

Hands-on Lab: Populating a Data Warehouse using PostgreSQL



Estimated time needed: 1 hour

Purpose of the Lab:

The lab is designed to provide hands-on experience in creating and managing a production database using PostgreSQL within the IBM Skills Network Labs (SN Labs) Cloud IDE. You will learn how to launch a PostgreSQL server instance, utilize the pgAdmin graphical user interface (GUI) for database operations, and execute essential tasks like creating a database, designing tables, and loading data. The lab focuses on building a foundation in database management by guiding learners through the process of setting up a 'Production' database and populating it with data following a star schema design.

Benefits of Learning the Lab:

Engaging in this lab offers significant benefits for learners seeking to deepen their understanding of database management systems, particularly PostgreSQL. By working through the lab, you will gain practical skills in SQL, database creation, table design, and data manipulation, which are crucial for roles in data engineering, database administration, and data science. The hands-on approach helps in consolidating knowledge of database schemas and SQL queries, thereby enhancing the learner's ability to manage and analyze data effectively in real-world scenarios. Additionally, familiarity with tools like pgAdmin and the Cloud IDE environment adds valuable experience to your skill set, preparing you for advanced database projects and tasks.

Software Used in this Lab

To complete this lab you will utilize the [PostgreSQL Database](#) relational database service available as part of IBM Skills Network Labs (SN Labs) Cloud IDE. SN Labs is a virtual lab environment used in this course.



Database Used in this Lab

Production database is used in this lab.
The production database contains:

- DimCustomer
- DimMonth
- FactBilling

Objectives

In this lab you will:

- Create production related database and tables in a PostgreSQL instance.
- Populate the production data warehouse byloading the tables from Scripts.

Lab Structure

In this lab, you will complete several tasks in which you will learn how to create tables and load data in the PostgreSQL database service using the pgAdmin graphical user interface (GUI) tool.

Data Used in this Lab

The following are the SQL data files used in this lab.

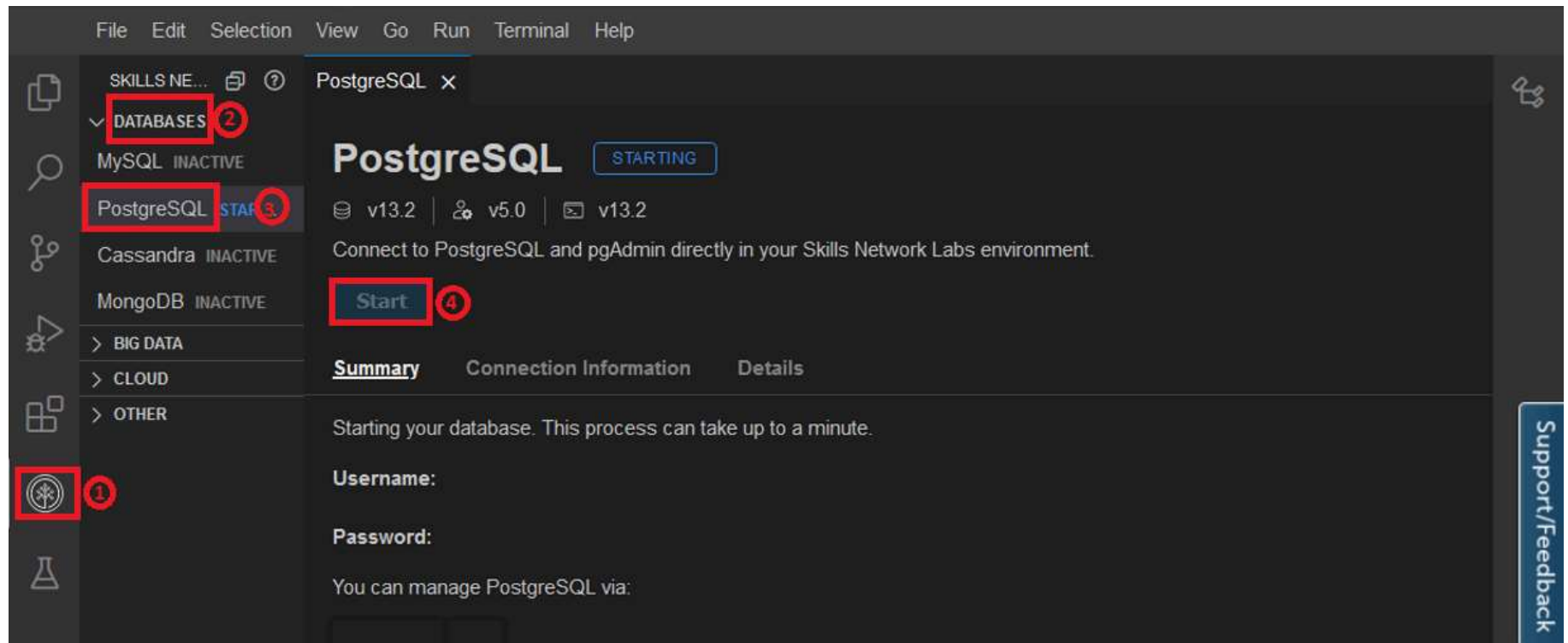
The production database contains:

- [DimCustomer](#)
- [DimMonth](#)
- [FactBilling](#)
- [Star Schema](#)

Task A: Create a database

First, to create a database on a PostgreSQL server instance, you'll first want to actually launch a PostgreSQL server instance on Cloud IDE and open up the pgAdmin Graphical User Interface.

1. Click on the Skills Network extension button on the left side of the window.
2. Open the **DATABASES** drop down menu.
3. Click on **PostgreSQL**
4. Click on the **Start** button. PostgreSQL may take a few moments to start.



5. Next, open the pgAdmin Graphical User Interface by clicking the **pgAdmin** button in the Cloud IDE interface.

The screenshot shows the IBM Cloud Skills Network Labs interface for a PostgreSQL instance. The top navigation bar includes 'Lab', 'IBMCLOUD', and 'Launch Application'. Below this is a menu with 'File', 'Edit', 'Selection', 'View', 'Go', 'Run', 'Terminal', and 'Help'. On the left sidebar, under 'SKILLS NETWORK', the 'DATABASES' section is expanded, showing 'MySQL INACTIVE', 'PostgreSQL ACTIVE', 'Cassandra INACTIVE', and 'MongoDB INACTIVE'. The 'PostgreSQL' instance is selected and shows its status as 'ACTIVE'. The main content area displays the PostgreSQL instance details, including version 'v13.2' and a 'Stop' button. Below this, there are tabs for 'Summary', 'Connection Information', and 'Details'. The 'Summary' tab is active, showing a message: 'Your database and pgAdmin server are now ready to use and available with the following login credentials. For more details on how to navigate PostgreSQL, please check out the Details section.' Below this message, there are input fields for 'Username' and 'Password', each with a copy icon. A section titled 'You can manage PostgreSQL via:' contains a red-bordered button labeled 'pgAdmin' with an external link icon, and a 'New Terminal' button. At the bottom, there are two buttons: 'PostgreSQL CLI' and 'New Terminal'. A vertical 'Support/Feedback' button is located on the right side of the interface.

Lab IBMCLOUD Launch Application

File Edit Selection View Go Run Terminal Help

SKILLS NETWORK

DATABASES

MySQL INACTIVE

PostgreSQL ACTIVE

Cassandra INACTIVE

MongoDB INACTIVE

> BIG DATA

> CLOUD

> OTHER

PostgreSQL

ACTIVE

v13.2 | v5.0 | v13.2

Connect to PostgreSQL and pgAdmin directly in your Skills Network Labs environment.

Stop

Summary Connection Information Details

Your database and pgAdmin server are now ready to use and available with the following login credentials. For more details on how to navigate PostgreSQL, please check out the Details section.

Username:

Password:

You can manage PostgreSQL via:

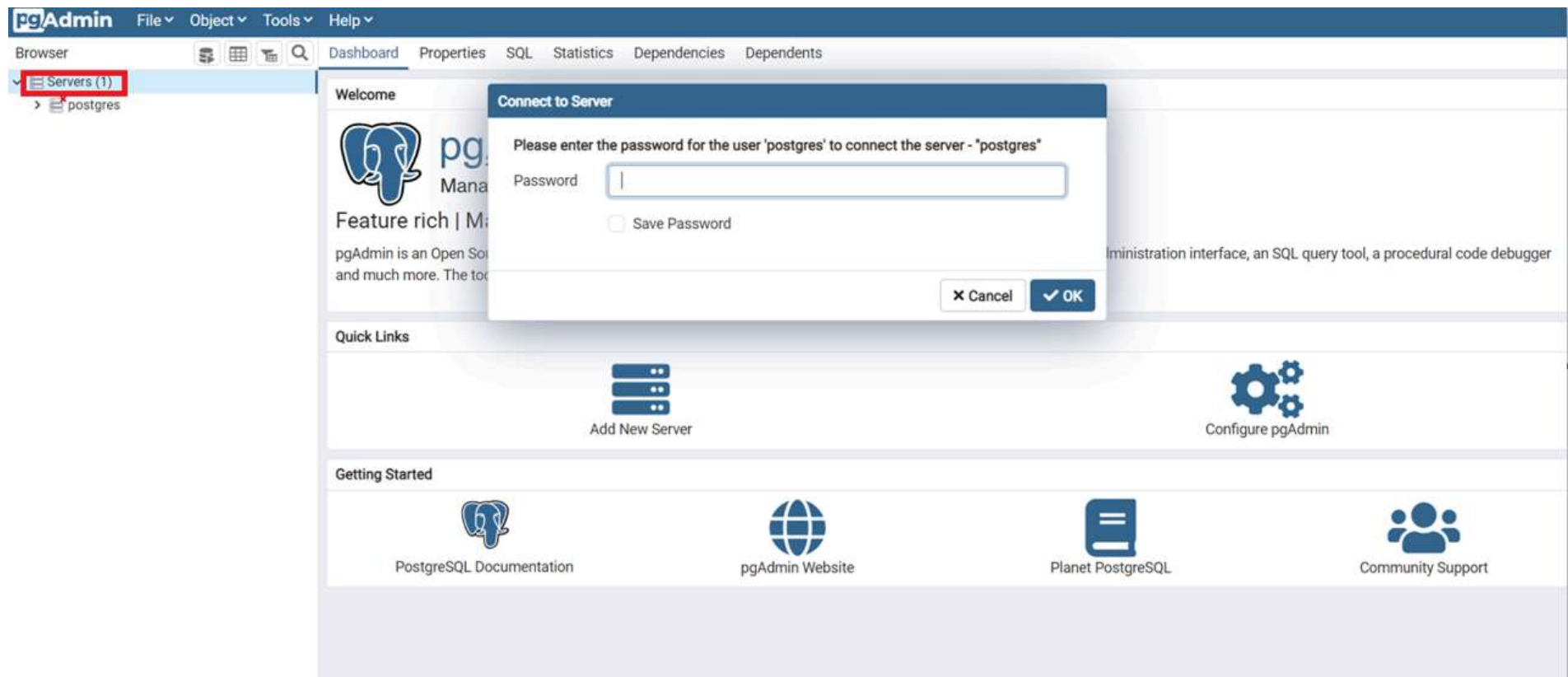
pgAdmin

Or to interact with the database in the terminal, select one of these options:

PostgreSQL CLI New Terminal

Support/Feedback

6. Once the pgAdmin GUI opens, click on the **Servers** tab on the left side of the page. You will be prompted to enter a password.



7. To retrieve your password, click on the **PostgreSQL** tab near the top of the interface.
8. Click on the **Copy** icon to the right of your password to copy the session password onto your clipboard.

Lab | IBMCloud | Launch Application

File Edit Selection View Go Run Terminal Help

SKILLS NE... ? PostgreSQL x

✓ DATABASES

MySQL INACTIVE

PostgreSQL ACTIVE

Cassandra INACTIVE

MongoDB INACTIVE

> BIG DATA

> CLOUD

> OTHER

PostgreSQL

ACTIVE

v13.2 | v5.0 | v13.2

Connect to PostgreSQL and pgAdmin directly in your Skills Network Labs environment.

Stop

Summary Connection Information Details

Your database and pgAdmin server are now ready to use and available with the following login credentials. For more details on how to navigate PostgreSQL, please check out the Details section.

Username: [REDACTED]

Password: [REDACTED]

You can manage PostgreSQL via:

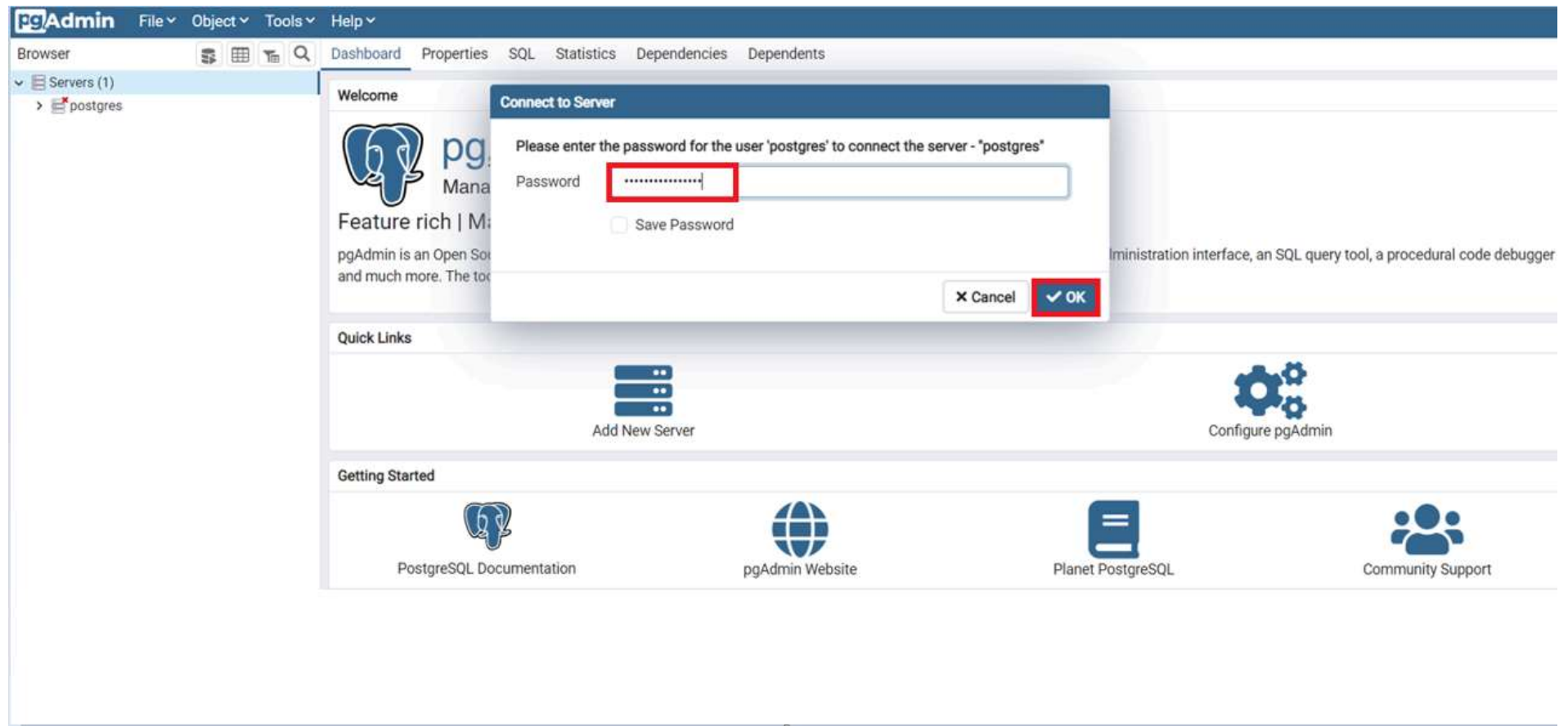
pgAdmin

Or to interact with the database in the terminal, select one of these options:

PostgreSQL CLI New Terminal

Support/Feedback

9. Navigate back to the **pgAdmin** tab and paste in your password, then click OK.



10. You will then be able to access the pgAdmin GUI tool.

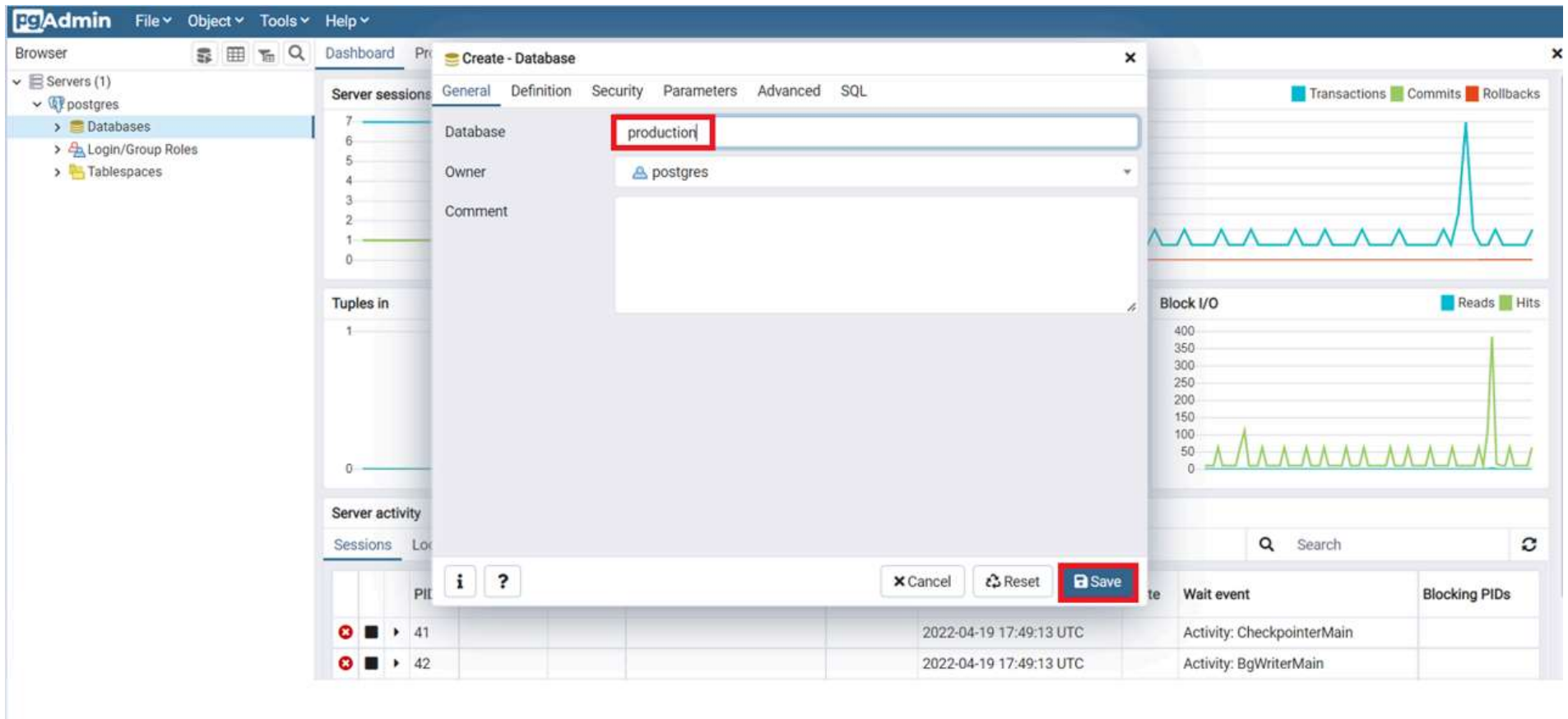
11. In the left tree-view, right-click on **Databases**> **Create** > **Database**.

The screenshot shows the pgAdmin interface with the following components:

- Browser:** A tree view on the left showing the hierarchy: Servers (1) > postgres > Databases. The 'Databases' folder is highlighted with a red circle (1). A context menu is open over it, with 'Create' highlighted with a red circle (2). The 'Create' menu item has a sub-menu 'Database...' which is also highlighted with a red circle (3).
- Server sessions:** A chart showing the number of sessions over time. The y-axis ranges from 0 to 7. The legend indicates Total (blue), Active (green), and Idle (red).
- Transactions per second:** A line chart showing the rate of transactions, commits, and rollbacks. The y-axis ranges from 0 to 3. The legend indicates Transactions (blue), Commits (green), and Rollbacks (red).
- Tuples in:** A chart showing the number of tuples inserted, updated, and deleted. The y-axis ranges from 0 to 1. The legend indicates Inserts (blue), Updates (green), and Delete (red).
- Tuples out:** A chart showing the number of tuples fetched and returned. The y-axis ranges from 0 to 1400. The legend indicates Fetched (blue) and Returned (green).
- Block I/O:** A chart showing the number of reads and hits. The y-axis ranges from 0 to 350. The legend indicates Reads (blue) and Hits (green).
- Server activity:** A table showing the current state of the server. It includes columns for PID, Database, User, Application, Client, Backend start, State, Wait event, and Blocking PIDs.

Server activity										
Sessions Locks Prepared Transactions Configuration										
		PID	Database	User	Application	Client	Backend start	State	Wait event	Blocking PIDs
✖	■	▶ 41					2022-04-19 17:49:13 UTC		Activity: CheckpointerMain	
✖	■	▶ 42					2022-04-19 17:49:13 UTC		Activity: BgWriterMain	
✖	■	▶ 43					2022-04-19 17:49:13 UTC		Activity: WalWriterMain	

In the Database box, type **Production** as the name for your new database, and then click **Save**. Proceed to Task B.

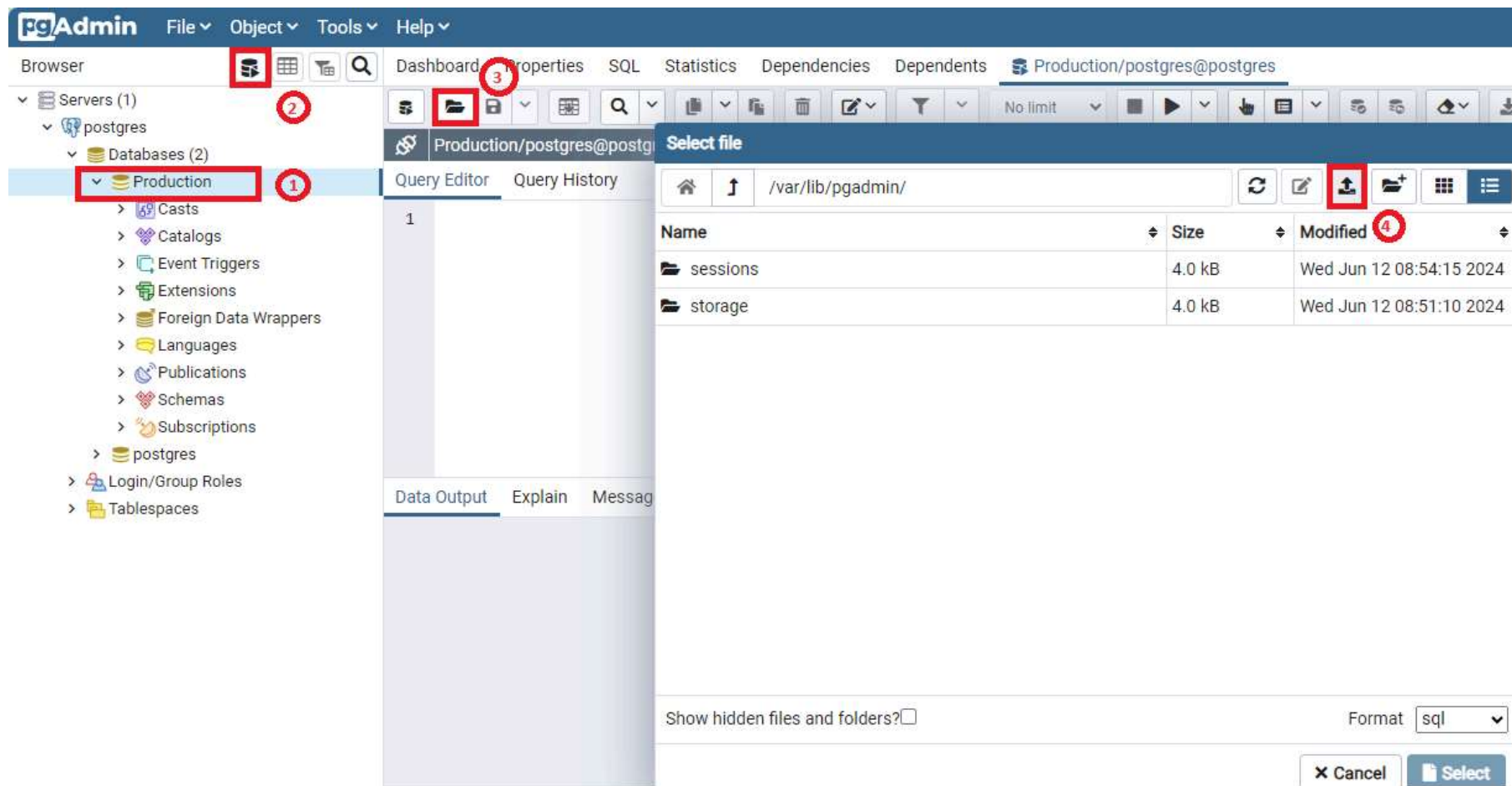


Task B: Create tables

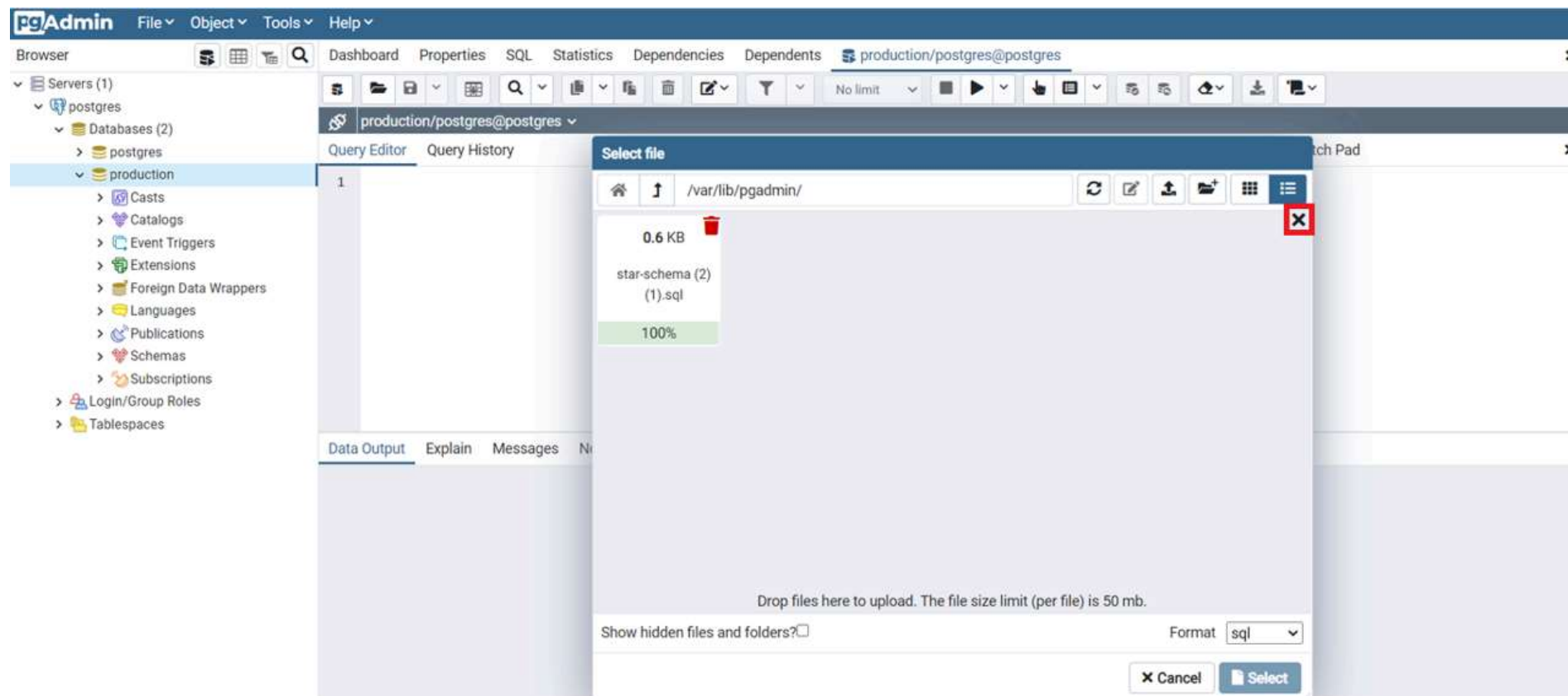
Now, that you have your PostgreSQL service active and have created the **Production database** using pgAdmin, let's go ahead and create a few tables to populate the database and store the data that we wish to eventually upload into it.

1. Click on the Production database and in the top of the page go to **Query tool** and then click on **Open File**. Next a new page pops up called **Select File**. Click on **Upload** icon as shown in the screenshot.

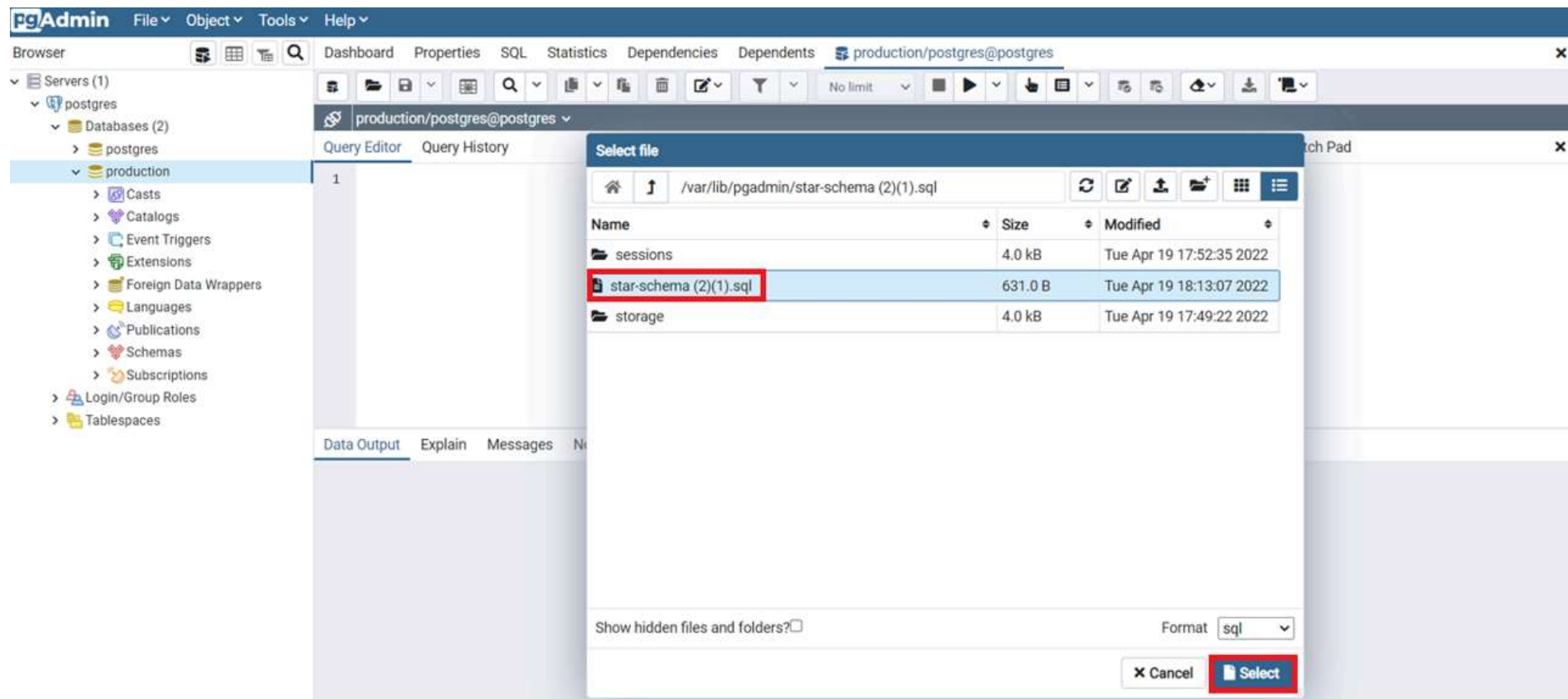
Note: Ensure that you upload the files to this path: /var/lib/pgadmin/



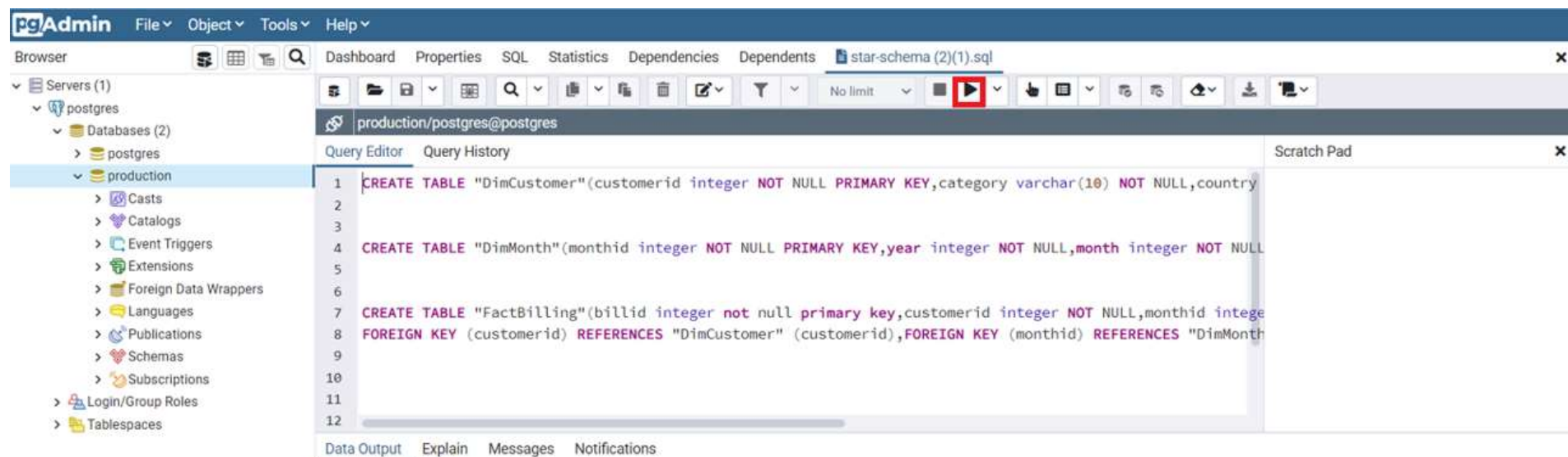
2. In the new blank page that appears drag and drop the **star-schema.sql** file inside the blank page. Once the **star-schema.sql** file is successfully loaded, click on the **X** icon on the right hand side of the page as shown in the screenshot.



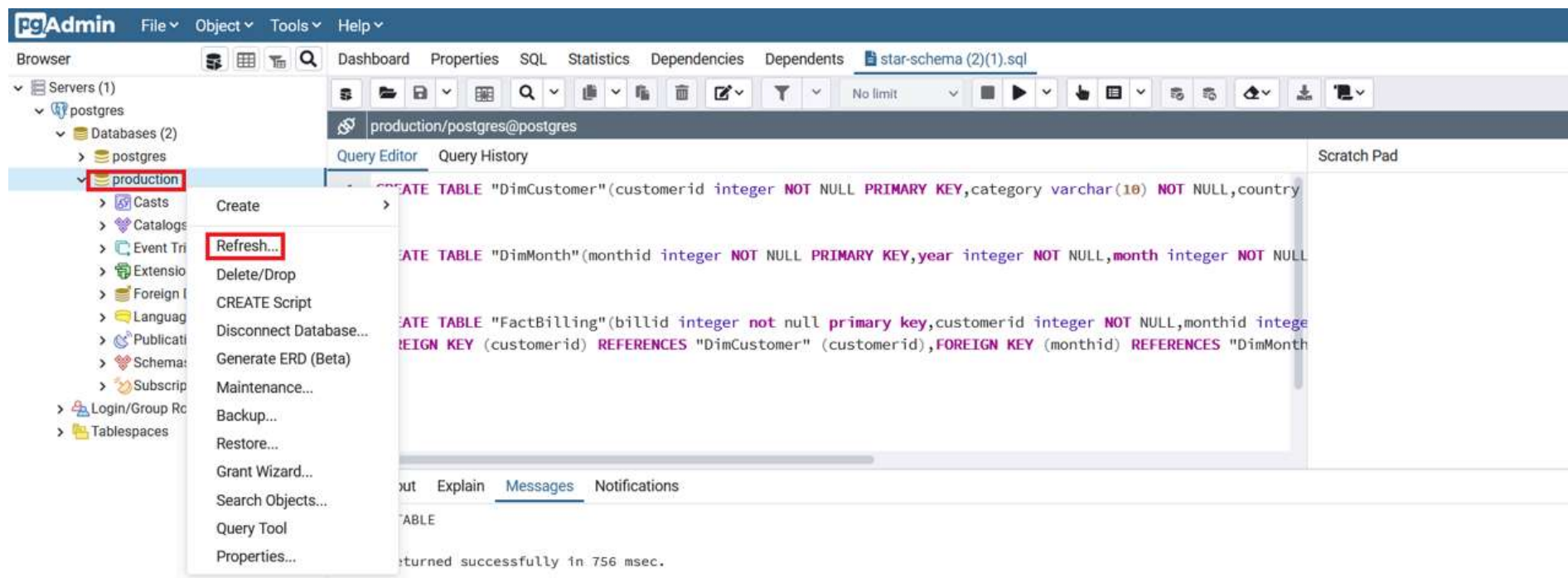
3. Once you click on the **X** icon a new page appears with the file **star-schema.sql**. Select the **star-schema.sql** file from the list and click the **Select** tab.



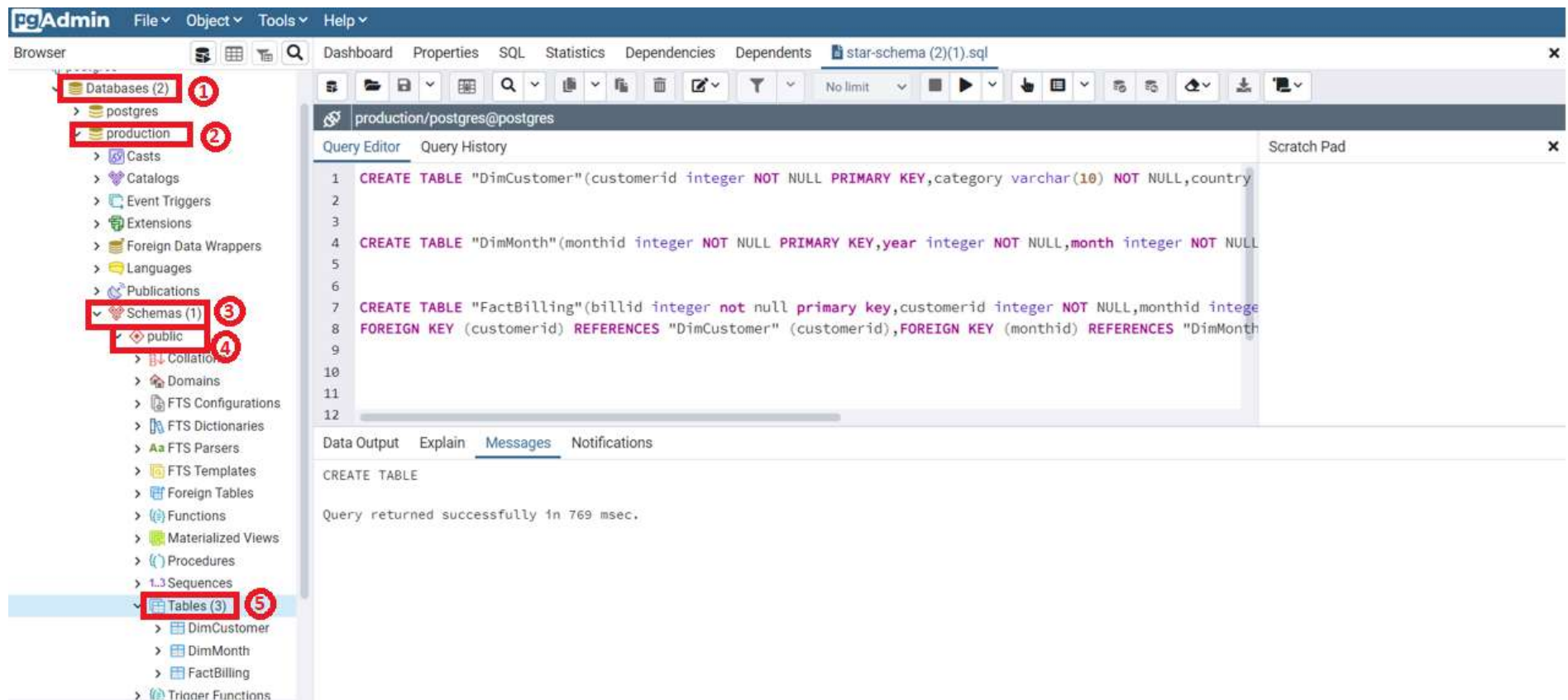
4. Once the file opens up click on the **Run** option to execute the **star-schema.sql** file.



5. Next, right-click on the **Production** database and click on the **Refresh** option from the dropdown.

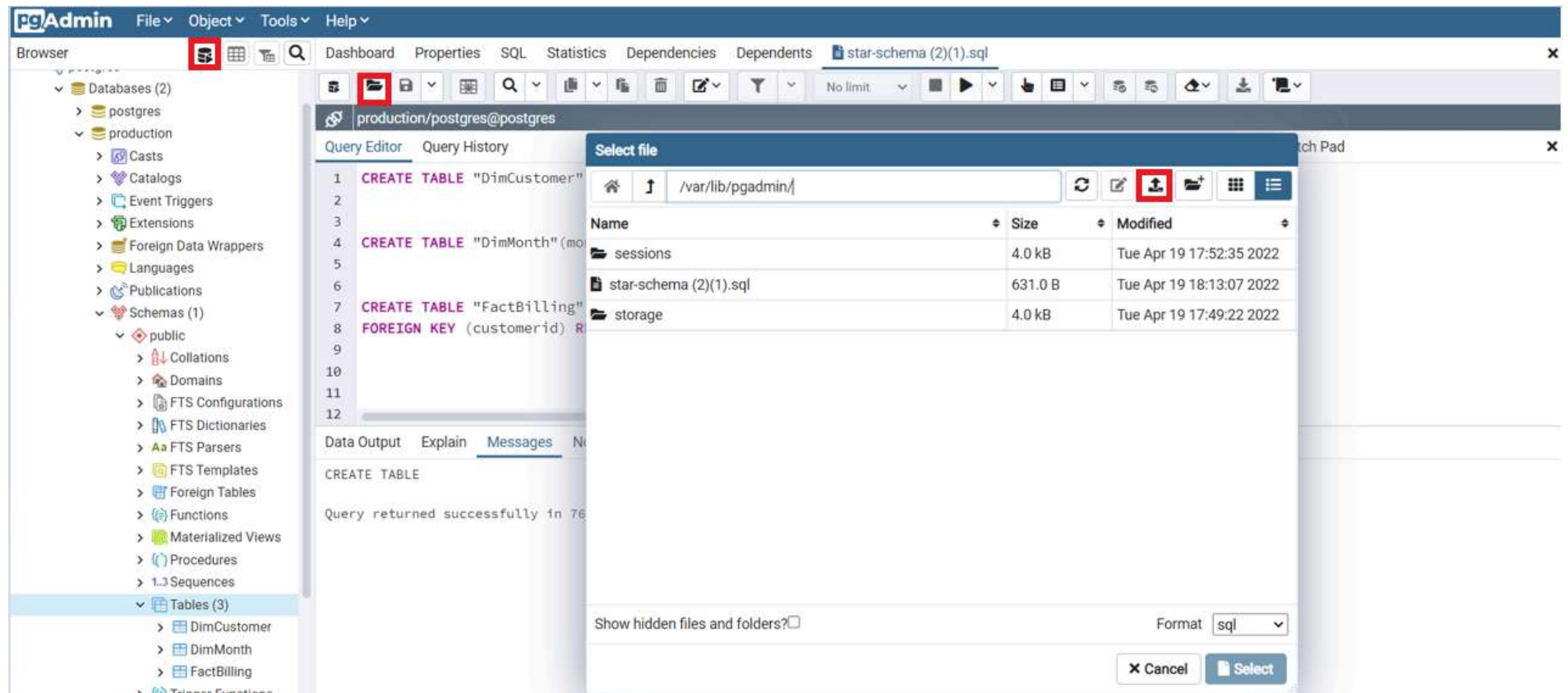


After the database is refreshed the 3 tables (DimCustomer, DimMonth, FactBilling) are created under the **Databases > Production > Schemas > Public > Tables**.



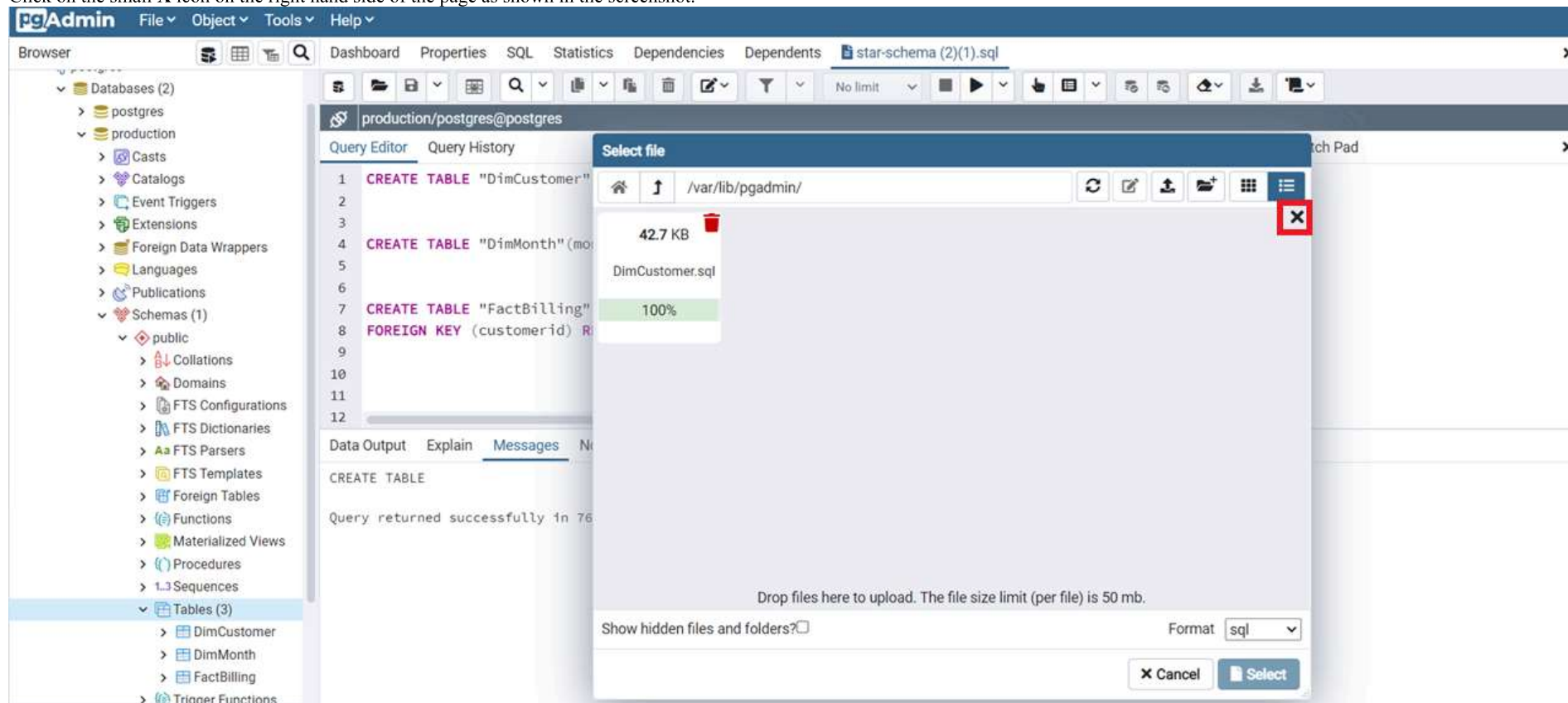
Task C: Load tables

1. Click on **Query tool** and then click **Open** file and click on **Upload** icon.

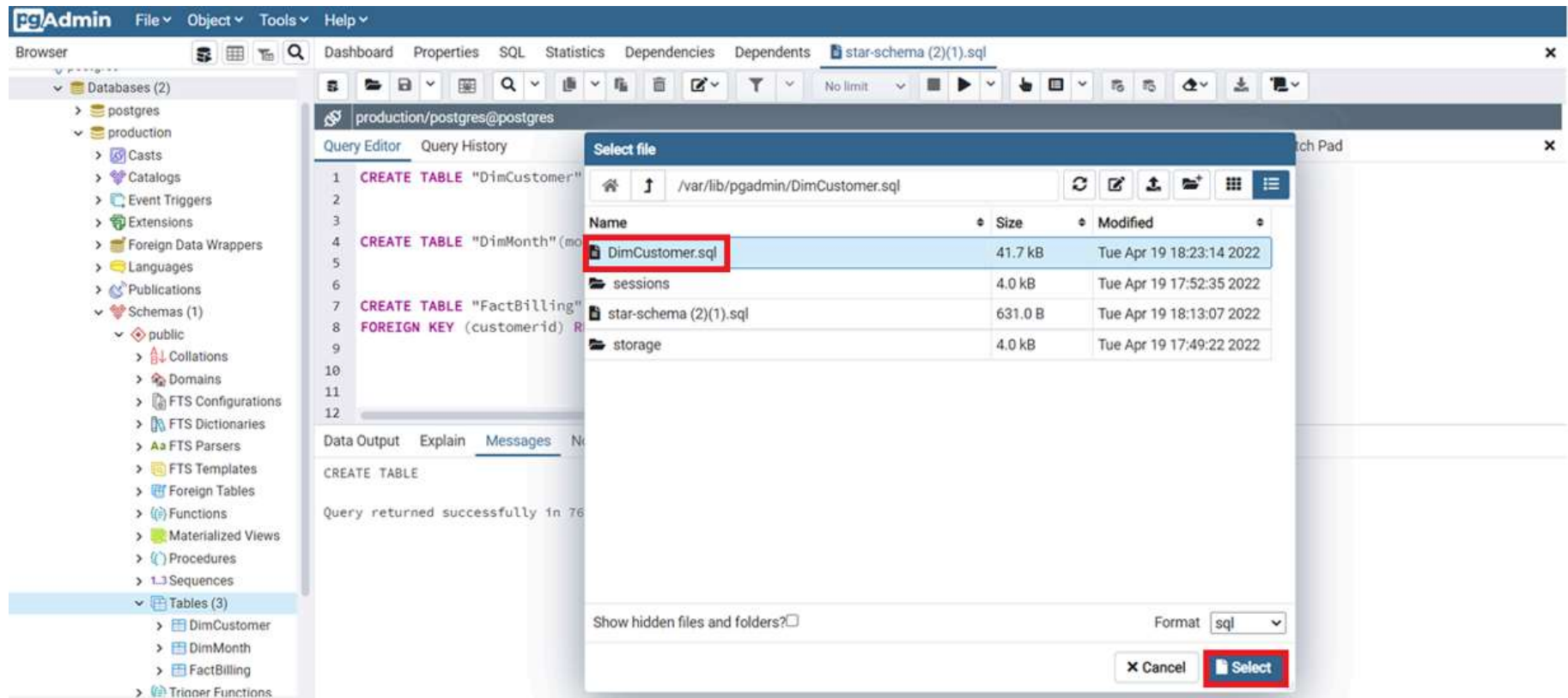


2. In the new blank page that appears drag and drop the **DimCustomer.sql** file inside the blank page. Once the **DimCustomer.sql** file is successfully loaded.

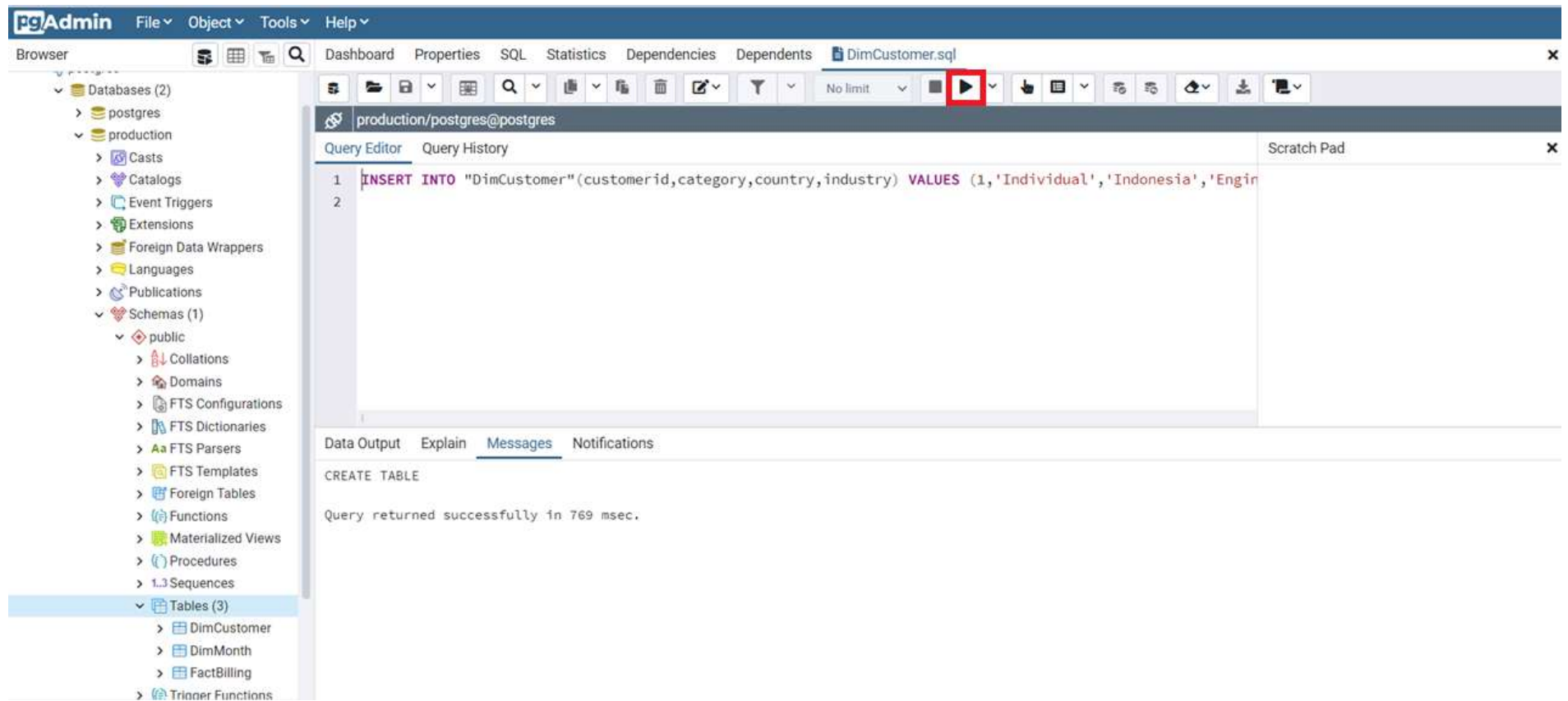
Click on the small **X** icon on the right hand side of the page as shown in the screenshot.



3. Once you click on the **X** icon a new page appears with the file **DimCustomer.sql**. Select the **DimCustomer.sql** file from the list and click on **Select** tab.



4. Once the file opens up, click on the **Run** option to execute the **DimCustomer.sql** file.



Note: Repeat the steps as given in Task C to upload the remaining sql files to insert data in **DimMonth** and **FactBilling**.

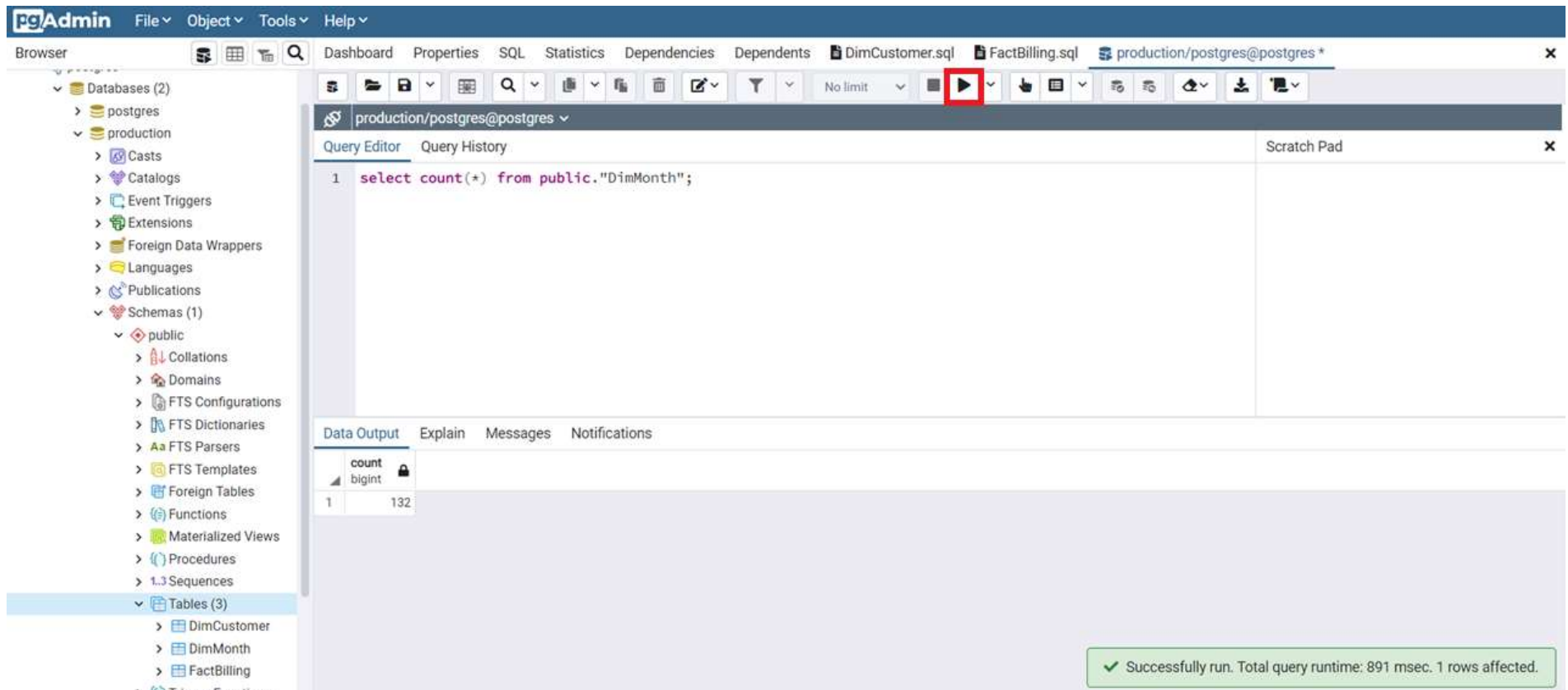
5. Let's run the command below on the PostgreSQL Tool.

1. 1

1. select count(*) from public."DimMonth";

Copied!

You should see an output as seen in the image below.



You are encouraged to run more SQL queries.

Practice exercises

Problem 1: Using the PostgreSQL tool, find the count of rows in the table FactBilling

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

Problem 2: Using the PostgreSQL tool, create a simple Materialized views named avg_customer_bill with fields customerid and averagebillamount.

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

Problem 3: Refresh the newly created Materialized views

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

Problem 4: Using the newly created Materialized views find the customers whose average billing is more than 11000.

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

Congratulations! You have successfully finished the Populating a Data Warehouse lab.

Author

Amrutha Rao

© IBM Corporation 2022. All rights reserved.