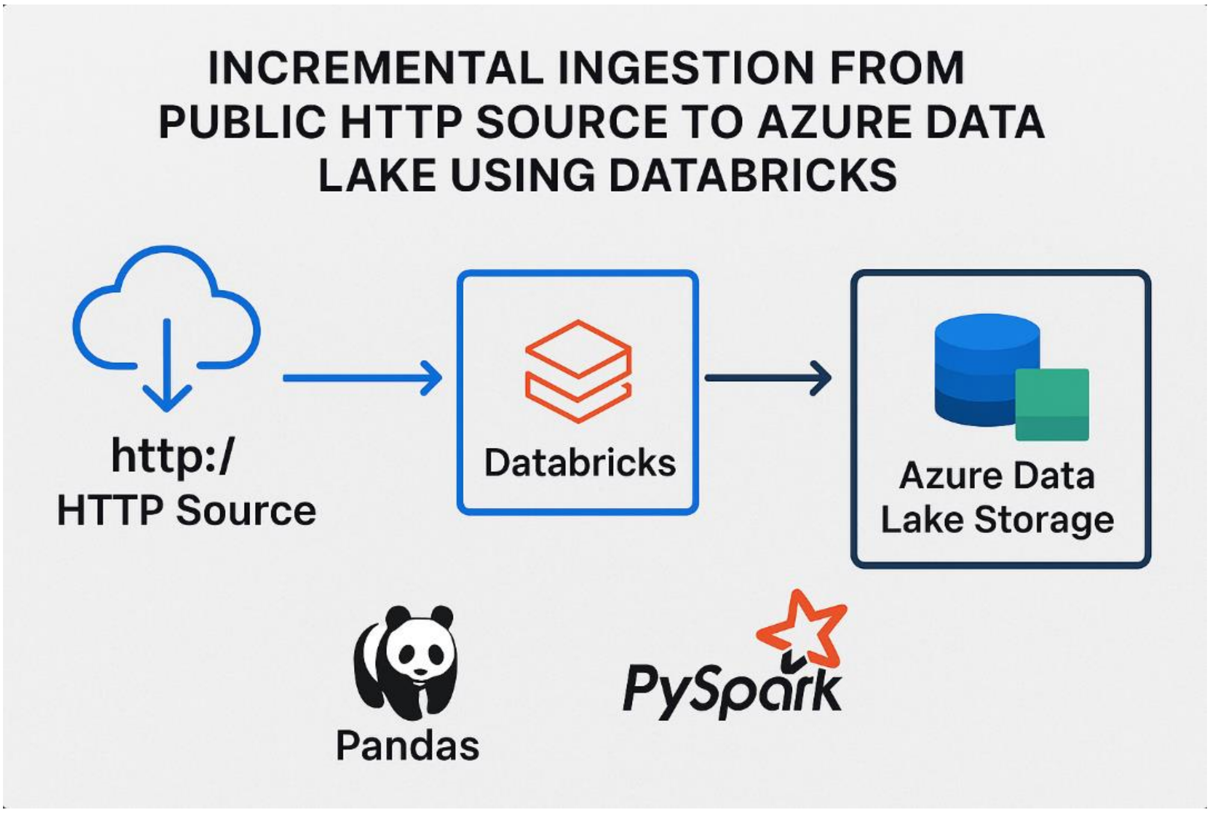
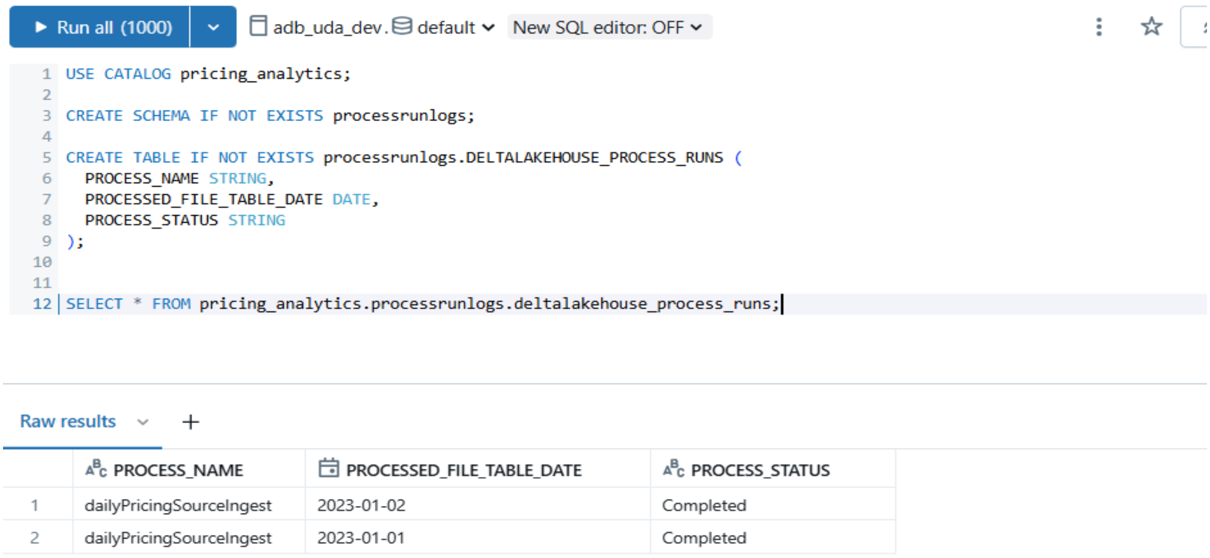
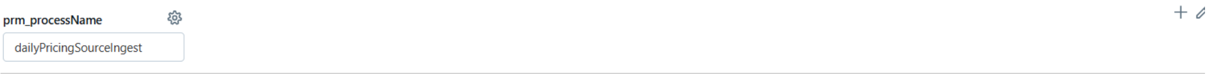


INCREMENTAL INGESTION FROM PUBLIC HTTP SOURCE TO AZURE DATA LAKE USING DATABRICKS

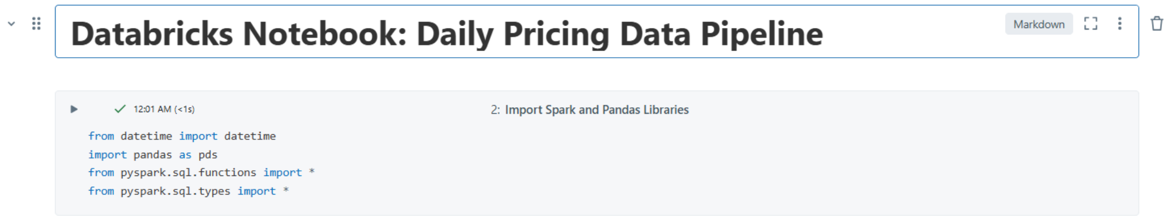


```sql
USE CATALOG pricing_analytics;

CREATE SCHEMA IF NOT EXISTS processrunlogs;

CREATE TABLE IF NOT EXISTS processrunlogs.DELTALAKEHOUSE_PROCESS_RUNS (
  PROCESS_NAME STRING,
  PROCESSED_FILE_TABLE_DATE DATE,
  PROCESS_STATUS STRING
);


SELECT * FROM pricing_analytics.processrunlogs.deltalakehouse_process_runs;
```

	PROCESS_NAME	PROCESSED_FILE_TABLE_DATE	PROCESS_STATUS
1	dailyPricingSourceIngest	2023-01-02	Completed
2	dailyPricingSourceIngest	2023-01-01	Completed



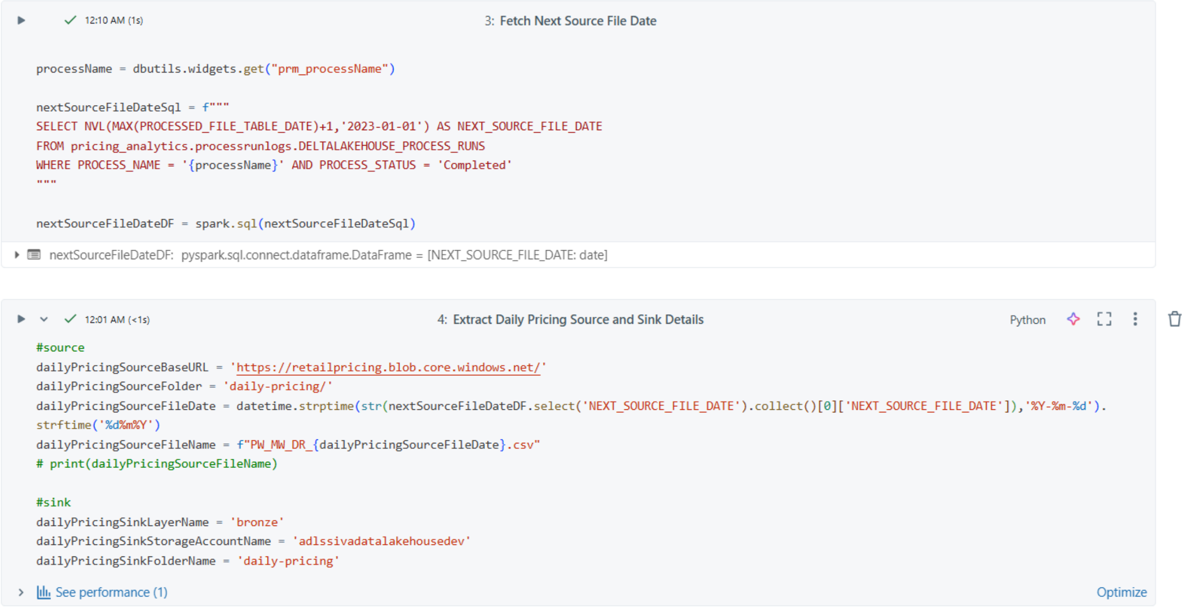
prm_processName: dailyPricingSourceIngest



Databricks Notebook: Daily Pricing Data Pipeline

2: Import Spark and Pandas Libraries

```python
from datetime import datetime
import pandas as pds
from pyspark.sql.functions import *
from pyspark.sql.types import *
```



12:10 AM (1s)

3: Fetch Next Source File Date

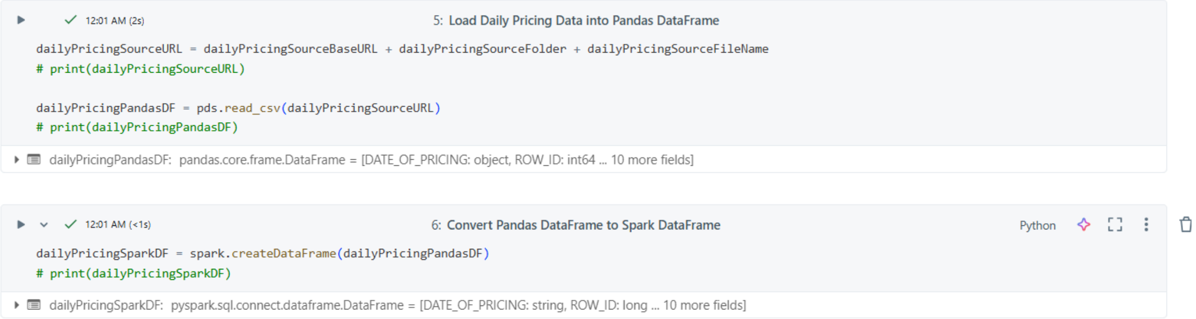
```
processName = dbutils.widgets.get("prm_processName")

nextSourceFileDateSql = f"""
SELECT NVL(MAX(PROCESSED_FILE_TABLE_DATE)+1,'2023-01-01') AS NEXT_SOURCE_FILE_DATE
FROM pricing_analytics.processrunlogs.DELTALAKEHOUSE_PROCESS_RUNS
WHERE PROCESS_NAME = '{processName}' AND PROCESS_STATUS = 'Completed'
"""

nextSourceFileDateDF = spark.sql(nextSourceFileDateSql)
```

nextSourceFileDateDF: pyspark.sql.connect.dataframe.DataFrame = [NEXT_SOURCE_FILE_DATE: date]

12:01 AM (<1s)

4: Extract Daily Pricing Source and Sink Details

```
#source
dailyPricingSourceBaseURL = 'https://retailpricing.blob.core.windows.net/'
dailyPricingSourceFolder = 'daily-pricing/'
dailyPricingSourceFileDate = datetime.strptime(str(nextSourceFileDateDF.select('NEXT_SOURCE_FILE_DATE').collect()[0]['NEXT_SOURCE_FILE_DATE']),'%Y-%m-%d').
strftime('%d%m%Y')
dailyPricingSourceFileName = f"PW_MW_DR_{dailyPricingSourceFileDate}.csv"
# print(dailyPricingSourceFileName)

#sink
dailyPricingSinkLayerName = 'bronze'
dailyPricingSinkStorageAccountName = 'adlssivadatalakehousedev'
dailyPricingSinkFolderName = 'daily-pricing'
```



12:01 AM (2s)

5: Load Daily Pricing Data into Pandas DataFrame

```
dailyPricingSourceURL = dailyPricingSourceBaseURL + dailyPricingSourceFolder + dailyPricingSourceFileName
# print(dailyPricingSourceURL)

dailyPricingPandasDF = pds.read_csv(dailyPricingSourceURL)
# print(dailyPricingPandasDF)
```

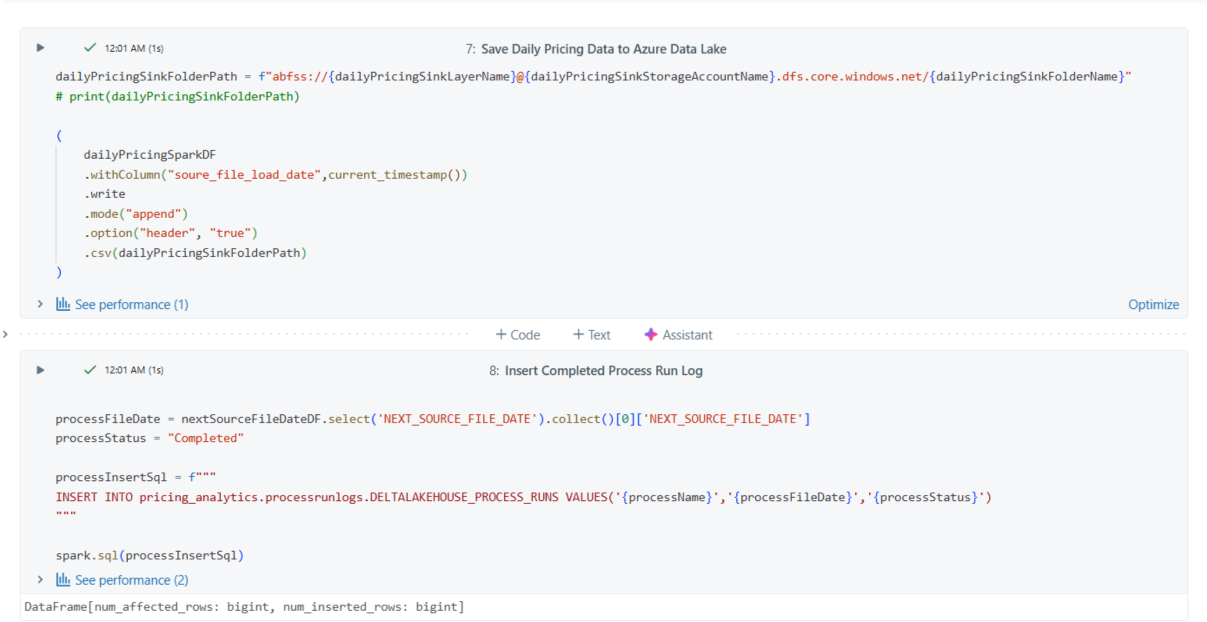
dailyPricingPandasDF: pandas.core.frame.DataFrame = [DATE_OF_PRICING: object, ROW_ID: int64 ... 10 more fields]

12:01 AM (<1s)

6: Convert Pandas DataFrame to Spark DataFrame

```
dailyPricingSparkDF = spark.createDataFrame(dailyPricingPandasDF)
# print(dailyPricingSparkDF)
```

dailyPricingSparkDF: pyspark.sql.connect.dataframe.DataFrame = [DATE_OF_PRICING: string, ROW_ID: long ... 10 more fields]



12:01 AM (1s)

7: Save Daily Pricing Data to Azure Data Lake

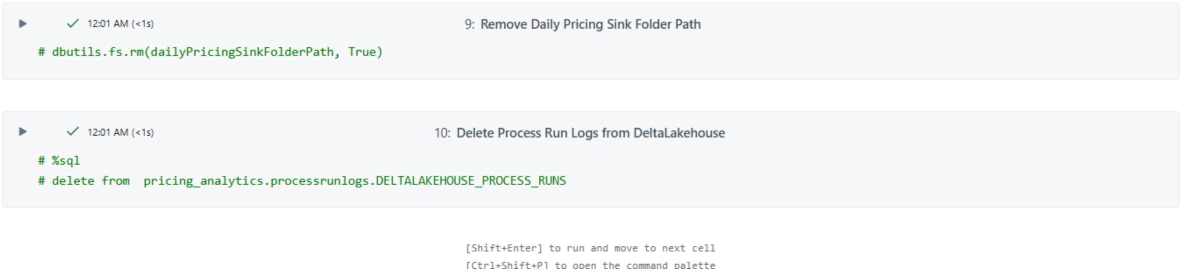
```
dailyPricingSinkFolderPath = f"abfss://{dailyPricingSinkLayerName}@{dailyPricingSinkStorageAccountName}.dfs.core.windows.net/{dailyPricingSinkFolderName}"
# print(dailyPricingSinkFolderPath)

(
    dailyPricingSparkDF
    .withColumn("soure_file_load_date",current_timestamp())
    .write
    .mode("append")
    .option("header", "true")
    .csv(dailyPricingSinkFolderPath)
)
```

12:01 AM (1s)

8: Insert Completed Process Run Log

```
processFileDate = nextSourceFileDateDF.select('NEXT_SOURCE_FILE_DATE').collect()[0]['NEXT_SOURCE_FILE_DATE']
processStatus = "Completed"

processInsertSql = f"""
INSERT INTO pricing_analytics.processrunlogs.DELTALAKEHOUSE_PROCESS_RUNS VALUES('{processName}','{processFileDate}','{processStatus}')
"""

spark.sql(processInsertSql)
```

```
DataFrame[num_affected_rows: bigint, num_inserted_rows: bigint]
```



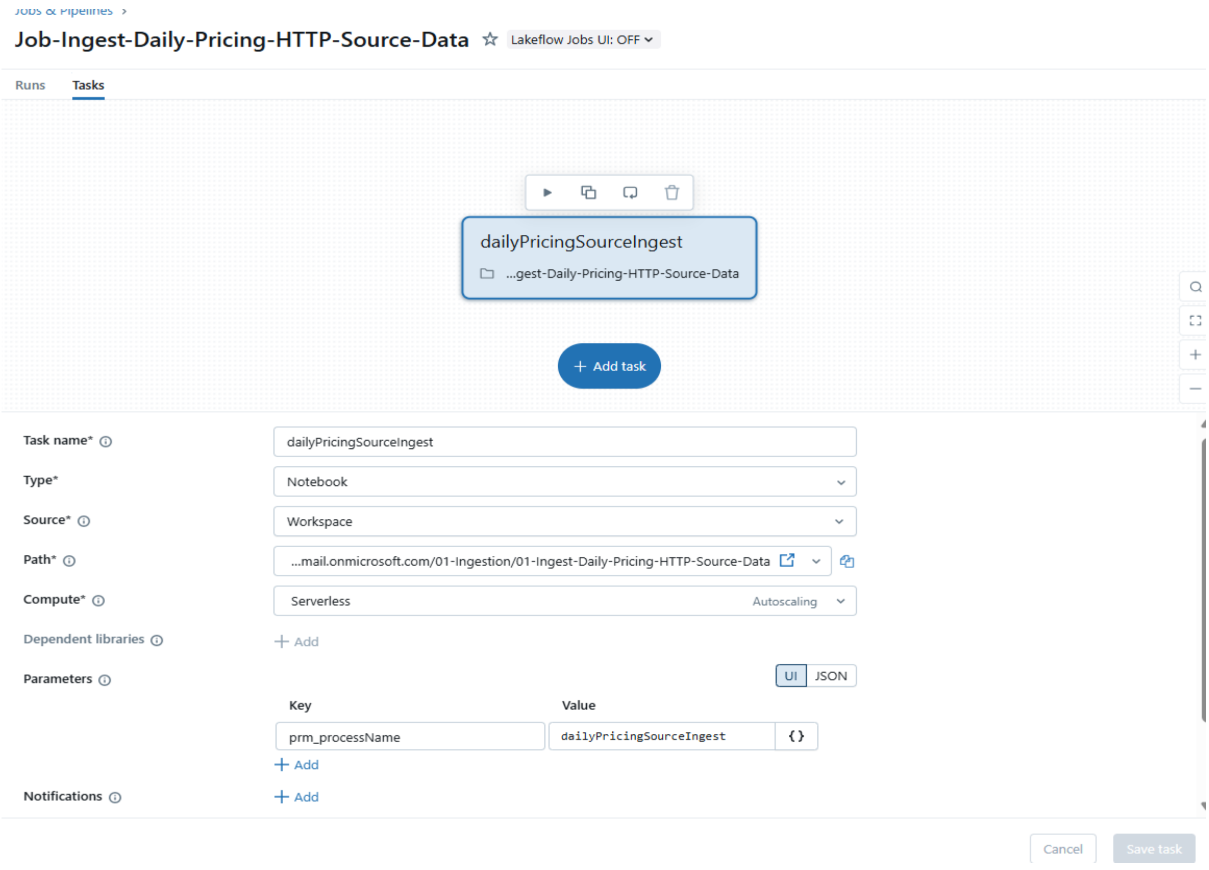
▶ ✓ 12:01 AM (<1s) 9: Remove Daily Pricing Sink Folder Path

```
# dbutils.fs.rm(dailyPricingSinkFolderPath, True)
```

▶ ✓ 12:01 AM (<1s) 10: Delete Process Run Logs from DeltaLakehouse

```
# %sql
# delete from  pricing_analytics.processrunlogs.DELTALAKEHOUSE_PROCESS_RUNS
```

```
[Shift+Enter] to run and move to next cell
[Ctrl+Shift+P] to open the command palette
```



Jobs & Pipelines ›

Job-Ingest-Daily-Pricing-HTTP-Source-Data ☆ Lakeflow Jobs UI: OFF

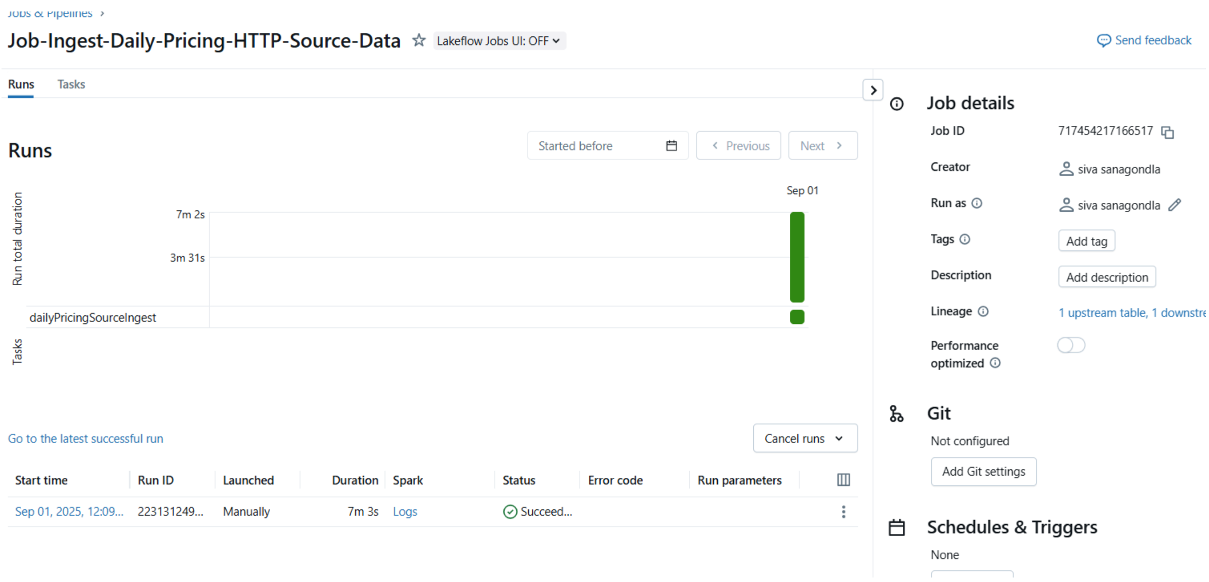
Task name* ⓘ	dailyPricingSourceIngest
Type*	Notebook
Source* ⓘ	Workspace
Path* ⓘ	...mail.onmicrosoft.com/01-Ingestion/01-Ingest-Daily-Pricing-HTTP-Source-Data
Compute* ⓘ	Serverless Autoscaling
Dependent libraries ⓘ	+ Add
Parameters ⓘ	UI JSON

Key	Value
prm_processName	dailyPricingSourceIngest

+ Add

Notifications ⓘ + Add

Cancel Save task



Jobs & Pipelines ›

Job-Ingest-Daily-Pricing-HTTP-Source-Data ☆ Lakeflow Jobs UI: OFF

Send feedback

Start time	Run ID	Launched	Duration	Spark	Status	Error code	Run parameters
Sep 01, 2025, 12:09...	223131249...	Manually	7m 3s	Logs	✓ Succeed...		

Job details

Job ID	717454217166517
Creator	siva sanagondla
Run as ⓘ	siva sanagondla
Tags ⓘ	Add tag
Description	Add description
Lineage ⓘ	1 upstream table, 1 downstre
Performance optimized ⓘ	

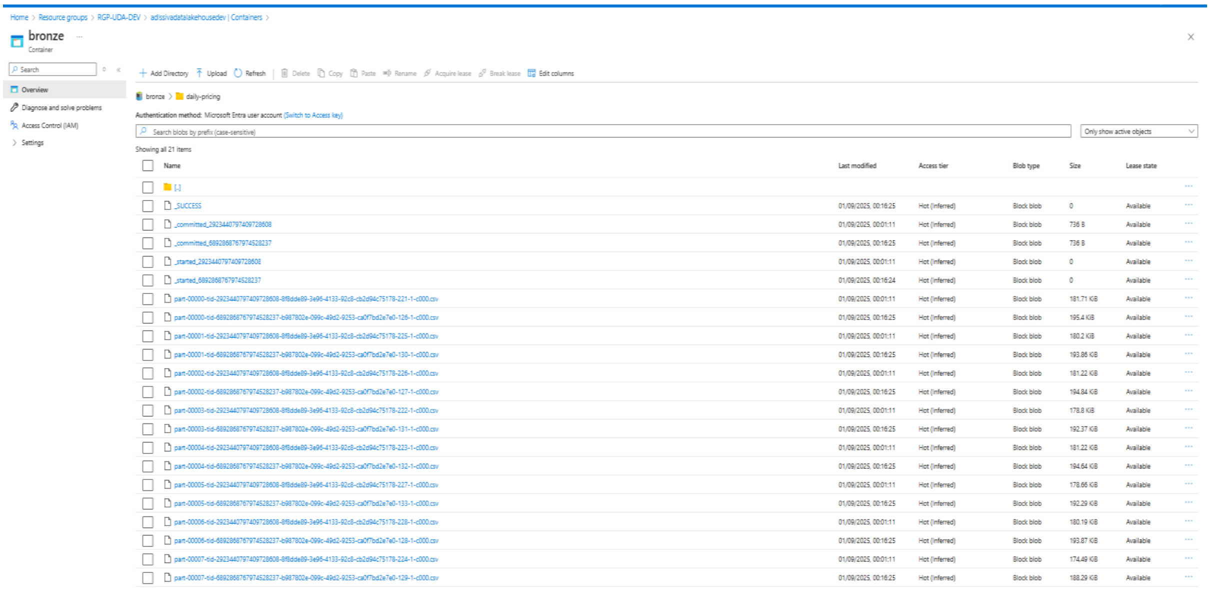
Git

Not configured

Add Git settings

Schedules & Triggers

None



Home > Resource groups > RGP-UDA-DEV > adlssivadatalakehousedev | Containers >

bronze

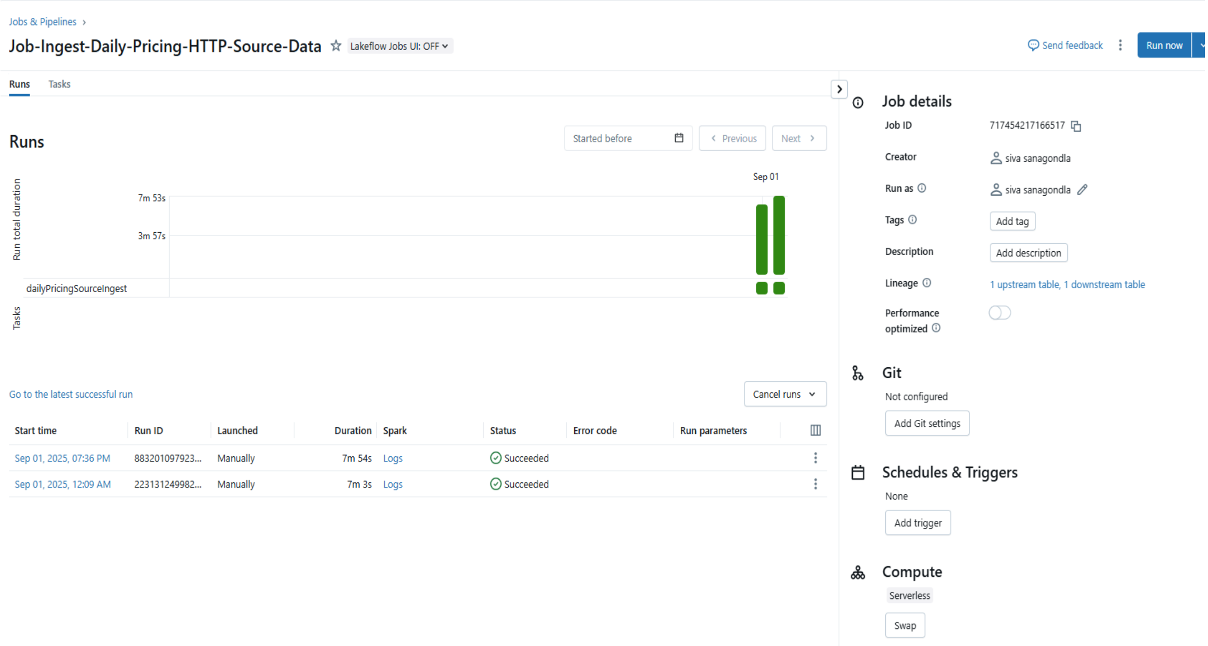
Container

bronze > daily-pricing

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

Showing all 21 items

□	Name	Last modified	Access tier	Blob type	Size	Lease state	
□	[..]						···
□	_SUCCESS	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	0	Available	···
□	_committed_2923440797409728608	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	736 B	Available	···
□	_committed_6892868767974528237	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	736 B	Available	···
□	_started_2923440797409728608	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	0	Available	···
□	_started_6892868767974528237	01/09/2025, 00:16:24	Hot (Inferred)	Block blob	0	Available	···
□	part-00000-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-221-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.71 KiB	Available	···
□	part-00000-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-126-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	195.4 KiB	Available	···
□	part-00001-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-225-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	180.2 KiB	Available	···
□	part-00001-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-130-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	193.86 KiB	Available	···
□	part-00002-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-226-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.22 KiB	Available	···
□	part-00002-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-127-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	194.84 KiB	Available	···
□	part-00003-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-222-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	178.8 KiB	Available	···
□	part-00003-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-131-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	192.37 KiB	Available	···
□	part-00004-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-223-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.22 KiB	Available	···
□	part-00004-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-132-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	194.64 KiB	Available	···
□	part-00005-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-227-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	178.66 KiB	Available	···
□	part-00005-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-133-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	192.29 KiB	Available	···
□	part-00006-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-228-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	180.19 KiB	Available	···
□	part-00006-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-128-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	193.87 KiB	Available	···
□	part-00007-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-224-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	174.49 KiB	Available	···
□	part-00007-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-129-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	188.29 KiB	Available	···



Job-Ingest-Daily-Pricing-HTTP-Source-Data

Runs

Start time	Run ID	Launched	Duration	Spark	Status	Error code	Run parameters
Sep 01, 2025, 07:36 PM	883201097923...	Manually	7m 54s	Logs	✓ Succeeded		
Sep 01, 2025, 12:09 AM	223131249982...	Manually	7m 3s	Logs	✓ Succeeded		

Job details

Job ID	717454217166517
Creator	siva sanagondla
Run as	siva sanagondla
Tags	Add tag
Description	Add description
Lineage	1 upstream table, 1 downstream table
Performance optimized	

Git

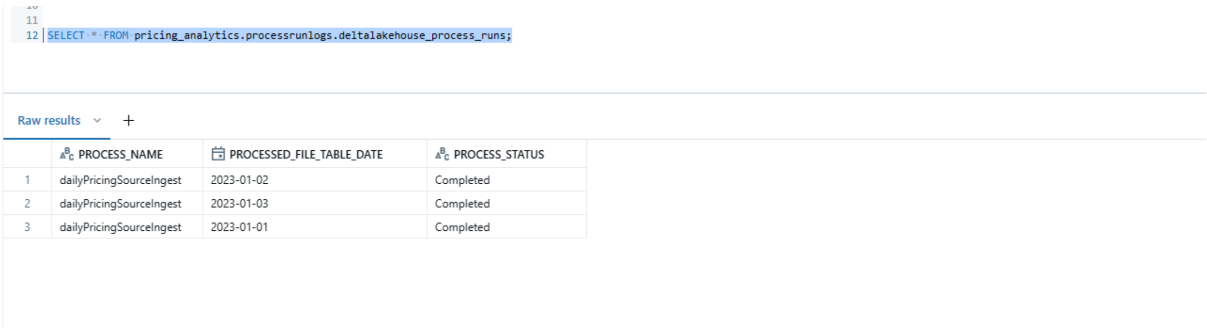
Not configured

Schedules & Triggers

None

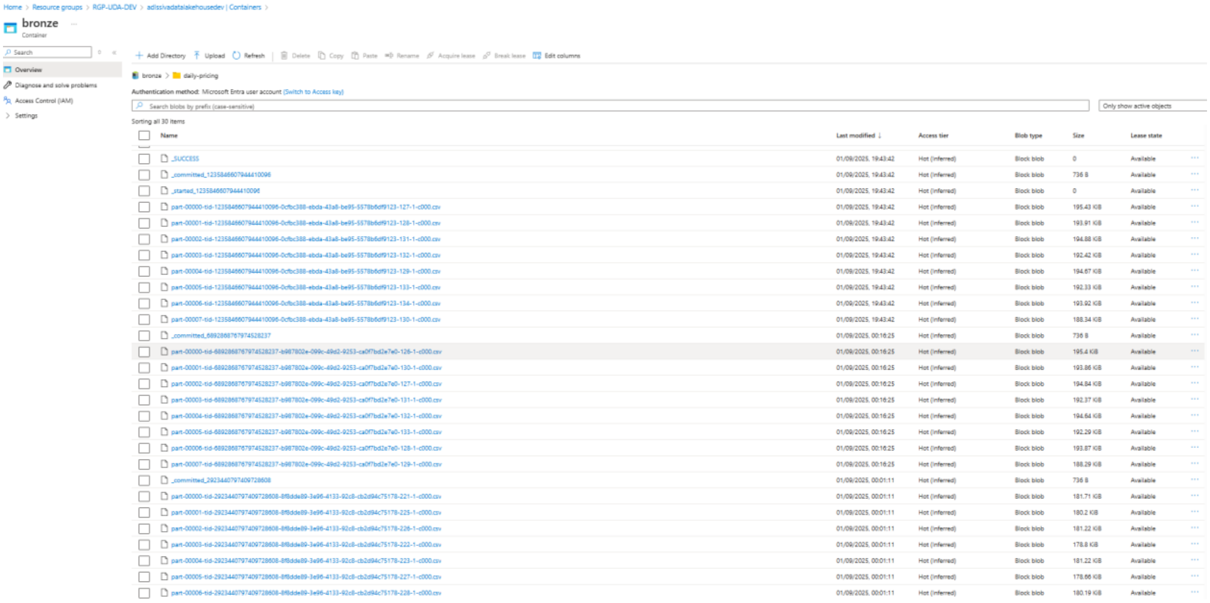
Compute

Serverless



```
SELECT * FROM pricing_analytics.processrunlogs.deltalakehouse_process_runs;
```

	PROCESS_NAME	PROCESSED_FILE_TABLE_DATE	PROCESS_STATUS
1	dailyPricingSourceIngest	2023-01-02	Completed
2	dailyPricingSourceIngest	2023-01-03	Completed
3	dailyPricingSourceIngest	2023-01-01	Completed


Home > Resource groups > RGP-UDA-DEV > adlssivadatalakehousedev | Containers >

bronze ···
Container

bronze > daily-pricing

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

Search blobs by prefix (case-sensitive)

Only show active objects

Sorting all 30 items

☐	Name	Last modified ↓	Access tier	Blob type	Size	Lease state	
☐	_SUCCESS	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	0	Available	···
☐	_committed_1235846607944410096	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	736 B	Available	···
☐	_started_1235846607944410096	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	0	Available	···
☐	part-00000-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-127-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	195.43 KiB	Available	···
☐	part-00001-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-128-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	193.91 KiB	Available	···
☐	part-00002-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-131-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	194.88 KiB	Available	···
☐	part-00003-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-132-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	192.42 KiB	Available	···
☐	part-00004-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-129-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	194.67 KiB	Available	···
☐	part-00005-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-133-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	192.33 KiB	Available	···
☐	part-00006-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-134-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	193.92 KiB	Available	···
☐	part-00007-tid-1235846607944410096-0cfbc388-ebda-43a8-be95-5578b6df9123-130-1-c000.csv	01/09/2025, 19:43:42	Hot (Inferred)	Block blob	188.34 KiB	Available	···
☐	_committed_6892868767974528237	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	736 B	Available	···
☐	part-00000-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-126-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	195.4 KiB	Available	···
☐	part-00001-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-130-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	193.86 KiB	Available	···
☐	part-00002-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-127-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	194.84 KiB	Available	···
☐	part-00003-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-131-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	192.37 KiB	Available	···
☐	part-00004-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-132-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	194.64 KiB	Available	···
☐	part-00005-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-133-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	192.29 KiB	Available	···
☐	part-00006-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-128-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	193.87 KiB	Available	···
☐	part-00007-tid-6892868767974528237-b987802e-099c-49d2-9253-ca0f7bd2e7e0-129-1-c000.csv	01/09/2025, 00:16:25	Hot (Inferred)	Block blob	188.29 KiB	Available	···
☐	_committed_2923440797409728608	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	736 B	Available	···
☐	part-00000-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-221-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.71 KiB	Available	···
☐	part-00001-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-225-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	180.2 KiB	Available	···
☐	part-00002-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-226-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.22 KiB	Available	···
☐	part-00003-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-222-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	178.8 KiB	Available	···
☐	part-00004-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-223-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	181.22 KiB	Available	···
☐	part-00005-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-227-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	178.66 KiB	Available	···
☐	part-00006-tid-2923440797409728608-8f8dde89-3e96-4133-92c8-cb2d94c75178-228-1-c000.csv	01/09/2025, 00:01:11	Hot (Inferred)	Block blob	180.19 KiB	Available	···