

# E-Commerce Data Lakehouse Implementation and Analysis

526-05

**Chaitanya Bejjanki**

**Simranjeet Kaur Gill**

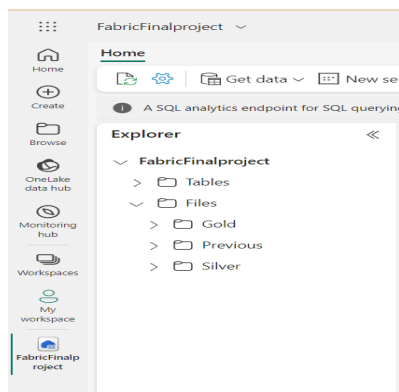
Firstly we have taken Data from Kagle has E-commerce website which it consists of sales & returns Sheets.

we have created a new Fabric Lake House and created sub folders as :

Previous

Silver

Gold



Created a Bronze\_Notebook in Microsoft \_Fabric to read my data set from my local system and used pandas to read my excel file and used pandas to join condition to combine Sales and Returns file.

And Saved it as Bronze\_Sales.parquet in Fabric Lakehouse

Later created a pipe line to copy a parquet file into azure blob storage container.

Now Created a Silver\_Notebook file and readed data from Azure blob storage (Bronze\_Sales.parquet)

Now I used Duckdb to read parquet file and added 2 new columns as Created\_date and Modified\_Date removed Duplicates from it

Saved it as Silver\_sales.parquet into my fabric lakehouse

Now created a pipeline to load this data into Azure blob storage.

Explorer

FabricFinalproject

Tables

bronze\_sales

silver\_sales

Files

Gold

Previous

bronze\_sales.parquet

Silver

Silver\_sales.parquet

bronze\_sales

	ABC	Order_ID	Order
1	FA-2015-417		1/12/202
2	FA-2015-3096		1/6/2023
3	FA-2015-373		1/18/202
4	FA-2015-3080		1/6/2023
5	FA-2015-1015		1/13/202
6	FA-2015-688		1/28/202
7	FA-2015-1138		1/2/2023
8	FA-2015-2012		1/30/202
9	FA-2015-12471		1/11/202
10	FA-2015-13277		1/3/2023
11	FA-2015-12581		1/25/202
12	FA-2015-13835		1/10/202

Microsoft Azure

Search resources, services, and docs (G+)

Home >

container-name

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

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

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

Location: container-name

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> _azuretmpfolder\$						... ..
<input type="checkbox"/> bronze-layer						... ..
<input type="checkbox"/> Gold						... ..
<input type="checkbox"/> bronze_sales.parquet	4/29/2024, 12:35:09 ...	Hot (Inferred)		Block blob	168.45 KiB	Available ...
<input type="checkbox"/> silver_sales.parquet	5/2/2024, 4:53:04 PM	Hot (Inferred)		Block blob	153.24 KiB	Available ...

Now created a Gold\_Notebook into fabric notebook I have used Duck DB to create dimensions and fact tables. Saved it into csv files,

Microsoft Azure

Search resources, services, and docs (G+I)

Home > container-name

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

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

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

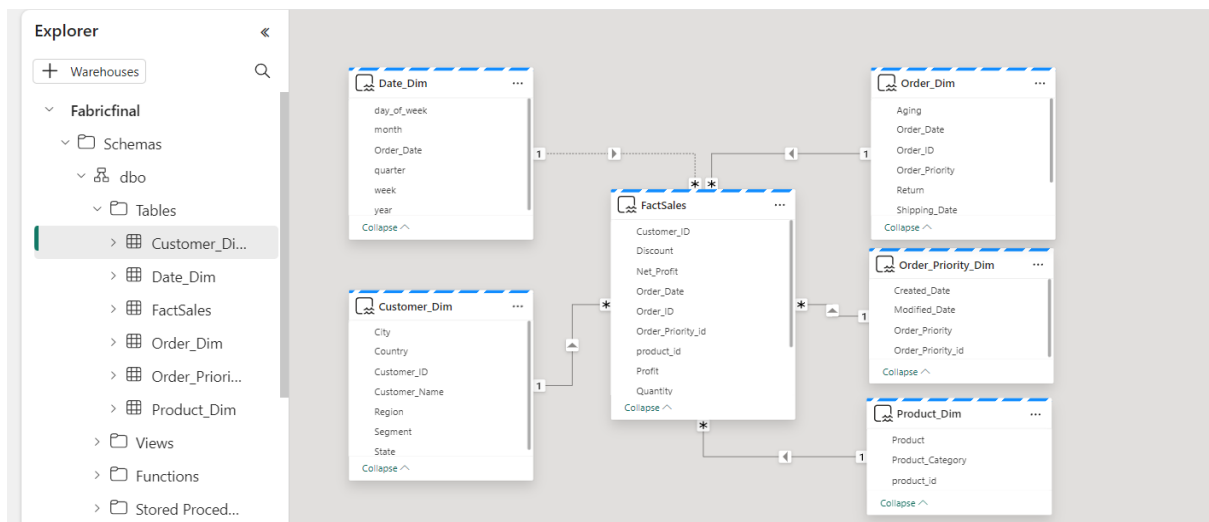
Location: container-name / Gold

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						...
Customer_Dim.csv	5/4/2024, 11:57:21 AM	Hot (Inferred)		Block blob	311.95 KiB	Available
Date_Dim.csv	5/4/2024, 11:57:21 AM	Hot (Inferred)		Block blob	1.05 KiB	Available
FactSales.csv	5/4/2024, 3:23:00 PM	Hot (Inferred)		Block blob	390.45 KiB	Available
Order_Dim.csv	5/4/2024, 11:57:21 AM	Hot (Inferred)		Block blob	268.09 KiB	Available
Order_Priority_Dim.csv	5/4/2024, 11:57:21 AM	Hot (Inferred)		Block blob	181 B	Available
Product_Dim.csv	5/4/2024, 11:57:21 AM	Hot (Inferred)		Block blob	1.13 KiB	Available

Now we have created a pipeline to load this data into fabric Data warehouse added all csv files into tables in Fabric warehouse



Later added all these warehouse into power bi through power BI semantic model in power bi visualization.

Added this ware house into power bi created reports.

