

A Complete Real Time End to End Azure Data Engineering Project 02

Lets build a complete End to End Azure Data Engineering Project in 3 hours. In this project we are going to create an end to end data platform right from Data Ingestion, Data Transformation, Data Loading and Reporting.

The tools that are covered in this project are,

1. Azure Data Factory
2. Azure Data Lake Storage Gen2
3. Azure Databricks
4. Azure Synapse Analytics
5. Azure Key vault
6. Azure Active Directory (AAD) and
7. Microsoft Power BI

The use case for this project is building an end to end solution by ingesting the tables from on-premise SQL Server database using Azure Data Factory and then store the data in Azure Data Lake. Then Azure databricks is used to transform the RAW data to the most cleanest form of data and then we are using Azure Synapse Analytics to load the clean data and finally using Microsoft Power BI to integrate with Azure synapse analytics to build an interactive dashboard. Also, we are using Azure Active Directory (AAD) and Azure Key Vault for the monitoring and governance purpose.

Chapters

TimeStamp:

0:00:00 - Project Overview

0:13:40 - Environment Setup

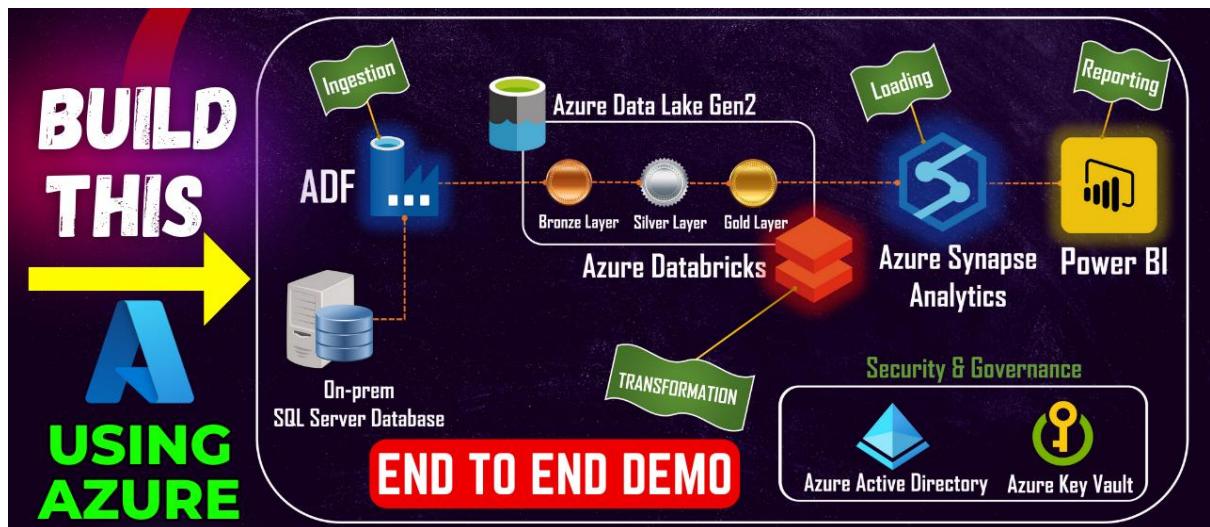
0:22:34 - Data Ingestion using ADF (Part 1)

0:45:48 - Data Ingestion using ADF (Part 2)

1:07:36 - Data Transformation using Databricks (Part 1)

1:22:57 - Data Transformation using Databricks (Part 2)
1:43:46 - Data Transformation using Databricks (Part 3)
1:58:59 - Data Loading using Azure Synapse Analytics
2:22:22 - Data Reporting using Microsoft Power BI
2:36:59 - Creating Security Group using Azure Active Directory (AAD)
2:44:00 - End to End Pipeline Testing

Project Data Pipeline Design



The screenshot shows the Microsoft Azure portal home page. The top navigation bar includes links for Home, A Complete Real Time End to E..., portal.azure.com/?pwa=1#home, Apps, Gmail, YouTube, Your Dashboard | Fo..., Dashboard - Great L..., Career path: Data E..., Career Path to Beco..., Novorésumé MYRE..., TNTBESC books por..., Microsoft Azure, Upgrade, Search resources, services, and docs (G+), and balachandar2014elu@g... DEFAULT DIRECTORY. Below the navigation bar, there are sections for Azure services (Create a resource, Resource groups, Azure Synapse Analytics, Data factories, Storage accounts, Azure Databricks, App registrations, Network security groups, Virtual machines, More services) and Resources (Recent, Favorite, Name, Type, Last Viewed). A message states 'No resources have been viewed recently' with a 'View all resources' button. At the bottom, there is a 'Navigate' bar with a search input, pinned icons, and system status indicators (0.2 KB/s, 2.2 KB/s, ENG, 06:56 AM, 17-04-2024).

My Azure portal...

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. In the Object Explorer, the database 'AdventureWorksLT2017' is selected. In the center pane, a query window titled 'CreateloginSQLQuery1.sql' is open, displaying the following T-SQL code:

```
CREATE LOGIN balachandar WITH PASSWORD = 'Highspeedengineer01x';
create user bala for login balachandar
```

The 'Messages' pane at the bottom shows the command completed successfully with a completion time of 2024-04-17T18:04:52.8954284+05:30. A status bar at the bottom indicates 'Query executed successfully.'

The screenshot shows the Microsoft Azure portal. The URL in the address bar is 'portal.azure.com/?pwa=1# create/Microsoft.KeyVault'. The page title is 'Create a key vault'. The 'Basics' tab is selected. The 'Subscription' dropdown is set to 'Free Trial' and the 'Resource group' dropdown is set to 'sbc-data-engineering-project-02'. At the bottom, there are 'Previous', 'Next', and 'Review + create' buttons, with 'Review + create' being the active button.

Create the key vault to store the username and password of the sql server(SSMS)

The screenshot shows two Microsoft Azure portal windows side-by-side.

Top Window: sbc-data-valut-001 | Secrets

- Left Sidebar:** Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Access policies, Events, Objects (Keys, Secrets, Certificates), Settings, Access configuration.
- Right Content Area:** Shows a table of secrets:

Name	Type	Status	Expiration date
password		✓ Enabled	
username		✓ Enabled	

A message at the top states: "The secret 'password' has been successfully created."

Bottom Window: Data factories - Microsoft Azure

- Left Sidebar:** Home, Data factories (selected), Create, Manage view, Refresh, Export to CSV, Open query, Assign tags.
- Content Area:** Displays a message: "No data factories to display. Try changing or clearing your filters." It includes a "Create data factory" button and a "Learn more" link.

Create the Data factory resource for Data Ingestion and Pipeline creation.

Microsoft.DataFactory-20240417185744 | Overview

Your deployment is complete

Deployment name : Microsoft.DataFactory-20240417185744 Start time : 17/4/2024, 7:01:07 pm
Subscription : Free Trial Correlation ID : 760a0747-c493-428a-9fcf-5a1678b514...
Resource group : sbc-data-engineering-project-02

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

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

0.1 KB/s 0.1 KB/s 07:02 PM 17-04-2024

adf-sbc-point-01 - Microsoft A | New Tab

Microsoft Azure | Upgrade

Home > Microsoft.DataFactory-20240417185744 | Overview

adf-sbc-point-01

Data factory (V2)

Search | Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Essentials

Resource group (move) : sbc-data-engineering-project-02

Status : Succeeded

Type : Data factory (V2)

Location : East US

Getting started : Quick_start

Subscription (move) : Free Trial

Subscription ID : 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6

Azure Data Factory Studio

Launch studio

1.1 KB/s 19.6 KB/s 07:02 PM 17-04-2024

Home - Microsoft Azure

portal.azure.com/?pwa=1#home

Microsoft Azure

Azure services

Create a resource Data factories Key vaults Resource groups Azure Synapse Analytics Storage accounts Databricks App registrations Network security groups More services

Resources

Recent Favorite

Name	Type	Last Viewed
adf-sbc-point-01	Data factory (V2)	a few seconds ago
sbc-data-valut-001	Key vault	17 minutes ago
sbc-data-engineering-project-02	Resource group	18 minutes ago

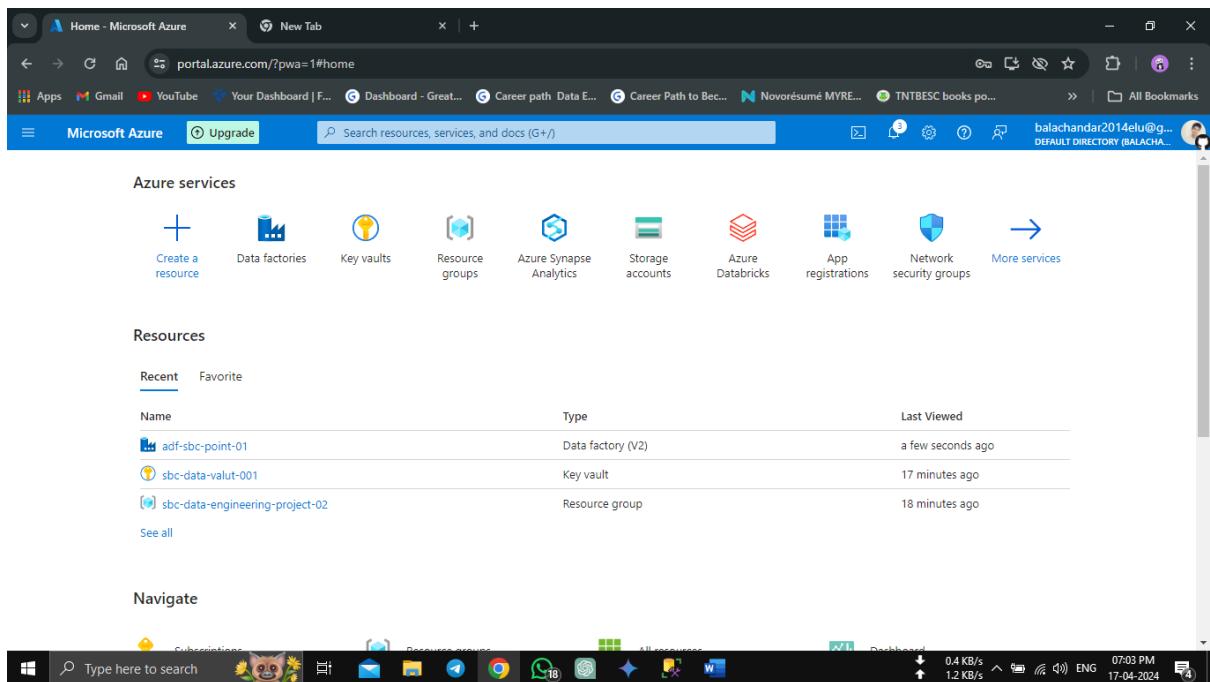
See all

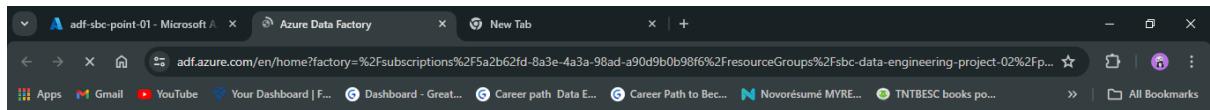
Navigate

Subscriptions

Dashboard

0.4 KB/s 1.2 KB/s ENG 07:03 PM 17-04-2024





adf-sbc-point-01 - Microsoft Edge

adf-sbc-point-01 - Azure Data

New Tab

adf.azure.com/en/home?factory=%2Fsubscriptions%2F5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6%2FresourceGroups%2Fsbcd-data-engineering-project-02%2Fp...

Apps Gmail YouTube Your Dashboard | F... Dashboard - Great... Career path Data E... Career Path to Bec... Novorésumé MYRE... TNTBESC books po... All Bookmarks

Microsoft Azure | Data Factory > adf-sbc-point-01

Search factory and documentation

balachandar2014elu@gmail.com

Data factory

adf-sbc-point-01

New

Ingest

Orchestrate

Transform data

Configure SSIS

Recent resources

No items to show

Type here to search

88.9 KB/s
12.9 KB/s
8.0 KB/s
0.0 KB/s
ENG
07:03 PM
17-04-2024

Azure Data Factory for Pipeline creation work ..

The screenshot shows the 'Integration runtime setup' page in the Azure Data Factory interface. On the left, a sidebar lists various settings like General, Factory settings, Connections, and Integration runtimes. The main area displays three integration runtime options:

- Azure, Self-Hosted**: Described as performing data flows, data movement and dispatch activities to external compute.
- Azure-SSIS**: Described as lift-and-shift existing SSIS packages to execute in Azure.
- Airflow (Preview)**: Described as using this for running your existing DAGs.

At the bottom right of the main area are 'Continue' and 'Cancel' buttons.

The screenshot shows the 'Integration runtime setup' page in the Azure Data Factory interface, specifically the 'Create' step. The 'Name' field is filled with 'SHIR'. The 'Description' field contains the text: 'Self-hosted integration runtime setup-machine. Used to connect the On-premises SQL Server.' The 'Type' field is set to 'Self-Hosted'. At the bottom right are 'Create' and 'Back' buttons.

Install the Self hosted runtime integration in Manage selection of ADF.

The screenshot shows two windows side-by-side. The left window is the 'Integration runtimes' page in the Microsoft Azure Data Factory portal. It lists one item: 'AutoResolveIntegrationR...' of type 'Azure'. The right window is a 'Microsoft Integration Runtime Express Setup' dialog box. It displays a progress bar with the following steps: 'Loading configuration', 'Downloading Integration Runtime (Self-hosted)', 'Installing Integration Runtime (Self-hosted)', and 'Registering Integration Runtime (Self-hosted)'. Below the progress bar, an error message is shown: 'Error Code: 1002' and 'Error: A problem occurred while receiving the configuration file from the server. Suggestion: You may retry the express setup on Azure portal, or switch over to manual setup.' Both windows have a Windows taskbar at the bottom.

Failed to install Express setup SHIR. So, I go for manual method installation.

The Microsoft Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory and Azure Synapse Analytics to provide data integration capabilities across different network environments.

Important! Selecting a language below will dynamically

Select language English

If your download does not start in 30 seconds, click here.

Install Instructions

Download and run the IntergrationRuntime.msi (64-bit) to install the IntegrationRuntime.msi to your local disk to install later.

Completed the Microsoft Integration Runtime Setup Wizard

Click the Finish button to exit the Setup Wizard.

Back Finish Cancel

[Expand all](#) | [Collapse all](#)

Back To Top

Details

Type here to search

adf-sbc-point-01 - Microsoft A... | adf-sbc-point-01 - Azure Data | Download Microsoft Integrati... | Downloads

0.2 KB/s 0.1 KB/s 07:44 PM 17-04-2024

Home - Microsoft Azure | adf-sbc-point-01 - Microsoft A... | adf-sbc-point-01 - Azure Data | Troubleshoot self-hosted integ...

adf.azure.com/en/management/integrationruntimes?factory=%2Fsubscriptions%2F5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6%2FResourceGroups%2Fsb...

Microsoft Integration Runtime Configuration Manager

Home Settings Diagnostics Update Help

Self-hosted node is connected to the cloud service

Data Factory: adf-sbc-point-01

Integration Runtime: SHIR

Node: SELVI

Stop Service

Data Source Credential

Credential store: On-premises

Credential status: In sync

Last backup time: N/A

Generate Backup Import Backup

Connected to the cloud service (Data Factory V2)

5.3 KB/s 4.8 KB/s 07:32 AM 18-04-2024

Self Hosted Integrated Runtime is running successfully in manual mode installation

adf.azure.com/en/management/integrationruntimes?factory=%2Fsubscriptions%2F5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6%2FresourceGroups%2Fsbcd...

Microsoft Azure | Data Factory > adf-sbc-point-01

Integration runtimes

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

Learn more [\[link\]](#)

+ New [\[link\]](#) Refresh

Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationR...	Azure	Public	Running	0	Auto Resolve	---
SHIR	Self-Hosted	---	Running	0	---	5.40.8856.1

Preview experience [\[link\]](#) Off

Home Author Monitor Manage Learning Center

Type here to search

adf.azure.com/en/authoring?factory=%2Fsubscriptions%2F5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6%2FresourceGroups%2Fsbcd...

Microsoft Azure | Data Factory > adf-sbc-point-01

Factory Resources

Filter resources by name

+ Pipeline

Pipelines

Change Data Capture (preview)

Datasets

Data flows

Power Query

Copy Data tool

Pipeline

Template gallery

Import from pipeline template

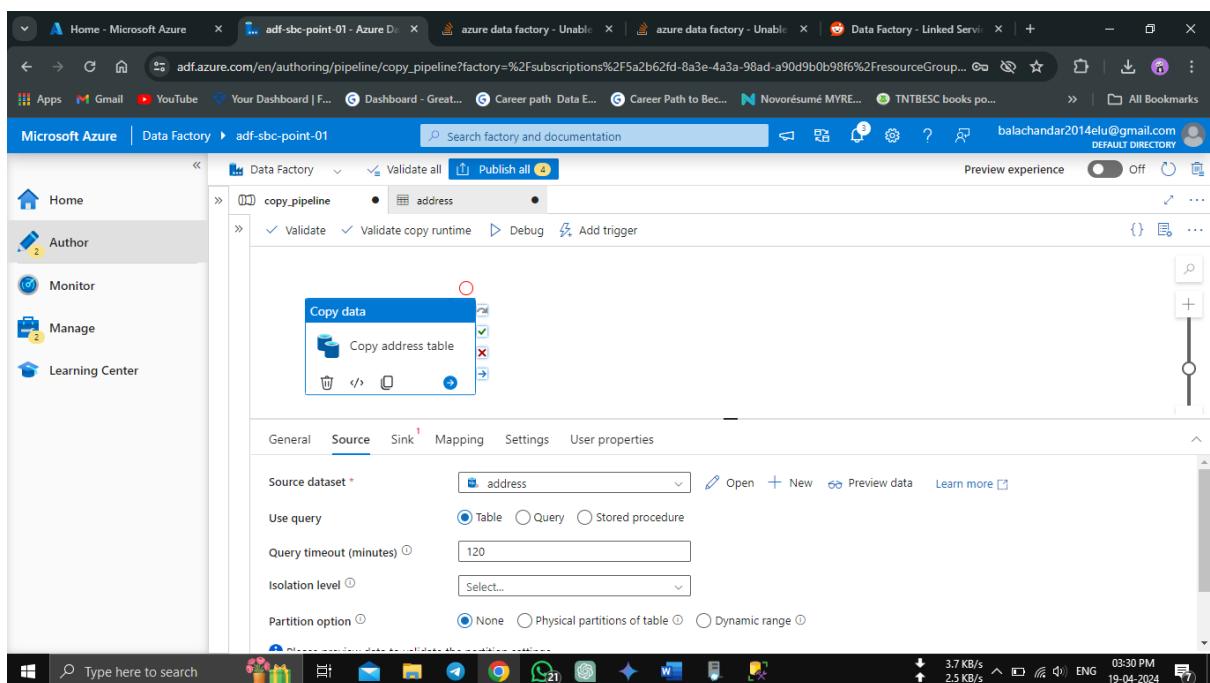
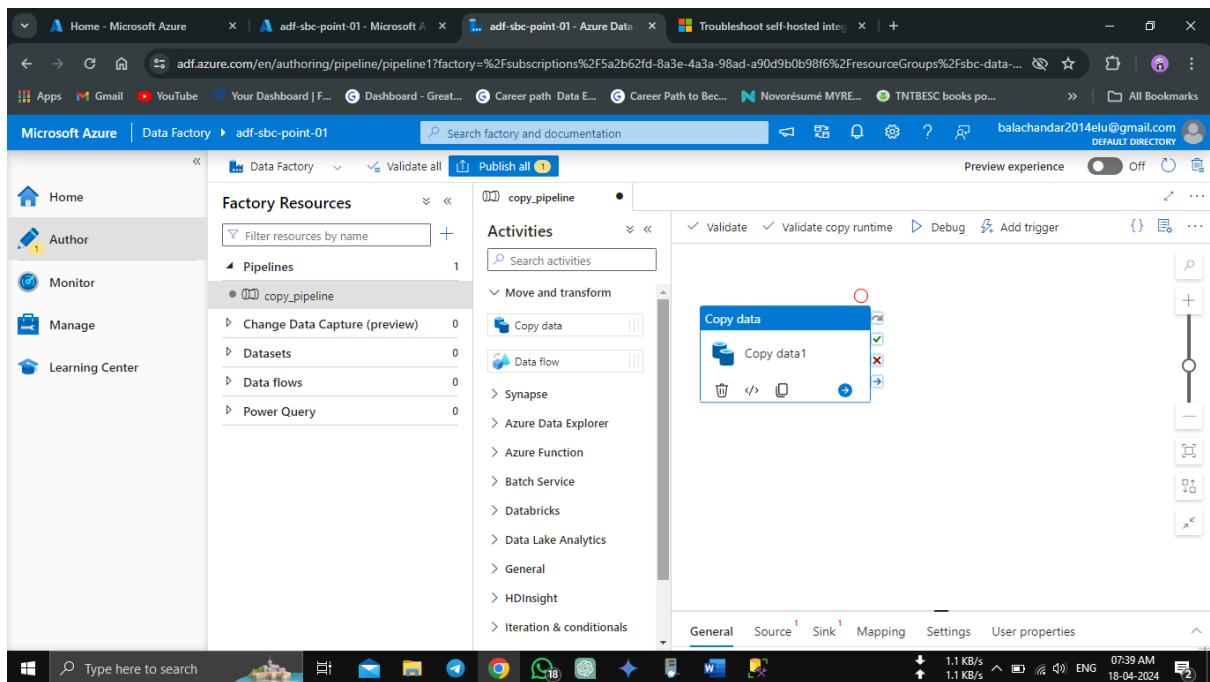
Select an item

Use the resource explorer to select or create a new item

Preview experience [\[link\]](#) Off

Home Author Monitor Manage Learning Center

Type here to search



Use the Copy activity in ADF and do Data ingestion from the source (on-prem SQL Server).

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a navigation sidebar with options like Home, Author, Monitor, Manage, and Learning Center. The main area is titled "Preview data" and shows a table of data from a linked service named "onpremssqlserver". The table has columns: AddressID, AddressLine1, AddressLine2, City, StateProvince, CountryRegion, PostalCode, rowguid, and ModifiedDate. There are four rows of data:

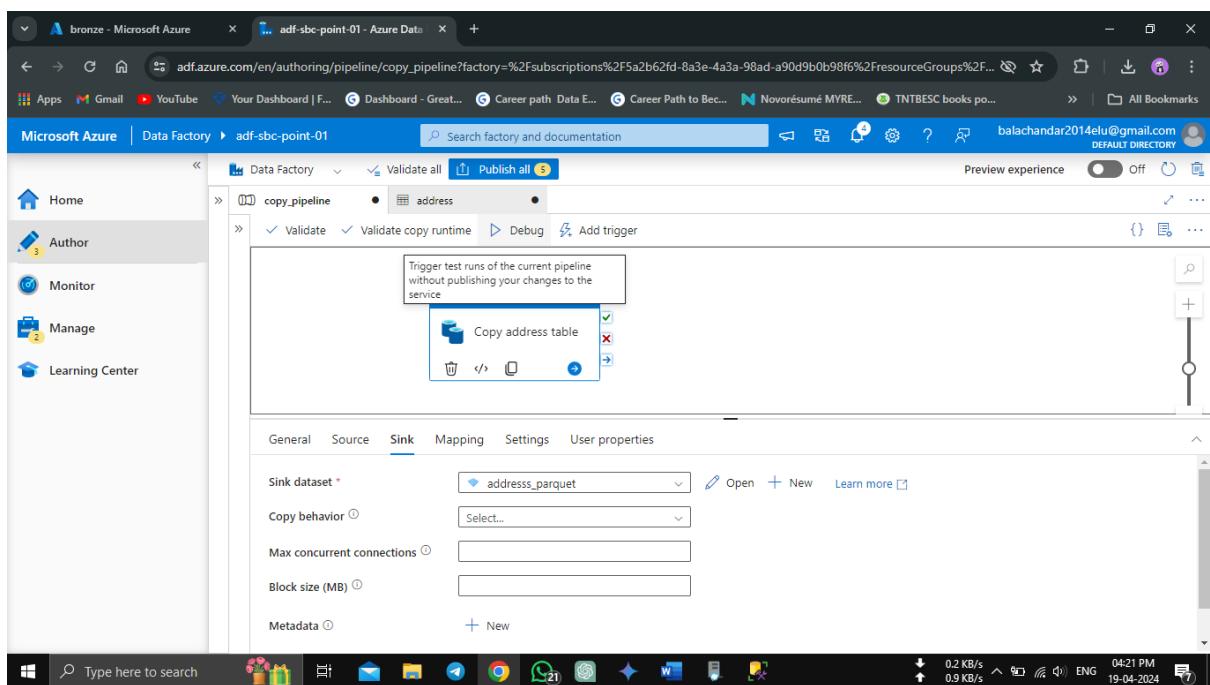
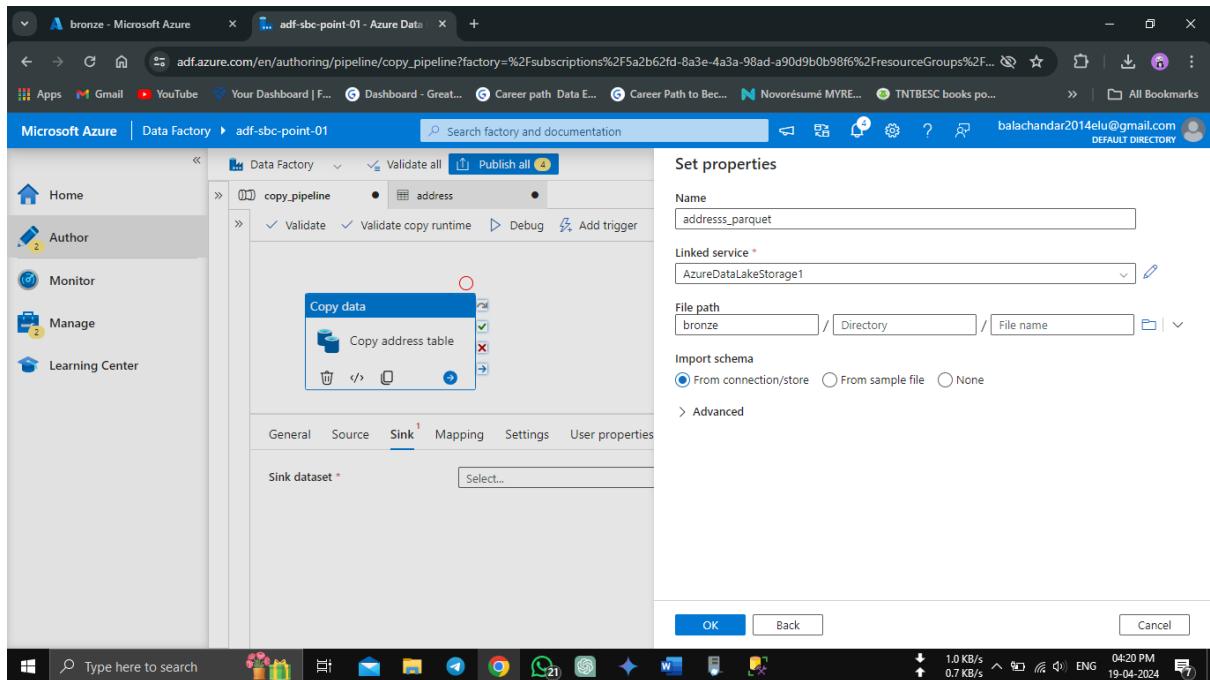
	AddressID	AddressLine1	AddressLine2	City	StateProvince	CountryRegion	PostalCode	rowguid	ModifiedDate
1	9	8713 Yosemite Ct.		Bothell	Washington	United States	98011	268af621-76d7-4c78-9441-144fd139821a	2006-07-01T00:00:00
2	11	1318 Lasalle Street		Bothell	Washington	United States	98011	981b5303-ac2-49c7-9a96-fb670785b269	2007-04-01T00:00:00
3	25	9178 Jumping St.		Dallas	Texas	United States	75201	c8df3bd9-48f0-4654-a8dd-14a67a84d3c6	2006-09-01T00:00:00
4	28	9228 Via Del Sol		Phoenix	Arizona	United States	85004	12ae5ee1-fc3e-468b-9b92-3b970b169774	2005-09-01T00:00:00

The screenshot shows the "Create a storage account" wizard on the Microsoft Azure portal. The page title is "Create a storage account". At the top, there are tabs for Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review + create. The Basics tab is selected. Below the tabs, there's a brief description of Azure Storage and a link to learn more. The "Project details" section asks for a subscription (Free Trial) and a resource group (sbc-data-engineering-project-02). The "Instance details" section shows a storage account name (shchatalaken?). At the bottom, there are "Previous", "Next", and "Review + create" buttons, along with a "Give feedback" link.

Create the Data lake Gen2 Storage ..

The screenshot shows the Microsoft Azure portal interface. The main title bar reads "sbcdatalakegen2_1713521915358 | Overview". On the left, there's a navigation menu with "Deployment" selected. The main content area displays a message: "Deployment is in progress". Below this, it shows deployment details: Deployment name: sbcdatalakegen2_1713521915358, Subscription: Free Trial, Resource group: sbc-data-engineering-project-02. It also lists "Deployment details" and "Give feedback". To the right, there are promotional banners for Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes "bronze - Microsoft Azure" and "adf-sbc-point-01 - Azure Data". The left sidebar has options like Home, Author, Monitor, Manage, and Learning Center. The main workspace shows a "copy_pipeline" pipeline with a "Copy data" activity. On the right, a "New linked service" dialog is open, prompting for "Name" (AzureDataLakeStorage1), "Description", "Connect via integration runtime" (AutoResolveIntegrationRuntime), "Authentication type" (Account key), "Account selection method" (From Azure subscription), "Azure subscription" (Free Trial (5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6)), and "Storage account name". At the bottom of the dialog are "Create" and "Cancel" buttons, along with a "Test connection" link.

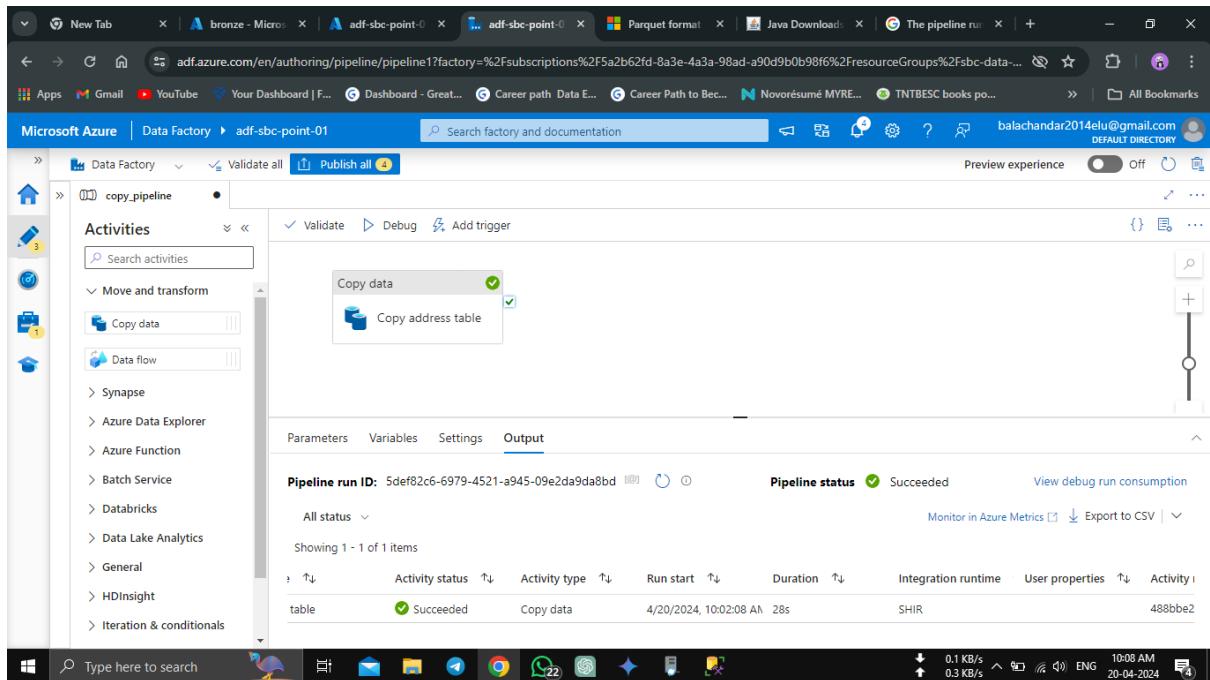


The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a navigation sidebar with options: Home, Author (selected), Monitor, Manage, and Learning Center. The main area displays a pipeline named "copy_pipeline". Inside the pipeline, there is one activity named "Copy address table" under the "Copy data" category. The pipeline run ID is c0e1a57a-c29b-4f6c-b2d3-352226826098, and the status is "In progress". Below the pipeline details, there's a table showing the activity status:

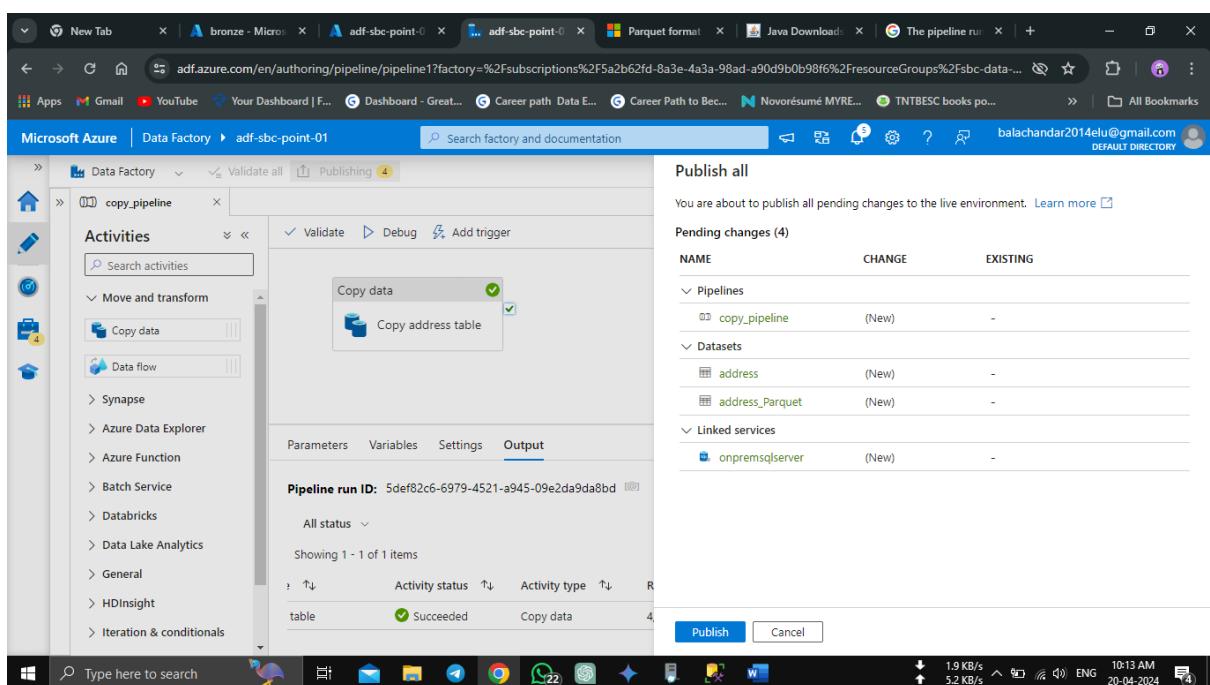
Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Copy address table	Queued	Copy data	4/19/2024, 4:25:03 PM	Less than 1s		

At the bottom of the screen, the Windows taskbar shows various pinned icons and the system tray indicates the date and time as 19-04-2024 at 04:25 PM.

This screenshot is nearly identical to the one above, showing the same Microsoft Azure Data Factory interface. The pipeline "copy_pipeline" is running, and the activity "Copy address table" is still in the "Queued" state. The pipeline run ID is c0e1a57a-c29b-4f6c-b2d3-352226826098, and the status is "In progress". The activity table shows the same information as the first screenshot. The Windows taskbar at the bottom is visible, showing the date and time as 19-04-2024 at 04:25 PM.



The screenshot shows the Microsoft Azure Data Factory interface. A pipeline named 'copy_pipeline' is selected. The 'Activities' pane on the left lists various data movement options like 'Copy data' and 'Data flow'. The main workspace displays a single activity: 'Copy data' with a sub-activity 'Copy address table'. Below this, the 'Output' tab shows a table of pipeline run details. The table has one item: 'table' with status 'Succeeded', activity type 'Copy data', run start at '4/20/2024, 10:02:08 AM', duration '28s', integration runtime 'SHR', and activity ID '488bbe2'. The pipeline run ID is '5def82c6-6979-4521-a945-09e2da9da8bd'. The pipeline status is 'Succeeded'.



This screenshot shows the 'Publish all' dialog box overlaid on the Data Factory interface. The dialog is titled 'Publish all' and informs the user they are about to publish pending changes to the live environment. It lists 'Pending changes (4)' under three categories: Pipelines, Datasets, and Linked services. Under Pipelines, 'copy_pipeline' is listed as (New). Under Datasets, 'address' and 'address_Parquet' are listed as (New). Under Linked services, 'onpremssqlserver' is listed as (New). At the bottom of the dialog are 'Publish' and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory interface. A pipeline named 'copy_pipeline' is selected. The 'Activities' pane on the left lists various options like Move and transform, Synapse, Azure Data Explorer, etc. The main area displays a 'Copy data' activity with a green checkmark. Below it, the 'Output' tab shows a table of pipeline run details:

	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity ID
table	Succeeded	Copy data	4/20/2024, 10:02:08 AM	28s	SHIR		488bbe2

A message at the top right indicates 'Publishing completed'.

The screenshot shows the Microsoft Azure Storage Container interface for the 'bronze' container. The 'Overview' tab is selected. It shows a single blob named 'SalesLT.Address.parq...' with a size of 34.74 KiB. The table below lists all blobs in the container:

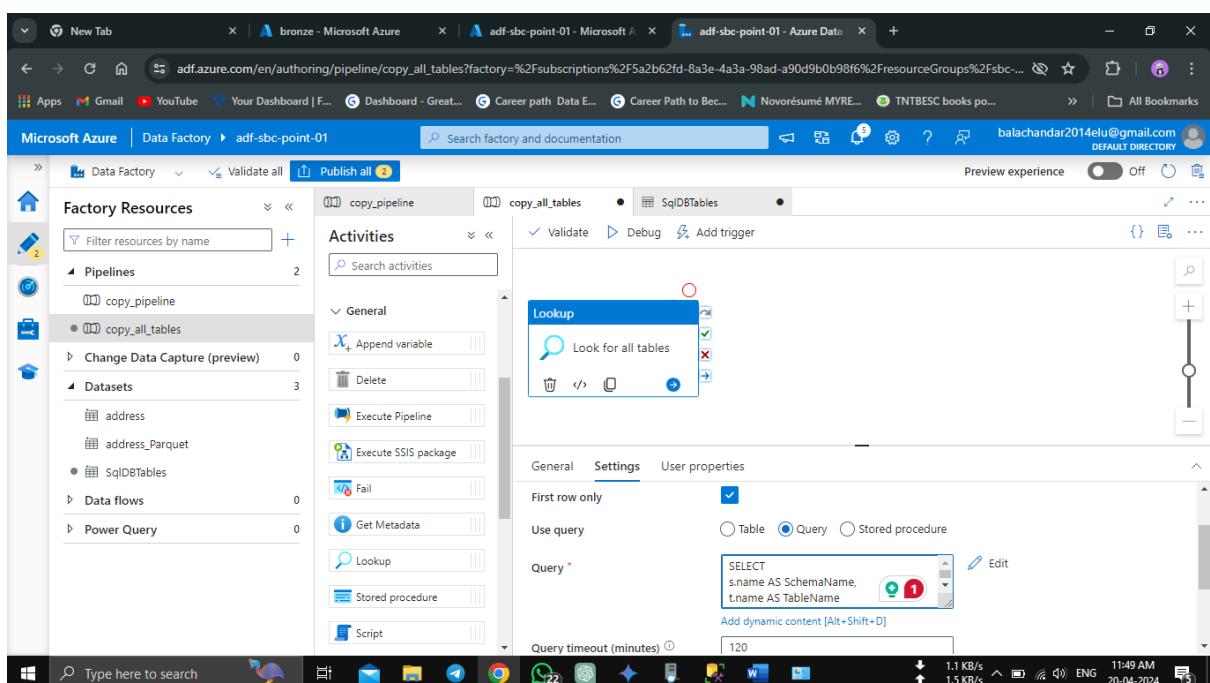
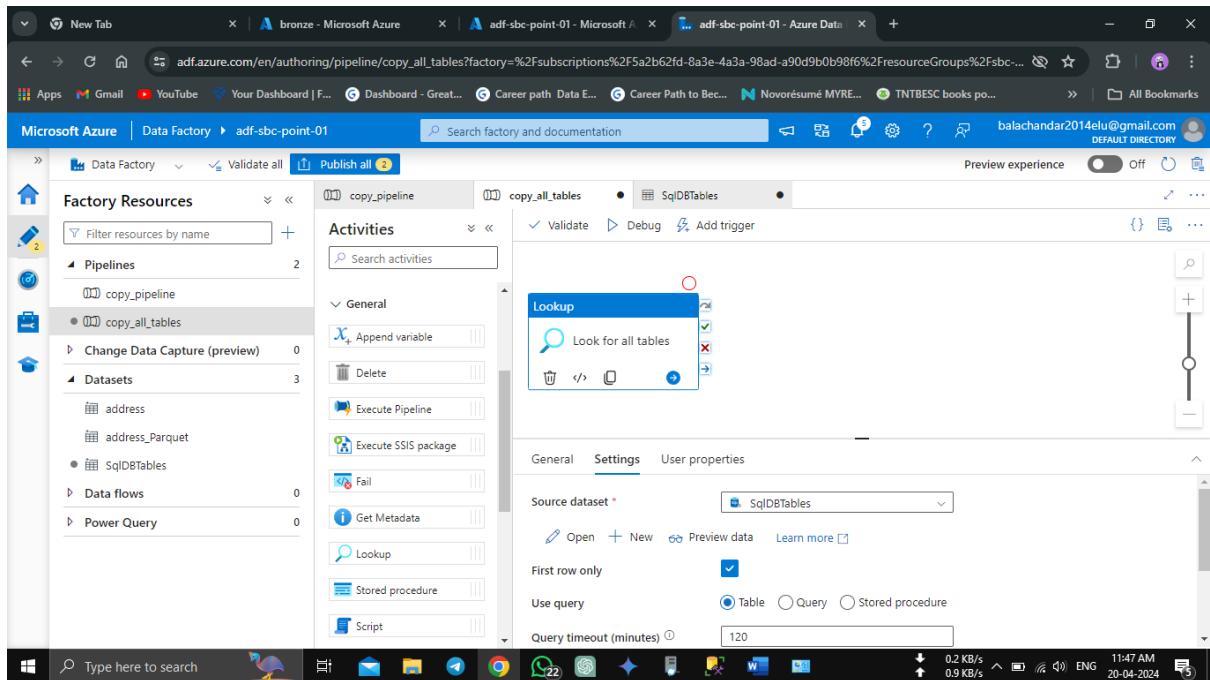
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
SalesLT.Address.parq...	20/04/2024, 10:02:34 ...	Hot (Inferred)		Block blob	34.74 KiB	Available



The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. The main area displays a pipeline named 'copy_pipeline' which contains an activity named 'copy_all_tables'. The 'Properties' panel on the right shows the activity's name as 'copy_all_tables' and its description as 'copy all tables'. The pipeline is currently in 'Preview experience' mode, which is turned off.

The screenshot shows Microsoft SQL Server Management Studio (SSMS) running a query titled 'GetSchema_SQLQuery1.sql'. The query retrieves table names from the 'SalesLT' schema in the 'AdventureWorksLT2017' database. The results are displayed in a table with columns 'SchemaName' and 'TableName', listing 10 rows. The SSMS interface includes an Object Explorer on the left and a toolbar at the bottom.

SchemaName	TableName
SalesLT	Address
SalesLT	Customer
SalesLT	CustomerAddress
SalesLT	Product
SalesLT	ProductCategory
SalesLT	ProductDescription
SalesLT	ProductModel
SalesLT	ProductModelProductDescription
SalesLT	SalesOrderDetail
SalesLT	SalesOrderHeader



Screenshot of the Microsoft Azure Data Factory preview interface showing the "copy_all_tables" pipeline.

Pipeline Details:

- Linked service:** onpremssqlserver
- Object:** copy_all_tables

Preview Data:

SchemaName	TableName
SalesLT	Address
SalesLT	Customer
SalesLT	CustomerAddress
SalesLT	Product
SalesLT	ProductCategory
SalesLT	ProductDescription
SalesLT	ProductModel
SalesLT	ProductModelProductDescription
SalesLT	SalesOrderDetail
SalesLT	SalesOrderHeader

Actions: Edit, Preview experience (Off), ...

Screenshot of the Microsoft Azure Data Factory pipeline editor showing the "copy_all_tables" pipeline.

Pipeline Activities:

- copy_pipeline
- copy_all_tables
- SqlDBTables

Activities Details:

- copy_all_tables Activity:** Validate, Cancel options, Add trigger
- Activities Submenu:** General, Append variable, Delete, Execute Pipeline, Execute SSIS package, Fail, Get Metadata, Lookup, Stored procedure, Script

Lookup Activity:

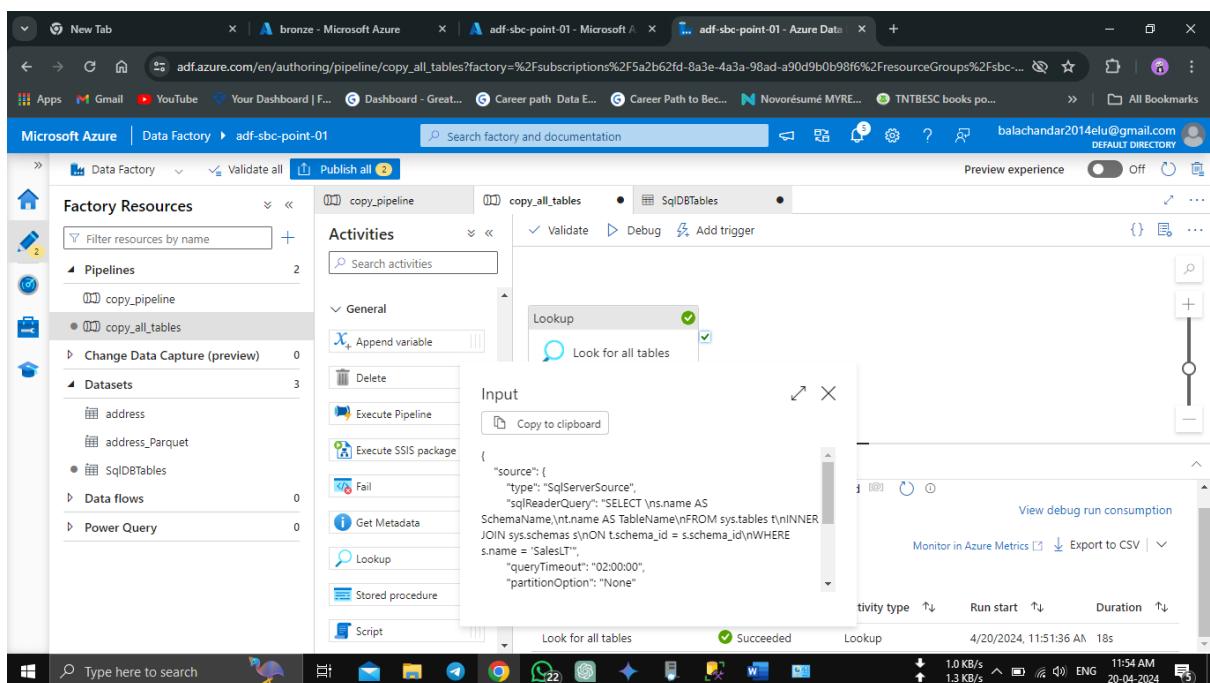
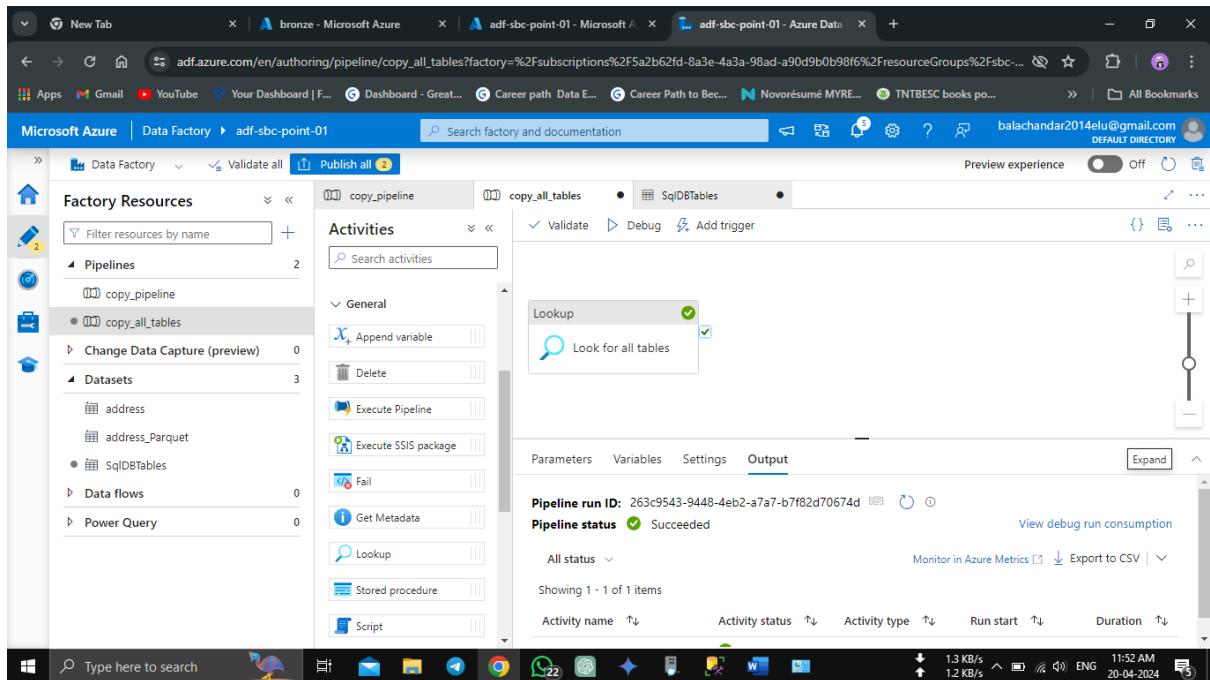
- Look for all tables

Pipeline Status: In progress

Metrics: Monitor in Azure Metrics, Export to CSV

Logs: Showing 0 - 0 of 0 items

Activity Headers: Activity name, Activity status, Activity type, Run start, Duration



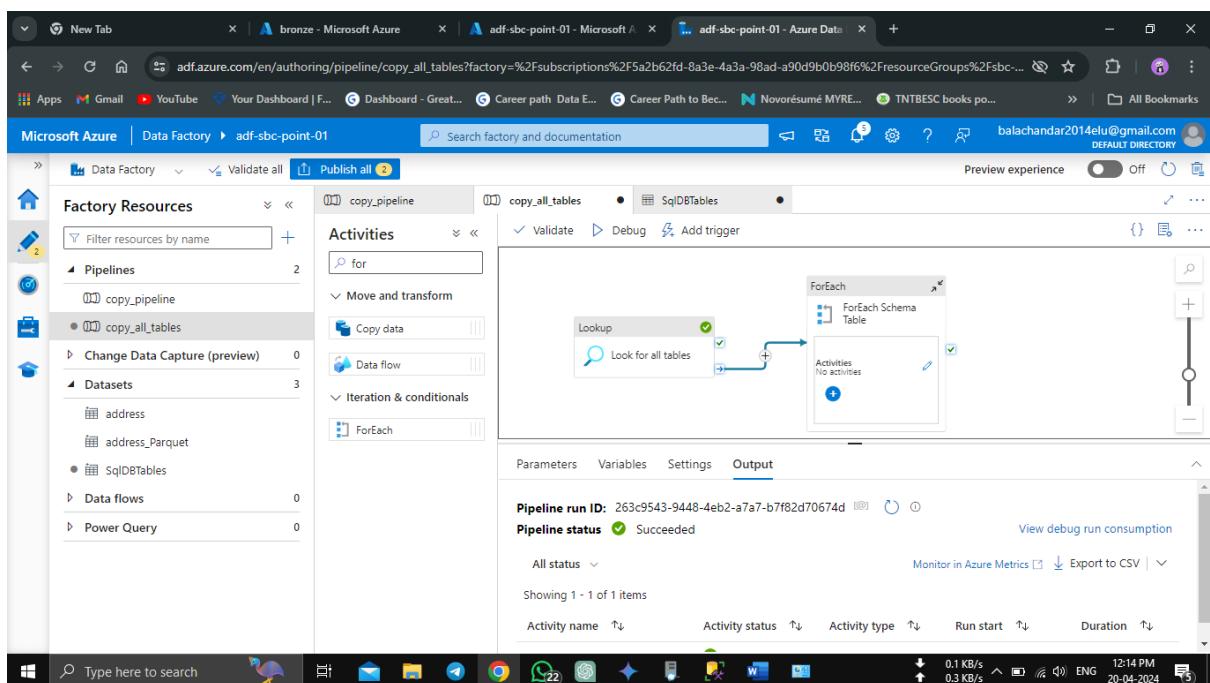
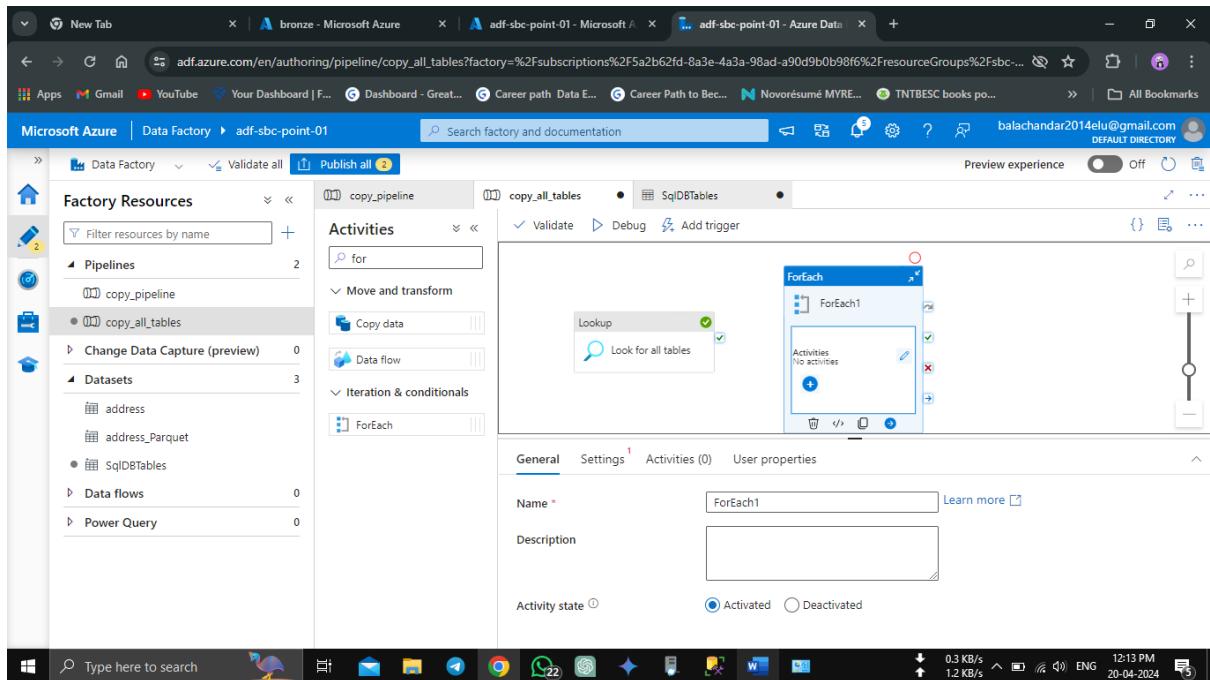
Use the Lookup activity ..

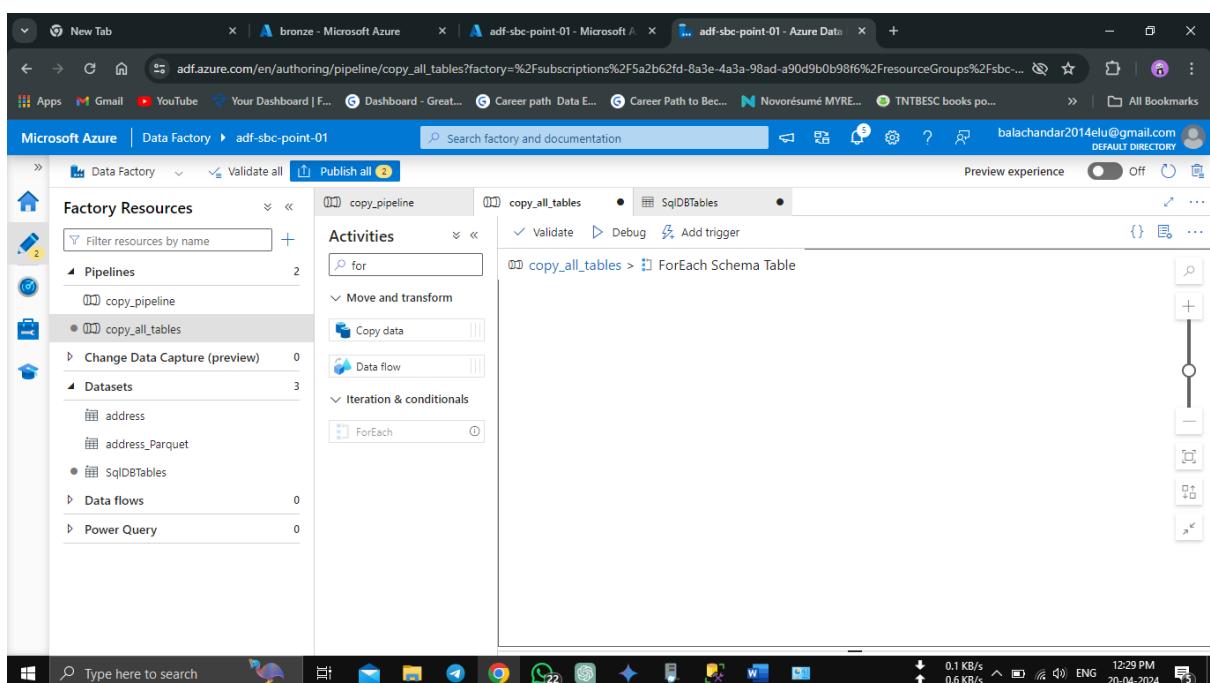
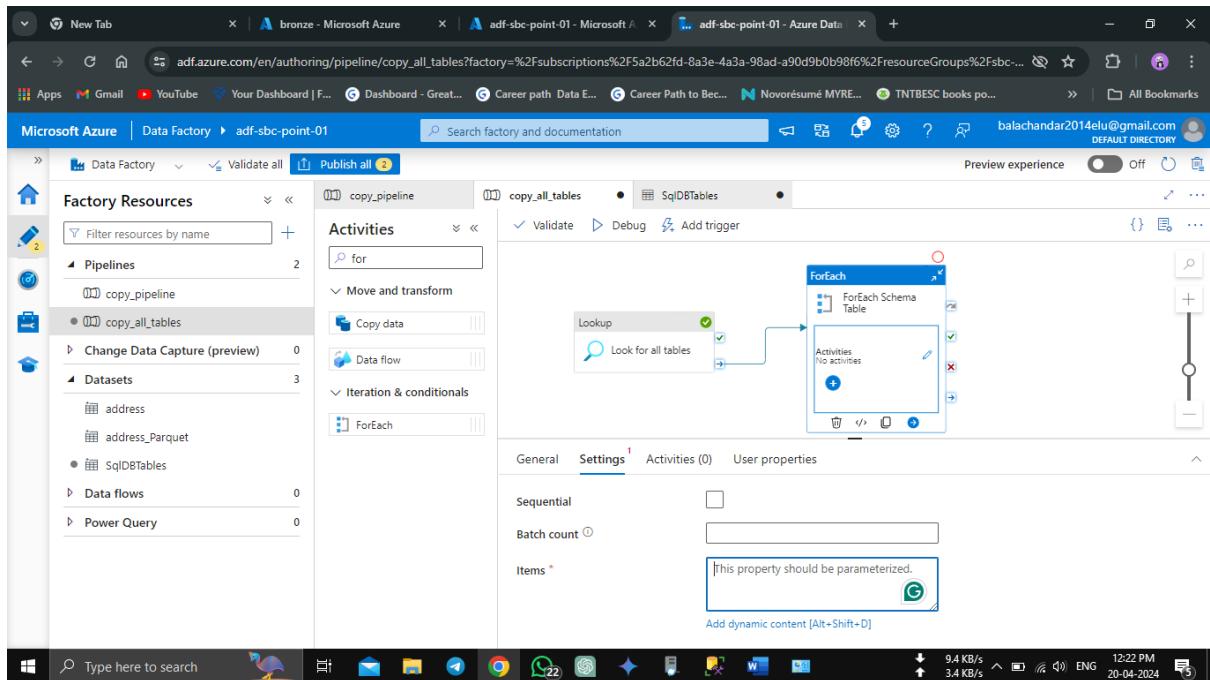
The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (copy_pipeline, copy_all_tables), 'Datasets' (address, address_Parquet, SqlDBTables), and 'Data flows'. The main workspace displays the 'copy_all_tables' pipeline. A 'Lookup' activity is selected, with its 'Output' pane expanded, showing the following JSON output:

```
{"count": 10, "value": [{"SchemaName": "SalesLT", "TableName": "Address"}, {"SchemaName": "SalesLT"}]}
```

The pipeline run details at the bottom indicate a successful run on 4/20/2024 at 11:51:36 AM, with a duration of 18s.

This screenshot shows the same Microsoft Azure Data Factory interface as the first one, but with a different configuration. The 'Activities' pane now shows a 'for' loop, which contains a 'Lookup' activity. The 'Pipeline status' is listed as 'Succeeded'. The pipeline run details at the bottom indicate a successful run on 4/20/2024 at 11:51:36 AM, with a duration of 18s.





The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and other resources. The main workspace displays a pipeline named 'copy_pipeline' containing a 'copy_all_tables' activity. The 'copy_all_tables' activity is configured to 'Copy data' from 'SqlDBTables' to 'SqlDBTables'. The 'Source' tab is selected, showing a 'Query' type source with the query:

```
SELECT * FROM [AdventureworksLT].[dbo].[Customer]
```

Other settings include a 'Query timeout (minutes)' of 120 and an 'Isolation level' of 'Select...'. The 'Sink' tab is also visible.

This screenshot is identical to the one above, but it includes a modal window titled 'Pipeline expression builder' over the bottom right of the workspace. The builder window contains a large text area for dynamic content and tabs for 'ForEach iterator', 'Activity outputs', 'Parameters', 'System variables', 'Functions', and 'Variables'. A search bar at the top of the builder window is empty.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (copy_pipeline, copy_all_tables), 'Datasets' (address, address_Parquet, SqlDBTables, SqlServercopy), and 'Data flows'. The main workspace displays the 'copy_all_tables' pipeline. Under 'Activities', a 'Copy data' task is selected, showing its configuration. The 'Sink dataset' tab is active, with the 'File path' set to 'bronze' and 'Linked service' set to 'AzureDataLakeStorage1'. A 'Set properties' dialog box is open, showing the 'Name' field as 'ParquetTables'. Other tabs in the dialog include 'General', 'Source', and 'Sink'. At the bottom right of the dialog are 'OK', 'Back', and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory interface. The 'Factory Resources' sidebar lists 'Pipelines' (copy_pipeline, copy_all_tables), 'Datasets' (address, address_Parquet, ParquetTables, SqlDBTables, SqlServercopy), and 'Data flows'. The main workspace displays the 'ParquetTables' dataset. The 'Properties' pane on the right shows the 'General' tab with 'Name' set to 'ParquetTables'. Other tabs include 'Related (1)' and 'Annotations'. Below the dataset name, there is a 'Parameters' section with two entries: 'schemaname' (String type, Default value: Value) and 'tablename' (String type, Default value: Value). The status bar at the bottom indicates '657.0 KB/s 14.0 KB/s' and the date '20-04-2024'.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and other resources. In the main workspace, a pipeline named 'copy_pipeline' is selected. A dataset named 'ParquetTables' is highlighted. The 'Properties' panel on the right shows the dataset's name as 'ParquetTables'. The 'Connection' tab indicates the connection is set to 'AzureDataLakeStorage1'. The 'Schema' tab shows the schema mapping, and the 'Parameters' tab displays parameters like 'bronze' and 'snappy'. The status bar at the bottom shows network activity and the date/time.

This screenshot shows the same Microsoft Azure Data Factory interface, but the pipeline 'copy_pipeline' has been renamed to 'copy_all_tables'. The 'Activities' pane is open, showing a 'ForEach' loop activity. Inside the loop, a 'Copy data' activity is configured to 'Copy Each Table'. The 'Source' tab of the 'Copy data' activity settings shows the 'Source dataset' as 'SqlServercopy' and the 'Query' as '@{concat('SELECT * FROM ', item().Sc...')}'. The status bar at the bottom shows network activity and the date/time.

The screenshot shows the Microsoft Azure Data Factory interface for publishing changes. On the left, the 'Factory Resources' sidebar lists Pipelines (copy_pipeline, copy_all_tables), Datasets (address, address_Parquet, ParquetTables, SqlDBTables, SqlServercopy), and other resources. The main workspace displays a pipeline named 'copy_pipeline' with an activity named 'copy_all_tables'. This activity is configured to 'Copy data' from a 'SqlDBTables' dataset to a 'ParquetTables' dataset. A 'Pending changes' table shows four items: 'copy_all_tables' (New), 'SqlDBTables' (New), 'SqlServercopy' (New), and 'ParquetTables' (New). A 'Publish' button is visible at the bottom right of the activity configuration area.

This screenshot shows the detailed configuration of the 'copy_all_tables' activity within the 'copy_pipeline' pipeline. The 'Source' tab is selected, showing the 'SqlServercopy' dataset as the source. The 'Query' field contains the dynamic SQL query: '@(concat('SELECT * FROM ', item().Sc...'))'. Other tabs include General, Sink, Mapping, Settings, and User properties. The 'Activities' pane on the right shows the pipeline structure: copy_all_tables > ForEach Schema Table. The 'ForEach Schema Table' activity is expanded, showing a 'Copy data' sub-activity with 'Copy Each Table' selected. The 'Preview experience' toggle switch is turned off.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and other components. In the center, the 'copy_pipeline' pipeline is selected, showing its activities. The 'copy_all_tables' activity is currently being edited. The 'Source' tab is active, showing the configuration for copying data from a SQL Server database. The 'Query' field contains the dynamic SQL query: `@{concat('SELECT * FROM ', item().SchemaName, '.', item().Name)}`. The 'Query timeout (minutes)' is set to 120.

The screenshot shows the Microsoft Azure Storage Container 'bronze' listing blobs. The container has a single directory named '[-]'. Inside this directory, there are several blobs with names like 'Address', 'Customer', 'CustomerAddress', 'Product', 'ProductCategory', 'ProductDescription', 'ProductModel', and 'ProductModelPro...'. The table below provides a detailed view of these blobs:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[-]					-	---
Address					-	---
Customer					-	---
CustomerAddress					-	---
Product					-	---
ProductCategory					-	---
ProductDescription					-	---
ProductModel					-	---
ProductModelPro...					-	---

Microsoft Azure | Overview

Deployment

Search

Overview

Inputs

Outputs

Template

Deployment is in progress

Deployment name : sbc-data-engineering-project-02_dbw-sbc-point-01 Start time : 21/4/2024, 8:52:17 am

Subscription : Free Trial Correlation ID : 868735c7-5d94-459e-ab8a-e278e11f0...

Resource group : sbc-data-engineering-project-02

Deployment details

Resource	Type	Status	Operation details
dbw-sbc-point-01	Azure Databricks Service	Created	Operation details

Give feedback

Tell us about your experience with deployment

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 >

1.5 KB/s
4.1 KB/s
08:53 AM
21-04-2024

Microsoft Azure | Overview

dbw-sbc-point-01

Azure Databricks Service

Search

Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Virtual Network Peering

Encryption

Networking

Security & compliance

Properties

Locks

Monitoring

Essentials

Status : Active

Managed Resource Group : sbc

Resource group : sbc-data-engineering-project-02

URL : <https://adb-1840545276150268.azuredatabricks.net>

Location : East US

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

Subscription : Free Trial

Subscription ID : 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6

Tags (edit) : Add tags

Launch Workspace

1.6 KB/s
1.8 KB/s
08:55 AM
21-04-2024

Sign In to Databricks

Sign in using Azure Active Directory Single Sign On.

Signing you in

Contact your site administrator to request access.

Type here to search

147 KB/s
3.2 KB/s

08:55 AM
21-04-2024

Get started

Import and transform data

Create a table by uploading local files, or create a pipeline for continuous data ingestion and transformation.

Create table Create pipeline

Notebook

Create a new notebook for data analysis, transformation, and machine learning.

Create notebook

SQL query editor

Create a new query and explore your data in the SQL Editor.

Create query

AutoML

Accelerate the training of ML models for efficient discovery and iteration.

Start AutoML

Pick up where you left off

No recent items

Start exploring and your recently viewed items will show up here.

Popular

No popular items

Start exploring and popular items in your workspace will show up here.

Type here to search

0.6 KB/s
1.7 KB/s

08:56 AM
21-04-2024

Do the Data transformation work in Data Bricks..

The screenshot shows the Databricks Compute interface for creating a new cluster named "data_transformation".

Policy: Unrestricted

Access mode: Single user access (Single user selected)

Runtime: Photon, Standard_DS3_v2 (1.5 DBU/h)

Summary:

- 1 Driver, 14 GB Memory, 4 Cores
- Runtime: 13.3.x-scala2.12
- Photon, Standard_DS3_v2, 1.5 DBU/h

Performance:

Databricks runtime version: 13.3 LTS (Scala 2.12, Spark 3.4.1)

Use Photon Acceleration (checked)

Node type: Standard_DS3_v2 (14 GB Memory, 4 Cores)

Termination settings: Terminate after 15 minutes of inactivity

Buttons: Create compute, Cancel

The screenshot shows the Databricks Compute interface for creating a new cluster named "data_transformation".

Azure Data Lake Storage credential passthrough: Enable credential passthrough for user-level data access (checked)

Spark config:

```
spark.master local["*"], 4  
spark.databricks.cluster.profile singleNode
```

Environment variables:

```
PYSPARK_PYTHON=/databricks/python3/bin/python3
```

Buttons: Create compute, Cancel

Microsoft Azure | sбcdatalakegen02 | Access Control (IAM)

Storage account

Search

Add Download role assignments Edit columns Refresh Remove Feedback

View assignments

All Job function (3) Privileged (0)

Search by name or email Type : All Role : All Scope : All scopes Group by : Role

Name	Type	Role	Scope	Condition
Storage Blob Data Contributor (3)				
dbmanagedidentity /subscriptions/f9994c66...	User-assigned Managed Identity	Storage Blob Data Contributor	This resource	Add
adf-sbc-point-01 /subscriptions/f9994c66...	Data Factory	Storage Blob Data Contributor	This resource	Add
unity-catalog-access-conn App	App	Storage Blob Data Contributor	This resource	Add

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Data storage Containers File shares Queues Tables

Type here to search

Windows Start button Taskbar icons Network connection speed: 146.0 KB/s (3.9 KB/s) Date: 09:11 AM (24-04-2024)

sбcdatalakegen02 - Microsoft | Compute - Databricks | adf-sbc-point-01 - Azure Data | New Tab

Microsoft Azure | databricks

Search data, notebooks, recents, and more... CTRL + P

New Workspace Recents Catalog Workflows Compute SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables

https://adb-1973240855609375.15.azuredatabricks.net/compute?o=1973240855609375

Compute

All-purpose compute Job compute SQL warehouses Vector Search Preview Pools Policies

Filter compute you have... Created by Only pinned Create with Personal Compute Create compute

State	Name	Policy	Runtime	Active m...	Active co...	Active D...	Source	Creator	Notebooks	...
Running	data_transformation	-	13.3	-	-	1.5	UI	balachandar202...	-	⋮

Previous Next 20 / page

Type here to search

Windows Start button Taskbar icons Network connection speed: 7.7 KB/s (11.5 KB/s) Date: 09:13 AM (24-04-2024)

The screenshot shows the Microsoft Edge browser interface with the following details:

- Address Bar:** adb-1973240855609375.15.azuredatabricks.net/browse?o=1973240855609375
- Tab Bar:** sbedatalakegen02 - Microsoft, Home - Workspace - Databricks, adf-sbc-point-01 - Azure Data, New Tab.
- Search Bar:** Search data, notebooks, recents, and more... (CTRL + P)
- Toolbar:** Apps, Gmail, YouTube, Your Dashboard | F..., Dashboard - Great..., Career path Data E..., Career Path to Bec..., Novorésumé MYRE..., TNTBESC books po..., All Bookmarks, dbw-sbc-point-01, Share, Create.
- Sidebar (Left):** Microsoft Azure | databricks. Options include: New, Workspace (selected), Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables.
- Main Content Area:** Workspace > Users > balachandar2024elu@gmail.com. A table lists the user's workspace. The table has columns: Name, Type, Owner, Created at. One row is shown: Name is 'User Shared', Type is 'Folder', Owner is 'balachandar2024elu@gmail.com', and Created at is '2024-04-24 09:14 AM'. Below the table, it says 'This folder is empty.'
- Bottom Status Bar:** Type here to search, followed by a Windows taskbar with various icons (File Explorer, Mail, Task View, Google Chrome, File Manager, Taskbar icons).

The screenshot shows the Microsoft Edge browser interface with the following details:

- Address Bar:** adb-1973240855609375.15.azuredatabricks.net/?o=1973240855609375#notebook/980621189977273/command/980621189977274
- Tab Bar:** sbedatalakegen02 - Microsoft, Stroagmount - Databricks, adf-sbc-point-01 - Azure Data, New Tab.
- Search Bar:** Search data, notebooks, recents, and more... (CTRL + P)
- Toolbar:** Apps, Gmail, YouTube, Your Dashboard | F..., Dashboard - Great..., Career path Data E..., Career Path to Bec..., Novorésumé MYRE..., TNTBESC books po..., All Bookmarks, dbw-sbc-point-01, Share.
- Sidebar (Left):** Microsoft Azure | databricks. Options include: New, Workspace (selected), Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables.
- Main Content Area:** Stroagmount Python. The notebook interface shows a single code cell with the placeholder text 'start typing or generate with AI (Ctrl + I)...'. The cell is set to Python. Navigation buttons for previous/next cell, run cell, and cell controls are visible.
- Bottom Status Bar:** Type here to search, followed by a Windows taskbar with various icons (File Explorer, Mail, Task View, Google Chrome, File Manager, Taskbar icons).

sbcdatalakegen02 - Microsoft | Stroagemount - Databricks | adf-sbc-point-01 - Azure Data | Access Azure Data Lake Storage | +

ad-1973240855609375.15.azuredatabricks.net/?o=1973240855609375#notebook/980621189977273/command/980621189977274

Apps Gmail YouTube Your Dashboard | F... Dashboard - Great... Career path Data E... Career Path to Bec... Novorésumé MYRE... TNTBESC books po... All Bookmarks

Microsoft Azure | databricks

Search data, notebooks, recents, and more... CTRL + P

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables

Stroagemount Python

File Edit View Run Help Last edit was 2 minutes ago New cell UI: ON

Run all Starting Schedule Share

1

```
configs = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://bronze@sbcdatalakegen02.dfs.core.windows.net/",
    mount_point = "/mnt/bronze",
    extra_configs = configs)
```

[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts

Python

0.1 KB/s 0.3 KB/s ENG 09:41 AM 24-04-2024

This screenshot shows a Databricks notebook titled 'Stroagemount' in Python mode. The notebook contains a single cell with the following code:configs = {
 "fs.azure.account.auth.type": "CustomAccessToken",
 "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")}

Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
 source = "abfss://bronze@sbcdatalakegen02.dfs.core.windows.net/",
 mount_point = "/mnt/bronze",
 extra_configs = configs)

The cell has been run successfully, as indicated by the green checkmark icon and the output 'True'. The status bar at the bottom right shows network activity (0.1 KB/s, 0.3 KB/s) and the date and time (09:41 AM, 24-04-2024).

Default Directory - M | sbcdatalakegen02 - | Stroagemount - Databricks | Compute - Databrick | adf-sbc-point-01 - A | Access Azure Data L | +

ad-1973240855609375.15.azuredatabricks.net/?o=1973240855609375#notebook/980621189977273/command/980621189977276

Apps Gmail YouTube Your Dashboard | F... Dashboard - Great... Career path Data E... Career Path to Bec... Novorésumé MYRE... TNTBESC books po... All Bookmarks

Microsoft Azure | databricks

Search data, notebooks, recents, and more... CTRL + P

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables

Stroagemount Python

File Edit View Run Help Last edit was 17 minutes ago New cell UI: ON

Run all data_transformation Schedule Share

1 09:46 AM (14s)

```
# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://bronze@sbcdatalakegen02.dfs.core.windows.net/",
    mount_point = "/mnt/bronze",
    extra_configs = config)
```

True

2 Just now (1s)

```
dbutils.fs.ls("/mnt/bronze")
```

[FileInfo(path='dbfs:/mnt/bronze/SalesLT/', name='SalesLT', size=0, modificationTime=1713927831000)]

3

```
Start typing or generate with AI (Ctrl + I)... Python
```

1.9 KB/s 2.1 KB/s ENG 10:05 AM 24-04-2024

This screenshot shows the same Databricks notebook after the first cell has been run. The cell output 'True' is visible. Below it, a second cell has been run, showing the output of the 'dbutils.fs.ls()' command. The status bar at the bottom right shows network activity (1.9 KB/s, 2.1 KB/s) and the date and time (10:05 AM, 24-04-2024).

```
extra_configs = configs
True

dbutils.fs.ls("/mnt/bronze/SalesLT")

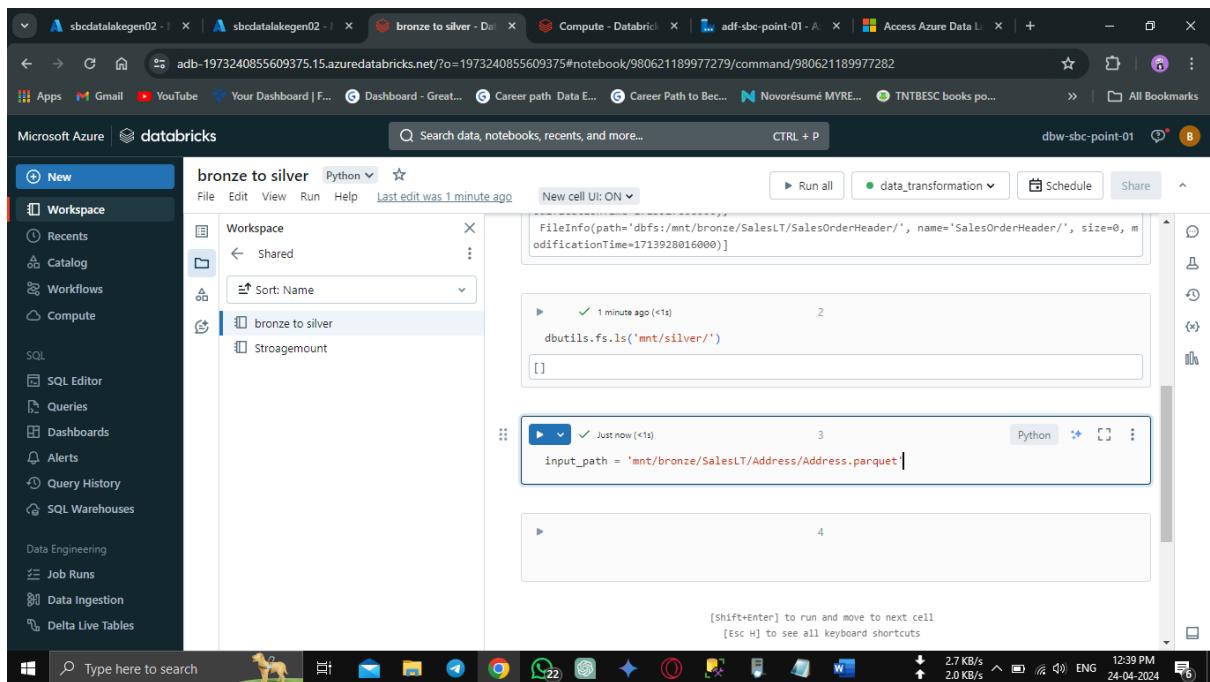
[FileInfo(path='dbfs:/mnt/bronze/SalesLT/Address', name='Address', size=0, modificationTime=1713928015000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/Customer', name='Customer', size=0, modificationTime=1713927949000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/CustomerAddress', name='CustomerAddress', size=0, modificationTime=1713927866000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/Product', name='Product', size=0, modificationTime=1713927934000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductCategory', name='ProductCategory', size=0, modificationTime=1713927873000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductDescription', name='ProductDescription', size=0, modificationTime=171392783100),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductModel', name='ProductModel', size=0, modificationTime=1713927968000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductModelDescription', name='ProductModelProductDescription', size=0, modificationTime=1713927948000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/SalesOrderDetail', name='SalesOrderDetail', size=0, modificationTime=1713927866000),
 FileInfo(path='dbfs:/mnt/bronze/SalesLT/SalesOrderHeader', name='SalesOrderHeader', size=0, modificationTime=1713928016000)]
```

sbcdatalakegen02 | Containers

Name	Last modified	Anonymous access level	Lease state
Slogs	23/4/2024, 6:01:38 pm	Private	Available
bronze	23/4/2024, 6:03:03 pm	Private	Available
gold	24/4/2024, 10:11:45 am	Private	Available
silver	24/4/2024, 10:10:58 am	Private	Available

```
bronze to silver Python 1
File Edit View Run Help Last edit was 5 minutes ago New cell UI: ON
Run all data_transformation Schedule Share
Workspace Shared Sort: Name
bronze to silver Stroagmount
[FileInfo(path='dbfs:/mnt/bronze/SalesLT/Address/', name='Address', size=0, modificationTime=13928015000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/Customer/', name='Customer', size=0, modificationTime=1713927949000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/CustomerAddress/', name='CustomerAddress', size=0, modificationTime=1713927866000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/Product/', name='Product', size=0, modificationTime=1713927934000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductCategory/', name='ProductCategory', size=0, modificationTime=1713927873000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductDescription/', name='ProductDescription', size=0, modificationTime=1713927831000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductModel/', name='ProductModel', size=0, modificationTime=1713927968000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/ProductModelDescription/', name='ProductModelDescription', size=0, modificationTime=1713927948000), FileInfo(path='dbfs:/mnt/bronze/SalesLT/SalesOrderDetail/', name='SalesOrderDetail', size=0, modificationTime=1713927866000)]
```

```
bronze to silver Python 2
File Edit View Run Help Last edit was 6 minutes ago New cell UI: ON
Run all data_transformation Schedule Share
Workspace Shared Sort: Name
bronze to silver Stroagmount
[FileInfo(path='dbfs:/mnt/silver/')] 3
Start typing or generate with AI (Ctrl + I)...
[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts
```



bronze to silver Python

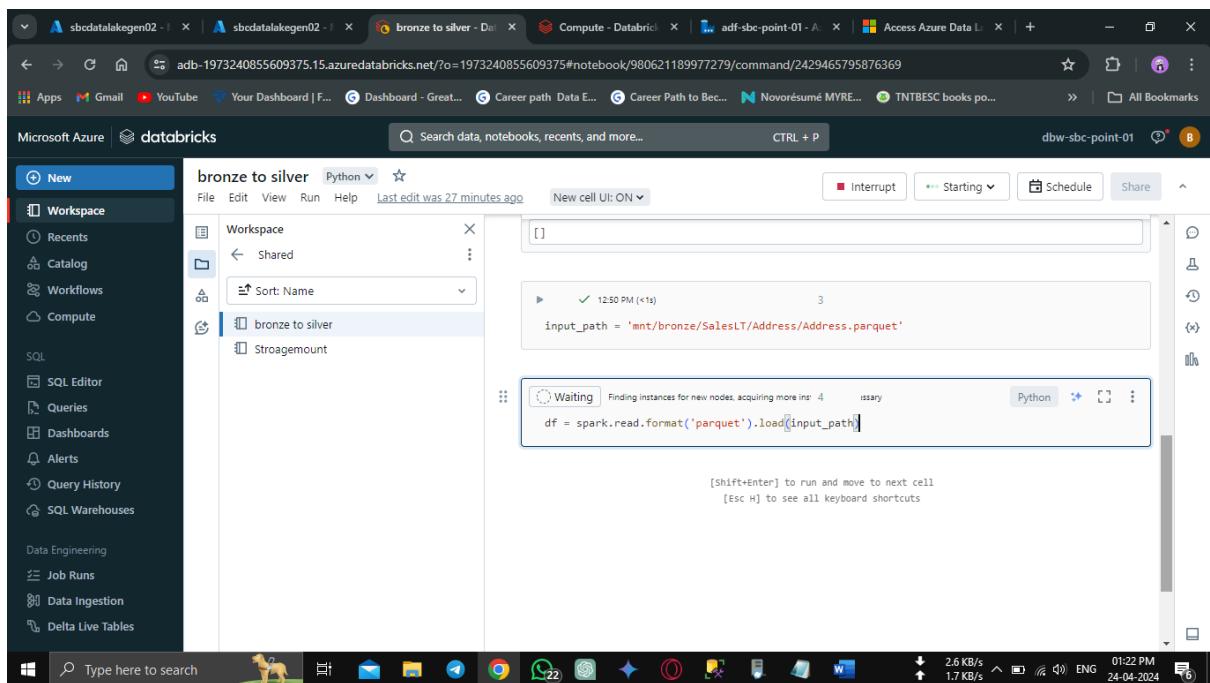
```
FileInfo(path='dbfs:/mnt/bronze/SalesLT/SalesOrderHeader/', name='SalesOrderHeader', size=0, modificationTime=17139280160000)
```

```
1 minute ago (<1s) 2
dbutils.fs.ls('mnt/silver/')

[]
```

```
Just now (<1s) 3
input_path = 'mnt/bronze/SalesLT/Address/Address.parquet'
```

```
4
```



bronze to silver Python

```
[]
```

```
12:50 PM (<1s) 3
input_path = 'mnt/bronze/SalesLT/Address/Address.parquet'
```

```
Waiting Finding instances for new nodes, acquiring more instances... 4
df = spark.read.format('parquet').load(input_path)
```

```
[Shift+Enter] to run and move to next cell  
[Esc H] to see all keyboard shortcuts
```

The screenshot shows a Microsoft Azure Databricks workspace. On the left, a sidebar lists various options like Workspace, Recents, Catalog, Workflows, Compute, SQL, and Data Engineering. The main area displays a Python notebook titled "bronze to silver". The notebook has three cells:

- Cell 1: An empty cell with the status "1 minute ago (<1s)".
- Cell 2: A code cell with the status "Just now (<1s)". It contains the command `input_path = '/mnt/bronze/SalesLT/Address/Address.parquet'`.
- Cell 3: A code cell with the status "Just now (3s)". It contains the command `df = spark.read.format('parquet').load(input_path)` followed by a note "(1) Spark Jobs". Below this, it shows the schema of the DataFrame: `df: pyspark.sql.dataframe.DataFrame` with fields: AddressID: integer, AddressLine1: string, AddressLine2: string, City: string, StateProvince: string, CountryRegion: string, PostalCode: string, rowguid: string, ModifiedDate: timestamp.

The bottom of the screen shows a Windows taskbar with icons for File Explorer, Mail, Task View, and others, along with system status information.

This screenshot continues the same Microsoft Azure Databricks session. The notebook "bronze to silver" is still open. The third cell from the previous screenshot is now expanded, showing the detailed schema of the DataFrame `df`:

```
df: pyspark.sql.dataframe.DataFrame
AddressID: integer
AddressLine1: string
AddressLine2: string
City: string
StateProvince: string
CountryRegion: string
PostalCode: string
rowguid: string
ModifiedDate: timestamp
```

A new cell, labeled "5", is visible below it, showing the command `display(df)` with a "Waiting" status. The bottom of the screen shows the Windows taskbar and system status.

bronze to silver Python

```
display(df)
```

	AddressID	AddressLine1	AddressLine2	City	State
1	9	8713 Yosemite Ct.	null	Bothell	Washington
2	11	1318 Lasalle Street	null	Bothell	Washington
3	25	9178 Jumping St.	null	Dallas	Texas
4	28	9228 Via Del Sol	null	Phoenix	Arizona
5	32	26910 Indela Road	null	Montreal	Quebec
6	185	2681 Eagle Peak	null	Bellevue	Washington
7	297	7943 Walnut Ave	null	Renton	Washington
8	445	6388 Lake City Way	null	Burnaby	British Columbia
9	446	5260 Free Street	null	Toronto	Ontario
10	447	22580 Free Street	null	Toronto	Ontario
11	448	2575 Bloor Street East	null	Toronto	Ontario
12	449	Station E	null	Chalk River	Ontario

```
from pyspark.sql.functions import from_utc_timestamp,date_format
from pyspark.sql.types import TimestampType

df = df.withColumn("ModifiedDate", date_format(from_utc_timestamp(df["ModifiedDate"].cast(TimestampType()), "UST"), "yyyy-MM-dd"))
```

[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts

bronze to silver Python

```
from pyspark.sql.functions import from_utc_timestamp, date_format
from pyspark.sql.types import TimestampType

df = df.withColumn("ModifiedDate", date_format(from_utc_timestamp(df["ModifiedDate"].cast(TimestampType()), "UTC"), "yyyy-MM-dd"))

df: pyspark.sql.dataframe.DataFrame = [AddressID: integer, AddressLine1: string ... 7 more fields]
```

bronze to silver Python

```
df.show()
```

(1) Spark Jobs

AddressID	AddressLine1	AddressLine2	City	StateProvince	CountryRegion	PostalCode	rowguid	ModifiedDate
9	8713 Yosemite Ct.		Bothell	Washington	United States	98011-268a	f621-76d7-4c78-9441-144fd139821a	2006-07-01

```
display(df)
```

(1) Spark Jobs

tryRegion	A ^B PostalCode	A ^B rowguid	A ^B ModifiedDate
1 tates	98011	268a	2006-07-01
2 tates	98011	b3303-ac	2007-04-01
3 tates	75201	c8df3bd9-48f0-4654-a8dd-14a67a84d3c6	2006-09-01
4 tates	85004	12ae5ee1-fc3e-468b-9b92-3b970b169774	2005-09-01
5 H1Y 2H5	84a95f62-3ae8-4e7e-bbd5-5a6f00cd982d	2006-08-01	
6 tates	98004	7bcc442-2268-46cc-8472-14c44c14e98c	2006-09-01
7 tates	98055	52410d84-2778-4b1d-a599-95746625ce6d	2006-08-01
8 VSA 3A6	5357225-9133-4ab0-a065-10cf35416e	2006-09-01	
9 M4B 1V7	801a1cf2-5125-48bb-aa84-ccb2ec7c4	2005-08-01	
10 M4B 1V7	88cee379-dbb8-433b-bb4e-a35e09435500	2006-08-01	
11 M4B 1V6	2df6d0ad-0926-4f54-a450-9b1083150cbf	2007-08-01	
12 K0J 1J0	Bb5a7729-cb75-4303-a607-799793bd94f	2005-08-01	

bronze to silver Python

```
from pyspark.sql.functions import from_utc_timestamp, date_format
from pyspark.sql.types import TimestampType

df = df.withColumn("ModifiedDate", date_format(from_utc_timestamp(df["ModifiedDate"].cast(TimestampType()), "UTC"), "yyyy-MM-dd"))

df.show()
```

df: pyspark.sql.dataframe.DataFrame = [AddressID: integer, AddressLine1: string ... 7 more fields]

(1) Spark Jobs

AddressID	AddressLine1	AddressLine2	City	StateProvince	CountryRegion	PostalCode	rowguid	ModifiedDate
9	8713 Yosemite Ct.	NULL	Bothell	Washington	United States	98011	268af621-76d7-4c7...	2006-07-01

bronze to silver Python

```
display(df)
```

(1) Spark Jobs

	StateProvince	CountryRegion	PostalCode	rowguid	ModifiedDate
1	Washington	United States	98011	268af621-76d7-4c78-9441-144fd139821a	2006-07-01
2	Washington	United States	98011	981b3303-aca2-49c7-9a96-fb670785b269	2007-04-01
3	Texas	United States	75201	c8df3bd9-48f0-4654-a8dd-14a67a84d3c6	2006-09-01
4	Arizona	United States	85004	12ae5ee1-fc3e-468b-9b92-3b970b169774	2005-09-01
5	Quebec	Canada	H1Y 2H5	84a95f62-3ae8-4e7e-bbd5-5a6f00cd982d	2006-08-01
6	Washington	United States	98004	7bccf442-2268-46cc-8472-14c44c14e98c	2006-09-01
7	Washington	United States	98055	52410da4-2778-4b1d-a599-95746625ce6d	2006-08-01
8	British Columbia	Canada	V5A 3A6	53572d25-9133-448b-a065-102ff5416ee	2006-09-01
9	Ontario	Canada	M4B 1V7	801a1dfc-5125-486b-aa84-ccb2ec57ca4	2005-08-01
10	Ontario	Canada	M4B 1V7	88cee379-dbb8-433b-b84e-a35e09435500	2006-08-01
11	Ontario	Canada	M4B 1V6	2df6d0ad-0926-4f54-a450-9b1083150cf0	2007-08-01
12	Ontario	Canada	K0J 1J0	8b5a7729-cb75-4303-a607-7f9793b4d94f	2005-08-01

bronze to silver Python

File Edit View Run Help Last edit was 8 minutes ago New cell UI: ON

(1) Spark Jobs

_sqldf: pyspark.sql.dataframe.DataFrame = [column1: integer]

Table +

1 column1

1 1

1 row | 0.46 seconds runtime Refreshed now

This result is stored as PySpark data frame _sqldf and in the IPython output cache as Out[33]. Learn more

Start typing or generate with AI (Ctrl + I)... 10 Python

Type here to search

Run all cells in this notebook.

bronze to silver Python

File Edit View Run Help Last edit was 1 minute ago New cell UI: ON

Doing Transformation for all tables

Just now (<1s)

```
table_name =[]  
for i in dbutils.fs.ls('mnt/bronze/SalesLT/'):  
    table_name.append(i.name.split('/')[0])
```

Just now (<1s)

```
table_name
```

['Address', 'Customer', 'CustomerAddress', 'Product', 'ProductCategory', 'ProductDescription', 'ProductModel', 'ProductModelProductDescription', 'SalesOrderDetail', 'SalesOrderHeader']

Run all cells in this notebook.

Type here to search

```
bronze to silver Python 13
from pyspark.sql.functions import from_utc_timestamp, date_format
from pyspark.sql.types import TimestampType

for i in table_name:
    path='/mnt/bronze/SalesLT/' + i + '/' + i + '.parquet'
    df = spark.read.format('parquet').load(path)
    column = df.columns

    for col in column:
        if "Date" in col or "date" in col:
            df = df.withColumn(col, date_format(from_utc_timestamp(df[col].cast(TimestampType()),"UTC"),"yyyy-MM-dd"))

    output_path = '/mnt/silver/SalesLT/' + i + '/'
    df.write.format('delta').mode("overwrite").save(output_path)
```

```
bronze to silver Python 13
from pyspark.sql.functions import from_utc_timestamp, date_format
from pyspark.sql.types import TimestampType

for i in table_name:
    path='/mnt/bronze/SalesLT/' + i + '/' + i + '.parquet'
    df = spark.read.format('parquet').load(path)
    column = df.columns

    for col in column:
        if "Date" in col or "date" in col:
            df = df.withColumn(col, date_format(from_utc_timestamp(df[col].cast(TimestampType()),"UTC"),"yyyy-MM-dd"))

    output_path = '/mnt/silver/SalesLT/' + i + '/'
    df.write.format('delta').mode("overwrite").save(output_path)
```

```
05:55 PM (39s) 13
from pyspark.sql.functions import from_utc_timestamp, date_format
from pyspark.sql.types import TimestampType

for i in table_name:
    path='/mnt/bronze/SalesLT/' + i + '/' + i + '.parquet'
    df = spark.read.format('parquet').load(path)
    column = df.columns

    for col in column:
        if "Date" in col or "date" in col:
            df = df.withColumn(col, date_format(from_utc_timestamp(df[col].cast(TimestampType()),"UTC"),"yyyy-MM-dd"))

    output_path = '/mnt/silver/SalesLT/' + i + '/'
    df.write.format('delta').mode("overwrite").save(output_path)
```

```
(50) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [SalesOrderID: integer, RevisionNumber: integer ... 20 more fields]
```

```
2 minutes ago (<1s) 14
display(df)
```

Screenshot of a Microsoft Azure Databricks workspace showing a Python notebook titled "bronze to silver".

The notebook displays the following code and output:

```
bronze to silver Python
File Edit View Run Help Last edit was 4 minutes ago New cell UI: ON
▶ Run all data_transformation Schedule Share
bronze to silver 3 minutes ago (<1s)
display(df)
(1) Spark Jobs
Table + 14
New result table: ON Search
SalesOrderID RevisionNumber OrderDate DueDate ShipDate Status Online
1 71774 2 2008-06-01 2008-06-13 2008-06-08 5 false
2 71776 2 2008-06-01 2008-06-13 2008-06-08 5 false
3 71780 2 2008-06-01 2008-06-13 2008-06-08 5 false
4 71782 2 2008-06-01 2008-06-13 2008-06-08 5 false
5 71783 2 2008-06-01 2008-06-13 2008-06-08 5 false
6 71784 2 2008-06-01 2008-06-13 2008-06-08 5 false
7 71796 2 2008-06-01 2008-06-13 2008-06-08 5 false
8 71797 2 2008-06-01 2008-06-13 2008-06-08 5 false
9 71815 2 2008-06-01 2008-06-13 2008-06-08 5 false
10 71816 2 2008-06-01 2008-06-13 2008-06-08 5 false
11 71831 2 2008-06-01 2008-06-13 2008-06-08 5 false
12 71832 2 2008-06-01 2008-06-13 2008-06-08 5 false
```

The notebook interface includes a sidebar with various Databricks services like Workspace, Recents, Catalog, Workflows, Compute, SQL, and Data Engineering. The status bar at the bottom shows network speed (1.0 KB/s), battery level (1.8 KB/s), and system time (06:03 PM, 24-04-2024).

silver to gold - Databricks

File Edit View Run Help Last edit was 1 minute ago New cell UI: ON

1
`dbutils.fs.ls('mnt/silver/SalesLT')`
`[FileInfo(path='dbfs:/mnt/silver/SalesLT/Address/', name='Address', size=0, modificationTime=1713961582000), FileInfo(path='dbfs:/mnt/silver/SalesLT/Customer/', name='Customer', size=0, modificationTime=1713961593000), FileInfo(path='dbfs:/mnt/silver/SalesLT/CustomerAddress/', name='CustomerAddress', size=0, modificationTime=1713961596000), FileInfo(path='dbfs:/mnt/silver/SalesLT/Product/', name='Product', size=0, modificationTime=1713961600000), FileInfo(path='dbfs:/mnt/silver/SalesLT/ProductCategory/', name='ProductCategory', size=0, modificationTime=1713961603000), FileInfo(path='dbfs:/mnt/silver/SalesLT/ProductDescription/', name='ProductDescription', size=0, modificationTime=1713961606000), FileInfo(path='dbfs:/mnt/silver/SalesLT/ProductModel/', name='ProductModel', size=0, modificationTime=1713961609000), FileInfo(path='dbfs:/mnt/silver/SalesLT/ProductModelProductDescription/', name='ProductModelProductDescription', size=0, modificationTime=1713961610000), FileInfo(path='dbfs:/mnt/silver/SalesLT/SalesOrderDetail/', name='SalesOrderDetail', size=0, modificationTime=1713961614000), FileInfo(path='dbfs:/mnt/silver/SalesLT/SalesOrderHeader/', name='SalesOrderHeader', size=0, modificationTime=1713961616000)]`

2
`dbutils.fs.ls('mnt/gold')`
`[]`

3
`input_path = '/mnt/silver/SalesLT/Address'`

4
`df = spark.read.format('delta').load(input_path)`
`df: pyspark.sql.dataframe.DataFrame = [AddressID: integer, AddressLine1: string ... 7 more fields]`

5
`display(df)`
`(1) Spark Jobs`

AddressID	AddressLine1	AddressLine2	City	StateProvince	CountryRegion
1	8713 Yosemite Ct.	null	Bothell	Washington	United States
2	1318 Lasalle Street	null	Bothell	Washington	United States

silver to gold Python

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, regexp_replace

#Get the list of column names
column_names = df.columns

for old_col_name in column_names:
    # Convert column name from ColumnName to Column_Name
    new_col_name = ''.join(['_'.join([char.lower() if char.isupper() and not old_col_name[i - 1].isupper() else char for i, char in enumerate(old_col_name)]).strip('_')

    #Change the column using WithColumnRenamed and regexp_replace
    df = df.withColumnRenamed(old_col_name, new_col_name)

df: pyspark.sql.dataframe.DataFrame = [Address_ID: integer, Address_Line1: string ... 7 more fields]
```

File Edit View Run Help Last edit was 1 minute ago New cell UI: ON

Run all data_transformation Schedule Share

11:26 AM (1s) 6

11:26 AM (1s) 7

display(df)

(1) Spark Jobs

0.1 KB/s 0.1 KB/s ENG 12:01 PM 25-04-2024

The screenshot shows two instances of the Microsoft Azure Databricks workspace. Both instances have a Python notebook titled "silver to gold" open.

Table Data:

	Address_ID	Address_Line1	Address_Line2	City	State_Province	Country_Region
1	9	8713 Yosemite Ct.	null	Bothell	Washington	United States
2	11	1318 Lasalle Street	null	Bothell	Washington	United States
3	25	9178 Jumping St.	null	Dallas	Texas	United States
4	28	9228 Via Del Sol	null	Phoenix	Arizona	United States
5	32	26910 Indela Road	null	Montreal	Quebec	Canada
6	185	2681 Eagle Peak	null	Bellevue	Washington	United States
7	297	7943 Walnut Ave	null	Renton	Washington	United States
8	445	6388 Lake City Way	null	Burnaby	British Columbia	Canada
9	446	52560 Free Street	null	Toronto	Ontario	Canada
10	447	22580 Free Street	null	Toronto	Ontario	Canada
11	448	2575 Bloor Street East	null	Toronto	Ontario	Canada
12	449	Station E	null	Chalk River	Ontario	Canada
13	450	575 Rue St Amable	null	Quebec	Quebec	Canada

Notebook Code:

```
display(df)

# (1) Spark Jobs
```

```
table_name = []

for i in dbutils.fs.ls('mnt/silver/SalesLT/'):
    table_name.append(i.name.split('/')[0])
```

Output:

Doing transformation for all tables(Changing column names)

```
[ 'Address',
  'Customer',
  'CustomerAddress',
  'Product',
  'ProductCategory',
  'ProductDescription',
  'ProductModel',
  'ProductModelProductDescription',
```

```
for name in table_name:
    path = '/mnt/silver/SalesLT/' + name
    print(path)
    df = spark.read.format('delta').load(path)

    # Get the list of column names
    columns_names = df.columns

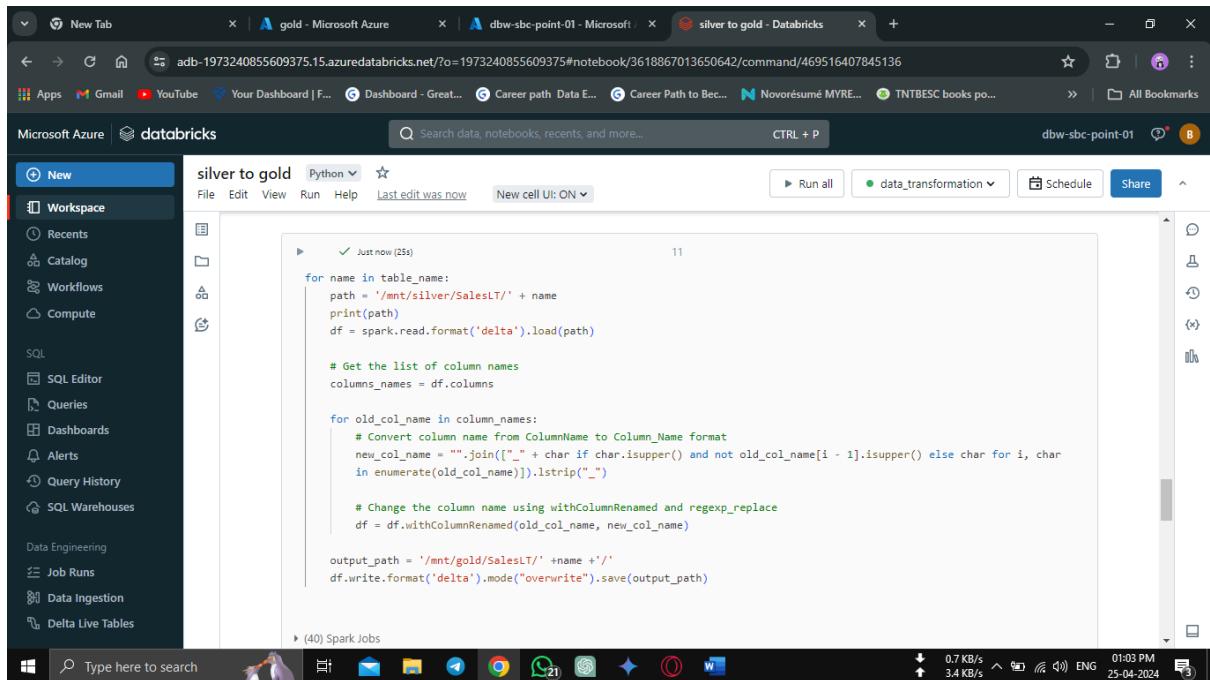
    for old_col_name in columns_names:
        # Convert column name from ColumnName to Column_Name format
        new_col_name = "".join(["_" + char if char.isupper() and not old_col_name[i - 1].isupper() else char for i, char in enumerate(old_col_name)]).lstrip("_")

        # Change the column name using withColumnRenamed and regexp_replace
        df = df.withColumnRenamed(old_col_name, new_col_name)

    output_path = '/mnt/gold/SalesLT/' + name + '/'
    df.write.format('delta').mode("overwrite").save(output_path)
```

```
# Change the column name using withColumnRenamed and regexp_replace
df = df.withColumnRenamed(old_col_name, new_col_name)

output_path = '/mnt/gold/SalesLT/' + name + '/'
df.write.format('delta').mode("overwrite").save(output_path)
```



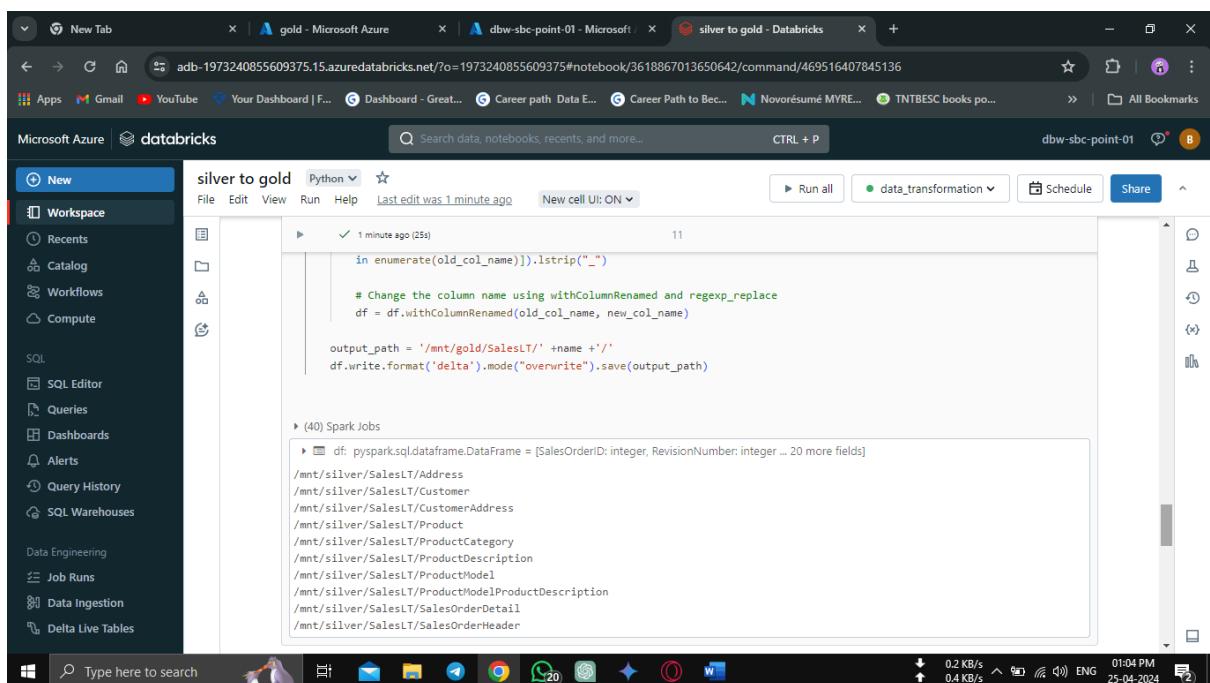
```
for name in table_name:
    path = f'/mnt/silver/SalesLT/{name}'
    print(path)
    df = spark.read.format('delta').load(path)

    # Get the list of column names
    columns_names = df.columns

    for old_col_name in columns_names:
        # Convert column name from ColumnName to Column_Name format
        new_col_name = ''.join(['_'.join([char if char.isupper() and not old_col_name[i-1].isupper() else char for i, char in enumerate(old_col_name)]).lstrip('_')])

        # Change the column name using withColumnRenamed and regexp_replace
        df = df.withColumnRenamed(old_col_name, new_col_name)

    output_path = f'/mnt/gold/SalesLT/{name}/'
    df.write.format('delta').mode('overwrite').save(output_path)
```



```
in enumerate(old_col_name)]).lstrip('_')

# Change the column name using withColumnRenamed and regexp_replace
df = df.withColumnRenamed(old_col_name, new_col_name)

output_path = f'/mnt/gold/SalesLT/{name}/'
df.write.format('delta').mode('overwrite').save(output_path)
```

df: pyspark.sql.dataframe.DataFrame = [SalesOrderID: integer, RevisionNumber: integer ... 20 more fields]

```
/mnt/silver/SalesLT/Address
/mnt/silver/SalesLT/Customer
/mnt/silver/SalesLT/CustomerAddress
/mnt/silver/SalesLT/Product
/mnt/silver/SalesLT/ProductCategory
/mnt/silver/SalesLT/ProductDescription
/mnt/silver/SalesLT/ProductModel
/mnt/silver/SalesLT/ProductModelProductDescription
/mnt/silver/SalesLT/SalesOrderDetail
/mnt/silver/SalesLT/SalesOrderHeader
```

Doing transformation for all tables(Changing column names)

```
08:08 PM (9)
table_name = []

for i in dbutils.fs.ls('mnt/silver/SalesLT'):
    table_name.append(i.name.split('/')[0])

10
table_name

['Address',
 'Customer',
 'CustomerAddress',
 'Product',
 'ProductCategory',
 'ProductDescription',
 'ProductModel',
 'ProductModelDescription',
 'SalesOrderDetail',
 'SalesOrderHeader']

11
for name in table_name:
    path = './mnt/silver/SalesLT/' + name
    print(path)
    df = spark.read.format('delta').load(path)
    # Get the list of column names
    column_names = df.columns
    for old_col_name in column_names:
        new_col_name = ''.join(['_' + char if char.isupper() and not old_col_name[i - 1].isupper() else char for i, char in enumerate(old_col_name)]).lstrip('_')
        df = df.withColumnRenamed(old_col_name, new_col_name)

df = pyspark.sql.DataFrame = [SalesOrderID: integer, RevisionNumber: integer ... 20 more fields]

/mnt/silver/SalesLT/Address
/mnt/silver/SalesLT/Customer
/mnt/silver/SalesLT/CustomerAddress
/mnt/silver/SalesLT/Product
/mnt/silver/SalesLT/ProductCategory
/mnt/silver/SalesLT/ProductDescription
/mnt/silver/SalesLT/ProductModel
/mnt/silver/SalesLT/ProductModelDescription
/mnt/silver/SalesLT/SalesOrderDetail
/mnt/silver/SalesLT/SalesOrderHeader
```

Screenshot 1: Databricks Notebook - silver to gold

The screenshot shows a Microsoft Edge browser window with two tabs open. The active tab is a Databricks notebook titled "silver to gold". The notebook interface includes a sidebar with various Databricks services like Workspace, Catalog, Workflows, Compute, and Data Engineering. The main area displays a Python code cell with the following content:

```
✓ 08:08 PM (1s)
display(df)

(1) Spark Jobs
```

Below the code cell is a table showing 12 rows of data from the SalesLT table. The columns are SalesOrderID, RevisionNumber, OrderDate, DueDate, ShipDate, Status, and Online.

SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	Online
1	71774	2008-06-01	2008-06-13	2008-06-08	5	false
2	71776	2008-06-01	2008-06-13	2008-06-08	5	false
3	71780	2008-06-01	2008-06-13	2008-06-08	5	false
4	71782	2008-06-01	2008-06-13	2008-06-08	5	false
5	71783	2008-06-01	2008-06-13	2008-06-08	5	false
6	71784	2008-06-01	2008-06-13	2008-06-08	5	false
7	71796	2008-06-01	2008-06-13	2008-06-08	5	false
8	71797	2008-06-01	2008-06-13	2008-06-08	5	false
9	71815	2008-06-01	2008-06-13	2008-06-08	5	false
10	71816	2008-06-01	2008-06-13	2008-06-08	5	false
11	71831	2008-06-01	2008-06-13	2008-06-08	5	false
12	71832	2008-06-01	2008-06-13	2008-06-08	5	false

Screenshot 2: Databricks Notebook - silver to gold

This screenshot shows the same Databricks notebook environment. The code cell has been modified to include a loop that processes multiple table names:

```
✓ 08:08 PM (2s)
for name in table_name:
    path = '/mnt/silver/SalesLT/' + name
    print(path)
    df = spark.read.format('delta').load(path)
    # Get the list of column names
    column_names = df.columns
    for old_col_name in column_names:
        # Convert column name from ColumnName to Column_Name format
        new_col_name = "".join(["_." + char if char.isupper() and not old_col_name[i - 1].isupper() else char for i,char in enumerate(old_col_name)]).lstrip("_")
        # Change the column name using withColumnRenamed and regexp_replace
        df = df.withColumnRenamed(old_col_name, new_col_name)

    output_path = '/mnt/gold/SalesLT/' + name + '/'
    df.write.format('delta').mode("overwrite").save(output_path)

(4) Spark Jobs
```

The notebook interface includes a search bar, a toolbar with "Run all", "Schedule", and "Share" buttons, and a status bar at the bottom showing network activity and system information.

silver to gold Python

```

for name in table_name:
    path = '/mnt/silver/SalesLT/' + name
    print(path)
    df = spark.read.format('delta').load(path)
    column_names = df.columns
    for old_col_name in column_names:
        new_col_name = "".join(["_" + char if char.isupper() and not old_col_name[i - 1].isupper() else char for i, char in enumerate(old_col_name)]).lstrip("_")
        df = df.withColumnRenamed(old_col_name, new_col_name)
    output_path = '/mnt/gold/SalesLT/' + name + '/'
    df.write.format('delta').mode("overwrite").save(output_path)

(40) Spark Jobs
+ df: pyspark.sql.dataframe.DataFrame = [Sales_Order_ID:integer, Revision_Number:integer ... 20 more fields]
/mnt/silver/SalesLT/Address
/mnt/silver/SalesLT/Customer
/mnt/silver/SalesLT/CustomerAddress
/mnt/silver/SalesLT/Product
/mnt/silver/SalesLT/ProductCategory
/mnt/silver/SalesLT/ProductDescription
/mnt/silver/SalesLT/ProductModel
/mnt/silver/SalesLT/ProductModelProductDescription

```

silver to gold Python

```

display(df)

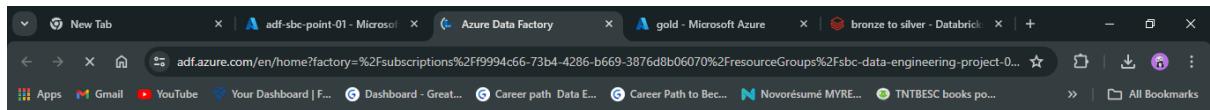
```

(1) Spark Jobs

Sales_Order_ID	Revision_Number	Order_Date	Due_Date	Ship_Date	Status
1	71774	2008-06-01	2008-06-13	2008-06-08	5 false
2	71776	2008-06-01	2008-06-13	2008-06-08	5 false
3	71780	2008-06-01	2008-06-13	2008-06-08	5 false
4	71782	2008-06-01	2008-06-13	2008-06-08	5 false
5	71783	2008-06-01	2008-06-13	2008-06-08	5 false
6	71784	2008-06-01	2008-06-13	2008-06-08	5 false
7	71796	2008-06-01	2008-06-13	2008-06-08	5 false
8	71797	2008-06-01	2008-06-13	2008-06-08	5 false
9	71815	2008-06-01	2008-06-13	2008-06-08	5 false
10	71816	2008-06-01	2008-06-13	2008-06-08	5 false
11	71831	2008-06-01	2008-06-13	2008-06-08	5 false
12	71832	2008-06-01	2008-06-13	2008-06-08	5 false
13	71845	2008-06-01	2008-06-13	2008-06-08	5 false

The screenshot shows the Microsoft Azure Storage Container 'gold'. The left sidebar includes 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', 'Settings' (Shared access tokens, Manage ACL, Access policy, Properties, Metadata), and 'Search' bar. The main area displays a table of blobs with columns: Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. The table lists items like Address, Customer, CustomerAddress, Product, ProductCategory, ProductDescription, ProductModel, ProductModelPro..., and SalesOrderDetail.

The screenshot shows the Microsoft Azure Data Factory 'adf-sbc-point-01'. The left sidebar includes 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Settings' (Networking, Managed identities, Properties, Locks), 'Getting started' (Quick start), and 'Monitoring'. The main area displays 'Essentials' details: Resource group (move), Status (Succeeded), Location (East US), Subscription (move) (Free Trial), and Subscription ID (f9994c66-73b4-4286-b669-3876d8b06070). It also features a 'Launch studio' button and a large Azure Data Factory Studio logo.



A screenshot of the Microsoft Azure portal. The left sidebar shows 'Data Factory' under 'adf-sbc-point-01'. The main area is titled 'Linked services' with the sub-header 'Linked service defines the connection information'. It lists three items: 'AzureDataLakeStorage1' (Type: Azure Data Lake Storage), 'AzureKeyVault1' (Type: Azure Key Vault), and 'onpremssqlserver' (Type: SQL server). To the right, there is a 'New linked service' dialog box. The 'Compute' tab is selected, showing icons for 'Azure Batch', 'Azure Data Lake Analytics', and 'Azure Databricks'. Below the icons are 'Continue' and 'Cancel' buttons. The taskbar at the bottom shows the same pinned icons as the first screenshot.

The image shows two side-by-side screenshots of the Microsoft Azure portal.

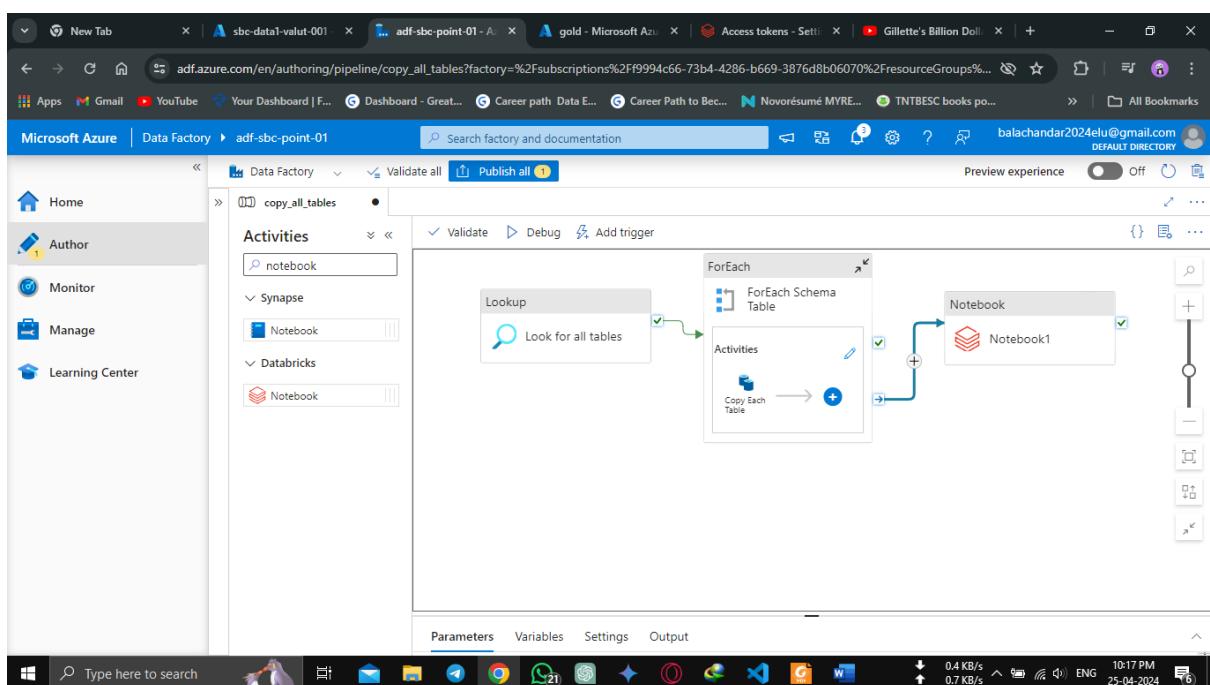
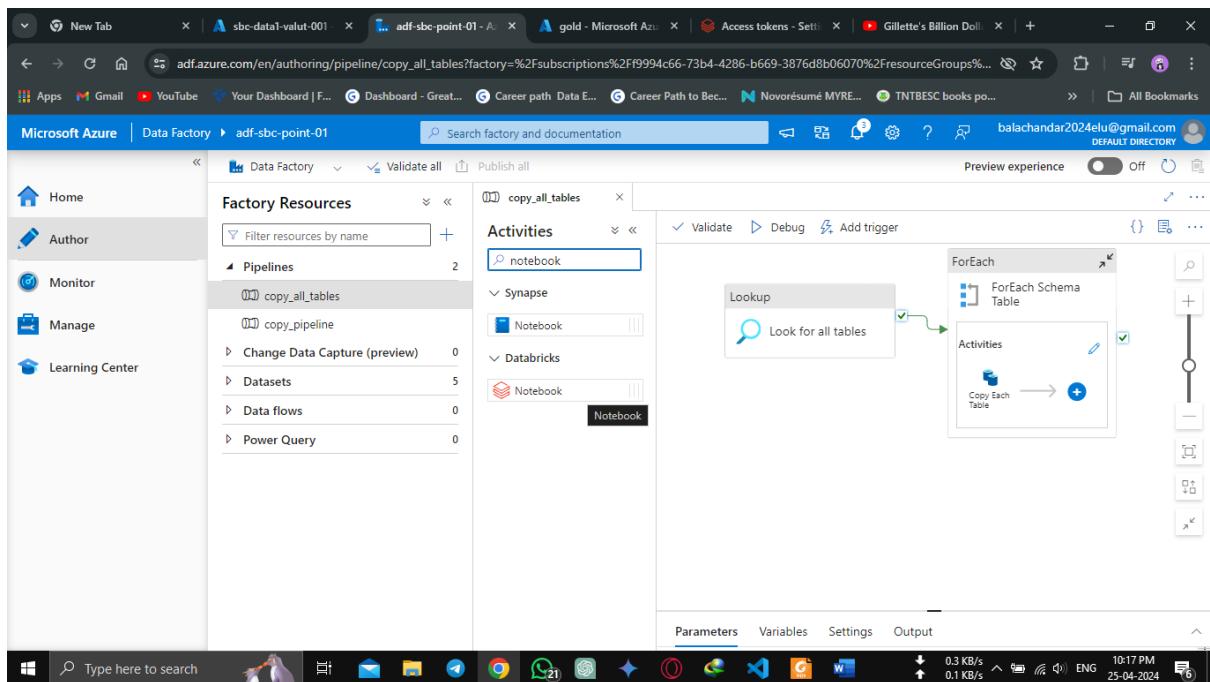
Left Screenshot (Databricks):

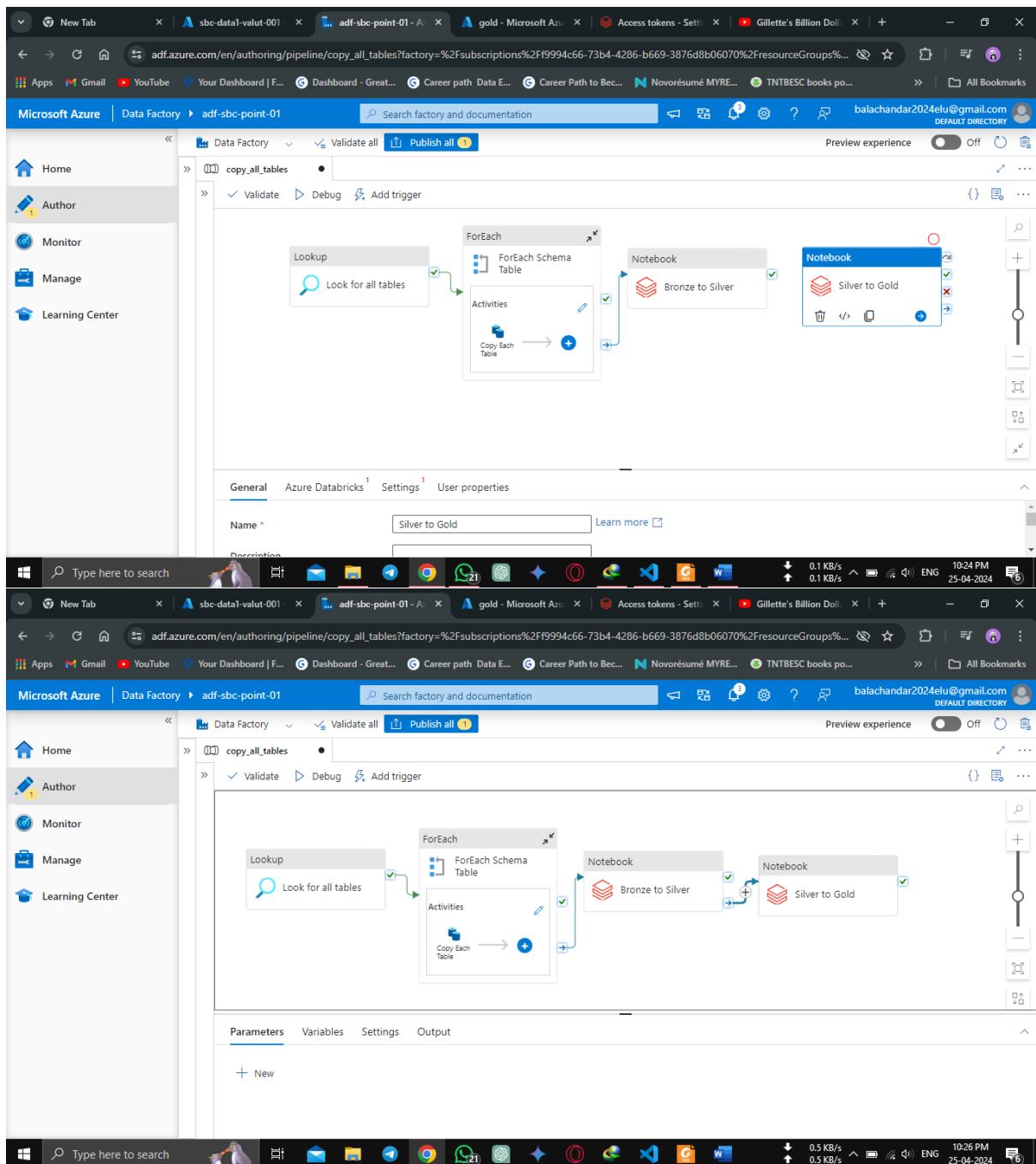
- Header:** Microsoft Azure | databricks
- Left Sidebar:** New, Workspace, Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables.
- Active Section:** Settings > Developer > Access tokens
- Content:** Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords. A "Generate new token" button is present. A message indicates "No tokens exist."
- Top Right:** Free trial ends in 13 days. Upgrade to Premium in Azure Portal.

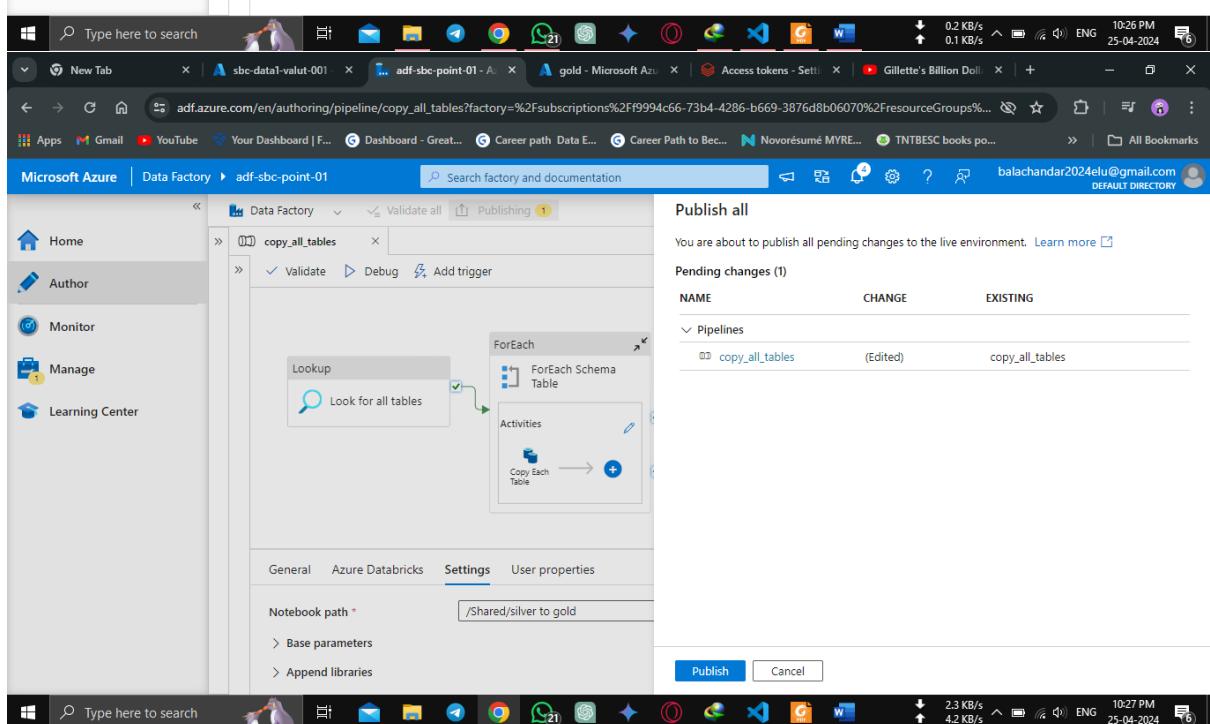
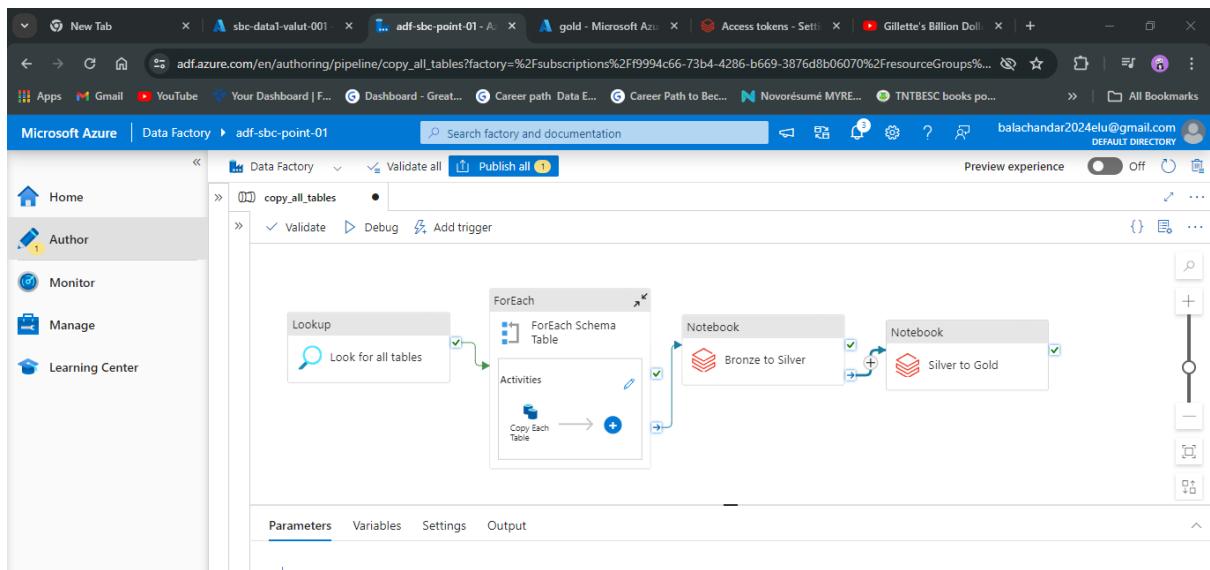
Right Screenshot (Data Factory):

- Header:** Microsoft Azure | Data Factory | adf-sbc-point-01
- Left Sidebar:** Home, Author, Monitor, Manage (selected), Learning Center.
- Active Section:** Data Factory > Validate all > Publish all (1)
- Content:** Linked services section. Sub-sections include General, Factory settings, Connections, and Linked services (selected).
 - General:** Integration runtimes, Microsoft Purview, Source control (Git configuration, ARM template).
 - Connections:** Author, Triggers, Global parameters, Data flow libraries, Security, Credentials.
 - Linked services:** Shows 4 items:

Name	Type	Related	Annotations
AzureDatabricks1	Azure Databricks	0	
AzureDataLakeStorage1	Azure Data Lake Storage Gen2	2	
AzureKeyVault1	Azure Key Vault	2	
onpremssqlserver	SQL server	3	
- Top Right:** Preview experience toggle (Off), Default directory.
- Bottom:** Taskbar showing various Azure services and system icons.







Screenshot of the Microsoft Azure Data Factory pipeline run details page.

The pipeline run ID is 2ca70e61-976a-4ccf-bec1-fee05ec9f617.

The pipeline run status is In progress.

The pipeline run duration is 12s.

The pipeline run start time is 4/25/2024, 11:04:06 PM.

The pipeline run end time is 4/25/2024, 11:04:18 PM.

The pipeline run user properties include:

- Integration runtime: SHIR
- User properties: 4.4 KB/s, 0.9 KB/s

The pipeline run details table shows the following activity runs:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Look for all tables	Succeeded	Lookup	4/25/2024, 11:03:36 PM	28s	SHIR	4.4 KB/s, 0.9 KB/s
ForEach Schema Table	In progress	ForEach	4/25/2024, 11:04:05 PM	12s		
Copy Each Table	Queued	Copy data	4/25/2024, 11:04:06 PM	11s		
Copy Each Table	Queued	Copy data	4/25/2024, 11:04:06 PM	11s		

Screenshot of Microsoft Azure Data Factory Pipeline Runs - copy_all_tables - Activity runs

The screenshot shows the Azure Data Factory interface for monitoring pipeline runs. The left sidebar is collapsed, and the main area displays the activity runs for the 'copy_all_tables' pipeline.

Pipeline Run Overview:

- Run ID:** 2ca70e61-976a-4ccf-bec1-fee05ec9f617
- Run Status:** Succeeded
- Run Start:** 4/25/2024, 11:04:06 PM
- Run Duration:** 2m 10s
- Integration Runtime:** SHIR
- Run Details:** Pipeline run ID 2ca70e61-976a-4ccf-bec1-fee05ec9f617

Activity Runs:

The pipeline consists of the following activities:

```

graph LR
    L1[Lookup: Look for all tables] --> F1[ForEach Schema Table]
    F1 --> A1[Activities]
    A1 --> C1[Copy Each Table]
    A1 --> C2[Copy Each Table]
    A1 --> C3[Copy Each Table]
    A1 --> C4[Copy Each Table]
    A1 --> C5[Copy Each Table]
    A1 --> C6[Copy Each Table]
    A1 --> C7[Copy Each Table]
    A1 --> C8[Copy Each Table]
    A1 --> C9[Copy Each Table]
    C1 --> N1[Notebook: Bronze to Silver]
    C2 --> N1
    C3 --> N1
    C4 --> N1
    C5 --> N1
    C6 --> N1
    C7 --> N1
    C8 --> N1
    C9 --> N1
    N1 --> N2[Notebook: Silver to Gold]
    C9 --> N2
  
```

Activity Log:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Look for all tables	Succeeded	Lookup	4/25/2024, 11:03:36 PM	28s	SHIR	
ForEach Schema Table	Succeeded	ForEach	4/25/2024, 11:04:05 PM	4m 7s	SHIR	
Copy Each Table	Succeeded	Copy data	4/25/2024, 11:04:06 PM	3m 21s	SHIR	
Copy Each Table	Succeeded	Copy data	4/25/2024, 11:04:06 PM	1m 29s	SHIR	
Copy Each Table	Succeeded	Copy data	4/25/2024, 11:04:06 PM	1m 30s	SHIR	

Screenshot of Microsoft Azure Data Factory Pipeline Runs - copy_all_tables - Activity runs

The screenshot shows the Azure Data Factory interface for monitoring pipeline runs. The pipeline, named "copy_all_tables", consists of the following steps:

- A "Lookup" activity named "Look for all tables" is connected to an "ForEach" activity.
- The "ForEach" activity has two parallel branches:
 - The first branch contains an "Activities" group with a "Copy Each Table" activity.
 - The second branch contains two "Notebook" activities: "Bronze to Silver" and "Silver to Gold".

The pipeline run ID is 2ca70e61-976a-4ccf-bec1-fee05ec9f617. The run status is successful, with a duration of 4m 7s. The run started at 4/25/2024, 11:03:36 PM.

Activity runs table:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Look for all tables	Succeeded	Lookup	4/25/2024, 11:03:36 PM	28s	SHIR	
ForEach Schema Table	Succeeded	ForEach	4/25/2024, 11:04:05 PM	4m 7s	SHIR	
Copy Each Table	Succeeded	Copy data	4/25/2024, 11:04:06 PM	3m 21s	SHIR	

Activity runs table (second screenshot):

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Look for all tables	Succeeded	Lookup	4/25/2024, 11:03:36 PM	28s	SHIR	
ForEach Schema Table	Succeeded	ForEach	4/25/2024, 11:04:05 PM	4m 7s	SHIR	

A gold - Microsoft Azure | A Create Synapse workspace | Databricks | A adf-sbc-point-01 - M | A adf-sbc-point-01 - A... | A gold - Microsoft Azure | +

portal.azure.com/#create/Microsoft.Synapse

Apps Gmail YouTube Your Dashboard | Fo... Dashboard - Great L... Career path Data E... Career Path to Beco... Novorésumé MYRE... TNTBESC books por... All Bookmarks

Microsoft Azure Upgrade Search resources, services, and docs (G+/)

Home > Azure Synapse Analytics > Create Synapse workspace ...

Basics *Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

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

Subscription * ⓘ Free Trial ⓘ The Synapse and SQL resource providers are now registered with this subscription.

Resource group * ⓘ sbc-data-engineering-project-02 ⓘ Create new

Managed resource group ⓘ Enter managed resource group name

Workspace details

Review + create < Previous Next: Security >

Type here to search

0.1 KB/s 0.0 KB/s ENG 07:13 AM 26-04-2024

Screenshot of the Microsoft Azure portal showing the creation of a Synapse workspace.

The browser tabs include:

- gold - Microsoft Azure
- Create Synapse workspace
- Databricks
- adf-sbc-point-01 - M
- adf-sbc-point-01 - A
- gold - Microsoft Azure

The main page shows the "Create Synapse workspace" wizard, step 2: Security. The "Security" tab is selected. The configuration includes:

- Authentication method: Use both local and Microsoft Entra ID authentication (selected)
- SQL Server admin login: sbcadminsql
- SQL Password: (redacted)
- Confirm password: (redacted)
- System assigned managed identity permission: (dropdown menu)

Buttons at the bottom: Review + create, < Previous, Next: Networking >

The status bar at the bottom shows: 1.4 KB/s, 1.6 KB/s, ENG, 07:20 AM, 26-04-2024.

Screenshot of the Microsoft Azure portal showing the deployment details of the Synapse workspace.

The browser tabs include:

- gold - Microsoft Azure
- Microsoft.Azure.Synapse
- Databricks
- adf-sbc-point-01 - M
- adf-sbc-point-01 - A
- gold - Microsoft Azure

The main page shows the "Microsoft.Azure.SynapseAnalytics-20240426071055 | Overview" page. The deployment status is "Deployment is in progress".

Deployment details:

Deployment name	Start time
Microsoft.Azure.SynapseAnalytics-202...	26/4/2024, 7:23:53 am

Resource group: sbc-data-engineering-project-02

Deployment details table:

Resource	Type	Status	Operation details
synw-sbc-point...	Synapse workspace	Created	Operation details

Give feedback: Tell us about your experience with deployment

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 >

The status bar at the bottom shows: 1.3 KB/s, 1.4 KB/s, ENG, 07:24 AM, 26-04-2024.

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#@balachandar2024elugmail.onmicrosoft.com/resource/subscriptions/f9994c66-73b4-4286-b669-3876d8b06070/resourceGroups/sbc-data-engineering-project-02/providers/Microsoft.Analytics/synapseWorkspaces/synw-sbc-point-01>. The page displays the 'Essentials' section of the workspace configuration, including details like Resource group, Status, Location, Subscriptions, and Networking. It also features a 'Getting started' section with links to Open Synapse Studio and Read documentation. The left sidebar shows navigation options for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Microsoft Entra ID, Properties, Locks), and Analytics pools (SQL pools, Apache Spark pools, Data Explorer pools (preview)). The bottom of the screen shows the Windows taskbar with various pinned icons.

This screenshot is nearly identical to the one above, showing the Microsoft Azure portal at the same URL. The main difference is in the 'Analytics pools' section, which now lists a single 'SQL pool' named 'Built-in' with a 'Serverless' type and 'Auto' size. The rest of the interface, including the 'Essentials' section and 'Getting started' area, remains the same. The Windows taskbar at the bottom is also visible.

Screenshot of the Microsoft Azure Synapse Analytics workspace page for 'synw-sbc-point-01'.

The page features a large circular network graph visualization on the right side. On the left, there's a sidebar with navigation links: Home, Data, Develop, Integrate, Monitor, and Manage. Below the sidebar, three cards are displayed: 'Ingest' (Perform a one-time or scheduled data load.), 'Explore and analyze' (Learn how to get insights from your data.), and 'Visualize' (Build interactive reports with Power BI capabilities.).

At the bottom, there's a 'Discover more' section with links to 'Knowledge center' and 'Browse partners'. The browser taskbar at the bottom shows various open tabs including Microsoft Edge, Gmail, YouTube, and Databricks.

Screenshot of the Microsoft Azure Synapse Analytics workspace page for 'synw-sbc-point-01', showing the 'Data' section.

The 'Data' section is currently selected in the sidebar. A dropdown menu is open under 'Workspace', showing options like 'SQL database', 'Lake database', 'Data Explorer database (preview)', 'Connect to external data', 'Integration dataset', and 'Browse gallery'. The 'Select an item' message is visible at the bottom of the dropdown.

The browser taskbar at the bottom shows various open tabs including Microsoft Edge, Gmail, YouTube, and Databricks.

The screenshot shows two side-by-side browser windows for Microsoft Azure Synapse Analytics.

Left Window: The URL is `web.azure-synapse.net/en/authoring/explore/workspace?workspace=%2Fsubscriptions%2F9994c66-73b4-4286-b669-3876d8b06070%2FresourceGroups...` . The page displays a 'Create SQL database' dialog. It includes a 'Select an item' section with two cylindrical icons, a 'Create' button, and a 'Cancel' button. The right pane shows a 'Data' workspace with tabs for 'Workspace' and 'Linked'. A sidebar on the left lists 'Home', 'Data', 'Develop', 'Integrate', 'Monitor', and 'Manage'.

Right Window: The URL is `web.azure-synapse.net/en/management/sqlpools?workspace=%2Fsubscriptions%2F9994c66-73b4-4286-b669-3876d8b06070%2FresourceGroups%2Fscb...` . The page displays the 'SQL pools' section. It shows a message about the serverless SQL pool being available. Below is a table listing one item:

Name	Type	Status	Size
Built-in	Serverless	Online	Auto

The screenshot shows a Microsoft Azure Synapse Analytics workspace titled "synw-sbc-point-01". On the left, a navigation menu includes Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays a "Data" section with "Workspace" and "Linked" tabs. Under "Linked", there is a "Filter resources by name" input field and a list of resources under "Azure Data Lake Storage Gen2": "synw-sbc-point-01 (Primary - sbcd...)" containing "gold (Primary)", "bronze", and "silver"; and "(Attached Containers)". A modal dialog titled "Select TOP 100 rows" is open over the workspace. It contains a "Source folder format settings" section with "Folder path" set to "https://sbcdatalakegen02.dfs.core.windows.net/gold/SalesLT/Address/" and "File type" set to "Delta format". At the bottom of the dialog are "Apply" and "Cancel" buttons.

Windows Type here to search □ Mail File Explorer Edge Google Chrome Microsoft Edge Microsoft Word Microsoft Excel 0.5 KB/s 7.6 KB/s ENG 07:56 AM 26-04-2024

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, there's a navigation sidebar with options like Home, Data, Develop, Integrate, Monitor, and Manage. The main area is titled 'Data' and shows a 'Linked' section. A tree view indicates an Azure Data Lake Storage Gen2 account named 'synw-sbc-point-01' with three containers: gold (Primary), bronze, and silver. Below this is a '(Attached Containers)' section.

In the center, there's a query editor window titled 'gold'. The tab 'SQL script 1' is selected, showing the following T-SQL code:

```
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://sbcdatalakegen02.dfs.core.windows.net/gold/SalesLT/Address/',
7         FORMAT = 'DELTA'
8     ) AS [result]
```

Below the code, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, displaying a table with two rows of data:

Address_ID	Address_Line1	Address_Line2	City	State_Province	Country_Region	Postal_C
9	8713 Yosemite ...	(NULL)	Bothell	Washington	United States	98011
11	1318 Lasalle Str...	(NULL)	Bothell	Washington	United States	98011

At the bottom of the results pane, it says '00:00:28 Query executed successfully.'

The bottom of the screen shows the Windows taskbar with various pinned icons and the system tray indicating network speed (0.3 KB/s up, 0.1 KB/s down) and date/time (07:57 AM, 26-04-2024).

A screenshot of the Microsoft Azure Synapse Analytics workspace. The left sidebar shows navigation options: Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays a SQL script editor titled 'gold' with a tab 'SQL script 1'. The script contains the following code:

```
1 CREATE VIEW address
2 AS
3 SELECT *
4 FROM
5 OPENROWSET(
6     BULK 'https://sbcdatalakegen02.dfs.core.windows.net/gold/SalesLT/Address/',
7     FORMAT = 'DELTA'
8 ) AS [result]
```

The 'Messages' section shows the following log entry:

8:00:31 AM Started executing query at Line 1
CREATE/ALTER VIEW is not supported in master database.
Visit this article to learn more about this error.
Total execution time: 00:00:03.534

At the bottom, a status bar indicates: 0.4 KB/s, 0.3 KB/s, ENG, 08:00 AM, 26-04-2024.

A screenshot of the Microsoft Azure Synapse Analytics workspace. The left sidebar shows navigation options: Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays a SQL script editor titled 'gold' with a tab 'SQL script 1'. The script is identical to the one in the previous screenshot:

```
1 CREATE VIEW address
2 AS
3 SELECT *
4 FROM
5 OPENROWSET(
6     BULK 'https://sbcdatalakegen02.dfs.core.windows.net/gold/SalesLT/Address/',
7     FORMAT = 'DELTA'
8 ) AS [result]
```

The 'Results' tab is selected, showing the following message:

No results to show
Your query yielded no displayable results

At the bottom, a status bar indicates: 1.2 KB/s, 0.6 KB/s, ENG, 08:02 AM, 26-04-2024.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar has 'Data' selected under 'Workspace'. In the main area, a SQL script titled 'gold' is open in 'SQL script 1'. The script creates a view named 'address' with a single column '*' and a query that reads from an OPENROWSET BULK source pointing to a URL. The results pane shows 'No results to show'.

```
CREATE VIEW address
AS
SELECT *
FROM
OPENROWSET(
    BULK 'https://sbcdatalakegen02.dfs.core.windows.net/gold/SalesLT/Address/',
    FORMAT = 'DELTA'
) AS [result]
```

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar has 'Data' selected under 'Workspace'. In the main area, a SQL script titled 'gold' is open in 'SQL script 2'. The script selects the top 100 rows from the 'address' table, listing columns such as Address_ID, Address_Line1, Address_Line2, City, State_Province, Country_Region, and Postal_Code. The results pane displays the first row of data.

```
SELECT TOP (100) [Address_ID]
,[Address_Line1]
,[Address_Line2]
,[City]
,[State_Province]
,[Country_Region]
,[Postal_Code]
,[rowguid]
,[Modified_Date]
FROM [dbo].[address]
```

Address_ID	Address_Line1	Address_Line2	City	State_Province	Country_Region	Postal_C
9	8713 Yosemite ...	(NULL)	Bothell	Washington	United States	98011

Screenshot of Microsoft Azure Synapse Analytics workspace showing Data exploration and Integration.

Data View:

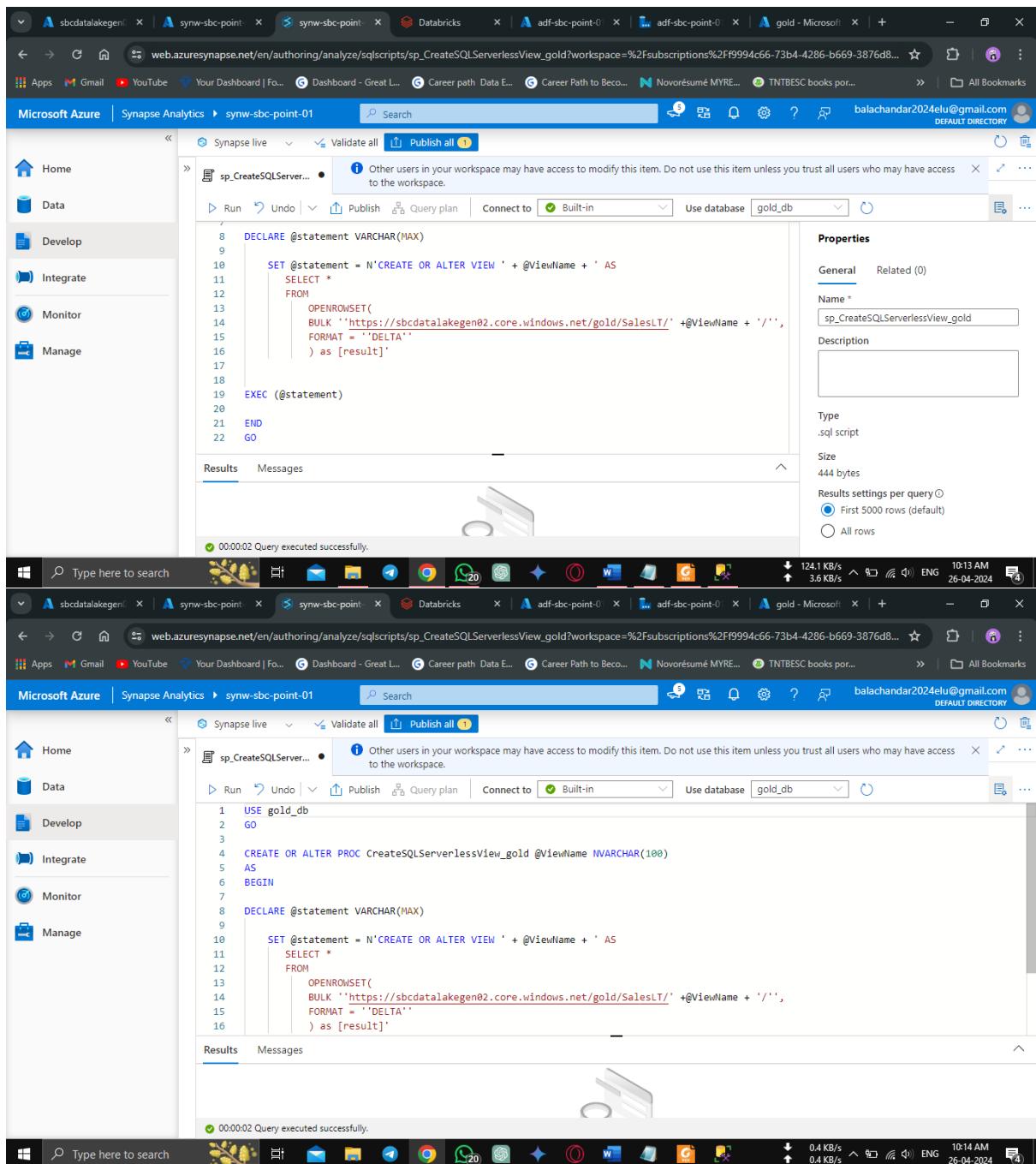
- Left sidebar: Home, Data, Develop, Integrate, Monitor, Manage.
- Central pane: Workspace (gold) selected.
- SQL database: gold_db (SQL) expanded, showing External tables, External resources, Views (dbo.address, System views), Schemas, Security.
- SQL script 1: SELECT TOP (100) [Address_ID], [Address_Line1], [Address_Line2], [City], [State_Province], [Country_Region].
- Results: Address_ID, Address_Line1, Address_Line2, City, State_Province, Country_Region, Postal_C.

Address_ID	Address_Line1	Address_Line2	City	State_Province	Country_Region	Postal_C
9	8713 Yosemite ...	(NULL)	Bothell	Washington	United States	98011
11	1318 Lasalle Str...	(NULL)	Bothell	Washington	United States	98011
25	9178 Jumping St.	(NULL)	Dallas	Texas	United States	75201
28	9228 Via Del Sol	(NULL)	Phoenix	Arizona	United States	85004

00:00:07 Query executed successfully.

Integrate View:

 - Left sidebar: Home, Data, Develop, Integrate, Monitor, Manage.
 - Central pane: Integrate (gold) selected.
 - No items to show. Try creating a new item using the + button above. [Learn more](#).
 - Select an item. Use the resource explorer to select or create a new item.



The screenshot shows two nearly identical SQL scripts being run in the Microsoft Azure Synapse Analytics workspace. Both scripts are intended to create a view named `CreateSQLServerlessView_gold` in the `gold_db` database.

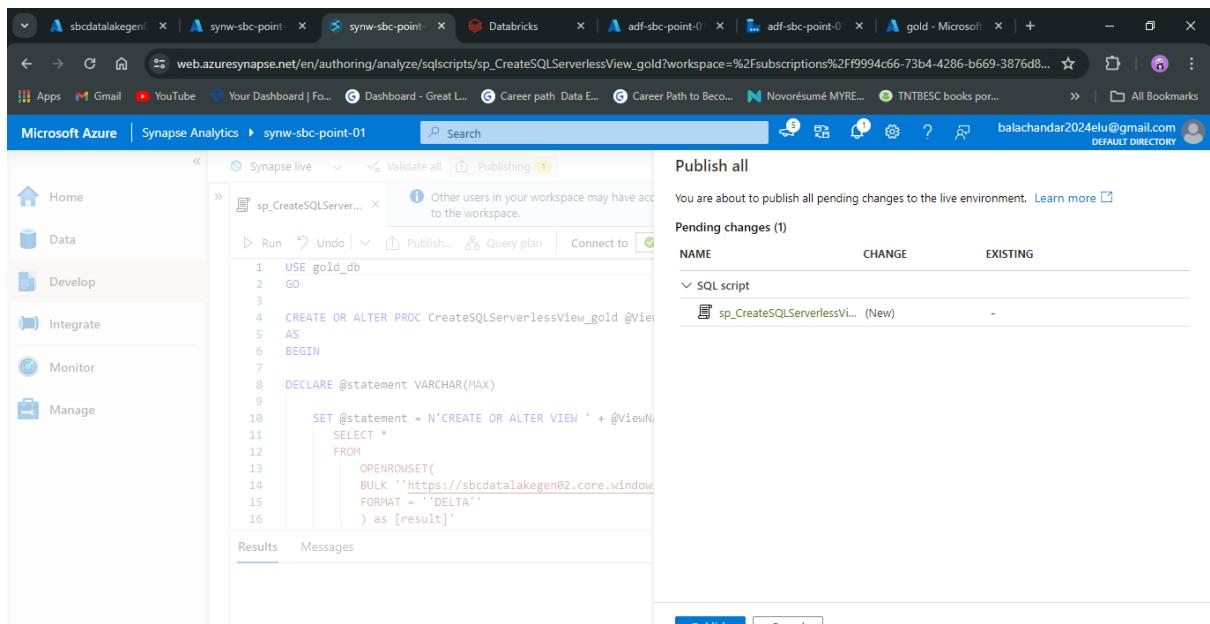
Script 1 (Top):

```
8 DECLARE @statement VARCHAR(MAX)
9
10 SET @statement = N'CREATE OR ALTER VIEW ' + @ViewName + ' AS
11     SELECT *
12     FROM
13         OPENROWSET(
14             BULK ''https://sbcdatalakegen02.core.windows.net/gold/SalesLT/' + @ViewName + '/',
15             FORMAT = 'DELTA'
16         ) as [result]
17
18 EXEC (@statement)
19
20 END
21 GO
```

Script 2 (Bottom):

```
1 USE gold_db
2 GO
3
4 CREATE OR ALTER PROC CreateSQLServerlessView_gold @ViewName NVARCHAR(100)
5 AS
6 BEGIN
7
8 DECLARE @statement VARCHAR(MAX)
9
10 SET @statement = N'CREATE OR ALTER VIEW ' + @ViewName + ' AS
11     SELECT *
12     FROM
13         OPENROWSET(
14             BULK ''https://sbcdatalakegen02.core.windows.net/gold/SalesLT/' + @ViewName + '/',
15             FORMAT = 'DELTA'
16         ) as [result]
17
18 EXEC (@statement)
19
20 END
21 GO
```

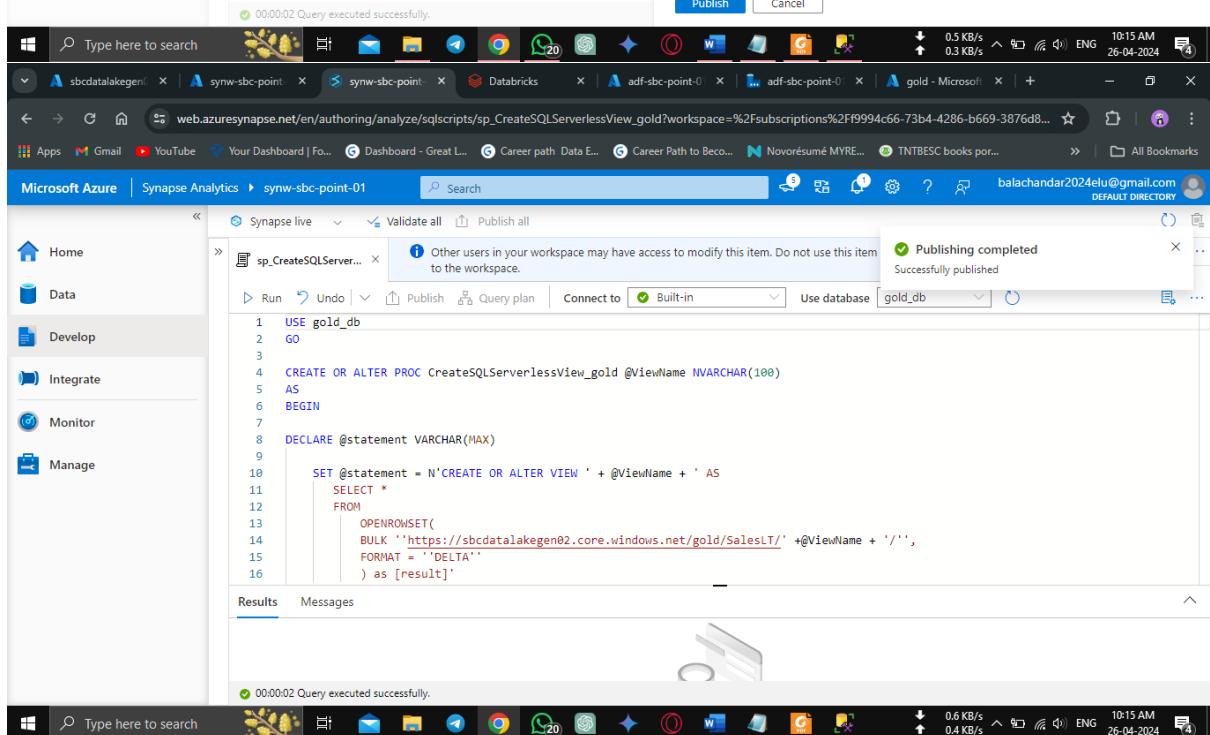
In both scripts, the URL used in the `OPENROWSET` command is `https://sbcdatalakegen02.core.windows.net/gold/SalesLT/`, followed by the view name `@ViewName`.



The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. A modal dialog titled "Publish all" is open, prompting the user to publish pending changes to the live environment. The dialog includes a "Pending changes (1)" section showing a single item: "sp_CreateSQLServerlessView_gold" (New). Below the dialog, the main workspace shows a query editor with the following T-SQL script:

```
1 USE gold_db
2 GO
3
4 CREATE OR ALTER PROC CreateSQLServerlessView_gold @viewName NVARCHAR(100)
5 AS
6 BEGIN
7
8 DECLARE @statement VARCHAR(MAX)
9
10 SET @statement = N'CREATE OR ALTER VIEW ' + @viewName +
11 'SELECT * '
12 FROM
13 OPENROWSET(
14 BULK ''https://sbcdatalakegen02.core.windows.net/gold/SalesLT/' + @viewName + '/',
15 FORMAT = ''DELTA''
16 ) as [result]'
```

The "Results" tab is selected, and a message at the bottom indicates "00:00:02 Query executed successfully."



The second screenshot shows the same workspace after the publish operation has completed. The "Publish all" dialog has closed, and the workspace now displays a success message: "Publishing completed" and "Successfully published". The T-SQL script remains the same as in the first screenshot.

The screenshot displays two windows from the Microsoft Azure Synapse Analytics workspace.

Top Window: A query editor titled "sp.CreateSQLServerlessView_gold". The code pane contains the following T-SQL script:

```
1 USE gold_db
2 GO
3
4 CREATE OR ALTER PROC CreateSQLServerlessView_gold @ViewName NVARCHAR(100)
5 AS
6 BEGIN
7
8 DECLARE @statement VARCHAR(MAX)
9
10 SET @statement = N'CREATE OR ALTER VIEW ' + @ViewName + ' AS
11     SELECT *
12     FROM
13         OPENROWSET(
14             BULK ''https://sbcdatalakegen02.core.windows.net/gold/SalesLT/' + @ViewName + '/',
15             FORMAT = ''DELTA''
16         ) as [result]'
```

The status bar at the bottom indicates "00:00:02 Query executed successfully."

Bottom Window: A page titled "Linked services". It shows a table of existing linked services:

Name	Type	Related	Annotations
synw-sbc-point-01-Workspa...	Azure Synapse Analytics	0	
synw-sbc-point-01-Workspa...	Azure Data Lake Storage Gen2	0	

New linked service

sql data

All Azure Compute Database File Generic protocol NoSQL

Azure SQL Database

Azure SQL Database Managed Instance

Continue Cancel

New linked service

Azure SQL Database Learn more

Name * serverlessSQLdb

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Account selection method Enter manually

Fully qualified domain name * synw-sbc-point-01-on-demand.sql.azuresynapse.net

Database name * gold_db

Create Back Test connection Cancel

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

New linked service

Azure SQL Database [Learn more](#)

Choose a name for your linked service. This name cannot be updated later.

Name *

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Account selection method Enter manually

Fully qualified domain name *

Connection successful

Create Back Test connection Cancel

Type here to search

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

Publish all

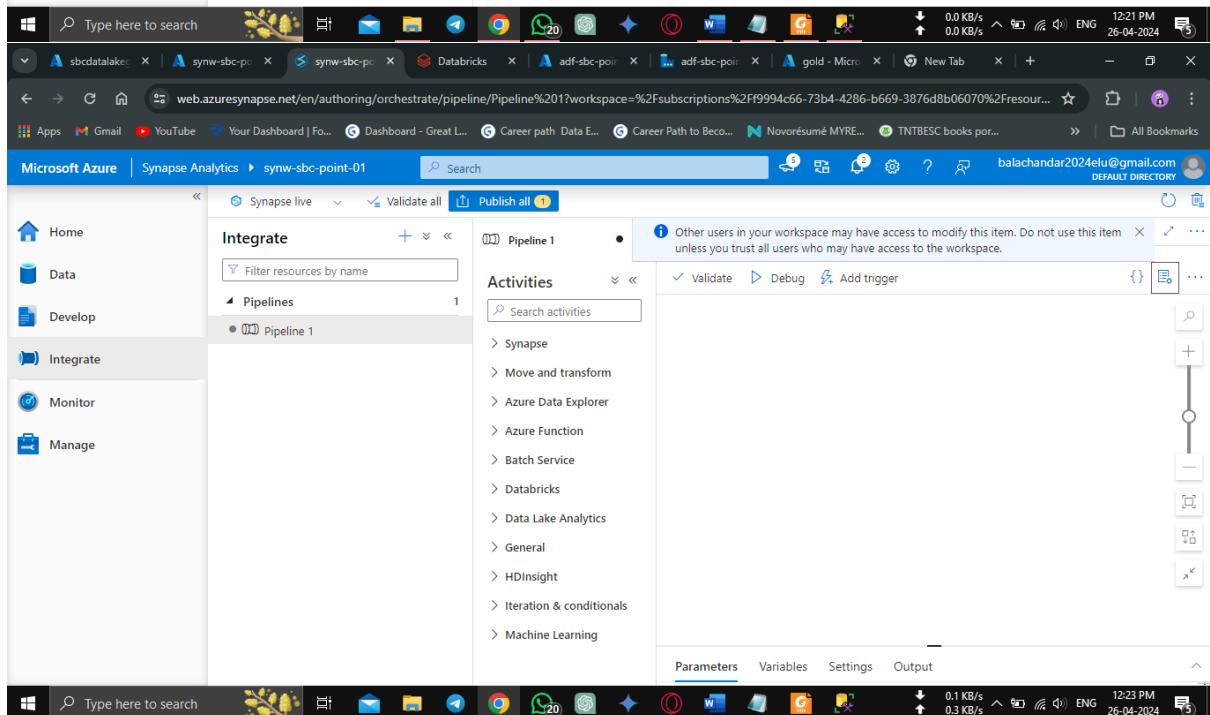
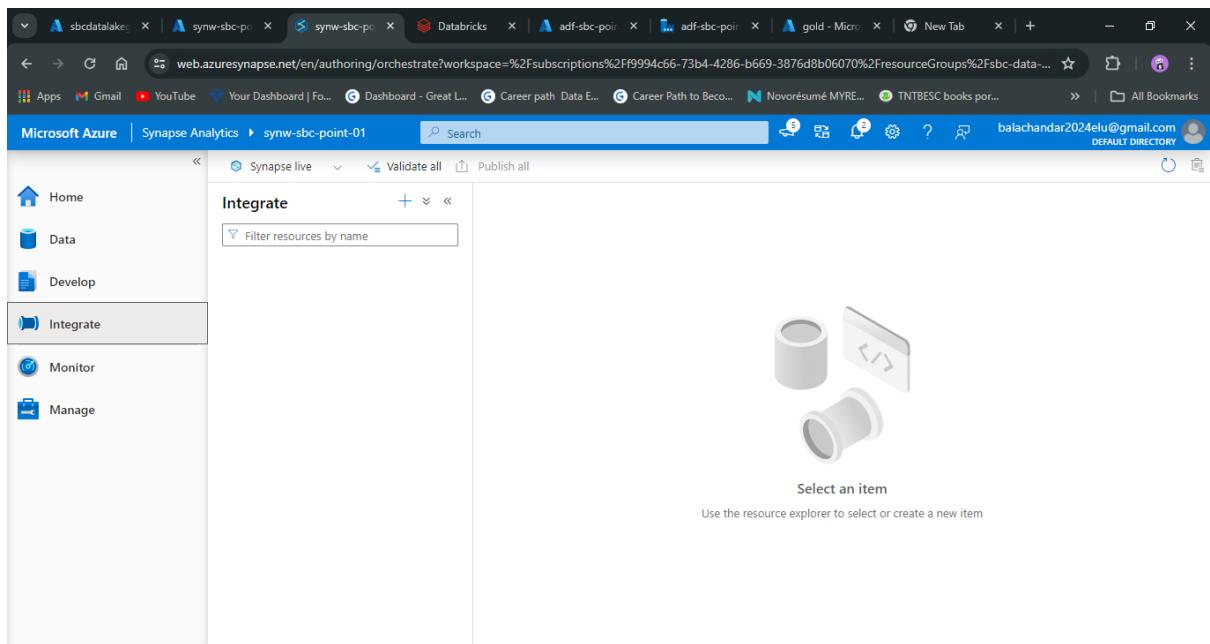
You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (1)

NAME	CHANGE	EXISTING
serverlessSQLdb	(New)	-

Publish Cancel

Type here to search



Microsoft Azure | Synapse Analytics > synw-sbc-point-01

Search

New integration dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

azur

All Azure Database File

Azure Blob Storage Azure Data Lake Storage Gen2 Azure Data Lake Storage Gen2

SQL SQL

Continue Cancel

The screenshot shows the Microsoft Azure Synapse Analytics interface. A dialog box titled 'New integration dataset' is open, prompting the user to select a data store. The search bar contains 'azur'. Below it, tabs for 'All', 'Azure', 'Database', and 'File' are shown. Under the 'Azure' tab, three options are listed: 'Azure Blob Storage', 'Azure Data Lake Storage Gen2' (which is highlighted with a blue border), and 'Azure Data Lake Storage Gen2'. At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

Browse

Select a file or folder.

Root folder > gold > SalesLT

- Address
- Customer
- CustomerAddress
- Product
- ProductCategory
- ProductDescription
- ProductModel
- ProductModelProductDescription
- SalesOrderDetail
- SalesOrderHeader

Showing 1 - 10 of 10 items

OK Cancel

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

Set properties

Name
goldtables

Linked service *
synw-sbc-point-01-WorkspaceDefaultStorage

Connect via integration runtime *
AutoResolveIntegrationRuntime

File path
gold / SalesLT / File name

Advanced

Open this dataset for more advanced configuration with parameterization.

OK Back Cancel

The screenshot shows the Microsoft Azure Synapse Analytics Pipeline Editor interface. On the left, a navigation sidebar includes Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays a pipeline named "Pipeline 1". The "Activities" section lists a single "Get Metadata" activity under the "General" category. The "Dataset" tab is selected, showing a dropdown menu set to "goldtables". The "Field list" tab is active, showing a table structure with one column labeled "Child items". A tooltip message at the top right states: "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." The status bar at the bottom indicates "12:28 PM 26-04-2024".

The screenshot shows the Microsoft Azure Synapse Analytics Pipeline Editor interface. On the left, a navigation sidebar includes Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays a pipeline named "Pipeline 1". The "Activities" pane shows a single activity: "Get Metadata" under the "General" category. The "Dataset" tab is selected, showing a dropdown menu set to "goldtables". The "Field list" section contains a "New" button and a "Delete" button. Below these are sections for "Argument" and "Child items". A status message at the top right indicates that other users in the workspace may have access to modify the item.

Pipeline 1

Integrate

Activities

Get Metadata

Get Tablenames

Dataset: goldtables

Field list

Argument

Child items

General Settings User properties

Validate Debug Add trigger

Get Tablenames

Get Metadata

Get Tablenames

Output

Pipeline run ID: 8eae2085-4148-4b66-812a-315f24ea71fd

Pipeline status: In progress

All status

Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start
Get Tablenames	Queued	Get Metadata	4/26/2024, 1:

0.3 KB/s
0.1 KB/s

12:31 PM
26-04-2024

The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor interface. On the left, a navigation sidebar lists Home, Data, Develop, Integrate, Monitor, and Manage. The main area is titled "Integrate" and shows a list of Pipelines. Under "Pipeline 1", there are two activities: "Get Metadata" and "Get Tablenames". The "Output" tab is selected, displaying the results of the "Get Tablenames" activity. The output JSON is:

```
{ "childItems": [ { "name": "Address", "type": "Folder" }, { "name": "Customer", "type": "Folder" } ] }
```

The pipeline status is "Succeeded".

The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. The pipeline is named "Pipeline 1". In the "Activities" pane, a single "Get Metadata" activity is selected, with its sub-activity "Get Tablenames" also highlighted. The pipeline status is "Succeeded".

Pipeline 1 Activities:

- Get Metadata (Get Tablenames)

Pipeline 1 Settings:

- General: Sequential
- Settings: Items * (This property should be parameterized)

Pipeline 1 Run History:

Activity name	Activity status	Activity type	Run start	Duration
Get Tablenames	Succeeded	Get Metadata	4/26/2024, 12:31:11 PM	24s

Screenshot of the Microsoft Azure Synapse Analytics Pipeline expression builder.

The pipeline is named "Pipeline 1". The "Activities" section shows a "Get Table" activity selected. The expression builder window contains the following code:

```
@activity('Get Tablenames').output.childItems
```

The "Activity outputs" tab shows the following items:

- Get Tablenames (Get Tablenames activity output)
- Get Tablenames (Get Tablenames pipeline return value)

Buttons at the bottom of the expression builder window include "OK" and "Cancel".

The screenshot shows the Microsoft Azure Synapse Analytics pipeline configuration interface. The pipeline is named "Pipeline 1". The activities pane shows a single activity: "For each table name" which contains a "Stored procedure" activity named "Stored procedure1". The "General" tab is selected, showing the activity state as "Activated". The "Settings" tab is also visible, providing options for linked services, integration runtime, and stored procedure name. A note in the settings tab indicates that to reference a SQL pool, a SQL pool stored procedure should be used instead.

Integrate

Pipelines

Pipeline 1

Activities

Stored procedure

Name: Stored procedure1

Description:

Activity state: Activated

General

Settings

Linked service: serverlessSQLdb

Integration runtime: AutoResolveIntegrationRuntime

Stored procedure name: [dbo].[CreateSQLServerlessView_gold]

The screenshot shows the Microsoft Azure Synapse Analytics pipeline configuration interface. The pipeline is named "Pipeline 1". The activities section displays a "For each table name" loop, which contains a "Stored procedure" activity named "Stored procedure1". The "General" tab of the stored procedure settings is selected, showing the stored procedure name as "[dbo].[CreateSQLServerlessView_gold]" and an "Enter manually" checkbox. Below this, the "Stored procedure parameters" section is visible, featuring an "Import" button and a "New" button. The pipeline expression builder window is open, showing the expression "@item().name" entered in the main text area. The "ForEach iterator" tab is selected in the pipeline expression builder. The system status bar at the bottom indicates "12:42 PM 26-04-2024".

The screenshot shows the Microsoft Azure Synapse Analytics interface for Pipeline 1. The pipeline consists of a single activity: "For each table name". This activity is configured to run a "Stored procedure" named "[dbo].[CreateSQLServerlessView_gold]". A parameter named "ViewName" is defined, with its value set to "@item().name". The pipeline has one run history entry, dated 26-04-2024 at 12:44 PM.

Pipeline 1 Activities

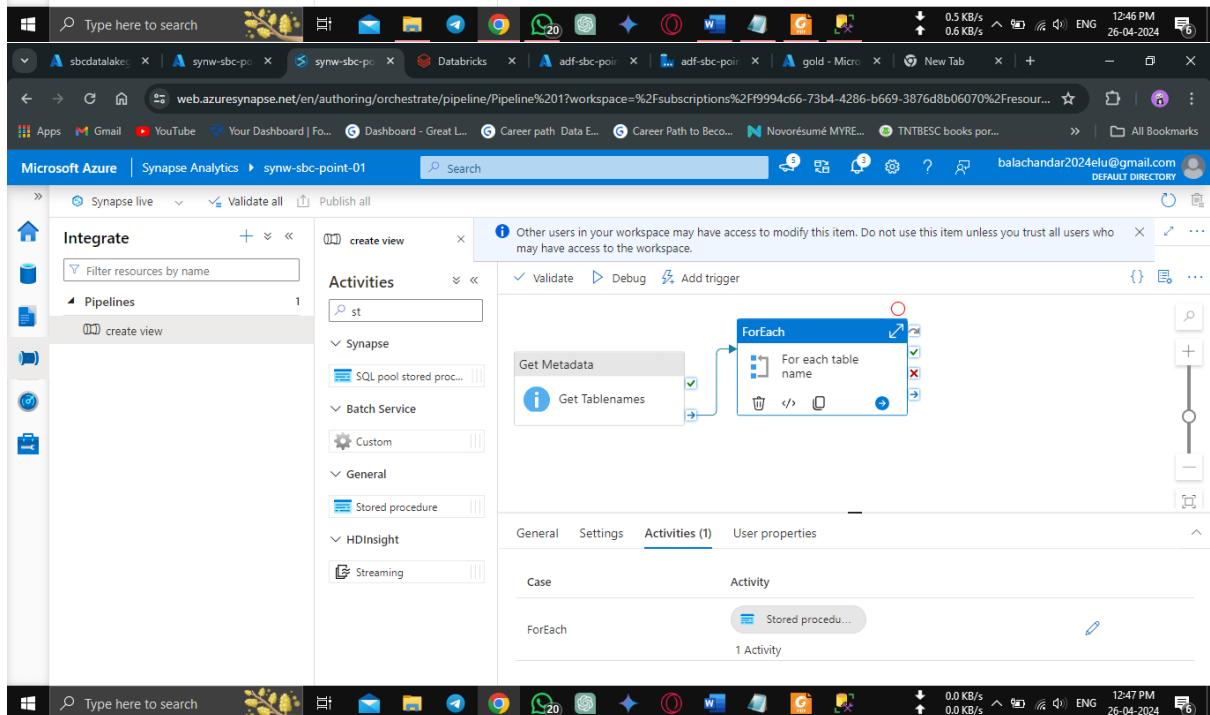
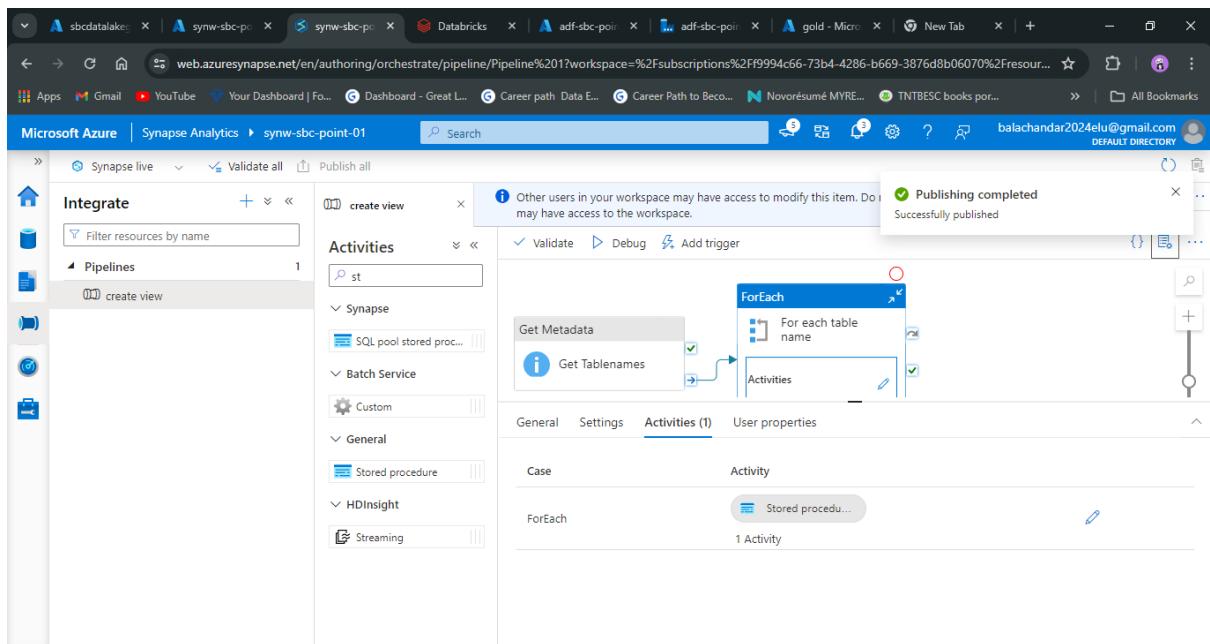
- For each table name
 - Stored procedure

General Settings User properties

Stored procedure name: [dbo].[CreateSQLServerlessView_gold]

Stored procedure parameters:

Name	Type	Value
ViewName	String	@item().name



Screenshot of Microsoft Azure Synapse Analytics Pipeline Run interface.

The pipeline run dialog shows:

- Pipeline run**: Trigger pipeline now using last published configuration.
- Parameters**: No records found.

The pipeline editor shows a workflow with the following activities:

```
graph TD; A[Get Metadata: Get Tablenames] --> B[ForEach: For each table name]
```

The pipeline run status is **Running**, indicating successful execution.

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

Pipeline runs

Triggered Debug Rerun Cancel options Refresh Edit columns List Gantt

Filter by run ID or name: Chennai, Kolkata, Mu... ; Last 24 hours Pipeline name: All Status: All

Runs: Latest runs Triggered by: All Add filter

Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run
create view	4/26/2024, 12:48:08 PM	--	1m 42s	Manual trigger	In progress	Original

Analytics pools: SQL pools, Apache Spark pools, Data Explorer pools (preview)

Activities: SQL requests, KQL requests, Apache Spark applications, Data flow debug

Integration: Pipeline runs, Trigger runs, Integration runtimes, Link connections

Microsoft Azure | Synapse Analytics > synw-sbc-point-01

All pipeline runs > create view - Activity runs

Re-run Cancel Refresh Update pipeline List Gantt

Get Metadata ForEach

Get Tablenames For each table name Activities

Monitor in Azure Metrics Export to CSV

All status List

Showing 1 - 12 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Get Tablenames	Succeeded	Get Metadata	4/26/2024, 1:00:39 PM	3s	AutoResolveIntegration	
For each table name	In progress	ForEach	4/26/2024, 1:00:43 PM	17s		
Stored procedure1	In progress	Stored procedure	4/26/2024, 1:00:44 PM	16s		
Stored procedure1	In progress	Stored procedure	4/26/2024, 1:00:44 PM	16s		
Stored procedure1	In progress	Stored procedure	4/26/2024, 1:00:44 PM	16s		
Stored procedure1	In progress	Stored procedure	4/26/2024, 1:00:44 PM	16s		

Analytics pools: SQL pools, Apache Spark pools, Data Explorer pools (preview)

Activities: SQL requests, KQL requests, Apache Spark applications, Data flow debug

Integration: Pipeline runs, Trigger runs, Integration runtimes, Link connections

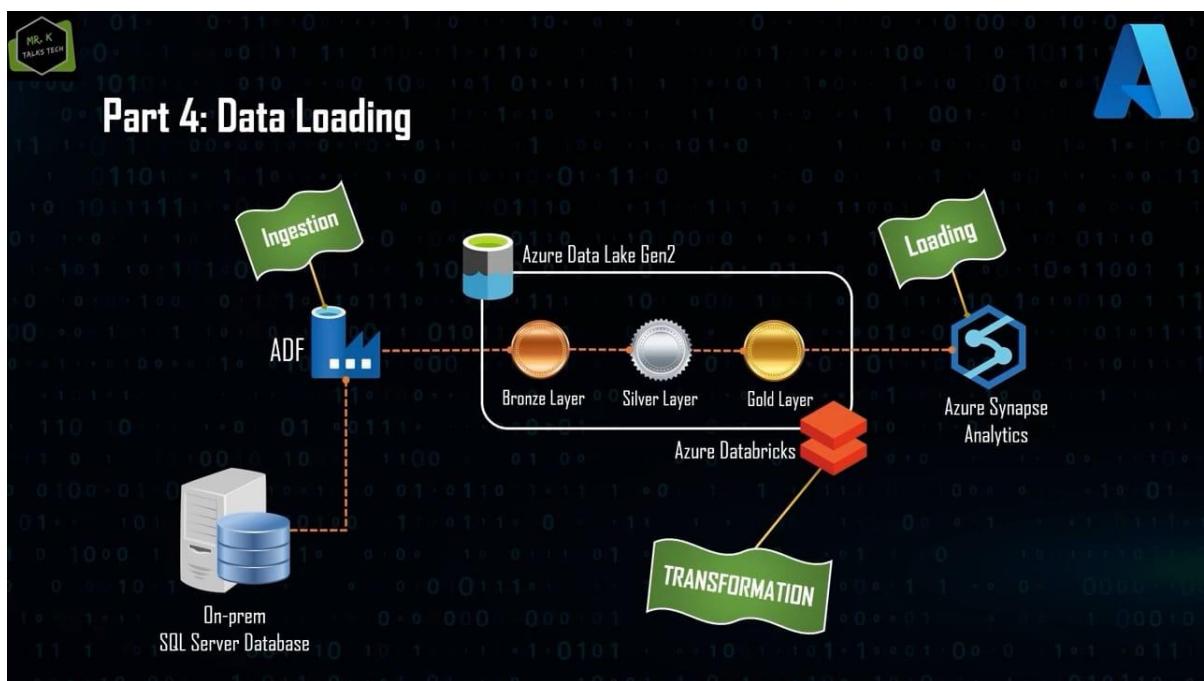
Screenshot 1: Azure Synapse Analytics Pipeline Runs

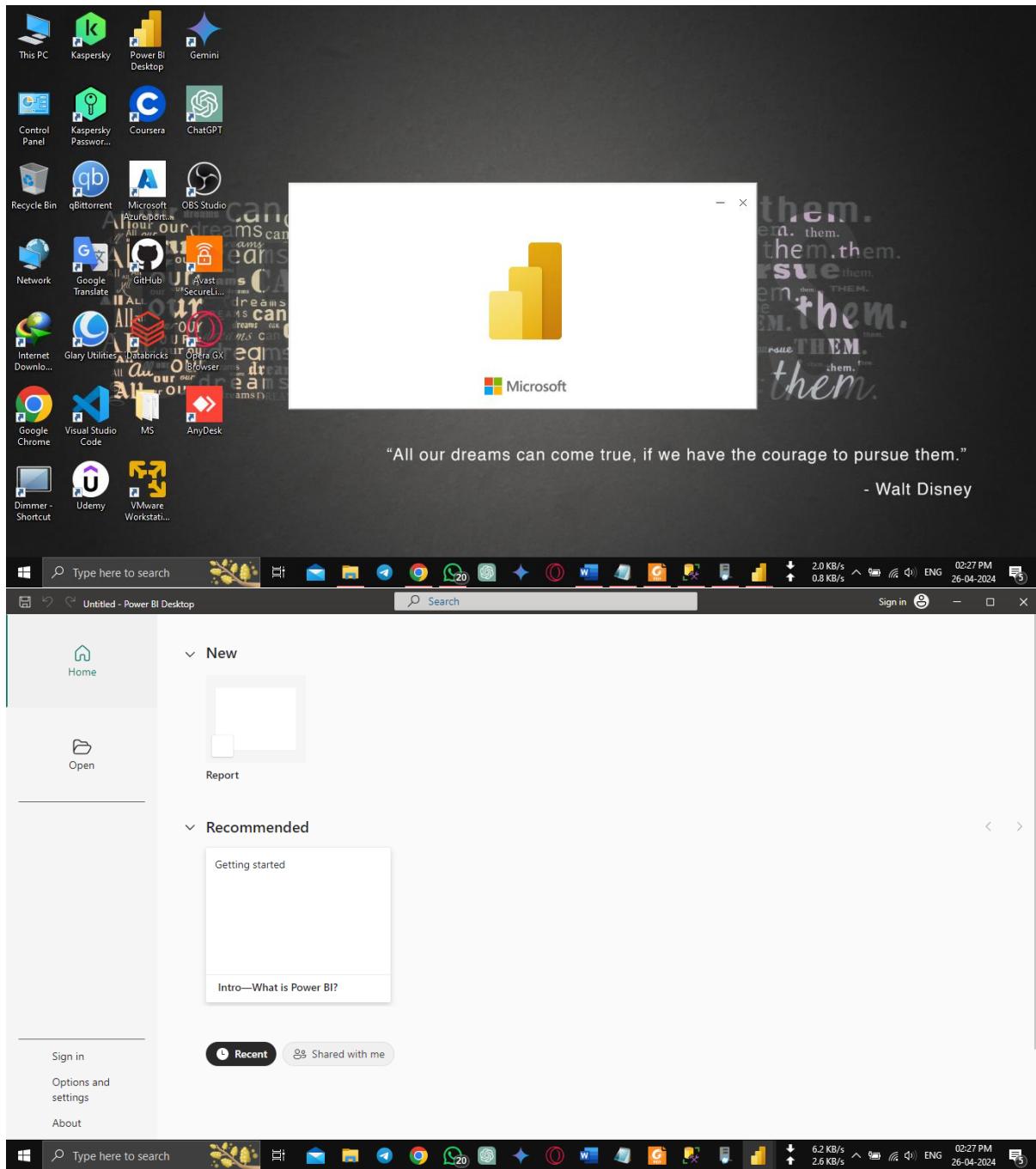
Screenshot 2: Azure Synapse Analytics Pipeline Authoring

The screenshot shows the Microsoft Azure Synapse Analytics workspace. A pipeline named "create view" is being edited. The pipeline structure is as follows:

- Get Metadata** activity (selected)
- Followed by a **ForEach** loop activity.
- Inside the **ForEach** loop:
 - Activity: **For each table name**
 - Activity: **Activities** (with a **Stored procedure...** option selected)

The pipeline is currently in **Validate** mode. The left sidebar shows the workspace structure under the "Data" category, including Views and Schemas.





The screenshot shows the Power BI Desktop interface. The main area displays the message "Add data to your report" with the sub-instruction "Once loaded, your data will appear in the Data pane." Below this are four buttons: "Import data from Excel", "Import data from SQL Server", "Paste data into a blank table", and "Use sample data". A link "Get data from another source →" is also present. The ribbon at the top has tabs for File, Home, Insert, Modeling, View, Optimize, and Help. The Home tab is selected. The Data pane on the left is visible, showing a list of items like "Clipboard", "Get data", "Excel", "OneLake", "Server", "Enter data", "Data", "Transform data", "Queries", "New visual", "Text box", "More visuals", "Insert", "Calculations", "Sensitivity", "Publish", and "Share". The Visualizations pane on the right lists various visualization types such as Bar chart, Line chart, Map, Gauge, etc., under categories like Build visual, Values, and Drill-through.

Add data to your report
Once loaded, your data will appear in the Data pane.

Import data from Excel Import data from SQL Server Paste data into a blank table Use sample data

Get data from another source →

File Home Insert Modeling View Optimize Help

Clipboard Get data Excel OneLake Server Enter data Data Transform data Queries New visual Text box More visuals Insert Calculations Sensitivity Publish Share

Visualizations Build visual

Filters

Data

Values

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Page 1 of 1

The screenshot shows the "Get Data" dialog box in Power BI Desktop. The left sidebar lists categories: All, File, Database, Microsoft Fabric, Power Platform, Azure, Online Services, and Other. The main pane shows a list of data sources under the "All" category, including Excel Workbook, Text/CSV, XML, JSON, Folder, PDF, Parquet, SharePoint folder, SQL Server database, Access database, SQL Server Analysis Services database, Oracle database, IBM Db2 database, IBM Informix database (Beta), IBM Netezza, and MySQL database. At the bottom of the dialog are buttons for "Certified Connectors", "Template Apps", "Connect", and "Cancel". The ribbon at the top is identical to the previous screenshot, showing the Home tab selected. The Data pane and Visualizations pane are also visible on the right.

Get Data

All

All

File

Database

Microsoft Fabric

Power Platform

Azure

Online Services

Other

Excel Workbook

Text/CSV

XML

JSON

Folder

PDF

Parquet

SharePoint folder

SQL Server database

Access database

SQL Server Analysis Services database

Oracle database

IBM Db2 database

IBM Informix database (Beta)

IBM Netezza

MySQL database

Certified Connectors

Template Apps

Connect

Cancel

Page 1 of 1

File Home Insert Modeling View Optimize Help

Clipboard Get data Excel OneLake Server Enter data Data Transform data Queries New visual Text box More visuals Insert Calculations Sensitivity Publish Share

Visualizations Build visual

Filters

Data

Values

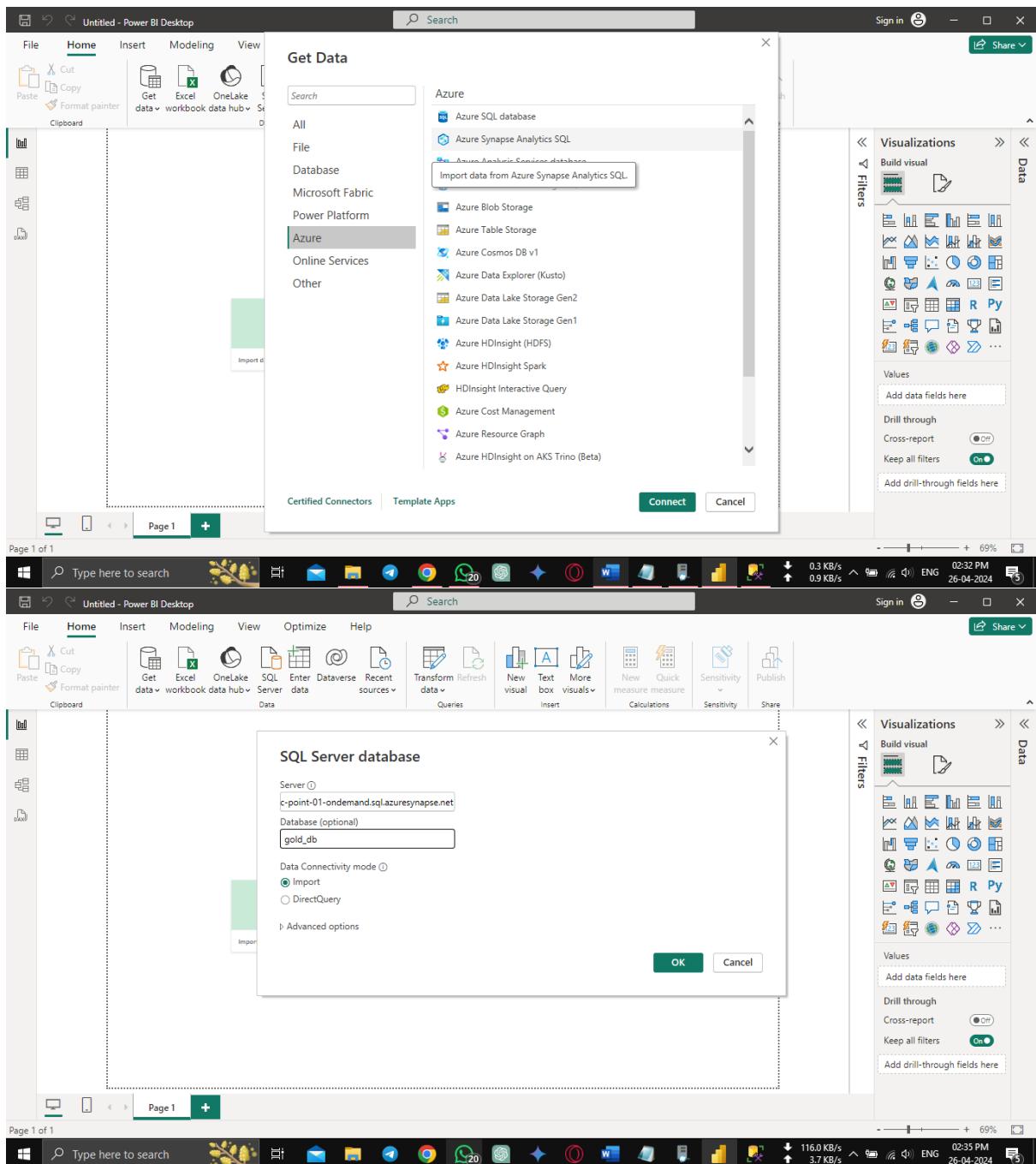
Drill through

Cross-report

Keep all filters

Add drill-through fields here

Page 1 of 1



The screenshot shows two instances of the Power BI Desktop application running side-by-side on a Windows desktop.

Top Window (Foreground):

- Home Tab:** Selected.
- Clipboard:** Shows a placeholder for data imports.
- Data Tab:** Shows options for "Get data", "OneLake", "SQL Server", "Enter data", "Dataverse", and "Recent sources".
- Queries Tab:** Shows a "Transform data" button and a "Refresh data" button.
- Visualizations Tab:** Shows a "Build visual" button and a grid of visualization icons.
- Filters Tab:** Shows sections for "Values", "Drill through", "Cross-report", and "Keep all filters".

A modal dialog box titled "Database" is open, prompting for sign-in information:

- Sign-in method: Microsoft account.
- Target database: synw-sbc-point-01-on-demand.sql.azuresynapse.net...
- Sign in as different user: Option available.
- Select which level to apply these settings to: synw-sbc-point-01-on-demand.sql.azuresynapse.net...
- Connect and Cancel buttons.

Bottom Window (Background):

- File Tab:** Selected.
- Clipboard:** Shows a placeholder for data imports.
- Navigator Tab:** Selected.
- Display Options:** Set to "synw-sbc-point-01-on-demand.sql.azuresynapse.net..."
- Content Area:** Shows a tree view of tables:
 - address
 - Customer
 - CustomerAddress
 - Product
 - ProductCategory
 - ProductDescription
 - ProductModel
 - ProductModelProductDescription
 - SalesOrderDetail
 - SalesOrderHeader
- Visualizations Tab:** Shows a "Build visual" button and a grid of visualization icons.
- Filters Tab:** Shows sections for "Values", "Drill through", "Cross-report", and "Keep all filters".

The taskbar at the bottom of both windows shows the Windows Start button, a search bar, and various pinned application icons.

Untitled - Power BI Desktop

Navigator

Display Options

- symw-sbc-point-01-on-demand.sql.azuresynaps...
- address
- Customer
- CustomerAddress
- Product
- ProductCategory
- ProductDescription
- ProductModel
- ProductModelProductDescription
- SalesOrderDetail
- SalesOrderHeader

SalesOrderHeader

Freight	TotalDue	Comment	rowguid	Modi...
22.0087	null	null	89e42cdc-8506-48a2-b89b-eb3e64e3554e	2008-
1.9703	null	null	8a3448c5-e677-4158-a29b-d3d3069b0e00	2008-
960.4672	null	null	a47665d2-7ac5-4cf3-8ab2-a3a83554284	2008-
994.6333	null	null	f1be45a5-5c57-4a50-93c6-5f8be44cb7cb	2008-
2096.4607	null	null	7db2329e-6446-42a8-8915-9c8370b68e08	2008-
2714.0458	null	null	ca31f324-2c32-4f8d-95e6-7f5a6b611c38	2008-
1440.8659	null	null	917ef5ba-f32d-4f63-8588-66db0bcd846	2008-
1950.7422	null	null	bb3fe84-c8bf-4d02-ccca-675ab6b11c38	2008-
28.5395	null	null	2aa5f39b-1096-4a4b-b17b-f10504a397ce	2008-
84.9541	null	null	e3c189e7-98de-4c40-b6c2-0d1d13f9b33	2008-
50.4085	null	null	625d76fc-c26f-4149-b2a4-939fb2bccd77	2008-
894.3803	null	null	addbd8620-432a-456e-8470-1beddd4bc3457	2008-

Visualizations

Build visual

Data

Filters

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Untitled - Power BI Desktop

File Home Insert Modeling View Optimize Help

Clipboard

Queries

Load

There are pending changes in your queries that haven't been applied.

Once loaded, you can apply them by clicking **Load**.

Import data from Excel

Import data from Power BI

Import data from OneLake

Import data from SQL Server

Import data from Data

Import data from Recent sources

New visual

Text box

More visuals

Refresh data

Transform data

Load

Transform Data

Cancel

Visualizations

Build visual

Data

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Page 1 of 1

Type here to search

18.6 KB/s
7.1 KB/s

02:44 PM
26-04-2024

Untitled - Power BI Desktop

File Home Insert Modeling View Optimize Help

Clipboard

Queries

Load

address
Waiting for other queries...

Customer
Waiting for other queries...

CustomerAddress
Evaluating...

Product
Evaluating...

ProductCategory
Evaluating...

Visualizations

Build visual

Data

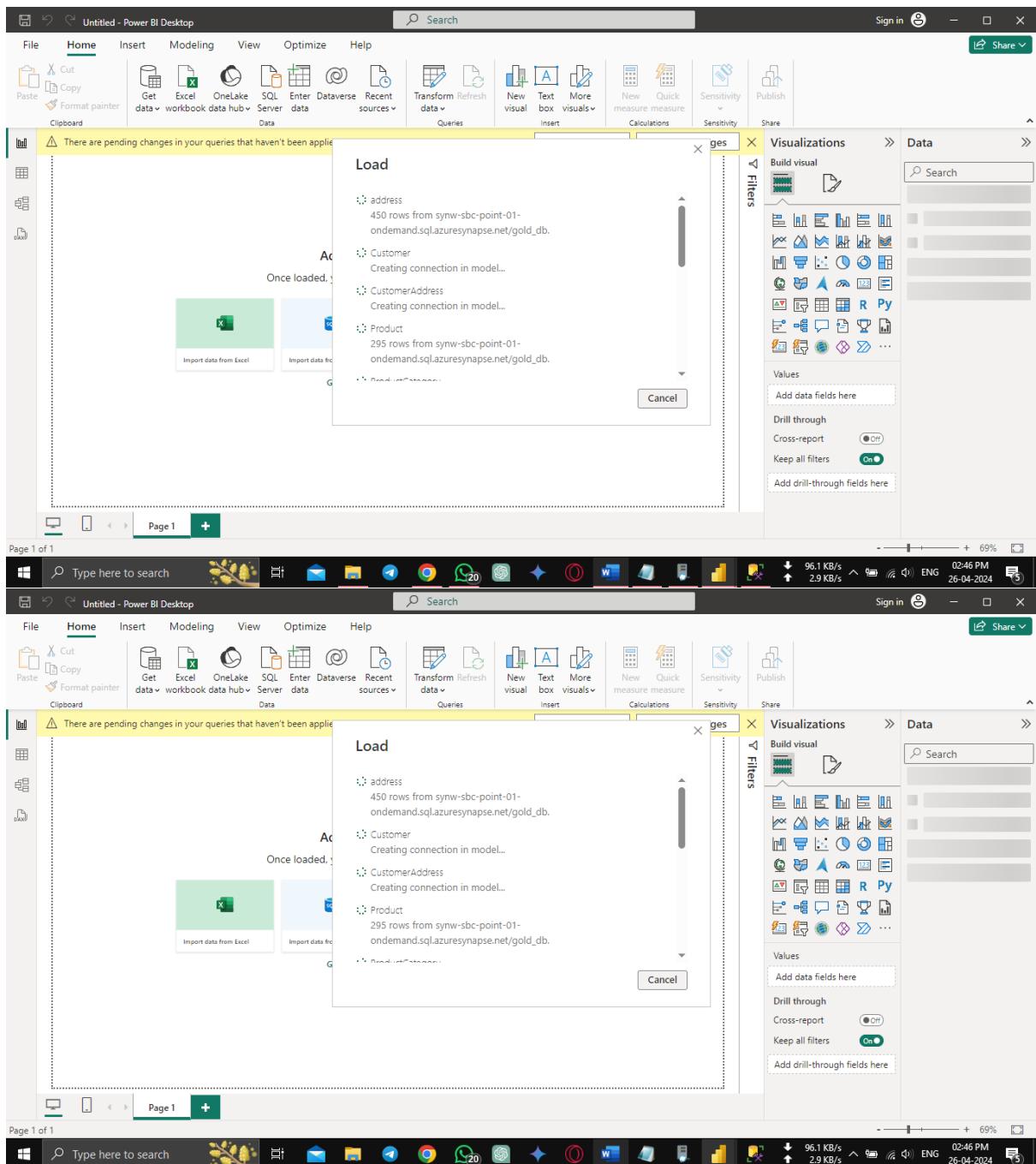
Search

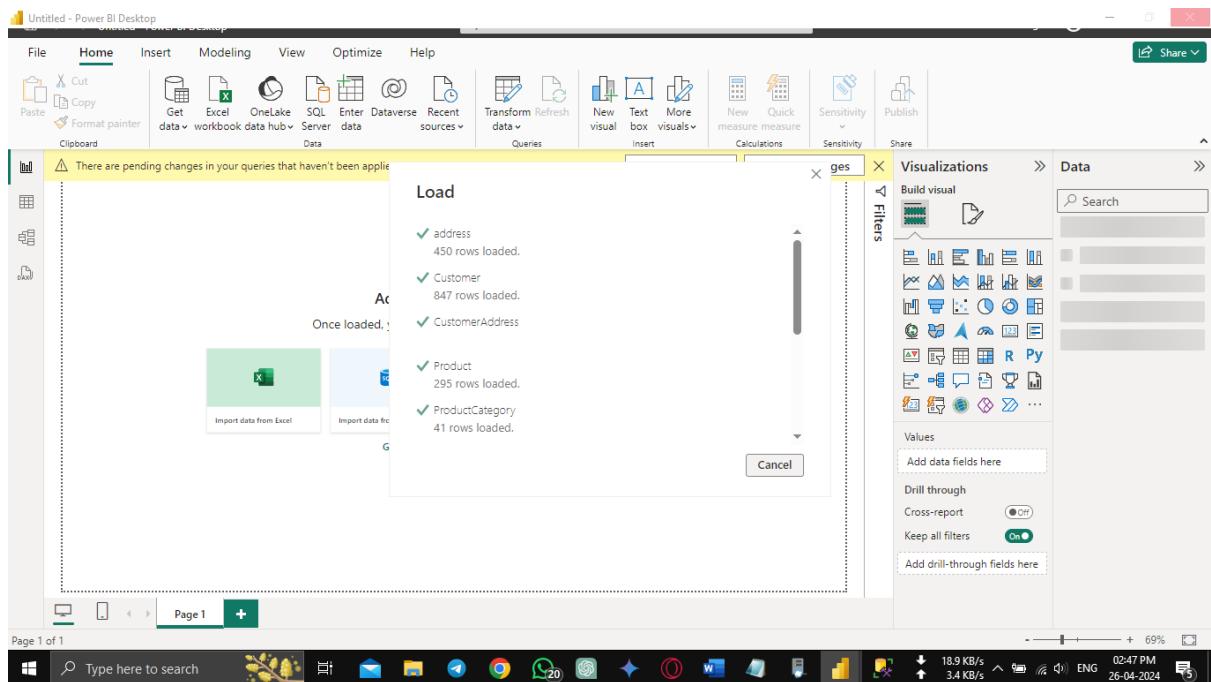
Page 1 of 1

Type here to search

12.8 KB/s
5.5 KB/s

02:45 PM
26-04-2024





Untitled - Power BI Desktop

File Home Help Table tools

Name Customer

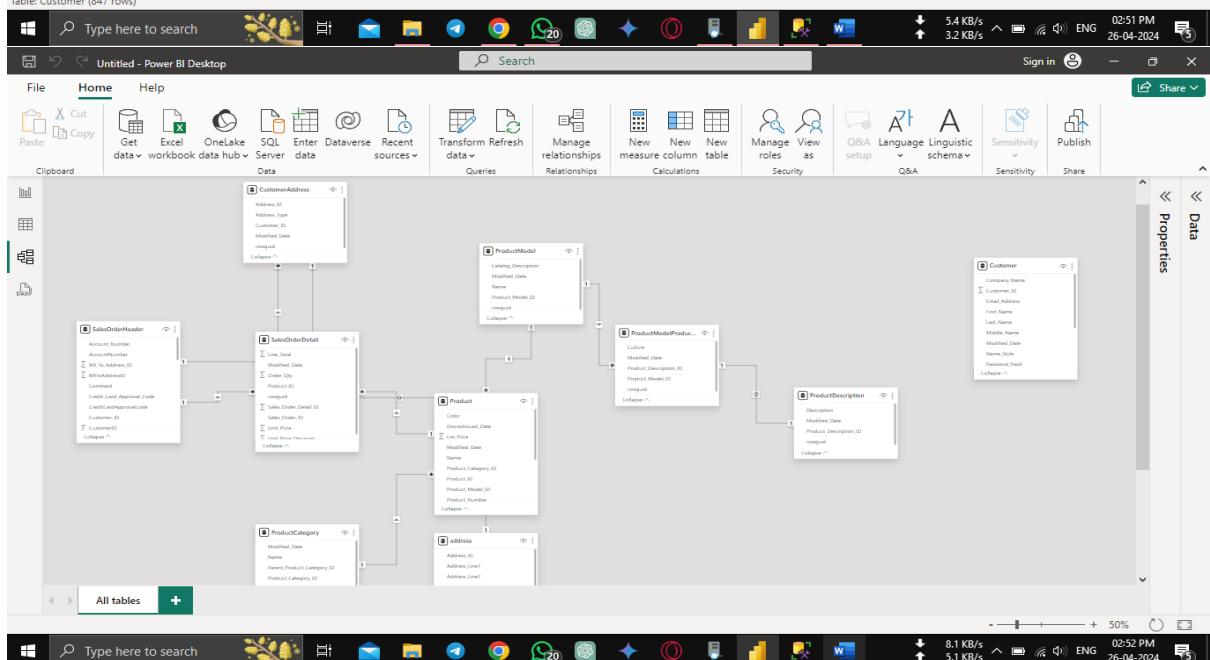
Structure

Mark as date table Calendars Relationships New measure Quick New measure column New table Calculations

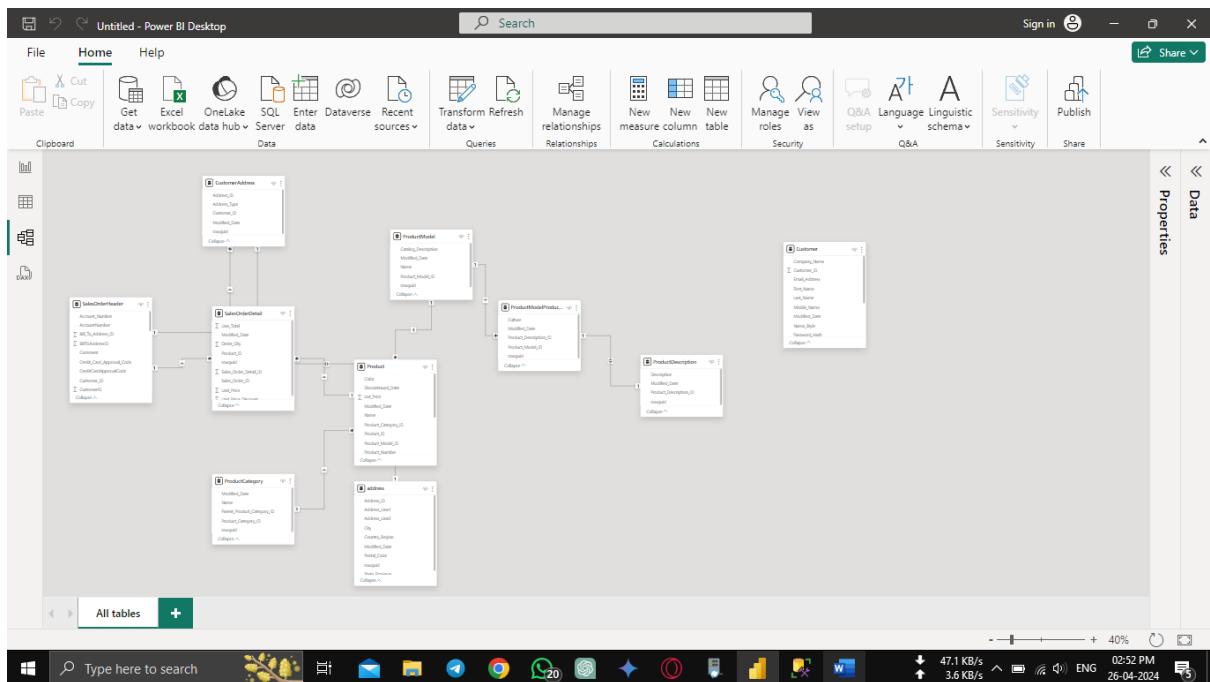
Customer_ID Name_Style Title First_Name Middle_Name Last_Name Suffix Company_Name Sales_Person Email_Address

Customer_ID	Name_Style	Title	First_Name	Middle_Name	Last_Name	Suffix	Company_Name	Sales_Person	Email_Address
2	False	Mr.	Keith		Harris		Progressive Sports	adventure-works'david8	keith@adventure
5	False	Mr.	Lucy		Harrington		Metropolitan Sports Supply	adventure-works'shu0	lucy0@adventure
23	False	Mr.	Kerim		Hanif		Bike World	adventure-works'shu0	kerim@adventure
24	False	Mr.	Kevin		Liu		Eastside Department Store	adventure-works'linda3	kevin5@adventure
29	False	Mr.	Bryan		Hamilton		Cross-Country Riding Supplies	adventure-works'jose1	bryan7@adventure
37	False	Mr.	Jim		Geist		Two Bike Shops	adventure-works'pamela0	jim1@adventure
40	False	Mr.	Darren		Gehring		Journey Sporting Goods	adventure-works'jillian0	daren0@adventure
42	False	Mr.	Jeremy		Los		Healthy Activity Store	adventure-works'linda3	jeremy0@adventure
55	False	Mr.	Eric		Lang		Kickstands and Accessories Company	adventure-works'jose1	eric6@adventure
56	False	Mr.	Brian		Groth		Latest Accessories Sales	adventure-works'david8	brian7@adventure
58	False	Mr.	Peter		Kurniawan		Largest Bike Store	adventure-works'jillian0	peter4@adventure
59	False	Mr.	Douglas		Groncki		Locks Company	adventure-works'shu0	douglas2@adventure
61	False	Mr.	Jeffrey		Kurtz		Many Bikes Store	adventure-works'shu0	jeffrey3@adventure
64	False	Mr.	Vamsi		Kuppa		Metal Clearing House	adventure-works'jose1	vamsi1@adventure
70	False	Mr.	Deepak		Kumar		Outdoor Aerobic Systems Company	adventure-works'jae0	deepak0@adventure
77	False	Mr.	Michael		Brundage		Mechanical Products Ltd.	adventure-works'shu0	michael13@adventure
78	False	Mr.	Stefan		Delmarco		Preferred Bikes	adventure-works'linda3	stefan0@adventure
79	False	Mr.	Mitch		Kennedy		Reasonable Bicycle Sales	adventure-works'shu0	mitch0@adventure
93	False	Mr.	Prashanth		Desai		Stationary Bikes and Stands	adventure-works'jillian0	prashanth0@adventure
94	False	Mr.	Scott		Konermann		Specialty Sports Store	adventure-works'jillian0	scott6@adventure
97	False	Mr.	Eugene		Kogan		Mountain Bike Center	adventure-works'shu0	eugen2@adventure
118	False	Mr.	Yale		Li		Rapid Bikes	adventure-works'jose1	yale0@adventure

Table: Customer (847 rows)



Data modelling was done for linked Tables.



Untitled - Power BI Desktop

File Home Help

Paste Cut Copy Get data Excel OneLake workbook data hub SQ Data

Clipboard

Manage relationships

Active From: Table (Column) To: Table (Column)

- CustomerAddress (Address_ID) address (Address_ID)
- CustomerAddress (Customer_ID) SalesOrderHeader (Customer_ID)
- Product (Product_Category_ID) ProductCategory (Product_Category_ID)
- Product (Product_Model_ID) ProductModel (Product_Model_ID)
- ProductModelProductDescription (Product_Description_ID) ProductDescription (Product_Description_ID)
- ProductModelProductDescription (Product_Model_ID) ProductModel (Product_Model_ID)
- SalesOrderDetail (Product_ID) Product (Product_ID)
- SalesOrderDetail (Sales_Order_ID) SalesOrderHeader (Sales_Order_ID)

New... Autodetect... Edit... Delete Close

All tables +

CustomerAddress (Address_ID) address (Address_ID)
Customer (Customer_ID) SalesOrderHeader (Customer_ID)
Product (Product_Category_ID) ProductCategory (Product_Category_ID)
Product (Product_Model_ID) ProductModel (Product_Model_ID)
ProductModelProductDescription (Product_Description_ID) ProductDescription (Product_Description_ID)
ProductModelProductDescription (Product_Model_ID) ProductModel (Product_Model_ID)
SalesOrderDetail (Product_ID) Product (Product_ID)
SalesOrderDetail (Sales_Order_ID) SalesOrderHeader (Sales_Order_ID)

CustomerAddress (Address_ID) address (Address_ID)
Customer (Customer_ID) Customer (Customer_ID)
SalesOrderHeader (Customer_ID) SalesOrderHeader (Customer_ID)
Product (Product_Category_ID) ProductCategory (Product_Category_ID)
Product (Product_Model_ID) ProductModel (Product_Model_ID)
ProductModelProductDescription (Product_Description_ID) ProductDescription (Product_Description_ID)
ProductModelProductDescription (Product_Model_ID) ProductModel (Product_Model_ID)
SalesOrderDetail (Product_ID) Product (Product_ID)
SalesOrderDetail (Sales_Order_ID) SalesOrderHeader (Sales_Order_ID)

File Home Help

Paste Cut Copy Get data Excel OneLake workbook data hub SQ Data

Clipboard

Manage relationships

Active From: Table (Column) To: Table (Column)

- CustomerAddress (Address_ID) address (Address_ID)
- CustomerAddress (Customer_ID) Customer (Customer_ID)
- CustomerAddress (Customer_ID) SalesOrderHeader (Customer_ID)
- Product (Product_Category_ID) ProductCategory (Product_Category_ID)
- Product (Product_Model_ID) ProductModel (Product_Model_ID)
- ProductModelProductDescription (Product_Description_ID) ProductDescription (Product_Description_ID)
- ProductModelProductDescription (Product_Model_ID) ProductModel (Product_Model_ID)
- SalesOrderDetail (Product_ID) Product (Product_ID)
- SalesOrderDetail (Sales_Order_ID) SalesOrderHeader (Sales_Order_ID)

New... Autodetect... Edit... Delete Close

All tables +

CustomerAddress (Address_ID) address (Address_ID)
Customer (Customer_ID) Customer (Customer_ID)
SalesOrderHeader (Customer_ID) SalesOrderHeader (Customer_ID)
Product (Product_Category_ID) ProductCategory (Product_Category_ID)
Product (Product_Model_ID) ProductModel (Product_Model_ID)
ProductModelProductDescription (Product_Description_ID) ProductDescription (Product_Description_ID)
ProductModelProductDescription (Product_Model_ID) ProductModel (Product_Model_ID)
SalesOrderDetail (Product_ID) Product (Product_ID)
SalesOrderDetail (Sales_Order_ID) SalesOrderHeader (Sales_Order_ID)

Power BI Desktop - Untitled

Data Model View

```

    graph TD
        SOH[SalesOrderHeader] --> SOD[SalesOrderDetail]
        SOD --> PMS[ProductModel]
        SOD --> PA[CustomerAddress]
        PA --> PM[ProductModel]
        PM --> P[Product]
        PC[ProductCategory] --> P
        P --> PD[ProductDescription]
        C[Customer] --> PA
    
```

Table Tools - Product

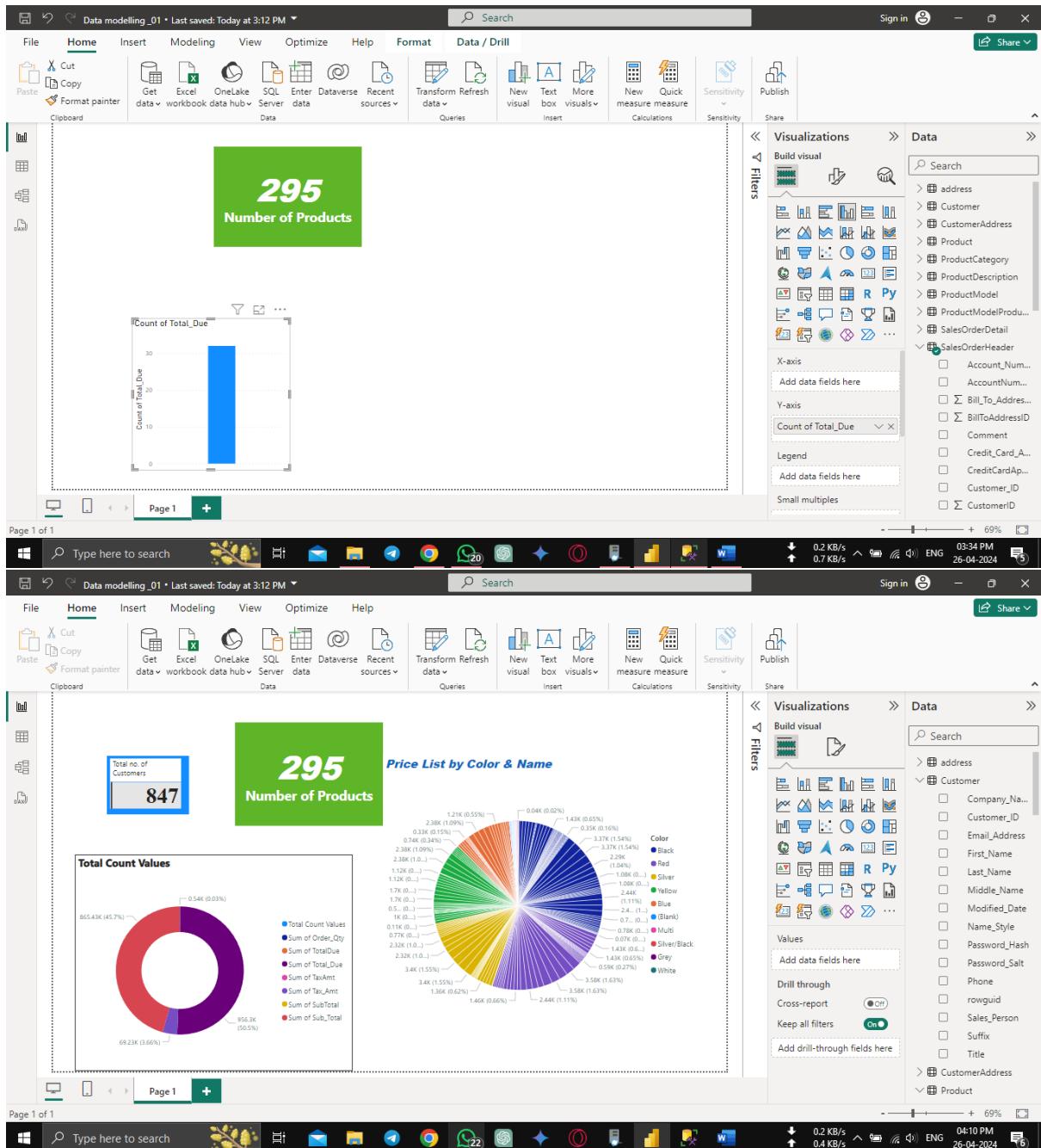
Data

Product_ID	Name	Product_Number	Color	Standard_Cost	List_Price	Size	Weight	Product_Category_ID	Product_Model_ID	Sell_ID
707	Sport-100 Helmet, Red	HL-U509-R	Red	13.0863	34.99			35	33 20	
708	Sport-100 Helmet, Black	HL-U509	Black	13.0863	34.99			35	33 20	
709	Mountain Bike Socks, M	SO-B909-M	White	3.3963	9.5	M		27	18 20	
710	Mountain Bike Socks, L	SO-B909-L	White	3.3963	9.5	L		27	10 20	
711	Sport-100 Helmet, Blue	HL-U509-B	Blue	13.0863	34.99			35	33 20	
712	AWC Logo Cap	CA-1098	Multi	6.9223	8.99			25	11 20	
713	Long-Sleeve Logo Jersey, S	LJ-0192-S	Multi	38.4923	49.99	S		25	11 20	
714	Long-Sleeve Logo Jersey, M	LJ-0192-M	Multi	38.4923	49.99	M		25	11 20	
715	Long-Sleeve Logo Jersey, L	LJ-0192-L	Multi	38.4923	49.99	L		25	11 20	
716	Long-Sleeve Logo Jersey, XL	LJ-0192-X	Multi	38.4923	49.99	XL		25	11 20	
802	LL Fork	FK-1639		65.8097	148.22			14	104 20	
803	ML Fork	FK-5136		77.9176	175.49			14	105 20	
804	HL Fork	FK-9939		101.8936	229.49			14	106 20	
805	LL Headset	HS-0296		15.1848	34.2			15	59 20	
806	ML Headset	HS-2451		45.4168	102.29			15	60 20	
807	HL Headset	HS-3479		55.3801	124.73			15	61 20	
808	LL Mountain Handlebars	HB-M243		19.7758	44.54			8	52 20	
809	ML Mountain Handlebars	HB-M763		27.4925	61.92			8	54 20	
810	HL Mountain Handlebars	HB-M918		53.3999	120.27			8	55 20	
811	LL Road Handlebars	HB-R504		19.7758	44.54			8	56 20	
812	ML Road Handlebars	HB-R720		27.4925	61.92			8	57 20	
813	HL Road Handlebars	HB-R956		53.3999	120.27			8	58 20	

Table: Product (295 rows)

The screenshot shows the Power BI desktop interface. A single value visual displays the number 295 with the text "Count of Product_ID". The ribbon menu is open, showing options like File, Home, Insert, Modeling, View, Optimize, Help, Format, and Data / Drill. The Data pane on the right lists fields from the "Product" table, including Product_ID (selected), Color, Discontinued..., List_Price, Modified_Date, Name, Product_Cat..., Product_ID, Product_Mod..., Product_Num..., rowguid, Sell_End_Date, Sell_Start_Date, Size, and Standard_Cost. The status bar at the bottom indicates the file was last saved at 3:12 PM on April 26, 2024.

This screenshot shows the same Power BI desktop environment as the first one. However, the context menu for the "Count of Product_ID" field in the Data pane is open, displaying various aggregation options such as Sum, Average, Minimum, Maximum, Count (Distinct), Count, Standard deviation, Variance, Median, Show value as, and New quick measure. The rest of the interface, including the visual on the canvas and the ribbon, remains identical to the first screenshot.



Dashboard created and pipeline executed after the dashboard creation and the results are reflecting after the pipeline run execution done..

Screenshot of Microsoft Azure Portal showing the Home page and a detailed view of a Resource Group.

Microsoft Azure Home Page:

- Recent resources: sbc-data-engineering-project-02, adf-sbc-point-01, synw-sbc-point-01, dbw-sbc-point-01, sbcdatalakegen02, sbc-data1-valut-001, Free Trial.
- Services: Create a resource, Azure Synapse Analytics, Microsoft Entra ID, Azure Databricks, Storage accounts, Data factories, Key vaults, Resource groups, Virtual machines, More services.

sbc-data-engineering-project-02 Resource Group Overview:

- Overview tab selected.
- Essentials section shows 5 resources:

Name	Type	Location
adf-sbc-point-01	Data factory (V2)	East US
dbw-sbc-point-01	Azure Databricks Service	East US
sbc-data1-valut-001	Key vault	East US
sbcdatalakegen02	Storage account	East US
synw-sbc-point-01	Synapse workspace	East US

The screenshot shows two stacked windows of the Microsoft Azure portal.

The top window displays the 'Default Directory | Groups' page. The left sidebar under 'Manage' has 'Groups' selected. The main area shows a list of groups: External identities, Roles and administrators, Administrative units, Delegated admin partners, Enterprise applications, Devices, App registrations, and Identity Governance. The 'External identities' group is currently expanded.

The bottom window shows the 'Add owners' blade for creating a new group. The 'Group type' field is set to 'data-engineer-group'. The 'Name' field contains 'data-engineer-group'. The 'Membership type' field is set to 'Assigned'. Under 'Owners', there is one result found: 'Balachandar Saravanan' (Type: User, Email: balachandar2024elu@gmail.com). A 'Select' button is at the bottom right of the blade.

Screenshot of Microsoft Azure portal showing the creation of a new group named "data-engineer-group".

New Group

Group name: data-engineer-group

Group description: Enter a description for the group

Membership type: Assigned

Owners: 1 owner selected

Members: 1 member selected

Create

Groups | All groups

Name	Object Id	Group type	Membership type
data-engineer-group	20de0259-c0db-4f06-b945-a18977d74593	Security	Assigned

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes links for Apps, Gmail, YouTube, Your Dashboard, Dashboard - Great L..., Career path Data E..., Career Path to Beco..., Novorésumé MYRE..., TNTBESC books por..., and All Bookmarks. The user's email address, balachandar2024elu@g... and DEFAULT DIRECTORY (BALACHA...) are also visible.

The main content area displays the "sbc-data-engineering-project-02 | Access control (IAM)" blade. The left sidebar lists options: Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings (Deployments, Security, Deployment stacks, Policies, Properties, Locks), and a search bar. The "Access control (IAM)" section is selected.

The main panel shows the "Number of role assignments for this subscription" (6) and "Privileged" (0). It includes a "View assignments" button and filters for All, Job function (0), and Privileged (0). A search bar allows filtering by name or email, and buttons for Type: All, Role: All, Scope: All scopes, and Group by: Role.

The "Role assignments" table is empty, showing the message "No user assignments exist".

The bottom navigation bar includes links for Add role, portal.azure.com/view/Microsoft_Azure_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Ff9994c66-73b4-4286-b669-3876d8b06070/resourceGroups/sbc-data-engineering-project-02/providers/Microsoft.Authorization/roleAssignments, and All Bookmarks. The system status bar at the bottom right shows 2.0 KB/s, 1.9 KB/s, ENG, 04:26 PM, and 26-04-2024.

The screenshot shows the "Add role assignment" blade. The top navigation bar and user information are identical to the previous screenshot.

The main content area shows the "Role" tab selected, with "Members" highlighted. Other tabs include Conditions and Review + assign. A note states: "A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. [Learn more](#)".

The "Privileged administrator roles" section is selected under Job function roles. It describes the "Privileged administrator access" role, which grants the ability to assign roles to other users. A warning message states: "⚠️ Can a job function role with less access be used instead?"

A search bar allows filtering by role name, description, permission, or ID, with Type: All and Category: All selected. The results table lists three roles:

Name	Description	Type	Category	Details
Owner	Grants full access to manage all resources, including the ability to assign roles in Azure RBAC.	BuiltinRole	General	View
Contributor	Grants full access to manage all resources, but does not allow you to assign roles in Azure RBAC, man...	BuiltinRole	General	View
Access Review Operator Service Role	Lets you grant Access Review System app permissions to discover and revoke access as needed by the...	BuiltinRole	None	View

Buttons at the bottom include Review + assign, Previous, Next, and Next. The bottom navigation bar and system status bar are identical to the first screenshot.

Add role assignment

Role **Members** **Conditions** **Review + assign**

Selected role Contributor

Assign access to User, group, or service principal Managed identity

Members [+ Select members](#)

Name	Object ID	Type
data-engineer-group	20de0259-c0db-4f06-b945-a18977d74593	Group

Description

Review + assign **Previous** **Next**

Review + assign

Role **Members** **Conditions** **Review + assign**

Role Contributor

Scope /subscriptions/f9994c66-73b4-4286-b669-3876d8b06070/resourceGroups/sbc-data-engineering-project-02

Members

Name	Object ID	Type
data-engineer-group	20de0259-c0db-4f06-b945-a18977d74593	Group

Description No description

Review + assign **Previous** **Next**

Review + assign

Role **Members** **Conditions** **Review + assign**

Role Contributor

Scope /subscriptions/f9994c66-73b4-4286-b669-3876d8b06070/resourceGroups/sbc-data-engineering-project-02

Members

Name	Object ID	Type
data-engineer-group	20de0259-c0db-4f06-b945-a18977d74593	Group

Description No description

Review + assign **Previous** **Next**

The image displays two side-by-side screenshots of Microsoft Azure interfaces.

Top Screenshot (Microsoft Azure IAM):

- URL:** portal.azure.com/#@balachandar2024elugmail.onmicrosoft.com/resource/subscriptions/f9994c66-73b4-4286-b669-3876d8b06070/resourceGroups/...
- Resource Group:** sbc-data-engineering-project-02
- Section:** Access control (IAM)
- Role Assignments:** Number of role assignments for this subscription: 7 / 4000. Privileged: 1.
- Table:** Shows one role assignment for 'data-engineer-group' as a Contributor to 'This resource'.

Bottom Screenshot (Microsoft Azure Data Factory):

- URL:** adf.azure.com/en/authoring/pipeline/copy_all_tables?factory=%2Fsubscriptions%2F9994c66-73b4-4286-b669-3876d8b06070%2FresourceGroups%2Fsb...
- Factory Resources:** Pipelines, Datasets, Data flows, Power Query.
- Pipeline:** copy_all_tables
- Trigger:** Trigger now
- Activities:** Lookup, Copy Each Table, Notebooks (Bronze to Silver and Silver to Gold).

The screenshot shows the Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. The main workspace displays a pipeline named 'copy_all_tables'. The pipeline consists of a 'Lookup' activity followed by a 'ForEach' loop. Inside the 'ForEach' loop is an 'Activities' section containing a 'Copy Each Table' activity. The 'Parameters' tab is selected. On the right, a 'New trigger' dialog is open, prompting for a name ('scheduled_trigger'), type ('Schedule'), start date ('4/26/2024, 1:20:53 PM'), time zone ('Eastern Time (US & Canada) (UTC-5)'), and recurrence ('Every 1 Day(s)').

This screenshot is identical to the one above, showing the Azure Data Factory pipeline editor and the 'New trigger' dialog. The difference is that the 'Advanced recurrence options' section is now expanded, revealing fields for 'Execute at these times' (Hours: 0, Minutes: 25), 'Schedule execution times' (00:25), and a checkbox for 'Specify an end date'. Other sections like 'Annotations' and 'Start trigger' are also visible.

Screenshot of Microsoft Azure Data Factory Pipeline Editor showing the 'copy_all_tables' pipeline.

The pipeline consists of the following steps:

- A **Lookup** activity named "Look for all tables".
- An **ForEach** loop activity.
- Inside the **ForEach** loop:
 - An **ForEach Table** activity.
 - An **Activities** group containing a **Copy Each Table** activity.

The pipeline is currently in the "Pending changes" state, with one trigger named "scheduled_trigger" listed.

Publish all dialog box is open, showing the pending changes and a "Publish" button.

Published Pipeline View:

- The pipeline is now published and running.
- A watermark message "EyeLeo" and "Short break! Blink your eyes." is displayed over the pipeline diagram.
- The pipeline status shows "8" tasks completed.
- A success message "Successfully published" is displayed in a toast notification.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar is titled "Author" and includes sections for Home, Monitor, Manage, and Learning Center. The main content area is titled "Pipeline runs" and displays a table of runs. The table has columns for Pipeline name, Run start, Run end, Duration, and Triggered by. One row is visible: "copy_all_tables" with a run start of 5/1/2024, 12:03:00 AM, a run end of 5/1/2024, 12:03:06 AM, a duration of 6s, and triggered by "scheduled_trigger". The status bar at the bottom indicates "Last refreshed 5 minutes ago".

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar is titled "Monitor" and includes sections for Home, Author, Manage, and Learning Center. The main content area is titled "Trigger runs" and displays a table of runs. The table has columns for Trigger name, Trigger type, Trigger time, Status, Pipelines, Run, and Message. One row is visible: "scheduled_trigger" with a trigger type of "Schedule trigger", a trigger time of 5/1/2024, 12:02:59 A, a status of "Succeeded", one pipeline run, and the message "Original". The status bar at the bottom indicates "Last refreshed 5 minutes ago".

Scheduled trigger also ran successfully .Results are visible in the Dashboard..

Power BI Desktop - Data Reporting

Last saved: 4/26/2024 at 12:43 PM

File Home Insert Modeling View Optimize Help Format Data / Drill

Cut Copy Format painter

Get data Excel OneLake SQL Server Data workbook data hub data recent sources Transform Refresh data New visual Text box More Quick measure Sensitivity Publish Share

Clipboard

Total no. of Customers: 847

Number of Products: 295

Price List by Color & Name

Total Count Values

Donut chart showing Total Count Values:

- Total Count Values: 865.4K (45.7%)
- Sum of Order_Qty: 69.23K (3.66%)
- Sum of TotalDue: 95.63K (50.5%)

Donut chart showing Price List by Color & Name:

Color	Value	Percentage
Black	1.43K	0.65%
Red	1.35K	0.61%
Silver	1.33K	0.55%
Yellow	1.31K	0.54%
Blue	1.29K	0.53%
(Blank)	1.28K	0.53%
Grey	1.27K	0.53%
Multi	1.26K	0.53%
Silver/Black	1.25K	0.53%
White	1.24K	0.53%

Visualizations

Build visual

Filters

Search

Customer CustomerAddress Product ProductCategory ProductDescription ProductModel ProductModelProduct SalesOrderDetail SalesOrderHeader

Columns

Total no. of Customers Last_Name

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Page 1

9.3 kB/s 2.6 kB/s ENG 04:58 AM 01-05-2024

The screenshot shows the Microsoft Azure Resource Groups page. At the top, there's a search bar and a navigation bar with links like 'All bookmarks' and the user's email 'balachandar2024delhi@g...'. Below the header, the main content area displays 'Resource groups' with a table listing two records. The table has columns for Name, Subscription, and Location. The first row shows 'NetworkWatcherRG' with 'Free Trial' subscription and 'East US' location. The second row shows 'sbc-data-engineering-project-02' with 'Free Trial' subscription and 'East US' location. The bottom of the table includes sorting options ('Name ↑', 'Subscription ↑↓', 'Location ↑↓') and a 'List view' button.

Name	Subscription	Location
NetworkWatcherRG	Free Trial	East US
sbc-data-engineering-project-02	Free Trial	East US

At the bottom left, there are navigation buttons for 'Previous', 'Page 1 of 1', and 'Next >'. On the right side, there's a 'Notifications' sidebar with a heading 'Notifications' and a message 'More events in the activity log →'. It lists three recent events: 'Deleting resource group sbc-data-engineering-project-02' (Running), 'Deleting resource group NetworkWatcherRG' (Running), and 'Deleted resource group databricks-rg-dbw-sbc-point-01' (Succeeded). There's also a message about credit remaining and a link to upgrade to a Pay-As-You-Go subscription.

Don't forget to delete the created resource and avoid the Over Billing...

Resources

Recent

Favorite

Name _____

Type

Last Viewed

[sbcdatalakegen2](#)

Storage account

10 hours ago

[dbw-sbc-point-01](#)

Azure Databricks Service

23 hours ago

[sbc-data-engineering-project-02](#)

Resource group

23 hours ago

[adf-sbc-point-01](#)

Data factory (V2)

2 days ago

[sbc-data-valut-001](#)

Key vault

4 days ago

[See all](#)

-----Project Over-----