https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/tree/master/Instructions

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/tree/master/Instructions/Labs

All DP-900 Labs

DP-900 trainer lab mapping

This course, DP-900: Microsoft Azure Data Fundamentals, has a set of labs designated for instructor-led training.

DP-900 module mapping to labs

Module	**Lab**	**Markdown file**
02-Explore relational data in Azure	Explore Azure SQL Database	dp900-01-sql-lab.md
03-Explore non-relational data in Azure	Explore Azure Storage	dp900-02-storage-lab.md
03-Explore non-relational data in Azure	Explore Azure Cosmos DB	dp900-03-cosmos-lab.md
04-Explore data analytics in Azure	Explore Azure Synapse Analytics	dp900-04-synapse-lab.md
04-Explore data analytics in Azure	Explore Azure Stream Analytics	dp900-05-stream-lab.md
04-Explore data analytics in Azure	Visualize data with Power BI	dp900-pbi-06-lab.md

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/blob/master/Instructions/01-Provision-Azure-relational-database-services.md

Instructions

As part of your role at Contoso as a data engineer, you've been asked to create and **configure SQL Server, PostgreSQL, and MySQL servers for Azure**.

In this lab you'll create and configure one of the following servers for Azure: SQL Server, PostgreSQL, or MySQL.

1. Go to the Microsoft Learn exercise at https://aka.ms/dp900lab01 and complete the unit in the browser:

Exercise: Provision Azure relational database services

Completed100 XP

- 15 minutes

Now it's your opportunity to explore relational database services in Azure.

This exercise can be completed using a Microsoft Learn *sandbox*, which provides a temporary Azure subscription. To activate the sandbox subscription, you must sign in using a Microsoft account. The sandbox subscription will be automatically deleted when you complete this module. After the sandbox has been activated, sign into the Azure portal using the credentials for your sandbox subscription. Ensure you're working in the **Microsoft Learn Sandbox** directory - indicated at the top right of the portal under your user ID. If not, select the user icon and switch directory.

The free sandbox allows you to create resources in a subset of the Azure global regions. Select a region from the following list when you create resources:

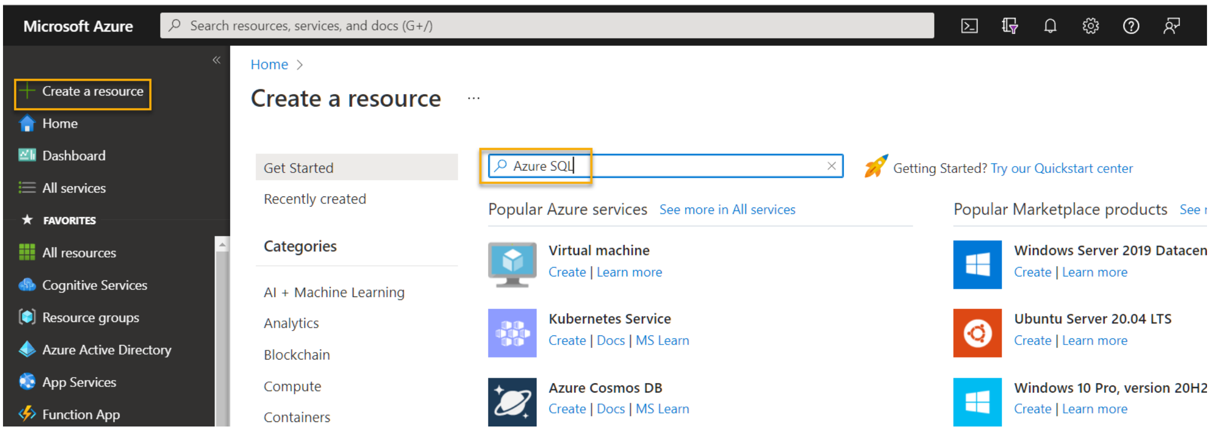
- West US 2
- South Central US
- Central US
- East US
- West Europe
- Southeast Asia
- Japan East
- Brazil South
- Australia Southeast
- Central India

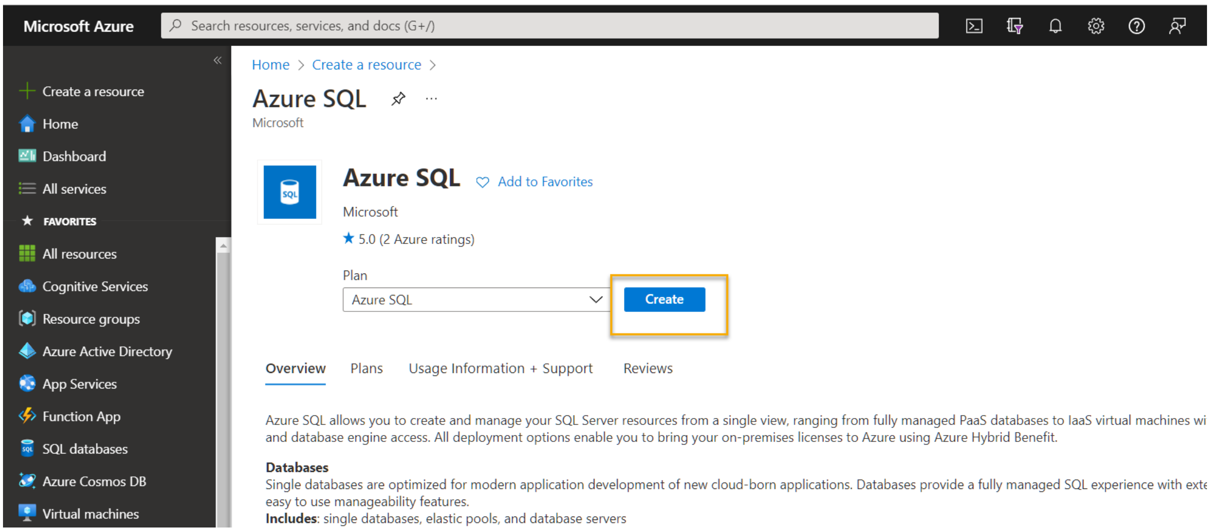
Lab 01

Provision an Azure SQL Database resource

In this exercise, you'll provision and test an Azure SQL Database resource.

1. In the Azure portal, select **+ Create a resource** from the upper left-hand corner and search for *Azure SQL*. Then in the resulting **Azure SQL** page, select **Create**.

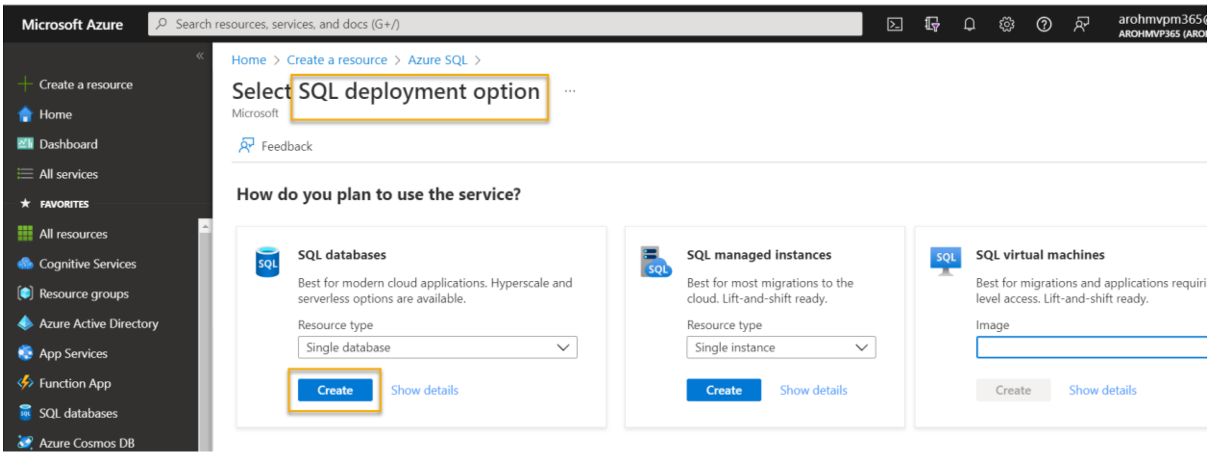




Azure SQL allows you to create and manage your SQL Server resources from a single view, ranging from fully managed PaaS databases to IaaS virtual machines wi and database engine access. All deployment options enable you to bring your on-premises licenses to Azure using Azure Hybrid Benefit.

Databases
Single databases are optimized for modern application development of new cloud-born applications. Databases provide a fully managed SQL experience with exte easy to use manageability features.
Includes: single databases, elastic pools, and database servers

2. Review the Azure SQL options that are available, and then in the **SQL databases** tile, ensure **Single database** is selected and select **Create**.



Best for modern cloud applications. Hyperscale and serverless options are available.

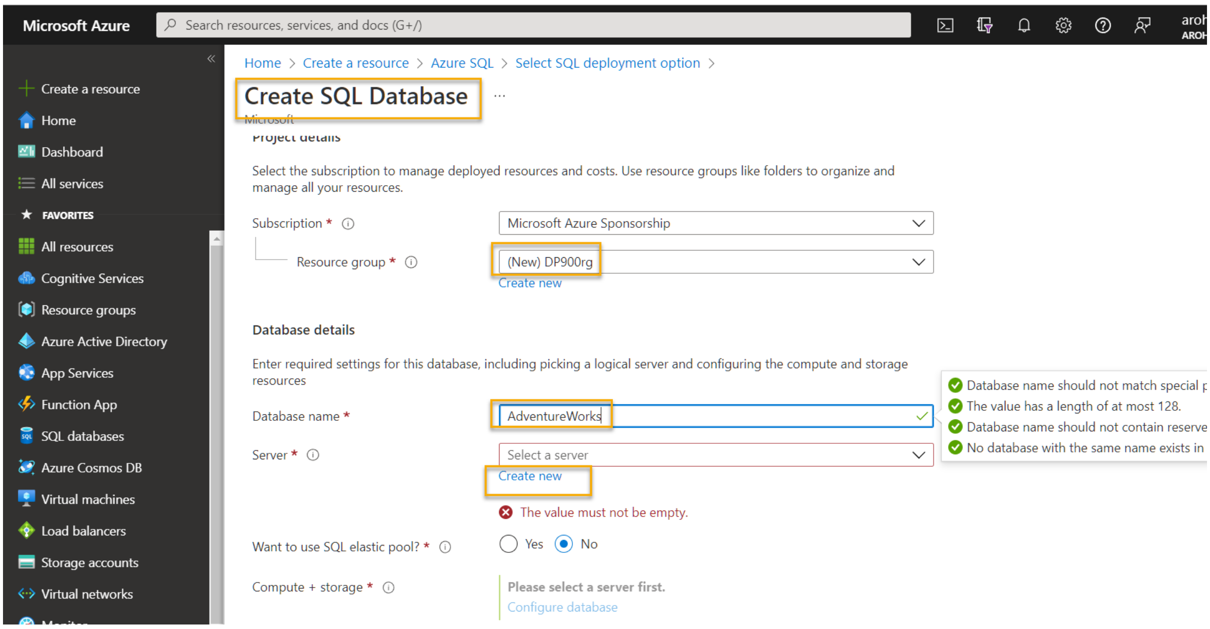
Best for most migrations to the cloud. Lift-and-shift ready.

Best for migrations and applications requiri level access. Lift-and-shift ready.

3. Enter the following values on the **Create SQL Database** page:

- **Subscription**: If you're using a *sandbox*, select *Concierge Subscription*. Otherwise, select your Azure subscription.

- **Resource group**: If you're using a *sandbox* subscription, select the existing resource group (which will have a name like *learn-xxxx...*). Otherwise, create a new resource group with a name of your choice.
- **Database name**: *AdventureWorks*



Home > Create a resource > Azure SQL > Select SQL deployment option >

Create SQL Database ...

Microsoft

Project details

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

Subscription * : Microsoft Azure Sponsorship

Resource group * : (New) DP900rg
Create new

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name * : AdventureWorks

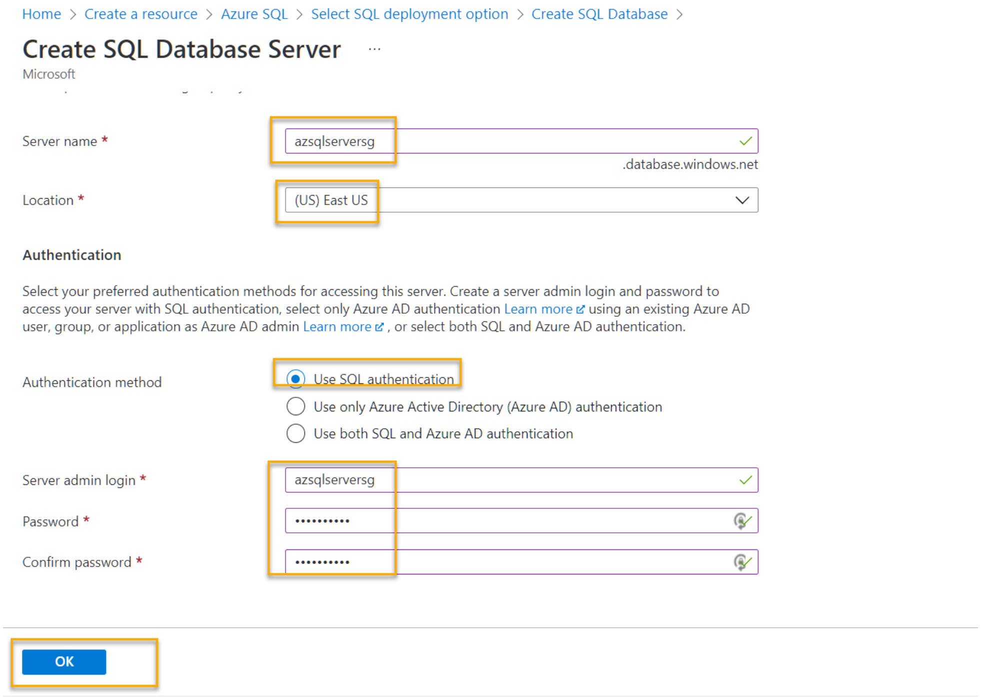
✓ Database name should not match special p
✓ The value has a length of at most 128.
✓ Database name should not contain reserve
✓ No database with the same name exists in

Server * : Select a server
Create new
✖ The value must not be empty.

Want to use SQL elastic pool? * : ○ Yes ● No

Compute + storage * : Please select a server first.
Configure database

- **Server**: Select **Create new** and create a new server with a unique name in any available location. Use **SQL authentication** and specify your name as the server admin login and a suitably complex password (remember the password - you'll need it later!)


Home > Create a resource > Azure SQL > Select SQL deployment option > Create SQL Database >

Create SQL Database Server ...

Microsoft

Server name * : azsqlserversg ✓
.database.windows.net

Location * : (US) East US

Authentication

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Azure AD authentication Learn more using an existing Azure AD user, group, or application as Azure AD admin Learn more , or select both SQL and Azure AD authentication.

Authentication method
● Use SQL authentication
○ Use only Azure Active Directory (Azure AD) authentication
○ Use both SQL and Azure AD authentication

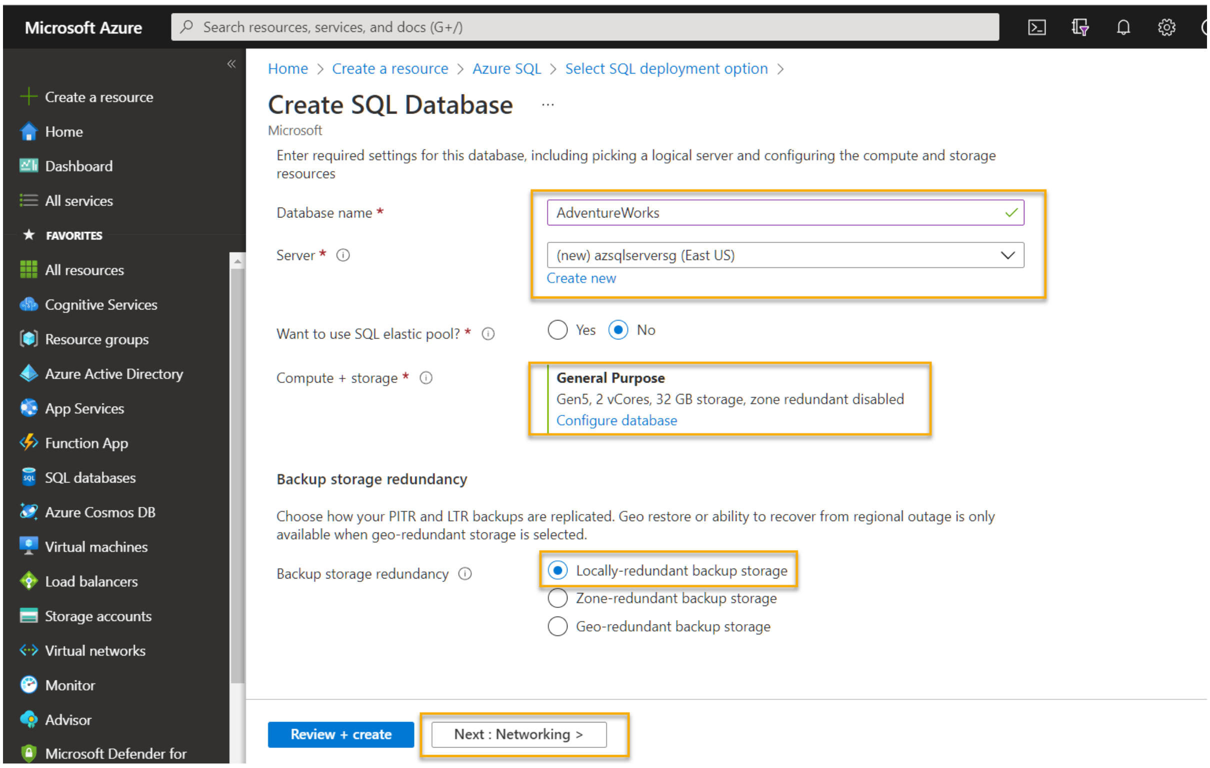
Server admin login * : azsqlserversg ✓

Password * : ••••••••••

Confirm password * : ••••••••••

- **Want to use SQL elastic pool?**: *No*

- **Compute + storage**: Leave unchanged
- **Backup storage redundancy**: *Locally-redundant backup storage*



Create SQL Database

Microsoft

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name * AdventureWorks

Server * (new) azsqlserversg (East US)

Create new

Want to use SQL elastic pool? * ○ Yes ● No

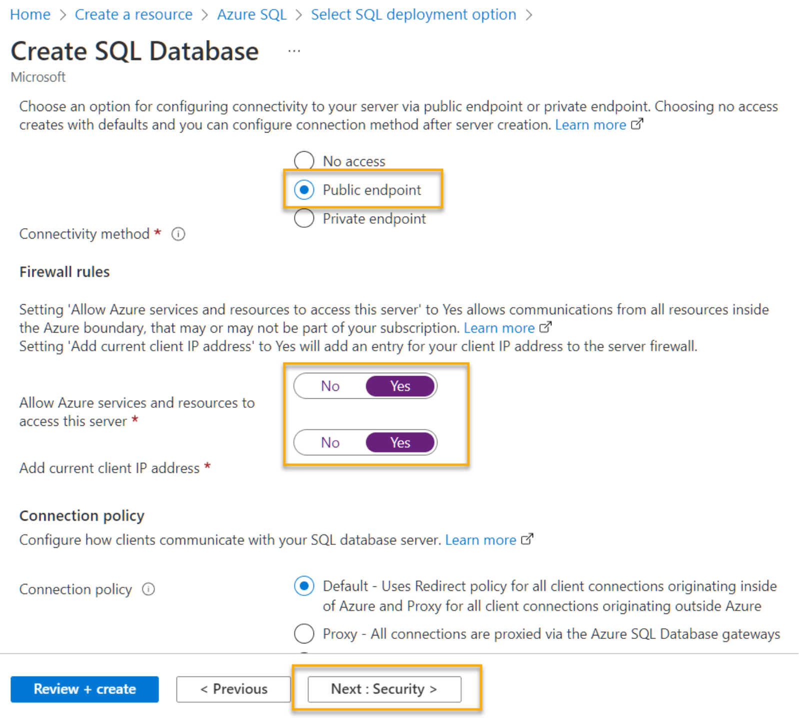
Compute + storage * **General Purpose**
Gen5, 2 vCores, 32 GB storage, zone redundant disabled
Configure database

Backup storage redundancy

Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only available when geo-redundant storage is selected.

Backup storage redundancy
● Locally-redundant backup storage
○ Zone-redundant backup storage
○ Geo-redundant backup storage

4. On the **Create SQL Database** page, select **Next :Networking** >, and on the **Networking page,** in the Network connectivity section, select Public endpoint. Then select **Yes** for both options in the **Firewall rules** section to allow access to your database server from Azure services and your current client IP address.


Home > Create a resource > Azure SQL > Select SQL deployment option >

Create SQL Database ···

Microsoft

Choose an option for configuring connectivity to your server via public endpoint or private endpoint. Choosing no access creates with defaults and you can configure connection method after server creation. Learn more

Connectivity method * ○ No access ● Public endpoint ○ Private endpoint

Firewall rules

Setting 'Allow Azure services and resources to access this server' to Yes allows communications from all resources inside the Azure boundary, that may or may not be part of your subscription. Learn more
Setting 'Add current client IP address' to Yes will add an entry for your client IP address to the server firewall.

Allow Azure services and resources to access this server * No [✓ Yes]

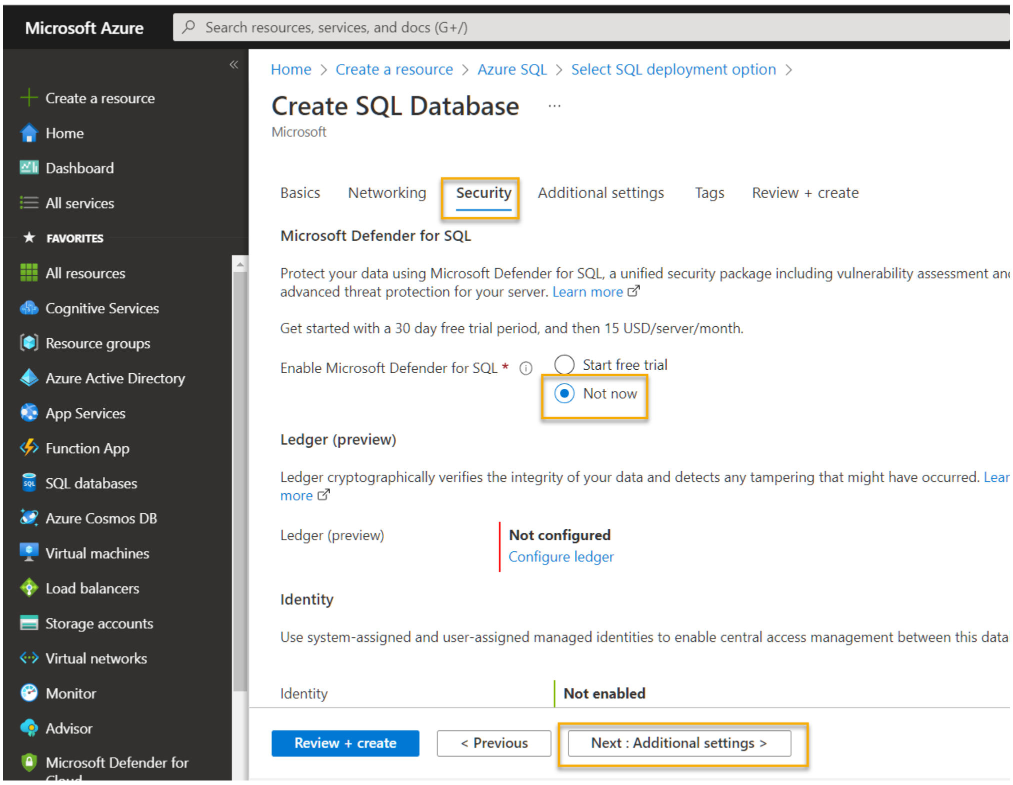
Add current client IP address * No [✓ Yes]

Connection policy

Configure how clients communicate with your SQL database server. Learn more

Connection policy
● Default - Uses Redirect policy for all client connections originating inside of Azure and Proxy for all client connections originating outside Azure
○ Proxy - All connections are proxied via the Azure SQL Database gateways

5. Select **Next: Security >** and set the **Enable Microsoft Defender for SQL** option to **Not now**.



Home > Create a resource > Azure SQL > Select SQL deployment option >

Create SQL Database ···

Microsoft

Microsoft Defender for SQL

Protect your data using Microsoft Defender for SQL, a unified security package including vulnerability assessment an advanced threat protection for your server. Learn more

Get started with a 30 day free trial period, and then 15 USD/server/month.

Enable Microsoft Defender for SQL * ○ Start free trial ● Not now

Ledger (preview)

Ledger cryptographically verifies the integrity of your data and detects any tampering that might have occurred. Lear more

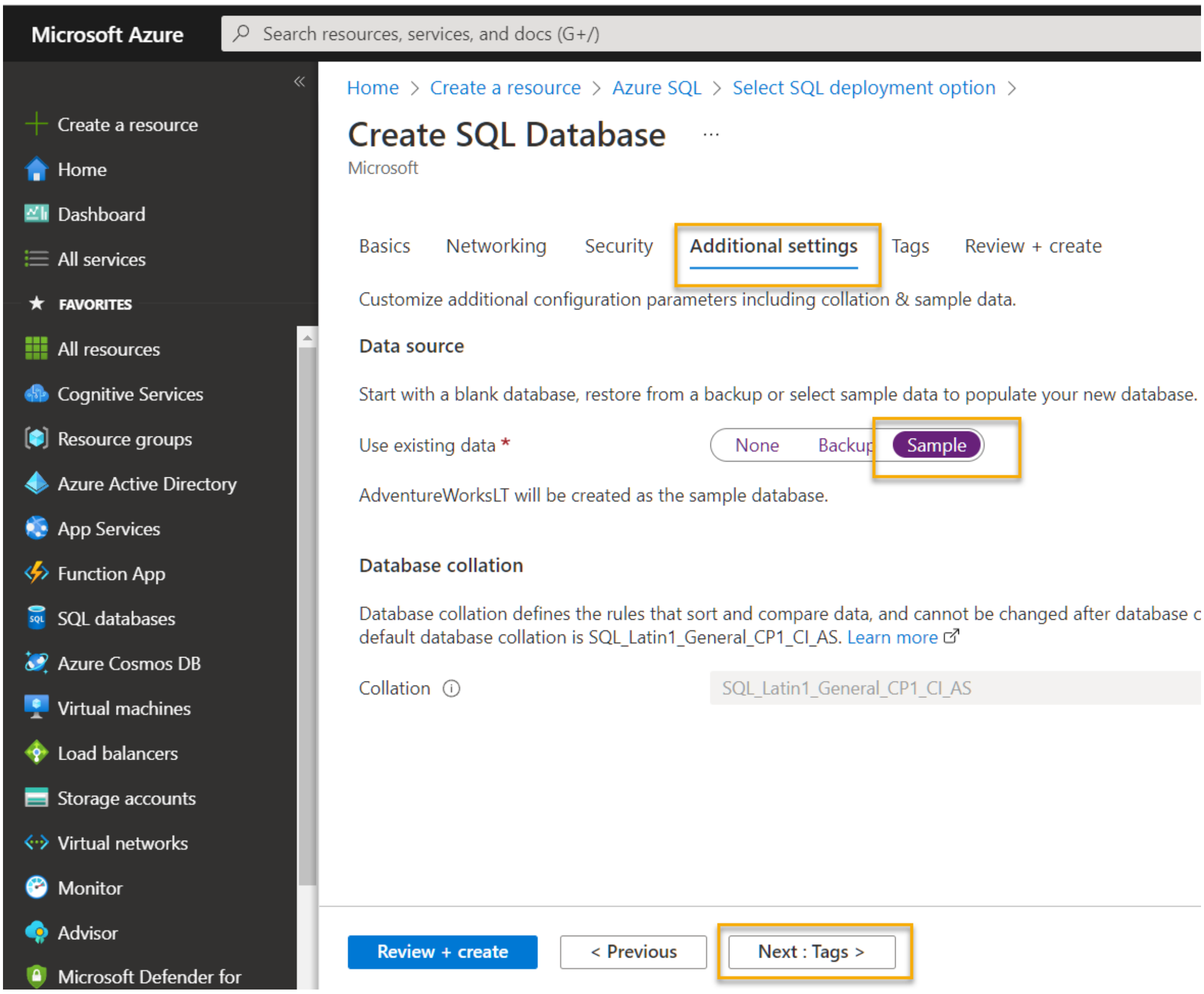
Ledger (preview) **Not configured** Configure ledger

Identity

Use system-assigned and user-assigned managed identities to enable central access management between this data

Identity **Not enabled**

6. Select **Next: Additional Settings >** and on the **Additional settings** tab, set the **Use existing data** option to **Sample** (this will create a sample database that you can explore later)



Create SQL Database

Microsoft

Customize additional configuration parameters including collation & sample data.

Data source

Start with a blank database, restore from a backup or select sample data to populate your new database.

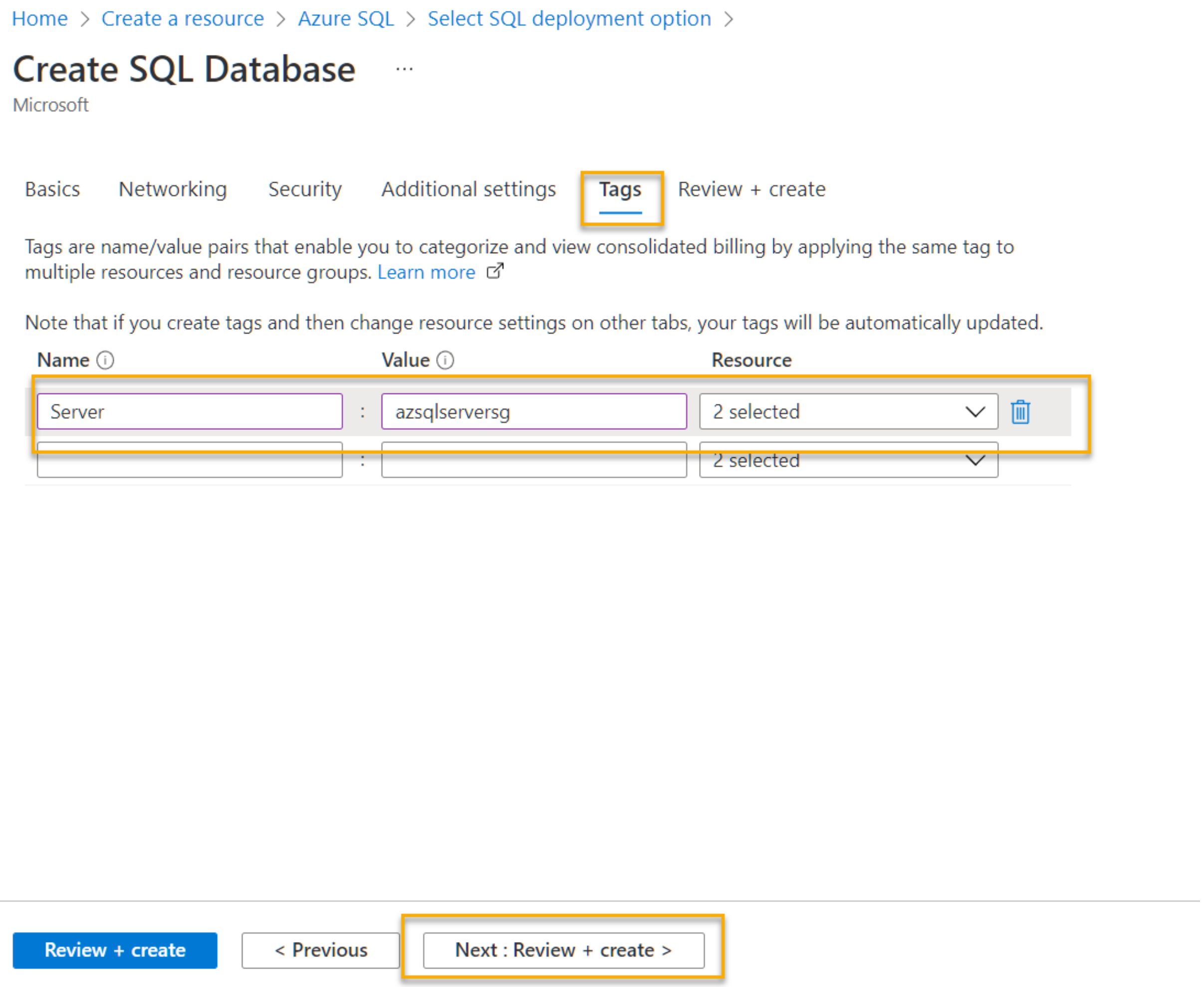
Use existing data * None Backup [✓ Sample]

AdventureWorksLT will be created as the sample database.

Database collation

Database collation defines the rules that sort and compare data, and cannot be changed after database c default database collation is SQL_Latin1_General_CP1_CI_AS. Learn more

Collation ⓘ SQL_Latin1_General_CP1_CI_AS

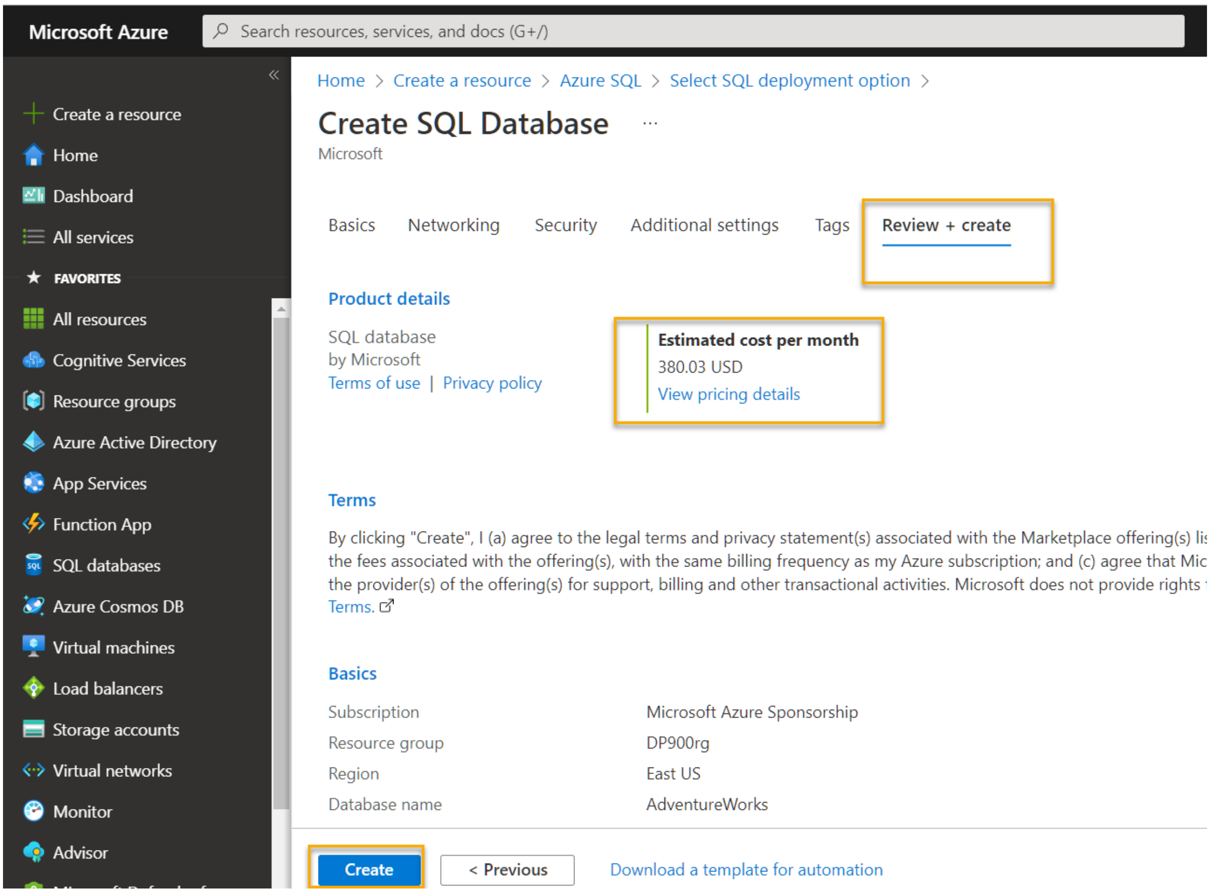


Tags are name/value pairs that enable you to categorize and view consolidated billing by applying the same tag to multiple resources and resource groups. Learn more

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

Name		Value	Resource
Server	:	azsqlserversg	2 selected
	:		2 selected

7. Select **Review + Create**, and then select **Create** to create your Azure SQL database.



Create SQL Database ···

Microsoft

Product details

SQL database
by Microsoft
Terms of use | Privacy policy

Estimated cost per month
380.03 USD
View pricing details

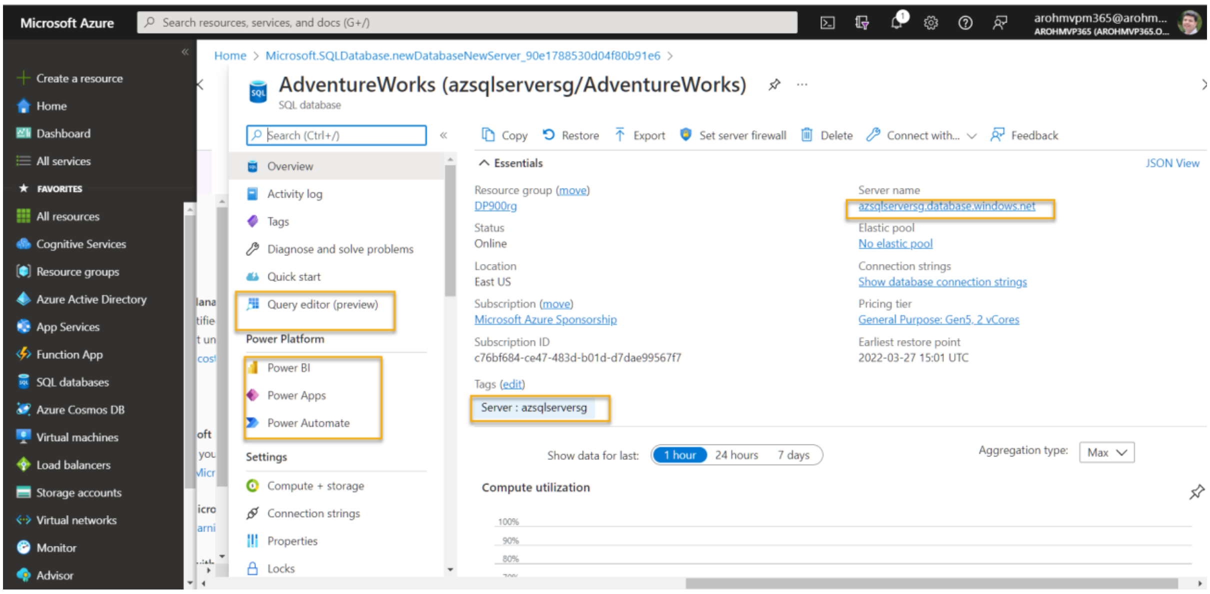
Terms

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) lis
the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Mic
the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights
Terms.

Basics

Subscription	Microsoft Azure Sponsorship
Resource group	DP900rg
Region	East US
Database name	AdventureWorks

8. Wait for deployment to complete. Then go to the resource that was deployed, which should look like this:



AdventureWorks (azsqlserversg/AdventureWorks)

SQL database

Resource group (move)
DP900rg

Status
Online

Location
East US

Subscription (move)
Microsoft Azure Sponsorship

Subscription ID
c76bf684-ce47-483d-b01d-d7dae99567f7

Tags (edit)
Server : azsqlserversg

Server name
azsqlserversg.database.windows.net

Elastic pool
No elastic pool

Connection strings
Show database connection strings

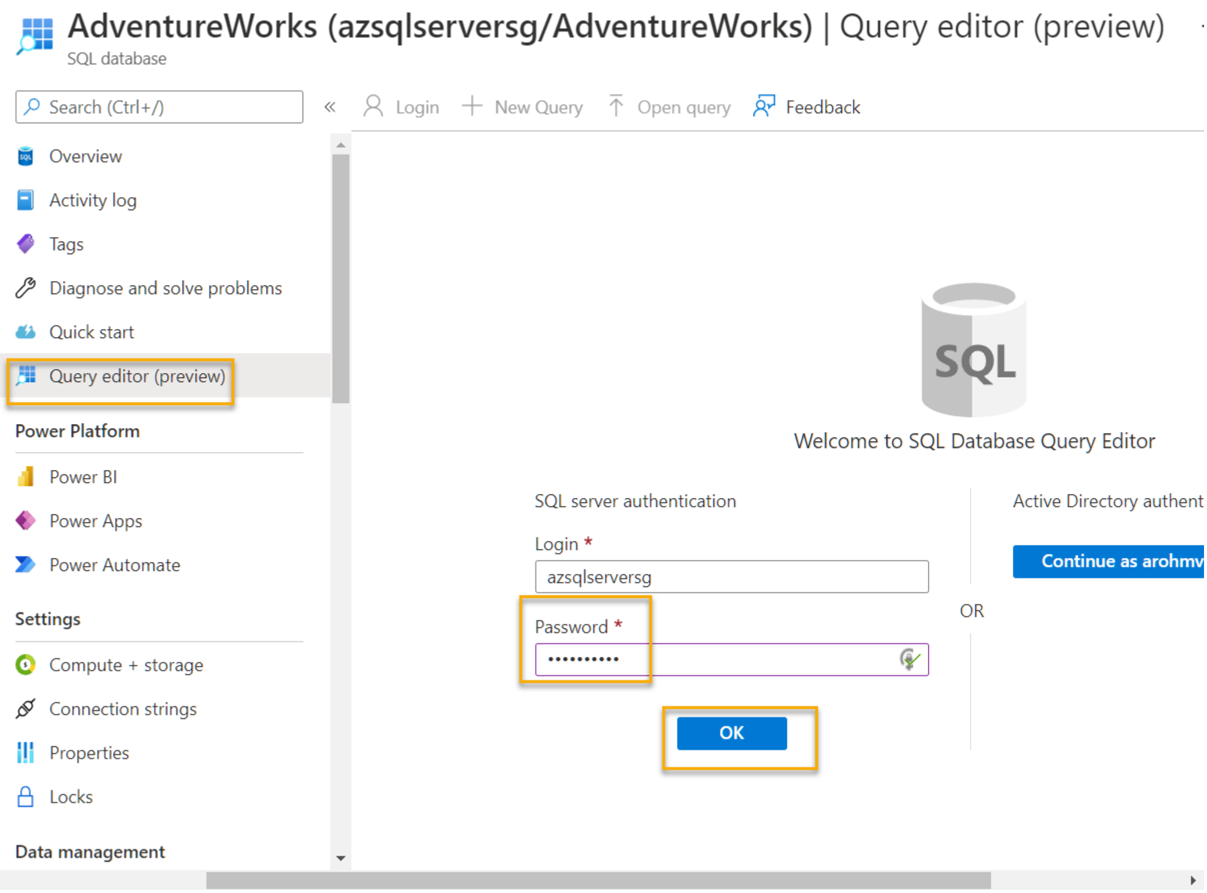
Pricing tier
General Purpose: Gen5, 2 vCores

Earliest restore point
2022-03-27 15:01 UTC

9. In the pane on the left side of the page, select **Query editor (preview)**, and then sign in using the administrator login and password you specified for your server.

*If an error message stating that the client IP address isn't allowed is displayed, select the **Allowlist IP ...** link at the end of the message to allow access and try to sign in again (you previously added you own computer's client IP address to the firewall rules, but the query editor may connect from a different address depending on your network configuration.)*

The query editor looks like this:



AdventureWorks (azsqlserversg/AdventureWorks) | Query editor (preview)

SQL database

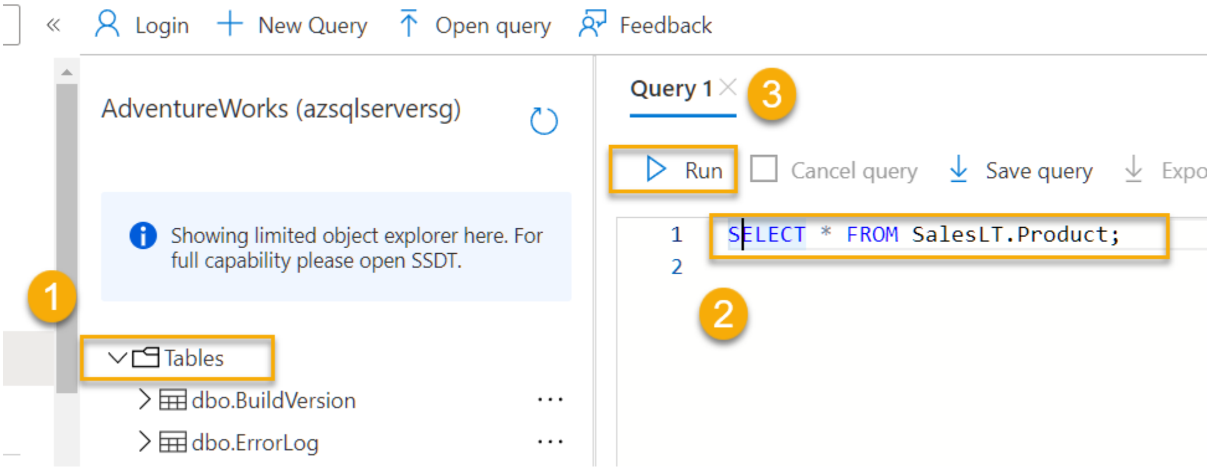
Login: azsqlserversg

Password: ••••••••••

10. Expand the **Tables** folder to see the tables in the database.

11. In the **Query 1** pane, enter the following SQL code:

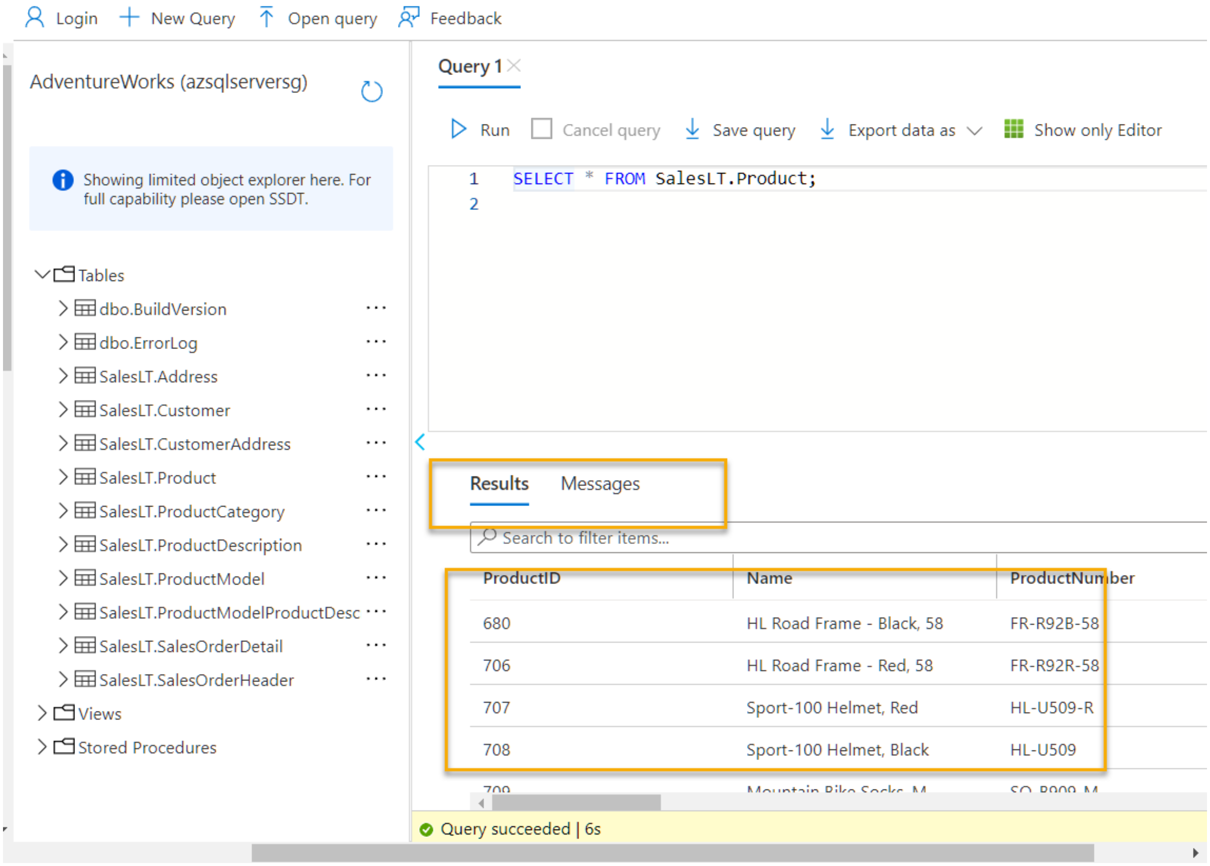
```
SELECT * FROM SalesLT.Product;
```



Showing limited object explorer here. For full capability please open SSDT.

```
SELECT * FROM SalesLT.Product;
```

12. Select ▷ **Run** above the query to run it and view the results, which should include all columns for all rows in the *SalesLT.Product* table as shown here:



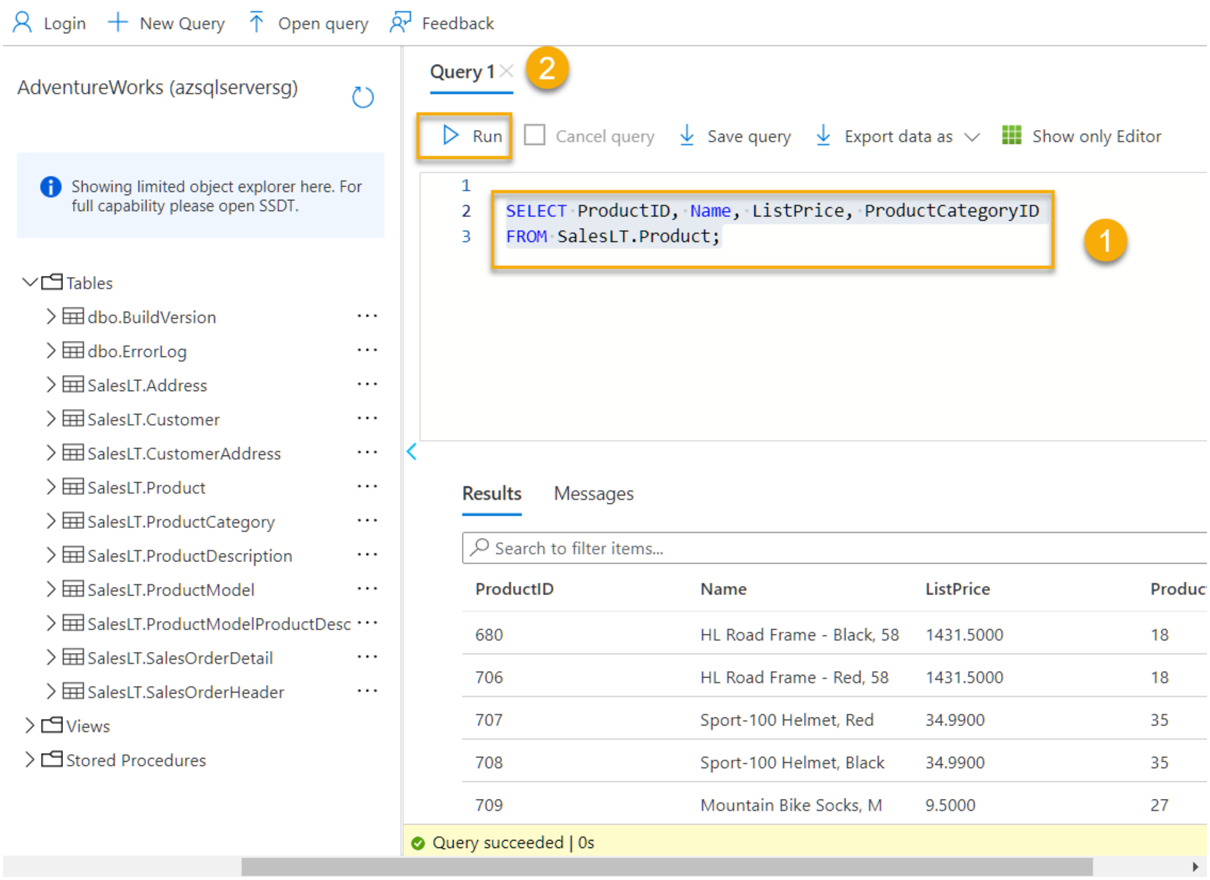
Showing limited object explorer here. For full capability please open SSDT.

```
SELECT * FROM SalesLT.Product;
```

ProductID	Name	ProductNumber
680	HL Road Frame - Black, 58	FR-R92B-58
706	HL Road Frame - Red, 58	FR-R92R-58
707	Sport-100 Helmet, Red	HL-U509-R
708	Sport-100 Helmet, Black	HL-U509
709	Mountain Bike Socks, M	SO-B909-M

Query succeeded | 6s

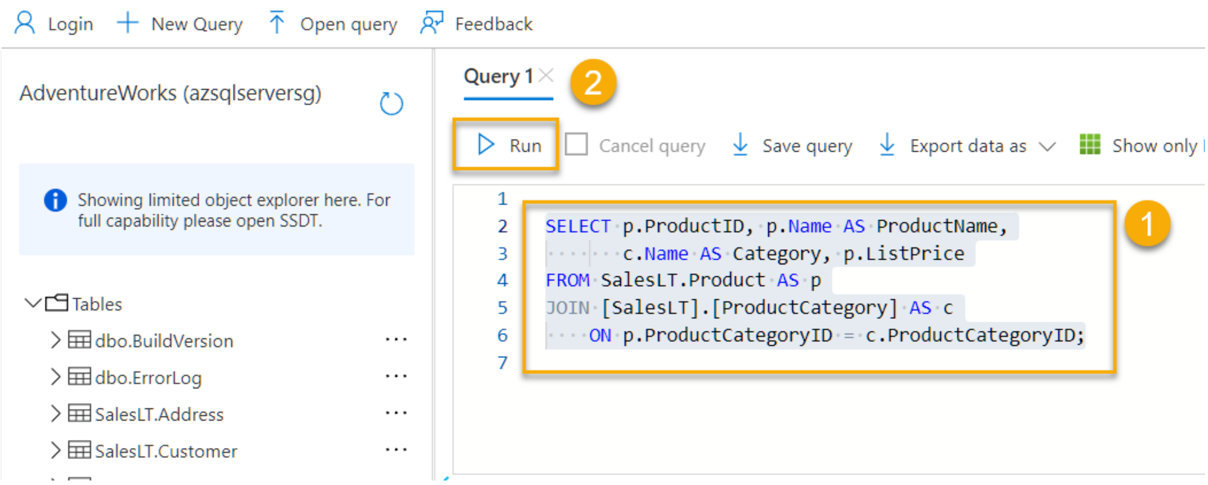
13. Replace the SELECT statement with the following code, and then select ▷ **Run** to run the new query and review the results (which includes only the **ProductID**, **Name**, **ListPrice**, **ProductCategoryID** columns):



```sql
SELECT ProductID, Name, ListPrice, ProductCategoryID
FROM SalesLT.Product;
```

ProductID	Name	ListPrice	Produc
680	HL Road Frame - Black, 58	1431.5000	18
706	HL Road Frame - Red, 58	1431.5000	18
707	Sport-100 Helmet, Red	34.9900	35
708	Sport-100 Helmet, Black	34.9900	35
709	Mountain Bike Socks, M	9.5000	27

Query succeeded | 0s



```sql
SELECT p.ProductID, p.Name AS ProductName,
       c.Name AS Category, p.ListPrice
FROM SalesLT.Product AS p
JOIN [SalesLT].[ProductCategory] AS c
    ON p.ProductCategoryID = c.ProductCategoryID;
```

14. Close the query editor pane, discarding your edits.
15.

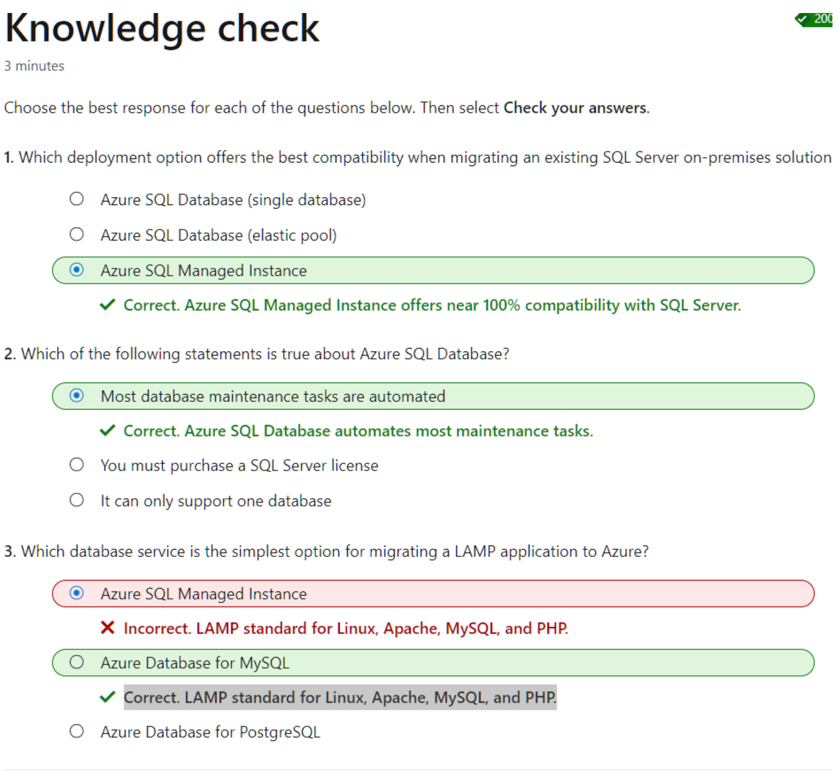
Knowledge Check

1. **Which deployment option** offers the **best compatibility** when migrating an **existing SQL Server on-premises solution**?
    a. **Azure SQL Managed Instance**

       Correct. Azure SQL Managed Instance offers **near 100% compatibility** with SQL Server.

2. Which of the following statements is true about Azure SQL Database?
    a. **Most database maintenance tasks are automated**
       Azure SQL Database automates most maintenance tasks.

3. Which database service is the **simplest option** for migrating a **LAMP** application to Azure?
    a. **Azure Database for MySQL**
       Correct. LAMP standard for Linux, Apache, MySQL, and PHP**.**



Knowledge check

✓ 200

3 minutes

Choose the best response for each of the questions below. Then select **Check your answers**.

1. Which deployment option offers the best compatibility when migrating an existing SQL Server on-premises solution

○ Azure SQL Database (single database)

○ Azure SQL Database (elastic pool)

● Azure SQL Managed Instance

✓ **Correct. Azure SQL Managed Instance offers near 100% compatibility with SQL Server.**

2. Which of the following statements is true about Azure SQL Database?

● Most database maintenance tasks are automated

✓ **Correct. Azure SQL Database automates most maintenance tasks.**

○ You must purchase a SQL Server license

○ It can only support one database

3. Which database service is the simplest option for migrating a LAMP application to Azure?

● Azure SQL Managed Instance

✗ **Incorrect. LAMP standard for Linux, Apache, MySQL, and PHP.**

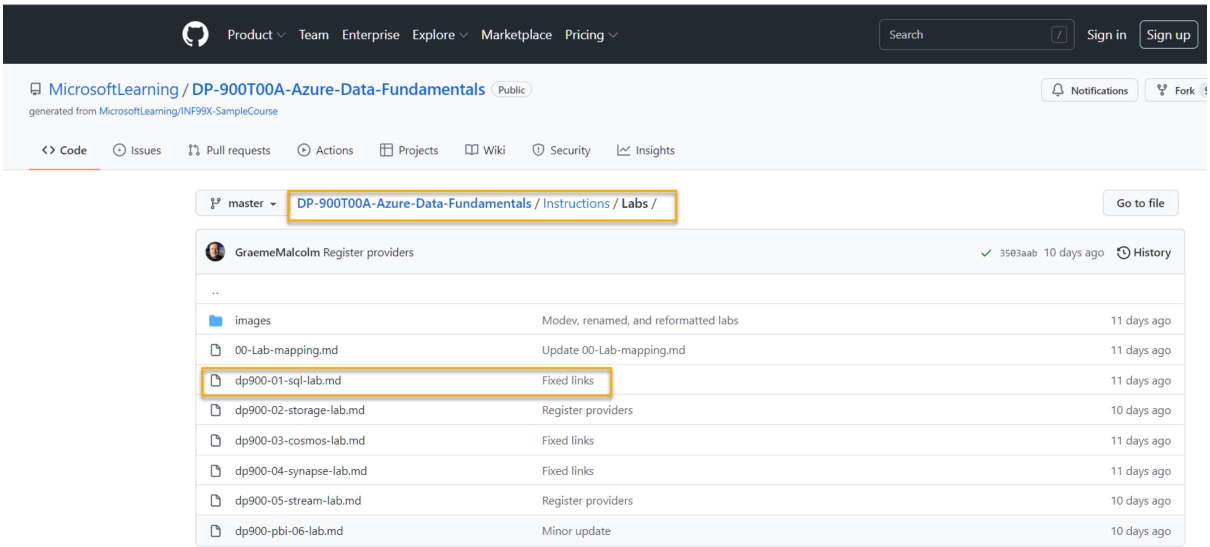
○ Azure Database for MySQL

✓ **Correct. LAMP standard for Linux, Apache, MySQL, and PHP.**

○ Azure Database for PostgreSQL

Labs

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/tree/master/Instructions/Labs



Product
Team
Enterprise
Explore
Marketplace
Pricing
Search
Sign in
Sign up
MicrosoftLearning / DP-900T00A-Azure-Data-Fundamentals Public
generated from MicrosoftLearning/INF99X-SampleCourse
Notifications
Fork
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
master
DP-900T00A-Azure-Data-Fundamentals / Instructions / Labs /
Go to file
GraemeMalcolm Register providers
3503aab 10 days ago
History

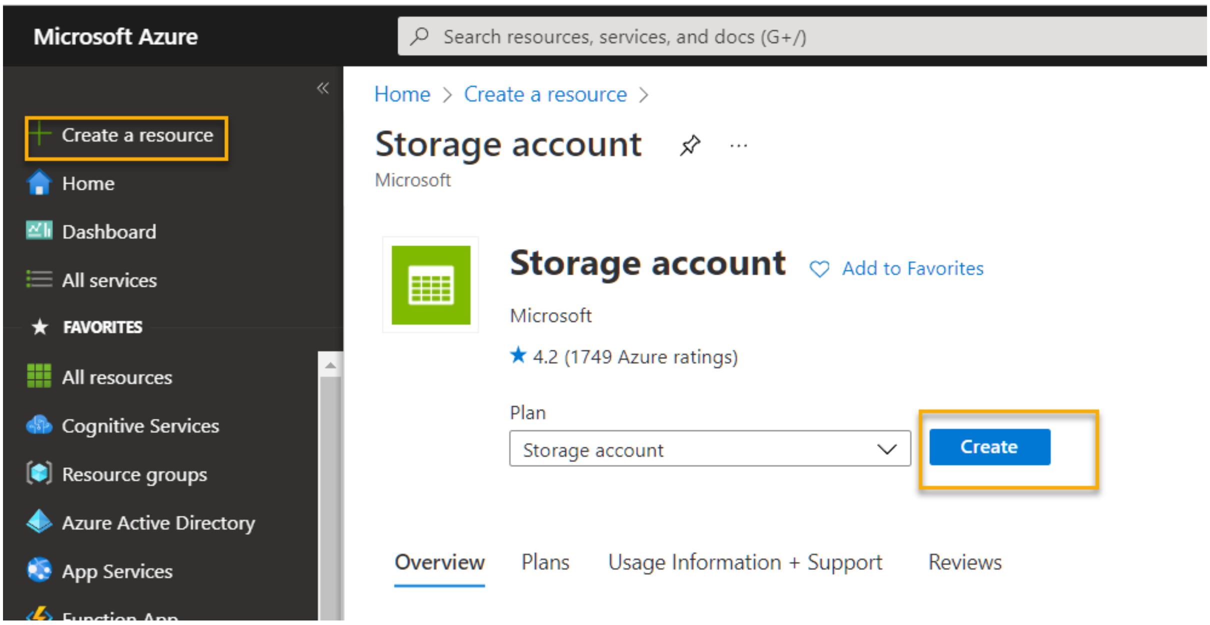
..		
images	Modev, renamed, and reformatted labs	11 days ago
00-Lab-mapping.md	Update 00-Lab-mapping.md	11 days ago
dp900-01-sql-lab.md	Fixed links	11 days ago
dp900-02-storage-lab.md	Register providers	10 days ago
dp900-03-cosmos-lab.md	Fixed links	11 days ago
dp900-04-synapse-lab.md	Fixed links	11 days ago
dp900-05-stream-lab.md	Register providers	10 days ago
dp900-pbi-06-lab.md	Minor update	10 days ago

Exercise: Explore Azure Storage

Provision an Azure Storage account

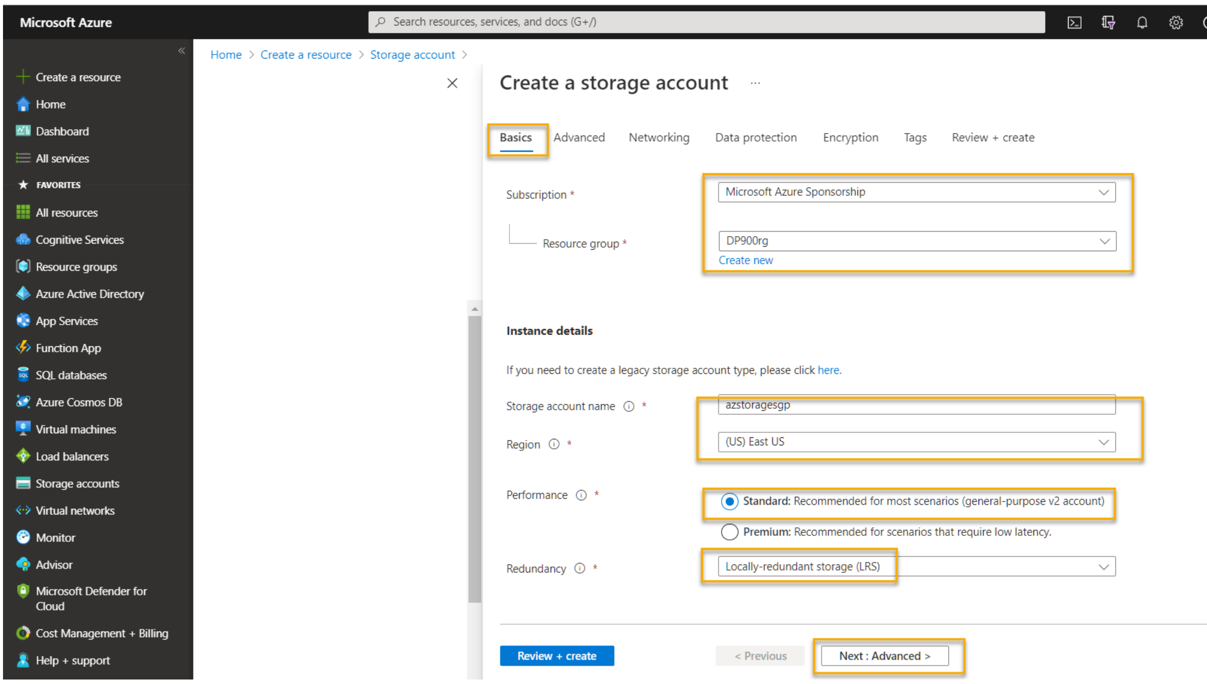
The first step in using Azure Storage is to provision an Azure Storage account in your Azure subscription.

1. If you haven't already done so, sign into the Azure portal at https://portal.azure.com. Then on the Azure portal home page, select **+ Create a resource** from the upper left-hand corner and search for *Storage account*. Then in the resulting **Storage account** page, select **Create**.



2. Enter the following values on the **Create a storage account** page:
    - **Subscription**: If you're using a sandbox, select *Concierge Subscription*. Otherwise, select your Azure subscription.

- **Resource group**: If you're using a sandbox, select the existing resource group (which will have a name like *learn-xxxx...*). Otherwise, create a new resource group with a name of your choice.
- **Storage account name**: Enter a unique name for your storage account using lower-case letters and numbers.
- **Region**: Select any available location.
- **Performance**: *Standard*
- **Redundancy**: *Locally-redundant storage (LRS)*



Create a storage account

Subscription *: Microsoft Azure Sponsorship

Resource group *: DP900rg

Instance details

If you need to create a legacy storage account type, please click here.

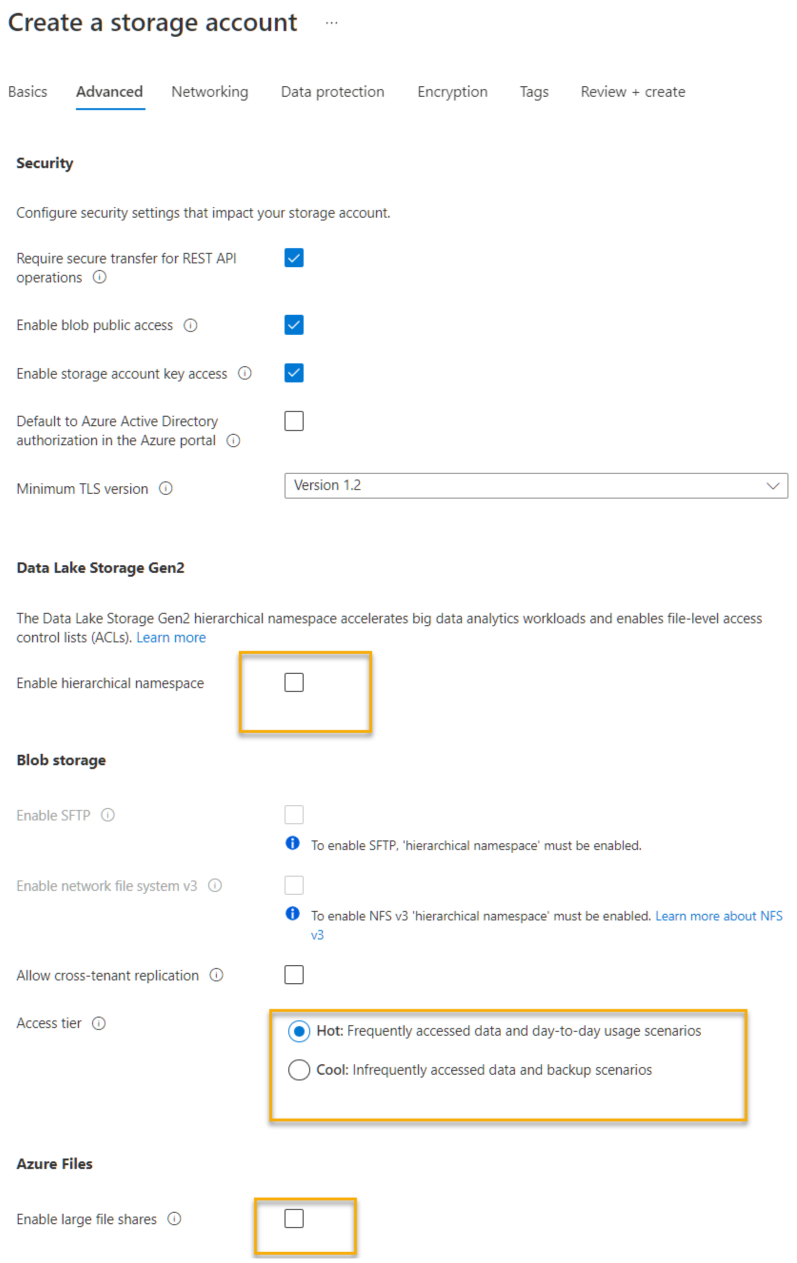
Storage account name *: azstoragesgp

Region *: (US) East US

Performance *:
● Standard: Recommended for most scenarios (general-purpose v2 account)
○ Premium: Recommended for scenarios that require low latency.

Redundancy *: Locally-redundant storage (LRS)

3. Select **Next: Advanced >** and view the advanced configuration options. In particular, note that this is where you can enable hierarchical namespace to support Azure Data Lake Storage Gen2. Leave this option **unselected** (you'll enable it later), and then select **Next: Networking >** to view the networking options for your storage account.



Create a storage account ···

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

Security

Configure security settings that impact your storage account.

Require secure transfer for REST API operations ⓘ ☑

Enable blob public access ⓘ ☑

Enable storage account key access ⓘ ☑

Default to Azure Active Directory authorization in the Azure portal ⓘ □

Minimum TLS version ⓘ Version 1.2

Data Lake Storage Gen2

The Data Lake Storage Gen2 hierarchical namespace accelerates big data analytics workloads and enables file-level access control lists (ACLs). Learn more

Enable hierarchical namespace □

Blob storage

Enable SFTP ⓘ □

To enable SFTP, 'hierarchical namespace' must be enabled.

Enable network file system v3 ⓘ □

To enable NFS v3 'hierarchical namespace' must be enabled. Learn more about NFS v3

Allow cross-tenant replication ⓘ □

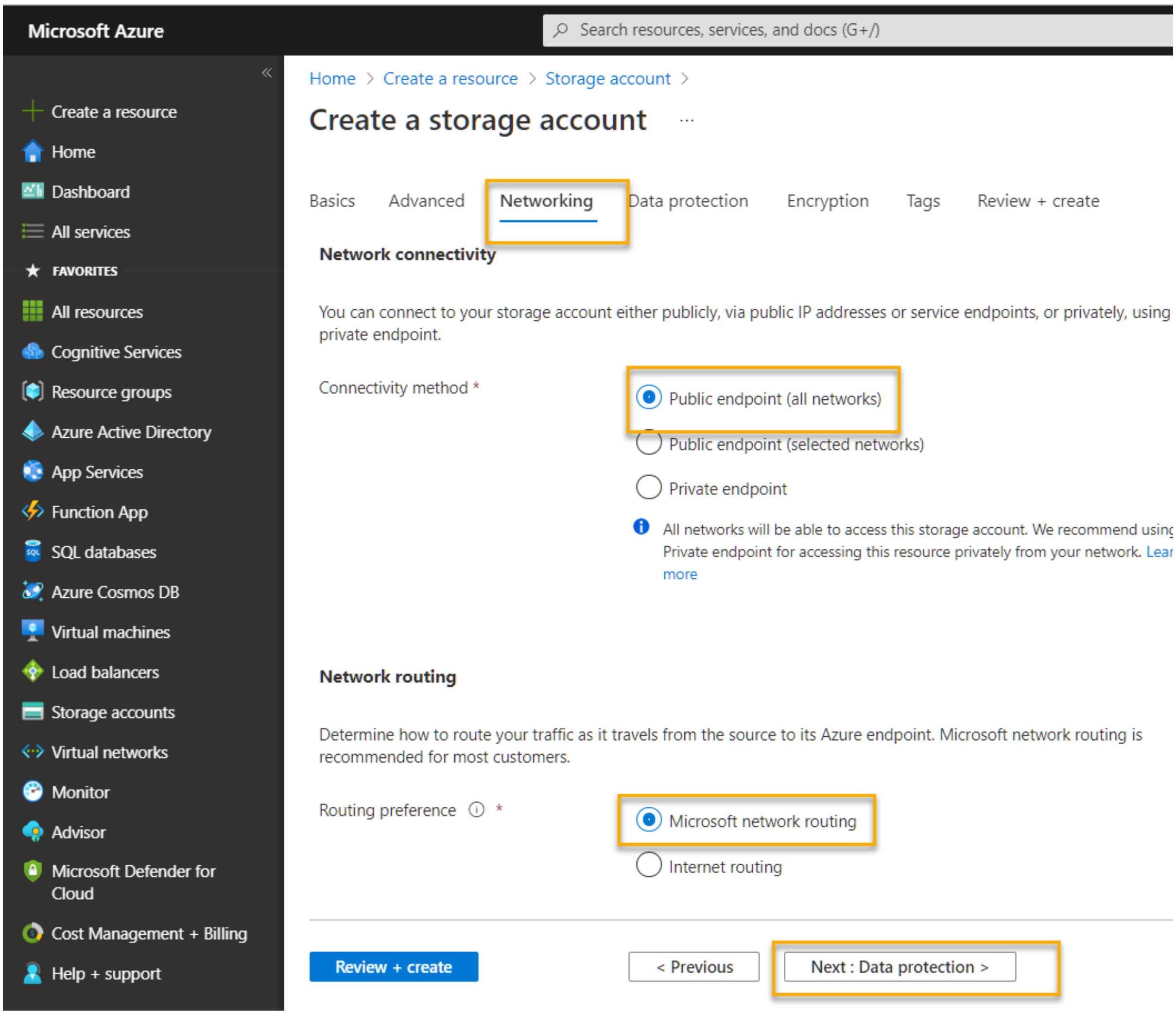
Access tier ⓘ

● **Hot:** Frequently accessed data and day-to-day usage scenarios

○ **Cool:** Infrequently accessed data and backup scenarios

Azure Files

Enable large file shares ⓘ □



Create a storage account ...

Network connectivity

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

Connectivity method *

● Public endpoint (all networks)

○ Public endpoint (selected networks)

○ Private endpoint

All networks will be able to access this storage account. We recommend using Private endpoint for accessing this resource privately from your network. Lear more

Network routing

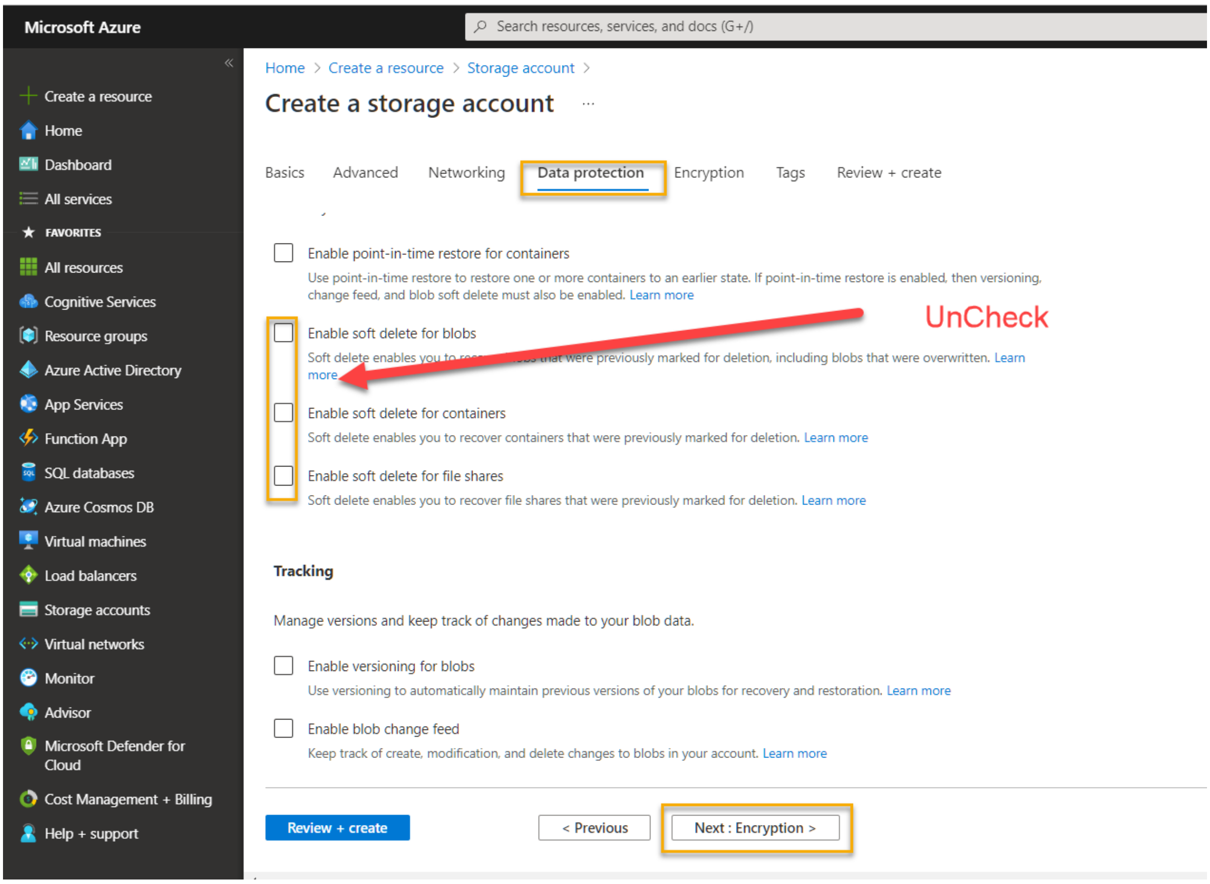
Determine how to route your traffic as it travels from the source to its Azure endpoint. Microsoft network routing is recommended for most customers.

Routing preference ⓘ *

● Microsoft network routing

○ Internet routing

4. Select **Next: Data protection >** and then in the **Recovery** section, deselect all of the **Enable soft delete...** options. These options retain deleted files for subsequent recovery, but can cause issues later when you enable hierarchical namespace.



Home > Create a resource > Storage account >

Create a storage account

□ Enable 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

UnCheck

□ Enable soft delete for blobs

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

□ Enable soft delete for containers

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

□ Enable soft delete for file shares

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

Tracking

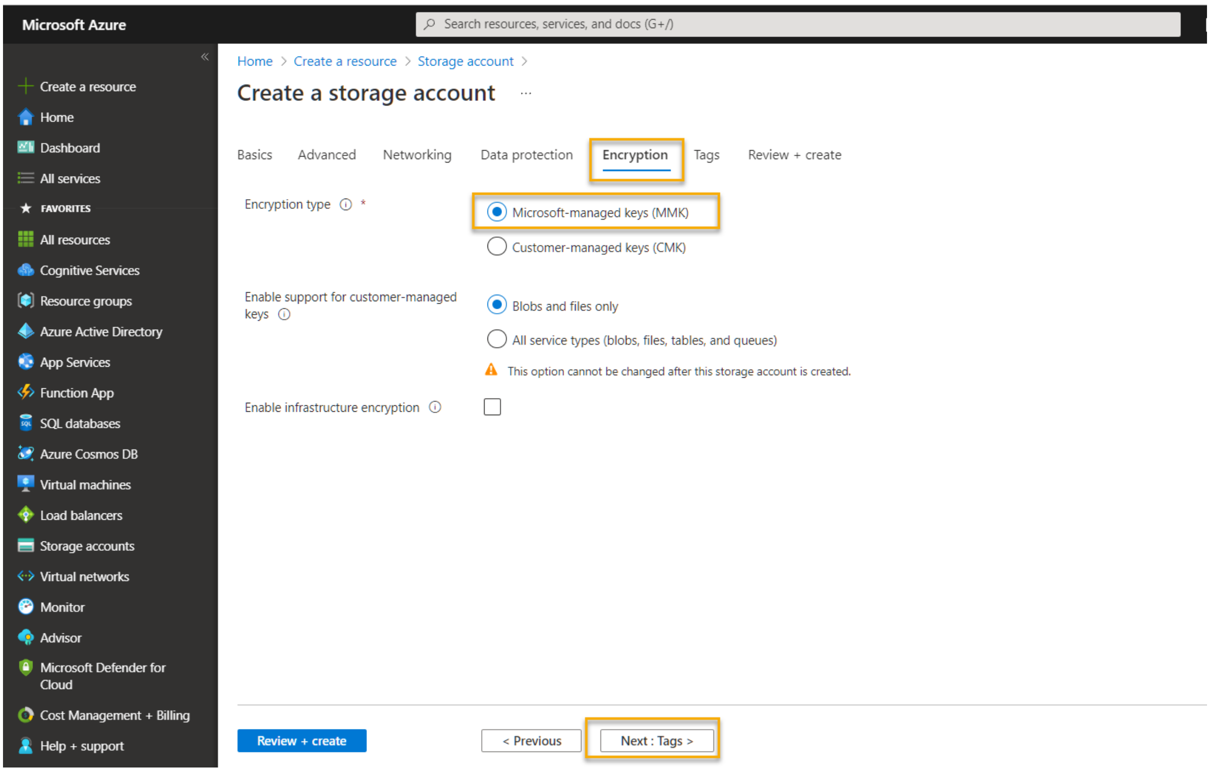
Manage versions and keep track of changes made to your blob data.

□ Enable versioning for blobs

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

□ Enable blob change feed

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



Home > Create a resource > Storage account >

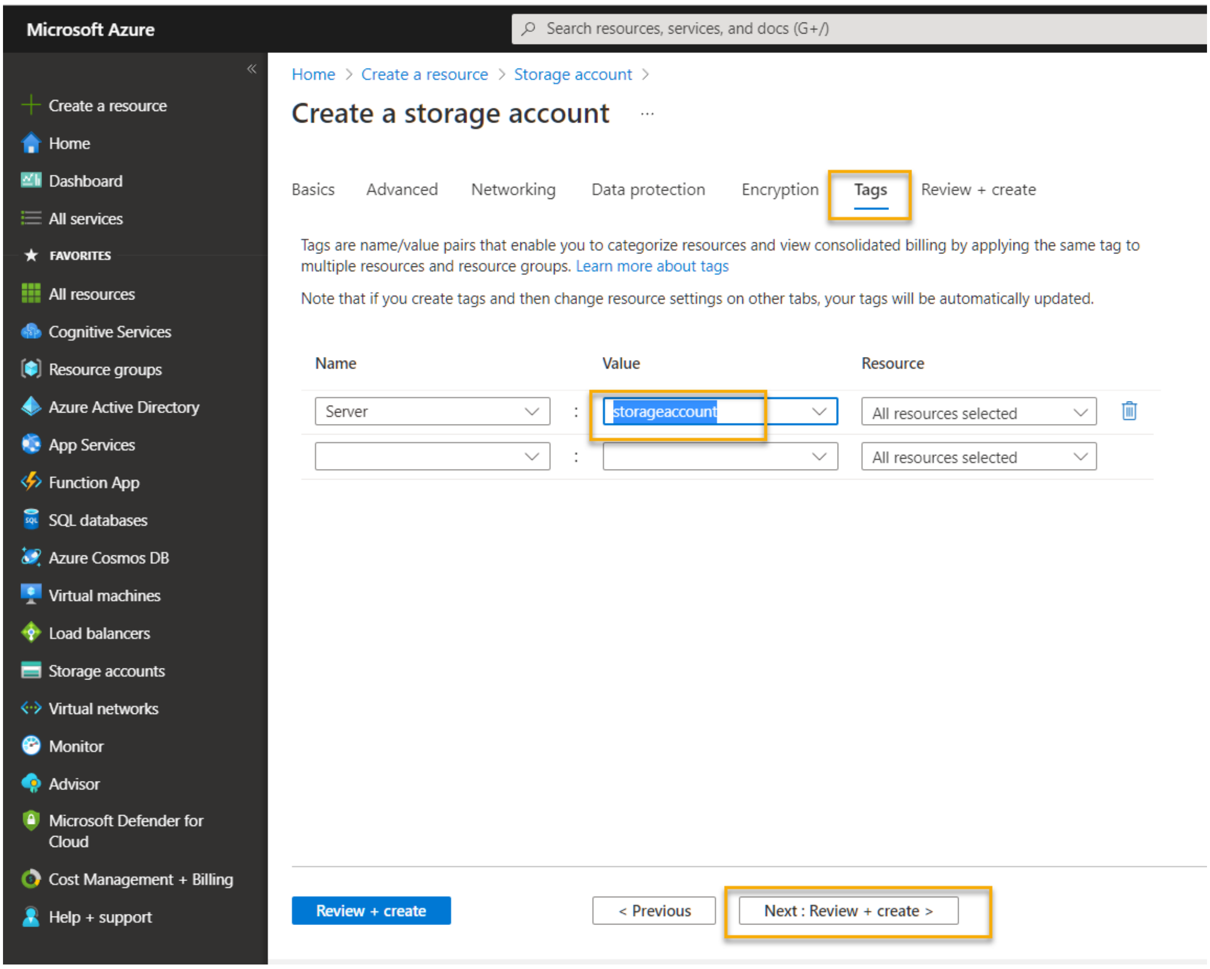
Create a storage account

Encryption type ⓘ *
● Microsoft-managed keys (MMK)
○ Customer-managed keys (CMK)

Enable support for customer-managed keys ⓘ
● Blobs and files only
○ All service types (blobs, files, tables, and queues)

⚠ This option cannot be changed after this storage account is created.

Enable infrastructure encryption ⓘ □



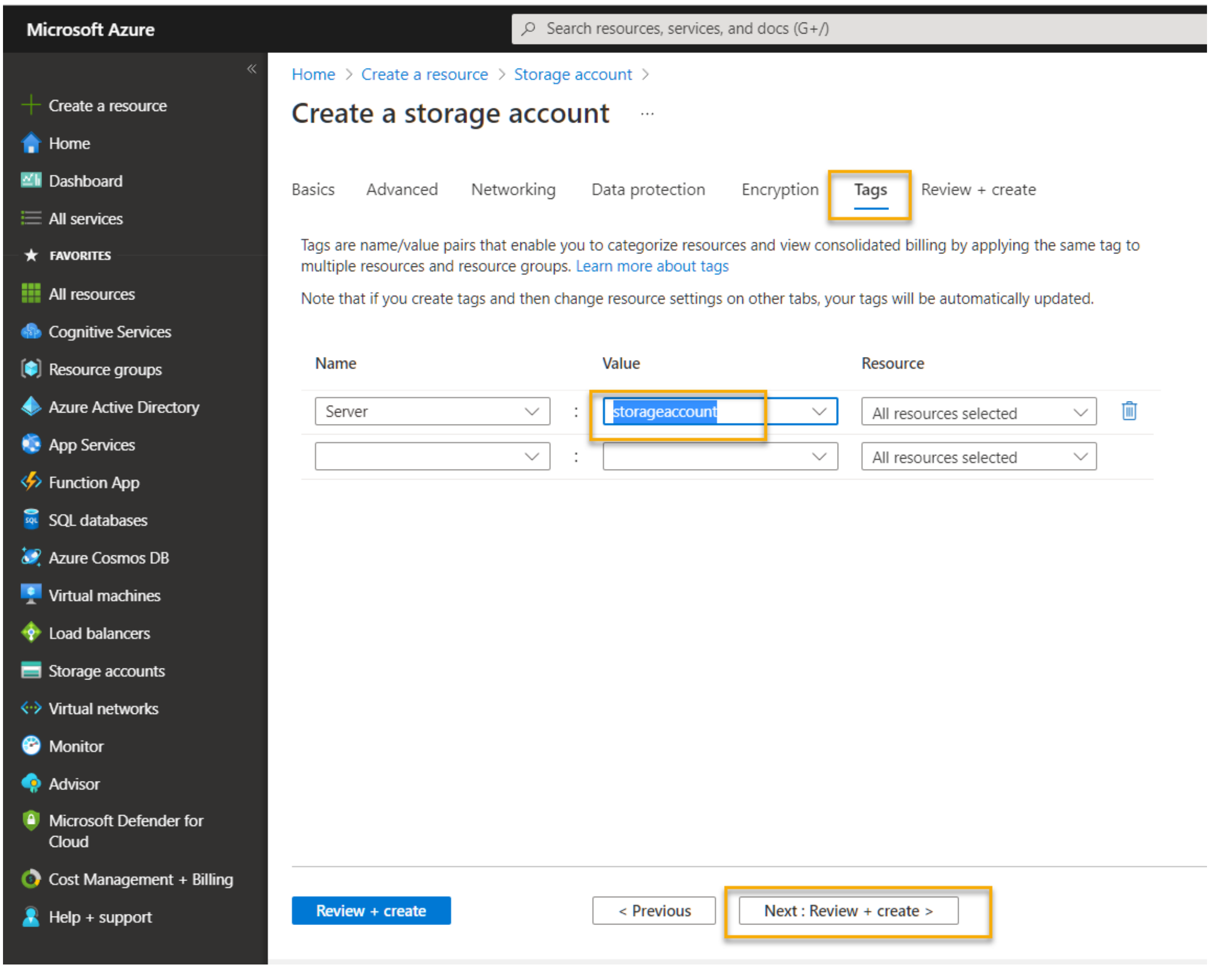
Create a storage account

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

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

Name		Value	Resource
Server	:	storageaccount	All resources selected
	:		All resources selected

5. Continue through the remaining **Next >** pages without changing any of the default settings, and then on the **Review + Create** page, wait for your selections to be validated and select **Create** to create your Azure Storage account.



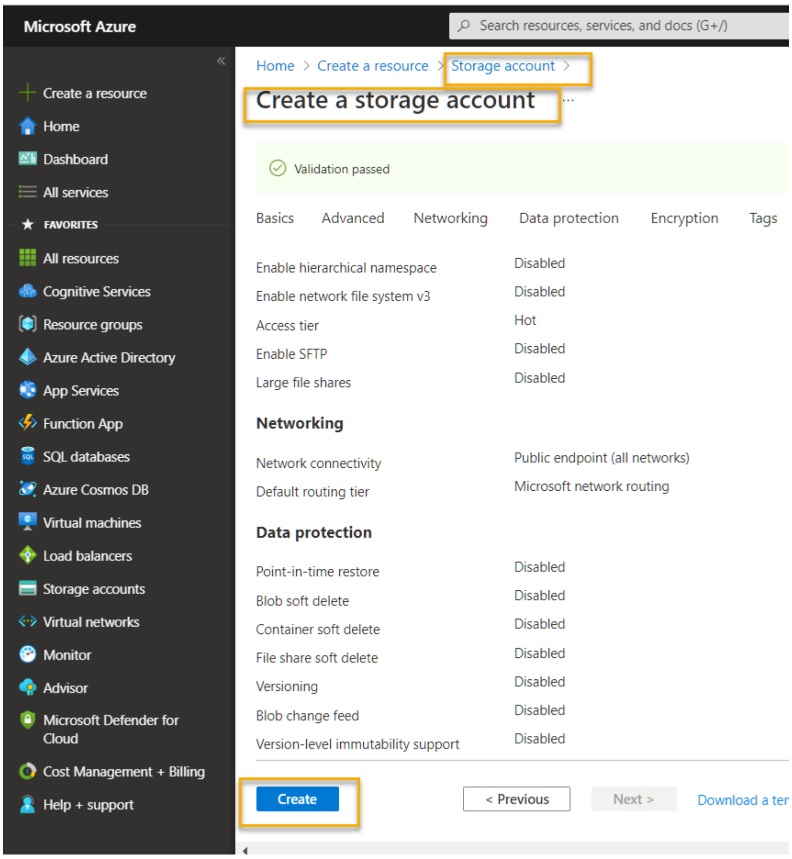
Create a storage account ...

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

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

Name		Value	Resource
Server	:	storageaccount	All resources selected
	:		All resources selected

6. Wait for deployment to complete. Then go to the resource that was deployed.



Create a storage account

Validation passed

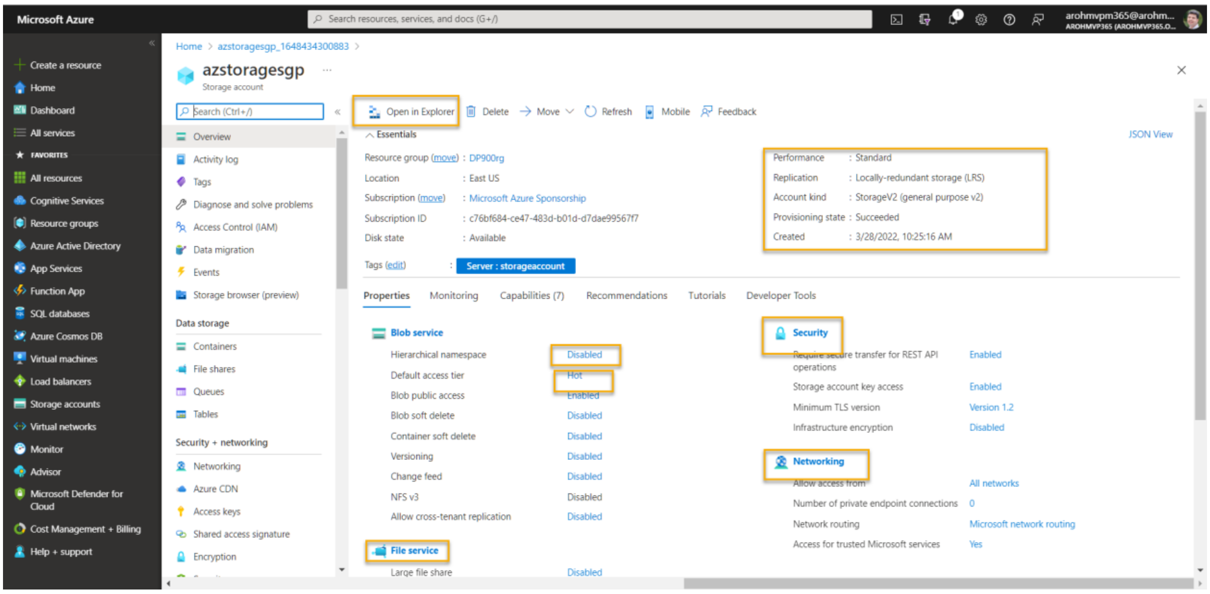
Enable hierarchical namespace: Disabled
Enable network file system v3: Disabled
Access tier: Hot
Enable SFTP: Disabled
Large file shares: Disabled

Networking

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

Data protection

Point-in-time restore: Disabled
Blob soft delete: Disabled
Container soft delete: Disabled
File share soft delete: Disabled
Versioning: Disabled
Blob change feed: Disabled
Version-level immutability support: Disabled



azstoragesgp
Storage account

Essentials

Resource group (move) : DP900rg
Location : East US
Subscription (move) : Microsoft Azure Sponsorship
Subscription ID : c76bf684-ce47-483d-b01d-d7dae99567f7
Disk state : Available
Tags (edit) : Server : storageaccount

Performance : Standard
Replication : Locally-redundant storage (LRS)
Account kind : StorageV2 (general purpose v2)
Provisioning state : Succeeded
Created : 3/28/2022, 10:25:16 AM

Blob service

Hierarchical namespace: Disabled
Default access tier: Hot
Blob public access: Enabled
Blob soft delete: Disabled
Container soft delete: Disabled
Versioning: Disabled
Change feed: Disabled
NFS v3: Disabled
Allow cross-tenant replication: Disabled

File service

Large file share: Disabled

Security

Require secure transfer for REST API operations: Enabled
Storage account key access: Enabled
Minimum TLS version: Version 1.2
Infrastructure encryption: Disabled

Networking

Allow access from: All networks
Number of private endpoint connections: 0
Network routing: Microsoft network routing
Access for trusted Microsoft services: Yes

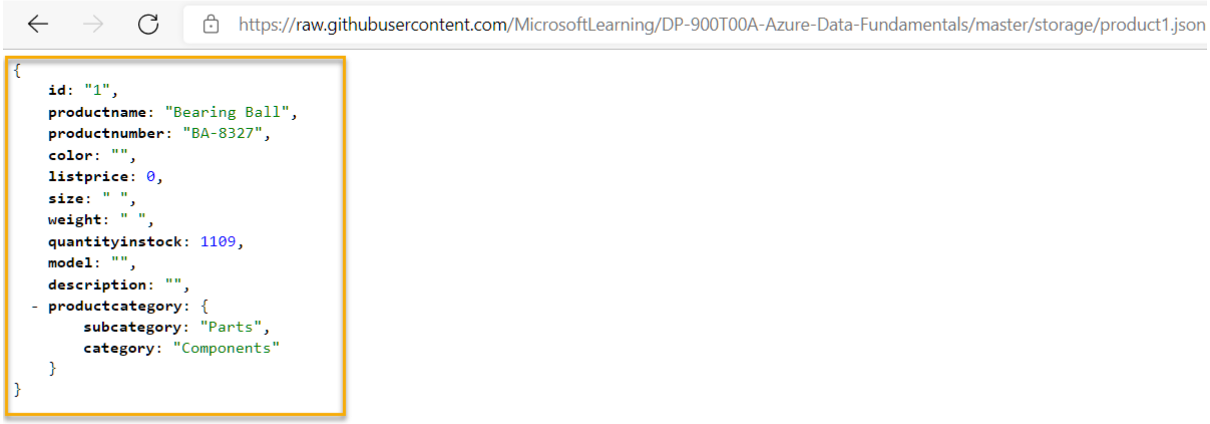
Explore blob storage

Now that you have an Azure Storage account, you can create a container for blob data.

1. Download the [product1.json](https://aka.ms/product1.json) JSON file from https://aka.ms/product1.json and save it on your computer (you can save it in any folder - you'll upload it to blob storage later).

*If the JSON file is displayed in your browser, save the page as **product1.json**.*

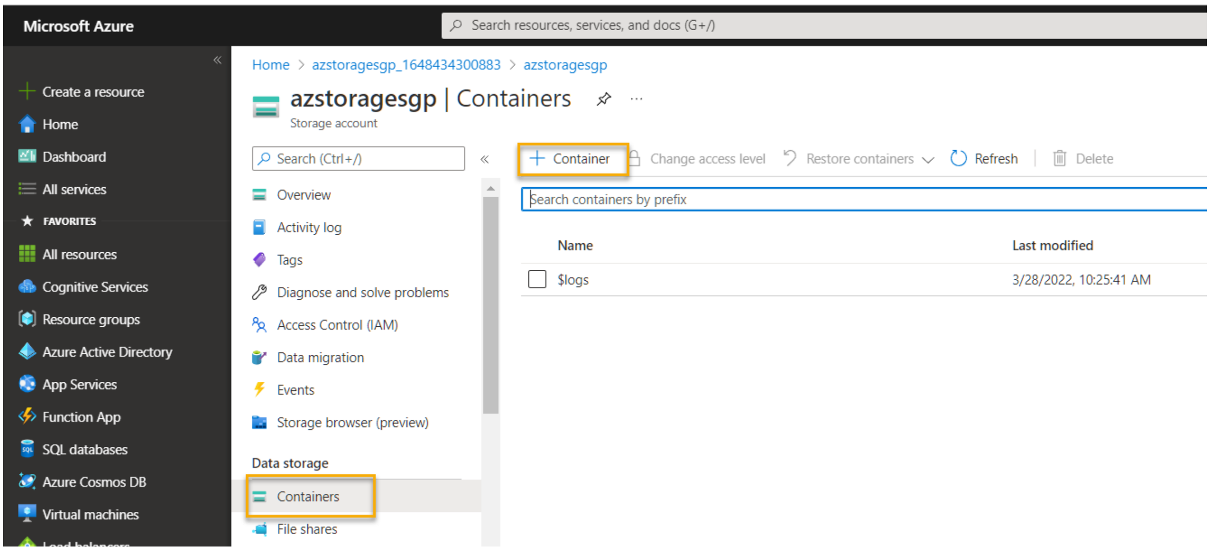
https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/master/storage/product1.json



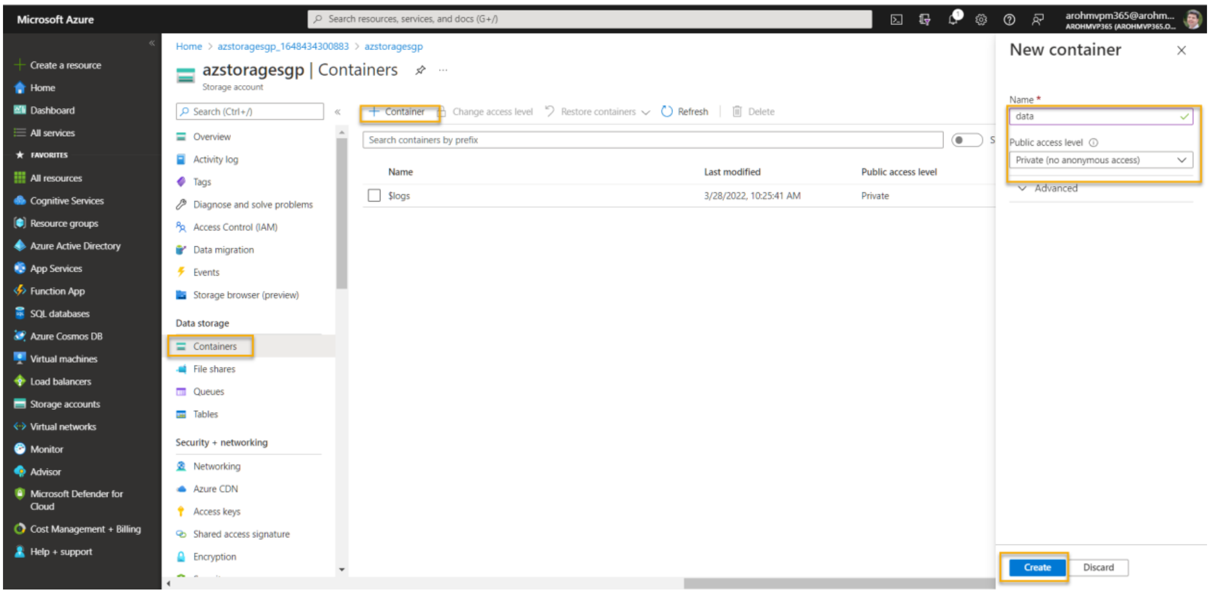
https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/master/storage/product1.json

```
{
    id: "1",
    productname: "Bearing Ball",
    productnumber: "BA-8327",
    color: "",
    listprice: 0,
    size: " ",
    weight: " ",
    quantityinstock: 1109,
    model: "",
    description: "",
  - productcategory: {
        subcategory: "Parts",
        category: "Components"
    }
}
```

2. In the Azure portal page for your storage container, on the left side, in the **Data storage** section, select **Containers**.



3. In the **Containers** page, select **+ Container** and add a new container named **data** with a public access level of **Private (no anonymous access)**.

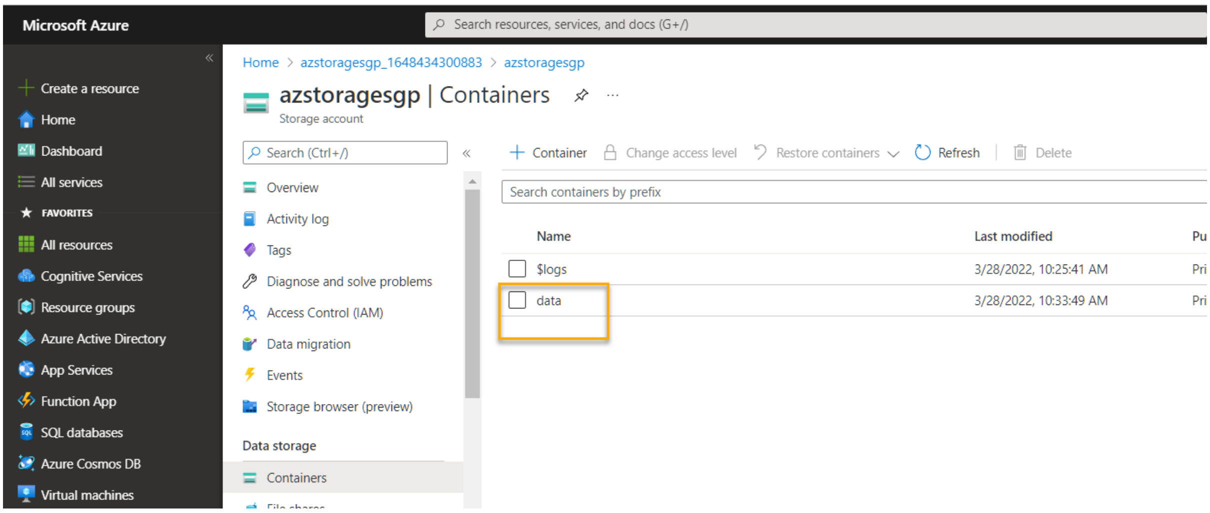


Name	Last modified	Public access level
□ $logs	3/28/2022, 10:25:41 AM	Private

Name * data

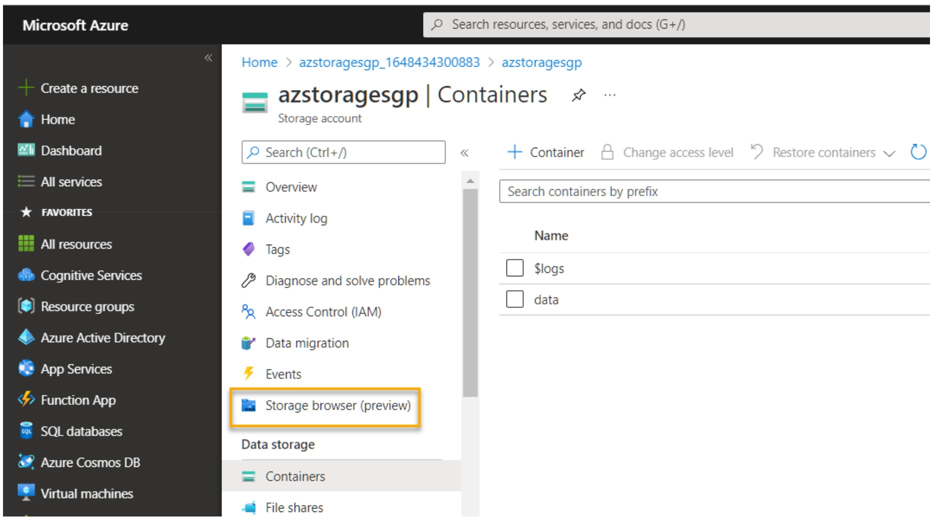
Public access level Private (no anonymous access)

4. When the **data** container has been created, verify that it's listed in the **Containers** page.

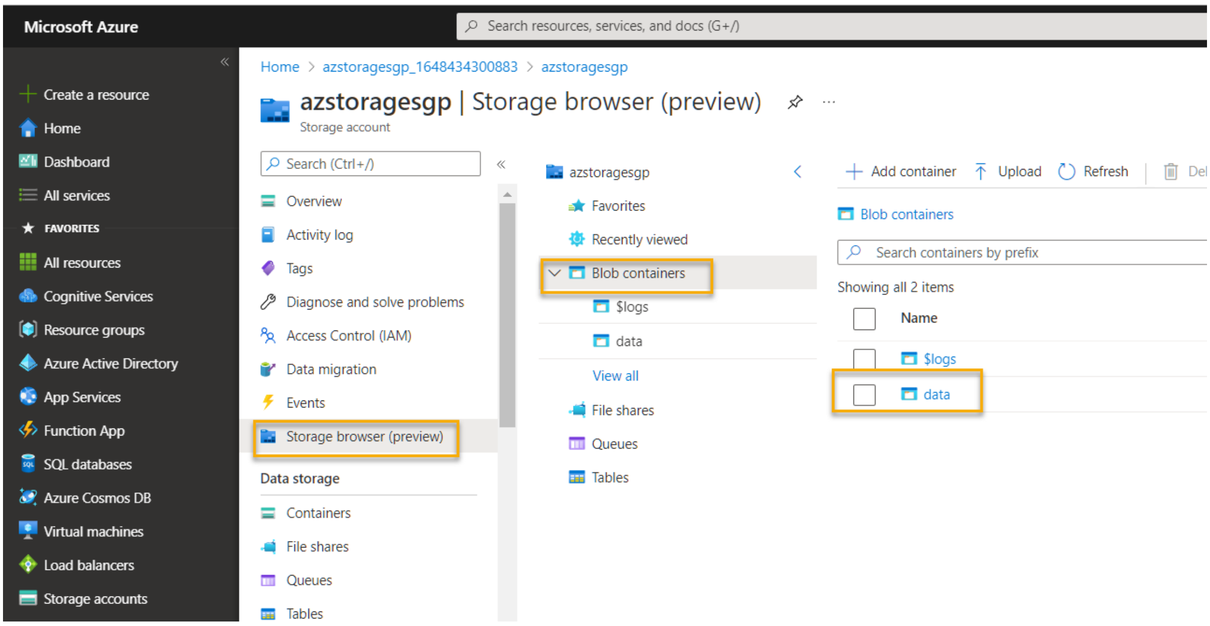


Name	Last modified	Pu
□ $logs	3/28/2022, 10:25:41 AM	Pri
□ data	3/28/2022, 10:33:49 AM	Pri

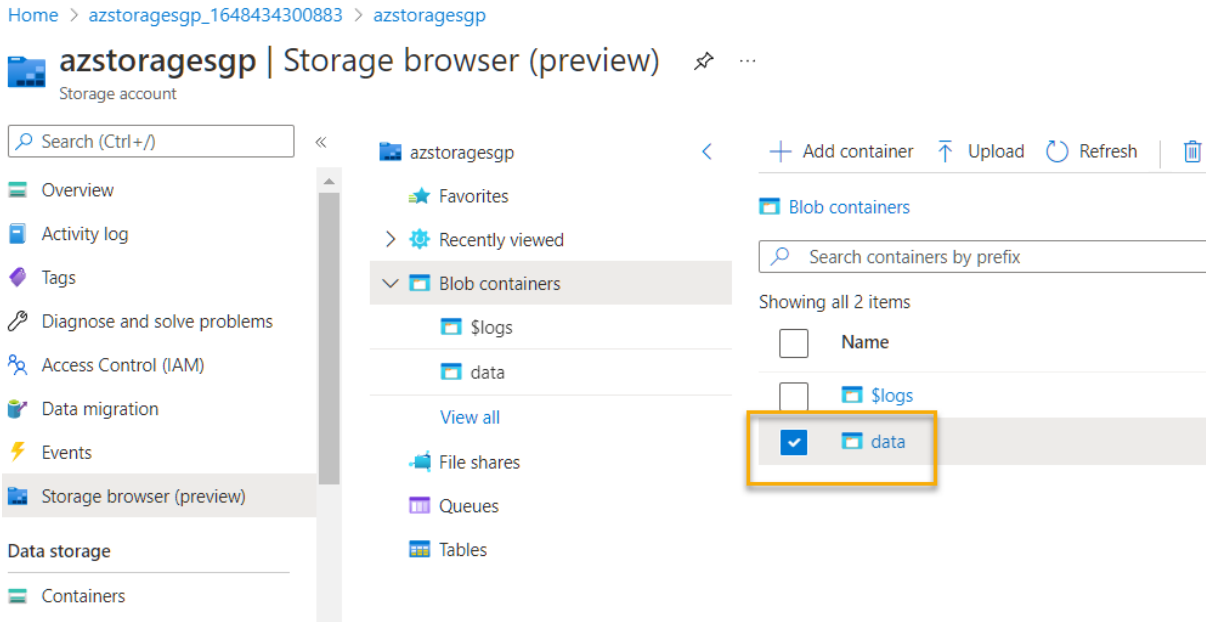
5. In the pane on the left side, in the top section, select **Storage browser (preview)**. This page provides a browser-based interface that you can use to work with the data in your storage account.



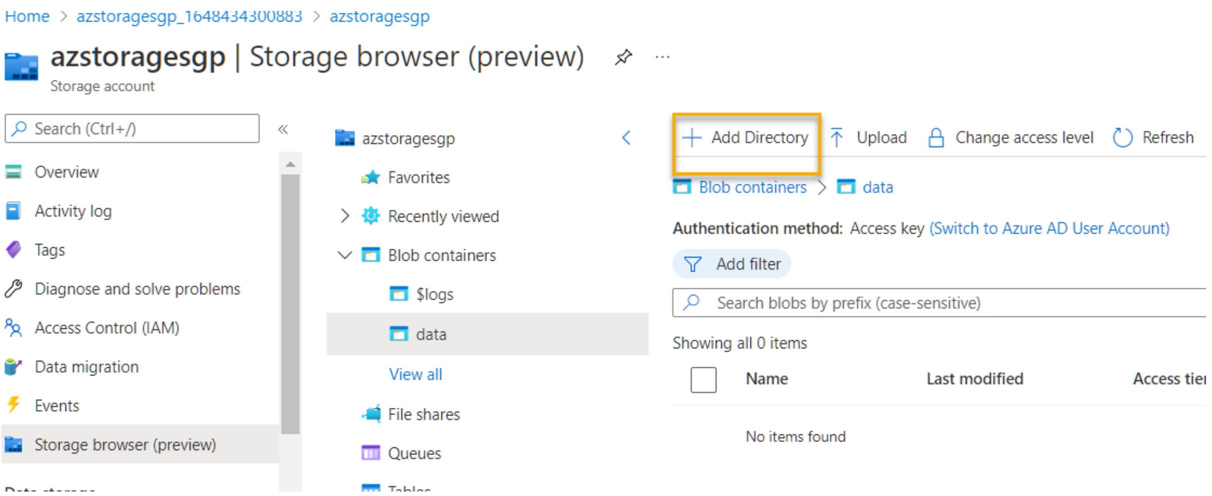
6. In the storage browser page, select **Blob containers** and verify that your **data** container is listed.

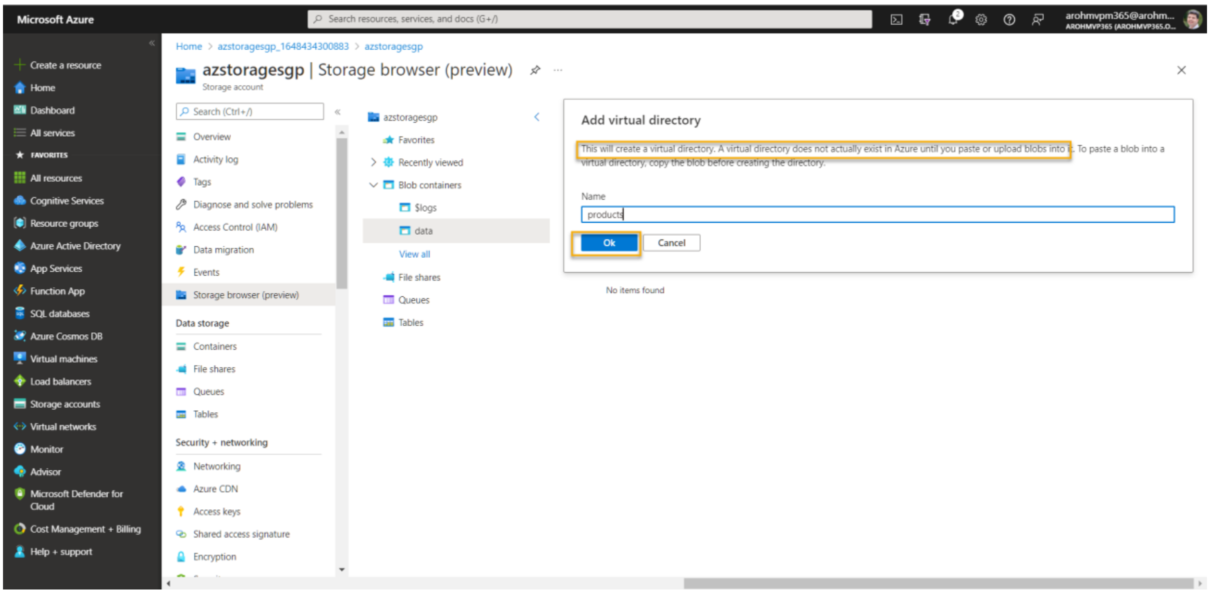


7. Select the **data** container, and note that it's empty.



8. Select + **Add Directory** and read the information about folders before creating a new directory named **products**.




Add virtual directory

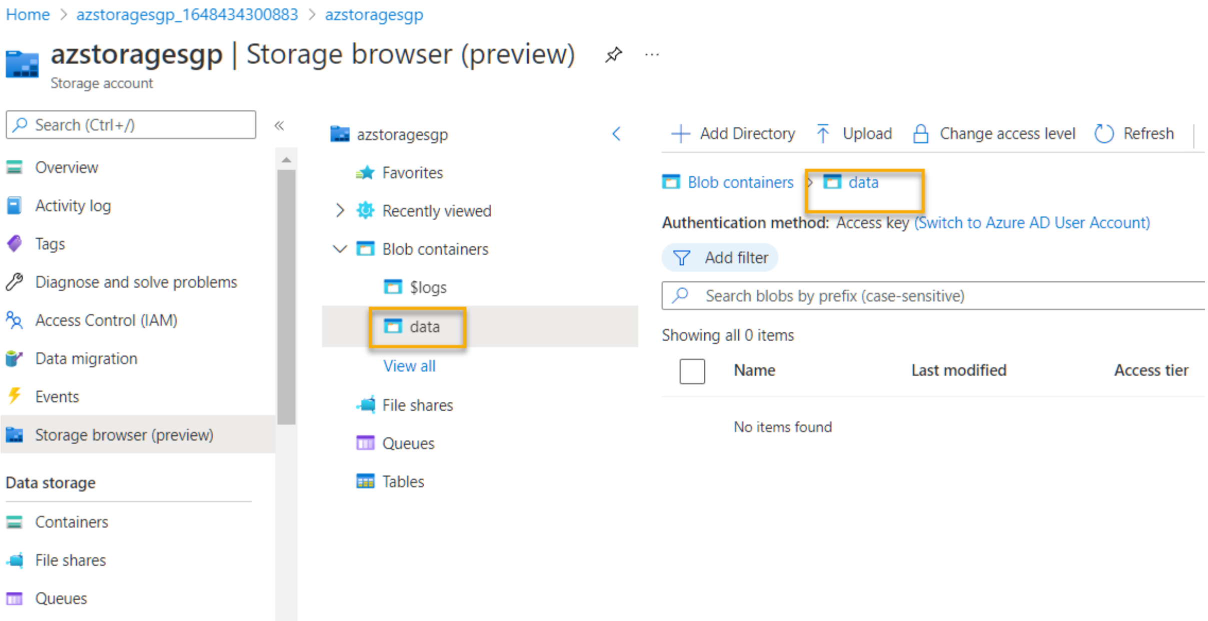
This will create a virtual directory. A virtual directory does not actually exist in Azure until you paste or upload blobs into it. To paste a blob into a virtual directory, copy the blob before creating the directory.

Name: products

No items found

9. In storage explorer, verify that the current view shows the contents of the **products** folder you just created - observe that the "breadcrumbs" at the top of the page reflect the path **Blob containers > data > products**.

10. In the breadcrumbs, select **data** to switch to the **data** container, and note that it does <u>not</u> contain a folder named **products**.

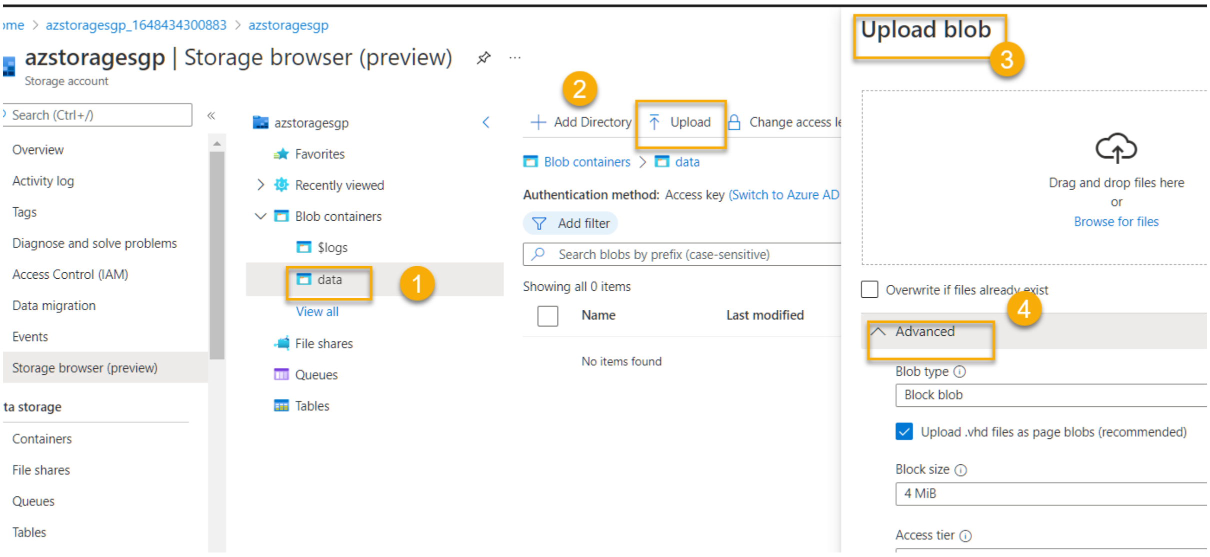

Authentication method: Access key (Switch to Azure AD User Account)

Showing all 0 items

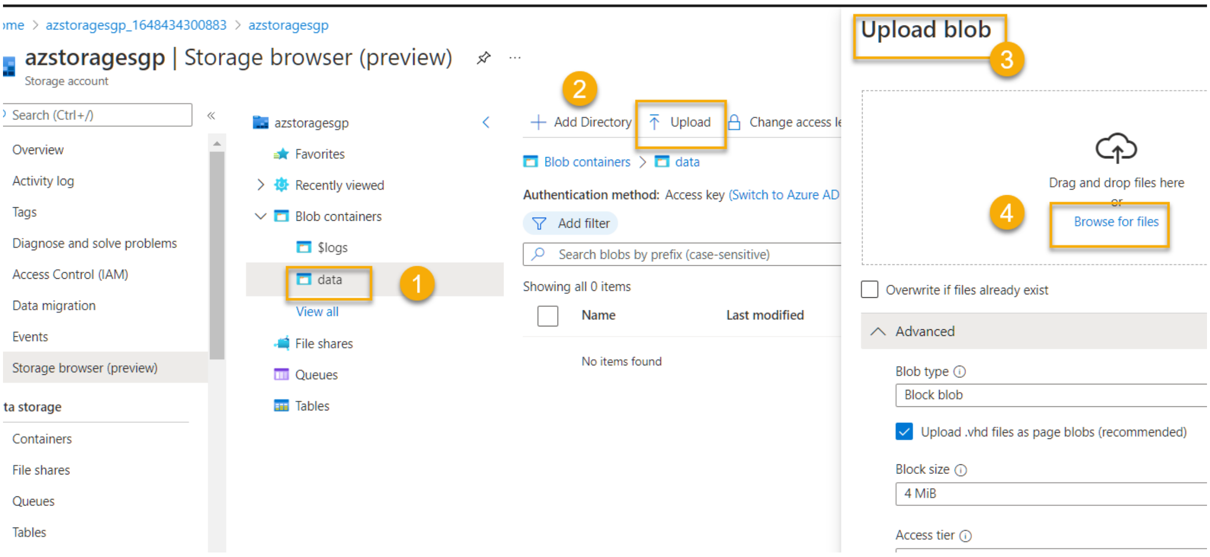
No items found

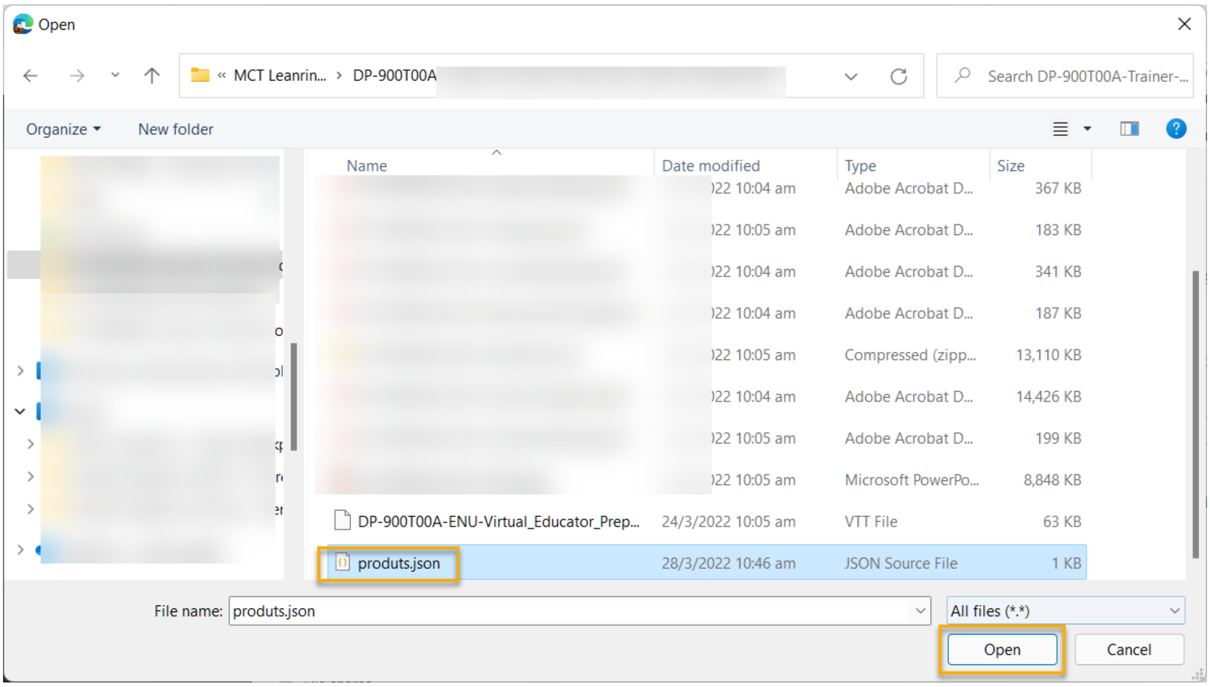
Folders in blob storage are virtual, and only exist as part of the path of a blob. Since the **products** folder contained no blobs, it isn't really there!

11. Use the ⤒ **Upload** button to open the **Upload blob** panel.



12. In the **Upload blob** panel, select the **product1.json** file you saved on your local computer previously.



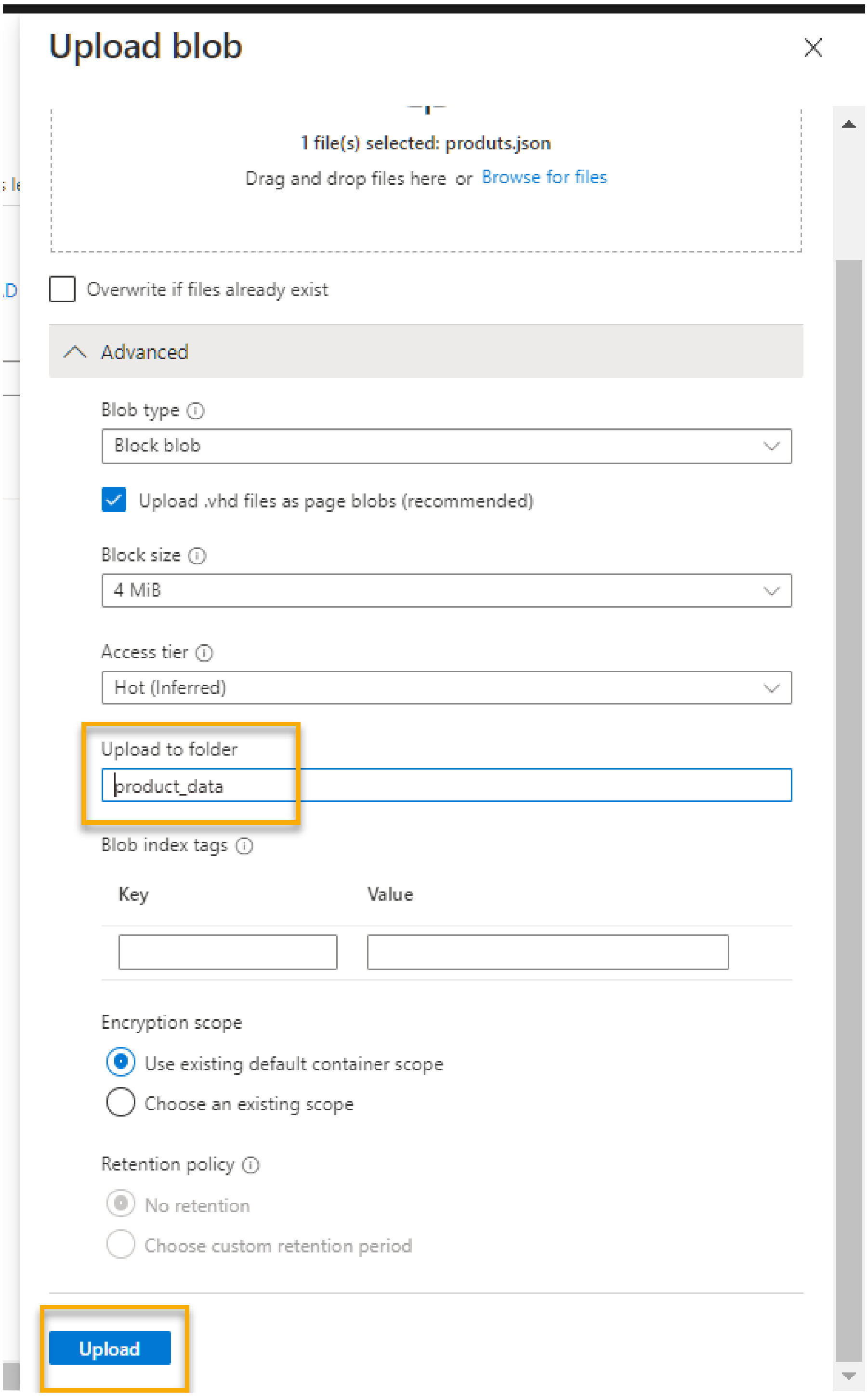


Name	Date modified	Type	Size
[redacted]	22 10:04 am	Adobe Acrobat D...	367 KB
[redacted]	22 10:05 am	Adobe Acrobat D...	183 KB
[redacted]	22 10:04 am	Adobe Acrobat D...	341 KB
[redacted]	22 10:04 am	Adobe Acrobat D...	187 KB
[redacted]	22 10:05 am	Compressed (zipp...	13,110 KB
[redacted]	22 10:04 am	Adobe Acrobat D...	14,426 KB
[redacted]	22 10:05 am	Adobe Acrobat D...	199 KB
[redacted]	22 10:05 am	Microsoft PowerPo...	8,848 KB
DP-900T00A-ENU-Virtual_Educator_Prep...	24/3/2022 10:05 am	VTT File	63 KB
produts.json	28/3/2022 10:46 am	JSON Source File	1 KB

File name: produts.json

All files (*.*)

Then in the **Advanced** section, in the **Upload to folder** box, enter **product_data** and select the **Upload** button.



Upload blob

1 file(s) selected: produts.json

Drag and drop files here or Browse for files

□ Overwrite if files already exist

Advanced

Blob type: Block blob

☑ Upload .vhd files as page blobs (recommended)

Block size: 4 MiB

Access tier: Hot (Inferred)

Upload to folder: product_data

Blob index tags

Key	Value

Encryption scope

● Use existing default container scope

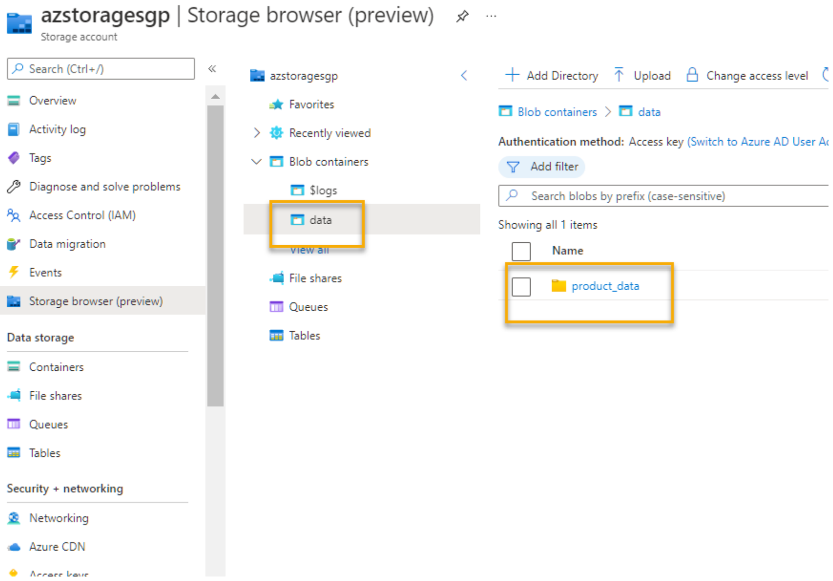
○ Choose an existing scope

Retention policy

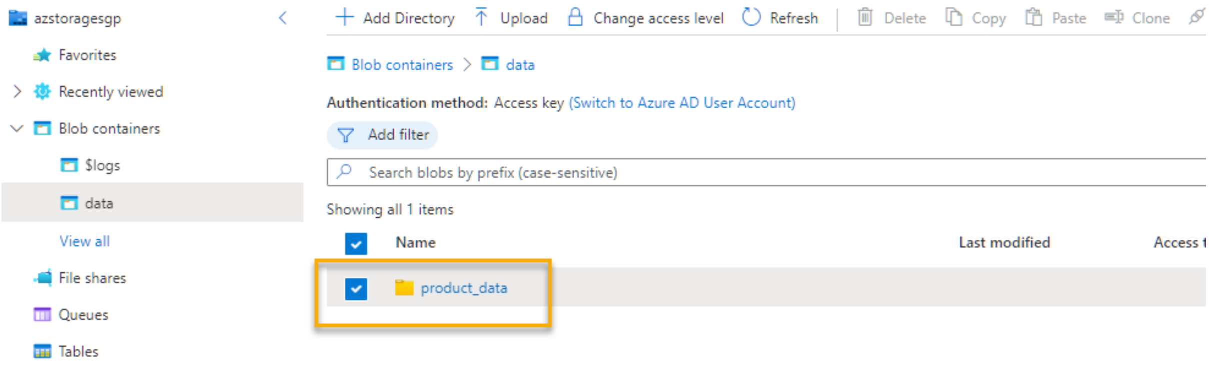
● No retention

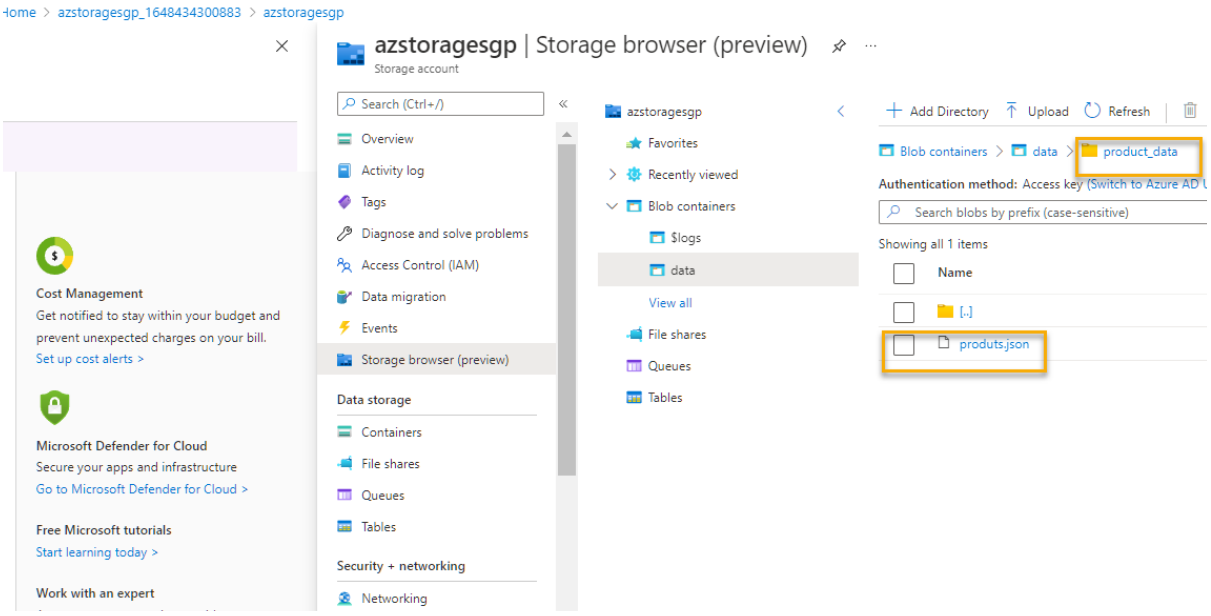
○ Choose custom retention period

13. Close the **Upload blob** panel if it's still open, and verify that a **product_data** virtual folder has been created in the **data** container.

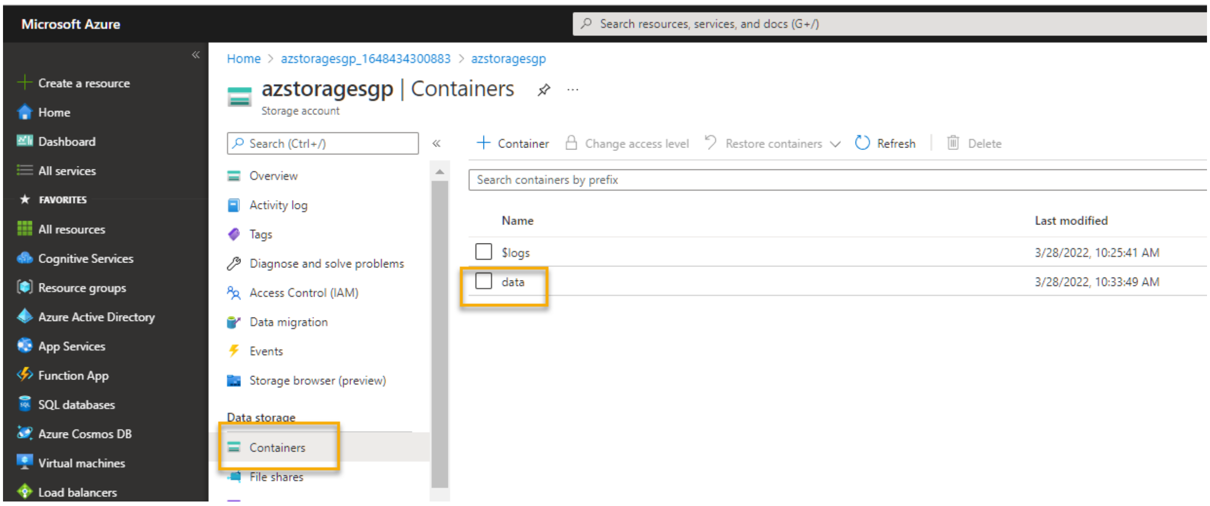


14. Select the **product_data** folder and verify that it contains the **product1.json** blob you uploaded.



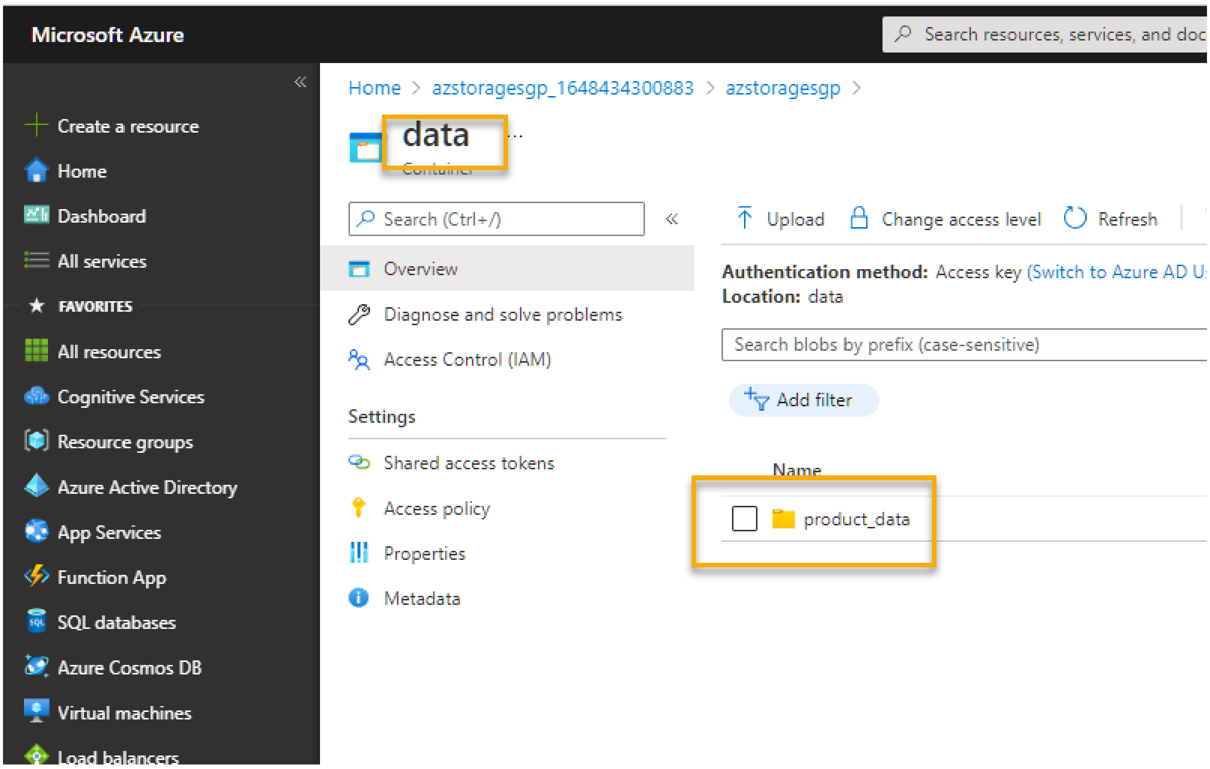


15. On the left side, in the **Data storage** section, select **Containers**.



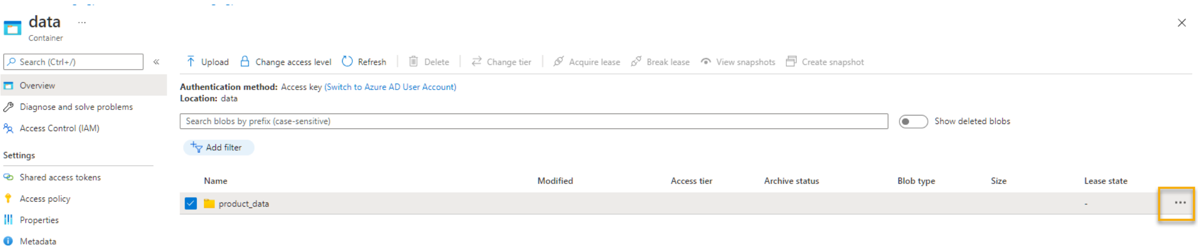
Name	Last modified
$logs	3/28/2022, 10:25:41 AM
data	3/28/2022, 10:33:49 AM

16. Open the **data** container, and verify that the **product_data** folder you created is listed.



Authentication method: Access key (Switch to Azure AD U
Location: data

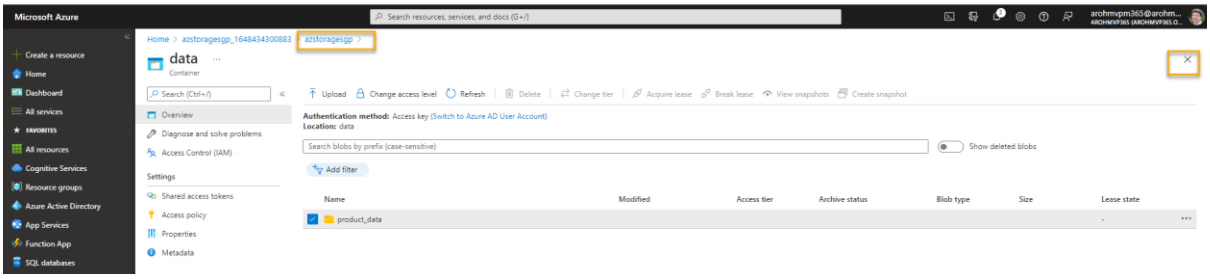
17. Select the ··· icon at the right-end of the folder, and note that it doesn't display any options. Folders in a flat namespace blob container are virtual, and can’t be managed.



Authentication method: Access key (Switch to Azure AD User Account)
Location: data

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
☑ product_data						-

18. Use the **X** icon at the top right in the **data** page to close the page and return to the **Containers** page.



Authentication method: Access key (Switch to Azure AD User Account)
Location: data

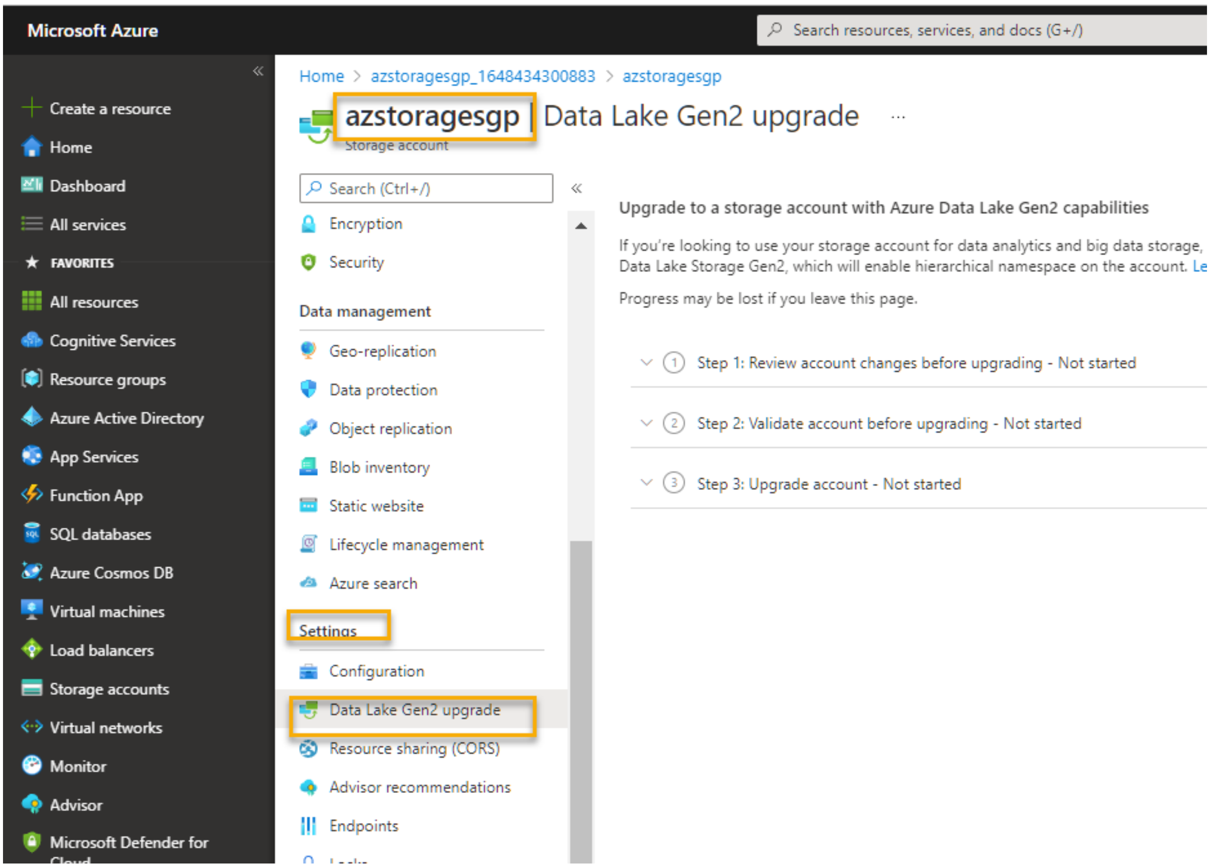
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
product_data						-

Explore Azure Data Lake Storage Gen2

Azure Data Lake Store Gen2 support enables you to use **hierarchical folders to organize** and manage **access to blobs**. It also enables you to use **Azure blob storage** to host **distributed file systems** for common big data analytics platforms.

1. Download the [product2.json](https://aka.ms/product2.json) JSON file from https://aka.ms/product2.json and save it on your computer in the same folder where you downloaded **product1.json** previously - you'll upload it to blob storage later).

2. In the Azure portal page for your storage container, on the left side, scroll down to the **Settings** section, and select **Data Lake Gen2 upgrade**.



Upgrade to a storage account with Azure Data Lake Gen2 capabilities

If you're looking to use your storage account for data analytics and big data storage, Data Lake Storage Gen2, which will enable hierarchical namespace on the account. Le

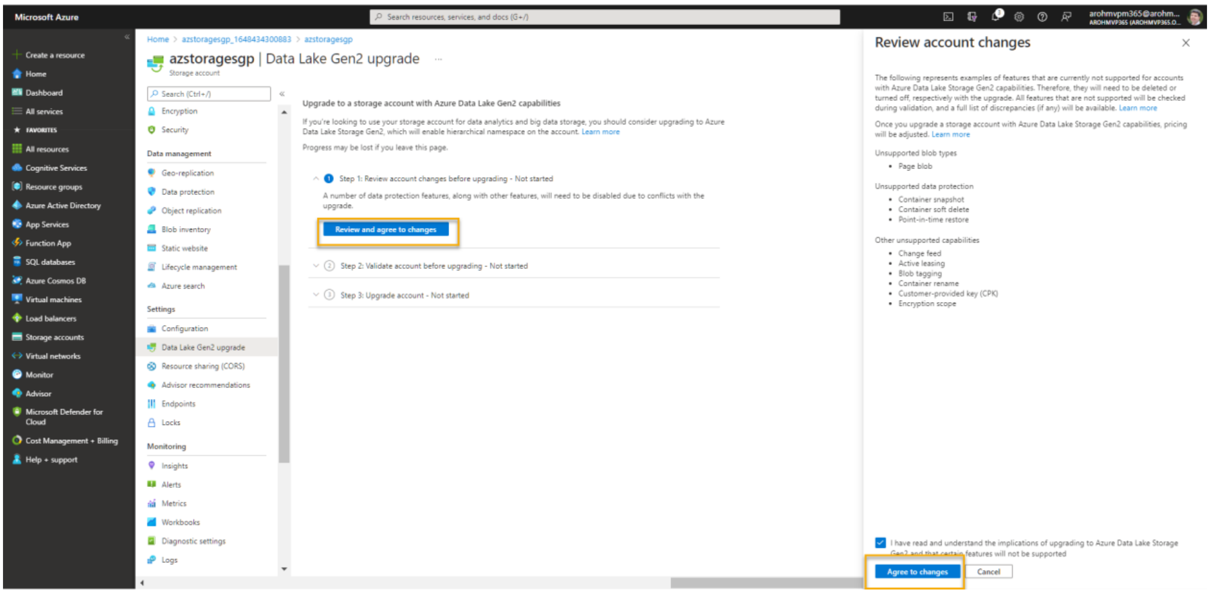
Progress may be lost if you leave this page.

1 Step 1: Review account changes before upgrading - Not started

2 Step 2: Validate account before upgrading - Not started

3 Step 3: Upgrade account - Not started

3. In the **Data Lake Gen2 upgrade** page, expand and complete each step to upgrade your storage account to enable hierarchical namespace and support Azure Data Lake Storage Gen 2. This may take some time.



Upgrade to a storage account with Azure Data Lake Gen2 capabilities

If you're looking to use your storage account for data analytics and big data storage, you should consider upgrading to Azure Data Lake Storage Gen2, which will enable hierarchical namespace on the account. Learn more

Progress may be lost if you leave this page.

1. Step 1: Review account changes before upgrading - Not started

   A number of data protection features, along with other features, will need to be disabled due to conflicts with the upgrade.

   Review and agree to changes

2. Step 2: Validate account before upgrading - Not started

3. Step 3: Upgrade account - Not started

Review account changes

The following represents examples of features that are currently not supported for accounts with Azure Data Lake Storage Gen2 capabilities. Therefore, they will need to be deleted or turned off, respectively with the upgrade. All features that are not supported will be checked during validation, and a full list of discrepancies (if any) will be available. Learn more

Once you upgrade a storage account with Azure Data Lake Storage Gen2 capabilities, pricing will be adjusted. Learn more

Unsupported blob types

- Page blob

Unsupported data protection

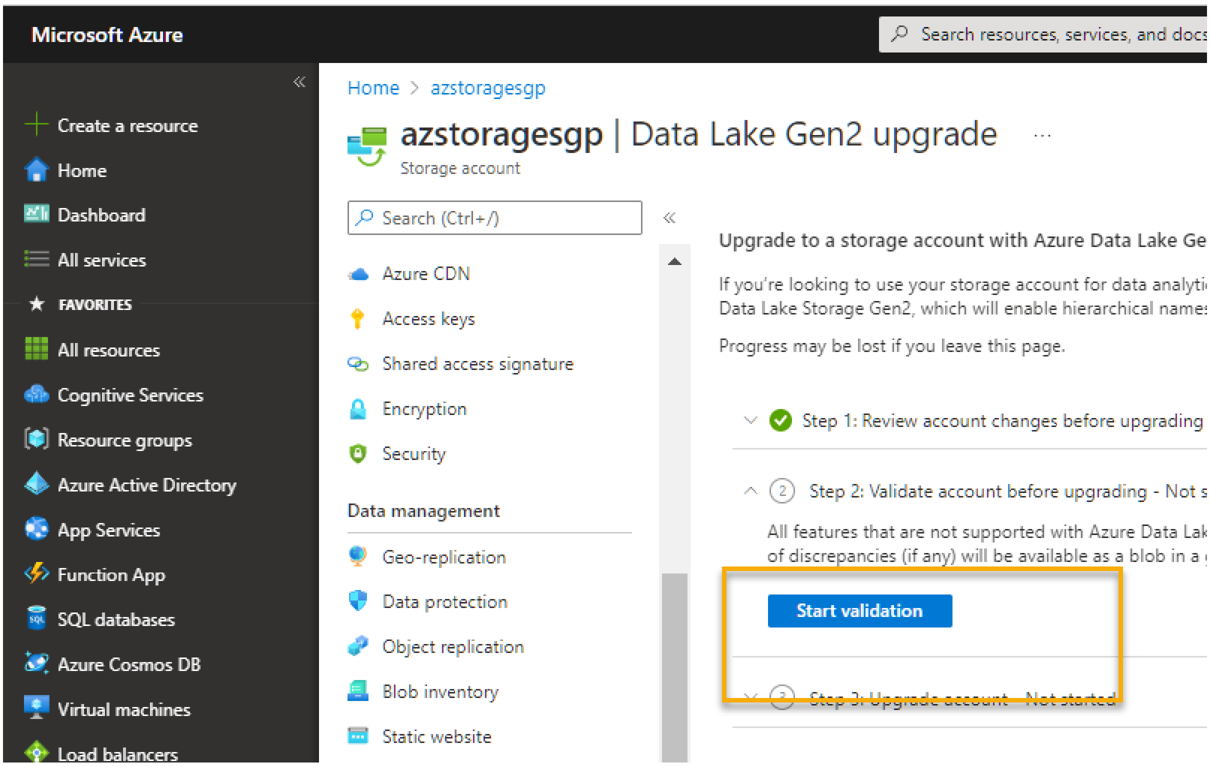
- Container snapshot
- Container soft delete
- Point-in-time restore

Other unsupported capabilities

- Change feed
- Active leasing
- Blob tagging
- Container rename
- Customer-provided key (CPK)
- Encryption scope

☑ I have read and understand the implications of upgrading to Azure Data Lake Storage Gen2 and that certain features will not be supported

Agree to changes | Cancel



Upgrade to a storage account with Azure Data Lake Ge

If you're looking to use your storage account for data analyti
Data Lake Storage Gen2, which will enable hierarchical names

Progress may be lost if you leave this page.

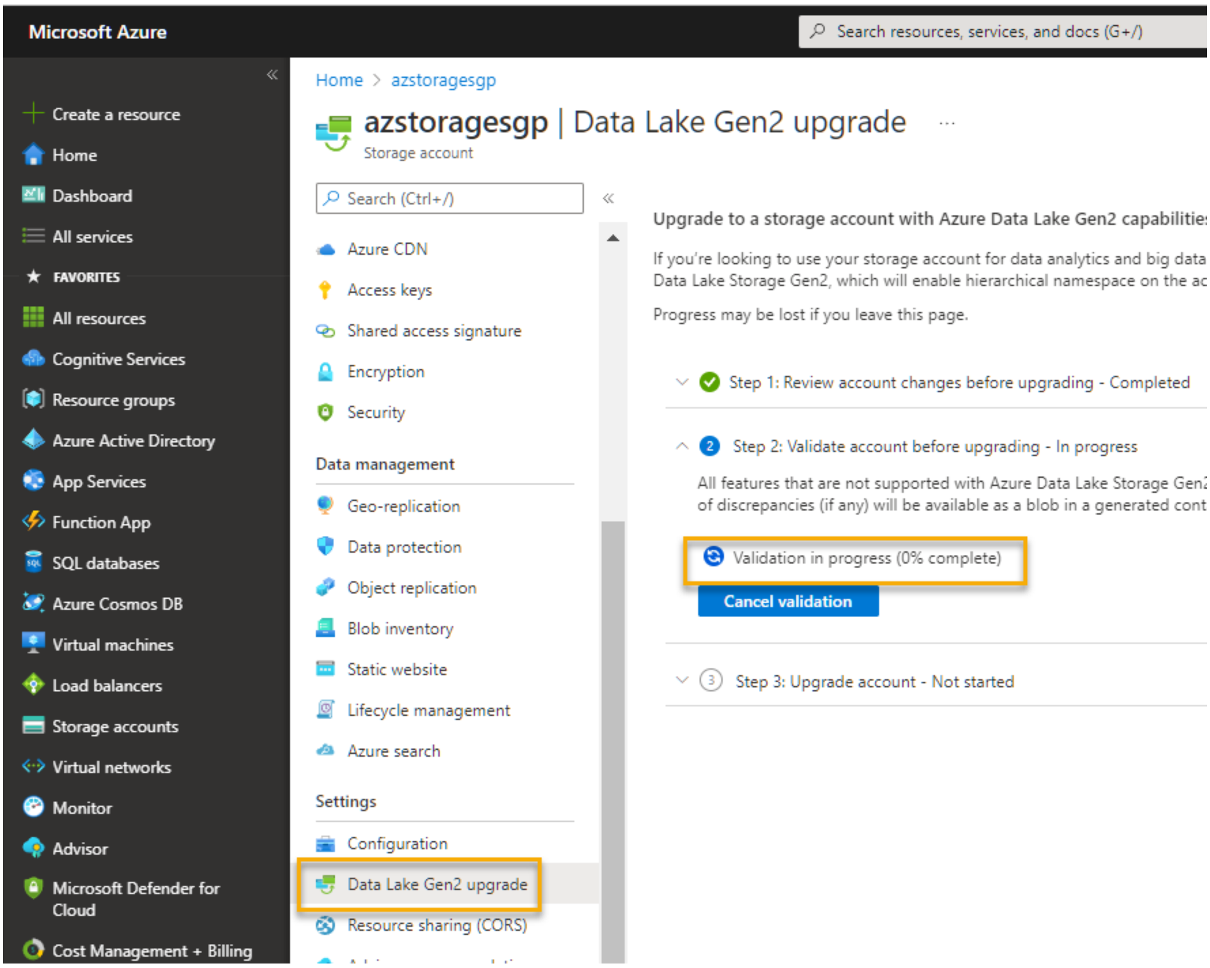
Step 1: Review account changes before upgrading

Step 2: Validate account before upgrading - Not s

All features that are not supported with Azure Data Lak
of discrepancies (if any) will be available as a blob in a

Start validation

Step 3: Upgrade account - Not started



azstoragesgp | Data Lake Gen2 upgrade

Storage account

Upgrade to a storage account with Azure Data Lake Gen2 capabilitie

If you're looking to use your storage account for data analytics and big data Data Lake Storage Gen2, which will enable hierarchical namespace on the ac

Progress may be lost if you leave this page.

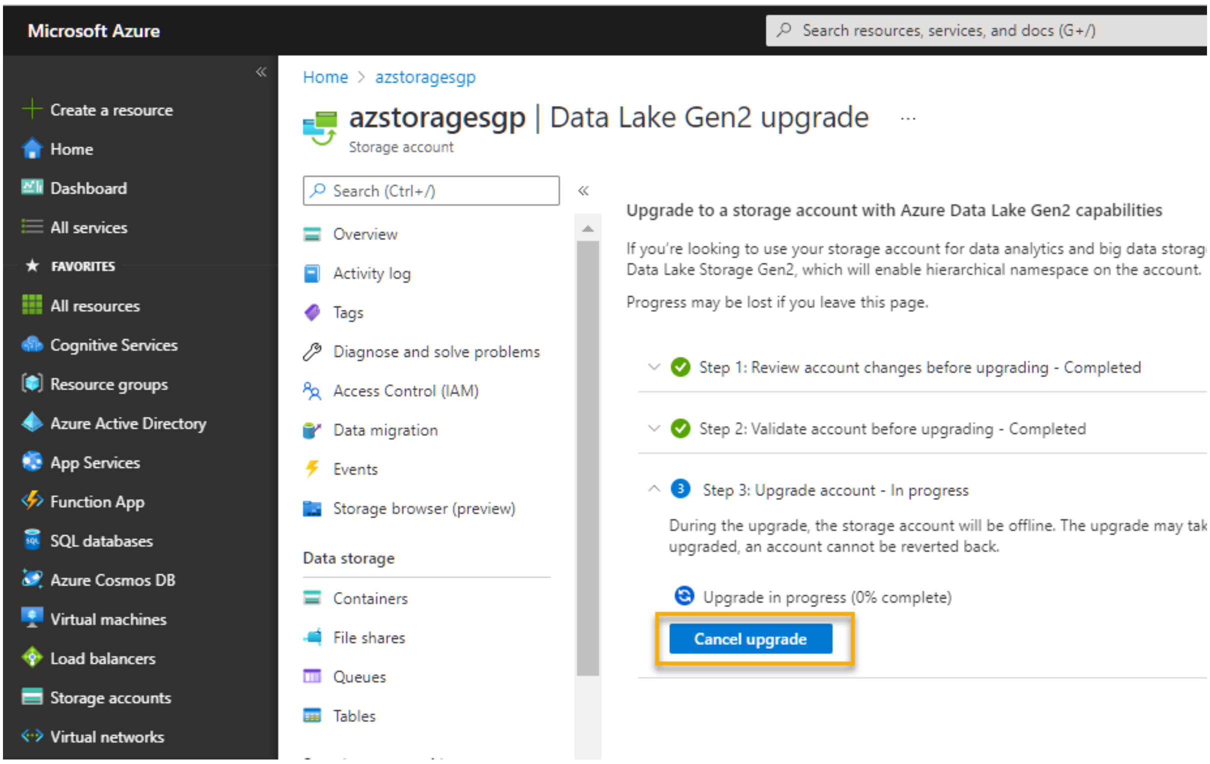
Step 1: Review account changes before upgrading - Completed

Step 2: Validate account before upgrading - In progress

All features that are not supported with Azure Data Lake Storage Gen2 of discrepancies (if any) will be available as a blob in a generated cont

Validation in progress (0% complete)

Step 3: Upgrade account - Not started



azstoragesgp | Data Lake Gen2 upgrade

Storage account

Upgrade to a storage account with Azure Data Lake Gen2 capabilities

If you're looking to use your storage account for data analytics and big data storag Data Lake Storage Gen2, which will enable hierarchical namespace on the account.

Progress may be lost if you leave this page.

Step 1: Review account changes before upgrading - Completed

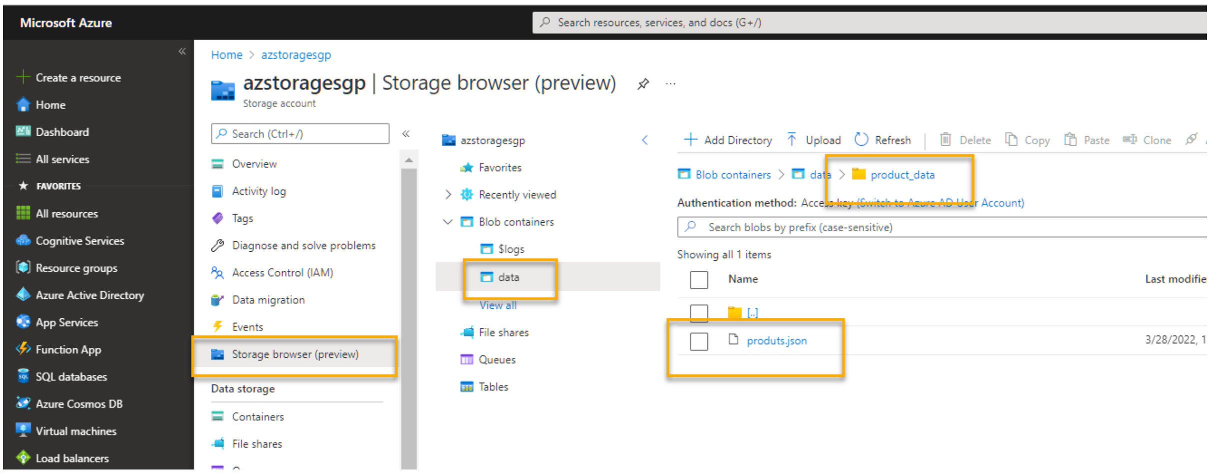
Step 2: Validate account before upgrading - Completed

Step 3: Upgrade account - In progress

During the upgrade, the storage account will be offline. The upgrade may tak upgraded, an account cannot be reverted back.

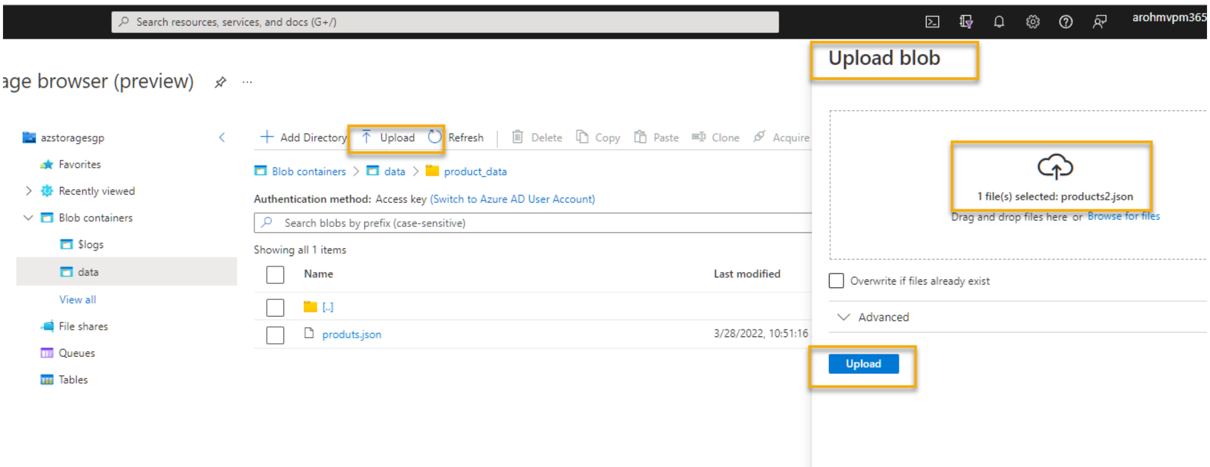
Upgrade in progress (0% complete)

4. When the upgrade is complete, in the pane on the left side, in the top section, select **Storage browser (preview)** and navigate back to the root of your **data** blob container, which still contains the **product_data** folder.



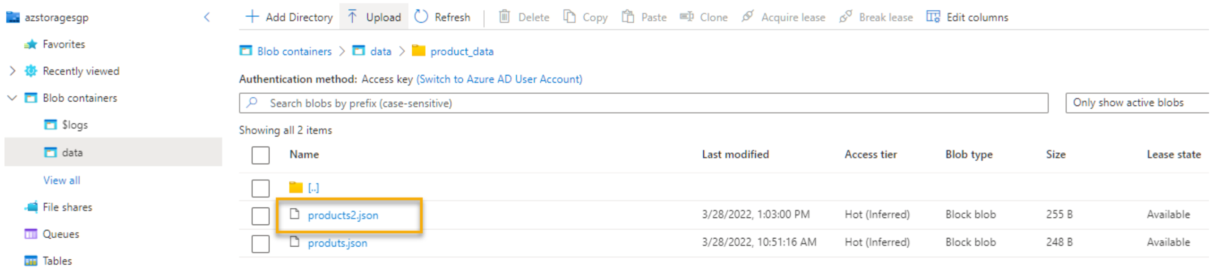
5. Select the **product_data** folder, and verify it still contains the **product1.json** file you uploaded previously.

6. Use the ⤒ **Upload** button to open the **Upload blob** panel.


Search resources, services, and docs (G+/)
arohmvpm365
age browser (preview)
azstoragesgp
Favorites
Recently viewed
Blob containers
$logs
data
View all
File shares
Queues
Tables
Add Directory
Upload
Refresh
Delete
Copy
Paste
Clone
Acquire
Blob containers > data > product_data
Authentication method: Access key (Switch to Azure AD User Account)
Search blobs by prefix (case-sensitive)
Showing all 1 items
Name
Last modified
[..]
produts.json
3/28/2022, 10:51:16
Upload blob
1 file(s) selected: products2.json
Drag and drop files here or Browse for files
Overwrite if files already exist
Advanced
Upload

7. In the **Upload blob** panel, select the **product2.json** file you saved on your local computer. Then select the **Upload** button.

8. Close the **Upload blob** panel if it's still open, and verify that a **product_data** folder now contains the **product2.json** file.



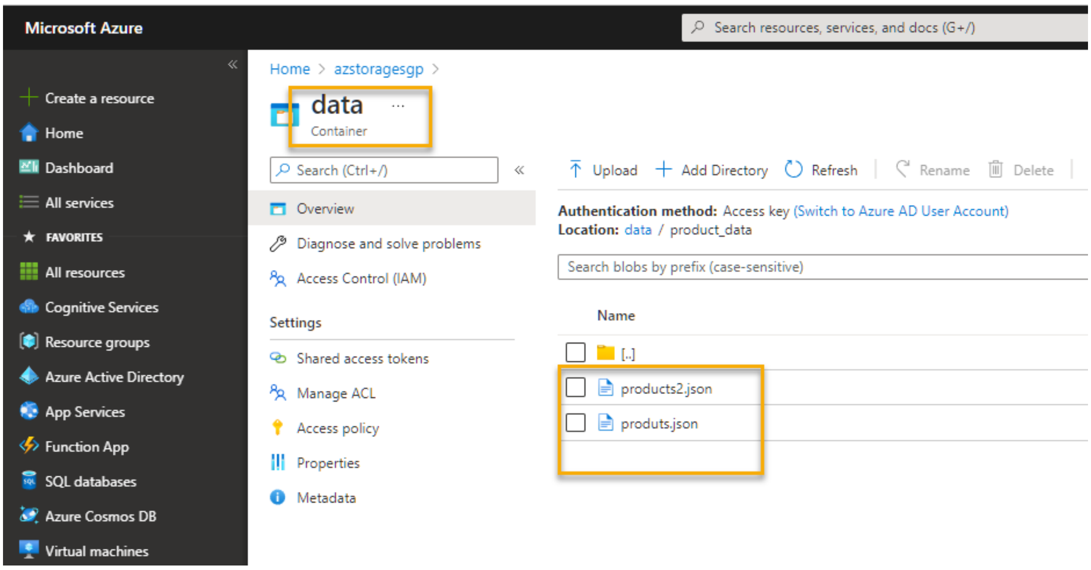
Authentication method: Access key (Switch to Azure AD User Account)

Showing all 2 items

Name	Last modified	Access tier	Blob type	Size	Lease state
[..]					
products2.json	3/28/2022, 1:03:00 PM	Hot (Inferred)	Block blob	255 B	Available
produts.json	3/28/2022, 10:51:16 AM	Hot (Inferred)	Block blob	248 B	Available

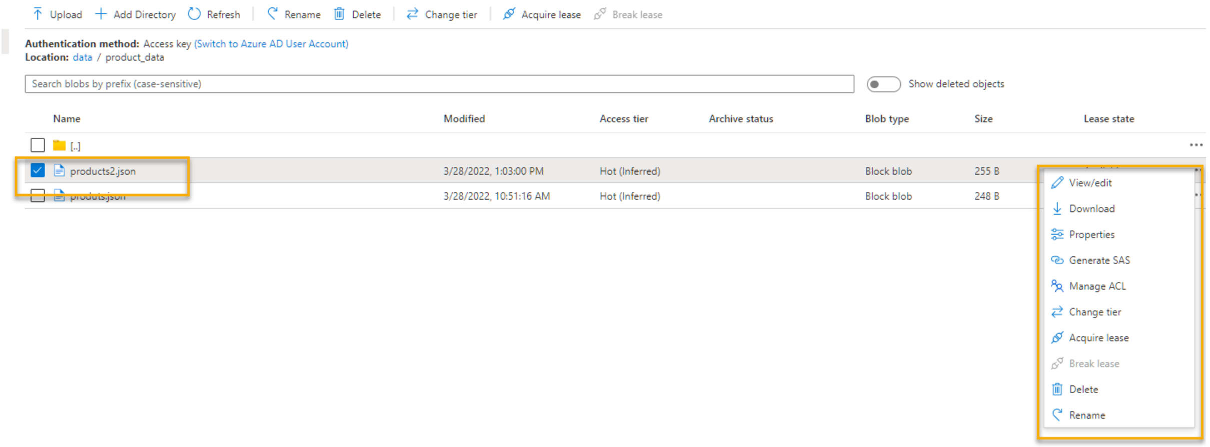
9. On the left side, in the **Data storage** section, select **Containers**.

10. Open the **data** container, and verify that the **product_data** folder you created is listed.



Authentication method: Access key (Switch to Azure AD User Account)
Location: data / product_data

11. Select the ··· icon at the right-end of the folder, and note that with hierarchical namespace enabled, you can perform configuration tasks at the folder-level; including renaming folders and setting permissions.



Authentication method: Access key (Switch to Azure AD User Account)
Location: data / product_data

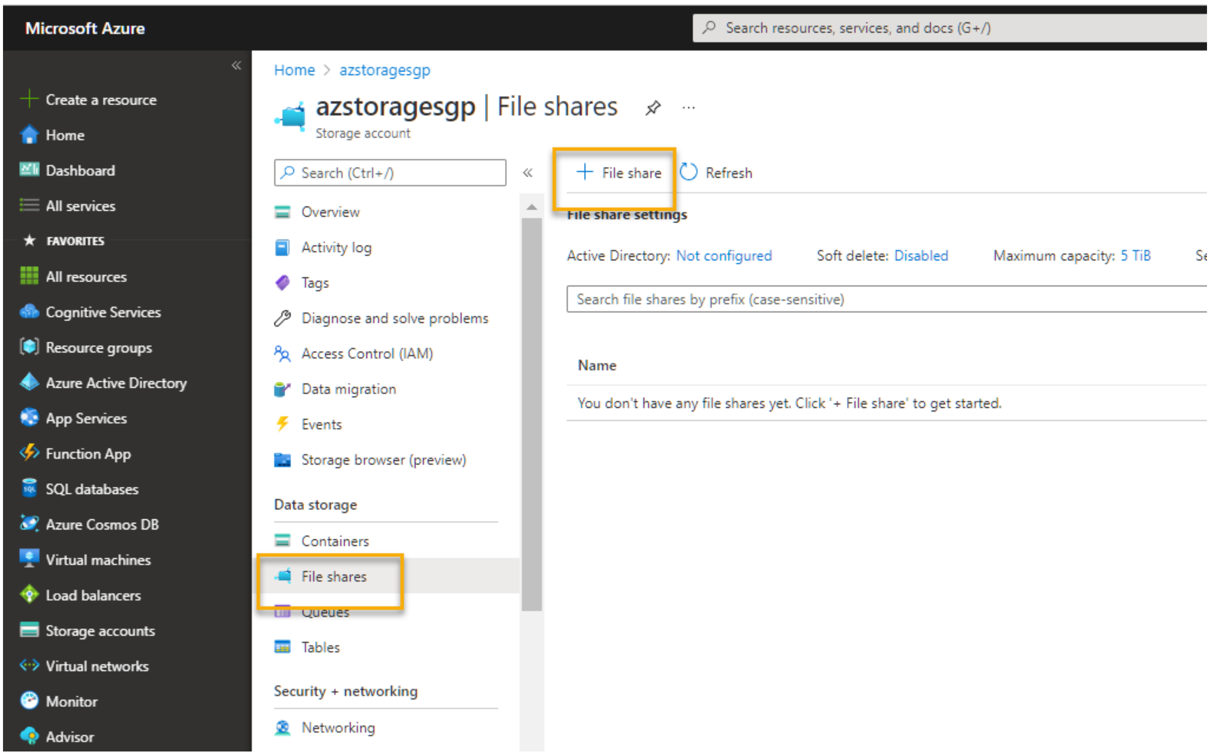
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[..]						
☑ products2.json	3/28/2022, 1:03:00 PM	Hot (Inferred)		Block blob	255 B	
□ produts.json	3/28/2022, 10:51:16 AM	Hot (Inferred)		Block blob	248 B	

12. Use the **X** icon at the top right in the **data** page to close the page and return to the **Containers** page.

Explore Azure Files

Azure Files provides a way to create cloud-based file shares.

1. In the Azure portal page for your storage container, on the left side, in the **Data storage** section, select **File shares**.

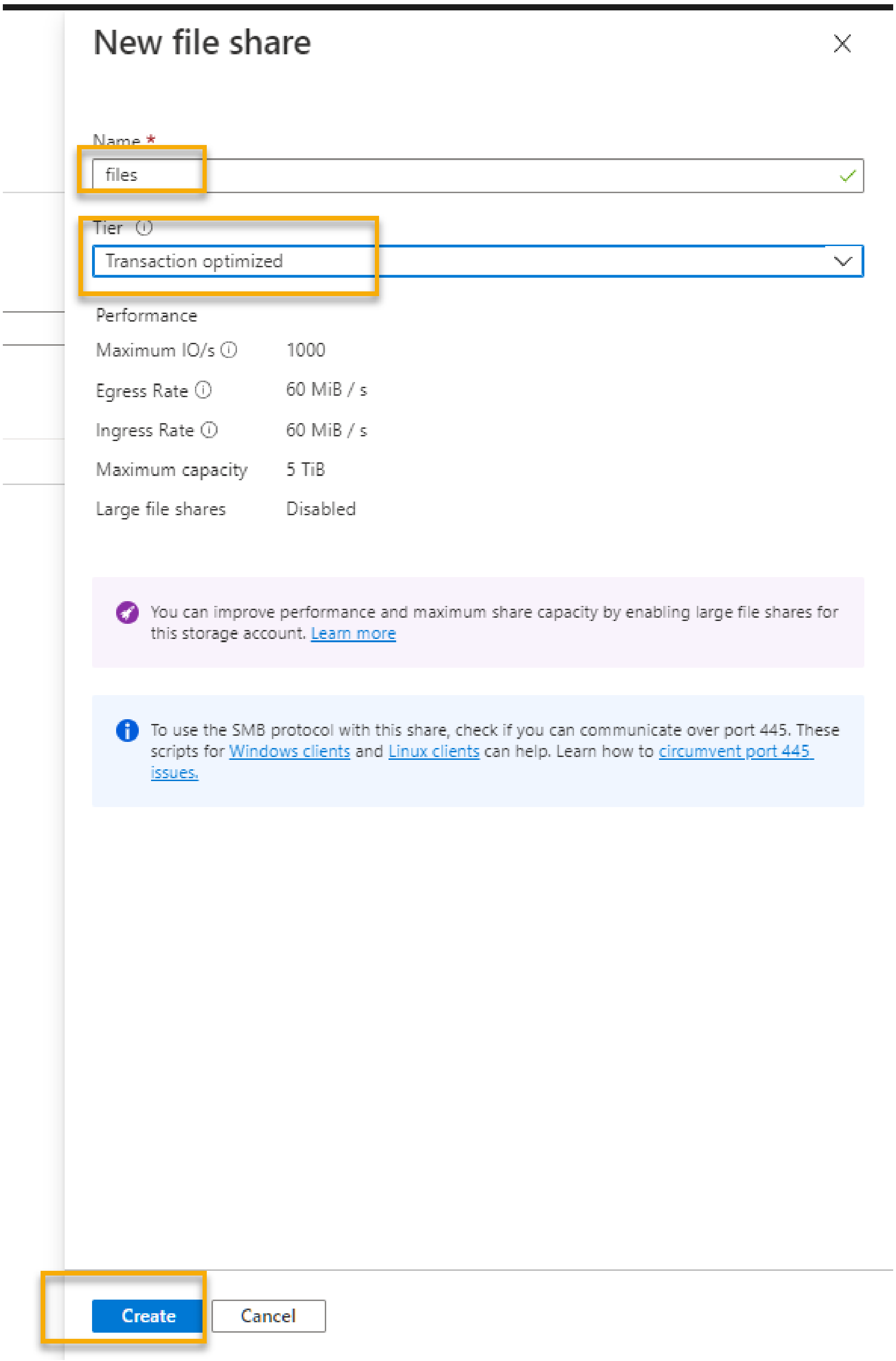


Active Directory: Not configured   Soft delete: Disabled   Maximum capacity: 5 TiB

You don't have any file shares yet. Click '+ File share' to get started.

2. In the File shares page, select **+ File share** and add a new file share named **files** using the **Transaction optimized** tier.

3. In the **File shares**, open your new **files** share.



New file share

Name * files

Tier Transaction optimized

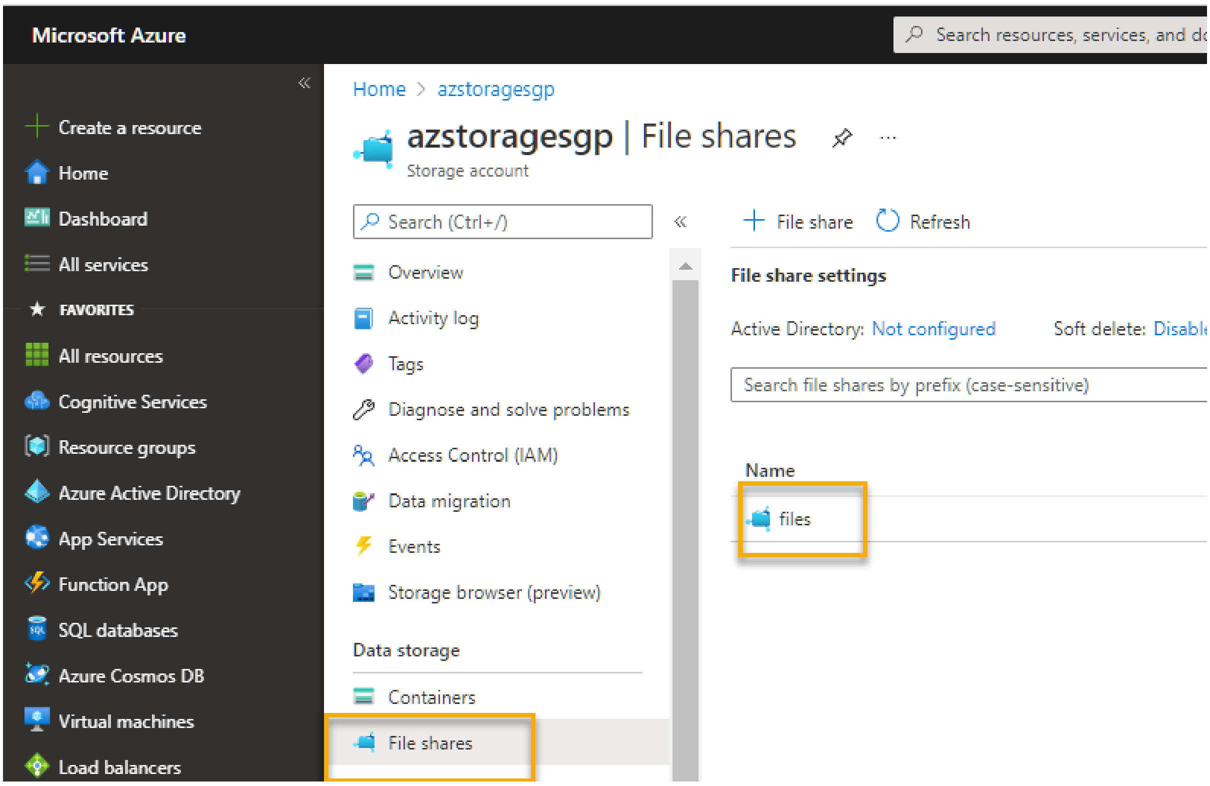
Performance

Maximum IO/s	1000
Egress Rate	60 MiB / s
Ingress Rate	60 MiB / s
Maximum capacity	5 TiB
Large file shares	Disabled

You can improve performance and maximum share capacity by enabling large file shares for this storage account. Learn more

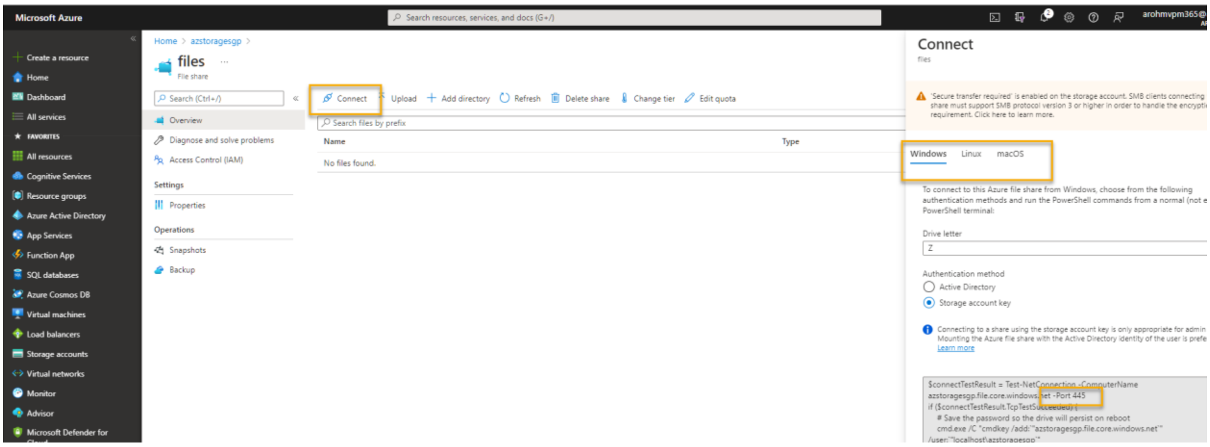
To use the SMB protocol with this share, check if you can communicate over port 445. These scripts for Windows clients and Linux clients can help. Learn how to circumvent port 445 issues.

4. Click on **Files**



Active Directory: Not configured   Soft delete: Disable

5. At the top of the page, select **Connect**. Then in the **Connect** pane, note that there are tabs for common operating systems (Windows, Linux, and macOS) that contain scripts you can run to connect to the shared folder from a client computer.



No files found.

'Secure transfer required' is enabled on the storage account. SMB clients connecting share must support SMB protocol version 3 or higher in order to handle the encrypti requirement. Click here to learn more.

To connect to this Azure file share from Windows, choose from the following authentication methods and run the PowerShell commands from a normal (not e PowerShell terminal:

Drive letter: Z

Authentication method
○ Active Directory
● Storage account key

Connecting to a share using the storage account key is only appropriate for admin Mounting the Azure file share with the Active Directory identity of the user is prefe Learn more

```
$connectTestResult = Test-NetConnection -ComputerName
azstoragesgp.file.core.windows.net -Port 445
if ($connectTestResult.TcpTestSucceeded) {
   # Save the password so the drive will persist on reboot
   cmd.exe /C "cmdkey /add:`"azstoragesgp.file.core.windows.net`"
/user:`"localhost\azstoragesgp`"
```

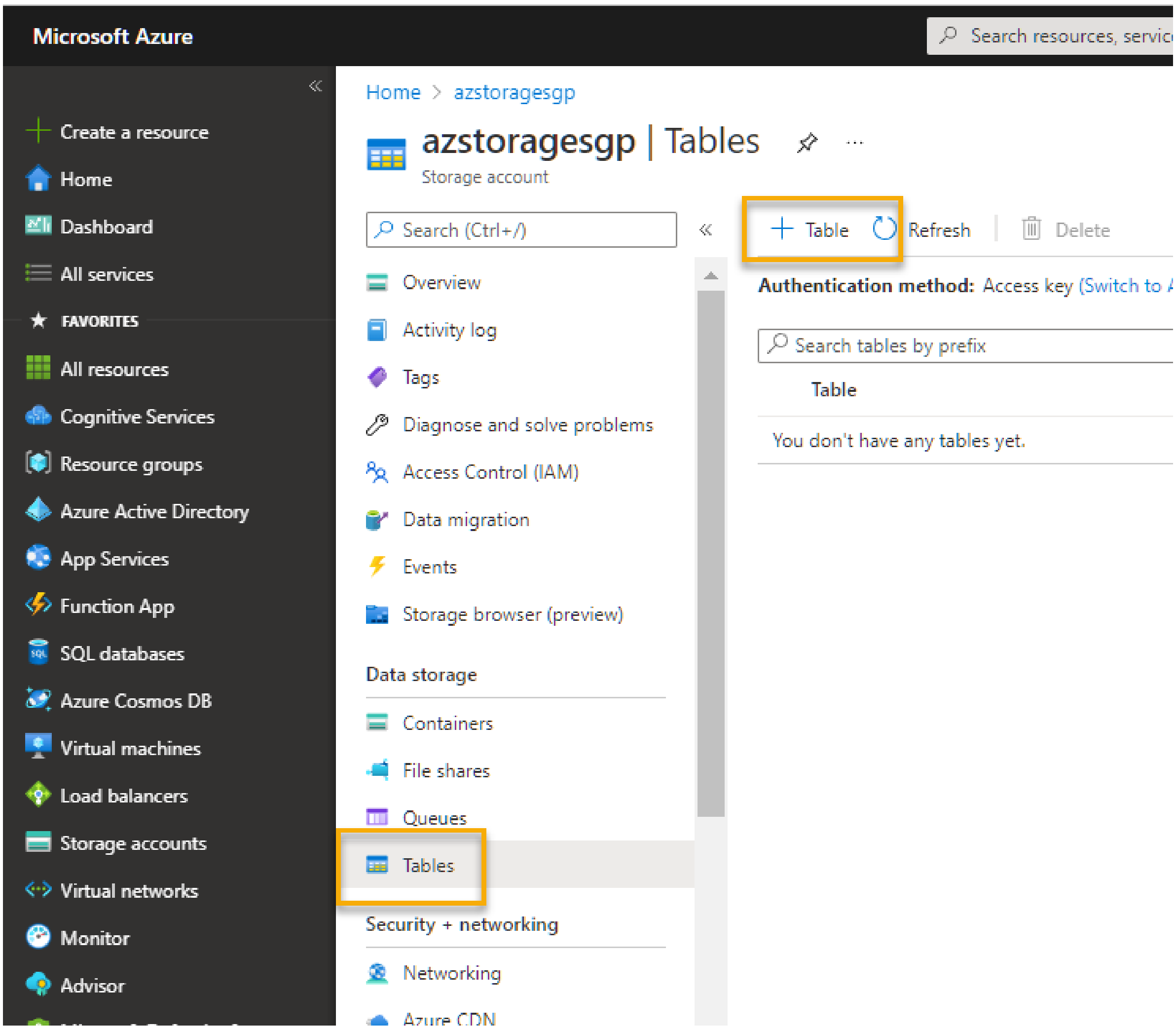
Port 445

6. Close the **Connect** pane and then close the **files** page to return to the **File shares** page for your Azure storage account.

Explore Azure Tables

Azure Tables provide a key/value store for applications that need to store data values, but don't need the full functionality and structure of a relational database.

1. In the Azure portal page for your storage container, on the left side, in the **Data storage** section, select **Tables**.



Authentication method: Access key (Switch to

You don't have any tables yet.

2. On the **Tables** page, select **+ Table** and create a new table named **products**.

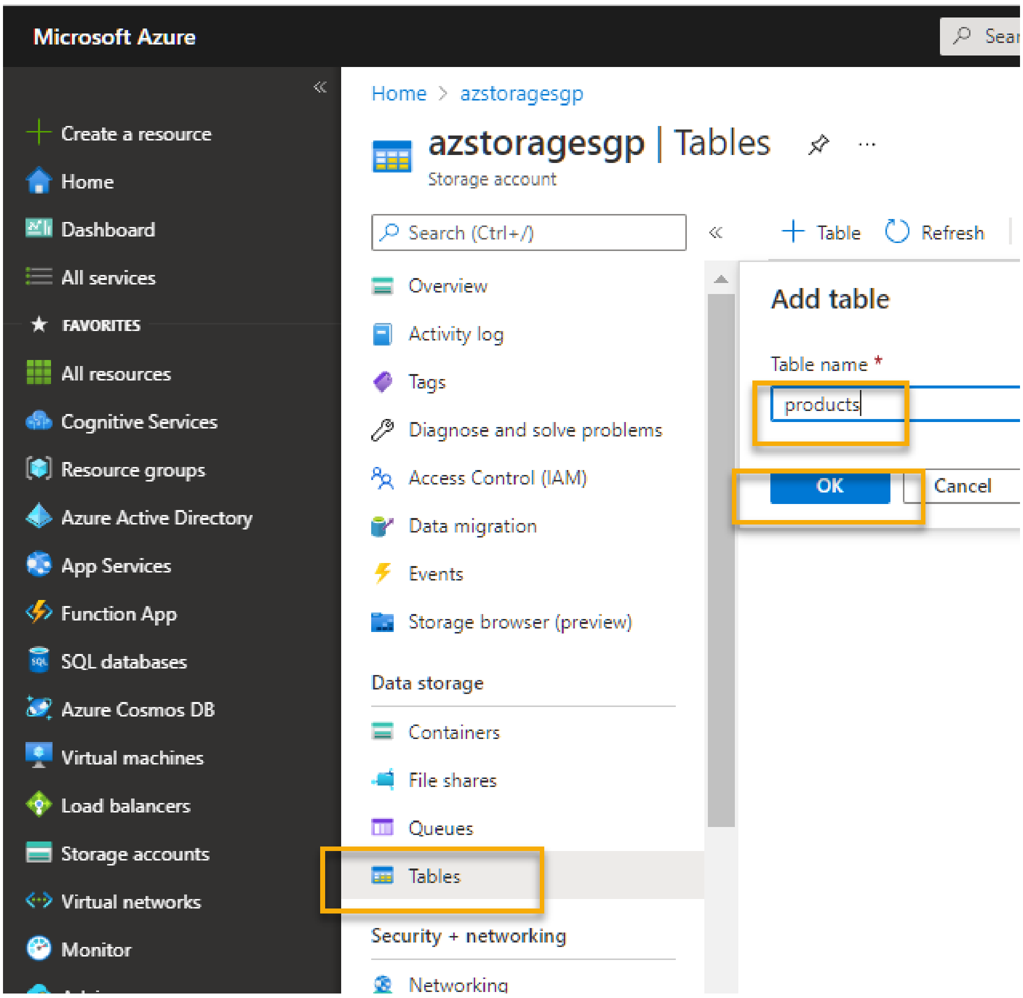
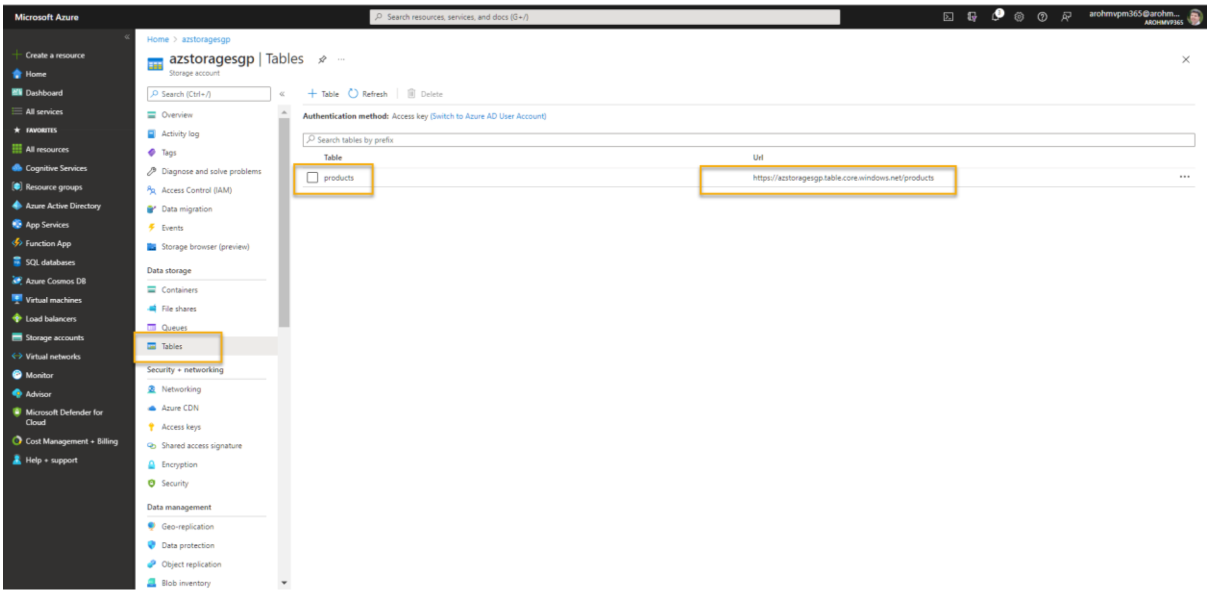


Table name * products

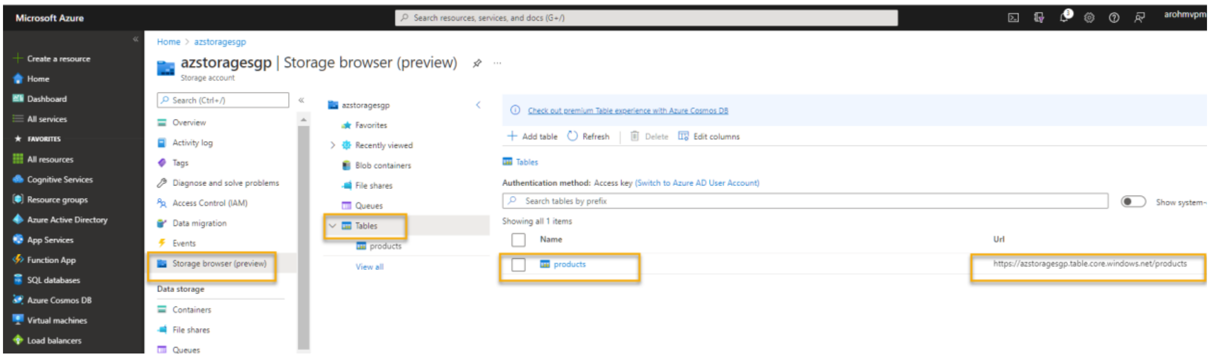


Authentication method: Access key (Switch to Azure AD User Account)

Table	Url
□ products	https://azstoragesgp.table.core.windows.net/products

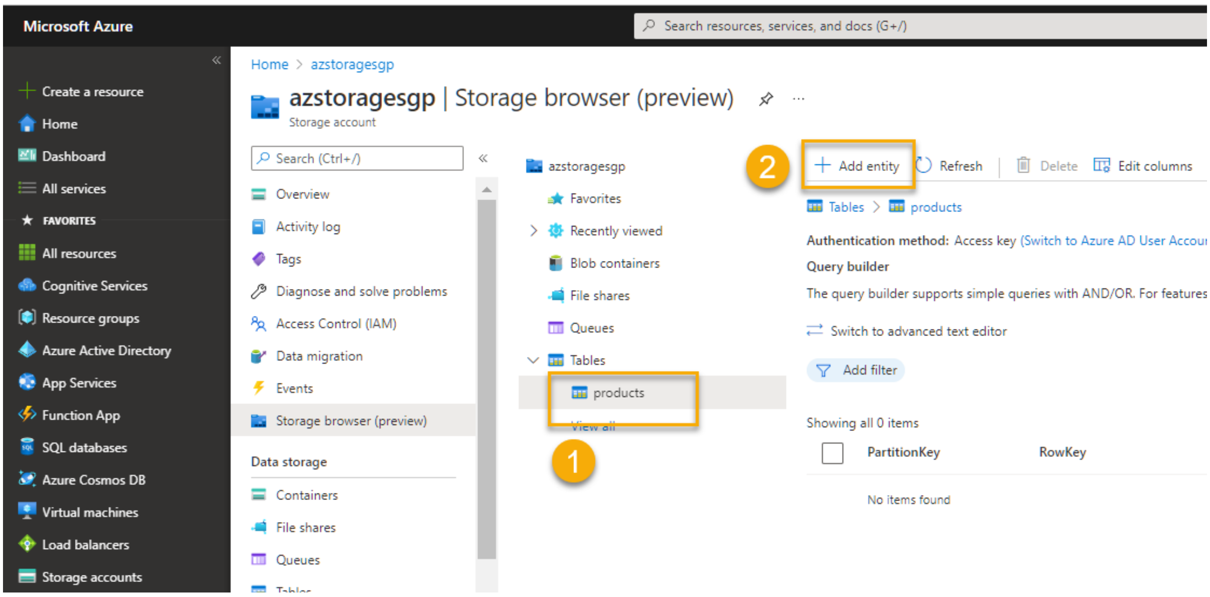
3. After the **products** table has been created, in the pane on the left side, in the top section, select **Storage browser (preview)**.

4. In storage explorer, select **Tables** and verify that the **products** table is listed.



5. Select the **products** table.

6. In the **product** page, select **+ Add entity**.



7. In the **Add entity** panel, enter the following key values:
   - **PartitionKey**: 1
   - **RowKey**: 1

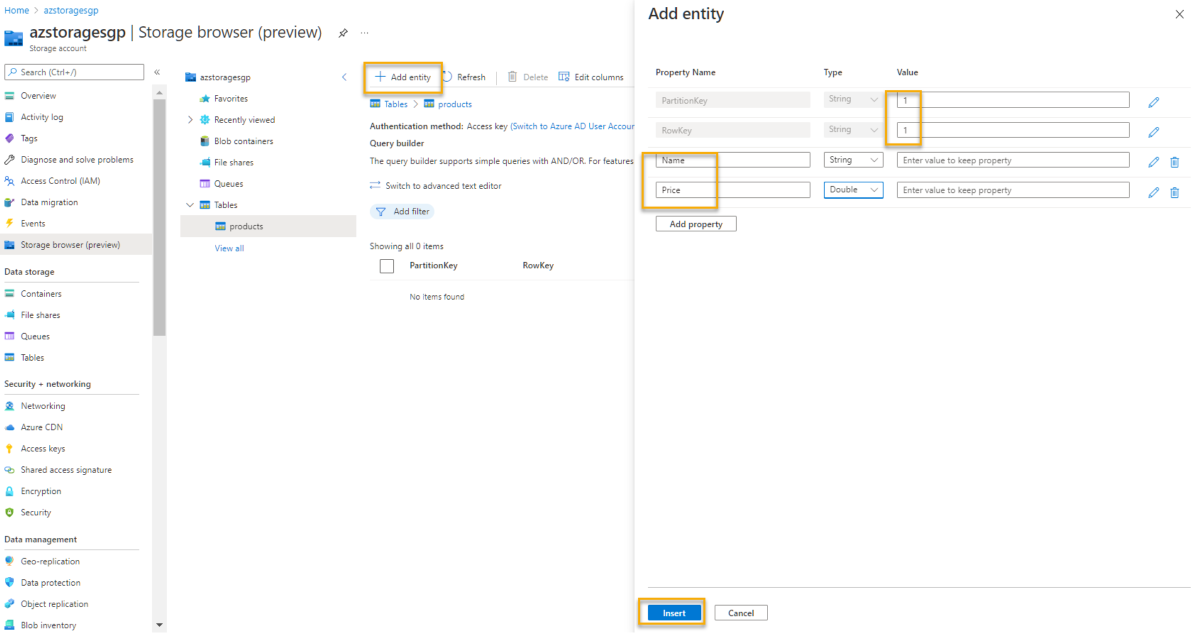
8. Select **Add property**, and create a new property with the following values:

Property name	**Type**
Name	String

9. Add a second property with the following values:

Property name	**Type**
Price	Double

10. Select **Insert** to insert a row for the new entity into the table.



Authentication method: Access key (Switch to Azure AD User Accour

Query builder

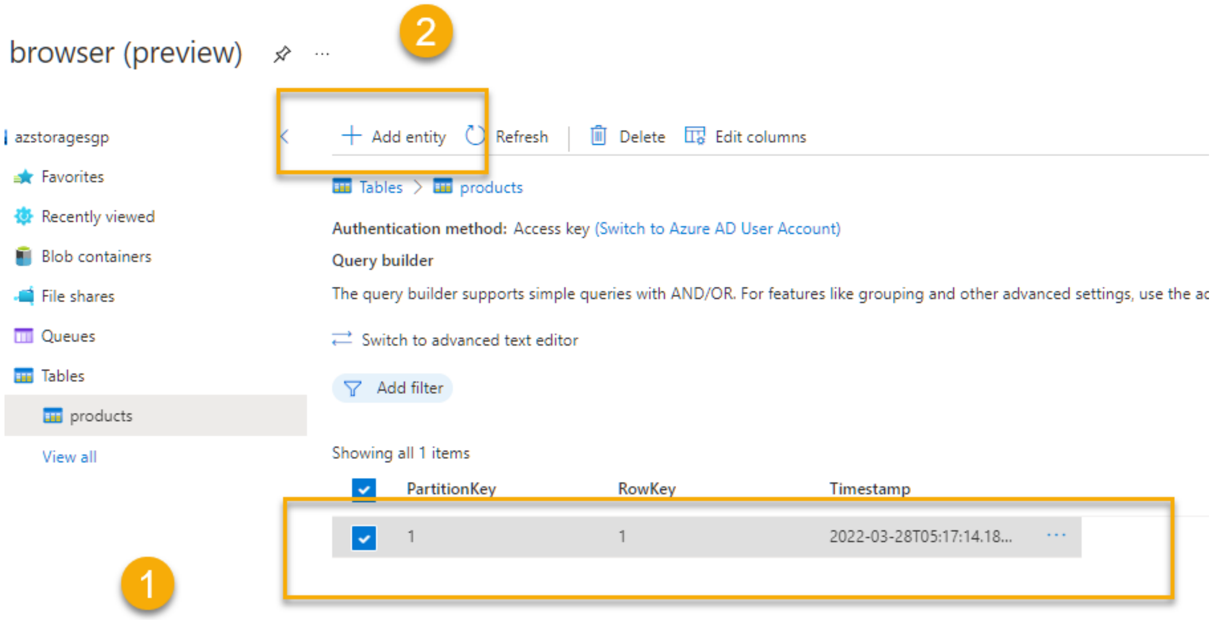
The query builder supports simple queries with AND/OR. For features

Showing all 0 items

PartitionKey	RowKey
No items found	

Property Name	Type	Value
PartitionKey	String	1
RowKey	String	1
Name	String	Enter value to keep property
Price	Double	Enter value to keep property

11. In storage browser, verify that a row has been added to the **products** table, and that a **Timestamp** column has been created to indicate when the row was last modified.



Authentication method: Access key (Switch to Azure AD User Account)

Query builder

The query builder supports simple queries with AND/OR. For features like grouping and other advanced settings, use the ac

Showing all 1 items

	PartitionKey	RowKey	Timestamp
☑	1	1	2022-03-28T05:17:14.18...

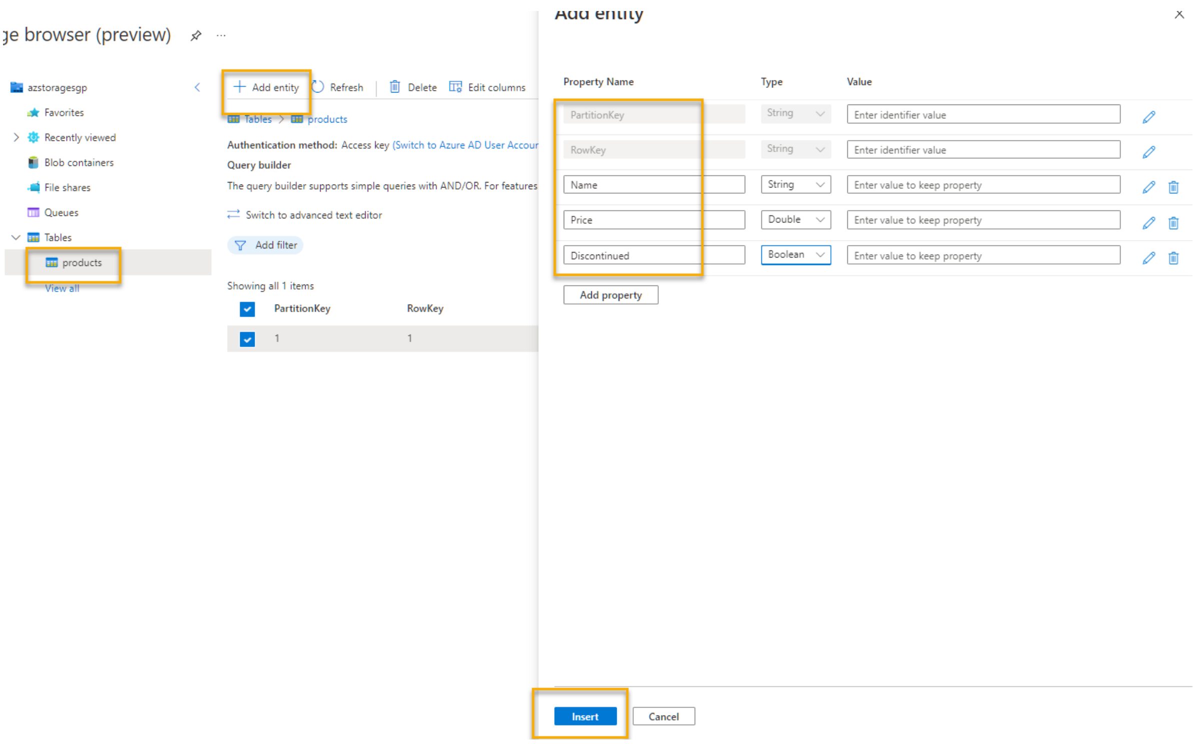
12. Add another entity to the **products** table with the following properties:

Property name	**Type**
PartitionKey	String
RowKey	String

Property name	**Type**
Name	String
Price	Double
Discontinued	Boolean

13. After inserting the new entity, verify that a row containing the discontinued product is shown in the table.

14.



Authentication method: Access key (Switch to Azure AD User Accour

Query builder

The query builder supports simple queries with AND/OR. For features

Showing all 1 items

☑	PartitionKey	RowKey
☑	1	1

Add entity

Property Name	Type	Value
PartitionKey	String	Enter identifier value
RowKey	String	Enter identifier value
Name	String	Enter value to keep property
Price	Double	Enter value to keep property
Discontinued	Boolean	Enter value to keep property

You have manually entered data into the table using the storage browser interface. In a real scenario, application developers can use the Azure Storage Table API to build applications that read and write values to tables, making it a cost effective and scalable solution for NoSQL storage.

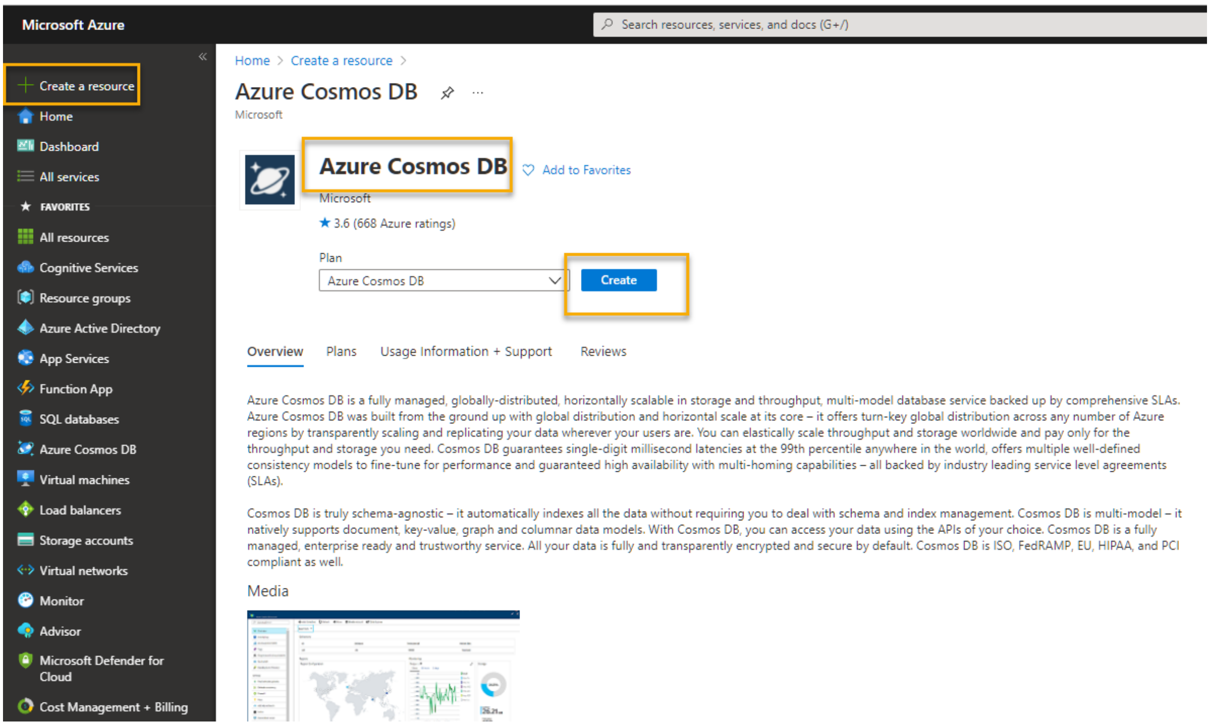
Exercise: Explore Azure Cosmos DB

https://docs.microsoft.com/en-gb/learn/modules/explore-non-relational-data-stores-azure/4-exercise-explore-cosmos-db#create-a-cosmos-db-account

Create a Cosmos DB account

To use Cosmos DB, you must provision a Cosmos DB account in your Azure subscription. In this exercise, you'll provision a Cosmos DB account that uses the core (SQL) API.

1. In the Azure portal, select **+ Create a resource** at the top left, and search for *Azure Cosmos DB*. In the results, select **Azure Cosmos DB** and select **Create**.



Azure Cosmos DB

Microsoft

Azure Cosmos DB

Microsoft

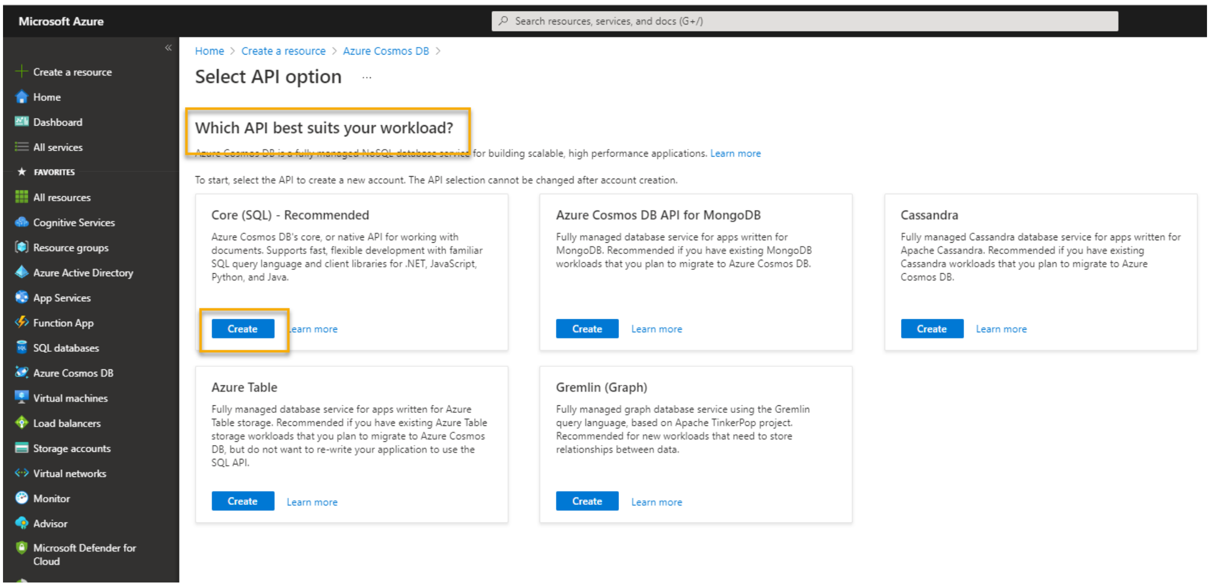
★ 3.6 (668 Azure ratings)

Plan: Azure Cosmos DB

Azure Cosmos DB is a fully managed, globally-distributed, horizontally scalable in storage and throughput, multi-model database service backed up by comprehensive SLAs. Azure Cosmos DB was built from the ground up with global distribution and horizontal scale at its core – it offers turn-key global distribution across any number of Azure regions by transparently scaling and replicating your data wherever your users are. You can elastically scale throughput and storage worldwide and pay only for the throughput and storage you need. Cosmos DB guarantees single-digit millisecond latencies at the 99th percentile anywhere in the world, offers multiple well-defined consistency models to fine-tune for performance and guaranteed high availability with multi-homing capabilities – all backed by industry leading service level agreements (SLAs).

Cosmos DB is truly schema-agnostic – it automatically indexes all the data without requiring you to deal with schema and index management. Cosmos DB is multi-model – it natively supports document, key-value, graph and columnar data models. With Cosmos DB, you can access your data using the APIs of your choice. Cosmos DB is a fully managed, enterprise ready and trustworthy service. All your data is fully and transparently encrypted and secure by default. Cosmos DB is ISO, FedRAMP, EU, HIPAA, and PCI compliant as well.

2. In the **Core (SQL) - Recommended** tile, select **Create**.



Select API option

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL 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.

Core (SQL) - Recommended

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.

Azure Cosmos DB API 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.

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.

Azure 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, but do not want to re-write your application to use the SQL API.

Gremlin (Graph)

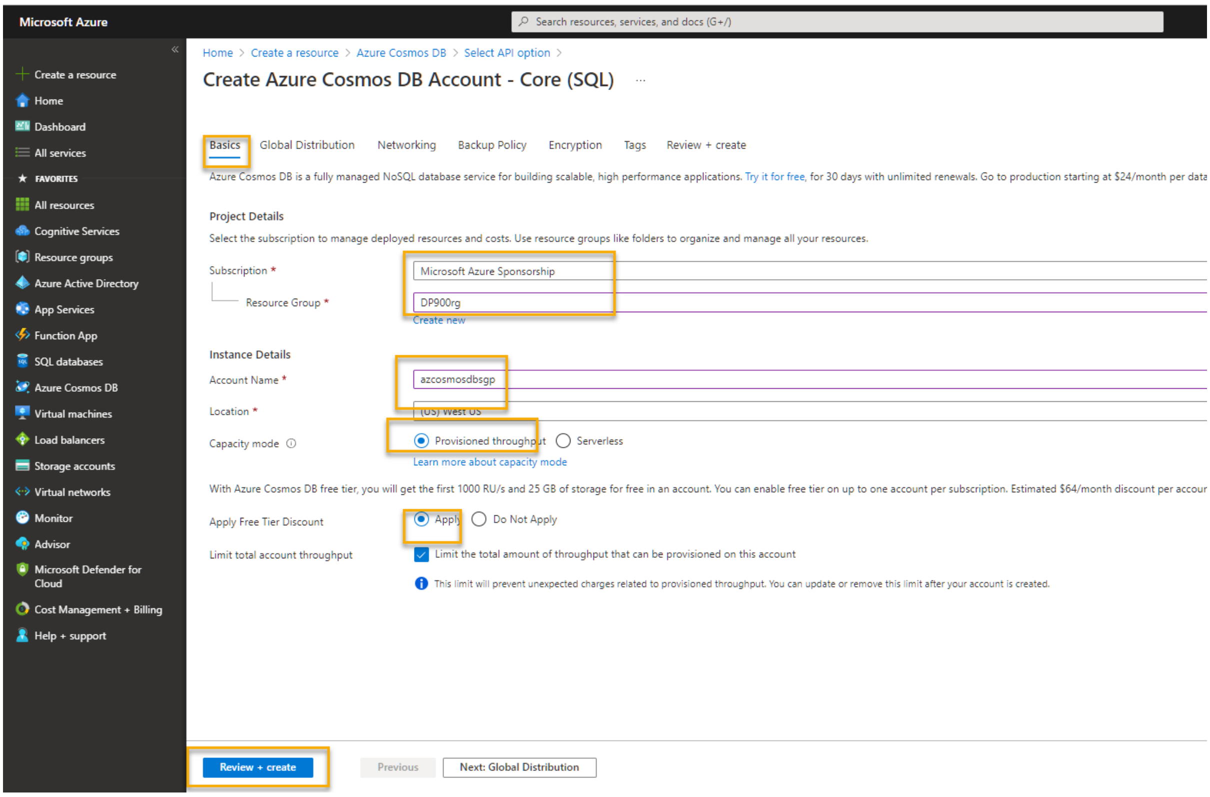
Fully managed graph database service using the Gremlin query language, based on Apache TinkerPop project. Recommended for new workloads that need to store relationships between data.

3. Enter the following details, and then select **Review + Create**:
   - **Subscription**: If you're using a sandbox, select *Concierge Subscription*. Otherwise, select your Azure subscription.
   - **Resource group**: If you're using a sandbox, select the existing resource group (which will have a name like *learn-xxxx...*). Otherwise, create a new resource group with a name of your choice.
   - **Account Name**: Enter a unique name
   - **Location**: Choose any available location

- **Capacity mode**: Provisioned throughput
- **Apply Free-Tier Discount**: Select Apply if available
- **Limit total account throughput**: Unselected

4. When the configuration has been validated, select **Create**.

5. Wait for deployment to complete. Then go to the deployed resource.



Create Azure Cosmos DB Account - Core (SQL) ...

Azure Cosmos DB is a fully managed NoSQL 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 data

Project Details

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

Subscription * Microsoft Azure Sponsorship

Resource Group * DP900rg

Create new

Instance Details

Account Name * azcosmosdbsgp

Location * (US) West US

Capacity mode ⓘ ● Provisioned throughput ○ Serverless

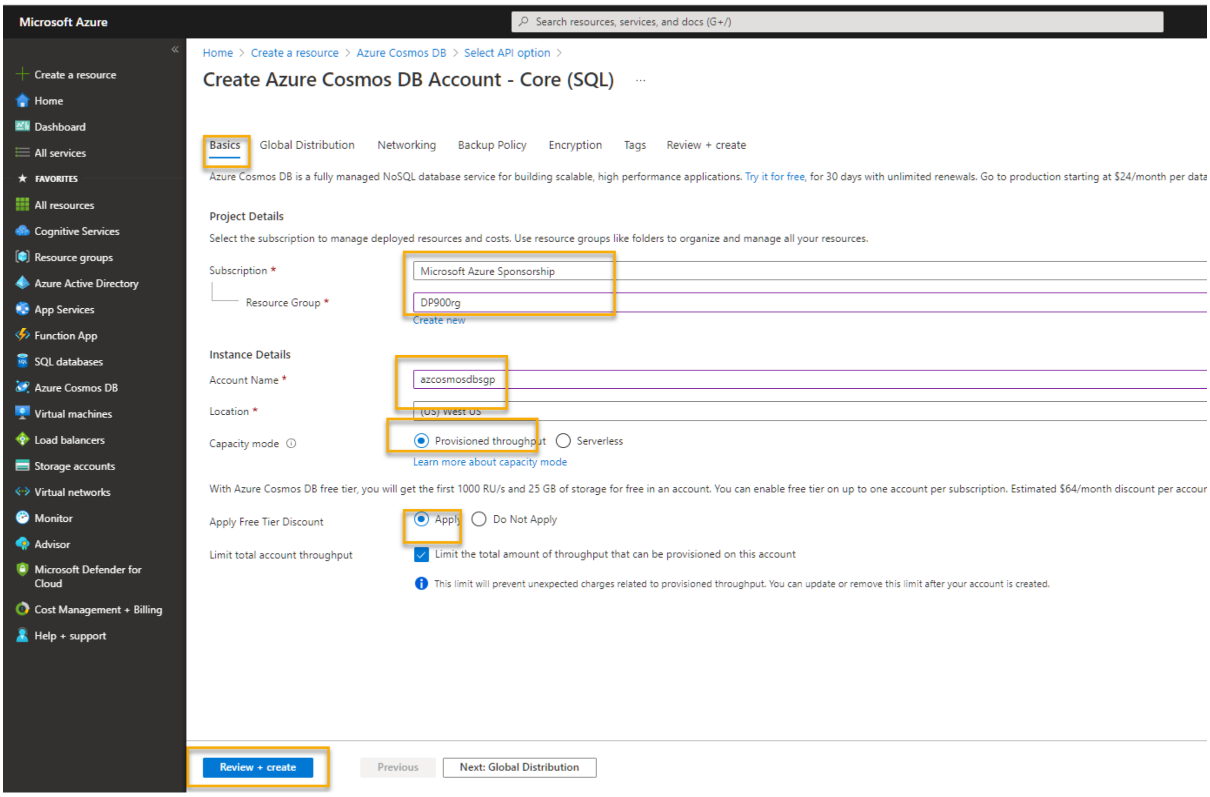
Learn more about capacity mode

With Azure Cosmos DB free tier, you will get the first 1000 RU/s 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 accour

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.



Create Azure Cosmos DB Account - Core (SQL)

Azure Cosmos DB is a fully managed NoSQL 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 data

Project Details

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

Subscription * Microsoft Azure Sponsorship

Resource Group * DP900rg

Create new

Instance Details

Account Name * azcosmosdbsgp

Location * (US) West US

Capacity mode ● Provisioned throughput ○ Serverless

Learn more about capacity mode

With Azure Cosmos DB free tier, you will get the first 1000 RU/s 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 accour

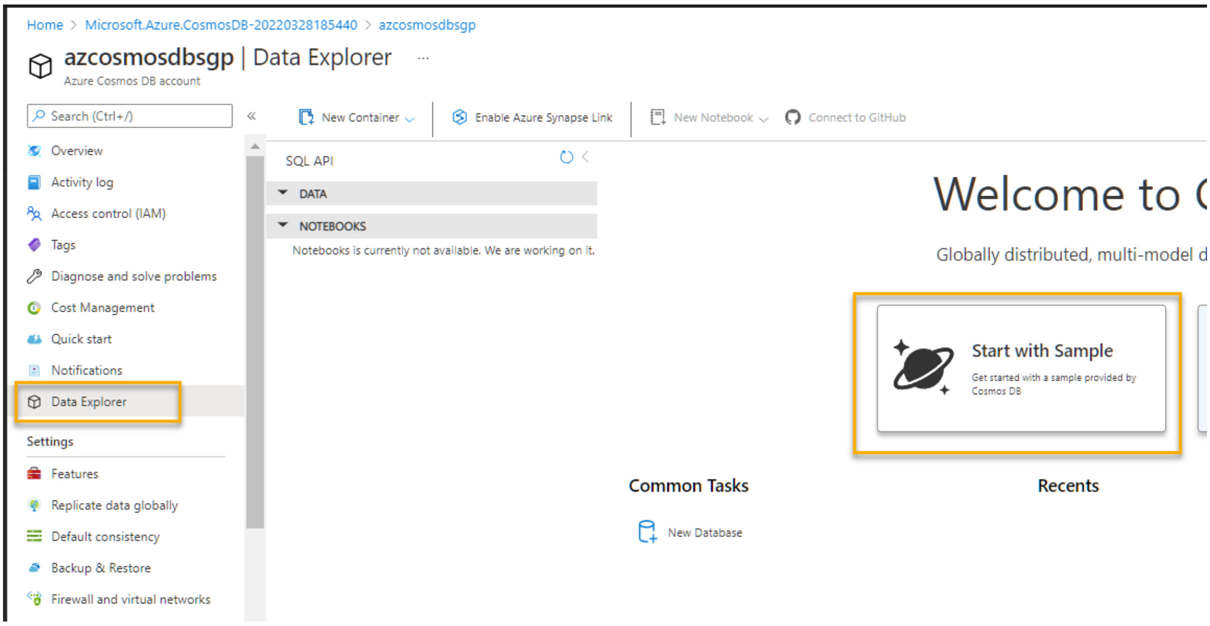
Apply Free Tier Discount ● Apply ○ Do Not Apply

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.

Create a sample database

1. On the page for your new Cosmos DB account, in the pane n the left, select **Data Explorer**.

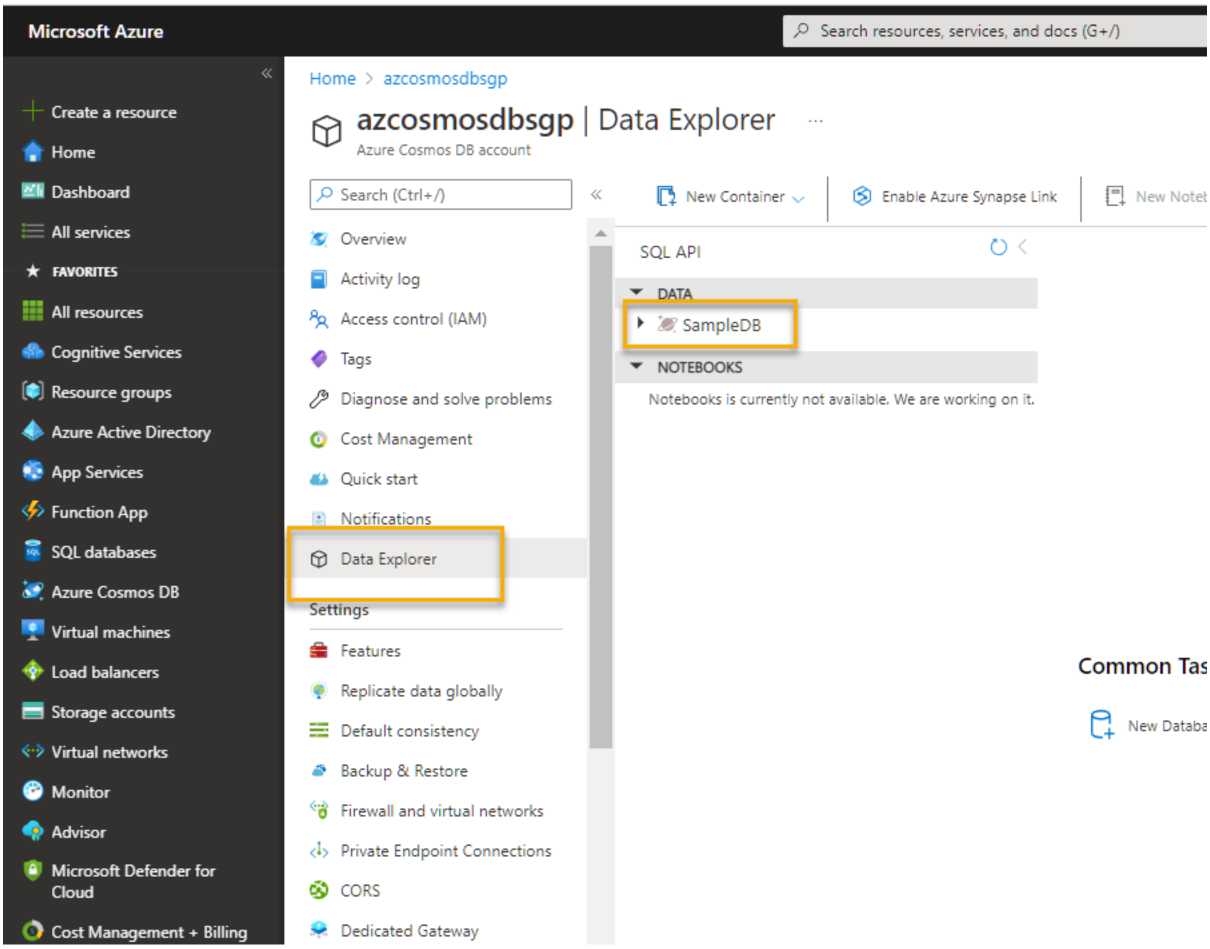


azcosmosdbsgp | Data Explorer

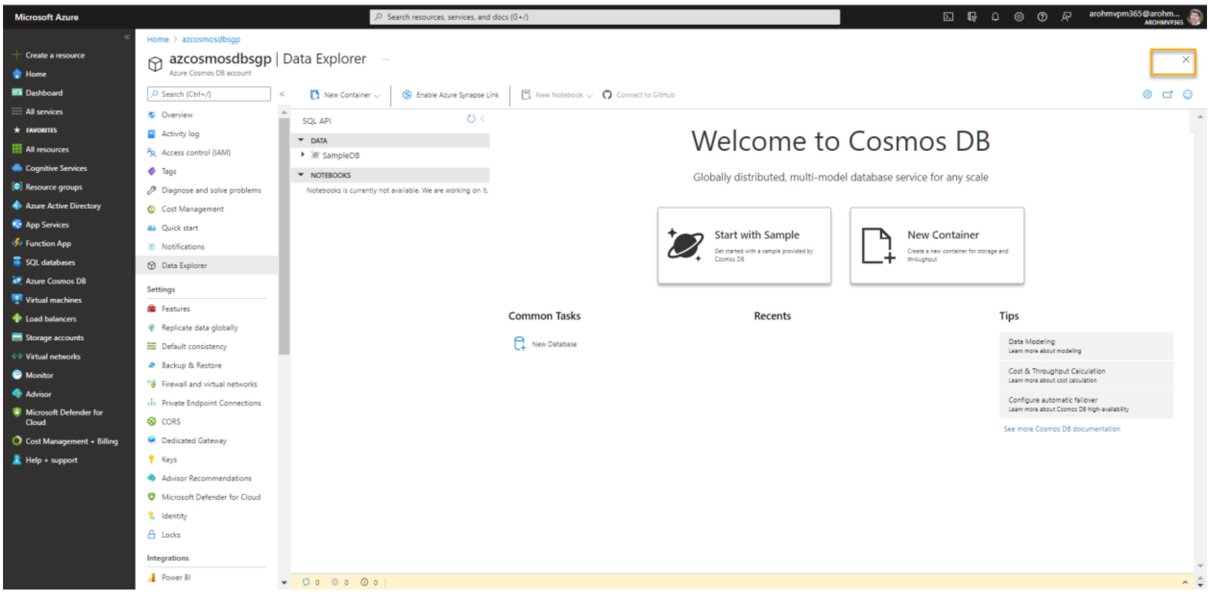
Azure Cosmos DB account

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

2. In the **Data Explorer** page, select **Start with Sample** and then observe the status in the panel at the bottom of the screen until the **SampleDB** database and the sample **Persons** container has been created (which may take a minute or so).



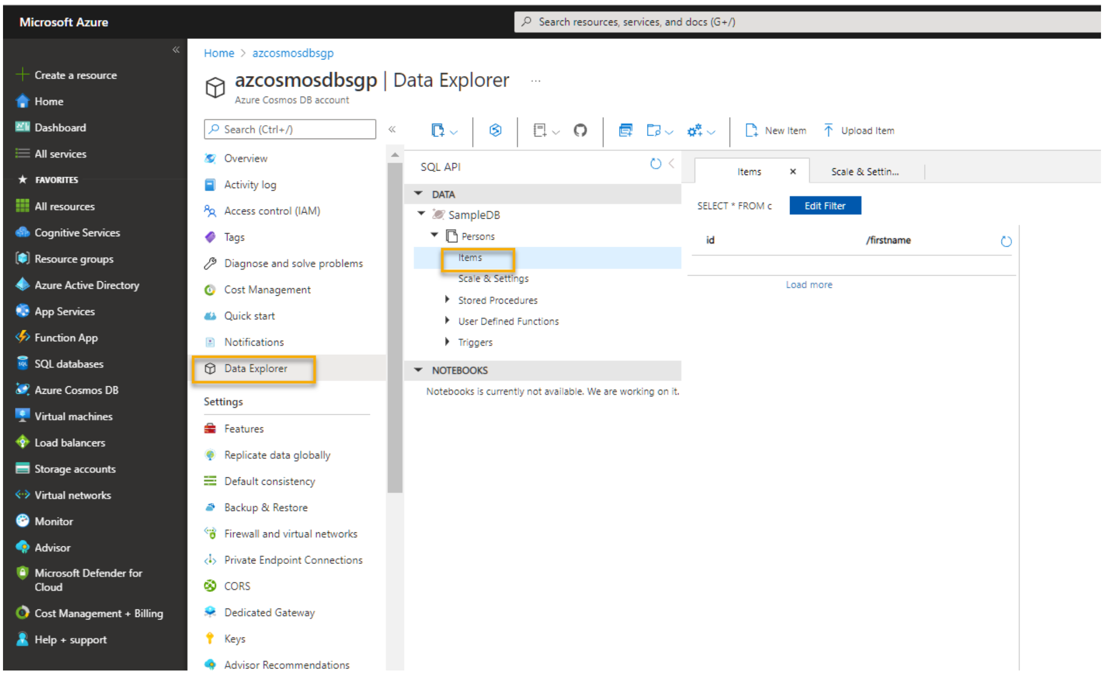
3. Select **Close** on the notification message.


Microsoft Azure
Search resources, services, and docs (G+/)
Home > azcosmosdbsgp
azcosmosdbsgp | Data Explorer
Azure Cosmos DB account
New Container
Enable Azure Synapse Link
New Notebook
Connect to GitHub
SQL API
DATA
SampleDB
NOTEBOOKS
Notebooks is currently not available. We are working on it.
Welcome to Cosmos DB
Globally distributed, multi-model database service for any scale
Start with Sample
Get started with a sample provided by Cosmos DB
New Container
Create a new container for storage and throughput
Common Tasks
New Database
Recents
Tips
Data Modeling
Learn more about modeling
Cost & Throughput Calculation
Learn more about cost calculation
Configure automatic failover
Learn more about Cosmos DB high-availability
See more Cosmos DB documentation

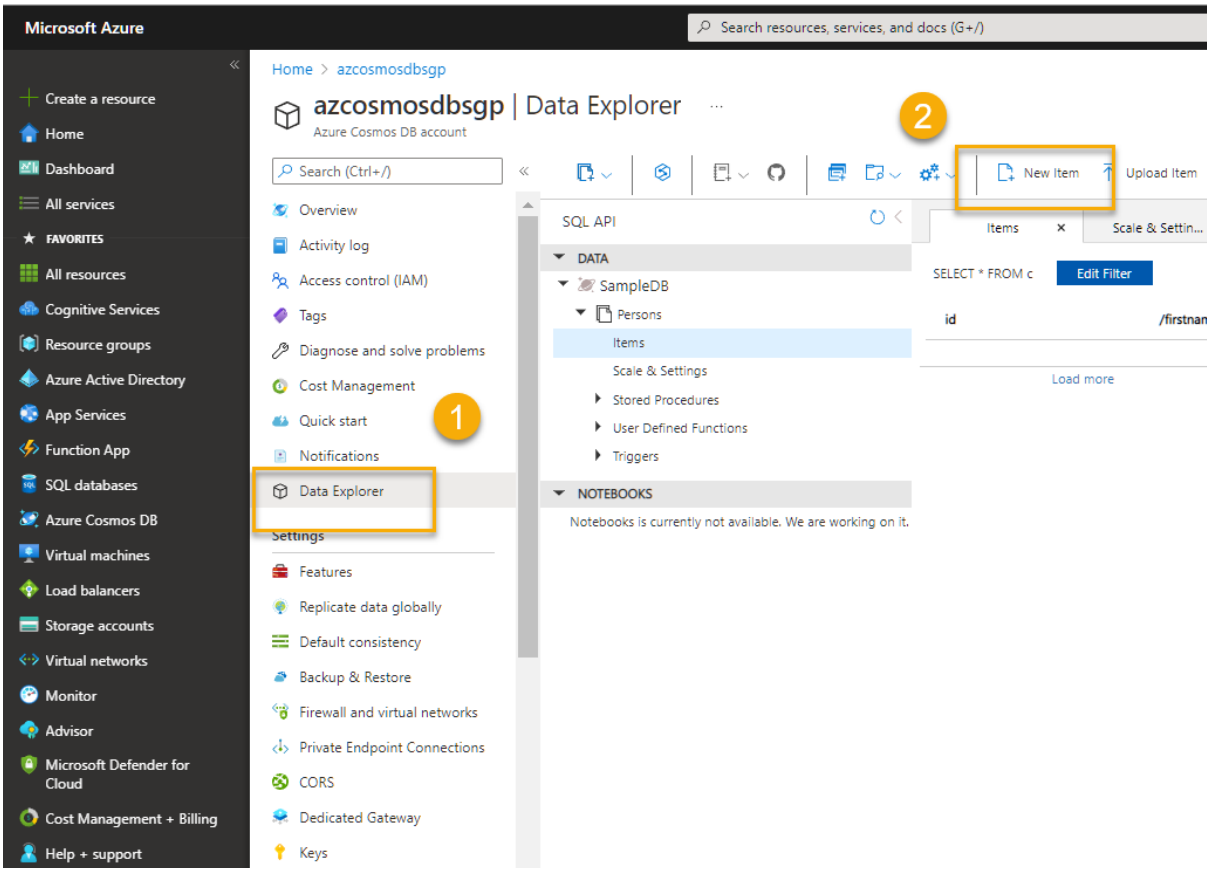
View and create items

1. In the Data Explorer page, expand the **SampleDB** database and the Persons container, and select **Items** to see a list of items in the container. The items represent people, each with a unique id, a firstname, an age, and other properties.

2. Select any of the items in the list to see a JSON representation of the item data.



3. At the top of the page, select **New Item** to create a new blank item.

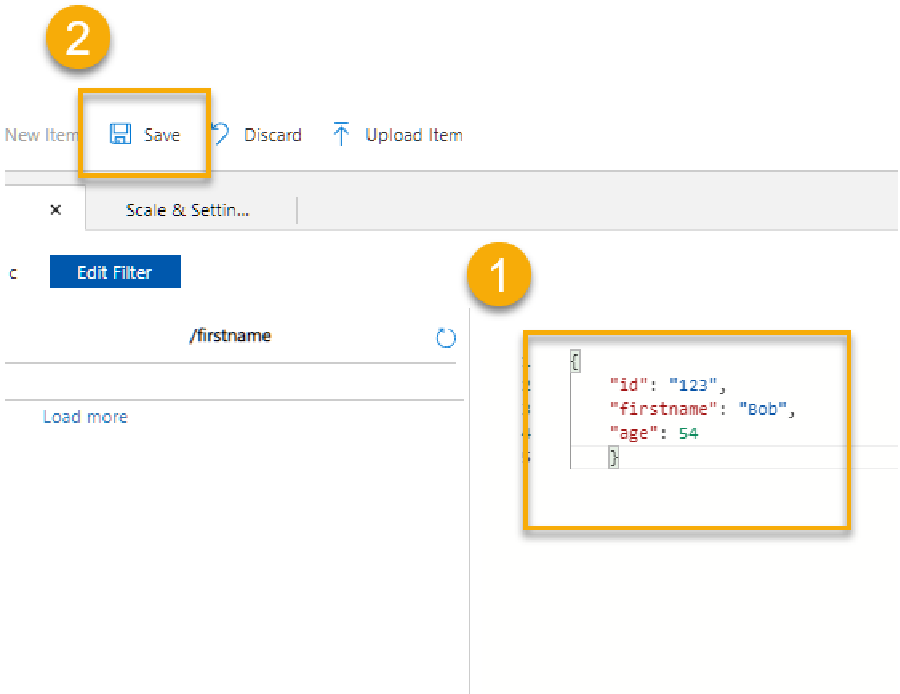


4. Modify the JSON for the new item as follows, and then select **Save**.

JSONCopy

```
{
    "id": "123",
    "firstname": "Bob",
    "age": 54
}
```

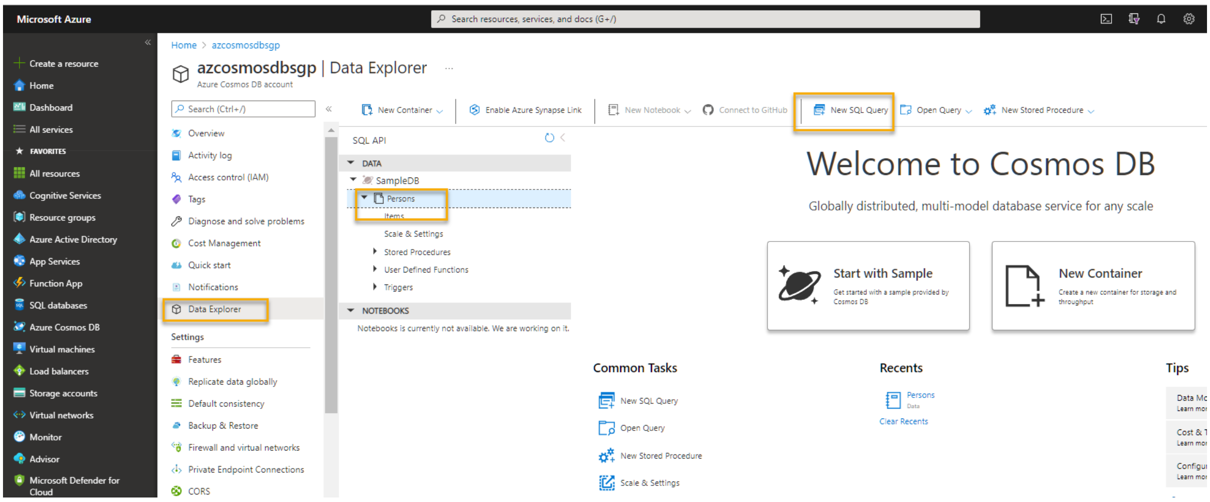
5. After **saving** the new item, notice that additional metadata properties are added automatically.



```
{
    "id": "123",
    "firstname": "Bob",
    "age": 54
}
```

Query the database

1. In the **Data Explorer** page, select the **New SQL Query** icon.



2. In the SQL Query editor, review the default query (SELECT * FROM c) and use the **Execute Query** button to run it.

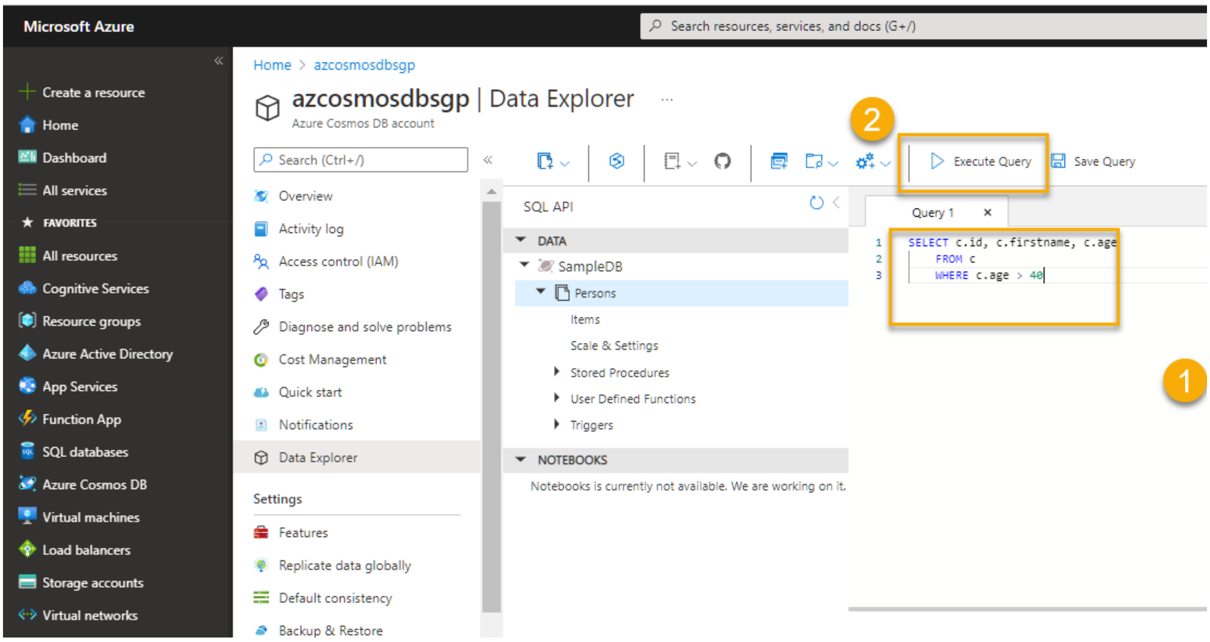
3. Review the results, which includes the full JSON representation of all items.
4. Modify the query as follows:

SQLCopy

SELECT c.id, c.firstname, c.age

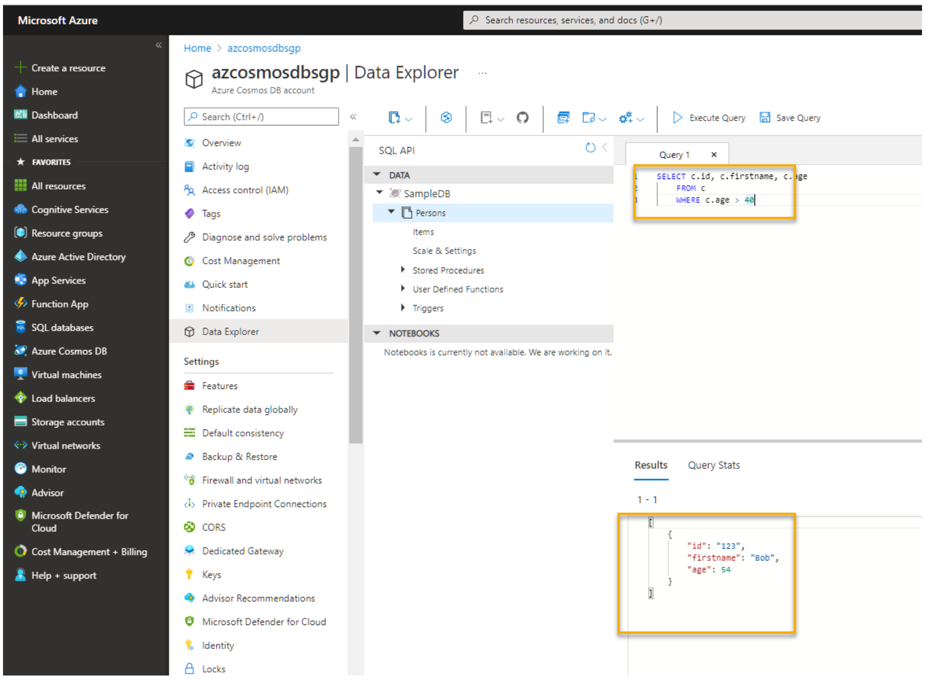
FROM c

WHERE c.age > 40



```
SELECT c.id, c.firstname, c.age
    FROM c
    WHERE c.age > 40
```

5. Use the **Execute Query** button to run the revised query and review the results, which includes JSON containing the id, firstname, and age fields for person items with an age greater than 40.



```
SELECT c.id, c.firstname, c.age
    FROM c
    WHERE c.age > 40
```

```
[
    {
        "id": "123",
        "firstname": "Bob",
        "age": 54
    }
]
```

6. Close the SQL Query editor, discarding your changes.

You've seen how to create and query JSON entities in a Cosmos DB database by using the data explorer interface in the Azure portal.

In **a real scenario,** an application developer would use one of the many programming language specific software development kits (SDKs) to call the core (SQL) API and work with data in the database.

Exercise: Explore Azure Synapse Analytics

https://docs.microsoft.com/en-gb/learn/modules/examine-components-of-modern-data-warehouse/5-exercise-azure-synapse#provision-an-azure-synapse-analytics-workspace

In this exercise, you'll create an **Azure Synapse Analytics workspace** and use it to **ingest** and **analyze** some data.

The exercise is designed to **familiarize you with some key elements of a modern data warehousing solution**, not as a comprehensive guide to performing advanced data analysis with Azure Synapse Analytics. The exercise should take around 30 minutes to complete.

Provision an Azure Synapse Analytics workspace

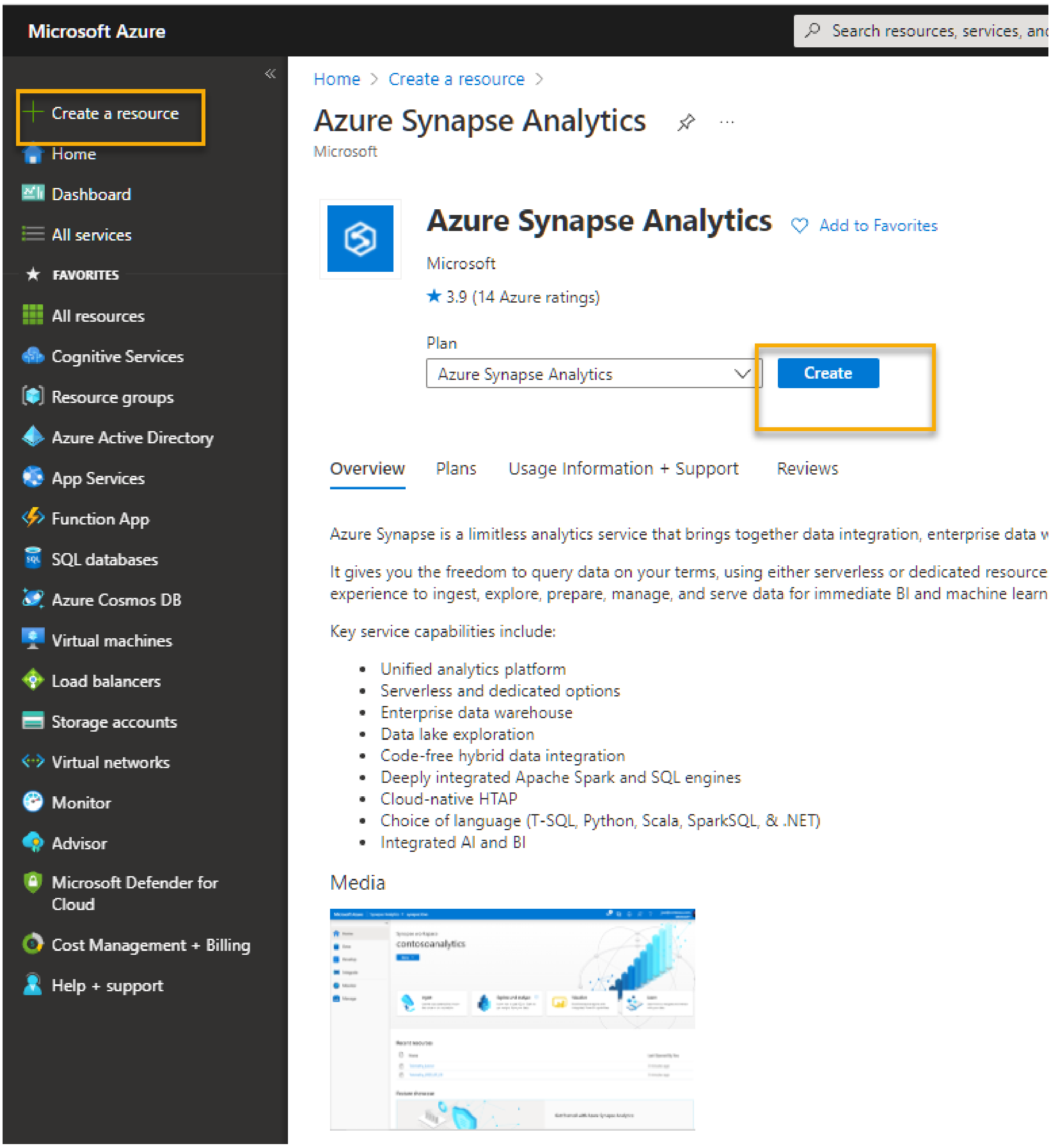
To use Azure Synapse Analytics, you must provision an Azure Synapse Analytics Workspace resource in your Azure subscription.

1. Open the Azure portal at https://portal.azure.com, and sign in using the credentials associated with your Azure subscription.

Tip

Ensure you are working in the directory containing your subscription - indicated at the top right under your user ID. If not, select the user icon and switch directory. Note that if you previously used a Microsoft Learn sandbox subscription, the portal may have defaulted to the *Microsoft Learn Sandbox* directory. If so, you'll need to switch to your own directory.

2. In the Azure portal, on the **Home** page, use the **+** **Create a resource** icon to create a new resource.



Azure Synapse is a limitless analytics service that brings together data integration, enterprise data w

It gives you the freedom to query data on your terms, using either serverless or dedicated resource experience to ingest, explore, prepare, manage, and serve data for immediate BI and machine learn

Key service capabilities include:

- Unified analytics platform
- Serverless and dedicated options
- Enterprise data warehouse
- Data lake exploration
- Code-free hybrid data integration
- Deeply integrated Apache Spark and SQL engines
- Cloud-native HTAP
- Choice of language (T-SQL, Python, Scala, SparkSQL, & .NET)
- Integrated AI and BI

3. Search for *Azure Synapse Analytics*, and create a new **Azure Synapse Analytics** resource with the following settings:

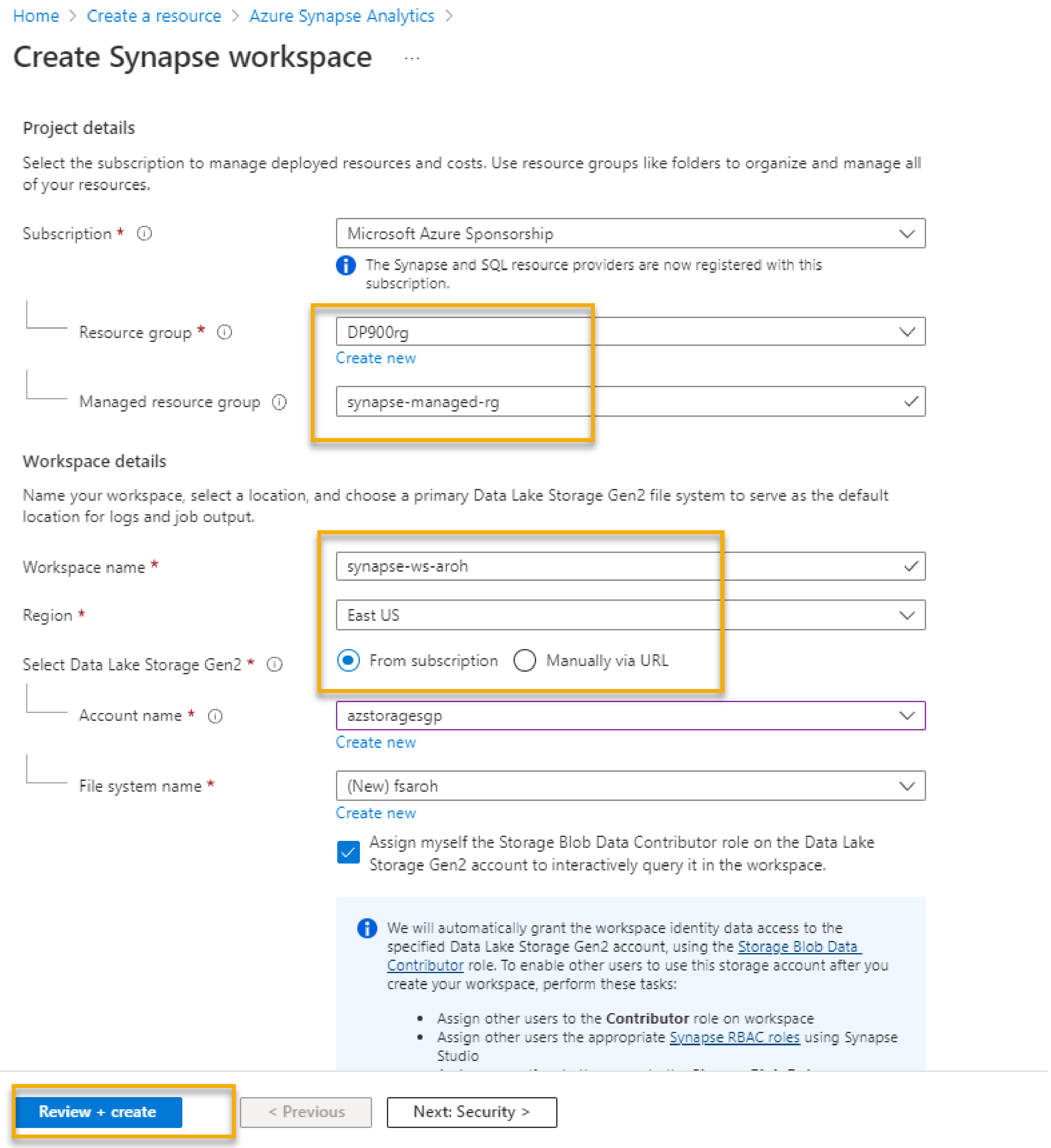
    - **Subscription**: *Your Azure subscription*
        - **Resource group**: *Create a new resource group with a suitable name, like "synapse-rg"*
        - **Managed resource group**: *Enter an appropriate name, for example "synapse-managed-rg"*.
    - **Workspace name**: *Enter a unique workspace name, for example "synapse-ws-<your_name>"*.

- **Region**: *Select any of the following regions*:
    - Australia East
    - Central US
    - East US 2
    - North Europe
    - South Central US
    - Southeast Asia
    - UK South
    - West Europe
    - West US
    - WestUS 2
- **Select Data Lake Storage Gen 2**: From subscription
    - **Account name**: *Create a new account with a unique name, for example "datalake<your_name>"*.
    - **File system name**: *Create a new file system with a unique name, for example "fs<your_name>"*.

Note

A Synapse Analytics workspace requires two resource groups in your Azure subscription; one for resources you explicitly create, and another for managed resources used by the service. It also requires a Data Lake storage account in which to store data, scripts, and other artifacts.

4. When you've entered these details, select **Review + create**, and then select **Create** to create the workspace.



Home > Create a resource > Azure Synapse Analytics >

Create Synapse workspace ...

Project details

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

Subscription * ⓘ: Microsoft Azure Sponsorship

ⓘ The Synapse and SQL resource providers are now registered with this subscription.

Resource group * ⓘ: DP900rg

Create new

Managed resource group ⓘ: synapse-managed-rg ✓

Workspace details

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

Workspace name *: synapse-ws-aroh ✓

Region *: East US

Select Data Lake Storage Gen2 * ⓘ: ● From subscription ○ Manually via URL

Account name * ⓘ: azstoragesgp

Create new

File system name *: (New) fsaroh

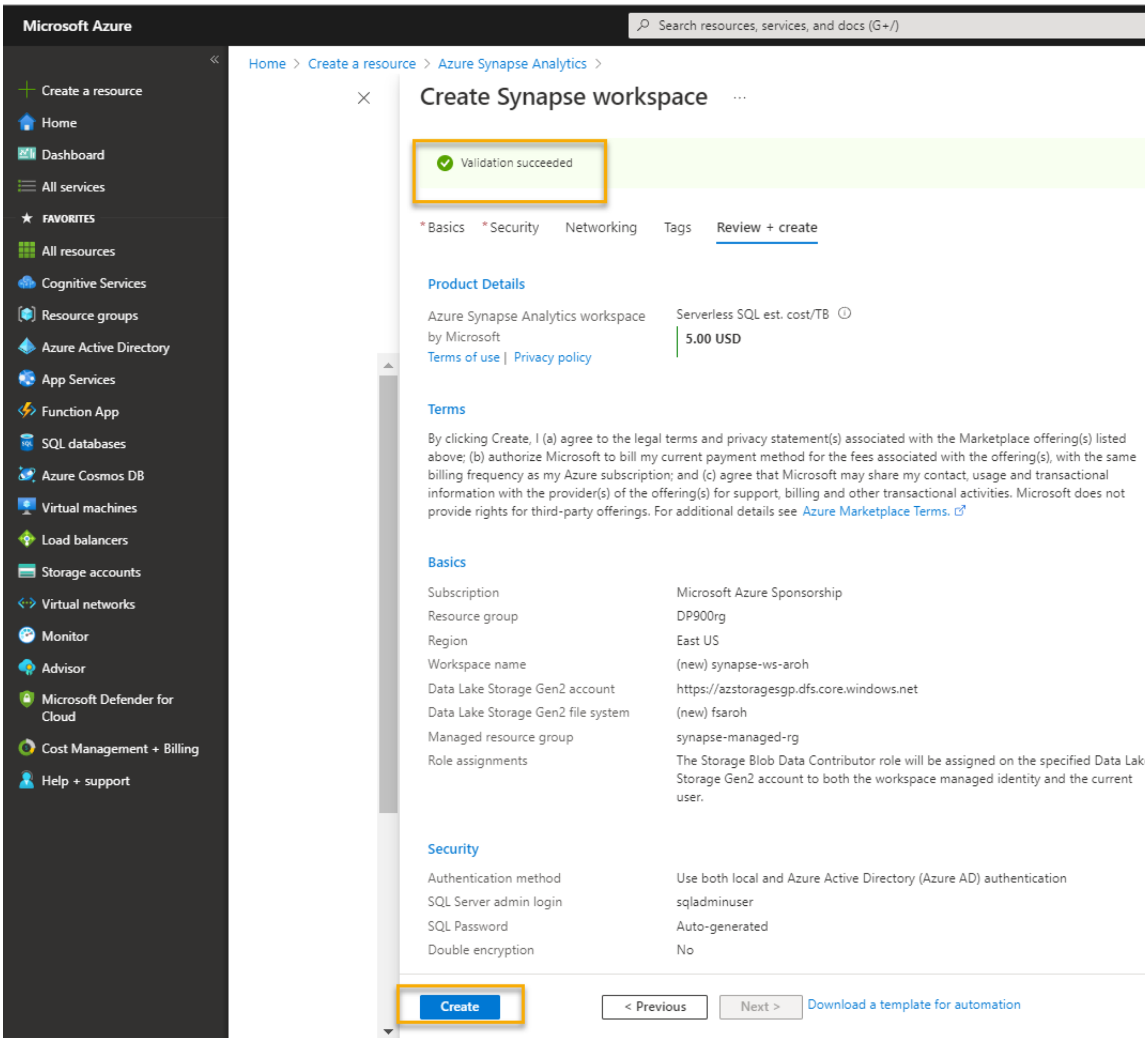
Create new

☑ 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

5. Wait for the workspace to be created - this may take five minutes or so.



Create Synapse workspace

Product Details

Azure Synapse Analytics workspace
by Microsoft
Terms of use | Privacy policy

Serverless SQL est. cost/TB
5.00 USD

Terms

By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see Azure Marketplace Terms.

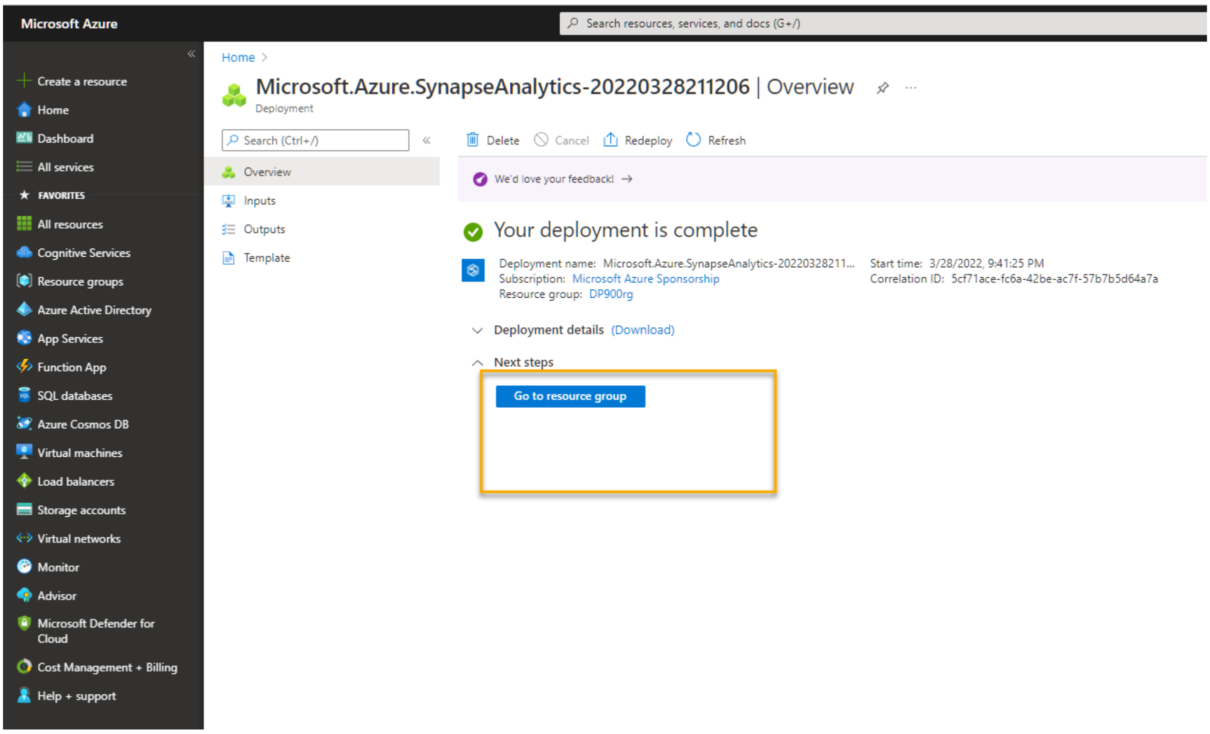
Basics

Subscription	Microsoft Azure Sponsorship
Resource group	DP900rg
Region	East US
Workspace name	(new) synapse-ws-aroh
Data Lake Storage Gen2 account	https://azstoragesgp.dfs.core.windows.net
Data Lake Storage Gen2 file system	(new) fsaroh
Managed resource group	synapse-managed-rg
Role assignments	The Storage Blob Data Contributor role will be assigned on the specified Data Lak Storage Gen2 account to both the workspace managed identity and the current user.

Security

Authentication method	Use both local and Azure Active Directory (Azure AD) authentication
SQL Server admin login	sqladminuser
SQL Password	Auto-generated
Double encryption	No

6. When deployment is complete, go to the resource group that was created and notice that it contains your **Synapse Analytics workspace** and a **Data Lake storage account**.

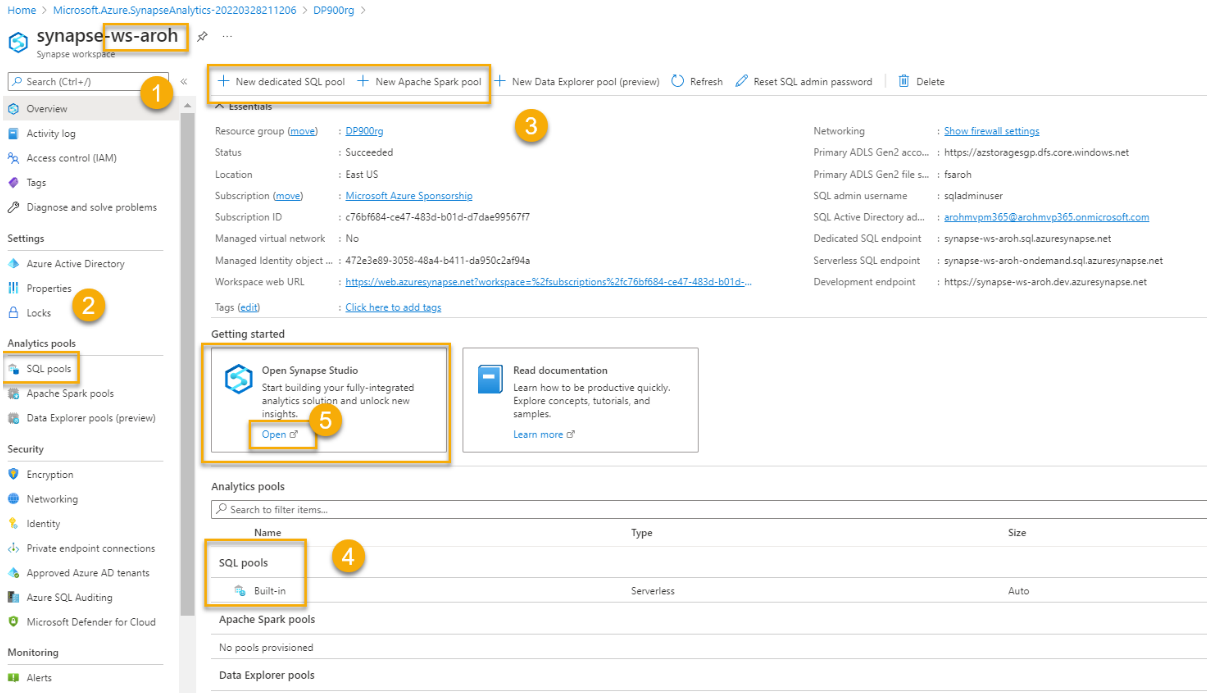


Your deployment is complete

Deployment name: Microsoft.Azure.SynapseAnalytics-20220328211...
Subscription: Microsoft Azure Sponsorship
Resource group: DP900rg

Start time: 3/28/2022, 9:41:25 PM
Correlation ID: 5cf71ace-fc6a-42be-ac7f-57b7b5d64a7a

7. Select your Synapse workspace, and in its **Overview** page, in **Open Synapse Studio** card, select **Open** to open Synapse Studio in a new browser tab. Synapse Studio is a **web-based interface** that you can use to work with your Synapse Analytics workspace.



Resource group (move) : DP900rg
Status : Succeeded
Location : East US
Subscription (move) : Microsoft Azure Sponsorship
Subscription ID : c76bf684-ce47-483d-b01d-d7dae99567f7
Managed virtual network : No
Managed Identity object ... : 472e3e89-3058-48a4-b411-da950c2af94a
Workspace web URL : https://web.azuresynapse.net?workspace=%2fsubscriptions%2fc76bf684-ce47-483d-b01d-...
Tags (edit) : Click here to add tags

Networking : Show firewall settings
Primary ADLS Gen2 acco... : https://azstoragesgp.dfs.core.windows.net
Primary ADLS Gen2 file s... : fsaroh
SQL admin username : sqladminuser
SQL Active Directory ad... : arohmvpm365@arohmvp365.onmicrosoft.com
Dedicated SQL endpoint : synapse-ws-aroh.sql.azuresynapse.net
Serverless SQL endpoint : synapse-ws-aroh-ondemand.sql.azuresynapse.net
Development endpoint : https://synapse-ws-aroh.dev.azuresynapse.net

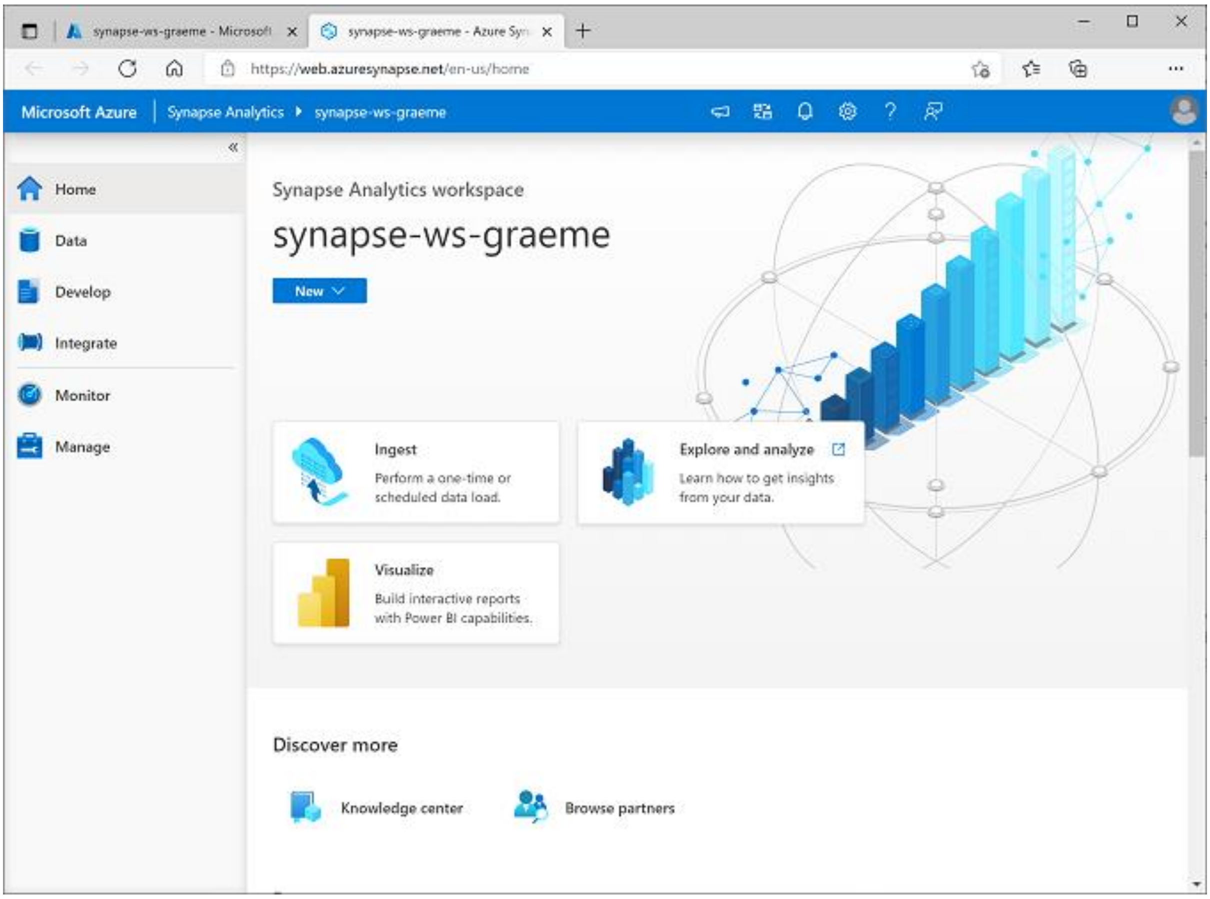
Getting started

Open Synapse Studio
Start building your fully-integrated analytics solution and unlock new insights.
Open

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

Name	Type	Size
SQL pools		
Built-in	Serverless	Auto
Apache Spark pools		
No pools provisioned		
Data Explorer pools		

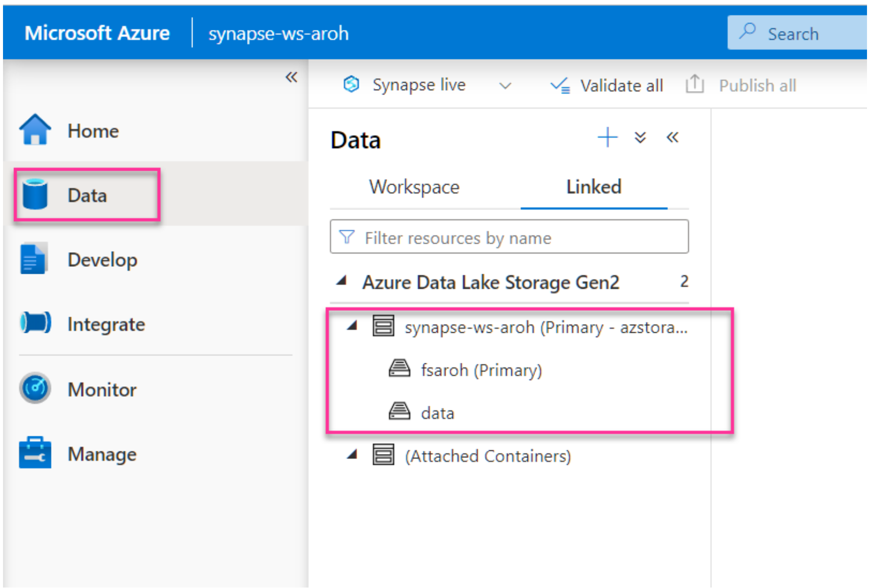
8. On the left side of Synapse Studio, use the ›› icon to expand the menu - this reveals the different pages within Synapse Studio that you'll use to manage resources and perform data analytics tasks, as shown here:

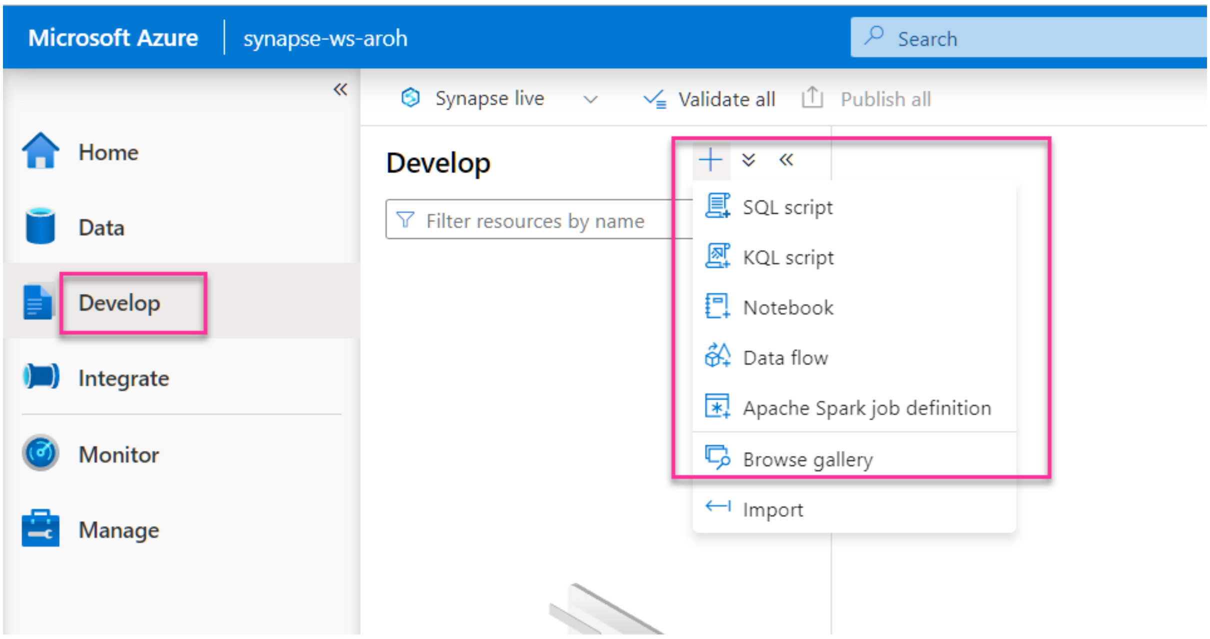


Perform a one-time or scheduled data load.

Learn how to get insights from your data.

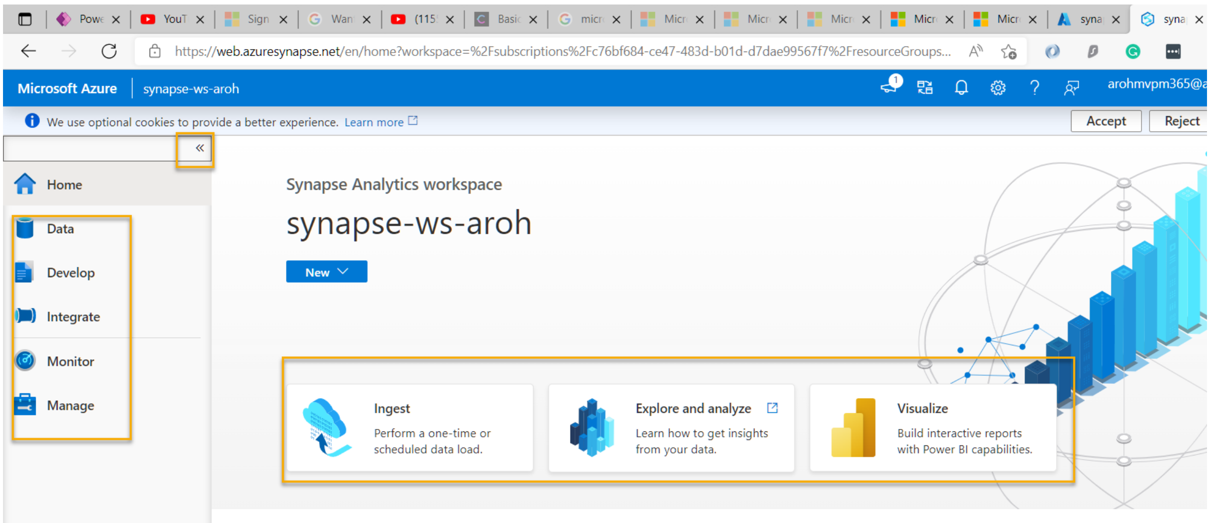
Build interactive reports with Power BI capabilities.

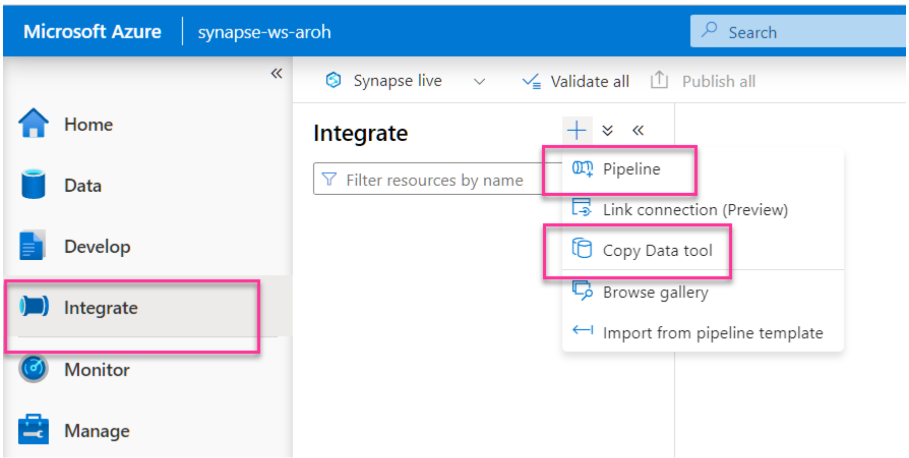


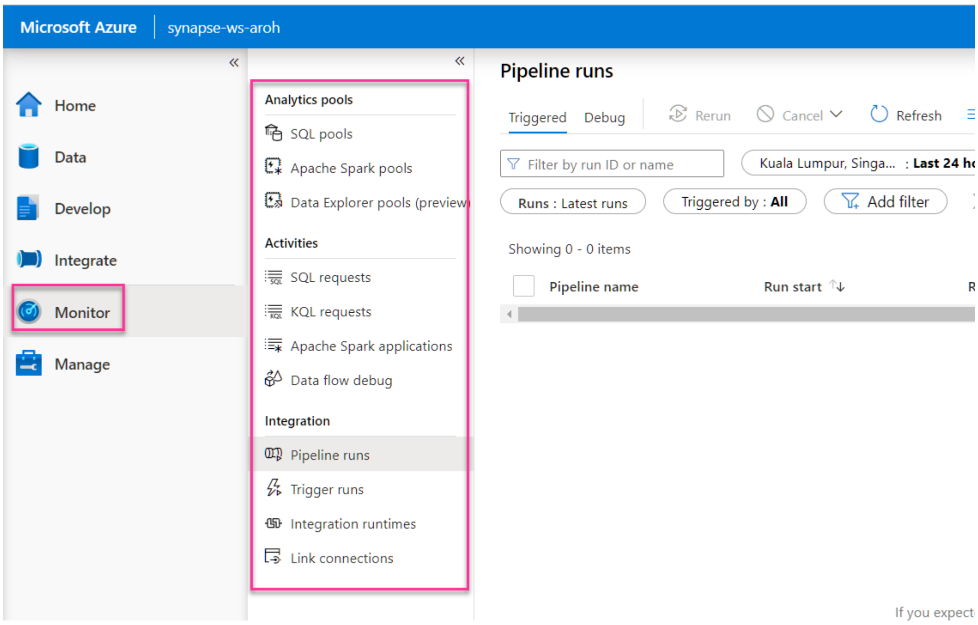


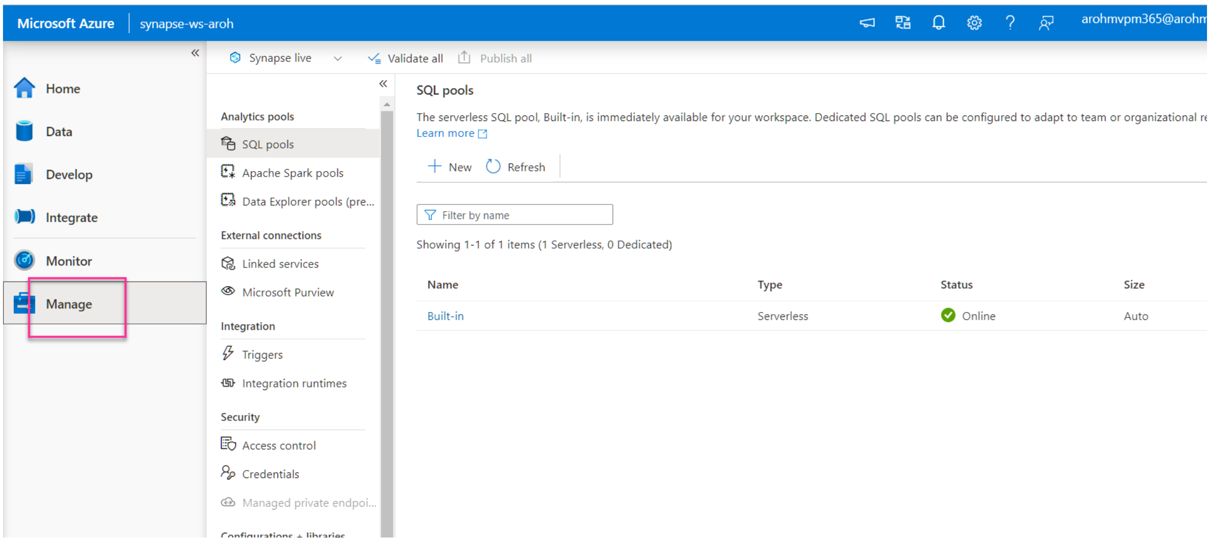
Keyword Query Language (KQL) syntax reference

[Keyword Query Language (KQL) syntax reference | Microsoft Docs]()


https://web.azuresynapse.net/en/home?workspace=%2Fsubscriptions%2Fc76bf684-ce47-483d-b01d-d7dae99567f7%2FresourceGroups...
Microsoft Azure
synapse-ws-aroh
arohmvpm365@a
We use optional cookies to provide a better experience. Learn more
Accept
Reject
Home
Data
Develop
Integrate
Monitor
Manage
Synapse Analytics workspace
synapse-ws-aroh
New
Ingest
Perform a one-time or scheduled data load.
Explore and analyze
Learn how to get insights from your data.
Visualize
Build interactive reports with Power BI capabilities.







The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational re

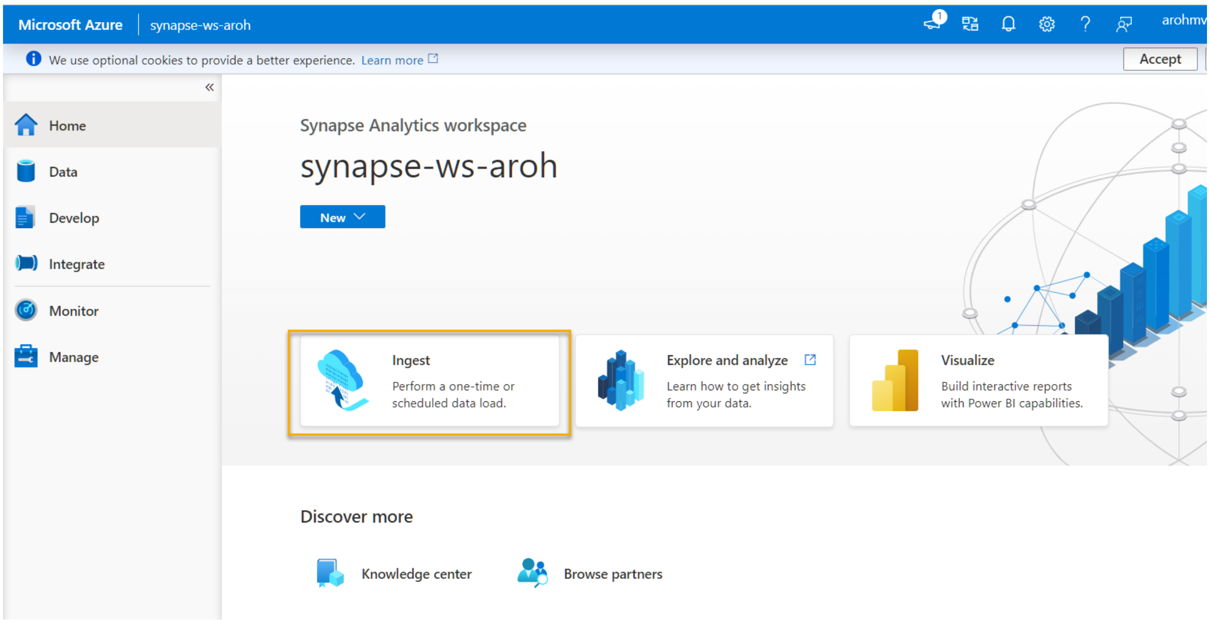
Showing 1-1 of 1 items (1 Serverless, 0 Dedicated)

Name	Type	Status	Size
Built-in	Serverless	✓ Online	Auto

Ingest data

One of the key tasks you can perform with Azure Synapse Analytics is to define *pipelines* that transfer (and if necessary, transform) data from a wide range of sources into your workspace for analysis.

1. In Synapse Studio, on the **Home** page, select **Ingest** to open the **Copy Data** tool



We use optional cookies to provide a better experience. Learn more

Synapse Analytics workspace

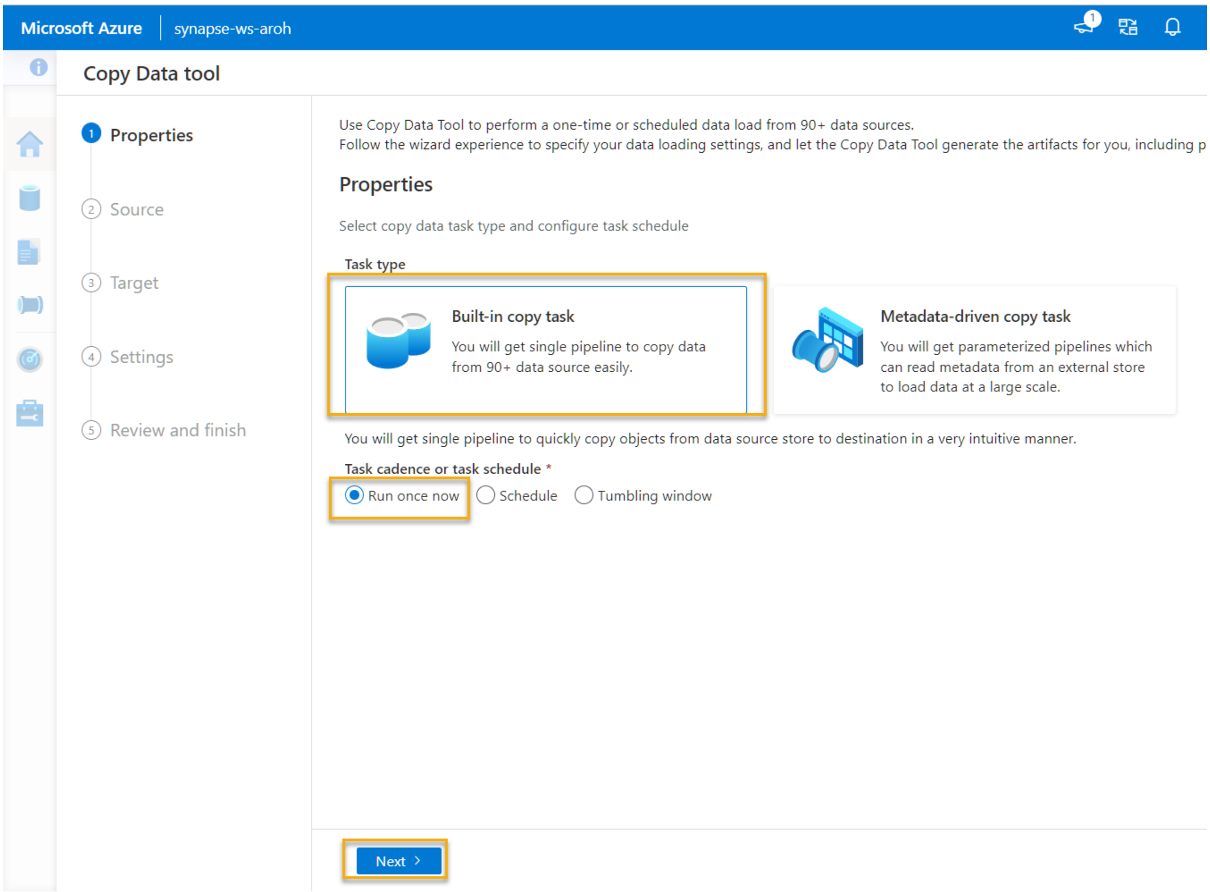
synapse-ws-aroh

Ingest: Perform a one-time or scheduled data load.

Explore and analyze: Learn how to get insights from your data.

Visualize: Build interactive reports with Power BI capabilities.

2. In the Copy Data tool, on the **Properties** step, ensure that **Built-in copy task** and **Run once now** are selected, and click **Next >**.



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 p

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.

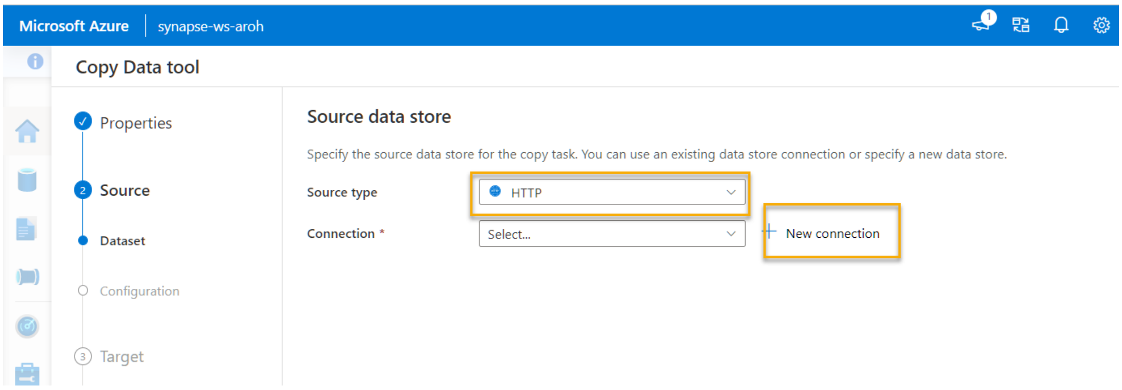
You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

● Run once now ○ Schedule ○ Tumbling window

3. On the **Source** step, in the **Dataset** substep, select the following settings:

- **Source type**: HTTP

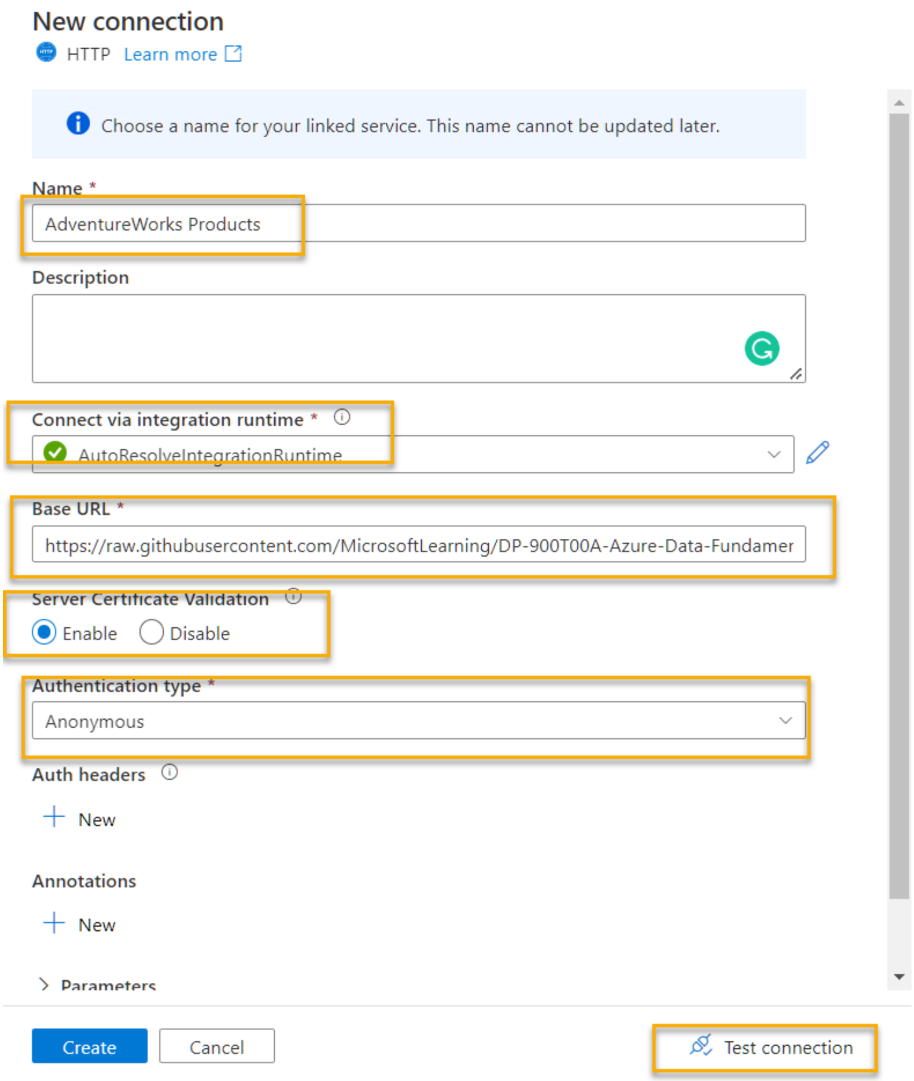

Source data store

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: HTTP

Connection *: Select...

- **Connection**: *Create a new connection with the following properties:*
    - **Name**: AdventureWorks Products
    - **Description**: Product list via HTTP
    - **Connect via integration runtime**: AutoResolveIntegrationRuntime
    - **Base URL**: https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/master/Azure-Synapse/products.csv
    - **Server Certificate Validation**: Enable
    - **Authentication type**: Anonymous



New connection

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

Name *

AdventureWorks Products

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Base URL *

https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamer

Server Certificate Validation

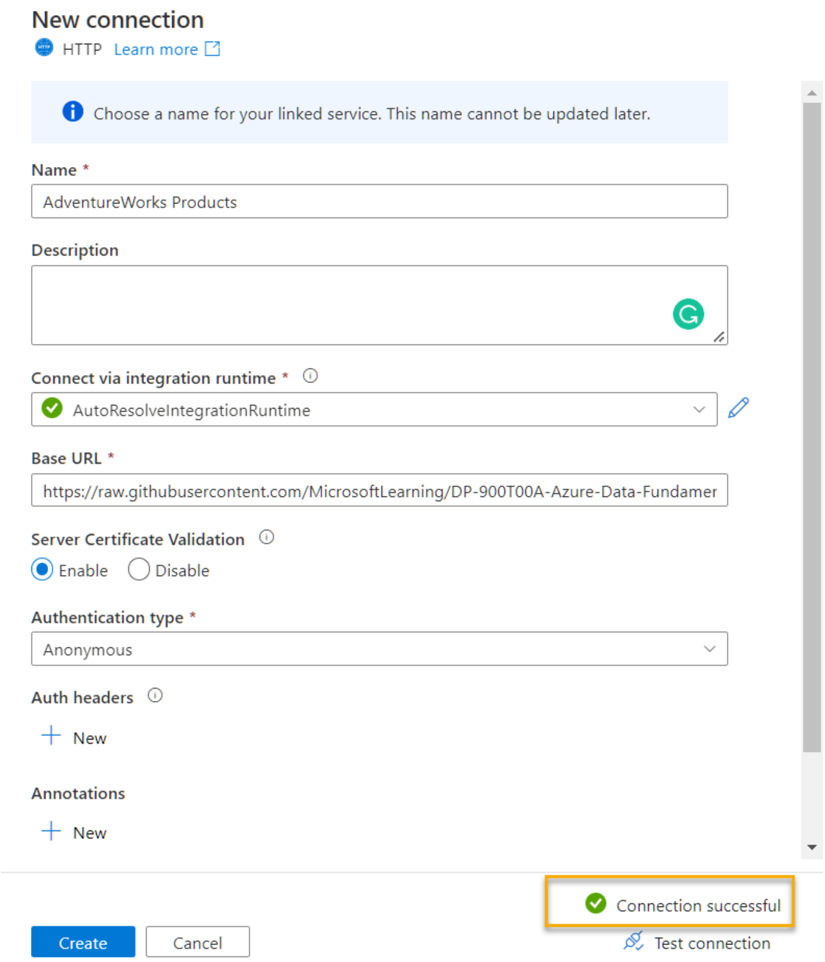
● Enable ○ Disable

Authentication type *

Anonymous

Auth headers

Annotations



New connection

HTTP Learn more

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

Name *

AdventureWorks Products

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Base URL *

https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamer

Server Certificate Validation

● Enable ○ Disable

Authentication type *

Anonymous

Auth headers

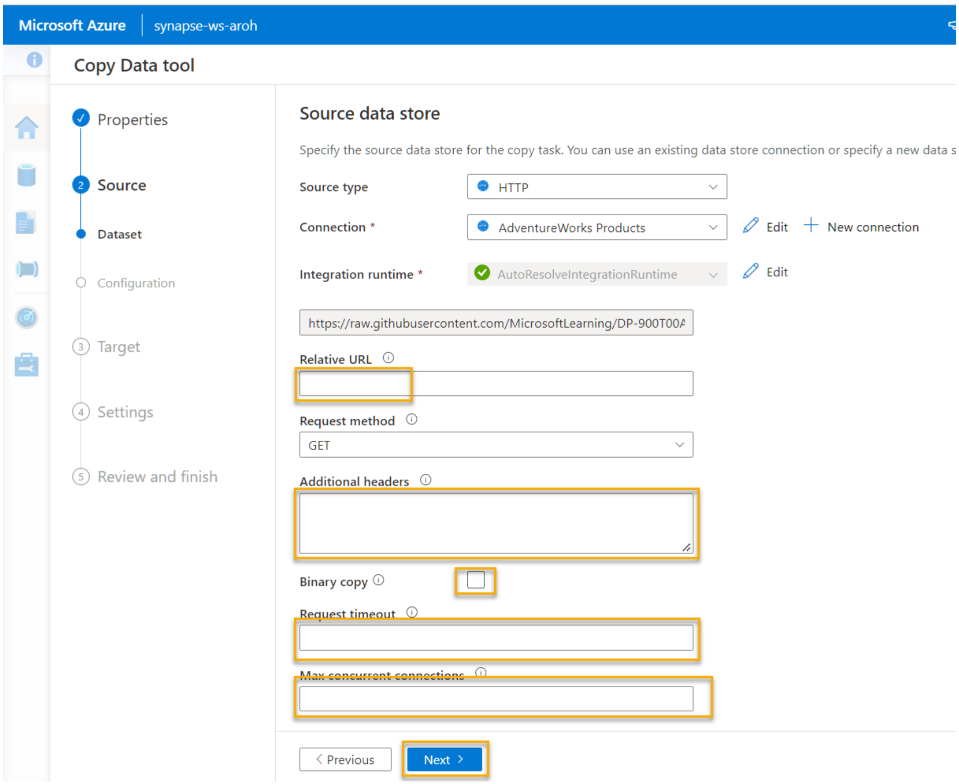
Annotations

Connection successful

4. After creating the connection, on the **Source/Dataset** substep, ensure the following settings are selected, and then select **Next >**:

   - **Relative URL**: *Leave blank*
   - **Request method**: GET
   - **Additional headers**: *Leave blank*
   - **Binary copy**: Unselected
   - **Request timeout**: *Leave blank*

- **Max concurrent connections**: *Leave blank*



Microsoft Azure | synapse-ws-aroh

Copy Data tool

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 s

Source type: HTTP

Connection *: AdventureWorks Products

Integration runtime *: AutoResolveIntegrationRuntime

https://raw.githubusercontent.com/MicrosoftLearning/DP-900T00A

Relative URL:

Request method: GET

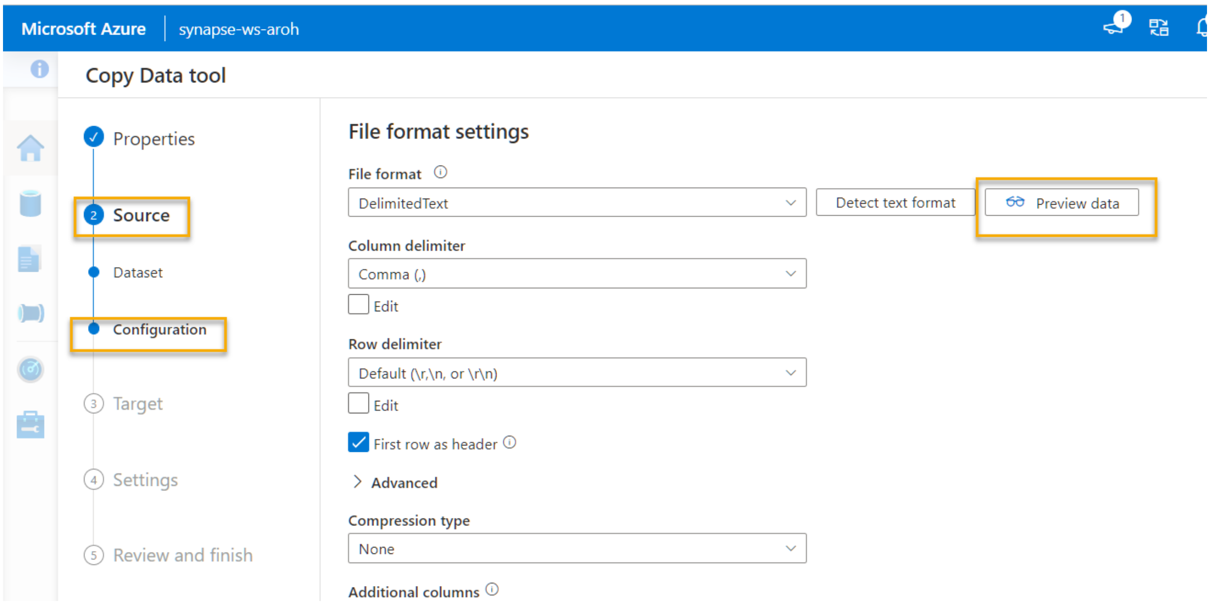
Additional headers:

Binary copy: □

Request timeout:

Max concurrent connections:

5. On the **Source** step, in the **Configuration** substep, select **Preview data** to see a preview of the product data your pipeline will ingest, then close the preview.



Microsoft Azure | synapse-ws-aroh

Copy Data tool

File format settings

File format: DelimitedText

Column delimiter: Comma (,)

□ Edit

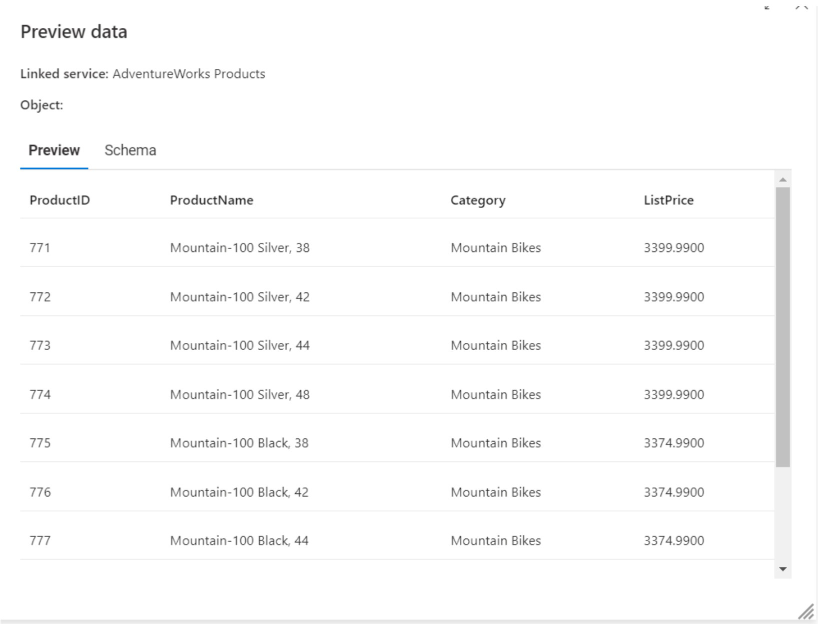
Row delimiter: Default (\r,\n, or \r\n)

□ Edit

☑ First row as header

Compression type: None

Additional columns



Preview data

Linked service: AdventureWorks Products

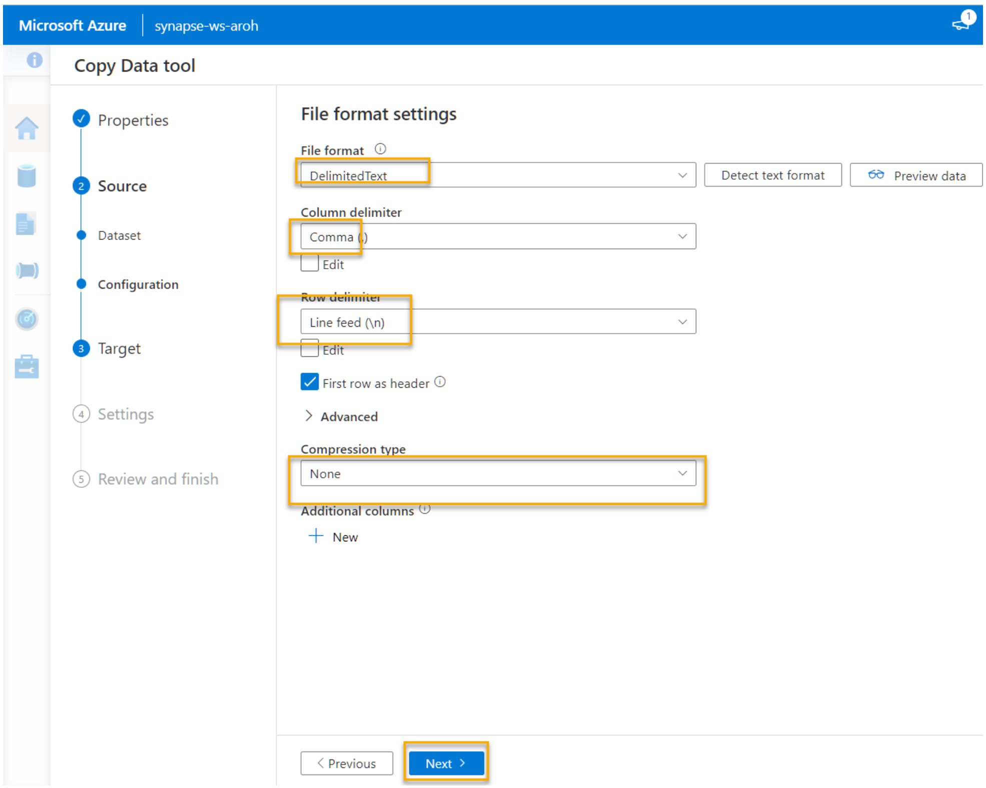
Object:

ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.9900
772	Mountain-100 Silver, 42	Mountain Bikes	3399.9900
773	Mountain-100 Silver, 44	Mountain Bikes	3399.9900
774	Mountain-100 Silver, 48	Mountain Bikes	3399.9900
775	Mountain-100 Black, 38	Mountain Bikes	3374.9900
776	Mountain-100 Black, 42	Mountain Bikes	3374.9900
777	Mountain-100 Black, 44	Mountain Bikes	3374.9900

6. After previewing the data, on the **Source/Configuration** step, ensure the following settings are selected, and then select **Next >**:

   - **File format**: DelimitedText
   - **Column delimiter**: Comma (,)
   - **Row delimiter**: Line feed (\n)
   - **First row as header**: Selected

- **Compression type**: None



File format settings

File format: DelimitedText

Column delimiter: Comma (,)

□ Edit

Row delimiter: Line feed (\n)

□ Edit

☑ First row as header

> Advanced

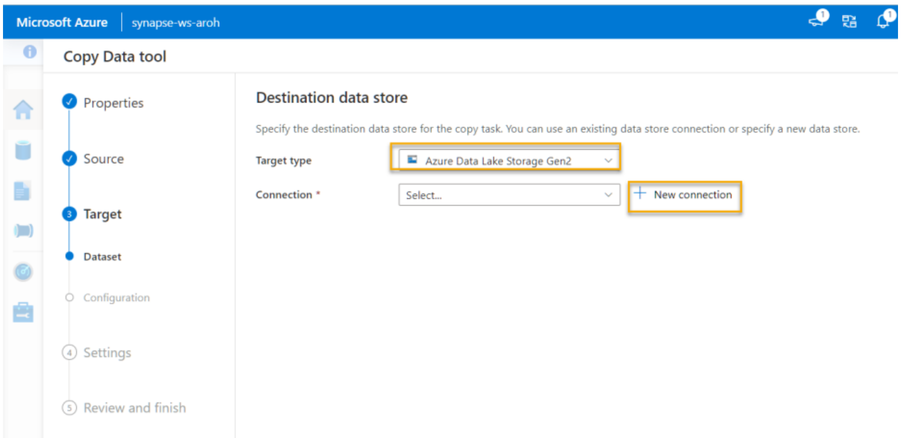
Compression type: None

Additional columns

+ New

7. On the **Target** step, in the **Dataset** substep, select the following settings:

   - **Target type**: Azure Data Lake Storage Gen 2 = **Connection**: *Create a new connection with the following properties:*



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.

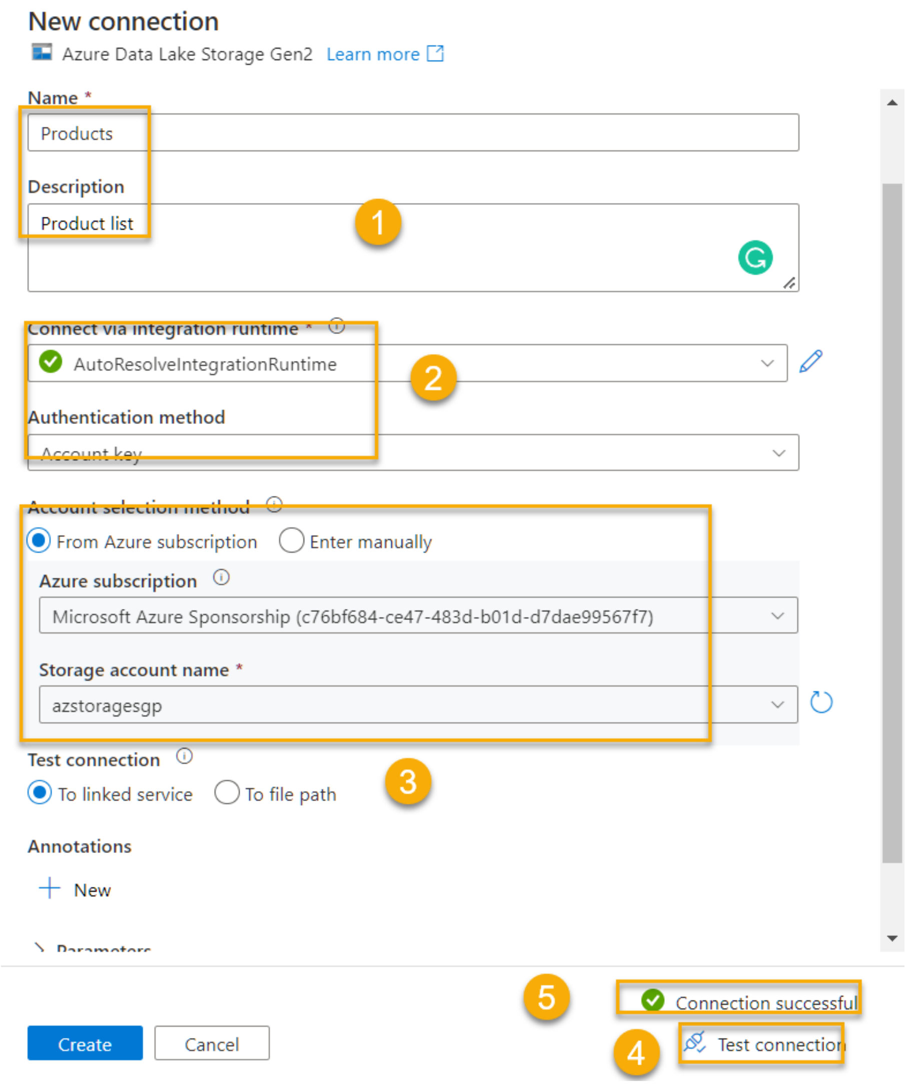
Target type: Azure Data Lake Storage Gen2

Connection *: Select...

     - **Name**: Products
     - **Description**: Product list
     - **Connect via integration runtime**: AutoResolveIntegrationRuntime
     - **Authentication method**: Account key
     - **Account selection method**: From subscription
       - **Azure subscription**: *select your subscription*

- **Storage account name**: *Select your storage account*

- **Test connection**: To linked service



New connection

Azure Data Lake Storage Gen2 Learn more

Name *

Products

Description

Product list

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication method

Account key

Account selection method

● From Azure subscription ○ Enter manually

Azure subscription

Microsoft Azure Sponsorship (c76bf684-ce47-483d-b01d-d7dae99567f7)

Storage account name *

azstoragesgp

Test connection

● To linked service ○ To file path

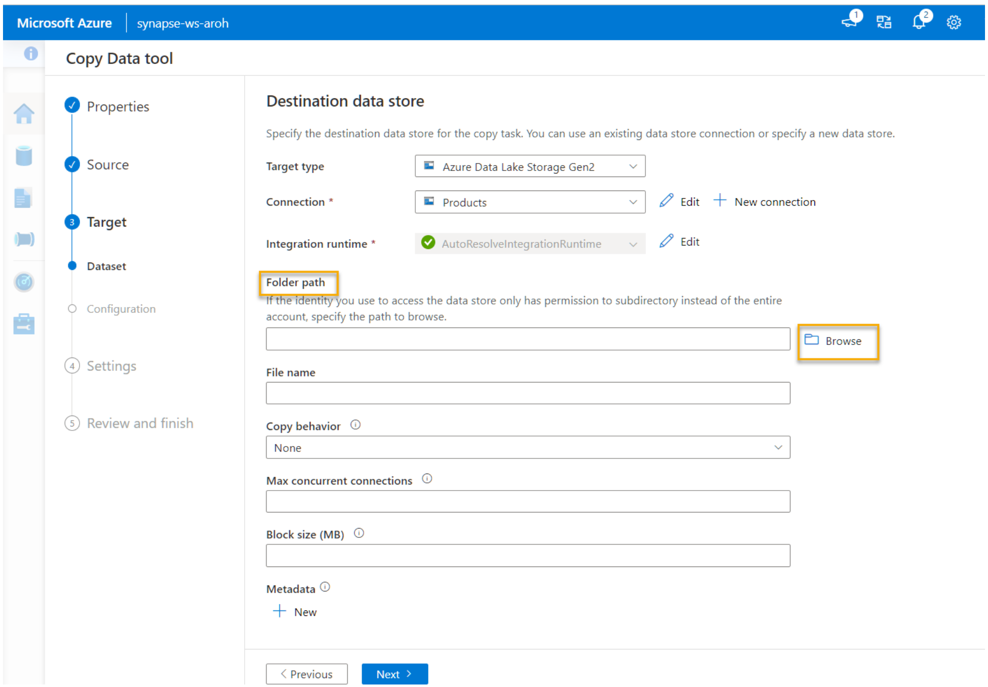
Annotations

Connection successful

Test connection

8. After creating the connection, on the **Target/Dataset** step, ensure the following settings are selected, and then select **Next >**:

   - **Folder path**: *Browse to your file system folder*



Microsoft Azure | synapse-ws-aroh

Copy Data tool

- Properties
- Source
- Target
- Dataset
- Configuration
- Settings
- 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.

Target type: Azure Data Lake Storage Gen2

Connection *: Products

Integration runtime *: AutoResolveIntegrationRuntime

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.

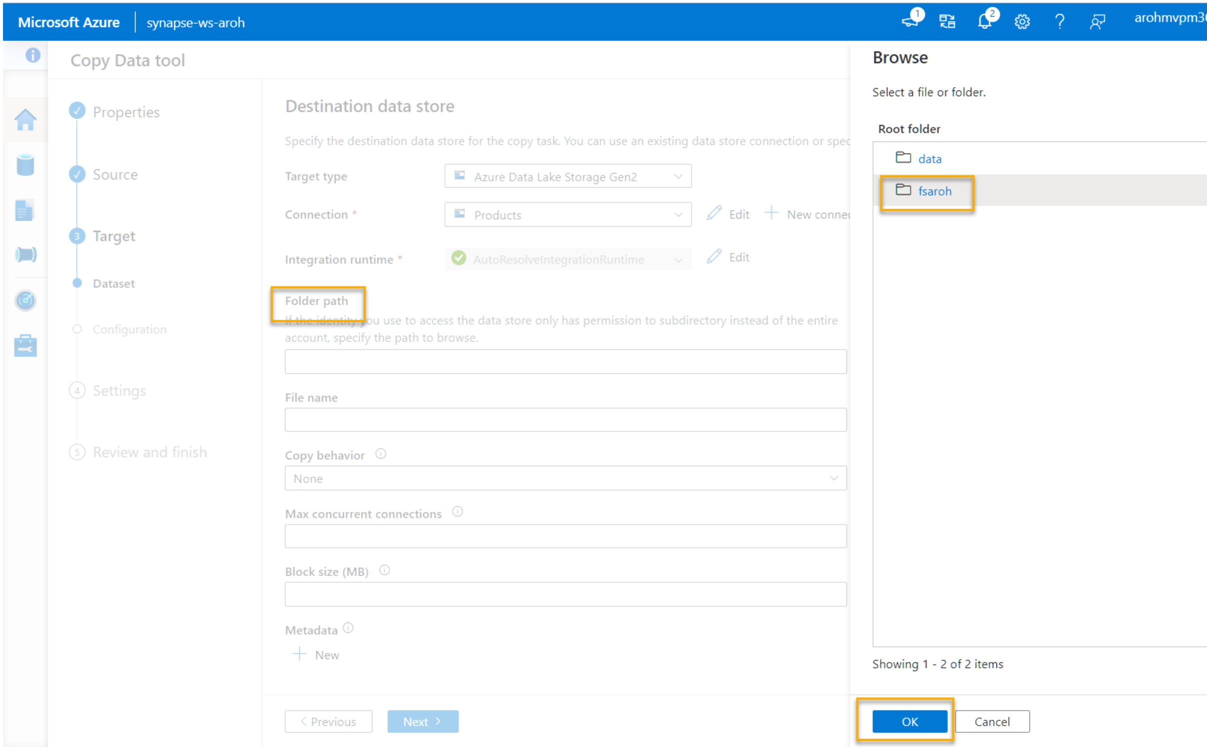
File name

Copy behavior: None

Max concurrent connections

Block size (MB)

Metadata



Microsoft Azure | synapse-ws-aroh

Browse

Select a file or folder.

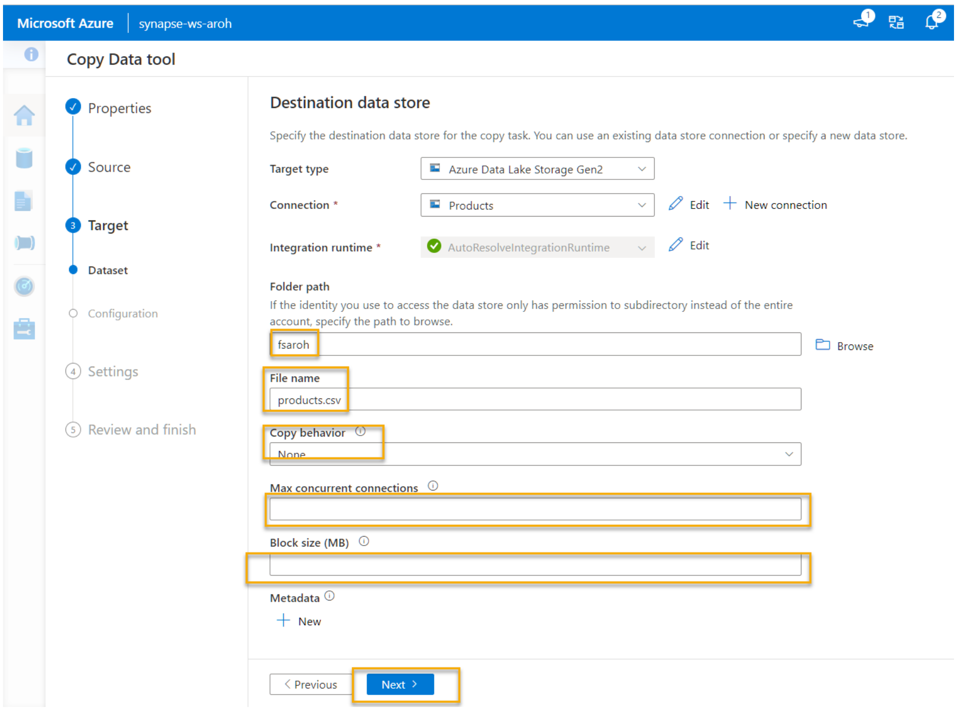
Root folder

- data
- fsaroh

Showing 1 - 2 of 2 items

- **File name**: products.csv
- **Copy behavior**: None
- **Max concurrent connections**: *Leave blank*

- **Block size (MB)**: *Leave blank*



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: Azure Data Lake Storage Gen2

Connection *: Products

Integration runtime *: AutoResolveIntegrationRuntime

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.

fsaroh

File name: products.csv

Copy behavior: None

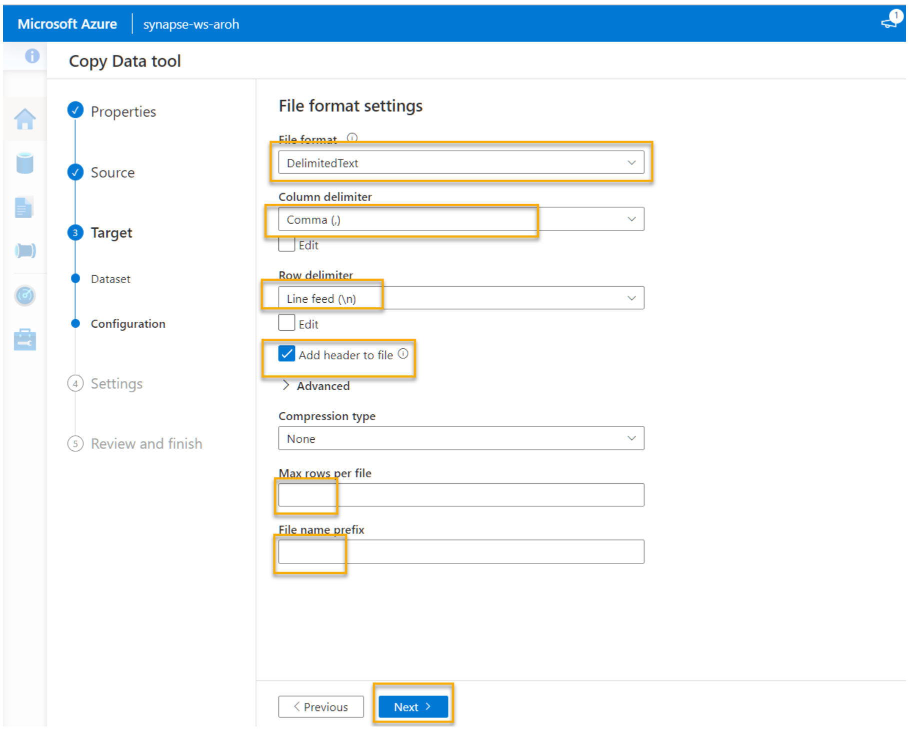
Max concurrent connections:

Block size (MB):

Metadata

9. On the **Target** step, in the **Configuration** substep, ensure that the following properties are selected. Then select **Next >**:

   - **File format**: DelimitedText
   - **Column delimiter**: Comma (,)
   - **Row delimiter**: Line feed (\n)
   - **Add header to file**: Selected
   - **Compression type**: None
   - **Max rows per file**: *Leave blank*
   - **File name prefix**: *Leave blank*



File format settings

File format: DelimitedText

Column delimiter: Comma (,)

□ Edit

Row delimiter: Line feed (\n)

□ Edit

☑ Add header to file

> Advanced

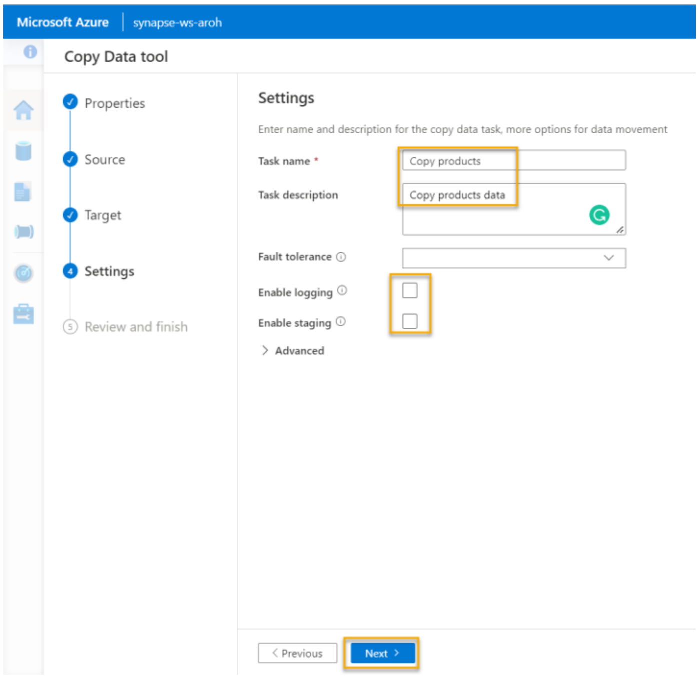
Compression type: None

Max rows per file:

File name prefix:

10. On the **Settings** step, enter the following settings and then click **Next >**:

   - **Task name**: Copy products
   - **Task description** Copy products data
   - **Fault tolerance**: *Leave blank*
   - **Enable logging**: Unselected
   - **Enable staging**: Unselected



Settings

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

Task name *: Copy products

Task description: Copy products data

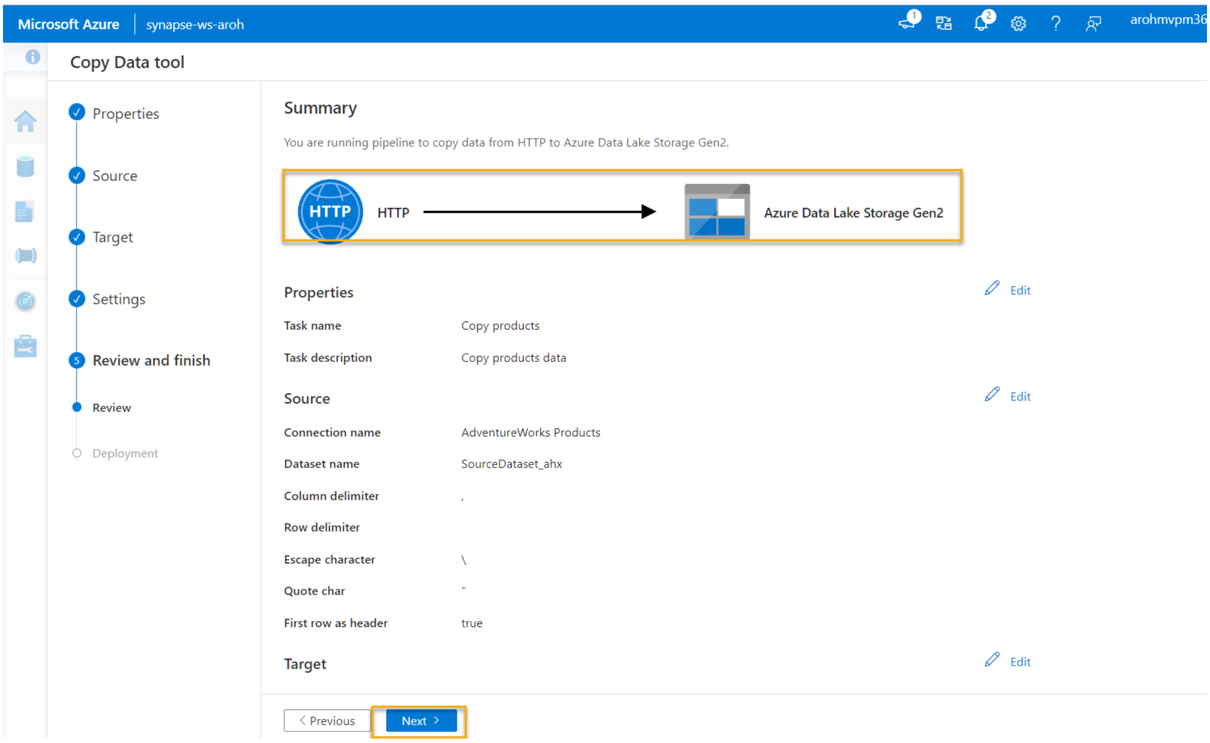
Fault tolerance:

Enable logging □

Enable staging □

> Advanced

11. On the **Review and finish** step, on the **Review** substep, read the summary and then click **Next >**.



Summary

You are running pipeline to copy data from HTTP to Azure Data Lake Storage Gen2.

Properties ✎ Edit

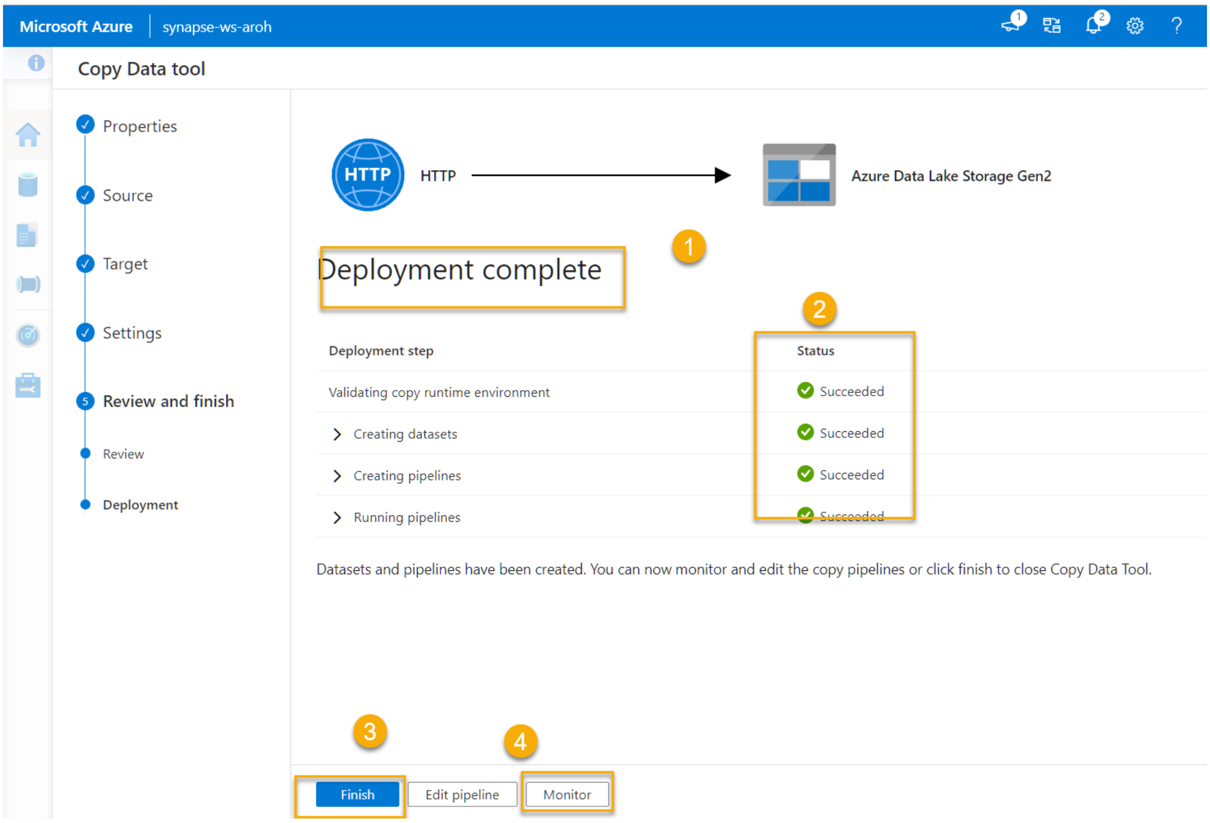
Task name	Copy products
Task description	Copy products data

Source ✎ Edit

Connection name	AdventureWorks Products
Dataset name	SourceDataset_ahx
Column delimiter	,
Row delimiter	
Escape character	\
Quote char	"
First row as header	true

Target ✎ Edit

12. On the **Deployment** step, wait for the pipeline to be deployed and then click **Finish**.

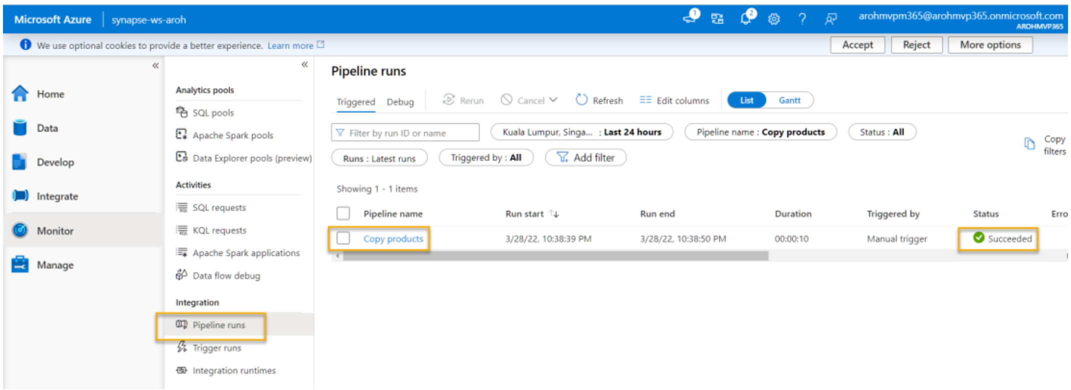


Deployment complete

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.

13. In Synapse Studio, select the **Monitor** page, and in the **Pipeline runs** tab, wait for the **Copy products** pipeline to complete with a status of **Succeeded** (you can use the ↻ **Refresh** button on the Pipeline runs page to refresh the status).



Pipeline runs

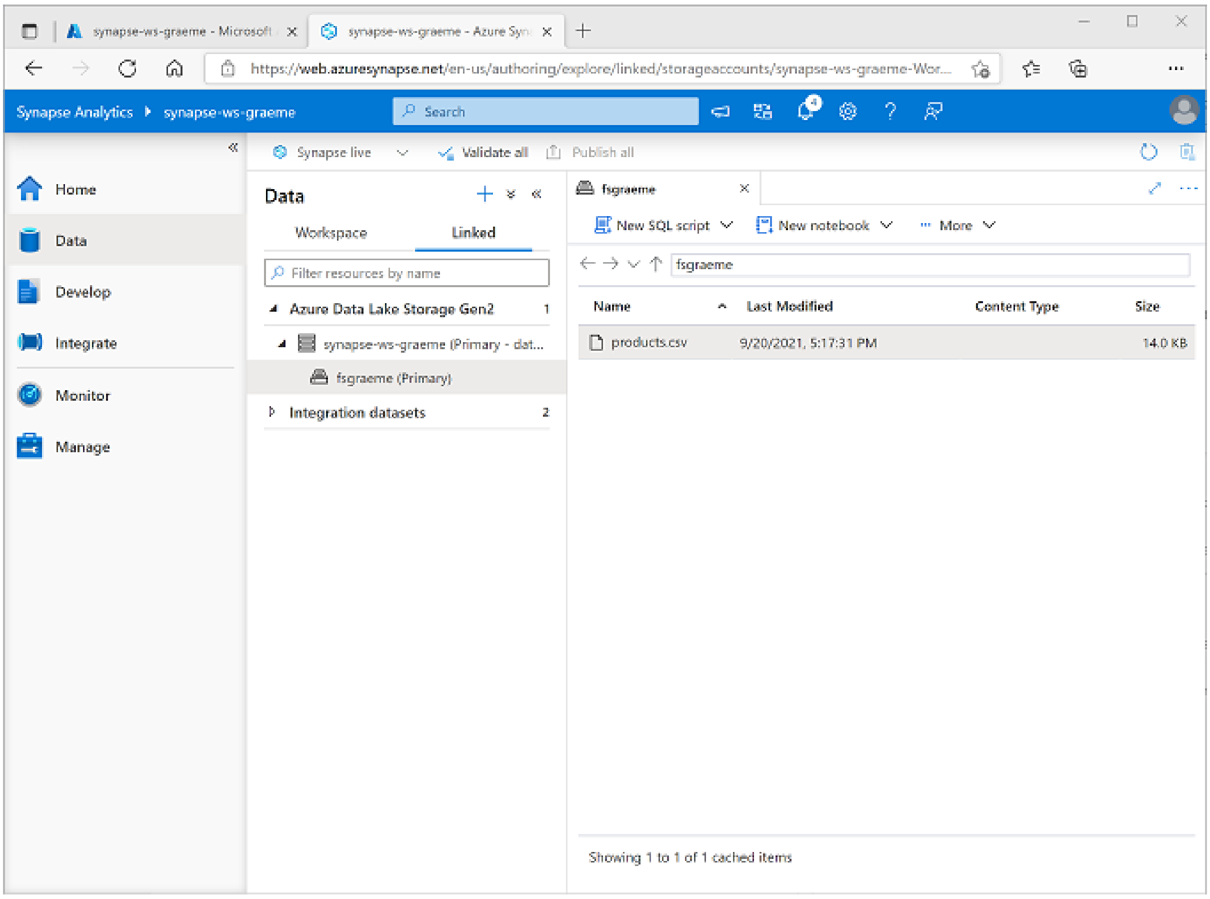
Kuala Lumpur, Singa... : **Last 24 hours**   Pipeline name : **Copy products**   Status : **All**

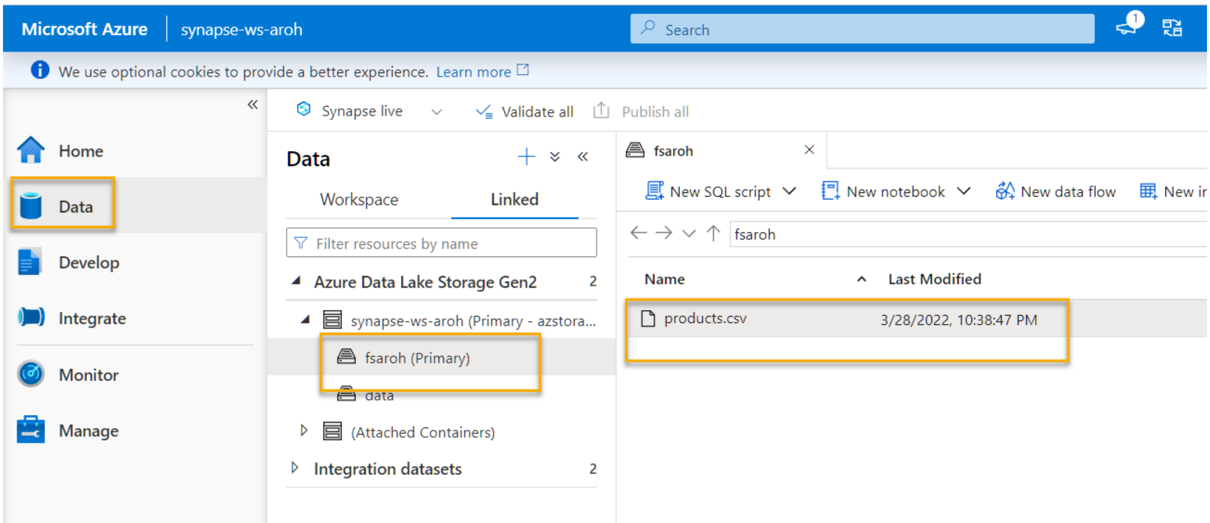
Runs : Latest runs   Triggered by : **All**

Showing 1 - 1 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Erro
Copy products	3/28/22, 10:38:39 PM	3/28/22, 10:38:50 PM	00:00:10	Manual trigger	✓ Succeeded	

14. On the **Data** page, select the **Linked** tab and expand the **Azure Data Lake Storage Gen 2** hierarchy until you see the file storage for your Synapse workspace. Then select the file storage to verify that a file named **products.csv** has been copied to this location, as shown here:

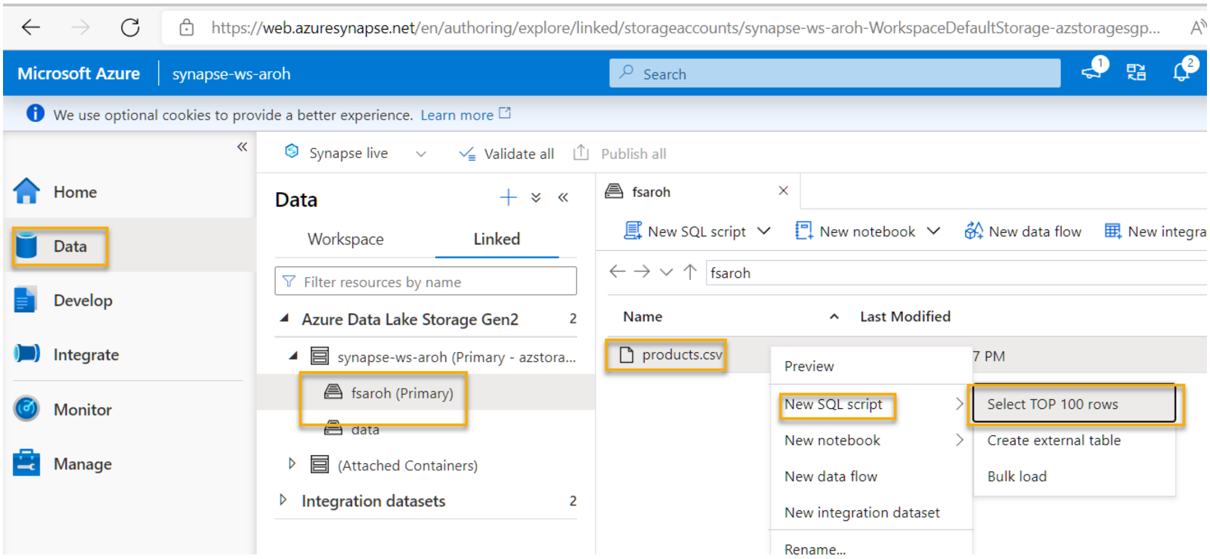




Use a SQL pool to analyze data

Now that you've ingested some data into your workspace, you can use Synapse Analytics to query and analyze it. One of the most common ways to query data is to use SQL, and in Synapse Analytics you can use a *SQL pool* to run SQL code.

1. In Synapse Studio, right-click the **products.csv** file in the file storage for your Synapse workspace, point to **New SQL script**, and select **Select TOP 100 rows**.



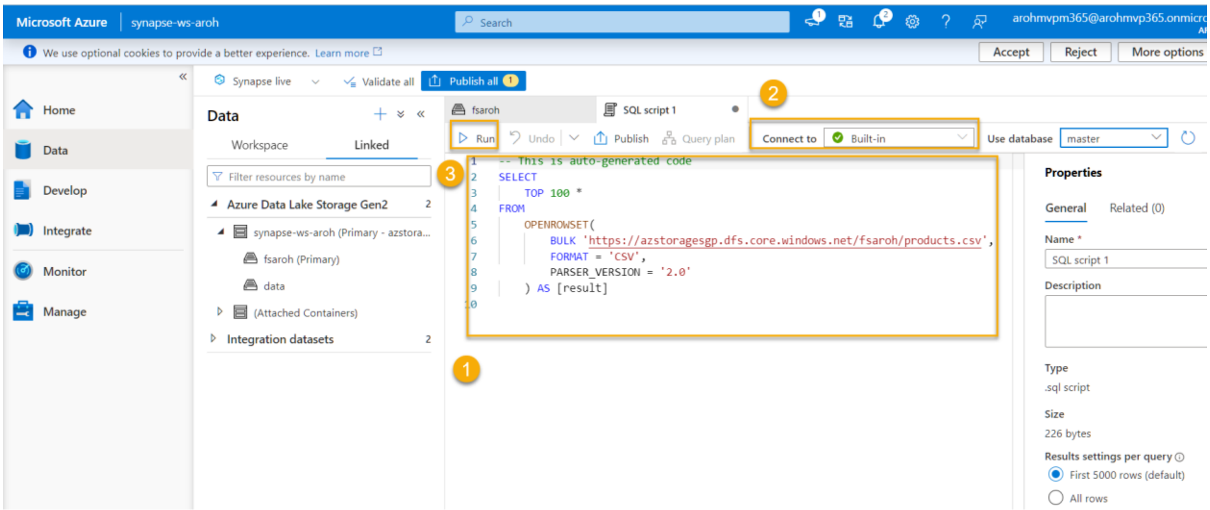
2. In the **SQL Script 1** pane that opens, review the SQL code that has been generated, which should be similar to this:

SQLCopy

```
-- This is auto-generated code
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://datalakexx.dfs.core.windows.net/fsxx/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION='2.0'
    ) AS [result]
```

This code opens a rowset from the text file you imported and retrieves the first 100 rows of data.

3. In the **Connect to** list, ensure **Built-in** is selected - this represents the built-in SQL Pool that was created with your workspace.



```
-- This is auto-generated code
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0'
    ) AS [result]
```

Name *: SQL script 1

Description:

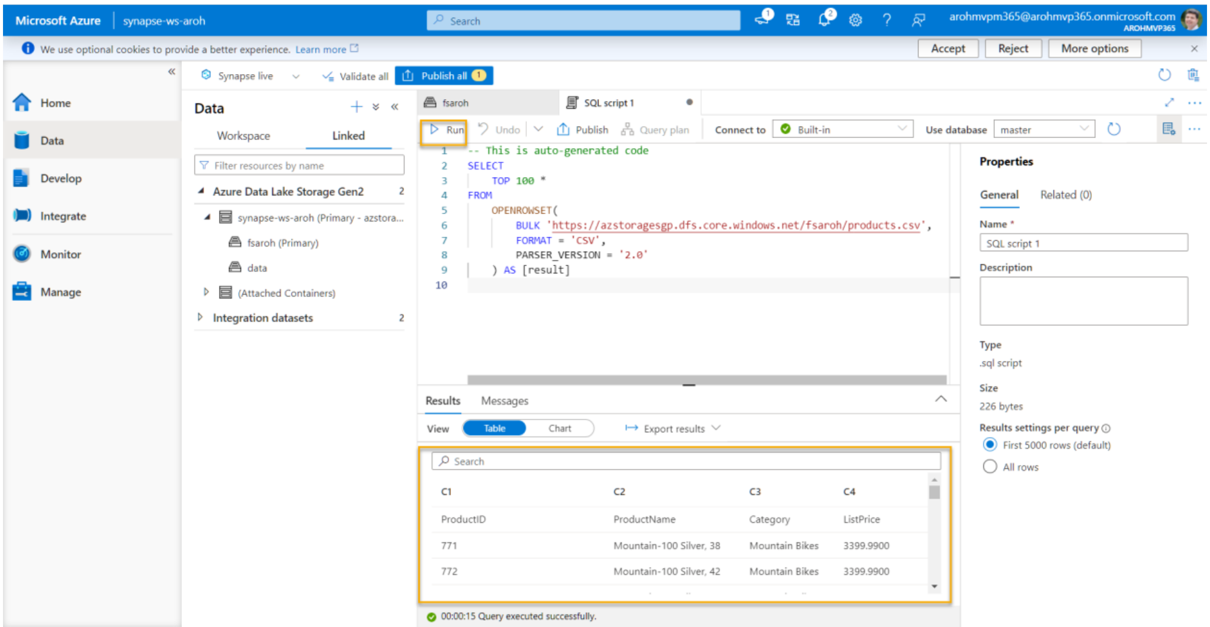
Type: .sql script

Size: 226 bytes

Results settings per query
● First 5000 rows (default)
○ All rows

4. On the toolbar, use the ▷ **Run** button to run the SQL code, and review the results, which should look similar to this:

C1	**c2**	**c3**
ProductID	ProductName	Category
771	Mountain-100 Silver, 38	Mountain Bikes
772	Mountain-100 Silver, 42	Mountain Bikes
...	...	...



```
-- This is auto-generated code
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0'
    ) AS [result]
```

C1	C2	C3	C4
ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.9900
772	Mountain-100 Silver, 42	Mountain Bikes	3399.9900

00:00:15 Query executed successfully.

Name *: SQL script 1

Description:

Type: .sql script

Size: 226 bytes

Results settings per query
● First 5000 rows (default)
○ All rows

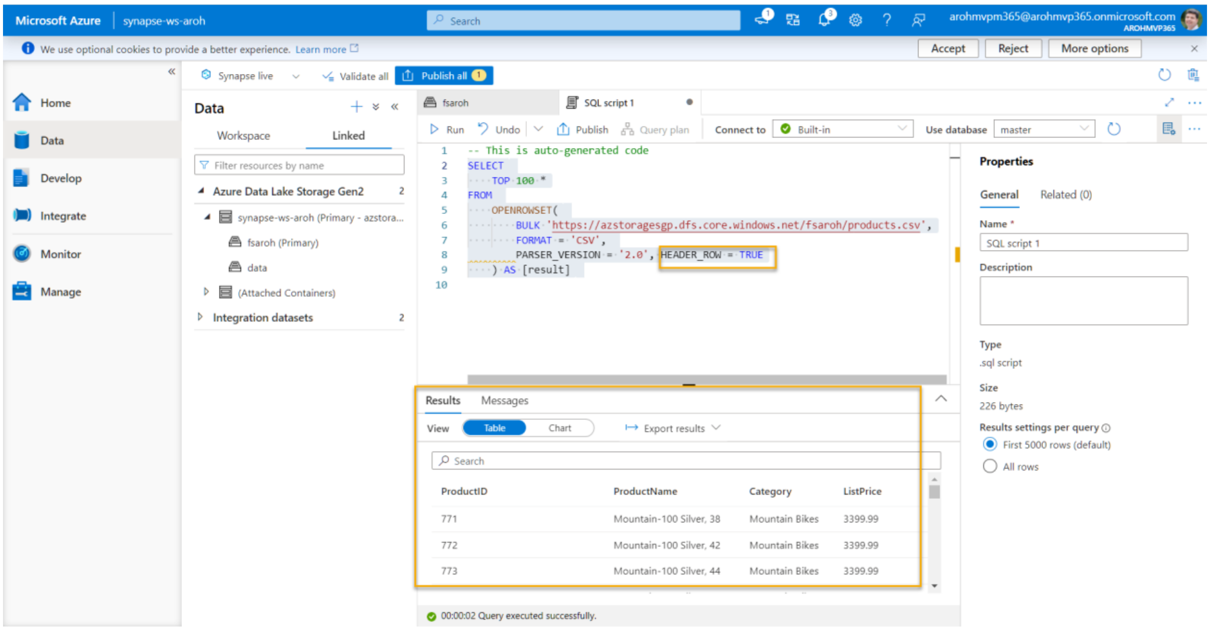
5. Note the results consist of four columns named **C1, C2, C3, and C4**; and that the first row in the results contains the names of the data fields. To fix this problem, add a HEADER_ROW = TRUE parameters to the **OPENROWSET** function as shown here (replacing *datalakexx* and *fsxx* with the names of your data lake storage account and file system), and then rerun the query:

SQLCopy

```
SELECT
   TOP 100 *
FROM
   OPENROWSET(
      BULK 'https://datalakexx.dfs.core.windows.net/fsxx/products.csv',
      FORMAT = 'CSV',
      PARSER_VERSION='2.0',
      HEADER_ROW = TRUE
   ) AS [result]
```

Now the results look like this:

ProductID	ProductName	Category
771	Mountain-100 Silver, 38	Mountain Bikes
772	Mountain-100 Silver, 42	Mountain Bikes
...	...	...



```
-- This is auto-generated code
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0', HEADER_ROW = TRUE
    ) AS [result]
```

ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.99
772	Mountain-100 Silver, 42	Mountain Bikes	3399.99
773	Mountain-100 Silver, 44	Mountain Bikes	3399.99

00:00:02 Query executed successfully.

6. Modify the query as follows (replacing *datalakexx* and *fsxx* with the names of your data lake storage account and file system):

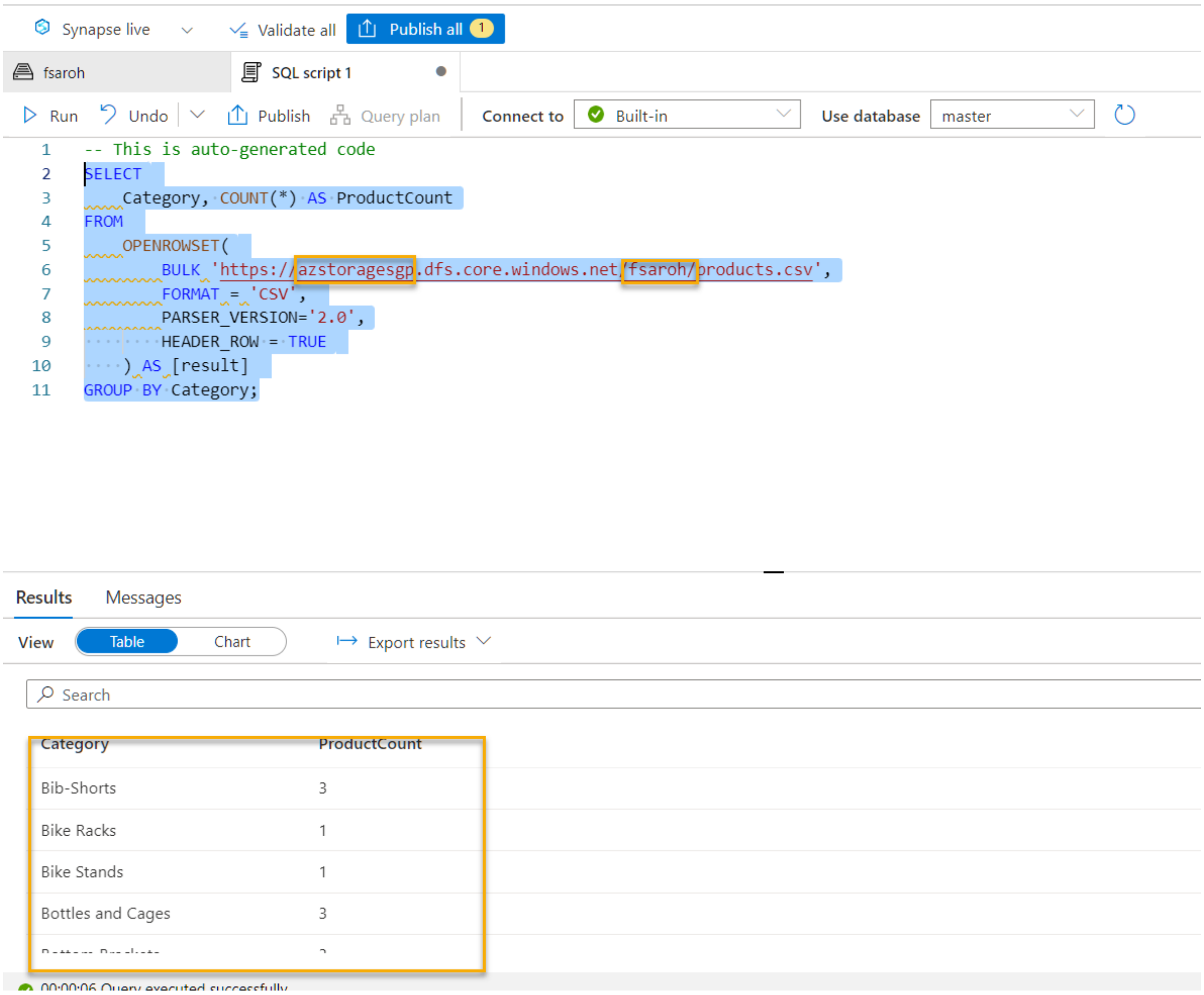
SQLCopy

```
SELECT
   Category, COUNT(*) AS ProductCount
```

```
FROM
  OPENROWSET(
    BULK 'https://datalakexx.dfs.core.windows.net/fsxx/products.csv',
    FORMAT = 'CSV',
    PARSER_VERSION='2.0',
    HEADER_ROW = TRUE
  ) AS [result]
GROUP BY Category;
```

7. Run the modified query, which should return a resultset that contains the number products in each category, like this:

Category	**ProductCount**
Bib Shorts	3
Bike Racks	1
...	...

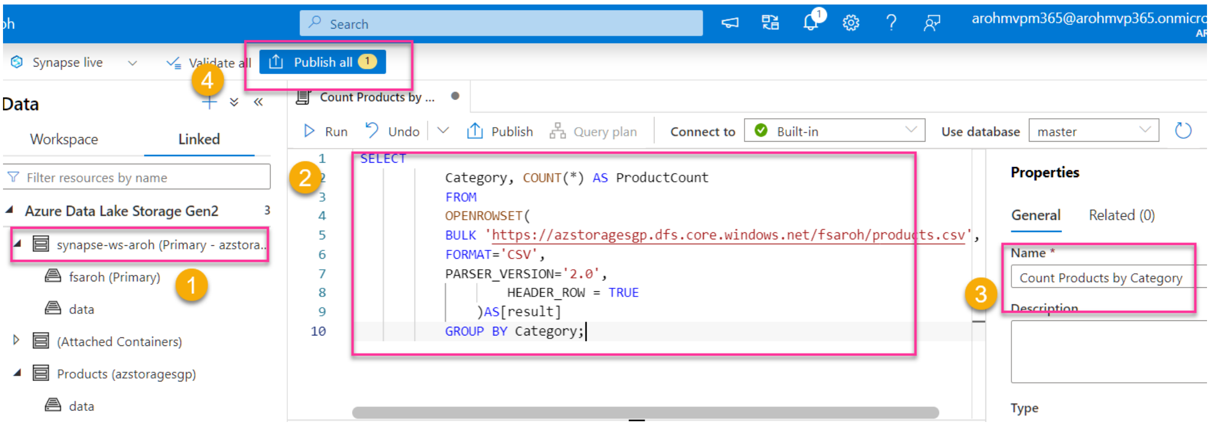


```sql
-- This is auto-generated code
SELECT
    Category, COUNT(*) AS ProductCount
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION='2.0',
        HEADER_ROW = TRUE
    ) AS [result]
GROUP BY Category;
```

Category	ProductCount
Bib-Shorts	3
Bike Racks	1
Bike Stands	1
Bottles and Cages	3
Bottom Brackets	3

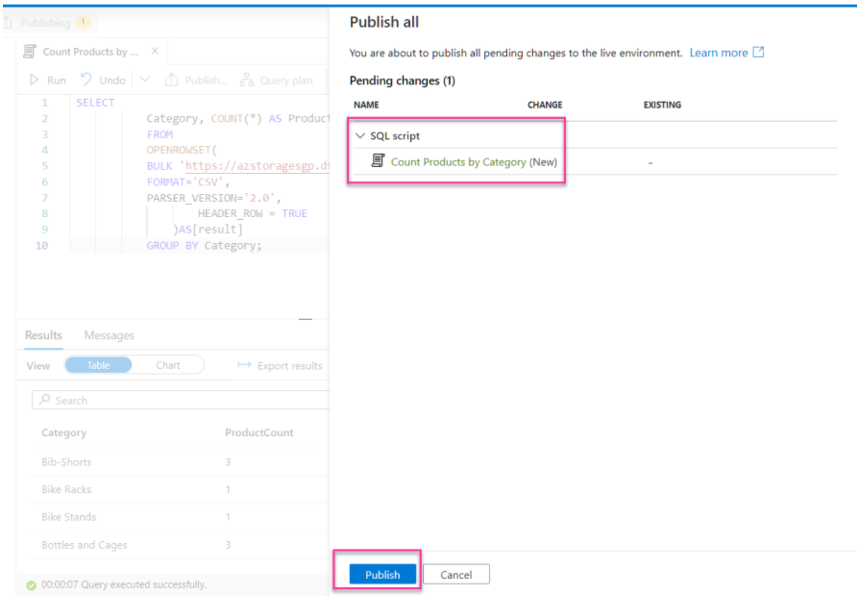
8. 00:00:06 Query executed successfully

9. In the **Properties** pane for **SQL Script 1**, change the **Name** to **Count Products by Category**. Then in the toolbar, select **Publish** to save the script.



```sql
SELECT
        Category, COUNT(*) AS ProductCount
        FROM
        OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT='CSV',
        PARSER_VERSION='2.0',
                HEADER_ROW = TRUE
            )AS[result]
        GROUP BY Category;
```

Name *: Count Products by Category



```
SELECT
        Category, COUNT(*) AS Product
    FROM
    OPENROWSET(
    BULK 'https://azstoragesgp.d
    FORMAT='CSV',
    PARSER_VERSION='2.0',
            HEADER_ROW = TRUE
        )AS[result]
    GROUP BY Category;
```

Category	ProductCount
Bib-Shorts	3
Bike Racks	1
Bike Stands	1
Bottles and Cages	3

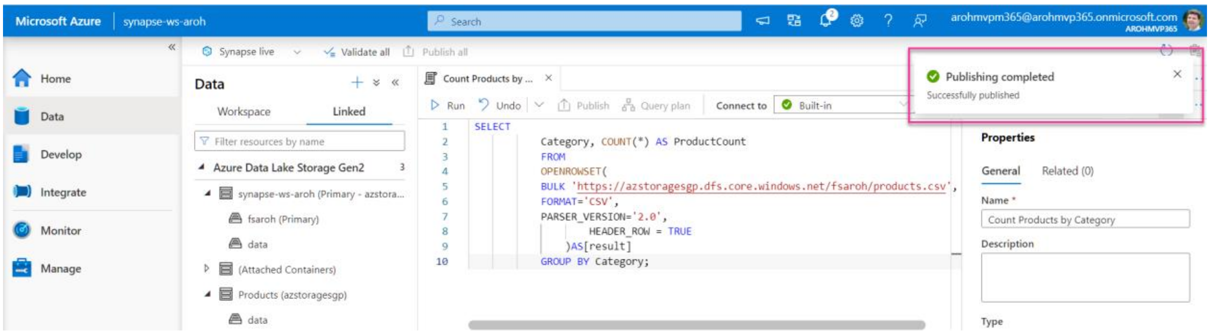
00:00:07 Query executed successfully.

Publish all

You are about to publish all pending changes to the live environment. Learn more

Pending changes (1)

NAME	CHANGE	EXISTING
SQL script		
Count Products by Category	(New)	-



```
SELECT
        Category, COUNT(*) AS ProductCount
    FROM
    OPENROWSET(
    BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
    FORMAT='CSV',
    PARSER_VERSION='2.0',
            HEADER_ROW = TRUE
        )AS[result]
    GROUP BY Category;
```

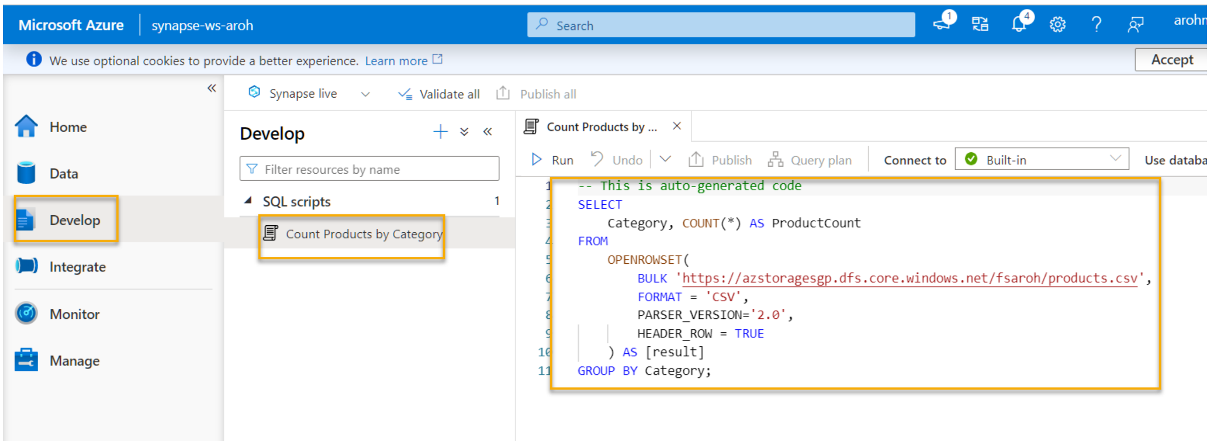
Publishing completed

Successfully published

Name *: Count Products by Category

10. Close the **Count Products by Category** script pane.

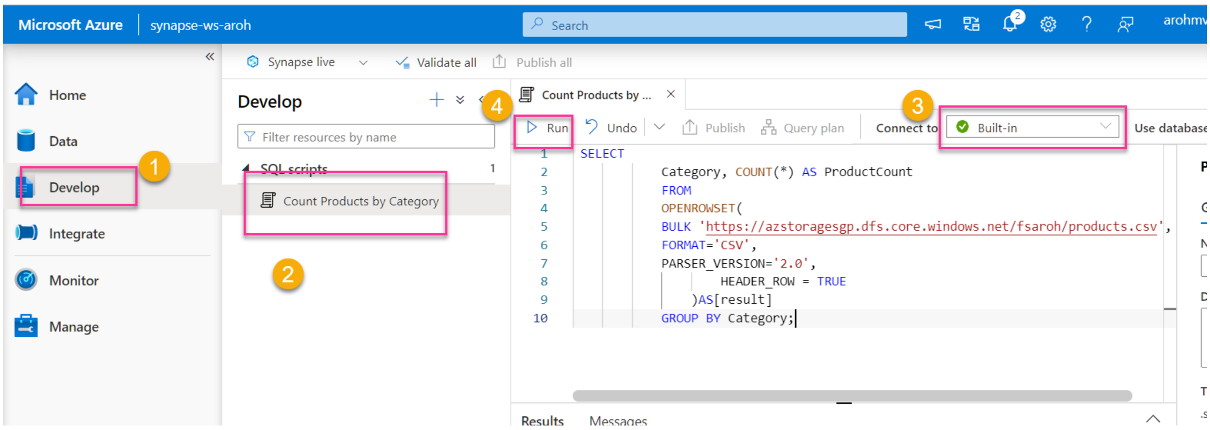
11. In Synapse Studio, select the **Develop** page, and notice that your published **Count Products by Category** SQL script has been saved there.



We use optional cookies to provide a better experience. Learn more

```
-- This is auto-generated code
SELECT
    Category, COUNT(*) AS ProductCount
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION='2.0',
        HEADER_ROW = TRUE
    ) AS [result]
GROUP BY Category;
```

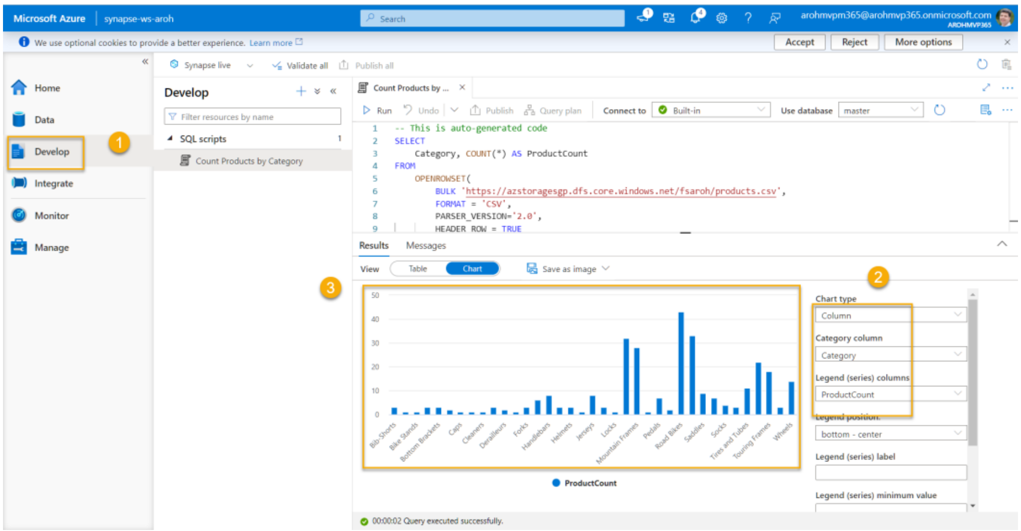
12. Select the **Count Products by Category** SQL script to reopen it. Then ensure that the script is connected to the **Built-in** SQL pool and run it to retrieve the product counts.



```
SELECT
        Category, COUNT(*) AS ProductCount
        FROM
        OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT='CSV',
        PARSER_VERSION='2.0',
               HEADER_ROW = TRUE
            )AS[result]
        GROUP BY Category;
```

13. In the **Results** pane, select the **Chart** view, and then select the following settings for the chart:

    - **Chart type**: Column
    - **Category column**: Category
    - **Legend (series) columns**: ProductCount
    - **Legend position**: bottom - center
    - **Legend (series) label**: *Leave blank*
    - **Legend (series) minimum value**: *Leave blank*
    - **Legend (series) maximum**: *Leave blank*
    - **Category label**: *Leave blank*

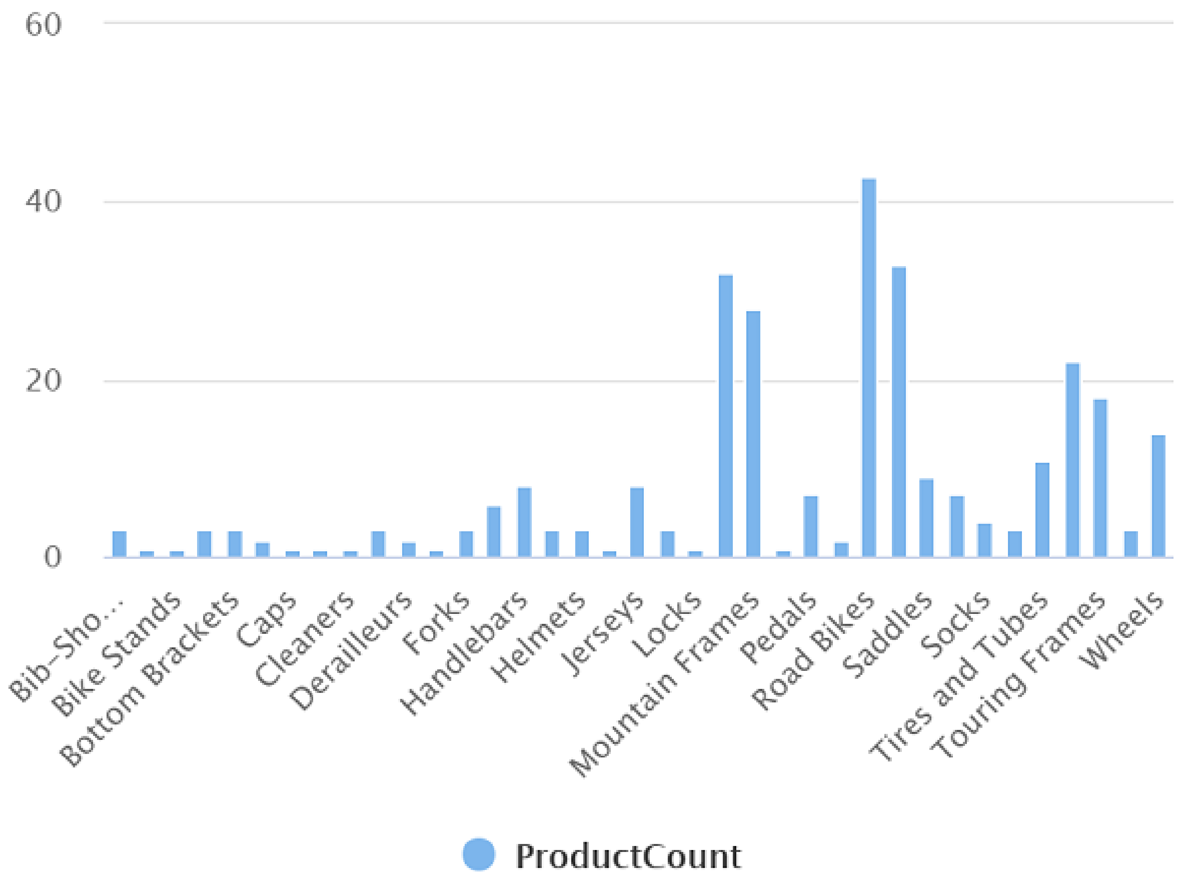


```
-- This is auto-generated code
SELECT
    Category, COUNT(*) AS ProductCount
FROM
    OPENROWSET(
        BULK 'https://azstoragesgp.dfs.core.windows.net/fsaroh/products.csv',
        FORMAT = 'CSV',
        PARSER_VERSION='2.0',
        HEADER_ROW = TRUE
```

Chart type: Column
Category column: Category
Legend (series) columns: ProductCount
Legend position: bottom - center
Legend (series) label:
Legend (series) minimum value:

00:00:02 Query executed successfully.

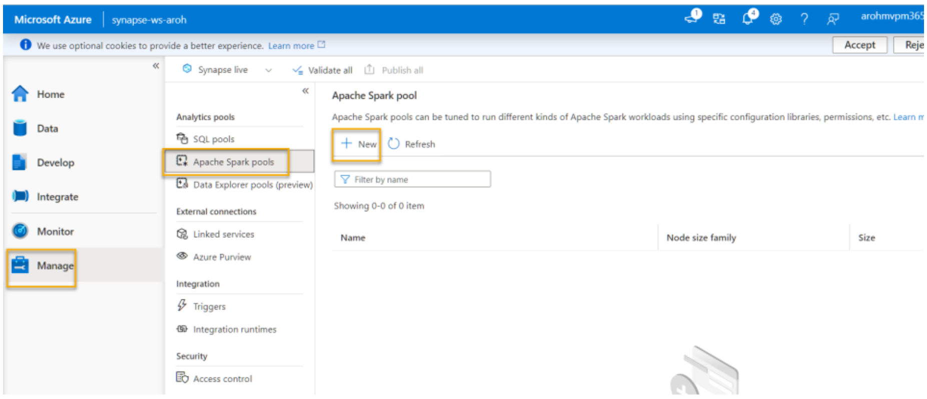
The resulting chart should resemble this:

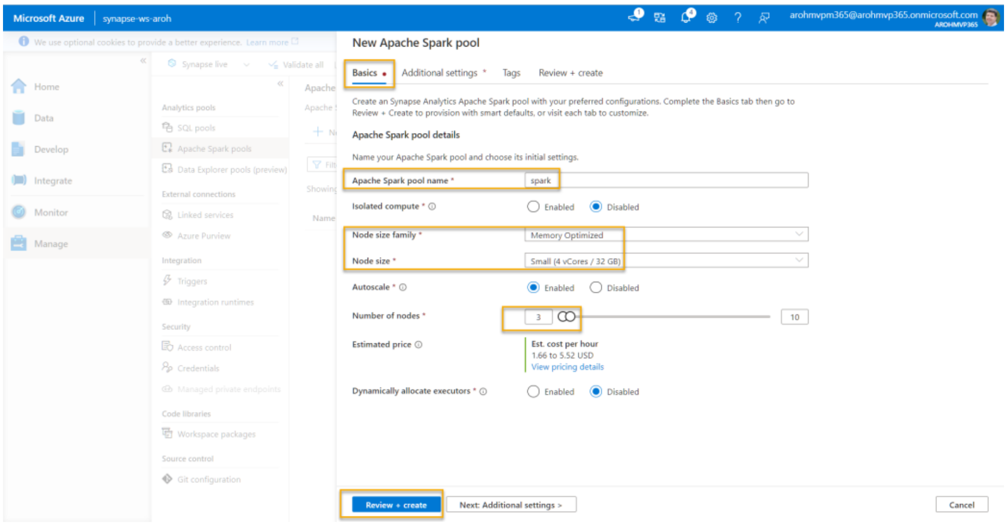


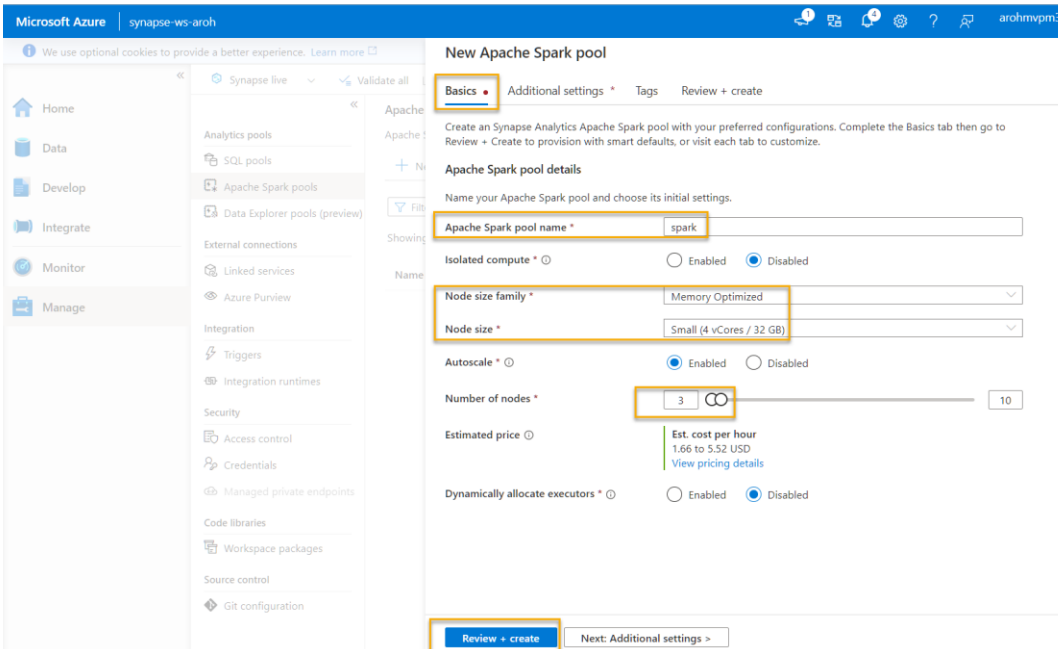
Use a Spark pool to analyze data

While **SQL is a common language for querying structured datasets**, many data analysts find languages like **Python useful to explore and prepare data** for analysis. In Azure Synapse Analytics, you can run Python (and other) code in a *Spark pool*; which uses a distributed data processing engine based on Apache Spark.

1. In Synapse Studio, select the **Manage** page.
2. Select the **Apache Spark pools** tab, and then use the **+ New** icon to create a new Spark pool with the following settings:
   - **Apache Spark pool name**: spark
   - **Node size family**: Memory Optimized
   - **Node size**: Small (4 vCores / 32 GB)
   - **Autoscale**: Enabled
   - **Number of nodes** 3----3







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 * spark

Isolated compute * ○ Enabled ● Disabled

Node size family * Memory Optimized

Node size * Small (4 vCores / 32 GB)

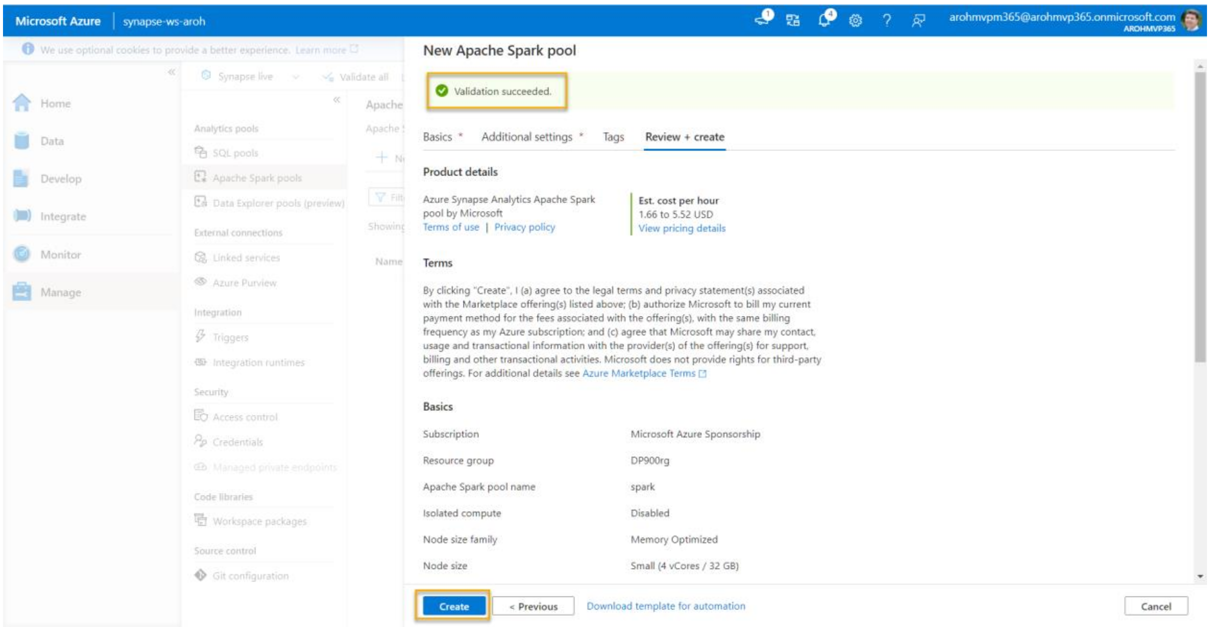
Autoscale * ● Enabled ○ Disabled

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

3. Review and create the Spark pool, and then wait for it to deploy (which may take a few minutes).



New Apache Spark pool

Validation succeeded.

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

Product details

Azure Synapse Analytics Apache Spark pool by Microsoft
Terms of use | Privacy policy

Est. cost per hour
1.66 to 5.52 USD
View pricing details

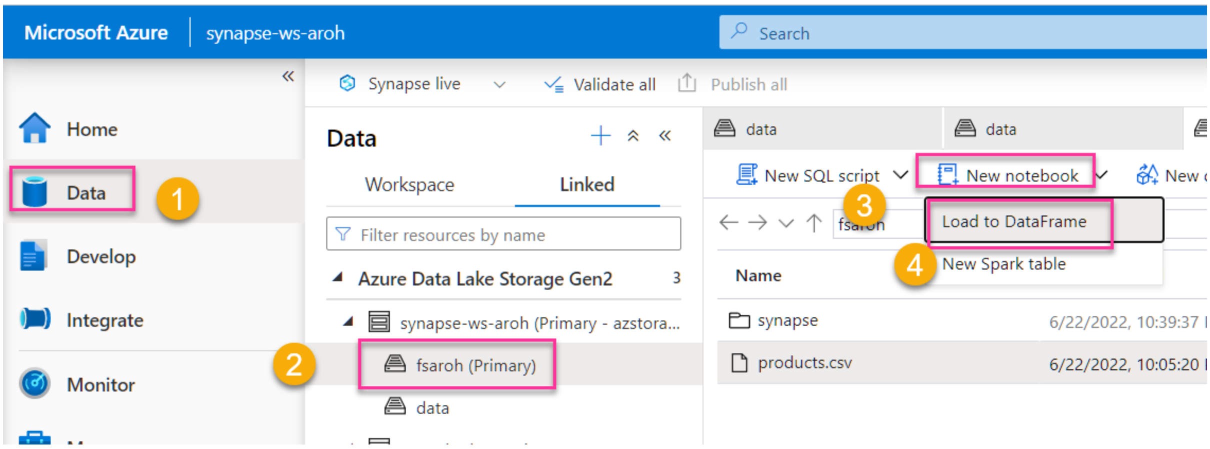
Terms

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see Azure Marketplace Terms

Basics

Subscription	Microsoft Azure Sponsorship
Resource group	DP900rg
Apache Spark pool name	spark
Isolated compute	Disabled
Node size family	Memory Optimized
Node size	Small (4 vCores / 32 GB)

4. When the Spark pool has been deployed, in Synapse Studio, on the **Data** page, browse to the file system for your Synapse workspace. Then right-click **products.csv**, point to **New notebook**, and select **Load to DataFrame**.



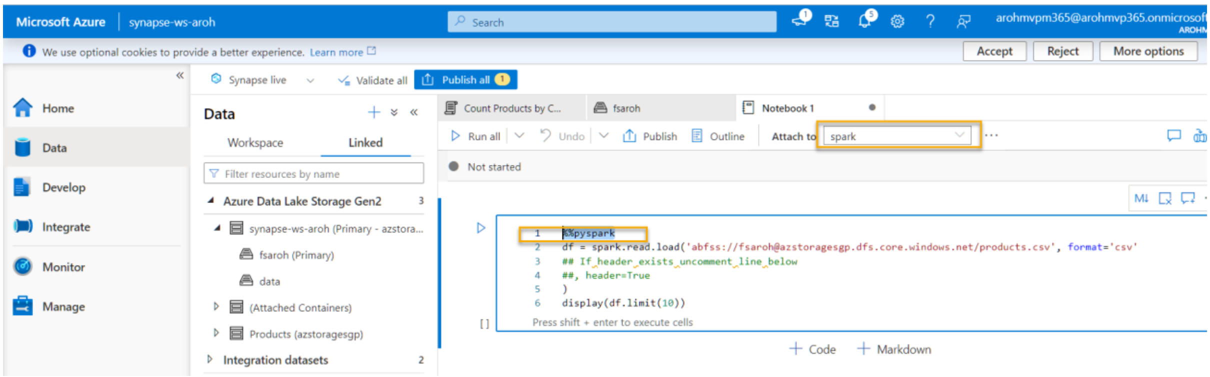
5. In the **Notebook 1** pane that opens, in the **Attach to** list, select the **spark** Spark pool to created previously and ensure that the **Language** is set to **PySpark (Python)**.

6. Review the code in the first (and only) cell in the notebook, which should look like this:

PythonCopy

```
%%pyspark

df = spark.read.load('abfss://fsxx@datalakexx.dfs.core.windows.net/products.csv', format='csv'

## If header exists uncomment line below

##, header=True

)

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

7. Use the ▷ icon to the left of the code cell to run it, and wait for the results. The first time you run a cell in a notebook, the Spark pool is started - so it may take a minute or so to return any results.

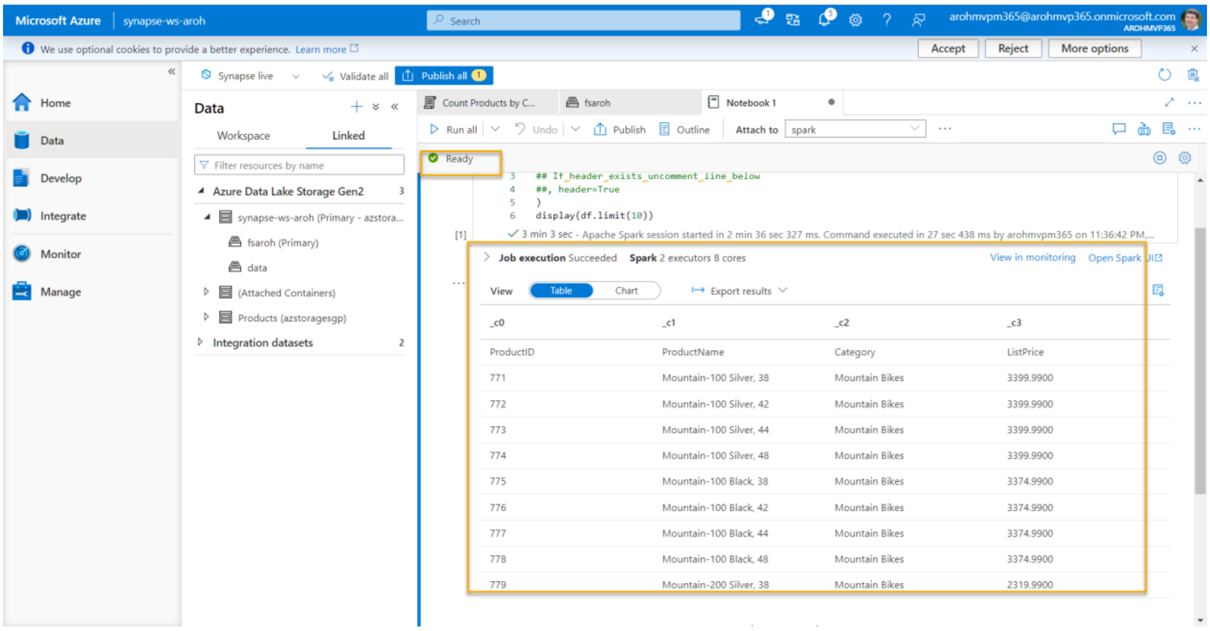


Not started

```
%%pyspark
df = spark.read.load('abfss://fsaroh@azstoragesgp.dfs.core.windows.net/products.csv', format='csv'
## If header exists uncomment line below
##, header=True
)
display(df.limit(10))
```

8. Eventually, the results should appear below the cell, and they should be similar to this:

c0	***c1***	***c2***
ProductID	ProductName	Category
771	Mountain-100 Silver, 38	Mountain Bikes
772	Mountain-100 Silver, 42	Mountain Bikes
...	...	...

9. 

Ready

```
## If header exists uncomment line below
##, header=True
)
display(df.limit(10))
```

[1] ✓ 3 min 3 sec - Apache Spark session started in 2 min 36 sec 327 ms. Command executed in 27 sec 438 ms by arohmvpm365 on 11:36:42 PM,...

Job execution Succeeded Spark 2 executors 8 cores

_c0	_c1	_c2	_c3
ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.9900
772	Mountain-100 Silver, 42	Mountain Bikes	3399.9900
773	Mountain-100 Silver, 44	Mountain Bikes	3399.9900
774	Mountain-100 Silver, 48	Mountain Bikes	3399.9900
775	Mountain-100 Black, 38	Mountain Bikes	3374.9900
776	Mountain-100 Black, 42	Mountain Bikes	3374.9900
777	Mountain-100 Black, 44	Mountain Bikes	3374.9900
778	Mountain-100 Black, 48	Mountain Bikes	3374.9900
779	Mountain-200 Silver, 38	Mountain Bikes	2319.9900

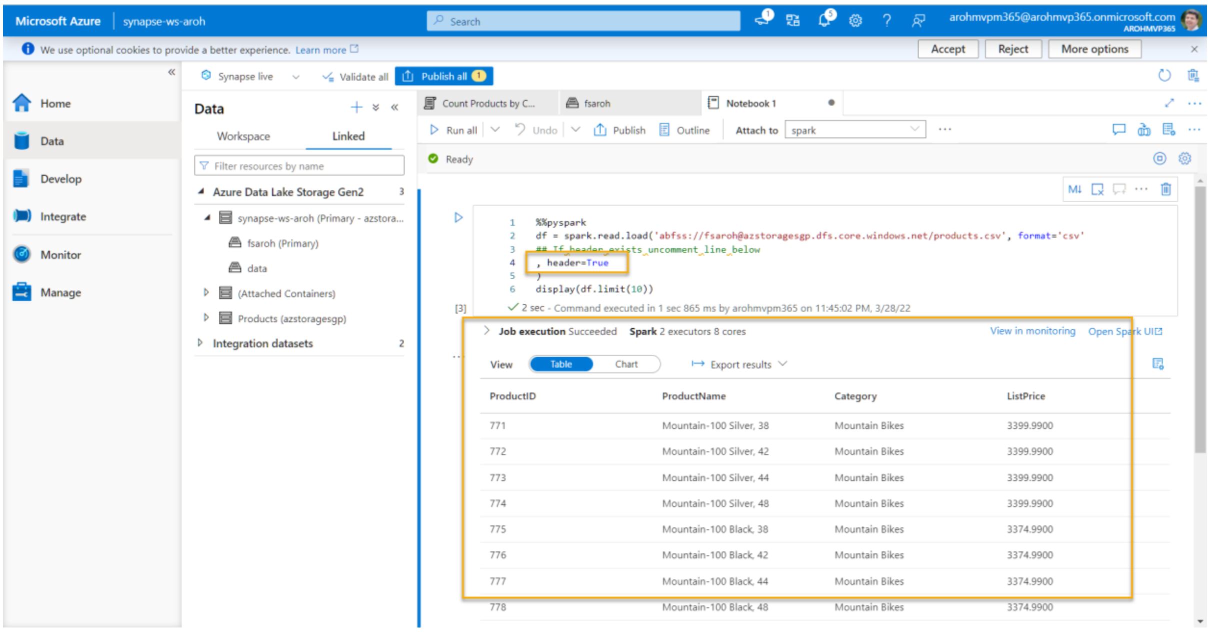
10. Uncomment the *,header=True* line (because the products.csv file has the column headers in the first line), so your code looks like this:

PythonCopy

```
%%pyspark

df = spark.read.load('abfss://fsxx@datalakexx.dfs.core.windows.net/products.csv', format='csv'

## If header exists uncomment line below

, header=True

)

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

11. Rerun the cell and verify that the results look like this:

ProductID	ProductName	Category
771	Mountain-100 Silver, 38	Mountain Bikes
772	Mountain-100 Silver, 42	Mountain Bikes
...	...	...



```
%%pyspark
df = spark.read.load('abfss://fsaroh@azstoragesgp.dfs.core.windows.net/products.csv', format='csv'
## If header exists uncomment line below
, header=True
)
display(df.limit(10))
```

[3] 2 sec - Command executed in 1 sec 865 ms by arohmvpm365 on 11:45:02 PM, 3/28/22

Job execution Succeeded Spark 2 executors 8 cores

ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.9900
772	Mountain-100 Silver, 42	Mountain Bikes	3399.9900
773	Mountain-100 Silver, 44	Mountain Bikes	3399.9900
774	Mountain-100 Silver, 48	Mountain Bikes	3399.9900
775	Mountain-100 Black, 38	Mountain Bikes	3374.9900
776	Mountain-100 Black, 42	Mountain Bikes	3374.9900
777	Mountain-100 Black, 44	Mountain Bikes	3374.9900
778	Mountain-100 Black, 48	Mountain Bikes	3374.9900

12. Notice that running the cell again takes less time, because the Spark pool is already started.

13. Under the results, use the **+ Code** icon to add a new code cell to the notebook.

14. In the new empty code cell, add the following code:

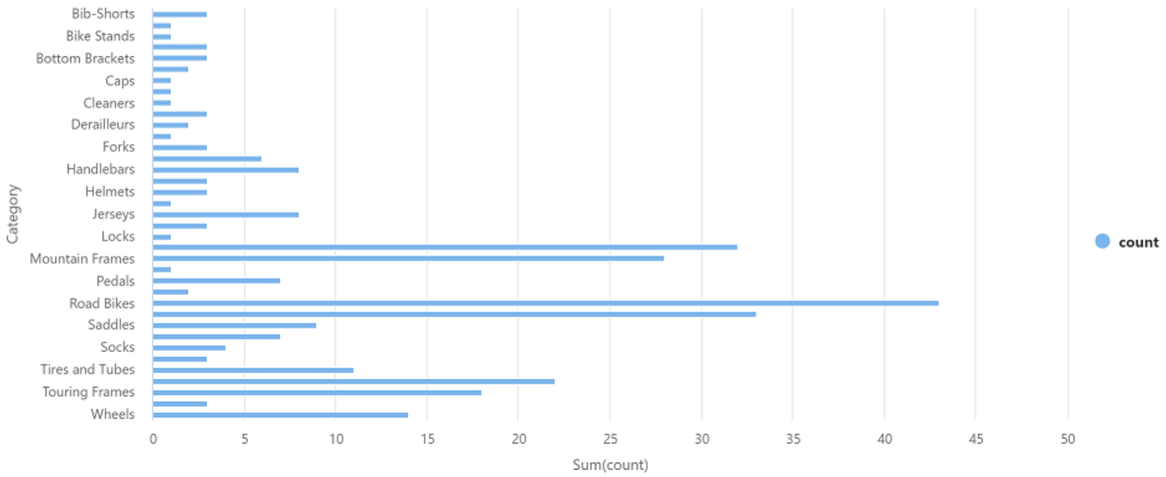
PythonCopy

```
df_counts = df.groupby(df.Category).count()

display(df_counts)
```

15. Run the new code cell by clicking its ▷ icon, and review the results, which should look similar to this:

Category	count
Headsets	3
Wheels	14
...	...

16. In the results output for the cell, select the **Chart** view. The resulting chart should resemble this:



17. Close the **Notebook 1** pane and discard your changes.

Delete Azure resources

If you've finished exploring Azure Synapse Analytics, you should delete the resources you've created to avoid unnecessary Azure costs.

1. Close the Synapse Studio browser tab and return to the Azure portal.
2. On the Azure portal, on the **Home** page, select **Resource groups**.
3. Select the resource group for your Synapse Analytics workspace (not the managed resource group), and verify that it contains the Synapse workspace, storage account, and Spark pool for your workspace.
4. At the top of the **Overview** page for your resource group, select **Delete resource group**.
5. Enter the resource group name to confirm you want to delete it, and select **Delete**.

After a few minutes, your Azure Synapse workspace and the managed workspace associated with it will be deleted.

Explore Azure Stream Analytics

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/blob/master/Instructions/Labs/dp900-05-stream-lab.md

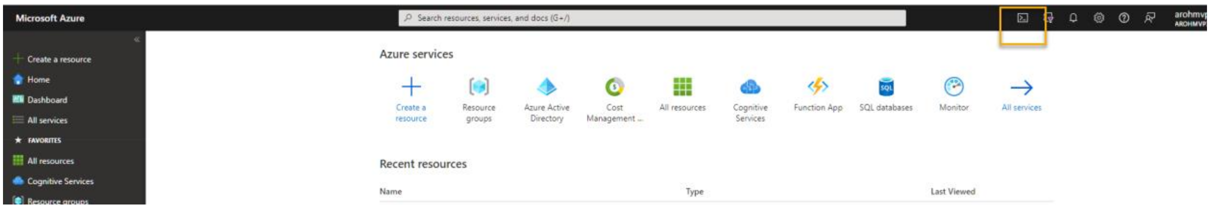
In this exercise you'll provision an Azure Cosmos DB database in your Azure subscription, and explore the various ways you can use it to store non-relational data.

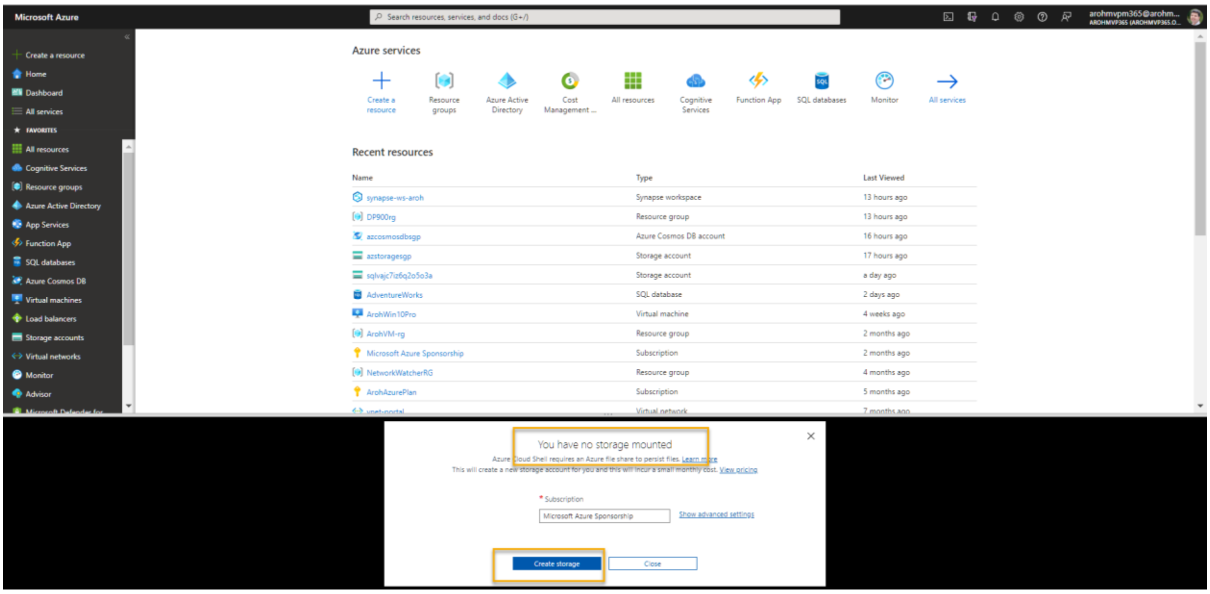
Note: The exercise is part of a module on Microsoft Learn, and includes an option to use a *sandbox* Azure subscription. However, if you are completing this exercise as part of an instructor-led class, you should use the Azure subscription provided as part of the class instead of the sandbox.

Before starting the exercise on Microsoft Learn, you'll need to prepare a cloud shell environment for your Azure subscription.

1. Sign into your Azure subscription in the Azure portal at https://portal.azure.com, using your Azure subscription credentials.

2. Use the **[>_]** button to the right of the search bar at the top of the page to create a new Cloud Shell in the Azure portal, selecting a ***Bash*** environment and creating storage if prompted. The cloud shell provides a command line interface in a pane at the bottom of the Azure portal, as shown here:





Recent resources

Name	Type	Last Viewed
synapse-ws-aroh	Synapse workspace	13 hours ago
DP900rg	Resource group	13 hours ago
azcosmosdbsgp	Azure Cosmos DB account	16 hours ago
azstoragesgp	Storage account	17 hours ago
sqlvajc7iz6q2o5o3a	Storage account	a day ago
AdventureWorks	SQL database	2 days ago
ArohWin10Pro	Virtual machine	4 weeks ago
ArohVM-rg	Resource group	2 months ago
Microsoft Azure Sponsorship	Subscription	2 months ago
NetworkWatcherRG	Resource group	4 months ago
ArohAzurePlan	Subscription	5 months ago
vnet-portal	Virtual network	7 months ago

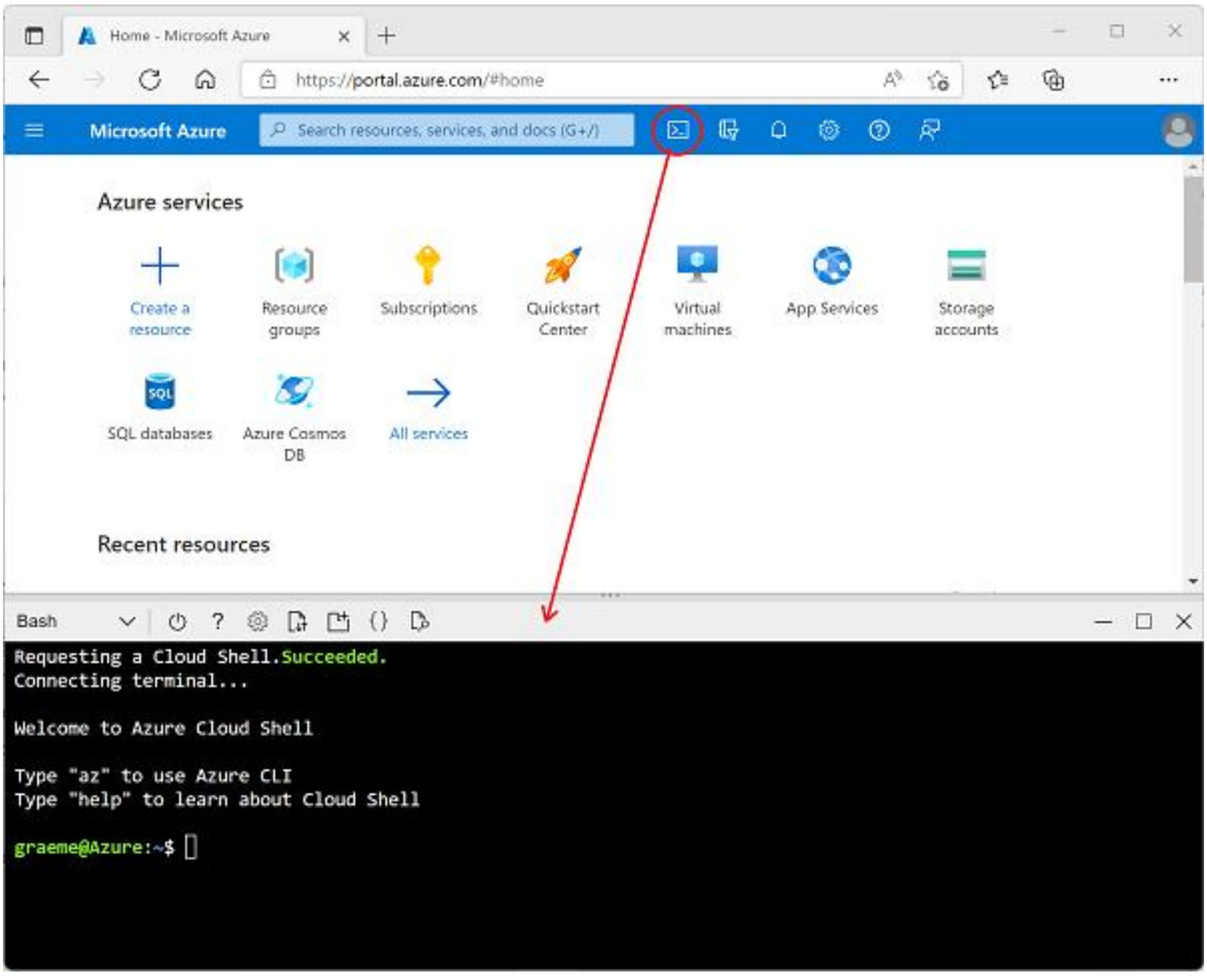
You have no storage mounted

Azure Cloud Shell requires an Azure file share to persist files. Learn more
This will create a new storage account for you and this will incur a small monthly cost. View pricing

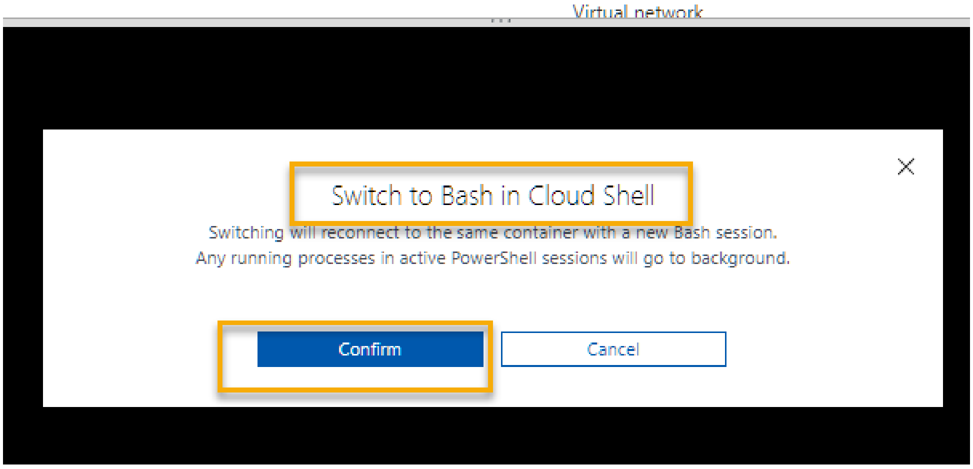
* Subscription: Microsoft Azure Sponsorship

Show advanced settings

Create storage | Close



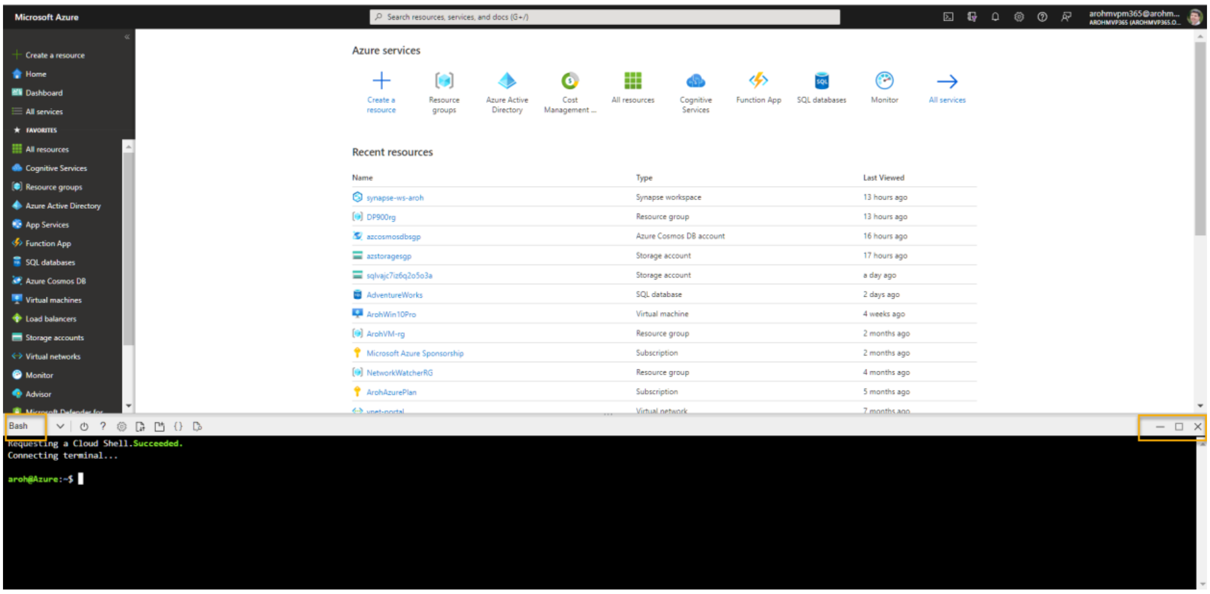
```
Requesting a Cloud Shell.Succeeded.
Connecting terminal...

Welcome to Azure Cloud Shell

Type "az" to use Azure CLI
Type "help" to learn about Cloud Shell

graeme@Azure:~$
```



Switch to Bash in Cloud Shell

Switching will reconnect to the same container with a new Bash session.
Any running processes in active PowerShell sessions will go to background.

3. Note that you can resize the cloud shell by dragging the separator bar at the top of the pane, or by using the —, ☐, and **X** icons at the top right of the pane to minimize, maximize, and close the pane. For more information about using the Azure Cloud Shell, see the Azure Cloud Shell documentation.



Recent resources

Name	Type	Last Viewed
synapse-ws-aroh	Synapse workspace	13 hours ago
DP900rg	Resource group	13 hours ago
azcosmosdbsgp	Azure Cosmos DB account	16 hours ago
azstoragesgp	Storage account	17 hours ago
sqlvajc7iz6q2o5o3a	Storage account	a day ago
AdventureWorks	SQL database	2 days ago
ArohWin10Pro	Virtual machine	4 weeks ago
ArohVM-rg	Resource group	2 months ago
Microsoft Azure Sponsorship	Subscription	2 months ago
NetworkWatcherRG	Resource group	4 months ago
ArohAzurePlan	Subscription	5 months ago
vnet-portal	Virtual network	7 months ago

```
Requesting a Cloud Shell.Succeeded.
Connecting terminal...

aroh@Azure:~$
```

4. Now you're ready to complete the exercise on Microsoft Learn - just be use to use the cloud shell in your Azure portal instead of the (blank) one in the Learn module (which is provided for self-paced learners using a sandbox subscription).

Use the link below to open the exercise on Microsoft Learn.

https://docs.microsoft.com/en-gb/learn/modules/explore-fundamentals-stream-processing/5-exercise-stream-analytics#create-azure-resources

Exercise: Analyze streaming data

Now it's your opportunity to explore **Azure Stream Analytics** in a sample solution that aggregates streaming data from a **simulated IoT device**.

This exercise can be completed using a free **Microsoft Learn *sandbox* subscription**, which provides an **Azure subscription and a Cloud Shell environment** integrated into this page. The **sandbox subscription** will be created when you click the button above, and automatically deleted when you complete this module.

Tip

If you prefer, you can use your own Azure subscription. To do this, before you start:

1. Sign into your Azure subscription in the **Azure portal**.

2. Ensure you are working in the directory containing your subscription - indicated at the top right under your user ID. If not, select the user icon and switch directory.

3. Use the **[>_]** button to the right of the search bar at the top of the page to create a new Cloud Shell in the Azure portal, creating storage if prompted and selecting a *Bash* environment.

4. Complete the exercise steps below in the Cloud Shell pane in the Azure portal.

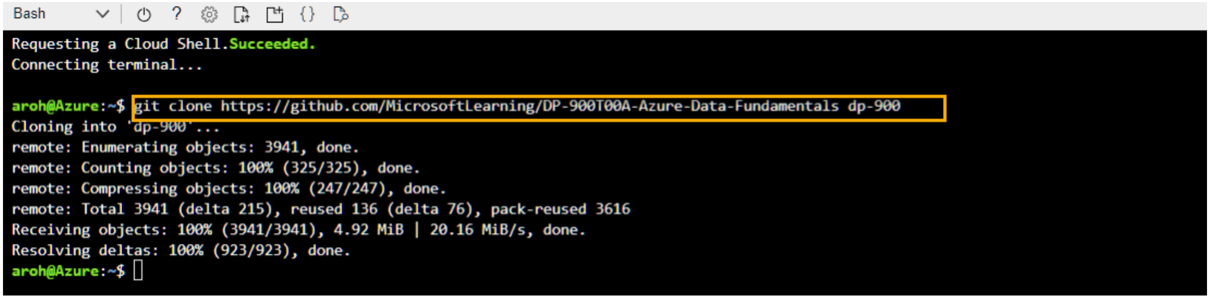
For more information about using the Azure Cloud Shell, see the **Azure Cloud Shell documentation**

Create Azure resources

1. In the Azure Cloud Shell, enter the following command to download the files you'll need for this exercise.

Copy git clone https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals dp-900

```
git clone https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals dp-900
```

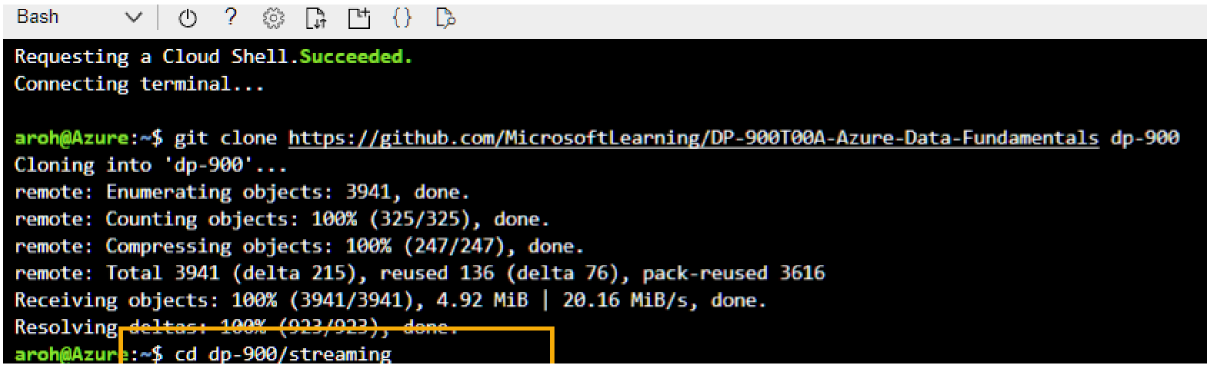


```
Requesting a Cloud Shell.Succeeded.
Connecting terminal...

aroh@Azure:~$ git clone https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals dp-900
Cloning into 'dp-900'...
remote: Enumerating objects: 3941, done.
remote: Counting objects: 100% (325/325), done.
remote: Compressing objects: 100% (247/247), done.
remote: Total 3941 (delta 215), reused 136 (delta 76), pack-reused 3616
Receiving objects: 100% (3941/3941), 4.92 MiB | 20.16 MiB/s, done.
Resolving deltas: 100% (923/923), done.
aroh@Azure:~$
```

2. Wait for the command to complete, and then enter the following command to change the current directory to the folder containing the files for this exercise.

Copy cd dp-900/streaming

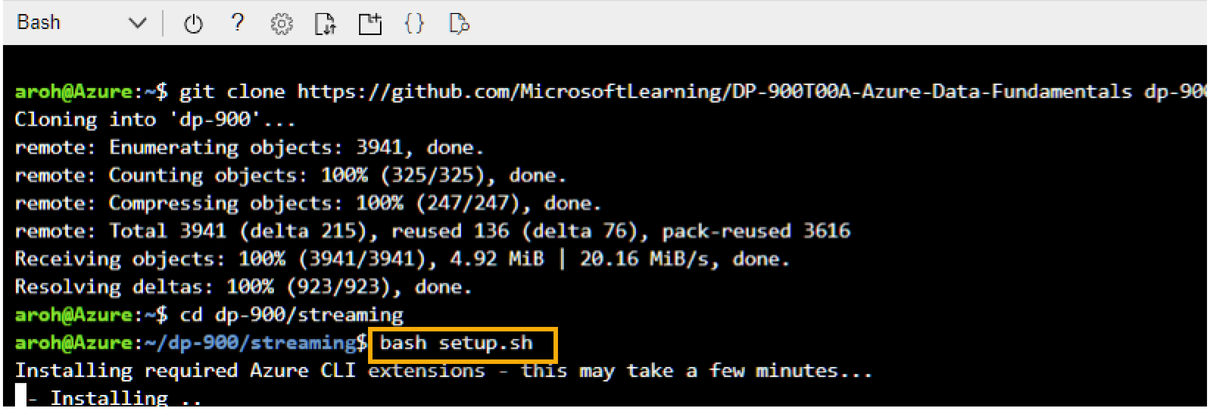
```
cd dp-900/streaming
```



```
Requesting a Cloud Shell.Succeeded.
Connecting terminal...

aroh@Azure:~$ git clone https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals dp-900
Cloning into 'dp-900'...
remote: Enumerating objects: 3941, done.
remote: Counting objects: 100% (325/325), done.
remote: Compressing objects: 100% (247/247), done.
remote: Total 3941 (delta 215), reused 136 (delta 76), pack-reused 3616
Receiving objects: 100% (3941/3941), 4.92 MiB | 20.16 MiB/s, done.
Resolving deltas: 100% (923/923), done.
aroh@Azure:~$ cd dp-900/streaming
```

3. Enter the following command to run a script that creates the Azure resources you will need in this exercise.

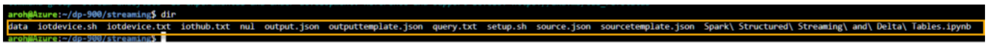
Copy bash setup.sh



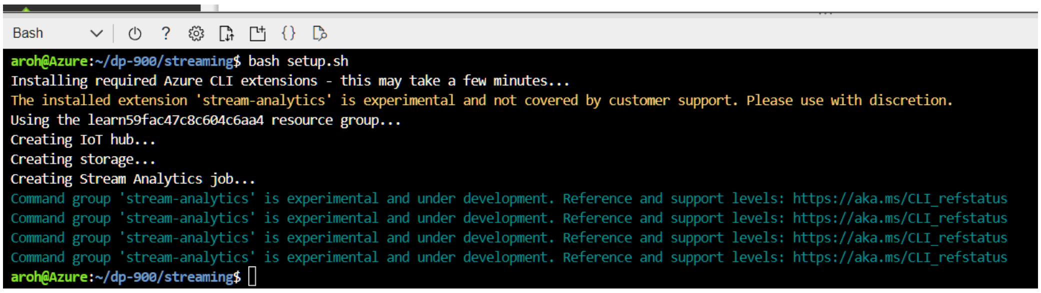
```
aroh@Azure:~$ git clone https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals dp-90
Cloning into 'dp-900'...
remote: Enumerating objects: 3941, done.
remote: Counting objects: 100% (325/325), done.
remote: Compressing objects: 100% (247/247), done.
remote: Total 3941 (delta 215), reused 136 (delta 76), pack-reused 3616
Receiving objects: 100% (3941/3941), 4.92 MiB | 20.16 MiB/s, done.
Resolving deltas: 100% (923/923), done.
aroh@Azure:~$ cd dp-900/streaming
aroh@Azure:~/dp-900/streaming$ bash setup.sh
Installing required Azure CLI extensions - this may take a few minutes...
 - Installing ..
```

Wait as the script runs and performs the following actions:

1. Installs the Azure CLI extensions needed to create resources (*you can ignore any warnings about experimental extensions*)
2. Identifies the Azure resource group provided for this exercise, which will have a name similar to **learn-*xxxxxxxxxxxxxxxxx*...**.
3. Creates an ***Azure IoT Hub* resource**, which will be used to **receive a stream of data from a simulated device.**
4. Creates a ***Azure Storage Account***, which will be used to **store processed data**.
5. Creates a *Azure Stream Analytics* job, which will process **the incoming device data in real-time, and write the results to the storage account.**



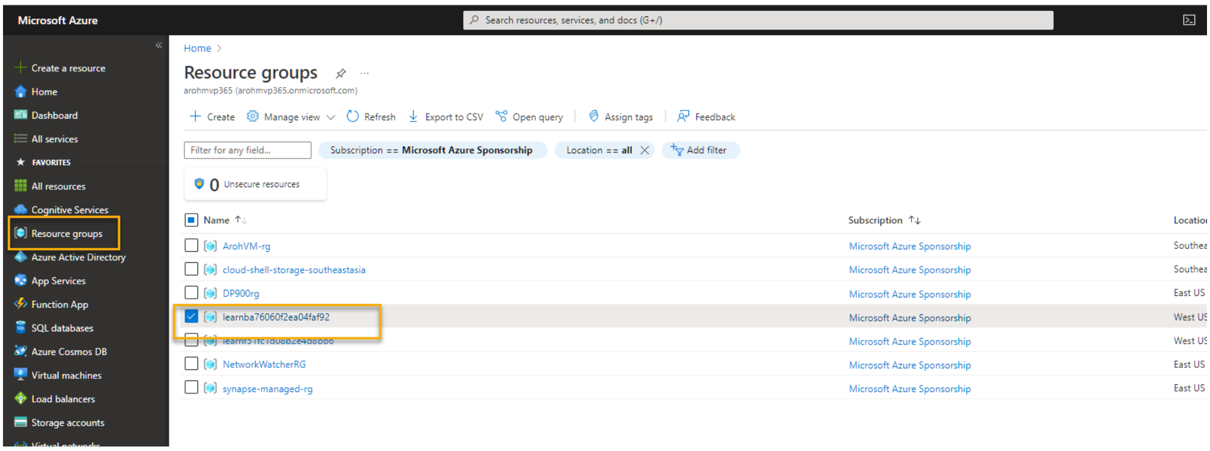
```
aroh@Azure:~/dp-900/streaming$ dir
data  iotdevice.sh  iotdevice.txt  iothub.txt  nul  output.json  outputtemplate.json  query.txt  setup.sh  source.json  sourcetemplate.json  Spark\ Structured\ Streaming\ and\ Delta\ Tables.ipynb
aroh@Azure:~/dp-900/streaming$
```



```
aroh@Azure:~/dp-900/streaming$ bash setup.sh
Installing required Azure CLI extensions - this may take a few minutes...
The installed extension 'stream-analytics' is experimental and not covered by customer support. Please use with discretion.
Using the learn59fac47c8c604c6aa4 resource group...
Creating IoT hub...
Creating storage...
Creating Stream Analytics job...
Command group 'stream-analytics' is experimental and under development. Reference and support levels: https://aka.ms/CLI_refstatus
Command group 'stream-analytics' is experimental and under development. Reference and support levels: https://aka.ms/CLI_refstatus
Command group 'stream-analytics' is experimental and under development. Reference and support levels: https://aka.ms/CLI_refstatus
Command group 'stream-analytics' is experimental and under development. Reference and support levels: https://aka.ms/CLI_refstatus
aroh@Azure:~/dp-900/streaming$
```

Explore the Azure resources

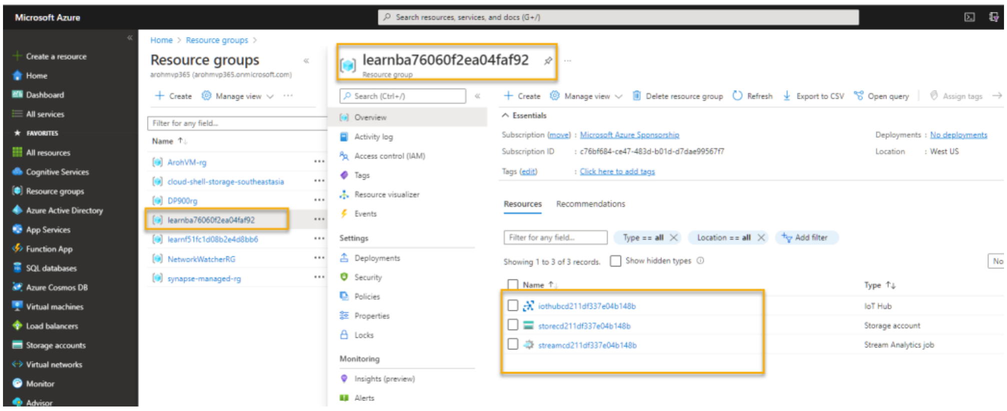
1. In the Azure portal, on the home page, select **Resource groups** to see the resource groups in your subscription. This should include the **learn-*xxxxxxxxxxxxxxxxx*...** resource group identified by the setup script.



Name	Subscription	Locatio
ArohVM-rg	Microsoft Azure Sponsorship	Southea
cloud-shell-storage-southeastasia	Microsoft Azure Sponsorship	Southea
DP900rg	Microsoft Azure Sponsorship	East US
☑ learnba76060f2ea04faf92	Microsoft Azure Sponsorship	West US
learnf51fc1d08b2e4d8bb6	Microsoft Azure Sponsorship	West US
NetworkWatcherRG	Microsoft Azure Sponsorship	East US
synapse-managed-rg	Microsoft Azure Sponsorship	East US

2. Select the **learn-*xxxxxxxxxxxxxxxxxxx*...** resource group, and review the resources it contains, which should include:

   - An *IoT Hub* named **iothub*xxxxxxxxxxxxx***, which is used to receive incoming device data.
   - A *Storage account* named **store*xxxxxxxxxxxxx***, to which the data processing results will be written.
   - A *Stream Analytics job* named **stream*xxxxxxxxxxxxx***, which will be used to process streaming data.



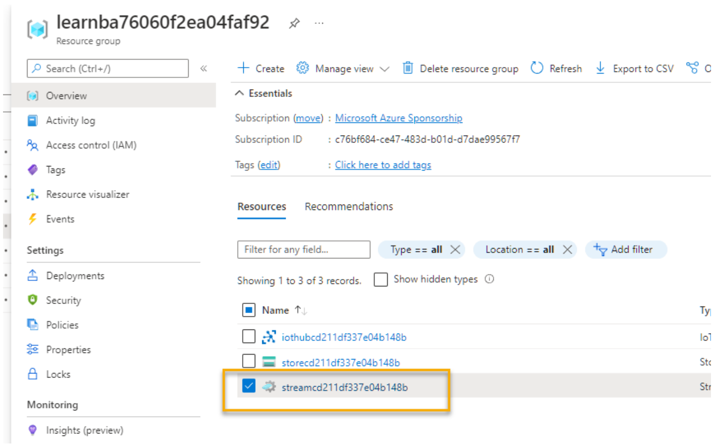
Name	Type
iothubcd211df337e04b148b	IoT Hub
storecd211df337e04b148b	Storage account
streamcd211df337e04b148b	Stream Analytics job

If all three of these resources are not listed, click the ↻ **Refresh** button until they appear.

Note

If you are using the learn sandbox, the resource group may also contain a second *Storage account* named **cloudshell*xxxxxxxx***, which is used to store data for the Azure Cloud Shell you used to run the setup script.

3. Select the **stream*xxxxxxxxxxxxx*** Stream Analytics job



Subscription (move) : Microsoft Azure Sponsorship

Subscription ID : c76bf684-ce47-483d-b01d-d7dae99567f7

Tags (edit) : Click here to add tags

Showing 1 to 3 of 3 records.

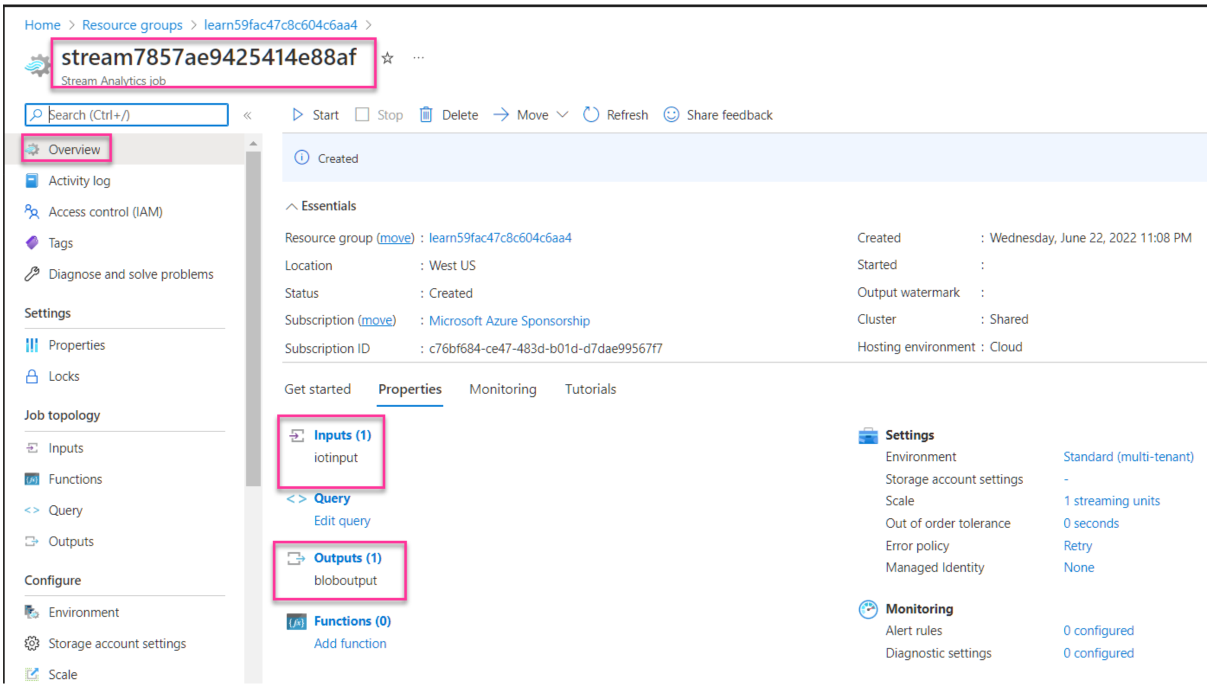
□ iothubcd211df337e04b148b Io1

□ storecd211df337e04b148b Sto

☑ streamcd211df337e04b148b Str

and view the information on its **Overview** page, noting the following details:

- The job has one *input* named **iotinput**, and one *output* named **bloboutput**. These reference the IoT Hub and Storage account created by the setup script.
- The job has a *query*, which reads data from the **iotinput** input, and aggregates it by counting the number of messages processed every 10 seconds; writing the results to the **bloboutput** output.



Created

Resource group (move) : learn59fac47c8c604c6aa4

Location : West US

Status : Created

Subscription (move) : Microsoft Azure Sponsorship

Subscription ID : c76bf684-ce47-483d-b01d-d7dae99567f7

Created : Wednesday, June 22, 2022 11:08 PM

Started :

Output watermark :

Cluster : Shared

Hosting environment : Cloud

Inputs (1)
iotinput

Query
Edit query

Outputs (1)
bloboutput

Functions (0)
Add function

Settings

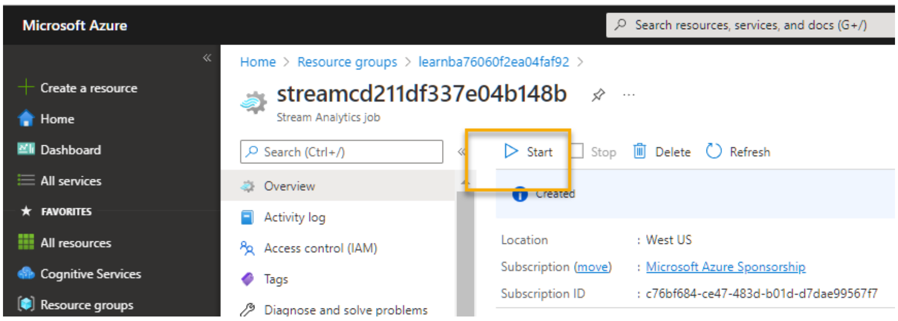
Environment	Standard (multi-tenant)
Storage account settings	-
Scale	1 streaming units
Out of order tolerance	0 seconds
Error policy	Retry
Managed Identity	None

Monitoring

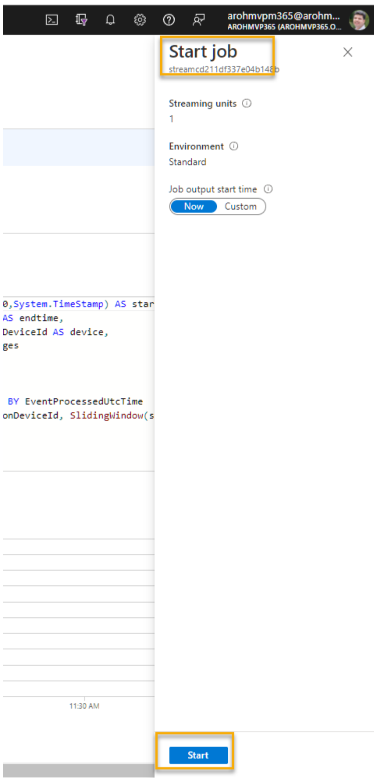
Alert rules	0 configured
Diagnostic settings	0 configured

Use the resources to analyze streaming data

1. At the top of the **Overview** page for the Stream Analytics job, select the ▷ **Start** button, and then in the **Start job** pane, select **Start** to start the job.
2. Wait for a notification that the streaming job started successfully.



Location : West US
Subscription (move) : Microsoft Azure Sponsorship
Subscription ID : c76bf684-ce47-483d-b01d-d7dae99567f7



Streaming units
1

Environment
Standard

Job output start time
Now

3. Switch back to the **Azure Cloud Shell**, and enter the following command to simulate a device that sends data to the IoT Hub.

Copy

```
bash iotdevice.sh
```

4. Wait for the simulation to start, which will be indicated by output like this:

Copy

```
Device simulation in progress: 6%|#    | 7/120 [00:08<02:21, 1.26s/it]
```

5. While the simulation is running, back in the Azure portal, return to the page for the **learn-*xxxxxxxxxxxxxxxxx...*** resource group, and select the **store*xxxxxxxxxxxxx*** storage account.

6. In the pane on the left of the storage account blade, select the **Containers** tab.

7. Open the **data** container.

8. In the **data** container, navigate through the folder hierarchy, which includes a folder for the current year, with subfolders for the month, day, and hour.

9. In the folder for the hour, select the file that has been created, which should have a name similar to **0_xxxxxxxxxxxxxxxx.json**.

10. On the page for the file, select **Edit**, and review the contents of the file; which should consist of a JSON record for each 10 second period, showing the number of messages received from IoT devices, like this:

Copy

```
{"starttime":"2021-10-23T01:02:13.2221657Z","endtime":"2021-10-23T01:02:23.2221657Z","device":"iotdevice","messages":2}

{"starttime":"2021-10-23T01:02:14.5366678Z","endtime":"2021-10-23T01:02:24.5366678Z","device":"iotdevice","messages":3}

{"starttime":"2021-10-23T01:02:15.7413754Z","endtime":"2021-10-23T01:02:25.7413754Z","device":"iotdevice","messages":4}

...
```

11. Use the ↻ **Refresh** button to refresh the file, noting that additional results are written to the file as Stream Analytics job processes the device data in real time as it is streamed from the device to the IoT Hub.

12. Return to the Azure Cloud Shell and wait for the device simulation to finish (it should run for around 3 minutes).

13. Back in the Azure portal, refresh the file one more time to see the full set of results that were produced during the simulation.

14. Return to the **learn-*xxxxxxxxxxxxxxxxx...*** resource group, and re-open the **stream*xxxxxxxxxxxxx*** Stream Analytics job.

15. At the top of the Stream Analytics job page, use the □ **Stop** button to stop the job, confirming when prompted.

Tip

If you are using your own Azure subscription, and you have finished exploring the streaming solution, you can delete the **learn-*xxxxxxxxxxxxxxxxx...*** resource group.

https://docs.microsoft.com/en-gb/learn/modules/explore-fundamentals-data-visualization/5-exercise-power-bi

Exercise – Visualize data with Power BI

Now it's your chance to explore **data modelling** and visualization with Microsoft Power BI.

Note

To complete this exercise, you will need a computer running Microsoft Windows.

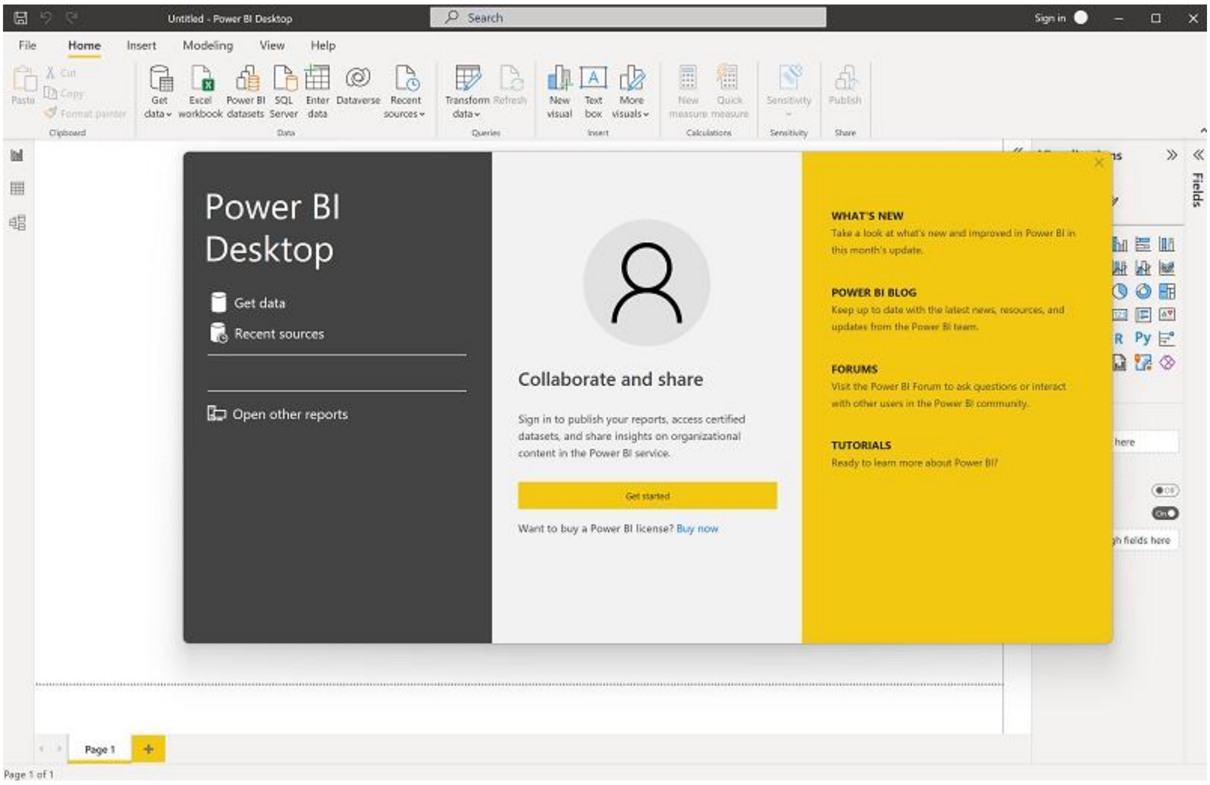
Install Power BI Desktop (if necessary)

If Microsoft Power BI Desktop is not already installed on your Windows computer, you can download and install it for free.

1. Download the Power BI Desktop installer from https://aka.ms/power-bi-desktop.
2. When the file has downloaded, open it, and use the setup wizard to install Power BI Desktop on your computer. This may take a few minutes.

Import data

1. Open Power BI Desktop. The application interface should look similar to this:



Power BI Desktop

Get data

Recent sources

Open other reports

Collaborate and share

Sign in to publish your reports, access certified datasets, and share insights on organizational content in the Power BI service.

Get started

Want to buy a Power BI license? Buy now

WHAT'S NEW

Take a look at what's new and improved in Power BI in this month's update.

POWER BI BLOG

Keep up to date with the latest news, resources, and updates from the Power BI team.

FORUMS

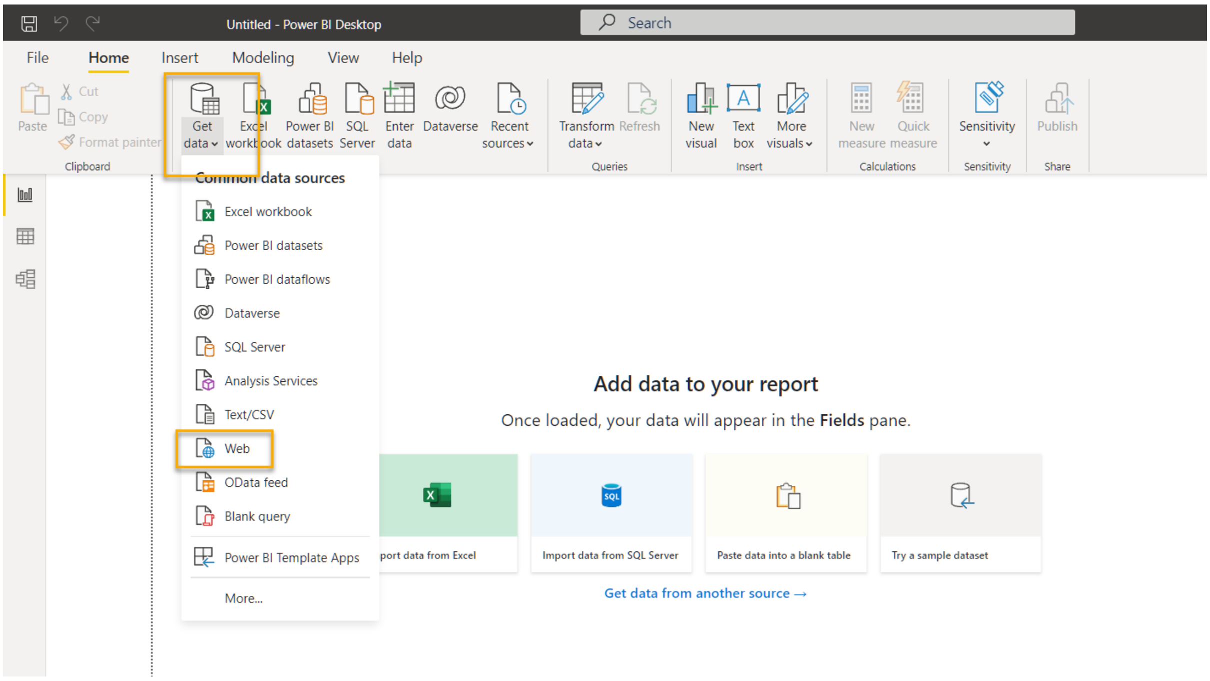
Visit the Power BI Forum to ask questions or interact with other users in the Power BI community.

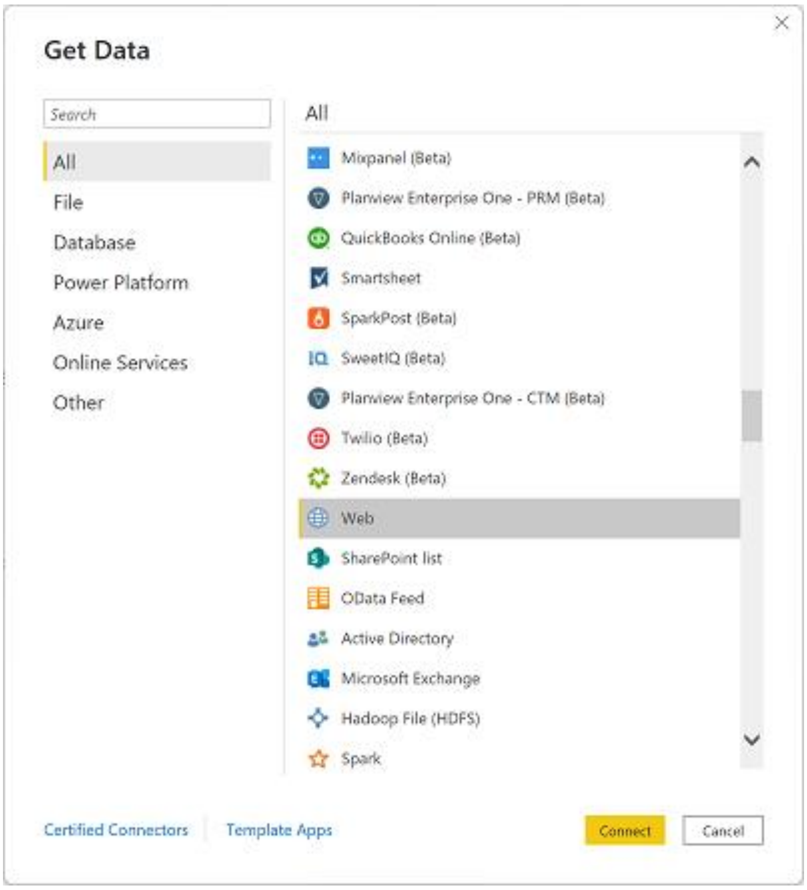
TUTORIALS

Ready to learn more about Power BI?

Now you're ready to import the data for your report.

2. On the Power BI Desktop welcome screen, select **Get data**, and then in the list of data sources, select **Web** and then select **Connect**.

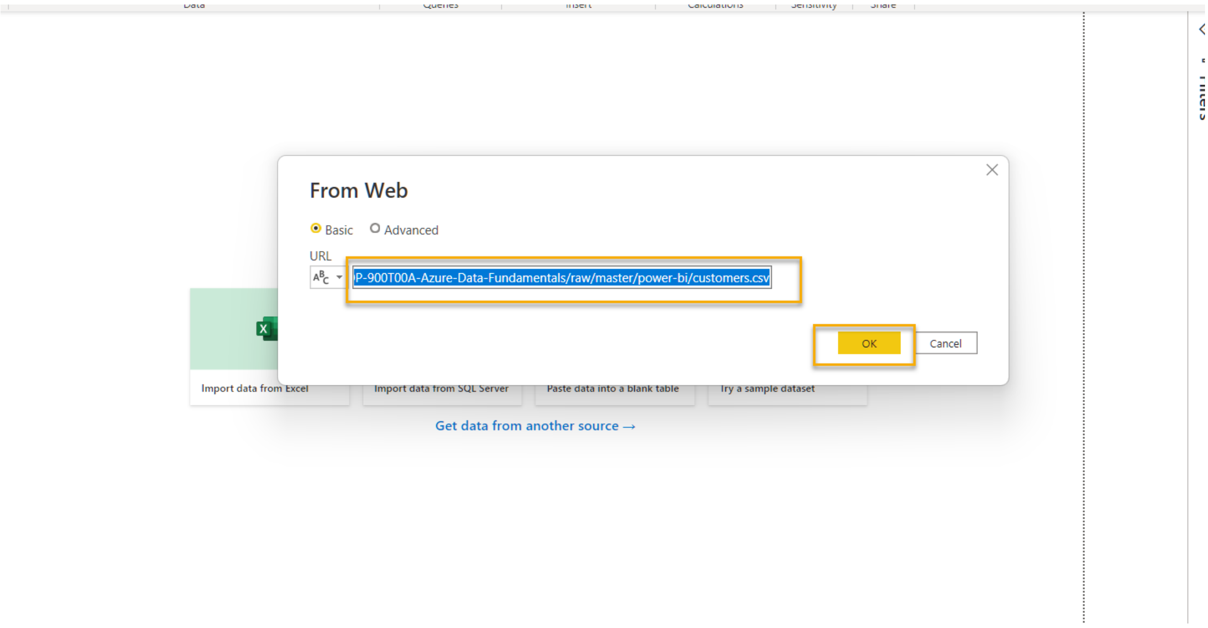




3. In the **From web** dialog box, enter the following URL and then select **OK**:

Copy

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/raw/master/power-bi/customers.csv



From Web

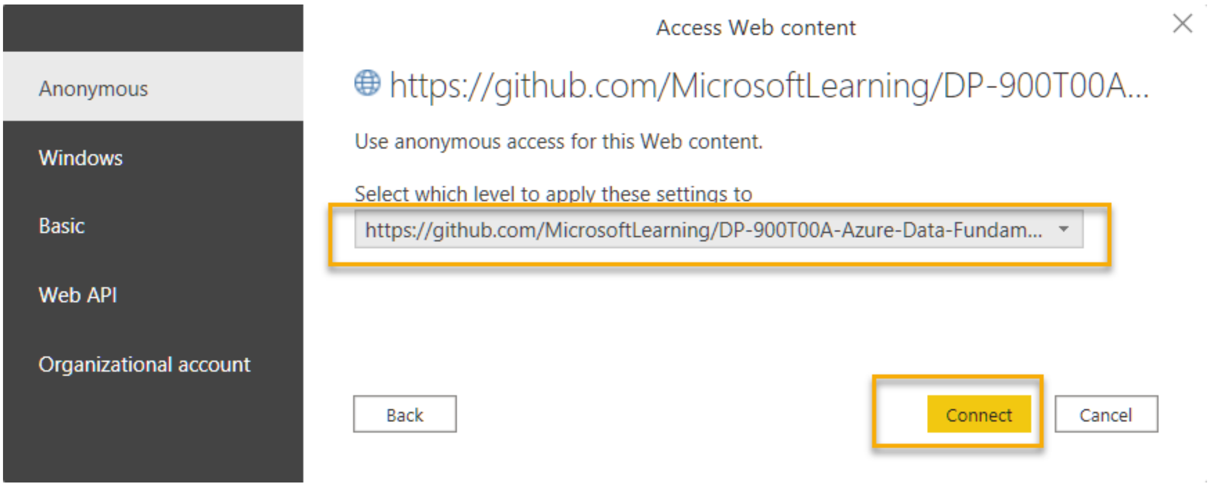
● Basic ○ Advanced

URL

P-900T00A-Azure-Data-Fundamentals/raw/master/power-bi/customers.csv

OK
Cancel
Import data from Excel
Import data from SQL Server
Paste data into a blank table
Try a sample dataset
Get data from another source →

4. Verify that the URL opens a dataset containing customer data, as shown below. Then select **Load** to load the data into the data model for your report.



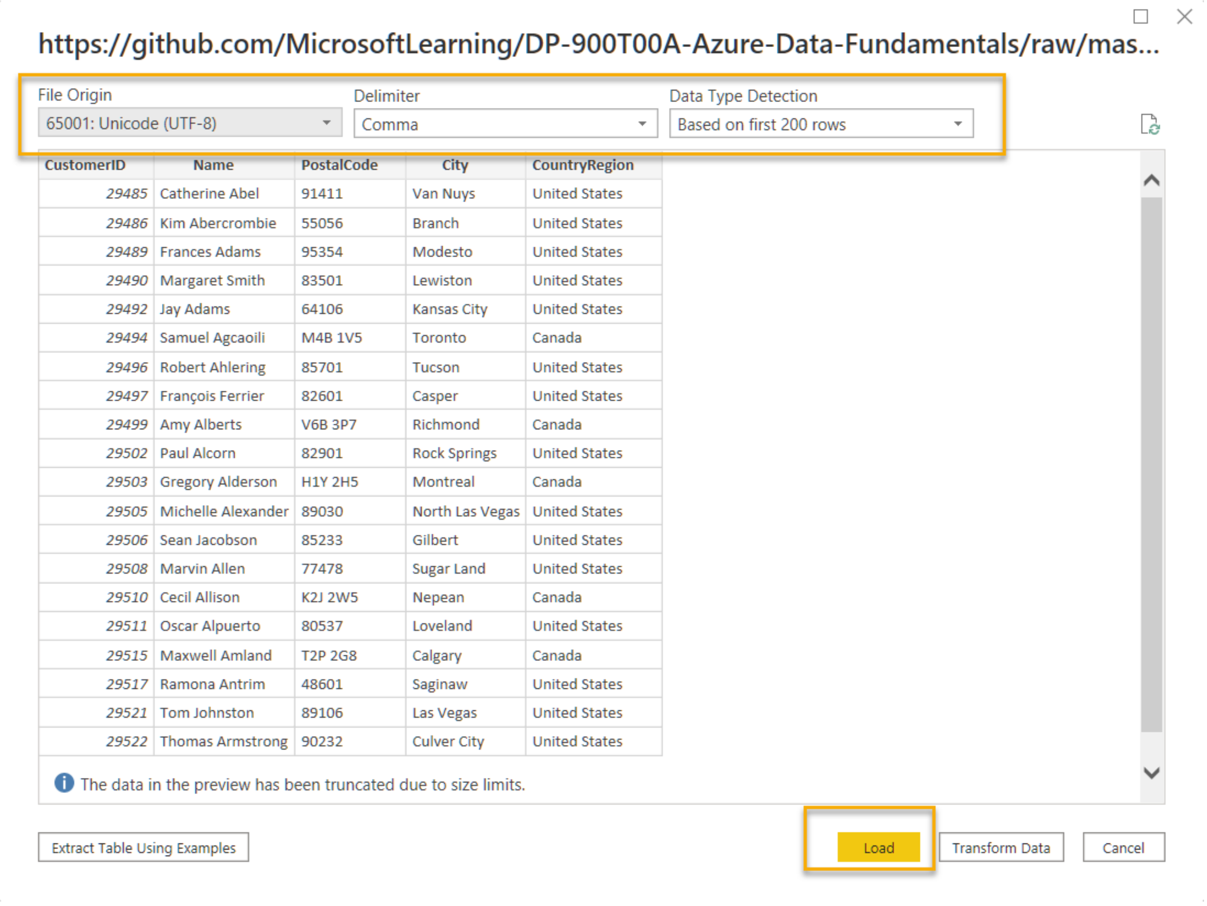
Access Web content

https://github.com/MicrosoftLearning/DP-900T00A...

Use anonymous access for this Web content.

Select which level to apply these settings to

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundam...



https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/raw/mas...

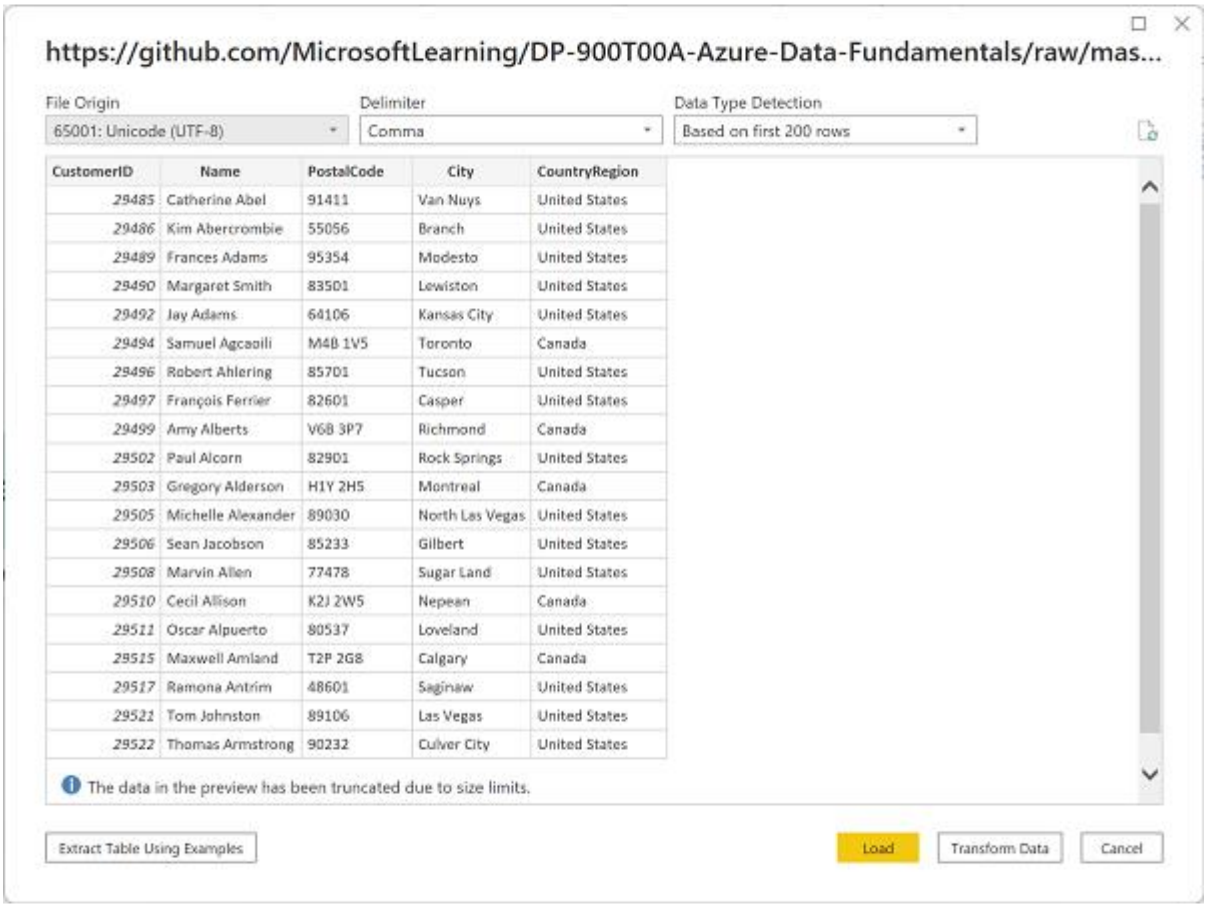
File Origin: 65001: Unicode (UTF-8)

Delimiter: Comma

Data Type Detection: Based on first 200 rows

CustomerID	Name	PostalCode	City	CountryRegion
29485	Catherine Abel	91411	Van Nuys	United States
29486	Kim Abercrombie	55056	Branch	United States
29489	Frances Adams	95354	Modesto	United States
29490	Margaret Smith	83501	Lewiston	United States
29492	Jay Adams	64106	Kansas City	United States
29494	Samuel Agcaoili	M4B 1V5	Toronto	Canada
29496	Robert Ahlering	85701	Tucson	United States
29497	François Ferrier	82601	Casper	United States
29499	Amy Alberts	V6B 3P7	Richmond	Canada
29502	Paul Alcorn	82901	Rock Springs	United States
29503	Gregory Alderson	H1Y 2H5	Montreal	Canada
29505	Michelle Alexander	89030	North Las Vegas	United States
29506	Sean Jacobson	85233	Gilbert	United States
29508	Marvin Allen	77478	Sugar Land	United States
29510	Cecil Allison	K2J 2W5	Nepean	Canada
29511	Oscar Alpuerto	80537	Loveland	United States
29515	Maxwell Amland	T2P 2G8	Calgary	Canada
29517	Ramona Antrim	48601	Saginaw	United States
29521	Tom Johnston	89106	Las Vegas	United States
29522	Thomas Armstrong	90232	Culver City	United States

The data in the preview has been truncated due to size limits.



https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/raw/mas...

File Origin: 65001: Unicode (UTF-8)

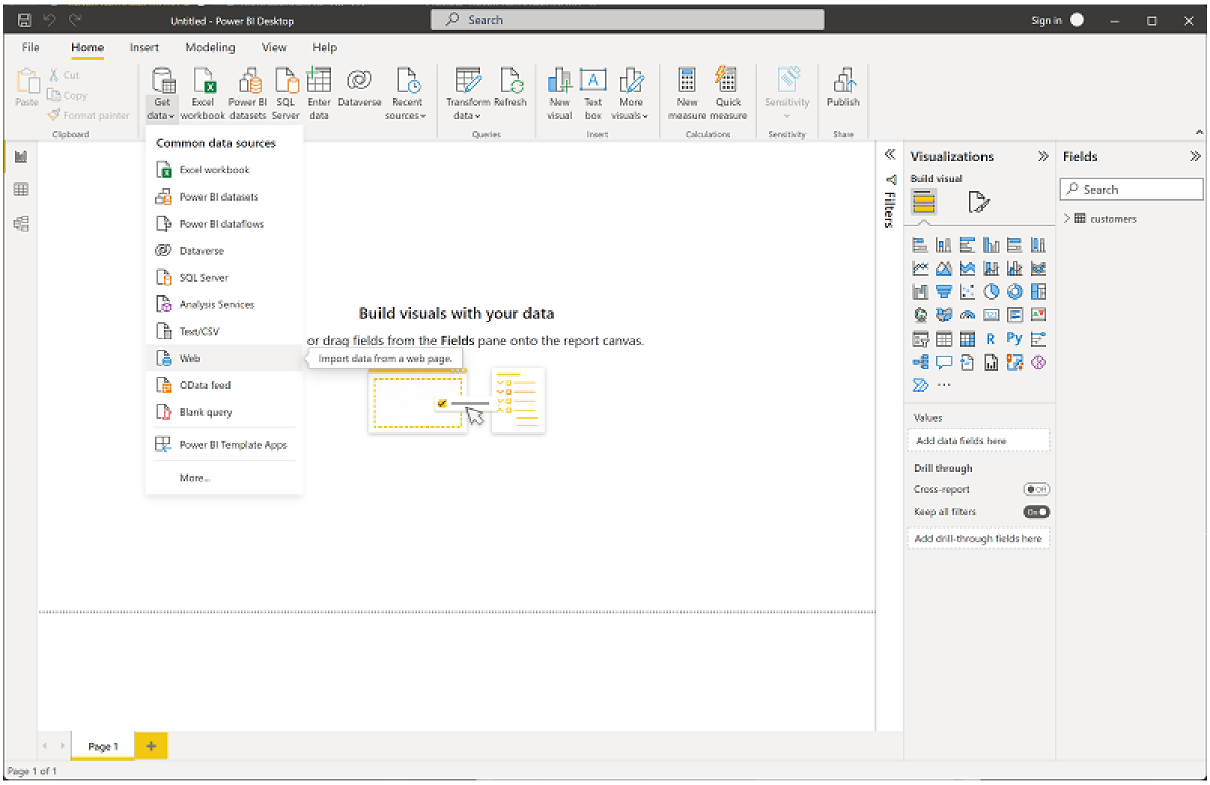
Delimiter: Comma

Data Type Detection: Based on first 200 rows

CustomerID	Name	PostalCode	City	CountryRegion
29485	Catherine Abel	91411	Van Nuys	United States
29486	Kim Abercrombie	55056	Branch	United States
29489	Frances Adams	95354	Modesto	United States
29490	Margaret Smith	83501	Lewiston	United States
29492	Jay Adams	64106	Kansas City	United States
29494	Samuel Agcaoili	M4B 1V5	Toronto	Canada
29496	Robert Ahlering	85701	Tucson	United States
29497	François Ferrier	82601	Casper	United States
29499	Amy Alberts	V6B 3P7	Richmond	Canada
29502	Paul Alcorn	82901	Rock Springs	United States
29503	Gregory Alderson	H1Y 2H5	Montreal	Canada
29505	Michelle Alexander	89030	North Las Vegas	United States
29506	Sean Jacobson	85233	Gilbert	United States
29508	Marvin Allen	77478	Sugar Land	United States
29510	Cecil Allison	K2J 2W5	Nepean	Canada
29511	Oscar Alpuerto	80537	Loveland	United States
29515	Maxwell Amland	T2P 2G8	Calgary	Canada
29517	Ramona Antrim	48601	Saginaw	United States
29521	Tom Johnston	89106	Las Vegas	United States
29522	Thomas Armstrong	90232	Culver City	United States

The data in the preview has been truncated due to size limits.

5. In the main Power BI Desktop window, in the **Get data** menu, select **Web**:



6. In the **From web** dialog box, enter the following URL and then select **OK**:

Copy

https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/raw/master/power-bi/products.csv

7. Load the product data in this file into the data model.
8. Repeat the previous three steps to import a third dataset containing order data from the following URL:

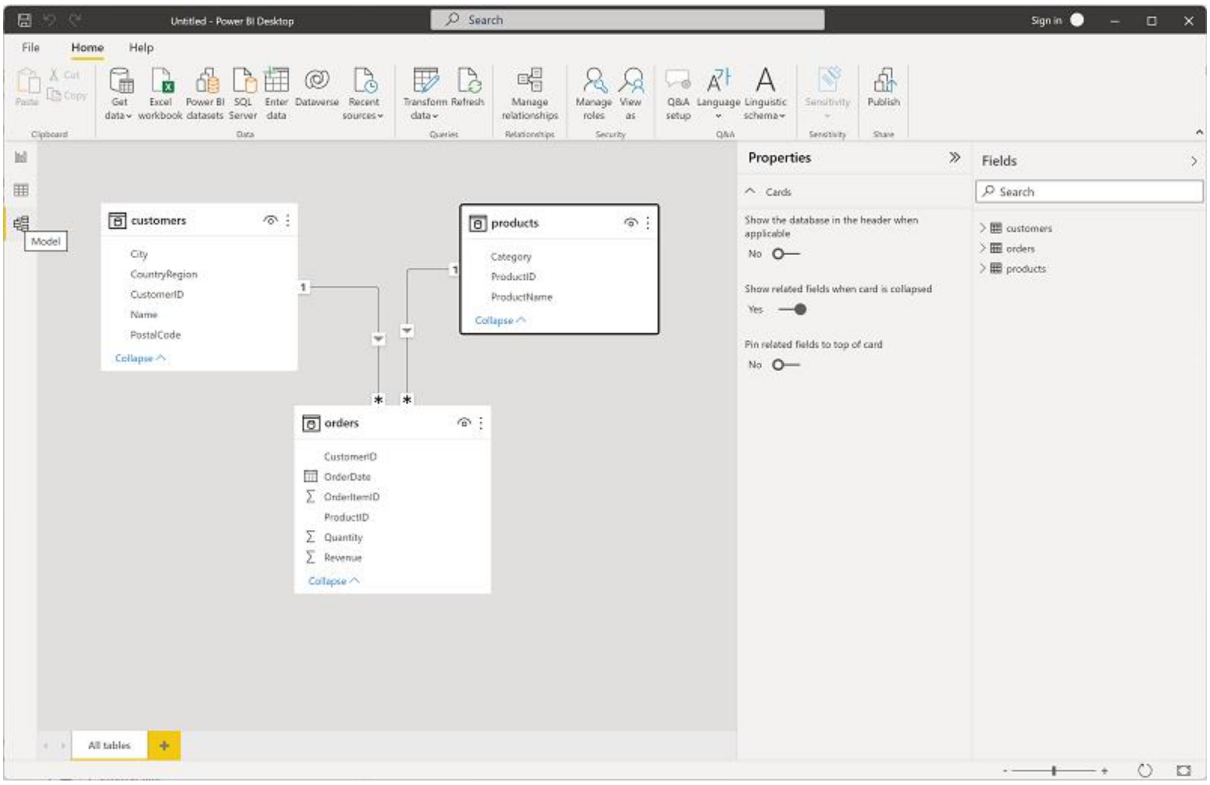
Copy

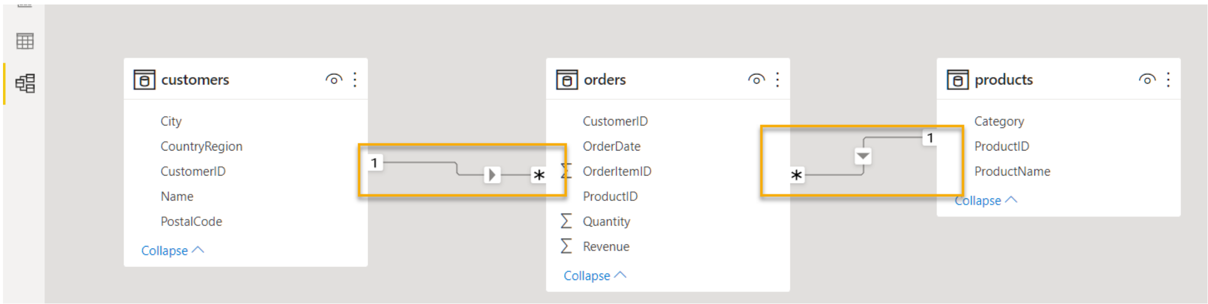
https://github.com/MicrosoftLearning/DP-900T00A-Azure-Data-Fundamentals/raw/master/power-bi/orders.csv

Explore a data model

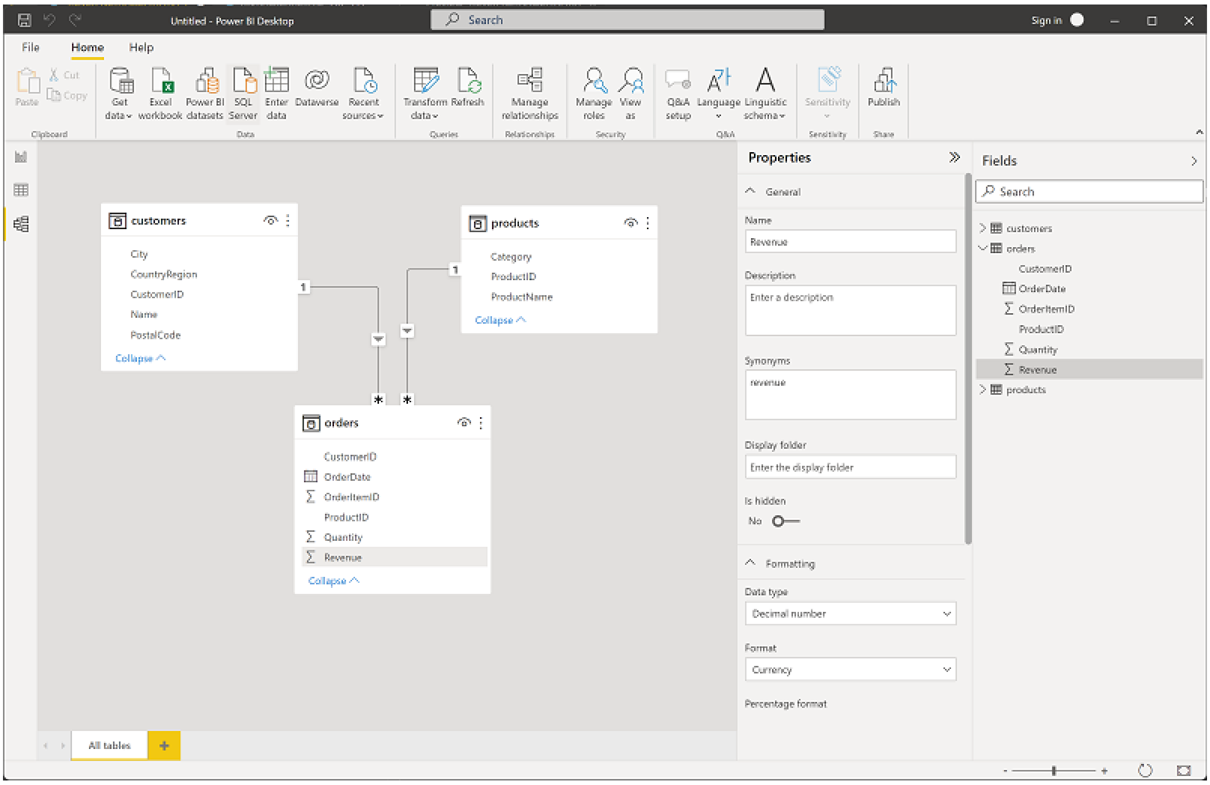
The three tables of data you've imported have been loaded into a data model, which you'll now explore and refine.

1. In Power BI Desktop, on the left-side edge, select the **Model** tab, and then arrange the tables in the model so you can see them (you can hide the panes on the right side by using the **>>** icons):





2. In the **orders** table, select the **Revenue** field and then in the **Properties** pane, set its **Format** property to **Currency**:



Properties

General

Name: Revenue

Description: Enter a description

Synonyms: revenue

Display folder: Enter the display folder

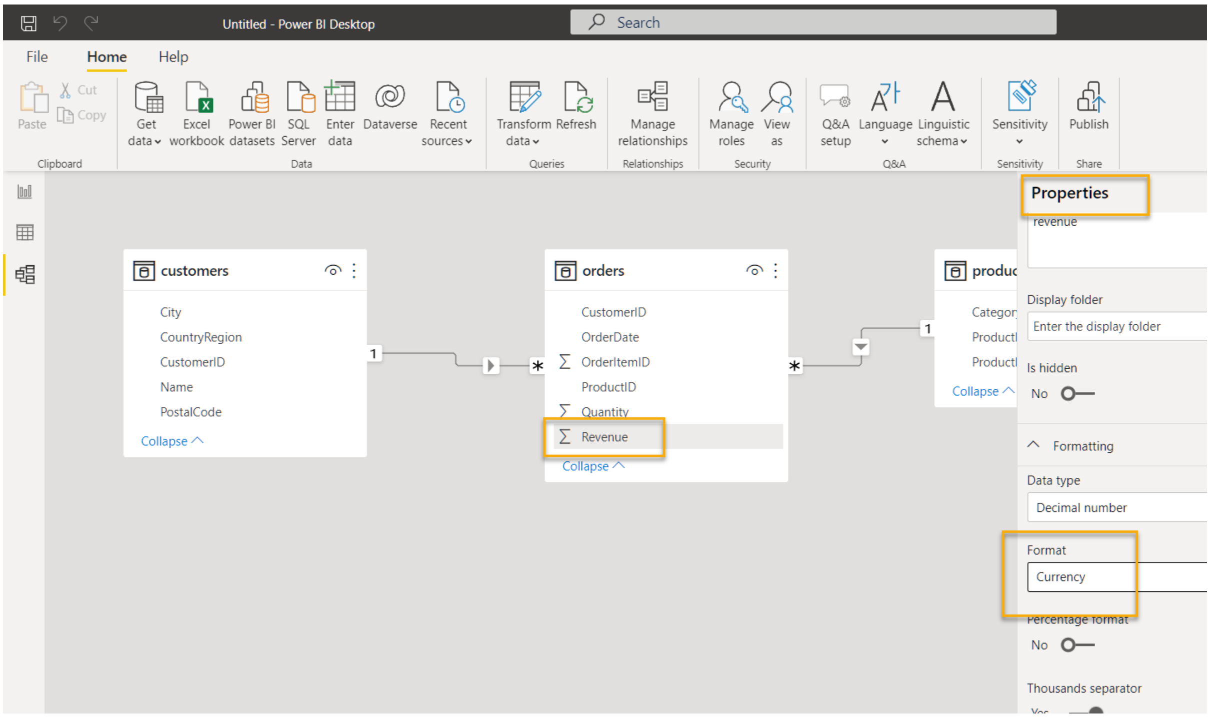
Is hidden: No

Formatting

Data type: Decimal number

Format: Currency

Percentage format



Properties

revenue

Display folder: Enter the display folder

Is hidden: No

Formatting

Data type: Decimal number

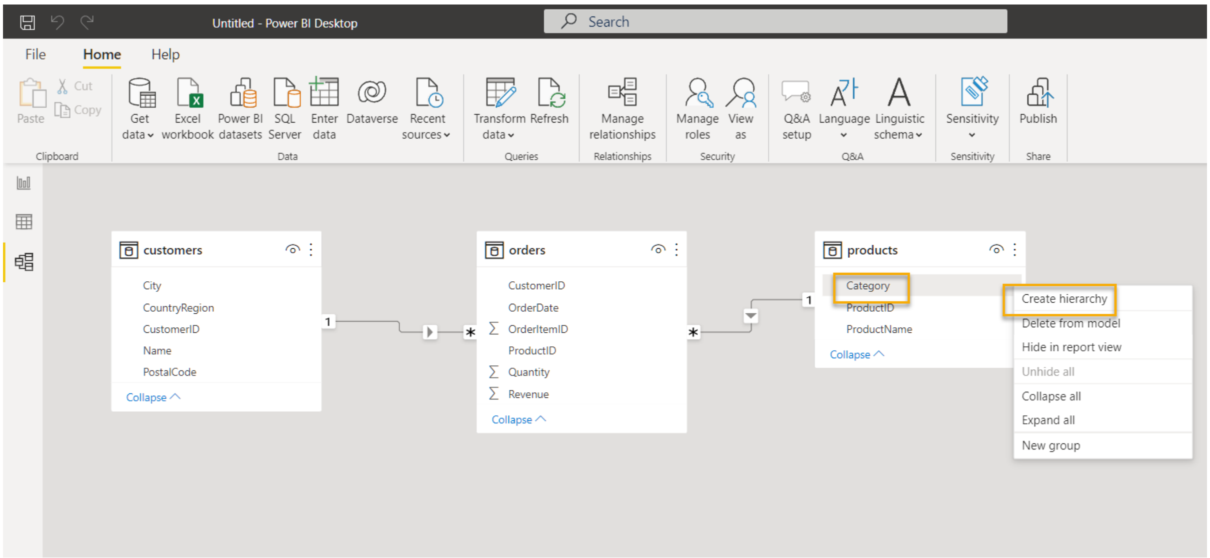
Format: Currency

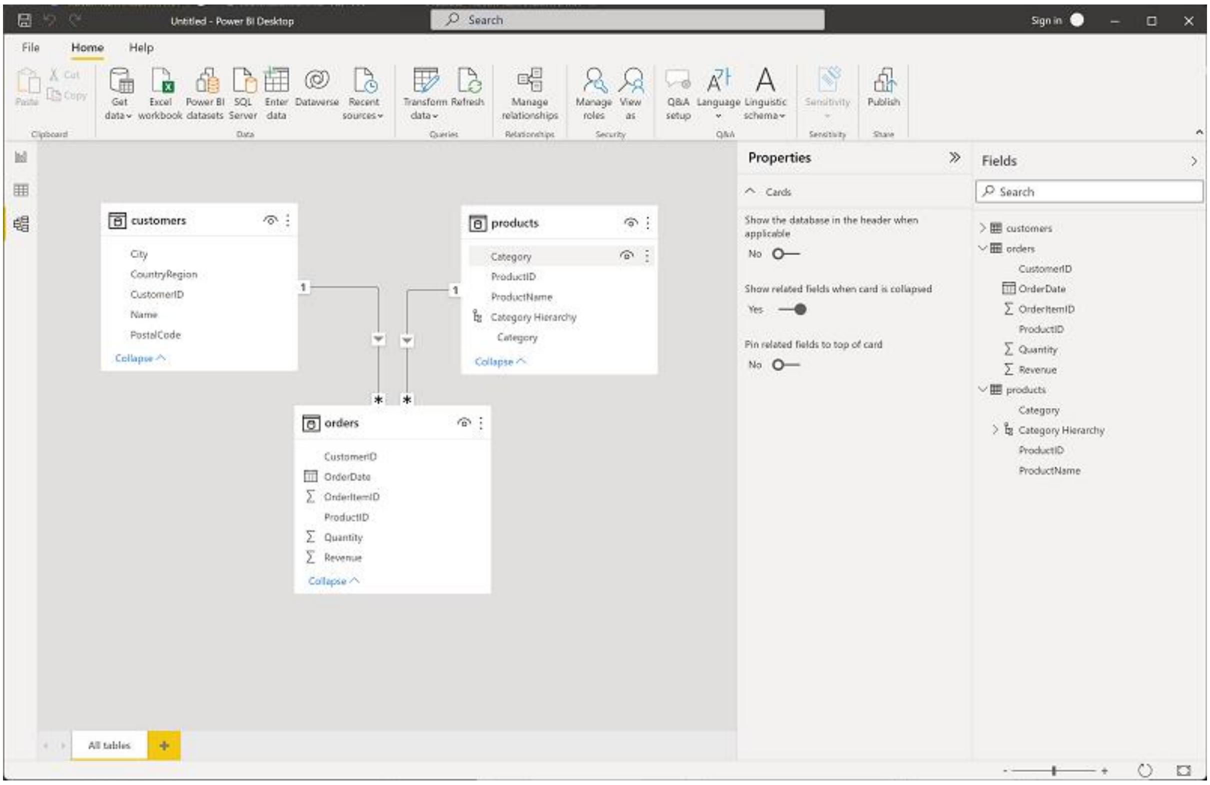
Percentage format: No

Thousands separator: Yes

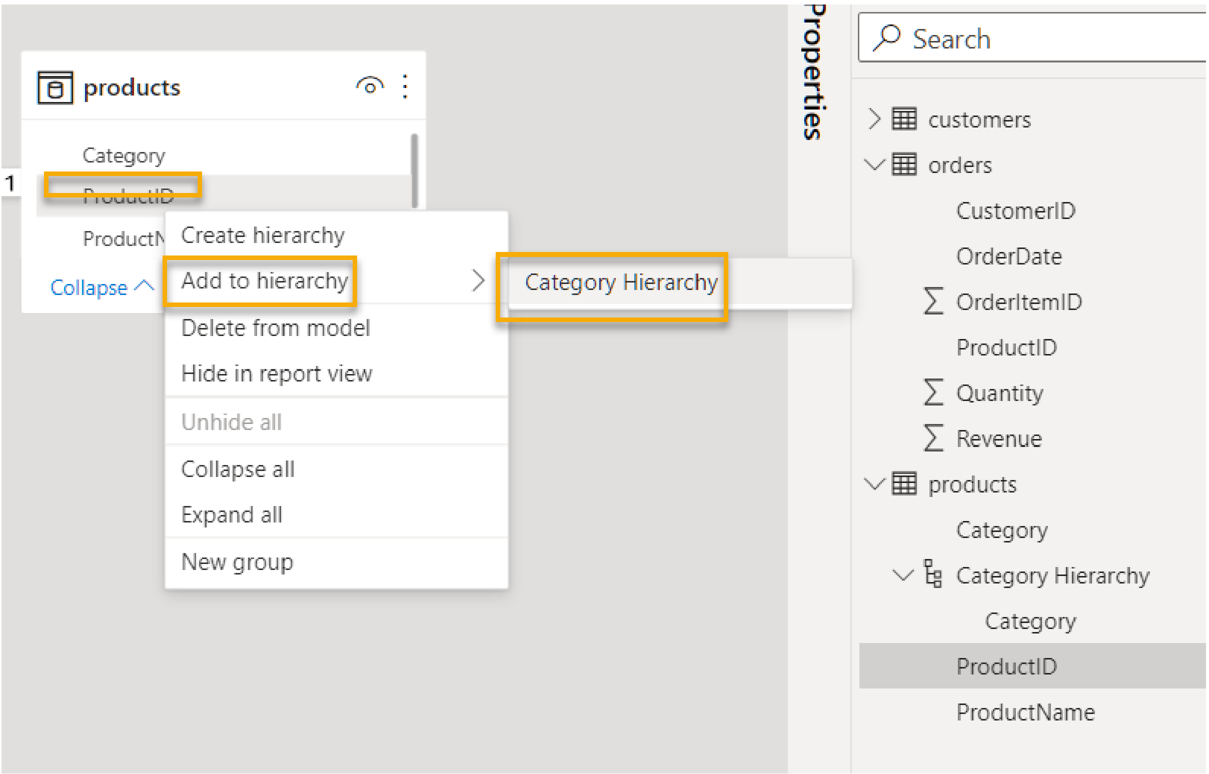
This will ensure that revenue values are displayed as currency in report visualizations.

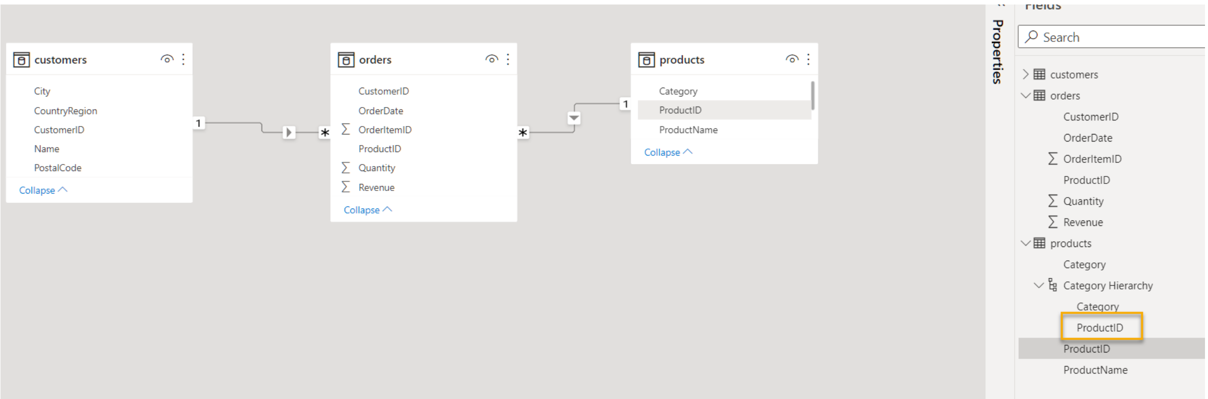
3. In the products table, right-click the **Category** field (or open its **⋮** menu) and select **Create hierarchy**. This creates a hierarchy named **Category Hierarchy** (you may need to expand or scroll in the **products** table to see this - you can also see it in the **Fields** pane)



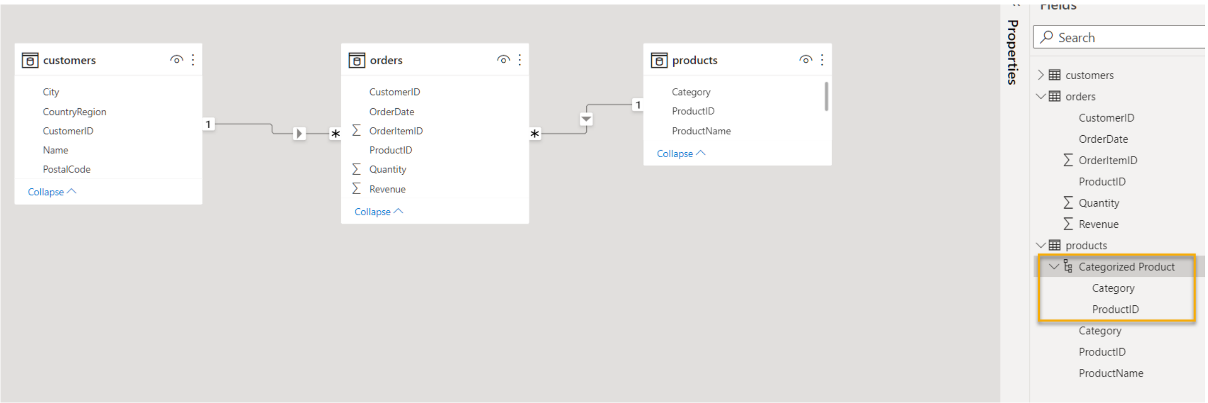


4. In the products table, right-click the **ProductName** field (or open its **⋮** menu) and select **Add to hierarchy** > **Category Hierarchy**. This adds the **ProductName** field to the hierarchy you created previously.



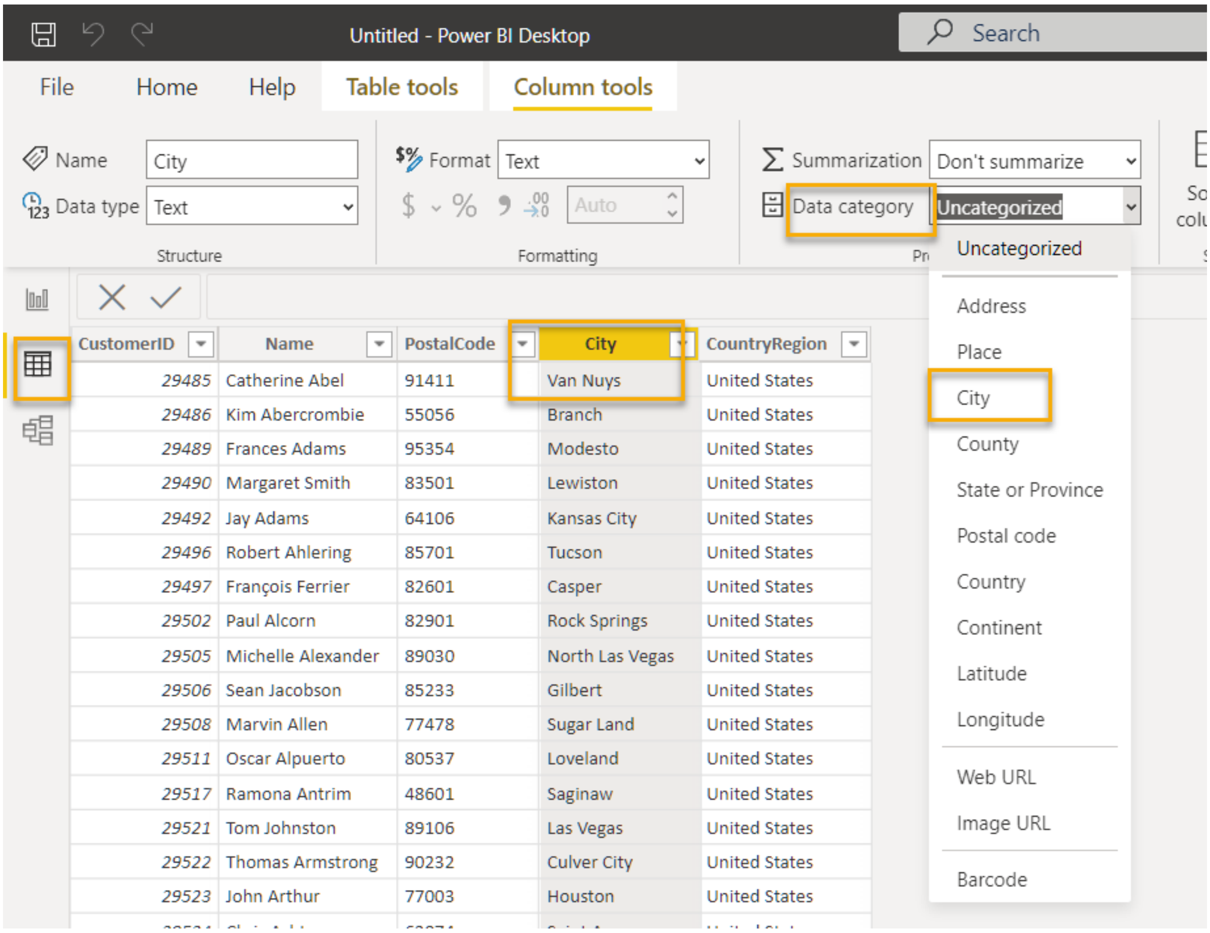


5. In the **Fields** pane, right-click **Category Hierarchy** (or open its **...** menu) and select **Rename**. The rename the hierarchy to **Categorized Product**.



6. On the left-side edge, select the **Data** tab, and then in the **Fields** pane, select the **customers** table.

7. Select the **City** column header, and then set its **Data Category** property to **City**:



Name: City
Data type: Text
Format: Text
Summarization: Don't summarize
Data category: Uncategorized

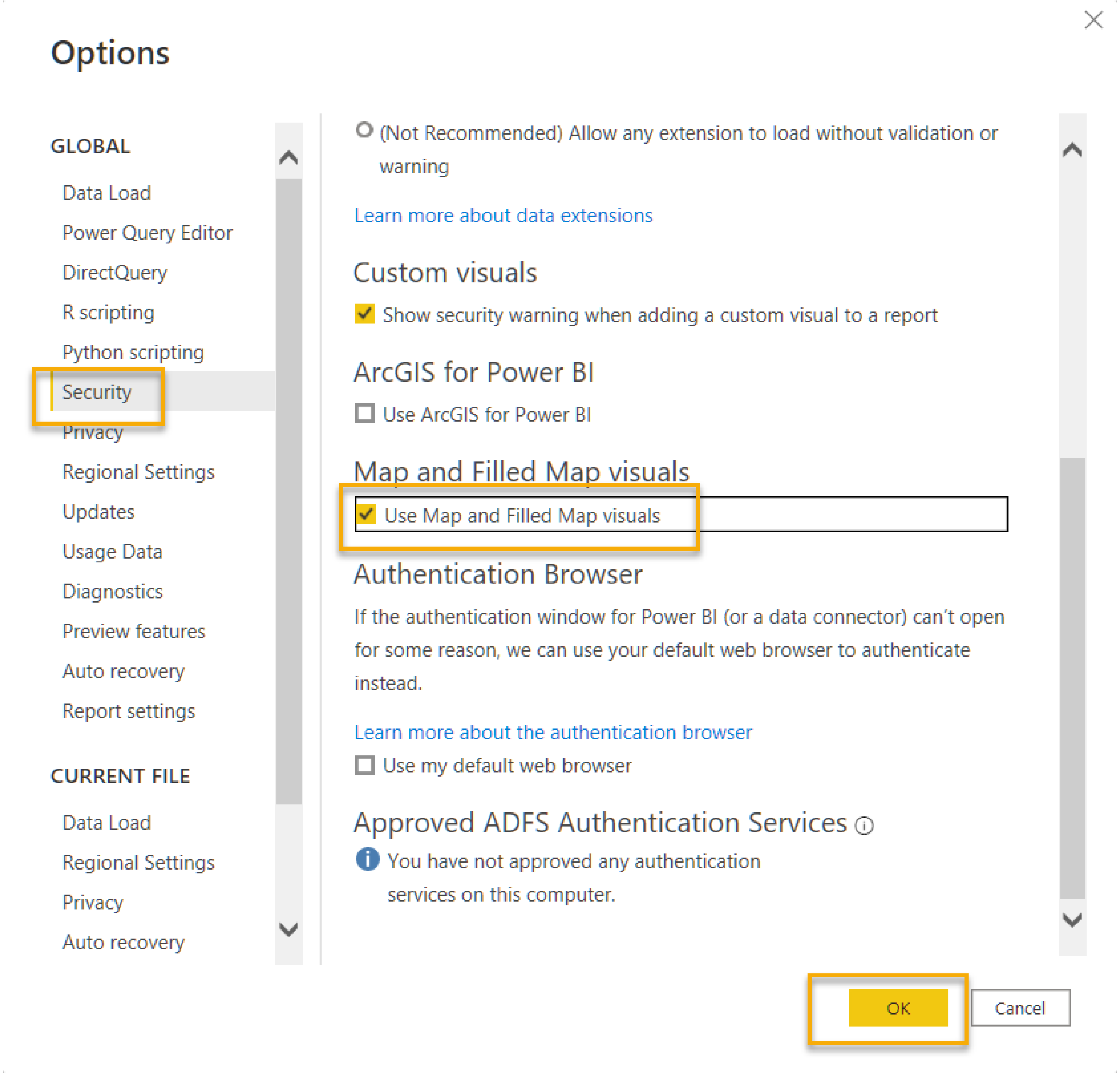
CustomerID	Name	PostalCode	City	CountryRegion
29485	Catherine Abel	91411	Van Nuys	United States
29486	Kim Abercrombie	55056	Branch	United States
29489	Frances Adams	95354	Modesto	United States
29490	Margaret Smith	83501	Lewiston	United States
29492	Jay Adams	64106	Kansas City	United States
29496	Robert Ahlering	85701	Tucson	United States
29497	François Ferrier	82601	Casper	United States
29502	Paul Alcorn	82901	Rock Springs	United States
29505	Michelle Alexander	89030	North Las Vegas	United States
29506	Sean Jacobson	85233	Gilbert	United States
29508	Marvin Allen	77478	Sugar Land	United States
29511	Oscar Alpuerto	80537	Loveland	United States
29517	Ramona Antrim	48601	Saginaw	United States
29521	Tom Johnston	89106	Las Vegas	United States
29522	Thomas Armstrong	90232	Culver City	United States
29523	John Arthur	77003	Houston	United States

This will ensure that the values in this column are interpreted as city names, which can be useful if you intend to include map visualizations.

Create a report

Now you're almost ready to create a report. First you need to check some settings to ensure all visualizations are enabled.

1. On the **File** menu, select **Options and Settings**. Then select **Options**, and in the **Security** section, ensure that **Use Map and Filled Map visuals** is enabled and select **OK**.



○ (Not Recommended) Allow any extension to load without validation or warning

Learn more about data extensions

Custom visuals

☑ Show security warning when adding a custom visual to a report

ArcGIS for Power BI

□ Use ArcGIS for Power BI

Map and Filled Map visuals

☑ Use Map and Filled Map visuals

Authentication Browser

If the authentication window for Power BI (or a data connector) can't open for some reason, we can use your default web browser to authenticate instead.

Learn more about the authentication browser

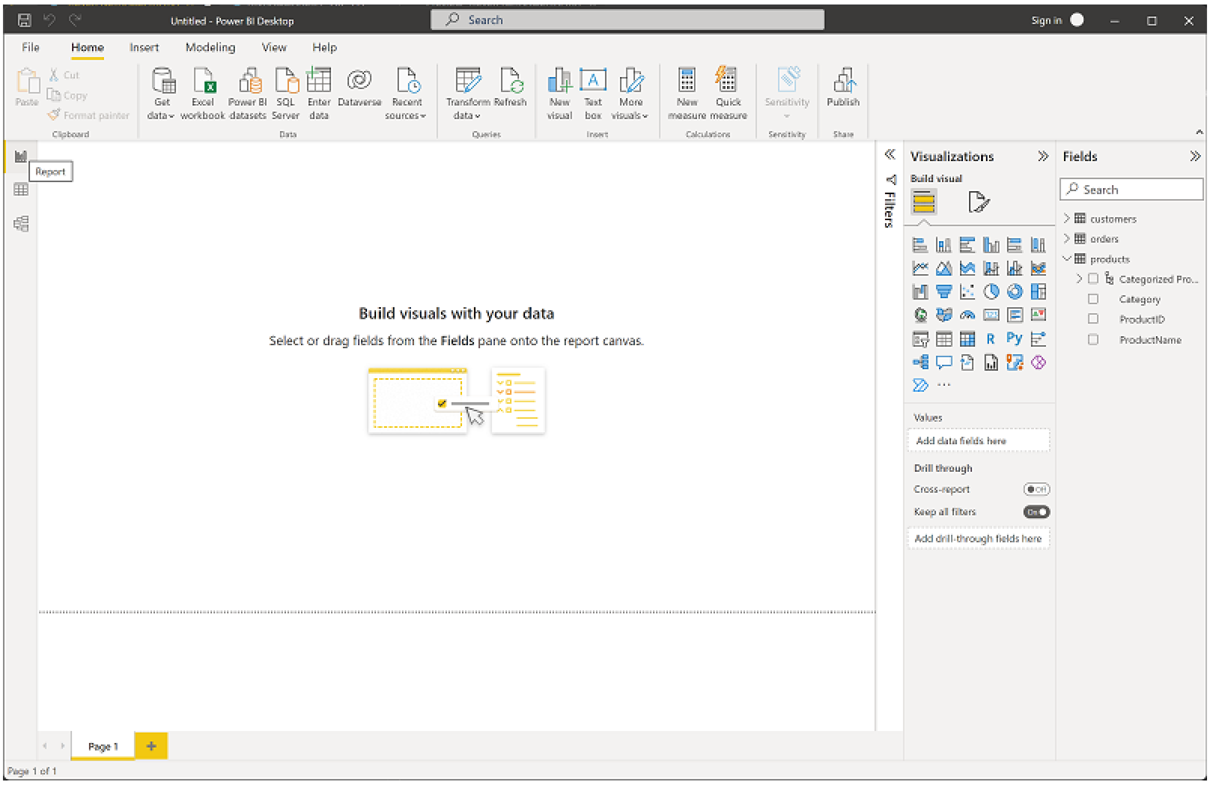
□ Use my default web browser

Approved ADFS Authentication Services ⓘ

ⓘ You have not approved any authentication services on this computer.

This ensures that you can include map visualizations in reports.

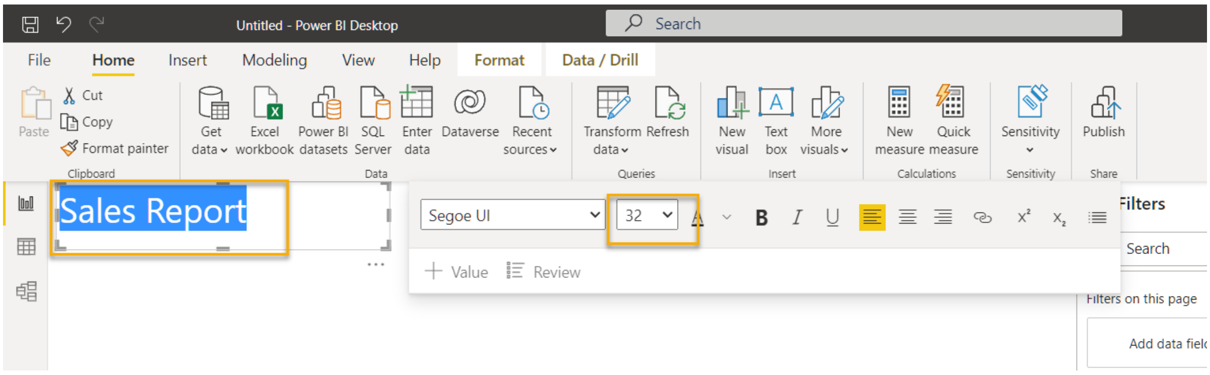
2. On the left-side edge, select the **Report** tab and view the report design interface.



Build visuals with your data

Select or drag fields from the **Fields** pane onto the report canvas.

3. In the ribbon, above the report design surface, select **Text Box** and add a text box containing the text **Sales Report** to the report. Format the text to make it bold with a font size of 32.

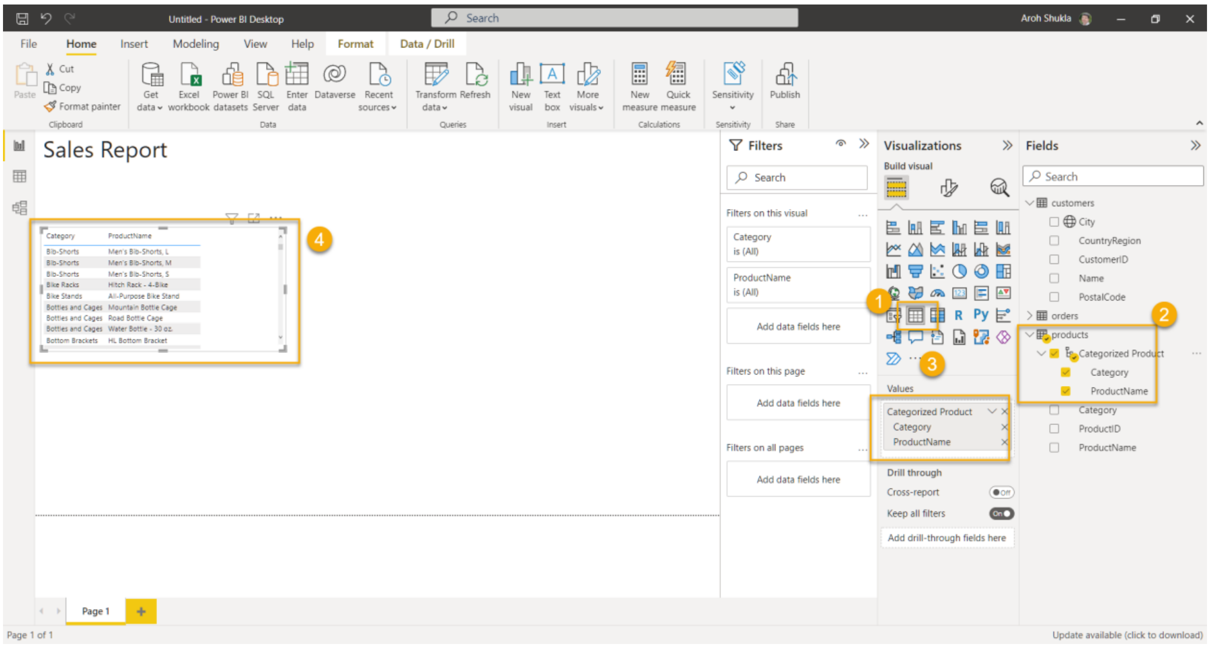


Sales Report

Segoe UI

32

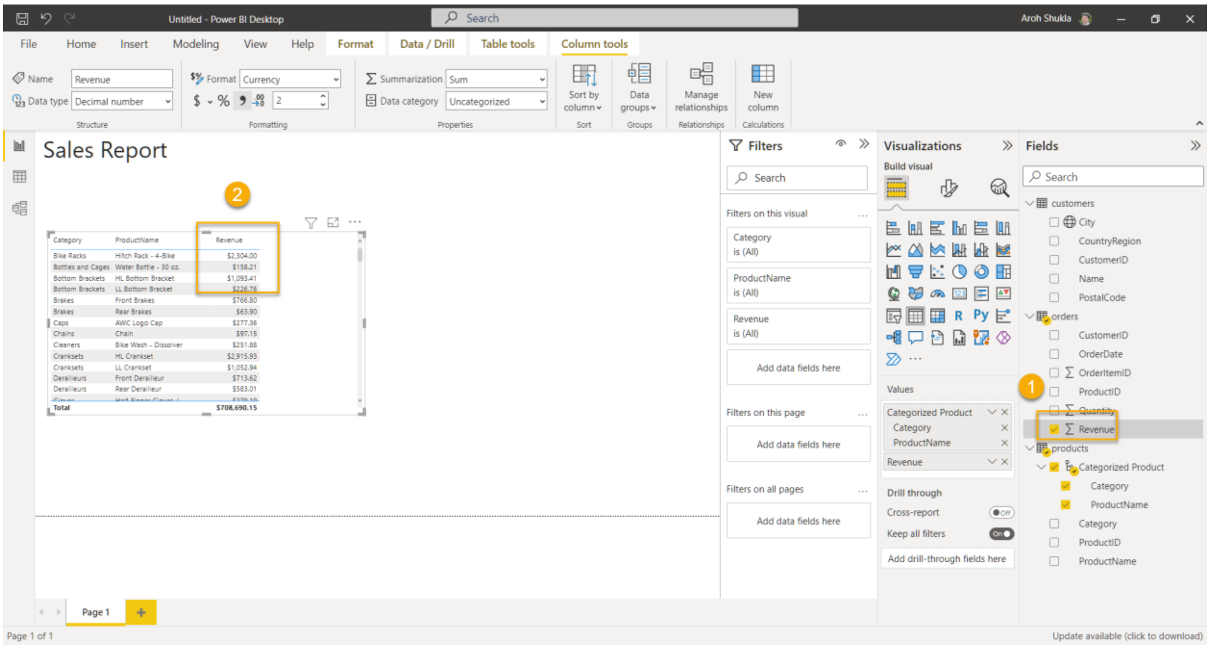
4. Select any empty area on the report to de-select the text box. Then in the **Fields** pane, expand **Products** and select the **Categorized Products** field. This adds a table to the report.



Sales Report

Category	ProductName
Bib-Shorts	Men's Bib-Shorts, L
Bib-Shorts	Men's Bib-Shorts, M
Bib-Shorts	Men's Bib-Shorts, S
Bike Racks	Hitch Rack - 4-Bike
Bike Stands	All-Purpose Bike Stand
Bottles and Cages	Mountain Bottle Cage
Bottles and Cages	Road Bottle Cage
Bottles and Cages	Water Bottle - 30 oz.
Bottom Brackets	HL Bottom Bracket

5. With the table still selected, in the **Fields** pane, expand **Orders** and select **Revenue**. A Revenue column is added to the table (you may need to expand the size of the table to see it).

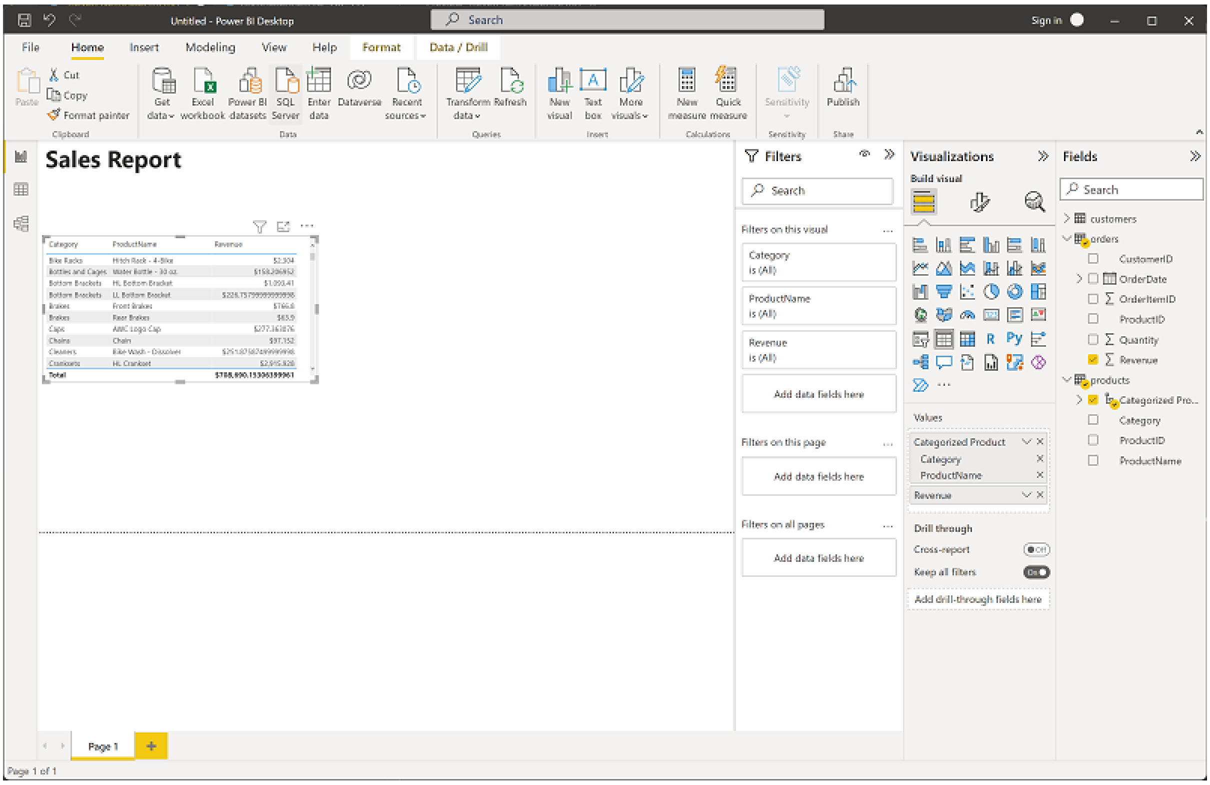


Sales Report

Category	ProductName	Revenue
Bike Racks	Hitch Rack - 4-Bike	$2,304.00
Bottles and Cages	Water Bottle - 30 oz.	$158.21
Bottom Brackets	HL Bottom Bracket	$1,093.41
Bottom Brackets	LL Bottom Bracket	$226.76
Brakes	Front Brakes	$766.80
Brakes	Rear Brakes	$63.90
Caps	AWC Logo Cap	$277.36
Chains	Chain	$97.15
Cleaners	Bike Wash - Dissolver	$251.88
Cranksets	HL Crankset	$2,915.93
Cranksets	LL Crankset	$1,052.94
Derailleurs	Front Derailleur	$713.62
Derailleurs	Rear Derailleur	$583.01
Total		$708,690.15

Name: Revenue
Data type: Decimal number
Format: Currency
Summarization: Sum
Data category: Uncategorized

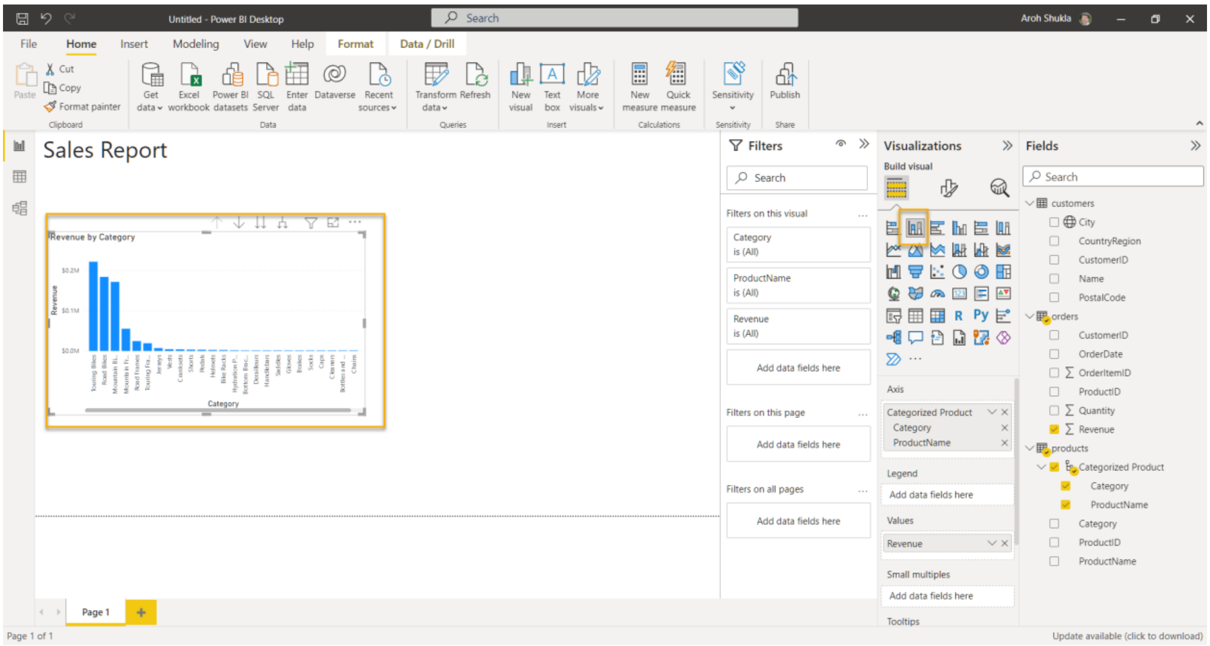
The revenue is formatted as currency, as you specified in the model. However, you didn't specify the number of decimal places, so the values include fractional amounts. It won't matter for the visualizations you're going to create, but you could go back to the **Model** or **Data** tab and change the decimal places if you wish!



Sales Report

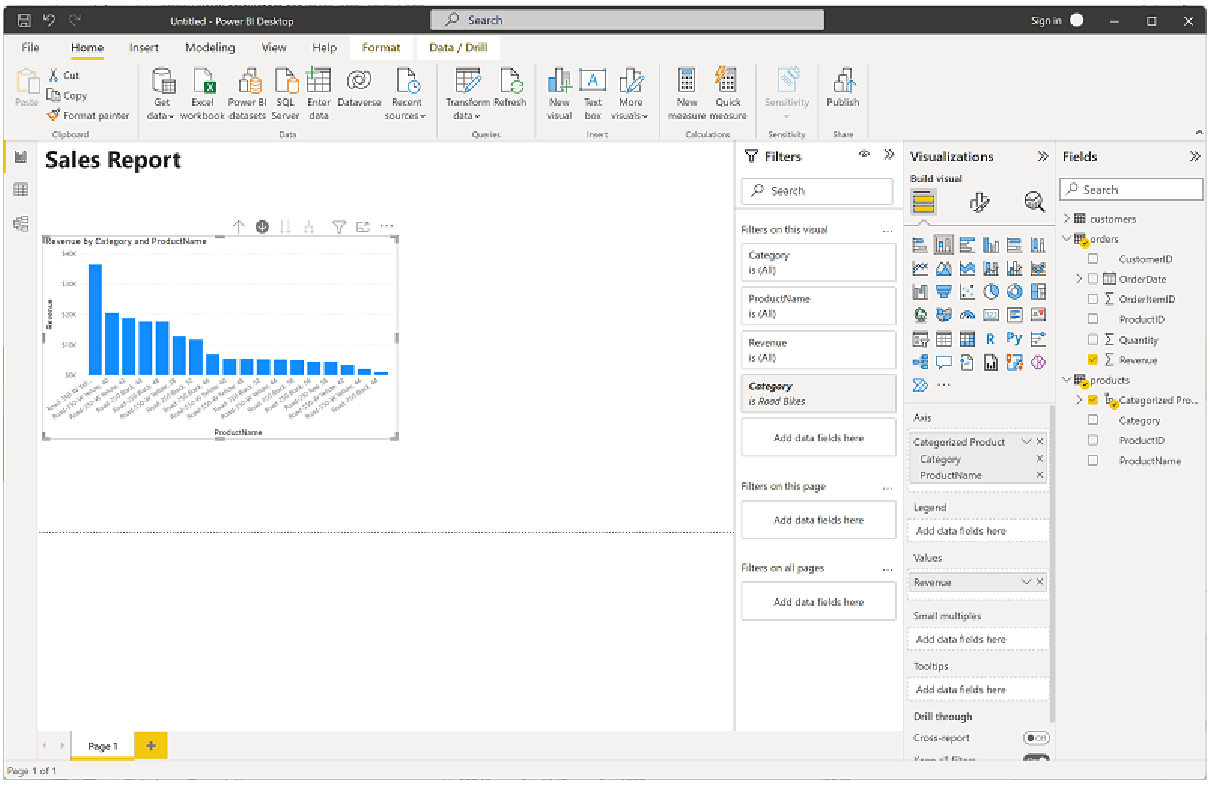
Category	ProductName	Revenue
Bike Racks	Hitch Rack - 4-Bike	$2,304
Bottles and Cages	Water Bottle - 30 oz.	$158.206952
Bottom Brackets	HL Bottom Bracket	$1,093.41
Bottom Brackets	LL Bottom Bracket	$228.7579999999998
Brakes	Front Brakes	$766.8
Brakes	Rear Brakes	$83.9
Caps	AWC Logo Cap	$277.363876
Chains	Chain	$97.152
Cleaners	Bike Wash - Dissolver	$251.875824999998
Cranksets	HL Crankset	$2,915.828
Total		$788,890.15306399961

6. With the table still selected, in the **Visualizations** pane, select the **Stacked column chart** visualization. The table is changed to a column chart showing revenue by category.



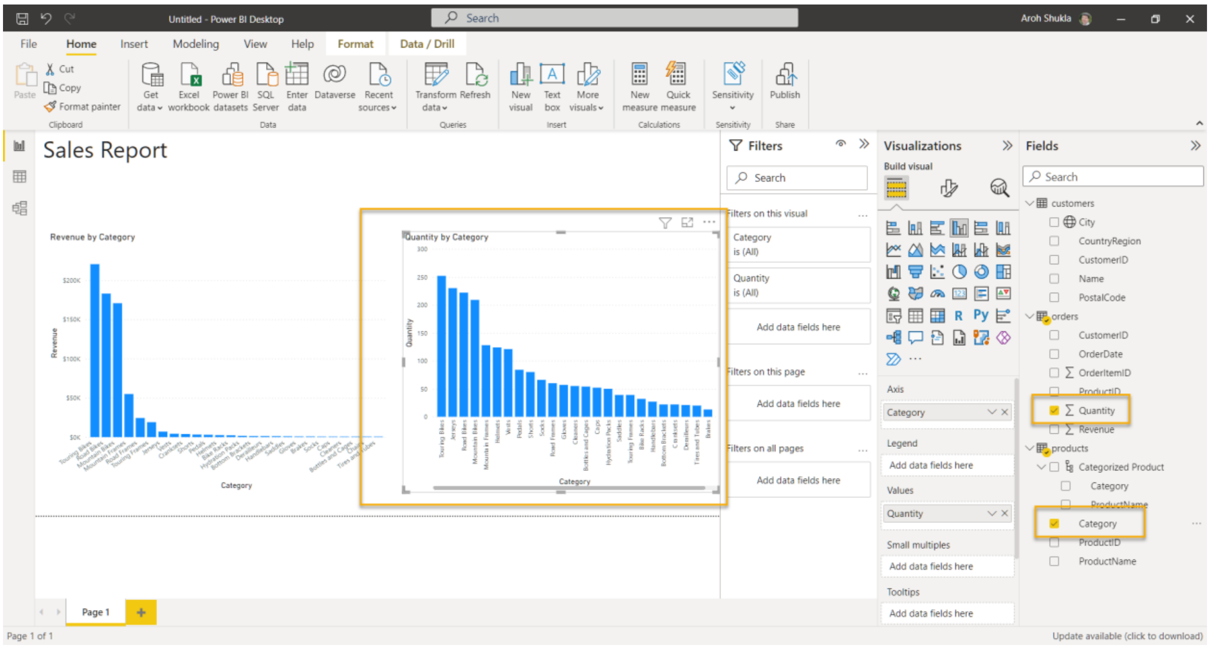
Sales Report

7. Above the selected column chart, select the ↓ icon to turn on drill-down. Then in the chart, select the second column (*Road Bikes*) to drill down and see the revenue for the individual products in this category. This capability is possible because you defined a hierarchy of categories and products.

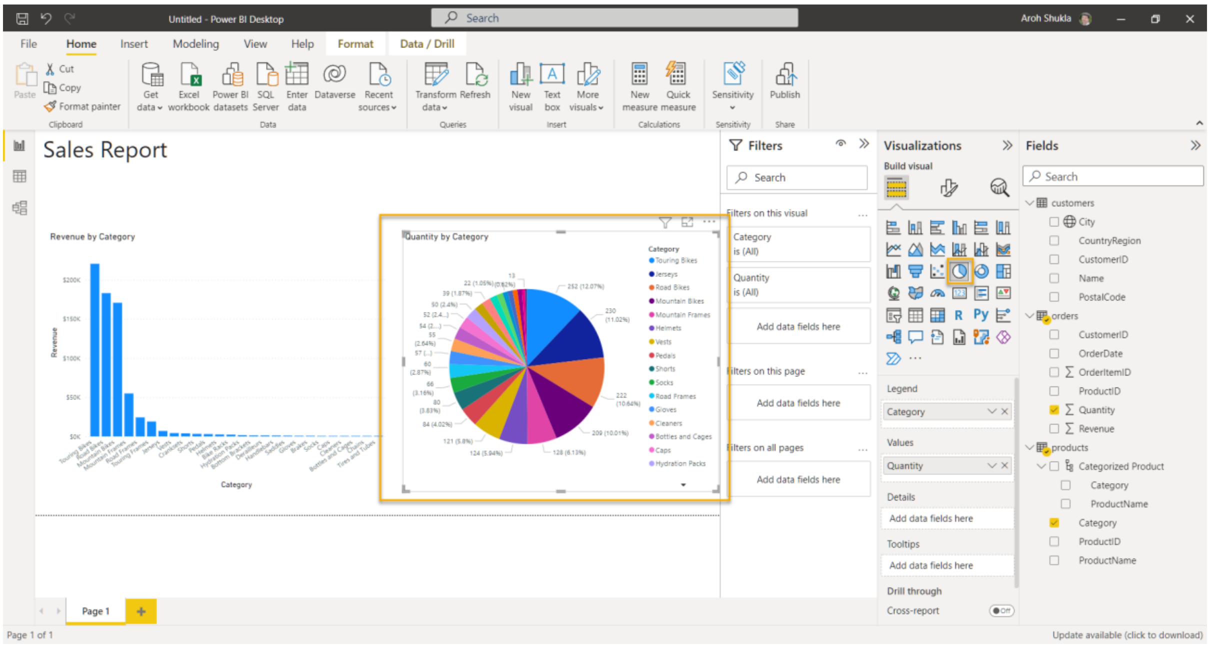


8. Use the ↑ icon to drill back up to the category level. Then select the **(↓)** icon to turn off the drill-down feature.

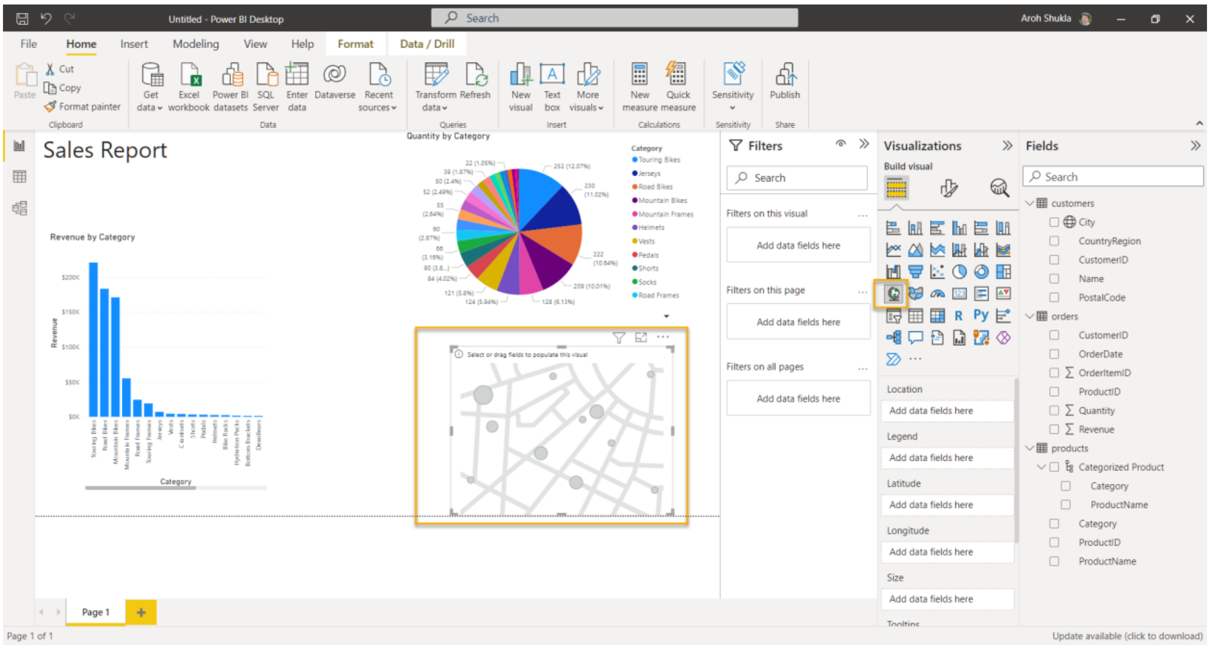
9. Select a blank area of the report, and then in the **Fields** pane, select the **Quantity** field in the **orders** table and the **Category** field in the **products** table. This results in another column chart showing sales quantity by product category.

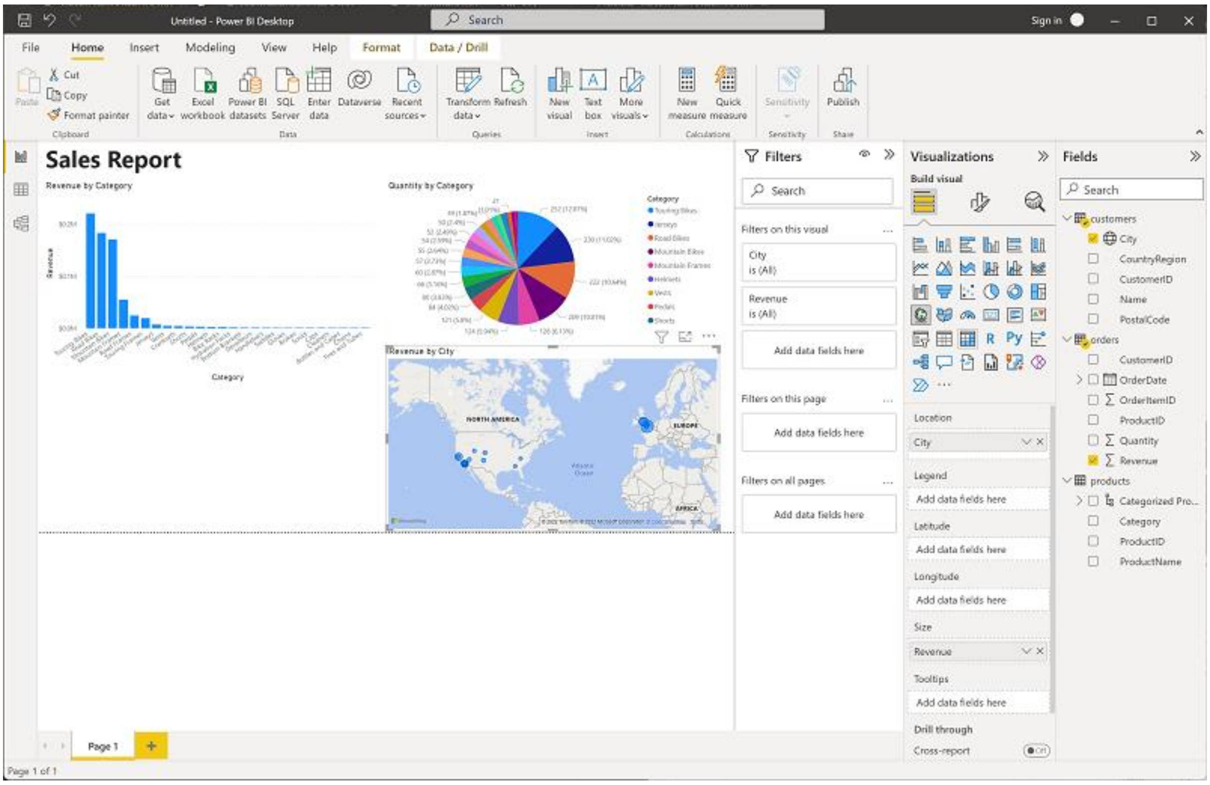


10. With the new column chart selected, in the **Visualizations** pane, select **Pie chart** and then resize the chart and position it next to the revenue by category column chart.

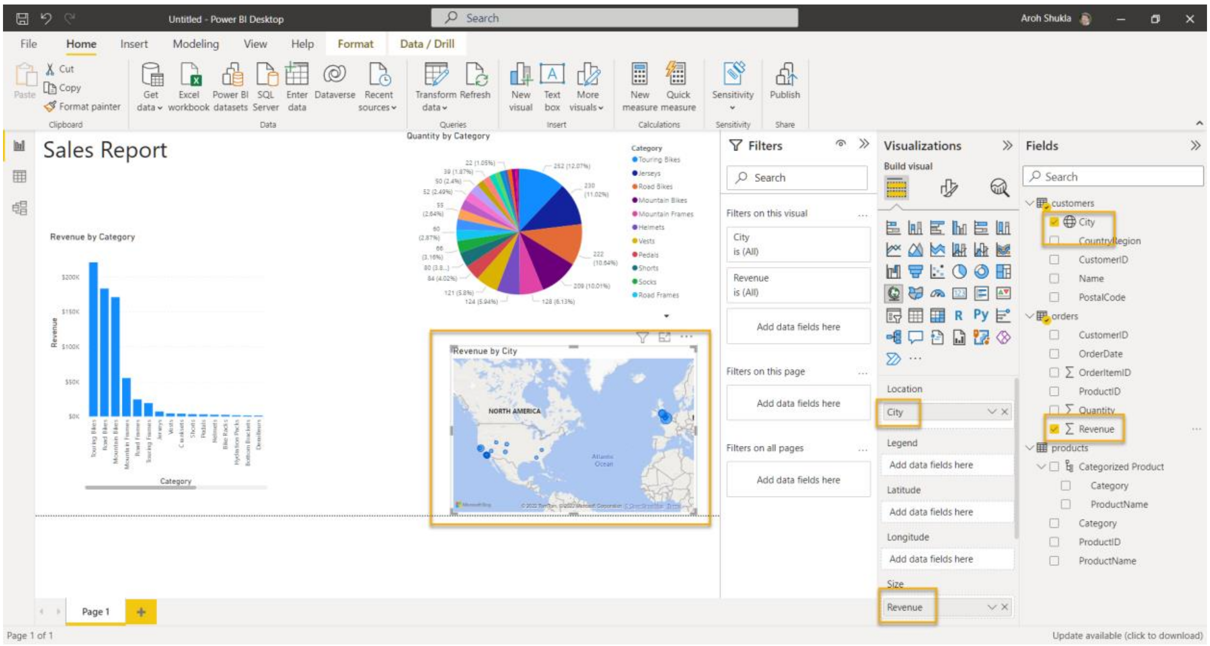


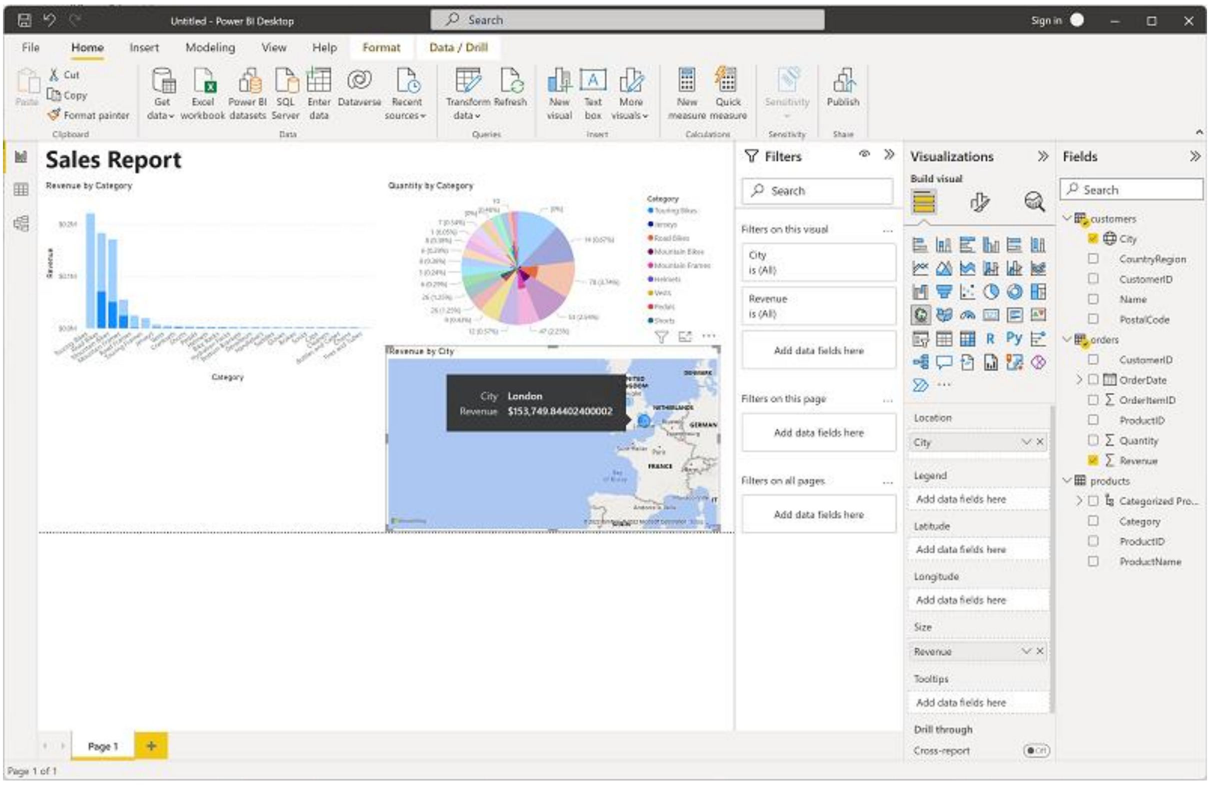
11. Select a blank area of the report, and then in the **Fields** pane, select the **City** field in the **customers** table and then select the **Revenue** field in the **orders** table. This results in a map showing sales revenue by city (rearrange and resize the visualizations as needed):





12. In the map, note that you can drag, double-click, use a mouse-wheel, or pinch and drag on a touch screen to interact. Then select a specific city, and note that the other visualizations in the report are modified to highlight the data for the selected city.





13. On the **File** menu, select **Save**. Then save the file with an appropriate .pbix file name. You can open the file and explore data modeling and visualization further at your leisure.

In this exercise, you have used Power BI Desktop to ingest data, create a data model, and use interactive visualizations to create a report. If you have a Power BI service subscription, you can sign into your account and publish the report to a Power BI workspace.