



# Microsoft Purview

## Data Governance



# Speaker

## Erwin de Kreuk

- Principal Consultant
- Lead Data & AI
- Data platform MVP



erwindekreuk/  
Erwindekreuk.com  
@erwindekreuk



InSpark



#ScottishSummit2022



Microsoft  
Partner

2021 Partner of the Year Finalist  
Analytics Award

# We Are InSpark

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

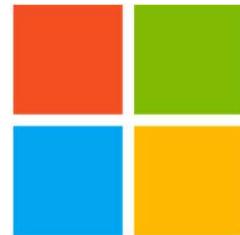


#ScottishSummit2022



# Thank You to our Sponsors...

Event Sponsor



# Microsoft

Platinum Sponsors



the results company



#ScottishSummit2022



# Thank You to our Sponsors...

## Gold Sponsors



Data Quality



Event Lunch



Accessibility



Data Analytics



#ScottishSummit2022

# Microsoft Purview



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



# Agenda

Introduction to Purview

Data Map

Data Catalog

Data Estate Insights



#ScottishSummit2022

# 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

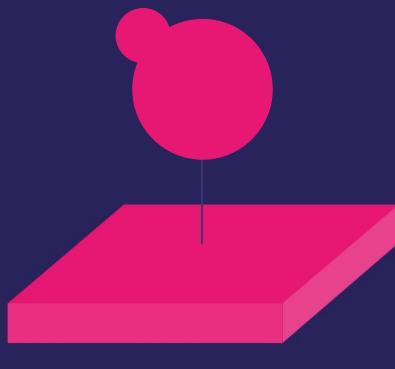
IDC

*IDC Data Age 2025, Dave Reinsel,*

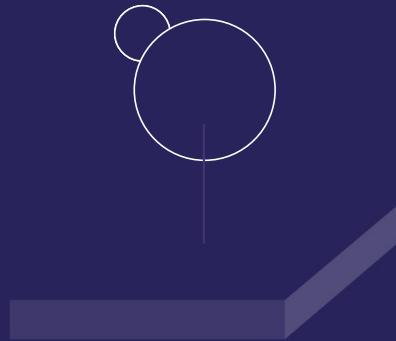




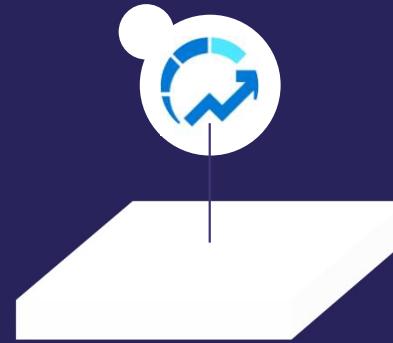
# Elements of successful data governance



- Manage growing data landscape



- Overcome operational silos



- Increase data agility



- Comply with industry regulations

#ScottishSummit2022

# Microsoft Purview

Unified data governance to maximize the business value of data

Reimagine **data governance** in the cloud

Set the foundation for effective **data governance**

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

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





Data governance is becoming increasingly interdisciplinary



#ScottishSummit2022



# Data governance is becoming increasingly interdisciplinary



## Chief Data Officer

**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 Management  
Data Stewardship  
Data Engineering  
Data Science  
Data Analyst

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.



# History



ADC Gen 1



BlueTalon Acquisition (June 2019)



ADC Gen 2



Azure Purview sept 2021

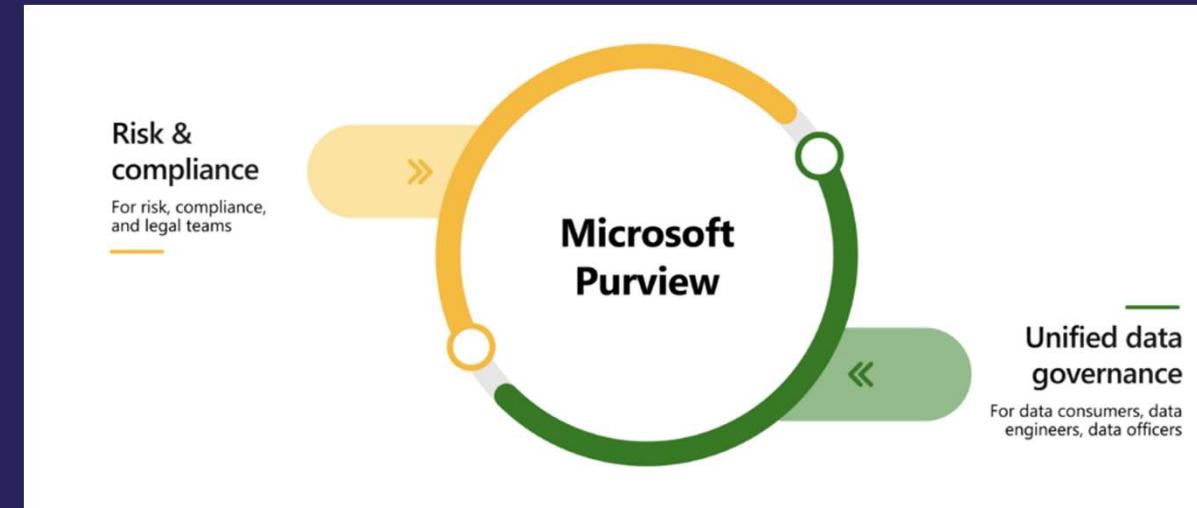


Microsoft Purview april 2022

#ScottishSummit2022

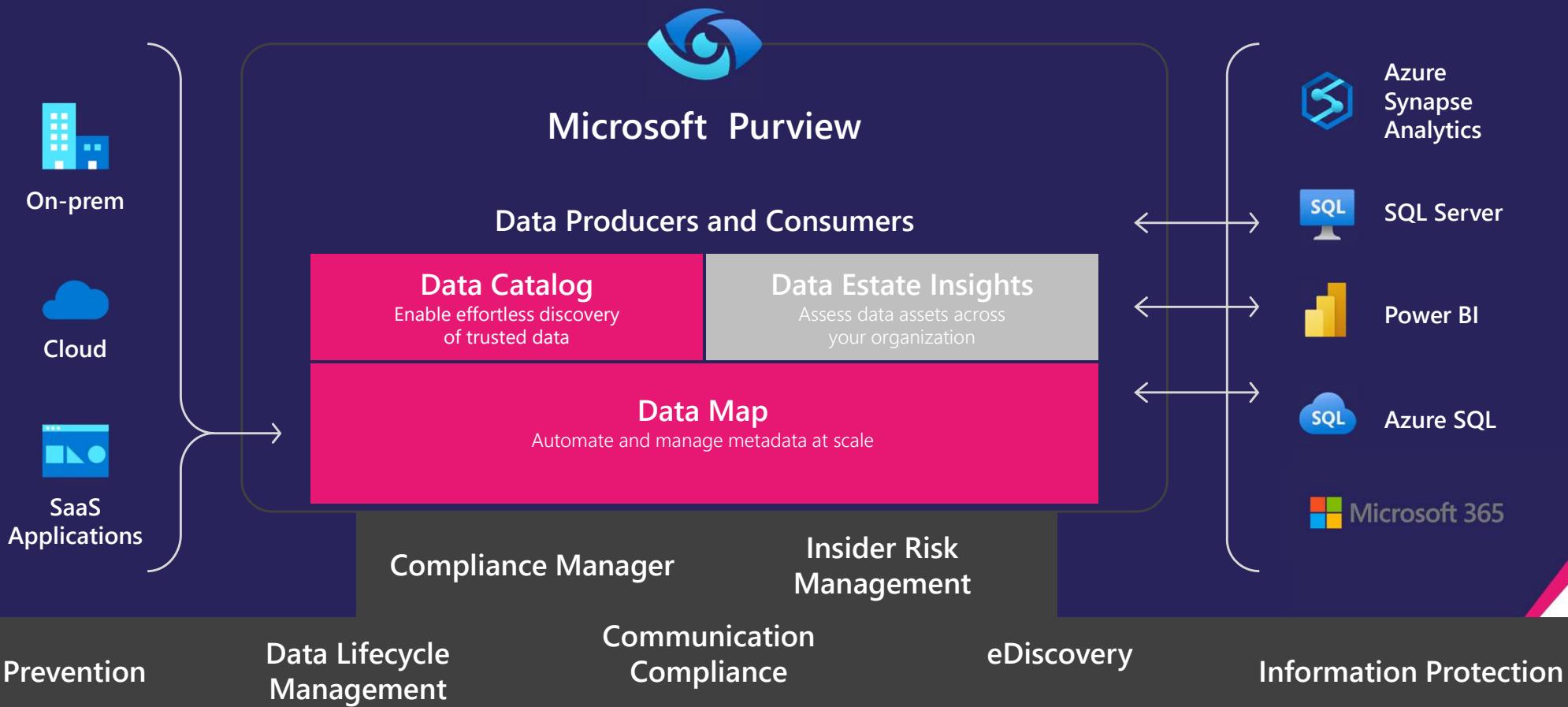
# The future of compliance and data governance

- 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.





# Unified Data Governance





# Microsoft Purview Features

## Microsoft Purview Studio

### Microsoft Purview Catalog included with Platform (C0)

- Search, browse, asset details
- Automated meta-data and lineage extraction
- Automated classification based on content inspection
- Private Endpoint
- Management center

### Microsoft Purview Platform Microsoft Purview Data Map

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

#### Automated Scanning & Classification

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

On-prem & Multi-cloud

Operational, Analytical, SaaS

#ScottishSummit2022

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Center

Open APIs  
(Apache Atlas 2.2)



# Microsoft Purview Features

## Microsoft Purview Studio

Data Producers & Consumers

### Microsoft Purview Catalog (C1)

- Business Glossary templates
- Lineage visualization & workflows

### Microsoft Purview Catalog included with Platform (C0)

- |   |  |
|---|--|
| <ul style="list-style-type: none"><li>• Search, browse, asset details</li><li>• Automated meta-data and lineage extraction</li><li>• Automated classification based on content inspection</li></ul> | <ul style="list-style-type: none"><li>• Private Endpoint</li><li>• Management center</li></ul> |
|---|--|

## Microsoft Purview Platform Microsoft Purview Data Map

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

### Automated Scanning & Classification

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

On-prem & Multi-cloud

Operational, Analytical, SaaS

#ScottishSummit2022

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Center

Open APIs  
(Apache Atlas 2.2)



# Microsoft Purview Features

## Microsoft Purview Studio

### Microsoft Purview Catalog (C1)

- Business Glossary templates
- Lineage visualization & workflows

### Microsoft Purview Catalog included with Platform (C0)

- Search, browse, asset details
- Automated meta-data and lineage extraction
- Automated classification based on content inspection

## Data Producers & Consumers

### Microsoft Purview Data Estate Insights (D1)

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

## Data Officers & Security Officers

## Microsoft Purview Platform

### Microsoft Purview Data Map

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

### Automated Scanning & Classification

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

On-prem & Multi-cloud

Operational, Analytical, SaaS

#ScottishSummit2022

Power BI

SQL Server on-prem

Azure Synapse

Azure Data Services

M365 Compliance Center

Open APIs  
(Apache Atlas 2.2)



# Microsoft Purview Studio

Navigation bar:

- Updates
- Accounts
- Notifications

Top right corner:

- Feedback
- Key Activities
- Usefull Links

Main content area:

**Microsoft Purview InSpark**  
(labseuwvdvImpurviewoxgn01)

17 sources | 1,950 assets | 56 glossary terms

Search catalog

Recently accessed entities:

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:

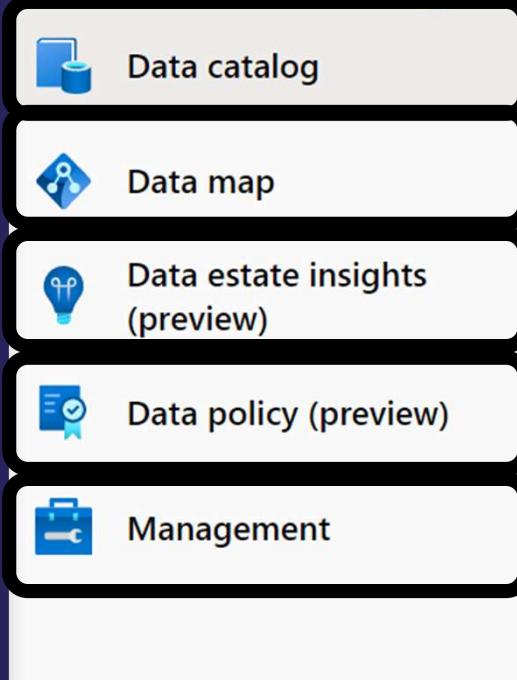
- Microsoft Purview overview
- Getting started
- Documentation

Bottom right corner:

22

# 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



#ScottishSummit2022

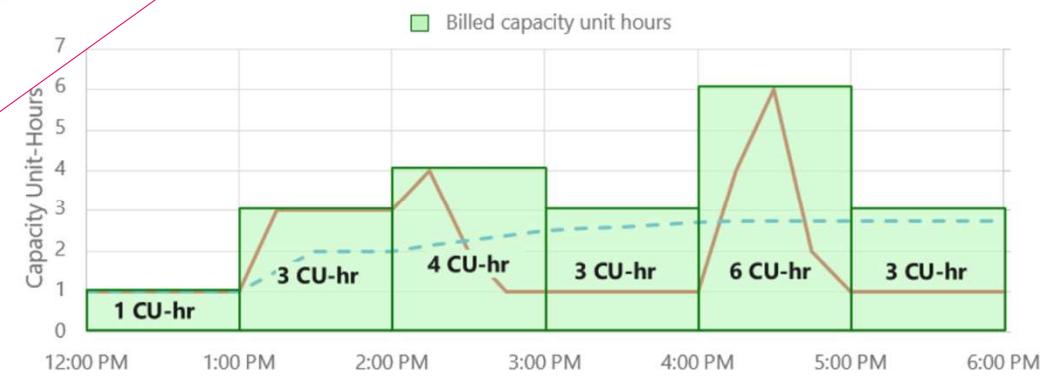
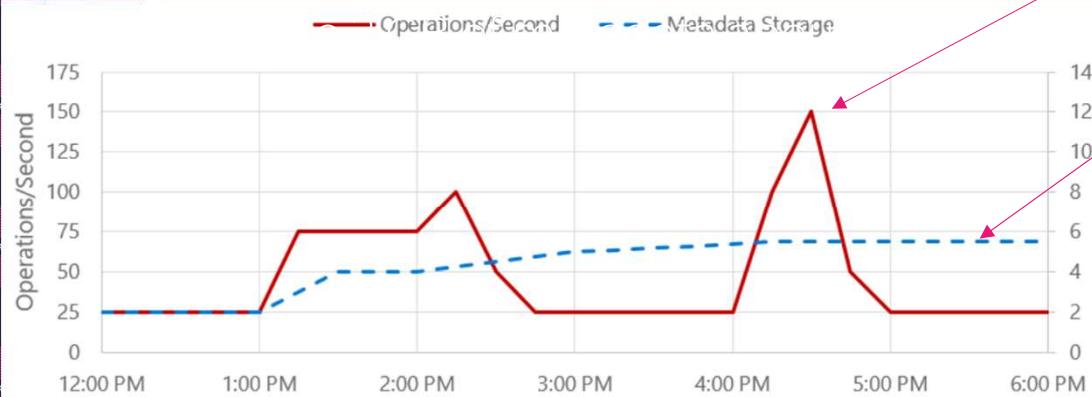
# Microsoft Purview



# 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

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



#ScottishSummit2022



# 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



# 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

Elastic Data Map

1CPU x €0.391 X 730 hours

€ 285,43

Scanning and Classification

4 scans x 2 hours x 32 vCore x €0.600 per vCore per hour

€ 153,60

Other Features

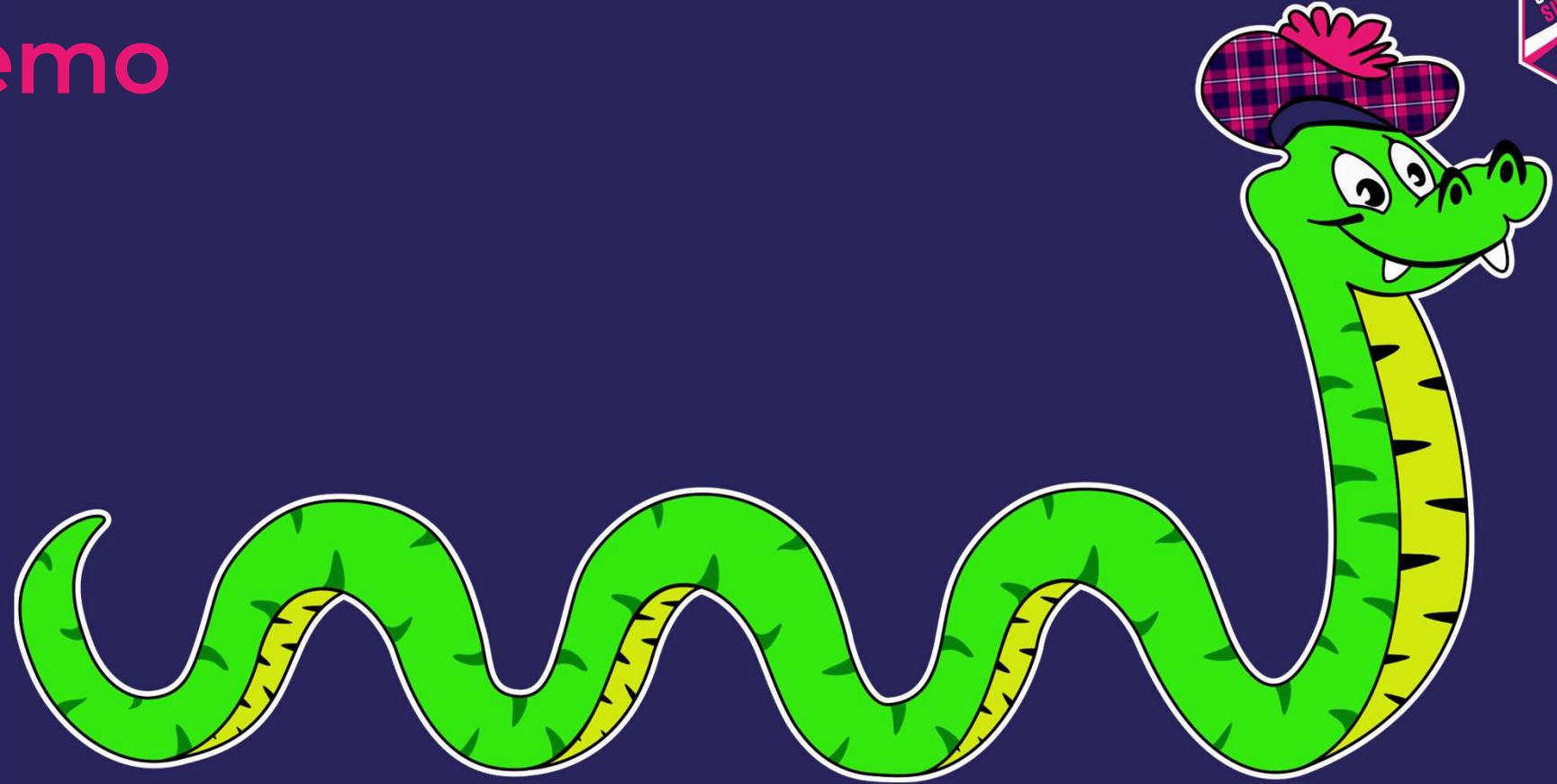
4 scans x 1 hour x €0.28 per vCore per hour

€ 0,80

€439,83

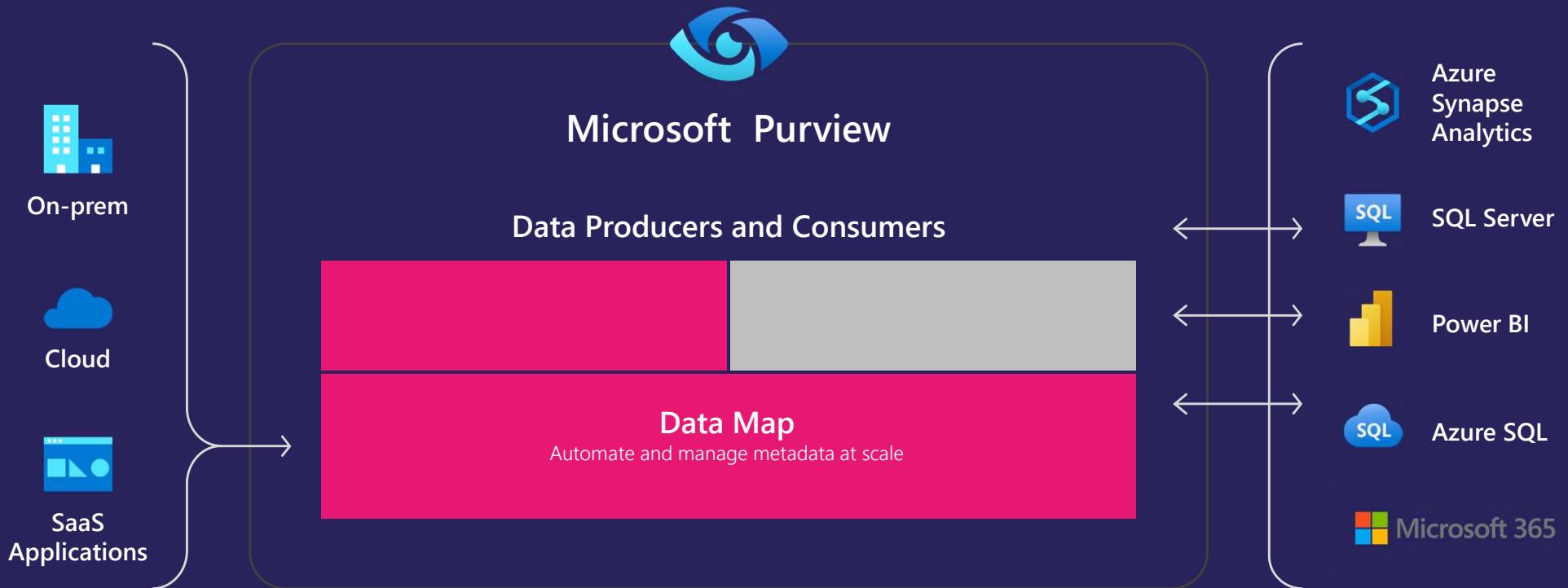


# Demo



#ScottishSummit2022

# Unified Data Governance

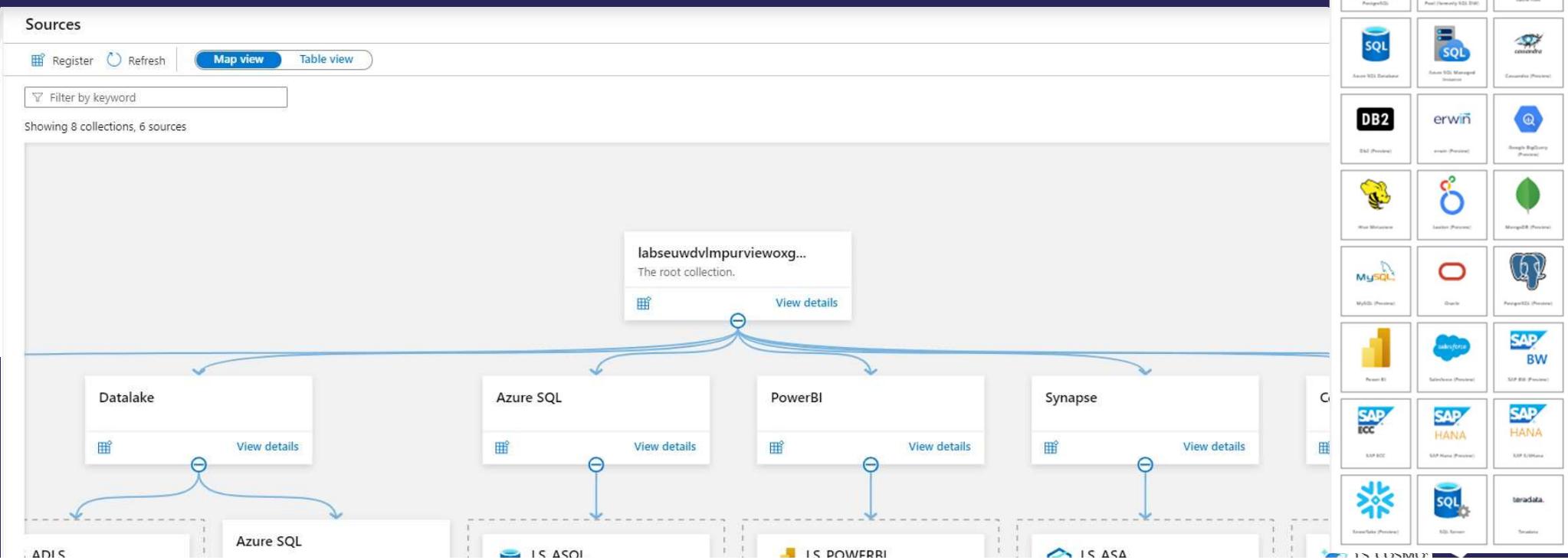


#ScottishSummit2022

## 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

<input 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

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



## Data Map

# Classifications

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

The screenshot shows the 'Classifications' section of the Data Map interface. On the left, a sidebar lists various management options: Sources, Collections, Monitoring, Source management, Scan rule sets, Pattern rules, Integration runtimes, Annotation management, Classifications (which is selected and highlighted in grey), and Classification rules. The main area is titled 'Classifications' and shows a list of system-provided classifications. It includes columns for 'Display name', 'Formal name', and 'Description'. The list contains entries such as ABA Routing Number, Age of an individual, Argentina National Identity (DNI) Number, Australia Bank Account Number, Australia Business Number, Australia Company Number, Australia Driver's License Number, Australia Medical Account Number, Australia Passport Number, Australia Tax File Number, Austria Driver's License Number, and Austria Identity Card, each with its corresponding formal name and description.

Display name	Formal name	Description
ABA Routing Number	MICROSOFT.FINANCIALUS.ABA_ROUTING_NUMB...	ABA Routing Number
Age of an individual	MICROSOFT.PERSONALAGE	Age of an individual
Argentina National Identity (DNI) Number	MICROSOFT.GOVERNMENT.ARGENITNA.DNI_NU...	Argentina National Identity (DNI) Number
Australia Bank Account Number	MICROSOFT.FINANCIALAUSTRALIA.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.LIC...	Austria Driver's License Number
Austria Identity Card	MICROSOFT.GOVERNMENT.AUSTRIA.IDENTITY.CA...	Austria Identity Card

The screenshot shows the schema view for the 'Cities' resource set in Azure Data Lake Storage Gen2. At the top, there are buttons for 'Edit', 'Select for bulk edit', 'Request access', 'Refresh', 'Delete', and 'Edit columns'. Below this, tabs for 'Overview', 'Properties', 'Schema' (which is selected and highlighted in blue), 'Lineage', 'Contacts', and 'Related' are visible. A search bar labeled 'Filter by name' is present. The main area displays 17 items, with headers for 'Column name', 'Classifications', 'Sensitivity label', and 'Glossary terms'. The 'Classifications' column shows that the 'CityName' column is classified as 'World Cities'. Other columns listed include 'CityID', 'LastEditedBy', 'LatestRecordedPopulation', and 'Location'.

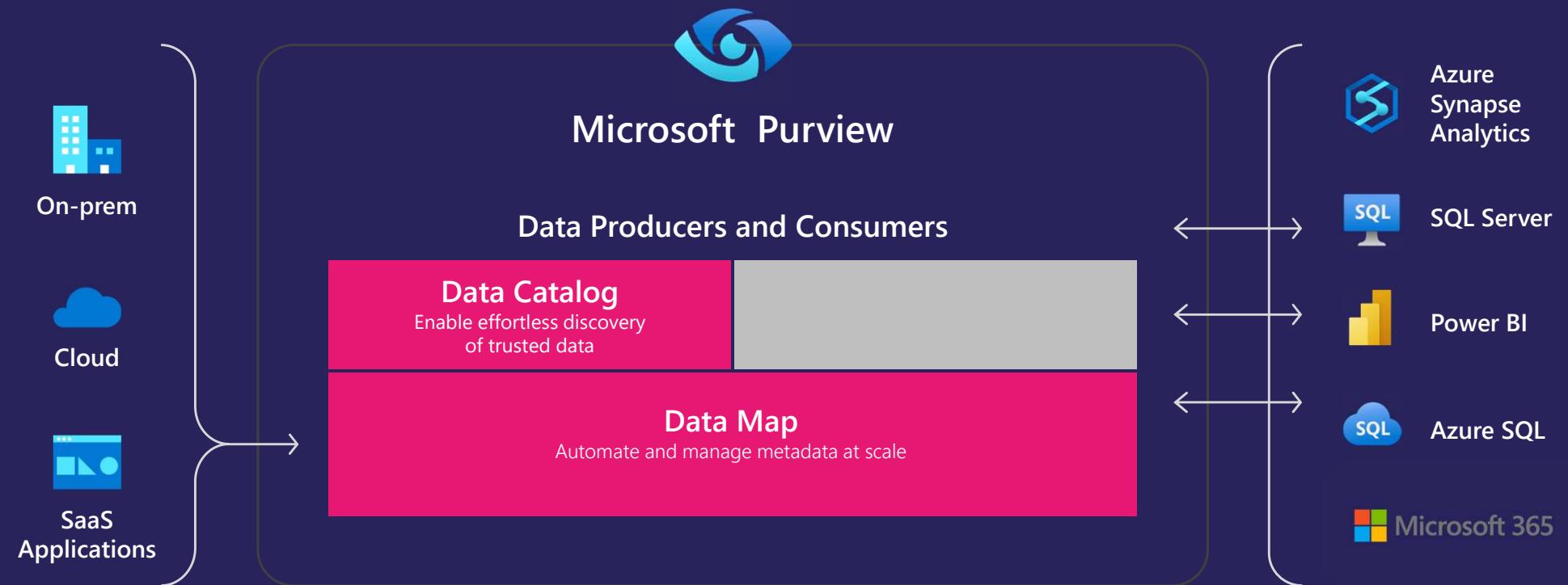
Column name	Classifications	Sensitivity label	Glossary terms
CityID			
CityName	World Cities		
LastEditedBy			
LatestRecordedPopulation			
Location			

#ScottishSummit2022



Microsoft Purview

# Unified Data Governance



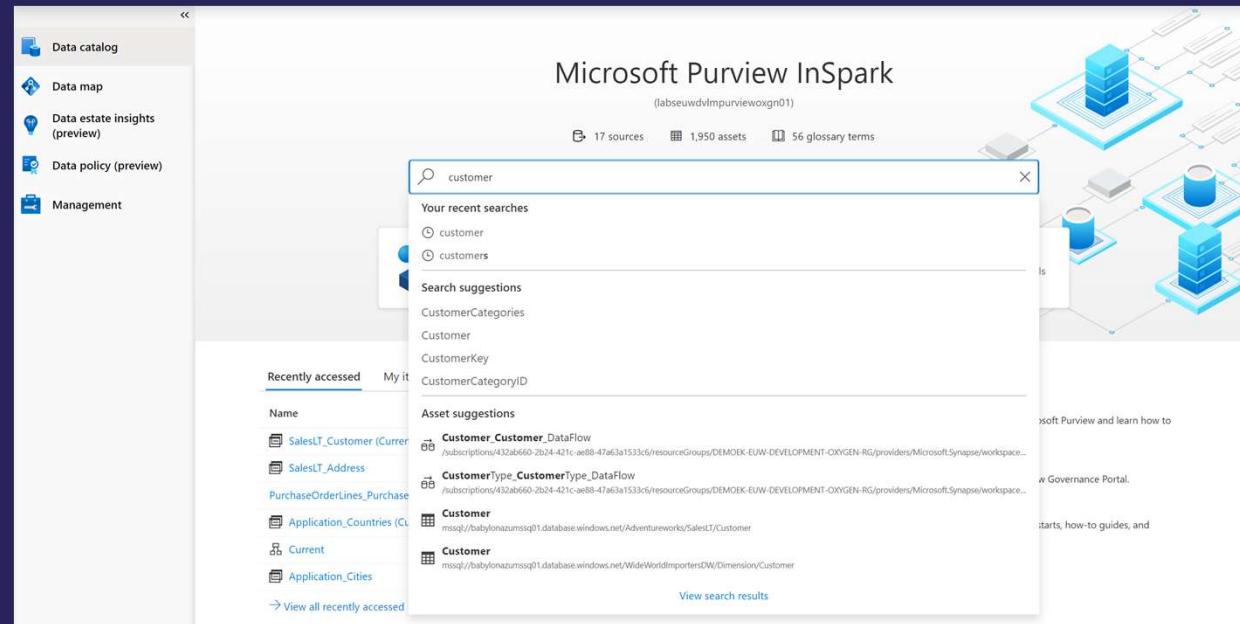
#ScottishSummit2022



## Data Catalog

# Search

- Semantic search and browse
- Business glossary and workflows



The screenshot shows the Microsoft Purview InSpark interface. On the left, a sidebar menu includes "Data catalog", "Data map", "Data estate insights (preview)", "Data policy (preview)", and "Management". Below this are sections for "Recently accessed" and "My it". The main area features a search bar with the placeholder "customer", showing recent searches for "customer" and "customers", and suggestions like "CustomerCategories", "Customer", "CustomerKey", and "CustomerCategoryID". It also lists asset suggestions such as "Customer\_Customer\_DataFlow", "CustomerType\_CustomerType\_DataFlow", "Customer", and "Customer". A large, colorful diagram on the right illustrates data flow and connectivity between various data sources and assets.

#ScottishSummit2022



## Data Catalog

# Assets

- Semantic search and browse
- Business glossary and workflows

The screenshot displays the Microsoft Data Catalog interface. At the top, a navigation bar includes links for "Data catalog", "Data map", "Data estate insights (preview)", "Data policy (preview)", and "Management". Below this, a main content area titled "Browse assets" shows a grid of data source types. The "Azure SQL Database" item is highlighted with a blue border. The interface then transitions into a detailed view of the "WideWorldImporters" database under the "Azure SQL Database" category. This detailed view shows a tree structure of database objects: Adventureworks, Datastore, Datastore, datastore, oxygenconfigdb, test\_purview, and WideWorldImporters. The "WideWorldImporters" node is expanded, revealing its schema: Application, dbo, Purchasing, Sales, and Warehouse. To the right, a table lists the objects under "WideWorldImporters" along with their types:

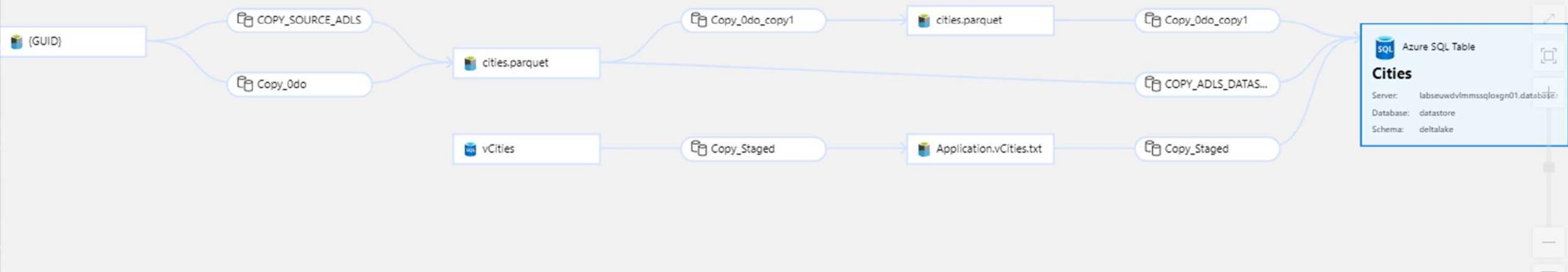
Name	Type
PurchaseOrderLines	Azure SQL Table
PurchaseOrders	Azure SQL Table
SupplierCategories	Azure SQL Table
SupplierCategories_Archive	Azure SQL Table
Suppliers	Azure SQL Table
Suppliers_Archive	Azure SQL Table
SupplierTransactions	Azure SQL Table



## Data Catalog

# Lineage

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

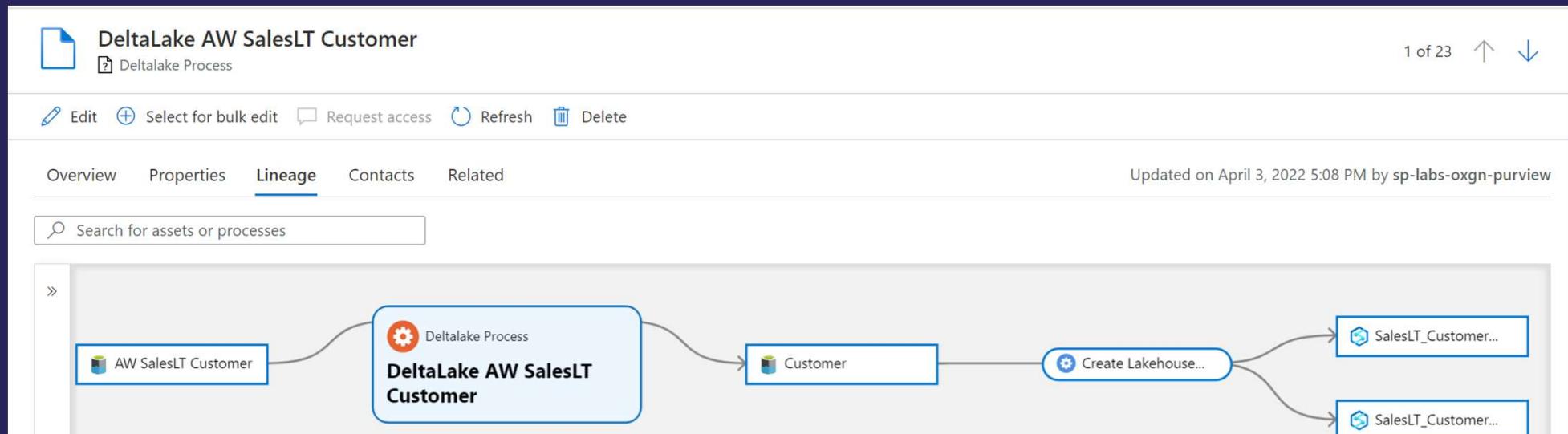




## Data Catalog

# Custom Lineage

- Create custom Lineage with Apache Spark 2.2
- <https://github.com/wjohnson/pyapacheatlas>

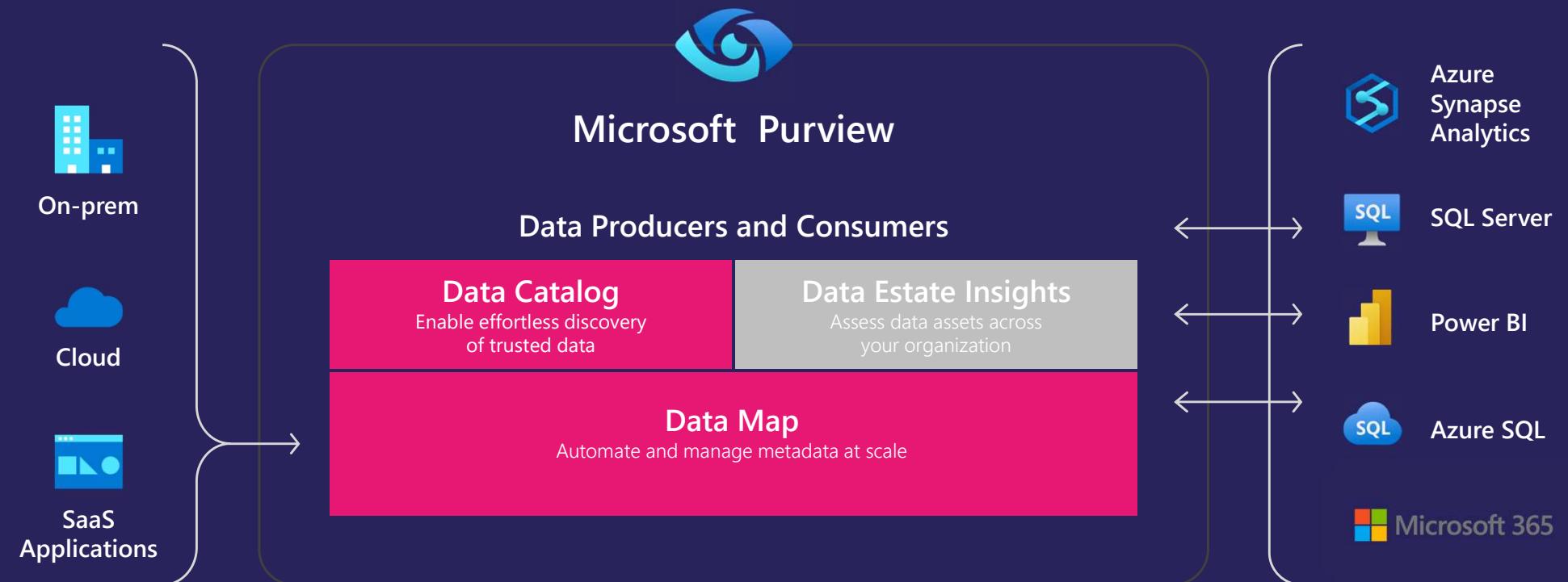


#ScottishSummit2022



Microsoft Purview

# Unified Data Governance

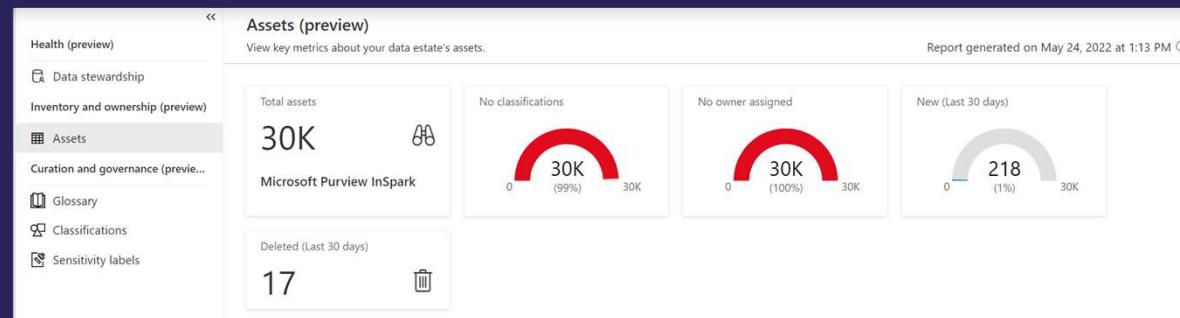
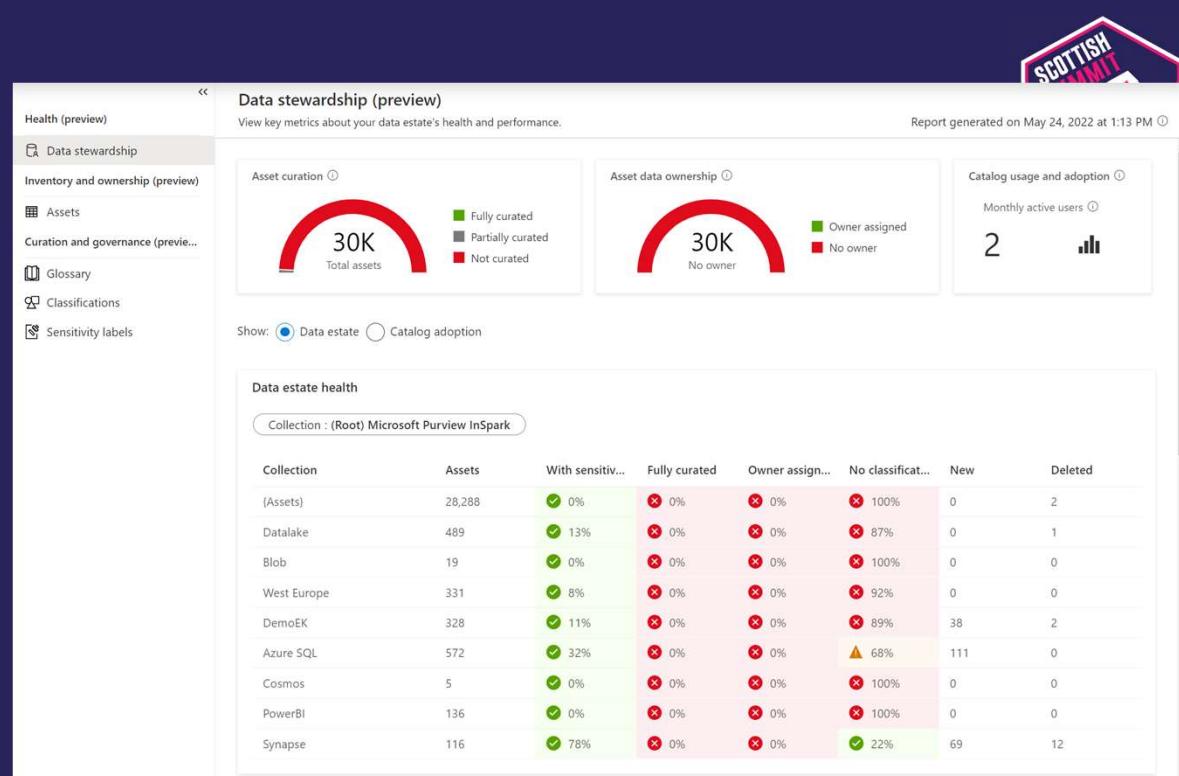


#ScottishSummit2022

## Data Estate Insight (preview)

# Metrics

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



#ScottishSummit2022

# ACCESS CONTROL





## Data Policy (preview)

# Data Owner

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

The screenshot shows the "Access Control Policy" page in the Azure Data Catalog. The left sidebar includes options for Data catalog, Data map, Data estate insights (preview), Data policy (preview) (which is 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." It then asks for a "Name" (Policy Name) and a "Description" (Describe the policy). Below this is a "Policy statements" section with an example: "Allow Read on data asset finance to group non-FTEs." A button "+ New policy statement" is available. A tooltip over the "Allow" dropdown shows options: Action (Allow, Deny), Data resources (data resources), Subjects (subjects). The "Data resources" section allows selecting a type (Data sources, Data sets, Data flows) and a specific source (Azure Subscription, Azure Resource Group, Azure Blob Storage, Azure Data Lake Storage Gen2, Azure SQL Server). The "Assets" section shows an asset named "labseuwvd1mdloxn02" of type "Azure Data Lake Storage Gen2".

#ScottishSummit2022

## Data Policy (preview)

# Self Service

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

The screenshot shows a user interface for creating a new data catalog workflow. On the left, a sidebar lists navigation options: Data catalog, Data map, Data estate insights (preview), Data policy (preview), and Management. The 'Management' option is highlighted. The main area is titled 'New data catalog workflow' and contains a heading 'Select a workflow template to get started:' followed by four listed options:

- 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



#ScottishSummit2022

# 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 Azure Synapse Analytics integration blade. On the left is a sidebar with the following navigation items:

- Analytics pools
- SQL pools
- Apache Spark pools
- Data Explorer pools (preview)
- External connections
- Linked services
- Microsoft Purview
- Integration
- Triggers
- Integration runtimes
- Security
- Access control
- Credentials
- Managed private endpoints
- Configurations + libraries
- Workspace packages
- Data flow libraries (preview)
- Apache Spark configurations
- Source control
- Git configuration

The main content area is titled "Microsoft Purview" and contains the following information:

- 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) [Edit](#) [Refresh](#) [Disconnect](#)
- 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	<span>Connected</span>

- Lineage status details (Status: Succeeded)
- Lineage status details (Status: Unsupported)

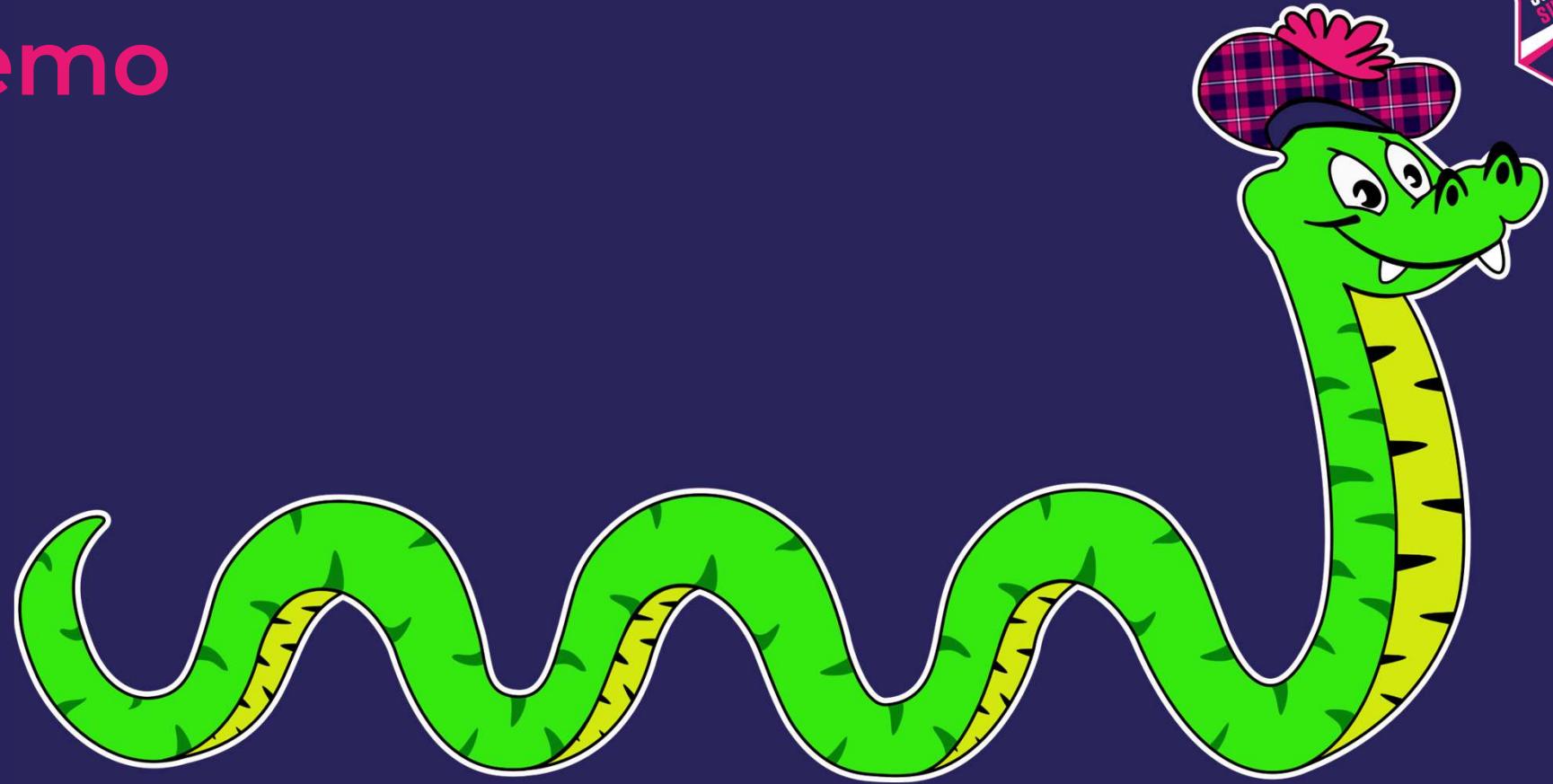
Details: The lineage of this copy activity is not reported because it went through the following unsupported scenario: stored procedure or query used in SQL source.

[Learn more about Microsoft Purview integration](#)

#ScottishSummit2022



# Demo



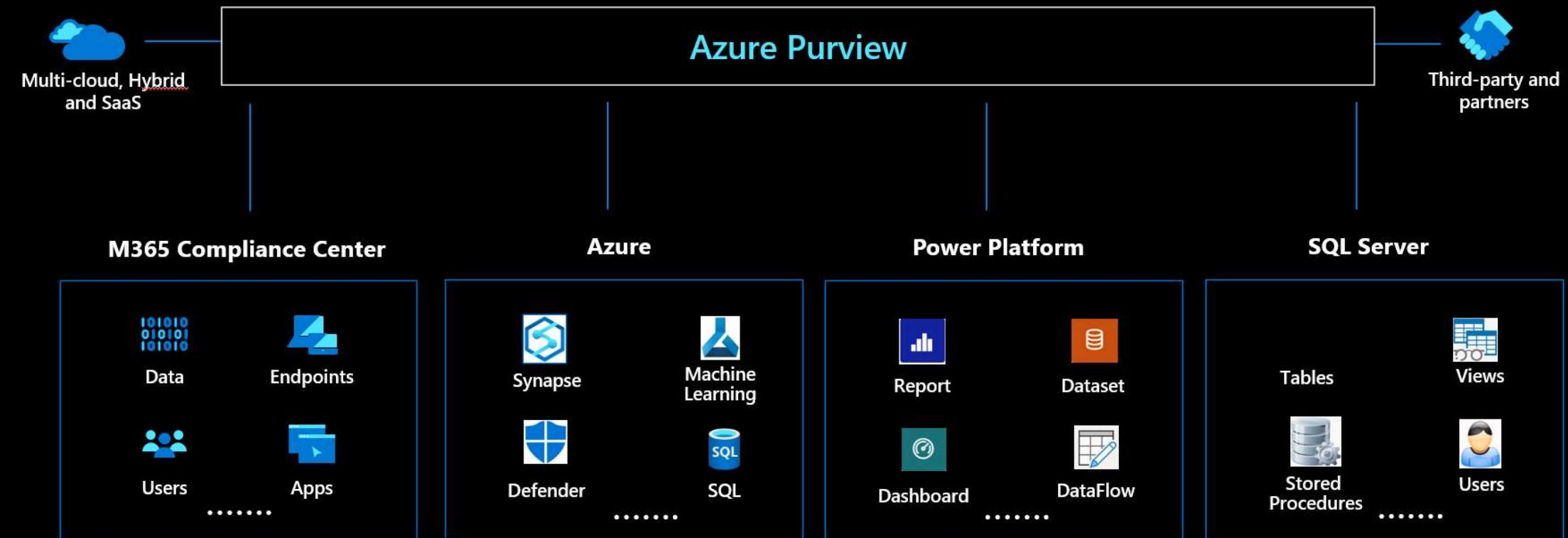
#ScottishSummit2022



# Recap

- Central location to register data sources
- Central location for documentation
- Easy overview of owners and expert of data
- Location to store your business glossary and to apply them to your data assets
- Automated scanning and classification of data, labelling of sensitive data
- Overview of which people or applications has access to which data assets and with which permissions
- Insight on your Data Estate

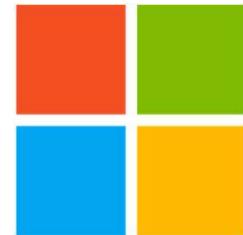
# Data Governance Opportunity





# Thank You to our Sponsors...

Event Sponsor



# Microsoft

Platinum Sponsors



the results company



#ScottishSummit2022



# Thank You to our Sponsors...

## Gold Sponsors



Data Quality



Event Lunch



Accessibility



Data Analytics



#ScottishSummit2022



# Feedback in the app!!



erwindekreuk/  
Erwindekreuk.com  
@erwindekreuk

A white rectangular card with a black header and footer. It features a circular profile picture of a man, the name "Erwin de Kreuk", and his title "Principal Consultant | Lead Data and AI | Data Platform MVP | Publ...". A QR code is at the bottom right.

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

A standard black and white QR code.

# Questions?

#ScottishSummit2022