

Data Engineering in the Cloud

Hybrid Transactional Analytical Processing

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- In the lab, we Implement Configuration of Synapse Link with Azure Cosmos DB

Prerequisites

- Prerequisites (5 minutes):
 - ▶ Create a datalake (storage account with hierarchical namespace enabled)
 - ★ To lower the cost, you may choose Redundancy as LRS
 - ▶ Create a container in the storage account
 - ▶ Create an Azure Synapse workspace and
 - ★ create a Apache spark pool in the Manage tab

Lab

- 1 Go to create “resource tab” and search for “Azure Cosmos DB”.



Azure Cosmos DB

- 2 Click on “Create” for Azure Cosmos DB for NoSQL.

[Home](#) > [Azure Cosmos DB](#) >

Create an Azure Cosmos DB account

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL and relational database service for building scalable, high performance applications. [Learn more](#)

To start, select the API to create a new account. The API selection cannot be changed after account creation.

Azure Cosmos DB for NoSQL

Azure Cosmos DB's core or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.

[Create](#) [Learn more](#)

Azure Cosmos DB for PostgreSQL

Fully-managed relational database service for PostgreSQL with distributed query execution, powered by the Citus open source extension. Build new apps on single or multi-node clusters—with support for JSONB, geospatial, rich indexing, and high-performance scale-out.

[Create](#) [Learn more](#)

Azure Cosmos DB for MongoDB

Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.

[Create](#) [Learn more](#)

Azure Cosmos DB for Apache Cassandra

Fully managed Cassandra database service for apps written for Apache Cassandra. Recommended if you have existing Cassandra workloads that you plan to migrate to Azure Cosmos DB.

[Create](#) [Learn more](#)

Azure Cosmos DB for Table

Fully managed database service for apps written for Azure Table storage. Recommended if you have existing Azure Table storage workloads that you plan to migrate to Azure Cosmos DB.

[Create](#) [Learn more](#)

Azure Cosmos DB for Apache Gremlin

Fully managed graph database service using the Gremlin query language, based on Apache TinkerPop project. Recommended for new workloads that need to store relationships between data.

[Create](#) [Learn more](#)

- 3 Then, enter the configuration details, account name (eg. esucosmos), location, and resource group.

[Home](#) > [Azure Cosmos DB](#)

Create Azure Cosmos DB Account - Azure Cosmos DB for NoSQL

[Basics](#) Global distribution Networking Backup Policy Encryption Tags Review & create

Azure Cosmos DB is a fully managed NoSQL and relational database service for building scalable, high performance applications. [Try it for free](#), for 30 days with unlimited renewals. Go to production starting at \$24/month per database, multiple containers included. [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 *

Resource Group *
[Create new](#)

Instance Details

Account Name *

Configure availability zone settings for your account. You cannot change these settings once the account is created.

Availability Zones ☐ Enable ☒ Disable

Location *

Available locations are determined by your subscription's access and availability zone support (if that is enabled). If you don't see or cannot select your desired location, please open a support request for region access. [Click here for more details on how to create a region access request](#)

Capacity mode ☒ Provisioned throughput ☐ Serverless
[Learn more about capacity mode](#)

With Azure Cosmos DB free tier, you will get the first 1000 RUs and 25 GB of storage for free in an account. You can enable free tier on up to one account per subscription. Estimated \$64/month discount per account.

Apply free tier discount ☒ Apply ☐ Do Not Apply

Limit total account throughput ☒ Limit the total amount of throughput that can be provisioned on this account

☒ This limit will prevent unexpected charges related to provisioned throughput. You can update or remove this limit after your account is created.

[Review & create](#) [Previous](#) [Next: Global distribution](#)

- 4 Add other configuration details as required. Then, go to “Review + create” and validate the configuration and click on “Create”.

[Home](#) > [Azure Cosmos DB](#) >

Create Azure Cosmos DB Account - Azure Cosmos DB for NoSQL

✓ Validation Success

[Basics](#) [Global distribution](#) [Networking](#) [Backup Policy](#) [Encryption](#) [Tags](#) [Review + create](#)

Creation Time

Estimated Account Creation Time (in minutes) 3

i The estimated creation time is calculated based on the location you have selected

Basics

Subscription	AzureSubscription
Resource Group	xzhang2
Location	East US 2
Account Name	(new) esucosmos
API	Azure Cosmos DB for NoSQL
Capacity mode	Provisioned throughput
Geo-Redundancy	Disable
Multi-region Writes	Disable
Availability Zones	Disable

Backup Policy

Backup policy	Periodic
Backup storage redundancy	Geo-redundant backup storage

Encryption

Data Encryption	Service Managed
-----------------	-----------------

Networking

Connectivity method	All networks
Minimum TLS Protocol	TLS 1.2

Create

Previous

Next

[Download a template for automation](#)

- 5 Go to the Data Explorer, create a new container called container2

The screenshot shows the Azure Cosmos DB Data Explorer interface. The left sidebar contains a navigation menu with options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Cost Management, Quick start, Notifications, and Data Explorer (which is selected). The main content area displays a welcome message: "Welcome to Azure Cosmos DB" and "Globally distributed, multi-model database service for any scale". Below this, there are three primary action cards: "Launch quick start" (with a lightning bolt icon), "New Container" (with a database icon), and "Connect" (with a plug icon). At the bottom, there are three sections: "Recents" showing a container named "container1", "Top 3 things you need to know" with links to "Advanced Modeling Patterns" and "Partitioning Best Practices", and "Learning Resources" with links to "Data Explorer keyboard shortcuts" and "Get Started using an SDK". A search bar and a "Home" button are visible at the top of the main content area.

Lab

- The settings of container2 are as the following

New Container

With free tier, you'll get the first 1000 RU/s and 25 GB of storage in this account for free. Billing will apply if you provision more than 1000 RU/s of manual throughput, or if the container scales beyond 1000 RU/s with autoscale. [Learn more](#)

Database id

Create new

Use existing

csdb1

Share throughput across containers

Database throughput (autoscale)

Autoscale

Manual

Estimate your required RU/s with [capacity calculator](#)

Database Max RU/s

1000

Your database throughput will automatically scale from 100 RU/s (10% of max RU/s) - 1000 RU/s based on usage.

Estimated monthly cost (USD) \approx \$8.76 - \$87.60 (1 region, 100 - 1000 RU/s, \$0.00012/RU)

Container id

container2

Indexing

Automatic

Off

All properties in your documents will be indexed by default for flexible and efficient queries. [Learn more](#)

Partition key

category

Add hierarchical partition key

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

- 6 Now, to configure the Azure Synapse Link with Azure Cosmos DB, we need to enable Azure Synapse Link. For this, go to the "Azure Cosmos DB account" and then search for Azure Synapse Link.

The screenshot shows the Azure portal interface for an Azure Cosmos DB account named 'esucosmos'. The breadcrumb navigation at the top indicates the path: Home > Microsoft Azure Cosmos DB-20240726143131 > Overview. The account name 'esucosmos' is displayed with a star icon and the text 'Azure Cosmos DB account'. Below this, there is a search bar containing 'syn' and a list of integrations with 'Azure Synapse Link' selected. A welcome message states: 'Welcome to your Azure Cosmos DB Free Tier account! Your first 1000 RU/s and 25 GB of storage will be free for the lifetime of this account. Click here to learn more.' Below this is a purple banner with a message: 'Query with AI using Microsoft Copilot for Azure in Cosmos DB: Ask your subscription admin to enable the preview today.' The 'Essentials' section provides account details: Status (Online), Resource group (fmcsp), Subscription (fmcsp), Subscription ID (c375ef05-a279-41f9-ba5e-23eebfb236d), Total throughput limit (1000 RU/s), Read Locations (East US 2), Write Locations (East US 2), URI (https://esucosmos.documents.azure.com/443/), Free Tier Discount (Opted in), and Capacity mode (Provisioned throughput). The 'Containers' section shows a message: 'Looks like you don't have any containers yet.' The 'Monitoring' section shows a graph for 'Requests' and 'Estimated Cost (hourly)' with a 'Show data for last' dropdown set to '24 hours'.

- Click on “Azure Synapse Link” and then click on “Enable”. It may take 1-2 minutes to be enabled.

The screenshot shows the Azure portal interface for an account named 'esucosmos'. The left-hand navigation pane is open, showing a search bar with 'Azure Synapse Link' entered. Below the search bar, the 'Integrations' section is expanded, and 'Azure Synapse Link' is selected. The main content area displays the 'Enable Azure Synapse Link' page. At the top of this page, there are tabs for 'Enable Azure Synapse Link' (which is active), 'Select Workspace', and 'Link Synapse Analytics'. The page title is 'Enable Azure Synapse Link'. Below the title, there is a brief description: 'Enable Azure Synapse Link to run near real-time analytics over operational data in Azure Cosmos DB. [Learn more.](#) Here are a few analytics use cases you can build on your Azure Cosmos DB data.' A prominent blue 'Enable' button is located below this text. Further down, under the heading 'Enable Azure Synapse Link for your containers', there is a note: 'After Synapse Link is enabled on your Azure Cosmos DB database account, you can choose which containers will be enabled for analytics with Synapse Link. Enabling Synapse Link on your containers will have cost implications. [Learn more.](#)' Below this note, there is a section titled 'Select containers to enable' with two checkboxes: 'cdb1' (which is checked) and 'container2' (which is unchecked). At the bottom of the page, there is a 'Give Feedback' section with a link to 'Help improve this page' and a 'Next' button.

Lab


- 8 Now go back to “Azure Cosmos DB account” home and click the “Data Explorer” tab to create a another container, called container1. However, enable the analytical store on the container.

The screenshot shows the Azure Cosmos DB Data Explorer interface for an account named 'esucosmos'. The left sidebar contains a navigation menu with the following items: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Cost Management, Quick start, Notifications, and Data Explorer (which is currently selected). Below the 'Data Explorer' item are sub-items: Settings, Integrations, Containers, Monitoring, and Automation. The main content area displays the 'Data Explorer' view. At the top, it shows 'esucosmos | Data Explorer' and 'Azure Cosmos DB account'. Below this is a search bar and a breadcrumb trail: Home > esucosmos. The main area shows a tree view of the database structure: NOSQL API > DATA > cdb1 > Scale > container2. A context menu is open over 'container2', showing options: 'New Container' and 'Delete Database'. A notification banner at the top right reads: 'To prevent queries from using exces'.

Lab

- 9 Click on “New container” and enter the details. Make sure the analytical store is on. Click on OK.

New Container



Your account currently has at least 1 database or container with provisioned RU/s. Billing will apply to this container if the total RU/s in your account exceeds 1000 RU/s. [Learn more](#)

* Database id ⓘ

☐ Create new ☒ Use existing

cdb1

* Container id ⓘ

container1

* Indexing

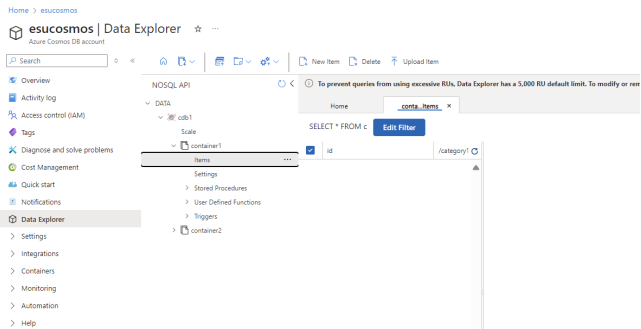
☒ Automatic ☐ Off

All properties in your documents will be indexed by default for flexible and efficient queries. [Learn more](#)


* Partition key ⓘ

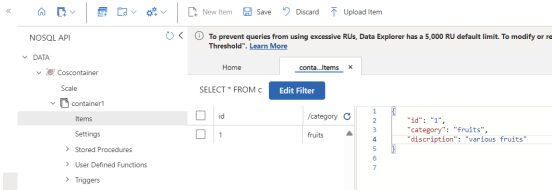
/category1

- 10 Go to “Items” under the container



Lab

-  And click on “New Item” to add a new item. We create two new items and enter the random data as shown below and click on “Save”.
 - ▶ Make sure the value of “id” is different in the two items. Write the code as given below in the script.

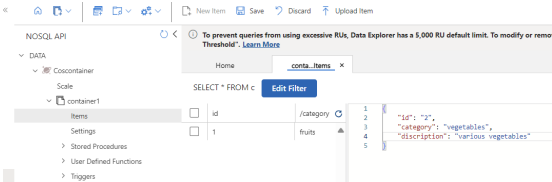


The screenshot shows the Azure Data Explorer interface. On the left, the 'DATA' pane shows a hierarchy: 'DATA' > 'CosmosContainer' > 'Scale' > 'container1' > 'Items'. The 'Items' collection is selected. The main pane shows a query editor with the following query:

```
SELECT * FROM c
```

The query results pane on the right shows a table with 7 rows. The first row is highlighted, showing the following JSON data:

```
{
  "id": "1",
  "category": "fruits",
  "discription": "various fruits"
}
```



The screenshot shows the Azure Data Explorer interface. On the left, the 'DATA' pane shows a hierarchy: 'DATA' > 'CosmosContainer' > 'Scale' > 'container1' > 'Items'. The 'Items' collection is selected. The main pane shows a query editor with the following query:

```
SELECT * FROM c
```

The query results pane on the right shows a table with 5 rows. The first row is highlighted, showing the following JSON data:

```
{
  "id": "2",
  "category": "vegetables",
  "discription": "various vegetables"
}
```

- 12 After you click on “Save”, you will see

Navigation: Home | New Item | Update | Discard | Delete | Upload Item

NOSQL API

DATA

- Coscontainer
 - Scale
 - container1
 - Items

	id	/category
<input type="checkbox"/>	1	fruits
<input checked="" type="checkbox"/>	2	veg...
 - Settings
 - Stored Procedures
 - User Defined Functions
 - Triggers

SELECT * FROM c [Edit Filter](#)

Home [conta_items](#) x


To prevent queries from using excessive RUs, Data Explorer has a 5,000 RU default limit. To modify or remove the limit, go to the Settings cog on the right and find 'Threshold'. [Learn More](#)

```

1 {
2   "id": "2",
3   "category": "vegetables",
4   "discription": "various vegetables",
5   "_rid": "wR9oAHf+J4ACAAAAAAAAA==",
6   "_self": "dbs/wR9oAHf+J4ACAAAAAAAAA==/colls/wR9oAHf+J4ACAAAAAAAAA==/",
7   "etag": "\1c004000-0000-0200-0000-06a3f2a70000\"",
8   "_attachments": "attachments/",
9   "_ts": 1722020519
10 }
  
```


- 13 Now, in the left pane, go to “Keys” under the Settings and note down the name of Azure Cosmos DB account and the primary key for later use.

Home > esucosmos

 **esucosmos | Keys** ☆ ...
Azure Cosmos DB account


Refresh Feedback


- Quick start
- Notifications
- Data Explorer
- Settings
 - Features
 - Replicate data globally
 - Default consistency
 - Backup & Restore
 - Networking
 - CORS
 - Dedicated Gateway
 - Keys**
 - Advisor
 - Recommendations
 - Microsoft Defender for Cloud
 - Identity


URI
`https://esucosmos.documents.azure.com:443/`


Read-write Keys

Read-only Keys

PRIMARY KEY 
`Im1RN8jTNIeZOMNTW0jKEsankALwECeloHmdC34KbclywPy4GHn3wlo7RJp2vMjIw1eCxoFOLYACDbKcNRjw==`

 Last regenerated: 7/26/2024 (0 days ago). [Learn more](#)

SECONDARY KEY 
`opO4T73DeWKGozH3vmheSfaDISIP85VrMLR42FvqFUSDOQNZJ60tfwMvnHRMRwbEWv8bzUM8SDrEACDbOagH5w==`

 Last regenerated: 7/26/2024 (0 days ago). [Learn more](#)

PRIMARY CONNECTION STRING
`AccountEndpoint=https://esucosmos.documents.azure.com:443/;AccountKey=Im1RN8jTNIeZOMNTW0jKEsankALwECeloHmdC34KbclywPy4GHn3wlo7RJp2vMjIw1eCxoFOLYACDbKcN...`

SECONDARY CONNECTION STRING
`AccountEndpoint=https://esucosmos.documents.azure.com:443/;AccountKey=opO4T73DeWKGozH3vmheSfaDISIP85VrMLR42FvqFUSDOQNZJ60tfwMvnHRMRwbEWv8bzUM8SDrEACD...`

Copy to clipboard

Lab

- For example, mine is

URI

`https://esucosmos.documents.azure.com:443/`

PRIMARY KEY

`Im1RN8JtTN1eZOMNTW0jKEsanKaLwECeIoHmdC34Kbc1lywPy4GHn3w1o7RJp2vMj1w1eCxofOLYACDbKcN`

SECONDARY KEY

`op04T73DeWKG0zH3vmheSfaDISiP85VrMLR42FVqFUSD0QNZJ60tfwMvnHRMRwbEWv8bzUM8SDrEACDb0agf`

PRIMARY CONNECTION STRING

`AccountEndpoint=https://esucosmos.documents.azure.com:443/;AccountKey=Im1RN8JtTN1eZ`

SECONDARY CONNECTION STRING

`AccountEndpoint=https://esucosmos.documents.azure.com:443/;AccountKey=op04T73DeWKG0z`

- Note: To create a linked service, there must be a synapse analytics workspace with an Apache spark pool. Create them if you do not have them yet.

- 14 Open “synapse studio” and go to the “Manage” section. Go to Linked services, and click on “+ New” to create a new linked service.

Microsoft Azure | Synapse Analytics | esusrnapsae

Synapse live Validate all Publish all

Home

Data

Develop

Integrate

Monitor

Manage

Analytics pools

SQL pools

Apache Spark pools

Data Explorer pools (preview)

External connections

Linked services

Microsoft Purview

Integration

Triggers

Integration runtimes

Security

Access control

Linked services

Linked services are much like connection strings, which define the connection information needed for Azure Synapse Analytics to connect to ext

[Learn more](#)

+ New

Filter by name

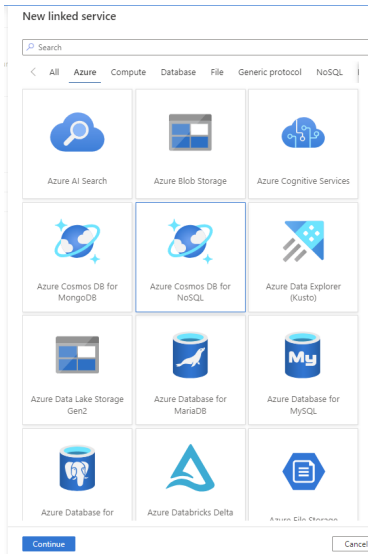
Annotations: Any

Showing 1 - 2 of 2 items

Name	Type	Related
esusrnapsae-WorkspaceDefaultSqlServer	Azure Synapse Analytics	0
esusrnapsae-WorkspaceDefaultStorage	Azure Data Lake Storage Gen2	0

Lab

- 15 Select “Azure Cosmos DB for NoSQL” and click on “Continue”.



- 16 Specify the details as shown below:

New linked service

[Azure Cosmos DB for NoSQL](#) [Learn more](#)

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

Name *

CosmosDbNoSql1

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 ⓘ

AzureSubscription (c375ef05-a27a-41f9-ba5e-20eebf29c6)

Azure Cosmos DB account name * ⓘ

esucosmos

Database name *

cdb1

Additional connection properties

+ New

Annotations

Create Back Test connection Cancel

- 17 Click on “Test connection” and click on “Create”

New linked service

Azure Cosmos DB for NoSQL. [Learn more](#)

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

Name *

CosmosDbNoSql1

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

AzureSubscription (c375ef05-a27a-41f9-ba5e-20eebfeb29c6)

Azure Cosmos DB account name *

esucosmos

Database name *

cdb1

Additional connection properties

+ New

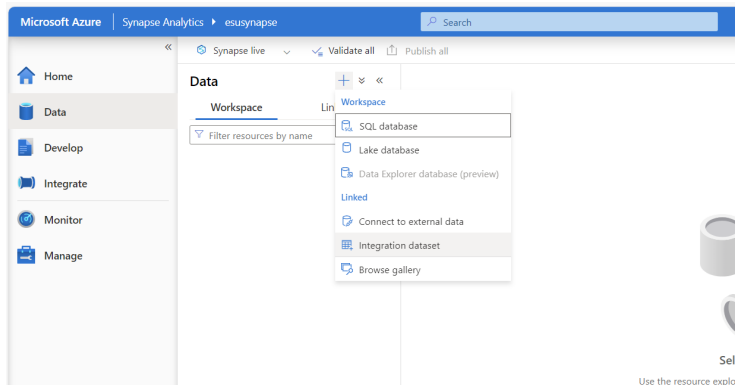
Connection successful

Test connection Cancel

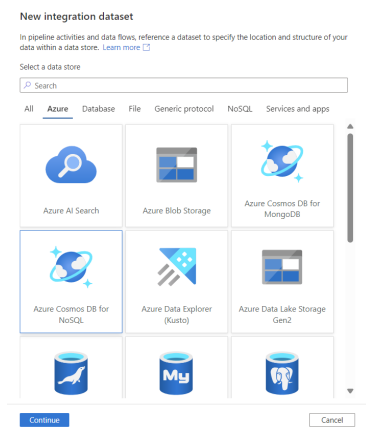
Create Back

Lab

- 18 Now, go to the “Data” section of the synapse studio, click on “+” and choose “Integration dataset” to create a dataset.



- 19 Select “Azure Cosmos DB for NoSQL” and click on “Continue”.



Lab

- 20 Set the properties as shown and click on "OK".

Set properties

Name

CosmosDbNoSqlContainer1

Linked service *

CosmosDbNoSql1

Connect via integration runtime * ⓘ



AutoResolveIntegrationRuntime

Container

container1

☐

Enter manually

Import schema



From connection/store



None

OK

Back

Cancel

- 21 In the below section, go to “Connections” and click on “Preview data” to view the data.

The screenshot displays the Synapse Studio interface. At the top, there are tabs for 'Synapse live', 'Validate all', and 'Publish all'. Below this, the 'Data' section is active, showing a 'Workspace' view with a search bar labeled 'Filter resources by name'. A notification banner at the top right states: 'Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.'

The main content area shows the configuration for a connection to 'CosmosDbNoSqlContainer1'. The 'Connection' tab is selected, showing the following settings:

- Linked service ***: A dropdown menu showing 'CosmosDbNoSql1'.
- Integration runtime ***: A dropdown menu showing 'AutoResolveIntegrationRuntime' with an 'Edit' link.
- Container ***: A text input field containing 'container1' and a 'Preview data' link.
- Enter manually**: A checked checkbox.

On the right side, the 'Properties' panel is visible, showing the 'General' tab with the following details:

- Name ***: 'CosmosDbNoSqlContainer1'
- Description**: An empty text area.
- Annotations**: A '+ New' button.

Synapse live
Validate all
Publish all

Data
+
✖
«

CosmosDbNoSqlCo...
Other users in your workspace may have access to modify this item. Do not use this item until the workspace.

Workspace
Linked

Filter resources by name

Preview data

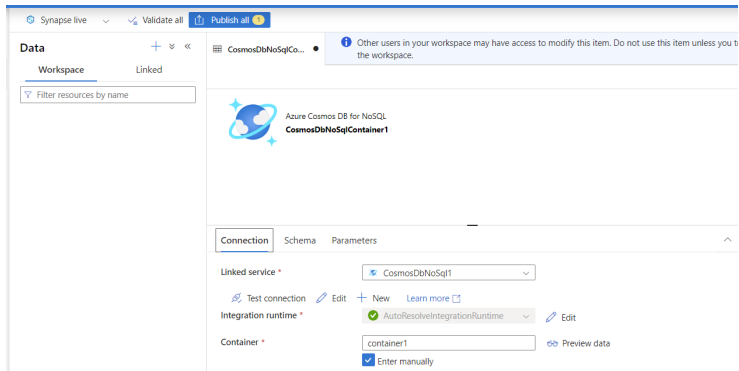
Linked service: CosmosDbNoSql1

Object: container1

```
[
  {
    "id": "1",
    "category": "fruits",
    "discription": "various fruits",
    "_rid": "wR9oAMf+J4ABAAAAAAAAA==",
    "_self": "dbs/wR9oAA==/colls/wR9oAMf+J4A=/docs/wR9oAMf+J4ABAAAAAAAAA==/",
    "_etag": "\"1c00240b-0000-0200-0000-66a3f1c20000\"",
    "_attachments": "attachments/",
    "_ts": 1722020290
  },
  {
    "id": "2",
    "category": "vegetables",
    "discription": "various vegetables",
    "_rid": "wR9oAMf+J4ACAAAAAAAAA==",
    "_self": "dbs/wR9oAA==/colls/wR9oAMf+J4A=/docs/wR9oAMf+J4ACAAAAAAAAA==/",
    "_etag": "\"1c00400b-0000-0200-0000-66a3f2a70000\"",
    "_attachments": "attachments/",
    "_ts": 1722020519
  }
]
```

Lab

- 22 Now, click on “Publish All”.
 - ▶ Then, next we create and run a pipeline.



Lab

- 23 Go to the “Integrate” tab of the “synapse studio”
- 24 Click on “+” and choose “Pipeline”.

Microsoft Azure | Synapse Analytics | esusynapse

Synapse live | Validate all | Publish all

Integrate

Filter resources by name

Other users in your workspace may have access to modify this item, the workspace.

Try creating a new item using the + button above. [Learn more](#)

Linked service * | CosmosDbNoSql1

Test connection | Edit | + New | [Learn more](#)

Integration runtime * | AutoResolveIntegrationRuntime | Edit

Container * | container1 | Preview data

☒ Enter manually

Lab

- 25 In the left pane, under “Move & transform”, drag and drop “Copy data” into the window.
- 26 Specify a name to the pipeline.

The screenshot displays the Azure Databricks workspace interface. On the left, the 'Integrate' sidebar is open, showing a search bar and a list of activities under 'Move and transform', including 'Copy data'. The main canvas shows 'Pipeline 1' with the 'Copy data' activity added and named 'Copydata1'. Below the canvas, the 'Properties' panel is visible, showing the 'General' tab with fields for 'Name' (set to 'Copydata1') and 'Description'. The 'Activity state' is set to 'Activated'.

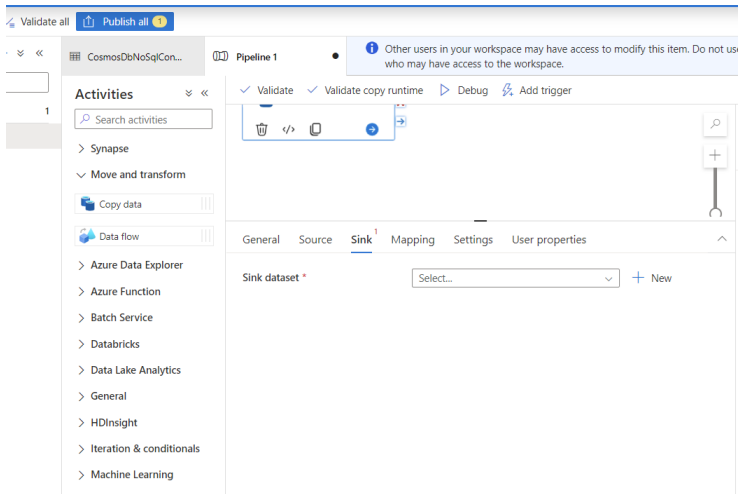
Lab

- 27 In the bottom pane, go to “Source” and select the “Source dataset” from the drop-down.

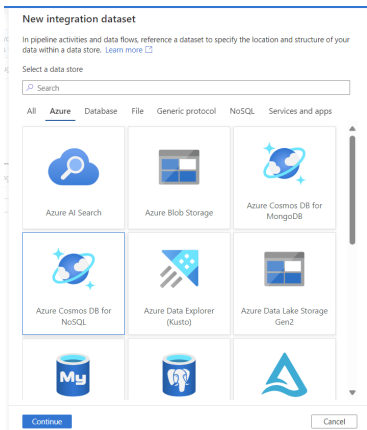
The screenshot displays the Azure Databricks workspace interface. At the top, there's a 'Publish all' button. Below it, the 'CosmosDbNoSqlCon...' workspace is selected. The 'Activities' pane on the left lists various tasks, with 'Data flow' currently selected. The main area shows 'Pipeline 1' with a warning message: 'Other users in your workspace may have access to modify this item. Do not us who may have access to the workspace.' Below the warning, there are buttons for 'Validate', 'Validate copy runtime', 'Debug', and 'Add trigger'. A toolbar contains icons for deleting, editing code, copying, and adding new components. The configuration tabs at the bottom are 'General', 'Source', 'Sink', 'Mapping', 'Settings', and 'User properties', with 'Source' being the active tab. In the 'Source' tab, the 'Source dataset' is set to 'CosmosDbNoSqlContainer1'. There are links for 'Open', 'New', 'Preview data', and 'Learn more'. The 'Use query' section has 'Container' selected. The 'Page size' is set to 10. 'Detect datetime' is checked. 'Preferred regions' has a 'New' button. 'Additional columns' also has a 'New' button.

Lab

- 28 Go to “Sink” and click on “+ New”.



- 29 Select “Azure Cosmos DB for NoSQL” and click on “Continue”.



Lab


- 30 Specify the property details as shown below, and click on “OK”.
 - ▶ We use the same container

Set properties


Name


Linked service *

▼




Connect via integration runtime * ⓘ

 AutoResolveIntegrationRuntime ▼



Container

▼


☐ Enter manually

Import schema

☒ From connection/store ☐ None

> Advanced

- 31 Select “Insert” in the below pane under Sink.

The screenshot displays the Synapse Studio interface for configuring a 'Copy data' activity within 'Pipeline 1'. The left sidebar shows the 'Integrate' workspace with 'Pipeline 1' selected. The main canvas shows the 'Copy data' activity, which is currently in the 'Sink' configuration tab. The 'Sink dataset' is set to 'CosmosDbNoSqlContainer2'. The 'Write behavior' dropdown menu is open, and 'Insert' is selected. Other configuration options like 'Write batch timeout', 'Write batch size', 'Max concurrent connections', and 'Disable performance metrics analytics' are visible but not yet configured.

Integrate + » «
 Filter resources by name
 Pipelines 1
 Pipeline 1

Activities » «
 Search activities
 > Synapse
 > Move and transform
 Copy data
 > Data flow
 > Azure Data Explorer
 > Azure Function
 > Batch Service
 > Databricks
 > Data Lake Analytics
 > General
 > HDInsight
 > Iteration & conditionals
 > Machine Learning

Pipeline 1
 Other users in your workspace may have access to modify this item. Or who may have access to the workspace.
 Validate Validate copy runtime Debug Add trigger

Copy data
 Copydata1
 General Source Sink Mapping Settings User properties

Sink dataset * CosmosDbNoSqlContainer2
 Open + New Learn more

Write behavior Insert
 Write batch timeout
 Write batch size
 Max concurrent connections
 Disable performance metrics analytics

- 32 Click on “Publish All” to save the pipeline.
- 33 Once the publish is successful, your pipeline will be ready to run.

The screenshot displays the Azure Synapse Studio interface. On the left, the 'Integrate' workspace is active, showing a list of pipelines with 'Pipeline 1' selected. The main canvas shows a pipeline named 'Pipeline 1' with a single activity, 'Copy data'. The 'Output' tab is selected, displaying the pipeline run ID: 9732c1a1-bbf1-43d5-8c9c-34e9fe0762cb. The pipeline status is 'In progress'. Below the status, a table shows the activity details:

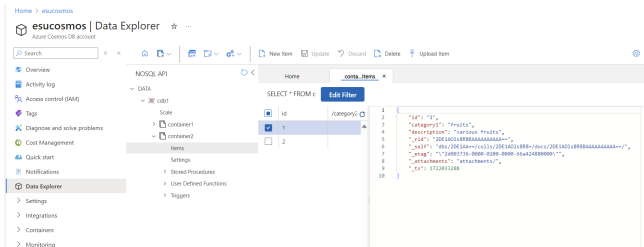
Activity name	Activity status	Activity type	Run start
Copydata1	Queued	Copy data	7/26/2024, 3:45:58

- 34 Check if the pipeline can be run successfully

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Integrate' tab is selected, showing a list of pipelines with 'Pipeline 1' highlighted. The main workspace displays the 'Pipeline 1' canvas, which includes a 'Copy data' activity. The 'Activities' pane on the left lists various activities, and the 'Copy data' activity is selected. The 'Output' tab at the bottom shows the pipeline run ID: 'b04554fc-870a-4b2f-bcf2-01eb8883a89f'. The pipeline status is 'Succeeded', and the run started on 7/26/2024 at 6:04:37 PM. The table below shows the activity details.

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	Use
Copydata1	Succeeded	Copy data	7/26/2024, 6:04:37 PM	13s	AutoResolveIntegrator	

- 35 After the pipeline is run, check container2 in the Azure Cosmos DB database.



License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).