

# Documentation Guide for Steps Taken

## SETUP GUIDE FOR DATA PIPELINE:

This project, while seemingly straightforward, took me some time to get implemented. I had to view documentation from Microsoft and watch various YouTube demonstration videos detailing setup methods. These resources are provided at the end of this documentation. Much of this information is included in the project report, however I also wanted to provide it separately as well. An example of my “coding”.

All this information is posted on my github at:  
[https://github.com/Jhansen19/Airline\\_Predictions](https://github.com/Jhansen19/Airline_Predictions).

## ENVIRONMENT SETUP:

After creating the Azure Cloud Services account and logging in.

Create a resource group ...

Step 1: Create Resource Group – click on Create Resource Group and fill out the following information.

Basics Tags Review + create

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Project details

Subscription \* ⓘ Azure for Students

Resource group \* ⓘ

Resource details

Region \* ⓘ (US) East US

### Resources

Recent Favorite

Name

datalakestorage-dep

jh-data-engineering-project

af-jhdep-01

adf-jhdep-01

db-jhdep-01

kv-jhdep-01

Step 2: After creating your Resource Group I created the Azure Services I wanted to use in my data pipeline.

Here is the list of resources I created for the pipeline. The ones highlighted in red are the ones used in this pipeline.

Step 3: Creating these resources is relatively straightforward. First, from the home page I clicked on create a resource and searched for the resource that I wanted to create. I clicked the ‘Create’ button next to the icon of the resource I wanted to create.

Fill out the necessary information from the screenshot.

For each resource there are differing configurations for the creation of the resources. It was important to know these differences, and, in some cases, why different selections would be more beneficial.

For azure data factory and Azure Key Vault some of these included whether to allow public or private endpoints, tags, instance region and version.

For Azure Storage Account Gen2 (Data Lake) the important configurations included:

- ‘Azure Blob Storage or Azure Data Lake Storage Gen 2’ – This gives the option to make the storage account a data lake that has a hierarchical namespace.
- ‘Region’ – select regions closest to where the Data Lake is most used.
- ‘Redundancy’ – Can choose locally-redundant storage, Geo-redundant storage, zone-redundant storage, or geo-redundant storage. For this project I chose Geo-redundant storage as it was a low-cost option that had a secondary regional backup. This way if there is a regional catastrophe in the primary region there is another secondary region that has a backup.
- Under the ‘Advanced Section’ be sure to select ‘enable hierarchy namespace’ this will create the datalake storage instead of a blob storage.

## INGESTION:

### Creating an Integration Runtime on local device:

There were a couple of methods I tried to implement. The first of these was ingesting from a local file.

Step 1: Open Azure Data Factory

Step 2: Go to the ‘Manage’ tab and click on ‘Integration Runtimes’.

Step 3: Click on ‘+ New’. This will open a panel on the right-hand side of the screen.

Step 4: Click on ‘Azure, Self-Hosted’ tab and continue.

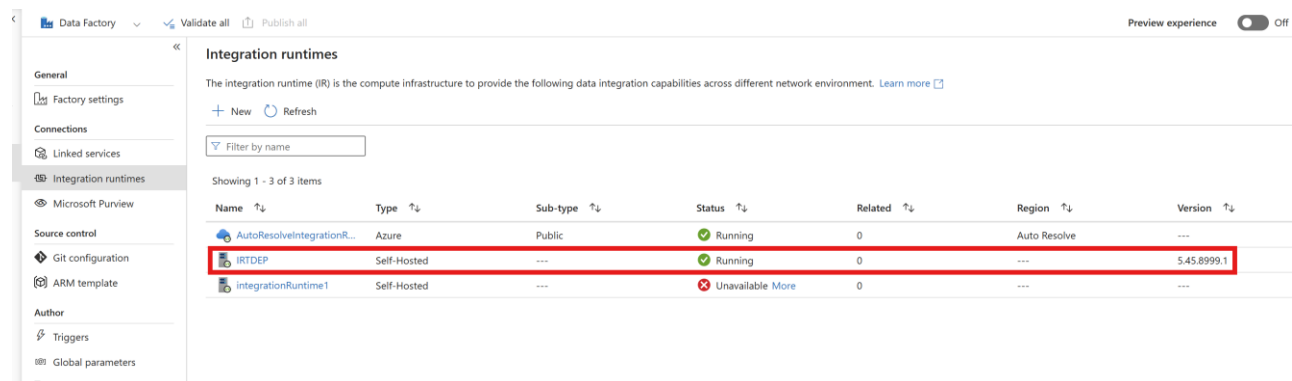
Step 5: Click on ‘Self-Hosted’ if using for an on-premises or private network or you can alternately use a existing self-hosted integration runtime that is used for another resource and click ‘Continue’.

Step 6: Click ‘Create’ on the next page. This will initiate the runtime and create the icon in your azure data factory. Note this does not connect your integration runtime to your data factory listing. You will set up the runtime integration on the next page.

Step 7: On this page you will select ‘Option 2: Manual setup’ and click on the link called ‘Download and install integration runtime’. Be sure to not close this page as you will need the listed ‘Key1’ and ‘Key2’ when setting up the integration runtime.

Step 8: You will follow the directions to download and install the integration runtime on your local device. Copy the authentication key from step 7 into the box before finishing.

This is what it will look like when complete:



Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationR...	Azure	Public	Running	0	Auto Resolve	---
IRTEP	Self-Hosted	---	Running	0	---	5.45.8999.1
integrationRuntime1	Self-Hosted	---	Unavailable	0	---	---

## Connecting local folder

Step 1: Go to ‘Linked Service’ under the ‘Manage’ tab. Then click on the ‘+ New’ button which will open panel on the righthand side.

Step 2: Select the ‘File system’ linked service.

Step 3: This will take you to the next page.

**New linked service**  
 File system [Learn more](#)

Name \*  
 FileServer1

Description

Connect via integration runtime \* ⓘ  
 ✓ AutoResolveIntegrationRuntime

Host \* ⓘ  
 e.g. \\ServerName\SharedFolder\[\Folder], \\<storage name>.file.core.windows.net\[\file servic

User name \*

Password Azure Key Vault

Password \*

Annotations  
 + New

> Parameters

> Advanced ⓘ

You will need to input the file pathway and assure proper sharing has been enabled on your file. Then you will add the username of your computer and the password for your computer.

(This is the part that did not work for me. When I input this information it gave me an error, however it gives very little information. Looking up the error code it looks like a problem with my username password combination. I could see this might be this issue as my login is facial recognition. I tried other methods to get this working with no success)

## TRANSFORMATIONS:

Initially I planned to mount my data lake from a Databricks compute cluster, however this did not end up work, I believe this was because the IU account I was using lacked proper administrator permissions to create a Microsoft Entra ID service principle and provide it access to the azure storage accounts. There was an additional issue with Databricks not having a compute cluster accessible via the student account credits I had available. Because of this issue I investigated other solutions Azure Functions App – which would allow me to write in a notebook but looked difficult to configure – and Azure Data Factory – which was easier to configure and commonly used, but I did not know how to use data processing and transformations in Data Factory. I selected the latter of these two as I was able to find a lot of good tutorials of how to implement data processing and transformation steps.

## Connecting to containers:

▲ Datasets	4
ImportfromBronze	
● ImportfromSilver1	
● SinktoGold	
● SinktoSilver	

I created 2 datasets. The first one 'ImportfromBronze' connects to the csv file in the datalake bronze container. The second one is connecting to the datalake silver container.

Here are the steps to connect to the datalake container. In this case I am Importing from the Silver Container.

Step 1: Click on the 'Author' tab on the left-hand side of the screen. Then go to datasets and click on 'ellipses' and click on New Dataset.

Step 2: A panel on the right-hand side will open and I clicked on the 'Azure Data Lake Storage Gen2'.

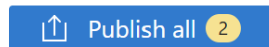
Step 3: Select the file format to import. I selected the 'DelimitedText'.

Step 4: I Filled out the information as listed in the corresponding picture with the Linked service and corresponding file path.

After pressing 'OK' I tested the connection which was successful.

Step 5: Can view schema by clicking on the 'Schema' tab.

Step 6: It is necessary to click on the publish tab in the top left-hand corner for changes to take effect.



This completes connecting a dataset to data factory and publishing to sink from data factory.

**Set properties**

Name  
ImportfromSilver

Linked service \*  
AzureDataLakeStorage1

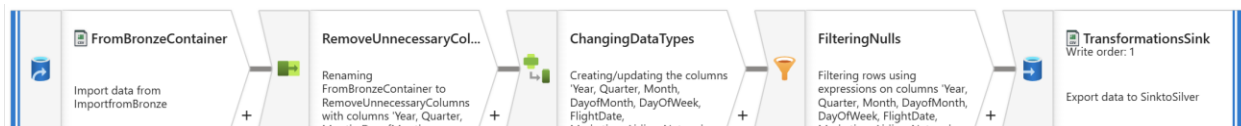
File path  
silver / Directory / File name

First row as header ☒

Import schema  
☒ From connection/store ☐ From sample file ☐ None

## Data Flows in Data Factory (Data Processing and Transformations):

The first data flow I worked on was the data processing step.



Step 1: To create this data flow I first connected my dataset called 'ImportfromBronze'

**Source settings** | Source options | Projection | Optimize | Inspect | Data

Output stream name \*  
FromBronzeContainer

Description  
Import data from ImportfromBronze

Source type \*  
☒ Dataset ☐ Inline

Dataset \*  
ImportfromBronze

Options  
☒ Allow schema drift ⓘ  
☐ Infer drifted column types ⓘ  
☐ Validate schema ⓘ

Skip line count  
[Empty field]

Sampling \* ⓘ  
☐ Enable ☒ Disable

Step 2: I removed all unnecessary columns that either had no information or very little information and contained mostly nulls values. I did this by clicking on the trash can button signified below. This step condensed my dataset from 119 columns to 66 columns. The goal is to

minimize this even more from the Silver to Gold processing, separating this dataset into specified sets for specific machine learning model predictions.

Select settings | Optimize | Inspect | Data preview

Output stream name \*  [Learn more](#)

Description  [Reset](#)

Incoming stream \*

Options

- ☒ Skip duplicate input columns
- ☒ Skip duplicate output columns

Input columns \* ☐ Auto mapping [Reset](#) [Add mapping](#) [Delete](#) 66 mappings

FromBronzeContainer's column	Name as	
<input type="checkbox"/> <input type="text" value="Year"/>	<input type="text" value="Year"/>	+
<input type="checkbox"/> <input type="text" value="Quarter"/>	<input type="text" value="Quarter"/>	+
<input type="checkbox"/> <input type="text" value="Month"/>	<input type="text" value="Month"/>	+
<input type="checkbox"/> <input type="text" value="DayOfMonth"/>	<input type="text" value="DayOfMonth"/>	+
<input type="checkbox"/> <input type="text" value="DayOfWeek"/>	<input type="text" value="DayOfWeek"/>	+

Step 3: In this step I changed the columns to the correct formatting. The majority of columns were changed to and integer format while date was changed to a date format.

Derived column's settings | Optimize | Inspect | Data preview

Output stream name \*  [Learn more](#)

Description  [Reset](#)

Incoming stream \*

[+ Add](#) [Clone](#) [Delete](#) [Open expression builder](#)

Column	Expression
<input type="checkbox"/> <input type="text" value="FlightDate"/>	<input type="text" value="toDate(FlightDate)"/>
<input type="checkbox"/> <input type="text" value="Year"/>	<input type="text" value="toInteger(Year)"/>
<input type="checkbox"/> <input type="text" value="Quarter"/>	<input type="text" value="toInteger(Quarter)"/>
<input type="checkbox"/> <input type="text" value="Month"/>	<input type="text" value="toInteger(Month)"/>
<input type="checkbox"/> <input type="text" value="DayOfMonth"/>	<input type="text" value="toInteger(DayOfMonth)"/>
<input type="checkbox"/> <input type="text" value="DayOfWeek"/>	<input type="text" value="toInteger(DayOfWeek)"/>
<input type="checkbox"/> <input type="text" value="Marketing_Airline_Network"/>	<input type="text" value="toInteger(Marketing_Airline_Network)"/>

Step 4: I next filtered null values. I did this using the expression builder this code 'isNull(column\_name)' to filter the null values in columns.

Filter settings | Optimize | Inspect | Data preview

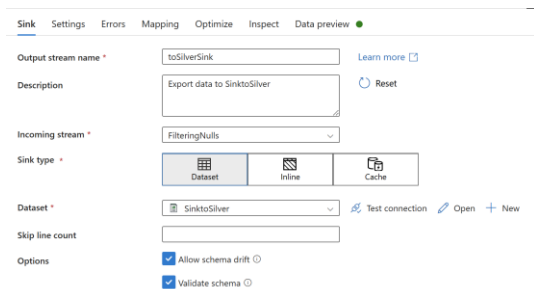
Output stream name \*  [Learn more](#)

Description  [Reset](#)

Incoming stream \*

Filter on \*

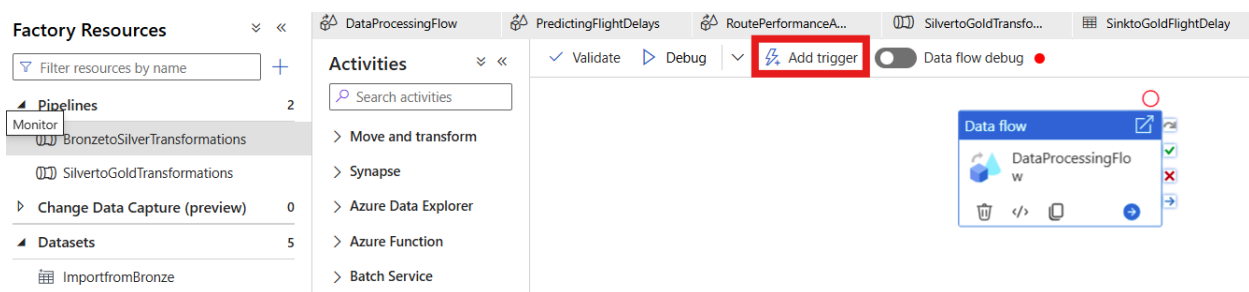
Step 5: In this step of the dataflow I ended the flow with a sink connecting the flow to the ‘SinktoSilver’ dataset, exporting the file in batch to the silver container.



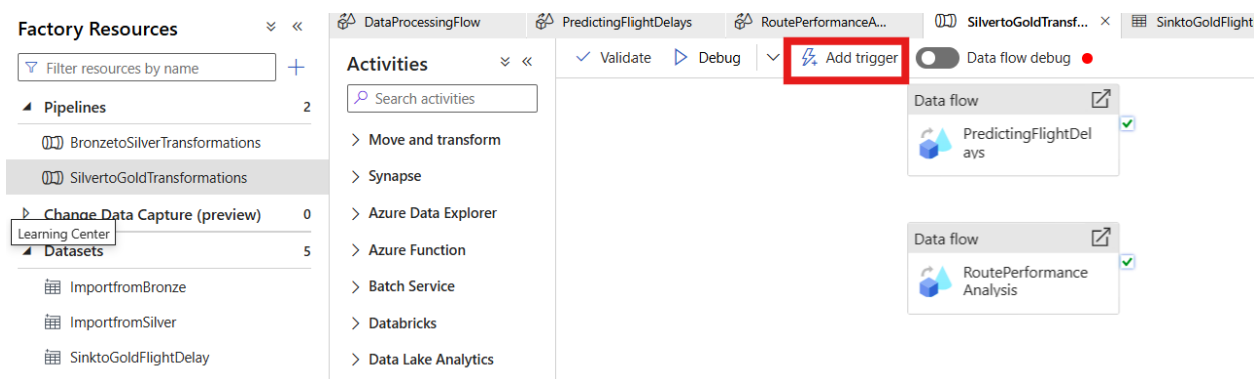
Step 6: Publish all changes.

Note: Starting the data preview helped to preview the data at each step to allow me to confirm the data processing steps before moving on.

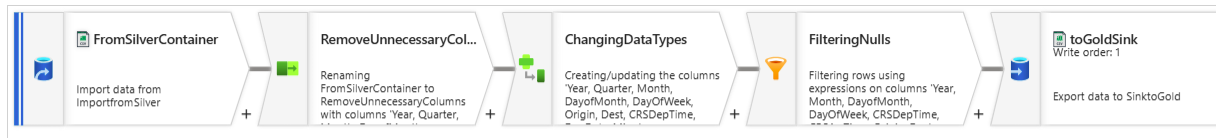
The pipeline run for transformation from bronze to silver (clicking the ‘Add trigger’ button will allow this pipeline to be triggered manually or setup to automatically trigger:



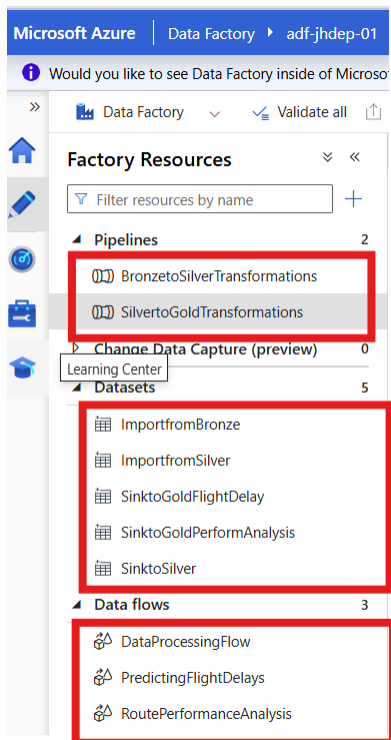
The pipeline run for transformation from silver to gold (clicking the ‘Add trigger’ button will allow this pipeline to be triggered manually or setup to automatically trigger:



The dataflow for each of these are clones of the data processing flow that have information changed on which columns to delete, datatypes, and assuring all nulls are filtered:

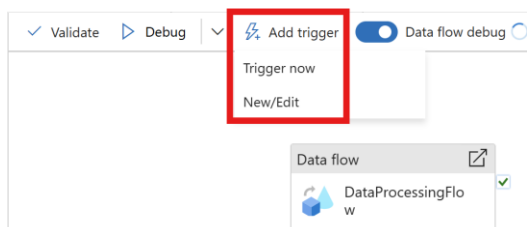


Listed below are the datasets, data flows, and pipelines created:



## PIPELINES IN DATA FACTORY:

To run a dataflow in Data Factory it is necessary to create a pipeline. This is easily done once the dataflow has been created. Simply select the dataflow (in my case 'DataProcessingFlow') from the 'Data Flows' from within the 'Author' tab and drag and drop it into the Pipelines builder. It is then easily possible to add triggers that are either specified by time or they can be manually triggered from the 'add trigger' button from the pipeline builder.





By pressing the trigger now button we can see the resulting output of 4 files into the silver container along with a success notification.

silver

Container

Search

Upload

+

Add Directory

Refresh

Rename

Delete

Change tier

Acquire lease

Break lease

Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: silver

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
.SUCCESS	11/20/2024, 3:37:26 PM	Hot (Inferred)		Block blob	0 B	Available
part-00000-c2db0340-7557-42e2-8a43-70f08270bd76-c000.csv	11/20/2024, 3:37:25 PM	Hot (Inferred)		Block blob	47.57 MiB	Available
part-00001-c2db0340-7557-42e2-8a43-70f08270bd76-c000.csv	11/20/2024, 3:37:26 PM	Hot (Inferred)		Block blob	47.67 MiB	Available
part-00002-c2db0340-7557-42e2-8a43-70f08270bd76-c000.csv	11/20/2024, 3:37:25 PM	Hot (Inferred)		Block blob	46.55 MiB	Available
part-00003-c2db0340-7557-42e2-8a43-70f08270bd76-c000.csv	11/20/2024, 3:37:25 PM	Hot (Inferred)		Block blob	45.14 MiB	Available

ADDITIONAL SCREENSHOTS

Dataflows setup information:

Factory Resources

Filter resources by name

Pipelines2

Change Data Capture (preview)0

Datasets5

Data flows3

Power Query0

ImportfromBronze

ImportfromSilver

SinktoGoldFlightDelay

SinktoGoldPerformAnalysis

SinktoSilver

DataProcessingFlow

PredictingFlightDelays

RoutePerformanceAnalysis

PredictingFlightDelays

RoutePerformanceA...

SilvertoGoldTransfo...

SinktoGoldFlightDelay

SinktoGold...

DelimitedText

ImportfromBronze

Connection

Schema

Parameters

Linked service \*

AzureDataLakeStorage1

Test connection

Edit

New

Learn more

File path

bronze

Directory

On\_Time\_Marketing\_Carrier...

Compression type

No compression

Column delimiter

Comma (,)

Row delimiter

Default (\r\n, or \n)

Encoding

Default(UTF-8)

Quote character

Double quote (")

Escape character

Backslash (\)

First row as header

✓

## Factory Resources

Filter resources by name

### Pipelines

BronzetoSilverTransformations

SilvertoGoldTransformations

### Change Data Capture (preview)

### Datasets

ImportfromBronze

ImportfromSilver

SinktoGoldFlightDelay

SinktoGoldPerformAnalysis

SinktoSilver

### Data flows

DataProcessingFlow

PredictingFlightDelays

RoutePerformanceAnalysis

### Power Query

SilvertoGoldTransfo...

SinktoGoldFlightDelay

SinktoGoldPerformA...

BronzetoSilverTrans...

Importfrom



DelimitedText  
**SinktoSilver**

#### Connection Schema Parameters

Linked service *	AzureDataLakeStorage1	Test connection	Edit	New	Learn more
File path	silver / Directory / File name				
Compression type	No compression				
Column delimiter	Comma (,)				
Row delimiter	Default (\r,\n, or \r\n)				
Encoding	Default(UTF-8)				
Quote character	Double quote (")				
Escape character	Backslash (\)				

## Factory Resources

Filter resources by name

### Pipelines

BronzetoSilverTransformations

SilvertoGoldTransformations

### Change Data Capture (preview)

### Datasets

ImportfromBronze

ImportfromSilver

SinktoGoldFlightDelay

SinktoGoldPerformAnalysis

SinktoSilver

### Data flows

DataProcessingFlow

PredictingFlightDelays

RoutePerformanceAnalysis

### Power Query

RoutePerformanceA...

SilvertoGoldTransfo...

SinktoGoldFlightDelay

SinktoGoldPerformA...

BronzetoSilverTrans...



DelimitedText  
**ImportfromSilver**

#### Connection Schema Parameters

Linked service *	AzureDataLakeStorage1	Test connection	Edit	New	Learn
File path	silver / Directory / File name				
Compression type	No compression				
Column delimiter	Comma (,)				
Row delimiter	Default (\r,\n, or \r\n)				
Encoding	Default(UTF-8)				
Quote character	Double quote (")				
Escape character	Backslash (\)				
First row as header	<input checked="" type="checkbox"/>				

Factory Resources

Filter resources by name

Pipelines2

BronzetoSilverTransformations

SilvertoGoldTransformations

Change Data Capture (preview)0

Datasets5

ImportfromBronze

ImportfromSilver

SinktoGoldFlightDelay...

SinktoGoldPerformAnalysis

SinktoSilver

Data flows3

DataProcessingFlow

PredictingFlightDelays

RoutePerformanceAnalysis

Power Query0

RoutePerformanceA...

SilvertoGoldTransfo...

SinktoGoldFlightDel...

SinktoGoldPerformA...

BronzetoS

DelimitedText

SinktoGoldFlightDelay

Connection

Schema

Parameters

Linked service \*AzureDataLakeStorage1Test connectionEdit+NewLearn mor

File pathgoldPredictingFlightDelaysFile name

Compression typeNo compression

Column delimiter ①Comma (,)

Row delimiter ①Default (\r,\n, or \r\n)

Encoding ①Default(UTF-8)

Quote character ①Double quote (")

Escape character ①Backslash (\)

Factory Resources

Authorer resources by name

Pipelines2

BronzetoSilverTransformations

SilvertoGoldTransformations

Change Data Capture (preview)0

Datasets5

ImportfromBronze

ImportfromSilver

SinktoGoldFlightDelay

SinktoGoldPerformAnalysis

SinktoSilver

Data flows3

DataProcessingFlow

PredictingFlightDelays

RoutePerformanceAnalysis

Power Query0

RoutePerformanceA...

SilvertoGoldTransfo...

SinktoGoldFlightDelay

SinktoGoldPerform...

BronzetoS

DelimitedText

SinktoGoldPerformAnalysis

Connection

Schema

Parameters

Linked service \*AzureDataLakeStorage1Test connectionEdit+NewLearn mor

File pathgoldRoutePerformanceAnalysisFile name

Compression typeNo compression

Column delimiter ①Comma (,)

Row delimiter ①Default (\r,\n, or \r\n)

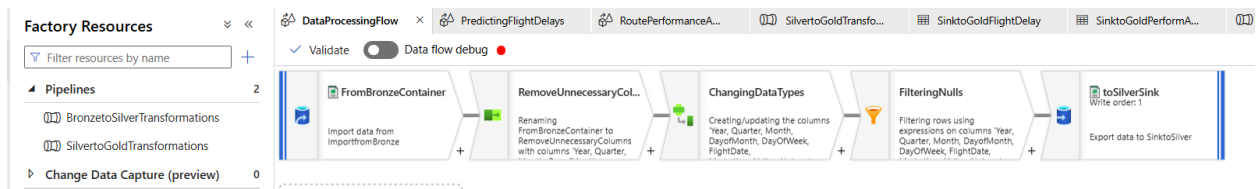
Encoding ①Default(UTF-8)

Quote character ①Double quote (")

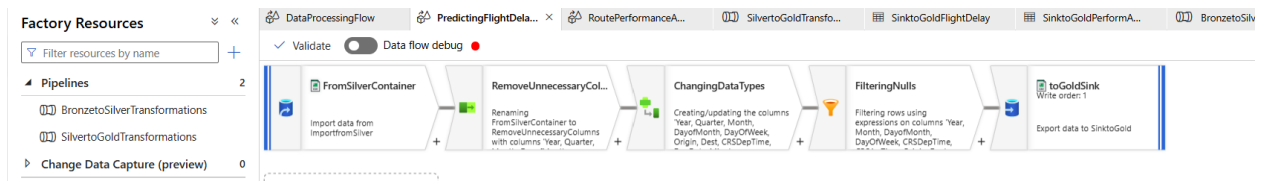
Escape character ①Backslash (\)

## Data Flows

DataProcessingFlow (Data Flow):



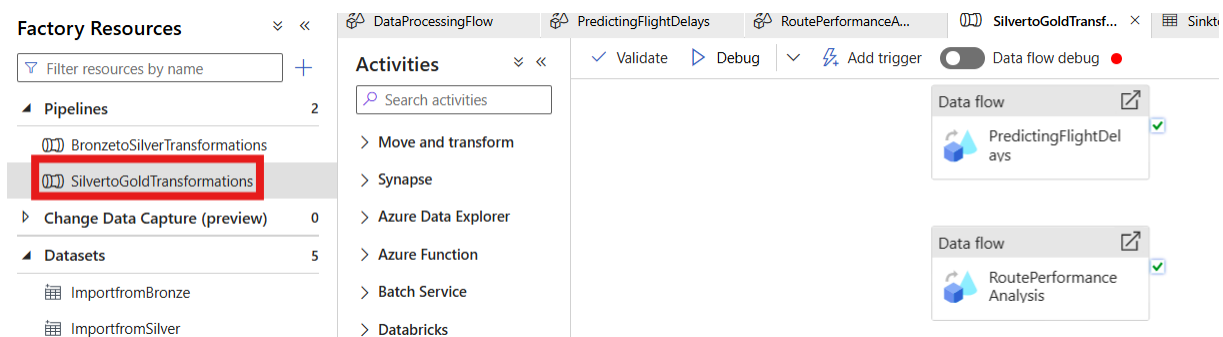
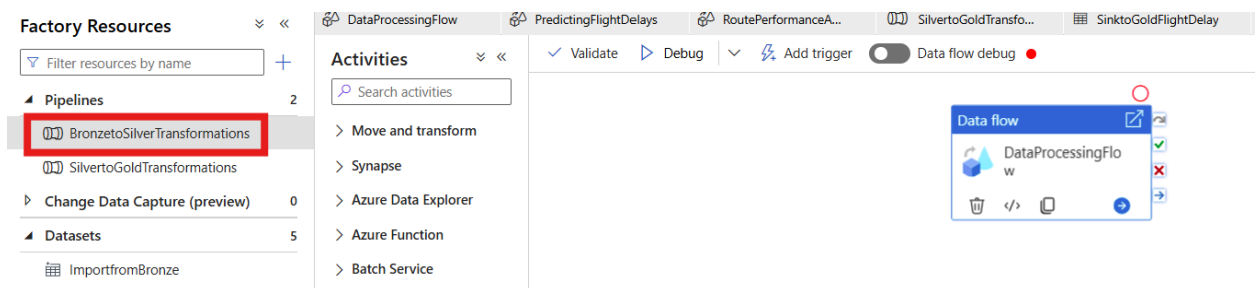
## PredictingFlightDelays (Data Flow):



## RoutePerformanceAnalysis (Data Flow):



## Pipelines










## Resources Used:

Microsoft Azure Search resources, services, and docs (G+7) Copilot

### Resources

Recent Favorite

Name	Type	Last Viewed
 datalakestorage01	Storage account	an hour ago
 jh-data-engineering-project	Resource group	2 days ago
 af-jhdep-01	Data factory (V2)	3 days ago
 ky-jhdep-01	Key vault	3 days ago
 af-jhdep-01	Function App	6 days ago
 db-jhdep-01	Azure Databricks Service	6 days ago
 datawhizzresources	Resource group	3 weeks ago

[See all](#)

Home > datalakestorage01

### datalakestorage01 | Containers

Storage account

Search

+ Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> Slogs	11/1/2024, 1:19:52 PM	Private	Available
<input type="checkbox"/> bronze	11/1/2024, 6:00:41 PM	Private	Available
<input type="checkbox"/> datalake	11/1/2024, 1:40:03 PM	Private	Available
<input type="checkbox"/> gold	11/1/2024, 6:01:19 PM	Private	Available
<input type="checkbox"/> silver	11/1/2024, 6:01:13 PM	Private	Available

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Partner solutions Data storage Containers

## Bronze container:

### bronze

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: bronze

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> On_Time_Marketing_Carrier_O...	11/1/2024, 6:02:03 PM	Hot (Inferred)		Block blob	275.31 MiB	Available ...
<input type="checkbox"/> readme.html	11/1/2024, 6:01:41 PM	Hot (Inferred)		Block blob	13.95 KiB	Available ...

## Silver container:

### silver

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)

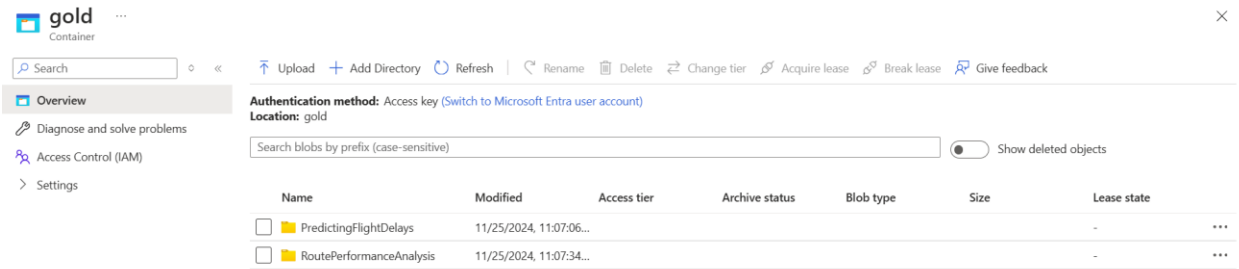
Location: silver

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> _SUCCESS	11/20/2024, 3:37:26 ...	Hot (Inferred)		Block blob	0 B	Available ...
<input type="checkbox"/> part-00000-c2db0340-7557-42...	11/20/2024, 3:37:25 ...	Hot (Inferred)		Block blob	47.57 MiB	Available ...
<input type="checkbox"/> part-00001-c2db0340-7557-42...	11/20/2024, 3:37:26 ...	Hot (Inferred)		Block blob	47.67 MiB	Available ...
<input type="checkbox"/> part-00002-c2db0340-7557-42...	11/20/2024, 3:37:25 ...	Hot (Inferred)		Block blob	46.55 MiB	Available ...
<input type="checkbox"/> part-00003-c2db0340-7557-42...	11/20/2024, 3:37:25 ...	Hot (Inferred)		Block blob	45.14 MiB	Available ...

## Gold container:



gold Container

Search

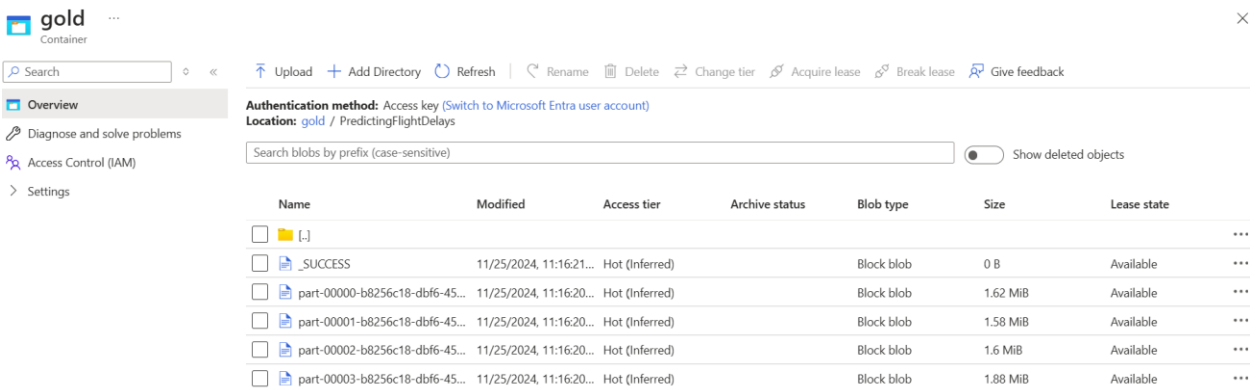
Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: gold

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
PredictingFlightDelays	11/25/2024, 11:07:06...					...
RoutePerformanceAnalysis	11/25/2024, 11:07:34...					...

## Gold container PredictingFlightDelays directory:



gold Container

Search

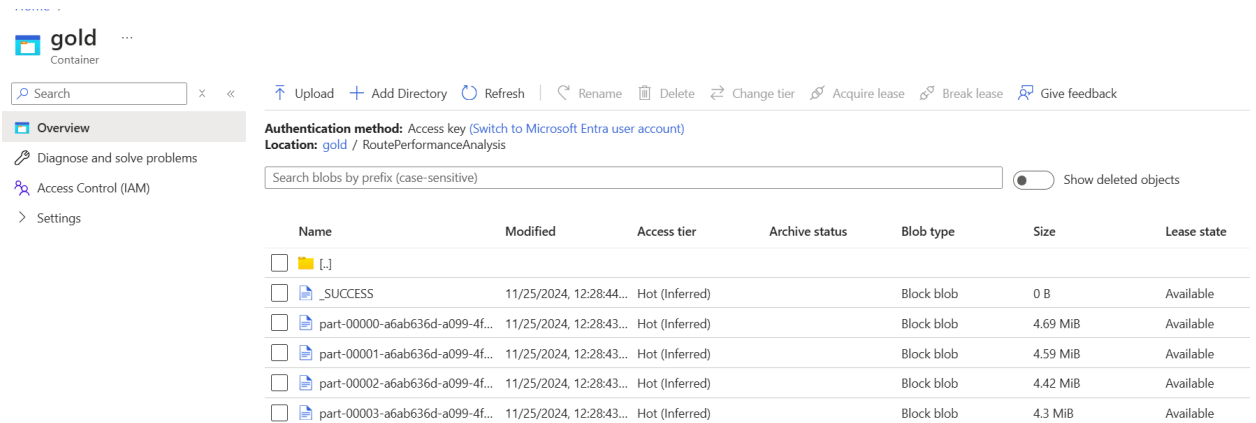
Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: gold / PredictingFlightDelays

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]						...
_SUCCESS	11/25/2024, 11:16:21...	Hot (Inferred)		Block blob	0 B	Available
part-00000-b8256c18-dbf6-45...	11/25/2024, 11:16:20...	Hot (Inferred)		Block blob	1.62 MiB	Available
part-00001-b8256c18-dbf6-45...	11/25/2024, 11:16:20...	Hot (Inferred)		Block blob	1.58 MiB	Available
part-00002-b8256c18-dbf6-45...	11/25/2024, 11:16:20...	Hot (Inferred)		Block blob	1.6 MiB	Available
part-00003-b8256c18-dbf6-45...	11/25/2024, 11:16:20...	Hot (Inferred)		Block blob	1.88 MiB	Available

## Gold container RoutePerformanceAnalysis directory:



gold Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

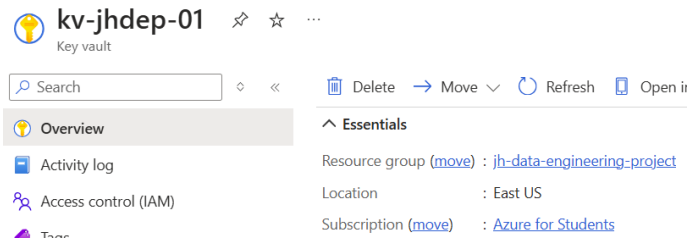
Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: gold / RoutePerformanceAnalysis

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]						...
_SUCCESS	11/25/2024, 12:28:44...	Hot (Inferred)		Block blob	0 B	Available
part-00000-a6ab636d-a099-4f...	11/25/2024, 12:28:43...	Hot (Inferred)		Block blob	4.69 MiB	Available
part-00001-a6ab636d-a099-4f...	11/25/2024, 12:28:43...	Hot (Inferred)		Block blob	4.59 MiB	Available
part-00002-a6ab636d-a099-4f...	11/25/2024, 12:28:43...	Hot (Inferred)		Block blob	4.42 MiB	Available
part-00003-a6ab636d-a099-4f...	11/25/2024, 12:28:43...	Hot (Inferred)		Block blob	4.3 MiB	Available

The first of the four files is posted on my GitHub at [https://github.com/Jhansen19/Airline\\_Predictions](https://github.com/Jhansen19/Airline_Predictions).

**Key vault** (I am not showing all this information for security reasons):



## RESOURCES:

1. Data Factory: [https://www.youtube.com/watch?v=Oggcq1euGJs&t=780s&ab\\_channel=BitsAnalytiCS](https://www.youtube.com/watch?v=Oggcq1euGJs&t=780s&ab_channel=BitsAnalytiCS)
2. Data Factory: [https://www.youtube.com/watch?v=ZILTV-BSAIE&t=2s&ab\\_channel=SkillCurb](https://www.youtube.com/watch?v=ZILTV-BSAIE&t=2s&ab_channel=SkillCurb)
3. Data Factory: [https://www.youtube.com/watch?v=Oggcq1euGJs&t=780s&ab\\_channel=BitsAnalytiCS](https://www.youtube.com/watch?v=Oggcq1euGJs&t=780s&ab_channel=BitsAnalytiCS)
4. Data Factory: [https://www.youtube.com/watch?v=Gt3XQ5Zx\\_iU&t=7s&ab\\_channel=AnalyticswithNags](https://www.youtube.com/watch?v=Gt3XQ5Zx_iU&t=7s&ab_channel=AnalyticswithNags)
5. Data Factory Transformations: [https://www.youtube.com/watch?v=75NTbT2QMaw&ab\\_channel=DataCafe](https://www.youtube.com/watch?v=75NTbT2QMaw&ab_channel=DataCafe)
6. Data Factory Filter, sort, sink Transformation: <https://www.youtube.com/playlist?list=PLcwrIWK7WBcTnbJludJBBHIKDvZaK3aQ>
7. Resources and Pipeline: [https://www.youtube.com/watch?v=ygJ11fzq\\_ik&ab\\_channel=Luke-CloudConsultant](https://www.youtube.com/watch?v=ygJ11fzq_ik&ab_channel=Luke-CloudConsultant)
8. Azure Data Factory: [https://youtube.com/watch?v=U5uYOM6j5II&ab\\_channel=LearnITEveryDay-AzureETLSolutionmadeEasy](https://youtube.com/watch?v=U5uYOM6j5II&ab_channel=LearnITEveryDay-AzureETLSolutionmadeEasy)
9. Mounting Azure Data Lake: [https://www.youtube.com/watch?v=ulueSLiQ8\\_o&ab\\_channel=SkillsPragati](https://www.youtube.com/watch?v=ulueSLiQ8_o&ab_channel=SkillsPragati)
10. Microsoft Documentation: <https://learn.microsoft.com/en-us/azure/?product=popular>