Cloud Only Data Pipeline Implemetation

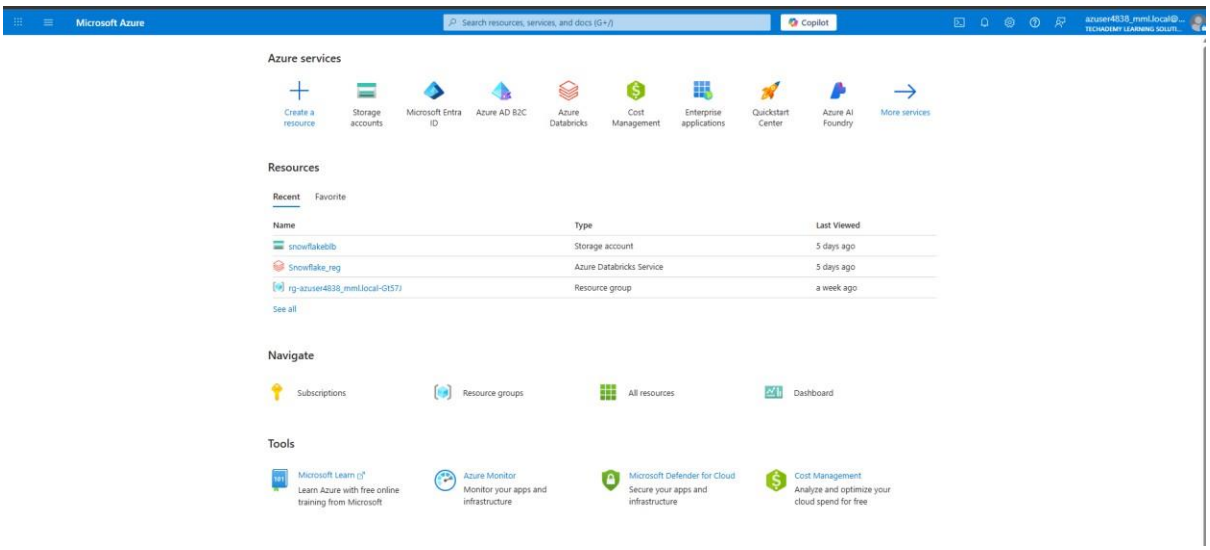
1. **Receives CSV files** from the sales team (monthly sales data).
2. **Stores them in Azure Blob Storage**.
3. **Processes and ingests them into Snowflake** using **Snowpark** (Snowflake's Python API).
4. **Transforms the data** into structured tables and views.
5. **Visualizes the data** in **Power BI** for business users.

Sales Team → Azure Blob Storage → Azure Databricks (Snowpark) → Snowflake → Power BI

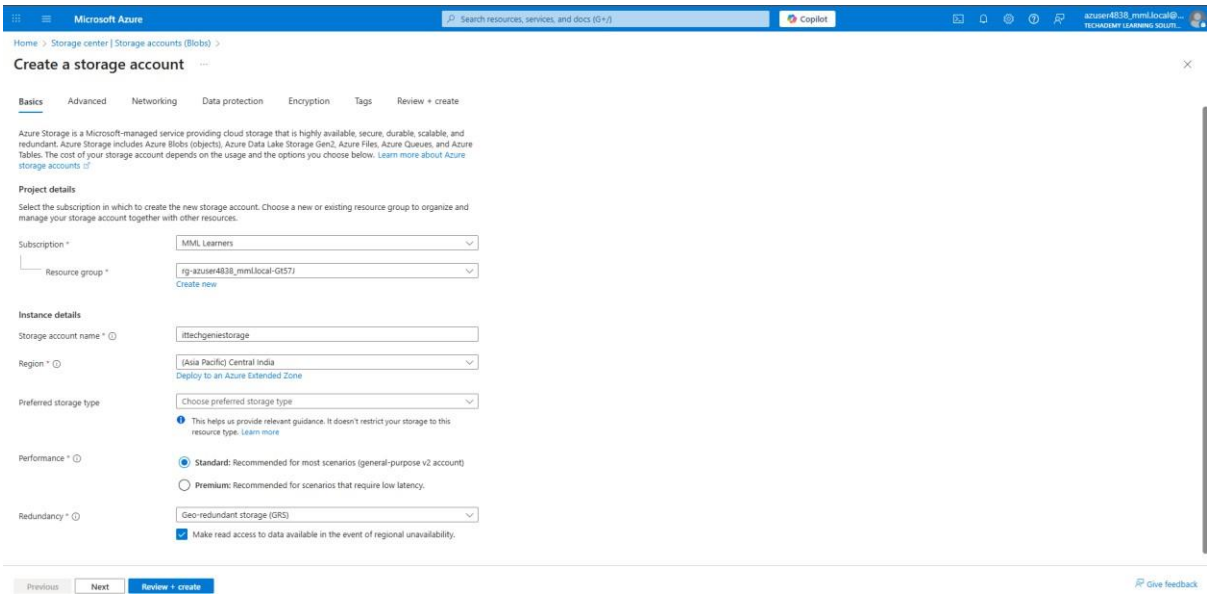
Phase 1: Setting Up Azure Blob Storage

Step 1: Create a Storage Account

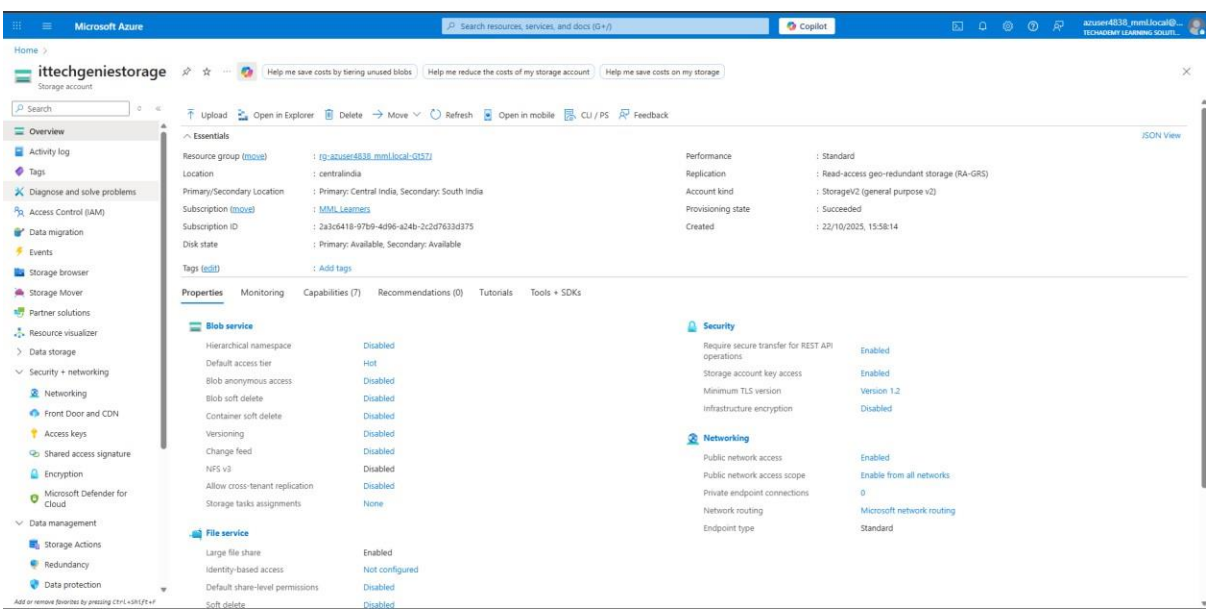
1. Sign in to the Azure Portal.



2. Click **"Create a resource"** and search for **"Storage account"**.



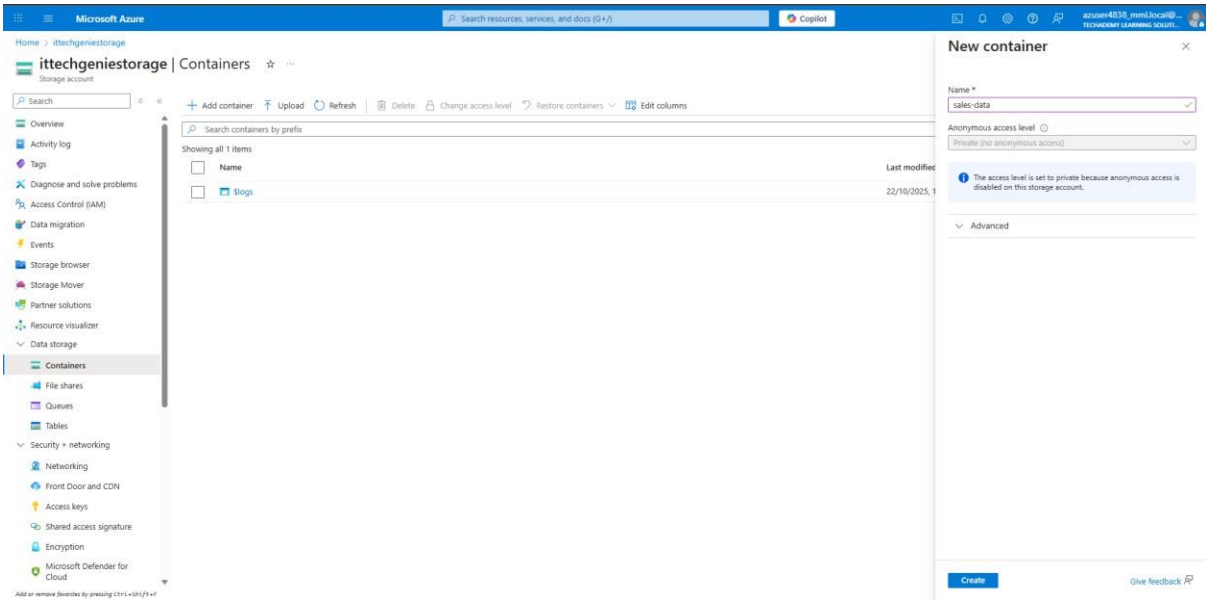
3. Choose your subscription and either select an existing resource group or create a new one (e.g., `ItTechGenie-RG`).
4. Set the storage account name (e.g., `itgretailstorage`) and region (e.g., East US).
5. Choose **Standard** performance and **Locally-redundant storage (LRS)** for redundancy.
6. Click **Review + Create**, then **Create**.



Step 2: Create a Container

1. Inside your storage account, go to **"Containers"**.

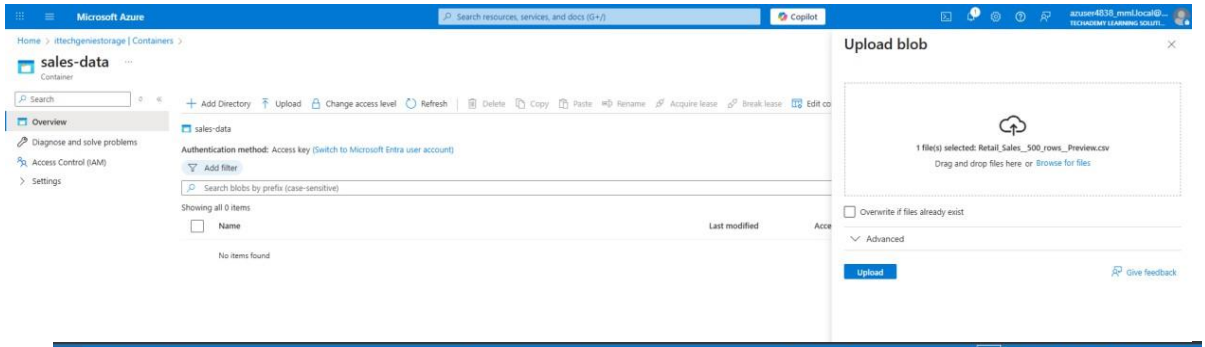
2. Click **"+ Container"** and name it something like `monthly-sales`.
3. Set the access level to **Private**.

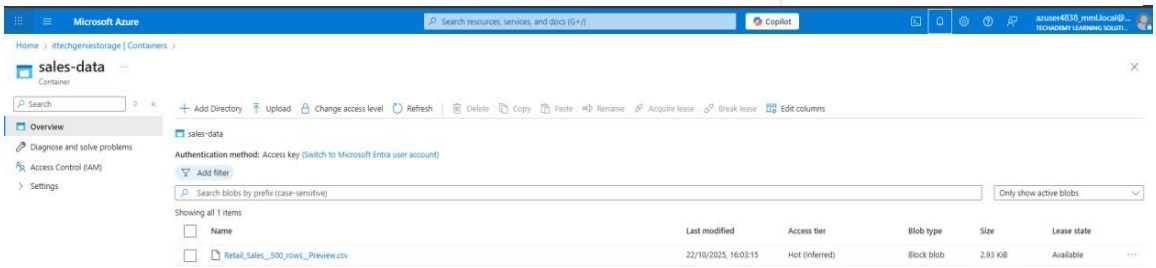


Step 3: Upload the CSV File

Option A: Azure Portal

- Navigate to the `monthly-sales` container.
- Click **Upload**, select your `.csv` file (e.g., `sales_october.csv`), and upload it.





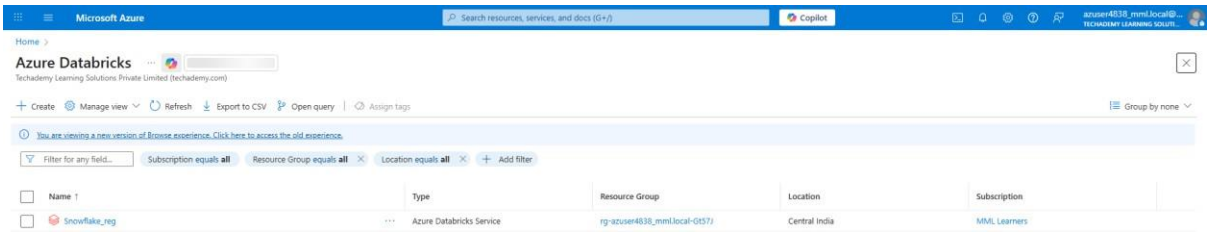
Option B: Azure Cloud Shell

```
az storage blob upload \
--account-name itgretailstorage \
 --container-name monthly-sales \
 --name sales_october.csv \
 --file sales_october.csv \
 --auth-mode login
```

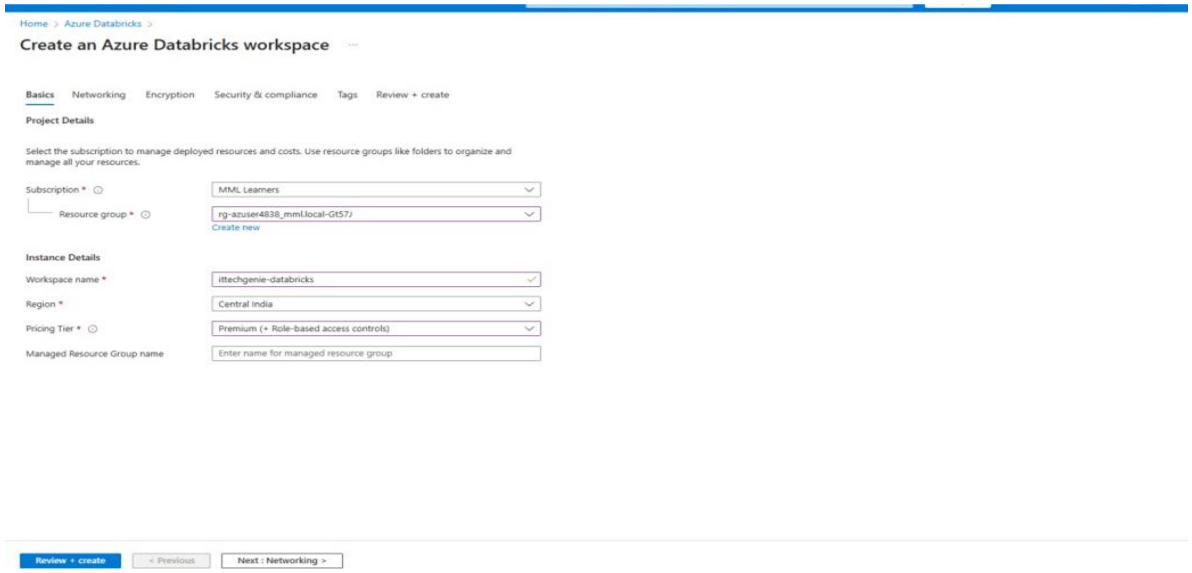
Phase 2: Provisioning Azure Databricks

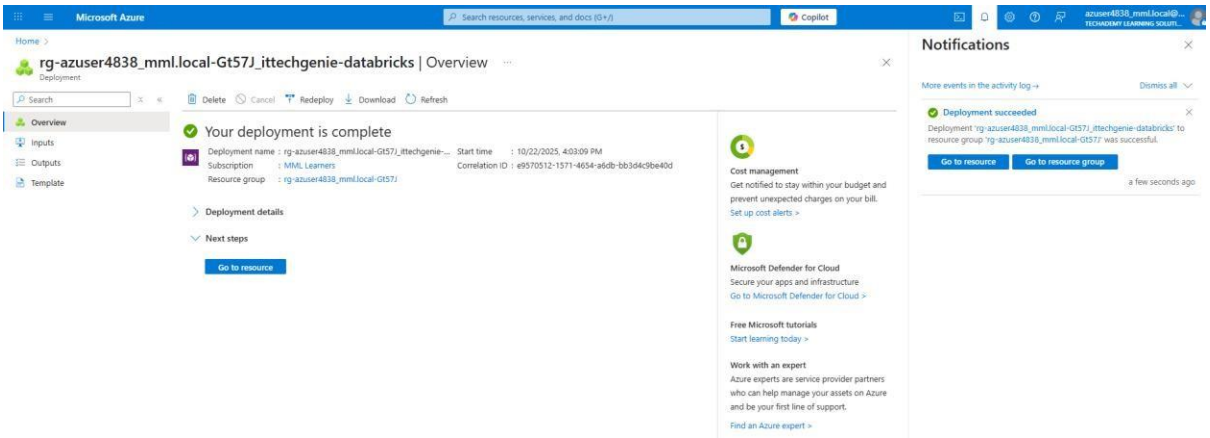
Step 1: Create a Databricks Workspace

1. In Azure Portal, click **Create a resource** and search for **Azure Databricks**.



2. Set the workspace name (e.g., `itgretail-databricks`) and region (same as storage).
3. Choose the **Premium** pricing tier.
4. Click **Review + Create**, then **Create**.

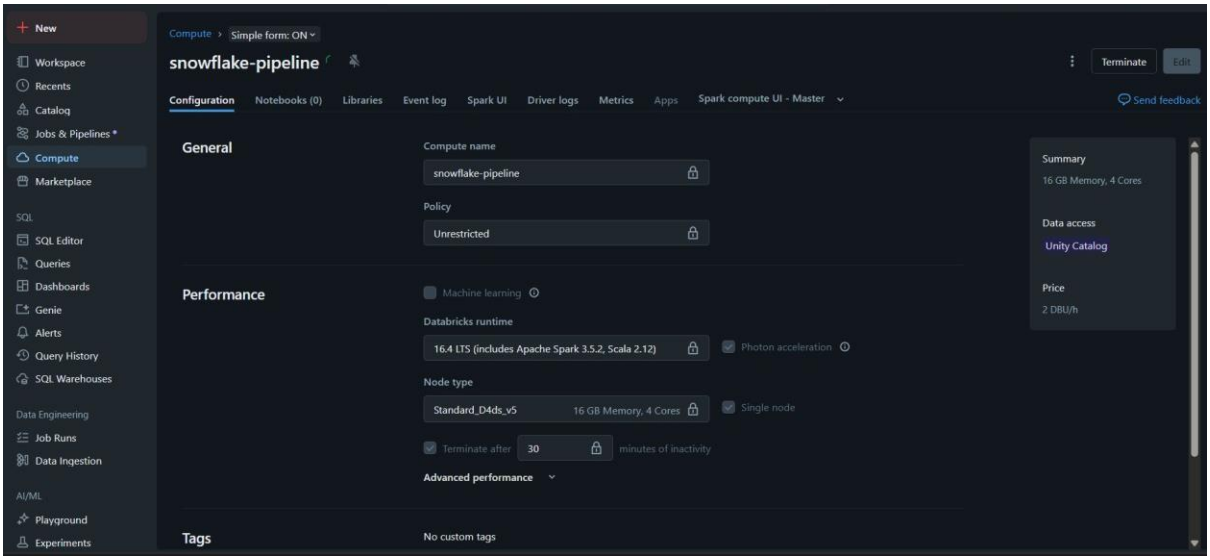




Step 2: Launch and Configure a Cluster

1. Open the Databricks workspace.
2. Go to **Compute** and click **Create Cluster**.
3. **Configure:**

- Name: `snowflake-pipeline`
- Mode: Single Node
- Runtime: 12.2 LTS
- Node Type: `Standard_DS3_v2`
- Auto-termination: 30 minutes



Phase 3: Snowflake Setup

Step 1: Create Snowflake Objects

Run the following SQL in Snowflake

```
CREATE WAREHOUSE ITG_WAREHOUSE WITH WAREHOUSE_SIZE = XSMALL
AUTO_SUSPEND = 300 AUTO_RESUME = TRUE;

CREATE DATABASE ITG_SALES_DB;
CREATE SCHEMA ITG_SALES_DB.RAW;
CREATE SCHEMA ITG_SALES_DB.MODELED;
CREATE SCHEMA ITG_SALES_DB.REPORTING;

CREATE ROLE DATA_ENGINEER;
GRANT USAGE ON WAREHOUSE ITG_WAREHOUSE TO ROLE DATA_ENGINEER;
GRANT ALL ON DATABASE ITG_SALES_DB TO ROLE DATA_ENGINEER;
```

Step 2: Connect Snowflake to Azure

```
CREATE STORAGE INTEGRATION azure_sales_integration
  TYPE = EXTERNAL_STAGE
  STORAGE_PROVIDER = AZURE
  ENABLED = TRUE
  AZURE_TENANT_ID = '<your-tenant-id>'
  STORAGE_ALLOWED_LOCATIONS =
('azure://itgretailstorage.blob.core.windows.net/monthly-sales/');
```

Then run:

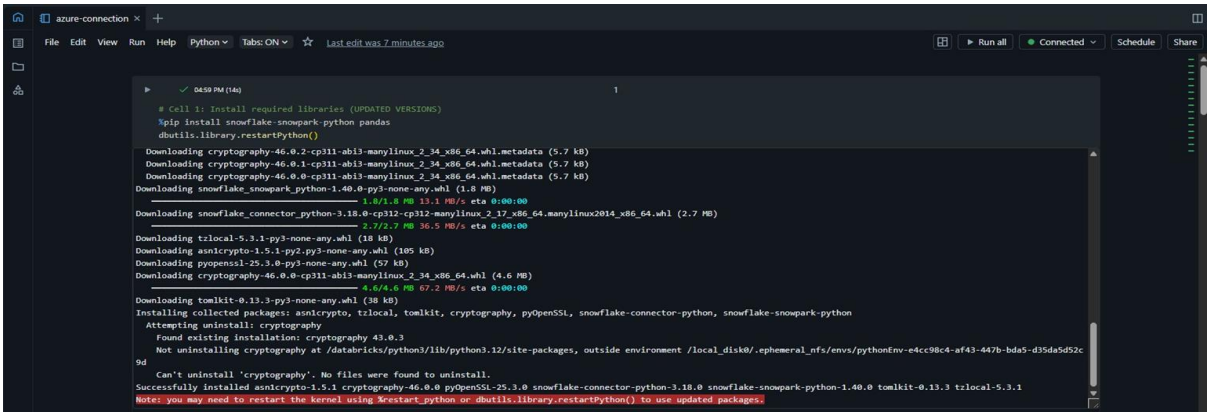
```
DESC STORAGE INTEGRATION azure_sales_integration;
```

Use the `AZURE_CONSENT_URL` to authorize Snowflake to access your Azure storage.

Phase 4: Data Ingestion with Snowpark in Databricks

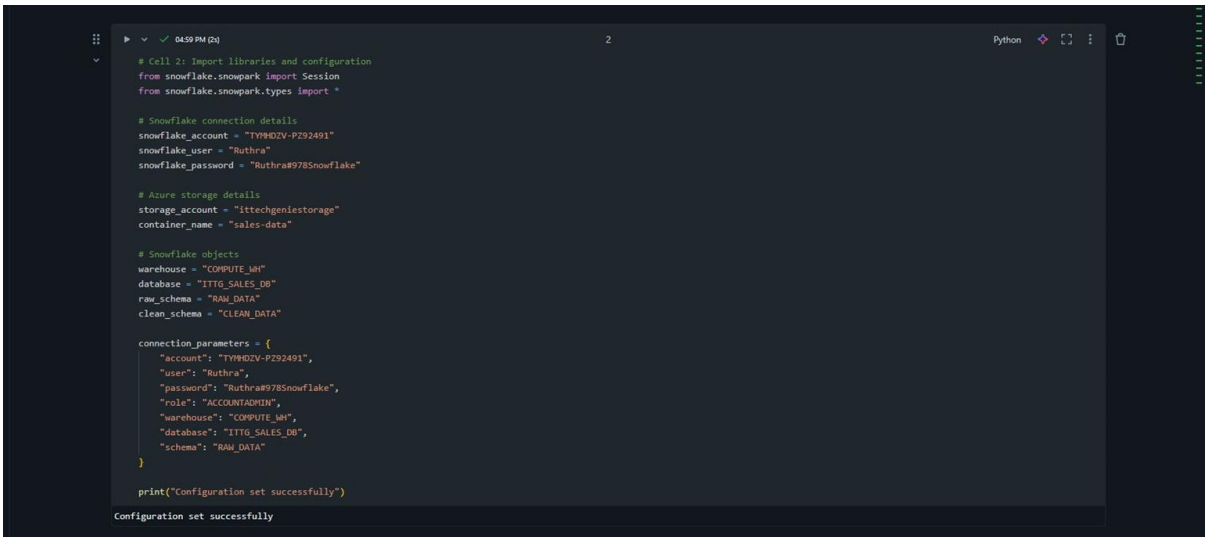
Step 1: Install Required Libraries

```
%pip install snowflake-snowpark-python pandas
dbutils.library.restartPython()
```



Step 2: Configure Snowflake Connection

```
from snowflake.snowpark import Session

connection_parameters = {
    "account": "<your_account>",
    "user": "<your_user>",
    "password": "<your_password>",
    "role": "ACCOUNTADMIN",
    "warehouse": "ITG_WAREHOUSE",
    "database": "ITG_SALES_DB",
    "schema": "RAW"
}

session = Session.builder.configs(connection_parameters).create()
```

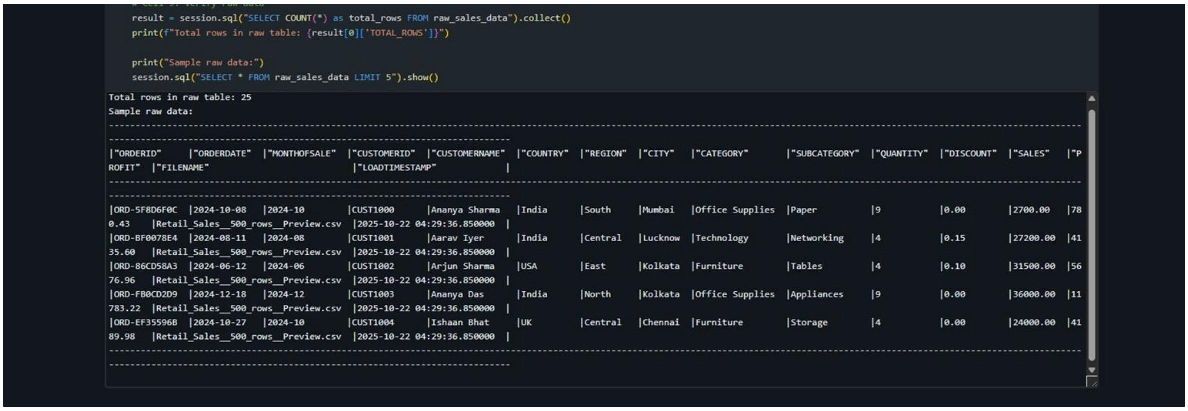


Step 3: Create Stage and File Format

```
session.sql("""
CREATE OR REPLACE FILE FORMAT csv_format
  TYPE = 'CSV'
  FIELD_DELIMITER = ','
  SKIP_HEADER = 1
  NULL_IF = ('NULL', 'null')
  EMPTY_FIELD_AS_NULL = TRUE;
""").collect()

session.sql("""
CREATE OR REPLACE STAGE azure_stage
  URL = 'azure://itgretailstorage.blob.core.windows.net/monthly-sales/'
  CREDENTIALS = (AZURE_SAS_TOKEN = '<your-sas-token>')
  FILE_FORMAT = csv_format;
""").collect()
```

Step 4: Load Data into Snowflake

```
session.sql("""
COPY INTO raw_sales_data (
  OrderID, OrderDate, MonthOfSale, CustomerID, CustomerName,
  Country, Region, City, Category, Subcategory,
  Quantity, Discount, Sales, Profit, FileName
)
FROM (
  SELECT $1, $2, $3, $4, $5, $6, $7, $8, $9, $10,
         $11, $12, $13, $14, METADATA$FILENAME
  FROM @azure_stage/sales_october.csv
)
FILE_FORMAT = (FORMAT_NAME = csv_format)
ON_ERROR = 'CONTINUE';
""").collect()
```



```
    result = session.sql("SELECT COUNT(*) as total_rows FROM raw_sales_data").collect()
    print(f"Total rows in raw table: {result[0]['TOTAL_ROWS']}")

    print("Sample raw data:")
    session.sql("SELECT * FROM raw_sales_data LIMIT 5").show()
```

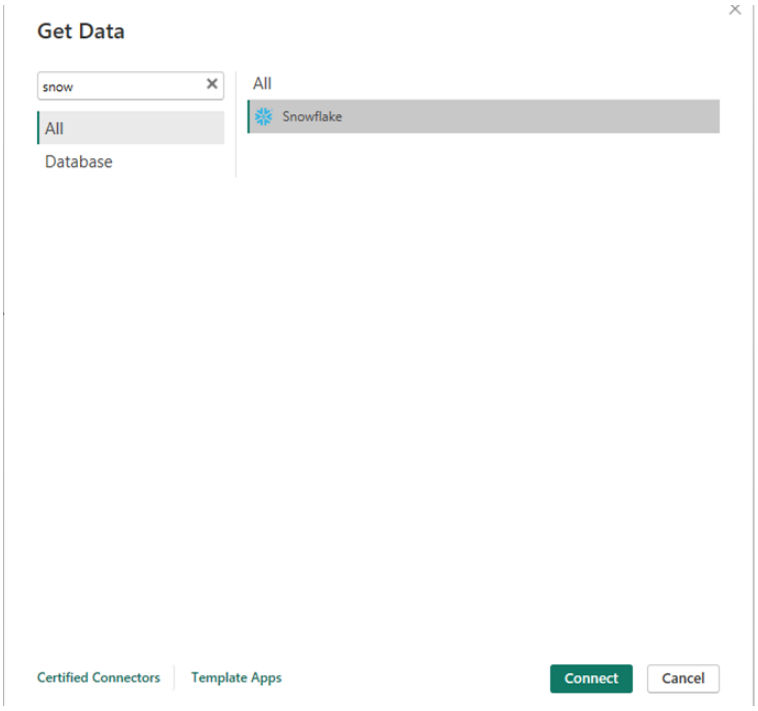
```
Total rows in raw table: 25
Sample raw data:
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------
|"ORDERID"     |"ORDERDATE"  |"MONTHOFSALE"  |"CUSTOMERID"  |"CUSTOMERNAME"  |"COUNTRY"  |"REGION"  |"CITY"   |"CATEGORY"      |"SUBCATEGORY"  |"QUANTITY"  |"DISCOUNT"  |"SALES"   |"P
ROFIT"  |"FILENAME"                          |"LOADTIMESTAMP"            |
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------
|ORD-5F8D6F0C  |2024-10-08   |2024-10        |CUST1000      |Ananya Sharma   |India      |South     |Mumbai   |Office Supplies |Paper          |9           |0.00        |2700.00   |78
0.43    |Retail_Sales__500_rows__Preview.csv  |2025-10-22 04:29:36.850000  |
|ORD-BF0078E4  |2024-08-11   |2024-08        |CUST1001      |Aarav Iyer      |India      |Central   |Lucknow  |Technology      |Networking     |4           |0.15        |27200.00  |41
35.60   |Retail_Sales__500_rows__Preview.csv  |2025-10-22 04:29:36.850000  |
|ORD-86CD58A3  |2024-06-12   |2024-06        |CUST1002      |Arjun Sharma    |USA        |East      |Kolkata  |Furniture       |Tables         |4           |0.10        |31500.00  |56
76.96   |Retail_Sales__500_rows__Preview.csv  |2025-10-22 04:29:36.850000  |
|ORD-FB0CD2D9  |2024-12-18   |2024-12        |CUST1003      |Ananya Das      |India      |North     |Kolkata  |Office Supplies |Appliances     |9           |0.00        |36000.00  |11
783.22  |Retail_Sales__500_rows__Preview.csv  |2025-10-22 04:29:36.850000  |
|ORD-EF35596B  |2024-10-27   |2024-10        |CUST1004      |Ishaan Bhat     |UK         |Central   |Chennai  |Furniture       |Storage        |4           |0.00        |24000.00  |41
89.98   |Retail_Sales__500_rows__Preview.csv  |2025-10-22 04:29:36.850000  |
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------
```

```
    session.close()
    print("Snowpark session closed")
    print("Pipeline execution completed successfully!")
    print("\nNext steps: Connect Power BI to Snowflake using:")
    print("Database: ITTG_SALES_DB")
    print("Schema: CLEAN_DATA")
    print("View: VW_POWERBI_SALES_DASHBOARD")
```

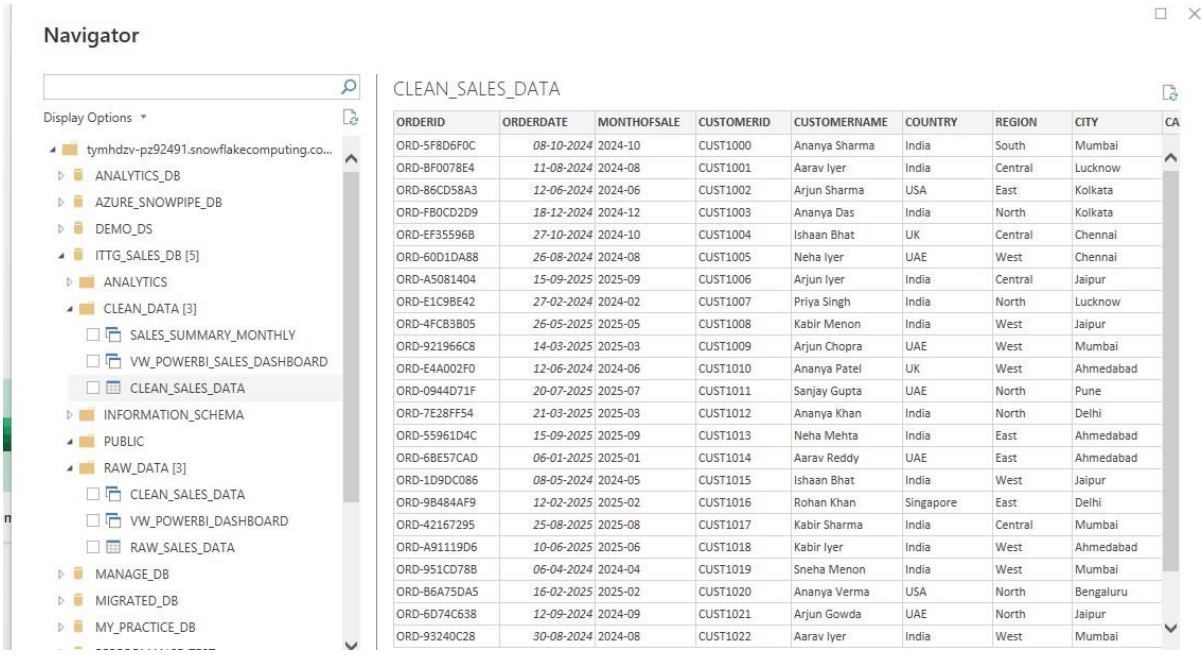
```
Snowpark session closed
Pipeline execution completed successfully!

Next steps: Connect Power BI to Snowflake using:
Database: ITTG_SALES_DB
Schema: CLEAN_DATA
View: VW_POWERBI_SALES_DASHBOARD
```

Phase 5: Power BI Reporting

1. Open Power BI Desktop.
2. Choose **Snowflake** as the data source.
3. Connect using your Snowflake credentials and select the `ITG_SALES_DB.REPORTING` schema.



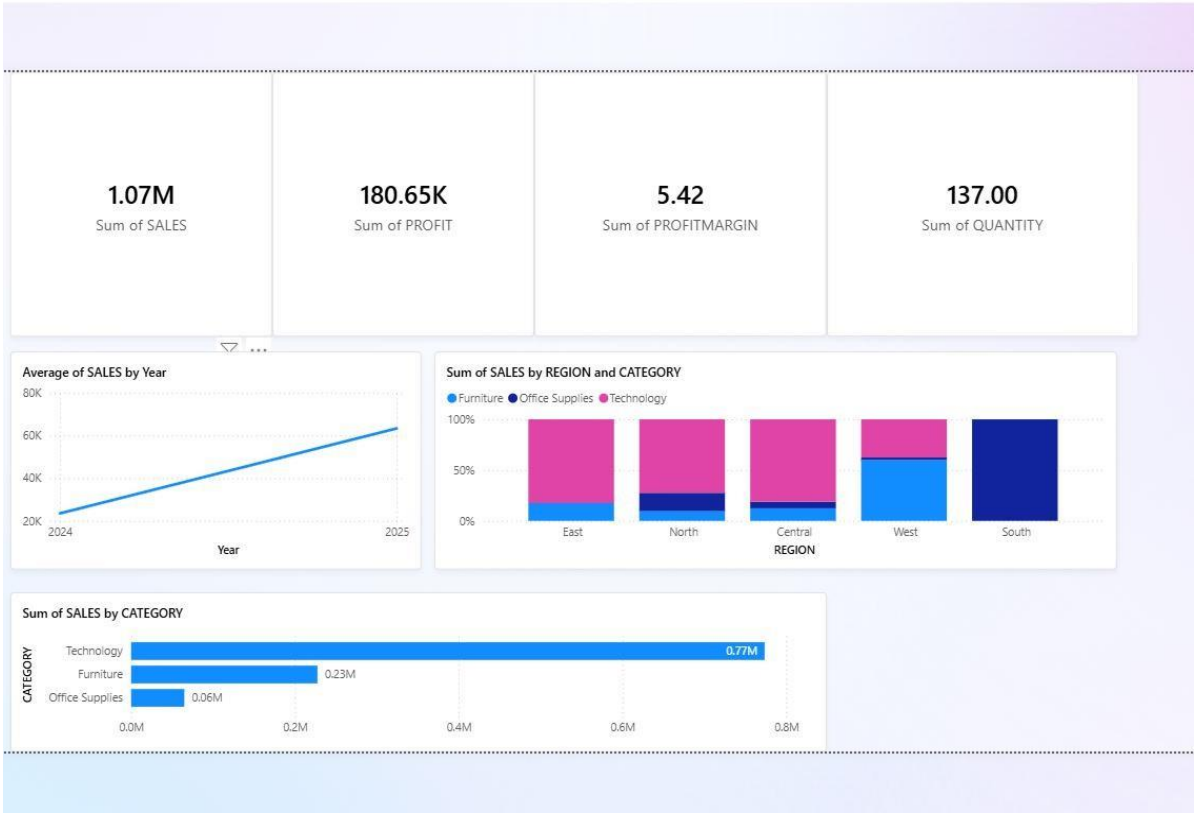
4. Import Data

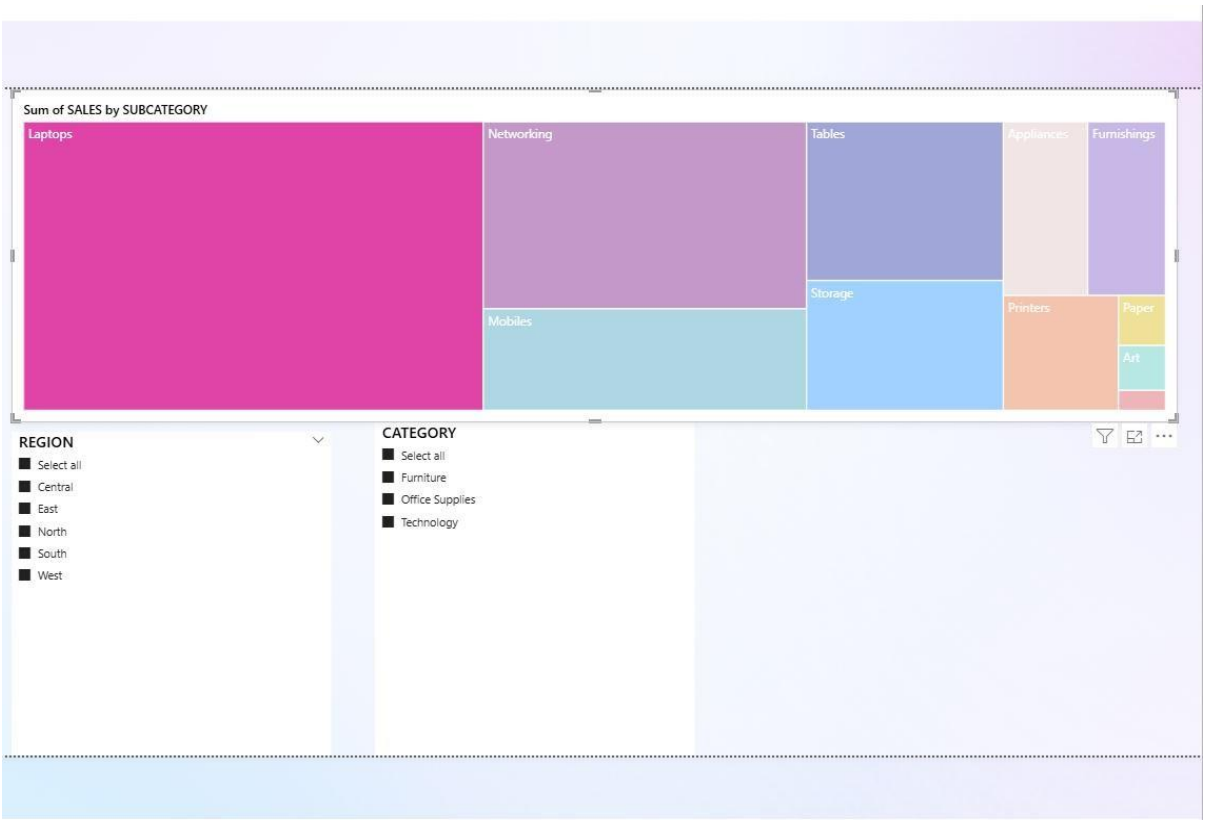


ORDERID	ORDERDATE	MONTHOFSALE	CUSTOMERID	CUSTOMERNAME	COUNTRY	REGION	CITY	CA
ORD-5F8D6F0C	08-10-2024	2024-10	CUST1000	Ananya Sharma	India	South	Mumbai	
ORD-BF0078E4	11-08-2024	2024-08	CUST1001	Aarav Iyer	India	Central	Lucknow	
ORD-86CD58A3	12-06-2024	2024-06	CUST1002	Arjun Sharma	USA	East	Kolkata	
ORD-FB0CD2D9	18-12-2024	2024-12	CUST1003	Ananya Das	India	North	Kolkata	
ORD-EF35596B	27-10-2024	2024-10	CUST1004	Ishaan Bhat	UK	Central	Chennai	
ORD-60D1DA88	26-08-2024	2024-08	CUST1005	Neha Iyer	UAE	West	Chennai	
ORD-A5081404	15-09-2025	2025-09	CUST1006	Arjun Iyer	India	Central	Jaipur	
ORD-E1C9BE42	27-02-2024	2024-02	CUST1007	Priya Singh	India	North	Lucknow	
ORD-4FCB3B05	26-05-2025	2025-05	CUST1008	Kabir Menon	India	West	Jaipur	
ORD-921966C8	14-03-2025	2025-03	CUST1009	Arjun Chopra	UAE	West	Mumbai	
ORD-E4A002F0	12-06-2024	2024-06	CUST1010	Ananya Patel	UK	West	Ahmedabad	
ORD-0944D71F	20-07-2025	2025-07	CUST1011	Sanjay Gupta	UAE	North	Pune	
ORD-7E28FF54	21-03-2025	2025-03	CUST1012	Ananya Khan	India	North	Delhi	
ORD-55961D4C	15-09-2025	2025-09	CUST1013	Neha Mehta	India	East	Ahmedabad	
ORD-6BE57CAD	06-01-2025	2025-01	CUST1014	Aarav Reddy	UAE	East	Ahmedabad	
ORD-1D9DC086	08-05-2024	2024-05	CUST1015	Ishaan Bhat	India	West	Jaipur	
ORD-9B484AF9	12-02-2025	2025-02	CUST1016	Rohan Khan	Singapore	East	Delhi	
ORD-42167295	25-08-2025	2025-08	CUST1017	Kabir Sharma	India	Central	Mumbai	
ORD-A91119D6	10-06-2025	2025-06	CUST1018	Kabir Iyer	India	West	Ahmedabad	
ORD-951CD78B	06-04-2024	2024-04	CUST1019	Sneha Menon	India	West	Mumbai	
ORD-B6A75DA5	16-02-2025	2025-02	CUST1020	Ananya Verma	USA	North	Bengaluru	
ORD-6D74C638	12-09-2024	2024-09	CUST1021	Arjun Gowda	UAE	North	Jaipur	
ORD-93240C28	30-08-2024	2024-08	CUST1022	Aarav Iyer	India	West	Mumbai	

5. Load the relevant tables or views.
6. Build visualizations like:

- Monthly sales trends
- Top-selling categories
- Regional performance


1.07M
Sum of SALES
180.65K
Sum of PROFIT
5.42
Sum of PROFITMARGIN
137.00
Sum of QUANTITY
Average of SALES by Year
80K
60K
40K
20K
2024
2025
Year
Sum of SALES by REGION and CATEGORY
Furniture
Office Supplies
Technology
100%
50%
0%
East
North
Central
West
South
REGION
Sum of SALES by CATEGORY
CATEGORY
Technology
Furniture
Office Supplies
0.77M
0.23M
0.06M
0.0M
0.2M
0.4M
0.6M
0.8M


Sum of SALES by SUBCATEGORY
Laptops
Networking
Tables
Appliances
Furnishings
Storage
Mobiles
Printers
Paper
Art
REGION
Select all
Central
East
North
South
West
CATEGORY
Select all
Furniture
Office Supplies
Technology