Retail Store Reporting Using Azure Data Engineering

Project Overview

This project is designed to generate **Revenue and Stock Wastage Reports** for an **FMCG Retail Company**. The **Revenue Reports** provide insights on a **monthly basis**, categorized by **product and city**. Additionally, the reporting includes **stock wastage analysis**, calculating the **financial loss** incurred due to different suppliers, which helps in determining the **supplier rating**.

The data pipeline is built using **Azure Data Factory (ADF)**, **Azure Data Lake Storage (ADLS)**, and **Azure Databricks Unity Catalog** for efficient data processing and management.

Requirement

The following reports need to be developed,

- Monthly Revenue for past one year.
- Top 5 City Revenue distribution.
- Revenue by different product category.
- Loss due to Stocky Wastage by different Supplier – The stock wastage is calculated by calculating the difference between the current day opening stock count and previous day closing stock count.
- Wastage over different days of the weekdays.
- List of people eligible for loyalty program – A loyalty score is calculated identifying the customer with most number of high value transactions over the period of one year

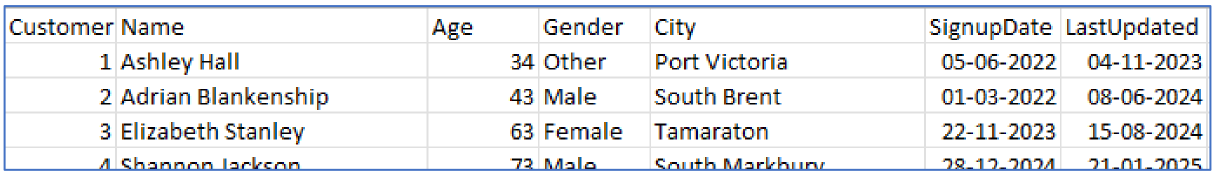
Tech Stack

- **Azure Data Factory (ADF)** – Data ingestion and orchestration.
- **Azure Data Lake Storage (ADLS)** – Raw, Bronze and Databricks unity catalog metastore.
- **Azure Databricks** – Data transformation (Silver Layer) and Aggregated data for reporting in Unity Catalog (Gold Layer).
- **Power BI** – Visualization and reporting.

Source Datasets Used:

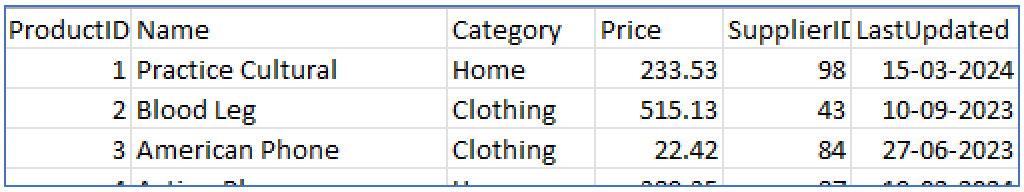
Dimension Datasets:

1. **Customer:**



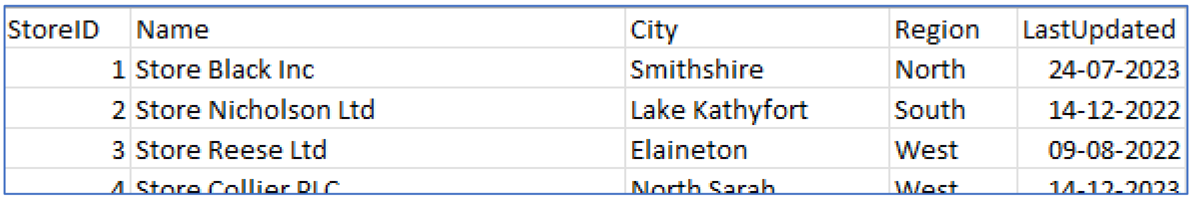
Customer	Name	Age	Gender	City	SignupDate	LastUpdated
1	Ashley Hall	34	Other	Port Victoria	05-06-2022	04-11-2023
2	Adrian Blankenship	43	Male	South Brent	01-03-2022	08-06-2024
3	Elizabeth Stanley	63	Female	Tamaraton	22-11-2023	15-08-2024
4	Shannon Jackson	73	Male	South Markbury	28-12-2024	21-01-2025

2. **Product**



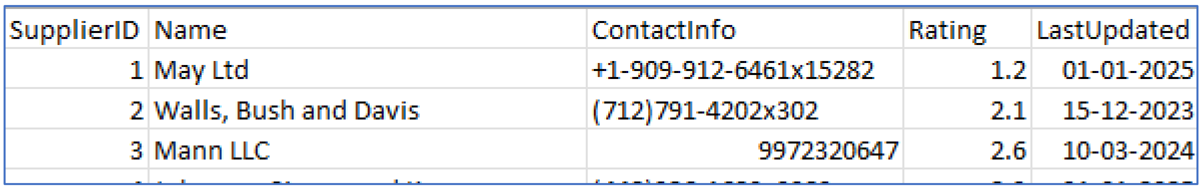
ProductID	Name	Category	Price	SupplierID	LastUpdated
1	Practice Cultural	Home	233.53	98	15-03-2024
2	Blood Leg	Clothing	515.13	43	10-09-2023
3	American Phone	Clothing	22.42	84	27-06-2023

3. **Store**



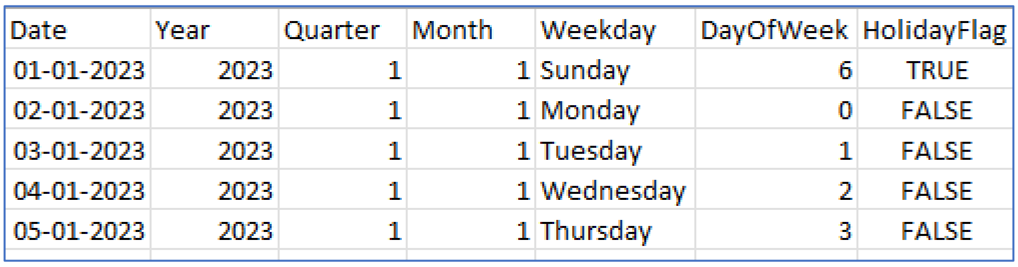
StoreID	Name	City	Region	LastUpdated
1	Store Black Inc	Smithshire	North	24-07-2023
2	Store Nicholson Ltd	Lake Kathyfort	South	14-12-2022
3	Store Reese Ltd	Elaineton	West	09-08-2022
4	Store Collier PLC	North Sarah	West	14-12-2023

4. **Supplier**



SupplierID	Name	ContactInfo	Rating	LastUpdated
1	May Ltd	+1-909-912-6461x15282	1.2	01-01-2025
2	Walls, Bush and Davis	(712)791-4202x302	2.1	15-12-2023
3	Mann LLC	9972320647	2.6	10-03-2024

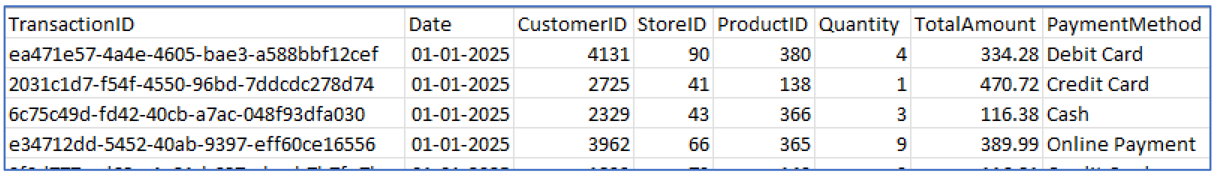
5. **Time**



Date	Year	Quarter	Month	Weekday	DayOfWeek	HolidayFlag
01-01-2023	2023	1	1	Sunday	6	TRUE
02-01-2023	2023	1	1	Monday	0	FALSE
03-01-2023	2023	1	1	Tuesday	1	FALSE
04-01-2023	2023	1	1	Wednesday	2	FALSE
05-01-2023	2023	1	1	Thursday	3	FALSE

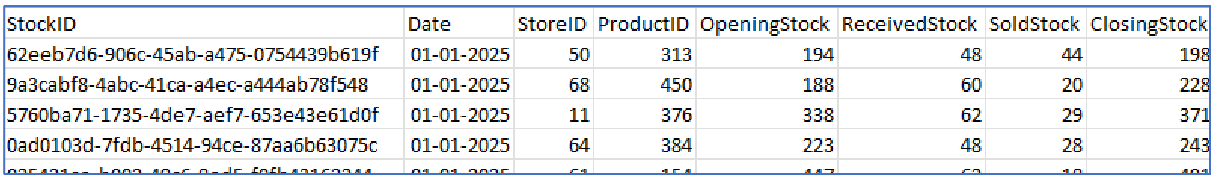
Transactional Datasets

1. **Sales Transactions**



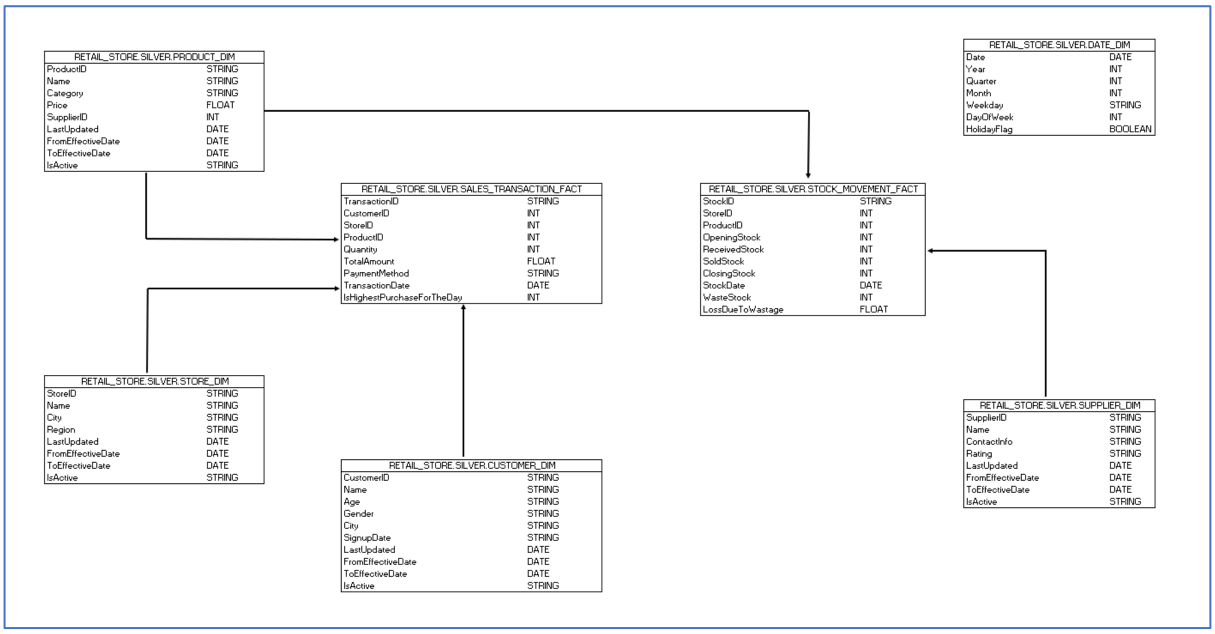
TransactionID	Date	CustomerID	StoreID	ProductID	Quantity	TotalAmount	PaymentMethod
ea471e57-4a4e-4605-bae3-a588bbf12cef	01-01-2025	4131	90	380	4	334.28	Debit Card
2031c1d7-f54f-4550-96bd-7ddcdc278d74	01-01-2025	2725	41	138	1	470.72	Credit Card
6c75c49d-fd42-40cb-a7ac-048f93dfa030	01-01-2025	2329	43	366	3	116.38	Cash
e34712dd-5452-40ab-9397-eff60ce16556	01-01-2025	3962	66	365	9	389.99	Online Payment

2. **Stock Movement**



StockID	Date	StoreID	ProductID	OpeningStock	ReceivedStock	SoldStock	ClosingStock
62eeb7d6-906c-45ab-a475-0754439b619f	01-01-2025	50	313	194	48	44	198
9a3cabf8-4abc-41ca-a4ec-a444ab78f548	01-01-2025	68	450	188	60	20	228
5760ba71-1735-4de7-aef7-653e43e61d0f	01-01-2025	11	376	338	62	29	371
0ad0103d-7fdb-4514-94ce-87aa6b63075c	01-01-2025	64	384	223	48	28	243

Data Model



RETAIL_STORE.SILVER.PRODUCT_DIM	
ProductID	STRING
Name	STRING
Category	STRING
Price	FLOAT
SupplierID	INT
LastUpdated	DATE
FromEffectiveDate	DATE
ToEffectiveDate	DATE
IsActive	STRING

RETAIL_STORE.SILVER.DATE_DIM	
Date	DATE
Year	INT
Quarter	INT
Month	INT
Weekday	STRING
DayOfWeek	INT
HolidayFlag	BOOLEAN

RETAIL_STORE.SILVER.SALES_TRANSACTION_FACT	
TransactionID	STRING
CustomerID	INT
StoreID	INT
ProductID	INT
Quantity	INT
TotalAmount	FLOAT
PaymentMethod	STRING
TransactionDate	DATE
IsHighestPurchaseForTheDay	INT

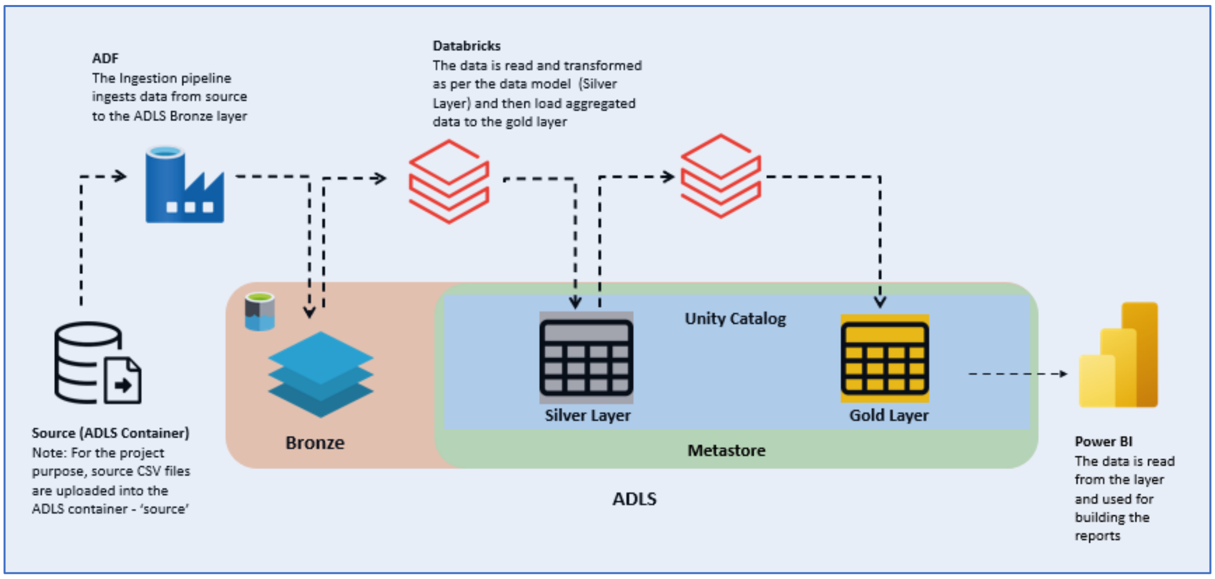
RETAIL_STORE.SILVER.STOCK_MOVEMENT_FACT	
StockID	STRING
StoreID	INT
ProductID	INT
OpeningStock	INT
ReceivedStock	INT
SoldStock	INT
ClosingStock	INT
StockDate	DATE
WasteStock	INT
LossDueToWastage	FLOAT

RETAIL_STORE.SILVER.STORE_DIM	
StoreID	STRING
Name	STRING
City	STRING
Region	STRING
LastUpdated	DATE
FromEffectiveDate	DATE
ToEffectiveDate	DATE
IsActive	STRING

RETAIL_STORE.SILVER.SUPPLIER_DIM	
SupplierID	STRING
Name	STRING
ContactInfo	STRING
Rating	STRING
LastUpdated	DATE
FromEffectiveDate	DATE
ToEffectiveDate	DATE
IsActive	STRING

RETAIL_STORE.SILVER.CUSTOMER_DIM	
CustomerID	STRING
Name	STRING
Age	STRING
Gender	STRING
City	STRING
SignupDate	STRING
LastUpdated	DATE
FromEffectiveDate	DATE
ToEffectiveDate	DATE
IsActive	STRING

Architecture


ADF
The Ingestion pipeline ingests data from source to the ADLS Bronze layer
Databricks
The data is read and transformed as per the data model (Silver Layer) and then load aggregated data to the gold layer
Unity Catalog
Silver Layer
Gold Layer
Bronze
Metastore
ADLS
Source (ADLS Container)
Note: For the project purpose, source CSV files are uploaded into the ADLS container - 'source'
Power BI
The data is read from the layer and used for building the reports

Detailed Implementation:

Azure Data Lake Storage:

The below containers are created for the respective usage,

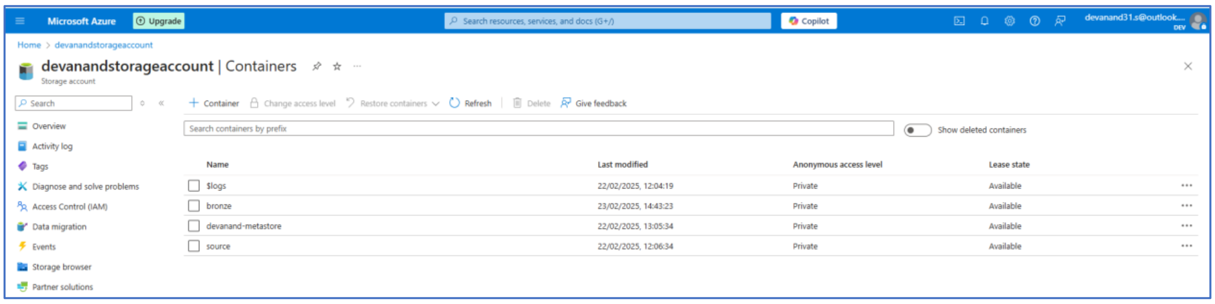
- Source – to store the raw csv source files.

NOTE: This action is implemented to mimic the data ingestion from source to ADLS

- Bronze – to store the parquet file ingested by ADF.

The transaction data files are stored in YYYY/MM/DD date folder format.

- devanand-metastore – Metastore for the Databricks to create the unity catalog.



devanandstorageaccount | Containers

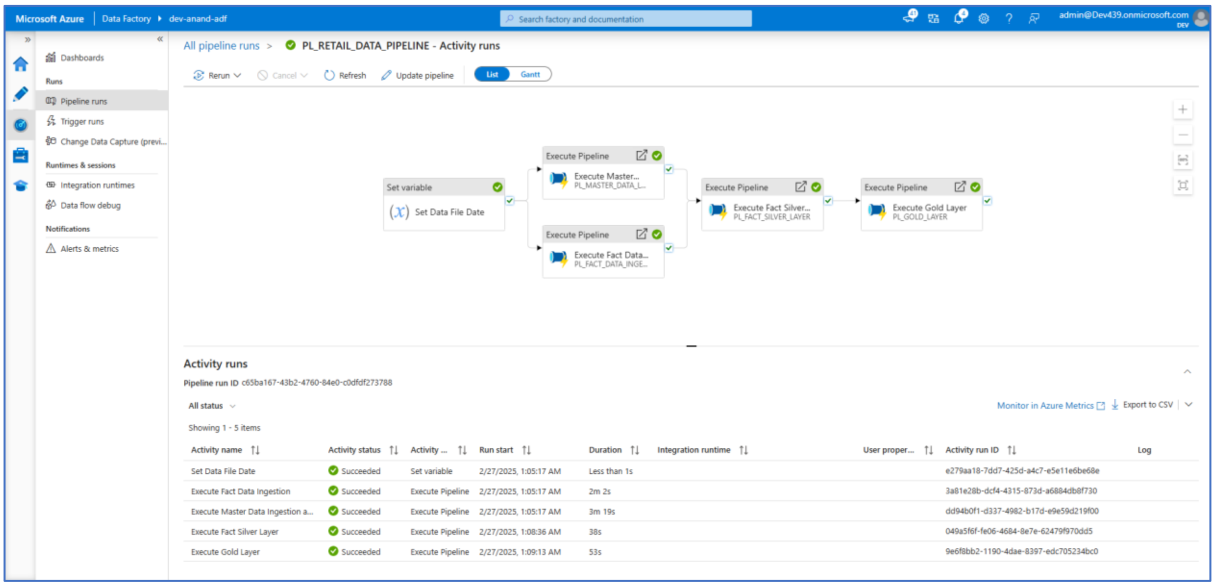
Storage account

Name	Last modified	Anonymous access level	Lease state
$logs	22/02/2025, 12:04:19	Private	Available
bronze	23/02/2025, 14:43:23	Private	Available
devanand-metastore	22/02/2025, 13:05:34	Private	Available
source	22/02/2025, 12:06:34	Private	Available

Azure Data Factory:

The ADF contains a main pipeline which starts with setting the load date (current date -1 day). It then executes Dimension data and Fact Data Ingestion and Curation pipelines.

PL_RETAIL_DATA_PIPELINE (main pipeline):



All pipeline runs > PL_RETAIL_DATA_PIPELINE - Activity runs

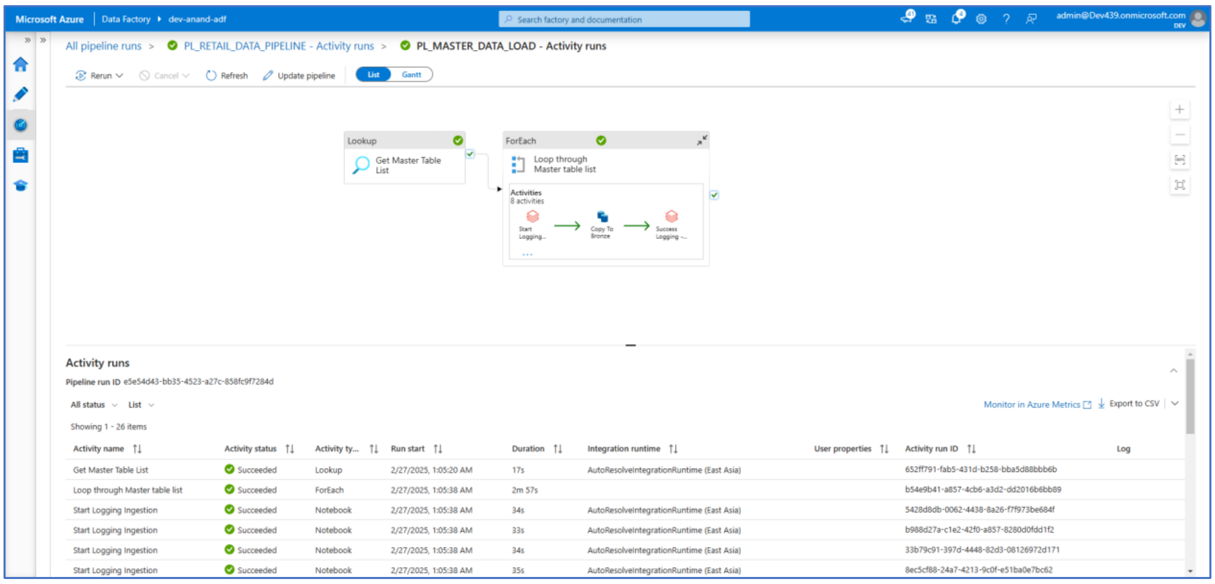
Activity runs

Pipeline run ID c65ba167-43b2-4760-84e0-c0dfdf273788

Showing 1 - 5 items

Activity name	Activity status	Activity ...	Run start	Duration	Integration runtime	User proper...	Activity run ID	Log
Set Data File Date	Succeeded	Set variable	2/27/2025, 1:05:17 AM	Less than 1s			e279aa18-7dd7-425d-a4c7-e5e11e6be68e	
Execute Fact Data Ingestion	Succeeded	Execute Pipeline	2/27/2025, 1:05:17 AM	2m 2s			3a81e28b-dcf4-4315-873d-a6884db8f730	
Execute Master Data Ingestion a...	Succeeded	Execute Pipeline	2/27/2025, 1:05:17 AM	3m 19s			dd94b0f1-d337-4982-b17d-e9e59d219f00	
Execute Fact Silver Layer	Succeeded	Execute Pipeline	2/27/2025, 1:08:36 AM	38s			049a5f6f-fe06-4684-8e7e-62479f970dd5	
Execute Gold Layer	Succeeded	Execute Pipeline	2/27/2025, 1:09:13 AM	53s			9e6f8bb2-1190-4dae-8397-edc705234bc0	

PL_MASTER_DATA_LOAD: Ingests all the dimension data on Incremental basis.



All pipeline runs > PL_RETAIL_DATA_PIPELINE - Activity runs > PL_MASTER_DATA_LOAD - Activity runs

Activity runs

Pipeline run ID e5e54d43-bb35-4523-a27c-858fc9f7284d

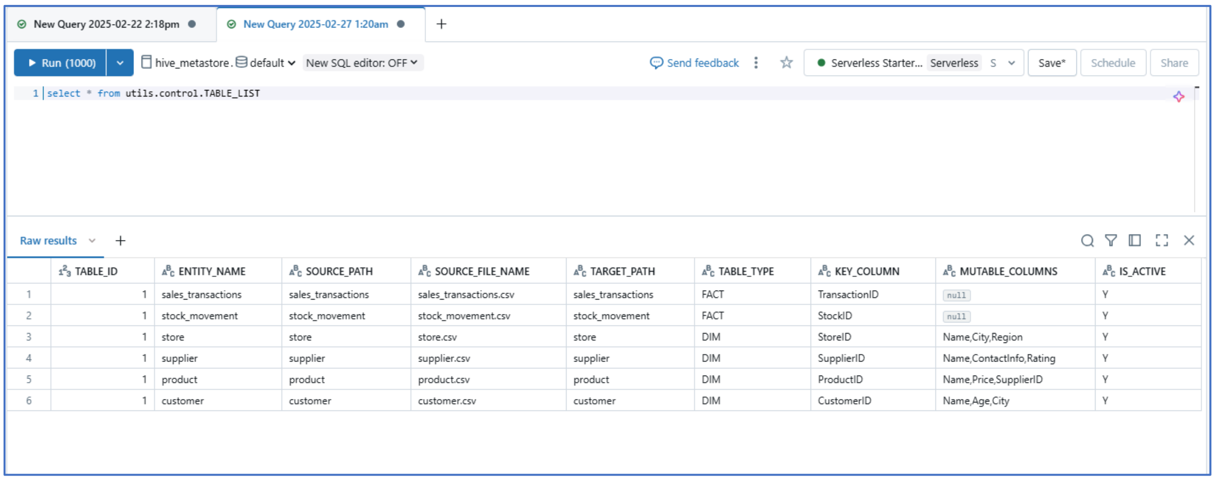
Showing 1 - 26 items

Activity name	Activity status	Activity ty...	Run start	Duration	Integration runtime	User properties	Activity run ID	Log
Get Master Table List	Succeeded	Lookup	2/27/2025, 1:05:20 AM	17s	AutoResolveIntegrationRuntime (East Asia)		652ff791-fab5-431d-b258-bba5d88bbb6b	
Loop through Master table list	Succeeded	ForEach	2/27/2025, 1:05:38 AM	2m 57s			b54e9b41-a857-4cb6-a3d2-dd2016b6bb89	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:38 AM	34s	AutoResolveIntegrationRuntime (East Asia)		5428d8db-0062-4438-8a26-f7f973be684f	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:38 AM	33s	AutoResolveIntegrationRuntime (East Asia)		b988d27a-c1e2-42f0-a857-8280d0fdd1f2	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:38 AM	34s	AutoResolveIntegrationRuntime (East Asia)		33b79c91-397d-4448-82d3-08126972d171	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:38 AM	35s	AutoResolveIntegrationRuntime (East Asia)		8ec5cf88-24a7-4213-9c0f-e51ba0e7bc62	

Time dimension is populated programmatically.

1. The Data from the source is first ingested in the bronze layer.
2. The Silver Layer notebook will read the data, performs cleanups and load the data to the silver layer tables on SCD2 basis.

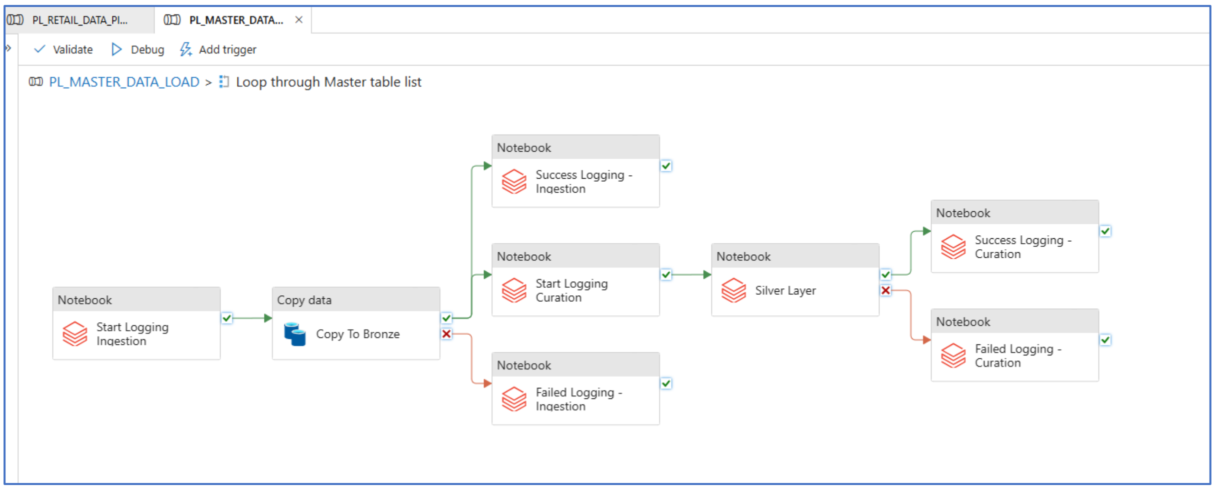
The above two steps are run in a loop with all the entities passed dynamically to the for each activity using the configuration table.



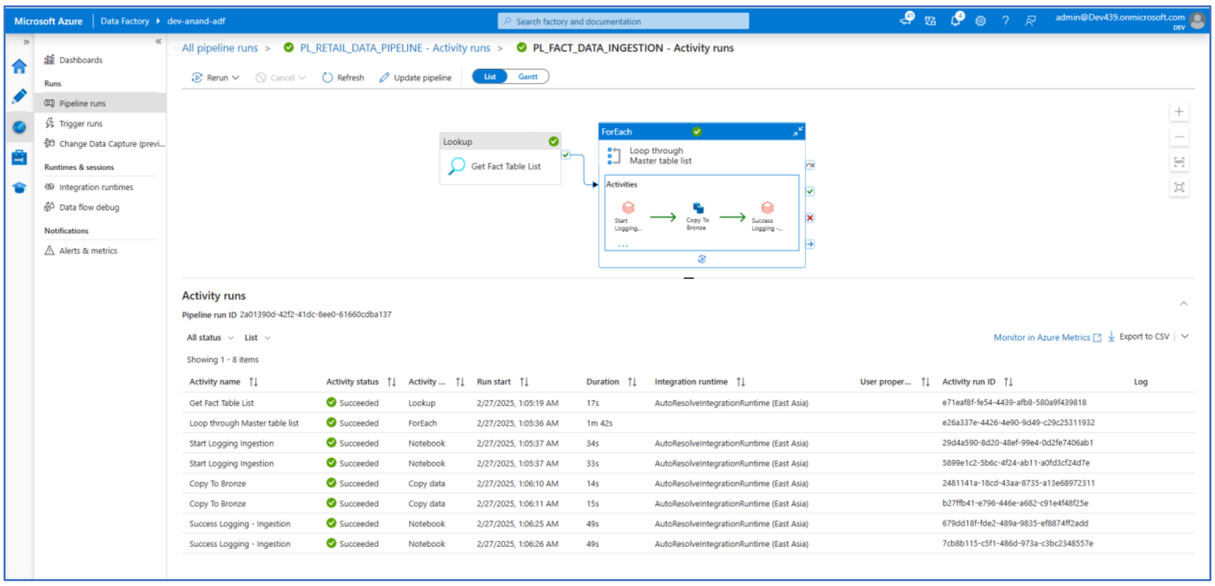
```sql
select * from utils.control.TABLE_LIST
```

	TABLE_ID	ENTITY_NAME	SOURCE_PATH	SOURCE_FILE_NAME	TARGET_PATH	TABLE_TYPE	KEY_COLUMN	MUTABLE_COLUMNS	IS_ACTIVE
1	1	sales_transactions	sales_transactions	sales_transactions.csv	sales_transactions	FACT	TransactionID	null	Y
2	1	stock_movement	stock_movement	stock_movement.csv	stock_movement	FACT	StockID	null	Y
3	1	store	store	store.csv	store	DIM	StoreID	Name,City,Region	Y
4	1	supplier	supplier	supplier.csv	supplier	DIM	SupplierID	Name,ContactInfo,Rating	Y
5	1	product	product	product.csv	product	DIM	ProductID	Name,Price,SupplierID	Y
6	1	customer	customer	customer.csv	customer	DIM	CustomerID	Name,Age,City	Y

The pipeline also performs the necessary logging. The Silver layer notebook is also developed in the form of framework to process each entity dynamically which will be discussed below in the Databricks sections.



PL_FACT_DATA_INGESTION: The Fact table load is also made dynamic. The fact tables are configured in the above configuration table. The load is incremental as the data at the source are stored in YYYY/MM/DD folder format


Microsoft Azure | Data Factory ▸ dev-anand-adf
Search factory and documentation
admin@Dev439.onmicrosoft.com
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
All pipeline runs > PL_RETAIL_DATA_PIPELINE - Activity runs > PL_FACT_DATA_INGESTION - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Lookup
Get Fact Table List
ForEach
Loop through Master table list
Activities
Start Logging...
Copy To Bronze
Success Logging -...

Activity runs

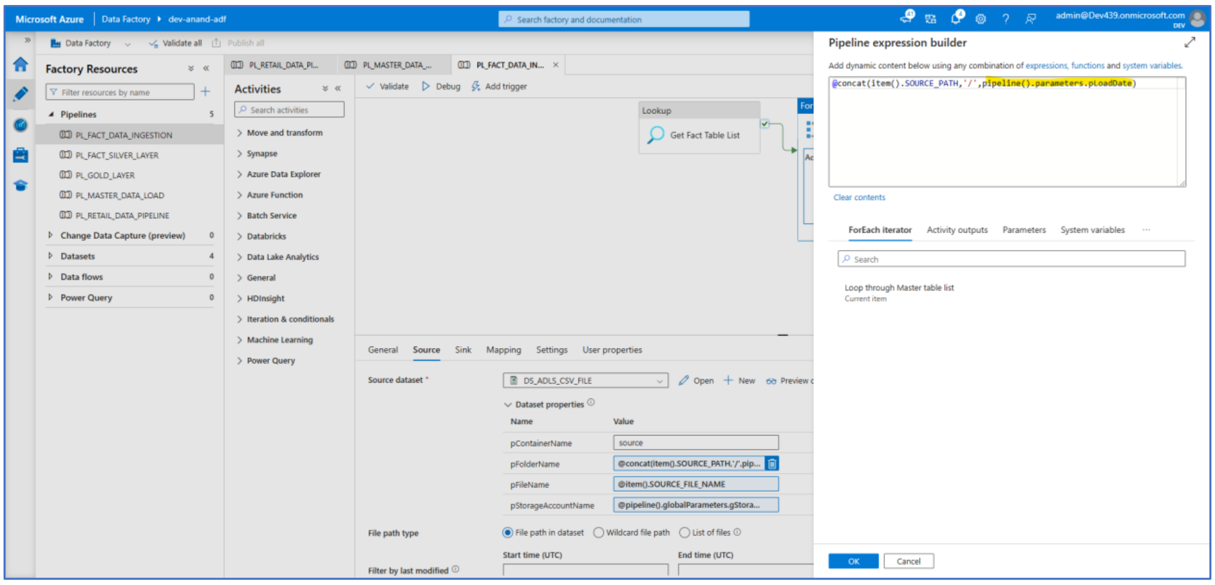
Pipeline run ID 2a01390d-42f2-41dc-8ee0-61660cdba137

All status | List

Showing 1 - 8 items

Activity name	Activity status	Activity ...	Run start	Duration	Integration runtime	User proper...	Activity run ID	Log
Get Fact Table List	Succeeded	Lookup	2/27/2025, 1:05:19 AM	17s	AutoResolveIntegrationRuntime (East Asia)		e71eaf8f-fe54-4439-afb8-580a9f439818	
Loop through Master table list	Succeeded	ForEach	2/27/2025, 1:05:36 AM	1m 42s			e26a337e-4426-4e90-9d49-c29c25311932	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:37 AM	34s	AutoResolveIntegrationRuntime (East Asia)		29d4a590-8d20-48ef-99e4-0d2fe7406ab1	
Start Logging Ingestion	Succeeded	Notebook	2/27/2025, 1:05:37 AM	33s	AutoResolveIntegrationRuntime (East Asia)		5899e1c2-5b6c-4f24-ab11-a0fd3cf24d7e	
Copy To Bronze	Succeeded	Copy data	2/27/2025, 1:06:10 AM	14s	AutoResolveIntegrationRuntime (East Asia)		2481141a-18cd-43aa-8735-a13e68972311	
Copy To Bronze	Succeeded	Copy data	2/27/2025, 1:06:11 AM	15s	AutoResolveIntegrationRuntime (East Asia)		b27ffb41-e796-446e-a682-c91e4f48f25e	
Success Logging - Ingestion	Succeeded	Notebook	2/27/2025, 1:06:25 AM	49s	AutoResolveIntegrationRuntime (East Asia)		679dd18f-fde2-489a-9835-ef8874ff2add	
Success Logging - Ingestion	Succeeded	Notebook	2/27/2025, 1:06:26 AM	49s	AutoResolveIntegrationRuntime (East Asia)		7cb8b115-c5f1-486d-973a-c3bc2348557e	

Incremental Logic:



Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
@concat(item().SOURCE_PATH,'/',pipeline().parameters.pLoadDate)
```

Clear contents

Source dataset * DS_ADLS_CSV_FILE | Open | New | Preview

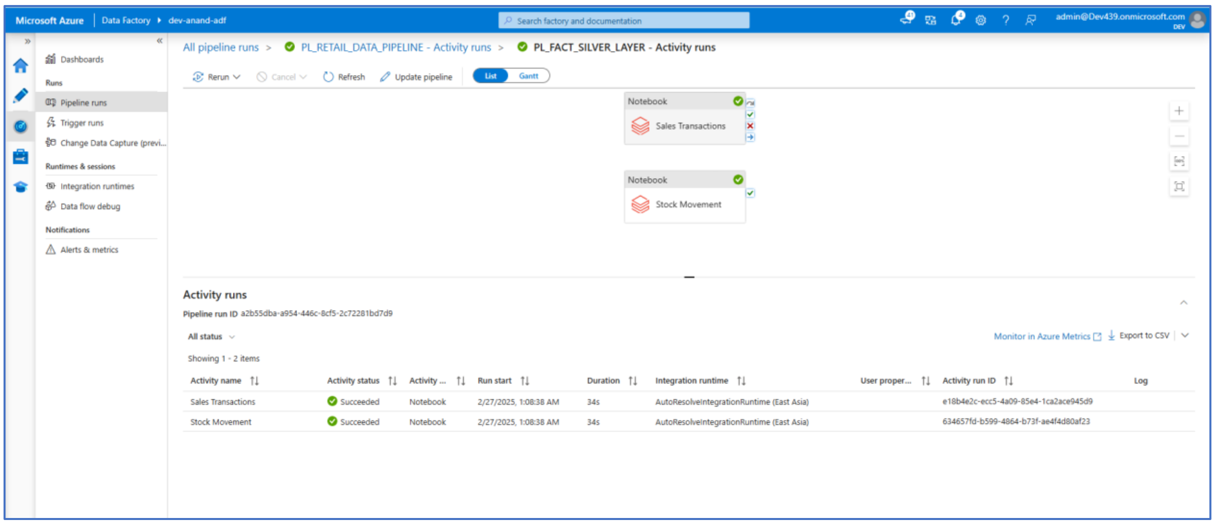
Dataset properties

Name	Value
pContainerName	source
pFolderName	@concat(item().SOURCE_PATH,'/',pip...
pFileName	@item().SOURCE_FILE_NAME
pStorageAccountName	@pipeline().globalParameters.gStora...

File path type: [✓ File path in dataset] Wildcard file path List of files

Filter by last modified: Start time (UTC) | End time (UTC)

PL_FACT_SILVER_LAYER – Then the fact tables are processed and loaded to the silver tables


Microsoft Azure | Data Factory ▸ dev-anand-adf
Search factory and documentation
admin@Dev439.onmicrosoft.com
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
All pipeline runs > PL_RETAIL_DATA_PIPELINE - Activity runs > PL_FACT_SILVER_LAYER - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Notebook
Sales Transactions
Notebook
Stock Movement

Activity runs

Pipeline run ID a2b55dba-a954-446c-8cf5-2c72281bd7d9

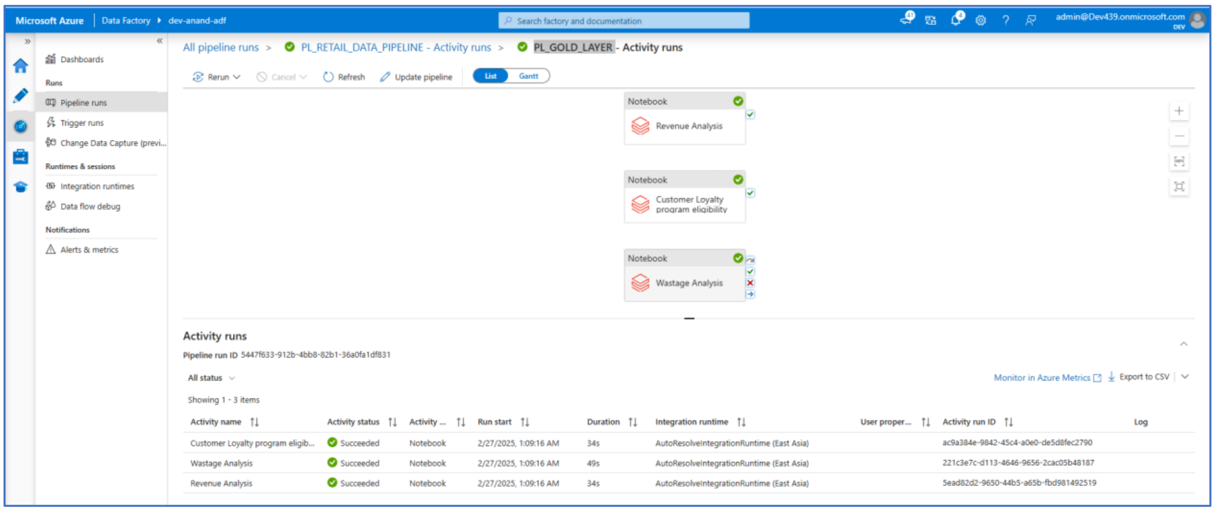
All status

Monitor in Azure Metrics | Export to CSV

Showing 1 - 2 items

Activity name	Activity status	Activity ...	Run start	Duration	Integration runtime	User proper...	Activity run ID	Log
Sales Transactions	Succeeded	Notebook	2/27/2025, 1:08:38 AM	34s	AutoResolveIntegrationRuntime (East Asia)		e18b4e2c-ecc5-4a09-85e4-1ca2ace945d9	
Stock Movement	Succeeded	Notebook	2/27/2025, 1:08:38 AM	34s	AutoResolveIntegrationRuntime (East Asia)		634657fd-b599-4864-b73f-ae4f4d80af23	

PL_GOLD_LAYER: The final pipeline will aggregate the data as per report requirement and load the data on full load basis to the gold layer tables.


Microsoft Azure | Data Factory ▸ dev-anand-adf
Search factory and documentation
admin@Dev439.onmicrosoft.com
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
All pipeline runs > PL_RETAIL_DATA_PIPELINE - Activity runs > PL_GOLD_LAYER - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Notebook
Revenue Analysis
Notebook
Customer Loyalty program eligibility
Notebook
Wastage Analysis

Activity runs

Pipeline run ID 5447f633-912b-4bb8-82b1-36a0fa1df831

All status

Monitor in Azure Metrics | Export to CSV

Showing 1 - 3 items

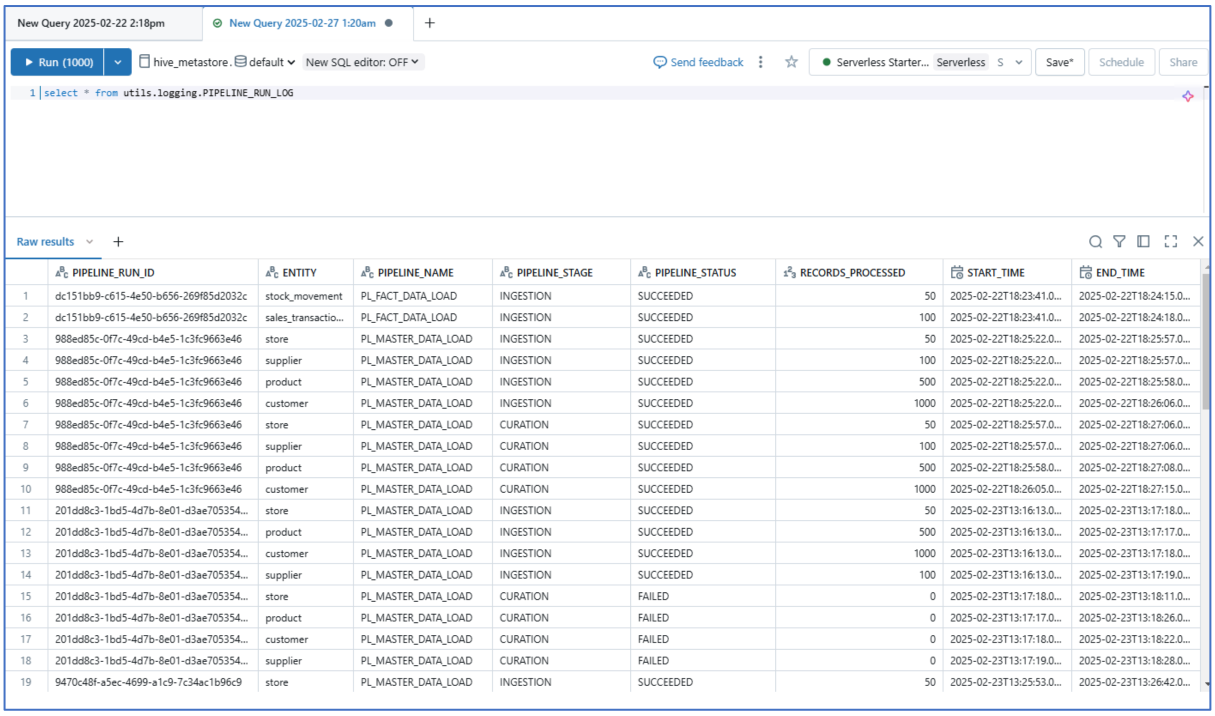
Activity name	Activity status	Activity ...	Run start	Duration	Integration runtime	User proper...	Activity run ID	Log
Customer Loyalty program eligib...	Succeeded	Notebook	2/27/2025, 1:09:16 AM	34s	AutoResolveIntegrationRuntime (East Asia)		ac9a384e-9842-45c4-a0e0-de5d8fec2790	
Wastage Analysis	Succeeded	Notebook	2/27/2025, 1:09:16 AM	49s	AutoResolveIntegrationRuntime (East Asia)		221c3e7c-d113-4646-9656-2cac05b48187	
Revenue Analysis	Succeeded	Notebook	2/27/2025, 1:09:16 AM	34s	AutoResolveIntegrationRuntime (East Asia)		5ead82d2-9650-44b5-a65b-fbd981492519	

Databricks:

Utils:

- Pipeline Logging Start - This notebook is called before the copy/notebook activity to start the logging.
- Pipeline Logging End – This notebook is called once the copy/notebook activity is completed to update the status of the activity.

The log data is stored in the below table,



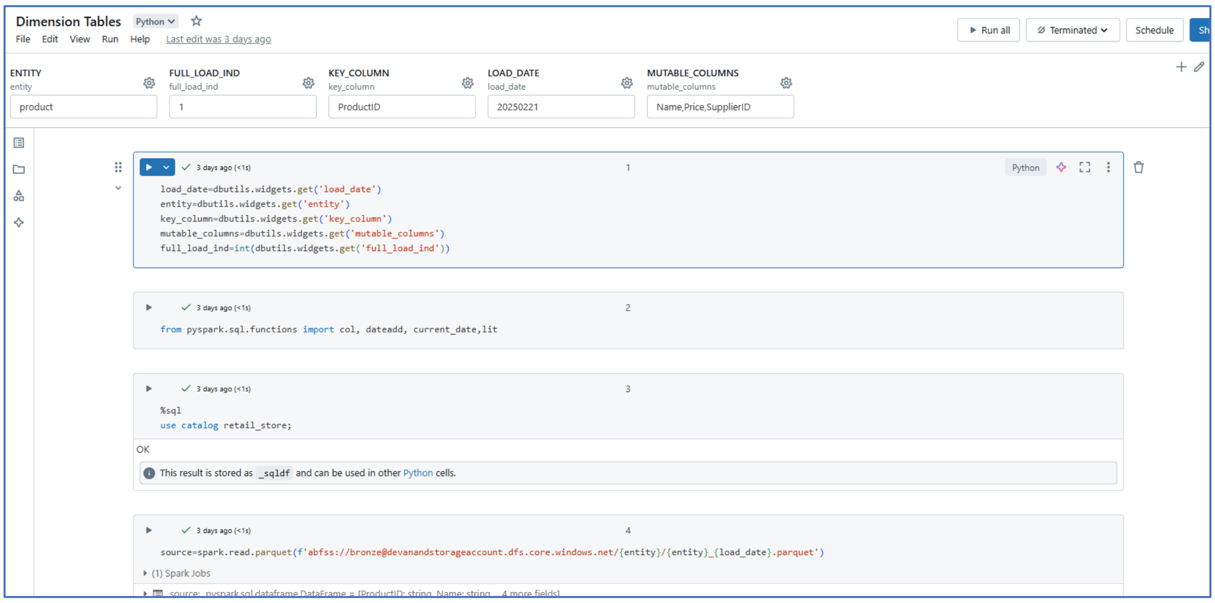
```sql
select * from utils.logging.PIPELINE_RUN_LOG
```

	PIPELINE_RUN_ID	ENTITY	PIPELINE_NAME	PIPELINE_STAGE	PIPELINE_STATUS	RECORDS_PROCESSED	START_TIME	END_TIME
1	dc151bb9-c615-4e50-b656-269f85d2032c	stock_movement	PL_FACT_DATA_LOAD	INGESTION	SUCCEEDED	50	2025-02-22T18:23:41.0...	2025-02-22T18:24:15.0...
2	dc151bb9-c615-4e50-b656-269f85d2032c	sales_transactio...	PL_FACT_DATA_LOAD	INGESTION	SUCCEEDED	100	2025-02-22T18:23:41.0...	2025-02-22T18:24:18.0...
3	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	store	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	50	2025-02-22T18:25:22.0...	2025-02-22T18:25:57.0...
4	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	supplier	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	100	2025-02-22T18:25:22.0...	2025-02-22T18:25:57.0...
5	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	product	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	500	2025-02-22T18:25:22.0...	2025-02-22T18:25:58.0...
6	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	customer	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	1000	2025-02-22T18:25:22.0...	2025-02-22T18:26:06.0...
7	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	store	PL_MASTER_DATA_LOAD	CURATION	SUCCEEDED	50	2025-02-22T18:25:57.0...	2025-02-22T18:27:06.0...
8	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	supplier	PL_MASTER_DATA_LOAD	CURATION	SUCCEEDED	100	2025-02-22T18:25:57.0...	2025-02-22T18:27:06.0...
9	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	product	PL_MASTER_DATA_LOAD	CURATION	SUCCEEDED	500	2025-02-22T18:25:58.0...	2025-02-22T18:27:08.0...
10	988ed85c-0f7c-49cd-b4e5-1c3fc9663e46	customer	PL_MASTER_DATA_LOAD	CURATION	SUCCEEDED	1000	2025-02-22T18:26:05.0...	2025-02-22T18:27:15.0...
11	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	store	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	50	2025-02-23T13:16:13.0...	2025-02-23T13:17:18.0...
12	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	product	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	500	2025-02-23T13:16:13.0...	2025-02-23T13:17:17.0...
13	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	customer	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	1000	2025-02-23T13:16:13.0...	2025-02-23T13:17:18.0...
14	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	supplier	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	100	2025-02-23T13:16:13.0...	2025-02-23T13:17:19.0...
15	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	store	PL_MASTER_DATA_LOAD	CURATION	FAILED	0	2025-02-23T13:17:18.0...	2025-02-23T13:18:11.0...
16	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	product	PL_MASTER_DATA_LOAD	CURATION	FAILED	0	2025-02-23T13:17:17.0...	2025-02-23T13:18:26.0...
17	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	customer	PL_MASTER_DATA_LOAD	CURATION	FAILED	0	2025-02-23T13:17:18.0...	2025-02-23T13:18:22.0...
18	201dd8c3-1bd5-4d7b-8e01-d3ae705354...	supplier	PL_MASTER_DATA_LOAD	CURATION	FAILED	0	2025-02-23T13:17:19.0...	2025-02-23T13:18:28.0...
19	9470c48f-a5ec-4699-a1c9-7c34ac1b96c9	store	PL_MASTER_DATA_LOAD	INGESTION	SUCCEEDED	50	2025-02-23T13:25:53.0...	2025-02-23T13:26:42.0...

Silver Layer:

- Dimension Tables – This notebook dynamically accepts the entity and the respective entity's key column and mutable columns and loads the data into the silver layer tables in SCD2 mode.

Dimension dynamic notebook:



ENTITY	FULL_LOAD_IND	KEY_COLUMN	LOAD_DATE	MUTABLE_COLUMNS
product	1	ProductID	20250221	Name,Price,SupplierID

```python
load_date=dbutils.widgets.get('load_date')
entity=dbutils.widgets.get('entity')
key_column=dbutils.widgets.get('key_column')
mutable_columns=dbutils.widgets.get('mutable_columns')
full_load_ind=int(dbutils.widgets.get('full_load_ind'))
```

```python
from pyspark.sql.functions import col, dateadd, current_date,lit
```

```
%sql
use catalog retail_store;
```

```
OK
This result is stored as _sqldf and can be used in other Python cells.
```

```python
source=spark.read.parquet(f'abfss://bronze@devanandstorageaccount.dfs.core.windows.net/{entity}/{entity}_{load_date}.parquet')
```

```
(1) Spark Jobs
source: pyspark.sql.dataframe.DataFrame = [ProductID: string, Name: string ... 4 more fields]
```

- Sales Transactions – The sales transaction fact table is processed in this notebook. The highest transaction for each day is identified here to calculate the customer eligibility score for loyalty programs

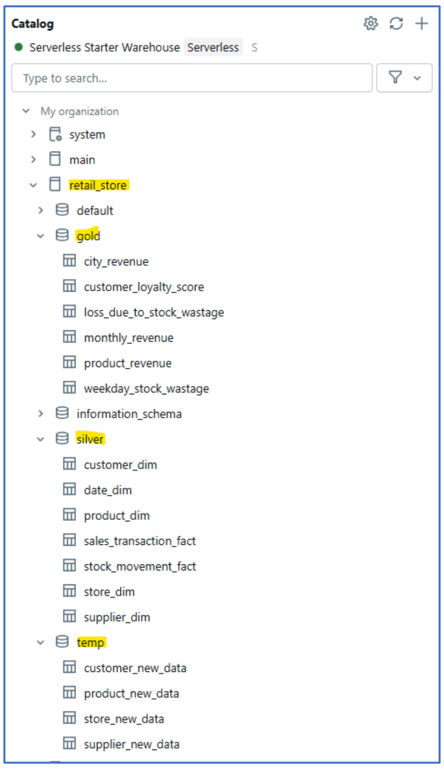
- Stock Movement – This Stock Movement fact table is processed to calculate the wastage on each day for each product.

Gold Layer:

- Revenue Analysis – This notebook calculates the Monthly revenue, Revenue by product category and revenue by cities
- Wastage Analysis – This notebook calculates the wastage on a daily basis - the difference between the current day opening stock count and previous day closing stock count
- Customer Loyalty program eligibility – This notebook calculates the score for each customers and picks the top 10 for loyalty programs.

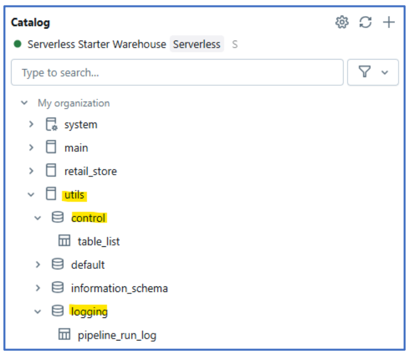
Unity Catalog:

- The Retail Store Catalog has the silver schema, gold schema and temp schema (required for curation process)



Catalog
Serverless Starter Warehouse Serverless S
Type to search...
My organization
system
main
retail_store
default
gold
city_revenue
customer_loyalty_score
loss_due_to_stock_wastage
monthly_revenue
product_revenue
weekday_stock_wastage
information_schema
silver
customer_dim
date_dim
product_dim
sales_transaction_fact
stock_movement_fact
store_dim
supplier_dim
temp
customer_new_data
product_new_data
store_new_data
supplier_new_data

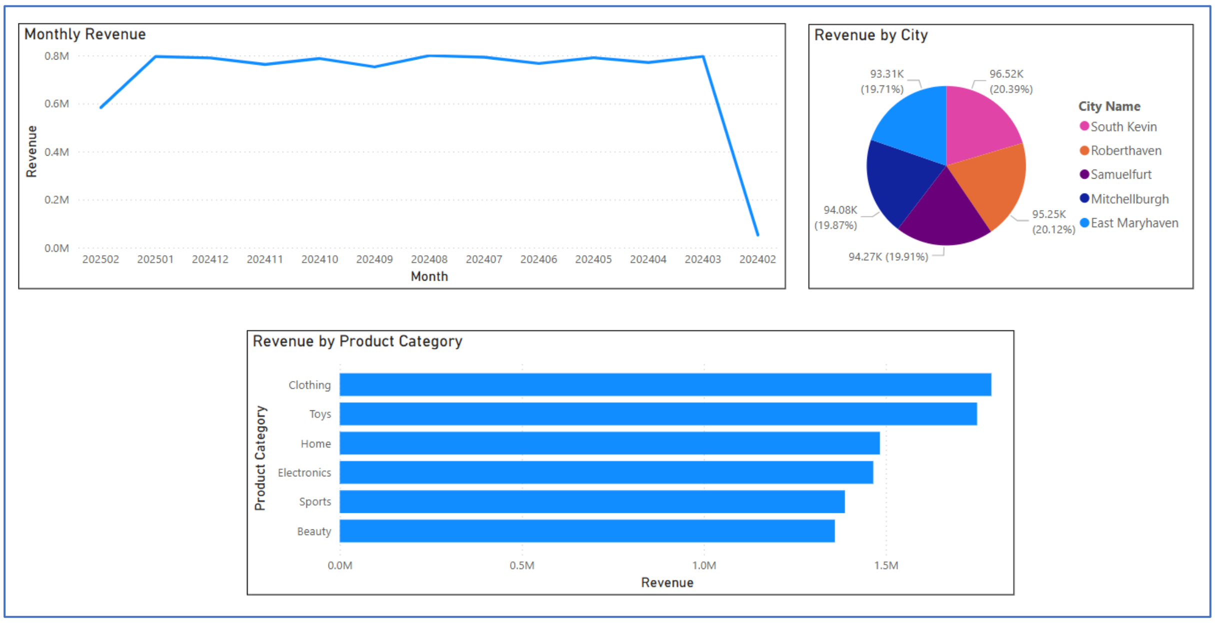
Then there is utils catalog which has logging schema and control schema (for configuration table).

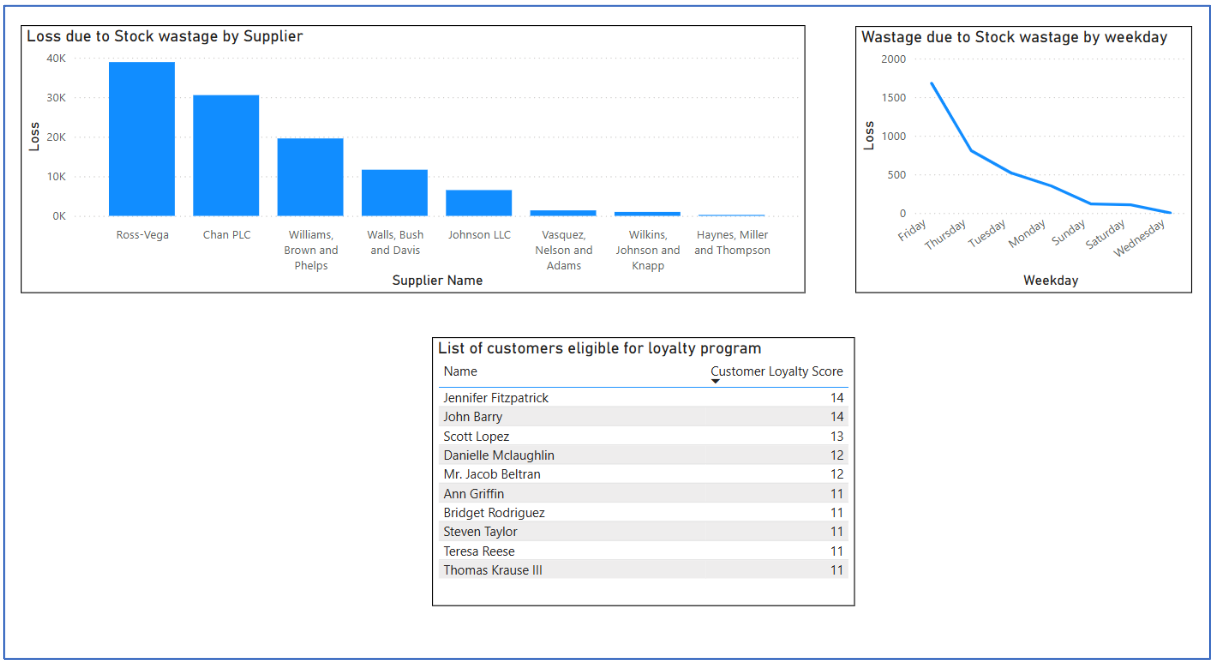


Catalog
Serverless Starter Warehouse Serverless S
Type to search...
My organization
system
main
retail_store
utils
control
table_list
default
information_schema
logging
pipeline_run_log

Power BI:

Finally, the Power BI is connected to the gold schema to fetch all the table data for the below reports.




Loss due to Stock wastage by Supplier
40K
30K
20K
10K
0K
Loss
Ross-Vega
Chan PLC
Williams, Brown and Phelps
Walls, Bush and Davis
Johnson LLC
Vasquez, Nelson and Adams
Wilkins, Johnson and Knapp
Haynes, Miller and Thompson
Supplier Name
Wastage due to Stock wastage by weekday
2000
1500
1000
500
0
Loss
Friday
Thursday
Tuesday
Monday
Sunday
Saturday
Wednesday
Weekday

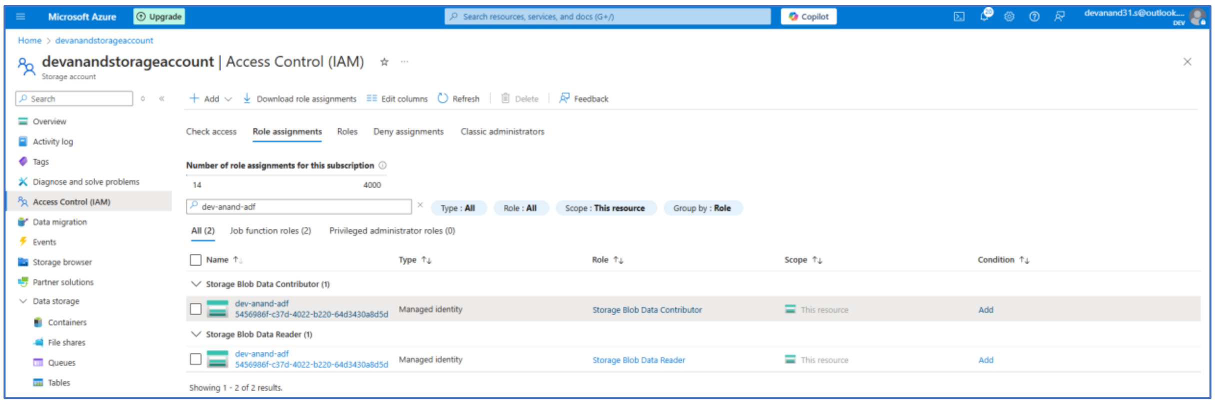
List of customers eligible for loyalty program

Name	Customer Loyalty Score
Jennifer Fitzpatrick	14
John Barry	14
Scott Lopez	13
Danielle Mclaughlin	12
Mr. Jacob Beltran	12
Ann Griffin	11
Bridget Rodriguez	11
Steven Taylor	11
Teresa Reese	11
Thomas Krause III	11

Security:

ADF <-> ADLS:

The Connection for the ADF to read and write files from and to ADLS is performed using managed identity with below access roles.


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
devanand31.s@outlook....
DEV
Home > devanandstorageaccount

devanandstorageaccount | Access Control (IAM)

Storage account

Number of role assignments for this subscription: 14 / 4000

Search: dev-anand-adf — Type : **All** · Role : **All** · Scope : **This resource** · Group by : **Role**

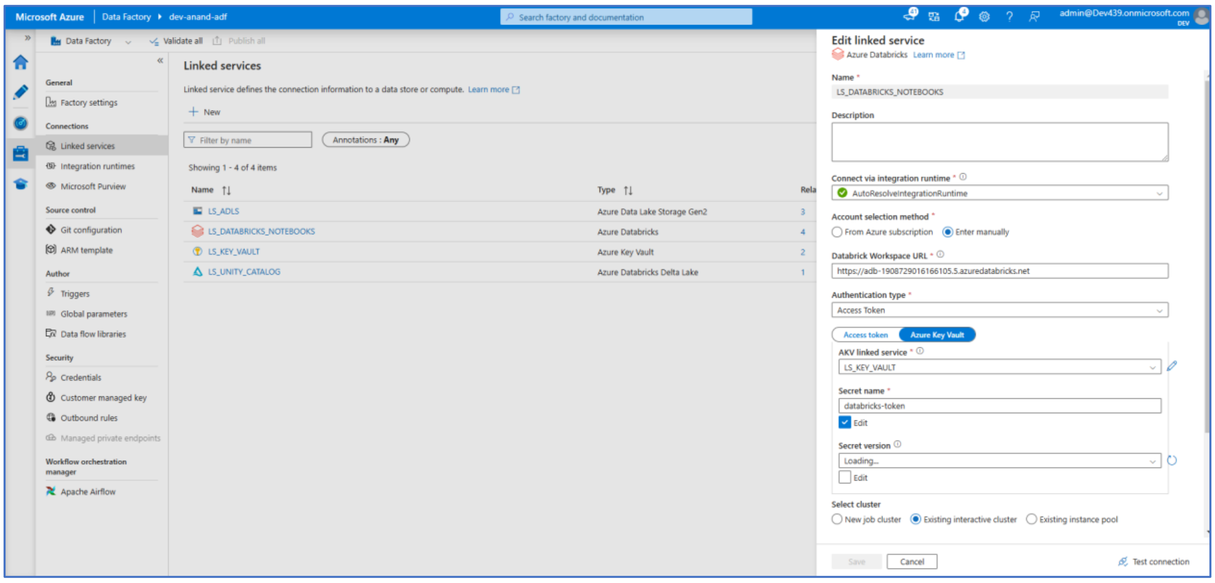
All (2) · Job function roles (2) · Privileged administrator roles (0)

Name	Type	Role	Scope	Condition
Storage Blob Data Contributor (1)				
dev-anand-adf 5456986f-c37d-4022-b220-64d3430a8d5d	Managed identity	Storage Blob Data Contributor	This resource	Add
Storage Blob Data Reader (1)				
dev-anand-adf 5456986f-c37d-4022-b220-64d3430a8d5d	Managed identity	Storage Blob Data Reader	This resource	Add

Showing 1 - 2 of 2 results.

ADF <-> Databricks:

The Connection for the ADF to Databricks to execute the notebook is created using the Access Token stored in the Key vault.



Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New · Filter by name · Annotations : **Any**

Showing 1 - 4 of 4 items

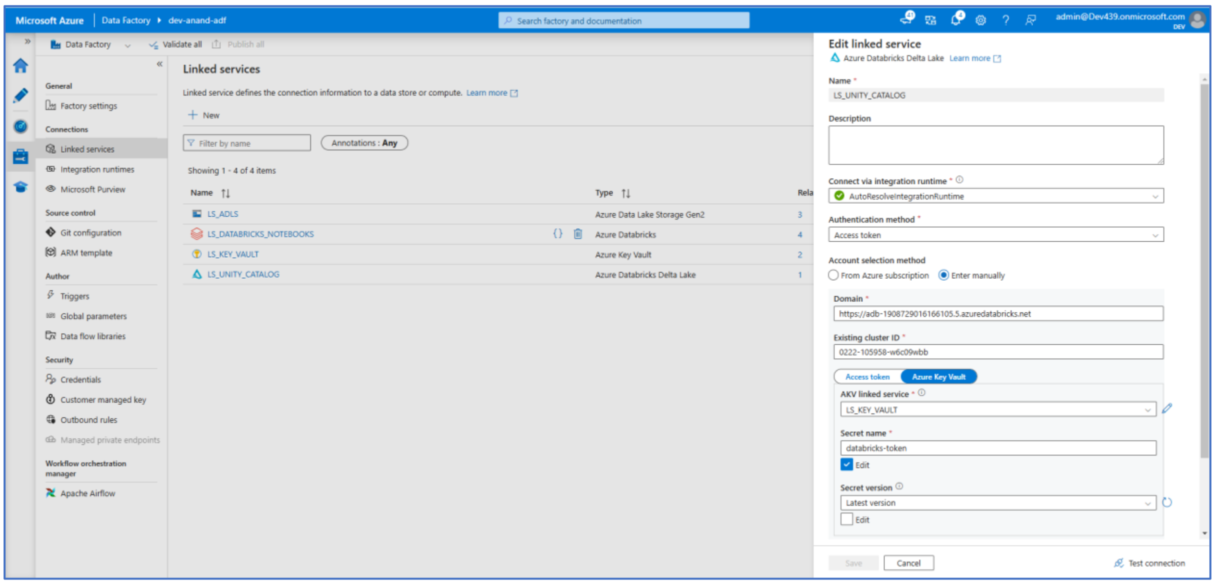
Name	Type	Rela
LS_ADLS	Azure Data Lake Storage Gen2	3
LS_DATABRICKS_NOTEBOOKS	Azure Databricks	4
LS_KEY_VAULT	Azure Key Vault	2
LS_UNITY_CATALOG	Azure Databricks Delta Lake	1

Edit linked service

Azure Databricks Learn more

Name *	LS_DATABRICKS_NOTEBOOKS
Description	
Connect via integration runtime *	AutoResolveIntegrationRuntime
Account selection method *	From Azure subscription [✓ Enter manually]
Databrick Workspace URL *	https://adb-1908729016166105.5.azuredatabricks.net
Authentication type *	Access Token
	Access token [✓ Azure Key Vault]
AKV linked service *	LS_KEY_VAULT
Secret name *	databricks-token
	[✓ Edit]
Secret version	Loading...
	Edit
Select cluster	New job cluster [✓ Existing interactive cluster] Existing instance pool

The connection to the unity catalog,



Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New · Filter by name · Annotations : **Any**

Showing 1 - 4 of 4 items

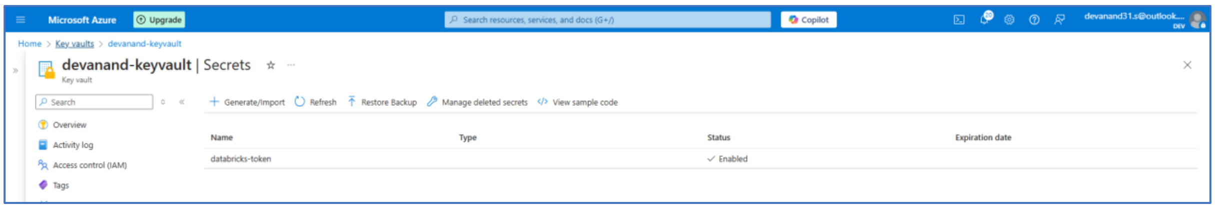
Name	Type	Rela
LS_ADLS	Azure Data Lake Storage Gen2	3
LS_DATABRICKS_NOTEBOOKS	Azure Databricks	4
LS_KEY_VAULT	Azure Key Vault	2
LS_UNITY_CATALOG	Azure Databricks Delta Lake	1

Edit linked service

Azure Databricks Delta Lake Learn more

Name *	LS_UNITY_CATALOG
Description	
Connect via integration runtime *	AutoResolveIntegrationRuntime
Authentication method *	Access token
Account selection method	From Azure subscription [✓ Enter manually]
Domain *	https://adb-1908729016166105.5.azuredatabricks.net
Existing cluster ID *	0222-105958-w6c09wbb
	Access token [✓ Azure Key Vault]
AKV linked service *	LS_KEY_VAULT
Secret name *	databricks-token
	[✓ Edit]
Secret version	Latest version
	Edit

Key vault secrets,



Home > Key vaults > devanand-keyvault

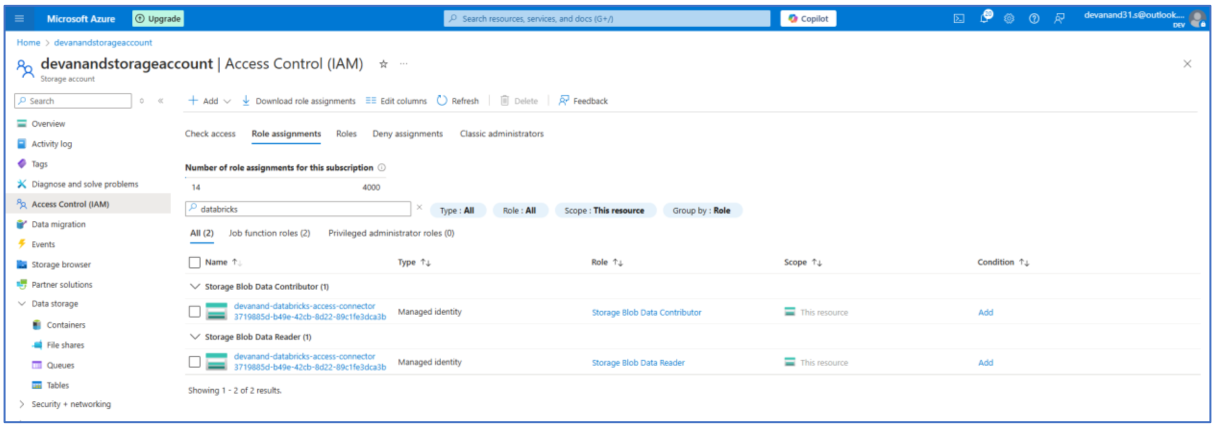
devanand-keyvault | Secrets

Key vault

Name	Type	Status	Expiration date
databricks-token		✓ Enabled	

Databricks <-> ADLS:

The External Storage is created in the Databricks using Databricks Access connector which in-turn is authorized by managed identity.



Home > devanandstorageaccount

devanandstorageaccount | Access Control (IAM)

Storage account

Number of role assignments for this subscription

14 4000

databricks

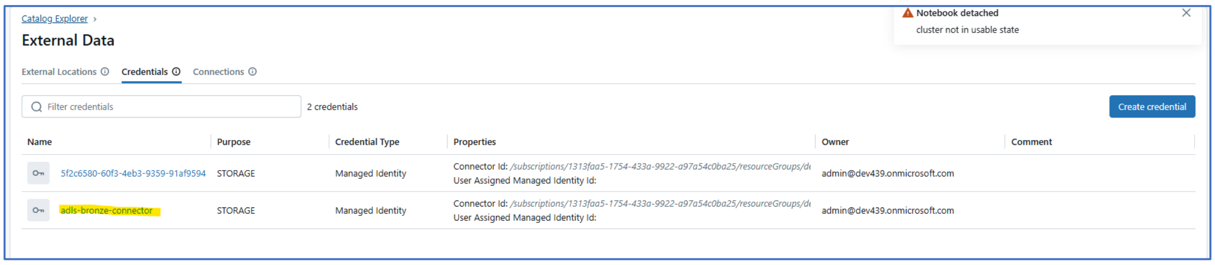
Type : All Role : All Scope : This resource Group by : Role

All (2) Job function roles (2) Privileged administrator roles (0)

Name	Type	Role	Scope	Condition
Storage Blob Data Contributor (1)				
devanand-databricks-access-connector 3719885d-b49e-42cb-8d22-89c1fe3dca3b	Managed identity	Storage Blob Data Contributor	This resource	Add
Storage Blob Data Reader (1)				
devanand-databricks-access-connector 3719885d-b49e-42cb-8d22-89c1fe3dca3b	Managed identity	Storage Blob Data Reader	This resource	Add

Showing 1 - 2 of 2 results.

Databricks credential created in the catalog,



Catalog Explorer >

External Data

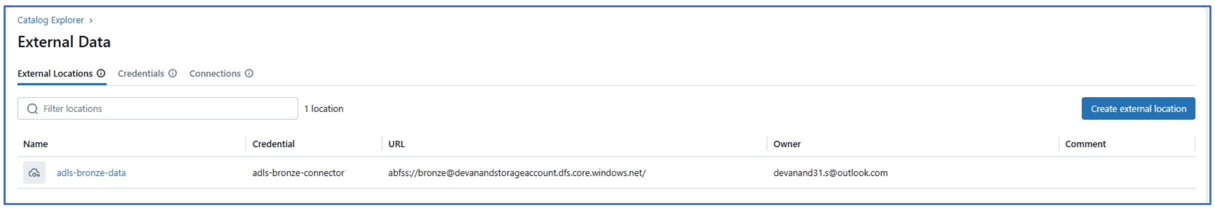
Notebook detached

cluster not in usable state

2 credentials

Name	Purpose	Credential Type	Properties	Owner	Comment
5f2c6580-60f3-4eb3-9359-91af9594	STORAGE	Managed Identity	Connector Id: */subscriptions/1313faa5-1754-433a-9922-a97a54c0ba25/resourceGroups/de* User Assigned Managed Identity Id:	admin@dev439.onmicrosoft.com	
adls-bronze-connector	STORAGE	Managed Identity	Connector Id: */subscriptions/1313faa5-1754-433a-9922-a97a54c0ba25/resourceGroups/de* User Assigned Managed Identity Id:	admin@dev439.onmicrosoft.com	

External location access the ADLS via databricks access connector,



Catalog Explorer >

External Data

1 location

Name	Credential	URL	Owner	Comment
adls-bronze-data	adls-bronze-connector	abfss://bronze@devanandstorageaccount.dfs.core.windows.net/	devanand31.s@outlook.com	

Power BI <-> Databricks unity catalog:

The PowerBI access the unity catalog gold schema using the Access Tokens.