LAB01 : Load Data Into Cosmos DB with ADF

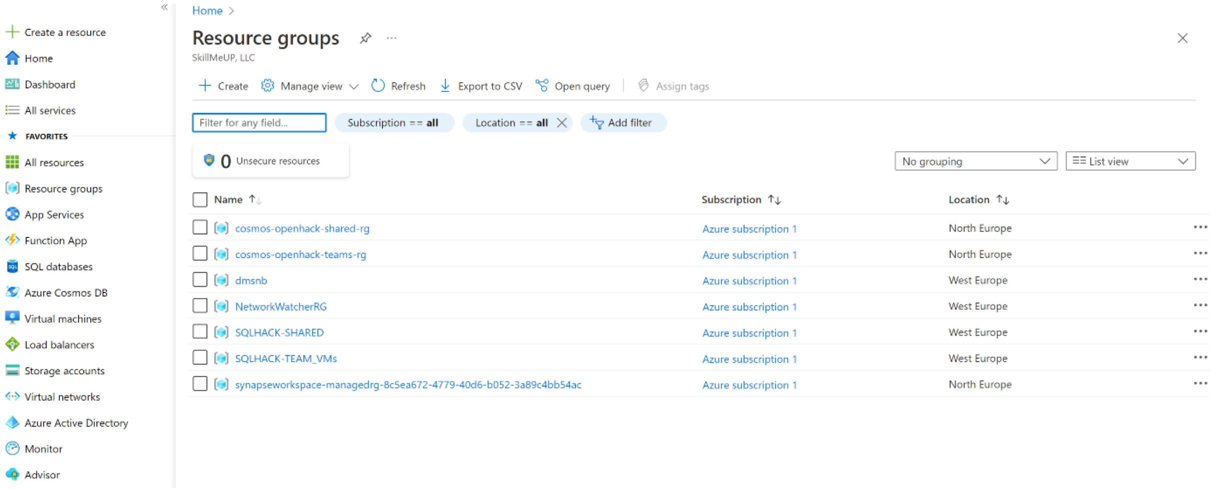
In this lab, you will populate an Azure Cosmos DB container from an existing set of data using tools built in to Azure. After importing, you will use the Azure portal to view your imported data.

Create Azure Cosmos DB Database and Container

You will now create a database and container within your Azure Cosmos DB account.

1. Navigate to the Azure Portal

2. On the left side of the portal, select the **Resource groups** link.



Resource groups

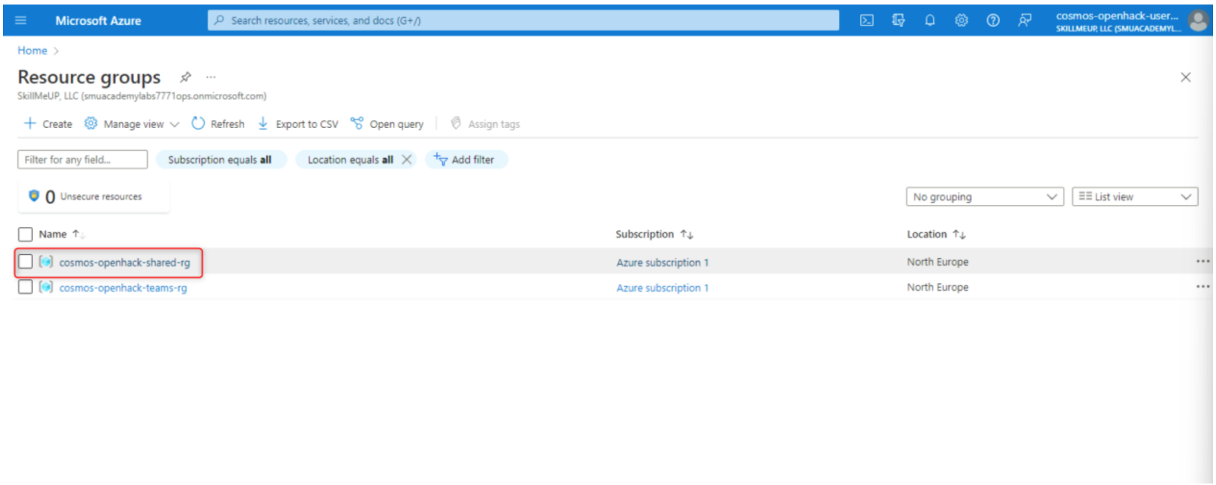
SkillMeUP, LLC

Filter for any field... | Subscription == all | Location == all

0 Unsecure resources

Name	Subscription	Location
cosmos-openhack-shared-rg	Azure subscription 1	North Europe
cosmos-openhack-teams-rg	Azure subscription 1	North Europe
dmsnb	Azure subscription 1	West Europe
NetworkWatcherRG	Azure subscription 1	West Europe
SQLHACK-SHARED	Azure subscription 1	West Europe
SQLHACK-TEAM_VMs	Azure subscription 1	West Europe
synapseworkspace-managedrg-8c5ea672-4779-40d6-b052-3a89c4bb54ac	Azure subscription 1	North Europe

3. In the **Resource groups** blade, locate and select the **cosmos-openhack-shared-rg** resource group.



Resource groups

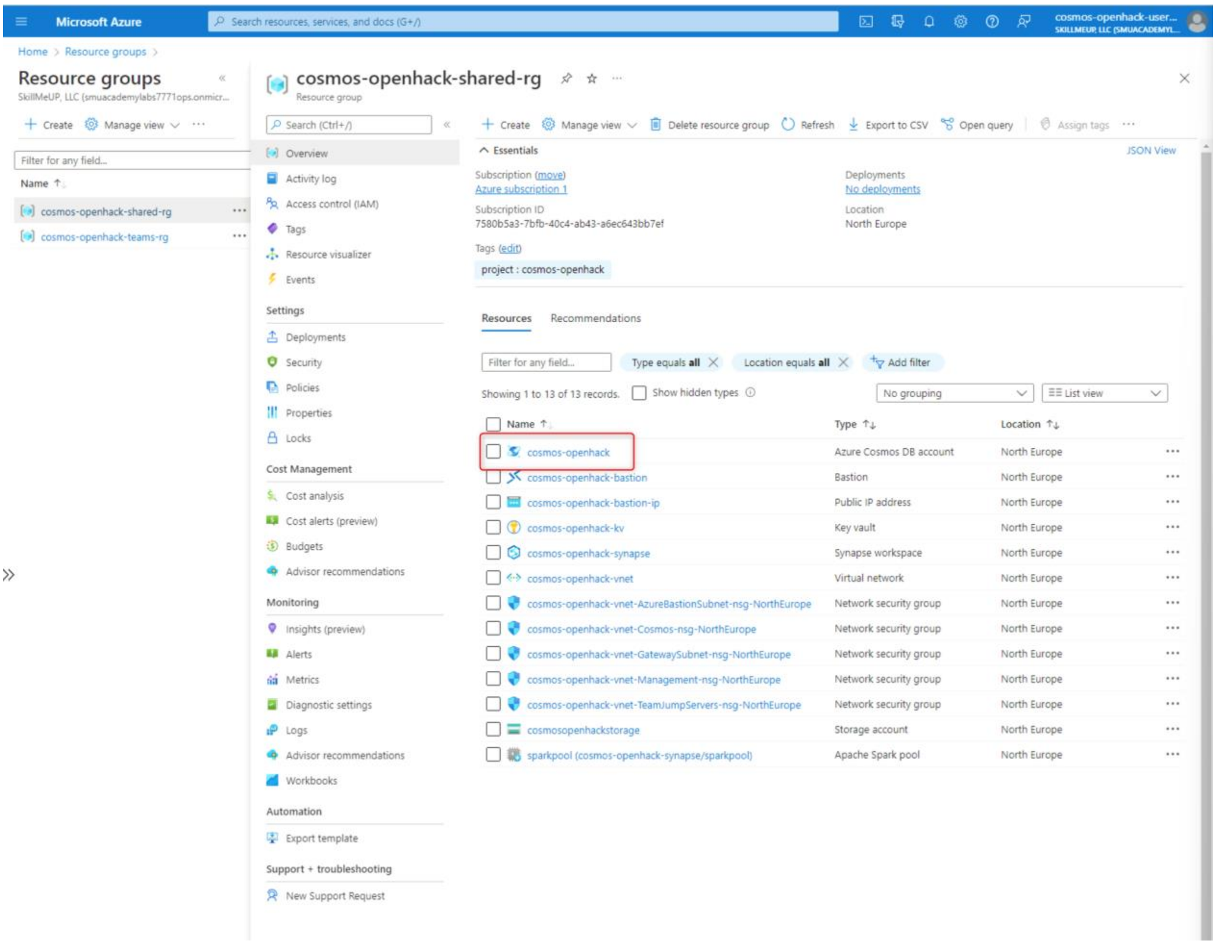
SkillMeUP, LLC (smuacademylabs7771ops.onmicrosoft.com)

Filter for any field... | Subscription equals all | Location equals all

0 Unsecure resources

Name	Subscription	Location
cosmos-openhack-shared-rg	Azure subscription 1	North Europe
cosmos-openhack-teams-rg	Azure subscription 1	North Europe

4. In the **cosmos-openhack-shared-rg** blade, select the **Azure Cosmos DB** account .



Microsoft Azure · Search resources, services, and docs (G+/) · cosmos-openhack-user... SKILLMEUP, LLC (SMUACADEMYL...

Home > Resource groups >

Resource groups SkillMeUP, LLC (smuacademylabs7771ops.onmicr...

+ Create · Manage view · ...

Filter for any field...

Name ↑

- cosmos-openhack-shared-rg
- cosmos-openhack-teams-rg

cosmos-openhack-shared-rg Resource group

Search (Ctrl+/)

Overview · Activity log · Access control (IAM) · Tags · Resource visualizer · Events

Settings: Deployments · Security · Policies · Properties · Locks

Cost Management: Cost analysis · Cost alerts (preview) · Budgets · Advisor recommendations

Monitoring: Insights (preview) · Alerts · Metrics · Diagnostic settings · Logs · Advisor recommendations · Workbooks

Automation: Export template

Support + troubleshooting: New Support Request

+ Create · Manage view · Delete resource group · Refresh · Export to CSV · Open query · Assign tags · ...

Essentials · JSON View

Subscription (move): Azure subscription 1

Subscription ID: 7580b5a3-7bfb-40c4-ab43-a6ec643bb7ef

Tags (edit): project : cosmos-openhack

Deployments: No deployments

Location: North Europe

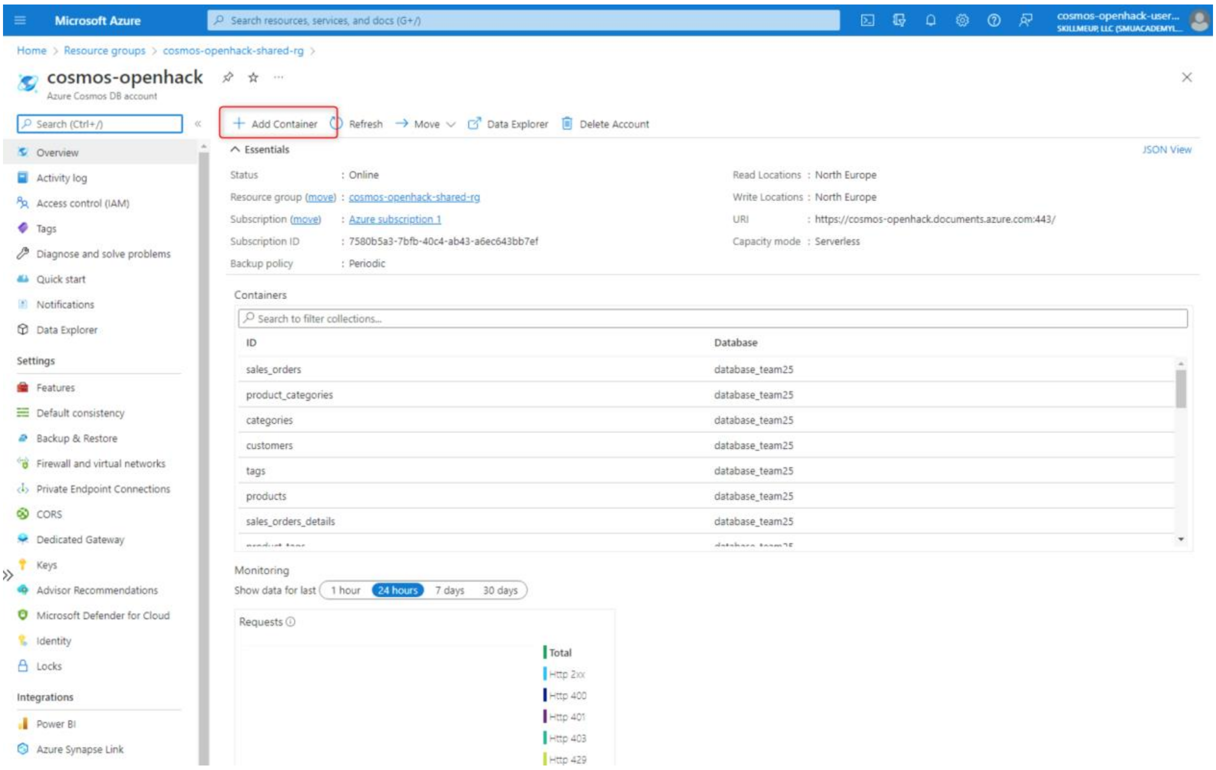
Resources · Recommendations

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

Showing 1 to 13 of 13 records. □ Show hidden types · No grouping · List view

Name ↑	Type ↑↓	Location ↑↓
cosmos-openhack	Azure Cosmos DB account	North Europe
cosmos-openhack-bastion	Bastion	North Europe
cosmos-openhack-bastion-ip	Public IP address	North Europe
cosmos-openhack-kv	Key vault	North Europe
cosmos-openhack-synapse	Synapse workspace	North Europe
cosmos-openhack-vnet	Virtual network	North Europe
cosmos-openhack-vnet-AzureBastionSubnet-nsg-NorthEurope	Network security group	North Europe
cosmos-openhack-vnet-Cosmos-nsg-NorthEurope	Network security group	North Europe
cosmos-openhack-vnet-GatewaySubnet-nsg-NorthEurope	Network security group	North Europe
cosmos-openhack-vnet-Management-nsg-NorthEurope	Network security group	North Europe
cosmos-openhack-vnet-TeamJumpServers-nsg-NorthEurope	Network security group	North Europe
cosmosopenhackstorage	Storage account	North Europe
sparkpool (cosmos-openhack-synapse/sparkpool)	Apache Spark pool	North Europe

5. In the **Azure Cosmos DB** blade, locate and select the **Overview** link on the left side of the blade. At the top select the **Add Container** button.



Microsoft Azure · Search resources, services, and docs (G+/) · cosmos-openhack-user... SKILLMEUP, LLC (SMUACADEMYL...

Home > Resource groups > cosmos-openhack-shared-rg >

cosmos-openhack Azure Cosmos DB account

Search (Ctrl+/)

Overview · Activity log · Access control (IAM) · Tags · Diagnose and solve problems · Quick start · Notifications · Data Explorer

Settings: Features · Default consistency · Backup & Restore · Firewall and virtual networks · Private Endpoint Connections · CORS · Dedicated Gateway · Keys · Advisor Recommendations · Microsoft Defender for Cloud · Identity · Locks

Integrations: Power BI · Azure Synapse Link

+ Add Container · Refresh · Move · Data Explorer · Delete Account

Essentials · JSON View

Status : Online

Resource group (move) : cosmos-openhack-shared-rg

Subscription (move) : Azure subscription 1

Subscription ID : 7580b5a3-7bfb-40c4-ab43-a6ec643bb7ef

Backup policy : Periodic

Read Locations : North Europe

Write Locations : North Europe

URI : https://cosmos-openhack.documents.azure.com:443/

Capacity mode : Serverless

Containers

Search to filter collections...

ID	Database
sales_orders	database_team25
product_categories	database_team25
categories	database_team25
customers	database_team25
tags	database_team25
products	database_team25
sales_orders_details	database_team25

Monitoring

Show data for last 1 hour · 24 hours · 7 days · 30 days

Requests

6. In the **Add Container** popup, perform the following actions:

   . In the **Database id** field, select the **use existing** option and select the database **database_teamxx (where xx represent your user number)**.

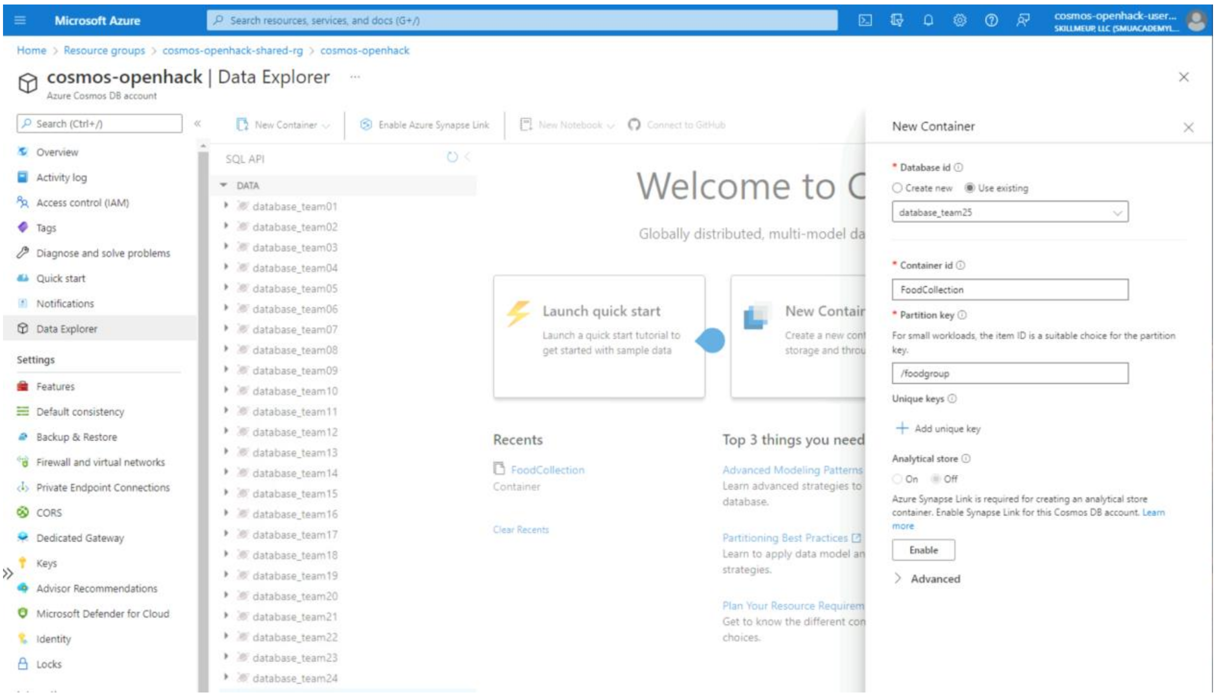
   i. *Note: For Cosmos DB accounts using Provisioned Throughput only*: Do not check the **Provision dedicated throughput** option.

   Provisioning throughput for a database allows you to share the throughput among all the containers that belong to that database. Within an Azure Cosmos DB database, you can have a set of containers which shares the throughput as well as containers, which have dedicated throughput.

   ii. In the **Container Id** field, enter the value **FoodCollection**.

   iii. In the **Partition key** field, enter the value /foodGroup.

   Select the **OK** button.



New Container

* Database id ⓘ

○ Create new ● Use existing

database_team25

* Container id ⓘ

FoodCollection

* Partition key ⓘ

For small workloads, the item ID is a suitable choice for the partition key.

/foodgroup

Unique keys ⓘ

+ Add unique key

Analytical store ⓘ

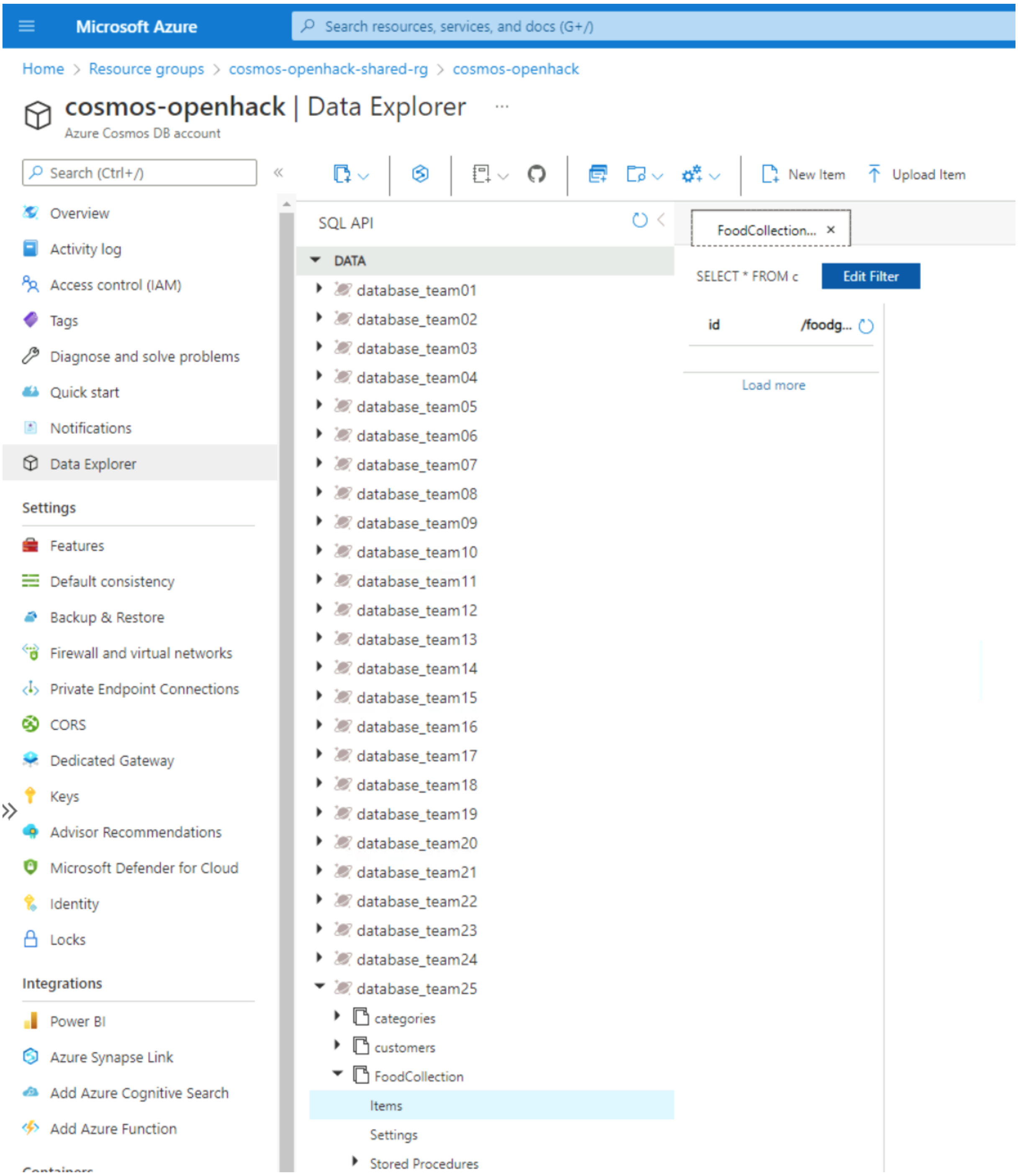
○ On ● Off

Azure Synapse Link is required for creating an analytical store container. Enable Synapse Link for this Cosmos DB account. Learn more

Enable

> Advanced

7. Wait for the creation of the new **container** and verify that the FoodCollection was created with no item for the moment.

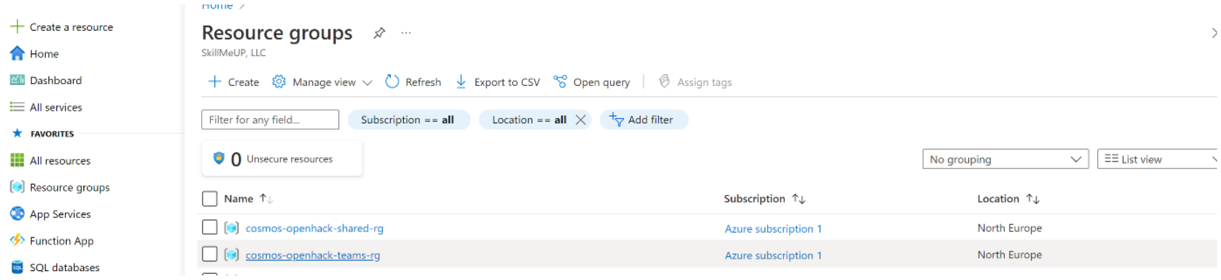


Import Lab Data Into Container

You will use **Azure Data Factory (ADF)** to import the JSON array stored in the **nutrition.json** file from Azure Blob Storage.

To learn more about copying data to Cosmos DB with ADF, please read ADF's documentation

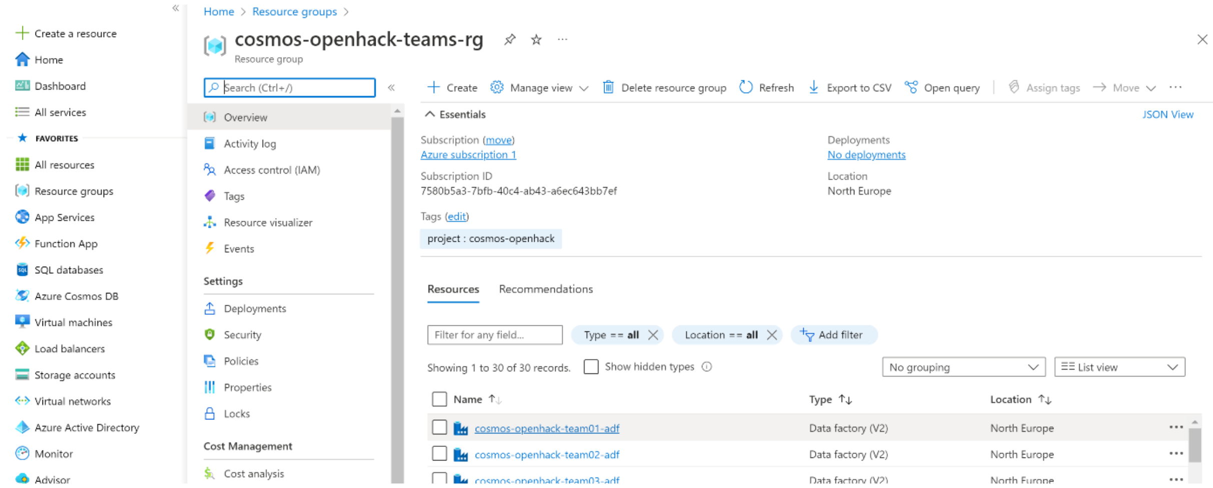
1. On the left side of the portal, select the **Resource groups** link.



Name	Subscription	Location
cosmos-openhack-shared-rg	Azure subscription 1	North Europe
cosmos-openhack-teams-rg	Azure subscription 1	North Europe

2. In the **Resource groups** blade, locate and select the **cosmos-openhack-teams-rg** resource group.

3. Search for the data factory ressource named cosmos-openhack-teamxx-adf where (xx is your user number)



Subscription (move)
Azure subscription 1

Subscription ID
7580b5a3-7bfb-40c4-ab43-a6ec643bb7ef

Tags (edit)
project : cosmos-openhack

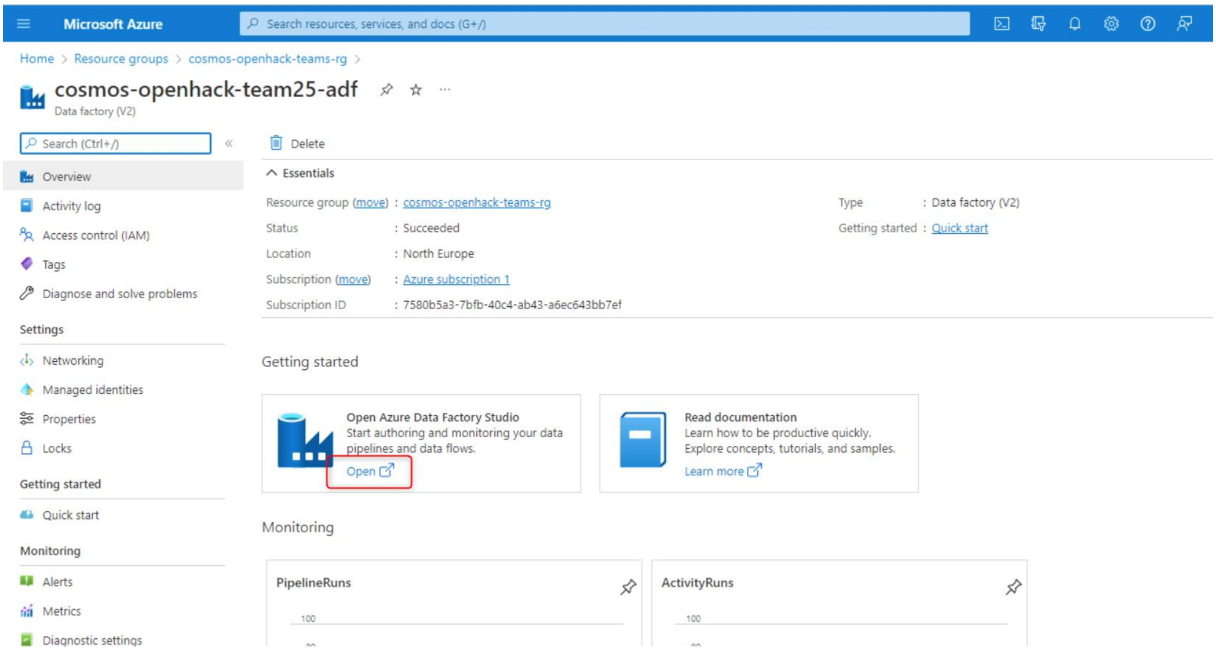
Deployments
No deployments

Location
North Europe

Showing 1 to 30 of 30 records.

Name	Type	Location
cosmos-openhack-team01-adf	Data factory (V2)	North Europe
cosmos-openhack-team02-adf	Data factory (V2)	North Europe
cosmos-openhack-team03-adf	Data factory (V2)	North Europe

4. Select **Open azure data Factory Studio** and you will launch ADF.



cosmos-openhack-team25-adf

Data factory (V2)

Resource group (move) : cosmos-openhack-teams-rg

Status : Succeeded

Location : North Europe

Subscription (move) : Azure subscription 1

Subscription ID : 7580b5a3-7bfb-40c4-ab43-a6ec643bb7ef

Type : Data factory (V2)

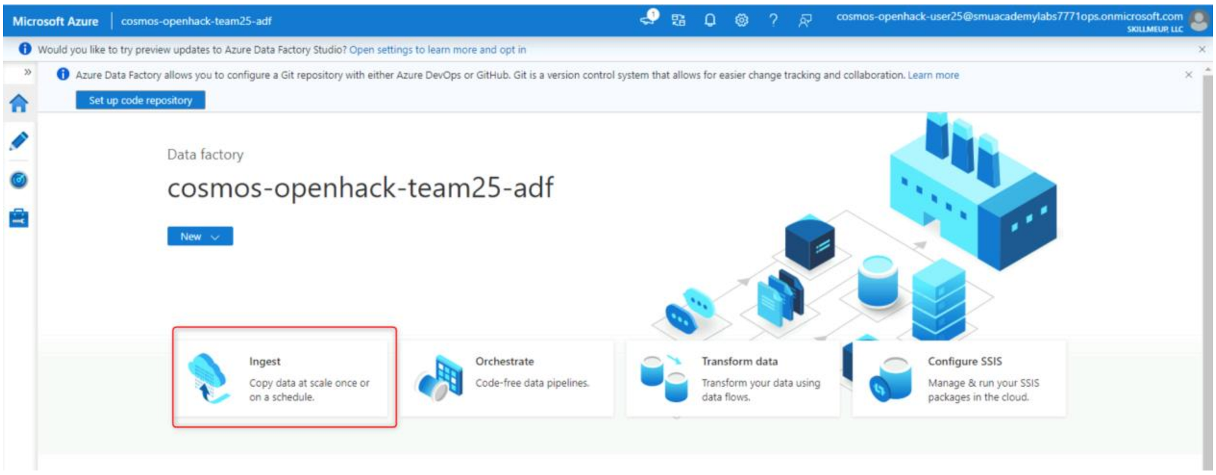
Getting started : Quick start

Open Azure Data Factory Studio: Start authoring and monitoring your data pipelines and data flows.

Read documentation: Learn how to be productive quickly. Explore concepts, tutorials, and samples.

5. Select **Ingest**

   ○ We will be using ADF for a one-time copy of data from a source JSON file on Azure Blob Storage to a database in Cosmos DB's SQL API. ADF can also be used for more frequent data transfers from Cosmos DB to other data stores



Would you like to try preview updates to Azure Data Factory Studio? Open settings to learn more and opt in

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Learn more

Data factory

cosmos-openhack-team25-adf

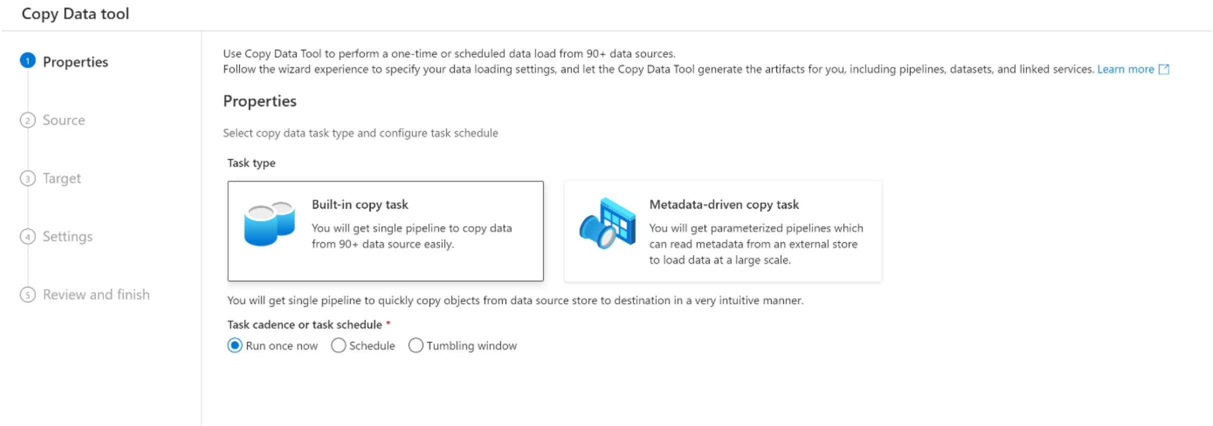
Ingest: Copy data at scale once or on a schedule.

Orchestrate: Code-free data pipelines.

Transform data: Transform your data using data flows.

Configure SSIS: Manage & run your SSIS packages in the cloud.

6. Select Built-in copy task ➔ next



Copy Data tool

1 Properties

2 Source

3 Target

4 Settings

5 Review and finish

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources.
Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. Learn more

Properties

Select copy data task type and configure task schedule

Task type

Built-in copy task
You will get single pipeline to copy data from 90+ data source easily.

Metadata-driven copy task
You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

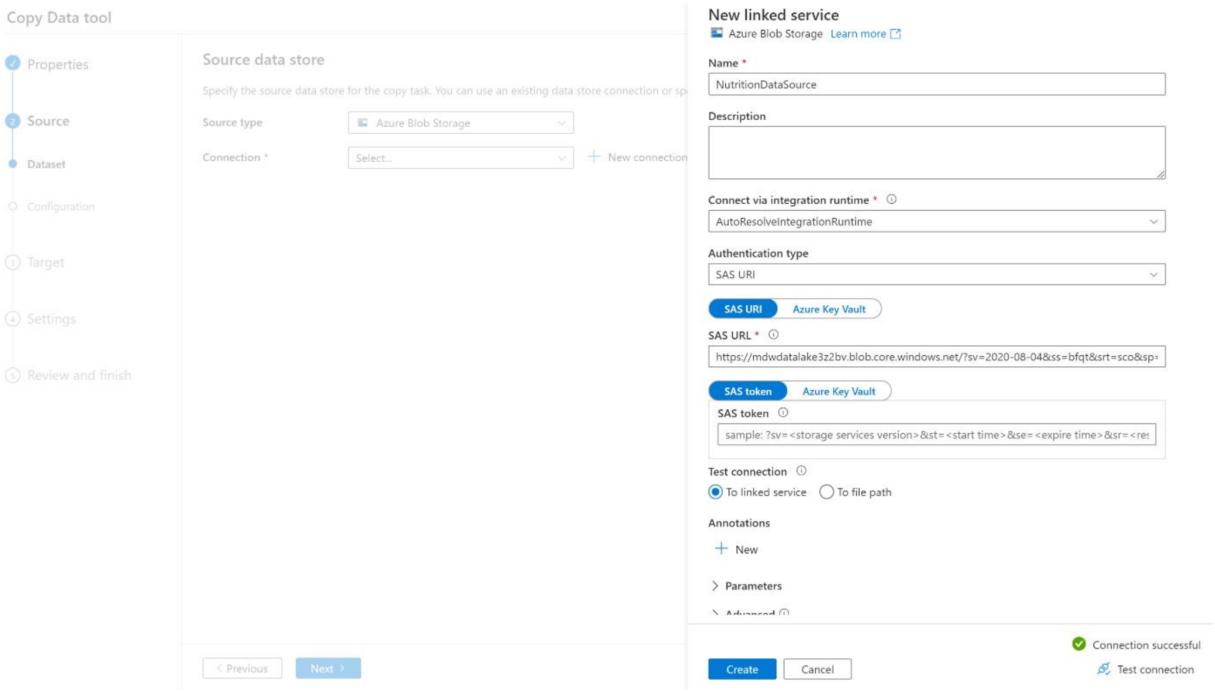
You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

● Run once now ○ Schedule ○ Tumbling window

7. We will import data from a json file on Azure Blob Storage. In addition to Blob Storage, you can use ADF to migrate from a wide variety of sources. We will not cover migration from these sources in this tutorial

   On the Source data store, select Azure Blob storage for the **Source Type** and New connection to create a linked service to the data source



Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or sp

Source type: Azure Blob Storage

Connection *: Select... + New connection

< Previous Next >

New linked service

Azure Blob Storage Learn more

Name *

NutritionDataSource

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication type

SAS URI

SAS URI | Azure Key Vault

SAS URL *

https://mdwdatalake3z2bv.blob.core.windows.net/?sv=2020-08-04&ss=bfqt&srt=sco&sp=

SAS token | Azure Key Vault

SAS token

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

Test connection

● To linked service ○ To file path

Annotations

+ New

> Parameters

> Advanced

Connection successful

Create Cancel Test connection

8. Name the source **NutritionDataSource** and select **SAS URI** as the Authentication method. Please use the following SAS URI for read-only access to this Blob Storage container :

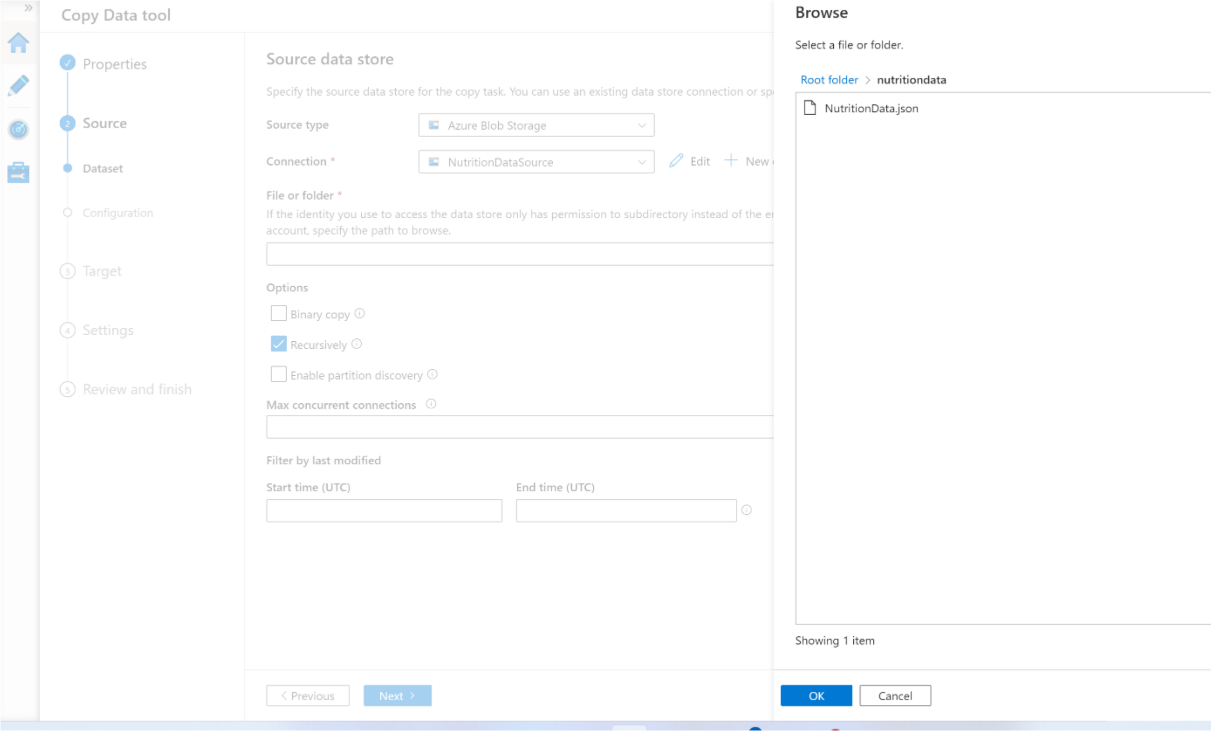
   https://mdwdatalake3z2bv.blob.core.windows.net/?sv=2021-06-08&ss=bfqt&srt=sco&sp=rwdlacupitfx&se=2022-07-15T21:41:03Z&st=2022-07-01T13:41:03Z&spr=https,http&sig=FO5%2FC%2B9FCJ%2F3MUtBKNMvpEFfbMR4t8SEn8xWB8IkvhI%3D

9. Select **Create**

10. Select **Next**

11. Select **Browse**, then double-click to open the **nutritiondata** folder

12. Select the **NutritionData.json** file, then select **Choose**



Copy Data tool

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or sp

Source type: Azure Blob Storage

Connection *: NutritionDataSource

File or folder *

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

Options

□ Binary copy

☑ Recursively

□ Enable partition discovery

Max concurrent connections

Filter by last modified

Start time (UTC)

End time (UTC)

Browse

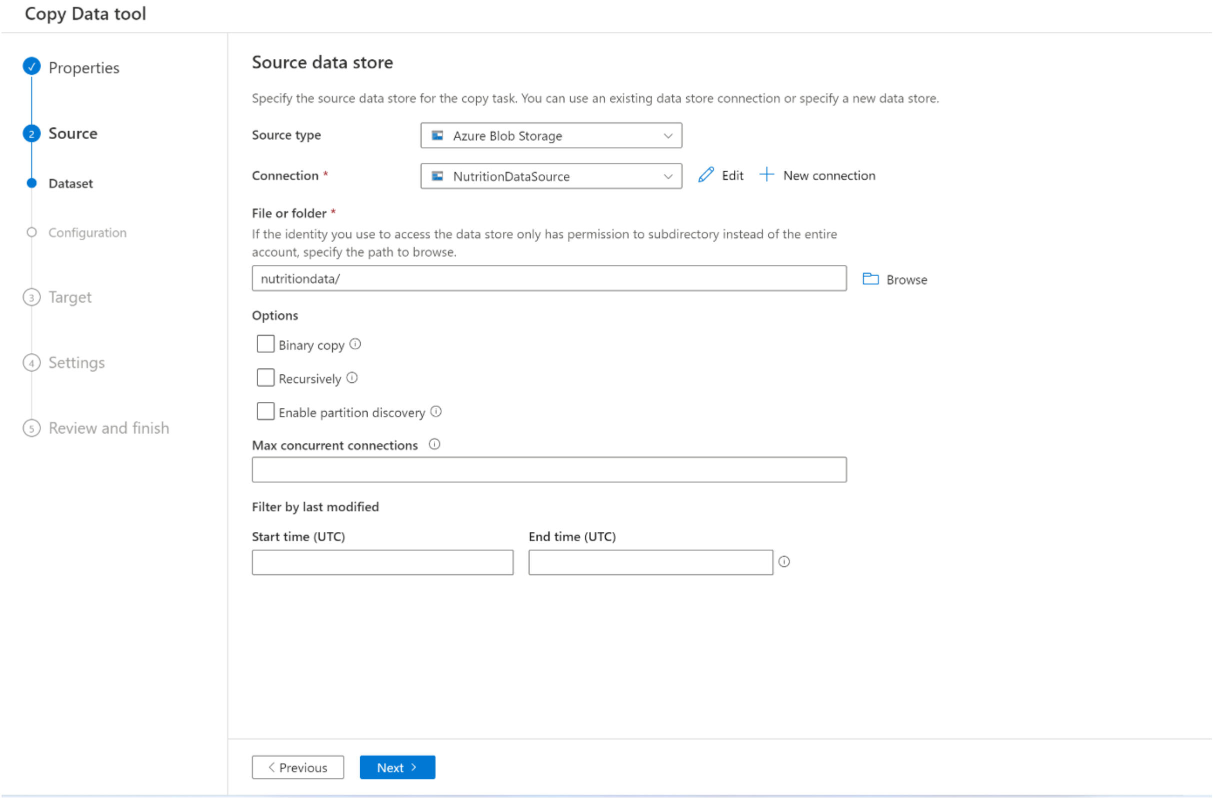
Select a file or folder.

Root folder > nutritiondata

NutritionData.json

Showing 1 item

13. Un-check **Copy file recursively** or **Binary Copy** if they are checked. Also ensure that other fields are empty. Click **Next**



Source data store

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

Source type: Azure Blob Storage

Connection *: NutritionDataSource

File or folder *

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.

nutritiondata/

Options

□ Binary copy

□ Recursively

□ Enable partition discovery

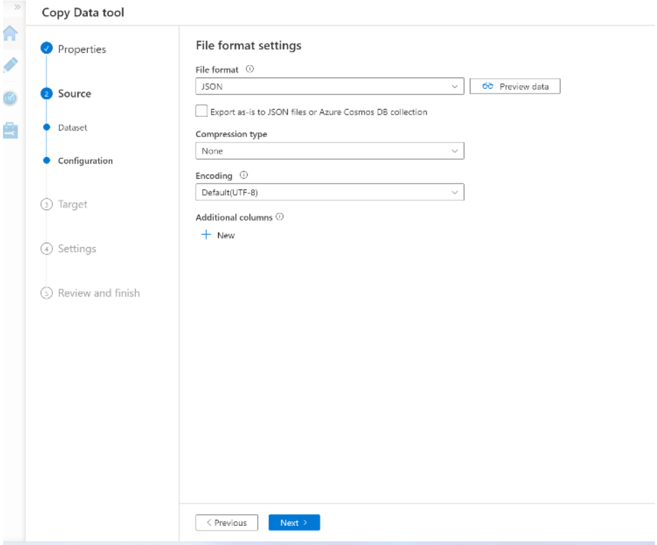
Max concurrent connections

Filter by last modified

Start time (UTC)

End time (UTC)

14. Select the file format as **JSON format**. Then select **Next**.



File format settings

File format: JSON

□ Export as-is to JSON files or Azure Cosmos DB collection

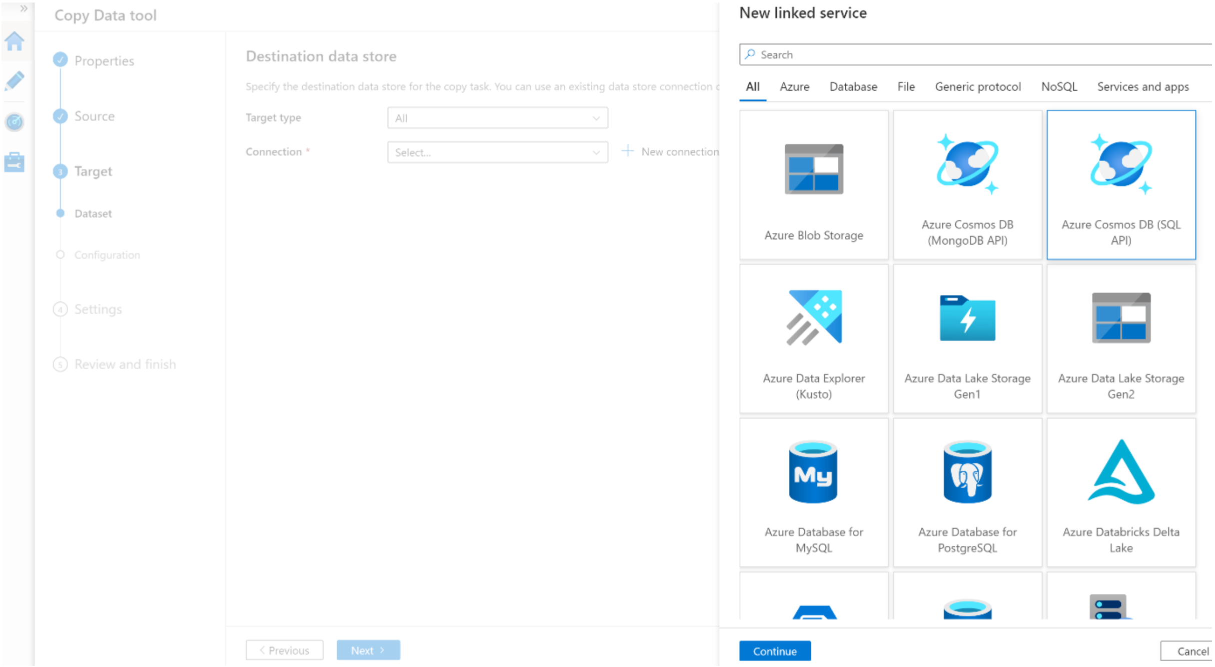
Compression type: None

Encoding: Default(UTF-8)

Additional columns

15. You have now successfully connected the Blob Storage container with the nutrition.json file as the source.

16. For the **Destination data store** add the Cosmos DB target data store by selecting **Create new connection** and selecting **Azure Cosmos DB (SQL API)**.



Destination data store

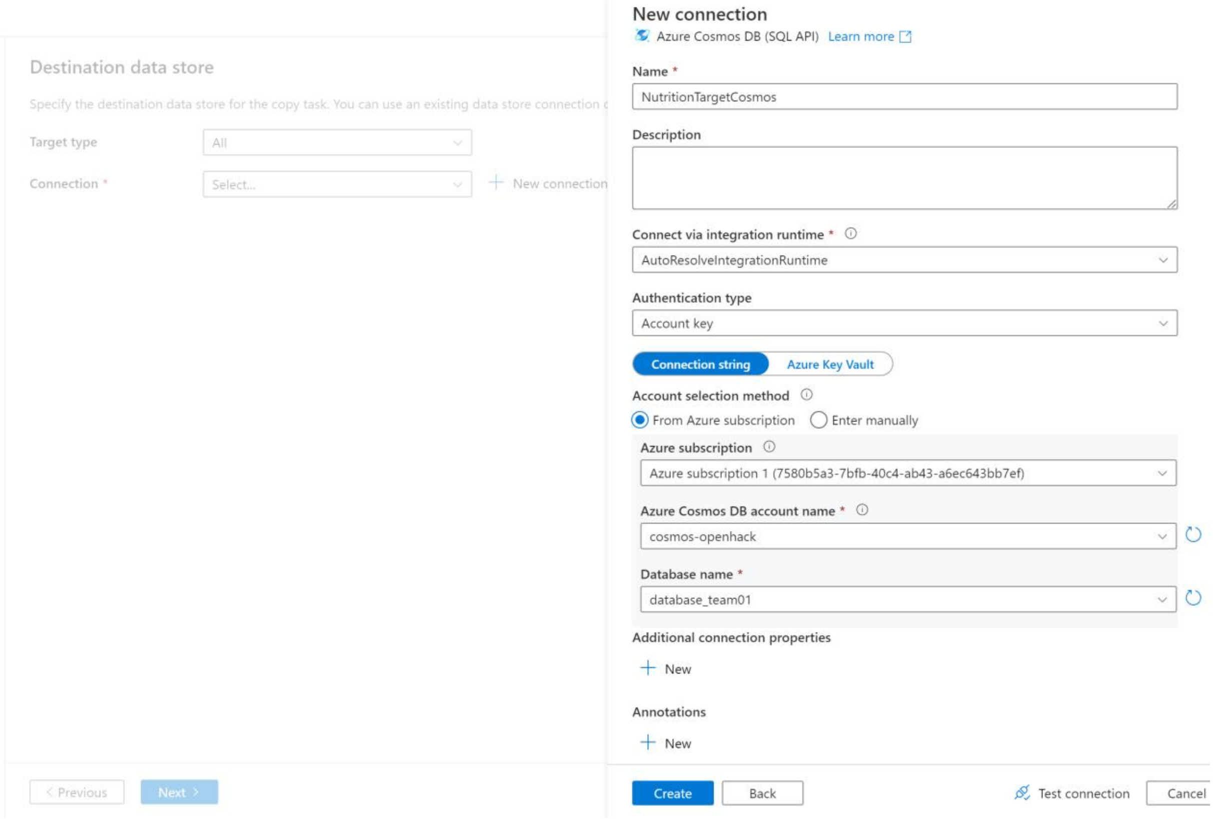
Specify the destination data store for the copy task. You can use an existing data store connection

Target type: All

Connection *: Select...

New connection

17. Name the linked service **NutritionTargetCosmos** and select your Azure subscription and Cosmos DB account. You should also select the Cosmos DB **ImportDatabase** that you created earlier.



Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection

Target type: All

Connection *: Select...

New connection

New connection

Azure Cosmos DB (SQL API)

Name *: NutritionTargetCosmos

Description:

Connect via integration runtime *: AutoResolveIntegrationRuntime

Authentication type: Account key

Account selection method: ● From Azure subscription ○ Enter manually

Azure subscription: Azure subscription 1 (7580b5a3-7bfb-40c4-ab43-a6ec643bb7ef)

Azure Cosmos DB account name *: cosmos-openhack

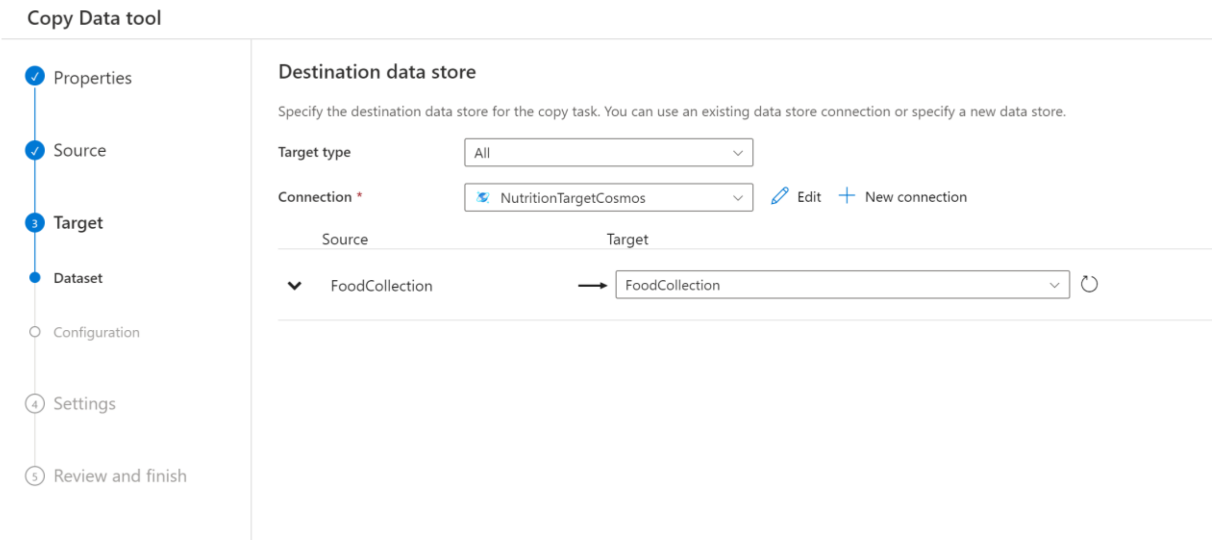
Database name *: database_team01

Additional connection properties

Annotations

18. Select your newly created **targetcosmosdb** connection as the Destination data store.

19. Select your **FoodCollection** container from the drop-down menu. You will map your Blob storage file to the correct Cosmos DB container. Select **Next** to continue.



Destination data store

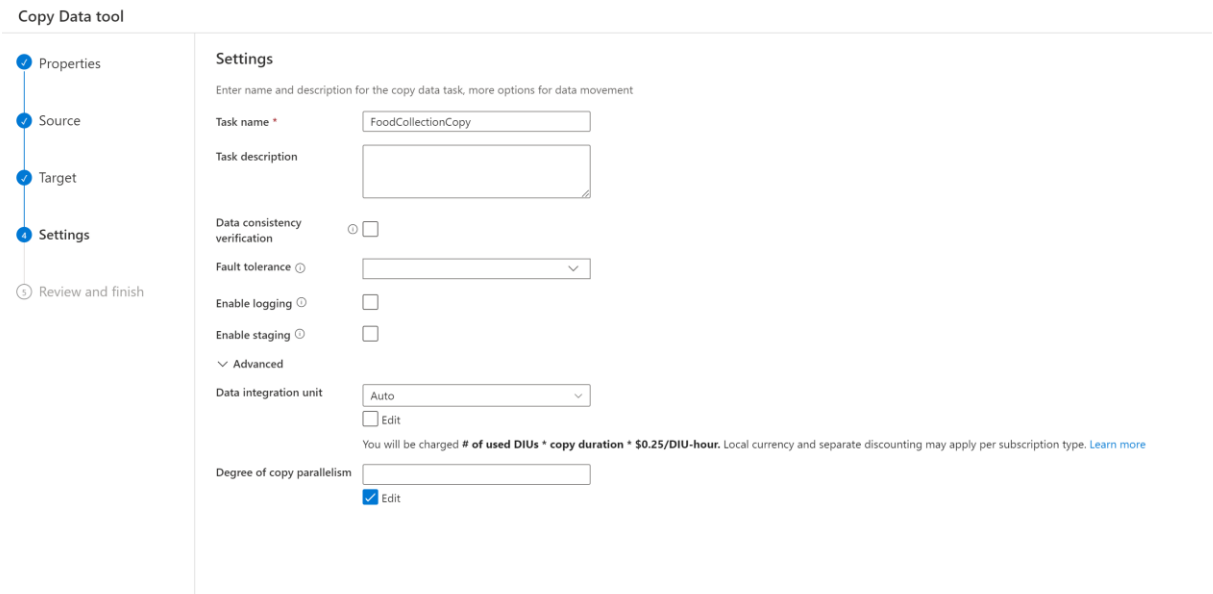
Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Target type: All

Connection *: NutritionTargetCosmos

Source	Target
FoodCollection	FoodCollection

20. Give a name to the copy task and select **next**.



Settings

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

Task name *: FoodCollectionCopy

Task description:

Data consistency verification: □

Fault tolerance:

Enable logging: □

Enable staging: □

Data integration unit: Auto

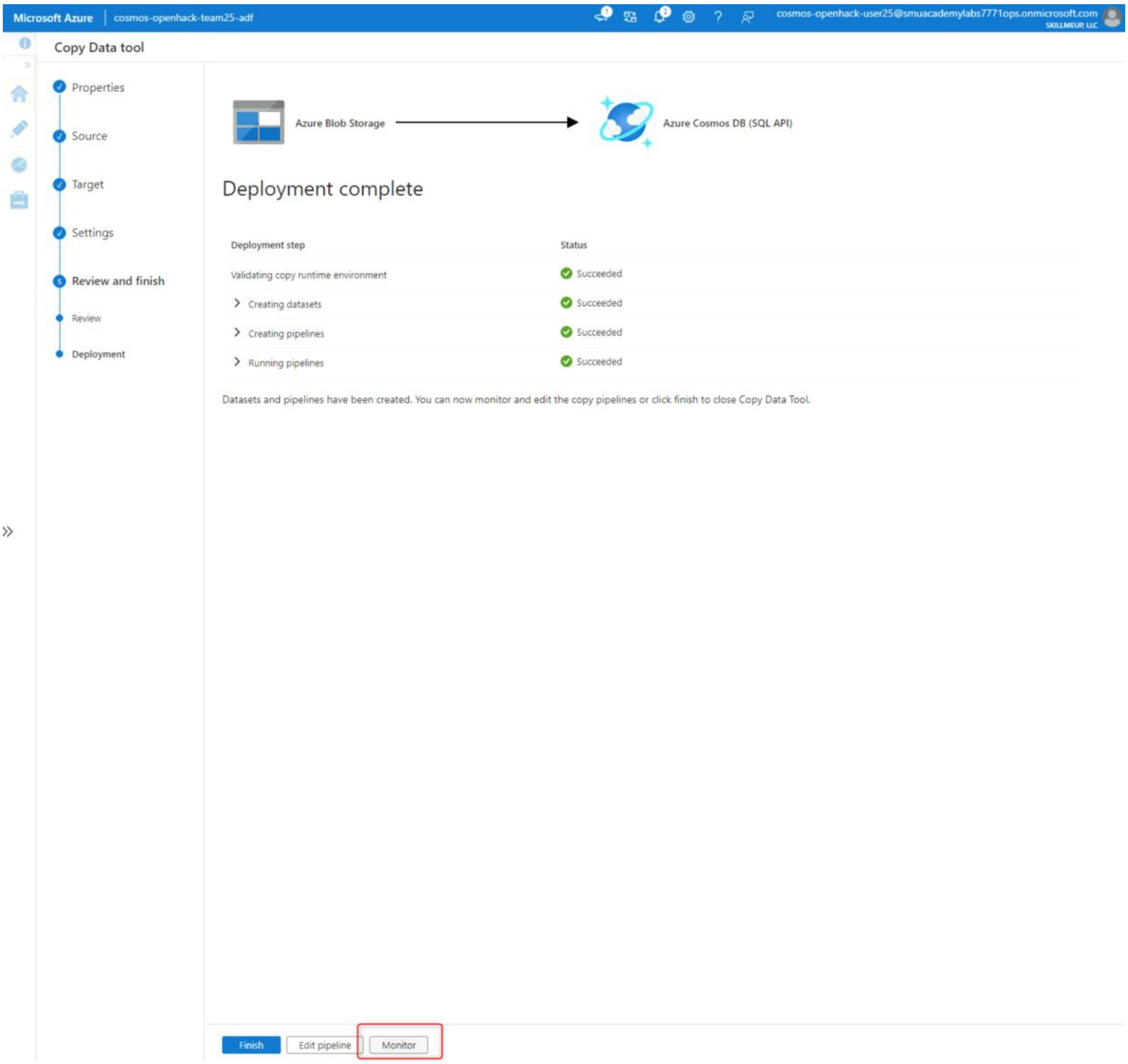
□ Edit

You will be charged **# of used DIUs * copy duration * $0.25/DIU-hour.** Local currency and separate discounting may apply per subscription type. Learn more

Degree of copy parallelism:

☑ Edit

21. Select **Next** to begin deployment After deployment is complete, select **Monitor**.

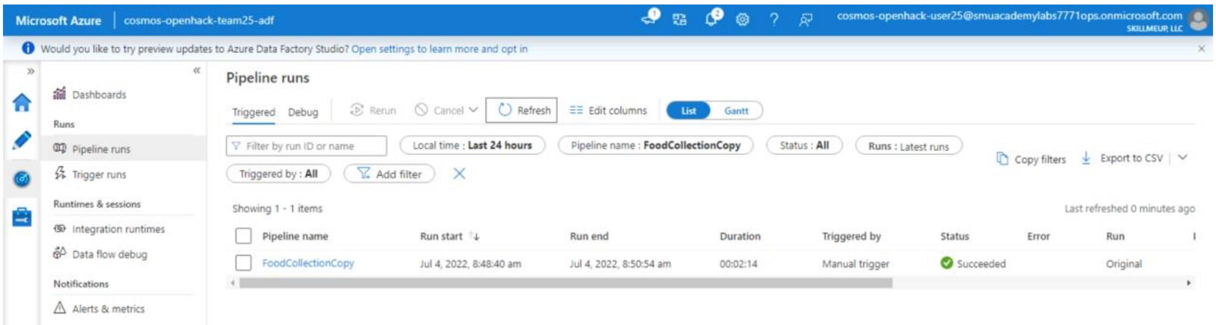


Deployment complete

Deployment step	Status
Validating copy runtime environment	Succeeded
Creating datasets	Succeeded
Creating pipelines	Succeeded
Running pipelines	Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

22. After a few minutes, refresh the page and the status for the ImportNutrition pipeline should be listed as **Succeeded**.



Would you like to try preview updates to Azure Data Factory Studio? Open settings to learn more and opt in

Pipeline runs

Showing 1 - 1 items

Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
FoodCollectionCopy	Jul 4, 2022, 8:48:40 am	Jul 4, 2022, 8:50:54 am	00:02:14	Manual trigger	Succeeded		Original

23. Once the import process has completed, close the ADF. You will now proceed to validate your imported data.

Validate Imported Data

The Azure Cosmos DB Data Explorer allows you to view documents and run queries directly within the Azure Portal. In this exercise, you will use the Data Explorer to view the data stored in our container.

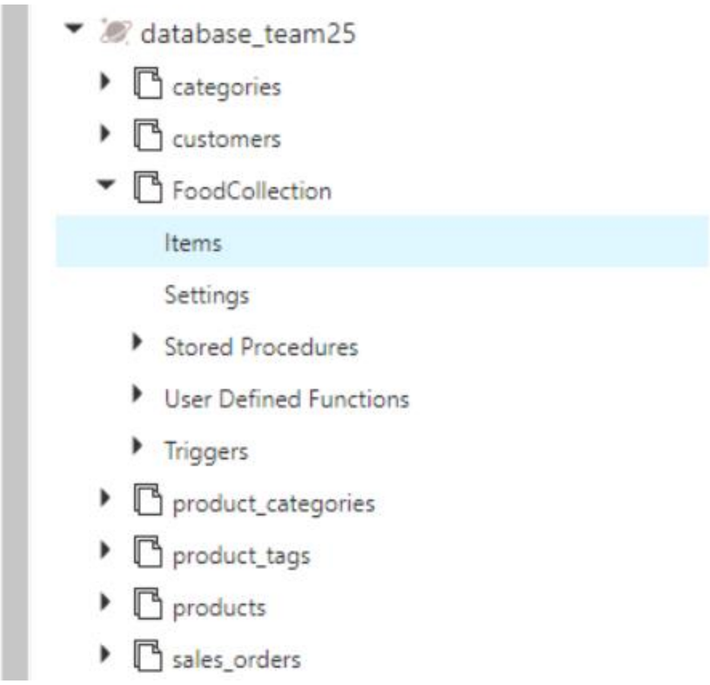
You will validate that the data was successfully imported into your container using the **Items** view in the **Data Explorer**.

1. Return to the **Azure Portal** (http://portal.azure.com).

2. In the **Resource groups** blade, locate and select the **cosmos-openhack-shared-rg** resource group.

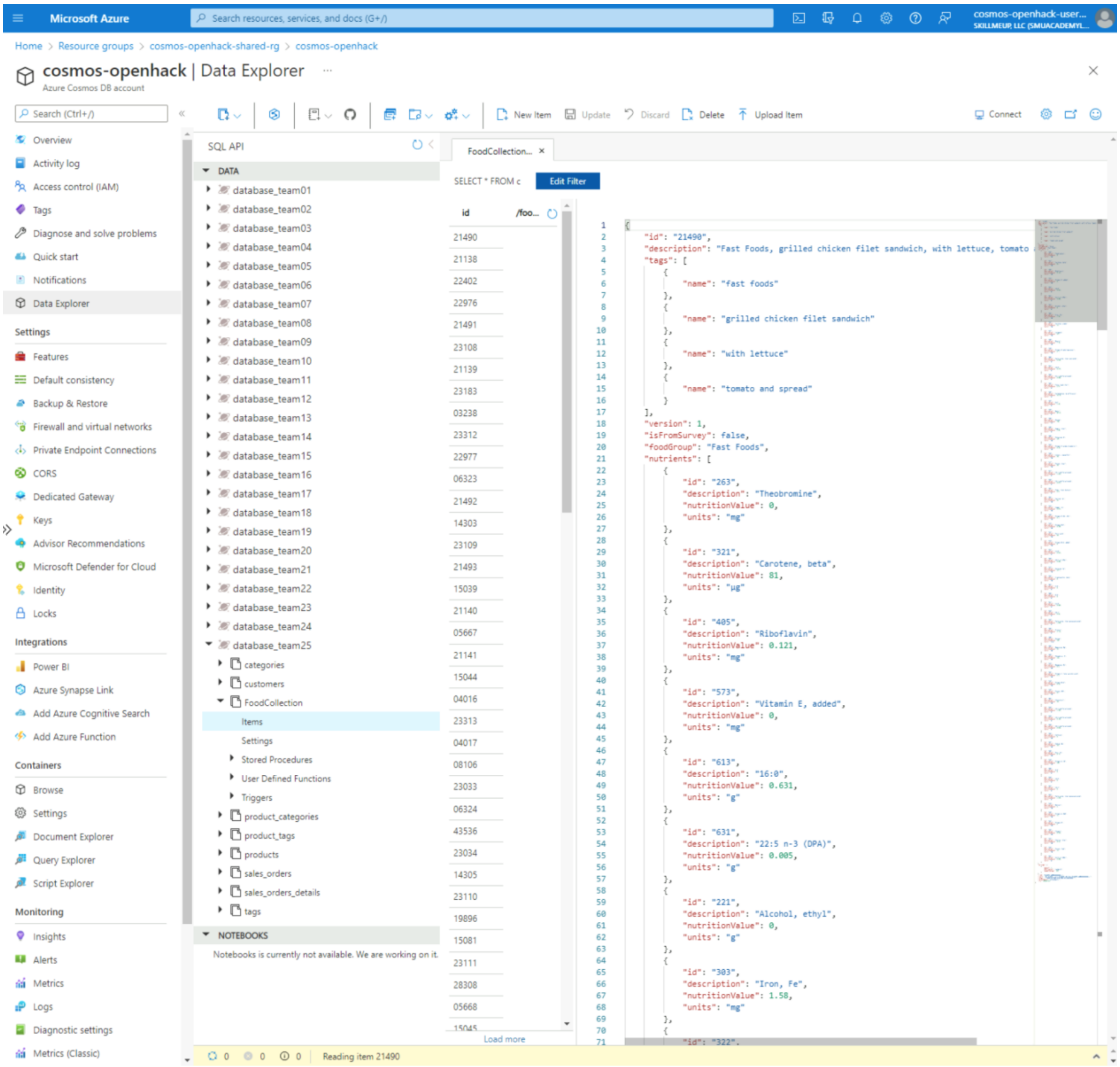
3. In the **Azure Cosmos DB** blade, locate and select the **Data Explorer** link on the left side of the blade.

4. In the **Data Explorer** section, expand the **database_teamxx** database node and then expand the **FoodCollection** container node.



- database_team25
  - categories
  - customers
  - FoodCollection
    - Items
    - Settings
    - Stored Procedures
    - User Defined Functions
    - Triggers
  - product_categories
  - product_tags
  - products
  - sales_orders

5. Within the **FoodCollection** node, select the **Items** link to view a subset of the various documents in the container. Select a few of the documents and observe the properties and structure of the documents.



Notebooks is currently not available. We are working on it.

SELECT * FROM c

id	/foo...
21490	
21138	
22402	
22976	
21491	
23108	
21139	
23183	
03238	
23312	
22977	
06323	
21492	
14303	
23109	
21493	
15039	
21140	
05667	
21141	
15044	
04016	
23313	
04017	
08106	
23033	
06324	
43536	
23034	
14305	
23110	
19896	
15081	
23111	
28308	
05668	
15045	

```
{
    "id": "21490",
    "description": "Fast Foods, grilled chicken filet sandwich, with lettuce, tomato
    "tags": [
        {
            "name": "fast foods"
        },
        {
            "name": "grilled chicken filet sandwich"
        },
        {
            "name": "with lettuce"
        },
        {
            "name": "tomato and spread"
        }
    ],
    "version": 1,
    "isFromSurvey": false,
    "foodGroup": "Fast Foods",
    "nutrients": [
        {
            "id": "263",
            "description": "Theobromine",
            "nutritionValue": 0,
            "units": "mg"
        },
        {
            "id": "321",
            "description": "Carotene, beta",
            "nutritionValue": 81,
            "units": "µg"
        },
        {
            "id": "405",
            "description": "Riboflavin",
            "nutritionValue": 0.121,
            "units": "mg"
        },
        {
            "id": "573",
            "description": "Vitamin E, added",
            "nutritionValue": 0,
            "units": "mg"
        },
        {
            "id": "613",
            "description": "16:0",
            "nutritionValue": 0.631,
            "units": "g"
        },
        {
            "id": "631",
            "description": "22:5 n-3 (DPA)",
            "nutritionValue": 0.005,
            "units": "g"
        },
        {
            "id": "221",
            "description": "Alcohol, ethyl",
            "nutritionValue": 0,
            "units": "g"
        },
        {
            "id": "303",
            "description": "Iron, Fe",
            "nutritionValue": 1.58,
            "units": "mg"
        },
        {
            "id": "322",
```

Reading item 21490