DataCo Supply Chain Documentation

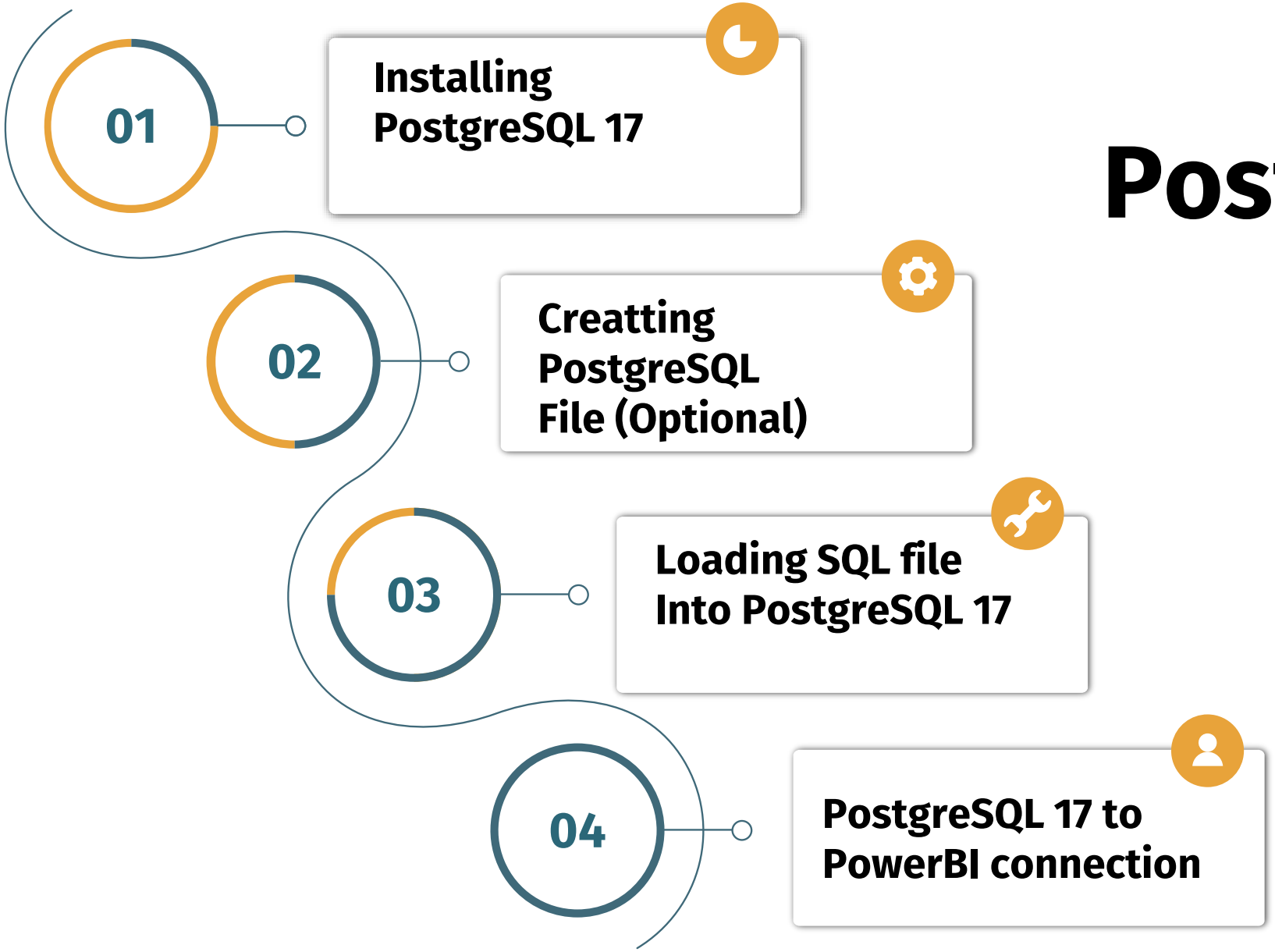
Loo Guan Yee

October 2024

Documentation Content

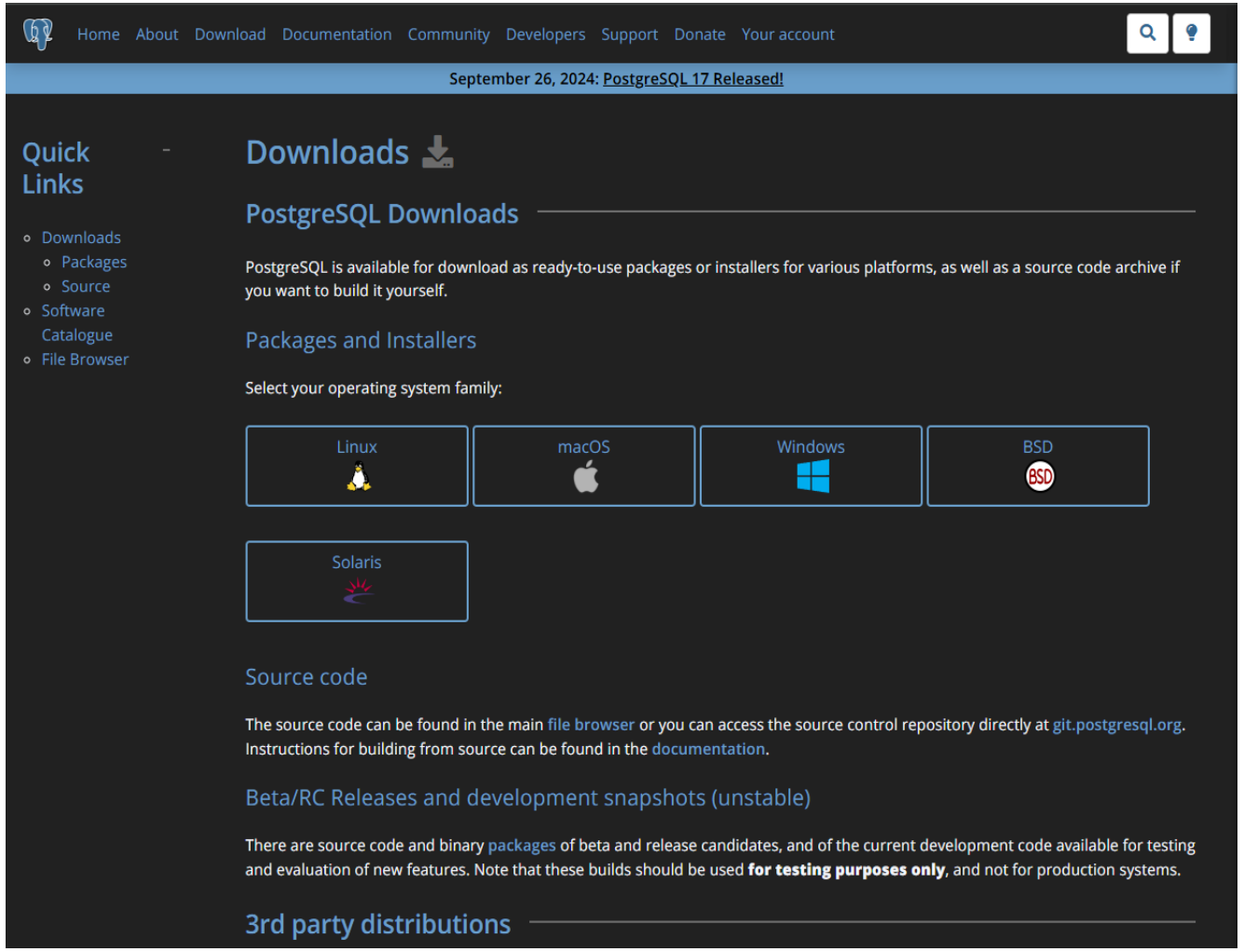
- **Installing PostgreSQL 17**
- **Creating the SQL file from scratch (Optional)**
  - Getting and preprocessing the DataCo Supply Chain Dataset
  - Creating the sql file in PostgreSQL 17
  - Backing up the sql file into PostgreSQL 17
- **Loading the sql file into PostgreSQL 17**
- **Connecting the PostgreSQL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

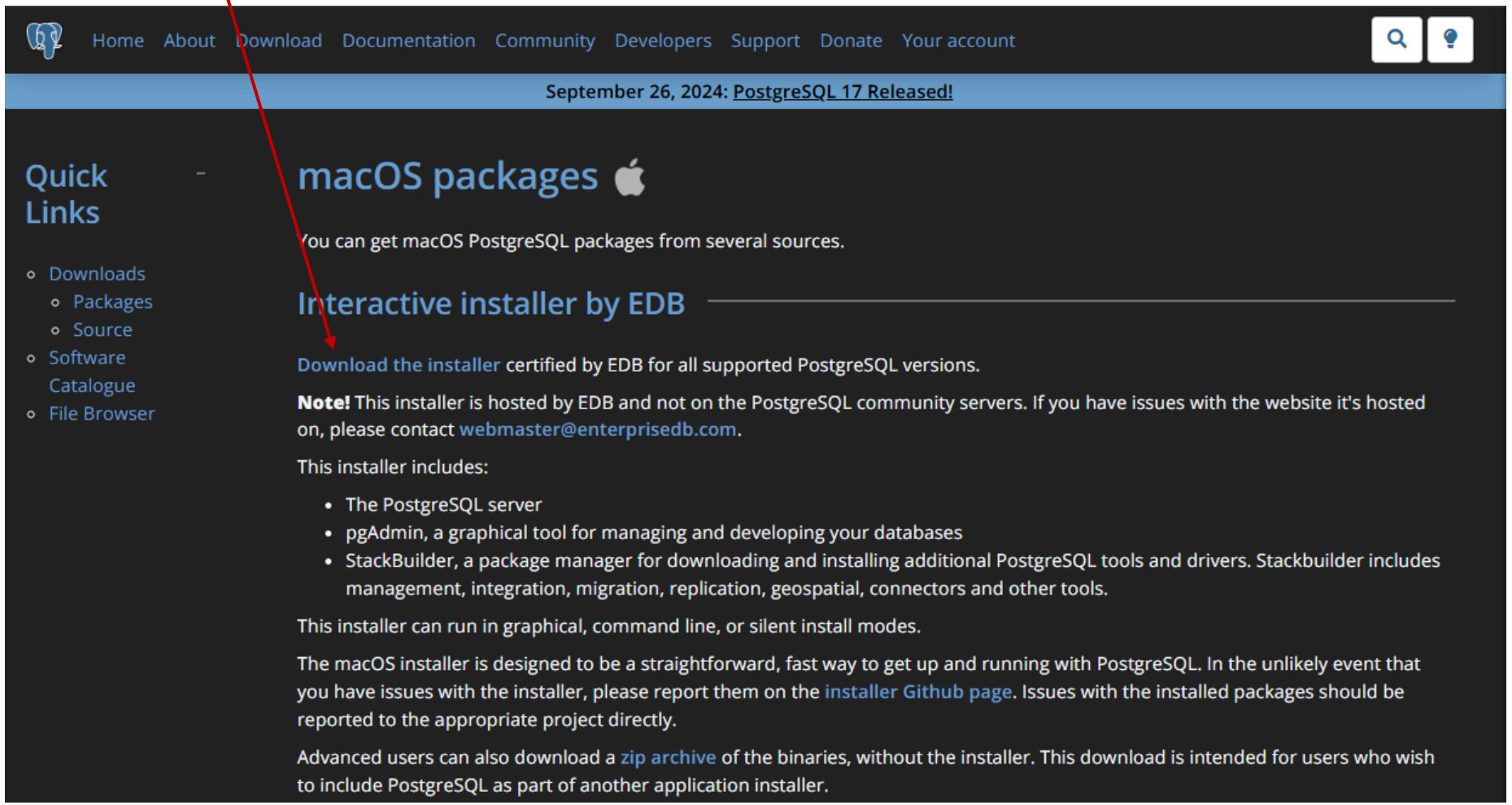


PostgreSQL Windows Installation

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

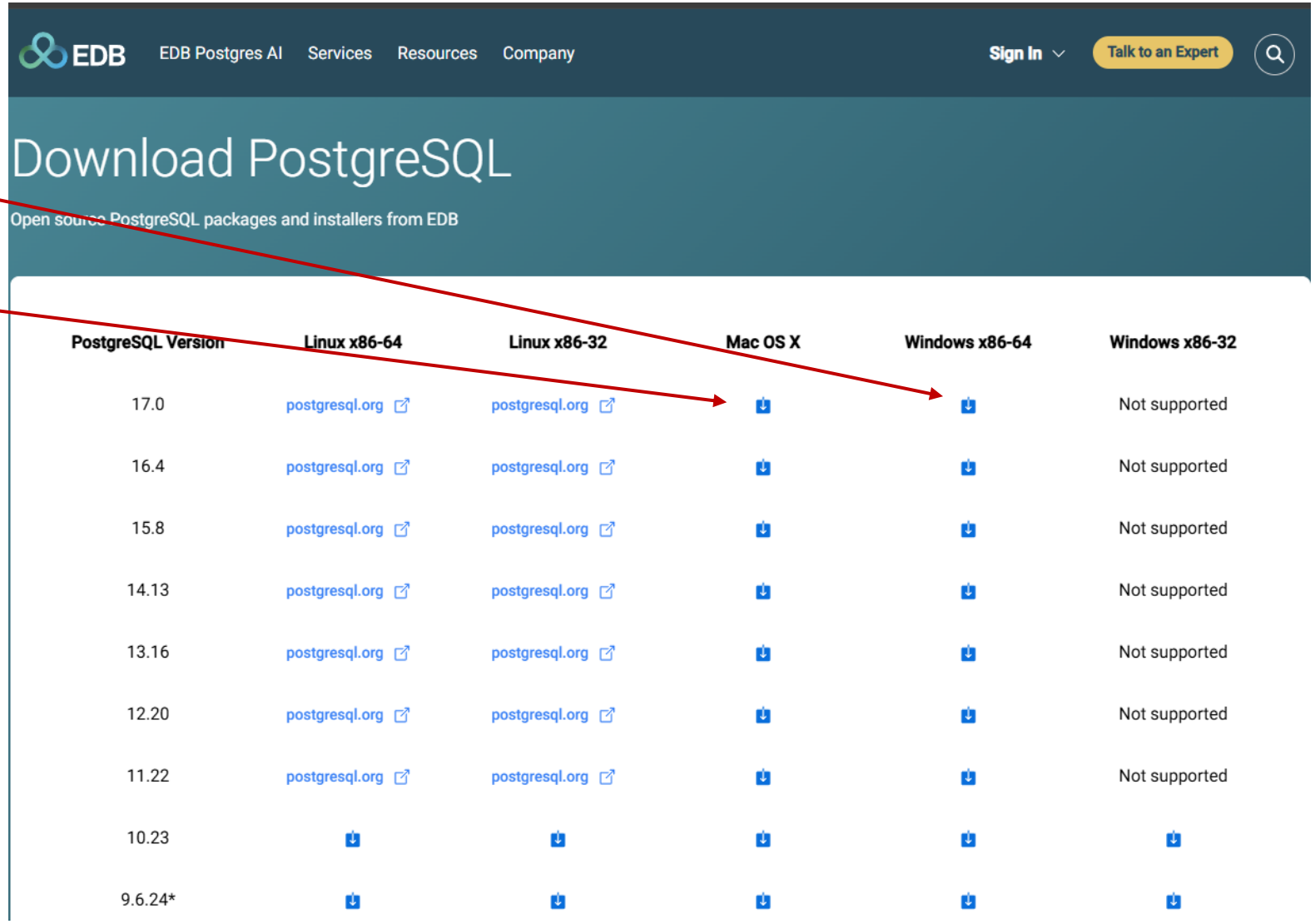
PostgreSQL MacOS Installation

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



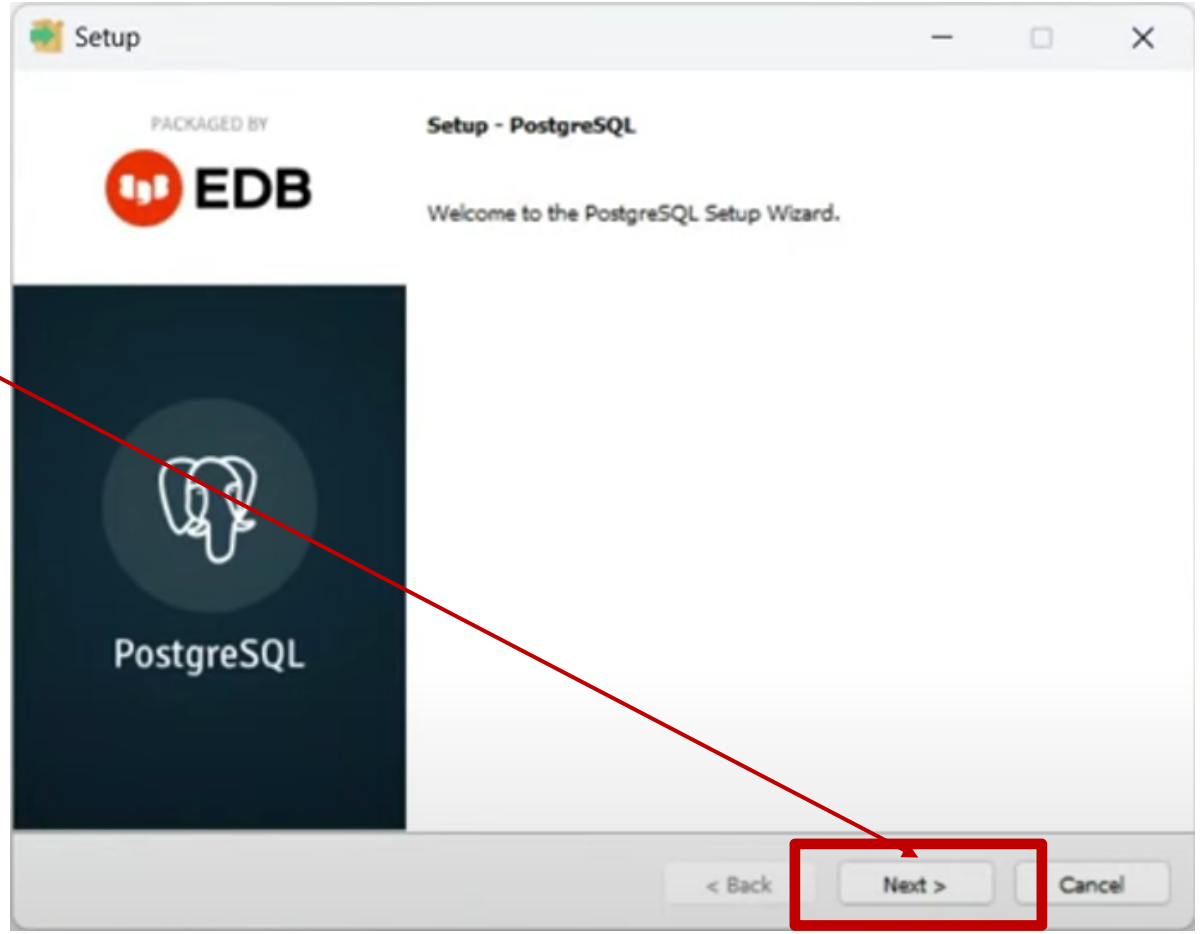
PostgreSQL Windows and Mac OS Installation

- 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



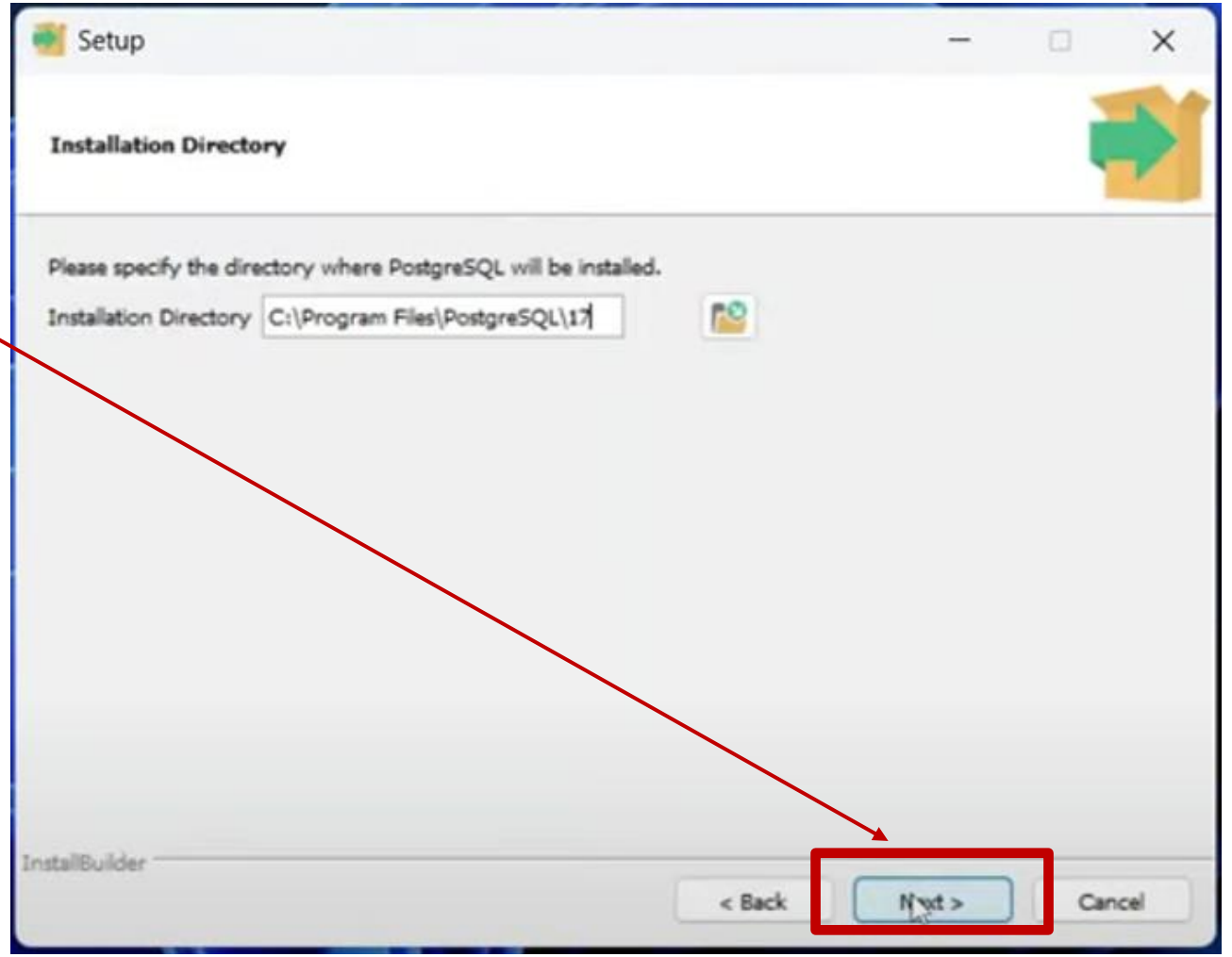
PostgreSQL.exe installation

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



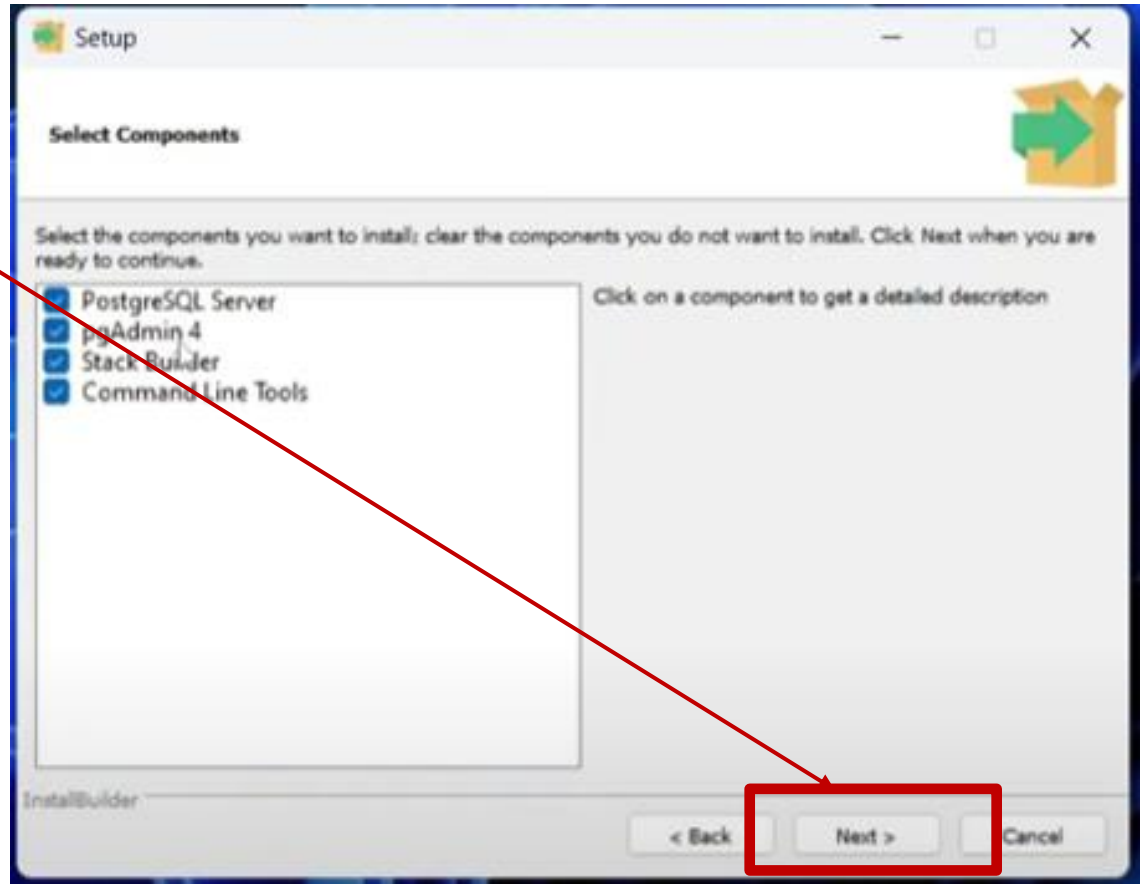
PostgreSQL.exe installation

- Click on next



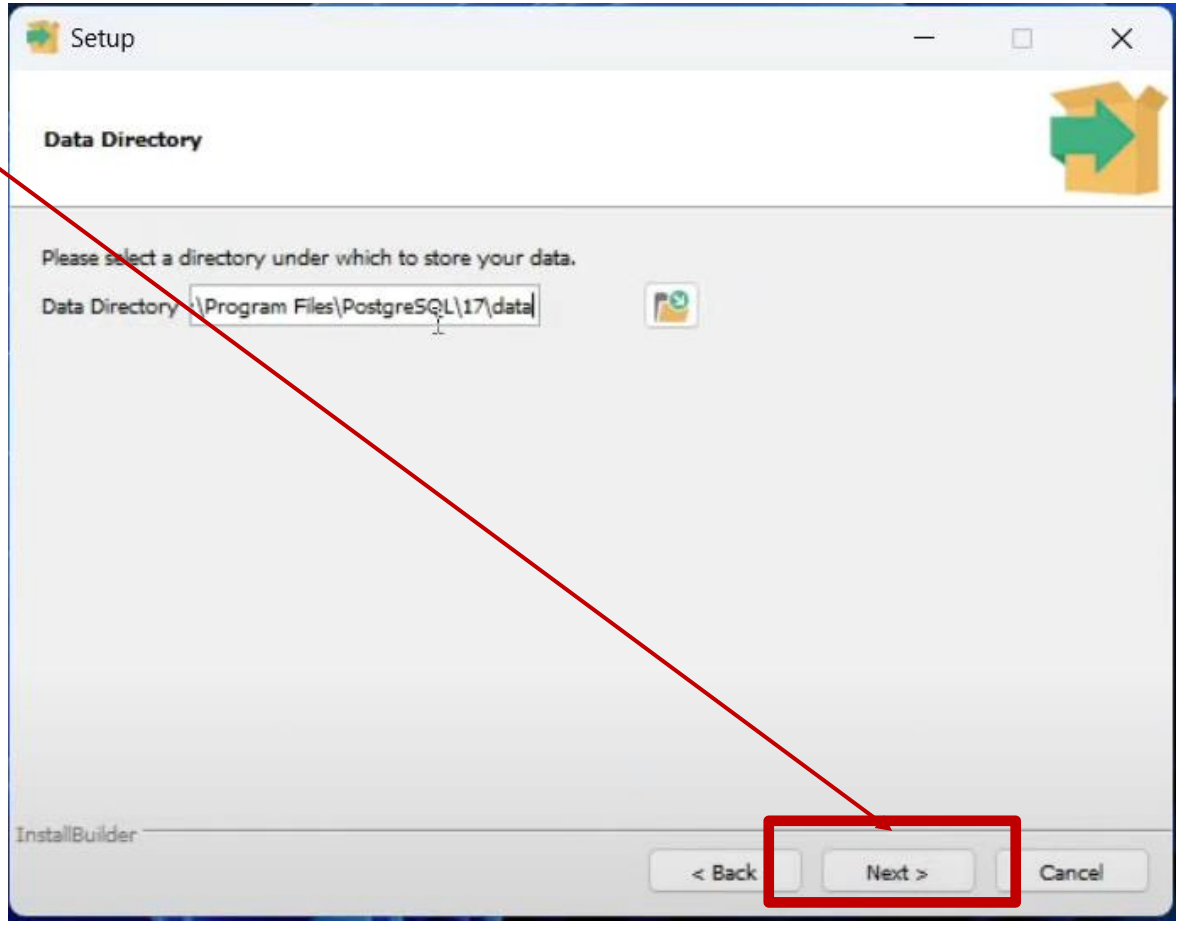
PostgreSQL.exe installation

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



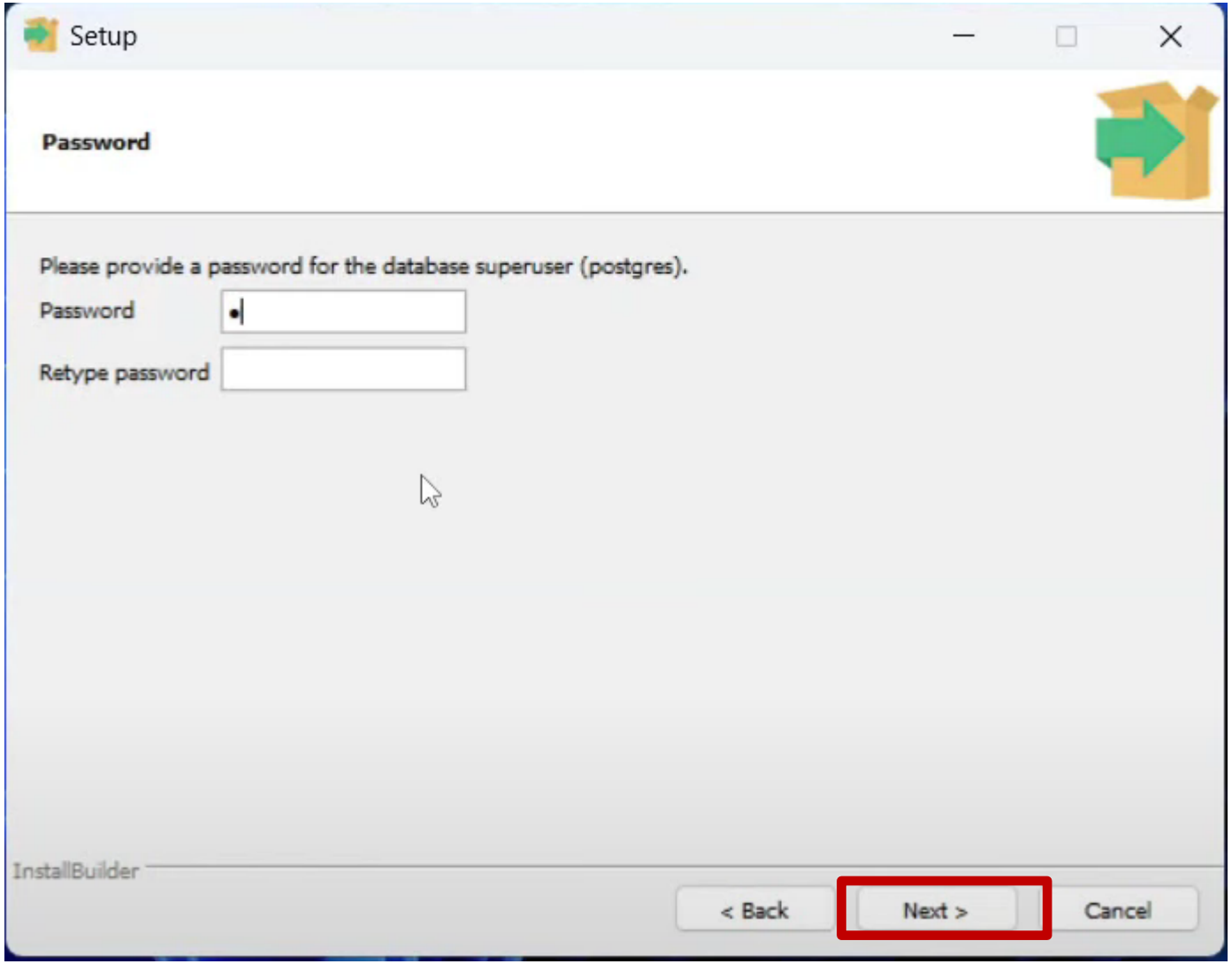
PostgreSQL.exe installation

- Click on next



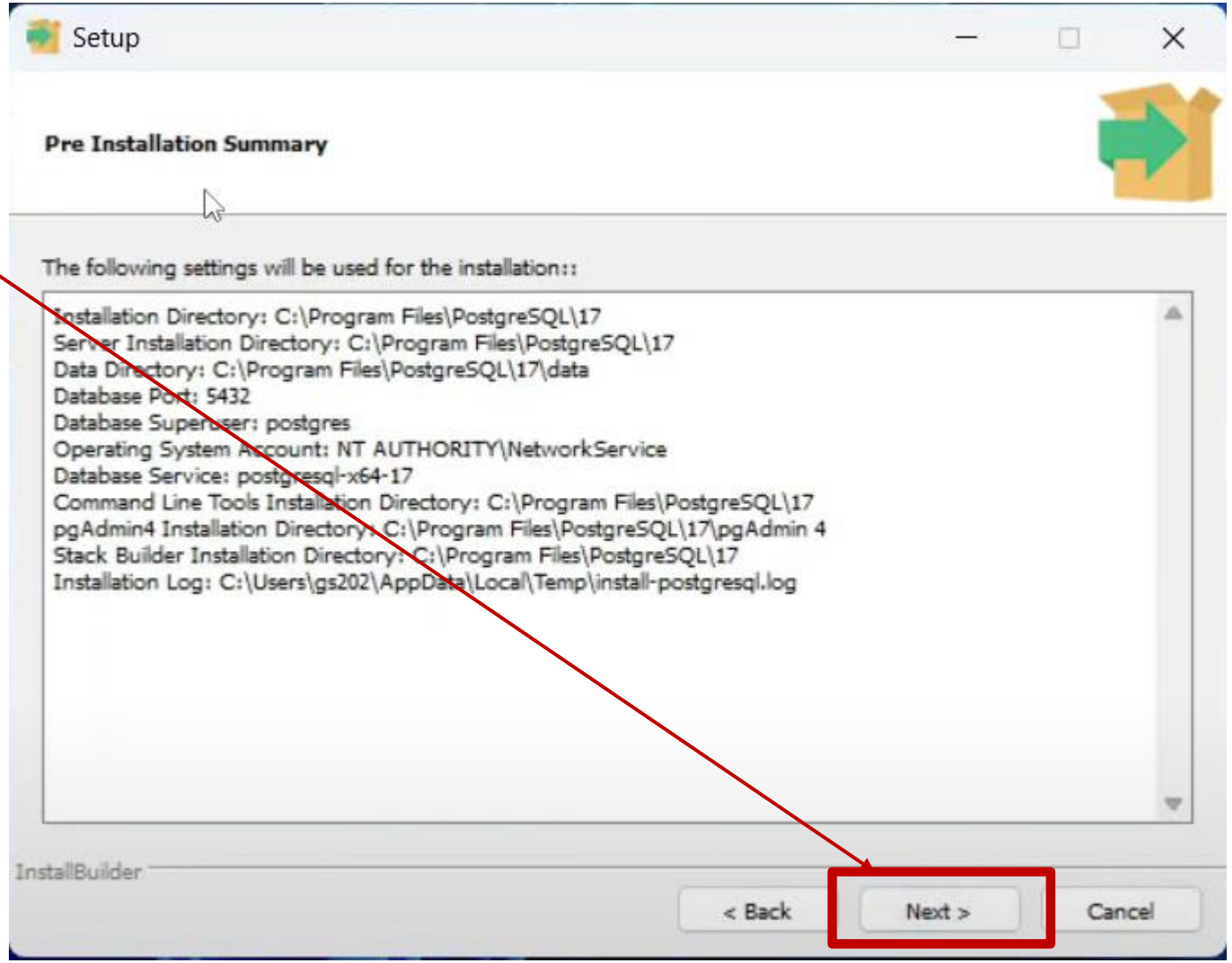
PostgreSQL.exe installation (Important!)

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



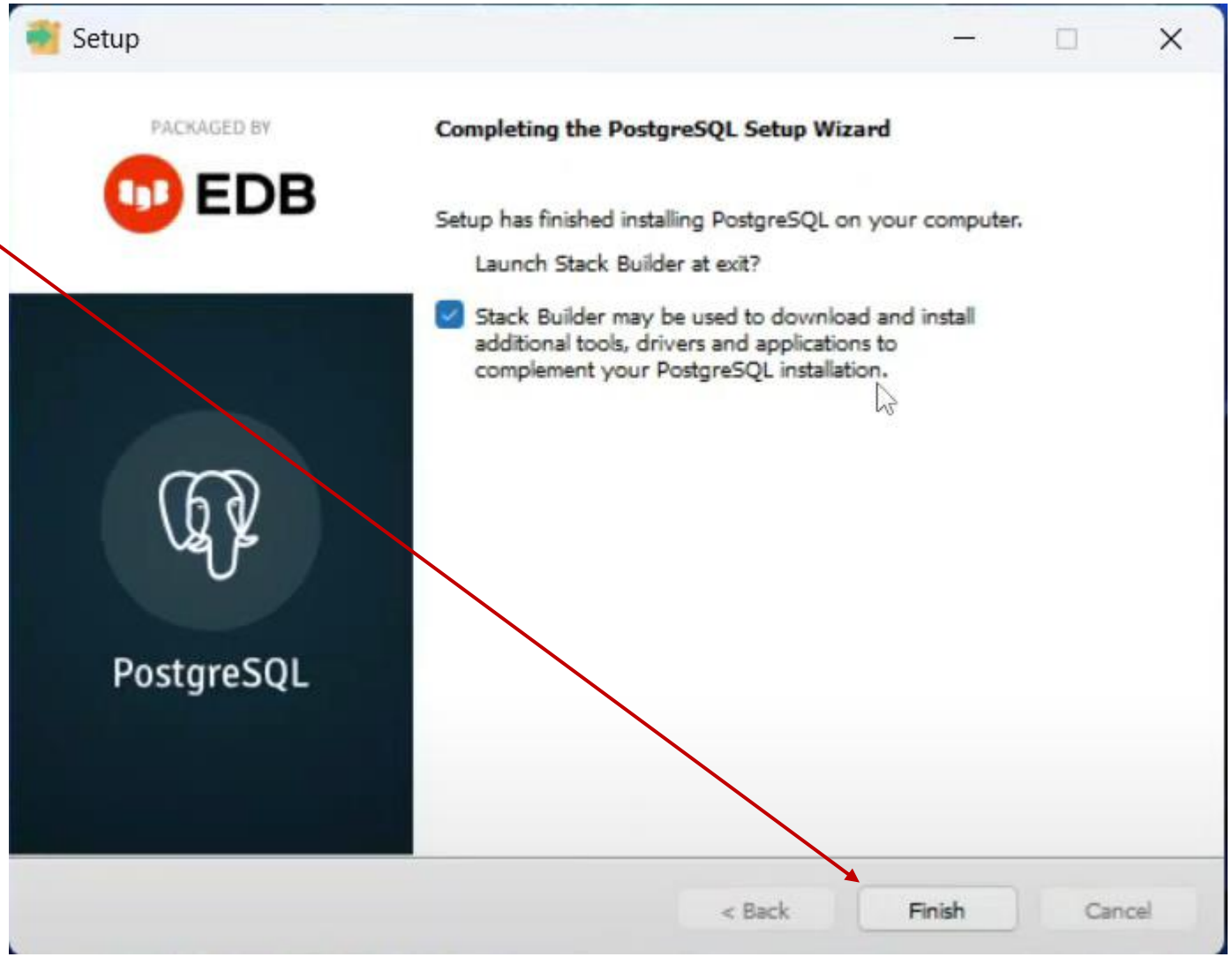
PostgreSQL.exe installation

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



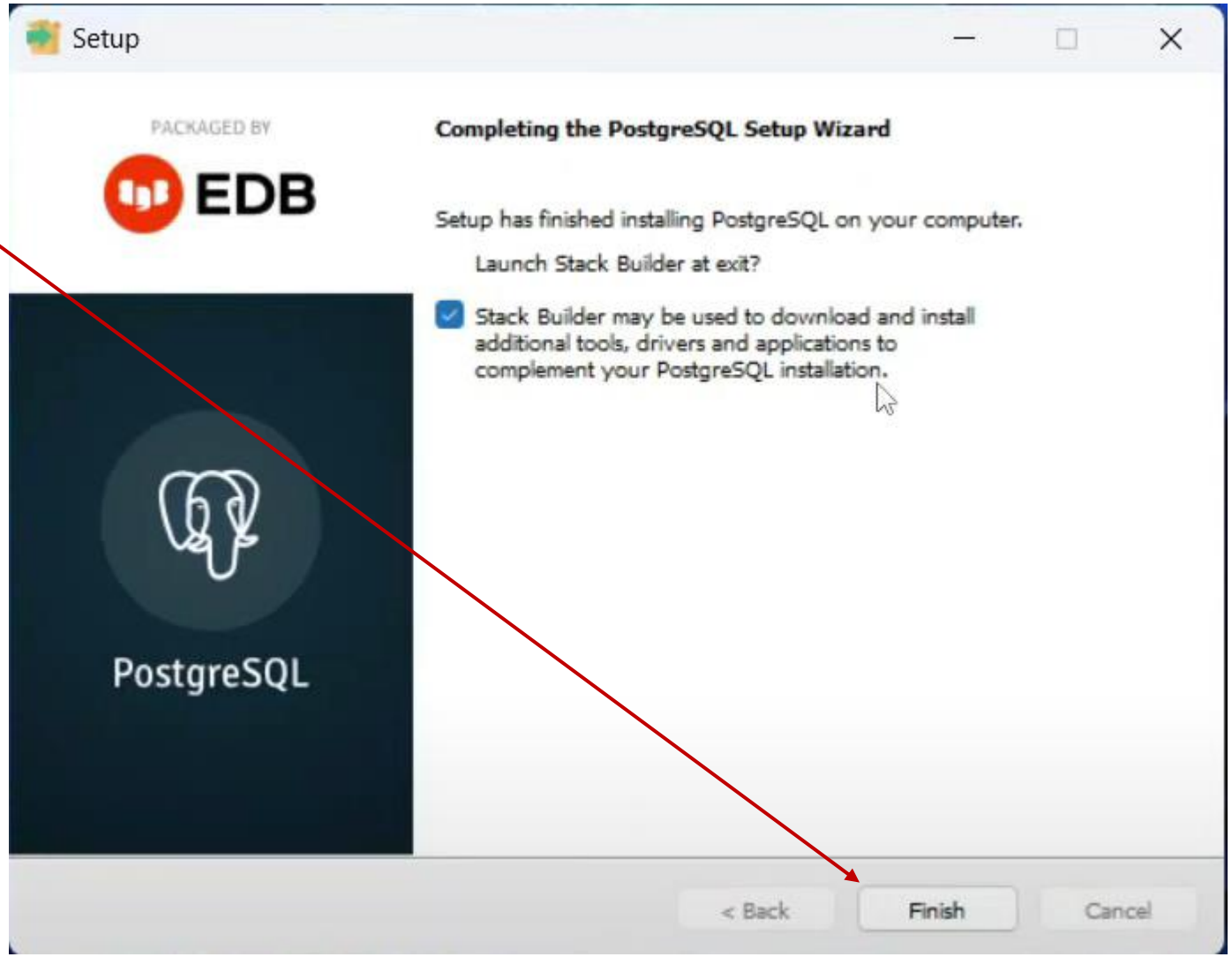
PostgreSQL.exe installation

- Click on “Finish”



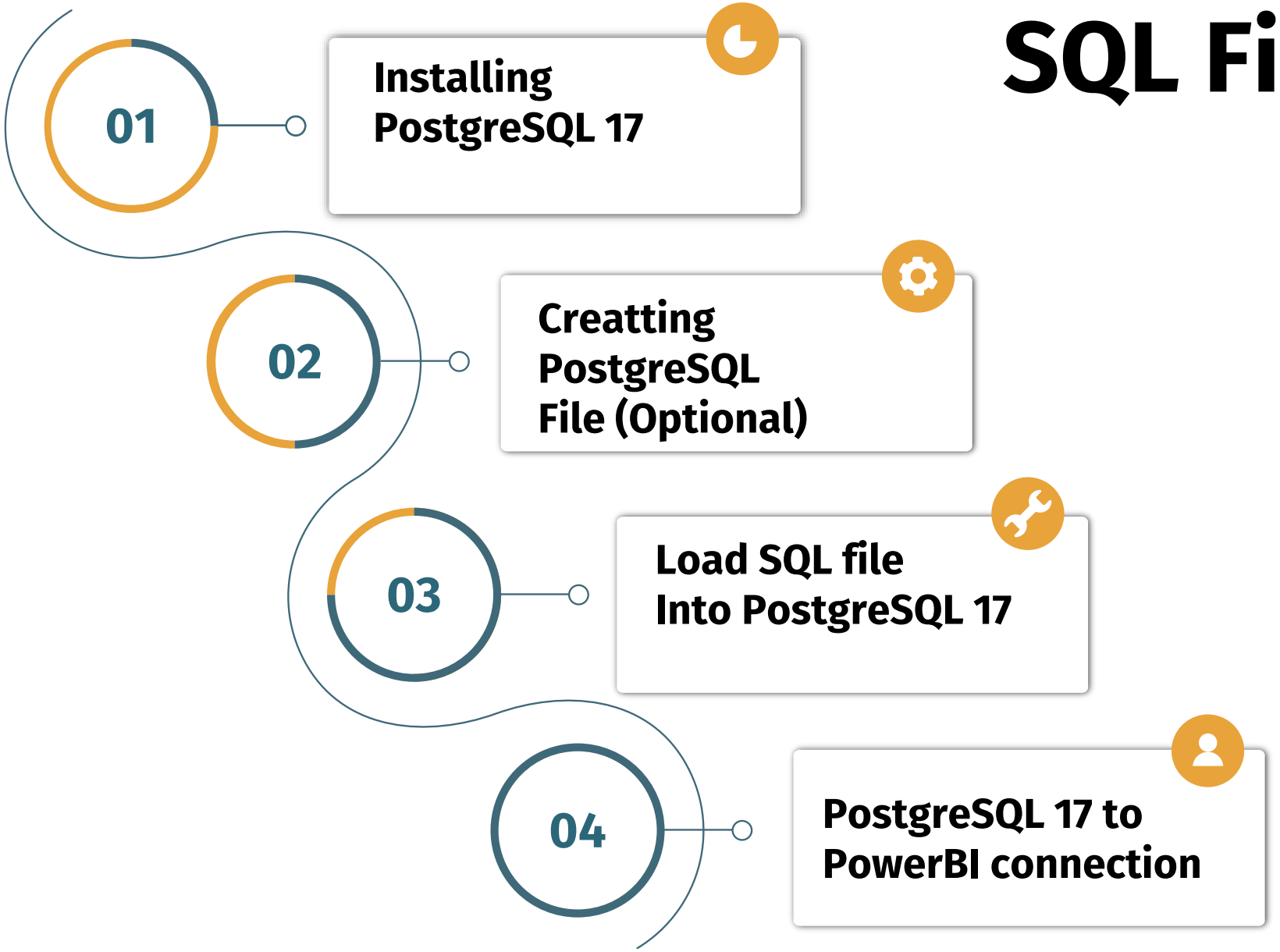
Open Pg admin4

- Click on “Finish”



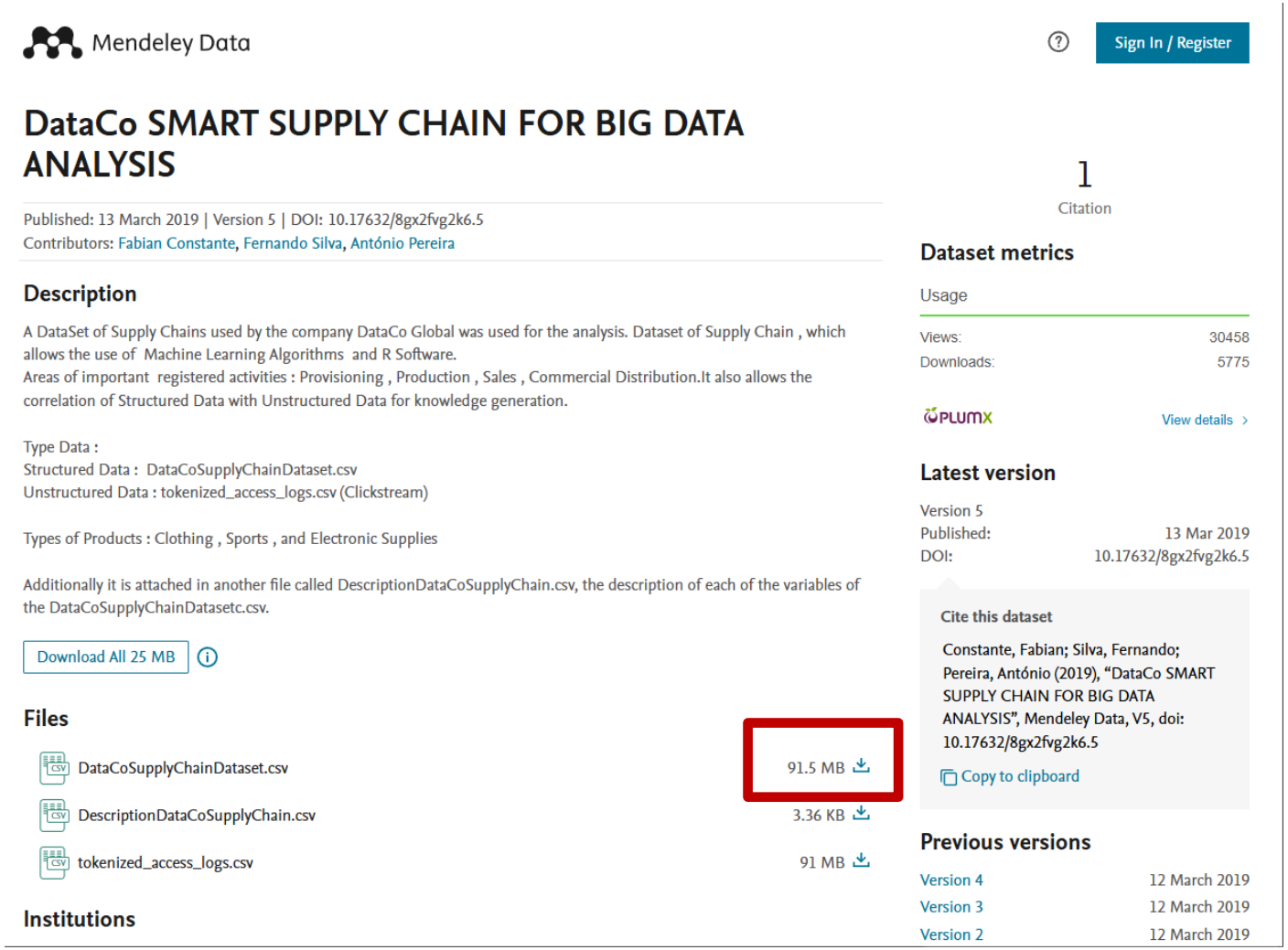
SQL File creation

Getting the DataCo Supply Chain Dataset



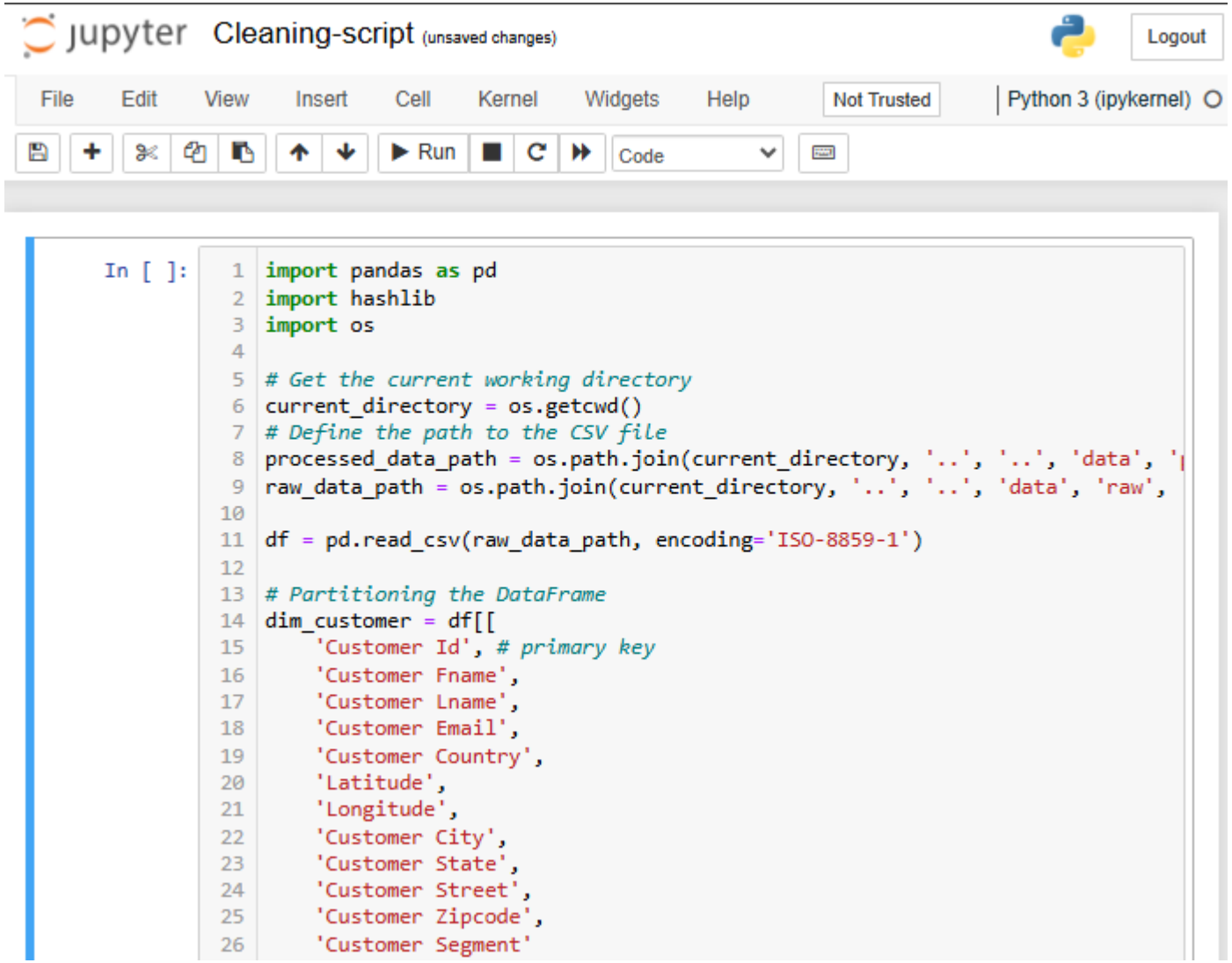
Obtaining the DataCo Supply Chain dataset

- 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*



Cleaning the DataCo Supply Chain dataset

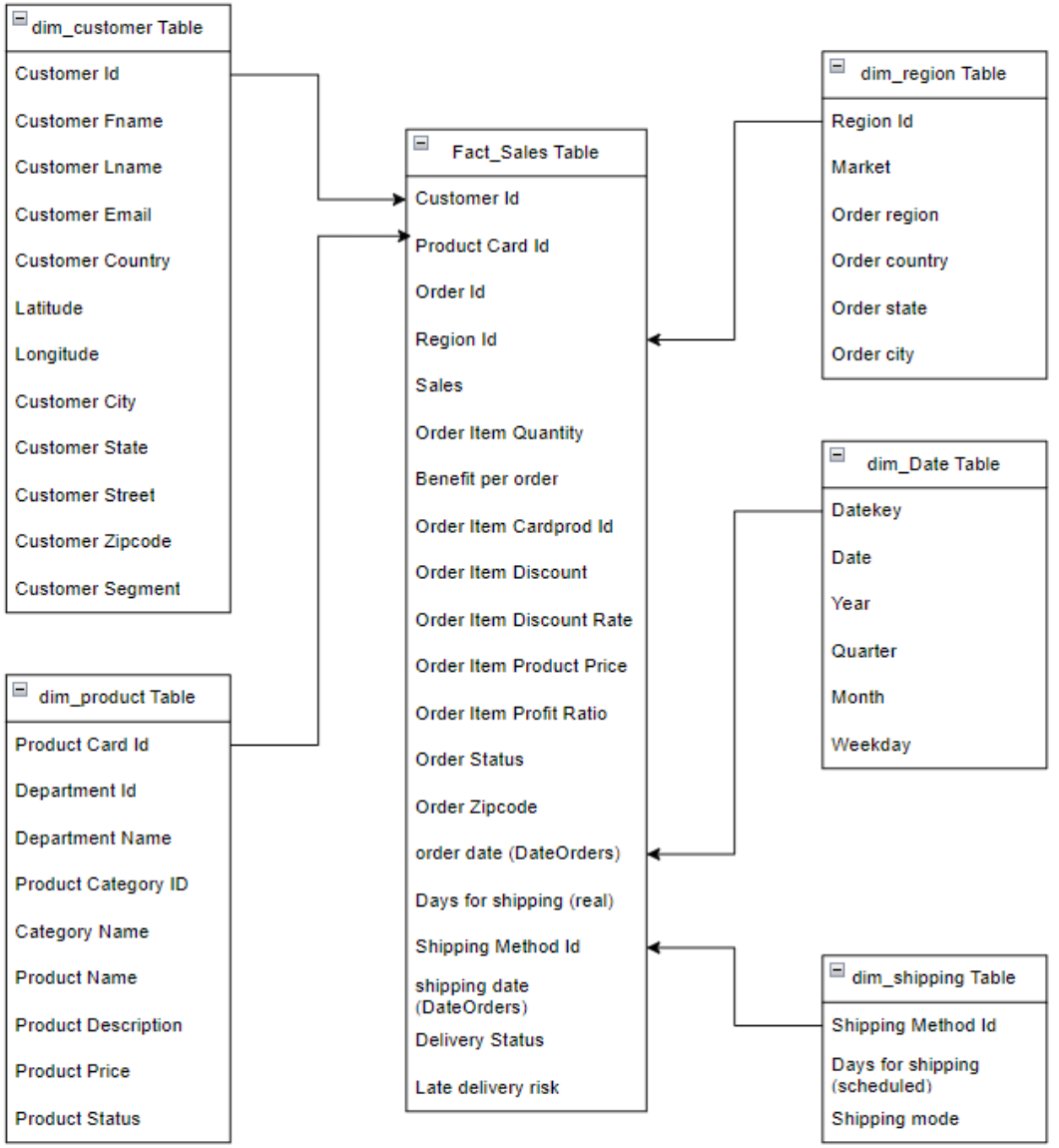
- Open your Powershell or Git Bash
- Navigate to the folder where you git clone the DataCo-supply-chain repository
- Cd to *scripts/python*
- Open the ipynb folder and click on run for this cell
    - It should take about 30 seconds to run this cell



Cleaning the DataCo Supply Chain dataset

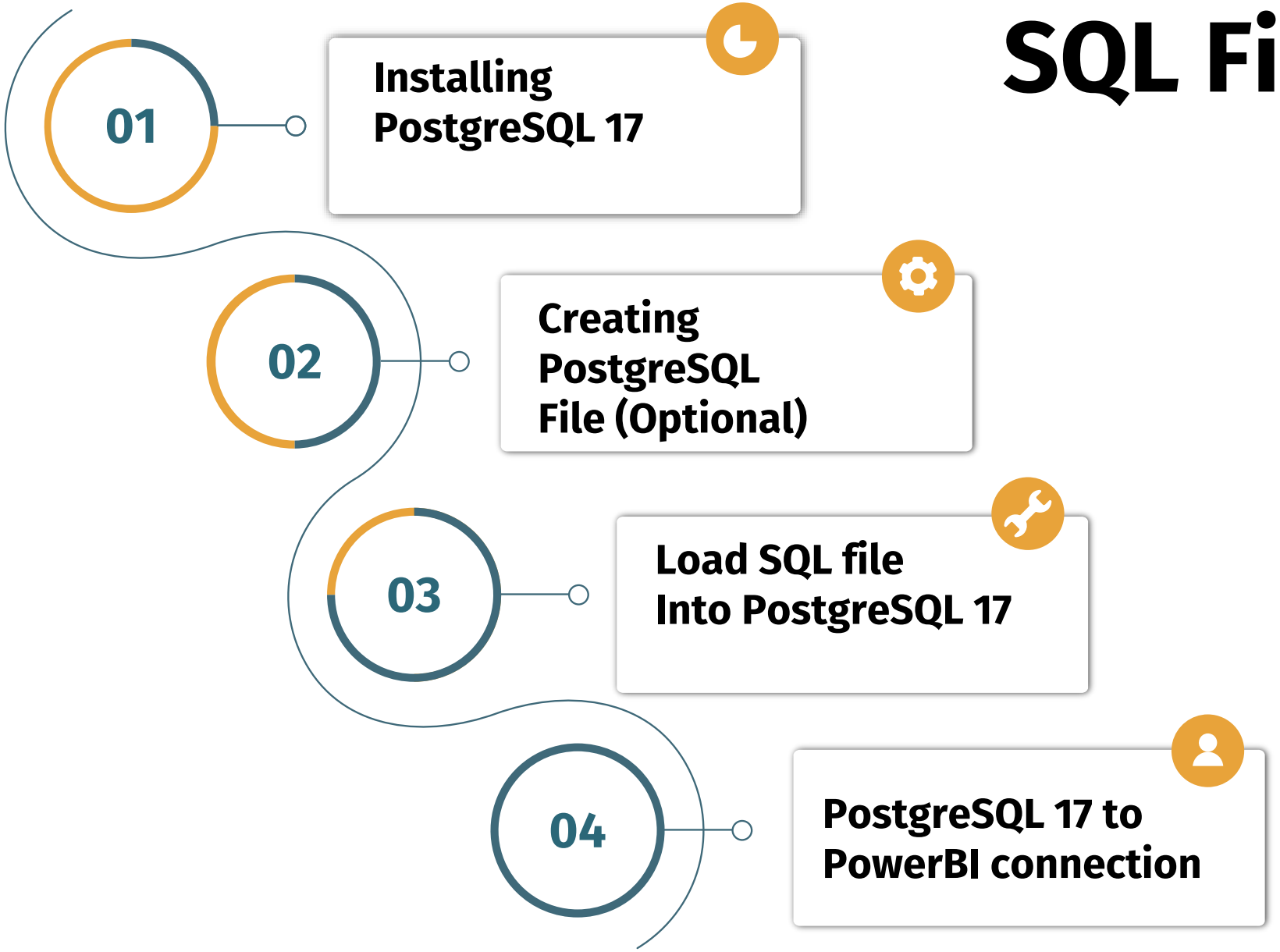
- You should be able to see the dimension and fact files based on this star schema in the data/processed directory
- *Shipping Method Id, Datekey, and Region Id are new columns created from this python script

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



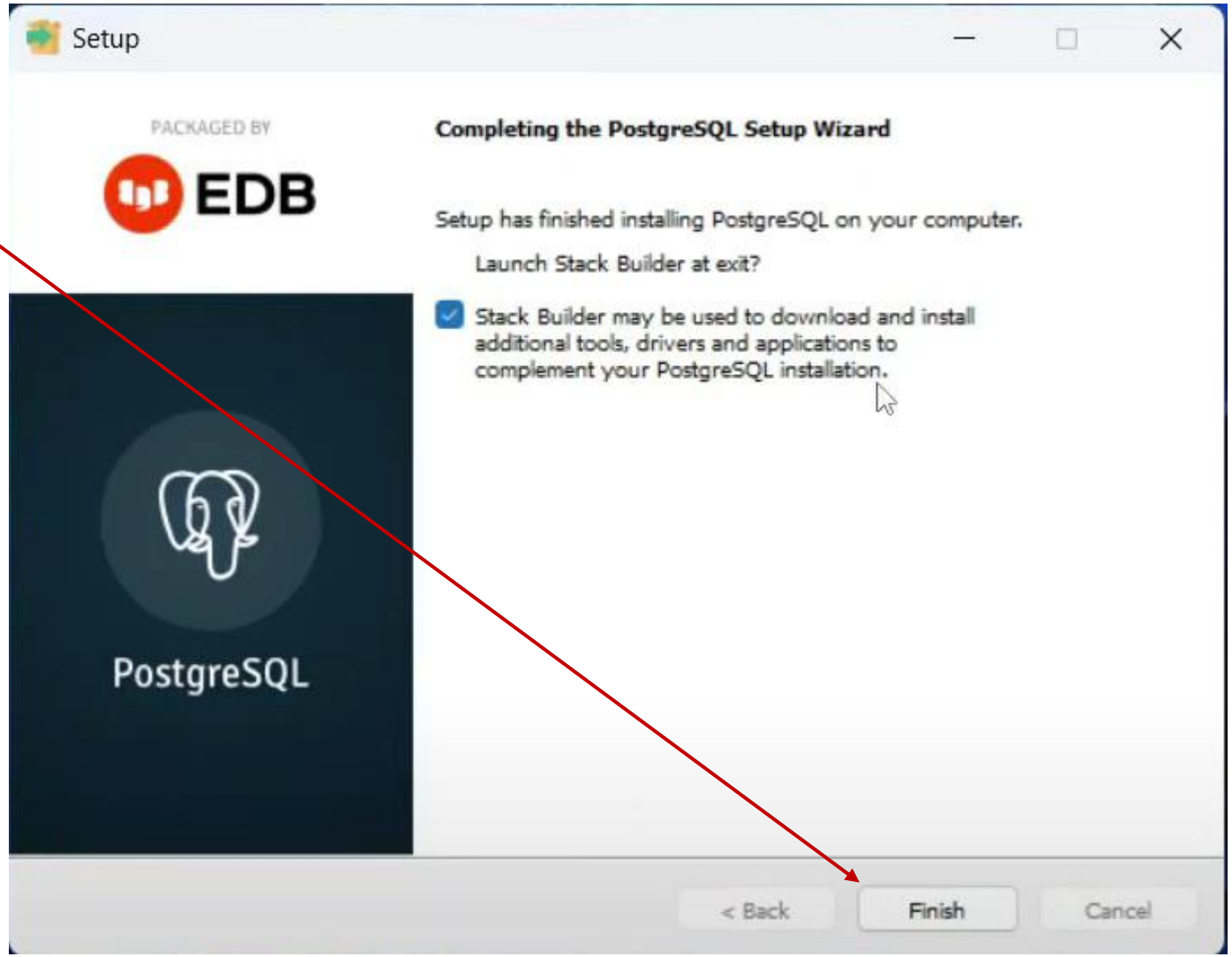
SQL File creation

Creating the SQL Database



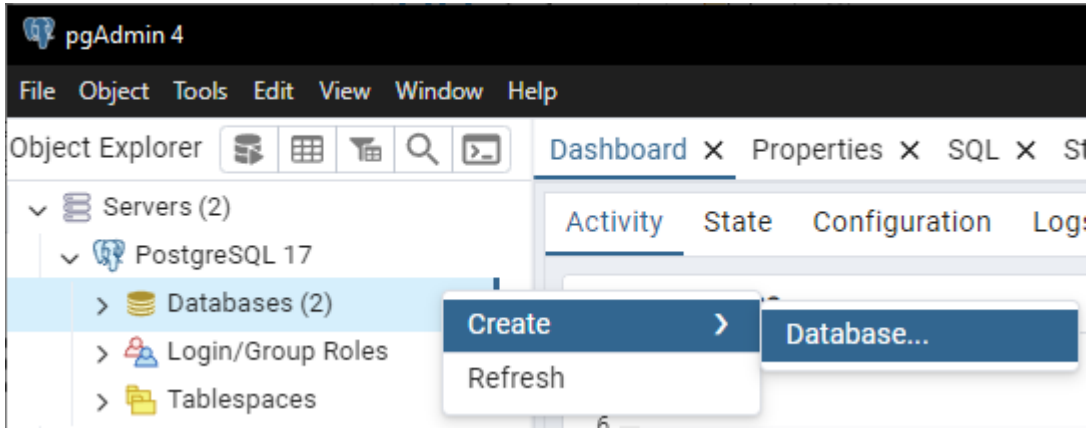
PostgreSQL.exe installation

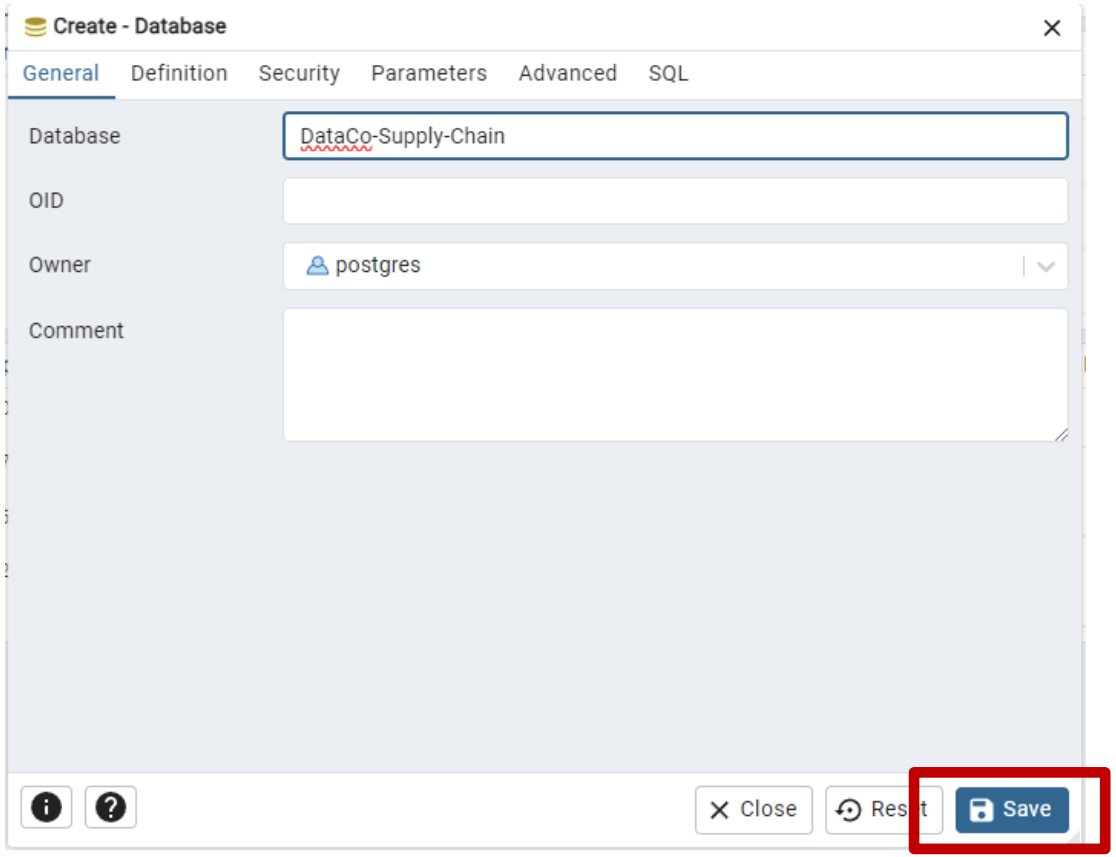
- Click on “Finish”



Creating the Database

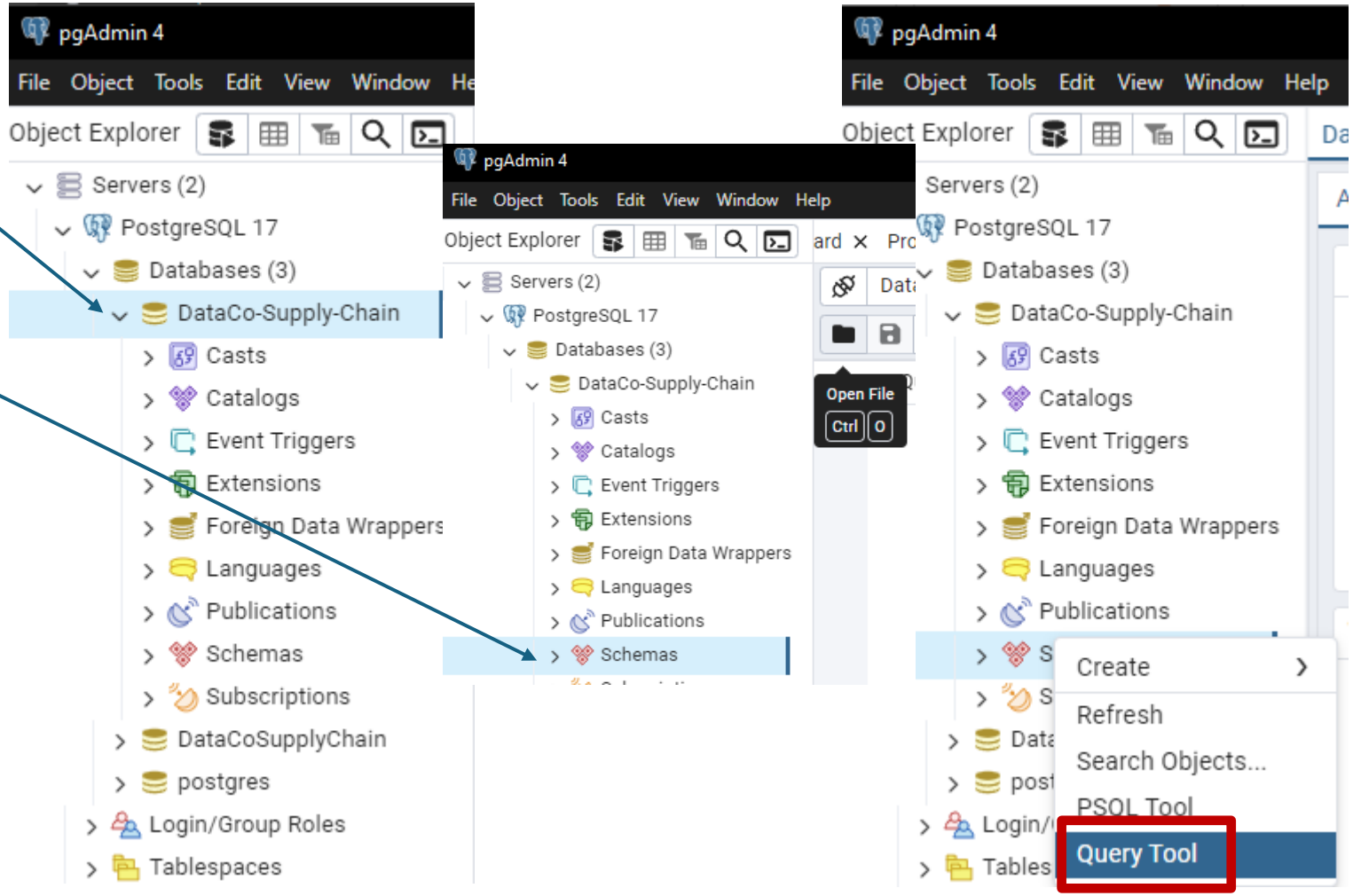
- Right click on “Databases”
- Click on “Create”
- Then click on “Database”
- A Create-Database window will appear
- Type in “DataCo-Supply-Chain”
- Click on “Save”





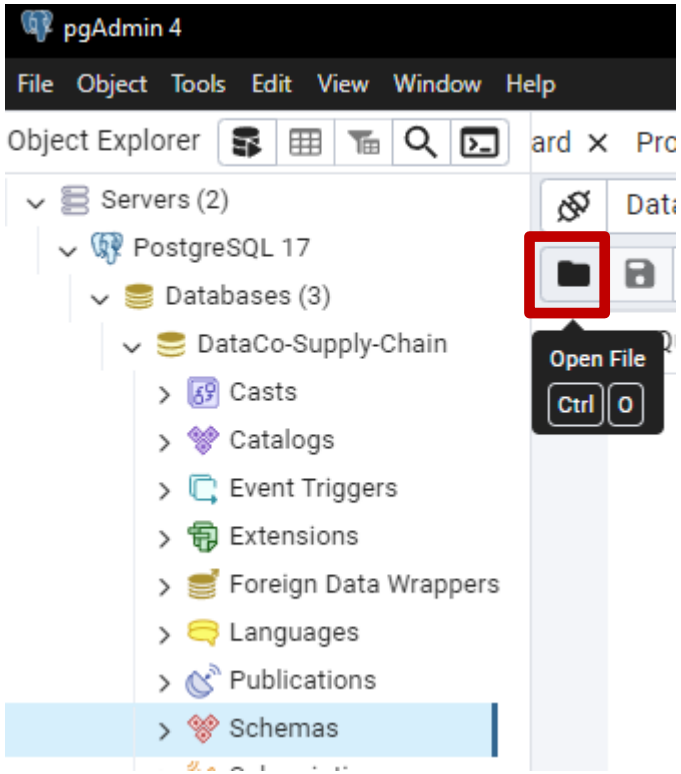
Creating the Dimension and Fact Tables

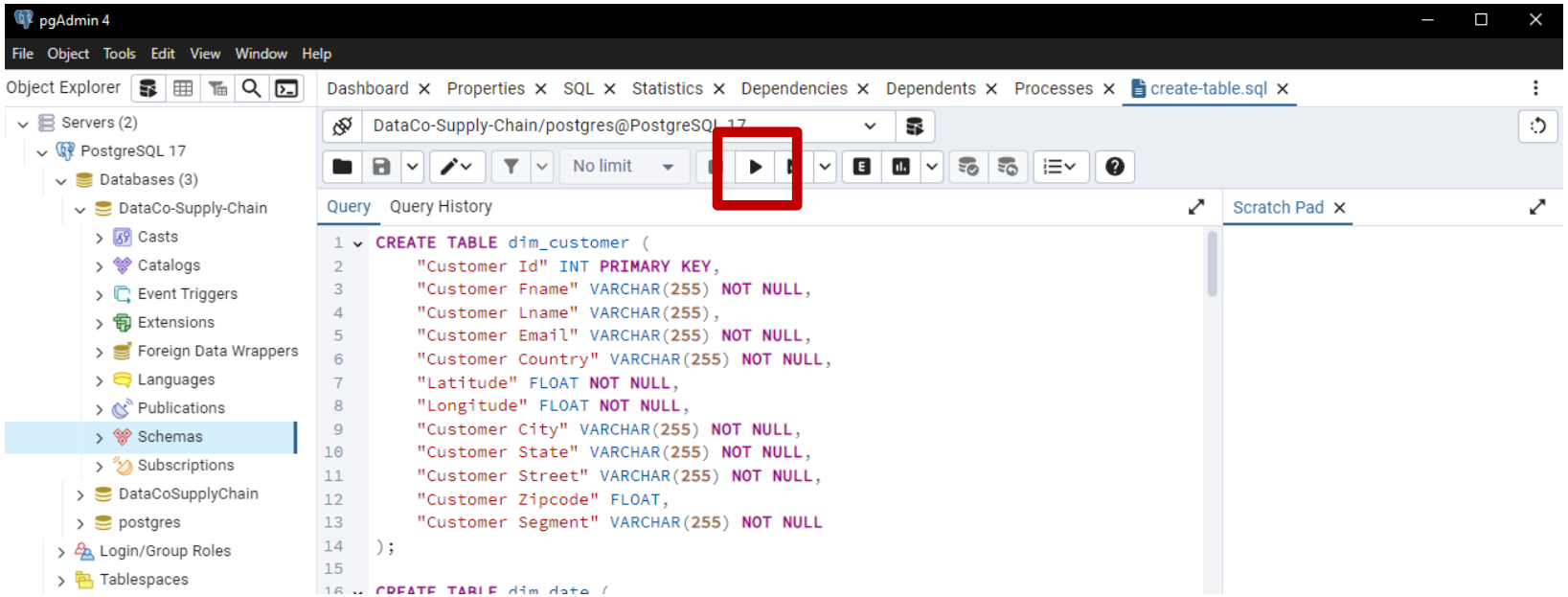
- Click on DataCo-Supply-Chain arrow to expand the list
- Right click on Schemas and then click on Query Tool



Creating the Dimension and Fact tables

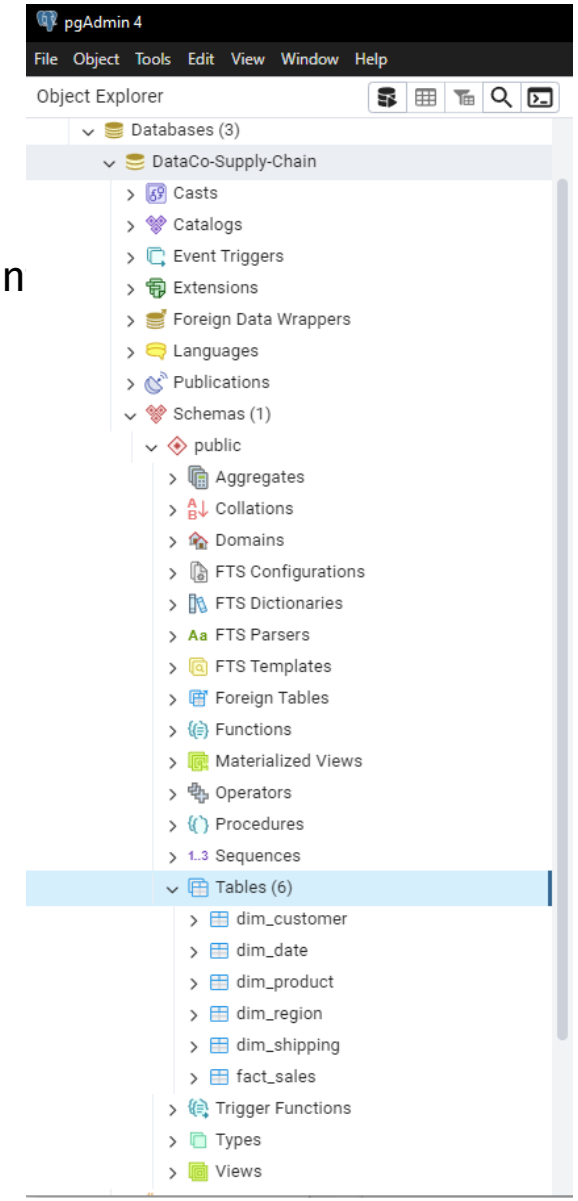
- Click on Open File
- Navigate to DataCo-Supply-Chain/Scripts/Sql/Create_Table.sql file
- Click on the play button to execute the script





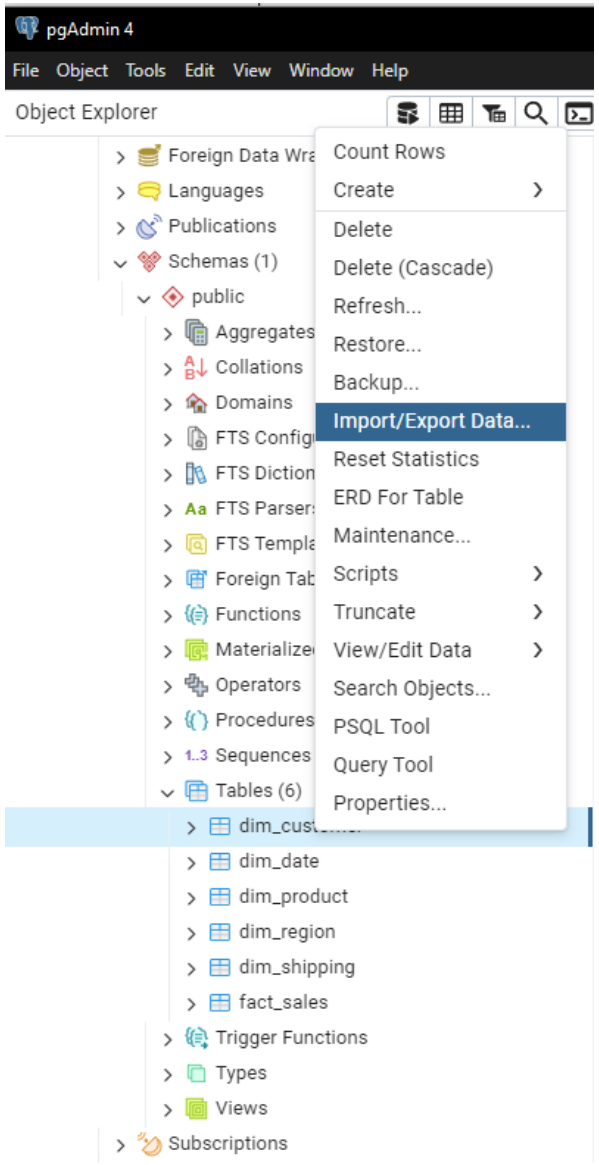
Creating the Dimension and Fact tables

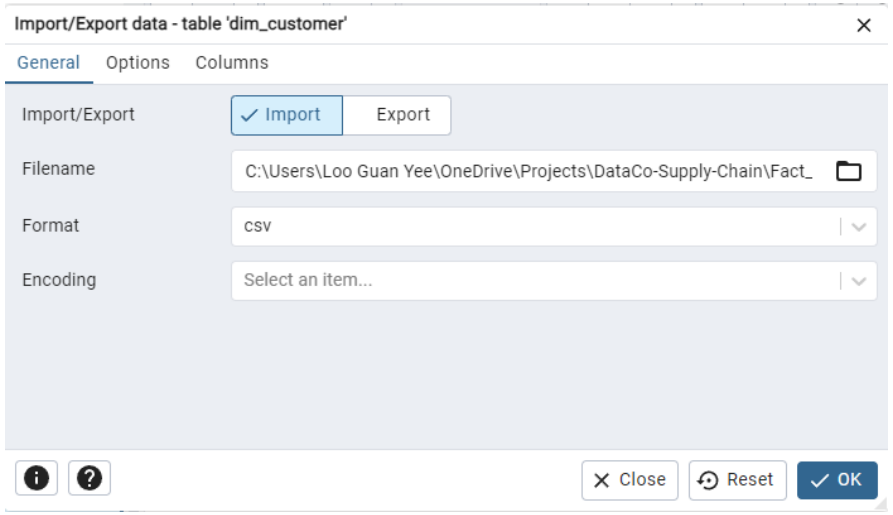
- Click on Schemas arrow
- Expand the Tables (6) button
    - You will see 6 dimension and fact tables presen



Loading the Dimension and Fact data into the tables

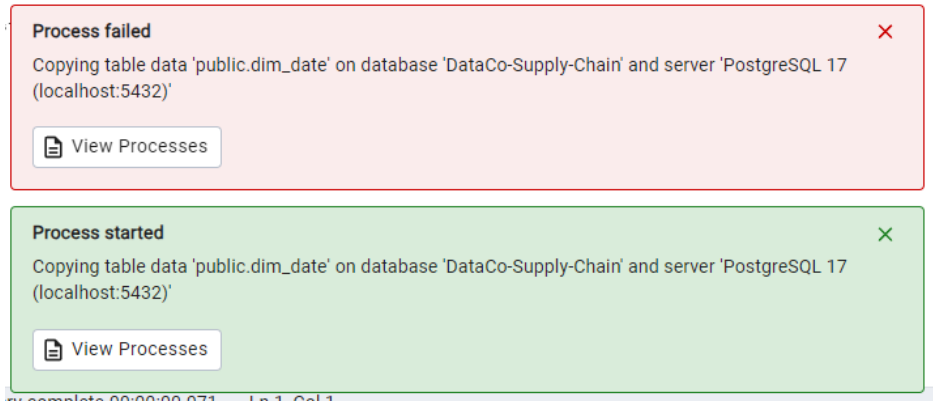
- Right click on each dim and Fact table
- Right click on Import/Export Data
- Type in the following file name for Dimension/Fact tables
  - Dim_customer
  - Yourpath\DataCo-Supply-Chain\data\processed\dim_customer.csv
  - Dim_date
  - Yourpath\DataCo-Supply-Chain\data\processed\dim_date.csv
  - Dim_product
  - 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

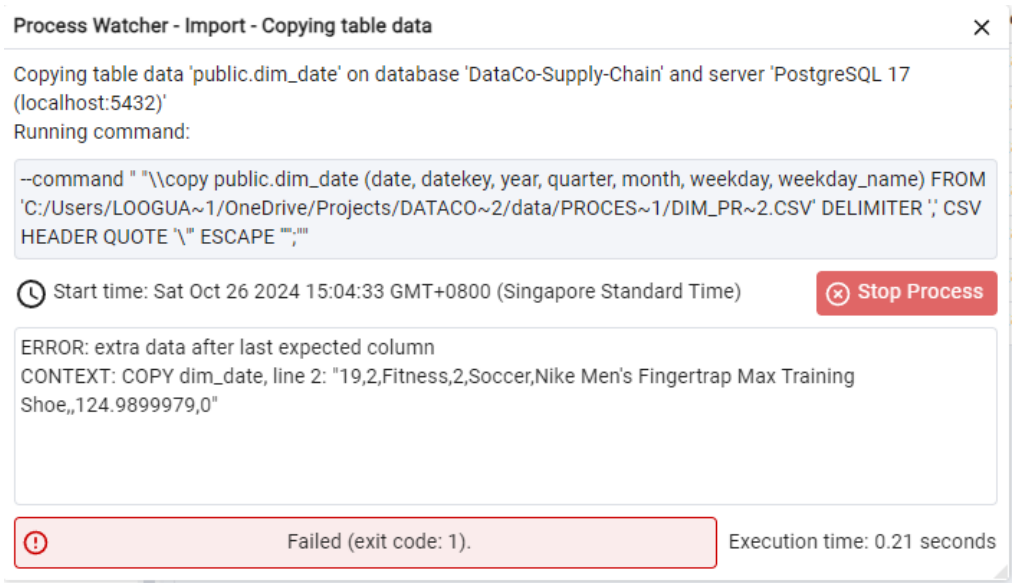




Loading the Dimension and Fact data into the tables

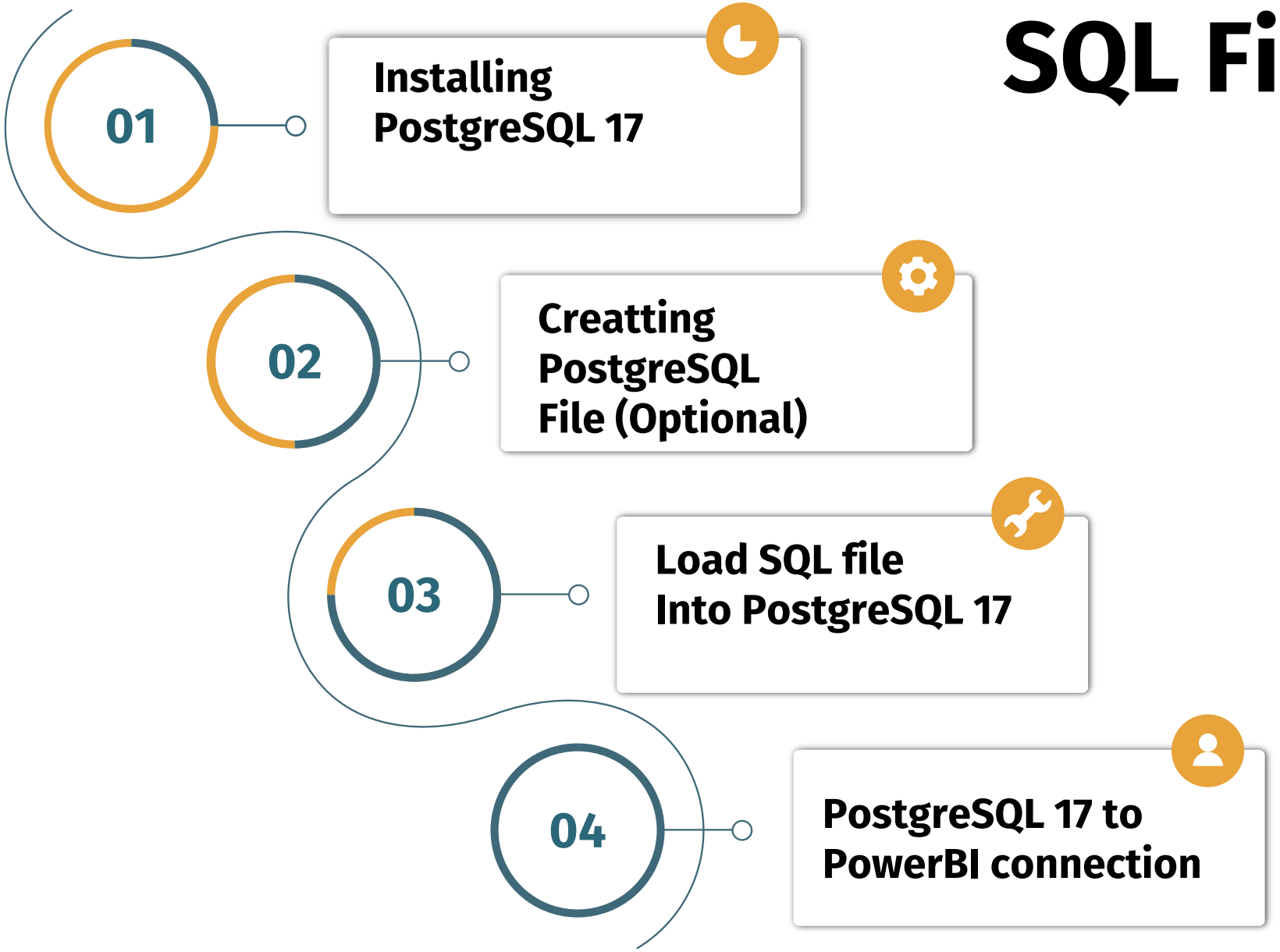
- You should be able see double green status with the process completed for the 6 tables
- If red status, check if the loaded data columns match with the dimension table columns





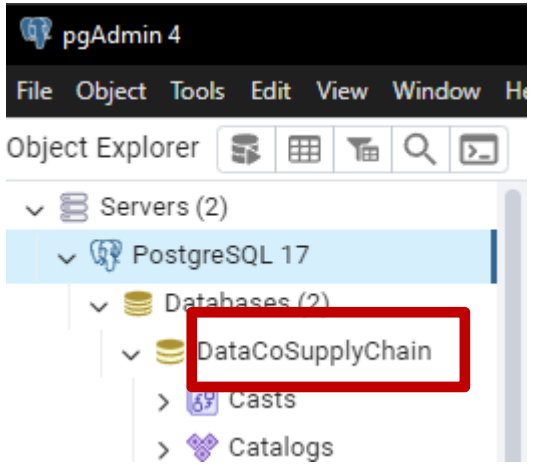
SQL File creation

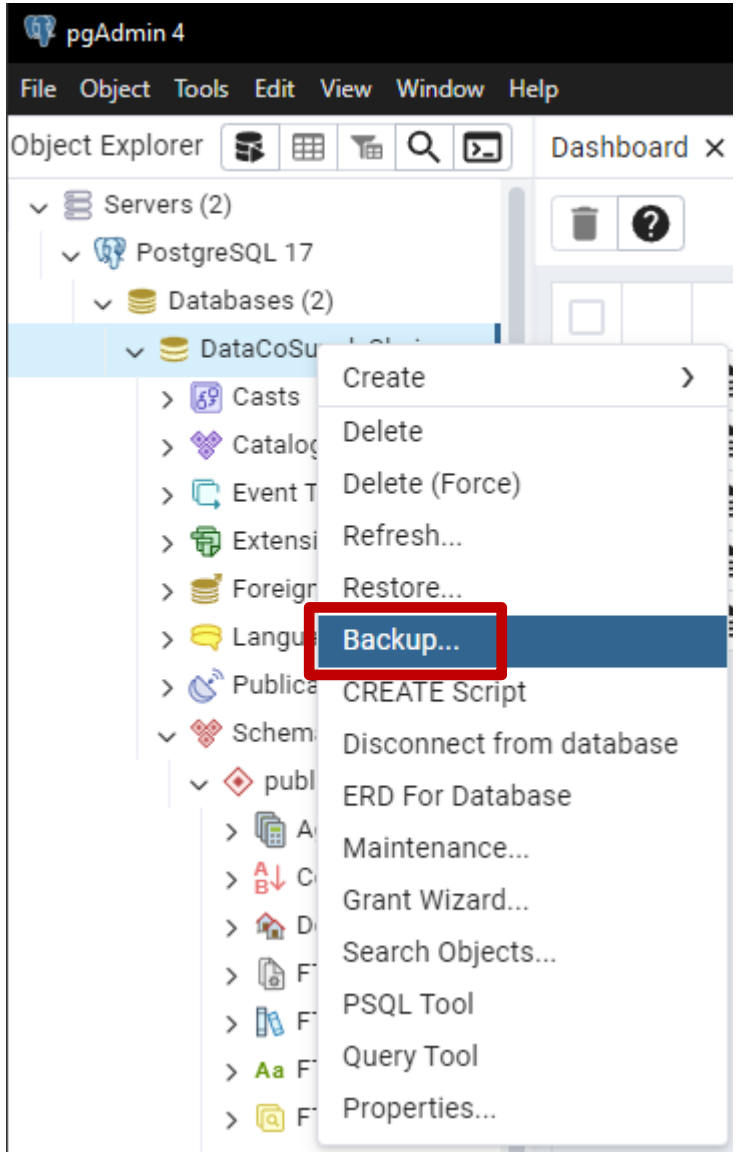
Backing up the SQL database



Backing up the SQL folder

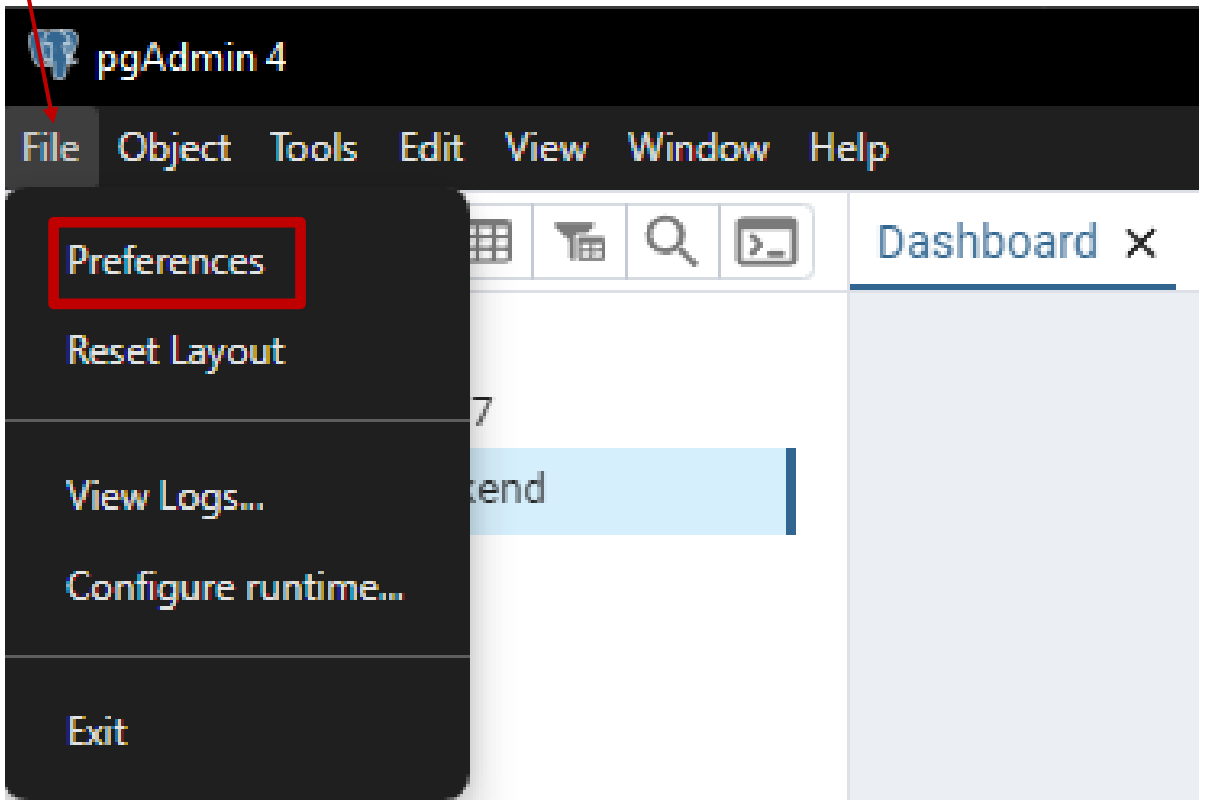
- Click on “DataCoSupplyChain” database
- Click on the backup button





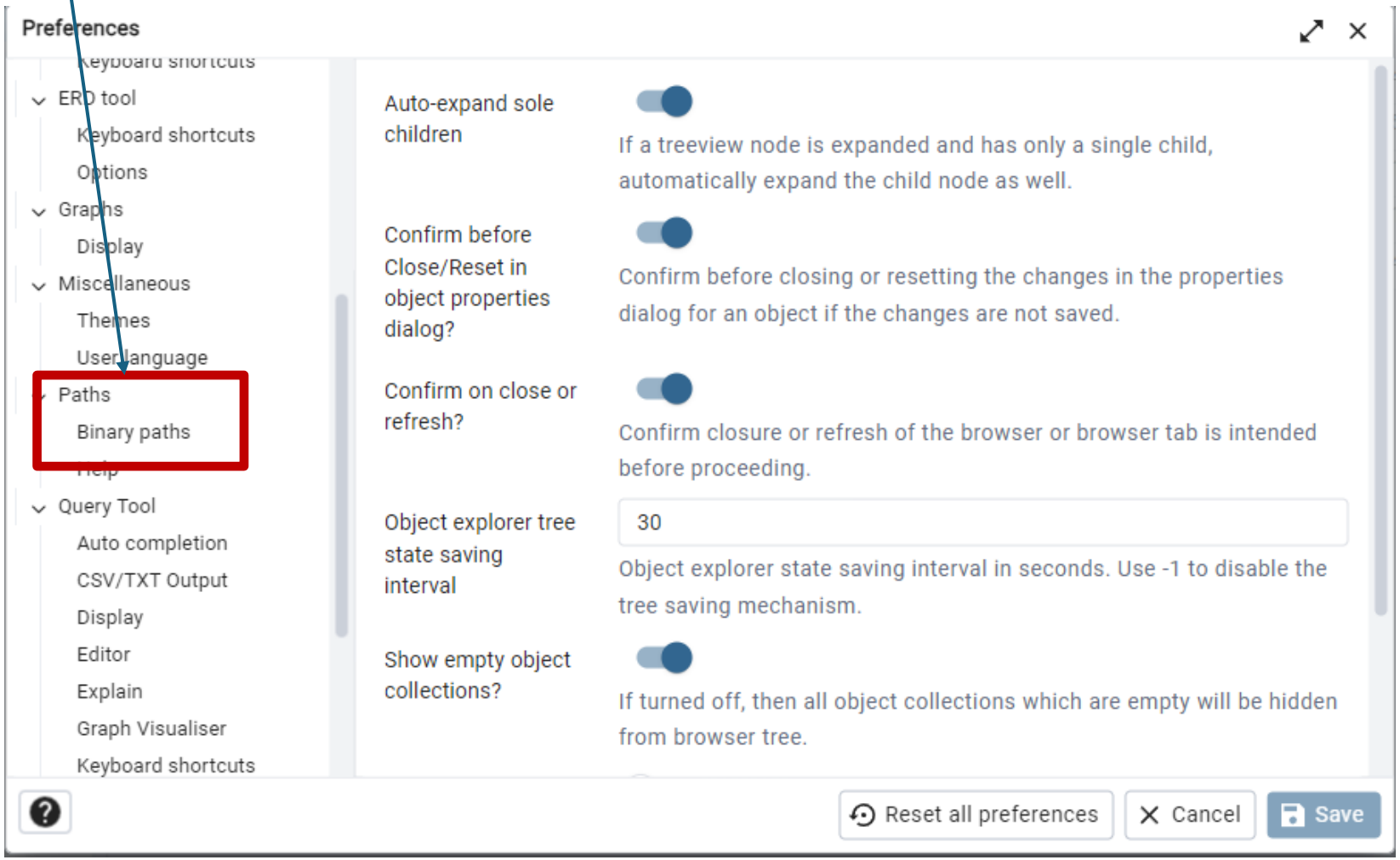
Open Pg admin4

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



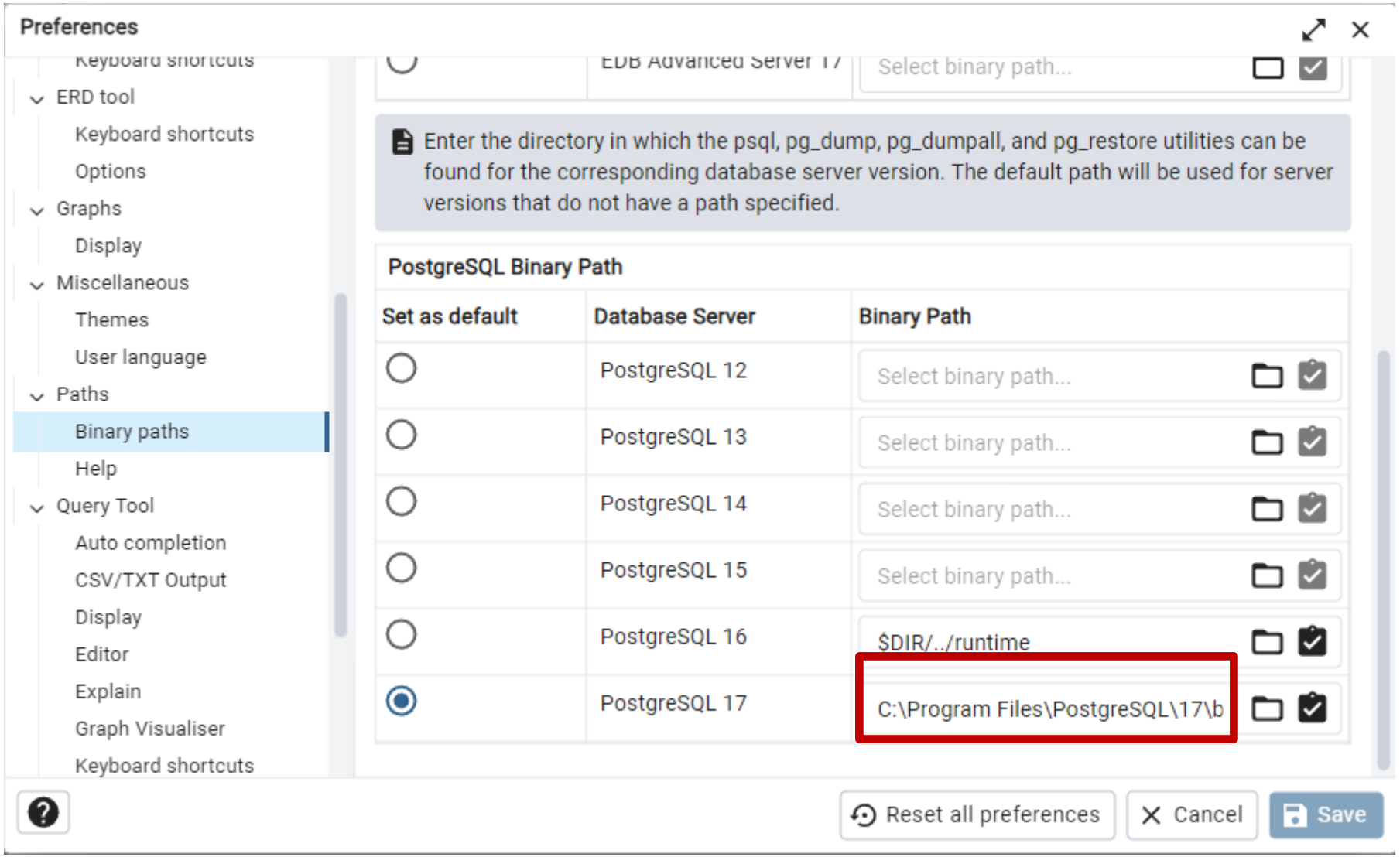
Setting PostgreSQL 17 as the default binary path

- Click on “Binary path”



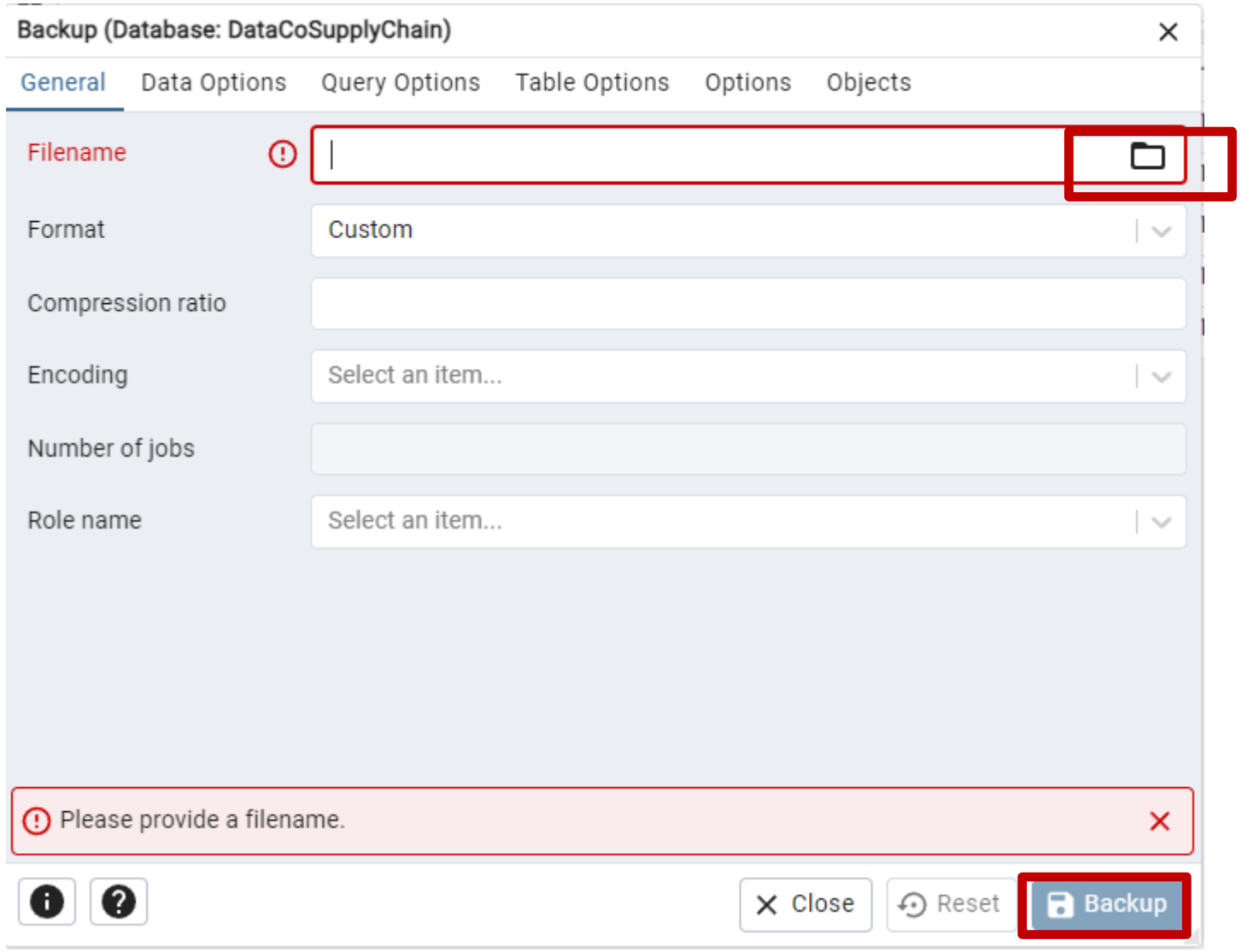
Setting PostgreSQL 17 as the default binary path

- 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

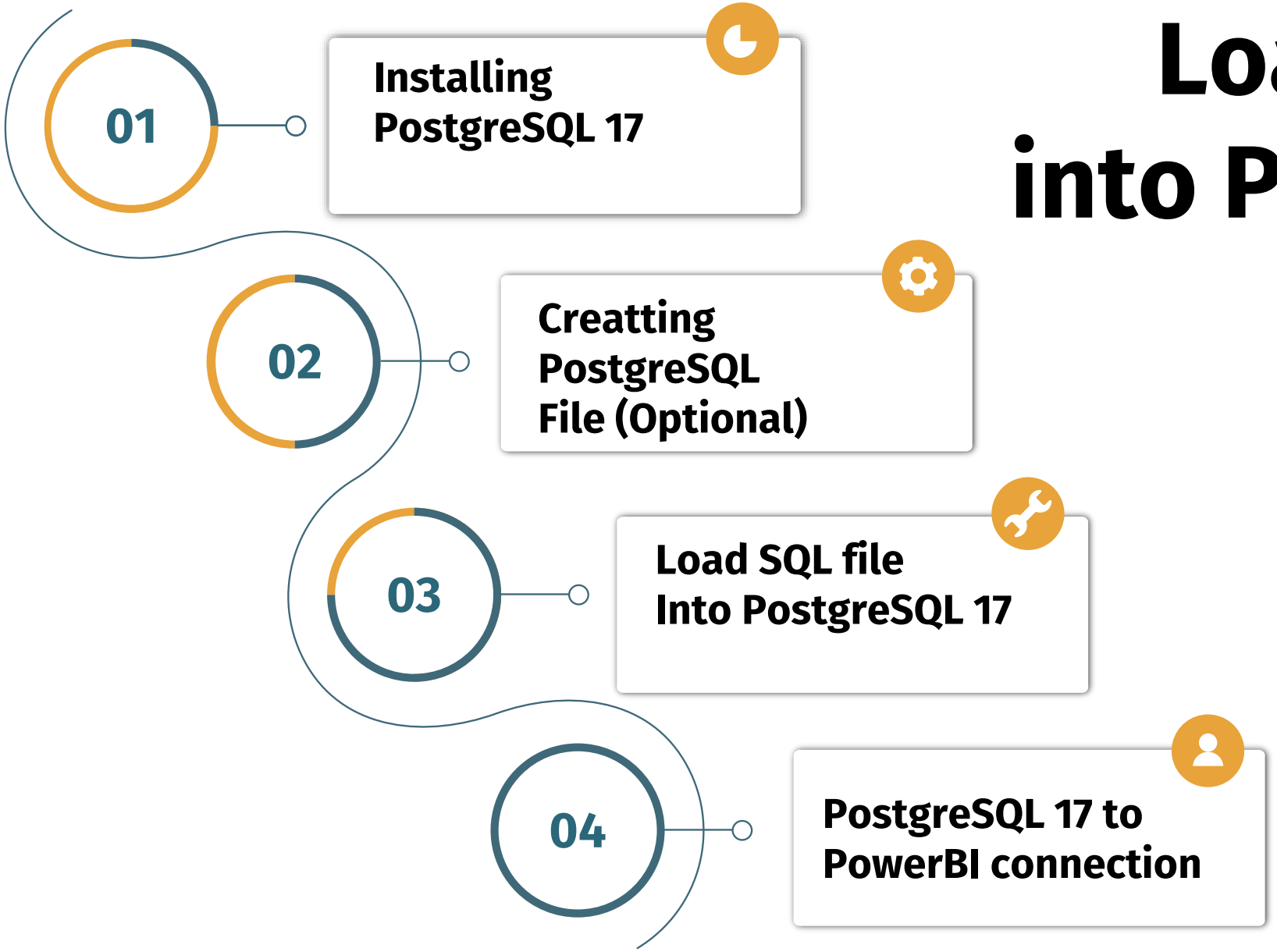


Backing up the SQL folder

- 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

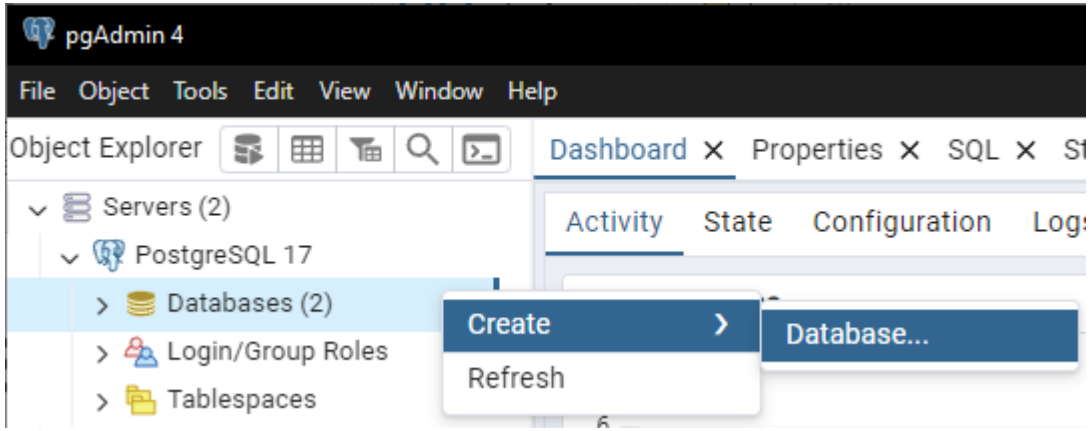


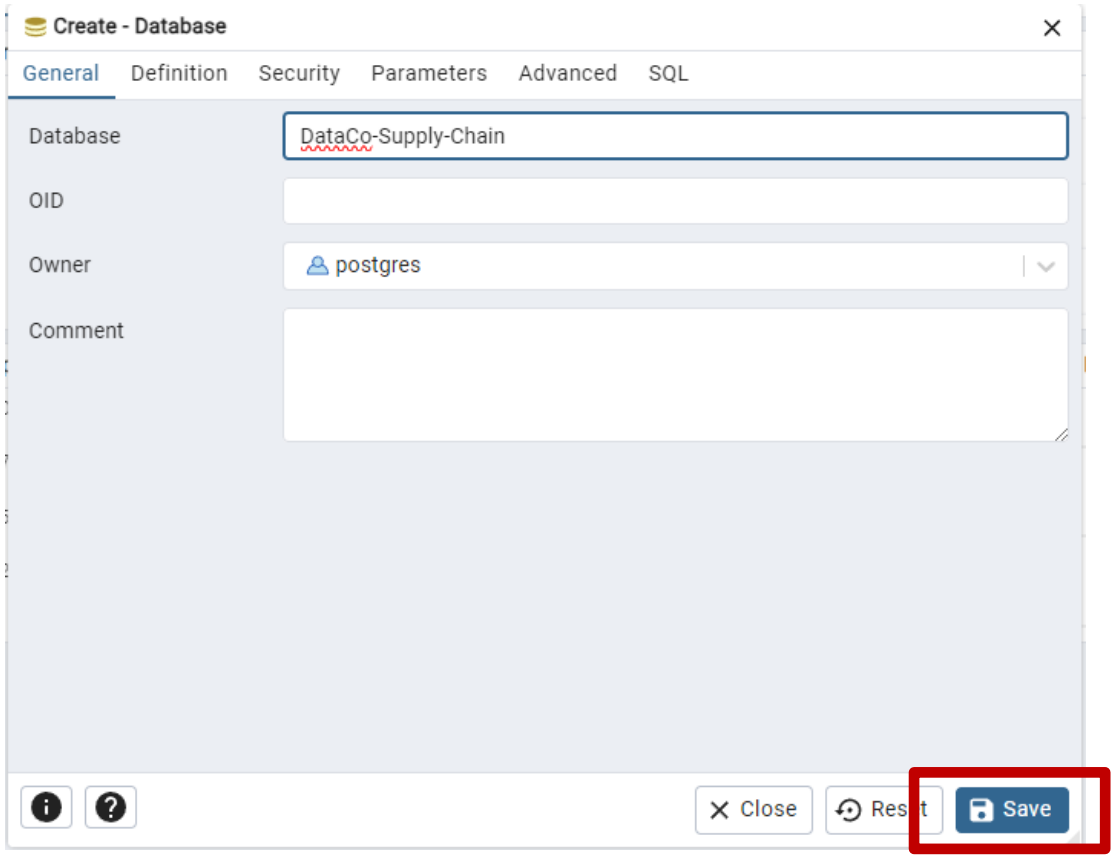
Load SQL file into PostgreSQL



Creating the Database

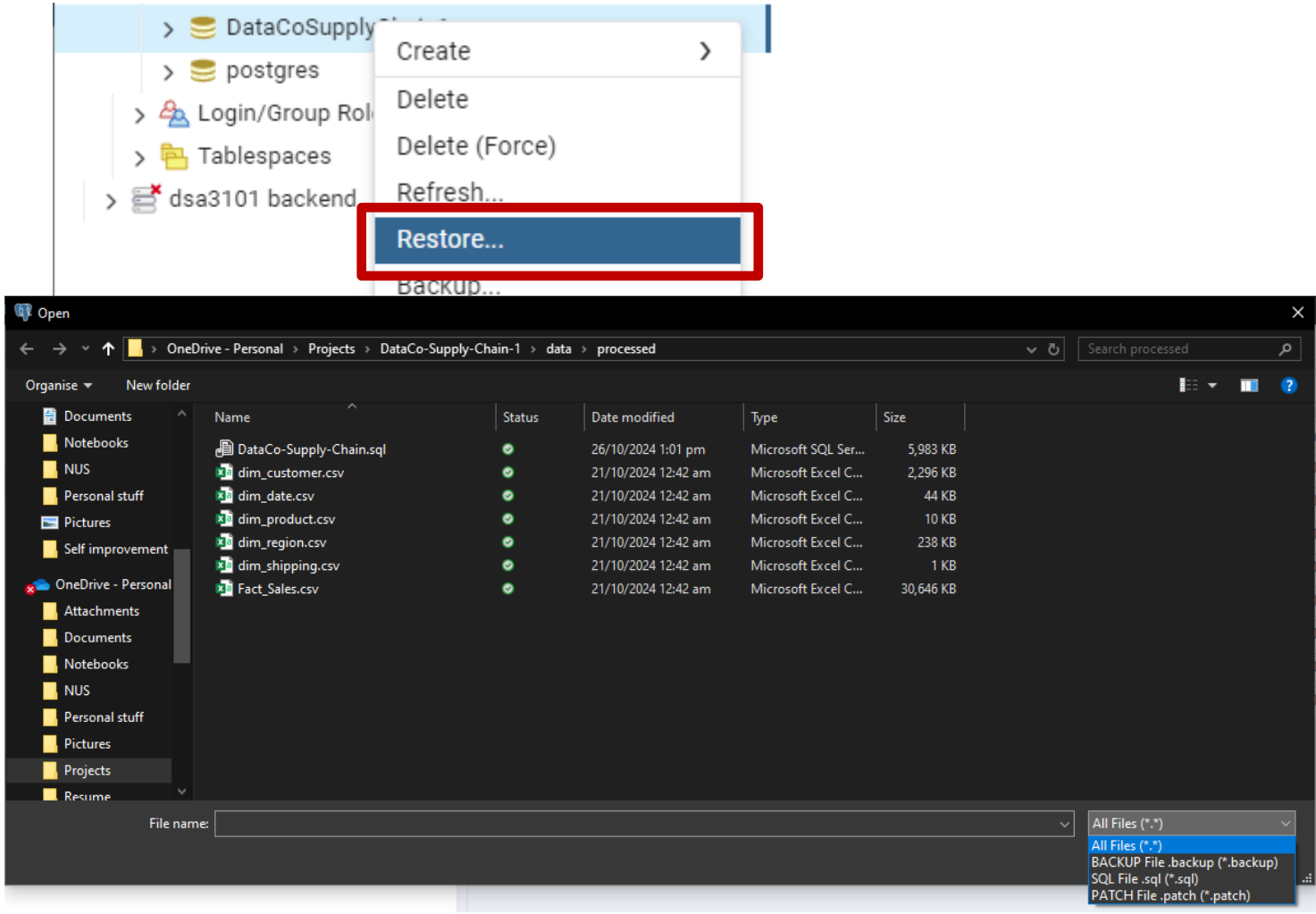
- *If you completed the steps in slide 22, skip this step.*
- Right click on “Databases”
- Click on “Create”
- Then click on “Database”
- A Create-Database window will appear
- Type in “DataCo-Supply-Chain”
- Click on “Save”





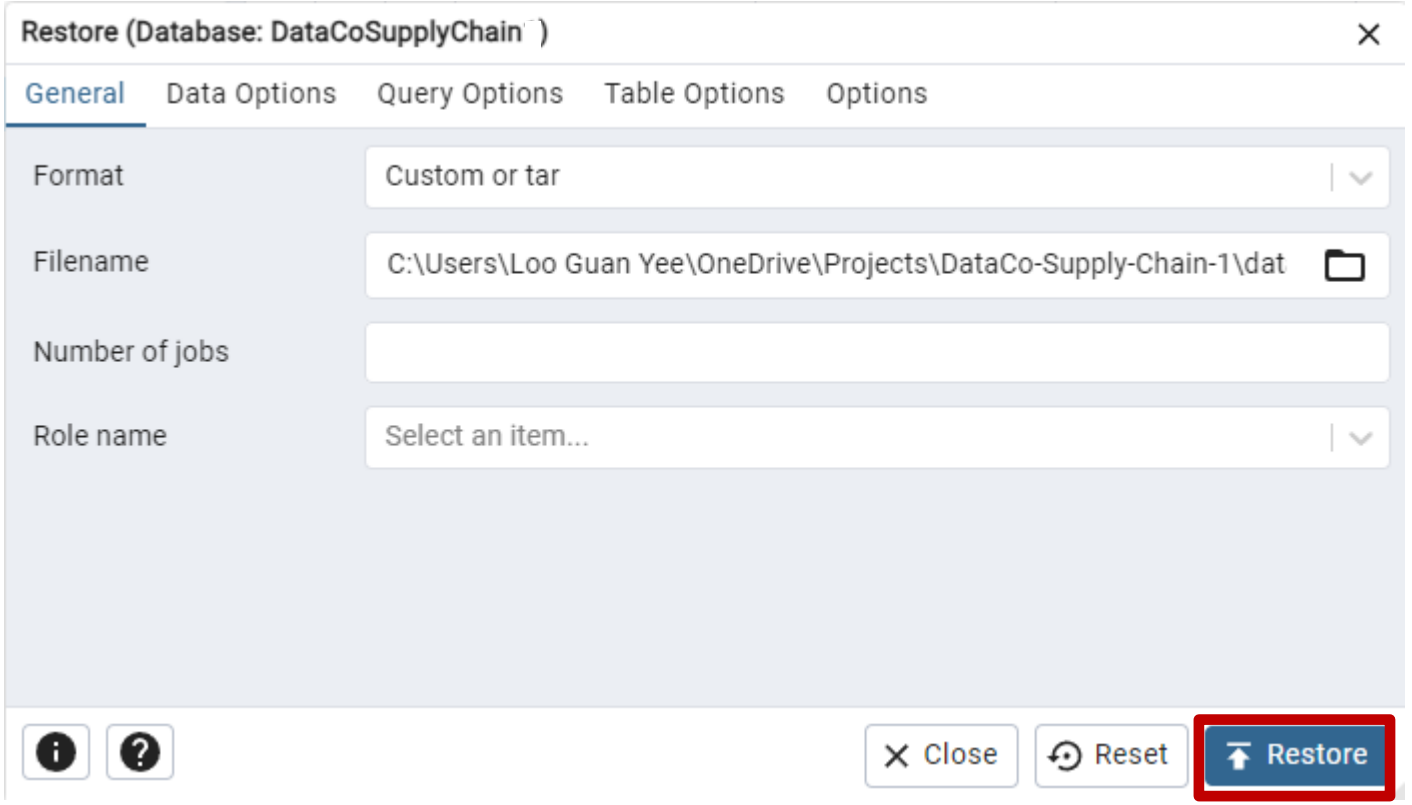
Restoring the database

- 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

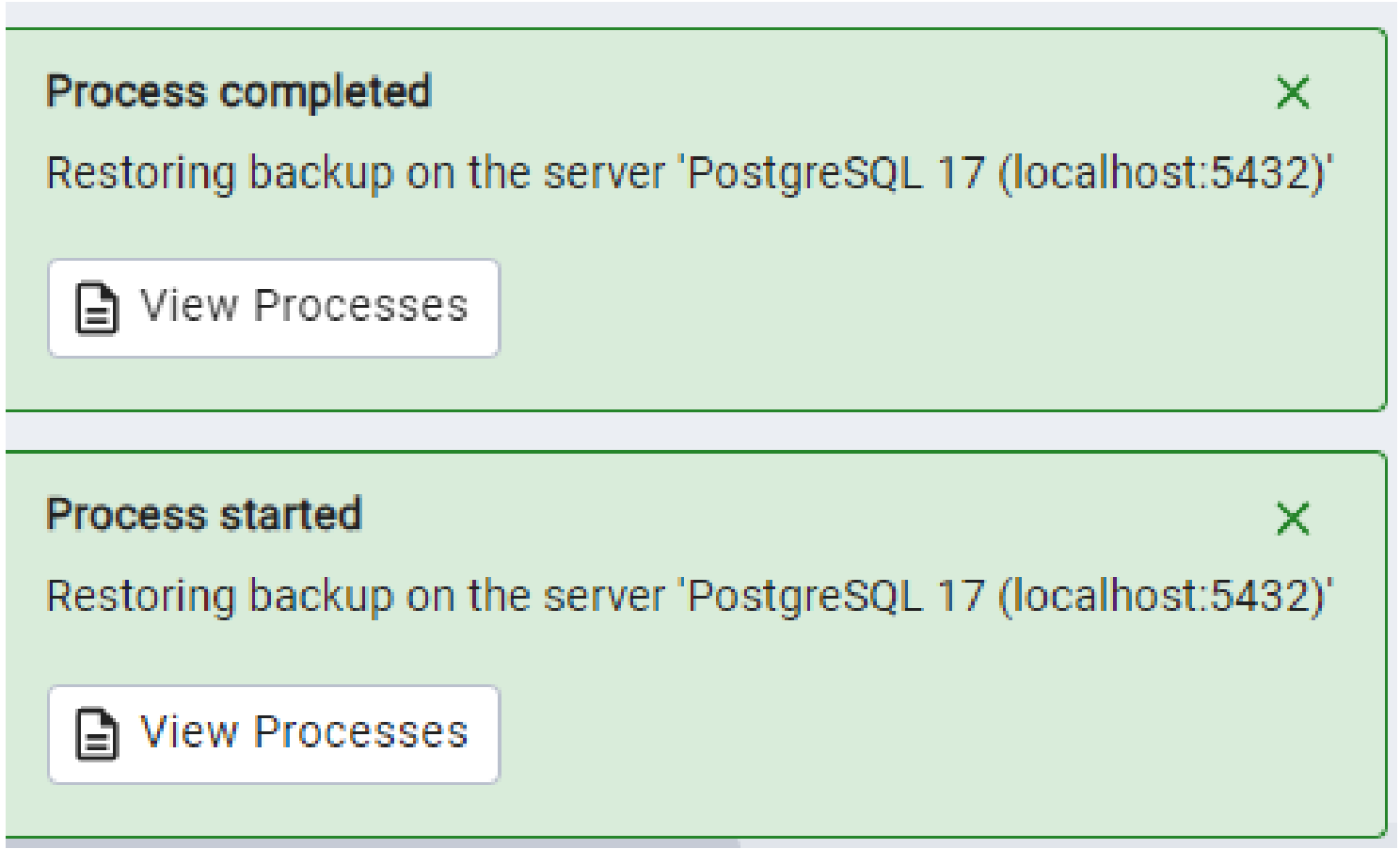


Restoring the database

- Click on ”Restore”

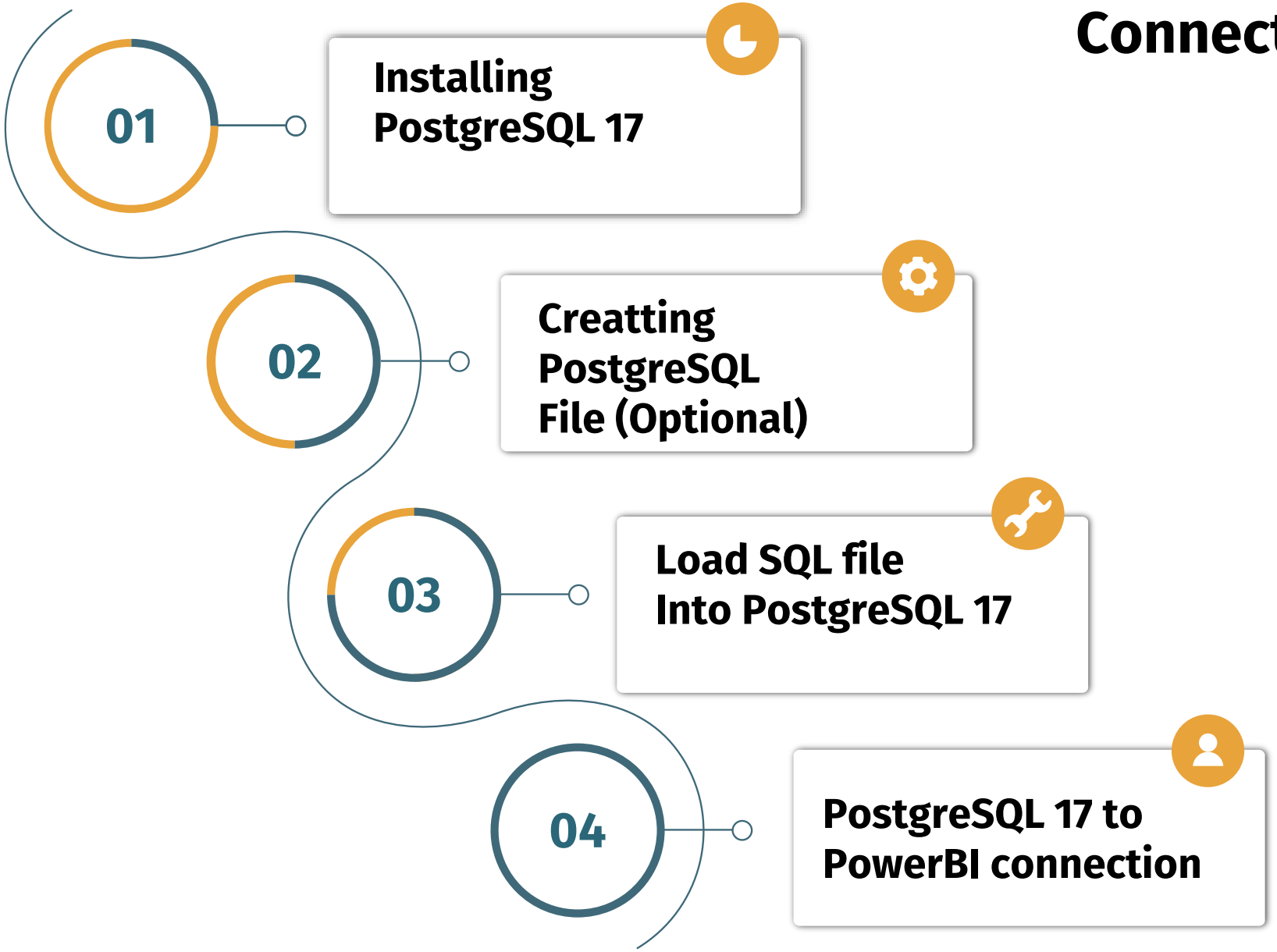


Check if you have double green status upon restoring the database



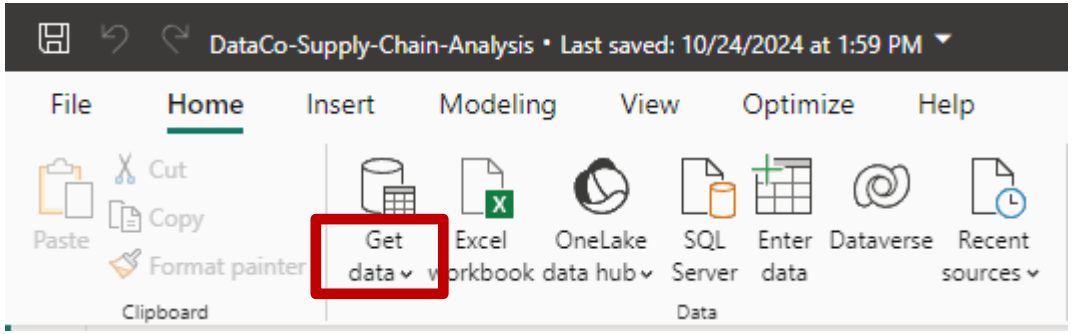
Connecting PostgreSQL to PowerBI

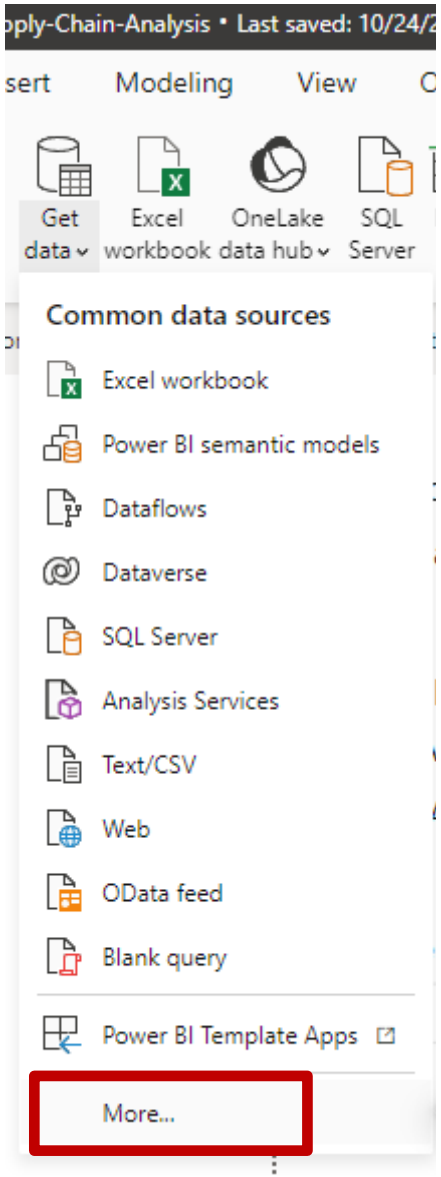
DataCo Eccomerce
Platfrom Sales



Open the Power Bi file from PowerBI folder

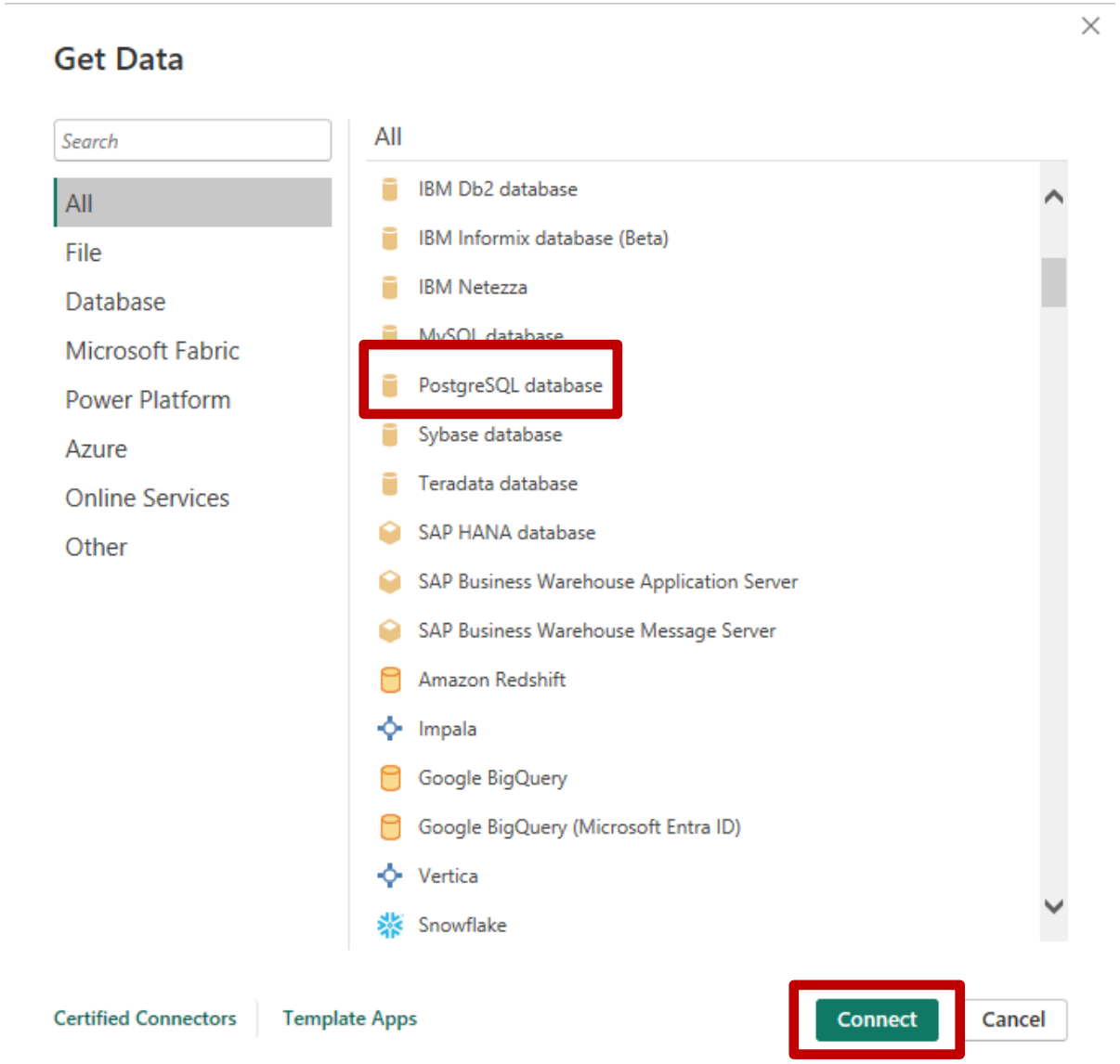
- Open the DataCo-Supply-Chain.pbix file from the PowerBI folder
- Click on Get data button
- Click on more button





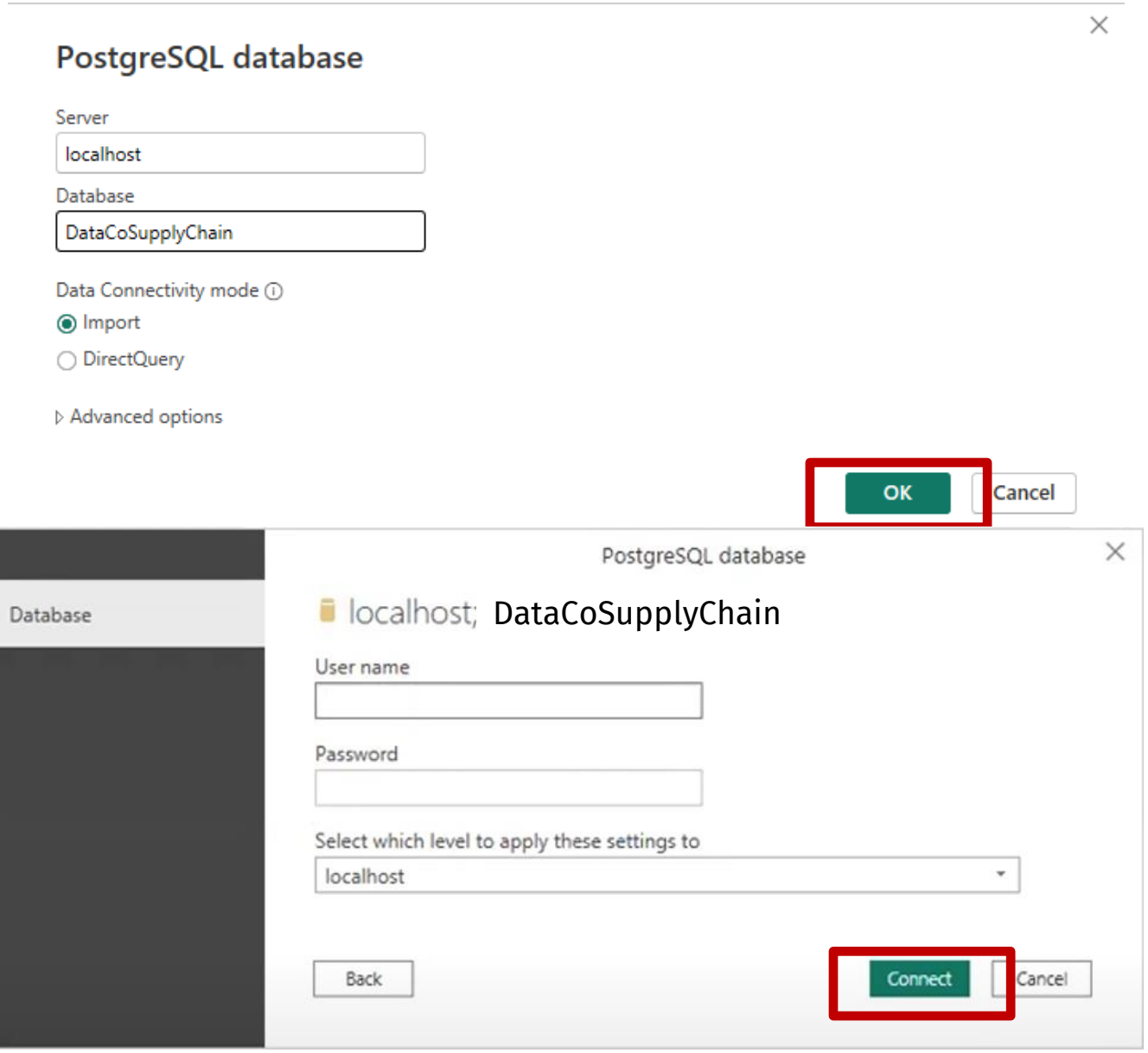
Connecting PostgreSQL to PowerBI

- 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



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