

CompSci 516 Database Systems

SQL Installation

Spring 2022

- Install postgres + Load data + practice SQL queries

Link to the MovieLens data:

<https://www2.cs.duke.edu/courses/spring22/compsci516/DataForClass/>

Install Postgres

- If you are using Mac or Windows, and would like to use the GUI by pgadmin, **follow slide #4-14**
- If you are using Linux, or if you want to use Virtual Machine (VM) on Mac or Windows, **follow slide #15-21**

Postgres GUI Installation and Data Import

Step 1: Follow the link to download PostgreSQL. Choose ~~10.10~~ the most updated version for you platform.

<https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>

PostgreSQL Version	Windows x86-64	Mac OS X	Linux x86-32	Linux x86-64	Windows x86-32
11.5	Download	Download	N/A	N/A	N/A
10.10	Download	Download	Download	Download	Download
9.6.15	Download	Download	Download	Download	Download
9.5.19	Download	Download	Download	Download	Download
9.4.24	Download	Download	Download	Download	Download
9.3.25 (Not Supported)	Download	Download	Download	Download	Download

Step 2: Install components. You need a server, a client pgAdmin4, and CLI tools. You can opt out Stack Builder. Keep your postgres password.

- PostgreSQL Server
- pgAdmin 4
- ~~Stack Builder~~
- Command Line Tools

Please provide a password for the database superuser (postgres).

Password

Retype password

Select the locale to be used by the new database cluster

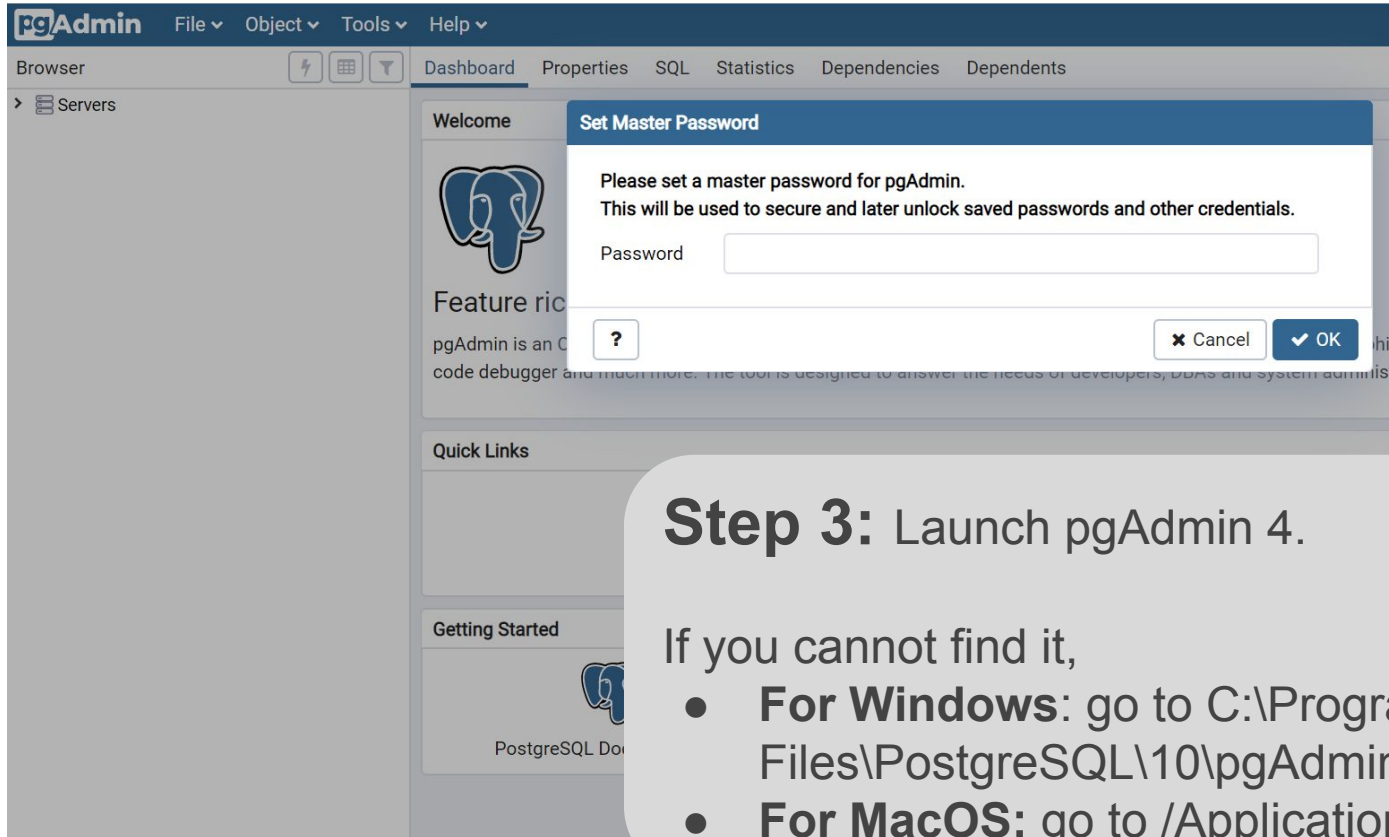
Locale

Launch Stack Builder at exit?

- Stack Builder may be used to download and install additional tools, drivers and applications to complement your PostgreSQL installation.

The following settings will be used for the installation::

Installation Directory: C:\Program Files\PostgreSQL\10
Server Installation Directory: C:\Program Files\PostgreSQL\10
Data Directory: C:\Program Files\PostgreSQL\10\data
Database Port: 5432
Database Superuser: postgres
Operating System Account: NT AUTHORITY\NetworkService
Database Service: postgresql-x64-10
Command Line Tools Installation Directory: C:\Program Files\PostgreSQL\10
pgAdmin4 Installation Directory: C:\Program Files\PostgreSQL\10\pgAdmin 4
Stack Builder Installation Directory: C:\Program Files\PostgreSQL\10



Step 3: Launch pgAdmin 4.

If you cannot find it,

- **For Windows:** go to C:\Program Files\PostgreSQL\10\pgAdmin 4\bin
- **For MacOS:** go to /Applications/PostgreSQL 10



Servers (1)

PostgreSQL 10

Step 4: Click here to connect to your local server.



Feature rich

pgAdmin is an C

code debugger a

Connect to Server

Please enter the password for the user 'postgres' to connect the server - "PostgreSQL 10"

Password

.....



Save Password

Enter the password you set for the user 'postgres'.

Cancel

OK

Quick Links



Add New Server

Getting Started



PostgreSQL Documentation

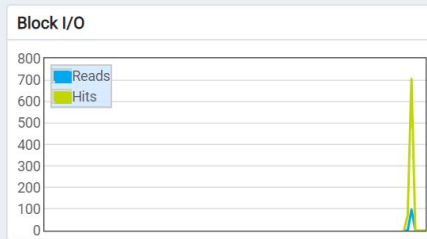
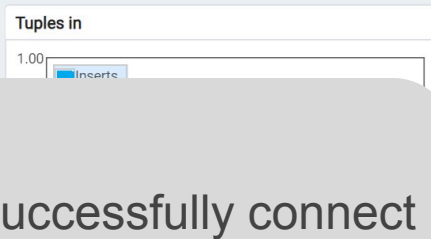
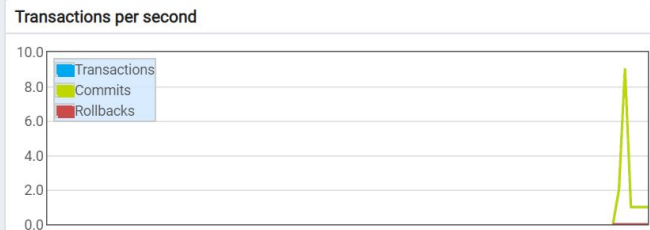
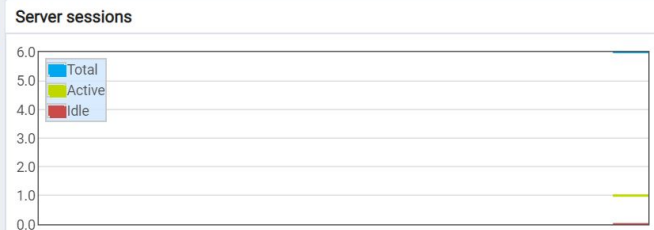


pgAdmin Website



Planet Postg

- Servers (1)
 - PostgreSQL 10
 - Databases
 - Login/Group Roles
 - Tablespaces



Step 5: If you successfully connect to the local server, you will see a group of panels.

Client	Backend start	State	Wait event	Blocking PIDs
	2019-08-22 18:47:25 EDT		Activity: BgWriterMain	
	2019-08-22 18:47:25 EDT		Activity: LogicalLauncherMain	
	2019-08-22 18:47:25 EDT		Activity: WalWriterMain	
13772				
15644	postgres	postgres	pgAdmin 4 - DB:postgres	::1
	2019-08-22 20:09:03 EDT	active		
15840			Activity: CheckpointerMain	
	2019-08-22 18:47:25 EDT		Activity: AutoVacuumMain	
16096				

Step 6: Create a new database called 'movielens'.

The screenshot displays the pgAdmin interface. At the top, the menu bar includes 'File', 'Object', 'Tools', and 'Help'. Below the menu is a toolbar with icons for refresh, grid, and filter. The main window is divided into several panes. On the left, the 'Browser' pane shows a tree view with 'Servers (1)' expanded to 'PostgreSQL 10', which is further expanded to 'Databases (1)'. A context menu is open over 'Databases (1)', with 'Create' selected. The 'Server sessions' pane shows a graph with a single bar for 'Idle' at 6.0. The 'Tuples in' pane shows a graph with bars for 'Inserts', 'Updates', and 'Deletes'. The 'Server activity' pane shows a table with columns for 'Sessions' and 'Locks', and a row with 'PID' 11176. The 'Create - Database' dialog box is open in the foreground, with the 'General' tab selected. The 'Database' field contains 'movielens', and the 'Owner' dropdown is set to 'postgres'. The 'Comment' field is empty. At the bottom of the dialog, there are buttons for 'Cancel', 'Reset', and 'Save'. The bottom right corner of the screenshot shows the date and time: '2019-08-22 18:47:2'.

pgAdmin File Object Tools Help

Browser Dashboard Properties SQL Statistics Dependencies Dependents

Servers (1)
PostgreSQL 10
Databases (2)
movielens

Database sessions

Restore (Database: movielens)

General Restore options

Format Custom or tar

Filename C:\Users\harry\Downloads\movielens.backup

Number of jobs

Role name Select an item...

Cancel Restore

Client Backend start State

:::1	2019-08-22 21:49:50 EDT	active
------	-------------------------	--------

6.1: Select the file you downloaded with the suffix “backup” to restore the table into the database.

Link to the data:
<https://www2.cs.duke.edu/courses/fall19/compsci516/DataForClass/>

Restore (Database: movielens)

General **Restore options**

Sections

Pre-data

Yes

Data

Yes

Post-data

No

Type of objects

Only data

No

Do not save

Owner

Yes

Privilege

No

Tablespace

No

6.2: Switch on 'Pre-data' (schema), 'Data' and 'Owner' options in the "Restore options" tab.

i

?

✕ Cancel

Restore

- PostgreSQL 10
 - Databases (2)
 - movielens
 - Cast
 - Catalogs
 - Event Triggers
 - Extensions
 - Foreign Data Wrappers
 - Languages
 - Schemas (1)
 - public
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Sequences
 - Tables (4)
 - movies
 - ratings
 - ~~universal table~~
 - users
 - Trigger Functions
 - Types
 - Views
 - postgres
 - Login/Group Roles
 - Tablespaces

```
1 SELECT * FROM public.movies
2
```

Step 7: Now you should be able to view the data if you select the table and click the button “View Data”.

There are three tables: “movies”, “ratings” and “users”

	movie_id integer	movie_title text	release_date character varying (40)	video_release_date character varying (40)	IMDb_URL text	unknown boolean	Action boolean	Adventure boolean	Animation boolean	Children's boolean	Comedy boolean	Crime boolean	Documentary boolean	Drama boolean	Fantasy boolean	Film-Noir boolean
1	1	Toy Story (19...	01-Jan-1995	[null]	http://us.imd...	false	false	false	true	true	true	false	false	false	false	false
2	2	GoldenEye (1...	01-Jan-1995	[null]	http://us.imd...	false	true	true	false	false	false	false	false	false	false	false
3	3	Four Rooms (...)	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	false	false	false	false	false
4	4	Get Shorty (1...	01-Jan-1995	[null]	http://us.imd...	false	true	false	false	false	true	false	false	true	false	false
5	5	Copycat (199...	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	true	false	true	false	false
6	6	Shanghai Tri...	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false
7	7	Twelve Monk...	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false
8	8	Babe (1995)	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	true	true	false	false	true	false	false
9	9	Dead Man W...	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false
10	10	Richard III (1...	22-Jan-1996	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false
11	11	Seven (Se7en...	01-Jan-1995	[null]	http://us.imd...	false	false	false	false	false	false	true	false	false	false	false
12	12	Usual Suspec...	14-Aug-1995	[null]	http://us.imd...	false	false	false	false	false	false	true	false	false	false	false
13	13	Mighty Aphro...	30-Oct-1995	[null]	http://us.imd...	false	false	false	false	false	true	false	false	false	false	false
14	14	Postino, Il (19...	01-Jan-1994	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false
15	15	Mr. Holland's...	06-Jan-1996	[null]	http://us.imd...	false	false	false	false	false	false	false	false	true	false	false

Browser



Servers (1)

localhost

Databases (4)

dblp

movielens

Casts

Catalogs

Event Triggers

Extensions

Foreign Data Wrapper

Languages

Schemas (1)

public

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Sequences

Tables (4)

movies

Query Tool

Reload Configuration

Pause Replay of WAL

Resume Replay of WAL

Add Named Restore Point...

Import/Export...

Maintenance...

Backup...

Backup Globals...

Backup Server...

Restore...

Grant Wizard...

Step 8: Click on “Query Tool” to start writing your own SQL queries!

es SQL Statistics Dependencies Dependents public.movies/... movielens/yuchao@localhost *



o@localhost

History

Click here to run

Data Output Explain Messages Notifications

Postgres Installation and data import on VM

Reserve a virtual machine

1. Go to <https://vcm.duke.edu>
2. Click “Reserve a VM”
3. Login with netid and password

community with easy access to virtual software packages, and semester, host your own server for development projects and coursework,



Virtual Machines (aka VMs)

Your Duke VM is like having a second computer that lives in OIT. You can log into and use your VM from your own machine.

- Run Windows or Linux
- Install zero, one or multiple apps for free

[Reserve a VM](#)

vm-manager-help@duke.edu for assistance

Reserve a virtual machine (Cont'd)

1. Click the drop down and select "Ubuntu 18.04" from the list
2. Read (or not) the agreement, agree and continue

New Virtual Machine Rese

Application and Operating System

✓ Please select

Plain VM: No Apps

RHEL 7

Ubuntu16.04

Ubuntu18.04

Windows 10

Linux AppStacks

COMPSCI 216 - Everything Data

Lamp Stack

Linux Matlab

Windows AppStacks

ArcGIS Desktop & Pro

Bio202 AppStack

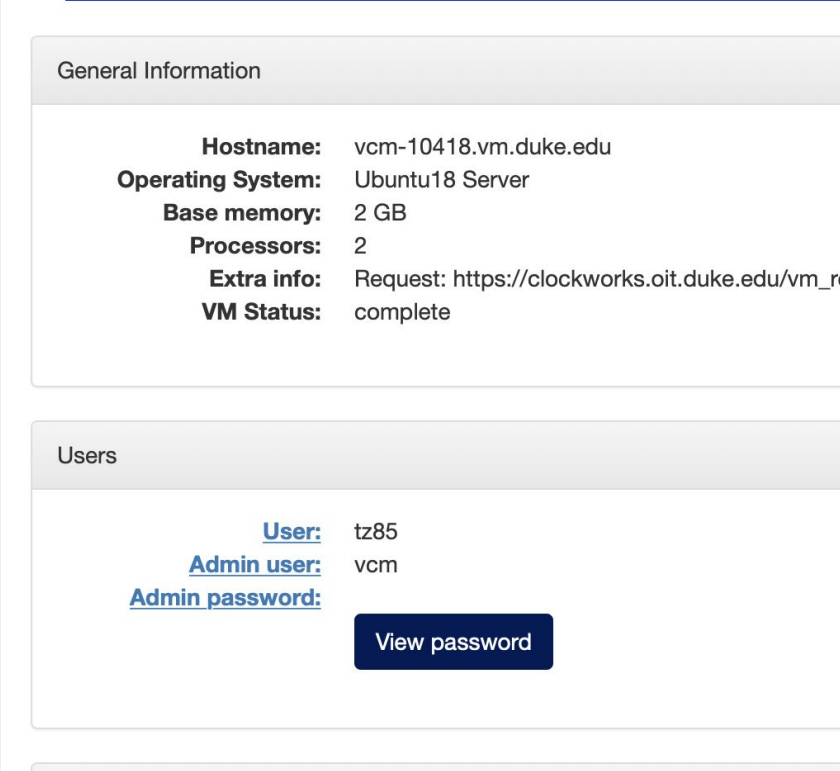
Bio212 AppStack

Bio201

Access your virtual machine

SSH into your new virtual machine using the hostname and your netid

```
$ ssh <netid>@<hostname>
```



The screenshot displays a web-based interface for managing virtual machines. It is divided into two main sections: 'General Information' and 'Users'. The 'General Information' section lists the following details: Hostname (vcm-10418.vm.duke.edu), Operating System (Ubuntu18 Server), Base memory (2 GB), Processors (2), Extra info (Request: https://clockworks.oit.duke.edu/vm_re...), and VM Status (complete). The 'Users' section lists the User (tz85), Admin user (vcm), and Admin password, with a 'View password' button next to the password field.

General Information	
Hostname:	vcm-10418.vm.duke.edu
Operating System:	Ubuntu18 Server
Base memory:	2 GB
Processors:	2
Extra info:	Request: https://clockworks.oit.duke.edu/vm_re...
VM Status:	complete

Users	
User:	tz85
Admin user:	vcm
Admin password:	<input type="password"/> View password

Installing Postgres on Ubuntu

1. Update your machine

- `sudo apt update`

2. Install latest version of psql

- `sudo apt install postgresql postgresql-contrib`

```
dpkg-query -f='${Package} ${Version} ${Architecture}\n'
```

The following NEW packages will be installed:
libpq5 libsensors4 postgresql postgresql-10 postgresql-client-10
postgresql-client-common postgresql-common postgresql-contrib ssl-cert
sysstat
0 upgraded, 10 newly installed, 0 to remove and 0 not upgraded.
Need to get 5,339 kB of archives.
After this operation, 21.1 MB of additional disk space will be used.
Do you want to continue? [Y/n]

```
syncing data to disk ... OK
```

Success. You can now start the database server using:

```
    /usr/lib/postgresql/10/bin/pg_ctl -D /var/lib/postgresql/10/main -l logfile  
start
```

Ver	Cluster	Port	Status	Owner	Data directory	Log file
10	main	5432	down	postgres	/var/lib/postgresql/10/main	/var/log/postgresql/postgresql-10-main.log

Create database and download dataset

1. Switch user to postgres

```
sudo su - postgres
```

2. Create database “movielens”

```
createdb movielens
```

3. Download dataset

```
wget
```

```
https://www2.cs.duke.edu/courses/fall19/compsci516/DataForClass/movielens.sql
```

```
tz85@vcm-10418:~$ sudo su - postgres
postgres@vcm-10418:~$ createdb movielens
postgres@vcm-10418:~$ wget "https://www2.cs.duke.edu/courses/fall19/compsci516/DataForClass/movielens.sql"
--2019-08-26 08:07:23-- https://www2.cs.duke.edu/courses/fall19/compsci516/DataForClass/movielens.sql
Resolving www2.cs.duke.edu (www2.cs.duke.edu)... 152.3.140.31
Connecting to www2.cs.duke.edu (www2.cs.duke.edu)|152.3.140.31|:443... connected
HTTP request sent, awaiting response... 200 OK
Length: 19636845 (19M) [application/x-sql]
Saving to: 'movielens.sql'

movielens.sql      100%[=====>] 18.73M  85.8MB/s   in 0.2s

2019-08-26 08:07:24 (85.8 MB/s) - 'movielens.sql' saved [19636845/19636845]

postgres@vcm-10418:~$ ls
10 movielens.sql
postgres@vcm-10418:~$
```

Load dataset into the new database

1. Create tables from sql file

```
psql -U postgres movielens <
movielens.sql
```

2. Connect to database

```
psql movielens
```

3. List all tables

```
\dt
```

```
[postgres@vcm-10418:~$ psql movielens
psql (10.10 (Ubuntu 10.10-0ubuntu0.18.04.1))
Type "help" for help.
```

```
[movielens=# \dt
```

List of relations			
Schema	Name	Type	Owner
public	movies	table	postgres
public	ratings	table	postgres
public	users	table	postgres

**Try queries similar to lecture slides on the
MovieLens dataset**

Schema

movie

- movie_id
- movie_title
- release_date
- Action
- Adventure
- Animation
-

ratings

- movie_id
- user_id
- rating
- timestamp

users

- user_id
- age
- gender
- occupation
- zipcode

Sample query:

```
SELECT AVG(ratings)
FROM movie M, ratings R
WHERE M.movie_id = R.movie_id
```

Note: for the genre columns in movies, like Action, Adventure, etc., use “..” in queries, e.g.,:

```
SELECT “Action”
FROM movies
```

Postgres converts all unquoted identifiers to lower case:

<https://stackoverflow.com/questions/37910287/sql-hint-to-reference-a-column>