

Cloud Only Data Pipeline Implementation

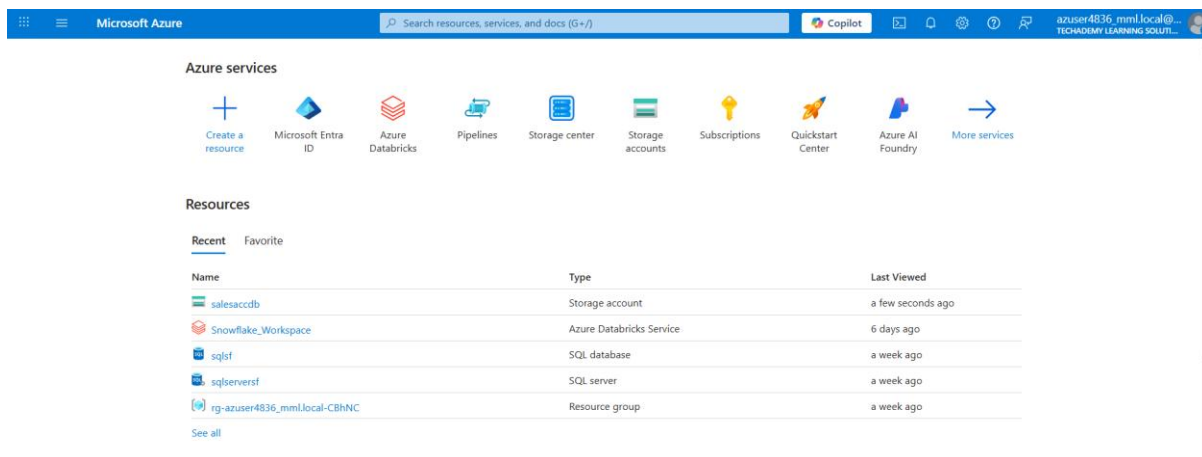
1. **Receives CSV files** from the sales team (monthly sales data).
2. **Stores them in Azure Blob Storage.**
3. **Processes and ingests them into Snowflake** using **Snowpark** (Snowflake's Python API).
4. **Transforms the data** into structured tables and views.
5. **Visualizes the data in Power BI**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

Microsoft Azure

Home > Storage center > Storage accounts (Blobs) >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group * [Create new](#)

Instance details

Storage account name *

Region * [Deploy to an Azure Extended Zone](#)

Preferred storage type

☒ This helps us provide relevant guidance. It doesn't restrict your storage to this resource type. [Learn more](#)

Performance * ☒ Standard: Recommended for most scenarios (general-purpose v2 account)
☐ Premium: Recommended for scenarios that require low latency.

Redundancy *

[Previous](#) [Next](#) [Review + create](#)

[Give feedback](#)

3. Choose your subscription and either select an existing resource group or create a new one
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**.

Microsoft Azure

Home > ittechgeniestorage > Storage account

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

Copilot

azuser4836_mml.local@...
TECHACADEMY LEARNING SOLUTIONS

Help me save costs by tiering unused blobs Help me reduce the costs of my storage account Help me save costs on my storage

Upload Open in Explorer Delete Move Refresh Open in mobile CU / PS Feedback

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Storage Mover Partner solutions Resource visualizer Data storage Security + networking Networking Front Door and CDN Access keys Shared access signature Encryption Microsoft Defender for Cloud Data management Storage Actions Redundancy Data protection

Essentials

Resource group (move) [rg-azuser4836_mml.local-Q157J](#)

Location [centralindia](#)

Primary/Secondary Location [Primary: Central India, Secondary: South India](#)

Subscription (move) [MML Learners](#)

Subscription ID [2a3c6418-97b9-4d96-a24b-2c2d7633d375](#)

Disk state [Primary: Available, Secondary: Available](#)

Tags (edit) [Add tags](#)

Performance [Standard](#)

Replication [Read-access geo-redundant storage \(RA-GRS\)](#)

Account kind [StorageV2 \(general purpose v2\)](#)

Provisioning state [Succeeded](#)

Created [22/10/2025, 15:58:14](#)

Properties Monitoring Capabilities (7) Recommendations (0) Tutorials Tools + SDKs

Blob service

Hierarchical namespace [Disabled](#)

Default access tier [Hot](#)

Blob anonymous access [Disabled](#)

Blob soft delete [Disabled](#)

Container soft delete [Disabled](#)

Versioning [Disabled](#)

Change feed [Disabled](#)

NFS v3 [Disabled](#)

Allow cross-tenant replication [Disabled](#)

Storage tasks assignments [None](#)

File service

Large file share [Enabled](#)

Identity-based access [Not configured](#)

Default share-level permissions [Disabled](#)

Soft delete [Disabled](#)

Security

Require secure transfer for REST API operations [Enabled](#)

Storage account key access [Enabled](#)

Minimum TLS version [Version 1.2](#)

Infrastructure encryption [Disabled](#)

Networking

Public network access [Enabled](#)

Public network access scope [Enable from all networks](#)

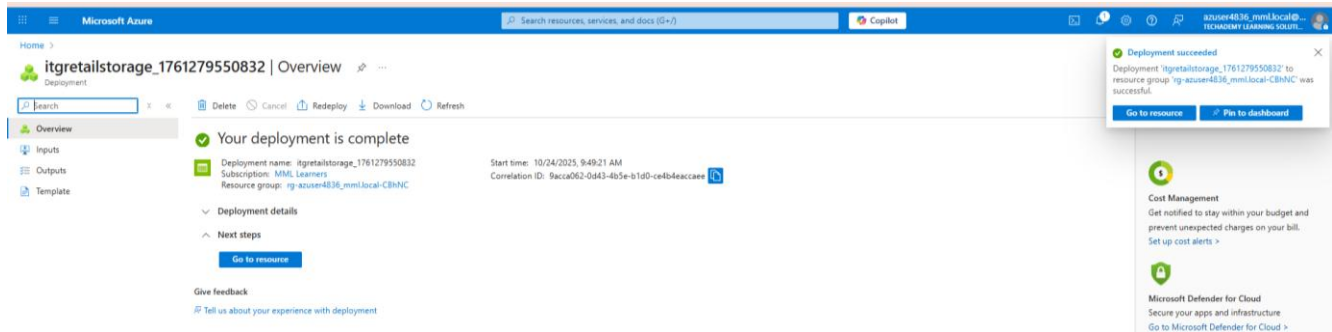
Private endpoint connections [0](#)

Network routing [Microsoft network routing](#)

Endpoint type [Standard](#)

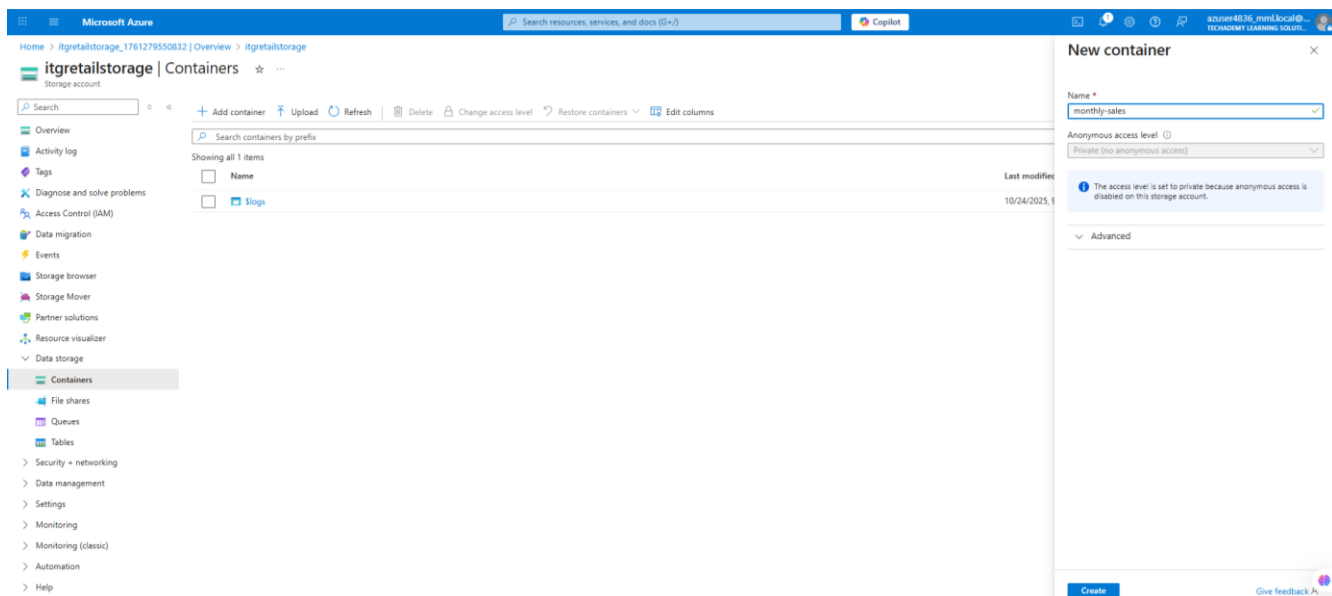
[JSON View](#)

Add or remove favorites by pressing Ctrl+SHIFT+F



Step 2: Create a Container

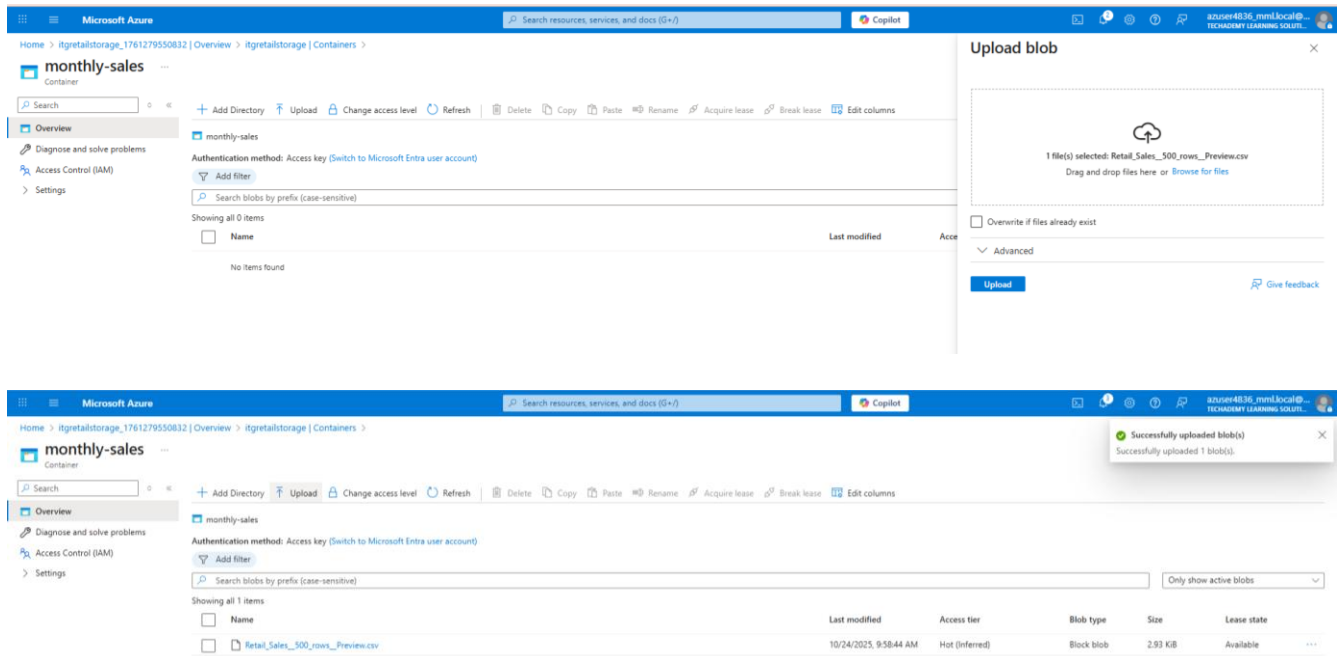
1. Inside your storage account, go to "**Containers**".
2. Click "**+ Container**" and name it something like monthly-sales.
3. Set the access level to **Private**.



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**.

Microsoft Azure

Home > Azure Databricks >

Create an Azure Databricks workspace

Basics Networking Encryption Security & compliance Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

Resource group * [Create new](#)

Instance Details

Workspace name *

Region *

Pricing Tier * ✕

We selected the recommended pricing tier for your workspace. You can change the tier based on your needs.

Managed Resource Group name

[Review + create](#) [< Previous](#) [Next: Networking >](#)

Microsoft Azure

Home > rg-azuser4838_mml-local-Gt57J_ittechgenie-databricks | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name : rg-azuser4838_mml-local-Gt57J_ittechgenie-... Start time : 10/22/2025, 4:03:09 PM

Subscription : MML Learners Correlation ID : e9570512-1371-4654-a6db-bb3d4c3be40d

Resource group : rg-azuser4838_mml-local-Gt57J

Deployment details

Next steps

[Go to resource](#)

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.

[Set up cost alerts >](#)

Microsoft Defender for Cloud

Secure your apps and infrastructure

[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials

[Start learning today >](#)

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

[Find an Azure expert >](#)

Notifications

More events in the activity log > [Dismiss all](#)

Deployment succeeded

Deployment 'rg-azuser4838_mml-local-Gt57J_ittechgenie-databricks' to resource group 'rg-azuser4838_mml-local-Gt57J' was successful.

[Go to resource](#) [Go to resource group](#)

a few seconds ago

Step 2: Launch and Configure a Cluster

1. Open the Databricks workspace.

Microsoft Azure

Home > rg-azuser4838_mml-local-CBHNC_itgretail-databricks | Overview

itgretail-databricks

Azure Databricks Service

Search Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Resource visualizer

Settings

Monitoring

Automation

Help

Essentials

Status : Active

Resource group : rg-azuser4838_mml-local-CBHNC

Location : Central India

Subscription : MML Learners

Subscription ID : 2a3c6418-97b9-4d96-a24b-2c2d7633d375

Tags [\(edit\)](#) [Add tags](#)

Managed Resource Group : databricks-rg-itgretail-databricks-dvjmszk5tmsc

URL : <https://adb-5794110499590008.8.azure.databricks.net>

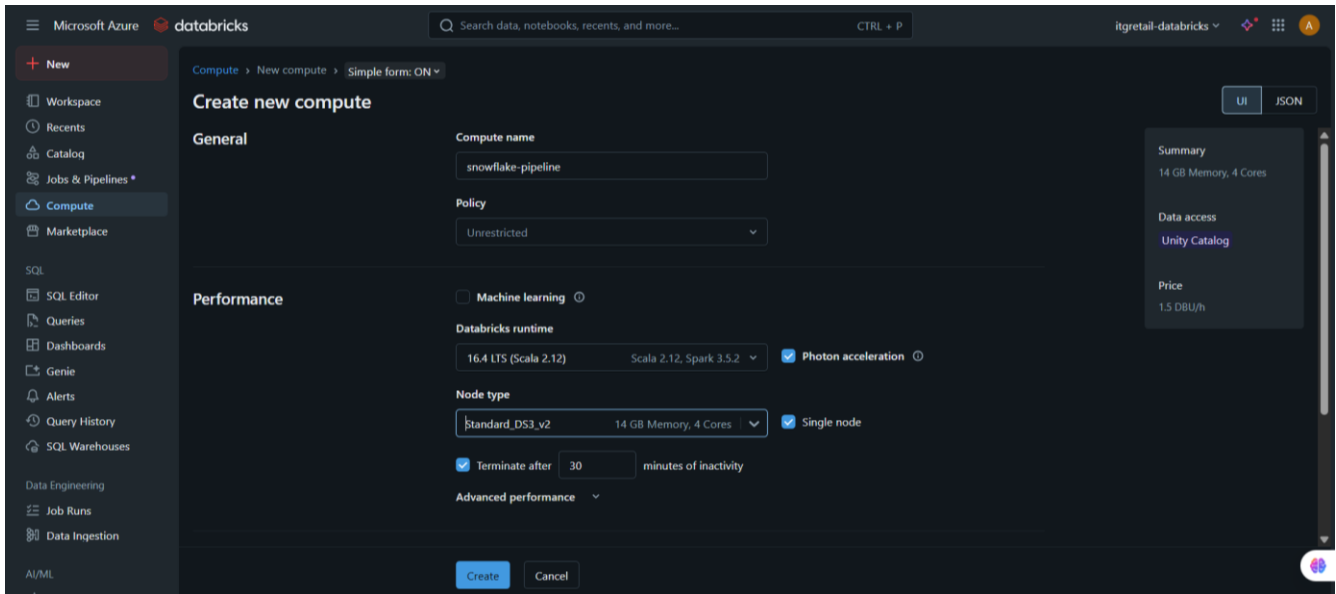
Pricing Tier : Premium (+ Role-based access controls) [Click to change](#)

Enable No Public IP : Yes

[Launch 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

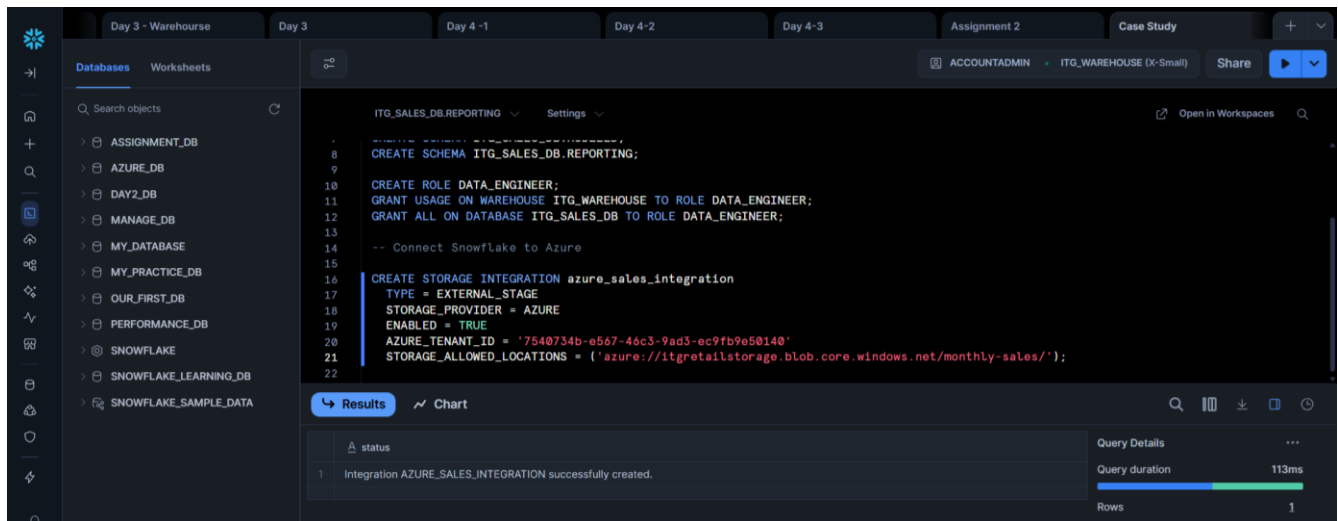
```
CREATE WAREHOUSE ITG_WAREHOUSE WITH WAREHOUSE_SIZE = XSMALL
AUTO_SUSPEND = 300 AUTO_RESUME = TRUE;
```

```
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;
```

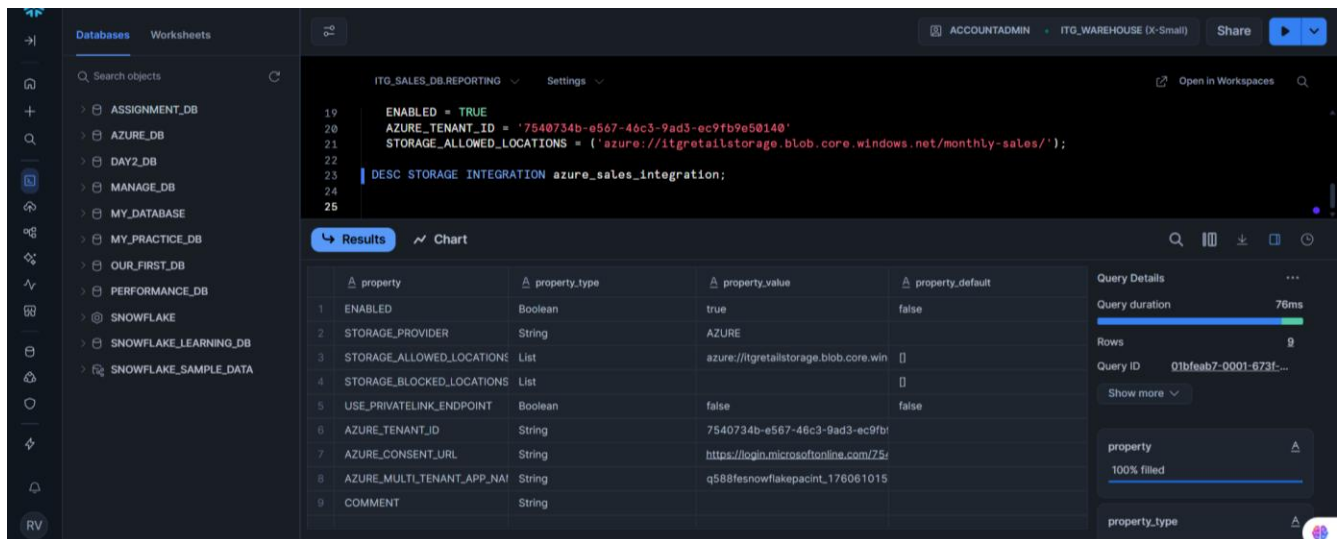
Step 2: Connect Snowflake to Azure

```
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;
```

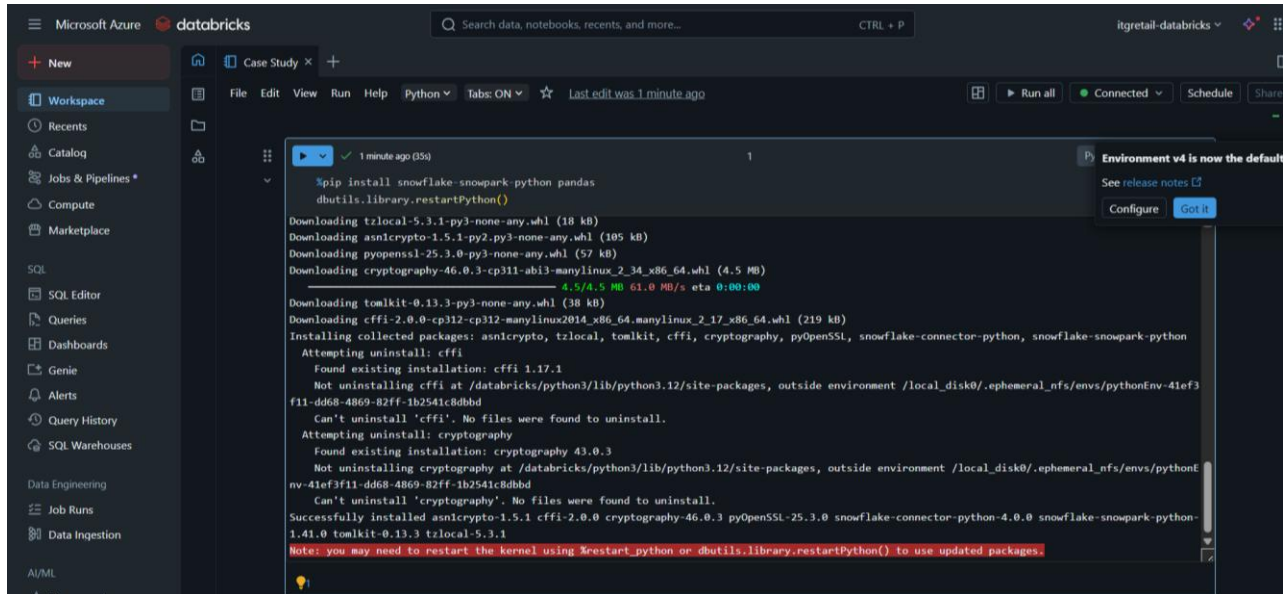


Use the AZURE_CONSENT_URL to authorize Snowflake to access your Azure storage.

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()
```



```
04:59 PM [D]
Python

# Cell 2: Import libraries and configuration
from snowflake.snowpark import Session
from snowflake.snowpark.types import *

# Snowflake connection details
snowflake_account = "TYV9HDZV-P292491"
snowflake_user = "Ruthra"
snowflake_password = "Ruthra#978Snowflake"

# Azure storage details
storage_account = "ittechgeniestorage"
container_name = "sales-data"

# Snowflake objects
warehouse = "COMPUTE_WH"
database = "ITIQ_SALES_DB"
raw_schema = "RAW_DATA"
clean_schema = "CLEAN_DATA"

connection_parameters = {
    "account": "TYV9HDZV-P292491",
    "user": "Ruthra",
    "password": "Ruthra#978Snowflake",
    "role": "ACCOUNTADMIN",
    "warehouse": "COMPUTE_WH",
    "database": "ITIQ_SALES_DB",
    "schema": "RAW_DATA"
}

print("Configuration set successfully")

Configuration set successfully
```

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()
```

```
session.sql("""
CREATE OR REPLACE FILE FORMAT csv_sales_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_sales_stage
URL = 'azure://ittechgeniestorage.blob.core.windows.net/sales-data/'
CREDENTIALS = (
    AZURE_SAS_TOKEN = 'sp=racwll&st=2025-10-22T10:47:09Z&se=2025-10-23T19:02:09Z&spr=https&sv=2024-11-04&sr=c&sig=HEF7601nEZPK2IyYbuk9f2FvtAouA2f3K2boDvfc3fKQ5fLbX30'
)
FILE_FORMAT = csv_sales_format;
""").collect()

print("File format and stage created successfully")

File format and stage created successfully
```

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"	"PROFIT"	"FILENAME"	"LOADTIMESTAMP"
ORD-SFBD0F0C	2024-10-08	2024-10	CUST1000	Ananya Sharma	India	South	Mumbai	Office Supplies	Paper	9	0.00	2700.00	78	Retail_Sales_500_rows_Preview.csv	2025-10-22 04:29:36.850000
ORD-BF0078E4	2024-05-11	2024-05	CUST1001	Aarav Iyer	India	Central	Lucknow	Technology	Networking	4	0.15	27200.00	41	Retail_Sales_500_rows_Preview.csv	2025-10-22 04:29:36.850000
ORD-B6C058A3	2024-06-12	2024-06	CUST1002	Arjun Sharma	USA	East	Kolkata	Furniture	Tables	4	0.10	31500.00	56	Retail_Sales_500_rows_Preview.csv	2025-10-22 04:29:36.850000
ORD-FBDC02D9	2024-12-18	2024-12	CUST1003	Ananya Das	India	North	Kolkata	Office Supplies	Appliances	9	0.00	36000.00	11	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	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")

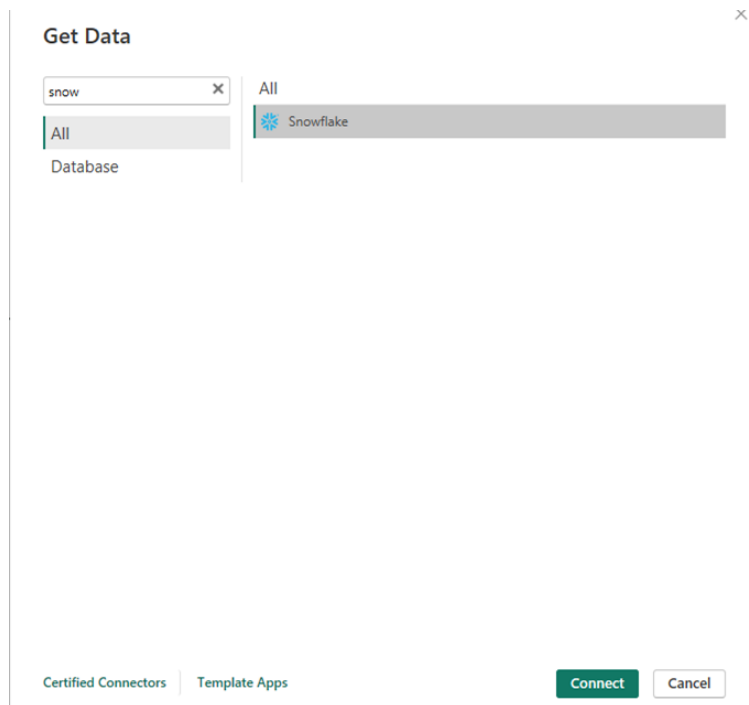
Snowpark session closed
Pipeline execution completed successfully!

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

5. Load the relevant tables or views.

6. Build visualizations like:

- Sum of SALES by CATEGORY
- Sum of SALES by REGION
- Sum of SALES by REGION and CATEGORY

1.07M

Sum of SALES

137.00

Sum of QUANTITY

180.65K

Sum of PROFIT

1.65

Sum of DISCOUNT

