

Microsoft Purview

Microsoft's answer to Data Governance and Data Lineage



Erwin de Kreuk

We Are InSpark

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

Microsoft
Partner

2021 Partner of the Year Finalist
Analytics Award



Microsoft Purview



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



Microsoft Purview

Topics

Introduction to Purview

Data Map

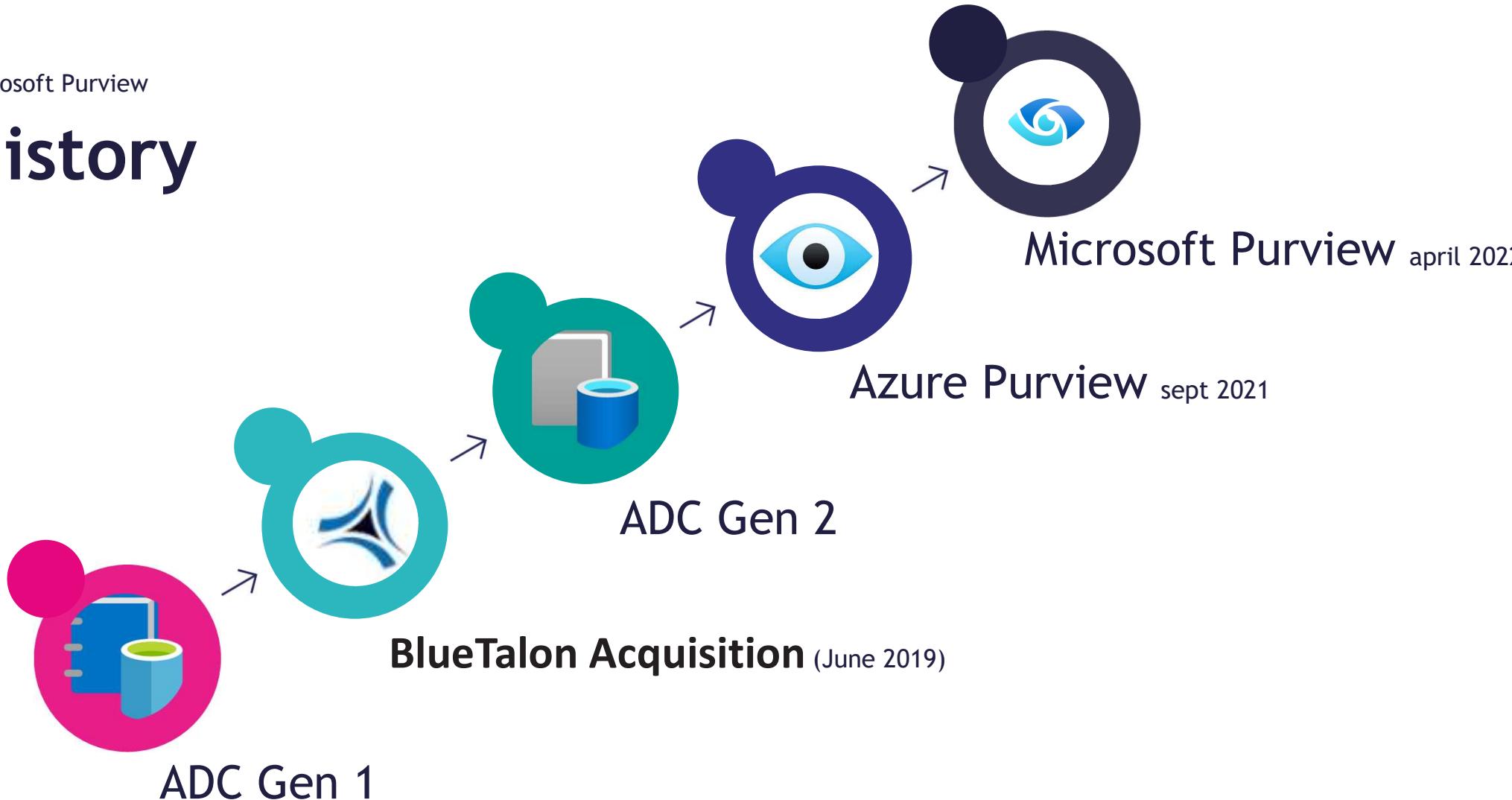
Data Catalog

Data Estate Insights



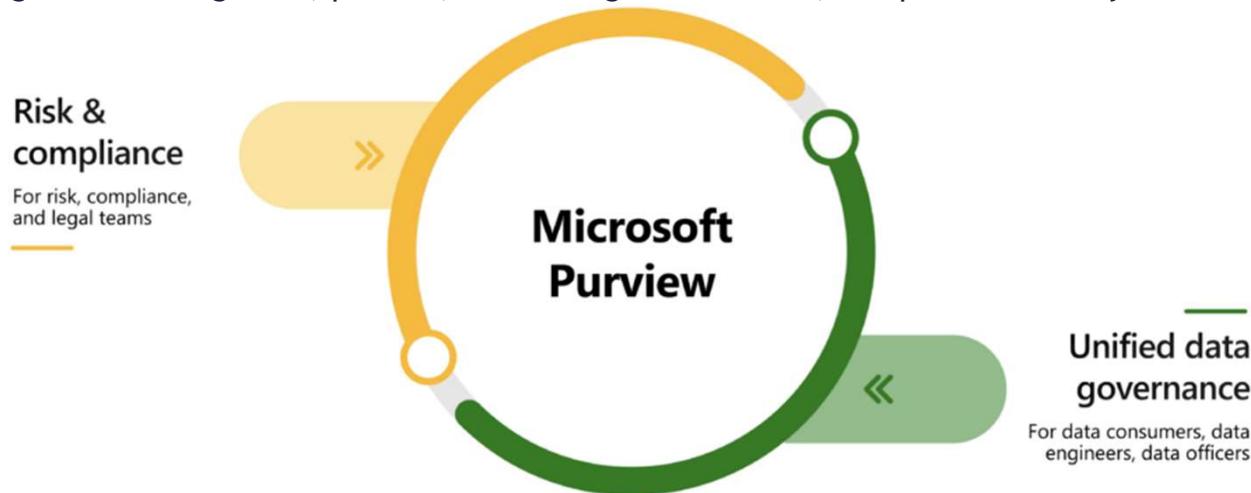
Microsoft Purview

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.

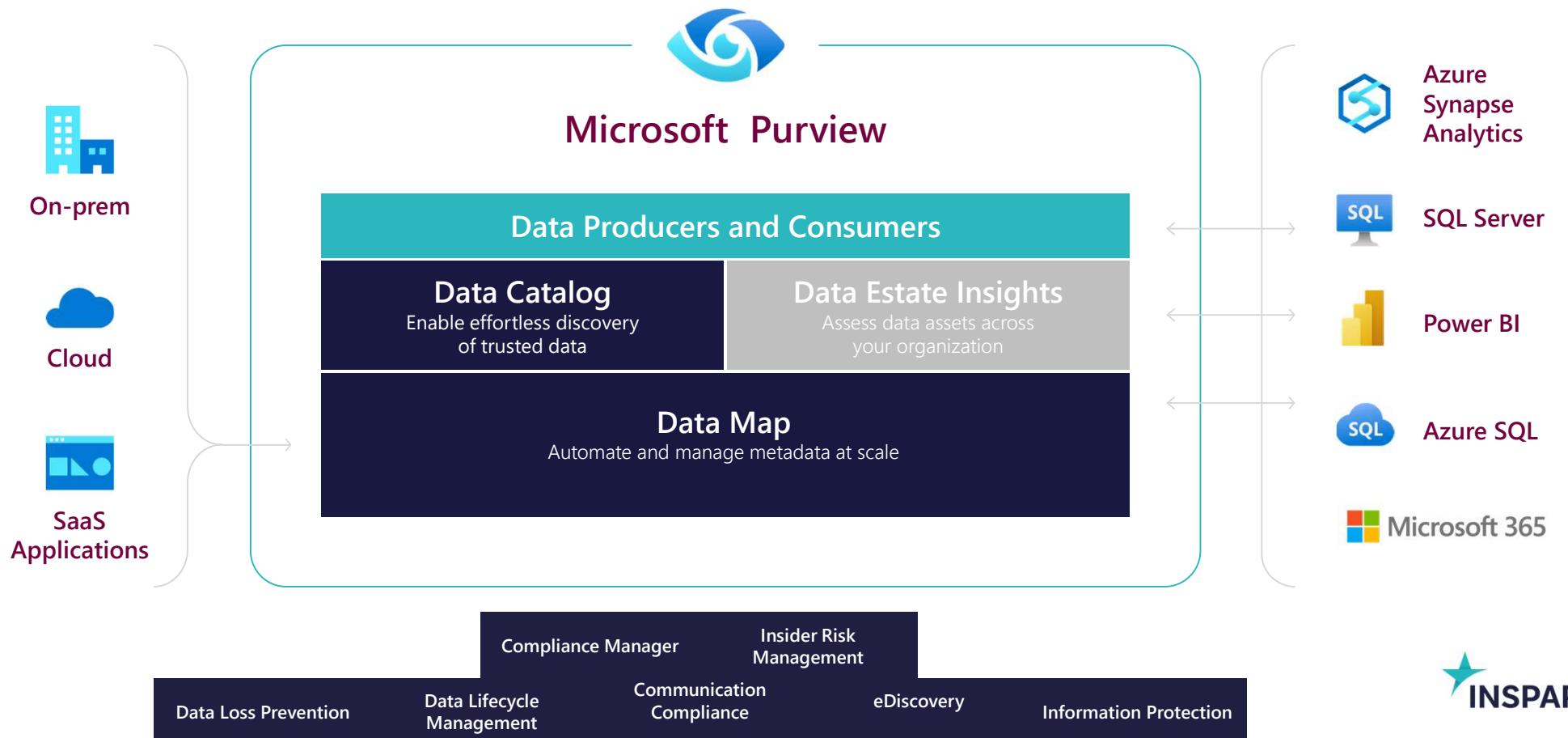


Microsoft Purview

Generally Available

Preview

Unified Data Governance



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 Purview

Data governance is becoming increasingly interdisciplinary



Data governance is becoming increasingly interdisciplinary



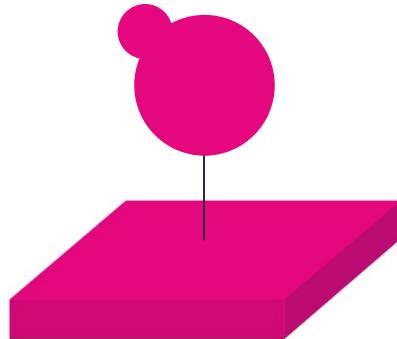
Chief Data Officer

Data Management
Data Stewardship
Data Engineering
Data Science
Data Analystist

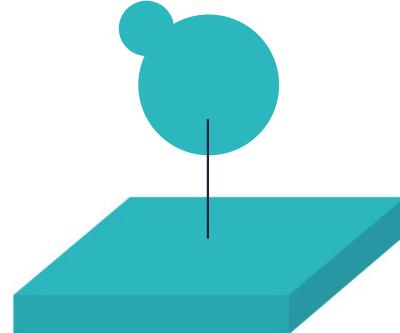
**What data do I have?
Where did the data originate?
Can I trust it?
What's my exposure to risk?
Is my usage compliant?
How do I control access & use?
What is required by regulation X?**

Data Compliancy
Data Access
Data Security
Data Privacy

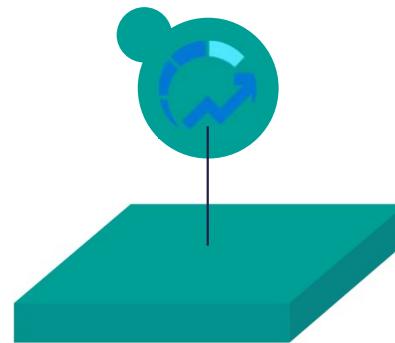
Elements of successful data governance



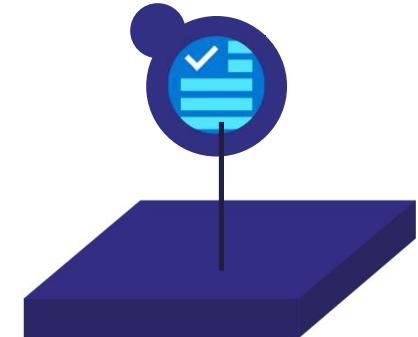
Manage growing
data landscape



Overcome
operational silos



Increase data
agility



Comply with
industry
regulations

Microsoft Purview

Unified data governance to
maximize the business
value of data





Reimagine data governance in the cloud

- ❑ Manage and govern operational, transactional and analytical data
- ❑ Cloud-native, purpose-built service to address discovery and compliance needs
- ❑ Fully managed, serverless, PaaS service
- ❑ Eliminate manual, ad-hoc and homegrown solutions





Set the foundation for effective data governance

- Automate discovery of data in on-premises, multicloud and SaaS sources
- Classify data at scale to specify sensitivity, compliance, industry, business and company-specific value
- Know where data came from and what was derived from it with data lineage





Maximize business value of **data** for data consumers

- Connect business and technical data analysts, data scientists, and data engineers to a trusted data catalog
- Enable users to quickly find data and view its lineage and sensitivity
- Deliver a curated and consistent glossary of business terms and definitions





Gain insight into **data use**
across the estate

- Understand at a glance how data is being created and used across your data estate
- Visually assess the state of data assets, scans, business glossary and sensitive data

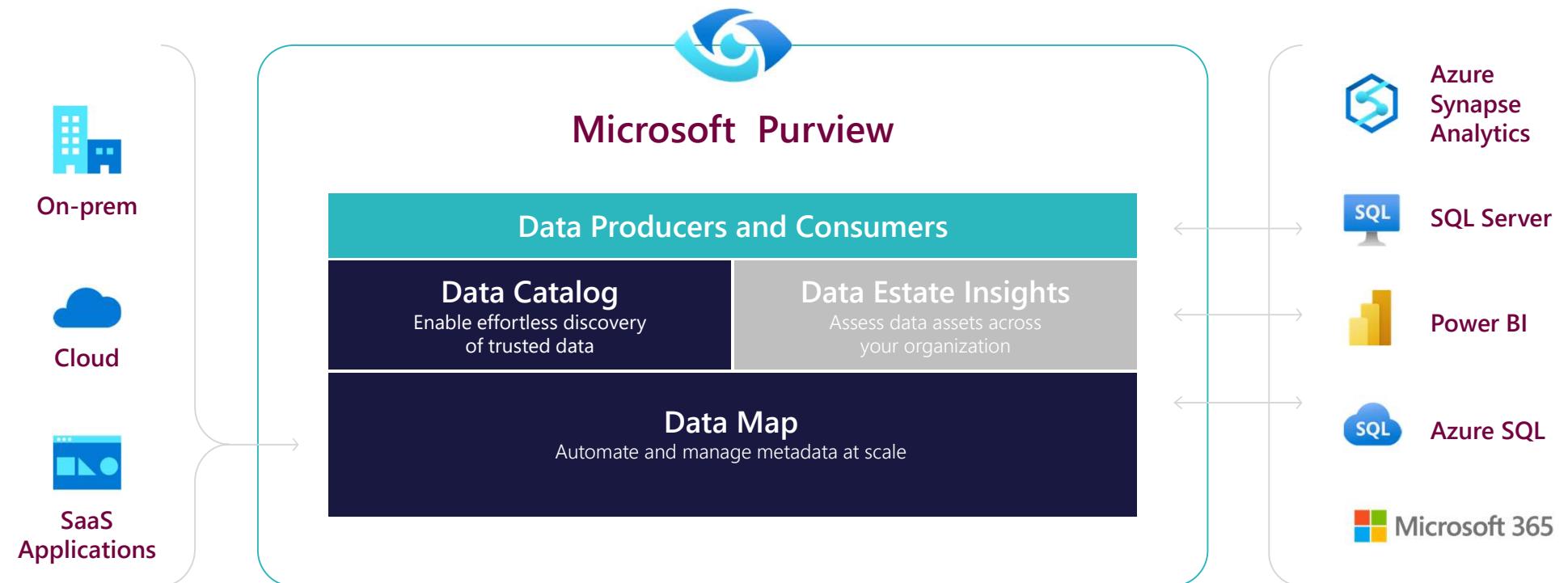


Microsoft Purview

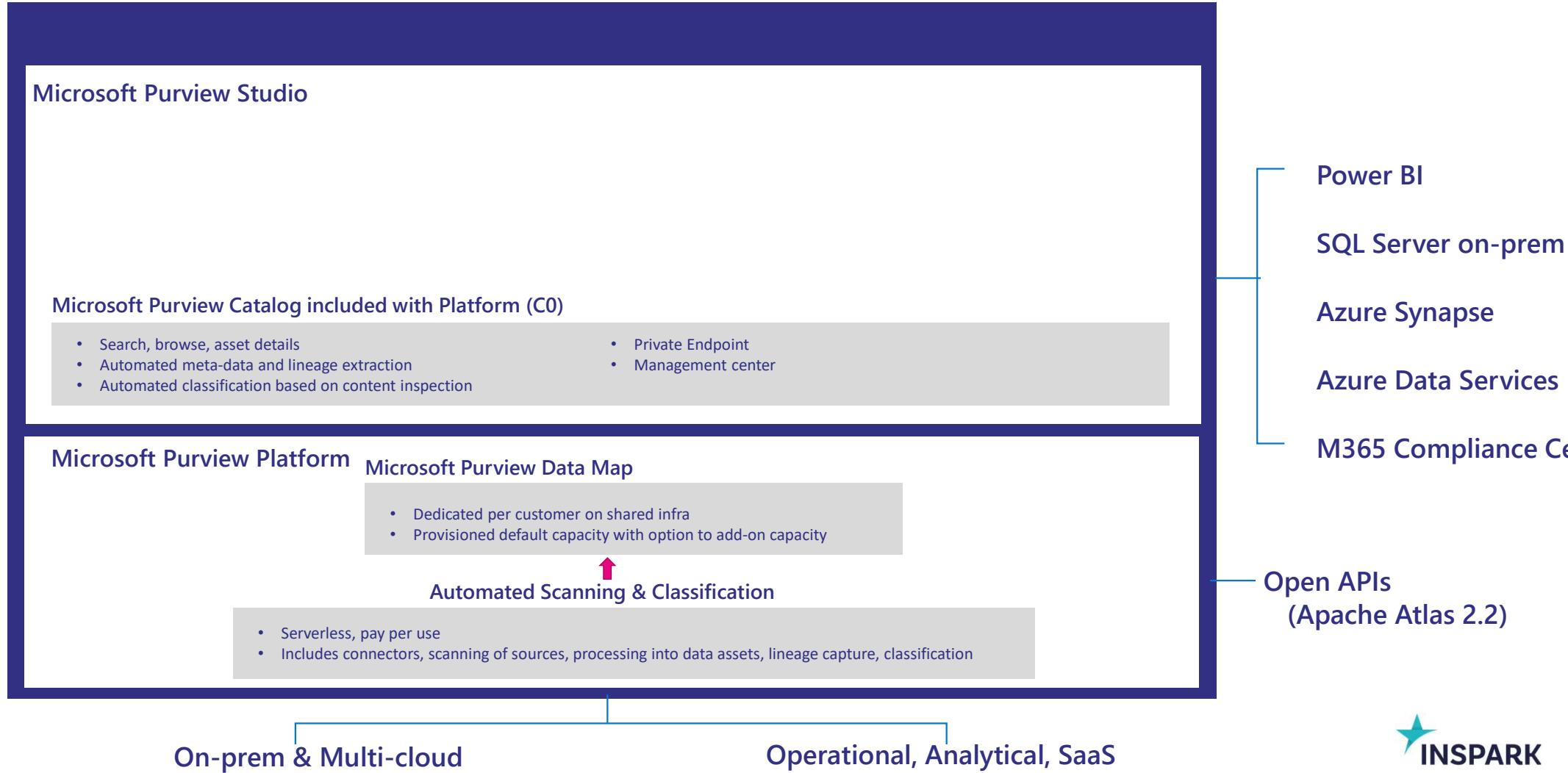
Generally Available

Preview

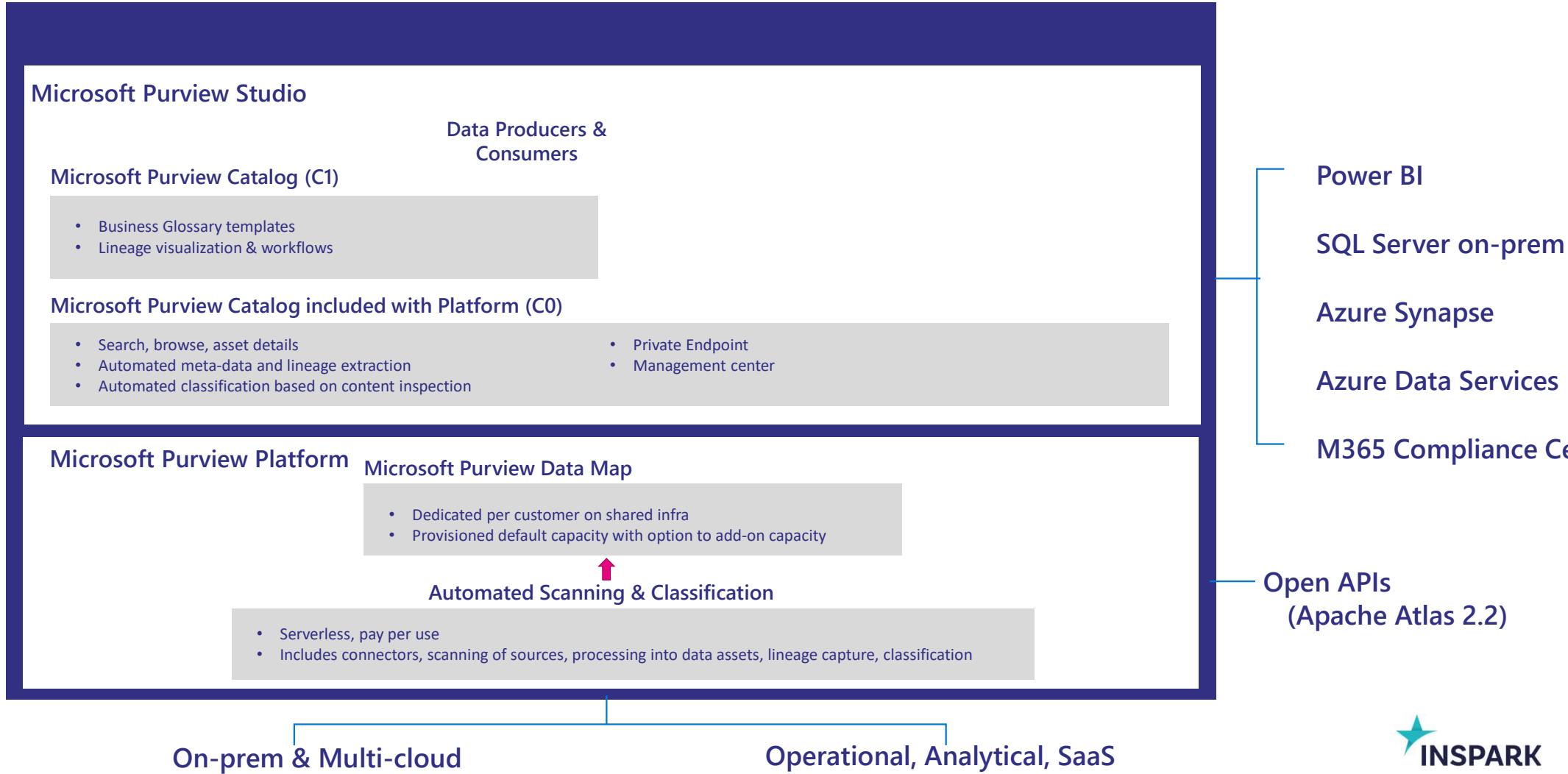
Unified Data Governance



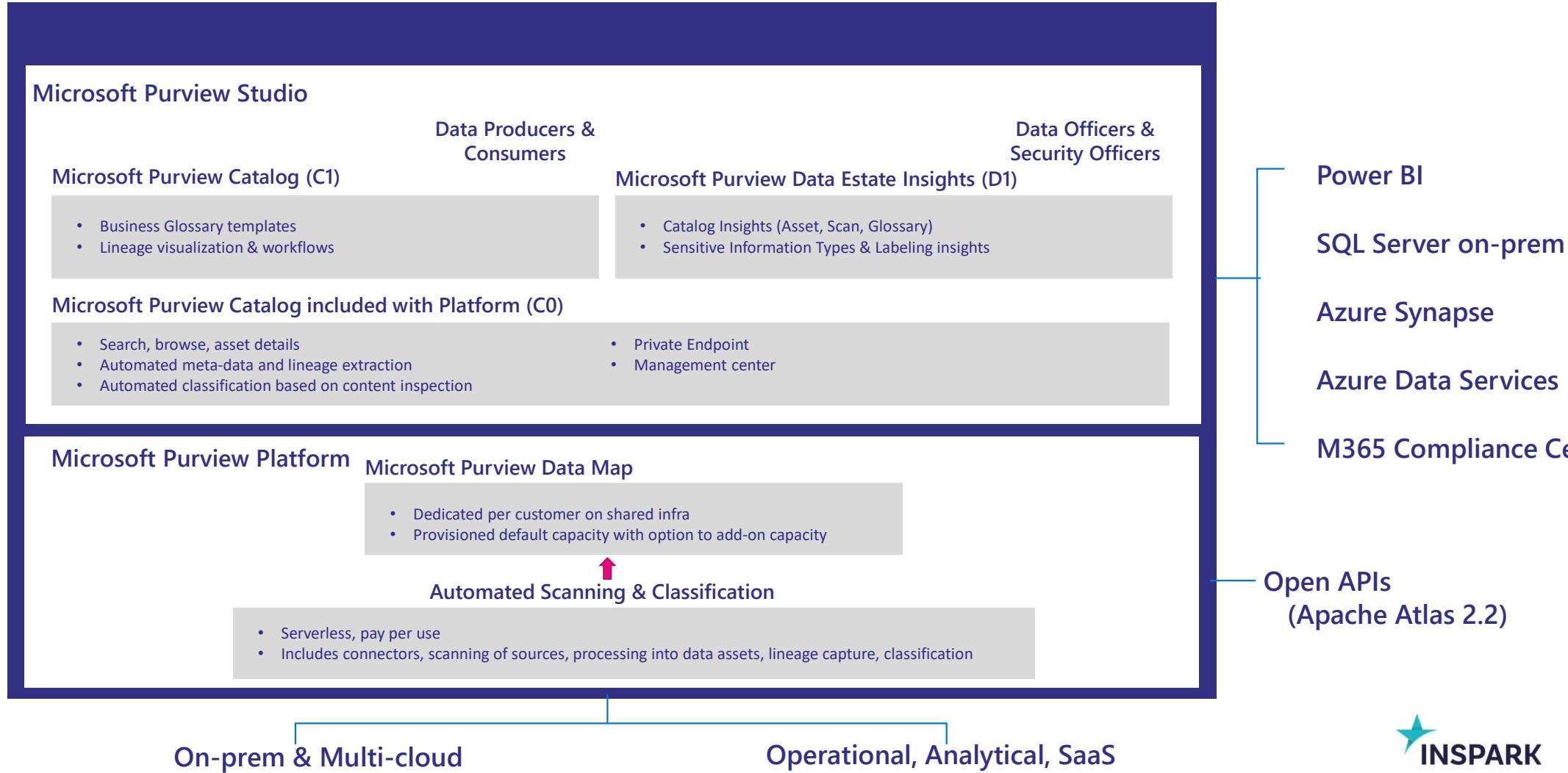
Microsoft Purview



Microsoft Purview



Microsoft Purview



Microsoft Purview

Purview Studio

The screenshot shows the Microsoft Purview InSpark studio interface. At the top, there's a navigation bar with 'Governance Portal' and 'Microsoft Purview InSpark'. On the right, there are profile details for 'Erwin.de.Kreuk@inspark...' and 'INSPARK LABS'. Below the navigation bar, there are three pink callout boxes: 'Updates', 'Accounts', and 'Notifications', each pointing to a corresponding icon in the top right corner of the interface. A large central area is labeled 'Microsoft Purview InSpark (labseuwvdvImpurviewoxgn01)' and displays metrics: 17 sources, 1,950 assets, and 56 glossary terms. It also features a 'Search catalog' bar and three main activity cards: 'Browse assets', 'Manage glossary', and 'Knowledge center'. Below these cards is a section titled 'Recently accessed' with a table of items like 'SalesLT_Customer (Current)', 'PurchaseOrderLines_PurchaseOrderLines', etc. To the right of this is a 'Links' section with links to 'Microsoft Purview overview', 'Getting started', 'Documentation', and 'Usefull Links'. Arrows from teal callout boxes on the left point to the 'Metrics' bar, the 'Search Bar', and the 'Recently Accessed Entities' table. Arrows from teal callout boxes on the right point to the 'Feedback' button, the 'Key Activities' section, and the 'Usefull Links' section.

Metrics

Search Bar

Recently Accessed Entities

Updates

Accounts

Notifications

Feedback

Key Activities

Usefull Links

Governance Portal > Microsoft Purview InSpark

Erwin.de.Kreuk@inspark... INSPARK LABS

Microsoft Purview InSpark (labseuwvdvImpurviewoxgn01)

17 sources 1,950 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.

Recently accessed My items

Name	Last update
SalesLT_Customer (Current)	2 months ago
SalesLT_Address	3 months ago
PurchaseOrderLines_PurchaseOrderLines	---
Application_Countries (Current)	6 days ago
Current	6 days ago
Application_Cities	6 days ago

Links

Microsoft Purview overview Discover the capabilities offered by Microsoft Purview and learn how to make the most of them.

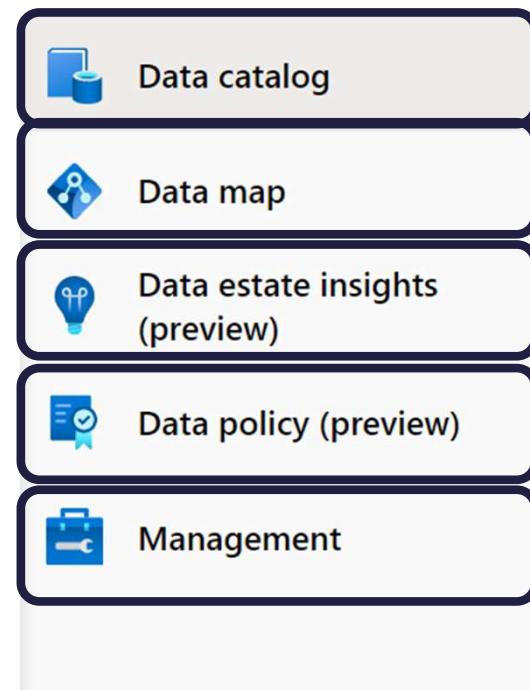
Getting started Learn about how to use Microsoft Purview Governance Portal.

Documentation Visit the documentation center for quickstarts, how-to guides, and references, etc.

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 and export Terms using csv
- Create collections, register data sources, setup Scans, Integration runtime
- 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



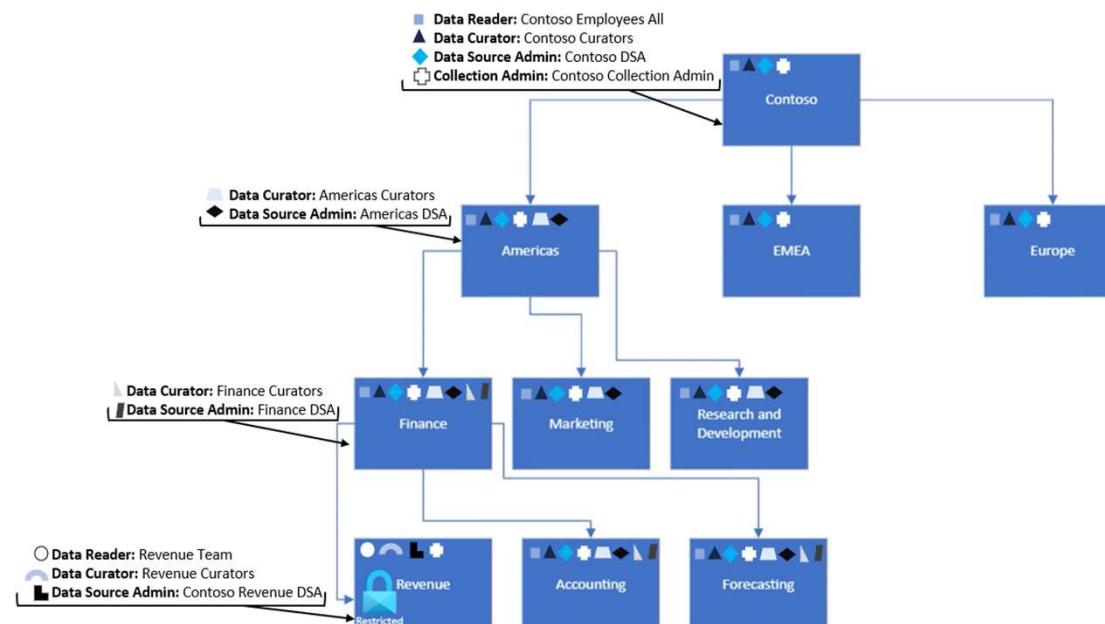
Microsoft Purview

ACCESS CONTROL



Collection Administrator

- Can edit the collection, its details, and add subcollections.
- Can also add data curators, data readers, and other Purview roles to a collection scope.
- Collection admins that are automatically inherited from a parent collection can't be removed.

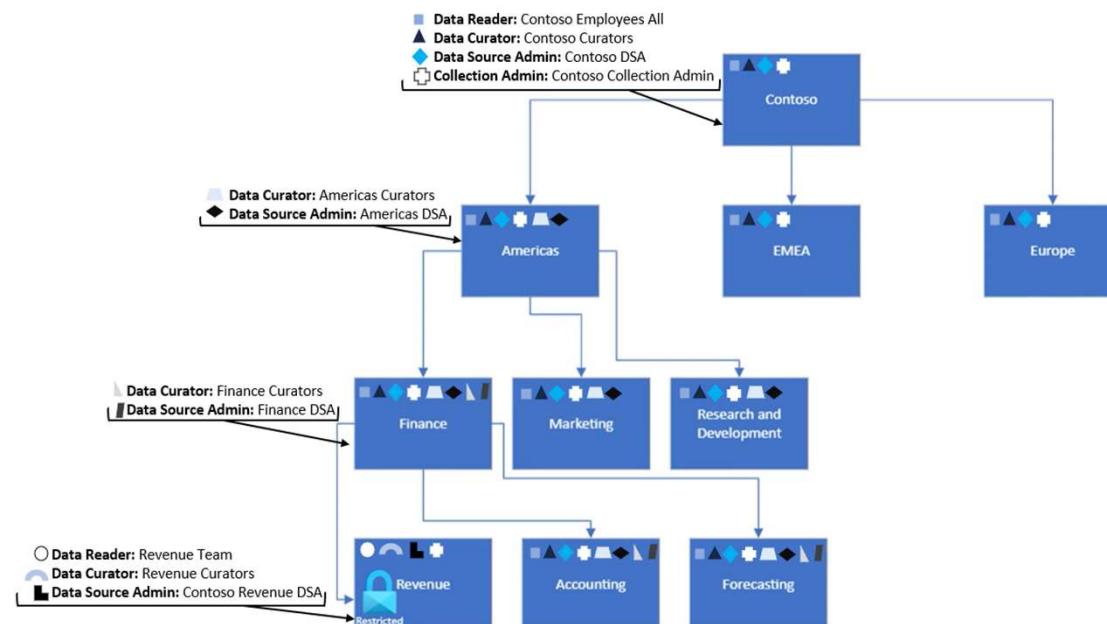


Data Curator

- Can perform create, read, modify, and delete actions on catalog data objects and establish relationships between objects.
- Can apply annotations to assets.

Collection administrator

Data curators



Microsoft Purview

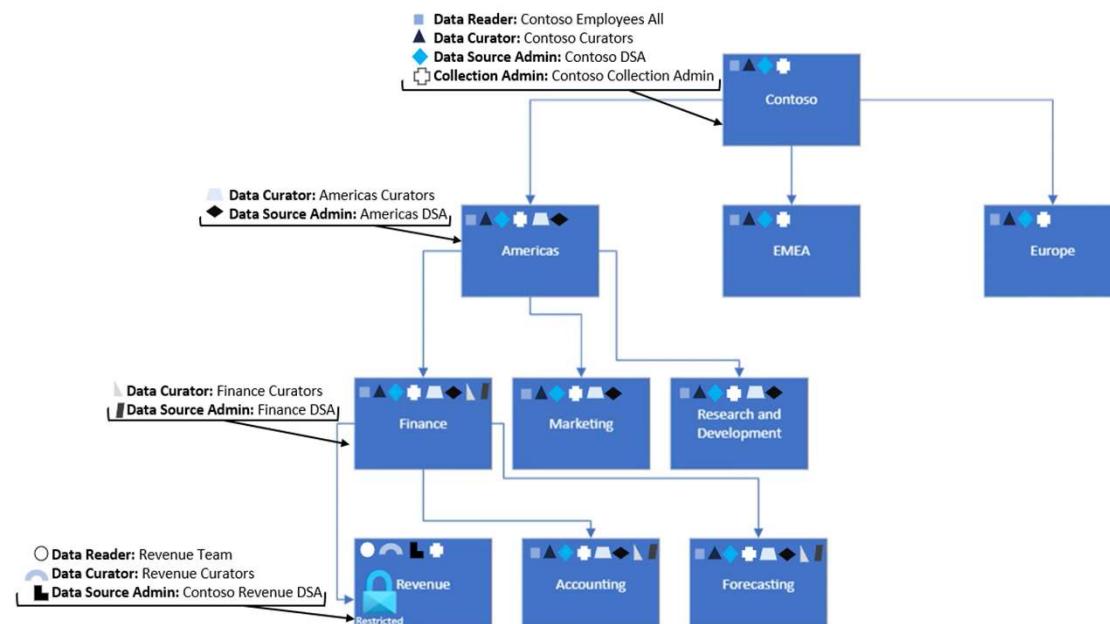
Data Reader

- Can access but can not modify catalog data objects

Collection administrator

Data curators

Data readers



Data Source administrator

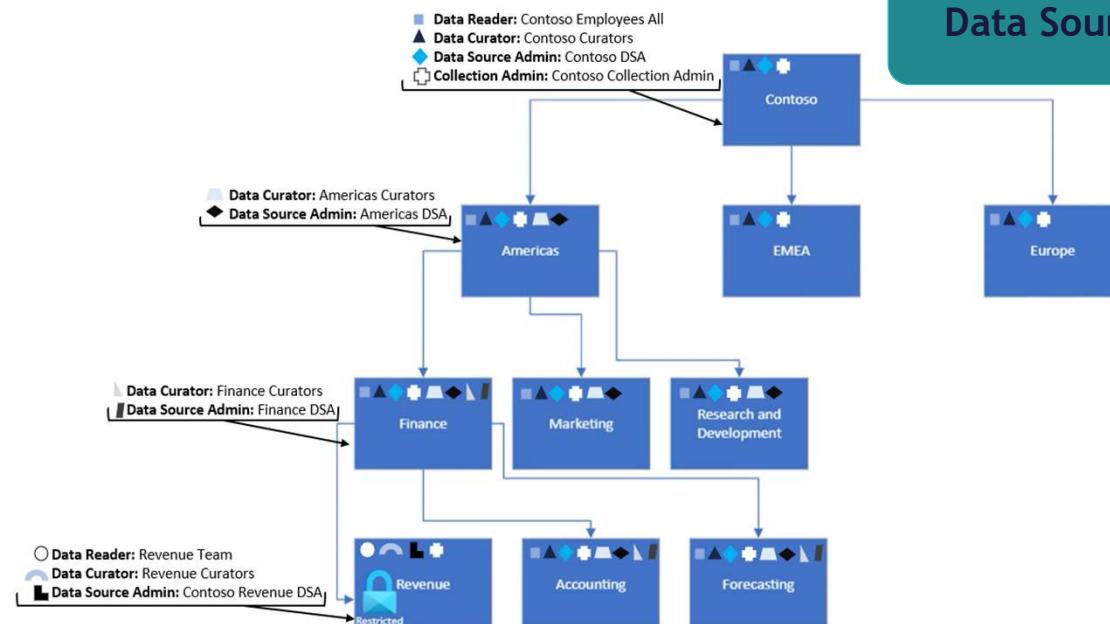
- Can manage data sources and data scans
- To add new scans Data Curator roles is needed

Collection administrator

Data curators

Data readers

Data Source Administrator



Microsoft Purview

Insights Reader (preview)

- Has read-only access to insights reports for collections where the insights reader also has at least the Data reader role.

Collection administrator

Data curators

Data readers

Data Source Administrator

Insight Reader



Microsoft Purview

Policy author

- Can view, update, and delete Microsoft Purview policies through the policy management app in Microsoft Purview
- Currently Data Source Administrator role is need to create policies



Collection administrator

Data curators

Data readers

Data Source Adminstrator

Insight Reader

Policy Author

Workflow administrator

- Has access to workflow authoring page
- Can publish workflows on collections with permissions
- User need to have Data reader to access the Microsoft Purview Portal

Collection administrator

Data curators

Data readers

Data Source Administrator

Insight Reader

Policy Author

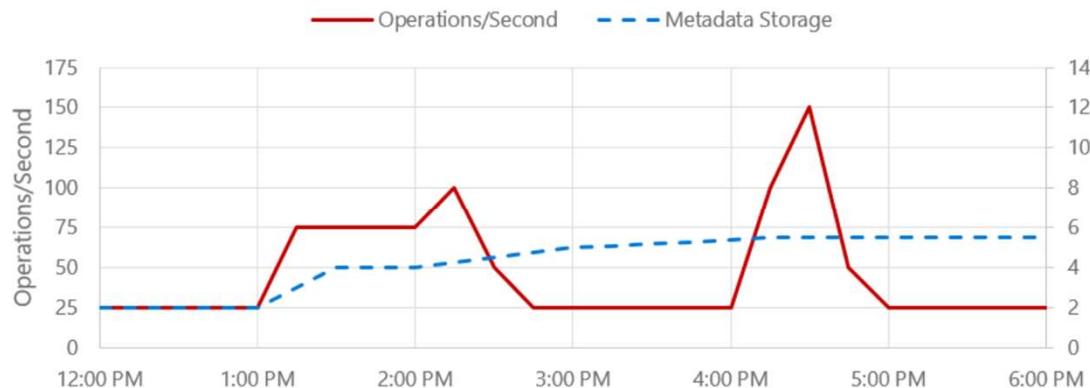
Workflow Adminstrator



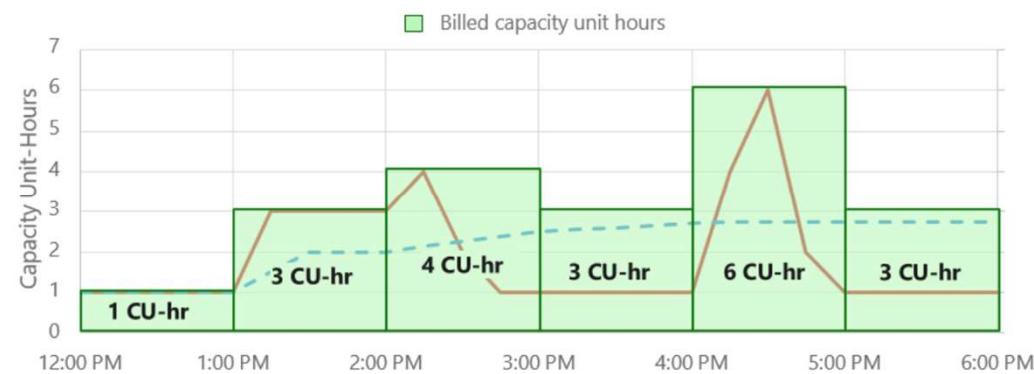
Microsoft Purview

Pricing - Data Map

- Capacity Unit
- €0.391 per 1 Capacity Unit Hour
- 1 Capacity Unit supports requests of up to 25 data map operations per second and includes storage of up to 2 GB of metadata about data assets.



Data Map Capacity Unit	Operations/Sec throughput	Storage capacity in GB
1	25	2
2	50	4
3	75	6
4	100	8
5	125	10
6	150	12
7	175	14
8	200	16
9	225	18
10	250	20



Microsoft Purview

Pricing - Automated Scanning & Classification

- Power BI Online

Free for a limited time

- SQL Server On Prem

Free for a limited time

- Other Data Sources

€0.600 per 1 vCore Hour

Microsoft Purview

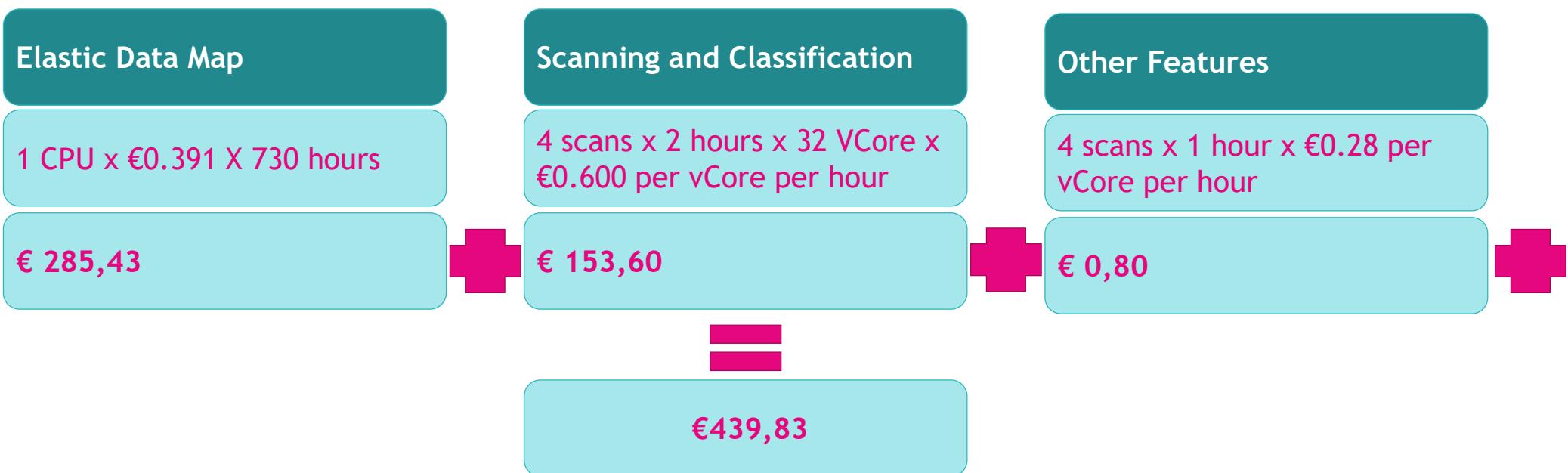
Pricing -Other features

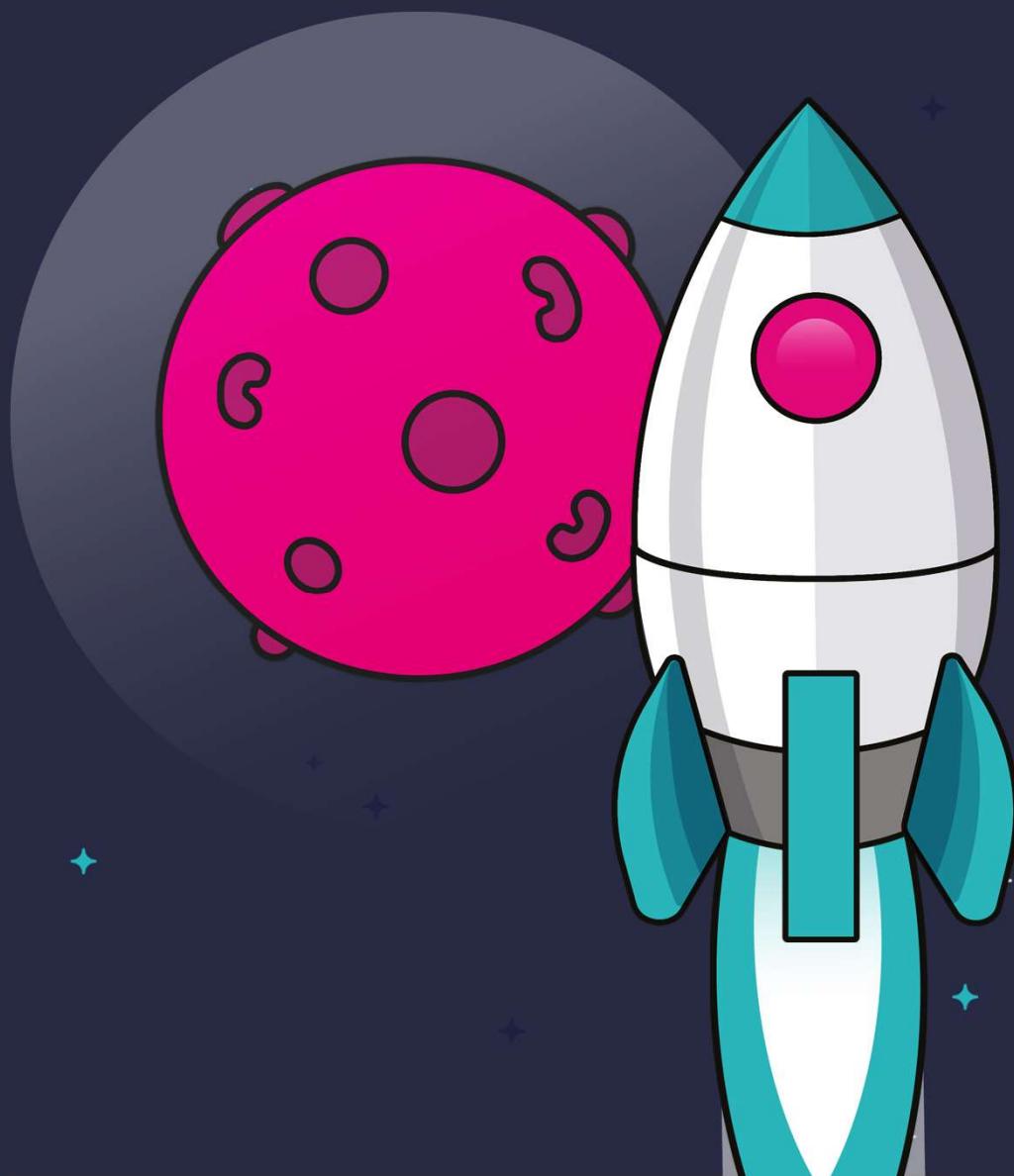
- Resource Set

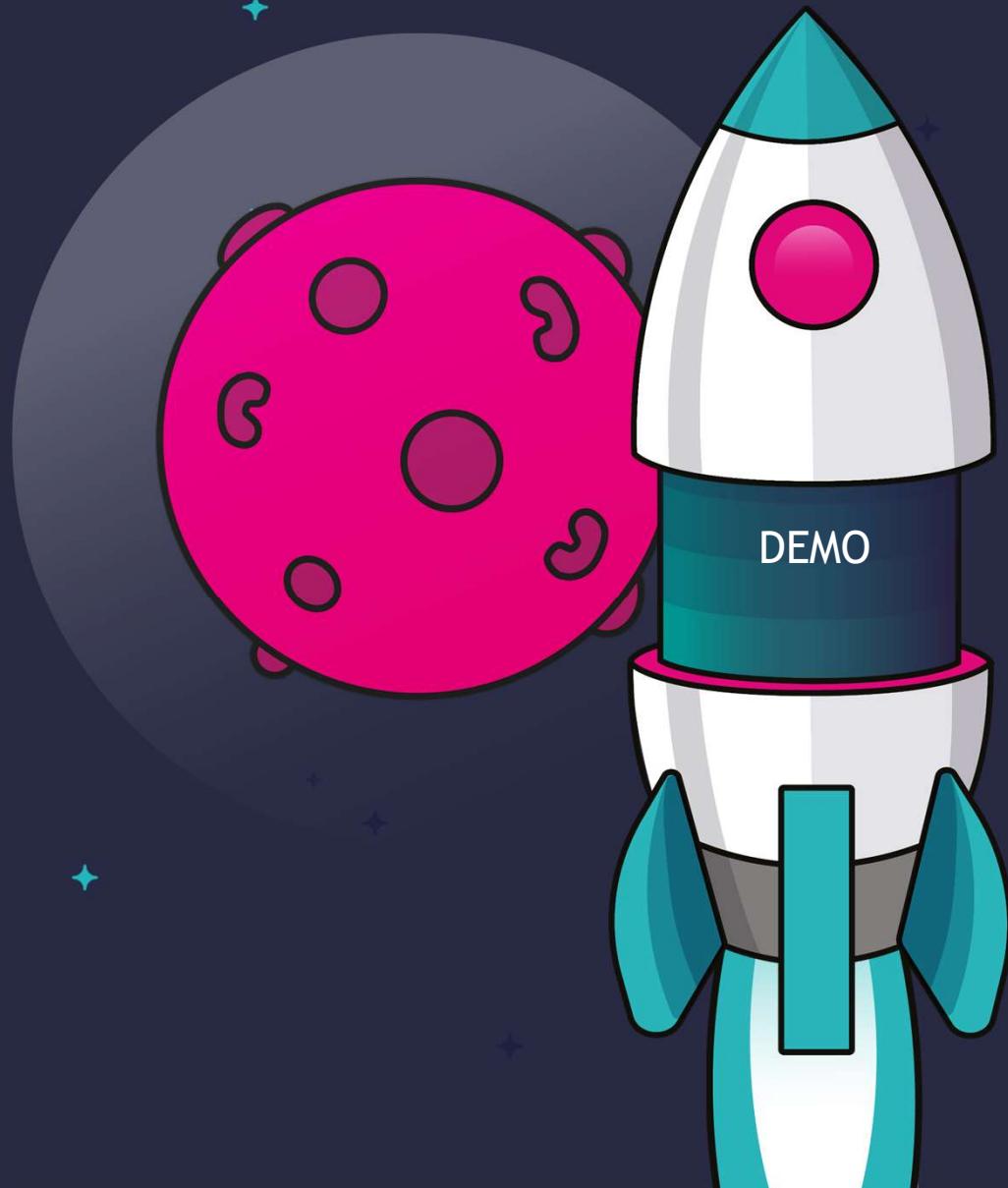
€0.20 per 1 vCore Hour

- 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

Pricing - Example







Innovate to accelerate

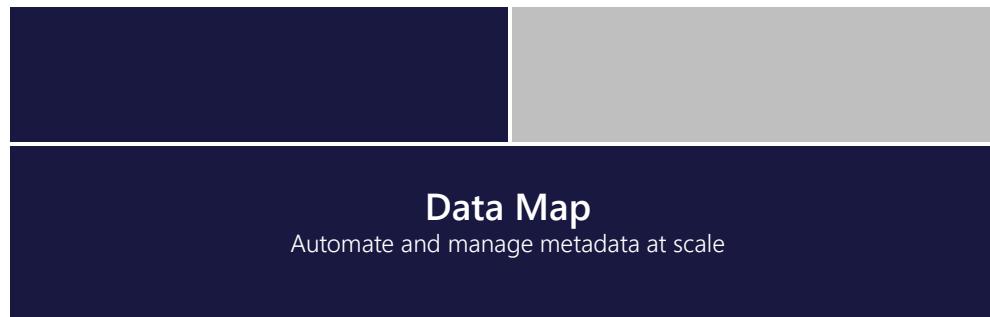
DEMO

Microsoft Purview

Unified Data Governance



Microsoft Purview

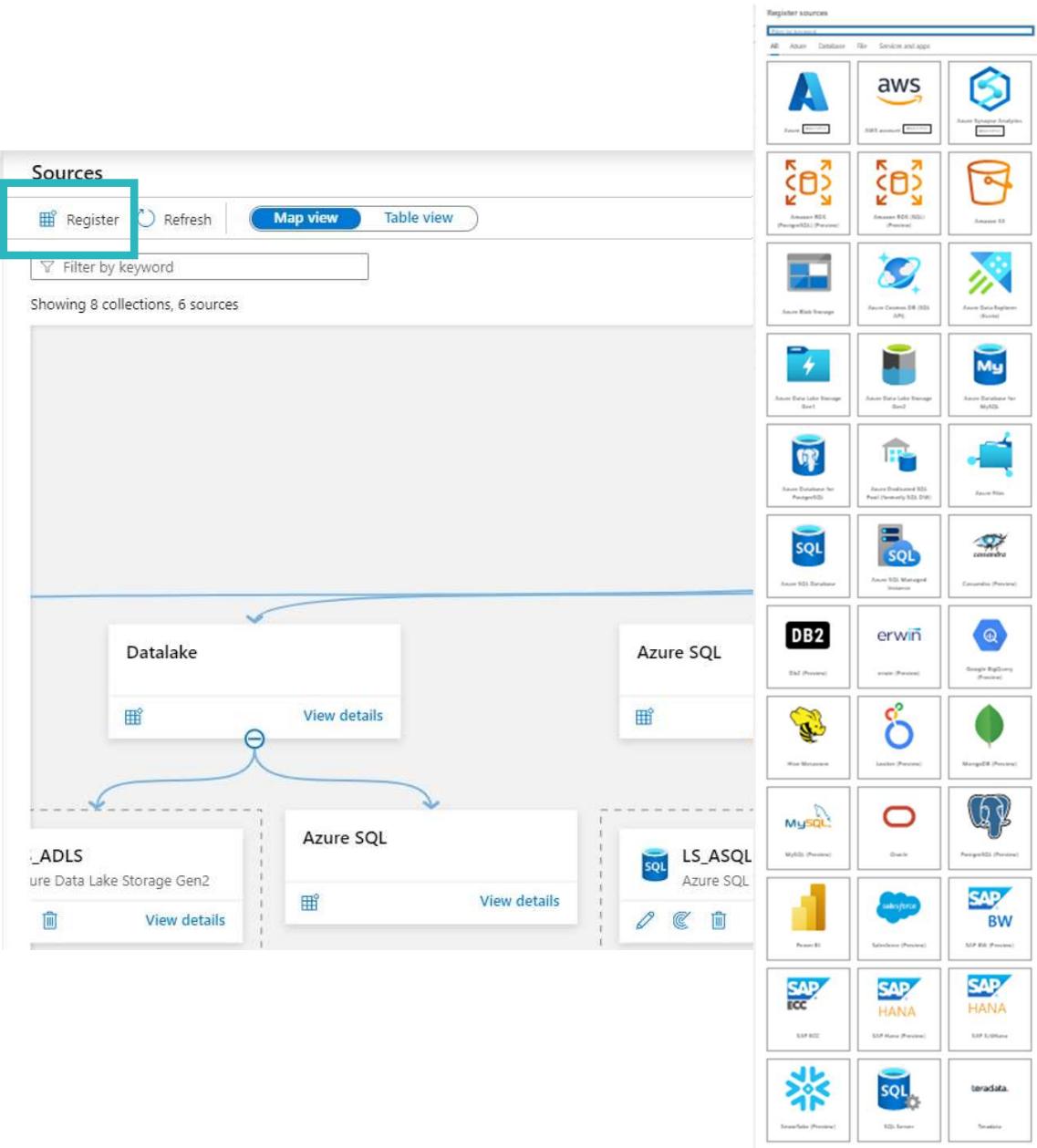


Data Map

Collections and Sources

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

Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIAL.US.ABA_ROUTING_NUMB...	ABA Routing Number
Age of an individual	MICROSOFT.PERSONAL AGE	Age of an individual
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGININA.DNI_NUM...	Argentina National Identity (DNI) Number
Australia Bank Account Number	MICROSOFT.FINANCIAL.AUSTRALIA.BANK_ACCOUNT...	Australia Bank Account Number
Australia Business Number	MICROSOFT.GOVERNMENT.AUSTRALIA.BUSINESS...	Australia Business Number
Australia Company Number	MICROSOFT.GOVERNMENT.AUSTRALIA.COMPANY...	Australia Company Number
Australia Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRALIA.DRIVERS_L...	Australia Driver's License Number
Australia Medical Account Number	MICROSOFT.GOVERNMENT.AUSTRALIA.MEDICAL...	Australia Medical Account Number
Australia Passport Number	MICROSOFT.GOVERNMENT.AUSTRALIA.PASSPORT...	Australia Passport Number
Australia Tax File Number	MICROSOFT.GOVERNMENT.AUSTRALIA.TAX_FILE...	Australia Tax File Number
Austria Driver's License Number	MICROSOFT.GOVERNMENT.AUSTRIA.DRIVERS.LC...	Austria Driver's License Number
Austria Identity Card	MICROSOFT.GOVERNMENT.AUSTRIA.IDENTITY.CA...	Austria Identity Card



Data Map

Scans

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

Sources > LS_ADLS > Weekly_Scan run history

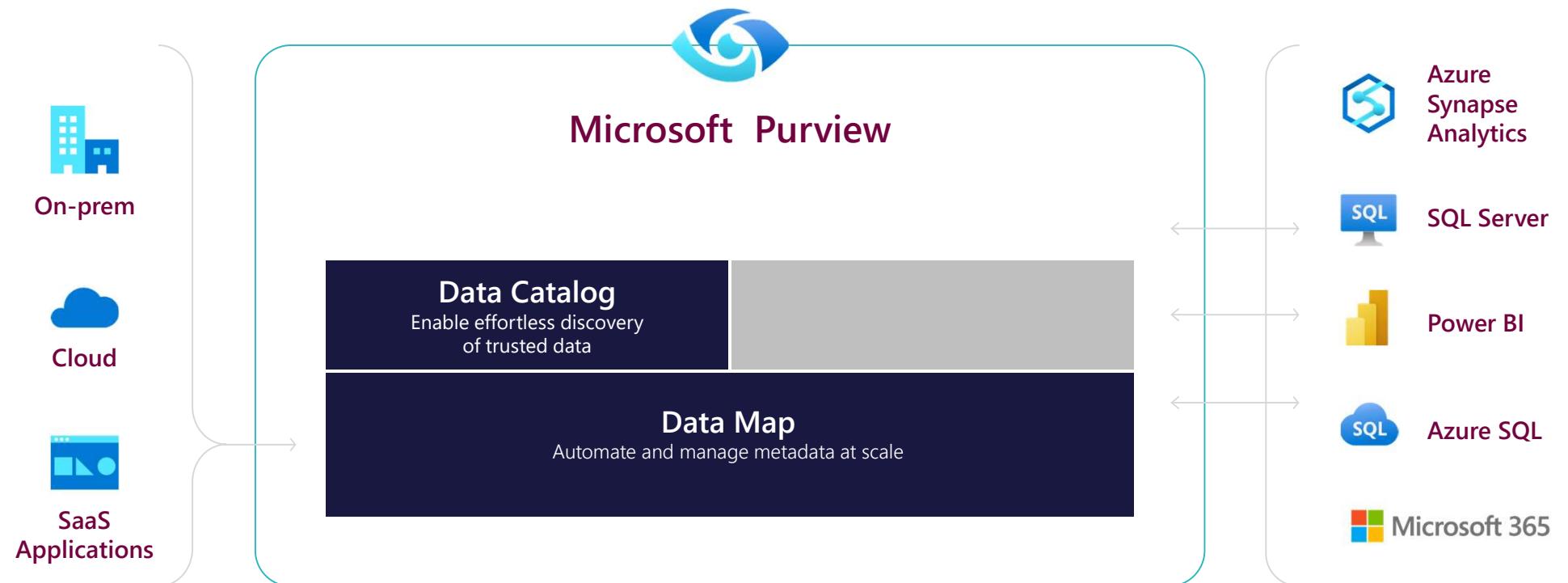
Weekly_Scan run history

 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 11:14 AM
 Completed	 Scheduled	172	432,117	01/11/21 11:00 AM	01/11/21 11:14 AM
 Completed	 Scheduled	199	431,156	01/04/21 11:00 AM	01/04/21 11:15 AM
 Completed	 Manual	171	429,874	12/30/20 02:47 PM	12/30/20 02:57 PM
 Completed	 Manual	192	429,496	12/29/20 02:57 PM	12/29/20 03:08 PM

Microsoft Purview

Unified Data Governance



Data Catalog

Search

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

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. Below it, 'Your recent searches' show 'customer' and 'customers'. 'Search suggestions' include 'CustomerCategories', 'Customer', 'CustomerKey', and 'CustomerCategoryID'. 'Asset suggestions' list several data flows and tables, such as 'Customer_Customer_DataFlow', 'CustomerType_CustomerType_DataFlow', 'Customer', and 'Customer'. At the bottom right of the search panel is a 'View search results' button. The top right of the screen displays the title 'Microsoft Purview InSpark' with a subtitle '(labseuvdvlmpurviewoxgn01)', and statistics: 17 sources, 1,950 assets, and 56 glossary terms. To the right of the search panel is a 3D diagram illustrating data lineage and relationships between various data assets.



Data Catalog

Assets

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

The screenshot illustrates the Azure Data Catalog interface, specifically focusing on the 'Azure SQL Database' asset.

Main Catalog View: The top section shows a grid of various Azure data assets. The 'Azure SQL Database' asset is highlighted with a red box. The grid includes other services like Azure Blob Storage, Azure Cosmos DB, Azure Data Factory, Azure Data Lake Storage Gen2, Azure Storage Account, Azure Synapse Analytics, Hive Metastore, and Power BI.

Name	Type
Azure Blob Storage	Container
Azure Cosmos DB	Database
Azure Data Factory	Integration Service
Azure Data Lake Storage Gen2	Storage
Azure Data Share	Share
Azure SQL Database	Database
Azure SQL Server	Server
Azure Storage Account	Storage
Azure Synapse Analytics	Analytics
Hive Metastore	Metastore
Power BI	Analytics

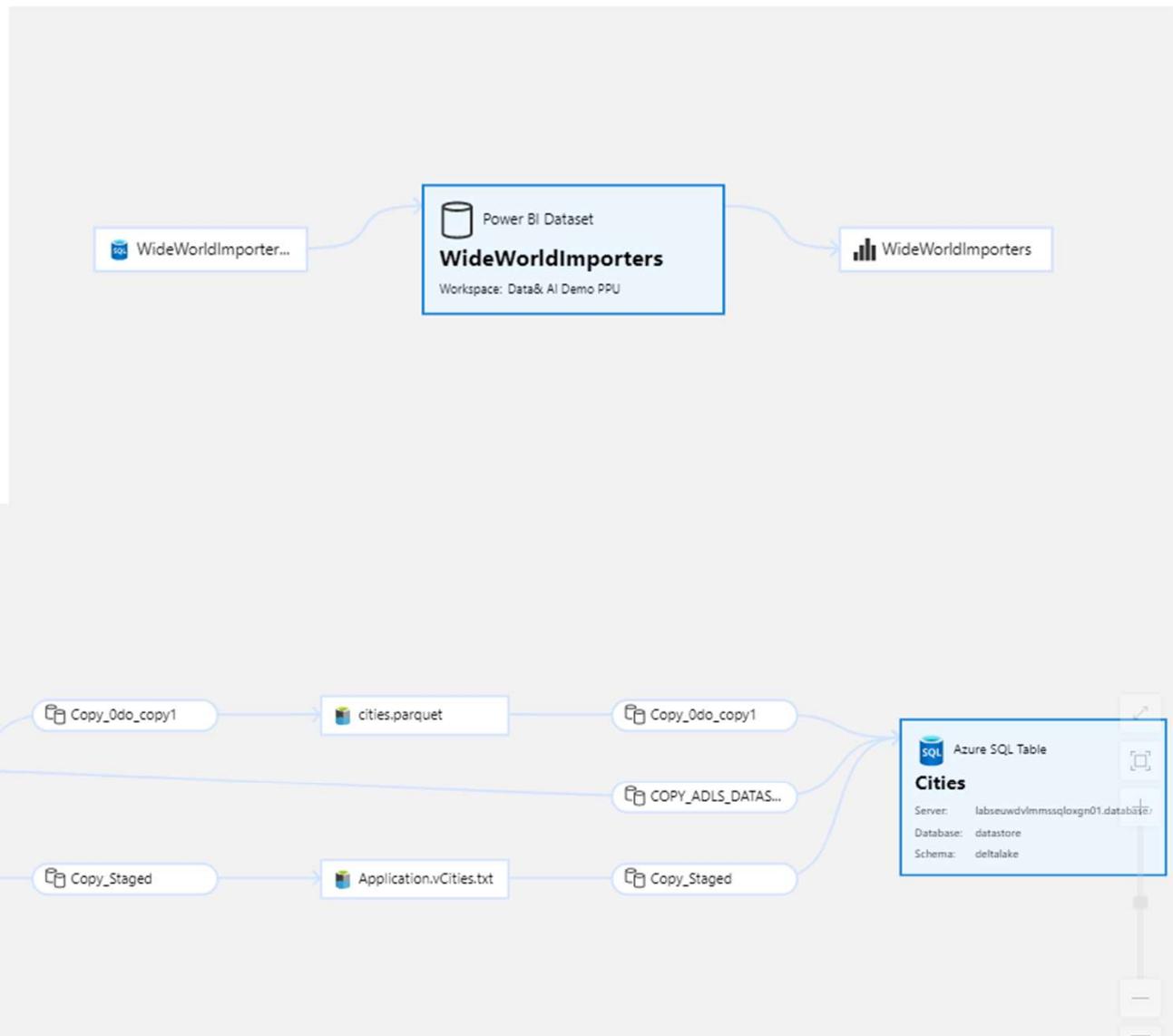
Detail View: Below the main catalog, three detailed views are shown:

- Browse assets > Azure SQL Database:** Shows a list of databases: Adventureworks, Datastore, Datastore, datastore, oxygenconfigdb, test_purview, and WideWorldImporters.
- Browse assets > Azure SQL Database > WideWorldImporters:** Shows the schema of the WideWorldImporters database, including Application, dbo, Purchasing, Sales, and Warehouse.
- Browse assets > Azure SQL Database > WideWorldImporters > Tables:** Shows a list of tables: PurchaseOrderLines, PurchaseOrders, SupplierCategories, SupplierCategories_Archive, Suppliers, Suppliers_Archive, and SupplierTransactions.

Data Catalog

Lineage

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

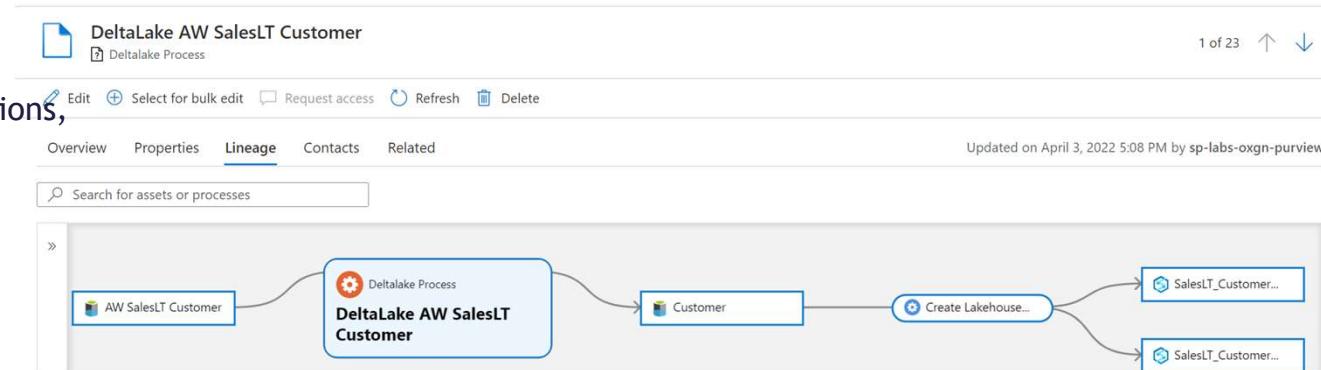


Data Catalog

Custom Lineage

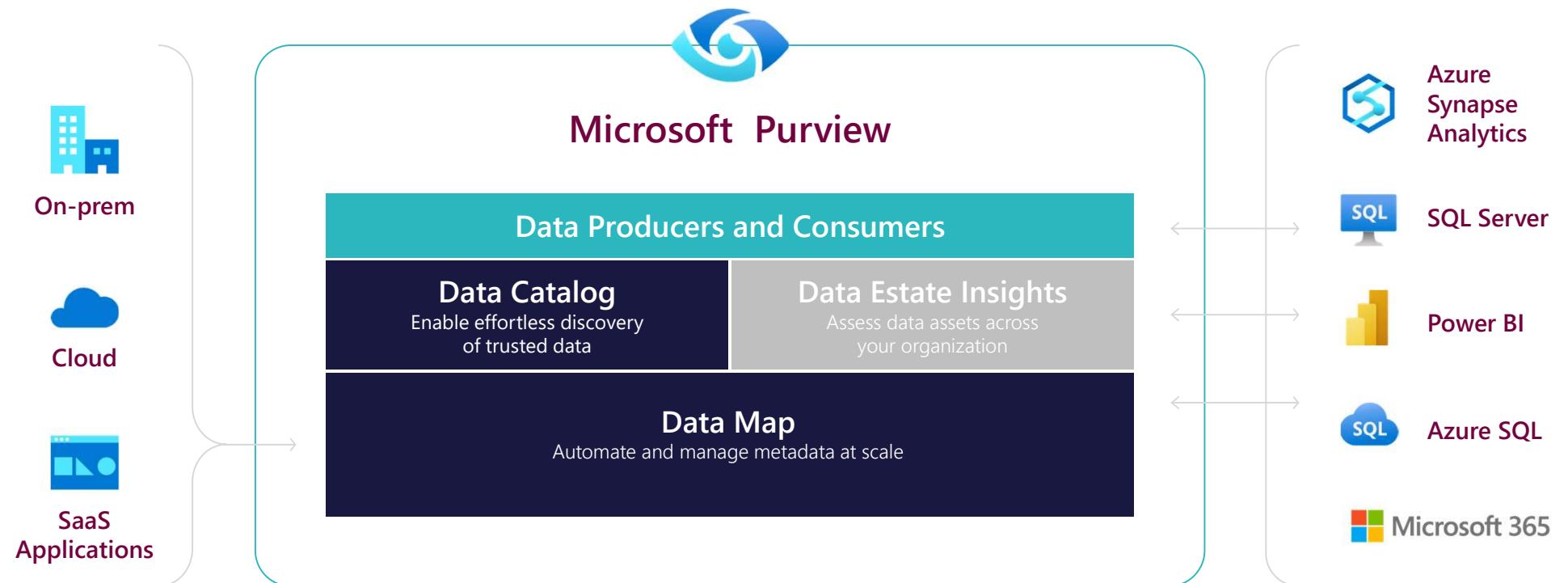
- Semantic search and browse
- Business glossary and workflows
- Data lineage with sources, owners, transformations, and lifecycle
- Create custom Lineage with Apache Spark 2.2

<https://github.com/wjohnson/pyapacheatlas>



Microsoft Purview

Unified Data Governance



Data Estate Insight (preview)

Metrics

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

Data stewardship (preview)

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

Report generated on May 24, 2022 at 1:13 PM ⓘ

Health (preview) ⟲

- Data stewardship
- Inventory and ownership (preview)
- Assets
- Curation and governance (previe...)
- Glossary
- Classifications
- Sensitivity labels

Data estate health

Collection : (Root) Microsoft Purview InSpark

Collection	Assets	With sensitiv...	Fully curated	Owner assign...	No classificat...	New	Deleted
[Assets]	28,288	✓ 0%	✗ 0%	✗ 0%	✗ 100%	0	2
Datalake	489	✓ 13%	✗ 0%	✗ 0%	✗ 87%	0	1
Blob	19	✓ 0%	✗ 0%	✗ 0%	✗ 100%	0	0
West Europe	331	✓ 8%	✗ 0%	✗ 0%	✗ 92%	0	0
DemoEK	328	✓ 11%	✗ 0%	✗ 0%	✗ 89%	38	2
Azure SQL	572	✓ 32%	✗ 0%	✗ 0%	⚠ 68%	111	0
Cosmos	5	✓ 0%	✗ 0%	✗ 0%	✗ 100%	0	0
PowerBI	136	✓ 0%	✗ 0%	✗ 0%	✗ 100%	0	0
Synapse	116	✓ 78%	✗ 0%	✗ 0%	✓ 22%	69	12

Assets (preview)

View key metrics about your data estate's assets.

Report generated on May 24, 2022 at 1:13 PM ⓘ

Health (preview) ⟲

- Data stewardship
- Inventory and ownership (preview)
- Assets
- Curation and governance (previe...)
- Glossary
- Classifications
- Sensitivity labels

Total assets

30K ⚡

Microsoft Purview InSpark

No classifications

30K (99%) ⚡

No owner assigned

30K (100%) ⚡

New (Last 30 days)

218 ⚡

Deleted (Last 30 days)

17 ⚡



Data Policy (preview)

Data Owner

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

The screenshot shows the 'Access Control Policy' creation page in the Azure Data Catalog. The left sidebar includes 'Data catalog', 'Data map', 'Data estate insights (preview)', 'Data policy (preview)' (selected), and 'Management'. The main area starts with a preview notice: 'This PREVIEW feature is licensed to you as part of your Azure Subscription. By proceeding you acknowledge the Preview Terms and Privacy Statement.' Below this is the 'Access Control Policy' section with fields for 'Name *' (Policy Name) and 'Description *' (Describe the policy). A 'Policy statements' section allows creating up to 5 statements. An example statement is shown: 'Allow Read on data asset finance to group non-FTEs'. A button '+ New policy statement' is present. A dropdown menu for 'action' is open, showing options: 'Allow', 'Read', 'Modify', 'SQL Performance Monitoring', and 'SQL Security Auditing'. Below this is the 'Data resources' section, which asks to select a type of data resource to apply the policy statement to. It lists 'Data sources' (systems where data is stored) and 'Assets' (individual objects stored within the data catalog). The 'Data sources' section includes a 'Data source type' dropdown with options: 'Select data source type', 'Azure Subscription', 'Azure Resource Group', 'Azure Blob Storage', 'Azure Data Lake Storage Gen2', and 'Azure SQL Server'. The 'Assets' section includes a 'Data source type' dropdown with 'Azure Data Lake Storage Gen2' selected, and a 'Name' field containing 'labseuwvd1mdloxgn02'. The bottom right corner features the INSPARK logo.

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 *
Policy Name

Description *
Describe the policy.

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 *

- Select 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 *

- Azure Data Lake Storage Gen2

Name *

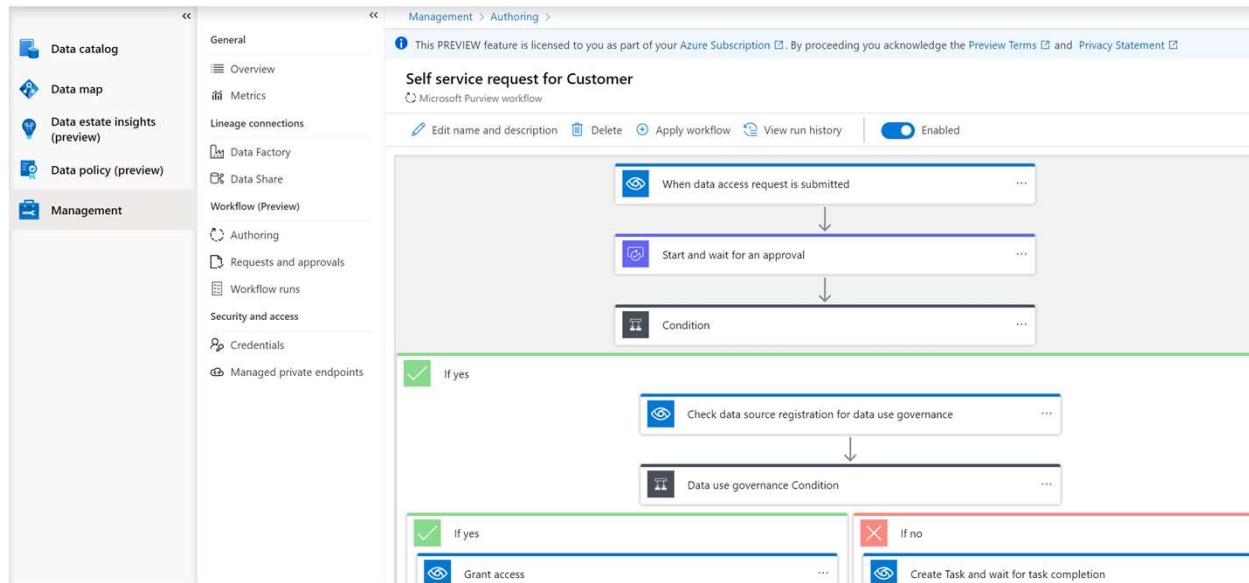
labseuwvd1mdloxgn02

INSPARK

Data Policy (preview)

Self Service

- Workflow Governance
 - Data Access
- Workflow Data Catalog
 - Create/Delete/Import/Update Glossary Terms





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

The screenshot shows the Microsoft Purview integration configuration page for an Azure Synapse workspace. The left sidebar lists various configuration sections: Analytics pools (SQL pools, Apache Spark pools, Data Explorer pools), External connections (Linked services, Microsoft Purview), Integration (Triggers, Integration runtimes), Security (Access control, Credentials, Managed private endpoints), Configurations + libraries (Workspace packages, Data flow libraries, Apache Spark configurations), and Source control (Git configuration). The main content area is titled "Microsoft Purview" and displays the following information:

- Purview account:** .01
- Default Purview integration capabilities:** Data Catalog - Discover enterprise data assets using Microsoft Purview powered search
- Additional Purview integration capabilities and status:**

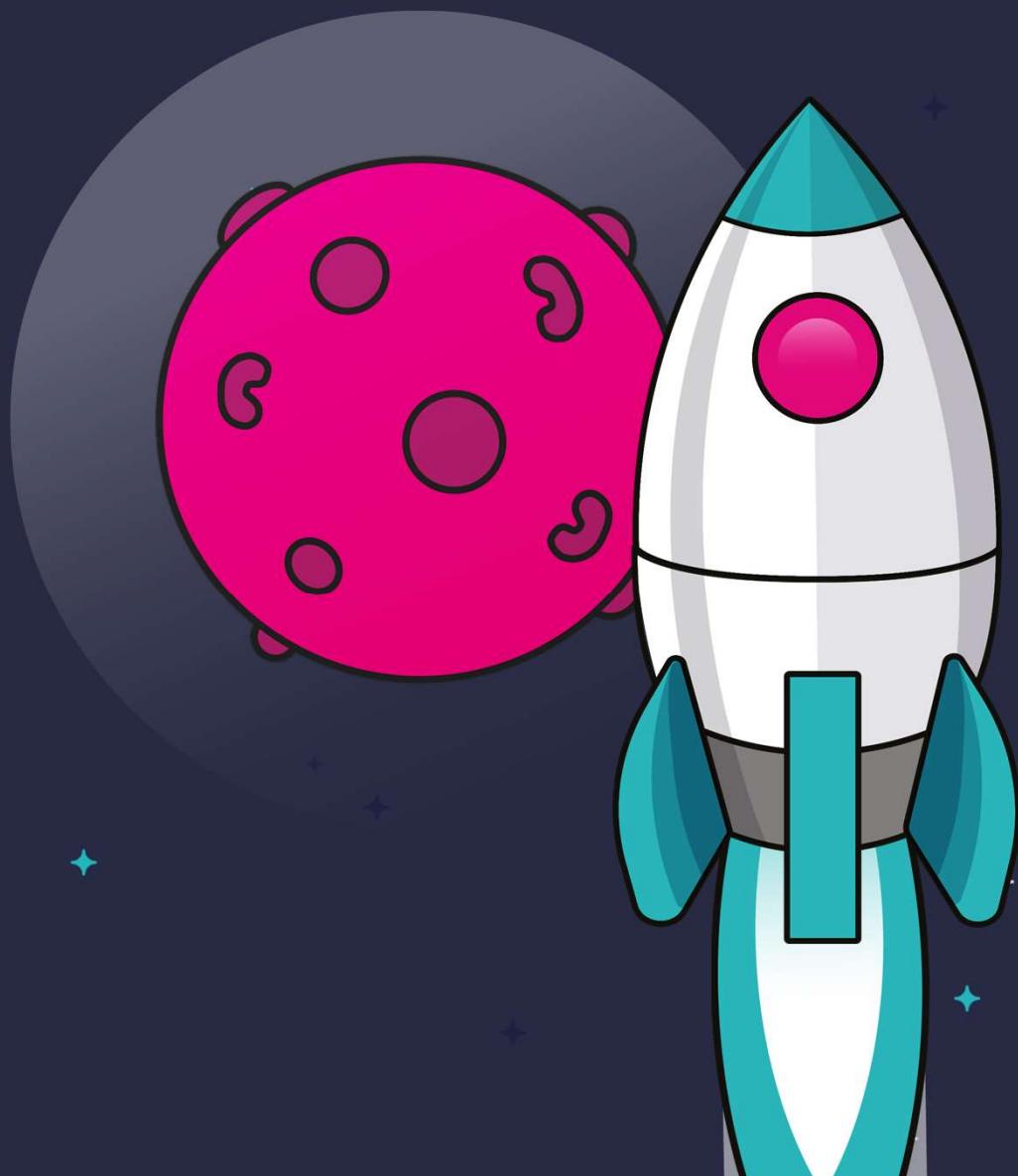
Integration	Status
Data Lineage - Synapse Pipeline	Connected
- Lineage status details:**

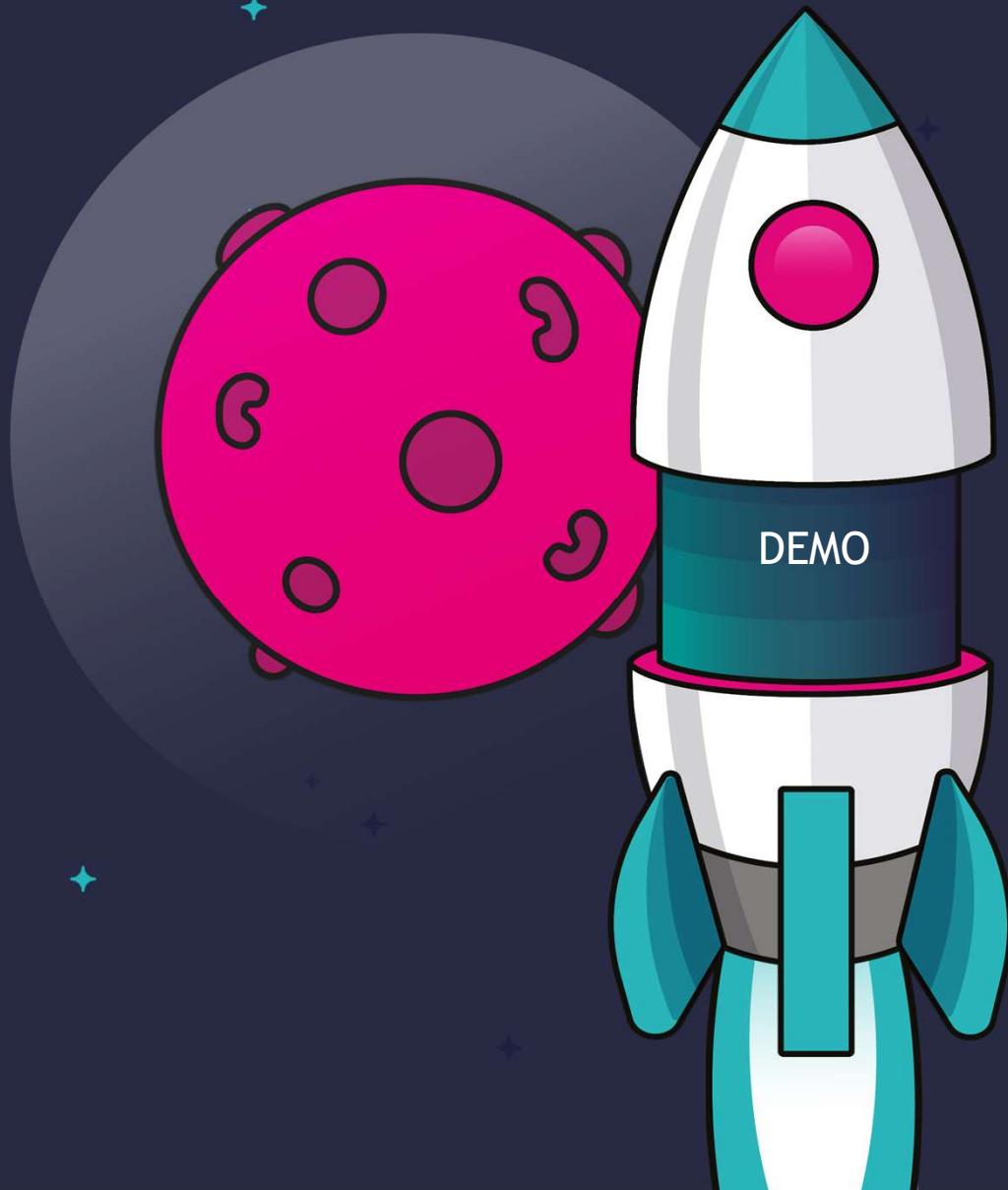
Status	Result
Succeeded	✓
- Lineage status details:**

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](#)



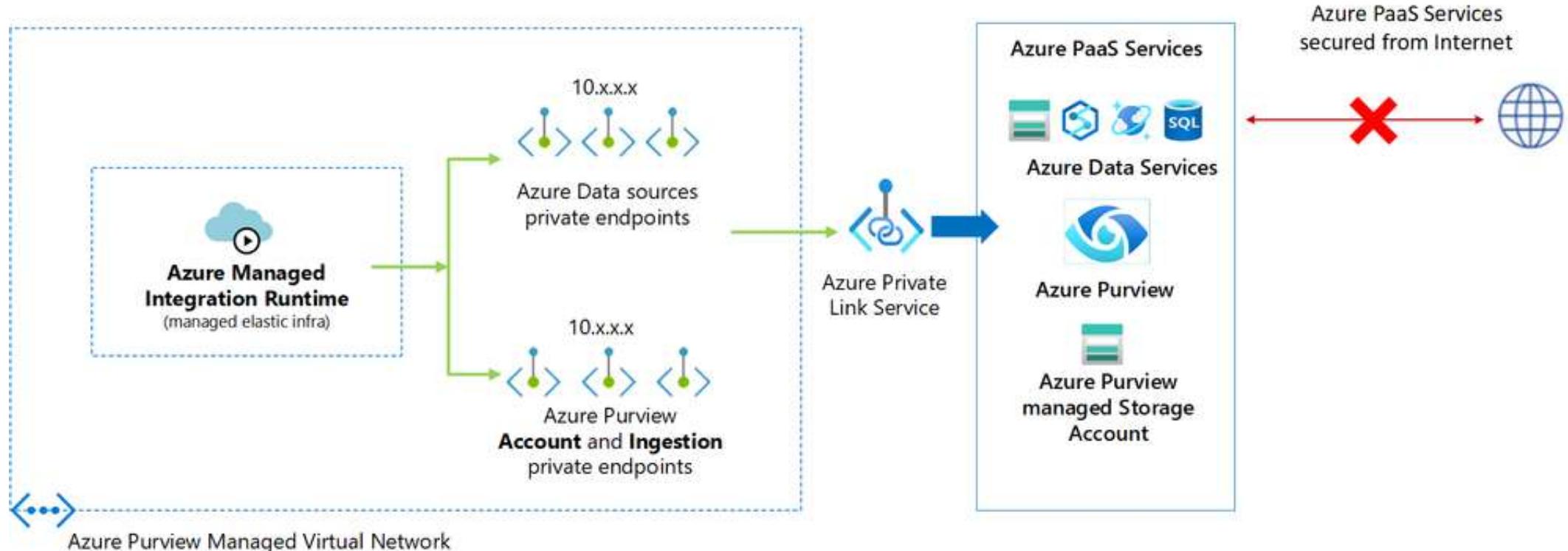


Innovate to accelerate

DEMO

Microsoft Purview

Managed Vnets

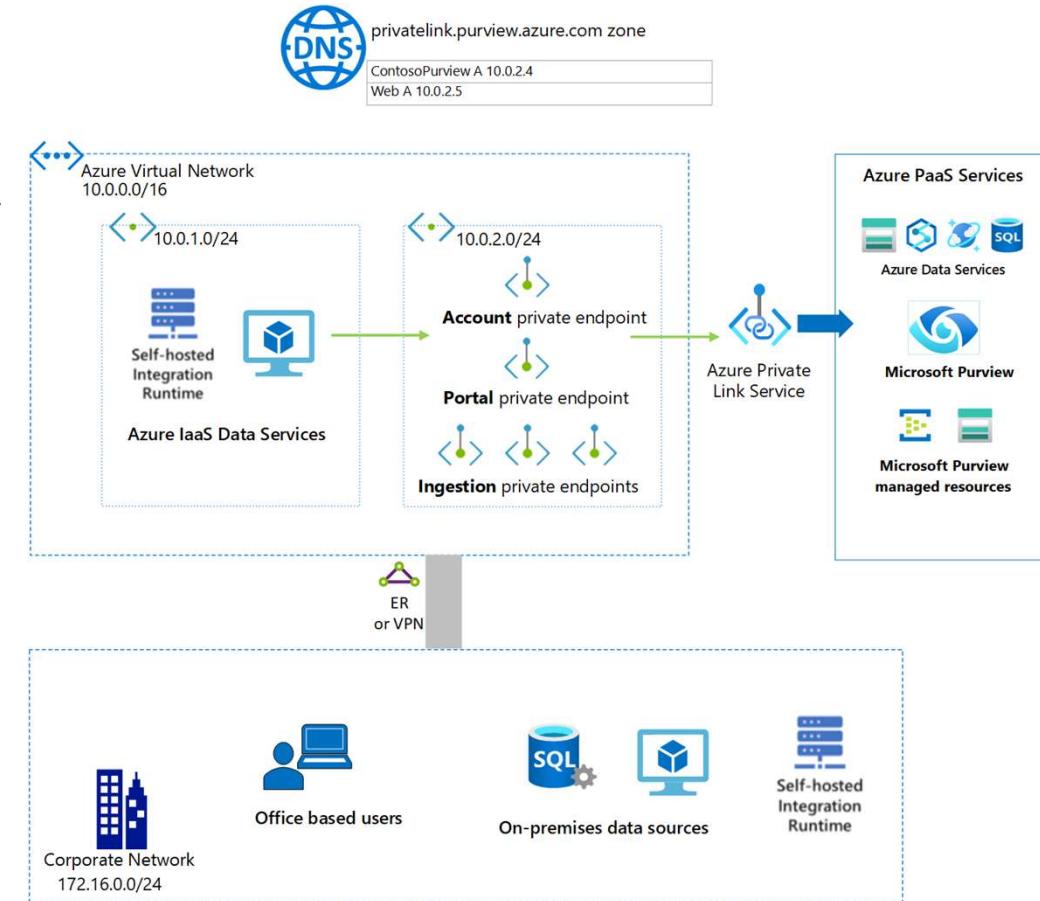


<https://docs.microsoft.com/en-us/azure/purview/catalog-managed-vnet>

Microsoft Purview

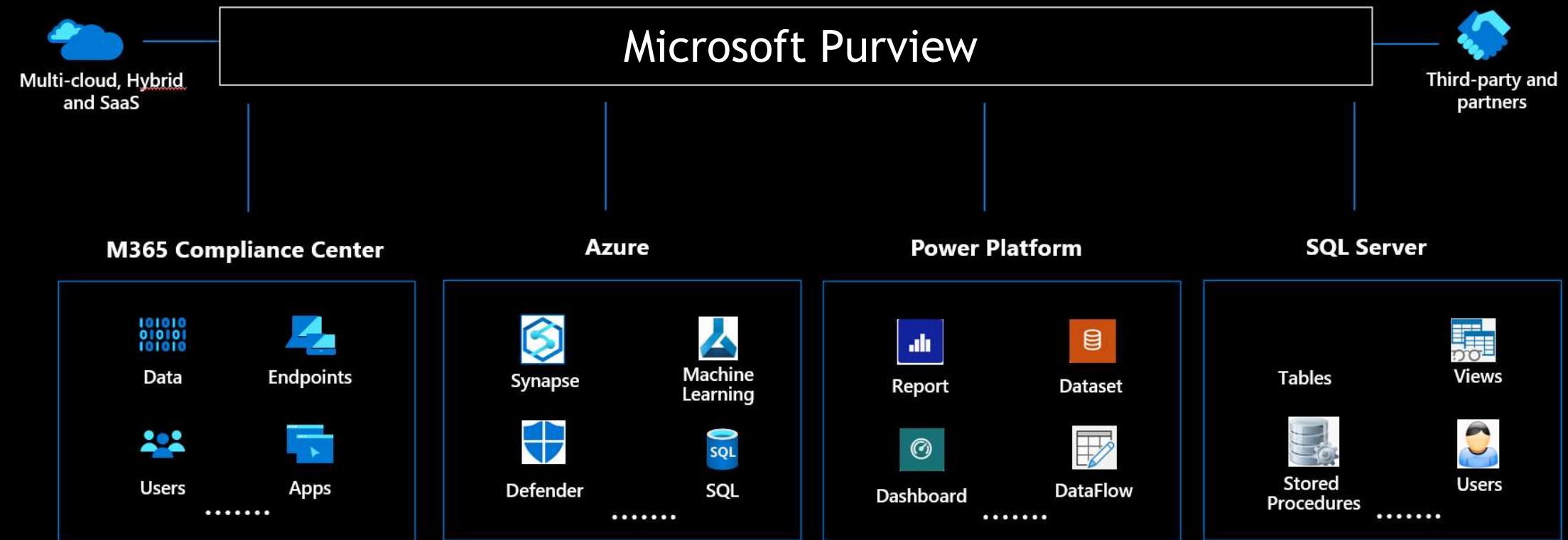
Private Endpoint

- Connect to your Microsoft Purview and scan data sources privately and securely
- Connect privately and securely to your Microsoft Purview account
- Scan data source securely using Managed Virtual Network



<https://docs.microsoft.com/en-us/azure/purview/catalog-private-link>

Data Governance Opportunity



Evaluation



<https://evals.datagrillen.com/evals.aspx>



Questions?



Erwin de Kreuk

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



@erwindekreuk



<https://erwindekreuk.com>

