

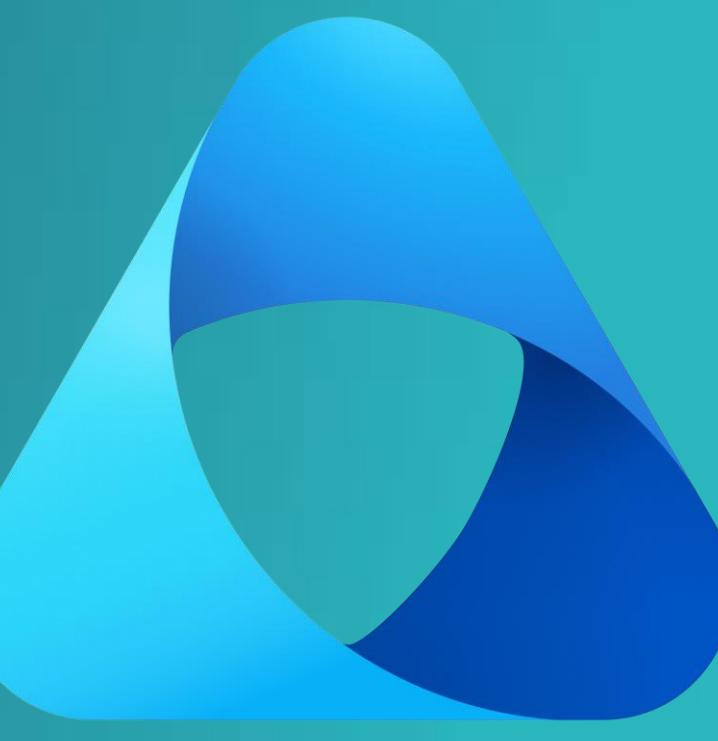


Feedback →



The Crucial Role of Data Quality in Your Data Estate

Innovate to Accelerate



We Are InSpark

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

Erwin de Kreuk

- ★ Principal Consultant
- ★ Lead Data & AI InSpark

Let's connect



@erwindekreuk



linkedin.com/in/erwindekreuk



erwindekreuk.com



github.com/edkreuk



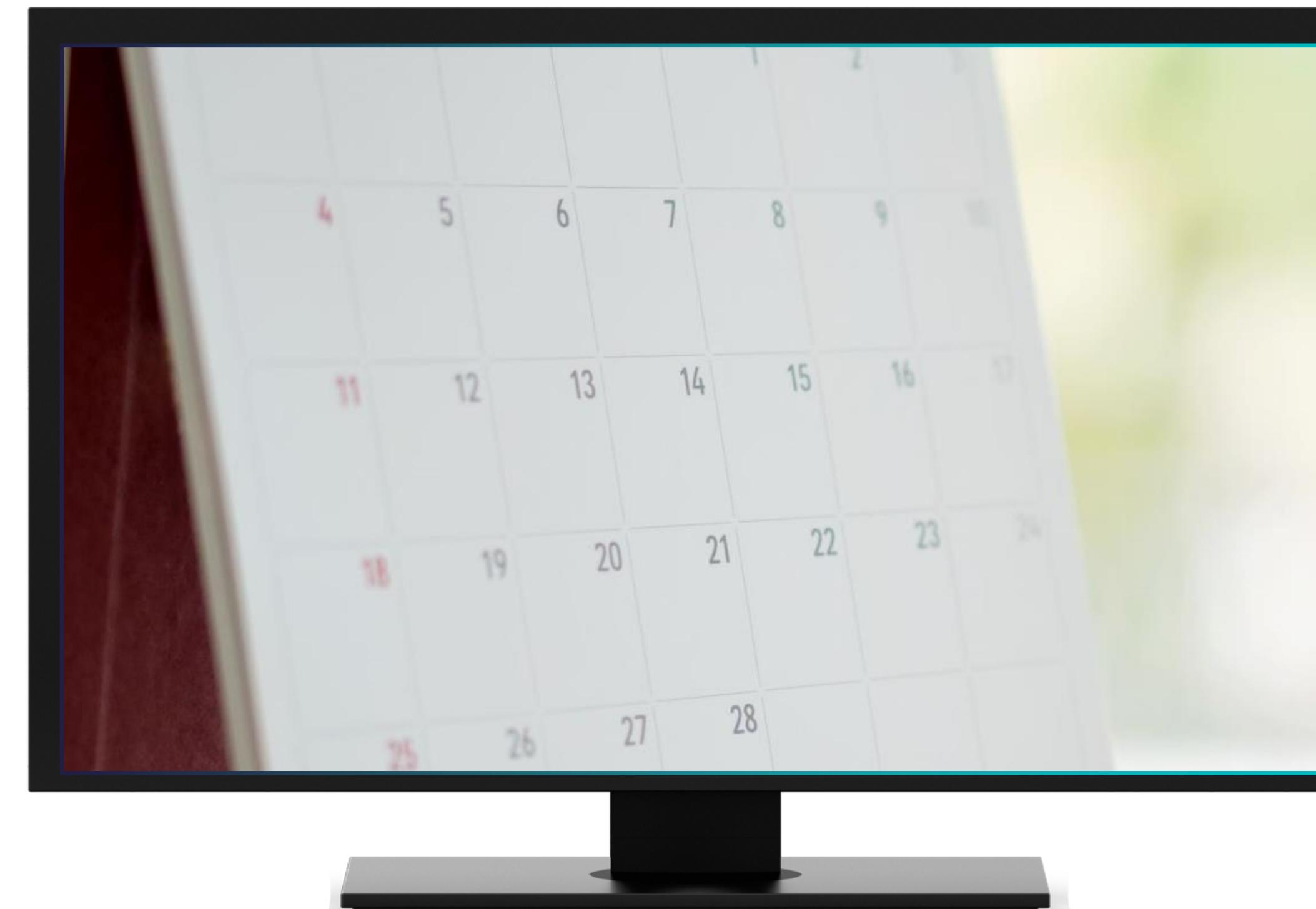
<https://sessionize.com/erwin-de-kreuk/>



Microsoft®
Most Valuable
Professional

Agenda

- ★ **Introduction**
- ★ **Data Quality**
- ★ **Data Cleansing**
- ★ **Microsoft Purview**
 - ★ Data Map
 - ★ Governance Domains and Products
 - ★ Data Quality
 - ★ Pricing
- ★ **Conclusion**
- ★ **Q & A**



Introduction

★ Impact on AI-driven Insights

- ★ High-quality data ensures accurate insights
- ★ Poor data quality leads to incorrect conclusions

★ Decision-Making

- ★ Reliable data supports better decision-making
- ★ Inaccurate data can result in poor decisions

★ Data-Driven Landscape

- ★ Data quality is crucial in a data-driven environment
- ★ Ensures the effectiveness of AI applications



sax now

Nieuws Achtergrond Mensen Opinie Studium Generale Video Over ons

Fout in excel heeft volgen; 875 medewerkers kregen in december te weinig geld

Bas Klaassen
18 januari 2023

Honderden medewerkers in december uitgekeerd de salarisbetaling kregen met het januarisalaris medewerkers kregen afgelopen woensdag

Alle medewerkers waar het Renske Bouwmeester, team volgens haar voor het overg enkele honderden euro's, m

345 medewerkers kregen te betaling van januari. Bouwmeester terugstorten. Waar nodig is medewerkers kregen juist te nabetaling gehad.

Vervelend

Bouwmeester noemt de problemen door een fout in de

Spreadsheet errors can have disastrous consequences – yet we keep making same mistakes

Published: January 25, 2024 5:01pm CET

Making mistakes with spreadsheets can not only cause us personal frustration but can also lead to some very serious consequences. In fact, it's estimated that 90% of spreadsheets contain at least one major mistake.

The dark matter of corporate IT

The above is just a fraction of the spreadsheet errors that are regularly made by professionals and hobbyists. Met Excel is rekenkracht, in combinatie met de flexibiliteit voor de boekhouding, voor financiële rapportage en projectmanagement.

Spreadsheets represent unknown risks in the form of errors, privacy violations, trade secrets and compliance violations. Yet they are also critical for the way organisations make their decisions. For this reason, they have been described by experts as the "dark matter" of corporate IT.

Industry studies show that 90% of spreadsheets containing more than 150 rows have at least one major mistake.

This is understandable because spreadsheet errors are easy to make but difficult to spot. My own research has shown that inspecting the spreadsheet's code is the most effective way of debugging them, but this approach still only catches between 60% and 80% of all errors.

EXCELWEB.NL
DE WEBSITE MET ALLES OVER EXCEL!

Home Kennisbank Formule Grafieken Tips Macro Download Contact

Home > Macro > Algemeen > Een foutje gemaakt in Excel: wie is verantwoordelijk voor de schade?

Een foutje gemaakt in Excel: wie is verantwoordelijk voor de schade?

Excel is een blijft een onmisbaar hulpmiddel voor professionals en hobbyisten. Met Excel is rekenkracht, in combinatie met de flexibiliteit voor de boekhouding, voor financiële rapportage en projectmanagement.

Een klein foutje is echter snel gemaakt en kan grote gevolgen hebben. We bespreken de juridische fouten in Excel, waarbij het noodzakelijk is om verschillende scenario's van elkaar te onderscheiden.

Excel-fout op het werk

Een eerste situatie: een werknemer maakt een fout in een bestand en veroorzaakt hierdoor schade aan een klant van de werkgever.

Neem bijvoorbeeld een accountant die via een ongeluk een fout maakt in een Excel-bestand. De financiële rapportage onjuist blijkt te zijn en de werkgever neemt verkeerde beslissingen.

Volgens het Nederlands recht is de werkgever verantwoordelijk voor fouten die worden gemaakt in het werk.

Fannie Mae Corrects Mistakes In Results

By Jonathan Glater
Oct. 30, 2003

Fannie Mae announced yesterday that it had corrected errors in its most recent financial results, which in some cases varied from the correct amounts by more than \$1 billion. The company attributed the errors to flawed application of new accounting standards.

As a result of the corrections reported by the company, which is the nation's largest buyer of home mortgages, its mortgage portfolio grew by \$1.7 billion; its total assets by \$1.04 billion; and its unrealized gains on certain securities, by \$1.3 billion. The changes do not affect the company's income statement, Fannie Mae said.

"In adopting a new and complex accounting standard in a short period of time, Fannie Mae had to put in place a system and process to capture all open commitments and mark them to market," the company said in a statement. "To implement this standard, Fannie Mae utilized information from its internal, automated systems in conjunction with spreadsheets that made additional calculations necessary under the new rule." A spokeswoman said that one of those spreadsheets contained an error.

Common Examples of Poor Data Quality

- ★ Poor decision-making
- ★ Business Inefficiencies
- ★ Missed Opportunities/Lost Reputation
- ★ Siloed Information
- ★ Lost Revenue



Data Quality Problems

Example of Poor Quality Data

ID	Last Name	First Name	Street	City	State	Zip	Phone	Fax	E-mail
113	Smith		123 S. Main	Denver	CO	80210	(303) 777-1258	(303) 777-5544	ssmith@aol.com
114	Jones	Jeff	12A	Denver	CO	80224	(303) 666-6868	(303) 666-6868	(303) 666-6868
115	Roberts	Jenny	1244 Colfax	Denver	CO	85231	759-5654	853-6584	jr@msn.com
116	Robert	Jenny	1244 Colfax	Denver	CO	85231	759-5654	853-6584	jr@msn.com

Annotations pointing to specific data points:

- 1. Missing information (no first name)
- 2. Incomplete information (no street)
- 3. Probable duplicate information (similar names, same address, phone number)
- 4. Potential wrong information (are the phone and fax numbers the same or is this an error?)
- 5. Inaccurate information (invalid e-mail)
- 6. Incomplete information (missing area codes)

Benefits of Clean, Well-Structured Data

- ★ Enhanced Decision-Making
- ★ Increased Efficiency / Reduced Costs
- ★ Improved Customer Experience
- ★ Enhanced Analytics and AI
- ★ Improved Collaboration
- ★ Increased Trust / Better Compliance



The Role of a Powerful Data Platform

- ★ **Importance of a Powerful Data Platform**
 - ★ Ensures high data quality
 - ★ Maintains data consistency and accuracy
- ★ **Benefits of a Robust Data Platform**
 - ★ Data readily available for various applications
 - ★ Supports effective data quality measures

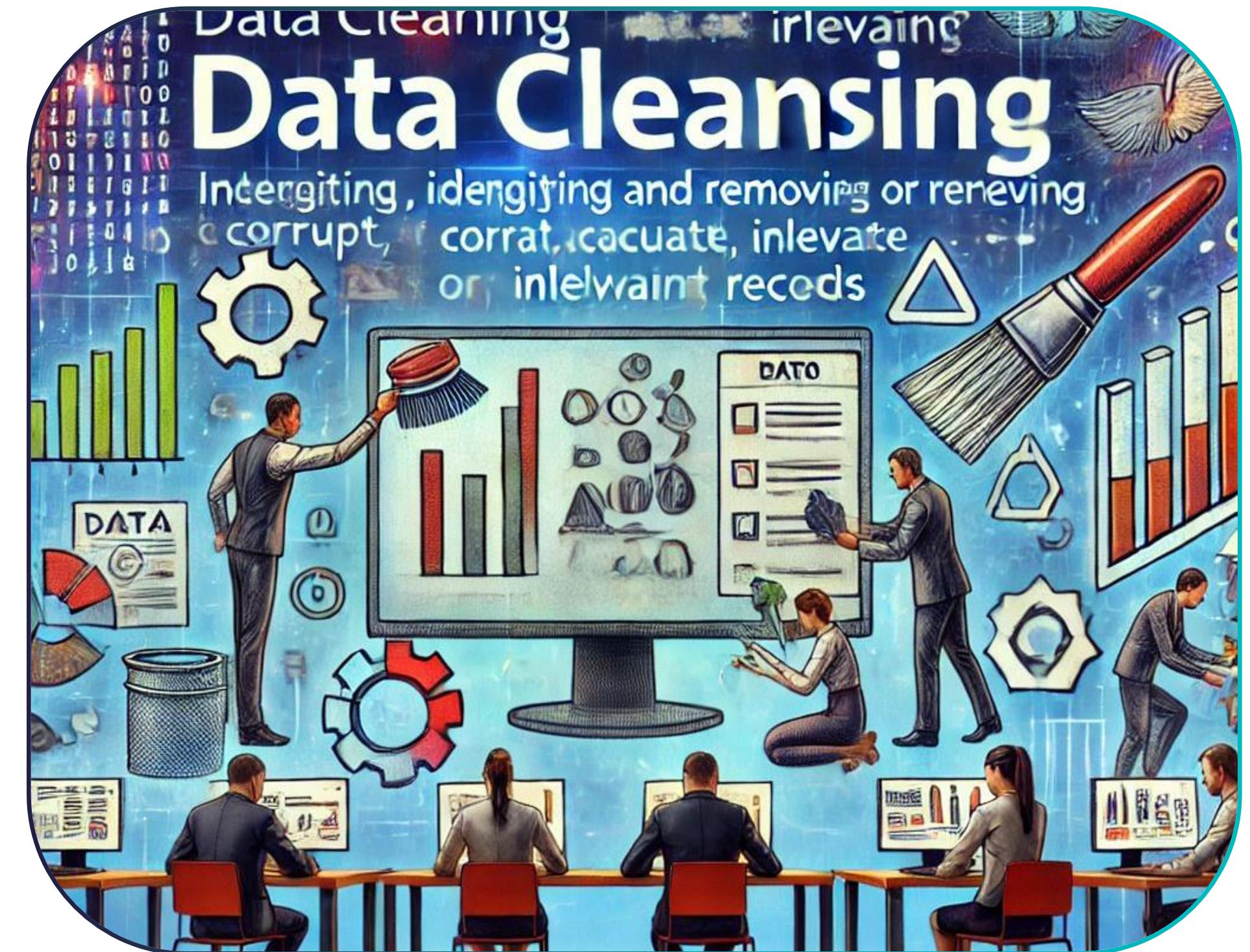


Data Cleansing

- ★ **Definition of Data Cleansing**
 - ★ Also known as data cleaning or data scrubbing
 - ★ Involves identifying and correcting or removing corrupt, inaccurate, or irrelevant records

- ★ **Purpose of Data Cleaning**
 - ★ Ensures data is accurate, complete, and reliable
 - ★ Eliminates errors and inconsistencies

- ★ **Methods of Data Cleaning**
 - ★ Interactive data wrangling tools
 - ★ Batch processing



Benefits of Data Cleansing



- ★ Improved Decision Making
- ★ Increased Efficiency
- ★ Cost Savings
- ★ Enhanced Customer Satisfaction
- ★ Regulatory Compliance
- ★ Better Analytics

Common Examples of Data Cleansing



- ★ Remove Duplicates
- ★ Correcting Errors
- ★ Standardizing Formats
- ★ Filling in Missing Values
- ★ Removing Irrelevant DataLost Reputation
- ★ Validating Data
- ★ Updating Outdated Information

Microsoft Purview

The screenshot displays the Microsoft Purview web interface. At the top, there's a dark blue header bar with the Microsoft Purview logo on the left, a search bar in the center, and a Copilot button on the right. The main content area has a pink gradient background. On the left, a sidebar lists various features: Home, Solutions, Learn, Settings, Data Map, Unified Catalog, Data policy, and Insider Risk Management. The main content area features a large section titled "Work faster and smarter with Copilot in Microsoft Purview" with a "Get started" button. To the right, two cards show how Copilot integrates with Data Loss Prevention and eDiscovery.

Microsoft Purview

Search

Copilot

Home

Solutions

Learn

Settings

Data Map

Unified Catalog

Data policy

Insider Risk Management...

Work faster and smarter with Copilot in Microsoft Purview

Discover, analyze, and understand data faster with the power of AI.

Get started

Copilot

Alert summaries in Data Loss Prevention

Organize, prioritize, and speed up your alert handling process.

Learn more

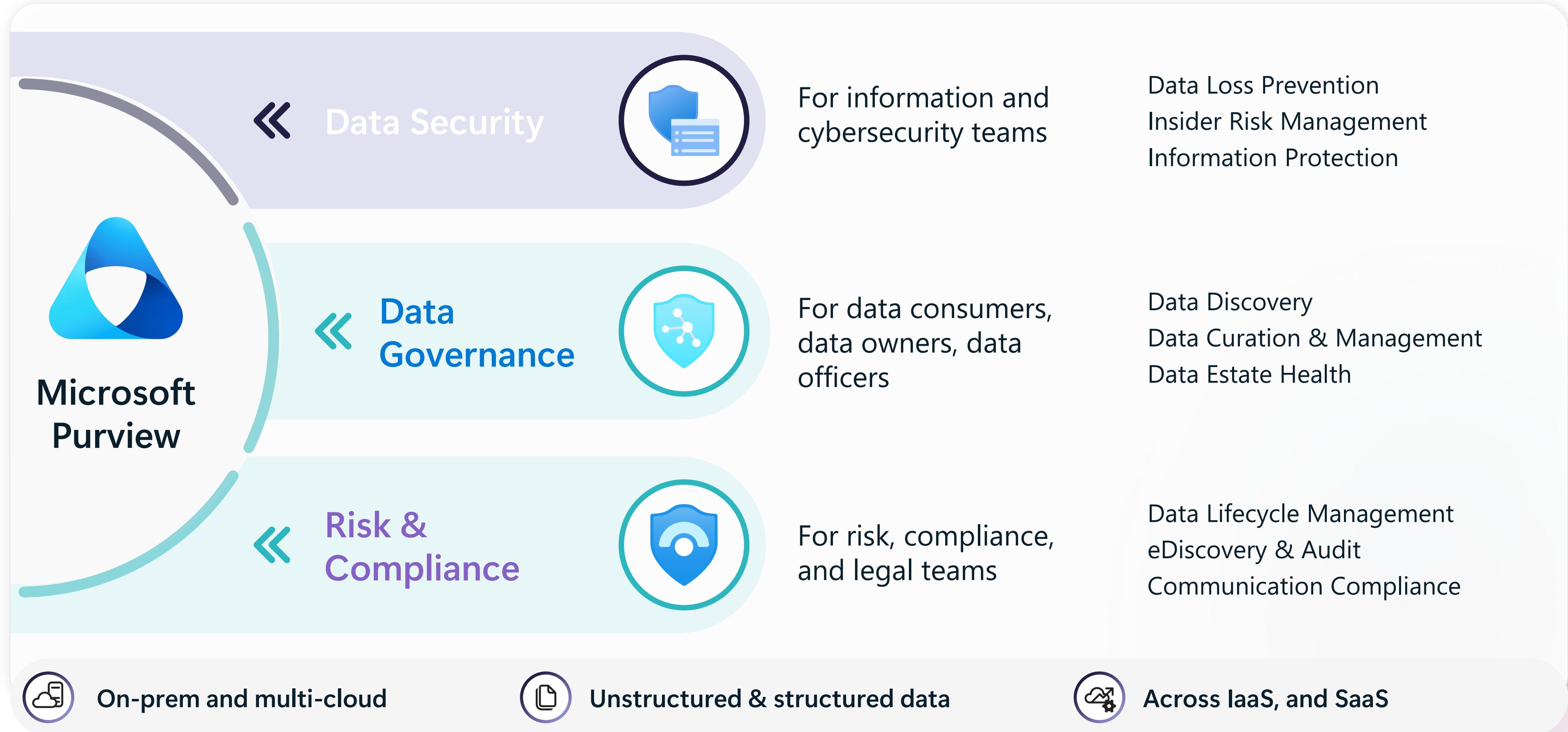
Copilot

Document summaries in eDiscovery

Improve the efficiency and accuracy of your document review process.

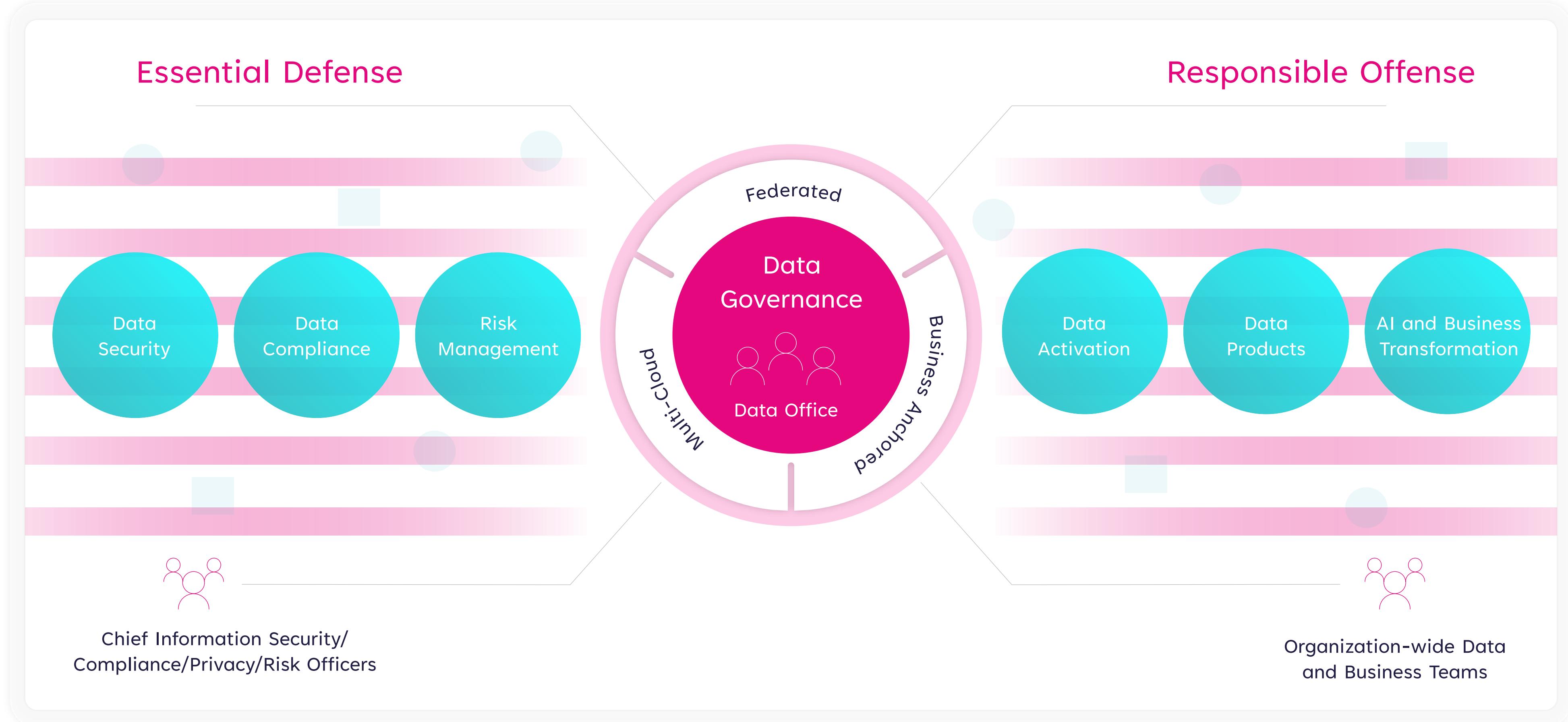
Learn more

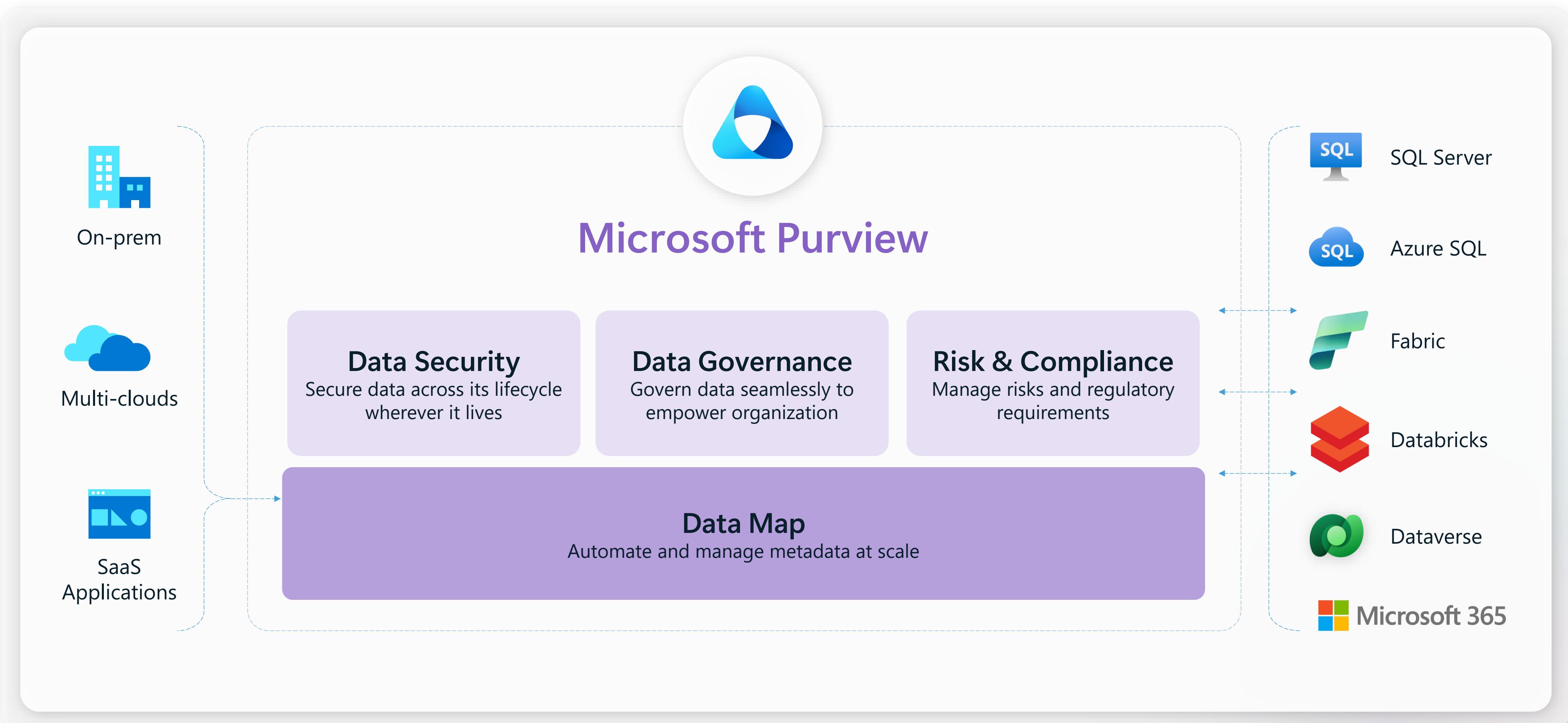
Bringing data security, governance, and compliance together



Data Governance for the Age of AI

The whole is greater than the sum of the parts





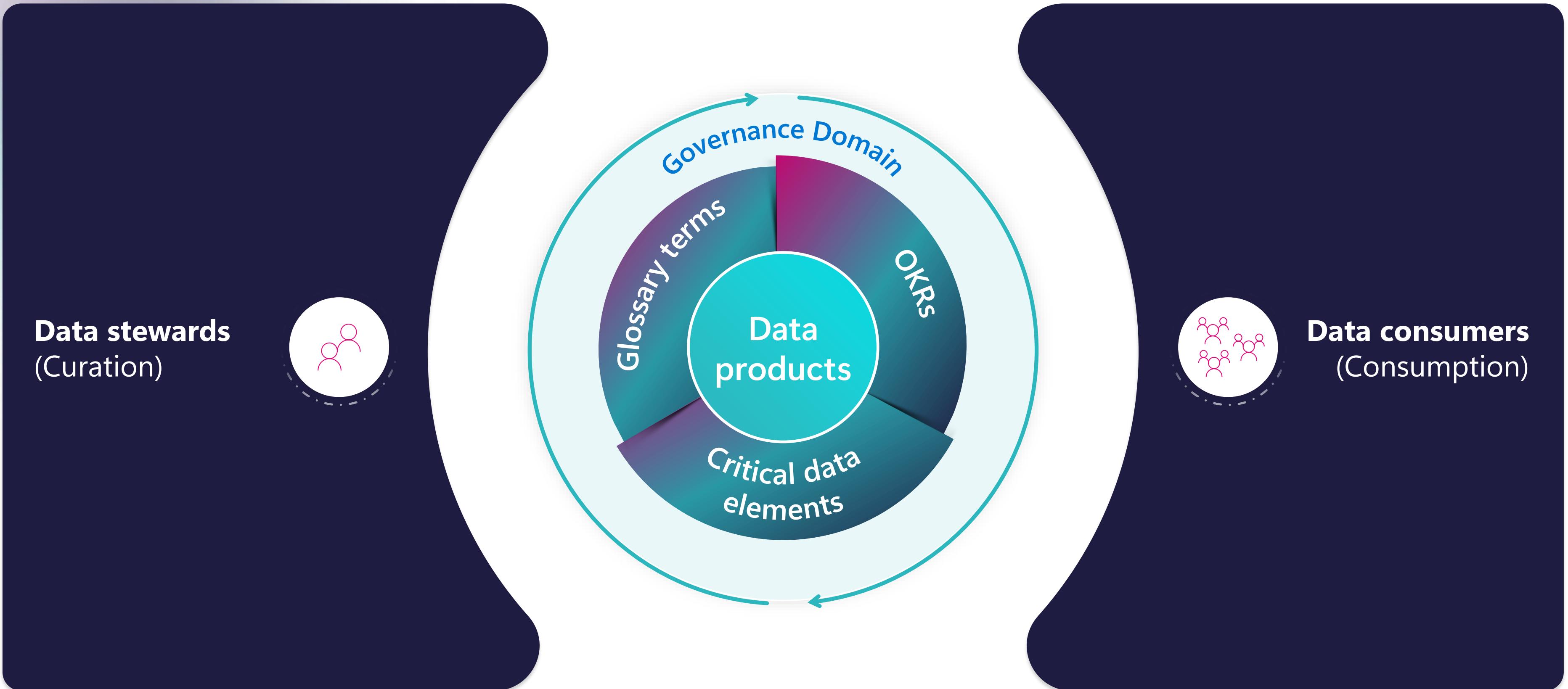
DATA GOVERNANCE FOR THE AGE OF **AI**

Your AI is only as good as your data

Your AI potential anchors on your people being able to responsibly use
your data

Data Governance makes your data ready for your people and your AI

Know your data with business concepts



Domains

Microsoft Purview

Search

Copilot 4

Home

Data Map

Domains

Data sources

Monitoring

Source management

Annotation management

Related solutions

Unified Catalog

Data policy

Insider Risk Management...

Domains

Create domains that match your org's key business segments as the top-level collections. With a collection, you can take action on all of its content at once.

Filter by name

ededeunpview... 3 collections ▾

Databases

Azure SQL

Fabric

ededeunpview01 (Default)

Domain

+ New collection Edit Refresh

Updated on December 9, 2024, 10:13 PM

Overview Role assignments

Description

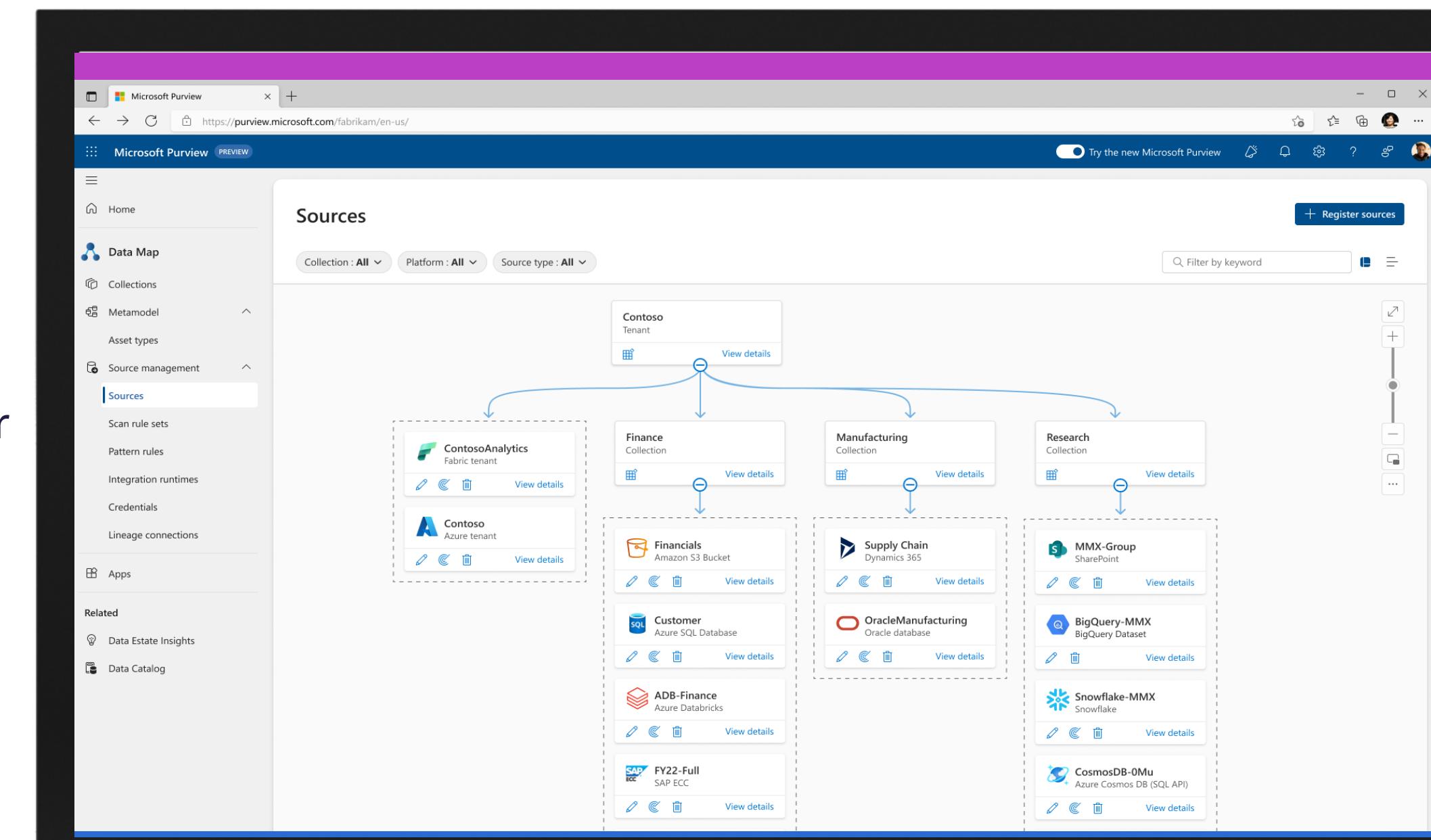
The default container.

Assets	Data sources	Scans
415	2	3

Data Map

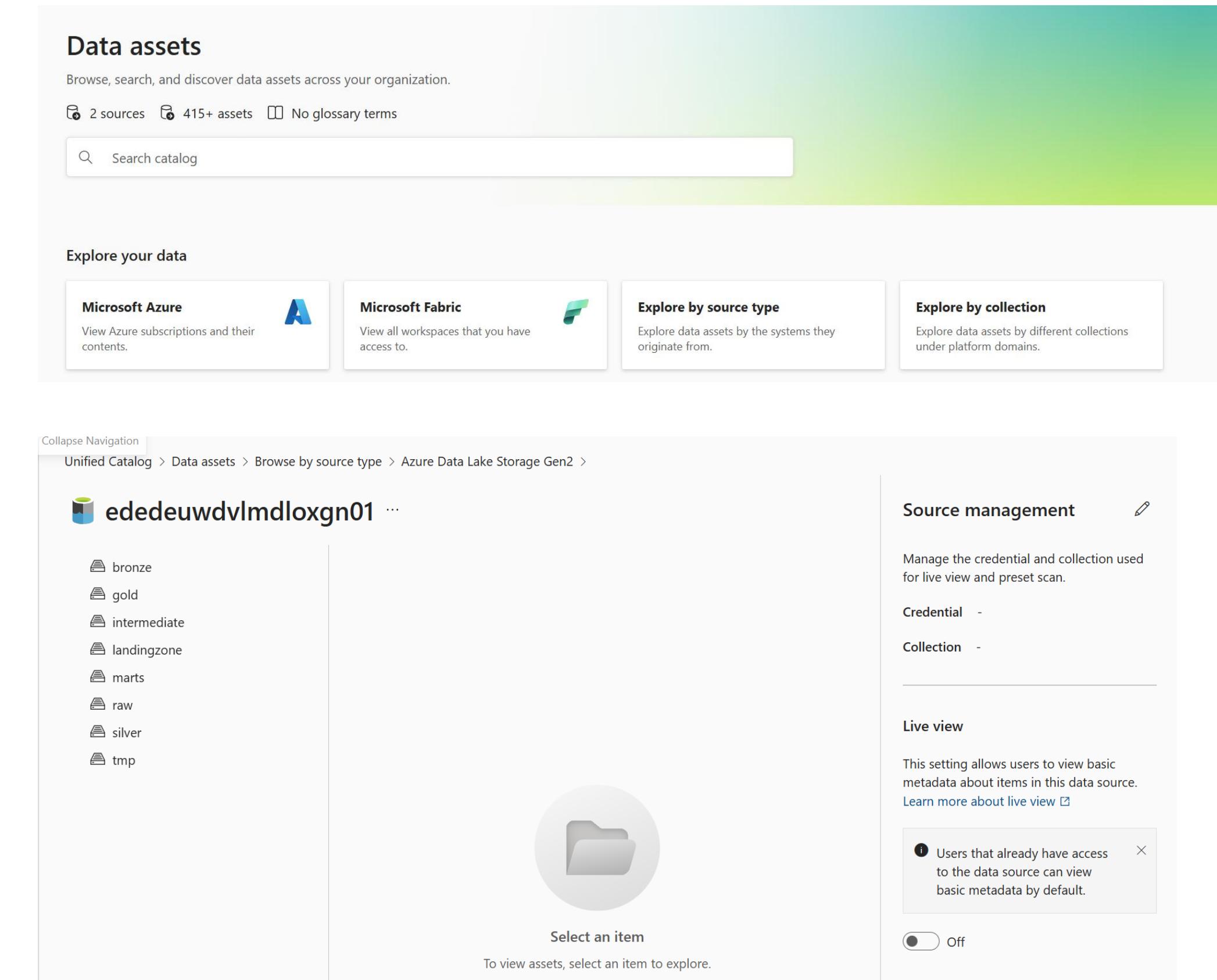
Create a unified map of data across your entire data estate by scanning and classifying all your data assets.

Scan Fabric tenants to discover data assets across your Fabric landscape with data discovery, sensitive data classification, and end-to-end data lineage.



Live View

- ★ Live view means that users with data access permissions can find these resources and their data assets in the catalog without setup or scanning.
- ★ These resources and their metadata are immediately available when you access the free or enterprise versions of Microsoft Purview, so you can start your governance journey quickly.



The screenshot shows the Microsoft Purview Data Catalog interface. At the top, there's a search bar labeled "Search catalog". Below it, a section titled "Explore your data" contains four cards: "Microsoft Azure" (View Azure subscriptions and their contents), "Microsoft Fabric" (View all workspaces that you have access to), "Explore by source type" (Explore data assets by the systems they originate from), and "Explore by collection" (Explore data assets by different collections under platform domains). The main content area is titled "Data assets" and includes a sub-section "Browse, search, and discover data assets across your organization". It shows statistics: 2 sources, 415+ assets, and No glossary terms. A navigation breadcrumb shows the path: Unified Catalog > Data assets > Browse by source type > Azure Data Lake Storage Gen2 >. Below this, a specific data asset named "ededeuwvdvldoxgn01" is listed with a preview icon. To the right of the asset, there are sections for "Source management" (Manage the credential and collection used for live view and preset scan) and "Live view" (This setting allows users to view basic metadata about items in this data source). A note says "Users that already have access to the data source can view basic metadata by default." At the bottom, there's a button labeled "Select an item" with the instruction "To view assets, select an item to explore." and a toggle switch labeled "Off".

Demo



Governance Domains

Business
Data

Microsoft Purview

Search

Home

Solutions

Learn

Settings

Data Catalog

Overview

Discovery

Catalog management

Governance domains

Data products

Data assets

Requests

Governance domains

New governance domain

Owner: All

Filter by keyword

10 governance domains

Corporate Functions

Information Technology

The screenshot shows the Microsoft Purview Data Catalog interface. The left sidebar has a navigation menu with icons and labels: Home, Solutions, Learn, Settings, Data Catalog (selected), and Audit. Under Data Catalog, 'Catalog management' is expanded, showing 'Governance domains' which is highlighted with a red box. Other options under Catalog management include 'Data products', 'Data assets', and 'Requests'. The main content area has a title 'Governance domains', a description about organizing data products into groups, and a 'New governance domain' button. It also features filters for 'Owner: All' and 'Filter by keyword'. Below that, it shows '10 governance domains' with categories like 'Corporate Functions' and 'Information Technology'.

Concept and Importance of Governance Domains

Definition of Governance Domains

- ★ Governance domains are specific areas of oversight that help manage data effectively within an organization.

Structuring Governance Efforts

- ★ These domains structure governance efforts, ensuring systematic management and clarity in data governance.

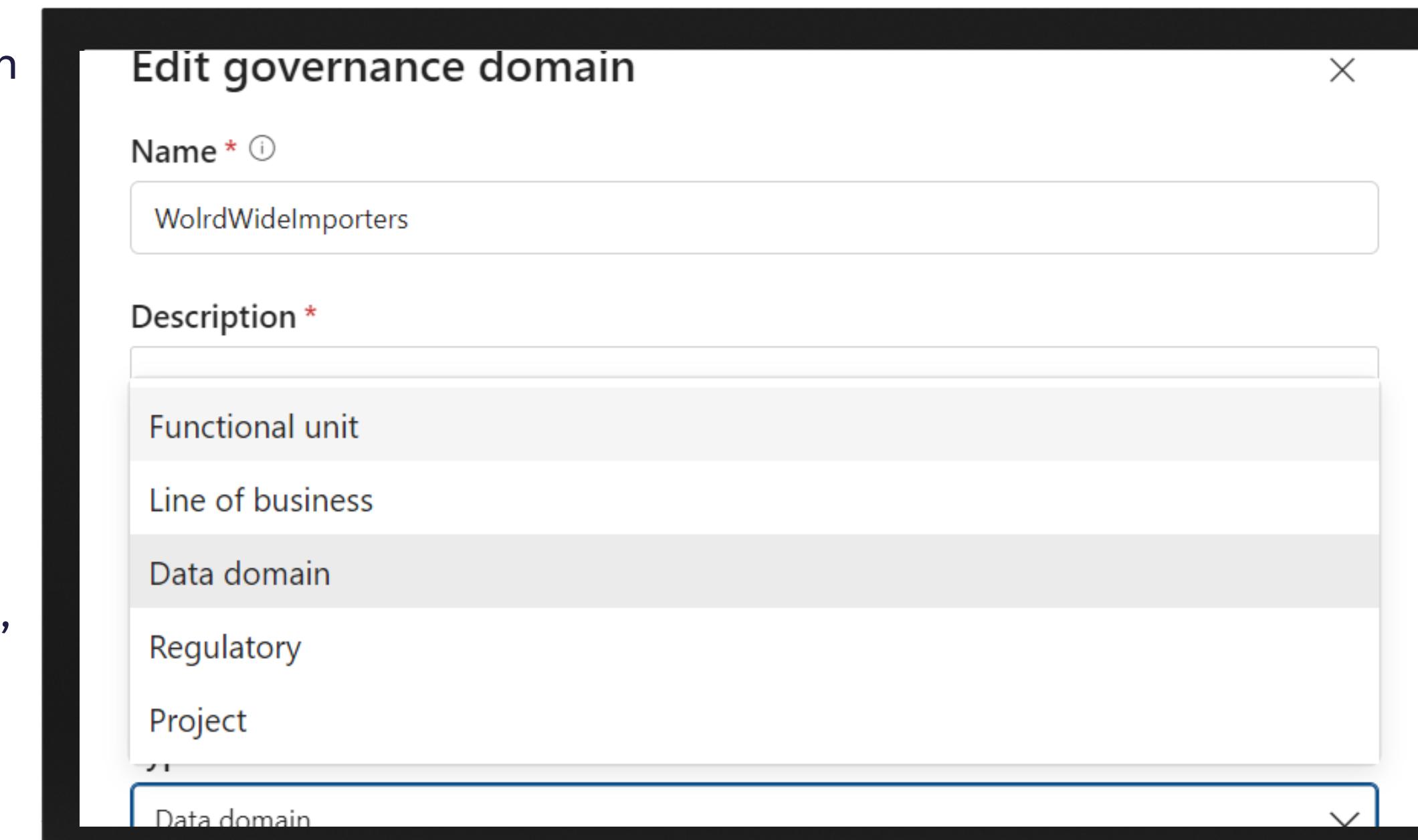
Ensuring Accountability

- ★ Governance domains ensure accountability in data management, creating clear roles and responsibilities across teams.



Types of governance domains

- **Functional unit** – organizations or business units such as Sales, Marketing, or Finance
- **Line of business** – products or services being sold such as Xbox, Office, or Azure and different markets or subsidiaries
- **Data domain** – key organization-wide entities such as customers or employees
- **Regulatory** – compliance related such as GDPR, SOX, or HIPPA
- **Project** – collaborative programs across the organization



Parts of a governance domain

- ❖ Data products
- ❖ Glossary terms
- ❖ Objectives and key results(OKRs)
- ❖ Critical data



Data Products

Business
Data

Microsoft Purview

Search

New Microsoft Purview portal

Notifications (1)

Help

Home

Solutions

Learn

Settings

Data Catalog

Overview

Discovery

Catalog management

Governance domains

Data products

Data assets

Requests

Classic types

Health management

Related solutions

Data Map

Data policy

Data products

All governance domains

Manage groups of data assets packaged together for specific use cases. [Learn more about data products](#)

+ New data product Refresh

Status: All Owner: All Type: All

Showing 1-42 of 42 items Filter by name Sort

Data product name ↑	Type	Governance domain	Data assets	Status	Data quality score
Building Locations Reference	Master data and reference data	C Corporate Functions	1	Published	--
Canada Sales Revenue Insights...	Dashboards/Reports	S Sales	7	Published	Fair 73.6
Candidate Details for CY2024	Dataset	H Human Resources	1	Published	Healthy 85.8
Candidate Details for H2CY2023	Dataset	H Human Resources	4	Expired	--
Candidate Position Fit Recom...	Dataset	H Human Resources	1	Published	Fair 62.4
Claim Center	Business System/Application	C Claims	1	Published	Fair 79.7

Microsoft Purview

- ★ A data product is a group of data assets packaged together for enterprise use, providing context and use cases for data consumers.

- ★ A data product includes:
 - ★ Name
 - ★ Description,
 - ★ Owners,
 - ★ Type
 - ★ Governance Domain and Use Case
 - ★ Data Assets

Unified Catalog > Data products >

Data Product

Data product | Dataset | Contains 1 data asset

[Edit](#) [Publish](#) [Manage policies](#) [Delete](#)

Governance domain	Status
 WorldWideImporters	 Draft
<hr/>	
Data product owner	Active Subscribers
 Erwin de Kreuk(MVP)	No active subscribers
<hr/>	
Update frequency	Data quality score
--	No score available
<hr/>	
Health actions	Terms of use
No active actions	No terms of use Add
<hr/>	
Documentation	No documentation Add

Details

Description
A description of Data Product

Use cases
To reduce customer churn by predicting which customers are likely to cancel their subscriptions and proactively engaging them with targeted retention strategies.

Data assets (1)

 Product	...
 Azure SQL Table	

[+ Add data assets](#) [View all data assets](#)

Data Quality

Microsoft Purview

Search

Data Quality

Identify and fix data quality issues by governance domain and data product.

Filter by name

Governance domain ↑	Data products	Data quality
Care Providers	3	Healthy 89.6
Claims	2	Fair 73.9
Corporate Functions	4	Healthy 86.5
Finance	6	Healthy 93.2
Fraud Services	3	Healthy 100
Human Resources	5	Fair 74.3
Sales	15	Healthy 90.5

Sales
Governance domain

Manage 86 action items

Type: All Status: All Owner: All Showing 15 item(s) Filter by

Data product name ↑	Type	Status	Data assets	Quality score	Last
Canada Sales Revenue...	DashboardsOrReports	Published	7	Fair 73.6	07/2
Commercial Customer...	MasterDataAndReferen...	Published	16	Healthy 92.7	08/2
Customer Master List	MasterDataAndReferen...	Draft	12	Healthy 95.2	08/2
DE Sales Revenue In...	DashboardsOrReports	Published	1	Healthy 100	03/2
EU Customer Churn ...	ModelTypes	Published	2	--	06/2
Global Sales Revenue...	DashboardsOrReports	Published	13	Healthy 100	08/2

Related solutions

Sales

Home
Solutions
Learn
Settings
Data Catalog
Audit
Information Protection
Data Map
Data policy

Data Quality Roles

★ Data quality steward

- ★ Able to use data quality features like data quality rule management, data quality scanning, browsing data quality insights, data quality scheduling, job monitoring, configuring threshold and alerts.

★ Data quality reader

- ★ Browse all data quality insight, data quality rules definition, and data quality error files. This role can't run data quality scanning and data profiling job, and this role won't have access to data profiling column level insight as column level insight.

★ Data quality metadata reader

- ★ Browse data quality insights (except profiling results column level insight), data quality rule definition, and rule level scores. This role won't have access to error records and can't run profiling and DQ scanning job.

This are sub roles, to perform this roles the individual also needs Governance domain reader and Data Catalog reader roles.

Data Quality features

- ★ Data source connection configuration
- ★ Data profiling
- ★ Data quality rules
- ★ Data quality scanning
- ★ Data quality job monitoring
- ★ Data quality scoring
- ★ Data quality for critical data elements (CDEs)
- ★ Data quality alerts
- ★ Data quality actions
- ★ Data quality managed virtual **network**

Data source connection configuration

- ★ **Azure Data Lake Storage Gen2**
 - ◆ File Types: Delta Parquet and Parquet
- ★ **Azure SQL Database**
- ★ **Fabric data estate in OneLake including shortcut and mirroring data estate. Data Quality scanning is supported only for Lakehouse delta tables and parquet files.**
 - ◆ Mirroring data estate: Cosmos DB, Snowflake, Azure SQL
 - ◆ Shortcut data estate: AWS S3, GCS, AdlsG2
- ★ **Azure Synapse serverless and data warehouse**
- ★ **Azure Databricks Unity Catalog**
- ★ **Snowflake**
- ★ **Google Big Query (Private Preview)**

Create connection

Overview

Display name *

Show ID

Description

Source type *

Select...

 Azure Data Lake Storage Gen2
 Azure SQL Database
 Azure Blob Storage
 Fabric
 Azure Databricks
 Google BigQuery
 Snowflake
 Azure Synapse Analytics

Data Profiling

- The AI recommendation engine suggests potentially important columns to run data profiling against

Index	Column ⓘ	Detected type	Minimum	Maximum	Distribution
1	Country	abc String	13	13	
2	CityID	123 Number	1	38,186	
3	City	abc String	3	35	
4	Continent	abc String	13	13	

Columns: All Detected type: All Pinned: All Filter by keyword

[View all](#)

Profile configurations

ⓘ System derived/inferred important columns to run Data Profiling against. Please update the Important column list if additional columns need to be profiled.

Select important columns

Data type: All

 Filter by name

<input type="checkbox"/> Column name ↑	Data type
<input checked="" type="checkbox"/> City	String
<input checked="" type="checkbox"/> CityID	Number
<input checked="" type="checkbox"/> Continent	String
<input checked="" type="checkbox"/> Country	String
<input type="checkbox"/> CityKey	Number
<input type="checkbox"/> Latest_Recorded_Population	Number
<input type="checkbox"/> Location	String
<input type="checkbox"/> Region	String
<input type="checkbox"/> Sales_Territory	String
<input type="checkbox"/> State_Province	String
<input type="checkbox"/> Subregion	String
<input type="checkbox"/> Valid_From	DateTime
<input type="checkbox"/> Valid_To	DateTime

Data Quality Rules

- ★ Freshness
- ★ Unique values
- ★ String format match
- ★ Data type match
- ★ Duplicate rows
- ★ Empty/blank fields
- ★ Table lookup
- ★ Custom

New rule

Rule
Details

What type of rule do you want to create?


Freshness
Confirms that all values are up to date.


Duplicate rows
Checks rows to find repeated values across two or more columns.


Empty/blank fields
Looks for blank and empty fields in a column where there should be values.


Unique values
Confirms that values in a column are unique.


Data type match
Confirms that values in a column match data type requirements.


String format match
Confirms that text values in a column match a specific format or other requirements.

Data Quality Rules

- ★ Freshness
- ★ Unique values
- ★ String format match
- ★ Data type match
- ★ Duplicate rows
- ★ Empty/blank fields
- ★ Table lookup
- ★ Custom

New rule

Rule

Details

Freshness

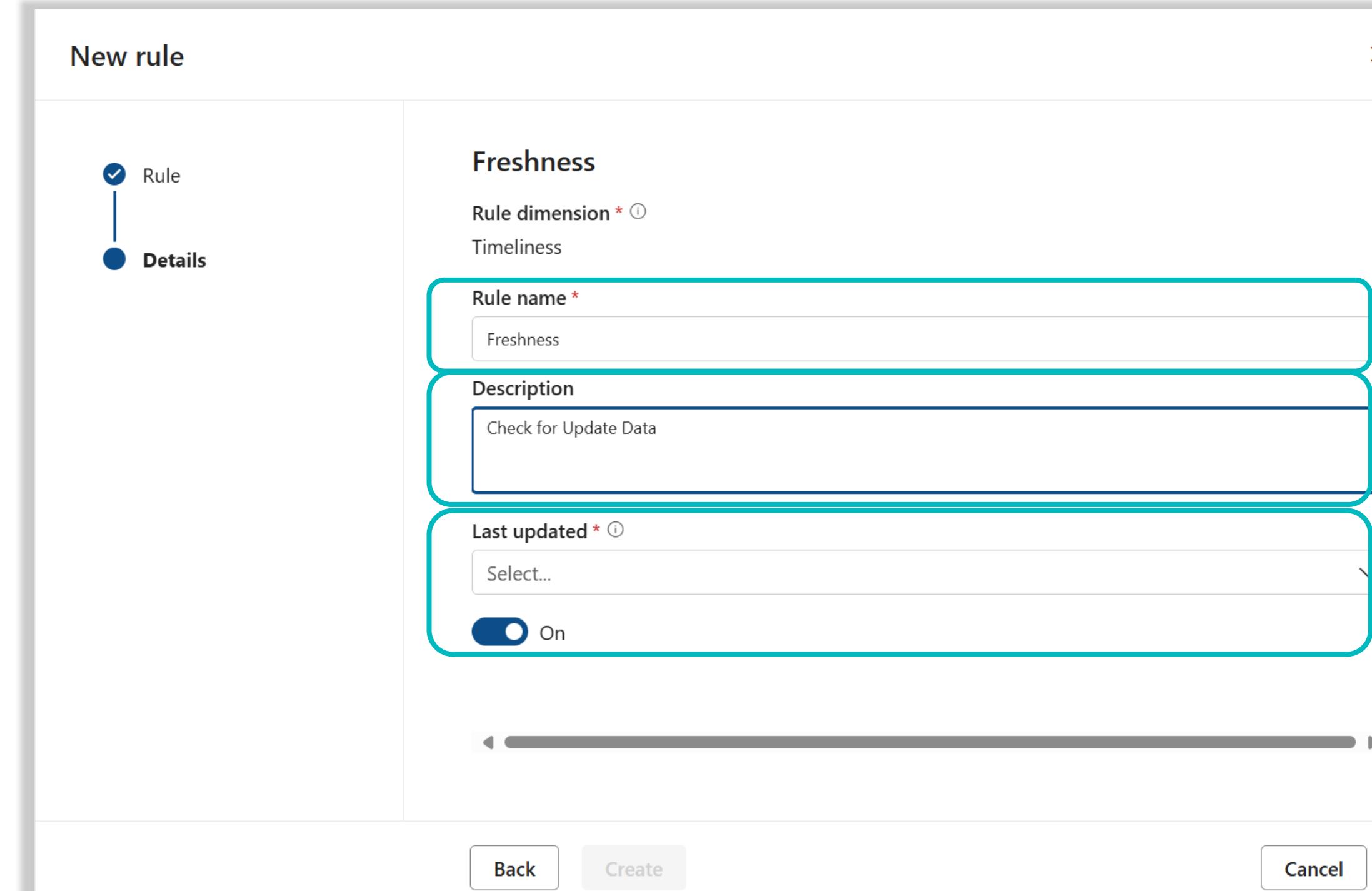
Rule dimension * ⓘ
Timeliness

Rule name *
Freshness

Description
Check for Update Data

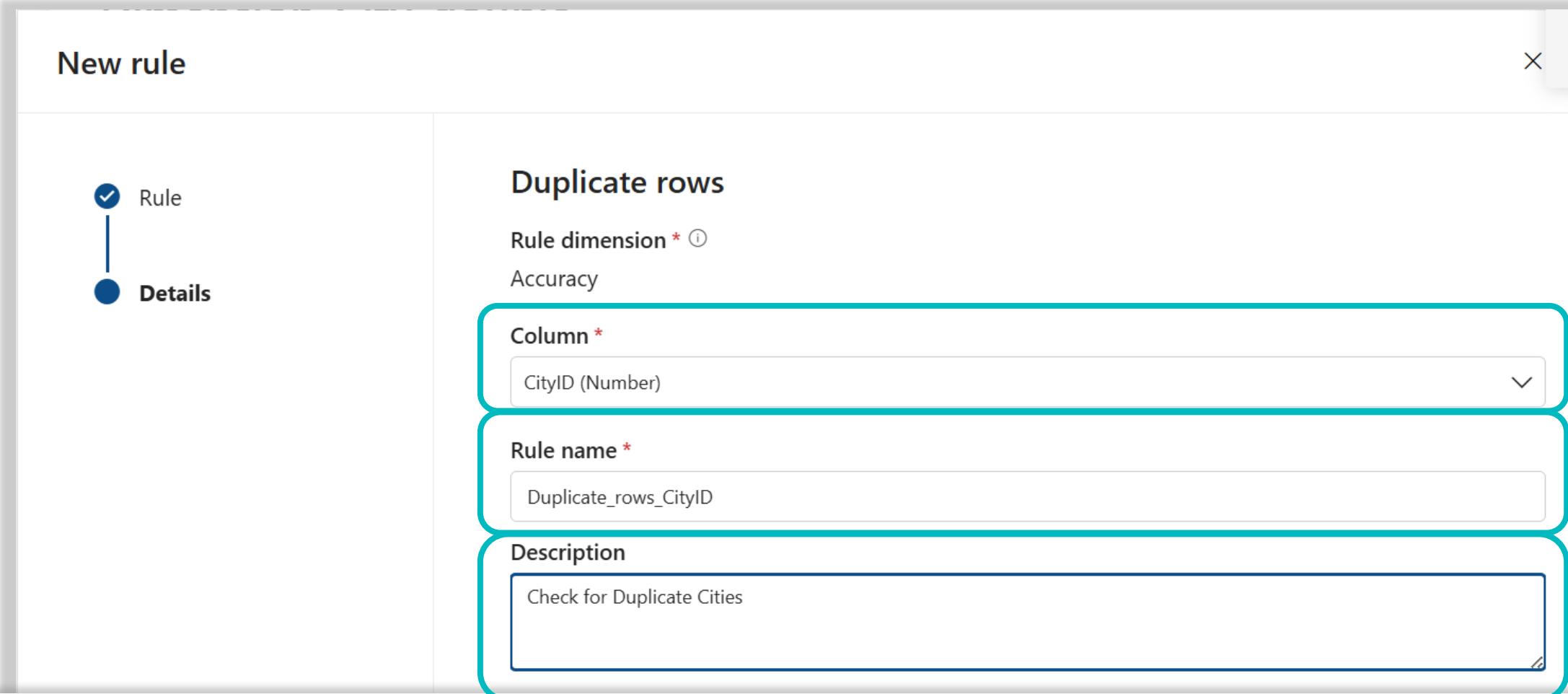
Last updated * ⓘ
Select...
 On

Back Create Cancel



Data Quality Rules

- ★ Freshness
- ★ Unique values
- ★ String format match
- ★ Data type match
- ★ [Duplicate rows](#)
- ★ Empty/blank fields
- ★ Table lookup
- ★ Custom



The screenshot shows the 'New rule' dialog for 'Duplicate rows'. It has two tabs: 'Rule' (selected) and 'Details'. Under 'Rule', the 'Rule dimension' is set to 'Accuracy'. Under 'Details', the 'Column' is 'CityID (Number)', the 'Rule name' is 'Duplicate_rows_CityID', and the 'Description' is 'Check for Duplicate Cities'.

Unified Catalog > Data Quality >

Dimension_City_double

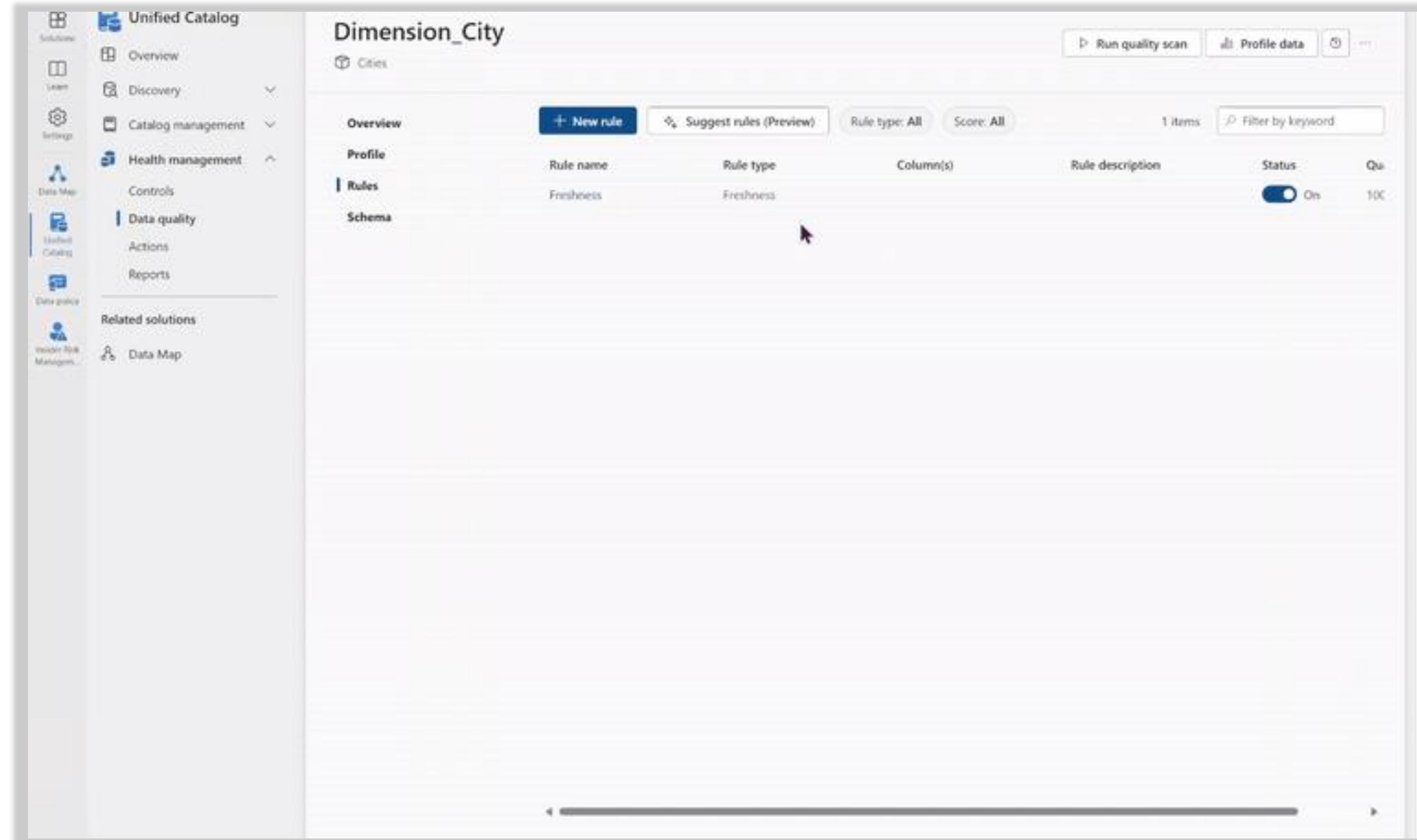
Cities

Overview + New rule Suggest rules (Preview) Rule type: All Score: All 2 items Filter by keyword

Rules	Rule name	Rule type	Column(s)	Rule description	Status	Q
	Duplicate_rows_CityID	Duplicate rows	CityID	Check for Duplicate Cities	<input checked="" type="checkbox"/> On	99%
	Freshness	Freshness			<input checked="" type="checkbox"/> On	--

Data Quality Rules

- ★ Freshness
- ★ Unique values
- ★ String format match
- ★ Data type match
- ★ Duplicate rows
- ★ Empty/blank fields
- ★ Table lookup
- ★ Custom



The screenshot shows the Microsoft Purview Data Quality Rules interface for a dimension named "Dimension_City". The left sidebar includes links for Unified Catalog, Overview, Discovery, Catalog management, Health management (selected), Controls, Data quality (selected), Actions, Reports, and Related solutions. The main area displays a table of rules for the "Dimension_City" table. The table has columns for Rule name, Rule type, Column(s), Rule description, Status, and Qu. A single rule is listed: "Freshness" (Rule type: Freshness). The interface also features a navigation bar with "Run quality scan", "Profile data", and other buttons.

Rule name	Rule type	Column(s)	Rule description	Status	Qu
Freshness	Freshness			<input checked="" type="checkbox"/> On	10%

Data Quality Rules

- ★ Freshness
- ★ Unique values
- ★ String format match
- ★ Data type match
- ★ Duplicate rows
- ★ Empty/blank fields
- ★ Table lookup
- ★ Custom

Custom

Rule dimension *

Accuracy

Data is error-free and exact. This rule checks for errors in copies due to normalization

Row expression *

Use Microsoft Purview Data Quality expression language.

```
1 Valid_From <= Valid_To
```

+ - * / || && ! ^ == === <=> != > < >= <= []

Expression elements	Expression values
All	<input type="text" value="Filter by keyword"/>
Functions	<code>1.2 varianceIf(boolean, number)</code> <code>1.2 variancePopulation(number)</code> <code>1.2 variancePopulationIf(boolean, number)</code> <code>1.2 varianceSample(number)</code> <code>1.2 varianceSampleIf(boolean, number)</code> <code>123 weekOfYear(datetime)</code> <code>121 weeks(integer)</code> <code>xor(boolean, boolean)</code> <code>123 year(datetime)</code>
Columns	

Schedule

- ★ Data quality scans review your data assets based on their applied data quality rules and produce a score.
- ★ Data stewards can use that score to assess the data health and address any issues that might be lowering the quality of your data.

Create scheduled scan

Overview Scope Schedule **Review**

Overview

Name Weekly_Cities
Description Weekly scan of the DQ Rules

Scope

Data asset	Data product
Dimension_City	Cities
Dimension_City_double	Cities

Schedule

Start At 02/19/2025, 12:45 PM (UTC)
Recurrence 8:00:00 AM (UTC), Monday, every 1 week(s)
End At -

Save

Back

Cancel

Monitoring

- ★ Data quality job monitoring enables data quality stewards to see the progress of data profiling, data quality rule generation, and data quality scanning jobs.
- ★ Both manually configured and scheduled jobs' progress can be viewed from the data quality monitoring page.

Unified Catalog > Data Quality >

Monitoring

W WorldWideImporters

Track data quality activities and scans

Activities Scans

Refresh Status: All Activity type: All Scan type: All Scan name: All Submit time: Last 24 Hours

Data asset	Data product	Status	Activity type	Scan type	Scan name
Dimension_City	Cities	✓ Completed	Assessment	Manual	--
Dimension_City_d...	Cities	✓ Completed	Assessment	Manual	--
Dimension_City_d...	Cities	✓ Completed	Profile	Manual	--
Dimension_City	Cities	✓ Completed	Rule suggestion	Manual	--

Alerts

- ★ Data quality alerts notify Microsoft Purview users about important events or unexpected behavior detected around the quality of the data.

Create alert

Overview Scope Review

Overview

Display name: Alerts on Product Cities
Description: Check if quality scores is not going under Treshold
Target: Score less than 80
Notifications: On
Turn on notifications for failed quality scans: On
Recipient: Erwin de Kreuk(MVP)

Scope

Data asset	Data product
Dimension_City	Cities
Dimension_City_double	Cities

Member	Type	X
 Erwin de Kreuk(MVP)	User	X

Actions

- ★ Rule has fallen below default thresholds
- ★ Global data quality score has fallen below threshold
- ★ Data quality scanning and/or data quality profiling job has failed
- ★ Data quality scanning and/or data quality profiling job has skipped due to no changes on data since last run
- ★ Outlier presence in profiled columns
- ★ Too many nulls in the profiled data asset

Data quality actions

Active In progress Resolved My items

Business domain: Sales X Target entity type: Data quality ... X Add

Finding name	Days active	Status
Data profile abnormally high null counts de...	18 days	Not started
Data profile abnormally high null counts de...	18 days	In progress
Data profile outlier values detected	18 days	Not started
Data profile outlier values detected	18 days	Not started
Data quality Assessment job has failed.	18 days	Not started
Data asset quality rule score has fallen belo...	18 days	Not started

Reports

- ★ The Data Quality Health Report has dependencies on Data Health Controls and Purview Metadata Self-serve analytics model. If customers do not use the Data Health Controls and do not subscribe Purview Unified Catalog metadata then the Data Quality Health Report will not be refreshed. Customers either need to use Data Health Controls or need to subscribe purview metadata for self-serve analytics to refresh Data Quality health report.



- ★ If customers don't use Data Quality feature, the Data Quality Health report will show blank

Microsoft Purview

Unified Catalog

Roles and permissions

Solution integrations (Preview)

Usage settings (Preview)

Self-serve analytics (Preview)

Enhance your data estate health with self-serve analytics and insights.

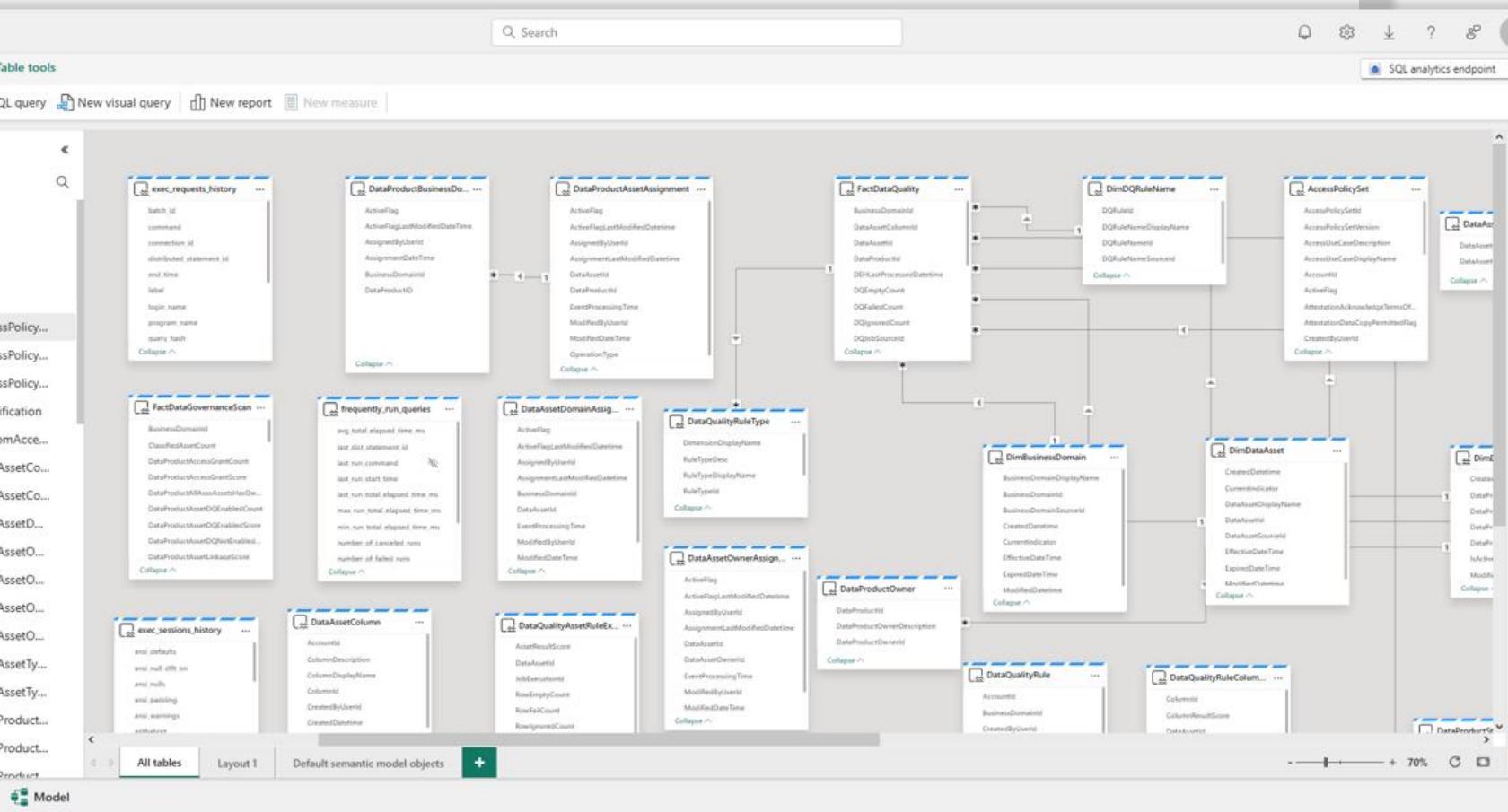
Configure storage

Status: **Disabled**

Storage type: -

Last published: -

Edit



Solution Integrations(Self-serve analytics (Preview))

Configure storage

Enabled

Storage type

Fabric

Location URL *

https://onelake.dfs.fabric.microsoft.com/WS_PURVIEW/LH_METADATA_PURVIEW.Lakehouse/Files/Purview

💡 You must grant the Microsoft Purview managed identity access to your Fabric Workspace. [Show more](#)

🔗 [Test connection](#)

✓ Connection successful.

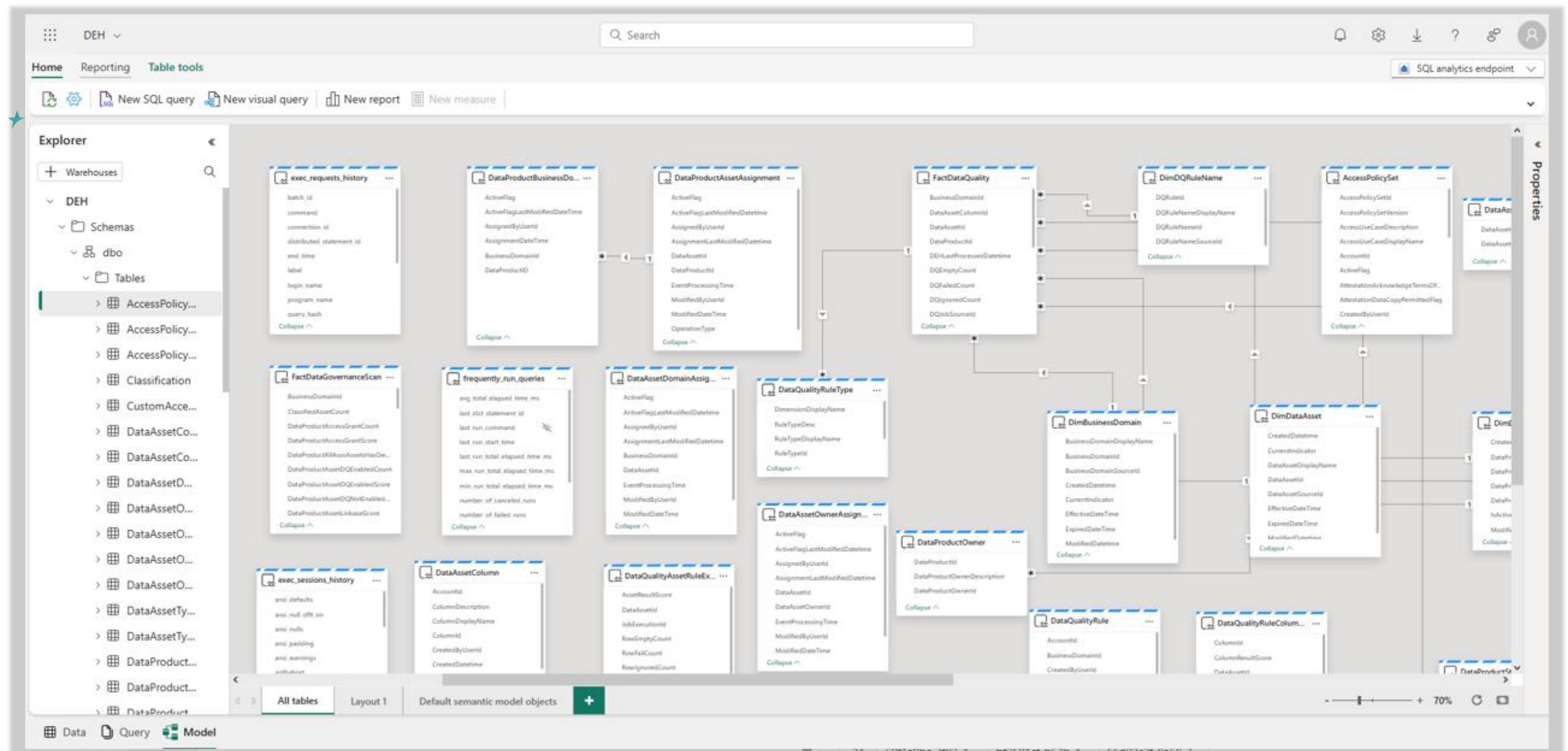
⬅ Add people WS_PURVIEW ✖

ℹ️ Admins, members, and contributors have edit and view access. Viewers only have view access. [Learn more](#) ↗

e ededeunpview01 ✖

edit Contributor Add

Microsoft Purview



Demo



Roadmap Data Quality

- ★ **Expanded Connectors Q3 2025**
 - ★ On-premises support for Oracle
 - ★ Fabric PBI and Warehouse workloads
 - ★ **Incremental Data Quality Q2 2025**
 - ★ **API to expose DQ scores and DQ actions and DQ rules. Q3 2025**
 - ★ **Scheduling, alerting and issue remediation Q2 2025**
 - ★ **Private endpoints and vNet availability for Data Quality Q2 2025**
 - ★ Data stored in Vnet storage Snowflake and Databricks
 - ★ Data stored behind the private endpoint (AzSQL, AdlsG2, Synapse dedicated and serverless, Databricks UC, Snowflake, and Fabric)
-
- ★ <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/MS-Purview-Data-Governance-Roadmap-detailed.pdf>

Conclusion

★ Importance of Data Quality Rules

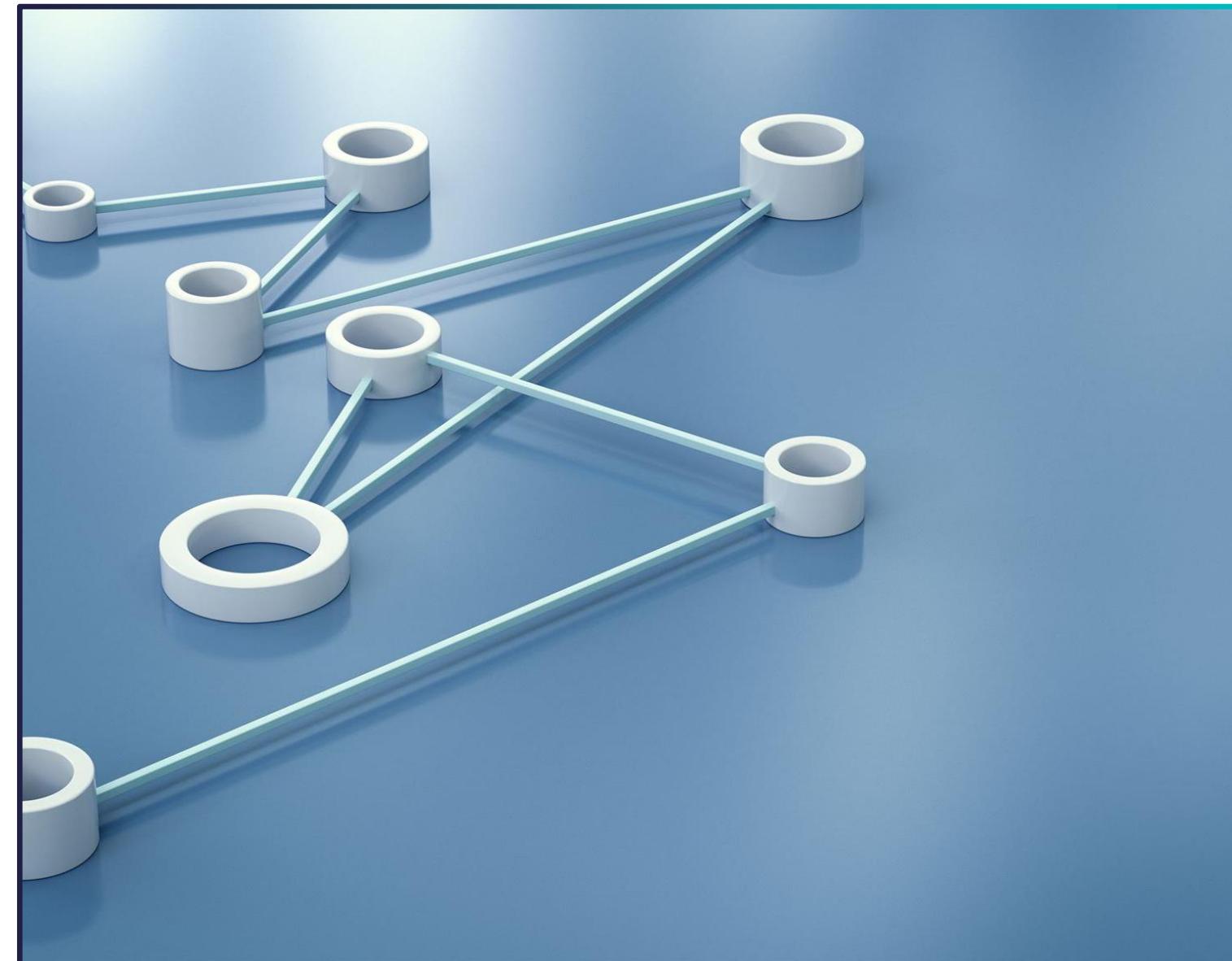
- ★ Defining data quality rules is essential for ensuring high standards and accuracy in data management across organizations.

★ Role of Microsoft Purview

- ★ Microsoft Purview simplifies the process of applying data quality rules, enabling organizations to manage their data effectively.

★ Setting Clear Expectations

- ★ Establishing clear expectations for data quality helps organizations maintain consistency and reliability in their data practices.



Maintaining High Data Quality

★ Data Quality Monitoring

- ★ Implementing tools to continuously monitor data quality helps in identifying potential issues before they escalate.

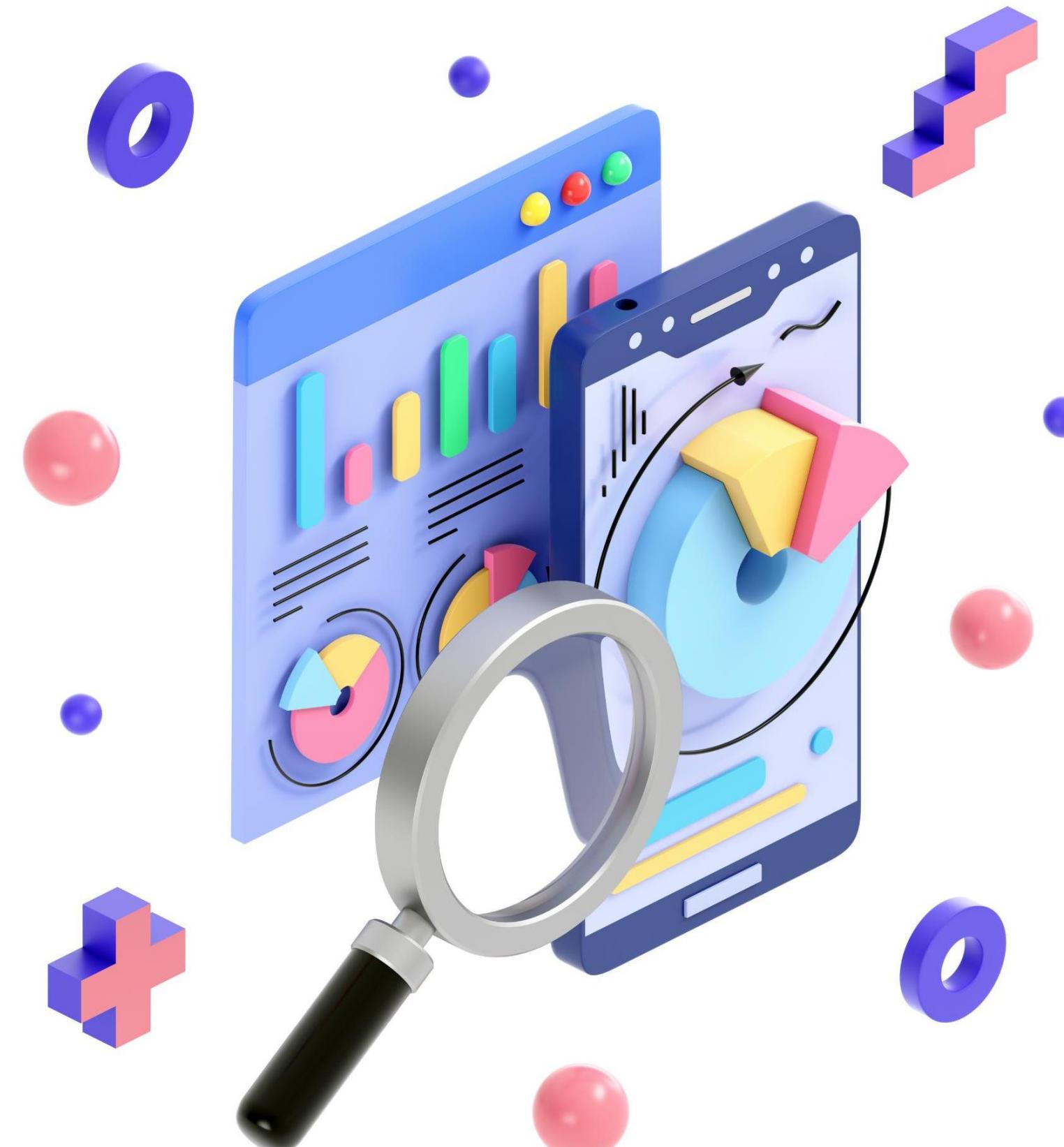
★ Preventive Measures

- ★ Establishing processes that prevent data quality issues ensures the reliability of insights derived from the data.

★ Accurate Insights

- ★ Maintaining high data quality leads to accurate insights, driving informed decision-making across the organization.

Maintaining High Data Quality is a continuous process and not a one stop. It must be embedded in the organization



Feedback →



Erwin de Kreuk

★ Thank You



@erwindekreuk



linkedin.com/in/erwindekreuk



erwindekreuk.com



github.com/edkreuk



<https://sessionize.com/erwin-de-kreuk/>

Feedback 



Let's connect



Microsoft®
Most Valuable
Professional





INSPARK