

Solve your Data  
Governance challenges  
with Microsoft Purview

Erwin de Kreuk

Principal Consultant – Lead Data & AI  
InSpark





# Erwin de Kreuk

Principal Consultant – Lead Data & AI  
InSpark

 @erwindekreuk

 [linkedin.com/in/erwindekreuk](https://linkedin.com/in/erwindekreuk)

 [erwindekreuk.com](http://erwindekreuk.com) [github.com/edkreuk](https://github.com/edkreuk)



# We Are InSpark

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



# 15 edycja konferencji SQLDay

8-10 maja 2023, WROCŁAW + ONLINE



---

partner złoty

---



---

partner srebrny

---

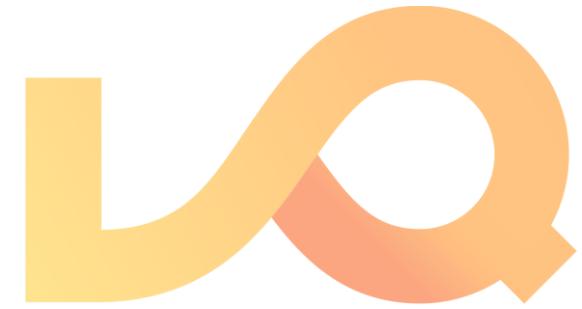


---

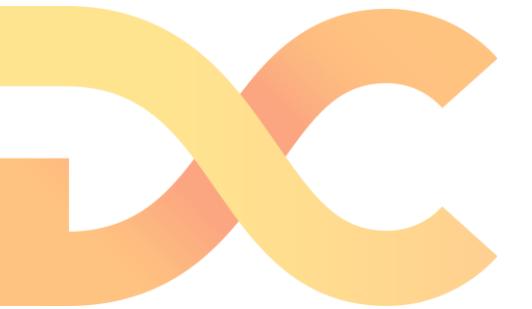
partner brązowy

---





# AGENDA



- Introduction to Microsoft Purview
- Data Map
- Microsoft Purview Apps
  - Data Catalog
  - Data Estate Insights
  - Data Use Management
  - Data Sharing
- Take aways



# Data is your most strategic asset

90%

Of corporate strategies will  
cite information as a critical  
enterprise asset by 2023

GARTNER

*Why Data and Analytics are key to digital  
transformation. Christy Pettey. Mar, 2019*

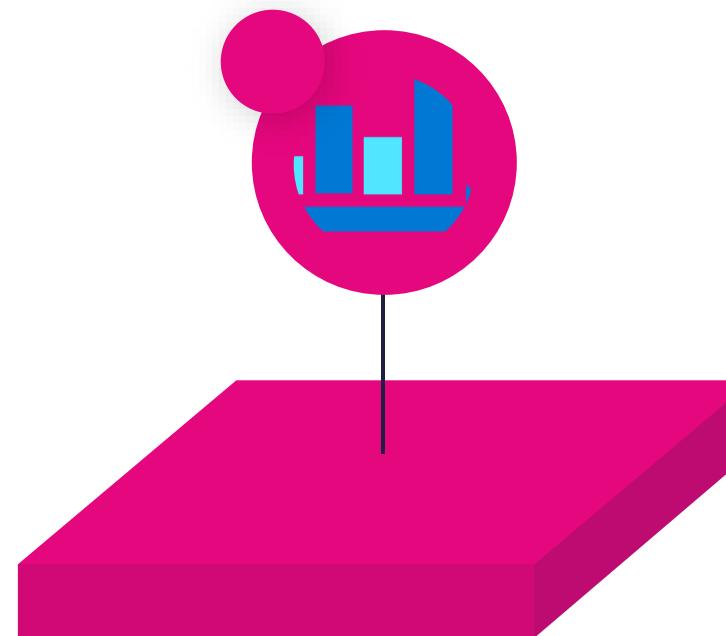
175 ZB

Expected global volume  
of data generated  
annually by 2025

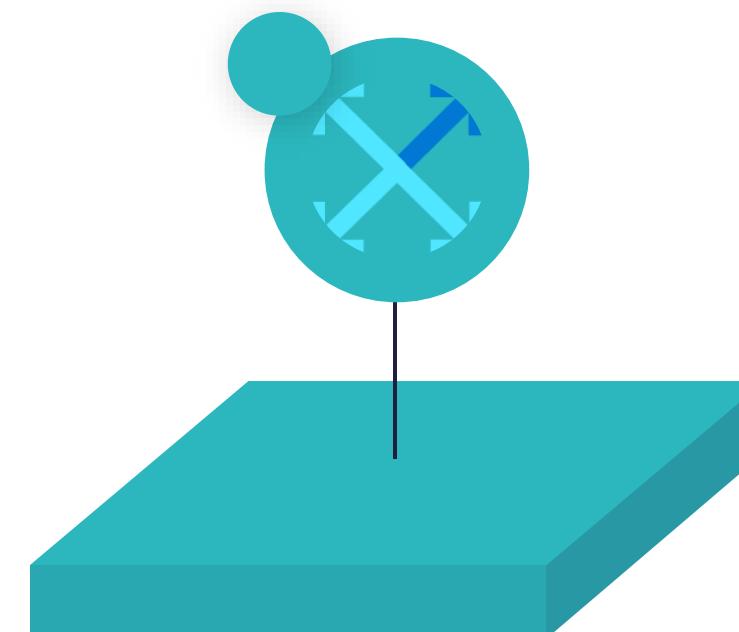
IDC

*IDC Data Age 2025, Dave Reinsel,*

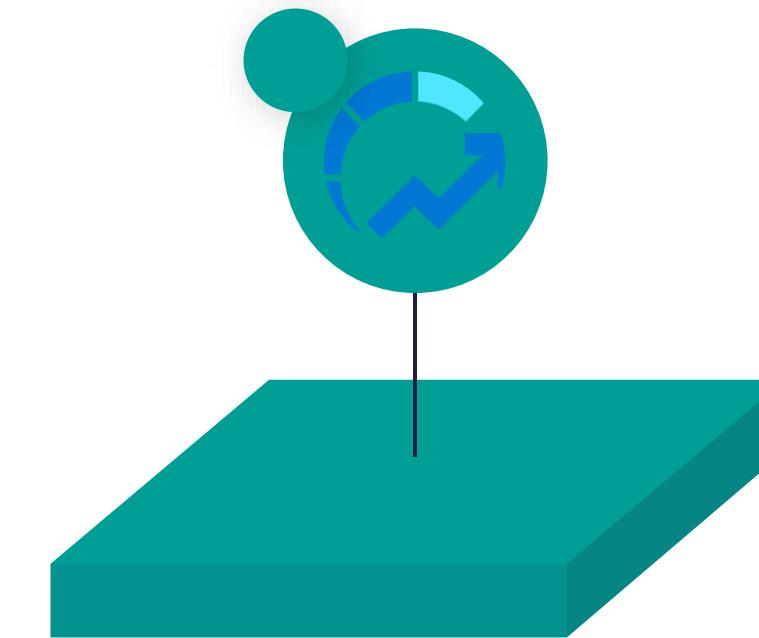
# Elements of successful data governance



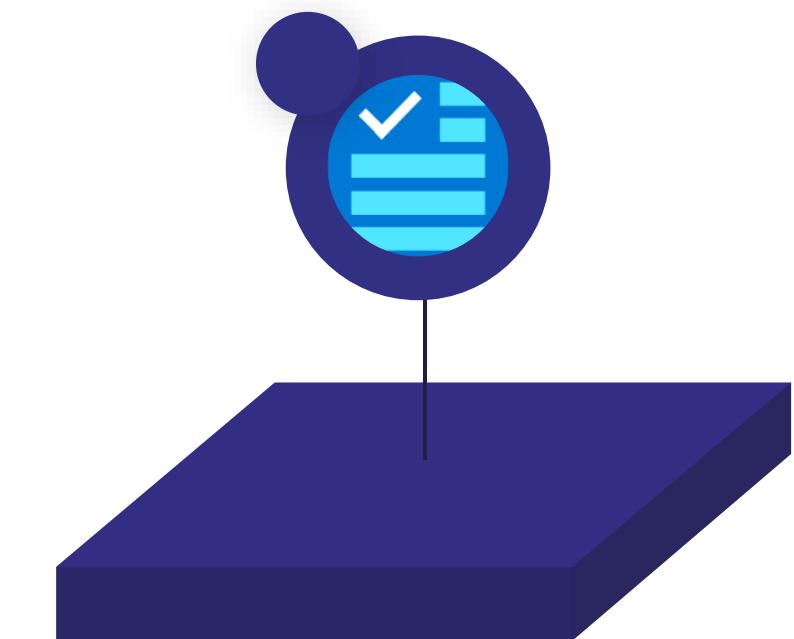
Manage growing  
data landscape



Overcome  
operational  
silos



Increase data  
agility



Comply with  
industry regulations

# Microsoft Purview

Unified data governance to  
maximize the business  
value of data

Reimagine **data governance** in the cloud

Set the foundation for  
effective **data governance**

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

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



# Data operation & governance is becoming increasingly interdisciplinary



# Challenges for Data Consumers

- There's no central location to register data sources.
- Data-consumption experiences require users to know the connection string or path.
- Data sources and documentation might live in several places.
- There's no explicit connection between the data and the experts that understand the data's context.

# Challenges for Data Producers

- Annotating data sources with descriptive metadata is often a lost effort. Client applications typically ignore descriptions that are stored in the data source.
- Creating documentation for data sources can be difficult and it's an ongoing responsibility to keep documentation in sync with data sources. Users might not trust documentation that's perceived as being out of date.
- Creating and maintaining documentation for data sources is complex and time-consuming.
- Restricting access to data sources and ensuring that data consumers know how to request access is an ongoing challenge.

# Challenges for Security Administrators

- Everything above and:
- An organization's data is constantly growing and being stored and shared in new directions. The task of discovering, protecting, and governing sensitive data is one that never ends.
- How to ensure that the organization's content is being shared with the correct people, applications, and with the correct permissions.
- Understanding the risk levels in an organization's data requires diving deep into the content, looking for keywords, RegEx patterns, and sensitive data types.
- Constantly monitor all data sources for sensitive content, as even the smallest amount of data loss can be critical to your organization.
- Ensuring that an organization continues to comply with corporate security policies is a challenging task as the content grows and changes.

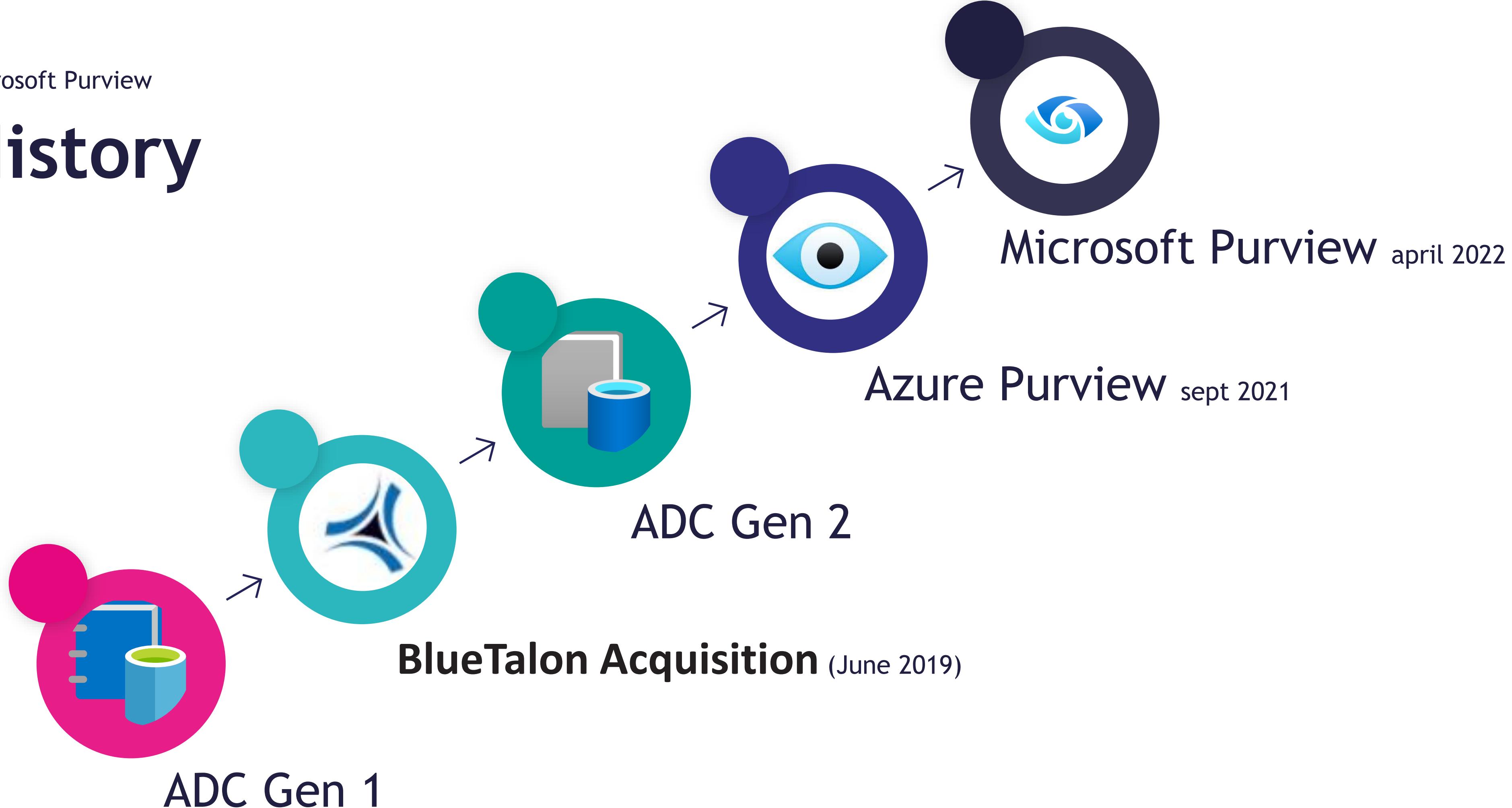
Unified data governance to  
maximize the business value of data

## Microsoft Purview



Microsoft Purview

# History



# Microsoft Purview

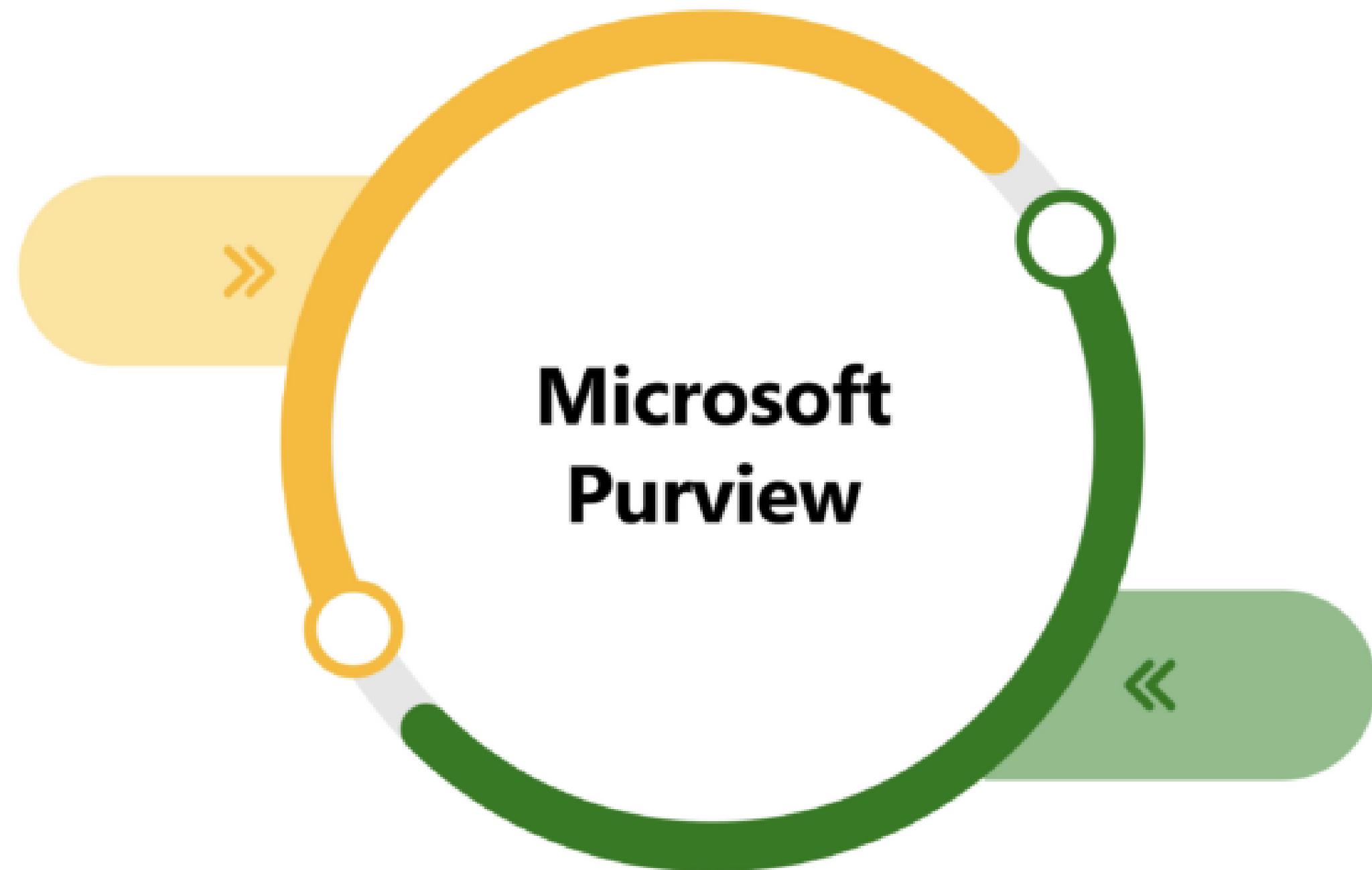
The future of compliance and data governance

## Risk & compliance

For risk, compliance, and legal teams

Former Name	New Name
Microsoft 365 Advanced Audit	Microsoft Purview Audit (Premium)
Microsoft 365 Basic Audit	Microsoft Purview Audit (Standard)
Microsoft 365 Communication Compliance	Microsoft Purview Communication Compliance
Microsoft Compliance Manager	Microsoft Purview Compliance Manager

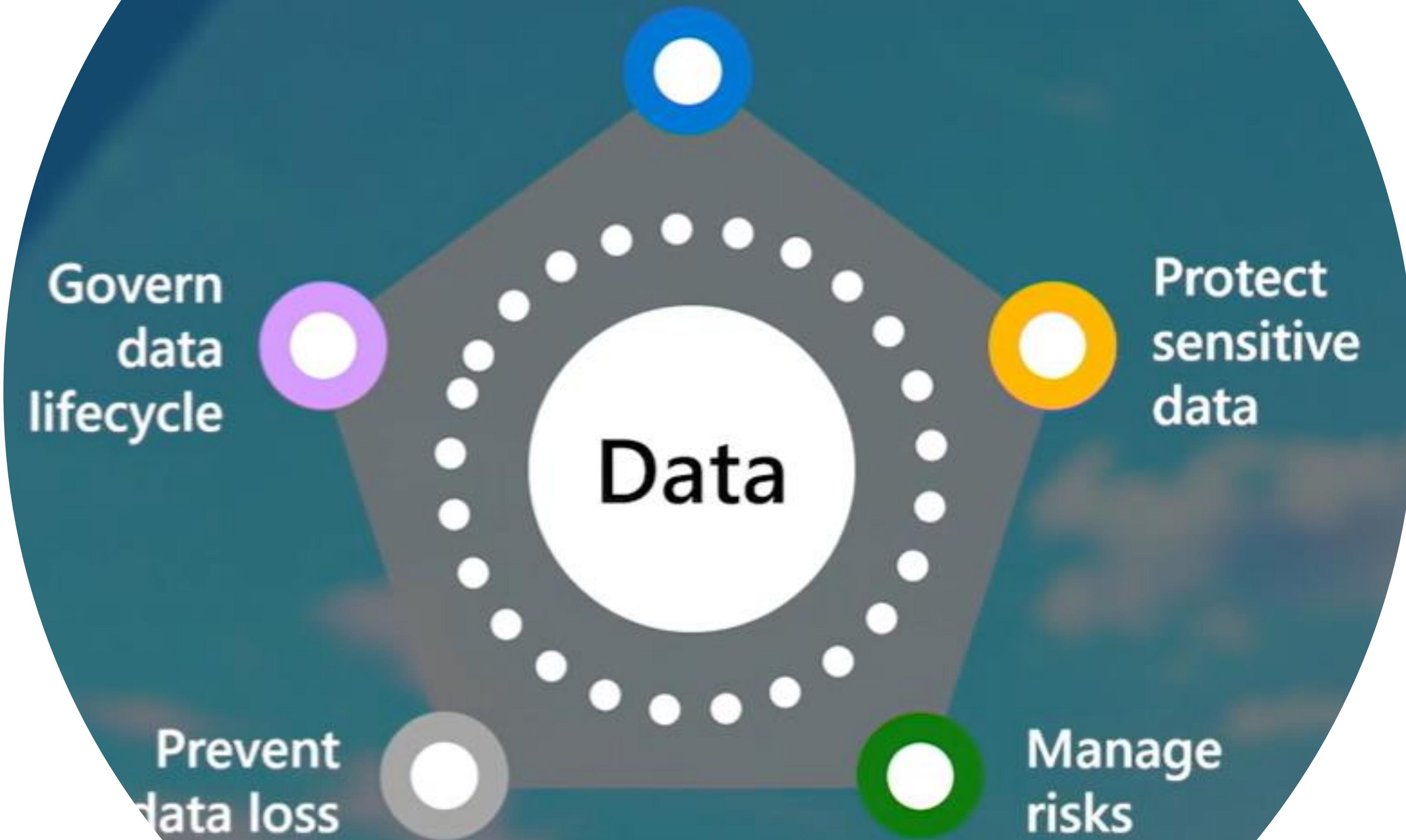
## Microsoft Purview



## Unified data governance

For data consumers, data engineers, data officers

# Identify data landscape



# Microsoft Purview

Generally Available

Preview

Unified Data Governance



Microsoft Purview



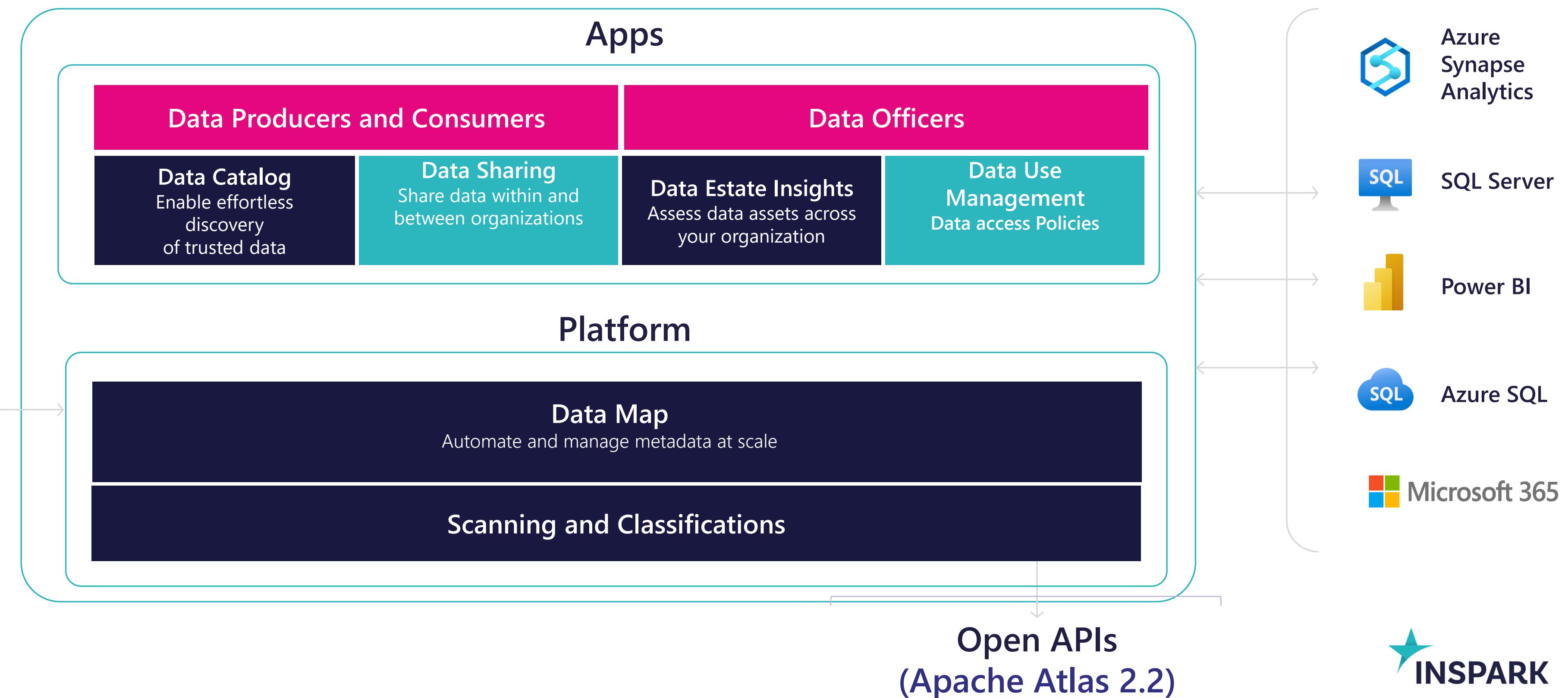
On-prem



Cloud



SaaS  
Applications



## Microsoft Purview

# Purview Studio

The screenshot shows the Microsoft Purview InSpark studio interface. Several UI elements are highlighted with callout boxes:

- Metrics**: A teal callout box pointing to the Metrics icon in the left sidebar.
- Search Bar**: A teal callout box pointing to the search bar at the top of the left sidebar.
- Recently Accessed Entities**: A teal callout box pointing to the "Recently accessed" section in the left sidebar.
- Updates**: A pink callout box pointing to the Updates icon in the top right corner of the header.
- Accounts**: A pink callout box pointing to the Accounts icon in the top right corner of the header.
- Notifications**: A pink callout box pointing to the Notifications icon in the top right corner of the header.
- Feedback**: A pink callout box pointing to the Feedback icon in the top right corner of the header.
- Key Activities**: A teal callout box pointing to the "Key activities" section in the bottom right corner.
- Usefull Links**: A teal callout box pointing to the "Links" section in the bottom right corner.

**Microsoft Purview InSpark**  
(labseuwvdvImpurviewoxgn01)

18 sources | 1,819 assets | 56 glossary terms

Search catalog

Browse assets | Manage glossary | Knowledge center

Recently accessed

Name	Last update
Product	16 days ago
Product	3 days ago
Copy_DeltaLake_Purview	2 days ago
Address	16 days ago
PL_EXECUTE_COPY_ADLS_TO_DELTA LAKE PURVIEWy2	---
Copy_DeltaLake_Purview	---

View all recently accessed

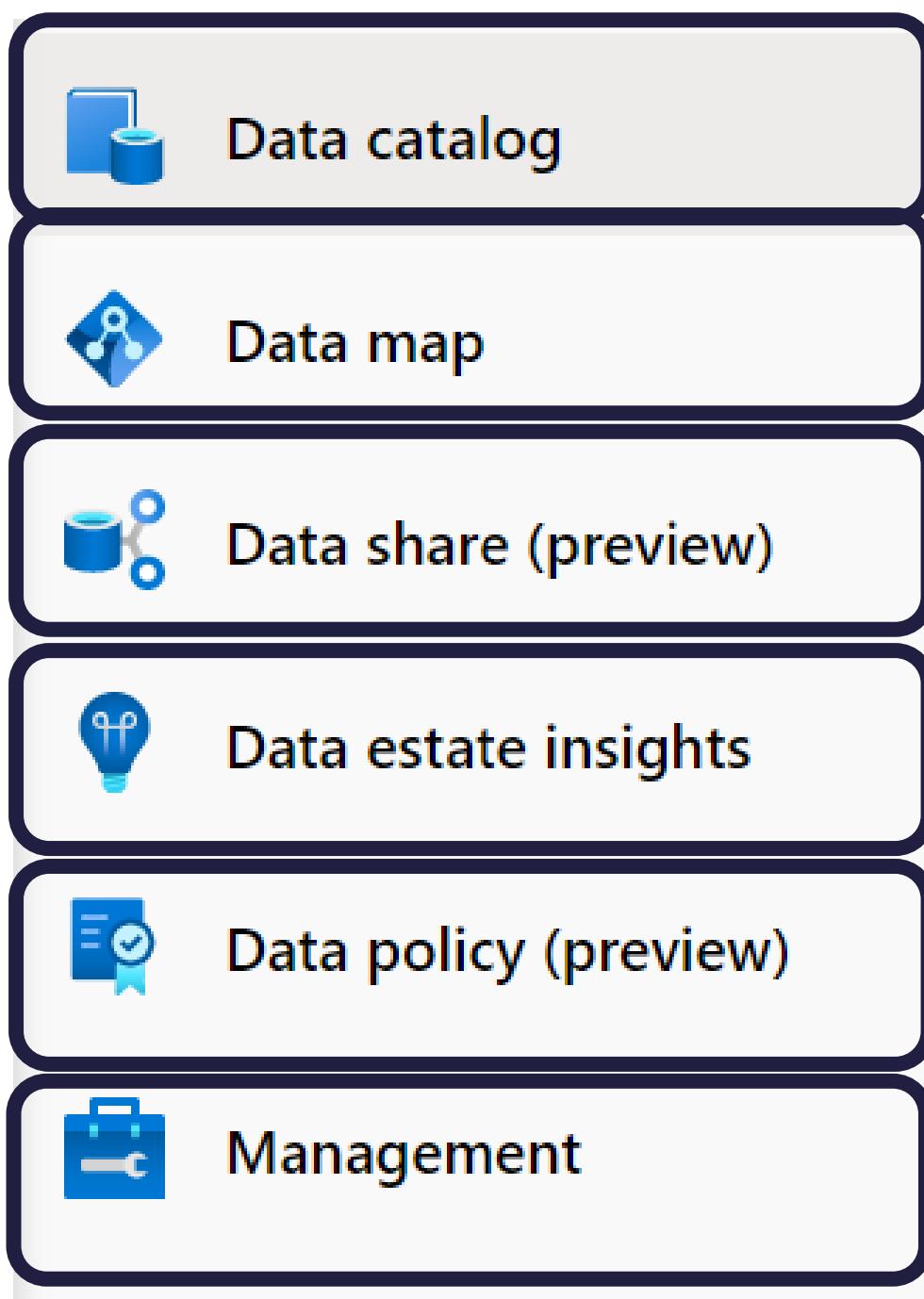
Links

- [Microsoft Purview overview](#)
- [Discover the capabilities offered by Microsoft Purview and learn how to make the most of them.](#)
- [Getting started](#)
- [Learn about how to use Microsoft Purview Governance Portal.](#)
- [Documentation](#)
- [Visit the documentation center for quickstarts, how-to guides, and references, etc.](#)

**INSPARK**

# Purview Studio - Activity hubs

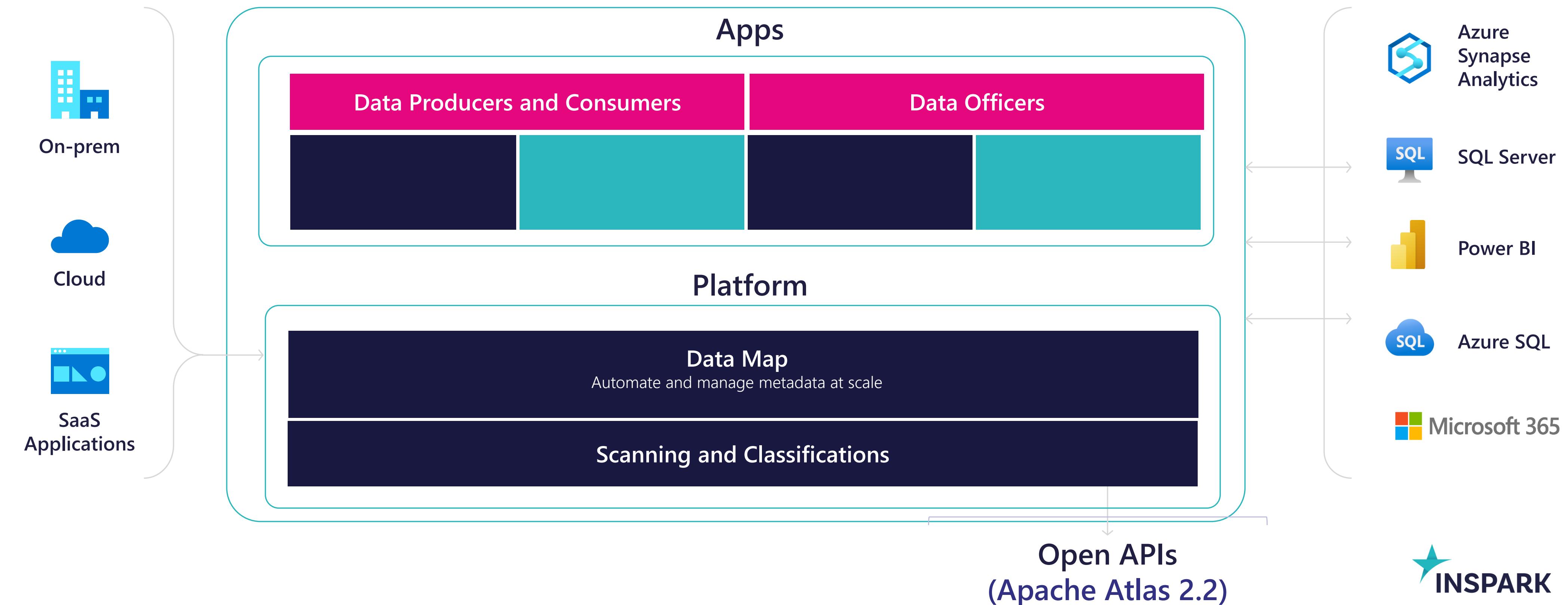
- Quick Actions, recently accessed items, owned Items, search bar and Documentation.
- Manage Glossary Items, search, manage terms templates and custom attributes, import and export Terms using csv.
- Create collections, register data sources, setup Scans, Integration runtime.
- Easily share data between organizations within the Microsoft Purview governance portal.
- Insights on your data Estate.
- Manage access to different data systems across your entire data estate.
- Meta Data Management, Security, Workflows, Managed private endpoints, ADF and data share Connections. Enable Feature options.



# Microsoft Purview

Generally Available  
Preview

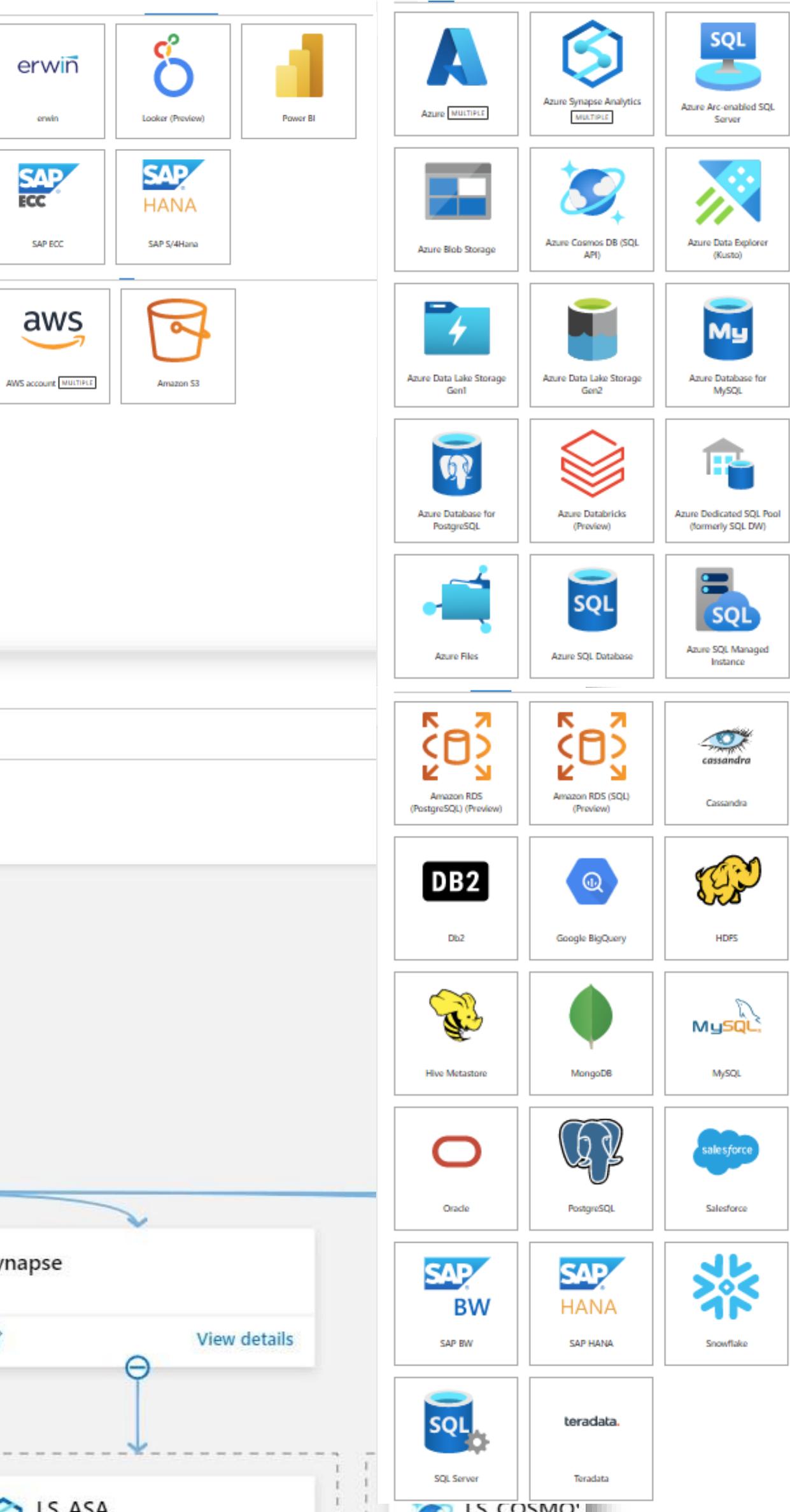
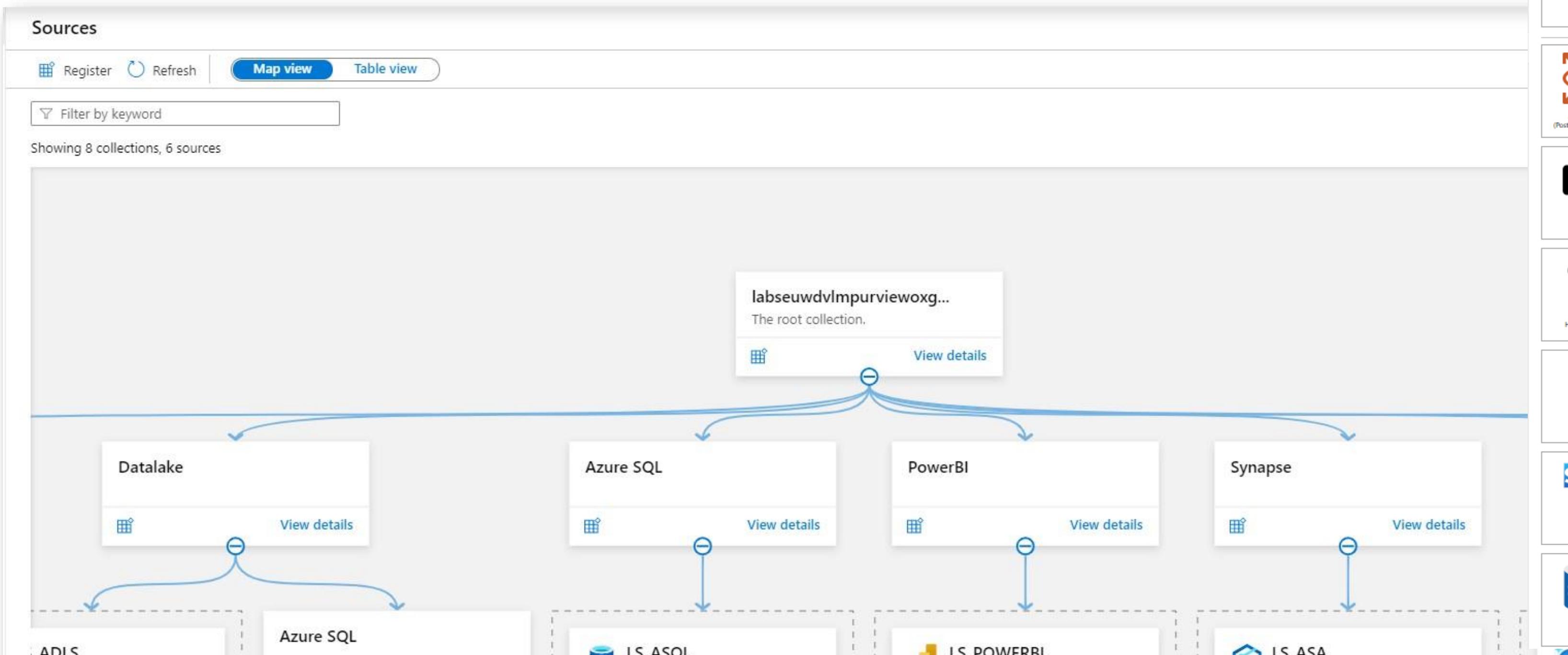
Unified Data Governance



## Data Map

# Collections and Sources

- A graph describing the data assets and their relationships across your data estate
- Support for 39+ data sources and growing



## Data Map

# Scans

- A graph describing the data assets and their relationships across your data estate
- Support for 39+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores

## Scope your scan

Refresh

Add new assets under partially selected assets to future scans (Preview)  On

Future scans will automatically include any new assets under the assets that you select.  
[Learn more](#)

New assets under partially selected assets will now be added to future scans automatically. If you don't want this, turn off the above toggle. [Learn more](#)

Search

- > LS\_ADLS
- > intermediate
- > marts
- > raw
- > tmp

## Set a scan trigger

Set a scan trigger to run the scan at specific dates and times. If once, the scan will start after set up is completed. If recurring, the scan will start at a date and time you choose. The initial scan is a full scan and every subsequent scan is incremental.

Recurring  Once

Time zone \*

(UTC+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna

This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Recurrence \*

Every  Month(s)

Month days  Week days

Select day of the month to scan

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	Last			

Schedule scan time (UTC+1)

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	2	432,119	01/18/21 11:00 AM	01/18/21 11:00 AM
Completed	Scheduled	172	432,117	01/11/21 11:00 AM	01/11/21 11:00 AM
Completed	Scheduled	199	431,156	01/04/21 11:00 AM	01/04/21 11:00 AM
Completed	Manual	171	429,874	12/30/20 02:47 PM	12/30/20 02:47 PM
Completed	Manual	192	429,496	12/29/20 02:57 PM	12/29/20 02:57 PM

## Data Map

# Classifications

- A graph describing the data assets and their relationships across your data estate
- Support for 39+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores
- 220+ built-in data classifiers

**Classifications**

+ New Edit Delete Refresh

System Custom

These are the system provided classifications.

Filter by name...

Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA R
All Full Names	MICROSOFT.PERSONAL.NAME	All Fu
All Physical Addresses	MICROSOFT.PERSONAL.PHYSICALADDRESS	All Ph
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGENТИNA.DNI_NU...	Argen
Australia Bank Account Number	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACCOU...	Austra
Australia Business Number	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Austra
Australia Company Number	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Austra
Australia Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...	Austra
Australia Medical Account Number	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL...	Austra
Australia Passport Number	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...	Austra
Australia Tax File Number	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE...	Austra
Austria Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LIC...	Austra

**Cities**  
Azure Data Lake Storage Gen2 Resource Set

Edit Select for bulk edit Request access Refresh Delete Edit columns

Overview Properties Schema Lineage Contacts Related

Filter by name

Showing 17 of 17 items

Column name	Classifications	Sensitivity label	Glossary terms
CityID	 World Cities		
CityName			
LastEditedBy			
LatestRecordedPopulation			
Location			

# Microsoft Purview

Generally Available  
Preview

Unified Data Governance



Apps

Data Producers and Consumers

**Data Catalog**  
Enable effortless discovery of trusted data

Data Officers

Platform

Data Map

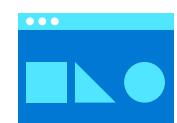
Automate and manage metadata at scale

Scanning and Classifications

On-prem



Cloud



SaaS Applications



Azure Synapse Analytics



SQL Server



Power BI



Azure SQL



Microsoft 365

Open APIs  
(Apache Atlas 2.2)



## Data Catalog

# Search

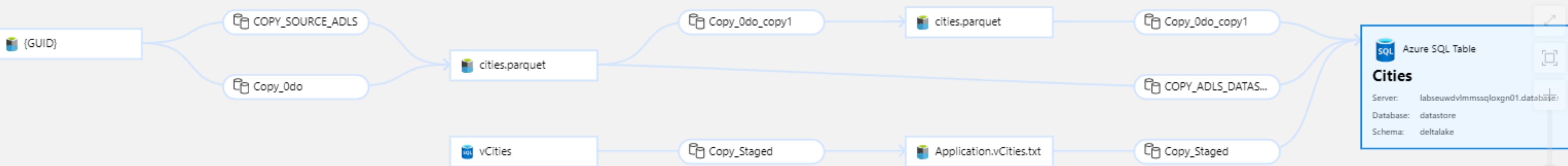
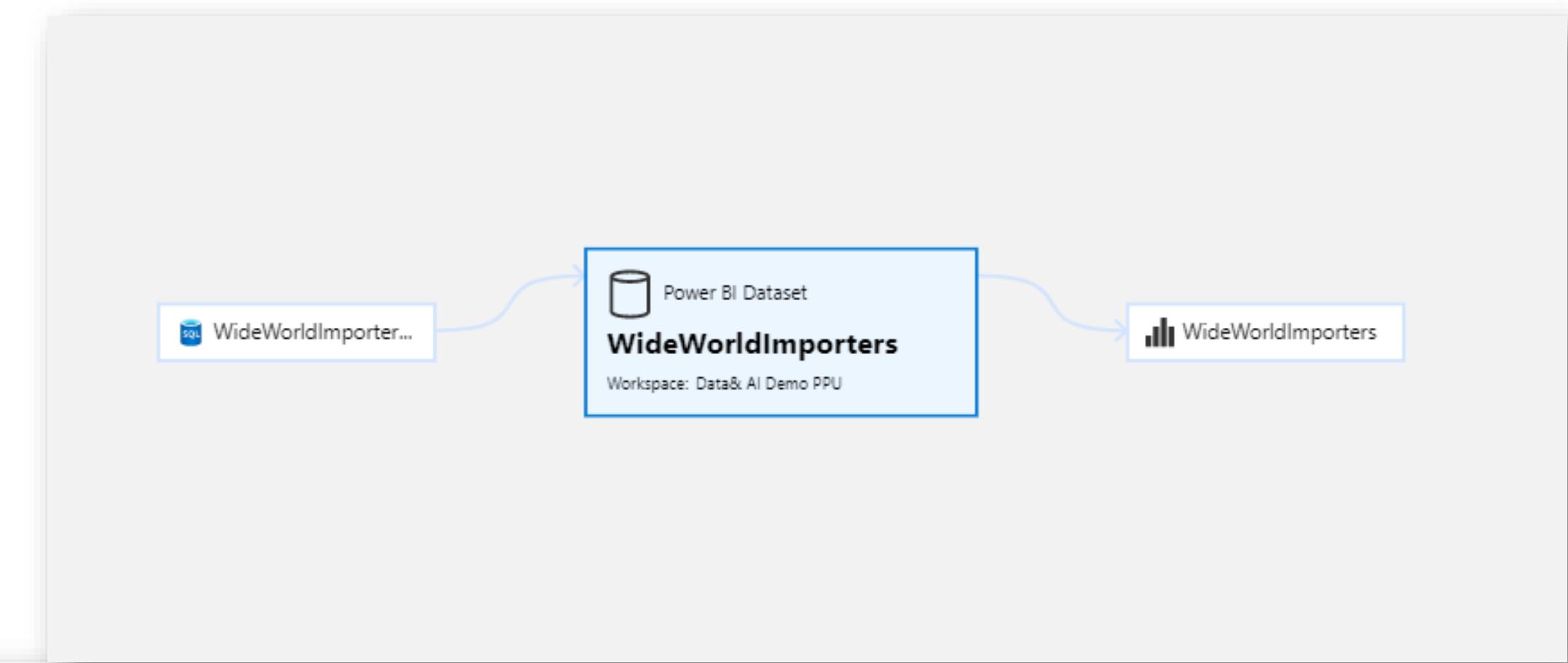
- Semantic search and browse
- Business glossary and workflows

The screenshot shows the Microsoft Purview InSpark Data Catalog interface. At the top right, it displays "Microsoft Purview InSpark" and "(labseuwvdylmpurviewoxgn01)" with statistics: 17 sources, 1,950 assets, and 56 glossary terms. A search bar contains the query "customer". Below the search bar, sections include "Your recent searches" (with entries for "customer" and "customers"), "Search suggestions" (CustomerCategories, Customer, CustomerKey, CustomerCategoryID), and "Asset suggestions" (Customer\_Customer\_DataFlow, CustomerType\_CustomerType\_DataFlow, Customer, Customer). On the left, a sidebar lists navigation options: Data catalog, Data map, Data estate insights (preview), Data policy (preview), and Management. The main content area shows a grid of data assets, with one asset titled "Customer" highlighted.

## Data Catalog

# Lineage

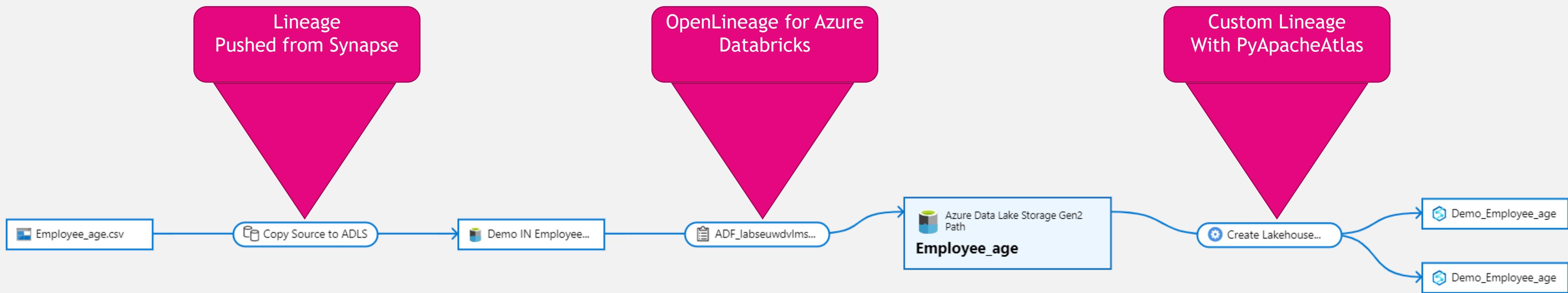
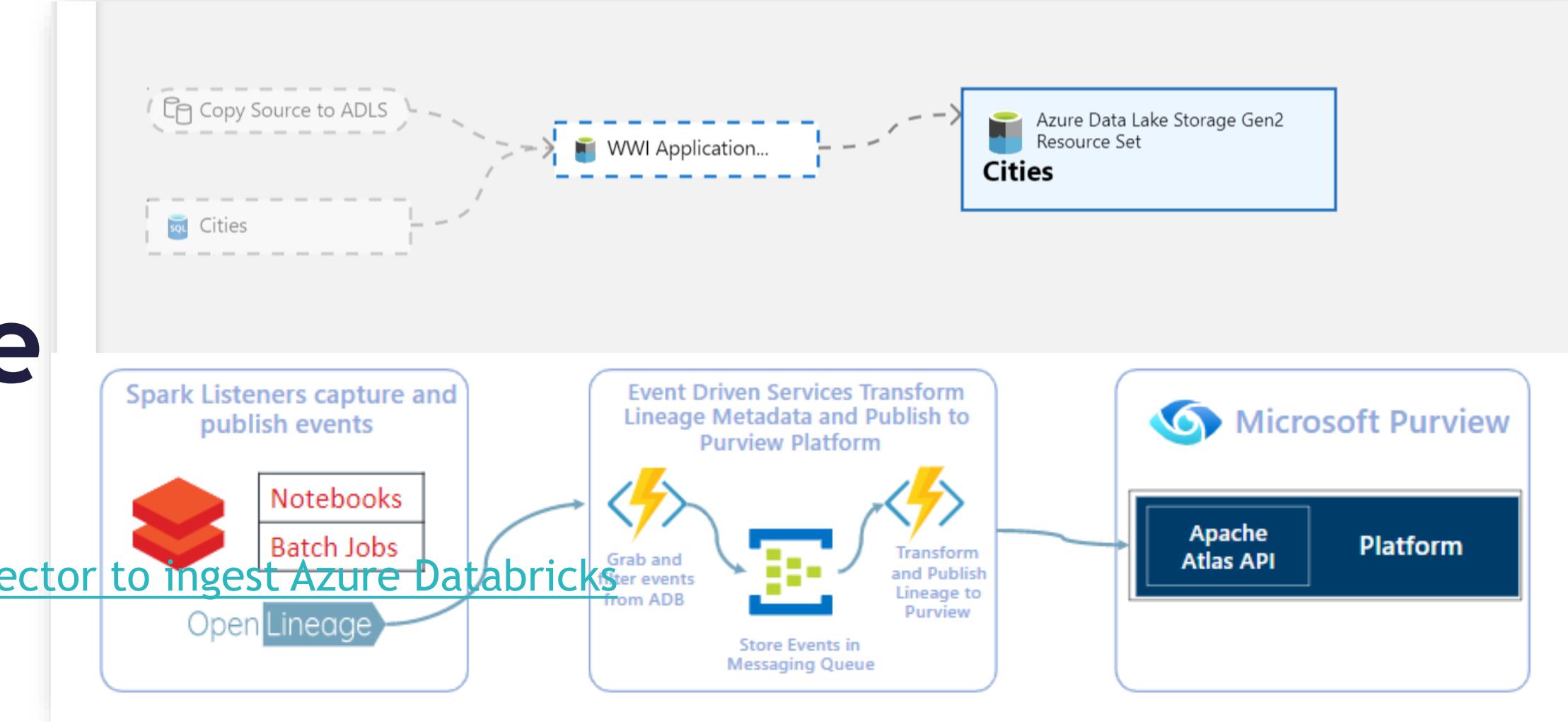
- Semantic search and browse
- Business glossary and workflows
- Data lineage with sources, owners, transformations, and lifecycle



Data Catalog

# Custom/Manual Lineage

- OpenLineage for Azure DataBricks  
[microsoft/Purview-ADB-Lineage-Solution-Accelerator: A connector to ingest Azure Databricks lineage into Microsoft Purview \(github.com\)](https://github.com/microsoft/Purview-ADB-Lineage-Solution-Accelerator)
- Create custom Lineage with Apache Spark 2.2  
<https://github.com/wjohnson/pyapacheatlas>



# Microsoft Purview

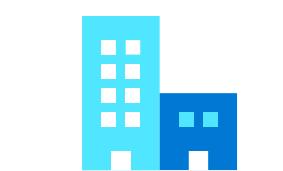
Generally Available

Preview

Unified Data Governance



Microsoft Purview



On-prem



Cloud



SaaS  
Applications

Apps

Data Producers and Consumers

**Data Catalog**  
Enable effortless discovery of trusted data

Data Officers

**Data Estate Insights**  
Assess data assets across your organization

Platform

Data Map

Automate and manage metadata at scale

Scanning and Classifications



Azure  
Synapse  
Analytics



SQL Server



Power BI



Azure SQL



Microsoft 365

Open APIs  
(Apache Atlas 2.2)

INSPARK

## Data Estate Insight

# Metrics

- Data Stewardship
- Assets
- Glossary
- Classification
- Sensitivity labels

**Data Estate Insights** On Automatic 7/18/2022, 9:48 PM Provides insight into your data estate.

Full report Report generated on July 1, 2022 at 1:10 PM

**Data stewardship**  
View key metrics about your data estate's health and performance.

**Asset curation** 29K Total assets  
Fully curated Partially curated Not curated

**Asset data ownership** 29K No owner  
Owner assigned No owner

**Catalog usage and adoption**  
Monthly active users 0 ↓100% in last month

Show:  Data estate  Catalog adoption

**Data estate health**  
Collection : (Root) Microsoft Purview InSpark

Collection	Assets	With sensitive classifications	Fully curated	Owner assigned
{Assets}	27,431	✓ 0%	✗ 0%	✗ 0%
PowerBI	153	✓ 0%	✗ 0%	✗ 0%
Synapse	116	✓ 78%	✗ 0%	✗ 0%
Cosmos	5	✓ 0%	✗ 0%	✗ 0%
DemoEK	330	✓ 12%	✗ 0%	✗ 0%
Blob	19	✓ 0%	✗ 0%	✗ 0%
Datalake	476	✓ 13%	✗ 0%	✗ 0%
Azure SQL	599	✓ 32%	✗ 0%	✗ 0%

**Assets**  
View key metrics about your data estate's assets.

**Total assets** 29K Microsoft Purview InSpark

**Asset classification** 29K No classification  
Applied Not applied

**Asset data ownership** 29K No owner  
Assigned Not assigned

**New assets (Last 30 days)** 81 New  
New Existing

**Deleted (Last 30 days)** 2 trash

...

# ACCESS CONTROL

Data Owner  
Dev Ops  
Self Service access



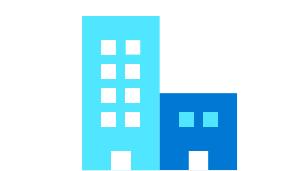
# Microsoft Purview

Generally Available  
Preview

Unified Data Governance



Microsoft Purview



On-prem



Cloud



SaaS  
Applications

Apps

Data Producers and Consumers

**Data Catalog**  
Enable effortless discovery of trusted data

Data Officers

**Data Estate Insights**  
Assess data assets across your organization

**Data Use Management**  
Data access Policies

Platform

**Data Map**

Automate and manage metadata at scale

Scanning and Classifications



Azure  
Synapse  
Analytics



SQL Server



Power BI



Azure SQL



Microsoft 365

Open APIs  
(Apache Atlas 2.2)

**INSPARK**

# Data Owner policies

## Currently in Preview

- Azure Storage
- Azure SQL
- Azure Arc Enabled
- Resource Groups
- Subscription

**Data resources**  
Select a type of data resource to apply your policy statement to.

 **Data sources**  
Systems where data is stored. Sources can be hosted in various places such as a cloud or on-premises.  
**Data source type \***

 **Assets**  
Individual objects stored within the data catalog.  
**Data source type \***  
  
 Azure Blob Storage  
 Azure Data Lake Storage Gen2  
 Azure SQL Database

# Dev Ops policies

## Manage Access to System Metadata

### SQL Performance Monitoring SQL Security Auditing

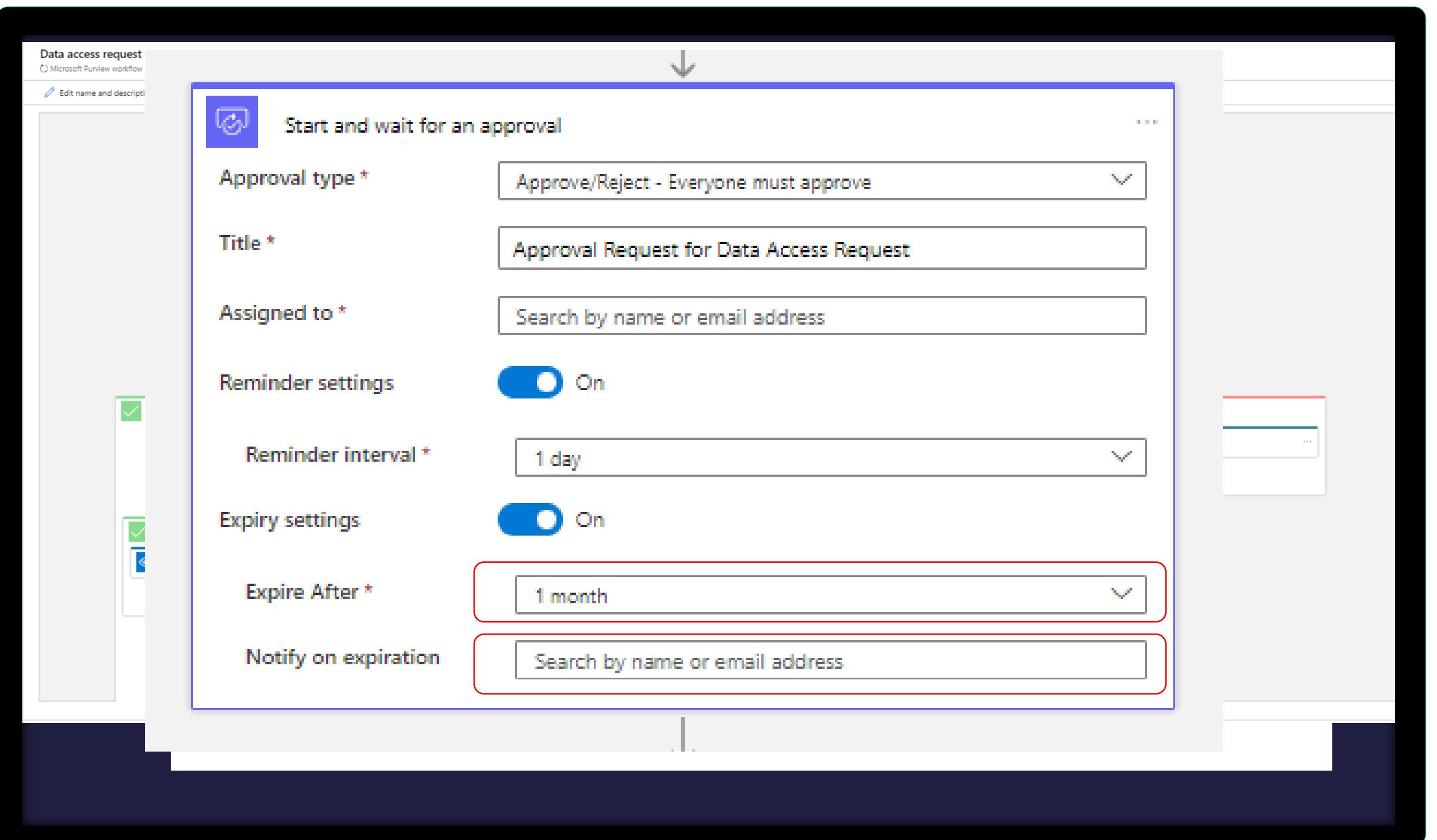
The screenshot shows the SSMS Object Explorer with the 'WideWorldImportersDW' database selected. The 'INFORMATION\_SCHEMA' node under 'Tables' is expanded, displaying a list of system views:

- INFORMATION\_SCHEMA.CHECK\_CONSTRAINTS
- INFORMATION\_SCHEMA.COLUMN\_DOMAIN\_USAGE
- INFORMATION\_SCHEMA.COLUMN\_PRIVILEGES
- COLUMNS
- CONSTRAINT\_COLUMN\_USAGE
- CONSTRAINT\_TABLE\_USAGE
- DOMAIN\_CONSTRAINTS
- DOMAINS
- KEY\_COLUMN\_USAGE
- PARAMETERS
- REFERENTIAL\_CONSTRAINTS
- ROUTINE\_COLUMNS
- ROUTINES
- SCHEMATA
- SEQUENCES
- TABLE\_CONSTRAINTS
- TABLE\_PRIVILEGES
- TABLES
- VIEW\_COLUMN\_USAGE
- VIEW\_TABLE\_USAGE
- VIEWS

# Self-service access policies

## Data Access (Governance)

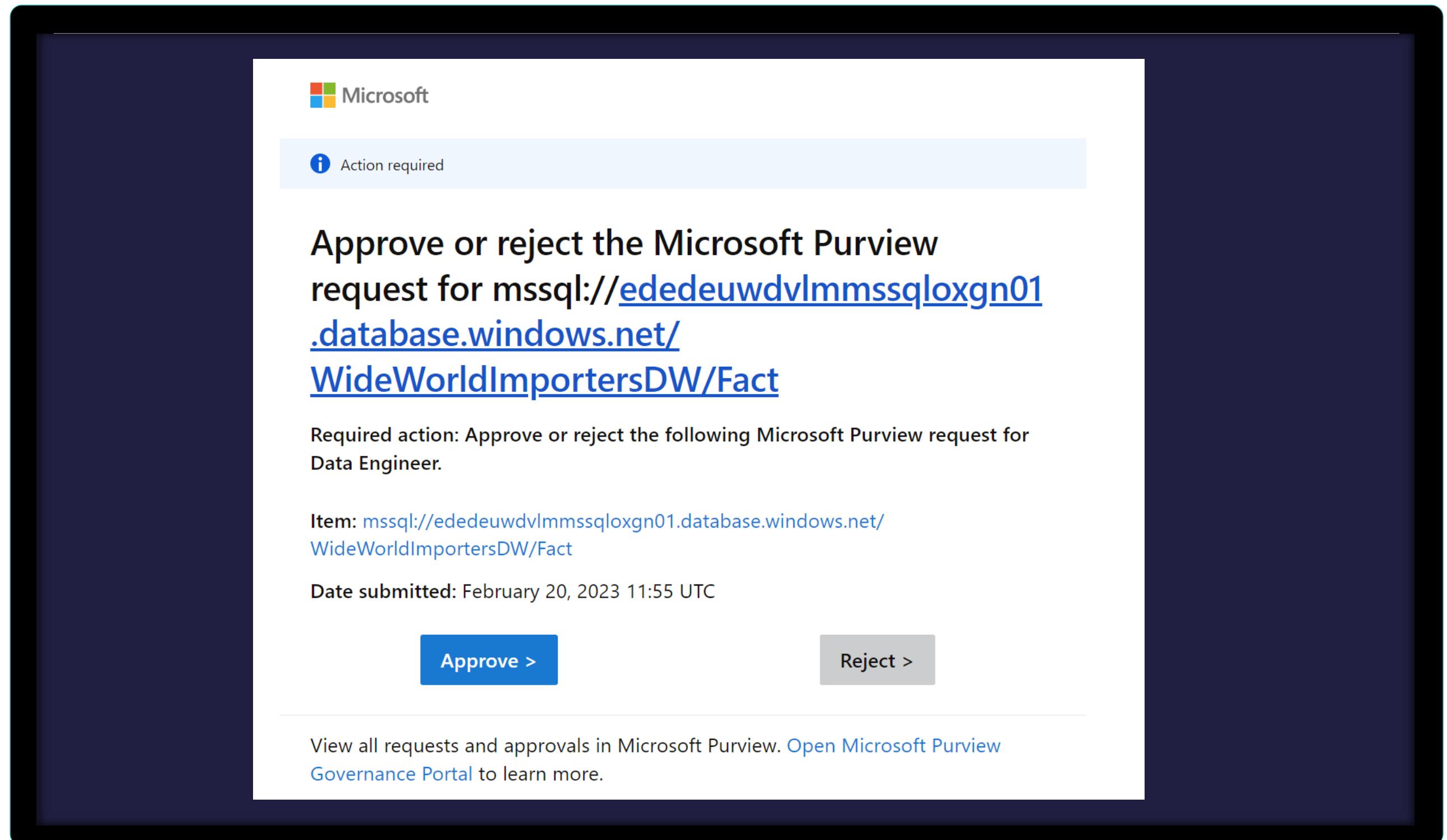
- Azure SQL Database
- Azure Storage
- Expiration Functionality



# Self-service access policies

## Data Access (Governance)

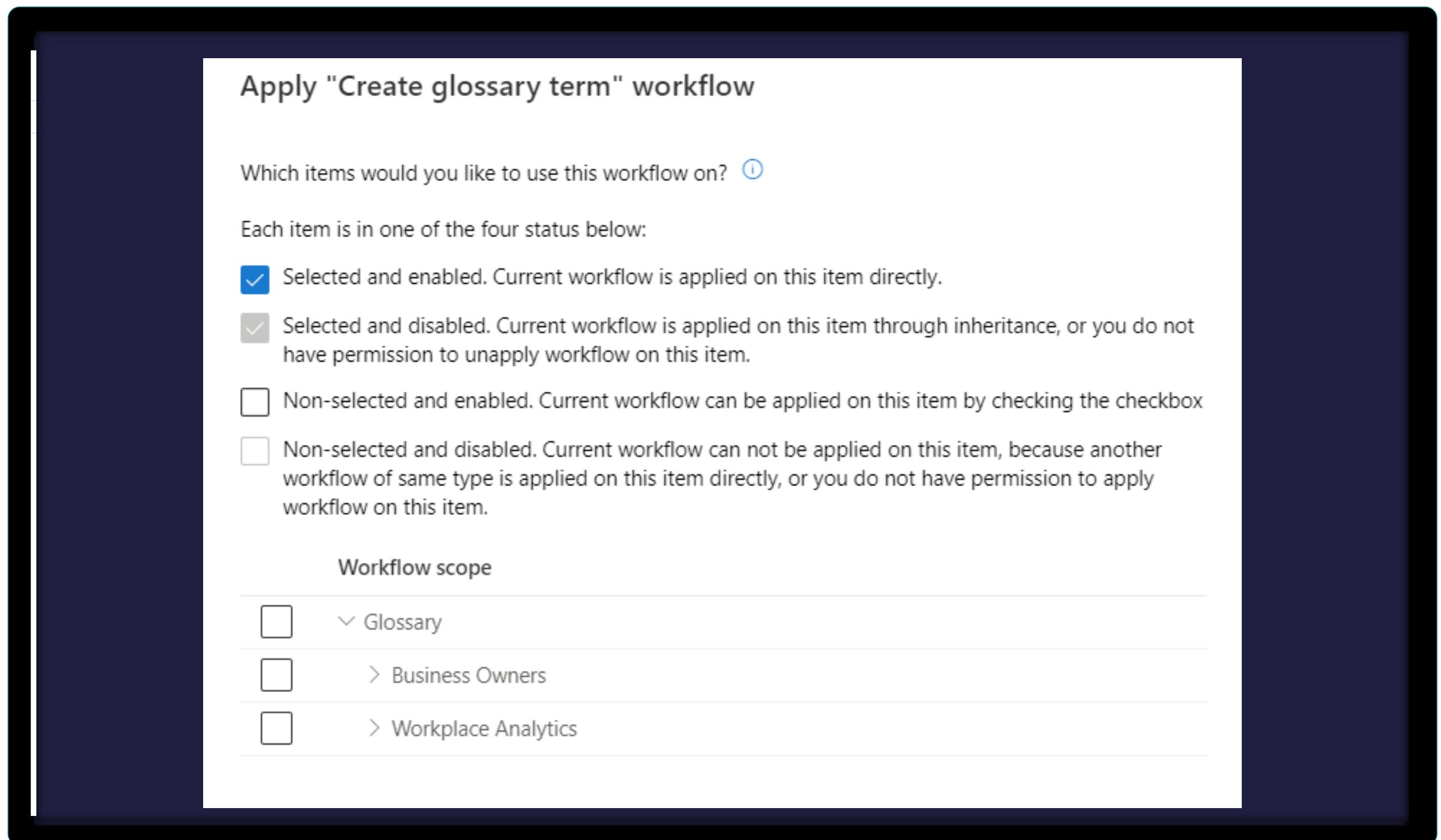
- Azure SQL Database
- Azure Storage
- Expiration Functionality
- Approval



# Workflows

## Data Catalog

- Create Glossary Term
- Delete Glossary Term
- Import Term
- Update Glossary Term
- Update Asset Attributes



# Microsoft Purview

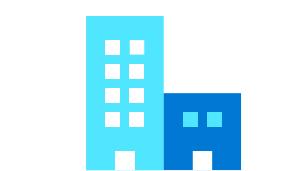
Generally Available

Preview

Unified Data Governance



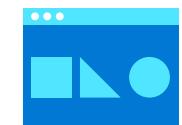
Microsoft Purview



On-prem



Cloud



SaaS  
Applications

Apps

Data Producers and Consumers

**Data Catalog**  
Enable effortless discovery of trusted data

Data Sharing

Share data within and between organizations

Data Officers

**Data Estate Insights**  
Assess data assets across your organization

**Data Use Management**  
Data access Policies

Platform

Data Map

Automate and manage metadata at scale

Scanning and Classifications



Azure  
Synapse  
Analytics



SQL Server



Power BI



Azure SQL



Microsoft 365

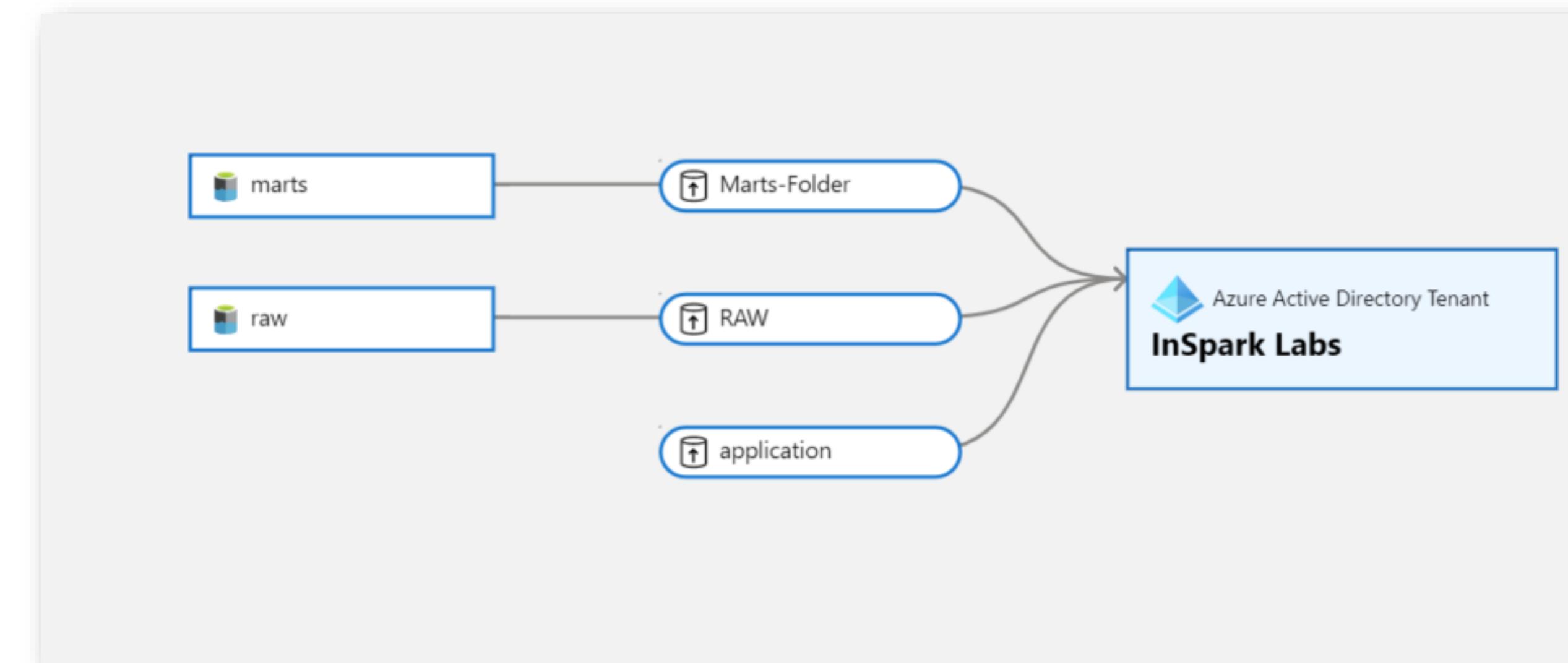
Open APIs  
(Apache Atlas 2.2)

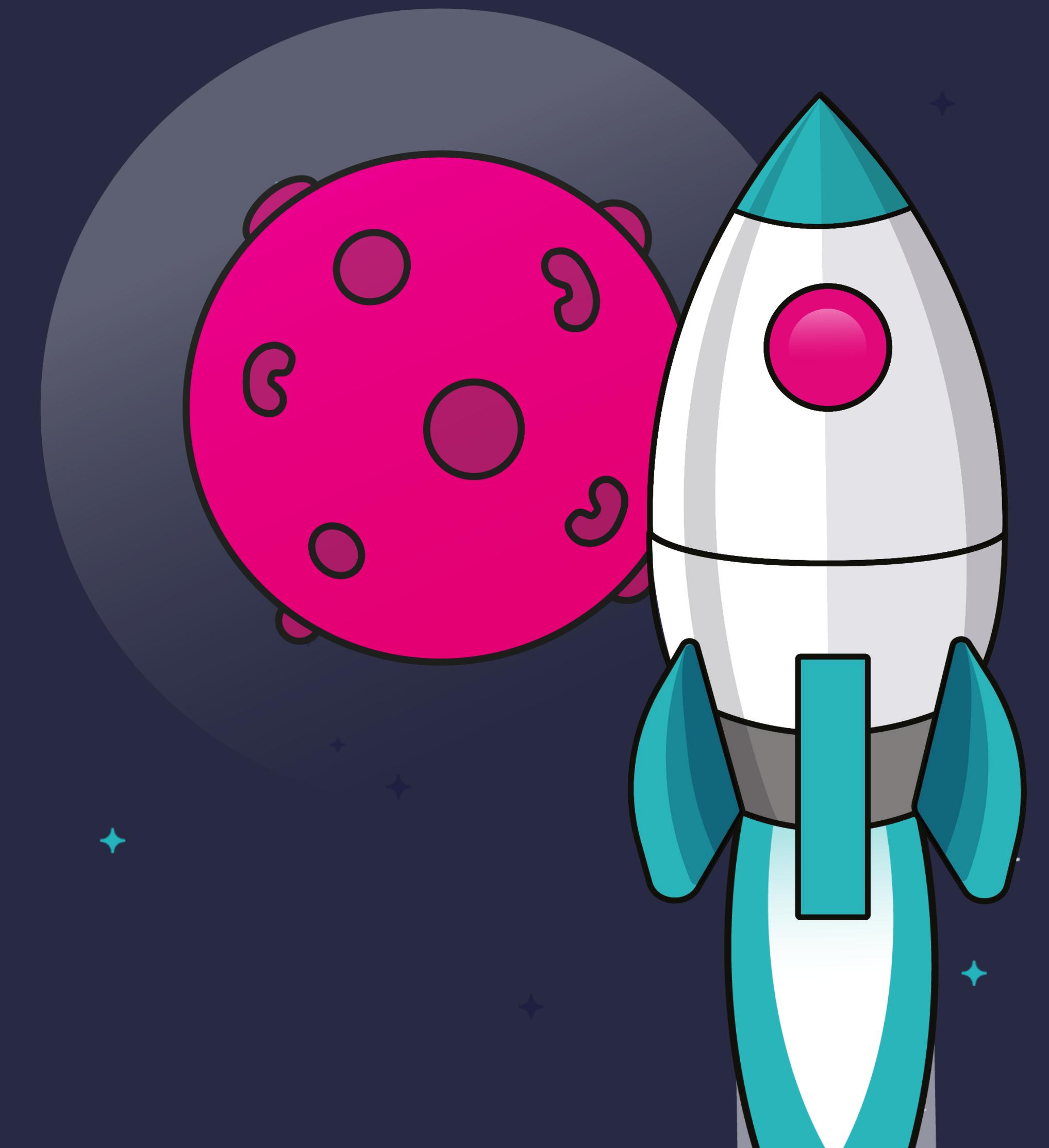
INSPARK

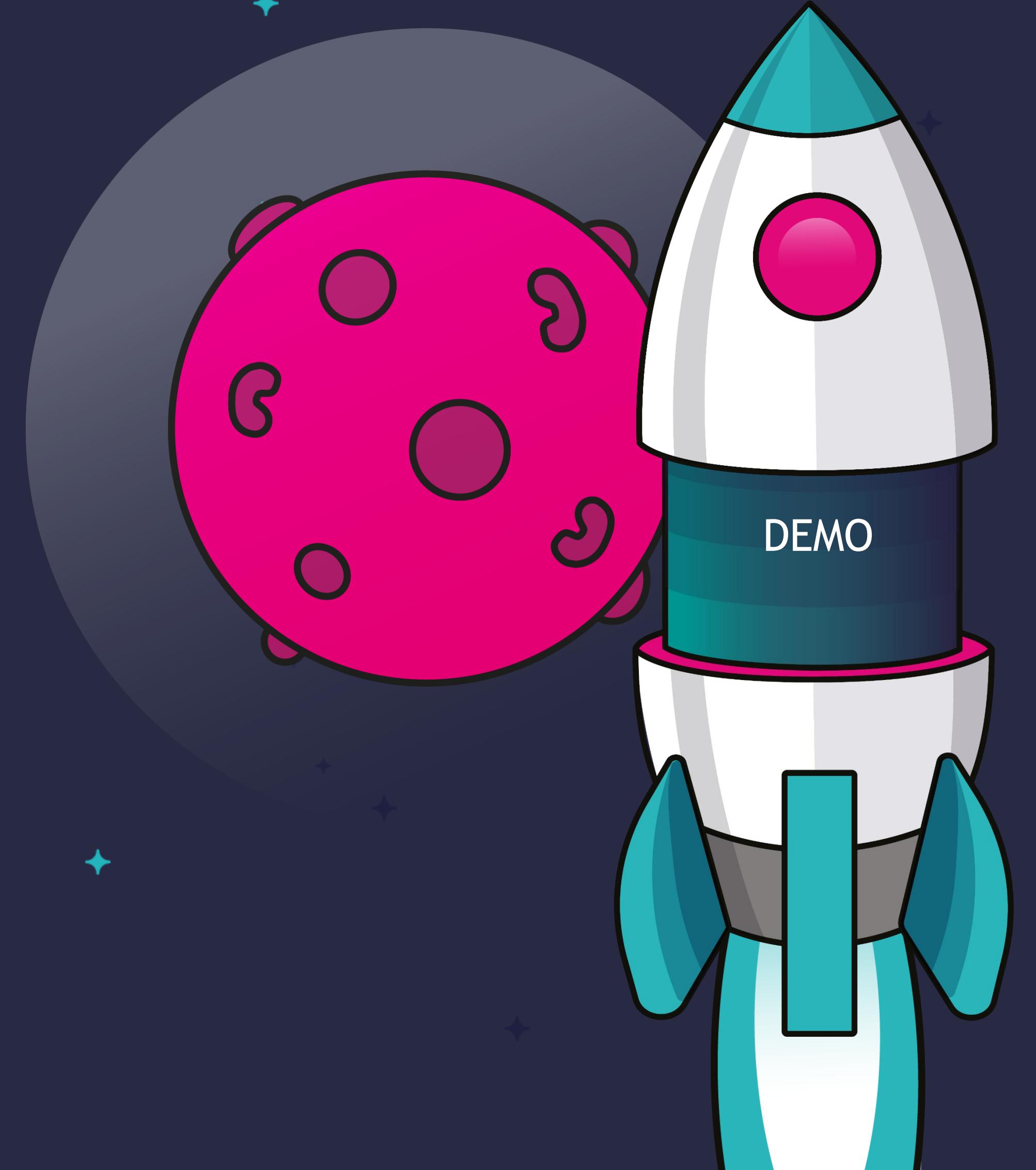
## Data Sharing

# Sharing

- Collaborate with external business partners while maintaining data security in your own environment.
- Outsource data transformation and processing to third party ISVs or data aggregators by sharing raw data and receiving normalized data and analytics results back.
- Automate sharing of big data in near real time and without data duplication.
- Share data between different departments within the organization.







Innovate to accelerate

# DEMO



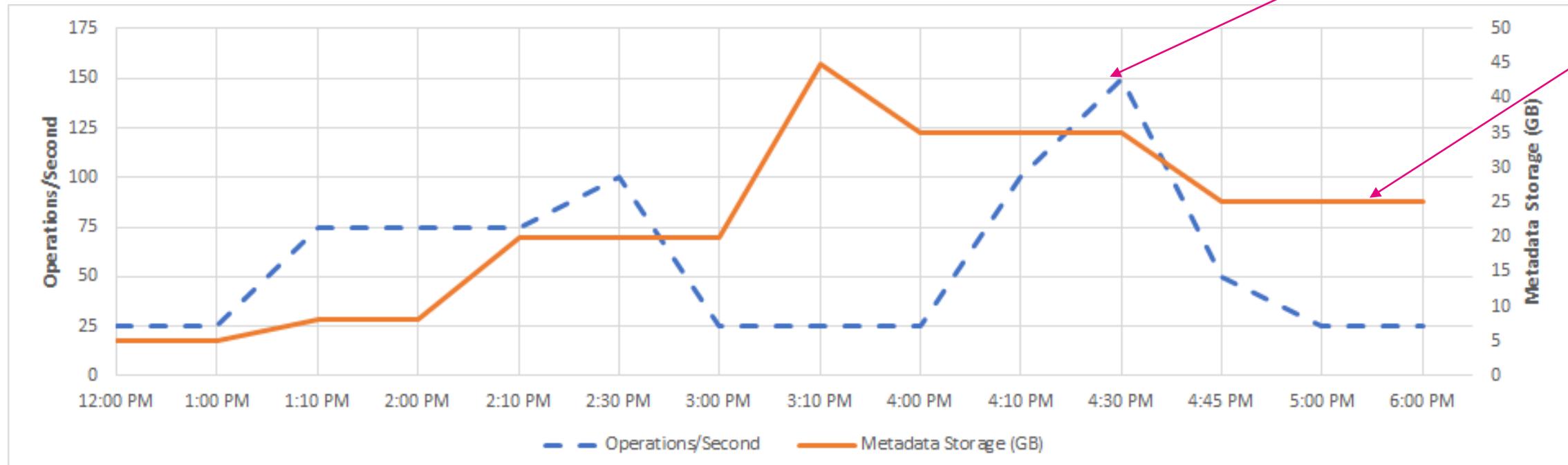
# Data Map Consumption

- Elastic Data Map

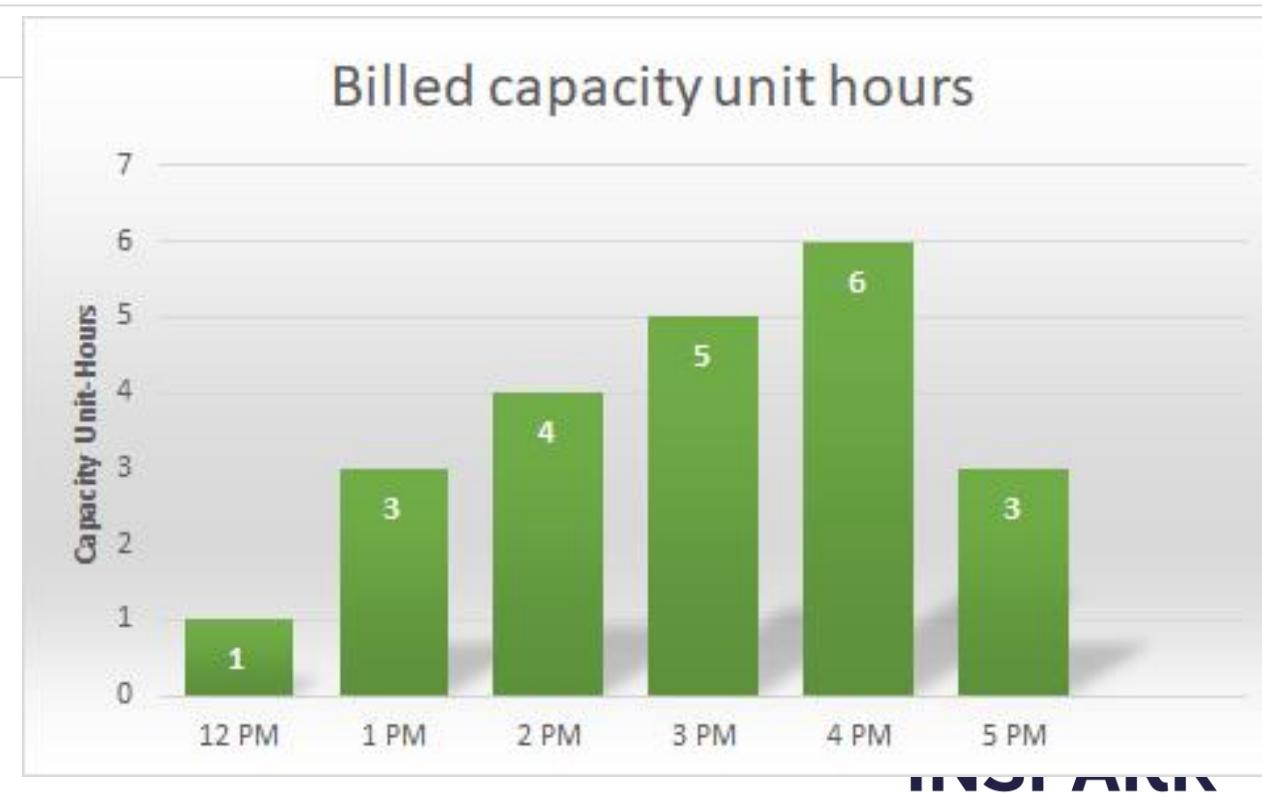
€0.373 per 1 Capacity Unit Hour

CU=Capacity Unit

1 Capacity Unit supports requests of up to 25 data map operations per second and includes storage of up to 10 GB of metadata about data assets.



Data Map Capacity Unit	Operations/Sec throughput	Storage capacity in GB
1	25	10
2	50	20
3	75	30
4	100	40
5	125	50
6	150	60
7	175	70
8	200	80
9	225	90
10	250	100
..	..	..
100	..	..

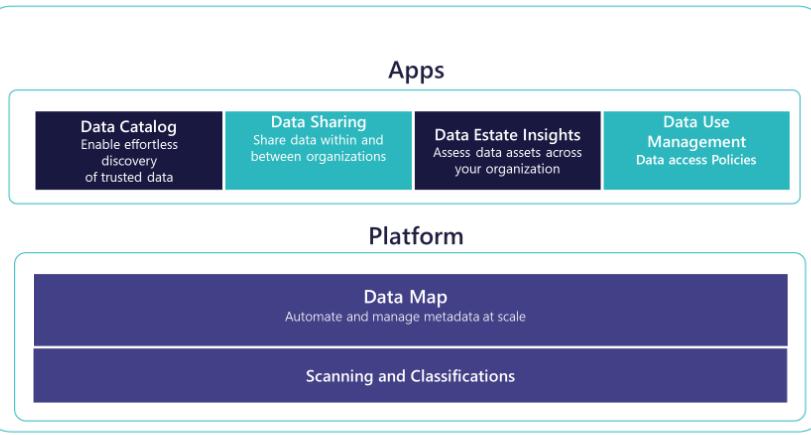


## Microsoft Purview Pricing

# Data Map Population

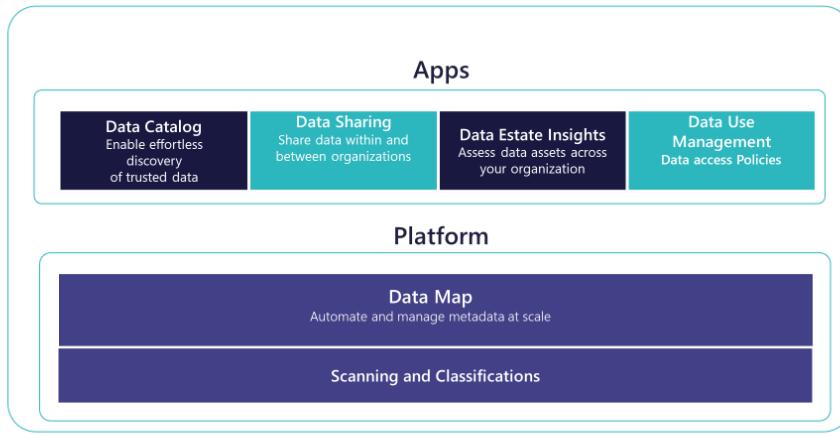
Automated Scanning, Ingestion & Classification

- Power BI Online  
Free for a limited time
- SQL Server On Prem  
Free for a limited time
- Other Data Sources  
€0.571 per 1 vCore Hour



## Microsoft Purview Pricing

# Data Map Enrichment



## Features

- Advanced Resource Set

**€0.191 per 1 vCore Hour**

Resource sets are a built-in feature of the Data Map used to optimize the storage and search of data assets associated with partitioned files in data lakes

- Insight Generation

**€0.743 per 1 vCore Hour**

Insights Generation aggregates metadata and classifications in the raw Data map into enriched, executive-ready reports that can be visualized in the Data Estate Insights application and granular asset level information in business-friendly format that can be exported.

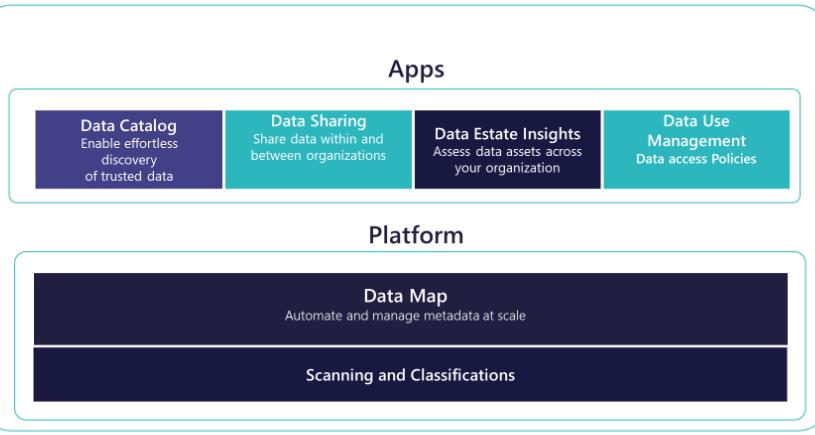
The screenshot shows a 'Feature options' interface with a teal border. On the left, there is a 'Refresh' button. Below it, a table lists features with their current state (On or Off) and refresh frequency. The 'Advanced Resource Sets' row shows 'On' and 'Off' for refresh frequency. The 'Data Estate Insights' row shows 'Off' and '7/25/2022, 2:03 PM' for refresh frequency.

Feature	State	Refresh frequency	Last updated	Description
Advanced Resource Sets	On	Off	9/28/2021, 10:00 AM	Allows customized resource sets with grouping patterns and runs aggregations on partitioned data.
Data Estate Insights	Off	Off	7/25/2022, 2:03 PM	Provides insight into your data estate.

## Microsoft Purview Pricing

# Data Catalog

- Search and browse of data assets  
Included with the Data Map
- Business Glossaries  
Included with the Data Map
- Lineage Visualization  
Included with the Data Map
- Self-Service Data Access  
Free in preview
- Business Workflows  
Free in preview

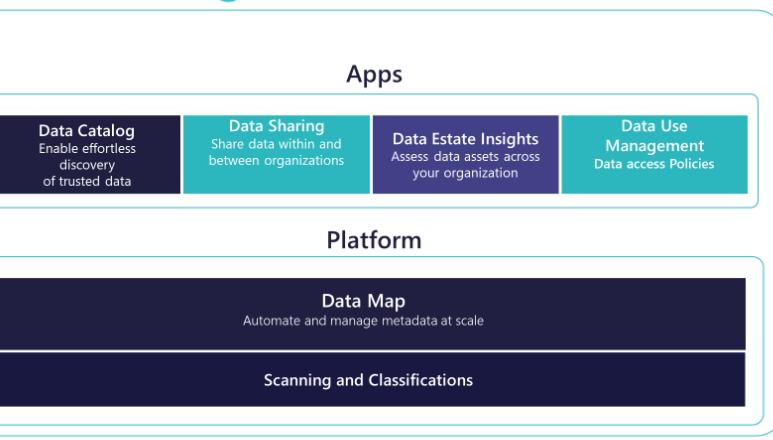


## Microsoft Purview Pricing

# Data Estate Insights

- Insights Consumption  
€0.191 per API call

One API call returns up to 10,000 rows of tabular result.



## Microsoft Purview Pricing

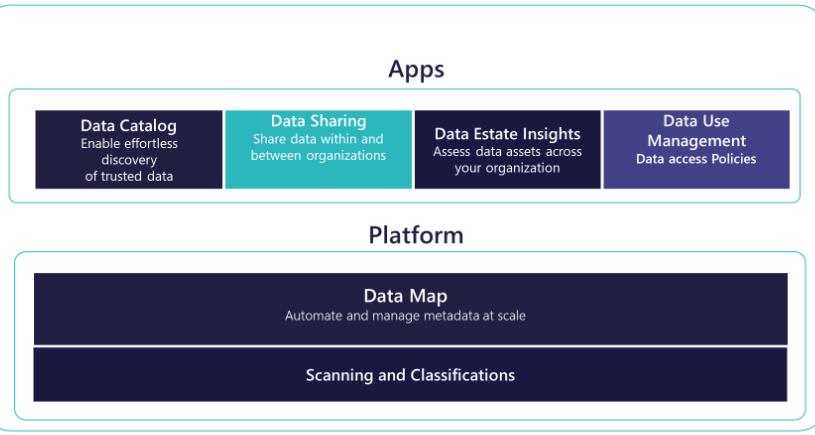
# Data Policy

- DevOps Policies

€0.003 per policy hour

## Data policies (data asset access)

Free in preview

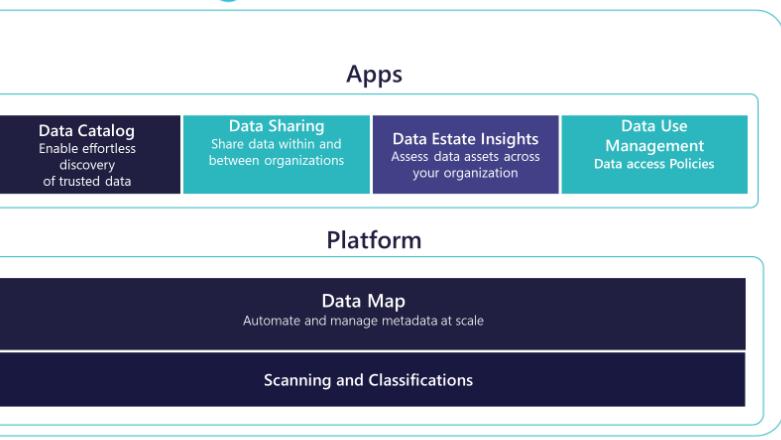


## Microsoft Purview Pricing

# Data Sharing

- In place sharing for ABLB and ADLS

Free in preview



# Pricing - Example

## Data Map Consumption

1,5 CPU x €0.407 X 730 hours

€ 445

## Data Map Population

4 scans x 4 hours x 32 VCore x  
€0.623 per vCore per hour

€ 320

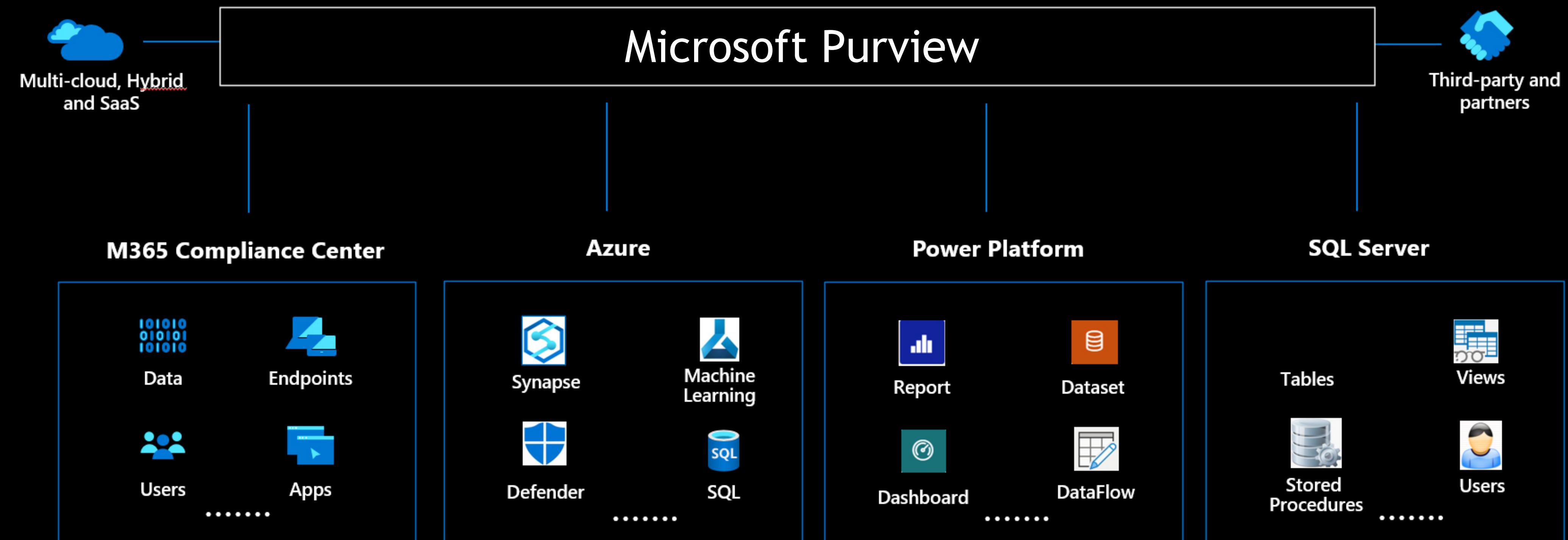
## Data Map Enrichement

60 scans x 1 hour x 8 VCore x  
€0.208 per 1vCore per  
hour(Resource set)

€ 100,00

€865,00

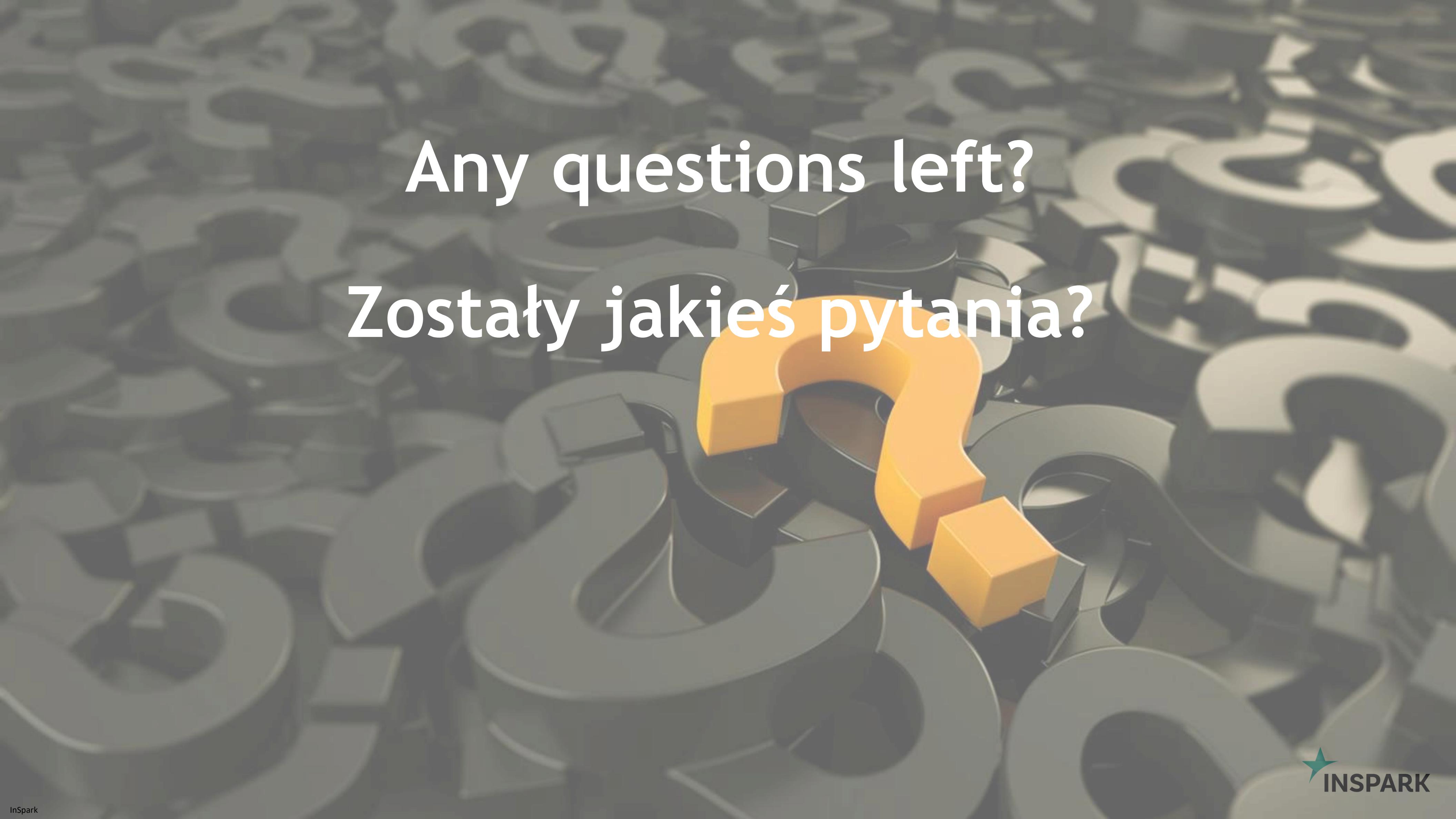
# Data Governance Opportunity



# Take aways / Roadmap

- Microsoft Purview will be added to Tenant Level
- Automatic detecting of new sources
- Automatic scanning (push)
- Tagging
- More reporting and In-product customization with Power BI Embedded
- Support for Databricks and Snowflake
  
- Define collections, before adding sources
- Start your project with Microsoft Purview readiness checklist for cloud-scale analytics
- <https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/scenarios/cloud-scale-analytics/best-practices/purview-checklist>





Any questions left?

Zostały jakieś pytania?

# Thank you for attending!



**Erwin de Kreuk**

Principal Consultant - Data & AI  
InSpark



@ErwinDeKreuk



[linkedin.com/in/ErwinDeKreuk](https://linkedin.com/in/ErwinDeKreuk)



[ErwinDeKreuk.com](http://ErwinDeKreuk.com) [github.com/edkreuk](http://github.com/edkreuk)

## FEEDBACK in the Whova APP

