Introduction

Michael Baker INTERNATIONAL

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

Introduction

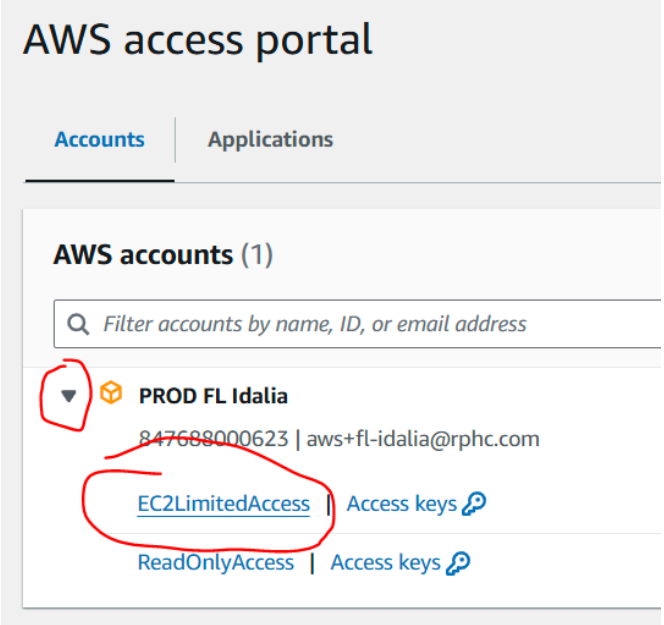
In this tutorial, we will cover the steps necessary to access the Tidal Basin EC2 machine in AWS and navigate to the Python Data Extraction jupyter notebook. The notebook will walk you through the steps of creating an SQL query, connecting to the PostgreSQL database hosted on in AWS RDS, converting that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

Table of Contents

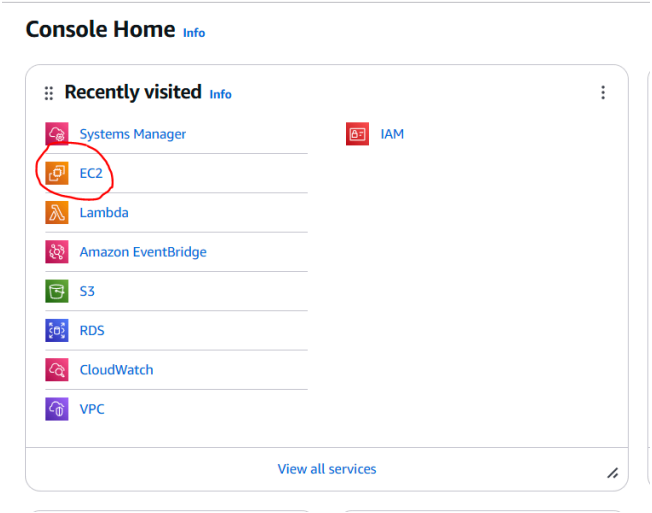
1: Access the EC2 Environment Through AWS

Logon to the AWS EC2LimitedAccess Environment:

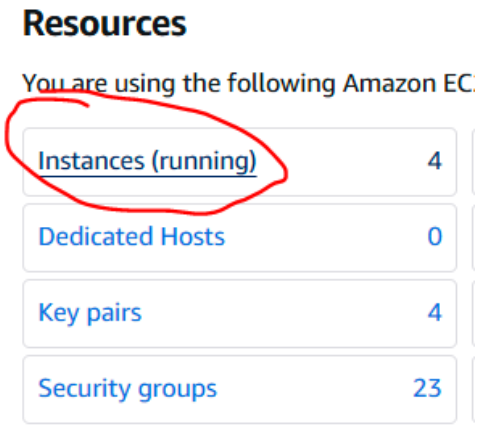
- Go to the AWS Access Portal here
- Select the drop down arrow and select **EC2LimitedAccess**



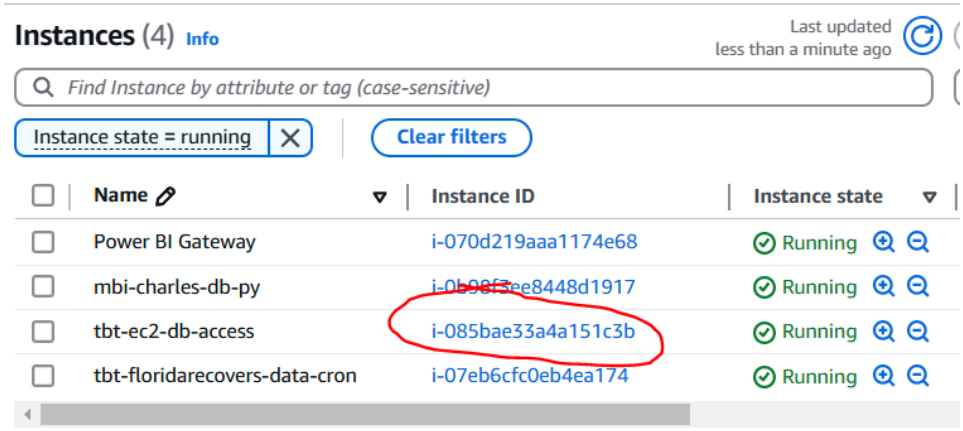
- On the main landing page, select **EC2**



- Select **Instances (running)**



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

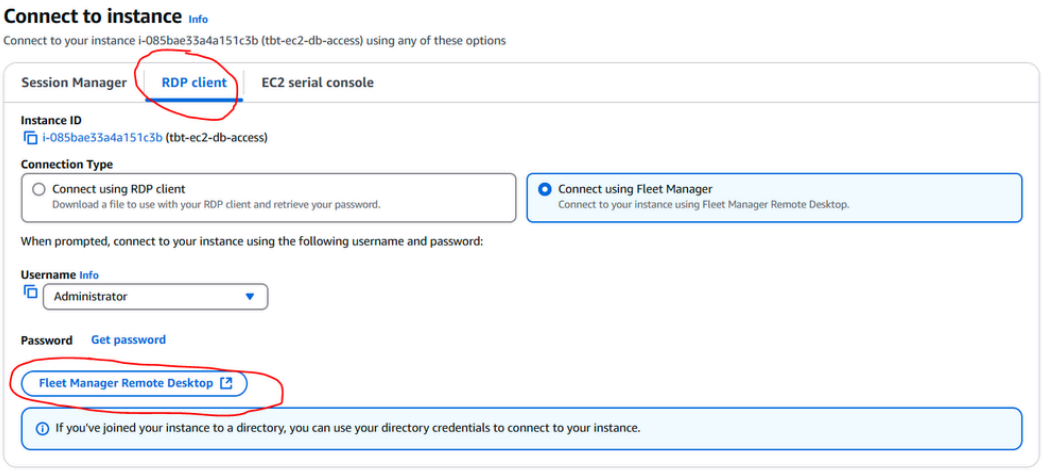


- Select **Connect**

- Select **RDP Client**

- Select Connect using **Fleet Manager**

- Select **Fleet Manager Remote Desktop**



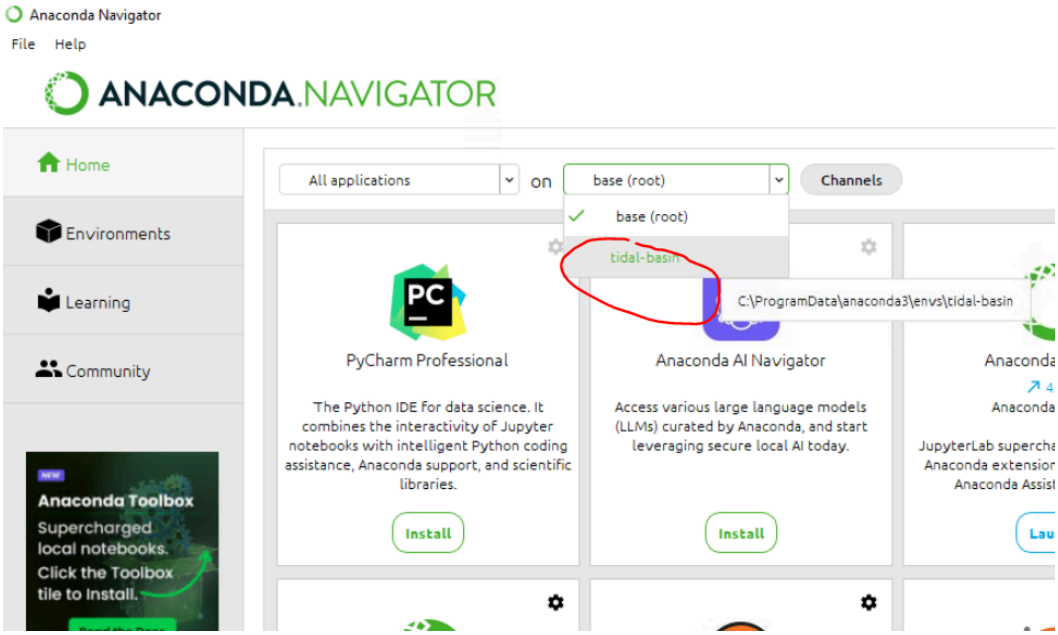
- Sign in using the username and password credentials
  - Username: *Administrator*
  - Password: *%=dNFybpf(c0s-0nMq896&P4A.KQAJKQ*

- Select **Connnect**

2: Open Anaconda and Select the Tidal Basin Environment

Open Anaconda Navigator and open the tidal-basin environment

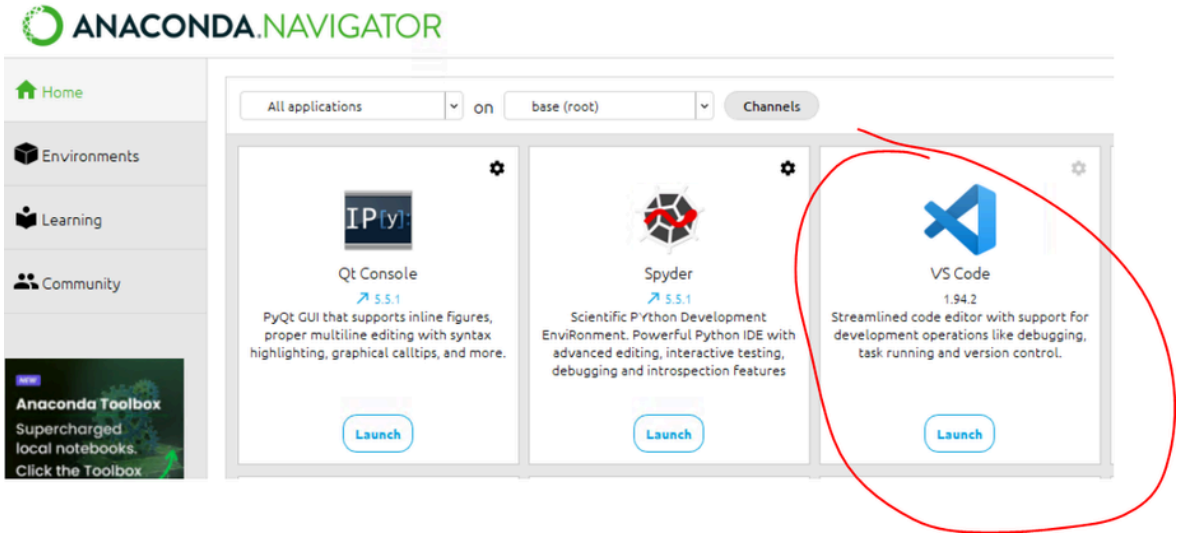
- Search for and open Anaconda Navigator application on the EC2 machine desktop
- Once on the application, locate in the top of the application where it says base (root) and select the dropdown arrow
- Select the option that says **tidal-basin**



3: Open Visual Code Studio on Anaconda Navigator

Open VS Code:

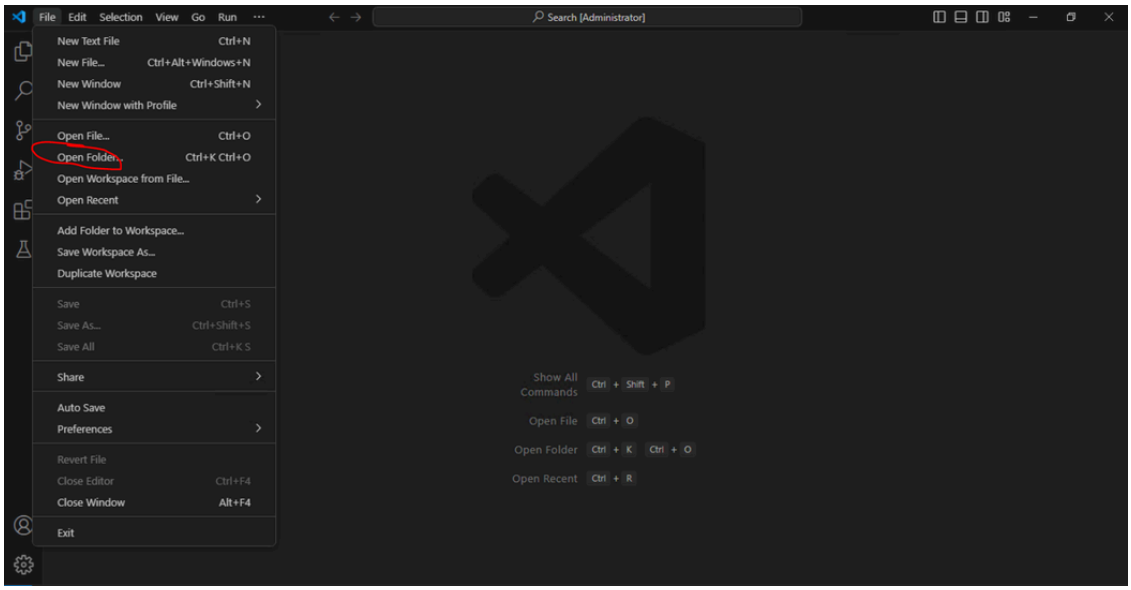
- When on the **tidal-basin** environment, scroll down the application placards till you see **VS Code**
- Select **Launch**

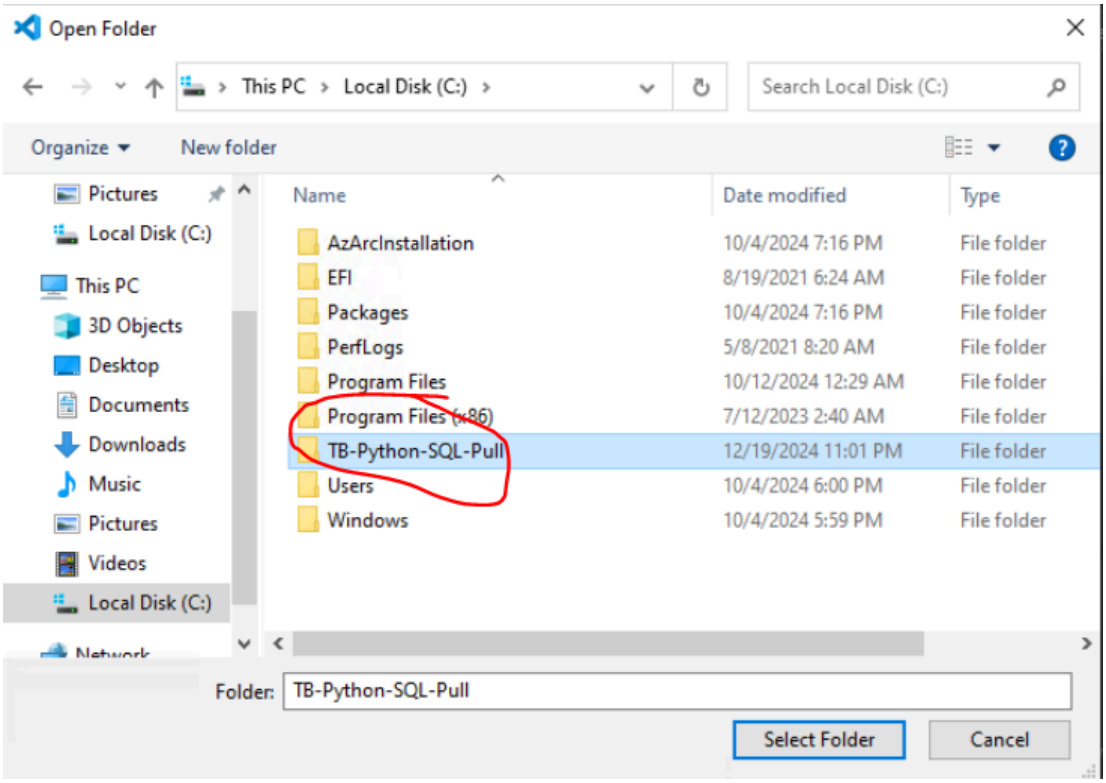


4: Open TB-Python-SQL-Pull Folder

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**

- **C**lick to open, begin running cells to complete SQL Pulls