Michael Baker
INTERNATIONAL

Tidal Basin: Power BI Python Connect to AWS S3 Report Database

Introduction

This tutorial will guide you through connecting Tidal Basin Power BI reports to an Amazon S3 bucket using Python. Ensure that you complete the previous tutorial, ***Power BI Python Environment Set Up***, as it is necessary for running Python code in Power BI. After enabling Python scripting, this tutorial provides a script to use boto3 to connect to your S3 bucket and retrieve a CSV file, which will be read into a pandas DataFrame. Once the DataFrame is imported into Power BI, you can adjust the data refresh rate to suit the specific report's needs. This step-by-step guide ensures a smooth setup and data import from S3 to Power BI.

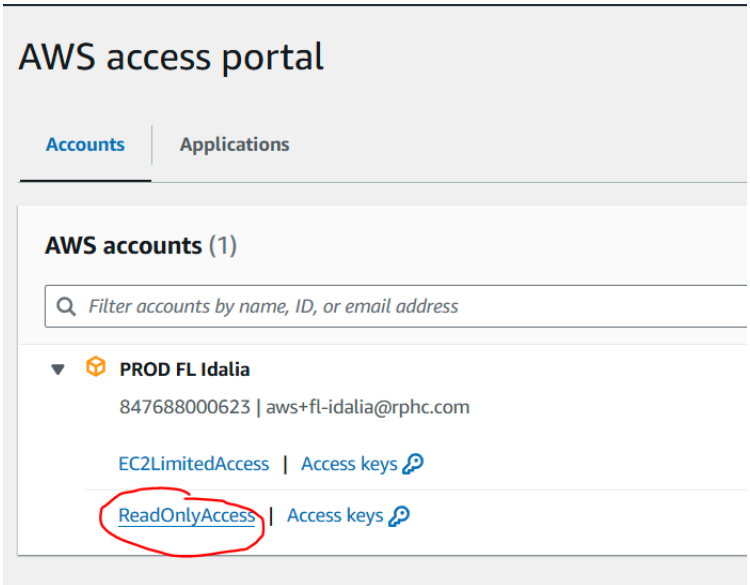
Table of Contents

1. Download Anaconda Navigator from the Anaconda website and install it on your computer.
2. Create a new environment in Anaconda Navigator called Tidal-Basin.
3. Install the CMD.exe prompt for the environment within Anaconda Navigator
4. Install required packages (boto3, pandas, matplotlib) in the Tidal-Basin environment using the command prompt.
5. Copy the Tidal-Basin environment folder path
6. Set up the Power BI Desktop python environment with the saved folder path

1: Navigate to the Reporting External AWS S3 Bucket

Sign in to AWS and navigate to the S3 Bucket :

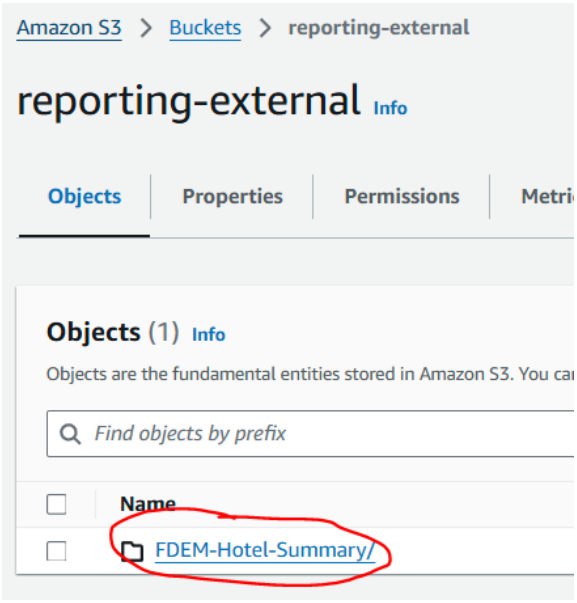
- Go to AWS Sign In link, sign in with Tidal Basin Credentials
- Click on the **ReadOnlyAccess** option
- Now open a new tab and go to the S3 reporting-external bucket

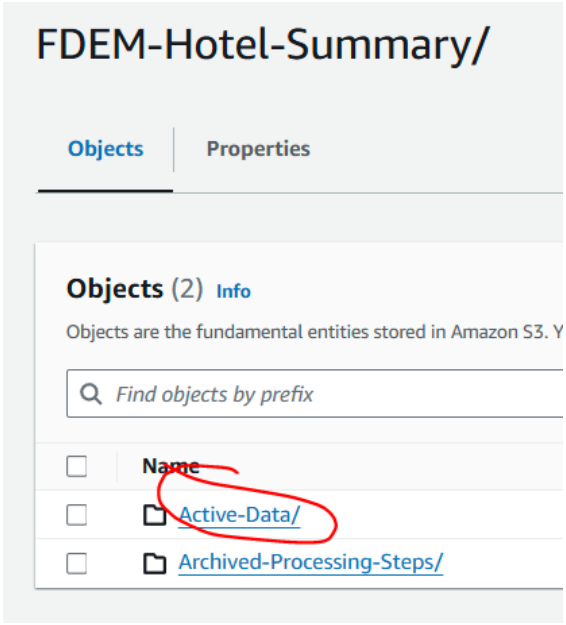


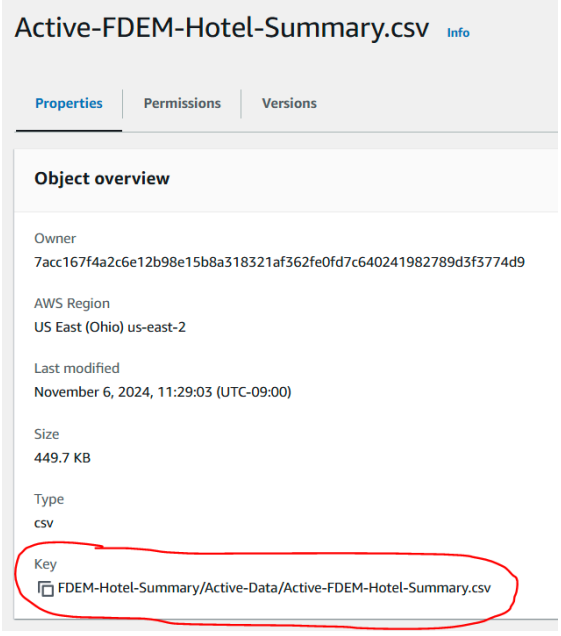
1: Locate the Report's Active Data Key

Find the report folder and locate the active data key :

- While on the reporting-external bucket main page, locate the folder with the same name as the report and select it
- Select the **Active-Data/** Option
- Click on the file stored in this folder, it will bring you to an information panel on the csv file
- Navigate to the bottom of the first panel to where it says **Key** and copy and store the key value, this will be used later when setting up our python file.


Amazon S3 > Buckets > reporting-external
reporting-external Info
Objects
Properties
Permissions
Objects (1) Info
Objects are the fundamental entities stored in Amazon S3. You ca
Find objects by prefix
Name
FDEM-Hotel-Summary/

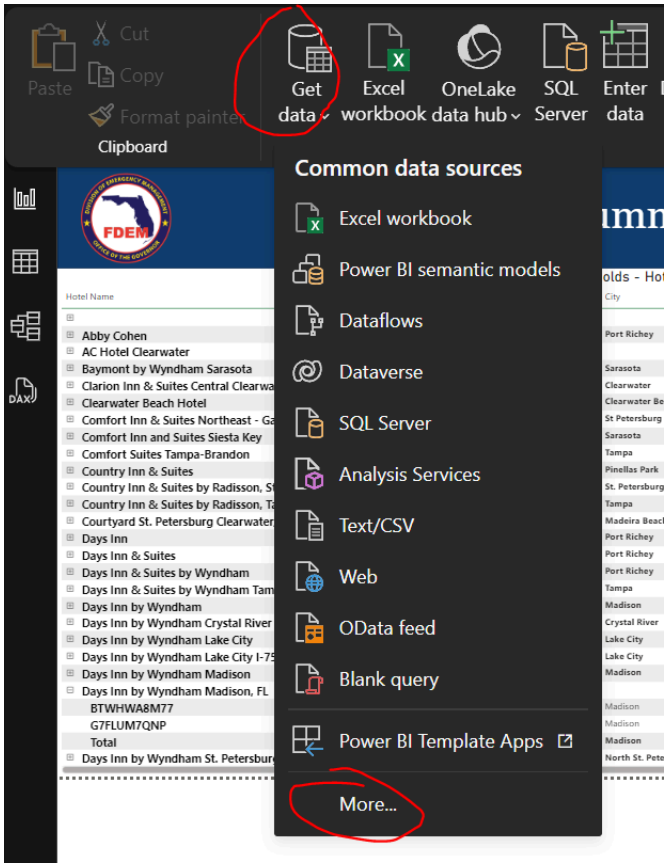

FDEM-Hotel-Summary/
Objects
Properties
Objects (2) Info
Objects are the fundamental entities stored in Amazon S3. Y
Find objects by prefix
Name
Active-Data/
Archived-Processing-Steps/


Active-FDEM-Hotel-Summary.csv Info
Properties
Permissions
Versions
Object overview
Owner
7acc167f4a2c6e12b98e15b8a318321af362fe0fd7c640241982789d3f3774d9
AWS Region
US East (Ohio) us-east-2
Last modified
November 6, 2024, 11:29:03 (UTC-09:00)
Size
449.7 KB
Type
csv
Key
FDEM-Hotel-Summary/Active-Data/Active-FDEM-Hotel-Summary.csv

2: Create New Data Connection in Power BI

Create a new data connection in the Power BI Report:

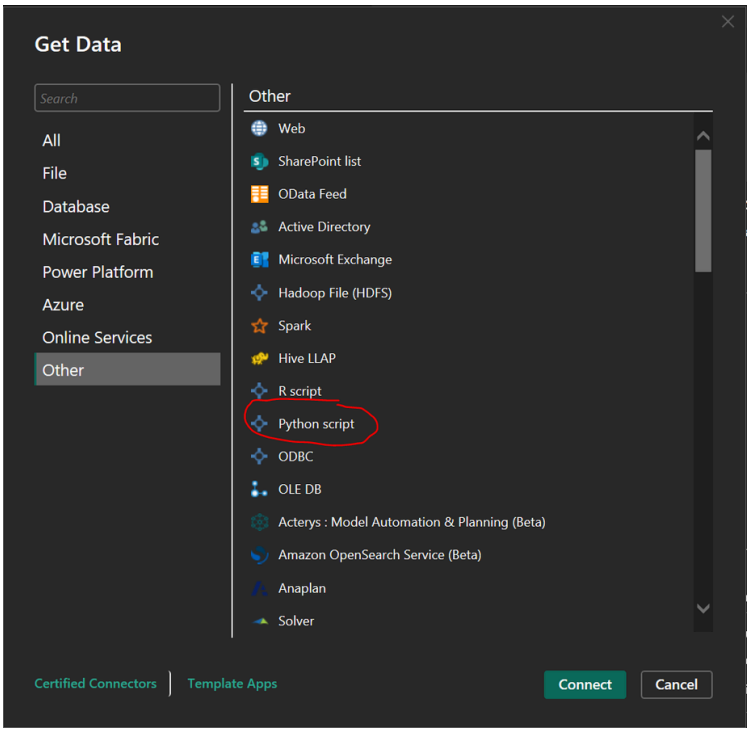
- Navigate to your Power BI Report and open it in the desktop application
- Click on the **Get Data** button on the top left
- Go to the bottom of the panel and select **More**



3: Create Python Data Connection

Paste Python Script with Key to Connect Data :

- On the More Connections Panel, Select the **Other** option
- Select **Python Script**
- Select **Connect**



4: Paste Python Script with Key

Paste The Data Connect Python Script with the Saved Key:

- Where it says **FILE KEY**, paste the key that was saved from the second step.
- Ensure the key is still wrapped by two quotation marks ex: "KEY"
- Copy and paste the python script below, this is the code that will connect and pull data from the S3 bucket.
- Click **OK**

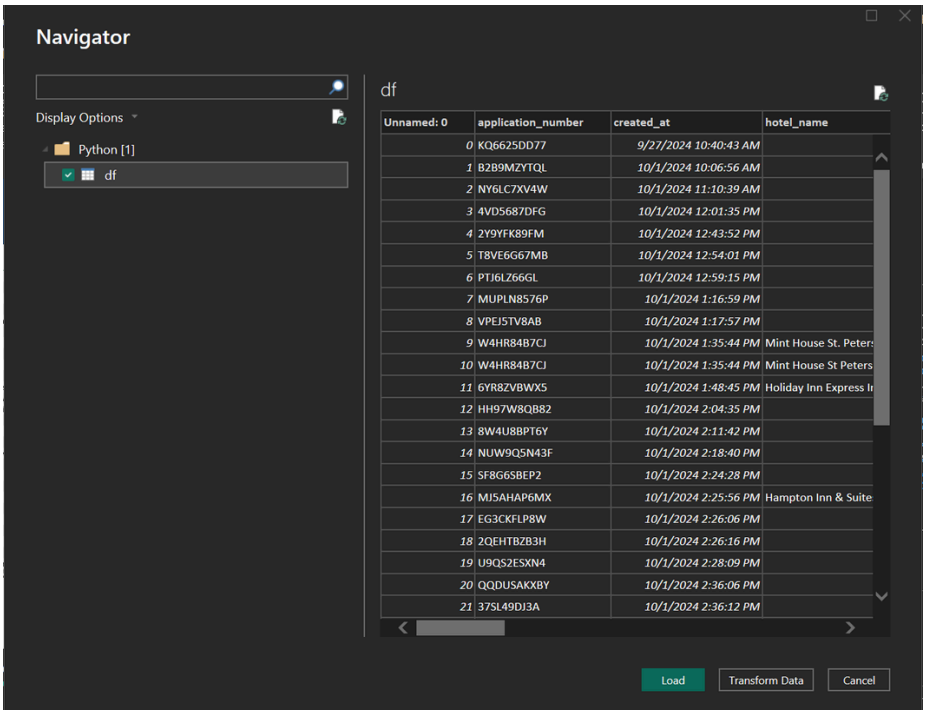
```
# Import Packages
import boto3
import pandas as pd

# Access Key
access = 'AKIA2NPV3FHVMR3TWB4N'

# Secret Key
secret = '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'

# Bucket Name
bucket_name = 'reporting-external'

# File Key (UPDATE THIS WHEN CONNECTING TO NEW REPORT)
key = "INSERT KEY HERE"  #Do not remove quotation marks around key !IMPORTANT

# S3 Connector
s3 = boto3.resource(
    service_name='s3',
    region_name='us-east-2',
    aws_access_key_id=access,
    aws_secret_access_key=secret
)

# Retrieve the CSV
obj = s3.Bucket(bucket_name).Object(key).get()

# Convert CSV to Pandas DF
df = pd.read_csv(obj['Body'])
```

5: Select and Load the DataFrame (DF)

Select the DataFrame that was Created from the Python Script:

- Once run, a window will appear saying the data is connecting

- If the connection was successful, a window will appear that will list a folder called **Python**, and a table within that folder called **df**
- Select **df**
- Review the table to ensure the data looks correct, then select **Load**



Complete

You have now successfully used python to add data from an AWS S3 Bucket to the Power BI Report.