

Introduction

Michael Baker
INTERNATIONAL

Tidal Basin: Access and Run the Python Data Extraction Jupyter Notebook

Introduction

In this tutorial, we will cover the steps necessary to access the Tidal Basin EC2 machine in AWS and navigate to the Python Data Extraction jupyter notebook. The notebook will walk you through the steps of creating an SQL query, connecting to the PostgreSQL database hosted on in AWS RDS, converting that query to a pandas table, then uploading that data to an S3 Bucket.

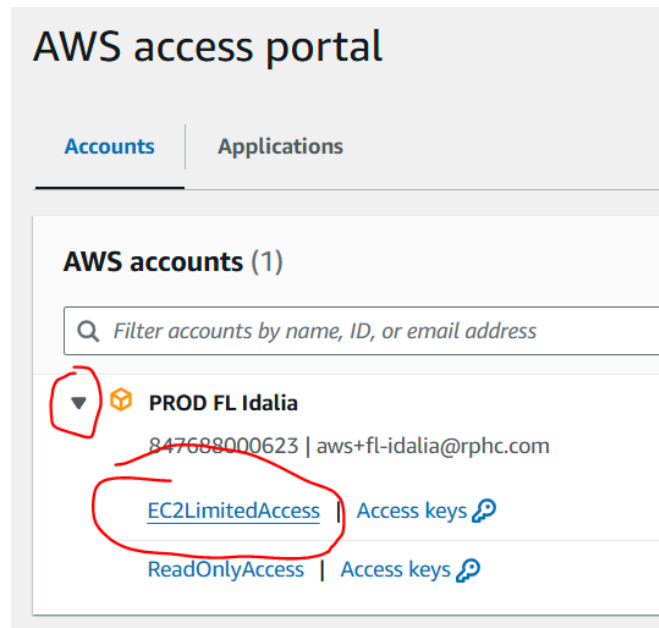
Table of Contents

1. Access the EC2 Environment Through AWS
2. Open Anaconda and Select the Tidal Basin environment
3. Open Visual Code
4. Open the TB-Python-SQL-Pull Folder
5. Open the Data Extraction Jupyter Notebook

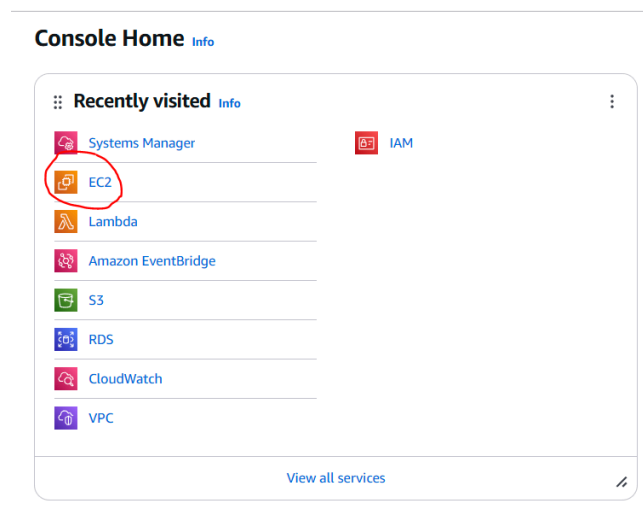
1: Access the EC2 Environment Through AWS

Logon to the AWS EC2LimitedAccess Environment:

- Go to the AWS Access Portal [here](#)
- Select the drop down arrow and select **EC2LimitedAccess**



- On the main landing page, select **EC2**



- Select **Instances (running)**

Resources

You are using the following Amazon EC2

<u>Instances (running)</u>	4
Dedicated Hosts	0
Key pairs	4
Security groups	23

- Select the instance labeled **tbt-ec2-db-access** by selecting the **Instance ID** next to the name

Instances (4) [Info](#) Last updated less than a minute ago

Instance state = running **Clear filters**

<input type="checkbox"/>	Name	Instance ID	Instance state
<input type="checkbox"/>	Power BI Gateway	i-070d219aaa1174e68	Running
<input type="checkbox"/>	mbi-charles-db-py	i-0b98f3ee8448d1917	Running
<input type="checkbox"/>	tbt-ec2-db-access	i-085bae33a4a151c3b	Running
<input type="checkbox"/>	tbt-floridarecovers-data-cron	i-07eb6cfc0eb4ea174	Running

- Select **Connect**
- Select **RDP Client**
- Select Connect using **Fleet Manager**
- Select **Fleet Manager Remote Desktop**

Connect to instance [Info](#)
Connect to your instance i-085bae33a4a151c3b (tbt-ec2-db-access) using any of these options

Session Manager **RDP client** EC2 serial console

Instance ID
 i-085bae33a4a151c3b (tbt-ec2-db-access)

Connection Type

☐ Connect using RDP client
Download a file to use with your RDP client and retrieve your password.

☒ **Connect using Fleet Manager**
Connect to your instance using Fleet Manager Remote Desktop.

When prompted, connect to your instance using the following username and password:

Username [Info](#)
 Administrator

Password [Get password](#)

Fleet Manager Remote Desktop

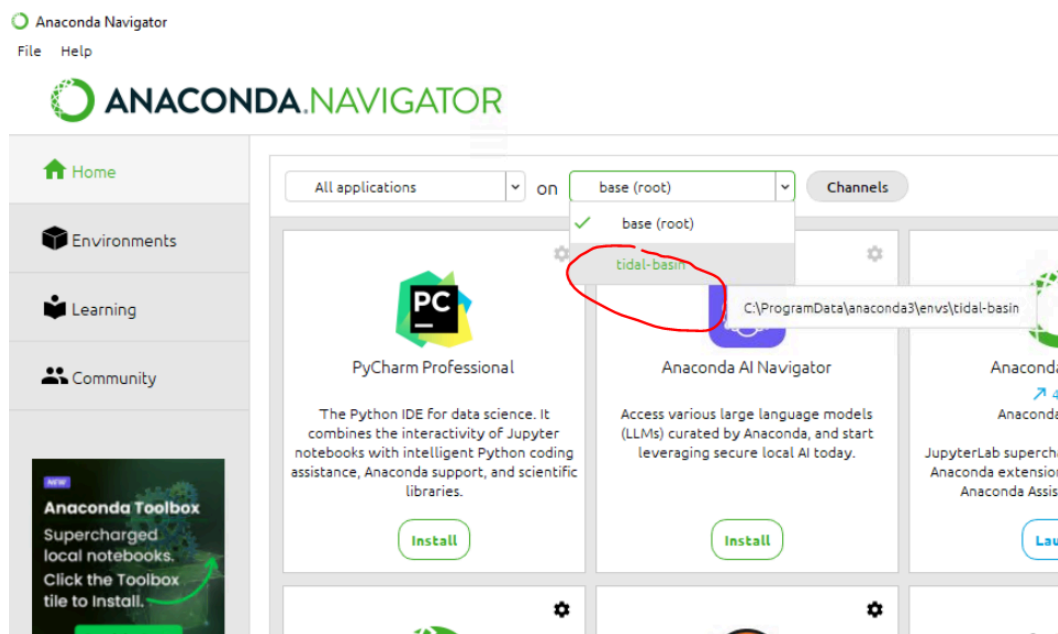
If you've joined your instance to a directory, you can use your directory credentials to connect to your instance.

- Sign in using the username and password credentials
 - Username: *Administrator*
 - Password: *%=dNFybpfc0s-0nMq896&P4A.KQAJKQ*
- Select **Connect**

2: Open Anaconda and Select the Tidal Basin Environment

Open Anaconda Navigator and open the tidal-basin environment

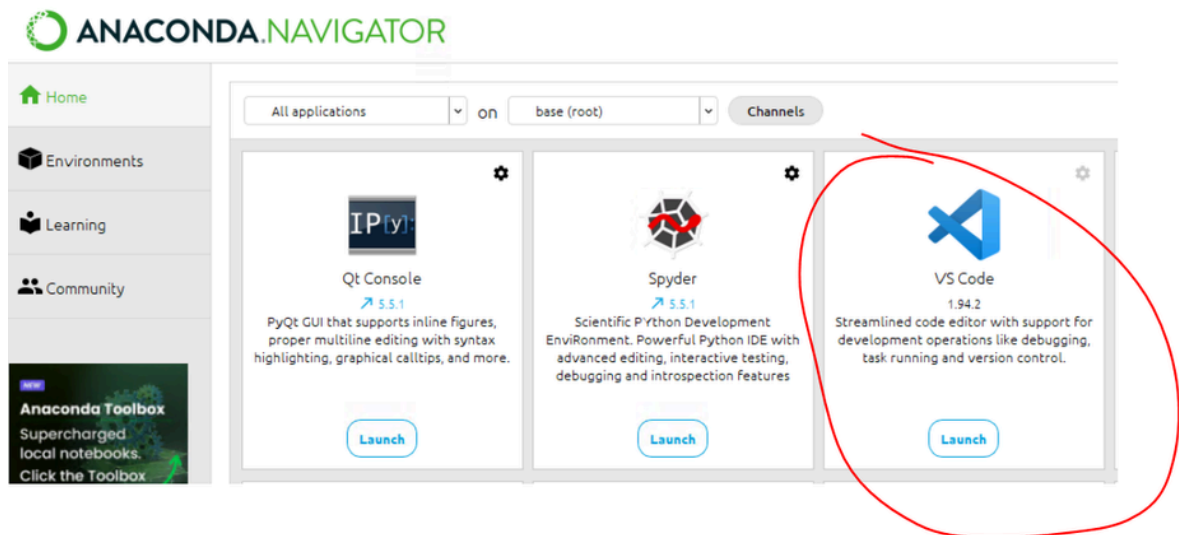
- Search for and open Anaconda Navigator application on the EC2 machine desktop
- Once on the application, locate in the top of the application where it says base (root) and select the dropdown arrow
- Select the option that says **tidal-basin**



3: Open Visual Code Studio on Anaconda Navigator

Open VS Code:

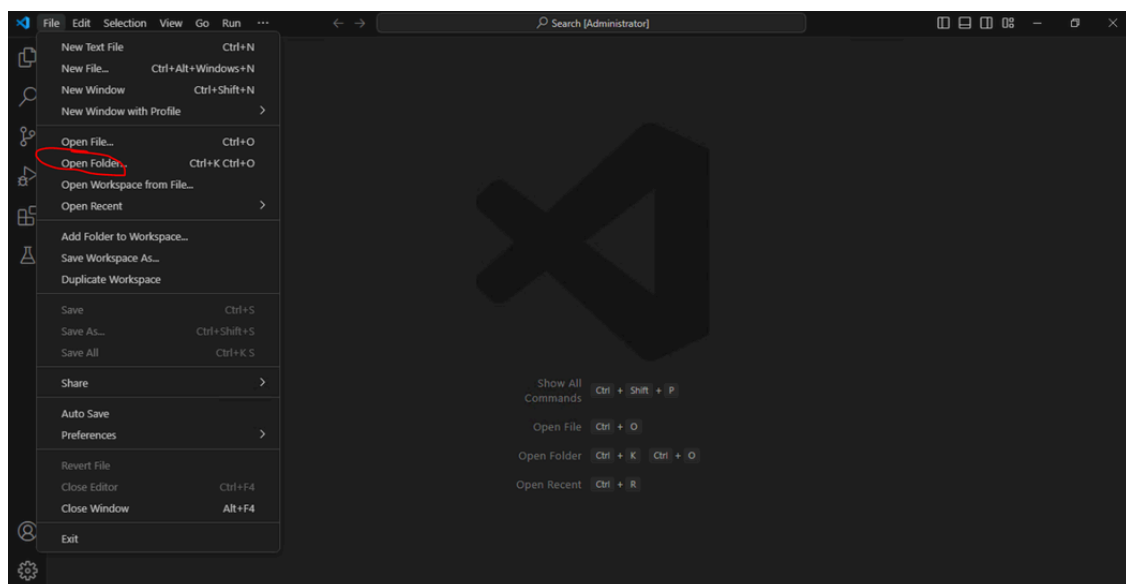
- When on the **tidal-basin** environment, scroll down the application placards till you see **VS Code**
- Select **Launch**

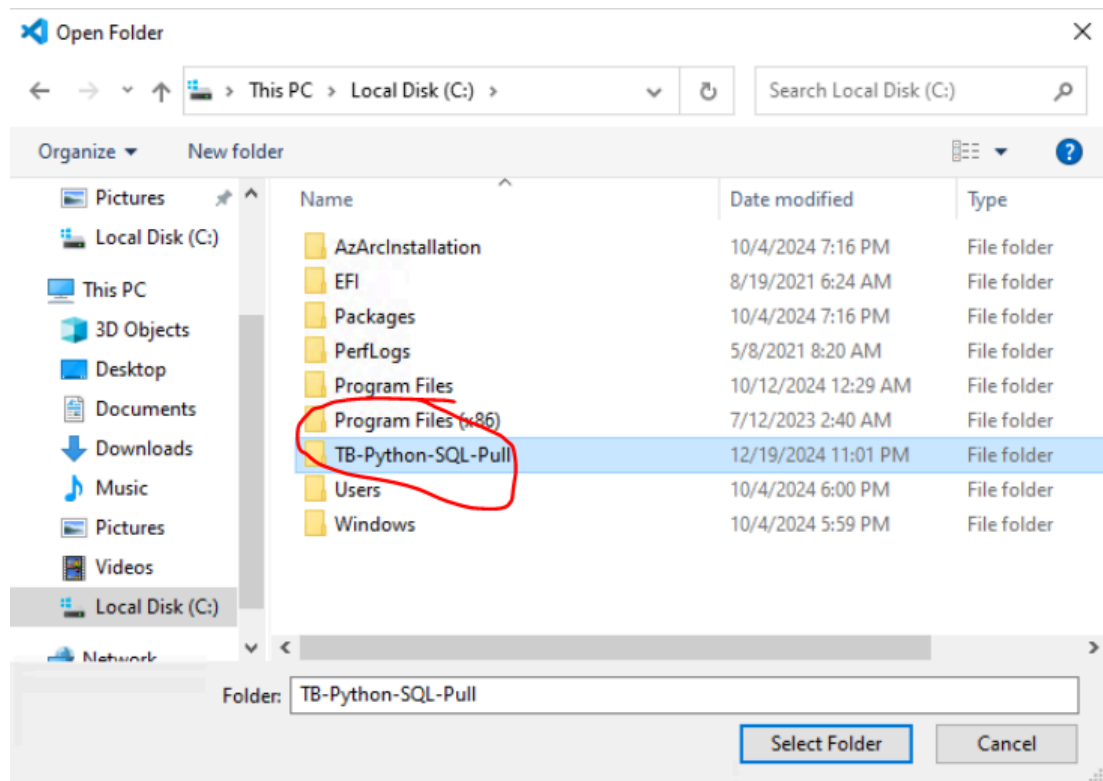


4: Open TB-Python-SQL-Pull Folder

Open the Notebook Folder:

- While in VS Code, navigate to the top toolbar and select **File**, then **Open Folder**
- Select **Launch**
- Select the **Local Disk (C:)**
- Select the folder **TB-Python-SQL-Pull**





5: Open the Data Extraction Notebook

Open the Data Extraction Notebook:

- Back in VS Code, look in the folder on the left hand side and find the file named **Data Extraction Transformation and Loading PostgreSQL to AWS S3.ipynb**
- Click to open, begin running cells to complete SQL Pulls

