

Solve your Data Governance challenges with Microsoft Purview



Erwin de Kreuk

Principal Consultant - Lead Data & AI

InSpark



INSPARK

A business card for Erwin de Kreuk, featuring a circular profile picture, his name, title, and a QR code. The card is set against a white background with a black border and is placed on a pink and white background.

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

A QR code located on the business card.

Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark



@ErwinDeKreuk



linkedin.com/in/ErwinDeKreuk



ErwinDeKreuk.com



github.com/edkreuk



We Are **InSpark**

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



twoday
kapacity



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

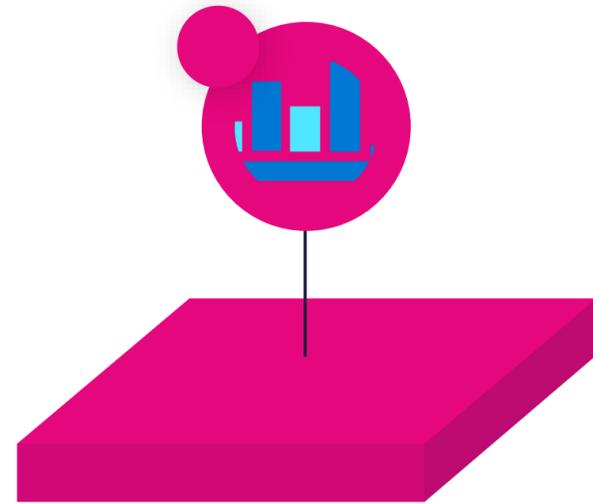
175 ZB

Expected global volume of data generated annually by 2025

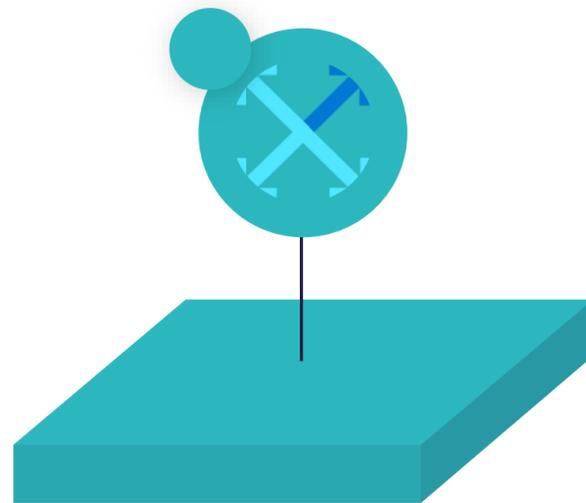
IDC

IDC Data Age 2025, Dave Reinsel,

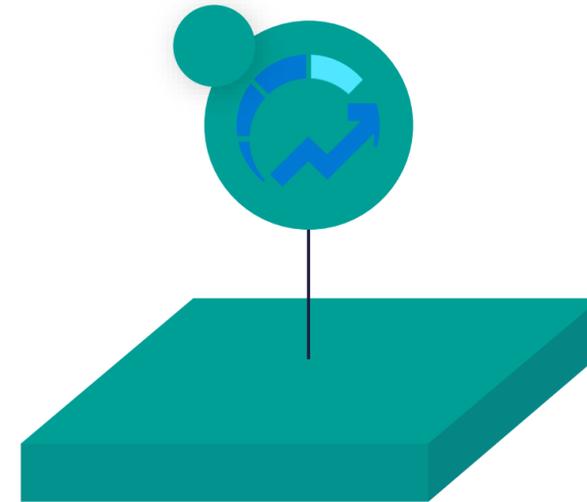
Elements of successful data governance



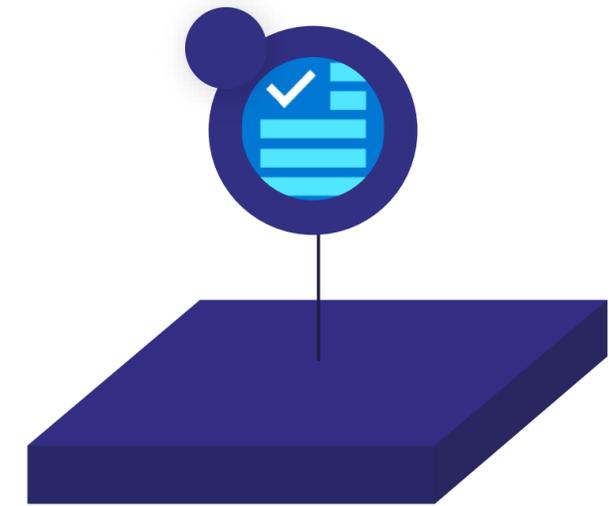
Manage growing data landscape



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.



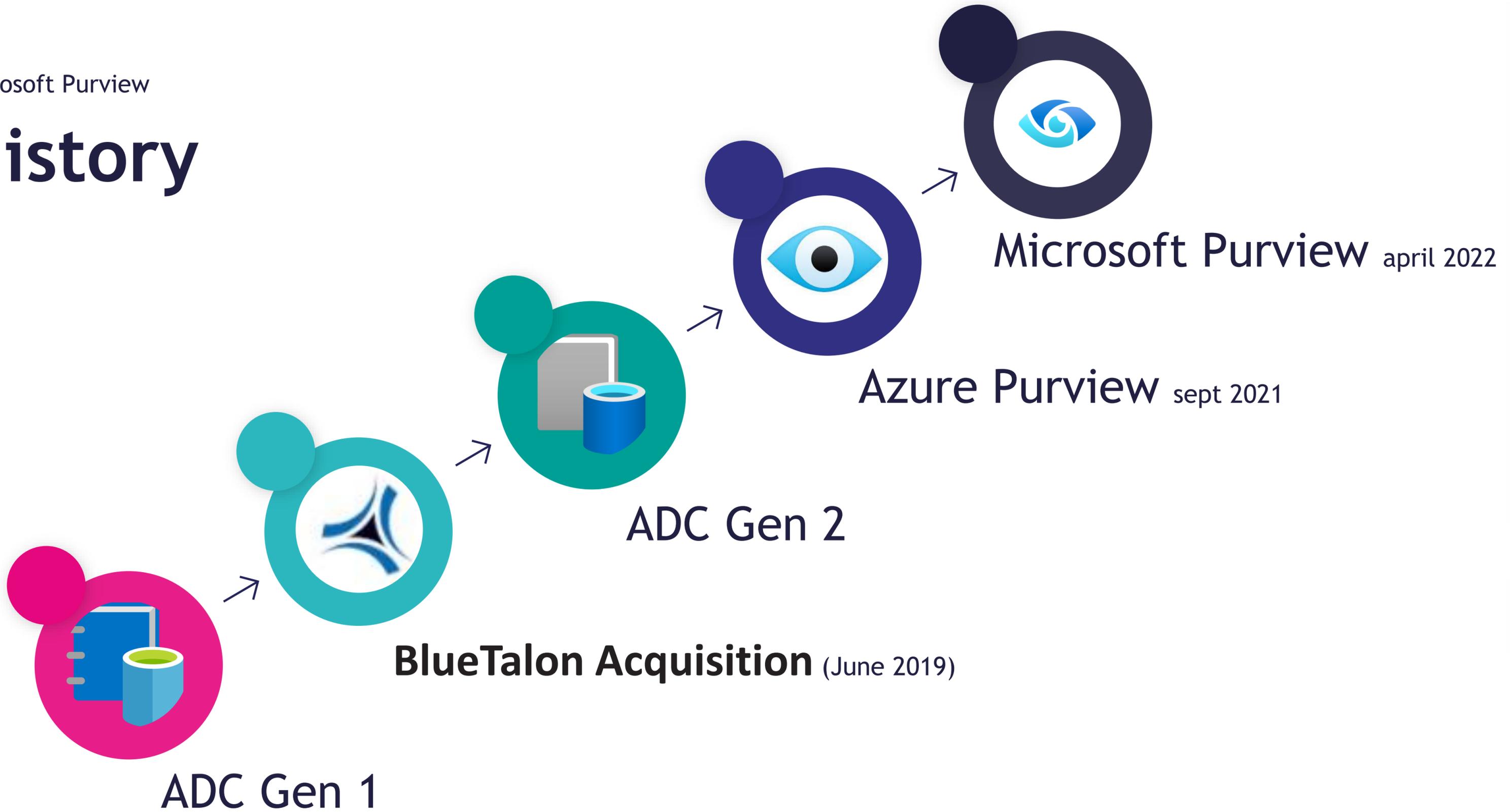
Unified **data governance** to
maximize the business value of data

Microsoft Purview



Microsoft Purview

History



BlueTalon Acquisition (June 2019)

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

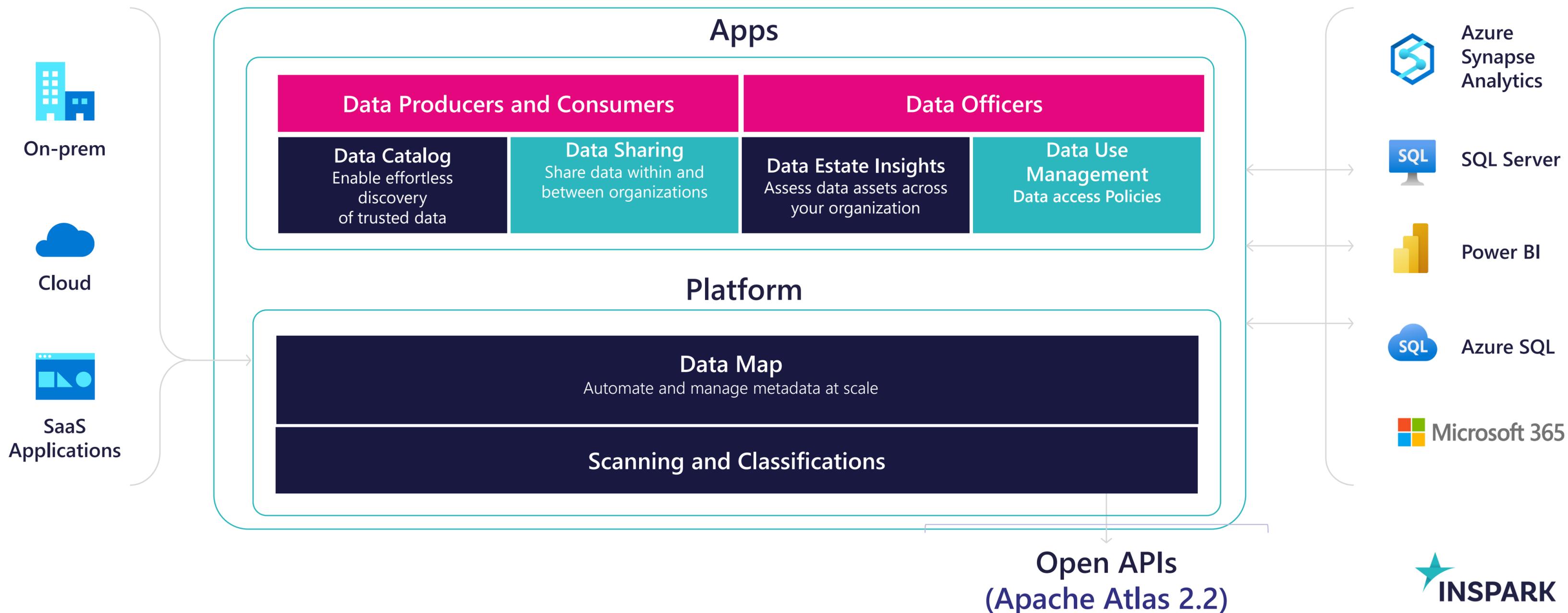
For data consumers, data engineers, data officers

Microsoft Purview

Generally Available

Preview

Unified Data Governance



Purview Studio

Updates Accounts Notifications

The screenshot shows the Microsoft Purview InSpark interface. At the top, there is a navigation bar with the Microsoft Purview logo and the user's name, Erwin.de.Kreuk@inspark... The main content area displays the title 'Microsoft Purview InSpark' and statistics: 18 sources, 1,819 assets, and 56 glossary terms. Below this is a search bar labeled 'Search catalog'. Three main activity cards are visible: 'Browse assets', 'Manage glossary', and 'Knowledge center'. At the bottom, there is a 'Recently accessed' table and a 'Links' section.

Name	Last update
Product	16 days ago
Product	3 days ago
Copy_DeltaLake_Purview	2 days ago
Address	16 days ago
PL_EXECUTE_COPY_ADLS_TO_DELTALAKE_PURVIEWy2	---
Copy_DeltaLake_Purview	---

Links

- Microsoft Purview overview
- Getting started
- Documentation

Metrics

Search Bar

Recently Accessed Entities

Feedback

Key Activities

Usefull Links



Purview Studio - Activity hubs

-  **Data catalog**
-  **Data map**
-  **Data share (preview)**
-  **Data estate insights**
-  **Data policy (preview)**
-  **Management**

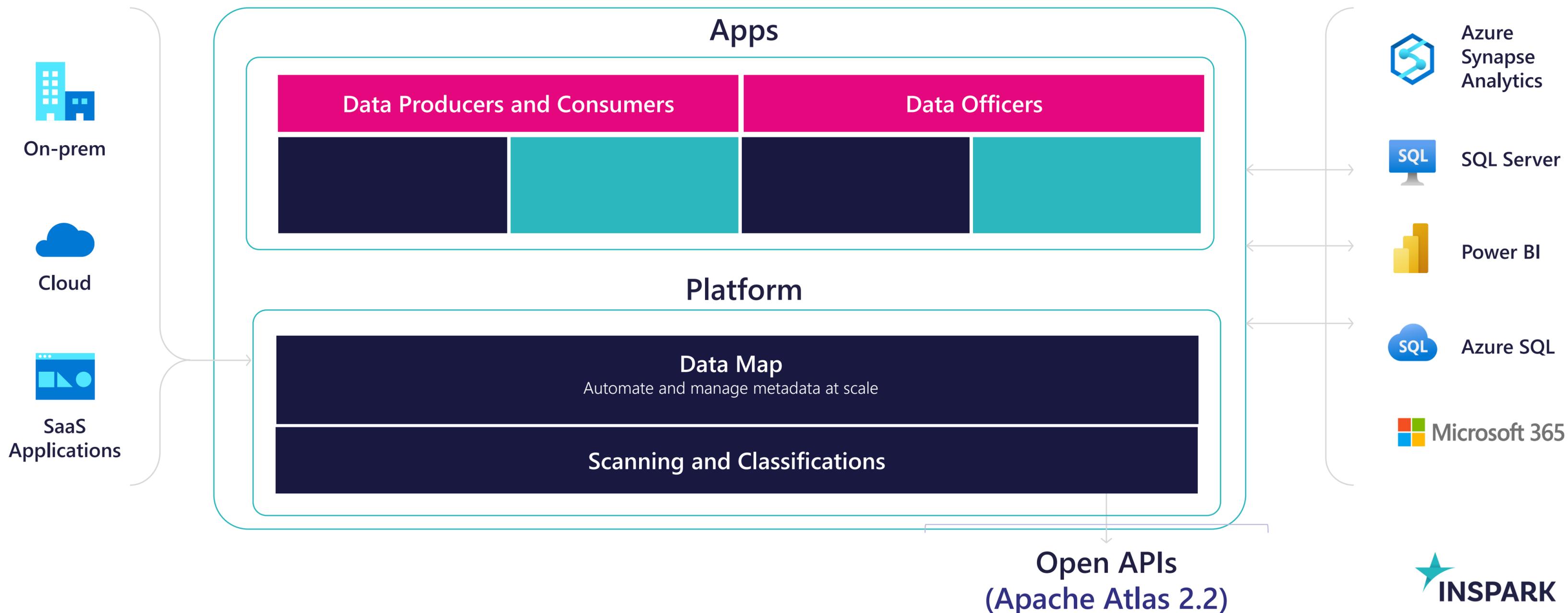
- Quick Actions, recently accessed items, owned Items, search bar and Documentation.
- Manage Glossary Items, search, manage terms templates and custom attributes, import/export Terms using csv.
- Create collections, register data sources, setup Scans, Integration runtime.
- Easily share data between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

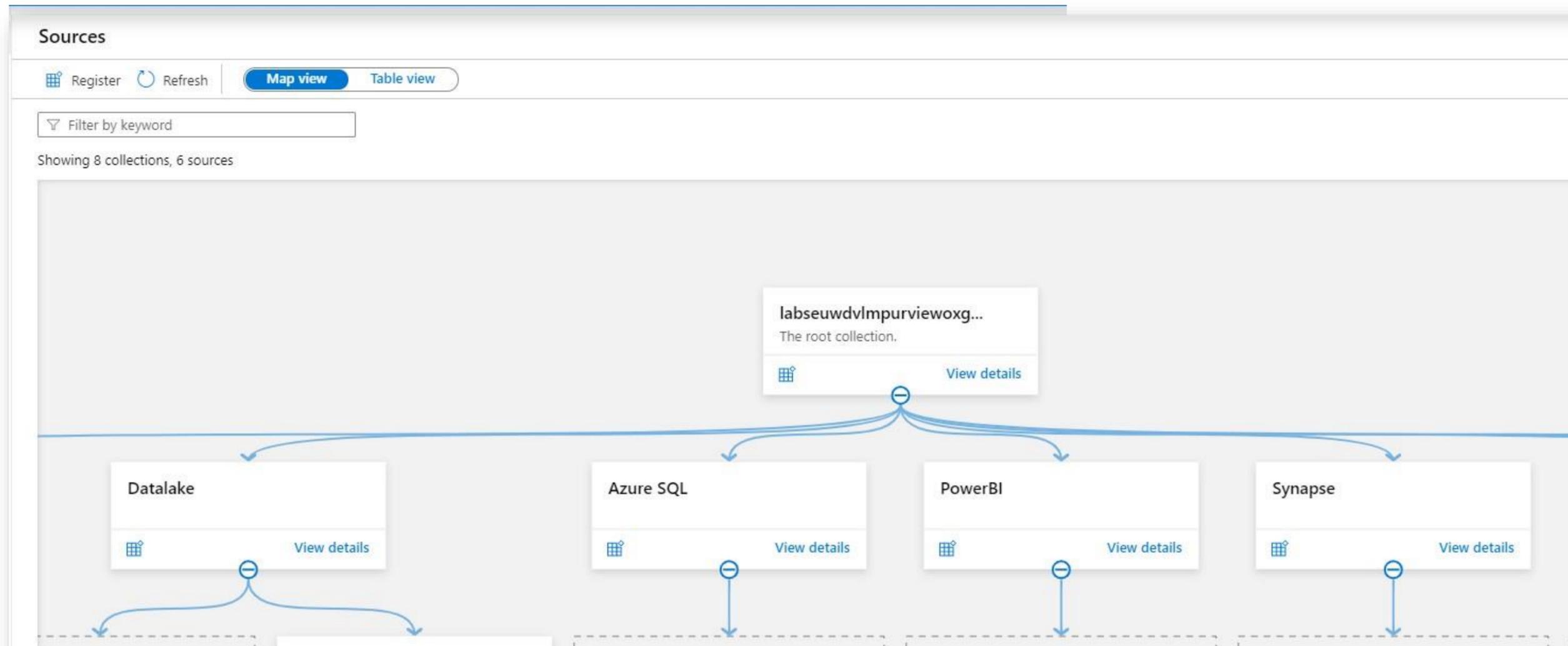
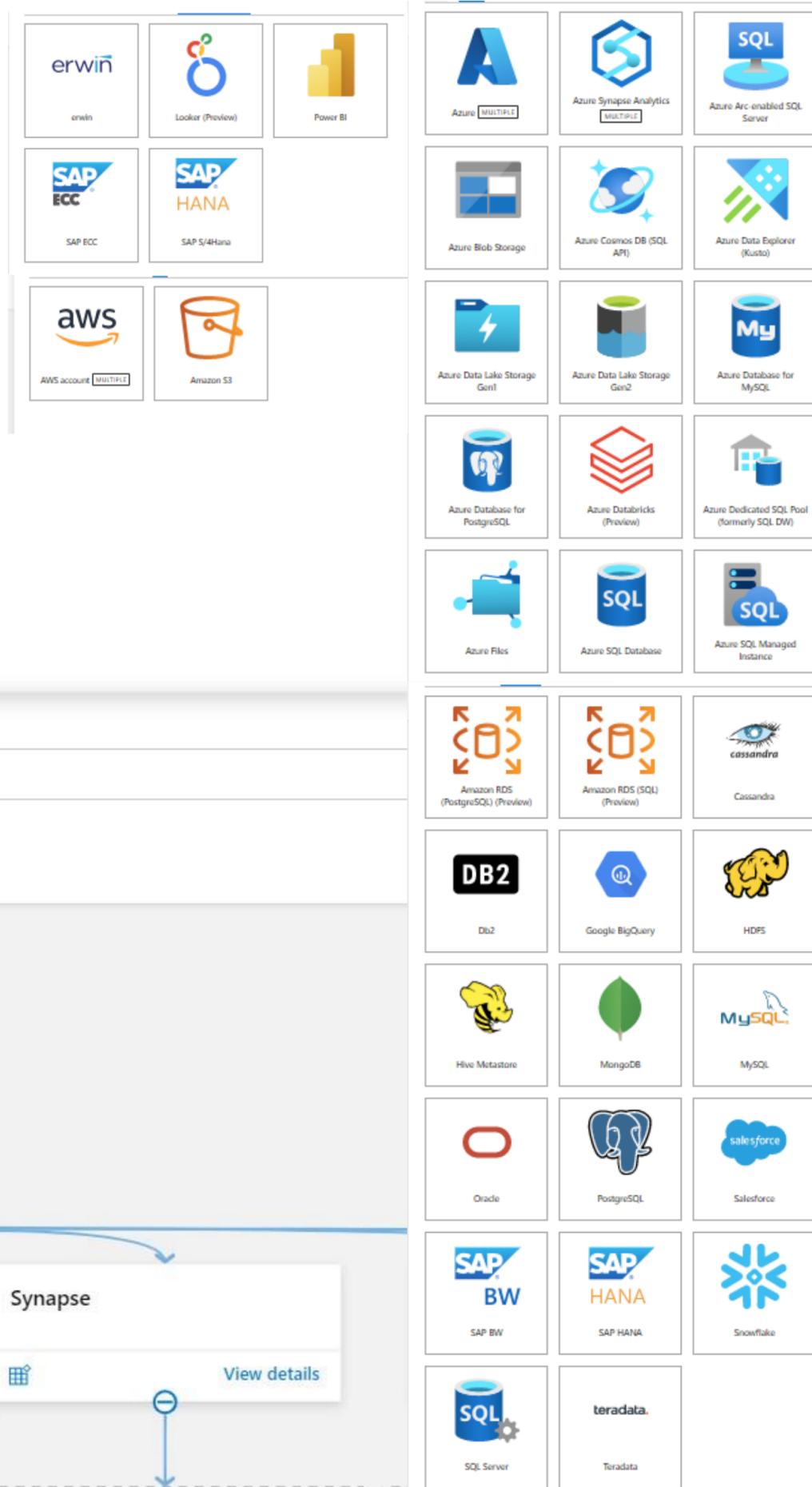
Unified Data Governance



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

Sources > LS_ADLS > Weekly_Scan run history

Weekly_Scan run history

▶ Run scan now ▾ ✎ Edit scan 🗑 Delete scan 🔄 Refresh

Status	Schedule	Assets classified	Assets discovered	Start time	End time
✔ Completed	🕒 Scheduled	2	432,119	01/18/21 11:00 AM	01/18/21
✔ Completed	🕒 Scheduled	172	432,117	01/11/21 11:00 AM	01/11/21
✔ Completed	🕒 Scheduled	199	431,156	01/04/21 11:00 AM	01/04/21
✔ Completed	👤 Manual	171	429,874	12/30/20 02:47 PM	12/30/20
✔ Completed	👤 Manual	192	429,496	12/29/20 02:57 PM	12/29/20

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

Classifications

+ New Edit Delete Refresh

System Custom

These are the system provided classifications.

Filter by name...

Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA F...
All Full Names	MICROSOFT.PERSONAL.NAME	All Fu...
All Physical Addresses	MICROSOFT.PERSONAL.PHYSICALADDRESS	All PH...
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGENTINA.DNI_NU...	Argen...
Australia Bank Account Number	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACCOU...	Austr...
Australia Business Number	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Austr...
Australia Company Number	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Austr...
Australia Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...	Austr...
Australia Medical Account Number	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL_...	Austr...
Australia Passport Number	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...	Austr...
Australia Tax File Number	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE_...	Austr...
Austria Driver's License Number	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
CityID			
CityName	World Cities		
LastEditedBy			
LatestRecordedPopulation			
Location			

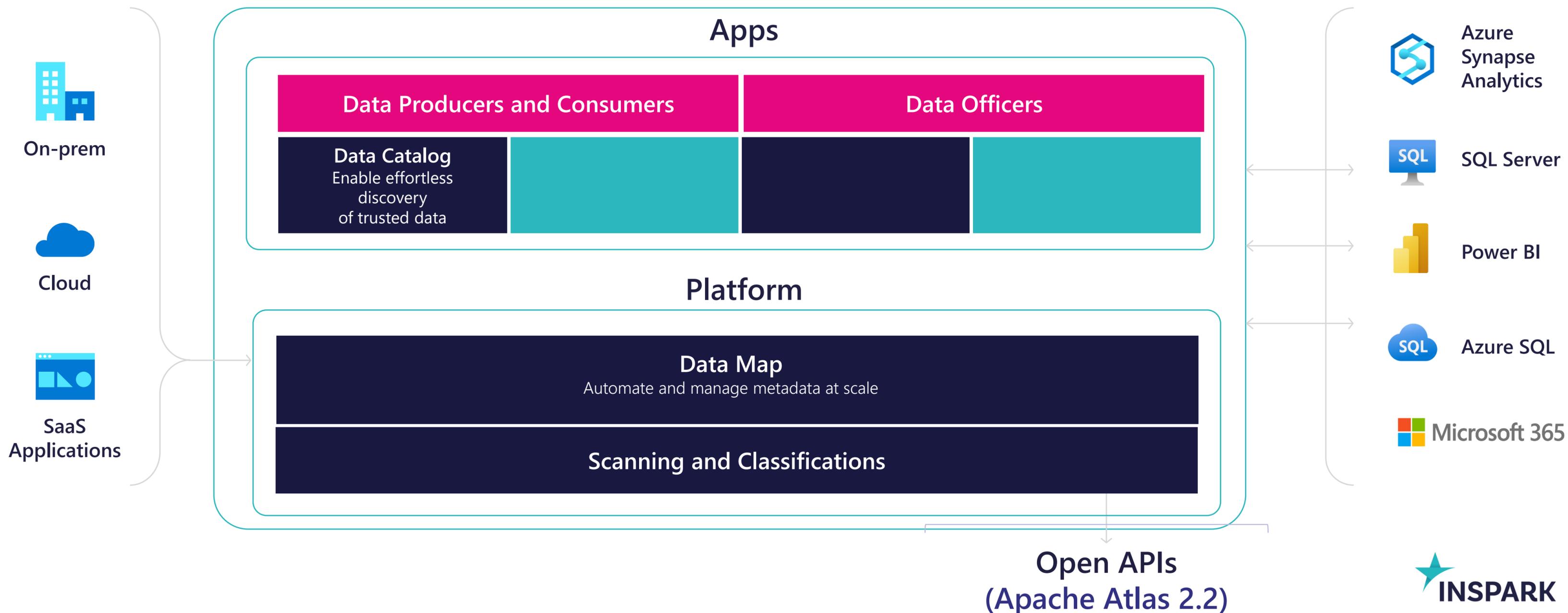
Microsoft Purview

Generally Available

Preview



Unified Data Governance



Data Catalog

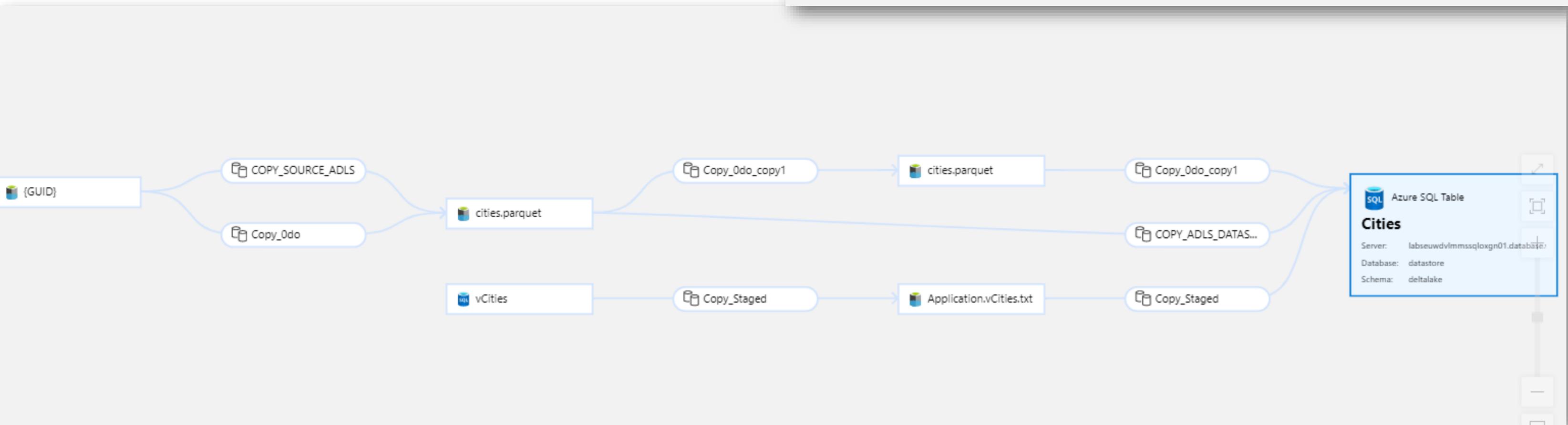
Search

- Semantic search and browse
- Business glossary and workflows

The screenshot displays the Microsoft Purview InSpark search interface. The main header reads "Microsoft Purview InSpark" with the identifier "(labseuwdvImpurviewoxgn01)". Below the header, it shows "17 sources", "1,950 assets", and "56 glossary terms". A search bar contains the text "customer". Below the search bar, there are sections for "Your recent searches" (listing "customer" and "customers"), "Search suggestions" (listing "CustomerCategories", "Customer", "CustomerKey", and "CustomerCategoryID"), and "Asset suggestions" (listing "Customer_Customer_DataFlow", "CustomerType_CustomerType_DataFlow", "Customer", and "Customer"). A "View search results" link is visible at the bottom right of the search results area. On the left side, a navigation menu includes "Data catalog", "Data map", "Data estate insights (preview)", "Data policy (preview)", and "Management". Below the search bar, a "Recently accessed" section lists various data assets such as "SalesLT_Customer (Current)", "SalesLT_Address", "PurchaseOrderLines_Purchase", "Application_Countries (Current)", "Current", and "Application_Cities".

Lineage

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



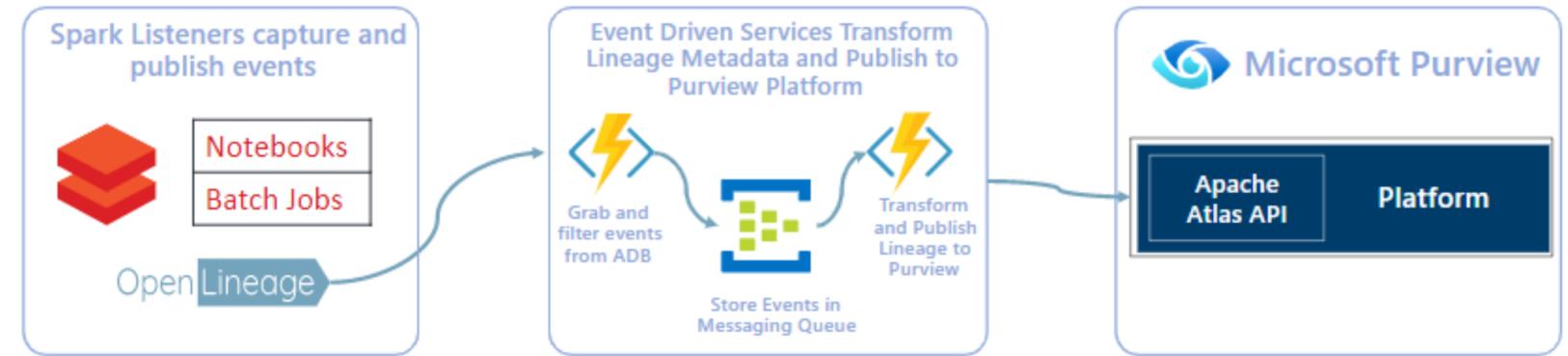
Manual Lineage

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

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



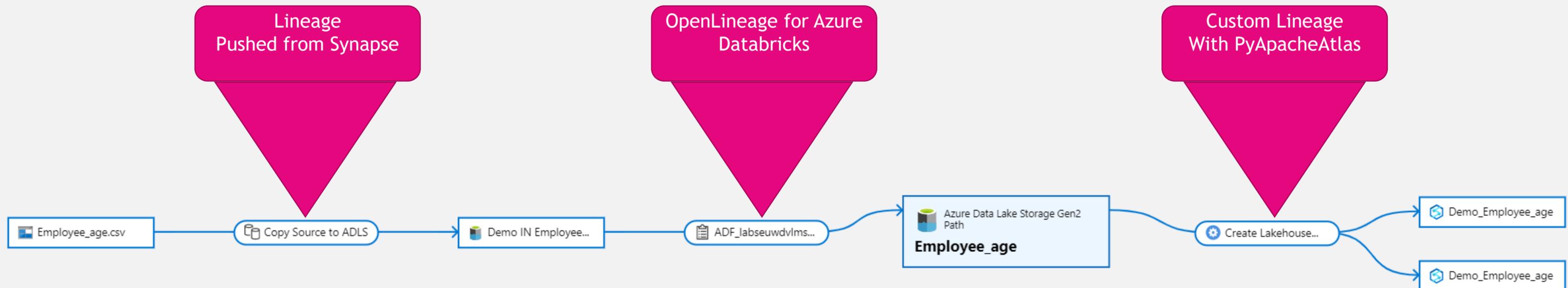
- OpenLineagefor Azure DataBricks

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

[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>

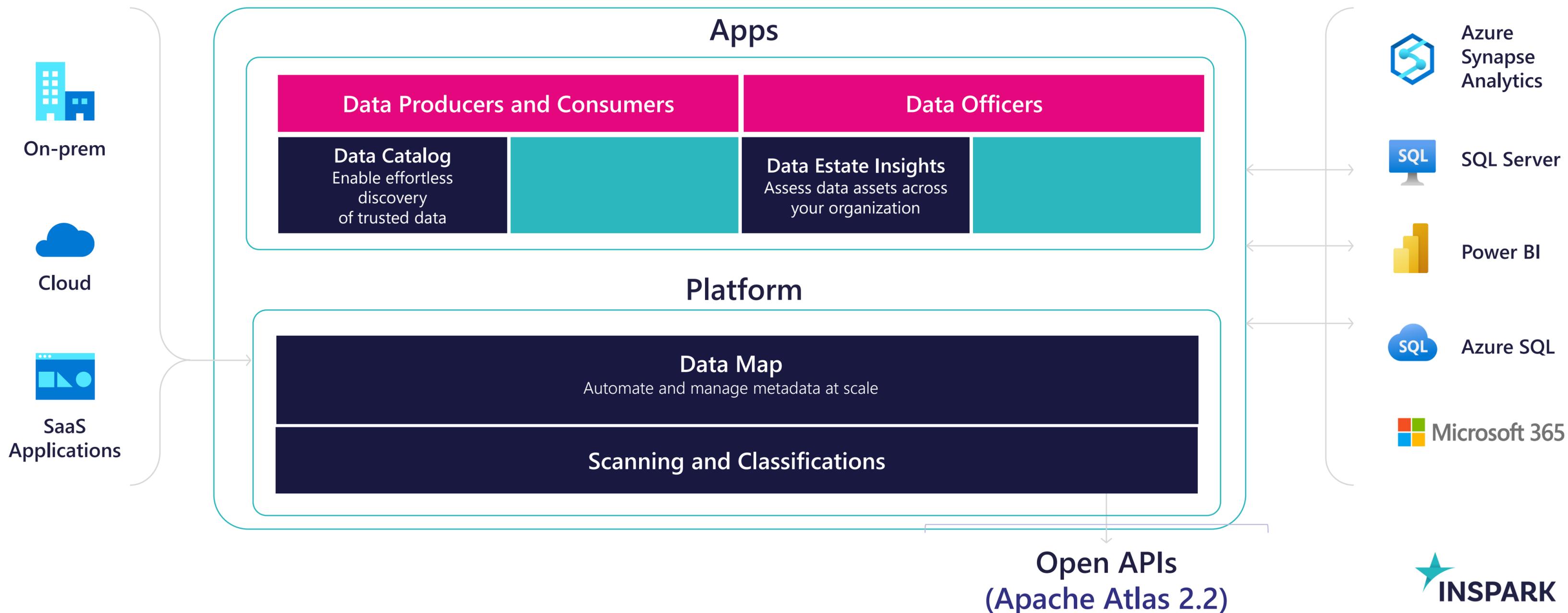


Microsoft Purview

Generally Available

Preview

Unified Data Governance



Data Estate Insight

Metrics

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

Data Estate Insights
 On

Automatic

7/18/2022, 9:48 PM
Provides insight into your data estate.

Health

- Data stewardship
- Inventory and ownership
- Assets
- Curation and governance
- Glossary
- Classifications
- Sensitivity labels

Data stewardship

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

Asset curation

29K
Total assets

Asset data ownership

29K
No owner

Catalog usage and adoption

Monthly active users

0

↓ 100% in last month

Show: Data estate Catalog adoption

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

Assets

View key metrics about your data estate's assets.

Total assets

29K
Microsoft Purview InSpark

Asset classification

29K
No classification

Asset data ownership

29K
No owner

New assets (Last 30 days)

81
New

Deleted (Last 30 days)

2

ACCESS CONTROL

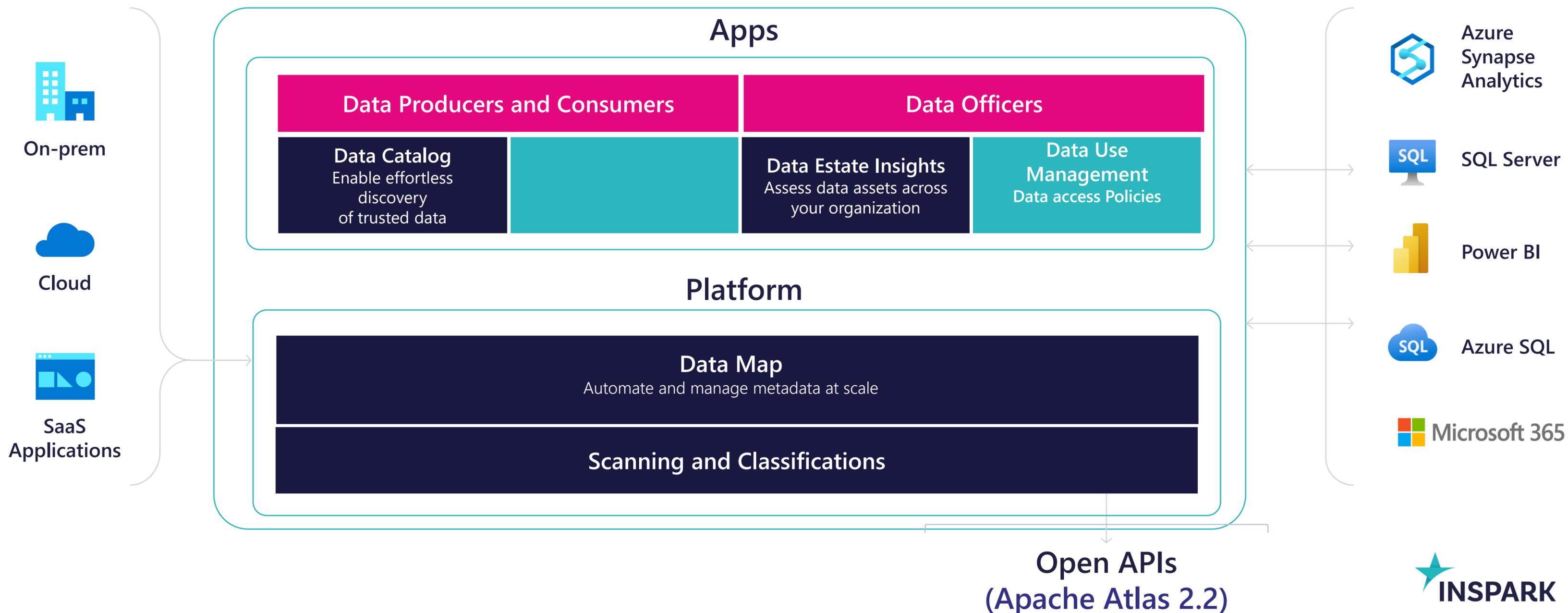
Data Owner
Dev Ops
Self Service access

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

The screenshot displays the 'DevOps roles' configuration page in SQL Server Enterprise Manager. It is divided into two main sections: 'SQL Performance monitor' and 'SQL Security auditor'. The 'SQL Performance monitor' section is currently selected and shows a description: 'SQL Performance Monitor allows read access to performance related system views'. Below this description is a blue button labeled 'Add/remove subjects'. Underneath the button is a section titled 'Subjects' with the text 'Selected users, groups, and applications will show up here.' To the right of the main configuration area is a tree view showing the database structure for 'WideWorldImportersDW'. The tree view is expanded to show 'System Views', which includes a list of system views such as 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, and INFORMATION_SCHEMA.VIEWS.

Self-service access policies

Data Access (Governance)

- Azure SQL Database
- Azure Storage
- Expiration Functionality

Data access request
Microsoft Purview workflow

Edit name and description

Start and wait for an approval

Approval type *
Approve/Reject - Everyone must approve

Title *
Approval Request for Data Access Request

Assigned to *
Search by name or email address

Reminder settings
 On

Reminder interval *
1 day

Expiry settings
 On

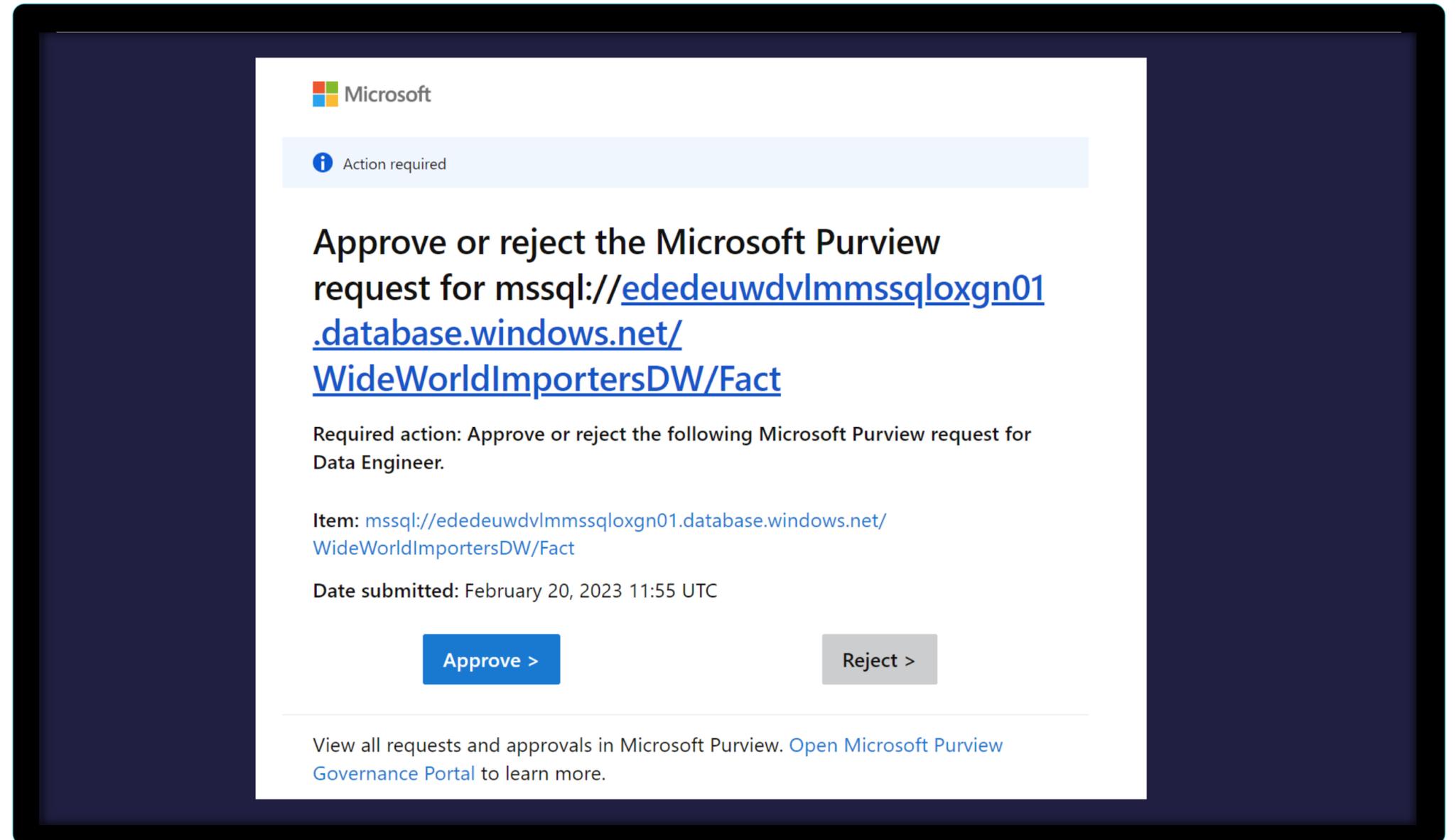
Expire After *
1 month

Notify on expiration
Search by name or email address

Self-service access policies

Data Access (Governance)

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



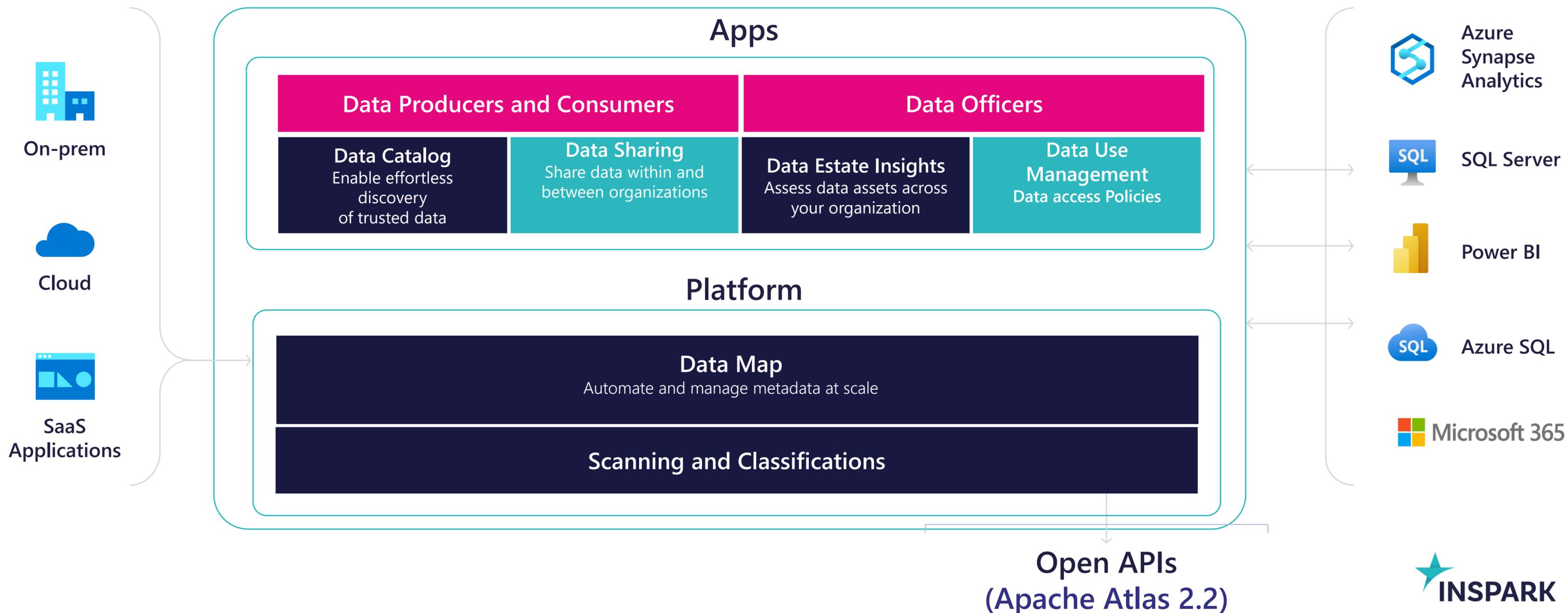
The screenshot shows a Microsoft notification card with a white background and a dark blue border. At the top left is the Microsoft logo. Below it is a light blue bar with an information icon and the text "Action required". The main text reads: "Approve or reject the Microsoft Purview request for mssql://[ededeuwdvImmsqlxgn01.database.windows.net/WideWorldImportersDW/Fact](#)". Below this, it says "Required action: Approve or reject the following Microsoft Purview request for Data Engineer." The "Item" is listed as "mssql://ededeuwdvImmsqlxgn01.database.windows.net/WideWorldImportersDW/Fact" and the "Date submitted" is "February 20, 2023 11:55 UTC". At the bottom are two buttons: "Approve >" (blue) and "Reject >" (grey). A footer link says "View all requests and approvals in Microsoft Purview. [Open Microsoft Purview Governance Portal](#) to learn more."

Microsoft Purview

Generally Available

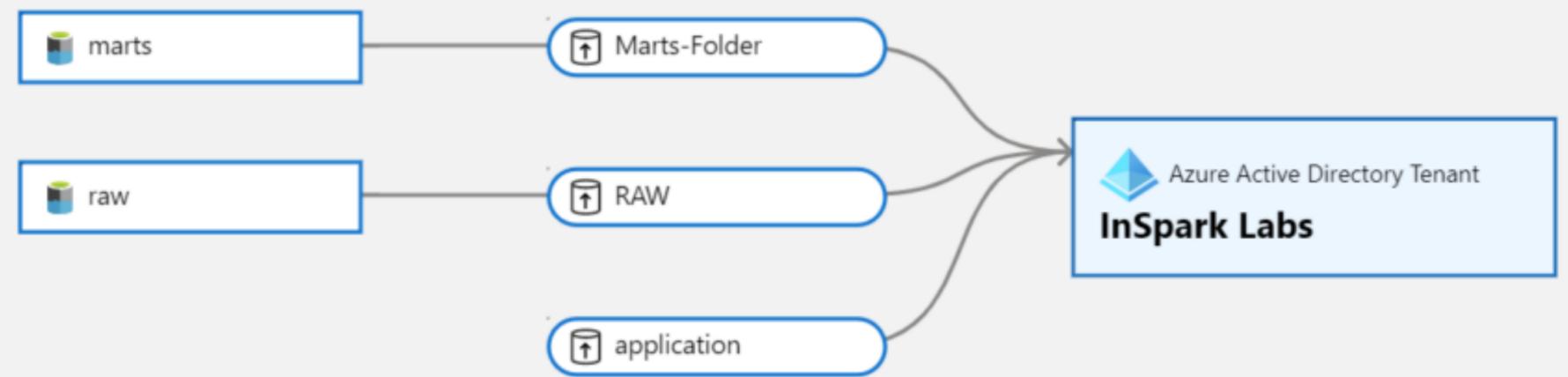
Preview

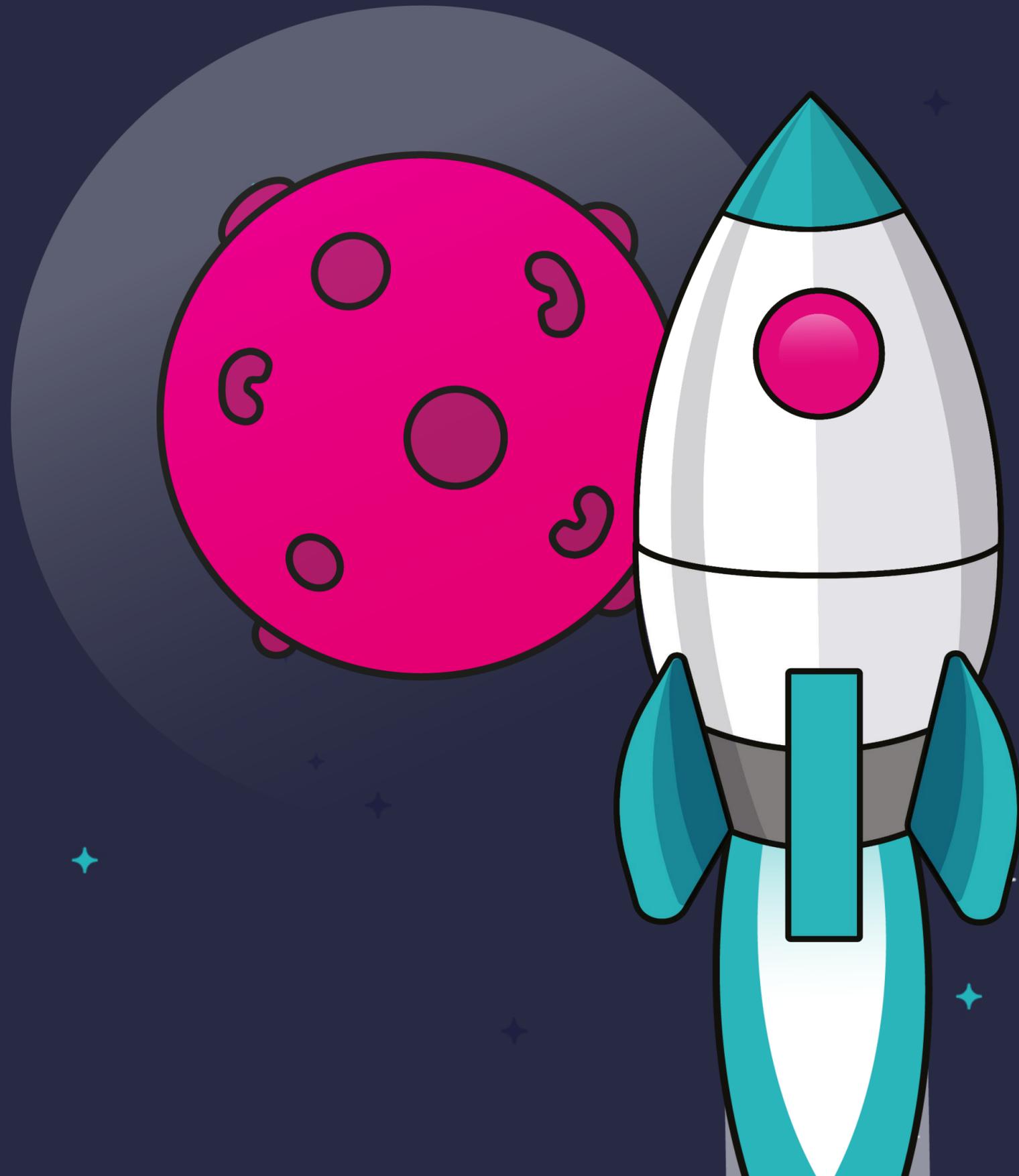
Unified Data Governance

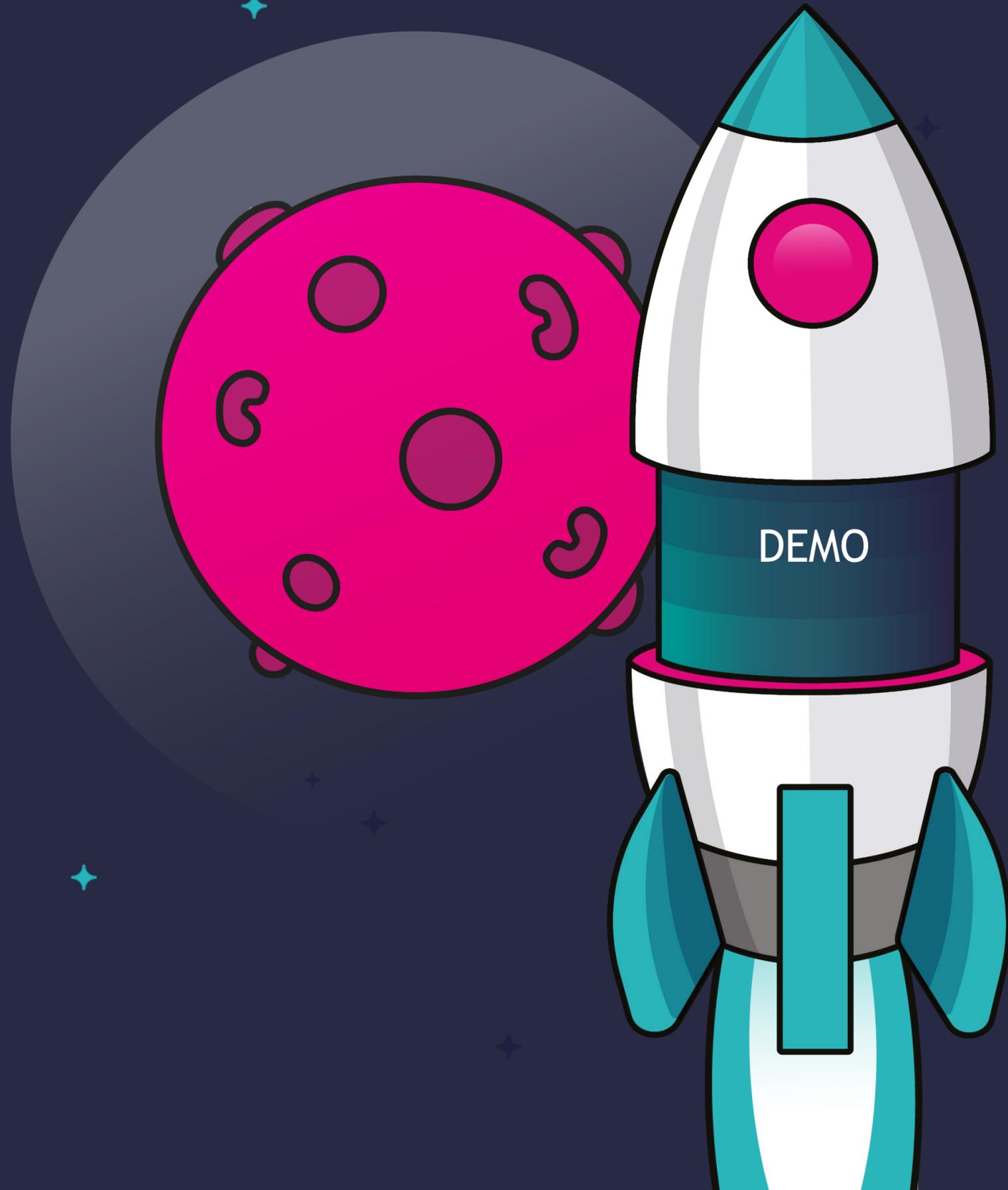


Sharing

- 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 piece of paper. The word "Pricing" is written on the paper in a stylized, hand-drawn font. The letter "P" is blue with diagonal hatching, while the rest of the word "ricing" is black with diagonal hatching. The background is a light gray brick wall.

Pricing

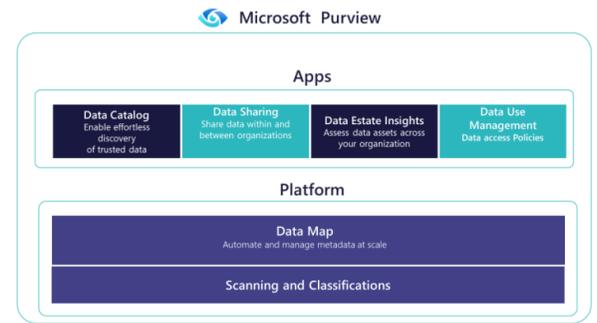
Data Map Consumption

- Elastic Data Map

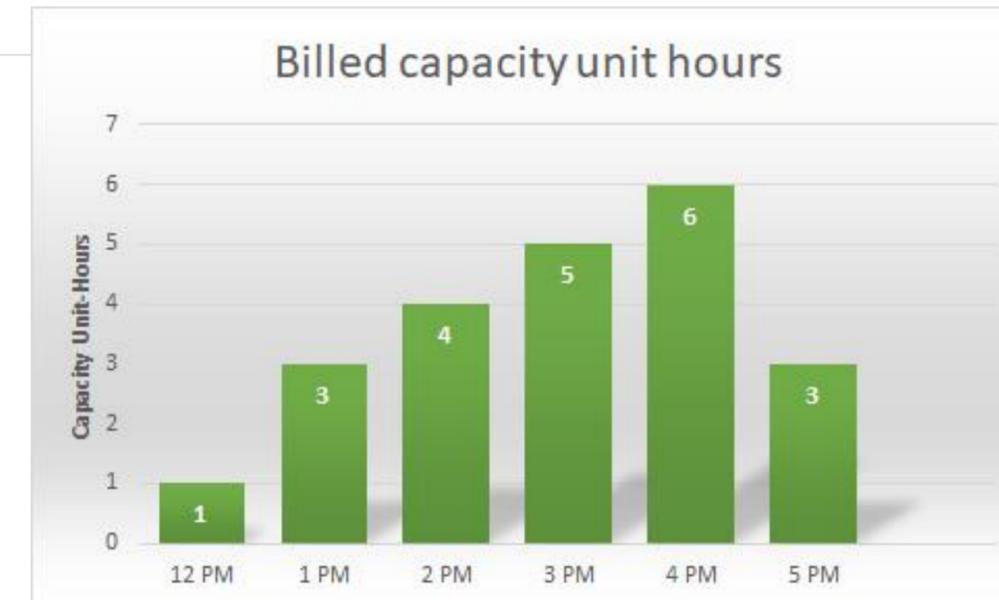
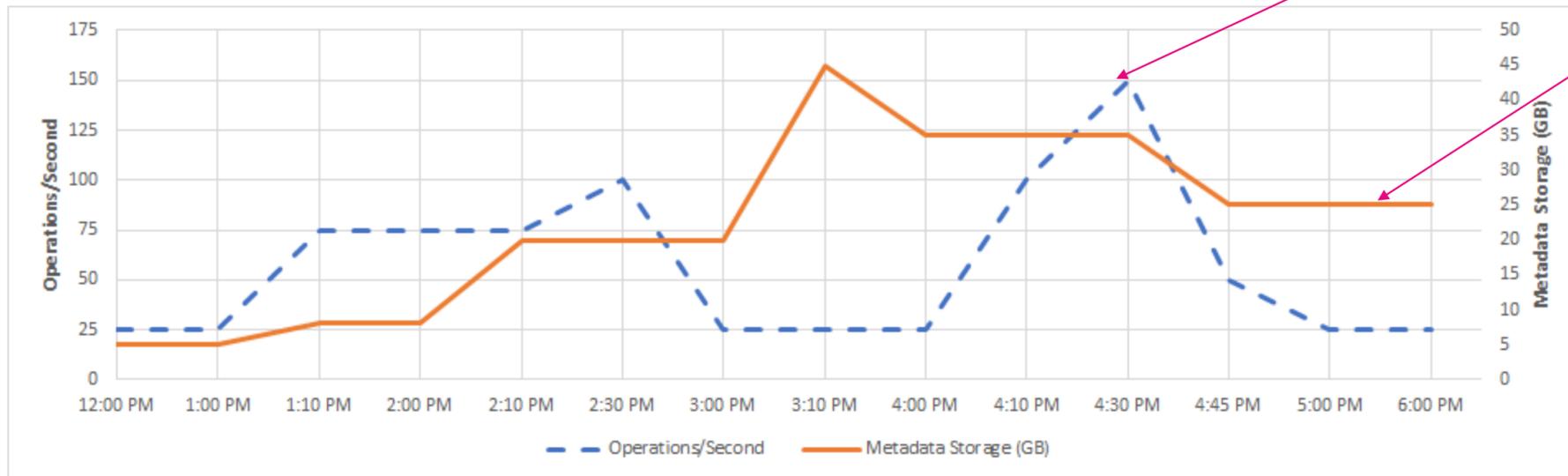
€0.373 per 1 Capacity Unit Hour

CU=Capacity Unit

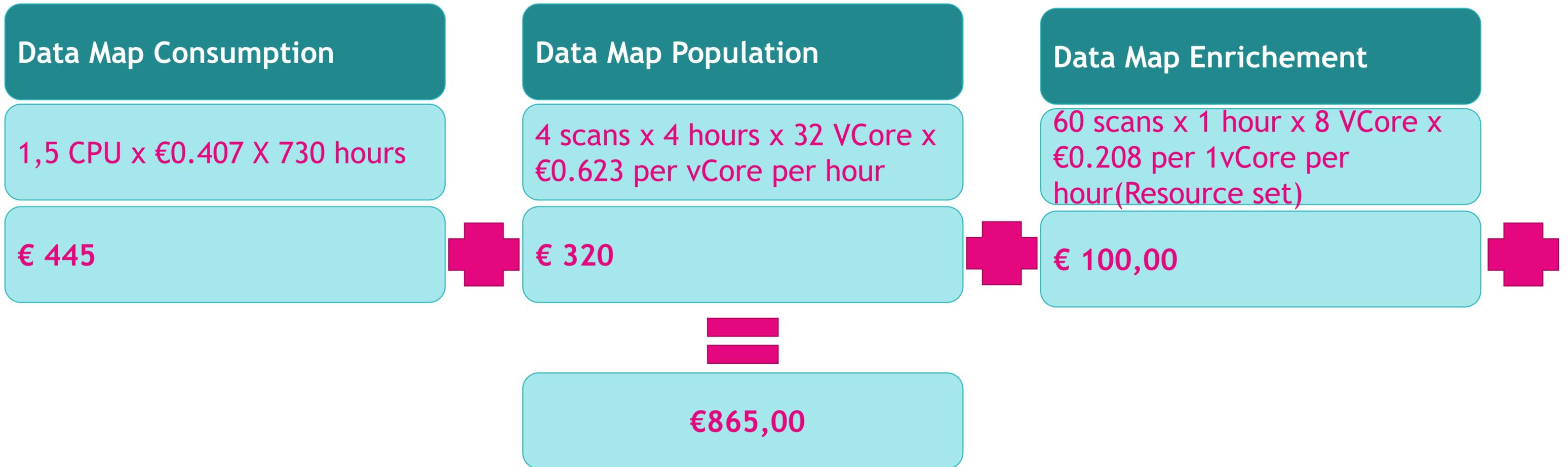
1 Capacity Unit supports requests of up to 25 data map operations per second and includes storage of up to 10 GB of metadata about data assets.



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



ErwinDeKreuk.com



github.com/edkreuk

