



Skills Network

Project : Database Design and Implementation

Estimated time needed: 90 minutes

Congratulations!! You have finished the modules. Now is the time to put your skills to test. Read through the scenario below.

Scenario

In this scenario, you have recently been hired as a Data Engineer by a New York based coffee shop chain that is looking to expand nationally by opening a number of franchise locations. As part of their expansion process, they want to streamline operations and revamp their data infrastructure.

Your job is to design their relational database systems for improved operational efficiencies and to make it easier for their executives to make data driven decisions.

Currently their data resides in several different systems: accounting software, suppliers databases, point of sales (POS) systems, and even spreadsheets. You will review the data in all of these systems and design a central database to house all of the data. You will then create the database objects and load them with source data. Finally, you will create subsets of data that your business partners require, export them, and then load them into staging databases that use different RDBMS.

Software used in this project

In this project, you will use [PostgreSQL Database](#), Datasette, and [MySQL](#). These are all relational database management systems (RDBMS) designed to efficiently store, manipulate, and retrieve the data.



Data used in this project

In this project, you will be working with a subset of data from the [Coffee shop sample data](#).

You will use a modified version of the data for the project, so to succeed in the project, download the linked files when prompted in the instructions. You do not need to use any data from the original source.

In your scenario, you will be working with data from the following sources:

- Staff information held in a spreadsheet at HQ
- Sales outlet information held in a spreadsheet at HQ
- Sales data output as a CSV file from the POS system in the sales outlets
- Customer data output as a CSV file from a bespoke customer relationship management system
- Product information maintained in a spreadsheet exported from your supplier's database

Note: Perform all the **tasks(1-10) except task 9** by launching this [lab](#) in a new browser tab and then complete **task 9** by following the instructions mentioned in this optional lab using **Datasette**.

Objectives

After completing this lab, you will be able to:

- Identify entities.
- Identity attributes.
- Create an entity relationship diagram (ERD) using the pgAdmin ERD Tool.
- Normalize tables.
- Define keys and relationships.
- Create database objects by generating and running the SQL script from the ERD Tool.
- Create a view and export the data.
- Create a materialized view and export the data.
- Import data into Datasette.
- Import data into a MySQL database.

Task 1: Identify entities

The first step when designing a new database is to review any existing data and identify the entities for your new system.

1. The following image shows sample data from each of the data sources that you will be working with to design your new central database. Review the image and identify the entities you plan to create.

staff					
staff_id	first_name	last_name	position	start_date	location
1	Sue	Tindale	CFO	08/03/2001	HQ
2	Ian	Tindale	CEO	3/8/2001	HQ
3	Marny	Hermione	Roaster	10/24/2007	WH
4	Chelsea	Claudia	Roaster	3/7/2003	WH
5	Alec	Isadora	Roaster	2/4/2008	WH
6	Xena	Rahim	Store Manager	7/24/2016	3
7	Kelsey	Cameron	Coffee Wrangler	10/18/2003	3
8	Hamilton	Emi	Coffee Wrangler	9/2/2005	3
9	Caldwell	Veda	Coffee Wrangler	9/9/2013	3
10	Ima	Winifred	Coffee Wrangler	10/12/2016	3

sales outlet						
sales outlet id	sales outlet type	address	city	telephone	postal code	manager
2	warehouse	164-14 Jamaica Ave	Jamaica	972-871-0402	11432	
3	retail	32-20 Broadway	Long Island City	777-718-3190	11106	6
4	retail	604 Union Street	Brooklyn	619-347-5193	11215	11
5	retail	100 Church Street	New York	343-212-5151	10007	16

sales transaction								
transaction id	transaction date	transaction time	sales outlet id	staff id	customer id	product id	quantity	price
1	27/04/2019	09:53:55	8	42	0	38	2	3.75
1	27/04/2019	09:53:55	8	42	0	84	1	0.8
2	27/04/2019	08:00:34	8	42	0	51	2	3
3	27/04/2019	09:04:58	8	42	0	33	1	3.5
4	27/04/2019	08:48:32	8	42	8232	27	1	3.5
5	27/04/2019	09:21:40	8	45	8223	24	1	3

customer						
customer id	customer name	customer_email	customer since	customer card number	birthdate	gender
3001	Kelly Key	Venus@adipiscing.edu	04/01/2017	908-424-2890	29/05/1950	M
3002	Clark Schroeder	Nora@fames.gov	07/01/2017	032-732-6308	30/07/1950	M
3003	Elvis Cardenas	Brianna@tellus.edu	10/01/2017	459-375-9187	30/09/1950	M
3004	Rafael Estes	Ina@non.gov	13/01/2017	576-640-9226	01/12/1950	M
3005	Colin Lynn	Dale@Integer.com	15/01/2017	344-674-6569	01/02/1951	M

product					
product id	product category	product type	product name	description	price
1	Coffee beans	Organic Beans	Brazilian - Organic	It's like Carnival in a cup. Clean and smooth.	18
2	Coffee beans	House blend Beans	Our Old Time Diner Blend	Our packed blend of beans that is reminiscent of the cup of coffee you used to get at a diner.	18
3	Coffee beans	Espresso Beans	Espresso Roast	Our house blend for a good espresso shot.	14.75
4	Coffee beans	Espresso Beans	Primo Espresso Roast	Our premium single source of hand roasted beans.	20.45
5	Coffee beans	Gourmet Beans	Columbian Medium Roast	A smooth cup of coffee any time of day.	15
6	Coffee beans	Gourmet Beans	Ethiopia	From the home of coffee.	21

- Note: You might find it useful to download a copy of this image or open it in another browser tab for reference later in the lab.

2. Make a list of the entities you have identified.

Task 2: Identify attributes

In this task, you will identify the attributes for one of the entities that you plan to create.

1. Using the information from the sample data in the image from Task 1, identify the attributes for the entity that will store the sales transaction data.
2. Make a list of the sales transaction attributes that you identified.

Task 3: Create an ERD

Now that you have defined some of your attributes and entities, you can determine the tables and columns for them and create an ERD.

1. Open a new terminal from the side-by-side Cloud IDE.
2. Use the `start_postgres` command to start a PostgreSQL service session in the Cloud IDE.
3. Use the pgAdmin weblink to open pgAdmin in a new tab in your browser.

4. Create a new database named **COFFEE**, view the schemas in the new **COFFEE** database, and then start a new ERD project.
5. Add a table to the ERD for the sale transactions entity using the information in the following table. Consider what naming convention to use so that your colleagues will be able to understand your data and to ensure that the names are valid in other RDBMS. And use the sample data shown in the image in Task 1 to determine appropriate data types for each column.

Sales transaction
Transaction id
Date
Time
Sales outlet
Staff
Customer
Product
Quantity
Price

6. Take a screenshot of your ERD and save it as Task3A.png or Task3A.jpg.
7. Add a table to the ERD for the product entity using the information in the following table. Consider what naming convention to use so that your colleagues will be able to understand your data and to ensure that the names are valid in other RDBMS. And use the sample data shown in the image in Task 1 to determine appropriate data types for each column.

Product
Product id
Category
Type
Product
Description
Price

8. Take a screenshot of your ERD and save it as Task3B.png or Task3B.jpg.

Task 4: Normalize tables

When reviewing your ERD you notice that it does not conform to second normal form. In this task, you will normalize some of the tables within the database.

1. Review the data in the sales transaction table. Note that the transaction id column does not contain unique values because some transactions include multiple products.
2. Determine which columns should be stored in a separate table to remove the repeating rows and to put this table into second normal form.
3. Add a new table named **sales_detail** to the ERD, define the columns in the new table, and delete the moved columns from the sales transaction table, leaving a matching column in each of two tables to later create a relationship between them.
4. Take a screenshot of your ERD and save it as Task4A.png or Task4A.jpg.
5. Review the data in the product table. Note that the product category and product type columns contain redundant data.
6. Determine which columns should be stored in a separate table to reduce redundant data and to put this table into second normal form.
7. Add a new table named **product_type** to the ERD, define the columns in the new table, and delete the moved columns from the product table, , leaving a matching column in each of two tables to later create a relationship between them.
8. Take a screenshot of your ERD and save it as Task4B.png or Task4B.jpg.

Task 5: Define keys and relationships

After normalizing your tables, you can define their primary keys and define relationships between the tables in your ERD.

1. Identify an appropriate column in each table to be a primary key and create the primary keys in the tables in your ERD.

2. Take a screenshot of your ERD and save it as Task5A.png or Task5A.jpg.
3. Identify the relationships between the following pairs of tables and then create the relationships in your ERD:
 - `sales_detail` to `sales_transaction`
 - `sales_detail` to `product`
 - `product` to `product_type`
4. Take a screenshot of your ERD and save it as Task5B.png or Task5B.jpg.

Task 6: Create database objects by generating and running the SQL script from the ERD Tool

Now that your design is complete, you will generate an SQL script from your ERD which you could use to create your database schema. For the purposes of this project, you will then use a provided SQL script to ensure that you will be able to successfully load the sample data into the schema. Finally, you will load the existing data from the various data sources into your new database schema.

1. Use the Generate SQL functionality in the ERD Tool to create an SQL script from your ERD.
2. Download the `GeneratedScript.sql` file below to your local computer storage.
 - [GeneratedScript.sql](#)
3. In pgAdmin, open the Query Tool, upload and open the `GeneratedScript.sql` file from your local computer storage, and then execute the script to create the tables defined in the ERD. Verify that the tables now exist in the `public` schema of the `COFFEE` database.
4. Take a screenshot of the tables shown in the tree-view pane on the left-hand side of the page and save it as Task6A.png or Task6A.jpg.
5. Download the `CoffeeData.sql` file below to your local computer storage.
 - [CoffeeData.sql](#)
6. In pgAdmin, open another instance of the Query Tool, upload and open the `CoffeeData.sql` file from your local computer storage, and then execute the script to populate the tables you just created.
7. In pgAdmin, view the first 100 rows of the `sales_detail` table.
8. Take a screenshot of the Data Output pane and save it as Task6B.png or Task6B.jpg.

Task 7: Create a view and export the data

The external payroll company have requested a list of employees and the locations at which they work. This should not include the CEO or CFO who own the company. In this task, you will create a view in your PostgreSQL database that returns this information and export the results to a CSV file.

1. In your `COFFEE` database, create a new view named `staff_locations_view` using the following SQL:

```
SELECT staff.staff_id,
staff.first_name,
staff.last_name,
staff.location
FROM staff
WHERE "position" NOT IN ('CEO', 'CFO');
```

2. View all the rows returned from the view.
3. Save the results of the query to a file named `staff_locations_view.csv` on your local computer storage.
4. Take a screenshot of the view shown in the tree-view pane on the left-hand side of the page alongside the results in the Data Output pane, and save it as Task7.png or Task7.jpg.

Task 8: Create a materialized view and export the data

A marketing consultant requires access to your product data in their MySQL database for a marketing campaign. You will create a materialized view in your PostgreSQL database that returns this information and export the results to a CSV file.

1. In your **COFFEE** database, create a new materialized view named **product_info_m-view** using the following SQL:

```
SELECT product.product_name, product.description, product_type.product_category
FROM product
JOIN product_type
ON product.product_type_id = product_type.product_type_id;
```

2. Refresh the materialized view with data.
3. View all the rows returned from the view.
4. Save the results of the query to a file named **product_info_m-view.csv** on your local computer storage.
5. Take a screenshot of the view shown in the tree-view pane on the left-hand side of the page alongside the results in the Data Output pane, and save it as Task8.png or Task8.jpg.

Task 9: Import data into Datasette

The external payroll company have asked you to upload the staff location information to their Datasette database.

1. Copy and enter this URL: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0110EN-SkillsNetwork/Datasetteoptionallabs/Week4/data/staff_locations_view.csv of the staff_locations_view.csv file, which contains the staff location information.
2. Explore the new table and then view the data in it.
3. Take a screenshot of the contents of the new table and save it as Task9.png or Task9.jpg.

Task 10: Import data into a MySQL database

The marketing consultant has asked you to upload the product information to their MySQL database.

1. In the terminal from the side-by-side Cloud IDE, use the **start_mysql** command to start a My SQL service session in the Cloud IDE.
2. Use the browser weblink to open phpMyAdmin in a new tab in your browser.
3. In phpMyAdmin, create a new database named **coffee_shop_products**, and then import the product information saved in the **product_info_m-view.csv** file from your materialized view into a new table in the **coffee_shop_products** database.
4. Browse the contents of the new table.
5. Take a screenshot of the contents of the new table and save it as Task10.png or Task10.jpg.

Author(s)

- Nivedita

Other Contributor(s)

-

Changelog

Date	Version	Changed by	Change Description
2022-09-29	1.2	Appalabhaktula Hema	Updated CSV link and instructions

Date	Version	Changed by	Change Description
29-06-2022	1.1	Niveditha	Converted intial version to DataSette
2021-04-01	1.0	Lin Joyner	Created initial version

© IBM Corporation 2021. All rights reserved.