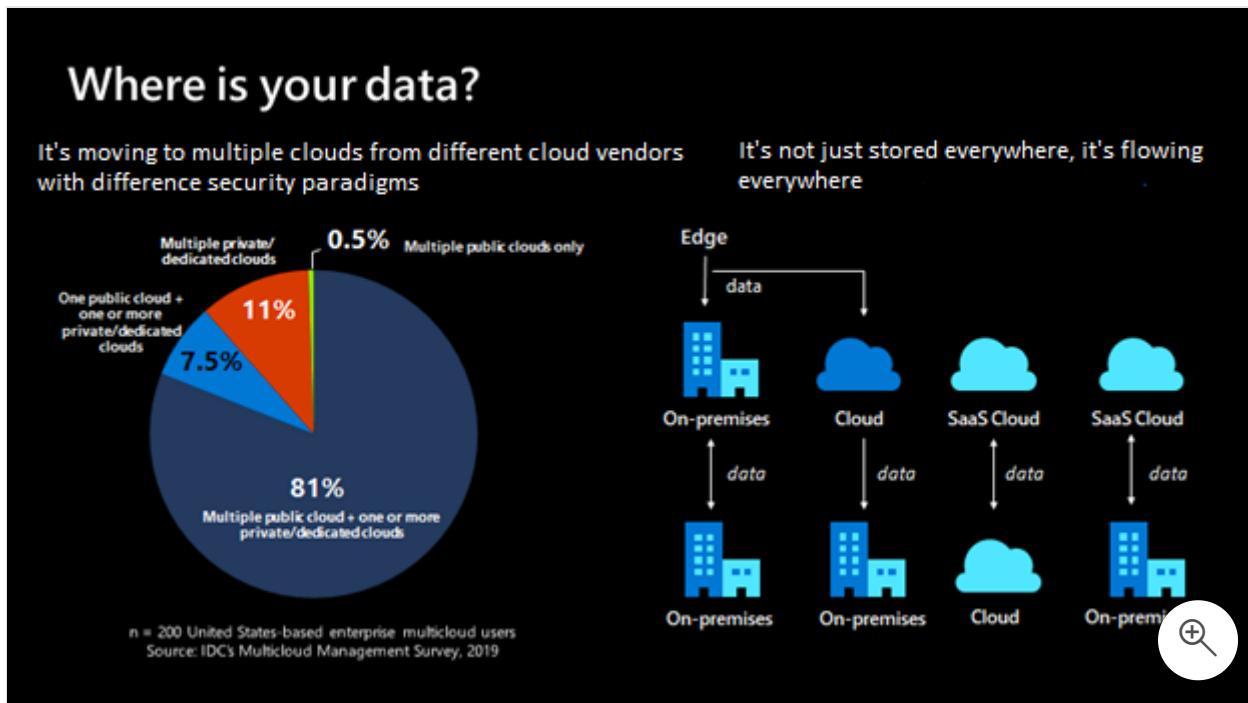


Introduction

2 minutes

As the volume and variety of data increases, the challenges of good data governance are likely to become more difficult. Digital transformation technologies have resulted in new data sources. How do users know what data is available? How do administrators manage data when they might not know what type of data exists and where it's stored? Does the data contain sensitive or personal information?

All these questions aren't easy to answer without insights into the data and the source of storage. Before you can develop data-governance plans for storage and usage, you need to understand the data your organization uses.



Example scenario

As a user or producer of data, you might be a business or technical data analyst, data scientist, or data engineer. You probably spend significant time on manual processes to annotate, catalog, and find trusted data sources.

Because there's no central location to register data sources, you might be unaware of a data source unless you come into contact with it as part of another process.

Writing metadata descriptions for data sources is often a wasted effort. Client applications typically ignore descriptions that are stored in the data source. Creating documentation for data sources is difficult because you must keep documentation in sync with data sources. Users also might not trust documentation that they think is out of date.

Without the ability to track data from end to end, you must spend time tracing problems created by data pipelines that other teams own. If you make changes to your datasets, you can accidentally affect related reports that are business or mission critical.

Microsoft Purview is designed to address these issues and help enterprises get the most value from their existing information assets. The catalog makes data sources easy to discover and understand by the users who manage the data.



What will we be doing?

This high-level overview of Microsoft Purview helps you discover the key aspects that make it the tool of choice for mapping out your enterprise data. You learn how it can help you:

- Manage and govern your data across various platforms and locations.
- Map out your data landscape.
- Classify sensitive data.
- Empower customers to find trustworthy data.

What's the main goal?

By the end of this session, you'll be able to decide if Microsoft Purview is the right choice to help you manage your enterprise data environment and your various data sources.

Next unit: What is Microsoft Purview?

[Continue >](#)

How are we doing?

100 XP

What is Microsoft Purview?

3 minutes

Let's start with a few definitions and a quick tour of the core features of Microsoft Purview.

What's Microsoft Purview?

Microsoft Purview is a unified data-governance service that helps you manage and govern your on-premises, multicloud, and software-as-a-service (SaaS) data. You can easily create a broad, up-to-date map of your data landscape with:

- Automated data discovery.
- Sensitive data classification.
- End-to-end data lineage.

You can also allow data users to find valuable, trustworthy data.

Microsoft Purview is designed to help enterprises get the most value from their existing information assets. With this cloud-based service, you can register your data sources to help you discover and manage them. Your data sources remain in place, but a copy of the metadata for the source is added to Microsoft Purview.

You can register a wide range of sources in Azure and across your multicloud data estate in Microsoft Purview. These sources include Azure Data Lake Storage, AWS, Azure SQL Database on-premises and in the cloud, and many more.

Microsoft Purview has three main elements:

Microsoft Purview Data Map: The data map provides a structure for your data estate in Microsoft Purview, where you can map your existing data stores into groups and hierarchies. In the data map, you can grant users and teams access to these groups so that they have access to find relevant data stores. The data map can then scan your data stores and gather metadata, like schema and data types. It can also identify sensitive data types so that you can keep track of them in your data estate.

The screenshot shows the Microsoft Purview Data Catalog interface. On the left is a vertical sidebar with icons for Home, Register, New collection, Refresh, Sources, Filter by name, and a gear icon. The main area is titled 'Sources' and shows a list of five data sources:

- AzureDataLakeStore-r6l (Azure Data Lake Storage Gen1)
- AzureFileStorage-pX1 (Azure Files)
- CosmosDB-P8b (Azure Cosmos DB (SQL API))
- AzureBlob-B5Z (Azure Blob Storage)
- AzureSqlDatabase-2vb (Azure SQL Database)

Each source entry includes a 'View details' button and edit/copy/delete icons.

Microsoft Purview Data Catalog: The data catalog allows your users to browse the metadata stored in the data map so that they can find reliable data and understand its context. For example, users can see where the data comes from and who are the experts they can contact about that data source. The data catalog also integrates with other Azure products, like the Azure Synapse Analytics workspace, so that users can search for the data they need from the applications they need it in.

Microsoft Purview Data Estate Insights: Insights offer a high-level view into your data catalog, covering these key facets:

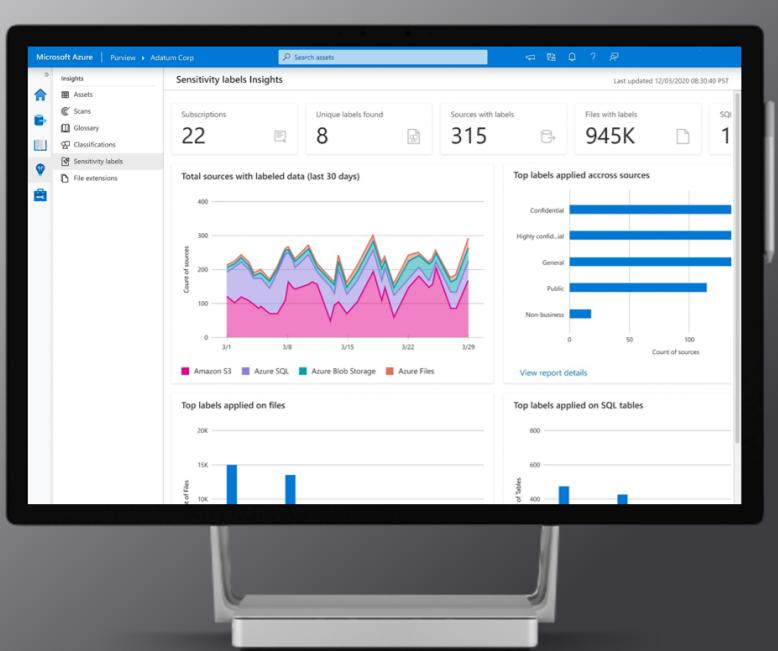
- **Data stewardship:** A report on how curated your data assets are so that you can track your governance progress.
- **Catalog adoption:** A report on the number of active users in your data catalog, their top searches, and your most viewed assets.
- **Asset insights:** A report on the data estate and source-type distribution. You can view by source type, classification, and file size. View the insights as a graph or key performance

indicators.

- **Scan insights:** A report that provides information on the health of your scans (successes, failures, or canceled).
- **Glossary insights:** A status report on the glossary to help users understand the distribution of glossary terms by status, or to view how the terms are attached to assets.
- **Classification insights:** A report that shows where classified data is located. It allows security administrators to understand the types of information found in their organization's data estate.
- **Sensitivity insights:** A report that focuses on sensitivity labels found during scans. Security administrators can make use of this information to ensure security is appropriate for the data estate.

Get a bird's-eye view of sensitive data with Data Estate Insights

- Reports on Assets, Scans, Glossary, Classification, and Labeling



Next unit: How Microsoft Purview works

[Continue >](#)

How are we doing? ☆ ☆ ☆ ☆ ☆

100 XP



How Microsoft Purview works

5 minutes

Here's where we take a look at how Microsoft Purview works. In this unit, you learn the core operational theory behind the functioning of Microsoft Purview for mapping and scanning your data sources. The key areas we focus on include how to:

- Load data in the data map.
- Browse and search information in the data catalog.

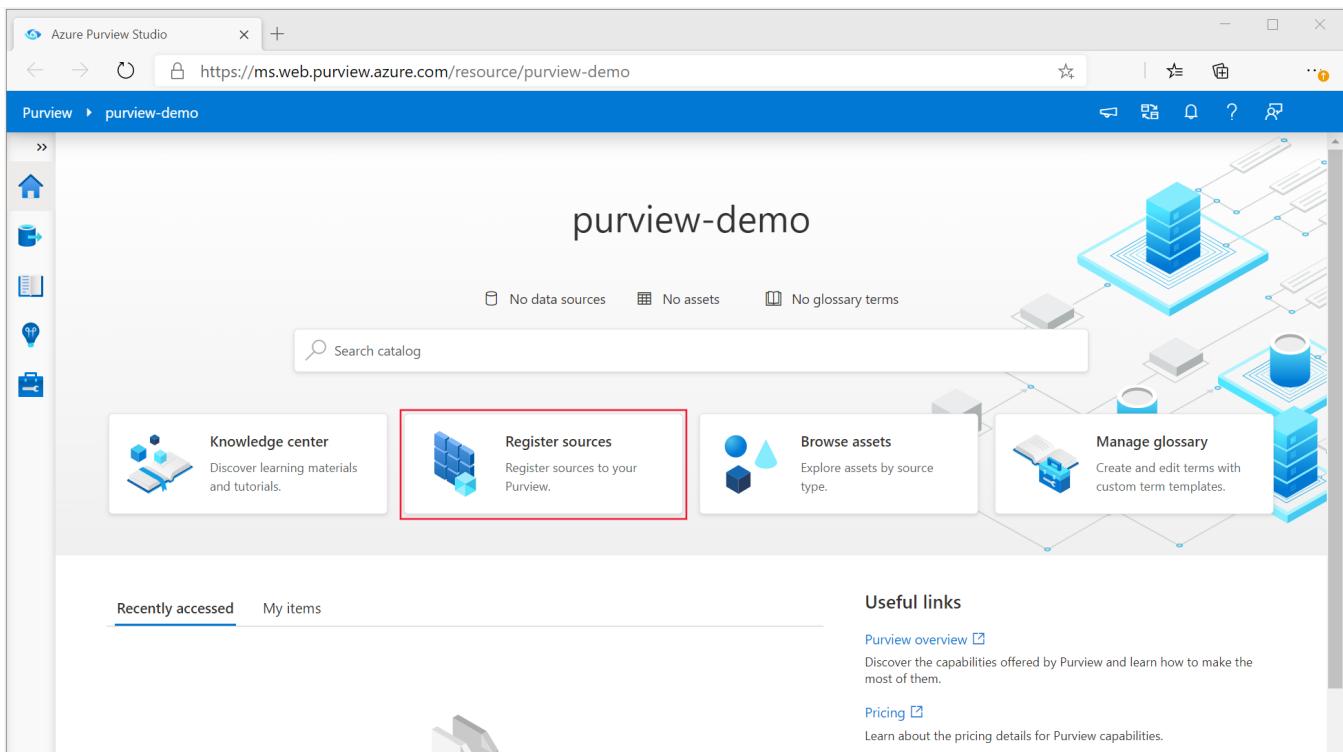
Load data in the data map

The Microsoft Purview Data Map is a unified map of your data assets and their relationships. As one cohesive map, it's easier for you and your users to visualize and govern. It also houses the metadata that underpins the Microsoft Purview Data Catalog and Data Estate Insights. It scales up and down to meet your enterprise compliance requirements. You can use it to govern your data estate in a way that makes the most sense for your business.

Source data

Sourcing your data starts with a process where you register data sources. Microsoft Purview supports an array of data sources that span on-premises, multicloud, and software-as-a-service (SaaS) options. You register the various data sources so that Microsoft Purview is aware of them. The data remains in its location and isn't migrated to any other platform.

After you have a Microsoft Purview service configured in Azure, you use the Microsoft Purview governance portal to register your data sources.



Each type of data source you choose requires specific information to complete the registration. For example, if your data sources reside in your Azure subscription, you choose the necessary subscription and storage account name. The following image is an example of choosing an Azure Blob Storage source.

Register sources (Azure Blob Storage)

Name *

The 'Name' field is highlighted with a red border.

Azure subscription

The 'Azure subscription' field is highlighted with a red border.

Storage account name *

The 'Storage account name' field is highlighted with a red border.

Endpoint

A small blue file icon is at the end of the input field.

Select a collection

A small blue arrow icon is at the end of the input field.

After registration, you scan the data source. Scanning ingests metadata about your data source into the Microsoft Purview Data Map. Each data source has specific requirements for authenticating and configuration to permit scanning of the assets in that data source.

For example, if you have data stored in an Amazon S3 standard bucket, you need to provide a configuration for the connection. For this service, you use Microsoft Purview to provide a Microsoft account with secure access to AWS, where the Microsoft Purview scanner will run. The Microsoft Purview scanner uses this access to your Amazon S3 buckets to read your data. The scanner then reports the results (including only the metadata and classification) back to Azure. You can use the Microsoft Purview classification and labeling reports to analyze and review your data scan results.

 **Note**

Check the [Microsoft Purview connector for Amazon S3 documentation](#) for region support related to AWS S3 sources.

In Microsoft Purview, there are a few options to use for authentication when the service needs to scan data sources. Some of these options are:

- Microsoft Purview managed identity
- Account key (using Azure Key Vault)
- SQL authentication (using Key Vault)
- Service principal (using Key Vault)

Map data

The data map is the foundational platform for Microsoft Purview. The data map consists of:

- Data assets.
- Data lineage.
- Data classifications.
- Business context.

Customers create a knowledge graph of data that comes in from a range of sources. Microsoft Purview makes it easy to register and automatically scan and classify data at scale. Within the data map, you can identify the type of data source, along with other details around security and scanning.

The data map uses collections to organize these details. Collections are a way of grouping data assets into logical categories to simplify management and discovery of assets within the catalog. You also use collections to manage access to the metadata that's available in the data map.

Select **Map view** in the Microsoft Purview governance portal to display the data sources in a graphical view, along with the collections you created for them.

The screenshot shows the 'Sources' section of the Microsoft Purview Governance portal. On the left, there's a sidebar with icons for Home, Data, Metrics, Classifications, and more. The main area displays three collections: 'Sales' (Collection), 'Web-store-front' (Collection), and 'HumanResources' (Collection). Each collection contains one or more data sources. The 'Sales' collection has 'AzureDataLakeStore-r6l' (Azure Data Lake Storage Gen1) and 'AzureSqlDatabase-2vb' (Azure SQL Database). The 'Web-store-front' collection has 'AzureFileStorage-pX1' (Azure Files). The 'HumanResources' collection has 'CosmosDB-P8b' (Azure Cosmos DB (SQL API)) and 'AzureBlob-B5Z' (Azure Blob Storage). Each data source item has a 'View details' button.

Scan data

After you register your data sources, you need to run a scan to access the metadata and browse the asset information. Before you can scan the data sources, you're required to enter the credentials for these sources. You can use Azure Key Vault to store the credentials for security and ease of access by your scan rules. The Microsoft Purview governance portal comes with existing system scan rule sets that you can select when you create a new scan rule. You can also specify a custom scan rule set.

A scan rule set is a container for grouping scan rules together to use the same rules repeatedly. A scan rule set lets you select file types for schema extraction and classification. It also lets you define new custom file types. You might create a default scan rule set for each of your data source types. Then you can use these scan rule sets by default for all scans within your company.

For example, you might want to scan only the .csv files in an Azure Data Lake Storage account. Or you might want to check your data only for credit card numbers rather than all the possible classifications. You might also want users with the right permissions to create other scan rule sets with different configurations based on business need.

The screenshot shows the 'Scan rule sets' section of the Microsoft Purview Governance portal. On the left, there's a sidebar with icons for General, Scan rule sets, Integration runtimes, Metrics, Metadata management, Classifications, Classification rules, Resource sets, Pattern rules, Lineage connections, Data Factory, Data Share, Security and access, Access control, and Credentials. The main area shows a table of scan rule sets. The 'Scan rule sets' tab is selected, showing a list of items: 'System' and 'Custom'. The 'Custom' tab is selected, displaying a table with columns: 'Name', 'Source type', and 'Version'. The table lists various scan rule sets with their corresponding source type and version:

Name	Source type	Version
AzureFileService	Azure Files	Version 2 (Update)
AzureDataExplorer	Azure Data Explorer (Kusto)	Version 1 (Update)
AdlsGen2	Azure Data Lake Storage Gen2	Version 2 (Update)
AdlsGen1	Azure Data Lake Storage Gen1	Version 2 (Update)
AzureSqlDatabase	Azure SQL Database	Version 1 (Update)
AzureSynapseSQL	Azure Synapse Analytics	Version 1 (Update)
AzureCosmosDb	Azure Cosmos DB (SQL API)	Version 1 (Update)
AzureSqlDW	Azure Dedicated SQL Pool (formerly SQL DW)	Version 1 (Update)
AzureSqlDatabaseManagedInstance	Azure SQL Database Managed Instance	Version 1 (Update)
AmazonS3	Amazon S3	Version 1 (Update)
AzureStorage	Azure Blob Storage	Version 1 (Update)
SqlServer	SQL Server	Version 1 (Update)

Classification

Metadata is used to help describe the data that's being scanned and made available in the catalog. During the configuration of a scan set, you can specify classification rules to apply during the scan that also serve as metadata. The classification rules fall under five major categories:

- **Government:** Attributes such as government identity cards, driver license numbers, and passport numbers.
- **Financial:** Attributes such as bank account numbers or credit card numbers.
- **Personal:** Personal information such as a person's age, date of birth, email address, and phone number.
- **Security:** Attributes like passwords that can be stored.
- **Miscellaneous:** Attributes not included in the other categories.

You can use several system classifications to classify your data. These classifications align with the sensitive information types in the Microsoft Purview compliance portal. You can also create custom classifications to identify other important or sensitive information types in your data estate.

General		Classifications		
	Account information	New	Edit	Delete
	Scan rule sets	Refresh		
	Integration runtimes			
	Metrics			
Metadata management		System		
	Classifications	Custom		
	Classification rules			
	Resource sets			
	Pattern rules			
	Lineage connections			
	Data Factory			
	Data Share			
Security and access		These are the system provided classifications.		
	Access control	<input type="text"/> Filter by name...		
	Credentials			
Display name				
ABA Routing Number				
MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...				
ABA Routing Number				
Age of an individual				
MICROSOFT.PERSONAL.AGE				
Age of an individual				
Argentina National Identity (DNI) Number				
MICROSOFT.GOVERNMENT.ARGENITNA.DNI_NU...				
Argentina National Identity (DNI) Number				
Australia Bank Account Number				
MICROSOFT.FINANCIALAUSTRALIA.BANK_ACCOU...				
Australia Bank Account Number				
Australia Business Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS....				
Australia Business Number				
Australia Company Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...				
Australia Company Number				
Australia Driver's License Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...				
Australia Driver's License Number				
Australia Medical Account Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL_...				
Australia Medical Account Number				
Australia Passport Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...				
Australia Passport Number				
Australia Tax File Number				
MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE...				
Australia Tax File Number				
Austria Driver's License Number				
MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LIC...				
Austria Driver's License Number				
Austria Identity Card				
MICROSOFT.GOVERNMENT.AUSTRIA.IDENTITY.CA...				
Austria Identity Card				

After you register a data source, you can enrich its metadata. With proper access, you can annotate a data source by providing descriptions, ratings, tags, glossary terms, identifying experts, or other metadata for requesting data-source access. This descriptive metadata supplements the structural metadata, such as column names and data types, that's registered from the data source.

Discovering and understanding data sources and their use is the primary purpose of registering the sources. If you're an enterprise user, you might need data for business intelligence, application development, data science, or any other task where the right data is

required. You can use the data catalog discovery experience to quickly find data that matches your needs. You can evaluate the data for its fitness for the purpose and then open the data source in your tool of choice.

At the same time, you can contribute to the catalog by tagging, documenting, and annotating data sources that have already been registered. You can also register new data sources, which are then discovered, evaluated, and used by the community of catalog users.

Browse and search

Microsoft Purview allows you to search information from the data map by using the Microsoft Purview Data Catalog. You can perform text-based search and browse through results by using filters like data source type, tags, ratings, or collection.

You can use business context to search information from the Microsoft Purview catalog. You can define business glossaries and bulk import existing ones, too. You can also apply business context onto assets in the data map. By using a metamodel, you can define business processes in your environment and associate your data sources with those processes. Users can then apply these business contexts to browse and search for information in the data catalog.

Discovery enables you to use:

- Semantic search and browse.
- Business glossary and workflows.
- Data lineage with sources, owners, transformations, and lifecycle.

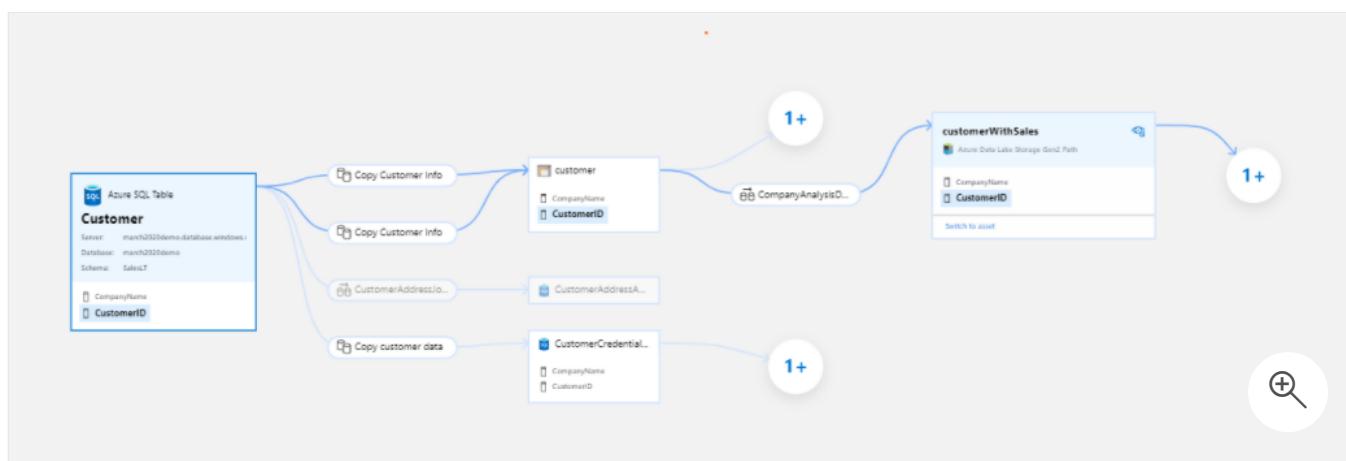
The screenshot shows the Microsoft Purview Data Catalog interface for Adatum Corp. The top navigation bar includes 'Microsoft Azure', 'Purview', and 'Adatum Corp'. The main dashboard displays statistics: 13,620 users, 1,134 sources, 76,392,971 assets, and 312 terms. A search bar at the top left contains the term 'Revenue'. Below it, a sidebar features a 'Knowledge center' icon and a 'Recently accessed' section listing items like 'Order', 'Percent Sales', 'SalesOrderHeader', 'ProductCategory', 'SSNNumber', and 'SalesLT.SalesOrderHeader.csv'. The main content area shows 'Search suggestions' for 'revenue customer', 'revenue finance', 'revenue glossary', and 'revenue information'. It also lists 'Asset suggestions' for 'TaxRevenue', 'CustomerRevenue_{N}', 'Revenue', and 'Revenue2020.tsv'. A 'View search results' button is located at the bottom of this list. To the right, there's a 'View insights' section with a callout: 'Set insights on your data.' and 'and learn how to make the most of our capabilities.' The overall interface is light-colored with blue and grey accents.

Data lineage

The concept of data lineage focuses on the lifecycle of data. The lifecycle concerns itself with the various stages data might go through. Data is sourced, moved, and stored throughout its lifecycle. Data might also undergo transformations in the extract, load, and transform/extract, transform, and load (ELT/ETL) operations.

Data lineage can offer insights into the data lifecycle by looking at the data pipeline. You can use the lineage to identify the root cause of data issues, perform data quality analysis, and verify compliance.

Microsoft Purview represents this data lineage in a visual form by showing data movement from source to destination.



100 XP

When to use Microsoft Purview

5 minutes

In this unit, we discuss how you can decide if Microsoft Purview is the right choice for your data governance and discovery needs. The criteria that indicate whether Microsoft Purview will meet your requirements are:

- Discovery
- Governance

Let's take a look at the criteria and see how Microsoft Purview can help address the needs in those specific areas.

Discovery

Without a central location to register data sources, you might be unaware of a data source unless you come into contact with it as part of another process.

Unless you know the location of a data source, you can't connect to the data by using a client application. You're required to know the connection string or path.

The intended use of the data is hidden to you unless you know the location of a data source's documentation. Data sources and documentation might live in several places and be utilized through different kinds of experiences.

Governance

As the data in your organization grows, the task of discovering, protecting, and governing that data becomes more difficult. Data is stored in different locations, which might be required for compliance reasons. The data might contain sensitive information such as credit card numbers, social security numbers, or other personal information.

Compliance with company security policies, government regulations, and customer needs are critical considerations for data governance. Understanding which data sources contain sensitive information is key to knowing where protections are needed and how to guard against access to this sensitive data.

The screenshot shows the Microsoft Purview governance portal. The left sidebar has a 'Classifications' item selected. The main area is titled 'Classifications' with tabs for 'System' and 'Custom'. It says 'These are the system provided classifications.' and includes a 'Filter by name...' search bar. The table below lists 14 system-provided classifications:

Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA Routing Number
Age of an individual	MICROSOFT.PERSONAL.AGE	Age of an individual
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGENТИNA.DNI_NU...	Argentina National Identity (DNI) Number
Australia Bank Account Number	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACCOU...	Australia Bank Account Number
Australia Business Number	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Australia Business Number
Australia Company Number	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Australia Company Number
Australia Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...	Australia Driver's License Number
Australia Medical Account Number	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL_...	Australia Medical Account Number
Australia Passport Number	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...	Australia Passport Number
Australia Tax File Number	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE_...	Australia Tax File Number
Austria Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LIC...	Austria Driver's License Number
Austria Identity Card	MICROSOFT.GOVERNMENT.AUSTRIA.IDENTITY.CA...	Austria Identity Card

Apply the criteria

Let's take a look at how Microsoft Purview can address the data discovery and governance criteria.

Does Microsoft Purview help with data discovery?

Do you require a solution or centralized location to register data sources? Often, users might be unaware of a data source unless they come into contact with it as part of another process. Microsoft Purview can help to provide a solution.

After you've registered data sources in the Microsoft Purview governance portal and displayed them in the data map, you can set up scanning of those data sources. The metadata that's returned catalogs the data in those sources. In this way, it's easier for users to discover what the data sources contain. The metadata is indexed to make each data source easy to discover via search. It's also more understandable to the users who discover it.

Users can contribute to the catalog by tagging, documenting, and annotating data sources that have already been registered. They can register new data sources so that other catalog users can discover, understand, and utilize them.

The screenshot shows the Microsoft Purview Sources page. At the top, there's a navigation bar with 'Purview' and 'purview-demo'. A search bar says 'Search assets'. On the right are icons for refresh, help, and more. Below the navigation is a sidebar with icons for Home, Sources, Collections, Assets, and Glossary. The 'Sources' icon is highlighted with a red box. The main area has a title 'Sources' with a back arrow. It includes buttons for 'Register', 'New collection', and 'Refresh'. A message says 'Showing 1 source'. On the right, there's a dropdown menu 'List view' with a red box around it. The table lists one source: 'blob-storage' (Name), 'Azure Blob Storage' (Source type), '-' (Collection), 'Source id' (empty), 'Scans' (with a red box around the 'New scan' button), and 'Registered on' (11/24/20 11:45 ...). There are edit, refresh, and delete icons next to the source name.

Does Microsoft Purview help with data governance?

Microsoft Purview can scan and automatically classify data in files and tables. Microsoft Purview classifies data by Bloom Filter and RegEx. Bloom Filter classifications include attributes for city, country/region, place, and person information. RegEx classifications cover attributes that include categories like bank information (ABA routing numbers or country/region-specific banking account numbers), passport numbers, and country/region-specific identification numbers. You can find the [full list of supported classifications](#) in the documentation for Microsoft Purview.

Microsoft Purview also uses predefined Data Plane roles to help control who has access to the information in Microsoft Purview. For access, users can use the Microsoft Purview governance portal only if they're placed in at least one of the three supported roles. When a Microsoft Purview account is created, no one but the creator can access the account or use its APIs. New users must be put in one or more of the following roles:

- **Purview Data Reader role:** Has access to the Microsoft Purview governance portal and can read all content in Microsoft Purview except for scan bindings.
- **Purview Data Curator role:** Has access to the Microsoft Purview governance portal and can read all content in Microsoft Purview except for scan bindings. Can edit information about assets, classification definitions, and glossary terms. Can also apply classifications and glossary terms to assets.
- **Purview Data Source Administrator role:** Doesn't have access to the Microsoft Purview governance portal because the user must also be in the Data Reader or Data Curator roles. Can manage all aspects of scanning data into Microsoft Purview. Doesn't have read or write access to content in Microsoft Purview beyond those tasks related to scanning.

These roles are assigned by using the collections where your data sources are registered. You can grant users access to the data they might need without granting them access to the entire

data estate. By assigning roles, you can promote resource discoverability while still protecting sensitive information.

Next unit: Knowledge check

[Continue >](#)

How are we doing?

✓ 200 XP



Knowledge check

15 minutes

Choose the best response for each of the following questions, and then select **Check your answers**.

1. What does Microsoft Purview do with the data it discovers from your registered sources? *

*



It catalogs and classifies the data that's scanned.

✓ Correct. It creates a broad, up-to-date map of your data landscape with automated data discovery, sensitive data classification, and end-to-end data lineage. It empowers data consumers to find valuable, trustworthy data.



It moves the data to your Azure subscription, automatically creating the necessary storage accounts.

✗ Incorrect. Microsoft Purview doesn't move the data. It leaves the data in its original storage location but maintains a link to the data.



It performs data transformations to match your on-premises schemas.

2. Where would you register your data sources for use in Microsoft Purview? *



On the Overview tab of the Microsoft Purview account page.

✗ Incorrect. The overview page for the Microsoft Purview account provides information about that account only.



On the Managed Resources tab of the Microsoft Purview account page.



In the Microsoft Purview governance portal.

✓ Correct. The Microsoft Purview governance portal is where you register your data sources.

3. What aspect of Microsoft Purview is used to configure the data discovery for your data sources? *



Scan rules

✓ Correct. In a Microsoft Purview catalog, you can create scan rule sets to enable you to quickly scan data sources in your organization.



Collections



Classifications

✗ Incorrect. Classifications are used to identify key attributes of the data in the data sources. The five categories are government, financial, personal, security, or miscellaneous.

Next unit: Summary

[Continue >](#)

How are we doing?

100 XP



Introduction

2 minutes

Microsoft Purview is a unified data governance service that helps you manage and govern your on-premises, multi-cloud, and software-as-a-service (SaaS) data. Data professionals can easily create a holistic, up-to-date map of the entire data landscape. Microsoft Purview includes automated data discovery, sensitive data classification, and end-to-end data lineage. Microsoft Purview can empower data analysts and other data consumers to find valuable, trustworthy data.

Imagine that you're a new Data Analyst at Contoso. In your second week, the sales manager desperately asks for the latest inventory and sales data for an impromptu review. After clarifying the requirements, you know you need to quickly find accurate assets to create a report. With the help of Microsoft Purview data catalog, you'll be able to search, browse, and discover assets. More importantly, you'll be able to validate that you're using the right data sources for your reports.

Learning objectives

In this module, you will:

- Browse and search data catalog assets.
- Use data catalog assets with Power BI.
- Use Microsoft Purview in Azure Synapse Studio.

Next unit: Search for assets

[Continue >](#)

How are we doing?

✓ 100 XP ➔

Search for assets

10 minutes

Microsoft Purview offers a central place to discover and understand assets to use in your day-to-day activities. This central place, Microsoft Purview data catalog, provides advanced search capabilities to quickly find the right assets and information. Using keywords, business terms and Microsoft Purview data catalog functionalities, you can find the assets needed to build and design reports.

As a data analyst looking for assets, you'll be searching the Microsoft Purview data catalog. This assumes that the *Microsoft Purview data Map* has been created by your organization. The data map provides the foundation for data discovery. The data map captures metadata about enterprise data in analytics and operations systems on-premises and in the cloud and must be established before the data catalog can be searched.

ⓘ Note

Learn more about the [Microsoft Purview data map components](#).

Search the Microsoft Purview data catalog

From the Microsoft Purview Studio home page, users can type relevant keywords to start discovering assets. In this scenario, you're looking for "product sales."

Azure Purview > Contoso Purview Catalog

CONTOSOHOTELS.COM

Contoso Purview Catalog

30 sources | 768 assets | 195 glossary terms

product sales

Search suggestions

- product Sales QuotaStats
- product sales **AccountRankCurve**
- product SalesWithCustomerID
- product sales

Asset suggestions

- ProductSalesReport

The screenshot below displays the search result, with all assets corresponding to the keywords entered in the search engine.

The screenshot shows the Microsoft Purview Data Catalog interface. The top navigation bar includes a back arrow, the title 'Data catalog >', and a search bar containing 'Search results for product sales'. Below the search bar are filters for 'Source type : all' and 'Instance : all', along with a 'Clear all filters' button. A 'Filter by keyword' input field is also present. On the left side, there are three vertical columns of filters: 'Object Type' (Dashboards, Data pipelines, Files, Folders, Reports, Stored procedures, Tables), 'Collection' (Sales, Contoso Purview Catalog, IT), and 'Classification' (Franmer Product, Credit Card Number, Country/Region, Person's Name). The main content area displays search results for 'product sales', showing four items: 'Product Mix.csv' (Azure Blob Path), 'Sales Dashboard-SQLDB' (Power BI Dataset, Secret), 'SalesWithCustomerID.csv' (Azure Blob Path), and 'SalesWithCustomerID' (Azure SQL Table, certified). The 'SalesWithCustomerID' item is described as a 'Table with sales raw information.' The 'ReportTable_Prod' item is described as a 'Report Table in production environment to create Power BI report for sales people. Ta...'. The results are sorted by Relevance.

You can fine-tune your search using the filters on the left side of the page.

You can filter by:

- Source type (and instance if needed)
- Object type
- Classification
- Glossary term
- If needed, more options are available like Collection, Contact and Label

The screenshot shows the Azure Purview Catalog interface. At the top, there's a search bar with the query "product sales". Below the search bar, a red box highlights the "Source type : Azure SQL Database" filter. Another red box highlights the "Tables" filter under the "Object Type" section. On the left, there are navigation icons and a sidebar with sections like "Classification", "Contact", and "Label". The main results area shows two items: "ReportTable_Prod" (certified) and "Product". The "ReportTable_Prod" item is described as a "Report Table in production environment to create Power BI report for sales people. Table de reporting ...". The "Product" item is described as an "Azure SQL Table".

Search results for product sales

Source type : Azure SQL Database

Instance : all Clear all filters

Filter by keyword

Showing 1-2 out of 2 results

Sort by: Relevance

Object Type

- Dashboards
- Data pipelines
- Files
- Folders
- Reports
- Stored procedures
- Tables

Collection

Classification

- Franmer Product (11)
- Franmer SIN (7)
- Person's Name (7)
- Canada Social Insurance Number (5)
- EU Phone Number (5)
- U.S. Phone Number (5)
- Country/Region (4)
- EU Mobile Phone Number (3)
- Credit Card Number (3)

Contact

-
-
-

Label

Glossary term

- Products (3)
- Sales (3)
- Customer Product (4)
- Product client (4)
- Consumer Products (3)
- Global Markets (3)
- Personal Information (3)
- Saleswithtax (3)
- Administrative and Support and ... (2)
- Ancillary (2)

You've been instructed to connect to sources like Azure SQL tables. In the result displayed below, two assets are displayed. To use the correct asset, it's possible to browse each asset to dig for more detailed information. Alternatively, you can rely on the work done by the data stewards who have labeled certified assets for the organization.

Microsoft Azure | Azure Purview > Contoso Purview Catalog

Search results for product sales

Source type : Azure SQL Database Instance : all Clear all filters

Filter by keyword

Object Type

- Dashboards
- Data pipelines
- Files
- Folders
- Reports
- Stored procedures
- Tables

Collection

Showing 1-2 out of 2 results

ReportTable_Prod (certified) Azure SQL Table Report Table in production environment to create Power BI report for sales people.

Product Azure SQL Table

Before using this asset to create your report, you need to verify more details and validate where data comes from to populate this asset. Select the asset to access more information.

Understand a single asset

Asset overview

Select an asset to see the *overview*. The overview displays information at a glance, including a description, asset classification, schema classification, collection path, asset hierarchy, and glossary terms.

Microsoft Azure | Azure Purview > Contoso Purview Catalog

Search results "product sales"

ReportTable_Prod (Certified) Azure SQL Table

Edit Select for bulk edit Request access Refresh Delete

Open in Power BI Desktop

Updated on March 15, 2022 10:16 PM UTC by

Overview Properties Schema Lineage Contacts Related

Asset description

Report Table in production environment to create Power BI report for sales people.

Classifications (4)

- Credit Card Number
- Person's Name
- Franmer Product
- Numero De Carte De Credit

Schema classifications (9)

- Canada Social Insurance Number
- Country/Region
- Credit Card Number
- EU Mobile Phone Number
- EU Phone Number
- Person's Name
- U.S. Phone Number
- Franmer Product
- Franmer SIN

Fully qualified name

Collection path

- Contoso Purview Catalog
- PurviewNinja
- Cloud_Data
- EMEA
- NinjaSales

Hierarchy

- Azure SQL Server
- Azure SQL Database
- Azure SQL Schema
- ReportTable_Prod

Glossary terms (6)

- Customer Product
- Produit client
- Products
- Sales
- Saleswithtax

The *asset description* provides a brief explanation of the purpose of an asset. Data stewards have made data analysts lives easier in the screenshot below, by noting that this is the correct resource to use for sales reporting.



ReportTable_Prod

Certified

Azure SQL Table



Edit



Select for bulk edit



Request access



Refresh



Delete

[Overview](#)[Properties](#)[Schema](#)[Lineage](#)[Contacts](#)[Related](#)

Asset description

Report Table in production environment to create Power BI report for sales people.

Beneath the description, you'll see the *asset classification* and *schema classification*.

Data classification, in the context of Microsoft Purview, is a way of categorizing data assets by assigning unique logical labels or classes. Classification is based on the business context of the data. For example, you might classify assets by Passport Number, Driver's License Number, Credit Card Number, SWIFT Code, Person's Name, and so on. Asset classifications can be automatically applied during a scan or applied manually.

Note

Microsoft Purview comes with more than 200 classifications out of the box. For a full list of classifications, see [System classifications in Microsoft Purview](#).

The overview tab reflects both asset level classifications and column level classifications that have been applied, which you can also view as part of the schema.



ReportTable_Prod

Certified

Azure SQL Table



Edit



Select for bulk edit



Request access



Refresh



Delete

[Overview](#)[Properties](#)[Schema](#)[Lineage](#)[Contacts](#)[Related](#)

Asset description

Report Table in production environment to create Power BI report for sales people.

Classifications (4)

Credit Card Number

Person's Name

Franmer Product

Numero De Carte De Credit

Schema classifications (9)

Canada Social Insurance Number

Country/Region

Credit Card Number

EU Mobile Phone Number

EU Phone Number

Person's Name

U.S. Phone Number

Franmer Product

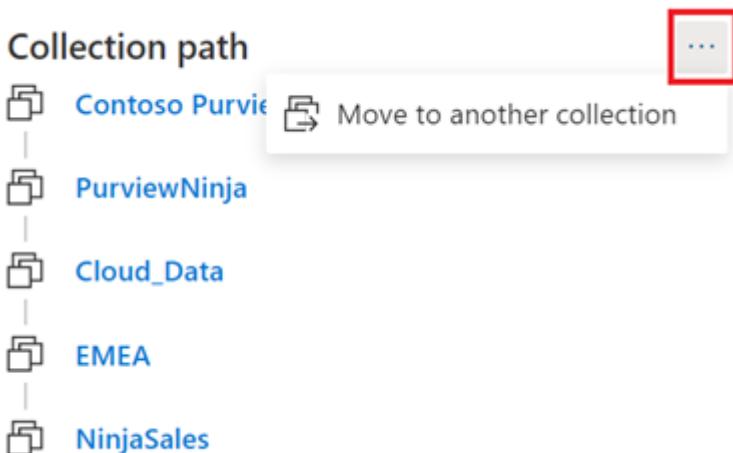
Franmer SIN

Important

You may notice that the classifications displayed above are sensitive or contain personally identifiable information (PII). data encryption is done at the source level, and Microsoft Purview stores only the metadata. It does not preview data.

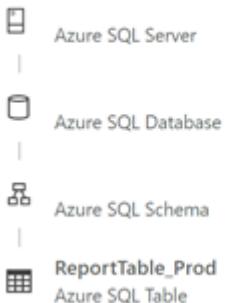
You can also view the *collection path*, *hierarchy* and *glossary terms* on the right side of the overview tab.

The *collection path* refers to the location of the asset inside Microsoft Purview. You have the option to move an asset to another collection.



You can view the full asset hierarchy within the overview tab. As an example: if you navigate to a SQL table, then you can see the schema, database, and the server the table belongs to.

Hierarchy



Glossary terms are a managed vocabulary for business terms that can be used to categorize and relate assets across your environment. For example, terms like 'customer,' 'buyer,' 'cost center,' or any terms that give your data context for your users. You can view the glossary terms for an asset in the overview section, and you can add a glossary term on an asset by editing the asset.

Glossary terms (6)

Customer Product

Produit client

Products

Sales

Saleswithtax

SSN

ⓘ Note

For more information, see the [business glossary page](#).

Asset schema

The *schema* view of the asset includes more granular details about the asset, such as column names, data types, column level classifications, terms, and descriptions.

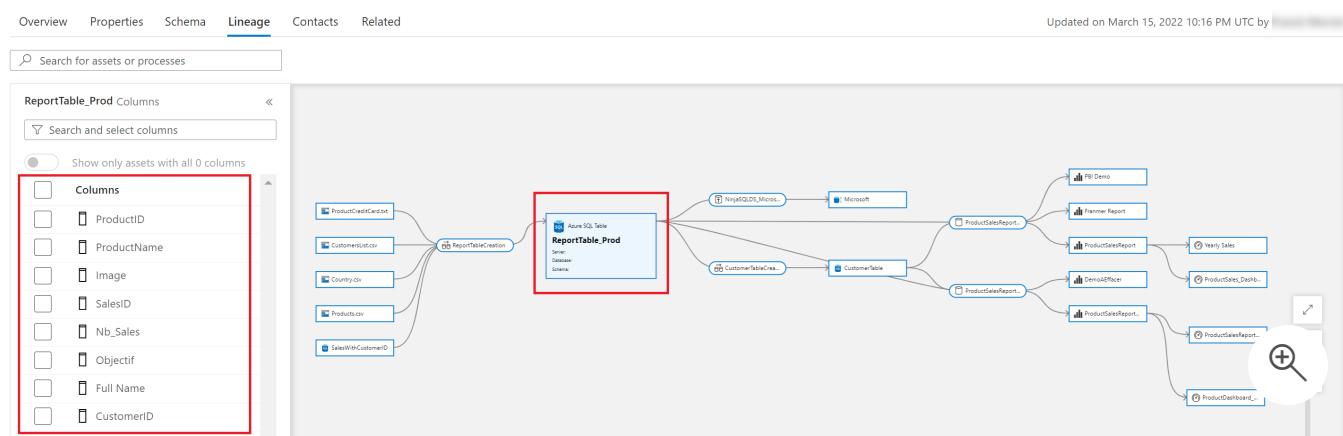
ReportTable_Prod Certified				
Azure SQL Table				
Edit Select for bulk edit Request access Refresh Delete Edit columns				
Overview	Properties	Schema	Lineage	Contacts
Related	Updated			
<p>ⓘ The schema was modified by on 03/15/2022, 10:16:53 PM UTC.</p> <p>Filter by name</p>				
Showing 17 of 17 items				
Column name	Classifications	Glossary terms	Data type	Asset description
CardNumber	Credit Card Number	Payments	nvarchar	Card number compliant with Luhn test
CardType		Payments	nvarchar	Card type (VISA, MASTERCARD, AMEX,...)
Country	Country/Region		nvarchar	Country Name
CountryID			nvarchar	Unique Identifier for country
CustomerID			nvarchar	Unique identifier for customer
Flag			nvarchar	Link to flag picture stored in data lake
Full Name	Person's Name	Personal Information	nvarchar	Full Name
Image		Consumer Products	nvarchar	Link to product picture stored in data lake
Nb_Sales		Sales	int	Technical name from database
Nom	Person's Name	Personal Information	nvarchar	Name
Objectif		Sales	int	Objective in CAD (with taxes)
Phone	EU Mobile Phone Number +2 More	Personal Information	nvarchar	Phone number
Prenom	Person's Name	Personal Information	nvarchar	First Name
ProductID			nvarchar	Unique Identifier for product

Asset lineage

Asset lineage gives you a clear view of how the asset is populated and where data comes from. *Data lineage* is broadly understood as the lifecycle that spans the data's origin, and where it moves over time across the data estate. Data lineage is important to analysts because it enables understanding of where data is coming from, what upstream changes may have occurred, and how it flows through the enterprise data systems.

A single view on the asset lineage tab displays the data flow to and from the asset. Asset lineage can also help you understand how the asset was built and how the asset is used inside the organization.

The columns pane on the left side of the lineage tab allows users to select and track columns as they flow through the lineage. For example, if you select the column Full Name, you can see how the Full Name field was created and where the information comes from.



! Note

The lineage view is a powerful way to understand the transformation process an asset has undergone. Learn more about the [lineage experience in Microsoft Purview data catalog](#)

Asset contacts and related assets

Asset *contacts* provide you contact details of experts or dataset owners with any questions. As a new analyst searching for the right data sources for your report, you may find these individuals helpful.

**ReportTable_Prod**

Certified

Azure SQL Table



Edit

Select for bulk edit



Request access



Refresh



Delete

Overview

Properties

Schema

Lineage

Contacts

Related

Here are the people you can contact about this asset:

Experts (2) ^①**Owners (2) ^①**

If needed, you can also navigate through the technical hierarchy of assets that are related to the current asset you're viewing.

The screenshot shows the Microsoft Purview Data Catalog interface. At the top, there's a search bar with the text "product sales". Below the search bar, the breadcrumb navigation shows "Data catalog > Search results "product sales"" and the current page "Contoso Purview Catalog". On the left, there's a vertical sidebar with icons for Home, Data Catalog, Data Governance, Data Quality, Data Privacy, and Data Integration. The main content area displays a table asset named "ReportTable_Prod" with a green "Certified" badge. Below the table are buttons for Edit, Select for bulk edit, Request access, Refresh, and Delete. A horizontal menu bar includes Overview, Properties, Schema, Lineage, Contacts, and Related, with "Related" being the active tab and highlighted with a red box. To the left of the main content, there's a tree view showing the hierarchy: Azure SQL Server, Azure SQL Database, and Azure SQL Schema. Under "Azure SQL Schema", "ReportTable_Prod" is listed. The right side of the screen shows a list of related items under the heading "Showing 1 to 4 of 4 items", with columns for Name and a small icon. The items listed are CustomerTable, ReportTable, ReportTable_Prod, and SalesWithCustomerID.

The ability to search the Microsoft Purview data catalog has the potential to break down data silos and enable the next level of enterprise analytics.

Next unit: Browse assets

[Continue >](#)

How are we doing? ☆ ☆ ☆ ☆ ☆

✓ 100 XP ➔

Browse assets

4 minutes

Searching a data catalog is a great tool for data discovery you know what you're looking for. Often, you may not know how your data estate is structured. The Microsoft Purview data catalog offers a browse experience that enables exploration of available data, either by collection or by exploring the hierarchy of each data source in the catalog.

Browse by collection or source type

If you're new to an organization or department, you may want to familiarize yourself with the contents of the data estate. From the Microsoft Purview Studio home page, select the "Browse assets" tile to browse either by collection or by source type.

Azure Purview ▶ Contoso Purview Catalog

Contoso Purview Catalog

30 sources | 768 assets | 195 glossary terms

Search catalog

Browse assets
Explore assets by source type and collection.

Manage glossary
Create and edit terms with custom term templates.

Knowledge center
Discover learning materials and tutorials.

Recently accessed

Name	Last update
ReportTable_Prod	a day ago

My items

Links

Azure Purview overview

Discover the capabilities offered by Azure Purview and learn the most of them.

Here you can specify whether you'd like to browse by collection or by source type.

Browse by collection allows you to explore the different collections you're a data reader or curator for. You'll only see collections you have access to. Select a collection to get a list of assets in that collection with the facets and filters available in search.

Microsoft Azure | Purview > AdventureWorks

„view“ AND categories X

Data catalog >

Browse assets

Refresh

By collection By source type

View collection tree

Source Systems

AdventureWorks > ... > North America >

Sub collection(s)

Related

North America (Parent)
Curated Zone
Raw Zone

Narrow results by:

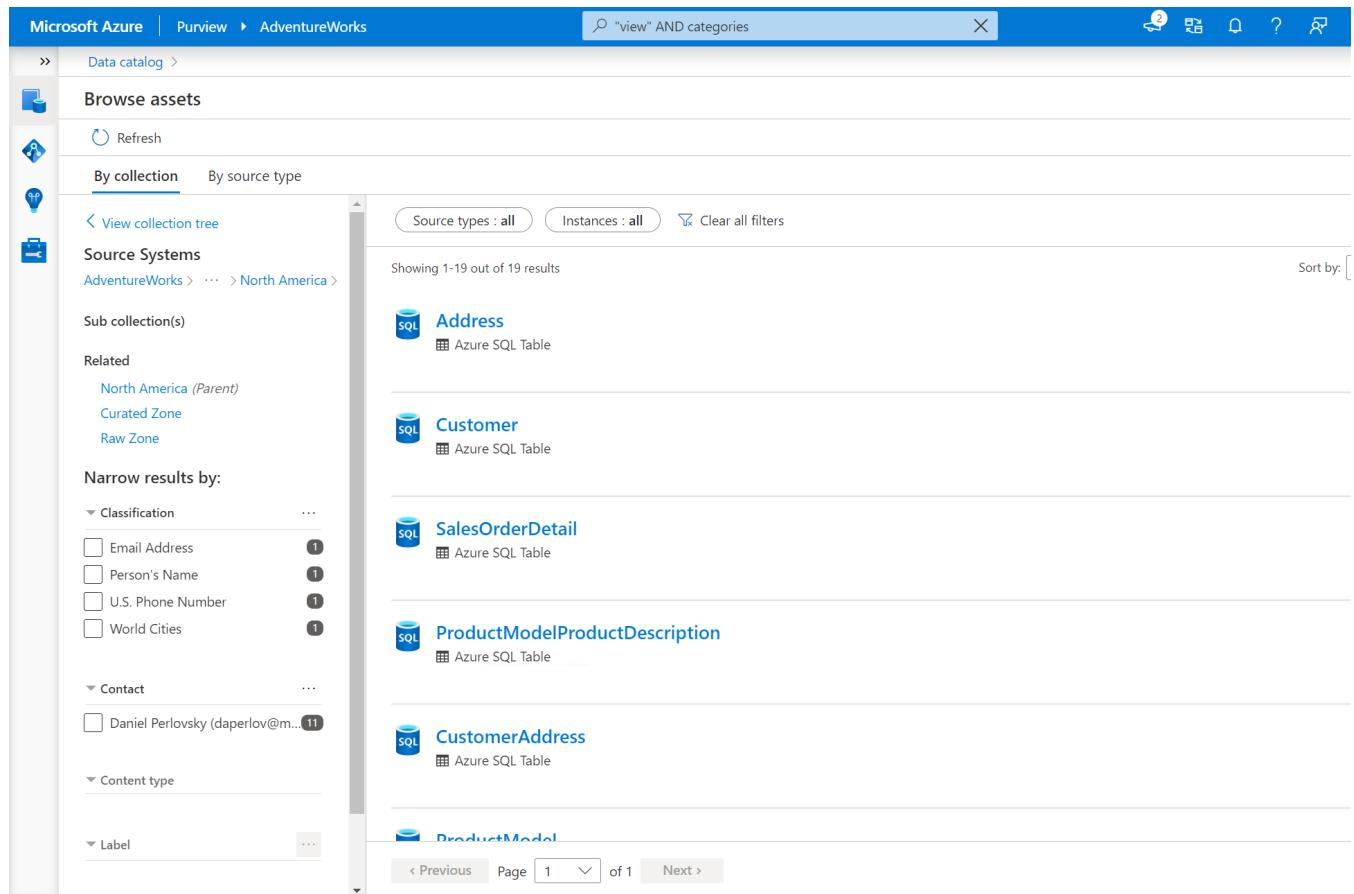
- Classification ...
 - Email Address 1
 - Person's Name 1
 - U.S. Phone Number 1
 - World Cities 1
- Contact ...
 - Daniel Perlovsky (daperlov@m... 11
- Content type
- Label ...

Source types : all Instances : all Clear all filters

Showing 1-19 out of 19 results Sort by: []

 Address	Azure SQL Table
 Customer	Azure SQL Table
 SalesOrderDetail	Azure SQL Table
 ProductModelProductDescription	Azure SQL Table
 CustomerAddress	Azure SQL Table
 ProductModel	Azure SQL Table

< Previous Page 1 of 1 Next >



Tip

Collections are a tool to manage ownership and access control across assets and data sources. They also organize assets and sources into categories that are customized to match the business flow. See [Create and manage collections in Microsoft Purview](#) to learn more.

Browse by source type allows you to explore the hierarchies of data sources using an explorer view.

After selecting a tile associated with a data source type, you'll see a list of assets belonging to that type. From there, you'll be able to use the explorer view to see parent and child assets.

Name	Type	Owner	Action
Address	Azure SQL Table	-	...
Address_SchemaChange	Azure SQL Table	-	...
Customer	Azure SQL Table	-	...
CustomerAddress	Azure SQL Table	-	...
Product	Azure SQL Table	-	...
ProductCategory	Azure SQL Table	-	...
ProductDescription	Azure SQL Table	-	...
ProductModel	Azure SQL Table	-	...
ProductModelProductDescription	Azure SQL Table	-	...
SalesOrderDetail	Azure SQL Table	-	...
SalesOrderHeader	Azure SQL Table	-	...
vGetAllCategories	Azure SQL View	-	...
vProductAndDescription	Azure SQL View	-	...
vProductModelCatalogDescription	Azure SQL View	-	...
VW_ColumnDescription	Azure SQL View	-	...
VW_ColumnDescriptionNew	Azure SQL View	-	...

The Microsoft Purview data catalog browse experience enables analysts or data consumers to explore what data is available in many different ways. Microsoft Purview has the ability to enable users access to data you may not have known about before. The possibilities are endless so long as your organization's data stewards have scanned and classified data across the estate.

✓ 100 XP

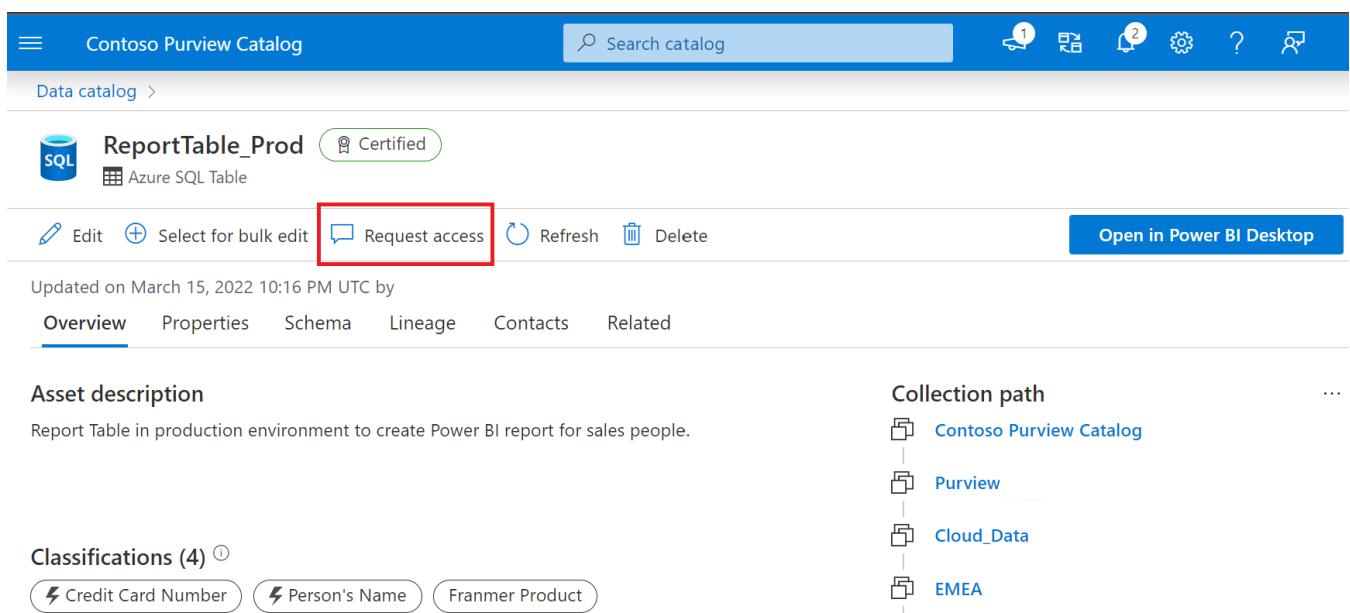
Use assets with Power BI

3 minutes

The integration of Microsoft Purview and Power BI makes it possible to gain a more complete understanding of the data across your estate.

Request access to assets

In your search or browsing session, you may come across assets that you don't have access to. Microsoft Purview makes it simple to request access directly from the Data Catalog by using the "Request access" button. Requesting access will kick off a workflow that manages requests and approvals.



The screenshot shows the Microsoft Purview Catalog interface. At the top, there's a blue header bar with the title 'Contoso Purview Catalog'. Below the header, a navigation bar includes 'Data catalog >', a search bar labeled 'Search catalog', and several icons. The main content area displays a table asset named 'ReportTable_Prod'. This asset is marked as 'Certified' and is identified as an 'Azure SQL Table'. Below the asset name, there are several action buttons: 'Edit', 'Select for bulk edit', 'Request access' (which is highlighted with a red box), 'Refresh', and 'Delete'. To the right of these buttons is a blue button labeled 'Open in Power BI Desktop'. Underneath the asset details, there's a note about the last update: 'Updated on March 15, 2022 10:16 PM UTC by'. Below this, a horizontal menu bar has 'Overview' underlined and other options like 'Properties', 'Schema', 'Lineage', 'Contacts', and 'Related'. The 'Asset description' section contains a brief description: 'Report Table in production environment to create Power BI report for sales people.' The 'Collection path' section shows the asset's hierarchy: 'Contoso Purview Catalog' > 'Purview' > 'Cloud_Data' > 'EMEA'. The 'Classifications (4)' section lists four categories: 'Credit Card Number', 'Person's Name', and 'Frammer Product'.

Build a Power BI report using data discovered in Purview

Working as a new analyst, you've taken the time to search and browse assets and now you'd like to use those trusted assets in a Power BI report. Purview makes it simple, with the ability to open the asset in Power BI desktop.

The screenshot shows the Microsoft Purview Catalog interface. At the top, there's a search bar labeled "Search catalog". Below the search bar, the title "Contoso Purview Catalog" is displayed, followed by "Data catalog >". On the left, there's a sidebar with a "SQL" icon and the name "ReportTable_Prod" next to a "Certified" badge. Below this, it says "Azure SQL Table". A toolbar at the top has icons for "Edit", "Select for bulk edit", "Request access", "Refresh", and "Delete". To the right of the toolbar is a blue button with white text that says "Open in Power BI Desktop", which is highlighted with a red box and a red arrow pointing to it from below. Below the toolbar, the text "Updated on March 15, 2022 10:16 PM UTC by" is shown. Underneath, there are tabs for "Overview" (which is underlined), "Properties", "Schema", "Lineage", "Contacts", and "Related".

Asset description

Report Table in production environment to create Power BI report for sales people.

Classifications (4) ①

Credit Card Number Person's Name Franmer Product
 Numero De Carte De Credit

Collection path

...
 Contoso Purview Catalog
|
 Purview
|
 Cloud_Data
|
 EMEA
|
 Sales

Selecting Open in Power BI Desktop initiates the download of a Power BI Data Source file (PBIDS) you can open with Power BI Desktop. PBIDS files contain a connection to the data source, so all you need to do is enter credentials upon opening the file and you're ready to start building.

Scan a Power BI tenant

In addition to using Purview to find trusted data sources in the data estate to build reports, you can also scan your Power BI tenant to manage and catalog assets. The metadata of Power BI assets, and information about their lineage across Power BI workspaces and their connections to data sources, are then available in Microsoft Purview.

Note

See [Connect to and manage a Power BI tenant in Microsoft Purview](#) for more details.

Next unit: Integrate with Azure Synapse Analytics

[Continue >](#)

How are we doing?

✓ 100 XP



Integrate with Azure Synapse Analytics

4 minutes

Microsoft Purview can be integrated directly into Azure Synapse. If Azure Synapse Studio is massively deployed in your organization, you can get the data catalog experience directly in Azure Synapse Studio.

This integrated experience allows you to discover Microsoft Purview assets, interact with them through Synapse capabilities, and push lineage information to Microsoft Purview.

ⓘ Note

To connect an Microsoft Purview Account to a Synapse workspace, you need 2 types of permissions. You need a contributor role in Synapse workspace from Azure portal identity and access management (IAM). You also need access to that Microsoft Purview Account. For more information, see [Microsoft Purview permissions](#).

Let's imagine you need to find and understand some assets before working with them in pipelines or notebooks. From Azure Synapse Studio, you can easily query your Microsoft Purview data catalog.

In Azure Synapse Studio, from the **Data** blade on the left, select **Purview** in the dropdown next to the search bar.

The screenshot shows the Microsoft Azure Synapse Studio interface. On the left, there is a sidebar with icons for Home, Data (which is selected and highlighted with a red box), Develop, Integrate, Monitor, and Manage. The main area has a header with 'Microsoft Azure' and 'azuresynapse'. Below the header, there is a message about optional cookies and a 'Learn more' link. To the right of the message is a search bar with the placeholder 'Search your organization data'. Above the search bar is a dropdown menu with 'Purview' selected (highlighted with a red box and a red arrow pointing to it). The main content area is titled 'Data' and shows two tabs: 'Workspace' (selected) and 'Linked'. Under 'Workspace', there is a 'Filter resources by name' input field and a list of databases: 'Lake database' (1 item) and 'SQL database' (2 items). To the right of the main content area, there is a sidebar titled 'Recently accessed' which lists files: 'Movies.csv', 'AllMovies', 'ActorsClean.csv', and 'Actors.csv'. At the bottom of the sidebar is a 'View search results' link.

Search for the asset that exists in Purview. Imagine you're looking for movie files. Enter the keyword **movie** in the search bar, and fine tune your search by selecting **Files** as the object

type and Raw as the collection.

The screenshot shows the Microsoft Azure Synapse Studio interface. In the top navigation bar, the workspace is set to 'azuresynapse'. The search bar contains 'Purview movie'. The left sidebar shows 'Data' with 'Workspace' selected. Under 'Object Type', 'Files' is checked. Under 'Collection', 'RAW' is checked. The main pane displays search results for 'movie', showing three items: 'Movies.csv' (Azure Blob Path), 'MovieActors.csv' (Azure Blob Path), and 'OnlineMovieMappings.csv' (Azure Blob Path). A red box highlights the 'Purview' search bar and the 'RAW' collection filter.

Select the first asset "Movies.csv" to get asset details. Because you are in Azure Synapse Studio, you can also leverage Azure Synapse capabilities.

The screenshot shows the Microsoft Azure Synapse Studio interface, focusing on the asset details for 'Movies.csv'. The top navigation bar shows 'azuresynapse'. The left sidebar shows 'Data' with 'Workspace' selected. The main pane shows the asset details for 'Movies.csv', which is located in 'Azure Blob Storage | Blob'. On the right, there is a 'Develop' dropdown menu with options: 'New SQL script', 'New notebook', and 'New data flow'. A red box highlights this 'Develop' menu. Below the asset details, sections for 'Asset description', 'Classifications', 'Schema classifications', and 'Fully qualified name' are visible.

For instance, you can use Azure Synapse serverless to query your assets. Select **Develop**, **New SQL Script** and **Select top 100**.

The screenshot shows the Microsoft Azure Synapse Analytics workspace. On the left, there's a sidebar with icons for Home, Data, Workbooks, Pipelines, Triggers, and Datasets. The main area is titled 'Purview search' and shows results for 'movie'. A specific asset, 'Movies.csv', is selected. The details pane shows it's an 'Azure Blob Storage | Blob'. Below this are tabs for Overview, Properties, Schema, Lineage, Contacts, and Related. A context menu is open over the 'Develop' tab, with options like 'New SQL script', 'New notebook', and 'New data flow'. The 'Overview' tab is currently selected.

Double check you're connected to your serveless instance and select **Run** to execute the script and get an overview of your data.

This screenshot shows the same workspace after running the SQL script. The 'Run' button is highlighted with a red box. The 'Connect to' dropdown is also highlighted with a red box, showing 'Built-in' and 'Use database master'. The results pane displays the output of the SQL query, which is a table of movie data. The entire results table is highlighted with a red box.

MovieID	MovieTitle	Category	Rating	RunTimeMin	ReleaseDate
2f49d5af-b5c1-4b87-b426-0028...	Card Spark	Romance	R	103	02-13-2018
6260de1c-81bc-48b8-be49-007d...	Cube Lake	Science Fiction	PG-13	186	03-13-2018
3f0e40a9-330f-40a9-968d-008a6...	Brick Destiny	Science Fiction	R	182	02-13-2018
de31d122-4694-4c89-8e91-00b8...	Magical Dogs	Comedy	R	128	12-05-2017
fce022f4-135c-4c3c-8d9b-00d4d...	Western Data	Family	PG	97	11-28-2017
ca1eb6ca-9d55-41c9-a9ff-0100a...	Data Destiny	Science Fiction	R	167	08-08-2017
672b5a1c-f5b3-45a0-9919-011b...	Space Case	Romance	PG	94	08-15-2017

After reviewing data, you can use the asset, for example, adding to a new dataflow in Azure Synapse.

The screenshot shows two windows side-by-side. On the left is the Microsoft Purview search interface, displaying a search result for 'Movies.csv' from 'Azure Blob Storage | Blob'. On the right is the 'Integrate' pipeline editor. In the pipeline editor, there is a 'Movies' dataset node. A red arrow points from the 'New data flow' button in the 'Develop' menu of the Purview search interface to the 'New data flow' button in the pipeline editor's context menu.

(!) Note

See [Connect an Microsoft Purview Account](#) for detailed information about integrating Microsoft Purview into Azure Synapse Analytics.

Next unit: Knowledge check

[Continue >](#)

How are we doing? ☆ ☆ ☆ ☆ ☆

✓ 200 XP



Knowledge check

3 minutes

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

Check your knowledge

1. What feature of Microsoft Purview can analysts and other data consumers use to find trustworthy data for reports? *

Data map.

Data catalog.

✓ Correct. Microsoft Purview data catalog helps users find trusted data sources by browsing and searching data assets across a data estate.

Data policies.

✗ Incorrect. Data access policies in Microsoft Purview enable you to manage access to different data systems.

2. If an analyst is looking for a specific asset by name and type, what is the most efficient way to find that asset in the data catalog? *

Browse assets.

Manage glossary.

✗ Incorrect. The business glossary is a definition of terms specific to a domain of knowledge that is commonly used, communicated, and shared in organizations as they are conducting business.

Search catalog.

✓ Correct. Searching the data catalog can help users quickly find assets they're looking for.

3. How can users download Power BI data source files that contain connections to assets discovered in Microsoft Purview? *



In the Microsoft Purview data catalog, in the asset view.

✓ Correct. There's a button in the asset view to open the asset in Power BI desktop.



In the Microsoft Purview insights report.



Users can't download Power BI data source files from Microsoft Purview.

Next unit: Summary

[Continue >](#)

How are we doing?

100 XP

Introduction

3 minutes

The Microsoft Purview Data Catalog offers a browse experience that enables users to explore available data. Users can explore the data catalog either by collection or through traversing the hierarchy of each data source. The first step in understanding the contents of your data map is registering and scanning data, after which you can classify data for easy identification of assets to use for reporting.

Learning objectives

In this module, you will:

- Describe asset classification in Microsoft Purview.

Next unit: Register and scan data

[Continue >](#)

How are we doing?

100 XP

Register and scan data

10 minutes

Registration and scanning of data enables discoverability of data across an estate.

Before you can register and scan data, it's important to understand the concept of collections. In Microsoft Purview Data Catalog, collections are key concept because they drive permissions and asset protection. Collections are also used to understand data estate health and catalog usage and adoption, as featured in the data stewardship section of your [Data Estate Insights](#).

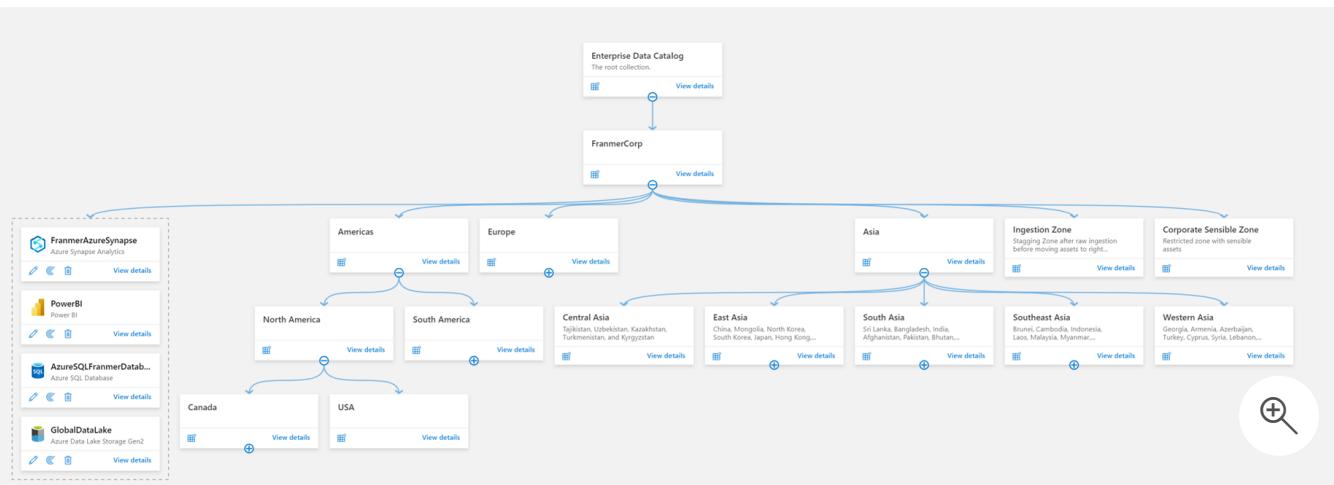
Collections

The data map is at the core of Microsoft Purview, which keeps an up-to-date map of assets and their metadata across your data estate. To hydrate the data map, you need to register and scan your data sources, which is done at the collection level. Collections support organizational mapping of metadata. By using collections, you can manage and maintain data sources, scans, and assets in a hierarchy instead of a flat structure. Collections allow you to build a custom hierarchical model of your data landscape based on how your organization plans to use Microsoft Purview to govern your landscape.

Collections also provide a security boundary for your metadata in the data map. Access to collections, data sources, and metadata is set up and maintained based on the collection's hierarchy in Microsoft Purview, following a least-privilege model:

- Users have the minimum amount of access they need to do their jobs.
- Users don't have access to sensitive data that they don't need.

Data sources are registered at the collection level. Scan results can then be sent to this collection or a sub collection. The image below displays the structure of a collection.



Tip

Learn more about [Microsoft Purview collections architectures and best practices](#).

Register and scan data sources

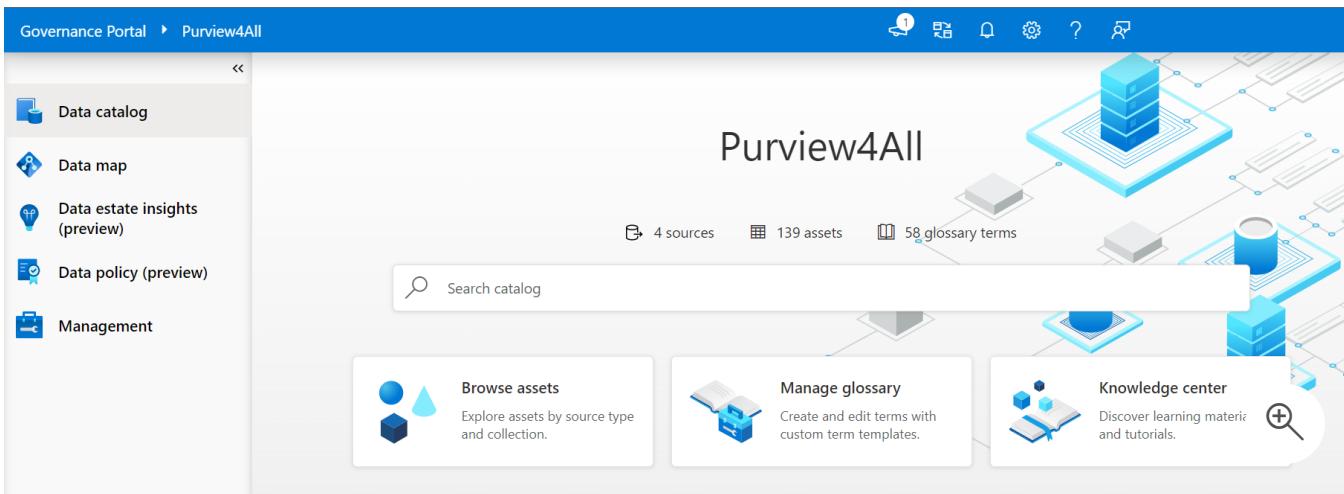
Data governance use begins at collection level, with the registration of data sources in Microsoft Purview governance portal. Microsoft Purview supports an array of data sources. Data teams (analysts, engineers, and scientists) may not be actively registering and scanning data in Microsoft Purview, but it's critical that data consumers understand governance efforts. Registering and scanning assets requires **Data Curator** permissions.

Important

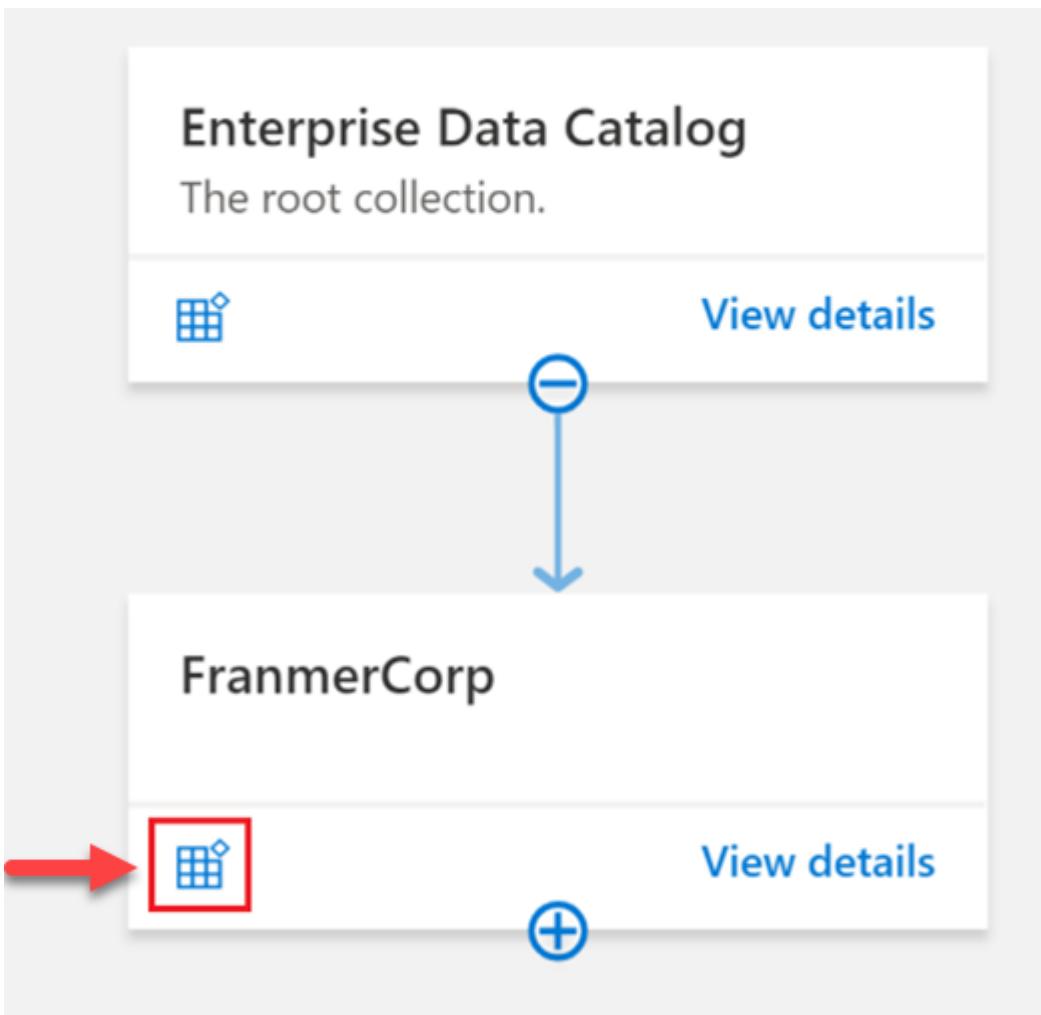
Data registered and scanned in Microsoft Purview only collects metadata information. Data remains in its location and isn't migrated to any other platform.

Register a data source

Registering a data source is done from within the Azure portal. Once you have a Microsoft Purview service configured in Azure, you use the Microsoft Purview governance portal to register your data sources.



To register a data source, you'll select the icon to register a data source as displayed in the image below. Selecting this icon will give you access to all data source connectors.



Below is a small sample of available connectors in Microsoft Purview Data Catalog. See [supported data sources and file types](#) for an up-to-date list of supported data sources and connectors.

The screenshot shows the Microsoft Purview Enterprise Data Catalog interface. On the left, there's a tree view of collections: 'Enterprise Data Catalog' (root) and 'FranmerCorp'. 'FranmerCorp' has a 'View details' button and a plus sign icon indicating it can be expanded. On the right, a grid of icons represents registered data sources under the 'Azure Database' tab. The sources include: Azure Synapse Analytics (multiple), Azure Blob Storage; Azure Cosmos DB (SQL API), Azure Data Explorer (Kusto), Azure Data Lake Storage Gen1; Azure Data Lake Storage Gen2, Azure Database for MySQL, Azure Database for PostgreSQL; Azure Dedicated SQL Pool (formerly SQL DW), Azure Files, Azure SQL Database; Azure SQL Managed Instance, and SQL Server on Azure Arc-enabled servers.

Registering a data source is straightforward, you need to complete the required fields. Authentication will be done during the scanning phase.

Each type of data source you choose will require specific information to complete the registration. For example, if your data sources reside in your Azure subscription, you'll choose the necessary subscription and storage account name.

Scan a data source

Once you have data sources registered in the Microsoft Purview governance portal and displayed in the data map, you can set up scanning. The scanning process can be triggered to run immediately or can be scheduled to run on a periodic basis to keep your Microsoft Purview account up to date.

Scanning assets is as simple as selecting **New scan** from the resource as displayed in the data map.

The screenshot shows the Azure Data Lake Storage Gen2 blade. At the top, there's a section for a scan named "GlobalDataLake" with the sub-type "Azure Data Lake Storage Gen2". Below this, there are three icons: a pencil for edit, a refresh symbol with a red border for view details, and a trash can. To the right of the icons is the text "View details".

You'll now need to configure your scan and assign the following details:

- Assign a friendly name.
- Define which [integration runtime](#) to use to perform the scan.
- [Create credentials](#) to authenticate to your registered data sources.
- Choose a collection to send scan results.

After the basic configuration, you'll *scope* your scan, which allows you to choose just a specific zone of your data source. For instance, if you have a collection called "Raw" in your data map, you can define the scope to scan only the raw container of your data lake.

After configuring and scoping your scan, you'll define the *scan rule set*. A scan rule set is a container for grouping a set of scan rules together so that you can easily associate them with a scan. For example, you might create a default scan rule set for each of your data source types, and then use these scan rule sets by default for all scans within your company. You might also want users with the right permissions to create other scan rule sets with different configurations based on business need.

Once a scan is complete, you can refer to the scan details to view information about the number of scans completed, assets detected, assets classified, Scan information. It's a good place to monitor scan progress, including success or failure.

The screenshot shows the Azure Purview Data Map blade. On the left, there's a navigation bar with "GlobalDataLake" and "Azure Data Lake Storage Gen2". Below it are buttons for "New scan", "Edit source", "Delete source", and "Refresh". The main area has tabs for "Overview" (which is selected) and "Scans". Under "Overview", there are summary cards for "Scans" (4), "Discovered assets" (108), and "Classified assets" (32). On the right, there are sections for "Source ID", "Collection path", "Source hierarchy", and a "Recent scans" table.

Scan name	Last run status	Scan rule set	Last scan time
Scan-Raw-NonSensibleZone	Completed	FranmerADLSGen2Rules	06/09/2022, 8:10 A...
Scan-IngestionZone	Completed	FranmerADLSGen2Rules	06/07/2022, 8:04 PM
Scan-Raw-SensibleZone	Completed	FranmerADLSGen2Rules	06/07/2022, 9:46 A...

[→ See all applied scans](#)



Tip

Refer to [Scanning best practices](#) for more information on scanning assets.

Roles and permissions

Permissions in Microsoft Purview are assigned at **collection** level. Collections are used to organize assets and sources and can be thought of as a logical grouping of data assets.

Data teams looking to discover and use data need to be assigned the **Data Reader** role in a collection in Microsoft Purview. The Data Reader role enables users to find assets, but doesn't enable users to edit anything. The **Data Curator** role is required to edit information about assets, assign classifications, and associate assets with glossary entries. To set up scans via the Microsoft Purview Governance Portal, individuals need to be either a data curator on the collection or data curator and data source administrator where the source is registered.

When a Microsoft Purview account is created, it starts with a root collection that has the same name as the Microsoft Purview account itself. The creator of the Microsoft Purview account is automatically added as a Collection Admin, who can then assign Data Source Admin, Data Curator, and Data Reader on this root collection, and can edit and manage this collection.



Tip

Learn more about [Microsoft Purview permissions and access](#).

Next unit: Classify and label data

[Continue >](#)

How are we doing?

✓ 100 XP



Classify and label data

6 minutes

Glossary terms, classifications and labels are all annotations to a data asset. Each of them have a different meaning in the context of the data catalog.

What is data classification?

Classifications are annotations that can be assigned to entities. The flexibility of classifications enables you to use them for multiple scenarios such as:

- understanding the nature of data stored in the data assets
- defining access control policies

Classification is based on the business context of the data. For example, you might classify assets by Passport Number, Driver's License Number, Credit Card Number, SWIFT Code, Person's Name, and so on. Microsoft Purview has more than 200 system classifiers today. Users can also define their own classifiers in the data catalog. As part of the scanning process, classifications are automatically detected and applied as metadata within the Purview Data Catalog.

Classification rules

In Microsoft Purview, you can apply system or custom classifications on a file, table, or column asset. Microsoft Purview makes use of Regex patterns and bloom filters to classify data. These classifications are then associated with the metadata discovered in the Azure Purview Data Catalog.

Metadata is used to help describe the data that is being scanned and made available in the catalog. During the configuration of a scan set, you can specify classification rules to apply during the scan that will also serve as metadata. The existing classification rules fall under five major categories:

- Government - covers attributes such as government identity cards, driver license numbers, passport numbers, etc.
- Financial - covers attributes such as bank account numbers or credit card numbers.
- Personal - personal information such as a person's age, date of birth, email address, phone number, etc.

- Security - attributes like passwords that may be stored.
- Miscellaneous - attributes not covered in the other categories.

Why classify data?

A good data governance strategy includes a process to classify data to understand its level of confidentiality, determine if the data source is compliant with various regulations, or how long to retain it for. Classification in Microsoft Purview makes data assets easier to understand, search, and govern. Classification can also help you implement measures to protect sensitive data.

Once a classification is tagged to a data source after a scan, you can generate reports and insights to gain a stronger understanding of your data estate. Because classification is based on the business context of the data, it can help bridge the gap between the business and the data team.

Data classification: system vs. custom classification

Microsoft Purview supports both system and custom classifications. There are over +200 system classifications available in Microsoft Purview today. Data teams need to know that if necessary classifications aren't available out of the box, they can work with the data stewards to create custom classifications, to meet their own organizational data governance requirements.

Important

For the entire list of available system classifications, see [Supported classifications in Microsoft Purview](#).

Who creates custom classifications?

Purview Data Curators can create, update, and delete custom classifiers and classification rules. Purview Data Readers can only view classifiers and classification rules.

In practical terms, Data Curators may not be members of the data team. It is however critical that data team members understand classification to be able to successfully work together and govern data across an organization.

What are data labels?

The Microsoft Purview Data Map supports labeling structured and unstructured data stored across various data sources. This may sound familiar to you from other Microsoft technologies - and may be known as sensitivity labels. The data map extends the use of sensitivity labels from Microsoft Purview Information Protection to assets stored in infrastructure cloud locations and structured data sources.

Labels are defined in Microsoft Purview Information Protection, and you can extend the application to Microsoft Purview Data Catalog.

The screenshot below shows both data classification and label in the Microsoft Purview Data Catalog. You can see that this Azure SQL table has a column called "CreditCard":

- Classified as "Credit Card Number" because scan detected numbers corresponding to credit card pattern rules.
- Labeled as "Confidential – Finance" because credit card number was defined in your organization as confidential information (and this label brings encryption).

Column name	Classifications	Sensitivity label
Franmer_Card		
CardNumber	<button>Credit Card Number</button>	<button>Confidential - Finance</button>
CardType		
CardVerification		
Country	<button>Country/Region</button>	
CountryName	<button>Country/Region</button>	
ExpiryMonth		

Next unit: Search the data catalog

100 XP



Search the data catalog

3 minutes

A data catalog search can empower business and data analysts to find and interpret data. The data catalog provides intelligent recommendations based on data relationships, business context, and search history. The Purview data catalog can assist data teams by adding business context to assets to drive analytics, AI and ML initiatives.

The data catalog can be searched by keyword, object type, collection, classification, contact, label, or assigned term. Results can then be sorted by relevance or name.

The screenshot shows the 'Search results for Bing' page in the Data Catalog. At the top, there are filters for 'Source type : all', 'Instance : all', and a 'Clear all filters' button. Below the filters, a red arrow points to the 'Filter by keyword' input field. Another red arrow points to the 'Object Type' section, which lists various data types like Dashboards, Data pipelines, Files, Folders, Glossary terms, Reports, Stored procedures, and Tables. A third red arrow points to the 'Collection' section, showing a single entry for 'Contoso Purview4All'. A fourth red arrow points to the 'Classification' section, listing 'Country/Region' (2 results) and 'World Cities' (1 result). A fifth red arrow points to the 'Contact' section, listing 'Paul' (3 results). A sixth red arrow points to the 'Label' section, listing 'Confidential' (2 results). A seventh red arrow points to the 'Assigned term' section, listing 'Contoso Child' (1 result). On the right side, the results are displayed with three items: 'QueriesByCountry', 'QueriesByState', and 'BingCoronavirusQuerySet'. Each item includes a preview icon, the dataset name, its type (Azure Data Lake Storage Gen2 Resource Set or Path), a brief description, a URL, and an 'Updated' timestamp. The results are sorted by 'Relevance'.

Object Type	Collection	Classification	Contact	Label	Assigned term
Dashboards	Contoso Purview4All	Country/Region	Paul	Confidential	Contoso Child
Data pipelines		World Cities			
Files					
Folders					
Glossary terms					
Reports					
Stored procedures					
Tables					

Showing 1-3 out of 3 results

Sort by: Relevance

QueriesByCountry
Azure Data Lake Storage Gen2 Resource Set
his dataset was curated from the Bing search logs (desktop users only) over the period of J...
https:// Updated 3 months ago

QueriesByState
Azure Data Lake Storage Gen2 Resource Set
This dataset was curated from the Bing search logs (desktop users only) over the period of ...
https:// Updated 3 months ago

BingCoronavirusQuerySet
Azure Data Lake Storage Gen2 Path
https:// Updated 21 days ago

✓ 200 XP ➔

Knowledge check

3 minutes

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

Check your knowledge

1. What level are user permissions set at in Microsoft Purview? *

Tenant.

✗ Incorrect. Permissions aren't set at tenant level in Microsoft Purview.

Data catalog.

Collection.

✓ Correct. User permissions are set at the collection level in Microsoft Purview.

2. What are the two types of classification in Microsoft Purview? *

System classifications and custom classifications.

✓ Correct. Microsoft Purview contains out of the box, system classifications and custom classifications.

Microsoft Information Protection Sensitivity Labels and system classifications.

✗ Incorrect. Microsoft Information Protection Sensitivity Labels can be applied in Microsoft Purview, but they aren't a type of classification.

Custom classifications and user-defined classifications.

3. If a data analyst is looking for a specific resource for reporting, what should they use? *

Purview Data Catalog to search.

✓ Correct. Searching the Purview Data Catalog will help analysts find resources.

The business glossary.

-
- Import into Power BI and create a custom report.
-

Next unit: Summary

[Continue >](#)

How are we doing?     

100 XP

Introduction

3 minutes

As the landscape of enterprise data continues to grow, it's critical to get an accurate view of your organization's data. Microsoft Purview and Power BI integration enables you to scan your entire Power BI tenant to search and browse Power BI assets, explore enhanced dataset metadata, trace end-to-end data lineage, and drill-down into datasets in Power BI for further analysis.

Learning objectives

In this module, you will:

- Register and scan a Power BI tenant.
- Use the search and browse functions to find data assets.
- Describe the schema details and data lineage tracing of Power BI data assets.

Next unit: Register and scan a Power BI tenant

[Continue >](#)

How are we doing?

✓ 100 XP



Register and scan a Power BI tenant

4 minutes

To get an understanding of what is going on in your Power BI tenant, you can perform a full scan in Microsoft Purview to view the schema and lineage of assets across all workspaces. After, you can schedule incremental scans on workspaces that have changed since the previous scan.

There are a few pre-requisite steps required to scan your Power BI tenant in Microsoft Purview.

💡 Tip

If you need to create a Microsoft Purview account, see the [quickstart guide](#) to create a Microsoft Purview account in the Azure Portal.

Establish a connection between Microsoft Purview and Power BI

Microsoft Purview can connect to and scan Power BI either in the same tenant or across tenants. You'll need to set up authentication either by using a Managed Identity or a Delegated Authentication.

❗ Note

See [Register and scan a Power BI tenant](#) to learn more about the set-up and authentication of Power BI connections in same and cross-tenant scenarios.

Authenticate to Power BI tenant

Give Microsoft Purview permissions to access your Power BI tenant.

If you're using **Managed Identity** to authenticate to Power BI, you'll need to create a security group in Azure Active Directory, and add your Microsoft Purview managed identity to this security group.

New Group

 Got feedback?

Group type * ⓘ

Security



Group name * ⓘ

Purview



Group description ⓘ

Enter a description for the group

Azure AD roles can be assigned to the group ⓘ

Yes

No

Membership type * ⓘ

Assigned



Owners

1 owner selected

Members

1 member selected

If a security group containing the Purview managed identity already exists, you can proceed to configuring the Power BI tenant.

Configure Power BI tenant

Next you need to enable access to Power BI by Microsoft Purview in Power BI itself. This is done by enabling *Allow service principals to use read-only Power BI admin APIs* in the Power BI admin portal.

Admin API settings

Allow service principals to use read-only Power BI admin APIs

Unapplied changes

Web apps registered in Azure Active Directory (Azure AD) will use an assigned service principal to access read-only Power BI Admin APIs without a signed in user. To allow an app to use service principal authentication, its service principal must be included in an allowed security group. By including the service principal in the allowed security group, you're giving the service principal read-only access to all the information available through Power BI admin APIs (current and future). For example, Power BI user names and emails, dataset and report detailed metadata. [Learn more](#)

Enabled

Apply to:

The entire organization

Specific security groups

Purview Enter security groups

✓ Applying changes
Tenant settings changes will be applied within the next 15 minutes.

Register and scan Power BI

Now that you've got access set up in both Microsoft Purview and Power BI, you can register and scan your Power BI tenant.

The screenshot shows the Microsoft Purview Data Catalog interface. On the left, there's a sidebar with options like Data catalog, Data map, Data estate insights (preview), Data policy (preview), and Management. The main area is titled 'Sources' and shows a list of collections and sources. A red box highlights the 'Register' button, which is located next to 'Sources' and 'Refresh'. Below it is a 'Filter by keyword' input field. A red arrow points from this main interface to a separate 'Register sources' dialog box on the right. This dialog box has a search bar at the top with 'Power BI' typed in. Below the search bar are tabs for All, Azure, Database, File, and Services and apps. Under the All tab, there's a chart showing three bars labeled 'Power BI'.

After registering the Power BI tenant, initiate the scan by selecting **New scan**. Give your scan a name and step through the interface, where you'll be able to exclude personal workspaces, confirm integration runtime and credentials, and select a collection. Test the connection to ensure authentication is set up properly.

Scan "PowerBI-Nod"

Name *
Scan-1

Personal workspaces *
 Include Exclude

Changes to the scan configuration will reset the upcoming scan to be a full scan for this data source. [Learn more](#)

Connect via integration runtime * ⓘ
Azure AutoResolveIntegrationRuntime

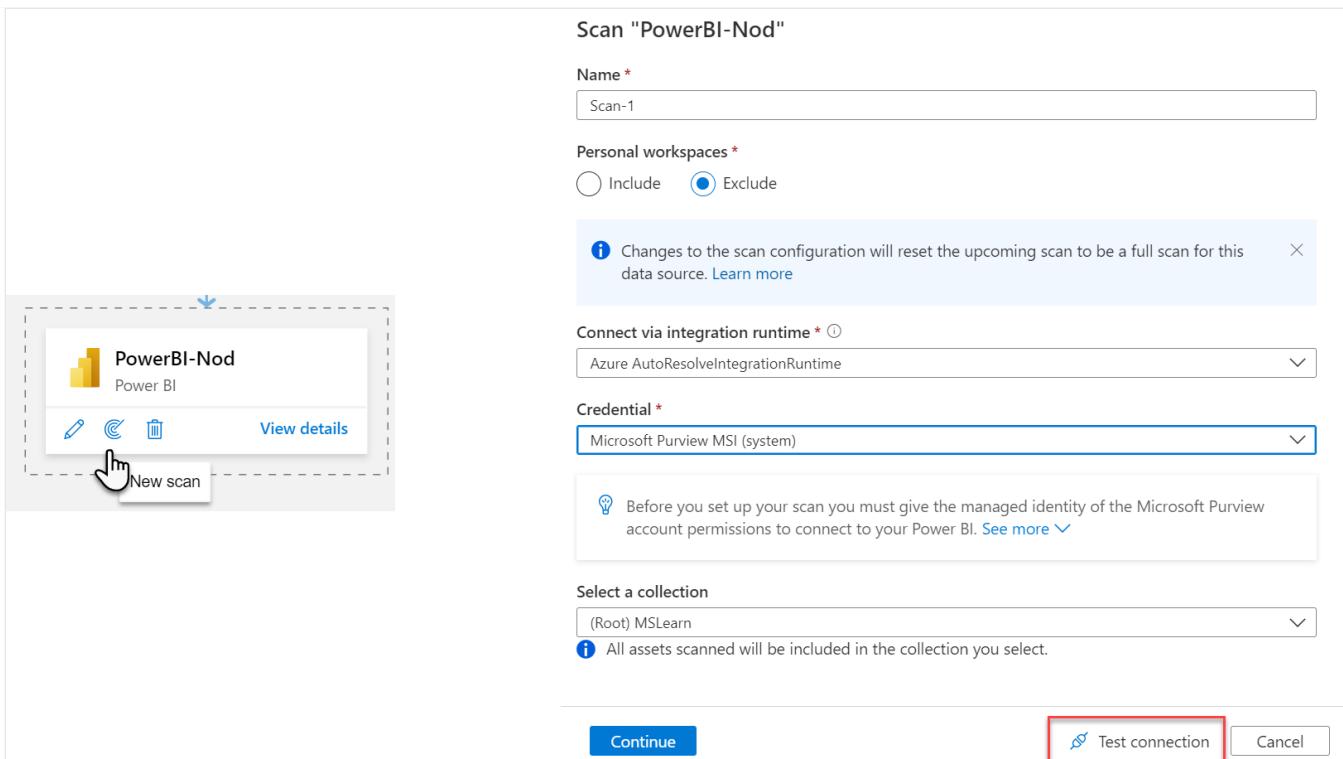
Credential *
Microsoft Purview MSI (system)

Before you set up your scan you must give the managed identity of the Microsoft Purview account permissions to connect to your Power BI. [See more](#)

Select a collection
(Root) MSLearn

All assets scanned will be included in the collection you select.

Continue **Test connection** **Cancel**



! Note

If you're performing the scan, you must be both a Data Source Administrator and a Data Reader. See **Access control in the Microsoft Purview Data Map** for details on permissions.

You're able to track the progress of the scan in the data map, and once the scan is complete, you'll be able to search and browse the contents of your entire Power BI tenant!

PowerBI-4Learn

 Power BI

 New scan  Edit source  Delete source  Refresh

[Overview](#) [Scans](#)

Source ID: <https://app.powerbi.com/>

Scans

1



Discovered assets 

1



Classified assets 

0



Recent scans

Scan name	Last run status	Scan rule set
Scan-C4B	 In progress	-
→ See all applied scans		

If you're having any issues with scanning your Power BI tenant, see [Troubleshoot Power BI tenant scans in Microsoft Purview](#) for details and helpful hints.

Next unit: Search and browse Power BI assets

[Continue >](#)

How are we doing?     

✓ 100 XP ➔

Search and browse Power BI assets

4 minutes

After data is registered and scanned, analysts and data consumers need to be able to find data, view enhanced metadata, and track data lineage. Search and browse in the Purview Data Catalog enables you to quickly find trustworthy data.

After scanning your Power BI tenant, you'll see those assets appear in the search results, including underlying data sources.

Search the Microsoft Purview Data Catalog

From the Microsoft Purview Governance Portal, you can type relevant keywords to start discovering assets. In this scenario, you're looking for "sales."

The screenshot shows the Microsoft Purview Data Catalog interface with the title "Purview4All" at the top. Below the title, there are three statistics: "6 sources", "139 assets", and "58 glossary terms". A search bar contains the keyword "sales". Below the search bar, under "Your recent searches", is a link to "sales". Under "Search suggestions", there are five items: "SalesLT SalesOrderDetail", "SalesLT vGetAllCategories", "SalesLT ProductModelProductDescription", and "SalesPerson".

The screenshot below displays the search result, with all assets corresponding to the keywords entered in the search engine. Notice the appearance of Power BI assets.

The screenshot shows the Microsoft Purview Data Catalog interface. On the left, there's a sidebar with filters for 'Source type : all', 'Instance : all', and a 'Clear all filters' button. Below these are sections for 'Object Type' (Dashboards, Data pipelines, Files, Folders, Reports, Stored procedures, Tables), 'Collection' (Sales, Contoso Purview Catalog, IT), and 'Classification' (Franmer Product, Credit Card Number, Country/Region, Person's Name). The main area displays search results for 'product sales'. It shows 1-25 out of 153 results, sorted by Relevance. The results include:

- Product Mix.csv** (Azure Blob Path)
- Sales Dashboard-SQLDB** (Secret, Power BI Dataset)
- SalesWithCustomerID.csv** (Azure Blob Path)
- SalesWithCustomerID** (certified, Azure SQL Table)
Table with sales raw information.
- ReportTable_Prod** (certified, Azure SQL Table)
Report Table in production environment to create Power BI report for sales people. Ta...

You can fine-tune your search using the filters on the left side of the page. Filters available include source type, keyword, object type, collection, classification, contact, label, and glossary term.

Search results for sales

Source type : all Instance : all

Filter by keyword

Object Type

- Dashboards
- Data pipelines
- Files
- Folders
- Glossary terms
- Reports
- Stored procedures
- Tables

Collection

- Finance Purview4All (3)
- Purview4All (3)

Classification

Contact

Label

Assigned term

Browse the Microsoft Purview Data Catalog

Searching for specific assets is great if you know what you're looking for, but analysts and data consumers may not know exactly how their data estate is structured. The browse experience enables you to explore what data is available, either by collection or through traversing the hierarchy of each data source in the catalog.

To access the browse experience, select **Browse assets** from the governance portal home page.

Purview4All

5 sources 139 assets 58 glossary terms

Search catalog

Browse assets
Explore assets by source type and collection.

Manage glossary
Create and edit terms with custom term templates.

Knowledge center
Discover learning materials and tutorials.

You can browse the data catalog either by collection or by source type, depending on your needs. Browsing by either collection or source type allows you to see assets you have access to. Once you find the asset you're looking for, you can select it to see details on schema, lineage, and a detailed classification list.

Browse assets

Refresh

By collection By source type

Filter by keyword

Showing 5 collections

Name	Description	Assets
Purview4All	The root collection.	12
Contoso		14
Finance	Logical collection to store Finance-related data	18
HR		67

Browse assets

Refresh

By collection By source type

Azure Data Factory 1 item	Azure Data Lake Storage Gen2 4 items	Azure SQL Database 3 items	Azure SQL Server 2 items	Azure Storage Account 5 items
------------------------------	---	-------------------------------	-----------------------------	----------------------------------

Custom source types

File 1 item	Metadata
----------------	----------

Uniquely, browsing by source type allows you to see the hierarchies of data sources using an explorer view. This is a helpful and familiar way to navigate to see lists of scanned assets.

! Note

Assets in Purview are organized by collection and permissions are granted at collection level. Both searching and browsing require data reader permissions. See [Access control in the Microsoft Purview Data Map](#) for details on permissions.

Select an asset to see details about the properties, schema, lineage, contacts, and related assets.

Data catalog > Browse assets >

 Customer
Azure SQL Table

[Edit](#) [Select for bulk edit](#) [Request access](#) [Refresh](#) [Delete](#)

[Overview](#) [Properties](#) [Schema](#) [Lineage](#) [Contacts](#) [Related](#)

Asset description
No description for this asset.

Classifications ⓘ
No classifications for this asset.

Schema classifications (3) ⓘ
[Email Address](#) [Person's Name](#) [U.S. Phone Number](#)

Fully qualified name ⓘ
mssql://

Collection path
 Purview4All
 Finance

Hierarchy
 pvlab
Azure SQL Server
 pvlab
Azure SQL Database
 SalesLT
Azure SQL Schema
 Customer
Azure SQL Table

Glossary terms
No glossary terms for this asset.

Next unit: View Power BI metadata and lineage

[Continue >](#)

How are we doing?     

✓ 100 XP



View Power BI metadata and lineage

4 minutes

Purview and Power BI together are powerful, enhancing the ability of the search and browse features to see both the schema and lineage of Power BI assets.

Extend your search with enhanced metadata

Metadata scanning facilitates governance by making it possible to catalog and report on the metadata of your organization's Power BI artifacts. The results of metadata scanning are displayed on the schema tab of the asset.

ⓘ Note

Metadata scanning must be enabled in the Power BI admin portal. See [Set up metadata scanning in your organization](#) to learn more.

After performing a search in the Purview Governance Portal, select a Power BI asset from your search result to see the sensitivity labels and endorsement metadata. Additional business metadata includes the dataset user configuration, create datetime, and description.

Under the Schema tab, you can see the list of all the tables, columns, and measures created inside the Power BI dataset.

Opportunity Report (PII-Personal Information) Certified

1 of 10 ↑ ↓

Edit Select for bulk edit Request access Refresh Delete Edit columns Open in Power BI

Overview Properties Schema Lineage Contacts Related Updated on June 9, 2022 9:14 PM by MOD Admin Administrator

Filter by name

Showing 125 of 125 items

Field name	Classifications	Sensitivity label	Glossary terms	Data type	Asset description
CountryRegionBudget			Sales without taxes	table	Budget per country without taxes
Actual Revenue			Sales without taxes	Double	Actual revenue
Actual Revenue MoM%					Actual Revenue MoM%
Country	⚡ Country/Region			String	Country name
Growth Year on Year			Sales without taxes	String	Growth Year on Year
Multiplier				Double	Accelerator
Region	⚡ Country/Region			String	Region name
Target Revenue			Sales without taxes	Int64	Target revenue without taxes
Variance				Double	Variance
Variance to Target Sales with Di...					Variance to Target Sales with Disc...
DimAccount				table	Account name
DimAccountOwner	⚡ Person's Name			table	Account owner name
DimDate				table	Date
Discount %				table	Discount in percent
Margin			Sales without taxes, Sales	table	Margin
prmAnalytic				table	Power BI field parameter
Product Mix	Franmer Product			table	Product mix
Actual Mix (\$)				Double	
Product	Franmer Product		Product, Customer product, Produit client	String	
Product Category	Franmer Product		Customer product, Produit client, Product	String	
Target Mix (%)				Int64	

For more detail, selecting a particular field within the schema tab will take you to the details for that field. You can then view the overview, properties, lineage, contacts, and related assets for that particular field.

Power BI Column

Edit Select for bulk edit Refresh Delete

Overview Properties Contacts

Filter by property key Show properties without a value

Showing 6 of 12 properties

Properties

dataType	String
displayName	Variance
isHidden	false
isMeasure	false
qualifiedName	https://app.powerbi.com/

Related assets

Table	dimAccount
-------	----------------------------

Metadata scanning requires no special license. It works for all of your tenant metadata, including items that are located in non-Premium workspaces.

If you'd like more information about assets, you also have the option open the Power BI dataset in the Power BI service for further analytics, root-cause investigation, impact analysis, management tasks, and dataset enrichment.

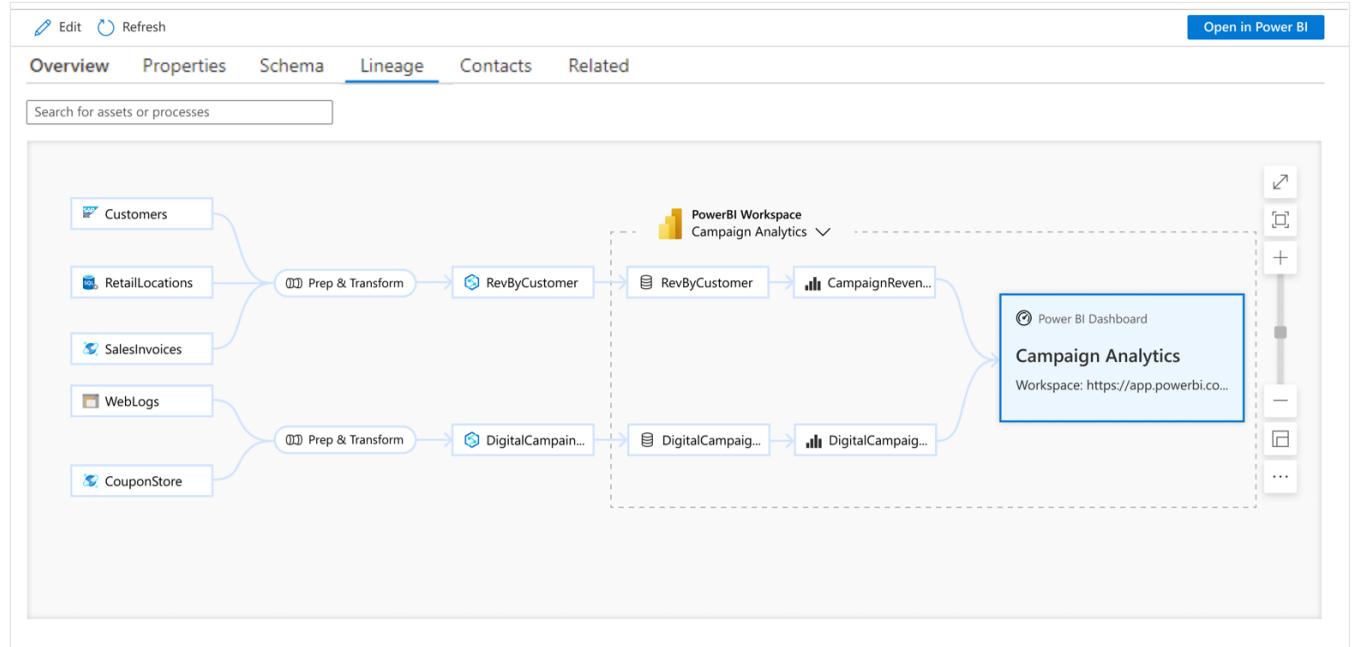
Extend your search with lineage

If you're using the search and browse features in Microsoft Purview to find assets for reporting or to troubleshoot existing assets, you likely need more information on where data actually comes from, and what transformation steps it has undergone. The lineage view displays the flow of data from the source through to Power BI assets, including dataflows, datasets, reports, and dashboards.

Although you can track [data lineage in Power BI](#), this information is limited to the items in a single workspace. Lineage in Purview enables you to view the movement of data across more than one workspace, in a single view.

Lineage enables easy troubleshooting and deeper analysis of analytics projects. You're able to look both up and down-stream, to perform either root cause or impact analysis.

For example, you can detect the Azure Synapse Analytics pipeline that is responsible for the transformation of the data upstream of Power BI.



In the Purview Governance Portal, lineage is displayed from the asset you're currently viewing.

Next unit: Knowledge check

[Continue >](#)

How are we doing? ☆ ☆ ☆ ☆ ☆

✓ 200 XP ➔

Knowledge check

3 minutes

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

Check your knowledge

1. What are the prerequisite steps to register and scan a Power BI tenant in Microsoft Purview? *

- Set up authentication between Purview and Power BI, and configure the Power BI tenant.

✓ **Correct. Authentication must be established using either Managed Identity or Delegated Authentication, and the Power BI tenant must be configured to allow service principles to use the read-only API.**

- Set up and deploy a Power BI data gateway.
 Configure the Power BI tenant only.

2. What steps are required in the Power BI admin portal to enable the scanning and display of enhanced metadata? *

- There are no other steps required. Enhanced metadata displays by default.

✗ **Incorrect. There is a required setting in the Power BI admin portal that needs to be enabled.**

- Enable enhanced metadata scanning in the Azure portal.
 Enable an admin API setting in the Power BI admin portal.

✓ **Correct. Enabling the admin API settings will enable enhanced metadata scanning.**

3. What details of an asset would be helpful in performing an impact analysis? *

- Lineage.

✓ Correct. Viewing lineage details will help users perform an impact analysis.



Properties.

✗ Incorrect. The properties of the asset aren't sufficient to perform an impact analysis.



Schema.

Next unit: Summary

[Continue >](#)

How are we doing?

100 XP

Introduction

1 minute

Microsoft Purview is a cloud service that provides the basis of a *data governance* solution in which you can catalog, classify, and track data assets across a large-scale data estate.

Azure Synapse Analytics is a cloud-scale data analytics suite that supports data ingestion and transformation, distributed big data processing and exploration with SQL and Spark, and enterprise data warehousing.

When combined, Microsoft Purview and Azure Synapse Analytics can be used to create a comprehensive solution for reliable, massively scalable data analytics with rich data asset discovery and lineage tracking capabilities.

In this module you'll learn how to:

- Catalog Azure Synapse Analytics database assets in Microsoft Purview.
- Configure Microsoft Purview integration in Azure Synapse Analytics.
- Search the Microsoft Purview catalog from Synapse Studio.
- Track data lineage in Azure Synapse Analytics pipelines activities.

Next unit: Catalog Azure Synapse Analytics data assets in Microsoft Purview

[Continue >](#)

How are we doing?

100 XP



Catalog Azure Synapse Analytics data assets in Microsoft Purview

8 minutes

Azure Synapse Analytics is a platform for cloud-scale analytics workloads that process data in multiple sources; including:

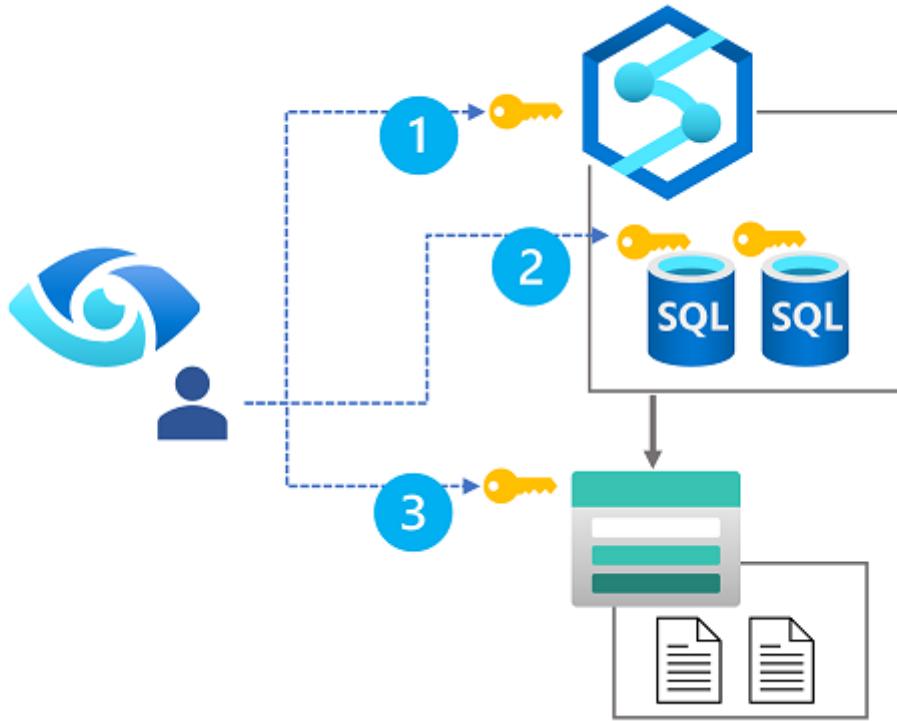
- Relational databases in serverless and dedicated SQL pools
- Files in Azure Data Lake Storage Gen2

A comprehensive data analytics solution can include many folders and files in a data lake, and multiple databases that each contain many tables, each with multiple fields. For a data analyst, finding and understanding the data assets associated with a Synapse Analytics workspace can present a significant challenge before any analysis or reporting can even begin.

Microsoft Purview can help in this scenario by cataloging the data assets in a *data map*, and enabling data stewards to add metadata, categorization, subject matter contact details, and other information that helps data analysts identify and understand data.

Configure data access for Microsoft Purview

In order to scan the data assets in the data lake storage and databases used in your Azure Synapse Workspace, Microsoft Purview must have appropriate permissions to read the data. In practice, this means that the account used by your Microsoft Purview account (usually a system-assigned managed identity that is created when Microsoft Purview is provisioned) needs to be a member of the appropriate role-based access control (RBAC) and database roles.



The diagram shows that Microsoft Purview requires role membership that permits the following access:

1. Read access to the Azure Synapse workspace (achieved through membership of the **Reader** role for the Azure Synapse Workspace resource in the Azure subscription).
2. Read access to each SQL database that will be scanned (achieved through membership of the **db_datareader** fixed database role in each database).
3. Read access to data lake storage (achieved through membership of the **Storage Blob Data Reader** role for the Azure Storage account hosting the Azure Data Lake Storage Gen2 container for the data lake).

💡 Tip

Learn more:

- For more information about RBAC in Microsoft Azure, see [What is Azure role-based access control \(Azure RBAC\)?](#)
- For more information about database-level roles in Azure Synapse Analytics SQL pools, see [Database-level roles](#).

You'll get a chance to assign RBAC and SQL database role membership to support Microsoft Purview data access for yourself in the exercise later in this module.

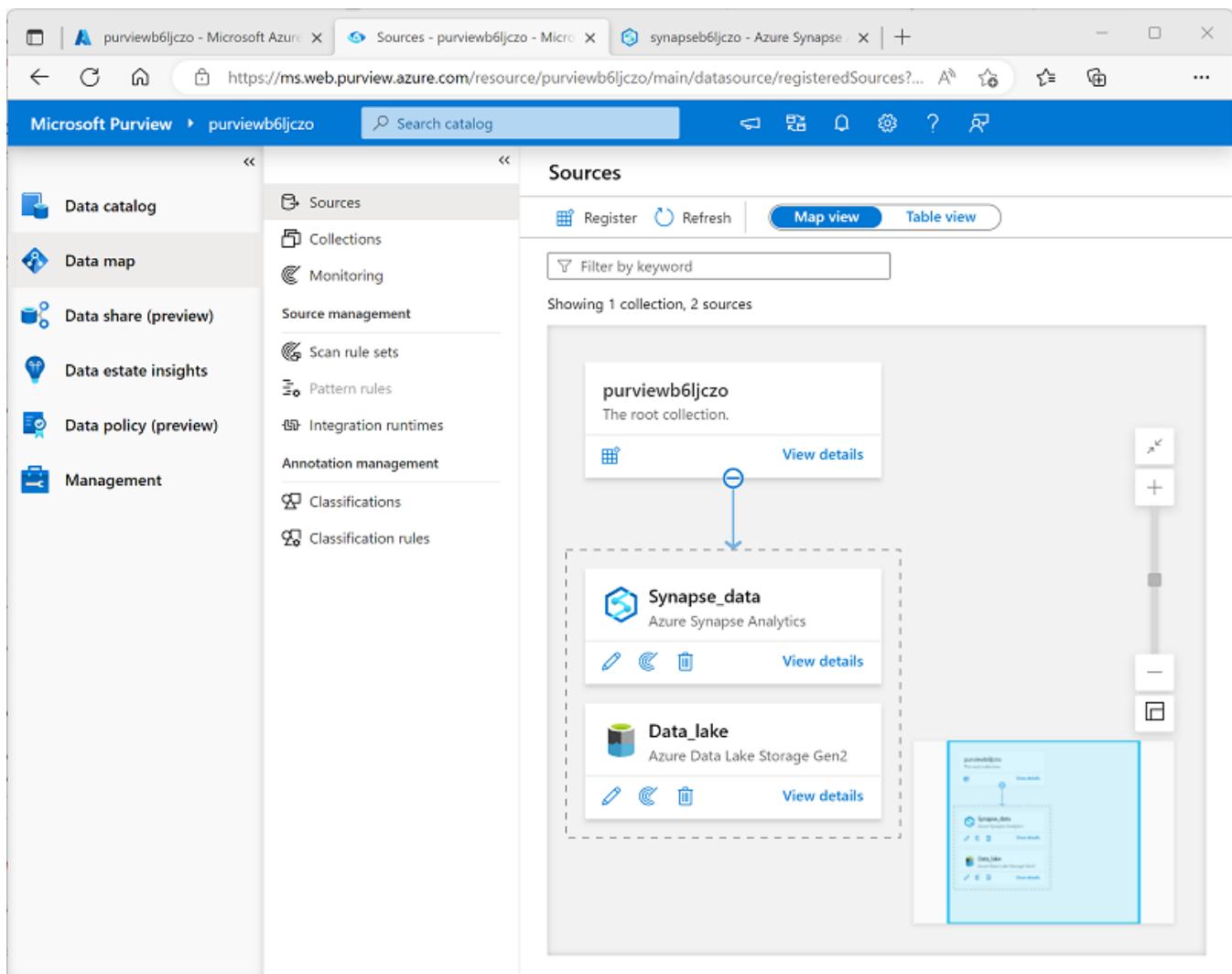
Register and scan data sources

Microsoft Purview supports the creation of a *data map* that catalogs data assets in *collections* by scanning registered *sources*. Collections form a hierarchy of logical groupings of related data assets, under a root collection that is created when you provision a Microsoft Purview account. You can use the Microsoft Purview Governance Portal to create and manage collections in your account.

To include assets from a particular data source, you need to register the source in a collection. Microsoft Purview supports many kinds of source, including:

- **Azure Synapse Analytics** - One or more SQL databases in a Synapse Analytics workspace.
- **Azure Data Lake Storage Gen2** - Blob containers used to host folders and files in a data lake.

To catalog assets used in an Azure Synapse Analytics workspace, you can register one or both of these sources in a collection, as shown here:



After registering the sources where your data assets are stored, you can scan each source to catalog the assets it contains. You can scan each source interactively, and you can schedule

period scans to keep the data map up to date.

Tip

To learn more about registering and scanning sources, see [Scans and ingestion in Microsoft Purview](#).

You'll get a chance to register and scan sources for an Azure Synapse Analytics workspace in the exercise later in this module.

View and manage cataloged data assets

As each scan finds data assets in the registered sources, they're added to the associated collection in the data catalog. You can query the data catalog in the Microsoft Purview Governance Portal to view and filter the data assets, as shown here:

The screenshot shows the Microsoft Purview Governance Portal interface. On the left, there's a sidebar with various navigation links: Data catalog, Data map, Data share (preview), Data estate insights, Data policy (preview), and Management. The main content area is titled "Browse assets" and shows a search bar with "Search catalog". Below the search bar, there are two tabs: "By collection" (which is selected) and "By source type". There are also buttons for "Refresh", "View collection tree", and "Clear all filters". A message indicates "Showing 1-3 out of 3 results". The results are listed as follows:

- products**: Azure Synapse Dedicated SQL Table
mssql://synapseb6ljczo.sql.azuresynapse.net/sql
- products.csv**: Azure Data Lake Storage Gen2 | File
https://datalakeb6ljczo.dfs.core.windows.net/file
- products_csv**: Azure Synapse Serverless SQL View
mssql://synapseb6ljczo-on-demand.sql.azuresynapse.net/sql

At the bottom, there are pagination controls: "Page 1 of 1".

Data assets include items in the registered data stores at multiple levels. For example, assets from an Azure Synapse Analytics source include databases, schemas, tables, and individual

fields; and assets from an Azure Data Lake Storage Gen 2 source include containers, folders, and files.

You can view and edit the properties of each asset to add contextual information such as descriptions, contacts for expert help, and other useful metadata. Data assets can also be classified using built-in or custom classifications that match specific patterns of data field to common types of data - for example, passport numbers, credit card numbers, and others.

 **Tip**

To learn more about data asset classification, see [Data classification in the Microsoft Purview governance portal](#).

Next unit: Connect Microsoft Purview to an Azure Synapse Analytics workspace

[Continue >](#)

How are we doing?     

✓ 100 XP

Connect Microsoft Purview to an Azure Synapse Analytics workspace

5 minutes

So far, you've learned how you can use Azure Synapse Analytics data stores as sources for a Microsoft Purview catalog; which is similar in most respects to using any other data source.

What sets Azure Synapse Analytics apart from many other data sources is the ability to configure direct integration between an Azure Synapse Analytics workspace and a Microsoft Purview account. By linking your workspace to a Purview account, you can:

- Search the Purview catalog in the Synapse Studio user interface.
- Push details of data pipeline activities to Purview in order to track data lineage information.

Connect a Purview account to a Synapse Analytics workspace

You connect a Microsoft Purview account to an Azure Synapse Analytics workspace on the [Manage](#) page of Synapse Studio, as shown here:

The screenshot shows the Microsoft Azure Synapse Analytics studio interface. On the left, there's a navigation bar with icons for Home, Data, Develop, Integrate, Monitor, and Manage. Under the 'Integrate' section, the 'Microsoft Purview' option is selected, highlighted with a blue background. The main content area is titled 'Microsoft Purview' and contains the following information:

- Text: "Connecting Azure Synapse workspace to Microsoft Purview will enable you to discover trusted and accurate data across your hybrid environment. [Learn more](#)"
- Buttons: 'About', 'Purview account' (which is currently selected), 'Edit', 'Refresh', and 'Disconnect'.
- Section: 'Purview account' with the value 'purviewvorsyde'.
- Text: 'Default Purview integration capabilities:' followed by a bullet point: '• Data Catalog - Discover enterprise data assets using Microsoft Purview powered search'.
- Text: 'Additional Purview integration capabilities and status:'
- Table: A two-column table with 'Integration' and 'Status' headers. It has one row: 'Data Lineage - Synapse Pipeline' with a green checkmark and the status 'Connected'.

Security considerations

To connect a Purview account by using the Synapse Studio interface, you require **Collection Administrator** access to the Purview account's root collection. After successfully connecting the account, the managed identity used by your Azure Synapse Analytics workspace will be added to the collection's **Data Curator** role.

If your Microsoft Purview account is behind a firewall, you need to create a managed endpoint, and configure the connection to access Purview using that endpoint. For more information, see [Access a secured Microsoft Purview account from Azure Synapse Analytics](#).

Tip

To learn more about connecting Azure Synapse Analytics to Microsoft Purview, see [QuickStart: Connect a Synapse workspace to a Microsoft Purview account](#).

You'll get a chance to connect an Azure Synapse Analytics workspace to a Microsoft Purview account in the exercise later in this module.

✓ 100 XP



Search a Purview catalog in Synapse Studio

4 minutes

After connecting an Azure Synapse Analytics workspace to a Microsoft Purview account, you can search the Purview catalog from Synapse Studio. This ability to discover and examine data assets from across the enterprise can greatly assist data engineers, data analysts, and other consumers of data by providing a curated catalog of documented data sources for analysis and reporting.

Search the Purview catalog in Synapse Studio

You can search the catalog from a connected Purview account by using the **Search** bar in the **Data**, **Develop**, or **Integrate** pages in Synapse Studio, as shown here:

The screenshot shows the Microsoft Synapse Studio interface. On the left, there's a navigation sidebar with icons for Home, Data, Develop, Integrate, Monitor, and Manage. The 'Integrate' tab is currently selected. In the center, there's a search bar with the placeholder 'products'. Below the search bar, there are two sections: 'Search suggestions' and 'Asset suggestions'. The 'Search suggestions' section lists 'products', 'products csv', 'products products csv', and 'products'. The 'Asset suggestions' section lists several items under 'products': a file link to 'https://datalakeb6ljczo.dfs.core.windows.net/files/products/' and three database links to 'mssql://synapseb6ljczo.sql.azuresynapse.net/sqlb6ljczo/dbo/products', 'products.csv' at 'https://datalakeb6ljczo.dfs.core.windows.net/files/products/products.csv', and 'products.csv' at 'mssql://synapseb6ljczo-on-demand.sql.azuresynapse.net/lakedb/dbo/produc...'. At the bottom of the search interface, there's a 'View search results' button. To the right of the search interface, there's a large preview pane showing a SQL script titled 'SQL script 2'. The script includes a 'Jndo' section with a dropdown menu, a 'Publish' button, a 'Query plan' button, and a '...'. The script itself starts with 'E USER purviewb6ljczo FROM EXTERNAL PROVIDER;'. Below the preview pane, there are two tabs: 'Results' and 'Messages'. The 'Results' tab displays a magnifying glass icon and the message 'No results to show. Your query yielded no displayable results.' The 'Messages' tab is empty. At the very bottom of the screen, there's a status bar with the message '00:00:15 Query executed successfully.'

The *search results* interface, and the details for each asset found reflect the user interface in the Microsoft Purview Governance Portal, ensuring that the data discovery and examination experience in Synapse Studio is consistent for users of Microsoft Purview in its own portal.

 **Tip**

For more information about searching the Purview catalog in Synapse Studio, see [Discover, connect, and explore data in Synapse using Microsoft Purview](#).

You'll get a chance to try searching a connected Purview account for yourself in the exercise later in this module.

Next unit: Track data lineage in pipelines

[Continue >](#)

How are we doing?     

100 XP

Track data lineage in pipelines

4 minutes

In a typical large-scale analytics solution, data is transferred and transformed across multiple systems until it's loaded into an analytical data store for reporting and analysis. Tracking the *lineage* of data as moves across the enterprise is an important factor in determining the provenance, trustworthiness, and recency of data assets used to inform analysis and decision making.

Generate and view data lineage information

In Azure Synapse Analytics, data movement and transformation is managed by using *pipelines*, which consist of an orchestrated set of *activities* that operate on data. The design and implementation of pipelines is too large a subject to cover in depth in this module, but a key point to be aware of is that there are two activity types available in Synapse Analytics pipelines that automatically generate data lineage information in a connected Purview catalog:

- The **Copy Data** activity
- The **Data Flow** activity

Running a pipeline that includes either of these activities in a workspace with a connected Purview account will result in the creation or update of data assets with lineage information. The assets recorded include:

- The source from which the data is extracted.
- The activity used to transfer the data.
- The destination where the data is stored.

In the Microsoft Purview Governance Portal, you can open the assets in the Purview catalog, and view the lineage information as shown here:

The screenshot shows the Microsoft Purview Data Catalog interface. On the left, there's a sidebar with icons for Data catalog, Data map, Data share (preview), Data estate insights (preview), Data policy (preview), and Management. The main area has a breadcrumb trail: Microsoft Purview > purviewvorsyde > Data catalog > Browse assets > Copy_lxr. The title bar says 'Copy_lxr - purviewvorsyde - Microsoft Azure' and the URL is 'https://web.purview.azure.com/resource/purviewvorsyde/main/catalog/entity?guid=d0740027-b16a-...'. Below the title bar are buttons for Edit, Select for bulk edit, Request access, Refresh, Delete, and Open in Azure Synapse Analytics. The main content area has tabs for Overview, Properties, Lineage, Contacts, and Related. The Lineage tab is selected, showing a search bar 'Search for assets or processes' and a list of columns under 'Input datasets > Columns'. The list includes 'products.csv' with sub-columns ProductID, ProductName, Category, and ListPrice. To the right, a diagram shows a flow from 'products.csv' to 'Copy_lxr' (Azure Synapse Copy Activity) to 'products'. The 'Copy_lxr' node has details: Workspace: synapsevorsyde.azuresynapse.net, Pipeline: /subscriptions/c1503d99-9ef7-4c79-4. There are buttons for See details and Open in Azure Synapse Analytics.

You can also view the lineage for a pipeline activity in Synapse Studio.

💡 Tip

For more information about tracking data lineage for Azure Synapse Analytics pipelines in Microsoft Purview, see [How to get lineage from Azure Synapse Analytics into Microsoft Purview](#).

You'll get a chance to generate and view data lineage from a Synapse Analytics pipeline in the exercise later in this module.

Next unit: Exercise - Integrate Azure Synapse Analytics and Microsoft Purview

[Continue >](#)

✓ 200 XP



Knowledge check

4 minutes

1. You want to scan data assets in a dedicated SQL pool in your Azure Synapse Analytics workspace. What kind of source should you register in Microsoft Purview? *

Azure Synapse Analytics

✓ Correct. An Azure Synapse Analytics source is used to scan database in SQL pools in a workspace.

Azure Data Lake Storage Gen2

Azure SQL Database

2. You want to scan data assets in the default data lake used by your Azure Synapse Analytics workspace. What kind of source should you register in Microsoft Purview? *

Azure Synapse Analytics

Azure Data Lake Storage Gen2

✓ Correct. Azure Synapse Analytics uses an Azure Data Lake Storage Gen2 storage account for the default data lake storage.

Azure Cosmos DB

3. You want data analysts using Synapse Studio to be able to find data assets that are registered in a Microsoft Purview collection. What should you do? *

Register an Azure Synapse Analytics source in the Purview account

Add a Data Explorer pool to the Synapse Workspace

Connect the Purview account to the Synapse analytics workspace

✓ Correct. Connecting the Purview account to the workspace makes it searchable in Synapse Studio.

4. Which of the following pipeline activities records data lineage data in a connected Purview account? *

Get Metadata

Copy Data

✓ Correct. A Copy Data activity and its data lineage from source to target is recorded in a connected Purview account.

Lookup

Next unit: Summary

[Continue >](#)

How are we doing?