Streamline Data Governance with Microsoft Purview

Erwin de Kreuk
Principal Consultant – Lead Data & AI
InSpark

@erwindekreuk
linkedin.com/in/erwindekreuk
erwindekreuk.com
github.com/edkreuk

Microsoft® Most Valuable Professional

Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark

Erwin de Kreuk
Principal Consultant - Lead Data & AI
InSpark

@erwindekreuk
linkedin.com/in/erwindekreuk
erwindekreuk.com
github.com/edkreuk
https://sessionize.com/erwin-de-kreuk/

Let's connect

Microsoft® Most Valuable Professional

We Are InSpark

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

Session Objectives

- 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

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,

Microsoft Azure

Elements of successful data governance

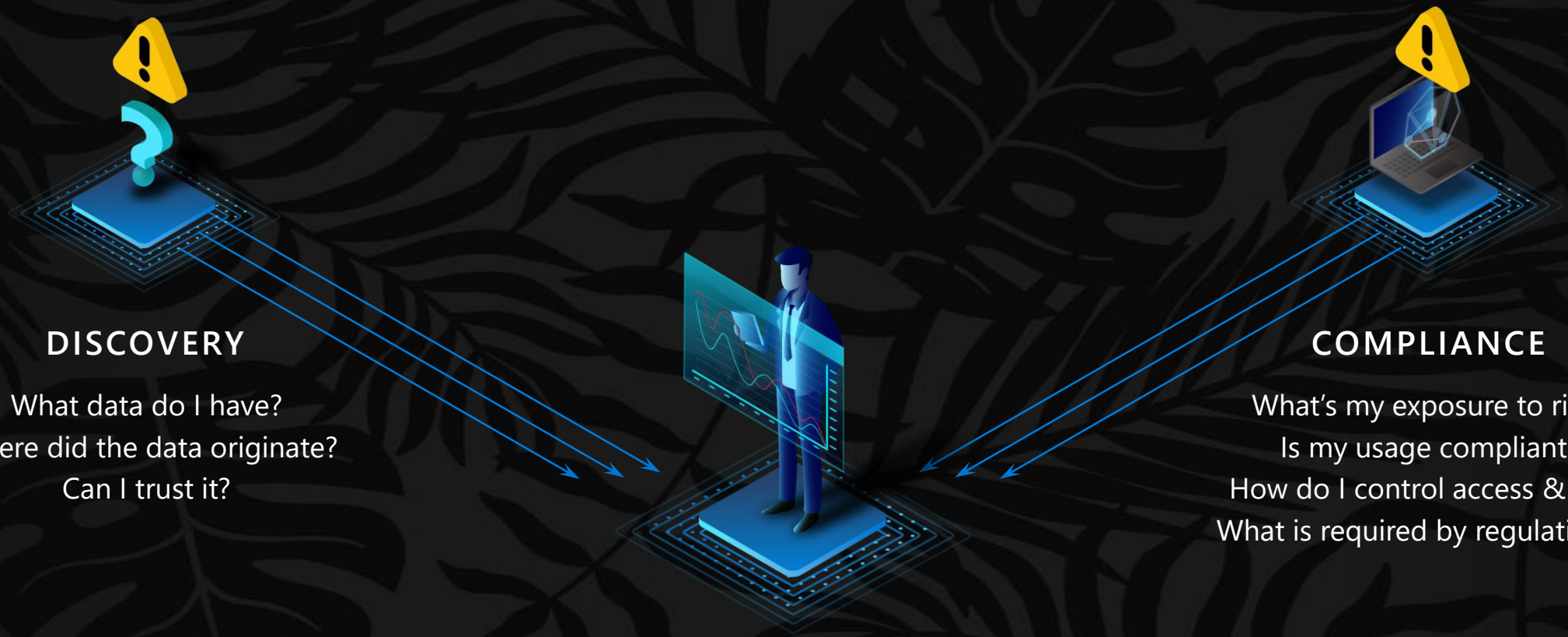
Manage growing data landscape

Overcome operational silos

Increase data agility

Comply with industry regulations

Data operation & governance
is becoming increasingly interdisciplinary



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

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.

History

- ADC Gen 1
- **BlueTalon Acquisition**

June 2019

- ADC Gen 2
- Azure Purview

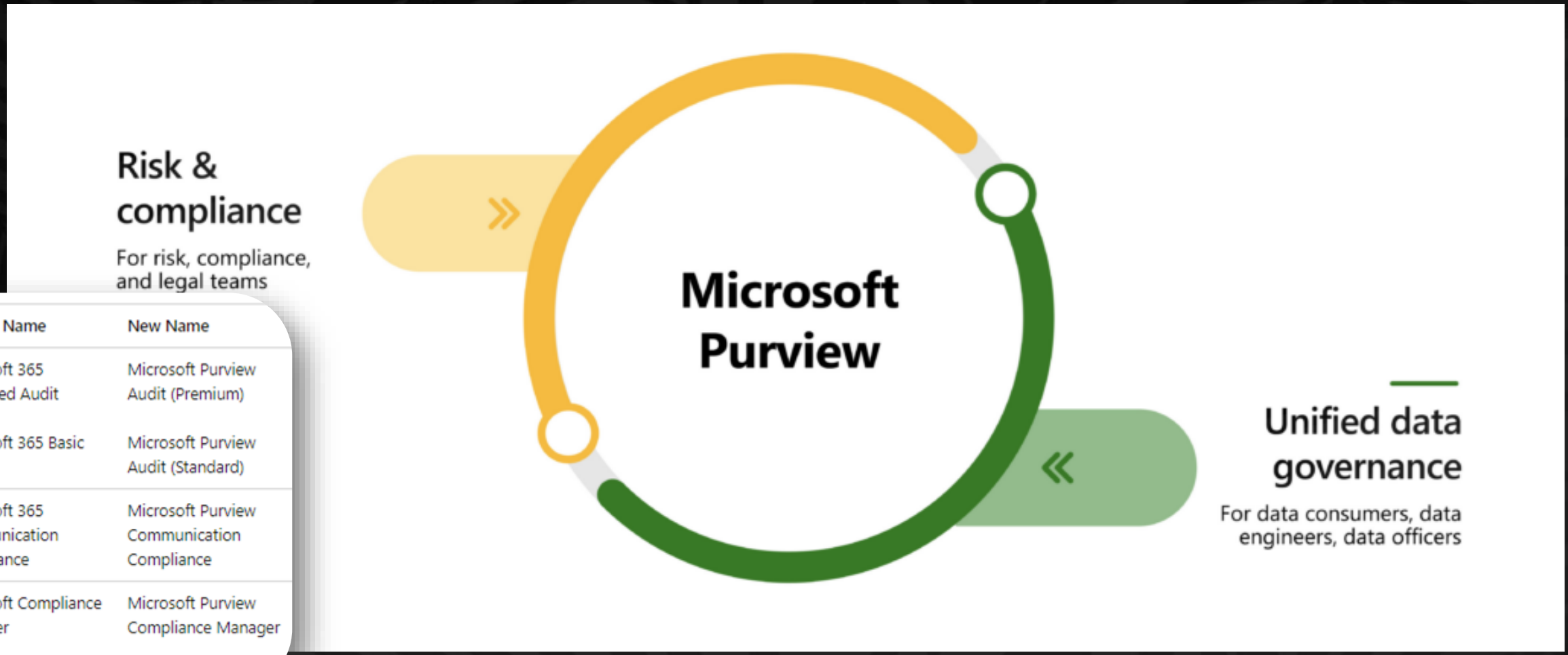
sept 2021

- Microsoft Purview

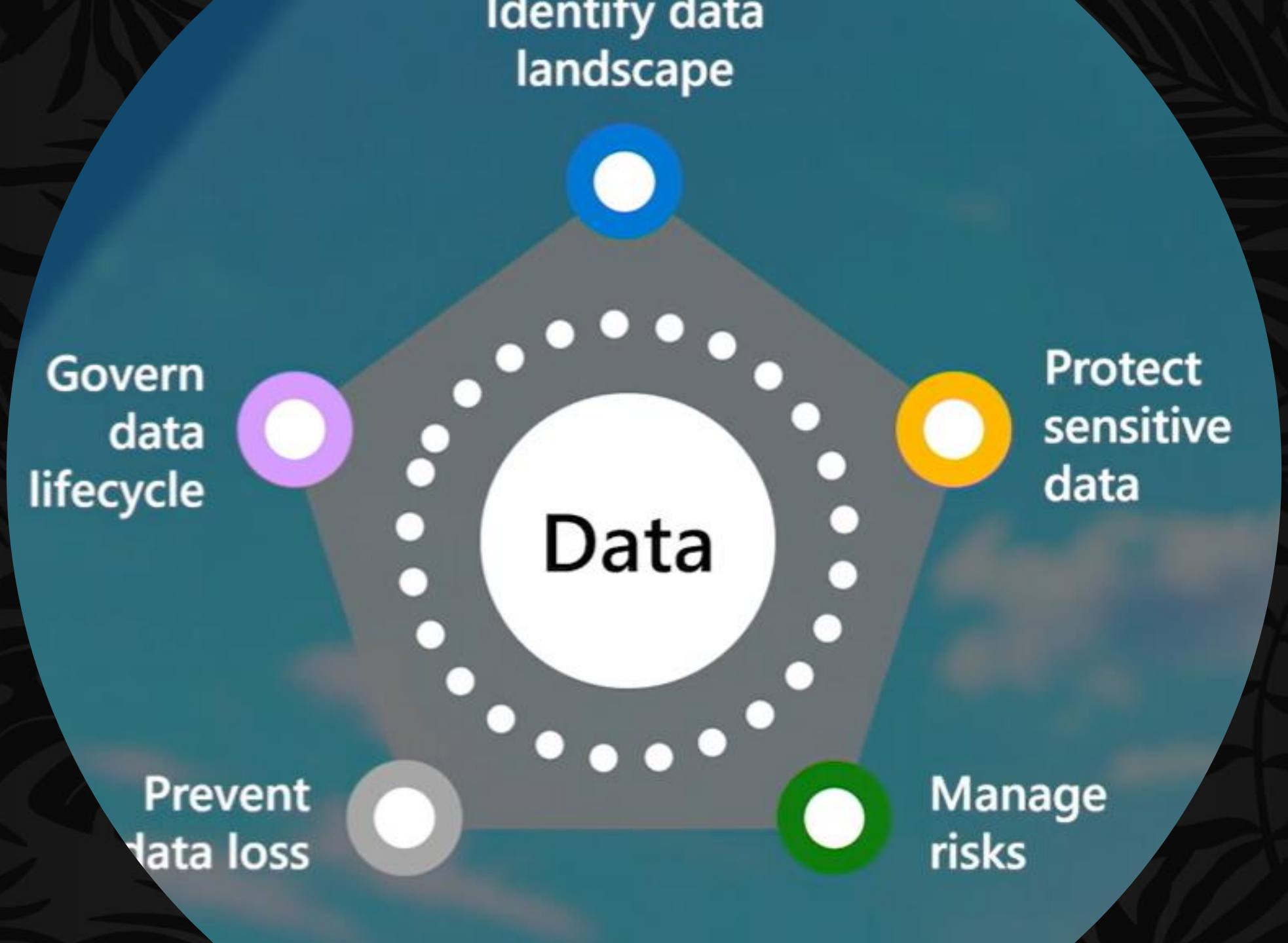
april 2022

Microsoft Purview

The future of compliance and data governance



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

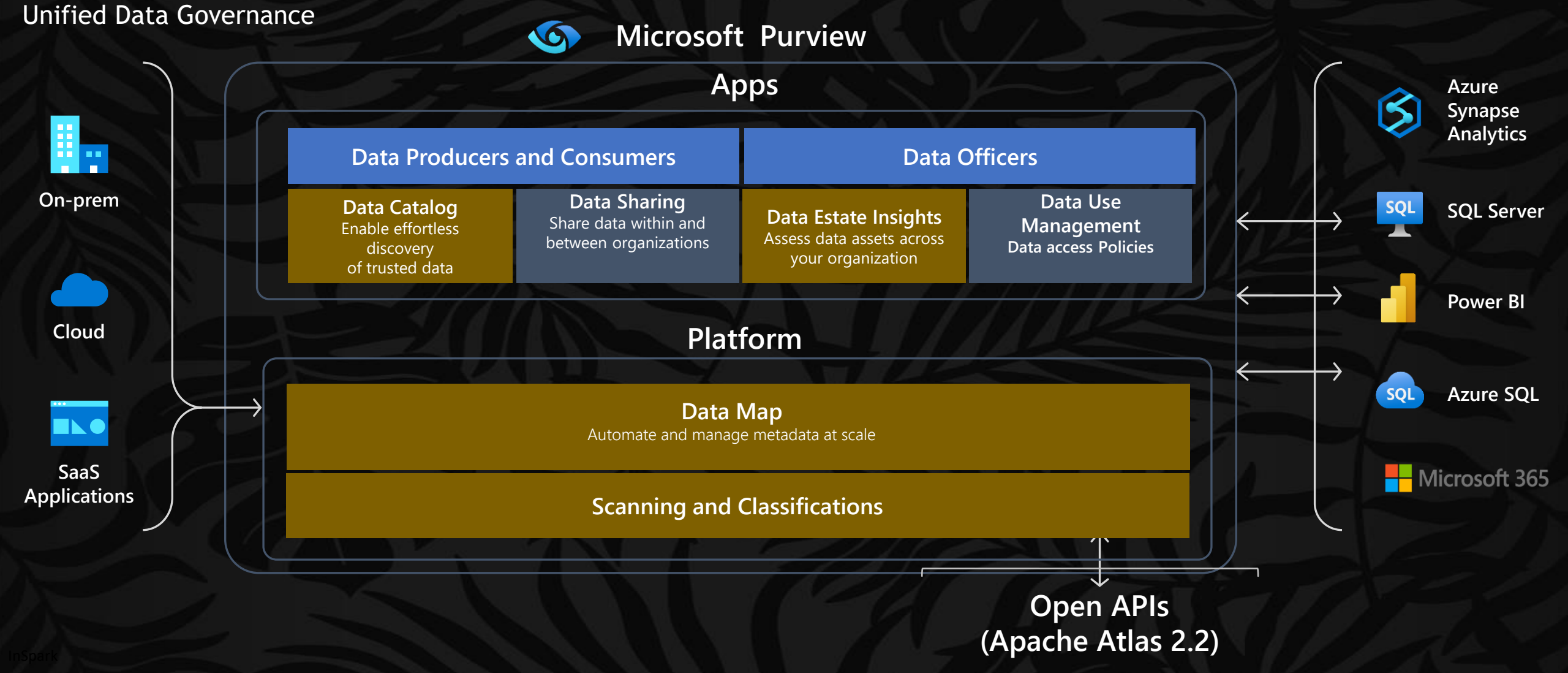

Identify data landscape
Govern data lifecycle
Protect sensitive data
Data
Prevent data loss
Manage risks

Microsoft Purview

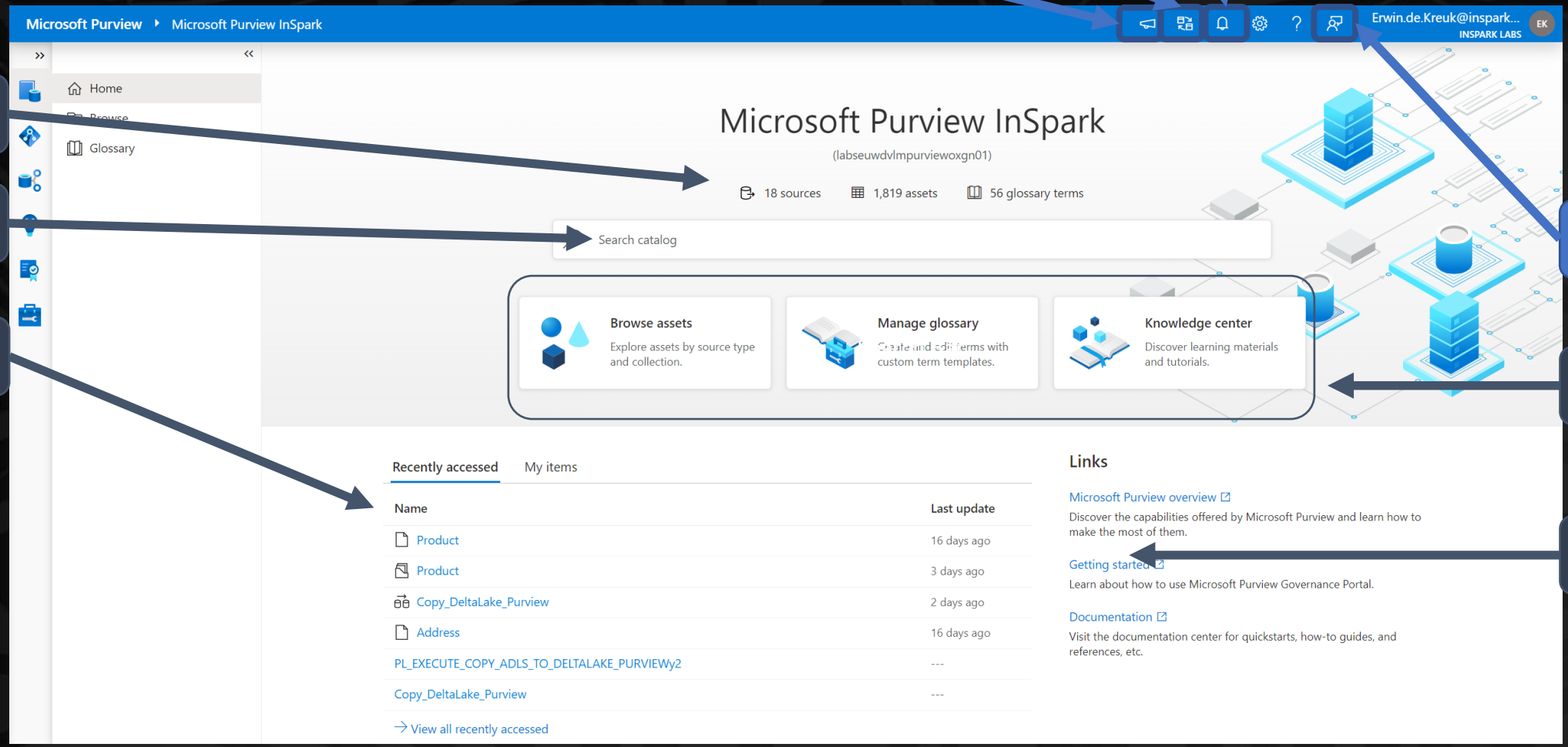
Generally Available

Preview

Unified Data Governance



Purview Studio



Updates

Accounts

Notifications

Metrics

Search Bar

Recently Accessed Entities

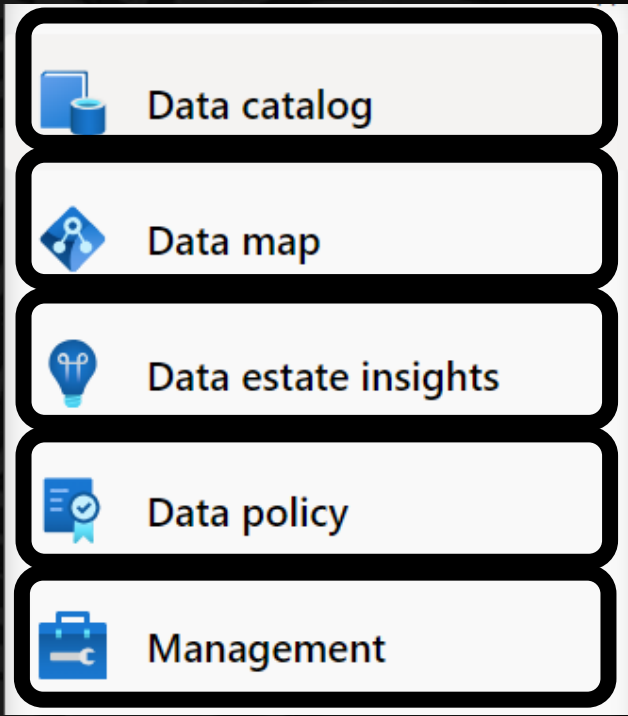
Feedback

Key Actvities

Usefull Links

Purview Studio

- 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.

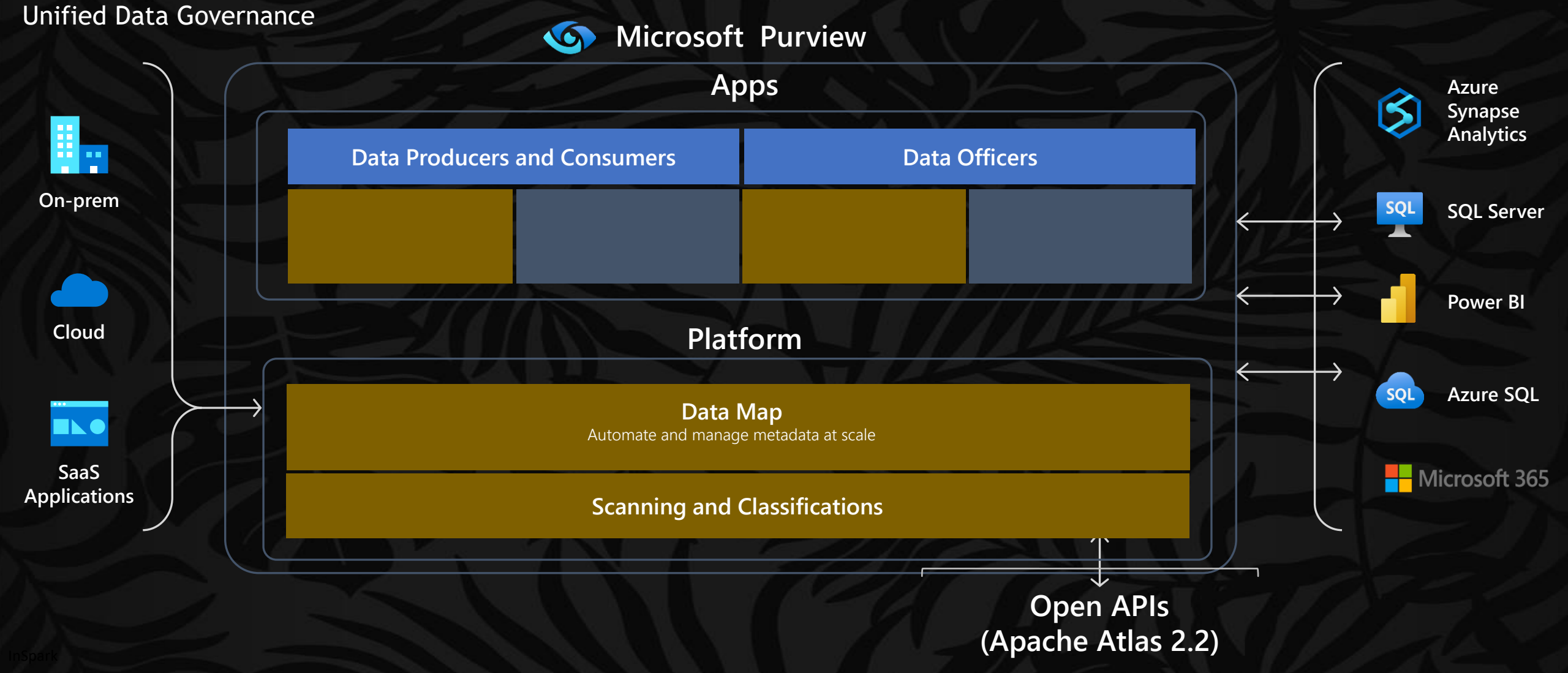


Microsoft Purview

Generally Available

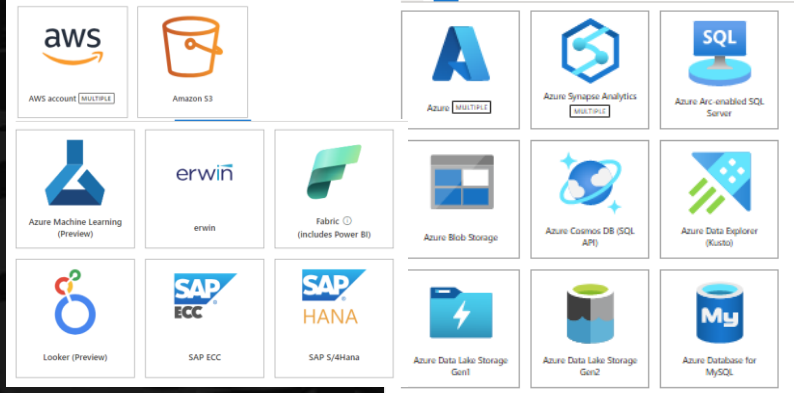
Preview

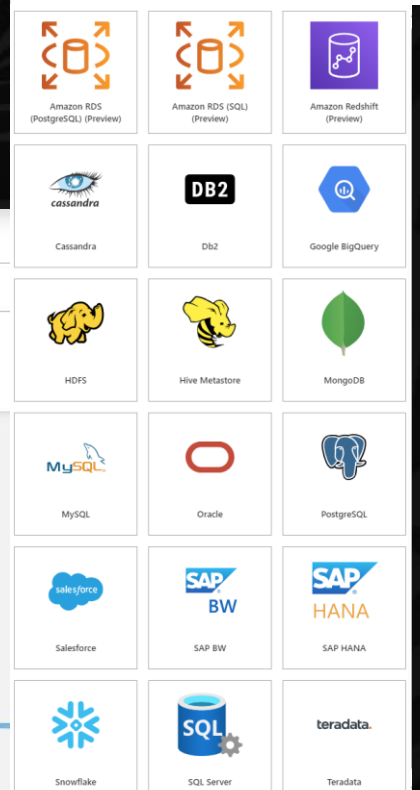
Unified Data Governance

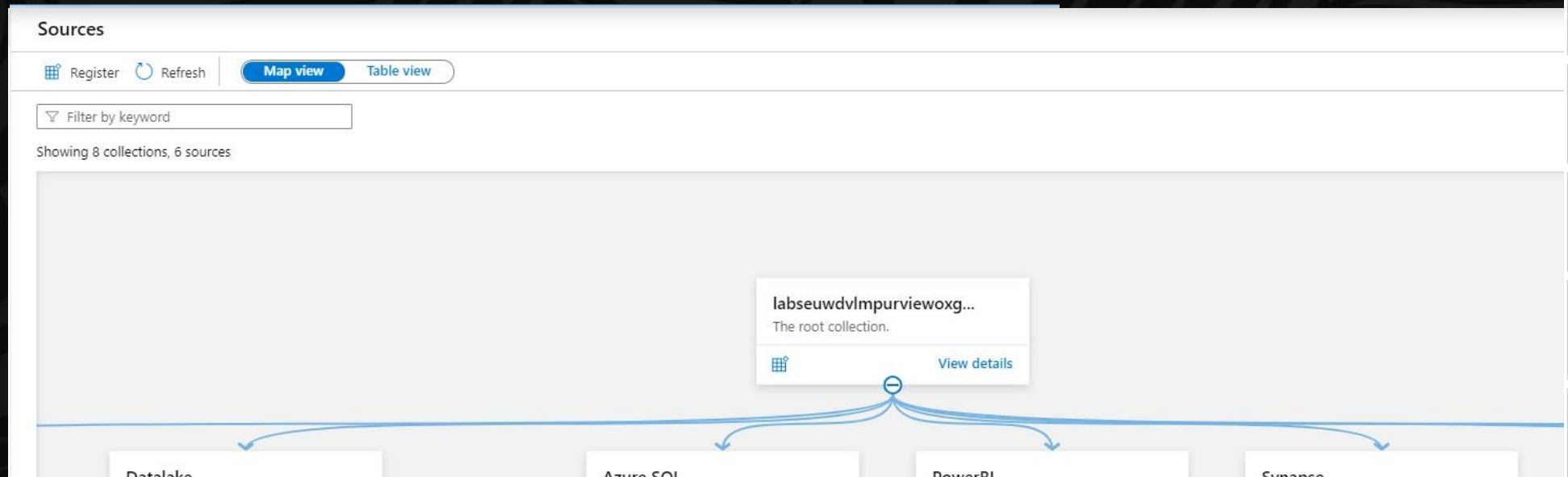


Collections and Sources

- A graph describing the data assets and their relationships across data estate
- Support for 41+ data sources and growing

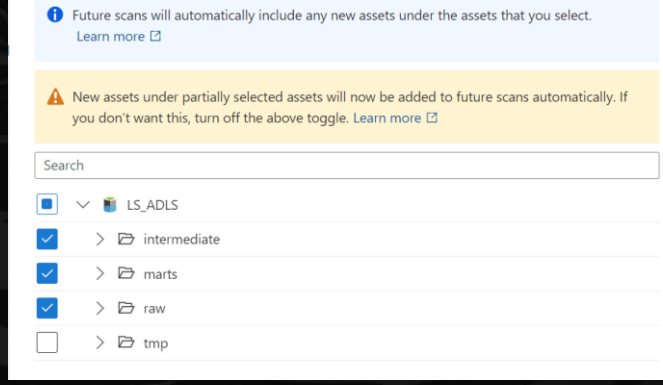


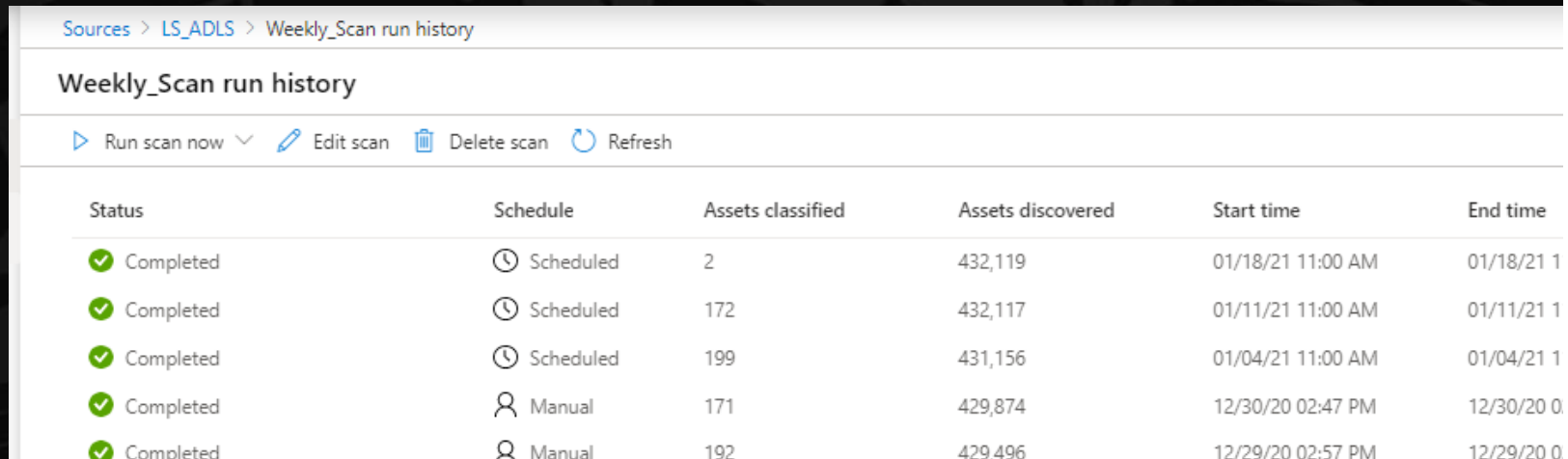


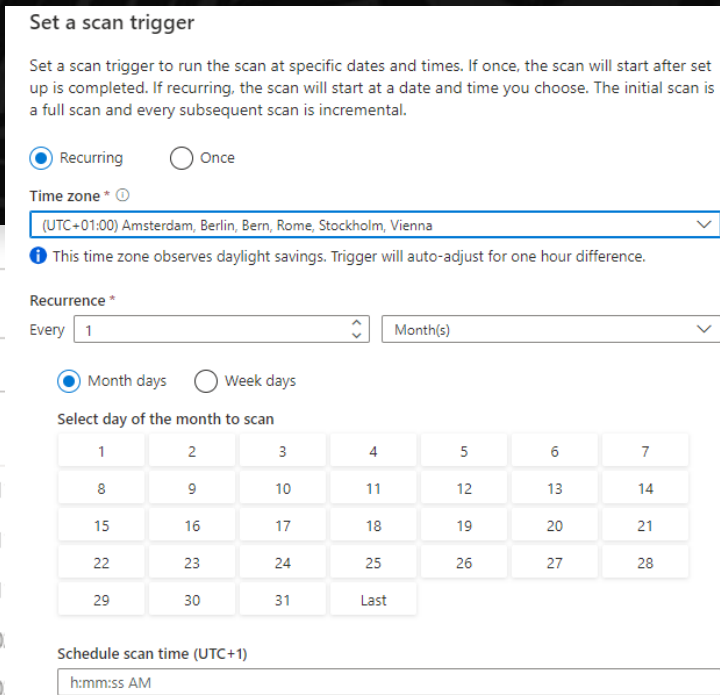


Scans

- A graph describing the data assets and their relationships across your data estate
- Support for 41+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores

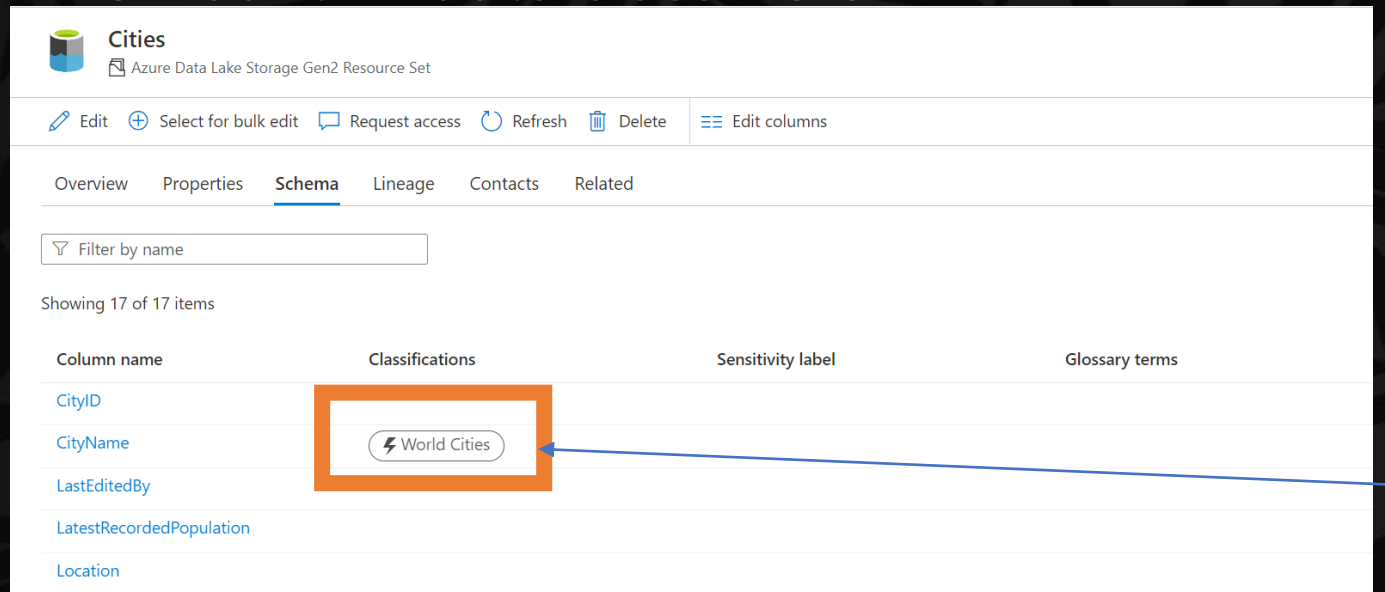


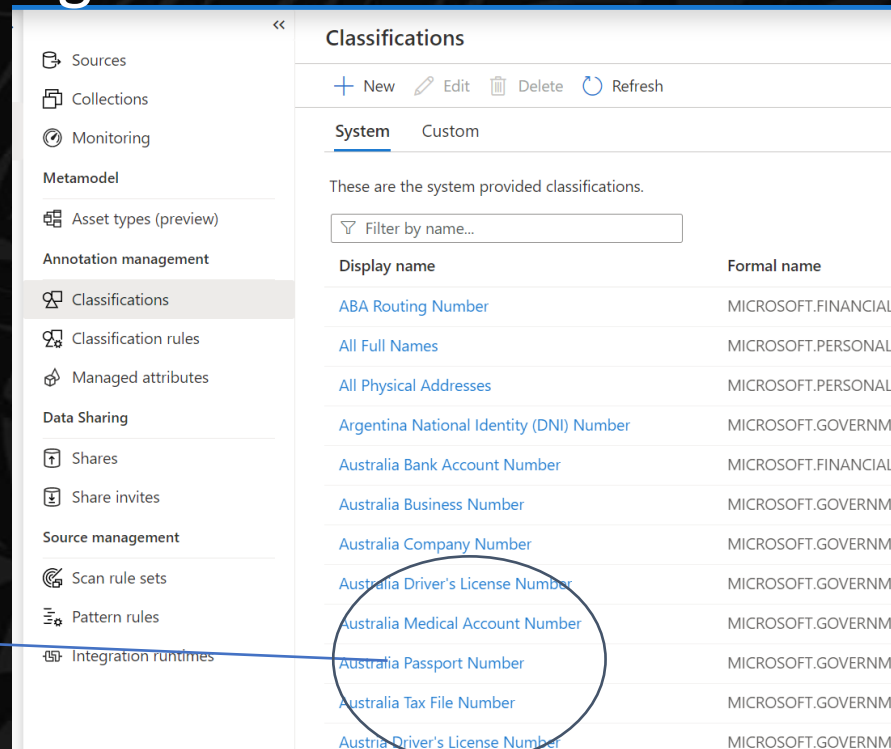




Classifications

- A graph describing the data assets and their relationships across your data estate
- Support for 41+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores
- 220+ built-in data classifiers



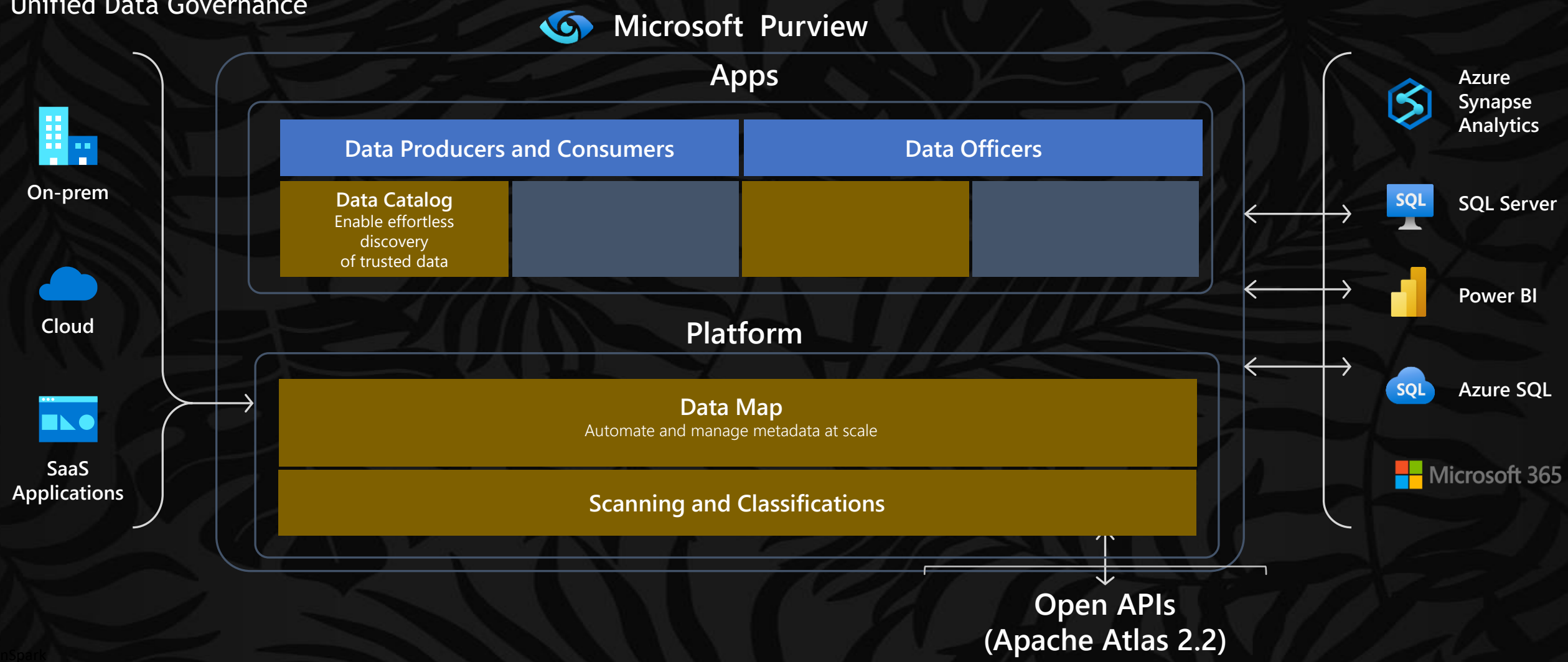


Microsoft Purview

Generally Available

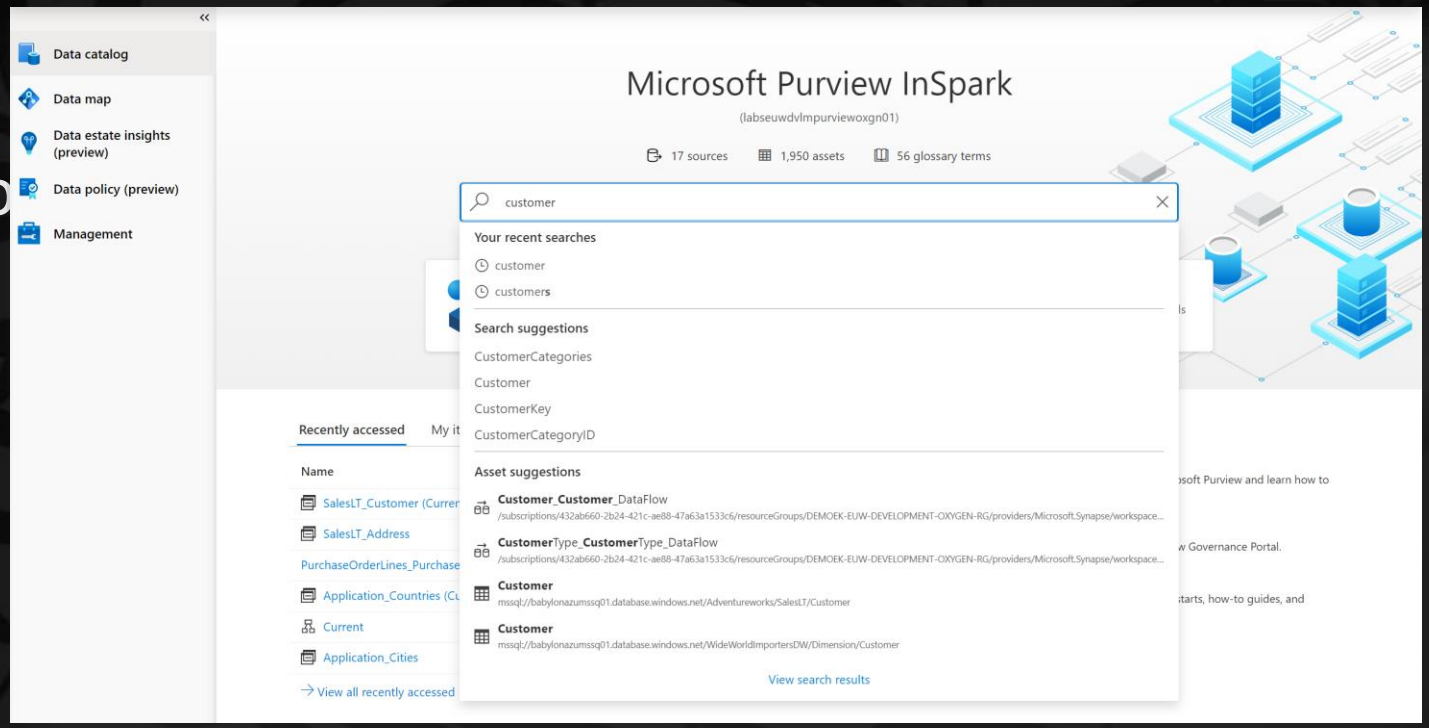
Preview

Unified Data Governance



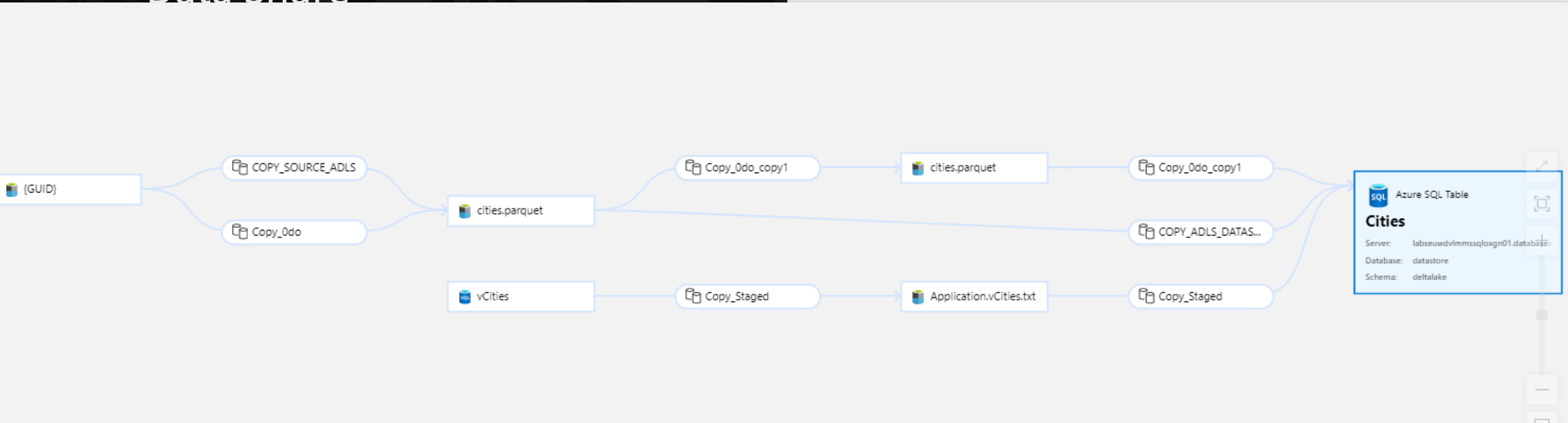
Search

- Semantic search and browse
- Business glossary and workflo



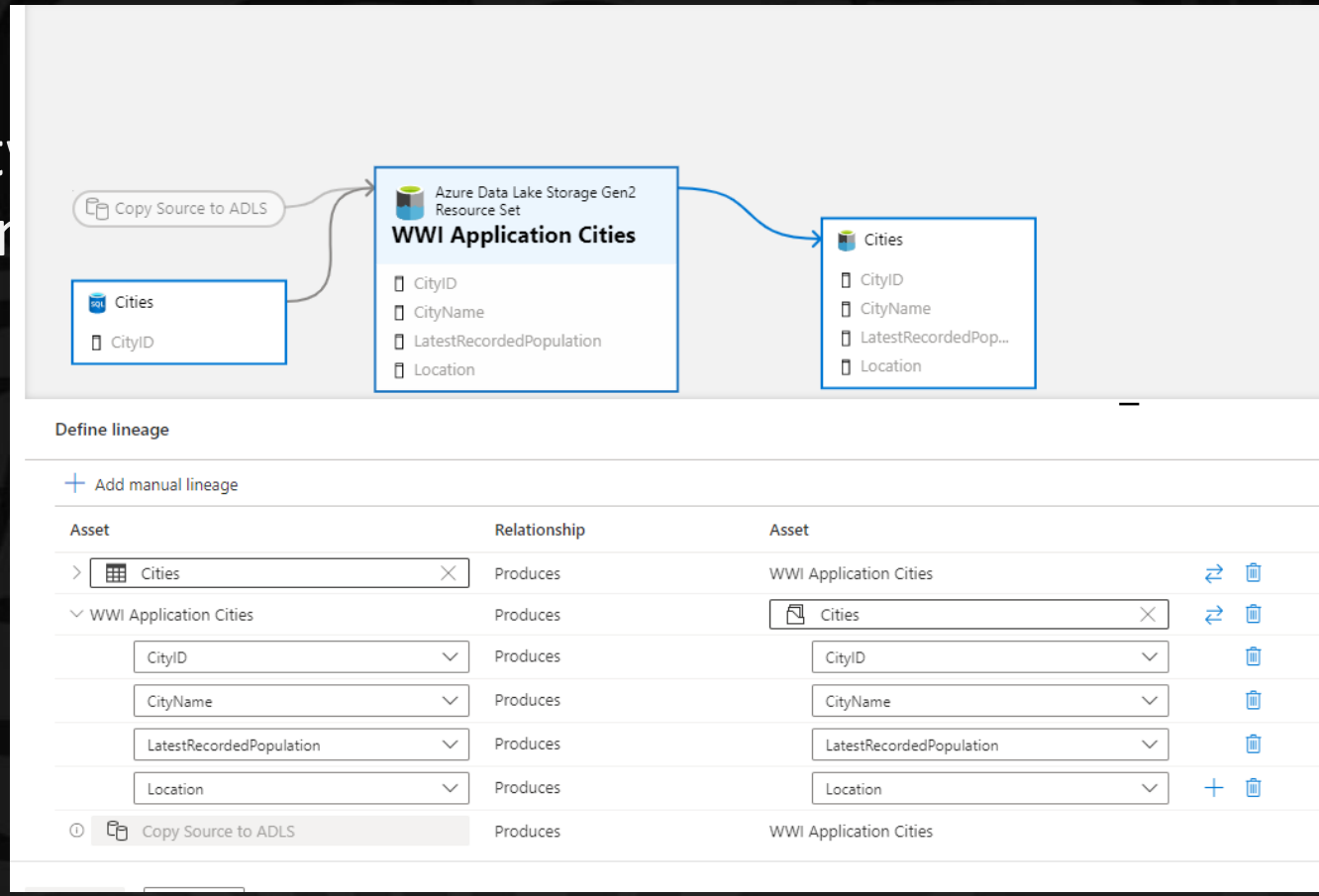
Lineage

- Semantic search and browse
- Business glossary and workflows
- Data lineage with sources, owners
- Pushed from Azure Synapse Anayt
  Data Share

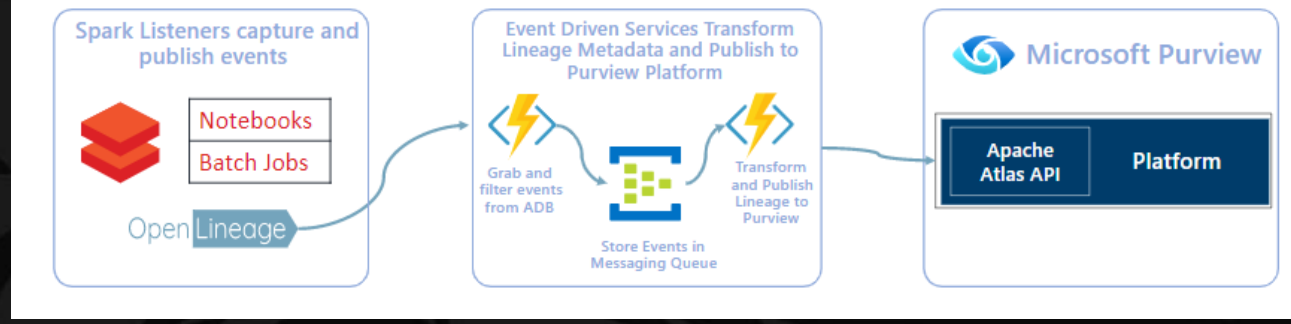


Manual Lineage

- Create Lineage in the Catalog bet
- Additionally configure the columr

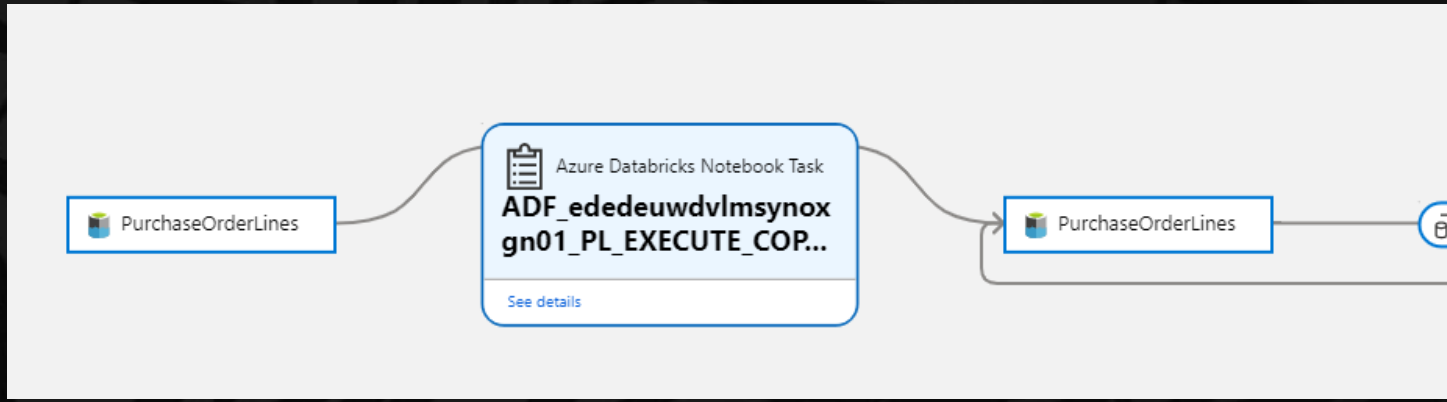


Custom Lineage



- OpenLineagefor Azure DataBricks

  Currently, the solution does not support pushing lineage to a Private Endpoint backed Microsoft Purview service.
- Column Level Mapping Supported Sources (ABFSS, WASBS, and default metastore hive tables)
- Column Mapping Support for Delta Format
  - Delta Merge statements are not supported at this time
  - Delta to Delta is NOT supported at this time

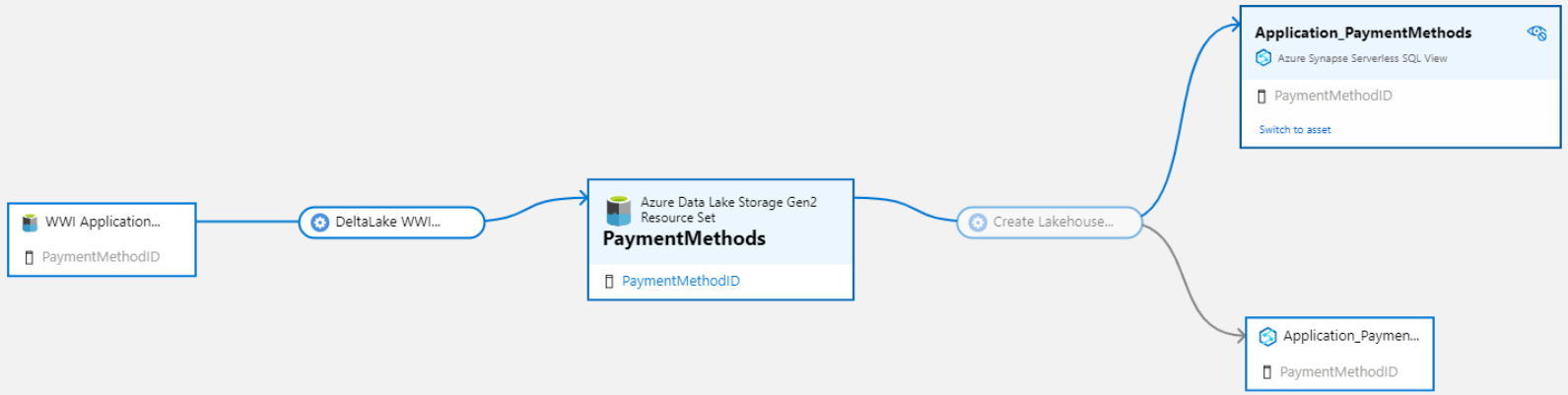


microsoft/Purview-ADB-Lineage-Solution-Accelerator: A connector to ingest Azure Databricks lineage into Microsoft Purview (github.com)

Custom Lineage

- Create custom Lineage with Apache Spark 2.2
- Connecting 2 entities with a process entity

- https://github.com/wjohnson/pyapacheatlas

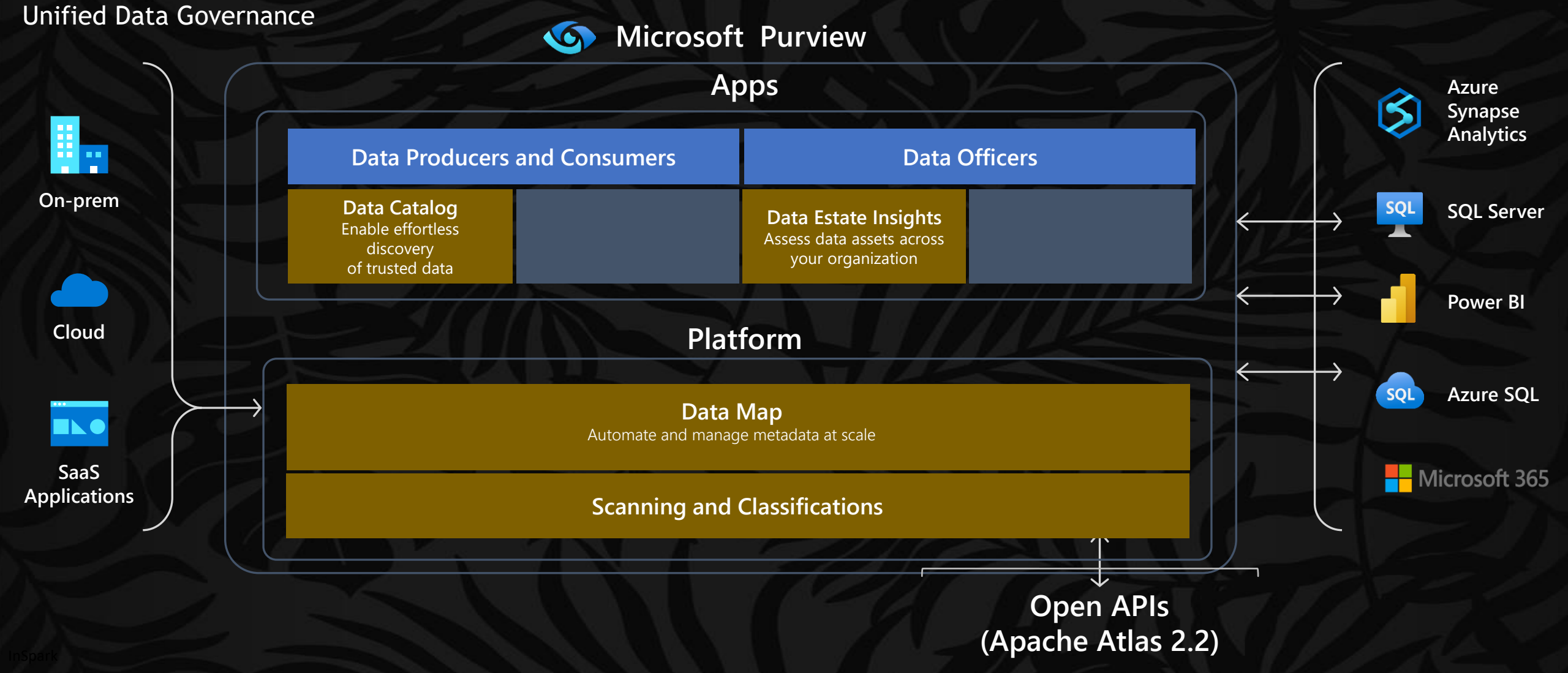


Microsoft Purview

Generally Available

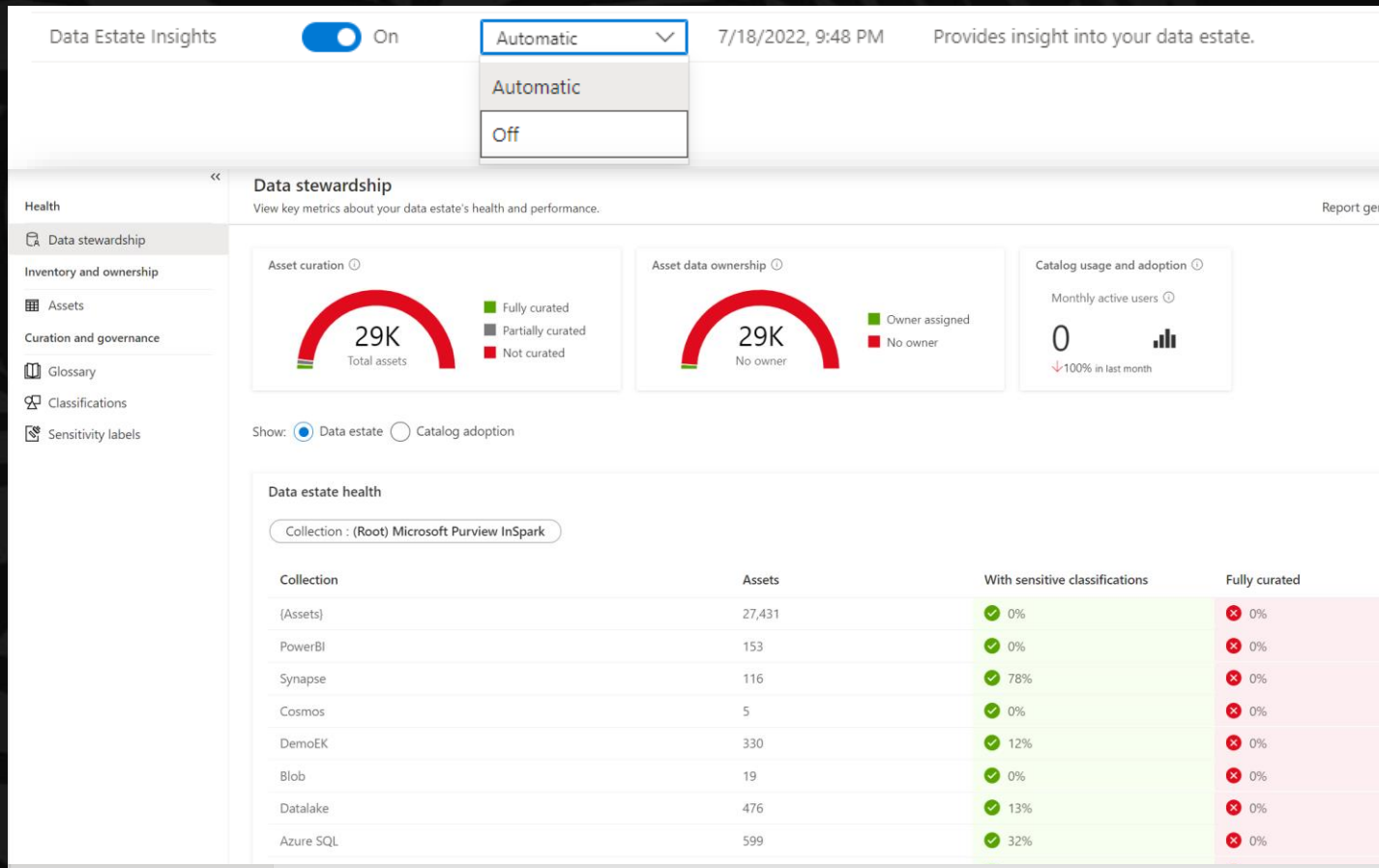
Preview

Unified Data Governance



Data Estate Insight

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



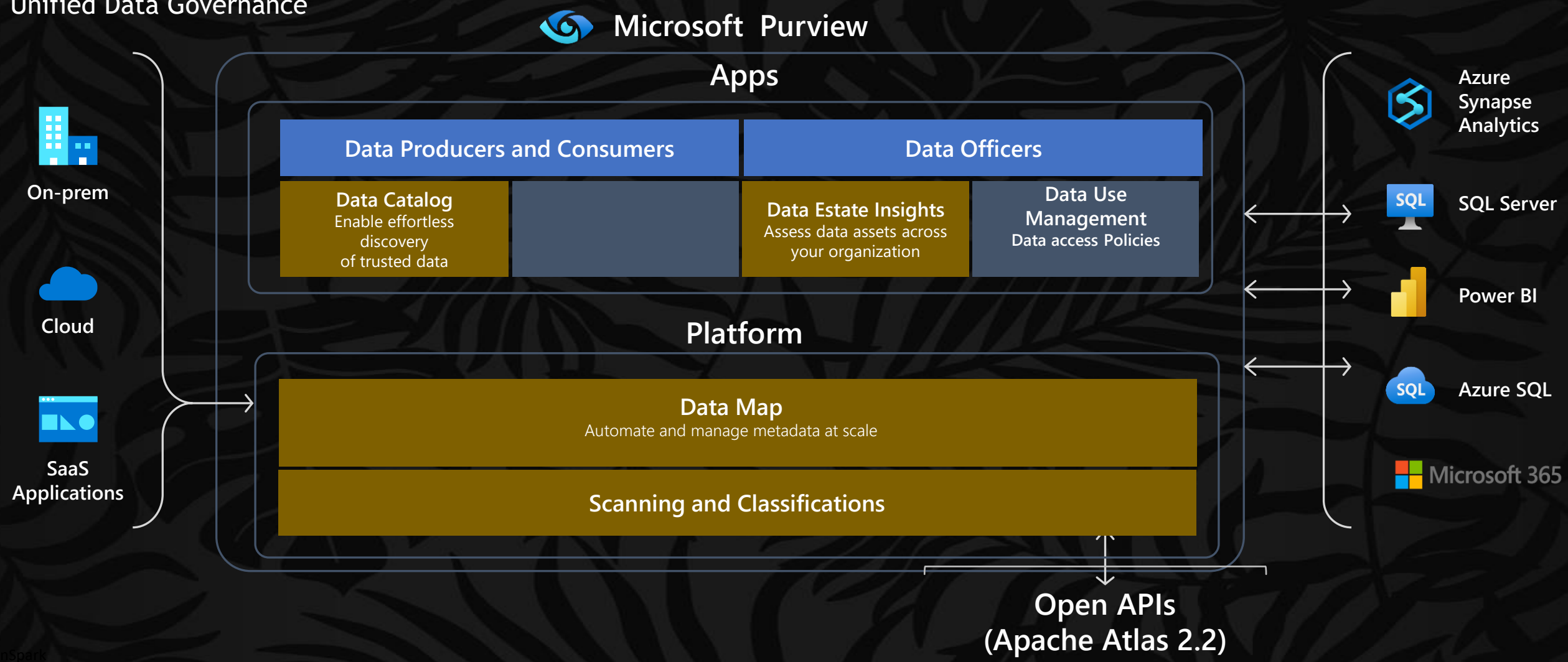
Data Owner
Dev Ops
Self Service access

Microsoft Purview

Generally Available

Preview

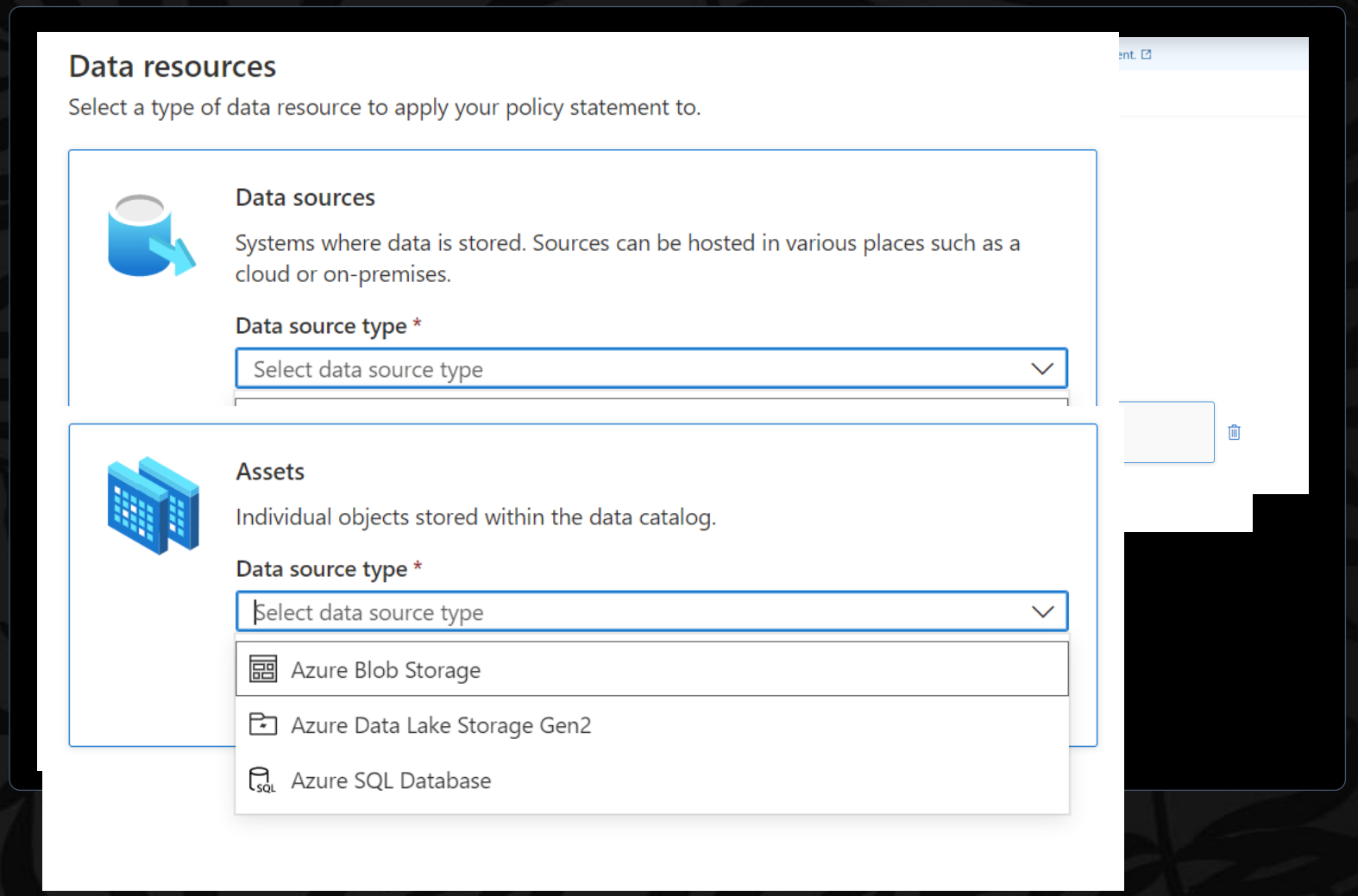
Unified Data Governance



Data Owner policies

Currently in Preview

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

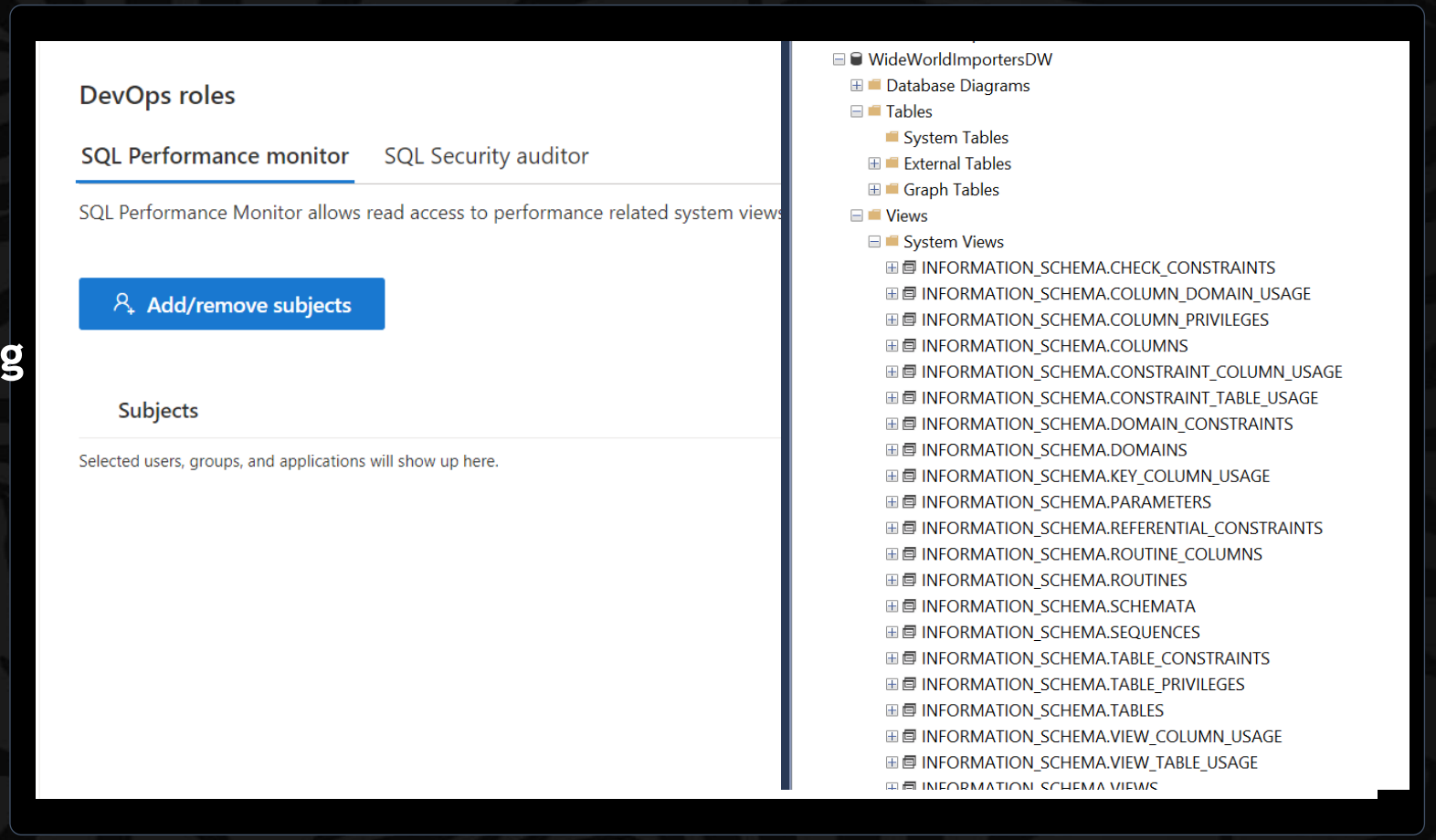


Dev Ops Policies

- **Manage Access to System Metadata**

SQL Performance Monitoring

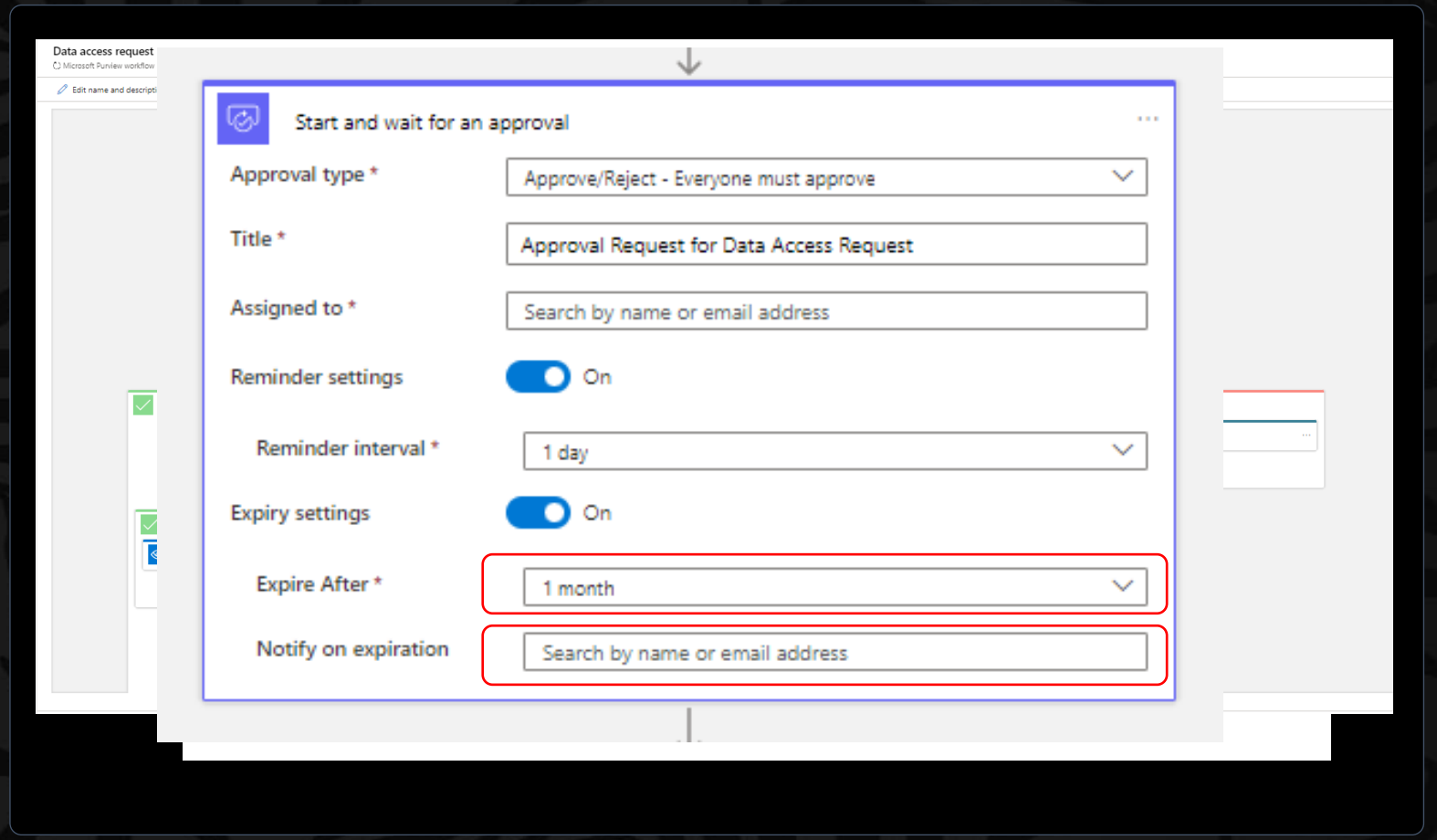
SQL Security Auditing



Self-service access policies

Data Access (Governance)

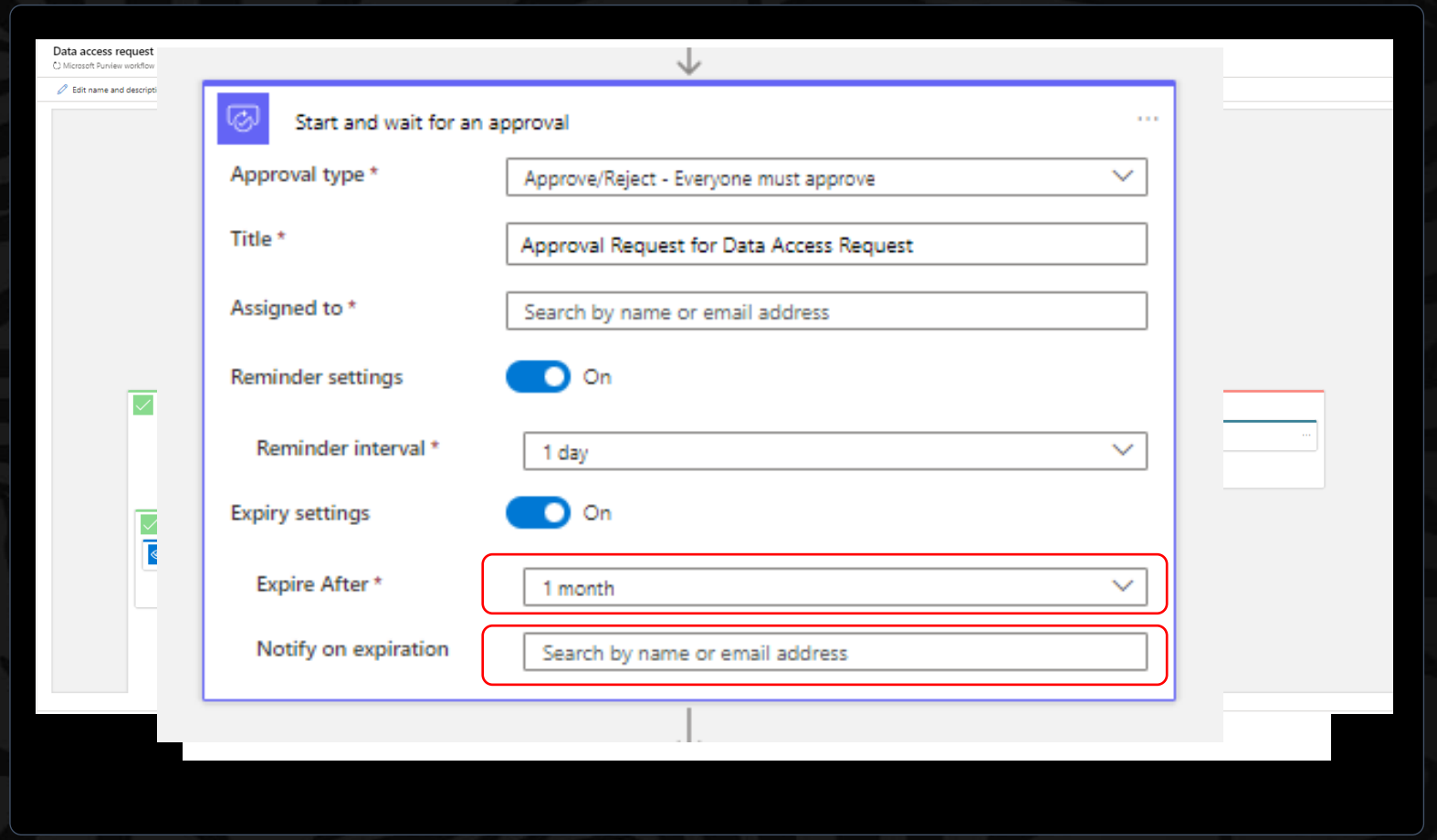
- **Azure SQL Database**
- **Azure Storage**
- **Expiration Functionality**



Self-service access policies

Data Access (Governance)

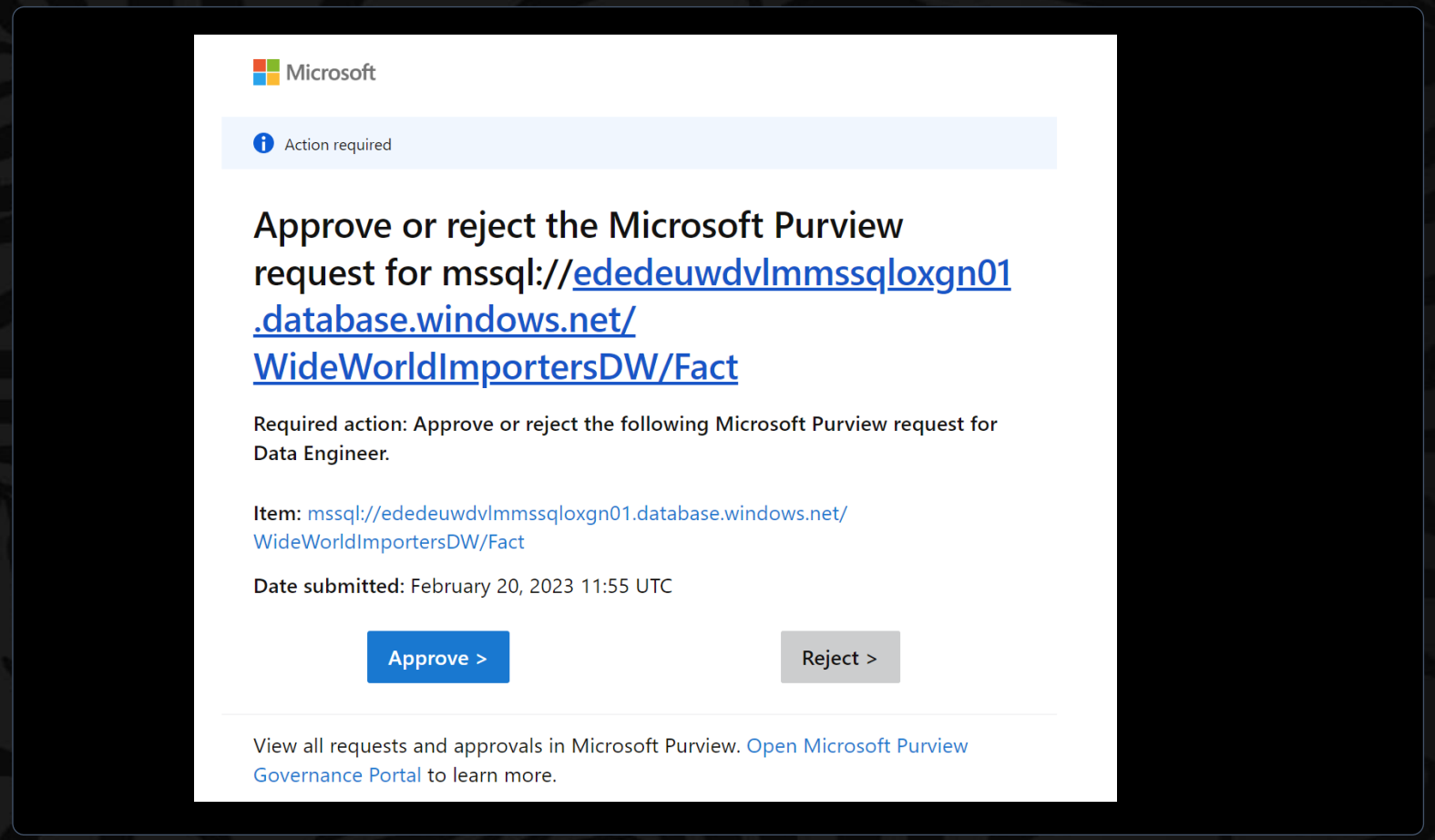
- **Azure SQL Database**
- **Azure Storage**
- **Expiration Functionality**



Workflows

Data Catalog

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

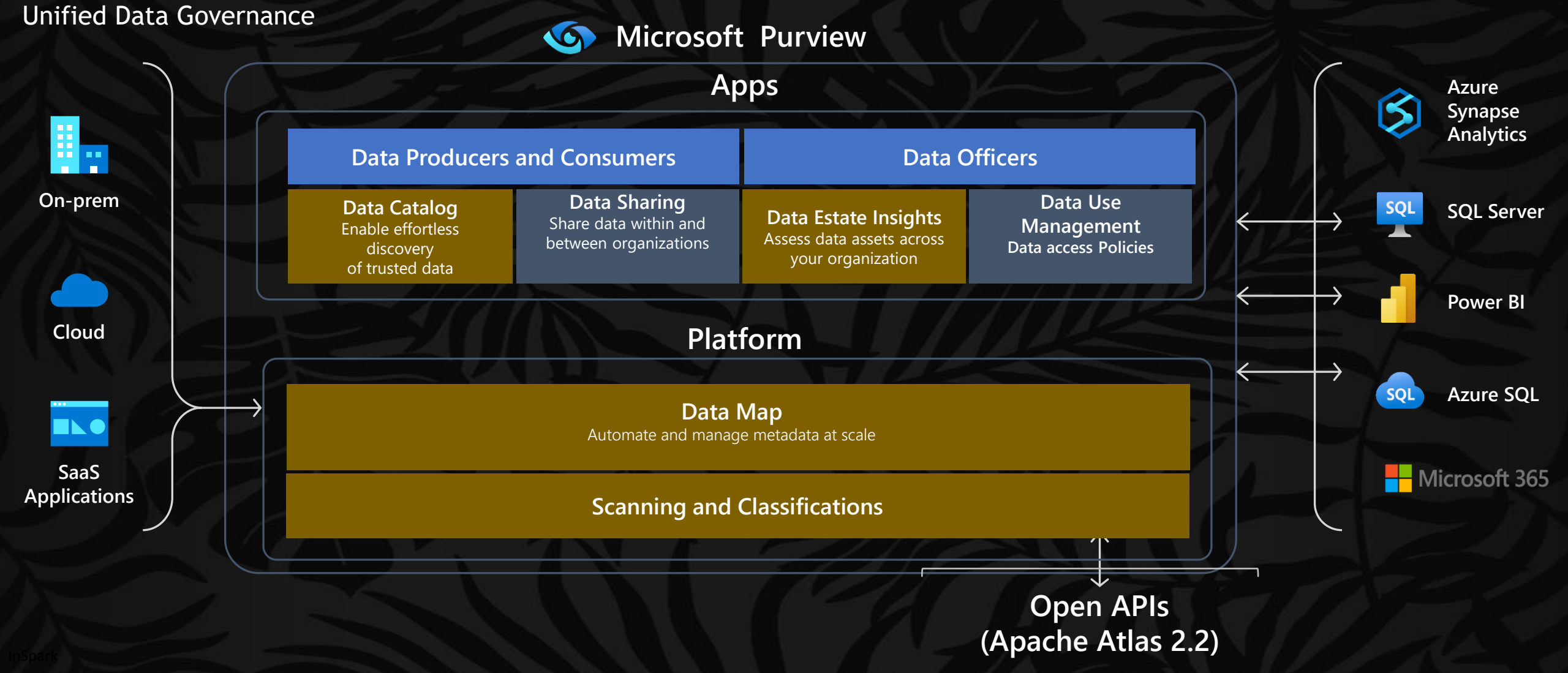


Microsoft Purview

Generally Available

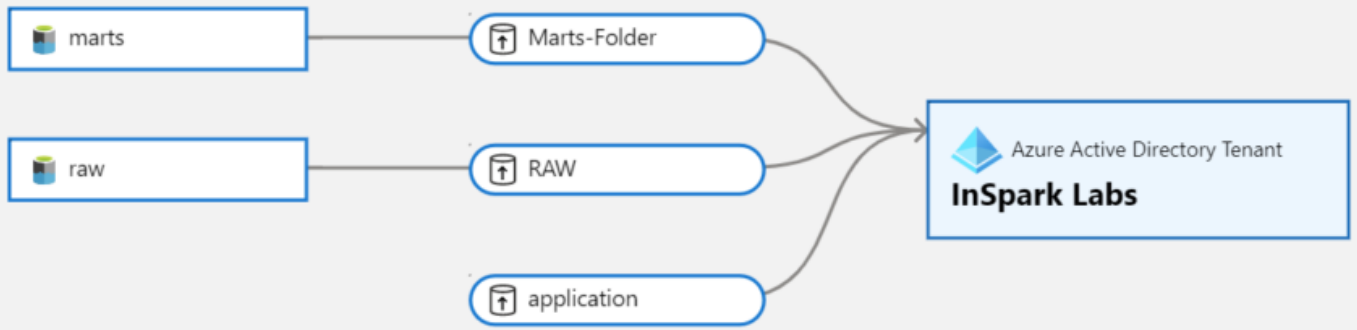
Preview

Unified Data Governance



Data 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.



Connect Fabric To Microsoft Purview

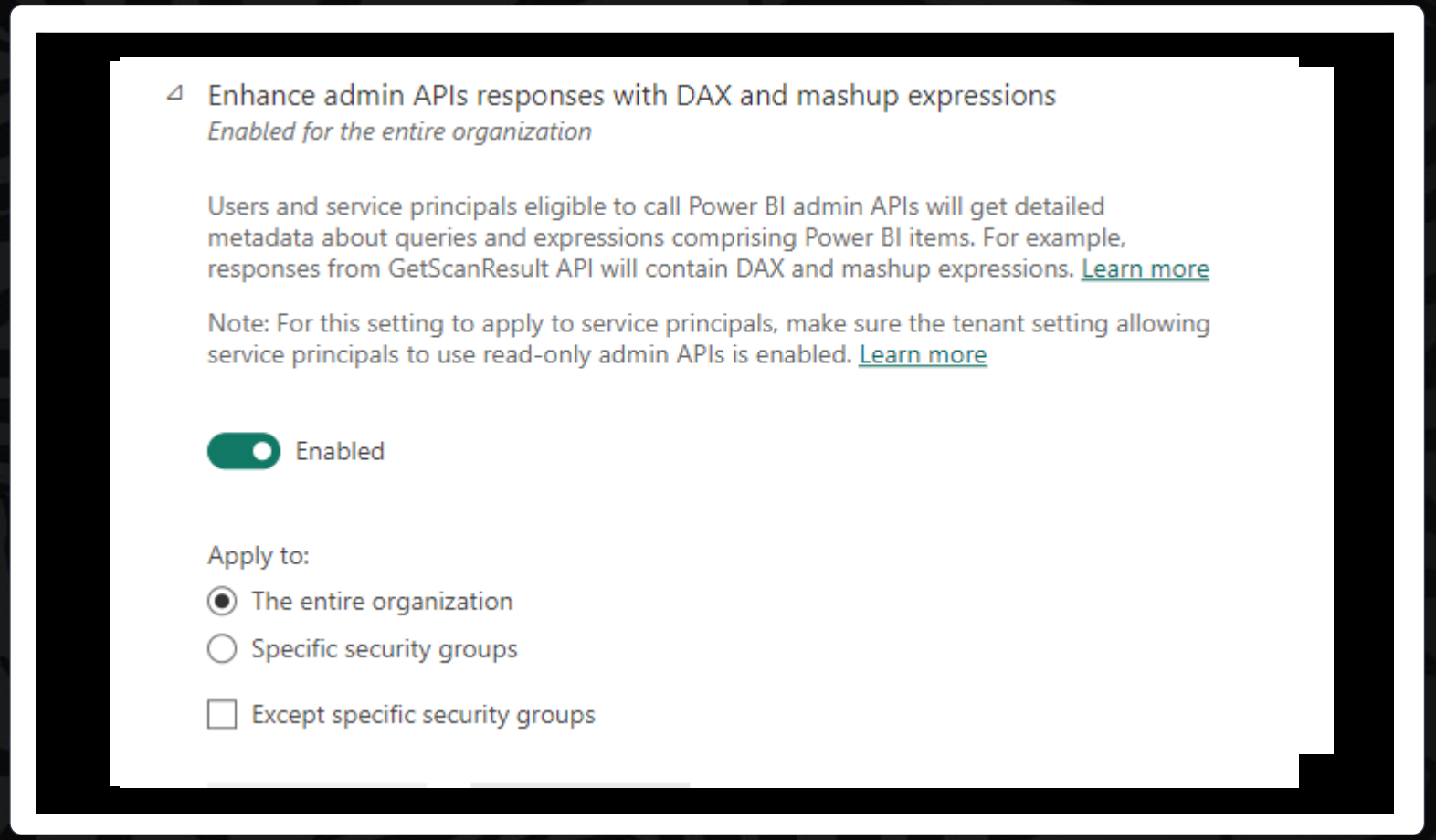
NEW

Same Tenant

Cross Tenant

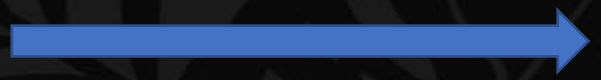
Scan Microsoft Fabric

- **Admin Portal**
- Allow Service Principal
- Detailed Metadata
- Dax and Mashup Expressions



Supported capabilities

- Metadata Extraction
- Full Scan
- Incremental Scan
- Lineage
- Scoped Scan
- Classification
- Access Policy
- Data Sharing



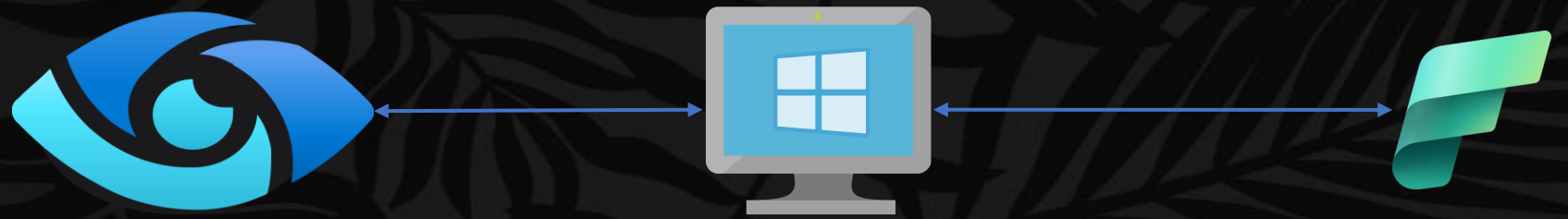
- Workspaces
- Dashboards
- Reports
- Datasets including tables and columns
- Dataflows
- Datamarts

Currently only Power BI assets are Available !

Supported Scenarios (security wise)

public access allowed

public access allowed



public access allowed/Denied

public access allowed/Denied

Disabled from all networks
Public network access will be disabled for Purview account, portal, and ingestion.

Azure Private Link
Enabled for the entire organization

Block Public Internet Access
Enabled for the entire organization

DEMO

Microsoft Purview

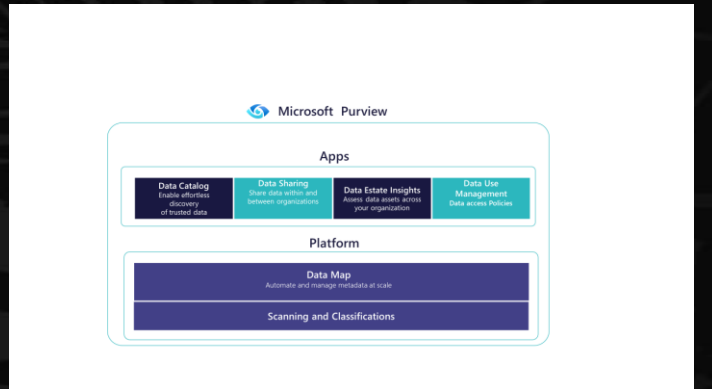
Data Map Consumption

- Elastic Data Map

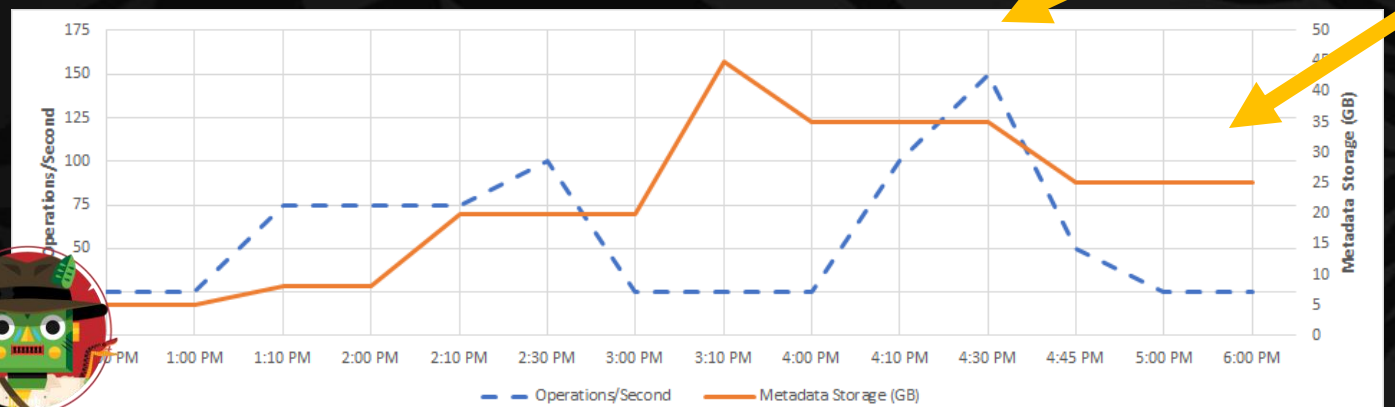
  €0.373 per 1 Capacity Unit Hour

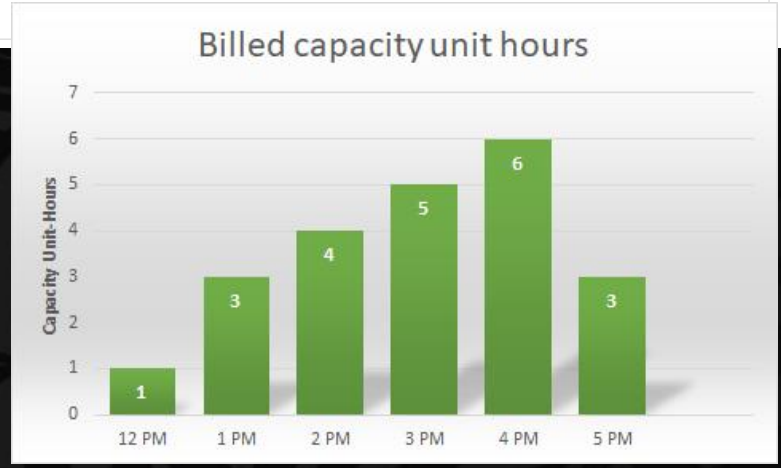
- CU=Capacity Unit

  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	[illegible]	80
[illegible]	225	90
10	250	100
..	..	..
100		





Pricing - Example

Data Map Consumption	Data Map Population	Data Map Enrichement
1,5 CPU x €0.407 X 730 hours	4 scans x 4 hours x 32 VCore x €0.623 per vCore per hour	60 scans x 1 hour x 8 VCore x €0.208 per 1vCore per hour(Resource set)
€ 445	€ 320	€ 100,00

€865,00

Microsoft Purview

Take aways

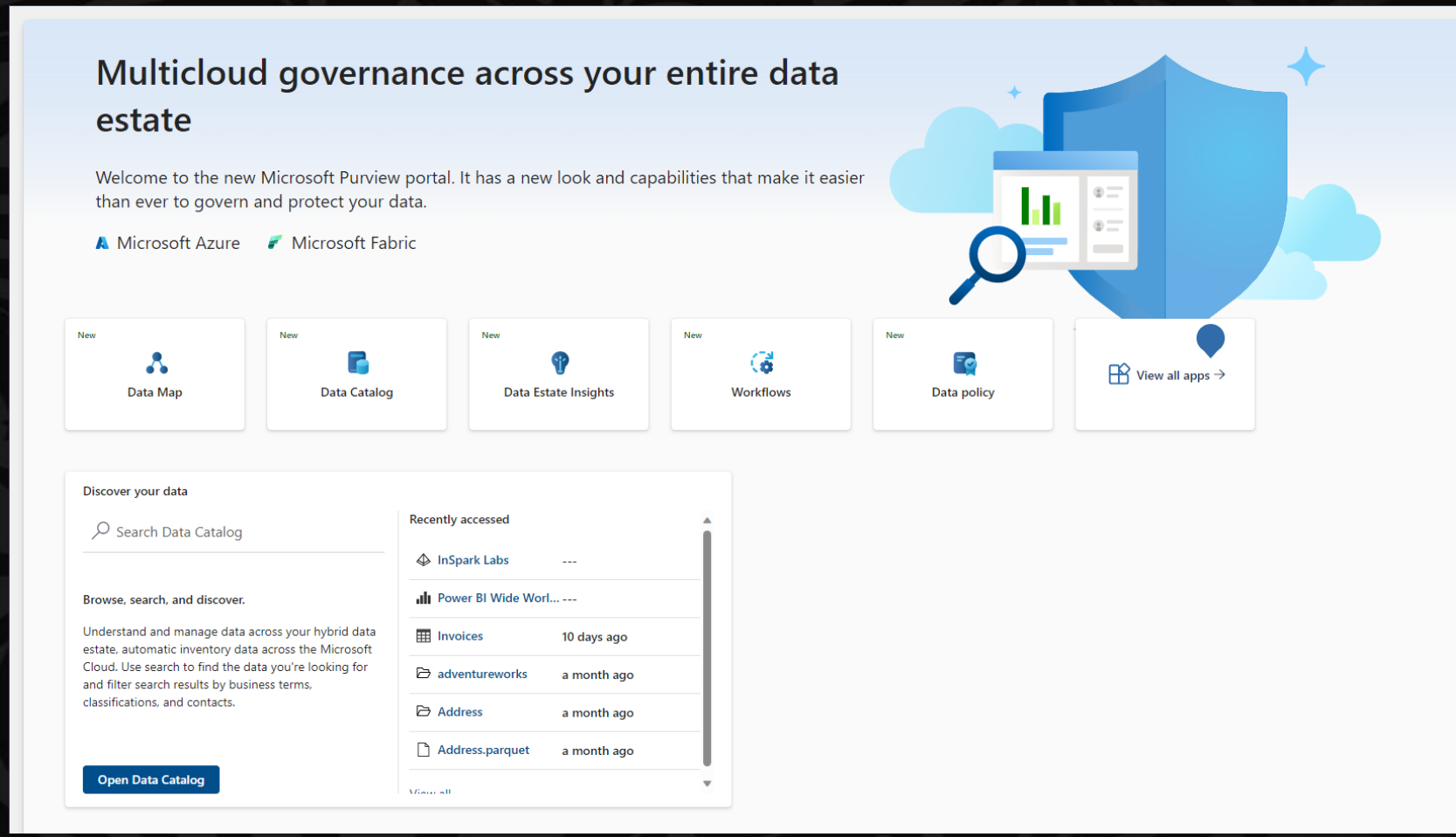
- Tagging
- More reporting and In-product customization with Fabric Embedded
- Support for Databricks Unity Catalog and Snowflake is now supported

- 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

INSPARK

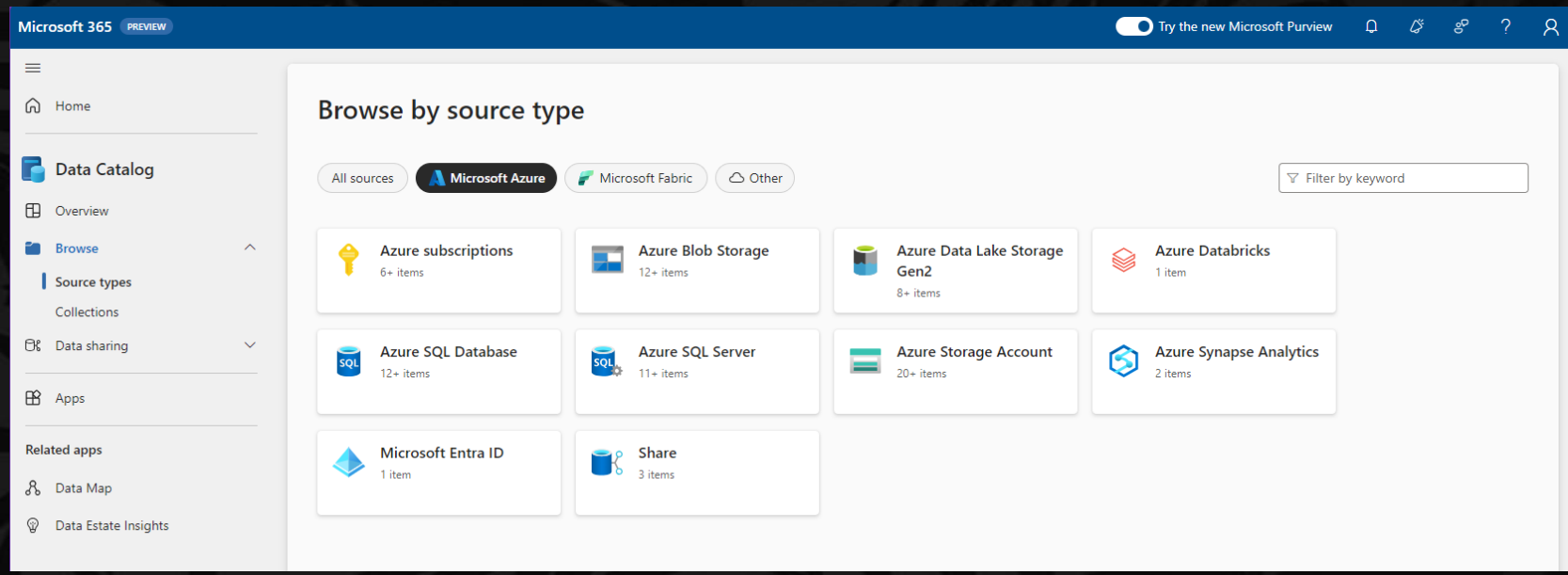
New Features

- Microsoft Purview will be added to Tenant Level

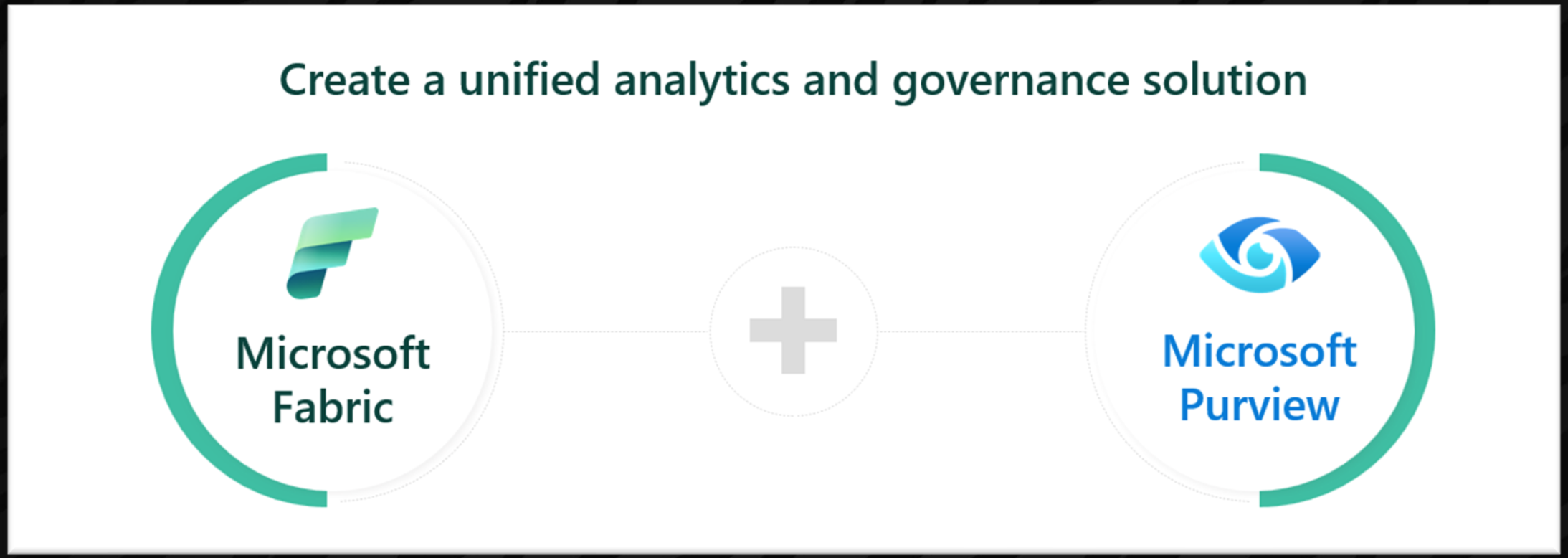


New Features

- Microsoft Purview will be added to Tenant Level
- Automatic detecting of new sources
- Automatic scanning (push)



Roadmap


Create a unified analytics and governance solution
Microsoft Fabric
Microsoft Purview

Simplify governance and security with Purview

Create a unified analytics and governance solution

Reshape how you access, manage, and act on data and insights by connecting every data source and analytics service together in Fabric

Manage, govern, and monitor your data estate with a governance solution designed for Microsoft Fabric

Unified governance | Unmatched compliance | Data ownership

Any questions left?

Thank you for attending!

Erwin de Kreuk
Principal Consultant - Lead Data & AI
InSpark

@erwindekreuk
linkedin.com/in/erwindekreuk
erwindekreuk.com
github.com/edkreuk
https://sessionize.com/erwin-de-kreuk/

Let's connect

Microsoft® MVP Most Valuable Professional