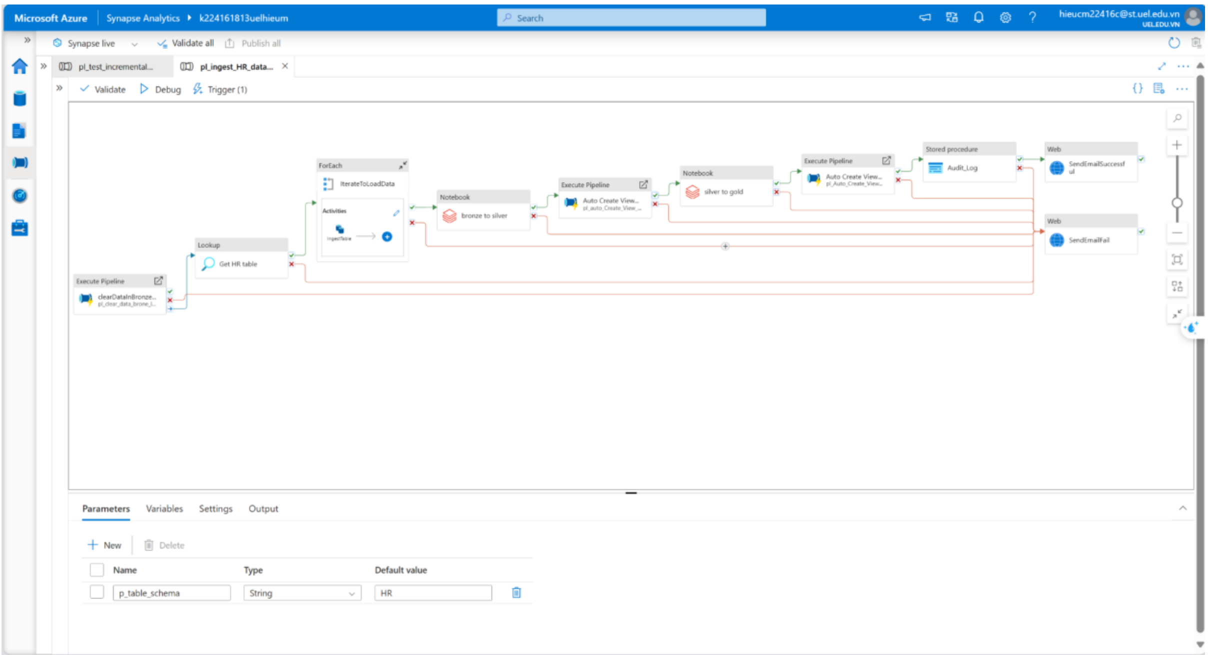
Lakehouse Data Pipeline Architecture

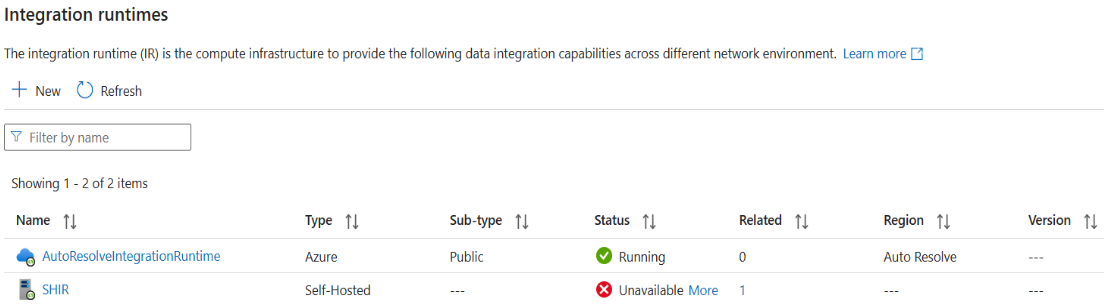
1. Data Pipeline Components



Name	Type	Default value
p_table_schema	String	HR

2. Ingestion Pipeline: Source to Bronze Layer

2.1 Configure Self-Hosted Integration Runtime


Integration runtimes

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

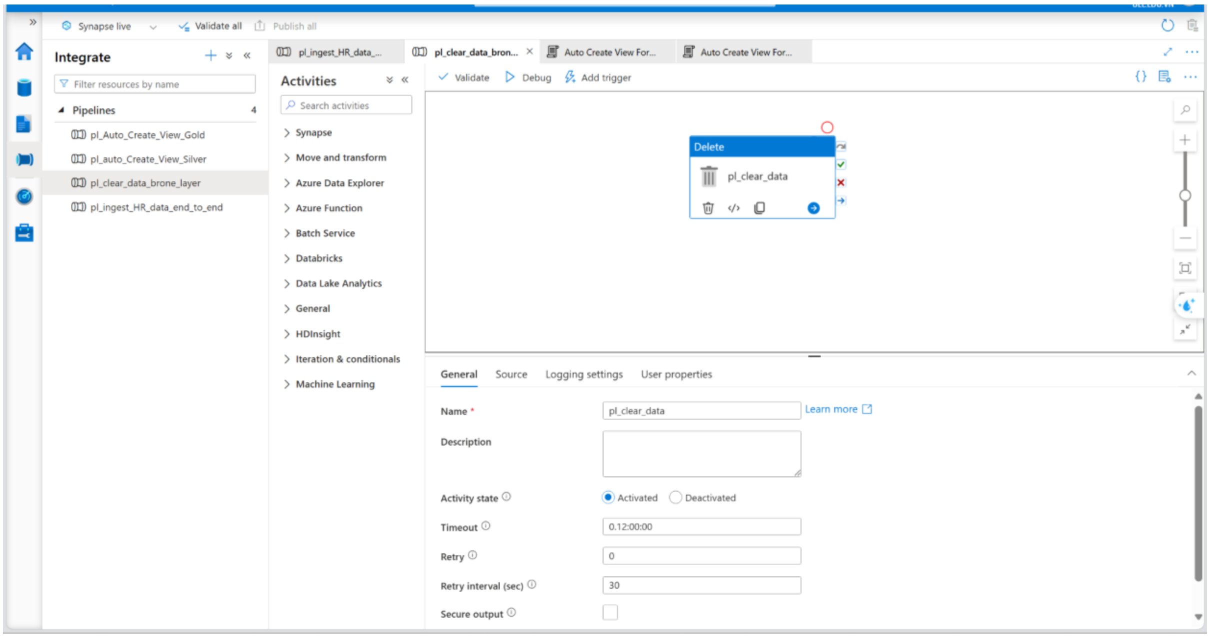
Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	✓ Running	0	Auto Resolve	---
SHIR	Self-Hosted	---	✗ Unavailable More	1	---	---

2.2 Source To BronzeLayer (Raw Data Ingestion)

a. Raw Data Ingestion (Bronze Layer)

ADF pipeline is triggered to **truncate** data in the **bronze layer**.



Name * pl_clear_data

Description

Activity state ● Activated ○ Deactivated

Timeout 0.12:00:00

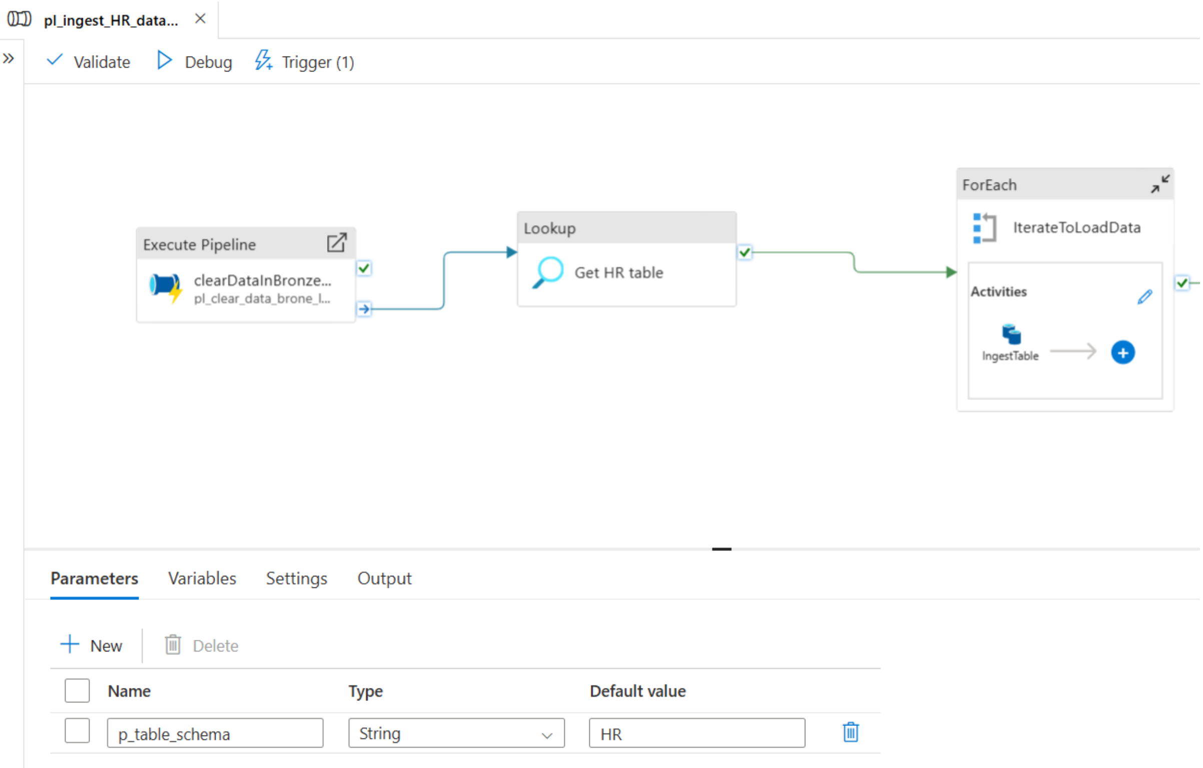
Retry 0

Retry interval (sec) 30

Secure output □

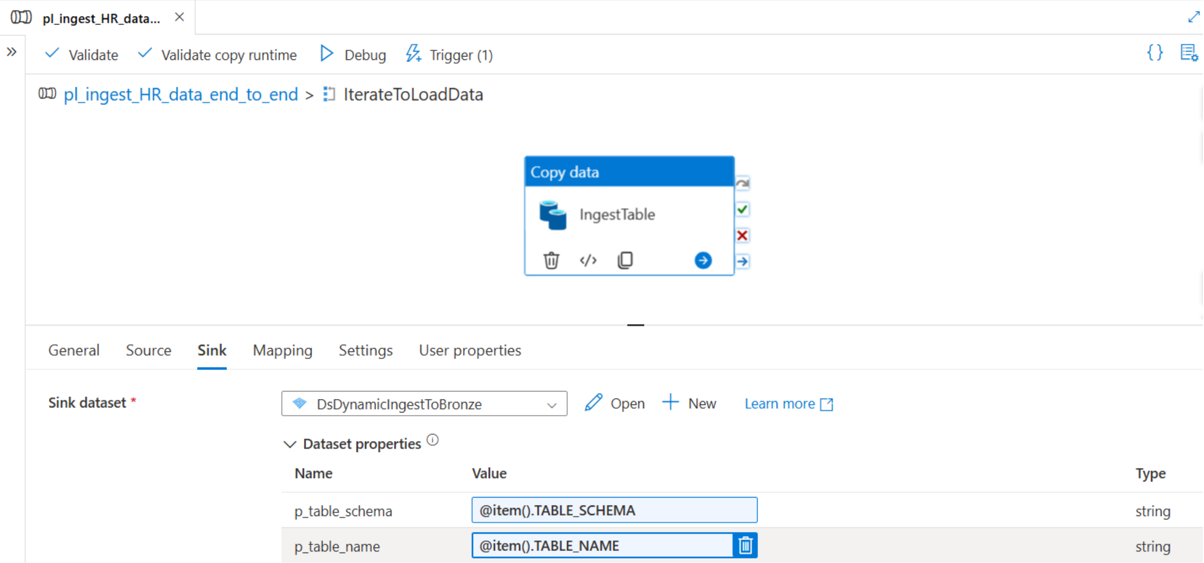
b. Full Load Implementation

ADF ingests data from on-prem SQL Server using SHIR.



□	Name	Type	Default value
□	p_table_schema	String	HR

Each table is stored in a corresponding folder in ADLS Gen2 under the Bronze Layer.

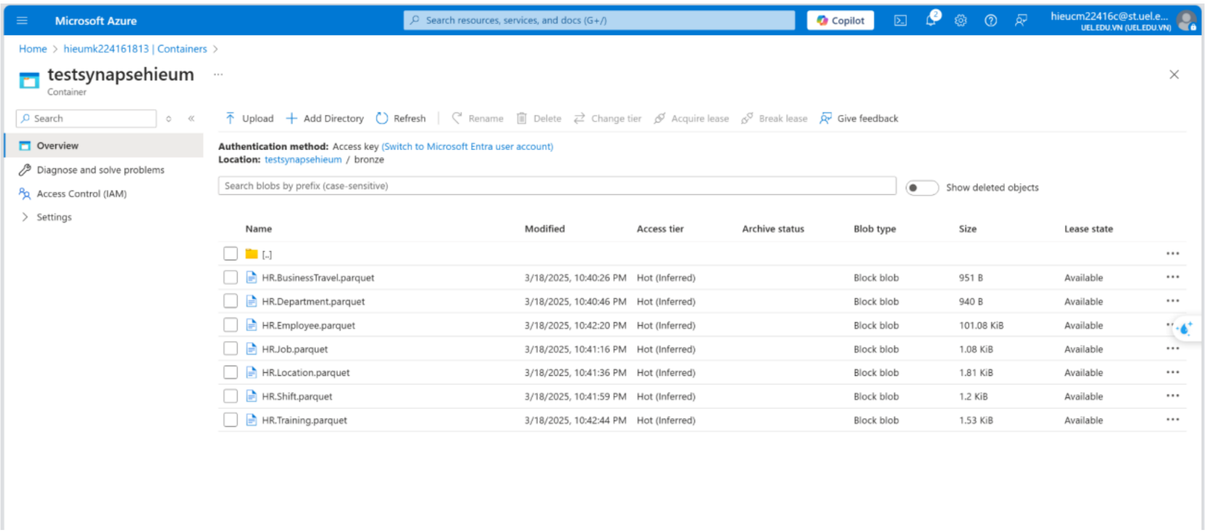


Sink dataset * DsDynamicIngestToBronze

Dataset properties

Name	Value	Type
p_table_schema	@item().TABLE_SCHEMA	string
p_table_name	@item().TABLE_NAME	string

Data have been loaded to Bronze Layer.



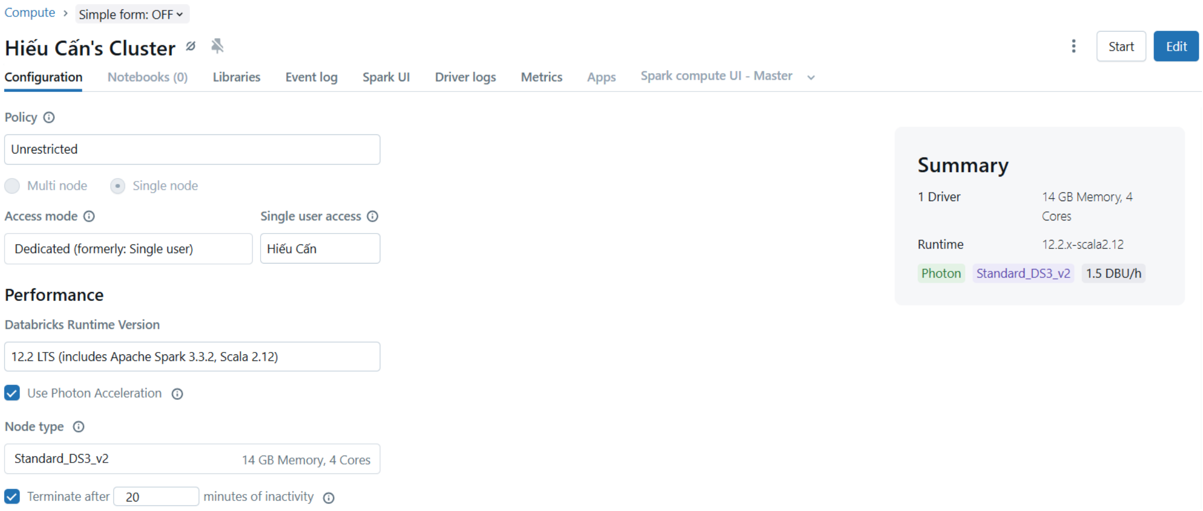
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: testsynapsehieum / bronze

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
HR.BusinessTravel.parquet	3/18/2025, 10:40:26 PM	Hot (Inferred)		Block blob	951 B	Available
HR.Department.parquet	3/18/2025, 10:40:46 PM	Hot (Inferred)		Block blob	940 B	Available
HR.Employee.parquet	3/18/2025, 10:42:20 PM	Hot (Inferred)		Block blob	101.08 KiB	Available
HR.Job.parquet	3/18/2025, 10:41:16 PM	Hot (Inferred)		Block blob	1.08 KiB	Available
HR.Location.parquet	3/18/2025, 10:41:36 PM	Hot (Inferred)		Block blob	1.81 KiB	Available
HR.Shift.parquet	3/18/2025, 10:41:59 PM	Hot (Inferred)		Block blob	1.2 KiB	Available
HR.Training.parquet	3/18/2025, 10:42:44 PM	Hot (Inferred)		Block blob	1.53 KiB	Available

3. Transformation Pipeline: Bronze to Silver Layer

3.1 Compute Initialization

Start a Databricks cluster to run transformation workloads.



Hiếu Cấn's Cluster

Policy: Unrestricted

○ Multi node ● Single node

Access mode: Dedicated (formerly: Single user)

Single user access: Hiếu Cấn

Performance

Databricks Runtime Version: 12.2 LTS (includes Apache Spark 3.3.2, Scala 2.12)

☑ Use Photon Acceleration

Node type: Standard_DS3_v2 14 GB Memory, 4 Cores

☑ Terminate after 20 minutes of inactivity

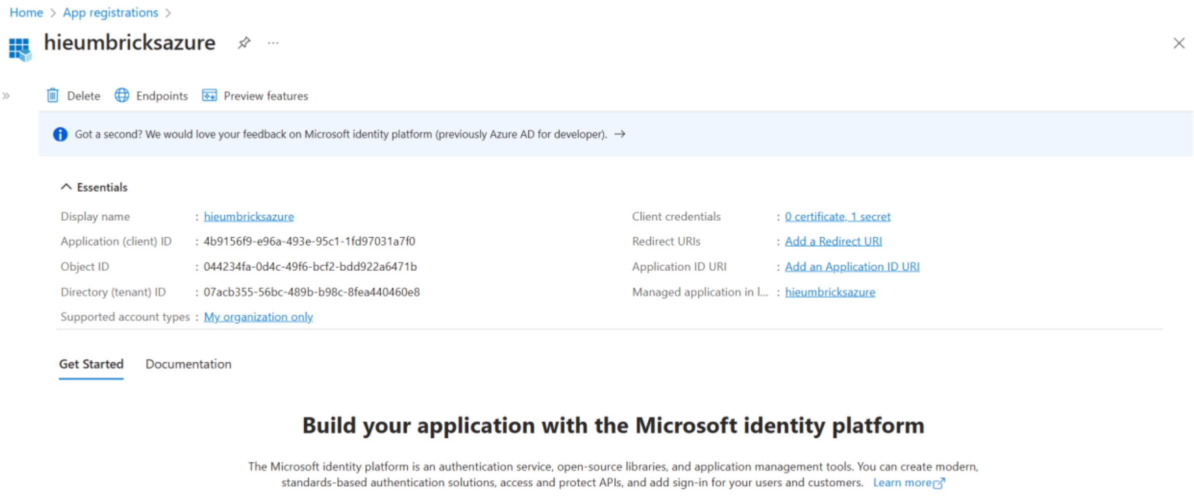
Summary

1 Driver	14 GB Memory, 4 Cores
Runtime	12.2.x-scala2.12

Photon Standard_DS3_v2 1.5 DBU/h

3.2 Configure Access to ADLS Gen2

Create App Registration To Get ObjectID, Tenant ID, Secret ID



hieumbricksazure

Got a second? We would love your feedback on Microsoft identity platform (previously Azure AD for developer). →

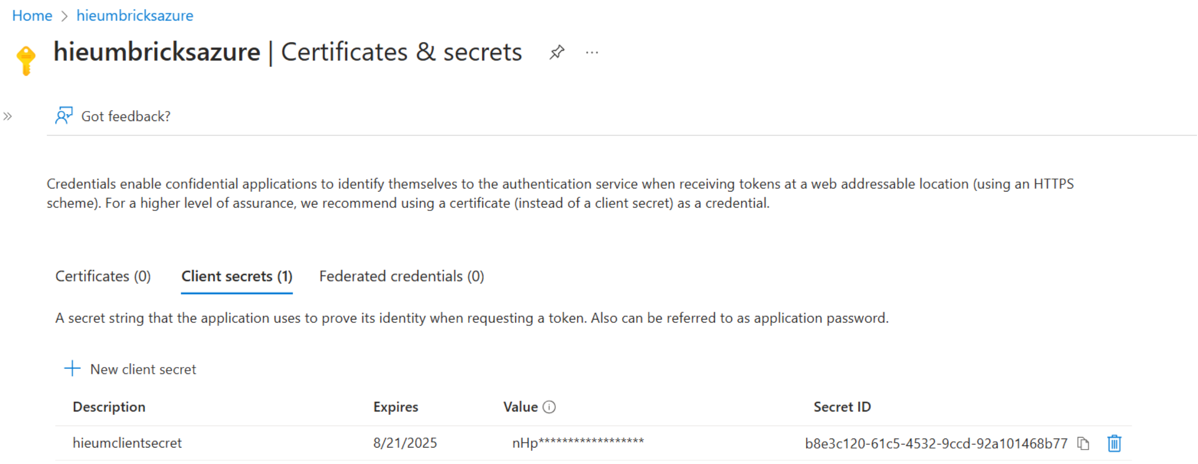
Essentials

Display name	: hieumbricksazure	Client credentials	: 0 certificate, 1 secret
Application (client) ID	: 4b9156f9-e96a-493e-95c1-1fd97031a7f0	Redirect URIs	: Add a Redirect URI
Object ID	: 044234fa-0d4c-49f6-bcf2-bdd922a6471b	Application ID URI	: Add an Application ID URI
Directory (tenant) ID	: 07acb355-56bc-489b-b98c-8fea440460e8	Managed application in l...	: hieumbricksazure
Supported account types	: My organization only		

Build your application with the Microsoft identity platform

The Microsoft identity platform is an authentication service, open-source libraries, and application management tools. You can create modern, standards-based authentication solutions, access and protect APIs, and add sign-in for your users and customers. Learn more

Store credentials securely in Azure Key Vault.



hieumbricksazure | Certificates & secrets

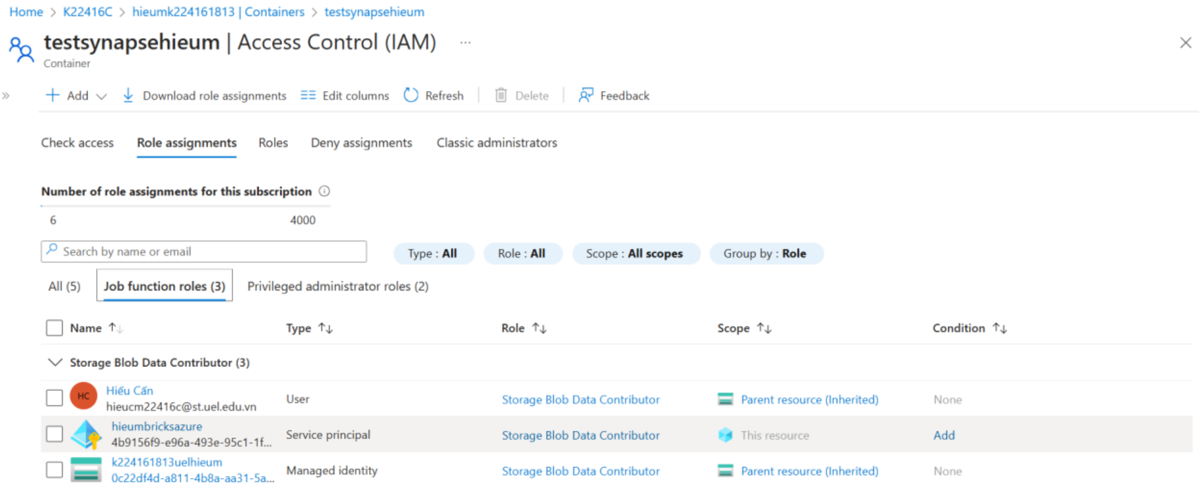
Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	Secret ID
hieumclientsecret	8/21/2025	nHp*****************	b8e3c120-61c5-4532-9ccd-92a101468b77

3.3 Assign Access Permissions

Provide RBAC permissions for the Databricks App to access ADLS Gen2 (Lakehouse Storage)



Home > K22416C > hieumk224161813 | Containers > testsynapsehieum

testsynapsehieum | Access Control (IAM)

Container

Number of role assignments for this subscription

6 4000

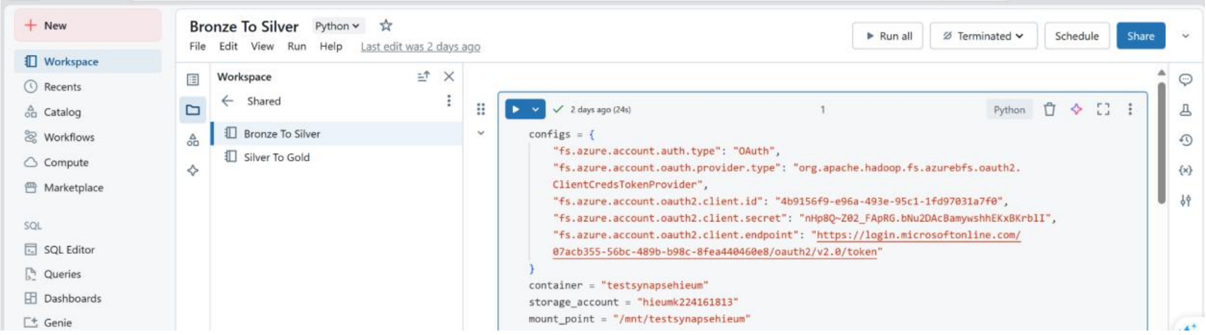
Type : **All** Role : **All** Scope : **All scopes** Group by : **Role**

All (5) **Job function roles (3)** Privileged administrator roles (2)

□ Name	Type	Role	Scope	Condition
Storage Blob Data Contributor (3)				
□ Hiếu Cấn hieucm22416c@st.uel.edu.vn	User	Storage Blob Data Contributor	Parent resource (Inherited)	None
□ hieumbricksazure 4b9156f9-e96a-493e-95c1-1f...	Service principal	Storage Blob Data Contributor	This resource	Add
□ k224161813uelhieum 0c22df4d-a811-4b8a-aa31-5a...	Managed identity	Storage Blob Data Contributor	Parent resource (Inherited)	None

3.4 OAuth2 Authentication

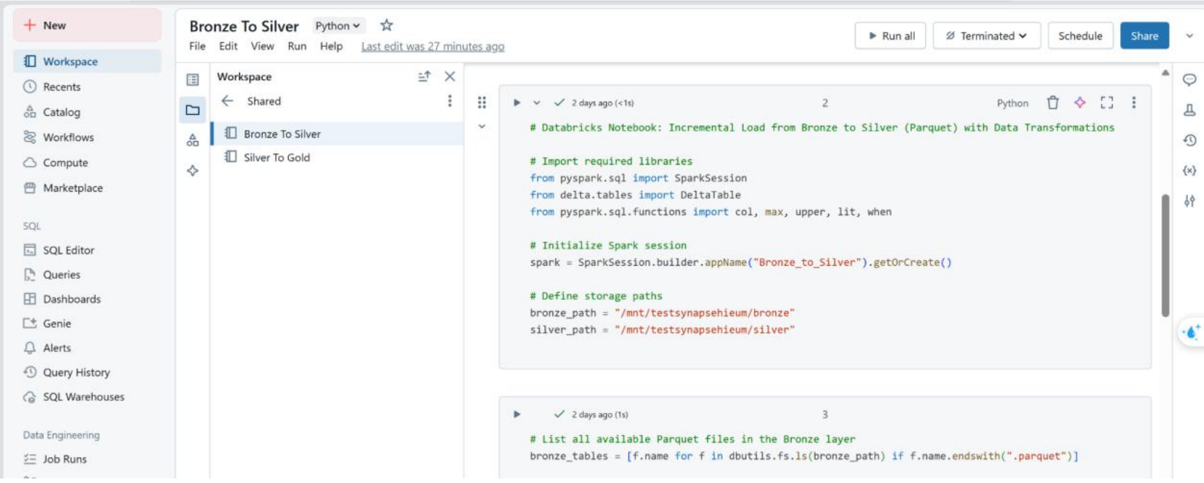
Use OAuth 2.0 flow to obtain Access Token for Databricks to securely access Lakehouse data



```python
configs = {
    "fs.azure.account.auth.type": "OAuth",
    "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.
    ClientCredsTokenProvider",
    "fs.azure.account.oauth2.client.id": "4b9156f9-e96a-493e-95c1-1fd97031a7f0",
    "fs.azure.account.oauth2.client.secret": "nHp8Q~Z02_FApRG.bNu2DAcBamywshhEKxBKrblI",
    "fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/
    07acb355-56bc-489b-b98c-8fea440460e8/oauth2/v2.0/token"
}
container = "testsynapsehieum"
storage_account = "hieumk224161813"
mount_point = "/mnt/testsynapsehieum"
```

3.5 Read from Bronze Layer

Databricks reads raw data from the Bronze Layer

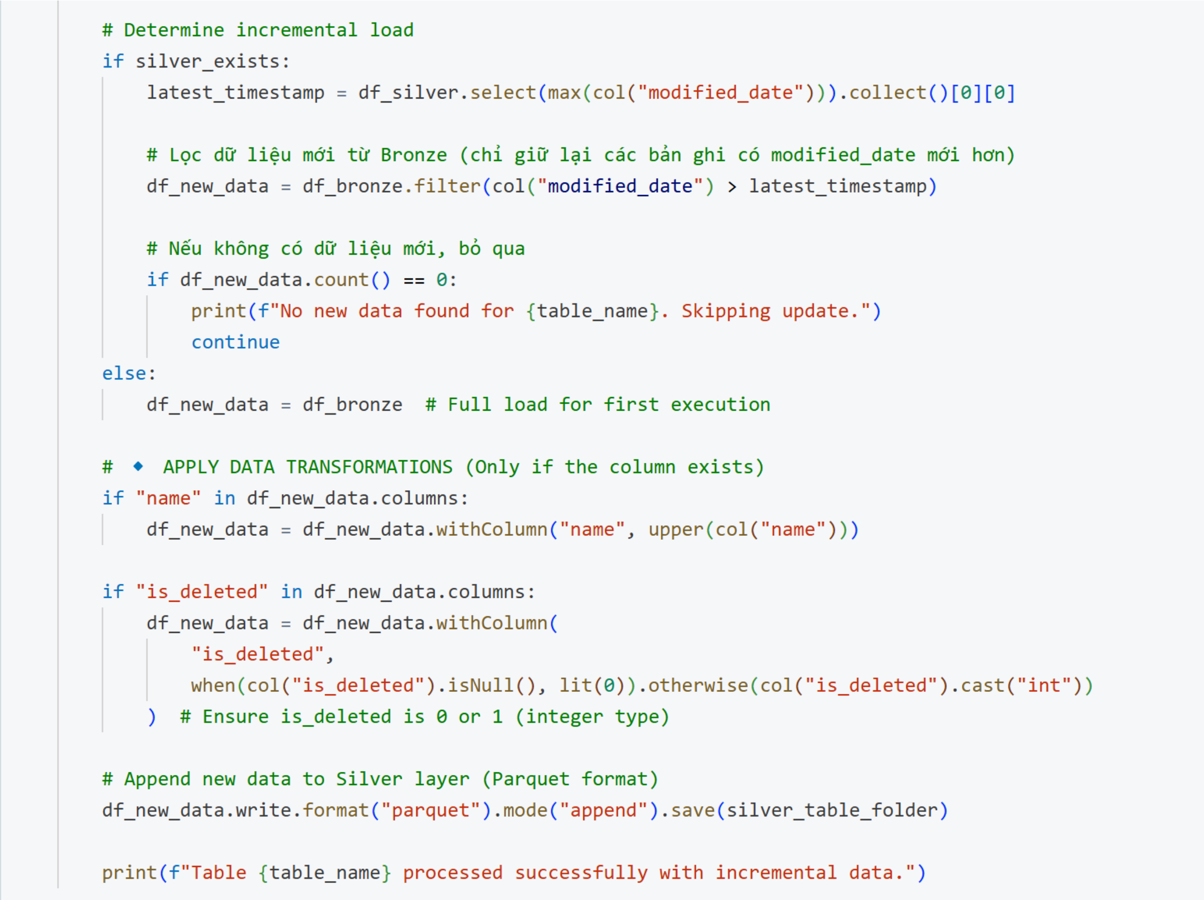


```python
# Databricks Notebook: Incremental Load from Bronze to Silver (Parquet) with Data Transformations

# Import required libraries
from pyspark.sql import SparkSession
from delta.tables import DeltaTable
from pyspark.sql.functions import col, max, upper, lit, when

# Initialize Spark session
spark = SparkSession.builder.appName("Bronze_to_Silver").getOrCreate()

# Define storage paths
bronze_path = "/mnt/testsynapsehieum/bronze"
silver_path = "/mnt/testsynapsehieum/silver"
```

```python
# List all available Parquet files in the Bronze layer
bronze_tables = [f.name for f in dbutils.fs.ls(bronze_path) if f.name.endswith(".parquet")]
```

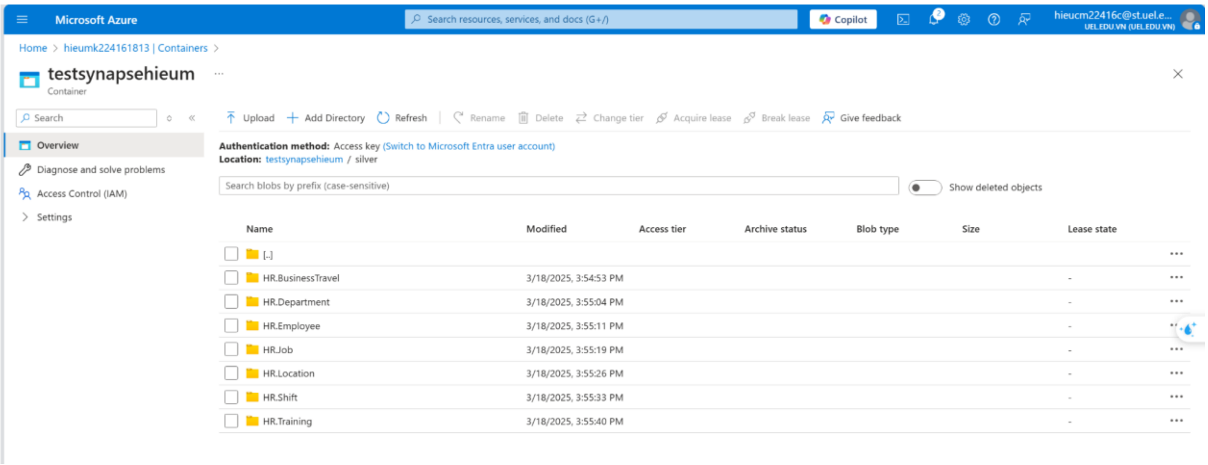
3.6 Perform Incremental Load + Transformations

Apply incremental loading techniques, apply necessary cleansing and transformations, processed data to Silver Layer in Parquet format.



```python
# Determine incremental load
if silver_exists:
    latest_timestamp = df_silver.select(max(col("modified_date"))).collect()[0][0]

    # Lọc dữ liệu mới từ Bronze (chỉ giữ lại các bản ghi có modified_date mới hơn)
    df_new_data = df_bronze.filter(col("modified_date") > latest_timestamp)

    # Nếu không có dữ liệu mới, bỏ qua
    if df_new_data.count() == 0:
        print(f"No new data found for {table_name}. Skipping update.")
        continue
else:
    df_new_data = df_bronze  # Full load for first execution

# ◆  APPLY DATA TRANSFORMATIONS (Only if the column exists)
if "name" in df_new_data.columns:
    df_new_data = df_new_data.withColumn("name", upper(col("name")))

if "is_deleted" in df_new_data.columns:
    df_new_data = df_new_data.withColumn(
        "is_deleted",
        when(col("is_deleted").isNull(), lit(0)).otherwise(col("is_deleted").cast("int"))
    )  # Ensure is_deleted is 0 or 1 (integer type)

# Append new data to Silver layer (Parquet format)
df_new_data.write.format("parquet").mode("append").save(silver_table_folder)

print(f"Table {table_name} processed successfully with incremental data.")
```

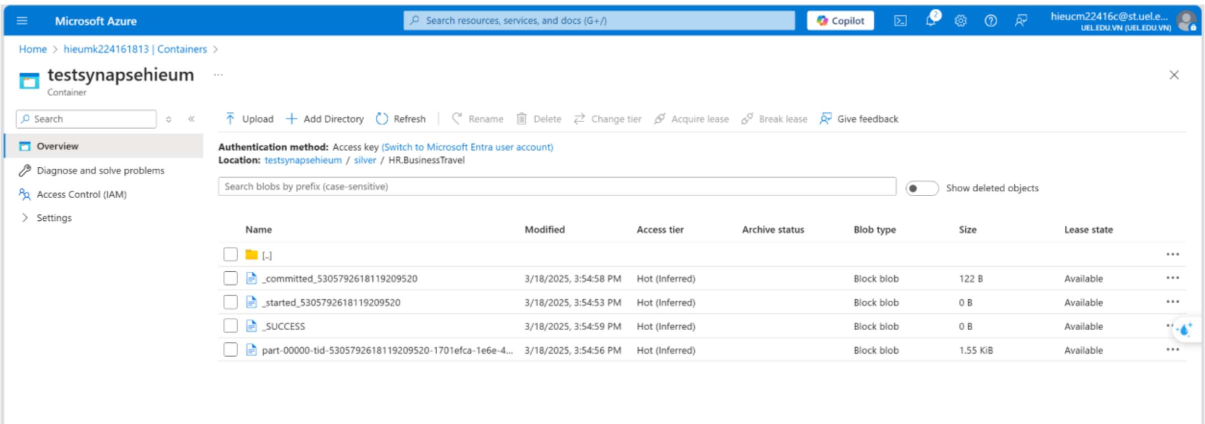
Data have been loaded to Silver Layer and stored as parquet file.



testsynapsehieum

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: testsynapsehieum / silver

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
HR.BusinessTravel	3/18/2025, 3:54:53 PM					-
HR.Department	3/18/2025, 3:55:04 PM					-
HR.Employee	3/18/2025, 3:55:11 PM					-
HR.Job	3/18/2025, 3:55:19 PM					-
HR.Location	3/18/2025, 3:55:26 PM					-
HR.Shift	3/18/2025, 3:55:33 PM					-
HR.Training	3/18/2025, 3:55:40 PM					-

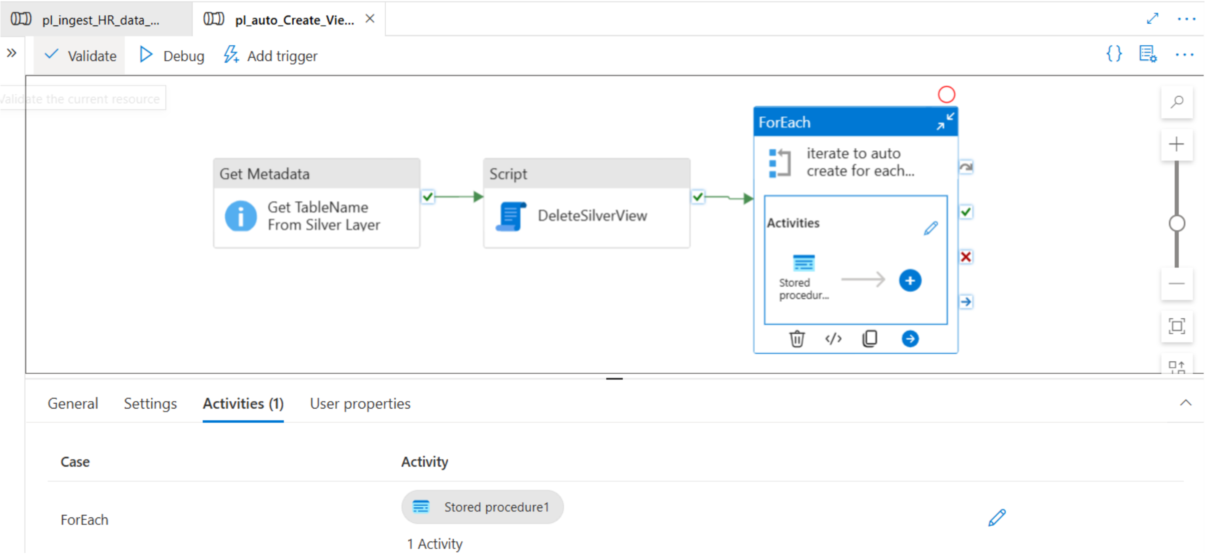


testsynapsehieum

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: testsynapsehieum / silver / HR.BusinessTravel

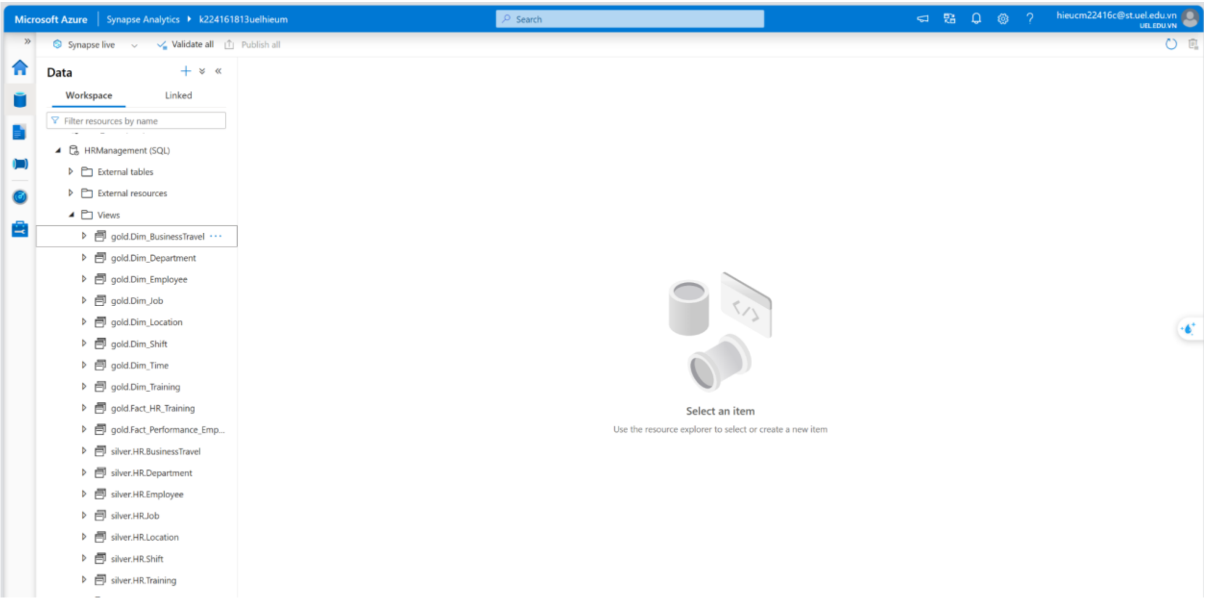
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
_committed_5305792618119209520	3/18/2025, 3:54:58 PM	Hot (Inferred)		Block blob	122 B	Available
_started_5305792618119209520	3/18/2025, 3:54:53 PM	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	3/18/2025, 3:54:59 PM	Hot (Inferred)		Block blob	0 B	Available
part-00000-tid-5305792618119209520-1701efca-1e6e-4...	3/18/2025, 3:54:56 PM	Hot (Inferred)		Block blob	1.55 KiB	Available

3.7 Auto create view cho Silver SQLServerLess(Synapse)



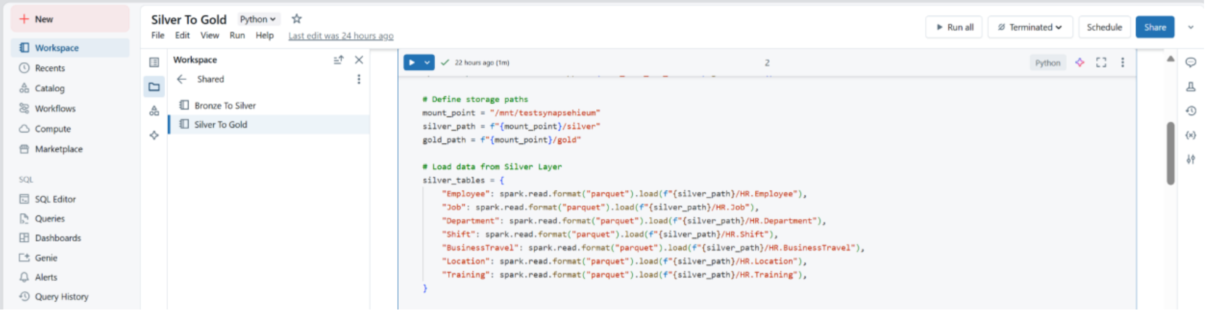
Case	Activity
ForEach	Stored procedure1 1 Activity

Automatically register views for Silver Layer parquet files in Synapse Analytics for querying via SQL



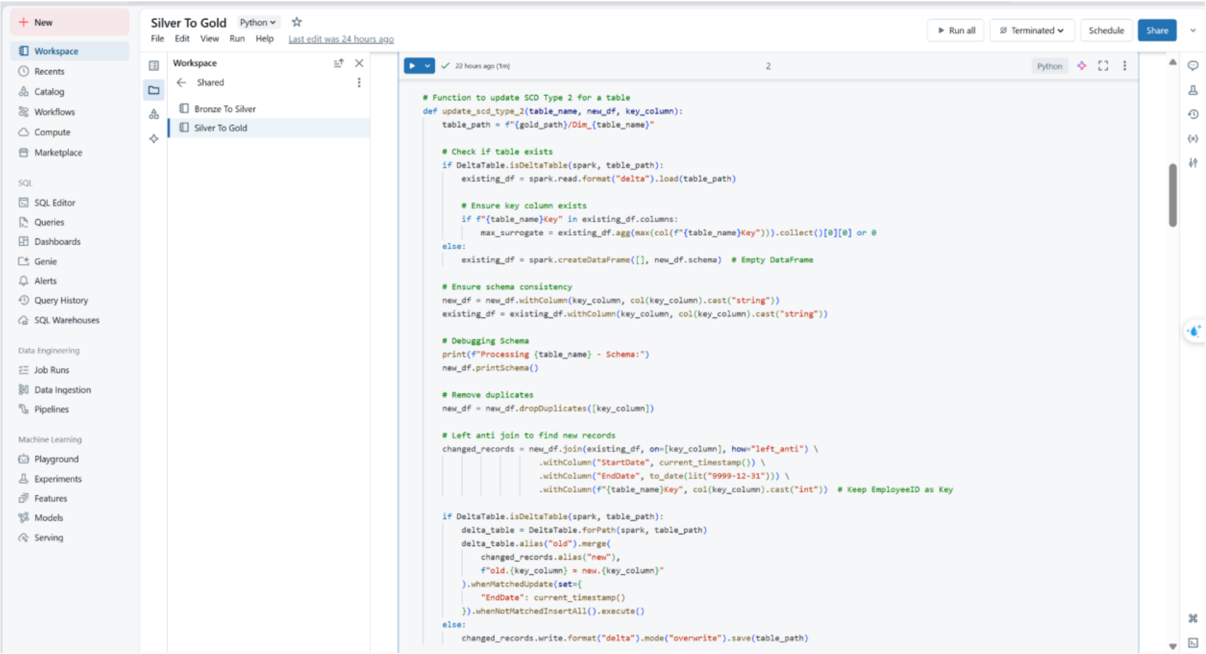
4. Business Logic Layer: Silver to Gold Layer

4.1 Read Data from Silver



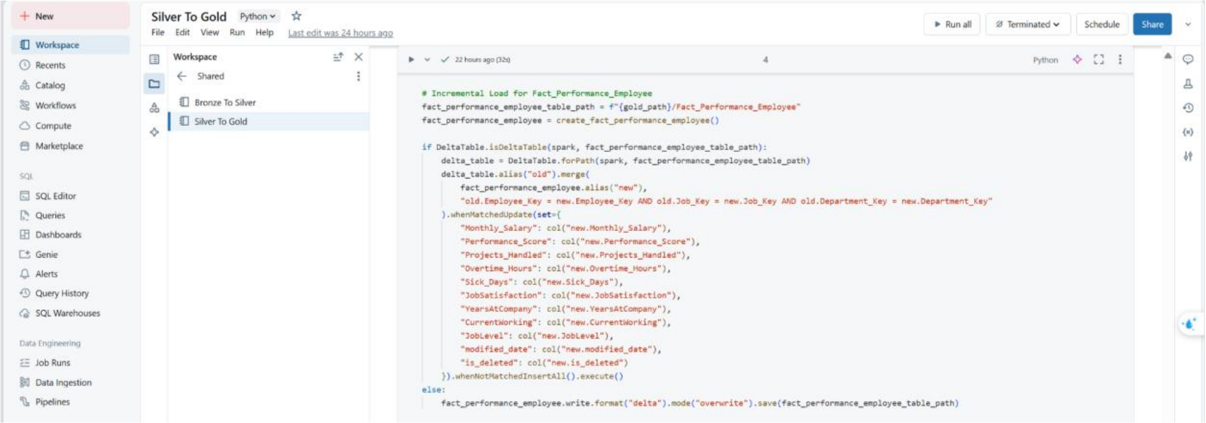
```python
# Define storage paths
mount_point = "/mnt/testsynapsehieum"
silver_path = f"{mount_point}/silver"
gold_path = f"{mount_point}/gold"

# Load data from Silver Layer
silver_tables = {
    "Employee": spark.read.format("parquet").load(f"{silver_path}/HR.Employee"),
    "Job": spark.read.format("parquet").load(f"{silver_path}/HR.Job"),
    "Department": spark.read.format("parquet").load(f"{silver_path}/HR.Department"),
    "Shift": spark.read.format("parquet").load(f"{silver_path}/HR.Shift"),
    "BusinessTravel": spark.read.format("parquet").load(f"{silver_path}/HR.BusinessTravel"),
    "Location": spark.read.format("parquet").load(f"{silver_path}/HR.Location"),
    "Training": spark.read.format("parquet").load(f"{silver_path}/HR.Training"),
}
```

4.2 Apply Slowly Changing for Dimension table



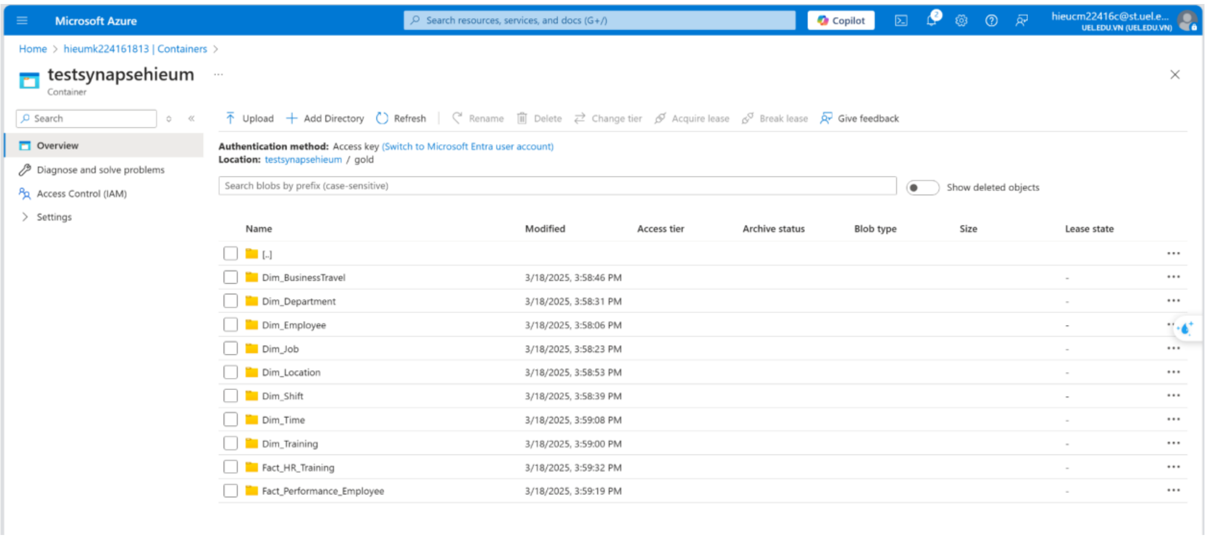
```python
# Function to update SCD Type 2 for a table
def update_scd_type_2(table_name, new_df, key_column):
    table_path = f"{gold_path}/Dim_{table_name}"

    # Check if table exists
    if DeltaTable.isDeltaTable(spark, table_path):
        existing_df = spark.read.format("delta").load(table_path)

        # Ensure key column exists
        if f"{table_name}Key" in existing_df.columns:
            max_surrogate = existing_df.agg(max(col(f"{table_name}Key"))).collect()[0][0] or 0
    else:
        existing_df = spark.createDataFrame([], new_df.schema)  # Empty DataFrame

    # Ensure schema consistency
    new_df = new_df.withColumn(key_column, col(key_column).cast("string"))
    existing_df = existing_df.withColumn(key_column, col(key_column).cast("string"))

    # Debugging Schema
    print(f"Processing {table_name} - Schema:")
    new_df.printSchema()

    # Remove duplicates
    new_df = new_df.dropDuplicates([key_column])

    # Left anti join to find new records
    changed_records = new_df.join(existing_df, on=[key_column], how="left_anti") \
                    .withColumn("StartDate", current_timestamp()) \
                    .withColumn("EndDate", to_date(lit("9999-12-31"))) \
                    .withColumn(f"{table_name}Key", col(key_column).cast("int"))  # Keep EmployeeID as Key

    if DeltaTable.isDeltaTable(spark, table_path):
        delta_table = DeltaTable.forPath(spark, table_path)
        delta_table.alias("old").merge(
            changed_records.alias("new"),
            f"old.{key_column} = new.{key_column}"
        ).whenMatchedUpdate(set={
            "EndDate": current_timestamp()
        }).whenNotMatchedInsertAll().execute()
    else:
        changed_records.write.format("delta").mode("overwrite").save(table_path)
```

4.3 Apply Incremental Load for Fact Tables



```python
# Incremental Load for Fact_Performance_Employee
fact_performance_employee_table_path = f"{gold_path}/Fact_Performance_Employee"
fact_performance_employee = create_fact_performance_employee()

if DeltaTable.isDeltaTable(spark, fact_performance_employee_table_path):
    delta_table = DeltaTable.forPath(spark, fact_performance_employee_table_path)
    delta_table.alias("old").merge(
        fact_performance_employee.alias("new"),
        "old.Employee_Key = new.Employee_Key AND old.Job_Key = new.Job_Key AND old.Department_Key = new.Department_Key"
    ).whenMatchedUpdate(set={
        "Monthly_Salary": col("new.Monthly_Salary"),
        "Performance_Score": col("new.Performance_Score"),
        "Projects_Handled": col("new.Projects_Handled"),
        "Overtime_Hours": col("new.Overtime_Hours"),
        "Sick_Days": col("new.Sick_Days"),
        "JobSatisfaction": col("new.JobSatisfaction"),
        "YearsAtCompany": col("new.YearsAtCompany"),
        "CurrentWorking": col("new.CurrentWorking"),
        "JobLevel": col("new.JobLevel"),
        "modified_date": col("new.modified_date"),
        "is_deleted": col("new.is_deleted")
    }).whenNotMatchedInsertAll().execute()
else:
    fact_performance_employee.write.format("delta").mode("overwrite").save(fact_performance_employee_table_path)
```

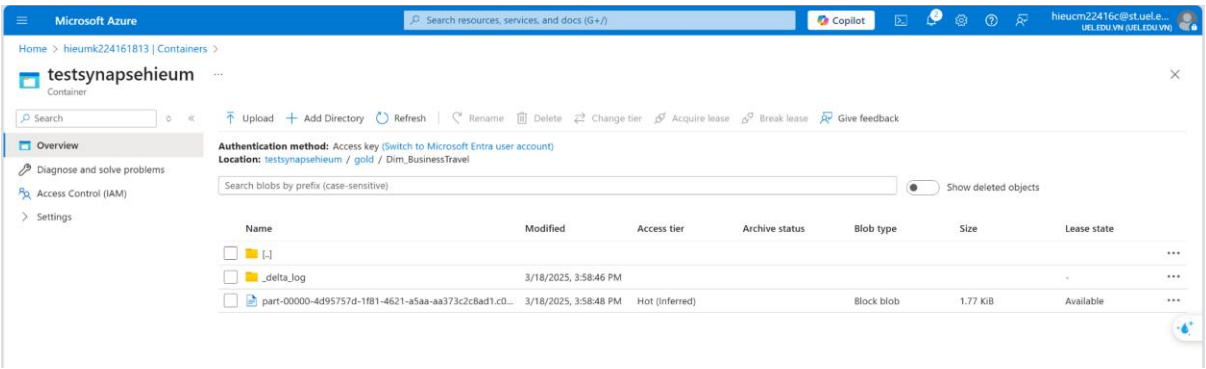
4.4 Data have been loaded to Gold Layer



Authentication method: Access key (Switch to Microsoft Entra user account)
Location: testsynapsehieum / gold

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
Dim_BusinessTravel	3/18/2025, 3:58:46 PM					-
Dim_Department	3/18/2025, 3:58:31 PM					-
Dim_Employee	3/18/2025, 3:58:06 PM					-
Dim_Job	3/18/2025, 3:58:23 PM					-
Dim_Location	3/18/2025, 3:58:53 PM					-
Dim_Shift	3/18/2025, 3:58:39 PM					-
Dim_Time	3/18/2025, 3:59:08 PM					-
Dim_Training	3/18/2025, 3:59:00 PM					-
Fact_HR_Training	3/18/2025, 3:59:32 PM					-
Fact_Performance_Employee	3/18/2025, 3:59:19 PM					-

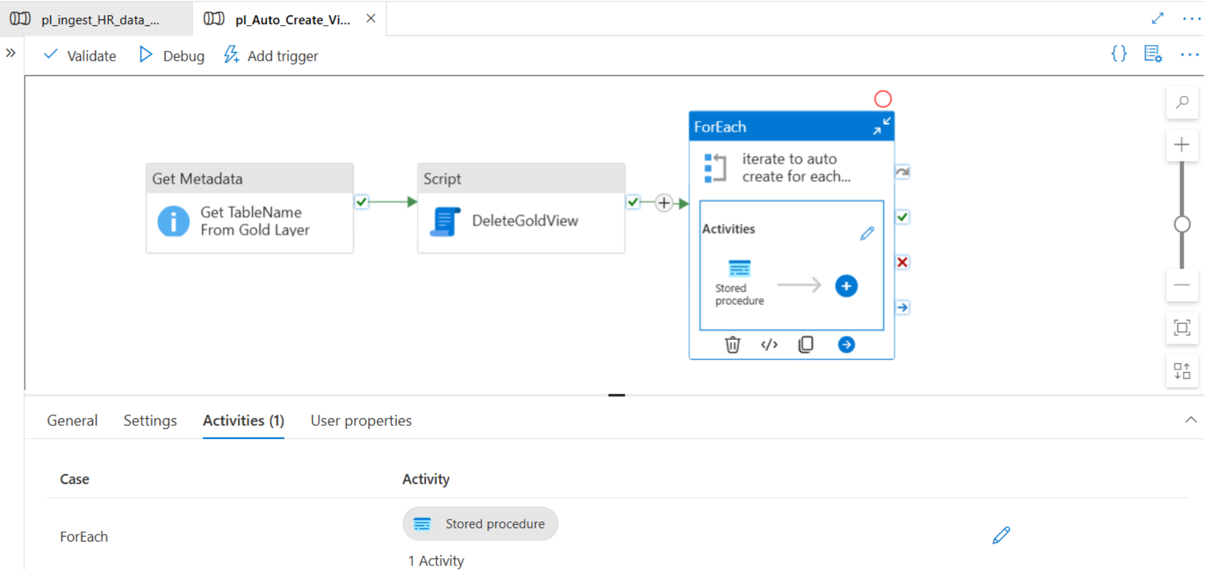
Data have been loaded to Gold Layer and stored as delta format file:



Authentication method: Access key (Switch to Microsoft Entra user account)
Location: testsynapsehieum / gold / Dim_BusinessTravel

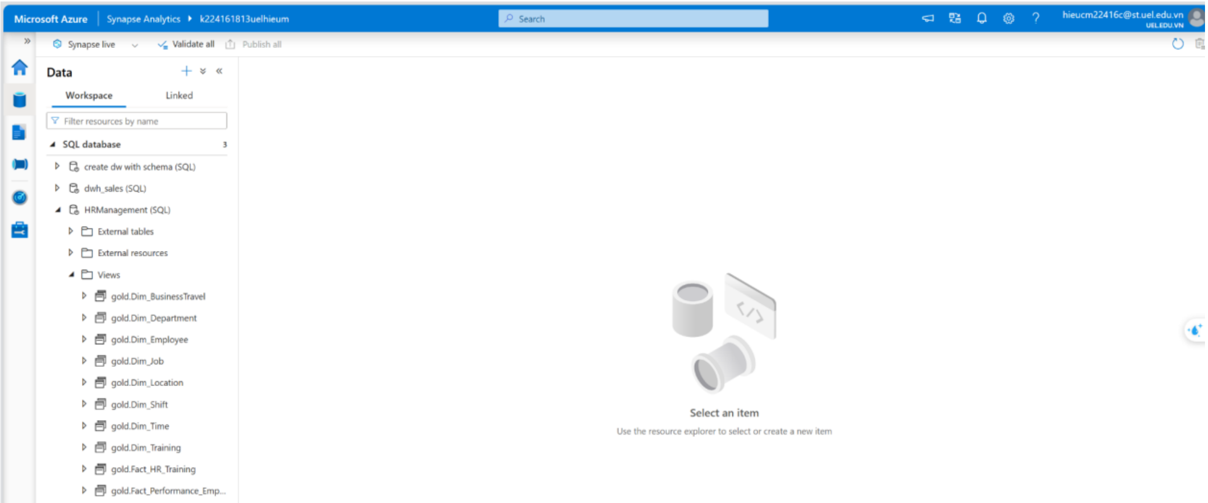
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
_delta_log	3/18/2025, 3:58:46 PM					-
part-00000-4d95757d-1f81-4621-a5aa-aa373c2c8ad1.c0...	3/18/2025, 3:58:48 PM	Hot (Inferred)		Block blob	1.77 KiB	Available

4.5 Auto-create Views in Synapse for Gold Layer



Case	Activity
ForEach	Stored procedure 1 Activity

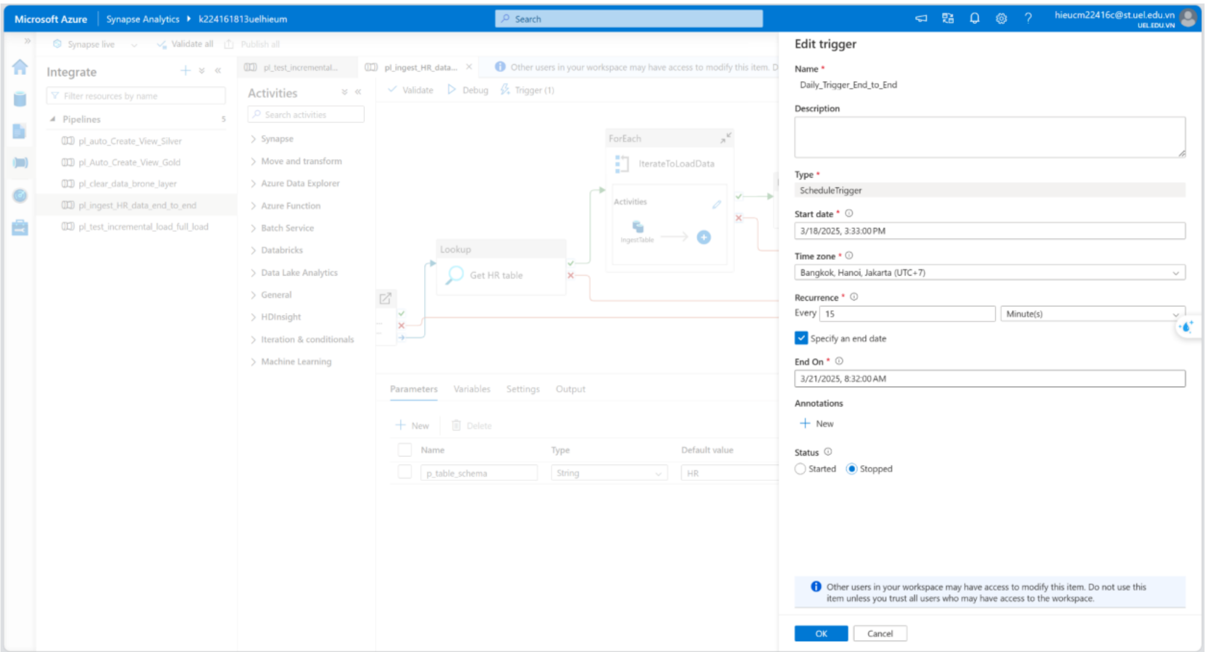
Automatically register views for Gold Layer delta files in Synapse Analytics for querying via SQL



- SQL database 3
  - create dw with schema (SQL)
  - dwh_sales (SQL)
  - HRManagement (SQL)
    - External tables
    - External resources
    - Views
      - gold.Dim_BusinessTravel
      - gold.Dim_Department
      - gold.Dim_Employee
      - gold.Dim_Job
      - gold.Dim_Location
      - gold.Dim_Shift
      - gold.Dim_Time
      - gold.Dim_Training
      - gold.Fact_HR_Training
      - gold.Fact_Performance_Emp...

5. Orchestration & Automation

5.1 Configure Triggers



Edit trigger

Name *
Daily_Trigger_End_to_End

Description

Type *
ScheduleTrigger

Start date *
3/18/2025, 3:33:00 PM

Time zone *
Bangkok, Hanoi, Jakarta (UTC+7)

Recurrence *
Every 15 Minute(s)

☑ Specify an end date

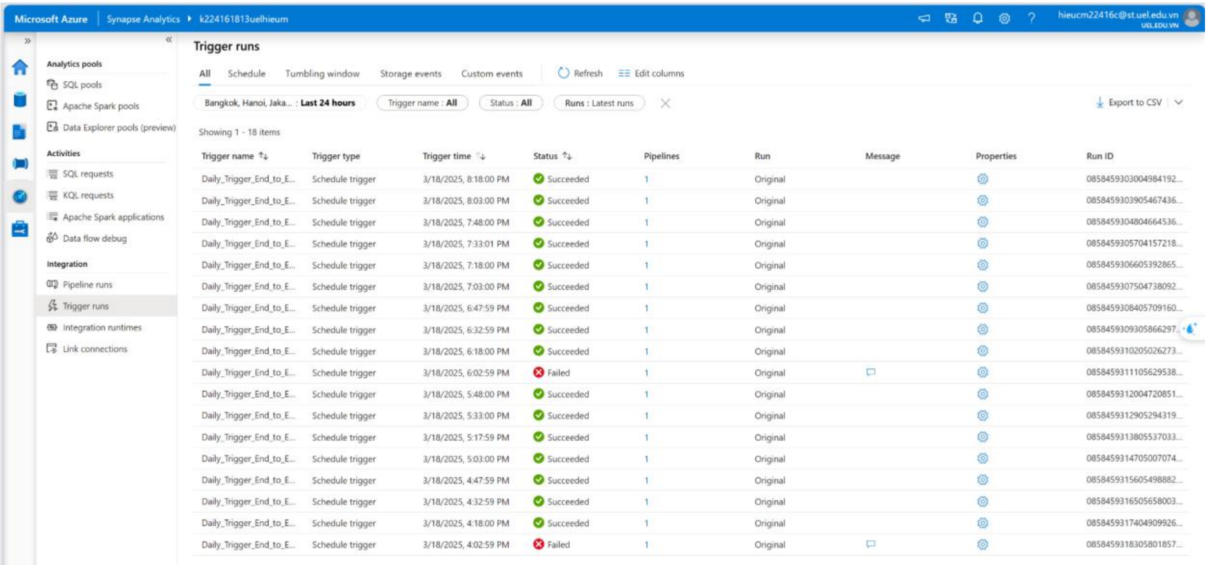
End On *
3/21/2025, 8:32:00 AM

Annotations
+ New

Status
○ Started ● Stopped

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

5.2 Trigger Result



Trigger runs

All | Schedule | Tumbling window | Storage events | Custom events | Refresh | Edit columns

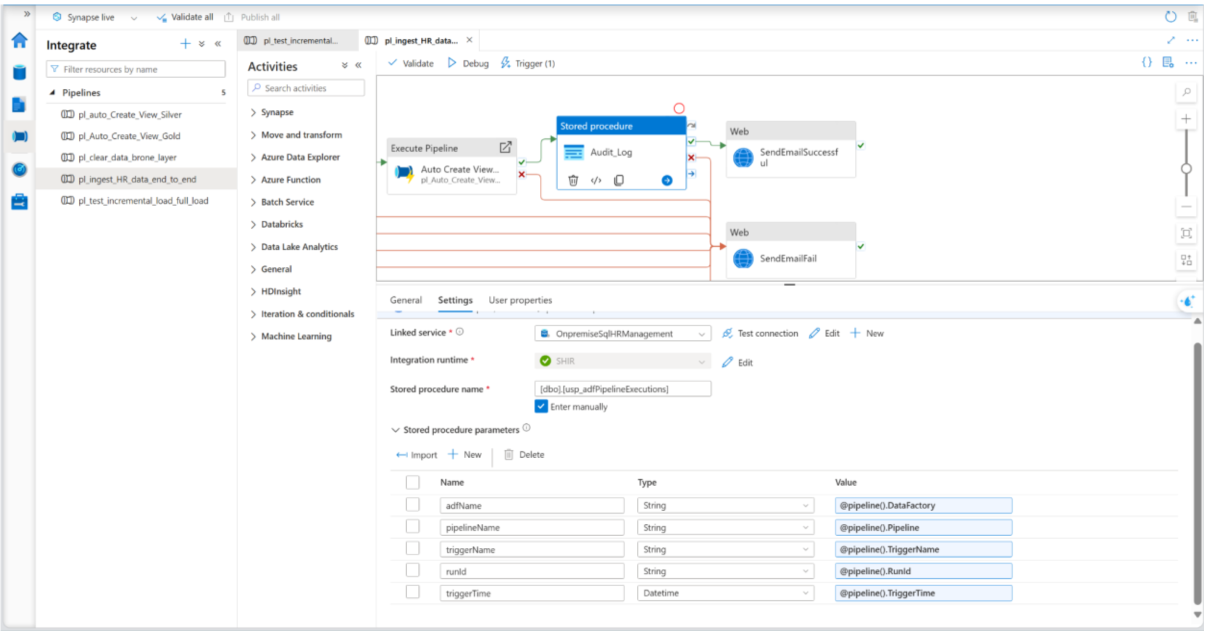
Bangkok, Hanoi, Jaka... : **Last 24 hours** | Trigger name : **All** | Status : **All** | Runs : Latest runs | Export to CSV

Showing 1 - 18 items

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run	Message	Properties	Run ID
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 8:18:00 PM	Succeeded	1	Original		⚙	0858459303004984192...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 8:03:00 PM	Succeeded	1	Original		⚙	0858459303905467436...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 7:48:00 PM	Succeeded	1	Original		⚙	0858459304804664536...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 7:33:01 PM	Succeeded	1	Original		⚙	0858459305704157218...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 7:18:00 PM	Succeeded	1	Original		⚙	0858459306605392865...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 7:03:00 PM	Succeeded	1	Original		⚙	0858459307504738092...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 6:47:59 PM	Succeeded	1	Original		⚙	0858459308405709160...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 6:32:59 PM	Succeeded	1	Original		⚙	0858459309305866297...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 6:18:00 PM	Succeeded	1	Original		⚙	0858459310205026273...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 6:02:59 PM	Failed	1	Original	🗨	⚙	0858459311105629538...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 5:48:00 PM	Succeeded	1	Original		⚙	0858459312004720851...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 5:33:00 PM	Succeeded	1	Original		⚙	0858459312905294319...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 5:17:59 PM	Succeeded	1	Original		⚙	0858459313805537033...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 5:03:00 PM	Succeeded	1	Original		⚙	0858459314705007074...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 4:47:59 PM	Succeeded	1	Original		⚙	0858459315605498882...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 4:32:59 PM	Succeeded	1	Original		⚙	0858459316505658003...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 4:18:00 PM	Succeeded	1	Original		⚙	0858459317404909926...
Daily_Trigger_End_to_E...	Schedule trigger	3/18/2025, 4:02:59 PM	Failed	1	Original	🗨	⚙	0858459318305801857...

6. Configure Audit Logging

6.1 Enable Audit Logs



General | Settings | User properties

Linked service *: OnpremiseSqlHRManagement | Test connection | Edit | New

Integration runtime *: SHIR | Edit

Stored procedure name *: [dbo].[usp_adfPipelineExecutions]

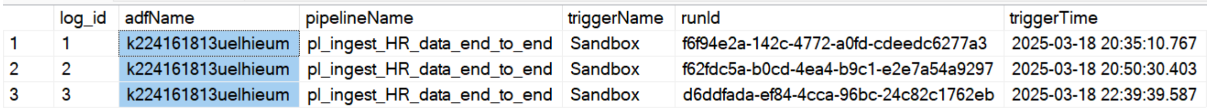
☑ Enter manually

Stored procedure parameters

Import | New | Delete

Name	Type	Value
adfName	String	@pipeline().DataFactory
pipelineName	String	@pipeline().Pipeline
triggerName	String	@pipeline().TriggerName
runId	String	@pipeline().RunId
triggerTime	Datetime	@pipeline().TriggerTime

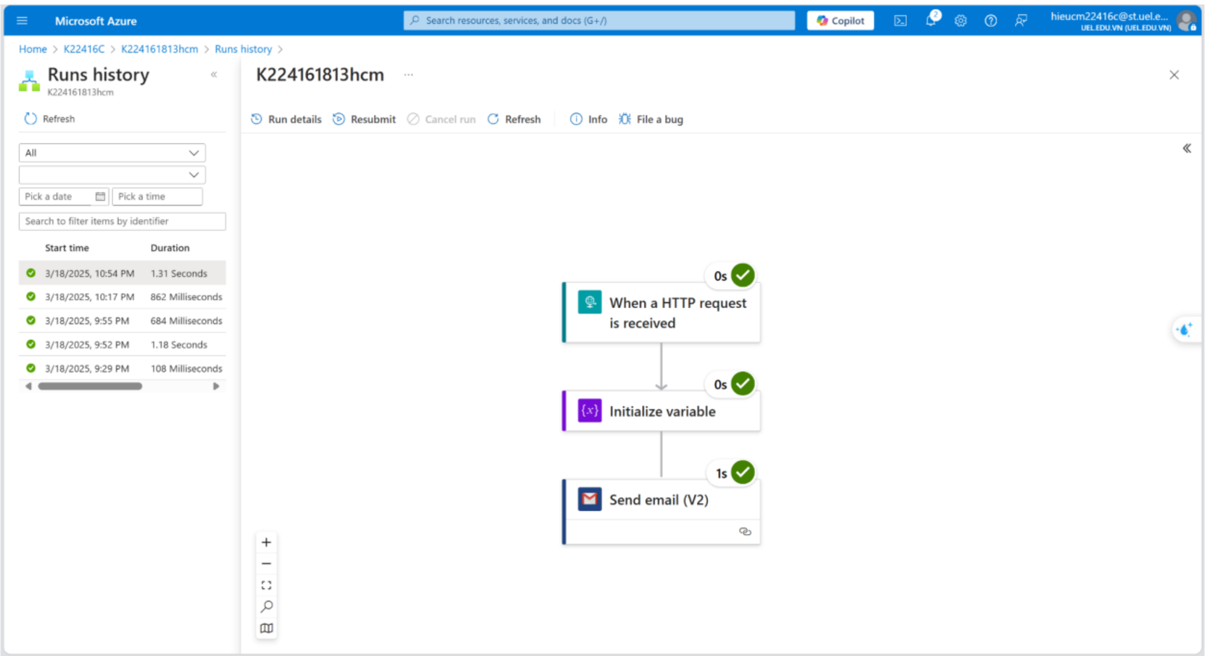
6.2 Results



	log_id	adfName	pipelineName	triggerName	runId	triggerTime
1	1	k224161813uelhieum	pl_ingest_HR_data_end_to_end	Sandbox	f6f94e2a-142c-4772-a0fd-cdeedc6277a3	2025-03-18 20:35:10.767
2	2	k224161813uelhieum	pl_ingest_HR_data_end_to_end	Sandbox	f62fdc5a-b0cd-4ea4-b9c1-e2e7a54a9297	2025-03-18 20:50:30.403
3	3	k224161813uelhieum	pl_ingest_HR_data_end_to_end	Sandbox	d6ddfada-ef84-4cca-96bc-24c82c1762eb	2025-03-18 22:39:39.587

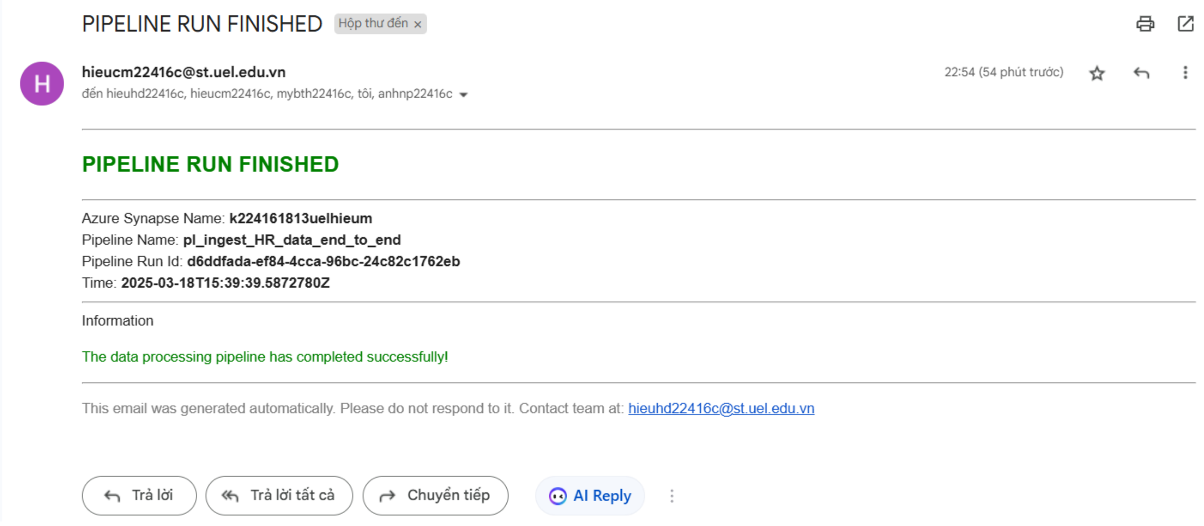
7. Configure Email Notifications

7.1 Configure Email Alerts



Start time	Duration
3/18/2025, 10:54 PM	1.31 Seconds
3/18/2025, 10:17 PM	862 Milliseconds
3/18/2025, 9:55 PM	684 Milliseconds
3/18/2025, 9:52 PM	1.18 Seconds
3/18/2025, 9:29 PM	108 Milliseconds

7.2 Result



PIPELINE RUN FINISHED Hộp thư đến ×

hieucm22416c@st.uel.edu.vn
đến hieuhd22416c, hieucm22416c, mybth22416c, tôi, anhnp22416c

22:54 (54 phút trước)

PIPELINE RUN FINISHED

Azure Synapse Name: **k224161813uelhieum**
Pipeline Name: **pl_ingest_HR_data_end_to_end**
Pipeline Run Id: **d6ddfada-ef84-4cca-96bc-24c82c1762eb**
Time: **2025-03-18T15:39:39.5872780Z**

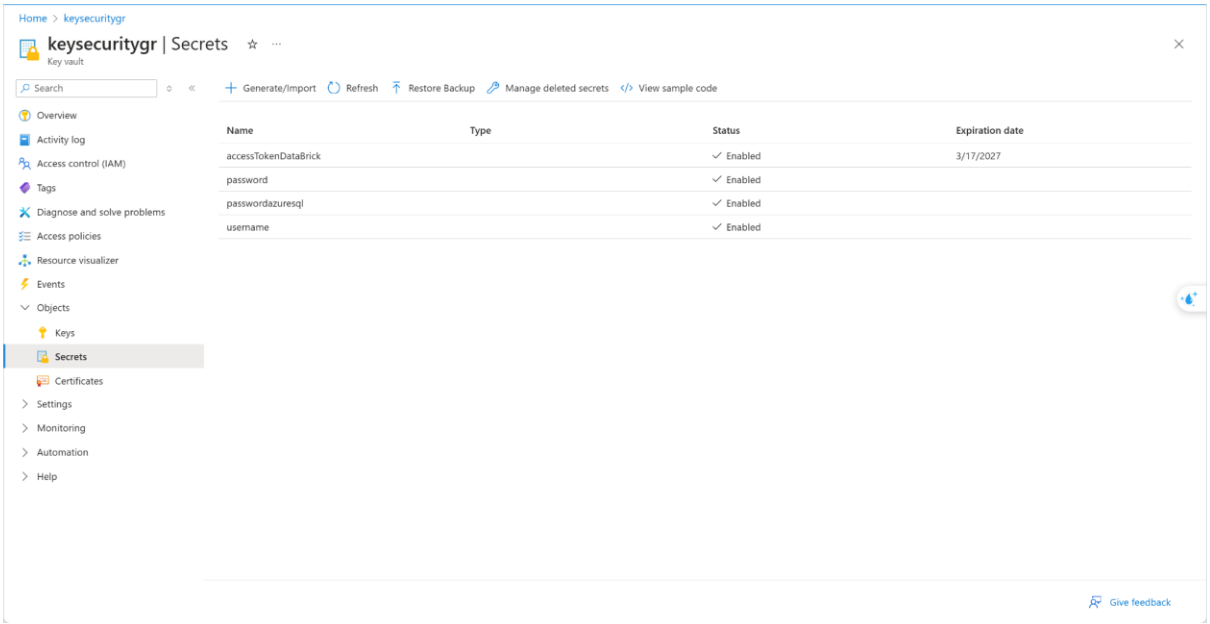
Information

The data processing pipeline has completed successfully!

This email was generated automatically. Please do not respond to it. Contact team at: hieuhd22416c@st.uel.edu.vn

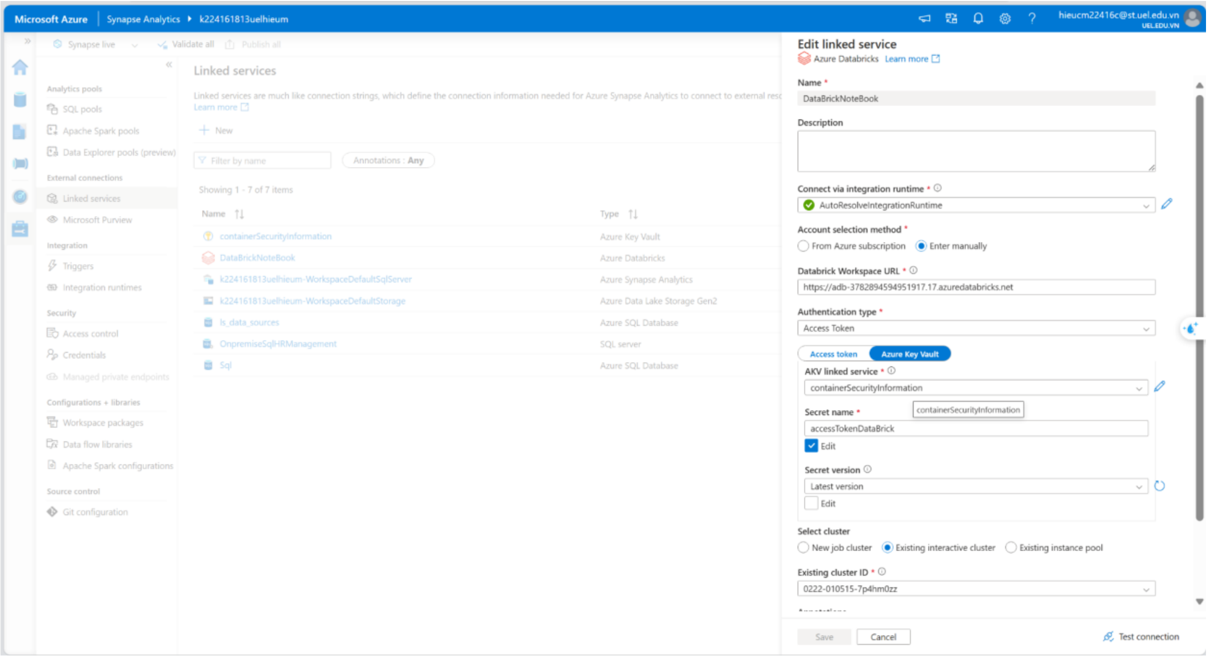
8. Azure Key Vault - Data Security

8.1 List of Encrypted Sensitive Information



Name	Type	Status	Expiration date
accessTokenDataBrick		✓ Enabled	3/17/2027
password		✓ Enabled	
passwordazuresql		✓ Enabled	
username		✓ Enabled	

8.2 Secret Management with Azure Key Vault



Linked services

Linked services are much like connection strings, which define the connection information needed for Azure Synapse Analytics to connect to external reso
Learn more

New

Filter by name   Annotations : Any

Showing 1 - 7 of 7 items

Name	Type
containerSecurityInformation	Azure Key Vault
DataBrickNoteBook	Azure Databricks
k224161813uelhieum-WorkspaceDefaultSqlServer	Azure Synapse Analytics
k224161813uelhieum-WorkspaceDefaultStorage	Azure Data Lake Storage Gen2
ls_data_sources	Azure SQL Database
OnpremiseSqlHRManagement	SQL server
Sql	Azure SQL Database

Edit linked service

Azure Databricks Learn more

Name *
DataBrickNoteBook

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Account selection method *
○ From Azure subscription ● Enter manually

Databrick Workspace URL *
https://adb-3782894594951917.17.azuredatabricks.net

Authentication type *
Access Token

Access token | Azure Key Vault

AKV linked service *
containerSecurityInformation

Secret name *
accessTokenDataBrick
☑ Edit

Secret version
Latest version
□ Edit

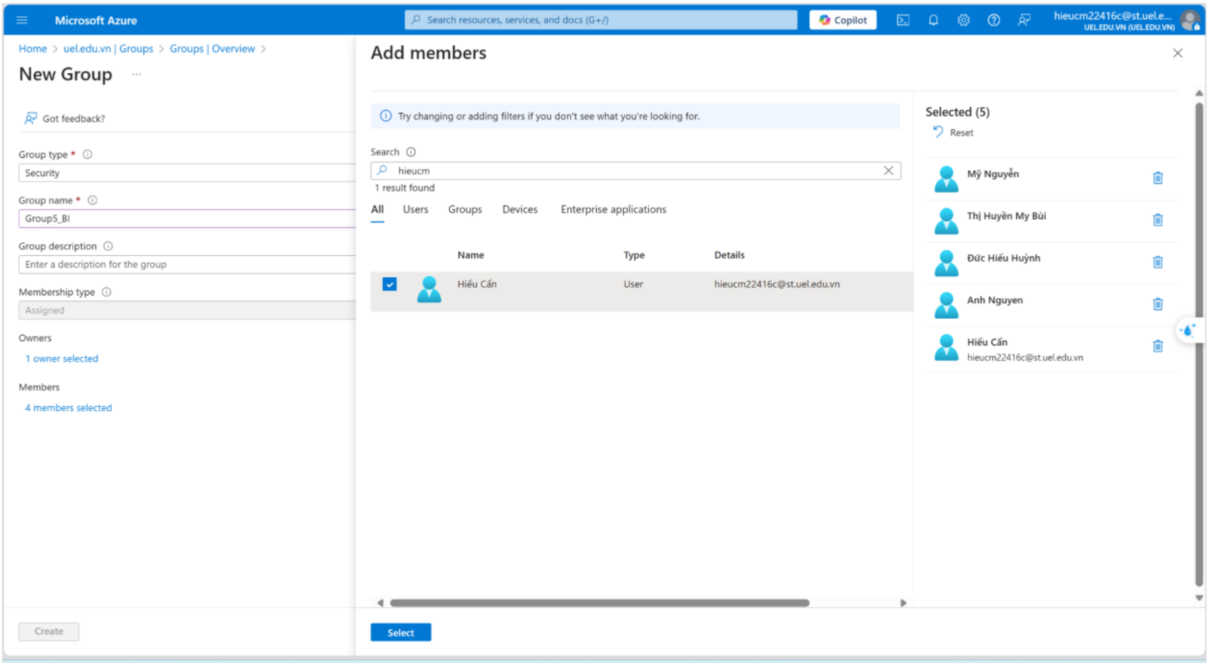
Select cluster
○ New job cluster ● Existing interactive cluster ○ Existing instance pool

Existing cluster ID *
0222-010515-7p4hm0zz

Save   Cancel   Test connection

9. Azure Active Directory – Data Governance

9.1 Create Azure AD Group



New Group

Group type * : Security

Group name * : Group5_BI

Group description : Enter a description for the group

Membership type : Assigned

Owners: 1 owner selected

Members: 4 members selected

Add members

Try changing or adding filters if you don't see what you're looking for.

Search: hieucm

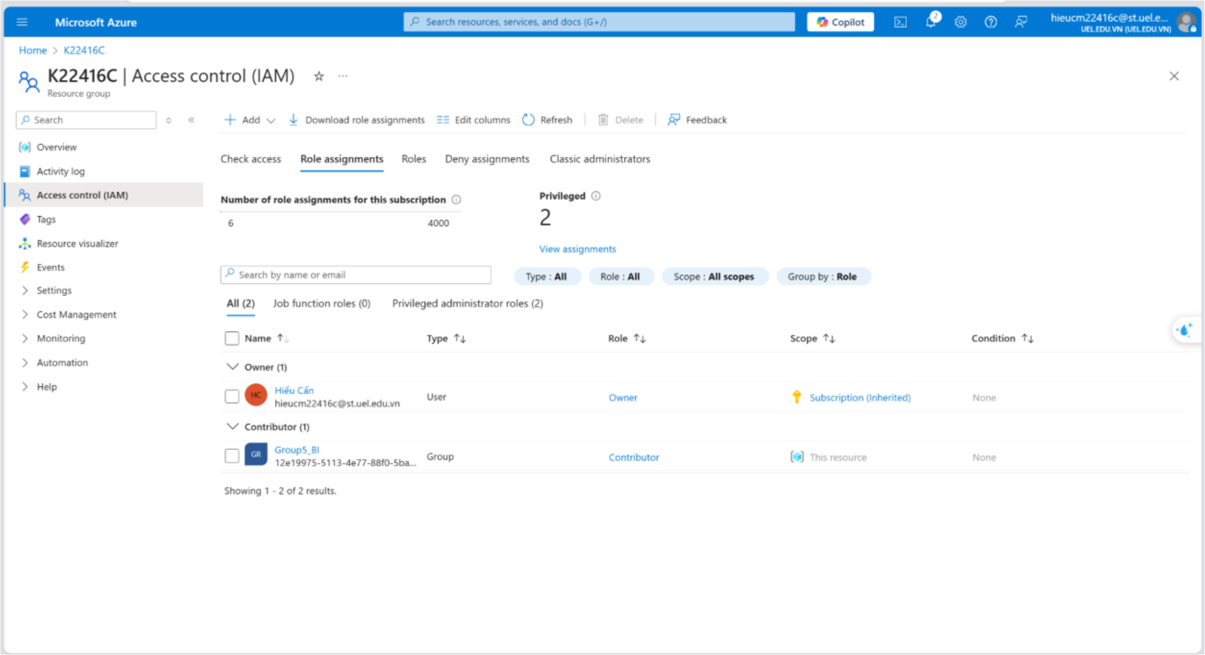
1 result found

	Name	Type	Details
☑	Hiếu Cấn	User	hieucm22416c@st.uel.edu.vn

Selected (5)

- Mỹ Nguyễn
- Thị Huyền My Bùi
- Đức Hiếu Huỳnh
- Anh Nguyen
- Hiếu Cấn
hieucm22416c@st.uel.edu.vn

9.2 Assign Role to Resource Group



K22416C | Access control (IAM)

Resource group

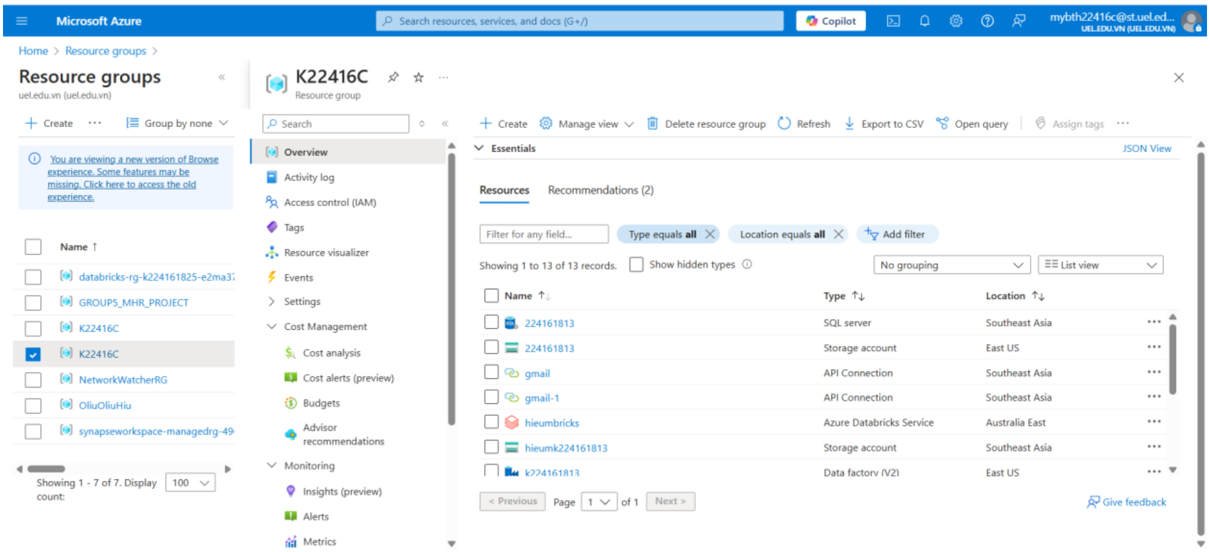
Number of role assignments for this subscription: 6 / 4000

Privileged: 2

Name	Type	Role	Scope	Condition
Owner (1)				
Hiếu Cấn hieucm22416c@st.uel.edu.vn	User	Owner	Subscription (Inherited)	None
Contributor (1)				
Group5_BI 12e19975-5113-4e77-88f0-5ba...	Group	Contributor	This resource	None

Showing 1 - 2 of 2 results.

9.3 Result



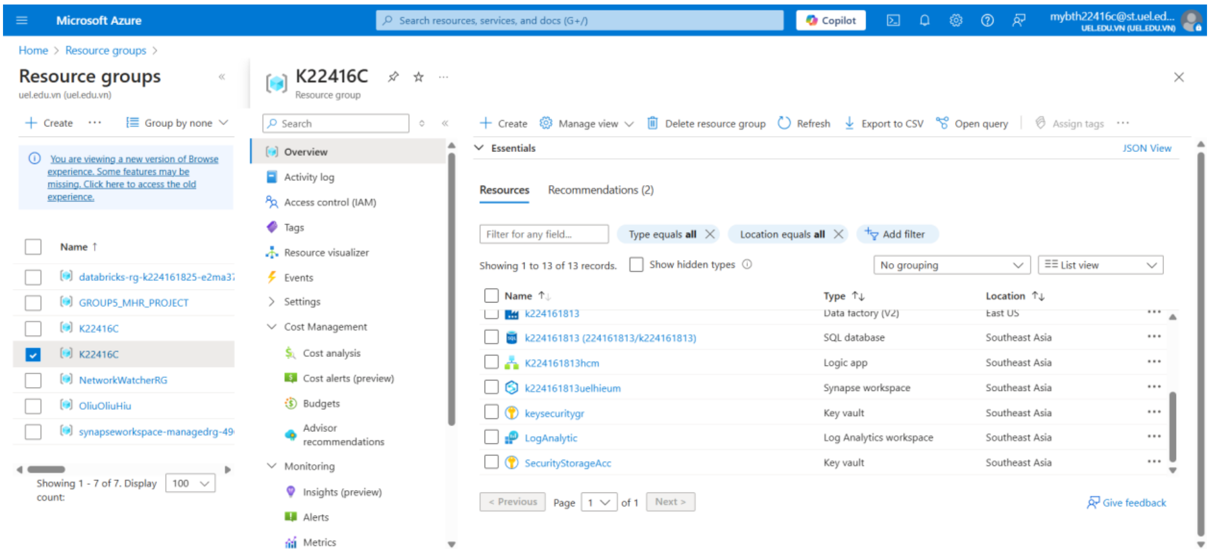
You are viewing a new version of Browse experience. Some features may be missing. Click here to access the old experience.

	Name ↑
□	databricks-rg-k224161825-e2ma3
□	GROUP5_MHR_PROJECT
□	K22416C
☑	K22416C
□	NetworkWatcherRG
□	OliuOliuHiu
□	synapseworkspace-managedrg-49

Showing 1 - 7 of 7. Display count: 100

Showing 1 to 13 of 13 records. □ Show hidden types

Name	Type	Location
224161813	SQL server	Southeast Asia
224161813	Storage account	East US
gmail	API Connection	Southeast Asia
gmail-1	API Connection	Southeast Asia
hieumbricks	Azure Databricks Service	Australia East
hieumk224161813	Storage account	Southeast Asia
k224161813	Data factory (V2)	East US



You are viewing a new version of Browse experience. Some features may be missing. Click here to access the old experience.

	Name ↑
□	databricks-rg-k224161825-e2ma3
□	GROUP5_MHR_PROJECT
□	K22416C
☑	K22416C
□	NetworkWatcherRG
□	OliuOliuHiu
□	synapseworkspace-managedrg-49

Showing 1 - 7 of 7. Display count: 100

Showing 1 to 13 of 13 records. □ Show hidden types

Name	Type	Location
k224161813	Data factory (V2)	East US
k224161813 (224161813/k224161813)	SQL database	Southeast Asia
K224161813hcm	Logic app	Southeast Asia
k224161813uelhieum	Synapse workspace	Southeast Asia
keysecuritygr	Key vault	Southeast Asia
LogAnalytic	Log Analytics workspace	Southeast Asia
SecurityStorageAcc	Key vault	Southeast Asia