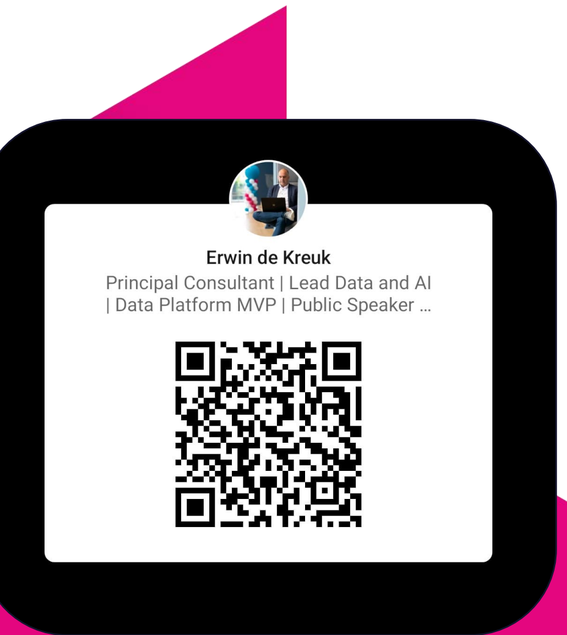
Solve your Data Governance challenges with Microsoft Purview

Erwin de Kreuk

Principal Consultant – Lead Data & AI

InSpark



Microsoft® Most Valuable Professional

INSPARK

INSPARK

Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark

@ErwinDeKreuk

linkedin.com/in/ErwinDeKreuk

ErwinDeKreuk.com

github.com/edkreuk

Microsoft® MVP Most Valuable Professional

We Are InSpark

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

INSPARK

snowflake

twoday
kapacity

ITPA
CONSULTING

Microsoft

dataon
Power your data

Premium Sponsors

Raffle Prizes

Standard Sponsors

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

INSPARK

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

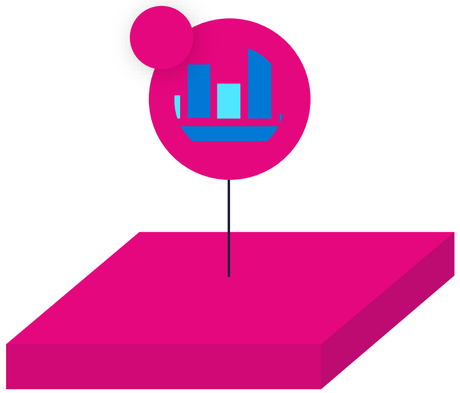
IDC

IDC Data Age 2025, Dave Reinsel,

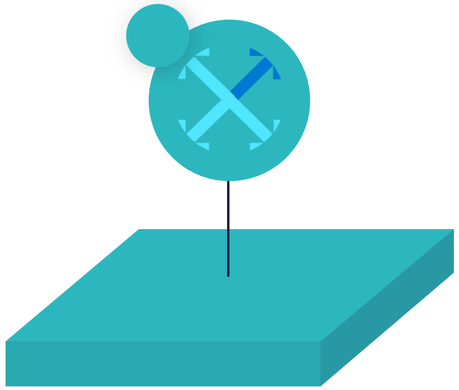
Microsoft Azure

Microsoft Purview

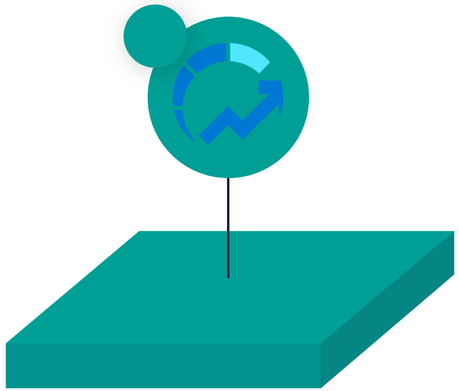
Elements of successful data governance



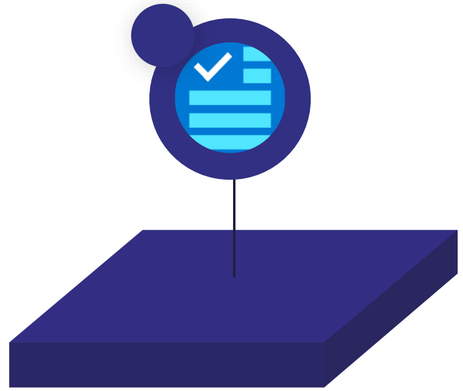
Manage growing data landscape



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

INSPARK

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.

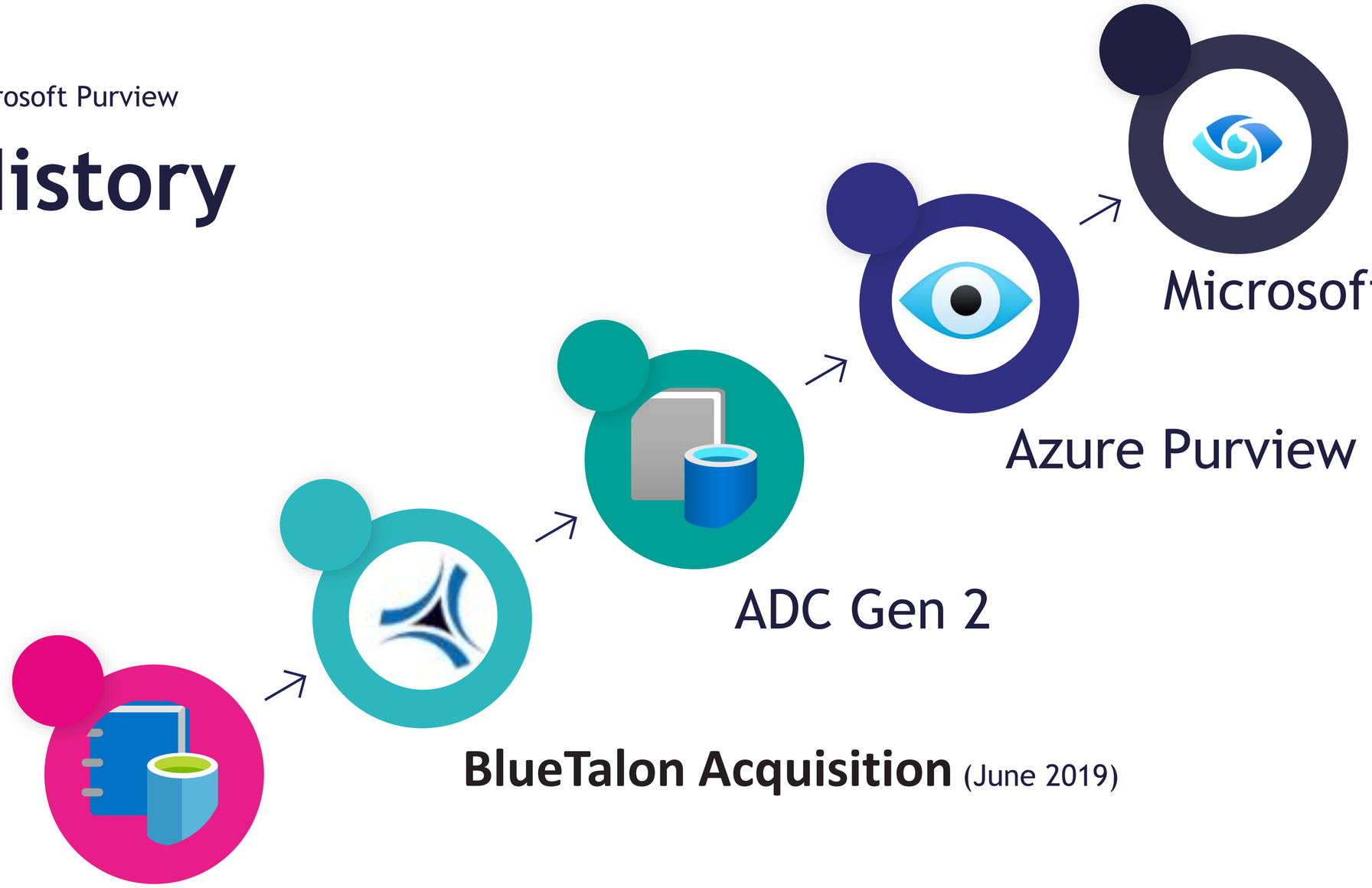
Unified data governance to maximize the business value of data

Microsoft Purview

INSPARK

Microsoft Purview

History



ADC Gen 1

BlueTalon Acquisition (June 2019)

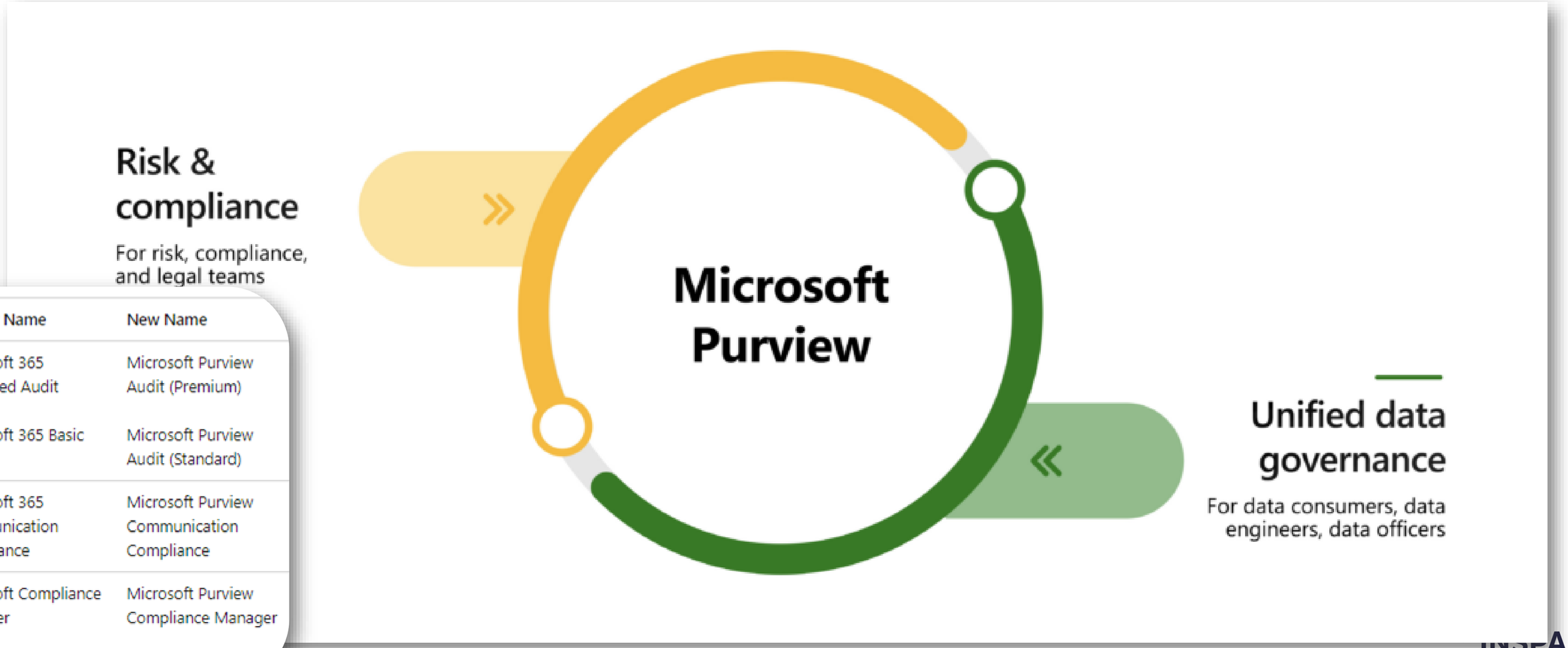
ADC Gen 2

Azure Purview sept 2021

Microsoft Purview april 2022

Microsoft Purview

The future of compliance and data governance

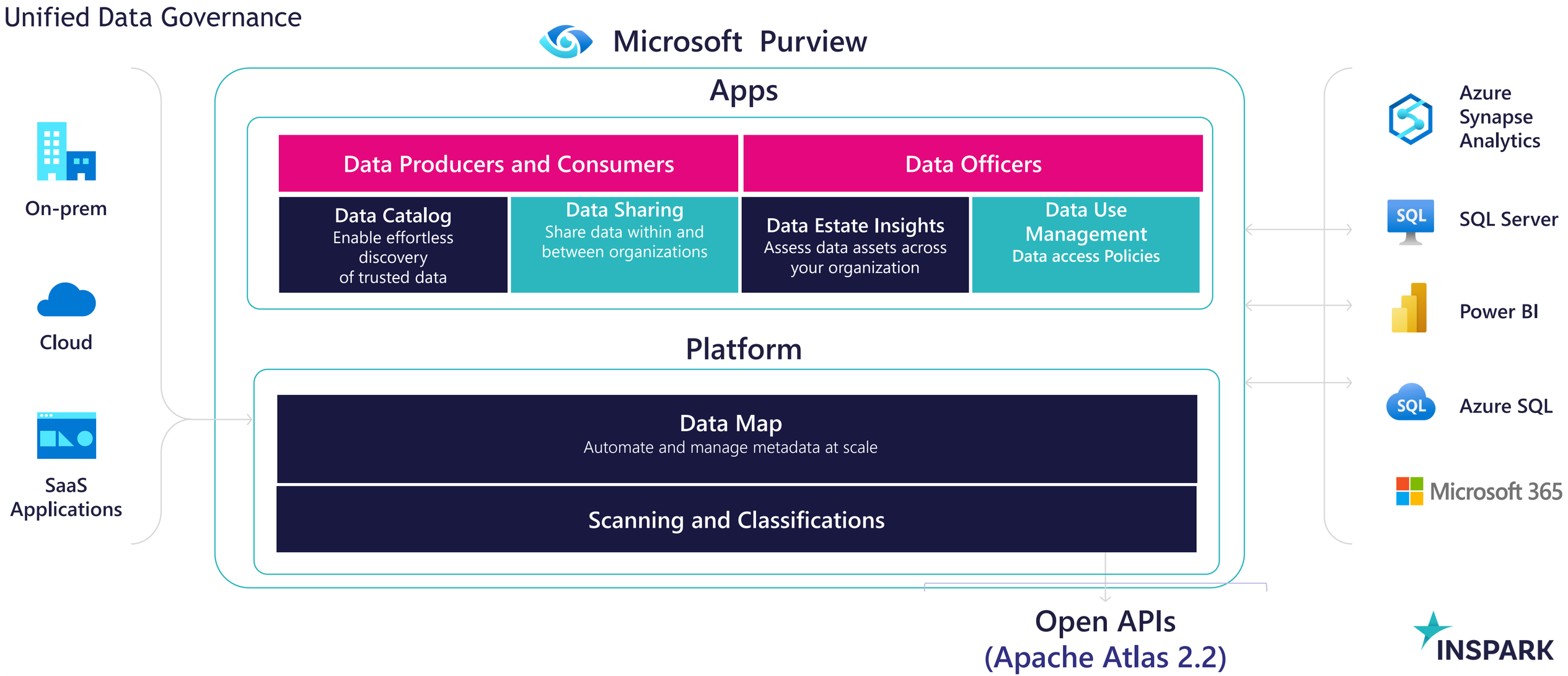


Former Name	New Name
Microsoft 365 Advanced Audit	Microsoft Purview Audit (Premium)
Microsoft 365 Basic Audit	Microsoft Purview Audit (Standard)
Microsoft 365 Communication Compliance	Microsoft Purview Communication Compliance
Microsoft Compliance Manager	Microsoft Purview Compliance Manager

Microsoft Purview

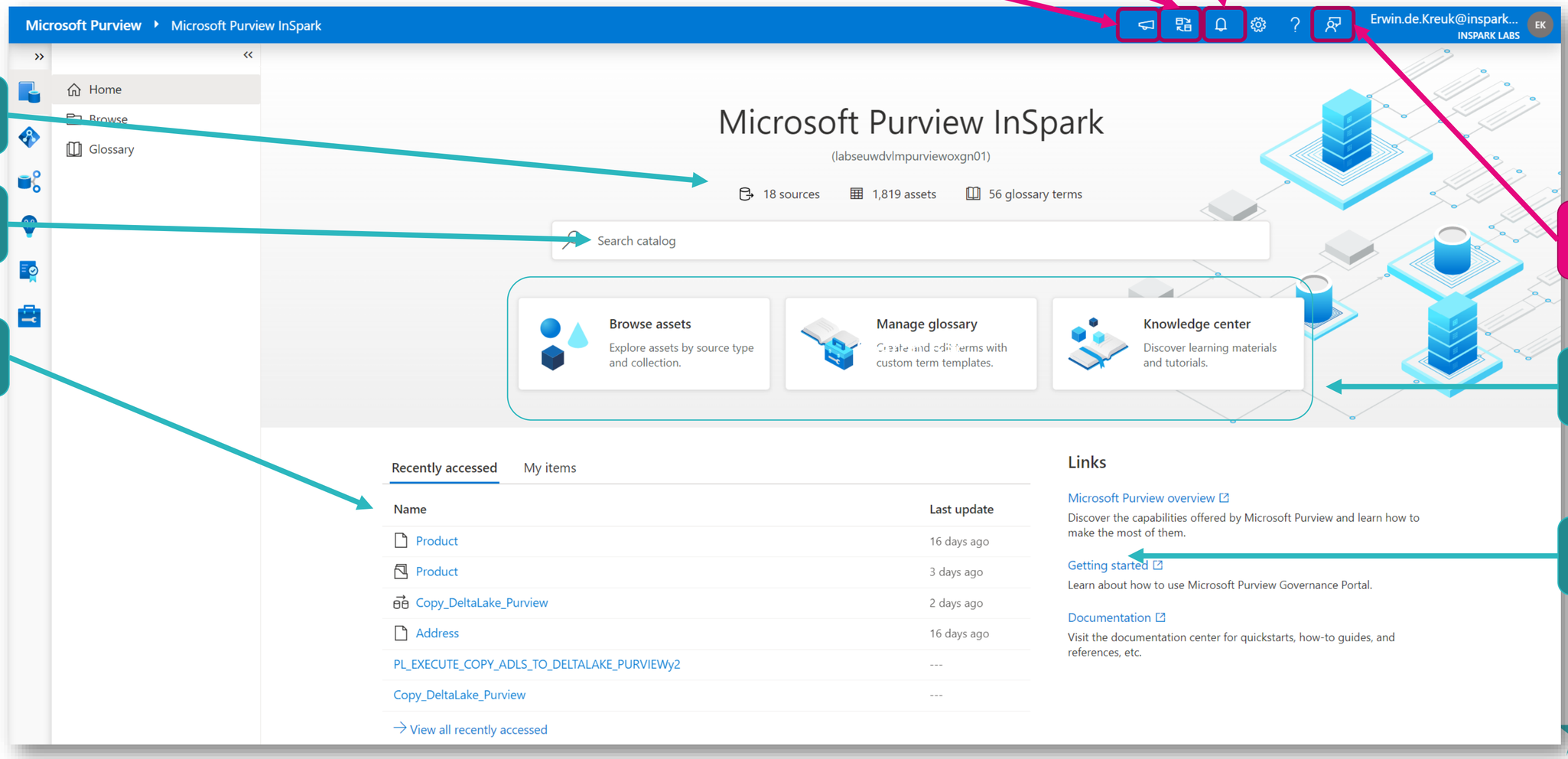
Generally Available

Preview



Microsoft Purview

Purview Studio



Updates

Accounts

Notifications

Metrics

Search Bar

Recently Accessed Entities

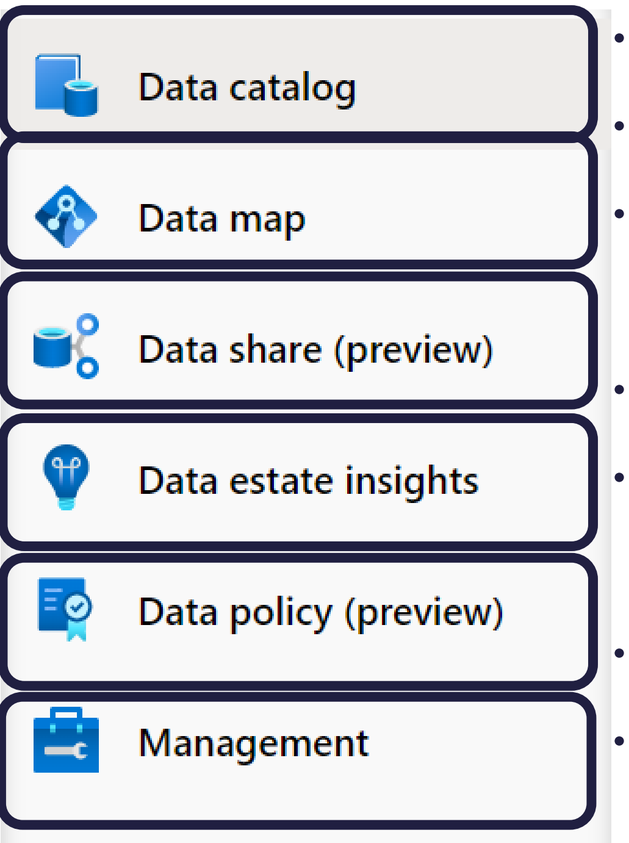
Feedback

Key Actvities

Usefull Links

INSPARK

Purview Studio - Activity hubs



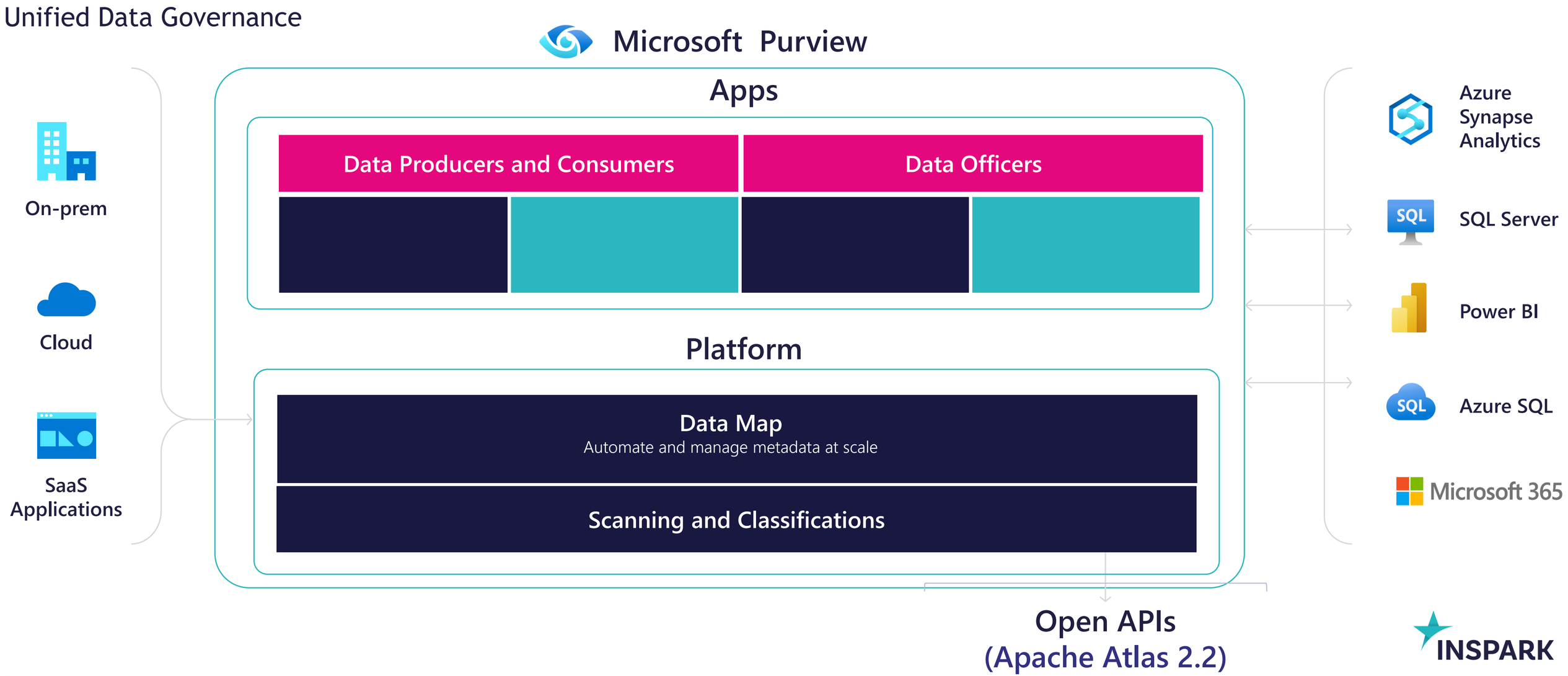
- Quick Actions, recently accessed items, owned Items, search bar and Documentation.
- Manage Glossary Items, search, manage terms templates and custom attributes, import/export Terms using csv.
- Create collections, register data sources, setup Scans, Integration runtime.
- Easily share data between 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. Enabl
  Feature options.

Microsoft Purview

Generally Available

Preview

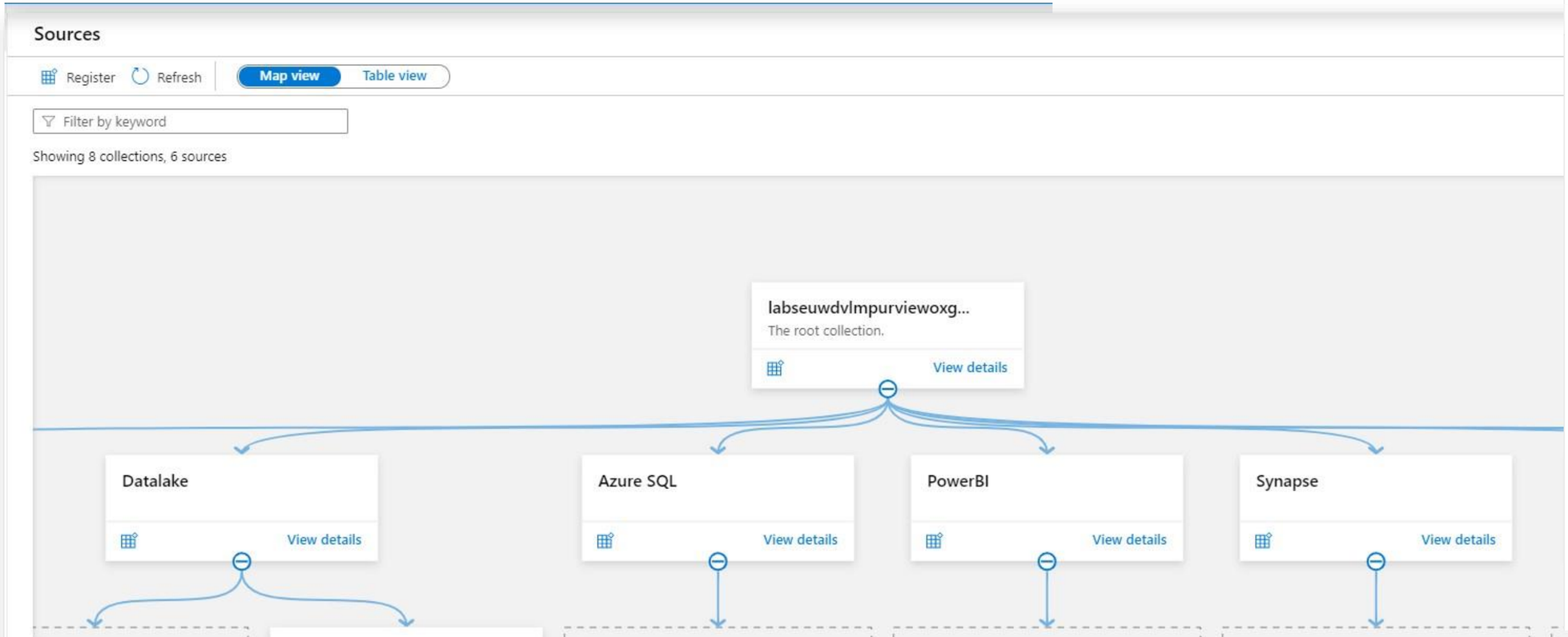
Unified Data Governance

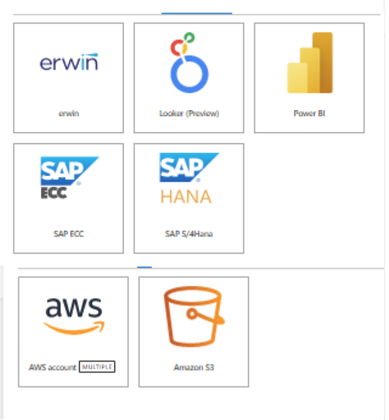


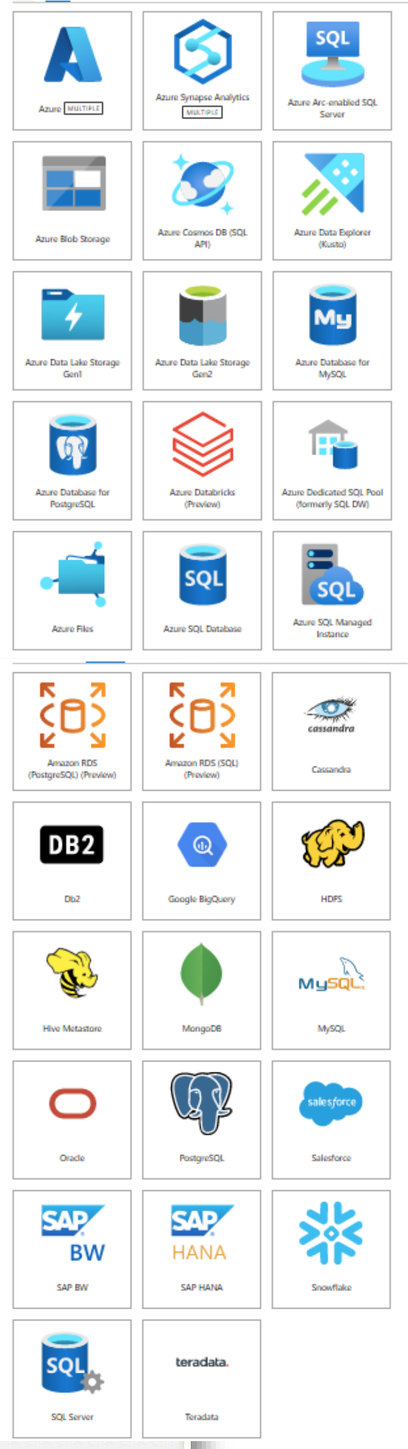
Data Map

Collections and Sources

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



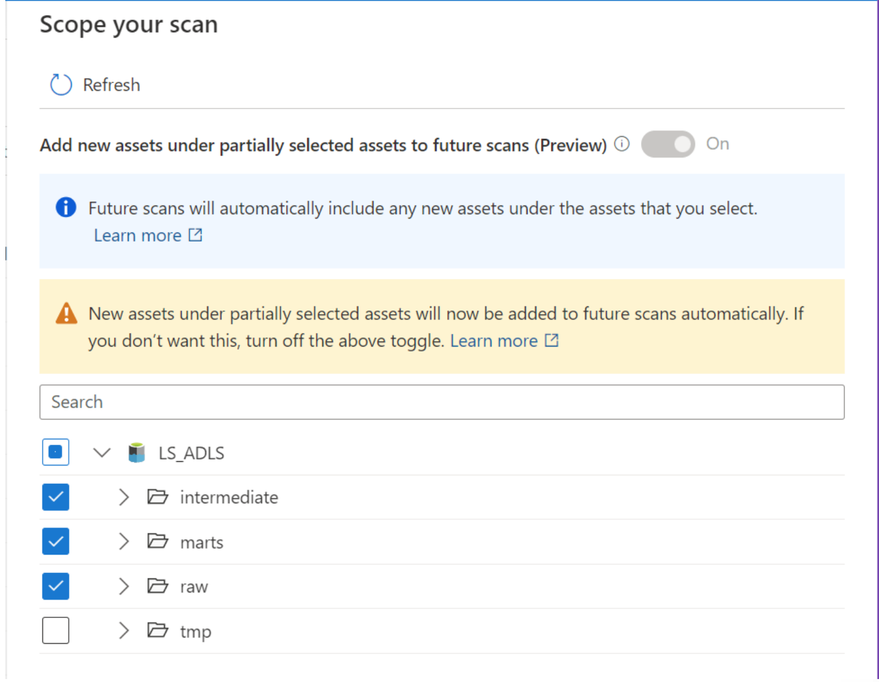


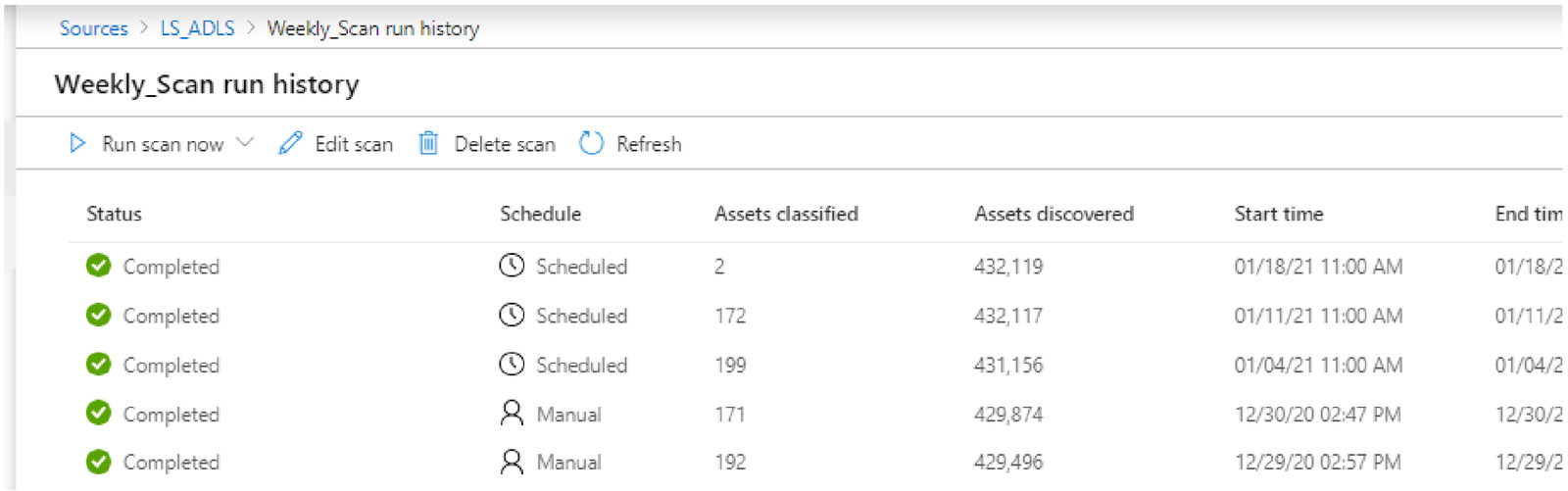


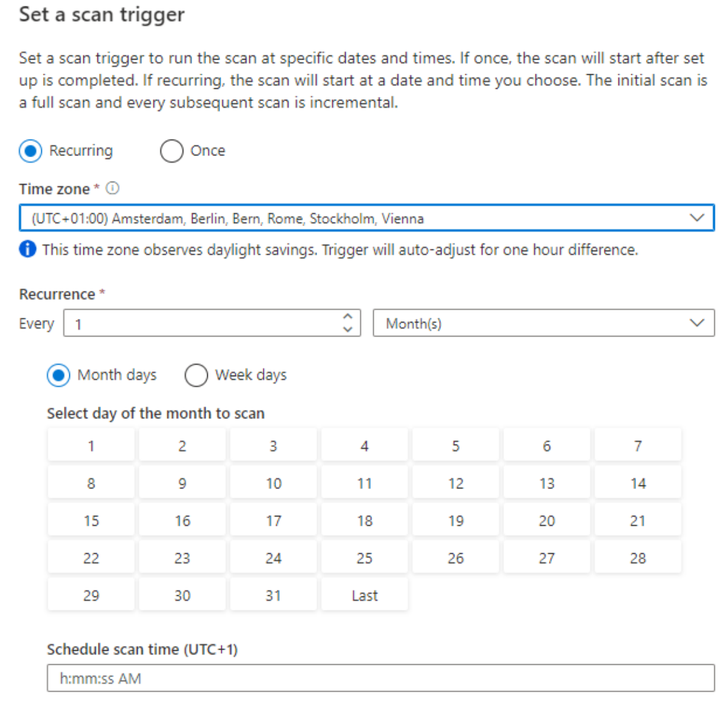
Data Map

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



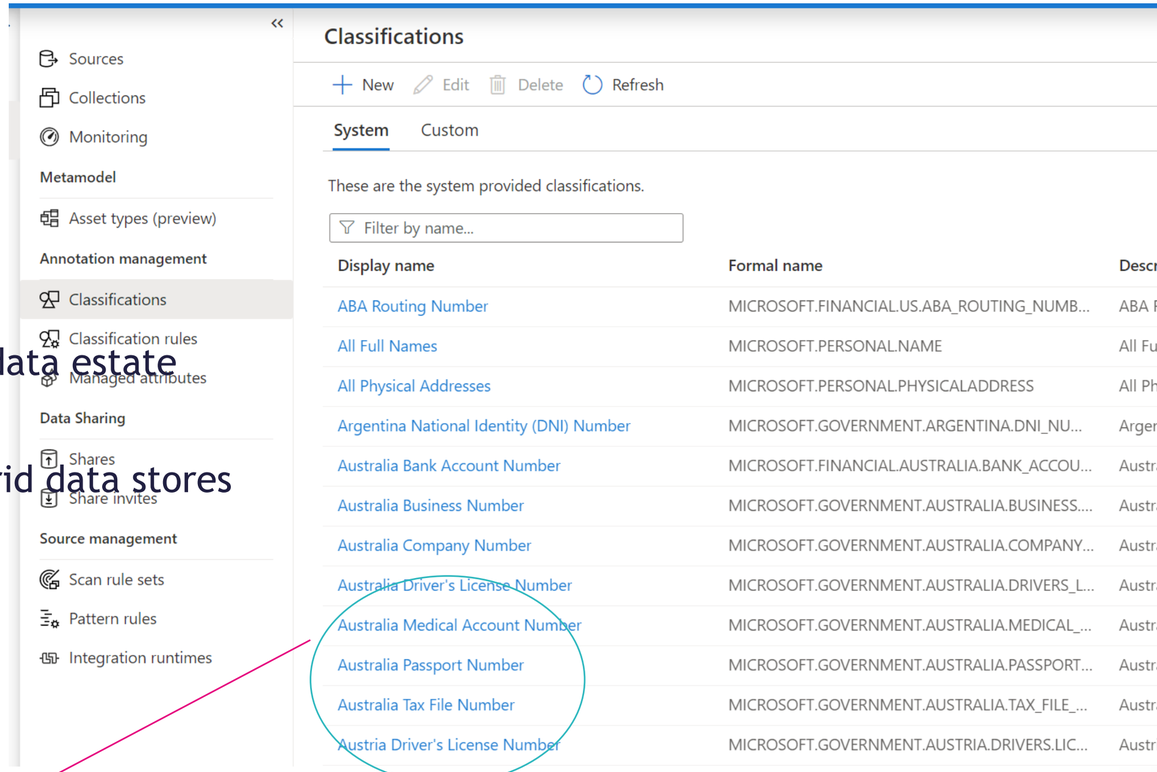


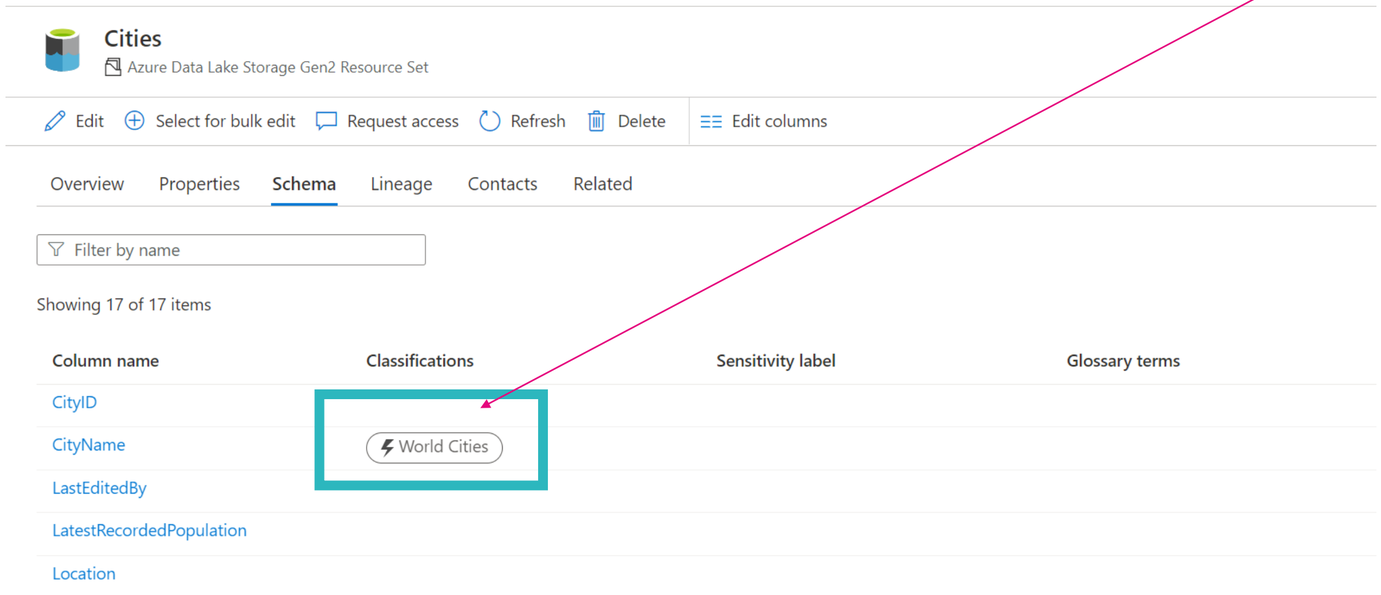


Data Map

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



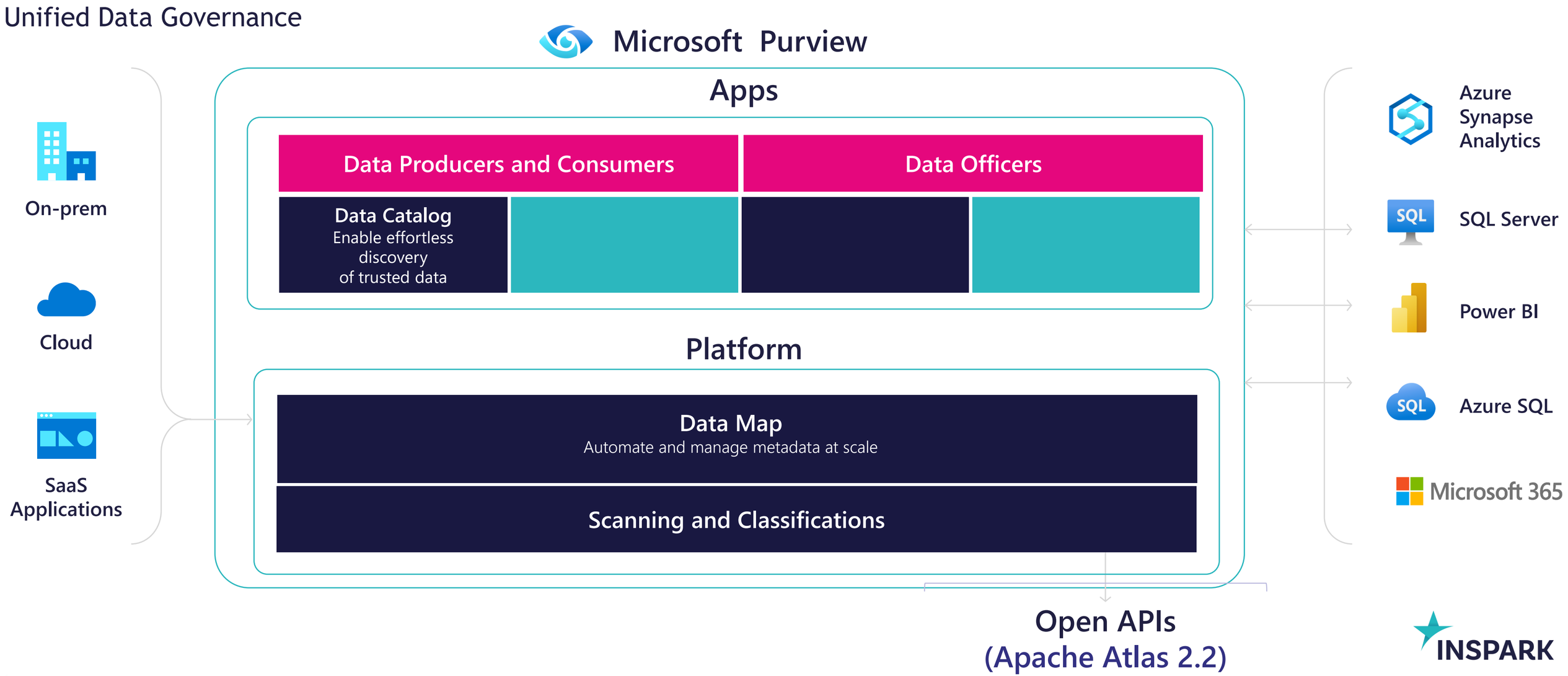


Microsoft Purview

Generally Available

Preview

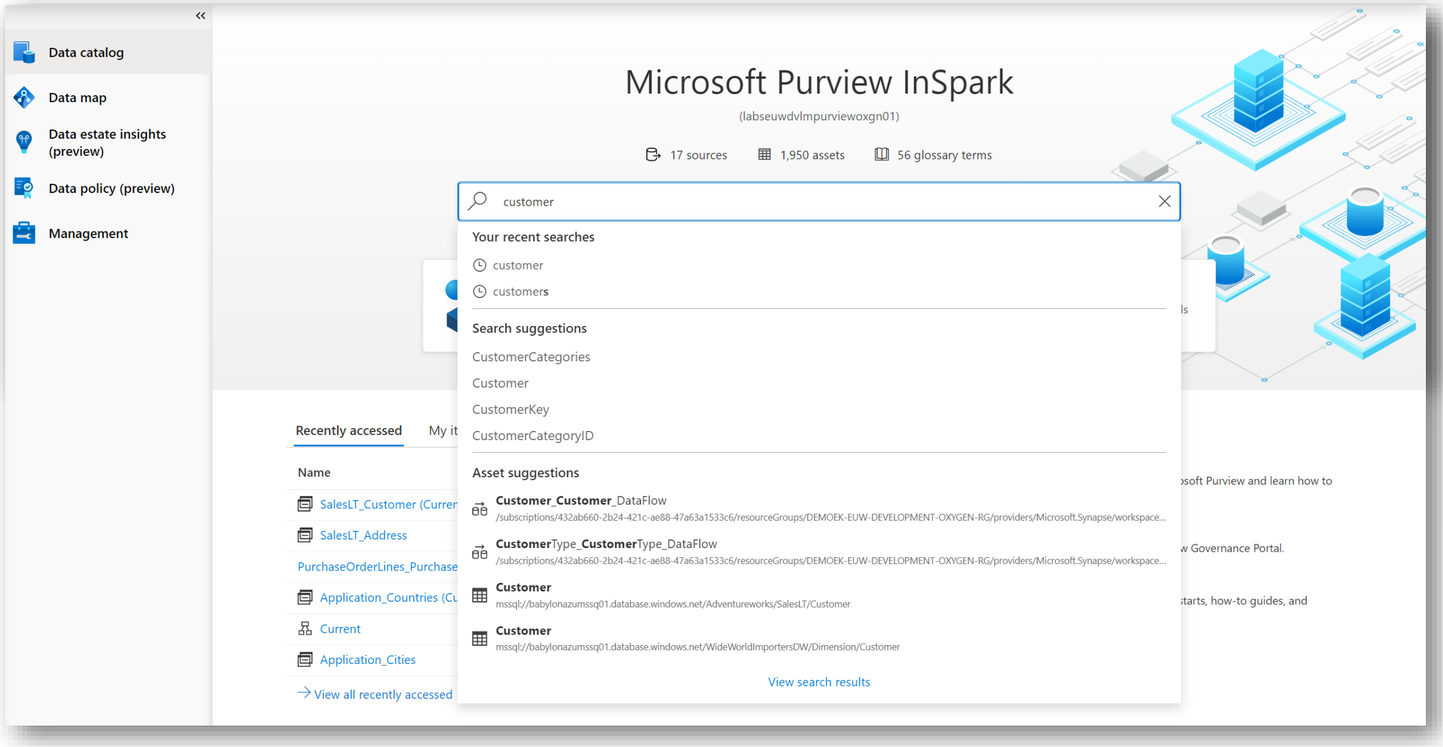
Unified Data Governance



Data Catalog

Search

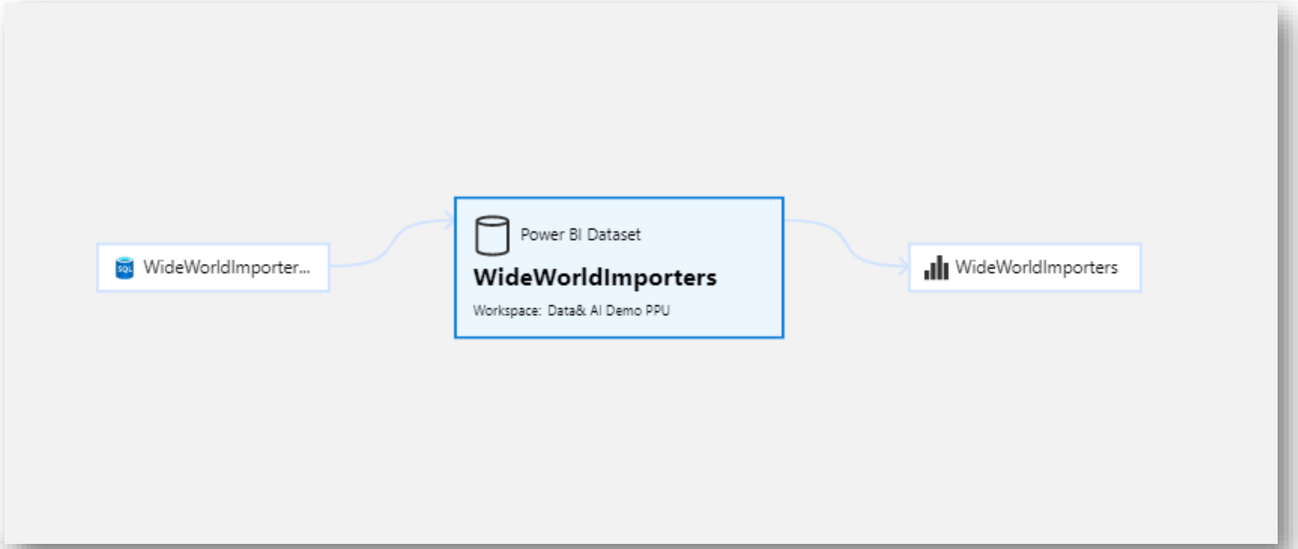
- Semantic search and browse
- Business glossary and workflows

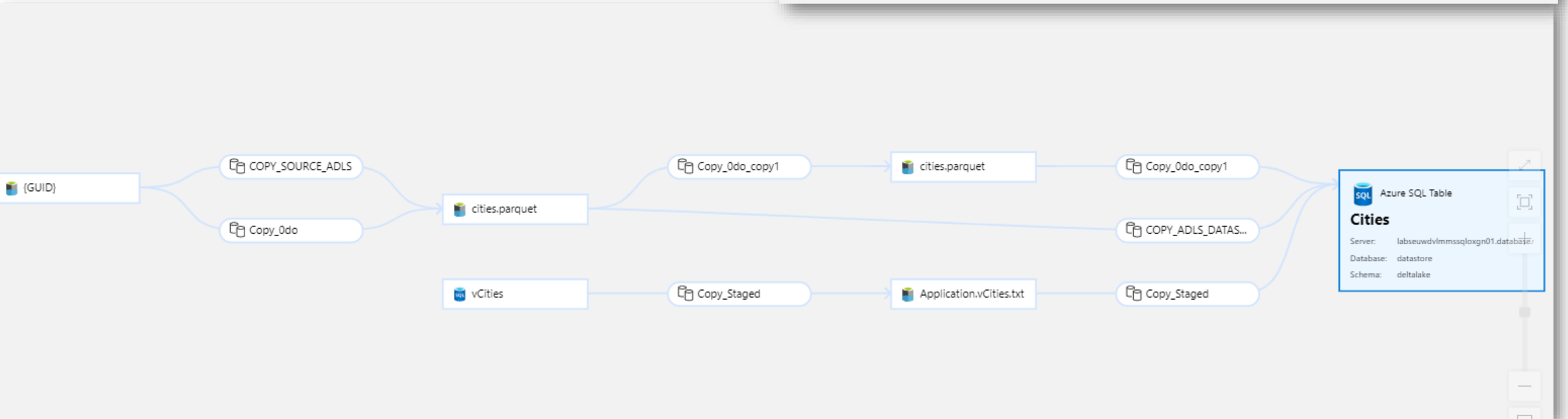


Data Catalog

Lineage

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

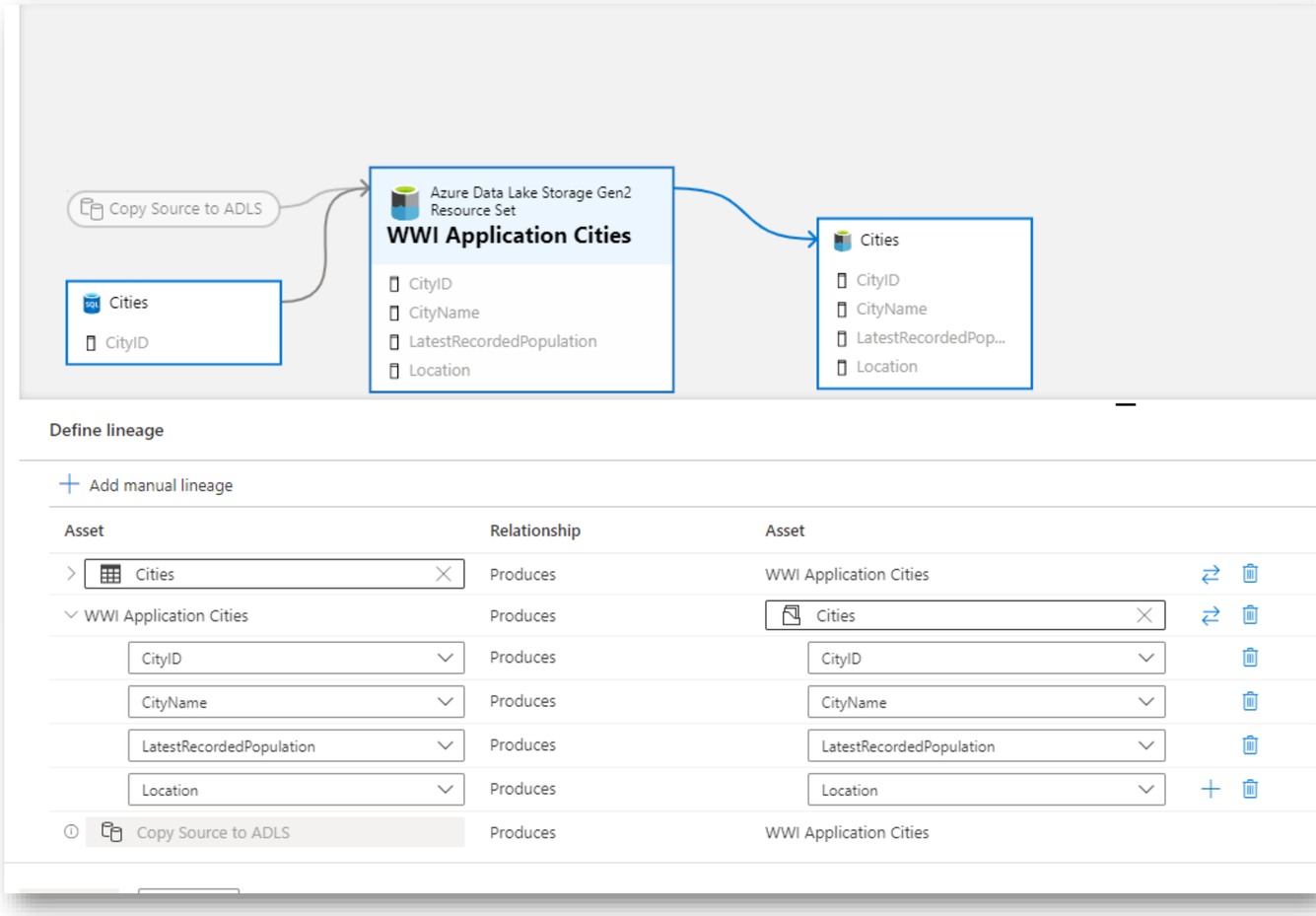




Data Catalog

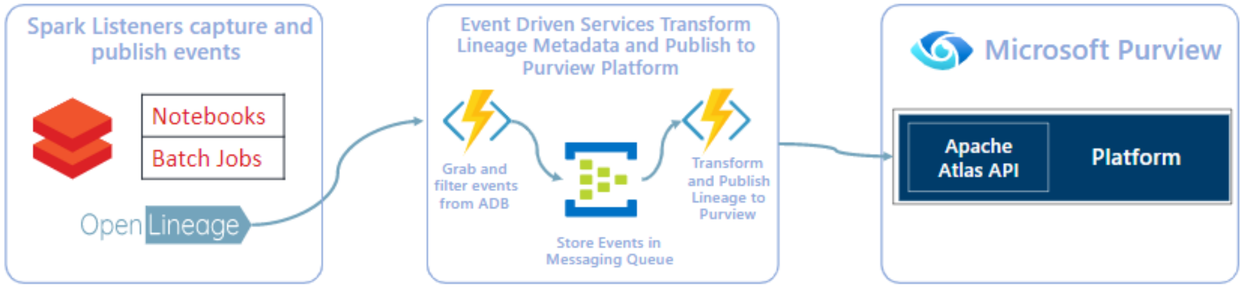
Manual Lineage

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



Data Catalog

Custom Lineage

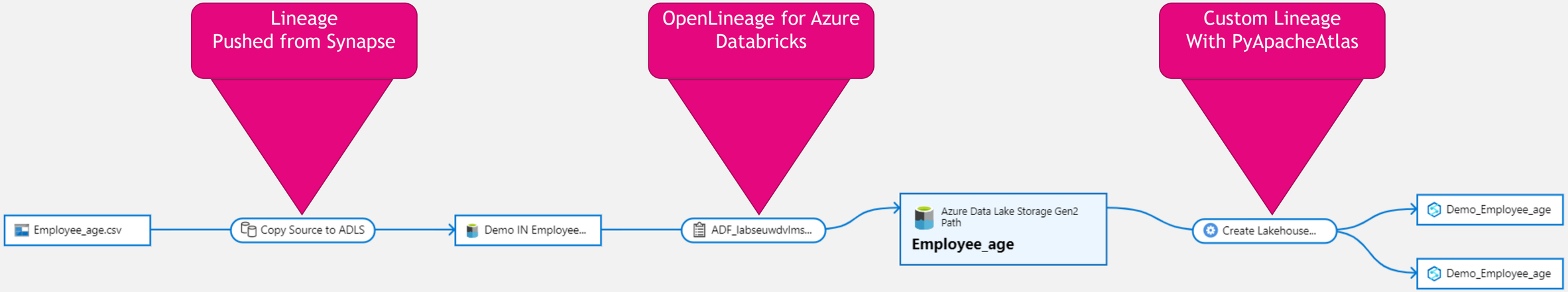


- OpenLineagefor Azure DataBricks

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

- Create custom Lineage with Apache Spark 2.2

https://github.com/wjohnson/pyapacheatlas

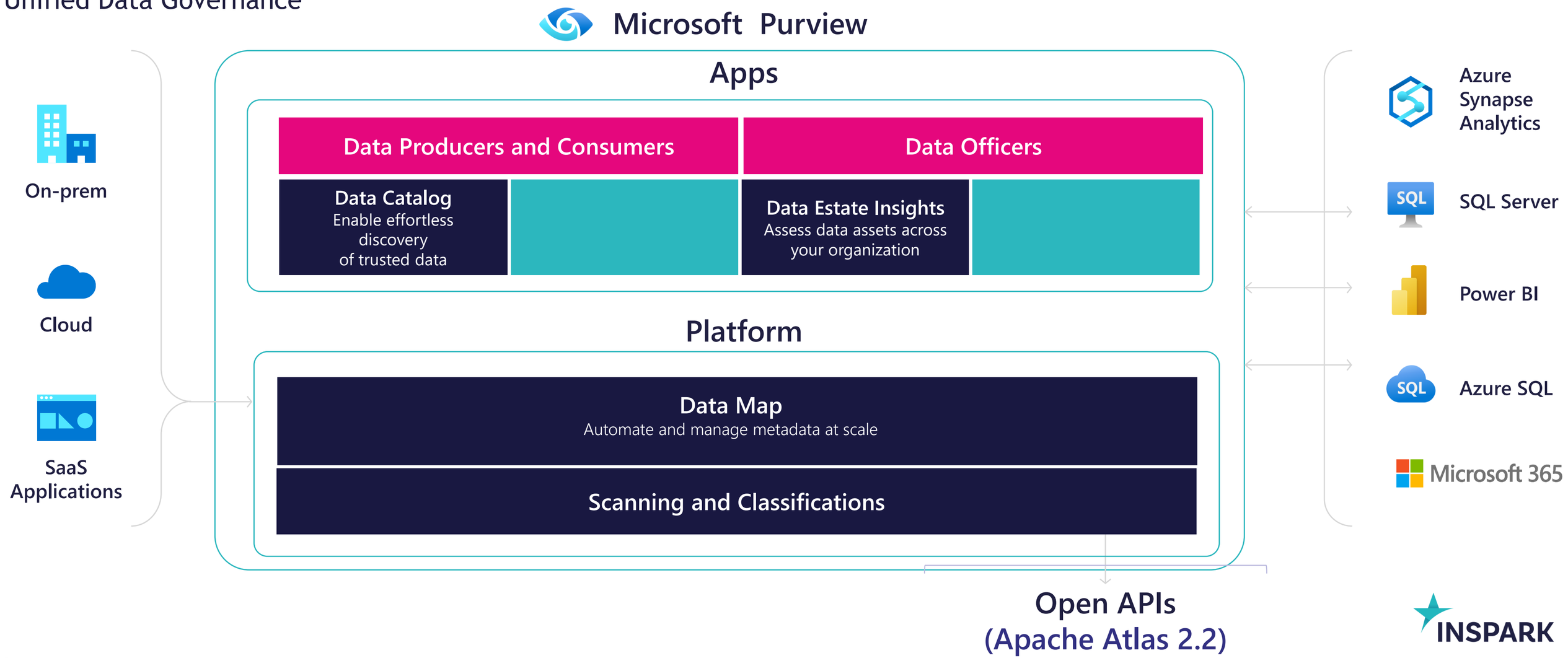


Microsoft Purview

Generally Available

Preview

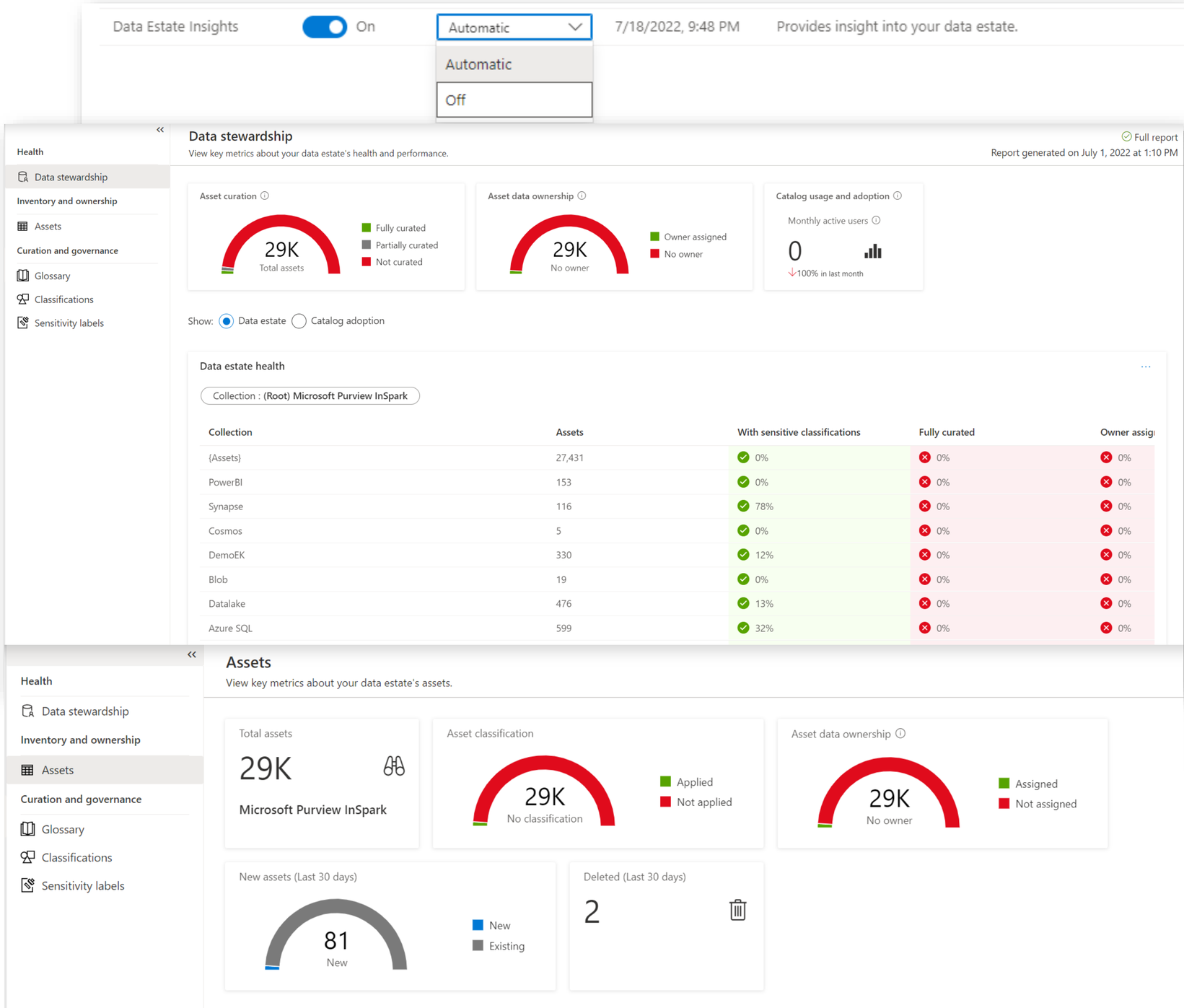
Unified Data Governance



Data Estate Insight

Metrics

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



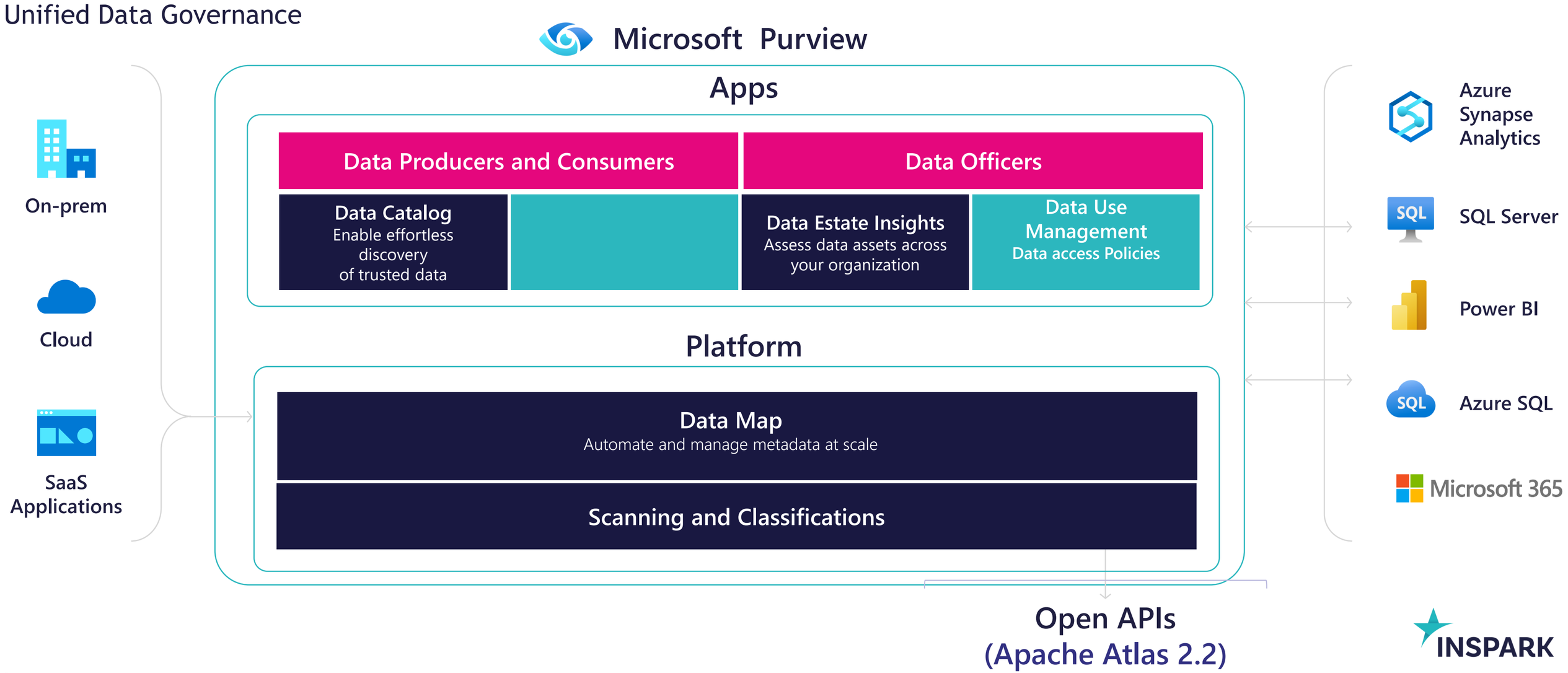
Data Owner
Dev Ops
Self Service access

Microsoft Purview

Generally Available

Preview

Unified Data Governance

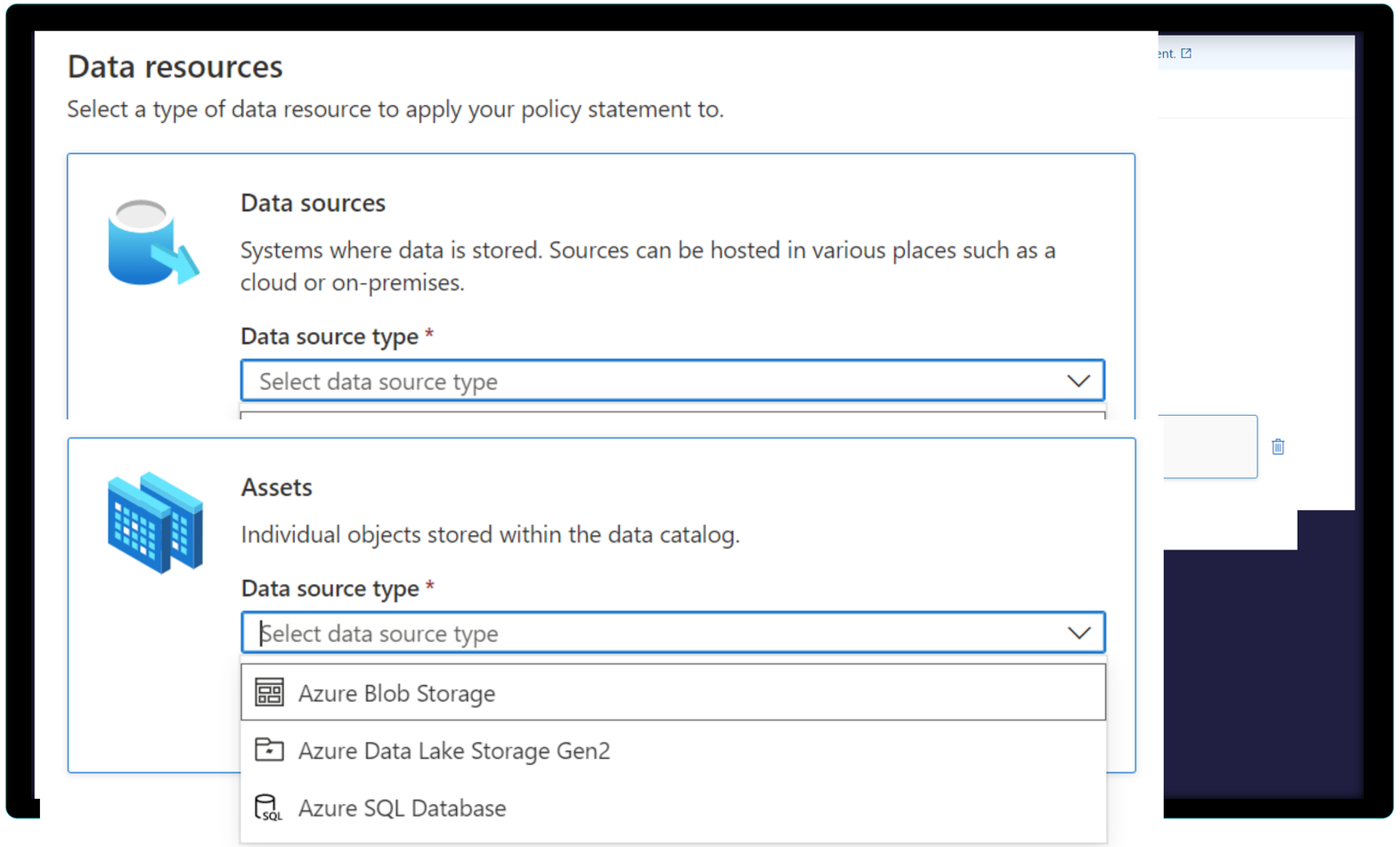


INSPARK

Data Owner policies

Currently in Preview

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

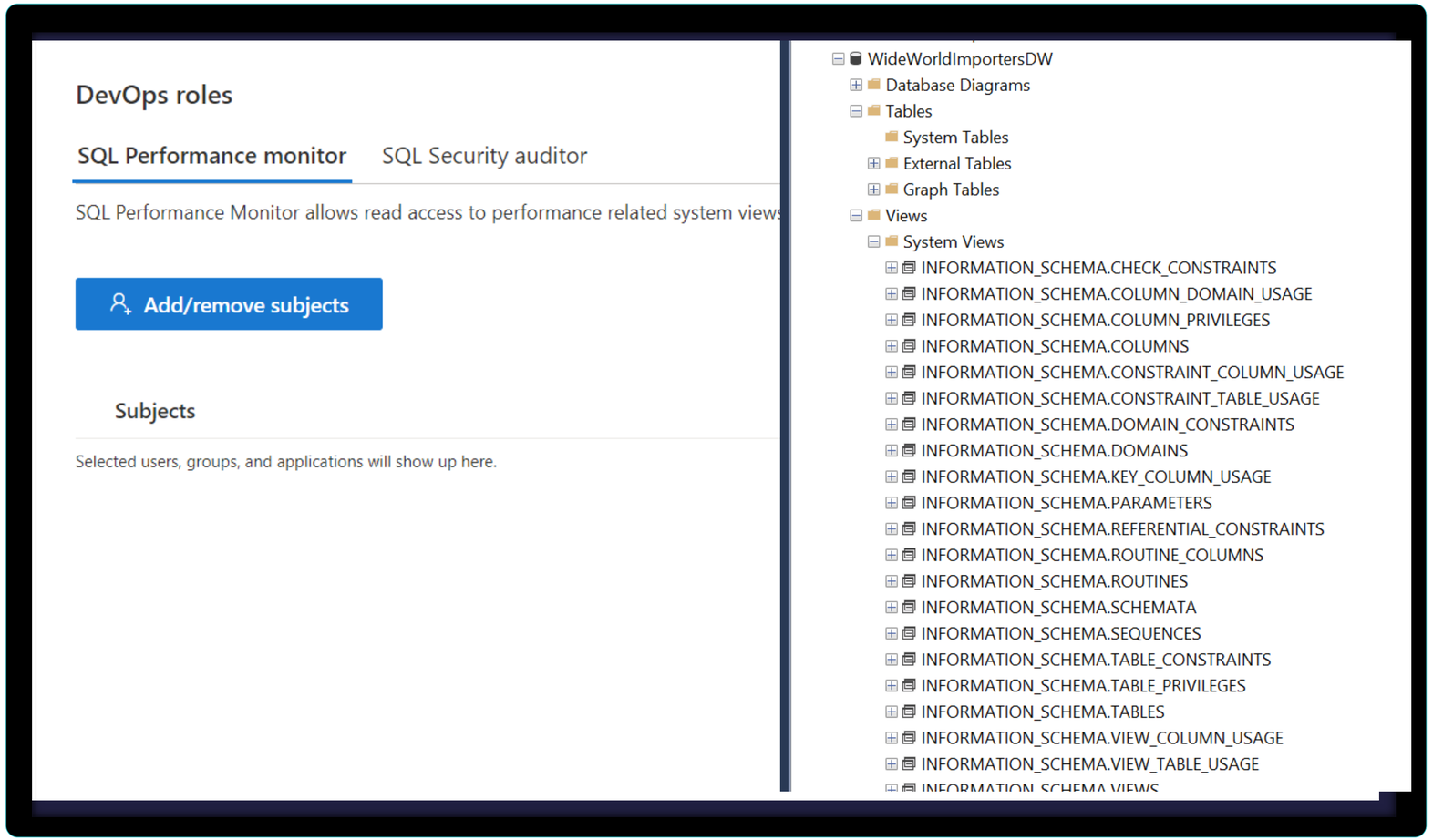


Dev Ops policies

Manage Access to System Metadata

SQL Performance Monitoring

SQL Security Auditing

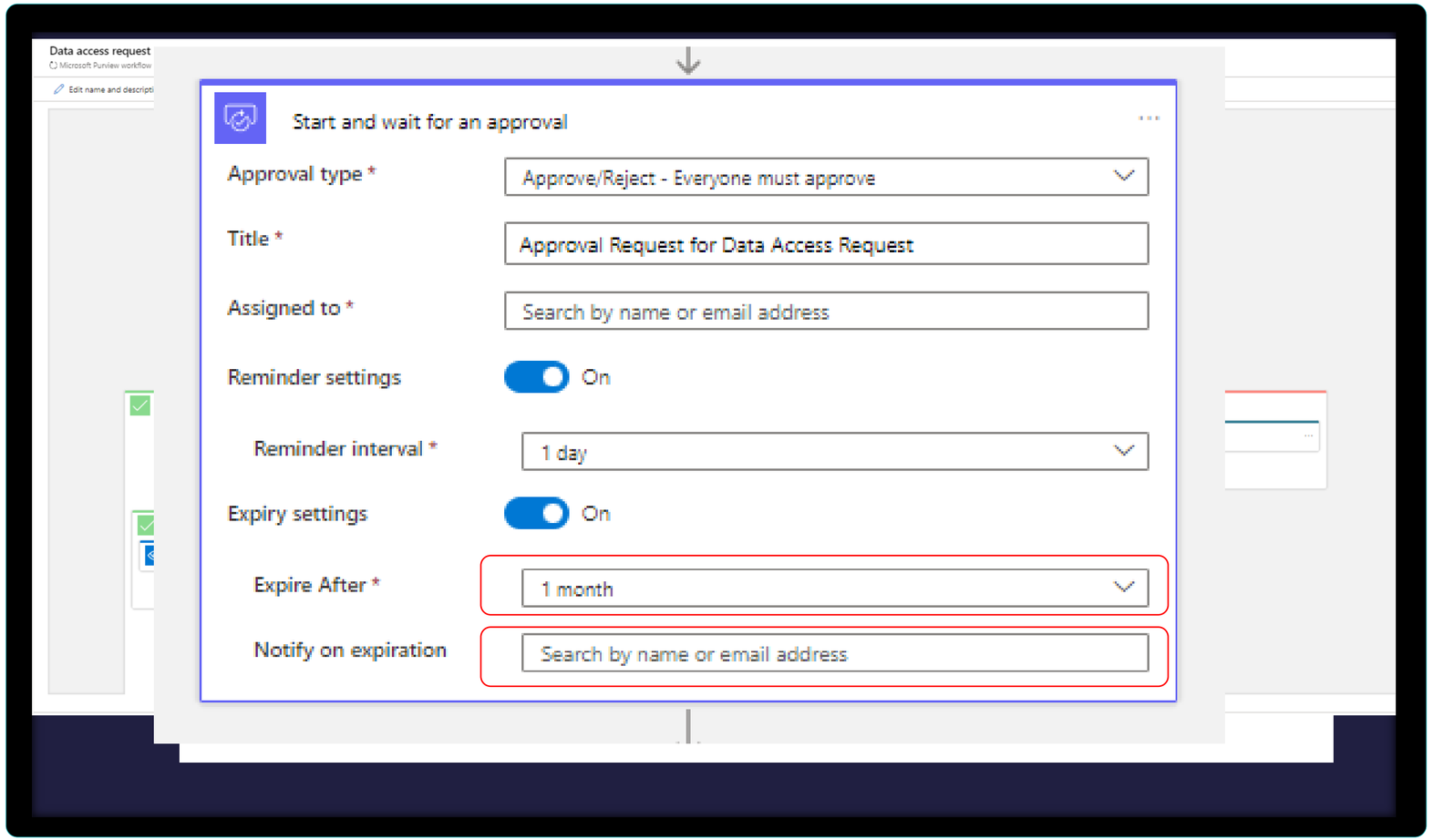


INSPARK

Self-service access policies

Data Access (Governance)

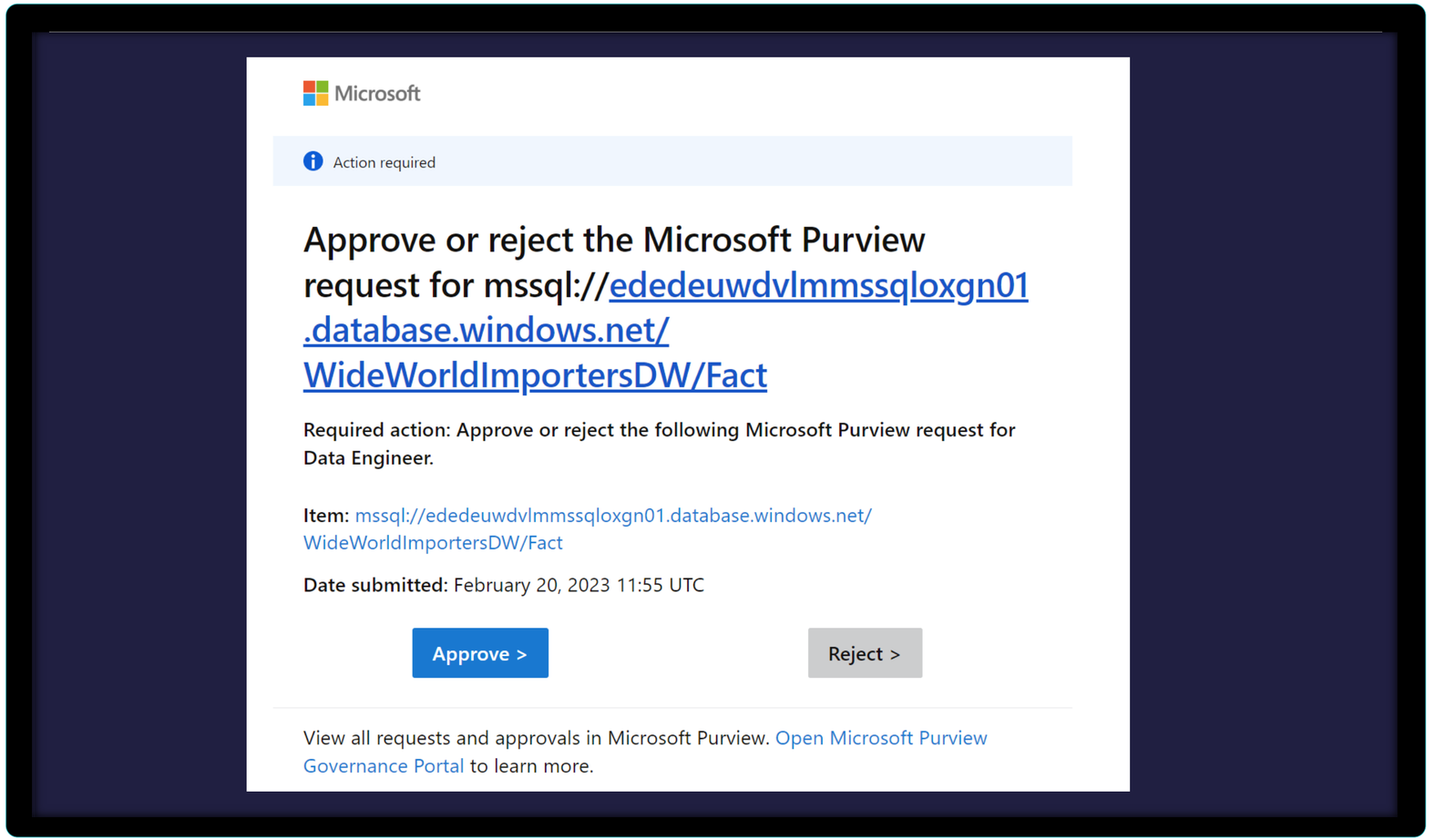
- **Azure SQL Database**
- **Azure Storage**
- **Expiration Functionality**



Self-service access policies

Data Access (Governance)

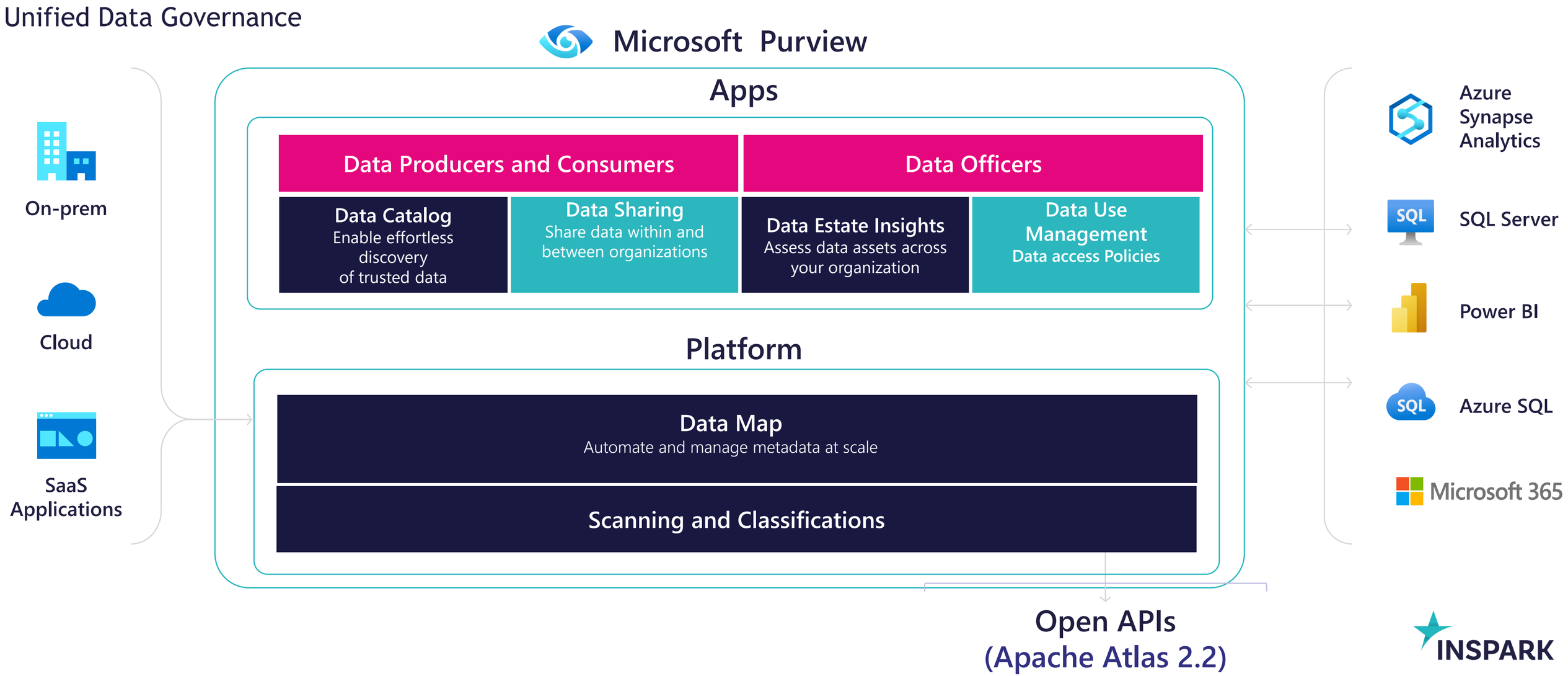
- Azure SQL Database
- Azure Storage
- Expiration Functionality
- Approval



Microsoft Purview

Generally Available

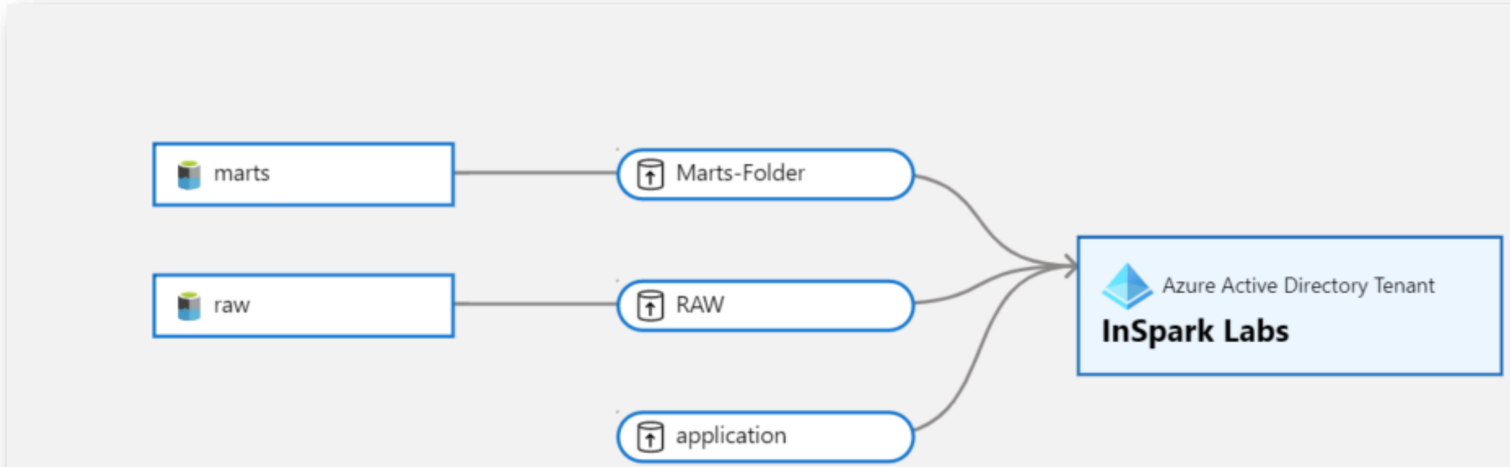
Preview

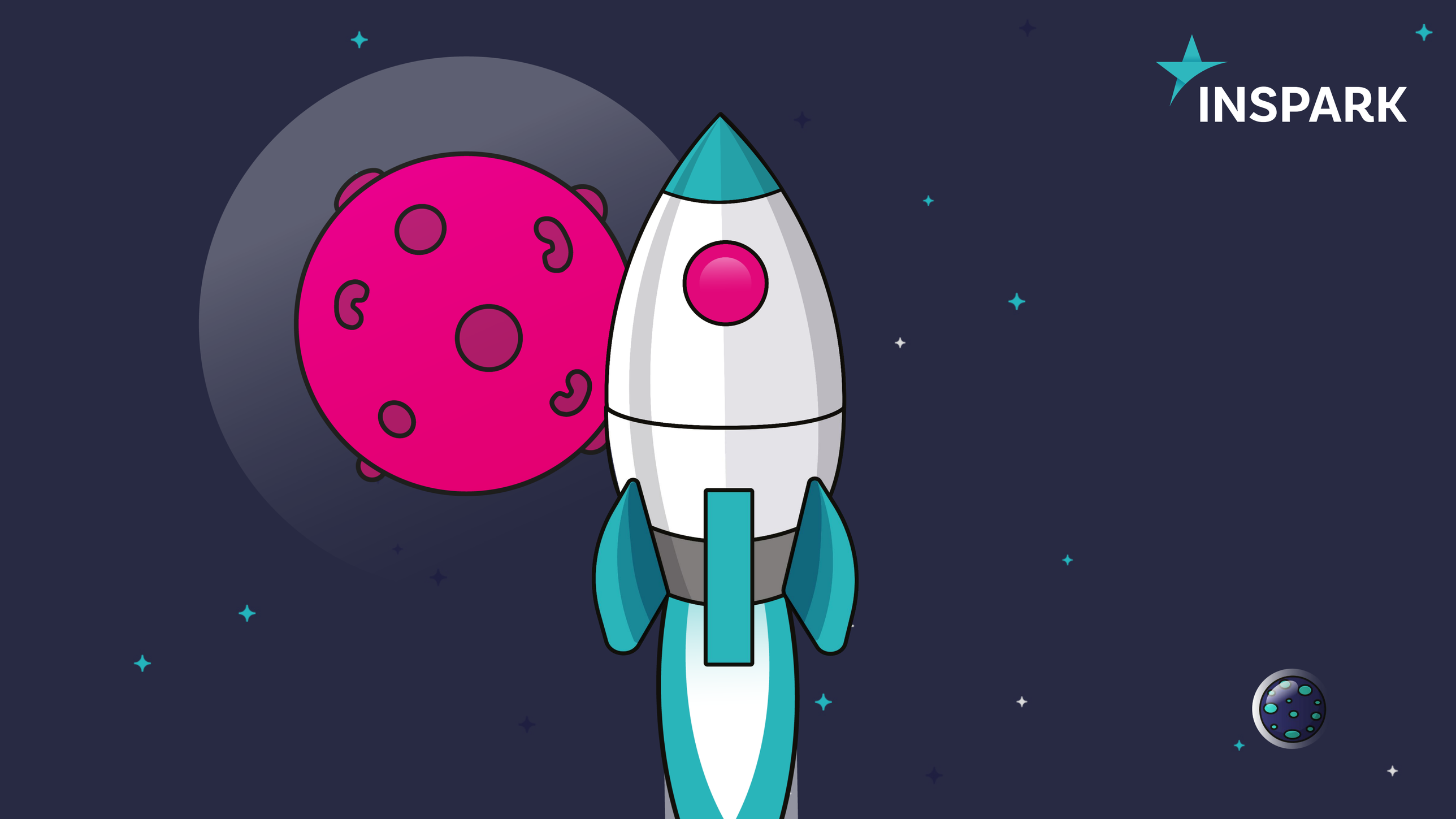


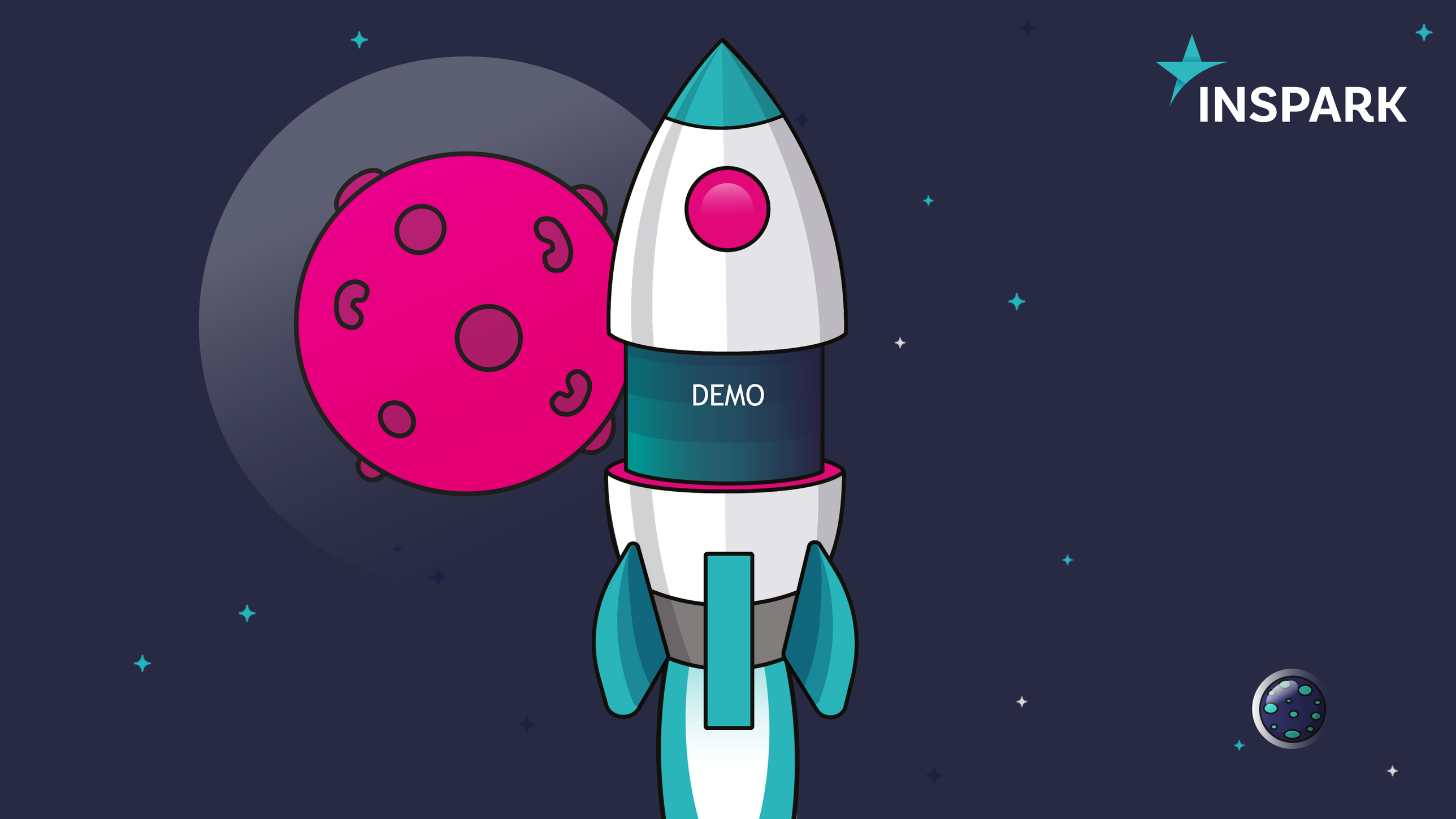
Data Sharing

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.




INSPARK


INSPARK
DEMO

Innovate to accelerate

DEMO

Microsoft Purview

Microsoft Purview Pricing

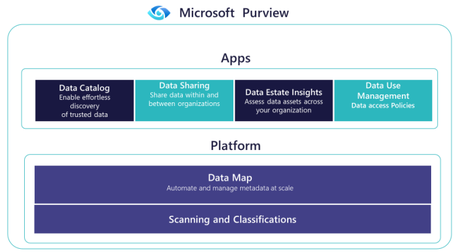
Data Map Consumption

- Elastic Data Map

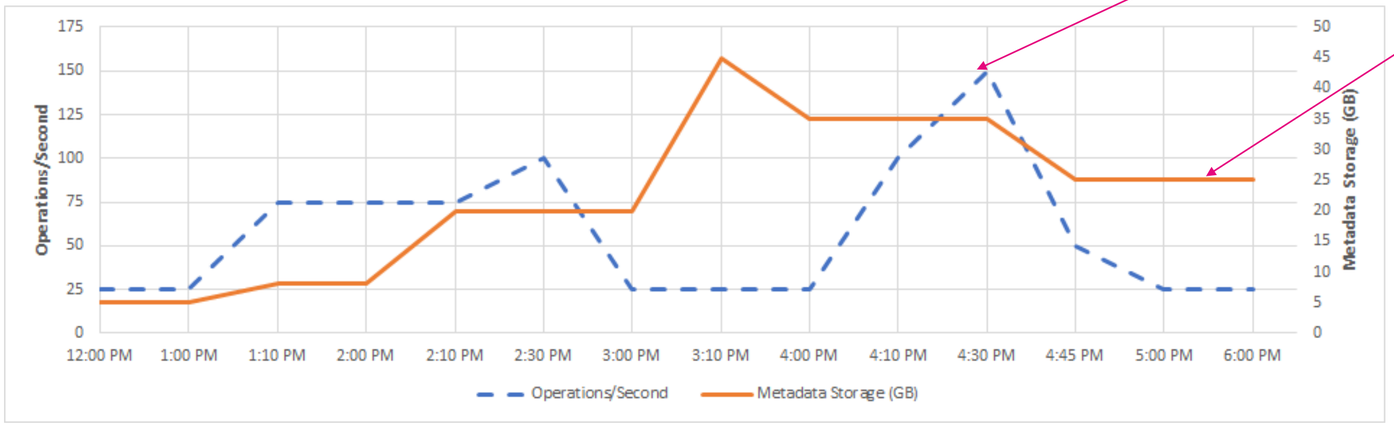
€0.373 per 1 Capacity Unit Hour

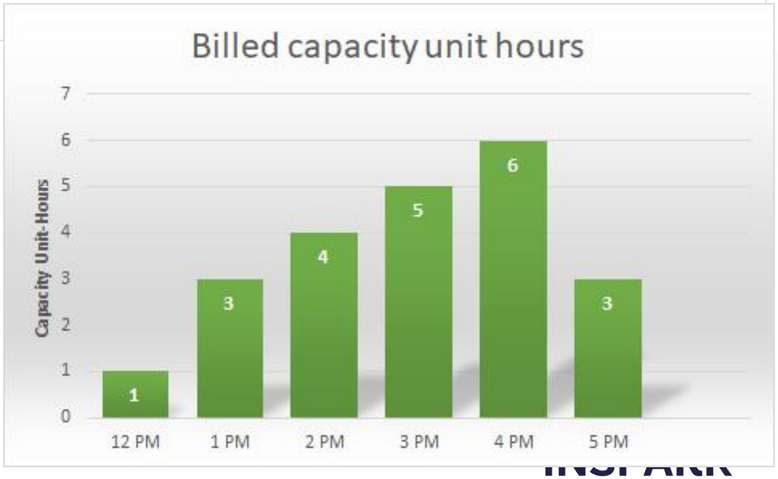
CU=Capacity Unit

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		





Microsoft Purview

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 / 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

INSPARK

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

ErwinDeKreuk.com

github.com/edkreuk

Microsoft® Most Valuable Professional

INSPARK