Data Governance With Microsoft Purview

Erwin de Kreuk

Microsoft® Most Valuable Professional

We Are InSpark

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

INSPARK

Microsoft Purview

Topics

- Introduction to Purview
- Data Map
- Data Catalog
- Data Estate Insights

INSPARK

Data is your most strategic asset

90%

Of corporate strategies will cite information as a critical enterprise asset by 2022

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

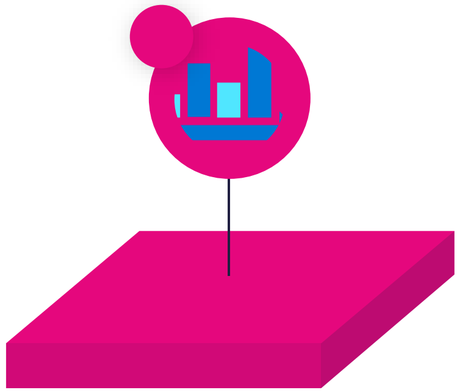
IDC

IDC Data Age 2025, Dave Reinsel,

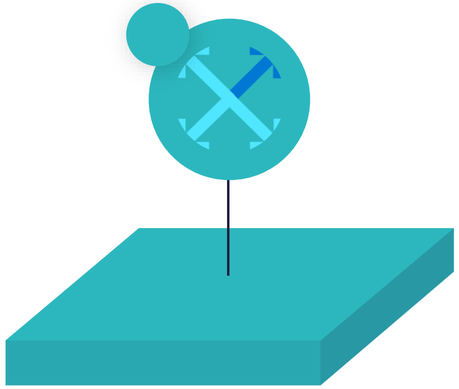
Microsoft Azure

Microsoft Purview

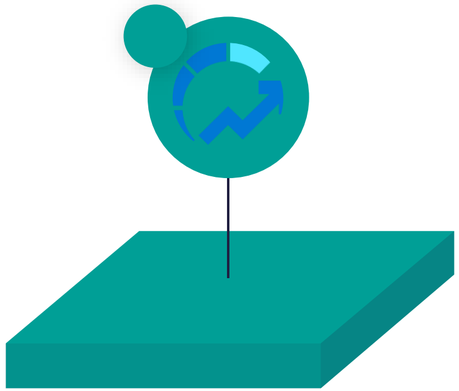
Elements of successful data governance



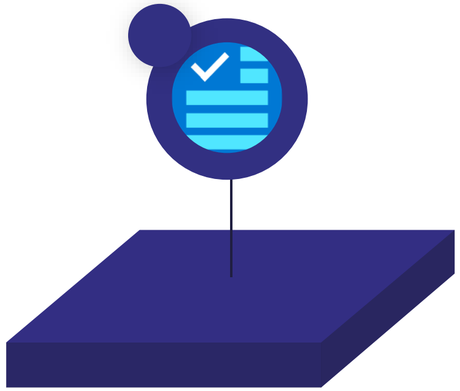
Manage growing data landscape



Overcome operational silos



Increase data agility



Comply with industry regulations

INSPARK

Microsoft Purview

Unified data governance to maximize the business value of data

Reimagine data governance in the cloud

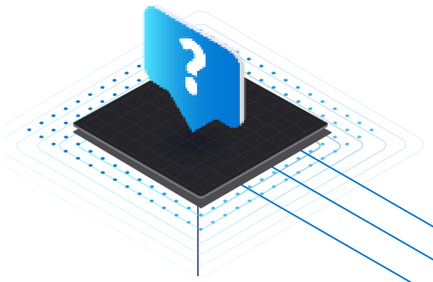
Set the foundation for effective data governance

Maximize business value of data for data consumers

Gain insight into data use across the estate

INSPARK

Data governance is becoming increasingly interdisciplinary



DISCOVERY

COMPLIANCE

Chief Data Officer

INSPARK

Microsoft Purview

Data governance is becoming increasingly interdisciplinary

Chief Data Officer

Data Management
Data Stewardship
Data Engineering
Data Science
Data Analysist

What data do I have?
Where did the data originate?
Can I trust it?
What's my exposure to risk?
Is my usage compliant?
How do I control access & use?
What is required by regulation X?

Data Compliancy
Data Access
Data Security
Data Privacy

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.

INSPARK

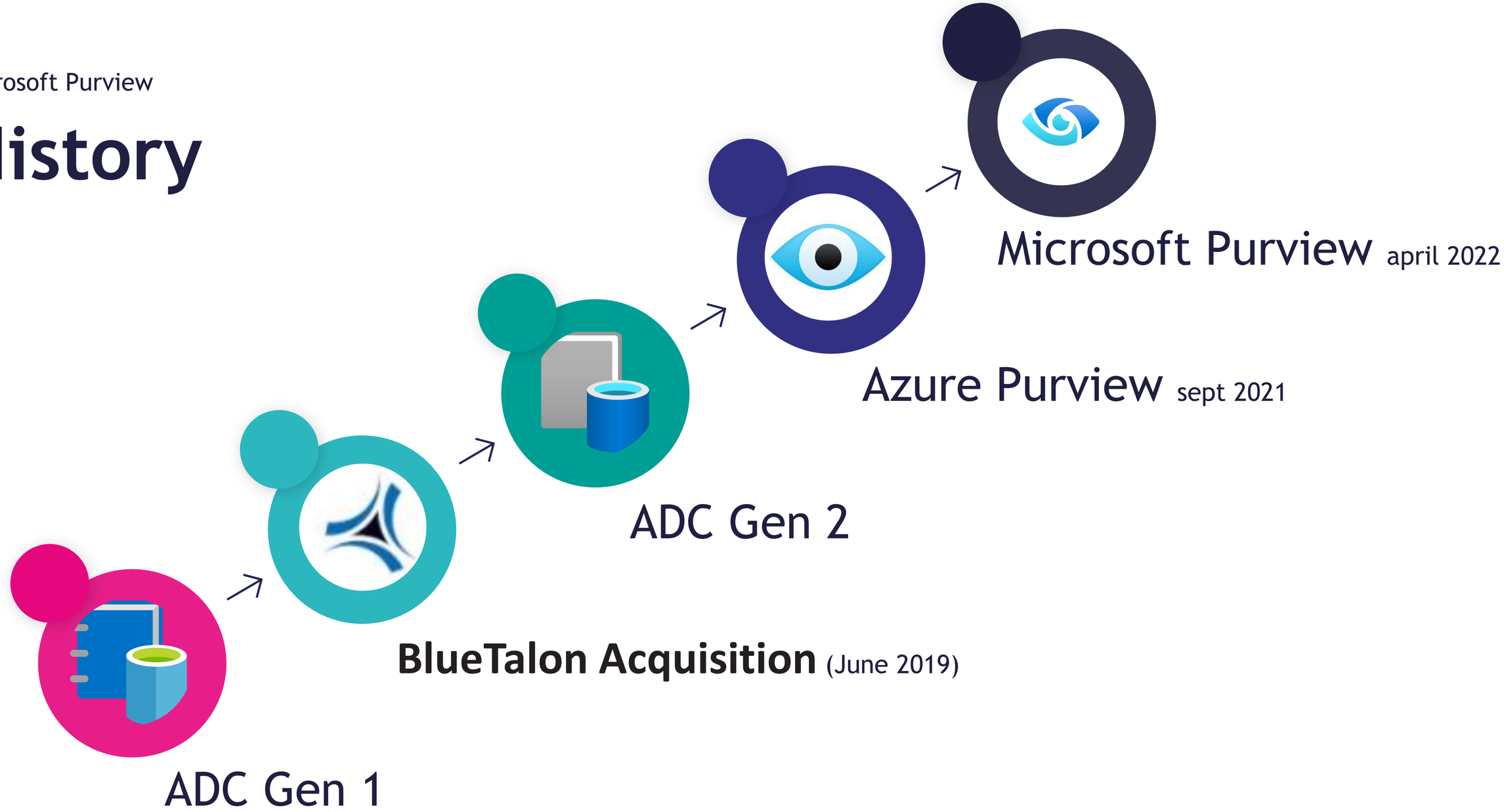
Unified data governance to maximize the business value of data

Microsoft Purview

INSPARK

Microsoft Purview

History



Microsoft Purview

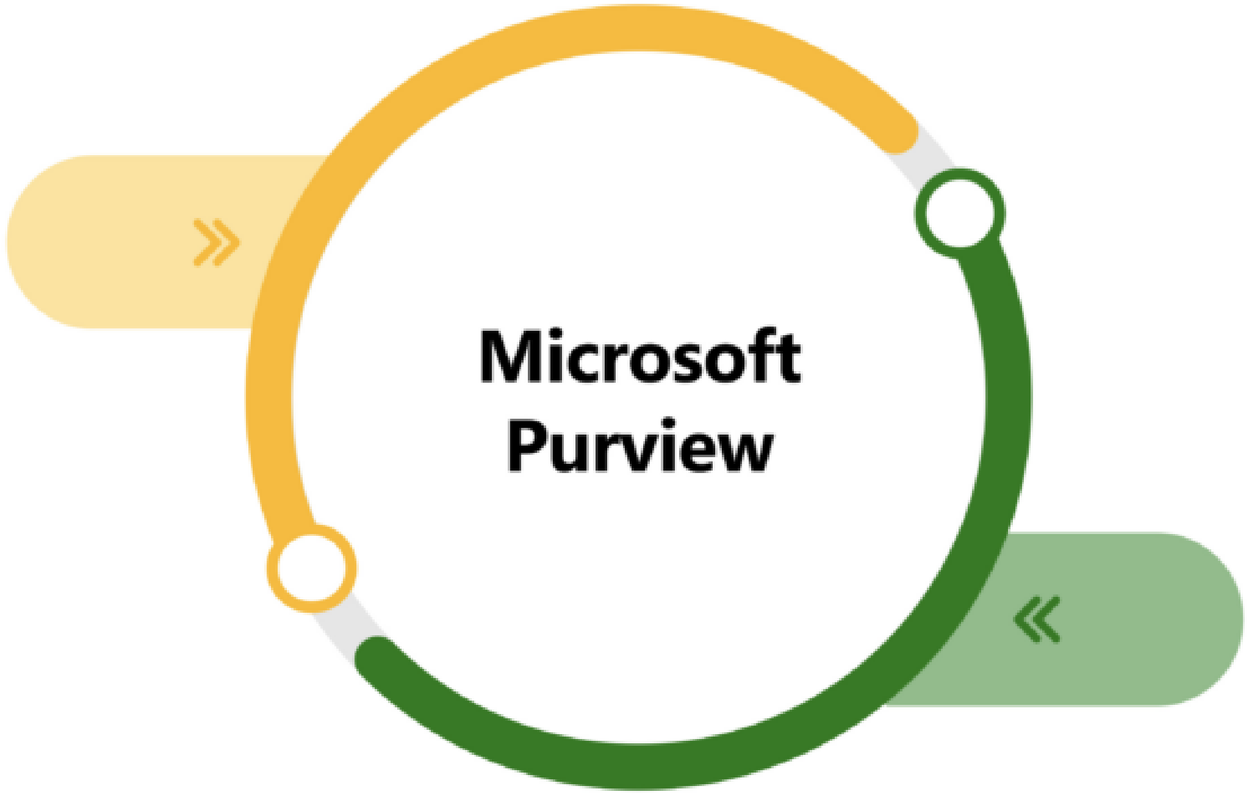
The future of compliance and data governance

Azure Purview is now Microsoft Purview

Microsoft Purview brings together trusted products for governance and compliance under one umbrella so it's easier to manage all of your data.

Risk & compliance

For risk, compliance, and legal teams



Unified data governance

For data consumers, data engineers, data officers

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

Unified Data Governance

On-prem

Cloud

SaaS Applications

Microsoft Purview

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

Open APIs (Apache Atlas 2.2)

Scanning and Classifications

Compliance Manager

Insider Risk Management

Data Loss Prevention

Data Lifecycle Management

Communication Compliance

eDiscovery

Information Protection

Azure Synapse Analytics

SQL Server

Power BI

Azure SQL

Microsoft 365

INSPARK

Microsoft Purview

Microsoft Purview Studio

Data Producers & Consumers

Data Officers & Security Officers

Microsoft Purview Platform

Microsoft Purview Data Map

- Dedicated per customer on shared infra
- Provisioned default capacity with option to add-on capacity

Automated Scanning & Classification

- Serverless, pay per use
- Includes connectors, scanning of sources, processing into data assets, lineage capture, classification

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Ce

Open APIs
(Apache Atlas 2.2)

On-prem & Multi-cloud

Operational, Analytical, SaaS

INSPARK

Microsoft Purview

Microsoft Purview Studio

Data Producers & Consumers

Data Officers & Security Officers

Microsoft Purview Data Catalog

- Business Glossary templates
- Lineage visualization & workflows
- Search, browse, asset details
- Automated meta-data and lineage extraction
- Automated classification based on content inspection
- Private Endpoint
- Management center

Microsoft Purview Platform

Microsoft Purview Data Map

- Dedicated per customer on shared infra
- Provisioned default capacity with option to add-on capacity

Automated Scanning & Classification

- Serverless, pay per use
- Includes connectors, scanning of sources, processing into data assets, lineage capture, classification

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Ce

Open APIs
(Apache Atlas 2.2)

On-prem & Multi-cloud

Operational, Analytical, SaaS

INSPARK

Microsoft Purview

Microsoft Purview Studio

Data Producers & Consumers

Microsoft Purview Data Catalog

- Business Glossary templates
- Lineage visualization & workflows
- Search, browse, asset details
- Automated meta-data and lineage extraction
- Automated classification based on content inspection
- Private Endpoint
- Management center

Data Officers & Security Officers

Microsoft Purview Data Estate Insights

- Catalog Insights (Asset, Scan, Glossary)
- Sensitive Information Types & Labeling insights

Microsoft Purview Platform

Microsoft Purview Data Map

- Dedicated per customer on shared infra
- Provisioned default capacity with option to add-on capacity

Automated Scanning & Classification

- Serverless, pay per use
- Includes connectors, scanning of sources, processing into data assets, lineage capture, classification

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Ce

Open APIs
(Apache Atlas 2.2)

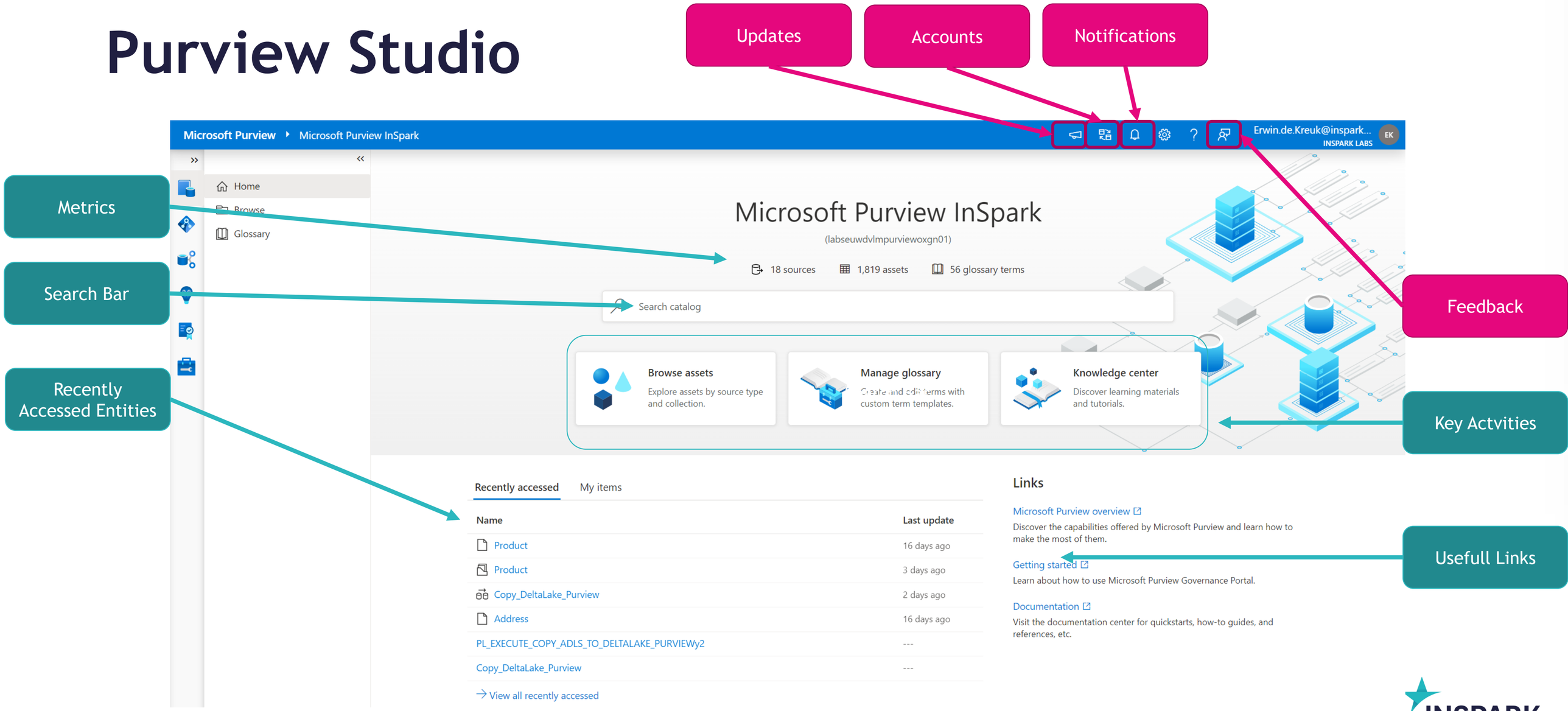
On-prem & Multi-cloud

Operational, Analytical, SaaS

INSPARK

Microsoft Purview

Purview Studio

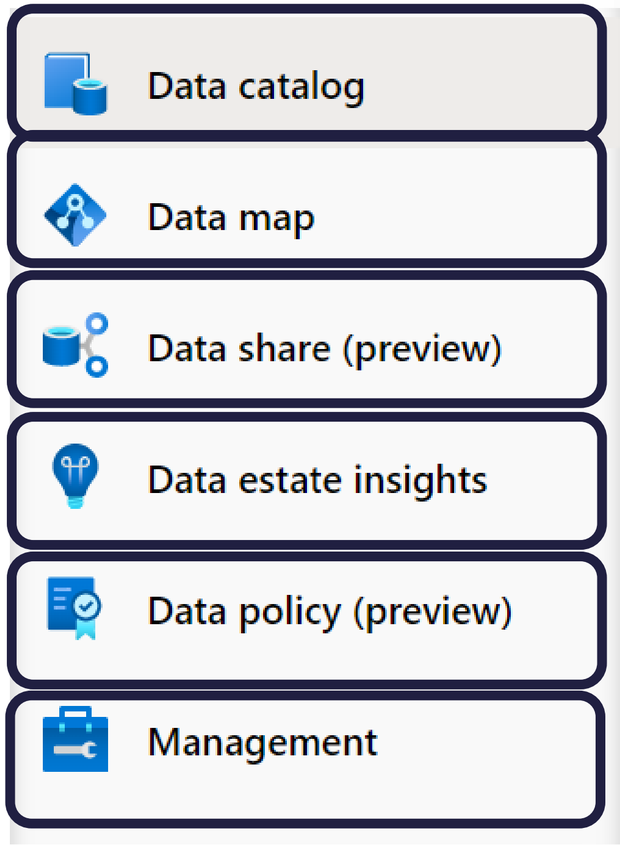


INSPARK

Microsoft Purview

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

Microsoft Purview Pricing

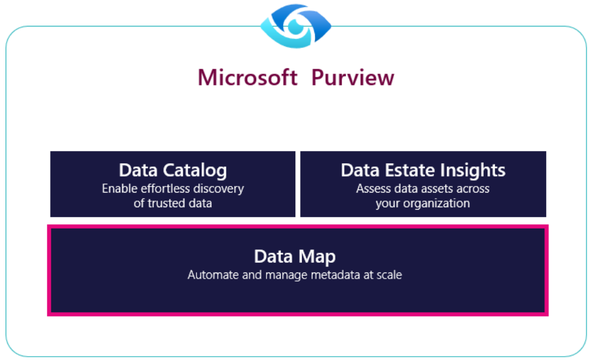
Data Map Consumption

- Elastic Data Map

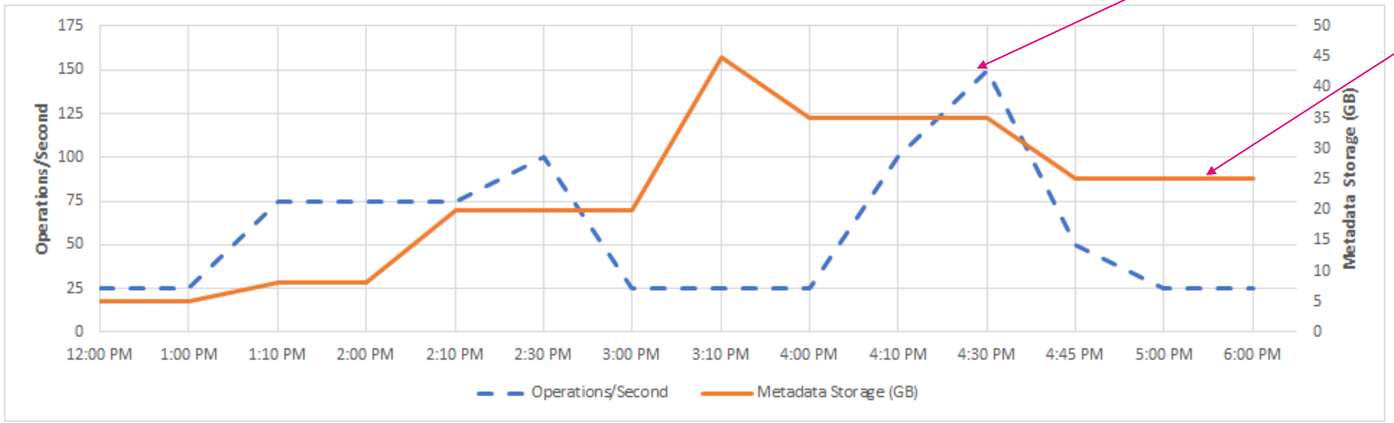
€0.407 per 1 Capacity Unit Hour

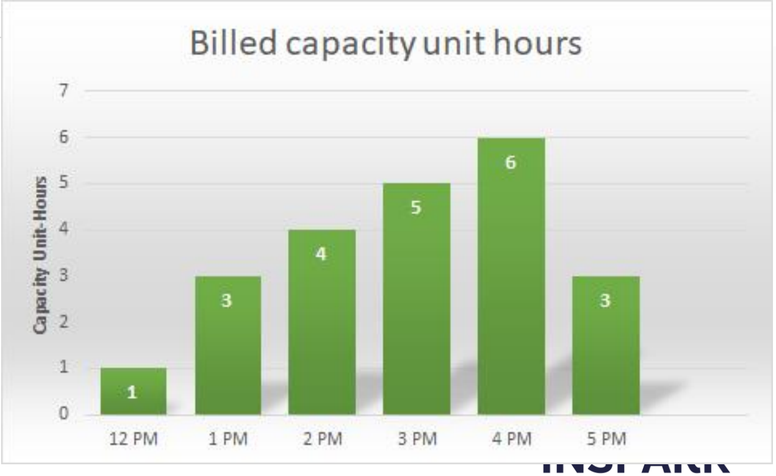
CU=Capacity Unit

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.



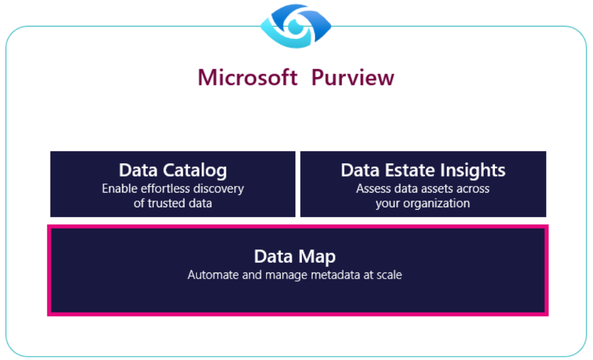
Data Map Capacity Unit	Operations/Sec throughput	Storage capacity in GB
1	25	10
2	50	20
3	75	30
4	100	40
5	125	50
6	150	60
7	175	70
8	200	80
9	225	90
10	250	100
..	..	..
100		





Microsoft Purview Pricing

Data Map Population



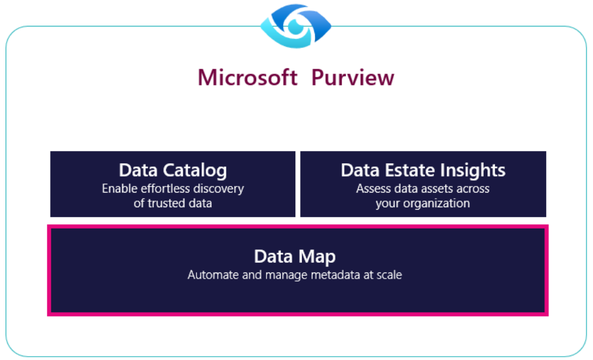
Automated Scanning, Ingestion & Classification

- Power BI Online
  Free for a limited time
- SQL Server On Prem
  Free for a limited time
- Other Data Sources
  €0.623 per 1 vCore Hour

INSPARK

Microsoft Purview Pricing

Data Map Enrichement



Features

- Advanced Resource Set

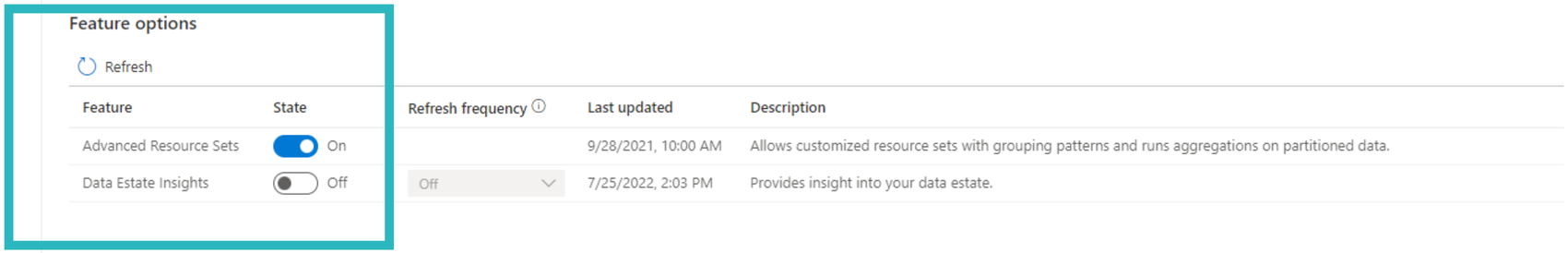
  €0.208 per 1 vCore Hour

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

  €0.811 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.

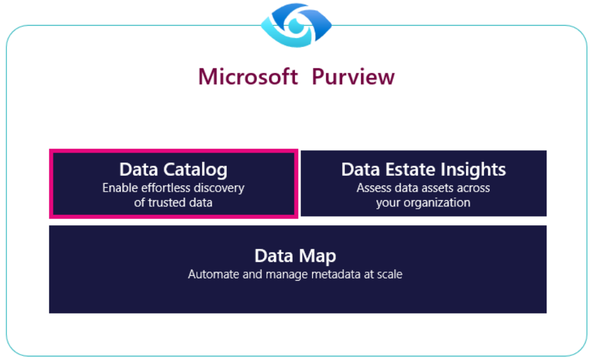


INSPARK

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



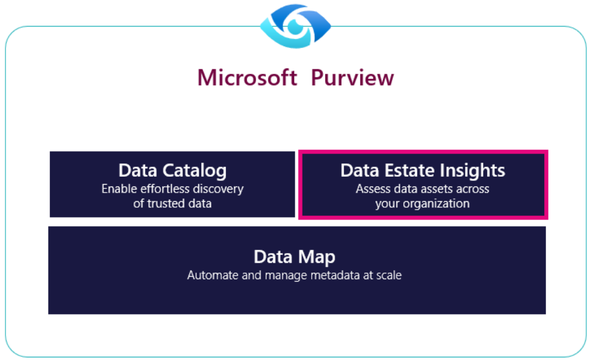
Microsoft Purview Pricing

Data Estate Insights

- Insights Consumption

  €0.200 per API call

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



INSPARK

Microsoft Purview

Preview Features

- Data Policy
  - SQL DevOps access
    Free in preview
  - Data Lake data asset access
    Free in preview
- Workflows
  - Business Workflows
    Free in preview
- Data Sharing
  - In place sharing for ABLB and ADLS
    Free in preview

INSPARK

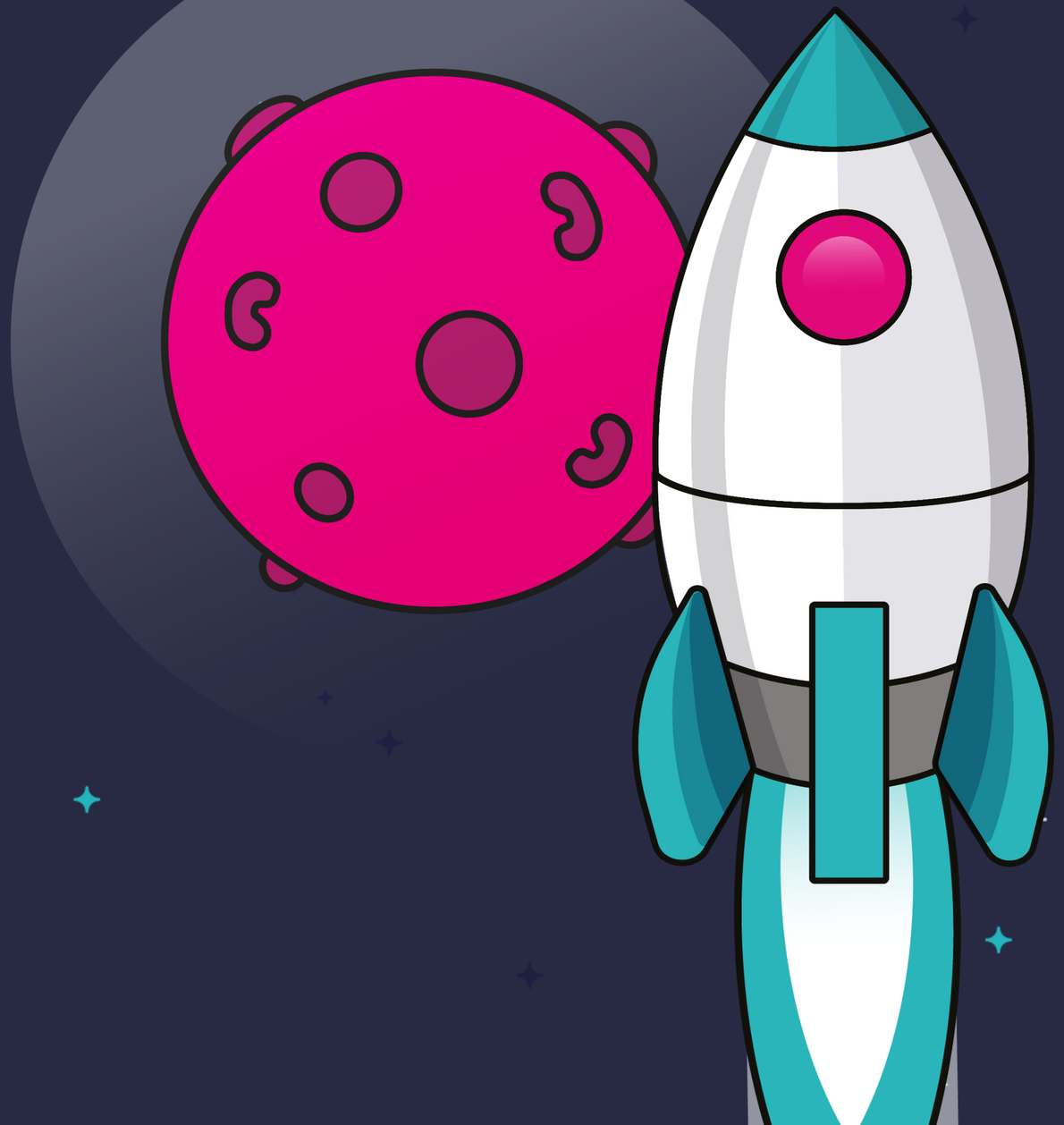
Microsoft Purview

Pricing - Example

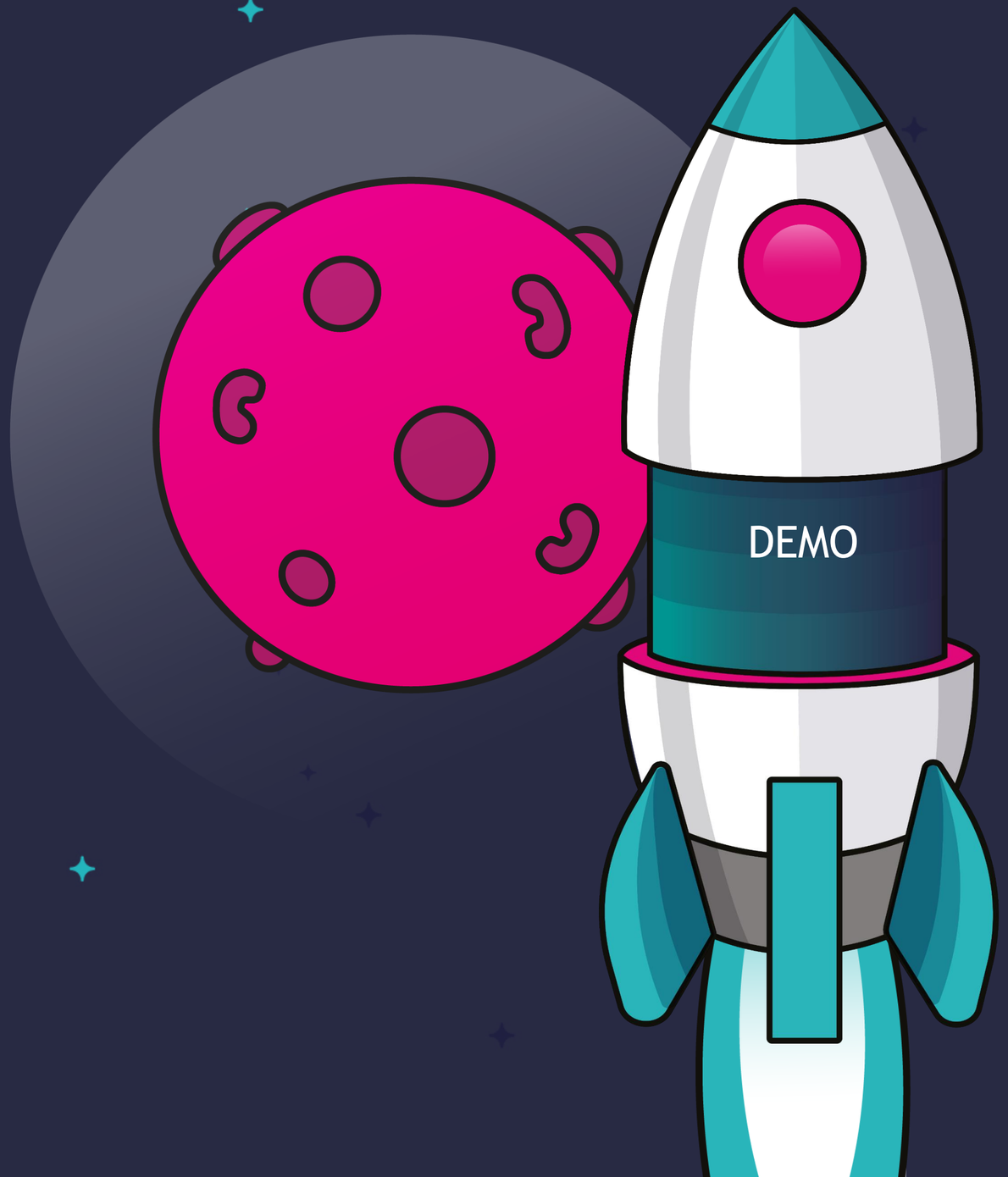
Data Map Consumption	Data Map Population	Data Map Enrichement
1,5 CPU x €0.407 X 730 hours	4 scans x 4 hours x 32 VCore x €0.623 per vCore per hour	60 scans x 1 hour x 8 VCore x €0.208 per 1vCore per hour(Resource set)
€ 445	€ 320	€ 100,00

€865,00

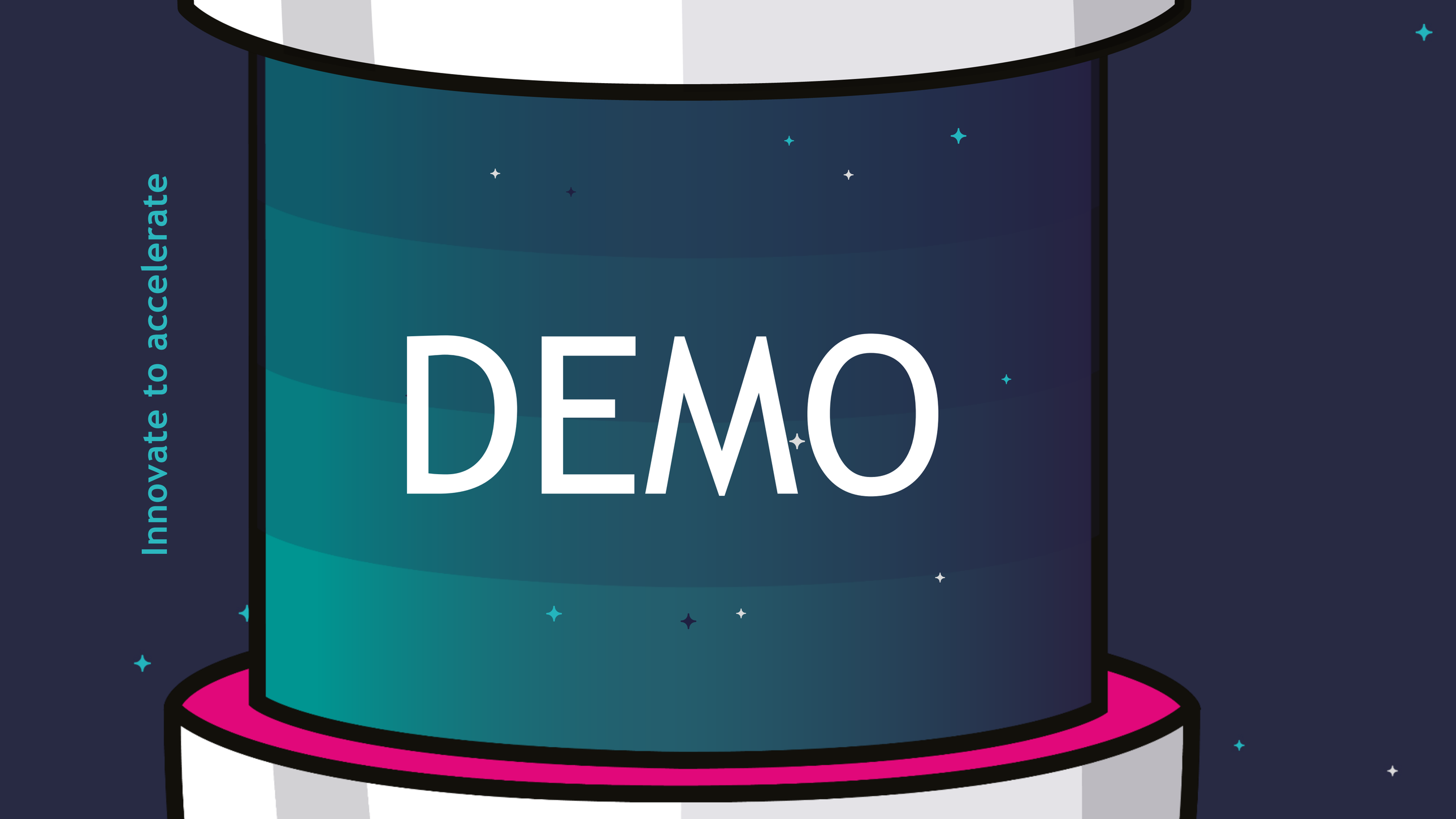
INSPARK



INSPARK



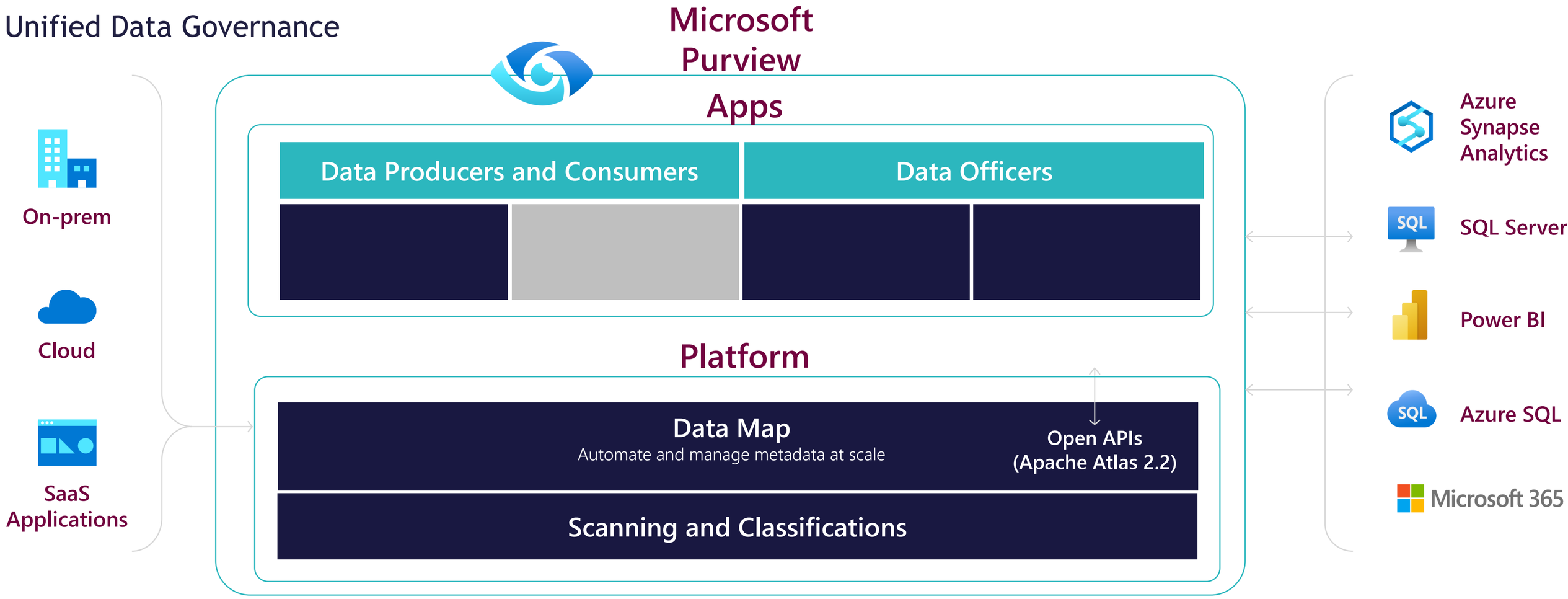
INSPARK


Innovate to accelerate
DEMO

Microsoft Purview

Generally Available

Preview

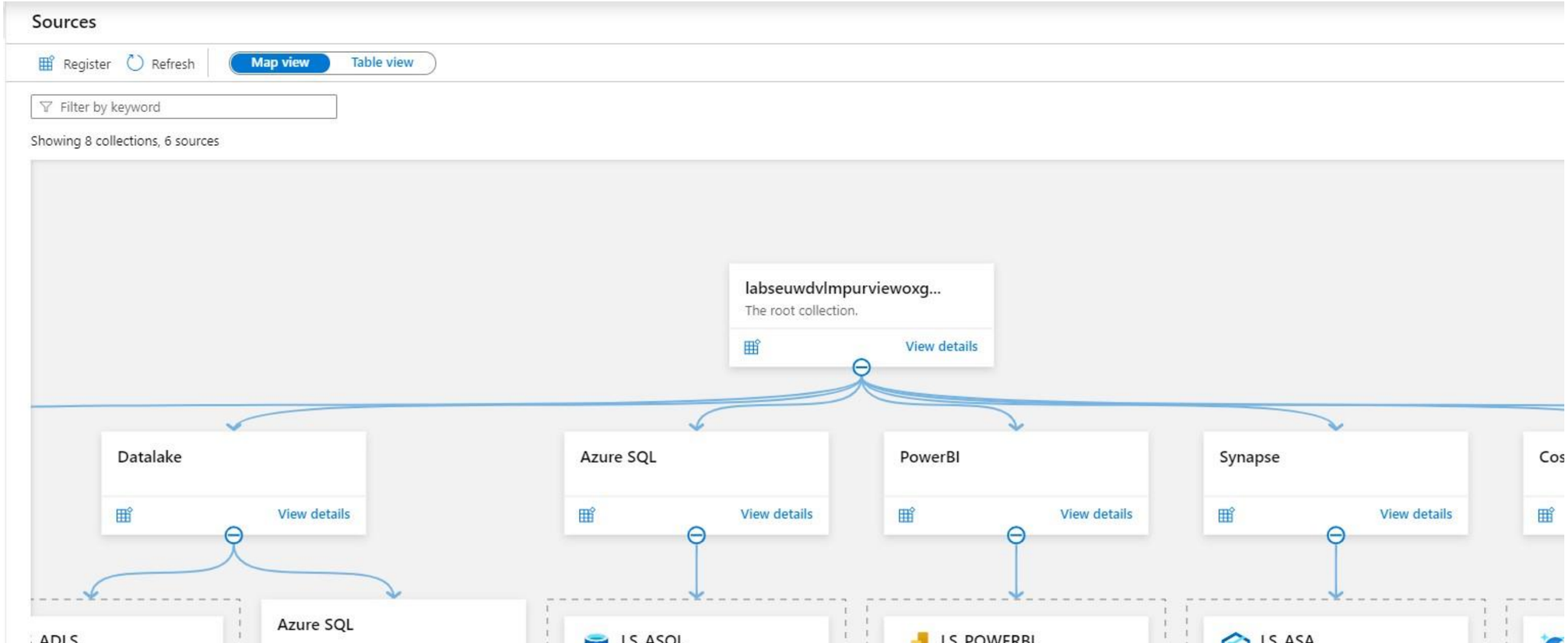


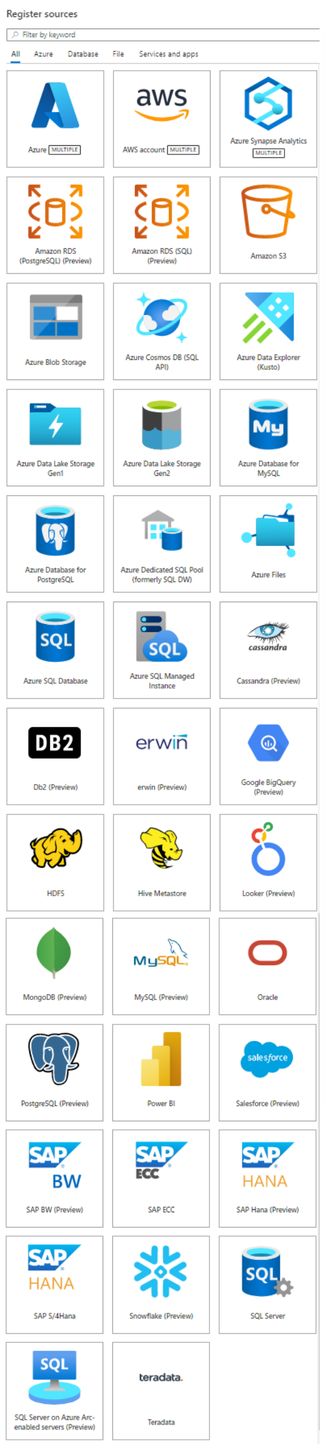
INSPARK

Data Map

Collections and Sources

- A graph describing the data assets and their relationships across your data estate
- Support for 35+ data sources and growing

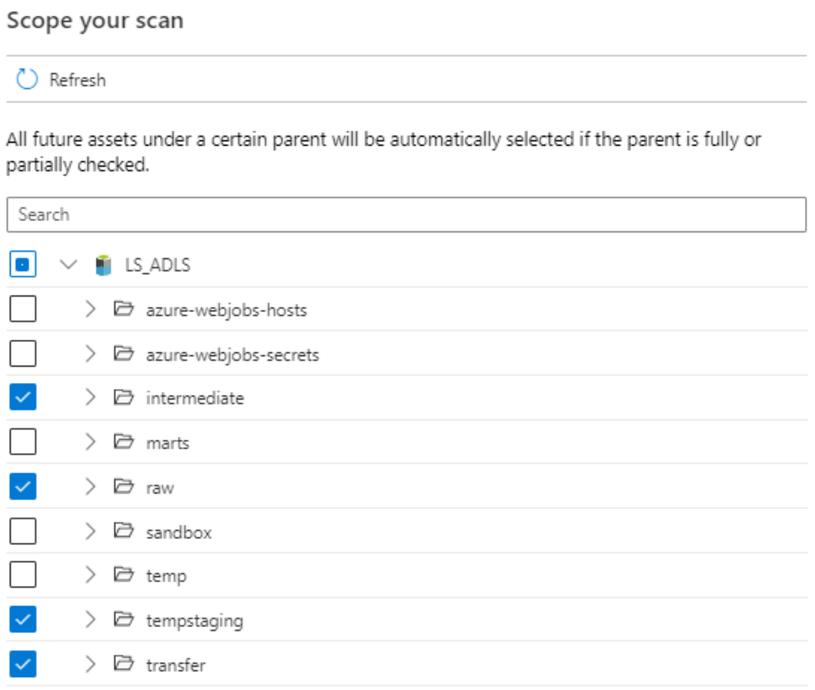


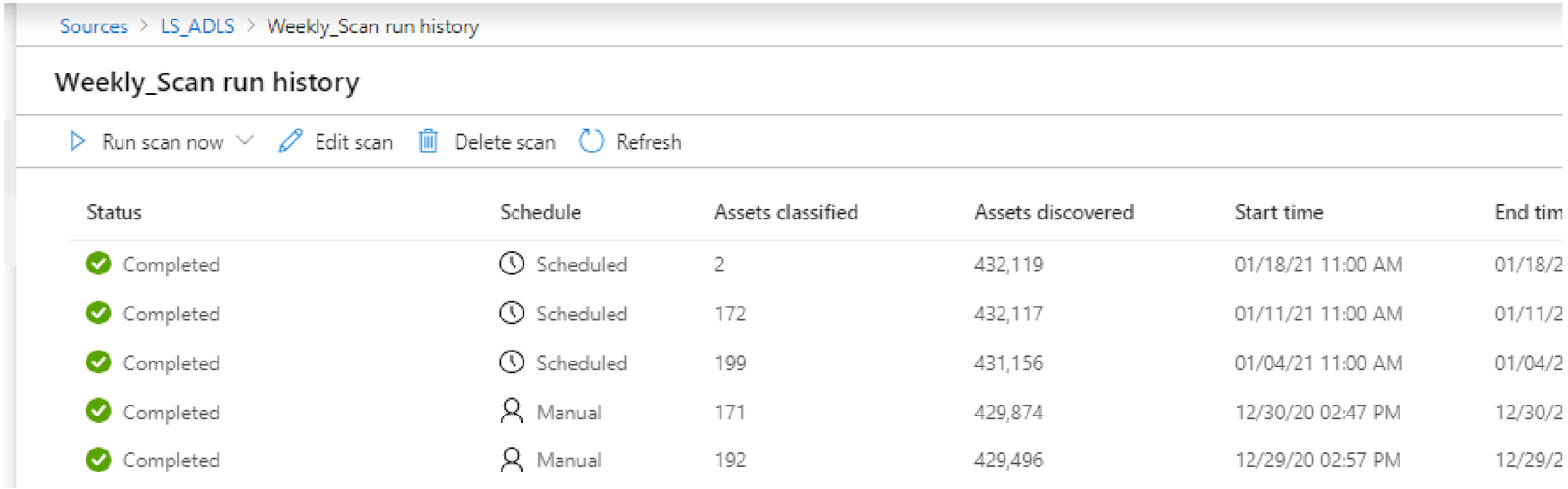


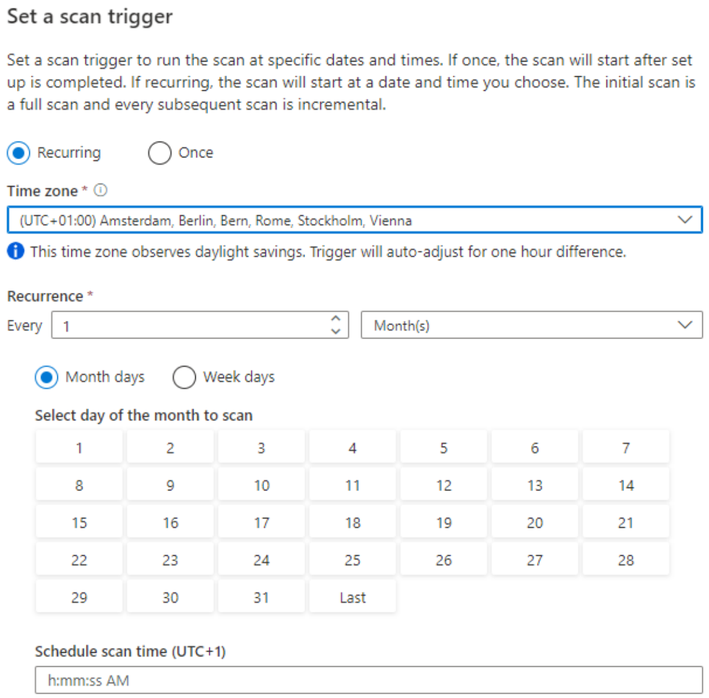
Data Map

Scans

- A graph describing the data assets and their relationships across your data estate
- Support for 35+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores



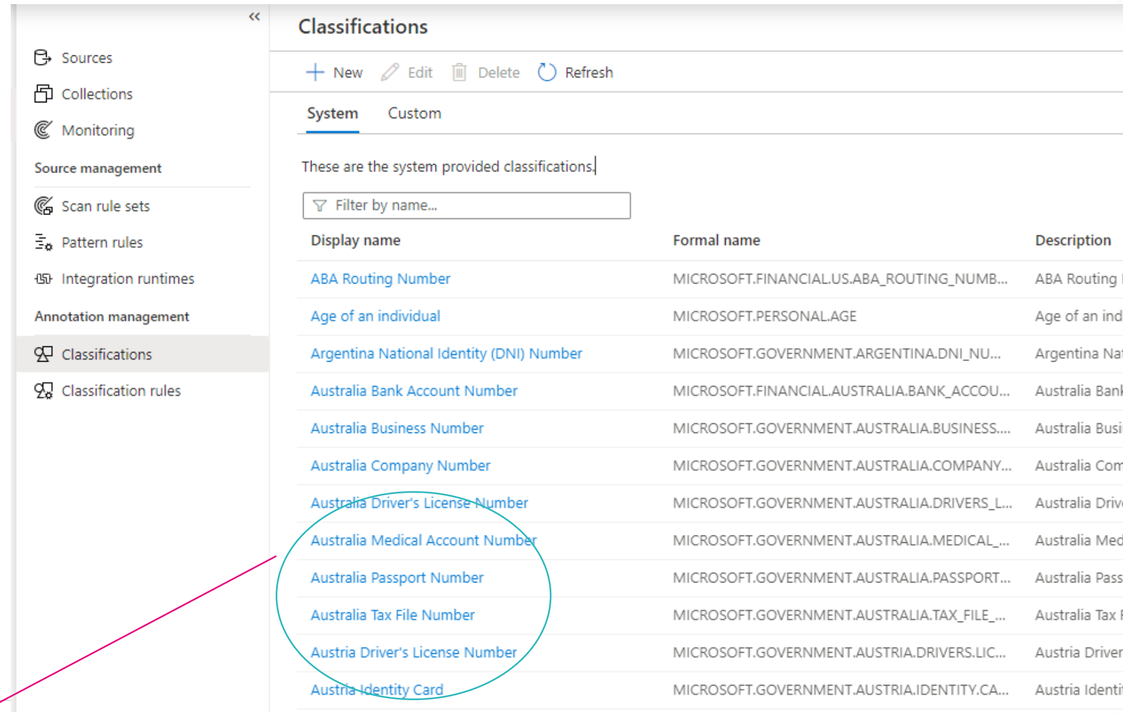


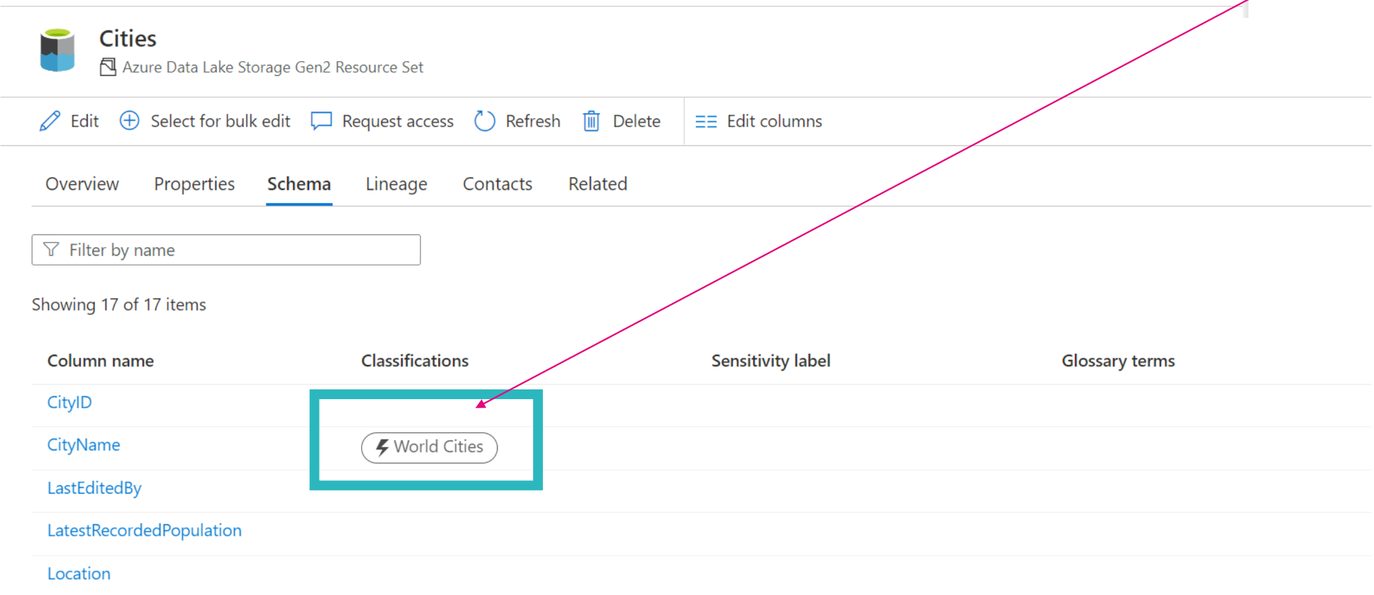


Data Map

Classifications

- A graph describing the data assets and their relationships across your data estate
- Support for 35+ data sources and growing
- Automated data scanning, classification and lineage extraction of hybrid data stores
- 200+ built-in data classifiers





Microsoft Purview

Generally Available

Preview

Unified Data Governance

On-prem

Cloud

SaaS Applications

Microsoft Purview

Apps

Data Producers and Consumers

Data Officers

Data Catalog
Enable effortless discovery of trusted data

Platform

Data Map
Automate and manage metadata at scale

Open APIs (Apache Atlas 2.2)

Scanning and Classifications

Azure Synapse Analytics

SQL Server

Power BI

Azure SQL

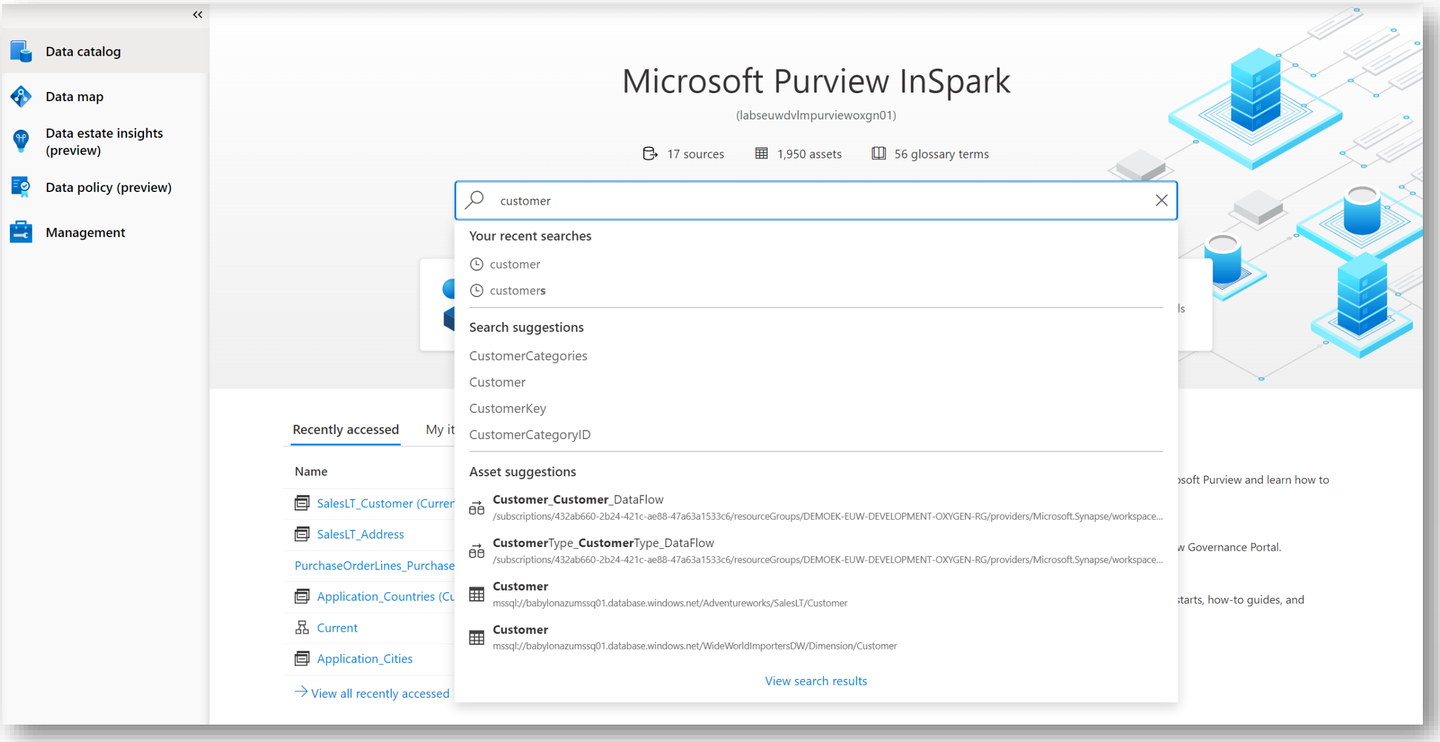
Microsoft 365

INSPARK

Data Catalog

Search

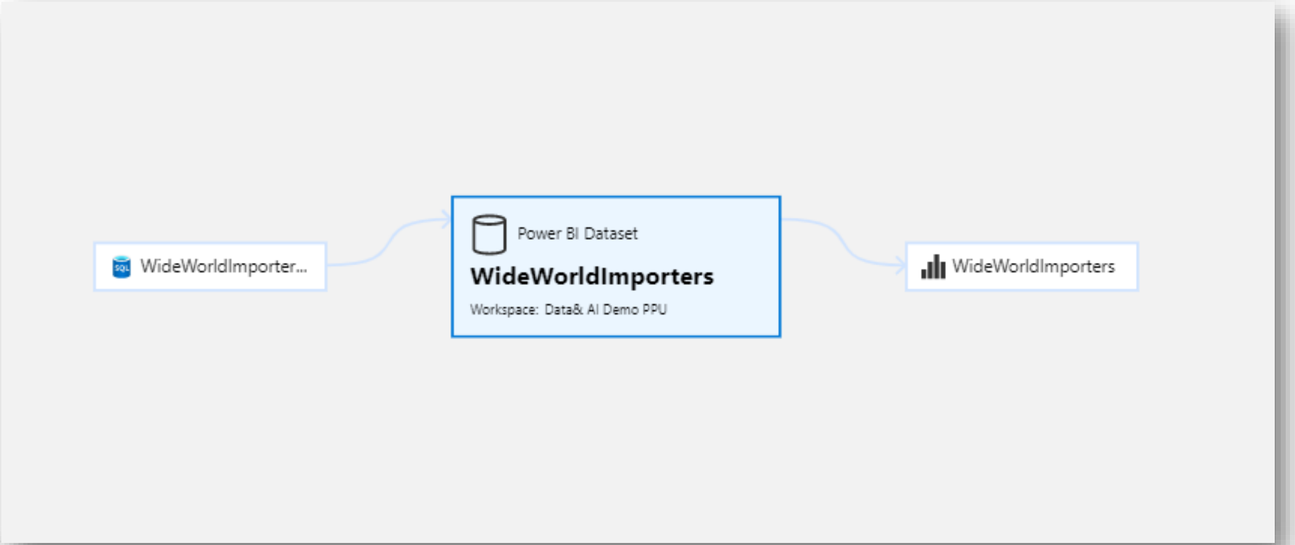
- Semantic search and browse
- Business glossary and workflows

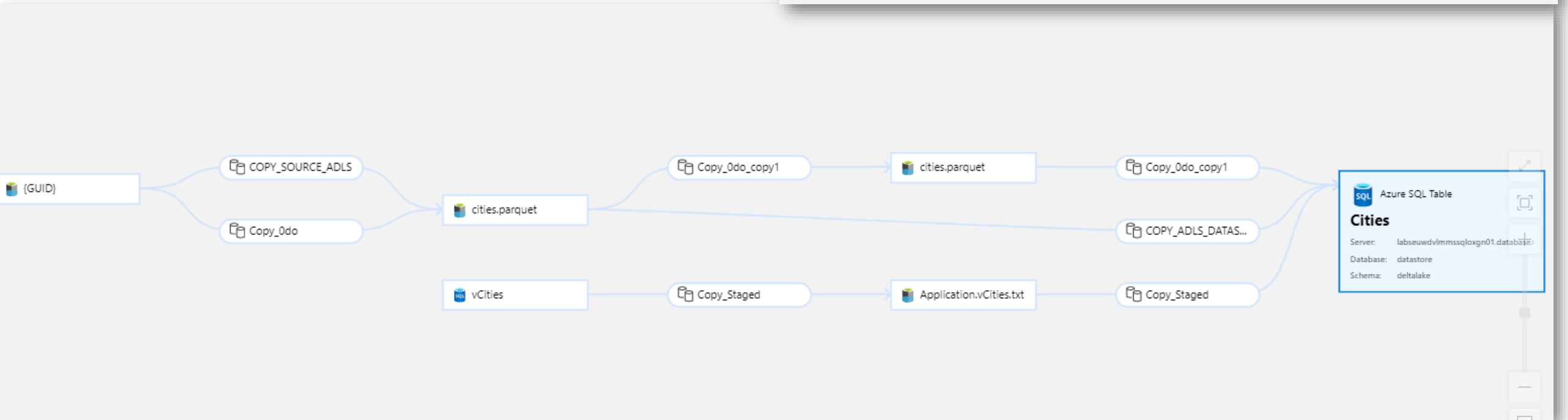


Data Catalog

Lineage

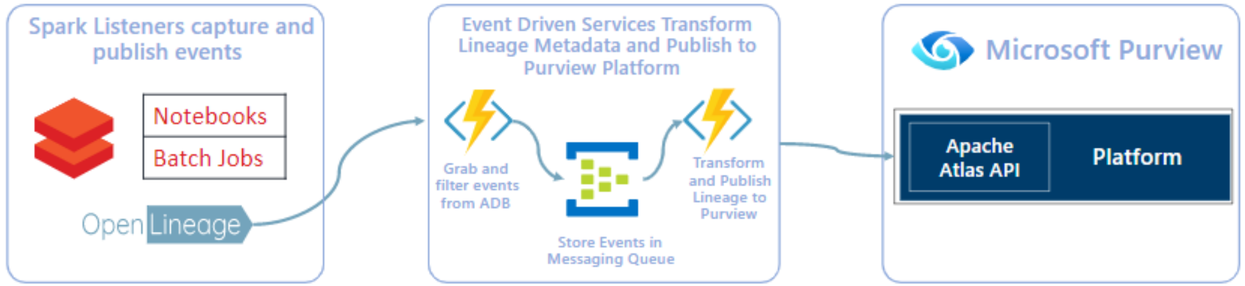
- Semantic search and browse
- Business glossary and workflows
- Data lineage with sources, owners, transformations, and lifecycle





Data Catalog

Custom Lineage

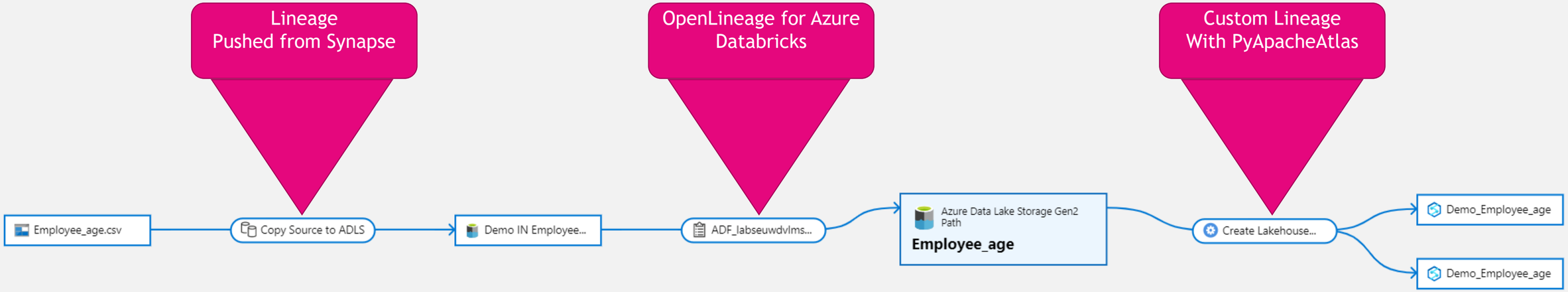


- OpenLineagefor Azure DataBricks

[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

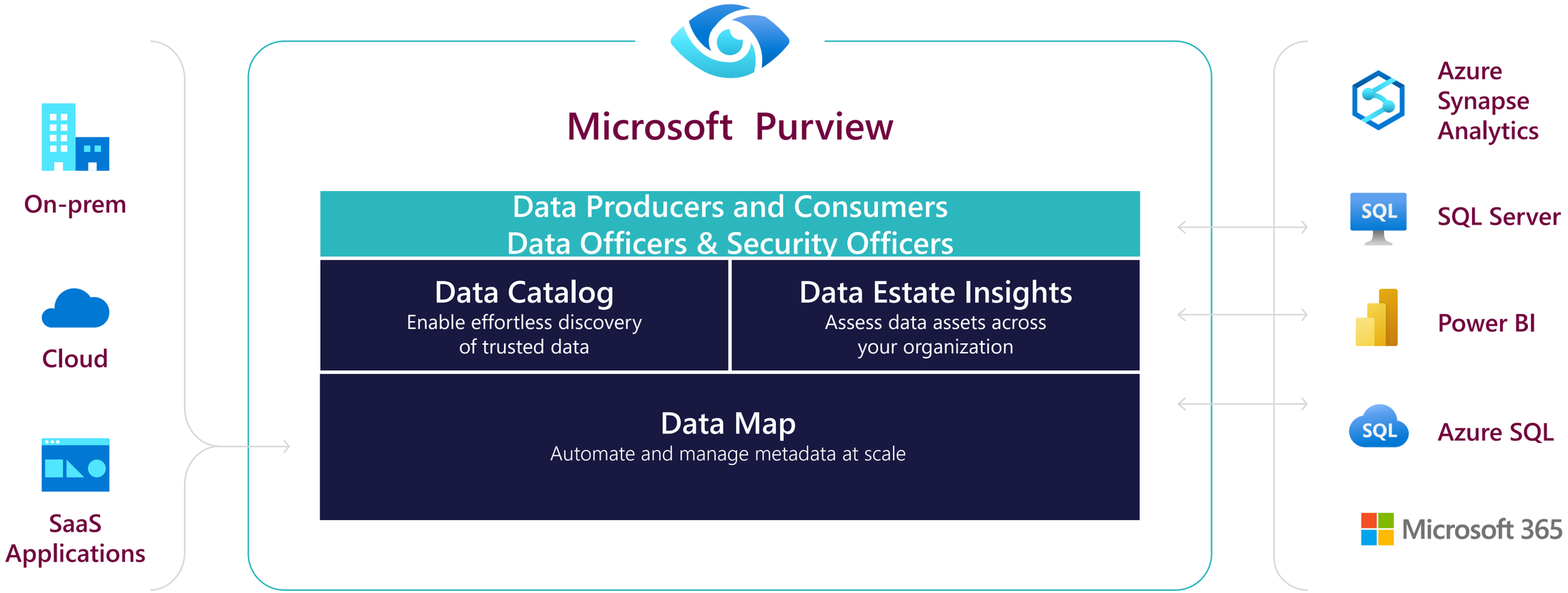


Generally Available

Preview

Microsoft Purview

Unified Data Governance



INSPARK

Microsoft Purview

Generally Available

Preview

Unified Data Governance

On-prem

Cloud

SaaS Applications

Microsoft Purview

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

Open APIs (Apache Atlas 2.2)

Scanning and Classifications

Azure Synapse Analytics

SQL Server

Power BI

Azure SQL

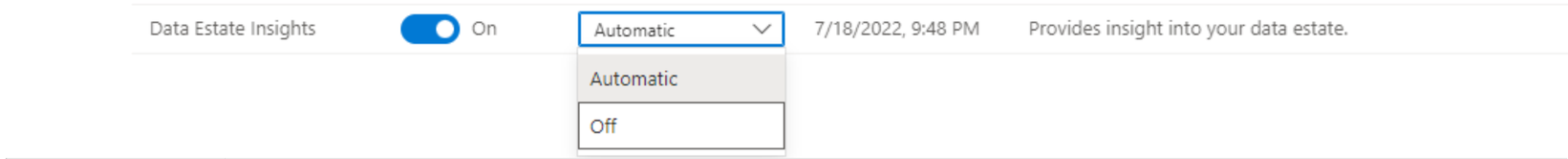
Microsoft 365

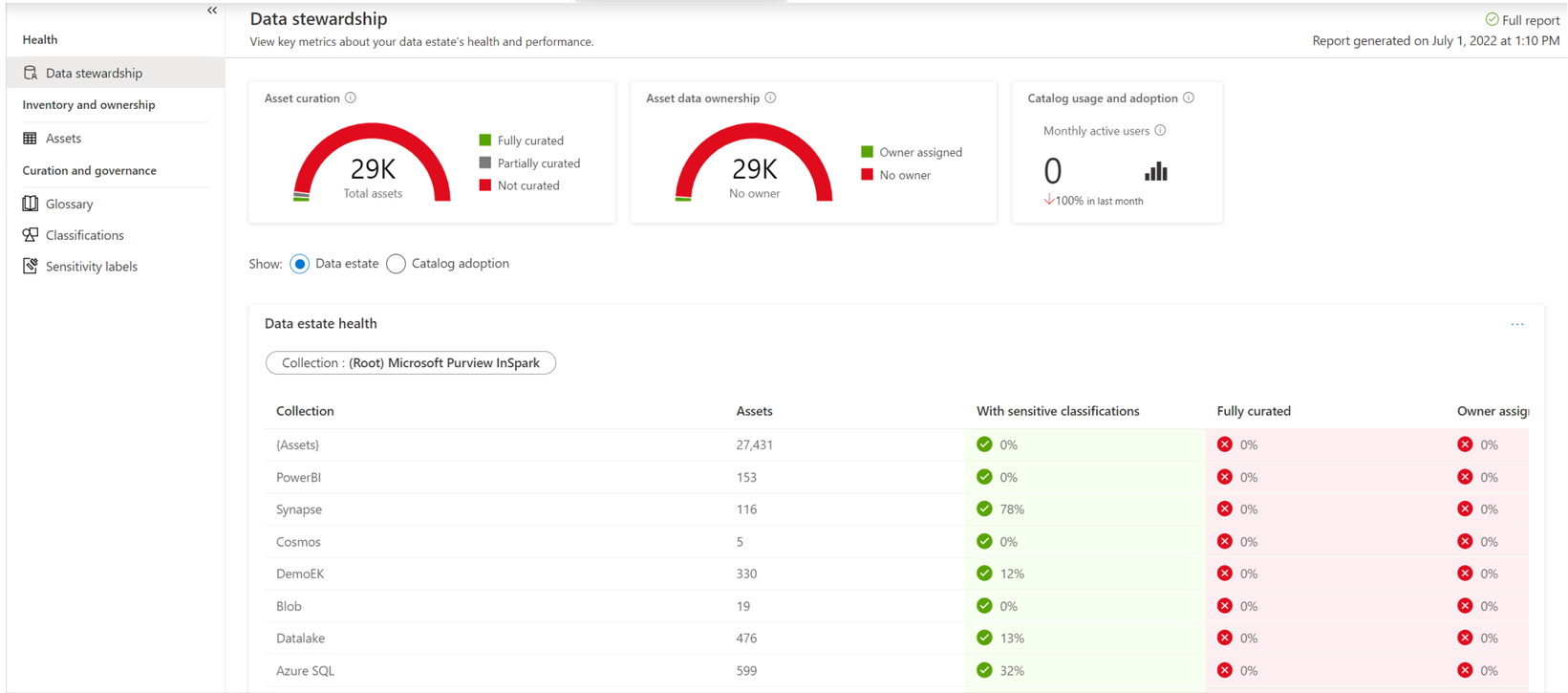
INSPARK

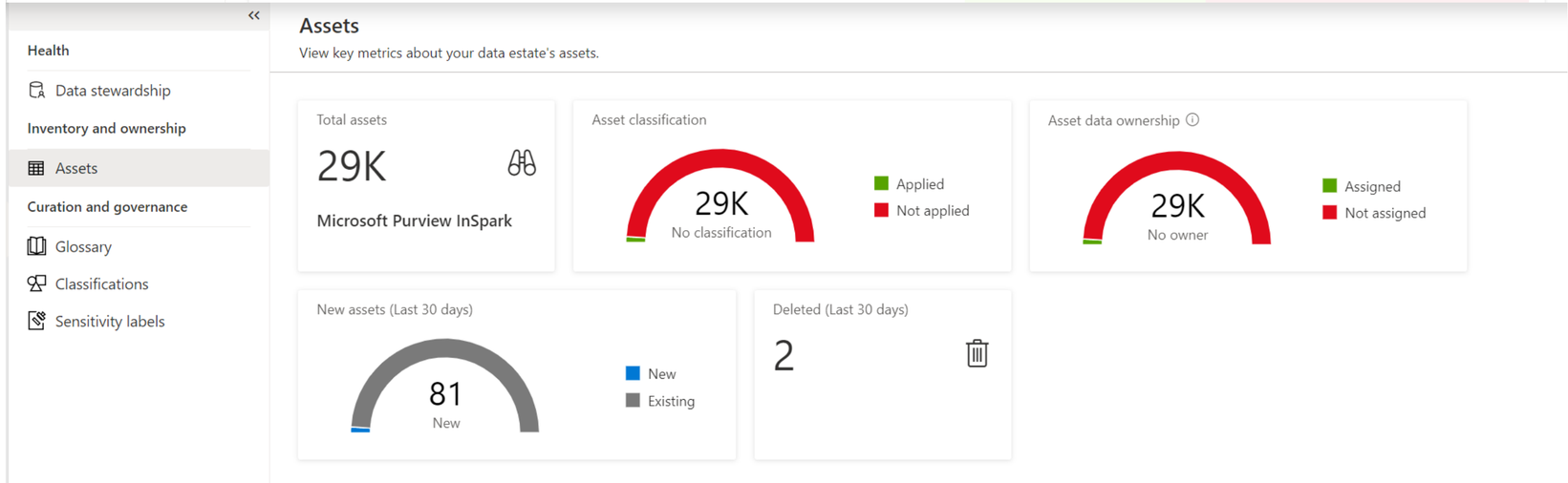
Data Estate Insight

Metrics

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





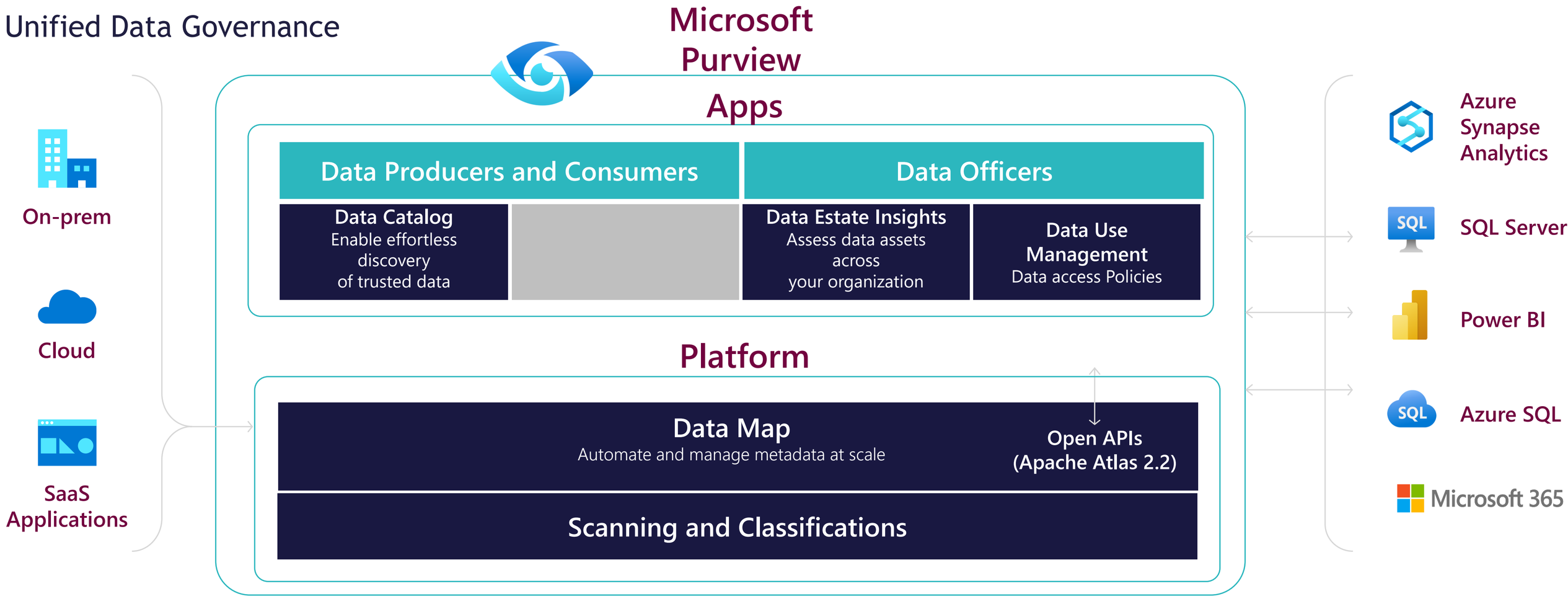


INSPARK

Microsoft Purview

Generally Available

Preview

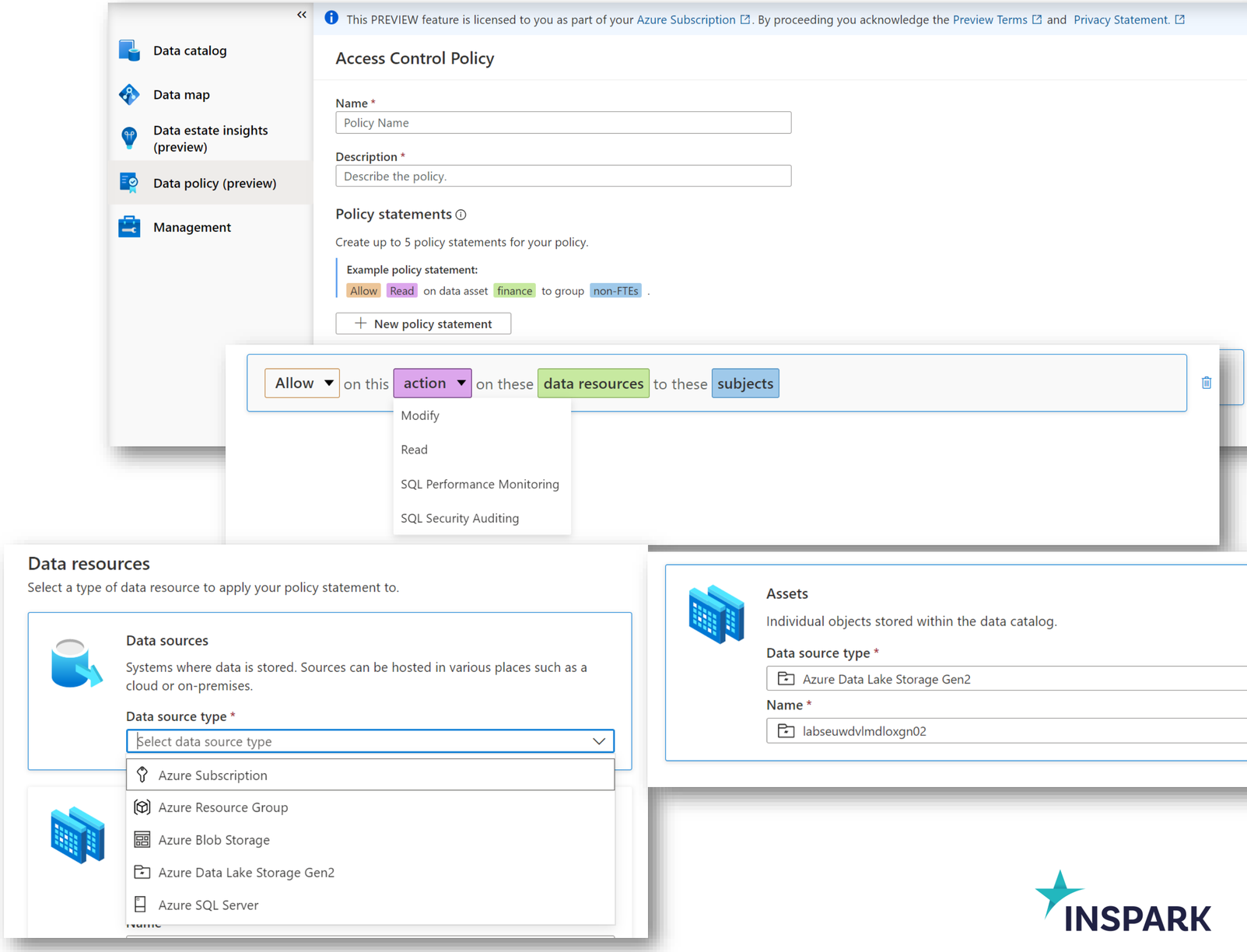


INSPARK

Data Use Management

Data Owner

- Azure Storage
- Azure SQL DB
- Azure Arc Enabled Servers
- Resource Groups - Subscriptions

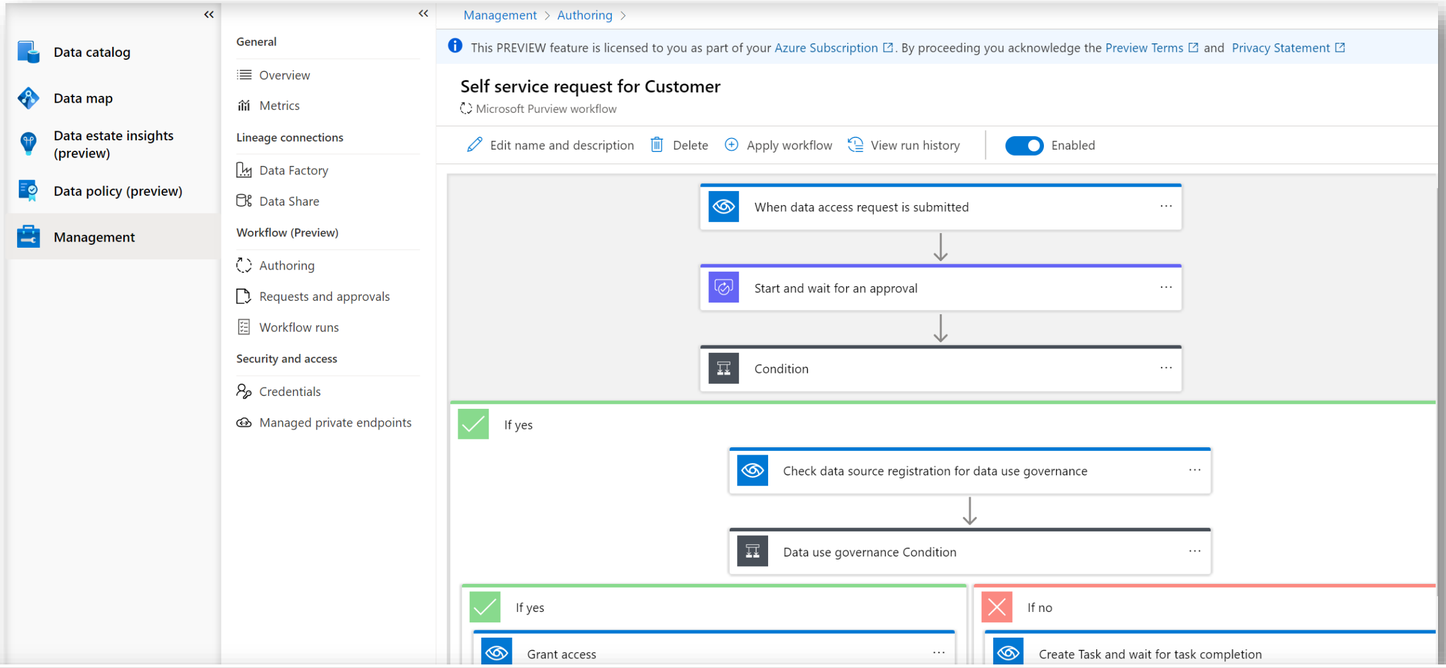


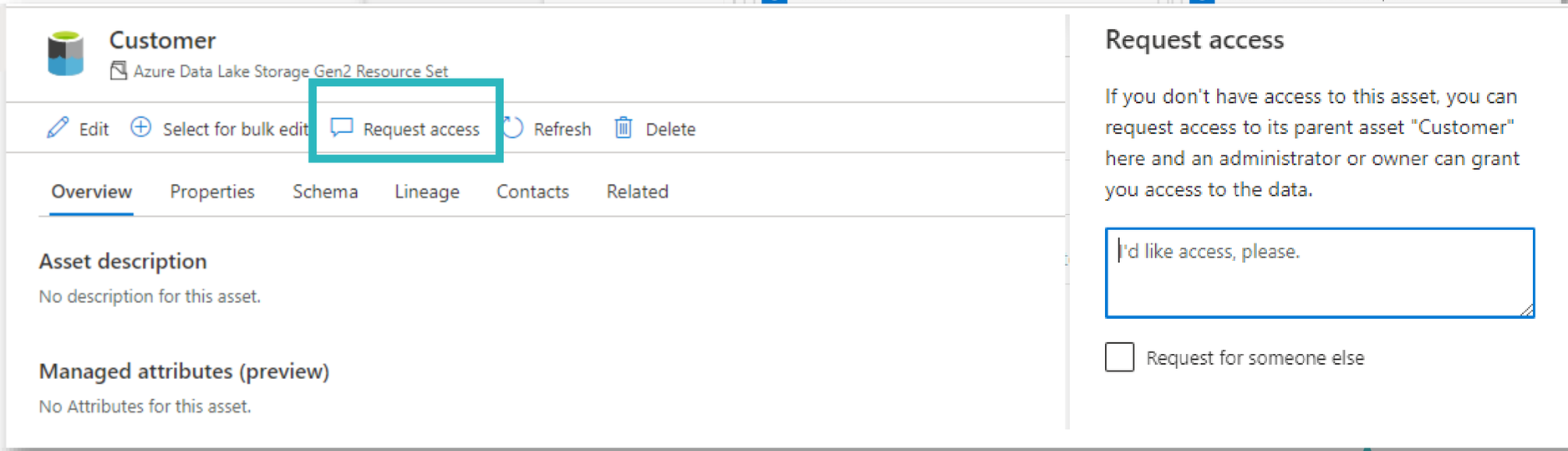
Data Use Management

Self Service

- Workflow Governance

  Data Access





INSPARK

Data Use Management

Self Service

- Workflow Governance

  Data Access

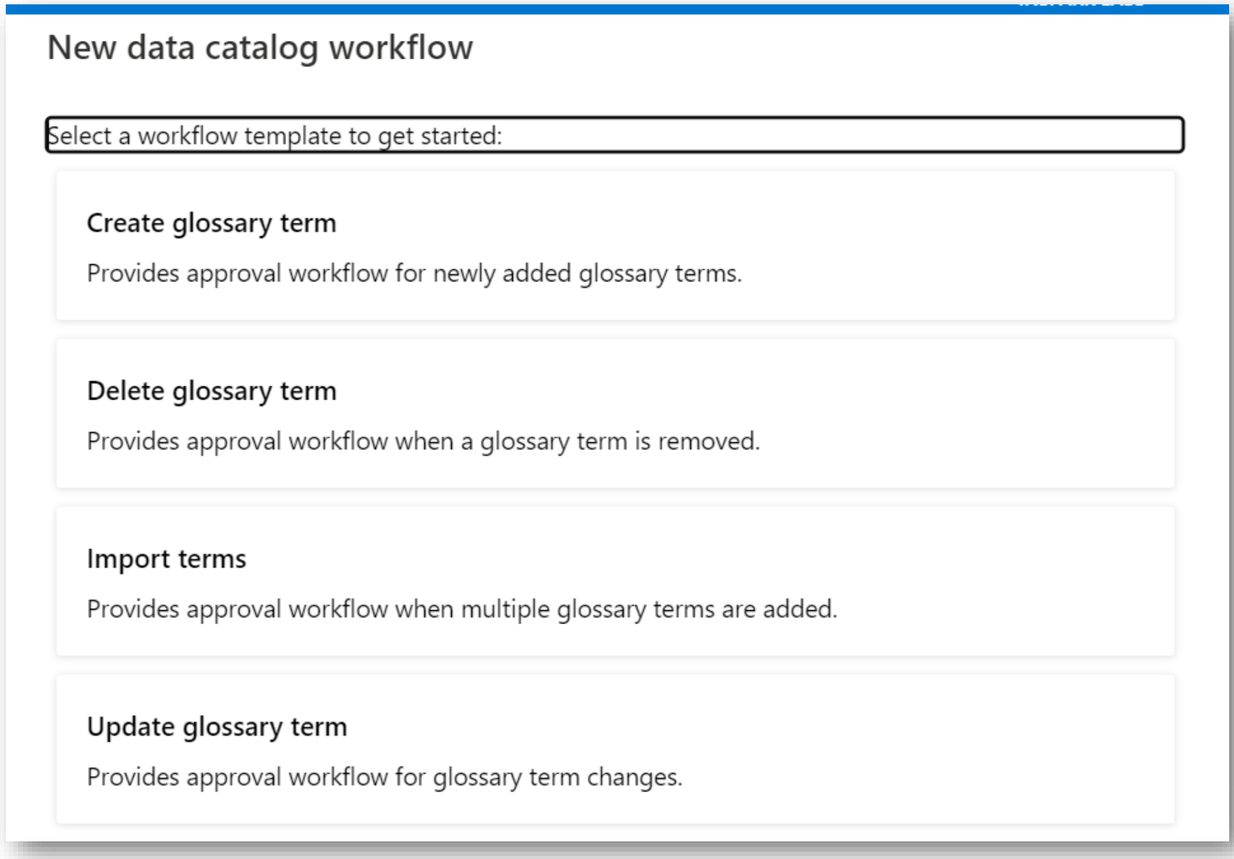
- Workflow Data Catalog

  Create Glossary Terms

  Delete Glossary Terms

  Import Terms

  Update Glossary Terms



INSPARK

Microsoft Purview

Generally Available

Preview

Unified Data Governance

On-prem

Cloud

SaaS Applications

Microsoft Purview

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

Open APIs (Apache Atlas 2.2)

Scanning and Classifications

Azure Synapse Analytics

SQL Server

Power BI

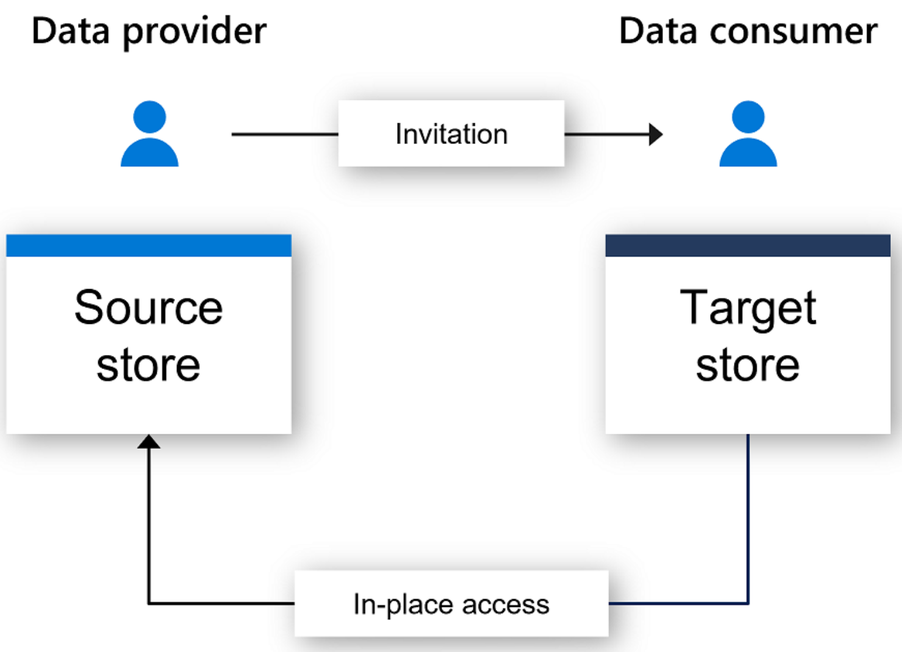
Azure SQL

Microsoft 365

INSPARK

Data Sharing

Data Share



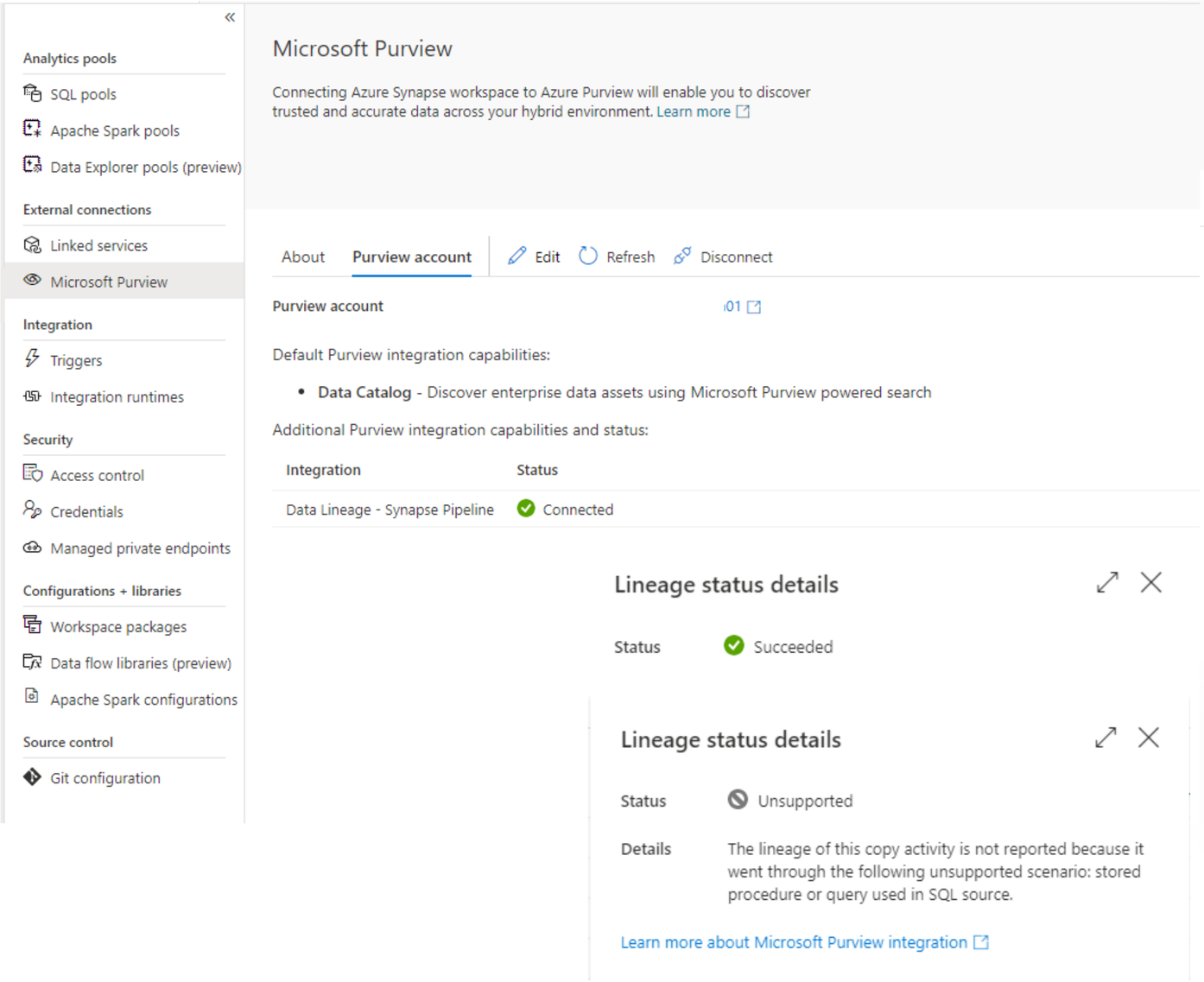
- Azure Storage
- Azure Data Lake Storage Gen 2
- Not available yet in Europe
- Sharing data in near real time
- Share data in the same Azure Tenant or across Azure Tenants
- Sharing is only supported for storage accounts with the following redundancy: LRS, GRS, RA-GRS.

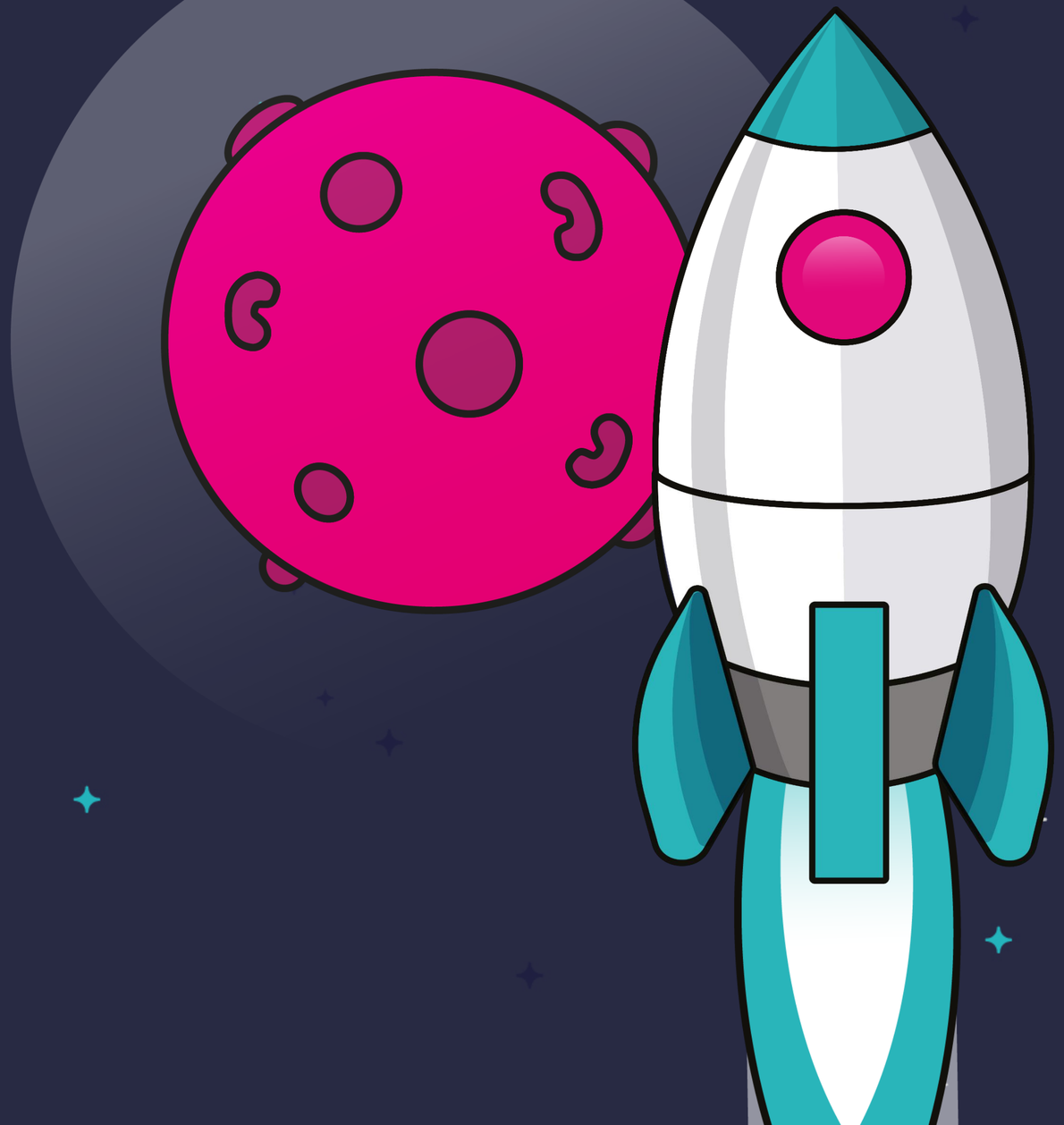
Microsoft Purview

Azure Synapse Analytics

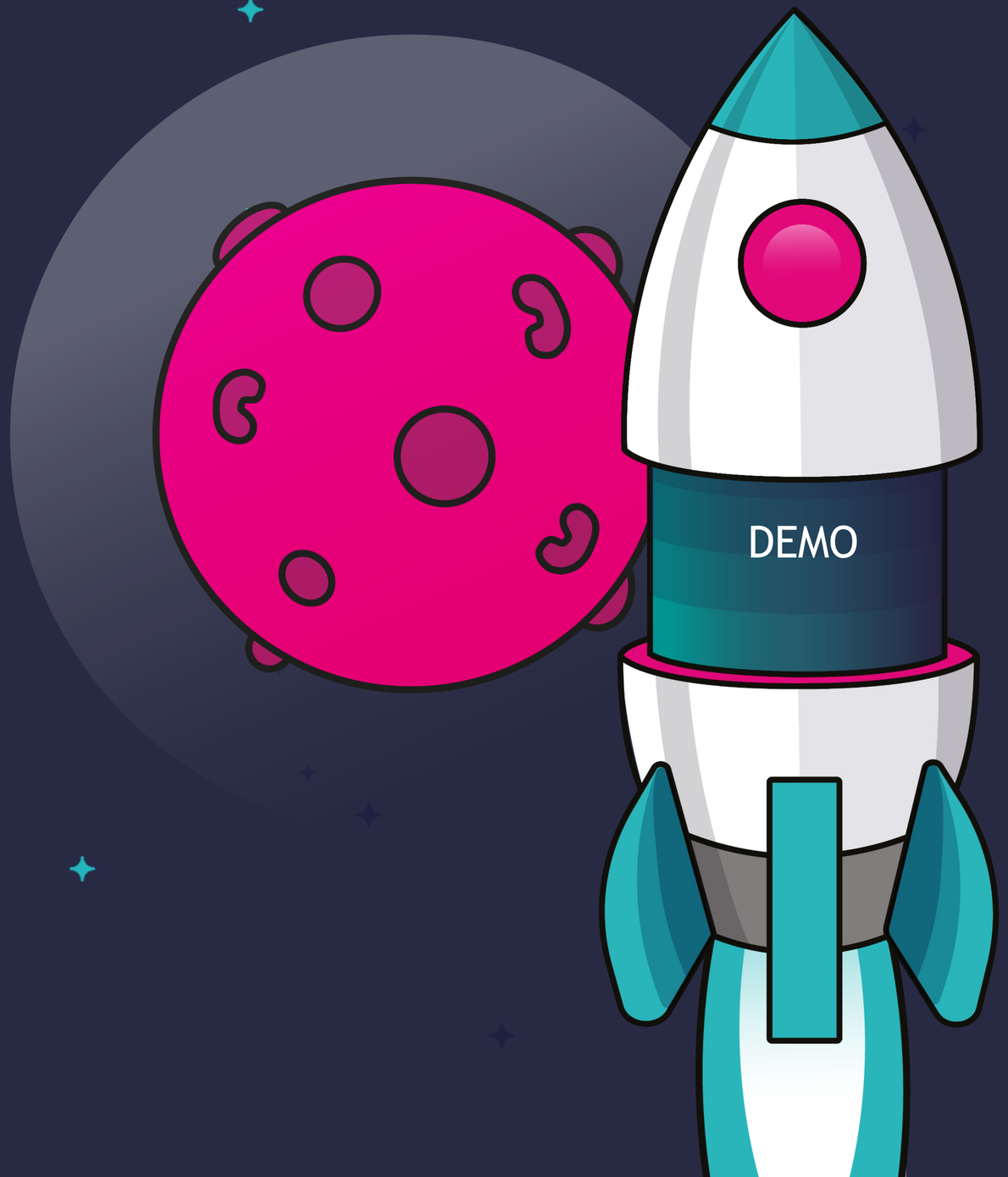
Integration

- Register Synapse Workspace with Purview
- Search datasets in Synapse workspace from Linked Purview Data Catalog
- View enhanced metadata, lineage from Purview for Synapse Tables and annotations

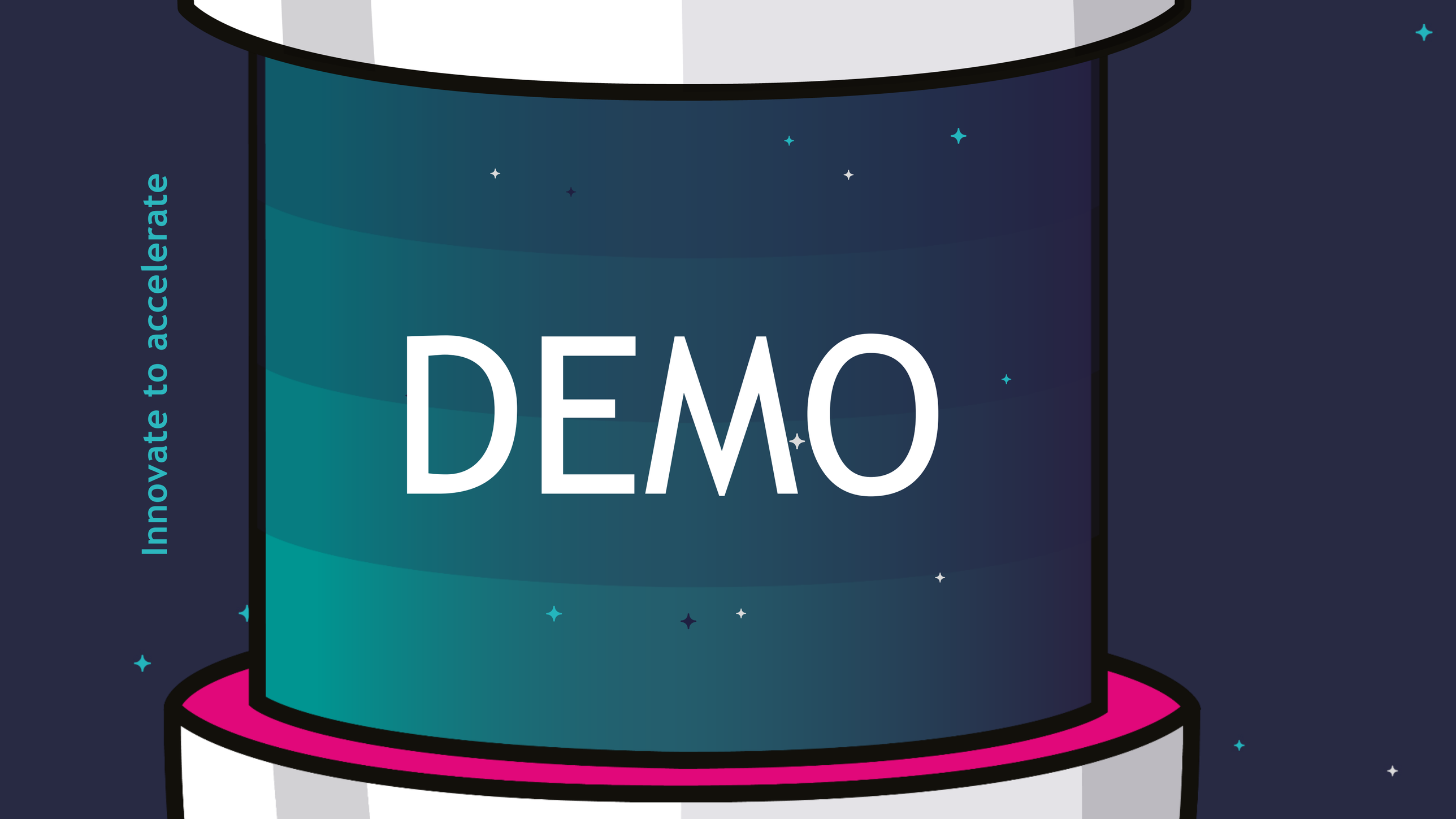




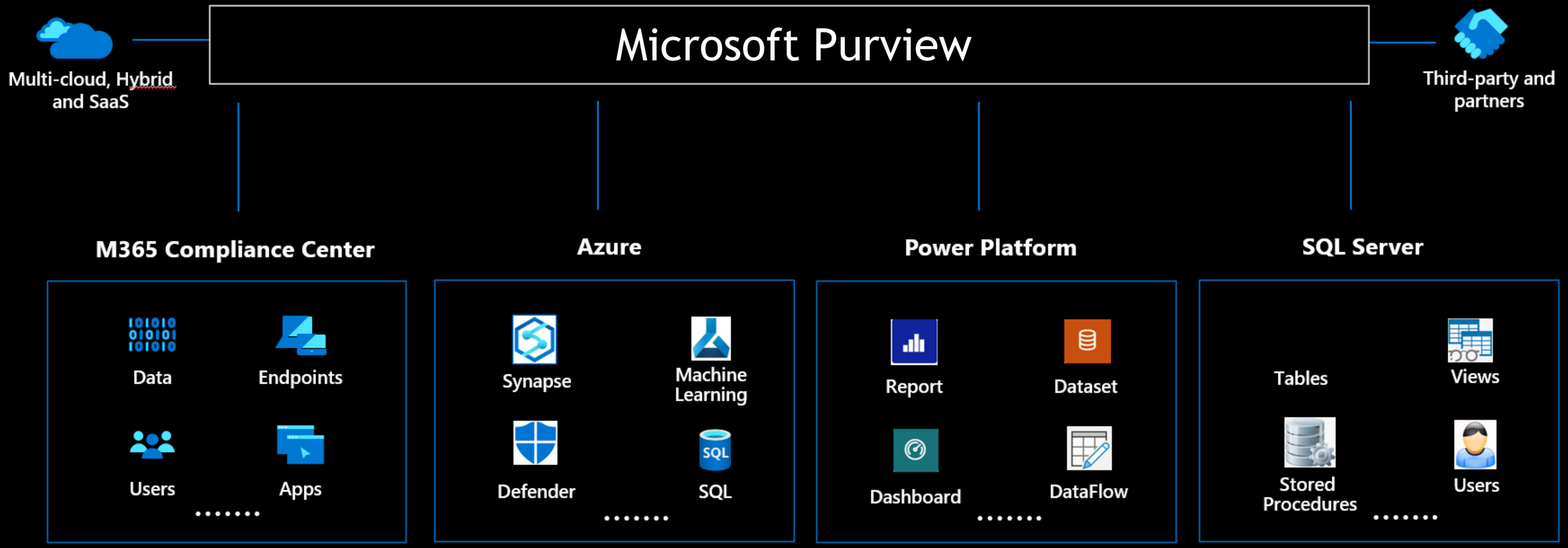
INSPARK



INSPARK


Innovate to accelerate
DEMO

Data Governance Opportunity



Questions?

@erwindekreuk

https://erwindekreuk.com

INSPARK