

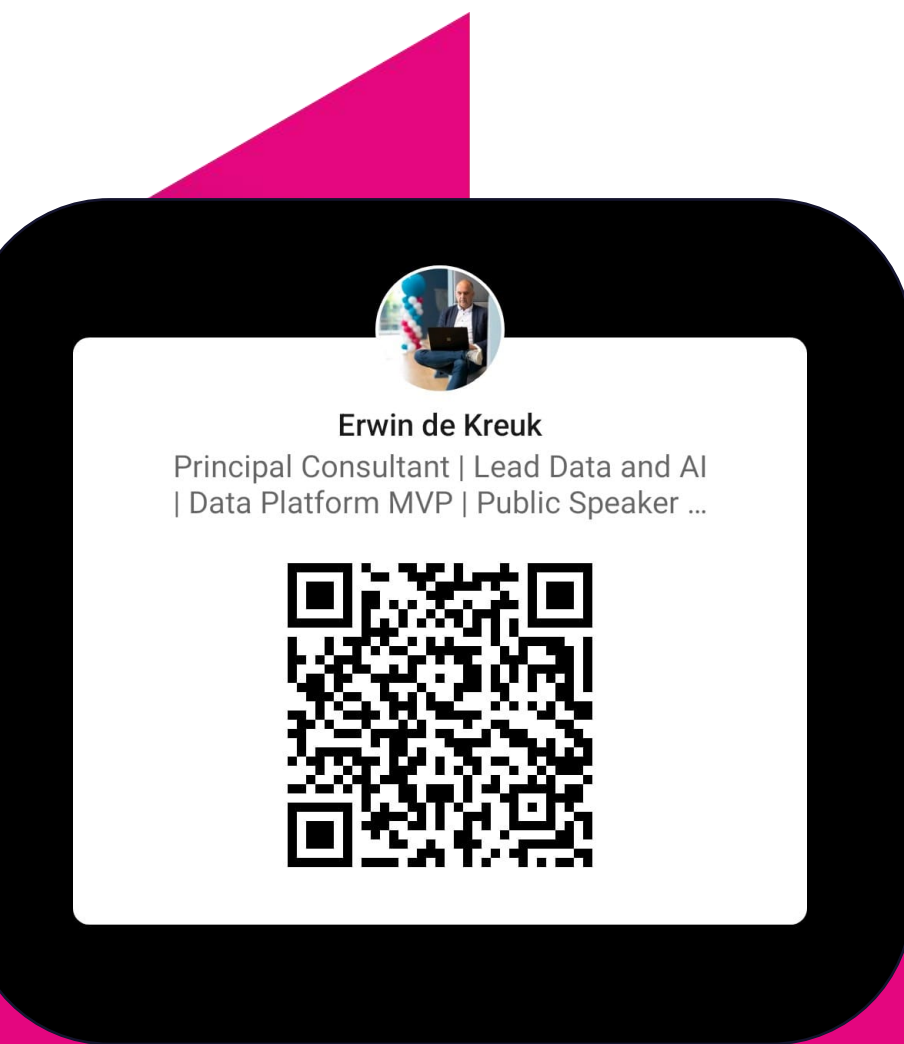
# Solve your Data Governance challenges with Microsoft Purview



Erwin de Kreuk

Principal Consultant - Lead Data & AI

InSpark



INSPARK



# Erwin de Kreuk

Principal Consultant – Lead Data & AI  
InSpark



@ErwinDeKreuk



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



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



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



# We Are InSpark

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





twoday  
capacity



# Premium Sponsors

---



# Raffle Prizes

---

## Standard Sponsors

TIME **X** TENDER

[stoltze][it]

ORANGEMAN

 **exmon**

 **KPMG**

**Fellowwind**



## AGENDA

### Introduction

### Challenges with Data Governance

### What is Microsoft Purview

- Data Map

#### Microsoft Purview Apps

- Data Catalog
- Data Estate Insights
- Data Use Management
- Data Sharing

### Road Map /Take aways

### Next Steps





# 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

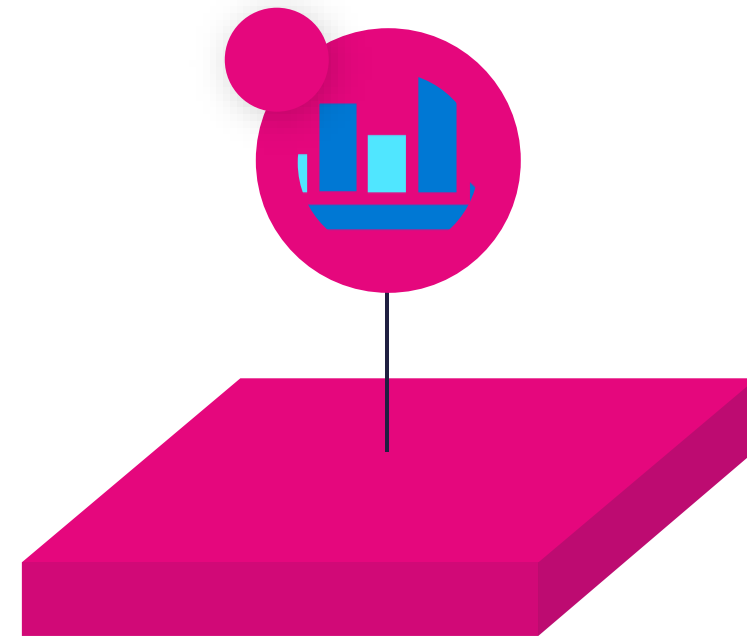
Expected global volume  
of data generated  
annually by 2025

IDC

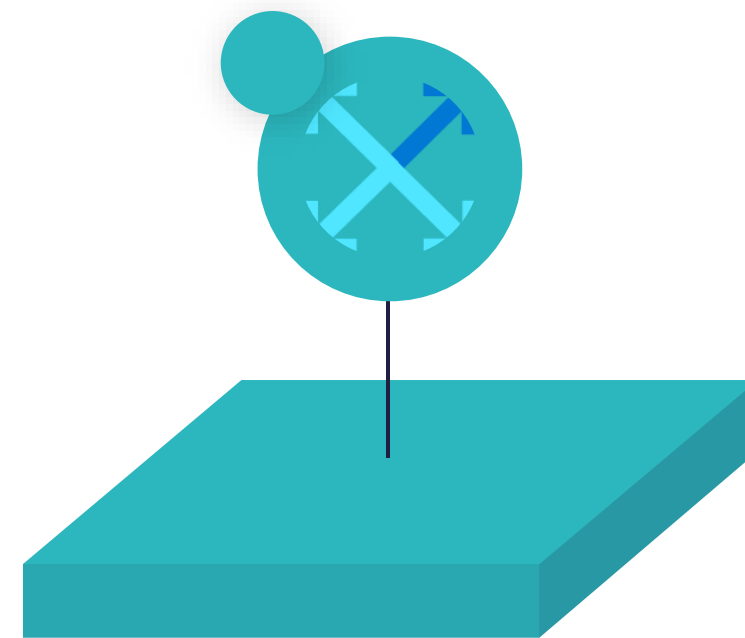
*IDC Data Age 2025, Dave Reinsel,*



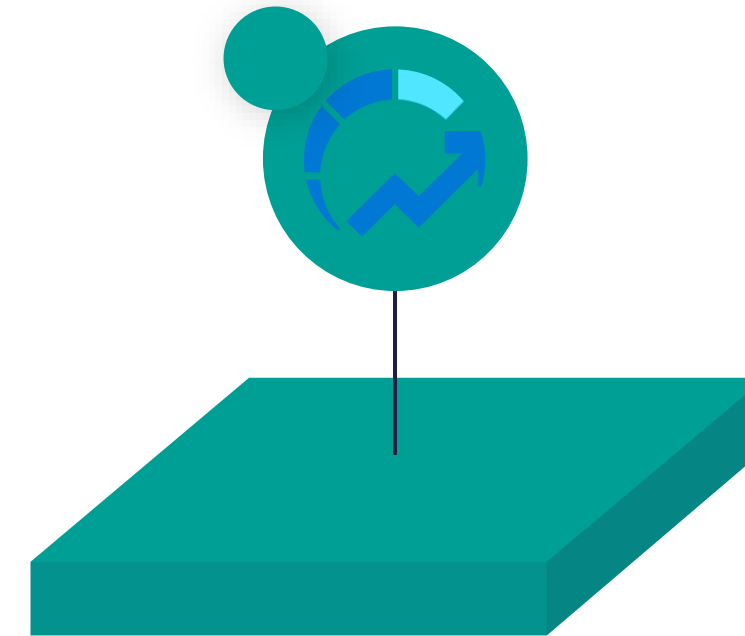
# Elements of successful data governance



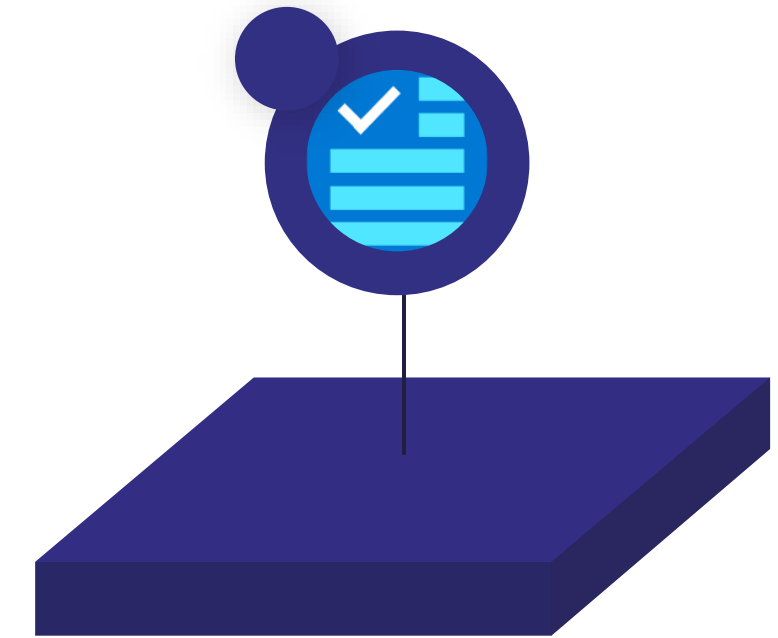
Manage growing  
data landscape



Overcome  
operational  
silos



Increase data  
agility



Comply with  
industry regulations

# 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

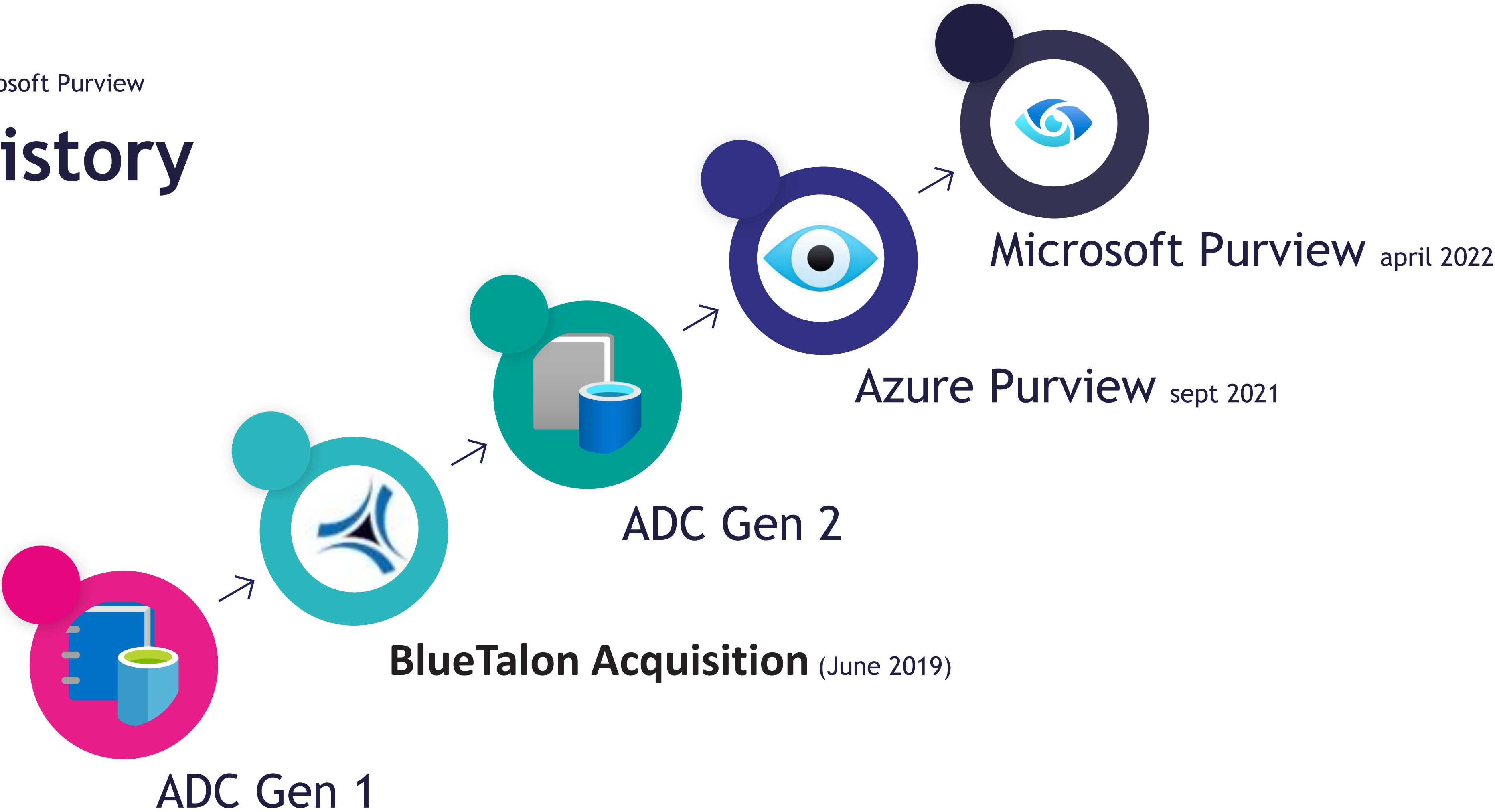
# Microsoft Purview





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

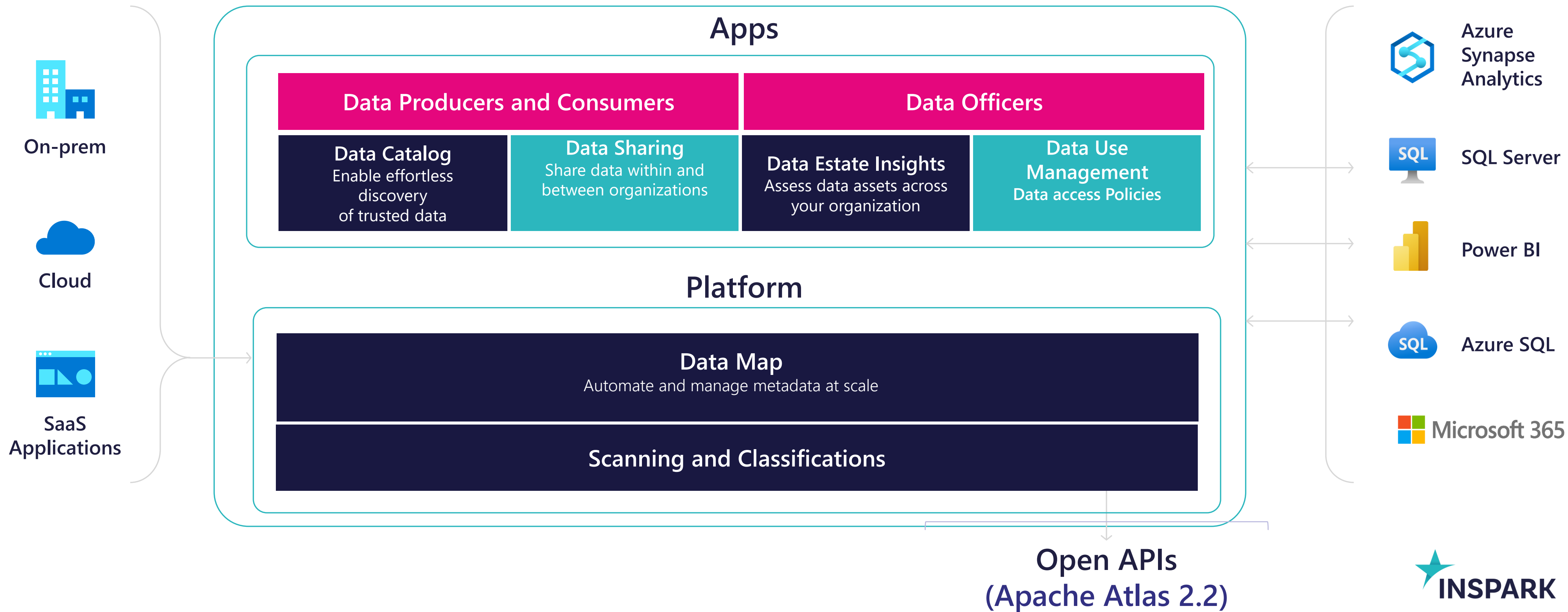
For data consumers, data engineers, data officers

# Microsoft Purview

Generally Available

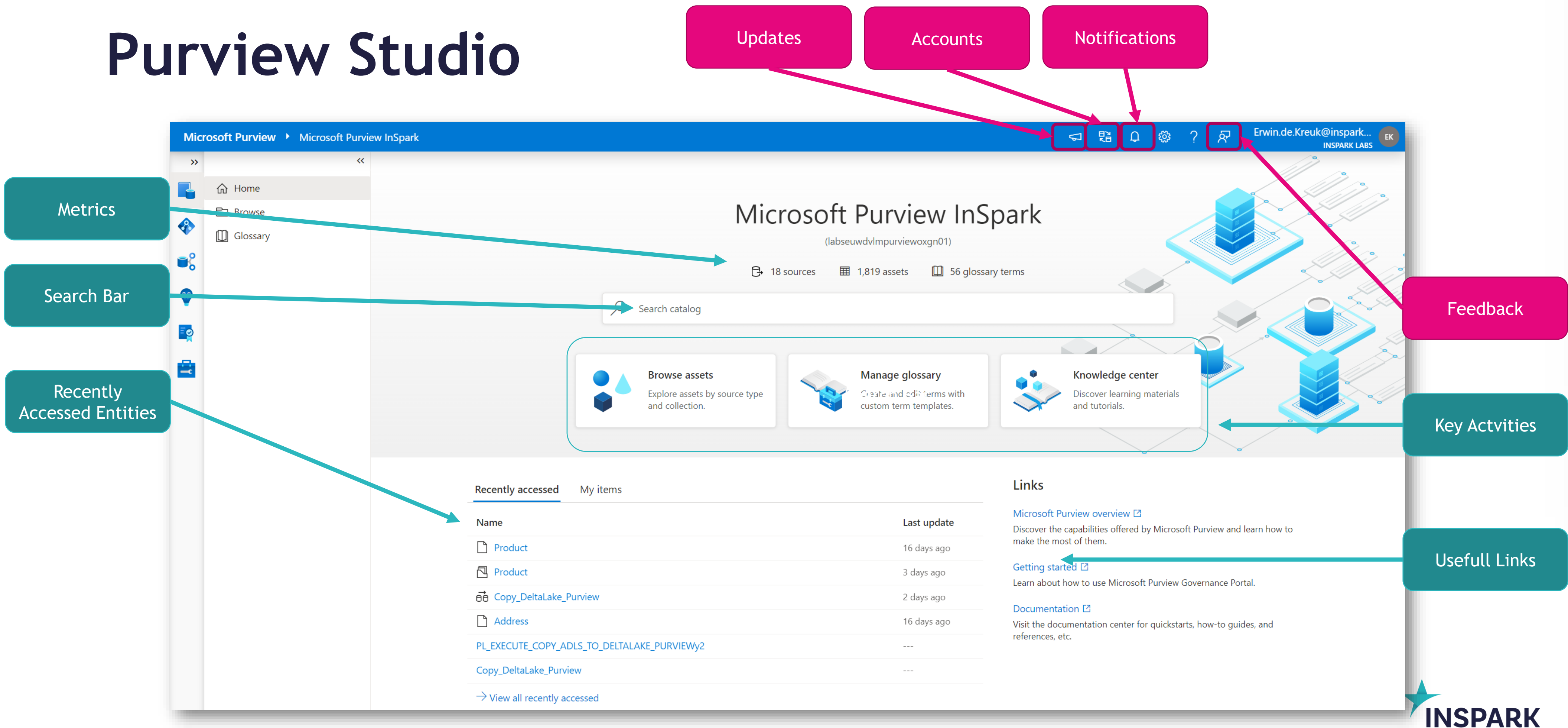
Preview

Unified Data Governance








# Purview Studio





# Purview Studio - Activity hubs


 **Data catalog**

 **Data map**

 **Data share (preview)**

 **Data estate insights**

 **Data policy (preview)**

 **Management**

- Quick Actions, recently accessed items, owned Items, search bar and Documentation.
- Manage Glossary Items, search, manage terms templates and custom attributes, import/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

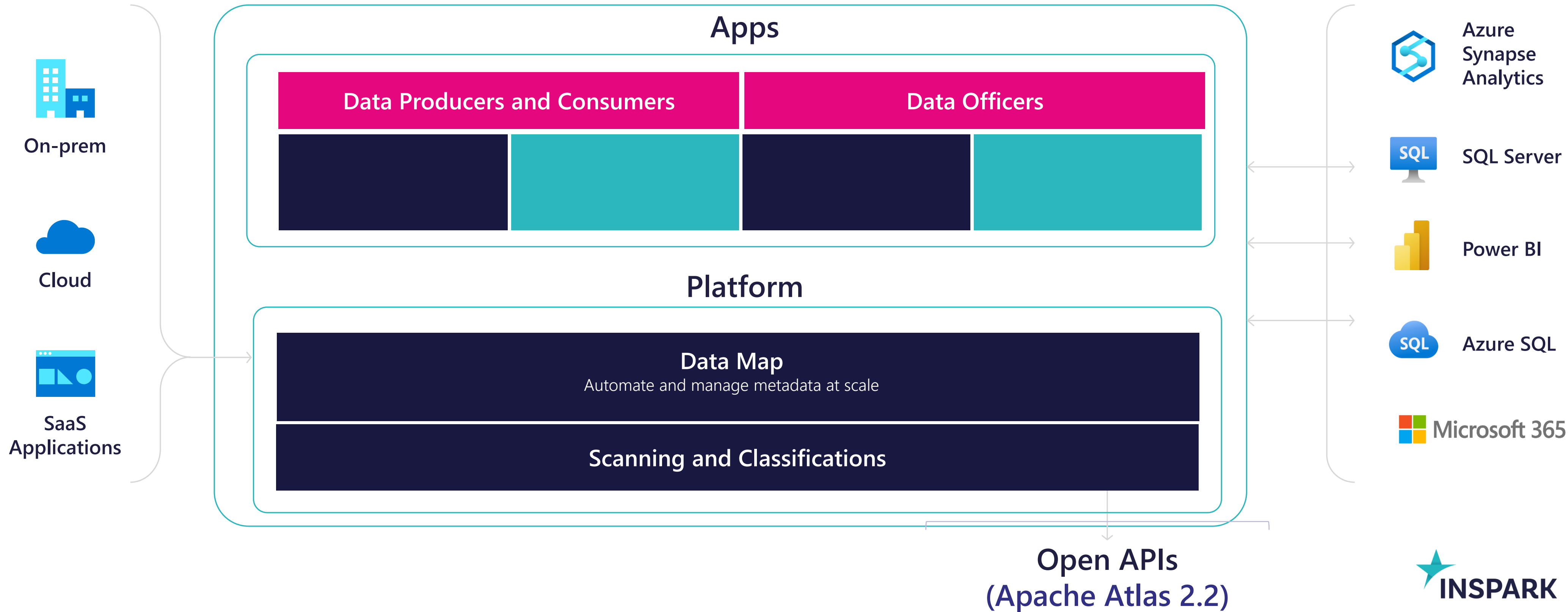
Generally Available

Preview

Unified Data Governance

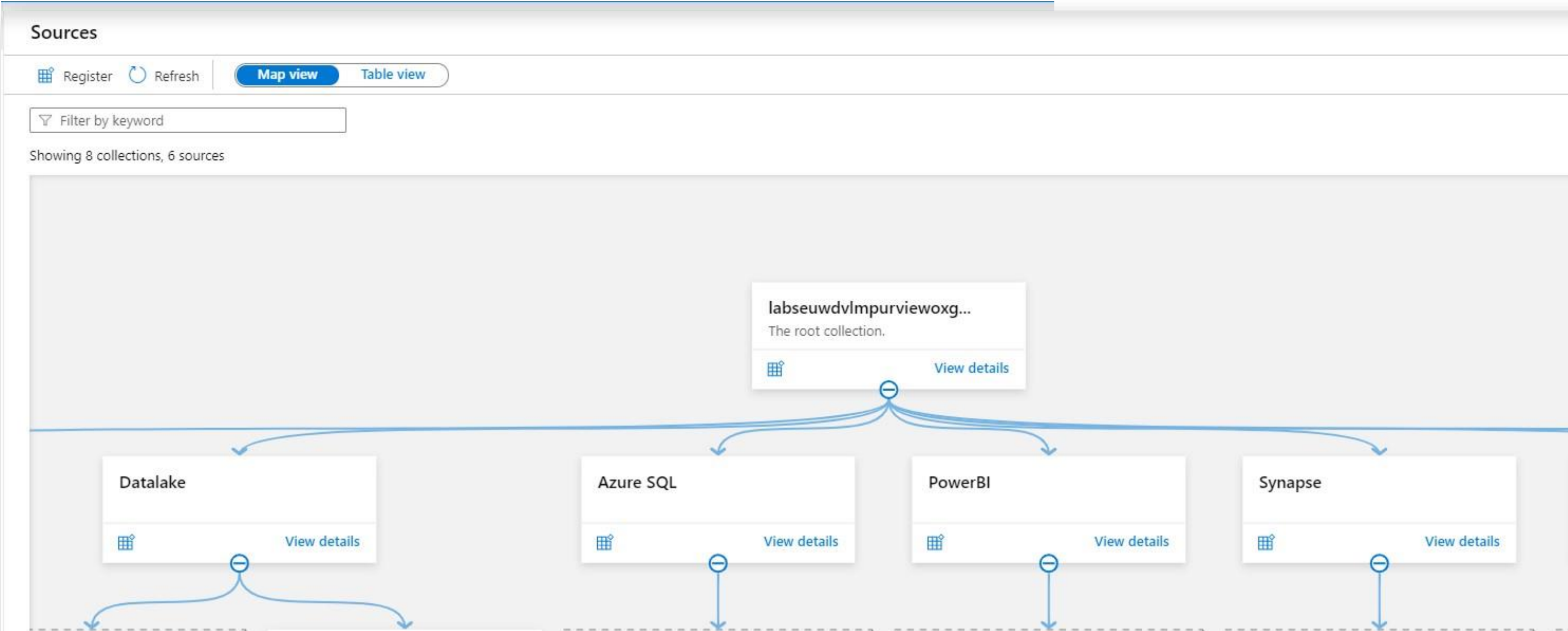
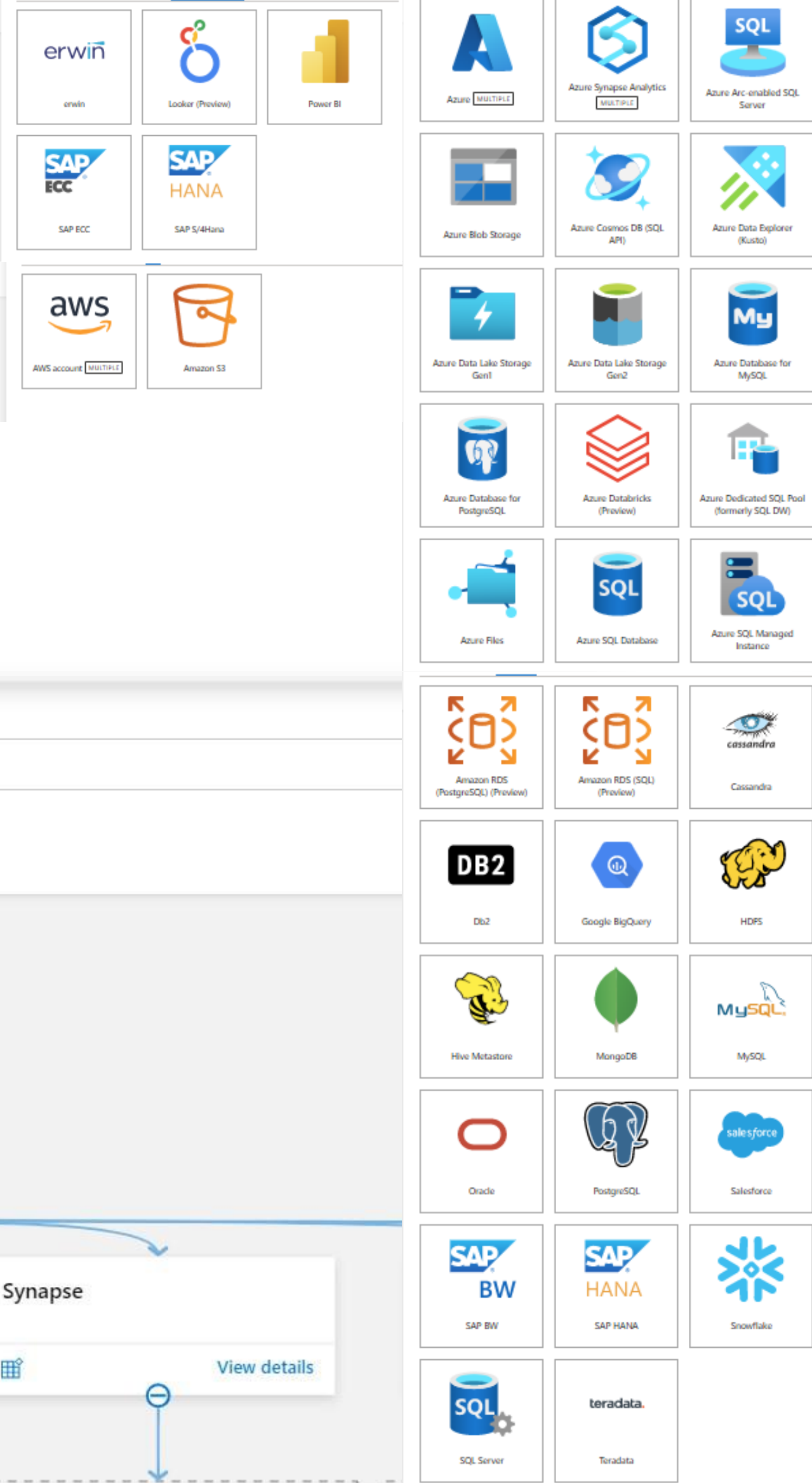


Microsoft Purview



# Collections and Sources

- A graph describing the data assets and their relationships across your data estate
- Support for 39+ data sources and growing.
- **NOTE:** Azure Machine Learning services arrived last week!





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

Sources > LS_ADLS > Weekly_Scan run history						
Weekly_Scan run history						
<div>▶ Run scan now</div> <div>✎ Edit scan</div> <div>🗑 Delete scan</div> <div>🔄 Refresh</div>						
Status	Schedule	Assets classified	Assets discovered	Start time	End tim	
✔ Completed	🕒 Scheduled	2	432,119	01/18/21 11:00 AM	01/18/2	
✔ Completed	🕒 Scheduled	172	432,117	01/11/21 11:00 AM	01/11/2	
✔ Completed	🕒 Scheduled	199	431,156	01/04/21 11:00 AM	01/04/2	
✔ Completed	👤 Manual	171	429,874	12/30/20 02:47 PM	12/30/2	
✔ Completed	👤 Manual	192	429,496	12/29/20 02:57 PM	12/29/2	

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 1 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)

h:mm:ss AM

# 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

Sources

Collections

Monitoring

Metamodel

Asset types (preview)

Annotation management

Classifications

Classification rules

Managed attributes

Data Sharing

Shares

Share invites

Source management

Scan rule sets

Pattern rules

Integration runtimes

Classifications

+ New

Edit

Delete

Refresh

System

Custom

These are the system provided classifications.

Filter by name...

Display name	Formal name	Descr
<a href="#">ABA Routing Number</a>	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA F
<a href="#">All Full Names</a>	MICROSOFT.PERSONAL.NAME	All Fu
<a href="#">All Physical Addresses</a>	MICROSOFT.PERSONAL.PHYSICALADDRESS	All PH
<a href="#">Argentina National Identity (DNI) Number</a>	MICROSOFT.GOVERNMENT.ARGENTINA.DNI_NU...	Argen
<a href="#">Australia Bank Account Number</a>	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACCOU...	Austr
<a href="#">Australia Business Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Austr
<a href="#">Australia Company Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Austr
<a href="#">Australia Driver's License Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...	Austr
<a href="#">Australia Medical Account Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL_...	Austr
<a href="#">Australia Passport Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...	Austr
<a href="#">Australia Tax File Number</a>	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE_...	Austr
<a href="#">Austria Driver's License Number</a>	MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LIC...	Austr

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
<a href="#">CityID</a>			
<a href="#">CityName</a>	<div>World Cities</div>		
<a href="#">LastEditedBy</a>			
<a href="#">LatestRecordedPopulation</a>			
<a href="#">Location</a>			



# Microsoft Purview

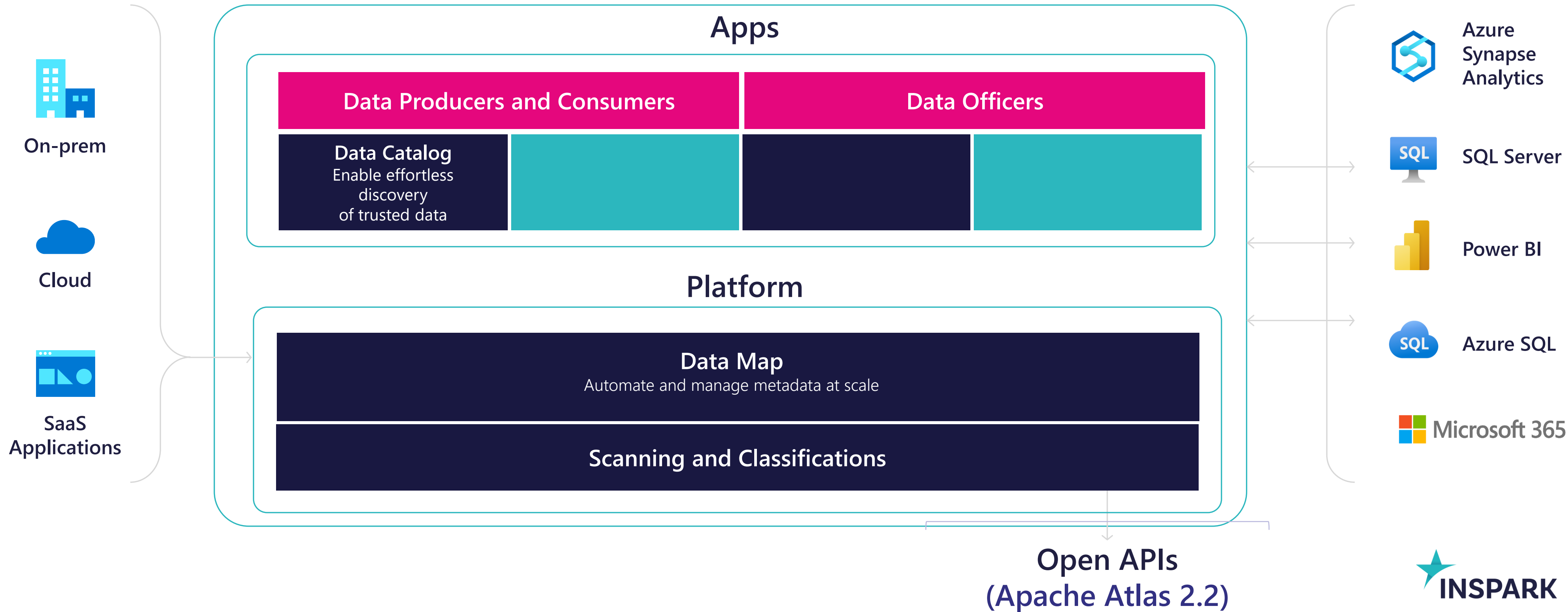
Generally Available

Preview

Unified Data Governance

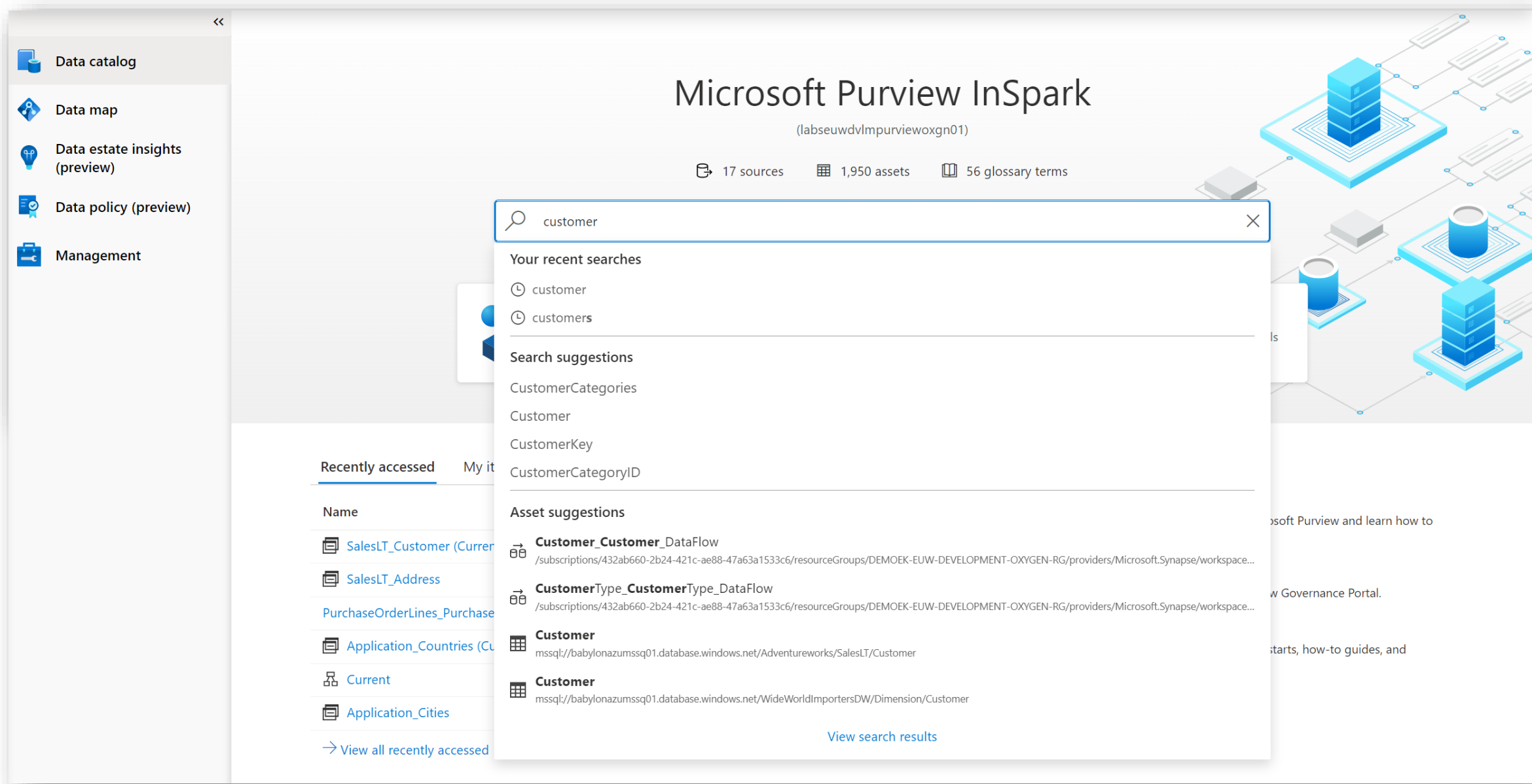


Microsoft Purview



# Search

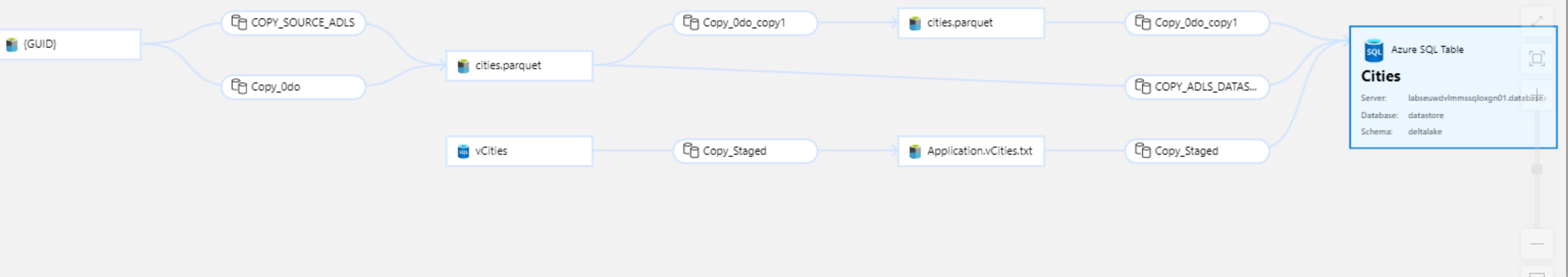
- Semantic search and browse
- Business glossary and workflows





# Lineage

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



# Manual Lineage

- Create Lineage in the Catalog between 2 assets
- Additionally configure the column level lineage

Copy Source to ADLS

Cities

CityID

Azure Data Lake Storage Gen2  
Resource Set

WWI Application Cities

CityID

CityName

LatestRecordedPopulation

Location

Cities

CityID

CityName

LatestRecordedPop...

Location

Define lineage

+ Add manual lineage

Asset	Relationship	Asset
> Cities	Produces	WWI Application Cities
WWI Application Cities	Produces	Cities
CityID	Produces	CityID
CityName	Produces	CityName
LatestRecordedPopulation	Produces	LatestRecordedPopulation
Location	Produces	Location
Copy Source to ADLS	Produces	WWI Application Cities

Data Catalog

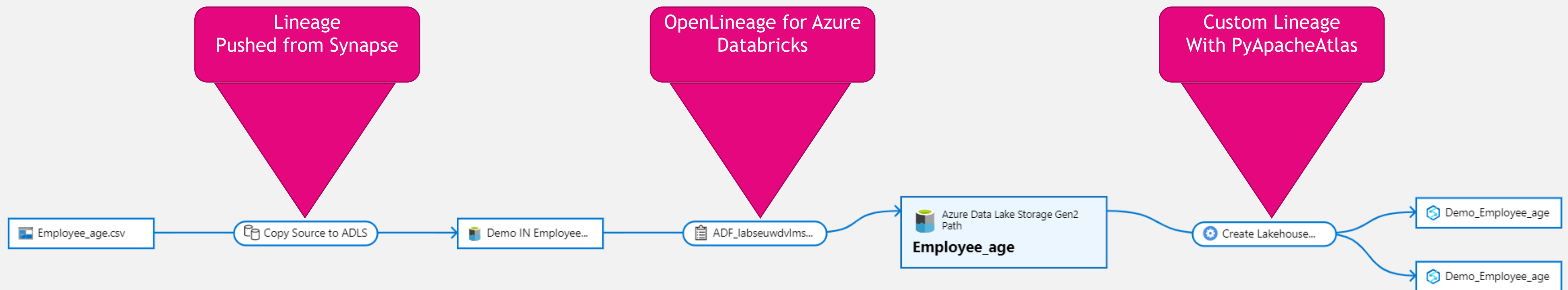
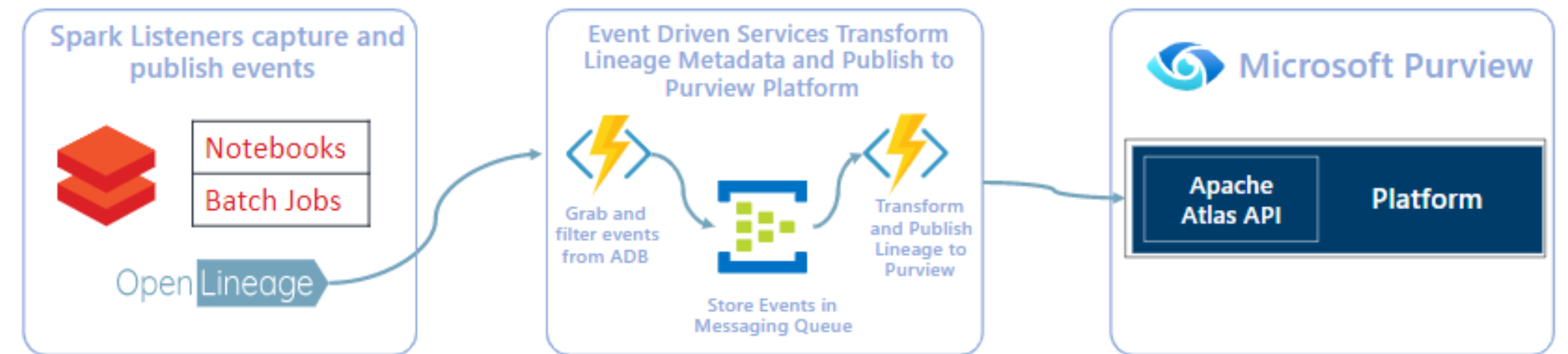
# Custom Lineage

- OpenLineagefor Azure DataBricks

[microsoft/Purview-ADB-Lineage-Solution-Accelerator: A connector to ingest Azure Databricks lineage into Microsoft Purview \(github.com\)](#)

- Create custom Lineage with Apache Spark 2.2

<https://github.com/wjohnson/pyapacheatlas>



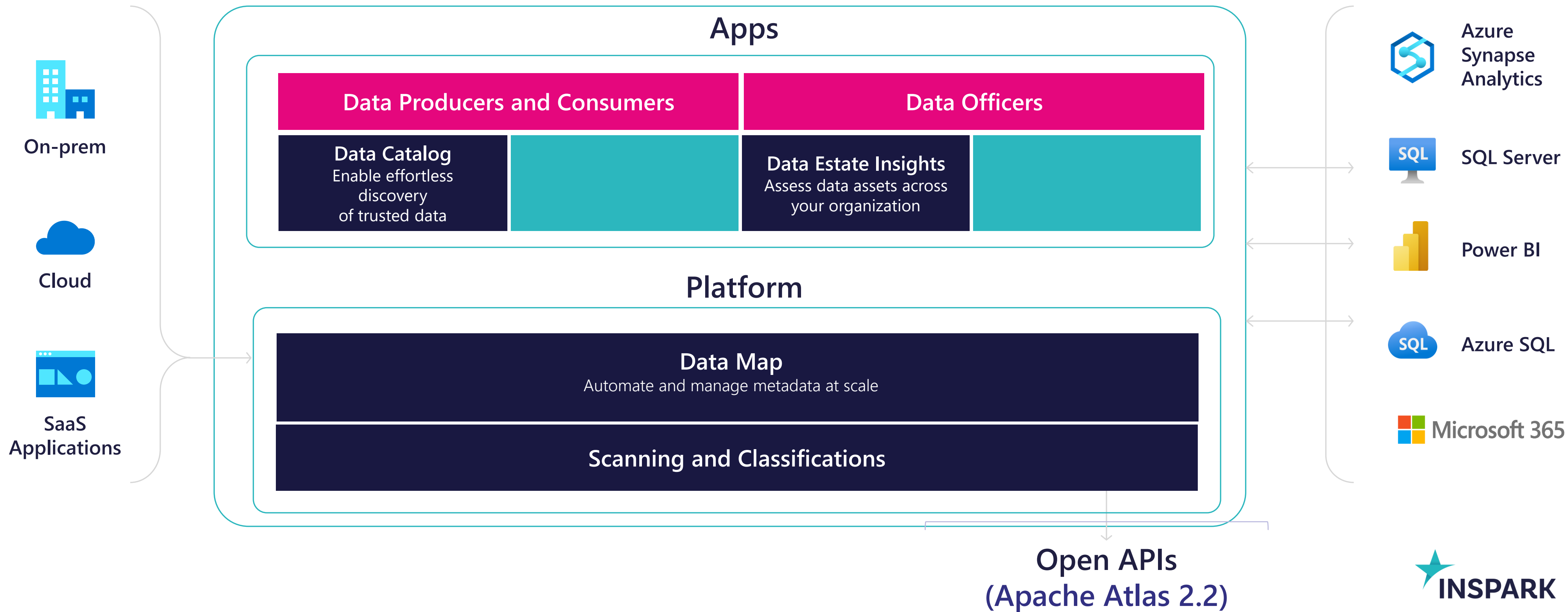


# Microsoft Purview

Generally Available

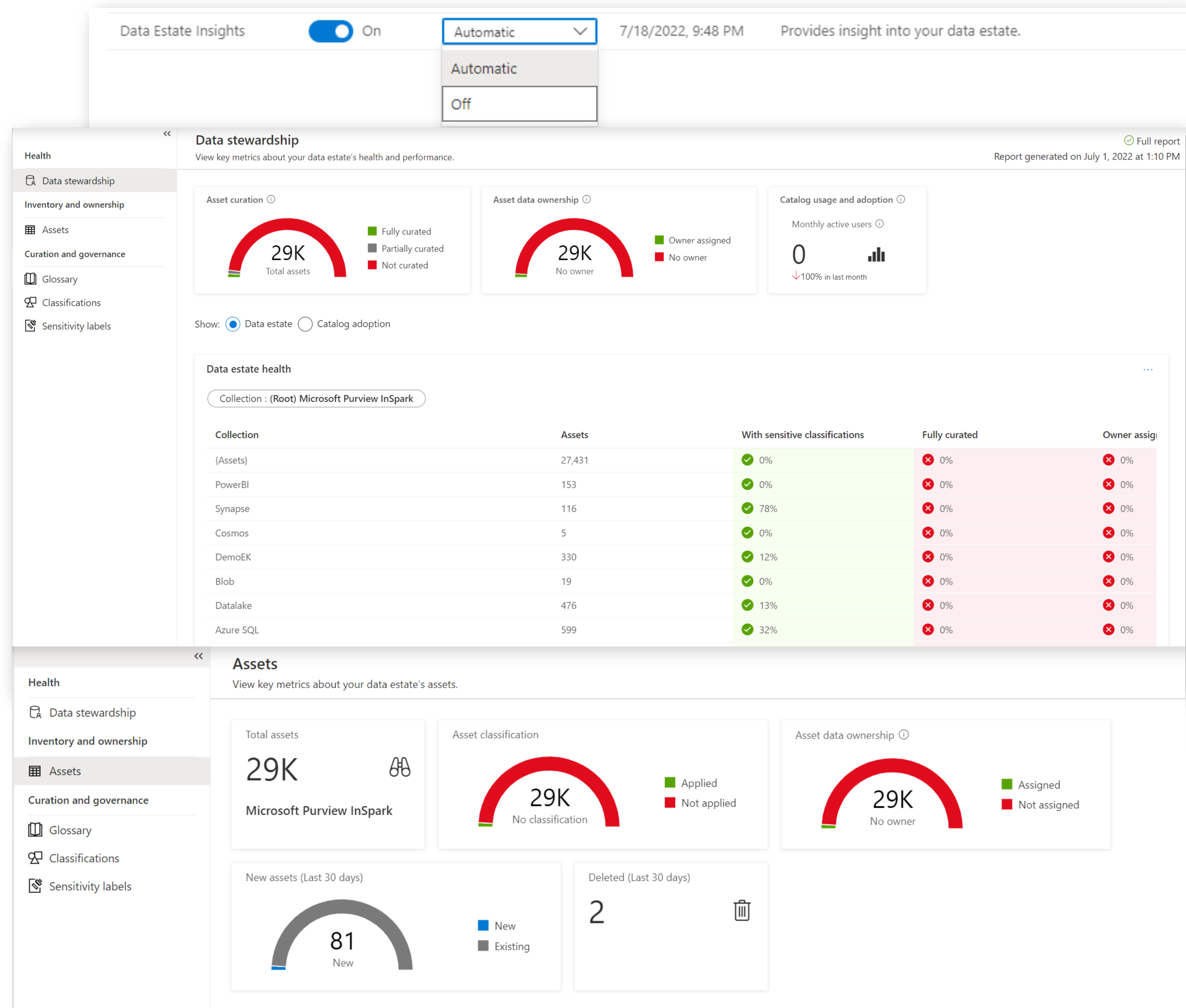
Preview

Unified Data Governance



# Metrics

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





# ACCESS CONTROL

Data Owner  
Dev Ops  
Self Service access

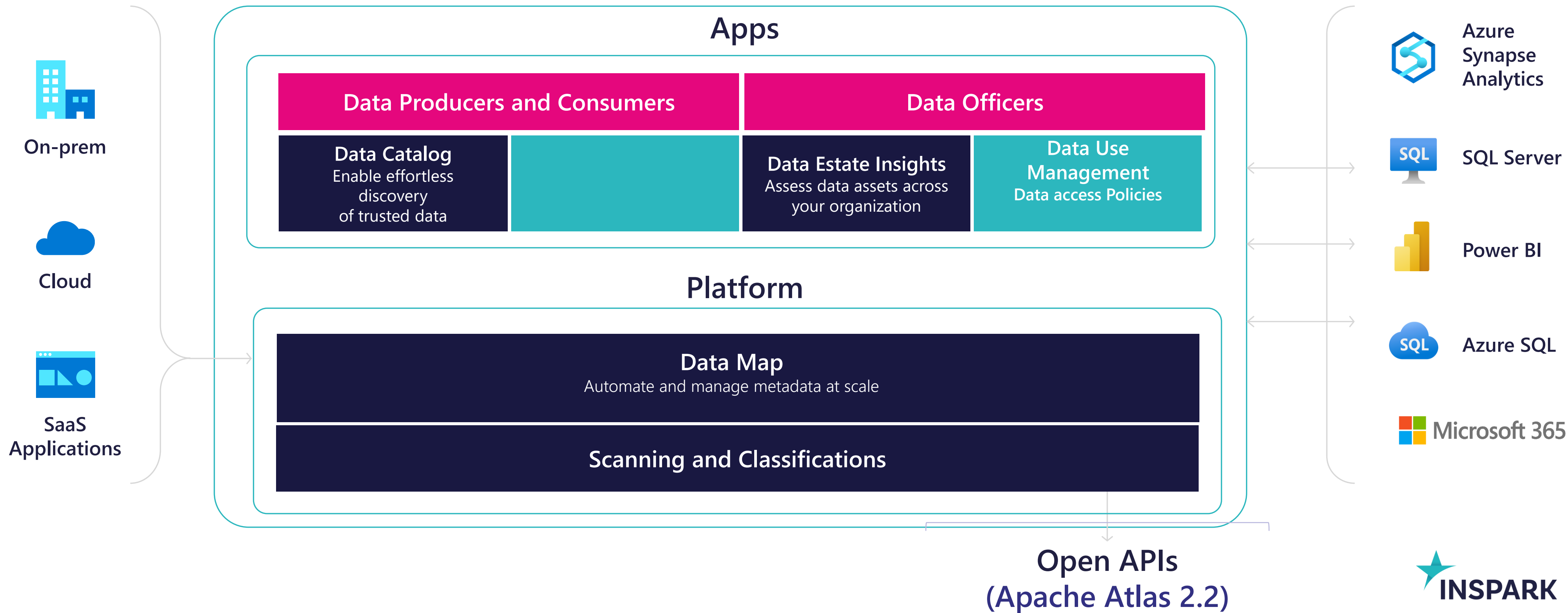


# Microsoft Purview

Generally Available

Preview

Unified Data Governance




# 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 \*

Select data source type



#### Assets

Individual objects stored within the data catalog.

Data source type \*

Select 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

DevOps roles

SQL Performance monitor

SQL Security auditor

SQL Performance Monitor allows read access to performance related system views

Add/remove subjects

Subjects

Selected users, groups, and applications will show up here.

WideWorldImportersDW

Database Diagrams

Tables

System Tables

External Tables

Graph Tables

Views

System Views

INFORMATION\_SCHEMA.CHECK\_CONSTRAINTS

INFORMATION\_SCHEMA.COLUMN\_DOMAIN\_USAGE

INFORMATION\_SCHEMA.COLUMN\_PRIVILEGES

INFORMATION\_SCHEMA.COLUMNS

INFORMATION\_SCHEMA.CONSTRAINT\_COLUMN\_USAGE

INFORMATION\_SCHEMA.CONSTRAINT\_TABLE\_USAGE

INFORMATION\_SCHEMA.DOMAIN\_CONSTRAINTS

INFORMATION\_SCHEMA.DOMAINS

INFORMATION\_SCHEMA.KEY\_COLUMN\_USAGE

INFORMATION\_SCHEMA.PARAMETERS

INFORMATION\_SCHEMA.REFERENTIAL\_CONSTRAINTS

INFORMATION\_SCHEMA.ROUTINE\_COLUMNS

INFORMATION\_SCHEMA.ROUTINES

INFORMATION\_SCHEMA.SCHEMATA

INFORMATION\_SCHEMA.SEQUENCES

INFORMATION\_SCHEMA.TABLE\_CONSTRAINTS

INFORMATION\_SCHEMA.TABLE\_PRIVILEGES

INFORMATION\_SCHEMA.TABLES

INFORMATION\_SCHEMA.VIEW\_COLUMN\_USAGE

INFORMATION\_SCHEMA.VIEW\_TABLE\_USAGE

INFORMATION\_SCHEMA.VIEWS



# Self-service access policies

## Data Access (Governance)

- Azure SQL Database
- Azure Storage
- Expiration Functionality

The screenshot displays the 'Data access request' configuration window in Microsoft Purview. The window is titled 'Data access request' and includes a subtitle 'Microsoft Purview workflow'. Below the title bar, there is a link to 'Edit name and description'. The main content area is titled 'Start and wait for an approval' and contains the following settings:

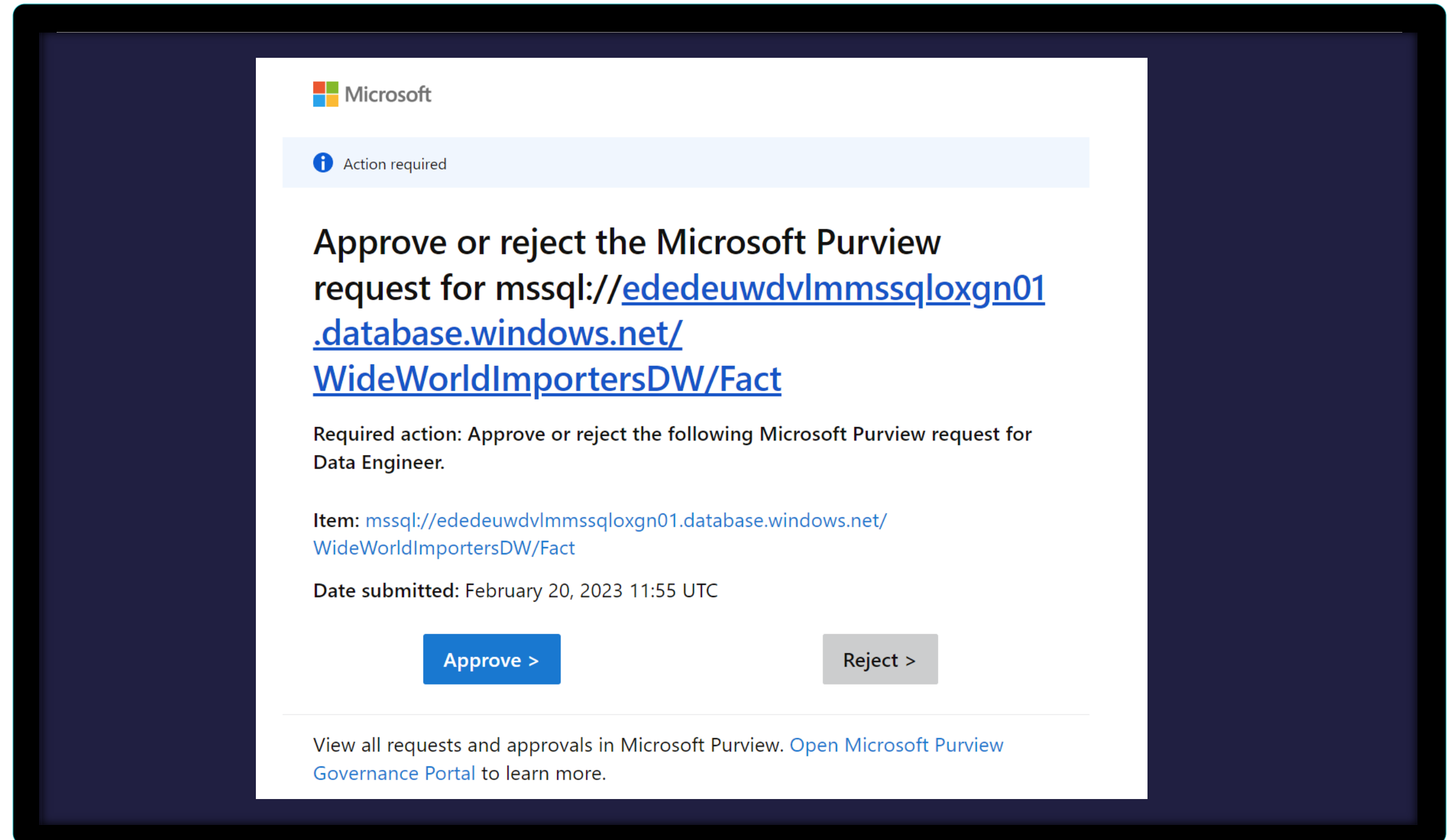
- Approval type \***: A dropdown menu set to 'Approve/Reject - Everyone must approve'.
- Title \***: A text input field containing 'Approval Request for Data Access Request'.
- Assigned to \***: A text input field with the placeholder 'Search by name or email address'.
- Reminder settings**: A toggle switch set to 'On'.
- Reminder interval \***: A dropdown menu set to '1 day'.
- Expiry settings**: A toggle switch set to 'On'.
- Expire After \***: A dropdown menu set to '1 month'.
- Notify on expiration**: A text input field with the placeholder 'Search by name or email address'.

The interface is designed with a clean, modern look, featuring a light gray background and a dark blue header. The settings are organized into a list of fields with labels and input controls. A red box highlights the 'Expire After' and 'Notify on expiration' fields, indicating the expiration functionality.

# Self-service access policies

## Data Access (Governance)

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

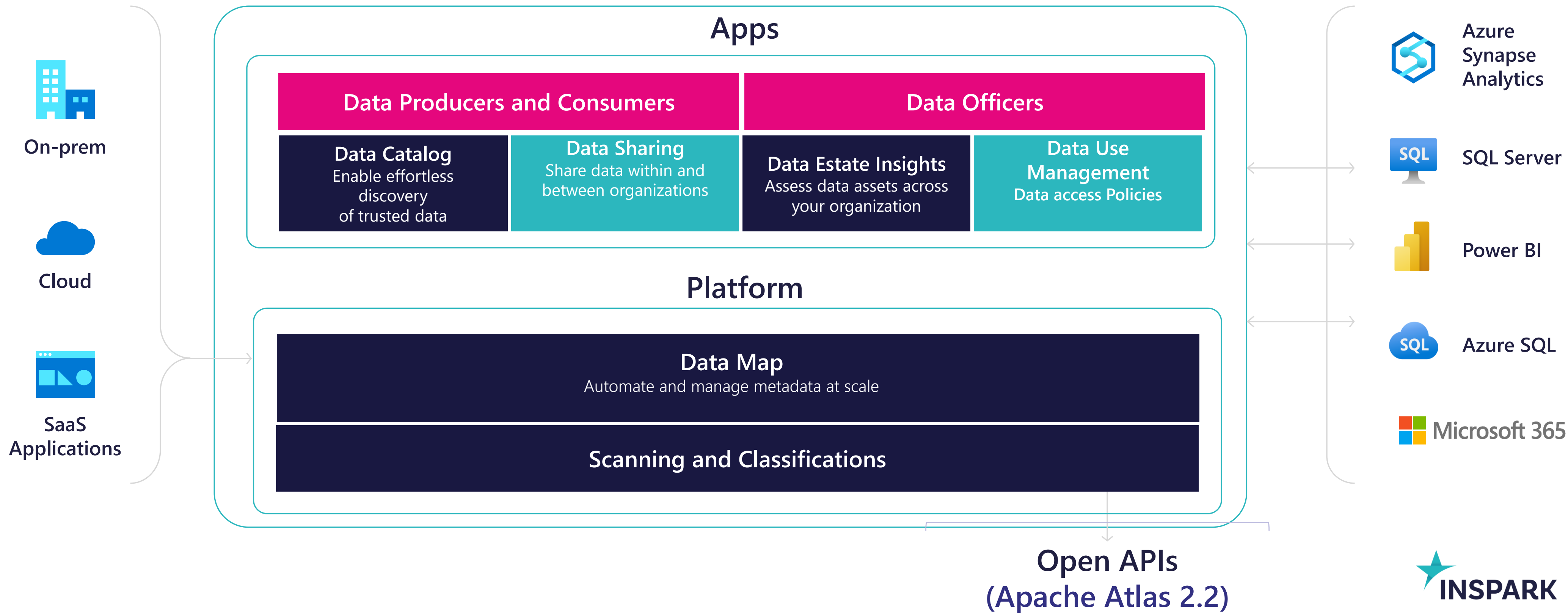


# Microsoft Purview

Generally Available

Preview

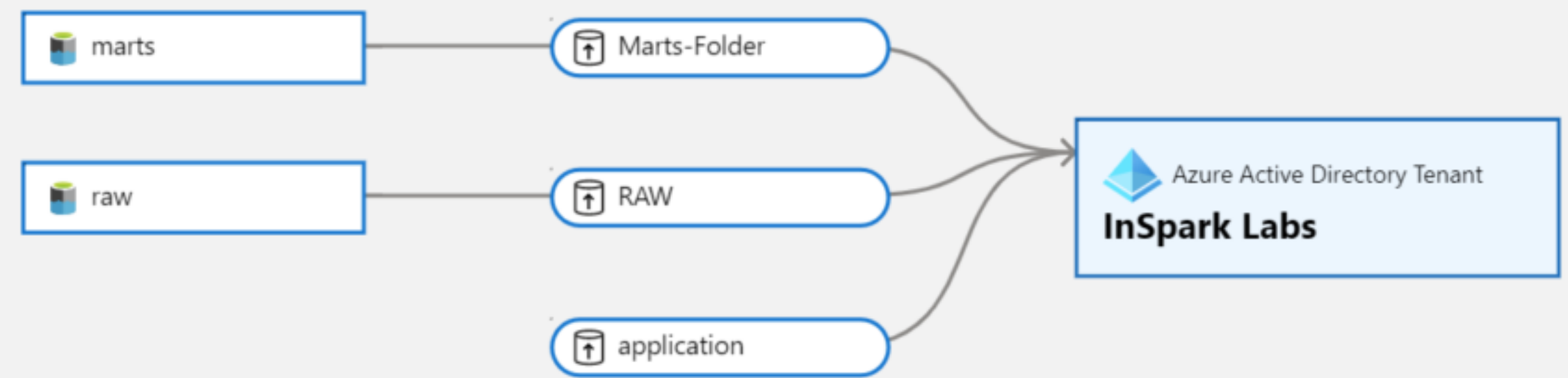
## Unified Data Governance

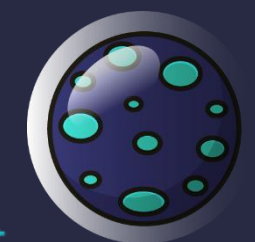
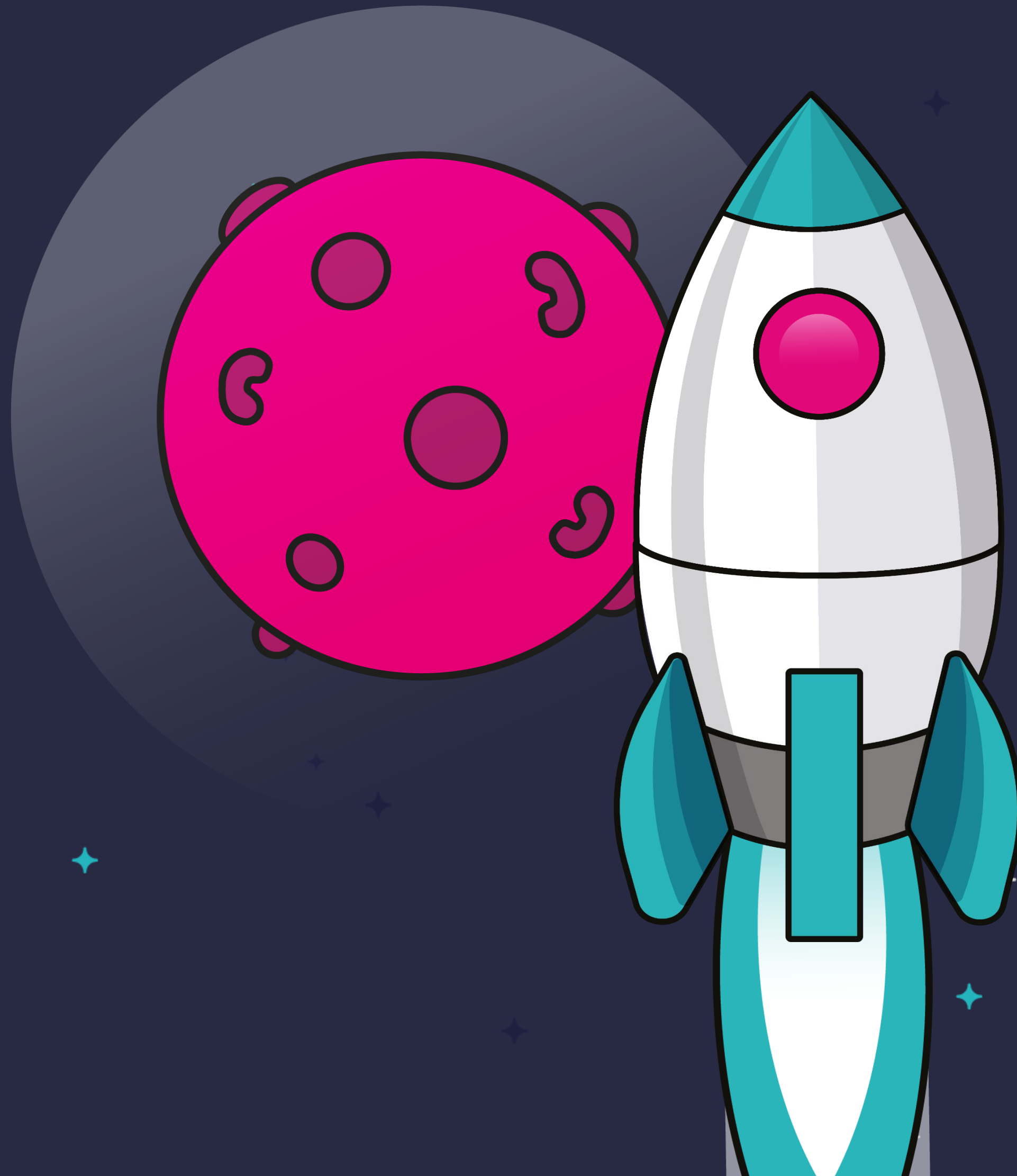


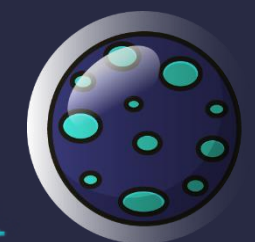
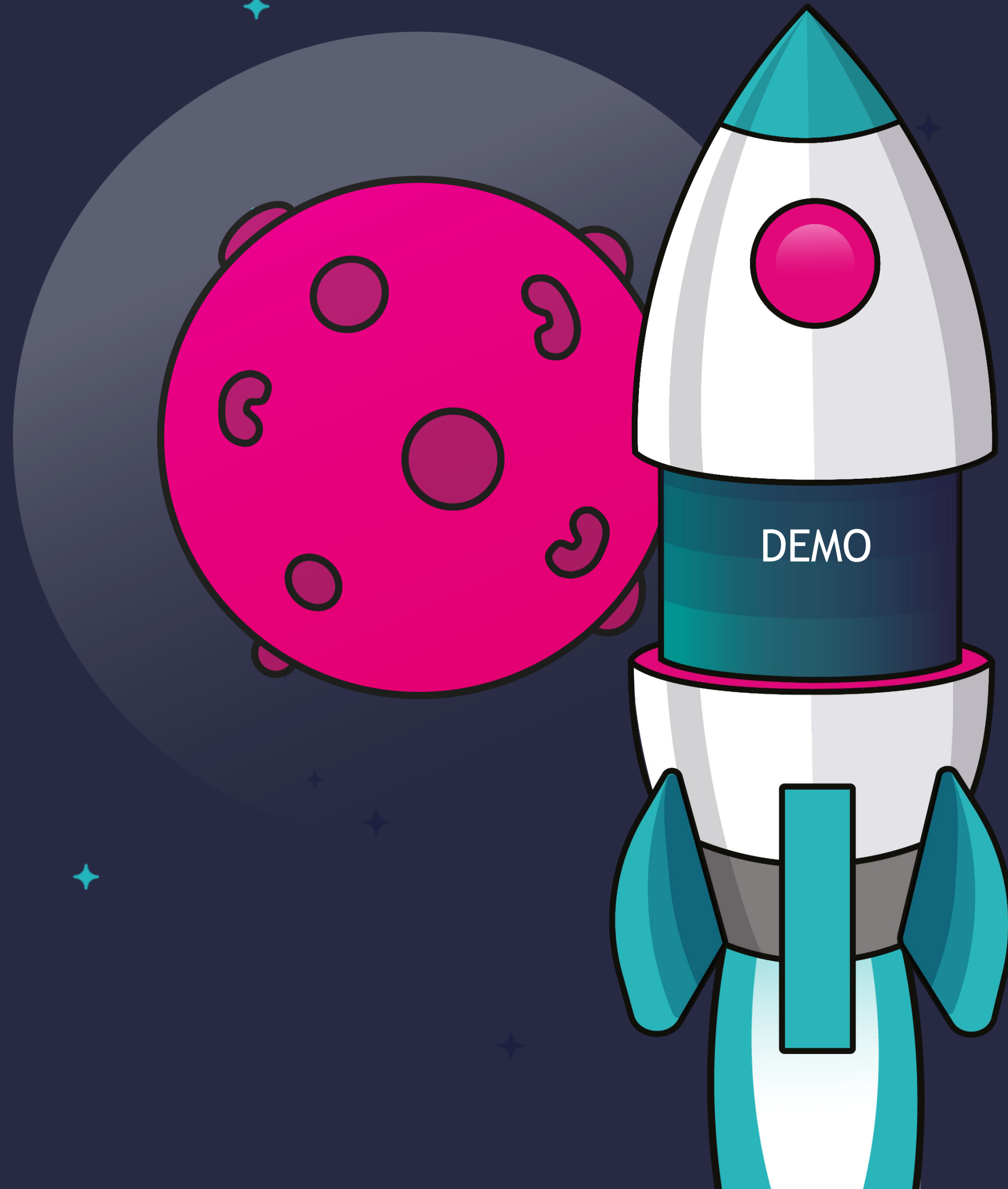


# Sharing

- 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

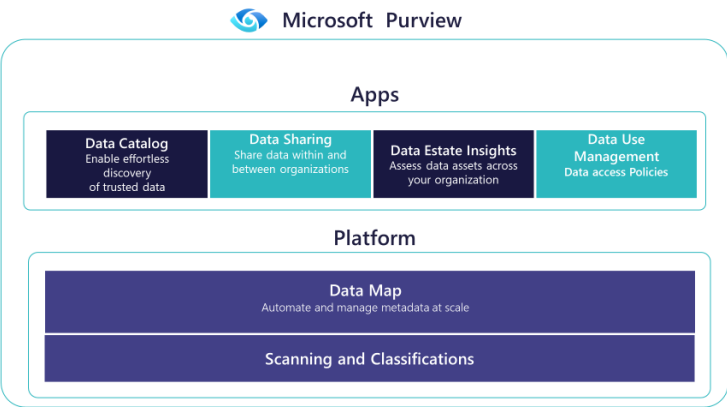
A hand holds a magnifying glass over a white rectangular piece of paper. The word "Pricing" is written on the paper in a stylized, hand-drawn font. The letters "Pri" are blue with a sketchy texture, while the letters "cing" are black with a similar texture. The background is a light gray surface with faint, hand-drawn lines and shapes, including a brick-like pattern in the upper right.

**Pricing**

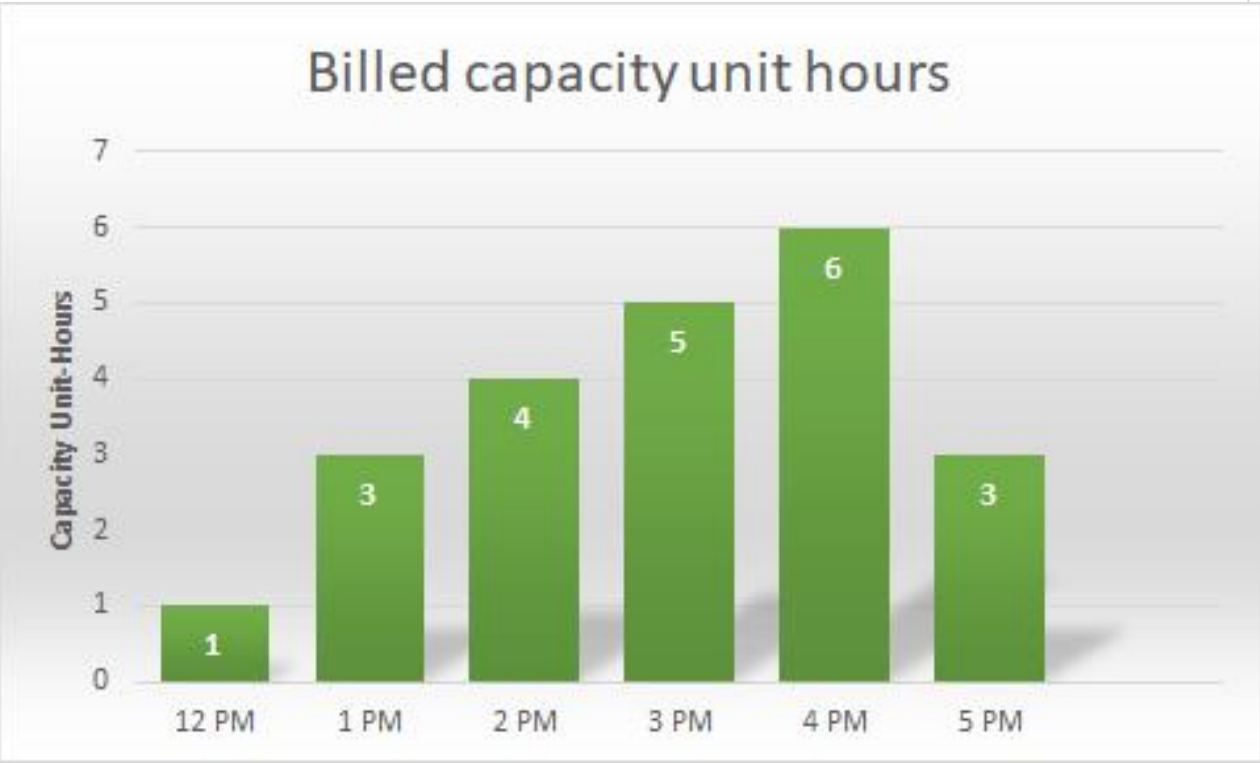
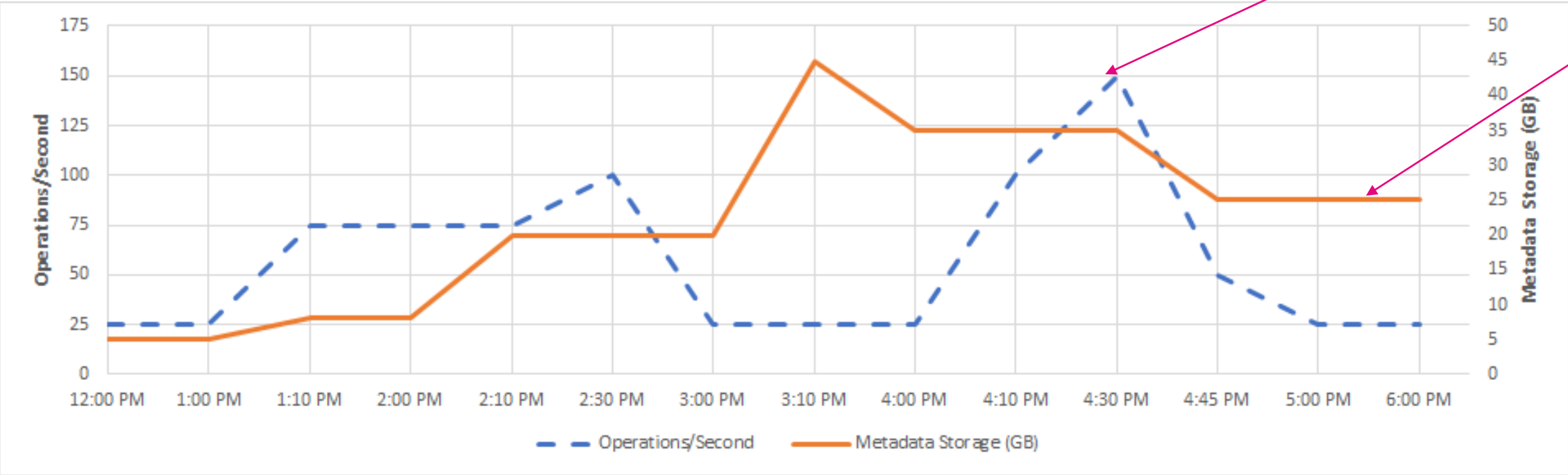
# Data Map Consumption

- 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.

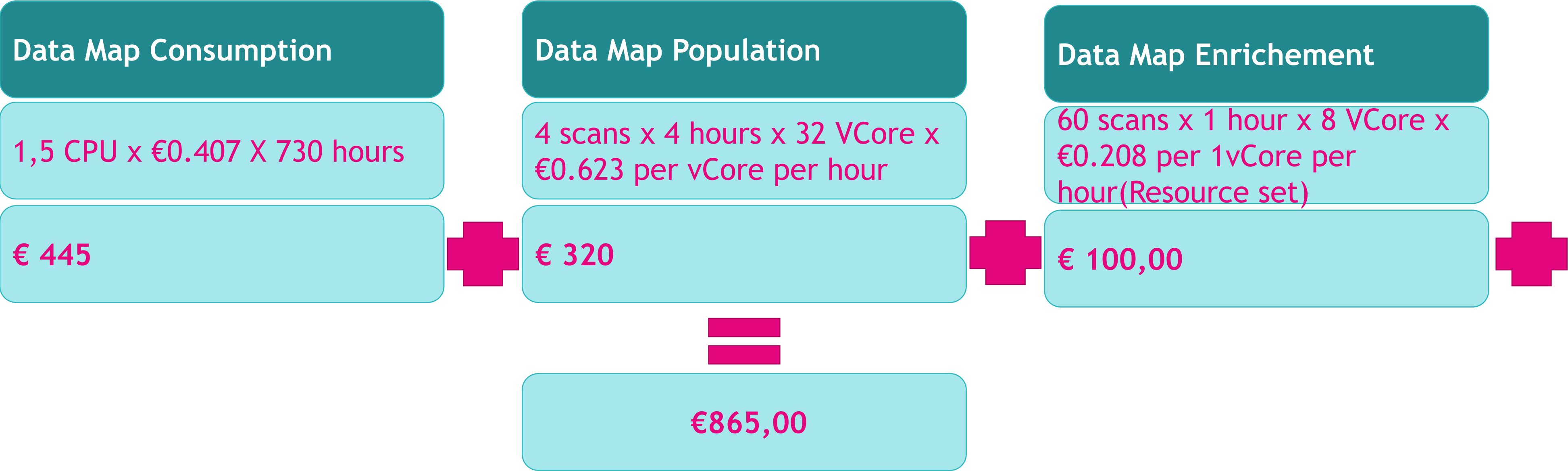


Data Map Capacity Unit	Operations/Sec throughput	Storage capacity in GB
1	25	10
2	50	20
3	75	30
4	100	40
5	125	50
6	150	60
7	175	70
8	200	80
9	225	90
10	250	100
..	..	..
100		





# Pricing - Example



# Take aways / Roadmap

- Microsoft Purview will be added to Tenant Level
- Tag => purview environment = production
- 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?

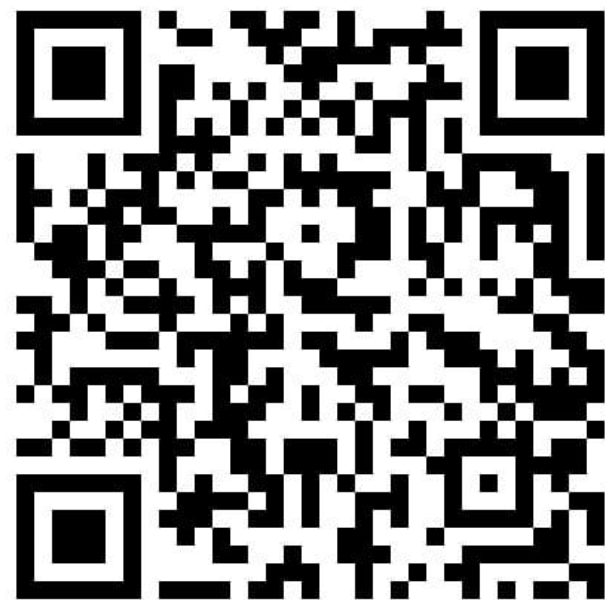


# Thank you for attending!



**Erwin de Kreuk**

Principal Consultant | Lead Data and AI  
| Data Platform MVP | Public Speaker ...



**Erwin de Kreuk**

Principal Consultant - Data & AI  
InSpark



@ErwinDeKreuk



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



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



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

