

DataSaturday Oslo

04 September 2021



Azure Purview

Microsoft's answer to Data Governance and Data Lineage

Microsoft
Partner

2021 Partner of the Year Finalist
Analytics Award



Erwin de Kreuk

Lead Data and AI



@erwindekreuk



<https://erwindekreuk.com>



@DataSatOslo



HELLHEIM
12:15 CET

DataSaturday Oslo

04 September 2021



Azure Purview

Microsoft's answer to Data Governance and Data Lineage

Microsoft
Partner

2021 Partner of the Year Finalist
Analytics Award

Erwin de Kreuk



@erwindekreuk

Lead Data and AI



<https://erwindekreuk.com>



HELLHEIM
12:15 CET

We Are InSpark

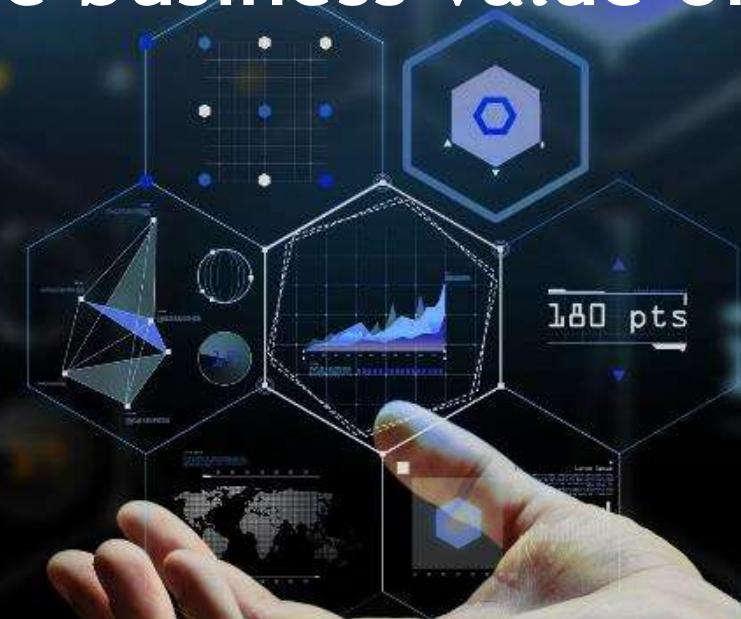
We help organizations
accelerating their digital
transformation with impactful
Microsoft solutions & expertise



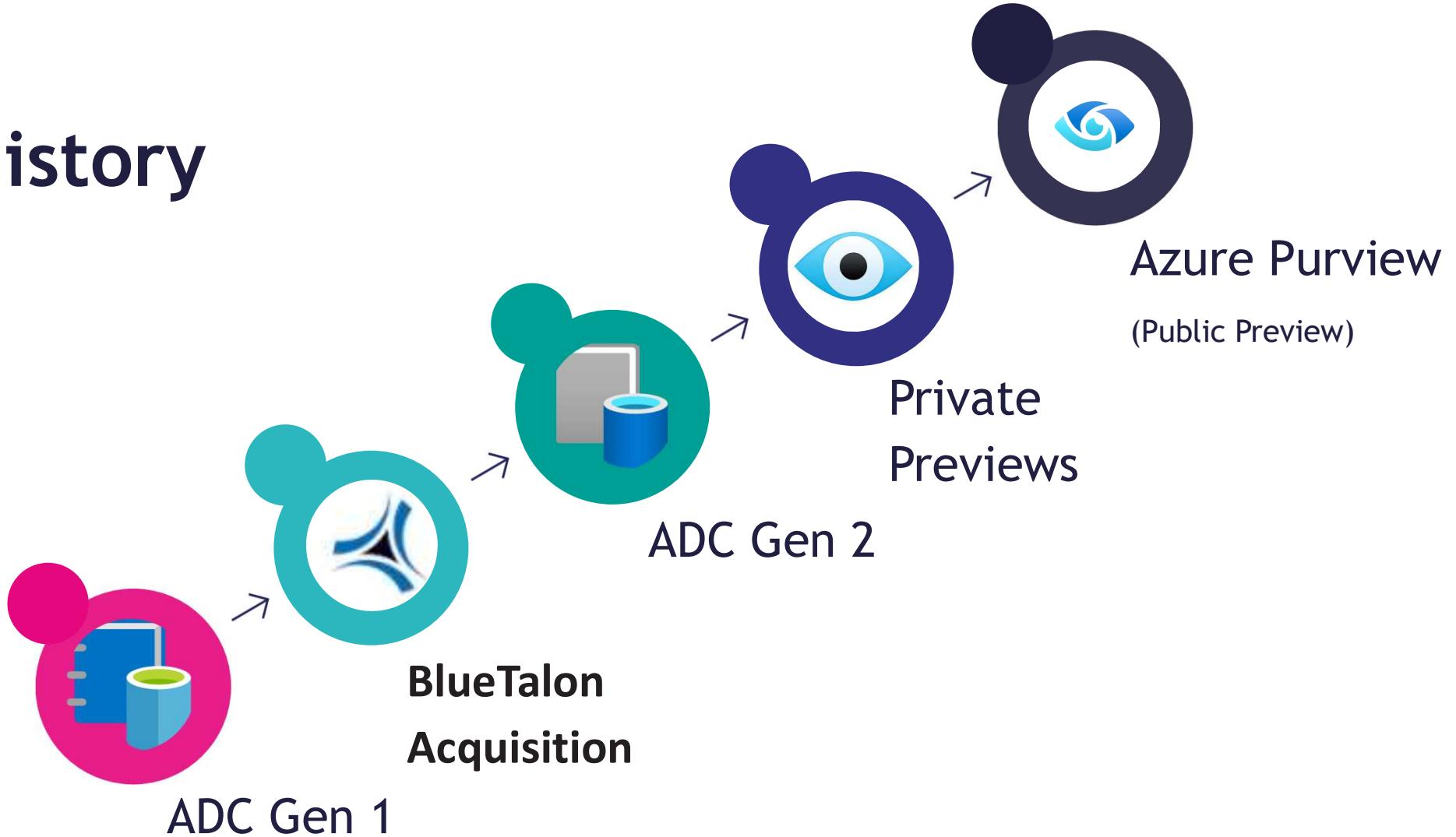
Azure Purview



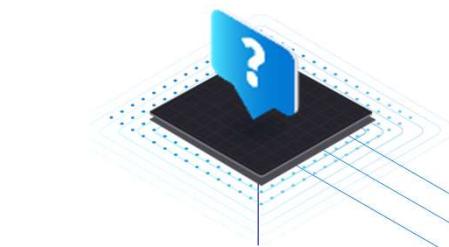
Unified **data governance** to
maximize the business value of data



History

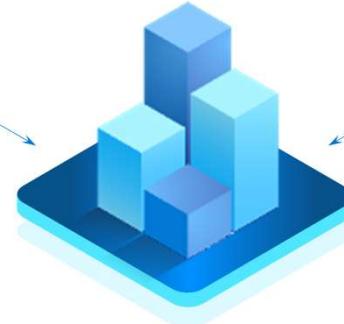


Data governance is becoming increasingly interdisciplinary



DISCOVERY

What data do I have?
Where did the data originate?
Can I trust it?

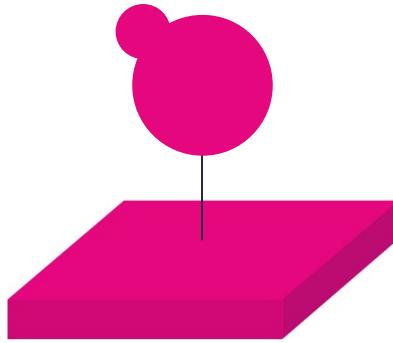


COMPLIANCE

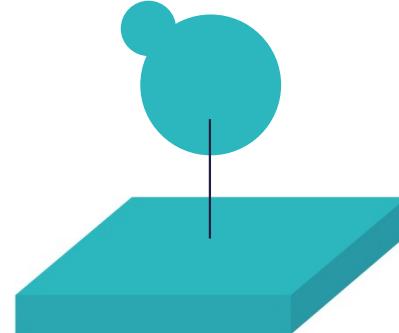
What's my exposure to risk?
Is my usage compliant?
How do I control access & use?
What is required by regulation X?

Chief Data Officer

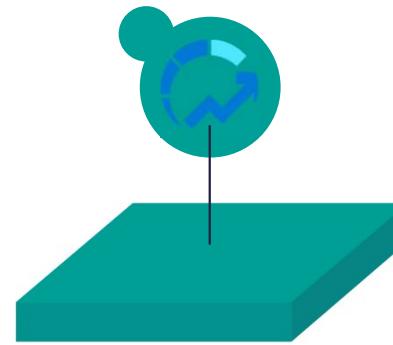
Elements of successful data governance



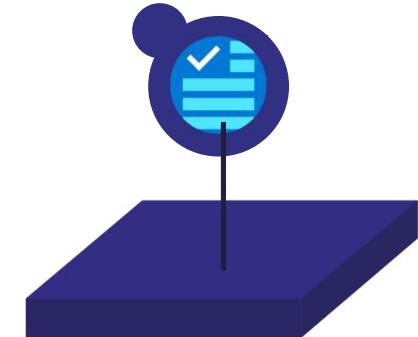
Manage growing
data landscape



Overcome
operational silos



Increase data
agility



Comply with
industry
regulations

Data Insights

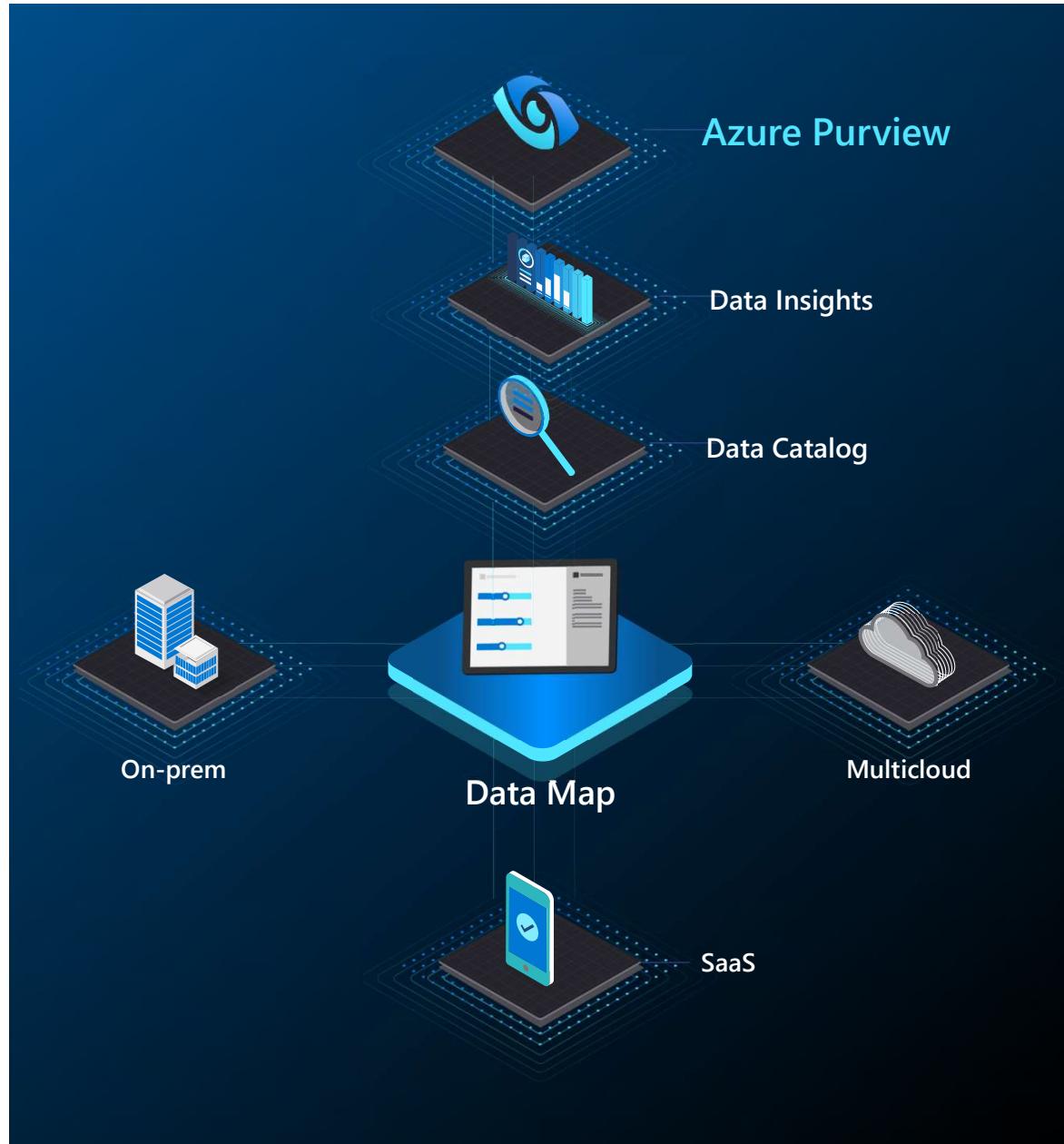
- Assess data usage across your organization

Data Catalog

- Enable effortless discovery for data consumers

Data Map

- Automate and manage metadata at scale



Azure Purview

Unified data governance to maximize the business value of data





Reimagine data governance in the cloud

- ❑ Manage and govern operational, transactional and analytical data
- ❑ Cloud-native, purpose-built service to address discovery and compliance needs
- ❑ Fully managed, serverless, PaaS service
- ❑ Eliminate manual, ad-hoc and homegrown solutions





Set the foundation for effective data governance

- Automate discovery of data in on-premises, multicloud and SaaS sources
- Classify data at scale to specify sensitivity, compliance, industry, business and company-specific value
- Know where data came from and what was derived from it with data lineage





Maximize business value of **data** for data consumers

- Connect business and technical data analysts, data scientists, and data engineers to a trusted data catalog
- Enable users to quickly find data and view its lineage and sensitivity
- Deliver a curated and consistent glossary of business terms and definitions



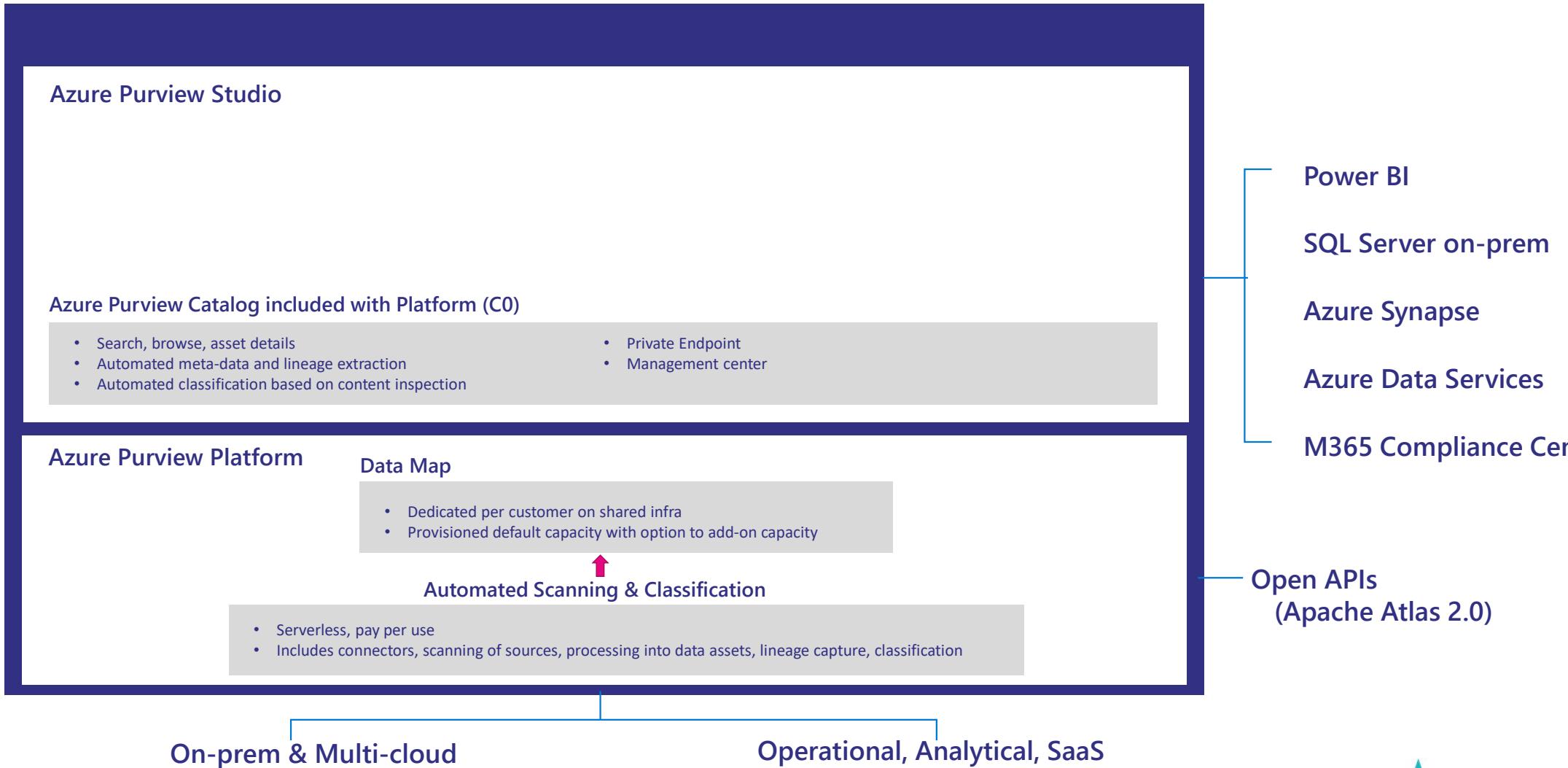


Gain insight into **data use**
across the estate

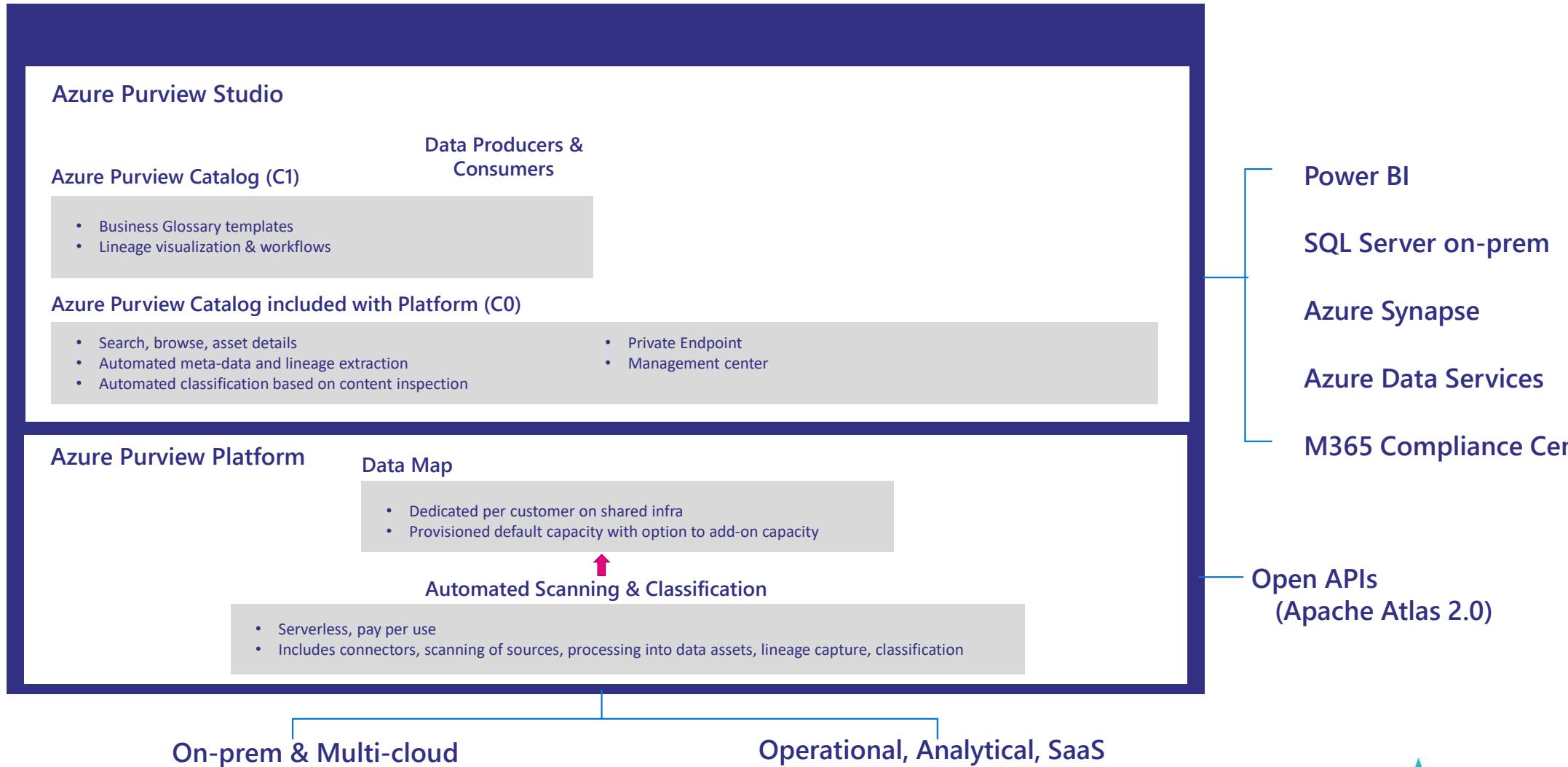
- Understand at a glance how data is being created and used across your data estate
- Visually assess the state of data assets, scans, business glossary and sensitive data



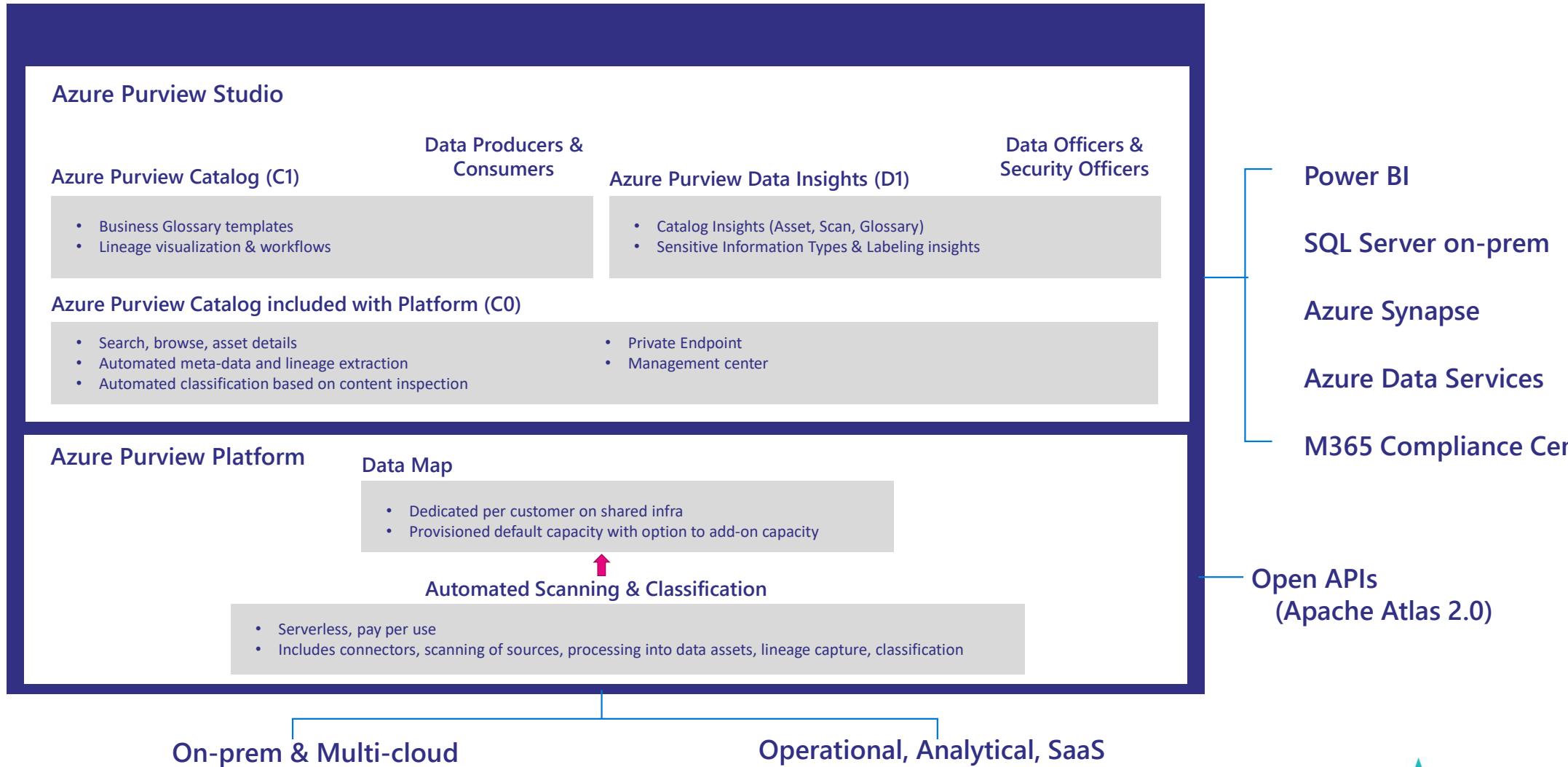
Azure Purview Features



Azure Purview Features



Azure Purview Features



Azure Purview - Access Control

Data Source Administrator

- No access to Purview Portal
- Can Manage all aspects of Scanning
- Ideal role for programmatic processes, such as service principals
- Can register Data Sources



Select a Purview account

Azure Purview is a unified data governance service that enables a foundational understanding of your on-premises, multi-cloud, operational and SaaS data. [Learn more](#)

Azure Active Directory

Account name

✖ You don't have access to this Purview account. Contact an Owner or User Access Administrator (UAA) for access in [Access control \(IAM\)](#)

[Continue](#)

Azure Purview - Access Control

Data Source Administrator

Data Reader

- Has access to Purview Portal
- Can read all content in Azure Purview

General

Account information

Metrics

Metadata management

Classifications

Classification rules

External connections

Data Factory

Data Share

Security and access

Access control

Account information

By participating in Purview Public Preview, you acknowledge that you have read and you agree to the Supplemental Terms of Use for Microsoft Azure Previews, including additional terms for Purview Public Preview. You can find those terms [here](#).

Overview

Account name labseuwvdvImpurviewoxgn01

Friendly name * Azure Purview Demo

Azure Purview Demo

1,138 assets 1 glossary term

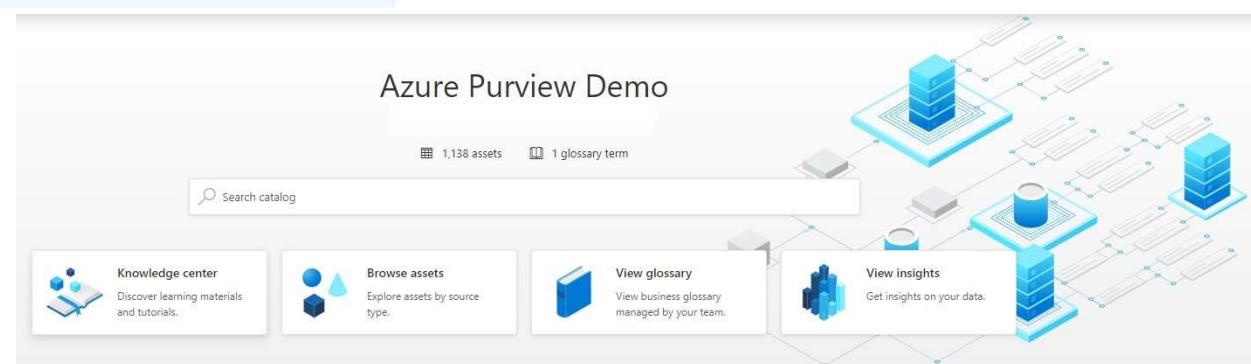
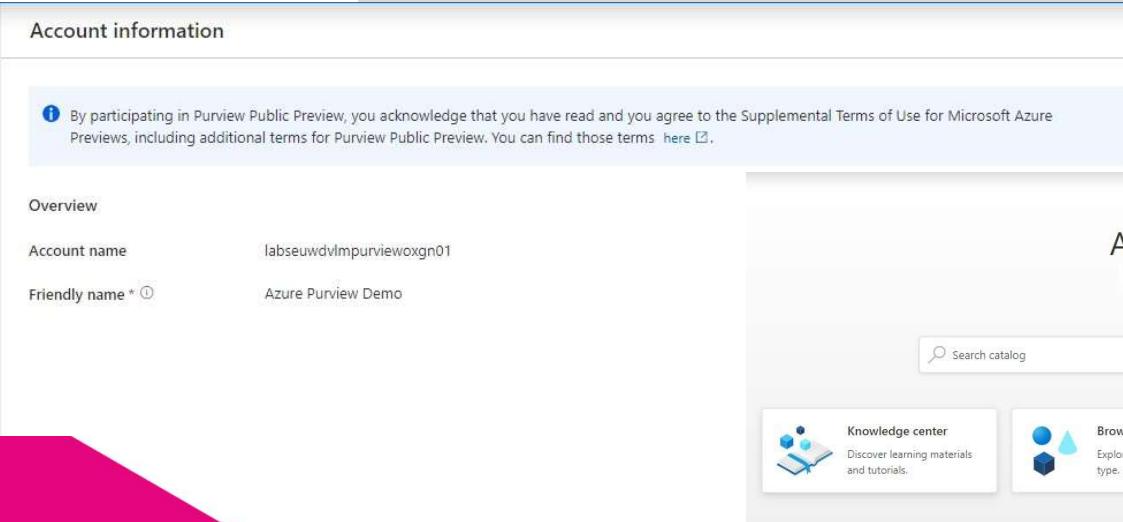
Search catalog

Knowledge center Discover learning materials and tutorials.

Browse assets Explore assets by source type.

View glossary View business glossary managed by your team.

View insights Get insights on your data.



Azure Purview - Roles

Data Source Administrator

- Has access to Purview Portal
- Can read all content in Azure Purview
- Can edit assets, classification and glossary terms
- Can apply classifications and glossary terms to assets.
- **Can not Register Data Sources, only read**

Data Reader

Data Curator

General

Account information

Data sources

Scan rule sets

Integration runtimes

Metrics

Metadata management

Classifications

Classification rules

External connections

Data Factory

Data Share

Security and access

Access control

Credentials

Account information

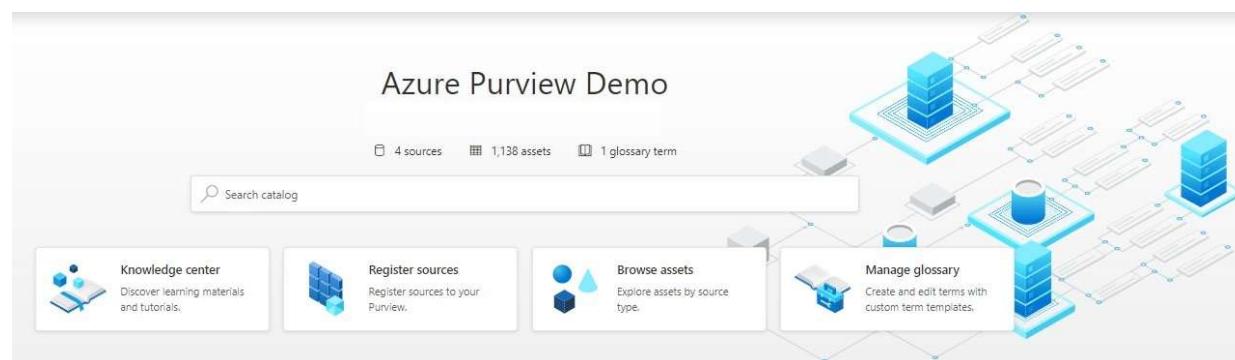
By participating in Purview Public Preview, you acknowledge that you have read and you agree to the Supplemental Terms of Use for Microsoft Azure Previews, including additional terms for Purview Public Preview. You can find those terms [here](#).

Overview

Account name

Friendly name *

Save



Error
Request failed with status code 403
Not authorized to access account

Azure Purview - Access Control

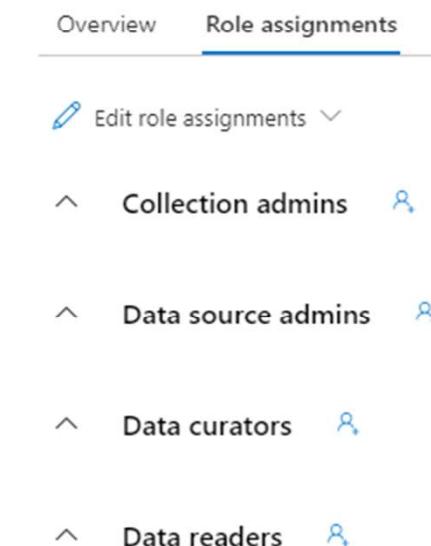
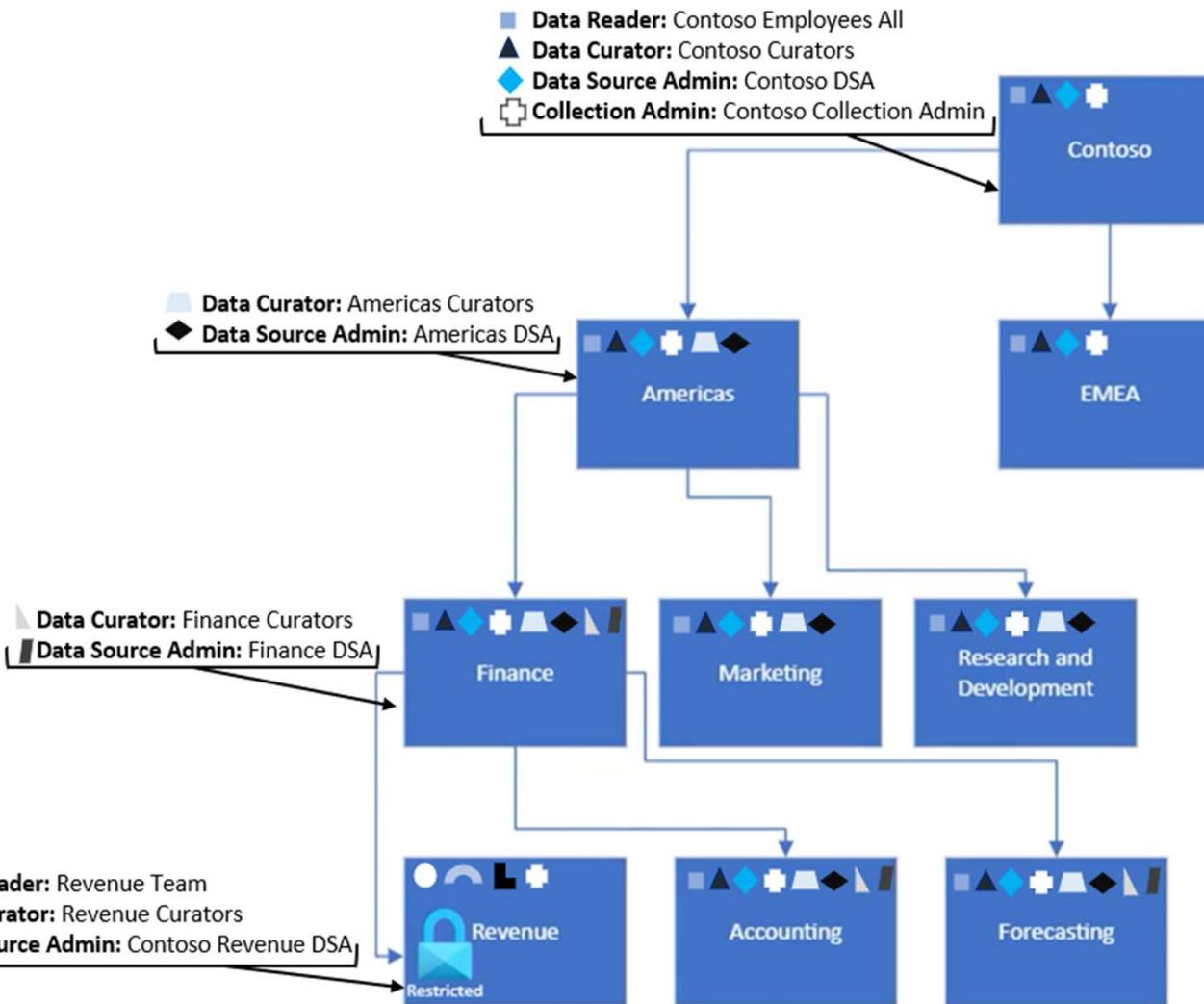
Data Source Administrator

Data Reader

Data Curator

	Sources Register and scan	Source management SRS, Credentials, IR, Pattern rules, External connections	Catalog Search, browse and asset details	Annotate Classification, *classification rules	Glossary	Collections Access control	Insights
 Account creator All access	✓	✓	✓	✓	✓	✓	✓
 Data source admin ScanRead, ScanWrite, CollectionRead	✓	✓	✗	✗	✗	 READ-ONLY	✓
 Collection admin CollectionRead, CollectionWrite	✗	✗	✗	✗	✗	✓	✓
 Data curator DataRead, DataWrite, CollectionRead	✗	✗	✓	✓	✓	 READ-ONLY	✓
 Data reader DataRead, CollectionRead	✗	✗	 READ-ONLY	 READ-ONLY	 READ-ONLY	 READ-ONLY	✓

Azure Purview - Access Control



Azure Purview - Pricing

Azure Purview Data Map

- Capacity Unit
 - €0.289 per 1 Capacity Unit Hour
 - Provisioned API throughput. 1 capacity unit = 1 API/sec
 - Includes 4 capacity units for free until February 28, 2021.
- Metadata Storage
 - Free in preview

Azure Purview - Pricing

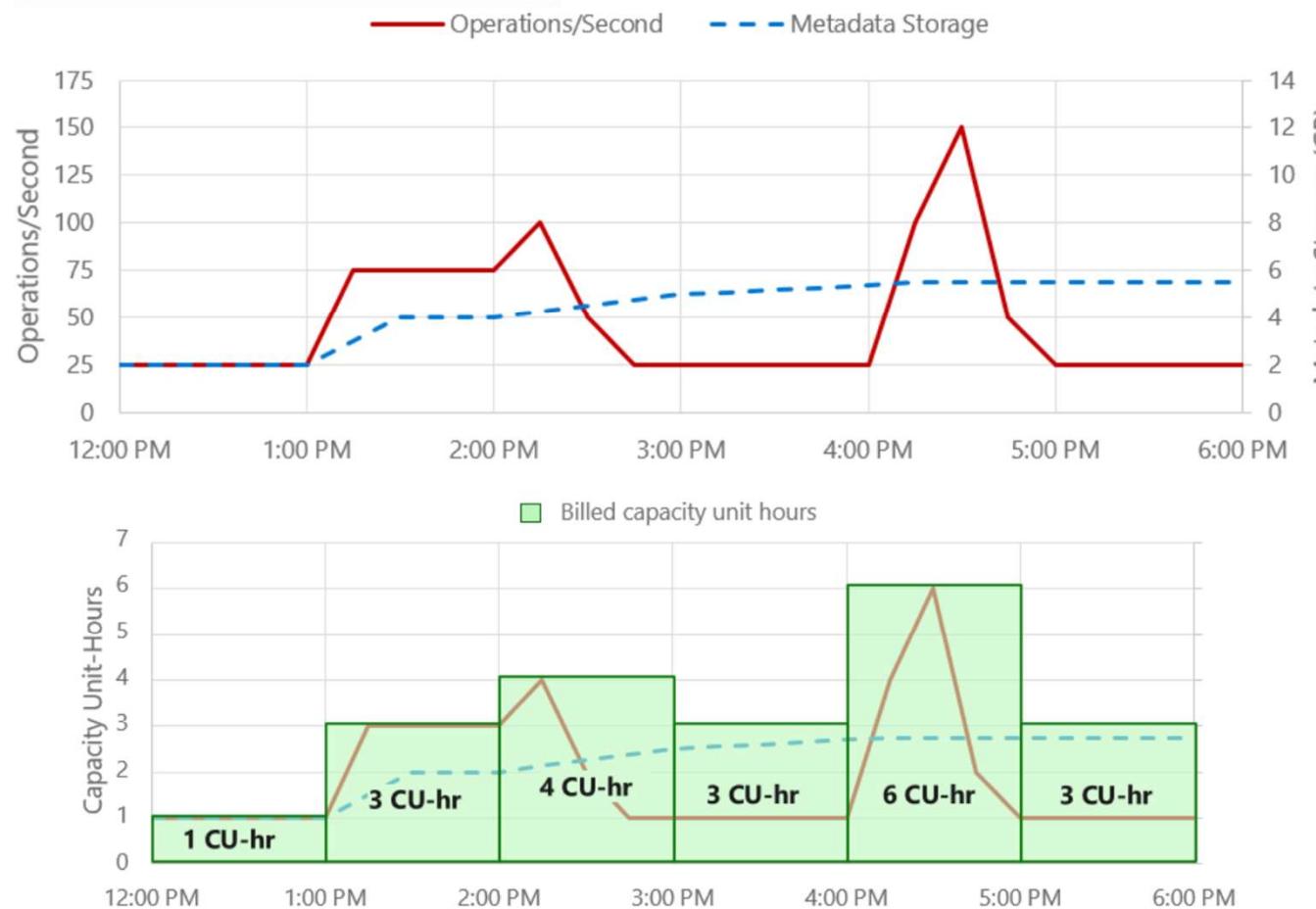
Azure Purview Data Map

- Capacity Unit
€0.289 per 1 Capacity Unit Hour
Unlimited provisioned API throughput. 1 capacity unit = 1000 QPS.
4 capacity units for free until Fall 2021.

Changed for all new Purview
Accounts created after or on
August 18th, 2021

Azure Purview - Pricing

Azure Purview Data Map



Data Map Capacity Unit	Operations/Sec throughput	Storage capacity in GB
1	25	2
2	50	4
3	75	6
4	100	8
5	125	10
6	150	12
7	175	14
8	200	16
9	225	18
10	250	20

Azure Purview - Pricing

Azure Purview Data Map

- Power BI Online
 - Free in Preview
 - SQL Server On Prem
 - Free in Preview
 - Other Data Sources
 - Free in Preview
 - €0.532 per 1 vCore Hour
- Includes 16 vCore-hours for Free every month until February 28, 2021

Scanning and Classification

Azure Purview - Pricing

Azure Purview Data Map

Scanning and Classification

Azure Purview Data Catalog

- C0
 - Included with the Data Map
Search and browse of data assets
- C1
 - Free in preview
 - Business glossary, lineage visualization and catalog insights
- D0
 - Free in preview
 - Sensitive data identification insights

<https://azure.microsoft.com/en-us/pricing/details/azure-purview>

Azure Purview Studio

The screenshot shows the Azure Purview Studio interface with several UI elements highlighted:

- Metrics**: Located in the top left corner.
- Search Bar**: Located in the top left corner.
- Recently Accessed Entities**: Located in the top left corner.
- Updates**, **Accounts**, **Notifications**: Located in the top right corner, connected to a blue bar with small icons.
- Feedback**: A pink callout box pointing to the feedback icon in the top right corner.
- Key Activities**: A teal callout box pointing to the "Knowledge center" section in the center of the screen.
- Usefull Links**: A teal callout box pointing to the "Links" section on the right side.

Azure Purview Demo

6 sources | 1,018 assets | No glossary terms

Search catalog

Browse assets | Manage glossary | Knowledge center

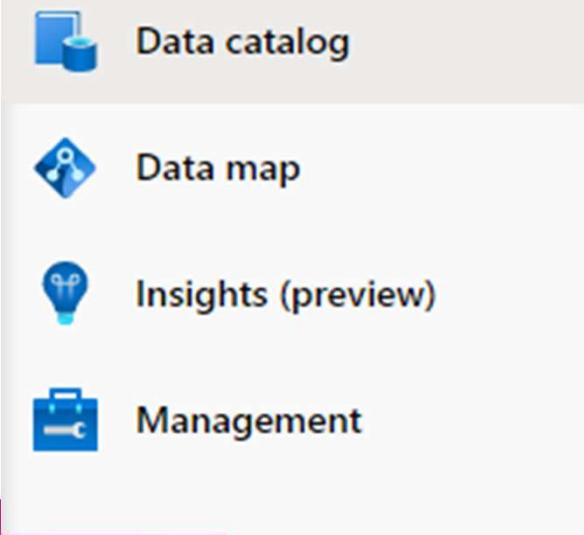
Recently accessed | My items

Name	Last update
Test_Purview	2 hours ago
cities	21 hours ago
Cities_staged	a day ago
telemetry	---

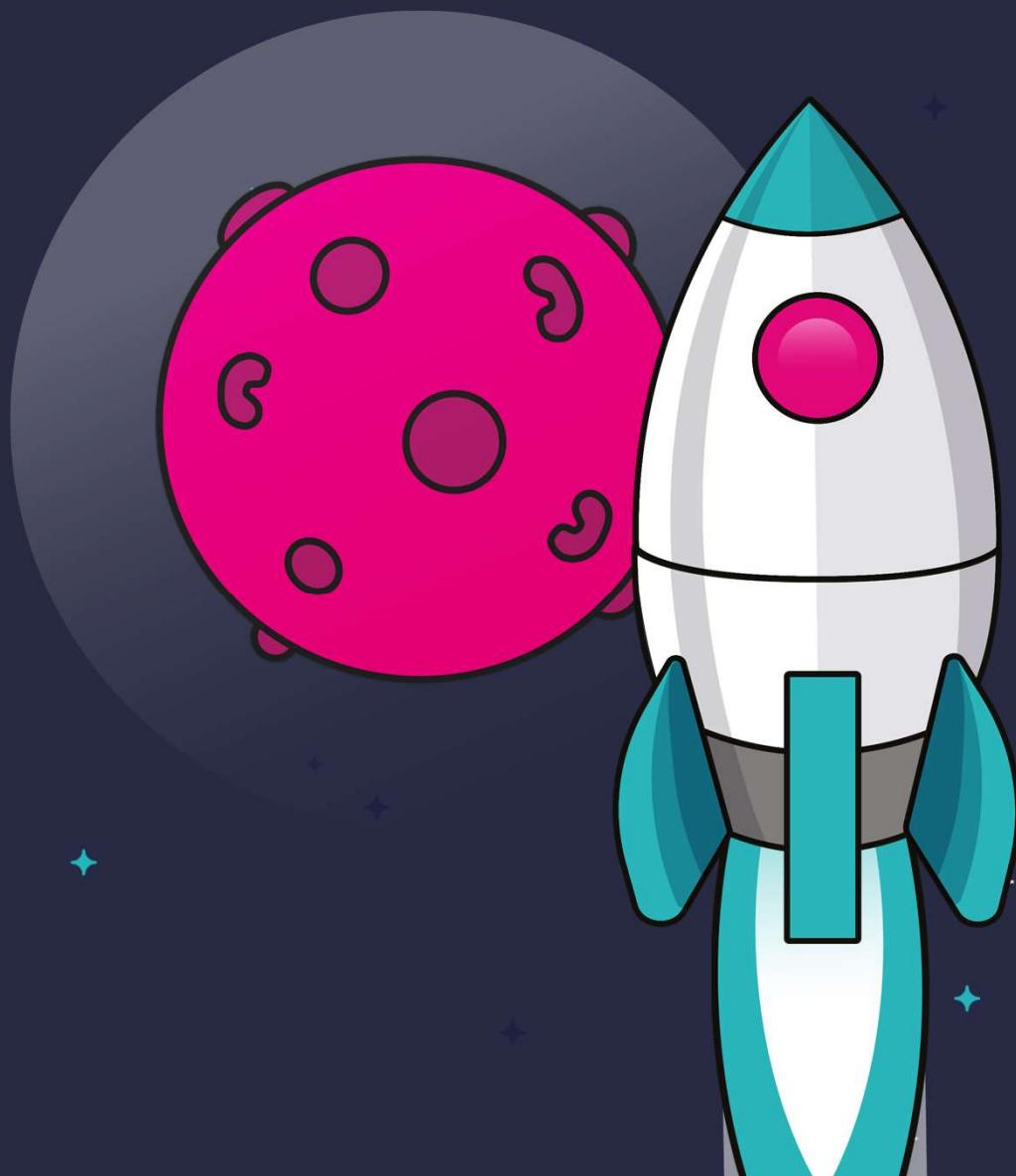
Links

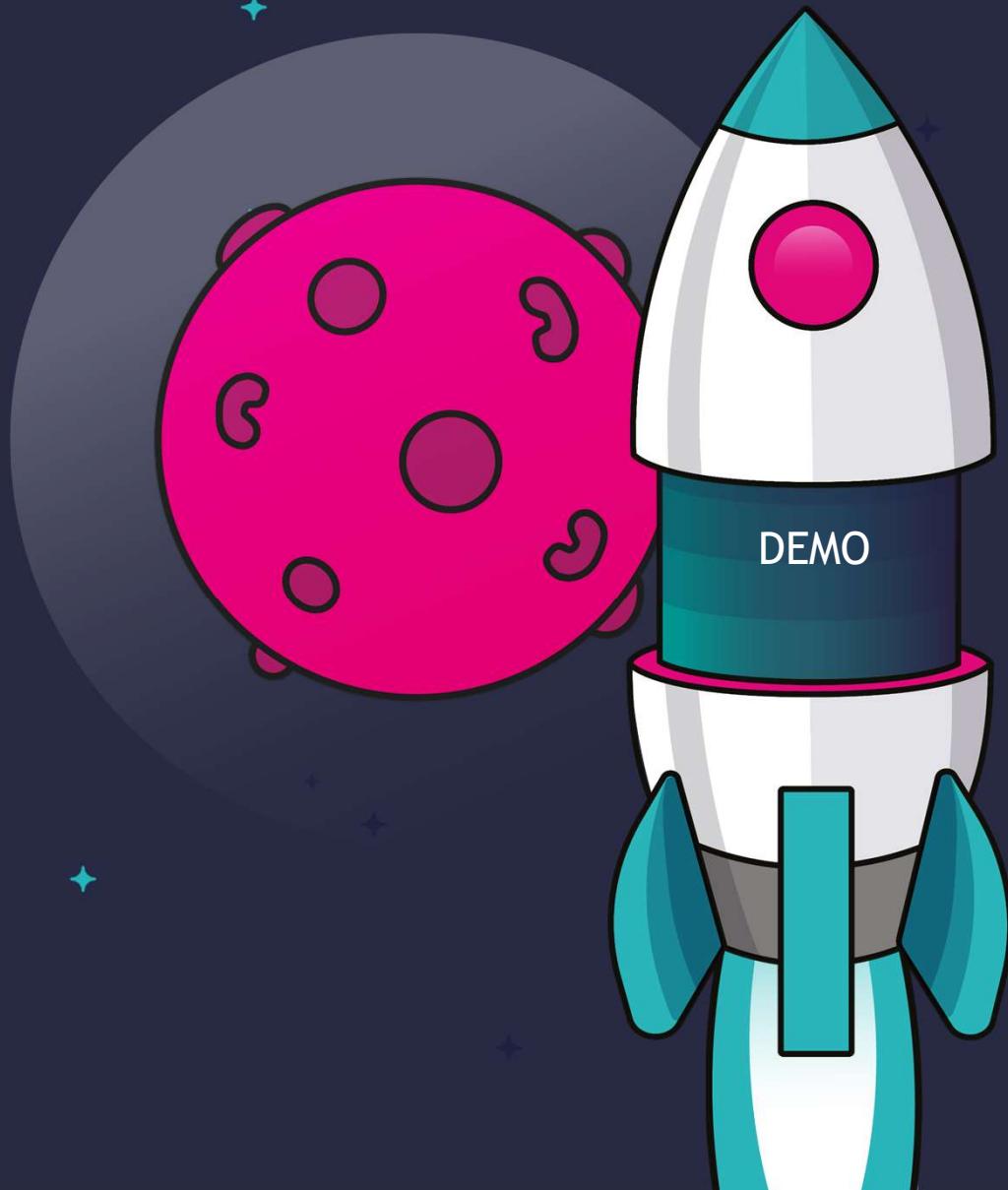
- Purview overview
- Getting started
- Documentation

Azure Purview Studio - Activity hubs



- Quick Actions, recently accessed items, owned Items, search bar and Documentation
- Manage Glossary Items, search, manage terms templates and custom attributes, import and export Terms using csv
- Create collections, register data sources, setup Scans, Integration runtime
- Insights on your data
- Meta Data Management Security, ADF and data share Connections





Innovate to accelerate

DEMO

Purview Data Map

Unify and make data meaningful

- ❑ Automated metadata scanning and lineage identification of hybrid data stores
- ❑ 100+ built-in and custom classifiers
- ❑ Microsoft Information Protection sensitivity labels



The screenshot shows the 'Sources' section of the Azure Purview Data Map. It includes a toolbar with 'Register' and 'Refresh' buttons, and tabs for 'Map view' (which is selected) and 'Table view'. A search bar is present. Below this, a message says 'Showing 8 collections, 6 sources'. The main area is a map view where data assets are represented as nodes connected by lines, indicating lineage.

Define the scope for this label

Labels can be applied directly to files, emails, containers like SharePoint sites and Teams, and more. Let us know what you would like to apply this label to so we can configure the applicable protection settings. [Learn more about label scopes](#)

Files & emails
Configure encryption and content marking settings to protect labeled emails and Office files. Also define auto-labeling rules to automatically apply this label to sensitive content in Office, files in Azure, and more.

(i) To set up auto-labeling for files in Azure, make sure you also scope this label to 'Azure Purview assets' below.

Groups & sites
Configure privacy, access control, and other settings to protect labeled Teams, Microsoft 365 Groups, and SharePoint sites.

Azure Purview assets (preview)
Apply label to assets in Azure Purview, including SQL columns, files in Azure Blob Storage, and more.

A grid of icons representing various data sources and services supported by Azure Purview:

- Azure (MULTIPLE)
- AWS account (MULTIPLE)
- Azure Synapse Analytics (MULTIPLE)
- Amazon S3
- 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 PostgreSQL
- Azure Dedicated SQL Pool (formerly SQL DW)
- Azure SQL Database
- Azure SQL Database Managed Instance

Purview Data Map

Unify and make data meaningful

- ❑ Automated metadata scanning and lineage identification of hybrid data stores
- ❑ 100+ built-in and custom classifiers
- ❑ Microsoft Information Protection sensitivity labels



Sources > LS_ADLS > Weekly_Scan run history

Weekly_Scan run history

Run scan now Edit scan Delete scan Refresh

Status	Schedule	Assets classified	Assets discovered	Start time	End time
Completed	Scheduled				
Completed	Scheduled				
Completed	Scheduled				
Completed	Manual				
Completed	Manual				

Scope your scan

Refresh

All future assets under a certain parent will be automatically selected if the parent is fully or partially checked.

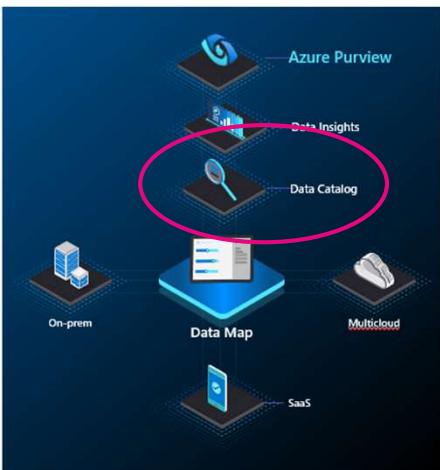
Search

<input checked="" type="checkbox"/> LS_ADLS
<input type="checkbox"/> > azure-webjobs-hosts
<input type="checkbox"/> > azure-webjobs-secrets
<input checked="" type="checkbox"/> > intermediate
<input type="checkbox"/> > marts
<input checked="" type="checkbox"/> > raw
<input type="checkbox"/> > sandbox
<input type="checkbox"/> > temp
<input checked="" type="checkbox"/> > tempstaging
<input checked="" type="checkbox"/> > transfer

Azure Purview Data Catalog

Enable effortless discovery

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



The screenshot shows the Azure Purview Demo interface. At the top, it displays '4 sources', '1,138 assets', and '1 glossary term'. A search bar shows the query 'customer'. Below the search bar, there are sections for 'Search suggestions' and 'Asset suggestions'. The 'Search suggestions' section lists: 'customer transactions original', 'customer changes', 'customer origin', and 'customer original'. The 'Asset suggestions' section lists several items under the 'Customer' category, each with a URL:

- Customer: <https://labseuwdvlmdfoxgn01.dfs.core.windows.net/intermediate/DeltaLake/AW/SalesLT/Customer/>
- Customer: <https://labseuwdvlmdfoxgn01.dfs.core.windows.net/intermediate/DeltaLake/WWI/Test/Customer/>
- Customer Profitability Sample: <https://app.powerbi.com/groups/3257d8ca-4da0-47fa-a172-e01d9b7d428a/reports/db6bc825-1e63-4c6a-9584-548f66d5c040>
- Customer Profitability Sample: <https://app.powerbi.com/groups/3257d8ca-4da0-47fa-a172-e01d9b7d428a/dashboards/b9a1cd08-3679-4192-91e3-5e6faf3ecbd>

Below the asset list, there is a 'View search results' button and a note about references, etc. To the right, there is a 'Manage glossary' section with the sub-instruction: 'Create and edit terms with custom term templates.'

Azure Purview Data Catalog

Unify and make data meaningful

The screenshot shows the Azure Purview Data Catalog interface. The top navigation bar says "Browse assets > Azure SQL Database". Below it, there's a "Refresh" button and a breadcrumb trail: "Browse assets > Azure SQL Database > WideWorldImporters".

The main area displays a list of databases under the "Name" column:

Name	Type
Adventureworks	Azure SQL Database
Datastore	Azure SQL Database
Datastore	Azure SQL Database
datastore	Azure SQL Database
oxygenconfigdb	Azure SQL Database
test_purview	Azure SQL Database
WideWorldImporters	Azure SQL Database
WideWorldImportersDW	Azure SQL Database

Under the "WideWorldImporters" database, there are several schema objects listed:

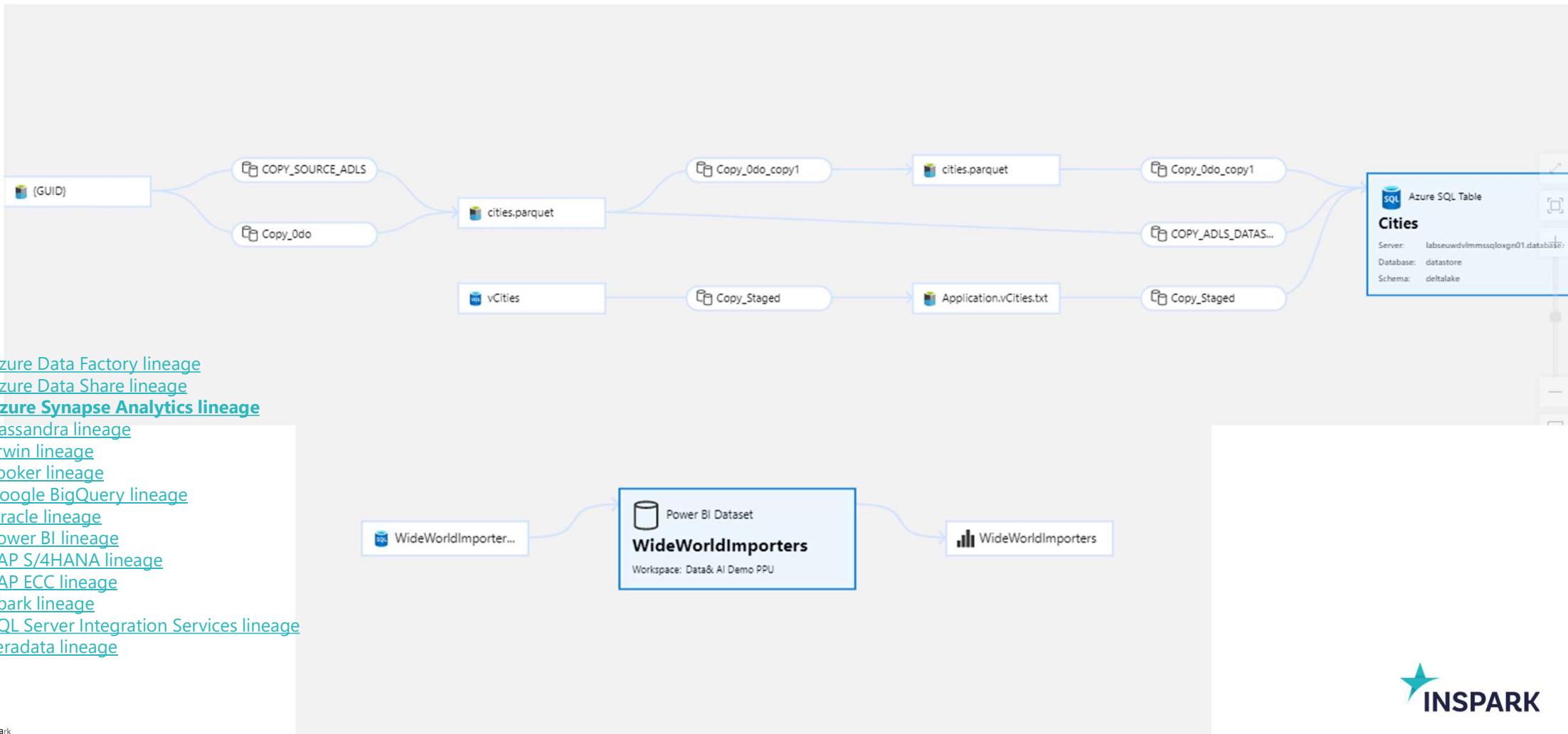
Name	Type
PurchaseOrderLines	Azure SQL Table
PurchaseOrders	Azure SQL Table
SupplierCategories	Azure SQL Table
SupplierCategories_Archive	Azure SQL Table
Suppliers	Azure SQL Table
Suppliers_Archive	Azure SQL Table
SupplierTransactions	Azure SQL Table

At the bottom, there are category cards for other services:

- Azure Data Factory: 4 items
- Azure Data Lake Storage Gen2: 1 item
- Azure SQL Database: 8 items
- Azure SQL Server: 3 items** (This card is highlighted with a teal border)
- Azure Storage Account: 1 item
- Power BI: 18 items

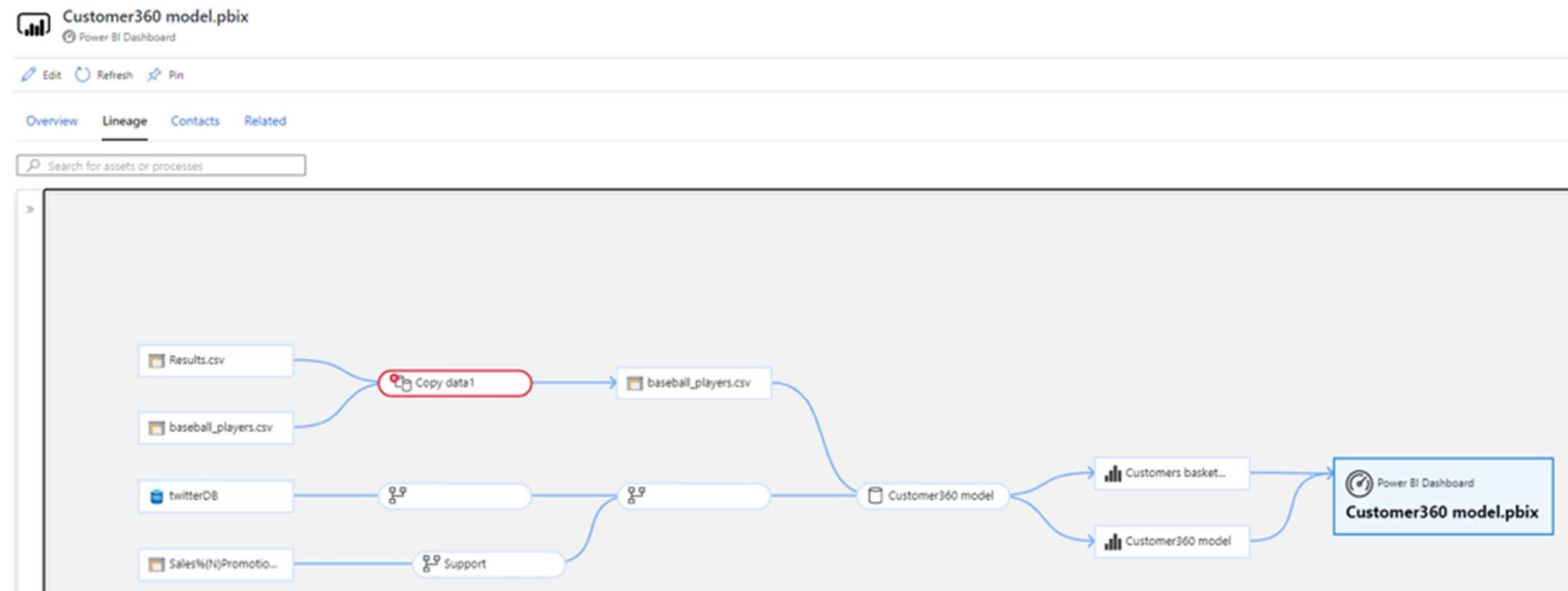
Azure Purview Data Catalog

Unify and make data meaningful



Azure Purview Data Catalog

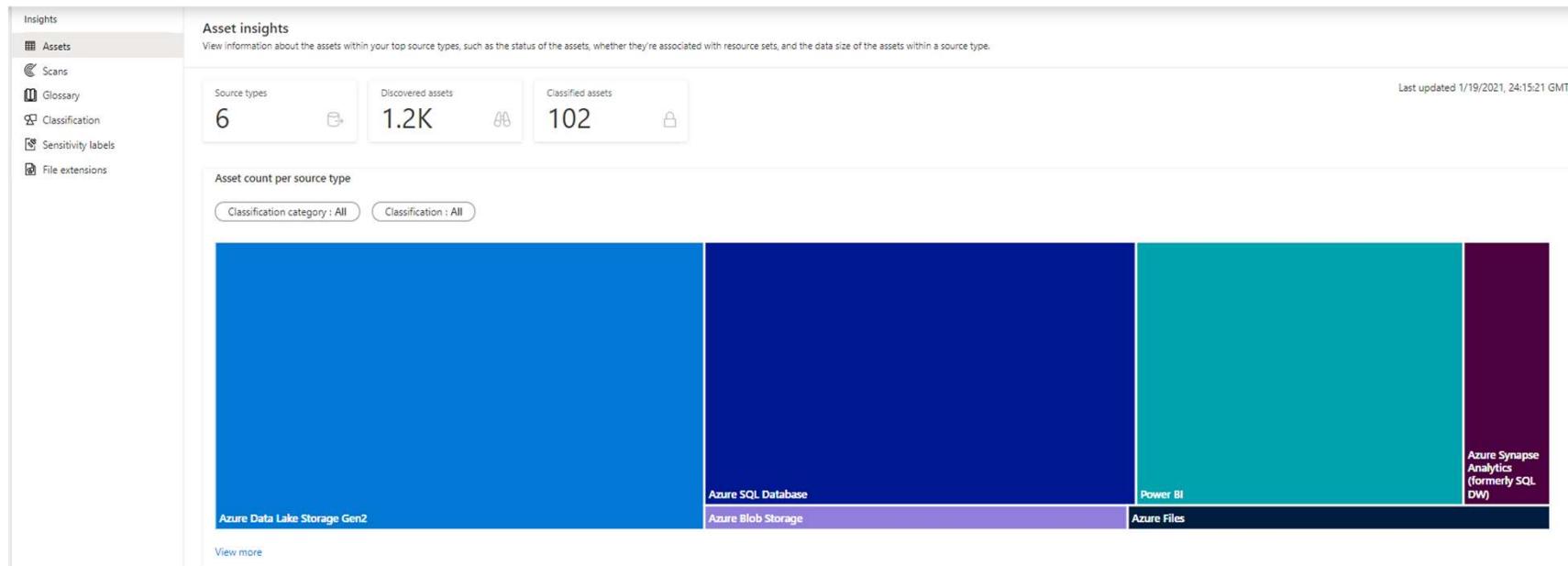
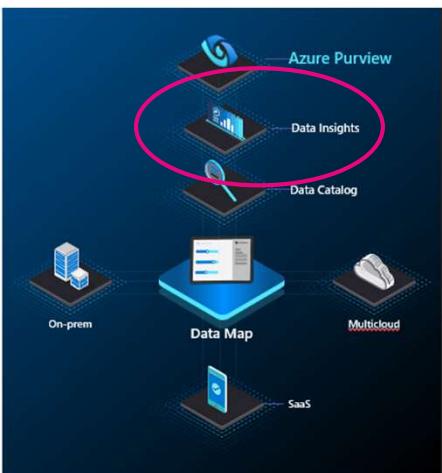
Unify and make data meaningful



Insights

Get a bird's-eye view of sensitive data

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



Register and Scan a Power BI Tenant

Discover data registered and scanned by Azure Purview

- Allow service principals to use read-only Power BI admin APIs
- Enhance admin APIs responses with detailed metadata

- Allow service principals to use read-only Power BI admin APIs
Enabled for a subset of the organization

Web apps registered in Azure Active Directory (Azure AD) will use an assigned principal to access read-only Power BI Admin APIs without a signed in user. If your 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 groups, 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 can scan emails, dataset and report detailed metadata. [Learn more](#)



Apply to:

- The entire organization
 Specific security groups

sg-oxgn-services-dvlm Enter security groups

- Enhance admin APIs responses with detailed metadata (Preview)
Enabled for a subset of the organization

Users and service principals allowed to call Power BI admin APIs may get detailed metadata about Power BI items. For example, responses from GetScanResult APIs will contain the names of dataset tables and columns. [Learn more](#)

Note: For this setting to apply to service principals, make sure the tenant setting allowing service principals to use read-only admin APIs is enabled. [Learn more](#)



Apply to:

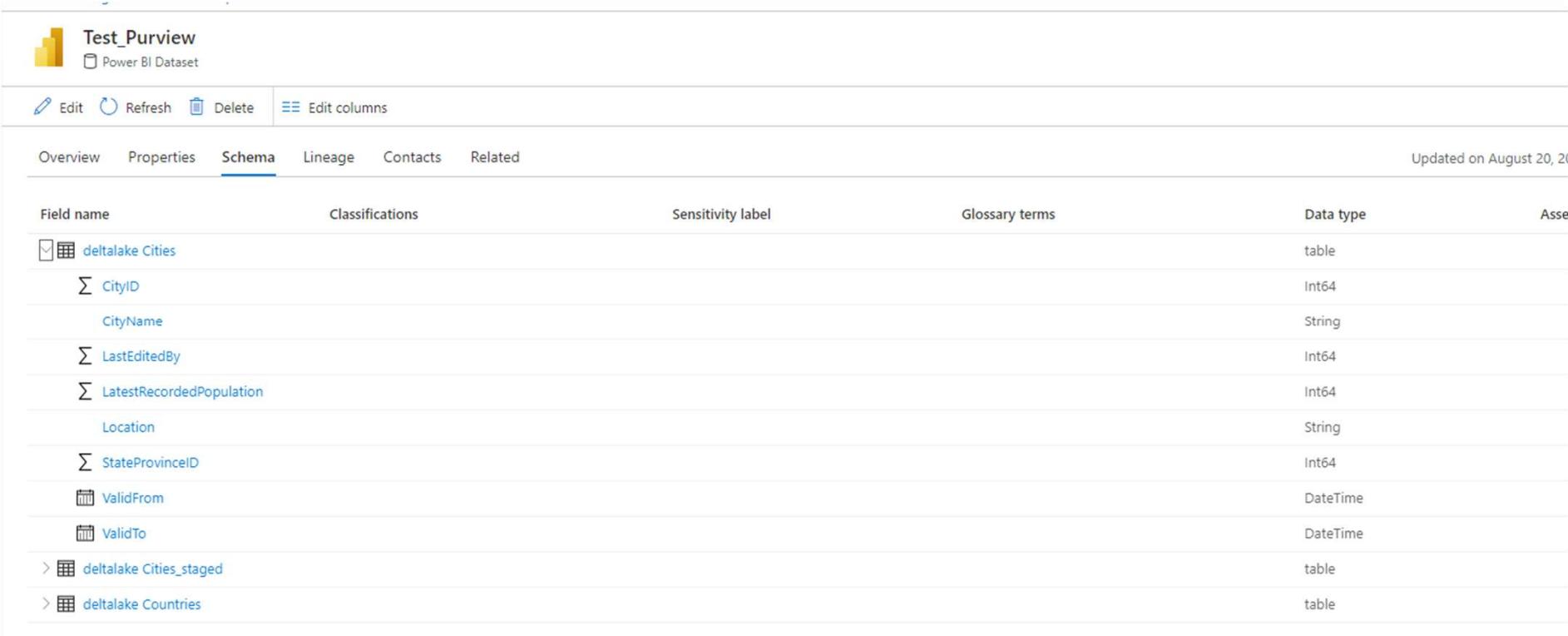
- The entire organization
 Specific security groups

sg-oxgn-services-dvlm Enter security groups

Register and Scan a Power BI Tenant

Discover data registered and scanned by Azure Purview

- Allow service principals to use read-only Power BI admin APIs
- Enhance admin APIs responses with detailed metadata



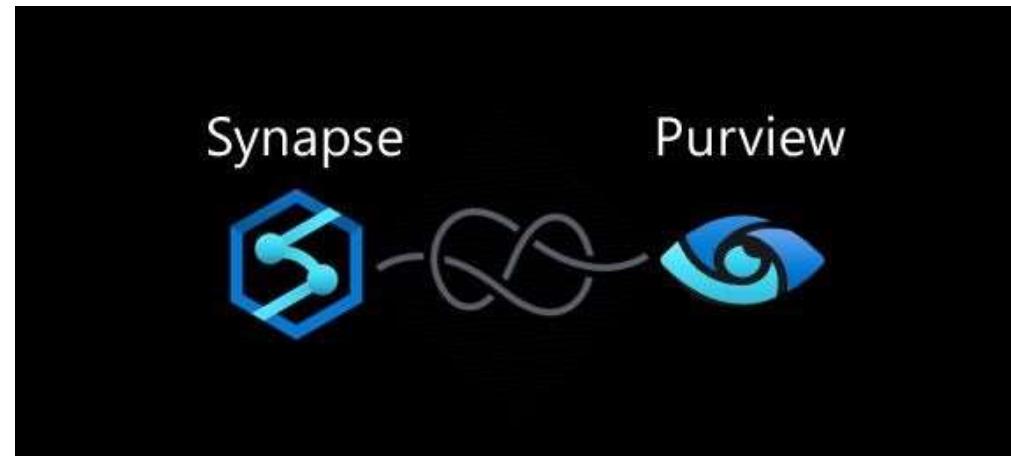
The screenshot shows the Azure Purview Data Catalog interface. At the top, there's a navigation bar with a yellow square icon, the text 'Test_Purview', and a 'Power BI Dataset' label. Below the navigation bar are buttons for 'Edit', 'Refresh', 'Delete', and 'Edit columns'. A horizontal menu bar includes 'Overview', 'Properties', 'Schema' (which is underlined), 'Lineage', 'Contacts', and 'Related'. On the right side of the menu bar, it says 'Updated on August 20, 2023'. The main area displays a table with the following columns: 'Field name', 'Classifications', 'Sensitivity label', 'Glossary terms', 'Data type', and 'Asse...'. The table lists several fields for a 'deltalake Cities' table:

Field name	Classifications	Sensitivity label	Glossary terms	Data type	Asse...
CityID				Int64	
CityName				String	
LastEditedBy				Int64	
LatestRecordedPopulation				Int64	
Location				String	
StateProvinceID				Int64	
ValidFrom				DateTime	
ValidTo				DateTime	
deltalake Cities_staged				table	
deltalake Countries				table	

Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

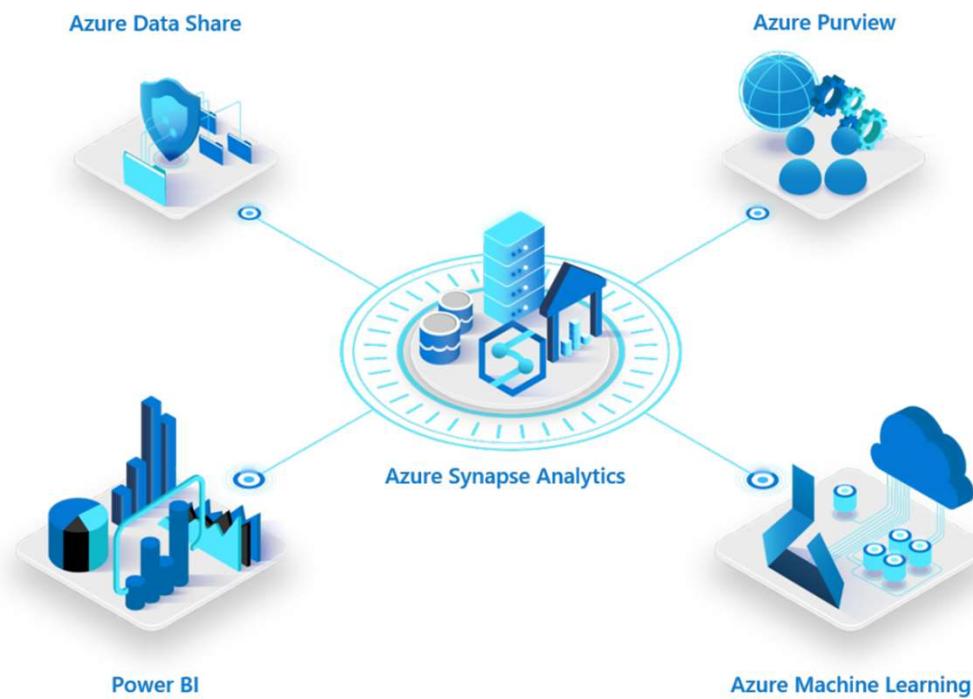
In Preview



Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

In Preview



Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

In Preview

The screenshot shows the Azure Synapse Analytics studio interface. On the left, there's a sidebar with various navigation items. The main area is divided into three cards:

- Register**: Shows a plug icon. Description: Register Synapse workspaces with Purview.
- Browse & Search**: Shows a search icon. Description: Search datasets in Synapse workspaces from linked Purview catalog.
- Enhanced metadata, lineage and annotation**: Shows a table icon. Description: View enhanced metadata, lineage from Purview for Synapse tables and add annotations.

Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

All pipeline runs > SQLServerDataIngestion - Activity runs

SQLServerDataIngestion

In Preview

List Gantt



Rerun



Rerun from activity



Rerun from failed activity



Refresh



Update pipeline



Activity runs

Pipeline run ID

All status ▾

Showing 1 - 33 of 33 items

Activity name	Activity type	Run start ↑	Duration	Status	Lineage status	Error	Log
CopyData	Copy data	8/10/21, 10:41:08 AM	00:00:44	Succeeded			
CopyData	Copy data	8/10/21, 10:41:00 AM	00:00:46	Succeeded			
CopyData	Copy data	8/10/21, 10:40:59 AM	00:00:44	Succeeded			RK

Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

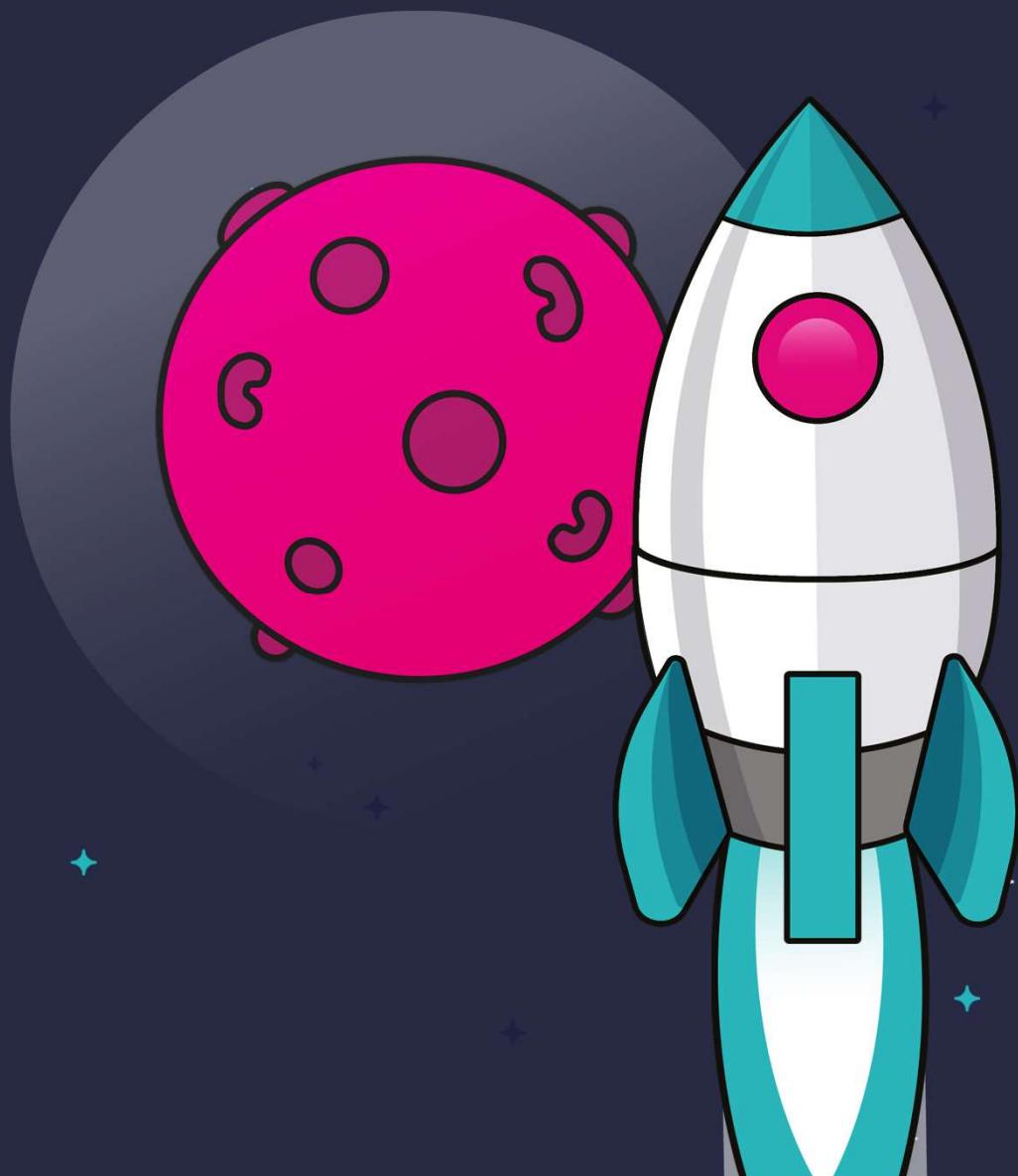
 In Preview

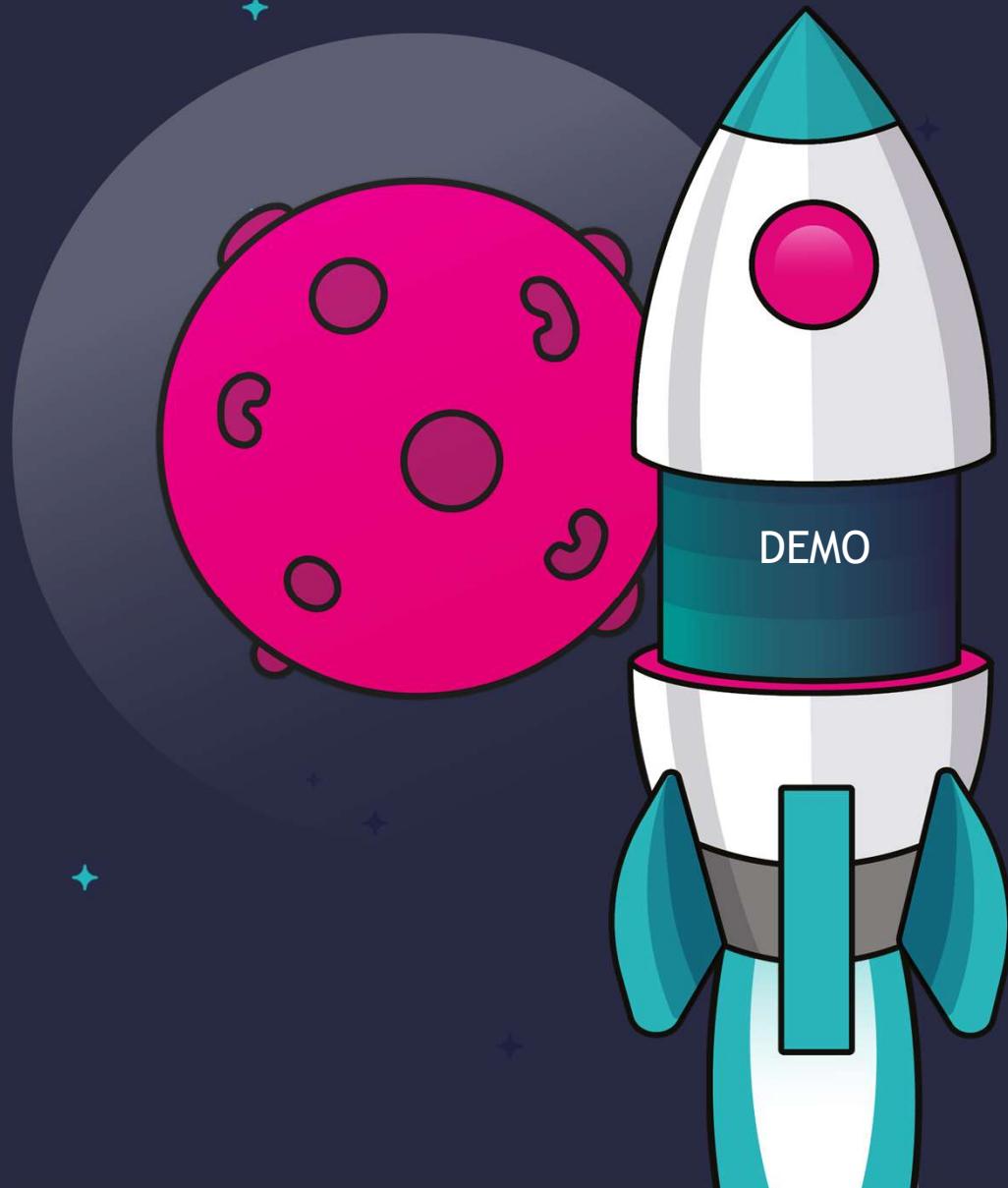


Unable to push lineage to Purview. [Learn more](#)

Azure Purview (Preview)

Connecting a Synapse workspace to Azure Purview will enable you to discover trusted and accurate data across your hybrid environment. [Learn more](#)



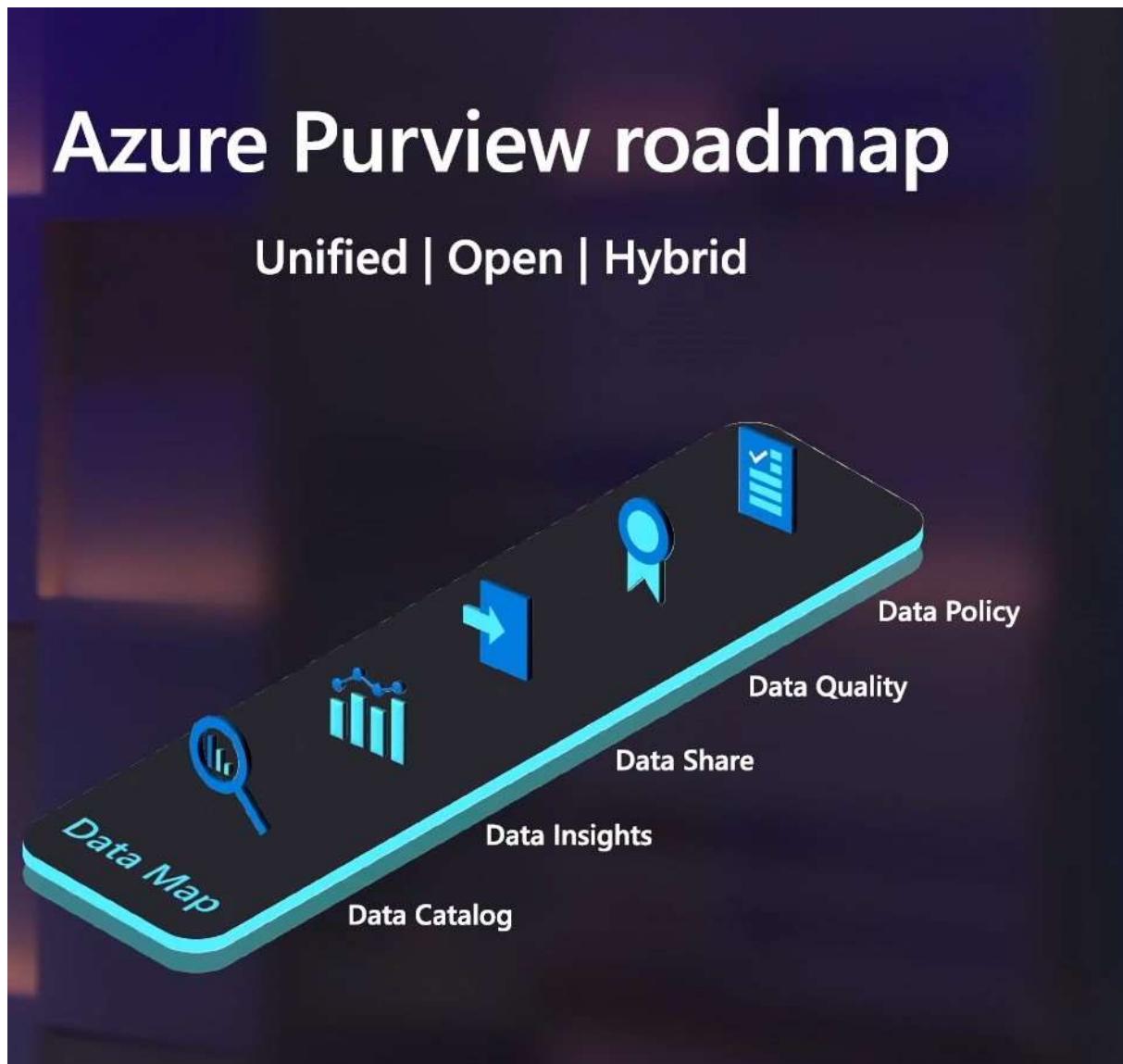


Innovate to accelerate

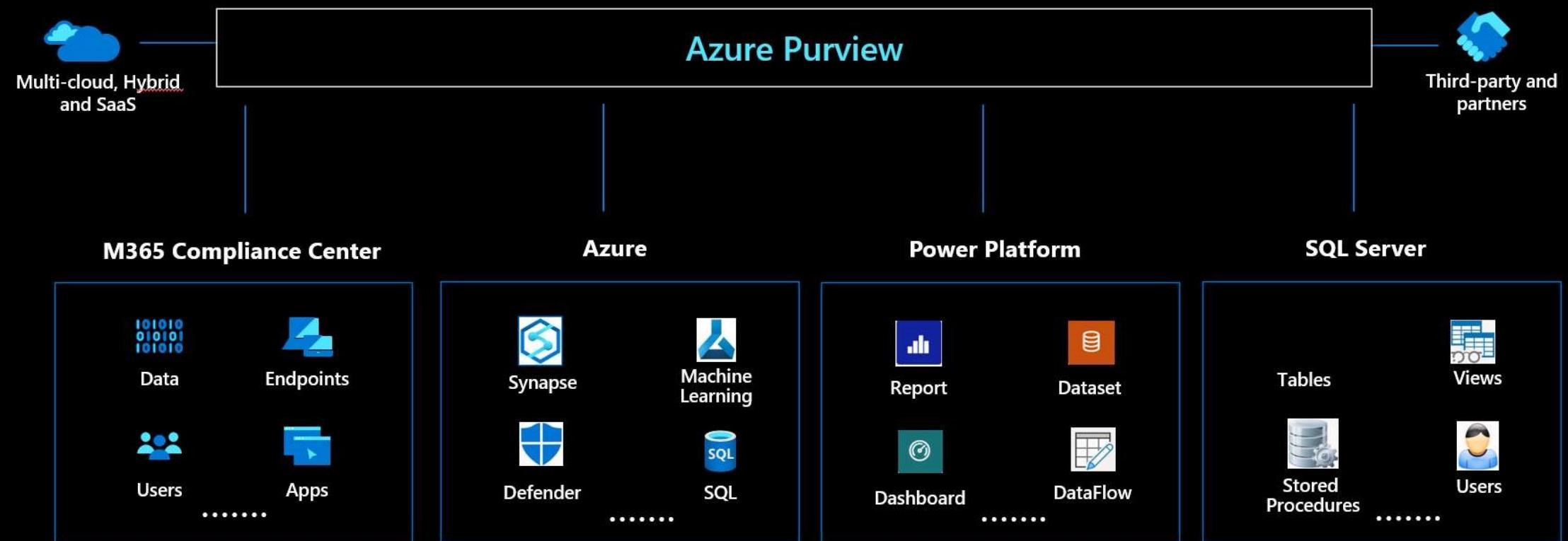
DEMO

Roadmap /Preview features

- Managed Vnets
- Data Lineage for Azure SQL
- Data Quality
- Data Privacy
- Master Data Management



Data Governance Opportunity





Maximize the Value of Your Data in the Cloud

Achieve unified data governance with Azure Purview

Digital event / September 28, 2021

<https://azuredatagovernance.eventcore.com/>

Questions?



<https://erwindekreuk.com>



@erwindekreuk



<https://www.linkedin.com/in/erwindekreuk/>

Slides will be available on my blog

