

Setup of PostgreSQL, pgAdmin and importing data

CS3200 Database design (fa18 s2)

<https://northeastern-datalab.github.io/cs3200/>

Version 9/6/2018

Overview

This document covers 2 issues:

1) How to install PostgreSQL:

- PostgreSQL is a popular open source database server. Unlike SQLite, PostgreSQL is a much more feature rich database management system. With PostgreSQL, you have 2 components, the server and the client. This isn't very different from your web server-browser model where the browser is your client: The web server services requests for fetching web pages whereas a database server services SQL queries on a database.
- This document will guide you through the process of setting up PostgreSQL on your machine. What this means is you will have a locally running instance of the PostgreSQL server on your machine.

Overview

- Just as a web browser helps make requests to a web server and displays the results of the request viz. a web page, similarly a database client helps you fire queries at a database server (PostgreSQL in our case), and displays the results that the database server sends over from processing those queries.
- The two most common clients that you will come across when using PostgreSQL are "psql" which is a command-line client and "pgAdmin" which is a graphical client.

2) How to import the IMDB data:

- Download the big ZIP file from our online directory. It contains a large collection of data from the IMDB movie website. This will create a folder named imdb2015, containing 6 .txt files. You will use them to create a database.

1. Setup PostgreSQL
(for MAC)

PostgreSQL



Postgres.app

The easiest way to get started with PostgreSQL on the Mac

Download

Documentation

Github

← 3905 Stars!

1) Click Here to Download

Postgres.app is a full-featured PostgreSQL installation packaged as a standard Mac app. It includes everything you need to get started: we've even included popular extensions like [PostGIS](#) for geo data and [plv8](#) for JavaScript.

Postgres.app has a beautiful user interface and a convenient menu bar item. You never need to touch the command line to use it – but of course we do include all the necessary [command line tools](#) and header files for advanced users.

Postgres.app updates automatically, so you get bugfixes as soon as possible.

The current version requires macOS 10.10 or later and comes with the latest PostgreSQL versions (10.1, 9.6.6, and 9.5.10), but we also maintain [other versions](#) of Postgres.app.

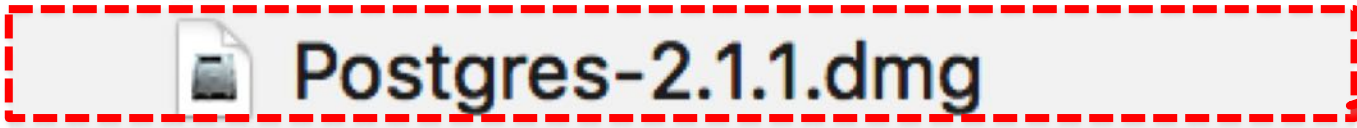
Installing Postgres.app



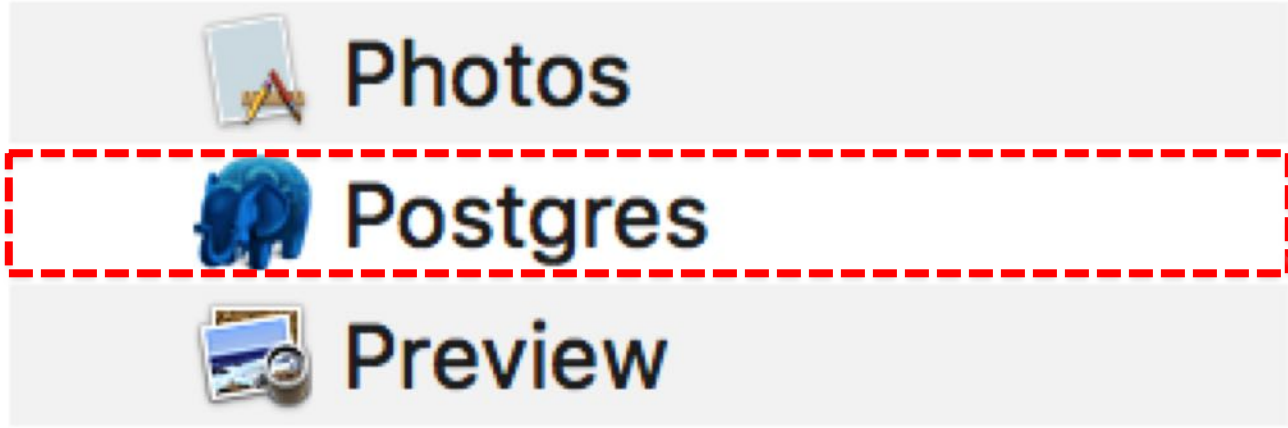
Download → Move to Applications folder → Double Click

If you don't move Postgres.app to the Applications folder, you will see a warning about an unidentified developer and won't be able to open it.

- Go to <http://postgresapp.com/>.
- Download the latest released version (not the prerelease!).



2) Extract the file you just downloaded. Typically the downloaded file should be in your downloaded folder.



3) The previous step will extract the "Postgres" application, typically in the same folder. Look for a file with a blue elephant icon.

4) Double click on this file and PostgreSQL server should be up and running.

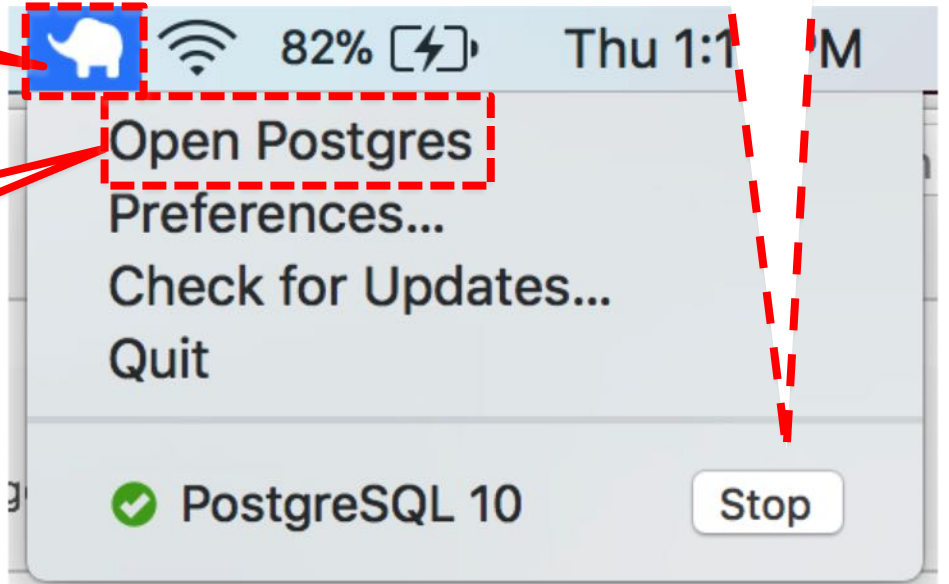


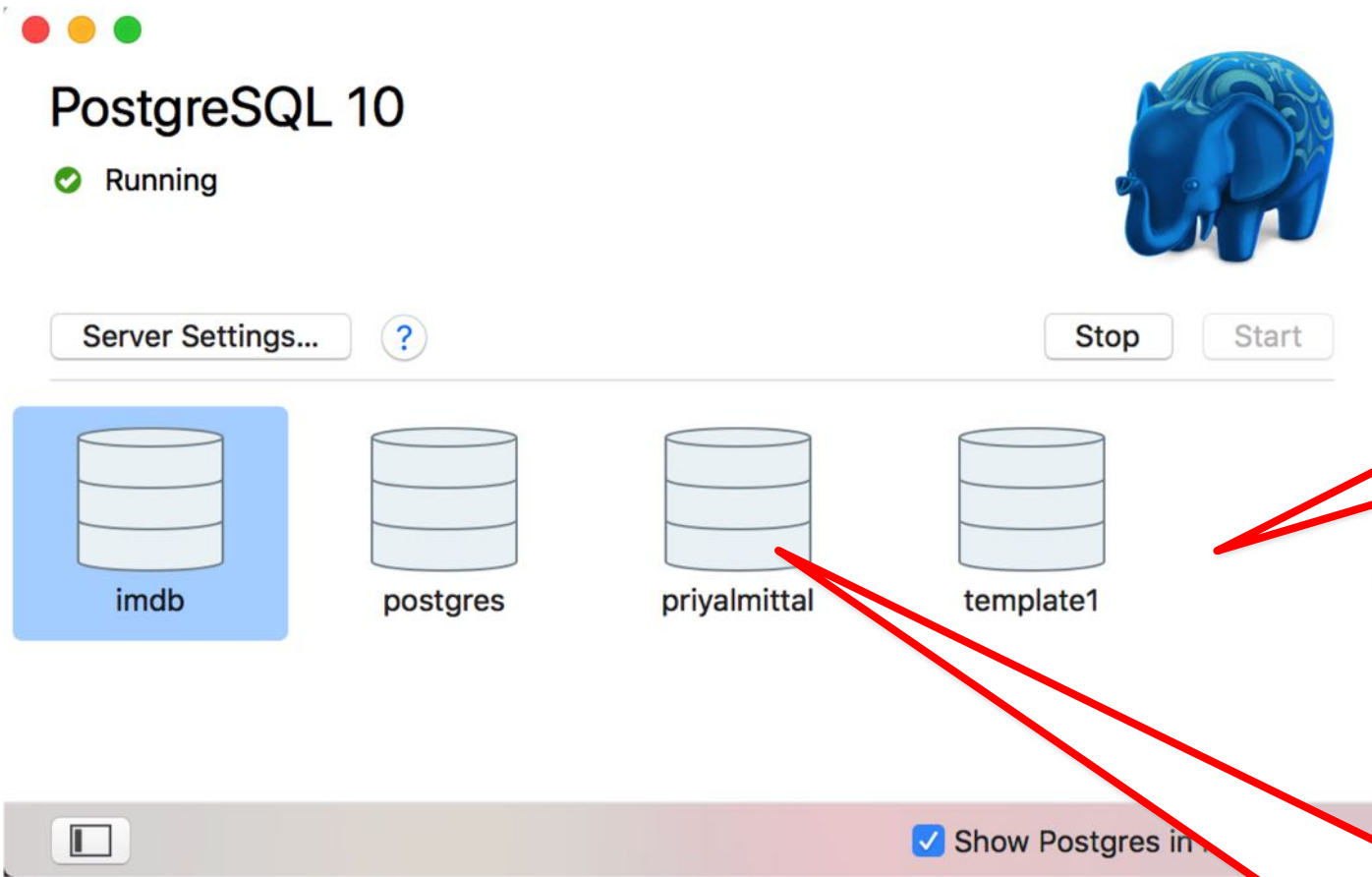
5) While the PostgreSQL server is up and running, you will see an icon show up in your menu bar at the top.

6) Click on the "Slonik" in the menu bar.

7) In the menu that shows up, select "Open Postgres".

If you restart your computer, the server may not be running. Thus verify that the server is running





9) After you “Open Postgres”, this window provides you with all your existing databases (in case you have already created some).

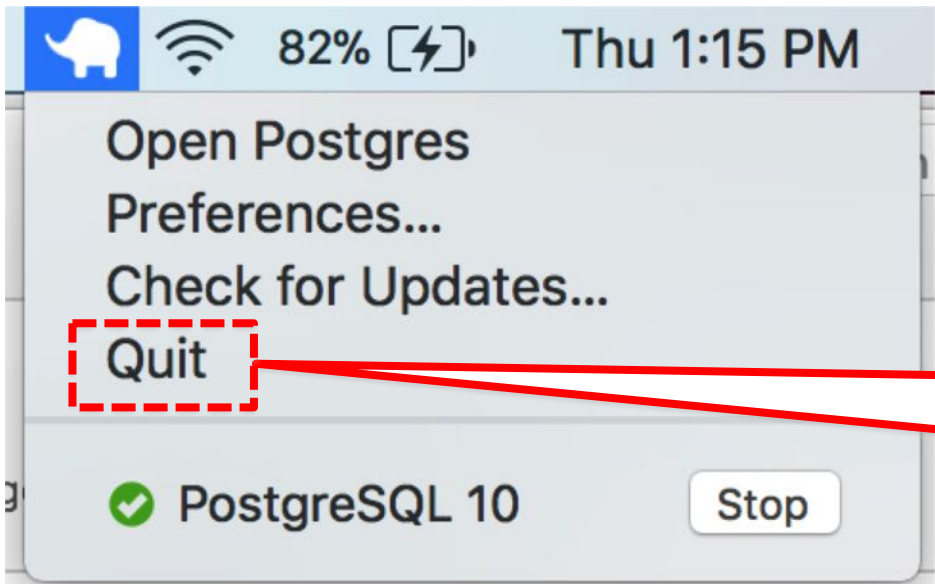
10) In case you have existing databases, then double-click on the database you want to work with.


```
Last login: Thu Jan 11 13:13:59 on ttys000
"/Applications/Postgres.app/Contents/Versions/10/bin/psql" -p5432 -d "priyalmittal"
Priyals-MacBook-Pro:~ priyalmittal$ "/Applications/Postgres.app/Contents/Versions/10/bin/psql" -p5432 -d "priyalmittal"
psql (10.1)
Type "help" for help.
```

```
priyalmittal=#
```

11) After clicking on the database you want to work with, the Postgres command prompt should open, that should look like <your_user_name>=#

12) This is the psql client that was mentioned earlier and where you will enter SQL commands to interact with the database server.



13) To stop the server, simply quit the application by clicking on the elephant icon in your menu bar and selecting Quit from the menu bar that shows up underneath the icon.

Command line Cheatsheet

Alternatively, start from terminal with command "psql"

```
gatter — psql — 80x24
Last login: Fri Feb  9 22:47:42 on ttys001
[Wolfgangs-MacBook-Pro-9:~ gatter] psql
psql (9.5.5, server 10.1)
WARNING: psql major version 9.5, server major version 10.
        Some psql features might not work.
Type "help" for help.

[gatter=# \c imdb
psql (9.5.5, server 10.1)
WARNING: psql major version 9.5, server major version 10.
        Some psql features might not work.
You are now connected to database "imdb" as user "gatter".
imdb=#
```

`\l` list existing databases
`\c` connect to a database
`\d` list tables in database
`\q` disconnect from psql

`\d <tablename>` view details of a table
create database <dbname> create DB

Problems

- One reason for starting problems with pgadmin can be that the user postgres does not have a password. In pgadmin, with the "New Server Registration" dialog form, it does not accept an empty password. In that case, perform the following steps :
 - Launch postgres command line interface using **sudo -u postgres psql**
 - Run the following command:
ALTER USER postgres WITH PASSWORD '<NewPostgresPassword>';
(FM: Remember to add the semi-colon)
 - Now setup a new server connection in pgadmin by clicking on **File -> New Server...**
 - Make sure the user is set to postgres and the password is set to <NewPostgresPassword>

2. Setup PostgreSQL (for Windows)

- Go to <http://www.postgresql.org/download/windows/>.

PostgreSQL The world's most advanced open source database.

Home About Download Documentation Community Developers Support Your account

» Downloads
Binary
Source
» Software Catalogue
» File Browser

Windows installers

Interactive installer by EnterpriseDB

[Download the installer](#) provided by EnterpriseDB for all supported PostgreSQL versions.

This installer includes the PostgreSQL server, pgAdmin; a graphical tool for managing and developing you databases, and StackBuilder; a package manager that can be used to download and install additional PostgreSQL tools and drivers. Stackbuilder includes management, integration, migration, replication, geospatial, connectors and other tools.

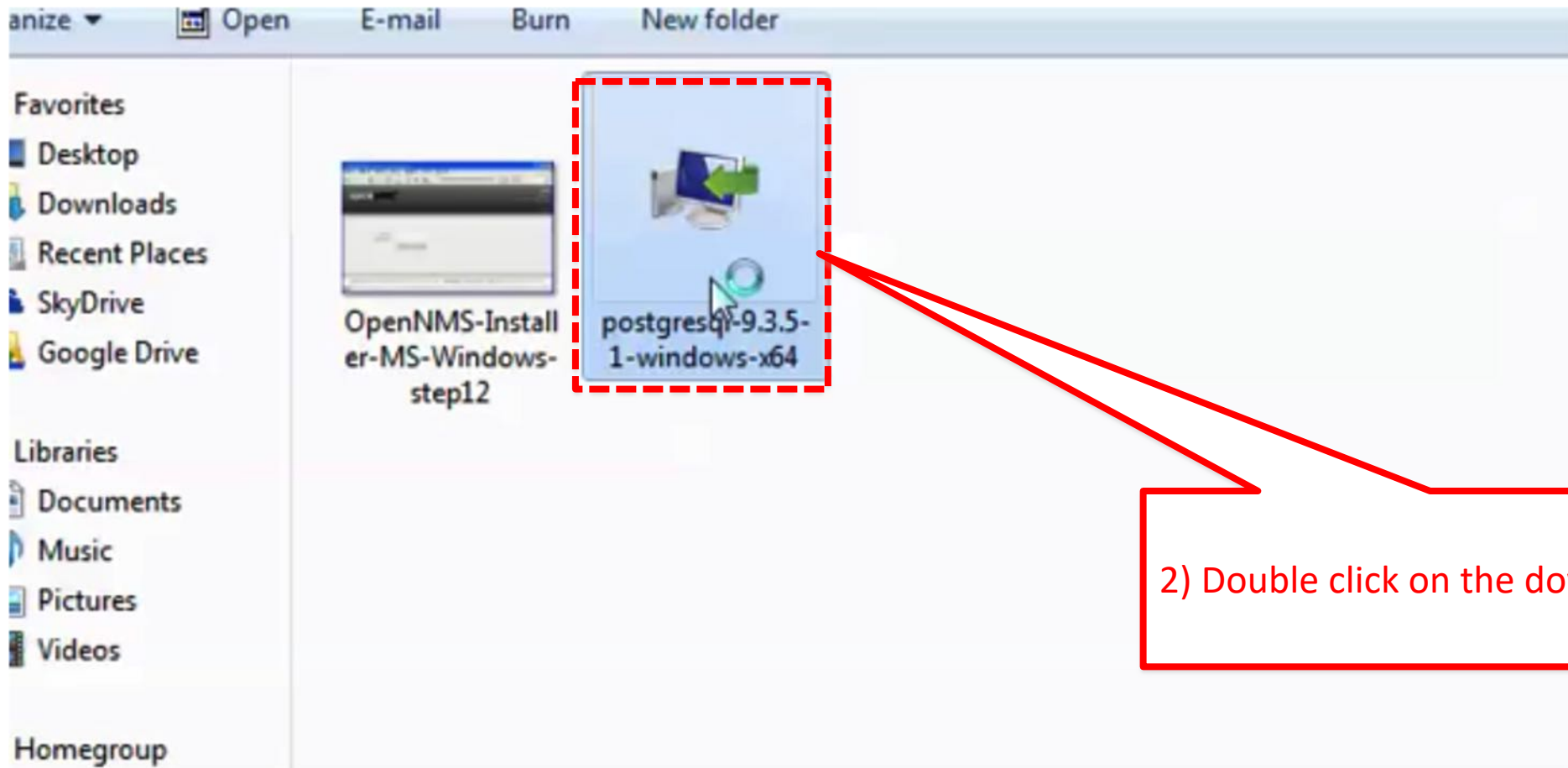
This installer can run in graphical or silent install modes.

The installer is designed to be a straightforward, fast way to get up and running with PostgreSQL on Windows.

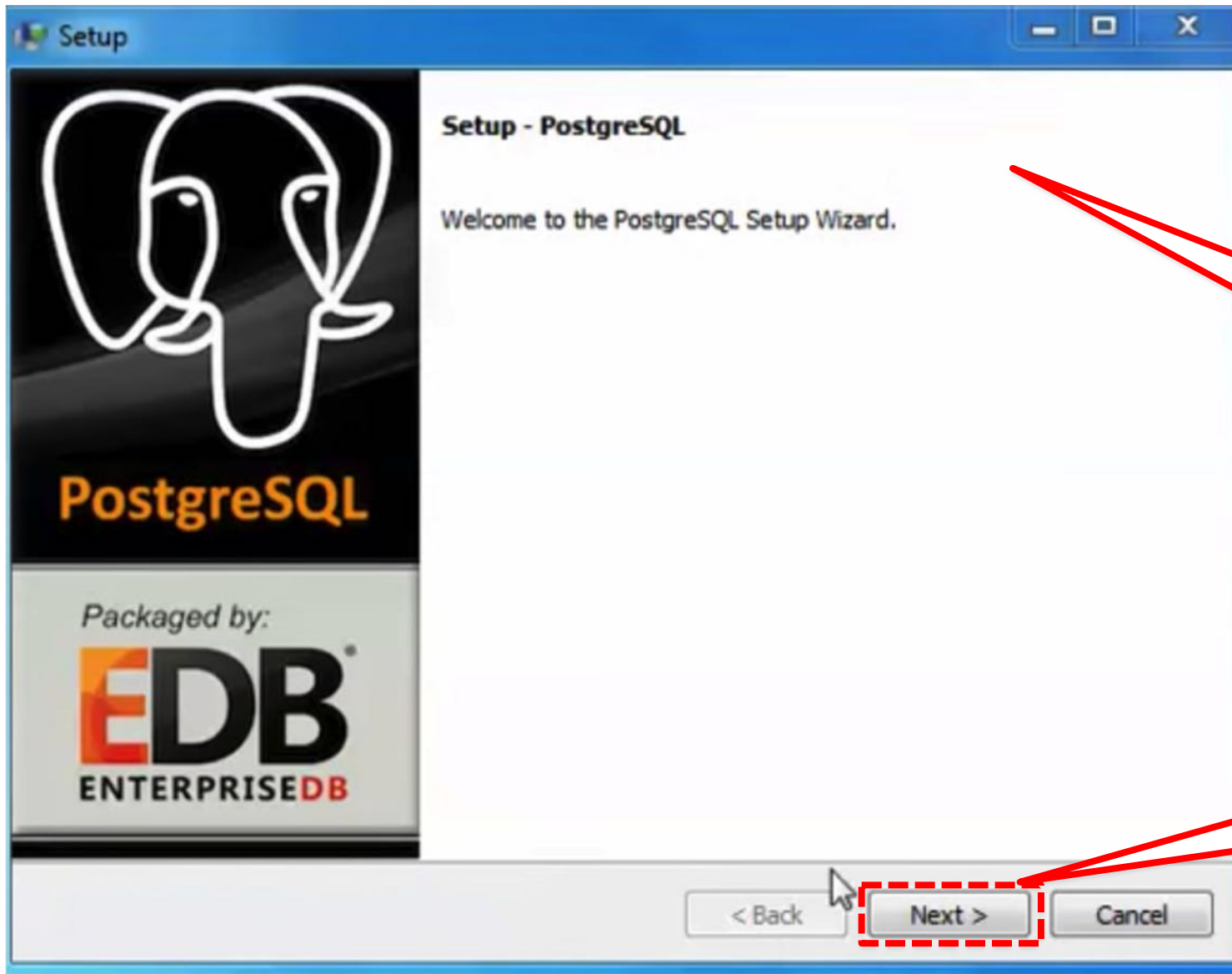
Advanced users can also download a [zip archive](#) of the binaries, without the installer. This download is intended for users who wish to include PostgreSQL as part of another application installer.

Platform support

- Double click on the downloaded file. A window will show up that will guide you through the installation.

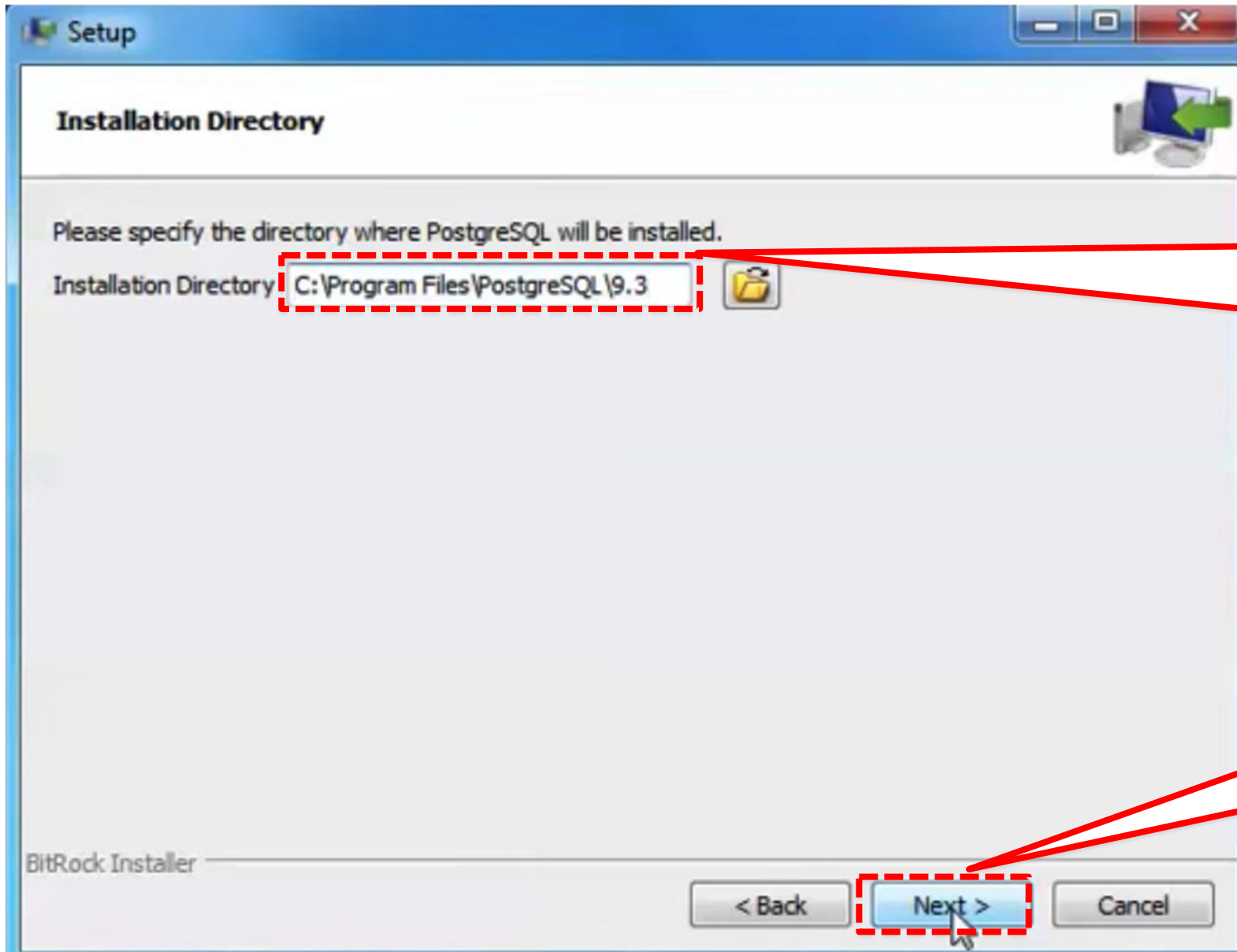


2) Double click on the downloaded file.



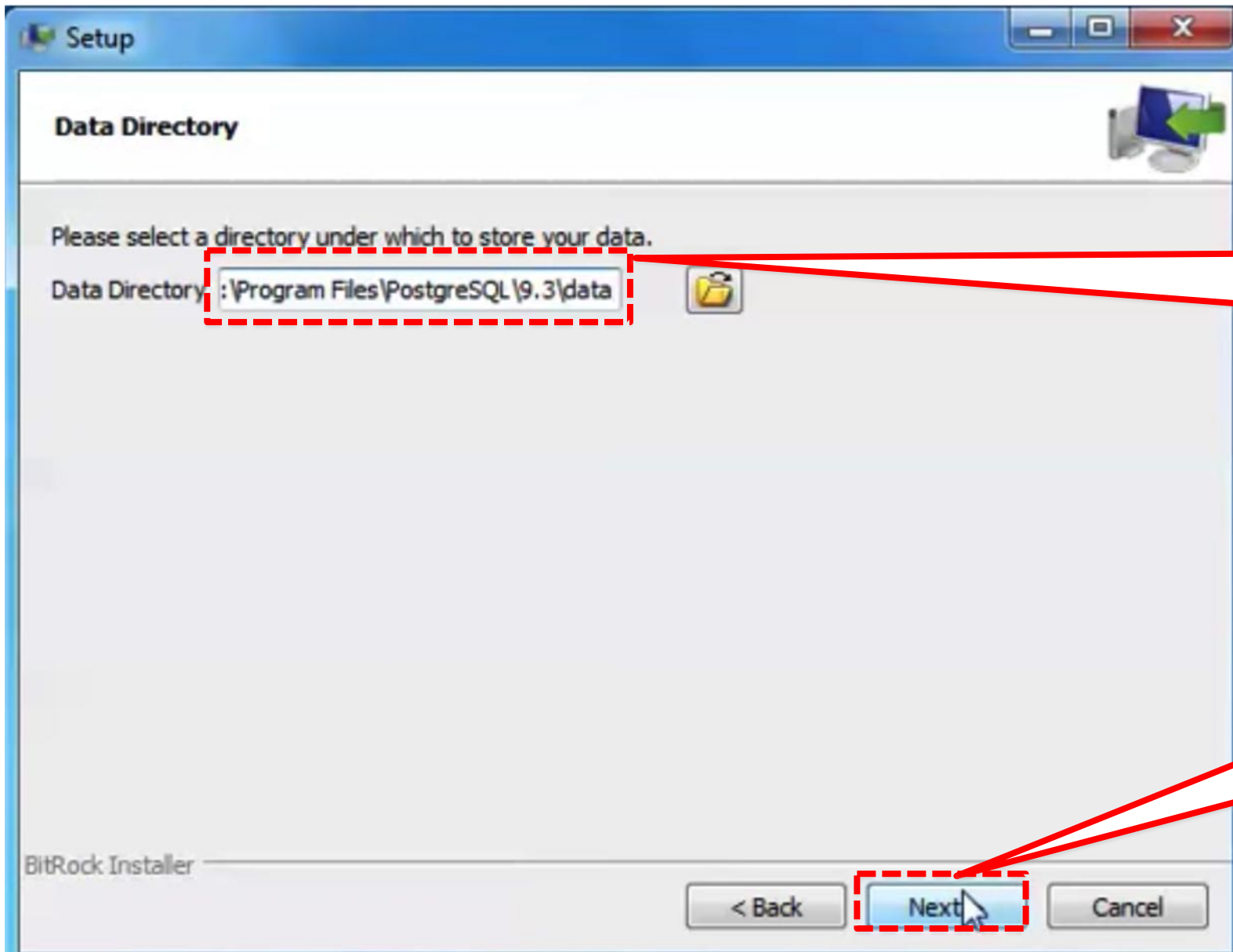
3) A window will show up that will guide you through the setup of PostgreSQL.

4) Click "Next" to continue the setup.



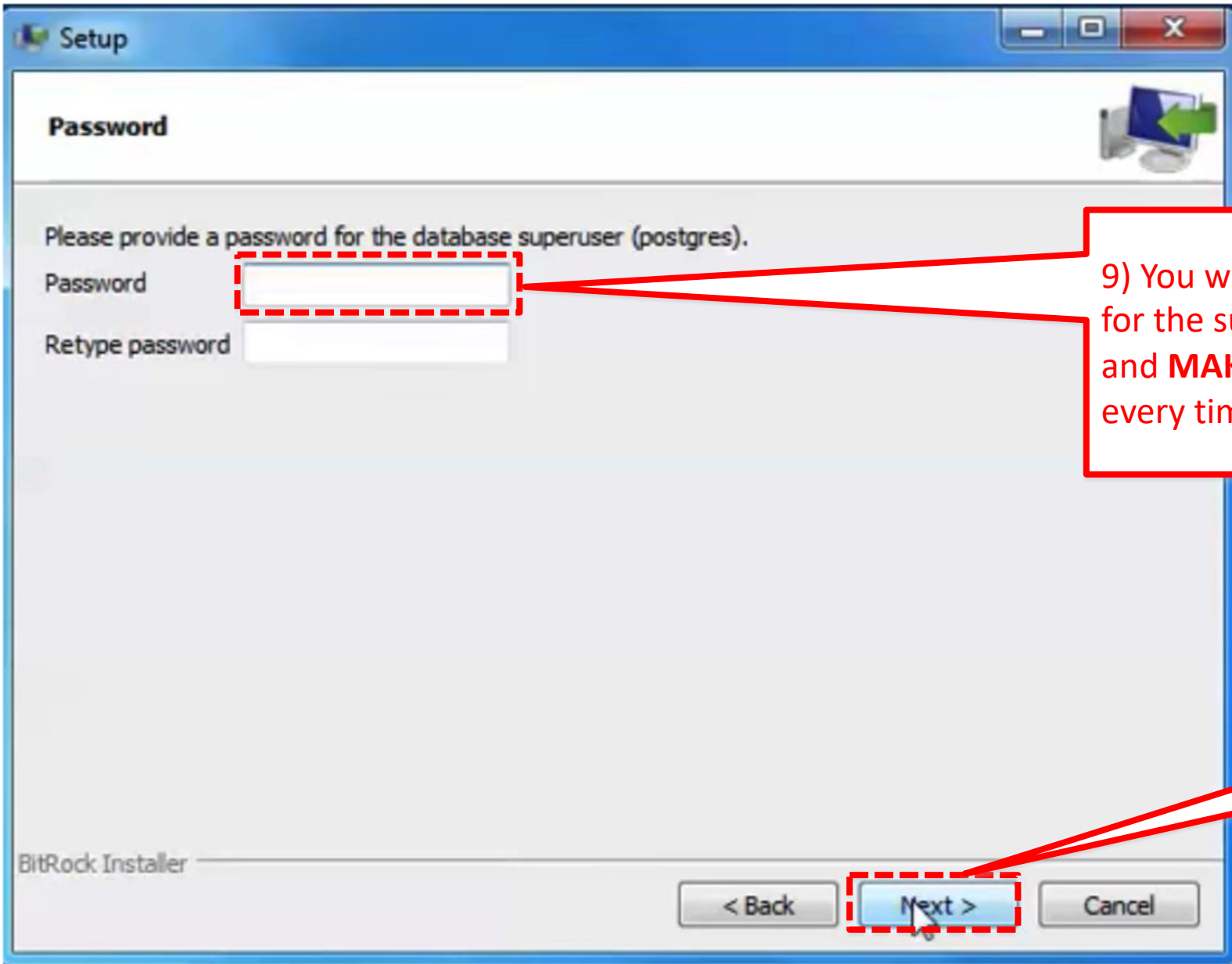
5) The installer wizard will ask you to specify a directory where PostgreSQL should be installed. It is right if you stick with the default option. So just click next.

6) Click "Next" to continue the setup.



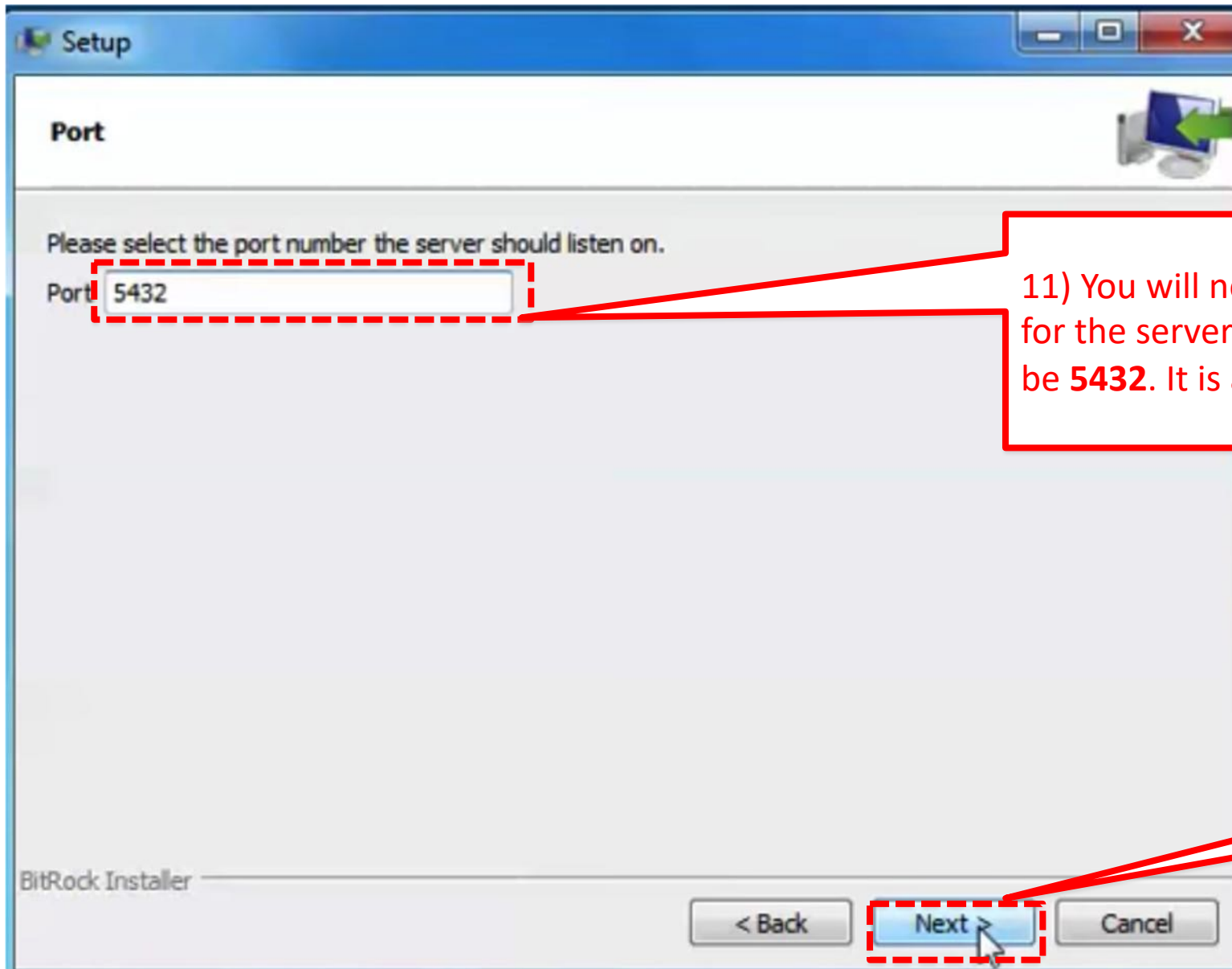
7) The installer wizard will then ask you to specify a data directory. Again, it is alright to stick with the default option.

8) Click "Next" to continue the setup.



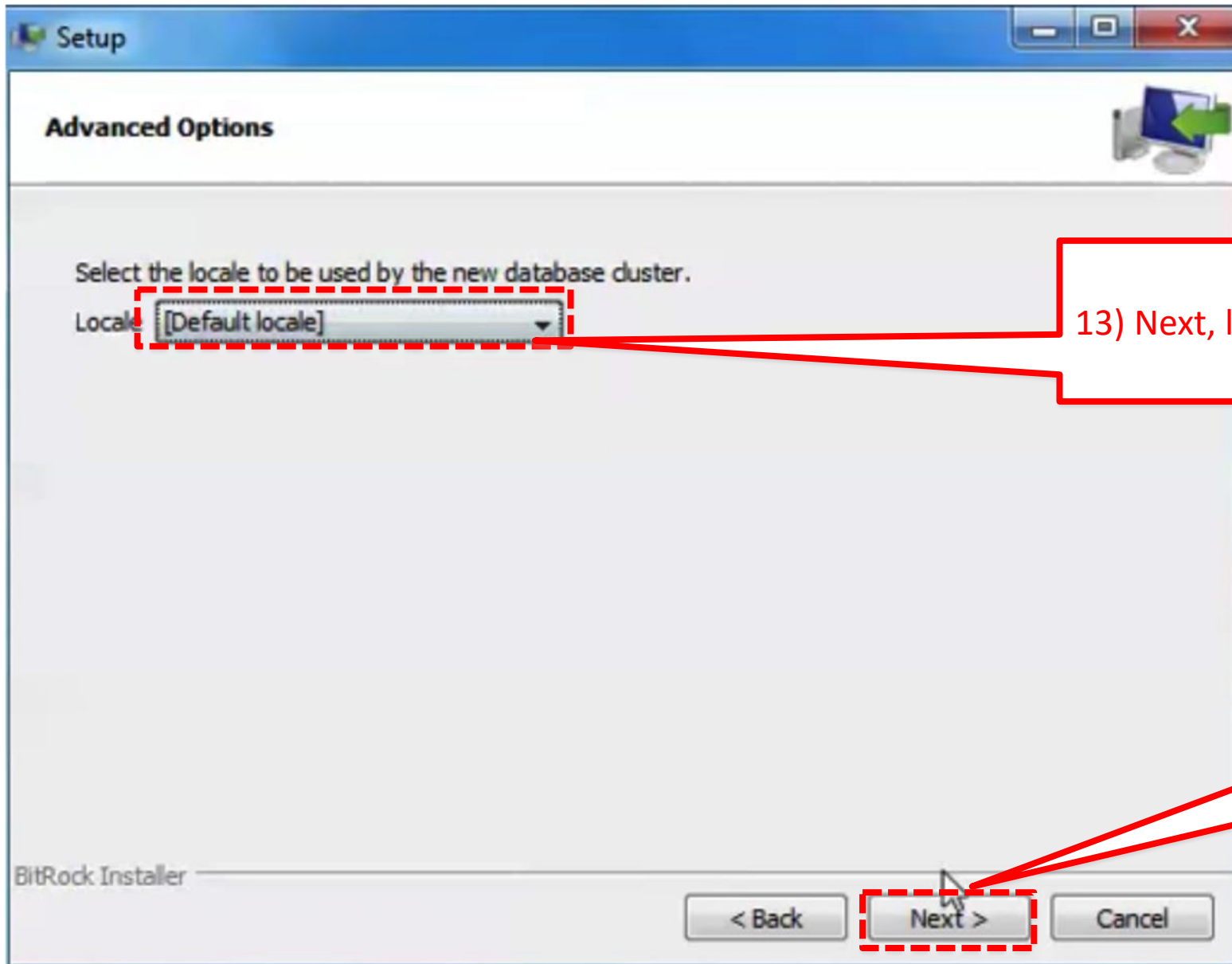
9) You will next be prompted to enter a password for the superuser "postgres". Enter the password and **MAKE A NOTE OF IT** as it will be required every time you want to work with postgres.

10) Click "Next" to continue the setup.



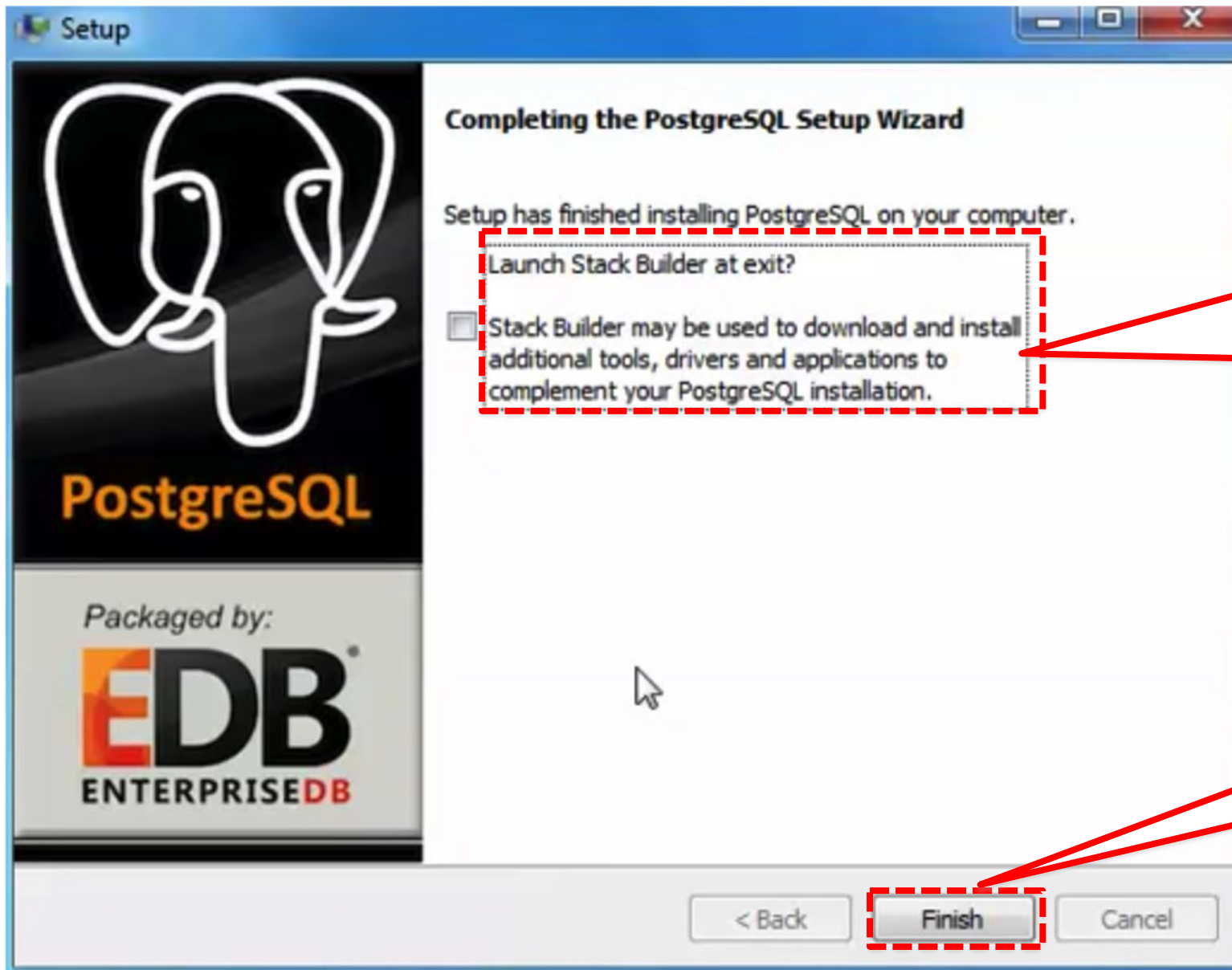
11) You will next be asked to enter a port number for the server to listen on. The default option should be **5432**. It is alright to keep it as it is.

12) Click "Next" to continue the setup.



13) Next, leave the locale selection at "Default locale"

14) Click "Next" to continue the setup.



15) Next, it will ask you if you want to install Stack Builder. You can check the box to install it, however you don't need Stack Builder right now and possibly not for the duration of this course. So, you uncheck the box.

14) Click "Finish" to complete the setup.

- To open psql now (which is the client we will use to create our database and interact with the database server) go to the command prompt and do the following:

17) Type this command, as it is.

```
C:\Users\Disha>psql -Upostgres
Password for user postgres: 
psql (10.1)
WARNING: Console code page (437) different from Windows code page (1252)
         8-bit characters might not work correctly. See psql reference
         page "Notes for Windows users" for details.
Type "help" for help.

postgres=#
```

18) It will then ask for the password you set during the PostgreSQL setup. Note: that the password you enter will not be visible to you, so just keep typing it CORRECTLY!

19) On executing the previous command, the prompt should change and should now look like "postgres=#"

- You are now in the psql program where you can enter queries to interact with the PostgreSQL server.
- Entering “\q” at this prompt should exit the psql program.

Example: postgres=# \q

3. Setup PostgreSQL (for Linux)

Step 1: Install PostgreSQL

Here are the installation steps on Ubuntu (this installation will also work on any Debian-based distribution):

1. Open a terminal window.
2. Issue the command *sudo apt-get install postgresql*.
3. Type the sudo password necessary to give you admin rights and hit Enter.
4. Allow apt to pick up any necessary dependencies.

Step 2: Change the default user password

If you don't follow this step, you will not be able to add databases and administer PostgreSQL, and the database will not be secure.

- Here's how to change the password for the default user. The user in question is postgres, and the password is changed like so:
 1. Open a terminal window.
 2. Issue the command *sudo passwd postgres*.
 3. Type (and confirm) the password to be used for this user.
- The postgres user will be the only user on your system that can open the PostgreSQL prompt without defining a database, which means postgres is the only user who can administer PostgreSQL.
- To test this, change to the postgres user with the command *su - postgres* and then enter the command *psql*. You should now be at the Postgres prompt, which looks like:

postgres=#

- All other users gain access to the prompt like so:

psql DB_NAME

Where, DB_NAME is the name of an existing database.

Step 3: Change the Postgres admin password

The administrator password must be set; otherwise, external applications will not be able to communicate with the database.

To change the admin password for Postgres, follow these steps:

1. Open a terminal window.
2. Change to the postgres user.
3. Log in to the postgres prompt.
4. Issue the command `\password postgres`.
5. Enter (and verify) the new password.
6. Exit the prompt with the command `\q`.

Step 4: Create your first database

This is where it gets exciting. Let's create a new database called testdb. To do this, follow these steps:

1. Open a terminal window.
2. Change to the postgres user.
3. Log in to the postgres prompt.
4. Issue the command *CREATE DATABASE ACTOR;* .

4. Importing the data










RESOURCES

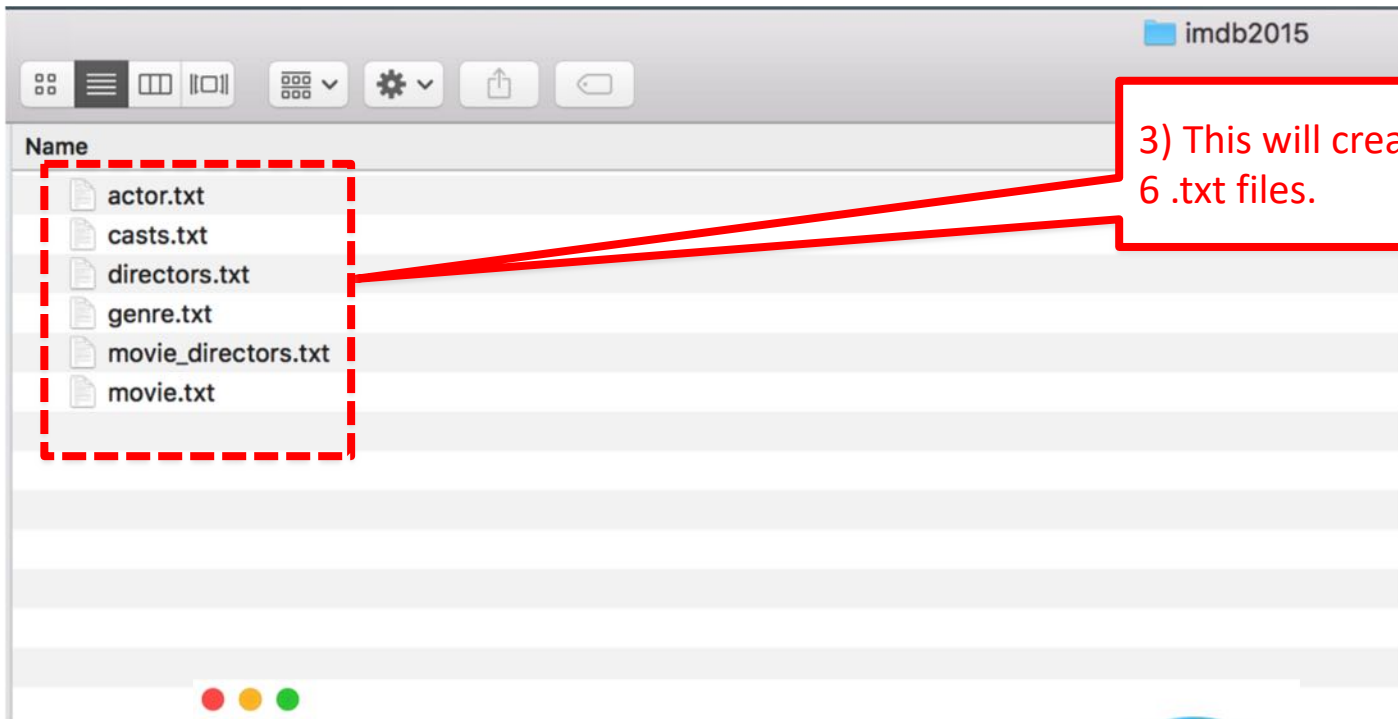
- Blackboard
- Piazza: access code posted on Blackboard
- Gradiance: class token posted on Blackboard, Setup Gradiance <http://www.gradiance.com>
- Lucidcharts: handy for drawing ER diagrams
- Textbooks
 - "GUW": Database textbook: Garcia, Widom, and Widom. Database Systems: The Complete Book (3rd international ed), [Textbook rental (30\$ rental)], [Amazon (30\$)], [EBSCOhost (30\$)]
 - "SAMS": Forta. SAMS Teach Yourself SQL in 10min. 4th ed. [Amazon (30\$)], [EBSCOhost (30\$)]
- Lecture slides: will be posted after each class, by end of the day
- SQL
 - SQL files: sql files to run and follow along our SQL lectures
 - SQL files for SAMS book
 - SQL files for Chinook
 - SQLite: in Firefox (recommended): Setup SQLite (slides), Firefox SQLite add-on
 - SQLite in Google Chrome (alternative if FF does not work for you): Setup SQLite (Chrome)
 - PostgreSQL: Setup PostgreSQL (slides)

1) Go to our sql directory on our website and a page with lots of files will pop-up.

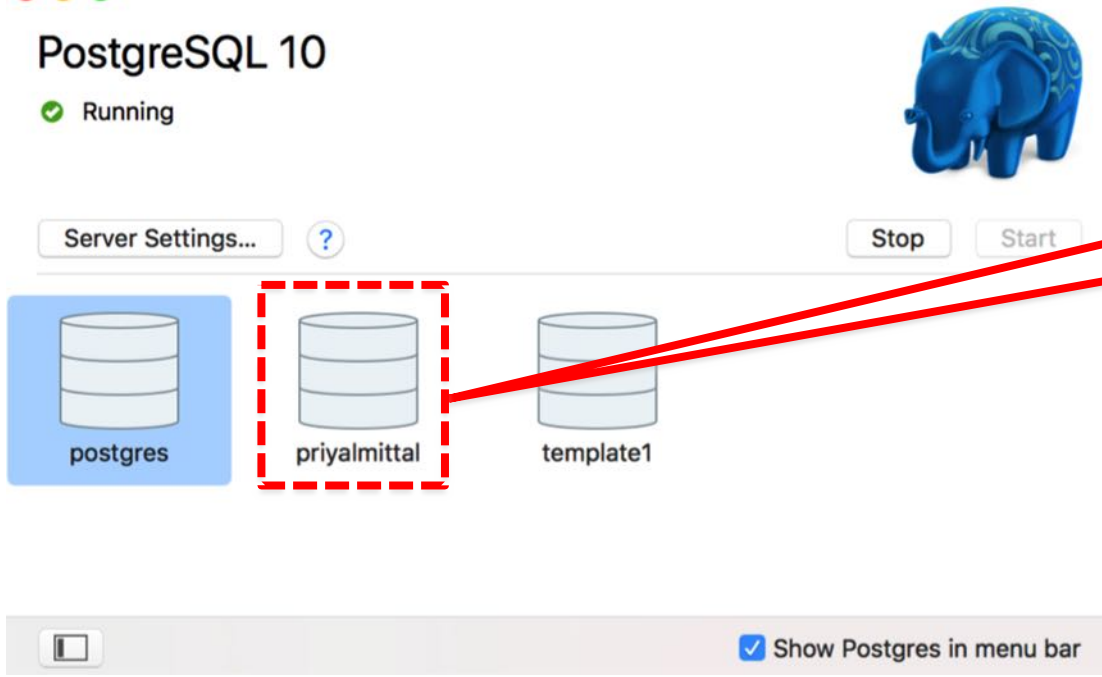
2) Download imdb-cs3200.zip and extract it.

Index of /cs3200sp18s2/sql

	Name	Last modified	Size	Desc
	Parent Directory		-	
	300 - Small IMDB for..>	2018-01-11 00:47	258K	
	302 - Simple product..>	2018-01-11 00:47	1.6K	
	305 - Conceptual eva..>	2018-01-11 00:47	827	
	308 - Purchase - SQL..>	2018-01-11 00:47	865	
	312 - Person and add..>	2018-01-11 19:19	1.1K	
	315 - Other company ..>	2018-01-11 19:19	1.4K	
	318 - Chinook - SQLi..>	2018-01-11 00:47	1.8M	
	322 - OLAP sales - S..>	2018-01-11 00:47	2.6K	
	331 - Person Bar dri..>	2018-01-11 00:47	1.5K	
	334 - Items - SQLite..>	2018-01-11 00:47	1.0K	
	340 - Sailors - SQLi..>	2018-01-11 00:47	2.5K	
	344 - Employees - SQ..>	2018-01-11 00:47	1.5K	
	348 - Car - SQLite.txt	2018-01-11 00:47	1.1K	
	361 - Translations -..>	2018-01-11 00:47	1.2K	
	380 - SAMS - SQLite..>	2018-01-11 00:47	14K	
	imdb-cs3200.zip	2018-01-11 01:08	264M	



3) This will create a folder named imdb2015, containing 6 .txt files.



4) Now, in order to create a database, open Postgres, make sure that the server is running and connect to psql.

```
Last login: Sun Jan 14 03:05:12 on ttys000
Priyals-MacBook-Pro:cs5200-spring2018-mittal priyalmittal$ "/Applications/Postgres.app/Contents/Versions/10/bin/psql" -p5432 -d "priyalmittal"
psql (10.1)
Type "help" for help.
```

```
priyalmittal=#
```

5) Once the server is running and connected to psql, you should see something like `<your_user_name>=#` as your prompt.

```
Last login: Sun Jan 14 03:05:12 on ttys000
Priyals-MacBook-Pro:cs5200-spring2018-mittal priyalmittal$ "/Applications/Postgres.app/Contents/Versions/10/bin/psql" -p5432 -d "priyalmittal"
psql (10.1)
Type "help" for help.
```

```
priyalmittal=# CREATE DATABASE imdb;
CREATE DATABASE
priyalmittal=# \c imdb
You are now connected to database "imdb" as user "priyalmittal".
imdb=#
```

6) To create a new database, enter the following commands:

```
priyalmittal=# CREATE DATABASE imdb;
priyalmittal=# \c imdb
```

Your prompt will then switch to `imdb=#`


```
priyalmittal=# create database imdb
CREATE DATABASE
priyalmittal=# \c imdb
You are now connected to database
imdb=# CREATE TABLE ACTOR (id INT
CREATE TABLE
imdb=# CREATE TABLE MOVIE (id INT
CREATE TABLE
imdb=# CREATE TABLE DIRECTORS (id
CREATE TABLE
imdb=# CREATE TABLE CASTS (pid INT
CREATE TABLE
imdb=# CREATE TABLE MOVIE_DIRECTOR
CREATE TABLE
imdb=# CREATE TABLE GENRE (mid INT
CREATE TABLE
```

7) You can now use the following command to create a table:
`imdb=# CREATE TABLE table_name(.....);`

You will have to decide on the appropriate data types 😊

8) Choose appropriate types for each column and specify all key constraints as described here. (part of your first homework)

9) Once tables are created, you can use the psql command `\copy`, to import data from the corresponding text files.

10) This command imports data from the "actor.txt" file, which is to be mentioned along with its complete path, into the ACTOR table.

```
imdb=# \copy ACTOR from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/actor.txt' with delimiter '|' null as ''
COPY 3442863
imdb=# \copy MOVIE from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/movie.txt' with delimiter '|' null as ''
COPY 2383721
imdb=# \copy DIRECTORS from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/directors.txt' with delimiter '|' null as ''
COPY 583174
imdb=# \copy CASTS from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/casts.txt' with delimiter '|' null as ''
COPY 16932400
imdb=# \copy MOVIE_DIRECTORS from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/movie_directors.txt' with delimiter '|' null as ''
COPY 3250604
imdb=# \copy GENRE from '/Users/priyalmittal/Documents/cs3200-Database Design/imdb2015/genre.txt' with delimiter '|' null as ''
COPY 956674
imdb=#
```

11) The delimiter specifies how the fields are separated, whereas the last part of the statement specifies that the empty string should be assigned as a NULL value.

NOTE: The `\copy` command may take a few minutes to complete for some tables (likely longer than 10min for casts). The exact speed of these operations may vary depending on your machine.

NOTE: You may get an error when running the copy command that looks similar to this:

ERROR: invalid byte sequence for encoding "UTF8": 0xc3 0x7c.

That means that psql is set to an encoding that does not match the file. Try changing the encoding, and try copying again.

```
imdb=# set client_encoding to 'latin1';
```

5. Creating the .sql files

```
priyalmittal=# create database imdb;
CREATE DATABASE
priyalmittal=# \c imdb
You are now connected to database "imdb" as user "priyalmittal".
imdb=# CREATE TABLE ACTOR (id
CREATE TABLE
imdb=# CREATE TABLE MOVIE (id
CREATE TABLE
imdb=# CREATE TABLE DIRECTORS
CREATE TABLE
imdb=# CREATE TABLE CASTS (pid
CREATE TABLE
imdb=# CREATE TABLE MOVIE_DIREC
CREATE TABLE
imdb=# CREATE TABLE GENRE (mid
CREATE TABLE
```

1) Copy the queries you wrote in the command prompt as it is, WITHOUT the "imdb=#".

```
IMDBqueries.sql
IMDBqueries.sql x
1  --Schema for each table is as follows:--
2
3  --ACTOR--
4  CREATE TABLE ACTOR (id INTEGER PRIMARY KEY, fname TEXT, lnmae TEXT, gender CHAR);
5  --MOVIE--
6  CREATE TABLE MOVIE (id INTEGER PRIMARY KEY, name TEXT, year INTEGER);
7  --DIRECTORS--
8  CREATE TABLE DIRECTORS (id INTEGER PRIMARY KEY, fname TEXT, lname TEXT);
9  --CASTS--
10 CREATE TABLE CASTS (pid INTEGER, mid INTEGER, role TEXT, FOREIGN KEY(pid)
11     REFERENCES ACTOR(id), FOREIGN KEY(mid) REFERENCES MOVIE(id));
12 --MOVIE_DIRECTORS--
13 CREATE TABLE MOVIE_DIRECTORS (did INTEGER, mid INTEGER, FOREIGN KEY(did) REFERENCES
14     DIRECTORS(id), FOREIGN KEY(mid) REFERENCES MOVIE(id));
15 --GENRE--
16 CREATE TABLE GENRE (mid INTEGER, genre TEXT);
```

2) You may add comments into the file as mentioned here.

3) Paste the queries into a text editor, AS IT IS, without making any changes to the font, color, size, etc. and save the file with the extension of .sql

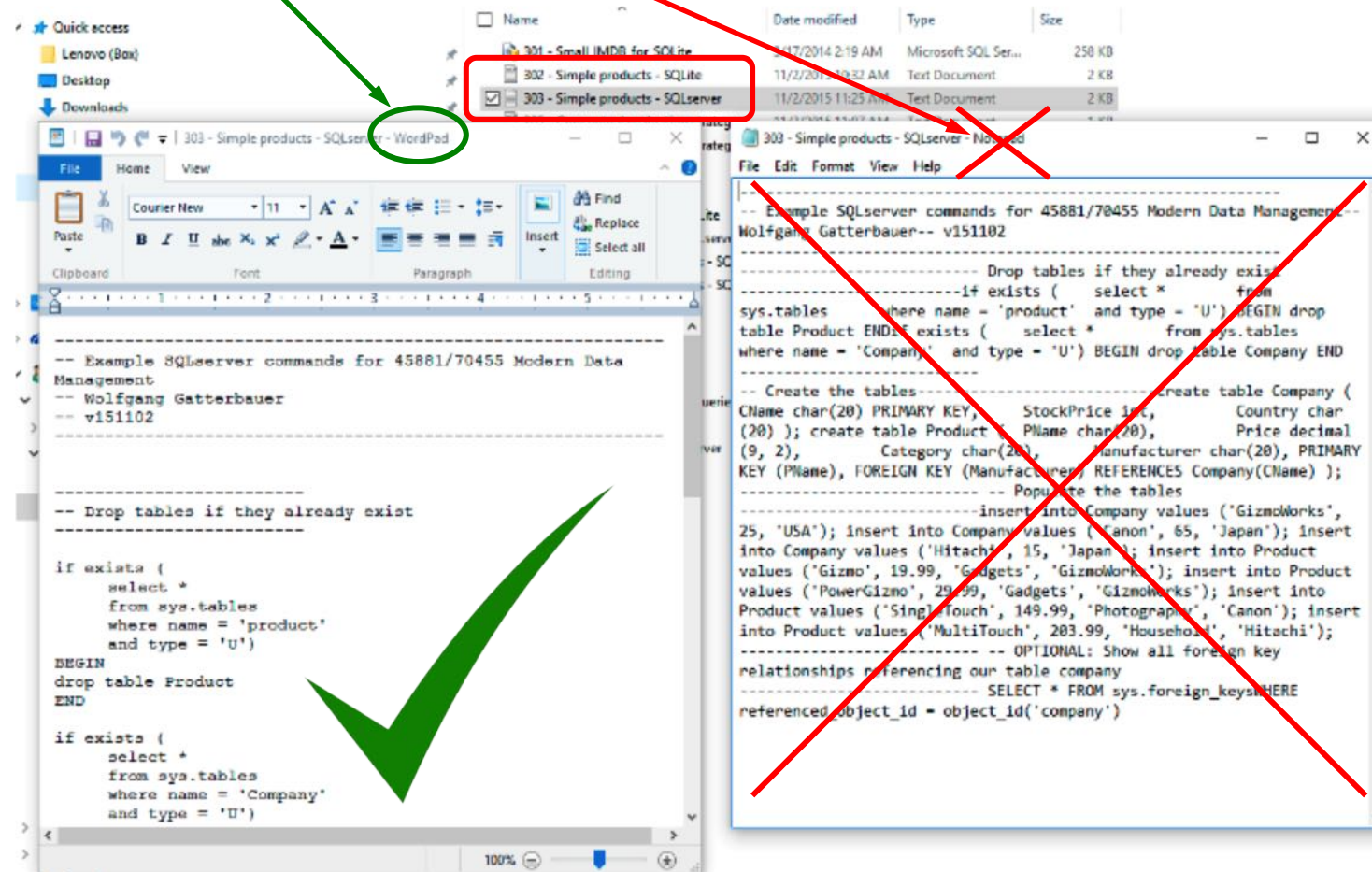
PLEASE NOTE: Adding any extra symbols into your .sql file, such as / at the end of each line or anywhere else, # , \$, etc. may result into errors which wouldn't let your file get executed.

The only added text permitted is any line that is preceeded by "--":
-- is used to add **comments** to the file.

Careful in the choice of your text editor

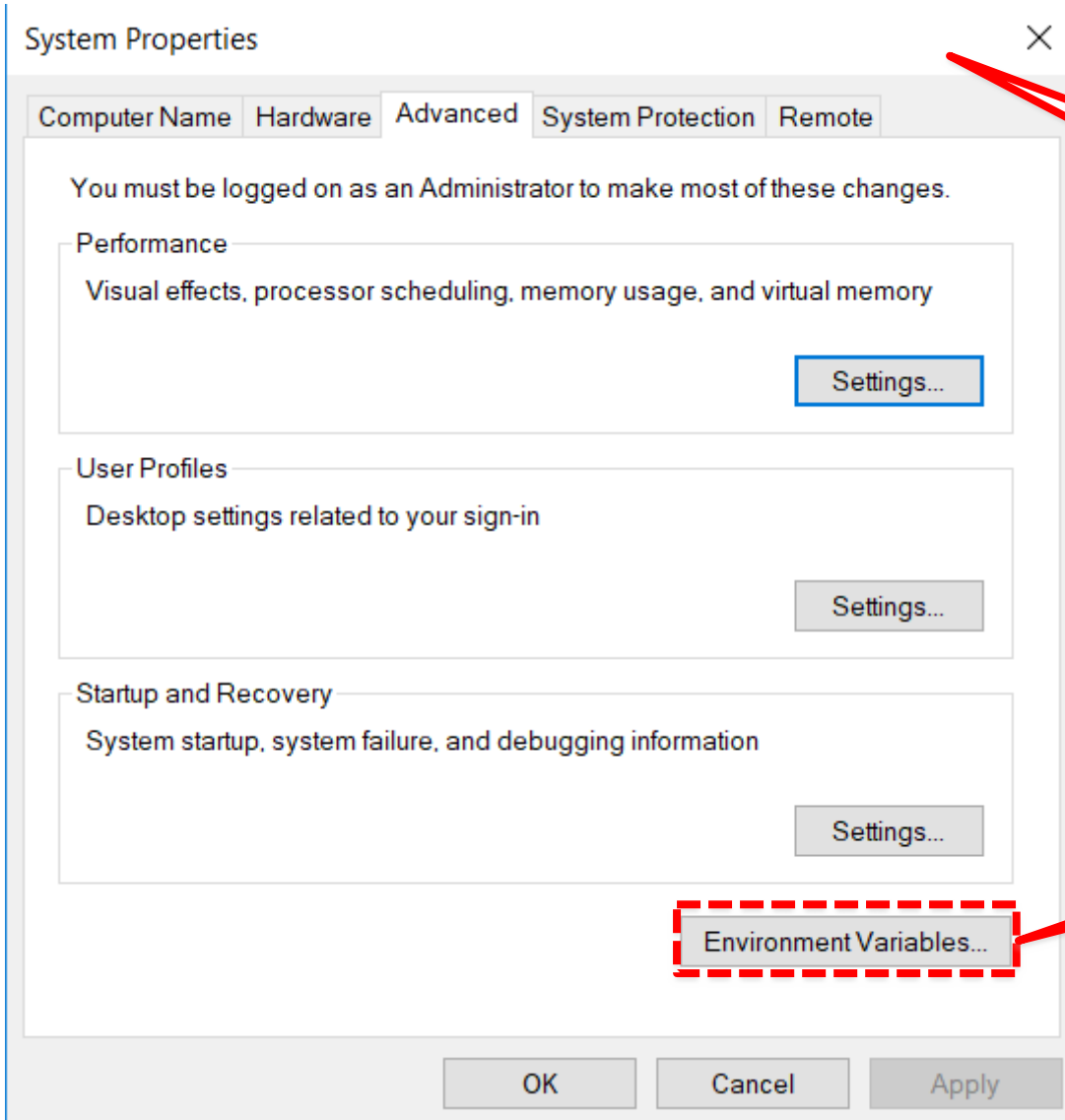
If you are using Windows:

1. Download the appropriate text files from our repository
2. Open them with **"Wordpad"** (not **"Notepad"** which messes up the text!)
3. Paste the SQL commands into your SQLite version, and execute



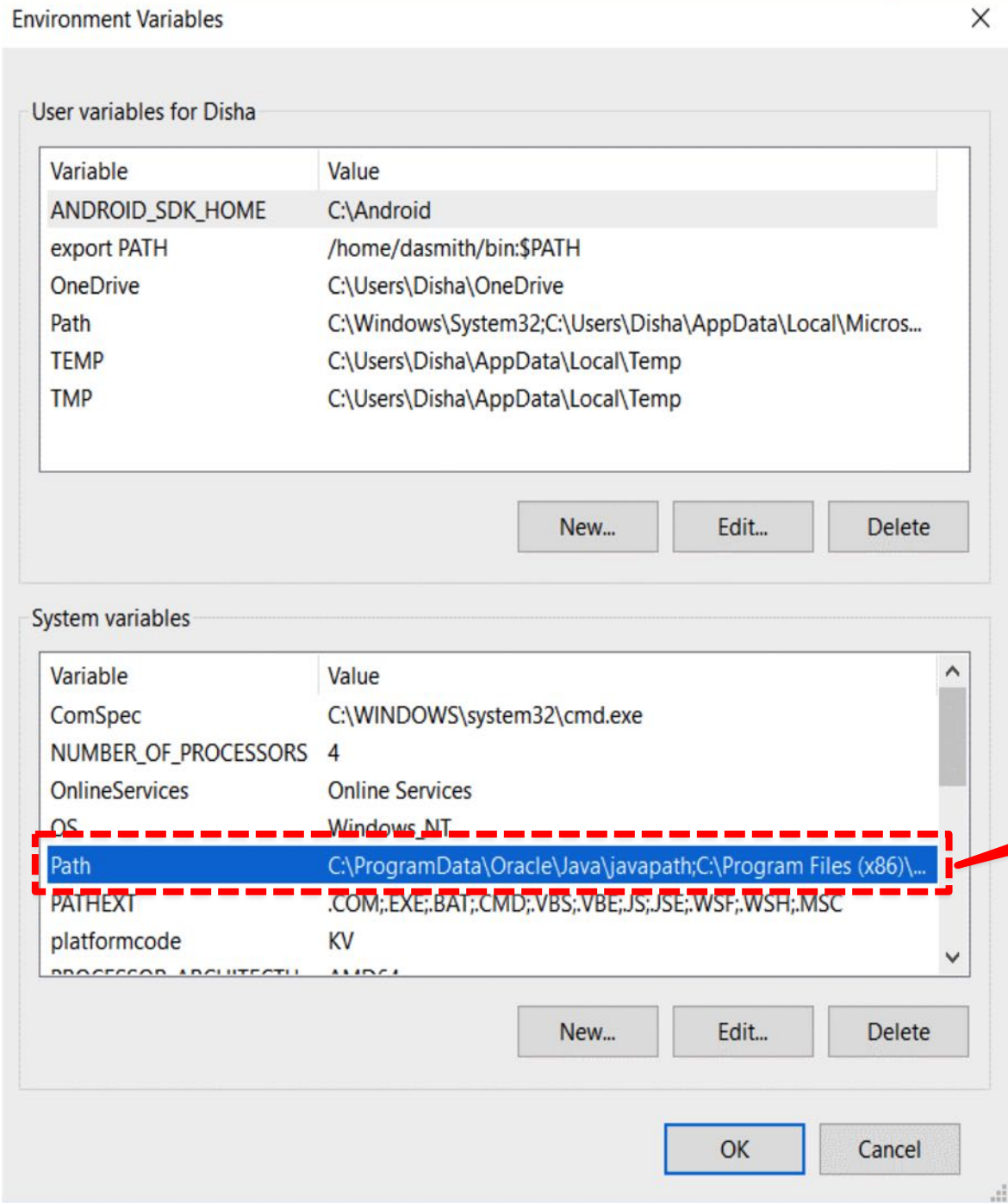
6. Environment Variables Setup

- If you encounter the error: **“psql not recognized as an internal or external command”** issue for PostgreSQL in Windows, then you are supposed to follow these steps:



1) When you search your machine with **“Edit the system environment variable”**, this window shows up.

2) Click on **“Environment Variables”**.



3) Double-click on the highlighted path.

Edit environment variable

- C:\Program Files (x86)\Intel\Intel(R) Management Engine Co...
- C:\Program Files\Intel\Intel(R) Management Engine Compon...
- C:\Program Files (x86)\Intel\Intel(R) Management Engine Co...
- C:\Program Files\Intel\Intel(R) Management Engine Compon...
- C:\Python27
- C:\Python27\Lib\site-packages
- C:\Program Files (x86)\PharosSystems\Core
- C:\Java\jdk1.5.0\bin
- C:\Java\jre1.5.0\bin
- C:\Android
- C:\Windows\System32
- C:\Program Files\Microsoft SQL Server\120\Tools\Binn\
- C:\Program Files\Microsoft SQL Server\130\Tools\Binn\
- C:\Program Files (x86)\Windows Kits\10\Windows Performan...
- C:\Program Files\PuTTY\
- C:\Program Files (x86)\Skype\Phone\
- C:\NLP\bookNLP
- C:\Program Files\PostgreSQL\10\bin**
- C:\Program Files\Intel\WiFi\bin\
- C:\Program Files\Common Files\Intel\WirelessCommon\
- C:\Program Files\Git\cmd

- New
- Edit
- Browse...
- Delete
- Move Up
- Move Down
- Edit text...

- OK
- Cancel

4) Click on "New" and add the highlighted path below.
If you followed the default, it should be the same.
Otherwise, you can check for the **bin** folder in the **PostgreSQL** folder and copy its path.

5) Click on "OK" and restart command prompt.
This should hopefully solve the issue.

7. How to import an SQL file

- You can run the chinook sql file using the following command in postgres prompt:

```
\i '319 - Chinook – PostgreSql.sql';
```

- **\i filename. sql** is the command.
- Also, use the complete path of the file

Example:

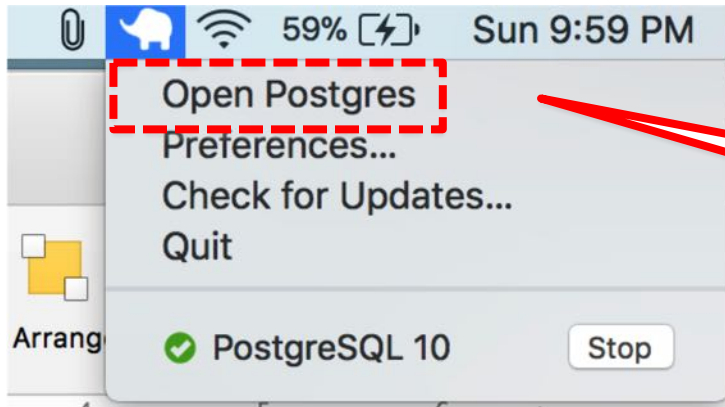
```
\i '/Users/...../foldername/319 – Chinook – PostgreSql.sql';
```

- You can also just copy and paste the file into PgAdmin.

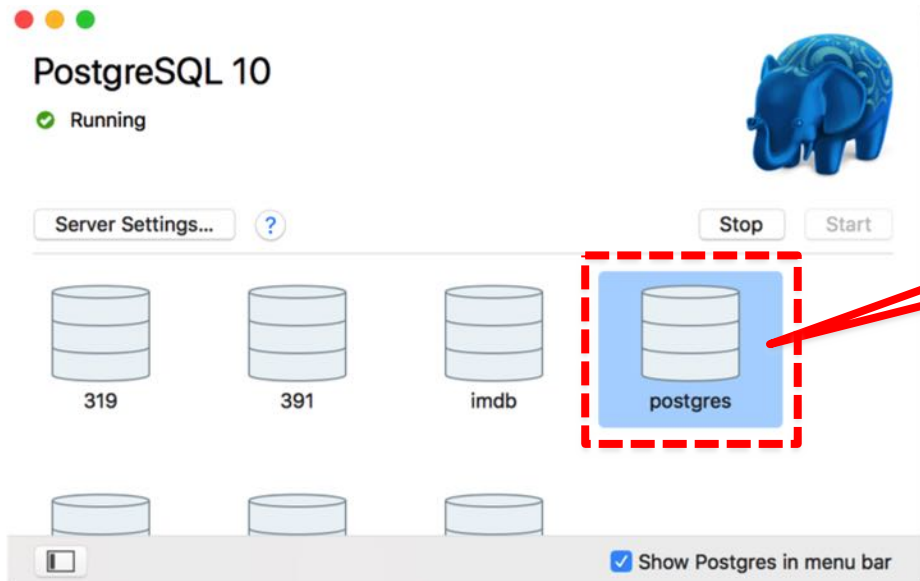
8. PgAdmin (version 3 or 4)

- pgAdmin is a visual client for your postgres database
- You can either use the latest version 4 or the older version 3
- Notice that – according to many users – version 4 is more cumbersome than version 3. We thus recommend that you download version 3 instead of version 4:
<https://www.pgadmin.org/download/>
- You may get some errors when first installing when working with postgres 10, but they should not prevent you from using it

Connecting to PostgreSQL database with pgAdmin 3 (after pgAdmin 3 is successfully installed)



1) Open Postgres, such that the slonik icon should appear on the top of your desktop.



2) Double click on the particular database, for whom the connection with pgAdmin 3 is to be established, to open its command prompt.


```
priyalmittal — psql -p5432 -d postgres — 80x24
Last login: Sun Feb  4 22:12:46 on ttys001
Priyals-MacBook-Pro:~ priyalmittal$ "/Applications/Postgres.app/Contents/Version
s/10/bin/psql" -p5432 -d "postgres"
psql (10.1)
Type "help" for help.
postgres=#
```

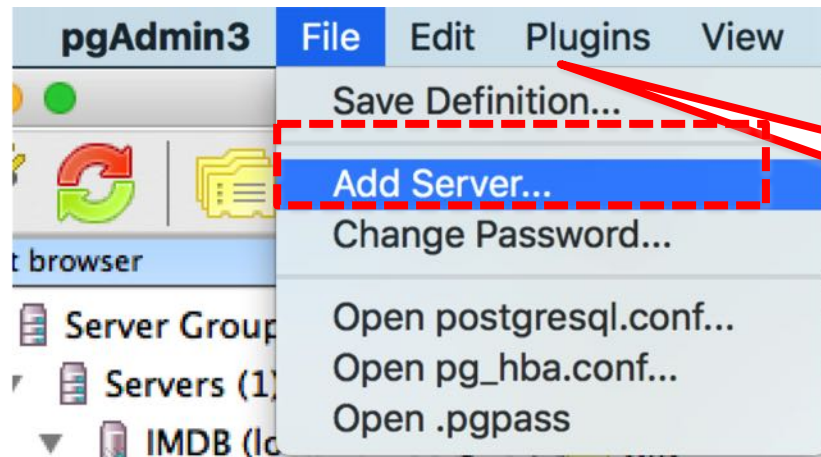
3) After clicking on the database, its corresponding command prompt should open up.

```
Last login: Sun Feb  4 22:11:56 on ttys001
Priyals-MacBook-Pro:~ priyalmittal$ "/Applications/Postgres.app/Contents/Versi
s/10/bin/psql" -p5432 -d "postgres"
psql (10.1)
Type "help" for help.
postgres=# ALTER USER postgres WITH PASSWORD 'NewPostgresPassword';
ALTER ROLE
postgres=#
```

4) Write this statement as it is with the PASSWORD of your choice.

NOTE:
(i) Please, note the password somewhere as you will need to enter it into pgAdmin, every time you want to run queries for a database.
(ii) REMEMBER TO ADD THE SEMI-COLON.

Now, to setup a new serve connection in pgAdmin, open pgAdmin and:

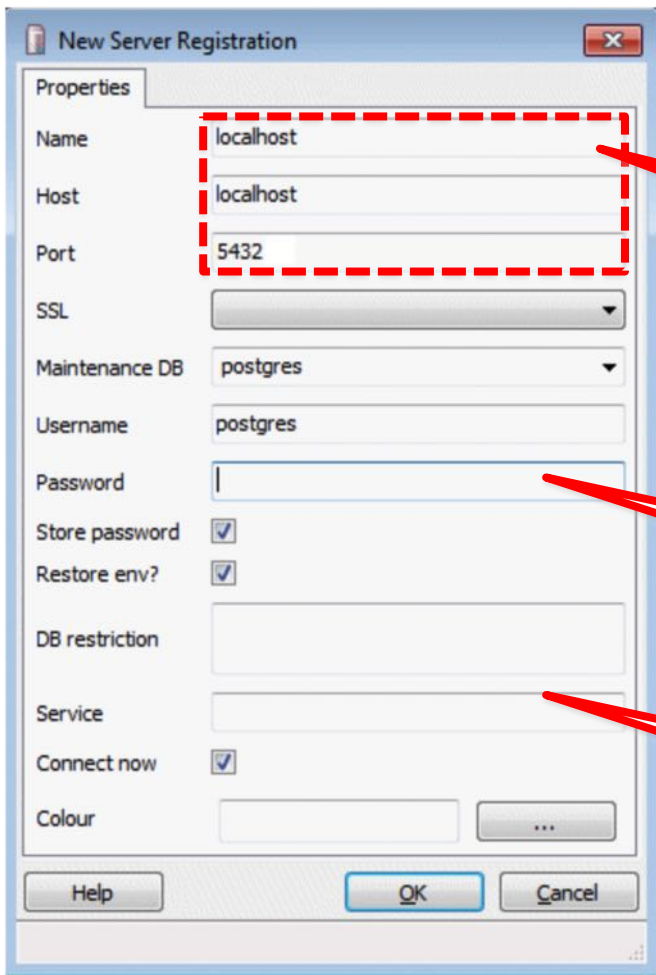


5) Go to File -> Add Server

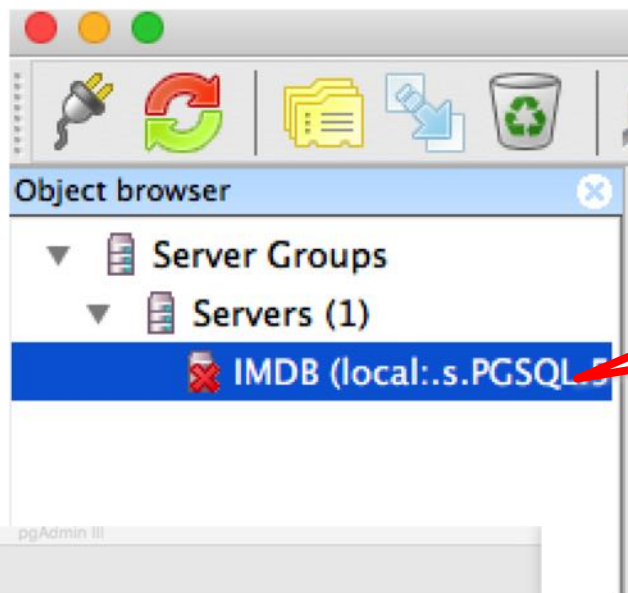
6) Ensure that the Name and Host is set to localhost and the Port to 5432 (unless you have configure PostgreSQL for a different port)

7) Enter the Password you just set for the database in the Postgres command prompt.

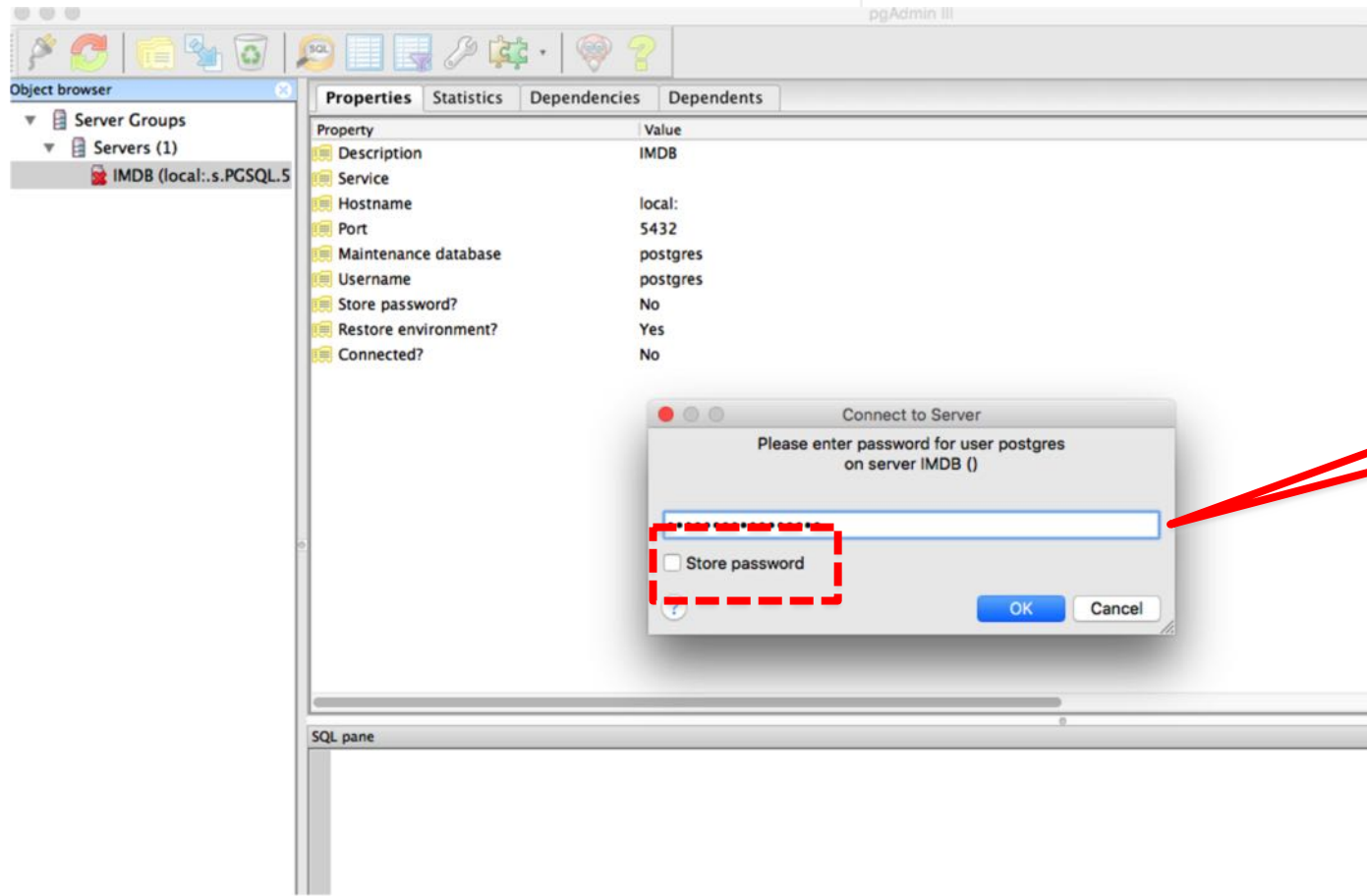
REST ALL REMAINS THE SAME!



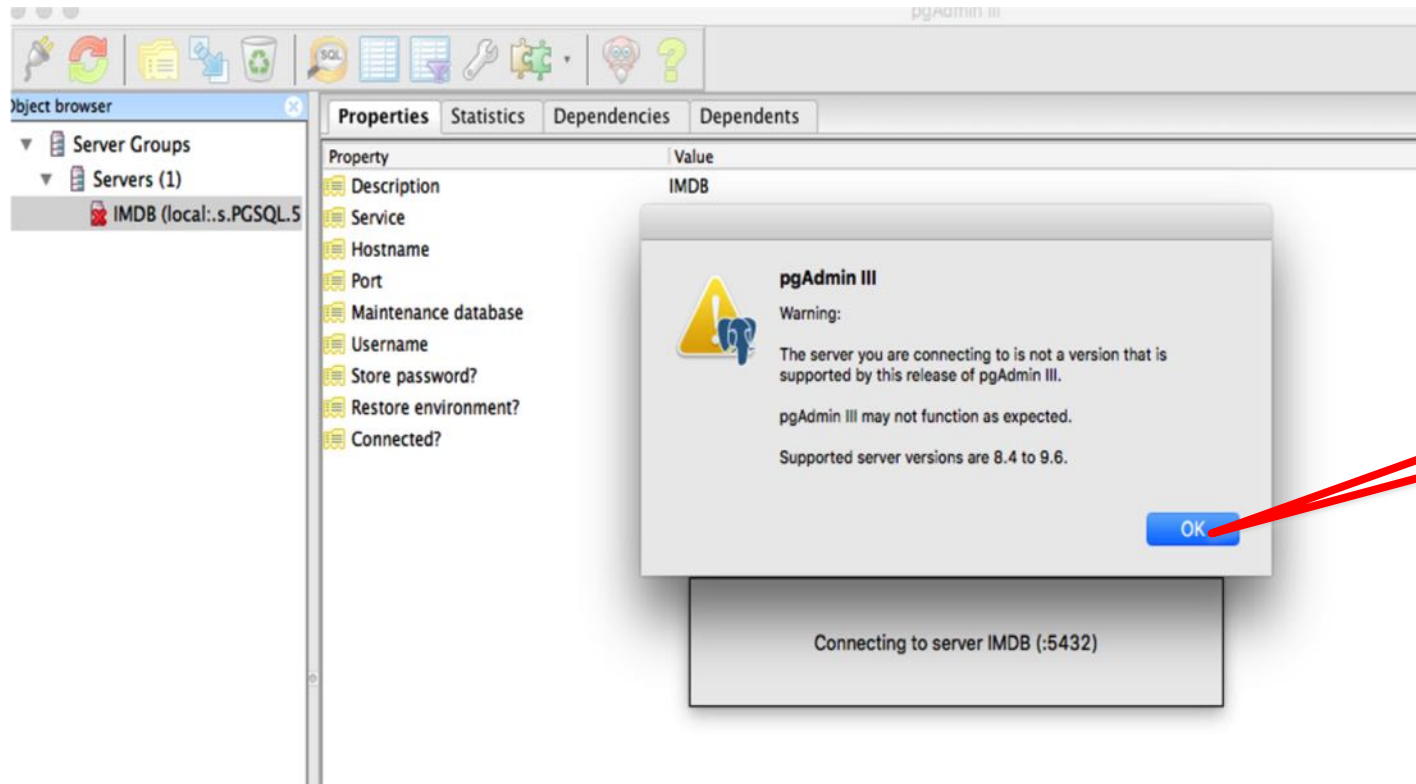
To connect to the server:



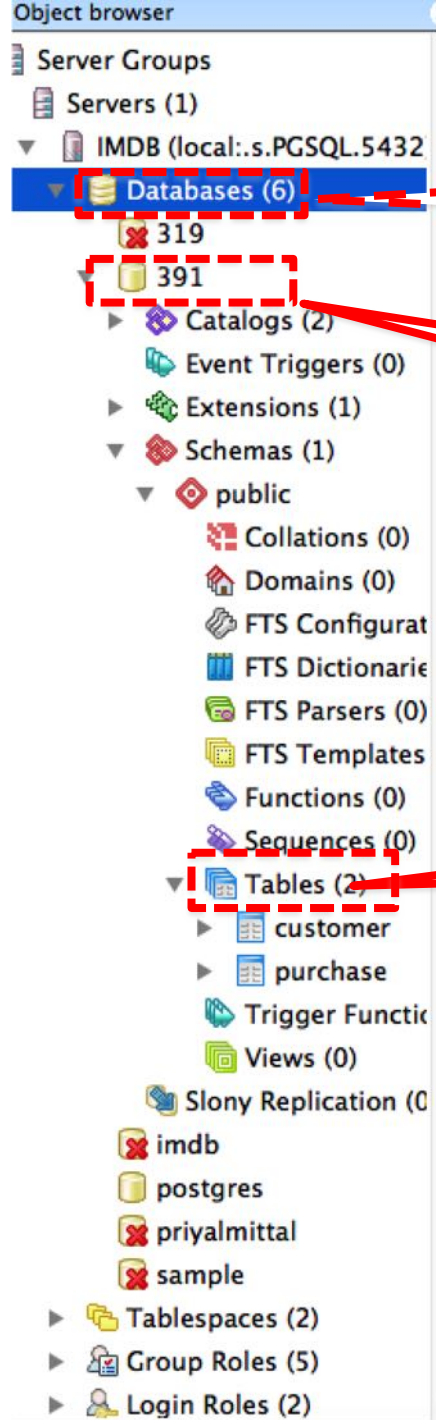
8) Double-click on the server item.



9) Provide the password when prompted.
You probably want to store it



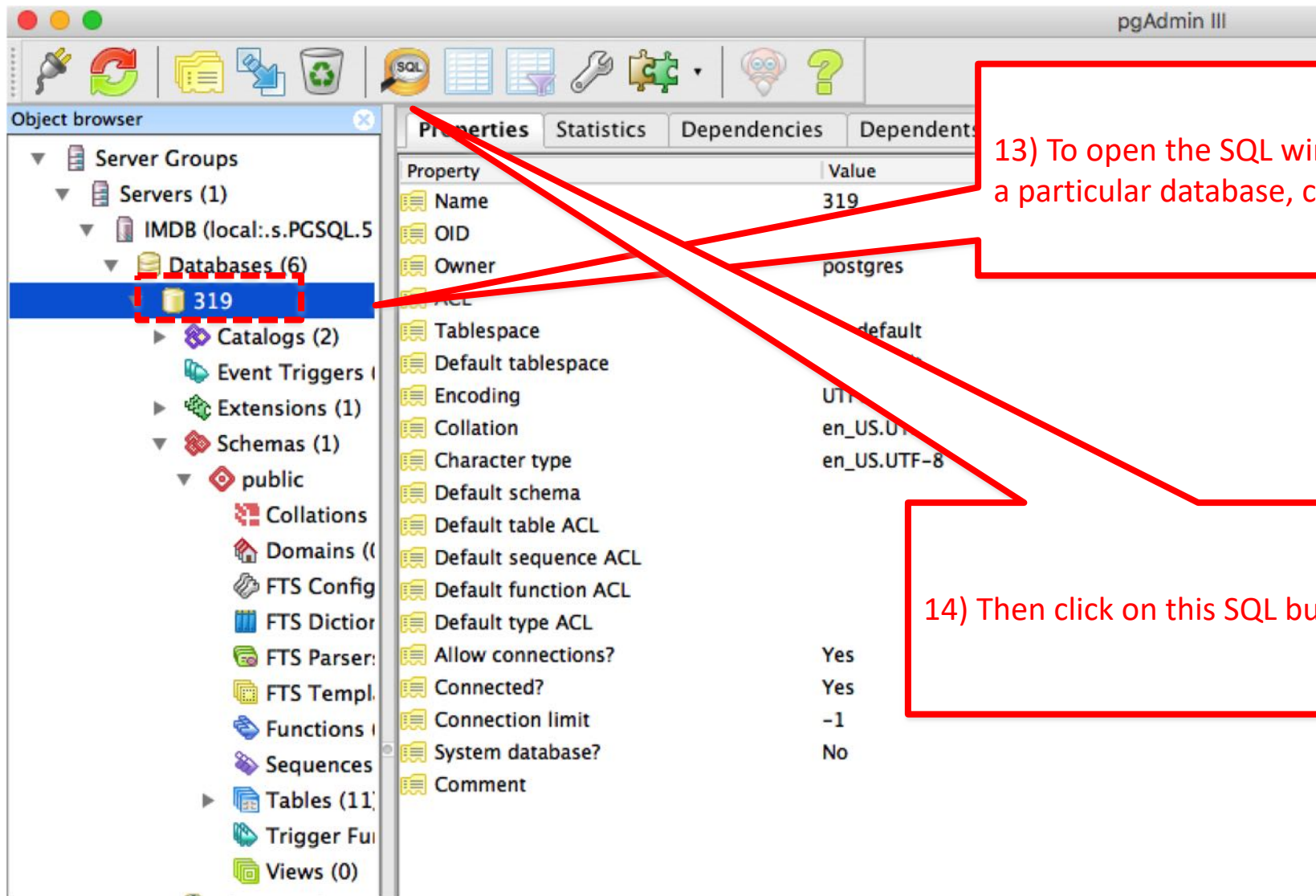
10) Click OK every time you get a pop-up.
You will have to click OK several times.



Right click on Databases in order to create and add a new database

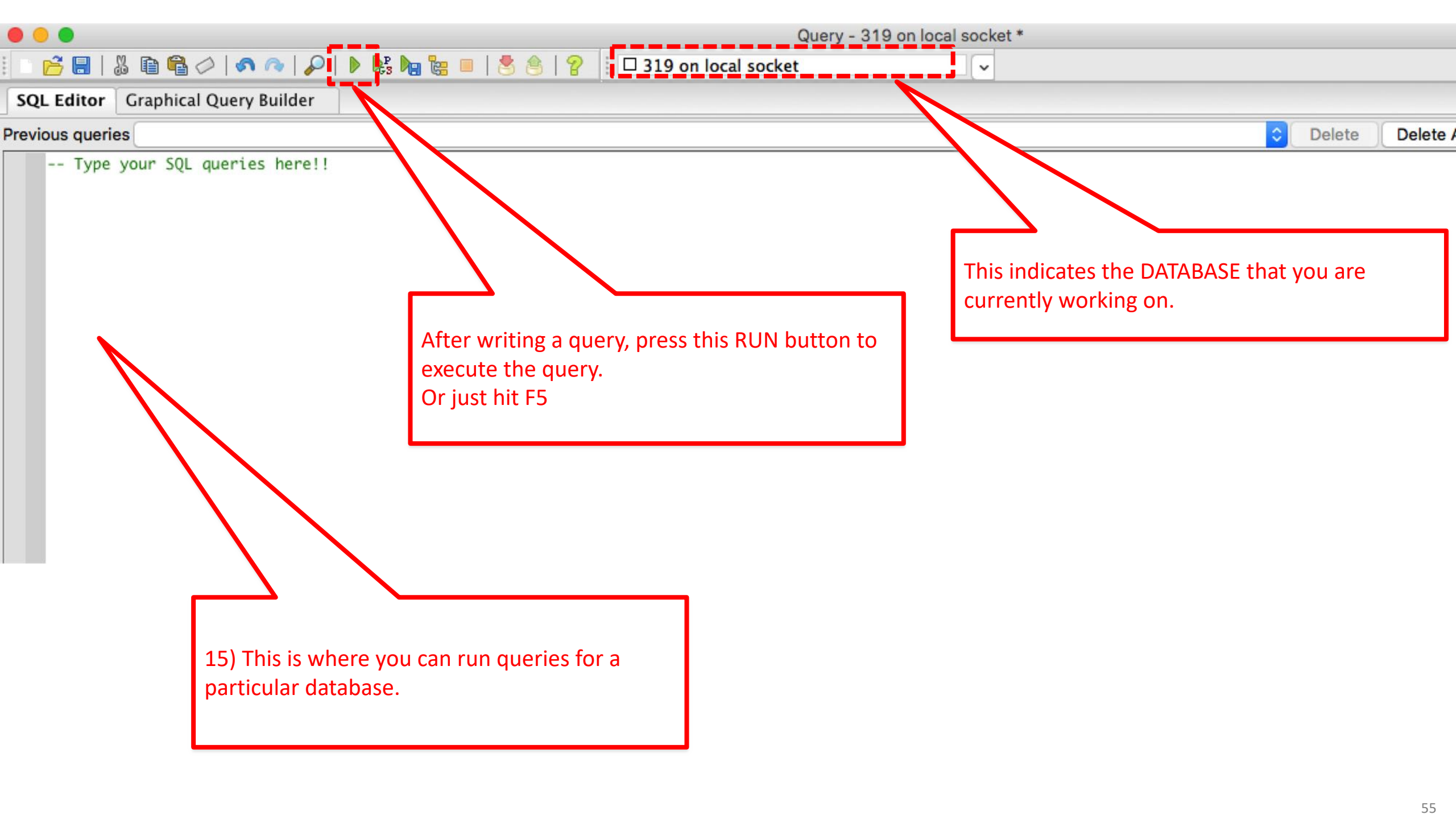
11) To check the Tables in the Database, expand the particular Database.

12) Expand the Tables to check what all tables are present in the database.



13) To open the SQL window to run queries over a particular database, click on the database.

14) Then click on this SQL button.



After writing a query, press this RUN button to execute the query.
Or just hit F5

This indicates the DATABASE that you are currently working on.

15) This is where you can run queries for a particular database.