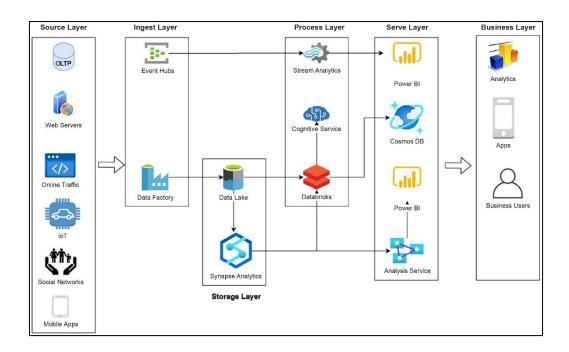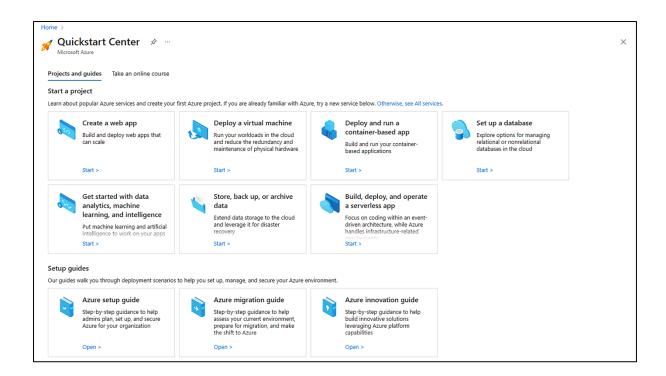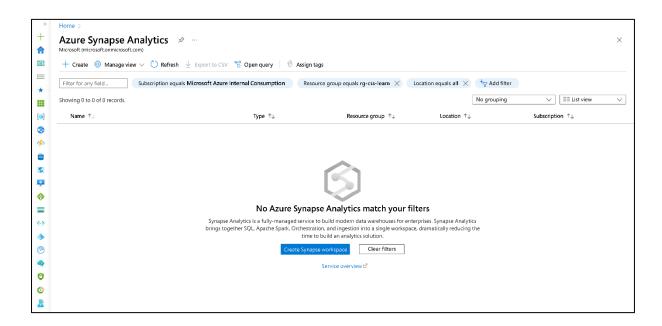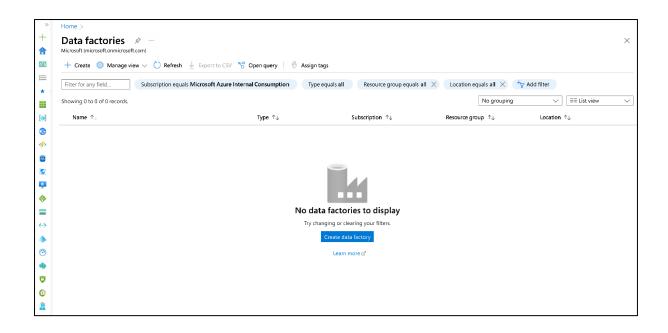# Chapter 1: Getting Started with ADF
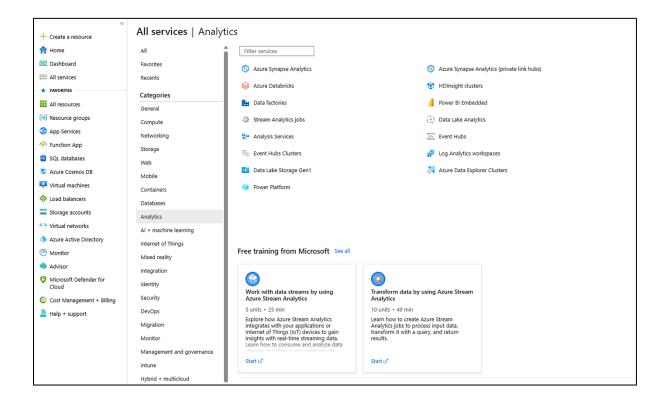
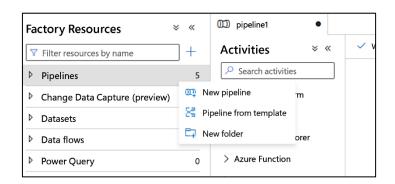| Service Name	Definition
Azure Synapse Analytics	A limitless analytics service with unmatched time to insight (formerly SQL Data Warehouse)
Power BI	A business intelligence solution for building reports and dashboards and data visualization
Azure Stream Analytics	Real-time analytics on fast-moving streams of data from applications and devices
ADF	Hybrid data integration at an enterprise scale made easy
Azure Databricks	Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform
Azure Cognitive Services	Cloud-based services with REST **application programming interfaces (APIs)** and client library **software development kits (SDKs)**, available to help developers build cognitive intelligence into applications without having direct **artificial intelligence (AI)** or data science skills or knowledge
Azure Event Hubs	Big data streaming platform and event ingestion service
Azure Data Lake Storage	Set of capabilities dedicated to big data analytics, built on Azure Blob storage
Azure HDInsight	Provisions cloud Hadoop, Spark, R Server, HBase, and Storm clusters
Azure Cosmos DB	Fast NoSQL database with open APIs for any scale
Azure SQL Database	Managed, intelligent SQL in the cloud

🚀 Quickstart Center

Microsoft Azure

Projects and guides   Take an online course

Start a project

Learn about popular Azure services and create your first Azure project. If you are already familiar with Azure, try a new service below. Otherwise, see All services.

Create a web app
Build and deploy web apps that can scale

Start >

Deploy a virtual machine
Run your workloads in the cloud and reduce the redundancy and maintenance of physical hardware

Start >

Deploy and run a container-based app
Build and run your container-based applications

Start >

Set up a database
Explore options for managing relational or nonrelational databases in the cloud

Start >

Get started with data analytics, machine learning, and intelligence
Put machine learning and artificial intelligence to work on your apps

Start >

Store, back up, or archive data
Extend data storage to the cloud and leverage it for disaster recovery

Start >

Build, deploy, and operate a serverless app
Focus on coding within an event-driven architecture, while Azure handles infrastructure-related

Start >

Setup guides

Our guides walk you through deployment scenarios to help you set up, manage, and secure your Azure environment.

Azure setup guide
Step-by-step guidance to help admins plan, set up, and secure Azure for your organization

Open >

Azure migration guide
Step-by-step guidance to help assess your current environment, prepare for migration, and make the shift to Azure

Open >

Azure innovation guide
Step-by-step guidance to help build innovative solutions leveraging Azure platform capabilities

Open >

Azure Synapse Analytics

Microsoft (microsoft.onmicrosoft.com)

+ Create   ⚙ Manage view ∨   ↻ Refresh   ↓ Export to CSV   Open query | Assign tags

| Filter for any field... | Subscription equals **Microsoft Azure Internal Consumption** | Resource group equals rg-css-learn ✕ | Location equals all ✕ | Add filter |

Showing 0 to 0 of 0 records.   No grouping ∨   List view ∨

Name ↑↓	Type ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓

No Azure Synapse Analytics match your filters

Synapse Analytics is a fully-managed service to build modern data warehouses for enterprises. Synapse Analytics brings together SQL, Apache Spark, Orchestration, and ingestion into a single workspace, dramatically reducing the time to build an analytics solution.

Create Synapse workspace | Clear filters

Service overview ⧉

Chapter 2: Orchestration and Control Flow

New linked service

Data store Compute

🔍 Search

All Azure Database File Generic protocol NoSQL Services and apps

Azure Blob Storage

Azure Cosmos DB for MongoDB

Azure Cosmos DB for NoSQL

Azure Data Explorer (Kusto)

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2

Continue Cancel

New linked service

🖼 Azure Blob Storage Learn more ⤢

Name *

OrchestrationAzureBlobStorage1

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime ⌄

Authentication type

Account key ⌄

Connection string   **Azure Key Vault**

Account selection method ⓘ

◉ From Azure subscription   ◯ Enter manually

Azure subscription ⓘ

Azure subscription 1 (1176c393-e348-4cdf-95b8-090923b3908e) ⌄

Storage account name *

adforchestrationstorage ⌄   ↻

Additional connection properties

✅ Connection successful

| Create | Back | 🔌 Test connection | Cancel |

New linked service

🗄 Azure SQL Database   Learn more ⧉

Name *

AzureSqlDatabase1

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime   ⌄

Connection string   **Azure Key Vault**

Account selection method ⓘ

🔘 From Azure subscription   ⚪ Enter manually

Azure subscription

Azure subscription 1 (1176c393-e348-4cdf-95b8-090923b3908e)   ⌄

Server name *

adf-orchestration-sql-server   ⌄   ↻

Database name *

AzureSQLDatabase   ⌄   ↻

Authentication type *

✅ Connection successful

Create   **Back**   ⚡ Test connection   **Cancel**

Data Factory ∨   ✓≡ Validate all   ⬆ Publish all

Factory Resources   ⌄ «

▽ Filter resources by name   +

▷ **Pipelines**   0

▷ **Change Data Capture (preview)**   0

▷ **Datasets**   0

    ⊞ New dataset

▷ **Data flows**   ⊟ New folder

▷ **Power Query**

Set properties

Name

CsvData

Linked service *

OrchestrationAzureBlobStorage1   ∨   ✏

File path

data   /   Directory   /   countries.csv   📁 | ∨

First row as header   ☑

Import schema

◉ From connection/store   ○ From sample file   ○ None

AzureSqlTables ●

Azure SQL Database
AzureSqlTables

Connection   Schema   **Parameters**

+ New   🗑 Delete

	Name	Type	Default value
☐	tableName	String ⌄	Value 🗑

AzureSqlTables ●

Azure SQL Database
AzureSqlTables

Connection   Schema   Parameters

Linked service *   AzureSqlDatabase1 ⌄

🔌 Test connection   ✏ Edit   + New   Learn more 🔗

Table   dbo   . @dataset().tableName
                                 ✓ Edit

👓 Preview data

Pipeline expression builder

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

```
@dataset().filename
```

Clear contents

Parameters   Functions

🔍 Search   +

filename

🏭 Data Factory ⌄   ✓≣ Validate all   ⬆ Publish all ①

Factory Resources   ⌄ «

▽ Filter resources by name   +

▷ **Pipelines**   0

▷ **Change Data Capture (preview)**   0

◢ **Datasets**   2

　　⊞ AzureSqlTables

● ⊞ CsvData

▷ **Data flows**   0

▷ **Power Query**   0

⊞ **CsvData**   ●

DelimitedText
CsvData

Connection   Schema   Parameters

Pipeline expression builder

↗

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

❌ Please specify an expression

Clear contents

| Parameters | System variables | **Functions** | Variables |

🔍 Search

⌄ Expand all

› **Collection Functions**

› **Conversion Functions**

› **Date Functions**

› **Logical Functions**

› **Math Functions**

⌄ **String Functions**

concat
Combines any number of strings together. For example, if parameter1 is foo, the following expressi...

endswith
Checks if the string ends with a value case insensitively. For example, the following expression retur...

[OK] [Cancel]
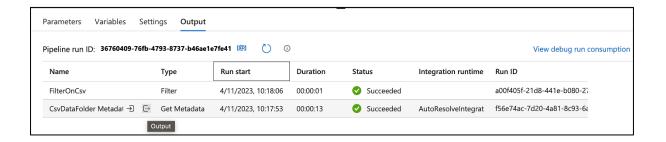
Metadata options
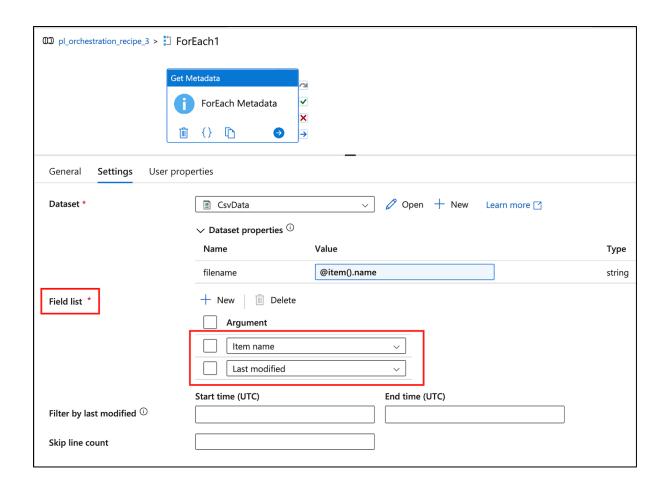
You can specify the following metadata types in the Get Metadata activity field list to retrieve the corresponding information:
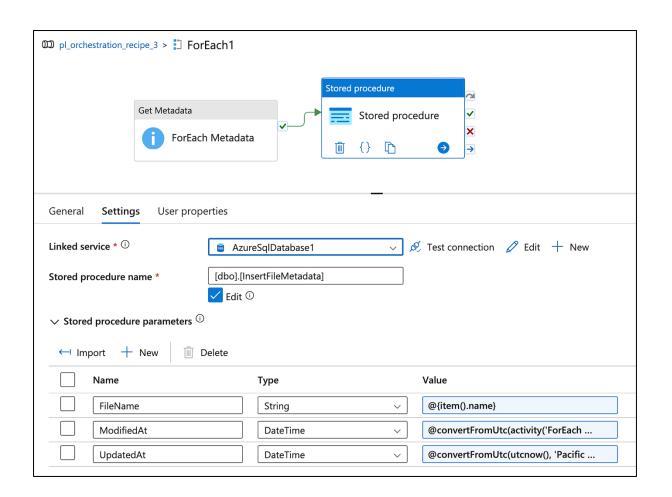
Metadata type	Description
itemName	Name of the file or folder.
itemType	Type of the file or folder. Returned value is `File` or `Folder`.
size	Size of the file, in bytes. Applicable only to files.
created	Created datetime of the file or folder.
lastModified	Last modified datetime of the file or folder.
childItems	List of subfolders and files in the given folder. Applicable only to folders. Returned value is a list of the name and type of each child item.
contentMD5	MD5 of the file. Applicable only to files.
structure	Data structure of the file or relational database table. Returned value is a list of column names and column types.
columnCount	Number of columns in the file or relational table.
exists	Whether a file, folder, or table exists. Note that if `exists` is specified in the Get Metadata field list, the activity won't fail even if the file, folder, or table doesn't exist. Instead, `exists: false` is returned in the output.

Parameters   Variables   Settings   **Output**

Pipeline run ID: **36760409-76fb-4793-8737-b46ae1e7fe41** [@] ⟳ ⓘ   View debug run consumption

Name	Type	Run start	Duration	Status	Integration runtime	Run ID
FilterOnCsv	Filter	4/11/2023, 10:18:06	00:00:01	✅ Succeeded		a00f405f-21d8-441e-b080-27
CsvDataFolder Metadat ⇥ ⇤	Get Metadata	4/11/2023, 10:17:53	00:00:13	✅ Succeeded	AutoResolveIntegrat	f56e74ac-7d20-4a81-8c93-6a

Output

Output
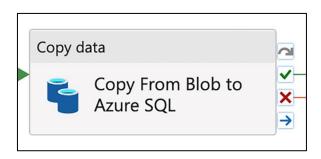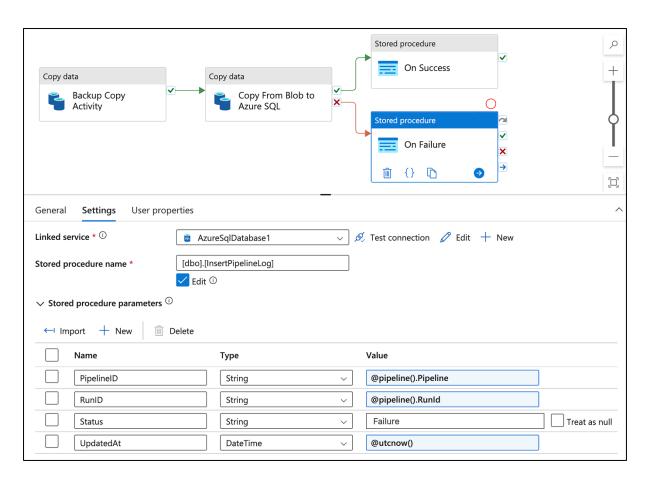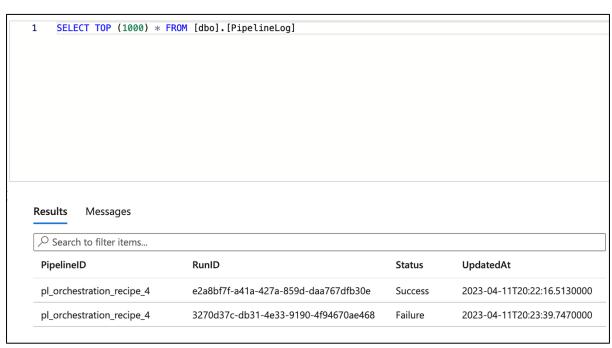
```json
{
    "itemName": "data",
    "lastModified": "2023-03-29T14:47:27Z",
    "childItems": [
        {
            "name": "Airlines-2023-04-11T19:21:59.4183732Z.backup",
            "type": "File"
        },
        {
            "name": "airlines.csv",
            "type": "File"
        },
        {
            "name": "countries.csv",
            "type": "File"
        }
    ],
    "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (East US)",
    "executionDuration": 0,
    "durationInQueue": {
        "integrationRuntimeQueue": 0
    },
    "billingReference": {
        "activityType": "PipelineActivity",
        "billableDuration": [
            {
                "meterType": "AzureIR",
                "duration": 0.016666666666666666,
                "unit": "Hours"
            }
        ]
    }
}
```

Get Metadata

ForEach Metadata

Stored procedure

Stored procedure

General | **Settings** | User properties

Linked service * ⓘ   AzureSqlDatabase1 ▾   ⚡ Test connection   ✏ Edit   ＋ New

Stored procedure name *   [dbo].[InsertFileMetadata]

☑ Edit ⓘ

∨ **Stored procedure parameters** ⓘ

←┤ Import   ＋ New   🗑 Delete

	Name	Type	Value
☐	FileName	String ▾	@{item().name}
☐	ModifiedAt	DateTime ▾	@convertFromUtc(activity('ForEach ...
☐	UpdatedAt	DateTime ▾	@convertFromUtc(utcnow(), 'Pacific ...

Copy data

Copy From Blob to Azure SQL

	Name	Type	Value	
☐	PipelineID	String	@pipeline().Pipeline	
☐	RunID	String	@pipeline().RunId	
☐	Status	String	Failure	☐ Treat as null
☐	UpdatedAt	DateTime	@utcnow()	

```
1    SELECT TOP (1000) * FROM [dbo].[PipelineLog]
```

Results   Messages

🔍 Search to filter items...

PipelineID	RunID	Status	UpdatedAt
pl_orchestration_recipe_4	e2a8bf7f-a41a-427a-859d-daa767dfb30e	Success	2023-04-11T20:22:16.5130000
pl_orchestration_recipe_4	3270d37c-db31-4e33-9190-4f94670ae468	Failure	2023-04-11T20:23:39.7470000

If Condition
Is CSV

True	
On Success	+

False	
On Failure	+

Get Metadata

CsvDataFolder Metadata

General   **Activities (2)**   User properties

Expression ⓘ   @not(endswith(activity('CsvDataFold...

Case	Activity	
True	On Success 1 Activity	✎
False	On Failure 1 Activity	✎

When a HTTP request is received

When a HTTP request is received

HTTP POST URL   URL will be generated after save

Request Body JSON Schema
```
{
    "type": "object",
    "properties": {
        "subject": {
            "type": "string"
        },
        "messageBody": {
            "type": "string"
```

Use sample payload to generate schema

Add new parameter

+ New step

Gmail

Search connectors and actions

Triggers   **Actions**

Delete email
Gmail

Get email details
Gmail

Move email to trash
Gmail

Reply to email (V2)
Gmail

Send email (V2)
Gmail

Don't see what you need?

Help us decide which connectors and triggers to add next with UserVoice

Gmail

Search connectors and actions

Triggers   **Actions**

Delete email
Gmail   ⓘ

Get email details
Gmail   ⓘ

Move email to trash
Gmail   ⓘ

Reply to email (V2)
Gmail   ⓘ

Send email (V2)
Gmail   ⓘ

Don't see what you need?

☺ Help us decide which connectors and triggers to add next with UserVoice

---

When a HTTP request is received   ···

⊕

Send email (V2)   ⓘ   ···

*To   A list of valid email addresses separated by a semicolon or a comma.

Add new parameter   ⌄

🔍 Search or filter parameters...

☑ Subject   Subject of the outgoing email.
☑ Body   Body of the outgoing email.
☐ Attachments   Details of attachments to be sent along with th...
☐ CC   A list of valid email addresses separated by a s...
☐ BCC   A list of valid email addresses separated by a s...
☐ Importance   Importance associated with the email message.

Lookup — Get Email Recipients

General | **Settings** | User properties

Source dataset *   AzureSqlTables   ✏ Open   + New   👓 Preview data

     ∨ Dataset properties ⓘ

Name	Value
tableName	dbo.EmailRecipients

First row only   ☐

Use query   ◯ Table   ◉ Query   ◯ Stored procedure

Query *   SELECT * FROM [dbo].[EmailRecipient]   ✏ Edit

Query timeout (minutes) ⓘ   120

Isolation level ⓘ   None

Partition option ⓘ   ◉ None   ◯ Physical partitions of table ⓘ   ◯ Dynamic range ⓘ

Web

Send Email

URL * ⓘ
https://prod-26.eastus.logic.azure.com:4...

⚠ Information will be sent to the URL
specified. Please ensure you trust the URL
entered.

Method * ⓘ
POST

Body
@json(concat('{"emailAddress": "', ite...

Authentication ⓘ
None

Headers * ⓘ
+ New   🗑 Delete

	Name	Value
☐	Content-Type	application/json

> Advanced

Web

Send Email

General **Settings** User properties

URL * ⓘ

`https://prod-26.eastus.logic.azure.com:4...`

⚠ Information will be sent to the URL specified. Please ensure you trust the URL entered.

Method * ⓘ

`POST`

Body

`@json(concat('{"emailAddress": "', ite...`

Authentication ⓘ

`None`

Headers * ⓘ

\+ New   🗑 Delete

☐	Name	Value
☐	Content-Type	application/json

\> **Advanced**

Copy data

Backup Copy Activity ✓ →

Copy data

Copy From Blob to Azure SQL ✓

Stored procedure

On Success ✓

Execute Pipeline ⧉

Send Email On Failure ✓

Filter

Filter for Backup

General	**Settings**	User properties

Items — @activity('CsvDataFolder Metadata')....

Condition — @endswith(item().name, '.backup')

ForEach

ForEach1

General	**Settings**	Activities (2)	User properties

Sequential — ☐

Batch count ⓘ —

Items — @activity('Filter For Backup').output....

Copy data
Copy from Data to Backup

Delete
Delete1

| General | Source | Logging settings | User properties |

Dataset *

CsvData   Open   New   Preview data

∨ **Dataset properties**

Name	Value
filename	@item().name

File path type

( ) File path in dataset   ( ) Wildcard file path   ( ) Prefix   ( ) List of files

Filter by last modified

Start time (UTC)

End time (UTC)

Recursively

✓

Name *

trigger_blob_added

Description

Type *

BlobEventsTrigger

Account selection method * ⓘ

◉ From Azure subscription   ◯ Enter manually

Azure subscription ⓘ

Azure subscription 1 (1176c393-e348-4cdf-95b8-090923b3908e)   ⌄

Storage account name * ⓘ

adforchestrationstorage   ⌄   ↻

Container name * ⓘ

data   ⌄

Blob path begins with ⓘ

Blob path ends with ⓘ

Event * ⓘ

☑ Blob created   ☐ Blob deleted

Ignore empty blobs * ⓘ

◉ Yes   ◯ No

Continue   Cancel

Chapter 3: Setting Up Synapse Analytics

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 *

| adfcookbookv2synapse | ✓ |

Region *

| East US 2 | ⌄ |

Select Data Lake Storage Gen2 * ⓘ   ◉ From subscription   ◯ Manually via URL

Account name * ⓘ

| adfcookbookv2 | ⌄ |

Create new

File system name *

| (New) adfcookbookv2synapsefs | ⌄ |

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 Contributor** role on the storage account
>
> Learn more

| **Review + create** | | < Previous | Next: Security > |

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 Azure Active Directory (Azure AD) authentication
- ○ Use only Azure Active Directory (Azure AD) authentication

SQL Server admin login * ⓘ | synapseadmin ✓

SQL Password ⓘ | ••••••••• ✓

Confirm 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

⚠ Double encryption configuration cannot be changed after opting into using a customer-managed key at the time of

[Review + create]   [< Previous]   [Next: Networking >]

Connection

Recent   **Browse**

Type here to filter the list

> 📁 Saved Connections
> 📁 Azure
  > 👤 _____ subscr...
    🗄 adfcookbookv2synapse (Azure Synapse Analytics, Azure subscription 1)
    🗄 adfcookbookv2synapse/adfcookbookv2sqlpool (Dedicated SQL Pools, Az...
    🗄
    🗄
    🗄
    🗄

Connection Details

Connection type	Microsoft SQL Server ⌄

● Parameters ○ Connection String

Server *	adfcookbookv2synapse.sql.azuresynapse.net
Authentication type	SQL Login ⌄
User name *	synapseadmin
Password	••••••••

☑ Remember password

Database	master ⌄
Encrypt ⓘ	Mandatory (True) ⌄
Trust server certificate ⓘ	False ⌄
Server group	<Default> ⌄
Name (optional)	

Advanced...

Connect   **Cancel**

Connection Details

Connection type	Microsoft SQL Server ⌄

◉ Parameters   ○ Connection String

Server *	adfcookbookv2synapse.sql.azuresynapse.net
Authentication type	SQL Login ⌄
User name *	adfuser
Password	••••••••••

☑ Remember password

Database	master ⌄
Encrypt ⓘ	Mandatory (True) ⌄
Trust server certificate ⓘ	False ⌄
Server group	<Default> ⌄
Name (optional)	

Advanced...

Connect   Cancel

Microsoft Azure | adfcookbookv2synapse

Search

Synapse live | Validate all | Publish all 1

Data

Workspace | Linked

Filter resources by name

▲ Azure Data Lake Storage Gen2 | 2
  ▲ adfcookbookv2synapse (Primar... ...
    adfcookbookv2synapsefs (Primary)
    data
    snowflakedata
    stage
  ▶ (Attached Containers) ...

SQL script 1 ● | data ✕

New SQL script | New notebook | New data flow | New integration dataset | Upload | Download | New folder

data

Name	Last Modified	Content Type
customers.csv	4/29/2023, 10:44:06 PM	
planes.csv	4/29/2023, 9:35:36 PM	

Preview
New SQL script ›   Select TOP 100 rows
New notebook ›   Create external table
New data flow   Bulk load
New integration dataset
Manage access...
Rename...
Download
Delete
Properties...

---

ublish all 1

SQL script 1 ● | data ✕

New SQL script | New notebook | New data flow | New integration

data

Name	Last Modified
customers.csv	4/29/2023, 10:44:06 PM
planes.csv	4/29/2023, 9:35:36 PM

Showing 1 to 2 of 2 cached items

Bulk load

Source data file format settings
Specify the format and layout of your data. Learn more ⧉

File path
data/customers.csv

Preview Data

Field terminator ⓘ
Default (comma ,)
☐ Edit

Row terminator ⓘ
Default (\n, or \r\n)
☐ Edit

First row ⓘ
0

☐ Infer column names ⓘ

Field quote ⓘ
Default (double quote ")
☐ Edit

Date format ⓘ
Session default

Compression ⓘ
None

Max string length * ⓘ
4000

Error settings
Specify your error tolerance and where rejected rows and the corresponding error file will be written. Learn more ⧉

Max errors ⓘ
0

Continue   Cancel

Bulk load

Select target SQL pool

Specify the target location for your load including the SQL pool, table, and column mapping. Learn more ⬚

Select SQL pool* ⓘ

| ✅ adfcookbookv2sqlpool | ⌄ | ↻ |

Select a database* ⓘ

| adfcookbookv2sqlpool | ⌄ |

Target table

◉ Existing table   ◯ Create new

Existing target table ⓘ

| dbo.Customer | ⌄ |

Configure column mapping

Load data

◯ Automatically   ◉ Using SQL script

ⓘ This will generate a SQL script and you will be required to run the SQL script.

| Open script | Back | | Cancel |

```sql
1   --Uncomment the 4 lines below to create a stored procedure for data pipeline orchestration
2   --CREATE PROC bulk_load_Customer
3   --AS
4   --BEGIN
5   COPY INTO dbo.Customer
6   (CustomerKey 1, CustomerID 2, Customer 3, BillToCustomer 4, Category 5, BuyingGroup 6, PrimaryContact 7, PostalCoo
7   FROM 'https://adfcookbookv2.dfs.core.windows.net/data/customers.csv'
8   WITH
9   (
10      FILE_TYPE = 'CSV'
11      ,MAXERRORS = 0
12      ,ERRORFILE = 'https://adfcookbookv2.dfs.core.windows.net/data/'
13  )
14  --END
15  GO
16
17  SELECT TOP 100 * FROM dbo.Customer
18  GO
```

Chapter 4: Working with Data Lake and Spark Pools

Create storage account

Basics   **Networking**   Data protection   Advanced   Tags   Review + create

Network connectivity

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method *

- ● Public endpoint (all networks)
- ○ Public endpoint (selected networks)
- ○ Private endpoint

ℹ All networks will be able to access this storage account.
Learn more about connectivity methods ⧉

Network routing

Determine how to route your traffic as it travels from the source to its Azure endpoint. Microsoft network routing is recommended for most customers.

Routing preference * ⓘ

- ● Microsoft network routing (default)
- ○ Internet routing

[**Review + create**]   [< Previous]   [Next : Data protection >]

Create storage account

Basics   Networking   **Data protection**   Advanced   Tags   Review + create

Recovery

☐ **Turn on point-in-time restore for containers**

Use point-in-time restore to restore one or more containers to an earlier state. If point-in-time restore is enabled, then versioning, change feed, and blob soft delete must also be enabled. Learn more ⧉

☐ **Turn on soft delete for blobs**

Soft delete enables you to recover blobs that were previously marked for deletion, including blobs that were overwritten. Learn more ⧉

☐ Turn on soft delete for containers

Soft delete enables you to recover containers that were previously marked for deletion. Learn more ⧉

ℹ Sign up is required on a per-subscription basis to use container soft delete. Sign up for container soft delete ⧉

☐ **Turn on soft delete for file shares**

Soft delete enables you to recover file shares that were previously marked for deletion. Learn more ⧉

Tracking

☐ **Turn on versioning for blobs**

Use versioning to automatically maintain previous versions of your blobs for recovery and restoration. Learn more ⧉

☐ **Turn on blob change feed**

Keep track of create, modification, and delete changes to blobs in your account. Learn more ⧉

[**Review + create**]   [< Previous]   [Next : Advanced >]

New Apache Spark pool

Basics •   Additional settings *   Tags   Review + create

Create an Synapse Analytics Apache Spark pool with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize.

Apache Spark pool details

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name *	adfcookbookv2
Isolated compute * ⓘ	◯ Enabled ⦿ Disabled
Node size family *	Memory Optimized ⌄
Node size *	Small (4 vCores / 32 GB) ⌄
Autoscale * ⓘ	⦿ Enabled ◯ Disabled
Number of nodes *	[3] ○━○────────── [10]
Estimated price ⓘ	**Est. cost per hour** 1.66 to 5.52 USD View pricing details
Dynamically allocate executors * ⓘ	◯ Enabled ⦿ Disabled

New Apache Spark pool

Basics *   **Additional settings** *   Tags   Review + create

Customize additional parameters including pause settings and component versions.

Automatic pausing

Configure the pause settings for the Apache Spark pool.

Automatic pausing * ⓘ	⦿ Enabled ◯ Disabled
Number of minutes idle *	15

Chapter 5: Working with Big Data and Databricks

Add role assignment ...

Role   Members   Conditions (optional)   **Review + assign**

Role   Storage Blob Data Owner

Scope   /subscriptions/c5389450-5c0d-459f-a7f6-cdc175d36cff/resourceGroups/adfcookbookv2/providers/Microsoft.Storage/storageAccounts/adfcookbookv2sa

Members

Name	Object ID	Type
managed_identity_adf	01a634a9-a06b-40b8-9c2b-3e2111cd0de7	Managed Identity ⓘ

Description   No description

Condition

1	

Review + assign

Review + assign   Previous   Next

Create HDInsight cluster ···

Basics   Storage   Security + networking   Configuration + pricing   Tags   Review + create

New to HDInsight? Get started with our training resources.
Create a managed HDInsight cluster. Select from Spark, Kafka, Hadoop, Storm, and more. Learn more

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 *	adfcookbookv2 ⌄
	Create new

Cluster details

Name your cluster, pick a region, and choose a cluster type and version. Learn more

Cluster name *	adfhdihadoop ✓
Region *	East US ⌄
Availability zone ⓘ	⌄
Cluster type *	**Hadoop**
	Change
Version *	Hadoop 3.1.0 (HDI 4.0) ⌄

Cluster credentials

Enter new credentials that will be used to administer or access the cluster.

Cluster login username * ⓘ	admin
Cluster login password *	•••••••••••••• ✓
Confirm cluster login password *	•••••••••••••• ✓
Secure Shell (SSH) username * ⓘ	sshuser
Use cluster login password for SSH	☑

[**Review + create**]   [« Previous]   [Next: Storage »]

Create HDInsight cluster ⋯

Basics   **Storage**   Security + networking   Configuration + pricing   Tags   Review + create

Select or create storage accounts that will be used for the cluster's logs, job input, and job output. Configure the cluster's access to these accounts, if needed.

Primary storage

Select or create a storage account that will be the default location for cluster logs and other output.

Primary storage type *	Azure Storage ⌄
Selection method * ⓘ	⦿ Select from list ◯ Use access key
Primary storage account *	adfcookbookv2sa ⌄
	Create new
Container * ⓘ	adfhdihadoop-2023-06-08t15-26-16-475z ✓

Identity

Select a user-assigned service identity to represent your cluster for enterprise security package or disk encryption. Learn More

User-assigned managed identity ⓘ	managed_identity_adf ⌄

Review + create   « Previous   Next: Configuration + pricing »

Create HDInsight cluster   ✕

ⓘ This configuration will use 40 of 40 available cores in the East US region.
View cores usage

+ Add application

Node type	Node size	Number of ...	Estimated cost/h...
Head node	D12 v2 (4 Cores, 28 GB RAM), 0.37 USD/hour ⌄	2	0.75 USD
Worker node	D4 v2 (8 Cores, 28 GB RAM), 0.59 USD/hour ⌄	4	2.36 USD

☐ Enable autoscale
Learn More

Total estimated cost/hour   3.11 USD

Home > HDInsight clusters >

afdhdihadoop 📌
HDInsight cluster

🔍 Search (Cmd+/) «   → Move ∨   🗑 Delete   ↻ Refresh

Overview

📋 Activity log   ⓘ Applying changes...

👥 Access control (IAM)   ∧ Essentials

🏷 Tags   Resource group (change)   Learn more
                                   ADF   Documentation

🩺 Diagnose and solve problems   Status   Cluster type, HDI version
                                   HDInsight configuration   Hadoop 2.7 (HDI 3.6)

⚡ Quick start   Location   URL
                                   East US   https://afdhdihadoop.azurehdinsight.net
🔲 Tools
                                   Subscription (change)   Cluster ID
Settings   Azure subscription 1   42cde877aef6450ab65a1465d7ae1298

📝 Cluster size   Subscription ID   Getting started
                                   1e70af7e-a8d1-4e0d-9a95-e6b389f01eb4   Quickstart
🔲 Quota limits
                                   Tags (change)
🔑 SSH + Cluster login   Click here to add tags

                                    « ⟳ Switch tenant   🗑 Delete tenant   + Create a tenant   ⬚ What's new   ⊞ Preview features   ⋯

🔲 External Identities

👥 Roles and administrators   ◈ **Tenant information**   💟 **Azure AD Connect**

🖼 Administrative units (Preview) **Your role**   **Status**
                                   Global administrator More info   Not enabled
🔲 Enterprise applications
                                   License   **Last sync**
🖥 Devices   Azure AD Free   Sync has never run

🔲 App registrations   **Tenant ID**
                                   ac76bf9e-678c-4ce2-bdb8-4a76... 🗐
🔲 Identity Governance
                                   Primary domain
🔲 Application proxy   kshpundragmail.onmicrosoft.com

🔑 Licenses

◈ Azure AD Connect

Register an application   ✕

* Name

The user-facing display name for this application (this can be changed later).

| adfbook   ✓ |

Supported account types

Who can use this application or access this API?

🔵 Accounts in this organizational directory only (Каталог по умолчанию only - Single tenant)

⚪ Accounts in any organizational directory (Any Azure AD directory - Multitenant)

⚪ Accounts in any organizational directory (Any Azure AD directory - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)

⚪ Personal Microsoft accounts only

By proceeding, you agree to the Microsoft Platform Policies ⬀

Register

Subscriptions 📌

Каталог по умолчанию

╳

\+ Add

View list of subscriptions for which you have role-based access control (RBAC) permissions to manage Azure resources. To view subscriptions for which you have billing access, click here

Showing subscriptions in Каталог по умолчанию directory. Don't see a subscription? Switch directories

My role ⓘ	Status ⓘ
8 selected ⌄	3 selected ⌄

Apply

Showing 1 of 1 subscriptions   ☑ Show only subscriptions selected in the global subscriptions filter ⓘ

🔍 Search to filter items...

Subscription name ↑↓	Subscription ID ↑↓	My role ↑↓	Current cost ↑↓	Status ↑↓	
🔵 Azure subscription 1	1e70af7e-a8d1-4e0d-9a95-e6b389f01e...	Owner	$173.11	✅ Active	...

M G M ◆ Add role ass ╳ A Azure subsc ╳ A dfhdi - Azure ╳ 🔒 Домашняя с ╳ 📄 Сторчак Ром ╳ 📄 Руководстве ╳ 📄 Приложение ╳ 📊 BI_BigData_: ╳ 🔗 How to Simp ╳ ╳ \+

← → C 🔒 portal.azure.com/?l=en-en-us#@kshpundragmail.onmicrosoft.com/resource/subscriptions/1e70af7e-a8d1-4e0d-9a95-e6b389f01eb4/users

≡   **Microsoft Azure**   ⊕ 🔍 Search resources, services, and docs (G+/)   ▣ ⬚ 🔔 ⚙ ? ☺ k.sh.pundra@gmail.com
КАТАЛОГ ПО УМОЛЧАНИЮ

Home > Subscriptions > Azure subscription 1

🔖 Azure subscription 1 | Access control (IAM)
Subscription

🔍 Search (Cmd+/) ≪

🏠 Overview

📋 Activity log

🔖 Access control (IAM)

🏷 Tags

🔧 Diagnose and solve problems

🛡 Security

⚡ Events

Cost Management

💰 Cost analysis

💲 Cost alerts

💲 Budgets

\+ Add   ⬇ Download role assignme

Check access   Role assignments

My access
View my level of access to this resource

View my access

Check access
Review the level of access a user, group, principal, or managed identity has to this resource. Learn more ⧉

Find ⓘ
Azure AD user, group, or service principa

🔍 Search by name or email address

Add role assignment ╳

Role ⓘ
Contributor ⓘ ⌄

Assign access to ⓘ
Azure AD user, group, or service principal ⌄

Select ⓘ
adfbook

🔲 adfbook

Selected members:

🔲 adfbook   Remove

Save   **Discard**

≫ 🏭 Data Factory ⌄ ⬆ Publish all ❶ ✓ Validate all ↻ Refresh 🗑 Discard all ◯⚫ Data flow debug 🔲 ARM template ⌄ ▷ ≪

🏠 » ▥ HDI_Hive   ●   ...

✏❶ **Activities** ≫ ≪   📷 Save as template ✓ Validate ▷ Debug ⚡ Add trigger   {} ⚙ ...

◉ 🔍 Search activities   **Properties**

💼 ▷ Move & transform   **General** Related

   ▷ Azure Data Explorer   **Name** *
   HDI_Hive
   ▷ Azure Function
   Description
   ▷ Batch Service   Runs Hive queries on HDI Hadoop cluster
   ▷ Databricks
   ▷ Data Lake Analytics
   ▷ General   **Concurrency** ⓘ
   ▷ HDInsight   1
   ▷ Iteration & conditionals   **Annotations**
   ▷ Machine Learning   \+ New

   🔍 \+ − 🔲 ▣ ⊞   **Name**

Parameters Variables Output   ⌃

adfbook 📌

Search (Cmd+/) «

🗑 Delete   🌐 Endpoints

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

- 📇 Overview
- ☁ Quickstart
- 🚀 Integration assistant | Preview

Manage

- 🖼 Branding
- 🔐 Authentication
- 🔑 Certificates & secrets
- ⑪ Token configuration
- ⊹ API permissions
- ☁ Expose an API
- 👥 App roles | Preview

∧ Essentials

Display name	Supported account types
adfbook	My organization only
Application (client) ID	Redirect URIs
d7955ea1-55b3-41c1-83c2-6e159f3d7f27	Add a Redirect URI 🗎
Directory (tenant) ID	Application ID URI
ac76bf9e-678c-4ce2-bdb8-4a7619dc1eb0	Add an Application ID URI
Object ID	Managed application in local directory
ed6e1846-dbd1-4463-b0fd-a67c125aeefc	adfbook

ℹ Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure AD Graph. We will continue to provide technical support and security updates but we will no longer provide feature updates. Applications will need to be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more

Activities
- ▷ Move & transform
- ▷ Azure Data Explorer
- ▷ Azure Function
- ▷ Batch Service
- ▷ Databricks
- ▷ Data Lake Analytics
- ▷ General
- ◢ HDInsight
  - Hive
  - MapReduce
  - Pig

General   HDI Cluster   Script   User properties

◢ Details

Script linked service *
lsadlv2   ⓘ 🔌 Test connection ✏ Edit + New

File path *
datalens/hive.hql   X ⓘ

Browse storage   Browse local   Add/Edit script   Preview script

▷ Advanced

Details   ↻ Refresh

Name	Value
Duration	00:07:14
Integration runtime	DefaultIntegrationRuntime (East US)
Log location	abfss://adfjobs@adfadlv2.dfs.core.windows.net/HiveQueryJobs/f1c65e62-c14b-4d2f-a74c-986105d7a13b/15_09_2020_07_38_49_335/Status
Compute information	https://hi-742f6925-b633-4d7c-ac9d-3c1049017c47.azurehdinsight.net:443/
Job ID	job_1600198594805_0001
Execution progress	Succeeded
YarnApplicationUri	https://hi-742f6925-b633-4d7c-ac9d-3c1049017c47.azurehdinsight.net//yarnui/hn/cluster/app/application_16001985948...
JobStatusDetailsFromHDI	application_1600198594805_0001
Duration in queue	{"integrationRuntimeQueue":10}
billingReference	{"activityType":"ExternalActivity","billableDuration":[{}]}

Home > Azure Databricks >

Azure Databricks Service   ✕

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 * ⓘ   ADF ⌄
    Create new

Instance Details

Workspace name *   adf_databricks ✓

Location *   East US ⌄

Pricing Tier * ⓘ   Standard (Apache Spark, Secure with Azure AD) ⌄

[Review + Create]   [Next : Networking >]

adfcookbookv2sa | Access keys

Storage account

🔍 Search

- Diagnose and solve problems
- Access Control (IAM)
- Data migration
- Events
- Storage browser
- Storage Mover

Data storage

- Containers
- File shares
- Queues
- Tables

Security + networking

- Networking
- Azure CDN
- Access keys
- Shared access signature
- Encryption
- Microsoft Defender for Cloud

🕐 Set rotation reminder   🔄 Refresh   🗨 Give feedback

Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location like Azure Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other.

Remember to update the keys with any Azure resources and apps that use this storage account.
Learn more about managing storage account access keys ⧉

Storage account name

adfcookbookv2sa

key1 🔄 Rotate key

Last rotated: 12/05/2023 (27 days ago)

Key

••   Show

Connection string

••   Show

key2 🔄 Rotate key

Last rotated: 12/05/2023 (27 days ago)

Key

••   Show

Connection string

••   Show

🏭 Data Factory   ✔ Validate all   📤 Publish all

General

- Factory settings

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 endpoints

Airflow (Preview)

- Airflow

Linked services

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

➕ New

Filter by name   Annotations : Any

Showing 1 - 3 of 3 items

Name	Type
AzureBlobStorage1	Azure Blob Storage
AzureDataLakeStorage1	Azure Data Lake Storage Gen2
SnowflakeLS	Snowflake

New linked service

Data store   Compute

🔍 Search

All   Azure   Compute

Azure Batch   Azure Data Lake Analytics   Azure Databricks

Azure Function   Azure HDInsight   Azure Machine Learning

Azure Machine Learning Studio (classic)   Azure Synapse Analytics (Artifacts)

Continue   Cancel

User Settings

Access tokens Git integration Editor settings Email preferences Language settings

Personal access tokens can be used for secure authentication to the Databricks API instead of passwords.

Generate new token

Comment	Creation	Expiration

No tokens exist.

---

pipeline1

Activities

not

∨ Synapse

Notebook

∨ Databricks

Notebook

✓ Validate ▷ Debug ⚡ Add trigger

Notebook ✓

Notebook1

General **Azure Databricks** Settings User properties

Databricks linked service * AzureDatabricks1 ∨ Test connection Edit + New

curated
Container

Search

«

- Overview
- Diagnose and solve problems
- Access Control (IAM)

Settings

- Shared access tokens
- Access policy
- Properties
- Metadata

↑ Upload   🔒 Change access level   ↻ Refresh | 🗑 Delete   ⇄ Change tier   ✏ Acquire lease   ✏ Break lease

Authentication method: Access key (Switch to Azure AD User Account)
Location: curated / movie_lens

Search blobs by prefix (case-sensitive)

+ Add filter

Name	Modified	Access tier	Archive status	Blob type
☐ 📁 [..]				
☐ 📁 _delta_log				
☐ 📁 rating_datetime=1996-04-17				
☐ 📁 rating_datetime=1996-05-14				
☐ 📁 rating_datetime=1996-08-17				
☐ 📁 rating_datetime=1996-10-14				
☐ 📁 rating_datetime=1996-10-17				
☐ 📁 rating_datetime=1996-11-05				
☐ 📁 rating_datetime=1996-11-26				
☐ 📁 rating_datetime=1996-12-15				
☐ 📁 rating_datetime=1997-01-22				
☐ 📁 rating_datetime=1997-02-14				
☐ 📁 rating_datetime=1997-02-19				
☐ 📁 rating_datetime=1997-02-23				
☐ 📁 rating_datetime=1997-02-25				

✓ Validate ⬤ Data flow debug ●

RawSource

Import data from
DelimitedText1

+

DeltaSink
Write order: 1

Columns:
0 total

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name *   DeltaSink   Learn more ⬈

Description   Add sink dataset   ↻ Reset

Incoming stream *   RawSource

Sink type * | Dataset | Inline | Cache |

Inline dataset type *   △ Delta

Linked service *   ▦ AzureBlobStorage1   ⬳ Test connection   ✎ Edit   + New

Options   ☑ Allow schema drift ⓘ
                            ☐ Validate schema ⓘ

Sink **Settings** Errors Mapping Optimize Inspect Data preview

Folder path *   staging   /   ratings   📁 Browse

Compression type   None

Vacuum ⓘ   0

Table action   ⦿ None   ◯ Overwrite ⓘ   ◯ Truncate ⓘ

Update method ⓘ   ☑ Allow insert

                       ☐ Allow delete

                       ☐ Allow upsert

                       ☐ Allow update

> **Delta options**

Activities

data

Move & transform

- Copy data
- Data flow

Azure Data Explorer

- Azure Data Explorer C...

Databricks

- Notebook
- Jar
- Python

Data Lake Analytics

- U-SQL

General

- Get Metadata

✓ Validate | ▷ Debug ⌄ | ⚡ Add trigger | ⬤ Data flow debug ●

Data flow ↗ ✓

Delta_DataFlow

🗑 {} ⧉ →

General | **Settings** | Parameters | User properties

Data flow *   Load_DeltaLake_DF   ✎ Open   + New

Run on (Azure IR) * ⓘ   AutoResolveIntegrationRuntime

Compute size * ⓘ   Small

∨ **Advanced**

  Compute type * ⓘ   General purpose

  Core count * ⓘ   4 (+ 4 Driver cores)

Logging level * ⓘ   ⦿ Verbose   ◯ Basic   ◯ None

  〉 Sink properties

  〉 Staging ⓘ

Turn on data flow debug

Integration runtime

AutoResolveIntegrationRuntime ⌄

⊡ AutoResolveIntegrationRuntime

Region	AutoResolve
Compute size	Small

Debug time to live ⓘ

1 hour ⌄

⊞ staging ⋯
Container

Search	«

⊡ Overview
🔧 Diagnose and solve problems
ஃ Access Control (IAM)

Settings

🔗 Shared access tokens
🔑 Access policy
⫴ Properties
ⓘ Metadata

⬆ Upload 🔓 Change access level ↻ Refresh | 🗑 Delete ⇄ Change tier ✏ Acquire lease

Authentication method: Access key (Switch to Azure AD User Account)
Location: staging / ratings

Search blobs by prefix (case-sensitive)

⊹ Add filter

	Name	Modified	Access tier	Archive status
☐	📁 [..]			
☐	📁 _delta_log			
☐	📄 _delta_log	6/8/2023, 11:16:18 AM	Hot (Inferred)	
☐	📄 part-00000-80aae3a7-afc0-43cd-...	6/8/2023, 11:16:17 AM	Hot (Inferred)	

Load_DeltaLake_DF (AddDatetime selected)

Save ✓ Validate   Data flow debug   Debug Settings

RawSource — Import data from DelimitedText1
AddDatetime — Columns: 4 total
AlterRow1 — Add expressions to alter rows
DeltaSink — Add sink dataset

Derived column's settings | Optimize | Inspect | Data preview ●

Output stream name * AddDatetime   Learn more ⧉

Description Creating/updating the column 'timestamp'   Reset

Incoming stream * RawSource

+ Add   Clone   Delete   ⧉ Open expression builder

Columns * ⓘ

☐	Column	Expression
☐	timestamp	toString(toTimestamp(toInteger(byName('TIMES... abc + 🗑

Load_DeltaLake_DF (AlterRow1 selected)

Save ✓ Validate   Data flow debug   Debug Settings

RawSource — Import data from DelimitedText1
AddDatetime — Creating/updating the column 'timestamp'
AlterRow1 — Columns: 4 total
DeltaSink — Add sink dataset

Alter row settings | Optimize | Inspect | Data preview ●

Output stream name * AlterRow1   Learn more ⧉

Description Add expressions to alter rows   Reset

Incoming stream * AddDatetime

Alter row conditions * ⓘ

| ☀ Update if | true() ✕ | + 🗑 |

RawSource — Import data from DelimitedText1

AddDatetime — Creating/updating the column 'timestamp'

AlterRow1 — Add expressions to alter rows

DeltaSink — Columns: 4 total

Sink | **Settings** | Errors | Mapping | Optimize | Inspect | Data preview ●

| Folder path * | staging | / | ratings | 📁 Browse |

Compression type: None

Vacuum ⓘ : 0

Table action: ● None ○ Overwrite ⓘ ○ Truncate ⓘ

Update method ⓘ:
- ☑ Allow insert
- ☐ Allow delete
- ☑ Allow upsert
- ☑ Allow update

Key columns * ⓘ: ● List of columns ○ Custom expression ⓘ

abc userId   ＋ 🗑
abc movieId   ＋ 🗑

🔍 Search

> ⊟ SNOWFLAKE
> ⊟ SNOWFLAKE_SAMPLE_DATA

Databases   ＋ Database

🔍 Search | Source All | ↻

New Database

Creating as 👤 ACCOUNTADMIN

Name
adfcookbookv2

Comment (optional)

Cancel | Create

...RCE	OWNER	CREATED
	—	5 minutes ago ...
	👤 ACCOUNTADMIN	5 minutes ago 💬 ...

🔍 Search

∨ ⊟ ADFCOOKBOOKV2
  > 🗄 INFORMATION_SCHEMA
  ∨ 🗄 PUBLIC
    > 🗄 Tables
> ⊟ SNOWFLAKE
> ⊟ SNOWFLAKE_SAMPLE_DATA

⊟ **ADFCOOKBOOKV2 / PUBLIC**   ... | Create ∨

🗄 Schema 👤 ACCOUNTADMIN 🕐 15 hours ago

Schema Details | Tables

Standard	Table >
As Select	View >
External	Stage >
	Storage Integration >
	File Format
	Sequence

Privileges   Group by Ro...

Add syntax for a: [Standard Table ⌄]   View Documentation   [📄 ACCOUNTADMIN • COMPUTE_WH]   **Create Table**

Pinned (0)

No pinned objects

🔍 All Objects   ⋯

∨ ⊟ ADFCOOKBOOKV2
   › ⬡ INFORMATION_SCHEMA
   ∨ ⬡ PUBLIC
     ∨ Tables
       ⊟ MOVIES
› ⊟ SNOWFLAKE
› ⊟ SNOWFLAKE_SAMPLE_DATA

```
1    create table tags (
2        userId INT,
3        movieId INT,
4        tag STRING NULL,
5        timestamp INT
6        )
```

🔍 Search   ⋯

∨ ⊟ ADFCOOKE
  › ⬡ INFORM
  ∨ ⬡ PUBLIC
   ∨ Tables
    ⊟ M
    ⊟ T
› ⊟ SNOWFLAK
› ⊟ SNOWFLAK

⊟ **ADFCOOKBOOKV2 / PUBLIC / TAGS**   ⋯   **Load Data**

Load Data into Table PREVIEW

📄 tags.csv → ⊟ TAGS

[● COMPUTE_WH]

File format

[Delimited Files (CSV or TSV) ⌄]

Select existing or create in Worksheets ↗

Learn more about format-specific configurations in Snowflake Docs ↗

Skip header ⓘ   [Skip one line ⌄]

Field delimiter ⓘ   [Comma (default) ⌄]

Trim space ⓘ   [False (default) ⌄]

Field optionally enclosed by ⓘ   [None (default) ⌄]

Date format ⓘ   [Auto* (default) ⌄]

Show SQL   [Cancel] [Back] [**Next**]

C.TAGS (

[Group by Role ⌄]   [+ Privilege]

ADFCOOKBOOKV2 / PUBLIC / TAGS

Table | ACCOUNTADMIN | 1 minute ago | 0 | 0.0B

Table Details | Columns | **Data Preview** | Copy History

COMPUTE_WH | 100 Rows • Updated just now

	USERID	MOVIEID	TAG	TIMESTAMP
1	2	60,756	funny	1,445,714,994
2	2	60,756	Highly quotable	1,445,714,996
3	2	60,756	will ferrell	1,445,714,992
4	2	89,774	Boxing story	1,445,715,207
5	2	89,774	MMA	1,445,715,200
6	2	89,774	Tom Hardy	1,445,715,205
7	2	106,782	drugs	1,445,715,054
8	2	106,782	Leonardo DiCaprio	1,445,715,051
9	2	106,782	Martin Scorsese	1,445,715,056
10	7	48,516	way too long	1,169,687,325

Load Data

New linked service

❄ Snowflake Learn more ↗

Name *

SnowflakeLS

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime ⌄

Connection string | **Azure Key Vault**

Account name * ⓘ

lv60010.east-us-2.azure

User name * ⓘ

ADFSNOW

Password | **Azure Key Vault**

Password * ⓘ

••••••••

Database * ⓘ

ADFCOOKBOOKV2

Warehouse * ⓘ

COMPUTE_WH

Role ⓘ

✅ Connection successful

| Create | Back | | ⚡ Test connection | Cancel |

Accounts   ➕ Account

1 Account in BSHHNRH ⓘ   🔍 Search   Edition All   Cloud All   Region All   ⟳

ACCOUNT ↑	EDITION	CLOUD	REGION		
				https://lv60010.east-us-2.azure.snowflakecomputing.com	
AJ38198	Standard	Azure	East US 2 (Virginia)	49 minutes ago	LV60010 🔗

Set properties

Name

SnowflakeTable1

Linked service *

SnowflakeLS ▽ ✎

Table name

PUBLIC.TAGS ▽ ↻

☐ Edit

Import schema

◉ From connection/store   ◯ None

| ⬗ SnowflakePL ● | ⊞ SnowflakeTable1 ● | ⊞ ADLSParquet ● |

✓ Validate ✓ Validate copy runtime ▷ Debug ⚡ Add trigger   { } ▤

Activities	⌄ «

🔍 Search activities

> Move & transform
> Synapse
> Azure Data Explorer
> Azure Function
> Batch Service
> Databricks
> Data Lake Analytics
> General
> HDInsight
> Iteration & conditionals
> Machine Learning
> Power Query

Copy data

🗄 Copy data from Snowflake

🗑 { } ⧉ ➔

Properties

General   Related

Name *

SnowflakePL

Description

Annotations

➕ New

General   Source   Sink   Mapping   **Settings**   User properties

Degree of copy parallelism ⓘ   ▽

☐ Edit

Fault tolerance ⓘ   ▽

Enable logging ⓘ   ☐

Enable staging ⓘ   ☑

⌄ Staging settings

Staging account linked service * ⓘ   🖫 AzureBlobStorage1 ▽

🔌 Test connection   ✎ Edit   ➕ New

Storage Path ⓘ   staging

📁 Browse

adfcookbookv2sa | Shared access signature ☆ ⋯
Storage account

🔍 Search «

📯 Give feedback

▣ Overview
▤ Activity log
🏷 Tags
🩺 Diagnose and solve problems
🔐 Access Control (IAM)
📁 Data migration
⚡ Events
🗄 Storage browser
🗄 Storage Mover

Data storage

▤ Containers
📄 File shares
▦ Queues
▦ Tables

Security + networking

🧑 Networking
☁ Azure CDN
🔑 Access keys
🔗 Shared access signature
🔒 Encryption

Preferred routing tier ⓘ
◉ Basic (default)   ○ Microsoft network routing   ○ Internet routing

ℹ️ Some routing options are disabled because the endpoints are not published.

Signing key ⓘ
[key1 ▾]

[**Generate SAS and connection string**]

Connection string
BlobEndpoint=https://adfcookbookv2sa.blob.core.windows.net/;QueueEndpoint=https://adfcookbookv2sa.queue.core.windows.net/;FileEndpoint=https://adfcookbookv2sa.file.core.windows...

SAS token ⓘ
?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdlacupiytfx&se=2024-04-01T19:17:48Z&st=2023-06-04T11:17:48Z&spr=https&sig=qZUEwH3RA8egCoVlwil5Het5ApUeerNDu%2B1BrZuDKuY%3D

Blob service SAS URL
https://adfcookbookv2sa.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdlacupiytfx&se=2024-04-01T19:17:48Z&st=2023-06-04T11:17:48Z&spr=https&sig=qZUEwH3RA8e...

File service SAS URL
https://adfcookbookv2sa.file.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdlacupiytfx&se=2024-04-01T19:17:48Z&st=2023-06-04T11:17:48Z&spr=https&sig=qZUEwH3RA8eg...

Queue service SAS URL
https://adfcookbookv2sa.queue.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdlacupiytfx&se=2024-04-01T19:17:48Z&st=2023-06-04T11:17:48Z&spr=https&sig=qZUEwH3RA...

Table service SAS URL
https://adfcookbookv2sa.table.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdlacupiytfx&se=2024-04-01T19:17:48Z&st=2023-06-04T11:17:48Z&spr=https&sig=qZUEwH3RA8e...

Edit linked service

🖼 Azure Blob Storage Learn more 🔗

Name *

AzureBlobStorage1

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime ⌄

Authentication type

SAS URI ⌄

| **SAS URI** | **Azure Key Vault** |

SAS URL * ⓘ

https://adfcookbookv2sa.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=co&sp=rwdla

| **SAS token** | **Azure Key Vault** |

SAS token ⓘ

sample: ?sv=<storage services version>&st=<start time>&se=<expire time>&sr=<resou

Test connection ⓘ

◉ To linked service ○ To file path

Annotations

➕ New

〉 Parameters

✅ Connection successful

🔌 Test connection

Apply   Cancel

Home > adfcookbookv2sa | Access Control (IAM) >

Add role assignment ...

Role   Members   Conditions (optional)   Review + assign

Selected role	Storage Blob Data Contributor
Assign access to	○ User, group, or service principal
	● Managed identity
Members	+ Select members

Name	Object ID	Type	
adfcookbookch5	842c44b8-f353-41f2-9ed7-56b8f5aff87a	Data factory (V2) ⓘ	🗑

Description   Optional

Details   ↻ Refresh

Snowflake	Succeeded →
	Azure Blob Storage Region: East US
	Succeeded →
	Azure Blob Storage Region: East US

Data read: ⓘ	122.356 KB	Data written: ⓘ	58.025 KB
Rows read:	3,683	Files written: ⓘ	1
		Rows written: ⓘ	3,683
		Peak connections: ⓘ	1

Copy duration	00:00:21
Throughput: ⓘ	17.479 KB/s

∨ Snowflake → Azure Blob Storage

Start time	6/7/2023, 10:24:11 AM
Used DIUs ⓘ	2
Used parallel copies ⓘ	1
∨ **Duration**	00:00:13

Details	Working duration	Total duration
✔ **Queue** ⓘ		00:00:07
✔ **Transfer** ⓘ		00:00:04

∨ Azure Blob Storage → Azure Blob Storage

Start time	6/7/2023, 10:24:25 AM
Used DIUs ⓘ	4
Used parallel copies ⓘ	1
∨ **Duration**	00:00:08

Details	Working duration	Total duration

Chapter 6: Data Migration – Azure Data Factory and Other Cloud Services

New connection

Amazon S3　Learn more [↗]

Name *
```
AmazonS31
```

Description
```

```

Connect via integration runtime * ⓘ
```
AutoResolveIntegrationRuntime                    ∨
```

Authentication type
```
Access key                                       ∨
```

| Access key ID | Azure Key Vault |

Access key ID *
```
AKIA3GYJ2UG4CPIHX67D
```

| Secret access key | Azure Key Vault |

Secret access key *
```
••••••••••••••••••••••••••••••••
```

Service URL ⓘ
```
https://s3.amazonaws.com
```

Test connection ⓘ
(●) To linked service   ( ) To file path

Annotations
＋ New

∨ **Parameters**
＋ New

〉 **Advanced** ⓘ

✓ Connection successful

| Create | Back | 🔌 Test connection   Cancel

File format settings

File format ⓘ

| DelimitedText | ⌄ |

[Detect text format] [👓 Preview data]

Column delimiter

| Semicolon (;) | ⌄ |

☐ Edit

Row delimiter

| Line feed (\n) | ⌄ |

☐ Edit

☑ First row as header ⓘ

⌄ Advanced

Compression type

| None | ⌄ |

Additional columns ⓘ

╋ New

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type | 🔲 Azure Blob Storage | ⌄ |

Connection * | 🔲 AzureBlobStorage1 | ⌄ | ✏ Edit ╋ New connection

Folder path *

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

| datafroms3 | 📁 Browse

File name

| |

Copy behavior ⓘ

| None | ⌄ |

Max concurrent connections ⓘ

| |

Block size (MB) ⓘ

| |

Metadata ⓘ

╋ New

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- ⑤ **Review and finish**
  - Review
  - ○ Deployment

Summary

You are running pipeline to copy data from Amazon S3 to Azure Blob Storage.

Amazon S3 → Azure Blob Storage

Properties   ✎ Edit

Task name	pl_copy_data_from_s3
Task description	

Source   ✎ Edit

Connection name	AmazonS31
Dataset name	SourceDataset_lfx
Column delimiter	;
Row delimiter	
Escape character	\
Quote char	"
First row as header	true
Folder path	data

Destination   ✎ Edit

Connection name	AzureBlobStorage1
Dataset name	DestinationDataset_lfx
Column delimiter	,
Escape character	\
Quote char	"

Copy settings   ✎ Edit

Timeout	0.12:00:00
Retry	0
Retry interval (sec)	30
Secure output	false
Secure input	false

[‹ Previous] [Next ›]

Binary
CommonCrawlDataS3

Connection   Parameters

Linked service *	▣ AmazonS31 ▾	⟋ Test connection ✎ Edit + New Learn more ⧉
File path *	commoncrawl / Directory / File name	🗁 Browse ⎮ ⌄
Compression type	None ▾	

Binary
CommonCrawlDataADLS

| Connection | Parameters |

Linked service *   [▣ AzureDataLakeStorage   ∨]   ✎ Test connection   ✎ Edit   + New   Learn more ⬏

File path *   [commoncrawl]   /   [@dataset().DirName]   /   [🗀 File name]   📁 Browse | ∨

Compression type   [None   ∨]

| Parameters | Variables | Settings | Output |

+ New | 🗑 Delete

	Name	Type	Default value	
☐	[Path]	[String ∨]	[Value]	🗑

Execute Pipeline ⬏

⚡ Execute Pipeline

🗑 {} 📄 ➡

| General | Settings | User properties |

Invoked pipeline *   [pl_Common Crawl_data_inner   ∨]   ✎ Open   + New

Wait on completion   ☐

Parameters

Name	Type	Value
Path	string	[@item().Path]

Activity runs
Pipeline run ID 911e527a-128e-492e-80e4-33f0cfb1ee14

All status ∨ List ∨   ⬇ Export to CSV | ∨
Showing 1 - 10 items

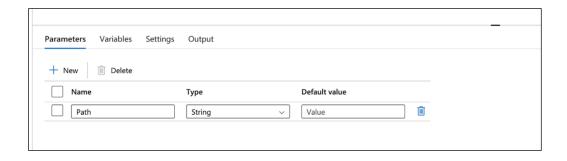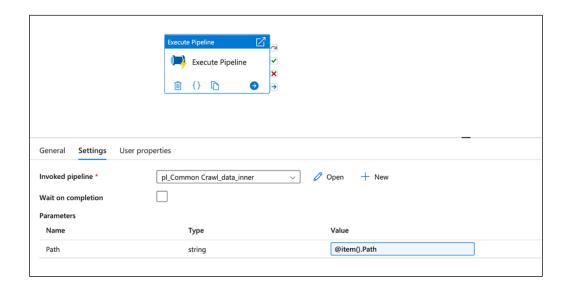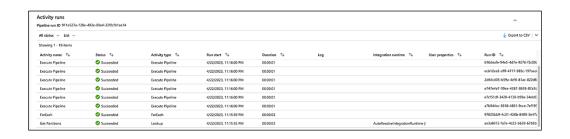| Activity name ↑↓	Status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log	Integration runtime ↑↓	User properties ↑↓	Run ID ↑↓
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				616deafe-94e5-4d7e-9276-f3c00a
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				ecb12ea3-dfff-4717-985c-197aac
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				2d64cd05-b59a-4ef6-81ac-822d8
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				e747e4e7-99ee-4267-8676-87a7c
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				e7cf51df-3420-4138-b99e-34eb0
Execute Pipeline	✅ Succeeded	Execute Pipeline	4/22/2023, 11:16:00 PM	00:00:01				a7b944cc-3618-4861-9cce-7ef19
ForEach	✅ Succeeded	ForEach	4/22/2023, 11:15:59 PM	00:00:03				97625bb9-4c31-428b-8409-5e47c
Get Partitions	✅ Succeeded	Lookup	4/22/2023, 11:15:55 PM	00:00:03		AutoResolveIntegrationRuntime ()		ae3a8912-fa7e-4c23-b639-67b5b

Settings

The Interoperability API allows Google Cloud Storage to interoperate with tools written for other cloud storage systems. This enables you to run migrations to Cloud Storage ⬀ and to authenticate both user and service accounts using keyed-hash message authentication codes (HMAC). Learn more ⬀

Request endpoint

Make sure the request endpoint in the tools or libraries you use with other cloud storage systems (e.g., Amazon S3) uses the Cloud Storage URI.

Storage URI

```
https://storage.googleapis.com
```
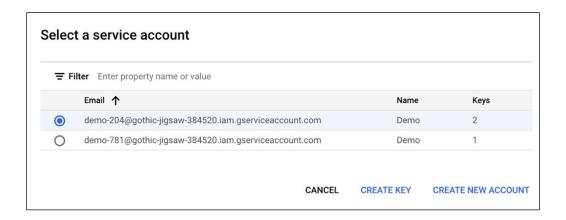
Service account HMAC

Use access keys with your organization's Cloud Platform service accounts when you don't want to tie HMAC authentication to specific user accounts. Recommended for production workloads. Learn more ⬀
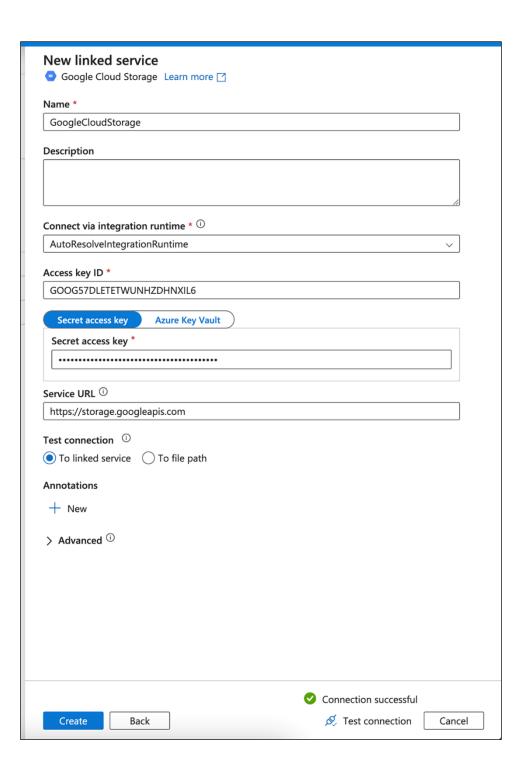
- Each service account can use up to five keys.
- Note that keys must be deactivated before they can be deleted.
- Grant your service accounts the required permissions for their intended operations – typically this is the IAM Storage Object Admin role.
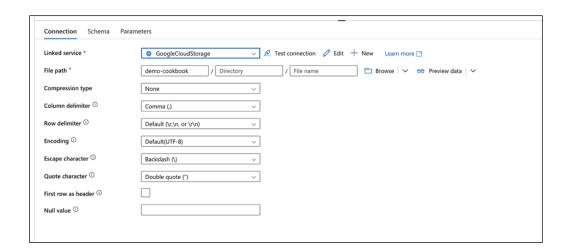
Access keys for service accounts
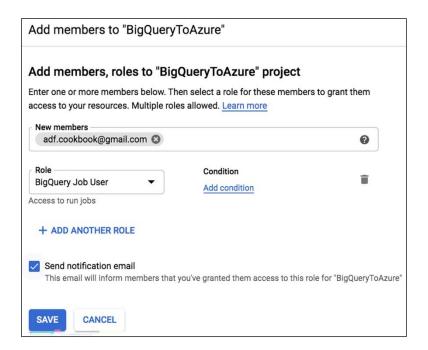
This project doesn't have any service account HMAC keys.
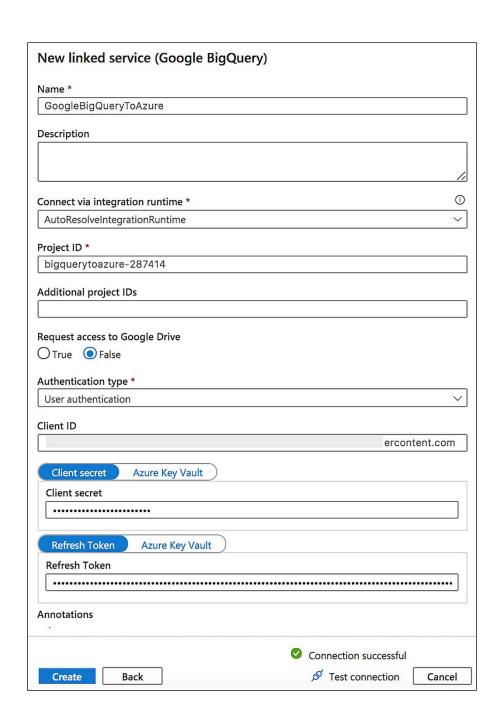
+ CREATE A KEY FOR A SERVICE ACCOUNT

Select a service account

≡ Filter   Enter property name or value

	Email ↑	Name	Keys
◉	demo-204@gothic-jigsaw-384520.iam.gserviceaccount.com	Demo	2
○	demo-781@gothic-jigsaw-384520.iam.gserviceaccount.com	Demo	1

CANCEL   CREATE KEY   CREATE NEW ACCOUNT

New linked service

⬡ Google Cloud Storage Learn more ⧉

Name *

GoogleCloudStorage

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime ⌄

Access key ID *

GOOG57DLETETWUNHZDHNXIL6

Secret access key | **Azure Key Vault**

Secret access key *

••••••••••••••••••••••••••••••

Service URL ⓘ

https://storage.googleapis.com

Test connection ⓘ
◉ To linked service   ○ To file path

Annotations

➕ New

❯ **Advanced** ⓘ

✓ Connection successful

Create Back   ⚡ Test connection Cancel

New linked service (Google BigQuery)

Name *

GoogleBigQueryToAzure

Description

Connect via integration runtime *   ⓘ

AutoResolveIntegrationRuntime   ⌄

Project ID *

bigquerytoazure-287414

Additional project IDs

Request access to Google Drive

◯ True   ⦿ False

Authentication type *

User authentication   ⌄

Client ID

ercontent.com

(**Client secret**   Azure Key Vault)

Client secret

••••••••••••••••••••

(**Refresh Token**   Azure Key Vault)

Refresh Token

••

Annotations

.

✅ Connection successful

Create   Back   🔌 Test connection   Cancel

Connection   Parameters

                        ✅ Connection successful

Linked service *   🔵 GoogleBigQueryToAzure2   ⌄   🔌 Test connection   ✏️ Edit   + New

Table   bigquery-public-data   .   github_repos.sample_files   👓 Preview data

                 ☑ Edit   Add dynamic content [Alt+P]

Chapter 7: Extending Azure Data Factory with Logic Apps and Azure Functions

When a HTTP request is received

Create a pipeline run

Initialize variable

Do ∧

Get a pipeline run

Set variable

Until ∧

Add an action

🔍 copy blob ✕
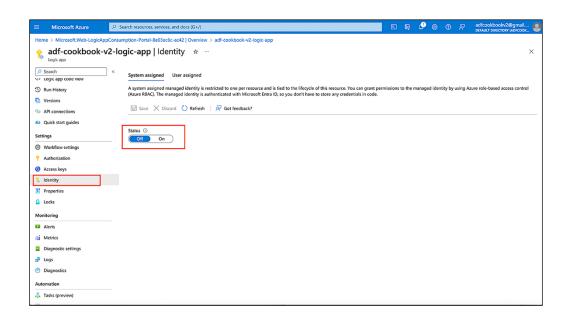
Runtime

Select a runtime ▾

Action Type

Actions ▾

☑ Group by Connector

Azure Blob Storage   ⓘ

See more

Copy blob (V2)   ⓘ

Azure Blob Storage   ⓘ
In App

See more

Copy blob from source location to destination   In App   ⓘ

Upload blob to storage container   In App   ⓘ

Elfsquad Product Configurator   ⓘ

See more

Copy quotation   ⓘ

Chapter 8: Microsoft Fabric and Power BI, Azure ML, and Cognitive Services

Copy data

Choose data source ✓

Connect to data source
Select, preview, and choose the data.

Choose data destination ○

Connect to data destination ○

Review + save ○

File format

DelimitedText ▾

[Detect text format] [⊖ Preview data]

Column delimiter

Comma (,) ▾

Row delimiter

Default (\r,\n, or \r\n) ▾

☑ First row as header ⓘ

Compression type

None ▾

[Back] [Next] [Cancel]

Copy data

Choose data source ✓

Connect to data source ✓

Choose data destination
Define the data store as destination.

Connect to data destination ○

Review + save ○

Data destinations

All categories Workspace Azure Database Generic protocol Services and apps

🔍 Search

Azure Blob Storage Azure	**Azure Cosmos DB for NoSQL** Azure	**Azure Data Explorer** Azure
Azure Data Lake Storage Gen1 Azure	**Azure Data Lake Storage Gen2** Azure	**Azure Database for PostgreSQL** Azure
Azure SQL Database Azure	**Azure SQL Database Managed Instance** Azure	**Azure Synapse Analytics** Azure
Azure Table Storage Azure	**Data Warehouse** Workspace	**Dataverse** Services and apps
Dynamics CRM Services and apps	**KQL Database** Workspace	**Lakehouse** Workspace
REST Generic protocol	**Snowflake** Services and apps	**SQL server** Database

[Back] [Next] [Cancel]

Copy data

- ✓ Choose data source
- ✓ Connect to data source
- ● **Choose data destination**
  Define the data store as destination.
- ○ Connect to data destination
- ○ Review + save

Lakehouse
Learn more

○ Existing Lakehouse   ○ Create new Lakehouse

Lakehouse *

[⌂ ADFCookbookV2_LH ⌄]   ↻ Refresh

Back   Next   Cancel

Copy data into Lakehouse

- ✓ Choose data source
- ✓ Connect to data source
- ✓ Choose data destination
- ● **Connect to data destination**
  Select and map to folder path or table.
- ○ Review + save

Root folder
○ Tables   ● Files

Folder path
If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

[data]   📁 Browse

File name
[Customers.csv]

Copy behavior ⓘ
[None ⌄]

Back   Next   Cancel

```
1   # Welcome to your new notebook
2   # Type here in the cell editor to add code!
3   table_name = "Customers"
```
PySpark (Python)

+ Code   + Markdown


```
1   # Welcome to your new notebook
2   # Type here in the cell editor to add code!
3   table_name = "Customers"
```
Parameters PySpark (Python)

```
1   from pyspark.sql.functions import *
2
3   # Read the Customers.csv file in Files/data folder
4   df = spark.read.format("csv").option("header","true").load("Files/data/Customers.csv")
5
6   # Split the CustomerName column into FirstName and LastName columns
7   df = df.withColumn("FirstName", split(col("CustomerName"), " ").getItem(0)).withColumn("La
8
9   # Include FirstName and LastName columns into the dataframe
10  # Reorder Columns in the data frame
11  df = df["CustomerId", "FirstName", "LastName", "Email", "PhoneNum", "Address1", "Address2"
12
13  # Load dataframe into the table
14  df.write.format("delta").mode("append").saveAsTable(table_name)
15
```
PySpark (Python)

Preview file data

URL: https://raw.githubusercontent.com/silverrosepetal/data/main/Orders.csv

File origin	Delimiter	Data type detection
65001: Unicode (UTF-8)	Comma	Based on first 200 rows

OrderId	Item	Quantity	CustomerId	OrderDate	DispatchDate	Status	CreatedAt	UpdatedAt
ACV2-SiUF-8485	Graphics Card	13	794717	11/10/2022, 11:20:04 ...	9/14/2022, 11:20:04 AM	Delivered	11/25/2022, 11:20:04 ...	10/14/2022, 11:20:04 ...
ACV2-ejpC-7532	Monitor	34	794717	4/8/2023, 11:20:04 AM	4/26/2023, 11:20:04 AM	Returned	6/30/2023, 11:20:04 AM	12/18/2022, 11:20:04 ...
ACV2-ttaM-4143	Keyboard	25	322808	1/23/2023, 11:20:04 AM	5/8/2023, 11:20:04 AM	Delivered	6/12/2023, 11:20:04 AM	11/27/2022, 11:20:04 ...
ACV2-ePYL-25...	Mouse	11	970802	5/5/2023, 11:20:04 AM	8/19/2022, 11:20:04 AM	Returned	1/30/2023, 11:20:04 AM	4/25/2023, 11:20:04 AM
ACV2-qCzd-25...	RAM	37	859145	1/9/2023, 11:20:04 AM	1/7/2023, 11:20:04 AM	Delivered	1/6/2023, 11:20:04 AM	7/14/2023, 11:20:04 AM
ACV2-lfRo-7654	USB Cable	33	859145	4/23/2023, 11:20:04 AM	5/18/2023, 11:20:04 AM	Returned	10/10/2022, 11:20:04 ...	7/1/2023, 11:20:04 AM
ACV2-pdDK-7...	Printer	17	859145	2/17/2023, 11:20:04 AM	3/13/2023, 11:20:04 AM	Delivered	3/10/2023, 11:20:04 AM	12/23/2022, 11:20:04 ...
ACV2-gdqy-34...	Hard Drive	24	859145	7/3/2023, 11:20:04 AM	10/6/2022, 11:20:04 AM	Returned	2/28/2023, 11:20:04 AM	1/14/2023, 11:20:04 AM
ACV2-sGkL-10...	Modem	12	859145	3/29/2023, 11:20:04 AM	11/10/2022, 11:20:04 ...	Returned	10/29/2022, 11:20:04 ...	12/21/2022, 11:20:04 ...
ACV2-jmHY-7...	Modem	34	859145	9/7/2022, 11:20:04 AM	2/12/2023, 11:20:04 AM	In transit	4/13/2023, 11:20:04 AM	6/27/2023, 11:20:04 AM
ACV2-iZkX-1542	Laptop	28	859145	3/20/2023, 11:20:04 AM	12/4/2022, 11:20:04 AM	Dispatched	1/13/2023, 11:20:04 AM	2/20/2023, 11:20:04 AM
ACV2-fqTz-1463	RAM	39	859145	8/21/2023, 11:20:04 AM	3/8/2023, 11:20:04 AM	Delivered	3/17/2023, 11:20:04 AM	10/29/2022, 11:20:04 ...
ACV2-UJbm-7...	Ethernet Cable	32	859145	11/2/2022, 11:20:04 AM	8/28/2022, 11:20:04 AM	Dispatched	5/27/2023, 11:20:04 AM	9/8/2022, 11:20:04 AM
ACV2-PQIJ-04...	RAM	23	859145	10/11/2022, 11:20:04 ...	7/14/2023, 11:20:04 AM	In transit	4/20/2023, 11:20:04 AM	8/31/2022, 11:20:04 AM
ACV2-FLvU-14...	RAM	18	859145	12/27/2022, 11:20:04 ...	7/16/2023, 11:20:04 AM	Dispatched	3/21/2023, 11:20:04 AM	2/28/2023, 11:20:04 AM
ACV2-fOxy-0717	Printer	7	859145	2/26/2023, 11:20:04 AM	10/5/2022, 11:20:04 AM	Returned	1/29/2023, 11:20:04 AM	10/30/2022, 11:20:04 ...
ACV2-oEBr-64...	SSD	23	859145	4/21/2023, 11:20:04 AM	6/4/2023, 11:20:04 AM	Returned	7/28/2022, 11:20:04 AM	10/2/2022, 11:20:04 AM
ACV2-kMzX-39...	Ethernet Cable	30	859145	6/3/2023, 11:20:04 AM	3/14/2023, 11:20:04 AM	Delivered	9/22/2022, 11:20:04 AM	5/30/2023, 11:20:04 AM
ACV2-DBUP-2...	Firewall	20	859145	4/14/2023, 11:20:04 AM	9/16/2022, 11:20:04 AM	Returned	4/10/2023, 11:20:04 AM	12/22/2022, 11:20:04 ...
ACV2-DqKh-3...	Keyboard	24	868390	1/20/2023, 11:20:04 AM	4/28/2023, 11:20:04 AM	In transit	8/25/2022, 11:20:04 AM	10/11/2022, 11:20:04 ...
ACV2-UcSZ-99...	USB Cable	12	868390	7/26/2023, 11:20:04 AM	9/1/2022, 11:20:04 AM	In transit	8/10/2022, 11:20:04 AM	9/14/2022, 11:20:04 AM
ACV2-IUYI-5073	Printer	11	868390	3/7/2023, 11:20:04 AM	4/29/2023, 11:20:04 AM	Dispatched	2/9/2023, 11:20:04 AM	11/6/2022, 11:20:04 AM

Back   Cancel   Add table using examples   Create

Screenshot 1 — Power Query (Load Order Data)

Load Order Data

Power Query

Home | Transform | Add column | View | Help

Search (Alt + Q)

Column from examples | Custom column | Invoke custom function | Conditional column | Index column | Rank column | Duplicate column | Format | Merge columns | Extract | Parse | Cluster values | Statistics | Standard | Scientific | Trigonometry | Rounding | Information | Date | Time | Duration

General | From text | From number | Date and time column

Queries [1] ①
- Orders

Orders ⑦
Source — Promoted headers — Changed column...

100%

```
1  let
2      Source = Csv.Document(Web.Contents("https://raw.githubusercontent.com/silverrosepetal/data/main/Orders.csv"), [Delimiter =
       ", Columns = 9, QuoteStyle = QuoteStyle.None]),
3      #"Promoted headers" = Table.PromoteHeaders(Source, [PromoteAllScalars = true]),
```

②

Tab moves focus.

rId	Item	Quantity	CustomerId	OrderDate	DispatchDate	Status	CreatedAt	UpdatedAt
1	Es-7603	Ethernet Cable	21	1932693	1/16/2023, 11:20:04 AM	3/5/2023, 11:20:04 AM	Delivered	6/4/2023, 11:20:04 AM	2/1/2023, 11:20:04 AM
2	X-8440	Desktop Comput...	39	1932693	12/23/2022, 11:20:04 AM	9/24/2022, 11:20:04 AM	Delivered	9/23/2022, 11:20:04 AM	12/29/2022, 11:20:04 AM
3	H-0981	Graphics Card	14	1932693	9/3/2023, 11:20:04 AM	1/9/2023, 11:20:04 AM	In transit	11/13/2023, 11:20:04 AM	10/16/2022, 11:20:04 AM
4	o-79...	Monitor	25	1932693	3/31/2023, 11:20:04 AM	5/23/2023, 11:20:04 AM	In transit	1/2/2023, 11:20:04 AM	3/23/2023, 11:20:04 AM
5	W-55...	RAM	12	1932693	1/13/2023, 11:20:04 AM	6/13/2023, 11:20:04 AM	Delivered	11/12/2022, 11:20:04 AM	6/7/2023, 11:20:04 AM
6	Oc-66...	SSD	21	1932693	5/5/2023, 11:20:04 AM	11/14/2022, 11:20:04 ...	Returned	2/12/2023, 11:20:04 AM	12/6/2022, 11:20:04 AM
7	vK-69...	Printer	6	1932693	8/23/2022, 11:20:04 AM	2/16/2023, 11:20:04 AM	Delivered	7/22/2023, 11:20:04 AM	12/19/2022, 11:20:04 AM
8	vT-77...	Printer	17	1932693	3/26/2023, 11:20:04 AM	11/7/2022, 11:20:04 AM	Returned	8/21/2022, 11:20:04 AM	3/4/2023, 11:20:04 AM
9	ic-3271	Router	19	1932693	12/23/2022, 11:20:04 ...	10/12/2022, 11:20:04 ...	Dispatched	8/14/2022, 11:20:04 AM	7/4/2023, 11:20:04 AM
10	G-0877	Keyboard	3	1932693	9/30/2022, 11:20:04 AM	6/27/2023, 11:20:04 AM	Dispatched	2/18/2023, 11:20:04 AM	9/2/2022, 11:20:04 AM
11	mS-8...	Laptop	33	1932693	10/5/2022, 11:20:04 AM	2/22/2023, 11:20:04 AM	Dispatched	4/15/2023, 11:20:04 AM	11/29/2022, 11:20:04 ...
12	QK-42...	Wi-Fi Extender	30	90862	4/4/2023, 11:20:04 AM	9/12/2022, 11:20:04 AM	Delivered	7/6/2023, 11:20:04 AM	11/8/2022, 11:20:04 AM
13	Oa-10...	Laptop	33	90862	9/8/2022, 11:20:04 AM	12/23/2022, 11:20:04 ...	Dispatched	1/27/2023, 11:20:04 AM	11/17/2022, 11:20:04 ...
14	V-3121	Keyboard	10	1817154	8/12/2022, 11:20:04 AM	9/4/2022, 11:20:04 AM	In transit	7/22/2023, 11:20:04 AM	8/1/2022, 11:20:04 AM
15	u-48...	Printer	37	1817154	10/20/2022, 11:20:04 ...	4/3/2023, 11:20:04 AM	Delivered	2/18/2023, 11:20:04 AM	4/26/2023, 11:20:04 AM
16	er-3499	Network Hub	27	1817154	5/26/2023, 11:20:04 AM	11/15/2022, 11:20:04 ...	Returned	9/9/2022, 11:20:04 AM	6/23/2023, 11:20:04 AM
17	-7089	Hard Drive	29	1817154	6/2/2023, 11:20:04 AM	1/17/2023, 11:20:04 AM	Dispatched	10/14/2022, 11:20:04 ...	3/2/2023, 11:20:04 AM
18	v-7360	Router	19	1817154	11/8/2022, 11:20:04 AM	8/8/2022, 11:20:04 AM	In transit	11/25/2022, 11:20:04 ...	8/21/2022, 11:20:04 AM

③

Completed (1.44 s) Columns: 9 Rows: 99+

⑥ Query

Query settings ›
Properties
Name ④
Orders
Entity type
Custom

▾ Applied steps
- Source
- Promoted h...
- Changed co...

▾ Data destination ⑤
Lakehouse

Publish

Screenshot 2 — Custom column dialog

Power Query

Home | Transform | Add column | View | Help

Search (Alt + Q)

Queries [1]

Orders
Source — Promoted headers — Changed column...

```
1  let
2      Source = Csv.Document(Web.Content...
       QuoteStyle = QuoteStyle.None]),
3      #"Promoted headers" = Table.Promo...
```

Tab moves focus.

Custom column ?

Add a column that is computed from other columns or values.

New column name *
ImportedAt

Data type
Date/Time

Custom column formula *
```
= DateTime.FixedLocalNow()
```

Available column(s)
- OrderId
- Item
- Quantity
- CustomerId
- OrderDate
- DispatchDate
- Status

Insert column

Learn more about Power Query formulas

Cancel | OK

	OrderId	Item	Quantity
1	ACV2-BeEs-7603	Ethernet Cable	
2	ACV2-TiCX-8440	Desktop Comput...	
3	ACV2-2YuH-0981	Graphics Card	
4	ACV2-Vgoo-79...	Monitor	
5	ACV2-LaEW-55...	RAM	
6	ACV2-IQOc-66...	SSD	
7	ACV2-qQvK-69...	Printer	
8	ACV2-ezWT-77...	Printer	
9	ACV2-iqGc-3271	Router	
10	ACV2-jlFG-0877	Keyboard	
11	ACV2-dQmS-8...	Laptop	
12	ACV2-qhQK-42...	Wi-Fi Extender	30
13	ACV2-gZOa-10...	Laptop	33

Screenshot 3 — Dataflow Gen2

Load Order Data

Refresh now

Dataflow Gen2

Set up pipeline job

Basics

Experiment name

◯ Select existing   ⦿ Create new

New experiment name *

Experiment1

Job display name

Recommendation - Movie Rating Tweets

Job description

This sample shows how to use the recommendation modules to train a movie recommender engine.

Job tags

| Name | : | Value | Add |

Set up pipeline job

Runtime settings

Default compute ⓘ

Select compute type

Compute cluster

Select Azure ML compute cluster

ADF-Cookbook-Cluster1

🖥 Create Azure ML compute cluster   ↻ Refresh Compute

Default datastore ⓘ

Select datastore *

workspaceblobstore

Advanced settings

☑ Continue on step failure ⓘ

Set up published pipeline ✕

PipelineEndpoint

○ Select existing

● Create new

New PipelineEndpoint name *

Recommendation - Movie Rating Tweets

PipelineEndpoint description (optional)

|

Published pipeline *

Recommendation - Movie Rating Tweets 02-04-2024-06-55

☑ Set as default pipeline for this endpoint.
☑ Continue on failure step
Pipeline inputs and default values

ⓘ No pipeline inputs are specified. Set a data node as a pipeline input to use different data input when invoking this endpoint.

Publish   Cancel

Pipeline endpoint overview ✕

Overview   Settings

Properties

Status
Active

REST endpoint
https://southindia.api.azureml.ms/pip
elines/v1.0/subscriptions/3c0cb7fd-
886b-4335-bb72-
bae0229792be/resourceGroups/revie
w-ch07-
rg/providers/Microsoft.MachineLearni
ngServices/workspaces/adf-
cookbook-v2-ML-
workspace/PipelineRuns/PipelineEnd
pointSubmit/Id/58934cb0-663a-
45c9-9640-864d99ee99ef

REST endpoint documentation
https://southindia.api.azureml.ms/pip
elines/swagger/pipelineendpointsub
mit/swagger.json

Published by
ADF Cookbook

Date published
Feb 4, 2024 6:57 PM

PipelineEndpoint ID
58934cb0-663a-45c9-9640-
864d99ee99ef

Default version
0

Default published pipeline
1ac3425a-c8e2-44b5-995d-

New linked service

🧪 Azure Machine Learning   Learn more ⬀

Name *

| AzureMLService1 |

Description

| |

Connect via integration runtime * ⓘ

| AutoResolveIntegrationRuntime | ⌄ |

Authentication method

| System Assigned Managed Identity | ⌄ |

Azure Machine Learning workspace selection method ⓘ

⦿ From Azure subscription   ◯ Enter manually

Azure subscription ⓘ

| Azure subscription 1 (▉▉▉▉▉▉▉▉▉▉▉▉▉) | ⌄ |

Azure Machine Learning workspace name *

| adfcookbook-v2-ml | ⌄ |

Managed identity name: **adf-cookbook-v2-ch07-df**
Managed identity object ID: **5e207c26-76e0-4d9f-8fdc-d6c0a1f39deb**
Grant Data Factory service managed identity access to your Azure Machine Learning.
Learn more ⬀

Annotations

╋ New

❯ **Advanced** ⓘ

| **Create** | Cancel   🔌 Test connection

General **Settings** User properties

Azure Machine Learning linked service * ⓘ   ⚗ AzureML_adf_cookbook_workspace ⌄   ✐ Test connection   ✎ Edit   + New

Machine Learning pipeline ID type   ⦿ Pipeline ID   ○ Pipeline endpoint ID

Machine Learning pipeline name * ⓘ   Recommendation - Movie Rating Tw... ⌄

Machine Learning pipeline ID * ⓘ   1ac3425a-c8e2-44b5-995d-0a54d1c... ⌄   ⚭ Open in Azure Portal

Experiment name ⓘ   [   ]

Add dynamic content [Alt+Shift+D]

> Machine Learning pipeline parameters

> Machine Learning data path assignments

Machine Learning parent run ID ⓘ   [   ]

Continue on step failure ⓘ   ☐

◗ Run Azure ML Exp... ●

✓ Validate   ⊘ Cancel options   ⚡ Add trigger   {}   ▤   ⋯

Machine Learning Execute.. →
⚗ Execute Azure ML Experiment ✓

Parameters   Variables   Settings   **Output**   ⌃

Pipeline run ID: **e650bbc3-28df-4635-89ac-6f3166119aaf** [@] 🔄 ⓘ

Expand toolbox pane   ↓ Export to CSV ⌄

Showing 1 - 1 of 1 items

Activity name ↑↓	Status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log
Execute Azure ML Experiment	● In progress	Machine Learning Execute	6/5/2023, 7:57:10 PM	00:00:22	

Azure AI | Machine Learning Studio   Azure subscription 1   adf-cookbook-v2-ML-workspace

Default Directory > adf-cookbook-v2-ML-workspace > Jobs

Jobs

All experiments   All jobs   All schedules

↻ Refresh   🗑 Archive experiment   ⚙ View options ⌄   ⦿ View archived experiments

🔍 Search   ▥ Columns

Experiment	☆	Latest job	Last submitted ↓	Created	Created by	Job types
Recommendation_-_Movie_Rating_Tweets_02-04-2024-06-5		silver_cushion_ysv89g4b	Feb 4, 2024 7:13 PM	Feb 4, 2024 7:13 PM	Service Principal	Pipeline
Experiment1		Recommendation - Movie Rating Tweets	Feb 4, 2024 6:51 PM	Feb 4, 2024 5:41 PM	ADF Cookbook	Pipeline

✓ Validate | ◉ Data flow debug ✓ | ⤷ Debug Settings

InputPatientData
Columns:
9 total

Source settings | Source options | **Projection** | Optimize | Inspect | Data preview ●

📋 Define default format | 🔍 Detect data type | ⟵ Import projection | ↻ Reset schema

Column name	Type	Format
age	abc string	Specify format
sex	abc string	Specify format
race	abc string	Specify format
ethnicity	abc string	Specify format
state	abc string	Specify format
county	abc string	Specify format
name	abc string	Specify format
email	abc string	Specify format
phone_number	abc string	Specify format

✓ Validate | ◉ Data flow debug ✓ | ⤷ Debug Settings

InputPatientData
Columns:
9 total

Source settings | Source options | Projection | Optimize | Inspect | **Data preview** ●

Number of rows ＋ INSERT 21 | ☀ UPDATE 0 | ✕ DELETE 0 | ✦ UPSERT 0 | 🔍 LOOKUP 0 | ❌ ERROR 0 | TOTAL N/A

↻ Refresh ∨ | Typecast ∨ | Modify ∨ | Map drifted | Statistics | ✕ Remove | ↓ Export to CSV ∨ | ⓘ

	Colum...	Colum...	Colum...	Colum...	Colum...	Colum...	Colum...	C
＋	age	sex	race	ethnicity	state	county	name	e
＋	15	Female	Black	Non-Hi...	CA	RAPIDES	Latoya J...	La
＋	65	Male	Multipl...	Missing	CA	RAPIDES	Miguel ...	M
＋	71	Female	Black	Non-Hi...	CA	LOS AN...	Tamika ...	Ta
＋	54	Male	Black	Non-Hi...	CA	LOS AN...	Tyrone ...	Ty
＋	21	Female	White	Non-Hi...	WA	KING	Charlot...	Ch
＋	33	Male	White	Non-Hi...	WA	SNOHO...	William...	W
＋	21	Male	Multipl...	Hispani...	TX	UPSHUR	Jesus R...	Je
＋	46	Female	White	Non-Hi...	WI	WINNE...	Adriana...	A
＋	41	Female	Multipl...	Missing	CA	LOS AN...	Ayumi Jo	A
＋	56	Male	Black	Non-Hi...	CA	LOS AN...	Jamal G...	Ja

New linked service

REST Learn more ⧉

Name *

AzureCognitiveSerices

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime   ⌄

Base URL *

https://adfcookbookv2ch07cogser.cognitiveservices.azure.com/

⚠ Information will be sent to the URL specified. Please ensure you trust the URL entered.

Authentication type *

Anonymous   ⌄

Server Certificate Validation ⓘ

◯ Enable   ⦿ Disable

Auth headers ⓘ

✛ New   🗑 Delete

☐	Name	Value
☐	Ocp-Apim-Subscription-Key ⌄	••••••••••••••••
		Add Azure Key Vault

Annotations

✛ New

❯ Parameters

| Create | Cancel | | 🔌 Test connection |

Chapter 9: Managing Deployment Processes with Azure DevOps

Microsoft Entra Connection
Connect your organization to a directory.

You are signed in as

AC

Microsoft Entra

Default Directory

Default Directory
Tenant Id: b269e661-3fcc-4f34-8ac4-a32126dce250

Cancel   Connect

Connection Failed

Your organization **adfcookbookv2** failed to connect to the **Default Directory** Microsoft Entra tenant.

⊗ User t[   ] is a guest in the target Microsoft Entra tenant Default Directory. The current organization policy does not allow guest users to access the organization. Change the policy setting to allow external guest access and try again.

Close

Configure a repository

⊘ Default Directory (b269e661-3fcc-4f34-8ac4-a32126dce250)

Specify the settings that you want to use when connecting to your repository.

◉ Select repository   ◯ Use repository link

Azure DevOps organization name * ⓘ

adfcookbook2	⌄

Project name * ⓘ

ADF Dev Ops	⌄

Repository name * ⓘ

ADF Devops	⌄

Collaboration branch * ⓘ

master

Publish branch * ⓘ

adf_publish	⌄

Root folder * ⓘ

/

Custom comment

☑ Use custom comment

Import existing resources

☑ Import existing resources to repository

Import resource into this branch ⓘ

	⌄

⟳ Applying   Back   Cancel

Create a key vault

Basics   Access configuration   Networking   Tags   Review + create

Azure Key Vault is a cloud service used to manage keys, secrets, and certificates. Key Vault eliminates the need for developers to store security information in their code. It allows you to centralize the storage of your application secrets which greatly reduces the chances that secrets may be leaked. Key Vault also allows you to securely store secrets and keys backed by Hardware Security Modules or HSMs. The HSMs used are Federal Information Processing Standards (FIPS) 140-2 Level 2 validated. In addition, key vault provides logs of all access and usage attempts of your secrets so you have a complete audit trail for compliance.

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 *	ADFBOOK-DEV-RG
	Create new

Instance details

Key vault name * ⓘ	ADFBOOK-DEV-AKV ✓
Region *	East US
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**

Create a key vault

Basics   **Access configuration**   Networking   Tags   Review + create

Configure data plane access for this key vault

To access a key vault in data plane, all callers (users or applications) must have proper aut... execute. Learn more

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 ...

+ Create   ✎ Edit   🗑 Delete

☐ Name ↑↓	Email ↑↓
∨ USER	
☐ Antonina Chernyshova	tonya.chernyshova_gmail.com#EXT#@

Previous   Next   Review + create

Create an access policy

✅ Permissions   ② Principal   ③ Application (optional)   ④ Review + create

Only 1 principal can be assigned per access policy.
Use the new embedded experience to select a principal. The previous popup experience can be accessed here. Select a principal

🔍 adf

🟦	ADFBOOK-DEV-ADF 62a6ffc8-42fa-4332-b838-cf5092cd2cca
🟦	ADFBOOK-UAT-ADF 324b12c4-0847-4983-97ea-fbdd7636c6b1

Selected item

🟦	ADFBOOK-DEV-ADF 62a6ffc8-42fa-4332-b838-cf5092cd2cca

Previous   **Next**

All pipelines > ⌄ New release pipeline

Pipeline   Tasks ⌄   Variables   Retention   Options   History

UAT
Deployment process

Agent job
Run on agent

ARM Template deploy...
ARM template deployment

Select a file or folder ✕

◢ 📁 Linked artifacts
  ◢ 📁 _ADF Devops (Azure Repos Git)
    ◢ 📁 ADFBOOK-DEV-ADF
      ▸ 📁 globalParameters
      ▸ 📁 linkedTemplates
        📄 ARMTemplateForFactory.json
        📄 ARMTemplateParametersForFactory.json
    ▸ 📁 Demo-data-factory-cookbook
      📄 readme.md

The artifacts published by each version will be available for deployment in release pipelines. The last successful version of **_ADF Devops (Azure Repos Git)** published the following artifacts: *ADFBOOK-DEV-ADF, Demo-data-factory-cookbook, readme.md.*

Location `_ADF Devops/ADFBOOK-DEV-ADF/ARMTemplateParametersForFact`

[OK]   [Cancel]

Location * ⓘ
[East US   ⌄] ↻

[⌄]

(...orkingDirectory)/_ADF Devops/ADFBOOK-DEV-
...orFactory.json   [...]

ⓘ
[   ] [...]

...ameters ⓘ
[   ] [...]

ⓘ
[⌄]

[   ]

Deployment outputs ⓘ
[   ]

☐ Access service principal details in override parameters ⓘ
☐ Use individual output values without JSON.Stringify applied ⓘ

New release pipeline   Save

Variables   Retention   Options   History

...   Create or update resource group

Override template parameters ✕

Name	Value
factoryName	"ADFBOOK-UAT-ADF"
ADLS_DEV_LS_propertie...	"asdfbookuatadsl-sercret"
AKV_DEV_LS_properties_...	"https://ADKCOOKBOOK-..."

[OK]   [Cancel]

deployment: Resource Gr...
ment

                    /_ADF DevOps/ADFBOOK-DEV-ADF/ARMTer...

                    /_ADF DevOps/ADFBOOK-DEV-ADF/ARMTer...

Advanced ⌃

Deployment name ⓘ

Visual Studio Code

Editing e

Start

New File...
Open...
Clone Git F
Connect to

Recent

b5datalake_qru
mlops-zoomca
Projects ~/De

Favorites
- Recents
- Applicati...
- Desktop
- Documents
- Downloads

Locations
- iCloud Dri...
- Subli...

Tags
- Red
- Orange
- Yellow
- Green
- Blue
- Purple

ADFDevOps

Name	Size	Kind	Date Added

New Folder   Cancel   Select as Repository Destination

Get Started with Jupyter Notebooks Updated

Restart Visual Studio Code to apply the latest update.
Update Now   Later   Release N

Show welcome page on startup

README.md — ADF DevOps VS

EXPLORER

ADF DEVOPS VS
- README.md

README.md > # Introduction

```
1   # Introduction
2   TODO: Give a short introduction of your project. Let this section explain the objectives or the
    motivation behind this project.
3
4   # Getting Started
5   TODO: Guide users through getting your code up and running on their own system. In this section you can
    talk about:
6   1.  Installation process
7   2.  Software dependencies
8   3.  Latest releases
9   4.  API references
10
11  # Build and Test
12  TODO: Describe and show how to build your code and run the tests.
13
14  # Contribute
15  TODO: Explain how other users and developers can contribute to make your code better.
16
17  If you want to learn more about creating good readme files then refer the following [guidelines](https://
    docs.microsoft.com/en-us/azure/devops/repos/git/create-a-readme?view=azure-devops). You can also seek
    inspiration from the below readme files:
18  - [ASP.NET Core](https://github.com/aspnet/Home)
19  - [Visual Studio Code](https://github.com/Microsoft/vscode)
20  - [Chakra Core](https://github.com/Microsoft/ChakraCore)
```

OUTLINE
TIMELINE

main   0 0   Ln 1, Col 1   Spaces: 4   UTF-8   CRLF   Markdown

(base) [redacted]-MacBook-Pro ADF DevOps VS % az login
A web browser has been opened at https://login.microsoftonline.com/organizations/oauth2/v2.0/authorize. Please continue the login in the web browser. If no web browser is available or if the web browser fails to open, use device code flow with `az login --use-device-code`.
[
  {
    "cloudName": "AzureCloud",
    "homeTenantId": "b269e661-3fcc-4f34-8ac4-a32126dce250",
    "id": "2381f82e-7229-48a8-a682-d6307a5be62f",
    "isDefault": true,
    "managedByTenants": [],
    "name": "Azure subscription 1",
    "state": "Enabled",
    "tenantId": "b269e661-3fcc-4f34-8ac4-a32126dce250",
    "user": {
      "name": "[redacted]",
      "type": "user"
    }
  }
]
(base) [redacted]-MacBook-Pro ADF DevOps VS % az datafactory create --location "East US" --name "ADFCookbook-VS" --resource-group "adfcookbookv2"
The command requires the extension datafactory. Do you want to install it now? The command will continue to run after the extension is installed. (Y/n):

ADFCookbook-VS.json — ADF DevOps VS

EXPLORER

ADF DEVOPS VS
- factory
  - ADFCookbook-VS.json
- publish_config.json
- README.md

factory > {} ADFCookbook-VS.json > {} identity > tenantId

```json
{
    "name": "ADFCookbook-VS",
    "location": "eastus",
    "identity": {
        "type": "SystemAssigned",
        "principalId": "207d4ac3-9d5a-4339-b25e-7f752bee9b72",
        "tenantId": "b269e661-3fcc-4f34-8ac4-a32126dce250"
    }
}
```

BlobStorage.json — ADF DevOps VS

EXPLORER

ADF DEVOPS VS
- dataset
  - Input_ds.json
  - Output_ds.json
- factory
  - ADFCookbook-VS.json
- linkedService
  - BlobStorage.json
- publish_config.json
- README.md

linkedService > {} BlobStorage.json > ...

```json
{
    "name": "BlobStorage",
    "type": "Microsoft.DataFactory/factories/linkedservices",
    "properties": {
        "annotations": [],
        "type": "AzureBlobStorage",
        "typeProperties": {
            "connectionString": "DefaultEndpointsProtocol=https;AccountName=adfcookbookv2;EndpointSuffix=core.windows.net;",
            "encryptedCredential": "ew0KICAiVmVyc2lvbiI6ICIyMDE3LTExLTMwIiwNCiAgIlByb3RlY3Rpb24iOiAiTktleSIsDQogICJTZWNyZXRDb..."
        }
    }
}
```

Input_ds.json — ADF DevOps VS

EXPLORER

{} Input_ds.json ×

ADF DEVOPS VS
∨ dataset
  {} Input_ds.json
  {} Output_ds.json
∨ factory
  {} ADFCookbook-VS.json
  > linkedService
  {} publish_config.json
  ① README.md

dataset > {} Input_ds.json > {} properties > {} typeProperties

```json
{
    "name": "Input_ds",
    "properties": {
        "linkedServiceName": {
            "referenceName": "BlobStorage",
            "type": "LinkedServiceReference"
        },
        "annotations": [],
        "type": "DelimitedText",
        "typeProperties": {
            "location": {
                "type": "AzureBlobStorageLocation",
                "fileName": "airlines.csv",
                "container": "data"
            },
            "columnDelimiter": ",",
            "escapeChar": "\\",
            "firstRowAsHeader": true,
            "quoteChar": "\""
        },
        "schema": [
            {
                "name": "ID",
                "type": "String"
            },
            {
                "name": "Name",
                "type": "String"
            },
            {
                "name": "Code",
                "type": "String"
            },
            {
                "name": "Country",
                "type": "String"
            }
        ]
    }
}
```


ADFCopyPipeline.json — ADF DevOps VS

EXPLORER

{} ADFCopyPipeline.json ×

ADF DEVOPS VS
∨ dataset
  {} Input_ds.json
  {} Output_ds.json
∨ factory
  {} ADFCookbook-VS.json
  > linkedService
∨ pipeline
  {} ADFCopyPipeline.json
  {} publish_config.json
  ① README.md

pipeline > {} ADFCopyPipeline.json > ...

```json
{
    "name": "ADFCopyPipeline",
    "properties": {
        "activities": [
            {
                "name": "Copy data1",
                "type": "Copy",
                "dependsOn": [],
                "policy": {
                    "timeout": "0.12:00:00",
                    "retry": 0,
                    "retryIntervalInSeconds": 30,
                    "secureOutput": false,
                    "secureInput": false
                },
                "userProperties": [],
                "typeProperties": {
                    "source": {
                        "type": "DelimitedTextSource",
                        "storeSettings": {
                            "type": "AzureBlobStorageReadSettings",
                            "recursive": true,
                            "enablePartitionDiscovery": false
                        },
                        "formatSettings": {
                            "type": "DelimitedTextReadSettings"
                        }
                    },
                    "sink": {
                        "type": "DelimitedTextSink",
                        "storeSettings": {
                            "type": "AzureBlobStorageWriteSettings"
                        },
                        "formatSettings": {
                            "type": "DelimitedTextWriteSettings",
                            "quoteAllText": true,
                            "fileExtension": ".txt"
                        }
                    },
                    "enableStaging": false,
                    "translator": {
                        "type": "TabularTranslator",
                        "typeConversion": true,
                        "typeConversionSettings": {
                            "allowDataTruncation": true,
                            "treatBooleanAsNumber": false
                        }
                    }
                },
                "inputs": [
                    {
                        "referenceName": "Input_ds",
                        "type": "DatasetReference"
```

Airflow environment setup

Use this interface to setup and create your Airflow integration runtime environment

ℹ️ Integration runtimes are published immediately to the Data Factory

Name * ⓘ

Airflow

Description

Enter description here...

Type

Airflow

Airflow auth type

🔘 Azure AD authentication ⭕ Basic

Region ⓘ

East US

Compute size * ⓘ

Large (scheduler:3, webserver:1, worker: 6)

Enable autoscale (Coming Soon) ☐

Extra nodes * ⓘ

0 — 0 ⇅

Airflow version * ⓘ

2.4.3

Enable git sync ⓘ ☐

Airflow configuration overrides ⓘ

＋ New

Environment variables ⓘ

＋ New

Airflow requirements ⓘ

apache-airflow-providers-microsoft-azure ✕

> Advanced

Create   Cancel

Airflow   DAGs   Datasets   Security ▾   Browse ▾   Admin ▾   Docs ▾   08:19 UTC ▾

Successfully connected to Azure Data Factory.

Add Connection

Connection Id *	airflow_adf
Connection Type *	Azure Data Factory ▾ Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.
Description	
Client ID	20b5af0a-97ff-4ea8-9fd8-a0d153937663
Secret	••••••••••••••••••••••••••••••••••
Tenant ID	b269e661-3fcc-4f34-8ac4-a32126dce250
Subscription ID	2381f82e-7229-48a8-a682-d6307a5be62f
Resource Group Name	adfcookbookv2
Factory Name	ADFCookbook-VS

Save 🖫   Test ➤   ←

```python
from datetime import datetime, timedelta

from airflow.models import DAG, BaseOperator

try:
    from airflow.operators.empty import EmptyOperator
except ModuleNotFoundError:
    from airflow.operators.dummy import DummyOperator as EmptyOperator  # type: ignore
from airflow.providers.microsoft.azure.operators.data_factory import AzureDataFactoryRunPipelineOperator
from airflow.providers.microsoft.azure.sensors.data_factory import AzureDataFactoryPipelineRunStatusSensor
from airflow.utils.edgemodifier import Label

with DAG(
    dag_id="example_adf_run_pipeline",
    start_date=datetime(2022, 5, 14),
    schedule_interval="@daily",
    catchup=False,
    default_args={
        "retries": 1,
        "retry_delay": timedelta(minutes=3),
        "azure_data_factory_conn_id": "airflow_adf", #This is a connection created on Airflow UI
        "factory_name": "ADFCookbook-VS",  # This can also be specified in the ADF connection.
        "resource_group_name": "adfcookbookv2",  # This can also be specified in the ADF connection.
    },
    default_view="graph",
) as dag:
    begin = EmptyOperator(task_id="begin")
    end = EmptyOperator(task_id="end")

    # [START howto_operator_adf_run_pipeline]
    run_pipeline1: BaseOperator = AzureDataFactoryRunPipelineOperator(
        task_id="run_pipeline1",
        pipeline_name="ADFCopyPipeline",
        parameters={"myParam": "value"},
    )
    # [END howto_operator_adf_run_pipeline]
```

Chapter 10: Monitoring and Troubleshooting Data Pipelines

Pipeline runs

Triggered Debug | ⟳ Rerun

▽ Filter by run ID or name | Loc

Runs : Latest runs | Triggered b

Showing 1 - 7 items

☐ **Pipeline name** ↑↓

☐ pl_orchestration_copy_from_mult

☐ pl_orchestration_copy_from_mult

☐ pl_orchestration_copy_from_mult

☐ pl_orchestration_recipe5_execute

☐ pl_orchestration_recipe_1_param

☐ pl_orchestration_troubleshooting

☐ pl_orchestration_recipe_1

Pipeline runs

Triggered Debug | ⟳ Rerun | ⊘ Cancel options ⌄ | ⟳ Refresh | ☰ Edit columns | **List** Gantt

▽ Filter by run ID or name | Local time : **Last 24 hours** | Pipeline name : **All** | Status : **All** | Runs : Latest runs

Triggered by : **All** | ▽ Add filter | ✕   ⎙ Copy filters   ⬇ Export to CSV ⌄

Showing 1 - 8 items   Last refreshed 3 minutes ago

	Pipeline name ↑↓	Status ↑↓	Duration	Run start ↑↓	Run end ↑↓	Run	Parameters	Annotations
☐	pl_orchestration_copy_from_mult_...	⊕ In progress	3s	8/14/2023, 7:04:23 PM	--	Original		
☐	pl_orchestration_copy_from_mult_...	✓ Succeeded	1m 14s	8/14/2023, 6:49:38 PM	8/14/2023, 6:50:51 PM	Original		
☐	pl_orchestration_copy_from_mult_...	✓ Succeeded	36s	8/14/2023, 6:47:14 PM	8/14/2023, 6:47:50 PM	Original		
☐	pl_orchestration_copy_from_mult_...	✓ Succeeded	38s	8/14/2023, 6:46:03 PM	8/14/2023, 6:46:40 PM	Original		
☐	pl_orchestration_recipe5_execute_...	✓ Succeeded	50s	8/14/2023, 6:45:56 PM	8/14/2023, 6:46:45 PM	Original		
☐	pl_orchestration_recipe_1_parame...	✓ Succeeded	15s	8/14/2023, 6:45:50 PM	8/14/2023, 6:46:04 PM	Original		
☐	pl_orchestration_troubleshooting	✓ Succeeded	1m 5s	8/14/2023, 6:45:43 PM	8/14/2023, 6:46:47 PM	Original		
☐	pl_orchestration_recipe_1	✓ Succeeded	19s	8/14/2023, 6:38:27 PM	8/14/2023, 6:38:45 PM	Original		

Pipeline name ↑↓	Status ↑↓	Duration	Run start ↑↓
pl_orchestration_copy_from_mult_sources	🔵 In progress	3s	8/14/2023, 7:04:23 PM
pl_orchestration_copy_from_mult... ↻ 📄	✅ Succeeded	1m 14s	8/14/2023, 6:49:38 PM
pl_orchestration_copy_from_mult_sources	✅ Succeeded	36s	8/14/2023, 6:47:14 PM
pl_orchestration_copy_from_mult_sources	✅ Succeeded	38s	8/14/2023, 6:46:03 PM
pl_orchestration_recipe5_execute_pipeline	✅ Succeeded	50s	8/14/2023, 6:45:56 PM

Pipeline run consumption

Name
pl_orchestration_copy_from_mult_sources

Status
✅ Succeeded

Run ID
bfdea7ba-d5c2-4774-9c86-e608f2f5df38

	Quantity	Unit
Pipeline orchestration		
Activity runs	2	Activity runs
Pipeline execution		
Azure integration runtime		
Data movement activities	0.0667	DIU-hour

Learn more 🗗

Pricing calculator 🗗

Close

🔄 Refresh ✏️ Edit **Details** Activities

STATUS ❌	TYPE	SUB-TYPE	VERSION
Unavailable ⓘ	Self-Hosted	Shared	4.11.7521.1

RUNNING / REGISTERED NODE(S)	HIGH AVAILABILITY ENABLED ❌	LINKED COUNT	QUEUE LENGTH
0 / 2	False ⓘ	3 ✓	0

AVERAGE QUEUE DURATION
0.00s

Node Details

Name	Status	Version	Available mem...	CPU utilization	Network (In/Out)	Concurrent job...	Role	Credential status	Message
ADFCookbookIRDev1	❌ Unavailable	4.11.7521.1	---	---	---/---	---/---/--- More	Worker	In sync	💬
ADFCookbookIRDev2	❌ Unavailable	4.11.7521.1	---	---	---/---	---/---/--- More	Dispatcher/wor...	In sync	💬

Home >

Custom deployment ⋯
Deploy from a custom template

🧭 New! Deployment Stacks let you manage the lifecycle of your deployments. Try it now →

Select a template   **Basics**   Review + create

Template

📋 **Customized template** ⤢
2 resources

✏️ Edit template   ✏️ Edit parameters   🔀 Visualize

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 * ⓘ | ⌄ |

Create new

Instance details

Region * ⓘ | Korea Central   ⌄ |

Data Factory Name ⓘ | [concat('datafactory', uniqueString(resourceGroup().id))] |

Location ⓘ | East US |

Storage Account Name ⓘ | [concat('storage', uniqueString(resourceGroup().id))] |

Blob Container ⓘ | [concat('blob', uniqueString(resourceGroup().id))] |

Previous   Next   **Review + create**

Activity name	Activity status	Activity type	Run start	Duration	Log
Record Status Failure	✅ Succeeded	Stored procedure	8/18/2023, 6:47:05 PM	5s	
Set Failure	✅ Succeeded	Set variable	8/18/2023, 6:47:04 PM	Less than 1s	
Copy From Bl...	❌ Failed	Copy data	8/18/2023, 6:46:44 PM	19s	

Error details

📋 Copy From Blob To Azure SQL

Error code [UserErrorSourceBlobNotExist](#) 🔗

Failure type User configuration issue

Activity ID 4edae43e-f619-4158-8a76-9c39e59b8d01

Details ErrorCode=UserErrorSourceBlobNotExist,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=The required Blob is missing. Please contact Blob Storage team if you need further support. ContainerName: https://storageaufszdu5xge7i.blob.core.windows.net/data, path: airlines.csv.,Source=Microsoft.DataTransfer.ClientLibrary,''Type=Microsoft.WindowsAzure.Storage.StorageException,Message=The remote server returned an error: (404) Not Found.,Source=Microsoft.WindowsAzure.Storage,StorageExtendedMessage=RequestId:825bd64f-001e-003f-053f-d21a0c000000 Time:Sat, 19 Aug 2023 01:46:59 GMT,,''Type=System.Net.WebException,Message=The remote server returned an error: (404) Not Found.,Source=Microsoft.WindowsAzure.Storage,'

Pipeline run ID: 5ef1c938-7eb9-4a53-94fe-5cde30df8920 [@] ↻ ⓘ   **Pipeline status** ✅ Succeeded   [View debug run consumption](#)

All status ⌄   ↓ Export to CSV | ⌄

Showing 1 - 3 of 3 items

Activity name ↑↓	Activity status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log	Integration runtime
Record Status Success	✅ Succeeded	Stored procedure	8/18/2023, 7:01:42 PM	7s		AutoResolveIntegratio
Set Success	✅ Succeeded	Set variable	8/18/2023, 7:01:41 PM	Less than 1s		
Copy From Bl... →] [→ 👓	✅ Succeeded	Copy data	8/18/2023, 7:01:22 PM	18s		AutoResolveIntegratio

Details

Details ↻ Refresh

[Learn more on copy performance details from here.](#)

Activity run id: 1734a092-2bc9-4016-b2ee-997aa23eb503

🗔 Azure Blob Storage   Succeeded →   🗄 SQL Azure SQL Database
   Region: East US   Region: Korea Central

Data read: ⓘ	233.849 KB	Data written: ⓘ	247.074 KB
Files read: ⓘ	1	Rows written: ⓘ	6,161
Rows read:	6,161	Peak connections: ⓘ	2
Peak connections: ⓘ	1		

Copy duration   00:00:12
Throughput: ⓘ   46.77 KB/s

⌄ Azure Blob Storage → Azure SQL Database

   Start time   8/18/2023, 7:01:25 PM
   Used DIUs ⓘ   4
   Used parallel copies ⓘ 1
   ⌄ Duration   00:00:12

Details	Working duration	Total duration
🟢 Queue ⓘ		00:00:06
🟢 Transfer ⓘ	Listing source ⓘ 00:00:01	00:00:05
	Reading from source ⓘ 00:00:00	
	Writing to sink ⓘ 00:00:00	

Data consistency verification ⓘ   Not verified

💬 How satisfied or dissatisfied are you with the performance of this copy activity?
   ★ ★ ★ ★ ★

Data Factory ▸ adf-cookbook-v2-ch10-df

🔍 Search factory and documentation

ⓘ Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build clo

Dashboards

Runs

Pipeline runs

Trigger runs

Change Data Capture (previ...

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

⚠ Alerts & metrics

Alerts & metrics

🔄 Refresh   📊 Metrics   🔔 New alert rule

Add criteria

Select one metric to set up the alert condition.

Metrics ↑↓

External available capacity percentage of MVNet integration runtime

External capacity utilization of MVNet integration runtime

External waiting queue length of MVNet integration runtime

Failed activity runs metrics

Failed pipeline runs metrics

Failed SSIS integration runtime start metrics

Failed SSIS package execution metrics

Failed trigger runs metrics

Integration runtime available memory

Integration runtime available node count

Integration runtime CPU utilization

Integration runtime queue duration

Integration runtime queue length

Maximum allowed entities count

Maximum allowed factory size (GB unit)

Pipeline available capacity percentage of MVNet integration runtime

Pipeline capacity utilization of MVNet integration runtime

Pipeline waiting queue length of MVNet integration runtime

Stuck SSIS integration runtime stop metrics

Succeeded activity runs metrics

Continue   Cancel

Configure alert logic

Show history

Over the last 6 hours ⌄

```
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
 0
   2PM      3PM      4PM      5PM      6PM      7PM
```

▌**Total**

Selecting the dimension values will help you filter to the right time series.

Dimension	Values
ActivityType	Select a value ⌄
ActivityName	Select a value ⌄
PipelineName	Select a value ⌄
FailureType	Select a value ⌄

Alert logic

Condition * ⓘ

Greater than ⌄

Time aggregation * ⓘ

Total ⌄

Threshold count * ⓘ

0 ⬍

Add criteria   **Back**   **Cancel**

Add notification

Learn more about **Pricing** ↗ and **Privacy statement.** ↗

Action name *

Notify On Call Data Engineers

Select which notifications you'd like to receive

☑ **Email**

oncall-data-engineers@my-company.com

☑ **SMS**

Country code	Phone number *
1 ⌄	8135551212

Carrier charges may apply.

☐ **Azure app push notifications**
Enter your email used to log into your Azure account. **Learn about connecting to your Azure resources using the Azure app.** ↗

email@example.com

☐ **Voice**

Country code | Phone number *

1 ⌄ | 1234567890

Add notification   Cancel

New alert rule

Alert rule name *

Pipeline Failures Detected

Description

This alert will fire when there are failures in our pipeline, i.e., if one or more of the activites fails

Severity *

Sev2

Target criteria	Actions
Whenever Activity Failed Runs metric is Greater T	✏️ 🗑️

+ Add criteria

ℹ️ *There will be a monthly rate for the configured criteria. Learn more about Pricing* 🔗

Notifications	Action group type	Actions
On Call Data Engineers		🗑️

+ Configure notification

Enable rule upon creation

Create alert rule   Cancel

Alerts & metrics

🔄 Refresh   📊 Metrics   🔔 New alert rule

ALERT	ENABLED	RESOURCE TYPE	RESOURCES	ACTIONS
Pipeline Failures Detec	🔵 On	Activity		✏️ 🗑️

Microsoft Azure

You've been added to an Azure Monitor action group

You are now in the DE on call action group and will receive notifications sent to the group.

View details on Azure Monitor action groups >

Account information

Subscription ID:

Resource group name: adf-cookbook-v2-ch10-rg

Action group name: Data Engineer On Call
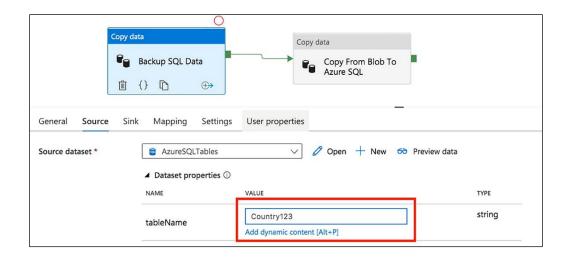
To unsubscribe from this action group, click here.

f 🐦 ▶ in

Privacy Statement

Microsoft Corporation, One Microsoft Way, Redmond, WA 98052

Microsoft

⚠ Your Azure Monitor alert was triggered
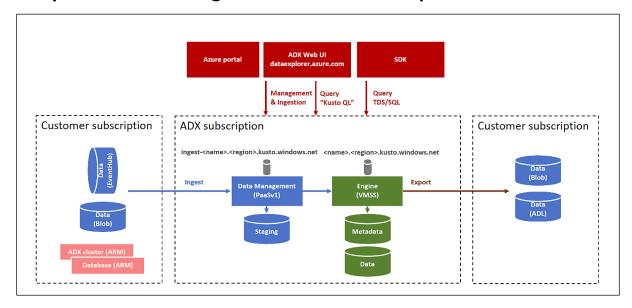
Azure monitor alert rule Pipeline Failures Detected was triggered for adf-cookbook-v2-ch10-df at August 25, 2023 3:06 UTC.
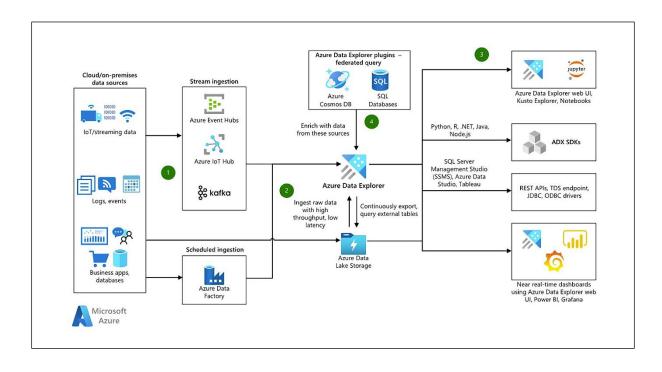
Rule ID	/subscriptions/░░░░░░░░░░░░░░░░░░░░░░░░░░░░be/resourcegroups/adf-cookbook-v2-ch10-rg/providers/Microsoft.Insights/metricalerts/Pipeline Failures Detected View Rule >
Resource ID	/subscriptions/░░░░░░░░░░░░░░░░░░░░░░░░░░░░be/resourcegroups/adf-cookbook-v2-ch10-rg/providers/Microsoft.DataFactory/factories/adf-cookbook-v2-ch10-df View Resource >
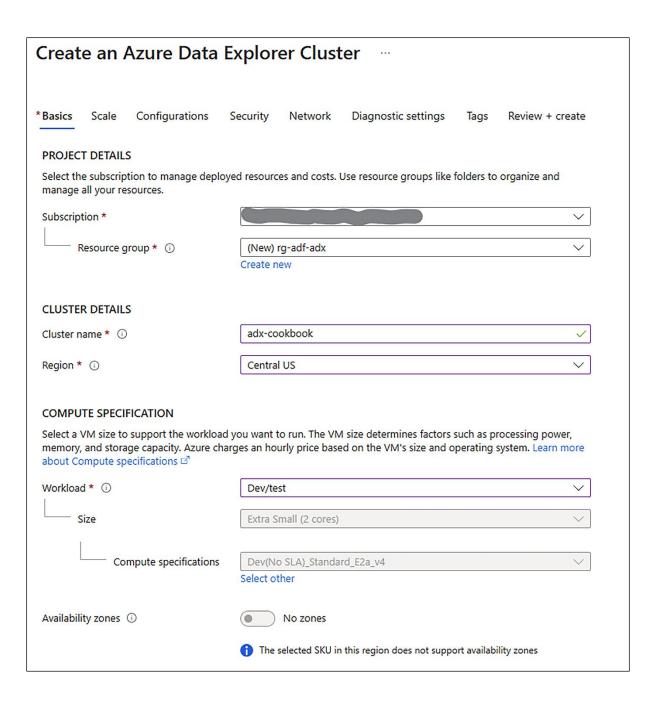
Alert Activated Because:
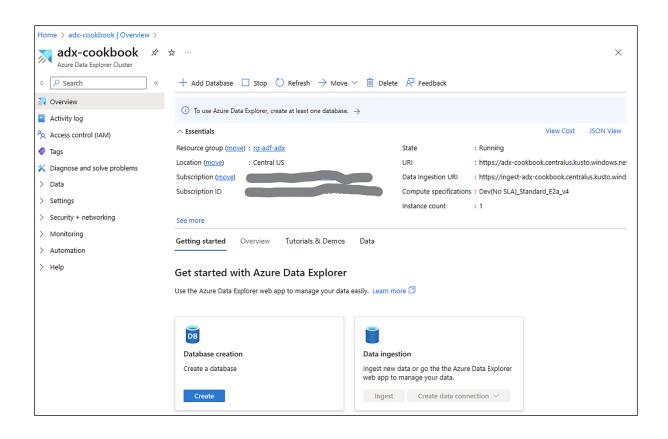
Metric name	ActivityFailedRuns
Metric namespace	factories/adf-cookbook-v2-ch10-df
Dimensions	ResourceId = /SUBSCRIPTIONS/░░░░░░░░░░░░░░░░░░░░░░░░░/RESOURCEGROUPS/ADF-COOKBOOK-V2-CH10-RG/PROVIDERS/MICROSOFT.DATAFACTORY/FACTORIES/ADF-COOKBOOK-V2-CH10-DF
Time Aggregation	Total
Period	Over the last 1 mins
Value	2
Operator	GreaterThan
Threshold	0
Criterion Type	StaticThresholdCriterion

Chapter 11: Working with Azure Data Explorer

Create an Azure Data Explorer Cluster ···

*Basics Scale Configurations Security Network Diagnostic settings Tags Review + create

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 * ⓘ   (New) rg-adf-adx   ▼

                    Create new

CLUSTER DETAILS

Cluster name * ⓘ   adx-cookbook   ✓

Region * ⓘ   Central US   ▼

COMPUTE SPECIFICATION

Select a VM size to support the workload you want to run. The VM size determines factors such as processing power, memory, and storage capacity. Azure charges an hourly price based on the VM's size and operating system. Learn more about Compute specifications ⧉

Workload * ⓘ   Dev/test   ▼

    Size   Extra Small (2 cores)   ▼

        Compute specifications   Dev(No SLA)_Standard_E2a_v4   ▼

                    Select other

Availability zones ⓘ   ⬤◯ No zones

            ⓘ The selected SKU in this region does not support availability zones

Command

| General | Connection | **Command** | User properties |

Command * ⓘ

```
.set-or-replace DeadlyStorms <|
StormEvents
| where DeathsDirect > 5
| order by StartTime desc
| project StartTime, EndTime, State, EventType, DeathsDirect,
BeginLocation, EndLocation
```

Command timeout ⓘ

```
00:20:00
```

« ▷ Run ⌄ | ⟲ Recall | 🖹 KQL tools ⌄ | adx-cookbook.centralus/cookbook

```
1
2  DeadlyStorms
3  | limit 5
```

⊞ Table 1 | + Add visual | ◎ Stats

StartTime	EndTime	State	EventType	DeathsDirect
2007-08-04 12:00:00.0000	2007-08-16 19:00:00.0000	MISSOURI	Excessive Heat	6
2007-08-05 11:00:00.0000	2007-08-25 18:00:00.0000	TENNESSEE	Heat	14
2007-09-01 00:00:00.0000	2007-09-03 20:00:00.0000	CALIFORNIA	Excessive Heat	10
2007-09-03 08:00:00.0000	2007-09-03 20:00:00.0000	CALIFORNIA	Excessive Heat	8
2007-10-21 08:23:00.0000	2007-10-31 00:00:00.0000	CALIFORNIA	Wildfire	8

✓ Validate   ▷ Debug   ⚡ Add trigger

Azure Data Explorer Comm... ✓

Azure Data
Explorer...

🗑 </> 📋   →

General   Connection   **Command**   User properties

Command * ⓘ

```
.create table bio (name: string, language: string, ['id']: string, bio:
string, ['version']: real)
```

Command timeout ⓘ

```
00:20:00
```

Azure Data Explorer Comm ✓

Azure Data Explorer...

Copy data

Copy data1

Preview data

Linked service: AzureBlobStorage1

Object: sample.json

```
[
  {
    "name": "Adeel Solangi",
    "language": "Sindhi",
    "id": "V59OF92YF627HFY0",
    "bio": "Donec lobortis eleifend condimentum. Cras dictum dolor lacinia lectus vehicula rutrum. Maec
    "version": 6.1
  },
  {
    "name": "Afzal Ghaffar",
    "language": "Sindhi",
    "id": "ENTOCR13RSCLZ6KU",
    "bio": "Aliquam sollicitudin ante ligula, eget malesuada nibh efficitur et. Pellentesque massa sem,
    "version": 1.88
  },
  {
    "name": "Aamir Solangi",
    "language": "Sindhi",
    "id": "IAKPO3R4761JDRVG",
    "bio": "Vestibulum pharetra libero et velit gravida euismod. Quisque mauris ligula, efficitur portt
    "version": 7.27
  },
  {
    "name": "Abla Dilmurat",
    "language": "Uyghur",
    "id": "5ZVOEPMJUI4MB4EN",
    "bio": "Donec lobortis eleifend condimentum. Morbi ac tellus erat."
```

Set properties

Name

```
AzureDataExplorerTable3
```

Linked service *

```
AzureDataExplorer1
```

Table

```
bio
```

☑ Edit

Import schema

⦿ From connection/store   ◯ None

⌄ **Advanced**

Open this dataset for more advanced configuration with parameterization.

✓ Validate   ▷ Debug   ⚡ Add trigger   { } ▤ ⋯

Azure Data Explorer Comm... ✓
  Azure Data
  Explorer...   Copy data ✓
                                   ▤ Copy data1

Parameters   Variables   Settings   **Output**   ⌃

Pipeline run ID: 0dff4fea-00c2-4906-8276-7a3f6e73b78f ▣ ↻ ⓘ   **Pipeline status** ✓ Succeeded   View debug run consumption

All status ⌄   ↓ Export to CSV | ⌄

Showing 1 - 2 of 2 items

Activity name ↑↓	Activity status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log	Integration runtime	User p
Copy data1	✓ Succeeded	Copy data	8/13/2023, 7:58:55 PM	41s		AutoResolveIntegratior	
Azure Data Explorer Comman...	✓ Succeeded	Azure Data Explorer Cc	8/13/2023, 7:58:53 PM	2s		AutoResolveIntegratior	

Chapter 12: The Best Practices of Working with ADF

Reader
BuiltInRole

Permissions   JSON   Assignments

Description: View all resources, but does not allow you to make any changes.

Data Factory

Type : **All**

⦿ Actions   ○ DataActions

Showing 59 of 6850 permissions

Type	Permissions	Description
	Microsoft.DataFactory	
Read	Read Data Factory ⓘ	Reads the Data Factory.
Read	Read Activity Windows in any Data Factory ⓘ	Reads Activity Windows in the Data Factory with specified parameters.
Read	Read diagnostic setting ⓘ	Gets the diagnostic setting for the resource
Read	Read datafactories metric definitions ⓘ	Gets the available metrics for datafactories
Read	Read Dataset ⓘ	Reads any Dataset.
Read	Read Activity Windows for any Dataset ⓘ	Reads Activity Windows for the Dataset with specified parameters.
Read	Read Data Slice Run ⓘ	Reads the Data Slice Run for the given dataset with the given start time.
Read	Read Data Slices ⓘ	Gets the Data Slices in the given period.
Read	Read Dataset ⓘ	Reads any Dataset.
Read	Read Gateway ⓘ	Reads any Gateway.
Read	Read Linked Service ⓘ	Reads any Linked Service.
Read	Read Pipeline ⓘ	Reads any Pipeline.
Read	Read Activity Windows for any Pipeline Activity ⓘ	Reads Activity Windows for the Pipeline Activity with specified parameters.
Read	Read Activity Windows for any Pipeline ⓘ	Reads Activity Windows for the Pipeline with specified parameters.
Read	Read Run Log Info ⓘ	Reads a SAS URI to a blob container containing the logs.
Read	Read Data Factory ⓘ	Reads Data Factory.
Read	Read ADF Change data capturen ⓘ	Reads ADF Change data capture.
Read	Read ADF Link Connection ⓘ	Reads ADF Link Connection.

Home > Data factories > adf-cookbook-v2-ch12-adf > adf-cookbook-v2-ch3-rg > Azure subscription 1 | Access control (IAM) >

Create a custom role ⋯

Basics   Permissions   Assignable scopes   JSON   Review + create

To create a custom role for Azure resources, fill out some basic information. Learn more ⧉

Custom role name * ⓘ	Data Engineer ✓
Description	Users chave same access as Data Contributor, but can not delete the data factory
Baseline permissions ⓘ	● Clone a role ○ Start from scratch ○ Start from JSON
Role to clone	Data Factory Contributor ⓘ ▼

Microsoft.DataFactory permissions   ✕

‹ All resource providers

ⓘ Exclude permissions enable you to subtract specific permissions from a wildcard (*) permission. If your custom role has a permission with a wildcard (*) and you don't want to include all the permissions under the wildcard, search for permissions to subtract from the wildcard. For example, search for "virtual machines" to find permissions related to virtual machines.

Data Factory

● Not Actions   ○ Not Data Actions

☑ Permission	Description
∨ **Microsoft.DataFactory**	
☐ Other : Register Data Factory Resource Provider ⓘ	Registers the subscription for the Data Factory Resource Provider.
☐ Other : Unregister Data Factory Resource Provider ⓘ	Unregisters the subscription for the Data Factory Resource Provider.
∨ **Microsoft.DataFactory/datafactories** : Unregister Data Factory Resource Provider	
☐ Read : Read Data Factory ⓘ	Reads the Data Factory.
☐ Write : Create or Update Data Factory ⓘ	Creates or Updates the Data Factory.
☑ Delete : Delete Data Factory ⓘ	Deletes the Data Factory.
∨ **Microsoft.DataFactory/datafactories/activitywindows**	
☐ Read : Read Activity Windows in any Data Factory ⓘ	Reads Activity Windows in the Data Factory with specified parameters.
∨ **Microsoft.DataFactory/datafactories/providers/Microsoft.Insights/diagnosticSettings**	
☐ Read : Read diagnostic setting ⓘ	Gets the diagnostic setting for the resource
☐ Write : Write diagnostic setting ⓘ	Creates or updates the diagnostic setting for the resource
∨ **Microsoft.DataFactory/datafactories/providers/Microsoft.Insights/metricDefinitions**	
☐ Read : Read datafactories metric definitions ⓘ	Gets the available metrics for datafactories
∨ **Microsoft.DataFactory/datafactories/datasets**	
☐ Read : Read Dataset ⓘ	Reads any Dataset.
☐ Delete : Delete Dataset ⓘ	Deletes any Dataset.
☐ Write : Create or Update Dataset ⓘ	Creates or Updates any Dataset.
∨ **Microsoft.DataFactory/datafactories/datasets/activitywindows**	

Add   **Cancel**

Copy Data tool

- ✓ Properties
- ✓ Source
- ③ **Destination**
  - Dataset
  - ○ Configuration
- ④ Settings
- ⑤ Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type
Azure Data Lake Storage Gen2 ▾

Connection *
LS_ADLS_Ingest ▾   ✏ Edit   + New connection

Folder path
If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

curated/movielens   📁 Browse

File name
File name is defined by source table name

> Advanced settings

File name suffix
.csv

Copy Data tool

- ✓ Properties
- ② Source
  - Dataset
  - Configuration
- ③ Destination
- ④ Settings
- ⑤ Review and finish

File format settings

File format
DelimitedText ▾   [Detect text format]   [👓 Preview data]

Column delimiter
Comma (,) ▾
☐ Edit

Row delimiter
Default (\r,\n, or \r\n) ▾
☐ Edit

☑ First row as header ⓘ

> Advanced

Compression type
Select... ▾

Additional columns ⓘ
+ New

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ④ Settings
- ⑤ Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task prefix name *
`Metadata_`

Task description
[   ]

Number of concurrent copy tasks ⓘ
`Auto` ▼

☐ Edit

> Advanced settings

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- ⑤ Review and finish
  - Review
  - Deployment

SQL → Azure SQL Database ────────▶ Azure Data Lake Storage Gen2

Deployment complete

Deployment step	Status
> Creating datasets	✓ Succeeded
> Creating pipelines	✓ Succeeded
> Generating SQL scripts	✓ Succeeded

Please run the following SQL script in your SQL server to create a control table. Then view your pipeline to execute a debug run.

⬇ Download SQL script

Generated SQL script for control table

⎘ Copy to clipboard

```
/****** Object:  Table [dbo].[MainControlTable] ******/
CREATE TABLE [dbo].[MainControlTable](
    [Id] [int] IDENTITY(1,1) NOT NULL PRIMARY KEY,
    [SourceObjectSettings] [nvarchar](max) NULL,
```

Finish   View pipeline + debug pipeline run

```sql
SELECT * FROM [dbo].[MainControlTable]
```

	Id	SourceObjectSettings			SourceConne...	CopySourceSettings		SinkObjectSettings	
1	1	{	"schema": "dbo",	"table": "movielens_links"	}	NULL	{	"partitionOption": "None", ...	{	"fileName": "dbomovielens_links.csv", ...
2	2	{	"schema": "dbo",	"table": "movielens_movies"	}	NULL	{	"partitionOption": "None", ...	{	"fileName": "dbomovielens_movies.csv", ...
3	3	{	"schema": "dbo",	"table": "movielens_ratings"	}	NULL	{	"partitionOption": "None", ...	{	"fileName": "dbomovielens_ratings.csv", ...
4	4	{	"schema": "dbo",	"table": "movielens_tags"	}	NULL	{	"partitionOption": "None", ...	{	"fileName": "dbomovielens_tags.csv", ...

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 * ⓘ | Azure subscription 1   ⌄ |

    Resource group * ⓘ | adfcookbookv2   ⌄ |

Create new

Instance details

Name * ⓘ | ADF-cookbook-restore   ✓ |

Region * ⓘ | Central US   ⌄ |

Version * ⓘ | V2   ⌄ |

Configure a repository

Default Directory (b269e661-3fcc-4f34-8ac4-a32126dce250)

Specify the settings that you want to use when connecting to your repository.

(●) Select repository   (○) Use repository link

Azure DevOps organization name * ⓘ

adfcookbook2ed	⌄

Project name * ⓘ

ADF DevOps	⌄

Repository name * ⓘ

ADF DevOps	⌄

Collaboration branch * ⓘ

master	⌄

Publish branch * ⓘ

adf_publish	⌄

Root folder * ⓘ

/

Custom comment

[✓] Use custom comment

Import existing resources

[ ] Import existing resources to repository

[Apply]   [Back]   [Cancel]

Integration runtime

Settings   Data flow runtime

The Data Factory manages the integration runtime in Azure to connect to required data source/destination or external compute in public network. The compute resource is elastic allocated based on performance requirement of activities.

Name * ⓘ

AutoResolveIntegrationRuntime

Description

Enter description here...

Status

Region *

Auto Resolve   ⌄

Choose Your Sources

Change Data Capture (CDC) automatically detects data changes at the source and then sends the updated data to the destination.

CDC name *

adfcdc1

Source type *

🗄 Azure SQL Database ⌄

Source linked service *

🗄 LS_SQL_IngestionConfiguration ⌄ ✎

Source tables

🔍 movi

Showing 1 - 1 of 1 tables (1 selected)

☑	Name	Incremental column		Preview
☑	▦ dbo.movielens_ratings	ConvertedDate ⌄	↻	▤

Continue   Cancel

Choose Your Targets

Change Data Capture (CDC) automatically detects data changes at the source and then sends the updated data to the destination.

Target type *

△ Delta ▾

Target linked service *

▣ LS_ADLS_Ingest ▾ ✎

Select target folder paths

Target base path ⓘ

curated ▭

curated/cdc/dbo.movielens_ratings ▭ + 🗑

Continue   Back   Cancel

curated
Container

Upload · Add Directory · Refresh · Rename · Delete · Change tier · Acquire lease · Break lease · Give feedback

Authentication method: Azure AD User Account (Switch to Access key)
Location: curated / cdc / dbo.movielens_ratings

Search blobs by prefix (case-sensitive) · Show deleted

	Name	Modified	Access tier	Archive status	Blob type	Size
	_delta_log					
	part-00000-7288da8b-b9cd-43...	10/23/2023, 12:23:16...	Hot (Inferred)		Block blob	1.38 KiB
	part-00000-aad8fe9f-bd0a-49d...	10/23/2023, 12:33:52...	Hot (Inferred)		Block blob	1.44 KiB

Settings
- Overview
- Diagnose and solve problems
- Access Control (IAM)
- Shared access tokens
- Manage ACL
- Access policy
- Properties
- Metadata

Transform your data into actionable insights using the best-in-class machine learning tools. This architecture allows you to combine any data at any scale, and to build and deploy custom machine learning models at scale.

- Advanced analytics on big data
- CI/CD for Azure Web Apps
- CI/CD for Containers
- Enterprise data warehouse
- Real-time analytics
- Azure Machine Learning Platform

Products
- Azure Synapse Analytics
- Azure Analysis Services
- Power BI Embedded
- Storage Accounts
- Azure Cosmos DB

Learn more →

Add to estimate

Your Estimate x | Advanced Analytics ... x | +

Advanced Analytics on Big Data

				Upfront	Monthly
∨ Azure Synapse Analytics	ⓘ	Tier: Compute Optimized Gen2, Dedicated SQL Pool...		Upfront: $0.00	Monthly: $6,418.94
∨ Azure Analysis Services	ⓘ	Developer (Hours), 1 Instance(s), 720 Hours		Upfront: $0.00	Monthly: $95.04
∨ Power BI Embedded	ⓘ	1 node(s) x 720 Hours, Node type: A2, 2 Virtual Cor...		Upfront: $0.00	Monthly: $1,445.83
∨ Storage Accounts	ⓘ	Data Lake Storage Gen2, Standard, LRS Redundancy...		Upfront: $0.00	Monthly: $15,818.40
∧ Azure Cosmos DB	ⓘ	Standard provisioned throughput (manual), Always-...		Upfront: $0.00	Monthly: $273.36

Get started   **Summary**   Tutorials & What's new

Amount due

$0.00

✓ No payment needed

[View invoices]

Upcoming invoices

Available on 10/9/2023

Billing period   Pre-tax total so far ⓘ
9/1/2023 - 9/30/2023 $387.42

Available on 11/9/2023

Billing period   Pre-tax total so far ⓘ
10/1/2023 - 10/31/2023 $13.08

Invoices over time

```
1.2K
1K
0.8K
0.6K
0.4K
0.2K
0K
     04/23  05/23  06/23  07/23  08/23
```

Total amount
$1.02 ₖ

Spending rate and forecast

```
300
200
100
0
  10/01  10/08  10/13  10/22  10/29
```

Current charges ⓘ   Forecast ⓘ
$13.44 $211.09

[View details]

This month's top products by charges

Premium SSD Managed Disks - …
$12.22
Standard SSD Managed Disks - …
$0.95
IP Addresses - Basic IPv4
$0.27
Other purchases
$0.00

[View details]

Billing alerts (2)

ⓘ Review and accept the updated terms

ⓘ Verify bill to and PO number before next invoice

[View alerts]

Microsoft Azure | Data Factory ▸ ADF-cookbook-restore   🔍 Search factory and documentation   tonya.chernyshova@gmail.com
                                                                                                  DEFAULT DIRECTORY

ⓘ Your insights matter! Participate in our brief survey about our CDC top-level resource, and help us enhance you

Pending changes

◇ / ⌥ master branch ▾   ✓ Validate all 💾 Save all ⬆ Publishing

«

General

 ▫ Factory settings

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 endpoints

Factory settings

In git mode, changes to these settings are automatically saved.

Show billing report ⓘ   ● by pipeline ○ by factory

  › Staging ⓘ

Publish branch   adf_publish   ⓘ

∨ Data Factory properties

  ▣ Global settings   (Edited)   –

[OK]   [Cancel]

ADFCOOKBOOK-V2 | Tags

Data factory (V2)

Home > ADFCOOKBOOK-V2

☑ Delete all

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. Tag names are case insensitive, but tag values are case sensitive. Learn more about tags

Do not enter names or values that could make your resources less secure or that contain personal/sensitive information because tag data will be replicated globally.

Name ⓘ		Value ⓘ	
Environment	:	Production	
	:		

ADFCOOKBOOK-V2 (Data factory (V2))

Environment : Production

No changes

Create budget ⋯
Budget

✓ Create a budget   ② Set alerts

Create a budget and set alerts to help you monitor your costs.

Budget scoping

The budget you create will be assigned to the selected scope. Use additional filters like resource groups to have your budget monitor with more granularity as needed.

Scope

Filters   ⁺▽ Add filter

Budget Details

Give your budget a unique name. Select the time window it analyzes during each evaluation period, its expiration date and the amount.

* Name   BudgetExample   ✓
* Reset period ⓘ   Monthly   ⌄
* Creation date ⓘ   2023 ⌄   October ⌄   1
* Expiration date ⓘ   2025 ⌄   September ⌄   30 ⌄

Budget Amount

Give your budget amount threshold

Amount ($) *   431   ✓

ⓘ Suggested budget: $431 based on forecast.

Create budget ···
Budget

✓ Create a budget   ✓ **Set alerts**

Configure alert conditions and send email notifications based on your spend.

* Alert conditions

Type	% of budget	Amount	
Actual	5	21.55	🗑
Forecasted ⌄	10 ✓	43.10	🗑
Select type ⌄	Enter %	-	

* Alert recipients (email)

Alert recipients (email)

your@email.com ✓	🗑
example@email.com	

It is recommended to add azure-noreply@microsoft.com to your email white list to ensure alert mails do not go to your spam folder.

Language preference

Select your preferred language for receiving the alert email for all recipients provided above. Default is the language associated to your enrollment.

Languages * | Default   ⌄ |

ℹ Your budget evaluation will begin in a few hours. Learn more

| Previous | **Create**