Tech Session 1: Getting Set Up

01 September 2021

Agenda

1. Introduction to course infrastructure
2. CMU VPN
3. SSH access to course server
4. Access to course database
5. GUI access to course database (via SSH tunneling)
6. Access to course github

INFRASTRUCTURE OVERVIEW

Why Remote Infrastructure?

1. Confidentiality
2. Computing Power and Scalability
3. Collaboration
4. Learning New Tools

Reminder: Data Confidentiality

1. Do not try to download data
2. Inform course instructors if you accidentally download data or your computer or key is compromised!
3. Get comfortable using the database!

Class Infrastructure Elements:

Your Laptop

GitHub Server

VPN

Amazon Web Services

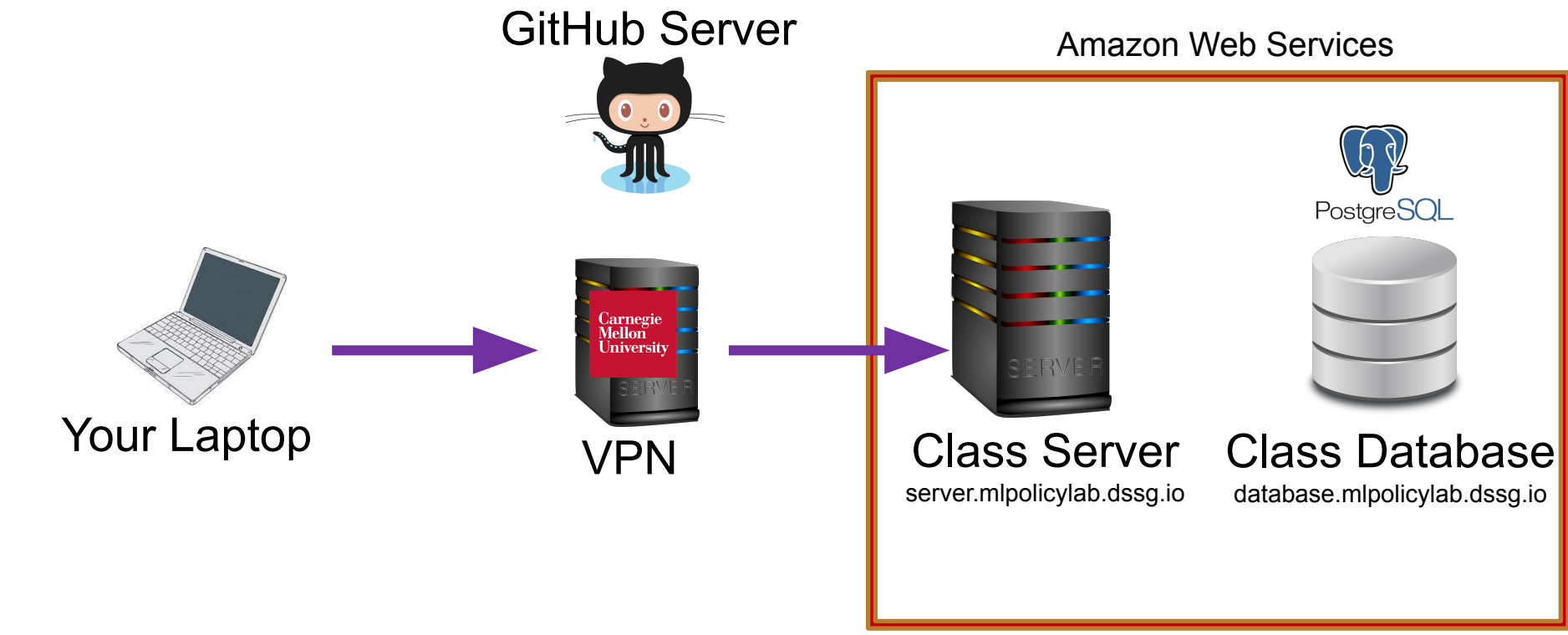
Class Server

server.mlpolicylab.dssg.io

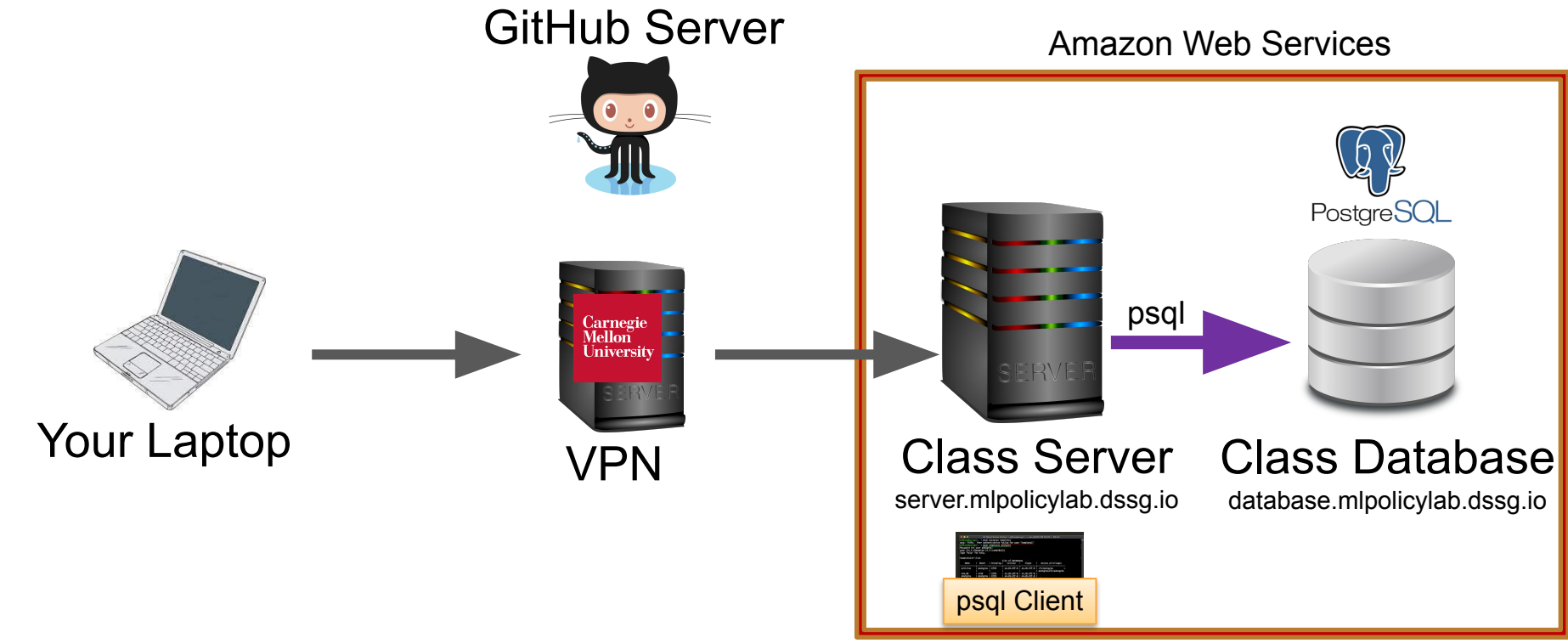
Class Database

database.mlpolicylab.dssg.io

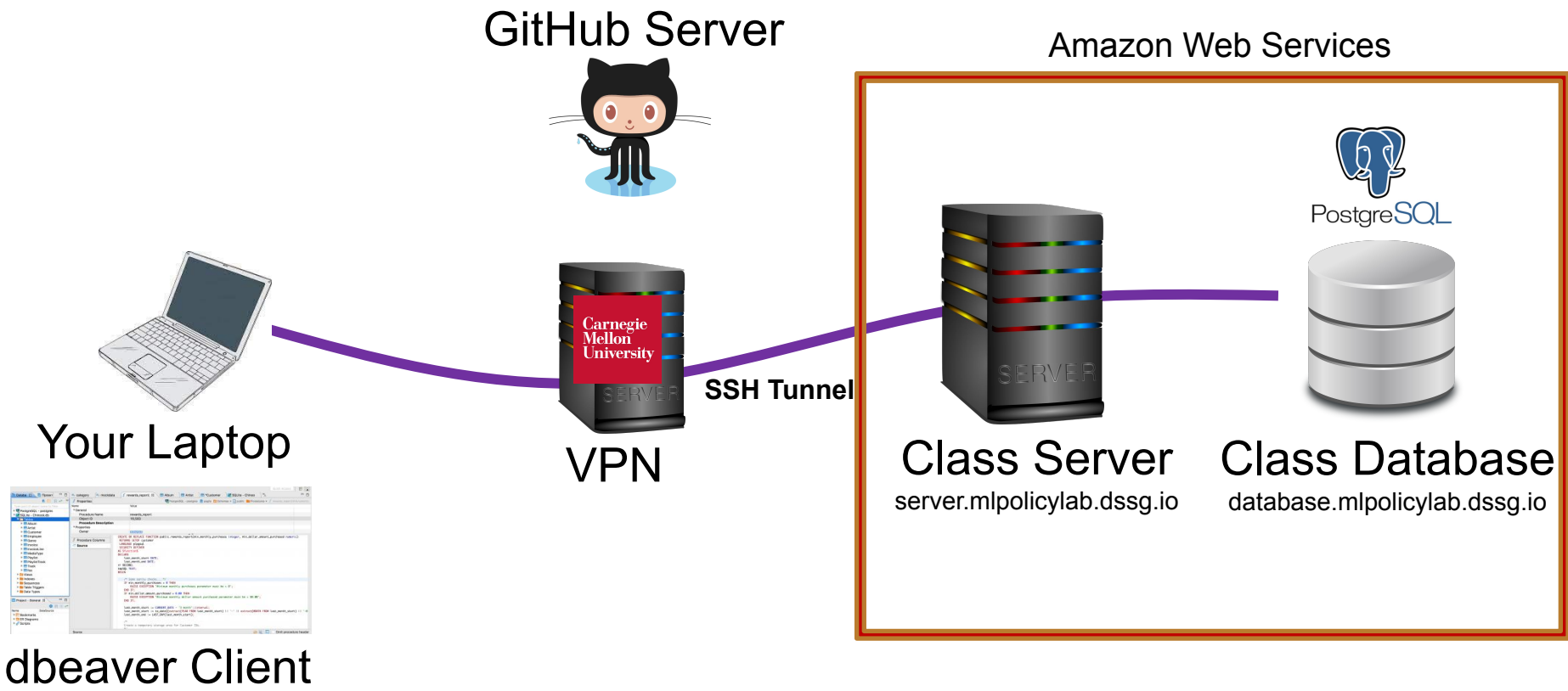
Class Infrastructure Elements:
When you ssh to server.mlpolicylab.dssg.io



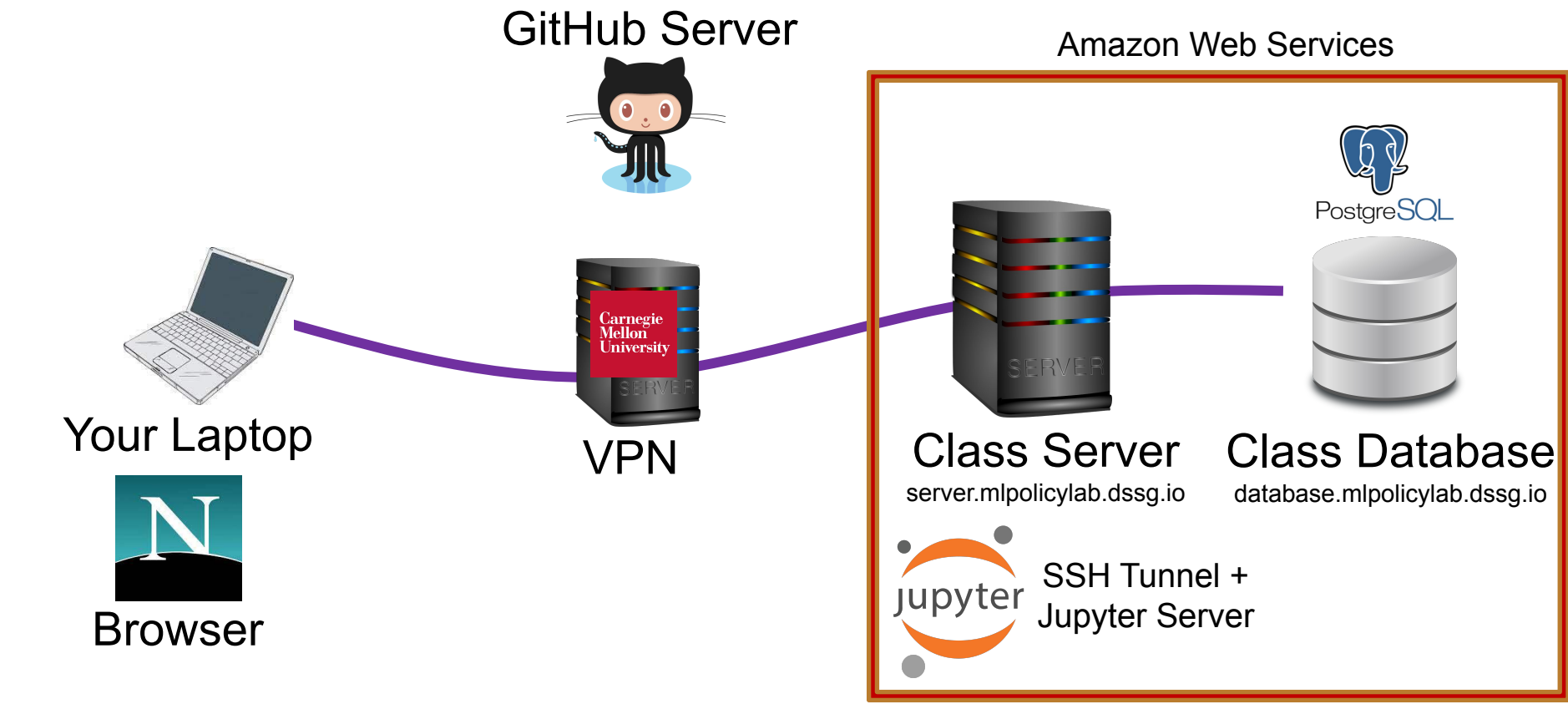
Class Infrastructure Elements:
When you use psql from the class server



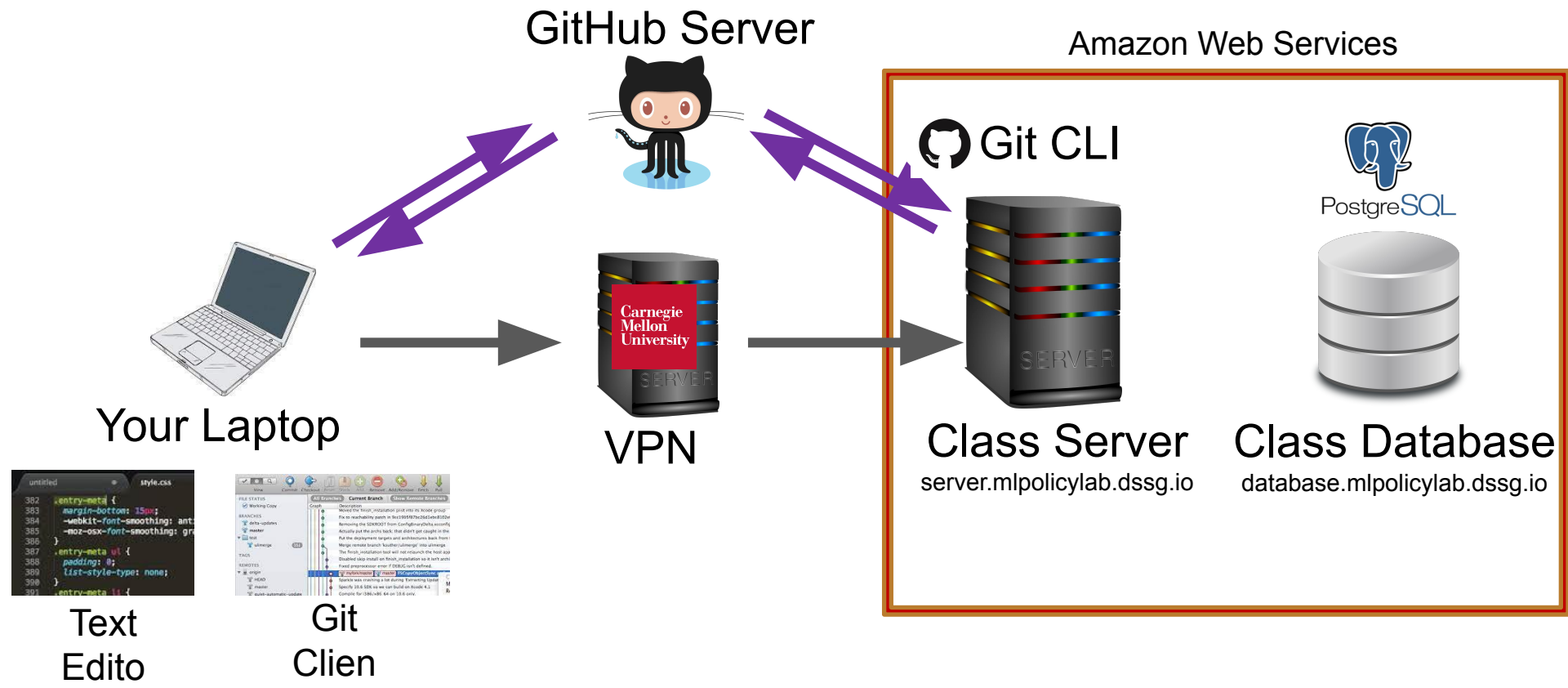
Class Infrastructure Elements:
When you use dbeaver to connect to the database



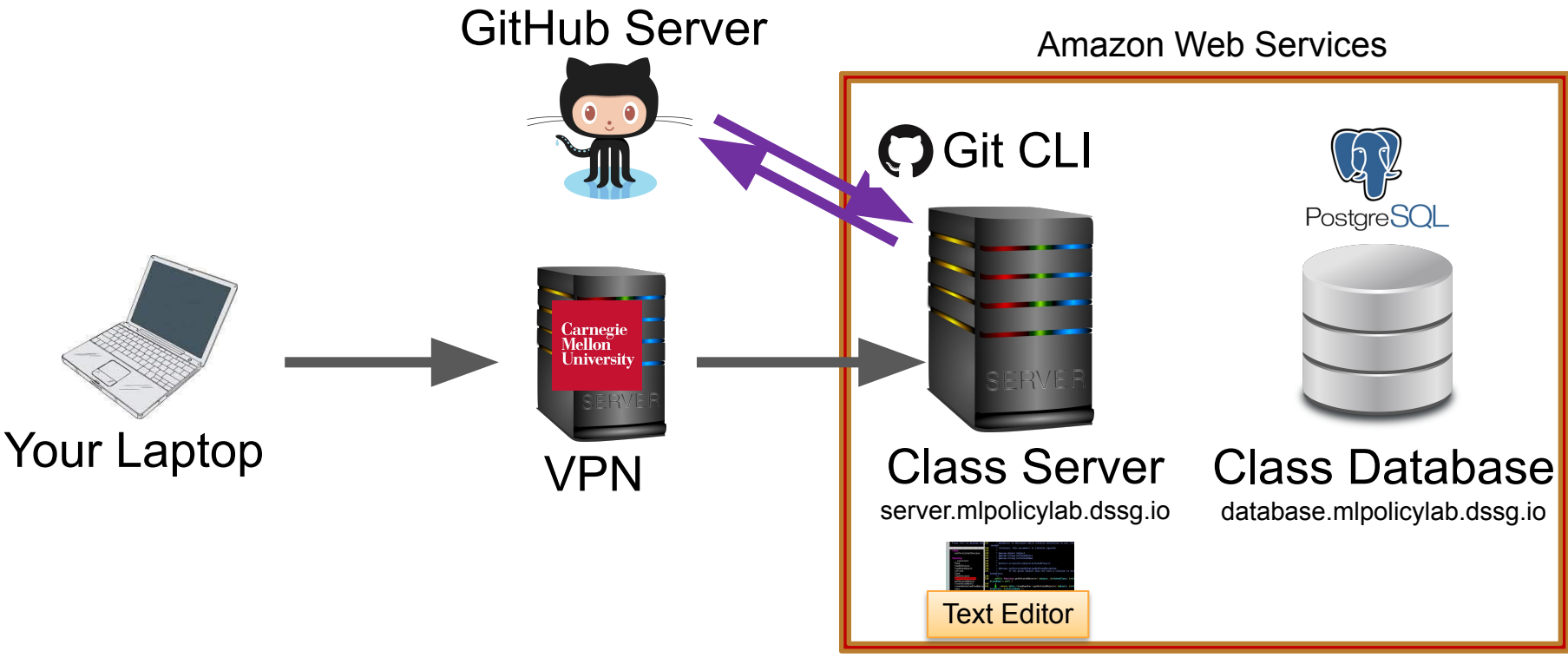
Class Infrastructure Elements: When you use Jupyter



Class Infrastructure Elements:
When edit code with a *local* text editor



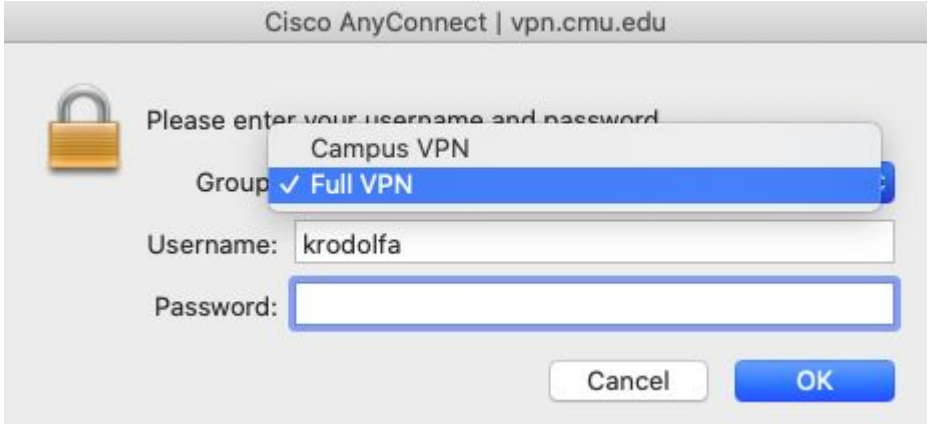
Class Infrastructure Elements:
When edit code with a *remote* text editor



CMU VPN

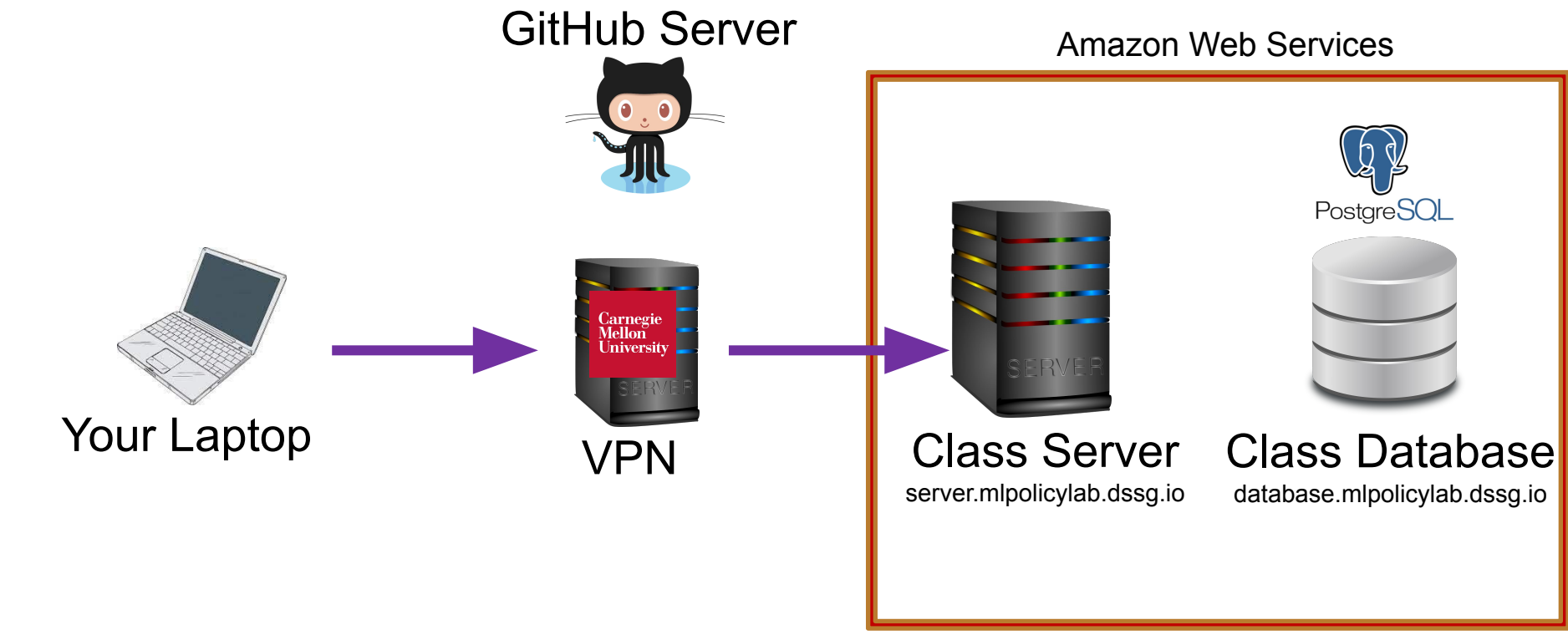
CMU VPN: For Connecting Off Campus

1. Download Cisco Anyconnect VPN client from here and install
2. Open Anyconnect client
3. Enter login credentials:
   - Connect to: vpn.cmu.edu
   - Group: “Full VPN”
   - Username: your andrewid
   - Password: your CMU password
4. Click connect

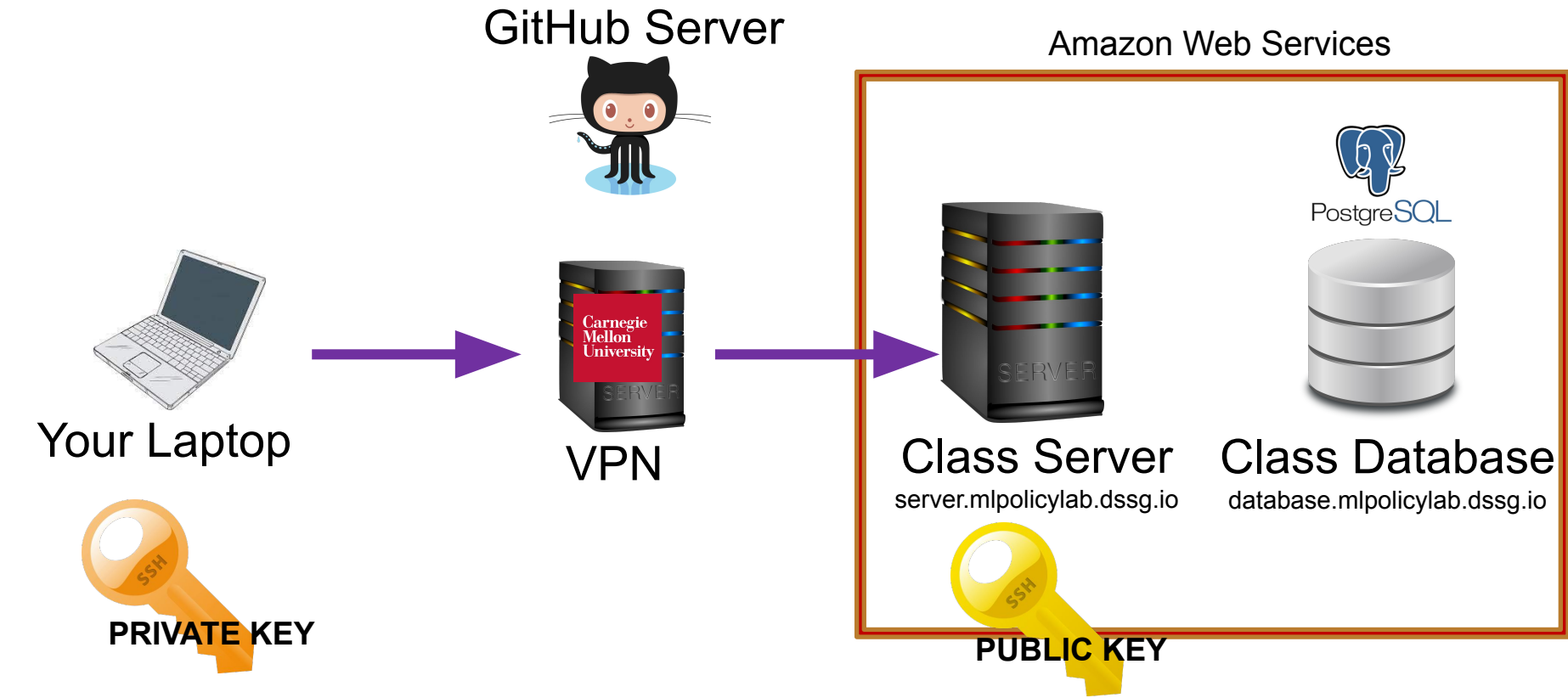


SSH ACCESS

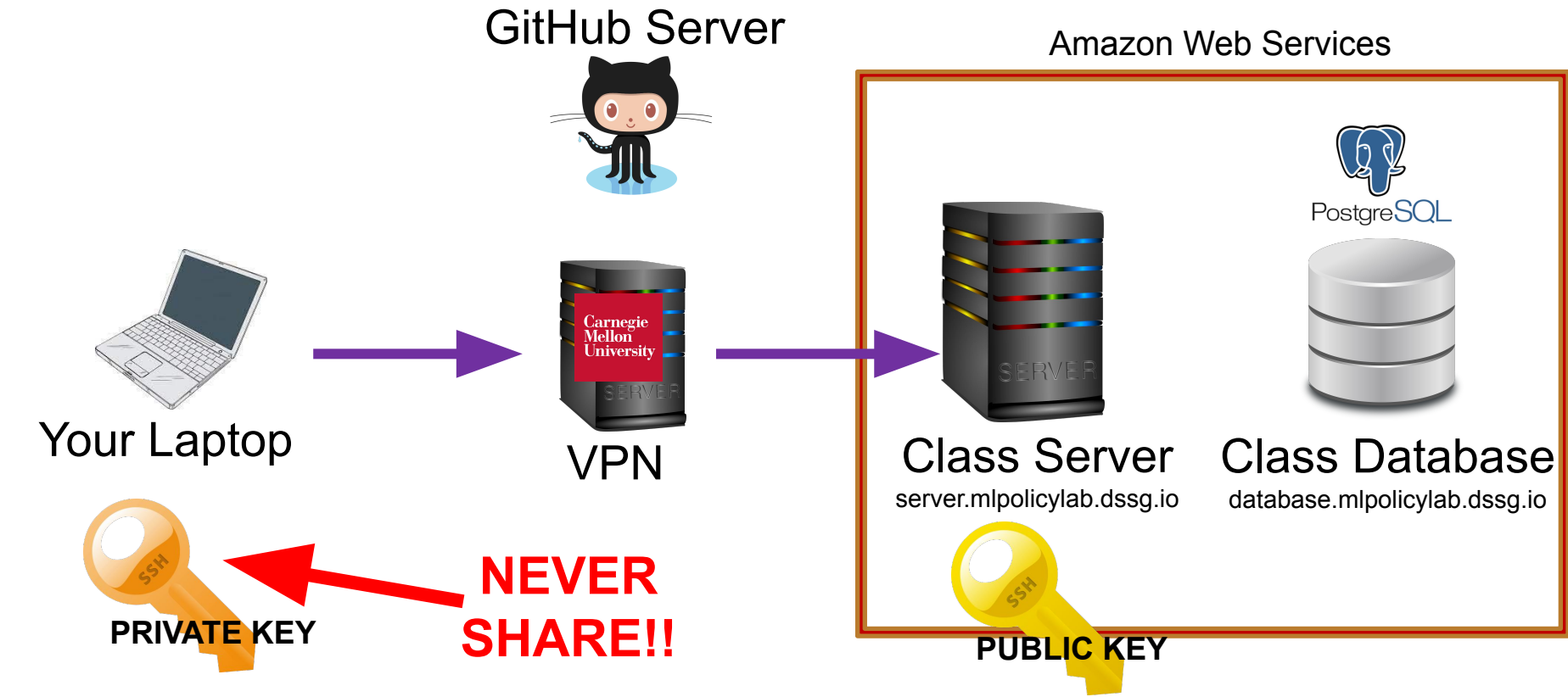
Class Infrastructure Elements:
When you ssh to server.mlpolicylab.dssg.io



Class Infrastructure Elements:
When you ssh to server.mlpolicylab.dssg.io



Class Infrastructure Elements:
When you ssh to server.mlpolicylab.dssg.io



Connecting

Windows PowerShell or Mac/*nix Terminal:

```
ssh {andrew_id}@server.mlpolicylab.dssg.io -i /path/to/your/private_key_file
```

Older versions of Windows: PuTTY or MobaXTerm

Were You Able to Connect?

Try Some Commands:

```
hostname
```

or

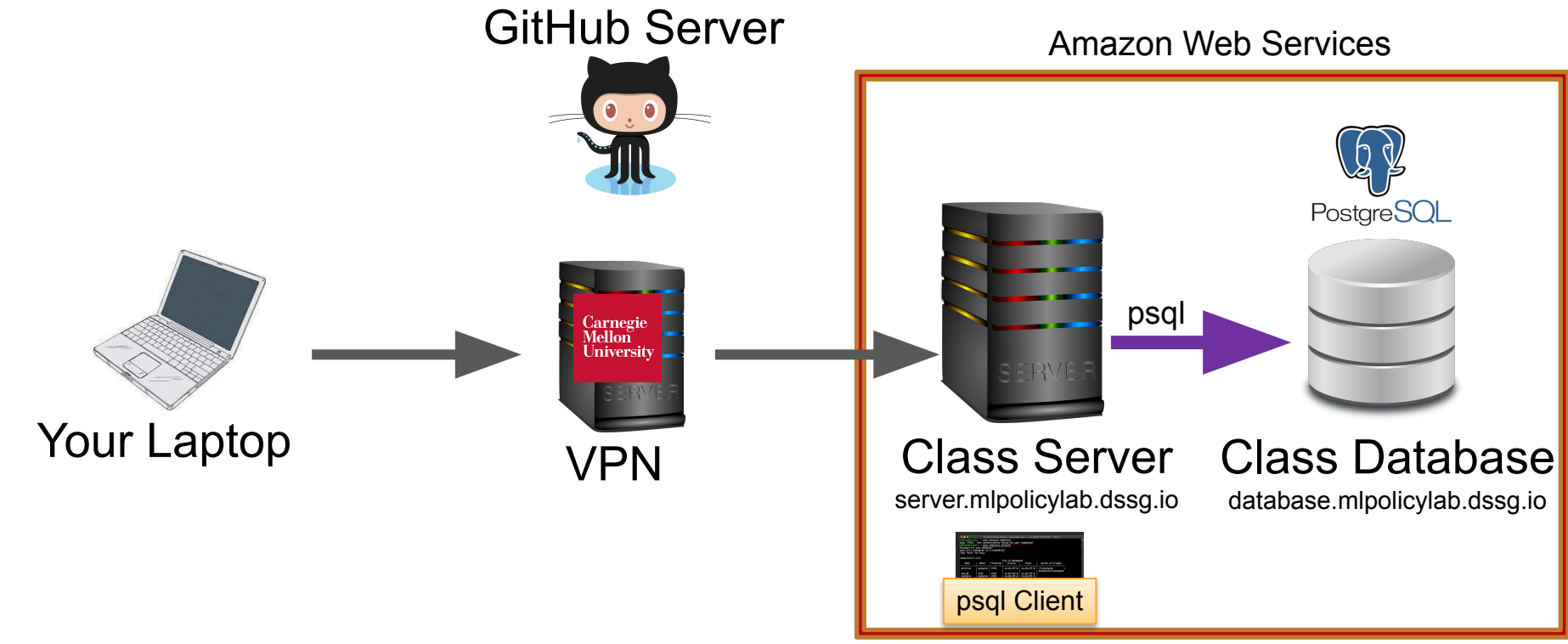
```
ec2metadata | grep "security-groups"
```

Not Familiar with the Linux CLI?

Don’t worry—we’ll walk through some basics in a future tech session...

DATABASE ACCESS
via psql

Class Infrastructure Elements:
When you use psql from the class server



Where is my password?

In a private file called **.pgpass** (note the “.”):

```
cat ~/.pgpass
```

should give something like:

```
database.mlpolicylab.dssg.io:5432:*:{andrew_id}:{YOUR_PASSWORD}
```

Connecting to the database

We'll first connect with the psql CLI just to be sure you can (type this all on one line):

```
PAGER='less -S' psql -h database.mlpolicylab.dssg.io -U
{YOUR_ANDREW_ID} group_students_database
```

Should give you a database prompt that looks like:

```
psql (11.6 (Ubuntu 11.6-1.pgdg18.04+1), server 11.5)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384,
bits: 256, compression: off)
Type "help" for help.

group_students_database=>
```

Were you able to connect?

Try some commands:

```
SELECT 1+1 AS foo;
```

or

```
SELECT CURRENT_USER;
```

To exit the psql client, type: `\q`

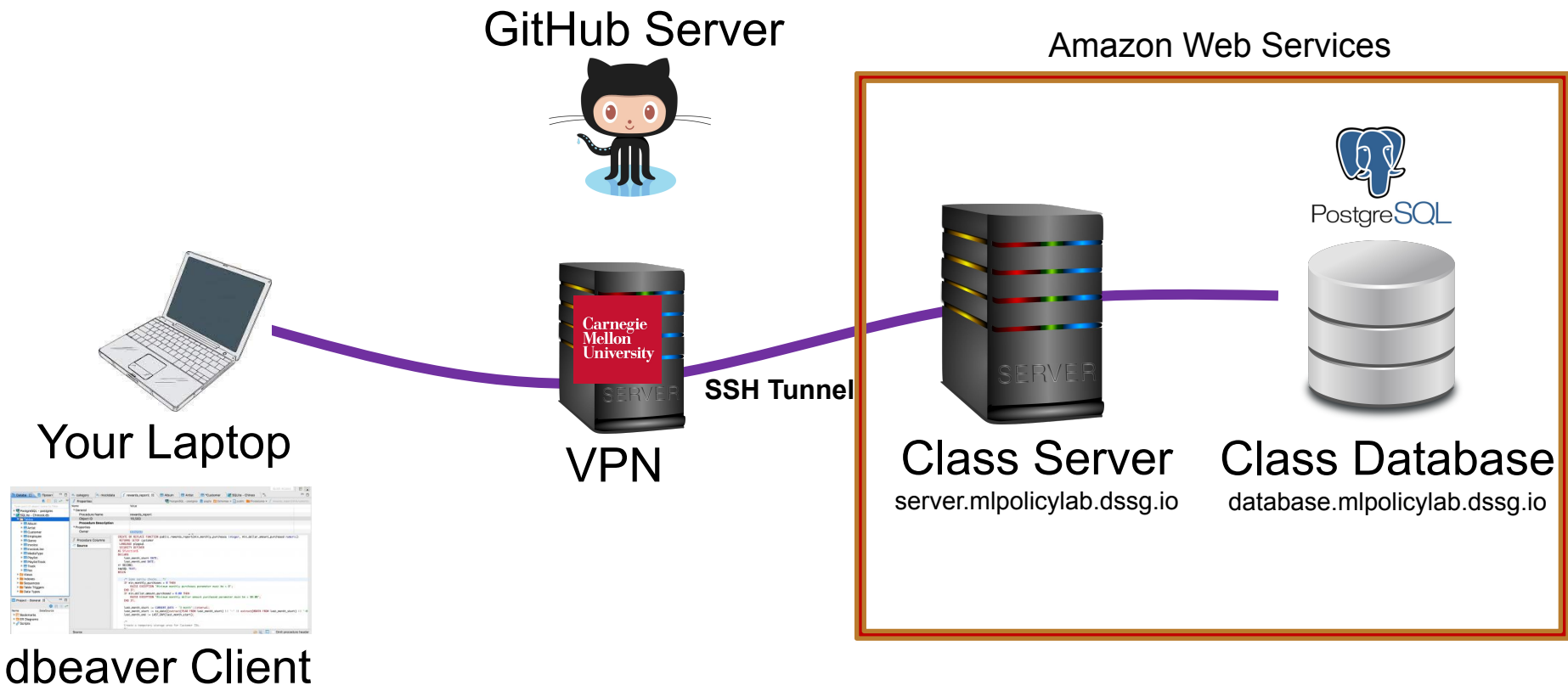
Reminder for Assignment!

This is the database server you'll be connecting to for your project

It is a **different server** than the one you're using for the data loading assignment (see the assignment for connection info) -- to keep things easier, that server can be accessed directly from your laptop!

DATABASE ACCESS
via DBeaver GUI

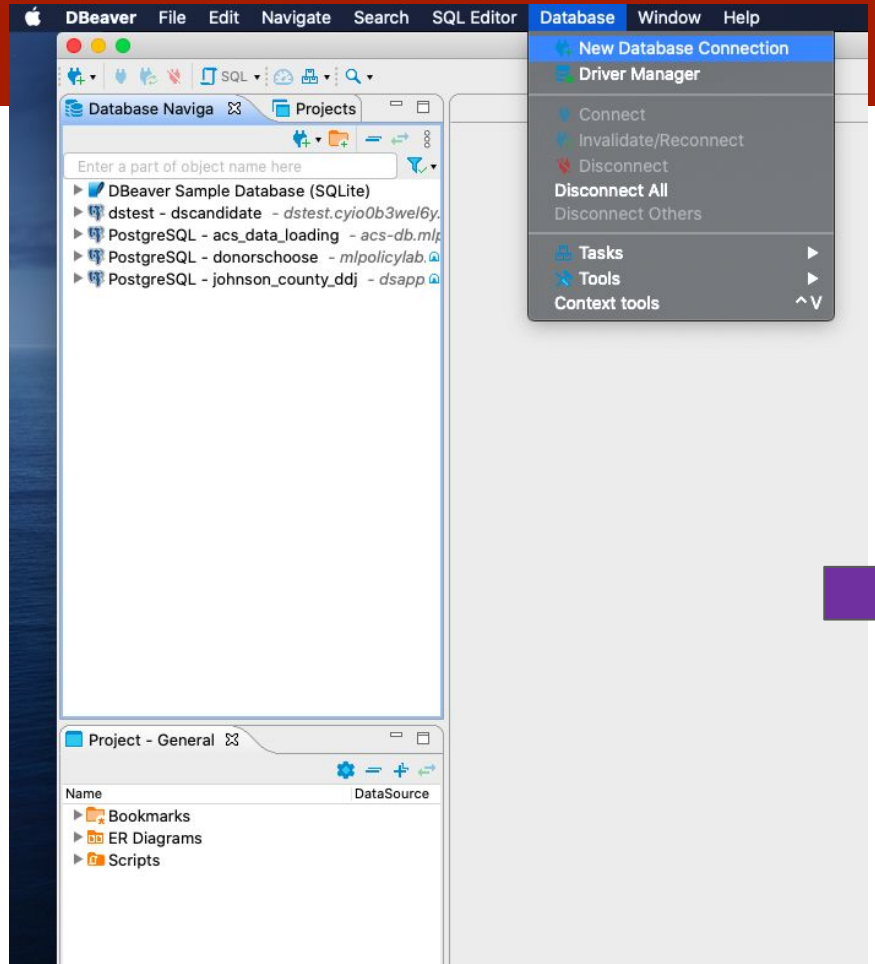
Class Infrastructure Elements:
When you use dbeaver to connect to the database

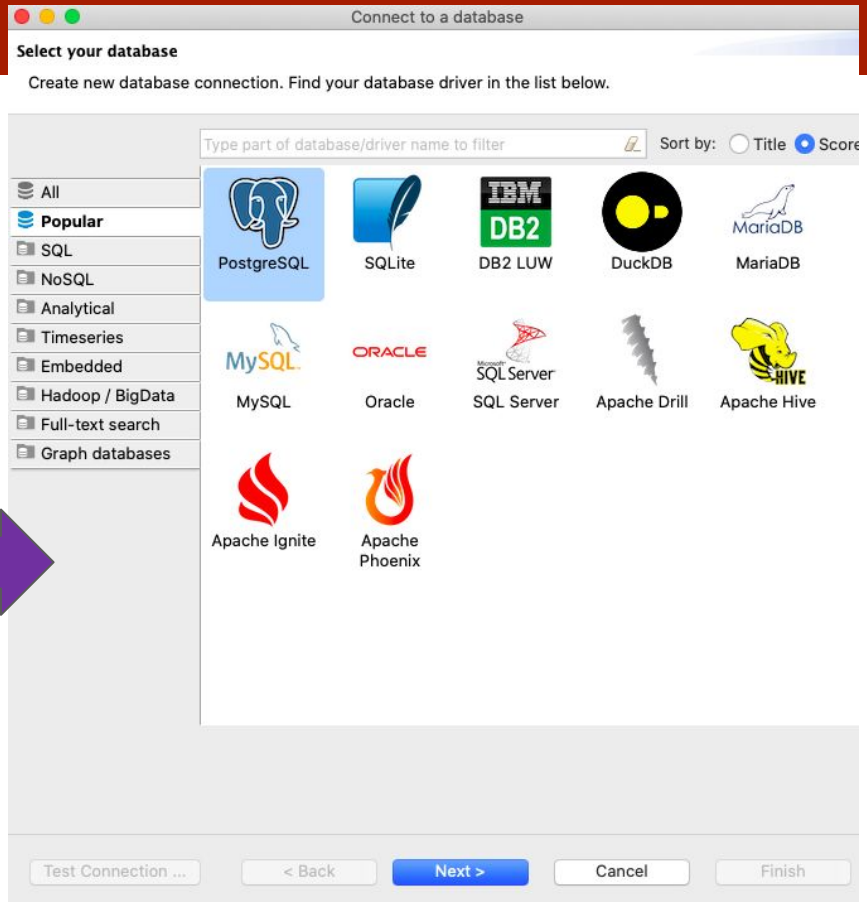


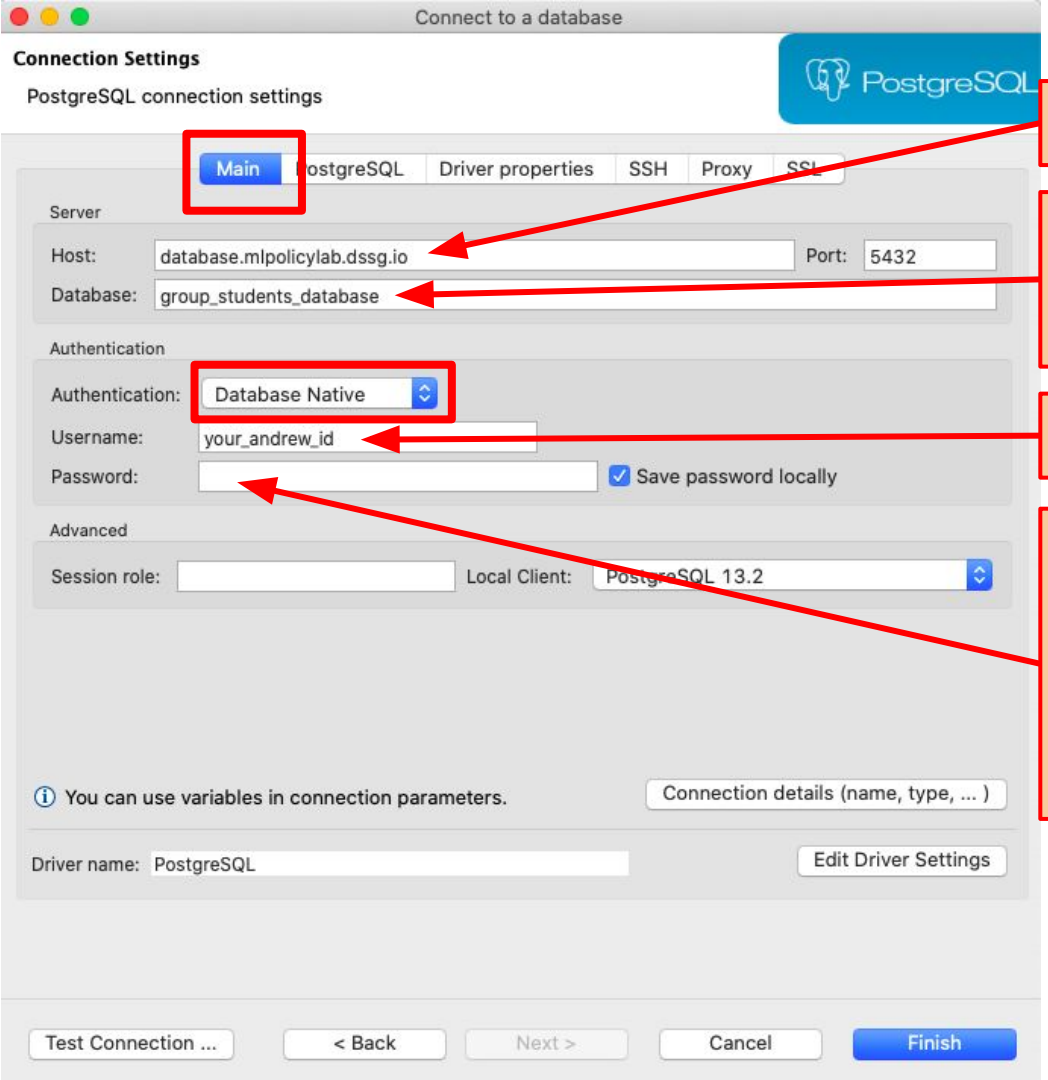
Getting DBeaver

- https://dbeaver.com/edition/community/
- Install the community edition (because it’s free)
- Install any drivers it’s asking you to install

Create a new connection and select postgres





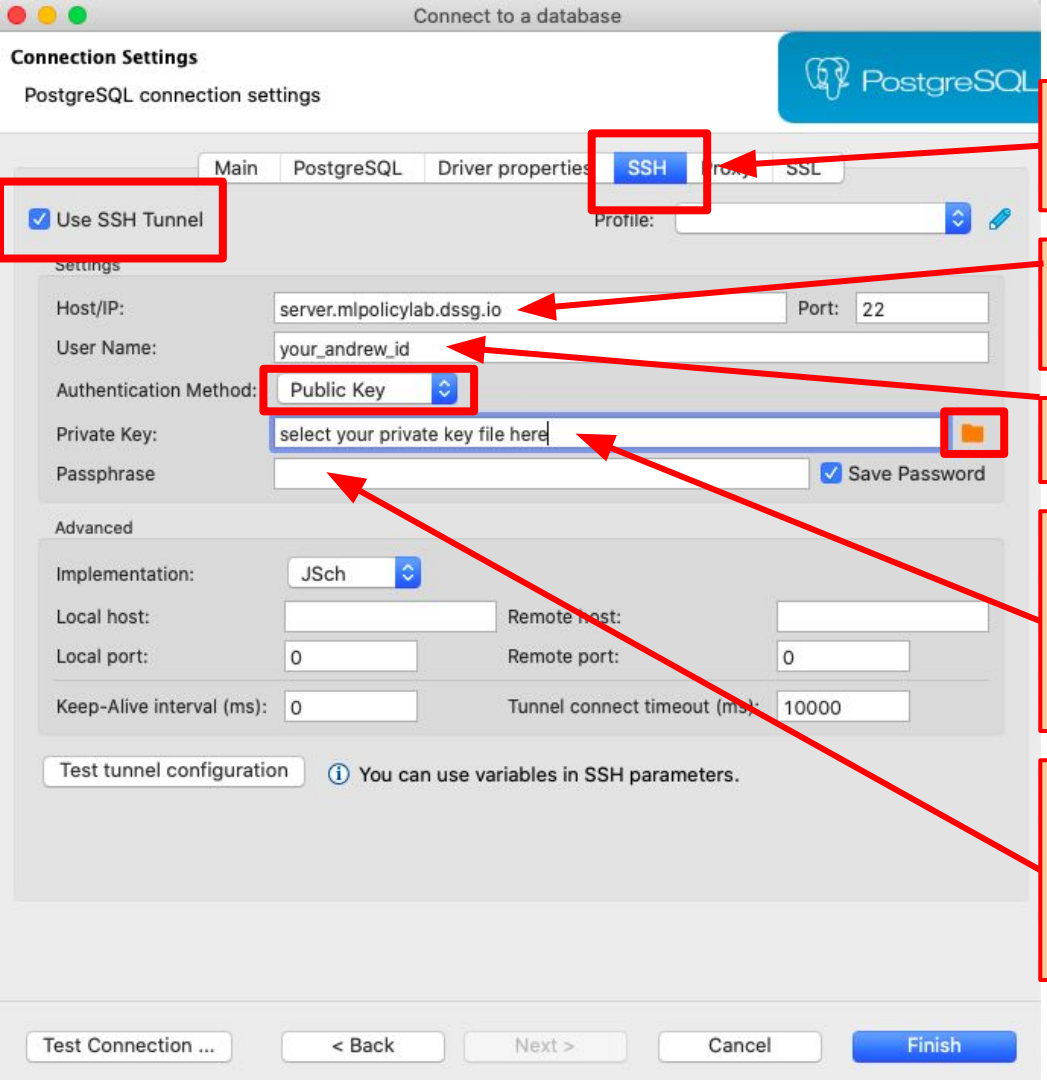


This is the *database server* we are using all semester

This is the test database name for now. Your project will have a different database name so you'll need to replace this later.

Enter your andrew id here

This is your *database* password. We have stored it in a file called **.pgpass** and placed it in your directory on the server - it will be a long string with lots of words combined. When you log on to the server through ssh, type cat .pgpass to see it (it will start after your andrew id in .pgpass, separated by a colon)



You need to “tunnel” through the class server to get to the database so this tab needs your server information

This is the *class server* we are using. – same as the one you ssh’ed to earlier

Enter your andrew id here

You can directly enter the path to your private key file or navigate to it using the folder icon. This will typically be ~/.ssh/id_rsa for mac/linux or C:\Users\username\.ssh\id_rsa for windows

When you generated your ssh key, you may have entered a passphrase to protect it. If you did, you will need to enter it here. If you left it blank, you can leave it blank here.

Debugging

If you get an error like invalid privatekey: [B@7696c31f:

- You may need to install a different SSH package called "SSHJ" -- under the Help menu, choose "Install new Software" then search for SSHJ and install the package (you'll need to restart dbeaver). After restarting, choose “SSHJ” in the drop-down under advanced (should be labeled either ”Implementation” or “Method”) when setting up the tunnel

- If that doesn’t work, you can try changing the format of your private key: go to the directory where your private key is and type ssh-keygen -p -m PEM -f id_rsa and try dbeaver again

Were you able to connect?

Try some commands:

```
SELECT 1+1 AS foo;
```

or

```
SELECT CURRENT_USER;
```

GITHUB ACCESS
via Web GUI

Accessing the Course Github

- We'll set up programmatic access on the server later. Right now, just want to make sure you can see a test **private** repo
- In your browser, try accessing: https://github.com/dssg/test-mlpolicylab-private