



# Azure Purview

Microsoft's answer to Data Governance and Data Lineage

## Erwin de Kreuk



@erwindekreuk



<https://erwindekreuk.com>





# Azure Purview

Microsoft's answer to Data Governance and Data Lineage



@erwindekreuk



<https://erwindekreuk.com>



# We Are InSpark

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



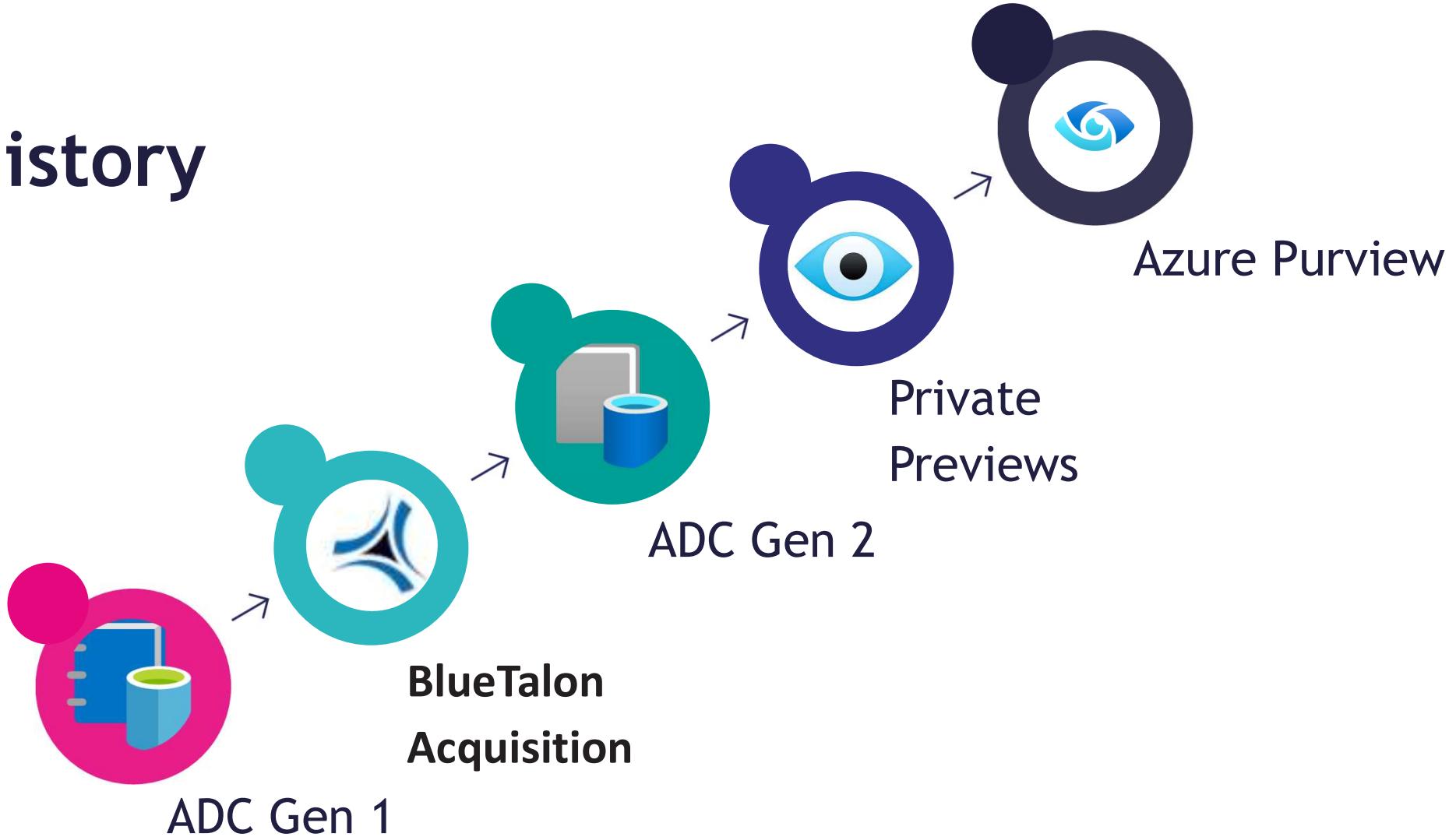
# Azure Purview



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



# History



# Data governance is becoming increasingly interdisciplinary



# Data governance is becoming increasingly interdisciplinary



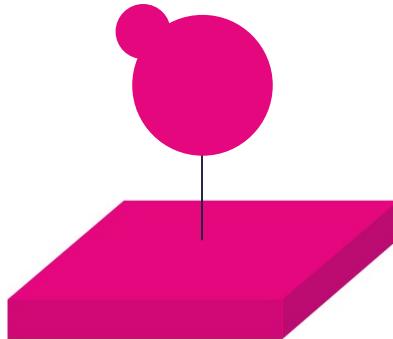
## Chief Data Officer

Data Management  
Data Stewardship  
Data Engineering  
Data Science  
Data Analyst

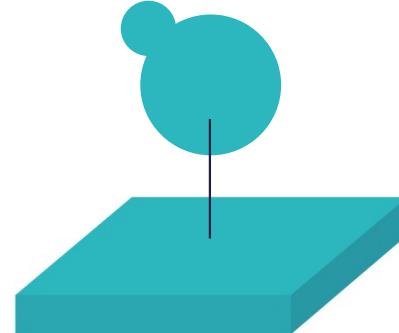
**What data do I have?  
Where did the data originate?  
Can I trust it?  
What's my exposure to risk?  
Is my usage compliant?  
How do I control access & use?  
What is required by regulation X?**

Data Compliancy  
Data Access  
Data Security  
Data Privacy

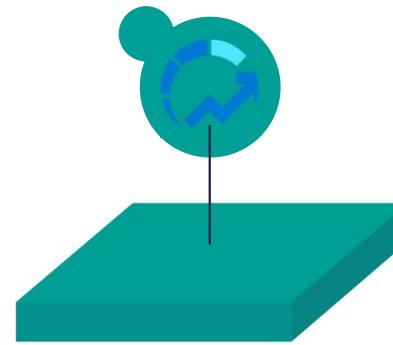
# Elements of successful data governance



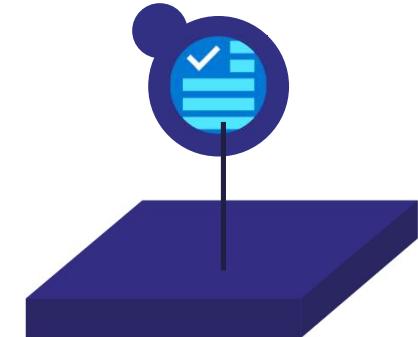
Manage growing  
data landscape



Overcome  
operational silos



Increase data  
agility



Comply with  
industry  
regulations

## Data Insights

- Assess data usage across your organization

## Data Catalog

- Enable effortless discovery for data consumers
  - Catalog search
  - Business glossary
  - Asset normalization

## Data Map

- Automate and manage metadata at scale
  - Elastic data map
  - Resource sets
  - Scans and ingestion
  - Data lineage



# Azure Purview

Unified data governance to maximize the business value of data





## Reimagine data governance in the cloud

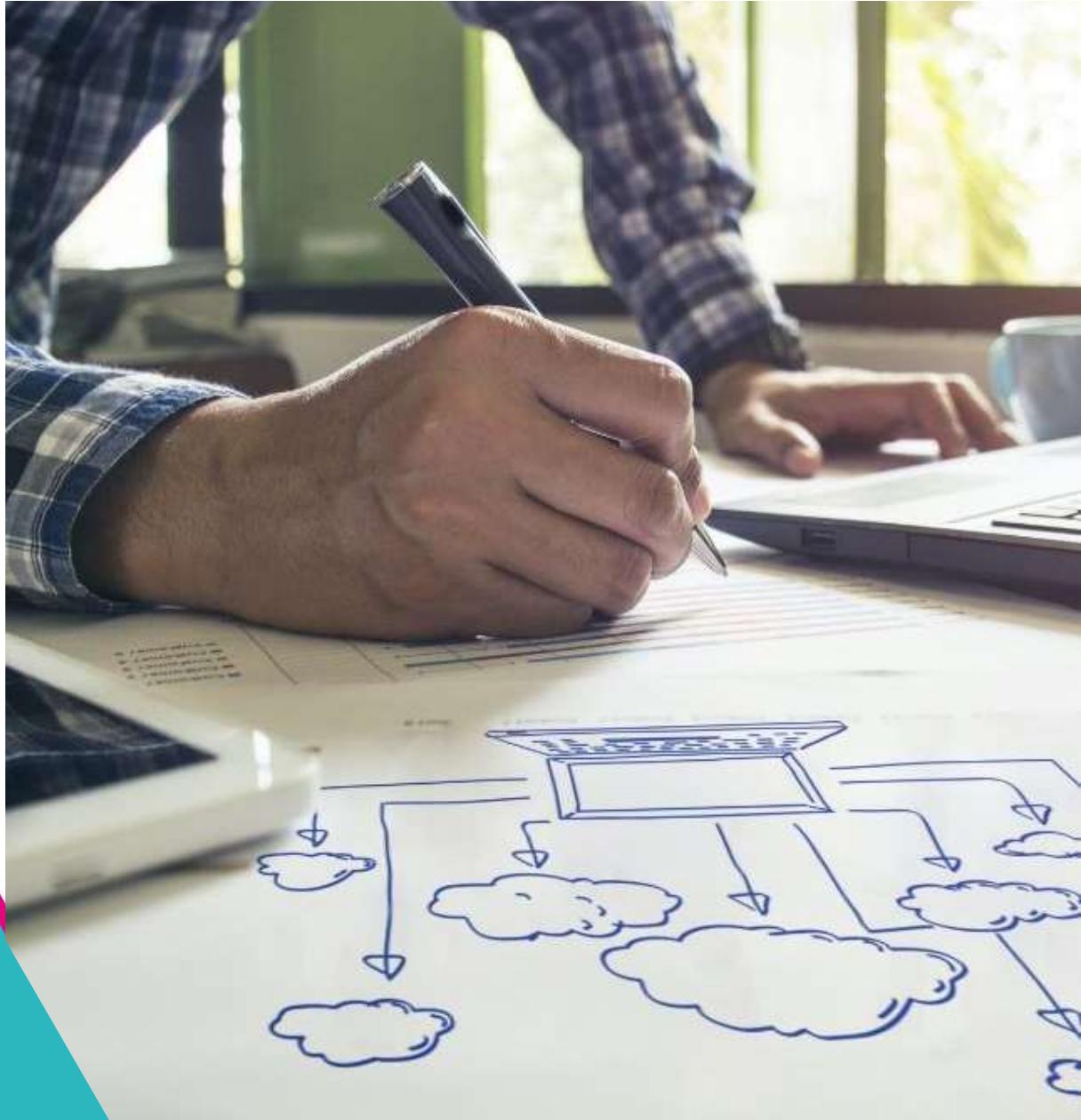
- ❑ Manage and govern operational, transactional and analytical data
- ❑ Cloud-native, purpose-built service to address discovery and compliance needs
- ❑ Fully managed, serverless, PaaS service
- ❑ Eliminate manual, ad-hoc and homegrown solutions





## 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



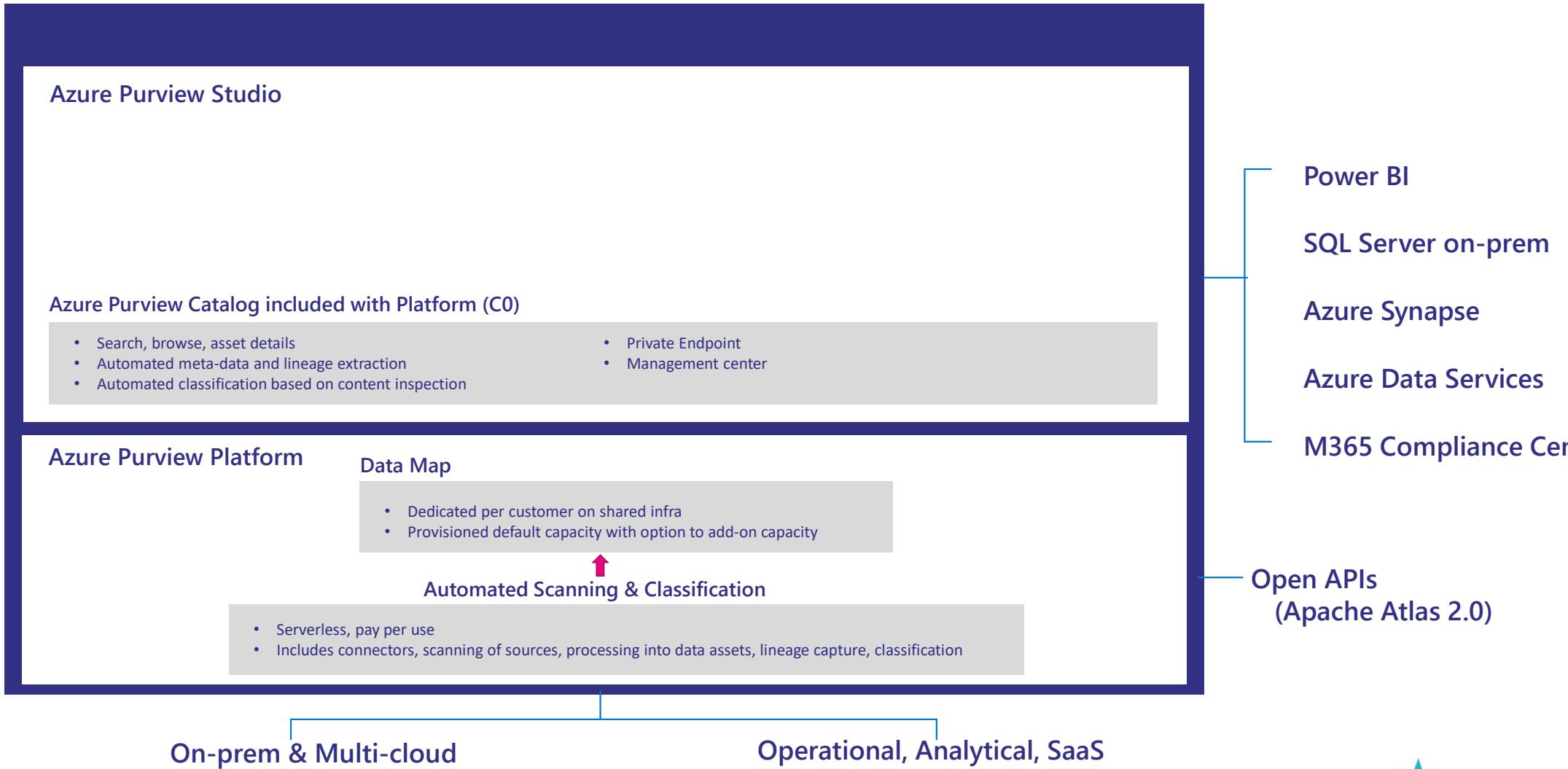


Gain insight into **data use**  
across the estate

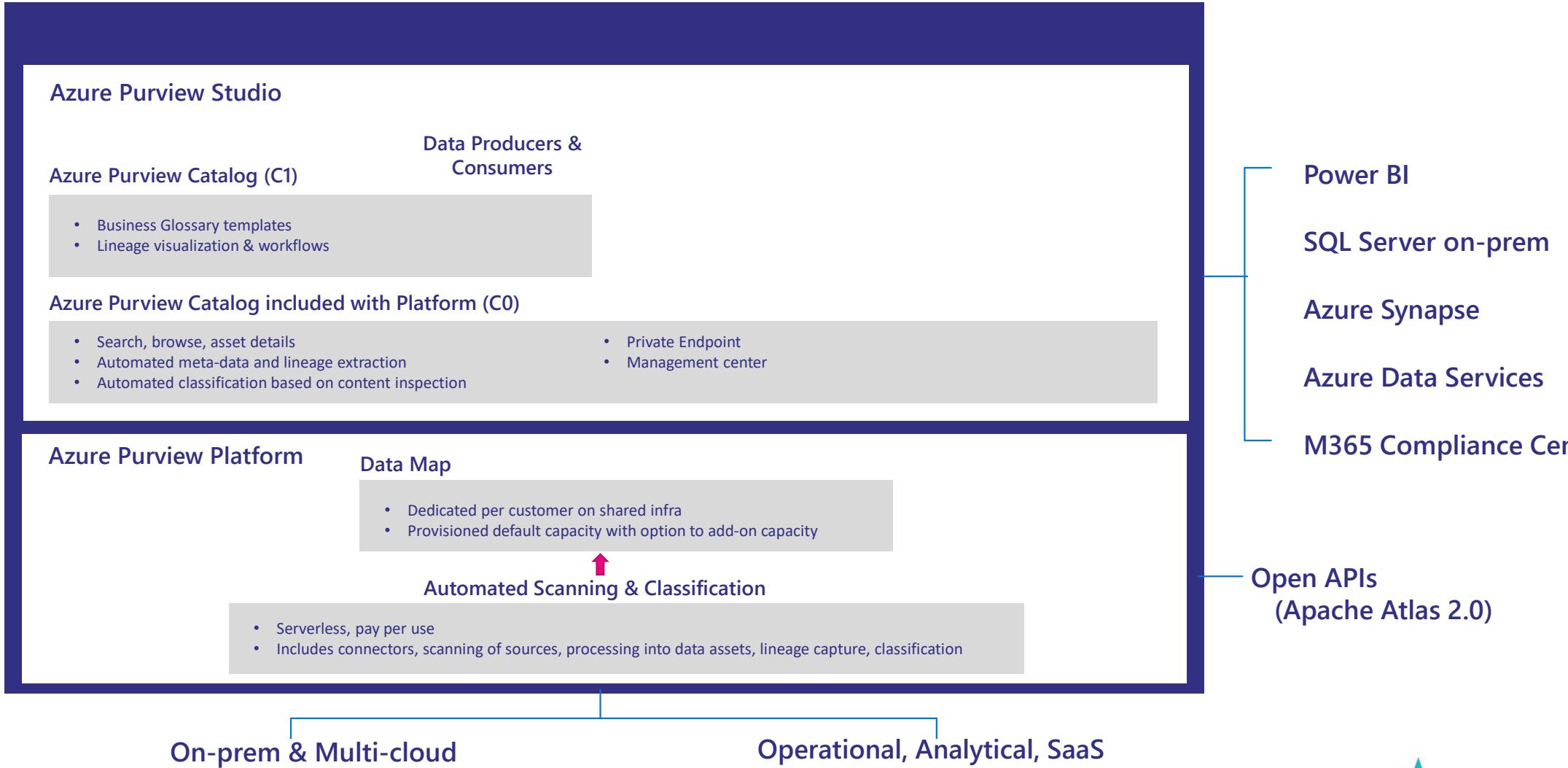
- Understand at a glance how data is being created and used across your data estate
- Visually assess the state of data assets, scans, business glossary and sensitive data



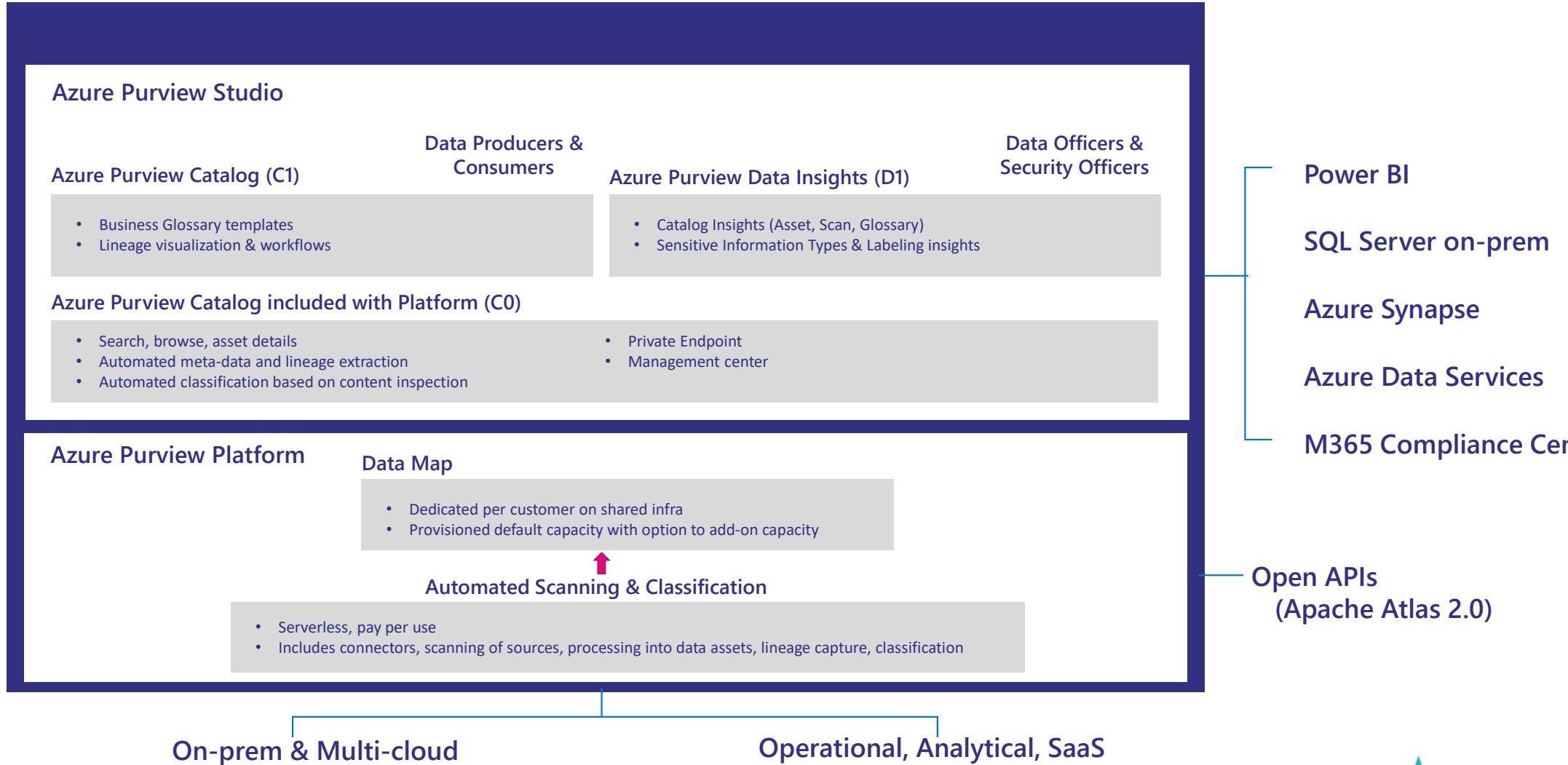
# Azure Purview Features



# Azure Purview Features



# Azure Purview Features



# Azure Purview - Access Control

## Collection Admins

- 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.

# Azure Purview - Access Control

Collection Admins

Data Source Admins

- can manage data sources and data scans

# Azure Purview - Access Control

Collection Admins

Data Source Admins

Data Curator

- can perform create, read, modify, and delete actions on catalog data objects and establish relationships between objects.

# Azure Purview - Access Control

Collection Admins

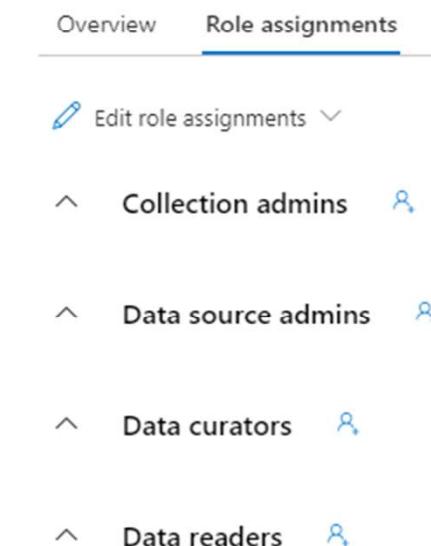
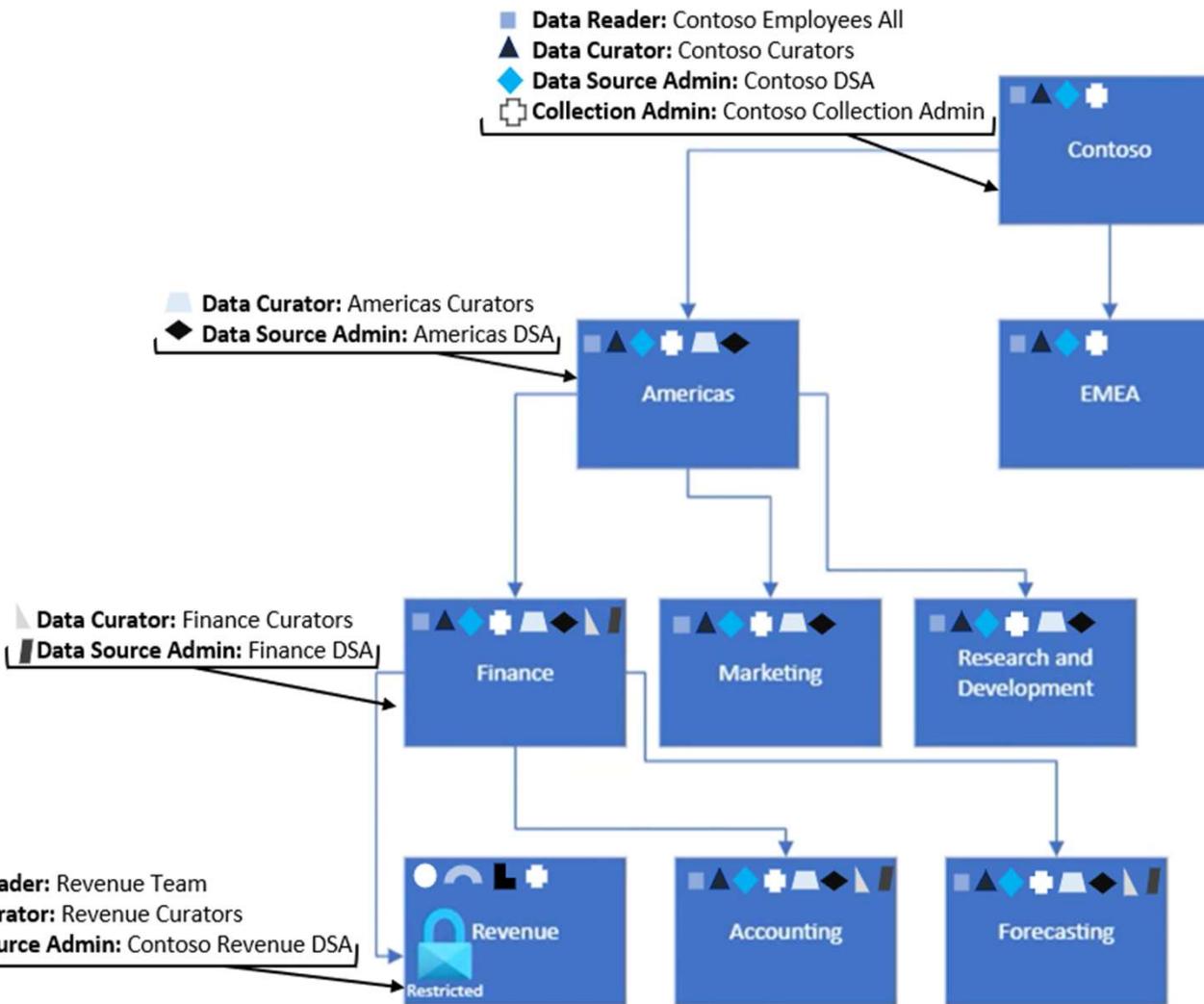
Data Source Admins

Data Curator

Data Reader

- can access but not modify catalog data objects.

# Azure Purview - Access Control

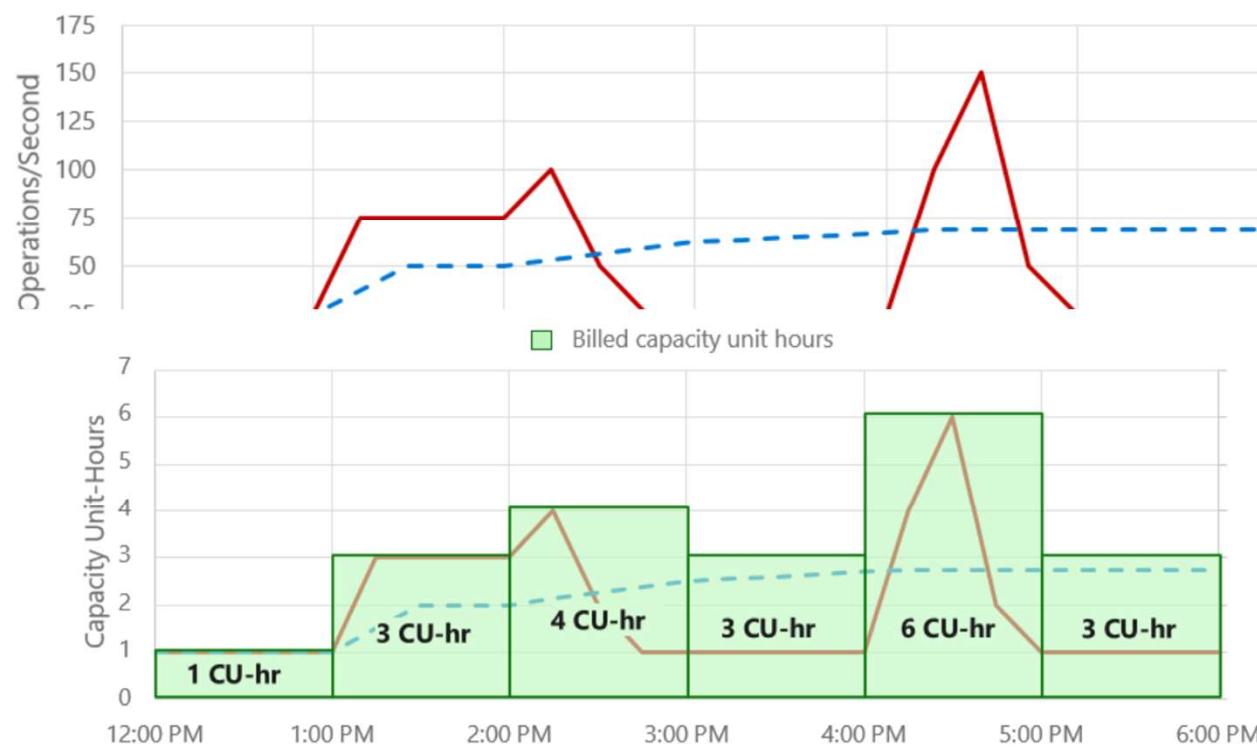


# Azure Purview - Pricing

## Azure Purview Data Map

- Capacity Unit
  - €0.353 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.

— Operations/Second    - - - Metadata Storage



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

# Azure Purview - Pricing

Azure Purview Data Map

- Power BI Online
  - Free in Preview
- SQL Server On Prem
  - Free in Preview
- Other Data Sources
  - Free in Preview
  - €0.540 per 1 vCore Hour

Scanning and Classification

# Azure Purview - Pricing

Azure Purview Data Map

- Power BI Online
  - Free in Preview
- SQL Server On Prem
  - Free in Preview
- Other Data Sources
  - Free in Preview
  - €0.540 per 1 vCore Hour

Scanning and Classification

# Azure Purview - Pricing-Example

Azure Purview Data Map

Scanning and Classification

Other Features

- Resource Set
  - €0.18 per 1 vCore Hour

# Azure Purview - Pricing

## Azure Purview Data Map

1 CPU x €0.353 X 730 hours

€ 257,69

## Scanning and Classification

4 scans x 2 hours x 32 VCore x  
€0.540 per vCore per hour

€ 138,24

## Other Features

4 scans x 1 hour x €0.18 per  
vCore per hour

€ 0,72

€396,65

# Azure Purview Studio

The screenshot shows the Azure Purview Studio interface with various UI elements highlighted by callout boxes and arrows:

- Metrics**: Points to the Metrics section in the top left.
- Search Bar**: Points to the Search catalog bar at the top.
- Recently Accessed Entities**: Points to the Recently accessed entities section at the bottom left.
- Updates**: Points to the Updates button in the top right.
- Accounts**: Points to the Accounts button in the top right.
- Notifications**: Points to the Notifications button in the top right.
- Feedback**: Points to the Feedback icon in the top right.
- Key Activities**: Points to the Key activities section in the middle right.
- Usefull Links**: Points to the Useful links section at the bottom right.

**Azure Purview Demo**

6 sources | 1,018 assets | No glossary terms

Search catalog

Browse assets | Manage glossary | Knowledge center

Recently accessed | My items

Name	Last update
Test_Purview	2 hours ago
cities	21 hours ago
Cities_staged	a day ago
telemetry	---

Links

- Purview overview
- Getting started
- Documentation

Discover the capabilities offered by Purview and learn how to make the most of them.  
Learn about how to use Purview Studio.  
Visit the documentation center for quickstarts, how-to guides, and references etc.

# Azure Purview Studio - Activity hubs



Data catalog



Data map

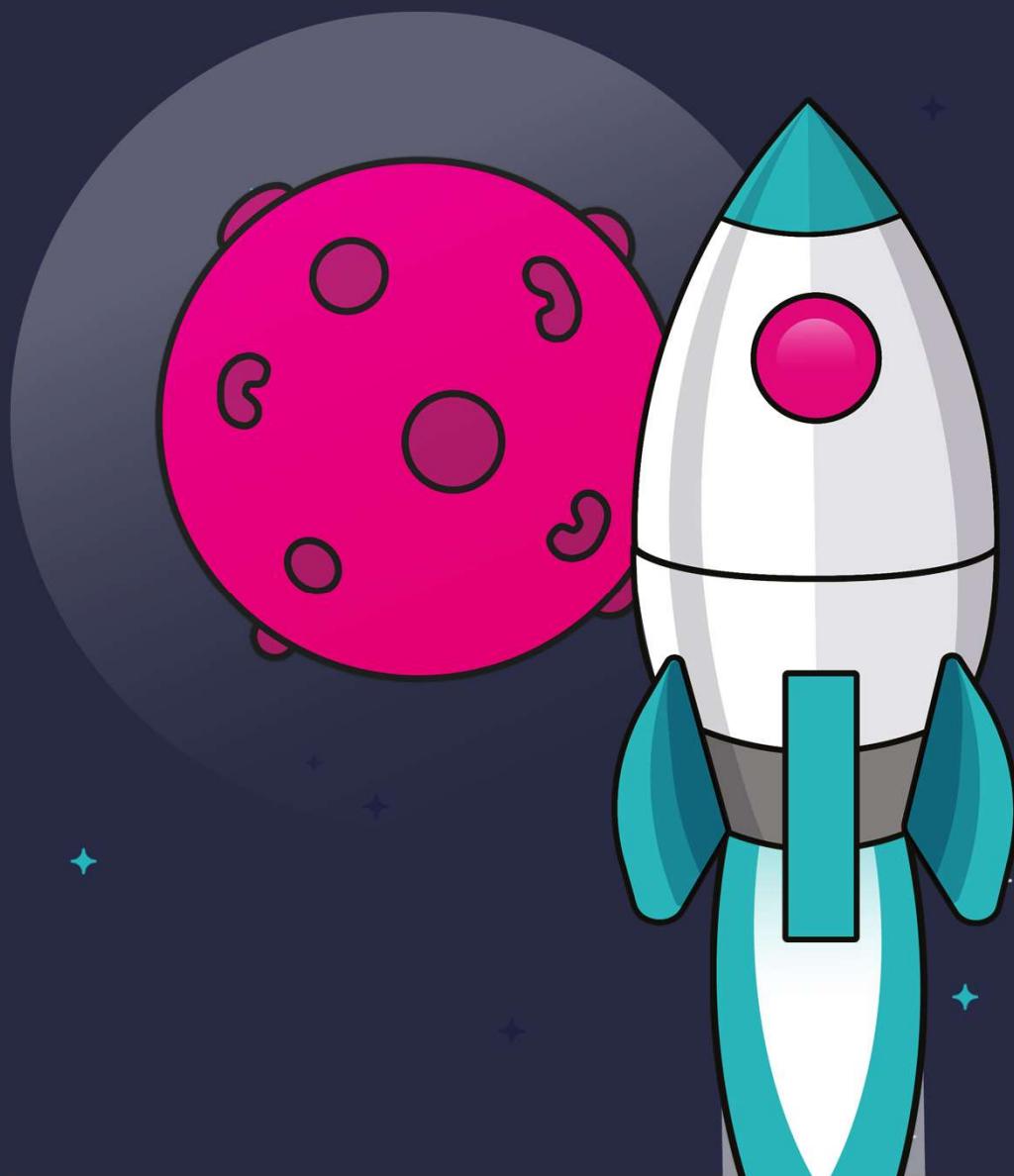


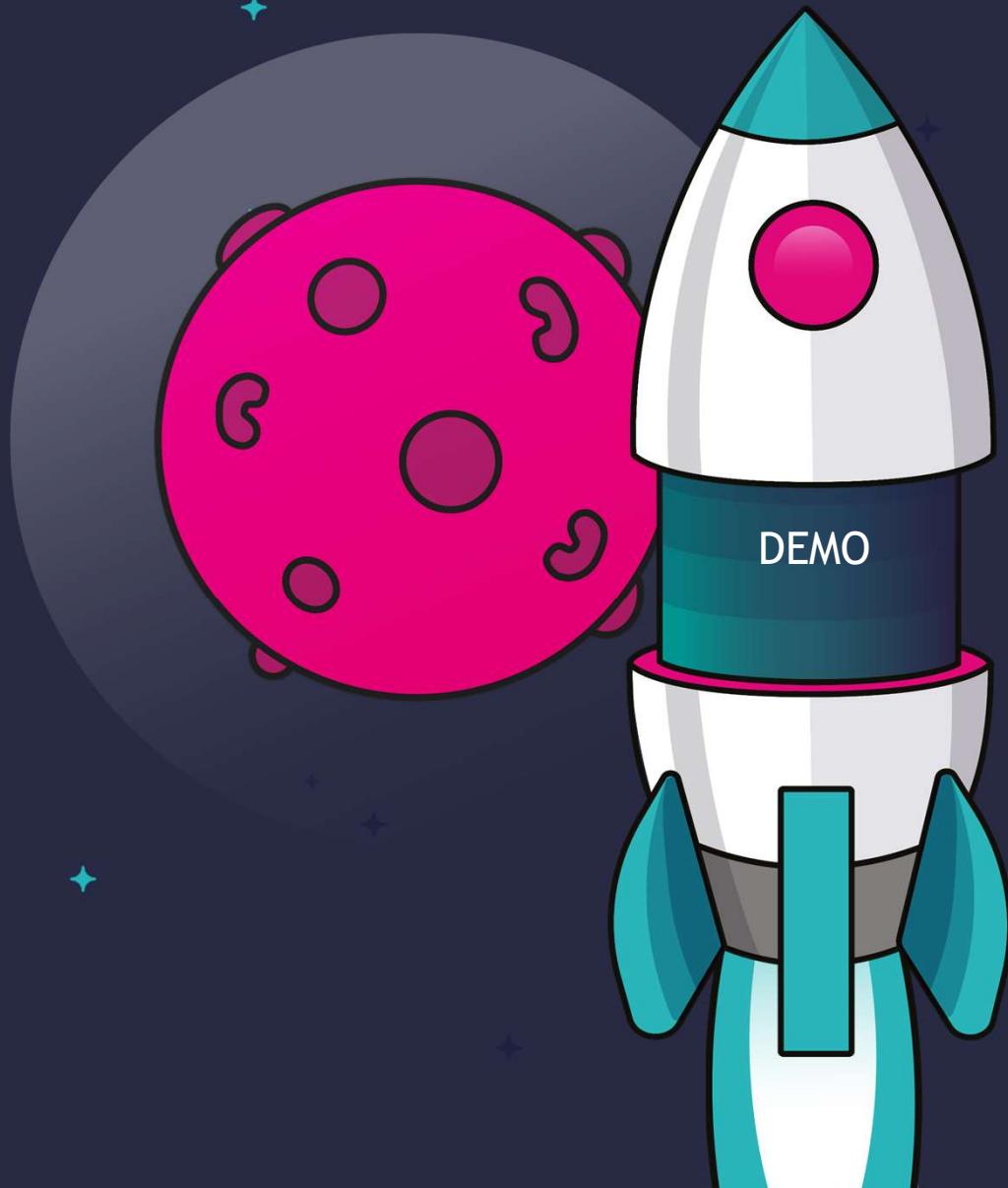
Insights



Management

- 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
- Meta Data Management Security, ADF and data share Connections





Innovate to accelerate

# DEMO

# Purview Data Map

Unify and make data meaningful

- ❑ Automated metadata scanning and lineage identification of hybrid data stores
- ❑ 100+ built-in and custom classifiers
- ❑ Microsoft Information Protection sensitivity labels



Sources

Register Refresh

Map view Table view

Showing 8 collections, 6 sources

Define the scope for this label

Labels can be applied directly to files, emails, containers like SharePoint sites and Teams, and more. Let us know so you can configure the applicable protection settings. [Learn more about label scopes](#)

**Files & emails**

Configure encryption and content marking settings to protect labeled emails and Office files. Also define auto-apply this label to sensitive content in Office, files in Azure, and more.

*(i) To set up auto-labeling for files in Azure, make sure you also scope this label to 'Azure Purview assets' below.*

**Groups & sites**

Configure privacy, access control, and other settings to protect labeled Teams, Microsoft 365 Groups, and SharePoint sites.

**Azure Purview assets (preview)**

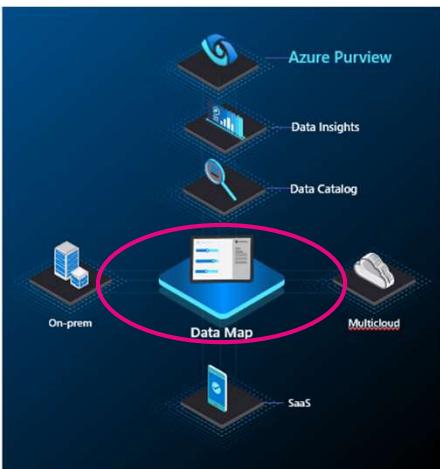
Apply label to assets in Azure Purview, including SQL columns, files in Azure Blob Storage, and more.

Azure (MULTIPLE)	aws (MULTIPLE)	Azure Synapse Analytics (MULTIPLE)
Amazon S3	Azure Blob Storage	Azure Cosmos DB (SQL API)
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen2
Azure Database for PostgreSQL	Azure Dedicated SQL Pool (formerly SQL DW)	Azure SQL Database
Azure SQL Database Managed Instance	erwin	INSPARK

# Purview Data Map

Unify and make data meaningful

- ❑ Automated metadata scanning and lineage identification of hybrid data stores
- ❑ 200+ built-in and custom classifiers
- ❑ Microsoft Information Protection sensitivity labels



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				
Completed	Scheduled				
Completed	Scheduled				
Completed	Manual				
Completed	Manual				

Scope your scan

Refresh

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

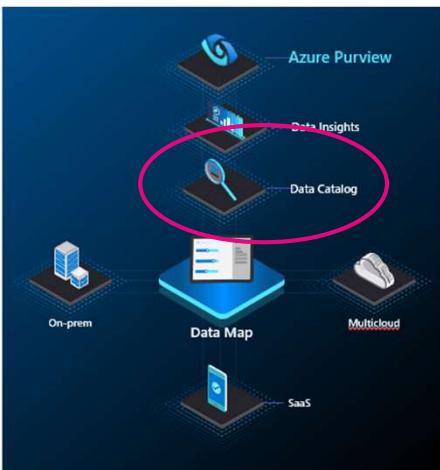
Search

<input checked="" type="checkbox"/> LS_ADLS
<input type="checkbox"/> > azure-webjobs-hosts
<input type="checkbox"/> > azure-webjobs-secrets
<input checked="" type="checkbox"/> > intermediate
<input type="checkbox"/> > marts
<input checked="" type="checkbox"/> > raw
<input type="checkbox"/> > sandbox
<input type="checkbox"/> > temp
<input checked="" type="checkbox"/> > tempstaging
<input checked="" type="checkbox"/> > transfer

# Azure Purview Data Catalog

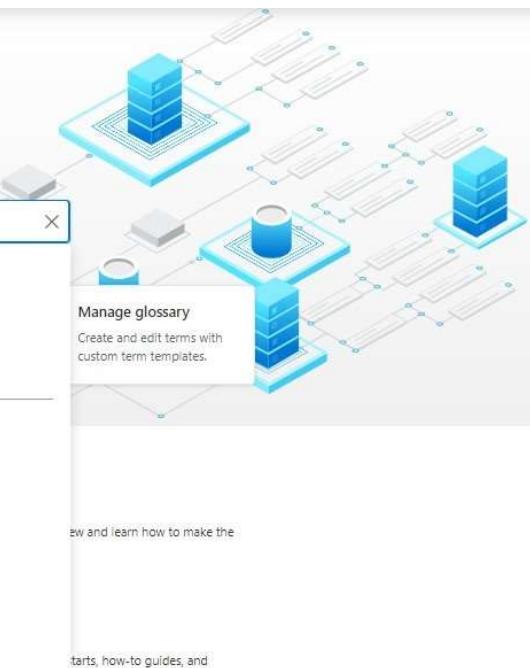
Enable effortless discovery

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



The screenshot displays the Azure Purview Demo interface. At the top, it shows "4 sources", "1,138 assets", and "1 glossary term". Below this is a search bar with the query "customer". The search results are categorized into "Search suggestions" and "Asset suggestions". Under "Search suggestions", there are four items: "customer transactions original", "customer changes", "customer origin", and "customer original". Under "Asset suggestions", there are five items, each with a preview icon and a link:

- Customer: <https://labseuwdvlmdfoxgn01.dfs.core.windows.net/intermediate/DeltaLake/AW/SalesLT/Customer/>
- Customer: <https://labseuwdvlmdfoxgn01.dfs.core.windows.net/intermediate/DeltaLake/WWI/Test/Customer/>
- Customer Profitability Sample: <https://app.powerbi.com/groups/3257d8ca-4da0-47fa-a172-e01d9b7d428a/reports/db6bc825-1e63-4c6a-9584-548f66d5c040>
- Customer Profitability Sample: <https://app.powerbi.com/groups/3257d8ca-4da0-47fa-a172-e01d9b7d428a/dashboards/b9a1cd08-3679-4192-91e3-5e6faf3ecbd>



# Azure Purview Data Catalog

Unify and make data meaningful

The screenshot shows the Azure Purview Data Catalog interface. The top navigation bar says "Browse assets > Azure SQL Database". Below it, there's a "Refresh" button and a breadcrumb trail: "Browse assets > Azure SQL Database > WideWorldImporters".

The main area displays a list of databases under the "Name" column:

Name	Type
Adventureworks	Azure SQL Database
Datastore	Azure SQL Database
Datastore	Azure SQL Database
datastore	Azure SQL Database
oxygenconfigdb	Azure SQL Database
test_purview	Azure SQL Database
WideWorldImporters	Azure SQL Database
WideWorldImportersDW	Azure SQL Database

Under the "WideWorldImporters" database, there are several schema objects listed:

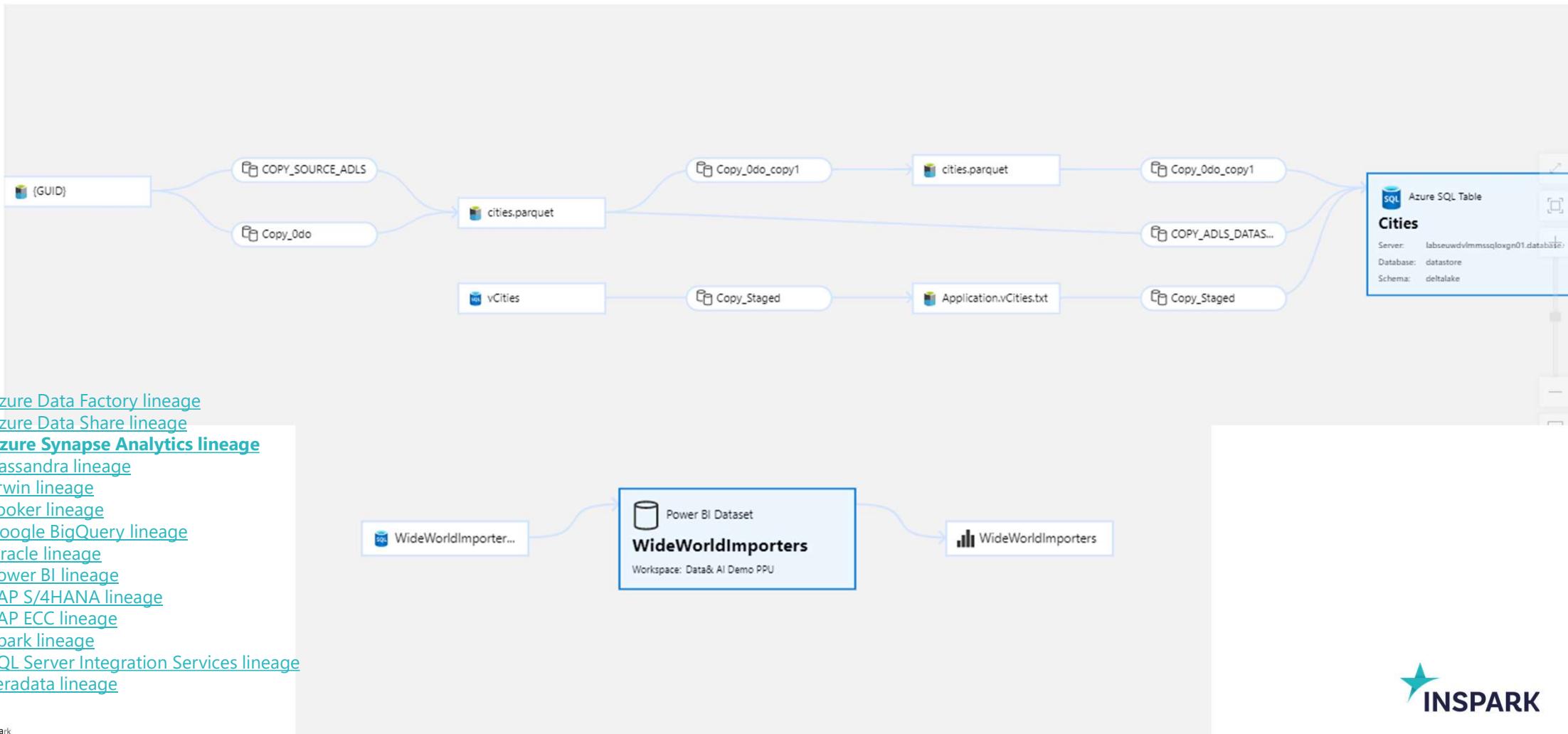
Name	Type
PurchaseOrderLines	Azure SQL Table
PurchaseOrders	Azure SQL Table
SupplierCategories	Azure SQL Table
SupplierCategories_Archive	Azure SQL Table
Suppliers	Azure SQL Table
Suppliers_Archive	Azure SQL Table
SupplierTransactions	Azure SQL Table

At the bottom, there are category cards:

- Azure Data Factory: 4 items
- Azure Data Lake Storage Gen2: 1 item
- Azure SQL Database: 8 items
- Azure SQL Server: 3 items** (This card is highlighted with a red box)
- Azure Storage Account: 1 item
- Power BI: 18 items

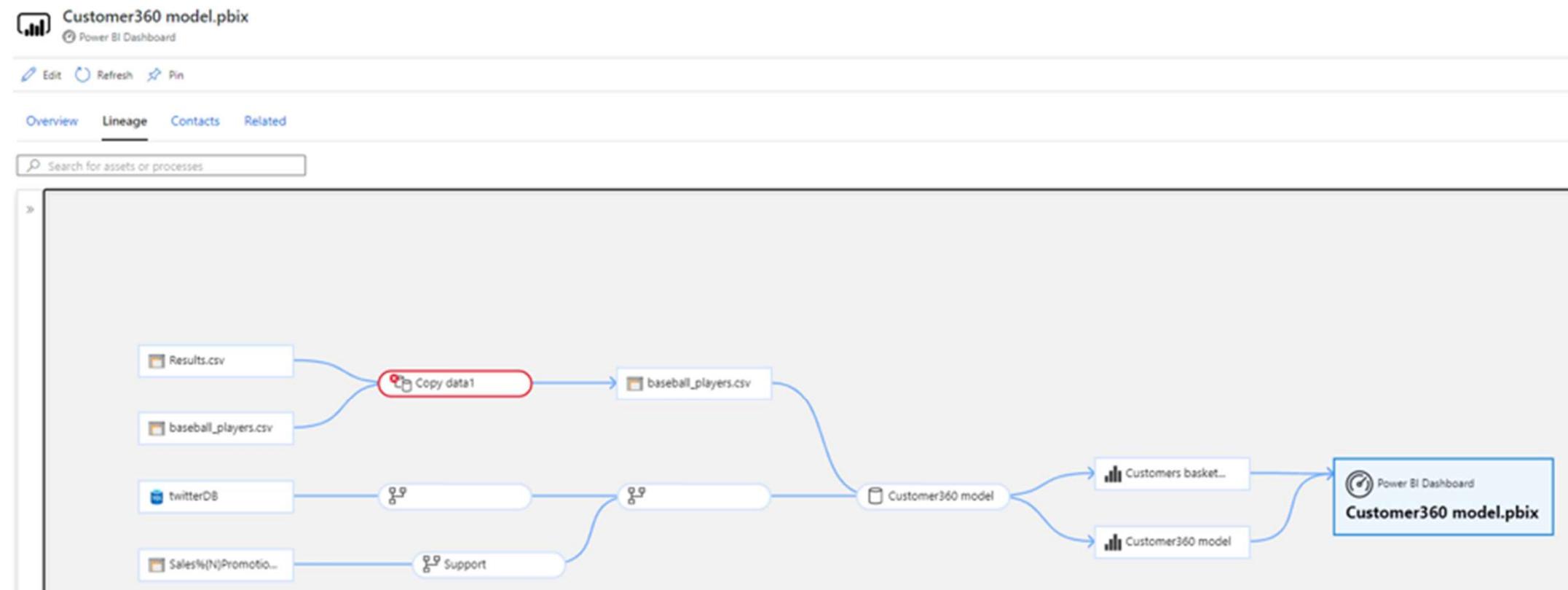
# Azure Purview Data Catalog

Unify and make data meaningful



# Azure Purview Data Catalog

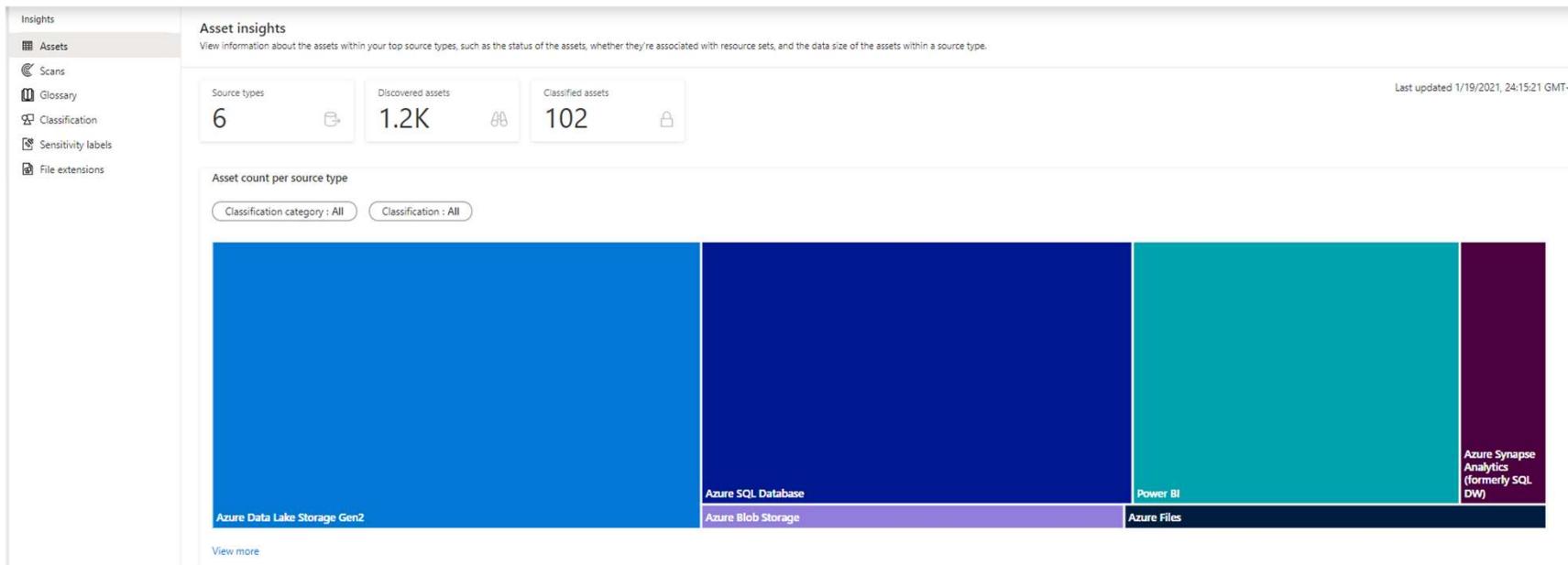
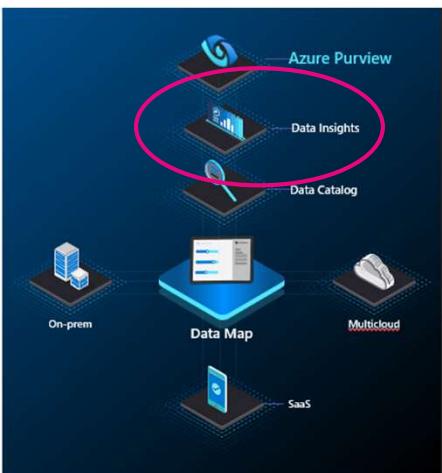
Unify and make data meaningful



# Insights

Get a bird's-eye view of sensitive data

- Reports on Assets, Scans, Glossary, Classification, and Labeling

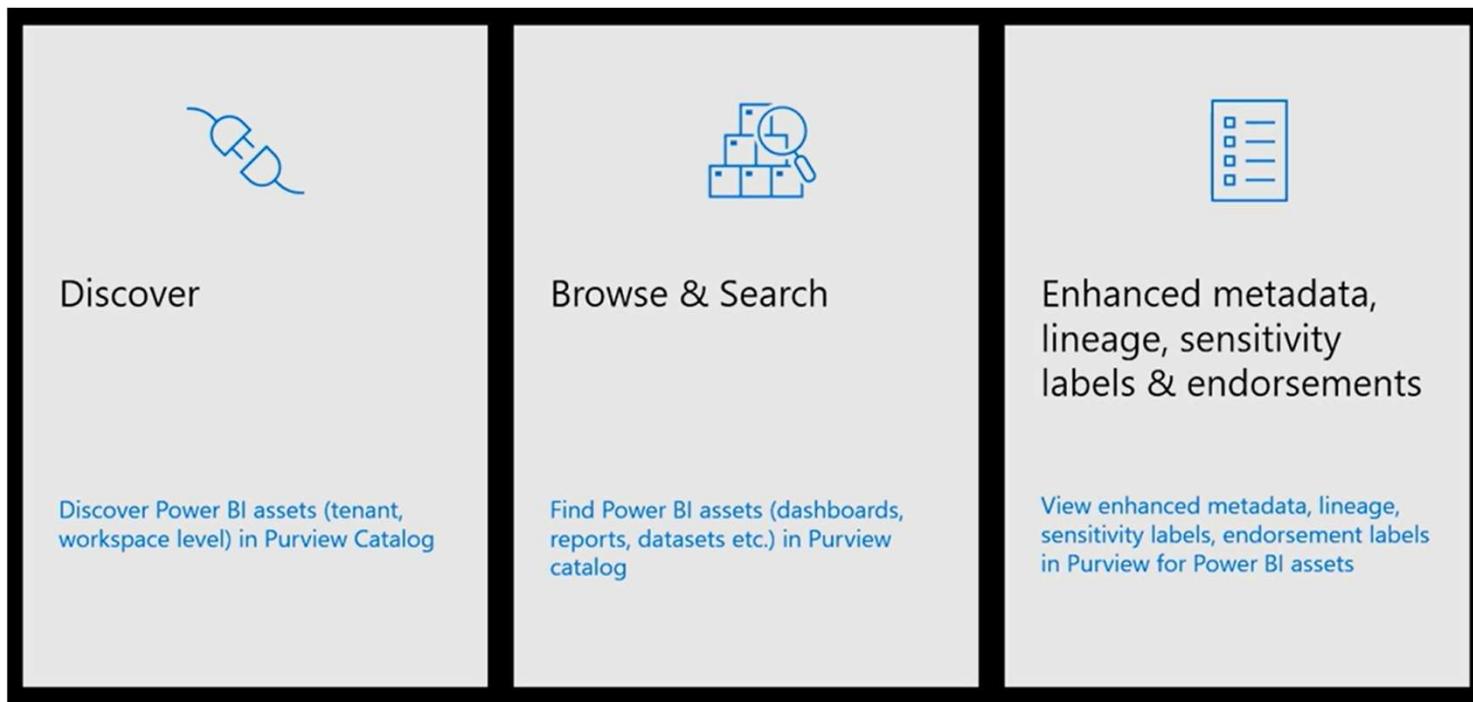




 INSPARK

# Register and Scan a Power BI Tenant

Discover data registered and scanned by Azure Purview



# Register and Scan a Power BI Tenant

Discover data registered and scanned by Azure Purview

- Allow service principals to use read-only Power BI admin APIs
- Enhance admin APIs responses with detailed metadata
- Enhance admin APIs responses with DAX and mashup expressions (Preview)  
*Enabled for a subset of the organization*

- ◀ Allow service principals to use read-only Power BI admin APIs  
*Enabled for a subset of the organization*

Web apps registered in Azure Active Directory (Azure AD) will use an assigned principal to access read-only Power BI Admin APIs without a signed in user. If your app to use service principal authentication, its service principal must be included in an allowed security group. By including the service principal in the allowed security groups, you're giving the service principal read-only access to all the information available through Power BI admin APIs (current and future). For example, Power BI can scan emails, dataset and report detailed metadata. [Learn more](#)



Apply to:

- The entire organization
- Specific security groups

sg-oxgn-services-dvlm

Enter security groups

- ◀ Enhance admin APIs responses with detailed metadata (Preview)  
*Enabled for a subset of the organization*

Users and service principals allowed to call Power BI admin APIs may get detailed metadata about Power BI items. For example, responses from GetScanResult APIs will contain the names of dataset tables and columns. [Learn more](#)

Note: For this setting to apply to service principals, make sure the tenant setting allowing service principals to use read-only admin APIs is enabled. [Learn more](#)



Apply to:

- The entire organization
- Specific security groups

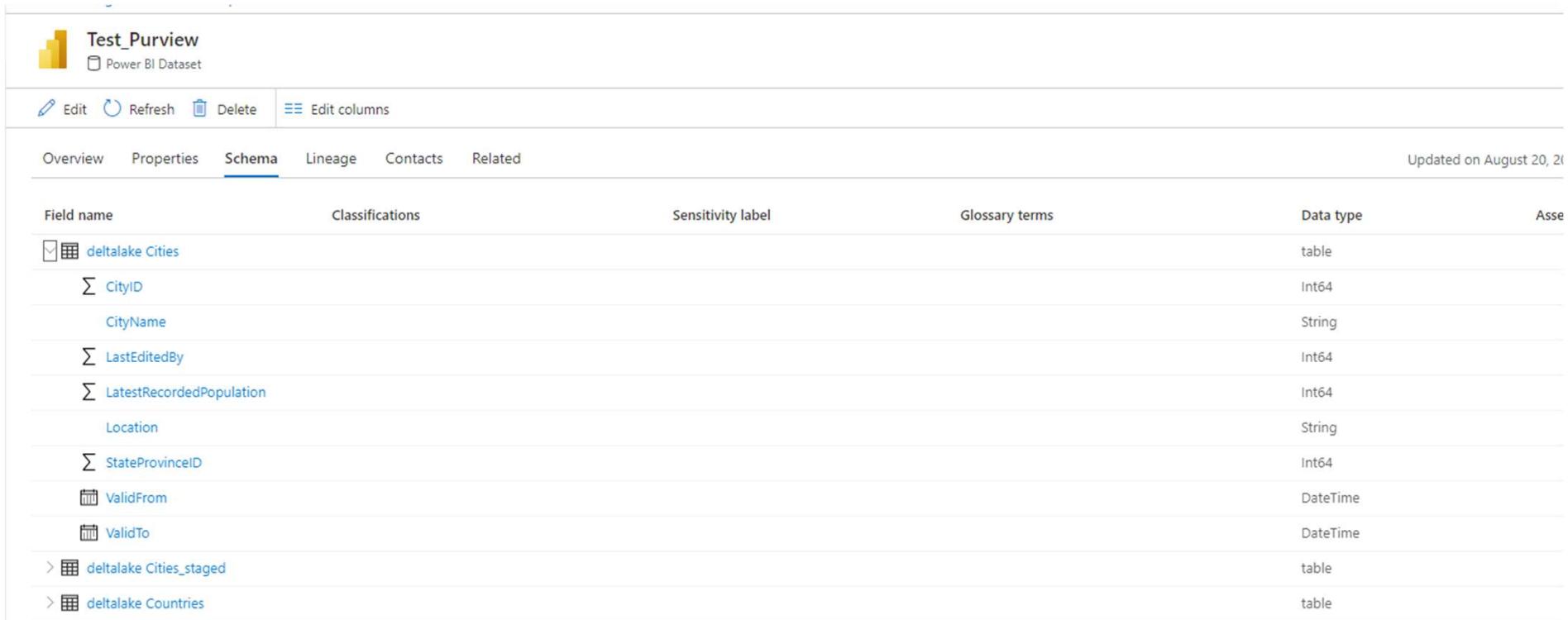
sg-oxgn-services-dvlm

Enter security groups

# Register and Scan a Power BI Tenant

Discover data registered and scanned by Azure Purview

- Allow service principals to use read-only Power BI admin APIs
- Enhance admin APIs responses with detailed metadata
- Enhance admin APIs responses with DAX and mashup expressions (Preview)



The screenshot shows the Azure Purview Data Catalog interface. At the top, there's a navigation bar with a yellow square icon, the text 'Test\_Purview', and a 'Power BI Dataset' link. Below the navigation bar are buttons for 'Edit', 'Refresh', 'Delete', and 'Edit columns'. A horizontal menu bar includes 'Overview', 'Properties', 'Schema' (which is underlined), 'Lineage', 'Contacts', and 'Related'. To the right of the menu bar, it says 'Updated on August 20, 2023'. The main area displays a table with columns: 'Field name', 'Classifications', 'Sensitivity label', 'Glossary terms', 'Data type', and 'Asse...'. The table lists several fields for a 'deltalake Cities' table:

Field name	Classifications	Sensitivity label	Glossary terms	Data type	Asse...
Δeltalake Cities				table	
Σ CityID				Int64	
CityName				String	
Σ LastEditedBy				Int64	
Σ LatestRecordedPopulation				Int64	
Location				String	
Σ StateProvinceID				Int64	
ValidFrom				DateTime	
ValidTo				DateTime	
> Δeltalake Cities_staged				table	
> Δeltalake Countries				table	

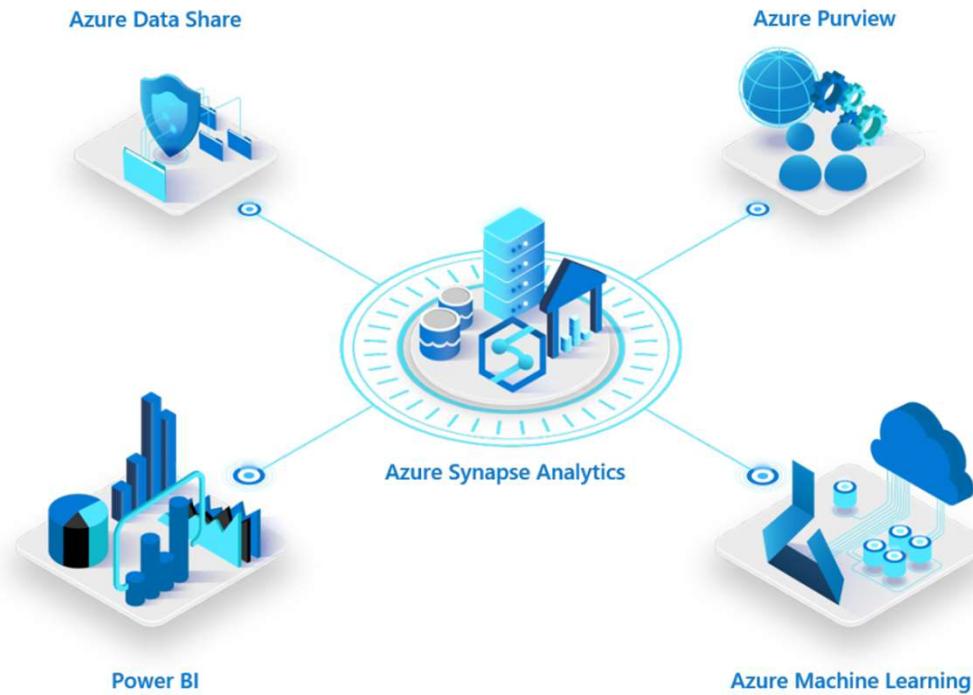
Synapse

Purview



# Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview



# Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

The screenshot shows the Azure Synapse Analytics workspace settings page. On the left, a sidebar lists various configuration options like Analytics pipelines, SQL pools, External connectors, Linked services, and Azure Purview. The main area displays three cards:

- Register**: Shows a plug icon and the text "Register Synapse workspaces with Purview".
- Browse & Search**: Shows a search icon and the text "Search datasets in Synapse workspaces from linked Purview catalog".
- Enhanced metadata, lineage and annotation**: Shows a table icon and the text "View enhanced metadata, lineage from Purview for Synapse tables and add annotations".

# Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

Azure 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**    Edit    Refresh    Disconnect

Purview account

Default Purview integration capabilities:

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

Additional Purview integration capabilities and status:

Integration	Status
Data Lineage - Synapse Pipeline	Connected

# Integrate Azure Purview in Azure Synapse Analytics

Discover data registered and scanned by Azure Purview

All pipeline runs > SQLServerDataIngestion - Activity runs

## SQLServerDataIngestion

List Gantt



Rerun



Rerun from activity



Rerun from failed activity



Refresh



Update pipeline



+ - [ ]

-

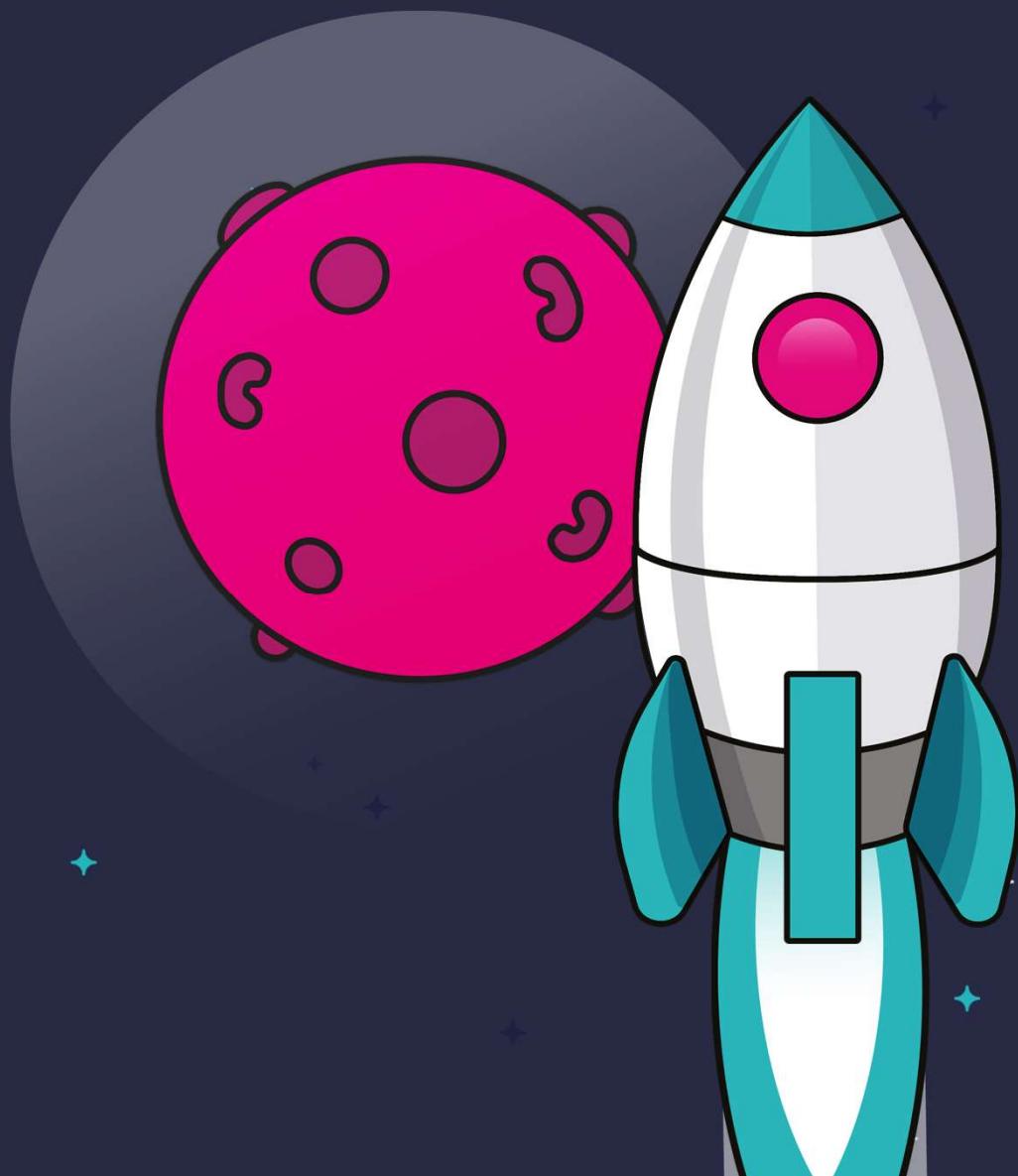
## Activity runs

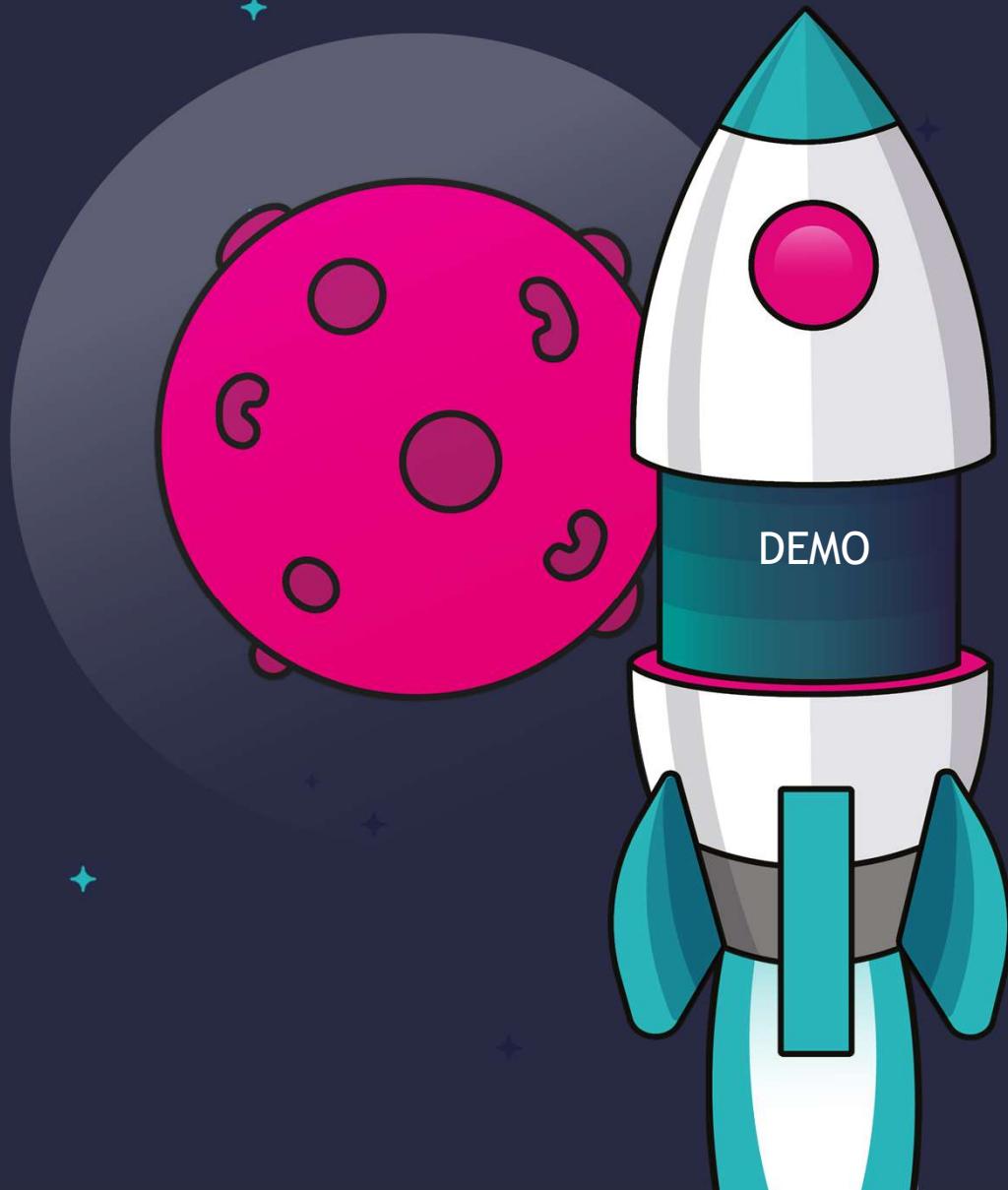
Pipeline run ID

All status

Showing 1 - 33 of 33 items

Activity name	Activity type	Run start ↑	Duration	Status	Lineage status	Error	Log
CopyData	Copy data	8/10/21, 10:41:08 AM	00:00:44	Succeeded			
CopyData	Copy data	8/10/21, 10:41:00 AM	00:00:46	Succeeded			
CopyData	Copy data	8/10/21, 10:40:59 AM	00:00:44	Succeeded			RK





Innovate to accelerate

# DEMO

## Roadmap /Preview features

- Data Quality
- Data Privacy
- Master Data Management
- Data Policy
- Microsoft Defender for Cloud



# Policy in Azure Purview

[Create policy](#)

**Name \***  
FinanaceLakePolicy

**Description \***  
Finance Lake policy

**Policy statements**

**Example policy statement:**

FinanaceLakePolicy

[Edit](#)  [Delete](#)

**Description**  
Finance Lake policy

**Policy statements**

**Allow Modify** on Data Asset "**Inventory**" To Principal "**Alice Beringer**"  
0a6c54c6-a72a-41d6-b0df-6464267d3cf  
0a6c54c6-a72a-41d6-b0df-6464267d3cf > prasad-demo-rg > findatalake > finance > Data/Inventory

**Allow Read** on Data Asset "**Customer**" To Group "**FinanceAnalysts**"  
0a6c54c6-a72a-41d6-b0df-6464267d3cf  
0a6c54c6-a72a-41d6-b0df-6464267d3cf > prasad-demo-rg > findatalake > finance > Data/Customer

[What's New in Azure Purview at Microsoft Ignite 2021 - Microsoft Tech Community](#)

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

**Data sources**  
Select the data source you want this policy to be applied to.  
 Data source Type \* Select...

**Assets**  
Select the data source to search the catalog

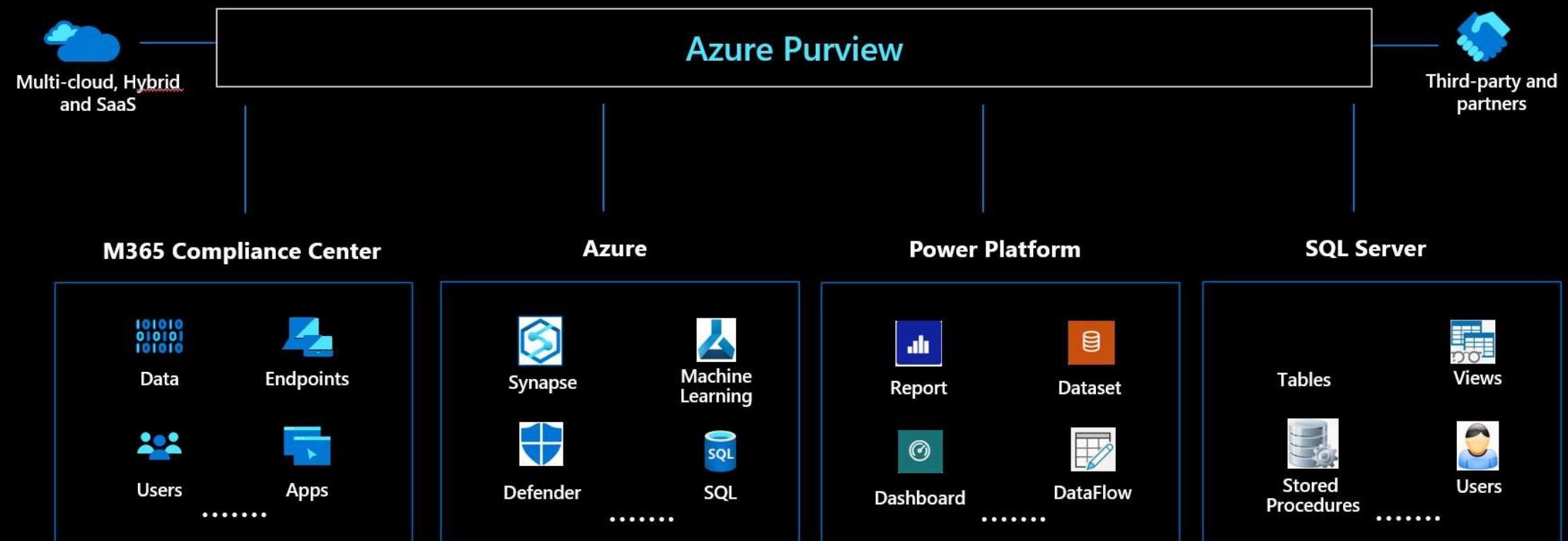
**Policy publish**  
Select sources to publish 'FinanaceLakePolicy' to.

Search for a data source

Selected (0)  
Showing 1 - 5 of 5 items

Data source name	Registered on
<input type="checkbox"/> FinanceDataLake	09/27/2021, 11:15 AM
<input type="checkbox"/> FinanceDataLake2	09/27/2021, 01:39 PM
<input checked="" type="checkbox"/> Finance-SQL-DB	10/14/2021, 04:47 PM
<input type="checkbox"/> JimGrayDB	10/15/2021, 01:37 PM
<input type="checkbox"/> MarketingLake	10/25/2021, 03:39 PM

# Data Governance Opportunity



# Questions?



<https://erwindekreuk.com>



@erwindekreuk



<https://www.linkedin.com/in/erwindekreuk/>

Slides will be available on my blog

