

DATA:Scotland

We are Scotland's Data Community Conference



Data Governance With Microsoft Purview



Erwin de Kreuk

We Are InSpark

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



DATA:Scotland 2022

is proudly supported by



THE
DATA LAB
value from data



**ADVANCING
ANALYTICS**



Microsoft Purview

Topics

Introduction to Purview

Data Map

Data Catalog

Data Estate Insights



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*

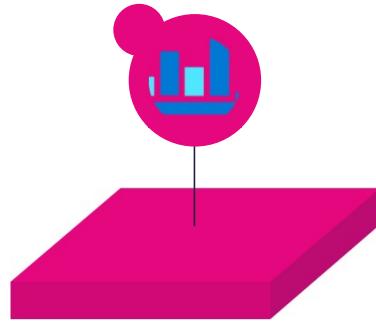
175 ZB

Expected global volume
of data generated
annually by 2025

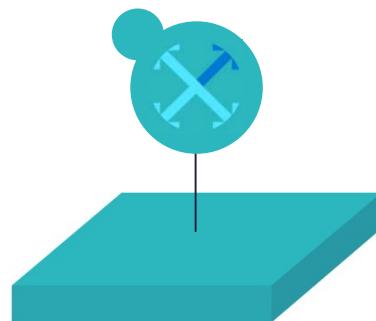
IDC

IDC Data Age 2025, Dave Reinsel,

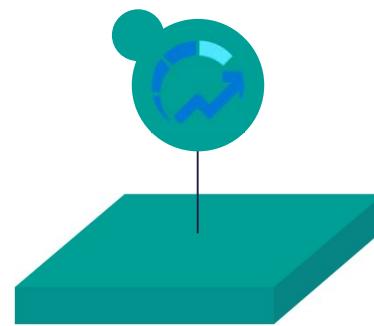
Elements of successful data governance



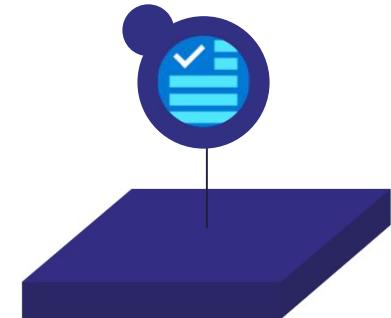
Manage growing
data landscape



Overcome
operational
silos



Increase data
agility



Comply with
industry regulations

Microsoft Purview

Unified data governance to
maximize the business
value of data



Microsoft Purview

Data governance is becoming increasingly interdisciplinary



Data governance is becoming increasingly interdisciplinary



Chief Data Officer

Data Management
Data Stewardship
Data Engineering
Data Science
Data Analystist

**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.

Microsoft Purview

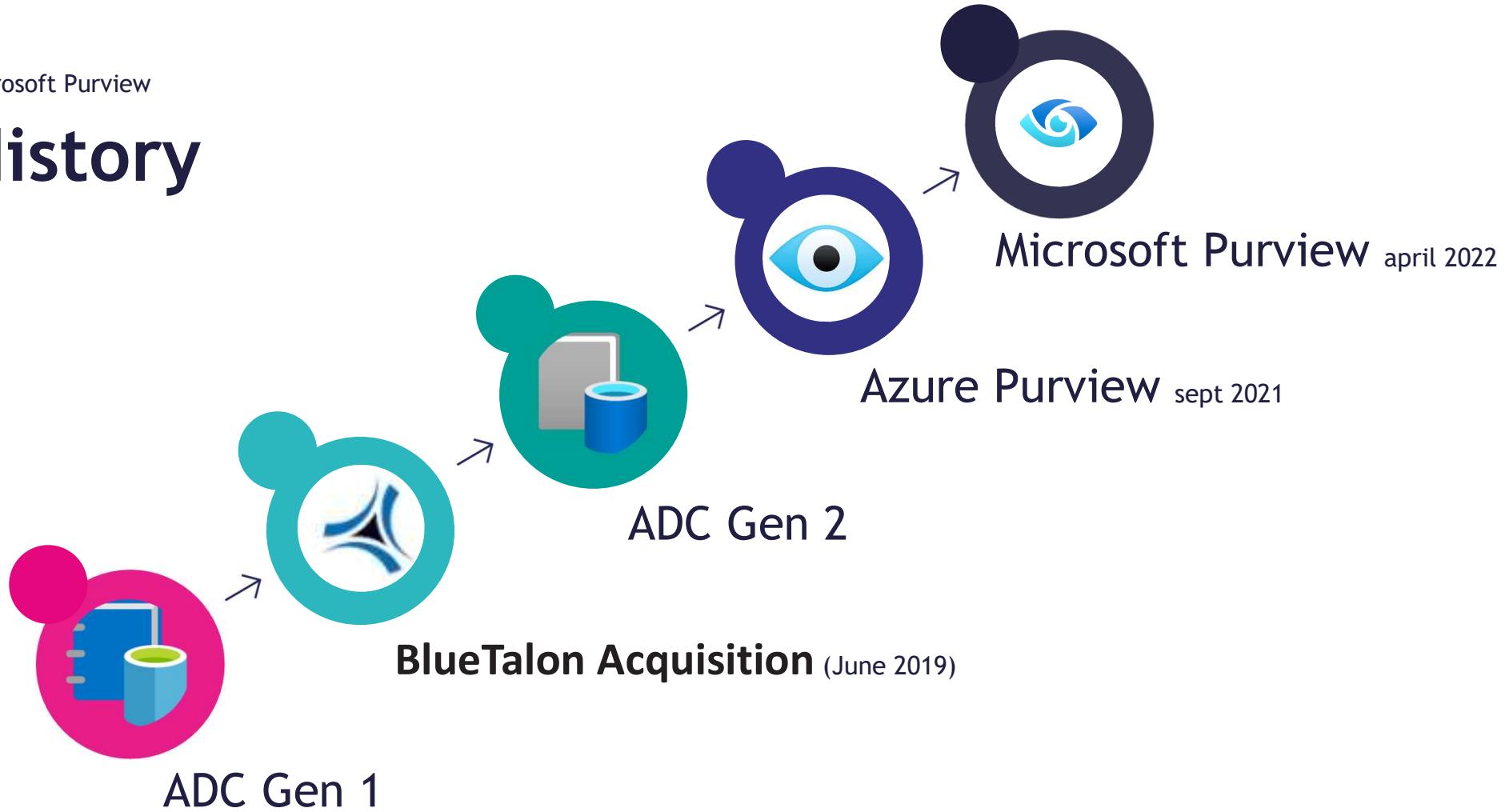


Unified **data governance** to
maximize the business value of data



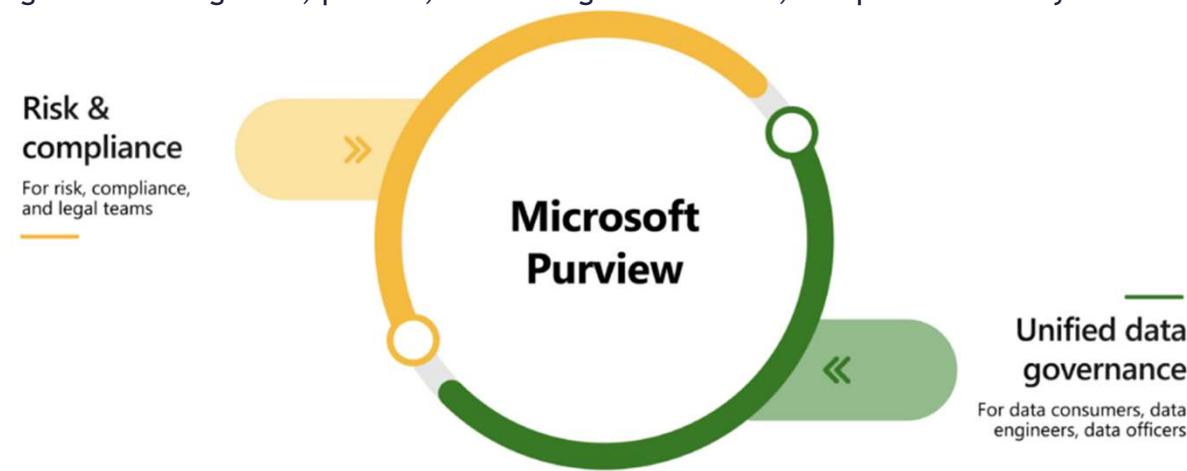
Microsoft Purview

History



The future of compliance and data governance

- Helps you gain visibility into assets across your entire data estate.
- Enables easy access to all your data, security, and risk solutions.
- Helps safeguard and manage sensitive data across clouds, apps, and endpoints.
- Manages end-to-end data risks and regulatory compliance.
- Empowers your organization to govern, protect, and manage data in new, comprehensive ways.

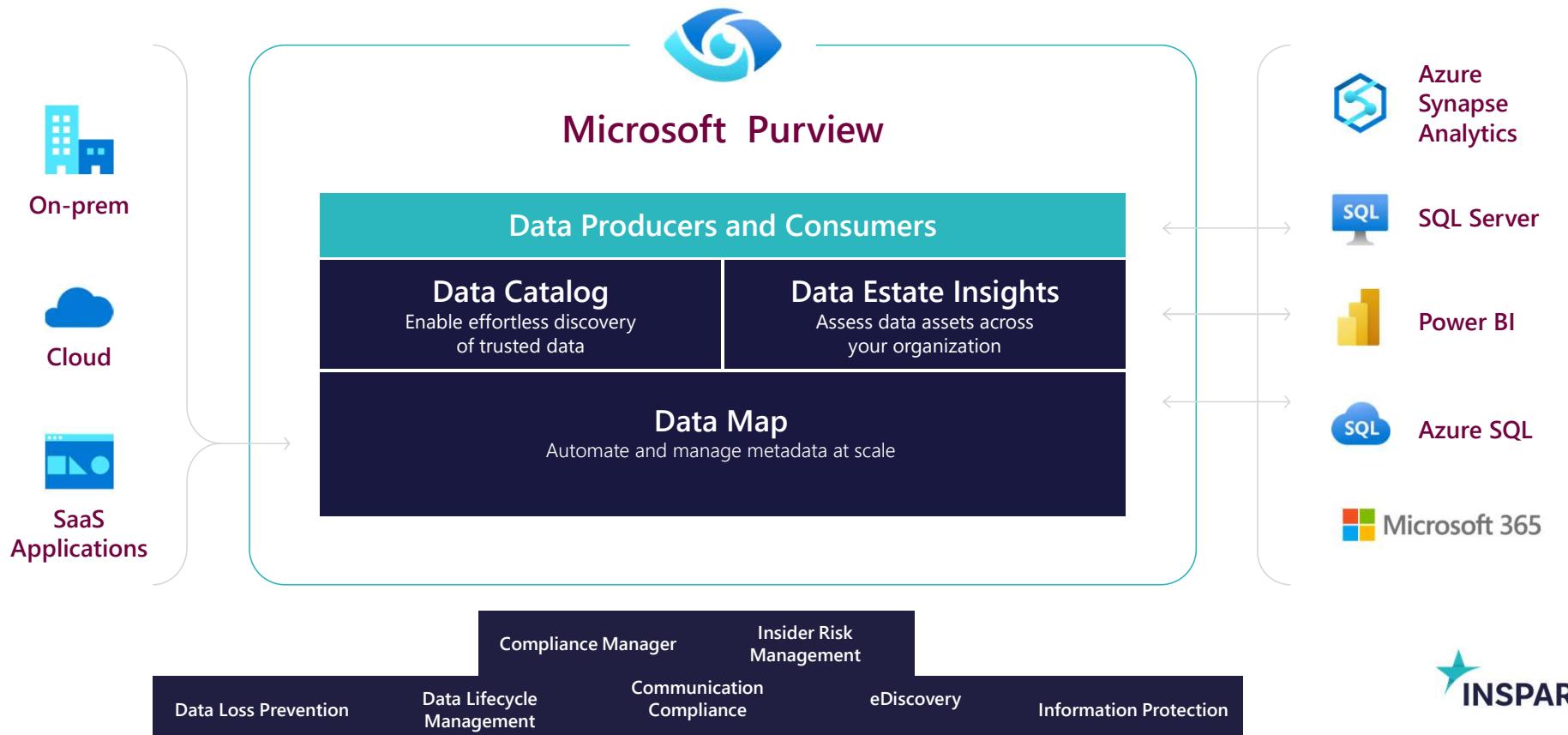


Microsoft Purview

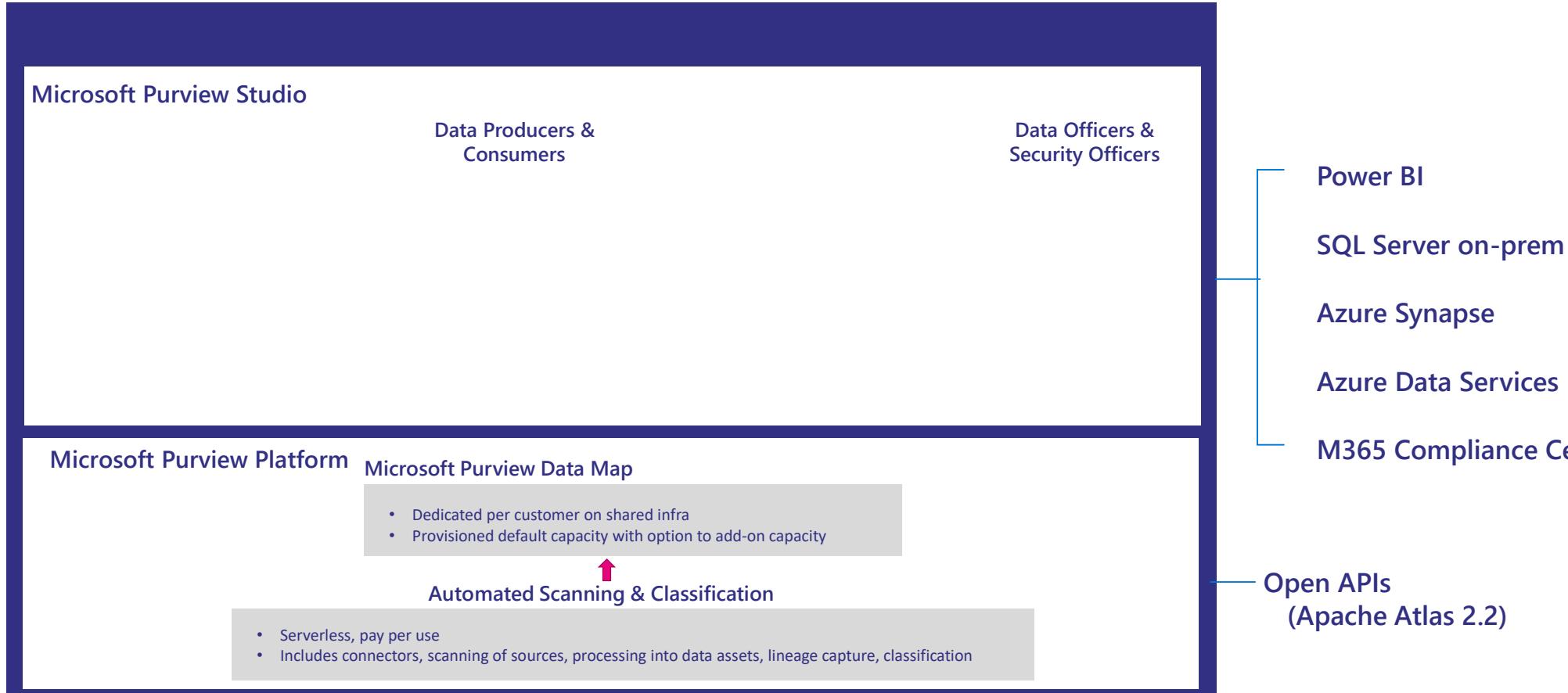
Generally Available

Preview

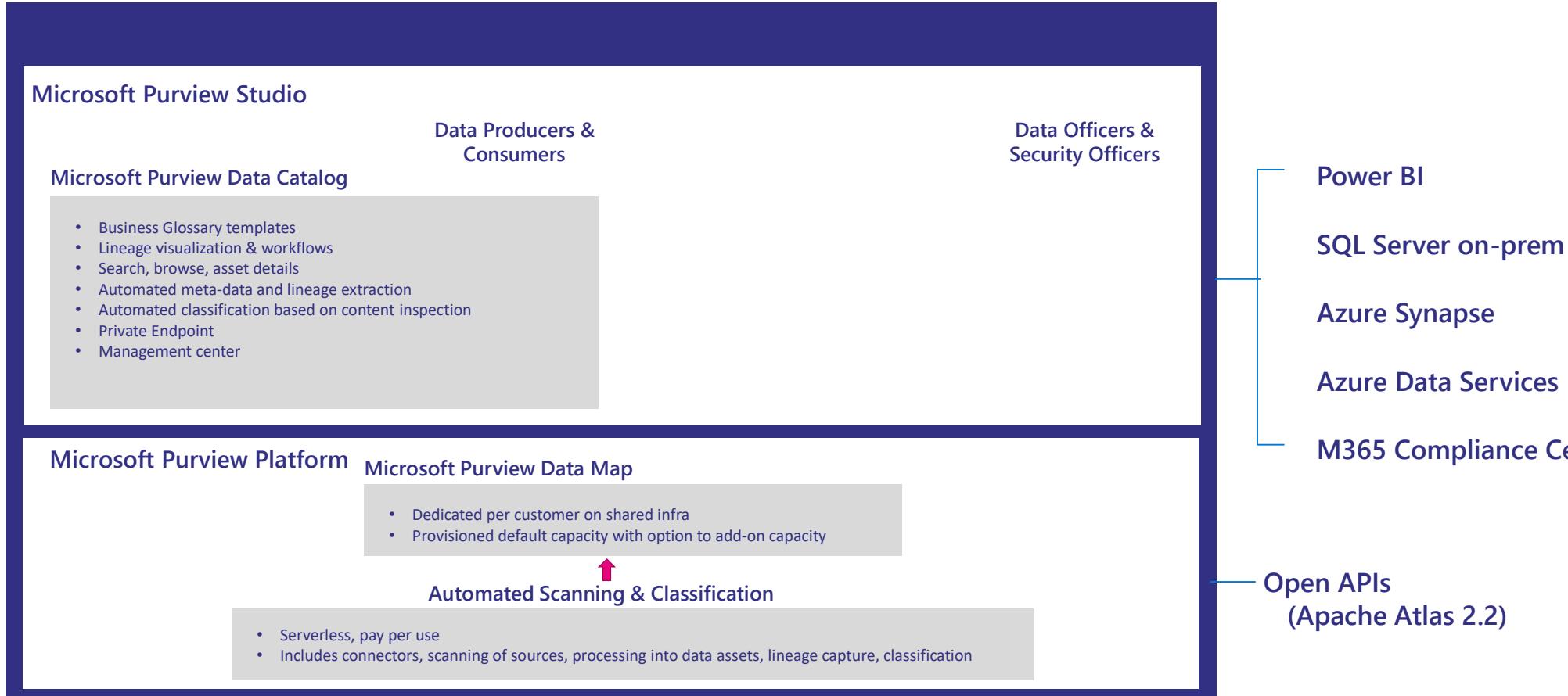
Unified Data Governance



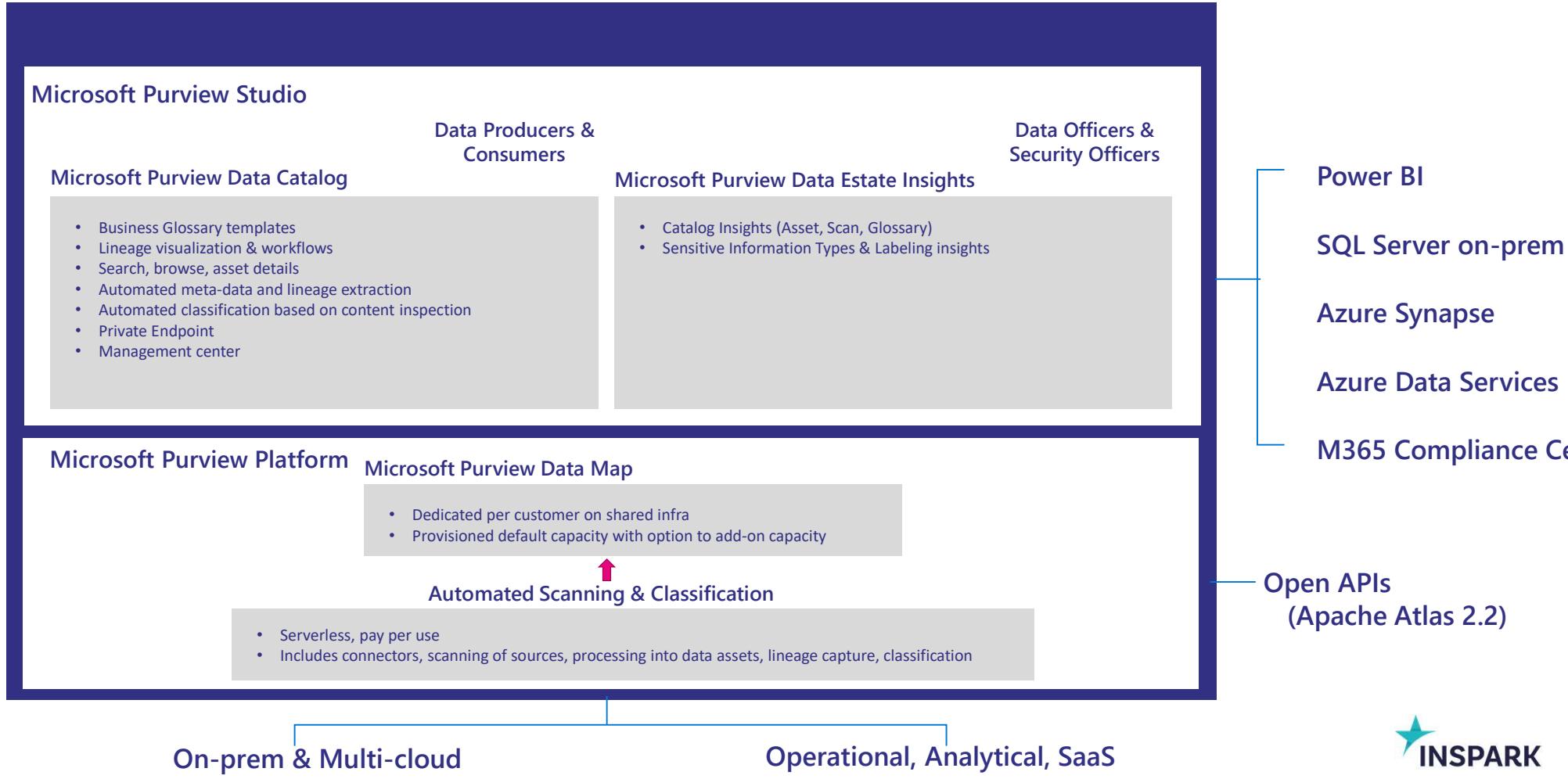
Microsoft Purview



Microsoft Purview



Microsoft Purview



Microsoft Purview

Purview Studio

The screenshot shows the Microsoft Purview InSpark studio interface. On the left, there's a sidebar with 'Metrics', 'Search Bar', and 'Recently Accessed Entities'. The main area displays 'Microsoft Purview InSpark' with stats: 18 sources, 1,819 assets, and 56 glossary terms. It features a search catalog, three key activities (Browse assets, Manage glossary, Knowledge center), and a 'Links' section with links to Purview overview, Getting started, and Documentation. A navigation bar at the top includes 'Updates', 'Accounts', 'Notifications', and other icons. The 'Recently Accessed Entities' section lists items like Product, Copy_DeltaLake_Purview, Address, and PL_EXECUTE_COPY_ADLS_TO_DELTA_LAKE_PURVIEWy2.

Updates

Accounts

Notifications

Feedback

Key Activities

Usefull Links

Metrics

Search Bar

Recently Accessed Entities

Microsoft Purview InSpark

18 sources | 1,819 assets | 56 glossary terms

Search catalog

Browse assets | Explore assets by source type and collection.

Manage glossary | Create and edit terms with custom term templates.

Knowledge center | Discover learning materials and tutorials.

Links

- Microsoft Purview overview
- Getting started
- Documentation

Recently accessed

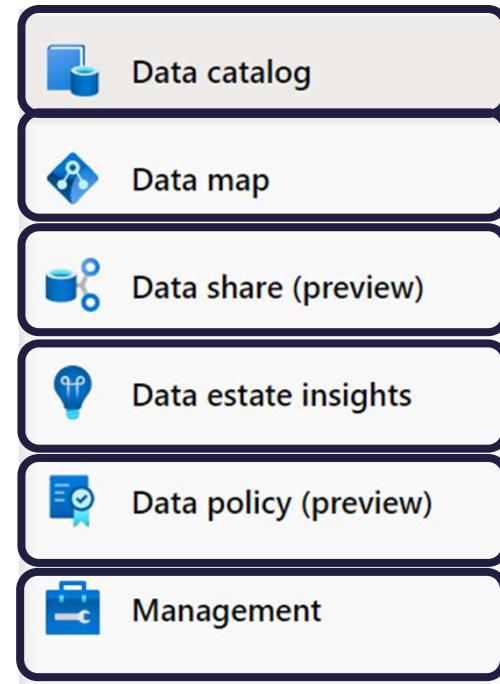
Name	Last update
Product	16 days ago
Product	3 days ago
Copy_DeltaLake_Purview	2 days ago
Address	16 days ago
PL_EXECUTE_COPY_ADLS_TO_DELTA_LAKE_PURVIEWy2	---
Copy_DeltaLake_Purview	---

→ View all recently accessed



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 Pricing

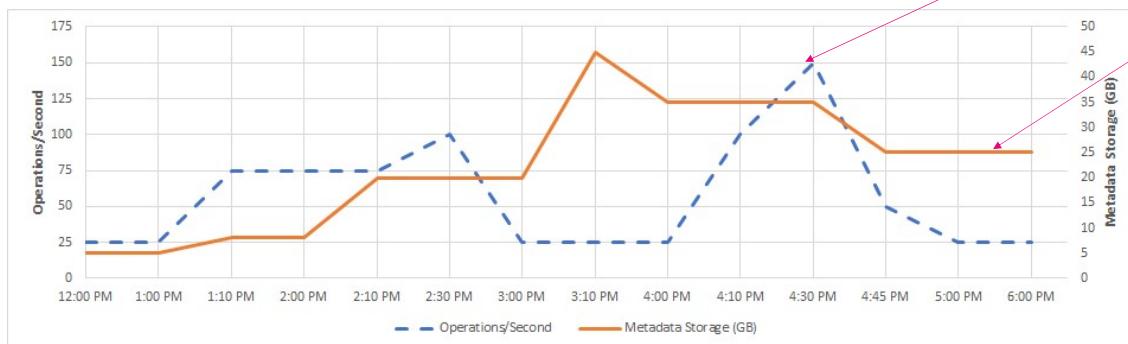
Data Map Consumption

- Elastic Data Map

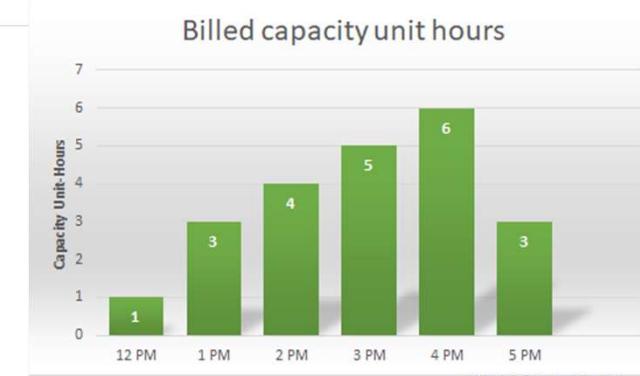
€0.407 per 1 Capacity Unit Hour

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

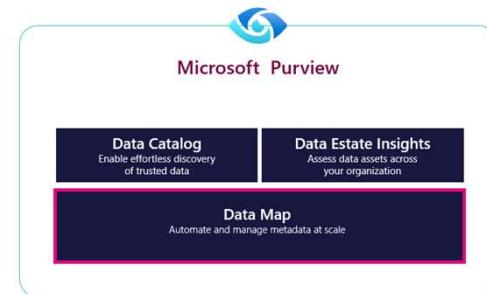


Microsoft Purview Pricing

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



Microsoft Purview Pricing

Data Map Enrichement

Features

- Advanced Resource Set

€0.208 per 1 vCore Hour

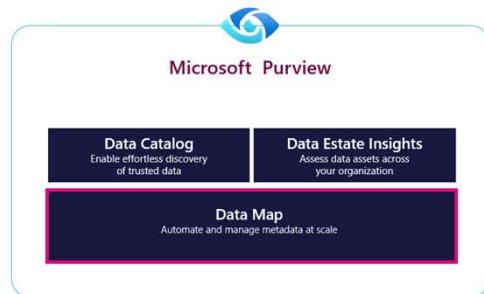
Resource sets are a built-in feature of the Data Map used to optimize the storage and search of data assets associated with partitioned files in data lakes

- 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.

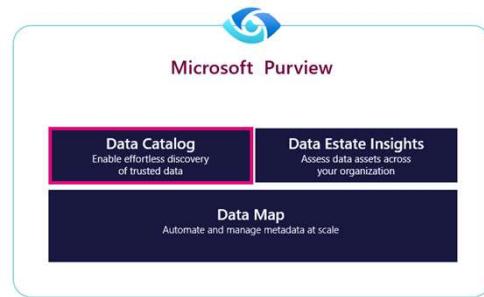
Feature options		Refresh frequency	Last updated	Description
<input checked="" type="checkbox"/>	Refresh	Off	9/28/2021, 10:00 AM	Allows customized resource sets with grouping patterns and runs aggregations on partitioned data.
<input checked="" type="checkbox"/>	Advanced Resource Sets	On	7/25/2022, 2:03 PM	Provides insight into your data estate.
<input checked="" type="checkbox"/>	Data Estate Insights	Off		



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

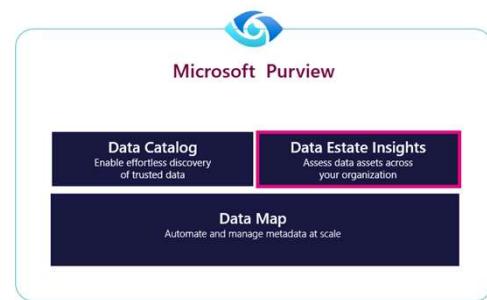


Microsoft Purview Pricing

Data Estate Insights

- Insights Consumption
€0.200 per API call

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



Preview Features

- Data Policy

SQL DevOps access

Free in preview

Data Lake data asset access

Free in preview

- Workflows

Business Workflows

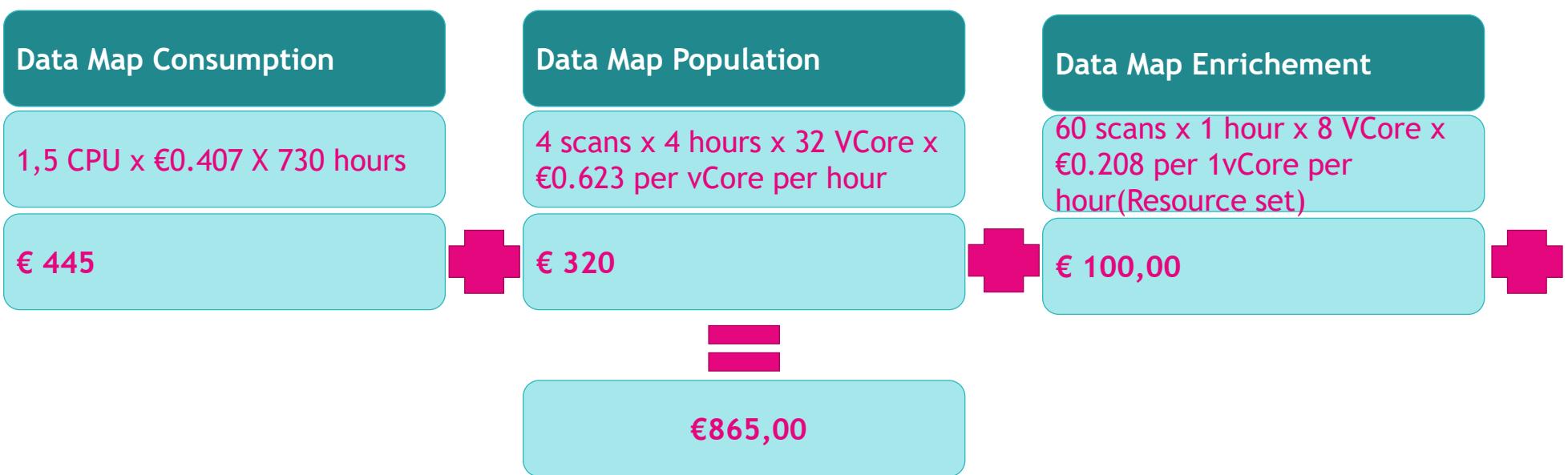
Free in preview

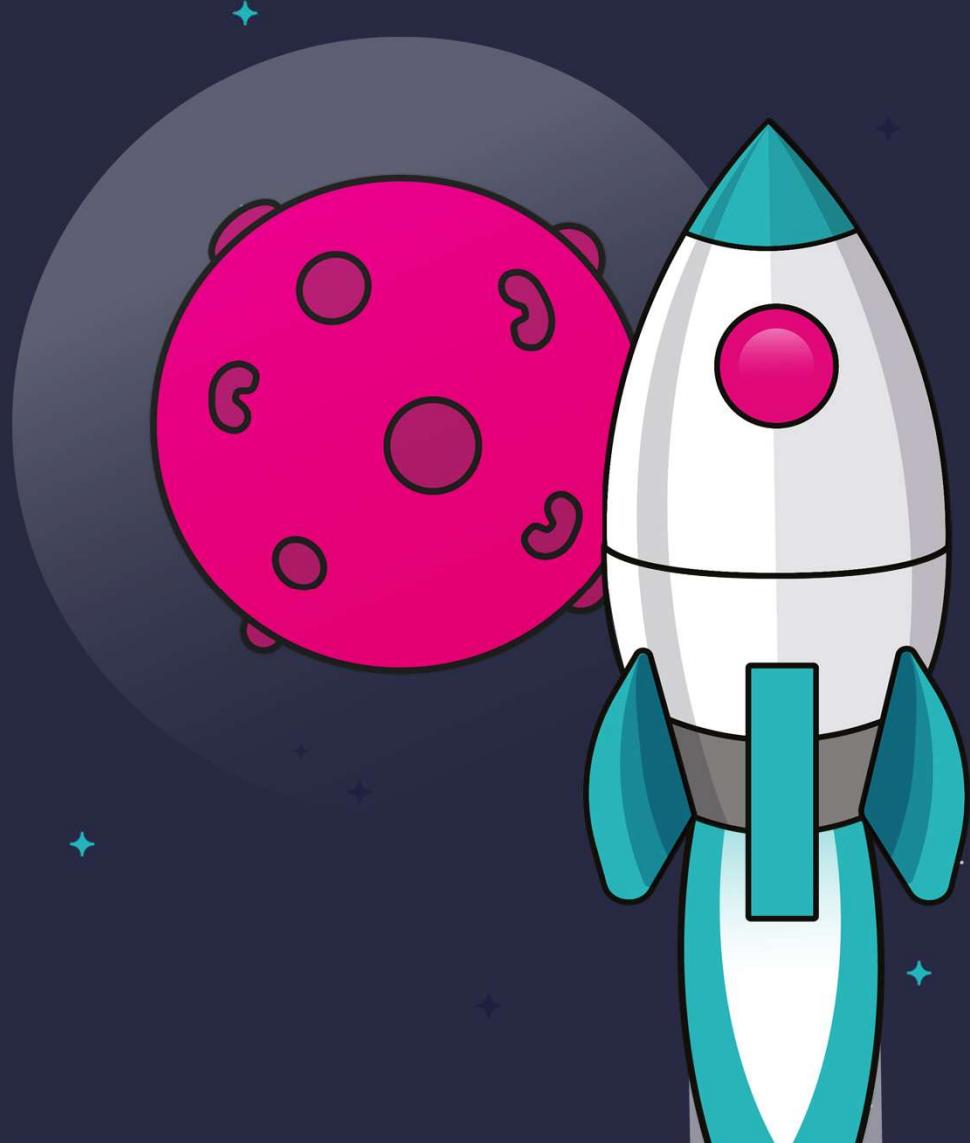
- Data Sharing

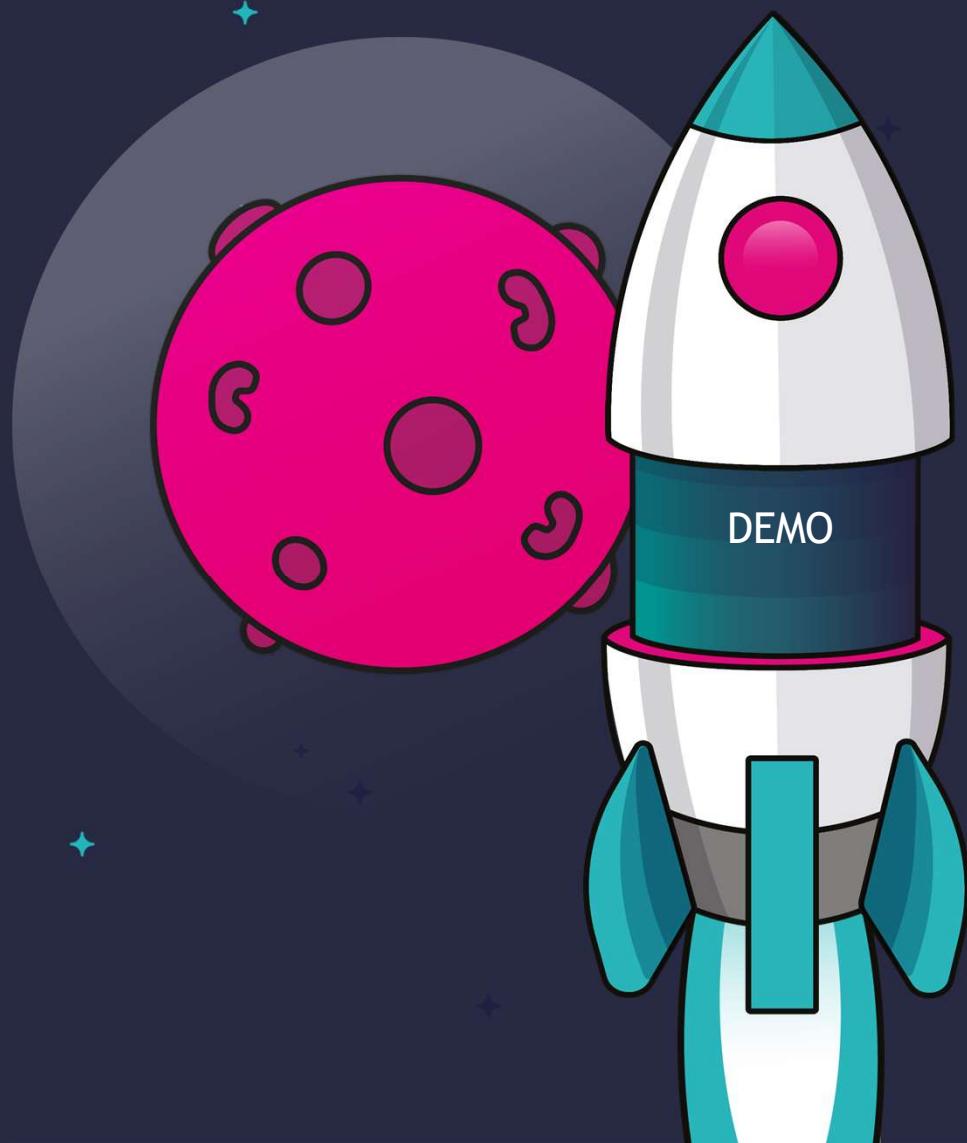
In place sharing for ABLB and ADLS

Free in preview

Pricing - Example







Innovate to accelerate



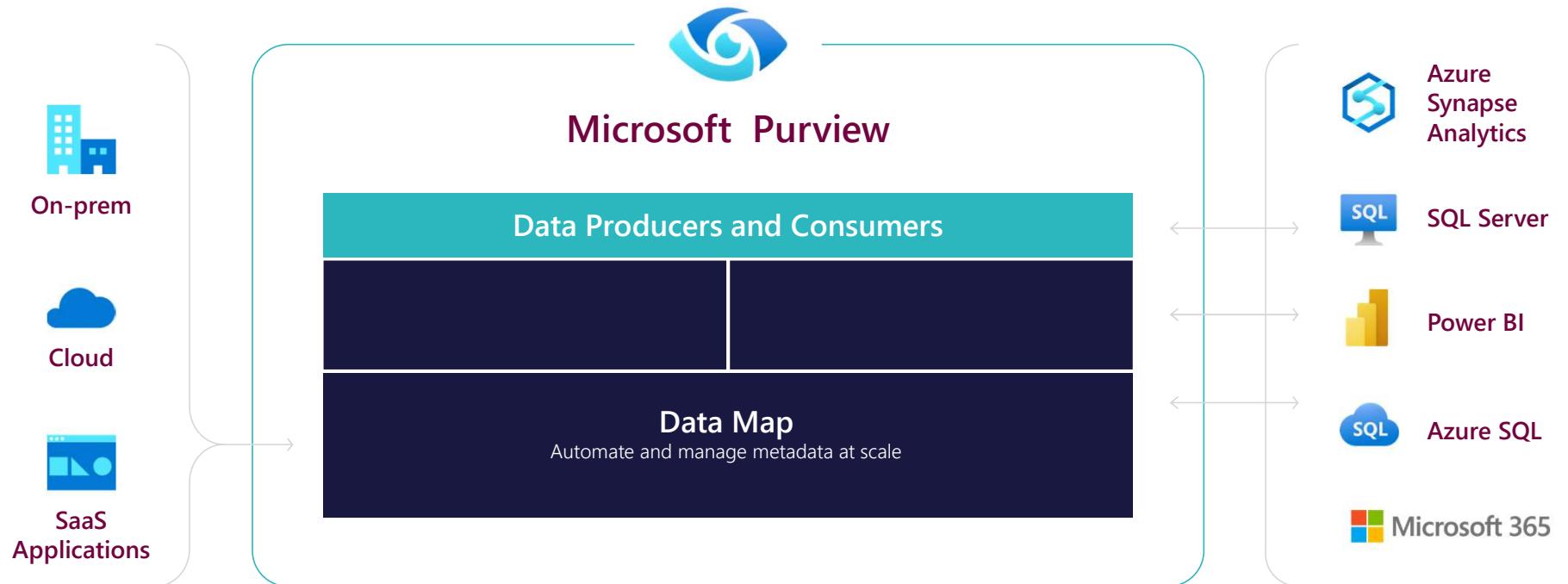
DEMO

Microsoft Purview

Generally Available

Preview

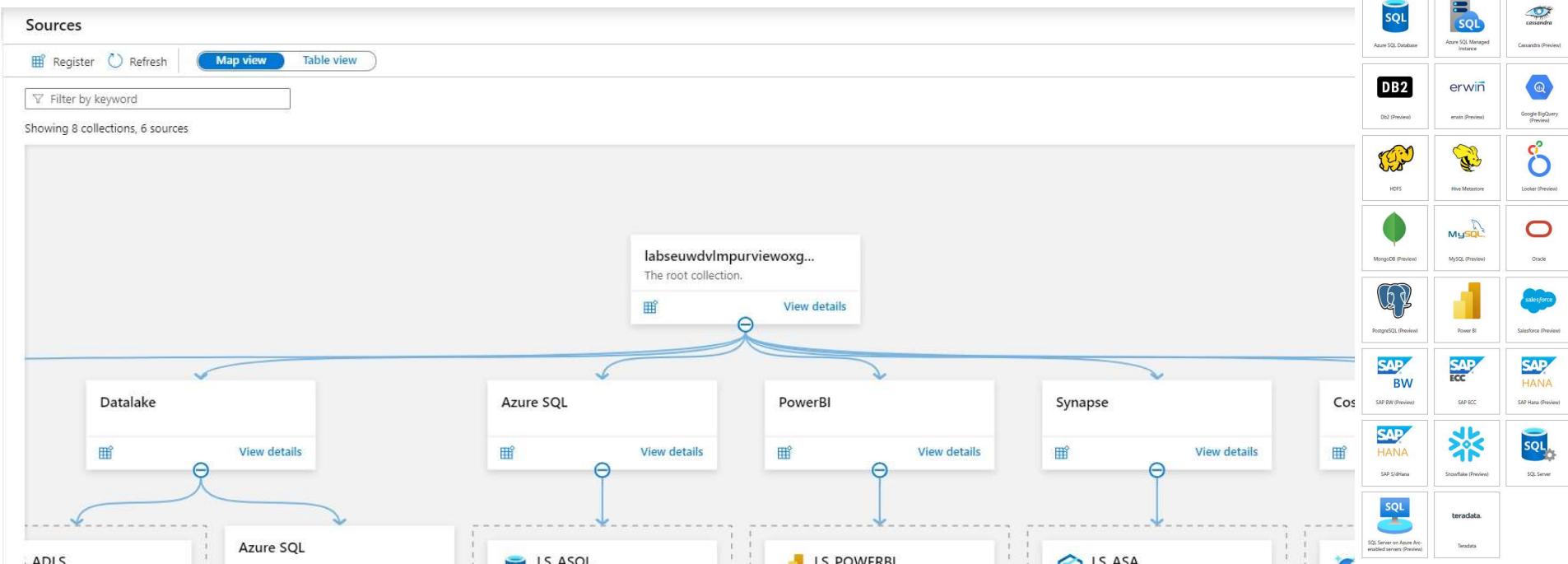
Unified Data Governance



Data Map

Collections and Sources

- A graph describing the data assets and their relationships across your data estate
- Support for 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

Scope your scan

Refresh

All future assets under a certain parent will be automatically selected if the parent is fully or partially checked.

Search

LS_ADLS

- azure-webjobs-hosts
- azure-webjobs-secrets
- intermediate
- marts
- raw
- sandbox
- temp
- tempstaging
- transfer

Set a scan trigger

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

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

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

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	World Cities		
CityName			
LastEditedBy			
LatestRecordedPopulation			
Location			

« Classifications

+ New Edit Delete Refresh

System Custom

These are the system provided classifications.

Filter by name...

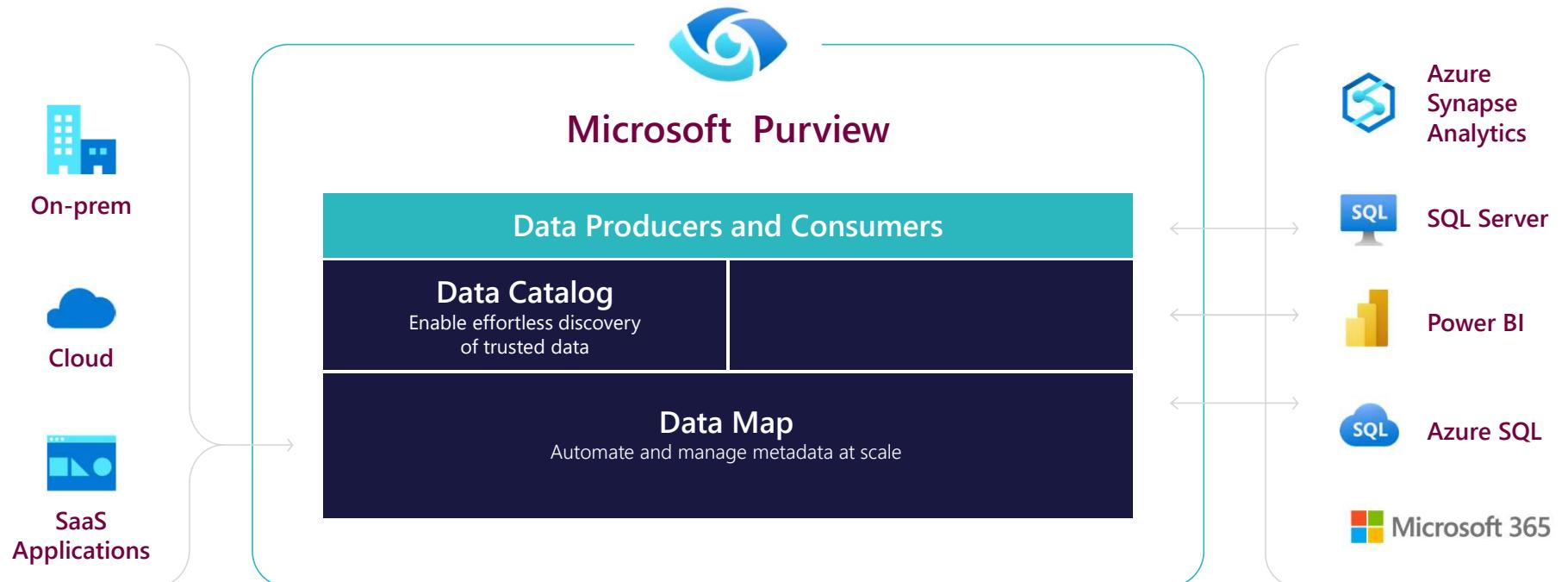
Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA Routing
Age of an individual	MICROSOFT.PERSONAL.AGE	Age of an ind...
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGINENTINA.DNI_NU...	Argentina Na...
Australia Bank Account Number	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACOUN...	Australia Ban...
Australia Business Number	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Australia Busi...
Australia Company Number	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Australia Com...
Australia Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_LIC...	Australia Driv...
Australia Medical Account Number	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL_AC...	Australia Medi...
Australia Passport Number	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT_N...	Australia Pass...
Australia Tax File Number	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE_N...	Australia Tax F...
Austria Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LIC...	Austria Driver...
Austria Identity Card	MICROSOFT.GOVERNMENT.AUSTRIA.IDENTITY.CA...	Austria Identit...

Microsoft Purview

Generally Available

Preview

Unified Data Governance



Data Catalog

Search

- Semantic search and browse
- Business glossary and workflows

The screenshot shows the Microsoft Purview InSpark Data Catalog interface. On the left, a sidebar menu includes 'Data catalog' (selected), 'Data map', 'Data estate insights (preview)', 'Data policy (preview)', and 'Management'. The main area has a search bar with 'customer' typed in, showing recent searches ('customer', 'customers') and suggestions ('CustomerCategories', 'Customer', 'CustomerKey', 'CustomerCategoryID'). Below the search bar, a list of recently accessed assets is shown under 'Recently accessed':

- SalesLT_Customer (Current)
- SalesLT_Address
- PurchaseOrderLines_Purchase
- Application_Countries (Current)
- Application_Cities

A link 'View all recently accessed' is at the bottom of this list. To the right, the main content area displays the search results for 'customer', showing 17 sources, 1,950 assets, and 56 glossary terms. The results include:

- Customer_Customer_DataFlow
- CustomerType_CustomerType_DataFlow
- Customer
- Customer

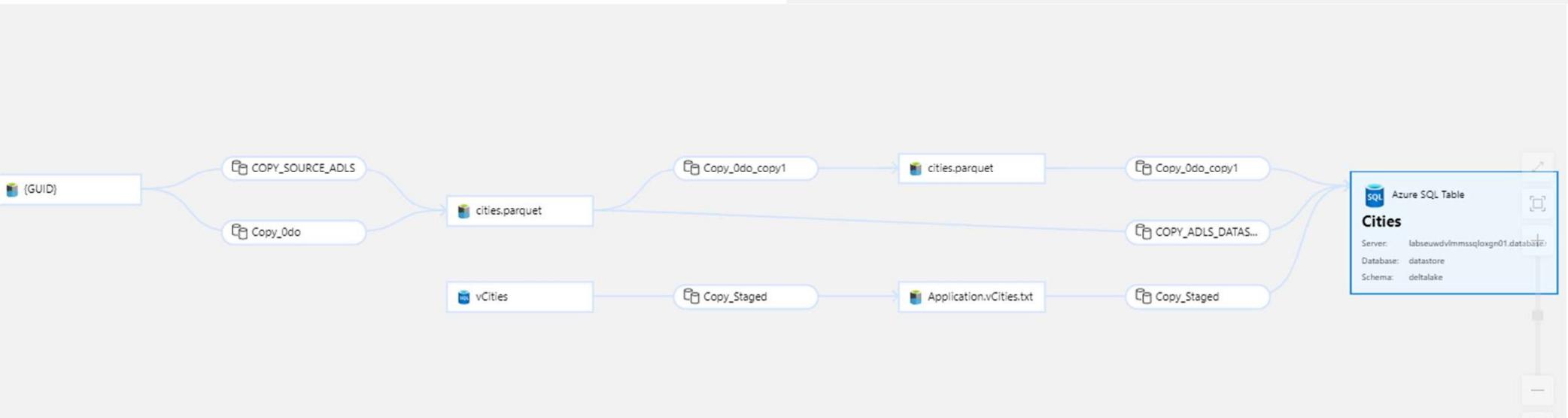
Each result entry includes a preview icon, a URL, and a 'View search results' button. The interface features a light blue and white color scheme with a 3D architectural background graphic on the right.



Data Catalog

Lineage

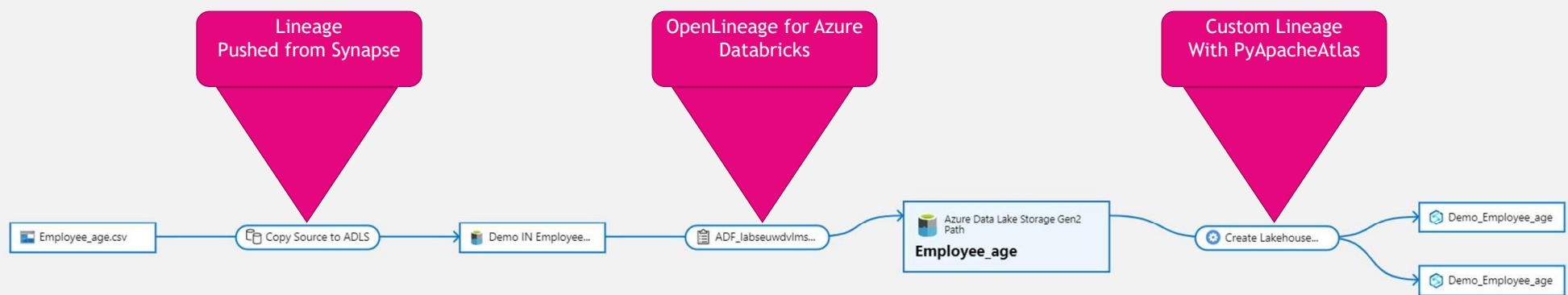
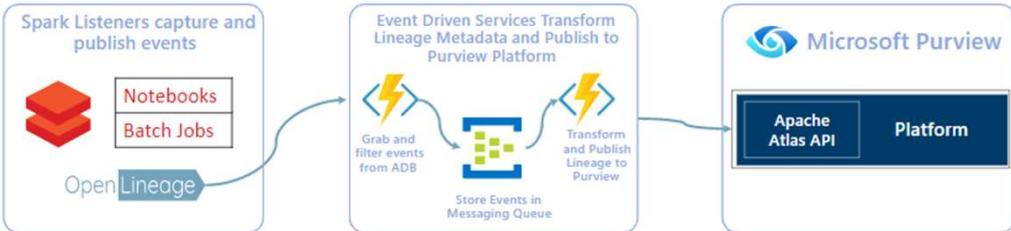
- Semantic search and browse
- Business glossary and workflows
- Data lineage with sources, owners, transformations, and lifecycle



Data Catalog

Custom Lineage

- OpenLineage for 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>

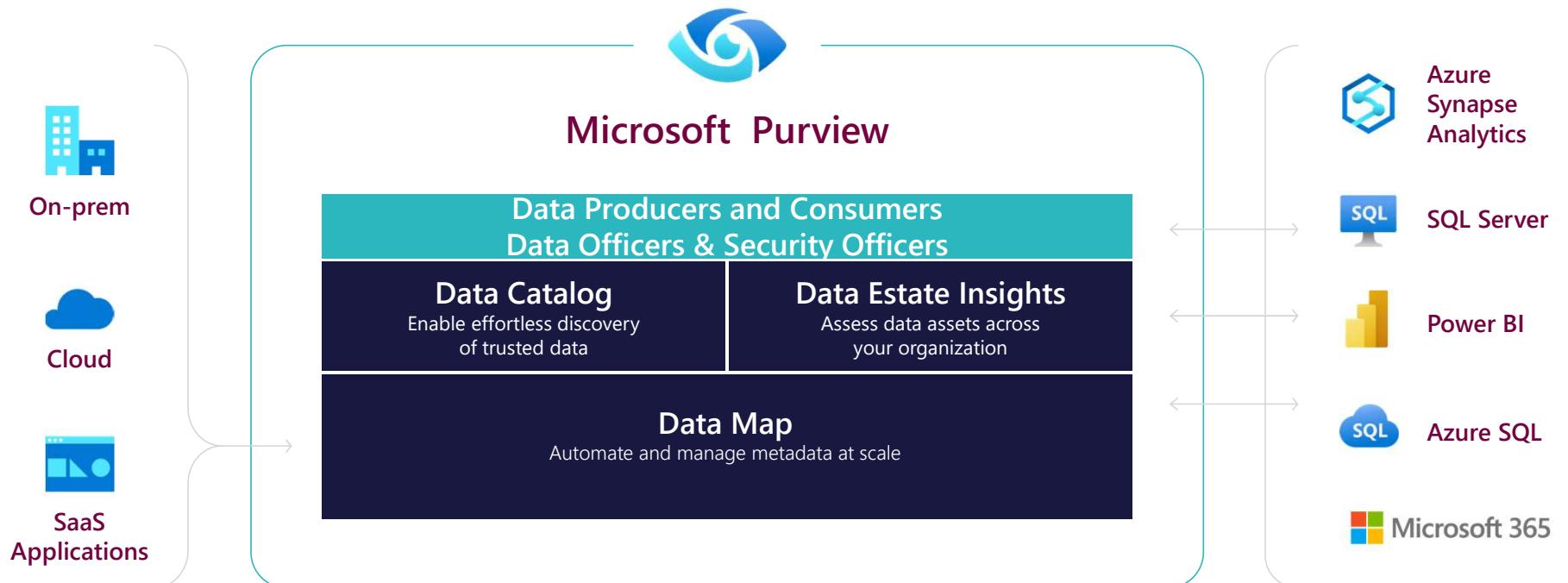


Microsoft Purview

Generally Available

Preview

Unified Data Governance



Data Estate Insight

Metrics

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

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

[Full report](#)

Report generated on July 1, 2022 at 1:10 PM

Data stewardship

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

Asset curation (29K Total assets)

- Fully curated (Green)
- Partially curated (Grey)
- Not curated (Red)

Asset data ownership (29K No owner)

- Owner assigned (Green)
- No owner (Red)

Catalog usage and adoption (0 Monthly active users)

↓100% in last month

Show: Data estate Catalog adoption

Data estate health

Collection : (Root) Microsoft Purview InSpark

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)

- Applied (Green)
- Not applied (Red)

Asset data ownership (29K No owner)

- Assigned (Green)
- Not assigned (Red)

New assets (Last 30 days) (81 New)

- New (Blue)
- Existing (Grey)

Deleted (Last 30 days) (2)

ACCESS CONTROL



Data Policy

Data Owner

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

This PREVIEW feature is licensed to you as part of your Azure Subscription. By proceeding you acknowledge the Preview Terms and Privacy Statement.

Access Control Policy

Name *

Description *

Policy statements ⓘ
Create up to 5 policy statements for your policy.

Example policy statement:
Allow Read on data asset finance to group non-FTEs .

+ New policy statement

Allow on this action on these data resources to these subjects

Modify
Read
SQL Performance Monitoring
SQL Security Auditing

Data resources
Select a type of data resource to apply your policy statement to.

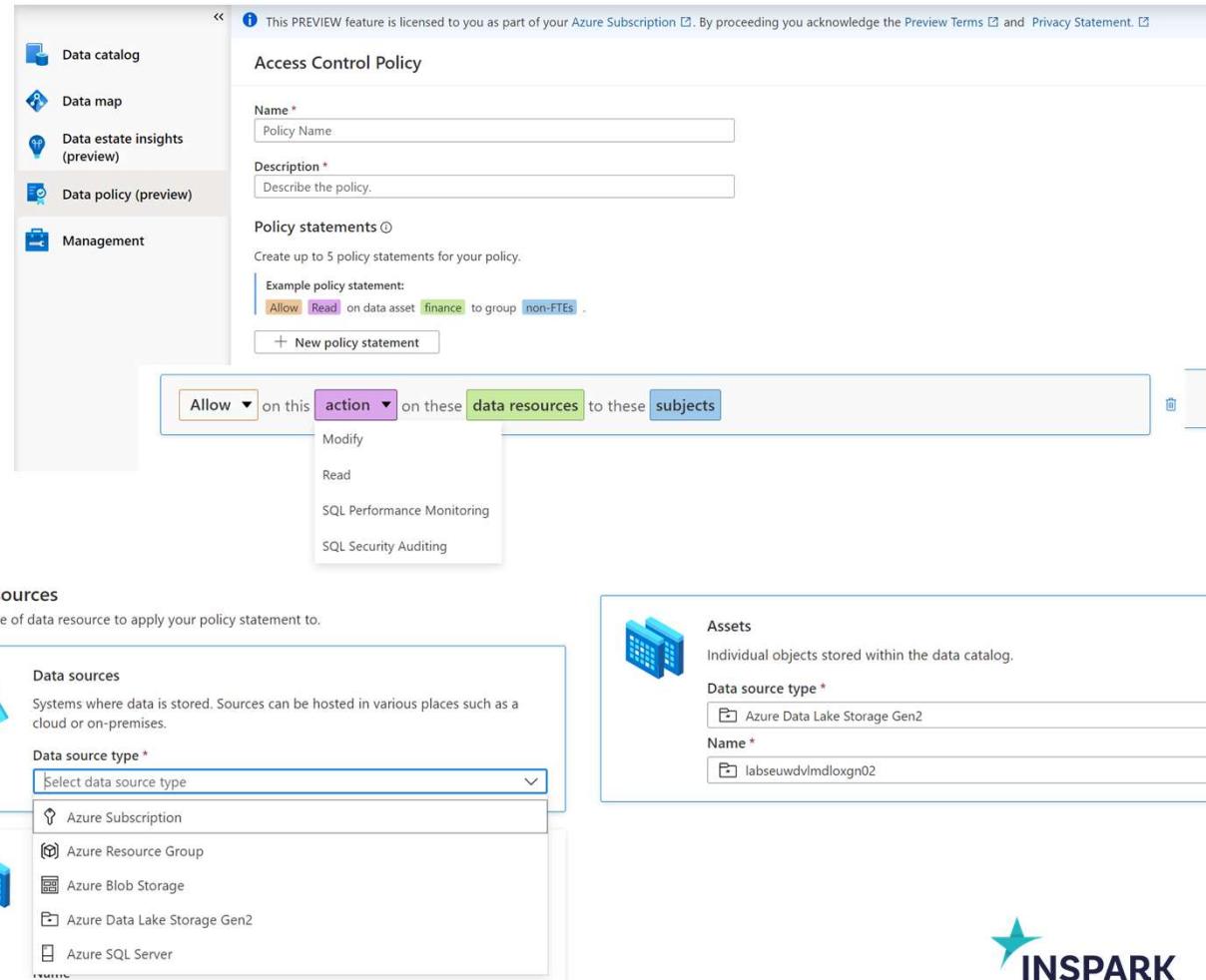
Data sources
Systems where data is stored. Sources can be hosted in various places such as a cloud or on-premises.

Data source type *

Azure Subscription
Azure Resource Group
Azure Blob Storage
Azure Data Lake Storage Gen2
Azure SQL Server

Assets
Individual objects stored within the data catalog.

Data source type *
Name *



Data Policy

Self Service

- Workflow Governance

Data Access

The screenshot shows the Microsoft Purview Workflow Authoring interface. A workflow named "Self service request for Customer" is displayed. The workflow starts with a trigger "When data access request is submitted", followed by a task "Start and wait for an approval". This is followed by a condition "Condition". If the condition is "If yes", it leads to a task "Check data source registration for data use governance", then a condition "Data use governance Condition", and finally a task "Grant access". If the condition is "If no", it leads to a task "Create Task and wait for task completion".

Customer
Azure Data Lake Storage Gen2 Resource Set

Edit Select for bulk edit Request access Refresh Delete

Overview Properties Schema Lineage Contacts Related

Asset description
No description for this asset.

Managed attributes (preview)
No Attributes for this asset.

Request access

If you don't have access to this asset, you can request access to its parent asset "Customer" here and an administrator or owner can grant you access to the data.

I'd like access, please.

Request for someone else



Data Policy

Self Service

- Workflow Governance

[Data Access](#)

- Workflow Data Catalog

[Create Glossary Terms](#)

[Delete Glossary Terms](#)

[Import Terms](#)

[Update Glossary Terms](#)

New data catalog workflow

Select a workflow template to get started:

Create glossary term

Provides approval workflow for newly added glossary terms.

Delete glossary term

Provides approval workflow when a glossary term is removed.

Import terms

Provides approval workflow when multiple glossary terms are added.

Update glossary term

Provides approval workflow for glossary term changes.



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

Microsoft Purview

Connecting Azure Synapse workspace to Azure Purview will enable you to discover trusted and accurate data across your hybrid environment. [Learn more](#)

About Purview account 01

Purview account

Default Purview integration capabilities:

- Data Catalog - Discover enterprise data assets using Microsoft Purview powered search

Additional Purview integration capabilities and status:

Integration	Status
Data Lineage - Synapse Pipeline	✓ Connected

Lineage status details

Status	Result
✓	Succeeded

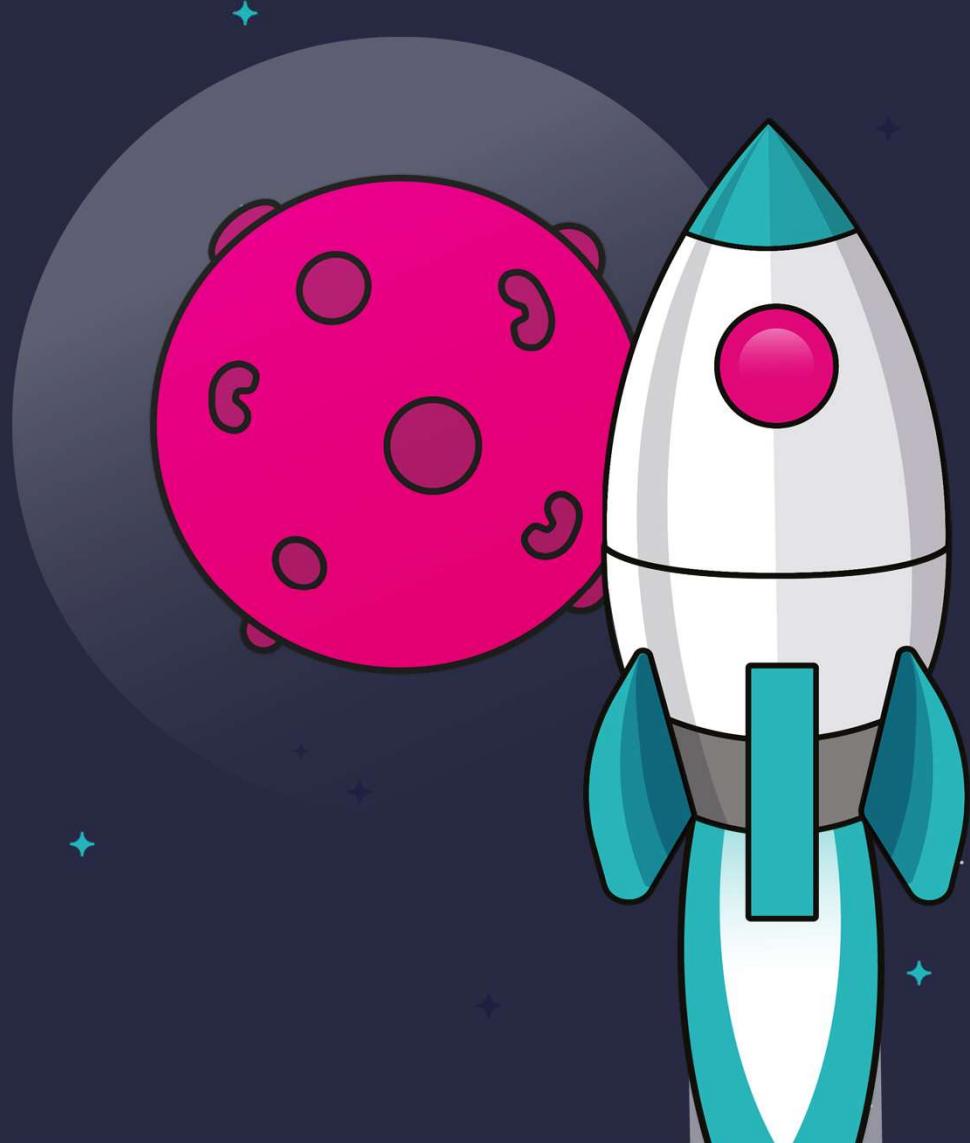
Lineage status details

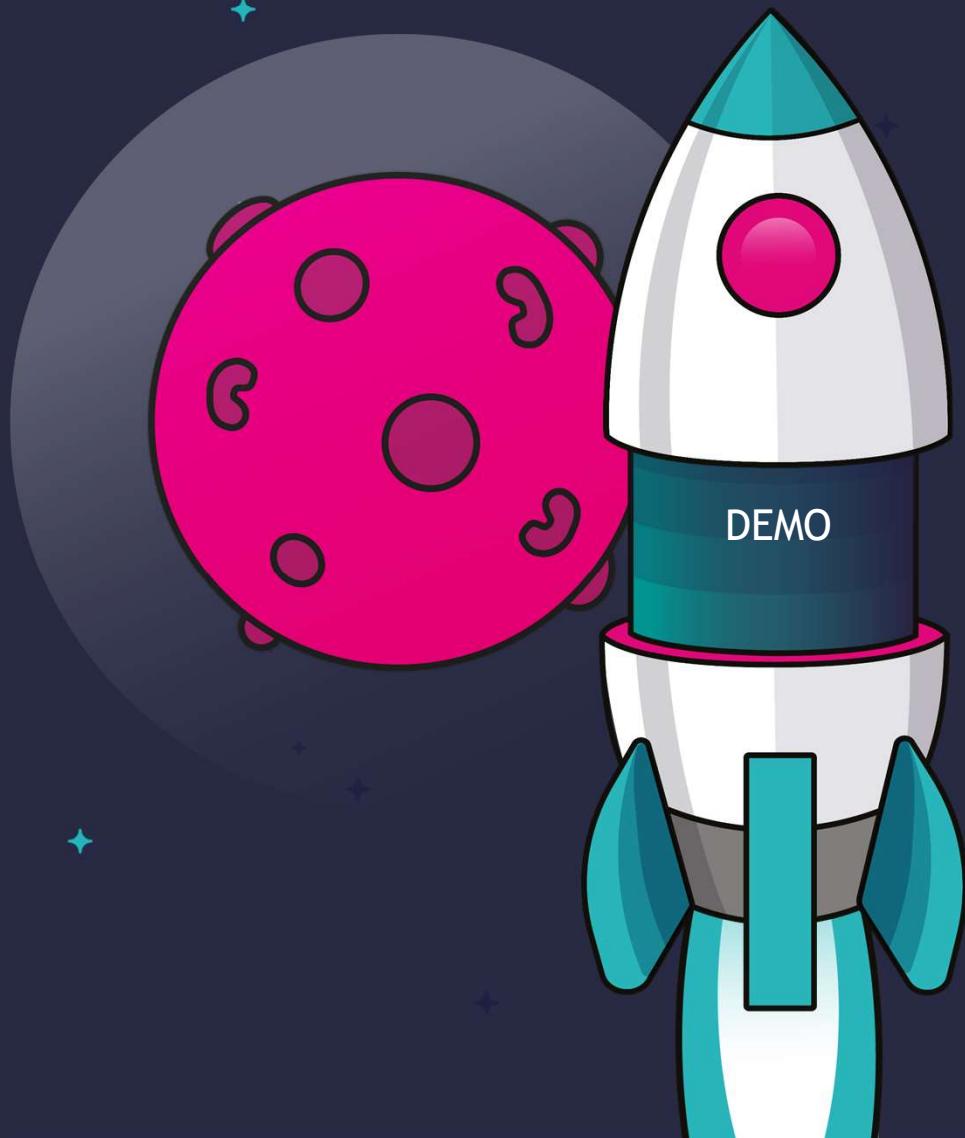
Status	Result
✗	Unsupported

Details The lineage of this copy activity is not reported because it went through the following unsupported scenario: stored procedure or query used in SQL source.

[Learn more about Microsoft Purview integration](#)





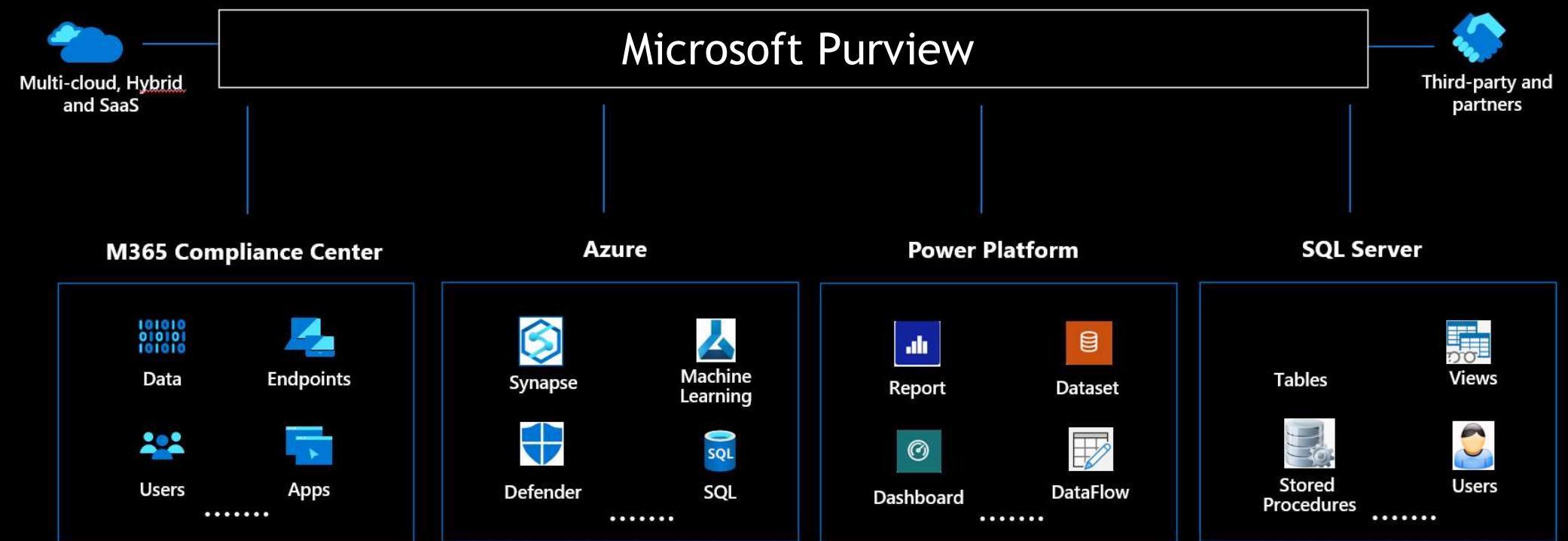


Innovate to accelerate



DEMO

Data Governance Opportunity

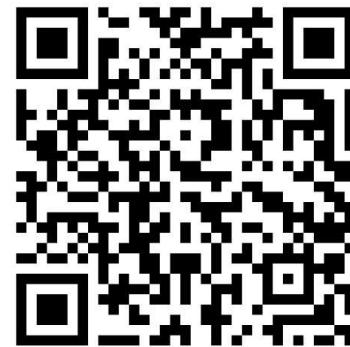


Questions?



Erwin de Kreuk

Principal Consultant | Lead Data
and AI | Data Platform MVP | Publ...



<https://erwindekreuk.com>



@erwindekreuk

Session Feedback



Event Feedback

Send your feedback and you will be entered in de prize draw



