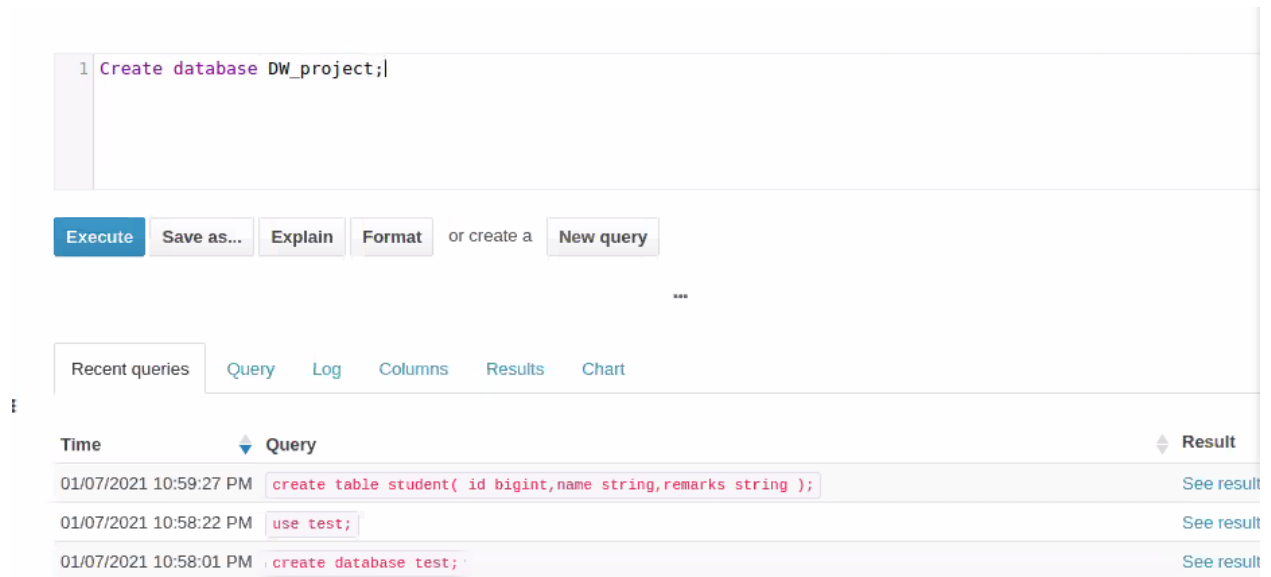


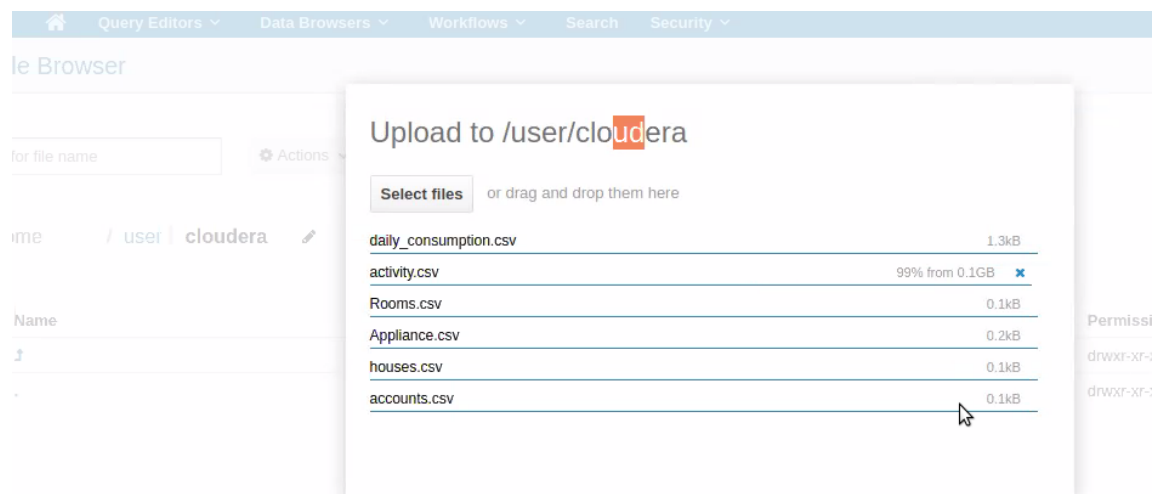
Dimensional Modeling in Hive



The screenshot shows a Hive query editor interface. At the top, there is a text area with the 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 table with the following columns: **Time**, **Query**, and **Result**. The table contains three rows of query execution history:

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

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 a Cloudera file browser interface. A dialog box titled "Upload to /user/cloudera" is open, displaying a list of files to be uploaded:

File 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 File Browser interface shows a search bar and a list of files in the /user/cloudera directory. The files are:

Name	Size	User	Group	Permissions
Appliance.csv	204 bytes	cloudera	cloudera	-rw-r--r--
Rooms.csv	143 bytes	cloudera	cloudera	-rw-r--r--
accounts.csv	146 bytes	cloudera	cloudera	-rw-r--r--
activity.csv	147.8 MB	cloudera	cloudera	-rw-r--r--
daily_consumption.csv	1.3 KB	cloudera	cloudera	-rw-r--r--
houses.csv	92 bytes	cloudera	cloudera	-rw-r--r--

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 Metastore Manager interface shows the 'Choose a Delimiter' step for creating a new table from a file. The delimiter is set to 'Comma (,)'.

Table preview:

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

HUE

Query Editors ▾

Data Browsers ▾

Workflows ▾

Search

Security ▾

Metastore Manager

SQL

dw_project

Tables

account

Databases > dw_project > Create a new table from a file

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

Name Your Table and Choose A File

Table Name

houses

Name of the new table. Table names must be globally unique. Table names tend to correspond to the data stored.

Description

Optional

Use a table comment to describe the table. For example, note the data's provenance and any caveats.

Input File

/user/cloudera/houses.csv

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

☒

Check this box to import the data in this file after creating the table definition. Leave it unchecked to defer.

Warning: The selected file is going to be moved during the import.

HUE

Query Editors ▾

Data Browsers ▾

Workflows ▾

Search

Security ▾

Metastore Manager

SQL

dw_project

Tables

account

Databases > dw_project > Create a new table from a file

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

Choose a Delimiter

Beeswax has determined that this file is delimited by commas.

Delimiter

Comma (,)

Preview

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

Table preview

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

Previous

Next

HUE

Query Editors ▾

Data Browsers ▾

Workflows ▾

Search

Security ▾

Metastore Manager

SQL

dw_project

Tables

account

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

Tables

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

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

dw_project

Tables

account

houses

Databases > dw_project > Create a new table from a file

Step 1: Choose File

Step 2: Choose Delimiter

Step 3: Define Columns

Name Your Table and Choose A File

Table Name

rooms

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

Optional

Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.

Input File

/user/cloudera/Rooms.csv

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

☒

Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.

Warning: The selected file is going to be moved during the import.

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

dw_project

Tables

account

houses

Databases > dw_project > Create a new table from a file

Step 1: Choose File

Step 2: Choose Delimiter

Step 3: Define Columns

Choose a Delimiter

Beeswax has determined that this file is delimited by commas.

Delimiter

Comma (,)

Preview

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

Table preview

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

PreviousNext

Metastore Manager

SQL

dw_project

Tables

account

houses

Define your columns

Use first row as column names

Bulk edit column names

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

Metastore Manager

SQL

dw_project

Tables

account

houses

rooms

Choose a Delimiter

Beeswax has determined that this file is delimited by commas.

Delimiter

Comma (,)

Preview

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

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	

PreviousNext

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

dw_project

Tables

account

houses

rooms

Define your columns

Use first row as column names

Bulk edit column names

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		

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

dw_project

Tables

account

appliances

houses

rooms

Step 1: Choose File

Step 2: Choose Delimiter

Step 3: Define Columns

Name Your Table and Choose A File

Table Name

daily_consumption

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

Optional

Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.

Input File

/user/cloudera/daily_consumption.csv

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

☒

Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.

Warning: The selected file is going to be moved during the import.

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

< dw_project

Tables (5)

account

appliances

daily_consumption

houses

rooms

Step 1: Choose File

Step 2: Choose Delimiter

Step 3: Define Columns

Name Your Table and Choose A File

Table Name

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

Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.

Input File

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 ☒

Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.

Warning The selected file is going to be moved during the import.

HUE

Query Editors

Data Browsers

Workflows

Search

Security

Metastore Manager

SQL

< dw_project

Tables (0)

The database has no tables

Databases > dw_project > Create a new table from a file

Step 1: Choose File

Step 2: Choose Delimiter

Step 3: Define Columns

Choose a Delimiter

Beeswax has determined that this file is delimited by commas.

Delimiter

Preview

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

Table preview

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

Previous Next

Queries on Hive:

Basic Queries:

1) Top 5 Power Consumption Activities in the House

The screenshot shows the Hue Hive Query 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 SORT BY activity.power_consumption DESC LIMIT 5
4
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 consumption activities, all of which are Fridges.

	appliances.appliance_name	appliances.rating	appliances.status	appliances.add_time	activity.activity_id	activity.appliance_id	activity.current_state
1	Fridge	0	active		882127	2	
2	Fridge	0	active		1841402	2	
3	Fridge	0	active		1690362	2	
4	Fridge	0	active		1683177	2	
5	Fridge	0	active		903647	2	

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

The screenshot shows the Hue Hive Query 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 electrical consumption for each appliance over the whole week.

	average	appliances.appliance_name
1	0.031367524981226927	Dishwasher
2	0.063556410069655483	Microwave
3	0.010982993450237016	Fridge
4	0.035312807287801264	Living_Room_Sensor
5	0.081286892420213808	Home_Office

3) Average Electrical Consumption of the appliances for the day

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

```
1 select SUM(activity.power_consumption)/(7*24) as AveragekwperHr, appliances.appliance_name from activity join appliances
2
3
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 two columns: **averagekwperhr** and **appliances.appliance_name**.

averagekwperhr	appliances.appliance_name
94.08577091244085	Dishwasher
190.63518213214343	Microwave
32.943096604219853	Fridge
105.91950428807104	Living_Room_Sensor
243.81713071113057	Home_Office

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

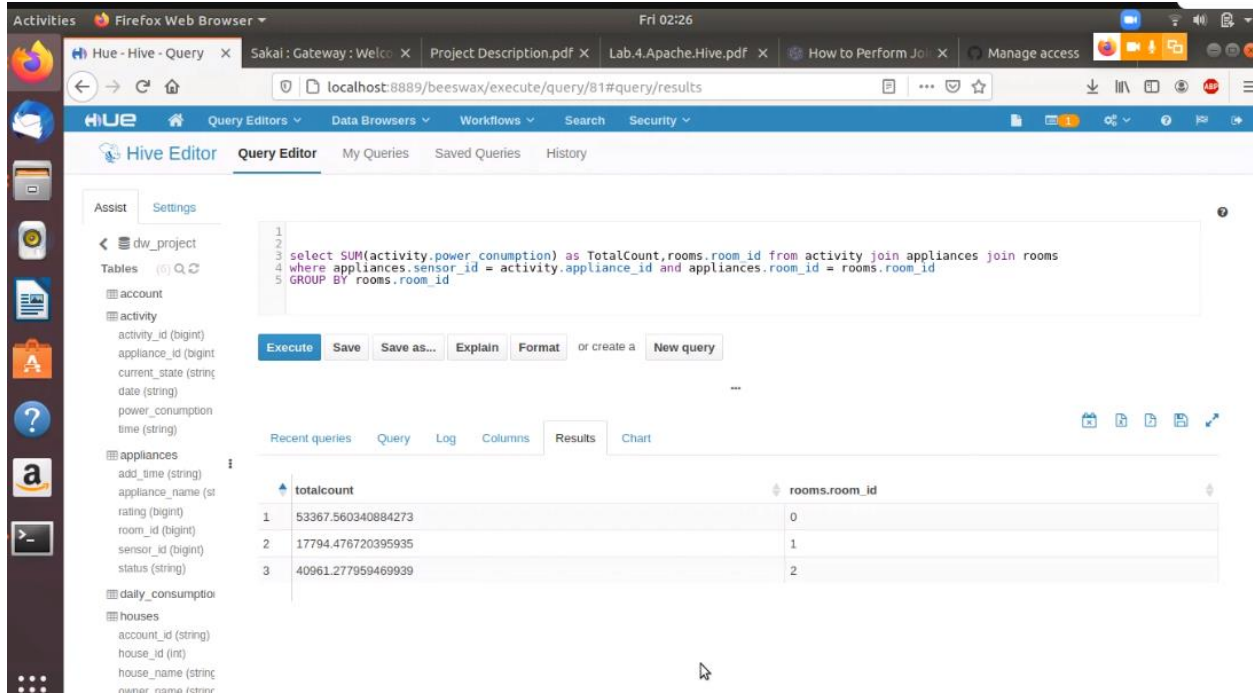
The screenshot shows the Hue Hive Query Editor interface. The query editor contains the following SQL query:

```
1
2
3 select SUM(activity.power_consumption)/(7*24) as AveragekwperHr, 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 query is executed, and the results are displayed in a table with two columns: **averagekwperhr** and **rooms.room_id**.

averagekwperhr	rooms.room_id
317.66404964812068	0
105.91950428807104	1
243.81713071113057	2

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



The screenshot displays the Hue Hive Editor interface within a Firefox Web Browser. The browser's address bar shows the URL `localhost:8889/ beeswax/execute/query/81#query/results`. The interface includes a sidebar with a table catalog on the left and a main workspace for editing queries. The query editor contains the following SQL code:

```
1 select SUM(activity.power_consumption) as TotalCount,rooms.room_id from activity join appliances join rooms
2
3
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 are buttons for `Execute`, `Save`, `Save as...`, `Explain`, `Format`, and `New query`. The `Results` tab is active, showing a table with two columns: `totalcount` and `rooms.room_id`. The table contains three rows of data:

	totalcount	rooms.room_id
1	53367.56034084273	0
2	17794.476720395935	1
3	40961.277959469939	2