

Azure Databricks Project

Resource Group and Resource Creation

1. Create Azure Account (Pay As You Go Subscription) - [Create Your Azure Free Account Or Pay As You Go | Microsoft Azure](https://azure.microsoft.com/en-us/pricing/purchase-options/azure-account/search?ef_id=_k_EAlalQobChMlwL_5y4zPiwMVYQfvAh3MmgfEAAYASAAEglo8vD_BwE_k_&C)

The screenshot shows a web browser window with the URL https://azure.microsoft.com/en-us/pricing/purchase-options/azure-account/search?ef_id=_k_EAlalQobChMlwL_5y4zPiwMVYQfvAh3MmgfEAAYASAAEglo8vD_BwE_k_&C. The page displays two main sections: 'Azure free account' and 'Pay as you go'.

Azure free account
Best for proof of concept and exploring capabilities

- Available only to new Azure customers
- Free monthly amounts of 20+ popular services for 12 months (new Azure customers only)
- Free monthly amounts of 65+ always-free services
- Access to full catalog of services up to free amounts and \$200 credit
- Spending protection—credit card won't be charged*
- No upfront commitment—cancel anytime
- Move to pay-as-you-go pricing to continue beyond 30 days or after credit is used up

\$200 credit
to use on Azure services within 30 days

Sign up

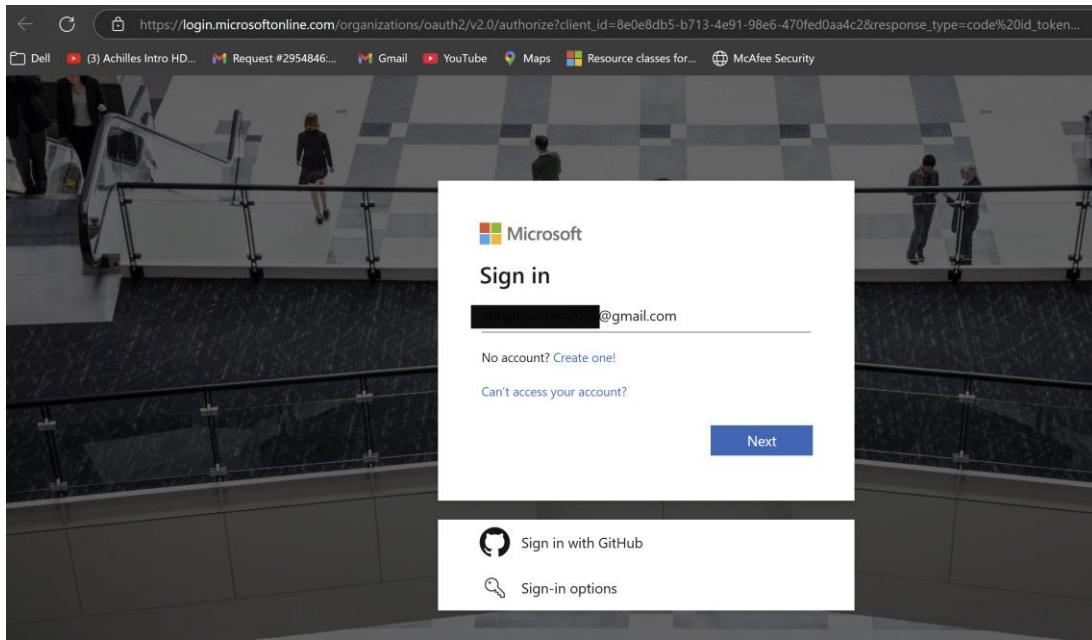
Pay as you go
Best for customers ready to start building workloads.

- Free monthly amounts of 20+ popular services for 12 months (new Azure customers only)
- Free monthly amounts of 65+ always-free services
- Access to full catalog of services with no cap on service usage
- Technical support options available
- No upfront commitment—cancel anytime
- No action required to continue beyond 30 days

Pay only for what you use
beyond free monthly amounts

Sign up

Provide your email-address, credit card and other details to complete the signup process



2. Now we should be able to login to Azure Portal

A screenshot of the Microsoft Azure portal home page. The browser tab shows "Home - Microsoft Azure". The main header includes the Microsoft logo, a search bar ("Search resources, services, and docs (G+)"), and various navigation icons like Copilot, Help + support, and Azure Synapse Analytics. On the left, there's a sidebar titled "Azure services" with icons for "Create a resource", "Resource groups", "Microsoft Entra ID", "Azure Databricks", "Subscriptions", "Quotas", "Help + support", "Azure Synapse Analytics", "Billing subscriptions", and "More services". Below this is a section titled "Resources" with a "Recent" tab selected, showing a single item: "Azure subscription 1" (Subscription type, last viewed a minute ago). There are also "See all" and "Favorite" tabs. At the bottom, there's a "Navigate" section with links to "Subscriptions", "Resource groups", "All resources", and "Dashboard".

3. Create Resource Group

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

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

Home > Resource groups >

Create a resource group

Basics Tags Review + create

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Subscription * ⓘ Azure subscription 1

Resource group name * ⓘ rg-de-databricks-projects

Region * ⓘ (US) West US 2

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

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

Home > Resource groups >

Create a resource group

Basics Tags Review + create

Automation Link

Basics

Subscription	Azure subscription 1
Resource group name	rg-de-databricks-projects
Region	West US 2

Tags

None

Previous Next Create

Microsoft Azure

Search resources, services, and docs (G+)

rg-de-databricks-projects

Resource groups

Default Directory

+ Create Manage view ...

Filter for any field... Name ↑ rg-de-databricks-projects ...

Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Events

Settings

Cost Management

Monitoring

Automation

Help

Essentials

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

Try changing or clearing your filters.

+ Create resources Clear filters

Learn more Give feedback

4. Create Azure Synapse Analytics and Storage Account (ADLS Gen2)

Microsoft Azure

Search resources, services, and docs (G+)

rg-de-databricks-projects

Marketplace

Get Started

Service Providers

Management

Private Marketplace

Private Offer Management

My Marketplace

Favorites

My solutions

Recently created

Private plans

Categories

Analytics (77)

Databases (32)

AI + Machine Learning (22)

synapse analytics

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

Azure services only

Publisher name : All ×

New! Get AI-generated suggestions for 'synapse analytics' View suggestions

Showing 1 to 20 of 95 results for 'synapse analytics': Clear search

Tile view

Image	Name	Provider	Description	Starts at	Subscribe
	Azure Synapse Analytics	Microsoft	Azure Service	\$10,000.00/month	Create
	Azure Synapse Analytics (private link hubs)	Microsoft	Connect to Azure Synapse Studio using private endpoints		Create
	Datometry Hyper-Q for Azure Synapse Analytics	Datometry	Re-platform from Teradata and Exadata to Azure Synapse at a fraction of the time, cost, and risk		Subscribe
	Synapse Data Fabric	Spektra Systems LLC	SaaS	\$0.001/month	Subscribe
	Datometry Hyper-Q for Azure Synapse Analytics	Datometry	Run existing Teradata applications natively on Azure Synapse Analytics	\$10,000.00/month	Subscribe

Synapse Analytics Workspace creation -> synapse-workspace-demo-dbs-01
Basic Page (work space details)

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

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

Home > Resource groups > rg-de-databricks-projects > Marketplace > Create Synapse workspace ...

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name * synapse-workspace-demo-dbs-01

Region * West US 2

Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name * (New) adlsgen2demodbs Create new

File system name * (New) bronze Create new

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the Storage Blob Data Contributor role. To enable other users to use this storage account after you create your workspace, perform these tasks:

- Assign other users to the **Contributor** role on workspace
- Assign other users the appropriate [Synapse RBAC roles](#) using Synapse Studio
- Assign yourself and other users to the **Storage Blob Data**

Review + create < Previous Next: Security >

Note: Make sure the "Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace." check box is enabled. This will make sure the Synapse Analytics have required access to the Storage account.

Security Page

← → ⌂ portal.azure.com/#create/Microsoft.Synapse

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

Home > Resource groups > rg-de-databricks-projects > Marketplace >

Create Synapse workspace

Configure security options for your workspace.

Authentication

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more](#)

Authentication method ⓘ

Use both local and Microsoft Entra ID authentication
 Use only Microsoft Entra ID authentication

SQL Server admin login * ⓘ

sqladminuser

SQL Password ⓘ

Enter server password ✓

Confirm password

Confirm the above password ✓

System assigned managed identity permission

Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity. [Learn more](#)

Allow network access to Data Lake Storage Gen2 account. ⓘ
 ⓘ The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more](#)

Workspace encryption

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

Networking Page

← → G portal.azure.com/#create/Microsoft.Synapse

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

Home > Resource groups > rg-de-databricks-projects > Marketplace >

Create Synapse workspace

Managed virtual network

Choose whether to set up a dedicated Azure Synapse-managed virtual network for your workspace. [Learn more](#)

Managed virtual network Enable Disable

ⓘ To control public network access to your Synapse workspace, you must enable managed virtual network.

Firewall rules

⚠ Azure Synapse Studio and other client tools will only connect to the workspace endpoints if this setting is selected. Connections from specific IP addresses or all Azure services can be allowed or disallowed after the workspace is provisioned.

Allow connections from all IP addresses to your workspace's endpoints. You can restrict these permissions to just Azure datacenter IP addresses and/or specific IP address ranges after creating the workspace.

Allow connections from all IP addresses

Encrypted connections

Azure Synapse Analytics applies TLS1.2 encryption protocol for all network connections. [Learn more](#)

Minimum TLS version

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

← → G portal.azure.com/#create/Microsoft.Synapse

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

Home > Resource groups > rg-de-databricks-projects > Marketplace >

Create Synapse workspace

* Basics * Security Networking Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [learn more](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ	Value ⓘ	Resource
<input type="text"/>	:	Synapse workspace

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

Now review and create the resource.

← → G portal.azure.com/#create/Microsoft.Synapse

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

Home > Resource groups > rg-de-databricks-projects > Marketplace >

Create Synapse workspace

Validation succeeded

* Basics * Security Networking Tags Review + create

Product Details

Azure Synapse Analytics workspace by Microsoft Serverless SQL est. cost/TB ⓘ
5.00 USD
[Terms of use](#) | [Privacy policy](#)

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. For additional details see [Azure Marketplace Terms](#).

Basics

Subscription	Azure subscription 1
Resource group	rg-de-databricks-projects
Region	West US 2

Create < Previous Next > Download a template for automation

← → G portal.azure.com/#create/Microsoft.Synapse

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

Home > Resource groups > rg-de-databricks-projects > Marketplace >

Create Synapse workspace

✓ Validation succeeded

Region	West US 2
Workspace name	(new) synapse-workspace-demo-dbs-01
Data Lake Storage Gen2 account	(new) https://adlsgen2demodbs.dfs.core.windows.net
Data Lake Storage Gen2 file system	(new) bronze
Managed resource group	None
Role assignments	The Storage Blob Data Contributor role will be assigned on the specified Data Lake Storage Gen2 account to both the workspace managed identity and the current user.

Security

Authentication method	Use both local and Microsoft Entra ID authentication
SQL Server admin login	sqladminuser
SQL Password	Auto-generated
Double encryption	No

Networking

Managed virtual network	No
Allow connections from all IP addresses	Yes

Create

< Previous

Next >

Download a template for automation

Microsoft Azure | Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview

Deployment

Deployment is in progress.

Deployment details:

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

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 >

Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview

Deployment name : Microsoft.Azure.SynapseAnalytics-20250218224016 Start time : 2/18/2025, 10:50:33 PM
Subscription : Azure subscription 1 Correlation ID : c38a3ce8-f72f-4ebe-bda8-e1b26f5ff1cb
Resource group : rg-de-databricks-projects

Deployment details

Resource	Type	Status	Operation details
synapse-workspace	Synapse workspace	Created	Operation details
adlsgen2demodbs	Microsoft.Storage/storageAccounts	Created	Operation details
adlsgen2demodbs	Storage account	OK	Operation details

[Give feedback](#)
[Tell us about your experience with deployment](#)

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20250218224016 Start time : 2/18/2025, 10:50:33 PM
Subscription : Azure subscription 1 Correlation ID : c38a3ce8-f72f-4ebe-bda8-e1b26f5ff1cb
Resource group : rg-de-databricks-projects

Deployment details

Resource	Type	Status	Operation details
StorageRoleDeploy	Deployment	OK	Operation details
synapse-workspace	Microsoft.Synapse/workspaces/1	OK	Operation details
synapse-workspace	Synapse workspace	OK	Operation details
synapse-workspace	Synapse workspace	OK	Operation details
adlsgen2demodbs	Microsoft.Storage/storageAccounts	Created	Operation details
adlsgen2demodbs	Storage account	OK	Operation details

After a success deployment, we can navigate to the resource.

rg-de-databricks-projects

Showing 1 to 2 of 2 records. Show hidden types

Name	Type	Location
adlsgen2demodbs	Storage account	West US 2
synapse-workspace-demo-dbs-01	Synapse workspace	West US 2

Use serverless endpoint to connect from SSMS (with MFA)

The screenshot shows the Microsoft Azure portal and Microsoft SQL Server Management Studio (SSMS).

Azure Portal Overview:

- Resource group: rg-de-databricks-projects
- Status: Succeeded
- Location: West US 2
- Subscription ID: 46ccf490-21e1-4902-95c-395310edbae
- Managed virtual network: No
- Managed Identity object ID: 2a1dd26e-6c4f-41d6-87b7-d5cc86e8e2ea
- Workspace web URL: https://web.azure.synapse.net/?workspace=%2fsubscriptions%2f46ccf490-21e1-4902-95c-395310edbae%2fresourcegroups%2frg-de-databricks-projects%2fworkspaces%2fsynapse-workspace-demo-dbs-01
- Tags: Add tags

Microsoft SQL Server Management Studio (SSMS) Object Explorer:

- Connected to synapse-workspace-demo-dbs-01-ondemand.sql.azuresynapse.net (SQL Server 12.0.2000.8 - live)
- Databases:
 - master
 - Tables
 - System Tables
 - Views
 - Programmability
 - Security
- Security

5. Create Azure Data Factory (ADF)

The screenshot shows the Microsoft Azure Marketplace search results for 'data factory'.

Search Filters:

- Get Started
- Service Providers
- Azure services only
- Pricing: All
- Operating System: All
- Publisher Type: All
- Publisher name: All

Marketplace Categories:

- Management
- Private Marketplace
- Private Offer Management
- My Marketplace
- Favorites
- My solutions
- Recently created
- Private plans
- Categories
- Analytics (51)
- AI + Machine Learning (28)

Showing 1 to 20 of 101 results for 'data factory':

Image	Name	Provider	Description	Starts at	Price varies
	Data Factory	Microsoft	Azure Service Hybrid data integration service that simplifies ETL at scale	\$1.00/1 year	Subscribe
	Sight Machine Factory CONNECT	Sight Machine, Inc.	SaaS Connecting Plant Data to Accelerate Industrial AI		
	Factory Namespace Manager	Sight Machine, Inc.	SaaS Create data dictionaries for renaming machine tags and variables for unified namespace		
	Modern Data Mart	Ceteris AG	Azure Application Small Data-Warehouse Architecture integrating a single source using Azure Data Factory		

Basic Page

Microsoft Azure

Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory ...

Basics Git configuration Networking Advanced Tags Review + create

One-click to create data factory with sample pipeline and datasets. [Try it](#)

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 *

Version *

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

Git Configuration (for code deployment – CI/CD)

← → ⌂ portal.azure.com/#create/Microsoft.DataFactory

Microsoft Azure

Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory ...

Basics **Git configuration** Networking Advanced Tags Review + create

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.
[Learn more about Git integration in Azure Data Factory](#)

Configure Git later

Networking Page

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

Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory

Basics Git configuration Networking **Advanced** Tags Review + create

Managed virtual network

Choose whether you want the default AutoResolveIntegrationRuntime to be provisioned on demand inside an ADF-managed virtual network. If this setting is disabled, after the data factory is created, you can still choose whether to provision explicitly created Azure integration runtime inside an ADF-managed virtual network.

[Learn more](#)

Enable Managed Virtual Network on the default AutoResolveIntegrationRuntime

Self-hosted integration runtime inbound connectivity to Azure Data Factory service

Choose whether to connect your self-hosted integration runtime to Azure Data Factory via public endpoint or private endpoint. This applies to self-hosted integration runtime running either on premises or inside customer managed Azure virtual network

[Learn more](#)

Connect via * Public endpoint Private endpoint

 You can change this or configure another connectivity method after this resource is created. [Learn more](#)

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

Advanced page for security and encryption

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

Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory

Basics Git configuration Networking **Advanced** Tags Review + create

Datafactory Encryption

By default, data is encrypted with Microsoft-managed keys. For additional control over encryption keys, you can supply customer-managed keys to use for encryption of blob and file data. Customer-managed keys must be stored in an Azure Key Vault. You can either create your own keys and store them in a key vault, or you can use the Azure Key Vault APIs to generate keys. The storage account and the key vault must be in the same region, but they can be in different subscriptions.

Enable encryption using a Customer Managed Key



Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory ...

Basics Git configuration Networking Advanced **Tags** Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more about tags](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ	Value ⓘ	Resource
<input type="text"/>	: <input type="text"/>	Data factory (V2)

Review and create resource.

Home > Microsoft.Azure.SynapseAnalytics-20250218224016 | Overview > rg-de-databricks-projects > Marketplace >

Create Data Factory ...

Basics Git configuration Networking Advanced Tags **Review + create**

[View automation template](#)

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 subscription 1
Resource group	rg-de-databricks-projects
Name	data-factory-databricks-demo-project
Region	West US 2
Version	V2

Networking

[Previous](#) [Next](#) **Create**

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

Home > Microsoft.DataFactory-20250218230351 | Overview

Deployment

Overview

Your deployment is complete

Deployment name : Microsoft.DataFactory-20250218230351 Start time : 2/18/2025, 11:08:26 PM
Subscription : Azure subscription 1 Correlation ID : 81535dee-9d93-4403-9428-c129cc62d2...
Resource group : rg-de-databricks-projects

Deployment details

Resource	Type	Status	Operation details
data-factory-datab	Data factory (V2)	OK	Operation details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

The screenshot shows the Microsoft Azure portal interface. It's a deployment overview for a Data Factory named 'Microsoft.DataFactory-20250218230351'. The deployment status is 'Your deployment is complete' with a green checkmark. Key details include the deployment name, start time (2/18/2025, 11:08:26 PM), subscription information (Azure subscription 1), and resource group ('rg-de-databricks-projects'). Below this, there's a table showing the deployment details for the 'data-factory-datab' resource, which is a Data factory (V2) in OK status. There are also sections for 'Deployment details' and 'Next steps' with a 'Go to resource' button.

After a successful deployment, we can verify the resource under the RG.

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

Home > Microsoft.DataFactory-20250218230351 | Overview

rg-de-databricks-projects

Resource group

Overview

Essentials

Resources Recommendations

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

Show 1 to 3 of 3 records. □ Show hidden types

Name	Type	Location
adigenzdemodbs	Storage account	West US 2
data-factory-databricks-demo-project	Data factory (V2)	West US 2
synapse-workspace-demo-dbs-01	Synapse workspace	West US 2

The screenshot shows the Microsoft Azure portal interface, specifically the 'rg-de-databricks-projects' resource group overview. It lists three resources: 'adigenzdemodbs' (Storage account, West US 2), 'data-factory-databricks-demo-project' (Data factory (V2), West US 2), and 'synapse-workspace-demo-dbs-01' (Synapse workspace, West US 2). The interface includes standard Azure navigation and search bars at the top.

6. Create Azure Databricks Workspace

Find Databricks in Azure Market Place and select create button.

The screenshot shows the Microsoft Azure Marketplace search results for 'databricks'. The search bar at the top contains 'databricks'. Below it, there are filters: 'Pricing : All' and 'Operating System : All', and a checkbox for 'Azure services only'. A button 'View suggestions' is also present. The results section shows 20 items out of 70, with three items displayed:

- Azure Databricks** (Microsoft, Azure Service, Azure Application) - Starts at Free.
- Lakehouse Optimizer for Azure Databricks** (Blueprint Technologies LLC, Azure Application) - Provides near realtime optimization and granular utilization and performance transparency.
- Access Connector for Azure Databricks** (Microsoft, Azure Service) - Leverages managed identities in Unity Catalog.

The sidebar on the left includes sections for Get Started, Management, My Marketplace, and Categories.

Basic Page

We need to use Azure Pay As You Go subscription so we can select Premium (+ Role-based Access Control). Without Premium plan, we won't be able to use Unity Catalog

This screenshot shows the 'Create an Azure Databricks workspace' wizard. The current step is 'Project Details'.

Project Details

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

Subscription * (dropdown): Azure subscription 1

Resource group * (dropdown): rg-de-databricks-projects

Instance Details

Workspace name * (input): dbs-demo-ws-01

Region * (dropdown): West US 2

Pricing Tier * (dropdown): Premium (+ Role-based access controls)

A tooltip for the Pricing Tier dropdown states: 'We selected the recommended pricing tier for your workspace. You can change the tier based on your needs.'

Managed Resource Group name (input): Enter name for managed resource group

At the bottom are buttons: 'Review + create', '< Previous', and 'Next : Networking >'.

Networking (for enterprises and organizations – use virtual networks (VNets) with no public IP configurations).

The screenshot shows the Microsoft Azure portal interface for creating a workspace. The top navigation bar includes 'Microsoft Azure' and a search bar. Below the navigation is a breadcrumb trail: Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >. The main title is 'Create an Azure Databricks workspace'. A horizontal navigation bar below the title has tabs: Basics, **Networking**, Encryption, Security & compliance, Tags, and Review + create. The 'Networking' tab is selected. Under the 'Networking' tab, there are two sections: 'Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)' and 'Deploy Azure Databricks workspace in your own Virtual Network (VNet)'. Both sections have 'Yes' and 'No' radio button options, with 'No' selected in both cases.

Encryption Page

The screenshot shows the Microsoft Azure portal interface for creating a workspace, specifically on the 'Encryption' page. The top navigation bar includes 'Microsoft Azure' and a search bar. Below the navigation is a breadcrumb trail: Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >. The main title is 'Create an Azure Databricks workspace'. A horizontal navigation bar below the title has tabs: Basics, Networking, **Encryption**, Security & compliance, Tags, and Review + create. The 'Encryption' tab is selected. The page contains several sections: 'Data Encryption' (with a note about customer-managed key encryption), 'Managed Disks' (with a checkbox for 'Use your own key'), 'Managed Services' (with a checkbox for 'Use your own key'), 'Double encryption for DBFS root' (with a note about infrastructure encryption and a checkbox for 'Enable Infrastructure Encryption'), and a note stating '⚠️This feature cannot be changed after this workspace is created.' At the bottom, there are navigation buttons: 'Review + create' (highlighted in blue), '< Previous', and 'Next : Security & compliance >'.

Security and Compliance Page

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

Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >

Create an Azure Databricks workspace

Basics Networking Encryption **Security & compliance** Tags Review + create

Enhanced Security & Compliance

Enhanced Security and Compliance Add-On helps simplify the complexity of meeting security and regulatory requirements.

Enable compliance security profile

⚠ This feature cannot be disabled once it is enabled.

Enable enhanced security monitoring

Enable automatic cluster update

Review + create < Previous Next : Tags >

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

Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >

Create an Azure Databricks workspace

Basics Networking Encryption Security & compliance **Tags** Review + create

Name	Value	Resource
<input type="text"/>	:	Azure Databricks Service

Review and Create the Databricks resource.



Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >

Create an Azure Databricks workspace

Validation Succeeded

Basics Networking Encryption Security & compliance Tags **Review + create**

Summary

Basics

Workspace name	dbs-demo-ws-01
Subscription	Azure subscription 1
Resource group	rg-de-databricks-projects
Region	West US 2
Pricing Tier	premium
Managed Resource Group name	

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

Encryption

Create

< Previous

Download a template for automation



Home > Microsoft.DataFactory-20250218230351 | Overview > rg-de-databricks-projects > Marketplace > Azure Databricks >

Create an Azure Databricks workspace

Validation Succeeded

Managed Resource Group name

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

Encryption

Enable Infrastructure Encryption	No
Enable CMK for Managed Disks	No
Enable CMK for Managed Services	No

Security & compliance

Compliance Security Profile	No
Compliance Standards	
Enhanced Security Monitoring	No
Automatic Cluster Update	No

Create

< Previous

Download a template for automation

Microsoft Azure | rg-de-databricks-projects_dbs-demo-ws-01 | Overview

Deployment name: rg-de-databricks-projects_dbs-demo-ws-01 | Start time: 2/18/2025, 11:16:45 PM | Correlation ID: e1237326-a6ae-4a78-a072-9c92bfad690d

Deployment details

Resource	Type	Status	Operation details
dbs-demo-ws-01	Azure Databricks Service	Created	Operation details

Your deployment is complete

Deployment name: rg-de-databricks-projects_dbs-demo-ws-01 | Start time: 2/18/2025, 11:16:45 PM | Correlation ID: e1237326-a6ae-4a78-a072-9c92bfad690d

Deployment details

Resource	Type	Status	Operation details
dbs-demo-ws-01	Azure Databricks Service	OK	Operation details

Next steps

[Go to resource](#)

Give feedback | Tell us about your experience with deployment

After a successful deployment, we should be able to see the resource. We can launch workspace by clicking the Launch button or selecting the URL

dbs-demo-ws-01 | Azure Databricks Service

Essentials

Status: Active	Managed Resource Group: databricks-rg-dbs-demo-ws-01-lonsc2x6c5t2g
Resource group: rg-de-databricks-projects	URL: https://adb-2903193773461431.11.core.databricks.net
Location: West US 2	Pricing Tier: Premium Y4
Subscription: Azure subscription 1	Role-based access controls (Click to change)
Subscription ID: 46ccf490-21e1-4902-a95c-39531f0ed8ae	
Tags (edit): Add tags	

Launch Workspace

Name	Type	Location
adlsgen2demodbs	Storage account	West US 2
data-factory-databricks-demo-project	Data factory (V2)	West US 2
dbs-demo-ws-01	Azure Databricks Service	West US 2
synapse-workspace-demo-dbs-01	Synapse workspace	West US 2

7. Azure Key Vault Creation

It is a cloud service used to securely store and manage sensitive information such as keys, secrets, and certificates. Its primary use case is to protect access to applications, services, and IT resources by centralizing credential management and ensuring secure access through role-based permissions.

Offer	Publisher	Description
Key Vault	Microsoft	Azure Service Safeguard cryptographic keys and other secrets used by cloud apps and services.
Azure Key Vault solution for Sentinel	Microsoft Sentinel, Microsoft Co...	Azure Application Azure Key Vault solution for Sentinel
Azure Key Vault Managed HSM	Microsoft	Azure Service Safeguard cryptographic keys used by cloud apps and services.
EJBCA SaaS - With Key Vault Backed CA Keys	Keyfactor, Inc.	SaaS EJBCA SaaS, PKI delivered as a service with Azure Key Vault key storage

Basic Page

← → ⌂ portal.azure.com/#create/Microsoft.KeyVault

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

Home > rg-de-databricks-projects_dbs-demo-ws-01 | Overview > rg-de-databricks-projects > Marketplace >

Create a key vault

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 subscription 1
Resource group *	rg-de-databricks-projects
	Create new

Instance details

Key vault name *	kv-dbs-demo-01
Region *	West US 2
Pricing tier *	Standard

Recovery options

Soft delete protection will automatically be enabled on this key vault. This feature allows you to recover or permanently delete a key vault and secrets for the duration of the retention period. This protection applies to the key vault and the secrets stored within the key vault.

To enforce a mandatory retention period and prevent the permanent deletion of key vaults or secrets prior to the retention period elapsing, you can turn on purge protection. When purge protection is enabled, secrets cannot be purged by users or by Microsoft.

Previous Next Review + create

Configure the options as per the below screenshot. (Note: In this course we are going with the "Vault Access Policy" Permission model, however it's not recommended currently).

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

Home > rg-de-databricks-projects_dbs-demo-ws-01 | Overview > rg-de-databricks-projects > Marketplace >

Create a key vault

Permission model

Grant data plane access by using a Azure RBAC or Key Vault access policy

Azure role-based access control (recommended)

Vault access policy

Resource access

Azure Virtual Machines for deployment

Azure Resource Manager for template deployment

Azure Disk Encryption for volume encryption

Access policies

Access policies enable you to have fine grained control over access to vault items. [Learn more](#)

Name	Email	Key Permissions	Secret Permissions	Certificate Permissions
Abhijit Das	[REDACTED]@ab...com#EXT#@ab...	Get, List, Update, Create, Import, Delete, R...	Get, List, Set, Delete, Recover, Backup, Res...	Get, List, Update, Create, Import, Delete, ...

Previous Next Review + create Give feedback

← → G portal.azure.com/#create/Microsoft.KeyVault

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

Home > rg-de-databricks-projects_dbs-demo-ws-01 | Overview > rg-de-databricks-projects > Marketplace >

Create a key vault ...

Basics Access configuration Networking Tags Review + create

You can connect to this key vault either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Enable public access

Public Access

Allow access from: All networks Selected networks

ⓘ Traffic from all public networks can access this resource. This is not recommended for private applications or environments. [Learn more](#)

Private endpoint

Create a private endpoint to allow a private connection to this resource. Additional private endpoint connections can be created within the key vault or private link.

+ Create a private endpoint

Name	Subscription	Resource group	Region	Subnet	Private DNS Zone
Click on add button to add private endpoint					

Previous Next Review + create

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

Home > rg-de-databricks-projects_dbs-demo-ws-01 | Overview > rg-de-databricks-projects > Marketplace >

Create a key vault ...

Basics Access configuration Networking Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups.

Name ⓘ	Value ⓘ	Resource
<input type="text"/>	:	<input type="text"/> Key vault

Previous Next Review + create

Create a key vault

[Basics](#) [Access configuration](#) [Networking](#) [Tags](#) [Review + create](#)[View Automation Template](#)

Basics

Subscription	Azure subscription 1
Resource group	rg-de-databricks-projects
Key vault name	kv-dbs-demo-01
Region	West US 2
Pricing tier	Standard
Soft-delete	Enabled
Purge protection during retention period	Disabled
Days to retain deleted vaults	90 days

Access configuration

Azure Virtual Machines for deployment	Disabled
Azure Resource Manager for template deployment	Disabled
Azure Disk Encryption for volume encryption	Disabled
Permission model	Vault access policy

[Previous](#)[Next](#)[Create](#)

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

Home > kv-dbs-demo-01 | Overview Deployment

Overview

Your deployment is complete

Deployment name : kv-dbs-demo-01 Start time : 2/18/2025, 11:30:25 PM

Subscription : Azure subscription 1 Correlation ID : 61e040ae-76c1-44c7-87ee-8617679b6ba3

Resource group : rg-de-databricks-projects

Deployment details

Resource	Type	Status	Operation details
kv-dbs-demo-01	Key vault	OK	Operation details

Next steps

[Go to resource](#)

Microsoft Azure Search resources, services, and docs (G+) abhijitaverc2023@gmail.com

Home > Resource groups >

Resource groups Default Directory

rg-de-databricks-projects

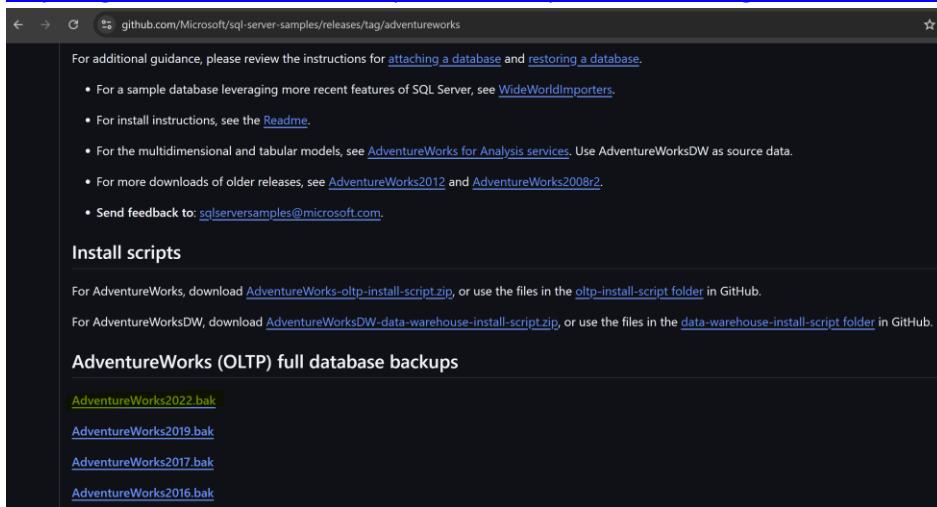
Overview

Resources Recommendations

Showing 1 to 5 of 5 records.

Name	Type	Location
adigen2demodbs	Storage account	West US 2
data-factory-databricks-demo-project	Data factory (V2)	West US 2
dbs-demo-ws-01	Azure Databricks Service	West US 2
kv-dbs-demo-01	Key vault	West US 2
synapse-workspace-demo-dbs-01	Synapse workspace	West US 2

8. It is an additional step to load a sample database in SSMS. Setting up data source - <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>



The screenshot shows a web browser window with the URL <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>. The page content includes:

- For additional guidance, please review the instructions for [attaching a database](#) and [restoring a database](#).
- For a sample database leveraging more recent features of SQL Server, see [WideWorldImporters](#).
- For install instructions, see the [Readme](#).
- For the multidimensional and tabular models, see [AdventureWorks for Analysis services](#). Use AdventureWorksDW as source data.
- For more downloads of older releases, see [AdventureWorks2012](#) and [AdventureWorks2008r2](#).
- Send feedback to: sqldatabasesamples@microsoft.com.

Install scripts

For AdventureWorks, download [AdventureWorks-oltp-install-script.zip](#), or use the files in the [oltp-install-script folder](#) in GitHub.

For AdventureWorksDW, download [AdventureWorksDW-data-warehouse-install-script.zip](#), or use the files in the [data-warehouse-install-script folder](#) in GitHub.

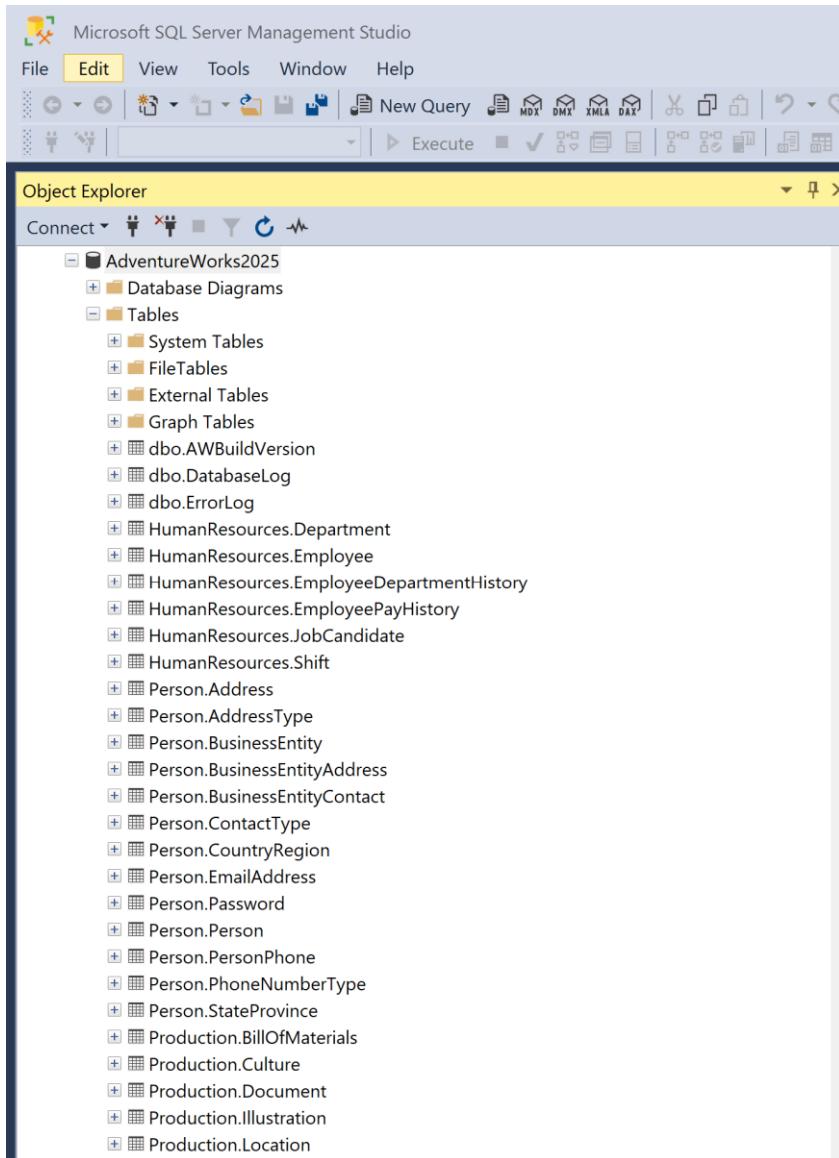
AdventureWorks (OLTP) full database backups

[AdventureWorks2022.bak](#)

[AdventureWorks2019.bak](#)

[AdventureWorks2017.bak](#)

[AdventureWorks2016.bak](#)



This user login will be used to setup SHIR in ADF. So we will be able to read on premise data and load to ADLS Gen2.

```
create login dbstest with password = 'P@ssw0rd';
create user dbstest for login dbstest;
```

The screenshot shows a Microsoft SQL Server Management Studio window with a query editor tab titled "SQLQuery1.sql". The code in the editor is:

```
create login dbstest with password = 'P@ssw0rd';
create user dbstest for login dbstest;
```

The "Object Explorer" pane on the left shows the database structure, including the "AdventureWorks2025" database and its tables.

```

SQLQuery1.sql - AB...IJTDAS\abhij (70)*  ×
[ create login dbstest with password = 'Test123abc'
[ create user dbstest for login dbstest;

```

177 % < Messages

Commands completed successfully.

Completion time: 2025-02-19T20:36:06.4285206-08:00

Microsoft Azure Search resources, services, and docs (G+) Copilot abhijitawscert2023@gmail.com DEFAULT DIRECTORY

Home > Resource groups > rg-de-databricks-projects > kv-dbs-demo-01

kv-dbs-demo-01 | Secrets ⋮

Key vault

Search Generate/Import Refresh Restore Backup View sample code

The secret 'password' has been successfully created.

Name	Type	Status	Expiration date
password		✓ Enabled	
user		✓ Enabled	

Access control (IAM) Tags Diagnose and solve problems Access policies Events Objects Keys Certificates

9. Data Ingestion Phase

First, we will launch ADF.

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

Home > Resource groups > rg-de-databricks-projects > data-factory-databricks-demo-project ⋮

Data factory (V2)

Search Delete

Overview

Activity log Status : Succeeded

Access control (IAM) Location : West US 2

Tags Subscription (move) : Azure subscription_1

Diagnose and solve problems Subscription ID : 46ccf490-21e1-4902-a95c-395310ed8ae

Essentials

Resource group (move) : rg-de-databricks-projects

Type : Data factory (V2)

Getting started : Quick start

Azure Data Factory Studio

Launch studio

Quick Starts Tutorials Template Gallery Training Modules

Creating a SHIR to connect to On Premise SQL Server (select Azure, Self-Hosted)

Integration runtime setup

Integration Runtime is the native compute used to execute or dispatch activities. Choose what integration runtime to create based on required capabilities. [Learn more](#)

Name	Type	Sub-type
AutoResolveIntegrationRuntime	Azure	Public

Select Self-Hosted

Integration runtime setup

Network environment:
Choose the network environment of the data source / destination or external compute to which the integration runtime will connect to for data flows, data movement or dispatch activities:

Azure Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.
Self-Hosted Use this for running activities in an on-premises / private network View more

External Resources:
You can use an existing self-hosted integration runtime that exists in another resource. This way you can reuse your existing infrastructure where self-hosted integration runtime is set up.

Linked Self-Hosted Learn more

Provide a name

Integration runtime setup

Private network support is realized by installing integration runtime to machines in the same on-premises network/VNET as the resource the integration runtime is connecting to. Follow below steps to register and install integration runtime on your self-hosted machines.

Name: SHR

Description: used to connect SQL Server

Type: Self-Hosted

Use Express setup for local computer

Integration runtime setup

Install integration runtime on Windows machine or add further nodes using the Authentication Key.

Name: SHR

Option 1: Express setup
[Click here to launch the express setup for this computer](#)

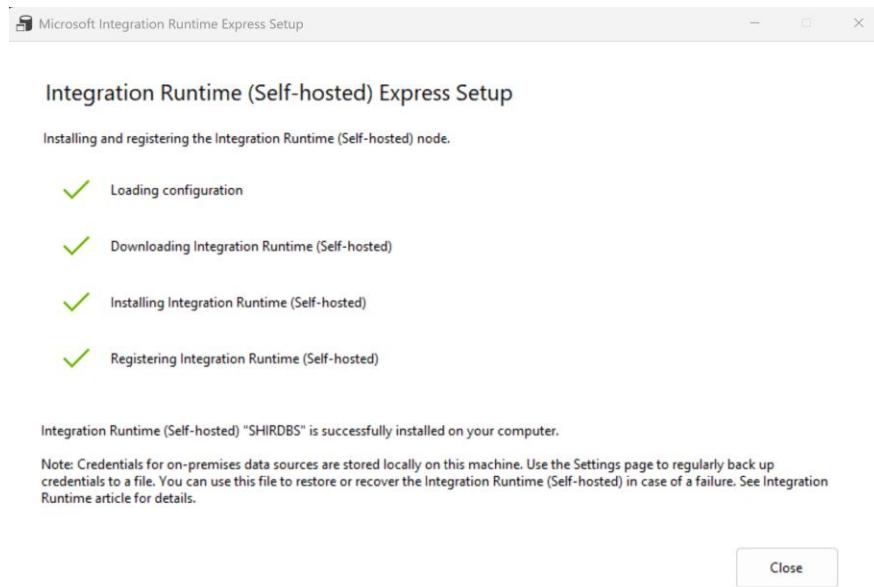
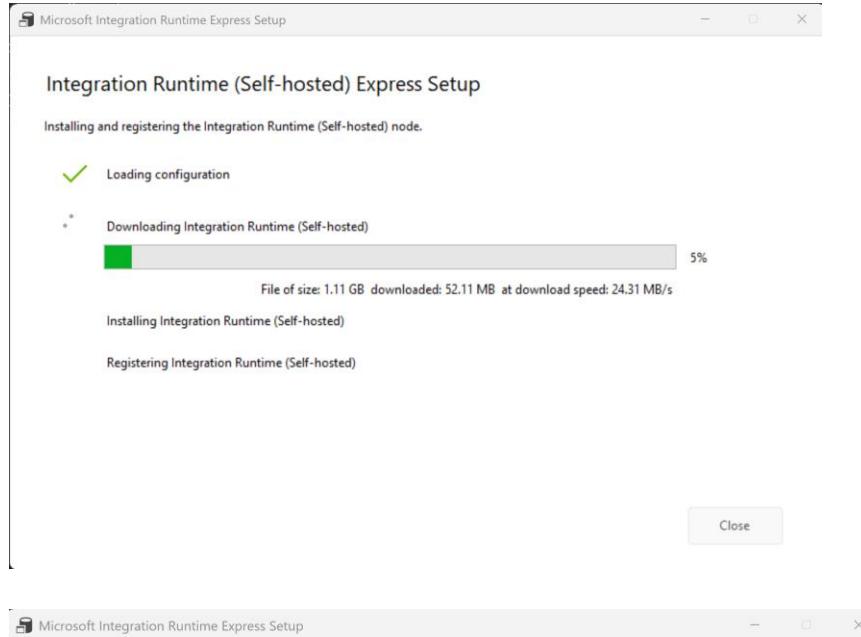
Option 2: Manual setup

Step 1: Download and install integration runtime

Step 2: Use this key to register your integration runtime

Name	Authentication key
Key1	@data-factory-databricks
Key2	@data-factory-databricks

Integration Runtime (Self-hosted) Express Setup will run in your local. We should be using either express or manual setup based on our client requirement.



After a success installation, we should be able to see the Self Hosted Integration Runtime under Integration Runtime. And it is ready to use.

The screenshot shows the Microsoft Azure Data Factory portal. The left sidebar shows navigation options: General, Factory settings, Connections, Linked services, Integration runtimes (which is selected and highlighted in yellow), Microsoft Purview, Source control, Git configuration, and ARM template. The main content area is titled "Integration runtimes" and contains the following information:

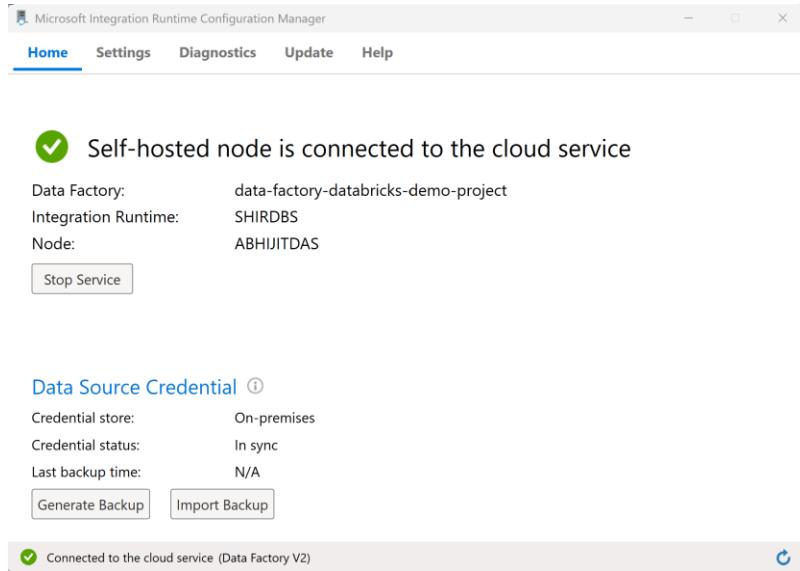
The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment. Learn more [\(link\)](#)

+ New Refresh

Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	0	Auto Resolve	---
SHIRDBS	Self-Hosted	---	Running	0	---	5.48.9076.1



Create an ADF pipeline

Add Copy Data activity for copying on premise data

The first screenshot shows the "New dataset" dialog where "sql server" is selected as the data store. The second screenshot shows the "Set properties" dialog for a dataset named "SQLTable" with "Linked service" set to "Select...".

10. Launch Azure Databricks workspace and use it:

Create Databricks Clusters:

11. Key Vault (I won't use key vault in this demo to run the ADF and Databricks, it is only for the information)

12. ADLS Storage Creation For source data store (this will be our source for ETL)

Microsoft Azure

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking Data protection Encryption Tags Review + create

[View automation template](#)

Basics

Subscription	Azure subscription 1
Resource group	rg-de-databricks-projects
Location	West US 2
Storage account name	rawsrstorage
Primary service	Azure Blob Storage or Azure Data Lake Storage Gen 2
Performance	Standard
Replication	Locally-redundant storage (LRS)

Advanced

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

[Previous](#)

[Next](#)

[Create](#)

Microsoft Azure Copilot

Home > Storage accounts >

Create a storage account ...

Network connectivity	Public endpoint (all networks)
Default routing tier	Microsoft network routing

Data protection

Point-in-time restore	Disabled
Blob soft delete	Enabled
Blob retention period in days	7
Container soft delete	Enabled
Container retention period in days	7
File share soft delete	Enabled
File share retention period in days	7
Versionsing	Disabled
Blob change feed	Disabled
Version-level immutability support	Disabled

Encryption

Encryption type	Microsoft-managed keys (MMK)
Enable support for customer-managed keys	Blobs and files only
Enable infrastructure encryption	Disabled

[Previous](#)

[Next](#)

[Create](#)

The screenshot shows the Microsoft Azure Deployment Overview page. At the top, it displays the deployment name: **rawsrcstorage_1740091565208** | Overview. Below the title, it says "Your deployment is complete". It provides deployment details: Deployment name: rawsrcstorage_1740091565208, Subscription: Azure subscription 1, Resource group: rg-de-databricks-projects. The status is listed as "OK" for three resources: rawsrcstorage/default, rawsrcstorage/default, and rawsrcstorage. There is a "Next steps" section with a "Go to resource" button.

We will be processing Person Details in our ETL pipeline. Uploading person_raw.csv file.

The screenshot shows the Microsoft Azure Storage Container Overview page for the 'person-details' container. It lists a single blob named 'person_raw.csv'. The blob details are: Name: person_raw.csv, Modified: 2/20/2025, 2:49:53 PM, Access tier: Hot (Inferred), Archive status: Not yet archived, Blob type: Block blob, Size: 12.71 MB, Lease state: Available.

13. ADF Pipeline Creation.

a. Linked Service Creation

Creating ADLS linked services to read and write data from source and target.

Edit linked service

Azure Data Lake Storage Gen2 [Learn more](#)

Name *
ADLStorageDB

Description

Connect via integration runtime * ⓘ
 AutoResolveIntegrationRuntime

Authentication type

Account selection method ⓘ
 From Azure subscription Enter manually

URL *
https://rawsrcstorage.dfs.core.windows.net/

Storage account key *
.....

Test connection ⓘ

Connection successful

adf.azure.com/en/authoring/pipeline/pipeline1?factory=%2Fsubscriptions%2F46ccf490-21e1-4902-a95c-39531f0ed8ae%2FresourceGroups%2Frg-de-databricks... [Incognito](#)

Microsoft Azure | Data Factory > data-factory-databricks-demo-project | Search factory and documentation

Factory Resources

- Pipelines
- copydata-pipeline
- Change Data Capture (preview)
- Datasets
- Data flows
- Power Query

copydata-pipeline Activities

- Copy data
- Move and transform

copy data

General Source Sink

Sink dataset *

New linked service

Azure Data Lake Storage Gen2 [Learn more](#)

Name *

Description

Connect via integration runtime * ⓘ
 AutoResolveIntegrationRuntime

Authentication type

Account selection method ⓘ
 From Azure subscription Enter manually

Azure subscription ⓘ
Azure subscription 1 (46ccf490-21e1-4902-a95c-39531f0ed8ae)

Storage account name *

Test connection ⓘ
 To linked service To file path

Connection successful

[Test connection](#)

Databricks Linked Service

Access tokens

Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords.

Comment	Creation	Expiration
ADF	2025-02-20 23:45:42 PST	2025-02-22 00:45:42 PDT

New linked service

From Azure subscription Enter manually

Azure subscription *

Databricks workspace *

Select cluster New job cluster Existing interactive cluster Existing instance pool

Databrick Workspace URL

Authentication type * Access Token Azure Key Vault

Access token *

Cluster version *

Cluster node type *

Create Cancel Test connection Connection successful

New linked service

From Azure subscription Enter manually

Name *

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Account selection method * From Azure subscription Enter manually

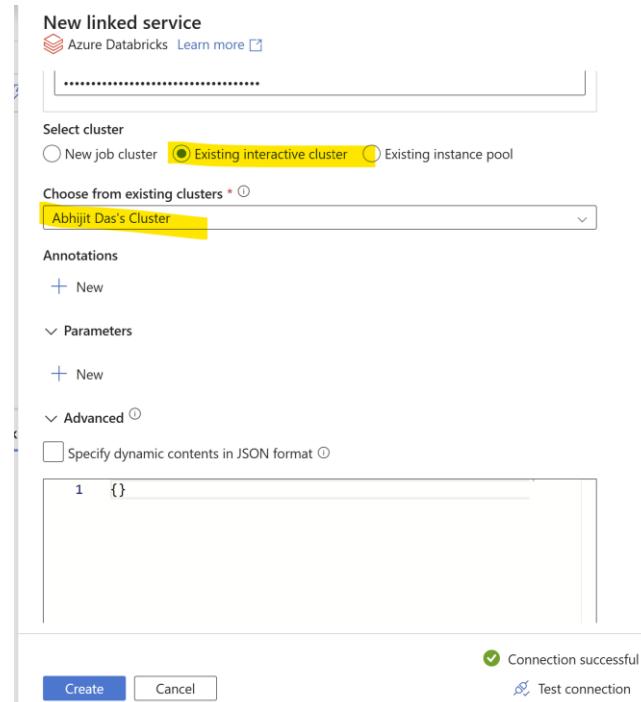
Databrick Workspace URL *

Authentication type * Access Token Azure Key Vault

Access token *

Select cluster

Create Cancel Test connection Connection successful



b. Creating Data Pipelines and Datasets

copydata-bronze-pipeline creation (extracting person, email, phone and phone types details to bronze layer)

Update name of the pipeline in General tab.

Update source details and include dataset with Parameters.

The screenshot shows the Azure Data Factory Copy Activity Editor. At the top, there are four parallel copy tasks:

- copy-person-adls**: Source dataset is `person`.
- phonenumbertype-adls copy**: Source dataset is `phonenumbertype`.
- email-adls_copy**: Source dataset is `email`.
- phone-adls_copy**: Source dataset is `phone`.

The **Source** tab is selected. Under **Source dataset ***, `person` is chosen. Below it, the **Dataset properties** table contains:

Name	Value	Type
pfolder	person	string
pfilename	person_raw.csv	string

Datasets

The screenshot shows the Azure Data Factory Dataset Properties page for the `person` dataset. It has tabs for **Connection**, **Schema**, and **Parameters**. The **Connection** tab is selected.

Connection settings:

- Linked service ***: ADLStorageDB
- File path**: `person-details` / `@dataset().pfolder` / `@dataset().pfilename`
- Compression type**: No compression
- Column delimiter**: Comma (,)
- Row delimiter**: Default (\r,\n, or \r\n)
- Encoding**: Default(UTF-8)
- Quote character**: Double quote ("")
- Escape character**: Backslash (\)

Parameters under datasets (so we can reuse the same linked services).

The screenshot shows the Azure Data Factory Dataset Parameters page for the `person` dataset. It has tabs for **Connection**, **Schema**, and **Parameters**. The **Parameters** tab is selected.

Parameters table:

Name	Type	Default value
pfolder	String	person
pfilename	String	person_raw.csv

Same steps will be repeated for other three datasets.

Databricks Data pipeline

dbs_bronze_silver and dbs_silver_gold pipelines will be created, once we will write those notebooks in Databricks Workspace.

The screenshot shows the Azure Data Factory Pipeline Editor interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (containing 'copydata-bronze-pipeline', 'dbs_bronze_silver' (selected), and 'dbs_silver_gold' (highlighted)), and 'Datasets' (containing 'email', 'emailaddressbrz', 'person', and 'personbrz'). The main workspace displays the 'Activities' pane, which includes options like 'Move and transform', 'Synapse', 'Azure Data Explorer', etc. A 'Notebook' section on the right shows a single entry: 'person'. The bottom navigation bar includes tabs for 'Parameters', 'Variables', 'Settings', and 'Output', with a 'New' button.

Before writing notebooks, we need to create Databricks Clusters.

14. Interactive Cluster creation for ADF data pipeline (can use job cluster as well).

The screenshot shows the Databricks Compute blade. The left sidebar includes 'Compute', 'SQL', 'Data Engineering', 'Job Runs', 'Data Ingestion', 'Pipelines', 'Machine Learning', and 'Playground'. The main area is titled 'Abhijit Das's Cluster' and contains configuration sections for 'Policy' (set to 'Unrestricted'), 'Access mode' (set to 'Single user access' with 'Dedicated (formerly: Single user)' selected), 'Performance' (set to 'Runtime: 15.4 LTS (Scala 2.12, Spark 3.5.0)'), 'Worker type' (set to 'Standard_D4ds_v5'), 'Workers' (set to '4'), and 'Driver type' (set to 'Same as worker'). A note at the bottom states: '⚠️ This account may not have enough CPU cores to satisfy this request. Estimated available: 0, requested: 20. Learn more about CPU quota.' At the bottom are 'Create compute' and 'Cancel' buttons.

For setting the mount-point – we will check the credential pass through option. If you will uncheck credential pass through, then we should be able to use default Unity Catalog.

Compute > New compute > Simple form: OFF ▾

Abhijit Das's Cluster

Advanced options

Azure Data Lake Storage credential passthrough ⓘ

Enable credential passthrough for user-level data access

⚠ Credential passthrough is deprecated starting with Databricks Runtime 15.0 and will be removed for future Databricks Runtime versions. [Learn more](#)

Spark Logging Init Scripts

Spark config ⓘ

Enter your Spark configuration options here. Provide only one key-value pair per line.
Example:
spark.sql.ansi.enabled true
spark.sql.files.maxPartitionBytes 134217728

Environment variables ⓘ

PYSPARK_PYTHON=/databricks/python3/bin/python3

Create compute Cancel

15. Setup a mount points to read data from ADLS paths

```
storagemount Python ⚡
File Edit View Run Help Last edit was 11 hours ago
Run all Terminated Schedule Share
ADLS Storage Mount on DBFS
Markdown
config = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

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

3 days ago (1h) 3
dbutils.fs.ls("/mnt/bronze/person/")
[FileInfo(path=dbfs:/mnt/bronze/person/person , name='person' , size=1029441, modificationTime=1740121732000)]

16. Create notebooks to transform bronze data to silver to gold.

```
bronze-to-silver Python ⚡
File Edit View Run Help Last edit was 23 hours ago
Run all Terminated Schedule Share
Bronze To Silver Data Transformation For Person Details
Markdown
from pyspark.sql.functions import lit, current_timestamp
from pyspark.sql.types import *

def bronzeToSilverDeltaTBLUWithDatabase(
    database_name: str,
    table_name: str,
    parquet_file_path: str,
    location: str,
    column_schemas: dict,
    source_to_target_mapping: dict = None
):
```

silver-to-gold Python

File Edit View Run Help Last edit was 23 hours ago

Run all Terminated Schedule (1) Share

Silver to Gold Data Transformation For Person Details

Markdown

```
%sql
CREATE TABLE IF NOT EXISTS person_db.person_gold
USING delta
LOCATION '/mnt/gold/person/'
AS
SELECT
    per.BusinessEntityID,
    per.PersonType,
    per.Title,
    per.FirstName,
    per.MiddleName,
    per.LastName,
    eml.EmailAddress,
    eml.EmailPromotion,
    phe.PhoneNumber
FROM person_db.person per
JOIN person_db.email eml
ON per.BusinessEntityID = eml.BusinessEntityID
JOIN person_db.phone phe
ON per.BusinessEntityID = phe.BusinessEntityID;
```

17. First, we will attach the Azure Databricks Linked Service that we created with Interactive Cluster

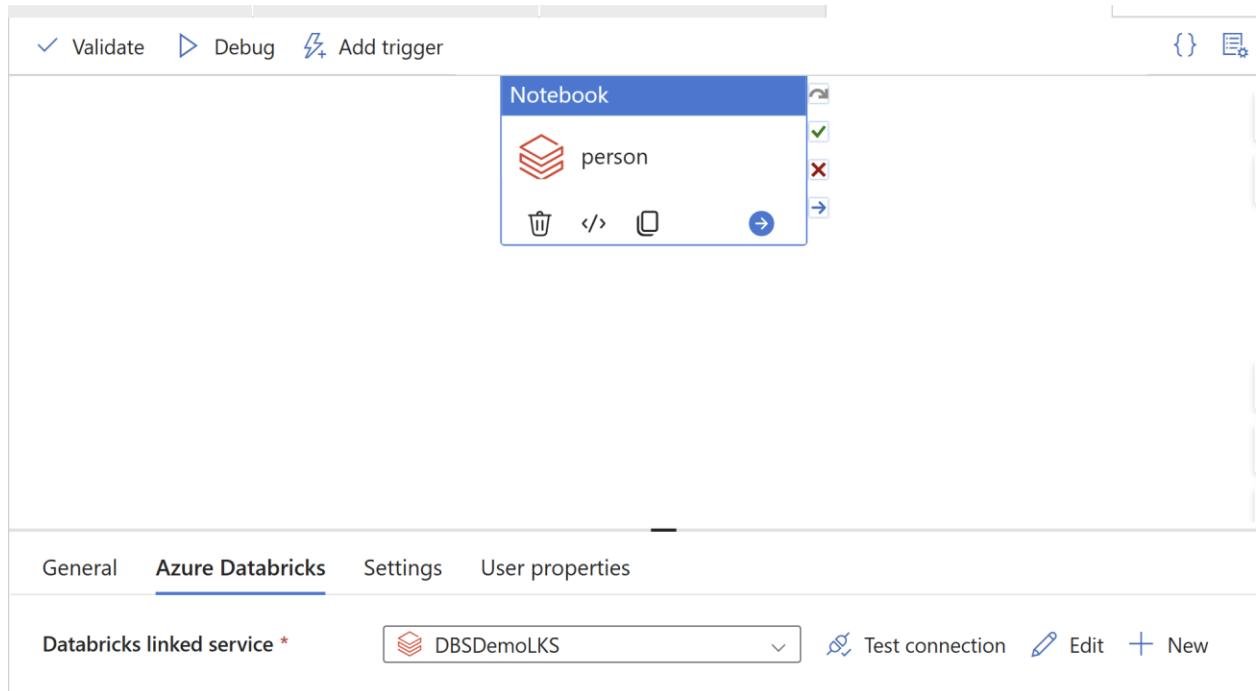
✓ Validate ⚡ Debug ⚡ Add trigger

Notebook

person

General Azure Databricks Settings User properties

Databricks linked service * DBSDemoLKS Test connection Edit New

The screenshot shows the Azure Data Factory interface. In the center, there's a 'Notebook' section containing a single cell named 'person'. Below this is a navigation bar with tabs: General, Azure Databricks (which is selected), Settings, and User properties. At the bottom, there's a search bar labeled 'Databricks linked service *' with a dropdown menu showing 'DBSDemoLKS', a 'Test connection' button, and 'Edit' and 'New' buttons.

Edit linked service

Azure Databricks [Learn more](#)

Name * DBSDemoLKS

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime

Account selection method *

From Azure subscription Enter manually

Databrick Workspace URL * ⓘ

https://adb-[REDACTED].azuredatabricks.net

Authentication type *

Access Token

Access token Azure Key Vault

Access token *

[REDACTED]

Select cluster

New job cluster Existing interactive cluster Existing instance pool

18. Now you can attach those notebooks to ADF pipelines.

✓ Validate ⚡ Debug ⏱ Add trigger

Notebook

person

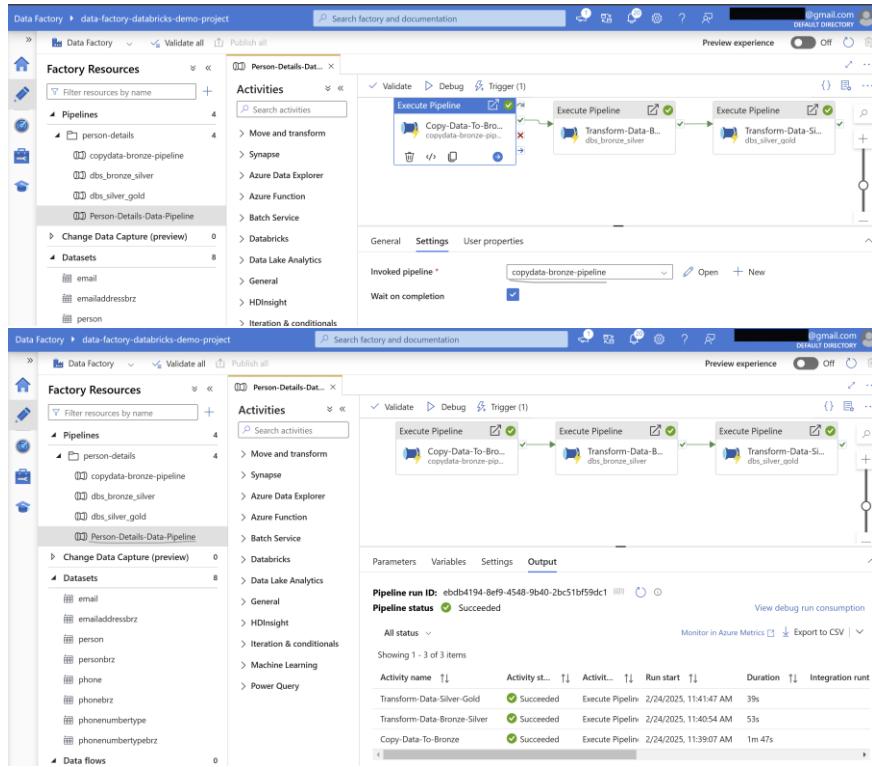
General Azure Databricks **Settings** User properties

Notebook path * /Shared/dbs-demo-project/silver-to-gold

> Base parameters

> Append libraries

19. Now will create a parent Pipeline to run those child pipeline in sequence.



20. Trigger pipeline from ADF or manually debug.

The screenshot shows the 'Add trigger' dialog for the 'Person-Details-Data-Pipeline'. The dialog fields are as follows:

- Name: person-datalinepipe-trig
- Description: To trigger person data pipeline
- Type: Schedule
- Start date: 2/23/2025, 12:10:24 PM
- Time zone: Pacific Time (US & Canada) (UTC-8)
- Recurrence: Every 1 Day
- Advanced recurrence options: Execute at these times, Hours, Minutes, Schedule execution times, 12:10, Specify an end date
- Annotations: New, Start trigger on creation (checked)

21. If you would like to skip ADF and only wants to use Databricks. We can use Databricks workflow for orchestration and scheduling.

It is preferred to use job-clusters. I am using interactive for this demo.

Workflows > Jobs >

New Job Feb 24, 2025, 12:13 PM ⭐

Runs Tasks

```
graph LR; A[persondata-bronze-silver<br/>...<br/>Abhijit Das's Cluster] --> B[persondata-silver-gold<br/>...<br/>Abhijit Das's Cluster]
```

+ Add task

Task name* ⓘ

Type* ⓘ

Source* ⓘ

Path* ⓘ

Cluster* ⓘ

Depends on

Run if dependencies ⓘ

Dependent libraries ⓘ

+ Add

Microsoft Azure databricks

Search data, notebooks, records, and more... (Ctrl + P)

Jobs

person-data-load

Runs Tasks

```
graph LR; A[persondata-bronze-silver<br/>...<br/>Abhijit Das's Cluster] --> B[persondata-silver-gold<br/>...<br/>Abhijit Das's Cluster]
```

+ Add task

Job details

- Job ID: 87009481099619
- Creator: Abhijit Das
- Run as: Abhijit Das
- Tags: Add tag
- Description: Add description
- Usage: No lineage information for this job. Learn more ↗

Git

not configured

Add Git settings

Schedules & Triggers

We can click on run-now to run the job and set trigger for scheduling

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar navigation includes 'New', 'Workspace', 'Recent', 'Catalog', 'Compute', 'Marketplace', 'SQL Editor', 'Overview', 'Dashboards', 'Metrics', 'Alerts', 'Query History', 'SQL Warehouses', 'Data Engineering', 'Job Runs', 'Data Injection', 'pipelines', 'Machine Learning', 'Playground', 'Experiments', 'Features', 'Models', and 'Services'. The main workspace displays a pipeline named 'person-data-load' under the 'Workflows' section. The pipeline has two runs: one from January 26, 2023, at 11:20 PM, which is 'Running' (Run ID: 81619176795582), and another from February 24, 2023, at 9:45 AM, which is 'Not configured'. A 'Runs' table provides a summary of these runs. On the right, there are sections for 'Job details', 'Git' (not configured), 'Schedules & Triggers' (None), and 'Compute' (Able to detect cluster). The top right corner features a 'Send feedback' button and a 'Help' link.

Workflows > Jobs >
person-data-load ☆

Runs Tasks

Runs

Start date: Feb 24 | < Previous | Next >

Run total duration: 59s

Feb 24

Tasks

persondata-bronze-silver

persondata-silver-gold

Go to the latest successful run

Cancel runs

Start time	Run ID	Launched	Duration	Status	Error code	Run parameters
Feb 24, 2025, 12:20 PM	81619765795582	Manually	59s	Succeeded		...

Schedules & Triggers

Trigger type: None (manual)

Scheduled

File arrival

Continuous

Cancel Save

Schedules & Triggers

Trigger Status: Active

Trigger type: Scheduled

Schedule type: Simple

Schedule: Every Day at 12 : 30 (UTC+00:00) UTC

Show cron syntax

Cancel Save

Workflows

Jobs Job runs Pipelines

Send feedback

Filter jobs Owned by me Accessible by me Favorites Tags

Load tutorial Create job

Name	Tags	Created by	Trigger	Recent runs
person-data-load		Abhijit Das	Scheduled	— — — (green)

Job details

- Job ID: 870090492999539
- Creator: Abhijit Das
- Run as: Abhijit Das
- Tags: Add tag
- Description: Add description
- Lineage: No lineage information for this job. Learn more

Git
Not configured
Add Git settings

Schedules & Triggers
At 12:30 PM (UTC+00:00 — UTC)
Edit trigger, Pause, Delete

Compute
Abhijit Das's Cluster
Single node: Standard_DS3_v2 - DBR: 15.4 LTS (includes Apache Spark 3.5.0, Scala 2.12)
View details, Swap, Spark UI, Logs, Metrics

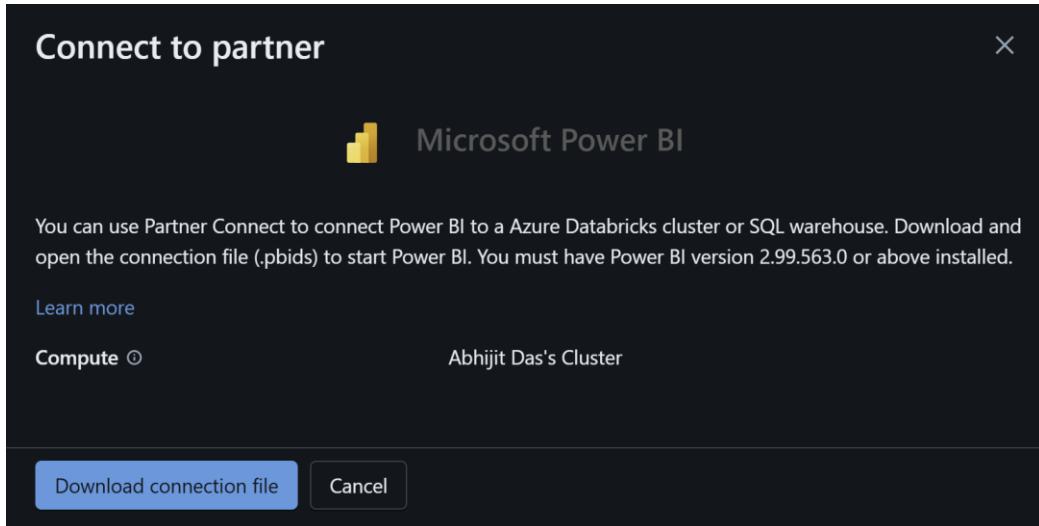
Job parameters
No job parameters are defined for this job

22. Integrate with Power BI and build a sample report.

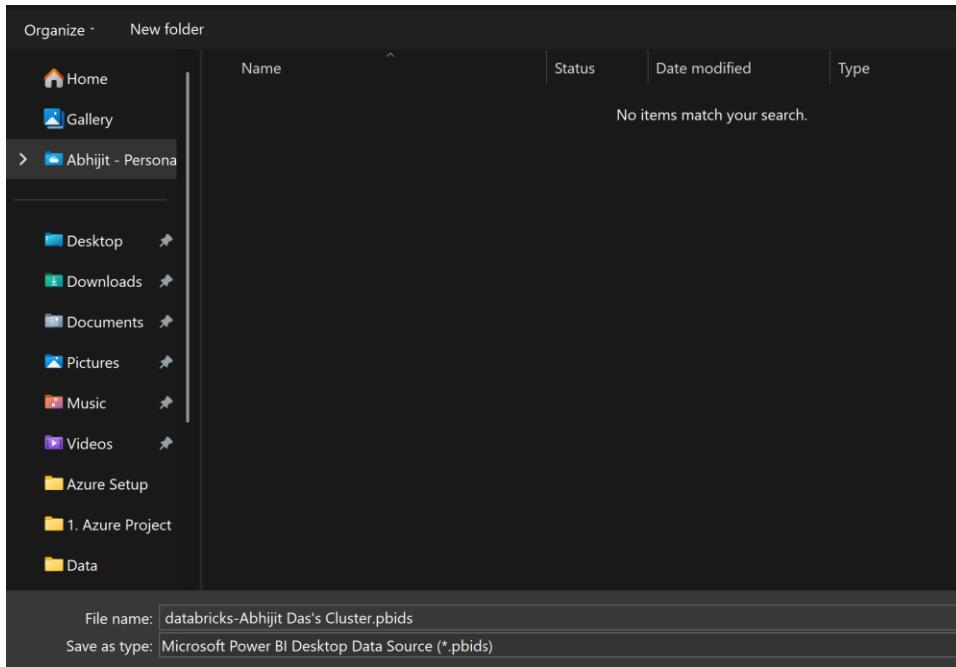
Navigate to Catalog -> Hive Metastore -> Click on Open in Power BI Desktop

Catalog Explorer > hive metastore > person_db > person_gold

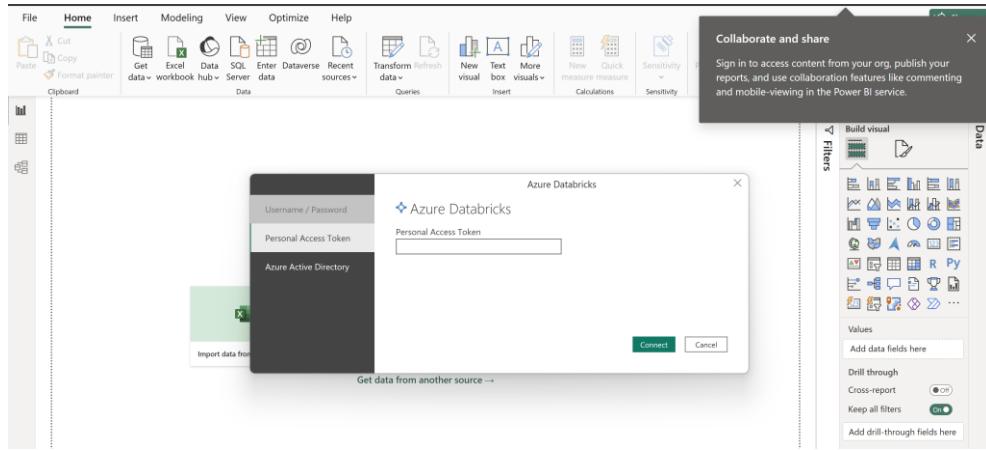
BusinessEntityID	PersonType	Title	FirstName	MiddleName	LastName
1	EM	NULL	Ken	J	Sánchez
2	EM	NULL	Terri	L	Duffy
3	EM	NULL	Roberto	NULL	Tamburello
4	EM	NULL	Rob	NULL	Walters
5	EM	Ms.	Gail	A	Erickson
6	EM	Mr.	Jossef	H	Goldberg
7	EM	NULL	Dylan	A	Miller
8	EM	NULL	Diane	L	Marghei
9	EM	NULL	Gigi	N	Matthew
10	EM	NULL	Michael	NULL	Rahiem
11	EM	NULL	Ovidiu	V	Craciun
12	EM	NULL	Thierry	B	D'Hers
13	EM	Ms.	Janice	M	Galvin
14	EM	NULL	Michael	I	Sullivan
15	EM	NULL	Sharon	B	Salavariz
16	EM	NULL	David	M	Bradley
17	EM	NULL	Kevin	F	Brown
18	EM	NULL	John	L	Wood



Save *.pbids file on your desktop



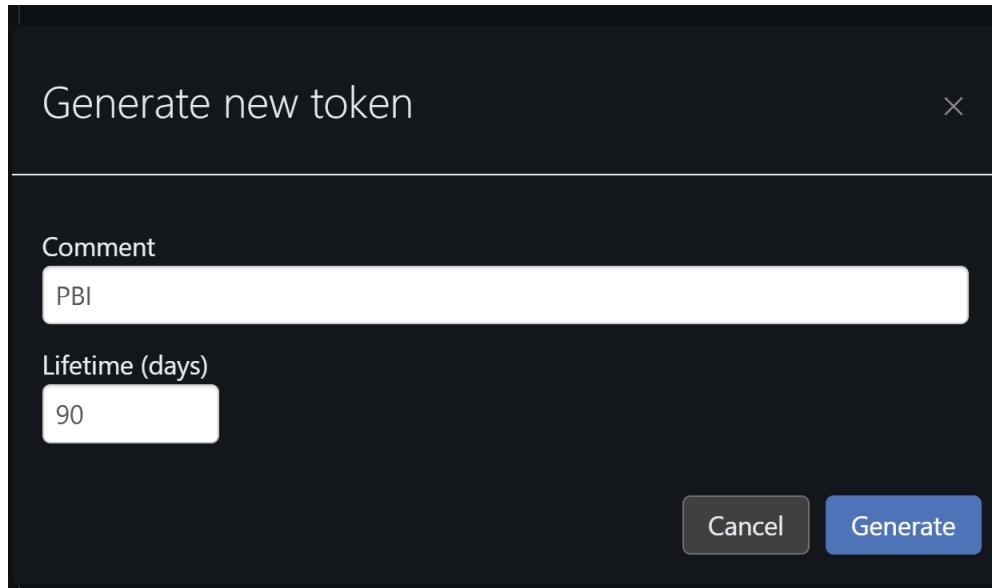
There are three different options to connect



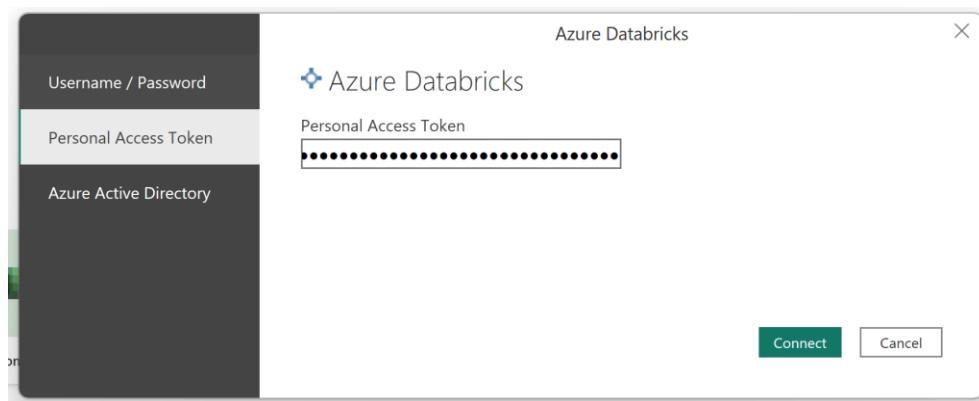
I will be using personal access token for easy access. We can use Entra-ID to connect. Or you can load the data to Synapse Analytics, SQL Server Managed Instance for presentation layer.

The screenshots show the Databricks developer settings page. The left sidebar includes options like Workspace, Catalog, Workflows, Compute, Marketplace, SQL, and SQL Editor. The main area shows 'Developer' settings under 'Settings'. The 'Access tokens' section contains a 'Manage' button. The bottom screenshot highlights the 'Generate new token' button, which is green, indicating it is the active or next step.

Generating Personal Access Token for Power BI.



Now we can use this token to connect to PBI.



Now you could see those hive metastore tables from Power BI

Navigator

Display Options

adb-1414194557308463.3.azuredatabricks.net...

- 'hive metastore'.'person_db'.'email'
- 'hive metastore'.'person_db'.'person_gold'
- 'hive metastore'.'person_db'.'person'
- 'hive metastore'.'person_db'.'phone'
- 'hive metastore'.'person_db'.'phonenumb...'

No items selected for preview

Load Cancel

Navigator

Display Options

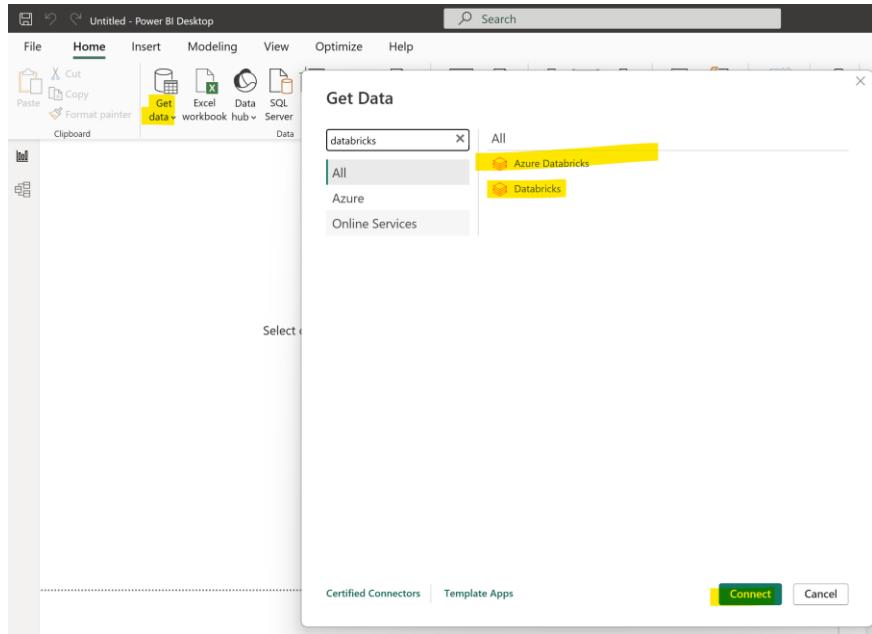
adb-1414194557308463.3.azuredatabricks.net...

- 'hive metastore'.'person_db'.'email'
- 'hive metastore'.'person_db'.'person_gold'
- 'hive metastore'.'person_db'.'person'
- 'hive metastore'.'person_db'.'phone'
- 'hive metastore'.'person_db'.'phonenumb...'

BusinessEntityID	PersonType	Title	FirstName	MiddleName	LastName	EmailAddress	EmailPromotion	PhoneNumber
1	EM	NULL	Ken	S	Ortiz	ken@adventure-works.com	0	609-555-0142
2	EM	NULL	Terri	L	Duffy	terri@adventure-works.com	1	819-555-0175
3	EM	NULL	Roberto	R	Tamburello	roberto@adventure-works.com	0	312-555-0187
4	EM	NULL	Rob	N	Walters	rob@adventure-works.com	0	612-555-0100
5	EM	M	Gail	A	Erickson	gail@adventure-works.com	0	849-555-0139
6	EM	M	Jossef	H	Goldberg	josef@adventure-works.com	0	312-555-0189
7	EM	NULL	Dylan	A	Miller	dylan@adventure-works.com	2	410-555-0156
8	EM	NULL	Diane	L	Marklein	diane@adventure-works.com	0	850-555-0180
9	EM	NULL	Gail	N	Mcmillan	gail@adventure-works.com	0	188-555-0186
10	EM	NULL	Michael	N	Rahim	michael@adventure-works.com	2	310-555-2568
11	EM	NULL	Ovidiu	V	Gruber	ovidiu@adventure-works.com	0	719-555-0181
12	EM	NULL	Thierry	B	D'Heris	thierry@adventure-works.com	2	168-555-0183
13	EM	M	Janice	M	Gahn	janice@adventure-works.com	2	473-555-0117
14	EM	NUL	Michael	I	Sullivan	michael@adventure-works.com	2	465-555-0196
15	EM	NULL	Sharon	B	Salavarria	sharon@adventure-works.com	2	970-555-0138
16	EM	NULL	David	M	Bradley	david@adventure-works.com	1	913-555-0172
17	EM	NULL	Kevin	F	Brown	kevin@adventure-works.com	2	150-555-0189
18	EM	NULL	John	L	Wood	john@adventure-works.com	2	486-555-0110
19	EM	NULL	Mary	A	Dempsey	mary@adventure-works.com	1	124-555-0114
20	EM	NULL	Wanda	M	Benítez	wanda@adventure-works.com	2	708-555-0141
21	EM	NULL	Terry	J	Eminizer	terry@adventure-works.com	2	138-555-0118
22	EM	NULL	Saranya	E	Hampdogdingayata	saranya@adventure-works.com	0	399-555-0176
23	EM	NULL	Mary	E	Gibson	mary@adventure-works.com	0	531-555-0183
24	EM	M	Jill	A	Williams	jill@adventure-works.com	0	510-555-0121
25	EM	NULL	James	R	Hamilton	james@adventure-works.com	0	870-555-0122
26	EM	NULL	Peter	J	Krebs	peter@adventure-works.com	0	913-555-0196
27	EM	NULL	Jo	A	Brown	jo@adventure-works.com	0	632-555-0129
28	EM	NULL	Guy	R	Gilbert	guy@adventure-works.com	0	320-555-0195
29	EM	NULL	Mark	K	McArthur	mark@adventure-works.com	1	417-555-0154

Load Cancel

We can use Power BI -> Get Data -> Type Databricks -> then connect to Azure Databricks.
 There are different ways to connect.



Sample Power BI Report

