

DATA REFRESH AND SCHEDULING

TOPIC: BRAZIL E-COMMERCE DATA ANALYSIS

TEAM: 3

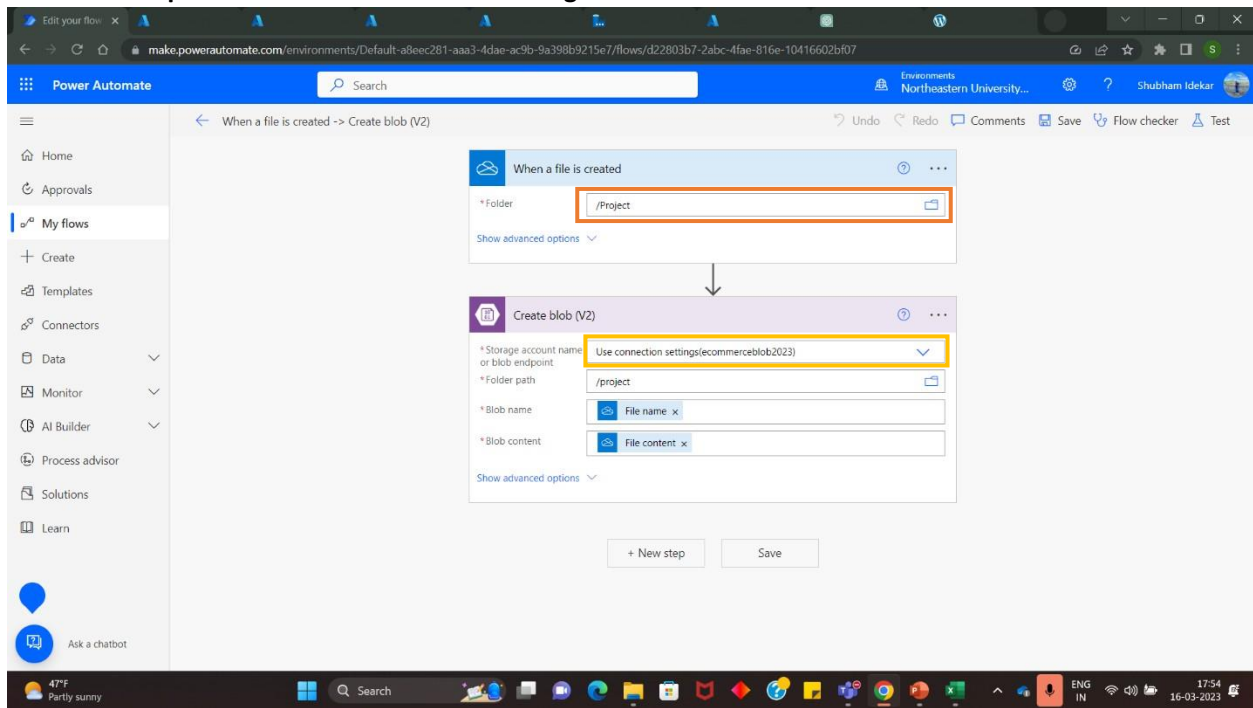
Abbas Furniturewala - NUID 002193272

Shubham Idekar - NUID 002776415

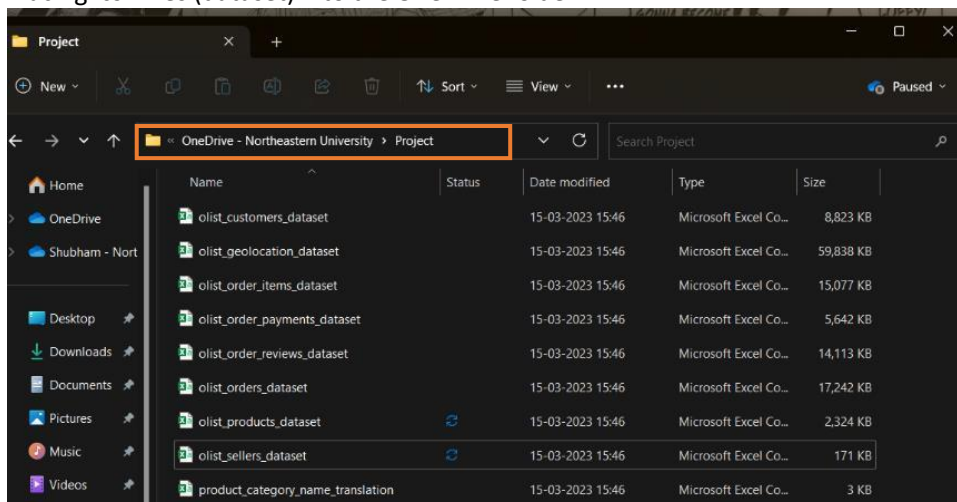
Shrutika Salian - NUID 002142365

Harshit Parikh - NUID 001044838

Using Power Automates Automated Cloud Flow to fetch files from OneDrive folder whenever a new file is created and upload the files to Azure Blob Storage:



Placing .csv files (dataset) into the OneDrive folder:



Successful execution of Power Automate flow:

The screenshot shows the Power Automate web interface. The left sidebar contains navigation options: Home, Approvals, My flows (selected), Create, Templates, Connectors, Data, Monitor, AI Builder, Process advisor, Solutions, and Learn. The main area displays a table of flow runs. The table has columns for Start time, Duration, and Status. All runs are marked as 'Succeeded'.

| Start time | Duration | Status |
|----------------------------|----------|-----------|
| Mar 15, 04:06 PM (1 d ago) | 00:00:04 | Succeeded |
| Mar 15, 04:06 PM (1 d ago) | 00:00:14 | Succeeded |
| Mar 15, 04:05 PM (1 d ago) | 00:00:14 | Succeeded |
| Mar 15, 04:05 PM (1 d ago) | 00:00:04 | Succeeded |
| Mar 15, 04:05 PM (1 d ago) | 00:00:03 | Succeeded |
| Mar 15, 04:00 PM (1 d ago) | 00:00:02 | Succeeded |
| Mar 15, 03:54 PM (1 d ago) | 881 ms | Succeeded |
| Mar 15, 03:49 PM (1 d ago) | 777 ms | Succeeded |

All files placed into Azure Blob Storage automatically once created in OneDrive Folder:

The screenshot shows the Microsoft Azure portal. The breadcrumb navigation is 'Home > ecommerceblob2023 | Containers >'. The container 'project' is selected. The left sidebar shows navigation options: Overview (selected), Diagnose and solve problems, Access Control (IAM), Settings, Shared access tokens, Access policy, Properties, and Metadata. The main area displays a table of blobs in the 'project' container. The table has columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. All blobs are 'Block blob' type and 'Available'.

| Name | Modified | Access tier | Archive status | Blob type | Size | Lease state |
|---------------------------------------|-----------------------|----------------|----------------|------------|------------|-------------|
| olist_customers_dataset.csv | 3/15/2023, 4:06:19 PM | Hot (Inferred) | | Block blob | 8.62 MiB | Available |
| olist_geolocation_dataset.csv | 3/15/2023, 6:51:16 PM | Hot (Inferred) | | Block blob | 58.44 MiB | Available |
| olist_order_items_dataset.csv | 3/15/2023, 4:05:37 PM | Hot (Inferred) | | Block blob | 14.72 MiB | Available |
| olist_order_payments_dataset.csv | 3/15/2023, 4:05:45 PM | Hot (Inferred) | | Block blob | 5.51 MiB | Available |
| olist_order_reviews_dataset.csv | 3/15/2023, 4:06:20 PM | Hot (Inferred) | | Block blob | 13.78 MiB | Available |
| olist_orders_dataset.csv | 3/15/2023, 4:06:08 PM | Hot (Inferred) | | Block blob | 16.84 MiB | Available |
| olist_products_dataset.csv | 3/15/2023, 4:00:07 PM | Hot (Inferred) | | Block blob | 2.27 MiB | Available |
| olist_sellers_dataset.csv | 3/15/2023, 3:54:23 PM | Hot (Inferred) | | Block blob | 170.61 KiB | Available |
| product_category_name_translation.csv | 3/15/2023, 3:49:17 PM | Hot (Inferred) | | Block blob | 2.55 KiB | Available |

Trigger to run the ADF pipeline:

The screenshot shows the 'Edit trigger' configuration for a pipeline named 'OrderItems' in the Microsoft Azure portal. The trigger is a 'ScheduleTrigger' with the following settings:

- Description:** (Empty text box)
- Type:** ScheduleTrigger
- Start date:** 4/10/2023, 10:00:00 PM (highlighted with a yellow box)
- Time zone:** Eastern Time (US & Canada) (UTC-5)
- Recurrence:** Every 1 Day(s)
- Advanced recurrence options:** (Expanded section with 'Specify an end date' unchecked)
- Annotations:** (Empty list)
- Status:** Started (Selected radio button)

The 'OK' button is visible at the bottom of the configuration panel.

The screenshot shows the 'Pipeline runs' list for the 'OrderItems' pipeline in the Microsoft Azure portal. The list displays the following data:

| Pipeline name | Run start | Run end | Duration | Triggered by | Status | Run | Parameters |
|---------------|------------------------|------------------------|----------|----------------|-----------|----------|------------|
| OrderItems | 4/10/2023, 10:00:00 PM | 4/10/2023, 10:27:02 PM | 00:27:01 | DailyTrigger | Succeeded | Original | |
| OrderItems | 4/10/2023, 6:47:29 PM | 4/10/2023, 7:13:58 PM | 00:26:29 | Manual trigger | Succeeded | Original | |
| OrderItems | 4/10/2023, 6:41:27 PM | 4/10/2023, 6:42:05 PM | 00:00:38 | Manual trigger | Cancelled | Original | |
| OrderItems | 4/10/2023, 5:19:50 PM | 4/10/2023, 5:46:26 PM | 00:26:36 | Manual trigger | Succeeded | Original | |
| OrderItems | 4/10/2023, 4:28:18 PM | 4/10/2023, 4:55:21 PM | 00:27:02 | Manual trigger | Succeeded | Original | |
| OrderItems | 4/10/2023, 3:43:49 PM | 4/10/2023, 4:26:34 PM | 00:42:45 | Manual trigger | Cancelled | Original | |
| OrderItems | 4/10/2023, 2:52:43 PM | 4/10/2023, 3:15:35 PM | 00:22:52 | Manual trigger | Failed | Original | |

The first row, representing the scheduled run, is highlighted with a yellow box. The table shows 7 items in total, with the first item being the scheduled run.

Automating the python code run:

Batch script to run python code automatically:

```
automate_py_code
File Edit View
@echo off
set PYTHON_EXE="C:\Python39\python.exe"
set PYTHON_SCRIPT="C:\Users\Harshit\Desktop\USA\NEU\4. Spring 2023\ADBMS\Project\azure-cosmos-db-graph-python\cosmos_project.py"
%PYTHON_EXE% %PYTHON_SCRIPT%
pause
```

"@echo off": To stop the console from showing each command as it is executed.

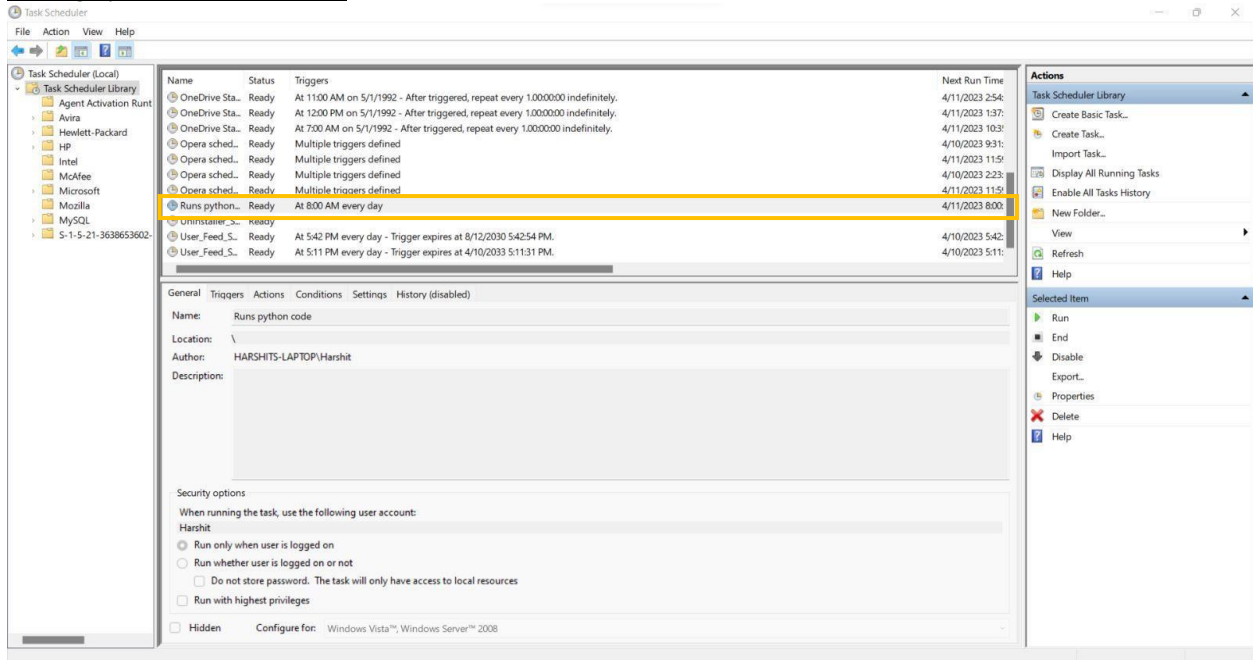
"set PYTHON EXE": sets the path to the Python executable in the PYTHON EXE variable.

"set PYTHON SCRIPT": sets the path of the Python script in the PYTHON SCRIPT variable.

"%PYTHON EXE%%PYTHON SCRIPT%": invokes the Python executable with the path to the script as an argument to run the Python script.

"pause": adds a pause to the script's end to stop the console window from shutting right away when the Python code has finished running.

Setting up the Task Scheduler:



1: Runs python code Properties (Local Computer)

X

General Triggers Actions Conditions Settings History (disabled)

When you create a task, you can specify the conditions that will trigger the task.

| Trigger | Details | Status |
|---------|----------------------|---------|
| Daily | At 8:00 AM Every day | Enabled |

" Runs python code Properties (Local Computer)

X

General Triggers Actions Conditions Settings History (disabled)

When you create a task, you must specify the condition that will occur when your task starts.

| Action | Details |
|-----------------|---|
| Start a program | "C:\Users\Harshit\Desktop\USA\NEU\4.Spring 2023\ADBMS\Project\azure-cosmos-db-graph-python" |

Desktop

Downloads

Documents

Pictures

All documents

IntelliJView

Behavioral question.i

Data Engineer

JOB

4, Spring 2023

Assignments

azure-cosmos-db-graph-python

Data Analyst

Project

OneDrive - Northeastern University

Attachments

dam97370_2022_01_5spring

Microsoft Team; Chat me:::

This PC

Desktop

Documents

Downloads

Name

.git

.github

.gitignore

automate_py_code

automate_py_code

CHANGELOG.md

CombineCSVs

connect

CONTRIBUTING.md

cosmos_project

gremlin_project

gremlin_project1

LICENSE.md

README.md

Requirements

test1

Date modified

3/25/2023 1:53 PM

5/25/2023 1:53 PM

3/25/2023 1:53 PM

4/10/2023 1:33 PM

4/10/2023 1:43 PM

3/25/2023 1:53 PM

4/9/2023 11:11 AM

WS/2023 2:01 PM

1/25/2023 1:53 PM

4/9/2023 11:13 AM

4/9/2023 9:37 AM

4/3/2023 12:31 AM

>125/2023 1:53 PM

1/25/2023 1:53 PM

1/25/2023 1:53 PM

4/10/2023 1:31 PM

Type

File folder

File folder

GITIGNORE File

Windows Batch file

Text Document

MD File

Microsoft Excel Co...

Python File

MD File

Microsoft Excel Co...

Microsoft Excel Co...

MO File

MD File

Text Document

Python File

Size

5 KB

1 KB

1 KB

58,149 KB

10KB

HB

11 KB

58,149 KB

47,458 KB

2 KB

4 KB

1KB

1KB

Runs python code Properties (Local Computer) ✕

General Triggers Actions Conditions Settings History (disabled)

Specify additional settings that affect the behavior of the task.

- ☒ Allow task to be run on demand
- ☒ Run task as soon as possible after a scheduled start is missed
- ☒ If the task fails, restart every: 1 minute ▼
Attempt to restart up to: 3 times
- ☒ Stop the task if it runs longer than: 12 hours ▼
- ☒ If the running task does not end when requested, force it to stop
- ☐ If the task is not scheduled to run again, delete it after: 30 days ▼

If the task is already running, then the following rule applies:

Do not start a new instance ▼

In this way, we have successfully scheduled our pipelines and python code.