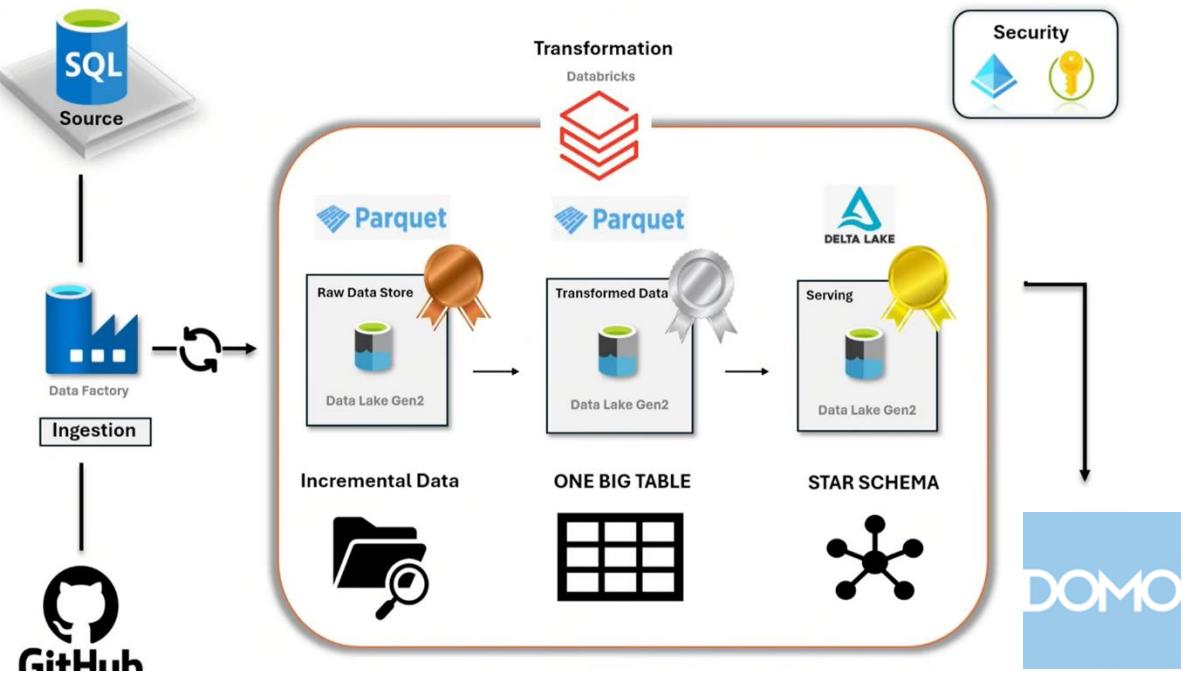


# Data Bricks Project Work

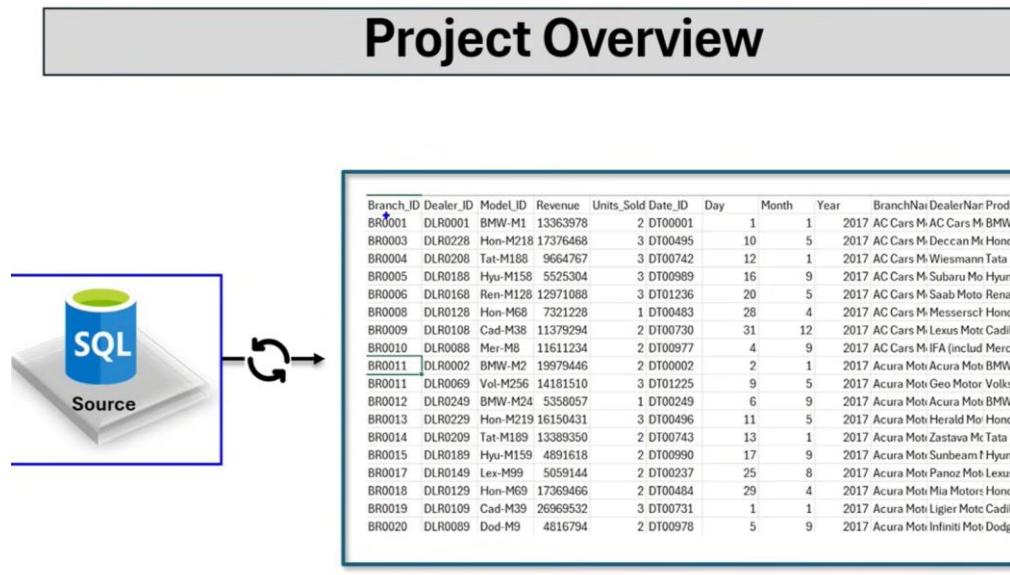
## 1) Architecture



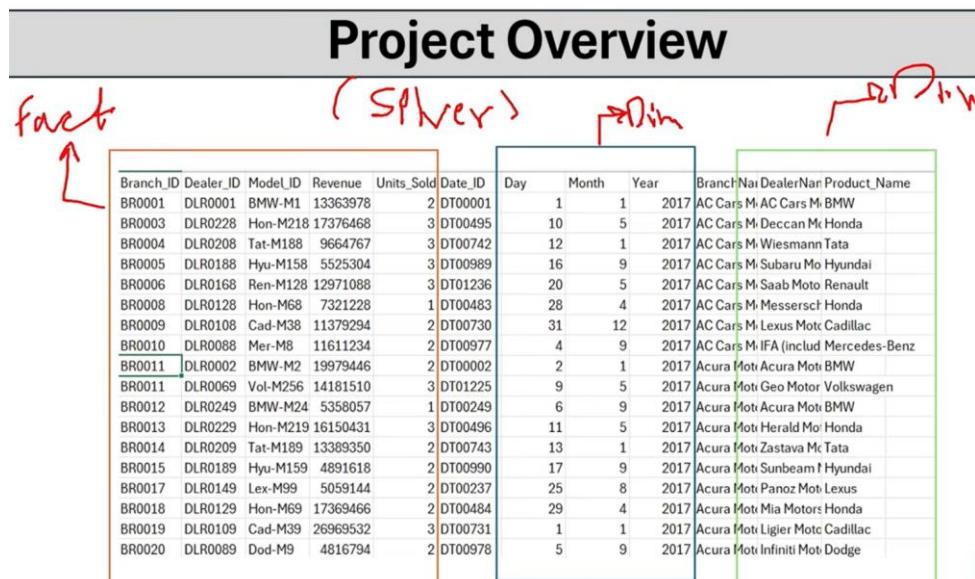
- 1) Data Source – Azure SQL Data Source
- 2) Git hub Repository -Pull the data from Github Repository to Azure SQL Data Source
- 3) ETL Tools -Azure Data Factory → After Azure SQL data source to connect pipeline to Azure data factory. It is advance Pipeline
- 4) Bronze layer→ Data Factory to add data in bronze layer with increment load data only. That is why is advance data pipeline create CDC(Change data capture) ,lookup files and store procedure
- 5) Parquet →it is file format
- 6) After land bronze layer then use Databricks. The databricks transform the data in silver layer and add data in Delta Lake format or gold layer
- 7) Unity Catalog → it is unity is the unified Solution govern of the data to apply data quality check and more. **It is hint**

Project Overview:

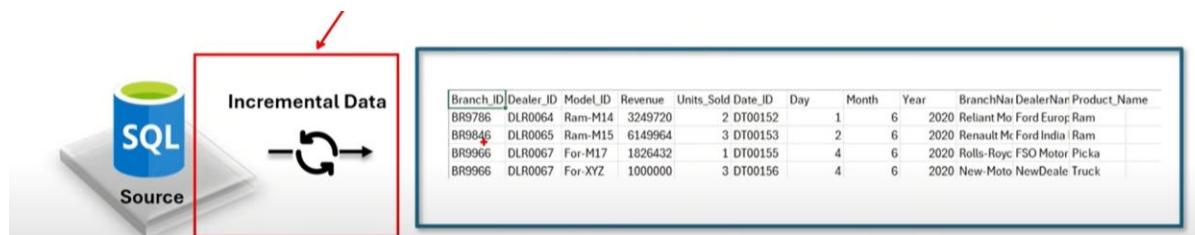
- 1) This is fact table data, which is present in the Bronze layer



- 2) Silver layer – fact table into fact and dimension table and finally connect each dimension table in one fact table



- 3) Data will load in Incremental load. Because past data cannot load again and again

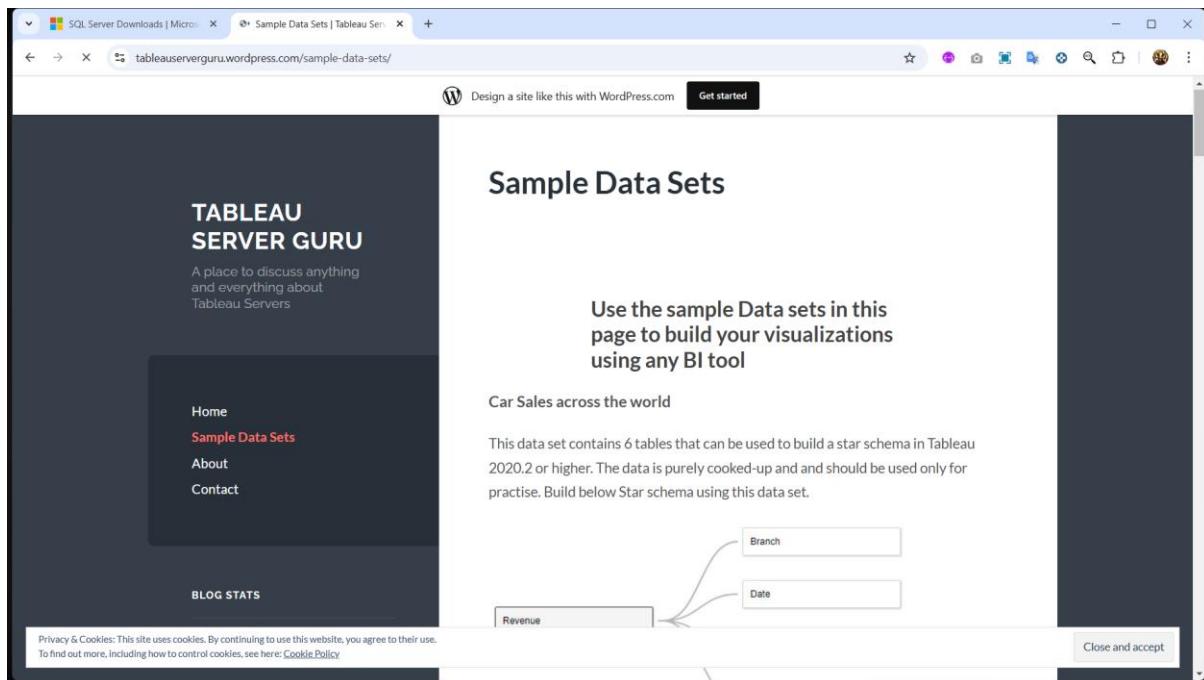


## Prerequisite:

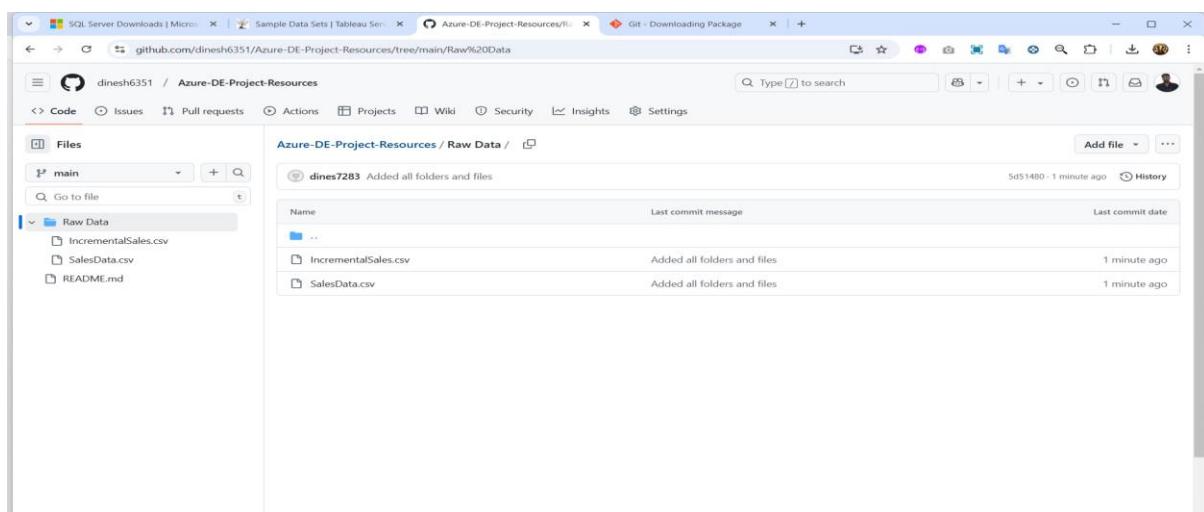
### Prerequisite

- Laptop or PC with stable internet connection
- Azure Account (Don't worry will create one)
- Excitement to learn Azure Data Solution

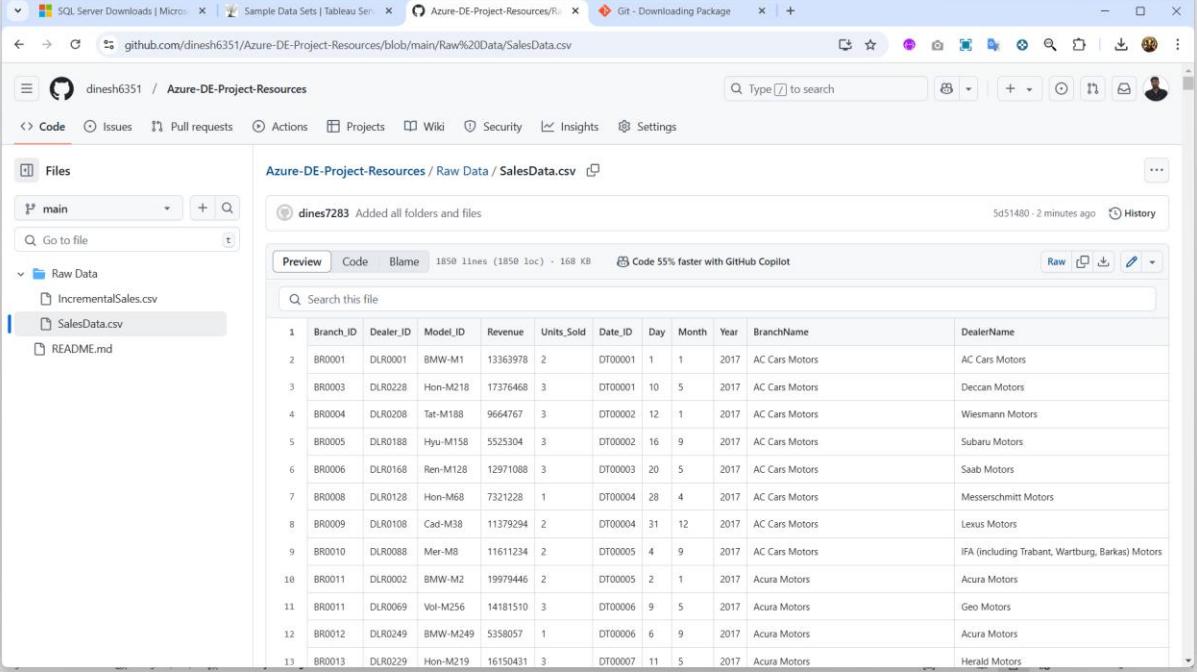
Sample Dataset Website-



--Github Dataset



- Sales Data have present all Transactions records
- Sample

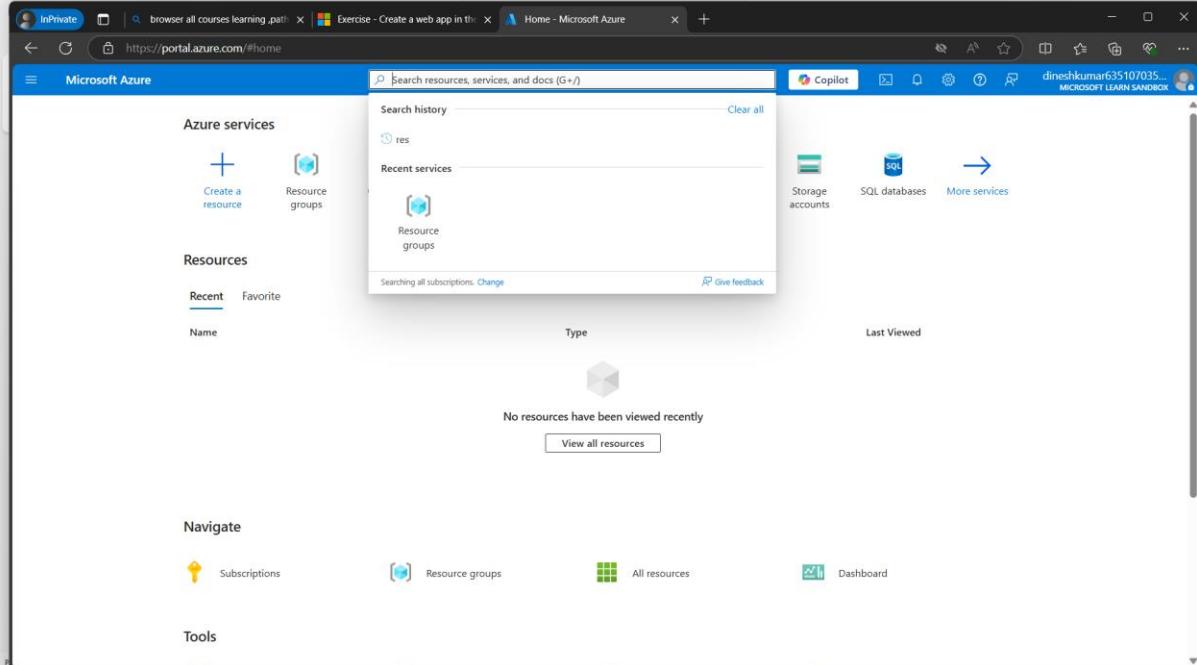


The screenshot shows a GitHub repository named "Azure-DE-Project-Resources". The "Raw Data" folder contains two files: "IncrementalSales.csv" and "SalesData.csv". The "SalesData.csv" file is displayed in a table format. The table has 13 columns: Branch\_ID, Dealer\_ID, Model\_ID, Revenue, Units\_Sold, Date\_ID, Day, Month, Year, BranchName, and DealerName. The data consists of 13 rows of sales information for various vehicles across different dealers.

	Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month	Year	BranchName	DealerName
2	BR0001	DLR0001	BMW-M1	13363978	2	DT00001	1	1	2017	AC Cars Motors	AC Cars Motors
3	BR0003	DLR0228	Hon-M218	17376468	3	DT00001	10	5	2017	AC Cars Motors	Deccan Motors
4	BR0004	DLR0208	Tat-M188	9664767	3	DT00002	12	1	2017	AC Cars Motors	Wiesmann Motors
5	BR0005	DLR0188	Hyu-M158	5525304	3	DT00002	16	9	2017	AC Cars Motors	Subaru Motors
6	BR0006	DLR0168	Ren-M128	12971088	3	DT00003	20	5	2017	AC Cars Motors	Saab Motors
7	BR0008	DLR0128	Hon-M68	7321228	1	DT00004	28	4	2017	AC Cars Motors	Messerschmitt Motors
8	BR0009	DLR0108	Cad-M38	11379294	2	DT00004	31	12	2017	AC Cars Motors	Lexus Motors
9	BR0010	DLR0088	Mer-M8	11611234	2	DT00005	4	9	2017	AC Cars Motors	IFA (including Trabant, Wartburg, Barkas) Motors
10	BR0011	DLR0002	BMW-M2	19979446	2	DT00005	2	1	2017	Acura Motors	Acura Motors
11	BR0011	DLR0069	Voi-M256	14181510	3	DT00006	9	5	2017	Acura Motors	Geo Motors
12	BR0012	DLR0249	BMW-M249	5358057	1	DT00006	6	9	2017	Acura Motors	Acura Motors
13	BR0013	DLR0229	Hon-M219	16150431	3	DT00007	11	5	2017	Acura Motors	Herald Motors

## Azure part

--Resource group – store all the information like data lake, catalog , unity and more



The screenshot shows the Microsoft Azure portal home page. The left sidebar includes sections for "Azure services" (Create a resource, Resource groups), "Recent" and "Favorite" resources, and "Tools" (Subscriptions, Resource groups, All resources, Dashboard). The main area displays a search bar, a "Search history" section with a recent entry for "res", and a "Recent services" section showing a "Resource groups" item. Below these are "Storage accounts", "SQL databases", and a "More-services" button. A message at the bottom states "No resources have been viewed recently".

Create resource group:

Home → search resource group

## By create resource group

The screenshot shows the Microsoft Azure portal interface for creating a new resource group. The top navigation bar includes links for 'Create a resource', 'dineshk...', 'Sign in...', 'ASURITI...', 'ASURITI...', 'My ASU...', 'Windows...', 'Sign in...', 'ASU Ho...', 'Microsoft...', and a Copilot button. The user's email, 'dineshkumar7283@gmail.com', is visible in the top right.

The main page title is 'Create a resource group'. Below it, there are three tabs: 'Basics' (selected), 'Tags', and 'Review + create'. A detailed description of a 'Resource group' is provided, stating it's a container for related resources. The 'Subscription' dropdown is set to 'Azure for Students'. The 'Resource group name' input field contains 'RG\_Azure\_Car\_Project'. The 'Region' dropdown is set to '(Europe) UK South'. At the bottom of this section are 'Previous', 'Next', and 'Review + create' buttons.

---

The screenshot shows the 'Resource groups' page. The top navigation bar is identical to the previous screen. The main title is 'Resource groups'. The page displays a single record:

Name	Subscription	Location	Actions
RG_Azure_Car_Project	Azure for Students	UK South	...

Below the table, there are filtering options: 'Filter for any field...' (dropdown), 'Subscription equals all', 'Location equals all', and 'Add filter'. The bottom of the page shows pagination controls: '< Previous', 'Page 1 of 1', 'Next >', and a 'Give feedback' link.

- 1) Create Data Lake
- 2) Not source → because source going to sql database
- 3) After create Data Lake → to create two resource data factory and sql Databases

--Let Start

1) click create → search storage account

Microsoft Azure

RG\_Azure\_Car\_Project

Resource groups

RG\_Azure\_Car\_Project

Overview

Essentials

Subscription (move)  
Azure for Students

Subscription ID  
d7ce8a1-8bd2-4126-ba88-75140d084a4d

Tags

Resource visualizer

Events

Settings

Cost Management

Monitoring

Automation

Help

Search resources, services, and docs (G+/-)

Manage view

Delete resource group

Refresh

Export to CSV

Open query

Assign tags

JSON View

Activity log

Access control (IAM)

Deployments  
No deployments

Location  
UK South

Tags (edit)  
Add tags

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Showing 0 to 0 of 0 records. Show hidden types No grouping List view

Name ↑ Type ↑ Location ↑

No resources match your filters

Give feedback

Microsoft Azure

Marketplace

Get Started

Service Providers

Management

Private Marketplace

Private Offer Management

My Marketplace

Favorites

My solutions

Recently created

Private plans

Categories

Storage (1711)

Compute (894)

IT & Management Tools (727)

Databases (523)

Search resources, services, and docs (G+/-)

Pricing : All Operating System : All Publisher Type : All Product Type : All Publisher name : All

storage

Azure services only

New! Get AI-generated suggestions for 'storage' View suggestions

Showing 1 to 20 of 3452 results for 'storage'. Clear search

IBM Storage Virtualize for Public Cloud Trial

IBM Storage Software

Azure Application

IBM Storage Virtualize for Public Cloud Trial

Price varies

Create

Storage task - Azure Storage Actions

Microsoft

Azure Service

Perform common operations on millions of objects based on logical conditions using object properties for Blobs and Data Lake Storage Gen2.

Price varies

Create

Pure Storage Solution for Microsoft Sentinel

Pure Storage, Inc.

Azure Application

Pure Storage Solution for Microsoft Sentinel helps to protect your array from security risks

Price varies

Create

IBM Storage Virtualize for Public Cloud

IBM Storage Software

Azure Application

IBM Storage Virtualize for Public Cloud

Price varies

Create

Storage account

Microsoft

Azure Service

Use Blobs, Tables, Queues, Files, and Data Lake Gen 2 for reliable, economical cloud storage.

Create

Tile view

Is Marketplace helpful?

## Create storage account and convert into data lake

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

**Project details**

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

Subscription \*

Resource group \*

**Instance details**

Storage account name \*

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

Primary service

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

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

I use LRS for reduce cost. Because data store in same data centre

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

**Project details**

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

Subscription \*

Resource group \*

**Instance details**

Storage account name \*

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

Primary service

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

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

Click next

And **Important** click Enable hierarchical namespace → it means create data lake instance of block storage

--cold check point enable → it is cost very very low it store one month only

→HOT- acces the data every day

→Cool → once a month

→Cold →6 month once

The screenshot shows the 'Create a storage account' wizard in the Azure portal. The current step is 'Configure security settings that impact your storage account'. Key configuration options include:

- Require secure transfer for REST API operations**: Checked.
- Allow enabling anonymous access on individual containers**: Unchecked.
- Enable storage account key access**: Checked.
- Default to Microsoft Entra authorization in the Azure portal**: Unchecked.
- Minimum TLS version**: Set to Version 1.2.
- Permitted scope for copy operations (preview)**: Set to 'From any storage account'.

**Hierarchical Namespace** section: Describes the benefit of enabling hierarchical namespaces for file and directory semantics. The 'Enable hierarchical namespace' checkbox is checked.

**Access protocols** section: Shows options for Blob and Data Lake Gen2 endpoints. The 'Enable SFTP' and 'Enable network file system v3' checkboxes are unchecked.

At the bottom, there are 'Previous' and 'Next' buttons, a 'Review + create' button, and a 'Give feedback' link.

We use HOT – now for learning

This screenshot shows the 'Create a storage account' wizard in the Azure portal, identical to the one above but with different configuration settings. The 'Hierarchical Namespace' section is present but lacks the explanatory text from the first screenshot. The 'Access protocols' section also lacks the explanatory text. The 'Blob storage' section includes a note about cross-tenant replication being disabled when hierarchical namespaces are enabled. The 'Access tier' section shows 'Hot' selected. The 'Azure Files' section has 'Enable large file shares' checked. The navigation and feedback buttons at the bottom are identical to the first screenshot.

## Finally review + create

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The 'Review + create' tab is selected. The configuration details are as follows:

Setting	Value
Subscription	Azure for Students
Resource group	RG_Azure_Car_Project
Location	UK South
Storage account name	cardineshdatastorage
Primary service	Standard
Performance	Locally-redundant storage (LRS)
Replication	Enabled
Advanced	Enabled
Enable hierarchical namespace	Enabled
Enable SFTP	Disabled
Enable network file system v3	Disabled
Allow cross-tenant replication	Disabled
Access tier	Hot
Enable large file shares	Enabled
Security	Enabled

At the bottom, there are 'Previous', 'Next', and 'Create' buttons. The 'Create' button is highlighted in blue.

Finally click create –it is data lake

The screenshot shows the 'cardineshdatastorage\_1741151583800 | Overview' page in the Microsoft Azure portal. The deployment status is 'Deployment is in progress'. Deployment details are listed as follows:

Resource	Type	Status	Operation details
cardineshdatastorage	Microsoft.Storage/storageAccounts	Accepted	<a href="#">Operation details</a>

On the right side, there are promotional links for Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

The screenshot shows the Microsoft Azure Deployment Overview page. The deployment is named "cardinehdatalakestorage\_1741151583800" and is marked as complete. Deployment details include a start time of 3/5/2025, 10:42:48 AM, and a correlation ID of e6bc3a23-a732-4e0a-8f70-1add6661d22d. A sidebar on the right provides links to Cost Management, Microsoft Defender for Cloud, and Azure experts.

Successfully create data lake

The screenshot shows the Microsoft Azure Resource Groups page. The resource group "RG\_Azure\_Car\_Project" is selected. It contains one storage account named "cardinehdatalakestorage". The page includes filtering and sorting options for resources.

Name	Type	Location
cardinehdatalakestorage	Storage account	UK South

## Now create Data factory

The screenshot shows the Microsoft Azure Marketplace search results for 'data factory'. The search bar at the top contains 'data factory'. Below the search bar, there are several filters: Pricing : All, Operating System : All, Publisher Type : All, Product Type : All, and Publisher name : All. A message box says 'New! Get AI-generated suggestions for 'data factory'' with a 'View suggestions' button. The results section shows 20 of 103 items:

Image	Name	Publisher	Description	Action Buttons
	Data Factory	Microsoft	Hybrid data integration service that simplifies ETL at scale	Create, Free trial
	Factory Operations Agent in Azure AI (preview)	Azure Service	Bring your manufacturing data into the era of AI with the Factory Operations Agent in Azure AI in Fabric. Full value from factory data is extracted by unifying and enriching	Create
	Astadia FastTrack Factory	Astadia	Astadia's Mainframe Migration FastTrack Factory is a set of automated refactoring and testing tools.	Subscribe
	Factory Namespace Manager	Sight Machine, Inc.	Create data dictionaries for renaming machine tags and variables for unified namespace	Subscribe
	Modern Data Mart	Ceteris AG	Small Data-Warehouse Architecture integrating a single source using Azure Data Factory	Create

The screenshot shows the Microsoft Azure Marketplace item details page for 'Data Factory'. The title is 'Data Factory' with an 'Add to Favorites' button. It is categorized as 'Microsoft | Azure Service' with a rating of '★ 3.6 (606 ratings)'. There is a 'Plan' dropdown set to 'Data Factory' and a 'Create' button. Below the title, there are tabs: Overview (selected), Plans, Usage Information + Support, Ratings + Reviews. The 'Overview' tab contains a brief description: 'Integrate data silos with Azure Data Factory, a service built for all data integration needs and skill levels. Easily construct ETL and ELT processes code-free within the intuitive visual environment, or write your own code. Visually integrate data sources using more than 90+ natively built and maintenance-free connectors at no added cost. Focus on your data - the serverless integration service does the rest.' It also lists benefits: 'No code or maintenance required to build hybrid ETL and ELT pipelines within the Data Factory visual environment', 'Cost-efficient and fully managed serverless cloud data integration tool that scales on demand', 'Azure security measures to connect to on-premises, cloud-based, and software-as-a-service apps with peace of mind', and 'SSIS integration runtime to easily rehost on-premises SSIS packages in the cloud using familiar SSIS tools'. The 'Media' section shows a screenshot of the Data Factory visual interface. At the bottom right, there is a 'Give feedback' button.

The screenshot shows the 'Create Data Factory' wizard on the 'Basics' step. The URL is [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education\\_asoS...](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education_asoS...). The page title is 'Create Data Factory'. The top navigation bar includes 'Create', 'dineshk...', 'Sign in', 'ASURITI...', 'ASURITI...', 'My ASU...', 'Windows...', 'Sign in...', 'ASU Ho...', 'Microsoft...', and 'Copilot'. The user 'dineshkumar7283@g...' is logged in.

**Project details**

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

**Subscription \***: Azure for Students

**Resource group \***: RG\_Azure\_Car\_Project

**Instance details**

**Name \***: dfs-cars-dinesh-datafactory

**Region \***: East US

**Version \***: V2

Buttons at the bottom: Previous, Next, Review + create, Give feedback.

#### Click review + create

The screenshot shows the 'Create Data Factory' wizard on the 'Review + create' step. The URL is [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education\\_asoS...](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education_asoS...). The page title is 'Create Data Factory'. The top navigation bar includes 'Create', 'dineshk...', 'Sign in', 'ASURITI...', 'ASURITI...', 'My ASU...', 'Windows...', 'Sign in...', 'ASU Ho...', 'Microsoft...', and 'Copilot'. The user 'dineshkumar7283@g...' is logged in.

**Project details**

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

**Subscription \***: Azure for Students

**Resource group \***: RG\_Azure\_Car\_Project

**Instance details**

**Name \***: dfs-cars-dinesh-datafactory

**Region \***: UK South

**Version \***: V2

Buttons at the bottom: Previous, Next, Review + create, Give feedback.

**TERMS**

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

**Basics**

Subscription	Azure for Students
Resource group	RG_Azure_Car_Project
Name	dfs-cars-dinesh-datafactory
Region	UK South
Version	V2

**Networking**

Connect via	Public endpoint
-------------	-----------------

**Create**

Finally click create to success full generate data factory

**Overview**

Subscription ([move](#)) : Azure for Students  
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d  
Tags ([edit](#)) : [Add tags](#)

**Resources**

Name	Type	Location	Actions
cardineshdatalakestorage	Storage account	UK South	...
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South	...

< Previous Page 1 of 1 Next > Give feedback

Now we Create third Resource Azure SQL DB

The screenshot shows the Microsoft Azure Marketplace search results for 'azure sql'. The search bar at the top contains 'azure sql'. Below the search bar, there are filters: 'Pricing : All', 'Operating System : All', 'Publisher Type : All', 'Product Type : All', and 'Publisher name : All'. A message box says 'New! Get AI-generated suggestions for 'azure sql''. The results section shows 1 to 20 of 682 results for 'azure sql'. Each result card includes a thumbnail, the service name, the publisher, a brief description, and a 'Create' button. The cards for 'Azure SQL', 'Azure SQL Managed Instance', 'Azure SQL Instance Pool', 'SQL Server - Azure Arc', and 'Azure SQL Analytics (Preview)' are visible.

The screenshot shows the Azure SQL product page. At the top, it says 'Azure SQL' and has a 'Create' button. Below that, it says 'Microsoft | Azure Service' and has a rating of '★ 4.7 (7 ratings)'. There is a dropdown menu for 'Plan' with options 'Azure SQL' and 'Azure SQL'. Below the plan section are tabs for 'Overview', 'Plans', 'Usage Information + Support', and 'Ratings + Reviews'. The 'Overview' tab is selected. It contains text about Azure SQL allowing you to create and manage your SQL Server resources from a single view, ranging from fully managed PaaS databases to IaaS virtual machines with direct OS and database engine access. It also mentions that all deployment options enable you to bring your on-premises licenses to Azure using Azure Hybrid Benefit. Under 'Databases', it says that single databases are optimized for modern application development of new cloud-born applications. Under 'Managed instances', it says that managed instances provide the PaaS benefits of SQL databases with added capabilities. Under 'SQL virtual machines', it says that SQL virtual machines offer an IaaS architecture with extensive control over SQL Server and the underlying OS. At the bottom, there is a 'More products from Microsoft' section with a 'See All' link and a 'Give feedback' button.

Click create

**Select SQL deployment option**

**How do you plan to use the service?**

**SQL databases**  
Best for modern cloud applications. Hyperscale and serverless options are available.  
Resource type: Single database  
**Create** **Show details**

**SQL managed instances**  
Best for most migrations to the cloud. Lift-and-shift ready.  
Resource type: Single instance  
**Create** **Show details**

**SQL virtual machines**  
Best for migrations and applications requiring OS-level access. Lift-and-shift ready.  
Image: **Create** **Show details**  High availability

Click sql databases for storage purpose only.

Click create

**Create SQL Database**

**Project details**

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

Subscription \* **Azure for Students**  
Resource group \* **RG\_Azure\_Car\_Project** [Create new](#)

**Database details**

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name \*   
Server \*  Select a server [Create new](#)

The value must not be empty.

Want to use SQL elastic pool?  Yes  No

Workload environment  Development

**Review + create** **Next : Networking >**

Fill database name and server

We donot have server. So create server So click new

The screenshot shows the 'Create SQL Database Server' wizard in the Microsoft Azure portal. The 'Server details' step is active, where the user has entered a server name ('carsalesserverdinesh') and selected a location ('(Europe) UK South'). In the 'Authentication' section, the 'Use both SQL and Microsoft Entra authentication' option is selected. A note indicates that Azure Active Directory (Azure AD) is now Microsoft Entra ID. The 'OK' button is visible at the bottom.

- Use both sql and Microsoft entra autentication in authentication method
- Set admin name and password
- Name→ admindinesh password →DIN123@kum

The screenshot shows the 'Create SQL Database Server' wizard in the Microsoft Azure portal. The 'Authentication' step is active, where the 'Use both SQL and Microsoft Entra authentication' option is selected. The 'Set Microsoft Entra admin' field is set to 'Not Selected' and contains the placeholder 'Set admin'. Below it, the 'Server admin login' field is filled with 'admindinesh' and the 'Password' field contains a masked password. The 'Confirm password' field also contains a masked password and includes a validation message: 'Password and confirm password must match.' The 'OK' button is visible at the bottom.

Than click set admin →

**Create SQL Database Server**

**Authentication**

Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#)

Select your preferred authentication methods for accessing this server. To access your server with SQL authentication, select only Microsoft Entra user, group, or application as Microsoft Entra admin for authentication.

**Authentication method**

- Use Microsoft Entra
- Use both SQL and Microsoft Entra
- Use SQL authentication

**Set Microsoft Entra admin \***

**Server admin login \***

**Password \***

**Confirm password \***

**Selected item (1)**

Dinesh km  
dineshdkumar7283@gmail.com

**OK**    **Select**

Final click ok

**Create SQL Database**

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

**Database name \*** carsalesdb

**Server \*** (new) carsalesserverdinesh (UK South)

**Want to use SQL elastic pool? \***  No  Yes

**Workload environment**  Development  Production

**Compute + storage \*** General Purpose - Serverless  
Standard-series (Gen5), 1 vCore, 32 GB storage, zone redundant disabled  
[Configure database](#)

**Backup storage redundancy**

Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only available for some storage accounts.

**Review + create**    **Next : Networking >**

When set up your performance

- 1) workload environment → Development
- 2) compute + storage → click configure database (for less storage)

**Service tier**: General Purpose (Most budget friendly)

**Compute tier**:  
 **Provisioned** - Compute resources are pre-allocated. Billed per hour based on vCores configured.  
 **Serverless** - Compute resources are auto-scaled. Billed per second based on vCores used.

**Compute Hardware**: Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.

**Hardware Configuration**: Standard-series (Gen5)  
up to 80 vCores, up to 240 GB memory  
Change configuration

**Max vCores**: 1

**Min vCores**: 0.5 vCores

**Cost summary**:

General Purpose (GP_S_Gen5_1)	11.00
Cost per GB (In INR)	x 41.6
ESTIMATED STORAGE COST / MONTH	457.70 INR
COMPUTE COST / VCORE SECOND <sup>1</sup>	0.015072 INR

**NOTES**: 1 Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. Learn more about serverless billing

**Apply**

Than click apply

Backup storage redundancy → check to Local-redundant backup storage(for less)

**Create SQL Database**

**Workload environment**: Development

**Compute + storage**: General Purpose - Serverless  
Standard-series (Gen5), 1 vCore, 32 GB storage, zone redundant disabled  
Configure database

**Backup storage redundancy**: Locally-redundant backup storage

**Review + create** | **Next : Networking >**

Click Next:Networking

Change connectivity method → public endpoint

Microsoft Azure

Home > RG\_Azure\_Car\_Project > Marketplace > Azure SQL > Select SQL deployment option >

### Create SQL Database

Networking

Basics Networking Security Additional settings Tags Review + create

Configure network access and connectivity for your server. The configuration selected below will apply to the selected server 'carsalesserverdinesh' and all databases it manages. [Learn more](#)

**Network connectivity**

Choose an option for configuring connectivity to your server via public endpoint or private endpoint. Choosing no access creates with defaults and you can configure connection method after server creation. [Learn more](#)

Connectivity method \*  Public endpoint  No access  Private endpoint

**Firewall rules**

Setting 'Allow Azure services and resources to access this server' to Yes allows communications from all resources inside the Azure boundary, that may or may not be part of your subscription. [Learn more](#)

Setting 'Add current client IP address' to Yes will add an entry for your client IP address to the server firewall.

Allow Azure services and resources to access this server \*  No  Yes

Add current client IP address \*  No  Yes

**Connection policy**

Review + create < Previous Next : Security >

**Cost summary**

General Purpose (GP_S_Gen5_1)	11.00
Cost per GB (in INR)	x 41.6
Max storage selected (in GB)	
ESTIMATED STORAGE COST / MONTH	457.70 INR
COMPUTE COST / VCORE SECOND <sup>1</sup>	0.015072 INR

**NOTES**

<sup>1</sup> Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Click Next: Securtiy

Microsoft Azure

Home > RG\_Azure\_Car\_Project > Marketplace > Azure SQL > Select SQL deployment option >

### Create SQL Database

Security

Basics Networking Security Additional settings Tags Review + create

**Microsoft Defender for SQL**

Protect your data using Microsoft Defender for SQL, a unified security package including vulnerability assessment and advanced threat protection for your server. [Learn more](#)

Get started with a 30 day free trial period, and then 1247.9202 INR/server/month.

Enable Microsoft Defender for SQL \*  Start free trial  Not now

**Ledger**

Ledger cryptographically verifies the integrity of your data and detects any tampering that might have occurred. [Learn more](#)

Ledger Not configured [Configure ledger](#)

**Server identity**

Use system assigned and user assigned managed identities to enable central access management between this database and other Azure resources. [Learn more](#)

Server identity Not enabled [Configure identities](#)

Review + create < Previous Next : Additional settings >

**Cost summary**

General Purpose (GP_S_Gen5_1)	11.00
Cost per GB (in INR)	x 41.6
Max storage selected (in GB)	
ESTIMATED STORAGE COST / MONTH	457.70 INR
COMPUTE COST / VCORE SECOND <sup>1</sup>	0.015072 INR

**NOTES**

<sup>1</sup> Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Click review+create

**Create SQL Database**

Microsoft

Basics Networking Security Additional settings Tags Review + create

**Product details**

SQL database by Microsoft

[Terms of use](#) | [Privacy policy](#)

**Estimated cost**

Storage cost 457.70 INR / month + Compute cost 0.015072 INR / vCore second

**Cost summary**

General Purpose (GP_S_Gen5_1)	
Cost per GB (in INR)	11.00
Max storage selected (in GB)	x 41.6
ESTIMATED STORAGE COST / MONTH	457.70 INR
COMPUTE COST / VCORE SECOND <sup>1</sup>	0.015072 INR

**NOTES**

<sup>1</sup> Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

**Basics**

Subscription: Azure for Students  
Resource group: RG\_Azure\_Car\_Project  
Region: UK South  
Database name: carsalesdb  
Server: (new) carsaleserverdinesh  
Authentication method: SQL and Microsoft Entra authentication  
Server admin login: admindinesh

[Create](#) < Previous Download a template for automation

Click create

**Microsoft.SQLDatabase.newDatabaseNewServer\_e01656c0545c41569633d | Overview**

Deployment

Search X

Delete Cancel Redeploy Download Refresh

**Deployment is in progress**

Deployment name : Microsoft.SQLDatabase.newDatabaseNewServer\_e01... Start time : 3/5/2025, 11:13:19 AM  
Subscription : Azure for Students Correlation ID : f476959e-c636-4608-ac1d-6498a0ee6edc  
Resource group : RG\_Azure\_Car\_Project

**Deployment details**

Resource	Type	Status	Operation details
carsaleserverdinesh/carsalesdb	Microsoft.Sql/servers/databases	Accepted	<a href="#">Operation details</a>
carsaleserverdinesh/Default	Microsoft.Sql/servers/connectio...	OK	<a href="#">Operation details</a>
carsaleserverdinesh	Microsoft.Sql/servers	Created	<a href="#">Operation details</a>

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 >](#)

Finally complete sql setup

Microsoft Azure | portal.azure.com/?Microsoft\_Azure\_Education\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\_Azure\_Education\_newA4E=true&Microsoft\_Azure\_Education\_asoS... | Microsoft | Copilot | Home | dineshkumar7283@gmail.com | DEFAULT DIRECTORY (DINESH...)

## Microsoft.SQLDatabase.newDatabaseNewServer\_e01656c0545c41569633d | Overview

Deployment

Search | Delete | Cancel | Redeploy | Download | Refresh

**Your deployment is complete**

Deployment name : Microsoft.SQLDatabase.newDatabaseNewServer\_e01... Start time : 3/5/2025, 11:13:19 AM  
Subscription : Azure for Students Correlation ID : f476959e-c636-4608-ac1d-6498a0ee6edc  
Resource group : RG\_Azure\_Car\_Project

Deployment details | Next steps | Go to resource | Give feedback | Tell us about your experience with deployment

Deployment succeeded Deployment 'Microsoft.SQLDatabase.newDatabaseNewServer\_e016...' to resource group 'RG\_Azure\_Car\_Project' was successful.

Cost management Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >

Microsoft Defender for Cloud Secure your apps and infrastructure. Go to Microsoft Defender for Cloud >

Free Microsoft tutorials Start learning today >

Work with an expert Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support. Find an Azure expert >

Microsoft Azure | portal.azure.com/?Microsoft\_Azure\_Education\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\_Azure\_Education\_newA4E=true&Microsoft\_Azure\_Education\_asoS... | Microsoft | Copilot | Home | dineshkumar7283@gmail.com | DEFAULT DIRECTORY (DINESH...)

## RG\_Azure\_Car\_Project | Resource group

Search | Create | Manage view | Delete resource group | Refresh | Export to CSV | Open query | Assign tags | Move | Delete | Export template | JSON View

Overview | Activity log | Access control (IAM) | Tags | Resource visualizer | Events | Settings | Cost Management | Monitoring | Automation | Help

Essentials

Subscription (move) : Azure for Students Deployments : 3 Succeeded  
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d Location : UK South  
Tags (edit) : Add tags

Resources Recommendations

Filter for any field... Type equals all × Location equals all × Add filter

Showing 1 to 4 of 4 records. □ Show hidden types ○

Name ↑	Type ↑	Location ↑
cardineshdatalakestorage	Storage account	UK South
carsalesdb (carsaleserverdinesh/carsalesdb)	SQL database	UK South
carsaleserverdinesh	SQL server	UK South
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South

< Previous Page 1 of 1 Next > | Give feedback

Now create datalake in broken,silver and Gold

Click resource manager → cardineshdatastorage

The screenshot shows the Azure Resource Group 'RG\_Azure\_Car\_Project'. The left sidebar has 'Overview' selected. The main area displays a table of resources:

Name	Type	Location	Actions
cardineshdatastorage	Storage account	UK South	...
carsalesdb (carsalesverdinesh/carsalesdb)	SQL database	UK South	...
carsaleserverdinesh	SQL server	UK South	...
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South	...

At the bottom, there are navigation buttons: < Previous, Page 1 of 1, Next >, and a feedback link.

The screenshot shows the Azure Storage Account 'cardineshdatastorage'. The left sidebar has 'Overview' selected. The main area shows general properties and two tabs: 'Data Lake Storage' and 'Security'.

**Properties**

Setting	Value
Resource group	: RG_Azure_Car_Project
Location	: ukouth
Subscription	: Azure for Students
Disk state	: Available
Tags	: Add tags

**Data Lake Storage**

Setting	Value
Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Disabled
Container soft delete	Disabled
Versioning	Unavailable due to migration
Change feed	Unavailable due to migration
NFS v3	Disabled
SFTP	Disabled
Storage tasks assignments	None

**Security**

Setting	Value
Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

**Networking**

Setting	Value
Allow access from	All networks
Private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes

Click container prepare data layers broken,silver,gold

Microsoft Azure

cardineshdatastorage | Containers

Search resources, services, and docs (G+)

Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state
\$logs	3/5/2025, 10:44:16 AM	Private	Available

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Partner solutions Resource visualizer Data storage

Containers File shares Queues Tables Security + networking Data management Settings Monitoring

Click container → write brozen

New container

Name \* brozen

Anonymous access level Private (no anonymous access)

The access level is set to private because anonymous access is disabled on this storage account.

Advanced

Create Give feedback

Microsoft Azure

cardineshdatastorage | Containers

Search resources, services, and docs (G+)

Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Name	Last modified	Anonymous
\$logs	3/5/2025, 10:44:16 AM	Private

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Partner solutions Resource visualizer Data storage

Containers File shares Queues Tables Security + networking Data management Settings Monitoring

Microsoft Azure

Home > RG\_Azure\_Car\_Project > cardineshdatastorage | Containers

**cardineshdatastorage | Containers**

Storage account

Search resources, services, and docs (G+)

Copilot

dineshkumar7283@gmail.com

Successfully created storage container  
Successfully created storage container 'silver'.

Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix Show deleted containers

Name	Last modified	Anonymous access level	Lease state
\$logs	3/5/2025, 10:44:16 AM	Private	Available
brozen	3/5/2025, 11:20:29 AM	Private	Available
gold	3/5/2025, 11:20:55 AM	Private	Available
silver	3/5/2025, 11:21:04 AM	Private	Available

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Partner solutions Resource visualizer Data storage Containers File shares Queues Tables Security + networking Data management Settings Monitoring

Click carsalesserverdinesh (SQL server)

Microsoft Azure

Home >

**RG\_Azure\_Car\_Project**

Resource group

Search resources, services, and docs (G+)

Copilot

dineshkumar7283@gmail.com

Create Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template ...

Subscription (move) : Azure for Students Deployment : 3 Succeeded

Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d Location : UK South

Tags (edit) : Add tags

Overview Essentials Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

No grouping List view

Showing 1 to 4 of 4 records. Show hidden types

Name	Type	Location
cardineshdatastorage	Storage account	UK South
carsalesdb (carsalesserverdinesh/carsalesdb)	SQL database	UK South
carsalesserverdinesh	SQL server	UK South
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South

< Previous Page 1 of 1 Next >

Give feedback

Microsoft Azure

Home > RG\_Azure\_Car\_Project >

**carsaleserverdinesh**

SQL server

Overview

Essentials

Create database New elastic pool New dedicated SQL pool (formerly SQL DW) Import database Reset password Move Delete Feedback

Resource group (move) : RG\_Azure\_Car\_Project

Status : Available

Location : UK South

Subscription (move) : Azure for Students

Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d

Tags (edit) : Add tags

Notifications (0)

Features (6)

All Security (4) Performance (1) Recovery (1)

**Microsoft Entra admin** Allows you to centrally manage identity and access to your Azure SQL databases. **CONFIGURED**

**Microsoft Defender for SQL** Vulnerability Assessment and Advanced Threat Protection. **NOT CONFIGURED**

**Automatic tuning** Monitors and tunes your database automatically to optimize performance. **CONFIGURED**

**Auditing** Track database events and writes them to an audit log in Azure storage. **NOT CONFIGURED**

**Failover groups** Automatically manages replication, connectivity and failover for a set of databases. **NOT CONFIGURED**

**Transparent data encryption** Encryption at rest for your databases, backups, and logs. **SERVICE-MANAGED KEY**

Available resources Filter by name All types

this method add more data bases as you preference

This is databased already create previse

Microsoft Azure

Home > RG\_Azure\_Car\_Project >

**carsaleserverdinesh**

SQL server

Overview

Create database New elastic pool New dedicated SQL pool (formerly SQL DW) Import database Reset password Move Delete Feedback

NOT CONFIGURED

NOT CONFIGURED

NOT CONFIGURED

NOT CONFIGURED

NOT CONFIGURED

NOT CONFIGURED

Available resources Filter by name All types

Name	Type	Status	Pricing tier
carsalesdb	SQL database	Online	General Purpose - Serverless: Standard-series (G...)

Click carsalesdb → then click query editor(preview)

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education\\_asoS...](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education_asoS...). The page title is "carsalesdb (carsalesserverdinesh/carsalesdb) | Query editor (preview)". The left sidebar shows navigation options like Overview, Activity log, Tags, Diagnose and solve problems, and Query editor (preview). The main area displays a "Welcome to SQL Database Query Editor" message. A "SQL server authentication" dialog box is open, prompting for "Login" (adminindinesh) and "Password". To the right, a "Microsoft Entra authentication" section shows an error message: "Cannot open server 'carsalesserverdinesh' requested by the login. Client with IP address '103.25.47.204' is not allowed to access the server. To enable access, use the Azure Management Portal or run sp\_set\_firewall\_rule on the master database to create a firewall rule for this IP address or address range. It may take up to five minutes for this change to take effect. Allowlist IP 103.25.47.204 on server carsalesserverdinesh". A "Continue as dineshkumar7283@..." button is visible.

Click allowlist IP (server less card)

After remove error than add admin and password

The screenshot shows the Microsoft Azure portal with the same URL as the previous screenshot. The page title is "carsalesdb (carsalesserverdinesh/carsalesdb) | Query editor (preview)". The left sidebar shows the same navigation options. The main area now displays a "Query 1" section with a query editor window. The editor shows a single line of code: "1". Below the editor are "Results" and "Messages" tabs, and a search bar. At the bottom of the screen, a yellow status bar says "Ready".

## Lets See Source data in github

Name	Last commit message	Last commit date
..	Added all folders and files	last week
IncrementalSales.csv	Added all folders and files	last week
SalesData.csv	Added all folders and files	last week

## Let start create source table in Azure SQL

```
1 CREATE TABLE source_car_data
2 (
3     Branch_ID varchar(200),
4     Dealer_ID varchar(200),
5     Model_ID varchar(200),
6     Revenue BIGINT,
7     Units_sold BIGINT,
8     Date_ID varchar(200),
9     Day INT,
10    Month INT,
```

Query succeeded: Affected rows: 0

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education..](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education..). The user is in the 'carsalesdb' database under the 'RG\_Azure\_Car\_Project' project. The left sidebar shows various Azure services like Overview, Activity log, Tags, and Query editor (preview). The Query editor (preview) tab is selected, showing two tabs: 'Query 1' and 'Query 2'. In 'Query 1', the user has run a script to create a table:

```

5 Model_ID varchar(200),
6 Revenue BIGINT,
7 Units_Sold BIGINT,
8 Date_ID varchar(200),
9 Day INT,
10 Month INT,
11 Year INT,
12 BranchName varchar(2000),
13 DealerName varchar(2000)
14 )

```

The results pane shows 'Query succeeded: Affected rows: 0'.

Check table create or not

The screenshot shows the Microsoft Azure portal with the same URL and context as the previous screenshot. The user has run a query to select all columns from the 'source\_car\_data' table:

```

1 SELECT * FROM [dbo].[source_car_data]

```

The results pane shows 'No results'.

```

CREATE TABLE source_car_data
(
Branch_ID varchar(200),
Dealer_ID varchar(200),
Model_ID varchar(200),
Revenue BIGINT,
Units_Sold BIGINT,
Date_ID varchar(200),
Day INT,
Month INT,
Year INT,
BranchName varchar(2000),
DealerName varchar(2000)
)

```

→data lake ready

→sql server ready

## Now Strat build Data factory

→pull the data from github to azure

→let's start

The screenshot shows the Azure Data Factory Studio interface. At the top, there's a navigation bar with tabs like Home, RG\_Azure\_Car\_Project, and a search bar. Below the navigation is a sidebar with links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Settings, Getting started, Monitoring, Automation, and Help. The main area has a section titled 'Essentials' with details about the resource group (RG\_Azure\_Car\_Project), status (Succeeded), location (UK South), subscription (Azure for Students), and subscription ID (d7c1e8a1-8bd2-4126-ba88-75140d084a4d). Below this is a large icon of a factory building with a blue cylinder next to it, labeled 'Azure Data Factory Studio'. There's a prominent 'Launch studio' button. At the bottom, there are four cards: 'Quick Starts' (cloud icon), 'Tutorials' (book icon), 'Template Gallery' (document icon), and 'Training Modules' (certificate icon). A 'Monitoring' section is also visible at the bottom.

→click launch studio

The screenshot shows the Azure Data Factory Authoring interface. On the left, there's a sidebar with links for Home, Author, Monitor, Manage, and Learning Center. The main content area is titled 'dfs-cars-dinesh-datafactory'. It features four large cards: 'Ingest' (Copy data at scale once or on a schedule), 'Orchestrate' (Code-free data pipelines), 'Transform data' (Transform your data using data flows), and 'Configure SSIS' (Manage & run your SSIS packages in the cloud). To the right of these cards is a 3D-style diagram of industrial buildings and pipes. Below the cards, there's a section for 'Recent resources' with a folder icon and the message 'No items to show'. At the bottom, there's a note: 'Your recently opened resources will show up here.'

1) Author →use for craft the pipe lines →main one

2) Monitor → Monitor the pipeline(all logs)

3) Manage → manage all the tools like github repository, connection, all the trigger and more

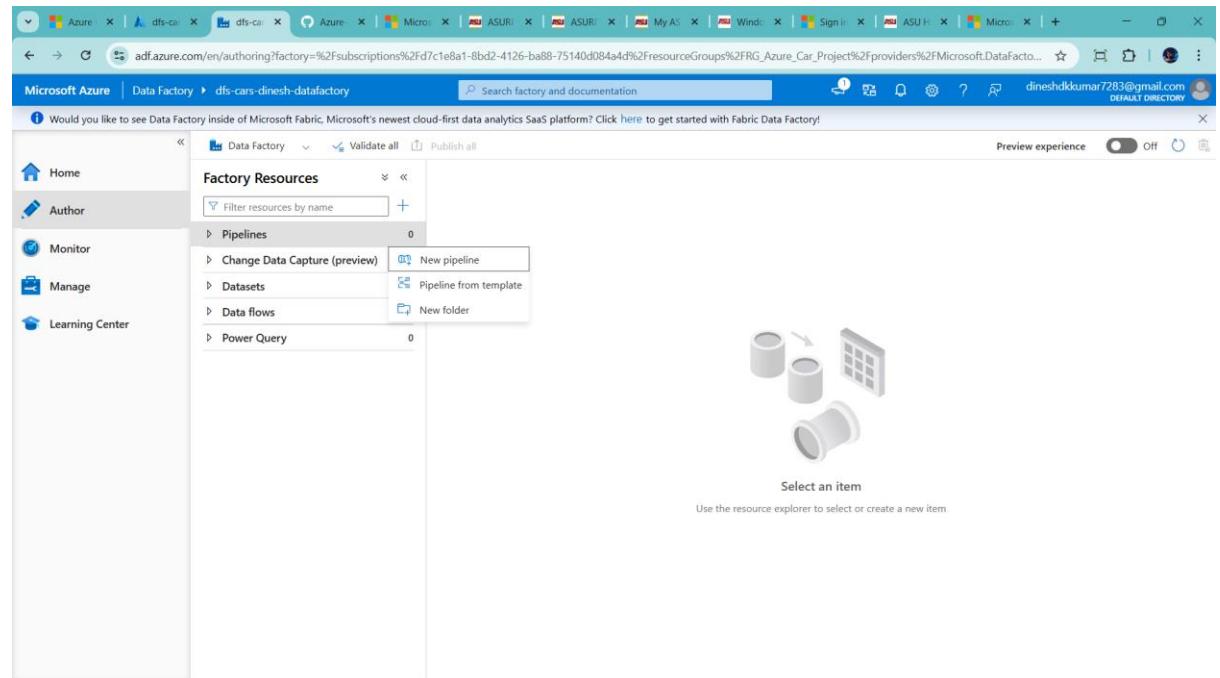
4) finally learning tap

---

Now Start with Author tap and start build pipeline

Step1) click author

Step2) click pipeline and add new pipeline

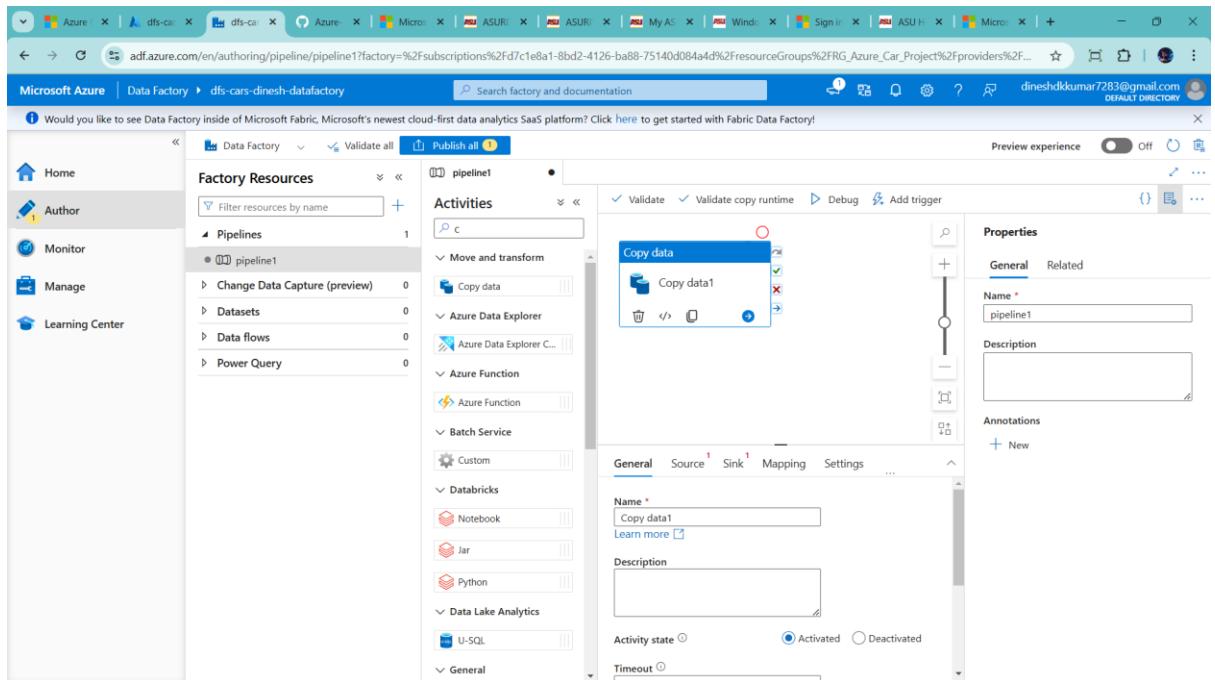


How to connection work

GitHub -----→ Data Factory -----→ SQL

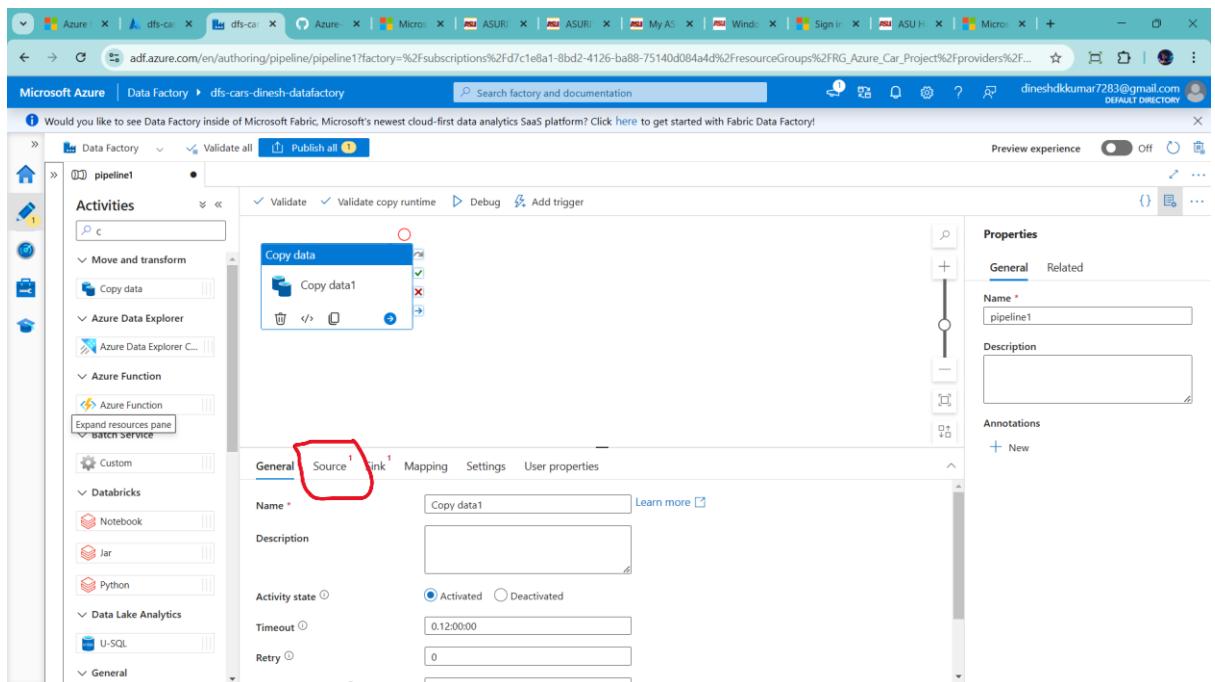
Load increment data

1) copy activity in data factory → migrate data one location another



It is connection with source

Two side to manage the copy data activity one is below source and another is manage tab'



Another one

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click [here](#) to get started with Fabric Data Factory!

Preview experience  Off

Microsoft Azure | Data Factory | dineshkumar7283@gmail.com | DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

General

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Outbound rules

Managed private endpoint

Workflow orchestration manager

Linked services

Validate all

Publish all

Search factory and documentation

+ New

Filter by name

Annotations : Any

No linked service to show

If you expected to see results, try changing your filters or create a new linked services.

Create linked service

Go to manage tab → click linked service → click + new

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click [here](#) to get started with Fabric Data Factory!

New linked service

Data store Compute

Search

All Azure Database File Generic protocol

Amazon RDS for Oracle	Amazon RDS for SQL Server	Amazon Redshift
Amazon S3	Amazon S3 Compatible	Apache Impala
af	asana	Cloud with magnifying glass

Continue Cancel

Search http → for github repository

The screenshot shows the Microsoft Azure Data Factory interface. On the left, a navigation sidebar lists various options like General, Connections, Source control, Author, Security, and Workflow orchestration manager. The main area is titled 'Linked services' and contains a sub-header 'Linked service defines the connection information to a data store or compute'. A search bar at the top right says 'Search factory and documentation'. On the right, a modal window titled 'New linked service' is open. It has tabs for 'Data store' (selected) and 'Compute'. In the 'Data store' tab, there is a search bar with 'http' typed into it. Below the search bar, there are tabs for 'All', 'Azure', 'Database', 'File', and 'Generic protocol'. Under the 'All' tab, there is a single entry for 'HTTP' with a blue icon. At the bottom of the modal are buttons for 'Continue', 'Cancel', and 'Create links'.

Click http → add name ls\_github → add base url (find raw url in github)

This screenshot shows the same 'Linked services' blade as the previous one, but with more detailed configuration for the 'ls\_github' linked service. The 'Name' field is filled with 'ls\_github'. The 'Description' field is empty. Under 'Connect via integration runtime', the 'AutoResolveIntegrationRuntime' option is selected. The 'Base URL' field is populated with a raw GitHub URL (e.g., https://raw.githubusercontent.com/). Other settings like 'Server certificate validation', 'Authentication type' (Basic), 'User name', and 'Password' are also visible. At the bottom of the modal are buttons for 'Create', 'Back', 'Test connection', and 'Cancel'.

In github

github.com/dinesh6351/Azure-DE-Project-Resources/blob/main/Raw%20Data/SalesData.csv

dinesh6351 / Azure-DE-Project-Resources

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Files

main Raw Data IncrementalSales.csv SalesData.csv README.md

Azure-DE-Project-Resources / Raw Data / SalesData.csv

dinesh2783 Added all folders and files 5d51480 · last week History

Preview Code Blame 1850 lines (1850 loc) · 168 KB Code 55% faster with GitHub Copilot

Search this file

1	Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month	Year	BranchName	DealerName
2	BR0001	DLR0001	BMW-M1	13363978	2	DT00001	1	1	2017	AC Cars Motors	AC Cars Motors
3	BR0003	DLR0228	Hon-M218	17376468	3	DT00001	10	5	2017	AC Cars Motors	Deccan Motors
4	BR0004	DLR0208	Tat-M188	9664767	3	DT00002	12	1	2017	AC Cars Motors	Wiesmann Motors
5	BR0005	DLR0188	Hyu-M158	5525304	3	DT00002	16	9	2017	AC Cars Motors	Subaru Motors
6	BR0006	DLR0168	Ren-M128	12971088	3	DT00003	20	5	2017	AC Cars Motors	Saab Motors
7	BR0008	DLR0128	Hon-M68	7321228	1	DT00004	28	4	2017	AC Cars Motors	Messerschmitt Motors
8	BR0009	DLR0108	Cad-M38	11379294	2	DT00004	31	12	2017	AC Cars Motors	Lexus Motors
9	BR0010	DLR0088	Mer-M8	11611234	2	DT00005	4	9	2017	AC Cars Motors	IFA (including Trabant, Wartburg, Barkas) Motors
10	BR0011	DLR0002	BMW-M2	19979446	2	DT00005	2	1	2017	Acura Motors	Acura Motors
11	BR0011	DLR0069	Voi-M26	14181510	3	DT00006	9	5	2017	Acura Motors	Geo Motors
12	BR0012	DLR0249	BMW-M249	5358057	1	DT00006	6	9	2017	Acura Motors	Acura Motors
						DT00007	11	5	2017	Acura Motors	Herald Motors

<https://github.com/dinesh6351/Azure-DE-Project-Resources/raw/refs/heads/main/Raw%20Data/SalesData.csv>

Click raw

raw.githubusercontent.com/dinesh6351/Azure-DE-Project-Resources/refs/heads/main/Raw%20Data/SalesData.csv

Google Lens

Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month	Year	BranchName	DealerName
BR0001	DLR0001	BMW-M1	13363978	2	DT00001	1	1	2017	AC Cars Motors	AC Cars Motors
BR0001	DLR0228	Hon-M218	17376468	3	DT00001	10	5	2017	AC Cars Motors	Deccan Motors
BR0004	DLR0208	Tat-M188	9664767	3	DT00002	12	1	2017	AC Cars Motors	Wiesmann Motors
BR0005	DLR0188	Hyu-M158	5525304	3	DT00002	16	9	2017	AC Cars Motors	Subaru Motors
BR0006	DLR0168	Ren-M128	12971088	3	DT00002	20	5	2017	AC Cars Motors	Saab Motors
BR0008	DLR0128	Hon-M68	7321228	1	DT00004	28	4	2017	AC Cars Motors	Messerschmitt Motors
BR0009	DLR0108	Cad-M38	11379294	2	DT00004	31	12	2017	AC Cars Motors	Lexus Motors
BR0010	DLR0088	Mer-M8	11611234	2	DT00005	4	9	2017	AC Cars Motors	IFA (including Trabant, Wartburg, Barkas) Motors
BR0011	DLR0002	BMW-M2	19979446	2	DT00005	2	1	2017	Acura Motors	Acura Motors
BR0011	DLR0069	Voi-M26	14181510	3	DT00006	9	5	2017	Acura Motors	Geo Motors
BR0012	DLR0249	BMW-M249	5358057	1	DT00006	6	9	2017	Acura Motors	Acura Motors
BR0013	DLR0229	Hon-M219	16150431	3	DT00006	11	5	2017	Acura Motors	Herald Motors
BR0013	DLR0229	Hon-M219	16150431	3	DT00006	12	1	2017	Acura Motors	Tata
BR0015	DLR0150	Hyu-M159	4891618	2	DT00006	17	9	2017	Acura Motors	Zunbeam Motors
BR0017	DLR0149	Mer-M9	5099144	2	DT00008	25	8	2017	Acura Motors	Panox Motors
BR0018	DLR0129	Hon-M69	17369466	2	DT00008	29	4	2017	Acura Motors	Lexus
BR0019	DLR0109	Cad-M39	26969532	3	DT00010	1	1	2017	Acura Motors	Ligier Motors
BR0020	DLR0089	Voi-M26	4816794	4	DT00011	5	2017	Acura Motors	Volkswagen	
BR0021	DLR0069	Voi-M26	14181510	3	DT00011	15	9	2017	Acura Motors	BMW
BR0021	DLR0070	Voi-M26	14181510	3	DT00011	16	9	2017	Acura Motors	Gilkorn Motors
BR0022	DLR0070	Voi-M26	14181510	3	DT00011	17	9	2017	Acura Motors	Acura Motors
BR0022	DLR0070	Voi-M26	14181510	3	DT00011	18	9	2017	Acura Motors	BMW
BR0025	DLR0190	Hyu-M159	6733530	2	DT00012	18	9	2017	Acura Motors	(including Arola) Motors
BR0027	DLR0191	Mer-M108	6594996	3	DT00013	26	8	2017	Aixam-Mega	(including Arola) Motors
BR0028	DLR0130	Hon-M72	8355774	3	DT00016	30	4	2017	Aixam-Mega	(including Arola) Motors
BR0029	DLR0110	Che-M42	6158643	1	DT00016	7	1	2017	Aixam-Mega	(including Arola) Motors
BR0031	DLR0040	IMM-M4	18118773	1	DT00017	4	1	2017	Alfa Romeo Motors	Alfa Romeo Motors
BR0032	DLR0040	IMM-M4	18118773	1	DT00017	14	5	2017	Alfa Romeo Motors	Volkswagen
BR0033	DLR0251	Mer-M221	8443706	1	DT00018	0	9	2017	Alfa Romeo Motors	Alfa Romeo Motors
BR0033	DLR0231	Mer-M221	4590128	1	DT00018	13	5	2017	Alfa Romeo Motors	Ford
BR0035	DLR0191	Hyu-M161	4225644	2	DT00019	19	9	2017	Aixam-Mega	(including Arola) Motors
BR0036	DLR0151	Toy-M101	7525079	1	DT00020	27	8	2017	Alfa Romeo Motors	ATLantic Motor Company
BR0038	DLR0131	Toy-M101	8074050	2	DT00022	1	5	2017	Alfa Romeo Motors	Peel Motors
BR0040	DLR0091	Jee-M11	4058632	1	DT00023	7	9	2017	Alfa Romeo Motors	Invicta Motors
BR0041	DLR0091	Jee-M11	4058632	1	DT00024	12	5	2017	Alfa Romeo Motors	Alfa Romeo Motors
BR0042	DLR0072	Voi-M259	13055810	2	DT00024	12	5	2017	Alfa Romeo Motors	Gin Motors
BR0042	DLR0252	Mer-M252	5616461	1	DT00024	9	9	2017	Alfa Romeo Motors	Volkswagen
BR0043	DLR0232	Mer-M222	18031048	2	DT00026	14	5	2017	Alpine Motors	2008 NRW-listed Ford
BR0044	DLR0222	Mer-M192	4061409	3	DT00026	16	1	2017	Alpine Motors	Tata
BR0045	DLR0192	Hyu-M162	2319938	3	DT00026	28	9	2017	Alpine Motors	Tata Motors
BR0047	DLR0152	Toy-M102	10000075	1	DT00027	28	9	2017	Alpine Motors	Perodua Motors
BR0049	DLR0112	Mer-M112	11300000	1	DT00028	4	1	2017	Alpine Motors	Hyundai
BR0050	DLR0092	Jee-M12	20808708	3	DT00028	4	1	2017	Alpine Motors	LTI Motors
BR0050	DLR0092	Jee-M12	17231670	3	DT00029	8	9	2017	Alpine Motors	Isdera Motors
BR0051	DLR0006	Mer-M6	1196976	1	DT00030	6	1	2017	Alvis Motors	Alvis Motors
BR0051	DLR0073	Vol-M260	2224368	1	DT00031	13	5	2017	Alvis Motors	Mercedes-Benz
BR0052	DLR0253	Jee-M23	16890826	2	DT00031	10	9	2017	Alvis Motors	Alvis Motors
BR0053	DLR0233	For-M223	7422324	1	DT00032	15	5	2017	Alvis Motors	Blankinship Motor Company Building
BR0054	DLR0042	Mer-M104	20808708	3	DT00032	15	5	2017	Alvis Motors	Mercedes-Benz
BR0055	DLR0193	Hyu-M163	309955	1	DT00033	21	9	2017	Alvis Motors	Jazzani Motors
BR0057	DLR0153	Toy-M103	3735254	2	DT00033	29	8	2017	Alvis Motors	Peugeot
										toyota

Than copy that url add in bases url in add link service

New linked service

Name \*  
ls\_github

Description

Connect via integration runtime \*  
AutoResolveIntegrationRuntime

Base URL \*  
https://raw.githubusercontent.com/

Server certificate validation  
Enable

Authentication type \*  
Basic

User name \*

Password  
Azure Key Vault

Create Back Test connection Cancel

Authentication type change Anonymous

And test the connection

New linked service

Name \*  
ls\_github

Description

Connect via integration runtime \*  
AutoResolveIntegrationRuntime

Base URL \*  
https://raw.githubusercontent.com/

Server certificate validation  
Enable

Authentication type \*  
Anonymous

Auth headers  
New

Annotations  
New

Connection successful  
Test connection Cancel

Click Create

Now create another connection to push data in azure sql db

Add name ls\_sqlDB → change azure subscription → azure of student (like) → add server name →

Add databases name and password (already created)

Microsoft Azure | Data Factory | dfs-cars-dinesh-datafactory | Search factory and documentation

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click here to get started with Microsoft Fabric.

Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New

Filter by name Annotations: Any

Showing 1 - 1 of 1 items

Name	Type
ls_github	HTTP

New linked service

Azure SQL Database Learn more

Name \* ls\_sqlDB

Description

Connect via integration runtime \* AutoResolveIntegrationRuntime

Version

Recommended (selected) Legacy

Import from connection string

Account selection method \* From Azure subscription Enter manually

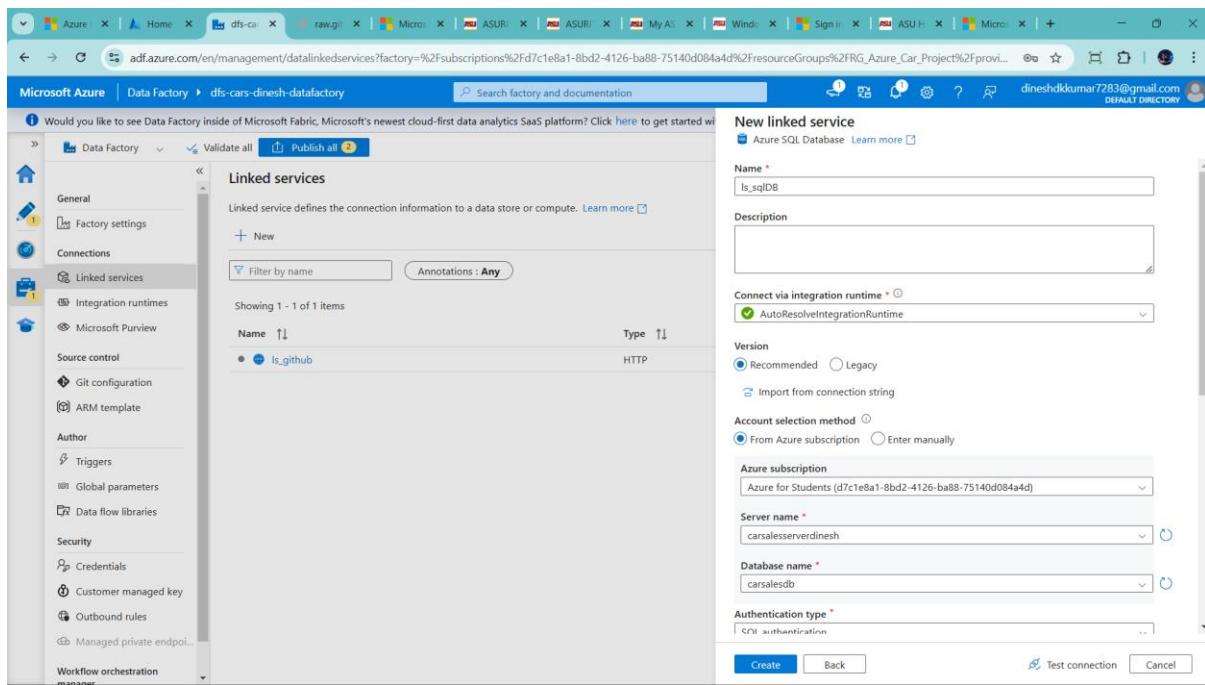
Azure subscription Azure for Students (d7c1e8a1-8bd2-4126-ba88-75140d084a4d)

Server name \* carsaleserverdinesh

Database name \* carsalesdb

Authentication type \* SQL authentication

Create Back Test connection Cancel



Microsoft Azure | Data Factory | dfs-cars-dinesh-datafactory | Search factory and documentation

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click here to get started with Microsoft Fabric.

Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New

Filter by name Annotations: Any

Showing 1 - 1 of 1 items

Name	Type
ls_github	HTTP

New linked service

Azure SQL Database Learn more

Server name \* carsaleserverdinesh

Database name \* carsalesdb

Authentication type \* SQL authentication

User name \* admindinesh

Password \* [REDACTED]

Azure Key Vault

Always encrypted \*

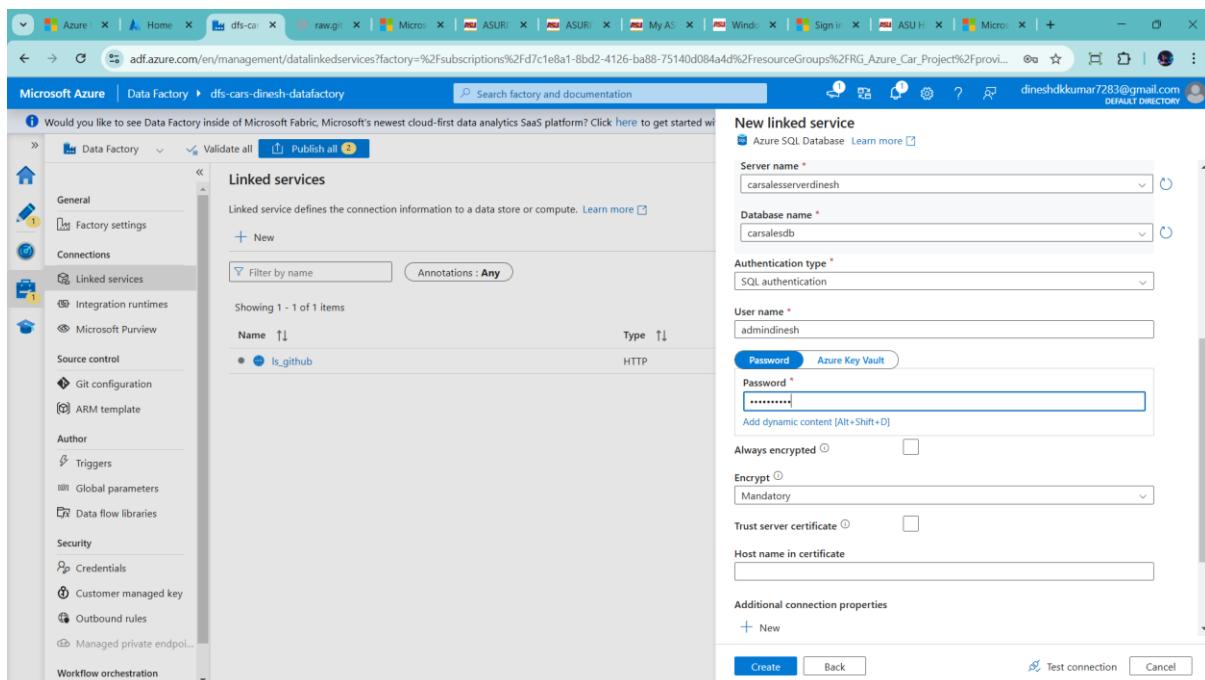
Encrypt \* Mandatory

Trust server certificate \*

Host name in certificate

Additional connection properties + New

Create Back Test connection Cancel



Test the connect show error → the give some security add in sqlserver

The screenshot shows the Microsoft Azure Data Factory interface. On the left, a sidebar lists various settings like General, Connections, and Source control. The main area is titled 'Linked services' and shows a single entry named 'ls\_github'. To the right, a form for creating a new linked service is displayed. It's configured to connect to a 'carsalesdb' database in 'Azure SQL Database'. The 'Authentication type' is set to 'SQL authentication' with 'adminindinesh' as the user name and a masked password. The 'Always encrypted' checkbox is unchecked. Under 'Additional connection properties', there is a 'Mandatory' dropdown set to 'Encrypt'. At the bottom, there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

Let change it

Click server

The screenshot shows the Microsoft Azure Resource Groups page. The resource group 'RG\_Azure\_Car\_Project' is selected. The 'Overview' tab is active, displaying basic information like the subscription (Azure for Students), location (UK South), and deployment status (3 Succeeded). The 'Resources' tab lists four resources: 'cardineshdatastorage' (Storage account), 'carsalesdb (carsaleserverdinesh/carsalesdb)' (SQL database), 'carsaleserverdinesh' (SQL server), and 'dfs-cars-dinesh-datafactory' (Data factory V2). Each resource has a 'More' button next to it. The URL in the browser bar is [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education...](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education...).

They show error because it is only selected networks

The screenshot shows the Azure portal interface for managing a SQL server named 'carsaleserverdinesh'. The left sidebar has a 'Networking' section selected. The main content area is titled 'Networking' and shows the 'Public access' tab is active. Under 'Public network access', the 'Selected networks' radio button is selected. A note below explains that connections from IP addresses in the Firewall rules section will have access. The 'Virtual networks' section shows an empty table for adding virtual network rules. The 'Firewall rules' section also shows an empty table. At the bottom, there are 'Save' and 'Discard' buttons.

So when click allow all the network ip address. So check the box in exception

And save it

This screenshot shows the same networking configuration for the 'carsaleserverdinesh' SQL server. The 'Selected networks' option remains selected. In the 'Exceptions' section, a checkbox labeled 'Allow Azure services and resources to access this server' is checked. The rest of the interface, including the 'Virtual networks' and 'Firewall rules' sections, is identical to the previous screenshot.

Now success full

Now create two activity in this pipeline

Now click Author to create pipeline

First click the copy data → then click source

Let create parameter dataset because it is dynamic dataset

So click +new

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Change Data Capture (preview)' (0), 'Datasets' (0), 'Data flows' (0), and 'Power Query' (0). The main workspace displays a 'pipeline1' pipeline with a single 'Copy data' activity. The 'Source' tab of the activity configuration is selected. The 'Properties' panel on the right shows the pipeline's name as 'pipeline1'. A 'New dataset' dialog is open on the right side of the screen.

The screenshot shows the 'New dataset' dialog from the previous step. It prompts the user to 'Select a data store' and provides a search bar with the prefix 'htt'. Below the search bar, there are tabs for 'All', 'Azure', 'Database', 'File', and 'Generic protocol'. Under the 'All' tab, a single dataset named 'HTTP' is listed, represented by a blue icon with a white globe and the word 'HTTP'. At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

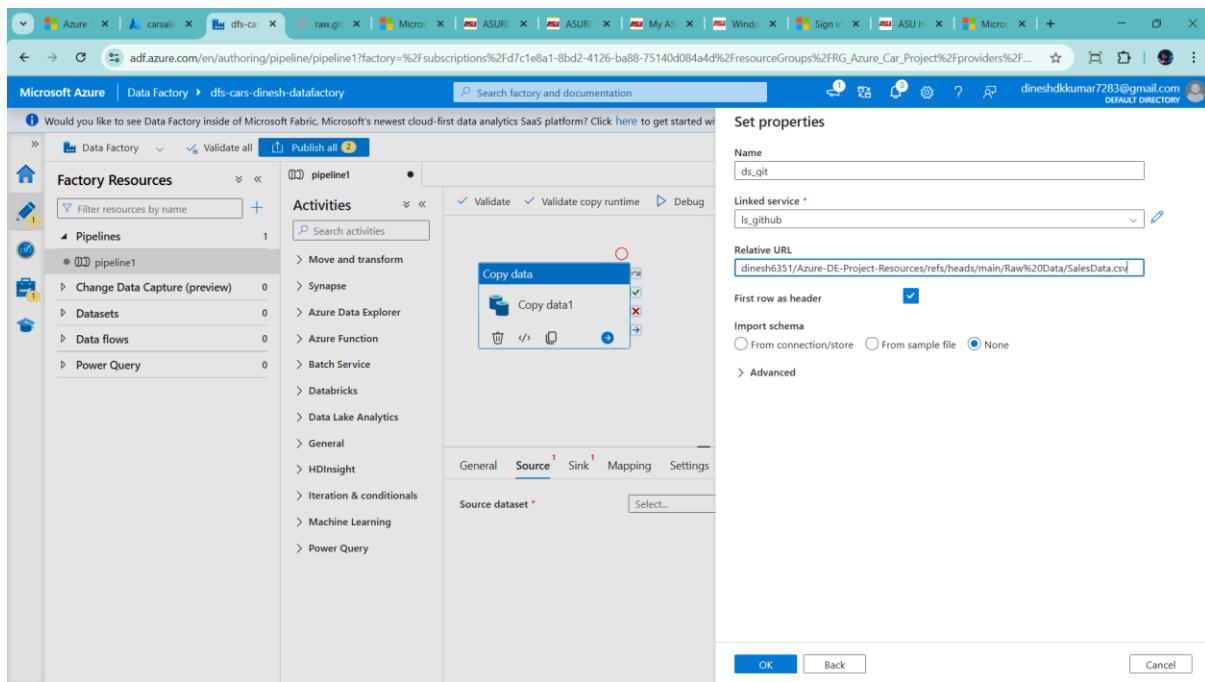
Csv

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Change Data Capture (preview)' (0), 'Datasets' (0), 'Data flows' (0), and 'Power Query' (0). The main area displays a pipeline named 'pipeline1'. A 'Copy data' activity is selected, with its properties pane visible. To the right, a 'Select format' dialog box is open, titled 'Choose the format type of your data'. It contains a grid of icons representing different file formats: Avro, Binary, DelimitedText, Excel, JSON, ORC, Parquet, and XML. At the bottom of the dialog are 'Continue', 'Back', and 'Cancel' buttons.

Now add name, add linked service, and reletative url

This reletative url

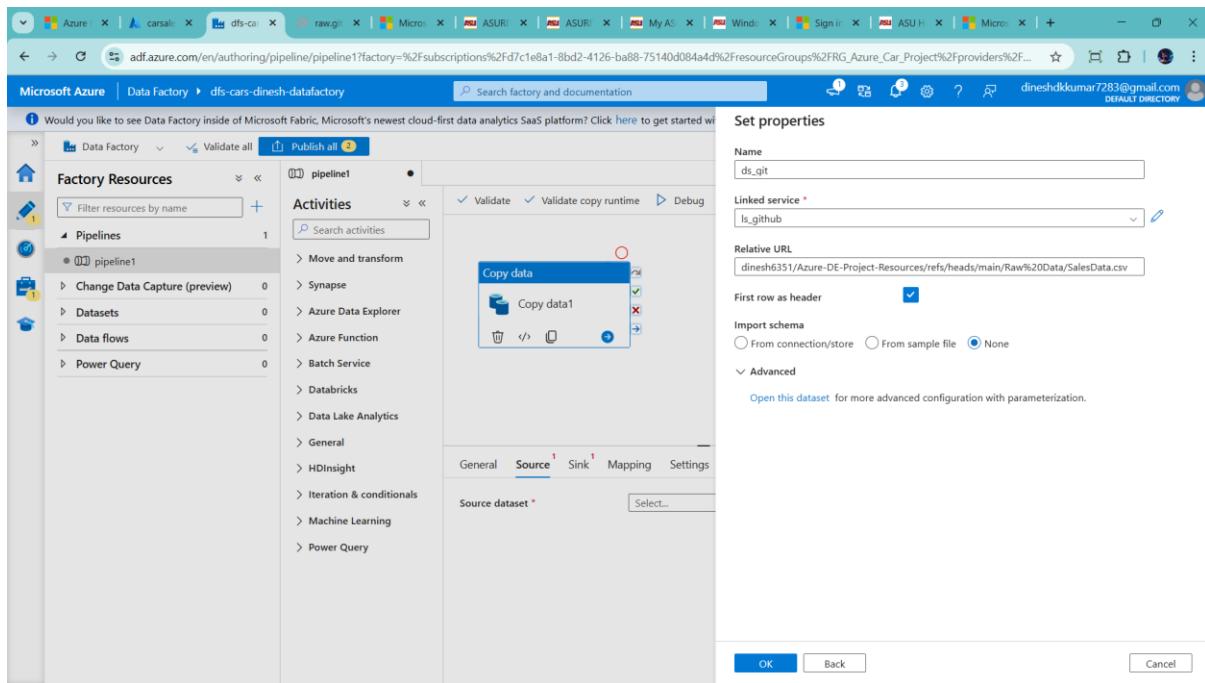
Branch_ID	Dealer_ID	ID	Model	ID	Revenue	Units	Sold	Date	ID	Day	Month	Year	BranchName	DealerName	Product_Name
BR0001	DLR0001	API-M22	13635979	D100001	1,1	2017	AC Cars Motors	Ac Cars Motors	BMW						
BR0002	DLR022	Ren-M22	17376468	D100001	1,1	2017	AC Cars Motors	Ac Cars Motors	Honda						
BR0004	DLR026	Tat-M188	9664767	D100002	12,1	2017	AC Cars Motors	Messmann Motors	Tata						
BR0005	DLR0188	Ayu-M158	5525304	D100002	16,9	2017	AC Cars Motors	Subaru Motors	Hyundai						
BR0006	DLR0168	Ren-M128	1297108	D100003	20,5	2017	AC Cars Motors	Saab Motors	Renault						
BR0008	DLR0128	Hon-M68	7321228	D100004	28,4	2017	AC Cars Motors	Messerschmitt Motors	Honda						
BR0009	DLR0089	Cad-M38	1811294	D100004	31,13	2017	AC Cars Motors	Lexus Motors	Cadillac						
BR0011	DLR0080	BMW-M2	15161220	D100005	17,9	2017	AC Cars Motors	BMW	BMW						
BR0011	DLR0002	BMW-M2	19979446	D100005	2,1	2017	Acura Motors	Acura Motors	BMW						
BR0012	DLR0249	BMW-M249	5358057	D100006	6,9	2017	Acura Motors	Acura Motors	Volkswagen						
BR0013	DLR0229	Non-M219	16150431	D100007	11,5	2017	Acura Motors	Herald Motors	Honda						
BR0014	DLR0209	Tat-M189	13389352	D100007	13,1	2017	Acura Motors	Zastava Motors	Tata						
BR0007	DLR0120	Ren-M20	10000000	D100008	17,9	2017	Acura Motors	Sunbeam Motors	Hyundai						
BR0017	DLR0145	Vol-M100	5959144	D100008	2,0	2017	Acura Motors	Panther Motors	BMW						
BR0018	DLR0129	Non-M69	17369466	D100009	29,4	2017	Acura Motors	Hia Motors	Honda						
BR0019	DLR0109	Cad-M39	26969532	D100010	1,1	2017	Acura Motors	Ligier Motors	Cadillac						
BR0020	DLR0089	BMW-M4	4816794	D100011	9,9	2017	Acura Motors	Infiniti Motors	Dodge						
BR0021	DLR0070	Vol-M257	7738896	D100011	10,5	2017	Acura Motors	Gilbern Motors	Volkswagen						
BR0024	DLR0210	Tat-M190	11038722	D100012	14,1	2017	Aixam-Mega	(including Arola) Motors	ZAZ Motors						
BR0025	DLR0150	Ren-M222	10000000	D100012	2,0	2017	Aixam-Mega	(including Arola) Motors	BMW						
BR0027	DLR0150	Vol-M100	6004096	D100013	12,6	2017	Aixam-Mega	(including Arola) Motors	Panther Motors						
BR0026	DLR0136	Non-M72	8355774	D100013	30,4	2017	Aixam-Mega	(including Arola) Motors	Micro Motors						
BR0029	DLR0110	Cad-M39	26969532	D100016	2,1	2017	Aixam-Mega	(including Arola) Motors	Lincoln Motors						
BR0031	DLR0004	BMW-M4	18118773	D100016	1,1	2017	Aixam-Mega	Alfa Romeo Motors	Chevrolet						
BR0031	DLR0004	BMW-M4	6158643	D100016	1,4	2017	Aixam-Mega	Alfa Romeo Motors	Alfa Romeo Motors						
BR0031	DLR0004	BMW-M4	18118773	D100017	4,1	2017	Aixam-Mega	Alfa Romeo Motors	BMW						
BR0031	DLR0004	BMW-M4	8443760	D100017	8,9	2017	Aixam-Mega	Alfa Romeo Motors	Alfa Romeo Motors						
BR0032	DLR0022	Ren-M222	10000000	D100018	12,0	2017	Aixam-Mega	Alfa Romeo Motors	Ford						
BR0033	DLR0191	Ayu-M161	4225644	D100019	19,9	2017	Aifa Romeo Motors	Tata Motors	Hyundai						
BR0037	DLR0151	Toy-M101	7525079	D100020	27,5	2017	Aifa Romeo Motors	Peel Motors	Toyota						
BR0038	DLR0131	Non-M71	8074050	D100022	1,5	2017	Alfa Romeo Motors	Mini Motors	Honda						
BR0040	DLR0091	Jee-M11	4058632	D100023	7,9	2017	Alfa Romeo Motors	Invicta Motors	Jeep						
BR0041	DLR0005	BMW-M5	1232500	D100023	5,1	2017	Alpine Motors	Alpine Motors	BMW						
BR0041	DLR0072	Ren-M20	10000000	D100023	2,0	2017	Alpine Motors	Alpine Motors	Glas Motors						
BR0042	DLR0072	Ren-M22	5014641	D100024	13,9	2017	Alpine Motors	Alpine Motors	Volkswagen						
BR0043	DLR0232	For-M222	18031408	D100025	14,5	2017	Alpine Motors	Alpine Motors	BMW						
BR0044	DLR0212	Tat-M192	4061400	D100026	16,1	2017	Alpine Motors	Tata Motors	BMW						
BR0045	DLR0192	Ayu-M162	2319930	D100026	20,9	2017	Alpine Motors	Tata Motors	Hyundai						
BR0047	DLR0152	Toy-M102	3120075	D100027	28,8	2017	Alpine Motors	Perodua Motors	Toyota						
BR0048	DLR0132	Ayu-M72	11630901	D100028	2,5	2017	Alpine Motors	Mitsubishi Motors	Hyundai						
BR0049	DLR0004	Che-M42	10000000	D100028	1,0	2017	Alpine Motors	Alfa Romeo	Chevrolet						
BR0050	DLR0091	Non-M71	17231670	D100028	8,9	2017	Alpine Motors	Isuzu Motors	Jeep						
BR0051	DLR0046	Mer-M6	1196976	D100029	6,1	2017	Alvis Motors	Alvis Motors	Mercedes-Benz						
BR0051	DLR0073	Vol-M260	2224368	D100031	13,5	2017	Alvis Motors	GMC Motors	Volkswagen						
BR0052	DLR0253	Jee-M253	16898082	D100031	10,9	2017	Alvis Motors	Alvis Motors	Jeep						
BR0053	DLR0233	For-M223	7422324	D100032	15,5	2017	Alvis Motors	Blankinship Motor Company Building	Ford						
BR0054	DLR0213	Tat-M193	5594044	D100032	17,1	2017	Alvis Motors	DKW Motors	Tata						
BR0056	DLR0193	Toy-M163	3000000	D100033	1,0	2017	Alvis Motors	Fazzari Motors	Hyundai						
BR0057	DLR0133	Toy-M103	1733254	D100033	19,9	2017	Alvis Motors	Peugeot Motors	Toyota						



Because create dynamic to add csv file as incremental load

Now add parameter in this dataset

So ,click advance option and open this dataset



Than click parameter tap and click new

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. Under 'Datasets', there is one entry: 'ds.git'. The main workspace displays a 'Pipeline1' pipeline with a single step named 'ds.git'. This step is a 'DelimitedText' type with a CSV file icon. Below the pipeline, the 'Parameters' tab is selected, showing a table with one parameter: 'load\_flag' of type String with a default value of 'Value'. To the right, the 'Properties' panel shows the dataset's name as 'ds.git' and its type as 'DelimitedText'. There are tabs for 'General' and 'Related (1)'. The 'Annotations' section is empty.

Go back to connection and click on relative url And click dynamic content show below

This screenshot shows the same dataset configuration page, but the 'Connection' tab is now selected. In the 'Linked service' dropdown, 'ls.github' is chosen. The 'Base URL' field contains the URL 'https://raw.githubusercontent.com/'. Below it, the 'Relative URL' field contains the path 'dineshkumar7283/Azure-DE-Project-Resources/ref/heads/main/Raw%20Data/SalesData.csv'. The 'Preview data' and 'Detect format' buttons are visible next to the URL input field. Other connection parameters like Compression type, Column delimiter, Row delimiter, Encoding, and Quote character are also shown.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar is open, showing one pipeline ('pipeline1') and one dataset ('ds.git'). The main area displays 'pipeline1' with a single step: 'DelimitedText ds.git'. Below this is the 'Connection' configuration for the dataset, which includes a 'Linked service' dropdown set to 'ls.github', a 'Base URL' of 'https://raw.githubusercontent.com/', and a 'Relative URL' of 'dinesh6351/Azure-DE-Project-Resources/refs/heads/main/Raw%20Data/SalesData.csv'. To the right is the 'Pipeline expression builder' window. It has tabs for 'Parameters' (selected) and 'Functions'. Under 'Parameters', there is a search bar and a list containing 'load\_flag'. The expression builder's main text area contains the expression '@dataset().load\_flag'. At the bottom right of the expression builder window are 'OK' and 'Cancel' buttons.

Select SalesData.csv → click load\_flag

This screenshot is identical to the previous one, showing the Microsoft Azure Data Factory Pipeline expression builder. The difference is in the expression being typed in the main text area: it now reads '@dataset().load\_flag' instead of '@{dataset().load\_flag}'. The rest of the interface, including the sidebar, dataset selection, connection details, and the 'OK' button, remains the same.

Now change string to variable

This screenshot shows the Microsoft Azure Data Factory Pipeline expression builder with the expression '@dataset().load\_flag' highlighted by a red rectangular box. A red arrow points from below towards this highlighted text. The rest of the interface is consistent with the previous screenshots.

The screenshot shows the 'Pipeline expression builder' interface for a 'DelimitedText' dataset named 'ds.git' in pipeline1. The 'Parameters' tab is active, showing the expression `@{dataset().load_flag}`. The 'Functions' tab is also visible.

Now click ok and go to pipeline1/source\_prep

The screenshot shows the 'Activities' section of pipeline1. A 'Copy data' activity named 'Copy data1' is selected. The 'Properties' pane on the right shows the pipeline name as 'pipeline1'.

Ask to add value

So I add SalesData.csv and click preview data

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (1), and 'Data flows' (0). The main workspace displays a pipeline named 'pipeline1' with a single 'ds.git' dataset. A 'Copy data' activity is selected, showing its properties: 'Source dataset' is set to 'ds.git'. The 'Properties' panel on the right shows the pipeline's name as 'pipeline1'.

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The pipeline 'pipeline1' now contains a preview step. The 'Source dataset' is still 'ds.git'. The preview data table shows the following rows:

	Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month	Year	BranchName
1	BR0001	DLR0001	BMW-M1	13363978	2	DT00001	1	1	2017	AC Cars Motors
2	BR0003	DLR0228	Hon-M218	17376468	3	DT00001	10	5	2017	AC Cars Motors
3	BR0004	DLR0208	Tat-M188	9664767	3	DT00002	12	1	2017	AC Cars Motors
4	BR0005	DLR0188	Hyu-M158	5525304	3	DT00002	16	9	2017	AC Cars Motors
5	BR0006	DLR0168	Ren-M128	12971088	3	DT00003	20	5	2017	AC Cars Motors
6	BR0008	DLR0128	Hon-M68	7321228	1	DT00004	28	4	2017	AC Cars Motors
7	BR0009	DLR0108	Cad-M38	11379294	2	DT00004	31	12	2017	AC Cars Motors

Now Click sink → click new → type azure sqldatabase → click ok

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (1), and 'Data flows' (0). The main workspace displays a pipeline named 'pipeline1' with a single 'ds\_git' dataset. A 'Copy data' activity is selected, showing its configuration pane. The 'Sink' tab is active, and the 'Sink dataset' dropdown is open, displaying three options: 'Azure SQL Database', 'Azure SQL Database Managed Instance', and 'Azure Synapse Analytics'. The 'Azure SQL Database' option is selected. At the bottom right of the configuration pane are 'Continue' and 'Cancel' buttons.

This screenshot shows the 'Set properties' dialog for the 'db\_sqldb' sink dataset. The 'Name' field is set to 'db\_sqldb'. The 'Linked service' dropdown is set to 'ls\_sqldb'. Under the 'Table name' section, 'dbo.source\_car\_data' is selected. The 'Import schema' section has 'From connection/store' checked. At the bottom are 'OK', 'Back', and 'Cancel' buttons.

**Set properties**

Name: db\_sqldb

Linked service: ls\_sqldb

Table name: dbo.source\_car\_data  
Enter manually

Import schema:  From connection/store  None

Advanced

OK Back Cancel

## Now click mapping and import schema

The screenshot shows the Microsoft Azure Data Factory pipeline editor. A pipeline named 'pipeline1' is selected. In the 'Activities' pane, a 'Copy data' activity is highlighted. The 'Mapping' tab is active, showing the data schema for the 'CopyGitData' source. The table lists four columns: 'Branch\_ID', 'Dealer\_ID', 'ModelID', and 'Revenue'. Each column is mapped to a destination column with the same name and type ('String' for Branch\_ID, Dealer\_ID, ModelID, and 'bigint' for Revenue). The 'Properties' pane on the right shows the pipeline's general settings, including its name 'pipeline1'.

Now load the data in sql db , so click Debug

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The pipeline 'pipeline1' is selected. The 'Activities' pane shows the 'Copy data' activity. The 'Debug' tab is active. The 'Parameters' tab is selected, showing a '+ New' button. The 'Properties' pane on the right shows the pipeline's general settings, including its name 'pipeline1'.

## Success

The screenshot shows the Microsoft Azure Data Factory pipeline status page. The pipeline, named 'pipeline1', contains a single activity named 'CopyGitData'. The status bar indicates the pipeline has succeeded. The table below provides detailed information about the activity run.

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime
CopyGitData	Succeeded	Copy data	3/5/2025, 12:52:59 PM	15s	AutoResolveIntegrationRuntime (UK)

## Dynamic ETL pipeline in Azure Data factory

Now the public the data , so click publishing and publish

→ it source-prep pipeline publish success

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The left sidebar lists 'Factory Resources' including Pipelines, Datasets, Data flows, and Power Query. The main workspace shows the 'source\_prep' pipeline with a single 'Copy data' activity named 'CopyGitData'. The 'Properties' pane on the right shows the pipeline name as 'pipeline1'. The 'Output' pane displays a successful run with a Pipeline run ID and Pipeline status. A detailed table shows the run history.

Now create new pipeline for incremental load

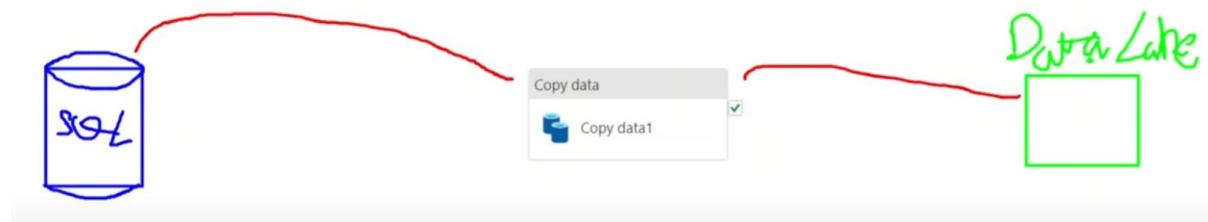
Click pipeline → new pipeline → name: incremental\_data\_pipeline

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The left sidebar lists 'Factory Resources' including Pipelines, Datasets, Data flows, and Power Query. The main workspace shows the creation of a new pipeline named 'increment\_data\_pipeline'. This pipeline contains two activities: 'source\_prep' and 'increment\_data\_pipeline'. The 'Properties' pane on the right shows the pipeline name as 'increment\_data\_pipeline'. The 'Output' pane is currently empty.

## Incremental loading in Azure data factory

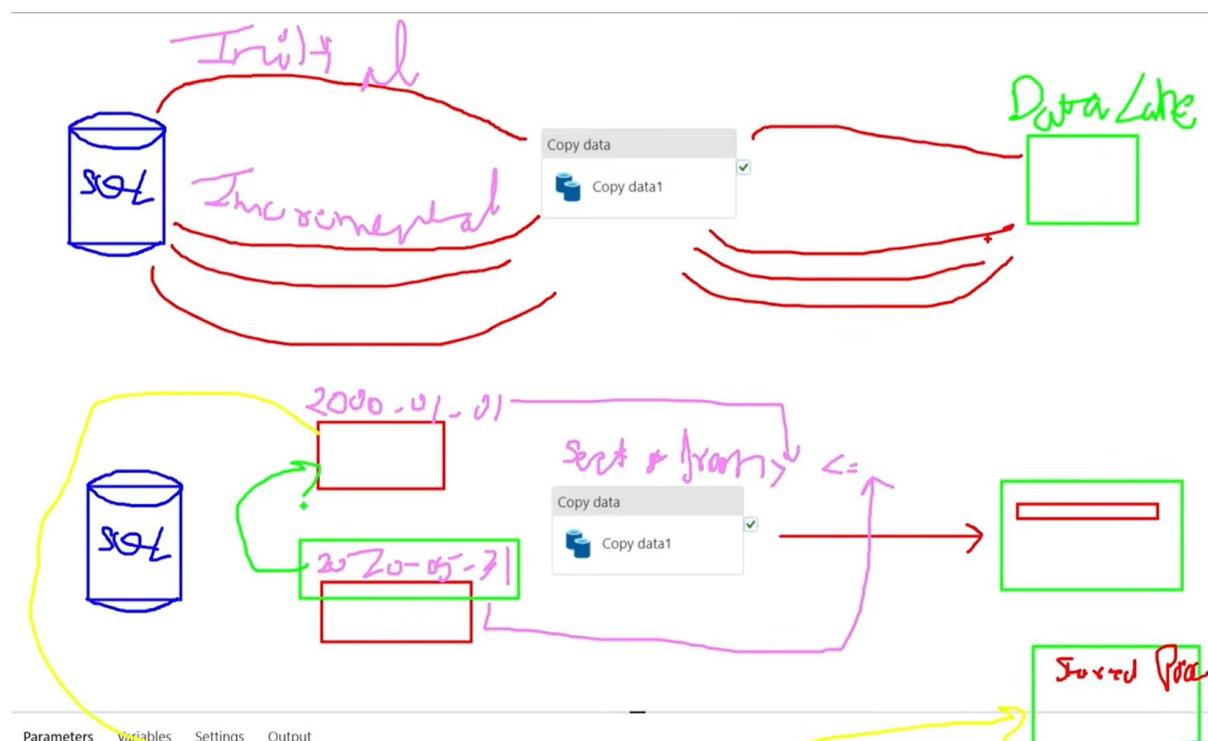
In this create two pipeline in incremental load

1) initial load pipeline



This is initial load . it load all the data to store in datalake

First time consider in instial load pipeline and after first time increment load pipeline use



Last date add in store procedure to use for incremental load

Let's create watermark table and store procedure

Click query section

carsalesdb (carsalessverdinesh/carsalesdb) | Query editor (preview)

Query 1

Run Cancel query Save query Export data as Show only Editor Open Copilot

Results Messages

Search to filter items...

Create water mark table for store last datetime of excute for store in this

carsalesdb (carsalessverdinesh/carsalesdb) | Query editor (preview)

Query 1

Run Cancel query Save query Export data as Show only Editor Open Copilot

```
1 CREATE TABLE water_table
2 (
3     last_load varchar(2000)
4 )
5
6 select * from water_table
```

Results Messages

Search to filter items...

last\_load

No results

Let check min of date id in source sql table

The screenshot shows the Microsoft Azure Query editor (preview) interface. The left sidebar lists database objects: Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview), Mirror database in Fabric (preview), Resource visualizer, Settings, Data management, Integrations, Power Platform, Security, Intelligent performance, Monitoring, Automation, Help, and a note about opening the full object explorer in Azure Data Studio. The main area displays a query in the 'Query 1' tab:

```

1 CREATE TABLE water_table
2 (
3     last_load varchar(2000)
4 )
5
6 select * from water_table
7
8 select min(Date_ID) from [dbo].[source_car_data]

```

The results pane shows the output of the query:

DT00001

Query succeeded | 0s

The screenshot shows the Microsoft Azure Query editor (preview) interface, identical to the first one but with a different query in the 'Query 1' tab:

```

1 CREATE TABLE water_table
2 (
3     last_load varchar(2000)
4 )
5
6 select * from water_table
7
8 select min(Date_ID) from [dbo].[source_car_data]
9
10 insert into [dbo].[water_table] values('DT00000')

```

The results pane shows the output of the query:

last\_load  
DT00000

Query succeeded | 0s

Create store procedure to add last date of data

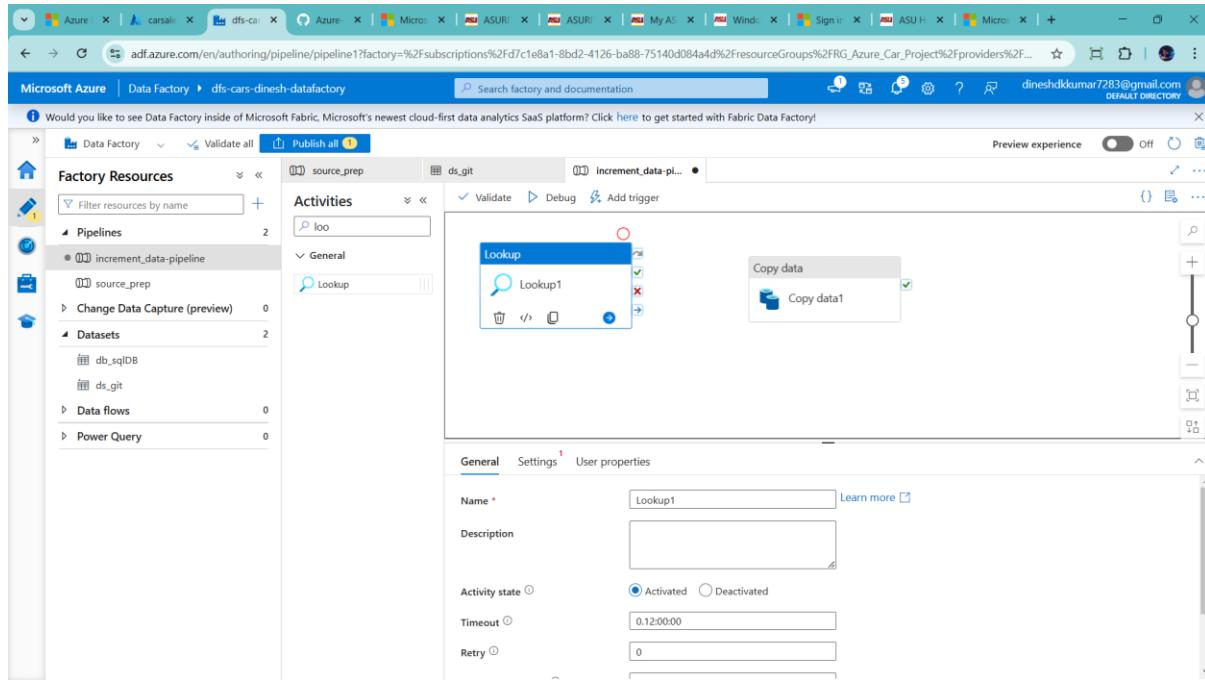
The screenshot shows the Microsoft Azure Query editor (preview) interface. The left sidebar lists database objects: Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview), Mirror database in Fabric (preview), Resource visualizer, Settings, Data management, Integrations, Power Platform, Security, Intelligent performance, Monitoring, Automation, Help, Tables, Views, and Stored Procedures. The 'Query editor (preview)' option is selected. The main area displays a query editor with two tabs: 'Query 1' and 'Query 2'. The 'Query 1' tab contains the following T-SQL script:

```
1 CREATE PROCEDURE UpdateWatermarkTable
2     @lastload varchar(200)
3     As
4     BEGIN
5         BEGIN TRANSACTION
6
7             UPDATE [dbo].[water_table]
8                 set last_load=@lastload
9         END;
```

And run it

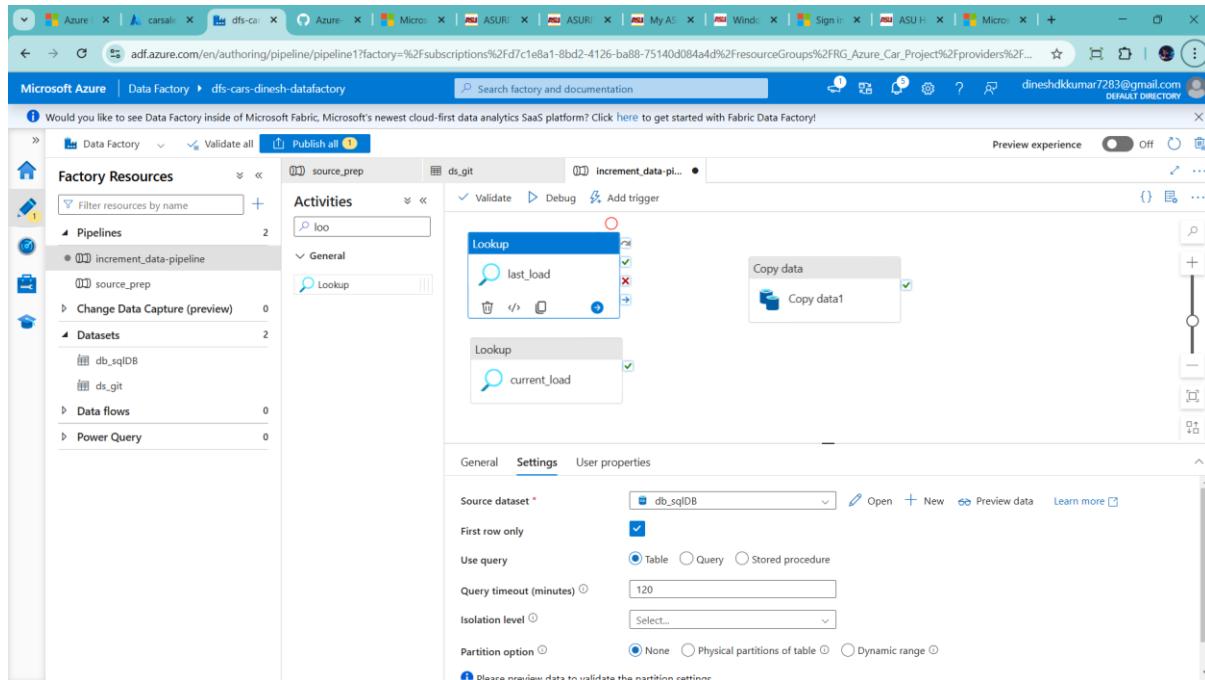
The screenshot shows the Microsoft Azure Query editor (preview) interface after running the stored procedure. The 'Results' tab shows the output: 'Query succeeded: Affected rows: 0'. The status bar at the bottom indicates 'Query succeeded | 0s'.

Go to data factory → click increment\_data\_pipeline (when you create it) → click activity use lookup



Use two lookup one is lastload and current load

Last\_load → use db\_sqldb → it is source dataset



The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. Under 'Datasets', 'db\_sqldb' is selected, which is an Azure SQL Database. The main panel displays the 'Connection' tab for 'db\_sqldb'. It shows a 'Linked service' dropdown set to 'ls\_sqldb', a 'Table' dropdown set to 'dbo.source\_car\_data', and a 'Test connection' button. A 'Parameters' tab is also visible.

Go to parameter → click new → give name as table name

The screenshot shows the Microsoft Azure Data Factory interface, similar to the previous one but with a focus on parameters. The 'Parameters' tab is selected in the 'db\_sqldb' dataset configuration. A new parameter is being added, with the 'Name' field set to 'table\_name', 'Type' set to 'String', and 'Default value' left empty. The 'New' and 'Delete' buttons are also visible.

Go to connection

The screenshot shows the Microsoft Azure Data Factory Pipeline expression builder. On the left, the 'Factory Resources' pane lists 'Pipelines' (increment\_data-pipeline, source\_prep), 'Datasets' (ds.git, db\_sqldb), and 'Data flows'. In the center, the 'source\_prep' pipeline is selected, showing its stages: source\_prep, ds.git, and increment\_data-pipe... The 'ds.git' stage is currently active. On the right, the 'Pipeline expression builder' interface is open, displaying the expression '@dataset().table\_name' in the main editor area. Below it, the 'Parameters' tab is selected, showing the 'table\_name' parameter with its value set to 'dbo'. There is also a 'Functions' tab. At the bottom are 'OK' and 'Cancel' buttons.

This screenshot is identical to the one above, but it includes a 'Preview experience' toggle switch at the top right of the interface. The switch is currently set to 'Off'. The rest of the interface, including the pipeline structure and the expression builder, remains the same.

Now go to increment\_data\_load → click last load flow → add table name water table

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The left sidebar lists 'Factory Resources' under 'Pipelines' (increment\_data\_pipeline), 'Datasets' (db\_sqlDB, ds\_git), and 'Data flows'. The main workspace shows a pipeline with three activities: 'Lookup' (last\_load), 'Lookup' (current\_load), and 'Copy data' (Copy data1). The 'Settings' tab for the first 'Lookup' activity is selected, showing the 'Source dataset' as 'db\_sqlDB' and the 'Name' field set to 'water\_table'.

Uncheck first row only → click query → select \* from water\_table

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The left sidebar lists 'Factory Resources' under 'Pipelines' (increment\_data\_pipeline), 'Datasets' (db\_sqlDB, ds\_git), and 'Data flows'. The main workspace shows a pipeline with three activities: 'Lookup' (last\_load), 'Lookup' (current\_load), and 'Copy data' (Copy data1). The 'Settings' tab for the first 'Lookup' activity is selected, showing the 'Query' field set to 'select \* from water\_table;' and the 'First row only' checkbox is unchecked.

Deactive both the debug it

Microsoft Azure | Data Factory > dfs-cars-dinesh-datafactory

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click [here](#) to get started with Fabric Data Factory!

Preview experience: Off

Factory Resources

- Pipelines: increment\_data\_pipeline (2)
- Datasets: db\_sqldb (1), ds\_git (1)
- Data flows: 0
- Power Query: 0

Activities

- source\_prep (1)
- increment\_data\_pipeline (1)
- General (1)

Lookup last\_load

Lookup current\_load

Copy data1

General

Name: Copy data1

Description:

Activity state: Deactivated

Mark activity as: Succeeded

Timeout: 0:12:00:00

Microsoft Azure | Data Factory > dfs-cars-dinesh-datafactory

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click [here](#) to get started with Fabric Data Factory!

Preview experience: On

Factory Resources

- Pipelines: increment\_data\_pipeline (2)
- Datasets: db\_sqldb (1), ds\_git (1)
- Data flows: 0
- Power Query: 0

Activities

- source\_prep (1)
- increment\_data\_pipeline (1)
- General (1)

Lookup last\_load

Lookup current\_load

Copy data1

Parameters

Pipeline run ID: 68588bc2-461e-4998-a013-c20e6a7cad37

Pipeline status: In progress

Showing 1 - 3 of 3 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime
last_load	Succeeded	Lookup	3/5/2025, 3:02:32 PM	5s	AutoResolveIntegrationRuntime (UK Sou...
current_load	Inactive	Lookup	3/5/2025, 3:02:32 PM	Less than 1s	Unknown
Copy data1	Inactive	Copy data	3/5/2025, 3:02:32 PM	Less than 1s	Unknown

Click output for success

The screenshot shows the Microsoft Azure Data Factory pipeline editor. A pipeline named 'increment\_data\_pipeline' is selected. The pipeline contains two activities: 'last\_load' (Lookup) and 'Copy data1'. The 'Output' tab is selected, displaying the results of the 'last\_load' activity. The output is a JSON object:

```
{
  "count": 1,
  "value": [
    {
      "last_load": "DT00000"
    }
  ]
}
```

The pipeline status is 'Succeeded'. The run details table shows the following data:

Activity name	Activity state	Activit...	Run start	Duration	Integration runtime
last_load	Succeeded	Lookup	3/5/2025, 3:02:32 PM	5s	AutoResolveIntegrationRuntime (UK Sou...
current_load	Inactive	Lookup	3/5/2025, 3:02:32 PM	Less than 1s	Unknown
Copy data1	Inactive	Copy data	3/5/2025, 3:02:32 PM	Less than 1s	Unknown

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The pipeline 'increment\_data\_pipeline' is selected. The pipeline contains two activities: 'last\_load' (Lookup) and 'Copy data1'. The 'Output' tab is selected, displaying the results of the 'last\_load' activity. The output is a JSON object:

```
{
  "count": 1,
  "value": [
    {
      "last_load": "DT00000"
    }
  ]
}
```

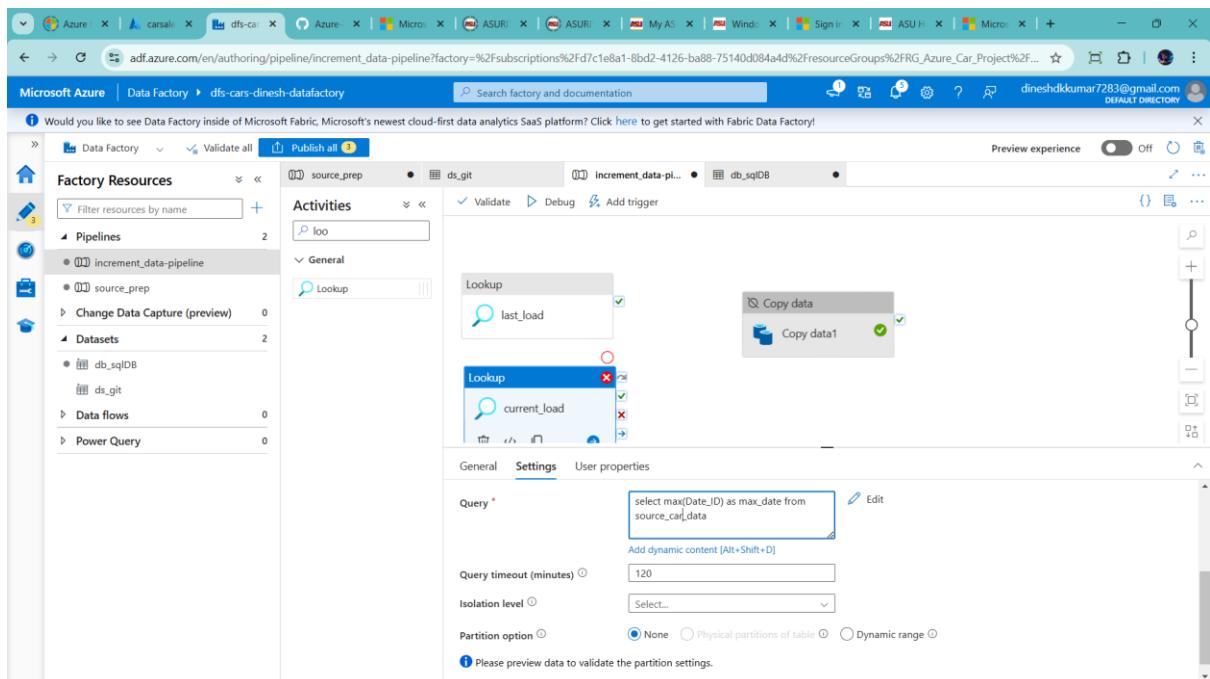
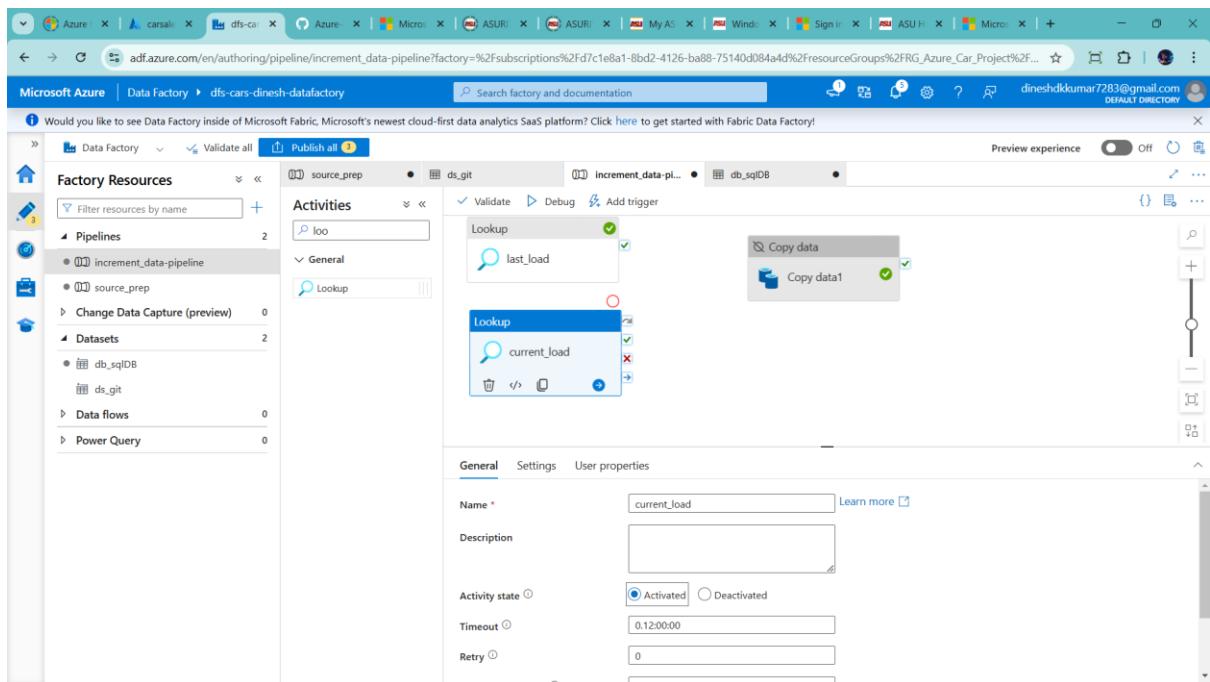
The pipeline status is 'Succeeded'. The run details table shows the following data:

Activity name	Activity state	Activit...	Run start	Duration	Integration runtime
last_load	Succeeded	Lookup	3/5/2025, 3:02:32 PM	5s	AutoResolveIntegrationRuntime (UK Sou...
current_load	Inactive	Lookup	3/5/2025, 3:02:32 PM	Less than 1s	Unknown
Copy data1	Inactive	Copy data	3/5/2025, 3:02:32 PM	Less than 1s	Unknown

Now change the name loopup card for current load → add parameter to pass name table name source\_cars\_data

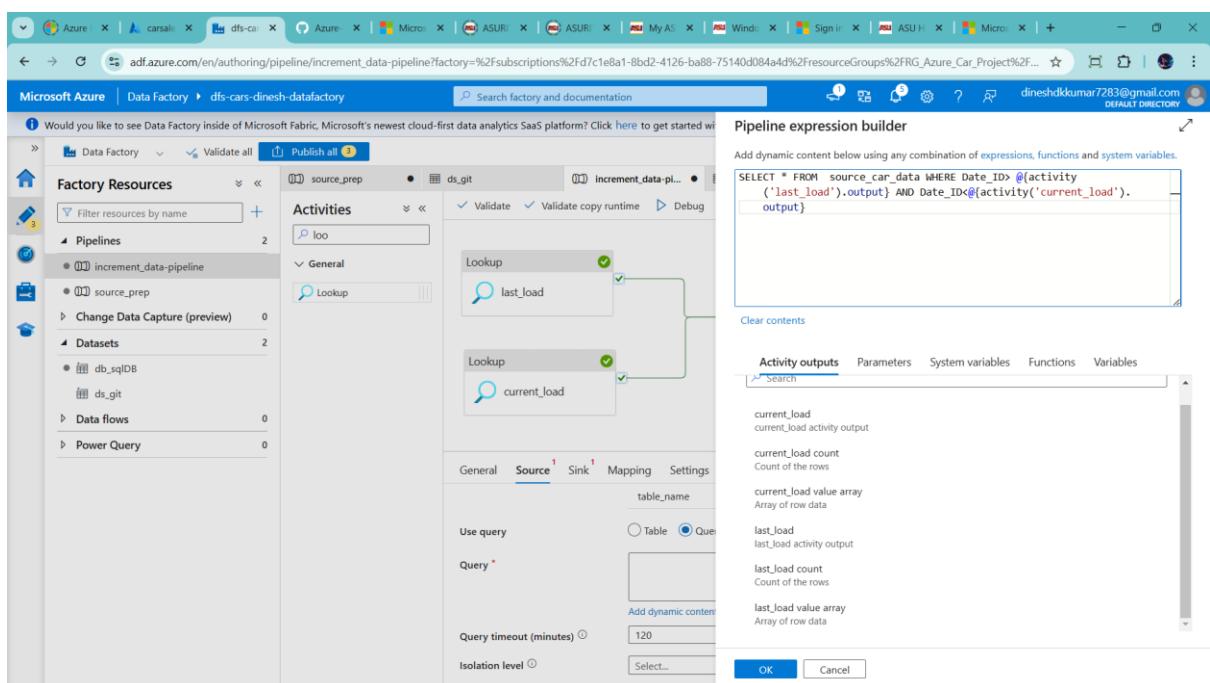
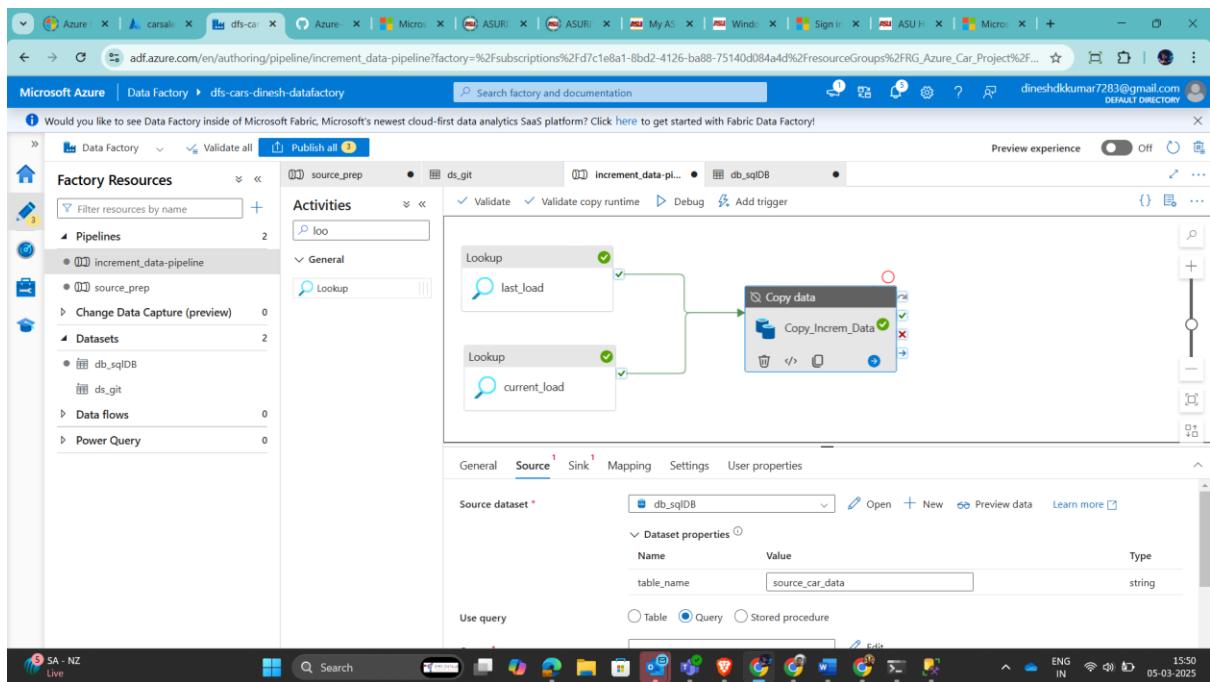
Uncheck first row only → use query -check Query

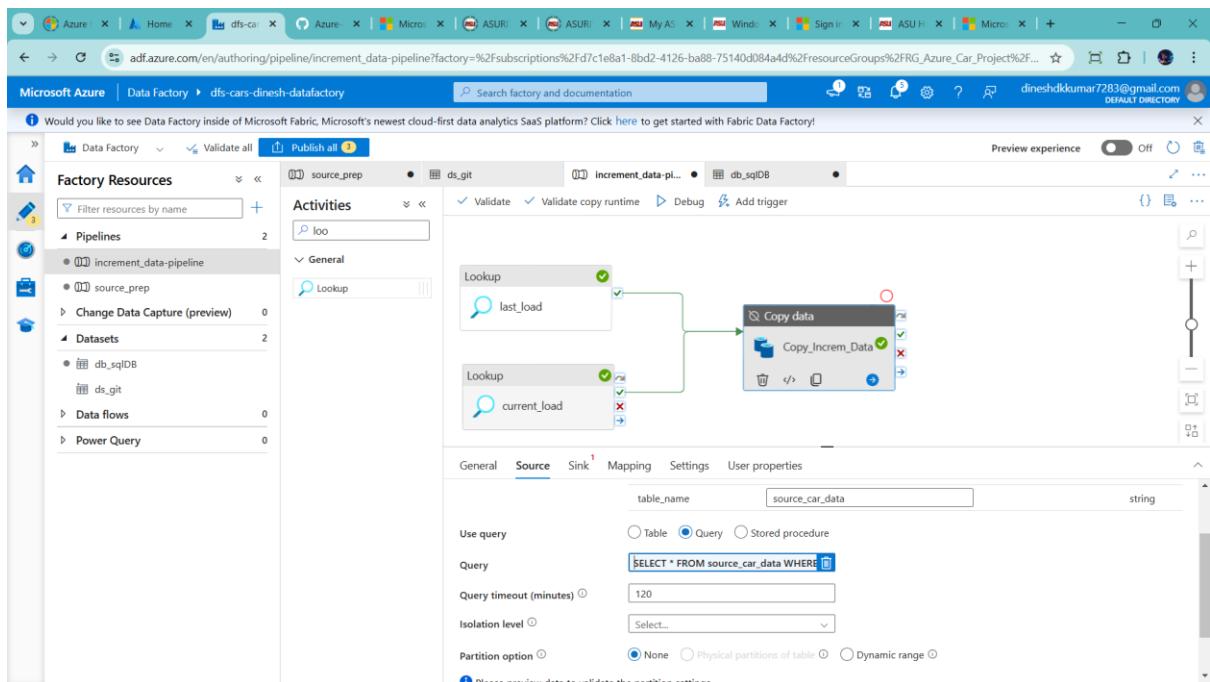
Final activate



Not change as well Copy Data1 to rename →Copy\_Increm\_Data→click source and dynamic name- :  
source\_car\_data → and click Query

→before write query →connect with two lookup to copy data/ sql

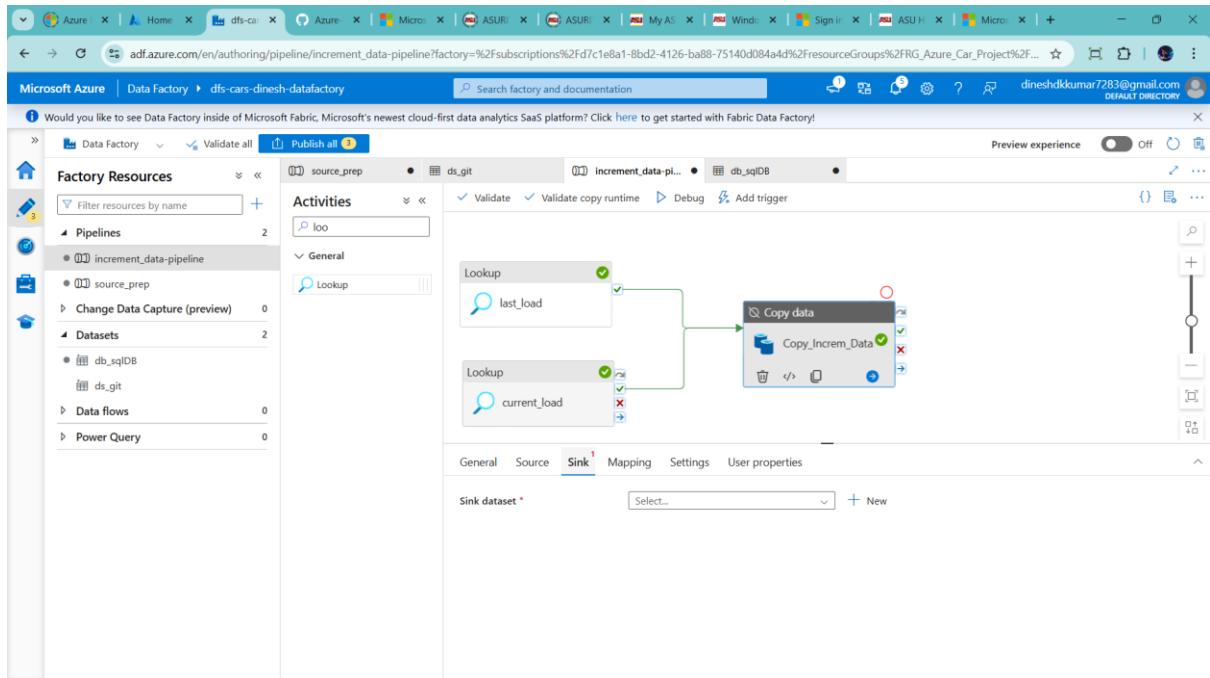




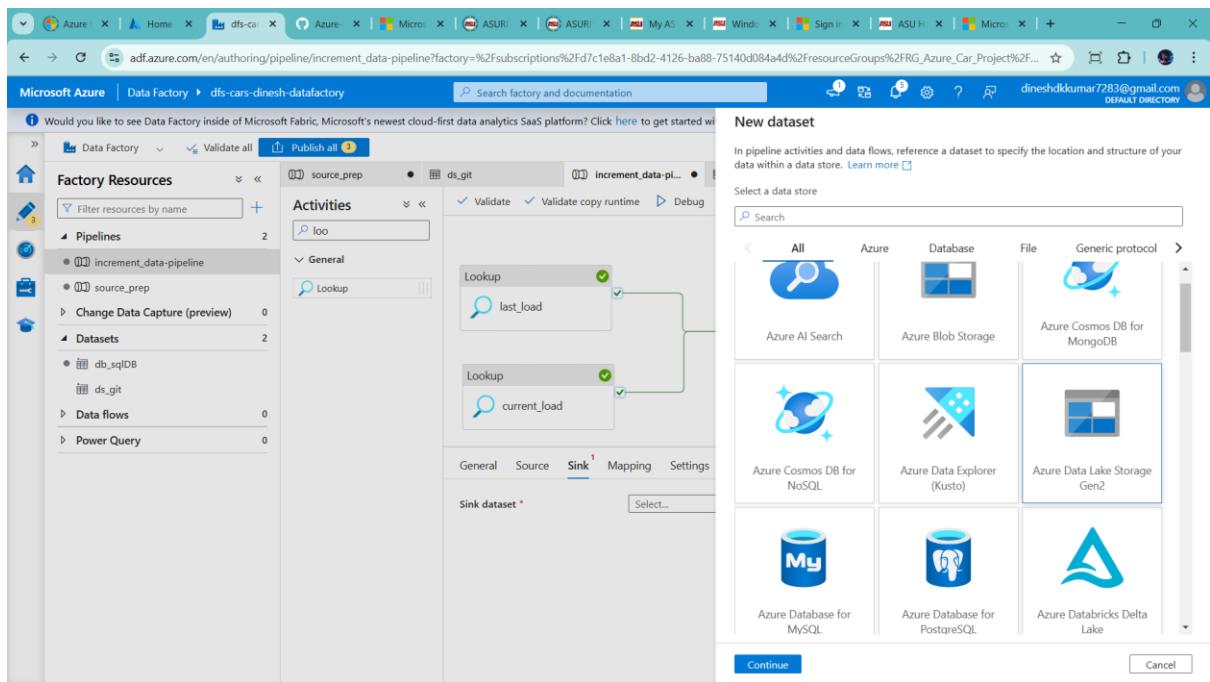
Let understood

SQL → Data Factory → Datalake (Brozen Container)

You can create data lake → based on sink /manager tap

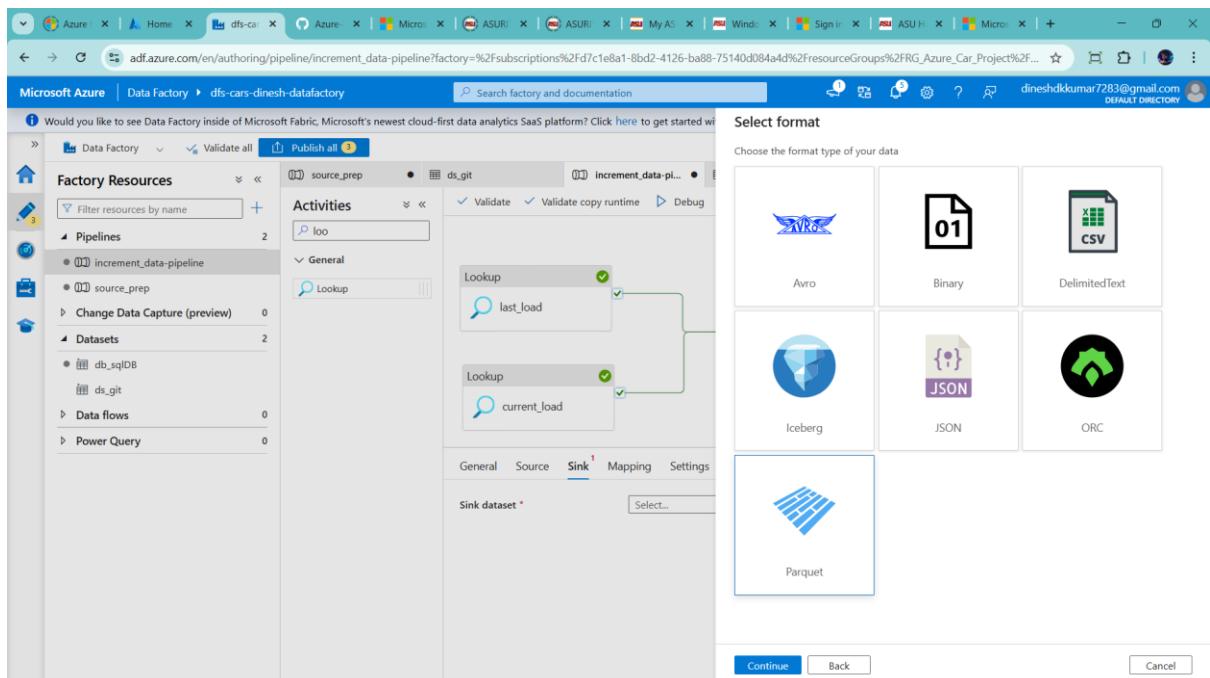


--in sink → click new →azure data lake gen 2



--click continue

→parquet formate to store data

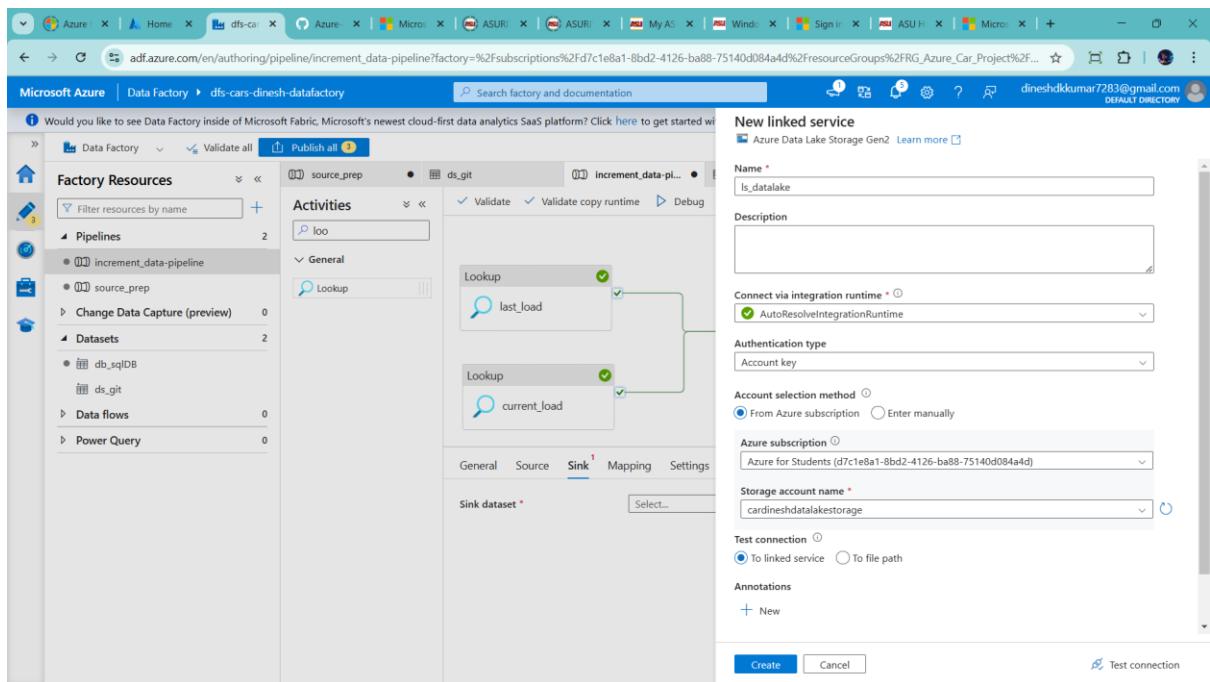


Click continue

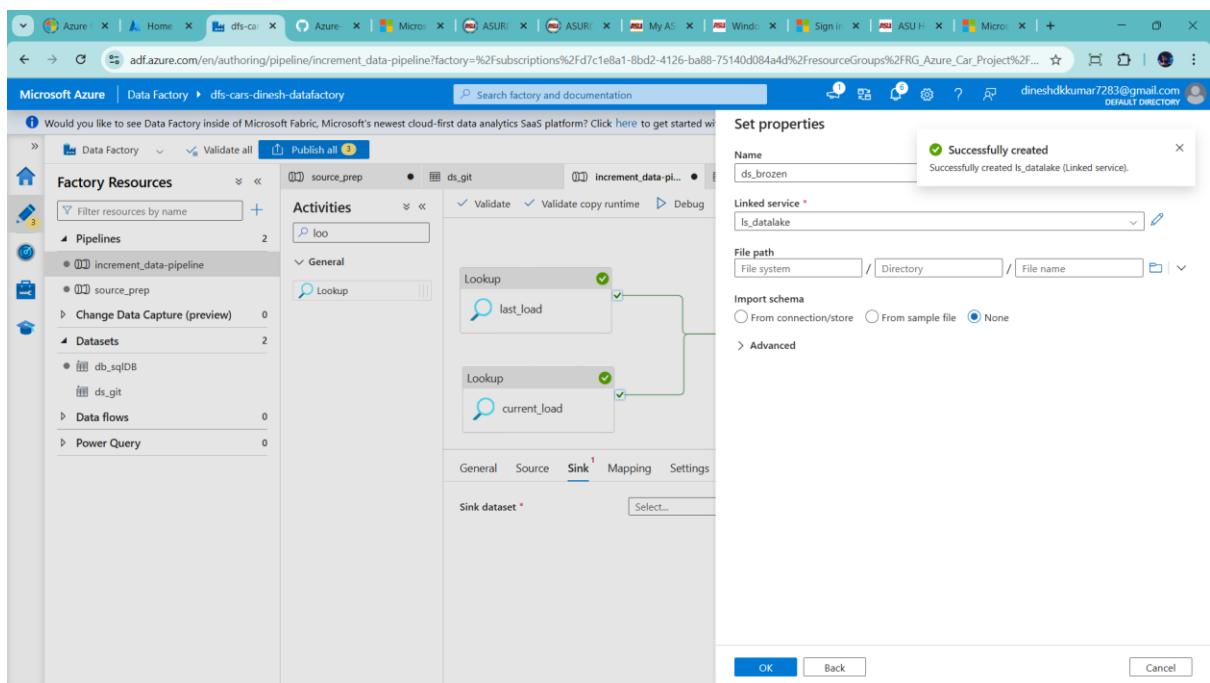
→name as ds\_brozen →linked service to ls\_datalake

→Azure subriscation chage to azure for student

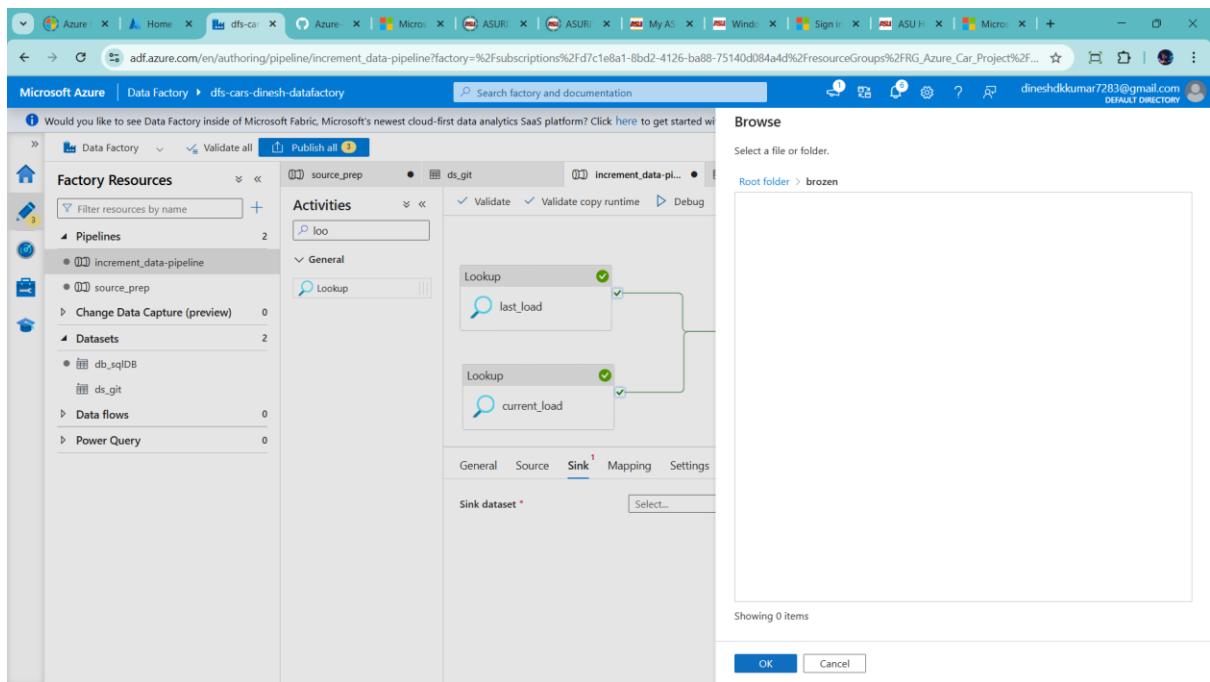
→storage account name



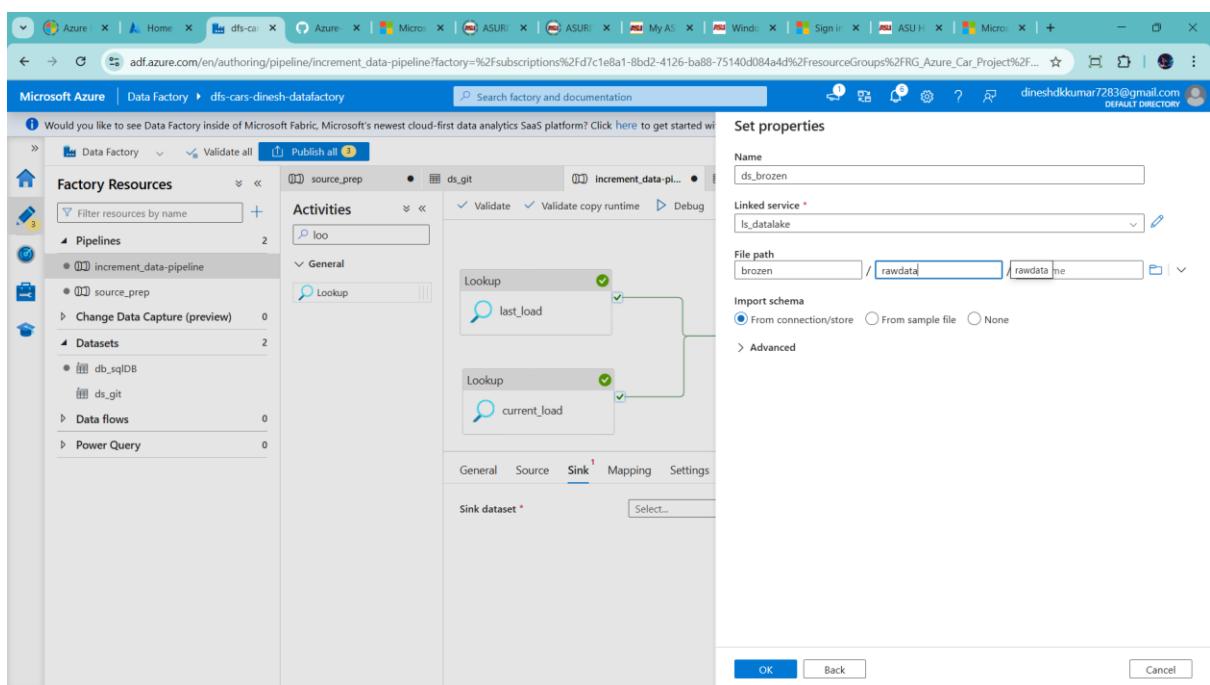
→click create



Click broken file path →click ok



Add directory → raw data



Click import schema none

The screenshot shows the Microsoft 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 'increment\_data\_pipeline'. It contains two 'Lookup' activities: 'last\_load' and 'current\_load', which are connected to a 'Sink' dataset named 'ds\_brozen'. The 'Sink' tab is selected. The 'File path' is set to 'brozen' / 'rawdata' / 'File name'. The 'Import schema' section shows 'None' selected. A warning message at the bottom states: 'Schema import failed: ADLS Gen2 operation failed for: Operation returned an invalid status code 'NotFound'. Account: 'cardineshdatalakestorage'. FileSystem: 'brozen', Path: 'rawdata'. ErrorCode: 'PathNotFound'. Message: 'The specified path does not exist.'. RequestId: 6d97f56a-401f-0071-73c4-8d5731000000. Timestamp: Wed, 05 Mar 2025 11:40:37 GMT.\nOperation returned an invalid status code 'NotFound'.'

→ now go mapping → click import schemas → add values based on query

The screenshot shows the Microsoft Azure Data Factory pipeline editor with the 'Mapping' tab selected. In the 'Import schemas' section, there is a table with two rows:

Name	Type	Value
@activity('last_load').output	st... ▾	<input type="text"/> Value Required
@activity('current_load').output	st... ▾	<input type="text"/> Value Required

At the bottom, there is a note: 'Add dynamic content [Alt+Shift+D]'.

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/?Microsoft\\_Azure\\_Education\\_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft\\_Azure\\_Education\\_newA4E=true&Microsoft\\_Azure\\_Education...](https://portal.azure.com/?Microsoft_Azure_Education_correlationId=993e75d6-70a9-47f3-a369-212146ab3e78&Microsoft_Azure_Education_newA4E=true&Microsoft_Azure_Education...). The user is signed in as dineshkumar7283@gmail.com. The left sidebar shows the 'carsalesdb' database with various management options like Overview, Activity log, Tags, and the selected 'Query editor (preview)'. The main area displays a query editor with the following content:

```
1 select * from [dbo].[water_table]
```

The results pane shows a single row:

last_load
DT00000

A status bar at the bottom indicates "Query succeeded | 0s".

This screenshot shows the same Microsoft Azure portal setup as the first one, but with two separate query executions in the editor.

The first query is:

```
1 select * from [dbo].[water_table]
```

The second query is:

```
2 select max(Date_ID) from [dbo].[source_car_data]
```

The results pane shows the output of the second query:

Date_ID
DT01245

A status bar at the bottom indicates "Query succeeded | 0s".

Please provide actual value of the parameters to get schema

Please provide actual value of the parameters for pipeline increment\_data-pipeline

Name	Type	Value
No records found		

Additional expressions

Name	Type	Value
@activity('last_load').output	String	DT00000
@activity('current_load').output	String	DT01245

--click ok → they show error change sql query

--change query as per my respectivity

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
SELECT * FROM source_car_data
WHERE Date_ID > '@{activity('last_load').output.value[0].last_load}'
AND Date_ID <= '@{activity('current_load').output.value[0].max_date}'
```

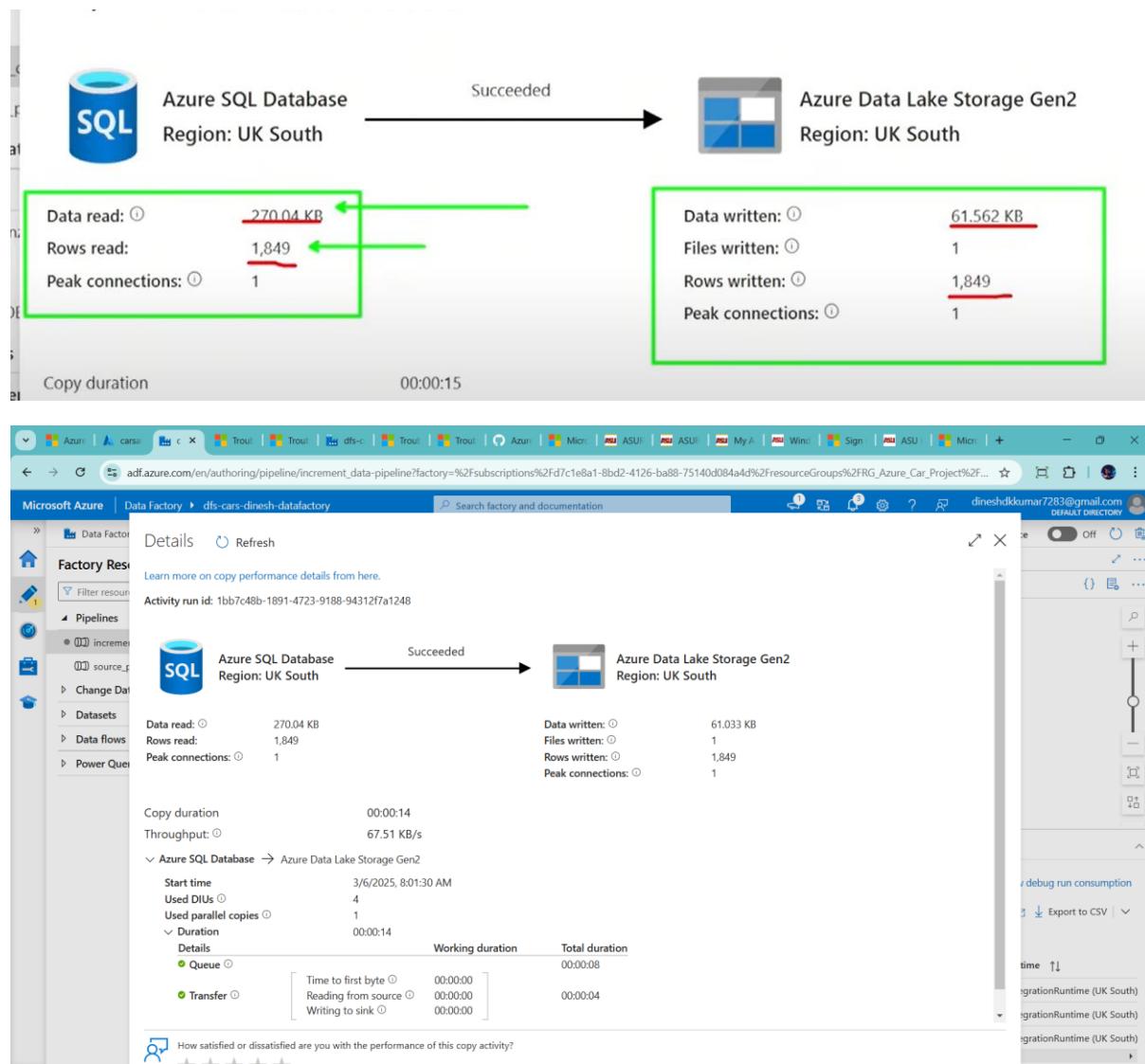
Clear contents

Activity outputs Parameters System variables Functions Variables

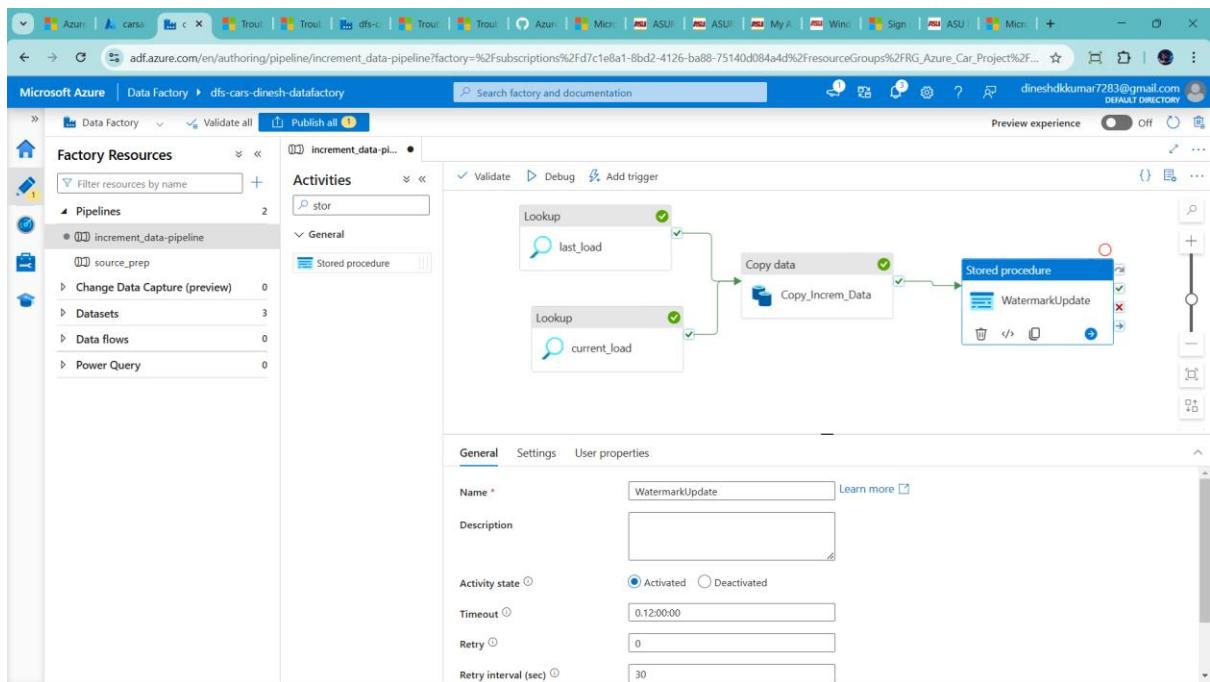
Search

current\_load  
current\_load.activity.output  
current\_load.count  
Count of the rows  
current\_load.value.array  
Array of row data  
last\_load  
last\_load.activity.output  
last\_load.count  
Count of the rows  
last\_load.value.array

## Debug it all activity

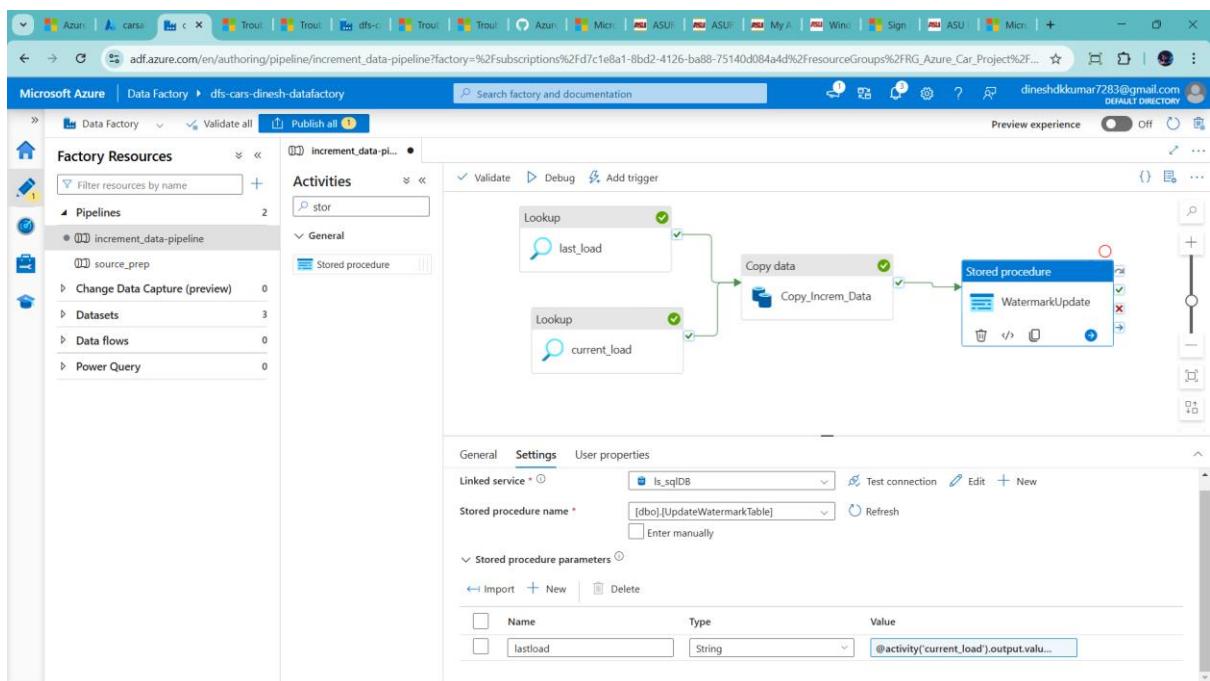


Add store procedure



Add in setting section

Linked service, storeprocedure name, click import parameter



Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
WatermarkUpdate	Failed	Stored procedure	3/6/2025, 8:13:57 AM	6s	AutoResolveIntegrationRuntime (UK South)
Copy_Increm_Data	Succeeded	Copy data	3/6/2025, 8:13:41 AM	16s	AutoResolveIntegrationRuntime (UK South)
current_load	Succeeded	Lookup	3/6/2025, 8:13:26 AM	15s	AutoResolveIntegrationRuntime (UK South)
last_load	Succeeded	Lookup	3/6/2025, 8:13:26 AM	7s	AutoResolveIntegrationRuntime (UK South)

Handle this error

Go edit store procedure

DROP PROCEDURE AND RECREATE

The screenshot shows the Microsoft Azure portal with the Query editor (preview) open for the database 'carsalesdb'. The left sidebar lists various database management options like Overview, Activity log, Tags, Diagnose and solve problems, and the current selected 'Query editor (preview)'. The main area displays a code editor with two tabs: 'Query 1' and 'Query 2'. 'Query 1' contains the following T-SQL script:

```
1 DROP PROCEDURE UpdateWatermarkTable;
2
3 CREATE PROCEDURE UpdateWatermarkTable
4     @lastload varchar(200)
5 As
6 BEGIN
7     BEGIN TRANSACTION;
8
9     UPDATE [dbo].[water_table]
10    set last_load=@lastload
```

The 'Messages' tab below the code editor shows the message: 'Query succeeded: Affected rows: 0'. A status bar at the bottom indicates 'Query succeeded | 0s'.

```
DROP PROCEDURE UpdateWatermarkTable;
```

```
CREATE PROCEDURE UpdateWatermarkTable
```

```
    @lastload varchar(200)
```

```
As
```

```
BEGIN
```

```
    BEGIN TRANSACTION;
```

```
    UPDATE [dbo].[water_table]
```

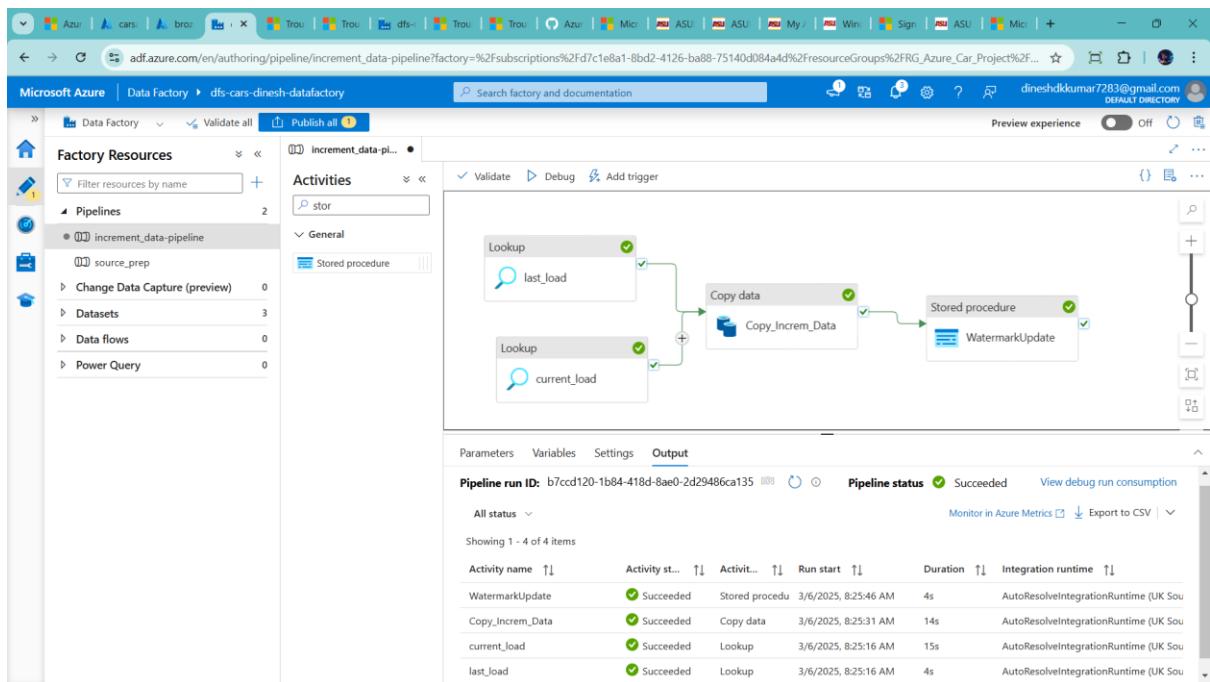
```
        set last_load=@lastload
```

```
    commit transaction;
```

```
END;
```

```
--delete data lake rawdata folder
```

```
And rerun pipeline
```



Successful updated watermark table

The screenshot shows the Microsoft Azure SQL Database Query Editor. The left sidebar shows the database structure with tables, views, and stored procedures. The main area displays a query in 'Query 1':

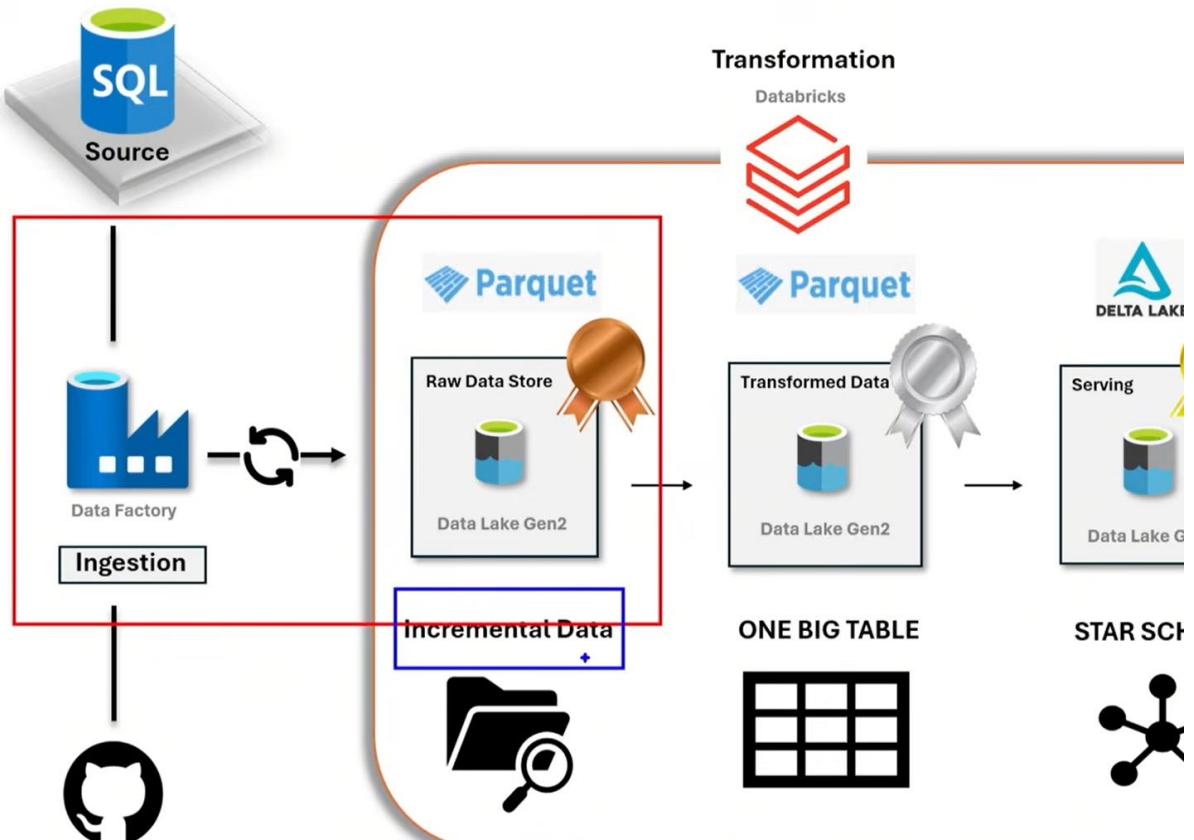
```
1 select * from [dbo].[water_table]
```

The results pane shows the output of the query:

Results
last_load
DT01245

A message at the bottom indicates: 'Query succeeded | 0s'.

Now Success full Complete First part



# Let Start Create Databricks Resource

Go resource → click create

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes links for Home, portal.azure.com, and various Azure services like Storage, Compute, and Monitoring. The main content area is titled 'RG\_Azure\_Car\_Project' under 'Resource group'. The 'Overview' tab is selected. Key details shown include:

- Subscription (move) : Azure for Students
- Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d
- Tags (edit) : Add tags
- Deployments : 3 Succeeded
- Location : UK South

The 'Resources' section displays a table of four resources:

Name	Type	Location	Actions
cardineshdatastorage	Storage account	UK South	...
carsalesdb	SQL database	UK South	...
carsalesserverdinesh	SQL server	UK South	...
dfscars-dinesh-datafactory	Data factory (V2)	UK South	...

At the bottom, there are navigation buttons for 'Previous', 'Page 1 of 1', and 'Next >'. A 'Give feedback' link is also present.

The screenshot shows the Microsoft Azure Marketplace search results for 'databricks'. The search bar at the top contains the query 'databricks'. Below the search bar, there are filters for Pricing, Operating System, Publisher Type, Product Type, and Publisher name, all set to 'All'. A checkbox for 'Azure services only' is checked. A message box says 'New! Get AI-generated suggestions for "databricks"' with a 'View suggestions' button. The results table shows 20 of 71 items:

Offer	Publisher	Description	Starts at	Price	Action
Azure Databricks	Microsoft	Azure Databricks is the fast, easy and collaborative Apache Spark-based analytics platform.	Free	Price varies	Create
Lakehouse Optimizer for Azure Databricks	Blueprint Technologies LLC	Provides near real-time optimization and granular utilization and performance transparency.	Free	Price varies	Create
Access Connector for Azure Databricks	Microsoft	Azure Service	Price varies	Price varies	Create
Unravel for Azure Databricks	Unravel Data	Azure Application	Price varies	Price varies	Create
Unravel for Azure Databricks (PayGo)	Unravel Data	SaaS	Price varies	Subscribe	

The left sidebar includes sections for Get Started, Service Providers, Management, My Marketplace, and Categories. The 'My Marketplace' section shows recent activity: Favorites (0), My solutions (0), Recently created (0), and Private plans (0). The 'Categories' section lists Analytics (58), AI + Machine Learning (22), Databases (16), IT & Management Tools (12), and Compute (4).

## → Cars databricks

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

Subscription \*  Resource group \*

Workspace name \*  Region \*  Pricing Tier \*   
We selected the recommended pricing tier for your workspace. You can change the tier based on your needs.

Managed Resource Group name

**Review + create** < Previous Next : Networking >

- Put workspace name
- Add change price Tier → trial 14 days
- Add manager resource group →(use for catalog,etc)
- Click review+create

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

Subscription \*  Resource group \*

Workspace name \*  Region \*  Pricing Tier \*   
We selected the recommended pricing tier for your workspace. You can change the tier based on your needs.

Managed Resource Group name

**Review + create** < Previous Next : Networking >

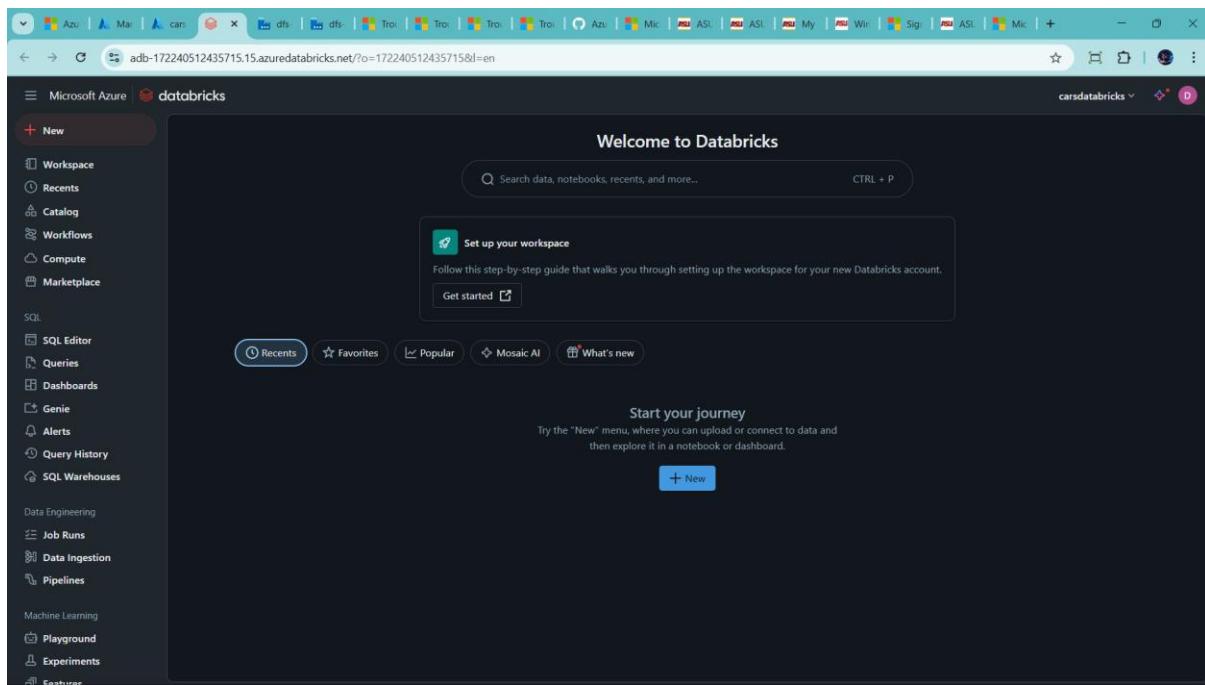
→ click create and completed

The screenshot shows the Microsoft Azure portal with a deployment overview page. The deployment name is "RG\_Azure\_Car\_Project\_carsdatabricks". The status is "Your deployment is complete". Deployment details include a start time of 3/6/2025, 10:12:49 AM, and a correlation ID of b0a168b2-a23f-405c-b80d-e171daaddd4. A sidebar on the right provides links to Cost management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

Click go to workspace

The screenshot shows the Microsoft Azure portal with the Azure Databricks service overview. The service name is "carsdatabricks". Key details include Status: Active, Resource group: RG\_Azure\_Car\_Project, Location: UK South, Subscription: Azure for Students, and Subscription ID: d7c1e8a1-8bd2-4126-ba88-75140d084a4d. A large red icon of three stacked cubes is displayed, along with "Launch Workspace" and "Upgrade to Premium" buttons. Below are links for Documentation, Getting Started, Import Data from File, Import Data from Azure Storage, Notebook, Admin Guide, and Link Azure ML workspace.

Click launch workspace

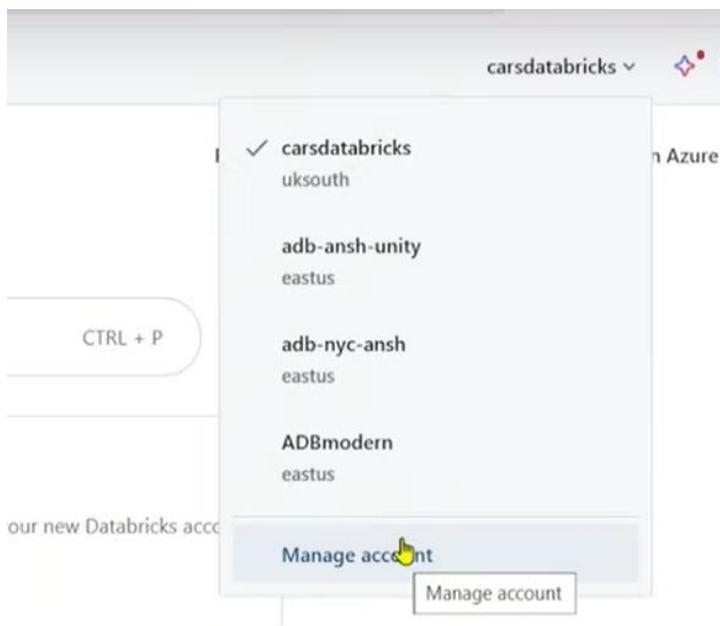


- workspace → manage folder, notebook and maintain hierarchy
- Recents → view all the recent items → notebook, delta table, live table and more in recent tap
- Catalog → manage databases/ table, schema and column
- Workflows → it like ETL frameworks, build pipeline low/no code,
- compute → create a clusters
- SQL → it is section used for data analysis
- data engineering → have 3 frameworks
  - One is pipelines/Delta live tables → it is ETL flow
  - Machine learning → build models
  - Marketplace → use service for 3<sup>rd</sup> party like azure, more

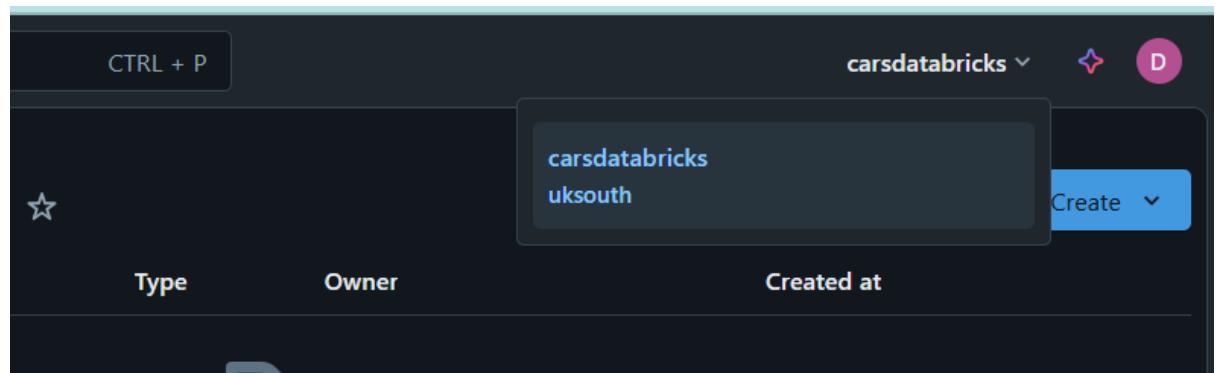
## First Create Unity Metastore

→ but in databricks not able to create unity metastore

→ first change in Admin in databricks



- But over databricks not show manage account option
- Let how to enable



- First go to azure account → click Microsoft entra id

The screenshot shows the Microsoft Azure portal homepage. The search bar at the top contains the text "entra". Below the search bar, there are sections for "Azure services" (Create a resource, Resource groups), "Recent" resources (RG\_Azure\_Car\_Project, carsalesdb, carsalesserverdinesh, dfs-cars-dinesh-datafactory, cardineshdatastorage), and "Navigate" (Subscriptions). On the right side, there is a sidebar with "Services (25)", "Marketplace (28)", "Workspaces", "Quickstart Center", and "More services". A "Last Viewed" section lists items from 2 hours ago to 23 hours ago. A "Frequently asked questions (FAQ) about Microsoft Entra, Microsoft Entra ID, and A..." link is also present.

The screenshot shows the Microsoft Entra ID Overview page. The URL is https://portal.azure.com/#view/Microsoft\_AAD\_IAM/ActiveDirectoryMenuBlade/~/Overview. The page includes a navigation menu on the left with options like Overview, Preview features, Diagnose and solve problems, Manage, Monitoring, and Troubleshooting + Support. The main content area shows basic information for the tenant (Name: Default Directory, Tenant ID: 1eadda0c-f714-4f5e-8af6-f435b63d1d8e, Primary domain: dineshdkumar7283@gmail.onmicrosoft.com, License: Microsoft Entra ID Free), users (1), groups (0), applications (0), and devices (0). There are two alerts at the bottom: "MSOnline PowerShell Retirement" (Learn more) and "Migrate to the converged Authentication methods policy" (Learn more).

→ go to manage → user → copy this user mail id with #tag

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes links for Home, Default Directory, and Users. The main content area is titled 'Users' under 'Default Directory'. A sub-menu 'All users' is open, showing a list of audit logs, sign-in logs, and other diagnostic tools. On the right, a detailed view of a single user named 'Dinesh km' is displayed. The user's principal name is 'dineshkumar7283@gmail.com#EXT#@dineshkumar7283@gmail.onmicrosoft.com'. The user type is 'Member', they are 'On-premises synchronized', and their identity provider is 'MicrosoftAccount'. There are also tabs for 'Audit logs', 'Sign-in logs', 'Diagnose and solve problems', 'Deleted users', 'Password reset', 'User settings', 'Bulk operation results', and 'New support request'.



Login in another portal → accounts.azuredatabricks.net

→ first reset the password → now login in the #tag mail id

→ success open Admin account

The screenshot shows the Databricks account management portal. The top navigation bar has links for Account, Workspaces, Catalog, Usage, User management, Cloud resources, Previews, and Settings. The main content area is titled 'Welcome to Databricks' and says 'Manage your Databricks account'. It features six cards: 'Workspaces' (Configure workspace settings), 'Catalog' (Manage metastores), 'Usage' (View usage details and graphs), 'Users & groups' (Manage identities), 'Cloud resources' (Manage network connectivity), and 'Settings' (Configure user provisioning). The URL in the address bar is 'accounts.azuredatabricks.net/?account\_id=242342a8-9b15-4cf8-a947-a6be29fa935b'.

→ go to user management section

The screenshot shows the 'User management' page in the Azure portal. The left sidebar includes options like Account, Workspaces, Catalog, Usage, User management (which is selected), Cloud resources, Previews, and Settings. The main area is titled 'User management' and contains tabs for 'Users', 'Service principals', and 'Groups'. A search bar at the top allows filtering by name or email. Below the search bar, there's a note about account users and account admins. A table lists two users: 'Dinesh km' and 'Dinesh km'. The first 'Dinesh km' has the email 'dineshdikumar7283@gmail.com' and the role 'Account admin'. The second 'Dinesh km' has the email 'dineshdikumar7283@gmail.com#ext#@dineshdikumar7283gmail.on...' and the role 'Account admin'. There are buttons for 'Add user' and navigation arrows at the bottom.

→ add new user or already present → click the name

The screenshot shows the 'User information' page for 'Dinesh km' in the Azure portal. The left sidebar is identical to the previous screenshot. The main area shows 'User information' and 'Roles' tabs. Under 'User information', there's a 'Role name' dropdown set to 'Account admin', a 'Display name' input field with 'Dinesh km' typed in, and a 'Email' input field with 'dineshdikumar7283@gmail.com' typed in. Under 'Roles', there are three listed: 'Account admin' (description: 'Can manage workspaces, users & groups, cloud resources and settings.'), 'Marketplace admin' (description: 'Can manage exchanges and listings on the Marketplace. This change might take a few minutes to update.'), and 'Billing admin' (description: 'Can view Budgets and create budget policies. This change might take a few minutes to update.').

→ click Admin account active

The screenshot shows the 'User management' section for the user 'Dinesh km'. The left sidebar includes options like Workspaces, Catalog, Usage, User management (which is selected), Cloud resources, Previews, and Settings. The main area displays the 'Roles' tab for 'Dinesh km', listing three roles: Account admin, Marketplace admin, and Billing admin. Each role has a description and a toggle switch.

Role name	Description
Account admin	Can manage workspaces, users & groups, cloud resources and settings.
Marketplace admin	Can manage exchanges and listings on the Marketplace. This change might take a few minutes to update.
Billing admin	Can view Budgets and create budget policies. This change might take a few minutes to update.

The screenshot shows the 'User management' list page. The left sidebar is identical to the previous screen. The main area lists two users: 'Dinesh km' and another entry for 'Dinesh km'. The table columns are Status, Name, Email, and Roles. Both entries are marked as 'Active' and have the 'Account admin' role assigned. A blue 'Add user' button is visible at the top right of the table.

Status	Name	Email	Roles
Active	Dinesh km	dineshdkumar7283@gmail.com	Account admin
Active	Dinesh km	dineshdkumar7283_gmail.com#ext#@dineshdkumar7283gmail.on...	Account admin

Now check catalog in admin console → the already create unity catalog

The screenshot shows the Azure DataBricks Catalog interface. On the left, a sidebar menu includes 'Workspaces', 'Catalog' (which is selected), 'Usage', 'User management', 'Cloud resources', 'Previews', and 'Settings'. The main content area is titled 'Catalog' and 'Metastores'. A sub-section titled 'A metastore is the top-level container for catalog in Unity Catalog. Within a metastore, Unity Catalog provides a 3-level namespace for organizing data: catalogs, schemas (also called databases), and tables / views. Learn more' is present. Below this is a search bar labeled 'Filter metastores' and a 'Create metastore' button. A table lists one metastore entry:

Name	Region	Path	Created at	Updated at
metastore_azure_uksouth	uksouth		today at 10:15 AM	today at 10:15 AM

→ now I deleted my unity catalog

-→ only one region same unity catalog possible to create

→ first click the name present in catalog

The screenshot shows the configuration page for the 'metastore\_azure\_uksouth' metastore. The left sidebar remains the same. The main content area shows the 'Configuration' tab selected. It includes fields for 'ADLS Gen 2 path' (set to '-'), 'Region' (set to 'uksouth'), 'Metastore Admin' (set to 'System user'), 'Delta Sharing' (checkbox for 'Allow Delta Sharing with parties outside your organization' is unchecked), and 'Workspace assignment' (checkbox for 'Automatically assign new workspaces in uksouth to this metastore' is checked). The URL in the browser bar is 'accounts.azuredatabricks.net/data/d7935ab6-6828-4e8c-9377-7ee3602b6103/configurations?account\_id=242342a8-9b15-4cf8-a947-a6be29fa935b'.

Click 3 dots → delete

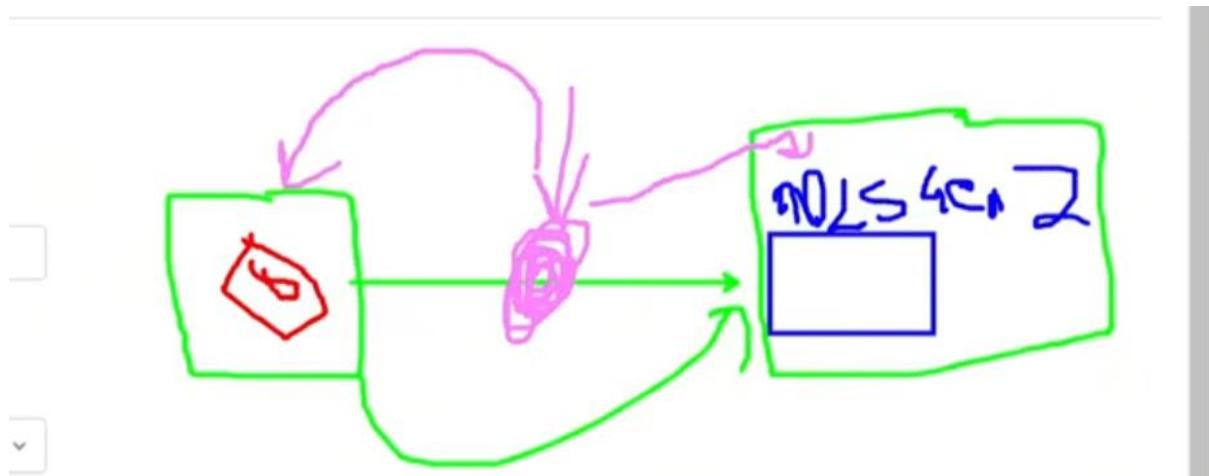
The screenshot shows the Azure Databricks Catalog Metastores page. On the left, there's a sidebar with options like Workspaces, Catalog (which is selected), Usage, User management, Cloud resources, Previews, and Settings. The main area is titled 'Catalog' and 'Metastores'. It contains a brief description of what a metastore is, a search bar labeled 'Filter metastores', and a 'Create metastore' button. A table below lists 'Name', 'Region', 'Path', 'Created at', and 'Updated at'. The table shows 'No metastores'.

- Now create manually
- Click create metasote

The screenshot shows the 'Create metastore' wizard, step 1: 'Create metastore'. It has two tabs: 'Create metastore' (selected) and 'Assign to workspaces'. The 'Create metastore' tab has fields for 'Name' (with placeholder 'My metastore') and 'Region' (with placeholder 'Select the region for your metastore. You will only be able to assign workspaces in this region to this metastore.'). Below these are optional fields: 'ADLS Gen 2 path (optional)' with placeholder '<container\_name> @ <storage\_account\_name>.dfs.core.windows.net/<path>' and 'Access Connector Id' with placeholder '/subscriptions/{sub-id}/resourceGroups/{rg-name}/providers/Microsoft.Databricks/accessConnect...'. There's also an 'Advanced options' button. At the bottom are 'Create' and 'Cancel' buttons.

→ fill it

→ before databricks → connector → ADLS Gen 2



- Databricks have permission to use it datalake
- That use access connector id in account admin
- Let create Access connector in azure account
- go to azure resource group

Screenshot of the Microsoft Azure portal showing the 'RG\_Azure\_Car\_Project' resource group overview. The page displays basic information like subscription details and a list of resources.

Essentials	
Subscription (more) : Azure for Students	Deployments : 4 Succeeded
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d	Location : UK South
Tags (edit) : Add tags	

Resources	
<input type="checkbox"/> Name ↑	Type ↑↓
<input type="checkbox"/> cardinesdatalakestorage	Storage account
<input type="checkbox"/> carsalesdb (carsaleserverdinesh/carsalesdb)	SQL database
<input type="checkbox"/> carsaleserverdinesh	SQL server
<input type="checkbox"/> carsdatabricks	Azure Databricks Service
<input type="checkbox"/> dts-cars-dinesh-datafactory	Data factory (V2)

--click create → search databricks → find access connector for azure databricks

→ click create

Microsoft Azure Marketplace

Search resources, services, and docs (G+)

dineshkumar7283@gmail.com

Home > RG\_Azure\_Car\_Project > Marketplace

Get Started

Service Providers

Management

Private Marketplace

Private Offer Management

My Marketplace

Favorites

My solutions

Recently created

Private plans

Categories

Analytics (58)

AI + Machine Learning (22)

Databases (16)

IT & Management Tools (12)

Compute (4)

databricks

Pricing : All

Operating System : All

Publisher Type : All

Product Type : All

Publisher name : All

New! Get AI-generated suggestions for 'databricks' View suggestions

Showing 1 to 20 of 71 results for 'databricks'. Clear search

Tile view

Azure Databricks

Lakehouse Optimizer for Azure Databricks

Access Connector for Azure Databricks

Unravel for Azure Databricks

Unravel for Azure Databricks (PayGo)

Microsoft Azure Marketplace

Search resources, services, and docs (G+)

dineshkumar7283@gmail.com

Home > RG\_Azure\_Car\_Project > Marketplace > Access Connector for Azure Databricks

Access Connector for Azure Databricks

Microsoft | Azure Service

★ 4.7 (216 ratings)

Plan

Create

Overview Plans Usage Information + Support Ratings + Reviews

Unity Catalog provides unified governance for all data and AI assets in your Lakehouse.

Unity Catalog can be configured to use an Azure managed identity to access storage containers on behalf of Unity Catalog users. Managed identities provide an identity for applications to use when they connect to resources that support Microsoft Entra ID authentication. You can use managed identities in Unity Catalog to support two primary use cases:

- As an identity to connect to the metastore's storage accounts (where managed tables are stored).
- As an identity to connect to other external storage accounts (either for file-based access or for external tables).

Configuring Unity Catalog with a managed identity has the following benefits over configuring Unity Catalog with a service principal:

- You can connect to an Azure Data Lake Storage Gen2 account that is protected by a storage firewall.
- Managed identities do not require you to maintain credentials or rotate secrets.

Media

https://portal.azure.com/#

Give feedback

The Azure Databricks Access Connector lets you connect managed identities to an Azure Databricks account for the purpose of accessing data registered in Unity Catalog.

**Project details**

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

Subscription \*  Resource group \*  Create new

**Instance details**

Name \*  Region \*

Previous Next Review + create Give feedback

Fill it

The Azure Databricks Access Connector lets you connect managed identities to an Azure Databricks account for the purpose of accessing data registered in Unity Catalog.

**Project details**

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

Subscription \*  Resource group \*  Create new

**Instance details**

Name \*  Region \*

Previous Next Review + create Give feedback

--review+create and click create

**AccessConnectorCreate-20250306112452 | Overview**

**Deployment**

**Deployment is in progress**

**Deployment name:** AccessConnectorCreate-20250306112452  
**Subscription:** Azure for Students  
**Resource group:** RG\_Azure\_Car\_Project

**Start time:** 3/6/2025, 11:26:39 AM  
**Correlation ID:** f8da0670-5214-4f27-9907-47ec849be038

**Deployment details**

Resource	Type	Status	Operation details
There are no resources to display.			

**RG\_Azure\_Car\_Project | Overview**

**Resource group**

**Subscription (move):** Azure for Students  
**Subscription ID:** d7c1e8a1-8bd2-4126-ba88-75140d084a4d  
**Tags (edit):** Add tags

**Deployments:** 5 Succeeded  
**Location:** UK South

**Resources** Recommendations (1)

Name	Type	Location	Actions
cardineshdatalakestorage	Storage account	UK South	...
carsaccessconnector	Access Connector for Azure Databricks	UK South	...
carsalesdb (carsalesserverdinesh/carsalesdb)	SQL database	UK South	...
carsaleserverdinesh	SQL server	UK South	...
carsdatabricks	Azure Databricks Service	UK South	...
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South	...

→ access connect use to storage whole contributor → it means contribute to cardineshdatalakestorage

→ it is indirect link and more data security

→ lets go to datalake

**RG\_Azure\_Car\_Project** Resource group

**Essentials**

Subscription (move) : Azure for Students  
Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d  
Tags (edit) : Add tags  
Deployments : 5 Succeeded  
Location : UK South

**Resources** Recommendations (1)

Name	Type	Location
cardineshdatalakestorage	Storage account	UK South
carsaccessconnector	Access Connector for Azure Databricks	UK South
carsalesdb (carsaleserverdinesh/carsalesdb)	SQL database	UK South
carsaleserverdinesh	SQL server	UK South
carsdatabricks	Azure Databricks Service	UK South
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South

< Previous Page 1 of 1 Next >

https://portal.azure.com/#@dineshdikumar7283@gmail.onmicrosoft.com/resource/subscriptio...

→ go to access control IAM

**cardineshdatalakestorage | Access Control (IAM)**

**Check access**

**My access**

**Check access**

**Grant access to this resource**

Grant access to resources by assigning a role. [Learn more](#)

**Add role assignment**

**View access to this resource**

View the role assignments that grant access to this and other resources. [Learn more](#)

**View**

**View deny assignments**

View the role assignments that have been denied access to specific actions at this scope. [Learn more](#)

**New! Permissions Management**

Discover, monitor and remediate unused permissions in your Azure environment with [Microsoft Entra Permission Management](#)

→ click +Add → Add role assignment

→ search storage blob data contributor → click storage blob data contributor → click next

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#view/Microsoft\\_Azure\\_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FRG...](https://portal.azure.com/#view/Microsoft_Azure_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FRG...). The page title is "Add role assignment". The "Members" tab is selected. A search bar at the top left contains "storage blob". Below it, a table lists various Azure roles:

Name	Description	Type	Category	Details
Defender CSPM Storage Data Scanner	Grants access to read blobs and files. This role is used by the data scanner of Defender CSPM.	BuiltinRole	None	View
Defender for Storage Data Scanner	Grants access to read blobs and update index tags. This role is used by the data scanner of Defender for Storage.	BuiltinRole	None	View
Storage Blob Data Contributor	Allows for read, write and delete access to Azure Storage blob containers and data	BuiltinRole	Storage	View
Storage Blob Data Owner	Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.	BuiltinRole	Storage	View
Storage Blob Data Reader	Allows for read access to Azure Storage blob containers and data	BuiltinRole	Storage	View
Storage Blob Delegator	Allows for generation of a user delegation key which can be used to sign SAS tokens	BuiltinRole	Storage	View

At the bottom of the table, it says "Showing 1 - 6 of 6 results." Navigation buttons "Review + assign", "Previous", and "Next" are at the bottom left, and a "Feedback" link is at the bottom right.

→click manage identity → now select members

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#view/Microsoft\\_Azure\\_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FRG...](https://portal.azure.com/#view/Microsoft_Azure_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2Fd7c1e8a1-8bd2-4126-ba88-75140d084a4d%2FresourceGroups%2FRG...). The page title is "Add role assignment". The "Members" tab is selected. Under "Selected role", "Storage Blob Data Contributor" is chosen. Under "Assign access to", "Managed identity" is selected. The "Members" section has a "+ Select members" button. To the right, a modal window titled "Select managed identities" is open:

Some results might be hidden due to your ABAC condition.

Subscription \*: Azure for Students

Managed identity: Select

Search by name

Selected members: No members selected. Search for and add one or more members you want to assign to the role for this resource.

Learn more about RBAC

Navigation buttons "Review + assign", "Previous", and "Next" are at the bottom left, and "Select" and "Close" buttons are at the bottom right of the modal.

→add like this

**Select managed identities**

Subscription \* Azure for Students

Managed identity Access Connector for Azure Databricks (2)

Search by name

Selected members:

Selected members: carsaccessconnector

Select Close Feedback

→click select

→review+assign

**Review + assign**

Role Storage Blob Data Contributor

Scope /subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG\_Azure\_Car\_Project/providers/Microsoft.Storage/storageAccounts/cardineshdatastorage

Members Name Object ID Type

carsaccessconnector 56ef03ab-e70f-42c7-a6f9-f653f01ce200 Access Connector for Azure Databricks

Description No description

Condition None

Review + assign Previous Next Feedback

→click one more time →add it

→one more step add container as well

The screenshot shows the Microsoft Azure Storage Container list page for the 'cardineshdatastorage' account. The left sidebar shows navigation options like Overview, Activity log, Tags, and Data storage (Containers, File shares, Queues, Tables). The main area displays a table of containers with columns: Name, Last modified, Anonymous access level, and Lease state. The table contains four rows: \$logs, brozen, gold, and silver, all listed as Private and Available.

Name	Last modified	Anonymous access level	Lease state
\$logs	3/5/2025, 10:44:16 AM	Private	Available
brozen	3/5/2025, 11:20:29 AM	Private	Available
gold	3/5/2025, 11:20:55 AM	Private	Available
silver	3/5/2025, 11:21:04 AM	Private	Available

→click + container →add name →create

The screenshot shows the 'New container' dialog box overlaid on the Azure Storage Container list page. The dialog has fields for 'Name' (set to 'unitymetastore') and 'Anonymous access level' (set to 'Private (no anonymous access)'). A note indicates that the access level is set to private because anonymous access is disabled on this storage account. At the bottom are 'Create' and 'Give feedback' buttons.

A screenshot of the Microsoft Azure portal showing the 'Containers' section of a storage account named 'cardineshdatastorage'. The left sidebar shows navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Partner solutions, Resource visualizer, and Data storage. Under Data storage, 'Containers' is selected. The main area displays a table of containers with columns: Name, Last modified, Anonymous access level, and Lease state. The containers listed are \$logs, brozen, gold, silver, and unitymetastore, all set to Private and Available.

→ go add Databricks Admin side

A screenshot of the Azure Databricks Catalog interface, specifically the 'Create metastore' page. The left sidebar shows 'Workspaces' (selected), Usage, User management, Cloud resources, Previews, and Settings. The main form has two steps: '1 Create metastore' and '2 Assign to workspaces'. Step 1 requires filling in 'Name' (carsproject) and 'Region' (uksouth). Step 2 is currently empty. Below the form, there's an optional 'ADLS Gen 2 path' field containing 'unitymetastore@cardineshdatastorage.dfs.core.windows.net' and an 'Access Connector Id' field with a placeholder URL. At the bottom are 'Advanced options', 'Create', and 'Cancel' buttons.

→ access connector id as well

→ go to azure → resource →

**RG\_Azure\_Car\_Project**

**Essentials**

Subscription (move)	Deployment	Location
Azure for Students	5 Succeeded	UK South
Subscription ID	d7c1e8a1-8bd2-4126-ba88-75140d084a4d	
Tags (edit)	Add tags	

**Resources**

Name	Type	Location	...
cardinestorage	Storage account	UK South	...
carsaccessconnector	Access Connector for Azure Databricks	UK South	...
carsalesdb	SQL database	UK South	...
carsaleserverdinesh	SQL server	UK South	...
carsdatabricks	Azure Databricks Service	UK South	...
dfs-cars-dinesh-datafactory	Data factory (V2)	UK South	...

[https://portal.azure.com/#@dineshdikkumar7283@gmail.onmicrosoft.com/resource/subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG\\_Azure\\_Car\\_Project/providers/Microsoft.DataFactory/DataFactories/dfs-cars-dinesh-datafactory](https://portal.azure.com/#@dineshdikkumar7283@gmail.onmicrosoft.com/resource/subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG_Azure_Car_Project/providers/Microsoft.DataFactory/DataFactories/dfs-cars-dinesh-datafactory)

→copy the resource id

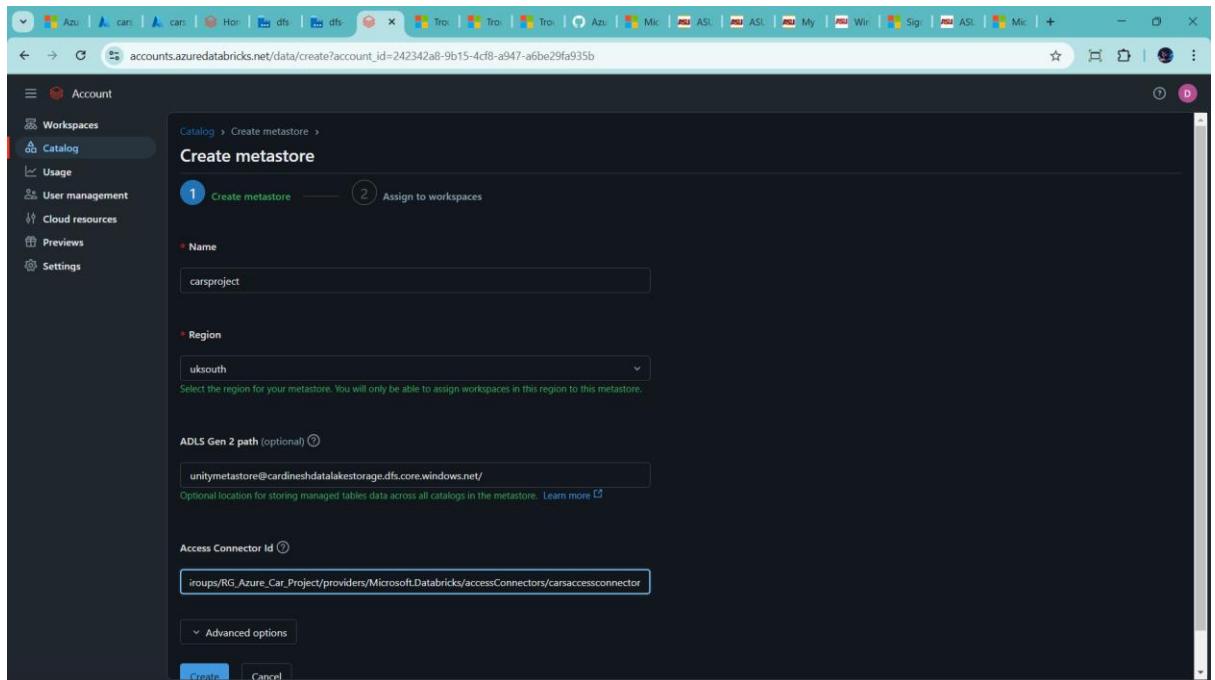
**carsaccessconnector**

**Overview**

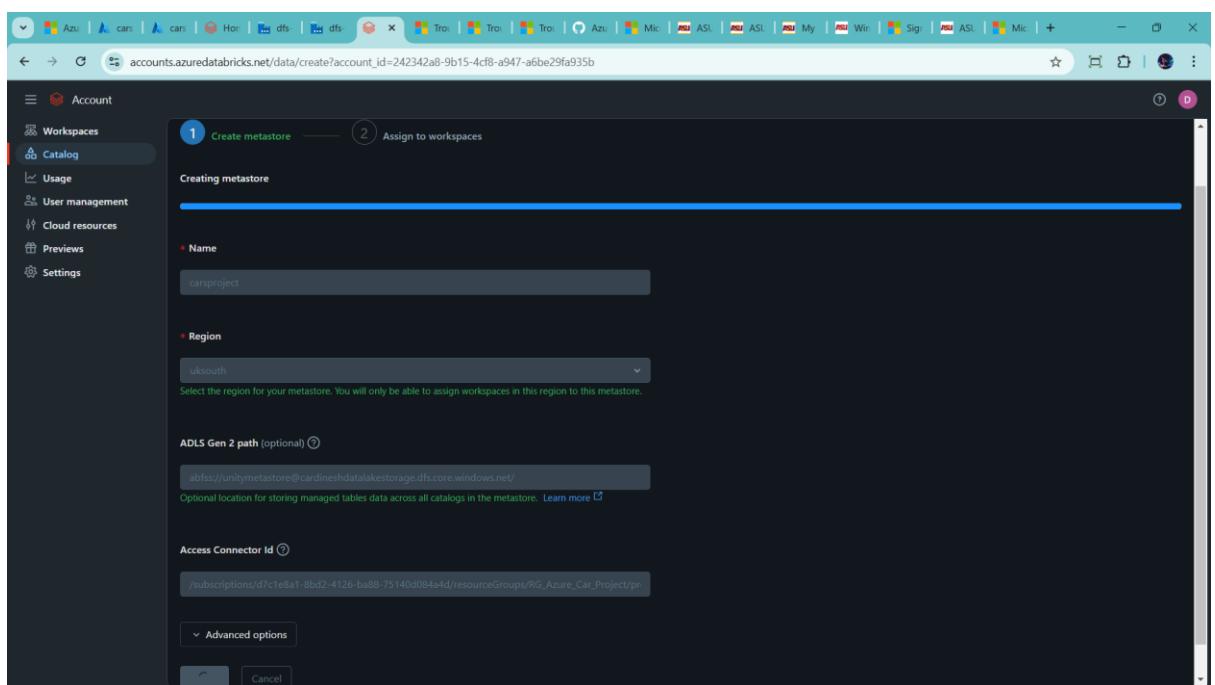
**Essentials**

Resource group (move)	State
RG_Azure_Car_Project	Succeeded
Location	UK South
Subscription (move)	Resource ID : /subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG_Az...
Subscription ID	d7c1e8a1-8bd2-4126-ba88-75140d084a4d
Tags (edit)	Add tags

→paste in databricks admin account



--Click create



→go to workspaces

The screenshot shows the 'Workspaces' section of the Azure Databricks portal. On the left, a sidebar menu includes 'Workspaces' (selected), 'Catalog', 'Usage', 'User management', 'Cloud resources', 'Previews', and 'Settings'. The main area displays a table titled 'Workspaces' with one entry:

Name	Status	Resource group	Region	Subscription	Created	Metastore
carsdatabricks	Running	RG_Azure_Car_Project	uksouth	d7c1e8a1-8bd2-4126-ba... 75140d084a4d	today at 10:15 AM	-

A 'Filter workspaces' input field and a search icon are also present.

→click name

The screenshot shows the configuration page for the 'carsdatabricks' workspace. The left sidebar remains the same. The main area has a breadcrumb path 'Workspaces > carsdatabricks >'. It features three main sections: 'Configuration' (selected), 'Permissions', and 'Security and compliance'. The 'Configuration' section contains tabs for 'Region', 'Metastore', and 'Resources'. The 'Region' tab shows 'uksouth'. The 'Metastore' tab is empty. The 'Resources' tab lists:

- Resource Group: RG\_Azure\_Car\_Project
- Subscription: d7c1e8a1-8bd2-4126-ba88-75140d084a4d

On the right, detailed information is provided:

- URL: <https://adb-172240512435715.15.azure.databricks.net>
- Workspace status: Running
- Workspace status message: Workspace is running.
- Pricing tier: Premium
- Identity federation: Enabled
- Created: today at 10:15 AM

Below the configuration tabs, there's a 'Network Policy' section with a 'default-policy' entry and an 'Update network policy' button.

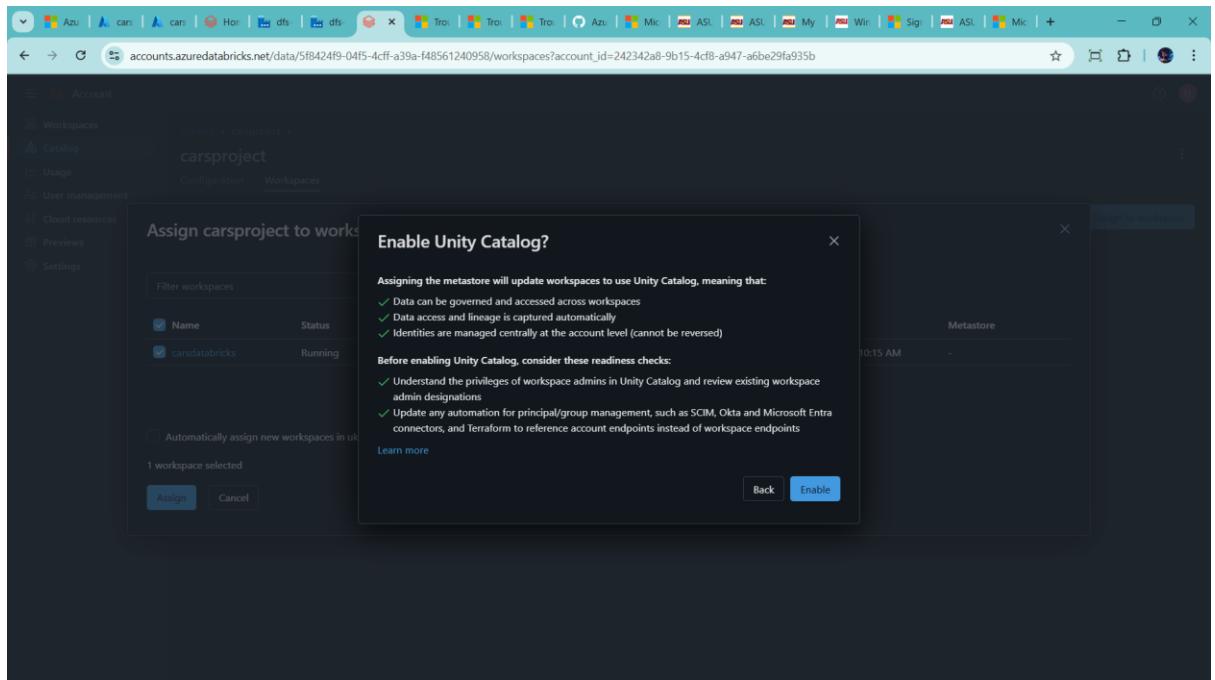
→now assign metastore

→go to catalog →click name →

The screenshot shows the Azure Databricks Catalog Metastores page. On the left, there's a sidebar with options like Workspaces, Catalog (which is selected), Usage, User management, Cloud resources, Previews, and Settings. The main area is titled 'Metastores' and contains a table with one row. The table columns are Name, Region, Path, Created at, and Updated at. The single entry is 'carsproject' in the 'carsproject' catalog, located in 'uksouth' with a path of 'abfss://unitymetastore@cardineshdatastorage.dfs.core.win...'. The 'Created at' and 'Updated at' times are 'today at 11:52 AM'. A 'Create metastore' button is visible in the top right of the table area.

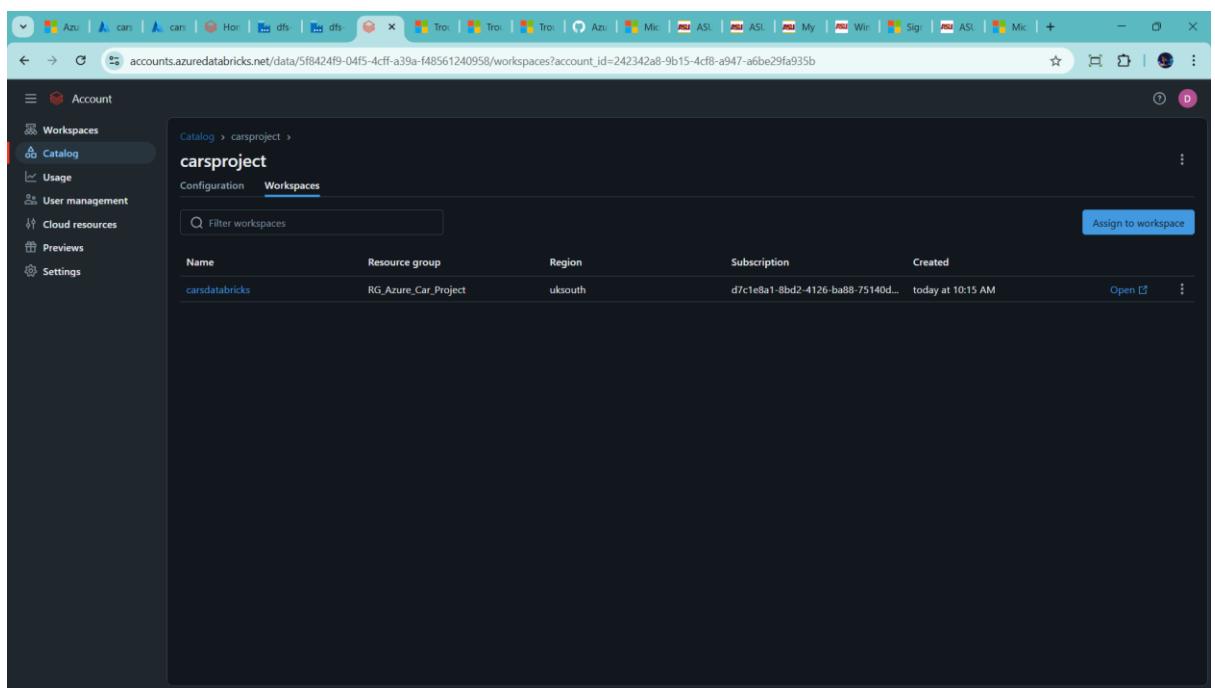
→ go to workspaces in catalog → assign workspace

The screenshot shows the 'Assign workspace' dialog box. At the top, it says 'Assign carsproject to workspaces'. Below that is a table showing one workspace: 'carsdatabricks' (Status: Running, Pricing tier: Premium, Resource group: RG\_Azure\_Car\_Project, Region: uksouth, Created: today at 10:15 AM). There's a checkbox for 'Automatically assign new workspaces in uksouth to this metastore'. At the bottom, there are 'Assign' and 'Cancel' buttons.



→unity catalog only use this workspace only in databricks

→click enable



→Now move forward azure databricks and refresh →check it is enable it

The screenshot shows the Microsoft Azure Databricks workspace. The left sidebar contains navigation links for Workspace, Recents, Catalog, Workflows, Compute, and Marketplace. Under Compute, there are sub-links for Job Runs, Data Ingestion, and Pipelines. The main workspace area shows a folder structure under 'dineshdkumar7283@gmail.com' with a message stating 'This folder is empty'. A user profile 'carsdatabricks' is visible in the top right corner.

→ now click compute → create compute

The screenshot shows the Databricks Compute interface. The left sidebar is identical to the previous workspace view. The main area is titled 'Compute' and features a tab bar with 'All-purpose compute' (selected), Job compute, SQL warehouses, Vector Search, Pools, and Policies. Below the tabs is a search bar and filter options. The central part of the screen displays a large plus sign icon and the text 'No compute'. It includes a note: 'Create compute to run workloads from your notebooks and jobs. Learn more about best practices for compute configuration'. At the bottom right of this section is a prominent 'Create compute' button.

**Dinesh km's Cluster**

**Policy**: Unrestricted

**Access mode**: Dedicated (formerly: Single user)

**Runtime**: Runtime: 15.4 LTS (Scala 2.12, Spark 3.5.0)

**Worker type**: Standard\_D4ds\_v5 (16 GB Memory, 4 Cores)

**Driver type**: Same as worker (16 GB Memory, 4 Cores)

**Tags**: Create compute Cancel

**Summary**

2-8 Workers	32-128 GB Memory
8-32 Cores	16 GB Memory, 4 Cores
1 Driver	Runtime 15.4-x-scala2.12
Unity Catalog	Photon
Standard_D4ds_v5	6-18 DBU/h

→ change it as your preference

→ first policy → change to personal computer → it is single node

→ change runtime version

→ change termite time

→ go to advance option → no change

**Dinesh km's Cluster**

**Policy**: Personal Compute

**Single user access**: Dinesh km

**Runtime**: Runtime: 15.4 LTS (Scala 2.12, Spark 3.5.0)

**Node type**: Standard\_DS3\_v2 (14 GB Memory, 4 Cores)

**Tags**: Add tags

Key	Value
> Automatically added tags	

**Summary**

1 Driver	14 GB Memory, 4 Cores
Runtime	15.4-x-scala2.12
Unity Catalog	Standard_DS3_v2
0.75 DBU/h	

Microsoft Azure | databricks

Compute > New compute > Simple form: OFF

Dinesh km's Cluster

Key Value Add

> Automatically added tags

Advanced options

Spark Logging Init Scripts

Spark config

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

Environment variables

```
MY_VAR=hello
MY_OTHER_VAR=$MY_VAR world
MY_SECRET_DB_PASSWORD=[{{secrets/prod/database_password}}]
```

Create compute Cancel

→ click create compute

→ why use create compute? → it act as a virtual machine

→ how to connect databricks and data lake container (bronze, silver, gold folder)

Ans) it is storage credential → it use access connector . to have access permission. To read the data at any container

→ lets go databricks → click catalog → click extranal data

Microsoft Azure | databricks

Catalog

Dinesh km's Cluster 14 GB, 4 Cores

Type to search... Delta Sharing Clean Rooms External Data Add data

Quick access

Recents Favorites Catalogs

Name Last viewed Type

No recent table available yet, start exploring to see your recent table here.

→ two ways create External location

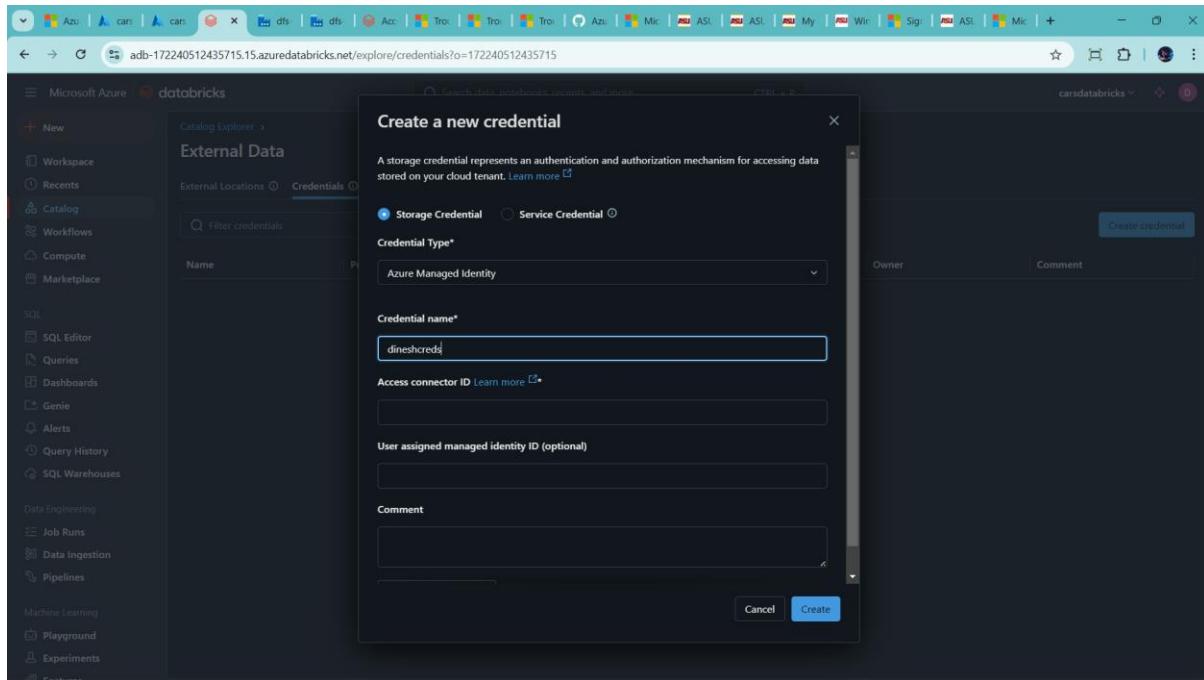
1) UI

2) Notebooks

→ prefer UI

→ click on credentials → create credentials

→ add name



→ add Access Connector ID → id present in Azure connector

→ copy the Resource id

Microsoft Azure

Home > RG\_Azure\_Car\_Project >

**carsaccessconnector** Access Connector for Azure Databricks

Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Settings

Automation

Help

Essentials

Resource group (move) : RG\_Azure\_Car\_Project

Location : UK South

Subscription (move) : Azure for Students

Subscription ID : d7c1e8a1-8bd2-4126-ba88-75140d084a4d

Tags (edit) : Add tags

State : Succeeded

Resource ID : /subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG\_Az... Copied

JSON View

→ paste in databricks create new credential → access connector id

Microsoft Azure > databricks

adb-172240512435715.15.azuredatabricks.net/explore/credentials?o=172240512435715

Create a new credential

A storage credential represents an authentication and authorization mechanism for accessing data stored on your cloud tenant. Learn more

Storage Credential Service Credential

Credential Type\*

Azure Managed Identity

Name

dineshcreds

Credential name\*

dineshcreds

Access connector ID Learn more

/subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d084a4d/resourceGroups/RG\_Azure\_Car\_Proj...

User assigned managed identity ID (optional)

Comment

Create

→ Click create

The screenshot shows the Databricks interface with the URL `adb-172240512435715.15.azuredatabricks.net/explore/credentials/dineshcreds?o=172240512435715`. The left sidebar is visible with various navigation options like Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Machine Learning, and Pipelines. The main area is titled 'Catalog Explorer > Credentials > dineshcreds'. The 'Overview' tab is selected. The credential details are as follows:

Credential Type	Managed Identity
Purpose	STORAGE
Connector Id	/subscriptions/d7c1e8a1-8bd2-4126-ba88-75140d984a4d/resourceGroups/RG_Azure.Car_Project/providers/Microsoft.Databricks/accessConnectors/carsaccessconnector
User Assigned Managed Identity Id	

Owner: Dinesh km

→ Now create same UI external credits

→ go catalog → click external data →

The screenshot shows the Databricks interface with the URL `adb-172240512435715.15.azuredatabricks.net/explore/data?o=172240512435715`. The left sidebar is visible with various navigation options. The main area is titled 'Catalog' and shows 'Dinesh km's Cluster'. The 'External Data' tab is selected. The interface includes a 'Quick access' section with 'Recents', 'Favorites', and 'Catalogs' buttons, and a search bar. Below is a table with columns 'Name', 'Last viewed', and 'Type'. A message at the bottom states: 'No recent table available yet, start exploring to see your recent table here.'

→ create external location

The screenshot shows the Databricks web interface. On the left, a sidebar menu includes options like Workspace, Recents, Catalog (which is selected), Workflows, Compute, Marketplace, SQL, and Data Engineering. The main content area is titled "Create a new external location". It explains that an external location allows access to data stored in ADLS. A form is being filled with the following details:

- External location name\***: bronzeext
- URL\***: abfss://bronze@cardineshdatalakestorage.dfs.core.windows.net
- Storage credential\***: dineshcreds (Managed Identity)

The status bar at the bottom shows the date as 06-03-2025.

--handle error →#tag account to normal account →save it

The screenshot shows the Azure Databricks account configuration page for a workspace named "carsproject". The sidebar on the left lists options: Workspaces, Catalog (selected), Usage, User management, Cloud resources, Previews, and Settings. The main content area shows the "Configuration" tab for the "carsproject" workspace. It displays the ADLS Gen 2 path: abfss://unitymetastore@cardineshdatalakestorage.dfs.core.windows.net/5f8424f9-04f5-4cff-a39a-f48561240958. Below this, there are sections for Region (set to "uksouth") and Metastore Admin (set to "Dinesh km"). Under Delta Sharing, there is an option to "Allow Delta Sharing with parties outside your organization". Under Workspace assignment, there is an option to "Automatically assign new workspaces in uksouth to this metastore".

The screenshot shows the 'Edit Metastore Admin' dialog box open over the 'Configuration' tab of a Databricks catalog. The dialog title is 'Edit Metastore Admin'. It contains a note about Metastore admins managing privileges for securable objects. A dropdown menu titled 'Select a user, group, or service principal as the metastore admin' has 'dinesh' selected. Below it, a list of results for 'dinesh' shows two entries: 'Dinesh km' (dineshdkumar7283@gmail.com#ext#@dineshdkumar7283@gmail.onmicrosoft.com) and 'Dinesh km' (dineshdkumar7283@gmail.com). A checkbox at the bottom left of the dialog is unchecked.

→now create

The screenshot shows the 'Create a new external location' form in the Databricks interface. The left sidebar is visible with various navigation options like Workspace, Catalog, Compute, Marketplace, and Data Engineering. The main form has a title 'Create a new external location'. It includes fields for 'External location name\*' (set to 'bronzeext'), 'URL\*' (set to 'abfss://bronze/cardineshdatastorage.dfs.core.windows.net/'), 'Storage credential\*' (set to 'dineshcreds (Managed Identity)'), and 'Comment'. At the bottom right are 'Cancel' and 'Create' buttons.

The screenshot shows the Databricks interface with the left sidebar expanded. Under the 'Catalog' section, a new external location named 'bronzeext' has been created. A success message at the top right indicates 'Your external location: bronzeext has been created'. The 'About this external location' panel shows the owner as 'Dinesh km'. The 'Fallback mode' toggle is turned off.

## Test the connection

The screenshot shows the 'Test connection' dialog for the 'bronzeext' location. It displays a list of successful permissions: Success - Read, Success - List, Success - Write, Success - Delete, Success - Path Exists, and Success - Hierarchical Namespace Enabled. A message at the bottom of the dialog box states 'All Permissions Confirmed' and 'The associated Storage Credential grants permission to perform all necessary operations.' A 'Done' button is visible at the bottom right of the dialog.

## Create same like silver and gold

The screenshot shows the Databricks Catalog interface. On the left sidebar, under the 'Catalog' section, there is a tree view of databases: 'My organization' (system, main, Delta Shares Received, samples, Legacy), and 'Dinesh km's Cluster' (hive\_metastore). The main area displays a table with columns 'Name', 'Last viewed', and 'Type'. A message at the bottom states: 'No recent table available yet, start exploring to see your recent table here.'

The screenshot shows the 'Create a new external location' dialog. The 'External location name\*' field contains 'silverext'. The 'URL\*' field contains 'abfss://silver@cardmeshdatalakestorage.dfs.core.windows.net'. The 'Storage credential\*' dropdown is set to 'dineshcreds (Managed identity)'. There is also a 'Comment' field and an 'Advanced Options' button. At the bottom right are 'Cancel' and 'Create' buttons.

→click create

The screenshot shows the Databricks Catalog Explorer interface. On the left is a sidebar with various navigation options like Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL, and Machine Learning. The Catalog option is currently selected. The main area is titled "External Data" and "External Locations". It displays two entries:

Name	Credential	URL	Owner	Comment
bronzeext	dineshcreds	abfss://brozen@cardinesthatalakestorage.dfs.core.windows.net/	dineshdkumar7283@gmail.com	
silverext	dineshcreds	abfss://silver@cardinesthatalakestorage.dfs.core.windows.net/	dineshdkumar7283@gmail.com	

A success message at the top right says "Successful! Your external location: silverext has been created".

→ same as gold

The screenshot shows the "Create a new external location" form in the Databricks Catalog Explorer. The sidebar on the left is identical to the previous screenshot. The main area is titled "Create a new external location". It includes fields for "External location name\*" (set to "goldext"), "URL\*○" (set to "abfss://gold@cardinesthatalakestorage.dfs.core.windows.net"), "Storage credential\* Learn more" (set to "dineshcreds (Managed identity)"), and "Comment". A "Copy from DBFS" button is also present.

The screenshot shows the Databricks interface with the URL [adb-172240512435715.15.azuredatabricks.net/explore/locations?o=172240512435715](https://adb-172240512435715.15.azuredatabricks.net/explore/locations?o=172240512435715). The left sidebar is collapsed, and the main area is titled "External Data". It displays a table of "External Locations" with three entries:

Name	Credential	URL	Owner	Comment
bronzeext	dineshcreds	abfss://brozen@cardineshdatalakestorage.dfs.core.windows.net/	dineshdkumar7283@gmail.com	
goldext	dineshcreds	abfss://gold@cardineshdatalakestorage.dfs.core.windows.net/	dineshdkumar7283@gmail.com	
silverext	dineshcreds	abfss://silver@cardineshdatalakestorage.dfs.core.windows.net/	dineshdkumar7283@gmail.com	

## Now Start Create WorkSpaces

→click workspaces →add folder name

The screenshot shows the Databricks interface with the URL [adb-172240512435715.15.azuredatabricks.net/browse/folders/15725494304152657o=172240512435715](https://adb-172240512435715.15.azuredatabricks.net/browse/folders/15725494304152657o=172240512435715). The left sidebar is collapsed, and the main area is titled "CarsProject". The workspace tree on the left shows a folder named "CarsProject" under "Workspace". The right panel shows the contents of the "CarsProject" folder, which is currently empty.

→click Carsproject → now create notebook

The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, a sidebar menu includes options like New, Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Machine Learning, and Experiments. The main workspace area is titled "CarsProject" and contains a single folder icon with the message "This folder is empty". A context menu on the right offers options such as Folder, Git folder, Query, Alert, Dashboard, Genie space, Notebook, MLflow experiment, and File.

The screenshot shows the Microsoft Azure Databricks notebook editor interface. The left sidebar is identical to the workspace view. The main area is titled "Silver\_Notebook" and is set to Python. It features a code cell with the placeholder text "Start typing or generate with AI (Ctrl + I)...". A message at the top of the code cell area states "Free trial ends in 14 days. Upgrade to Premium in Azure Portal". The interface includes standard notebook controls like Run all, Starting, Schedule, and Share, along with a command palette and keyboard shortcut information.

→ rename db\_Notebook

```
XSQL
--it is simillar with databases with match schema in hieraachey
2
XSQL
create catalog cars_catalog;
OK
This result is stored as __sqldf and can be used in other Python cells.

3
XSQL
## now create two schema use in catalog
## 1) silver
## 2) gold
Python
```

### Create Catalog

--it is simillar with databases with match schema in hieraachey

### Create Schema

→ create another notebook for reading data

```
DATA READING
2
df=spark.read.format("parquet")
.opti
.load("abfss://brozner@cardinesdatalakestorage.dfs.core.windows.net/rawdata")
(1) Spark Jobs
3
df.show()
(1) Spark Jobs
BR0004 | DLR0208 | Tat-M188 | 9664767 | 3|DTR0002| 12 | 1|2017 | AC Cars Motors | Wiesmann Motors |
BR0005 | DLR0188 | Hyu-M158 | 5525384 | 3|DTR0002| 16 | 9|2017 | AC Cars Motors | Subaru Motors |
BR0006 | DLR0168 | Ren-M128 | 12971088 | 3|DTR0003| 20 | 5|2017 | AC Cars Motors | Saab Motors |
BR0008 | DLR0128 | Hon-M68 | 7321228 | 1|DTR0004| 28 | 4|2017 | AC Cars Motors | Messerschmitt Motors |
BR0009 | DLR0108 | Cad-M38 | 11379294 | 2|DTR0004| 31 | 12|2017 | AC Cars Motors | Lexus Motors |
BR0010 | DLR0088 | Mer-M6 | 11611234 | 2|DTR0005| 4 | 9|2017 | AC Cars Motors | IFA (including Tr... |
BR0011 | DLR0062 | BMW-HO | 19979446 | 2|DTR0005| 21 | 1|2017 | Acura Motors | Acura Motors |
BR0011 | DLR0050 | Vol-M05E | 14181518 | 3|DTR0006| 9 | 5|2017 | Acura Motors | Geo Motors |
BR0012 | DLR0048 | Vol-M05E | 14181518 | 1|DTR0006| 1 | 9|2017 | Acura Motors | Acura Motors |
BR0013 | DLR0225 | Hon-A229 | 16539431 | 3|DTR0007| 11 | 5|2017 | Acura Motors | Honda Motors |
BR0014 | DLR0099 | Tat-A180 | 13389358 | 2|DTR0007| 13 | 1|2017 | Acura Motors | Zastava Motors |
BR0015 | DLR0189 | Hyu-M159 | 4891618 | 2|DTR0008| 17 | 9|2017 | Acura Motors | Sunbeam Motors |
BR0017 | DLR0149 | Lex-M99 | 5859144 | 2|DTR0008| 25 | 8|2017 | Acura Motors | Panos Motors |
BR0018 | DLR0129 | Hon-M69 | 117369466 | 2|DTR0009| 29 | 4|2017 | Acura Motors | Mia Motors |
BR0019 | DLR0109 | Cad-M39 | 2696532 | 3|DTR0010| 1 | 1|2017 | Acura Motors | Ligier Motors |
BR0020 | DLR0089 | Dod-M9 | 4816794 | 2|DTR0011| 5 | 9|2017 | Acura Motors | Infiniti Motors |
BR0021 | DLR0070 | Vol-M05T | 7738896 | 1|DTR0011| 10 | 5|2017 | Aixam-Mega | (Inclu...) | Gilbern Motors |
```

### DATA READING

df.show()

(1) Spark Jobs

Branch_ID	Dealer_ID	Model	Manufacture
BR0004	DLR0208	Tat-M188	Wiesmann Motors
BR0005	DLR0188	Hyu-M158	Subaru Motors
BR0006	DLR0168	Ren-M128	Saab Motors
BR0008	DLR0128	Hon-M68	Messerschmitt Motors
BR0009	DLR0108	Cad-M38	Lexus Motors
BR0010	DLR0088	Mer-M6	IFA (including Tr...)
BR0011	DLR0062	BMW-HO	Acura Motors
BR0011	DLR0050	Vol-M05E	Geo Motors
BR0012	DLR0048	Vol-M05E	Acura Motors
BR0013	DLR0225	Hon-A229	Honda Motors
BR0014	DLR0099	Tat-A180	Zastava Motors
BR0015	DLR0189	Hyu-M159	Sunbeam Motors
BR0017	DLR0149	Lex-M99	Panos Motors
BR0018	DLR0129	Hon-M69	Mia Motors
BR0019	DLR0109	Cad-M39	Ligier Motors
BR0020	DLR0089	Dod-M9	Infiniti Motors
BR0021	DLR0070	Vol-M05T	Gilbern Motors

Screenshot of Databricks Notebook titled "silver\_notebook" showing a Python script for data transformation.

```
from pyspark.sql.functions import *
df=withColumn("model_category",split(col("Model_ID"),"-")[0])
df.display()
```

The notebook displays a table with the following schema and data:

Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Data_ID	Day	Month
BR0004	DLR0208	Tat-M188	9664767	552504	DT00002	12	16
BR0005	DLR0188	Hyu-M158	12971088	12971088	DT00003	20	
BR0006	DLR0168	Ren-M128	732128	1	DT00004	28	
BR0008	DLR0128	Hon-M68	11379294	2	DT00004	31	
BR0009	DLR0108	Cad-M38	11611234	2	DT00005	4	
BR0010	DLR0068	Mer-M8	14181510	3	DT00006	9	
BR0011	DLR0002	BMW-M2	5358057	1	DT00006	6	
BR0012	DLR0249	BMW-M249	6733530	2	DT00007	11	
BR0013	DLR0229	Hon-M219	16150431	3	DT00007		

Screenshot of Databricks Notebook titled "silver\_notebook" showing a Python script for data transformation.

```
df.withColumn("Units_Sold",col("Units_Sold").cast(StringType())).display()
```

The notebook displays a table with the following schema and data:

Model_ID	Revenue	Units_Sold	Data_ID	Day	Month	Year	BranchName
Mer-M8	11611234	2	DT00005	4	9	2017	Acura Motors
BMW-M2	19978446	2	DT00005	2	1	2017	Acura Motors
Vol-M256	14181510	3	DT00006	9	5	2017	Acura Motors
BMW-M249	5358057	1	DT00006	6	9	2017	Acura Motors
Hon-M219	16150431	3	DT00007	11	5	2017	Acura Motors
Tat-M188	13389350	2	DT00007	13	1	2017	Acura Motors
Hyu-M158	4891618	2	DT00008	17	9	2017	Acura Motors
Lex-M99	5059144	2	DT00008	25	8	2017	Acura Motors
Hon-M68	17369466	2	DT00009	29	4	2017	Acura Motors
Cad-M38	26969532	3	DT00010	1	1	2017	Acura Motors
Dod-M9	4816794	2	DT00011	5	9	2017	Acura Motors
Vol-M257	7738896	1	DT00011	10	5	2017	Acura-Mega (inclu
Tat-M190	11038722	3	DT00012	14	1	2017	Acura-Mega (inclu
Hyu-M160	6733530	2	DT00012	18	9	2017	Acura-Mega (inclu

Screenshot of a Databricks notebook titled "silver\_notebook" running on Python 3. The notebook shows the schema of a DataFrame named "df" and displays its first two rows.

```

df.printSchema()
root
 |-- Branch_ID: string (nullable = true)
 |-- Dealer_ID: string (nullable = true)
 |-- Model_ID: string (nullable = true)
 |-- Revenue: long (nullable = true)
 |-- Units_Sold: long (nullable = true)
 |-- Date_ID: string (nullable = true)
 |-- Day: integer (nullable = true)
 |-- Month: integer (nullable = true)
 |-- Year: integer (nullable = true)
 |-- BranchName: string (nullable = true)
 |-- DealerName: string (nullable = true)
 |-- model_category: string (nullable = true)

df=df.withColumn("RevPerUnit",col("Revenue")/col("Units_Sold"))
df.display()

```

	Year	BranchName	DealerName	model_category	RevPerUnit
1	2017	AC Cars Motors	AC Cars Motors	BMW	6681989
2	2017	AC Cars Motors	Deccan Motors	Hon	5792156

Screenshot of a Databricks notebook titled "silver\_notebook" running on Python 3. The notebook displays a table with 1,849 rows and 0.83s runtime.

### AD-HOC

Units_Sold	Date_ID	Day	Month	Year	BranchName	DealerName	
11	1	DT00006	6	9	2017	Acura Motors	Acura Motors
12	3	DT00007	11	5	2017	Acura Motors	Herald Motors
13	2	DT00007	13	1	2017	Acura Motors	Zastava Motors
14	2	DT00008	17	9	2017	Acura Motors	Sunbeam Motors
15	2	DT00008	25	8	2017	Acura Motors	Panzo Motors
16	2	DT00009	29	4	2017	Acura Motors	Mia Motors
17	3	DT00010	1	1	2017	Acura Motors	Liger Motors
18	2	DT00011	5	9	2017	Acura Motors	Infiniti Motors
19	1	DT00011	10	5	2017	Axam-Mega (including Arola Mot...)	Gibson Motors
20	3	DT00012	14	1	2017	Axam-Mega (including Arola Mot...)	ZAF Motors
21	2	DT00012	18	9	2017	Axam-Mega (including Arola Mot...)	Suzuki Motors
22	3	DT00013	26	8	2017	Axam-Mega (including Arola Mot...)	Panther Motors
23	3	DT00016	30	4	2017	Axam-Mega (including Arola Mot...)	Micro Motors
24	1	DT00016	2	1	2017	Axam-Mega (including Arola Mot...)	Lincoln Motors

Screenshot of Databricks Notebook (silver\_notebook) showing a table and Python code.

**Table:**

year	BranchName	Total_Units_Sold
2017	Alvis Motors	19
2017	Tazzari Motors	9
2017	Blankship Motor Company	5
2017	DKW Motors	1
2017	Taziari Motors	15
2017	Alvis Motors	17
2017	Alvis Motors	21

**Python Code:**

```
df.groupby('year','BranchName').agg(sum('Units_Sold').alias('Total_Units_Sold')).sort('year','Total_Units_Sold',ascending=[1,0]).display()
```

Screenshot of Databricks Notebook (silver\_notebook) showing a pie chart visualization.

**Pie Chart Data:**

Year	Percentage
2019	29.34%
2017	29.99%
2018	29.99%
2020	11.91%

**Python Code:**

```
df.groupby('year','BranchName').agg(sum('Units_Sold').alias('Total_Units_Sold')).sort('year','Total_Units_Sold',ascending=[1,0]).display()
```

The screenshot shows the Microsoft Azure Databricks workspace. On the left, the sidebar includes sections for Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL (SQL Editor, Queries, Dashboards), Genie, Alerts, Query History, and SQL Warehouses. Under Data Engineering, Job Runs, Data Ingestion, and Pipelines are listed. Machine Learning sections like Playground, Experiments, Features, Models, and Serving are also present. The main area displays a pie chart visualization titled "DATA WRITING" with a refresh indicator. Below it is a code cell in Python mode containing the following code:

```
df.write.format("parquet")\n    .mode("append")\n    .option("path", "abfss://silver@cardineshdatastorage.dfs.core.windows.net/carsales")\n    .save()
```

Folder and file success added

The screenshot shows the Azure Storage Container blade for the "silver" container. The left sidebar has links for Overview, Diagnose and solve problems, Access Control (IAM), and Settings. The main area shows the "Overview" tab with an authentication method set to "Access key". It lists a single blob named "carsales" with a size of 0 B and an "Available" lease state. A table below shows the contents of the container:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
_committed_8475640370714188844	3/6/2025, 3:59:28 PM	Hot (Inferred)		Block blob	123 B	Available
_started_8475640370714188844	3/6/2025, 3:59:27 PM	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	3/6/2025, 3:59:28 PM	Hot (Inferred)		Block blob	0 B	Available
part-00000-tid-8475640370714188844-01b65b34...	3/6/2025, 3:59:28 PM	Hot (Inferred)		Block blob	58.99 KIB	Available

Microsoft Azure | databricks

silver\_notebook Python

Catalog

DATA WRITING

```
# data write in silver layer
df.write.format("parquet")
    .mode("overwrite")
    .option("path", "abfss://silver@cardineshdatalakestorage.dfs.core.windows.net/carsales")
    .save()
```

Querying Silver Data

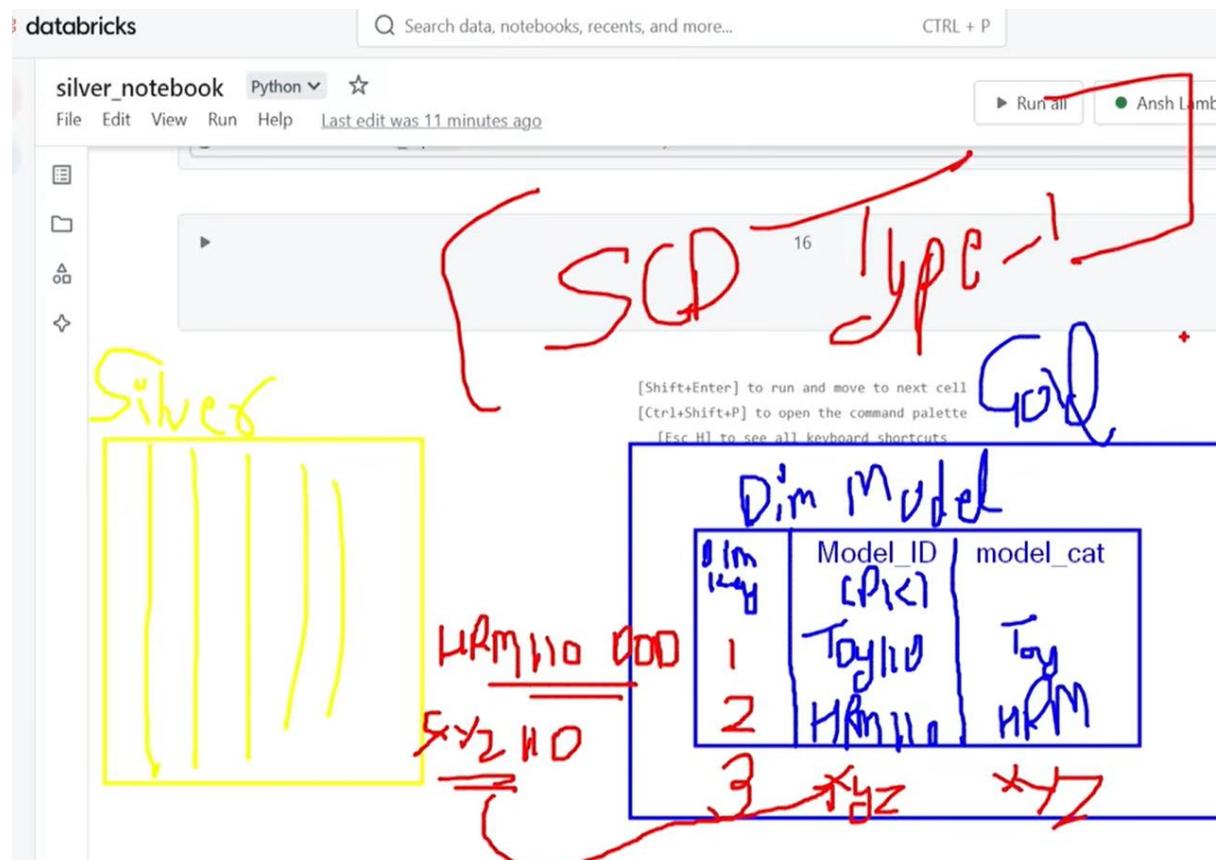
```
Just now (16)
Xsq1
SELECT * FROM parquet."abfss://silver@cardineshdatalakestorage.dfs.core.windows.net/carsales"
(2) Spark Jobs
sqldf: pyspark.sql.dataframe.DataFrame = [Branch_ID: string, Dealer_ID: string ... 11 more fields]
```

Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month
BR0001	DLR001	BMW-M1	13563978	2	DT00001	1	
BR0003	DLR0228	Hon-M218	17376468	3	DT00001	10	
BR0004	DLR0208	Tat-M188	9664767	3	DT00002	12	
BR0005	DLR0188	Hya-M158	5525304	3	DT00002	16	
BR0006	DLR0168	Ren-M128	12971088	3	DT00003	20	

→ Now Create Data Modelling

→ Star Schema → so Create New Notebook → for gold Layer

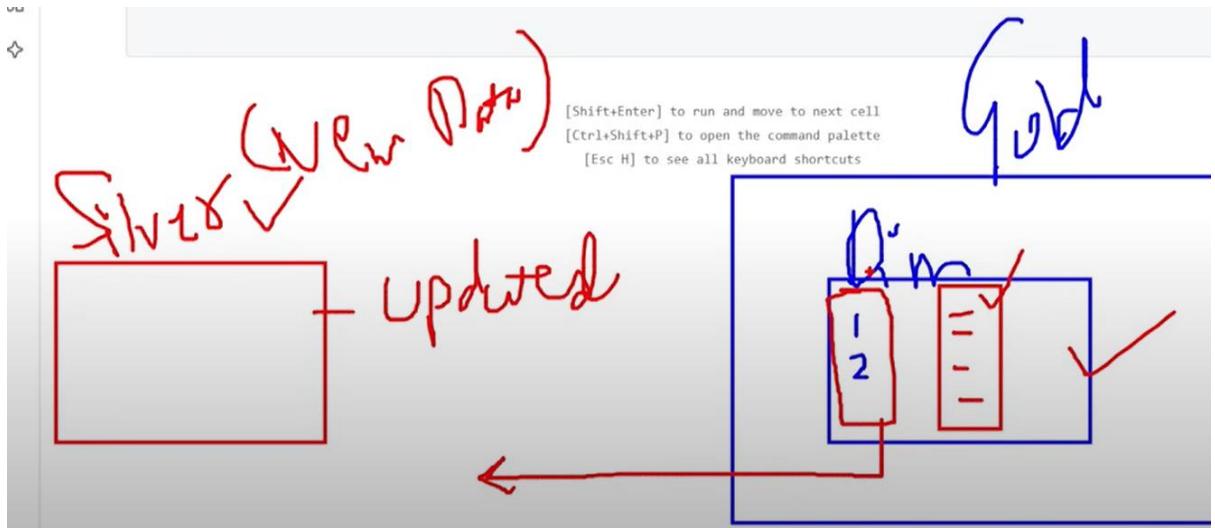
→ slow change dimension



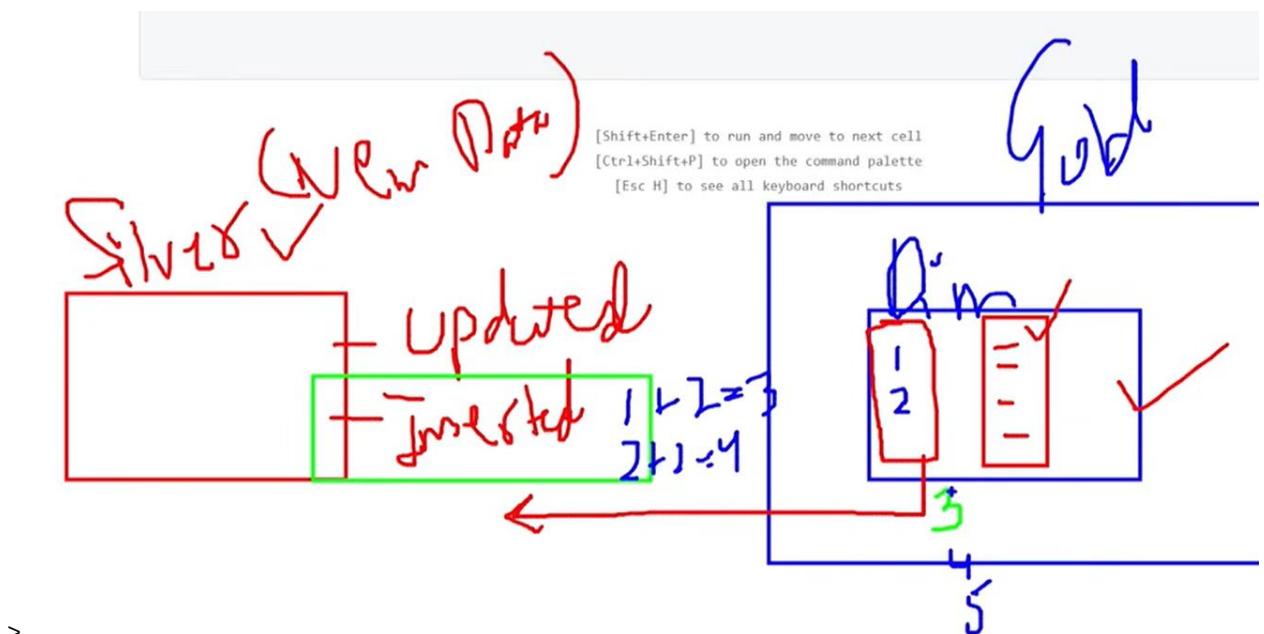
→ gold table have dim id, id, name

→ the silver table have new data → let check gold id but dim id, because dim id not present in silver table

→ Then id of gold table join with silver table. If left join both table new records in silver but not in gold table → it means dim id return null to inform insert new records present



→ dim id is surrogate key



→ first time run gold table use left join because gold layer not available table

# TYPE – 1 (UPsert)

Product_ID	Name	Prod_Cat
1	Honey	Food
2	Shirt	Clothing
3	Comb	Clothing

Product_ID	Name	Prod_Cat
1	Honey	Food
2	Shirt	Clothing
3	Comb	Hair

# Before

# AFTER

The screenshot shows the Microsoft Azure Databricks workspace. The left sidebar contains links for New, Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL Editor, Queries, Dashboards, Genie, Alerts, Query History, and SQL Warehouses. Under Data Engineering, there are Job Runs, Data Ingestion, and Pipelines. Under Machine Learning, there are Playground, Experiments, Features, Models, and Serving. The main area shows a notebook titled "gold\_notebook" in Python. The notebook has two cells: the first cell is empty with the placeholder "Start typing or generate with AI (Ctrl + I)...", and the second cell has the number "2". A status bar at the bottom indicates keyboard shortcuts: "[Shift+Enter] to run and move to next cell", "[Ctrl+Shift+P] to open the command palette", and "[Esc H] to see all keyboard shortcuts".

**gold\_notebook** Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

**CATALOG**

incremental\_flag 0

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

## CREATE FLAG PARAMETER

```

1 ✓ 3 minutes ago (46)
#Create incremental_flag parameter by default 0
dbutils.widgets.text("incremental_flag","0")

2
3 ✓ 2 minutes ago (<1s)
incremental_flag = dbutils.widgets.get('incremental_flag')
print(incremental_flag)

4 ✓ Just now (<1s)
print(type(incremental_flag))

<class 'str'>

```

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

**gold\_notebook** Python

File Edit View Run Help Last edit was 1 minute ago

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

Run all Dinesh km's Cluster Schedule Share

**CATALOG**

incremental\_flag 0

## CREATE DIMENSION Model

SQL

```

1 ✓ 1 minute ago (19s)
Sql
SELECT * FROM parquet.`abfss://silver@cardinesdatalakestorage.dfs.core.windows.net/carsales`

```

2 (2) Spark Jobs

3 \_sqldf: pyspark.sql.dataframe.DataFrame = [BranchID: string, DealerID: string ... 11 more fields]

#	Year	BranchName	DealerName	model_category	RevPerUnit
1	2017	AC Cars Motors	BMW	6681989	
2	2017	AC Cars Motors	Deccan Motors	5792156	
3	2017	AC Cars Motors	Wiesmann Motors	3221589	
4	2017	AC Cars Motors	Subaru Motors	1841768	
5	2017	AC Cars Motors	Saab Motors	4523696	
6	2017	AC Cars Motors	Messerschmitt Motors	7321228	
7	2017	AC Cars Motors	Lexus Motors	5689647	
8	2017	AC Cars Motors	IFA (including Trabant, Wartburg, Barkas) Motors	5805617	
9	2017	Acura Motors	Acura Motors	9989723	
10	2017	Acura Motors	Geo Motors	4727170	
11	2017	Acura Motors	Acura Motors	5358057	
12	2017	Acura Motors	Herald Motors	5383477	

Microsoft Azure | databricks

gold\_notebook Python

File Edit View Run Help Last edit was now

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

Dinesh km's Cluster Schedule Share

New Workspace Recents Catalog Workflows Compute Marketplace

SQL SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses

Data Engineering Job Runs Data Ingestion Pipelines

Machine Learning Playground Experiments Features Models Serving

Catalog Type to search... For you All

My organization system cars\_catalog main Delta Shares Received samples Legacy hive\_metastore

Just now (1) df\_src=spark.sql(''') SELECT DISTINCT(Model\_ID) as Model\_ID,model\_category FROM parquet.'abfss://silver@cardinesthatalakestorage.core.windows.net/carsales' ''')

(1) Spark Jobs df\_src.pyspark.sql.DataFrameDataframe = (Model\_ID:string, model\_category:string)

Just now (1) df\_src.display()

(2) Spark Jobs

Table +

Model_ID	model_category
Mar-M159	Mar
Ren-M207	Ren
Mar-M146	Mar
Mar-M141	Mar
Lin-M28	Lin
Hon-M69	Hon
Toy-M202	Toy

Microsoft Azure | databricks

gold\_notebook Python

File Edit View Run Help Last edit was now

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

Dinesh km's Cluster Schedule Share

New Workspace Recents Catalog Workflows Compute Marketplace

SQL SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses

Data Engineering Job Runs Data Ingestion Pipelines

Machine Learning Playground Experiments Features Models Serving

Catalog Type to search... For you All

My organization system cars\_catalog main Delta Shares Received samples Legacy hive\_metastore

05:52 PM (1) #add surrogate key for first load df\_sink=spark.sql(''') SELECT 1 as dim\_model\_key,Model\_ID,model\_category FROM parquet.'abfss://silver@cardinesthatalakestorage.dfs.core.windows.net/carsales' where 1=0 ''')

(1) Spark Jobs df\_sink.pyspark.sql.DataFrameDataframe = (dim\_model\_key:integer, Model\_ID:string - 1 more field)

06:00 PM (1) df\_sink.display()

Table +

dim_model_key	Model_ID	model_category

No rows returned

0 rows | 0.07s runtime Refreshed 20 minutes ago

**gold\_notebook** Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search...

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
- samples
- Legacy
- hive\_metastore

**dim\_model sink -- Initial and Incremental**

```

14
if spark.catalog.tableExists('cars_catalog.gold.dim_model'):
    df_sink=spark.sql("""
        SELECT dim_model_key,Model_ID,model_category FROM parquet:'abfss://silver@cardineshdatalakestorage.dfs.core.windows.net/carsales'
        """)
else:
    df_sink=spark.sql("""
        SELECT 1 as dim_model_key,Model_ID,model_category FROM parquet:'abfss://silver@cardineshdatalakestorage.dfs.core.windows.net/carsales'
        where 1=0
        """)

```

(1) Spark Jobs

```

15
df_sink=pyspark.sql.DataFrame = [dim_model_key:integer, Model_ID:string ... 1 more field]
df_sink.display()

```

Table +

**gold\_notebook** Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search...

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
- samples
- Legacy
- hive\_metastore

**Filtering New records and old Records**

```

17
df_filter=df_src.join(df_sink,df_src['Model_ID']==df_sink['Model_ID'],how='left').select(df_src['Model_ID'],df_src['model_category'],df_sink['dim_model_key'])

```

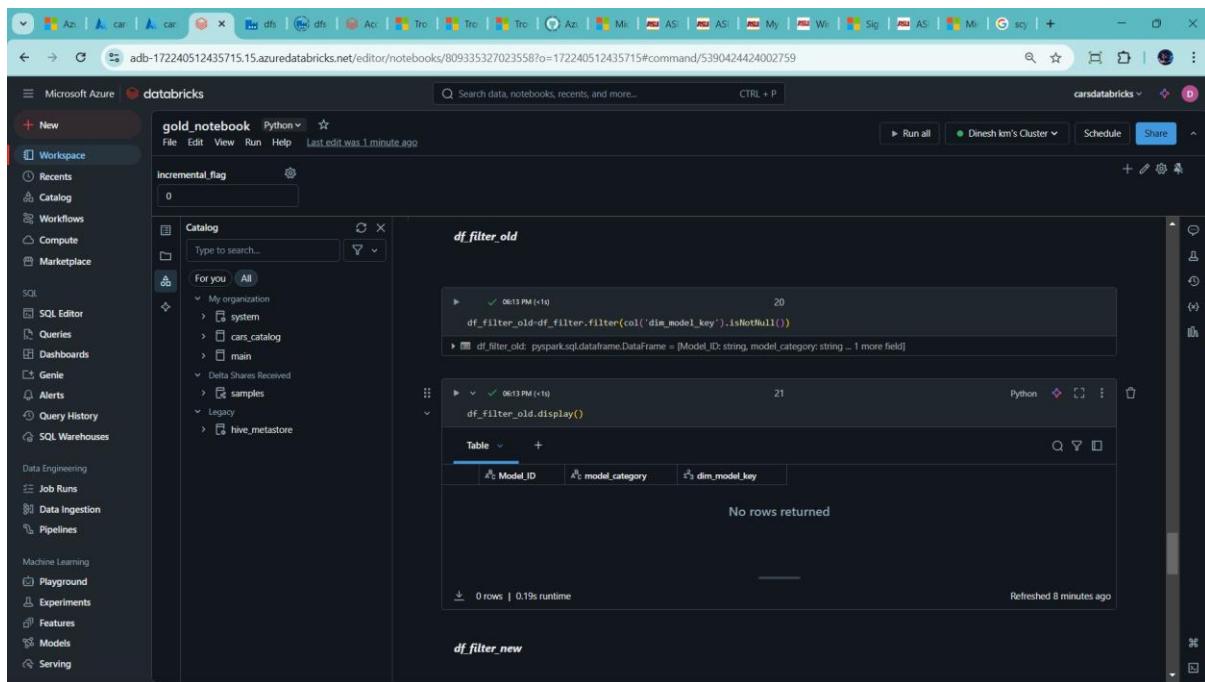
```

18
df_filter=pyspark.sql.DataFrame = [Model_ID:string, model_category:string ... 1 more field]
df_filter.display()

```

Table +

#	Model_ID	model_category	dim_model_key
24	Mar-M151	Mar	null
25	Acu-M59	Acu	null
26	Hon-M70	Hon	null
27	Mah-M174	Mah	null
28	Dod-M9	Dod	null
29	Tes-M96	Tes	null
30	Tat-M194	Tat	null
31	For-M23	For	null
32	BMW-M3	BMW	null



gold\_notebook Python

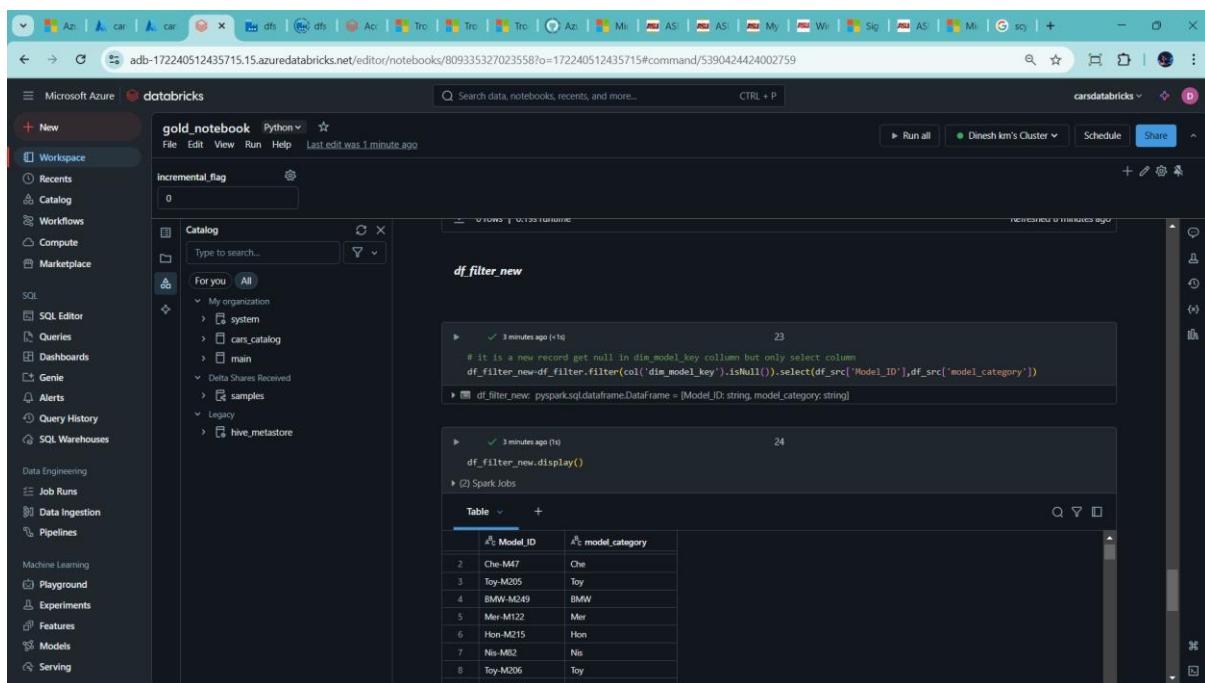
```
06/13 PM (1s) 20
df_filter_old=df_filter.filter(col('dim_model_key').isNotNull())
06/13 PM (1s) 21
df_filter_old.display()
```

No rows returned

df\_filter\_new

0 rows | 0.19s runtime

Refreshed 8 minutes ago



gold\_notebook Python

```
3 minutes ago (1s) 23
# it is a new record get null in dim model key column but only select column
df_filter_new=df_filter.filter(col('dim_model_key').isNull()).select(df_src['Model_ID'],df_src['model_category'])
3 minutes ago (1s) 24
df_filter_new.display()
```

Model_ID	model_category
Che-M47	Che
Toy-M205	Toy
BMW-M249	BMW
Mer-M122	Mer
Hon-M215	Hon
Nis-M82	Nis
Toy-M206	Toy

**gold\_notebook** Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... Create Surrogated key

Fetch the max Surrogate key from existing table

```
Just now <1s 27
if incremental_flag=='0':
    max_value=0
else:
    max_value_df=spark.sql("select max(dim_model_key) from cars_catalog.gold.dim_model")
    # convert to list
    print(max_value)
    max_value=max_value_df.collect()[0][0]
```

Create Surrogate key column and ADD the max surrogated key

```
Just now <1s 29
df_filter_new=df_filter_new.withColumn('dim_model_key',max_value+monotonically_increasing_id())
df_filter_row=pyspark.sql.DataFrame = [Model_ID: string, model_category: string ... 1 more field]
```

```
Just now (1s) 30
```

**gold\_notebook** Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... Create Surrogate key column and ADD the max surrogated key

```
Just now <1s 29
df_filter_new=df_filter_new.withColumn('dim_model_key',max_value+monotonically_increasing_id())
df_filter_row=pyspark.sql.DataFrame = [Model_ID: string, model_category: string ... 1 more field]
```

```
Just now (1s) 30
df_filter_row.display()
```

(2) Spark Jobs

Table

Model_ID	model_category	dim_model_key
Mah-M167	Mah	0
Che-M47	Che	1
Toy-M205	Toy	2
BMW-M249	BMW	3
Mer-M122	Mer	4
Hon-M215	Hon	5
Nis-M2	Nis	6
Toy-M206	Toy	7
Mar-M139	Mar	8
Ren-A207	Ren	9

--change to max\_value= 1 because surrogate key start 0

Microsoft Azure | databricks

gold\_notebook Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... All

For you All

- My organization
  - system
  - cars\_catalog
  - default
  - information\_schema
  - silver
  - main
  - Delta Shares Received
  - samples
  - legacy
  - hive\_metastore

Create Surrogated key

Fetch the max Surrogate key from existing table

```

2 minutes ago (<1s)
if incremental_flag=='0':
    max_value=[]
else:
    max_value_df=spark.sql("select max(dim_model_key) from cars_catalog.dim_model")
    # convert too list
    print(max_value)
    max_value=max_value_df.collect()[0][0]

```

Create Surrogate key column and ADD the max surrogated key

```

2 minutes ago (<1s)
df_filter_new=df_filter_new.withColumn("dim_model_key",max_value+monotonically_increasing_id())
df_filter_new.show()

```

df\_filter\_new: pyspark.sql.dataframe.DataFrame = [Model\_ID: string, model\_category: string ... 1 more field]

Microsoft Azure | databricks

gold\_notebook Python

File Edit View Run Help Last edit was now

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

Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... All

For you All

- My organization
  - system
  - cars\_catalog
  - default
  - information\_schema
  - silver
  - main
  - Delta Shares Received
  - samples
  - legacy
  - hive\_metastore

print(max\_value)
max\_value=max\_value\_df.collect()[0][0]

Create Surrogate key column and ADD the max surrogated key

```

Just now (<1s)
df_filter_new=df_filter_new.withColumn("dim_model_key",max_value+monotonically_increasing_id())
df_filter_new.show()

```

df\_filter\_new: pyspark.sql.dataframe.DataFrame = [Model\_ID: string, model\_category: string ... 1 more field]

Just now (<1s)
df\_filter\_new.display()

(2) Spark Jobs

Table +

#	Model_ID	model_category	dim_model_key
1	Mah-M167	Mah	1
2	Che-M17	Che	2
3	Toy-M205	Toy	3
4	BMW-M249	BMW	4
5	Mer-M122	Mer	5

The screenshot shows a Databricks notebook titled "gold\_notebook" running in Python. On the left, the sidebar includes sections like Workspace, Catalog, and Data Engineering. The main area displays a table of data with columns "ID", "Model", and "Category". Below the table, the code for creating a final DataFrame is shown:

```

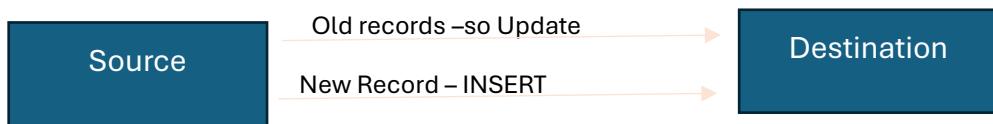
Create final DF - df_filter_old + df_filter_new

df_final = df_filter_new.union(df_filter_old)
df_final.show()

```

A preview of the resulting DataFrame is shown below the code:

ID	Model	Category
1	Mah-M167	Mah



The screenshot shows a Databricks notebook titled "gold\_notebook" running in Python. The main area displays the title "SCD Type -1 (UPSET)" and the subtitle "UPSET = INSERT + UPDATE". Below the subtitles, the code for performing an upsert operation is shown:

```

# import delta table
from delta.tables import DeltaTable

if spark.catalog.tableExists('cars_catalog.gold.dim_model'):
    delta_table = DeltaTable.forName(spark, 'abfss://gold@cardimeshdatastorage.dfs.core.windows.net/dim_model')

    # merge both table and compare automatically insert and update
    delta_table.alias('tgt').merge(df_final.alias('src'), 'tgt.dim_model_key = src.dim_model_key')\
        .whenMatchedUpdateAll()\
        .whenNotMatchedInsertAll()

else: # Initial Run
    df_final.write.format('delta')\
        .mode('overwrite')\
        .option('path', 'abfss://gold@cardimeshdatastorage.dfs.core.windows.net/dim_model')\
        .saveAsTable('cars_catalog.gold.dim_model')

```

The screenshot shows a Databricks notebook titled "gold\_notebook" in Python. The code implements an upsert operation using Delta tables:

```

UPSPERT = INSERT + UPDATE

# Import delta table
from delta.tables import DeltaTable

# Check if table exists
if spark.catalog.tableExists('cars_catalog.gold.dim_model'):
    # Merge both table and compare automatically insert and update
    delta_table = DeltaTable.forName(spark, 'abfss://gold@cardineshdatalakestorage.dfs.core.windows.net/dim_model')
    .whenMatchedUpdateAll()
    .whenNotMatchedInsertAll()
    .execute()

else: # Initial RUN
    df_final.write.format('delta')\
        .mode('overwrite')\
        .option('path', 'abfss://gold@cardineshdatalakestorage.dfs.core.windows.net/dim_model')\
        .saveAsTable('cars_catalog.gold.dim_model')

```

The notebook also shows a sidebar with a catalog view and a sidebar with recent items.

--change name gold\_dim\_model

The screenshot shows a Databricks notebook titled "gold\_dim\_model" in Python. The code creates a table named "gold.dim\_model" and displays its contents:

```

XSQL
select * from cars_catalog.gold.dim_model

```

The resulting table "gold.dim\_model" contains the following data:

Model_ID	model_category	dim_model_key
Mah-M167	Mah	1
Che-M47	Che	2
Toy-M205	Toy	3
BMW-M249	BMW	4
Mer-M122	Mer	5
Hon-M215	Hon	6
Nis-M82	Nis	7
Toy-M206	Toy	8
Mar-M139	Mar	9
Ren-M207	Ren	10
Mar-M146	Mar	11
Mar.M141	Mar	12

-→ copy this gold\_dim\_model to gold\_dim\_branch in clone method

Screenshot of the Databricks workspace showing a context menu for a notebook named "gold\_dim\_model". The menu includes options like Open in new tab, Clone, Download as, Copy URL/path, Rename, Share (Permissions), Move, Add to favorites, and Move to Trash.

Name	Type	Owner	Created at
db_Notebook	Notebook	Dinesh km	Mar 06, 2025, 01:03 PM
gold_dim_model	Notebook	Dinesh km	Mar 06, 2025, 04:58 PM
silver_notebook	Notebook	Dinesh km	Mar 06, 2025, 01:19 PM



Screenshot of the Databricks workspace showing the same context menu for the "gold\_dim\_model" notebook, but with a different set of items visible, likely due to a UI refresh or selection change.

Name	Type	Owner	Created at
db_Notebook	Notebook	Dinesh km	Mar 06, 2025, 01:03 PM
gold_dim_branch	Notebook	Dinesh km	Mar 07, 2025, 11:27 AM
gold_dim_model	Notebook	Dinesh km	Mar 06, 2025, 04:58 PM
silver_notebook	Notebook	Dinesh km	Mar 06, 2025, 01:19 PM

→change accouding

**CREATE DIMENSION Model**

```

df_src=spark.sql("""
SELECT * FROM parquet.`abfss://silver@cardinleshdatalestage.dfs.core.windows.net/carsales`""")
df_src.display()

```

#	Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month
1	BR0001	DLR0001	BMW-M1	1363978	2	DT00001	1	
2	BR0003	DLR0208	Hon-M218	17376468	3	DT00001	10	
3	BR0004	DLR0206	Tat-M188	9664767	3	DT00002	12	
4	BR0005	DLR0168	Hyu-M158	5525304	3	DT00002	16	
5	BR0006	DLR0168	Ren-M128	12971088	3	DT00003	20	
6	BR0008	DLR0128	Hon-M68	7321238	1	DT00004	28	
7	BR0009	DLR0108	Cad-M38	11379294	2	DT00004	31	
8	BR0010	DLR0068	Mer-M8	11611234	2	DT00005	4	
9	BR0011	DLR0002	BMW-M2	19979446	2	DT00005	2	
10	BR0011	DLR0069	Vol-M256	14181510	3	DT00006	9	

This result is stored as `_sql1df` and can be used in other Python and SQL cells.

```

df_src=spark.sql("""
SELECT DISTINCT(Branch_ID) as Branch_ID,BranchName FROM parquet.`abfss://silver@cardinleshdatalestage.dfs.core.windows.net/carsales`""")
df_src.display()

```

#	Branch_ID	BranchName
1	BR0131	Audi Motors
2	BR0760	Healey Motors
3	BR0769	Hillman Motors
4	BR0938	Isotta Fraschini Motors
5	BR1040	Lada Motors

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

Search data, notebooks, recents, and more... CTRL + P model 1/43 Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... All

For you All

- My organization
  - system
  - cars\_catalog
  - main
  - DeltaSharesReceived
  - samples
  - Legacy
  - hive\_metastore

**dim\_branch sink -- Initial and Incremental**

```

Just now (10) 13
if spark.catalog.tableExists('cars_catalog.gold.dim_branch'):
    df_sink=spark.sql('''
        SELECT dim_branch_key,Branch_ID,BranchName FROM parquet.`abfss://silver@cardinesthdatastorage.dfs.core.windows.net/carsales` ''')
else:
    df_sink=spark.sql('''
        SELECT 1 as dim_branch_key,Branch_ID,BranchName FROM parquet.`abfss://silver@cardinesthdatastorage.dfs.core.windows.net/carsales` where 1=0 ''')

(1) Spark Jobs
df_sink.pyspark.sql.DataFrame = {dim_branch_key:integer,Branch_ID:string ... 1 more field}

```

```

Just now (<1s) 14
df_sink.display()

Table + dim_branch_key Branch_ID BranchName

```

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

Search data, notebooks, recents, and more... CTRL + P model 1/43 Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search... All

For you All

- My organization
  - system
  - cars\_catalog
  - main
  - DeltaSharesReceived
  - samples
  - Legacy
  - hive\_metastore

**Filtering New records and old Records**

```

Just now (<1s) 16
df_filter=df_src.join(df_sink,df_src['Branch_ID']==df_sink['Branch_ID'],how='left').select(df_src['Branch_ID'],df_src['BranchName'],df_sink['dim_branch_key'])

(2) Spark Jobs
df_filter.pyspark.sql.DataFrame = [Branch_ID:string,BranchName:string ... 1 more field]

Just now (1s) 17
df_filter.display()

Table + Branch_ID BranchName dim_branch_key

```

Branch_ID	BranchName	dim_branch_key
BR0131	Audi Motors	null
BR0760	Healey Motors	null
BR0789	Hillman Motors	null
BR0938	Isotta Fraschini Motors	null
BR1040	Lada Motors	null
BR1693	Saleen Motors	null

<adb-172240512435715.15.azuredatabricks.net/editor/notebooks/397306497209653?o=172240512435715#command/3973064972096556>

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

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

model 1/43 Run all Dinesh km's Cluster Schedule Share

New Recent Catalog Workflows Compute Marketplace

SQL SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses

Data Engineering Job Runs Data Ingestion Pipelines

Machine Learning Playground Experiments Features Models Serving

Catalog Type to search... Refreshed 1 minute ago

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

df\_filter\_old

```
Just now (1s) 19
df_filter_old=df_filter.filter(col('dim_branch_key').isNotNull())
df_filter_old: pyspark.sql.DataFrameDataframe = [Branch_ID: string, BranchName: string ... 1 more field]
```

df\_filter\_old.display()

Table Branch\_ID BranchName dim\_branch\_key

No rows returned

0 rows | 0.08s runtime Refreshed now

This screenshot shows a Databricks notebook titled "gold\_dim\_branch" in Python. The sidebar contains various workspace sections like SQL, Data Engineering, and Machine Learning. The main area has a catalog sidebar. A code cell titled "df\_filter\_old" is running, showing a DataFrame named "df\_filter\_old" with columns Branch\_ID, BranchName, and dim\_branch\_key. The output shows "No rows returned". The status bar at the bottom indicates 0 rows and 0.08s runtime.

<adb-172240512435715.15.azuredatabricks.net/editor/notebooks/397306497209653?o=172240512435715#command/3973064972096556>

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

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

model 1/43 Run all Dinesh km's Cluster Schedule Share

New Recent Catalog Workflows Compute Marketplace

SQL SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses

Data Engineering Job Runs Data Ingestion Pipelines

Machine Learning Playground Experiments Features Models Serving

Catalog Type to search... Refreshed now

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

df\_filter\_new

```
Just now (1s) 22
# it is a new record get null in dim model key column but only select column
df_filter_new=df_filter.filter(col('dim_branch_key').isNull()).select(df_src['Branch_ID'],df_src['BranchName'])

df_filter_new: pyspark.sql.DataFrameDataframe = [Branch_ID: string, BranchName: string]
```

df\_filter\_new.display()

(2) Spark Jobs

Table Branch\_ID BranchName

Branch_ID	BranchName
BR0131	Audi Motors
BR0760	Healey Motors
BR0769	Hillman Motors
BR0938	Isotta Fraschini Motors
BR1040	Lada Motors
BR1693	Saleen Motors
BR1792	Simca do Brasil Motors
BR1799	Simca do Brasil Motors

This screenshot shows the same Databricks notebook continuing the process. The code cell "df\_filter\_new" is running, showing a DataFrame named "df\_filter\_new" with columns Branch\_ID and BranchName. The output displays 8 rows of data from the source. The status bar at the bottom indicates 8 rows and 0.08s runtime.

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

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

model 1/43 Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search...

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

Create Surrogated key

Fetch the max Surrogate key from existing table

```
Just now (<1s) 26
if incremental_flag==0:
    max_value=1
else:
    max_value_df=spark.sql("select max(dim_branch_key) from cars_catalog.gold.dim_branch")
    # convert to list
    print(max_value)
    max_value=max_value_df.collect()[0][0]
```

Create Surrogate key column and ADD the max surrogated key

```
10:49 AM (<1s) 28
df_filter_new=df_filter_new.withColumn('dim_branch_key',max_value+monotonically_increasing_id())
df_filter_new.show()
```

Python

10:49 AM (<1s) 28
df\_filter\_new=pyspark.sql.dataframe.DataFrame = [BranchID: string, BranchName: string ... 1 more field]

**gold\_dim\_branch** Python

File Edit View Run Help Last edit was now

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

model 1/43 Run all Dinesh km's Cluster Schedule Share

incremental\_flag 0

Catalog Type to search...

For you All

- My organization
  - system
  - cars\_catalog
  - main
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

Create Surrogate key column and ADD the max surrogated key

```
1 minute ago (<10) 26
df_filter_new=df_filter_new.withColumn('dim_branch_key',max_value+monotonically_increasing_id())
df_filter_new.show()
```

1 minute ago (<10) 28
df\_filter\_new=df\_filter\_new.withColumn('dim\_branch\_key',max\_value+monotonically\_increasing\_id())
df\_filter\_new.show()

Just now (<1s) 29
df\_filter\_new.display()

Spark Jobs

Table

Branch_ID	BranchName	dim_branch_key
BR0131	Audi Motors	1
BR0760	Healey Motors	2
BR0769	Hillman Motors	3
BR0938	Isotta Fraschini Motors	4
BR1040	Lada Motors	5
BR1693	Saab Motors	6
BR1792	Simsa do Brasil Motors	7

The screenshot shows a Databricks notebook titled "gold\_dim\_branch" in Python. The notebook contains the following code:

```

incremental_flag
0

Create final DF - df_filter_old + df_filter_new

Just now (1s)
df_final = df_filter_new.union(df_filter_old)

df_final: pyspark.sql.DataFrame = [BranchID: string, BranchName: string ... 1 more field]

Just now (1s)
df_final.display()

(2) Spark Jobs

```

Below the code, there is a table preview showing 10 rows of data:

BranchID	BranchName	dim_branch_key
BR0131	Audi Motors	1
BR0760	Healey Motors	2
BR0789	Hillman Motors	3
BR0938	Isotta Fraschini Motors	4
BR1040	Lada Motors	5
BR1693	Saun Motors	6
BR1792	Simca do Brasil Motors	7
BR1799	Simca do Brasil Motors	8
BR1955	Toyota Motors	9
BR1978	Turner Motors	10

The screenshot shows a Databricks notebook titled "gold\_dim\_branch" in Python. The notebook contains the following code:

## SCD Type -1 (UPsert)

**UPSERT = INSERT + UPDATE**

```

incremental_flag
0

Just now (1s)
# Import delta table
from delta.tables import DeltaTable

Just now (6s)
if spark.catalog.tableExists('cars_catalog.gold.dim_branch'):
    delta_table = DeltaTable.forPath(spark, 'abfss://gold@cardinehdatalekestorage.dfs.core.windows.net/dim_branch')

    # merge both table and compare automatically insert and update
    delta_table.alias('trg').merge(df_final.alias('src'), 'trg.dim_branch_key = src.dim_branch_key')\
        .whenMatchedUpdateAll()\
        .whenNotMatchedInsertAll()\
        .execute()

    else: # Initial RUN
        df_final.write.format('delta')\
            .mode('overwrite')\
            .option('path', 'abfss://gold@cardinehdatalekestorage.dfs.core.windows.net/dim_branch')\
            .saveAsTable('cars_catalog.gold.dim_branch')

```

-→ Now Create Another Dimension table – dim\_dealer

→ copy the clone the another gold\_dim\_branch

The screenshot shows the Azure Databricks workspace interface. On the left, there's a sidebar with various navigation options like Workspace, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Machine Learning, and more. The main area displays a list of notebooks under the 'CarsProject' workspace. A modal dialog is open titled 'Clone "gold\_dim\_branch"', prompting the user to enter a new name ('gold\_dim\_dealer') and select a workspace ('CarsProject').

The screenshot shows the Azure Databricks notebook editor for the 'gold\_dim\_dealer' notebook. The notebook is in Python mode. A code cell contains the following SQL query:
`df_src=spark.sql("""SELECT DISTINCT(Dealer_ID) as Dealer_ID,DealerName FROM parquet.`abfss://silver@cardinehdata.azuredatalakestorage.dfs.core.windows.net/carsales`""")
df_src.display()`

The results of the query are displayed in a table:

Dealer_ID	DealerName
DLR005	Fiat do Brasil Motors
DLR0107	Land Rover Motors
DLR0129	Mia Motors

--Change dim\_date as well create another clone file for dim\_dealer

The screenshot shows the Azure Databricks workspace interface. On the left, there's a sidebar with various navigation options like Workspace, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Machine Learning, and more. The main area is titled 'CarsProject' and lists several notebooks. A modal dialog is open in the center, titled 'Clone "gold\_dim\_dealer"', with a text input field containing 'gold\_dim\_date'. Below it, there's a 'Clone to' dropdown set to 'Workspace/CarsProject' and a 'Browse' button. At the bottom of the dialog are 'Cancel' and 'Clone' buttons.

--use only one column date id

The screenshot shows the Azure Databricks notebook editor. The left sidebar shows the workspace navigation. The main area has a tab for 'gold\_dim\_date' in Python. The notebook contains two cells. The top cell has code: 

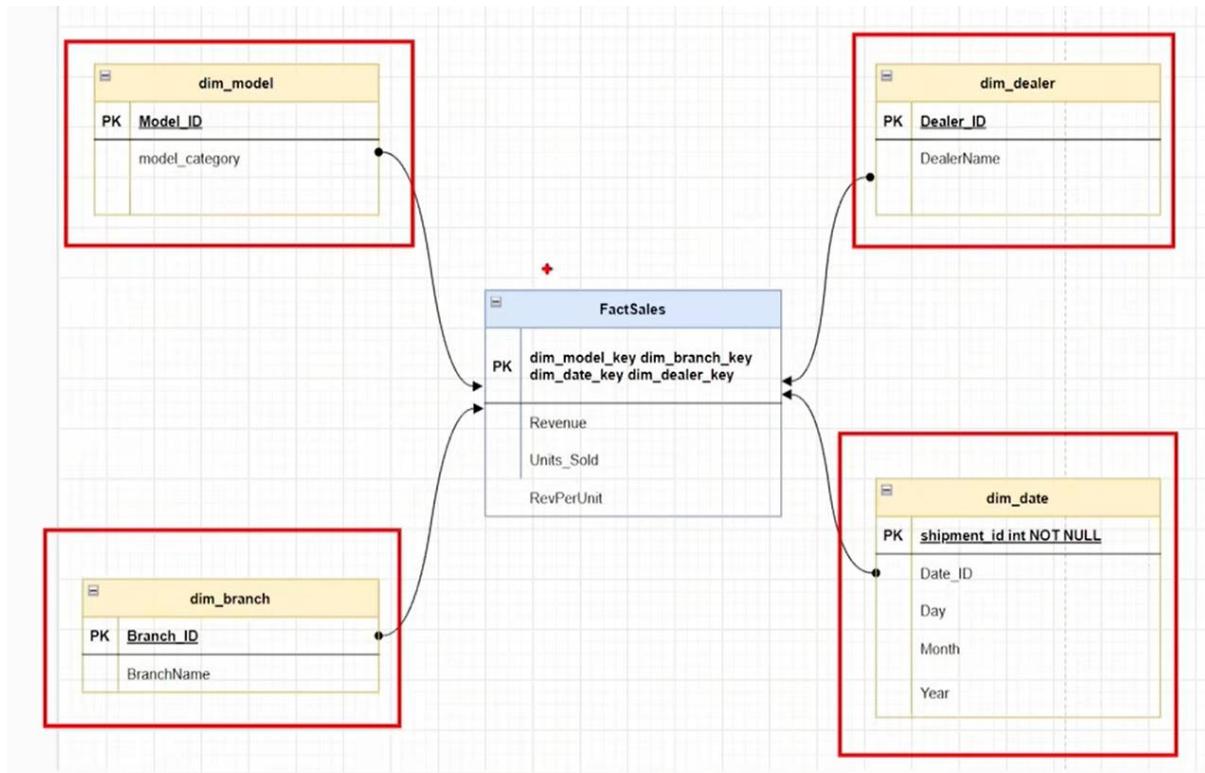
```
df_src=spark.sql("SELECT DISTINCT(Date_ID) as Date_ID FROM parquet.`abfss://silver@cardineshdatastorage.dfs.core.windows.net/carsales`")
```

. The bottom cell has code: 

```
df_src.pyspark.sql.DataFrame.show()
```

. The results of the bottom cell show a table with columns 'Dealer\_ID' and 'DealerName', containing 10 rows of data.

Dealer_ID	DealerName
DLR0058	Fiat do Brasil Motors
DLR0107	Land Rover Motors
DLR0129	Mia Motors
DLR0111	Lotus Motors
DLR0085	Humber Motors
DLR0001	AC Cars Motors
DLR0218	Lagonda Motors
DLR0082	Honda Motors
DLR0063	Ford do Brasil Motors
DLR0193	Tazzari Motors



--Create all dim table as above → now create Fact table

→ create new notebook → name as gold\_fact\_sales

The screenshot shows the Databricks SQL Editor interface with the following details:

- Left Sidebar:** Shows the workspace navigation bar and the current notebook titled "gold\_fact\_sales" in Python.
- Top Bar:** Includes the Azure logo, a search bar, and various navigation links.
- Middle Area:** A catalog sidebar on the left lists "For you" and "All" sections, including "My organization", "Delta Shares Received", and "Legacy".
- Right Area:**
  - Section Header:** CREATE A FACT TABLE
  - Reading Silver Data:** A section showing the code to read data from the silver layer.
  - Code Editor:** Two code cells are visible:
    - Cell 3: `df\_silver = spark.sql("select \* from parquet.`abfss://silver@cardinesthdatalakestorage.dfs.core.windows.net/carsales`")`
    - Cell 4: Placeholder text: "Start typing or generate with AI (Ctrl + I)..."
  - Bottom Status Bar:** Includes keyboard shortcuts for running cells and opening the command palette.

→ retrvie the source data/silver

```

gold_fact_sales Python
File Edit View Run Help Last edit was now
Catalog
Type to search...
15 | 1,849 rows | 0.41s runtime
Refreshed 3 minutes ago
Reading All The Dimensions

Just now (1s)
df_dealer=spark.sql("select * from cars_catalog.gold.dim_dealer")
df_branch=spark.sql("select * from cars_catalog.gold.dim_branch")
df_model=spark.sql("select * from cars_catalog.gold.dim_model")
df_date=spark.sql("select * from cars_catalog.gold.dim_date")

df_branch: pyspark.sql.DataFrame = [Branch_ID: string, BranchName: string ... 1 more field]
df_date: pyspark.sql.DataFrame = [Date_ID: string, dim_date_key: long]
df_dealer: pyspark.sql.DataFrame = [Dealer_ID: string, DealerName: string ... 1 more field]
df_model: pyspark.sql.DataFrame = [Model_ID: string, model_category: string ... 1 more field]

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

```

```

gold_fact_sales Python
File Edit View Run Help Last edit was now
Brining Keys to the FACT Tables

Just now (<1s)
df_fact=df_silver.join(df_branch,df_silver['Branch_ID']==df_branch['Branch_ID'],'left')
.join(df_dealer,df_silver['Dealer_ID']==df_dealer['Dealer_ID'],'left')
.join(df_model,df_silver['Model_ID']==df_model['Model_ID'],'left')
.join(df_date,df_silver['Date_ID']==df_date['Date_ID'],'left')
.select(df_silver['Revenue'],df_silver['Units_Sold'],df_silver['RevPerUnit'],df_branch['dim_branch_key'],df_dealer['dim_dealer_key'],df_model['dim_model_key'],df_date['dim_date_key'])

df_fact: pyspark.sql.DataFrame = [Revenue: long, Units_Sold: long ... 5 more fields]

df_fact.display()
(5) Spark Jobs
Table +
```

	Revenue	Units_Sold	RevPerUnit	dim_branch_key	dim_dealer_key	dim_model_key	dim_date_key
1	13363978	2	6681989	418	6	155	825
2	17376468	3	579156	1557	197	252	825
3	9664767	3	3221589	1058	104	199	752
4	5525304	3	1841768	789	95	183	752
5	12971088	3	4323696	497	231	106	882
6	7321228	1	7321228	1804	41	41	988
7	11379294	2	5689647	734	177	107	988
8	11611234	2	5805617	1211	182	110	1043

**Write the Fact Table**

```

11
from delta.tables import DeltaTable

12
if spark.catalog.tableExists('factsales'):
    deltaTable = DeltaTable.forName(spark, 'cars_catalog.gold.factsales')
    deltaTable.alias('trg').merge(deltaTable.alias('src'), trg.dim_branch_key==src.dim_branch_key and trg.dim_dealer_key==src.dim_dealer_key and trg.dim_model_key==src.dim_model_key)
    .whenMatchedUpdateAll()
    .whenNotMatchedInsertAll()
    .execute()

else:
    df_fact.write.format("delta")
        .mode("overwrite")
        .option("path", "abfss://gold@cardineshdatalestage.dfs.core.windows.net/factsales")
        .saveAsTable("cars_catalog.gold.factsales")

13
Xsql
select * from cars_catalog.gold.factsales

```

#	Revenue	Units_Sold	RevPerUnit	dim_branch_key	dim_dealer_key	dim_model_key	dim_date_key
14	4891618	2	244809	1290	198	192	648
15	5059144	2	252957	365	254	45	648
16	17369466	2	8664733	419	3	14	187
17	26969532	3	8988044	27	166	138	1012
18	4816794	2	2406397	954	71	28	729
19	7738896	1	7738896	751	142	248	729
20	11038722	3	3679574	1809	116	106	251
21	6733530	2	3366765	1103	189	214	251
22	6594996	3	2198332	150	27	230	169
23	8355774	3	2785259	1792	115	26	538
24	6158643	1	6158643	879	36	265	538
25	18118773	3	6039591	542	216	70	527
26	49997908	5	9999581	643	173	376	477

The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, there's a sidebar with various navigation options like Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Machine Learning, and more. The main area is titled 'CarsProject' and contains a list of notebooks. The table has columns for Name, Type, Owner, and Created at. The notebooks listed are:

Name	Type	Owner	Created at
db_Notebook	Notebook	Dinesh km	Mar 06, 2025, 01:03 PM
gold_dim_branch	Notebook	Dinesh km	Mar 07, 2025, 11:27 AM
gold_dim_date	Notebook	Dinesh km	Mar 07, 2025, 11:59 AM
gold_dim_dealer	Notebook	Dinesh km	Mar 07, 2025, 11:48 AM
gold_dim_model	Notebook	Dinesh km	Mar 06, 2025, 04:58 PM
gold_fact_sales	Notebook	Dinesh km	Mar 07, 2025, 12:12 PM
silver_notebook	Notebook	Dinesh km	Mar 06, 2025, 01:19 PM

## →Now Create WorkFlow in DataBricks

The screenshot shows the Databricks workflows page. The sidebar on the left is identical to the previous workspace screenshot. The main area is titled 'Workflows' and shows a table with columns for Name, Tags, Created by, Trigger, and Recent runs. A message at the bottom says 'No jobs found.'

→ Create job

**Job details**

- Job ID: 1055093289869141
- Creator: Dinesh km
- Run as: Dinesh km
- Tags: Add tag
- Description: Add description
- Lineage: No lineage information for this job. Learn more

**Schedules & Triggers**

- None
- Add trigger

**Job parameters**

- No job parameters are defined for this job
- Edit parameters

**Job notifications**

- No notifications
- Edit notifications

→ name as Data-model

- Task name – Silver\_data
- Click path – choose silver\_notebook

**Select Notebook**

Workspace Recents

- Repos
- Shared
- Users
- CarsProject

- db\_Notebook
- gold\_dim\_branch
- gold\_dim\_date
- gold\_dim\_dealer
- gold\_dim\_model
- gold\_fact\_sales
- silver\_notebook**

**Job details**

- 1055093289869141
- Dinesh km
- Dinesh km
- Add tag
- Add description
- No lineage information for this job. Learn more

**Schedules & Triggers**

- None
- Add trigger

**Job parameters**

- No job parameters are defined for this job
- Edit parameters

**Job notifications**

- No notifications
- Edit notifications

→ chose cluse as well

→ click create task

Microsoft Azure | databricks

Workflows > Jobs > Data-Model

Runs Tasks

Silver\_Data

Task name\*: Silver\_Data

Type\*: Notebook

Source\*: Workspace

Path\*: /Workspace/CarsProject/silver\_notebook

Cluster\*: Dinesh km's Cluster

Dependent libraries: + Add

Parameters: UI JSON

Job details

Job ID: 1055093289869141

Creator: Dinesh km

Run as: Dinesh km

Tags: + Add tag

Description: + Add description

Lineage: No lineage information for this job. Learn more

Schedules & Triggers: None

Add trigger

Job parameters: No job parameters are defined for this job

Edit parameters

Job notifications: No notifications

Edit notifications

Duration and streaming backlog thresholds

--click add task dim model

Microsoft Azure | databricks

Workflows > Jobs > Data-Model

Runs Tasks

Silver\_Data

+ Add task

Task name\*: Silver\_Data

Type\*: Notebook

Source\*: Workspace

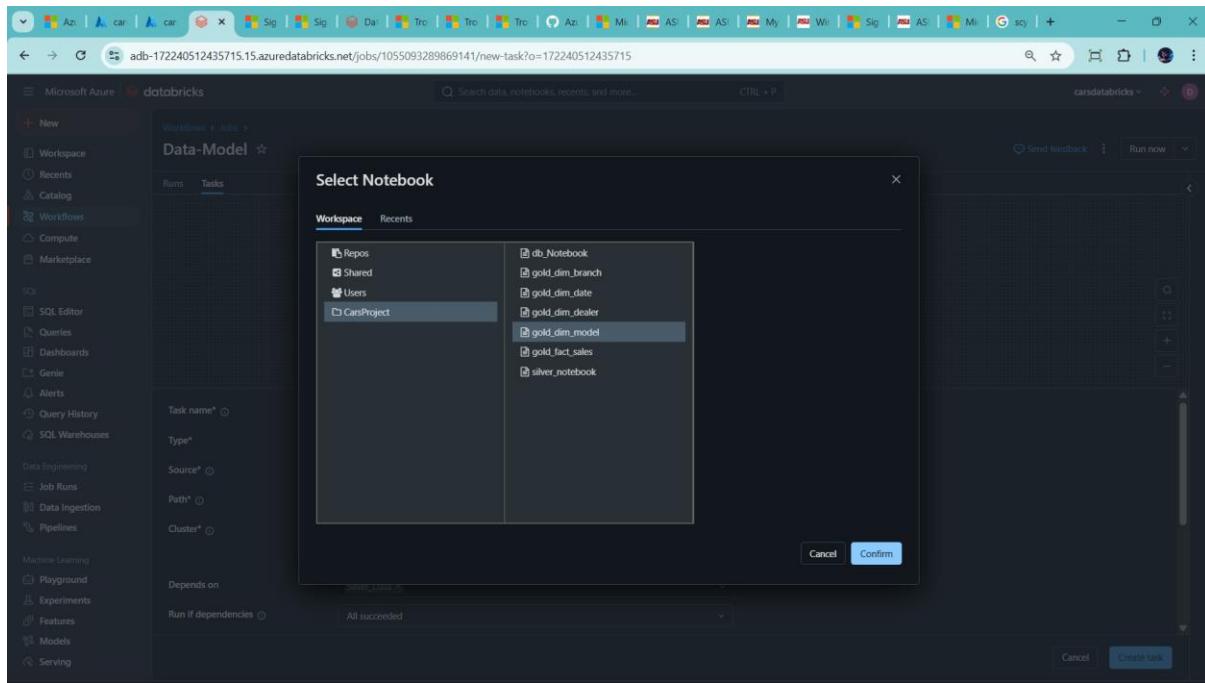
Path\*: /Workspace/CarsProject/silver\_notebook

Cluster\*: Dinesh km's Cluster

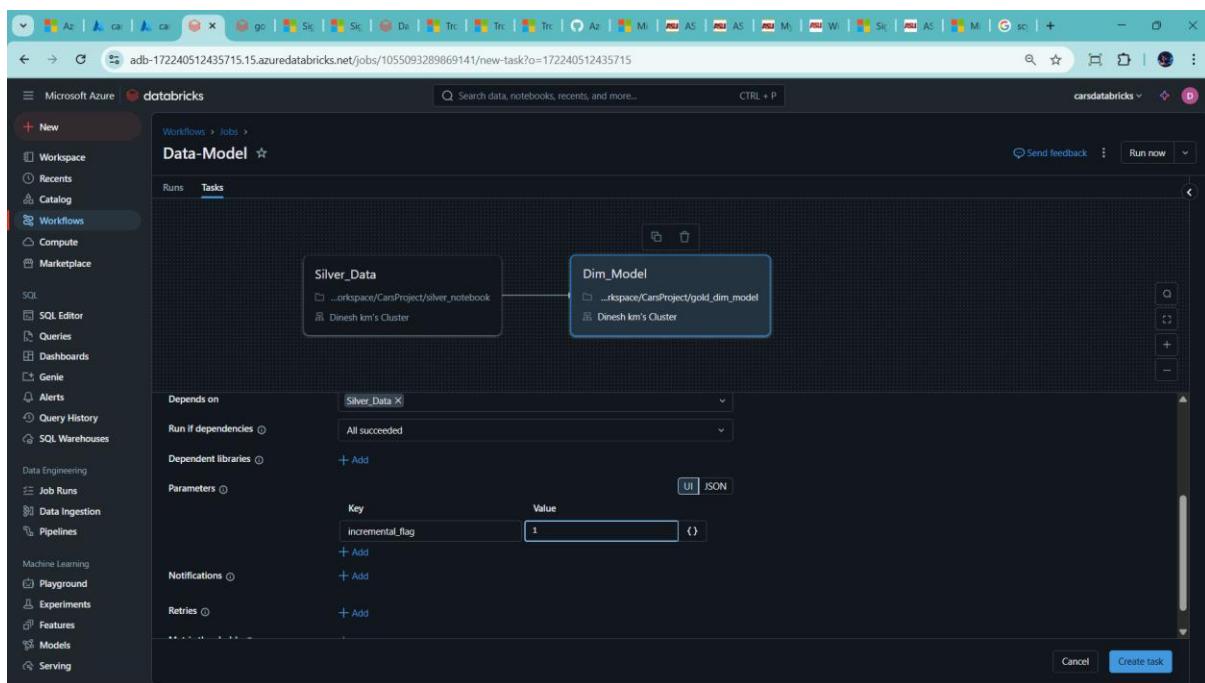
Dependent libraries: + Add

Parameters: UI JSON

Cancel Save task

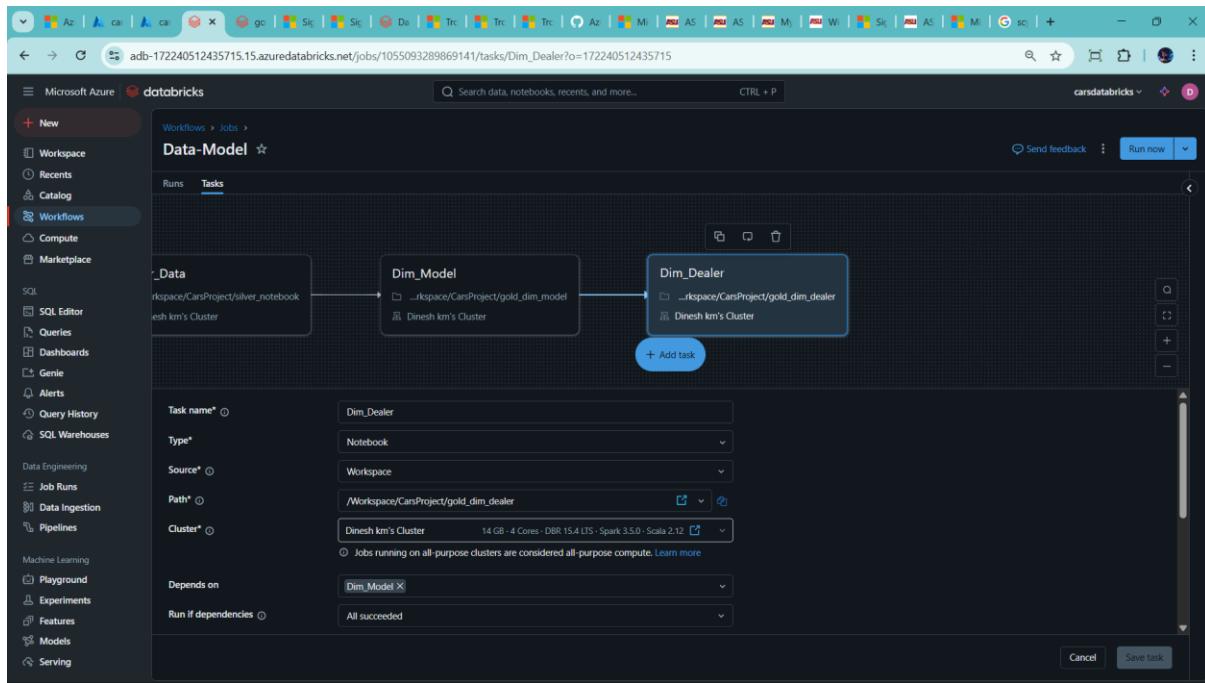


→add parameter →incremental\_flag

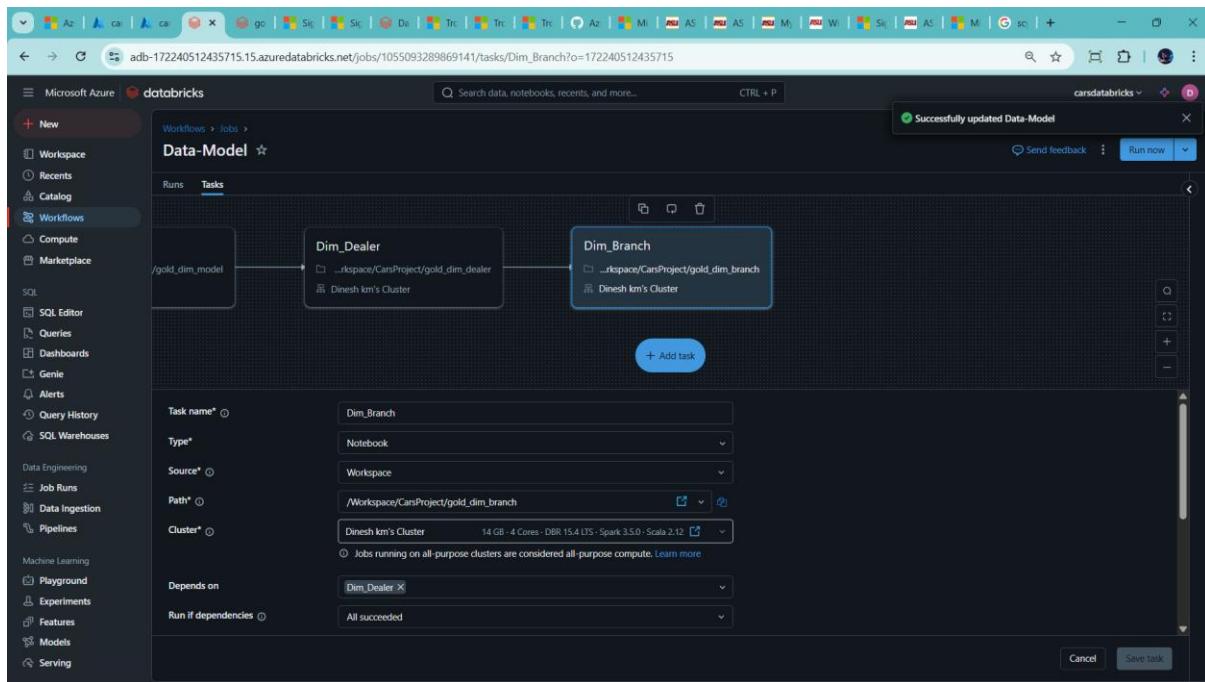


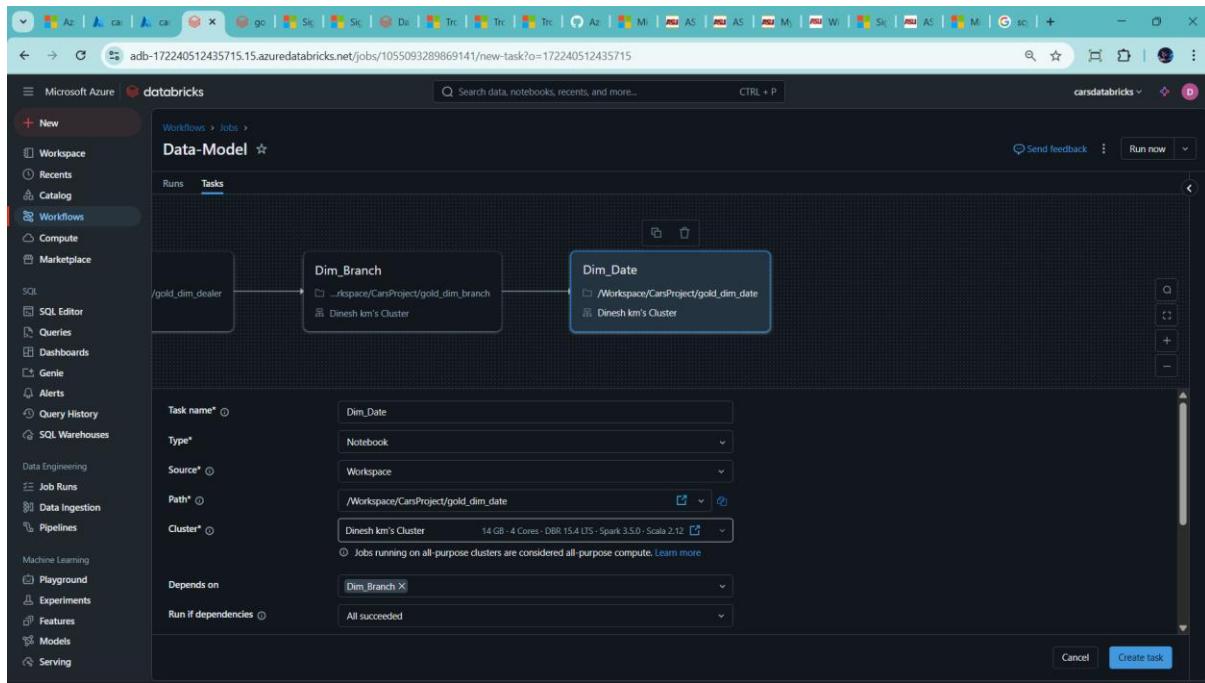
--click create

→same as other Dim assign task

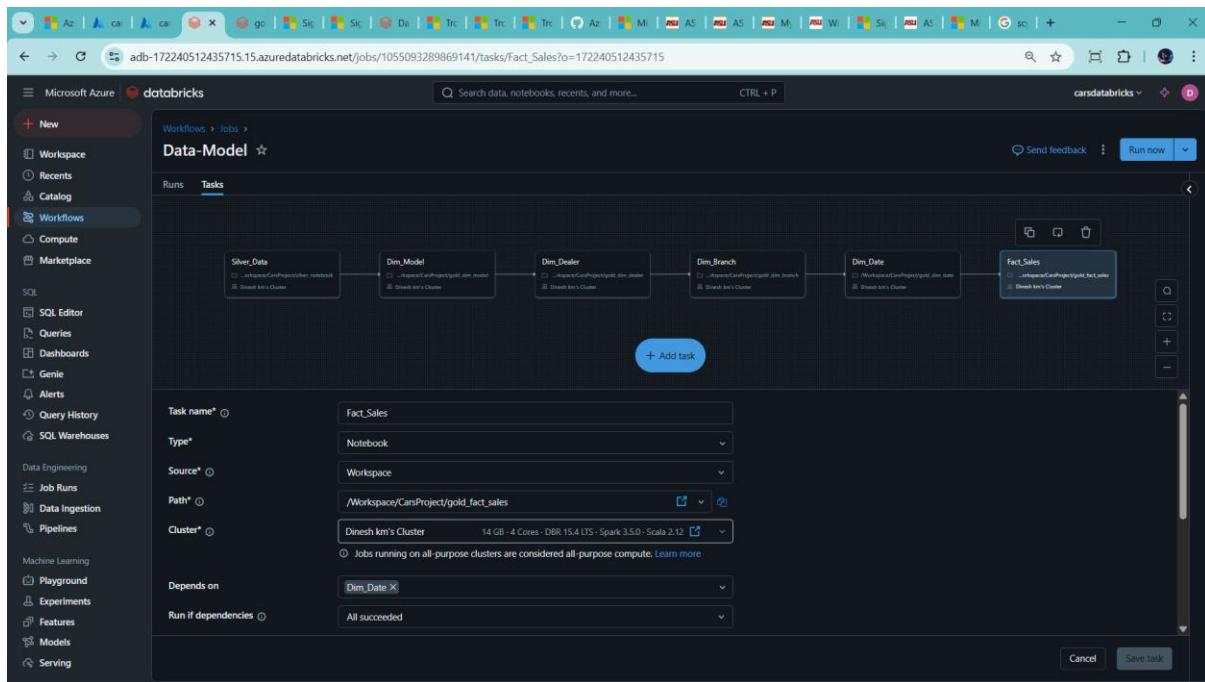


--create another task Dim\_Branch

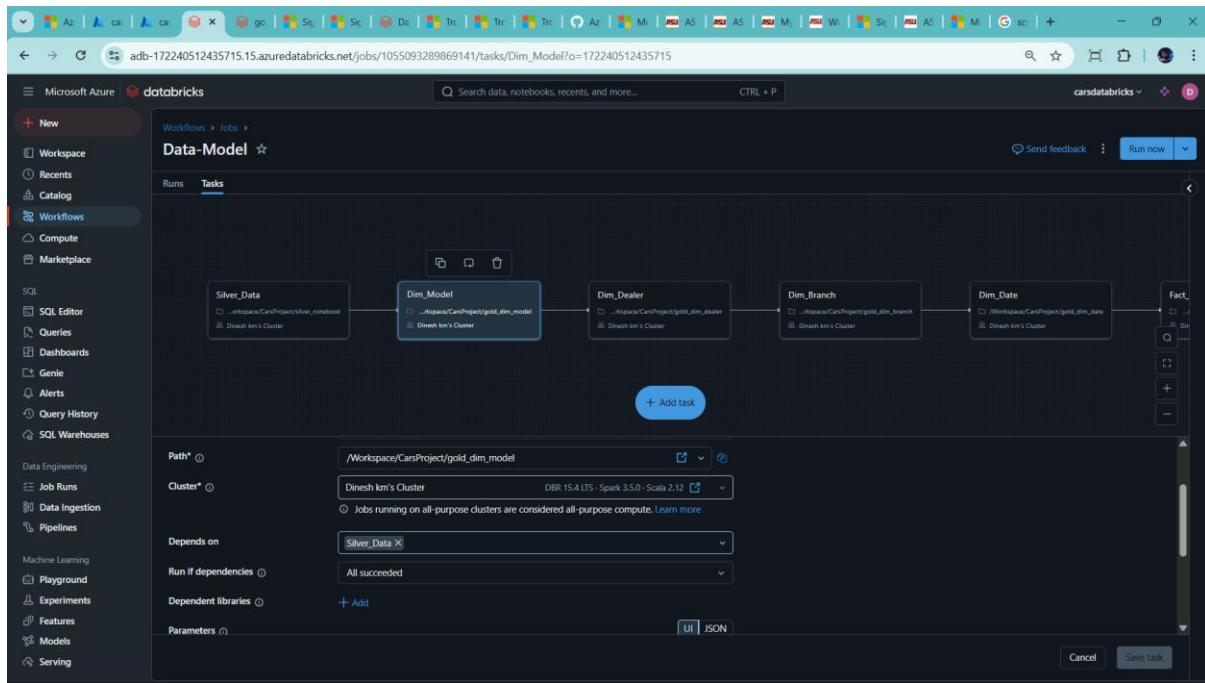




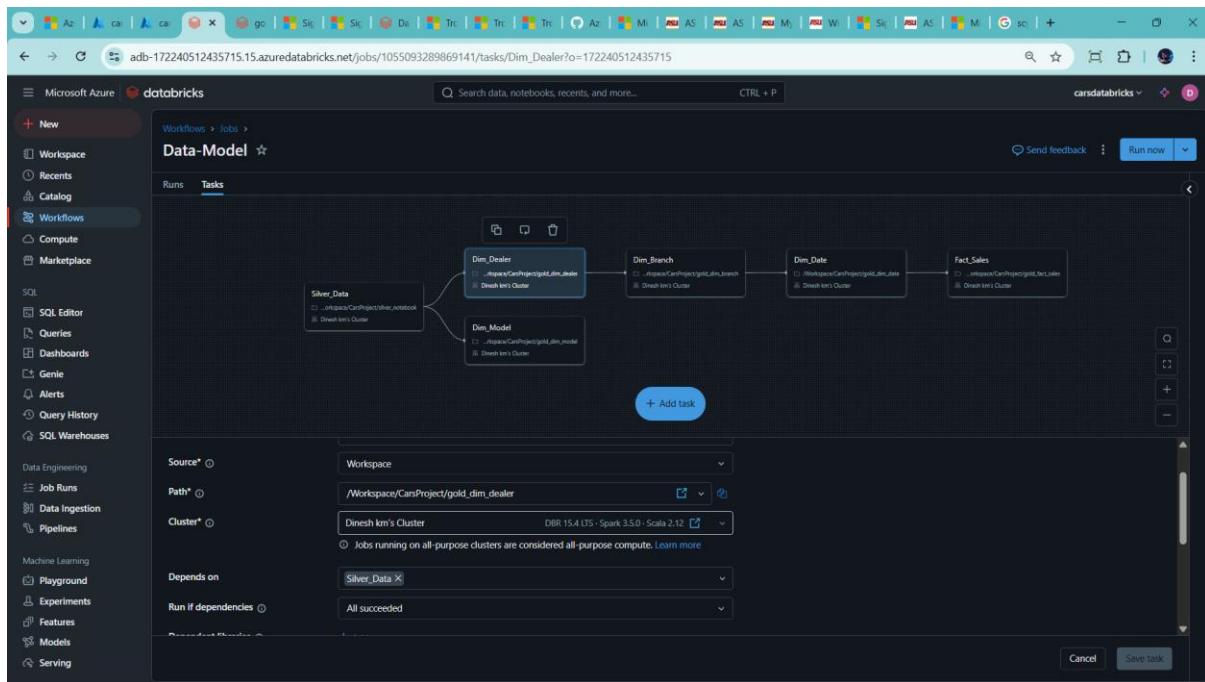
--Create task as fact\_sales notebook as well

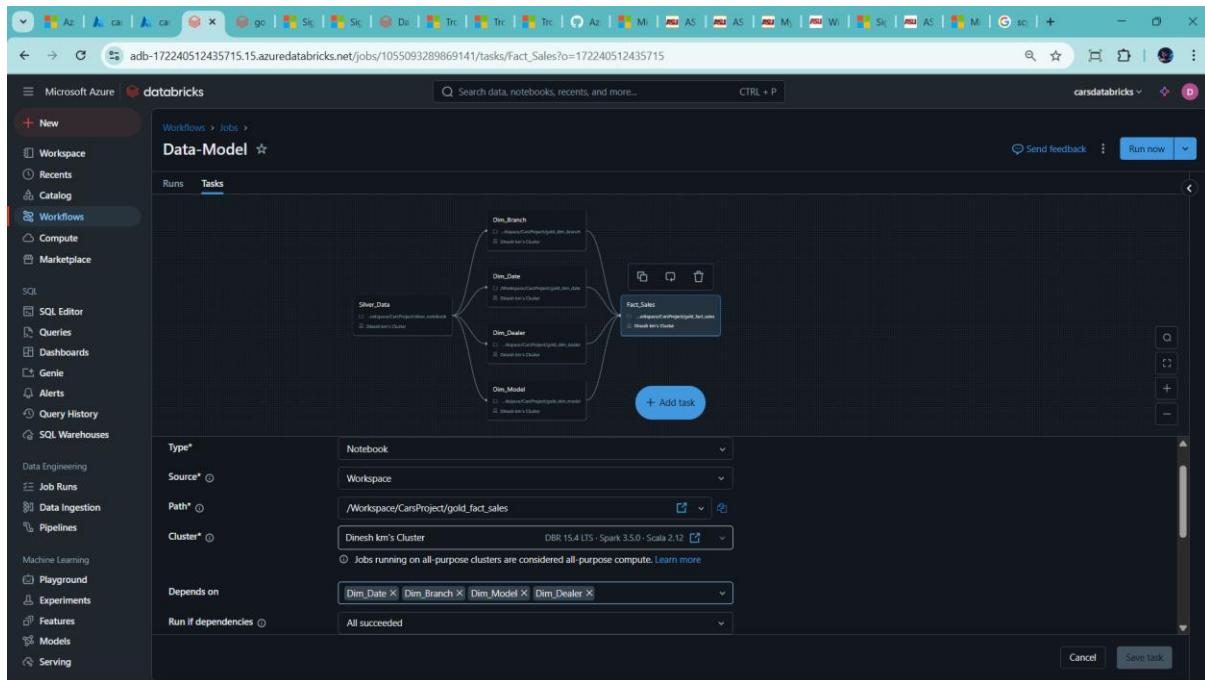


→ I want some flow run in parallel



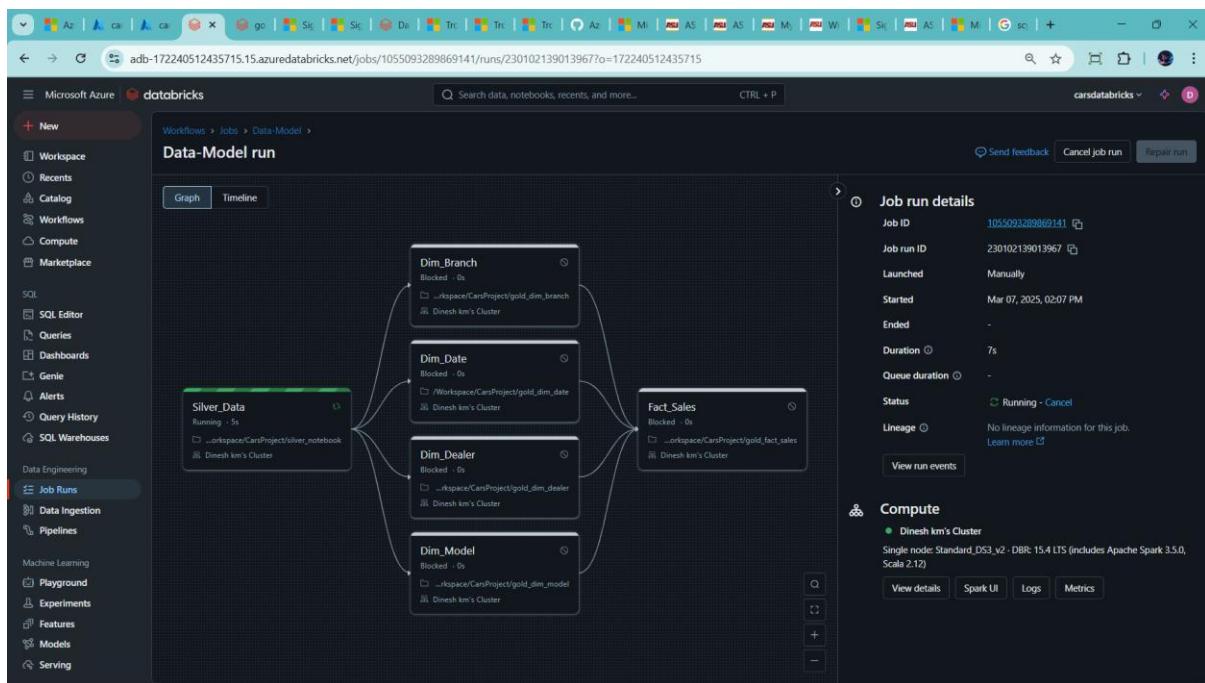
--click Depends on → choose silver data





→ merge all dim task in fact task

--finally click run



→ go to SQL query editor check table

→ Success

→ Try incremental load method

Go data factory

The screenshot shows the Microsoft Azure Data Factory Pipeline Editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (increment\_data-pipeline, source\_prep), 'Datasets', 'Data flows', and 'Power Query'. The main workspace displays a pipeline named 'source\_prep' with a single 'Copy' activity named 'CopyGitData'. The 'Source' tab is selected, showing the configuration for the 'ds\_git' dataset. The 'Request method' is set to 'GET'. The 'Additional headers' section is empty.

They have new data and old data with updated

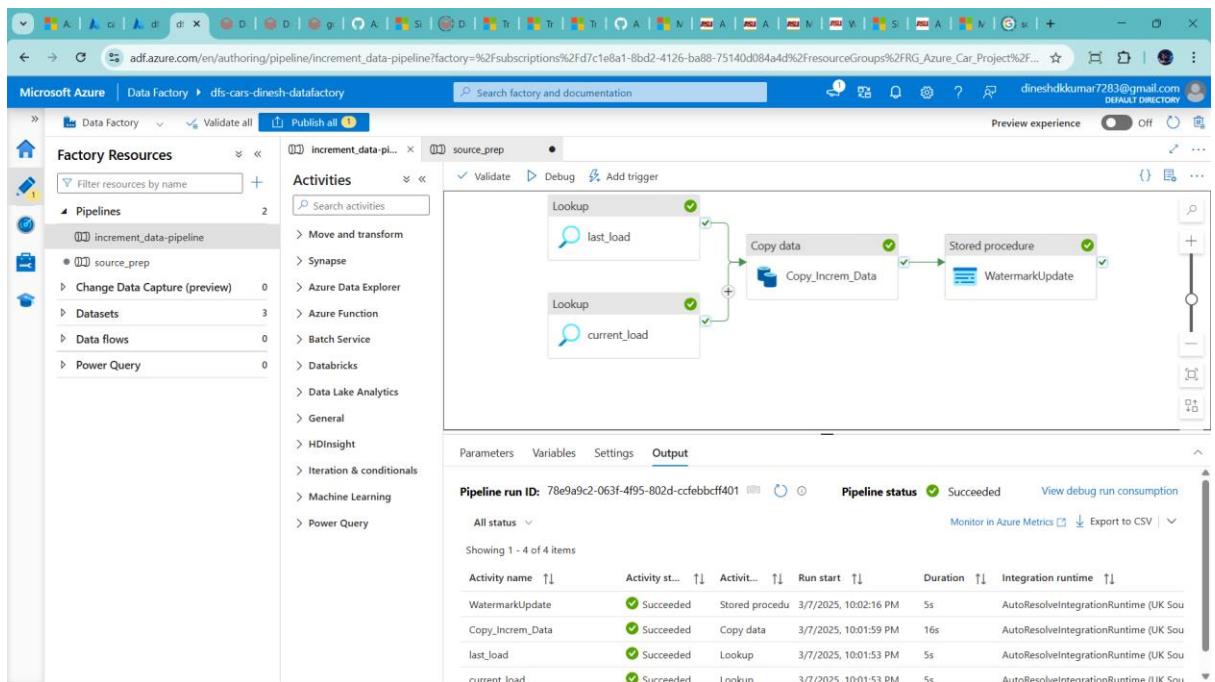
The screenshot shows a GitHub repository page for 'Azure-DE-Project-Resources'. The 'Raw Data' folder contains the 'IncrementalSales.csv' file. The file content is displayed as a table:

	Branch_ID	Dealer_ID	Model_ID	Revenue	Units_Sold	Date_ID	Day	Month	Year	BranchName	DealerName	Product_Name
1	BR9546	DLR0060	Jee-M10	7223451	1	DT01246	28	5	2020	Premier Motors	Fisker, Karma Motors	Jeep
2	BR9666	DLR0062	Jee-M12	22093020	3	DT01246	30	5	2020	Puma Motors	Ford Australia Motors	Jeep
3	BR9726	DLR0063	Jee-M13	22372413	3	DT01247	31	5	2020	Power Ranger Motors	Ford do Brasil Motors	Jeep
4	XYZ29726	XYZ0063	ZYX0M13	22372413	3	DT01247	31	5	2020	Datafarm Motors	Datafarm Dealers	Surprise

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar is open, showing a tree structure with 'Pipelines' expanded, containing 'increment\_data\_pipeline' and 'source\_prep'. The 'source\_prep' pipeline is selected. In the main workspace, a 'Copy data' activity named 'CopyGitData' is selected. Below it, the 'Source' tab is active in the configuration pane. The 'Source dataset' dropdown is set to 'ds\_qgit'. Under 'Dataset properties', there is a single entry: 'load\_flag' with a value of 'IncrementalSales.csv'. The 'Request method' is set to 'GET'. There are no additional headers defined.

This screenshot shows the same Azure Data Factory pipeline editor interface, but the 'Sink' tab is now active in the configuration pane. The 'Sink dataset' dropdown is set to 'db\_sqlDB'. Under 'Dataset properties', there is a single entry: 'table\_name' with a value of 'source\_car\_data'. The 'Write behavior' section shows 'Insert' selected. The 'Bulk insert table lock' section has 'No' selected. The 'Table option' section shows 'Use existing' selected.

Now run



**Query 1**

```
1 select * from [dbo].[water_table]
```

**Results**

last_load
DT01247

Query succeeded | 0s

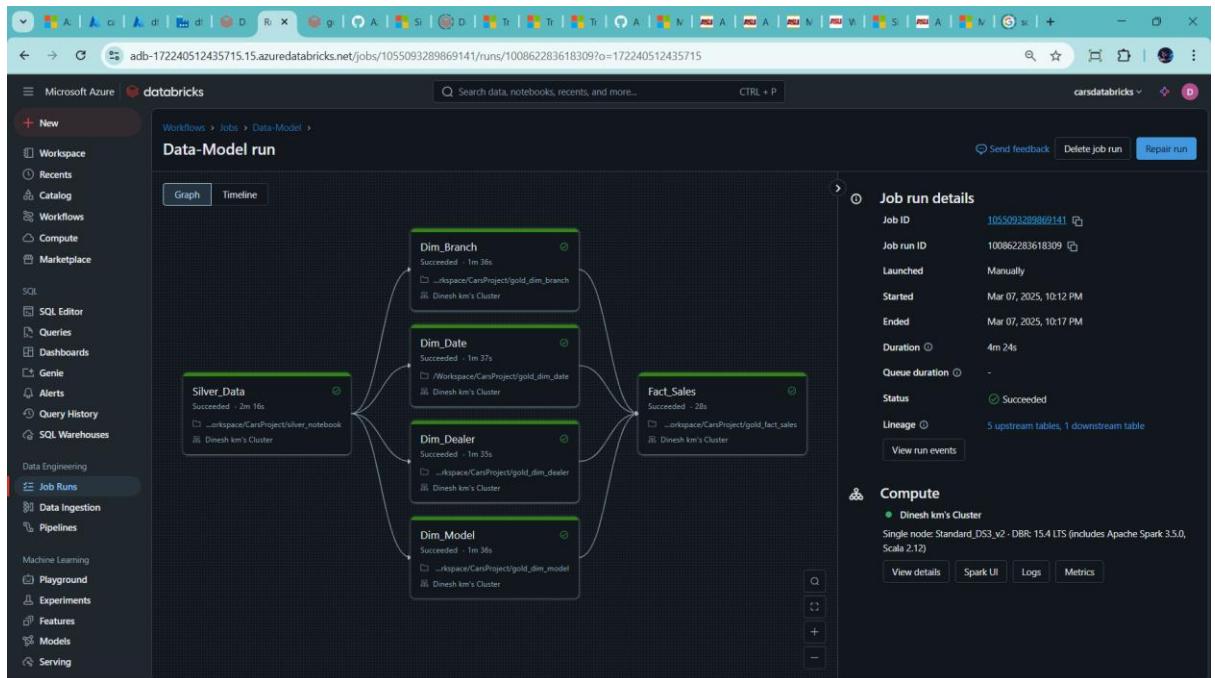
→ Run work flow in databricks

- Small mistake to miss add +1 in else as well
- Apply all dim table

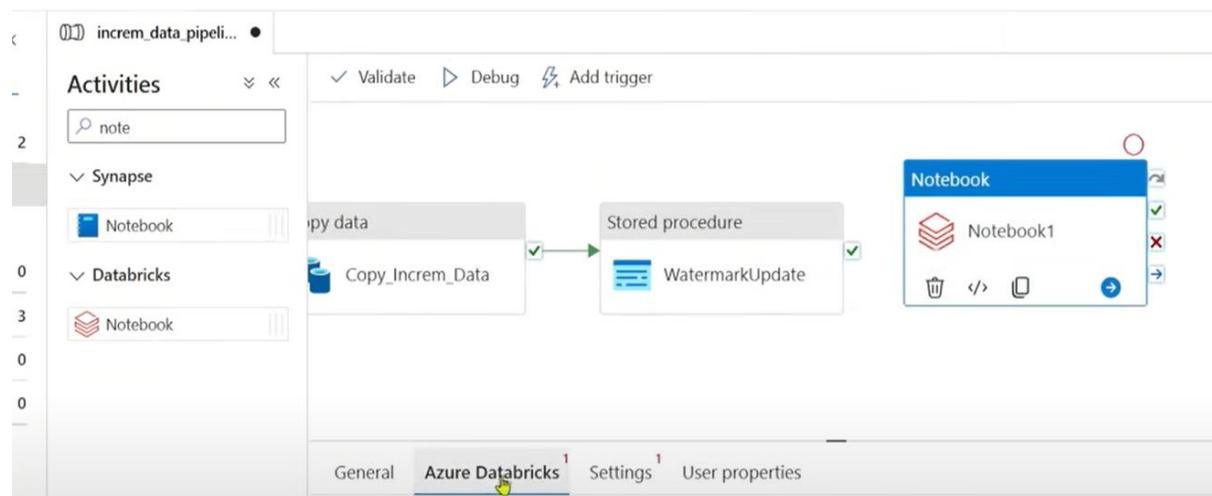
```

gold_dim_branch Python
File Edit View Run Help Last edit was now
Search data, notebooks, recents, and more... CTRL + P
Run all Terminated Schedule (1) Share
+ ⚙️ 🌐 🔍
New Workspace Recents Catalog Workflows Compute Marketplace
SQL SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses
Data Engineering Job Runs Data Ingestion Pipelines
Machine Learning Playground Experiments Features Models Serving
Create Surrogated key
Fetch the max Surrogate key from existing table
26
if incremental_flag==0:
    max_value=1
else:
    max_value_df=spark.sql("select max(dim_branch_key) from cars_catalog.gold.dim_branch")
    # convert too list
    max_value=max_value_df.collect()[0][0]+1
Python
Create Surrogate key column and ADD the max surrogated key
28
df_filter_new=df_filter_new.withColumn('dim_branch_key',max_value+monotonically_increasing_id())
df_filter_new=pyspark.sql.dataframe.DataFrame = [BranchID:string, BranchName:string ... 1 more field]

```



→ Also use notebook in Data factory as well



The screenshot shows the 'New linked service' configuration dialog for 'Azure Databricks'. The 'Azure Databricks' tab is selected. The configuration fields include:

- AutoResolveIntegrationRuntime**: A checked checkbox.
- Account selection method \***: A radio button group where 'From Azure subscription' is selected.
- Azure subscription \***: A dropdown menu showing 'Azure subscription 1 (0c02fec8-befb-4569-98ee-ef57af8ad2)'.
- Databricks workspace \***: An input field containing 'carsdatabricks'.
- Select cluster**: A radio button group where 'Existing interactive cluster' is selected.
- Databrick Workspace URL \***: An input field containing 'https://adb-1400347041729222.2.azuredata.databricks.net'.
- Authentication type \***: A radio button group where 'Access Token' is selected.
- Access token**: A selected tab under the authentication section.
- Azure Key Vault**: A tab under the authentication section.
- Access token \***: An input field for the access token.

Find access token

Microsoft Azure | databricks

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

+ New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

## Settings

Workspace admin

Appearance

Identity and access

Security

Compute

Development

Notifications

Advanced

User

Profile

Preferences

Developer

Linked accounts

Notifications

User settings

Access

Generate new token

Comment: What's this token for?

Lifetime (days): 90

Cancel Generate

Expiration

Successful connect

https://gwcteq-partner.domo.com/datasources/a57aa9cf-a150-40dd-8f4b-2235379a4436/details/data/table

DOMO + Databricks FLOW 1,853 rows

OVERVIEW AI READINESS DATA CARDS SETTINGS LINEAGE HISTORY PDP ALERTS AutoML OPEN WITH

Revenue	Units_Sold	RevPerUnit	dim_branch_key	dim_dealer_key	dim_model_key
123	Integer 123	Floating Decimal	123	Integer 123	Integer 123
1	13,363,978	2	6,681,989.00	418	6
2	17,376,468	3	5,792,156.00	1,557	197
3	9,664,767	3	3,221,589.00	1,058	104
4	5,525,304	3	1,841,768.00	789	95
5	12,971,088	3	4,323,696.00	497	231
6	7,321,228	1	7,321,228.00	1,804	41
7	11,379,294	2	5,689,647.00	734	177
8	11,611,234	2	5,805,617.00	1,211	182
9	19,979,446	2	9,989,723.00	116	204
10	14,181,510	3	4,727,170.00	116	160
11	5,358,057	1	5,358,057.00	1,423	103
12	16,150,431	3	5,383,477.00	432	141
13	13,389,350	2	6,694,675.00	661	89
14	4,891,618	2	2,445,809.00	1,290	198
15	5,059,144	2	2,529,572.00	365	254
16	17,369,466	2	8,684,733.00	419	3

→Follow this step connect databrick in DOMO

## 1. Enable Databricks SQL Warehouse

Ensure that your **Databricks SQL Warehouse** (formerly SQL Endpoint) is enabled:

- Go to **Databricks Workspace > SQL > SQL Warehouses**.
  - Start or create a new SQL Warehouse.
  - Copy the **Server Hostname, Port**, and **HTTP Path** from the **Connection Details**.
- 

## 2. Get Authentication Credentials

- You need an **Access Token** from Databricks.
  - In **Databricks**, go to **User Settings > Access Tokens > Generate New Token**.
  - Copy and save the token securely.
- 

## 3. Connect Databricks to Domo

1. **Go to Domo > Data Center > Connectors.**
  2. Search for **Databricks Connector** and select it.
  3. Click **Get the Data**.
  4. Fill in the connection details:
    - **Host:** <Databricks Server Hostname>
    - **Port:** 443
    - **HTTP Path:** <Databricks HTTP Path>
    - **Token:** <Databricks Access Token>
  5. Click **Connect**.
- 

## 4. Configure Data Import

- Choose the **Databricks database** and **table** you want to pull into Domo.
  - Set up **Query Mode** (Predefined Query or Custom SQL).
  - Define a **Schedule** for data updates.
  - Click **Save & Run**.
- 

## 5. Verify and Use the Data

- Once the data is imported, it appears as a **Domo dataset**.
- Use it in **Domo Analyzer, Cards, or Dashboards**.

<https://gwcteq-partner.domo.com/datacenter/datasources>

**DataSets**

CONNECT DATA TRANSFORM DATA ADVANCED TOOLS

QUICK FILTERS: All DataSets, Recently run, Owned by you, Needs attention.

FAVORITE FILTERS: + Add a saved filter

RESULTS (151) Created Date (newest to oldest)

Name	Owner	Rows	Cards	Last Run
dim_model	KM Dinesh	278	0	Mar 10, 2025 11:29 AM
dim_dealer	KM Dinesh	268	0	Mar 10, 2025 11:27 AM
dim_branch	KM Dinesh	1.8K	0	Mar 10, 2025 11:26 AM
dim_date	KM Dinesh	1.2K	0	Mar 10, 2025 11:25 AM
factsales	KM Dinesh	1.9K	0	Mar 10, 2025 11:29 AM
ROLES	Ajith Chandran	31	0	Mar 4, 2025 4:59 PM

<https://gwcteq-partner.domo.com/datacenter/datallows/new/graph?type=ETL>

**CarSalesDatabrickFlow**

Description (Optional): Changes not yet saved. CANCEL SAVE

Search for an action: 10K ROW LIMIT 100% RUN PREVIEW

**DATASETS**: Input DataSet, Output DataSet, Writeback

**TEXT**, **DATES AND NUMBERS**, **UTILITY**, **FILTER**, **COMBINE DATA**: Append Rows, Join Data

**CarSalesDatabrickDataset** (CONFIGURATION):

- CarSalesDatabrickDataset
- Edit Description
- Update Method: Replace

Dashboard

<https://gwcteq-partner.domo.com/page/2002150061>

## Car Sales Dashboard

Total Units Sold: 3.75K

Total Revenue: 19.32B

Branch Name Filter: Select

Dealer Name Filter: Select

Model Filter: Select

**Top 5 Branches by Units Sold**  
 Highlights the top 5 performing branches based on total units sold

BranchName	Units_Sold
Herald Motors	25
Alpine Motors	24
BMW Motors	24
Acura Motors	23
Aston Martin Motors	23
Tazzari Motors	23
Izh Motors	23
DAF Motors	22
DeLorean Motors	22
Gilbern Motors	22
Lamborghini Motors	22
Maserati Motors	22
Jennings Ford Automobile Dealership	22
Ariel Motors	21
Auto-Union Motors	21

**Units Sold by Model Category**  
 Shows the distribution of units sold across different model categories

ModelCategory	Units_Sold
Toy	285
Nis	267
For	234
Hyu	232
Hon	217
Mar	213
Aud	213
Tat	209
Mah	195
BMW	182
Che	176
Vid	150

<https://gwcteq-partner.domo.com/page/2002150061>

## Car Sales Databricks Dashboard

Add filters to your Dashboard to find new insights.

**Revenue Distribution by Model Category**  
 Shows the proportion of revenue generated by each model category

ModelCategory	Revenue
Toy	1,462,284,358
Nis	1,341,099,624
For	1,277,177,477
Hyu	1,193,778,032
Hon	1,181,142,484
Mar	1,170,415,017
Tat	1,056,588,532
Mah	924,613,296
Che	822,401,296
Vid	782,705,857
Other	6,070,521,48

**Top 5 Branches by Units Sold**

BranchName	sum_of_Units_Sold
Audi Motors	30
Bristol Motors	28
Herald Motors	30
Alpine Motors	28
BMW Motors	28

**Units Sold vs Revenue per Unit by Model Category**  
 Compares unit sales and average revenue per unit across model categories

model_category	sum_of_Units_Sold	RevPerUnit
Toy	~700M	~1.4M
Nis	~650M	~1.3M
For	~600M	~1.2M
Hyu	~550M	~1.1M
Hon	~500M	~1.0M
Mar	~450M	~0.9M
Tat	~400M	~0.8M
Mah	~350M	~0.7M
Che	~300M	~0.6M
Vid	~250M	~0.5M
Other	~200M	~0.4M

<https://gwcteq-partner.domo.com/page/2002150061>

**CarSalesDatabrickDashbaord**

Add filters to your Dashboard to find new insights.

**Dealer Performance Overview**

**Top 5 Dealers by Total Revenue**  
Identifies the highest-performing dealers based on total revenue

DealerName	sum of Revenue
Bentley Motors	200M
Asia Motors Motors	180M
BMW Motors	160M
Alfa Romeo Motors	140M
Acura Motors	120M

**Dealer Performance Summary**  
Provides a detailed summary of each dealers performance including revenue and units sold

DealerName	Revenue ↓	Units_Sold
Ariel Motors	112,476,647	24
Bitter Motors	111,496,713	21
Santana Motors	111,355,615	20
Anadol Motors	110,573,515	28
Austin-Healey Motors	110,547,199	19
Ginetta Motors	109,655,975	19
Borgward (including Goliath and Lloyd) Motors	109,646,064	18
Spyker Motors	108,762,831	16
Holden Motors	108,079,210	22
Chevrolet India Motors	107,858,563	14

**Monthly Units Sold Trend**  
Shows the trend of units sold over time to identify sales performance patterns

**CarSalesDatabrickDashbaord**

Add filters to your Dashboard to find new insights.

**Average Revenue per Unit by Model Category**  
Compares average revenue per unit across different model categories

model_category	Avg Revenue per Unit
Z100M13	7,457,471
Fiat	7,434,208,47
Opel	6,975,210,21
Cit	6,176,195,33
MG	6,110,086,36
Volvo	5,877,079,33
Renault	5,857,037,33
Rover	5,521,613,115
Skoda	5,495,625,115
Che	5,414,660,07
Saab	5,363,284,79
BMW	5,250,250,00
Other	119,170,513,12

**Revenue Comparison by Model Category**  
Compares total revenue generated by each model category

model_category	sum of Revenue
Z100M13	1.5B
Fiat	1.4B
Opel	1.3B
Cit	1.2B
MG	1.1B
Volvo	1.0B
Renault	900M
Rover	800M
Skoda	700M
Che	600M
Saab	500M
BMW	400M
BMW	300M
BMW	200M
BMW	100M
BMW	50M
BMW	10M
BMW	5M
BMW	2M
BMW	1M
BMW	0.5M

**Branch Performance Summary**

BranchName	Revenue ↓	Units_Sold
Bristol Motors	179,960,947	25
Audi Motors	171,342,111	29
Tazzari Motors	163,209,329	23
Mastretta Motors	151,455,831	22
Jeep Motors	145,904,655	19
Gilbern Motors	145,457,488	22
Delorean Motors	144,998,617	22
DAF Motors	142,148,873	22
Auto-Union Motors	138,448,557	21
Opel Motors	136,882,657	19
Alpine Motors	136,350,305	24

<https://gwcteq-partner.domo.com/page/2002150061>

**CarSalesDatabrickDashbaord**

Add filters to your Dashboard to find new insights.

**Branch Performance Summary**

BranchName	Revenue ↓	Units_Sold
Bristol Motors	179,960,947	25
Audi Motors	171,342,111	29
Tazzari Motors	163,209,329	23
Mastretta Motors	151,455,831	22
Jeep Motors	145,904,655	19
Gilbern Motors	145,457,488	22
Delorean Motors	144,998,617	22
DAF Motors	142,148,873	22
Auto-Union Motors	138,448,557	21
Opel Motors	136,882,657	19
Alpine Motors	136,350,305	24



