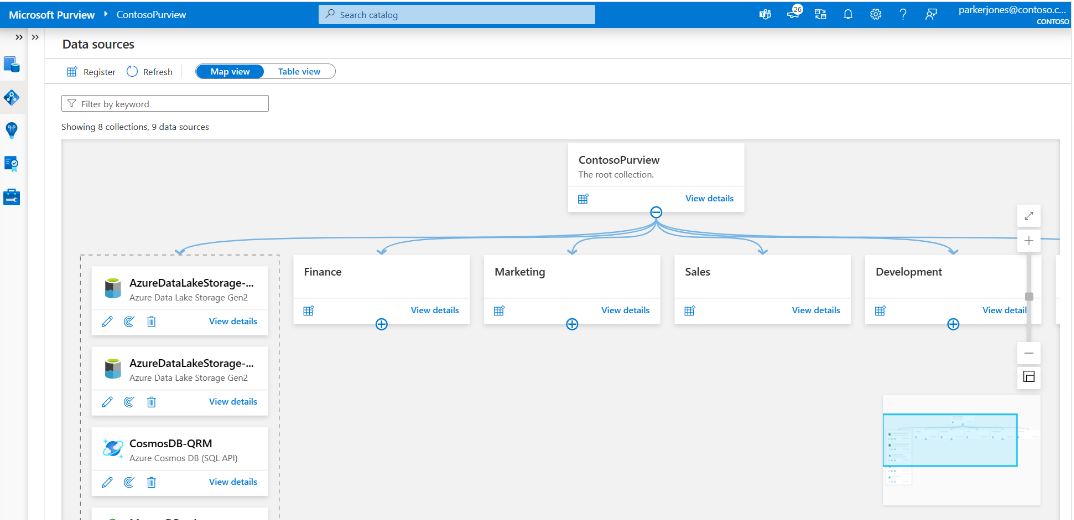
**Govern data across an Enterprise**

1. **Introduction to Microsoft Purview**

* Microsoft Purview is designed to help enterprise get the most value from their existing information assets.
* Its catalog makes data sources easy to discover and understand by the users who manage the data:
  + Mange and govern your data across various platforms and locations
  + Map out your data landscape
  + Classify sensitive data
  + Empower customers to find trustworthy data

1.1 What is Microsoft Purview

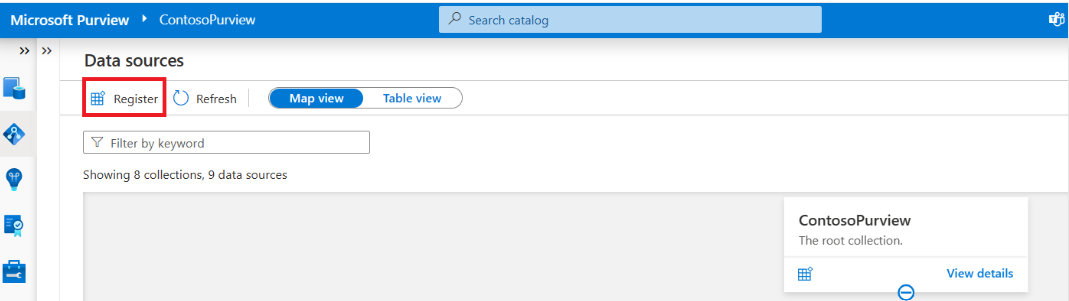
* It is a unified data-governance service that helps you manage and govern your on-premises, multi-cloud, and SaaS data.
* You can easily create a broad, up-to-date map of your data landscape with:
  + Automated data discovery
  + Sensitive data classification
  + End-to-end data lineage
* You can register 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, not limited to Azure services (e.g. AWS, Azure SQL Databases on-premises)
* Microsoft Purview has three main elements:
  + Microsoft Purview Data Map
    - It provides a structure for your data estate, divided in groups and hierarchies.
    - You can grant users and teams access to those groups so that they have access to find relevant data stores.
    - The data map can scan your data stores and gather metadata such as schemas and data types.



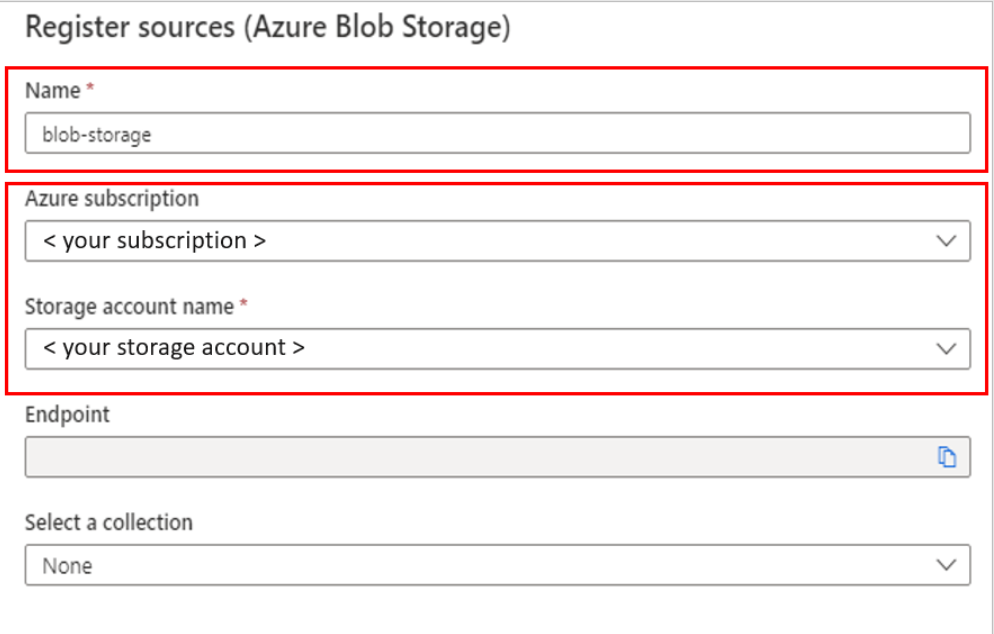
* + Microsoft Purview Data Catalog
    - It allows users to browse the metadata stored in the data map so they can find reliable data and understand its context.
    - For instance, users can see where the data comes from and who are the experts they can contact about the data source.
    - The data catalog also integrates with other Azure products, like the Azure Synapse Analytics workspace
  + Microsoft Purview Data Estate Insights. It offers a high-level view into your data catalogue, covering these key facets:
    - Data stewardship
      * A report which tracks governance progress
    - Catalog adoption
      * A report on the number of active users in your data catalog, top searches, and mostly viewed assets.
    - Asset insights
      * A report on the data estate and source-type distributions
    - Scan insights
      * A report that provides information on the health of your scans (success, failures, or cancelled).
    - Glossary insights
      * A report on the distribution of glossary terms by status, and view how the terms are attached to assets.
    - Classification insights
      * A report that shows where classified data is located (for security administrators).
    - 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 data estate.
  1. How Microsoft Purview works

*Load data in the data map*

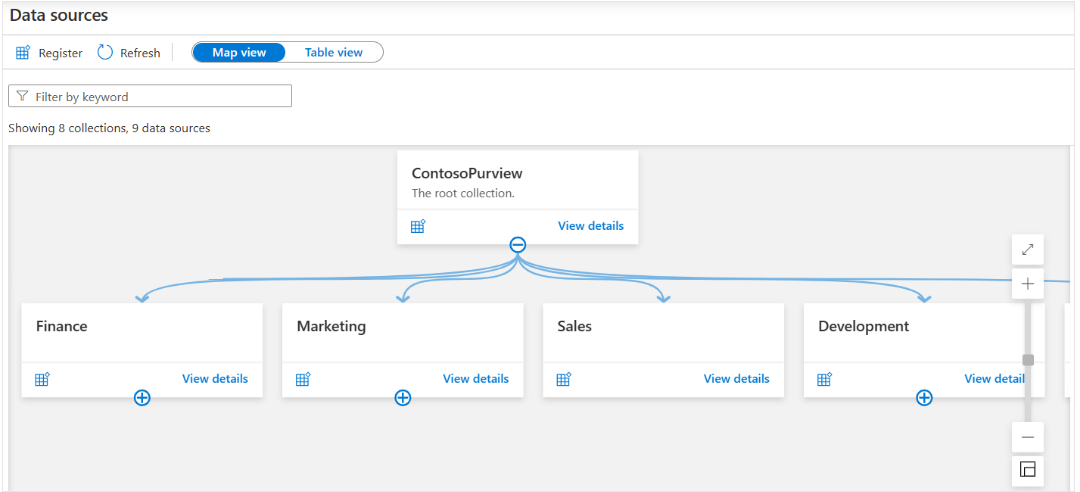
* Microsoft Purview Data Map is a unified map of your data assets and their relationships.
* It also houses the metadata that underpins the Microsoft Purview Data catalogue and Data Estate Insights.
* It scales up and down to meet your enterprise compliance requirements.
* Source data
  + Sourcing data starts with the registration of the data sources (on-premises, multi-cloud, and SaaS).
  + Data remains in its location and is not migrated to any other platform.



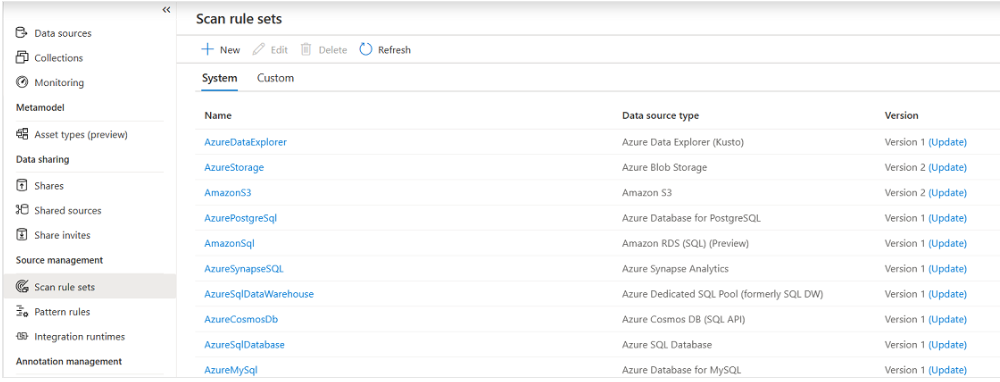
* + Each type of data source you choose requires specific information to complete the registration (e.g., subscription, storage account name).



* + After registration, you scan the data source. Scanning ingests metadata about your data source into Microsoft Purview Data Map.
  + Each data source has specific requirements for authenticating and configuration to permit scanning of the assets in that data source.
    - 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
  + It is the foundational platform for Microsoft Purview.
  + Customers create a knowledge graph of data that comes from a range of sources.
  + It consists of:
    - Data assets
    - Data lineage
    - Data classification
    - Business context
  + The data map uses collections to organize data assets into logical categories to simplify management and discovery of assets within the catalog.



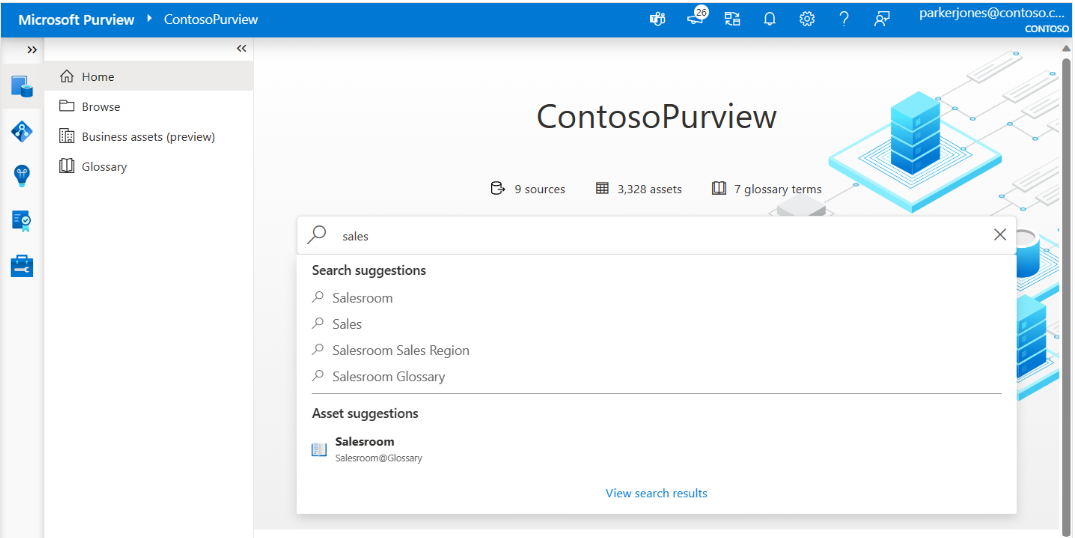
* Scan data
  + After you register your data sources, you'll need to run a scan to access their metadata and browse the asset information.
  + Before you can scan the data sources, you're required to enter the credentials for these sources.
  + You can use Azure Key vault to store the credential for security and ease access by your scan rules.
  + A *scan rule set* is a container for grouping scan rules together to use the same rules repeatedly (e.g., select file types for schema extraction and classification).



* 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
    - Financial
    - Personal
    - Security
    - Miscellaneous
  + You can use several system classifications to classify your data. The classifications align with the sensitive information types in the Microsoft Purview compliance portal.
  + 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 datatypes; that’s registered from the data source.

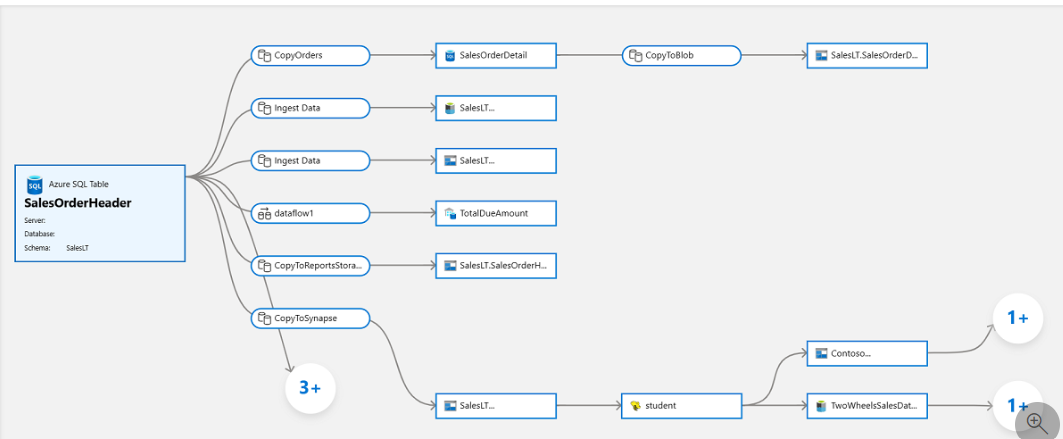
Browse and search

* 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 types, tags, ratings, or collection.
* Discovery enables you to use:
  + Sematic search and browse
  + Business glossary and workflows
  + Data lineage with sources, owners, transformations, and lifecycle



Data lineage

* The concept of data lineage focuses on the lifecycle of data. 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.



* 1. When to use Microsoft Purview

Discovery

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

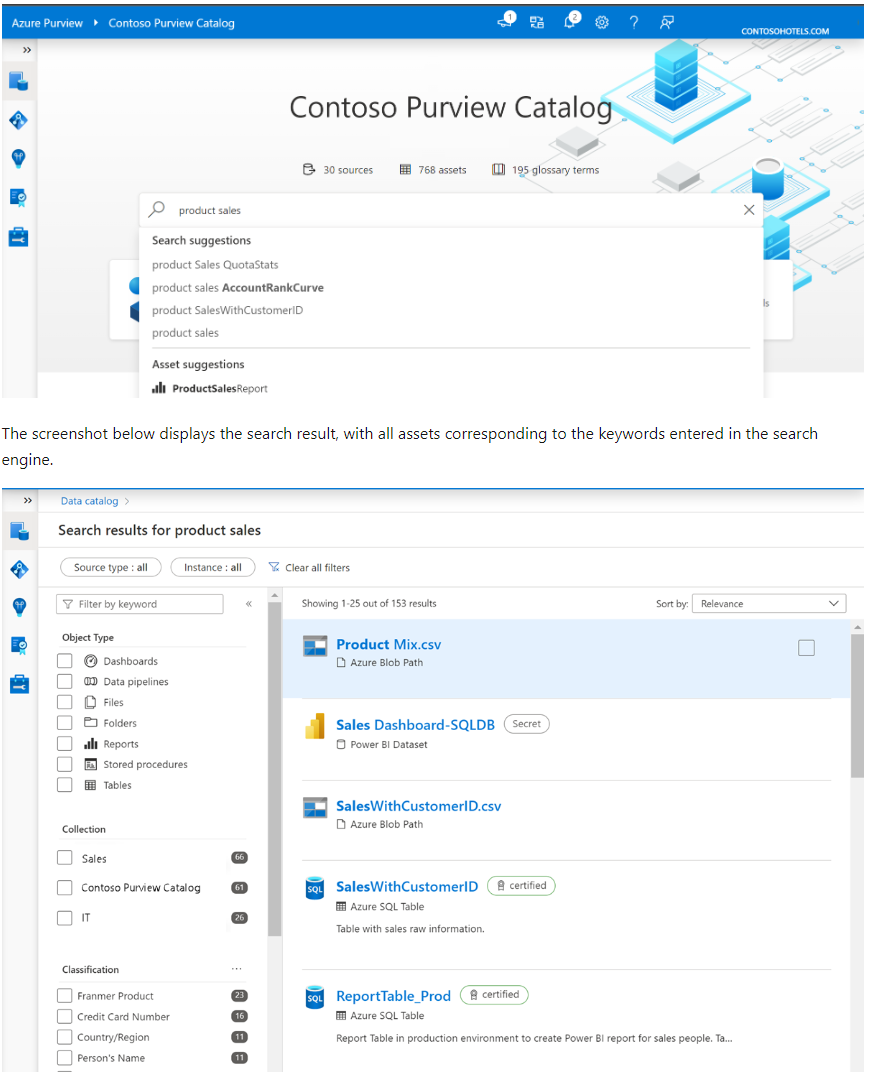
Governance

* Microsoft Purview classifies data by
  + Bloom Filter
    - It includes attributes for city, country/region, place, and person information.
  + RegEx
    - It cover attributes that include categories like bank information
      * ABA routing numbers or county/region-specific banking account numbers.
      * passport numbers, and country/region-specific identification numbers.
* Microsoft Purview also uses predefined Data plane roes o help control who has access to the information in Microsoft Purview.
* For access, users can use the Microsoft Purview governance portal only if they are placed in at least one of the three supported roles:
  + Purview Data Reader role
    - Access + Reading all content, except for scan bindings
  + Purview Data Curator
    - Access + Reading all content, except for scan bindings
    - Edit info information about assets, classification definitions, and glossary terms
    - Apply classifications and glossary terms to assets
  + Purview Data Source Administrative role
    - No access to the Microsoft Purview governance portal
    - Can manage all aspects of scanning data into Microsoft Purview
    - Does not have read or write access to content in Microsoft Purview beyond those tasks related to scanning.

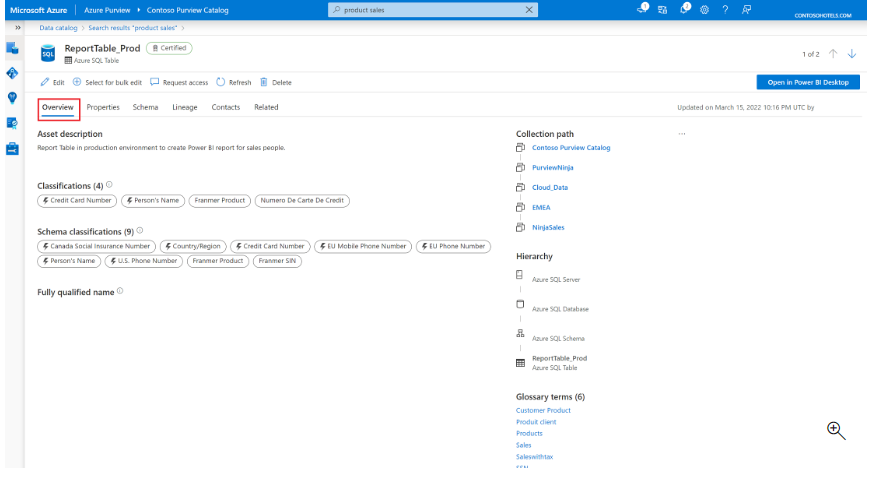
1. **Discover trusted data using Microsoft Purview**

2.1 Search for assets

* As a data analyst looking for assets, you will be searching the Microsoft Purview data catalog.
* This assumes that Microsoft Purview data Map has been created by your organization.



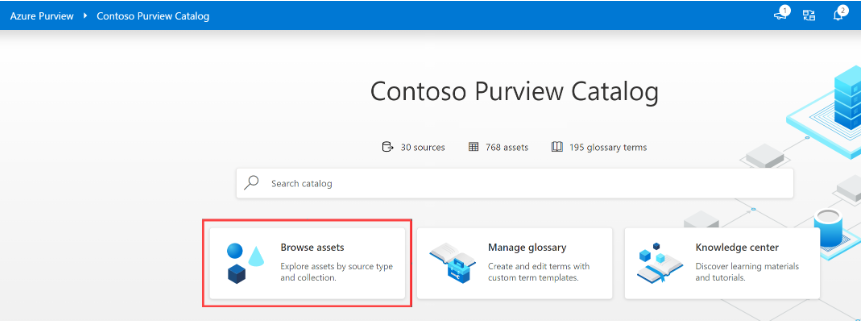
* Search the Microsoft Purview data catalog
  + From the home page, users can type relevant keywords to start discovering assets.
  + You can fine-tune your search using the filters on the left side of the page, e.g.
    - Source type
    - Object type
    - Classification
    - Glossary term
* Understand a single asset
  + Asset overview



* + - The overview displays information at a glance, including description, asset classifications, schema classification, collection path, asset hierarchy, and glossary terms.
    - The *asset description* provides a brief explanation of the purpose of an asset.
    - Beneath the description, you’ll see the *asset classification* and *schema classification.*
    - You can also view the *collection path*, *hierarchy*, and *glossary terms* on the right side of the overview tab.
      * *Collection path* refer to the location of the asset inside Microsoft Purview.
      * *Hierarchy* within the overview tab
      * *Glossary terms* are a managed vocabulary for business terms that can be used to categorize and relate assets across your environment.
  + Asset schema
    - The schema view of the asset includes more granular details about the asset, such as column names, data types, column level classification, terms, and descriptions.
  + Asset lineage
    - Asset lineage gives you a clear view of how the asset is populated and where data comes from (data lifecycle).
  + Asset contacts and related assets
    - Asset contacts provide you contact details of experts or datasets owners.

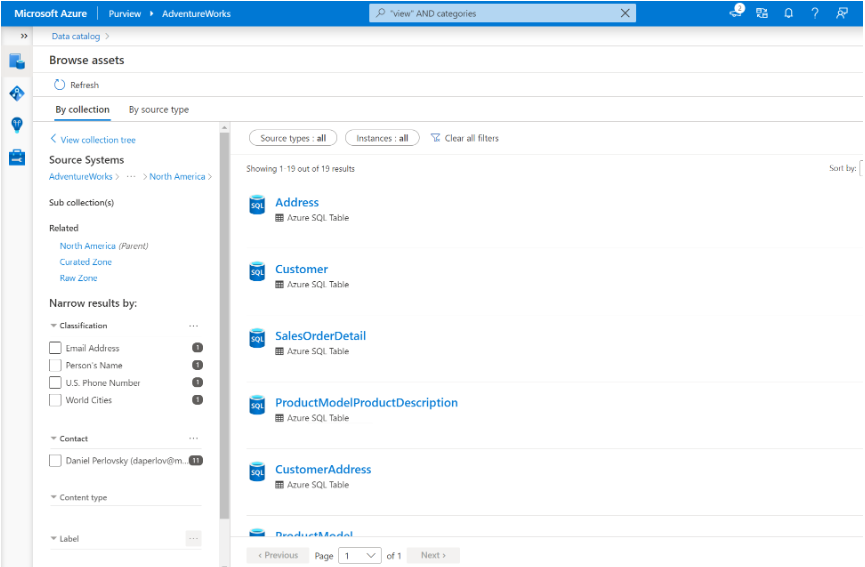
2.2 Browse assets

* The Microsoft Purview data catalog offers a browse experience that enables exploration of available data, either by
  + collection,
  + exploring the hierarchy of each data source in the catalog.



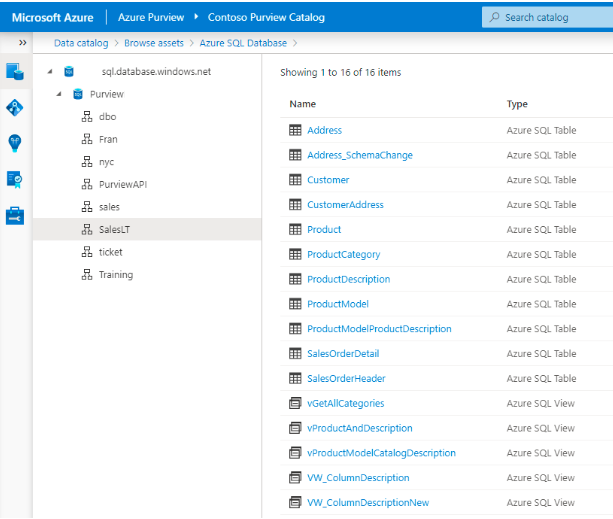
*Browse by collection*

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



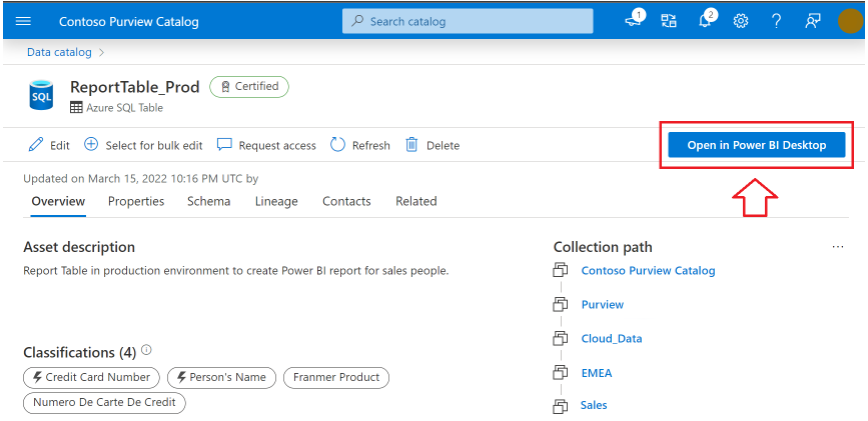
*Browse by source type*

* 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.



2.3 Use assets with Power BI

* 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
  + Microsoft Purview makes it simple to request access directly from the Data Catalog by using the “Request access” button.
  + Request access will kick off a workflow that manages requests and approvals.
* Build a Power BI report using data discovered in Purview
  + Purview makes it simple to open an asset in Power BI desktop.
  + Selecting Open in Power BI Desktop initiates the download of a Power BI Data Source file (PBIDS) you open with Power BI Desktop.
  + PBIDS files contain a connection to the data source, so all you need to do it enter credentials upon the file
  + 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 info about their lineage across Power BI workspace and their connections to data sources, are then available in Purview.



2.4 Integrate with Azure Synapse Analytics

* Microsoft Purview can be integrated directly into Azure Synapse, so you can get the data catalog experience directly into Azure Synapse Studio.
* To connect Purview-Synapse, you need 2 types of permissions:
  + Contributor role in Synapse workspace from Azure portal identity and access management (IAM).
  + You also need access to that Microsoft Purview Account.
* In Azure Synapse Studio, from the **Data** blade on the left, select **Purview** in the dropdown next to the search bar**.**
* Search for the asset that exists in Purview, through the search bar and fine tune your search by **Object Type** and **Collection**.

**3 Catalog data artifacts by using Microsoft Purview**

* 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 hierarchy of each data source.
* The first step is registering and scanning data, after which you can classify data for easy identification of assets to use for reporting.

3.1 Register and scan data

* Registration and scanning data of data enables discoverability of data across an estate.
* Collections
  + In Microsoft Purview Data Catalog, collections:
    - Drive permissions and asset protections
    - Use to understand data estate and catalog usage and adoption, as featured in the data stewardship section of your Data Estate Insights.
  + The data map is a the core of Microsoft Purview, which keeps an up-to-date map of assets and their metadata across the data estate.