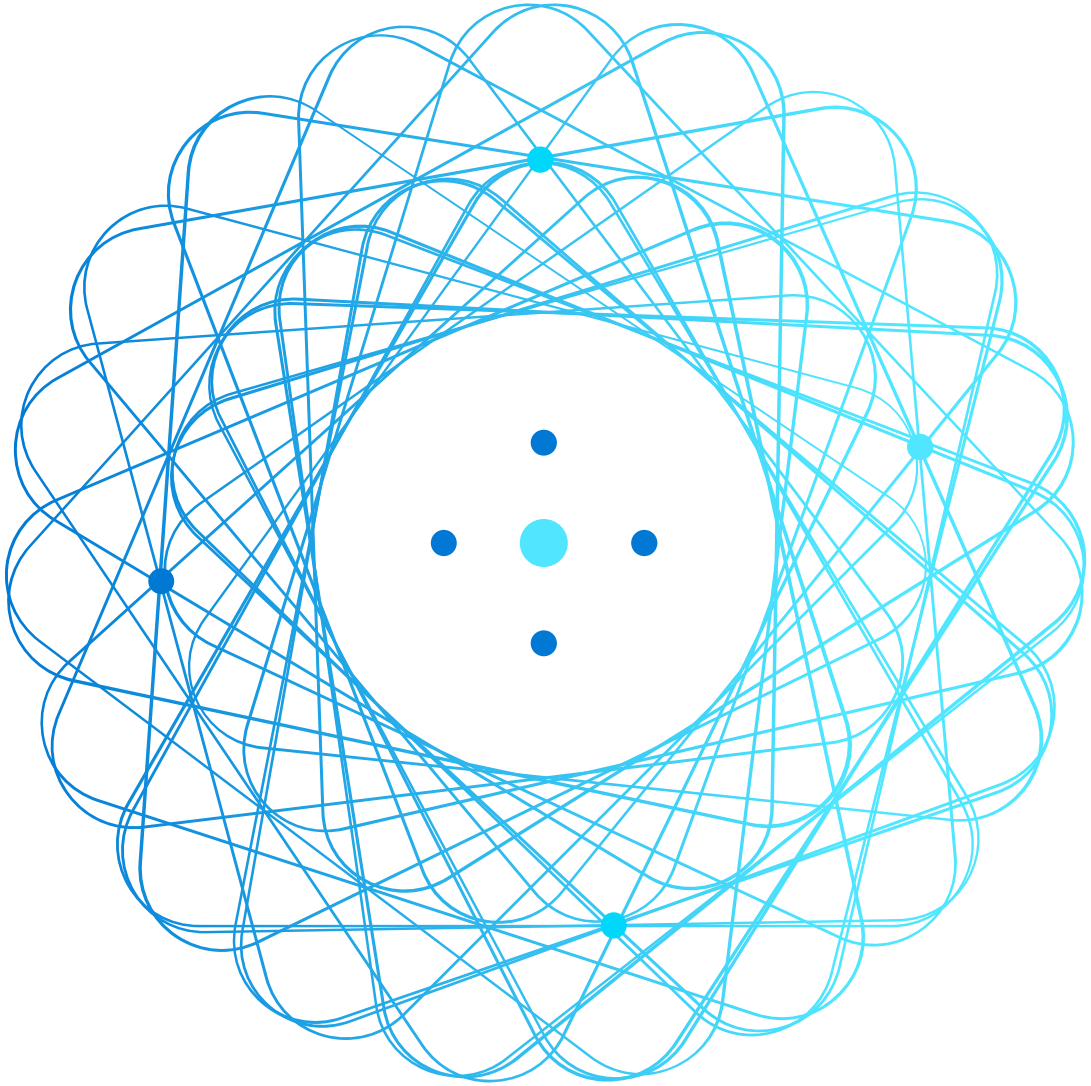
Microsoft Azure

Work with hybrid transactional and analytical processing (HTAP) Solutions using Azure Synapse Analytics



Agenda

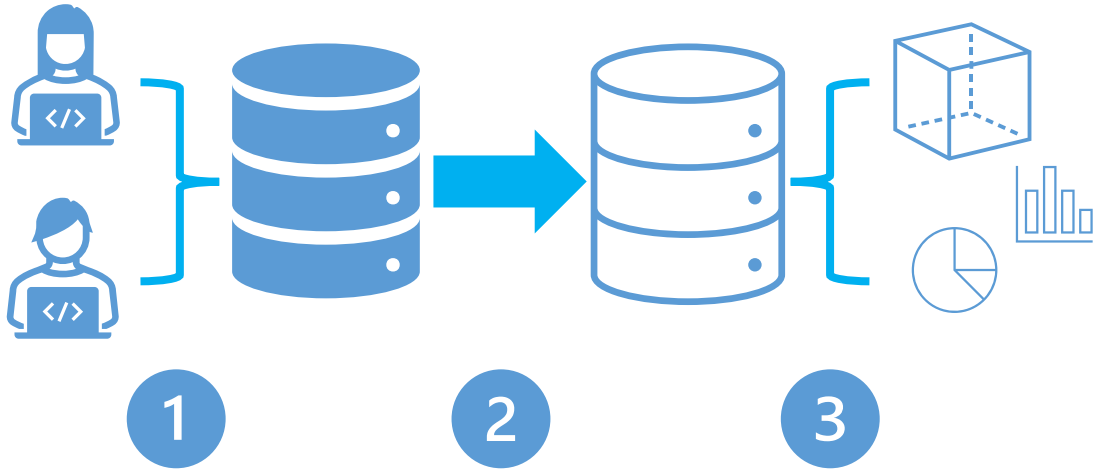
Plan hybrid transactional and analytical processing

Implement Azure Synapse Link with Azure Cosmos DB

Implement Azure Synapse Link for SQL

Plan hybrid transactional and analytical processing

Hybrid transactional and analytical processing patterns



1. A business application processes user input and stores data in a transactional database that is optimized for a mix of data reads and writes
2. The application data is automatically replicated to an analytical store with low latency
3. The analytical store supports data modeling, analytics, and reporting without impacting the transactional system

Azure Synapse Link

Azure Cosmos DB

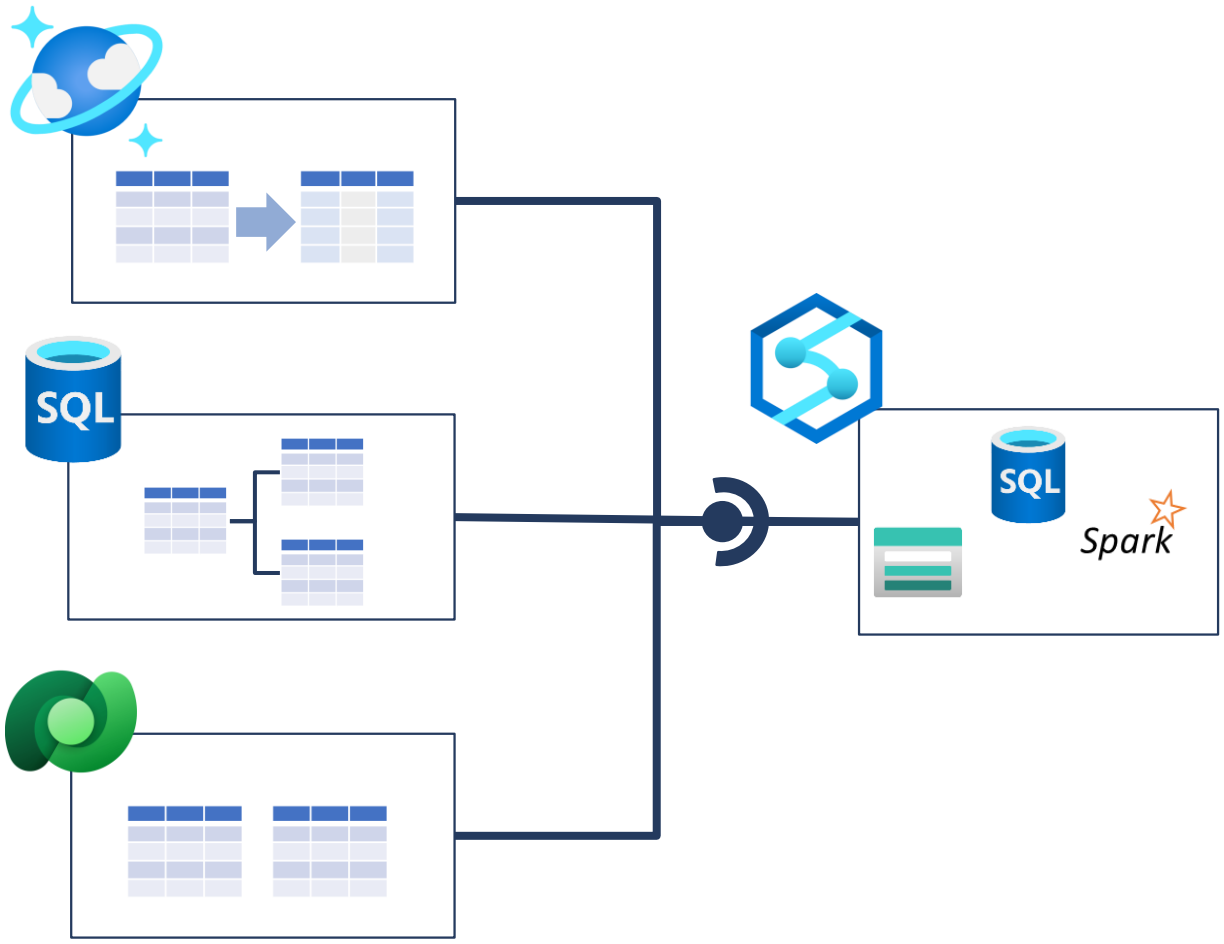
- Transactional data stored in Azure Cosmos DB container
- Data is synced to an analytical store in the container, and queried via linked service in Azure Synapse Analytics

Azure SQL

- Transactional data stored in Azure SQL Database or SQL Server
- Transactions are sync'd to a dedicated SQL pool in Azure Synapse Analytics

Microsoft Dataverse

- Transactional data stored in Dataverse tables
- Data is sync'd to data lake and queries via linked service in Azure Synapse Analytics



Knowledge check

Which of the following descriptions matches a hybrid transactional/analytical processing (HTAP) architecture?

- ❑ Applications store data in a transactional data store, which is also used to support analytical queries for reporting
- ☑ Application transactional data stores are synchronized to a separate analytical store for reporting and analysis
- ❑ Applications store transactional data in an analytical data store that is optimized for reporting and analysis

You want to use Azure Synapse Analytics to analyze operational data stored in a Cosmos DB for NoSQL container. Which Azure Synapse Link service should you use?

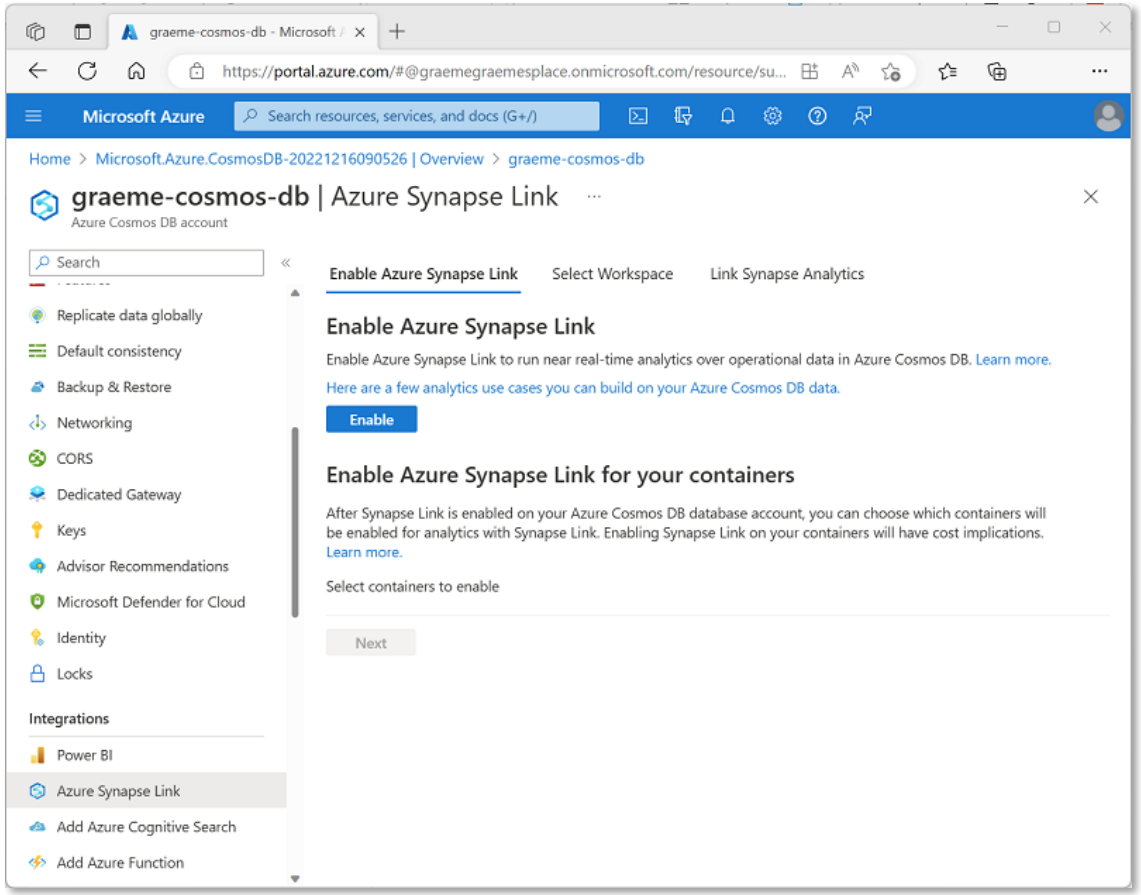
- ❑ Azure Synapse Link for SQL
- ❑ Azure Synapse Link for Dataverse
- ☑ Azure Synapse Link for Azure Cosmos DB

Implement Azure Synapse Link with Azure Cosmos DB

Enable Azure Synapse Link in Azure Cosmos DB

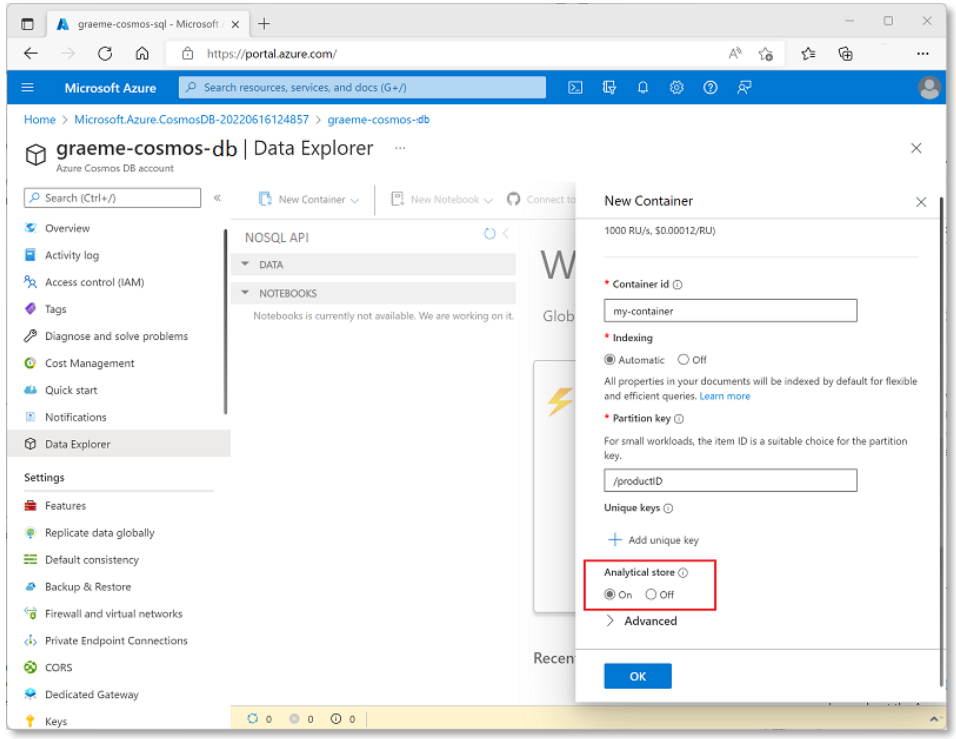
1. Create an Azure Cosmos DB account that supports Synapse Link:
   - Azure Cosmos DB for NoSQL
   - Azure Cosmos DB for MongoDB
   - Azure Cosmos DB for Apache Gremlin
2. Enable Azure Synapse Link
   - Azure portal
   - Azure PowerShell
   - Azure CLI

After enabling Azure Synapse Link, you can't disable it!

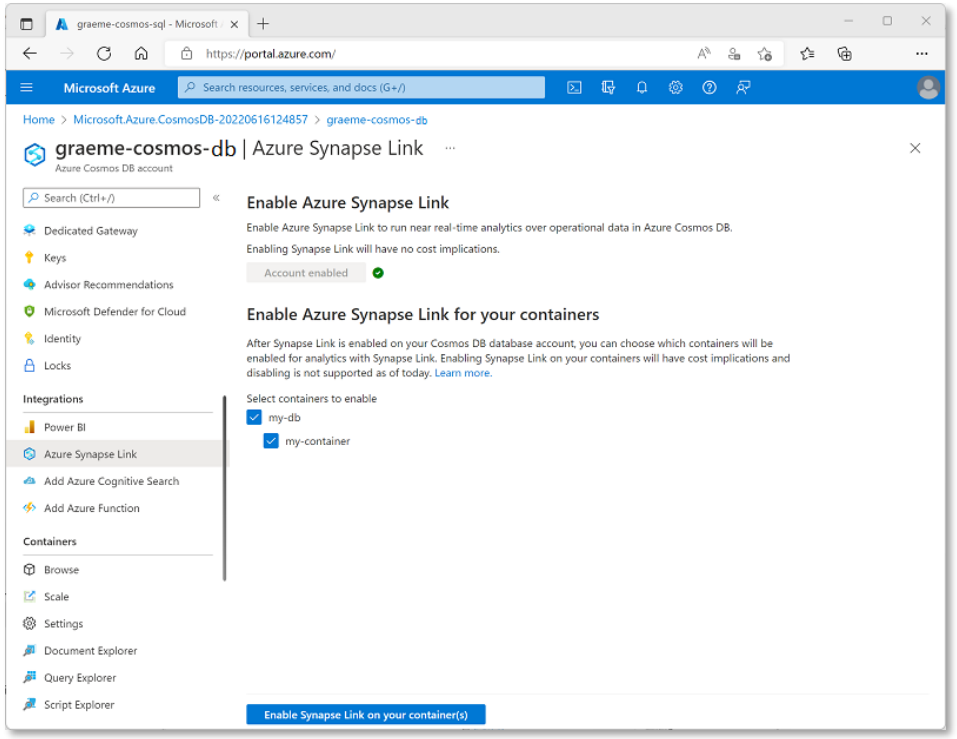


Create an analytical store enabled container

Create a new container with Analytical store enabled



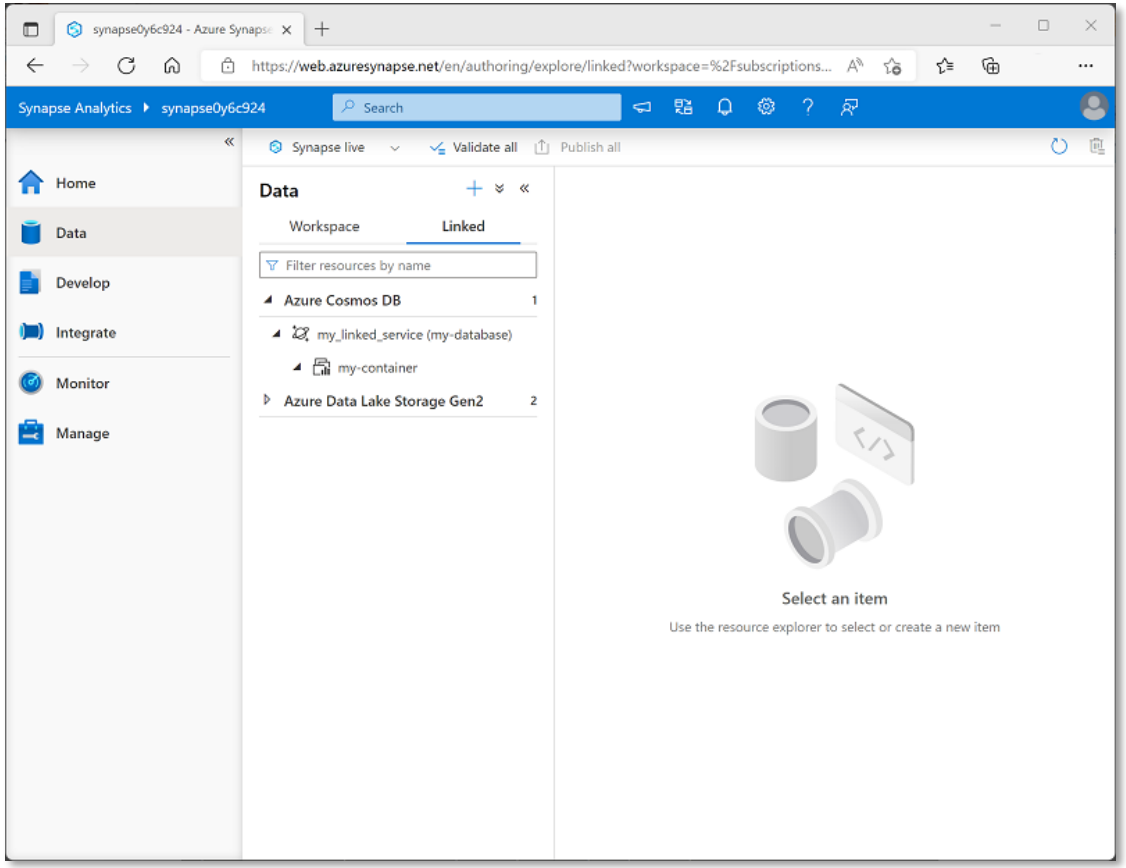
Enable Analytical store in an existing container



Create a linked service in Azure Synapse Analytics

Create an Azure Cosmos DB linked service with one of the following authentication types:

- Account key
- Service Principal
- System Assigned Managed Identity
- User Managed Identity



Query Cosmos DB

Spark pool

```
df = spark.read
     .format("cosmos.olap")\
     .option("spark.synapse.linkedService", "my_linked_service")\
     .option("spark.cosmos.container", "my-container")\
     .load()

display(df.limit(10))
```

SQL pool

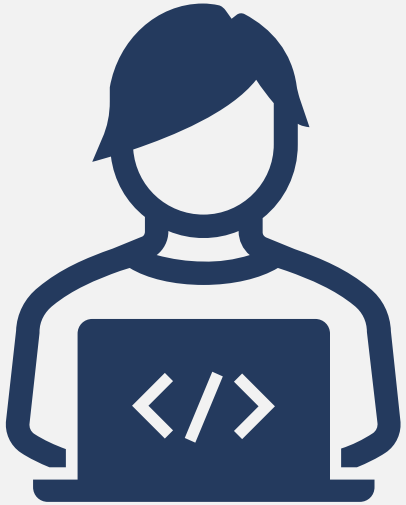
```
CREATE CREDENTIAL my_credential
 WITH IDENTITY = 'SHARED ACCESS SIGNATURE',
 SECRET = 'abcd1234....==';

SELECT *
FROM OPENROWSET(PROVIDER = 'CosmosDB',
                CONNECTION = 'Account=my-cosmos-db;Database=my-db',
                OBJECT = 'my-container',
                SERVER_CREDENTIAL = 'my_credential'
) AS cosmos_data
```

Exercise: Implement Azure Synapse Link for Cosmos DB

Use the hosted lab environment provided, or view the lab instructions at the link below:

https://aka.ms/mslearn-synapse-cosmos



Knowledge check

You have an Azure Cosmos DB for NoSQL account and an Azure Synapse Analytics workspace. What must you do first to enable HTAP integration with Azure Synapse Analytics?

- ❑ Configure global replication in Azure Cosmos DB
- ❑ Create a dedicated SQL pool in Azure Synapse Analytics
- ☑ Enable Azure Synapse Link in Azure Cosmos DB

You have an existing container in a Cosmos DB core (SQL) database. What must you do to enable analytical queries over Azure Synapse Link from Azure Synapse Analytics?

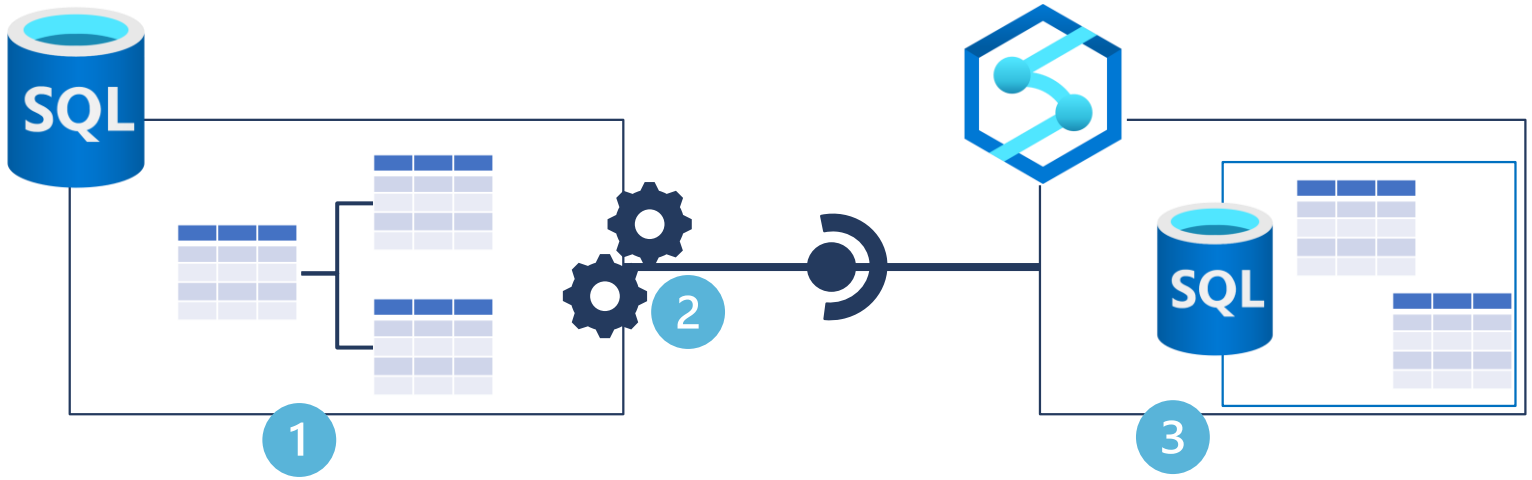
- ❑ Delete and recreate the container
- ☑ Enable Analytical Store support in the container
- ❑ Add an item to the container

You plan to use a Spark pool in Azure Synapse Analytics to query an existing analytical store in Azure Cosmos DB. What must you do?

- ☑ Create a linked service for the Azure Cosmos DB database where the analytical store enabled container is defined
- ❑ Disable automatic pausing for the Spark pool in Azure Synapse Analytics
- ❑ Install the Azure Cosmos DB SDK for Python package in the Spark pool

Implement Azure Synapse Link for SQL

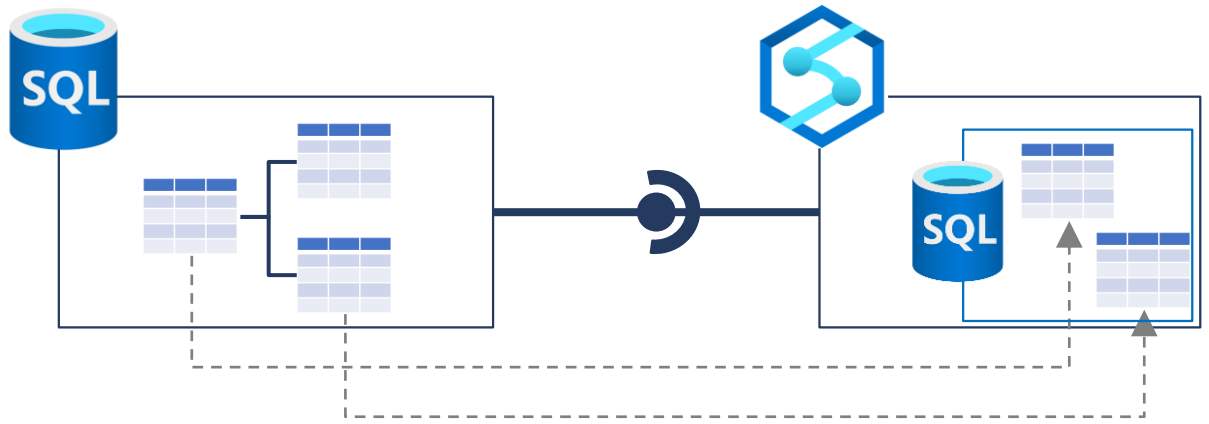
Azure Synapse Link for SQL



1. Transactional is data stored in the *source* database (Azure SQL Database or SQL Server 2022)
   - Azure SQL Managed Instance is not supported by Azure Synapse Link for SQL
2. The *change feed* service captures each transaction and forwards to the *target* database
3. Transactions are replayed in the *target* database (dedicated SQL pool in Azure Synapse Analytics)

Configure Azure Synapse Link for Azure SQL Database

1. Configure Azure SQL Database:
   - Select *System-assigned managed identity*
   - Set firewall rules
   - Create an external user and assign *db_owner* role
2. Configure the target database
   - Create a dedicated SQL pool
   - Create custom schema in the SQL pool database as required
3. Create a link connection in Azure Synapse Analytics:
   - Select or create a linked service
   - Select tables to synchronize
   - Configure mappings



After initialization, transactions in the source tables are automatically replayed in the target tables

Configure Azure Synapse Link for SQL Server 2022

1. Create landing zone storage
   - Use a new Azure Data Lake Store Gen2 container
   - You can't use the default data lake store for your Azure Synapse Analytics workspace
2. Configure SQL Server
   - Create a master key in the source database
3. Configure the target database
   - Create a dedicated SQL pool
   - Create a master key in the SQL pool database
4. Create a link connection in Azure Synapse Analytics:
   - Create linked services for SQL Server and the landing zone storage account
   - Select tables to synchronize
   - Specify file paths in the landing zone

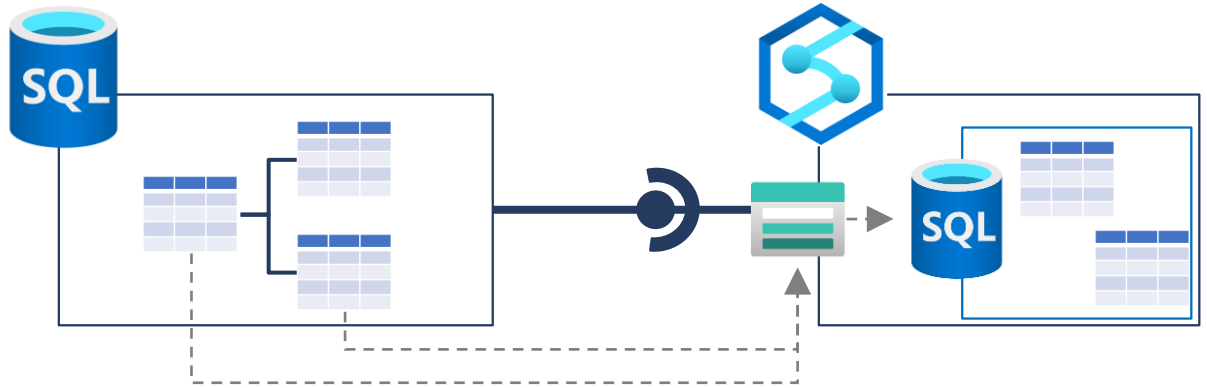


Table are initialized by copying and bulk loading Parquet files

Updates are applied by copying and ingesting CSV files

Demo: Implement Azure Synapse Link for SQL

You can try this for yourself later by following the instructions at the link below:

https://aka.ms/mslearn-synapse-link-sql

Knowledge check

From which of the following data sources can you use Azure Synapse Link for SQL to replicate data to Azure Synapse Analytics?

- ❑ Azure Cosmos DB
- ☑ SQL Server 2022
- ❑ Azure SQL Managed Instance

What must you create in your Azure Synapse Analytics workspace as a target database for Azure Synapse Link for Azure SQL Database?

- ❑ A serverless SQL pool
- ❑ An Apache Spark pool
- ☑ A dedicated SQL pool

You plan to use Azure Synapse Link for SQL to replicate tables from SQL Server 2022 to Azure Synapse Analytics. What additional Azure resource must you create?

- ☑ An Azure Storage account with an Azure Data Lake Storage Gen2 container
- ❑ An Azure Key Vault containing the SQL Server admin password
- ❑ An Azure Application Insights resource

Further reading

Work with Hybrid Transactional and Analytical Processing Solutions using Azure Synapse Analytics
https://aka.ms/mslearn-htap-synapse