

Import IMDB database to Postgres

Please read this guide carefully and follow the instructions in order. Before starting these instructions, you must have [installed PostgreSQL](#) and you must have the backup files for the [IMDB database](#). It is recommended to follow the process using the Terminal, since it's the format supported in the labs, but it can also be done through pgAdmin. You can find a section with sample queries below.

Terminal:

Windows

Press **Windows** key and type **cmd**.

Go to PostgreSQL location: Type **cd** and the location where PostgreSQL was installed, the default location is **C:\Program Files\PostgreSQL\the number of version you installed\bin**

Create an empty database where the database will be restored.

```
psql.exe -p 5432 -U postgres -c "CREATE DATABASE imdb"
```

Important! The default port of installation for postgres is **5432, but it might be 5433.*

Type the password from the installation process, and the database will be created.

Then type the following command to restore from the backup file:

```
psql.exe -p 5432 -U postgres --dbname "imdb" -f "C:\<FILE-LOCATION>\imdb_full_wdt.sql"
```

This might take a while (~30 minutes +/- depending on your machine)

Now the databases are ready, you can connect by typing:

- `psql -U postgres -p 5432 -d imdb`

Mac

Press **Command + spacebar** type **Terminal**

Go to PostgreSQL installation location: **/Library/PostgreSQL/the number of version you installed/bin/**

Create an empty database where the database will be restored.

```
psql -p 5432 -U postgres -c "CREATE DATABASE imdb"
```

Important! The default port of installation for postgres is **5432, but it might be 5433.*

Type the password from the installation process, and the database will be created.

Then type the following command to restore from the backup file for the IMDB database:

```
psql -p 5432 -U postgres --dbname "imdb" -f "<FILE-LOCATION>/imdb_full_wdt.sql"
```

Now the databases are ready, you can connect by typing:

- `psql -U postgres -p 5432 -d imdb`

Sample Queries:

Try the following queries in the *imdb* database. Pay attention that the quotes (') in the queries are in the correct format. *Remember that you have to connect with **psql -U postgres** and your password, you can type **\c imdb** to use the database*

Alternatively, use pgadmin (the GUI software which comes with postgresSQL, which is much more convenient).

1. Find the names of all people in the database (**Warning:** remember there are ~ 12 million people in the database, so it's recommended to put a **limit** on the query):

```
SELECT full_name
FROM person;
```

```
SELECT full_name
FROM persons
LIMIT 5;
```

2. Return the start_year of all titles called The Matrix (primary_title)

```
SELECT start_year
FROM titles
WHERE primary_title = 'The Matrix';
```

3. Return the id, original_title and millenium_age (the start_year minus 2000) of all items called The Matrix (primary_title)

```
SELECT title_id, original_title, start_year - 2000 AS millenium_age
FROM titles
WHERE primary_title = 'The Matrix';
```

4. Return all possible professions and order them (ascending)

```
SELECT DISTINCT profession
FROM persons_professions
ORDER BY profession;
```

5. Return the primary_title of all titles that are comedies (10 of them)

```
SELECT primary_title
FROM titles t, titles_genres tg
```

```
WHERE t.title_id = tg.title_id AND tg.genre = 'Comedy'  
(you can also do FROM titles t JOIN titles_genres tg ON t.title_id =  
tg.title_id)  
LIMIT 10;
```