a href="#">See result</a>
01/07/2021 10:58:22 PM	<code>use test;</code>	<a href="#">See result</a>
01/07/2021 10:58:01 PM	<code>create database test;</code>	<a href="#">See result</a>

To do dimensional modeling in Hive first we need to load our data from csv into the dw\_project tables. All the csv files were uploaded to Cloudera directory.

The screenshot shows the Cloudera File Browser interface. A modal dialog titled "Upload to /user/cloudera" is open, displaying a list of files to be uploaded:

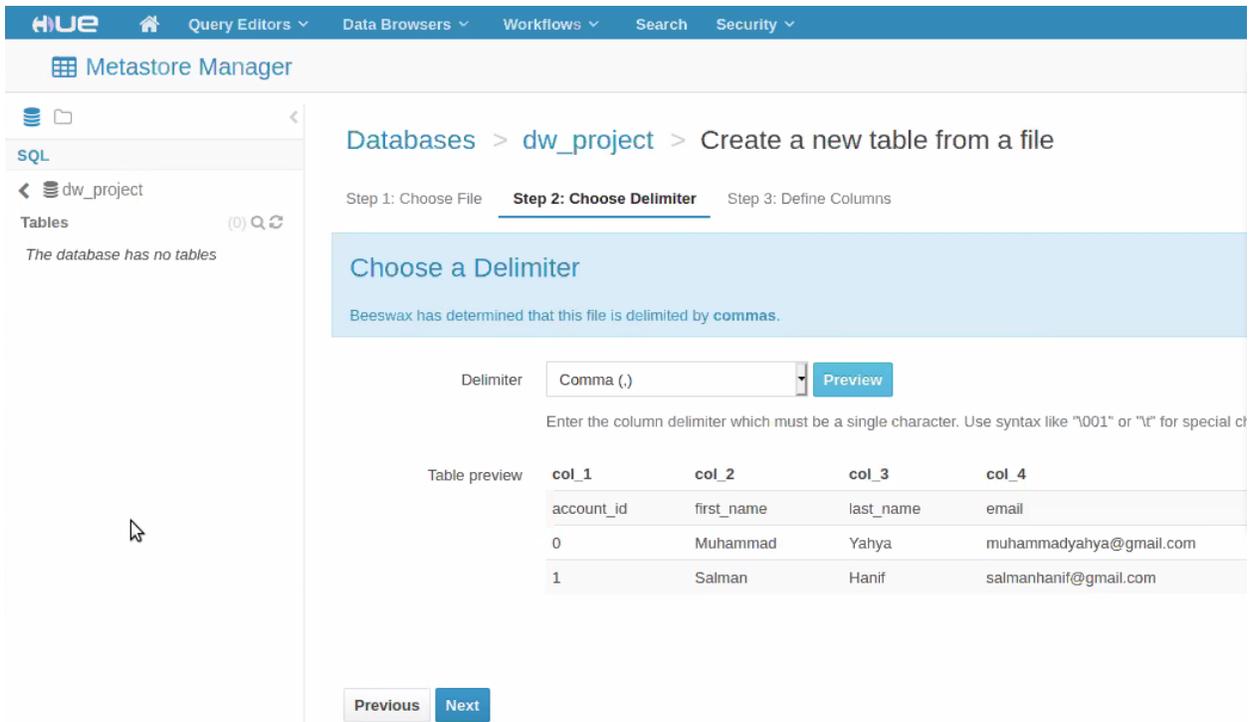
Name	Size
daily_consumption.csv	1.3kB
activity.csv	99% from 0.1GB
Rooms.csv	0.1kB
Appliance.csv	0.2kB
houses.csv	0.1kB
accounts.csv	0.1kB

The dialog also includes a "Select files" button and a "Permissions" column on the right side of the file list.



Creating Tables after uploading data CSV:

The following snapshots are how we imported the data into hive from our CSV files. Each csv file was assigned a individual table.



The screenshot shows the Hue Metastore Manager interface for creating a new table from a file. The breadcrumb path is 'Databases > dw\_project > Create a new table from a file'. The interface is in 'Step 1: Choose File'. The left sidebar shows a tree view with 'SQL', 'dw\_project', and 'Tables' containing 'account'. The main content area has a title 'Name Your Table and Choose A File'. It includes a 'Table Name' field with 'houses', a 'Description' field with 'Optional', and an 'Input File' field with '/user/cloudera/houses.csv'. There is a checked checkbox for 'Import data from file' and a yellow warning box at the bottom stating 'Warning: The selected file is going to be moved during the import.'

The screenshot shows the Hue Metastore Manager interface for creating a new table from a file, now in 'Step 2: Choose Delimiter'. The breadcrumb path is 'Databases > dw\_project > Create a new table from a file'. The left sidebar is the same as in the previous screenshot. The main content area has a title 'Choose a Delimiter'. A message states 'Beeswax has determined that this file is delimited by commas.' The 'Delimiter' dropdown is set to 'Comma (,)' with a 'Preview' button. Below this is a 'Table preview' table showing the data from the file.

col_1	col_2	col_3	col_4	col_5
hosue_id	account_id	house_name	owner_name	rooms
0	0	The Villa	Rasheeda Abbas	10

At the bottom, there are 'Previous' and 'Next' buttons.

HUE Query Editors Data Browsers Workflows Search Security

### Metastore Manager

Step 1: Choose File Step 2: Choose Delimiter **Step 3: Define Columns**

#### Define your columns

Use first row as column names Bulk edit column names

Column name	Column Type	Sample Row #1	Sample Row #2
hosue_id	string	0	
account_id	string	0	
house_name	string	The Villa	
owner_name	string	Rasheeda Abbas	
rooms	string	10	

### Metastore Manager

Databases > dw\_project > houses

Add a description...

**Overview** Columns (5) Sample Details

**PROPERTIES** **STATS**

Table	Location
cloudera	1 files
Thu Jan 07 18:31:40 UTC 2021	40 bytes
text	Not compressed

**COLUMNS (5)**

	Name	Type	Comment
1	house_id	int	Add a comment...
2	account_id	string	Add a comment...
3	house_name	string	Add a comment...

View more...

SAMPLE

The screenshot shows the HUE Metastore Manager interface. The breadcrumb navigation is 'Databases > dw\_project > Create a new table from a file'. The current step is 'Step 1: Choose File', with 'Step 2: Choose Delimiter' and 'Step 3: Define Columns' visible as subsequent steps. The main heading is 'Name Your Table and Choose A File'. There are three input fields: 'Table Name' with the value 'rooms', 'Description' with the value 'Optional', and 'Input File' with the value '/user/cloudera/Rooms.csv'. A checkbox for 'Import data from file' is checked. A warning message at the bottom states: 'Warning: The selected file is going to be moved during the import.'

The screenshot shows the HUE Metastore Manager interface at 'Step 2: Choose Delimiter'. The breadcrumb navigation is 'Databases > dw\_project > Create a new table from a file'. The main heading is 'Choose a Delimiter'. A message states: 'Beeswax has determined that this file is delimited by commas.' The 'Delimiter' dropdown is set to 'Comma (,)' with a 'Preview' button next to it. Below this is a 'Table preview' section with a table showing the first few rows of data. At the bottom, there are 'Previous' and 'Next' buttons.

col_1	col_2	col_3	col_4	col_5
room_id	house_id	room_name	description	sensor_count
0	0	Kitchen	Kitchen has many sensors ...	3
1	0	Living Room		1
2	0	Home Office		1

The screenshot shows the Hue Metastore Manager interface. The left sidebar displays a tree view with 'SQL' and 'dw\_project' folders, and a 'Tables' section containing 'account', 'houses', and 'rooms'. The main content area is titled 'Define your columns'. It features two tabs: 'Use first row as column names' (selected) and 'Bulk edit column names'. Below the tabs is a table with columns for 'Column name', 'Column Type', 'Sample Row #1', and 'Sample Row #2'. The table contains five rows of data for column definition.

Column name	Column Type	Sample Row #1	Sample Row #2
room_id	bigint	0	1
house_id	bigint	0	0
room_name	string	Kitchen	Living Room
description	string	Kitchen has many sensors in it	
sensor_count	bigint	3	1

The screenshot shows the Hue Metastore Manager interface at the 'Choose a Delimiter' step. The left sidebar is similar to the previous screenshot but includes 'rooms' in the 'Tables' list. The main content area is titled 'Choose a Delimiter'. A message states: 'Beeswax has determined that this file is delimited by commas.' Below this, there is a 'Delimiter' dropdown menu set to 'Comma (,)' and a 'Preview' button. A note below the dropdown says: 'Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.' A 'Table preview' section shows a table with 6 columns: 'col\_1', 'col\_2', 'col\_3', 'col\_4', 'col\_5', and 'col\_6'. The table contains 5 rows of data.

Table preview

col_1	col_2	col_3	col_4	col_5	col_6
sesnor_id	room_id	appliance_name	rating	status	add_time
0	0	Dishwasher	1.5	active	
1	0	Microwave	1	active	
2	0	Fridge	0.25	active	
3	1	Living_Room_Sensor	0.2	active	
4	2	Home_Office	0.6	active	

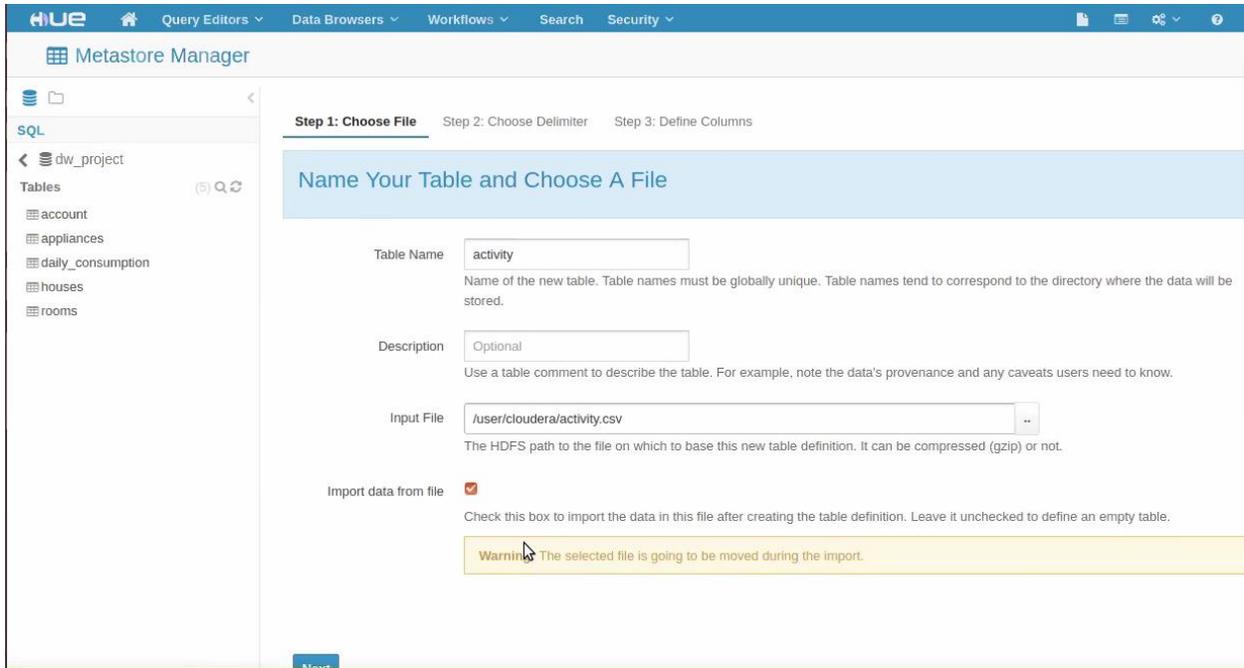
The screenshot shows the Hue Metastore Manager interface. The top navigation bar includes 'HUE', 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. The main header is 'Metastore Manager'. On the left, a sidebar shows a tree view with 'SQL' and 'dw\_project' containing tables 'account', 'houses', and 'rooms'. The main content area is titled 'Define your columns'. It has two buttons: 'Use first row as column names' and 'Bulk edit column names'. Below is a table with columns: 'Column name', 'Column Type', 'Sample Row #1', and 'Sample Row #2'. The table contains the following data:

Column name	Column Type	Sample Row #1	Sample Row #2
sensor_id	bigint	0	1
room_id	bigint	0	0
appliance_name	string	Dishwasher	Microwave
rating	bigint	1.5	1
status	string	active	active
add_time	string		

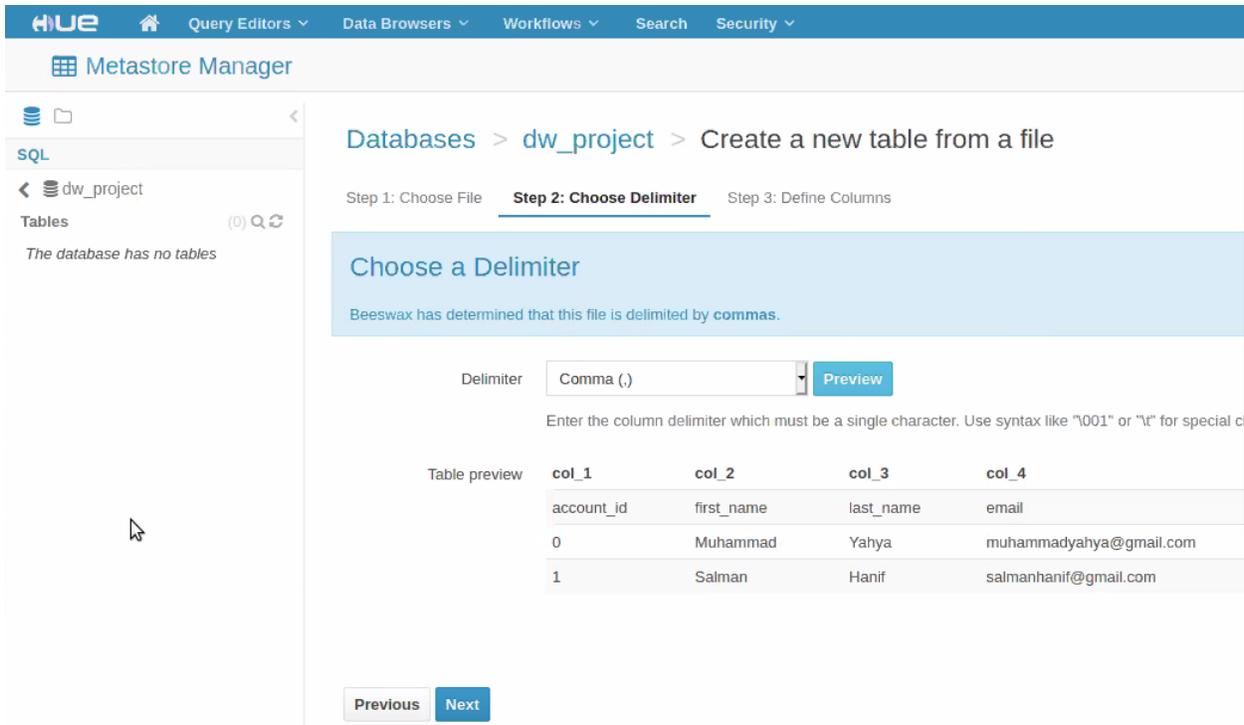
The screenshot shows the Hue Metastore Manager interface at the 'Name Your Table and Choose A File' step. The top navigation bar is the same as the previous screenshot. The main header is 'Metastore Manager'. The sidebar is the same. The main content area is titled 'Name Your Table and Choose A File' and has three steps: 'Step 1: Choose File', 'Step 2: Choose Delimiter', and 'Step 3: Define Columns'. The 'Step 1: Choose File' section contains the following form fields and text:

- Table Name:** . Text below: 'Name of the new table. Table names must be globally unique. Table names tend to correspond to the directory where the data will be stored.'
- Description:** . Text below: 'Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.'
- Input File:** . Text below: 'The HDFS path to the file on which to base this new table definition. It can be compressed (gzip) or not.'
- Import data from file:** . Text below: 'Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.'

A yellow warning box at the bottom states: 'Warning: The selected file is going to be moved during the import.'



The screenshot shows the Hue Metastore Manager interface. The top navigation bar includes 'HUE', 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. The main header is 'Metastore Manager'. On the left, a sidebar shows a tree view with 'SQL' and 'dw\_project' (containing tables: account, appliances, daily\_consumption, houses, rooms). The main content area is titled 'Step 1: Choose File' and 'Name Your Table and Choose A File'. It contains form fields for 'Table Name' (activity), 'Description' (Optional), and 'Input File' (/user/cloudera/activity.csv). There is a checked 'Import data from file' checkbox and a yellow warning box stating 'Warning: The selected file is going to be moved during the import.'



The screenshot shows the Hue Metastore Manager interface at 'Step 2: Choose Delimiter'. The breadcrumb path is 'Databases > dw\_project > Create a new table from a file'. The main content area is titled 'Choose a Delimiter' and states 'Beeswax has determined that this file is delimited by commas.' A dropdown menu shows 'Comma (,)' selected, with a 'Preview' button next to it. Below this is a 'Table preview' table with columns 'col\_1', 'col\_2', 'col\_3', and 'col\_4'. The table contains three rows of data. At the bottom, there are 'Previous' and 'Next' buttons.

col_1	col_2	col_3	col_4
account_id	first_name	last_name	email
0	Muhammad	Yahya	muhammadyahya@gmail.com
1	Salman	Hanif	salmanhanif@gmail.com

## Queries on Hive:

### Basic Queries:

#### 1) Top 5 Power Consumption Activities in the House

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL query:

```
1 select appliances.*,activity.* from activity join appliances where appliances.sensor_id = activity.appliance_id
2
3
4
5
6 SORT BY activity.power_consumption DESC LIMIT 5
```

The query is executed, and the results are displayed in a table with the following columns: appliances.appliance\_name, appliances.rating, appliances.status, appliances.add\_time, activity.activity\_id, activity.appliance\_id, and activity.current\_state. The results show the top 5 power-consuming appliances, all of which are Fridges.

appliance_name	rating	status	add_time	activity_id	appliance_id	current_state
Fridge	0	active		882127	2	
Fridge	0	active		1841402	2	
Fridge	0	active		1690362	2	
Fridge	0	active		1683177	2	
Fridge	0	active		903647	2	

#### 2) Average Electrical Consumption of the appliances for the whole week

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL query:

```
1
2
3 select AVG(activity.power_consumption) as Average,appliances.appliance_name from activity join appliances
4 where appliances.sensor_id = activity.appliance_id
5 GROUP BY appliances.sensor_id,appliances.appliance_name
```

The query is executed, and the results are displayed in a table with the following columns: average and appliances.appliance\_name. The results show the average power consumption for five different appliances.

average	appliance_name
0.031367524981226927	Dishwasher
0.063556410069655483	Microwave
0.010982993450237016	Fridge
0.035312807287801264	Living_Room_Sensor
0.081286892420213808	Home_Office

3) Average Electrical Consumption of the appliances for the day

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL query:

```
1  
2  
3 select SUM(activity.power_consumption)/(7*24) as AveragekwhperHr,appliances.appliance_name from activity join appliances  
4 where appliances.sensor_id = activity.appliance_id  
5 GROUP BY appliances.sensor_id,appliances.appliance_name
```

The results table shows the following data:

averagekwhperhr	appliances.appliance_name
1 94.08577091244085	Dishwasher
2 190.63518213214343	Microwave
3 32.943096604219853	Fridge
4 105.91950428807104	Living_Room_Sensor
5 243.81713071113057	Home_Office

4) Average of the total electric consumption with respect to Room

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL query:

```
1  
2  
3 select SUM(activity.power_consumption)/(7*24) as AveragekwhperHr,rooms.room_id from activity join appliances join rooms  
4 where appliances.sensor_id = activity.appliance_id and appliances.room_id = rooms.room_id  
5 GROUP BY rooms.room_id
```

The results table shows the following data:

averagekwhperhr	rooms.room_id
1 317.66404964812068	0
2 105.91950428807104	1
3 243.81713071113057	2

5) Sum of the total electric consumption with respect to Room

The screenshot shows the Hive Editor interface in a Firefox browser. The query editor contains the following SQL code:

```
1  
2  
3 select SUM(activity.power_consumption) as TotalCount,rooms.room_id from activity join appliances join rooms  
4 where appliances.sensor_id = activity.appliance_id and appliances.room_id = rooms.room_id  
5 GROUP BY rooms.room_id
```

Below the query editor, there are buttons for 'Execute', 'Save', 'Save as...', 'Explain', 'Format', and 'New query'. The 'Results' tab is active, displaying the following data:

	totalcount	rooms.room_id
1	53367.560340884273	0
2	17794.476720395935	1
3	40961.277959469939	2

The left sidebar shows a table catalog with categories like 'dw\_project', 'account', 'activity', 'appliances', 'daily\_consumption', and 'houses', each with its respective columns and data types listed.