

# End-To-End Data Engineering Project using Microsoft Azure Data Factory and Azure Synapse Analytics

## Business Problem

### Context:

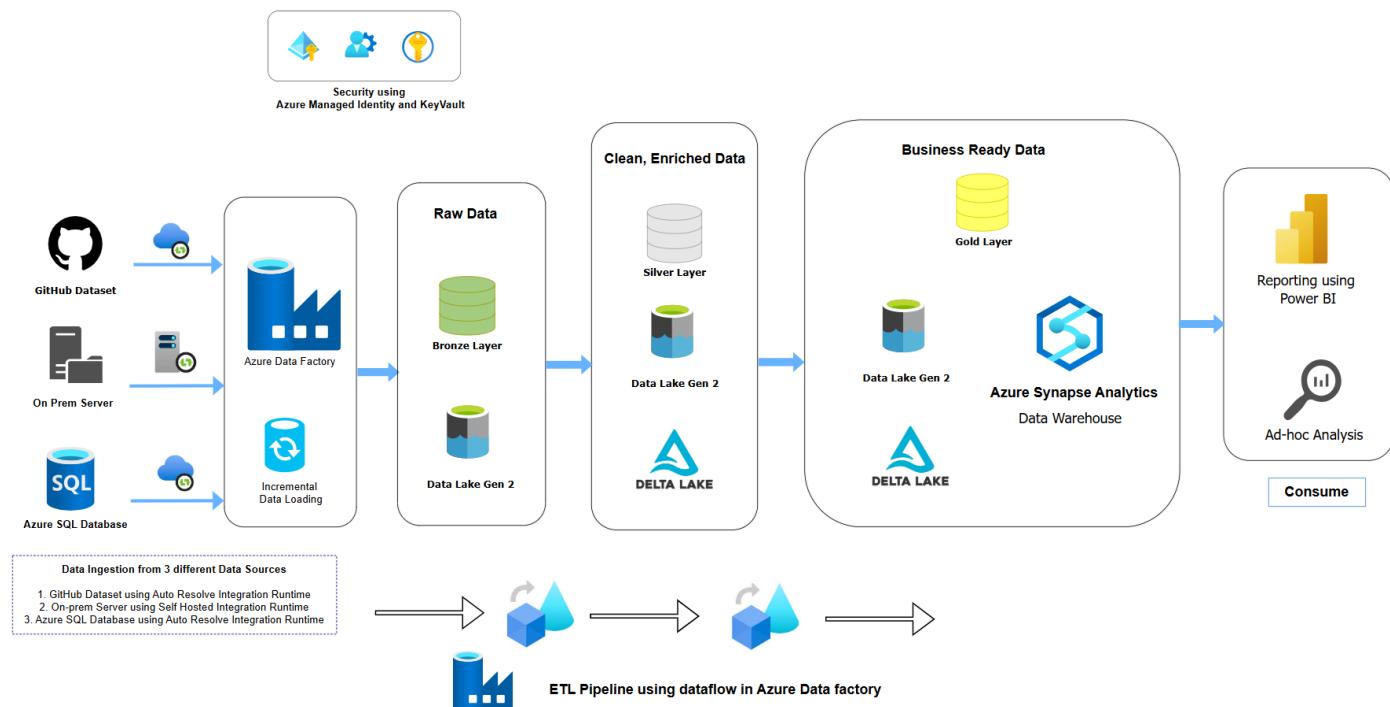
A retail company (or a similar enterprise in this case study) operates in multiple regions with diverse product lines and a network of resellers. Sales are driven by a distributed salesforce, and the company sets quarterly/annual sales targets at different levels (by product, region, or salesperson).

### Key Business Problems:

- No unified view of sales across category, brand, salesperson.
- Difficulty in tracking sales performance vs. targets
- Inefficient allocation of resources to underperforming areas
- No end-to-end data pipeline to process the data.

### Proposed Data-Driven Solution:

- Ingested raw CSV datasets into **Azure Data Lake Gen2** using **Azure Data Factory**
- Stored raw, untransformed data in the **Bronze Layer** in Azure Data Lake
- Cleaned, standardized, and enriched data using **ADF Data Flows** in the **Silver Layer**
- Applied business logic and store data in Delta format in **Gold Layer**. Build a Data Warehouse Solution in **Azure Synapse Analytics**
- Secured the pipeline with **Azure Managed Identity & Key Vault**
- Delivered analytics-ready data for visualization in **Power BI**



# Data Ingestion

Data have been stored in 3 different data sources.

1. GitHub Storage in CSV format
2. On-Prem Server in CSV format
3. Azure SQL in structured format

Here I have used two different Integration Runtime:

1. Self-Hosted Integration Runtime for migrating data from on-prem to cloud
2. Auto Resolve Integration Runtime for any other data operation

## Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment. [Learn more](#)

[+ New](#) [Refresh](#)

[Filter by name](#)

Showing 1 - 2 of 2 items

Name ↑↓	Type ↑↓	Sub-type ↑↓	Status ↑↓	Related ↑↓	Region ↑↓	Version ↑↓
AutoResolveIntegrationRuntime	Azure	Public	<span style="color: green;">✓</span> Running	0	Auto Resolve	---
PranoySelfHostedIntegrationRuntime	Self-Hosted	---	<span style="color: green;">✓</span> Running	1	---	5.54.9271.2

# Parent Pipeline Run

All pipeline runs > ✓ data-ingestion-parent-pipeline - Activity runs

[Rerun](#) [Cancel](#) [Refresh](#) [Update pipeline](#) [List](#) [Gantt](#)

**Activity runs**

All status ▾ [Monitor in Azure Metrics](#)

Showing 1 - 4 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓	User prop... ↑↓	Activity run ID ↑↓
SendNotificationViaLogicApp	<span style="color: green;">✓</span> Succeeded	Web	8/23/2025, 10:21:55 PM	13s	AutoResolveIntegrationRuntime (West US)		97d176fb-4dfd-4e46-8e91-df49a3a69f80
ExecutePipelineAzureSqlToDa...	<span style="color: green;">✓</span> Succeeded	Execute Pipeline	8/23/2025, 10:20:28 PM	1m 27s			0bcbd67c-3d0f-405f-9ef7-0a74e1aa8a14
ExecutePipelineGithubToData...	<span style="color: green;">✓</span> Succeeded	Execute Pipeline	8/23/2025, 10:20:00 PM	28s			fbb9430f-a7e0-43a8-8555-b01bcb14809f
ExecutePipelineOnPremToDa...	<span style="color: green;">✓</span> Succeeded	Execute Pipeline	8/23/2025, 10:18:36 PM	1m 24s			7029fe56-9125-4f4b-b93d-d146d65c5c52

## Migrating Data from On-Prem to Azure Data Lake

All pipeline runs > ✓ data-ingestion-onprem-to-datalake - Activity runs

↻ Rerun 🕒 Cancel ⟳ Refresh ✍ Update pipeline List Gantt

**Pipeline run details**

- Run by: 7029fe56-9125-4f4b-b93d-d146d65c5c52
- Start time: 8/23/2025, 10:18:36 PM
- End time: 8/23/2025, 10:19:59 PM
- Status: ✓ Succeeded
- Parameters: 0
- Pipeline run ID: 0a8ef549-d943-4457-b852-7651ccfb445d

**Activity runs**

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...
MigrateOnPremData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:18:39 PM	1m 19s	PranoySelfHostedIntegrationRuntime	

Monitor in Azure Metrics 🕒 Export to CSV

## Migrating Data from GitHub to Azure Data Lake

All pipeline runs > ✓ data-ingestion-github-to-datalake - Activity runs

↻ Rerun 🕒 Cancel ⟳ Refresh ✍ Update pipeline List Gantt

**Pipeline run details**

- Run by: fb9430f-a7e0-43a8-8555-b01bcb14809f
- Start time: 8/23/2025, 10:20:00 PM
- End time: 8/23/2025, 10:20:26 PM
- Status: ✓ Succeeded
- Parameters: 1
- Pipeline run ID: 056c78a0-4266-434b-9dfa-ec0777391c65

**Activity runs**

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...
ForEachFileinGitHubRepo	<span style="color: green;">✓ Succeeded</span>	ForEach	8/23/2025, 10:20:02 PM	24s		
MigrateGitHubData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:20:03 PM	14s	AutoResolveIntegrationRuntime (West US)	
MigrateGitHubData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:20:03 PM	15s	AutoResolveIntegrationRuntime (West US)	
MigrateGitHubData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:20:03 PM	14s	AutoResolveIntegrationRuntime (West US)	
MigrateGitHubData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:20:03 PM	22s	AutoResolveIntegrationRuntime (West US)	
MigrateGitHubData	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:20:03 PM	14s	AutoResolveIntegrationRuntime (West US)	

Monitor in Azure Metrics 🕒 Export to CSV

## Migrating Data from Azure SQL to Azure Data Lake + Incremental Data Loading

All pipeline runs > ✓ data-ingestion-azuresql-to-datalake - Activity runs

↻ Rerun 🕒 Cancel ⟳ Refresh ✍ Update pipeline List Gantt

**Pipeline run details**

- Run by: 0bcbd67c-3d0f-405f-9ef7-0a74e1aa8a14
- Start time: 8/23/2025, 10:20:28 PM
- End time: 8/23/2025, 10:21:54 PM
- Status: ✓ Succeeded
- Parameters: 0
- Pipeline run ID: ce521eb7-ea74-49fa-8227-ce63bacbd7dc

**Activity runs**

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...
LastLoadFromJson	<span style="color: green;">✓ Succeeded</span>	Lookup	8/23/2025, 10:20:32 PM	5s	AutoResolveIntegrationRuntime (West US)	
LatestLoadFromSQL	<span style="color: green;">✓ Succeeded</span>	Lookup	8/23/2025, 10:20:32 PM	46s	AutoResolveIntegrationRuntime (West US 2)	
CopyDataFromAzureSQL	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:21:19 PM	17s	AutoResolveIntegrationRuntime (West US)	
Watermark	<span style="color: green;">✓ Succeeded</span>	Copy data	8/23/2025, 10:21:38 PM	15s	AutoResolveIntegrationRuntime (West US)	

Monitor in Azure Metrics 🕒 Export to CSV

## Alert Notification using Azure Logic App

Home > la-alert-data-ingestion-pipeline | Run history >

**la-alert-data-ingestion-pipeline | Run history**

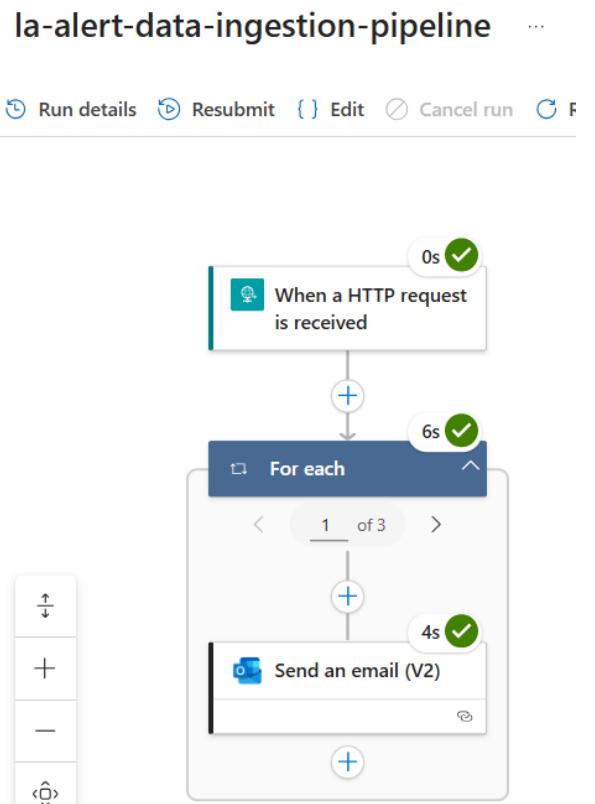
Logic app

Search  Refresh

Overview  
Activity log  
Access control (IAM)  
Tags  
Diagnose and solve problems  
Resource visualizer  
Development Tools  
Logic app designer  
Logic app code view  
Logic app templates  
**Run history**  
Versions  
API connections  
Quick start guides  
Settings  
Monitoring  
Automation

All  
Pick a date  Pick a time   
Search to filter items by identifier

Start time	Duration
8/23/2025, 10:22 PM	6.13 Seconds
8/21/2025, 11:46 PM	3.1 Seconds
8/21/2025, 11:42 PM	1.3 Seconds
8/21/2025, 11:36 PM	2.08 Seconds



## Raw Data stored in Bronze Layer in Azure Data Lake (CSV format)

Home > pranoyadlsgen2 | Containers >

**retail-bronze** Container

Search Add Directory Upload Refresh Delete Copy Paste Rename Acquire lease Break lease Edit columns

Overview  
Diagnose and solve problems  
Access Control (IAM)  
Settings

Search blobs by prefix (case-sensitive)

Showing all 7 items

	Name	Last modified	Access tier
<input type="checkbox"/>	campaigns	8/21/2025, 10:05:17 PM	
<input type="checkbox"/>	customer	8/21/2025, 6:55:32 PM	
<input type="checkbox"/>	dates	8/21/2025, 10:05:15 PM	
<input type="checkbox"/>	products	8/21/2025, 10:05:15 PM	
<input type="checkbox"/>	sales	8/21/2025, 10:49:14 PM	
<input type="checkbox"/>	salespersons	8/21/2025, 10:05:17 PM	
<input type="checkbox"/>	stores	8/21/2025, 10:05:16 PM	

# Data Transformation

Data Transformation has been performed by using Data flow in Azure Data Factory. The data from **Bronze layer** has been extracted, cleaned, transformed and stored in **Silver Layer** in Delta Format.

## 1. ACID Transactions on Data Lakes

- Delta brings **Atomicity, Consistency, Isolation, Durability** (ACID) to data lakes.
- Prevents issues like partial writes, duplicate data, or corrupted files common with CSV/Parquet.

## 2. Schema Enforcement & Evolution

- Enforces column types and structures to avoid “bad data” landing in production.
- Supports **schema evolution** when new columns are added — without breaking existing pipelines.

## 3. Time Travel (Data Versioning)

- Every write creates a new version of the table.
- You can query older snapshots for **debugging, audits, or rollback** (e.g., “see data as of yesterday”).

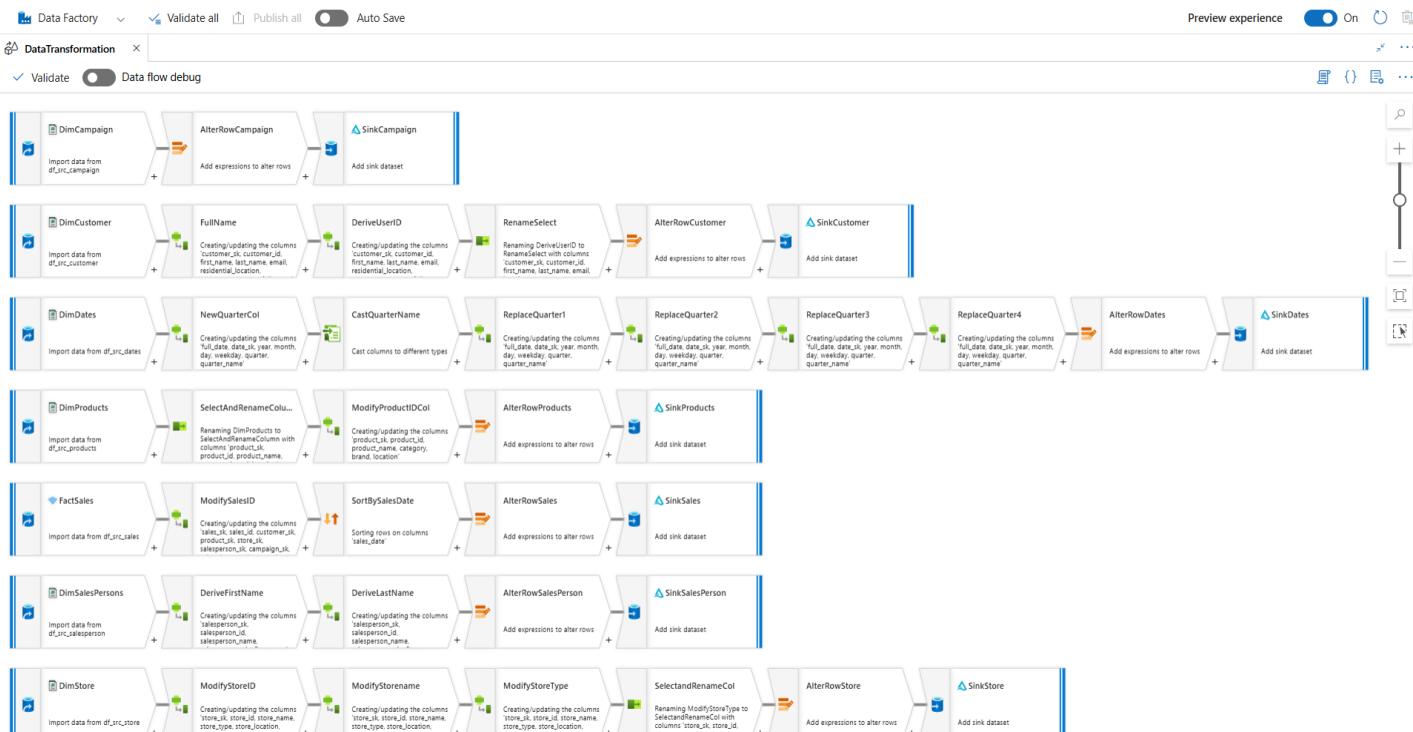
## 4. High Performance with Indexing & Caching

- Delta automatically creates metadata (transaction logs, indexes).
- Faster queries compared to plain Parquet because it avoids scanning entire datasets.

## 5. Unified Batch + Streaming

- Same Delta table can serve **streaming and batch pipelines** without separate architecture.
- Great for real-time analytics and incremental data processing.

**Delta makes data lake reliable, consistent, and production-ready**, solving the weaknesses of raw CSV/Parquet.

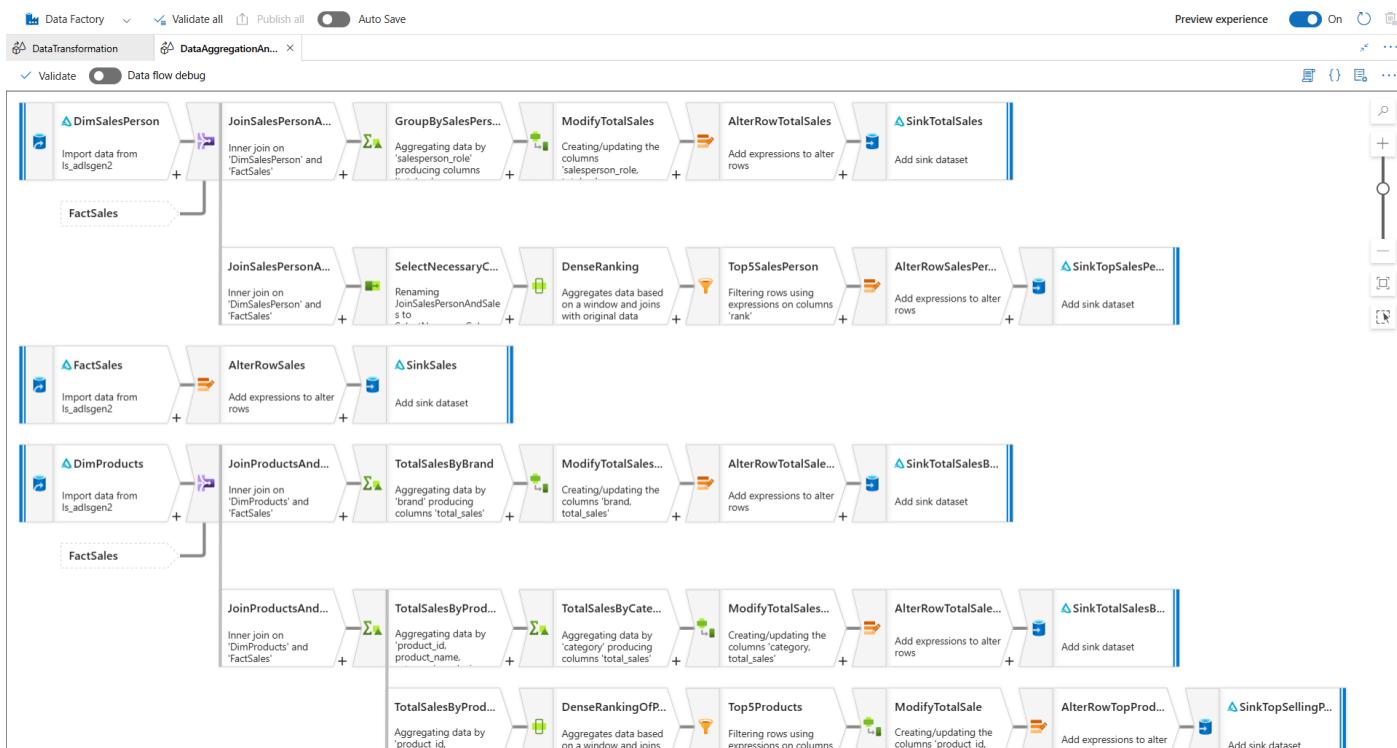


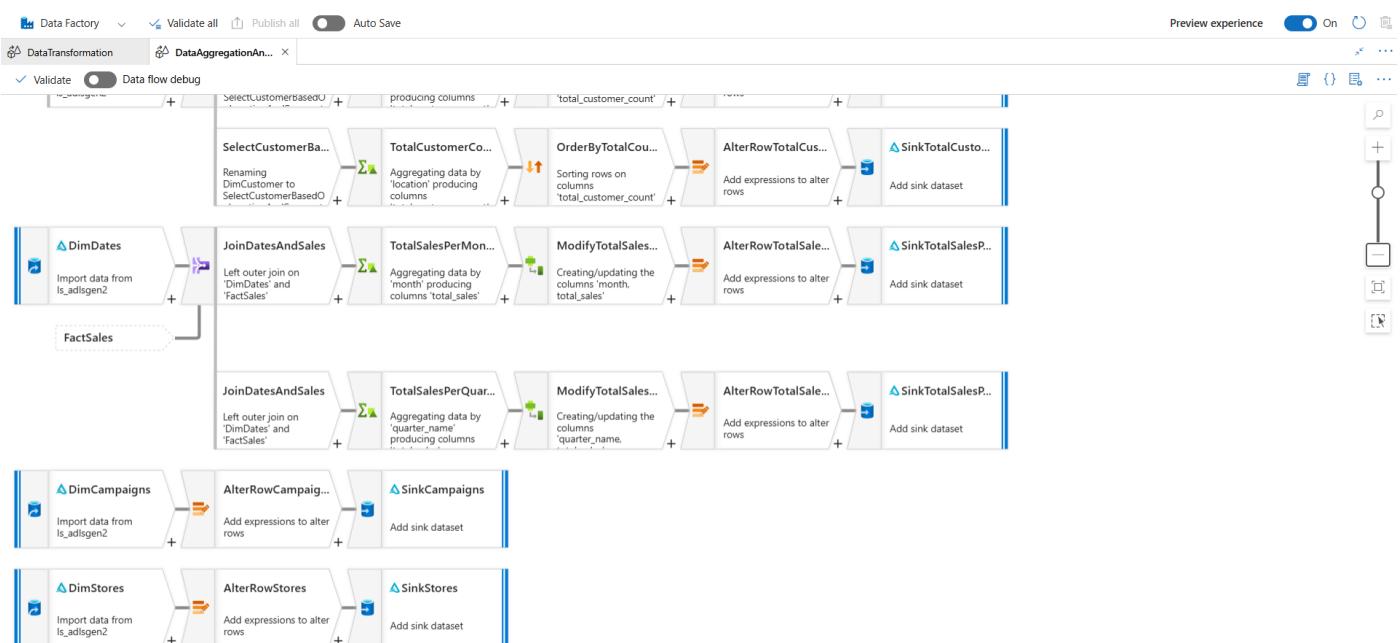
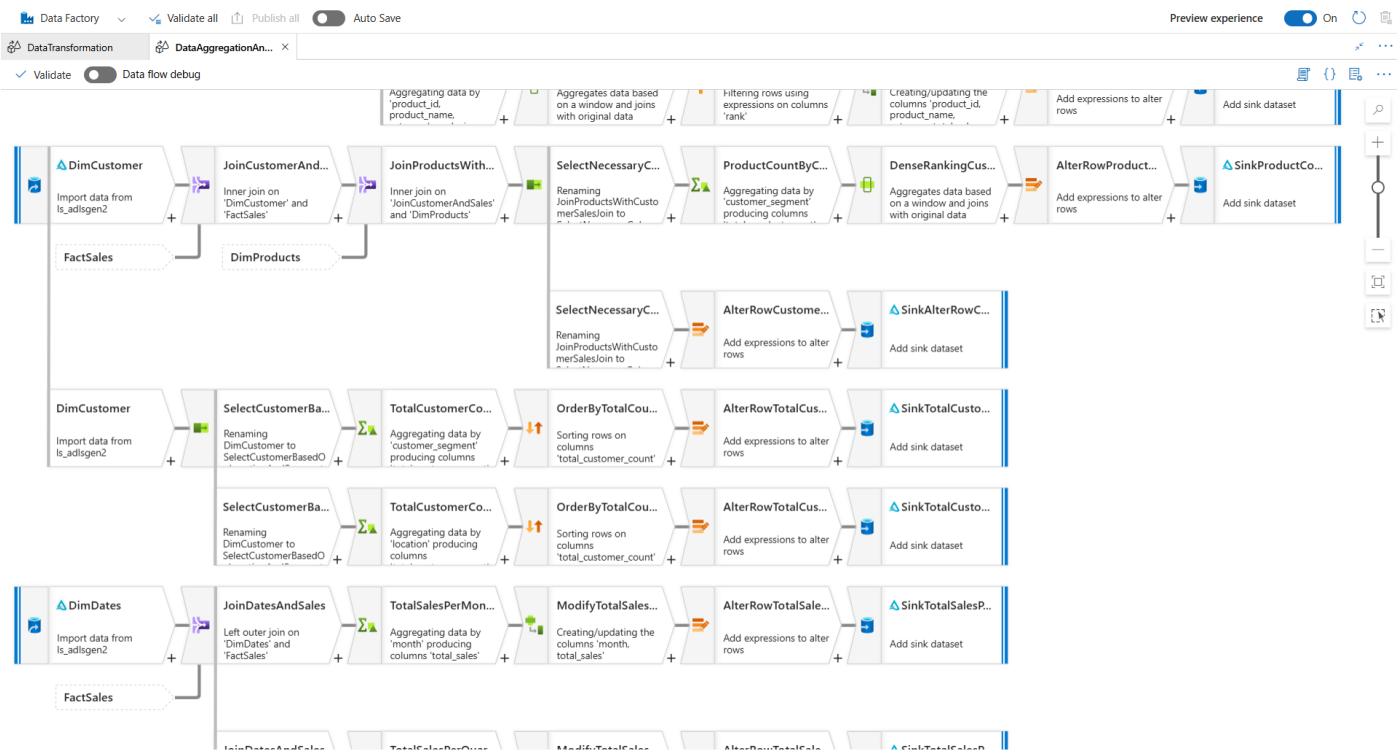
# Data Aggregation and Serving

The **Gold Layer** represents the final curated and business-ready data model designed for reporting, dashboarding, and advanced analytics. Data from the **Silver Layer** is aggregated, enriched, and structured into fact and dimension models that directly support business KPIs and decision-making.

## Key Features of the Gold Layer:

- **Business-Ready Data** – Provides clean, transformed, and aggregated datasets that can be directly consumed by BI tools like Power BI or Tableau.
- **Fact & Dimension Modeling** – Data is shaped into fact tables (sales, revenue, transactions) and dimension tables (customers, products, stores, regions) for intuitive analysis.
- **Advanced Transformations** – Includes **joins, ranking, grouping, and window functions** to derive insights such as *top-performing salespersons, best-selling products, or customer churn indicators*.
- **KPI Calculation** – Pre-computed measures (e.g., total sales, top products etc.) speed up dashboard performance and ensure consistency across analytics.
- **Actionable Insights** – Enables stakeholders to monitor trends, anomalies, and performance metrics that directly guide strategic decisions.





# Data Transformation and Aggregation Pipeline

## Parent Pipeline Run

All pipeline runs > ✓ data-transformation-aggregation-pipeline - Activity runs

↻ Rerun ✖ Cancel ⟳ Refresh ✎ Update pipeline List Gantt

Activity runs

All status ✓ Monitor in Azure Metrics 🔗 View run detail ⬇️ Export to CSV CSV

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Execute Pipeline Data Transfo...	<span style="color: green;">✓ Succeeded</span>	Execute Pipeline	8/24/2025, 12:12:07 AM	3m 59s			45971841-2aee-48a0-9294-d87086d09527
SendNotificationViaLogicApp1	<span style="color: green;">✓ Succeeded</span>	Web	8/24/2025, 12:16:06 AM	3s	AutoResolveIntegrationRuntime (West US)		36e19aaef-4047-4bf7-ac55-6edbb625b828
Execute Pipeline Data Aggreg...	<span style="color: green;">✓ Succeeded</span>	Execute Pipeline	8/24/2025, 12:16:09 AM	1m 31s			94b581d4-79a5-4029-a2e4-4449808abada
SendNotificationViaLogicApp2	<span style="color: green;">✓ Succeeded</span>	Web	8/24/2025, 12:17:40 AM	3s	AutoResolveIntegrationRuntime (West US)		459a23f8-7754-484b-8bfd-cd3b9767648e

## Data Transformation Pipeline Run

All pipeline runs > data-transformation-bronze-to-silver - Activity runs > **Data Transformation**

✓ Data Transformation  
Cluster startup time: 2m 5s Number of transformations: 40 Data flow status: Success

⟳ Refresh ⟳ Auto refresh On ✎ Edit dataflow

Sinks All streams

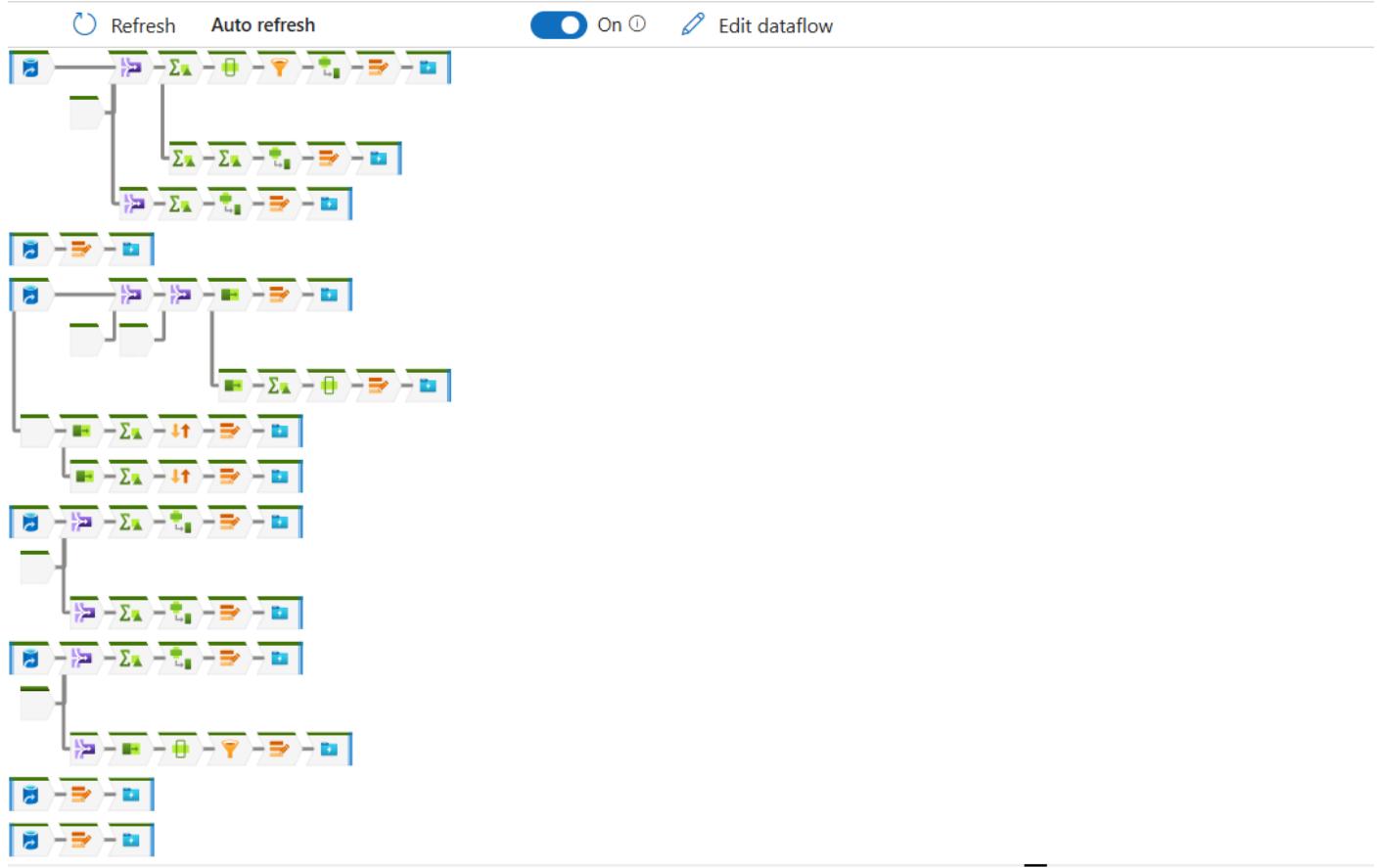
Sink	Status	Processing time	Highest processing time	Rows written	Stages	Lineage
SinkStore	<span style="color: green;">✓ Succeeded</span>	13s	658ms	500	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkCustomer	<span style="color: green;">✓ Succeeded</span>	7s	636ms	100k	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkProducts	<span style="color: green;">✓ Succeeded</span>	5s	54ms	210	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkCampaign	<span style="color: green;">✓ Succeeded</span>	5s	49ms	50	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkSales	<span style="color: green;">✓ Succeeded</span>	6s	416ms	2004	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkDates	<span style="color: green;">✓ Succeeded</span>	5s	119ms	366	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>
SinkSalesPerson	<span style="color: green;">✓ Succeeded</span>	5s	73ms	2000	<span style="color: green;">🔗</span>	<span style="color: green;">🔗</span>

## Data Aggregation Pipeline Run

All pipeline runs > data-aggregation-serving-silver-to-gold - Activity runs > Data Aggregation and Serving

### ✓ Data Aggregation and Serving

Cluster startup time: 4s 904ms Number of transformations: 65 Data flow status: Success



[Sinks](#) [All streams](#)

Sink	Status	Processing time ↑↓	Highest processing time ↑↓	Rows written ↑↓
SinkCampaigns	<span>✓ Succeeded</span>	3s 454ms	279ms	50
SinkTotalCustomerCountBySegm	<span>✓ Succeeded</span>	3s 226ms	565ms	10
SinkTotalCustomerCountByLocat	<span>✓ Succeeded</span>	2s 401ms	565ms	50
SinkProductCountByCustomerSe	<span>✓ Succeeded</span>	3s 775ms	565ms	10

## Alert Notification using Azure Logic App

Home > Microsoft.Web-LogicAppConsumption-Portal-bfa6d766-988a | Overview > la-alert-data-transformation-pipeline | Run history >

la-alert-data-transformation-pipeline | ...

Search Refresh

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Resource visualizer Development Tools Logic app designer Logic app code view Logic app templates Run history Versions API connections Quick start guides Settings

All Pick a date Pick a time Search to filter items by identifier

Start time Duration

8/24/2025, 12:16 AM 2.52 Seconds  
8/23/2025, 11:24 PM 1.9 Seconds

0s ✓ 2s ✓ 0.4s ✓ 0.7s ✓ 0.3s ✓

When a HTTP request is received  
For each  
Send an email (V2)

Home > la-alert-data-aggregation-pipeline | Run history >

la-alert-data-aggregation-pipeline | Run history

Search Refresh

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Resource visualizer Development Tools Logic app designer Logic app code view Logic app templates Run history Versions API connections Quick start guides Settings Monitoring

All Pick a date Pick a time Search to filter items by identifier

Start time Duration

8/24/2025, 12:17 AM 1.06 Seconds

0s ✓ 0.7s ✓ 0.3s ✓

When a HTTP request is received  
For each  
Send an email (V2)

## Cleaned, Transformed Data stored in Silver Layer in Azure Data Lake

Home > pranoyadlsgen2 | Containers >

**retail-silver** Container

Search Overview Diagnose and solve problems Access Control (IAM) Settings

Add Directory Upload Refresh Delete Copy Paste Rename Acquire lease Break lease Edit columns

retail-silver Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Search blobs by prefix (case-sensitive)

Showing all 7 items

Name	Last modified	Access tier
campaign	8/22/2025, 3:12:56 PM	
customer	8/22/2025, 3:13:30 PM	
dates	8/22/2025, 3:13:23 PM	
products	8/22/2025, 3:12:42 PM	
sales	8/22/2025, 3:13:08 PM	
salesperson	8/22/2025, 3:13:16 PM	
store	8/22/2025, 3:13:02 PM	

## Data stored in Delta Format

Home > pranoyadlsgen2 | Containers >

**retail-silver** Container

Search Overview Diagnose and solve problems Access Control (IAM) Settings

Add Directory Upload Refresh

retail-silver > sales Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Search blobs by prefix (case-sensitive)

Showing all 7 items

Name
[ ]
_delta_log
.temporary
<input checked="" type="checkbox"/> part-00000-17cd1ceb-46a1-4409-9377-0f6f437c694e-c000.snappy.parquet
part-00000-2a335417-4ab8-470a-9d4
part-00000-582e1652-6750-4dc8-977
part-00000-61fb9bad-fb2f-4b31-8d0
part-00000-e3df6a22-e188-4b7f-aa7c

**sales/part-00000-17cd1ceb-46a1-4409-9377-0f6f437c694e-c000.snappy.parquet**

Blob

Save Discard Download Refresh Delete Change tier Acquire lease Break lease Give feedback

Overview Versions Edit Generate SAS

Properties

URL	<a href="https://pranoyadlsgen2...">https://pranoyadlsgen2...</a>
LAST MODIFIED	8/24/2025, 12:15:31 AM
CREATION TIME	8/24/2025, 12:15:30 AM
VERSION ID	-
TYPE	Block blob
SIZE	1.01 KB
ACCESS TIER	Hot (Inferred)
ACCESS TIER LAST MODIFIED	N/A
ARCHIVE STATUS	-
REHYDRATE PRIORITY	-
SERVER ENCRYPTED	true
ETAG	0x8DDE2753C888A88
VERSION-LEVEL IMMUTABILITY POLICY	Disabled
CACHE-CONTROL	
CONTENT-TYPE	application/octet-stream
CONTENT-MD5	

## Aggregated, Business-ready Data stored in Gold Layer in Azure Data Lake

Home > pranoyadlsgen2 | Containers >

**retail-gold** ...

Container

Search

Add Directory Upload Refresh Delete Copy Paste Rename Acquire lease Break lease Edit columns

**Overview**

Diagnose and solve problems

Access Control (IAM)

Settings

Authenticaton method: Access key ([Switch to Microsoft Entra user account](#))

Search blobs by prefix (case-sensitive)

Showing all 13 items

	Name	Last modified	Access tier
<input type="checkbox"/>	Campaigns	8/24/2025, 12:17:18 AM	
<input type="checkbox"/>	CustomerProduct	8/24/2025, 12:17:07 AM	
<input type="checkbox"/>	ProductCountByCustomerSegment	8/24/2025, 12:17:02 AM	
<input type="checkbox"/>	Sales	8/24/2025, 12:17:05 AM	
<input type="checkbox"/>	Top5Products	8/24/2025, 12:17:13 AM	
<input type="checkbox"/>	Top5SalesPerson	8/24/2025, 12:17:10 AM	
<input type="checkbox"/>	TotalCustomerCountByLocation	8/24/2025, 12:16:59 AM	
<input type="checkbox"/>	TotalCustomerCountBySegment	8/24/2025, 12:16:41 AM	
<input type="checkbox"/>	TotalSales	8/24/2025, 12:16:54 AM	
<input type="checkbox"/>	TotalSalesByBrand	8/24/2025, 12:16:56 AM	
<input type="checkbox"/>	TotalSalesByCategory	8/24/2025, 12:16:45 AM	
<input type="checkbox"/>	TotalSalesPerMonth	8/24/2025, 12:16:51 AM	
<input type="checkbox"/>	TotalSalesPerQuarter	8/24/2025, 12:16:48 AM	

Add or remove favorites by pressing Ctrl+Shift+F

## Data stored in Delta Format

Home > pranoyadlsgen2 | Containers >

**retail-gold** ...

Container

Search

Add Directory Upload Refresh

**Overview**

Diagnose and solve problems

Access Control (IAM)

Settings

Authenticaton method: Access key ([Switch to Microsoft Entra user account](#))

Search blobs by prefix (case-sensitive)

Showing all 2 items

	Name
<input type="checkbox"/>	[.]
<input type="checkbox"/>	_delta_log
<input checked="" type="checkbox"/>	part-00000-35264091-1498-491d-b3e0-

**TotalSalesPerMonth/part-00000-35264091-1498-491d-b3e0-**

Blob

Save Discard Download Refresh Delete Change tier Acquire lease

Overview Versions Edit Generate SAS

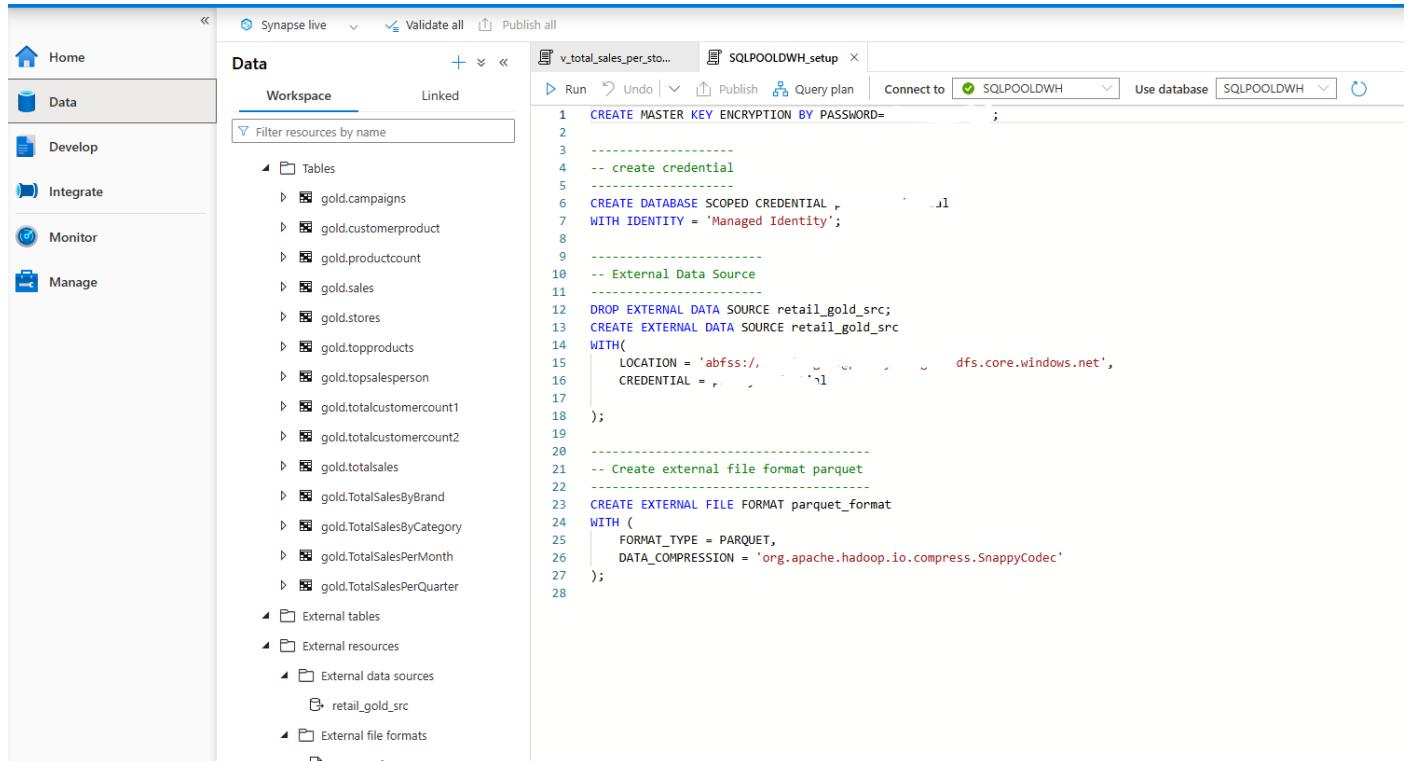
Properties

URL	<a href="https://pranoyadlsgen2....">https://pranoyadlsgen2....</a>
LAST MODIFIED	8/24/2025, 12:16:52 AM
CREATION TIME	8/24/2025, 12:16:52 AM
VERSION ID	-
TYPE	Block blob
SIZE	805 B
ACCESS TIER	Hot (Inferred)
ACCESS TIER LAST MODIFIED	N/A
ARCHIVE STATUS	-
REHYDRATE PRIORITY	-
SERVER ENCRYPTED	true
ETAG	0x8DDE2756CD6216C
VERSION-LEVEL IMMUTABILITY POLICY	Disabled
CACHE-CONTROL	
CONTENT-TYPE	application/octet-stream
CONTENT-MD5	
CONTENT-ENCODING	

# Data Warehousing and Serving using Azure Synapse Analytics

Setup **Dedicated SQL Pool** and complete the setup

- Create MASTER KEY ENCRYPTION
- Create CREDENTIAL
- Create EXTERNAL DATA SOURCE
- CREATE EXTERNAL FILE FORMAT



The screenshot shows the Azure Synapse Analytics Data workspace interface. On the left, the 'Data' blade is open, displaying a list of tables under the 'Tables' section, including gold.campaigns, gold.customerproduct, gold.productcount, gold.sales, gold.stores, gold.topproducts, gold.topsalesperson, gold.totalcustomercount1, gold.totalcustomercount2, gold.totalsales, gold.TotalSalesByBrand, gold.TotalSalesByCategory, gold.TotalSalesPerMonth, and gold.TotalSalesPerQuarter. Below the tables are sections for 'External tables', 'External resources', 'External data sources' (with 'retail\_gold\_src' listed), and 'External file formats'. On the right, a query editor window titled 'v\_total\_sales\_per\_sto...' is open, showing T-SQL code for setting up a master key, creating a credential, dropping and creating an external data source for 'retail\_gold\_src' using an ABFS location and managed identity, and creating an external file format named 'parquet\_format' for Parquet files with specific compression settings. The code is as follows:

```

1 CREATE MASTER KEY ENCRYPTION BY PASSWORD='';
2
3 -----
4 -- create credential
5 -----
6 CREATE DATABASE SCOPED CREDENTIAL retail_gold_src
7 WITH IDENTITY = 'Managed Identity';
8
9 -----
10 -- External Data Source
11 -----
12 DROP EXTERNAL DATA SOURCE retail_gold_src;
13 CREATE EXTERNAL DATA SOURCE retail_gold_src
14 WITH(
15     LOCATION = 'abfss://retail_gold_src@12345678901234567890123456789012. dfs.core.windows.net',
16     CREDENTIAL = retail_gold_src_credential
17 );
18
19 -----
20 -- Create external file format parquet
21 -----
22 CREATE EXTERNAL FILE FORMAT parquet_format
23 WITH (
24     FORMAT_TYPE = PARQUET,
25     DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
26 );
27
28 );

```

## Use COPY INTO to copy data from Azure Data Lake Gold Layer to Azure Synapse Analytics Dedicated SQL Pool

```

1  -----
2  -- TopSalesPerson Table
3  -----
4  DROP TABLE gold.topsalesperson;
5  CREATE TABLE gold.topsalesperson(
6      salesperson_id VARCHAR(100),
7      salesperson_name VARCHAR(100),
8      salesperson_role VARCHAR(100),
9      first_name VARCHAR(100),
10     last_name VARCHAR(100),
11     total_amount FLOAT,
12     rank INT
13 )
14 WITH(
15     DISTRIBUTION = ROUND_ROBIN
16 );
17
18 --load data using COPY INTO
19 COPY INTO gold.topsalesperson(
20     salesperson_id 1,
21     salesperson_name 2,
22     salesperson_role 3,
23     first_name 4,
24     last_name 5,
25     total_amount 6,
26     rank 7
27 )
28 FROM 'https://pranoyadlsgen2.dfs.core.windows.net/retail-gold/Top5SalesPerson'
29 WITH(
30     FILE_TYPE = 'PARQUET',
31     CREDENTIAL = (IDENTITY= 'Managed Identity')
32 );
33
34 -- Read Data
35 SELECT * FROM gold.topsalesperson;

```

## Query Data using SQL Server Management Studio

The screenshot shows the Object Explorer on the left, displaying the database structure for 'syn-project-learning-prd.sql.azuresynapse.net'. The 'Tables' node under 'SQLPOOL01' contains tables like gold.campaigns, gold.customerproduct, gold.productcount, gold.sales, gold.stores, gold.topproducts, gold.topsalesperson, gold.totalcustomercount1, gold.totalcustomercount2, gold.totalsales, gold.TotalSalesByBrand, gold.TotalSalesByCategory, gold.TotalSalesPerMonth, and gold.TotalSalesPerQuarter. The 'Views' node contains 'gold.v\_total\_sales\_per\_store'. The 'External Resources' node includes 'External Data Sources' with a connection to 'retail\_gold\_src'.

The main window shows a query editor with the following SQL code:

```

1  SELECT * FROM gold.topsalesperson;
2  SELECT * FROM gold.v_total_sales_per_store;

```

The results pane displays two tables:

	salesperson_id	salesperson_name	salesperson_role	first_name	last_name	total_amount	rank
1	SP00187	Tiffany Mitchell	Manager	Tiffany	Mitchell	4995.39	1
2	SP01103	Benjamin Rivas	Manager	Benjamin	Rivas	4992.32	2
3	SP00434	Jeffrey Hurst	Manager	Jeffrey	Hurst	4991.32	3
4	SP01347	Pamela Howell	Manager	Pamela	Howell	4988.21	4
5	SP00562	Samuel Shepherd	Manager	Samuel	Shepherd	4987.52	5
6	SP00178	Sarah Brown	Sales Associate	Sarah	Brown	4982.13	1
7	SP01447	Stephen Sanchez	Sales Associate	Stephen	Sanchez	4977.25	2
8	SP01511	Richard Scott	Sales Associate	Richard	Scott	4973.24	3
9	SP00170	Reginald Joseph	Sales Associate	Reginald	Joseph	4973.22	4
10	SP01360	Brandon Armstrong	Sales Associate	Brandon	Armstrong	4971.72	5

	store_name	store_type	location	total_sales
1	VALUEDEPOT	Large Malls or Complexes	Atlanta	19619.7
2	NOVAMART	Large Malls or Complexes	Las Vegas	577.73
3	DAILYMART	Large Malls or Complexes	Minneapolis	23363.14
4	HOMECENT...	Large Malls or Complexes	Chicago	8712.9
5	BUDGETBUY	Large Malls or Complexes	Milwaukee	8674.96
6	ACTIVELIFE	Large Malls or Complexes	Austin	17218.32
7	FRESHFLOW	Large Malls or Complexes	Baltimore	5606.07
8	VIVASTORES	Large Malls or Complexes	Los Angel...	7936.01
9	GROCEREA...	Large Malls or Complexes	Charlotte	12776.74
10	SHOPEASE	Large Malls or Complexes	Arlington	8689.8
11	ROOMIFY	Large Malls or Complexes	Albuquerque...	16603.28
12	FRESHNEST	Small Stores or Shops	Austin	15152.71
13	FURNICART	Large Malls or Complexes	Indianapolis	16934.91
14	OLTOUCH	Large Malls or Complexes	Orlando	21555.1