

# DataCo Supply Chain Documentation

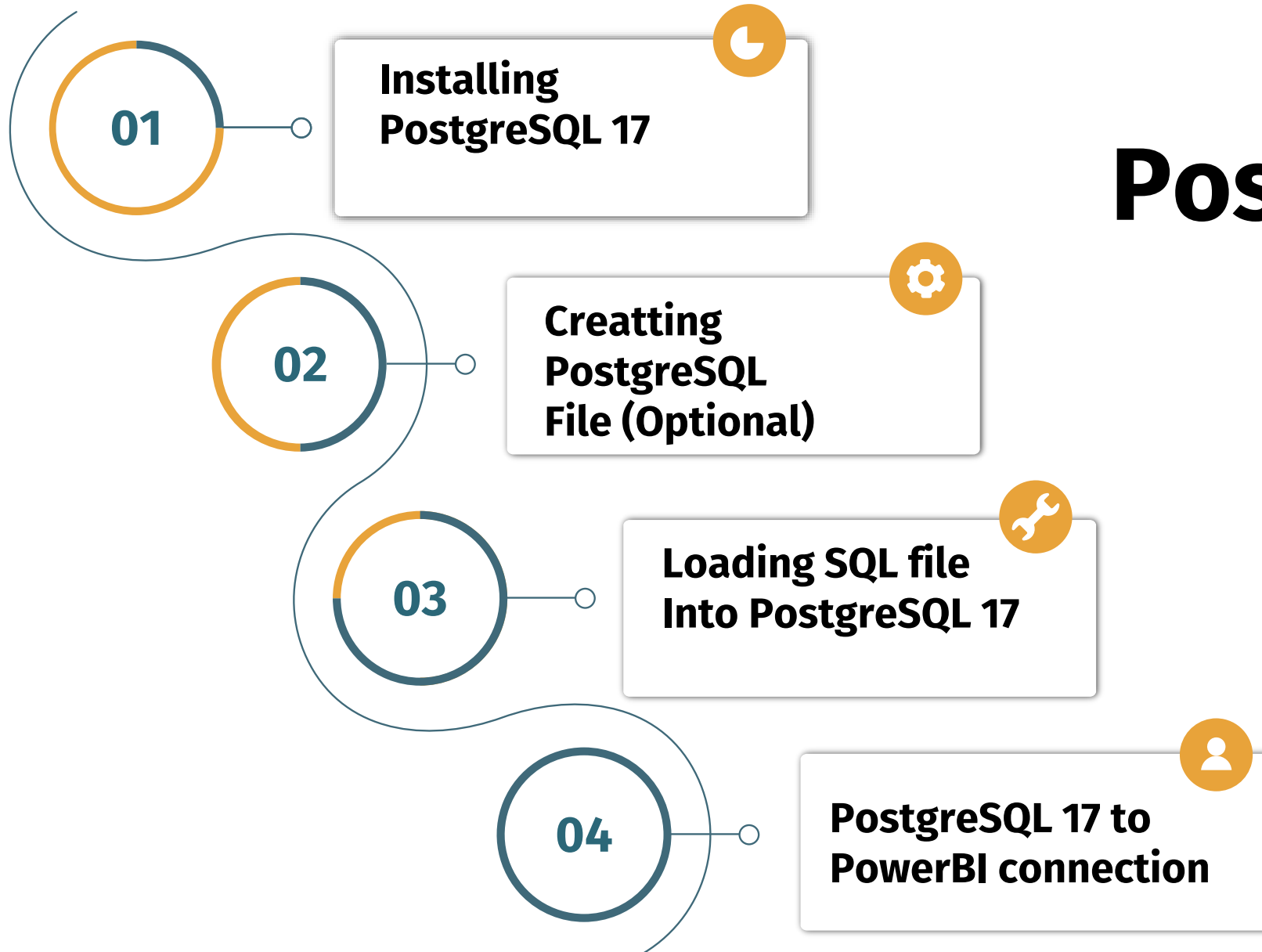
Loo Guan Yee  
October 2024



# Documentation Content

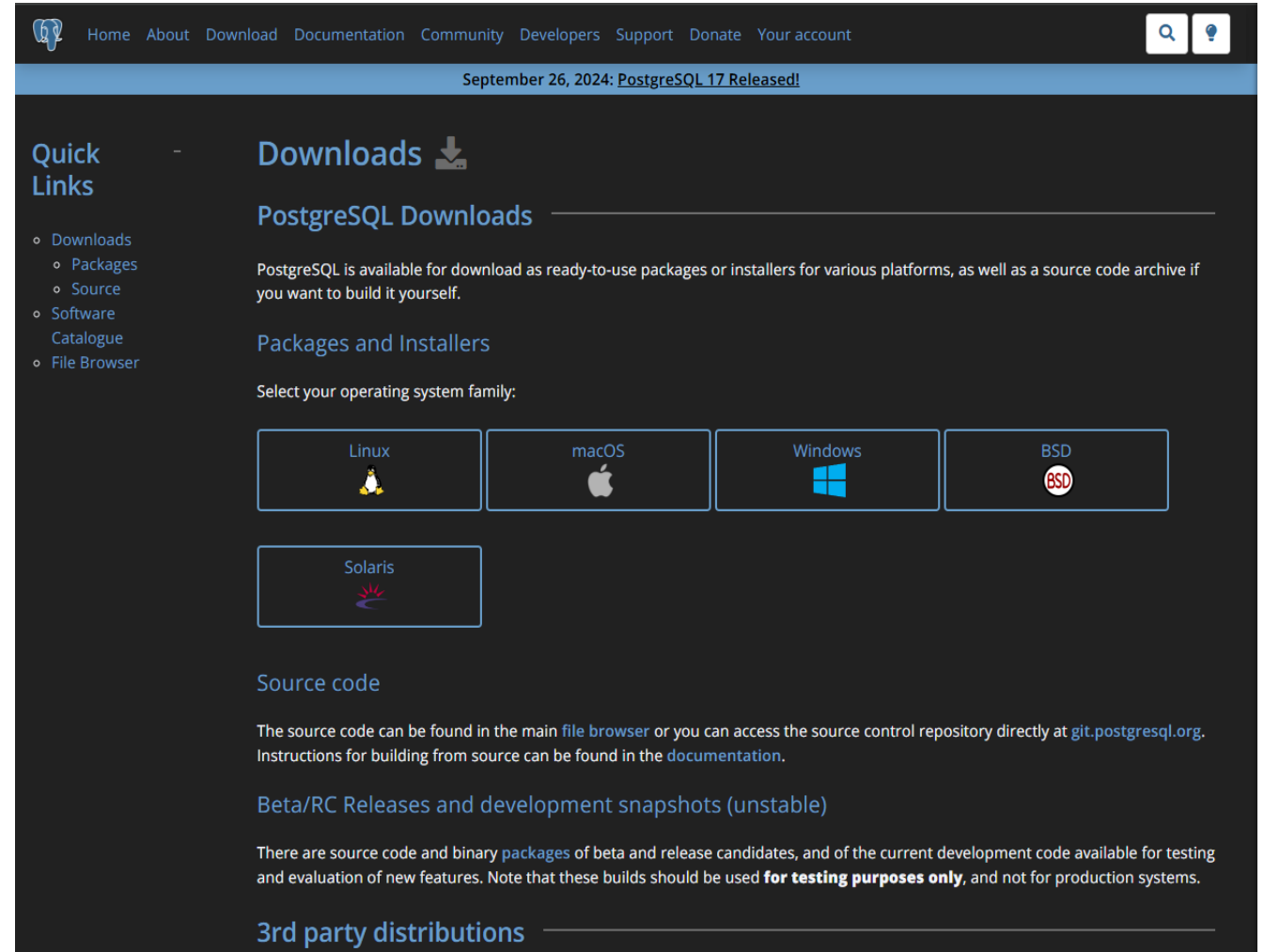
- **Installing PostgreSQL 17**
- **Creating the SQL file from scratch (Optional)**
  - Getting and preprocessing the DataCo Supply Chain Dataset
  - Creating the sql file in PostgreSQL 17
  - Backing up the sql file into PostgreSQL 17
- **Loading the sql file into PostgreSQL 17**
- **Connecting the PostgreSQL database to PowerBI**

# Installing PostgreSQL 17



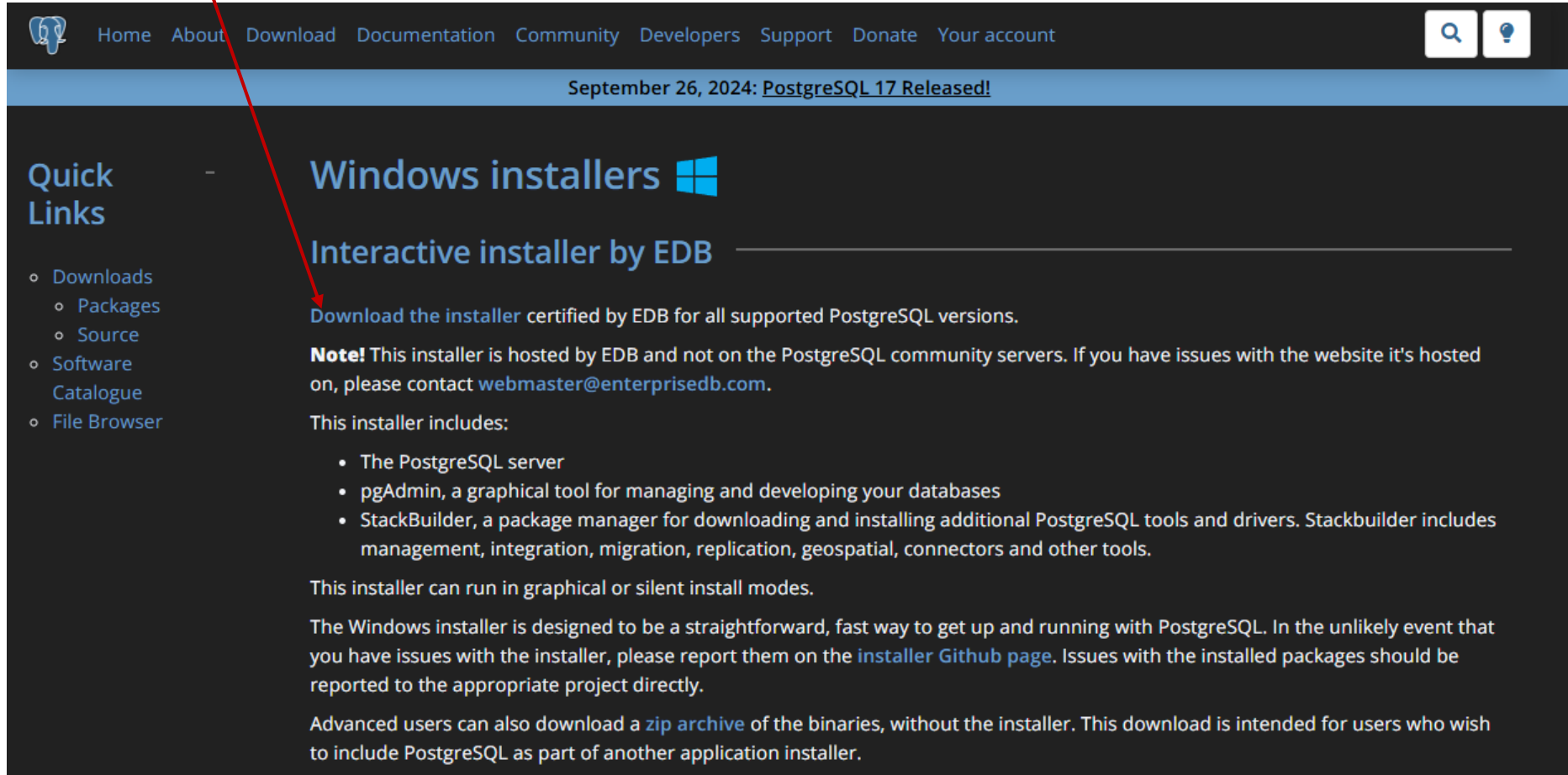
# PostgreSQL 17 Installation

- This documentation focuses on the installation of PostgreSQL 17 realised on 26 September 2024
- Go to <https://www.postgresql.org/download>
- Click on the your OS under Packages and Installers



# PostgreSQL Windows Installation

- For Windows, Click on the “Download the Installer” icon



The screenshot shows the PostgreSQL website's "Windows installers" page. The navigation bar at the top includes links for Home, About, Download, Documentation, Community, Developers, Support, Donate, and Your account. A blue banner below the navigation bar announces "September 26, 2024: PostgreSQL 17 Released!". On the left, a "Quick Links" sidebar lists Downloads, Packages, Source, Software Catalogue, and File Browser. The main content area is titled "Windows installers" with a Windows logo. Below this, it says "Interactive installer by EDB". A red arrow points from the bullet point above to the link "Download the installer", which is described as being certified by EDB for all supported PostgreSQL versions. A note states that the installer is hosted by EDB and not on the PostgreSQL community servers, and provides the contact email [webmaster@enterprisedb.com](mailto:webmaster@enterprisedb.com). The page lists the included components: the PostgreSQL server, pgAdmin, and StackBuilder. It also mentions that the installer can run in graphical or silent modes and provides instructions on where to report issues. Finally, it notes that advanced users can download a zip archive of the binaries without the installer.

Home About Download Documentation Community Developers Support Donate Your account

September 26, 2024: [PostgreSQL 17 Released!](#)

## Quick Links

- Downloads
- Packages
- Source
- Software Catalogue
- File Browser

## Windows installers

### Interactive installer by EDB

[Download the installer](#) certified by EDB for all supported PostgreSQL versions.

**Note!** This installer is hosted by EDB and not on the PostgreSQL community servers. If you have issues with the website it's hosted on, please contact [webmaster@enterprisedb.com](mailto:webmaster@enterprisedb.com).

This installer includes:

- The PostgreSQL server
- pgAdmin, a graphical tool for managing and developing your databases
- StackBuilder, a package manager for downloading and installing additional PostgreSQL tools and drivers. Stackbuilder includes management, integration, migration, replication, geospatial, connectors and other tools.

This installer can run in graphical or silent install modes.

The Windows installer is designed to be a straightforward, fast way to get up and running with PostgreSQL. In the unlikely event that you have issues with the installer, please report them on the [installer Github page](#). Issues with the installed packages should be reported to the appropriate project directly.

Advanced users can also download a [zip archive](#) of the binaries, without the installer. This download is intended for users who wish to include PostgreSQL as part of another application installer.

# PostgreSQL MacOS Installation

- For MacOS, Click on the “Download the Installer” icon



The screenshot shows the PostgreSQL website's 'macOS packages' page. The navigation bar at the top includes links for Home, About, Download, Documentation, Community, Developers, Support, Donate, and Your account. A blue banner below the navigation bar announces 'September 26, 2024: PostgreSQL 17 Released!'. The main content area is titled 'macOS packages' with an Apple logo. It states that users can get macOS PostgreSQL packages from several sources. A section titled 'Interactive installer by EDB' contains a link 'Download the installer' which is highlighted by a red arrow. Below this link, a note states that the installer is hosted by EDB and not on the PostgreSQL community servers. The page also lists the components included in the installer: the PostgreSQL server, pgAdmin, and StackBuilder. Finally, it mentions that the installer can run in graphical, command line, or silent install modes, and provides information on where to report issues and how advanced users can download binaries without the installer.

Quick Links

- Downloads
  - Packages
  - Source
- Software Catalogue
- File Browser

macOS packages 

You can get macOS PostgreSQL packages from several sources.

### Interactive installer by EDB

[Download the installer](#) certified by EDB for all supported PostgreSQL versions.

**Note!** This installer is hosted by EDB and not on the PostgreSQL community servers. If you have issues with the website it's hosted on, please contact [webmaster@enterprisedb.com](mailto:webmaster@enterprisedb.com).

This installer includes:

- The PostgreSQL server
- pgAdmin, a graphical tool for managing and developing your databases
- StackBuilder, a package manager for downloading and installing additional PostgreSQL tools and drivers. Stackbuilder includes management, integration, migration, replication, geospatial, connectors and other tools.

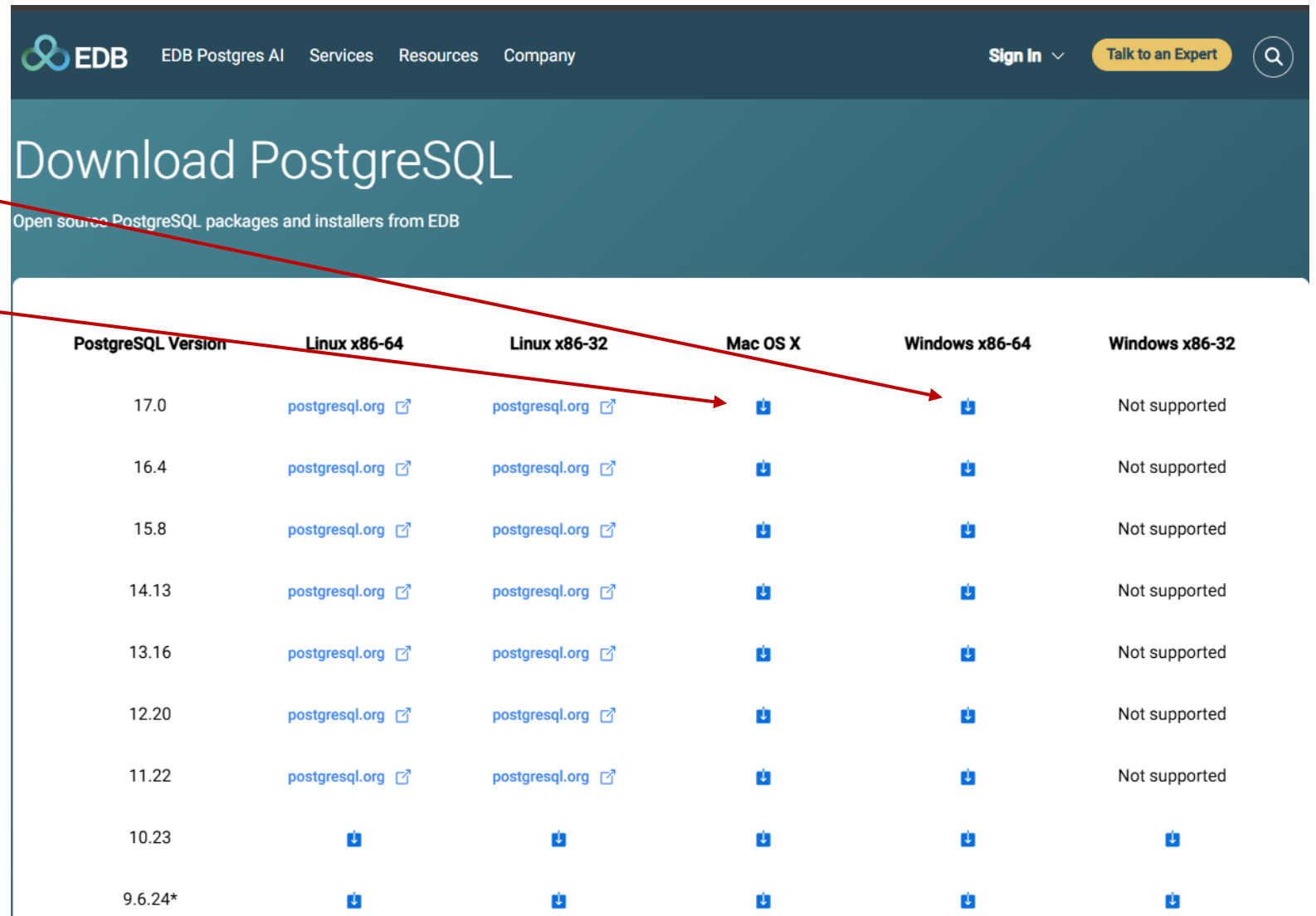
This installer can run in graphical, command line, or silent install modes.

The macOS installer is designed to be a straightforward, fast way to get up and running with PostgreSQL. In the unlikely event that you have issues with the installer, please report them on the [installer Github page](#). Issues with the installed packages should be reported to the appropriate project directly.

Advanced users can also download a [zip archive](#) of the binaries, without the installer. This download is intended for users who wish to include PostgreSQL as part of another application installer.

# PostgreSQL Windows and Mac OS Installation

- You will be redirected to a new page hosted by [EDB](#)
- For Windows, click on this.
- For Mac OS, click on this
- Download will start shortly

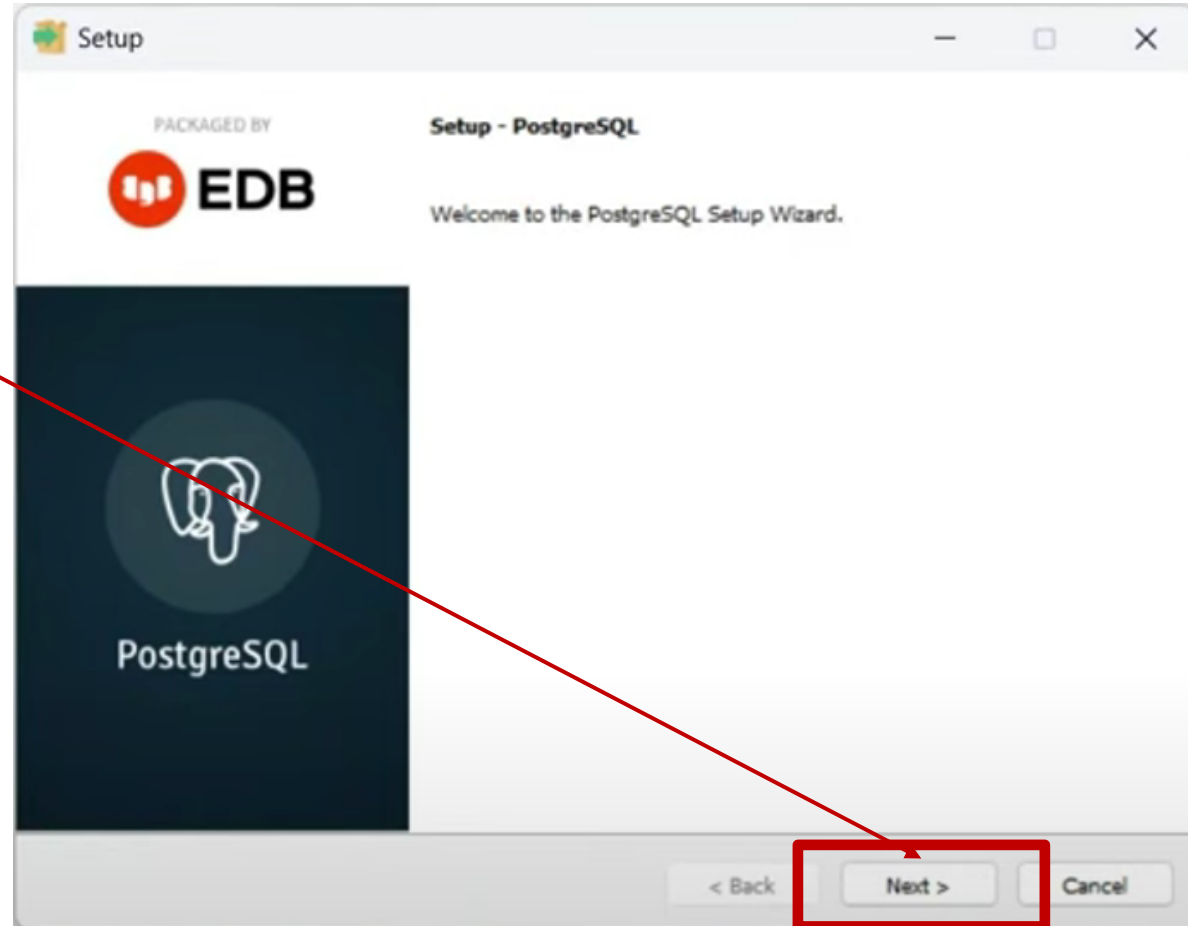


The screenshot shows the EDB PostgreSQL download page. The header includes the EDB logo and navigation links: EDB Postgres AI, Services, Resources, and Company. There are also links for 'Sign In' and 'Talk to an Expert'. The main heading is 'Download PostgreSQL' with the subtitle 'Open source PostgreSQL packages and installers from EDB'. Below this is a table of download links for various PostgreSQL versions across different operating systems.

PostgreSQL Version	Linux x86-64	Linux x86-32	Mac OS X	Windows x86-64	Windows x86-32
17.0	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
16.4	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
15.8	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
14.13	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
13.16	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
12.20	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
11.22	<a href="#">postgresql.org</a>	<a href="#">postgresql.org</a>			Not supported
10.23					
9.6.24*					

# PostgreSQL.exe installation

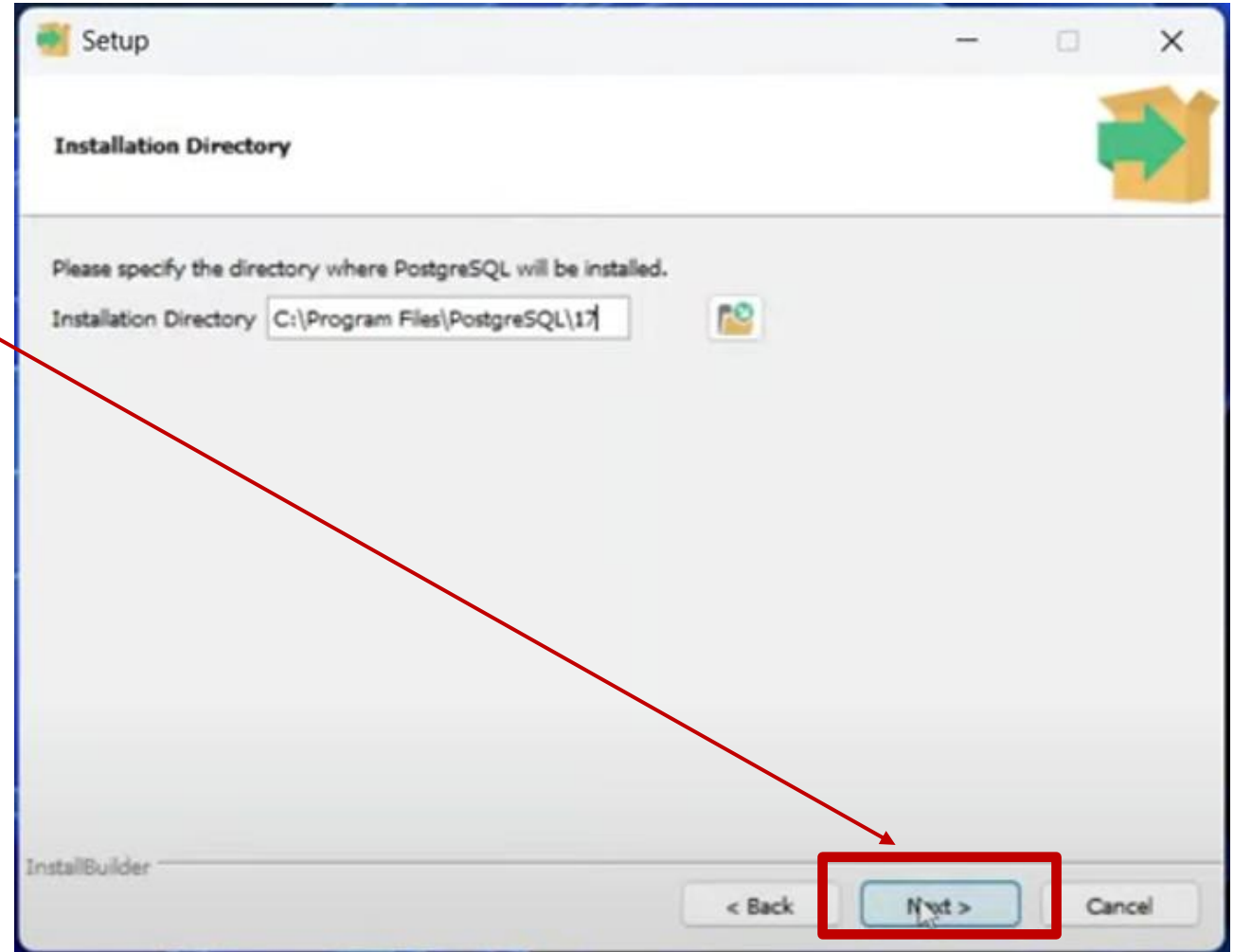
- Open the postgresql-17.0-windowsx64.exe found in your download folder
- Click on next





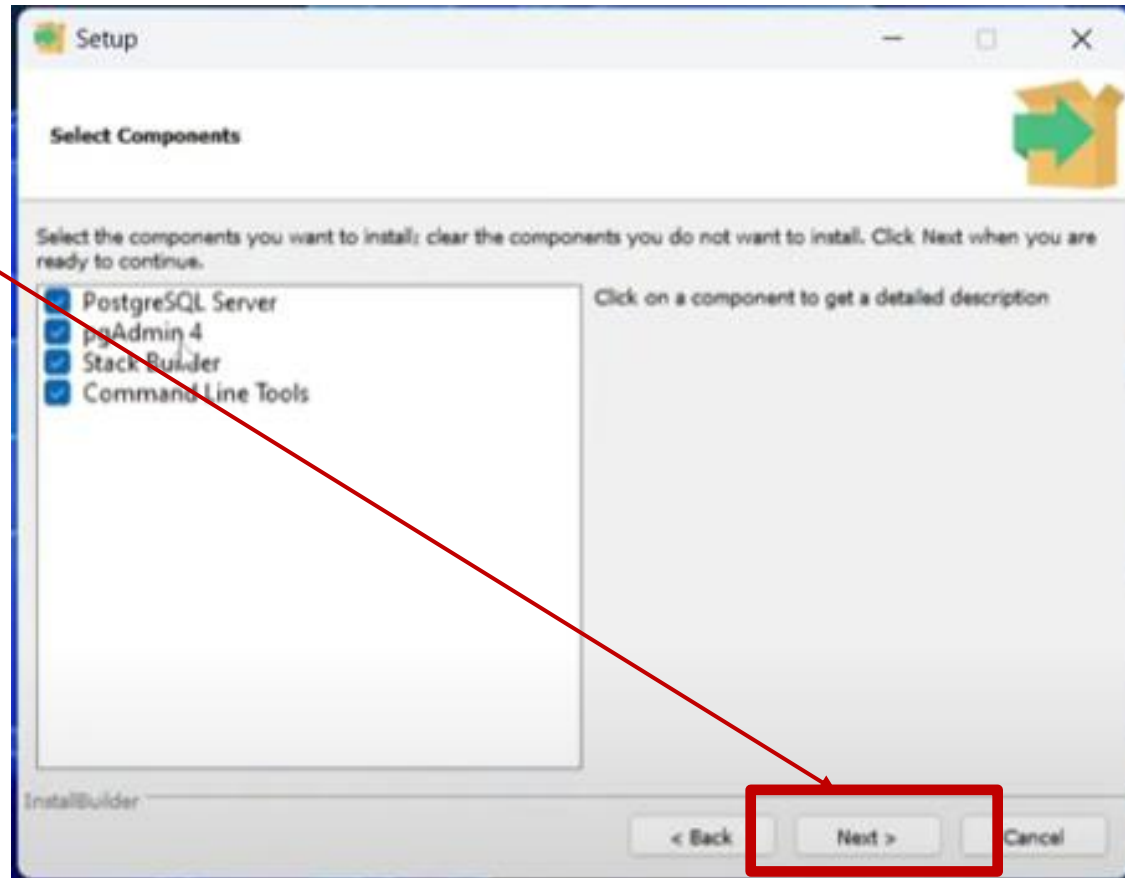
# PostgreSQL.exe installation

- Click on next



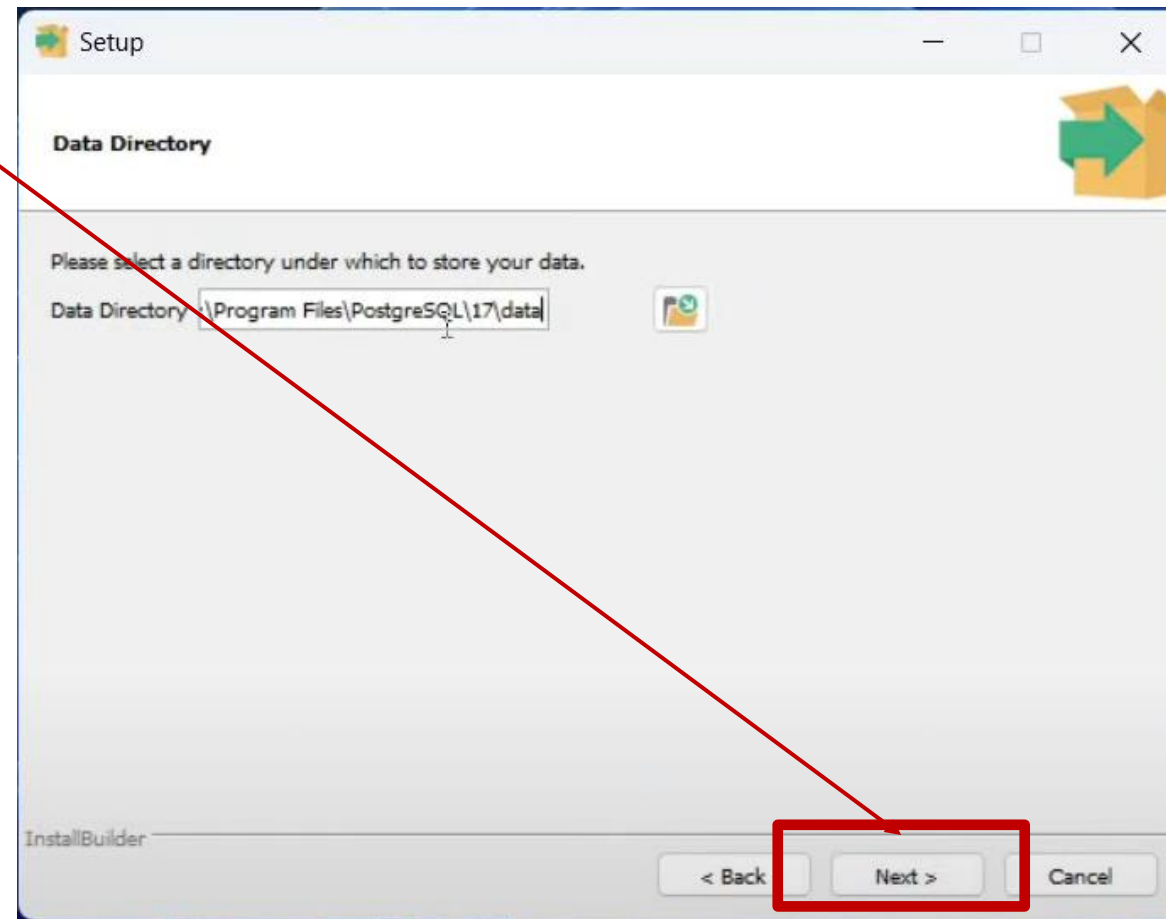
# PostgreSQL.exe installation

- Check all the ticks available in the screen
- Click on next



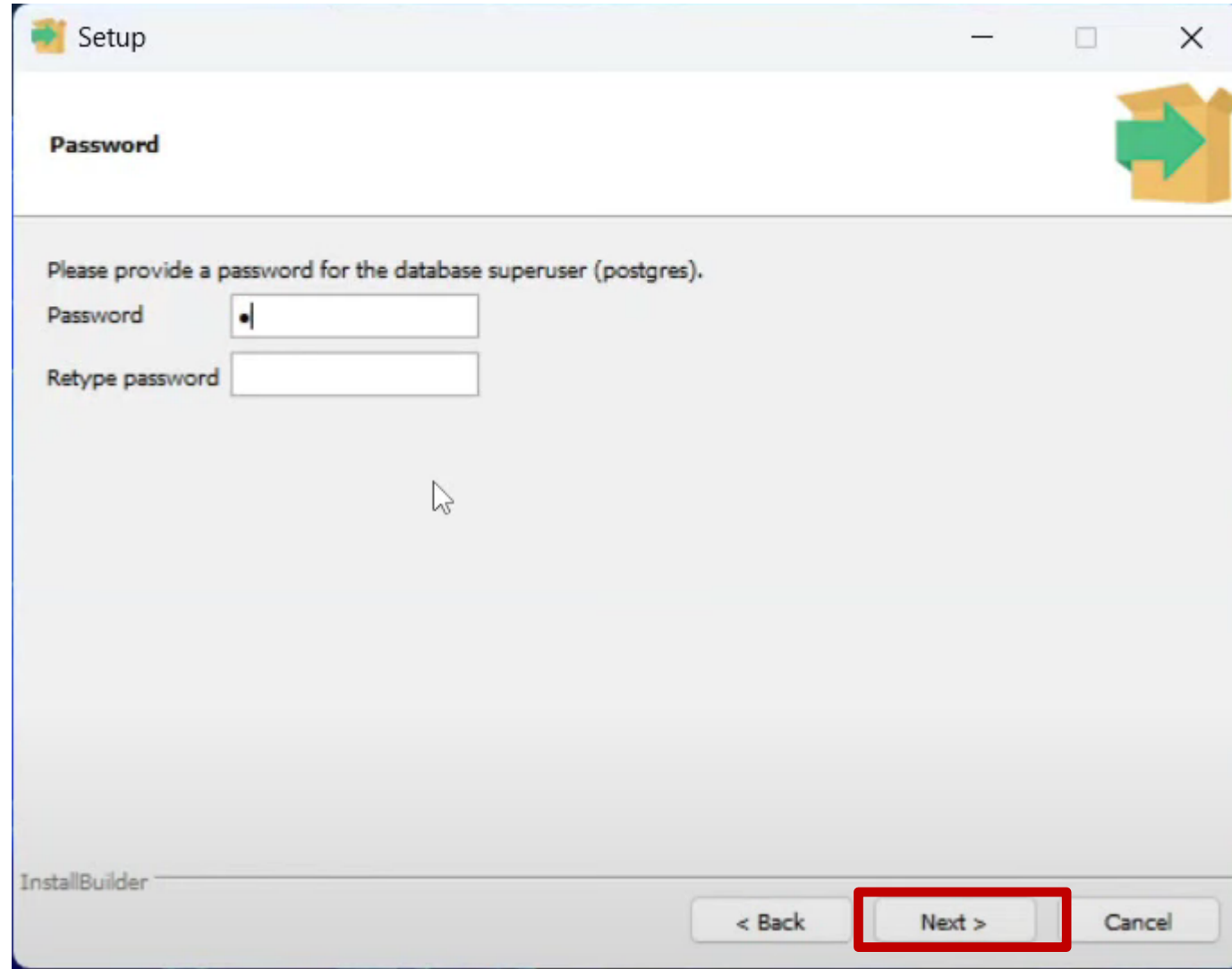
# PostgreSQL.exe installation

- Click on next



# PostgreSQL.exe installation (Important!)

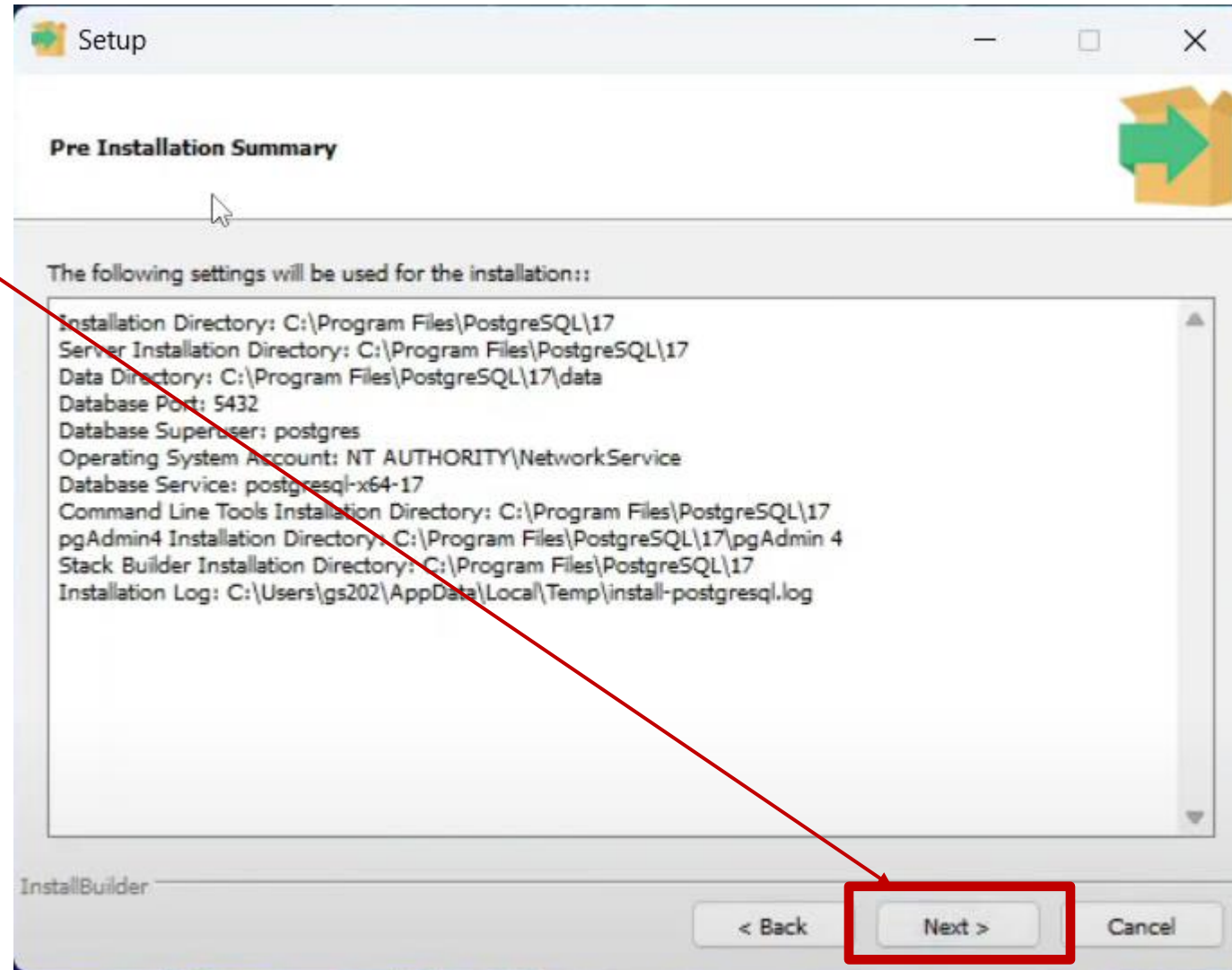
- Type in your password and Retype password
- Then, click on “Next”
- This step is important for the connecting PostgreSQL to PowerBI



The screenshot shows a Windows-style 'Setup' window for PostgreSQL. The title bar says 'Setup' with standard minimize, maximize, and close buttons. The main content area is titled 'Password' and contains the instruction: 'Please provide a password for the database superuser (postgres)'. Below this, there are two text input fields: 'Password' and 'Retype password'. A mouse cursor is visible over the 'Retype password' field. In the bottom right corner, there are three buttons: '< Back', 'Next >', and 'Cancel'. The 'Next >' button is highlighted with a red rectangular border. At the bottom left, the text 'InstallBuilder' is visible. A large green arrow icon is positioned in the top right corner of the main content area.

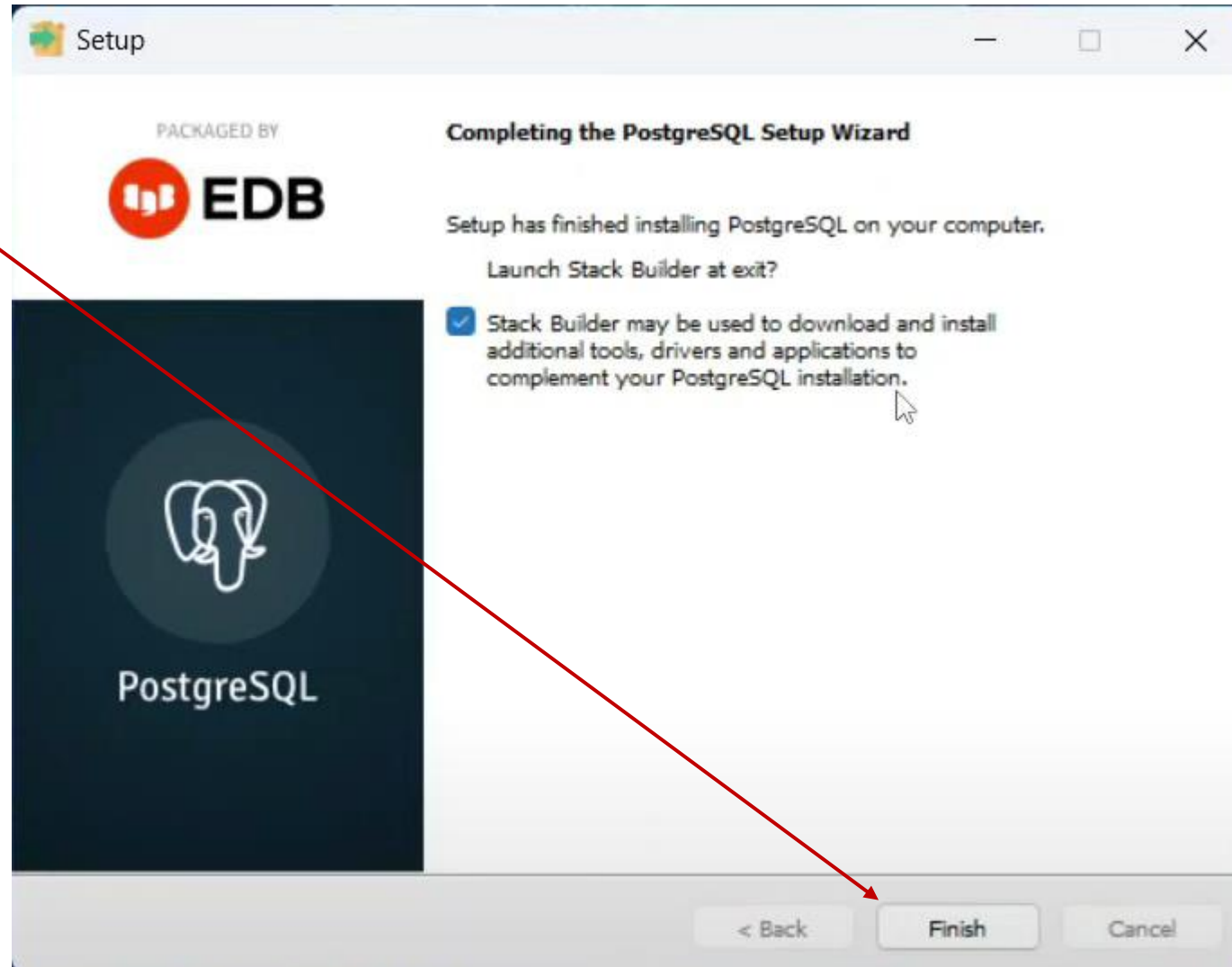
# PostgreSQL.exe installation

- Click on “Next”
- Wait for the installation to complete



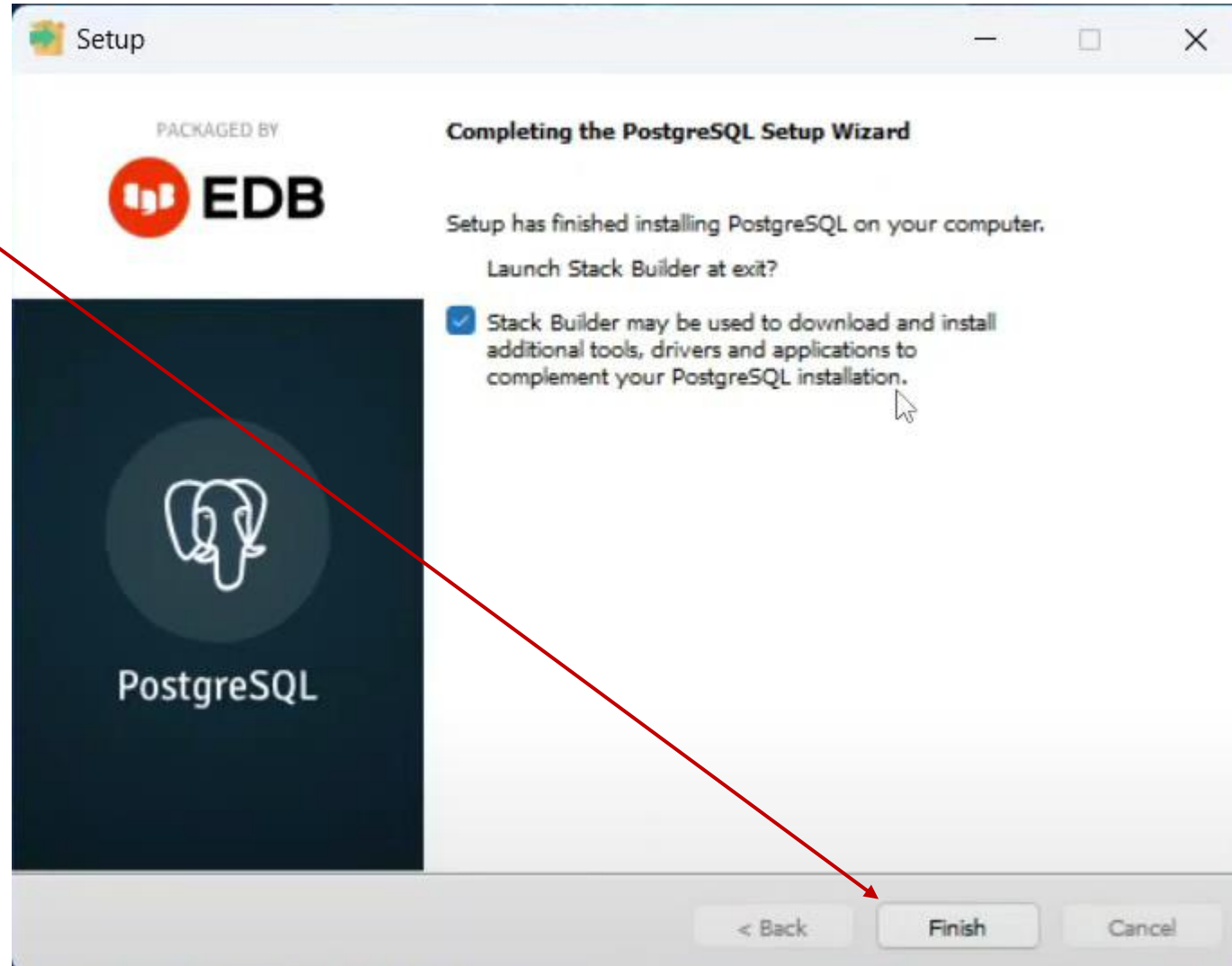
# PostgreSQL.exe installation

- Click on “Finish”



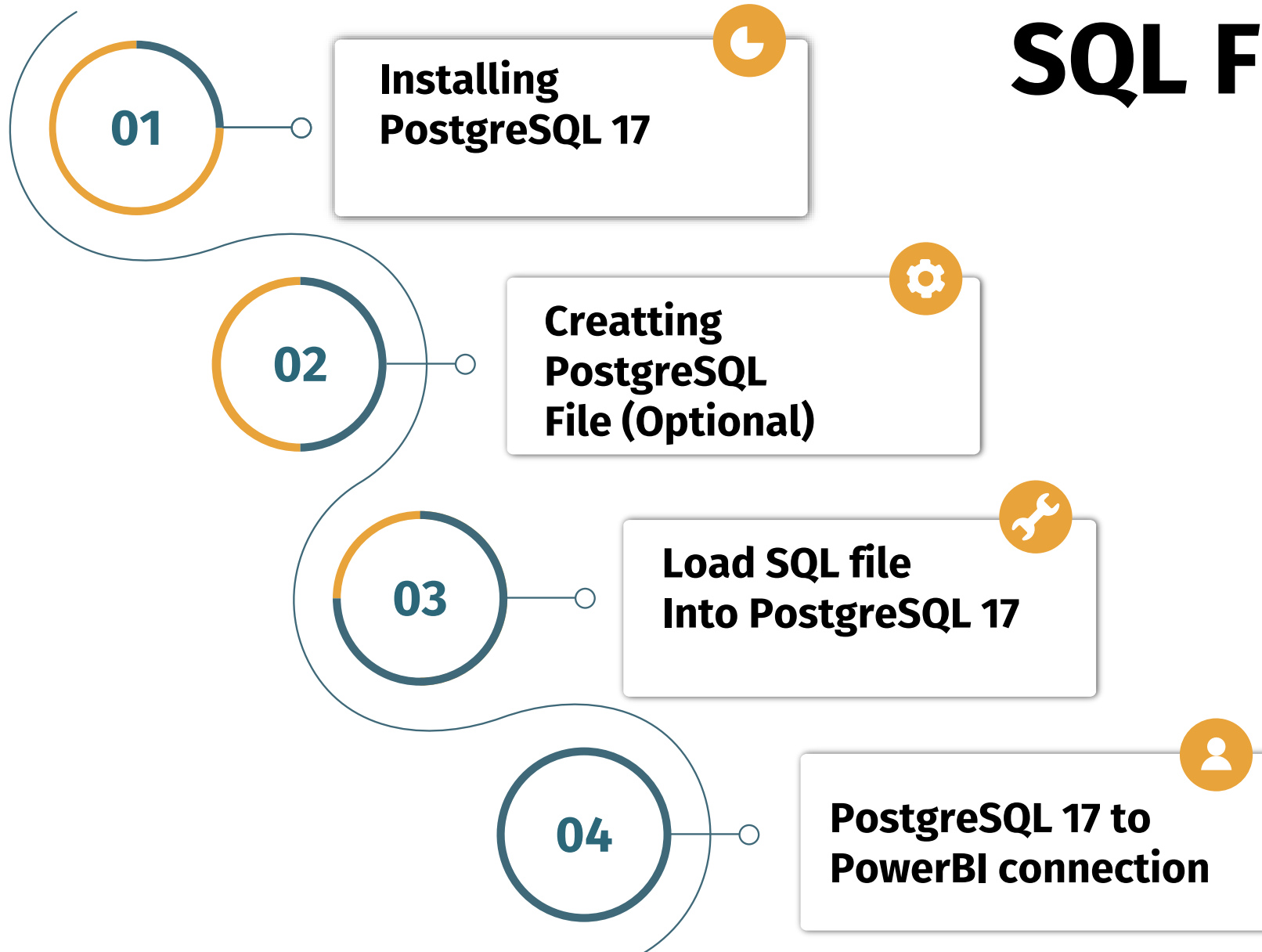
# Open Pg admin4

- Click on “Finish”



# SQL File creation


Getting the DataCo Supply  
Chain Dataset





# Obtaining the DataCo Supply Chain dataset

- Two methods
- First method:
  - Download from this link and click on
  - 91.5MB “DataCoSupplyChainDataset.csv: file
- Second method (Recommended):
  - Open Powershell or Git Bash
  - Use cd to navigate to the directory where you want to clone the repository
  - Clone the repository with this command
  - `git clone https://github.com/Guan-Yee/DataCo-Supply-Chain`

 Mendeley Data

?

Sign In / Register

## DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS

Published: 13 March 2019 | Version 5 | DOI: 10.17632/8gx2fvg2k6.5  
Contributors: [Fabian Constante](#), [Fernando Silva](#), [António Pereira](#)

### Description

A DataSet of Supply Chains used by the company DataCo Global was used for the analysis. Dataset of Supply Chain , which allows the use of Machine Learning Algorithms and R Software. Areas of important registered activities : Provisioning , Production , Sales , Commercial Distribution.It also allows the correlation of Structured Data with Unstructured Data for knowledge generation.







Type Data :  
Structured Data : DataCoSupplyChainDataset.csv  
Unstructured Data : tokenized\_access\_logs.csv (Clickstream)

Types of Products : Clothing , Sports , and Electronic Supplies

Additionally it is attached in another file called DescriptionDataCoSupplyChain.csv, the description of each of the variables of the DataCoSupplyChainDataset.csv.

Download All 25 MB ⓘ

### Files

	DataCoSupplyChainDataset.csv	91.5 MB 
	DescriptionDataCoSupplyChain.csv	3.36 KB 
	tokenized_access_logs.csv	91 MB 


### Institutions

1  
Citation

### Dataset metrics

Usage

Views:	30458
Downloads:	5775

 [View details >](#)

### Latest version

Version 5  
Published: 13 Mar 2019  
DOI: 10.17632/8gx2fvg2k6.5

#### Cite this dataset

Constante, Fabian; Silva, Fernando; Pereira, António (2019), "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS", Mendeley Data, V5, doi: 10.17632/8gx2fvg2k6.5

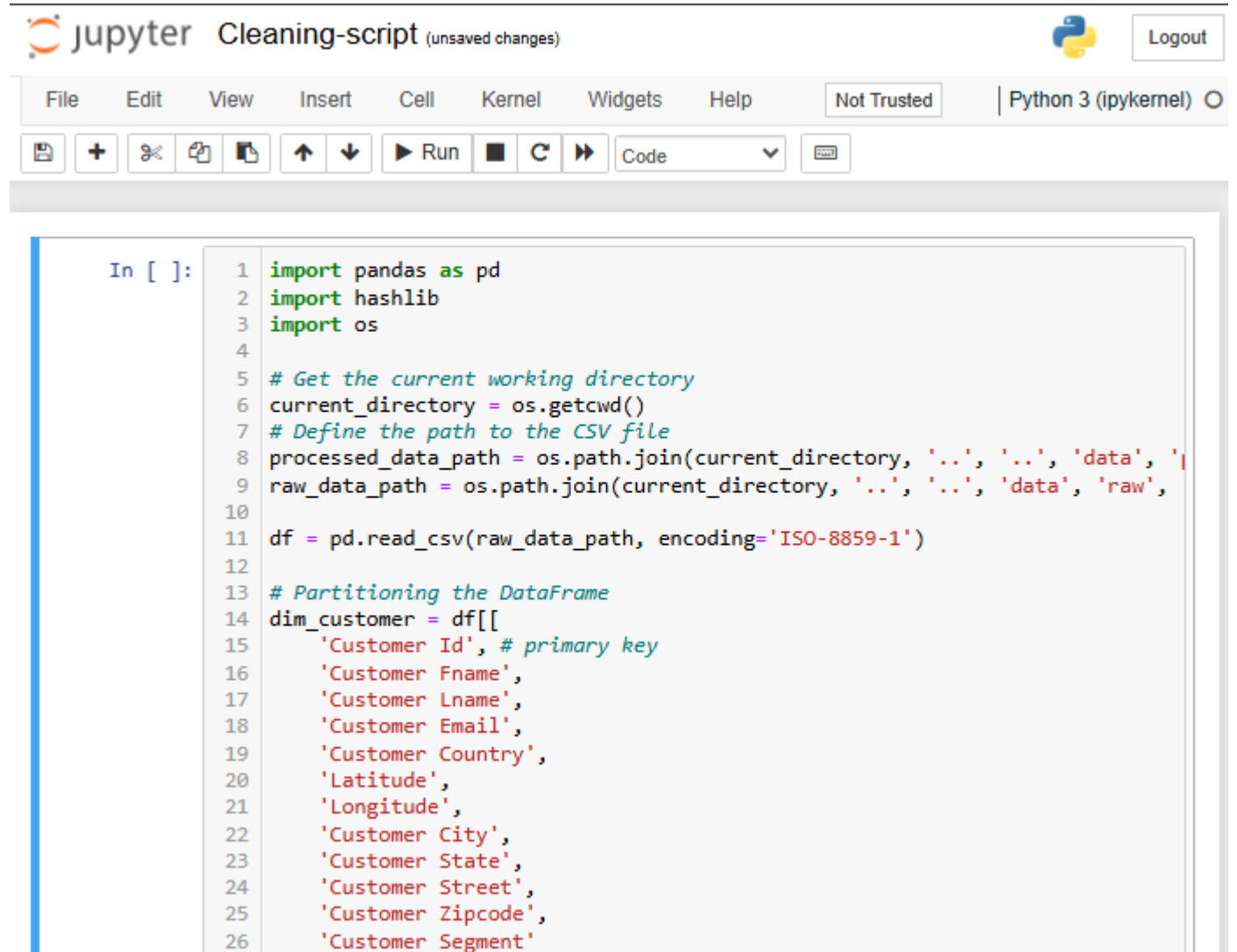
Copy to clipboard

### Previous versions

Version 4	12 March 2019
Version 3	12 March 2019
Version 2	12 March 2019

# Cleaning the DataCo Supply Chain dataset

- Open your Powershell or Git Bash
- Navigate to the folder where you git clone the DataCo-supply-chain repository
- Cd to *scripts/python*
- Open the ipynb folder and click on run for this cell
  - It should take about 30 seconds to run this cell



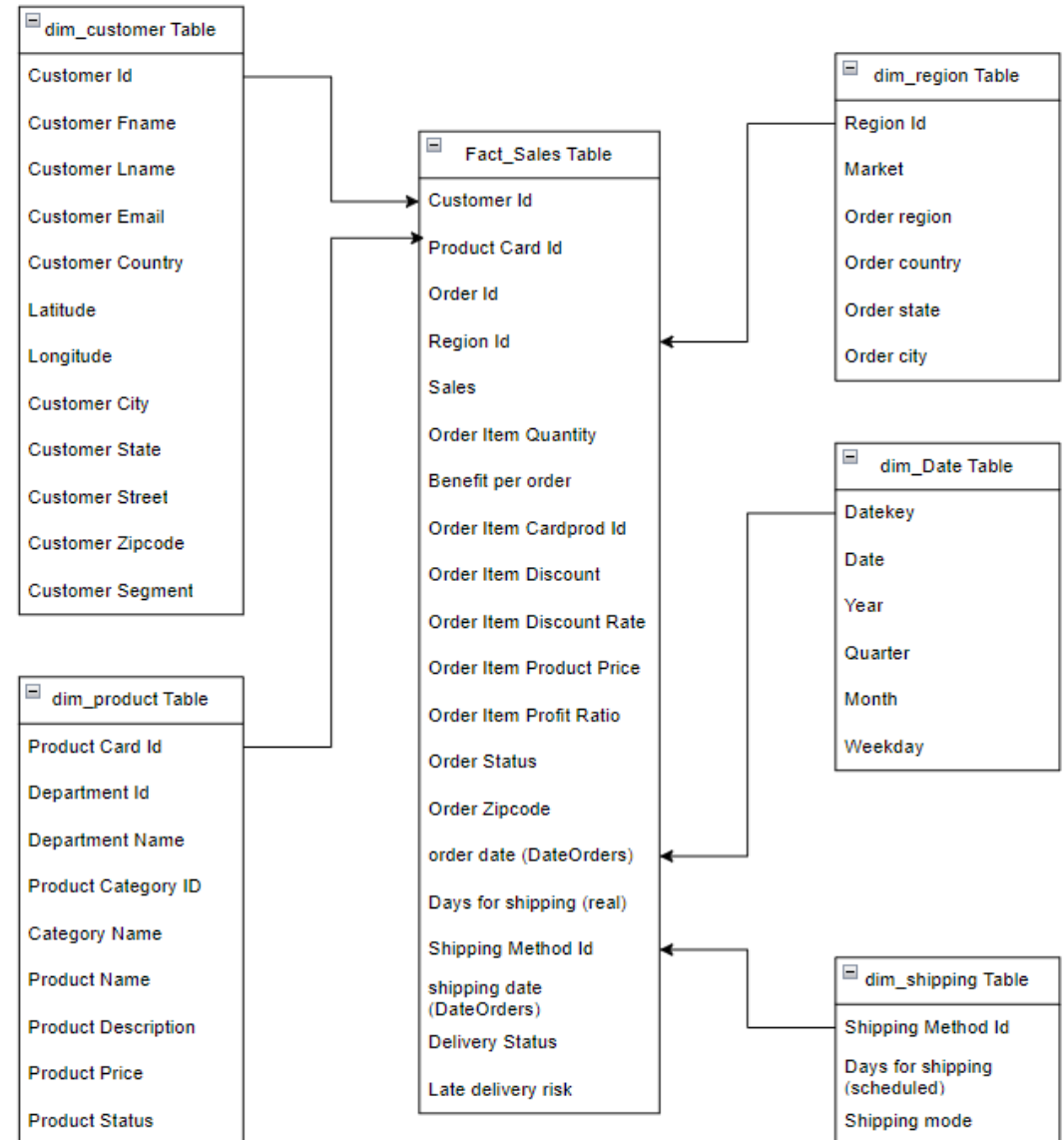
The screenshot shows a Jupyter Notebook titled "Cleaning-script (unsaved changes)". The interface includes a top bar with the Jupyter logo, the title, and a "Logout" button. Below the top bar is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, Help. To the right of the menu bar is a "Not Trusted" warning and the text "Python 3 (ipykernel)". Below the menu bar is a toolbar with icons for saving, adding, deleting, and running code. The main area of the notebook displays a Python script with line numbers 1 through 26. The script imports pandas, hashlib, and os. It then defines the current directory and the path to the CSV file. It reads the CSV file into a DataFrame and partitions it based on customer information.

```
In [ ]: 1 import pandas as pd
        2 import hashlib
        3 import os
        4
        5 # Get the current working directory
        6 current_directory = os.getcwd()
        7 # Define the path to the CSV file
        8 processed_data_path = os.path.join(current_directory, '..', '..', 'data', 'processed_data.csv')
        9 raw_data_path = os.path.join(current_directory, '..', '..', 'data', 'raw_data.csv')
        10
        11 df = pd.read_csv(raw_data_path, encoding='ISO-8859-1')
        12
        13 # Partitioning the DataFrame
        14 dim_customer = df[[
        15     'Customer Id', # primary key
        16     'Customer Fname',
        17     'Customer Lname',
        18     'Customer Email',
        19     'Customer Country',
        20     'Latitude',
        21     'Longitude',
        22     'Customer City',
        23     'Customer State',
        24     'Customer Street',
        25     'Customer Zipcode',
        26     'Customer Segment']]
```

# Cleaning the DataCo Supply Chain dataset

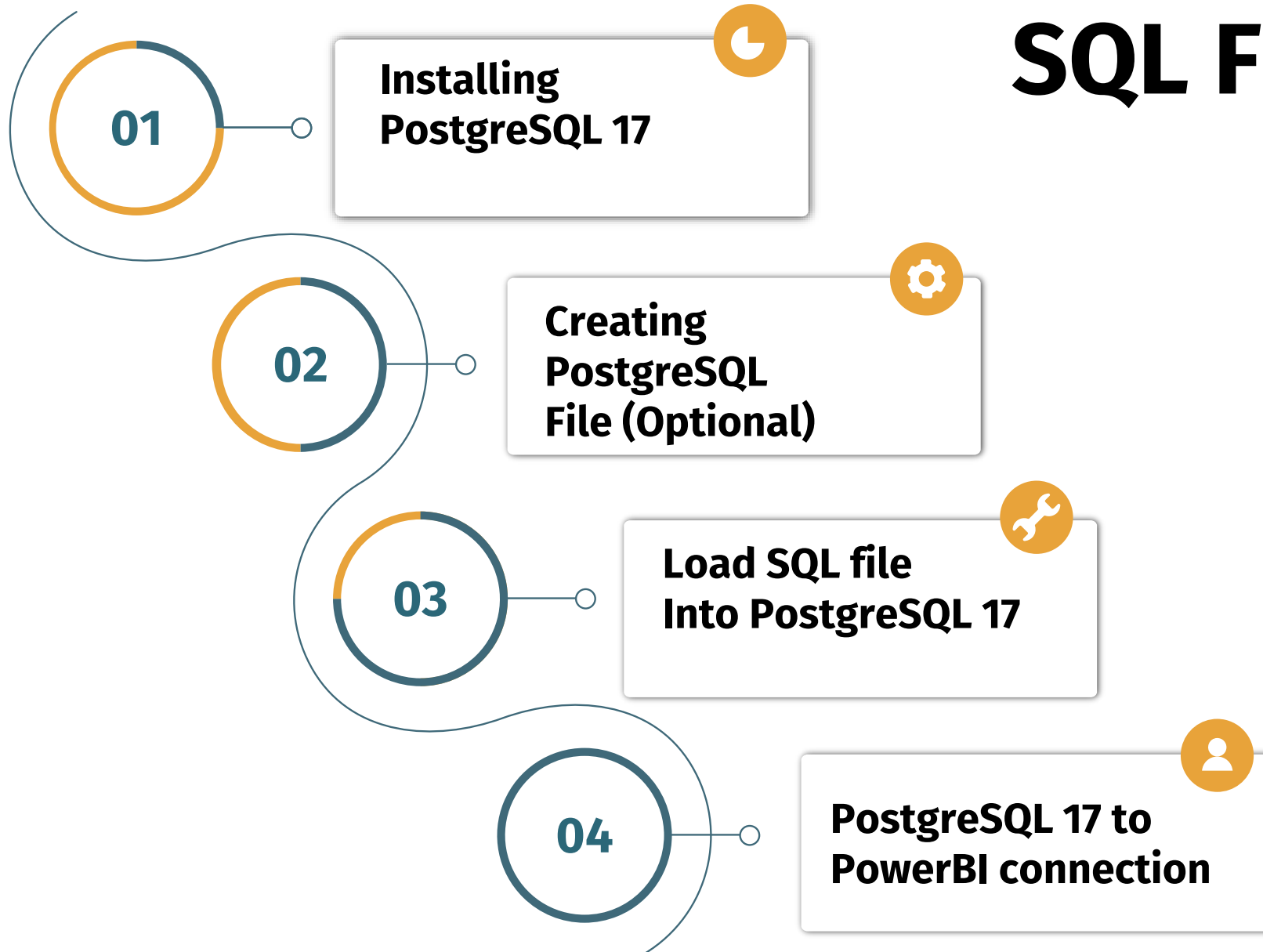
- You should be able to see the dimension and fact files based on this star schema in the data/processed directory
- \*Shipping Method Id, Datekey, and Region Id are new columns created from this python script

Name	Status	Date modified	Type	Size
dim_customer.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	2,296 KB
dim_date.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	44 KB
dim_product.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	10 KB
dim_region.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	238 KB
dim_shipping.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	1 KB
Fact_Sales.csv	✓	21/10/2024 12:42 am	Microsoft Excel C...	30,646 KB



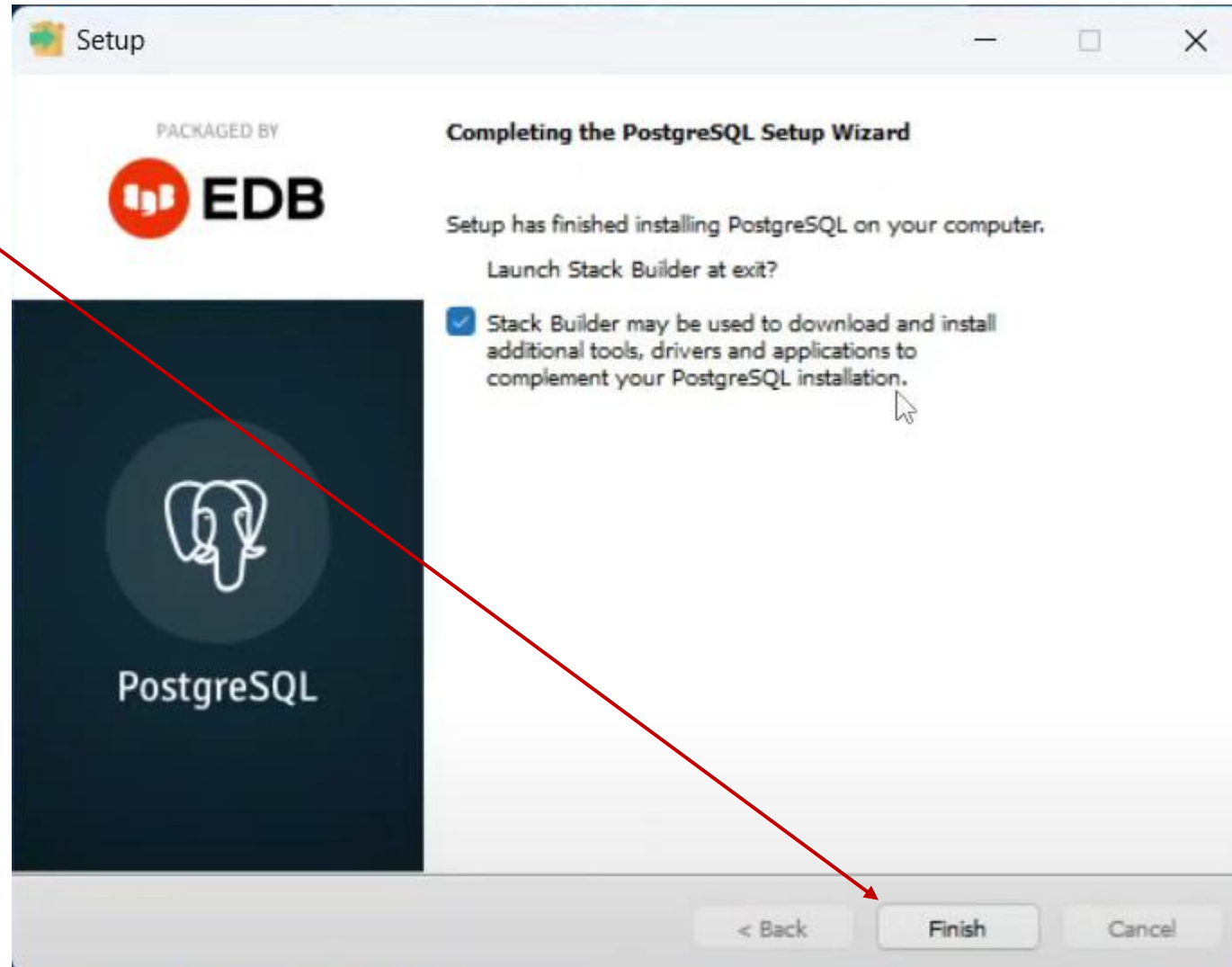
# SQL File creation

Creating the SQL Database



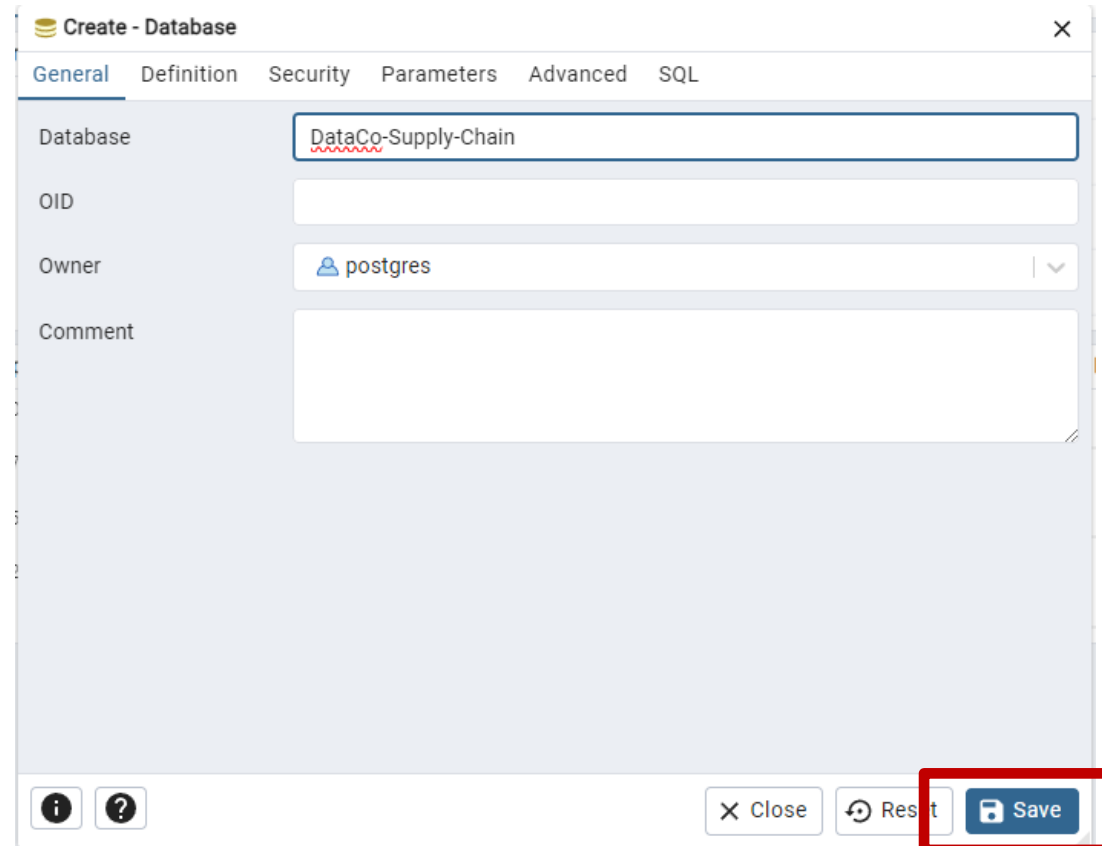
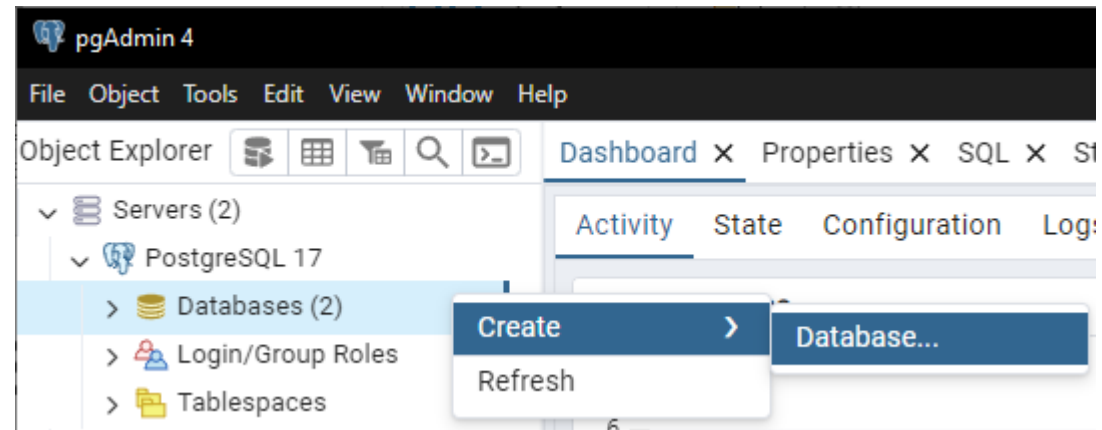
# PostgreSQL.exe installation

- Click on “Finish”



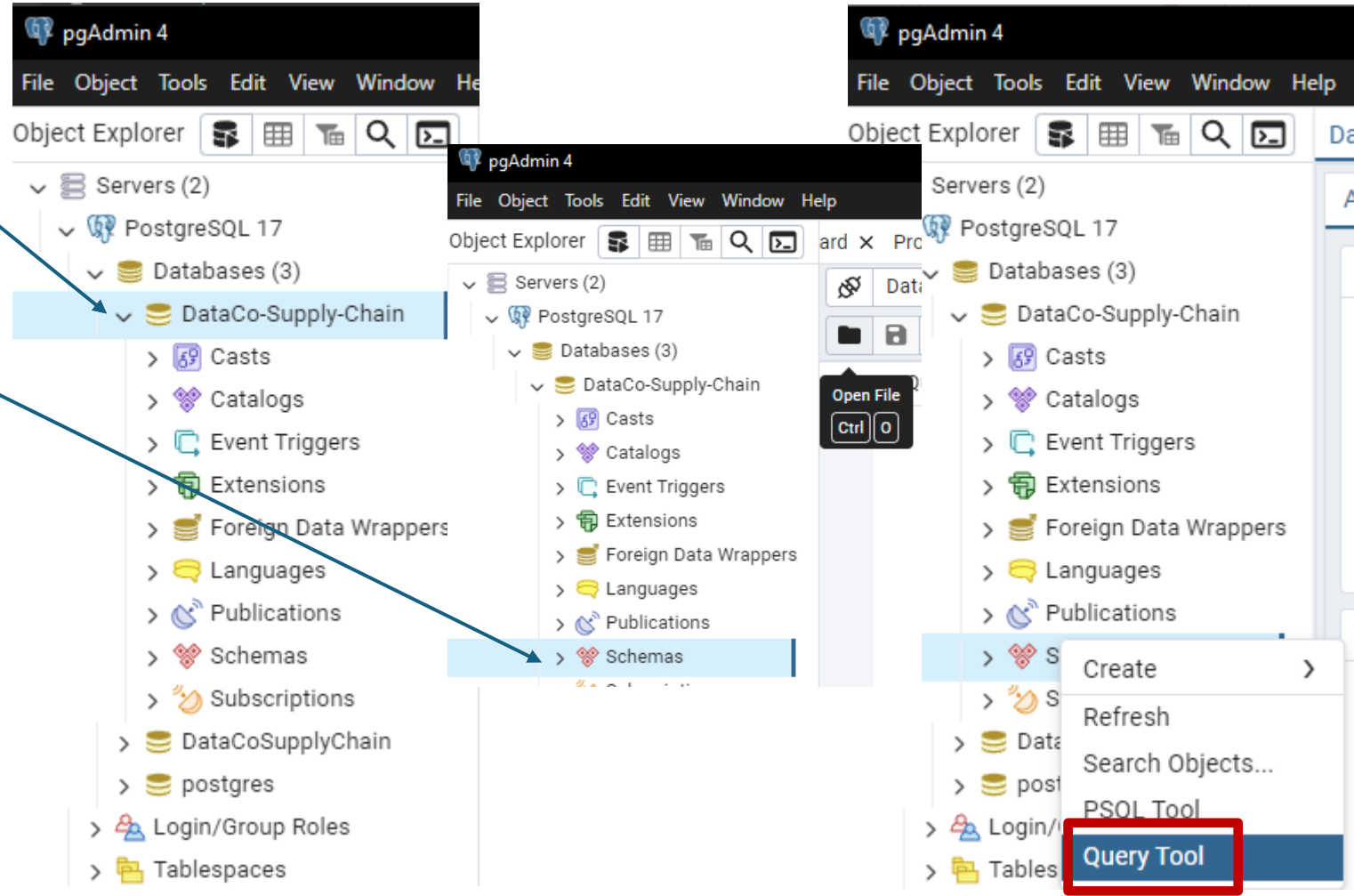
# Creating the Database

- Right click on “Databases”
- Click on “Create”
- Then click on “Database”
- A Create-Database window will appear
- Type in “DataCo-Supply-Chain”
- Click on “Save”



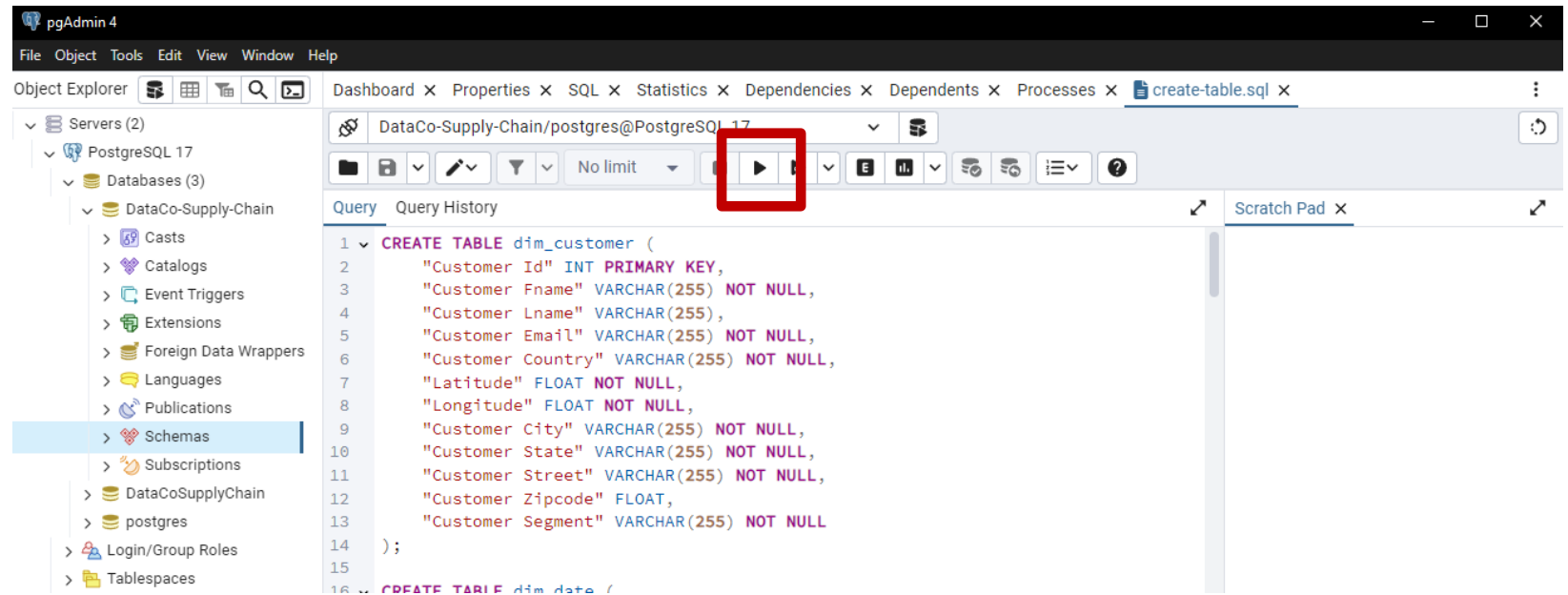
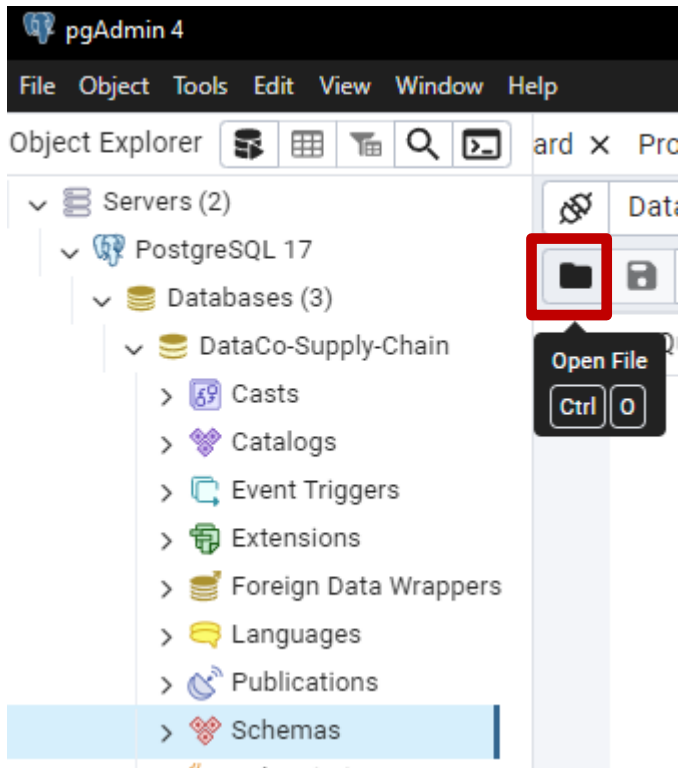
# Creating the Dimension and Fact Tables

- Click on DataCo-Supply-Chain arrow to expand the list
- Right click on Schemas and then click on Query Tool



# Creating the Dimension and Fact tables

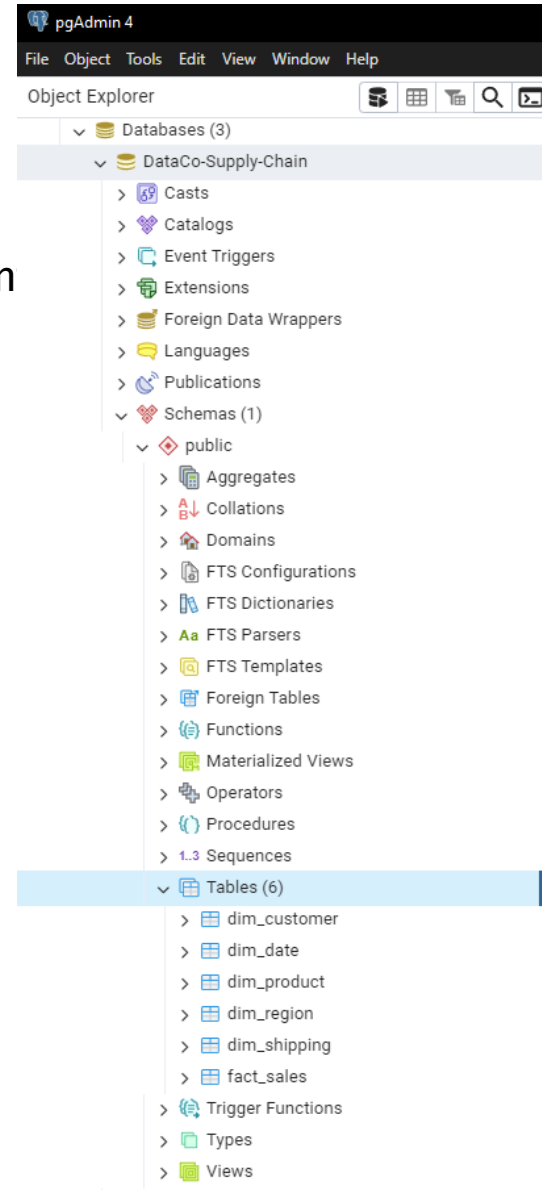
- Click on Open File
- Navigate to DataCo-Supply-Chain/Scripts/Sql/Create\_Table.sql file
- Click on the play button to execute the script





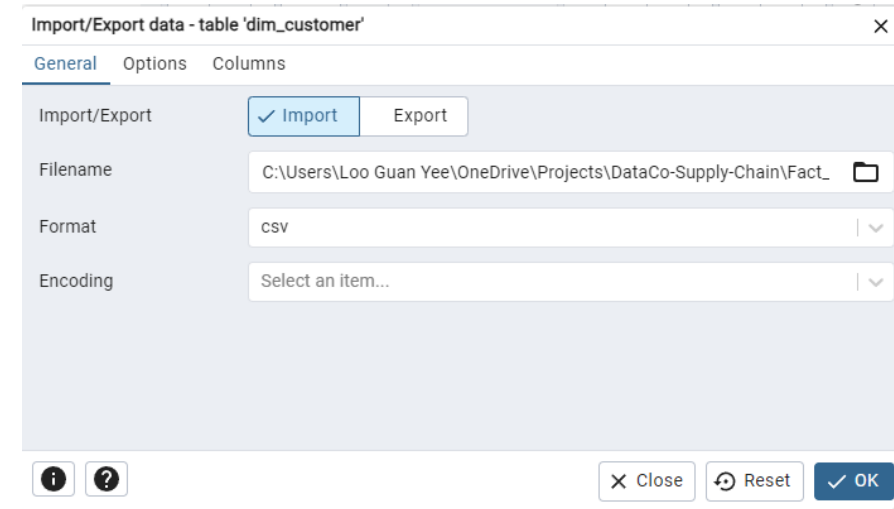
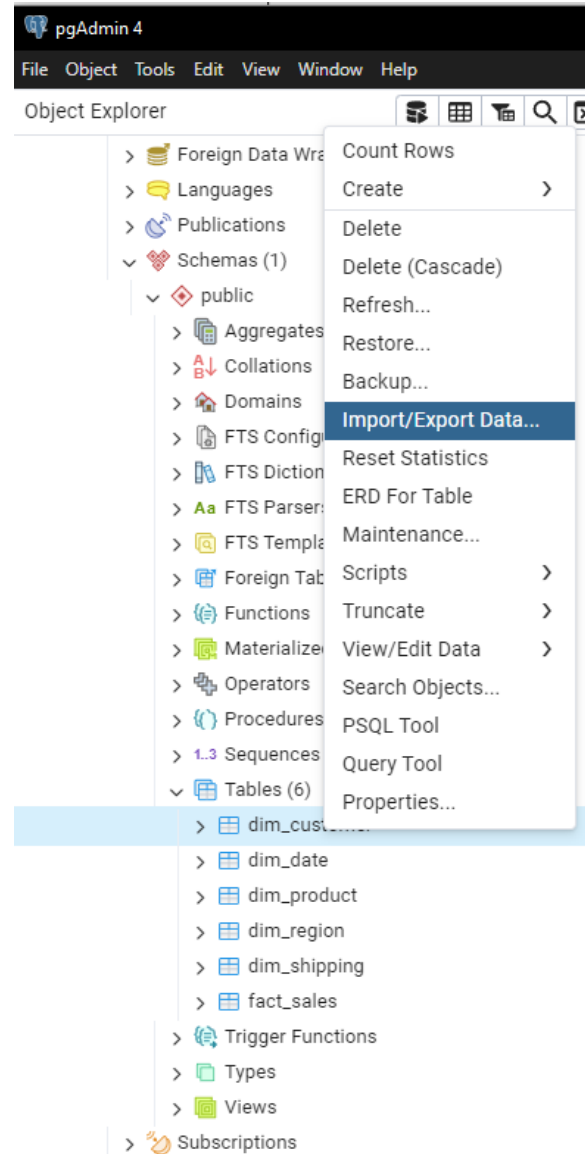
# Creating the Dimension and Fact tables

- Click on Schemas arrow
- Expand the Tables (6) button
  - You will see 6 dimension and fact tables present



# Loading the Dimension and Fact data into the tables

- Right click on each dim and Fact table
- Right click on Import/Export Data
- Type in the following file name for Dimension/Fact tables
  - Dim\_customer
    - Yourpath\DataCo-Supply-Chain\data\processed\dim\_customer.csv
  - Dim\_date
    - Yourpath\DataCo-Supply-Chain\data\processed\dim\_date.csv
  - Dim\_product
    - Yourpath\DataCo-Supply-Chain\data\processed\dim\_product.csv
  - Dim\_region
    - Yourpath\DataCo-Supply-Chain\data\processed\dim\_region.csv
  - Dim\_shipping
    - Yourpath\DataCo-Supply-Chain\data\processed\dim\_shipping.csv
  - Fact\_sales
    - Yourpath\DataCo-Supply-Chain\data\processed\FactSales.csv



# Loading the Dimension and Fact data into the tables

- You should be able to see double green status with the process completed for the 6 tables
- If red status, check if the loaded data columns match with the dimension table columns

**Process failed** ✖  
Copying table data 'public.dim\_date' on database 'DataCo-Supply-Chain' and server 'PostgreSQL 17 (localhost:5432)'  
[View Processes](#)

**Process started** ✖  
Copying table data 'public.dim\_date' on database 'DataCo-Supply-Chain' and server 'PostgreSQL 17 (localhost:5432)'  
[View Processes](#)

**Process Watcher - Import - Copying table data** ✖

Copying table data 'public.dim\_date' on database 'DataCo-Supply-Chain' and server 'PostgreSQL 17 (localhost:5432)'  
Running command:  

```
--command " \\copy public.dim_date (date, datekey, year, quarter, month, weekday, weekday_name) FROM 'C:/Users/LOOGUA~1/OneDrive/Projects/DATACO~2/data/PROCES~1/DIM_PR~2.CSV' DELIMITER ',' CSV HEADER QUOTE \" \" ESCAPE \"\",\""
```

🕒 Start time: Sat Oct 26 2024 15:04:33 GMT+0800 (Singapore Standard Time) ⛔ Stop Process

ERROR: extra data after last expected column  
CONTEXT: COPY dim\_date, line 2: "19,2,Fitness,2,Soccer,Nike Men's Fingertrap Max Training Shoe,,124.9899979,0"

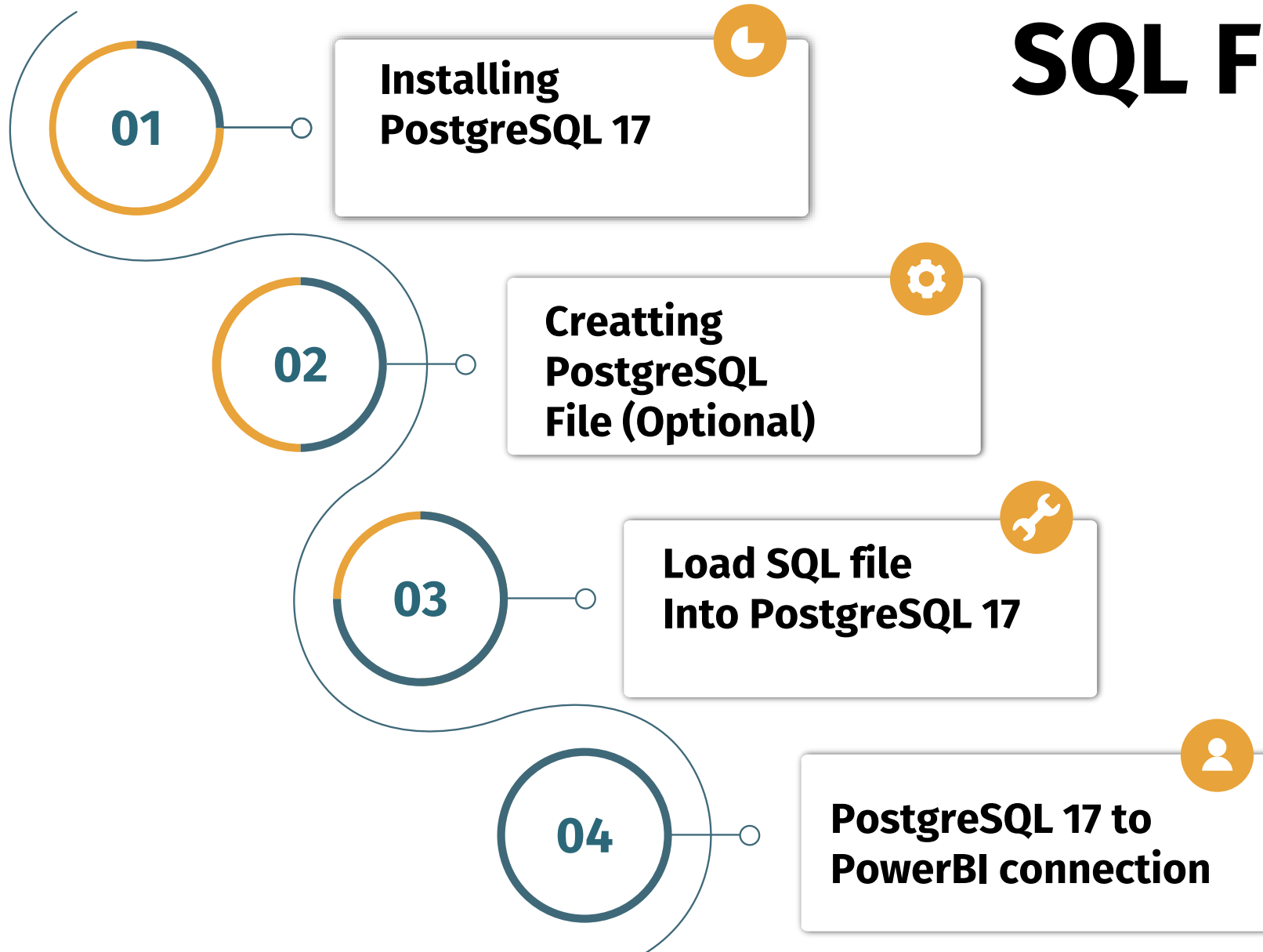
❗ Failed (exit code: 1). Execution time: 0.21 seconds

**Process completed** ✖  
Copying table data 'public.dim\_date' on database 'DataCo-Supply-Chain' and server 'PostgreSQL 17 (localhost:5432)'  
[View Processes](#)

**Process started** ✖  
Copying table data 'public.dim\_date' on database 'DataCo-Supply-Chain' and server 'PostgreSQL 17 (localhost:5432)'  
[View Processes](#)

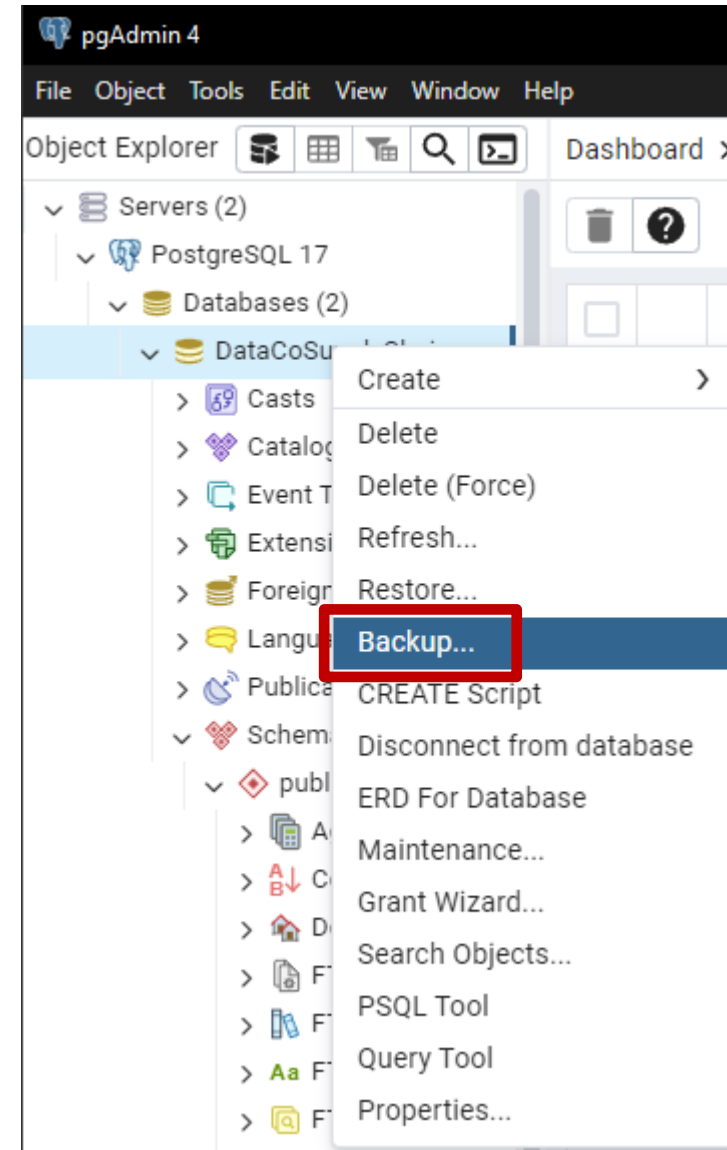
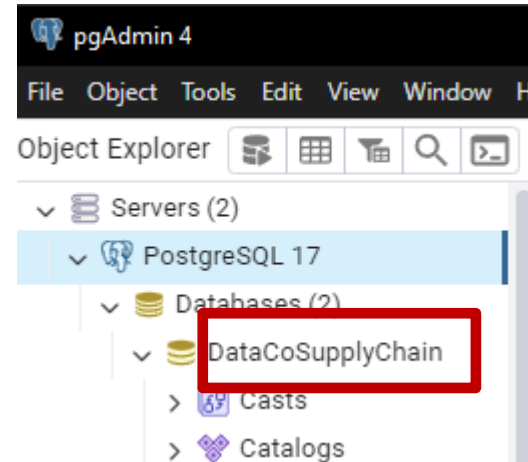
# SQL File creation

Backing up the SQL database



# Backing up the SQL folder

- Click on “DataCoSupplyChain” database
- Click on the backup button



# Open Pg admin4

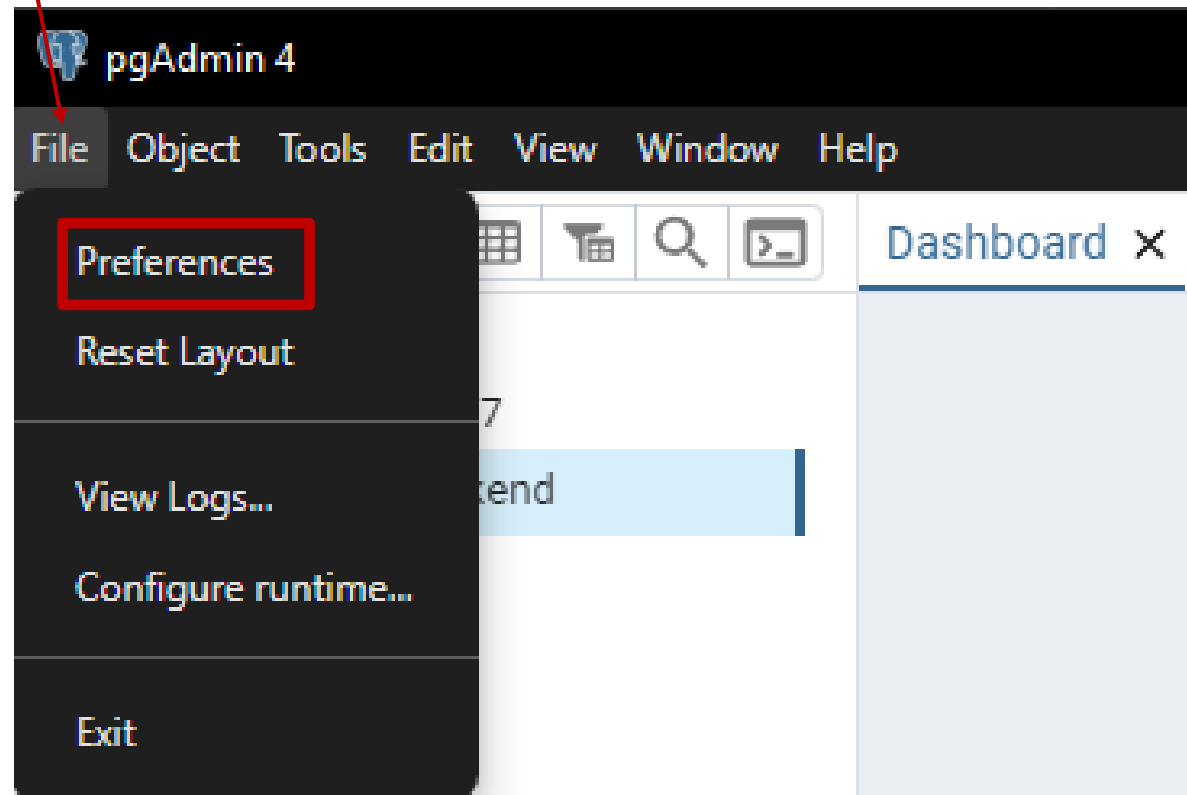
- Rationale to prevent server mismatch error
- Click on “File”
- Click on “Preferences”

pg\_dump: error: aborting because of server version mismatch  
pg\_dump: detail: server version: 17.0; pg\_dump version: 16.4



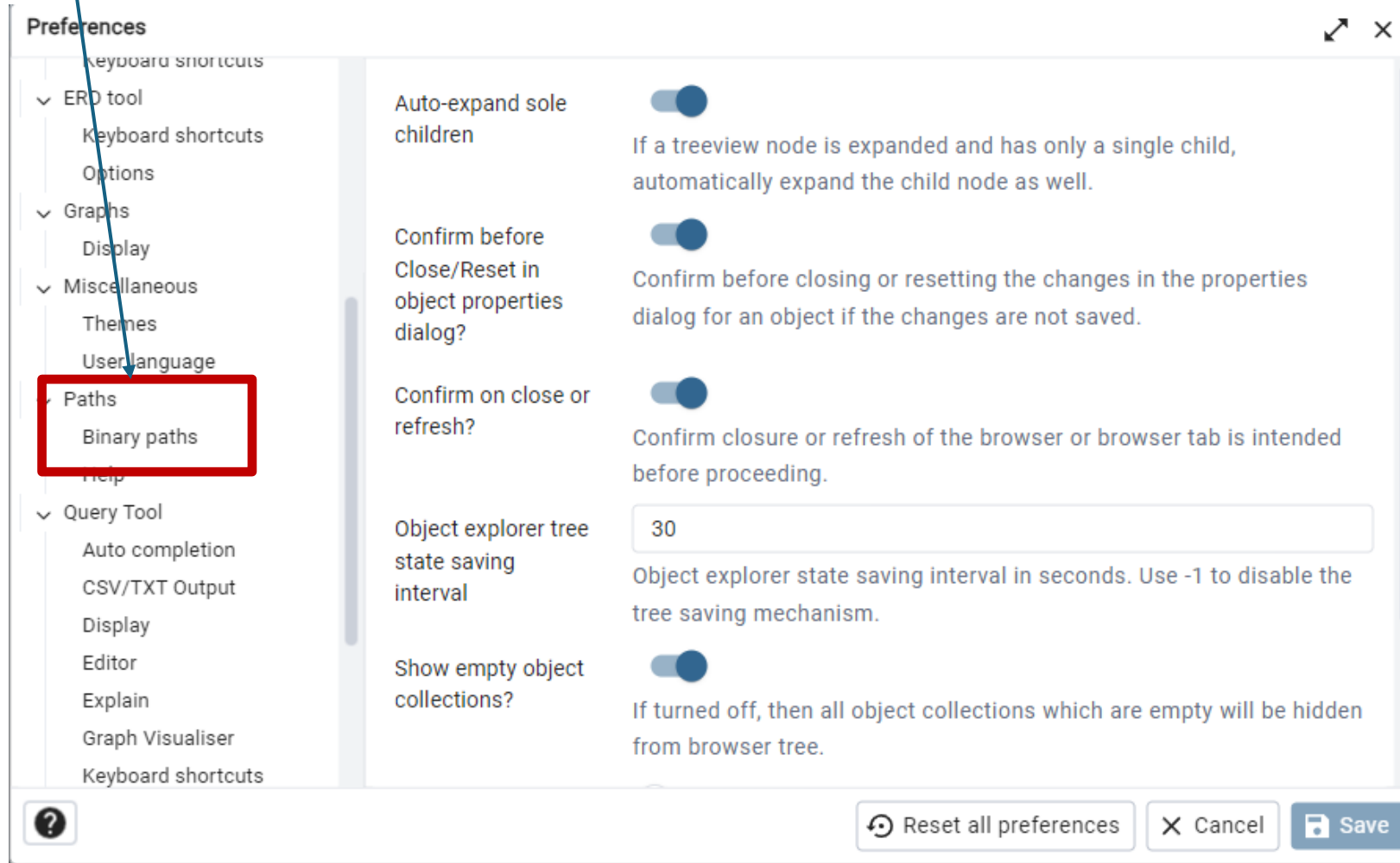
Failed (exit code: 1).

Execution time: 0.33 seconds



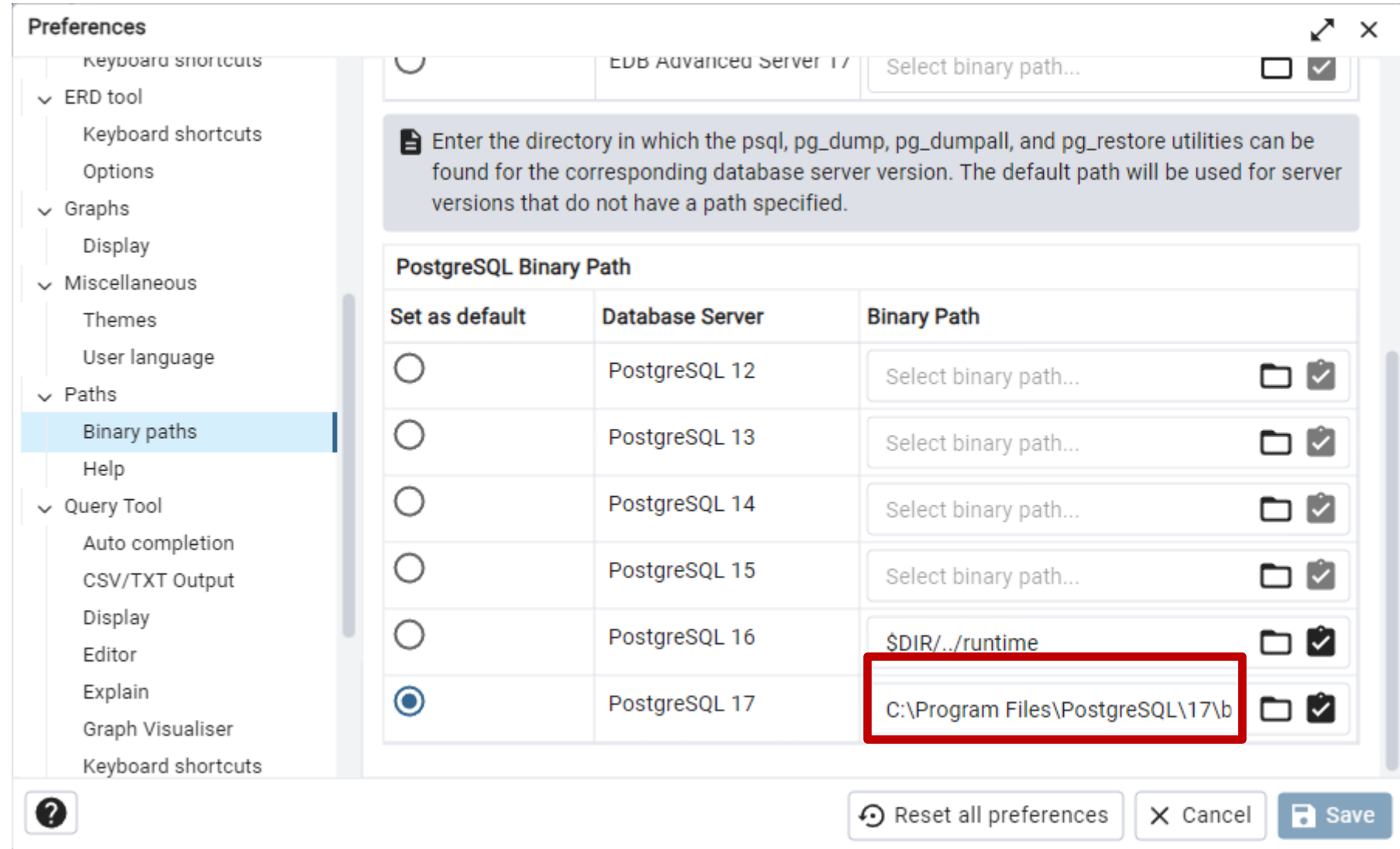
# Setting PostgreSQL 17 as the default binary path

- Click on “Binary path”



# Setting PostgreSQL 17 as the default binary path

- Scroll down and click on empty space.
- Type in “C:\Program Files\PostgreSQL\17\bin:
- Then click on the button “Set as default”
- Click on the “Save” button






# Backing up the SQL folder

- Type in the file name for this backup file
- Then click “Backup” button
- You should see double green rectangles if the process is successful
- The sql file will now appear in your directory where you save it

Backup (Database: DataCoSupplyChain)

General Data Options Query Options Table Options Options Objects

Filename 


Format Custom



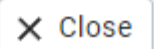


Compression ratio

Encoding Select an item...

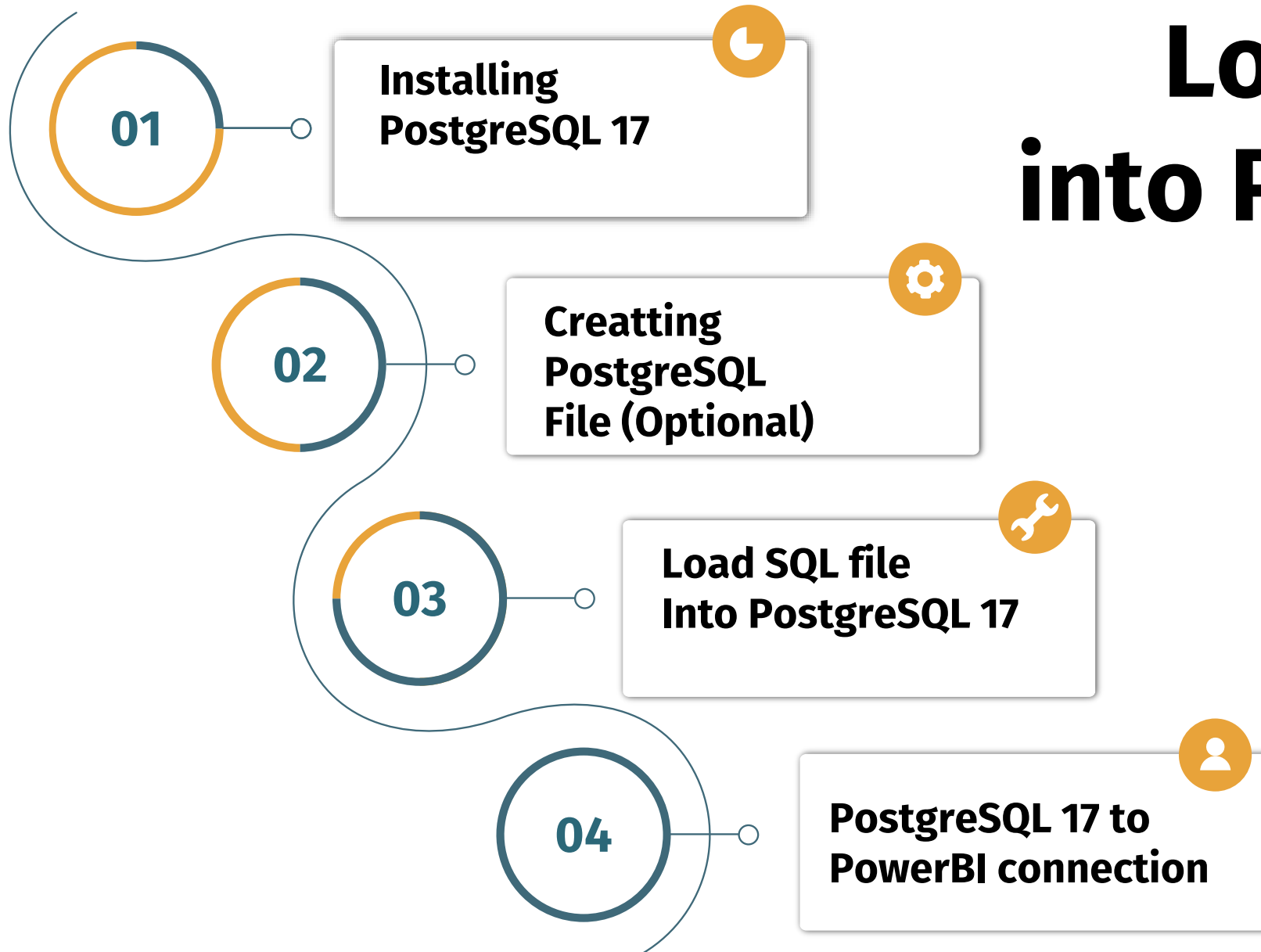
Number of jobs

Role name Select an item...

 Please provide a filename.

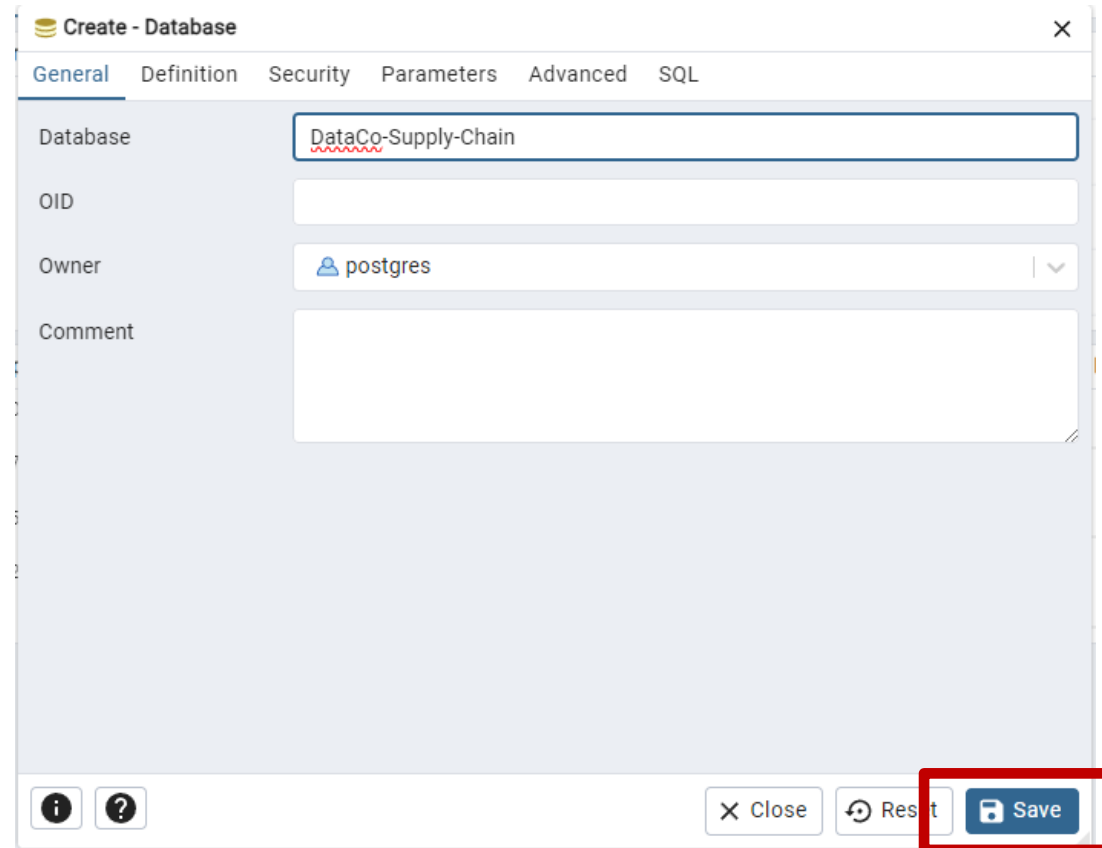
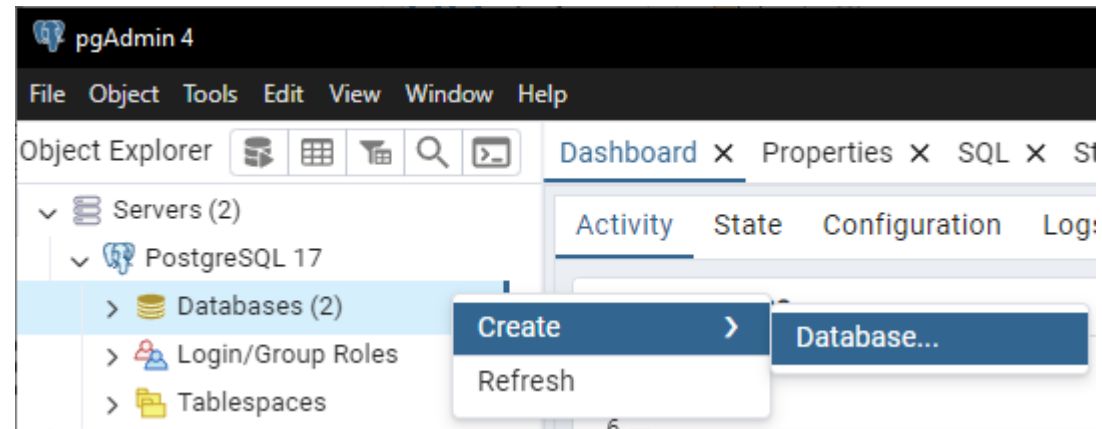
    

# Load SQL file into PostgreSQL



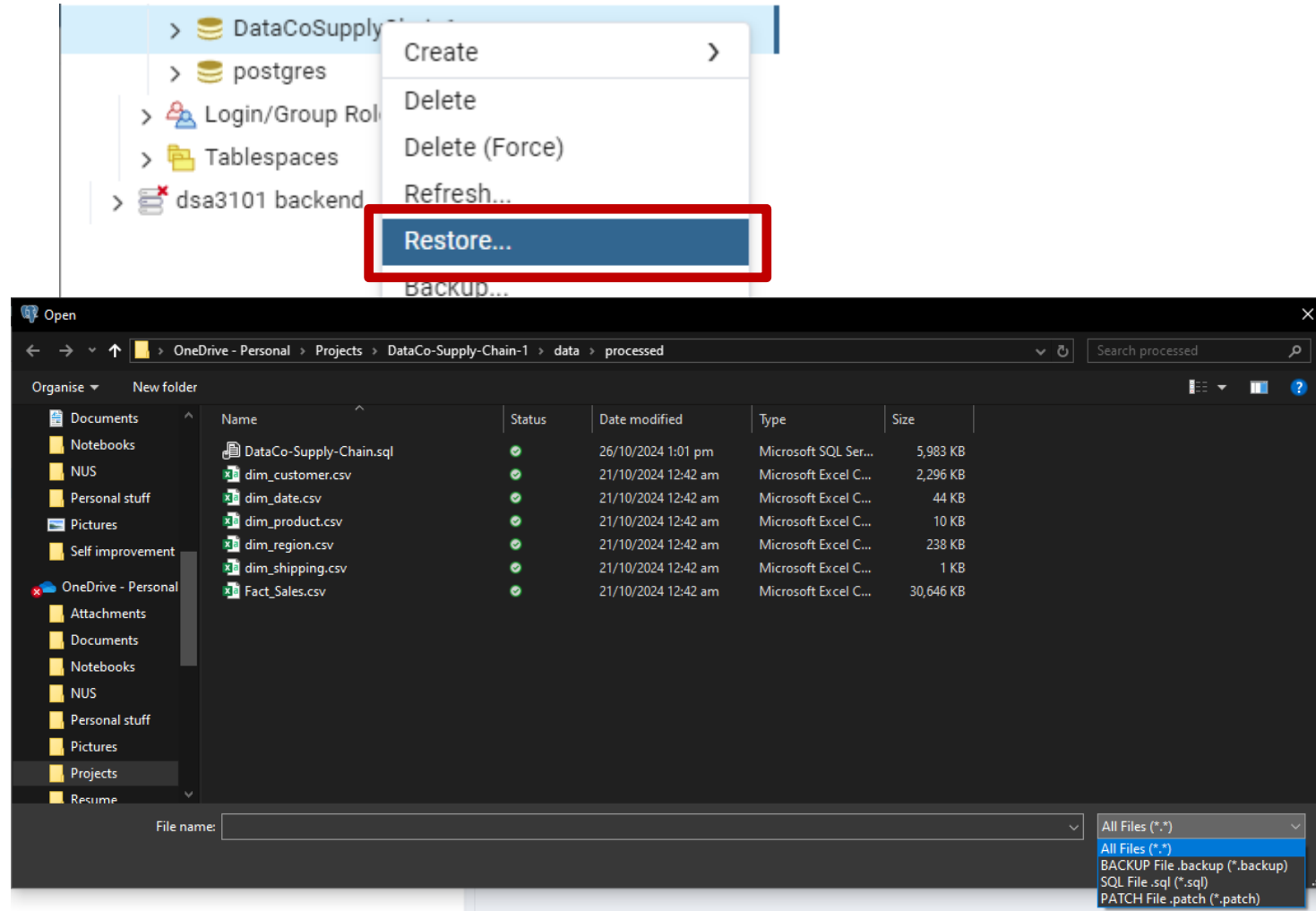
# Creating the Database

- *If you completed the steps in slide 22, skip this step.*
- Right click on “Databases”
- Click on “Create”
- Then click on “Database”
- A Create-Database window will appear
- Type in “DataCo-Supply-Chain”
- Click on “Save”



# Restoring the database

- Right Click on DataCoSupplyChain database
- Click on the “Restore” button
- Go to the directory where you stored your repository.
- Cd to C:\<your directory>\data\processed\DataCo-Supply-Chain.sql
- Click on All Files (\*.\*) to view all the files



# Restoring the database

- Click on "Restore"

Restore (Database: DataCoSupplyChain)

General

Data Options

Query Options

Table Options

Options

Format

Custom or tar

Filename

C:\Users\Loo Guan Yee\OneDrive\Projects\DataCo-Supply-Chain-1\dat

Number of jobs

Role name

Select an item...

i

?

Close

Reset

Restore

# Check if you have double green status upon restoring the database

Process completed



Restoring backup on the server 'PostgreSQL 17 (localhost:5432)'



View Processes

Process started



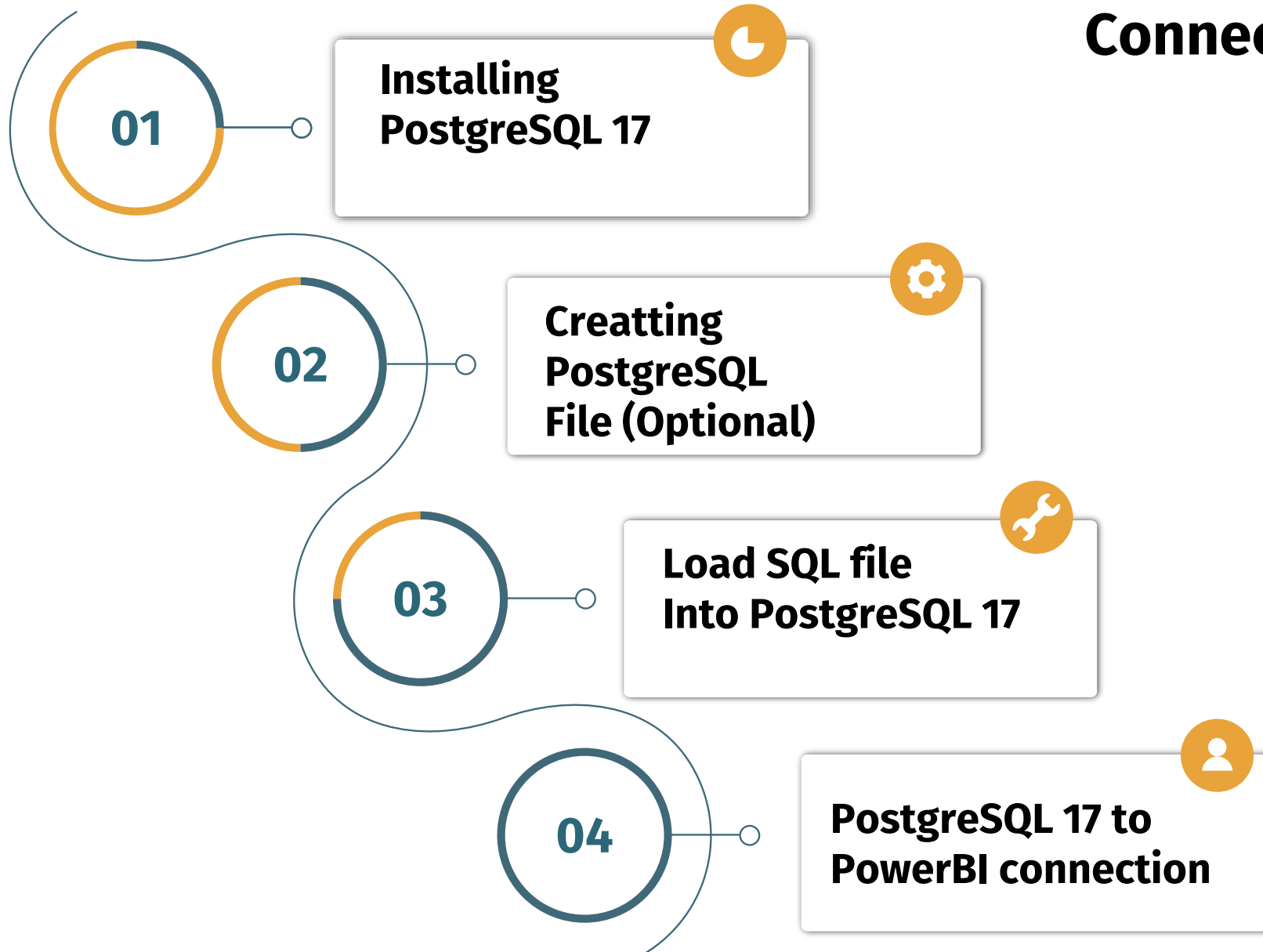
Restoring backup on the server 'PostgreSQL 17 (localhost:5432)'



View Processes

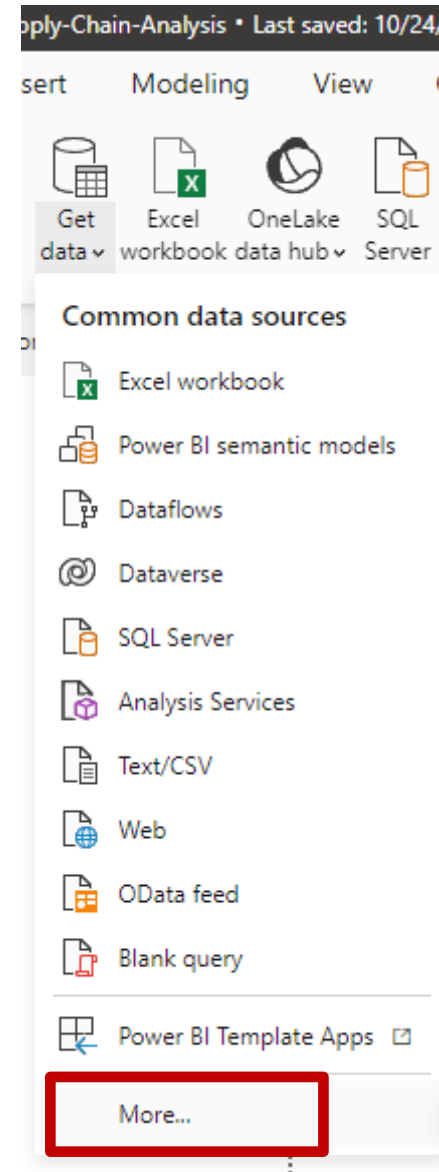
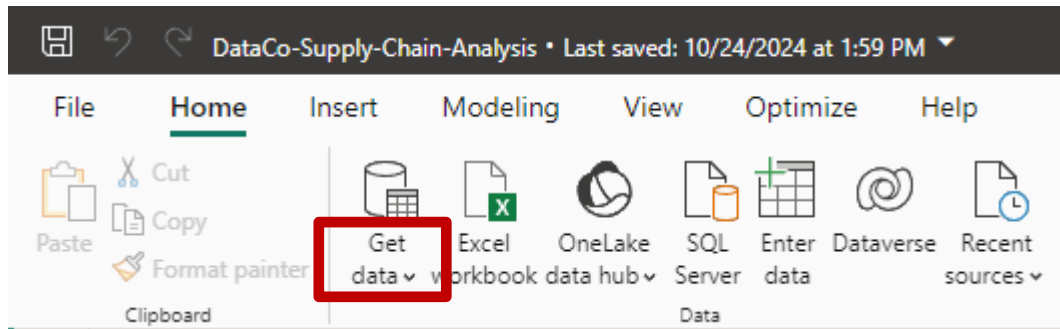
# Connecting PostgreSQL to PowerBI

DataCo Ecommerce  
Platform Sales



# Open the Power Bi file from PowerBI folder

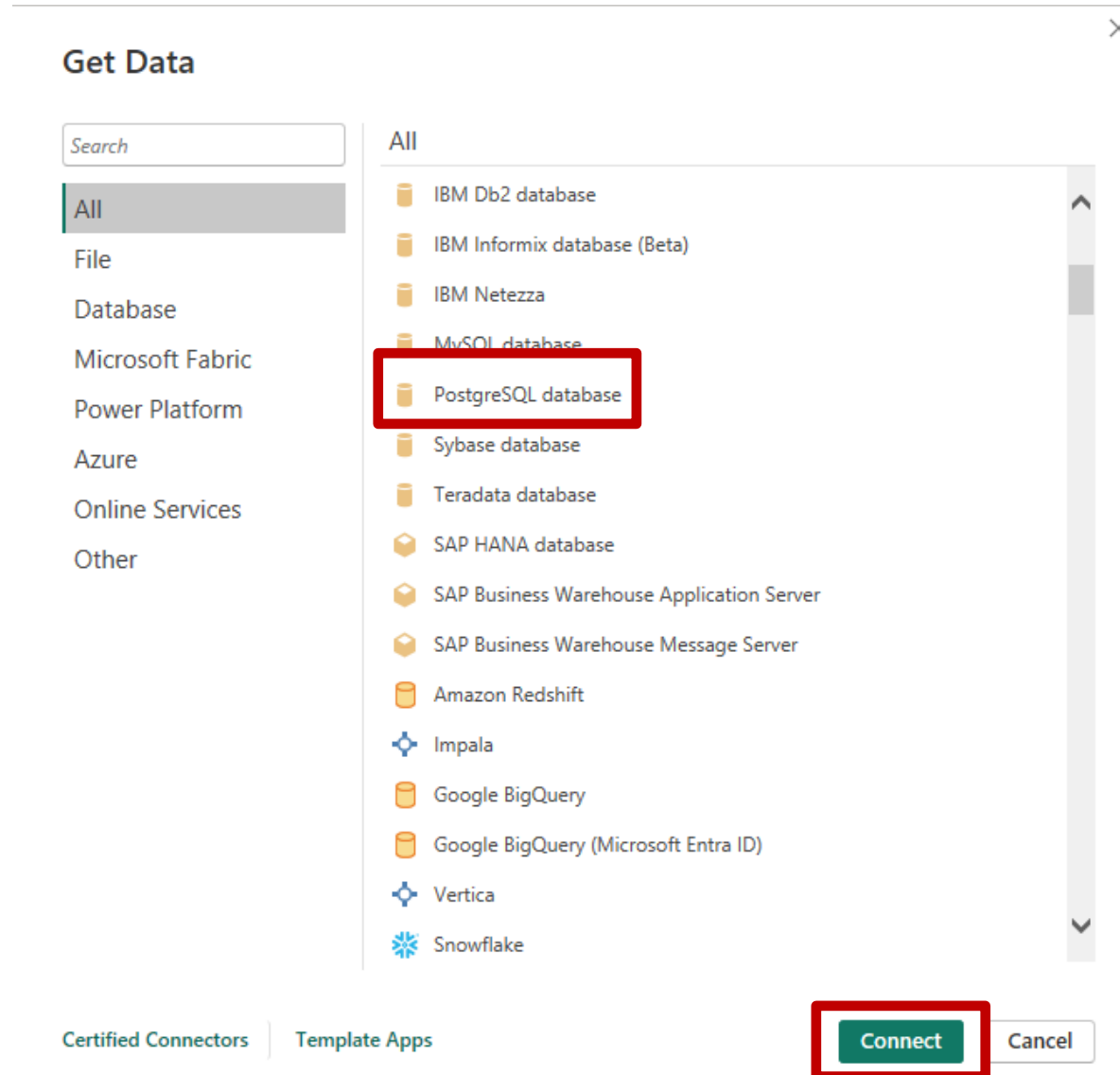
- Open the DataCo-Supply-Chain.pbix file from the PowerBI folder
- Click on Get data button
- Click on more button





# Connecting PostgreSQL to PowerBI

- Scroll down the Get Data window
- Click on PostgreSQL database
- Click on “Connect”



# Open the Power Bi file from PowerBI folder

- For *Server*: Type in “localhost”
- For *Database*: Type in “DataCoSupplyChain”
- For *User name*: Type in “postgre”
- For *Password*: Type in the password that you set when installing PostgreSQL earlier.
- Then click on Connect

The image displays two screenshots of the PostgreSQL database connection wizard in Power BI.

**Top Screenshot: PostgreSQL database configuration**

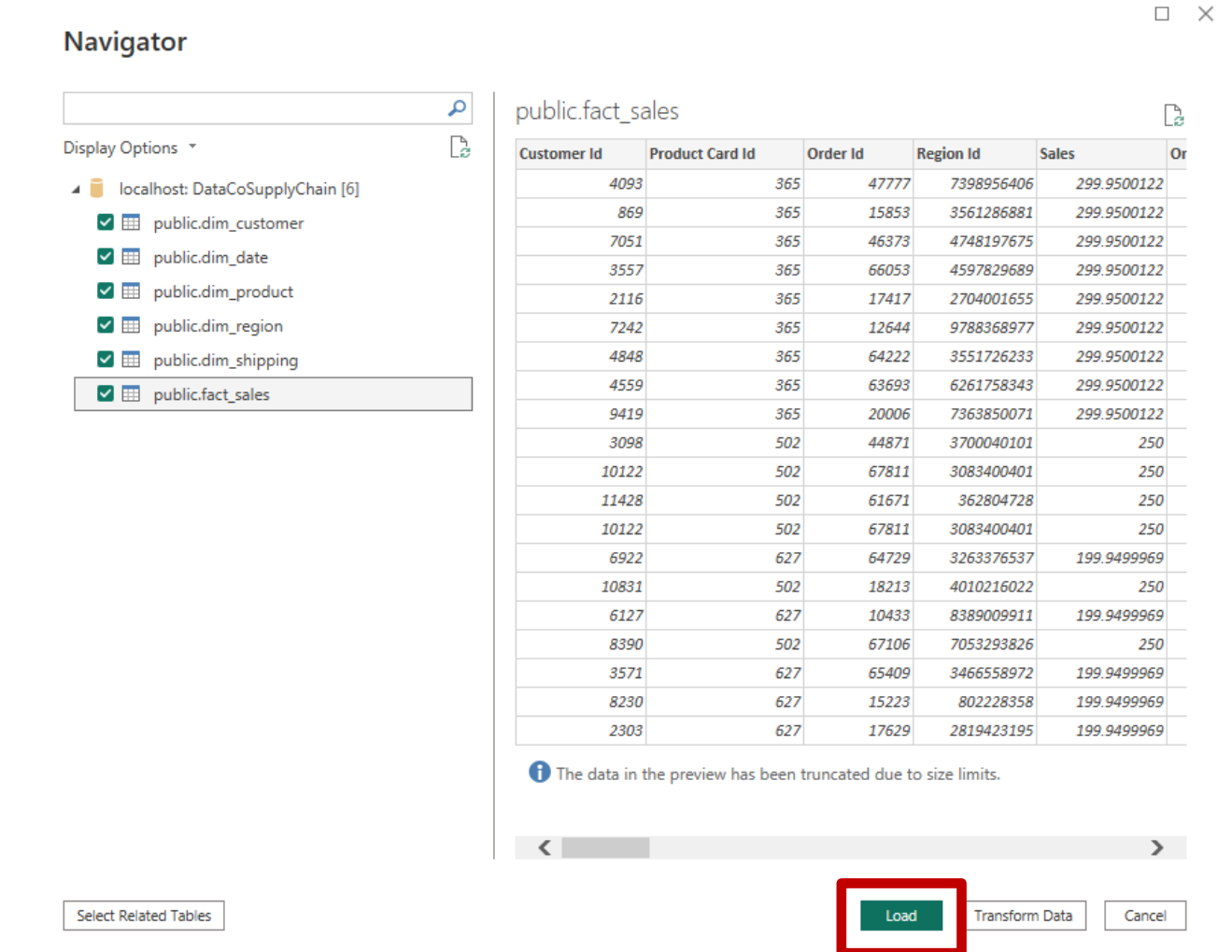
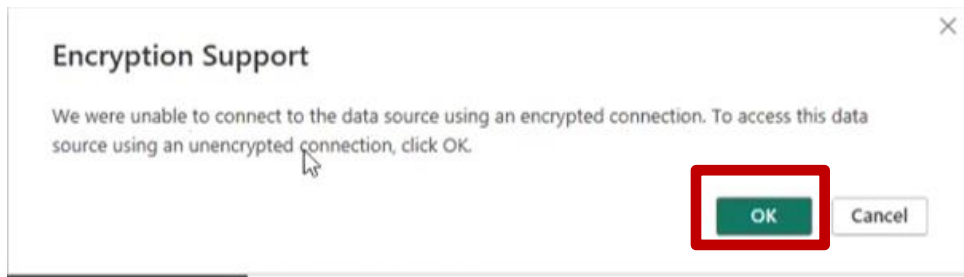
- Server:** localhost
- Database:** DataCoSupplyChain
- Data Connectivity mode:** Import (selected), DirectQuery
- Buttons:** OK (highlighted with a red box), Cancel

**Bottom Screenshot: PostgreSQL database connection confirmation**

- Database:** localhost; DataCoSupplyChain
- User name:** (empty field)
- Password:** (empty field)
- Select which level to apply these settings to:** localhost
- Buttons:** Back, Connect (highlighted with a red box), Cancel

# Open the Power Bi file from PowerBI folder

- For *Encryption Support*, Click on Ok
- Check all the ticks for the boxes under the localhost:DataCoSupplyChain
- Click on the Load button



The image shows the Power BI interface. On the left is the "Navigator" pane, which lists data sources under "localhost: DataCoSupplyChain [6]". The following tables are listed with checkboxes: public.dim\_customer, public.dim\_date, public.dim\_product, public.dim\_region, public.dim\_shipping, and public.fact\_sales. The "public.fact\_sales" table is selected and highlighted. Below the list is a "Select Related Tables" button. On the right is the "Data Preview" pane for the "public.fact\_sales" table. It shows a table with columns: Customer Id, Product Card Id, Order Id, Region Id, Sales, and Order Id (partially visible). The table contains 20 rows of data. Below the table is a message: "The data in the preview has been truncated due to size limits." At the bottom of the interface, there are three buttons: "Load", "Transform Data", and "Cancel". The "Load" button is highlighted with a red rectangular box.

Customer Id	Product Card Id	Order Id	Region Id	Sales	Order Id
4093	365	47777	7398956406	299.9500122	
869	365	15853	3561286881	299.9500122	
7051	365	46373	4748197675	299.9500122	
3557	365	66053	4597829689	299.9500122	
2116	365	17417	2704001655	299.9500122	
7242	365	12644	9788368977	299.9500122	
4848	365	64222	3551726233	299.9500122	
4559	365	63693	6261758343	299.9500122	
9419	365	20006	7363850071	299.9500122	
3098	502	44871	3700040101	250	
10122	502	67811	3083400401	250	
11428	502	61671	362804728	250	
10122	502	67811	3083400401	250	
6922	627	64729	3263376537	199.9499969	
10831	502	18213	4010216022	250	
6127	627	10433	8389009911	199.9499969	
8390	502	67106	7053293826	250	
3571	627	65409	3466558972	199.9499969	
8230	627	15223	802228358	199.9499969	
2303	627	17629	2819423195	199.9499969	