

Lunch: 12:30 – 13:30

Snacks: 15:00

End + Drinks: 17:00

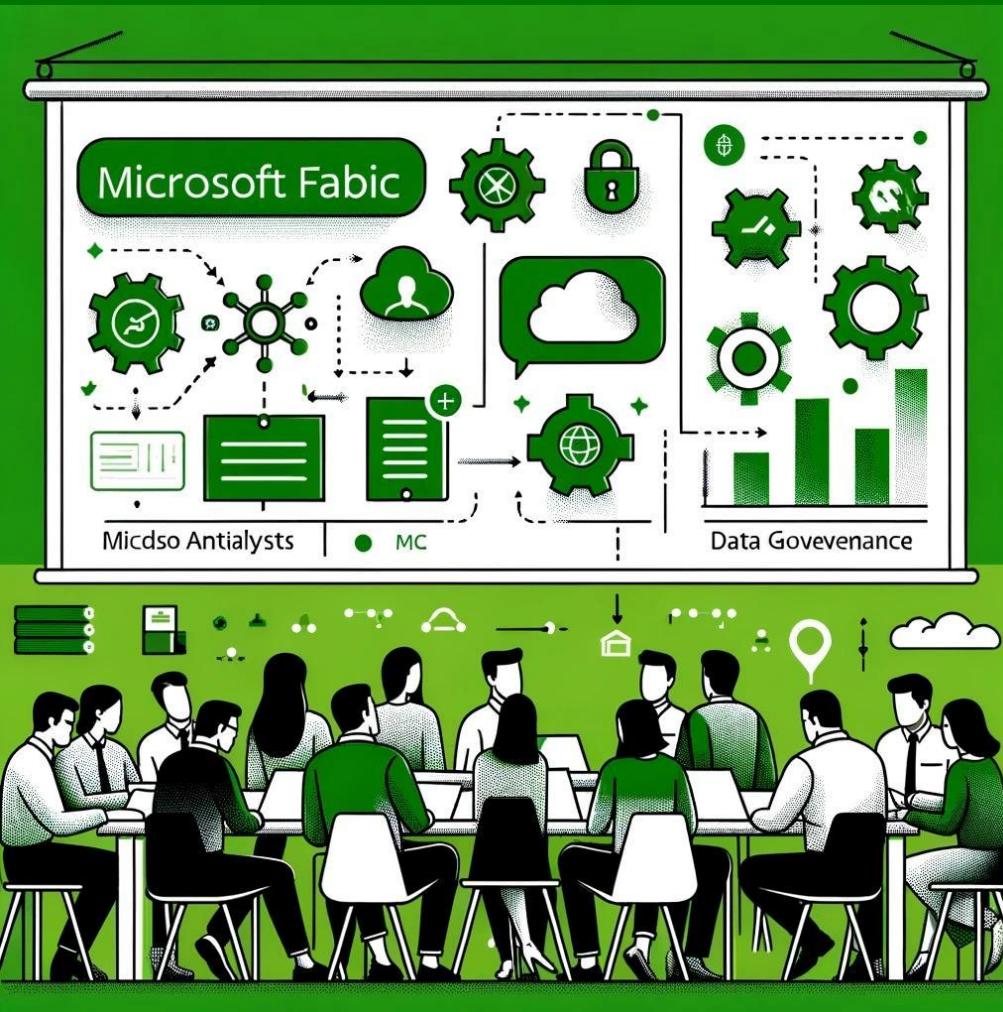
AMA: 17:30

Data Governance with Microsoft Fabric and Microsoft Purview

Erwin de Kreuk
Wolfgang Strasser



Data Platform Next Step
June 2024





Servus! Hi! Welcome!



Wolfgang Strasser

Data Consultant @  **CUBIDO**

Digital Solutions

workingondata.wordpress.com

<https://www.youtube.com/@WolfgangStrasser>



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/>



Objectives for Today

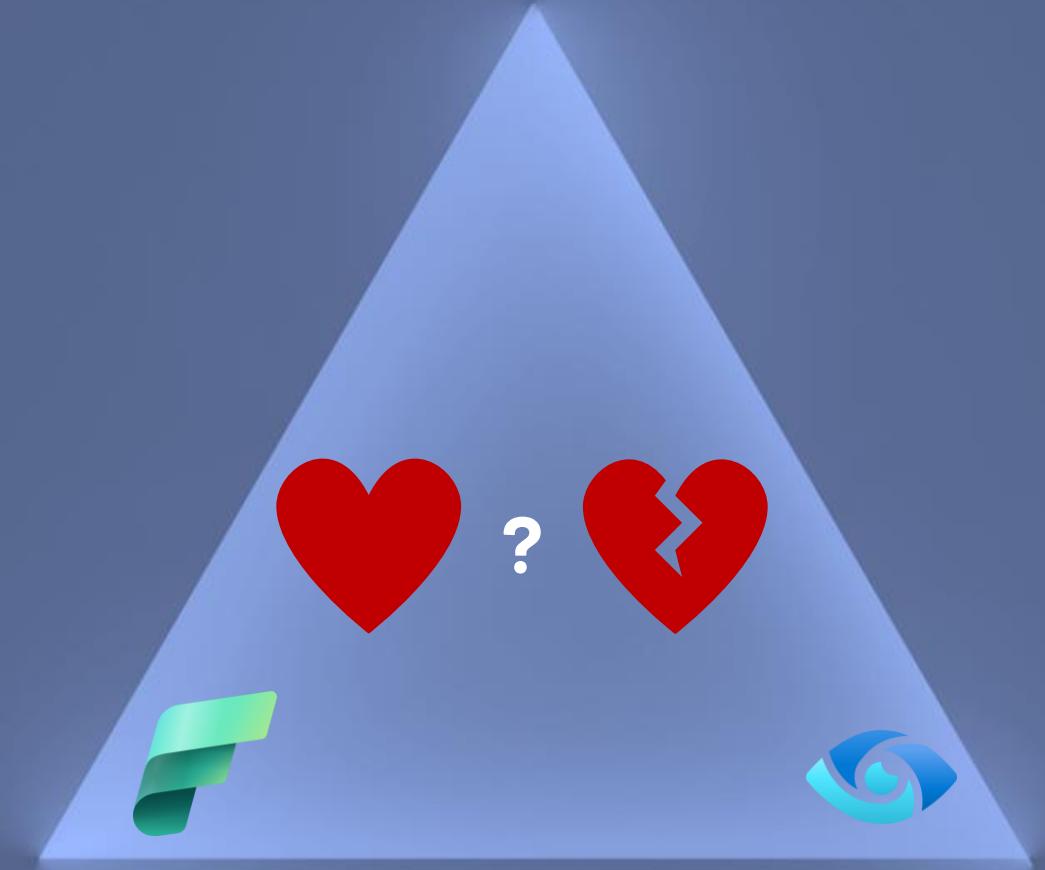
- Data Governance
- Microsoft Fabric
- Microsoft Fabric Governance
- Microsoft Purview
- Microsoft Purview
 - Demos / Hands-On / Click-Through

Feedback





Data Governance



**Microsoft
Fabric**

**Microsoft
Purview**

The book of Data Governance, Microsoft Purview & Microsoft Fabric

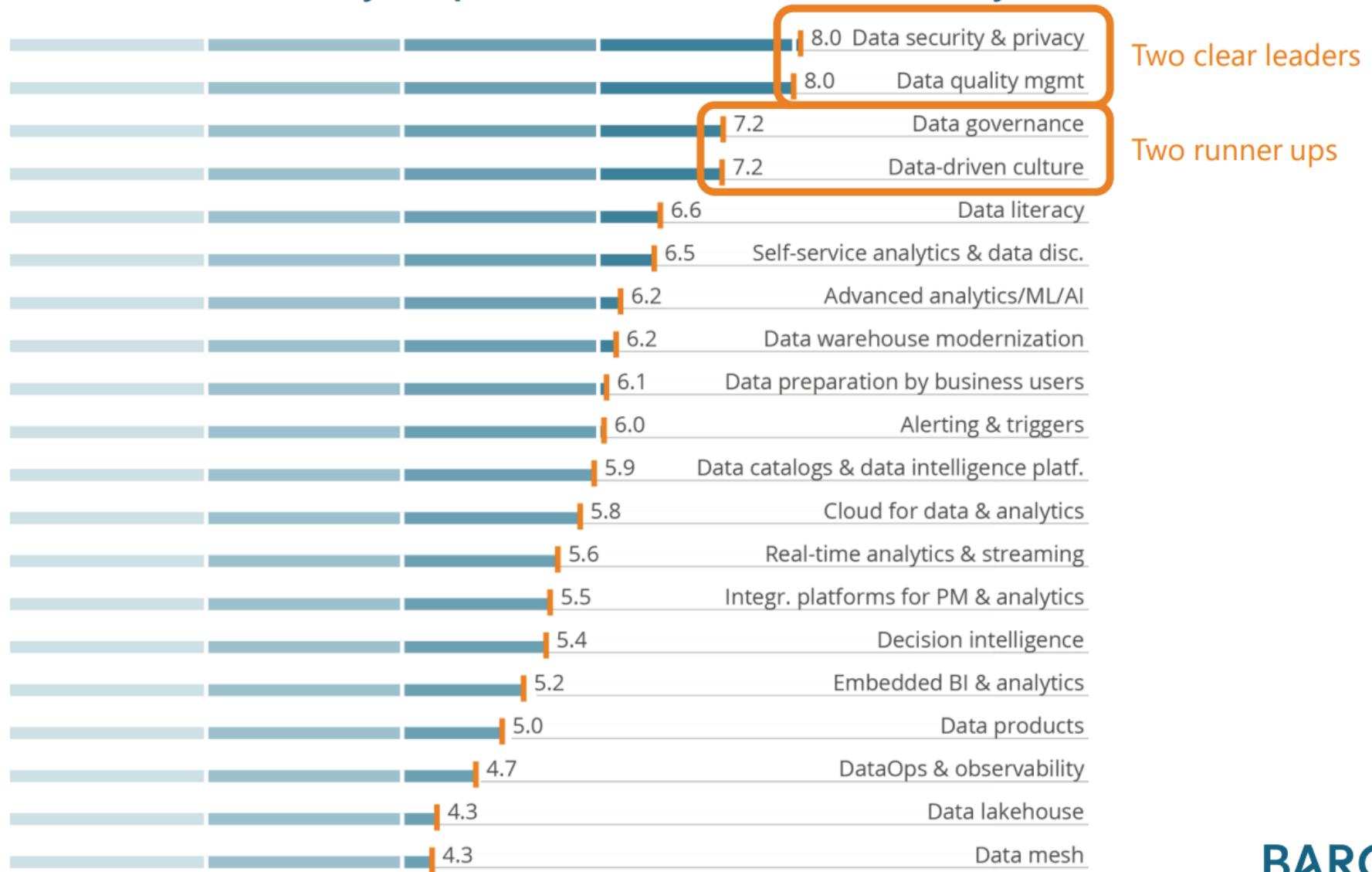




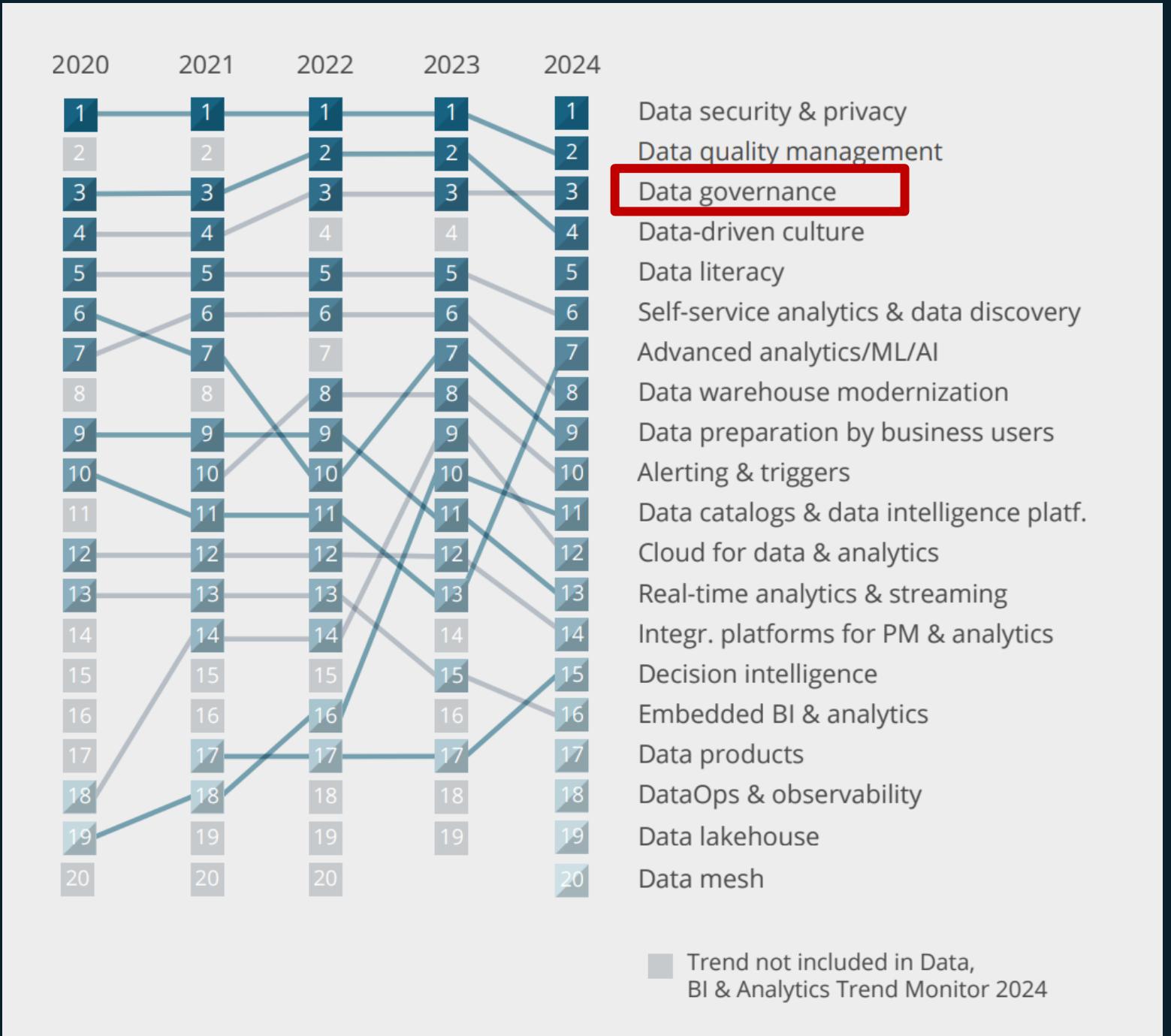
#DataGovernance

Why Data Governance?

2,398 BI practitioners have their say: Importance of Data, BI and Analytics Trends



Source: BARC Data, BI & Analytics
Trend Monitor 2024, n=2,398



Ask yourself (or in the name of your org)...

... what you cannot do?

- *We cannot compare costs across regions..*
- *We cannot track students' progress..*
- *We cannot match the records for an individual across our touch points to our organisation...*
- *We cannot maximize our decision-making capabilities based on the data we have.*

⇒ All of the *cannot* responses hinder how an organization grows...

Why is Data Governance so important?

It brings meaning to an organization's data

Increase your revenue (by improving data quality)

Reduce costs (less work to do with your data)

Lower risks (because you know your data)

Stronger regulatory compliance (data catalog)

It just makes sense..

What is Data Governance?

what is data governance

All

Images

Videos

News

Books

More

About 273.000.000 results (0,66 seconds)

what is data governance

ALL

WORK

IMAGES

VIDEOS

MAPS

NEWS

15 100 000 Results

Date ▾

Open links in new tab



Governance / To Govern

 **govern**

/'gʌv(ə)n/

See definitions in:

All Law Grammar

verb

verb: **govern**; 3rd person present: **governs**; past tense: **governed**; past participle: **governed**; gerund or present participle: **governing**

1. conduct the policy, actions, and affairs of (a state, organization, or people) with authority.
"he was incapable of governing the country"
Similar: rule preside over be in power over reign over control
- control, influence, or regulate (a person, action, or course of events).
"the future of Jamaica will be governed by geography not history"
Similar: determine decide control regulate direct rule dictate
- conduct oneself, especially with regard to controlling one's emotions.
"he does not have the ability to govern himself or others successfully"
Similar: control restrain keep in check check curb hold back
- serve to decide (a legal case).

2. **GRAMMAR**
(of a word) require that (another word or group of words) be in a particular case.
"the Latin preposition 'cum' governs nouns in the ablative"

guide, steer, manage



Definition Attempt #1

Data governance (DG) is the **process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies** that also control data usage.

Effective data governance ensures that **data is consistent and trustworthy** and **doesn't get misused**. It's increasingly critical as organizations face new data privacy regulations and rely more and more on data analytics to help optimize operations and drive business decision-making.

Definition Attempt #2

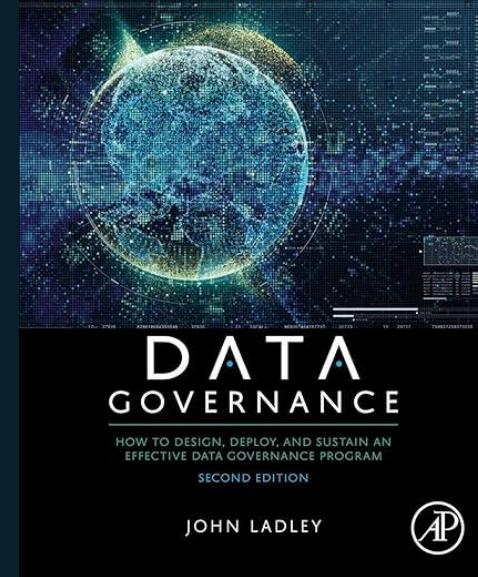
Data governance is a **collection of processes, roles, policies, standards, and metrics** that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

It establishes the processes and responsibilities that ensure the quality and security of the data used across a business or organization. Data governance defines who can take what action, upon what data, in what situations, using what methods.

Definition (Ladley 2019)

Data governance is a required business capability if you

want to get value from your data



Definition Attempt #xx

#TreatYourDataBetter

Wolfgang Strasser



How can I help you today?

Data Governance

Vision and Strategy



Roles



Processes



Policies



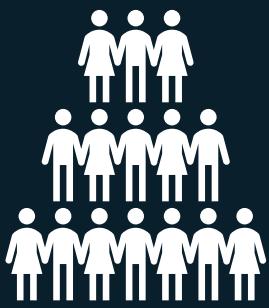
Tools



Data Lifecycle

People & Roles

DG Roles & Involved People



Data Owners

System Owners

Subject Matter Experts

Data Stewards

Data Custodians

Data Governance Council

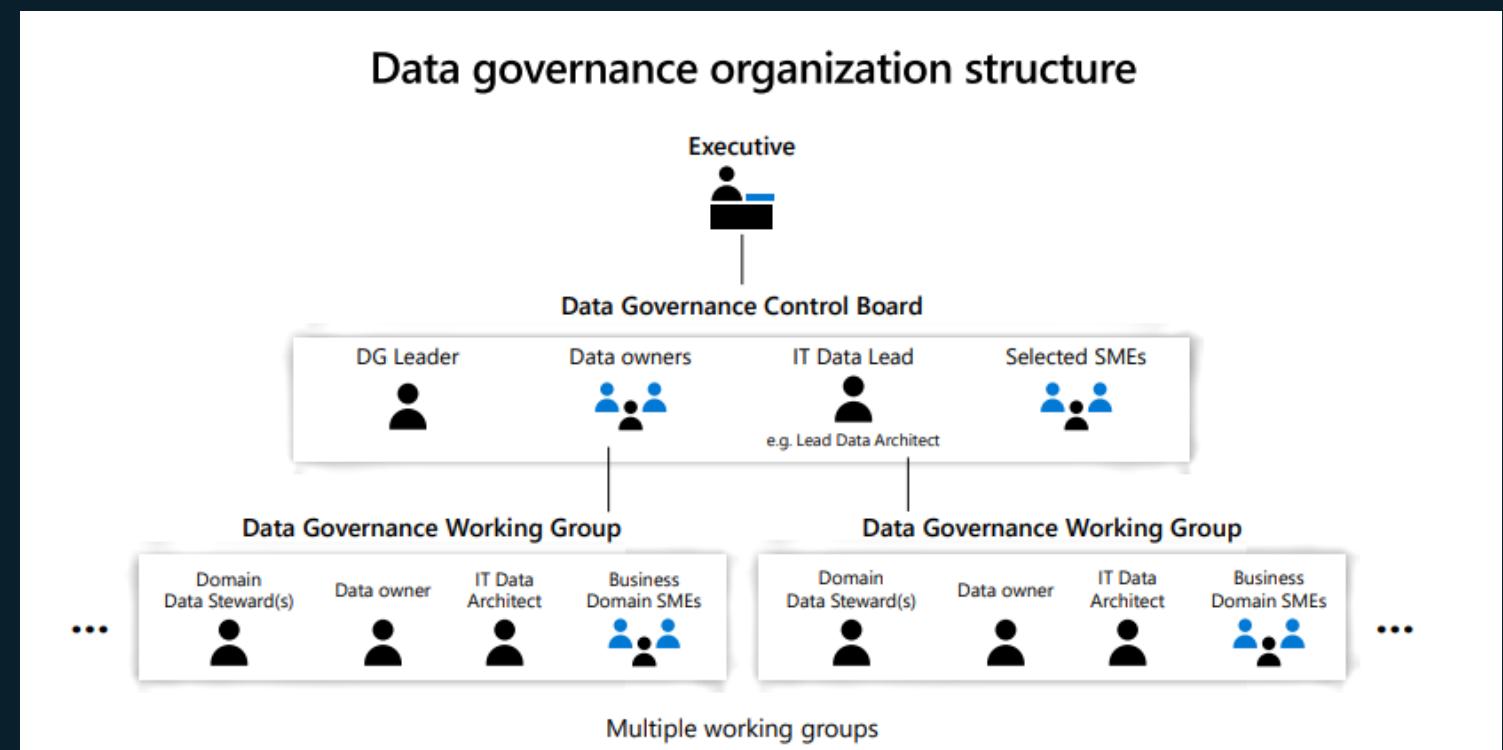
Data Consumers

Data Analysts

Auditors

Data Security Team

....



Data Responsibility / Accountability

- Who sets success metrics and monitors how well DQ is working?
- Who are the data owners?
- Who defines the business glossary?
- Who is protecting PII data privacy?
- Who is looking after data quality?
- Who ensures customer / product data is consistent across all systems?

Is it C-level? Is it a department head?

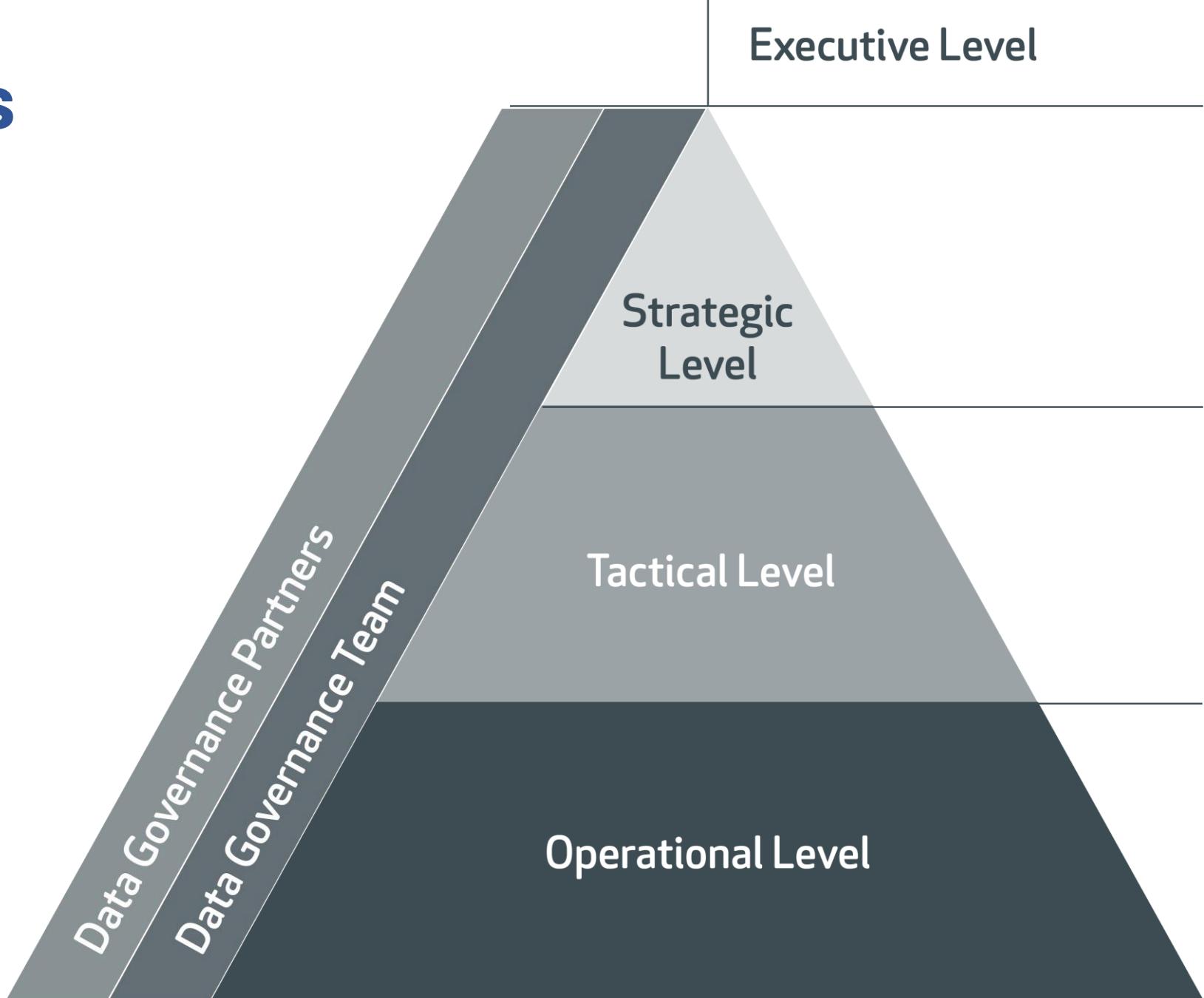
Is it the head of governance and risk?

Is it the legal department?

Is it IT?

=> Definition is required!

DG Roles & Responsibilities





#MicrosoftFabric



Microsoft Fabric

Data analytics for the era of AI

From

Isolated component

Single database

Gen AI Bolted-in

To

Unified stack

All the data

Gen AI built in



Microsoft Fabric

Data analytics for the era of AI

**Complete
Analytics
Platform**

**Lake Centric
and Open**

**Empower
Every
Business User**

**Secure and
governed**



Microsoft Fabric

The unified data platform for the era of AI



Data
Factory



Synapse Data
Engineering



Synapse Data
Science



Synapse Data
Warehousing



Synapse Real
Time Analytics



Power BI



Data
Activator



AI



OneLake



Purview

Unified
architecture

Unified
experience

Unified
governance

Unified
business model



Microsoft Fabric

The data platform for the era of AI

Complete Analytics Platform

Everything, unified

SaaS-ified

Secured and governed

Lake centric and open

OneLake

One Copy

Open at every tier

Empower Every Business User

Familiar and intuitive

Built into Microsoft 365

Insight to action

Secure and governed

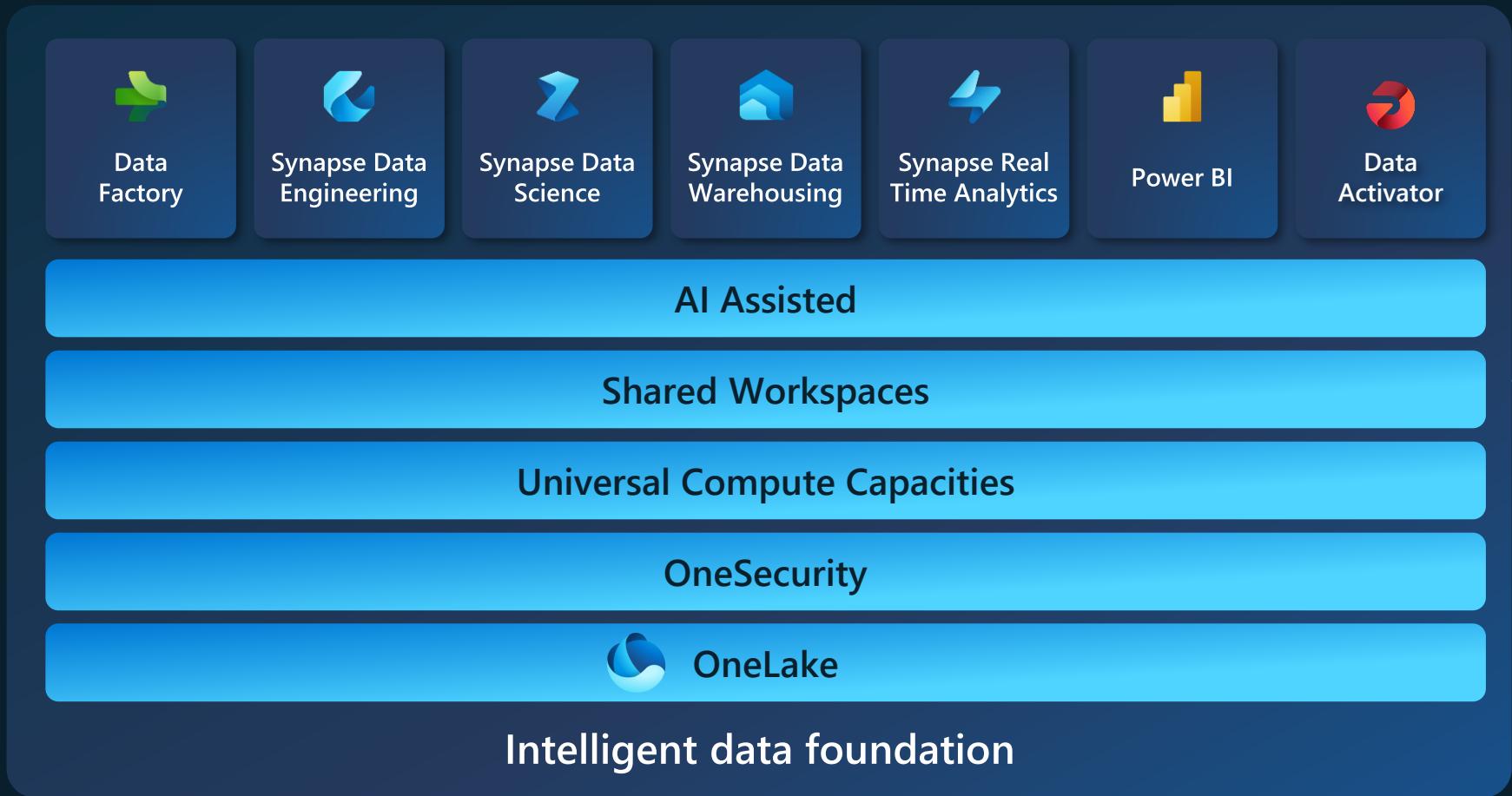
Manage your data estate

Secure, Protect & comply

Encourage data discovery, trust and use



Microsoft Fabric



Single...

- Onboarding and trials
- Sign-on
- Navigation model
- UX model
- Workspace organization
- Collaboration experience
- Data Lake
- Storage format
- Data copy for all engines
- Security model
- CI/CD
- Monitoring hub
- Data Hub
- Governance & compliance

SaaS

"It just works"

5x5

Frictionless onboarding

Instant Provisioning

Quick results w/ Intuitive UX

**Success
by default**

Minimal knobs

Auto optimized

Auto Integrated

**Centralized
administration**

Tenant-wide governance

Centralized
security management

Compliance built-in



Microsoft Fabric

The data platform for the era of AI

Complete Analytics Platform

Everything, unified

SaaS-ified

Secured and governed

Lake centric
and open

OneLake

One Copy

Open at every tier

Empower Every Business User

Familiar and intuitive

Built into Microsoft 365

Insight to action

Secure and governed

Manage your data estate

Secure, Protect & comply

Encourage data discovery,
trust and use

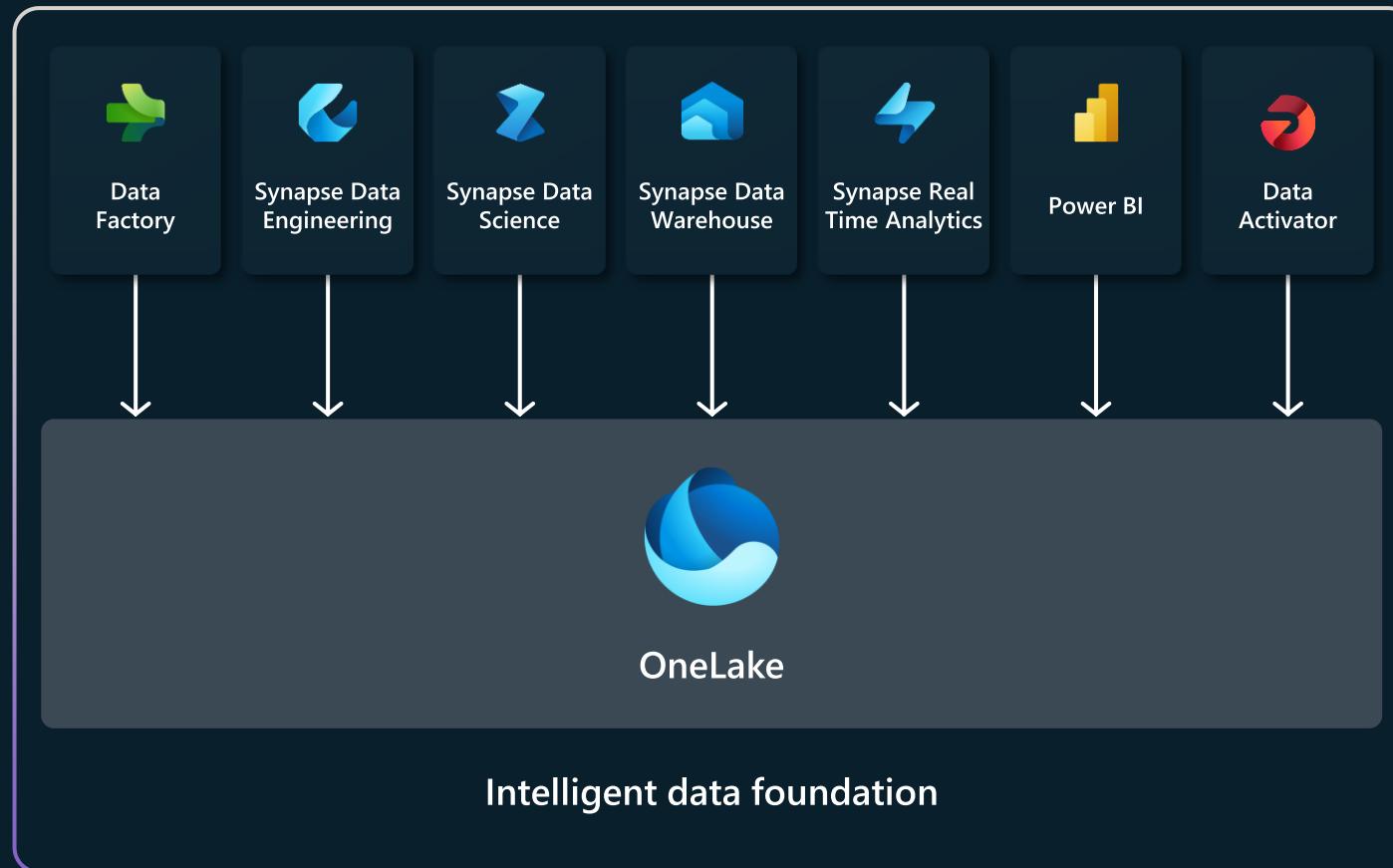
OneLake which logically spans the world

"The OneDrive for Data"



OneLake for all Data

“The OneDrive for Data”



A single SaaS lake for the whole organization

Provisioned automatically with the tenant

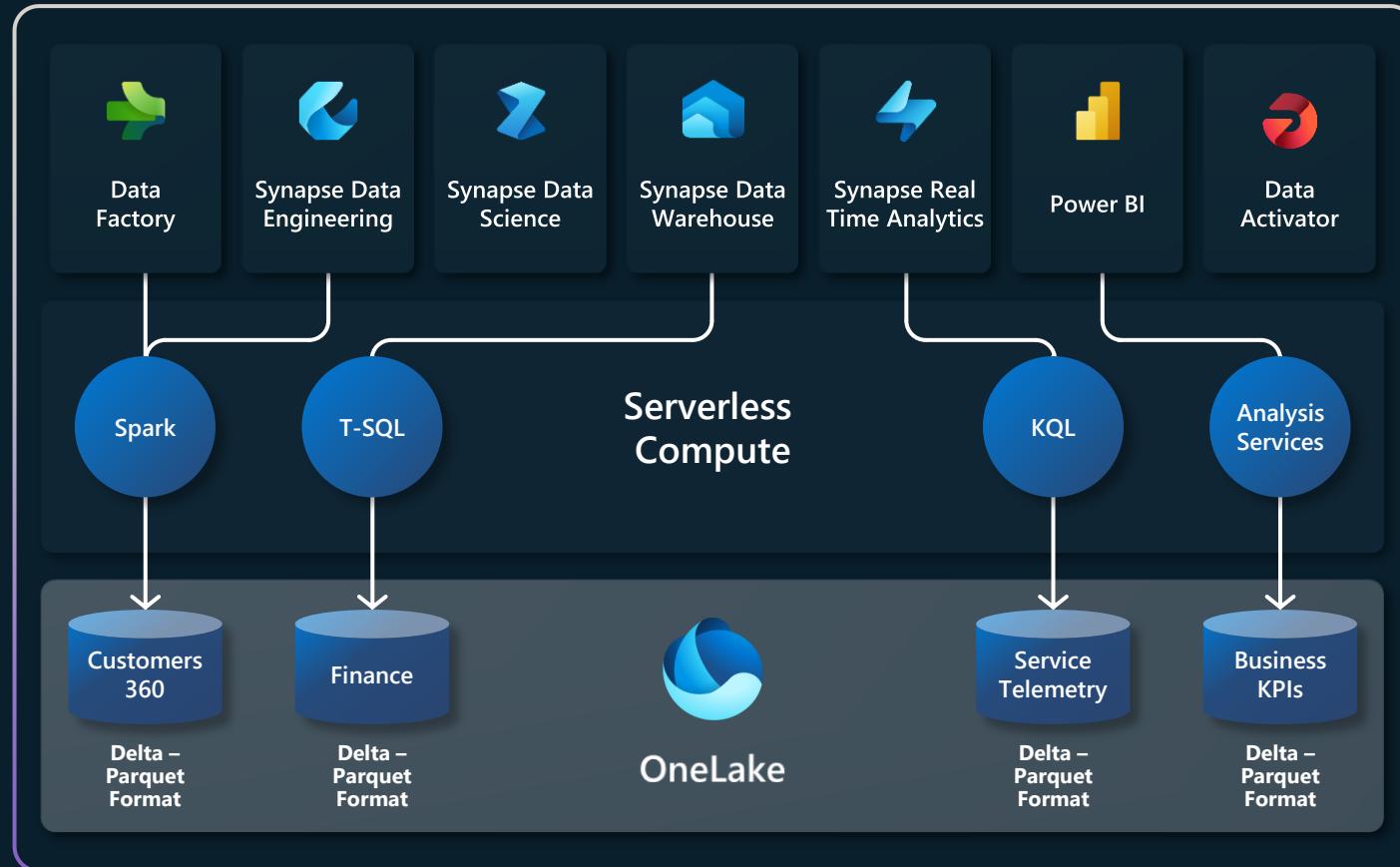
All workloads automatically store their data in the OneLake workspace folders

All the data is organized in an intuitive hierarchical namespace

The data in OneLake is automatically indexed for discovery, MIP labels, lineage, PII scans, sharing, governance and compliance

One Copy for all computers

Real separation of compute and storage



All the compute engines store their data automatically in OneLake

The data is stored in a single common format

Delta – Parquet, an open standards format, is the storage format for all tabular data in Analytics vNext

Once data is stored in the lake, it is directly accessible by all the engines without needing any import/export

All the compute engines have been fully optimized to work with Delta Parquet as their native format

Unifying data in OneLake

Shortcuts



Sharing data in OneLake is as easy as sharing files in OneDrive, removing the needs for data duplication

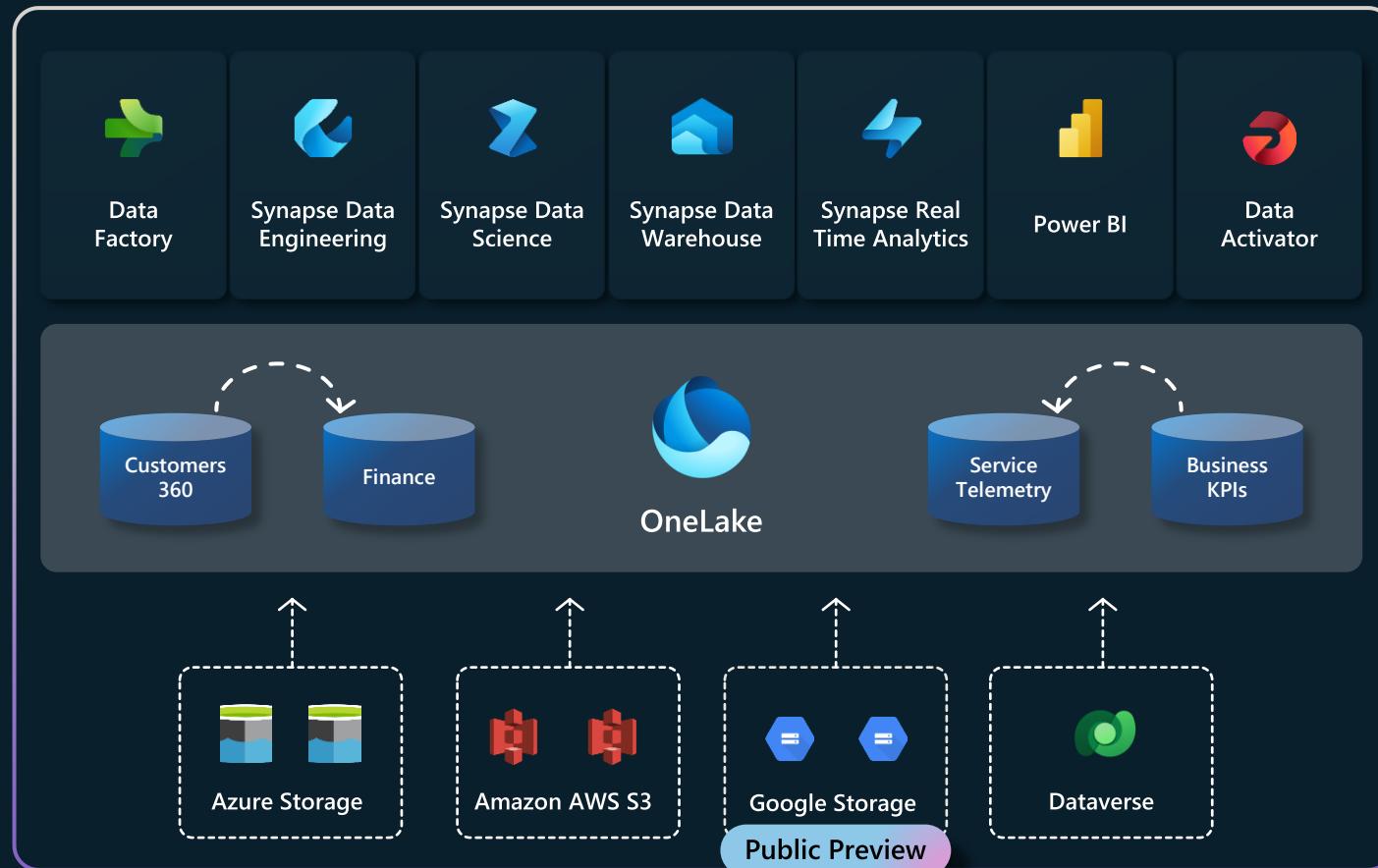
With shortcuts, data throughout OneLake can be composed together without any data movement

Shortcuts also allow instant linking of data already existing in Azure and in other clouds, without any data duplication and movement, making OneLake the first multi-cloud data lake

With support for industry standard APIs, OneLake data can be directly accessed by any application or service

Unifying data in OneLake

Shortcuts



Sharing data in OneLake is as easy as sharing files in OneDrive, removing the needs for data duplication

With shortcuts, data throughout OneLake can be composed together without any data movement

Shortcuts also allow instant linking of data already existing in Azure and in other clouds, without any data duplication and movement, making OneLake the first multi-cloud data lake

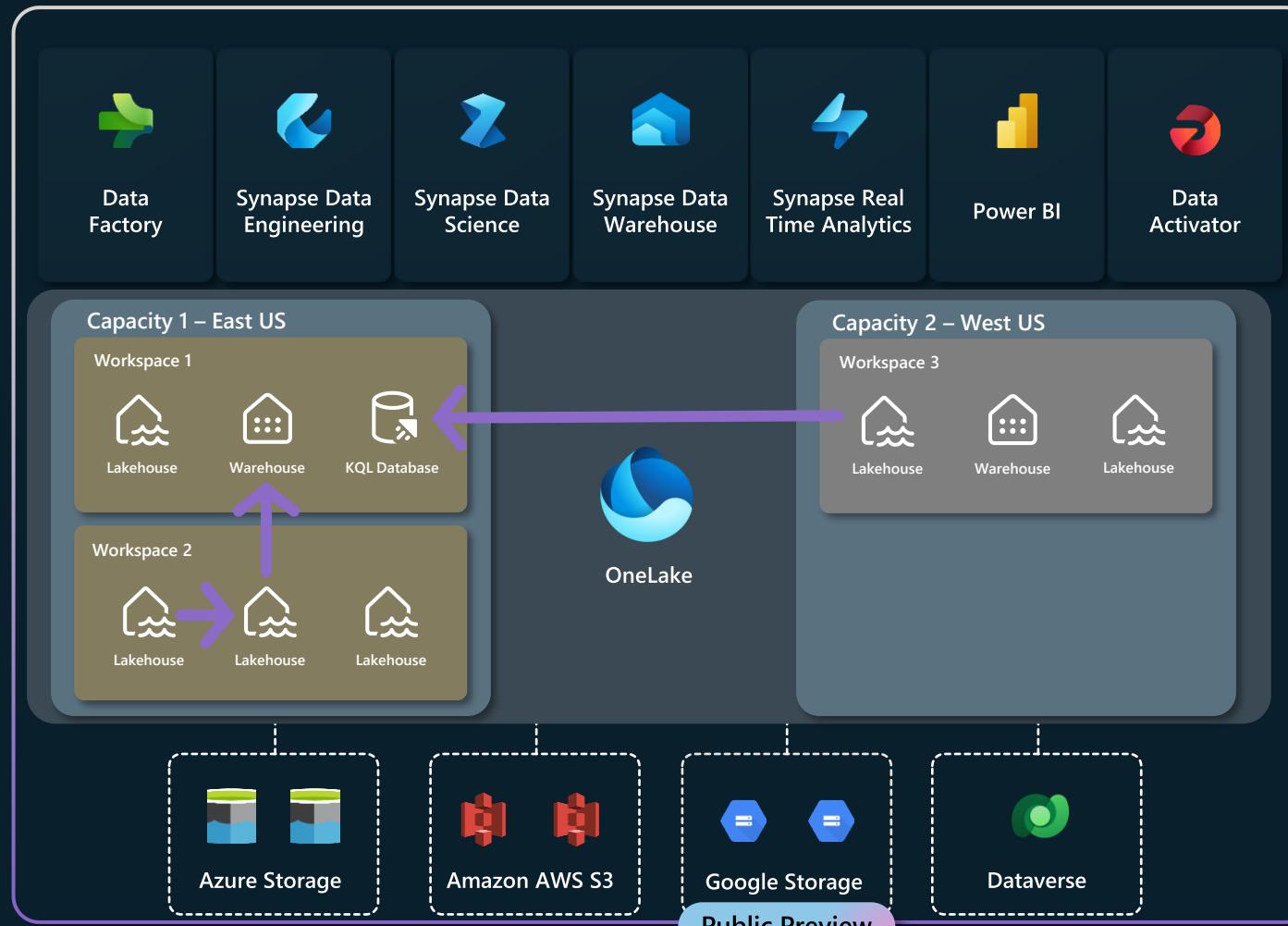
With support for industry standard APIs, OneLake data can be directly accessed by any application or service



Any questions so far?

Unifying data in OneLake

Shortcuts



Sharing data in OneLake is as easy as sharing files in OneDrive, removing the needs for data duplication

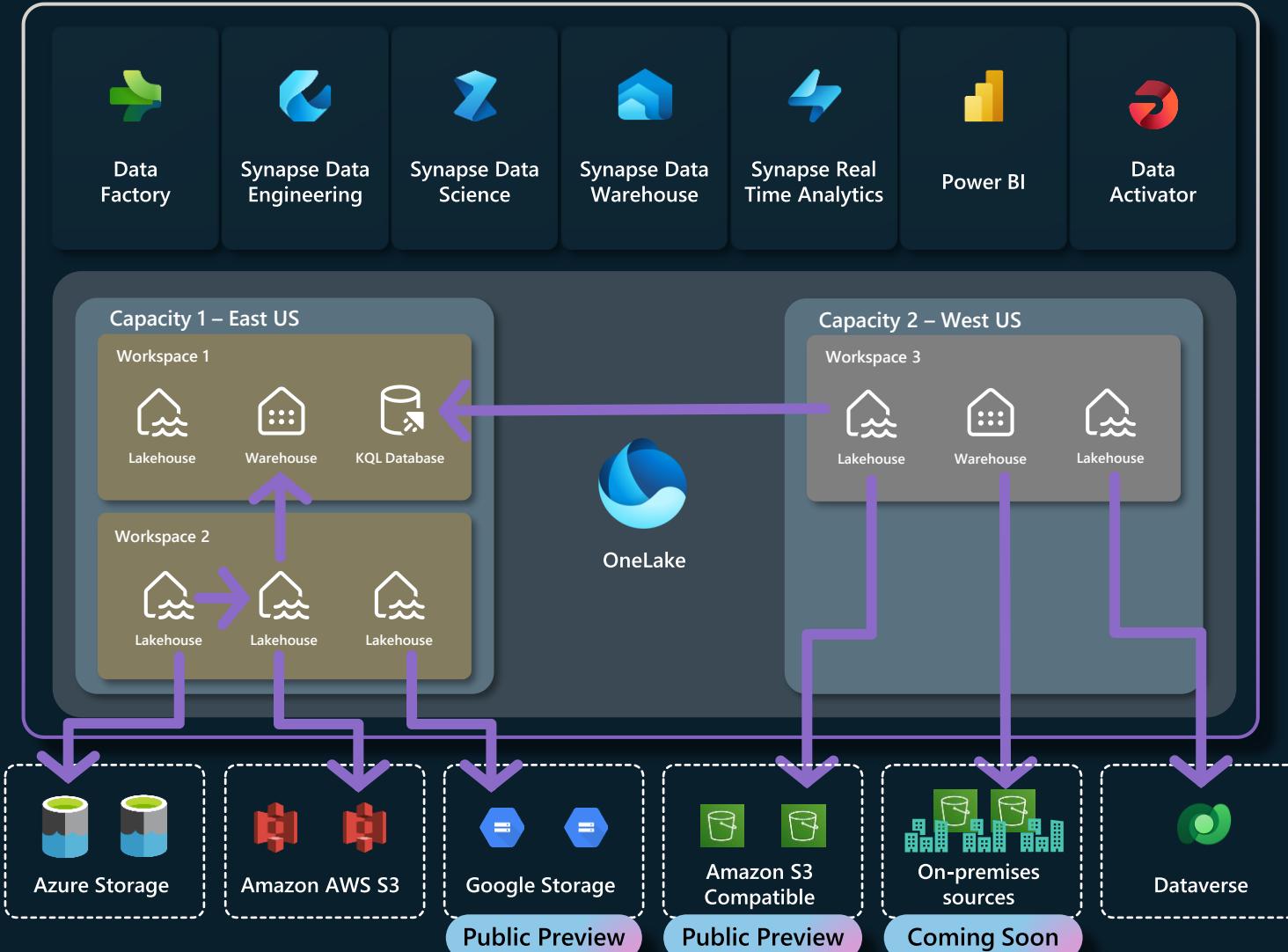
With shortcuts, data throughout OneLake can be composed together without any data movement

Shortcuts also allow instant linking of data already existing in Azure and in other clouds, without any data duplication and movement, making OneLake the first multi-cloud data lake

With support for industry standard APIs, OneLake data can be directly accessed by any application or service

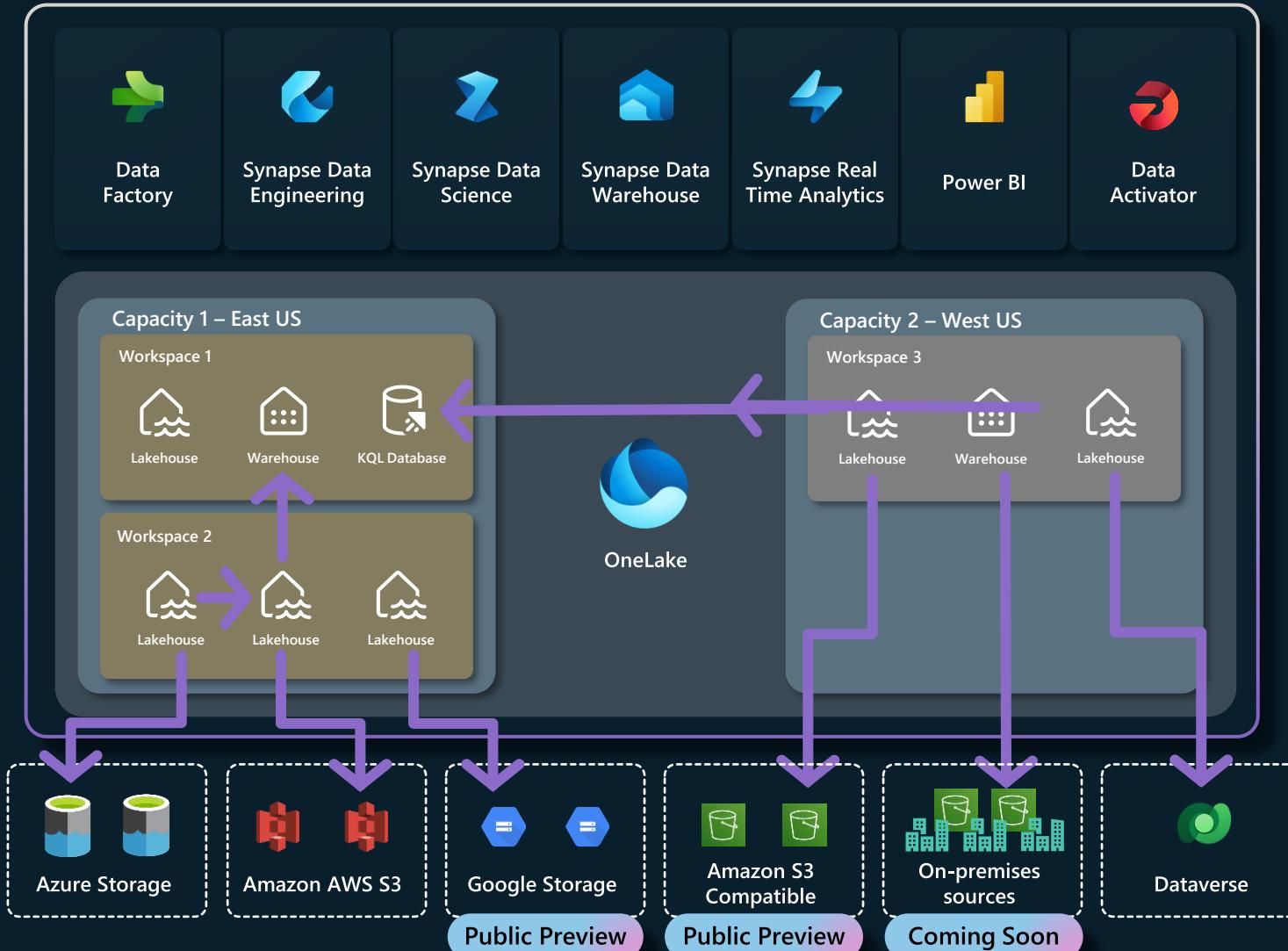
Unifying data in OneLake

Shortcuts



Unifying data in OneLake

External Data Sharing (Preview)





Microsoft Fabric

The data platform for the era of AI

Complete Analytics Platform

Everything, unified

SaaS-ified

Secured and governed

Lake centric and open

OneLake

One Copy

Open at every tier

Empower Every Business User

Familiar and intuitive

Built into Microsoft 365

Insight to action

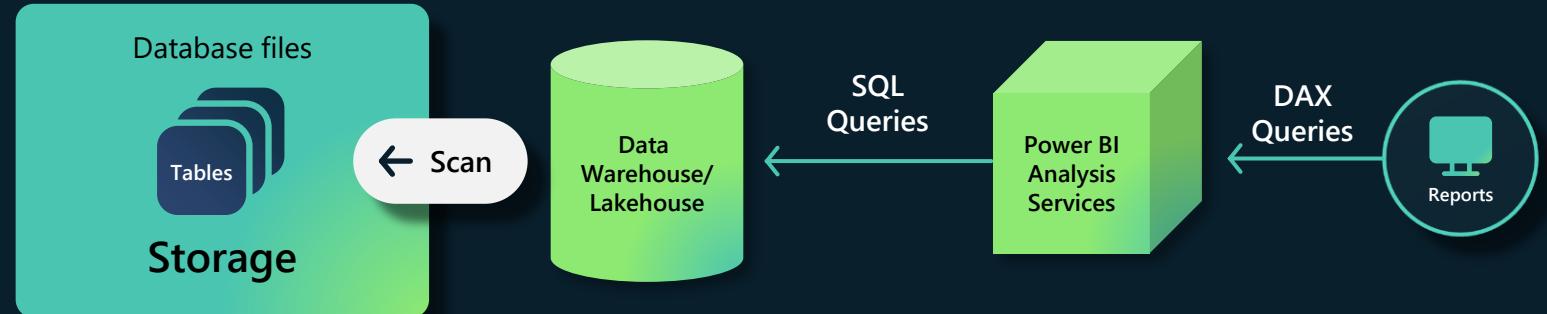
Secure and governed

Manage your data estate

Secure, Protect & comply

Encourage data discovery, trust and use

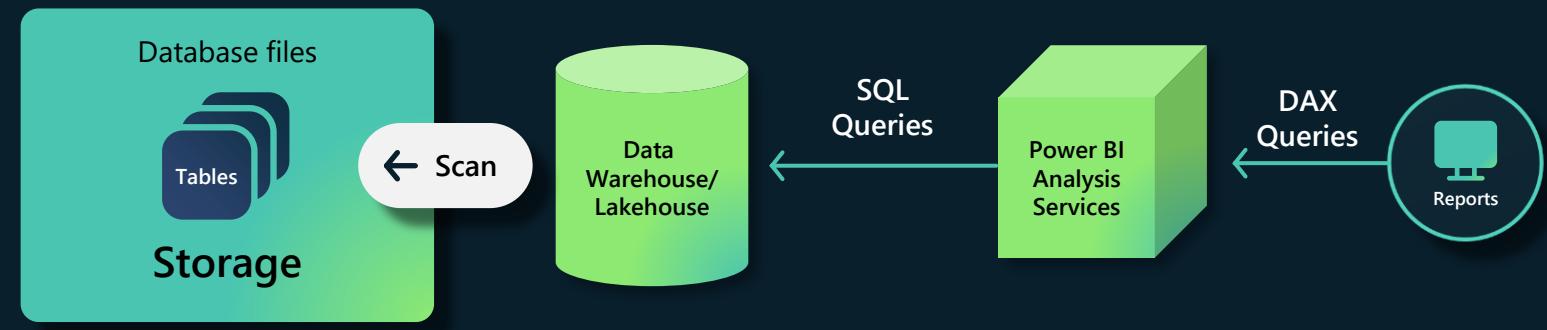
“Direct Query Mode” Slow, but real time



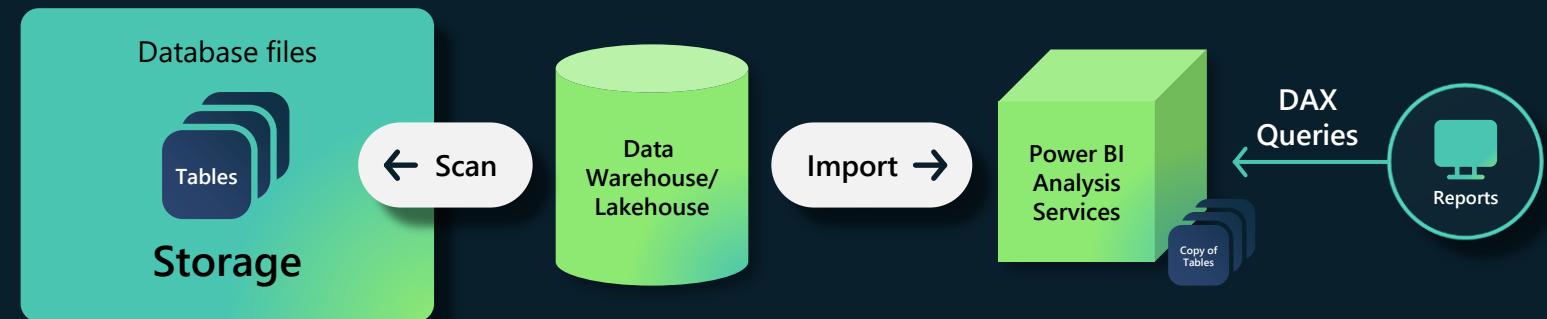
“Import Mode” Latent & duplicative but fast



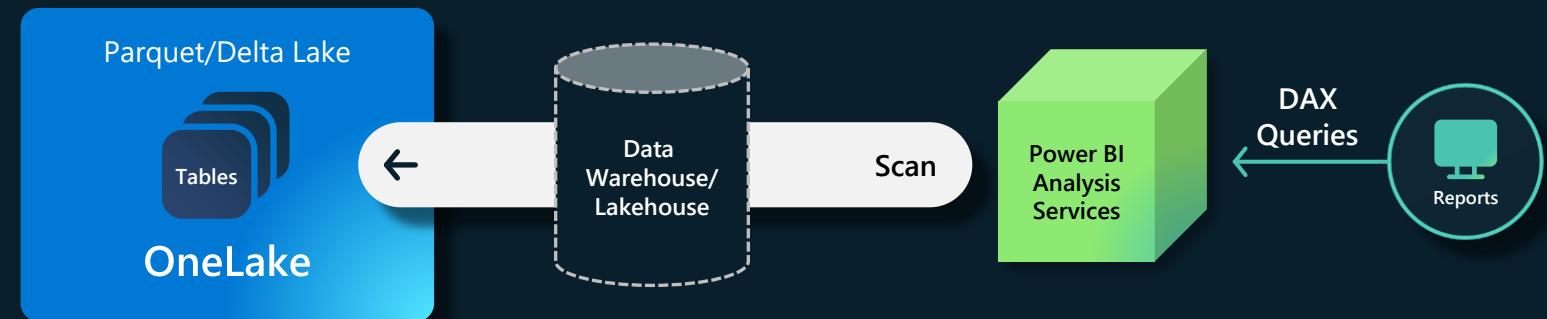
“Direct Query Mode”
Slow, but real time



“Import Mode”
Latent & duplicative but fast



“Direct Lake Mode”
Perfect!





Microsoft Fabric

The data platform for the era of AI

Complete Analytics Platform

Everything, unified

SaaS-ified

Secured and governed

Lake centric and open

OneLake

One Copy

Open at every tier

Empower Every Business User

Familiar and intuitive

Built into Microsoft 365

Insight to action

Secure and governed

Manage your data estate

Secure, Protect & comply

Encourage data discovery, trust and use

Manage, govern, and secure data in Microsoft Fabric



Manage your data estate

Create a logical data mesh with domains to simplify management and increase data usage

Gain fine-grained control over your tenant, capacity, and workspace settings



Secure, protect & comply

Secure inbound and outbound network connections

Ensure only the right people have access internally

Maintain compliance with the strictest requirements



Encourage data discovery, trust, and use

Manage, catalog, and discover data in a logical, searchable data hub

Endorse and enrich high-quality data to create & promote curated, discoverable sources of truth

View the lineage for data items to identify origin and find other relevant sources



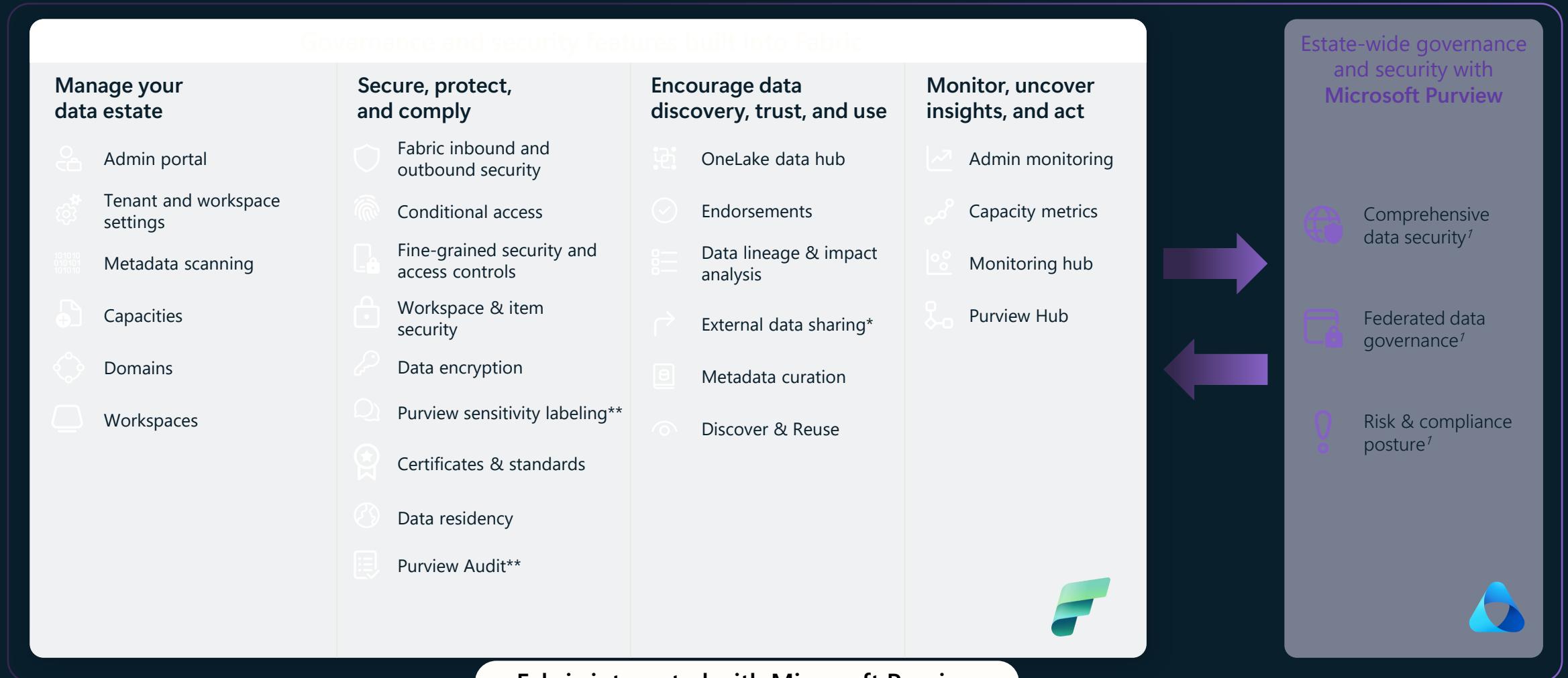
Monitor, uncover insights, and act

Track metrics important to admins such as usage, billing, and access

See relevant insights about sensitive data and item endorsement

Give users the ability to closely monitor their Fabric activities

Gain full visibility and control with industry-leading features



*Features coming soon



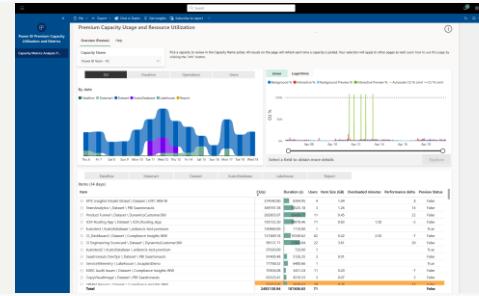
**Additional Microsoft Purview purchase required





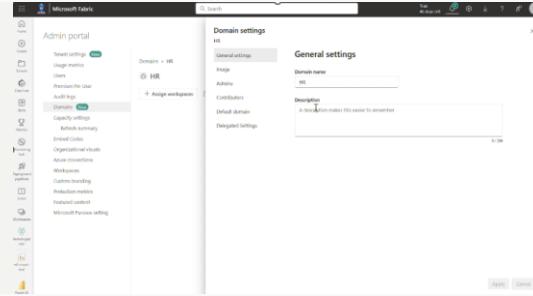
Manage your data estate

Create a logical data mesh to simplify management and increase usage



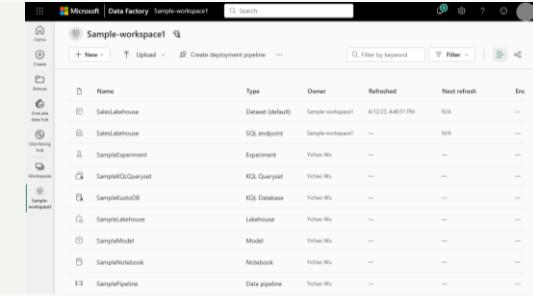
Capacities

Easily manage multiple capacities in a single tenant to **better manage cost**



Domains and subdomains

Group data and workspaces together into a domain and sub domains to establish a **data mesh architecture**



Workspaces

Group **items** like lakehouses, warehouses, and reports into workspaces to organize and more easily manage items together



Domains and subdomains

Create a federated data mesh by grouping together data relevant to a particular area or field (e.g., department) in domains and subdomains.

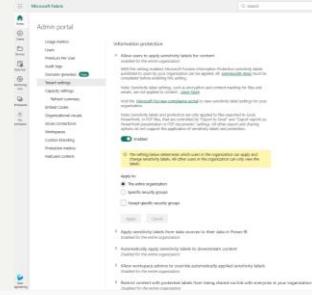
Find relevant data faster by filtering and finding content by domain and subdomains.

Assign domain roles and admins and delegate some governance control to the domain-level, enabling each business unit to define its own rules and restrictions according to its specific business needs.

The screenshot shows the Microsoft Power BI OneLake data hub interface. The left sidebar contains navigation links: Home, Create, Browse, Data hub (selected), Apps, Metrics, Monitoring hub, Deployment pipelines, Learn, Workspaces, testedogadmin (highlighted in green), rdl-visual-test, and Power BI. The main area is titled "OneLake data hub" and displays "Recommended" content cards for "omop_2", "DataflowsStagingLakehouse", "Lake123Sergey", and "testlhra". Below these cards is a search bar with filters: All, My data, Endorsed in your org, Favorites. At the bottom is a table of data assets:

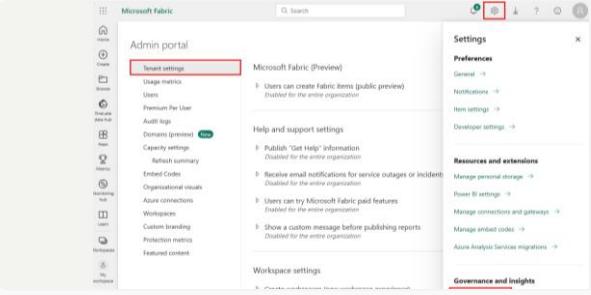
Name	Type	Owner	Endorsement
SU02 smoke	Semantic model	Admin Edog	-
Selection Perf Test Sample	Semantic model	Admin Edog	-
test-report	Semantic model	Admin Edog	-
Grocery Location Analytics v3_tienyu	Semantic model	Admin Edog	Certified

Gain fine-grained control over settings across your tenant



Admin portal

Centrally manage, review and apply tenant, domain and capacity settings for the entire Fabric tenant



Tenant and workspace settings

Gain fine-grained control over the features that are made available to your organization and govern workspaces, domains, and capacities across your tenant



Connect to external tools

Easily extract information such as item name, owner, domain id, lineage, sensitivity label, and endorsement status for all items across Fabric, and connect to your 3rd party or homegrown tools



Admin portal

View usage metrics, users, audit logs, domains, workspace settings, connections, and much more all from a single, unified admin portal.

Centrally manage, review and **apply tenant, domain and capacity settings** for the entire Fabric tenant.

Automate admin operations using nearly a hundred APIs.

The screenshot shows the Microsoft Fabric Admin portal interface. On the left, there's a sidebar with icons for Home, Create, Browse, Data hub, Monitoring hub, Workspaces, and My workspace. The 'Admin portal' section is selected. The main content area has a title 'Admin portal' and a sidebar with links: Usage metrics, Users, Premium Per User, Audit logs, Domains (preview) (New), Tenant settings (selected), Capacity settings, Refresh summary, Embed Codes, Organizational visuals, Azure connections, Workspaces, Custom branding, Protection metrics, and Featured content. The 'Information protection' section contains the following settings:

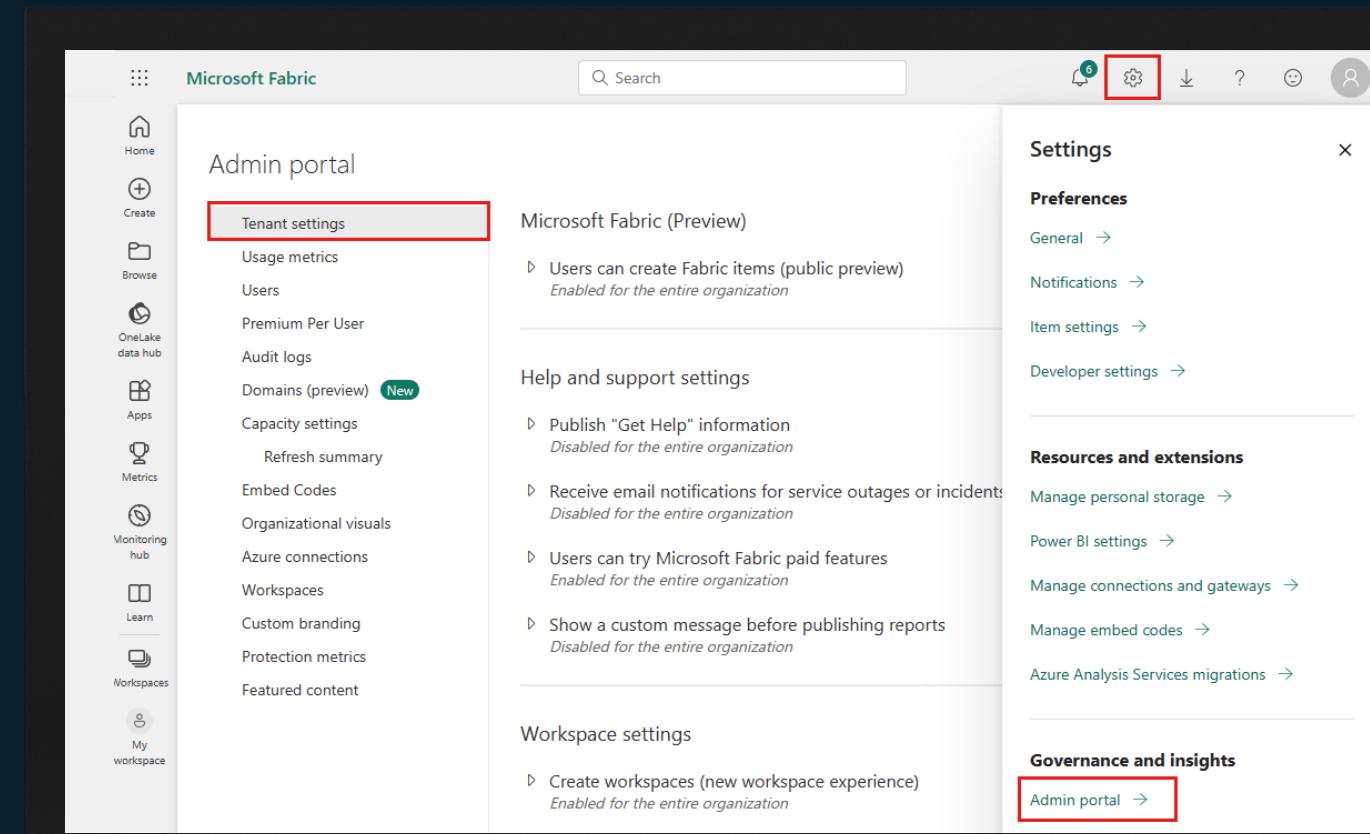
- Allow users to apply sensitivity labels for content**: Enabled for the entire organization. Note: Sensitivity label settings, such as encryption and content marking for files and emails, are not applied to content. [Learn More](#). Visit the [Microsoft Purview compliance portal](#) to view sensitivity label settings for your organization.
- Enabled**: A toggle switch is set to 'Enabled'. A note says: 'The setting below determines which users in the organization can apply and change sensitivity labels. All other users in the organization can only view the labels.'
- Apply to:**
 - The entire organization
 - Specific security groups
 - Except specific security groups
- Apply** and **Cancel** buttons.
- Apply sensitivity labels from data sources to their data in Power BI**: Enabled for the entire organization.
- Automatically apply sensitivity labels to downstream content**.

Tenant and workspace settings

Gain fine-grained control over the **features** that are made available to your organization.

Manage settings for preview features, trial and support, workspaces, information protection, export and sharing, integration, and much more

See detailed descriptions of each workspace and configure the workspace's access list, assigned capacity, description, current state, and more



Metadata scanning via API

Fabric's Scanner APIs enables effective scanning of Fabric's tenant metadata and lineage, including item's name, owner, sensitivity label, endorsement, data source, etc.

Robust enterprise capabilities like authentication with service principal and incremental scans.

Supports 1st and 3rd party catalog tool integration.

Sample Response
Status code: 200

JSON

```
{
  "workspaces": [
    {
      "id": "d507422c-8d6d-4361-ac7a-30074a8cd0a1",
      "name": "V2 shared",
      "type": "Workspace",
      "state": "Active",
      "isOnDedicatedCapacity": false,
      "reports": [
        {
          "id": "c6d072d1-ed20-4b60-8329-16c4b934203b",
          "name": "CompositeModelParams-RLS",
          "datasetId": "132593c4-bf8d-4548-8f25-1ebb16a1613c",
          "createdDateTime": "2020-06-16T08:22:49.14",
          "modifiedDateTime": "2020-06-16T08:22:49.14",
          "modifiedBy": "john@contoso.com",
          "reportType": "PaginatedReport",
          "endorsementDetails": {
            "endorsement": "Certified"
          }
        }
      ]
    }
  ]
}
```

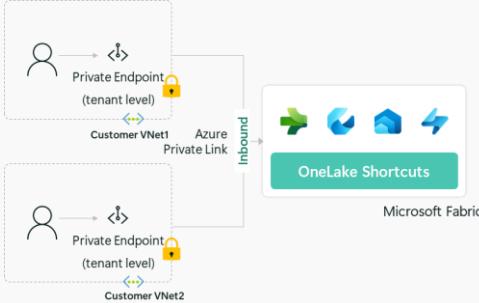


Unmatched security and
compliance

Security layers in Microsoft Fabric

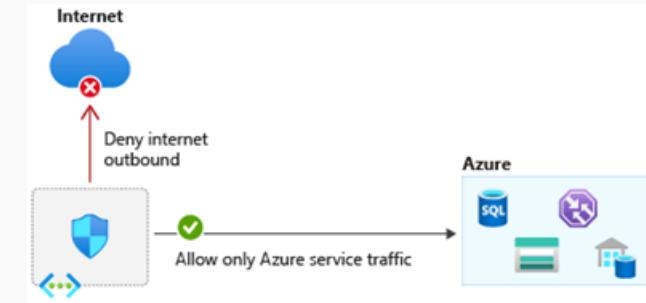


Secure inbound and outbound network connections



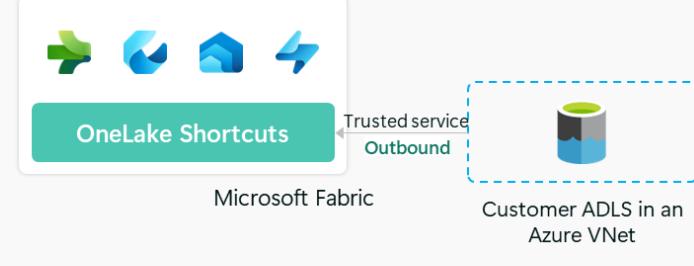
Fabric inbound security

Seamlessly secure [inbound data access](#) scenarios with Fabric's holistic identity and network security



Service tags

Minimize the complexity of updating network security rules using Azure service tags to [group](#) and [manage IP addresses](#) for a service



Fabric outbound security

Securely [connect to your protected data sources](#) and/or ingest data into OneLake



*Features in Public Preview

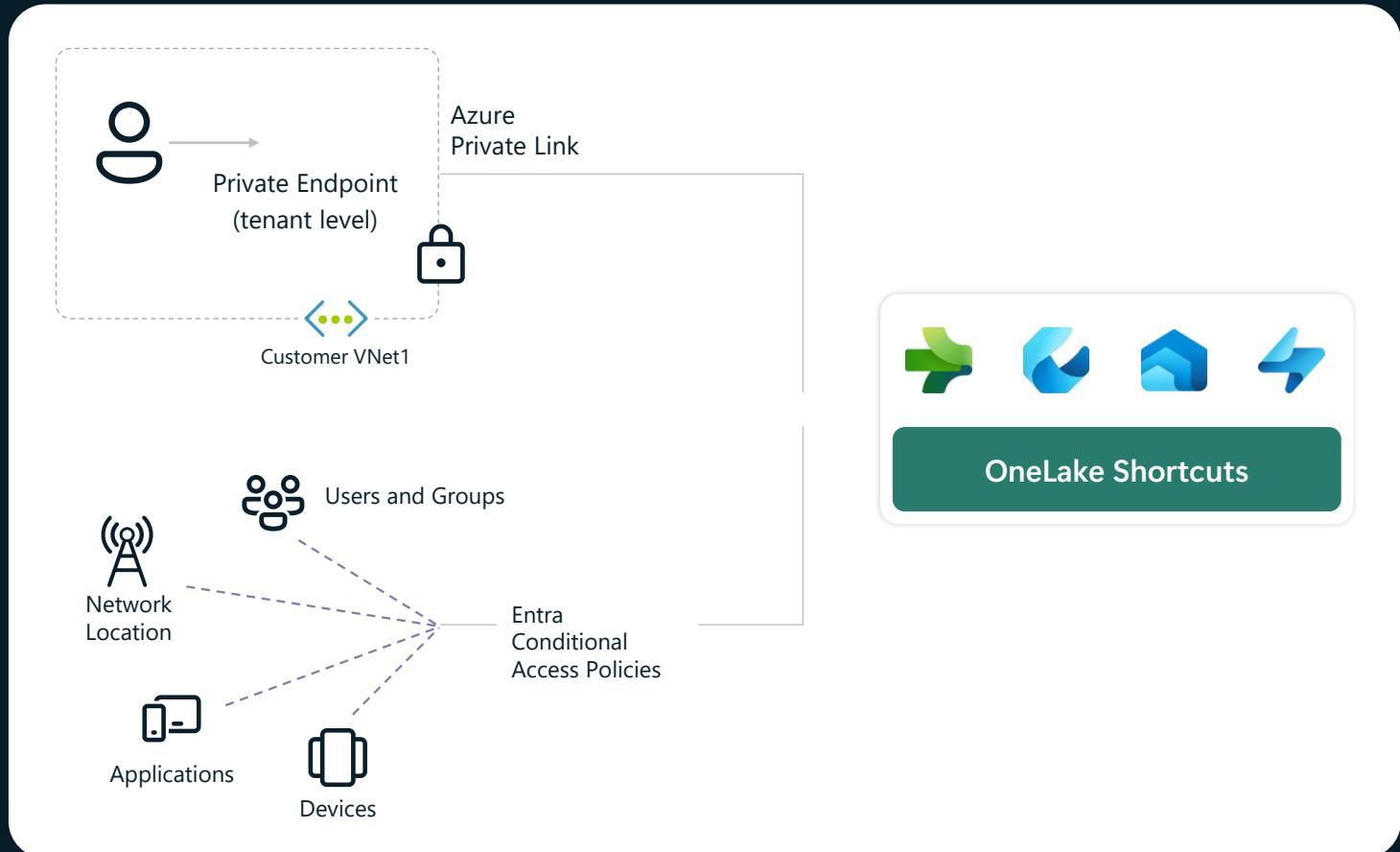


Fabric inbound security

Implement **Conditional Access** policies through Microsoft Entra which restrict access based on user, group, network location, application, device, and risk detection.

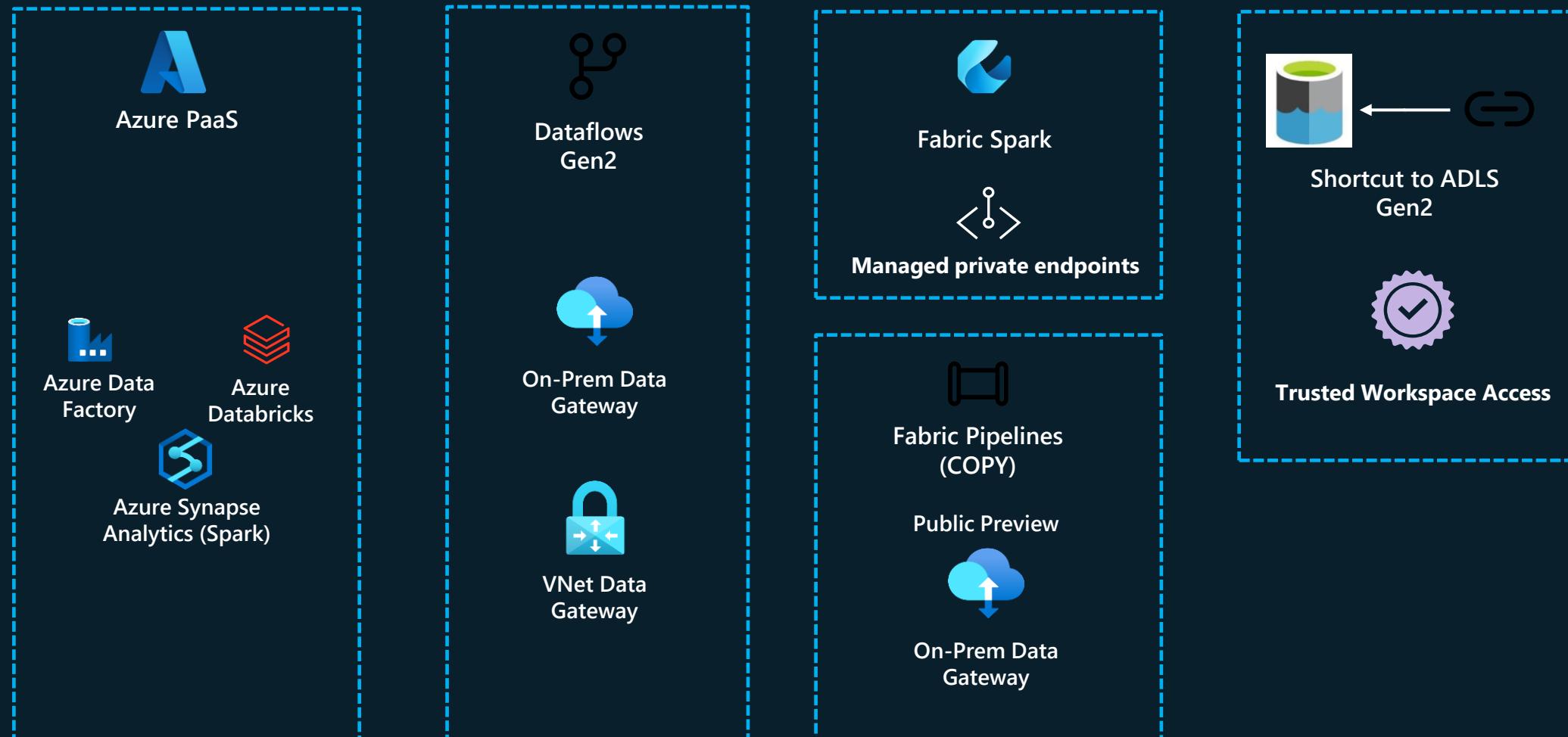
Connect to **datasets and pipelines** using Fabric tenant-level Azure Private Links which allow you to limit access to Fabric from specific VNets*

Use **Azure service tags** to easily manage connections to Microsoft Fabric



Getting data into Fabric

How to connect/load data in the VNET from Fabric



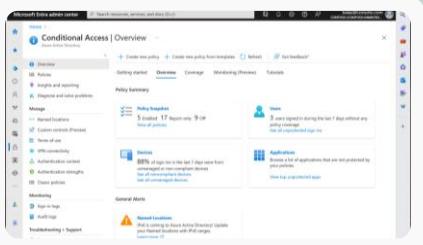
Fabric outbound security

Securely connect to your **protected data sources** and/or ingest data into OneLake:

- Using **shortcuts** to ADLS Gen2 that can connect to data behind VNets without copying the data*
- Using **Fabric Pipelines (COPY)** that can connect to data sources that are deployed in an Azure VNet, such as: Azure SQL VM, Azure SQL MI, etc.
- Using **Fabric Dataflows Gen2** and customer-managed Enterprise Data Gateways to connect to customer databases behind VNets (e.g., Azure SQL DB, Cosmos DB, Snowflake, or even on-prem)
- Using **non-Fabric services** like Azure Data Factory or Azure Databricks that can work with data behind a VNet



Ensure only the right people have access to the right data



Conditional access

Use the industry-leading cloud-based identity and access management service, Microsoft Entra ID, to secure access to Microsoft Fabric





Workspace and item security

Easily manage security at the [workspace and item level](#), ensuring only the right people have access

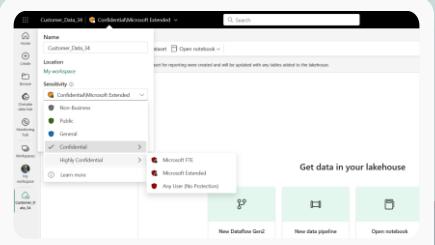




Fine-grained security and access controls

Define granular access to data in Fabric items using [fine-grained access control](#) like object and row level security

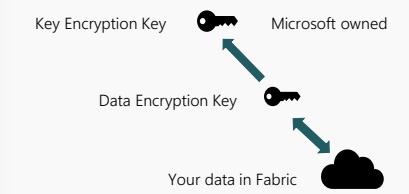




Information protection labels

Classify sensitive Fabric data using the same [sensitivity labels](#) that are used in Microsoft 365—enforced even when the data is exported





Data encryption

All Fabric data stores are [encrypted at rest and in transit](#) including system data and metadata



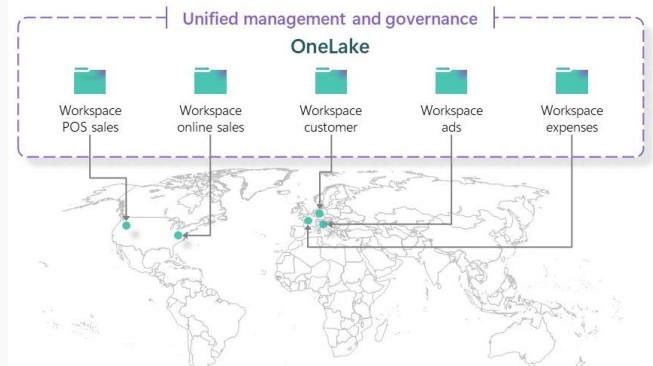
*Additional purchase required

Remain in compliance with even the strictest requirements



Certifications and standards

Microsoft Fabric, as a core online service, supports a wide range of compliance standards including GDPR, EUDB, HIPAA, ISO certifications and more



Data residency

Control where your data is stored and meet data residency requirements with over 54 data centers world-wide



A screenshot of the Microsoft Purview Audit interface. The top navigation bar includes "Audit", "New Search", "Classic Search", and "Audit retention policies". Below this, there are sections for "Active unified searches" and "Audit retention policies". A main table lists audit logs with columns for "Search name", "Job status", "Progress %", "Search #", "Total results", and "Created time". The table shows several entries, such as "Mar 23 - Apr 24 createwarehouse.createwarehouse.cs..." and "Apr 23 - Apr 24 AdminUser". The bottom right corner of the screenshot indicates "31 items".

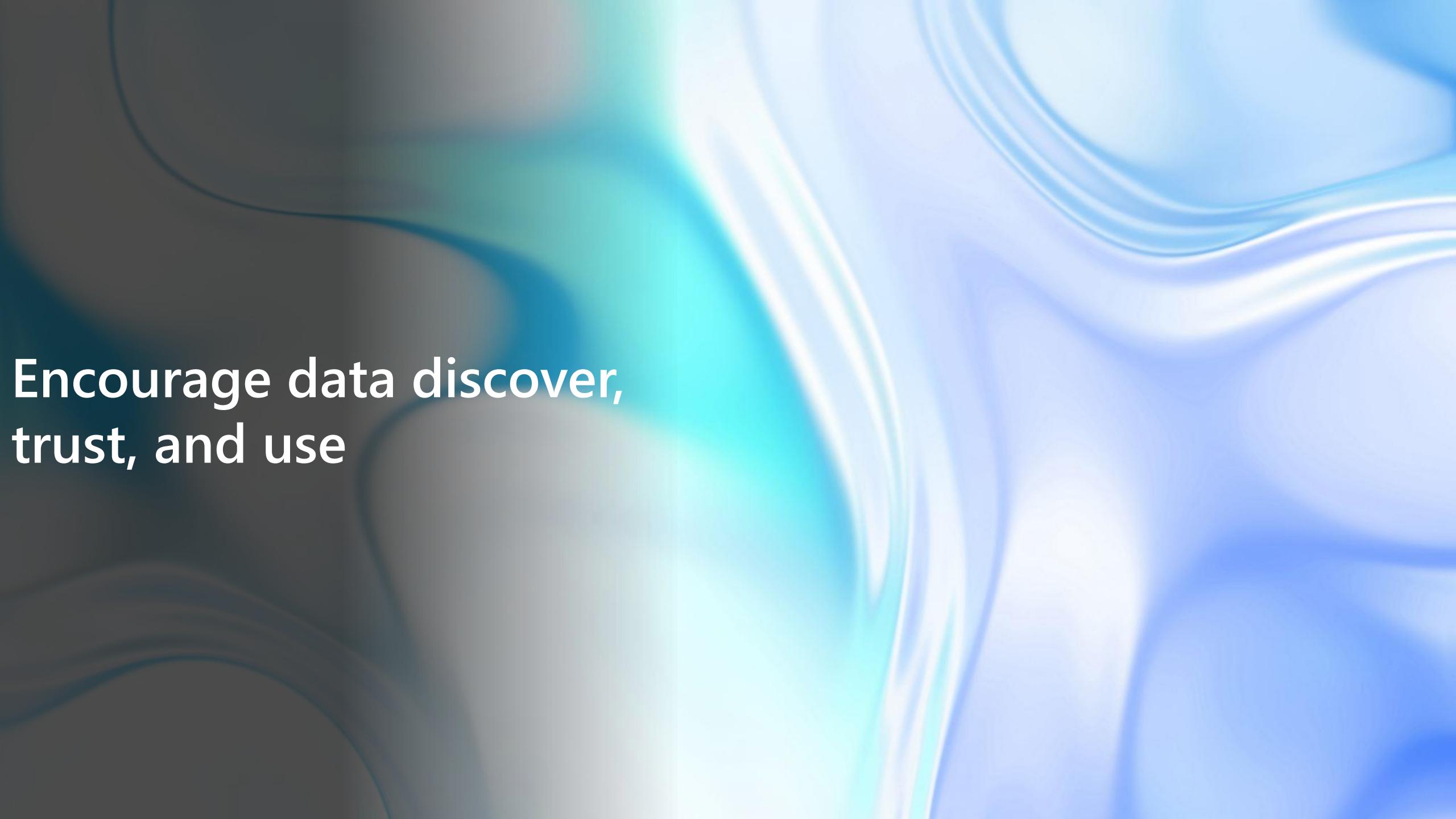
Microsoft Purview Audit*

Log user activities from Microsoft Fabric to support security, forensic, and internal investigations



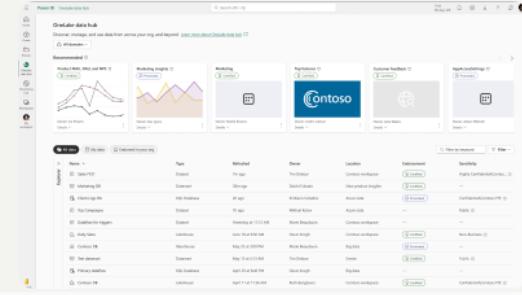
*Additional Microsoft Purview purchase required



The background of the slide features a dynamic, abstract design composed of several overlapping, curved bands of color. The colors transition from dark teal on the left to bright cyan in the center, and then to light blue and white on the right. These curves create a sense of motion and depth, resembling waves or energy flows.

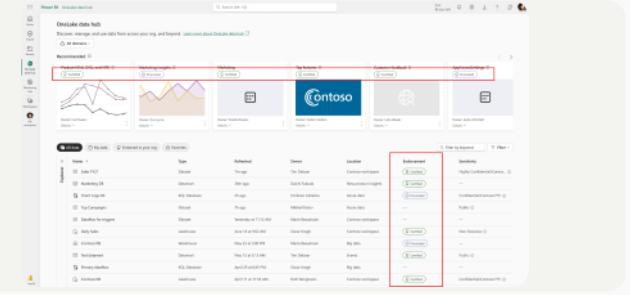
Encourage data discover,
trust, and use

Effectively manage your data to create a trusted data hub



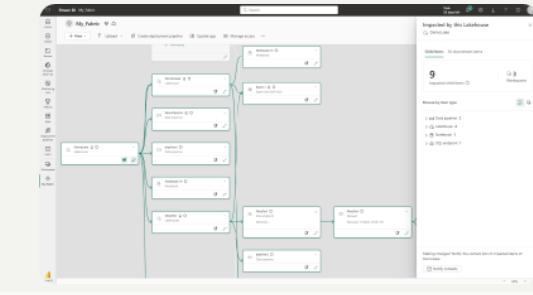
OneLake data hub

Find, explore, and [use](#) the Fabric data items in your organization that you have access to in one, central hub



Endorsements

Endorse or mark items as discoverable high-quality content in Fabric to [create curated sources of truth](#) and increase discoverability and reuse



Data lineage and impact analysis

See lineage view of analytical projects to [see how data flows](#) through items and perform impact analysis to [assess impact of changes](#)



OneLake Data Hub

Easily [discover, manage, and reuse](#) all your data from a single hub that's automatically indexed for searchability, governance, and compliance.

Quickly find the data you are looking for using the [search, filter, and sort](#) capabilities including specifying domain, workspace, artifact and more.

Discover data via the OneLake data hub across 15 Fabric experiences and in [Microsoft Teams](#), [Excel](#), and soon in the [Azure Portal](#).

The screenshot shows the Power BI OneLake data hub interface. On the left, a sidebar menu includes Home, Create, Browse, OneLake data hub (which is selected and highlighted), Monitoring hub, Metrics, Apps, Deployment pipelines, Learn, Workspaces, and My workspace. The main area has a header "Power BI OneLake data hub" with a search bar and a "Trial 59 days left" message. Below the header, it says "OneLake data hub" and "Discover, manage, and use data from across your org." with a link to "Learn more about OneLake data hub". A "Recommended" section displays cards for "DemoLake", "LHDemo1", "DemoLake2", "DemoLakehouse2", "DemoLake3", and "KustoDB Io". Each card shows a small icon, the dataset name, and a "Certified" badge. Below this is an "Explorer" table with columns: Explorer, Name, Type, Owner, Refreshed, Location, Endorsement, and Sensitivity. The table lists various datasets and lakehouses, each with a "Certified" badge and specific details like owner names (Yaron Canari, Omer Sade) and locations (My_Fabric, BizHealth). The "Endorsement" column shows "Certified" for most entries, while "Sensitivity" includes "Confidential/FTE", "Public", "General", and "Non-Business".

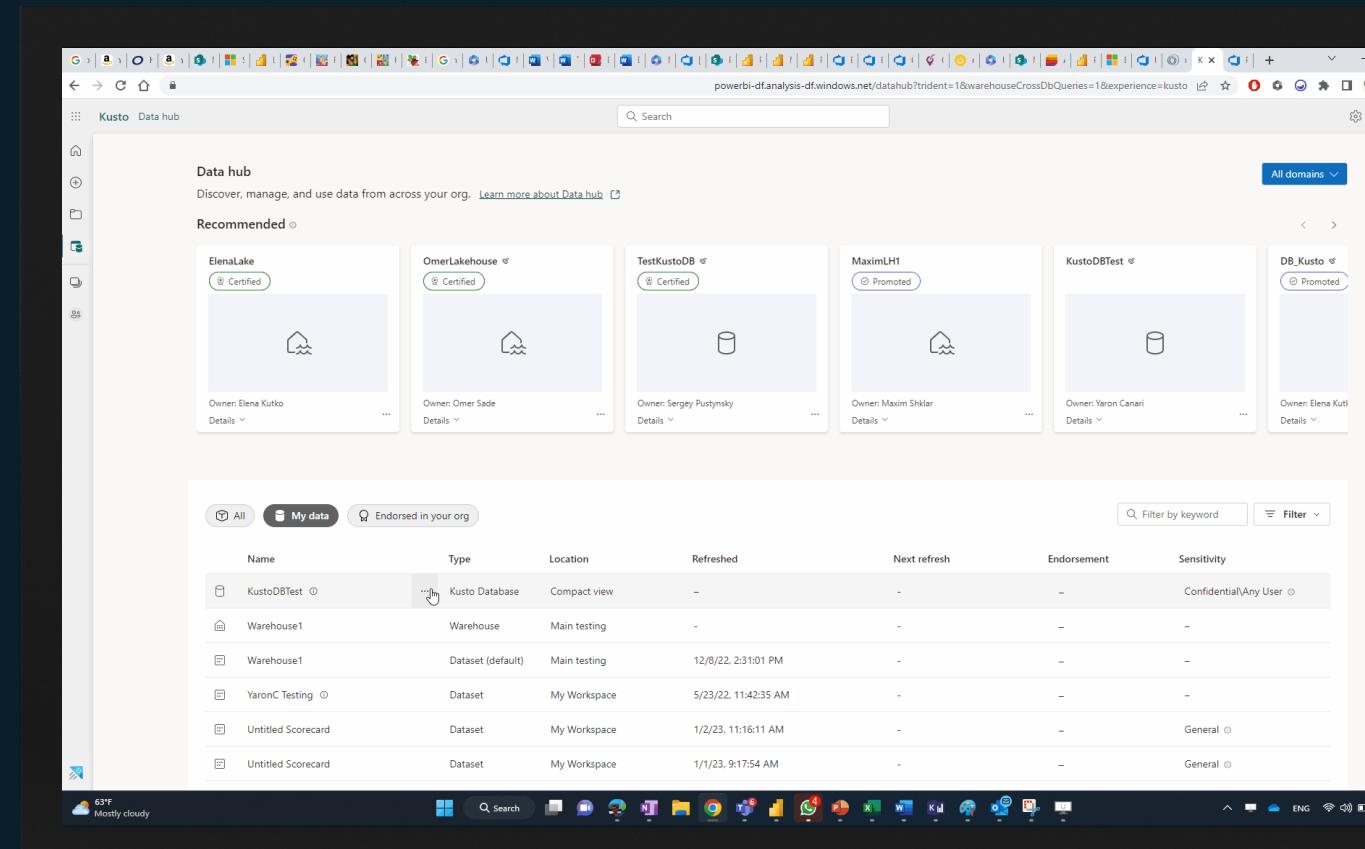
Explorer	Name	Type	Owner	Refreshed	Location	Endorsement	Sensitivity
All	DemoLake	Lakehouse	Yaron Canari	-	My_Fabric	Certified	Confidential/FTE
My Workspace	Sales health	Dataset	Yaron Canari	1/22/23, 6:27:31 PM	BizHealth	Certified	Public
I_Impact_analysis_1							
akj deployment 2							
Alexandr Gavrishev's w...							
Assafshemeshfsdfsdsg...							
bbash							
BizHealth							
Boba Inventory							
CDS for Apps raw data							

Endorsements

Data owners can **endorse** valuable, high-quality content in Fabric to create curated sources of truth, increase their discoverability, and encourage reuse.

Authorized users can **certify data** that meets the organization's quality standards and anyone can **promote data** they find valuable and recommend for others to use.

Endorsed items get **higher visibility** in Fabric data discovery experiences.

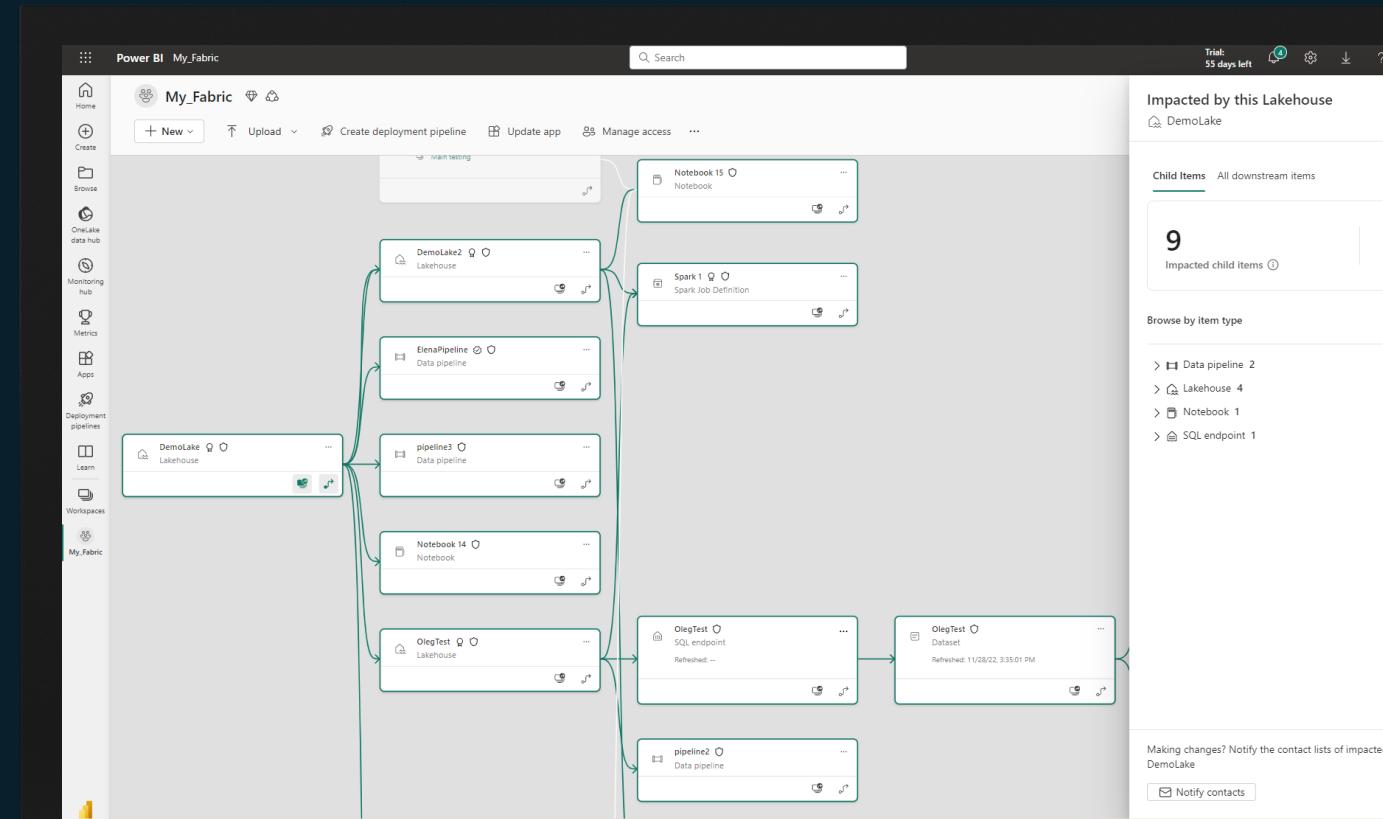


Data Lineage and impact analysis

See [lineage view](#) of complex analytical projects to understand how the [data flows](#) through all Fabric items, including lakehouses, data warehouses, pipelines, datasets and reports.

Troubleshoot data refresh issues while tracing the [data](#) back from the destination to the source and monitor refresh status.

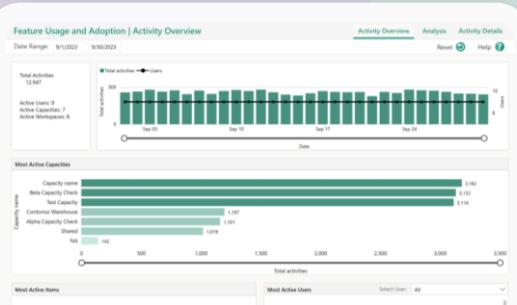
Perform [impact analysis](#) to assess impact for upcoming changes.





Monitor, uncover insights
and act

Easily monitor activity across your tenant and uncover relevant insights



Admin monitoring

Uncover insights for effective governance with pre-built reports and datasets built for admins

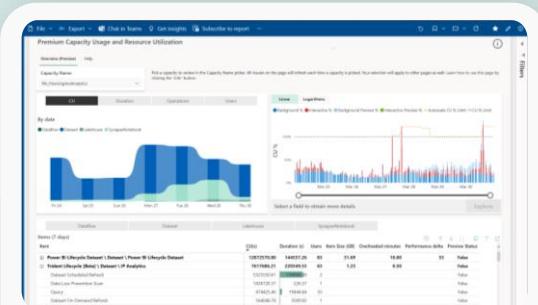


The Admin center interface allows users to manage domains. It features a sidebar with options like Tenant settings, Usage metrics, Users, Premium Per User, Audit logs, and Capacity settings. The main area shows a list of domains with columns for Name, Description, Domain admins, and Delete.

Name	Description	Domain admins	Delete
Finance	Finance related data	-	-
Health	Health related data	-	-
Education	Education related data	-	-

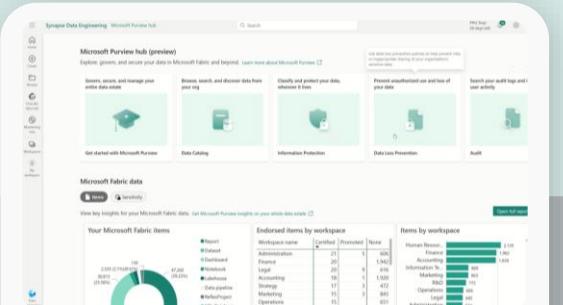
Monitor

Monitor Fabric activities, such as semantic model refresh and Spark Job runs and many others, from a central location



Capacity metrics

Gain visibility into capacity utilization and usage trends with daily aggregation so you can plan and scale your capacity accordingly



Purview hub

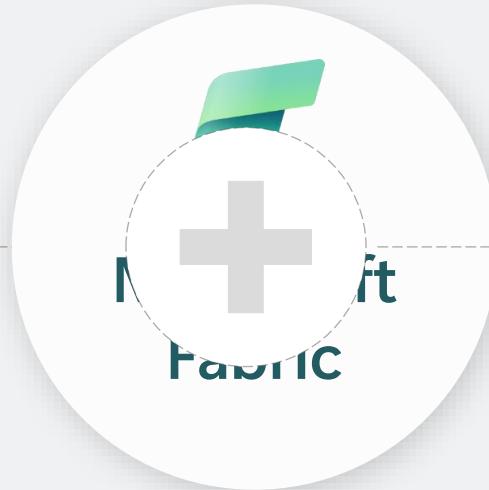
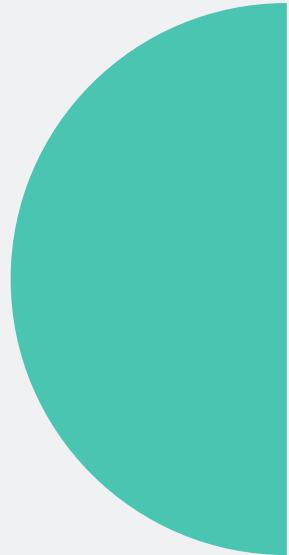
See relevant insights about sensitive data and item endorsement across the Fabric data estate and get easy links to more advanced Purview capabilities





Govern and secure your
entire data estate

Manage, govern, and secure data in Microsoft Fabric



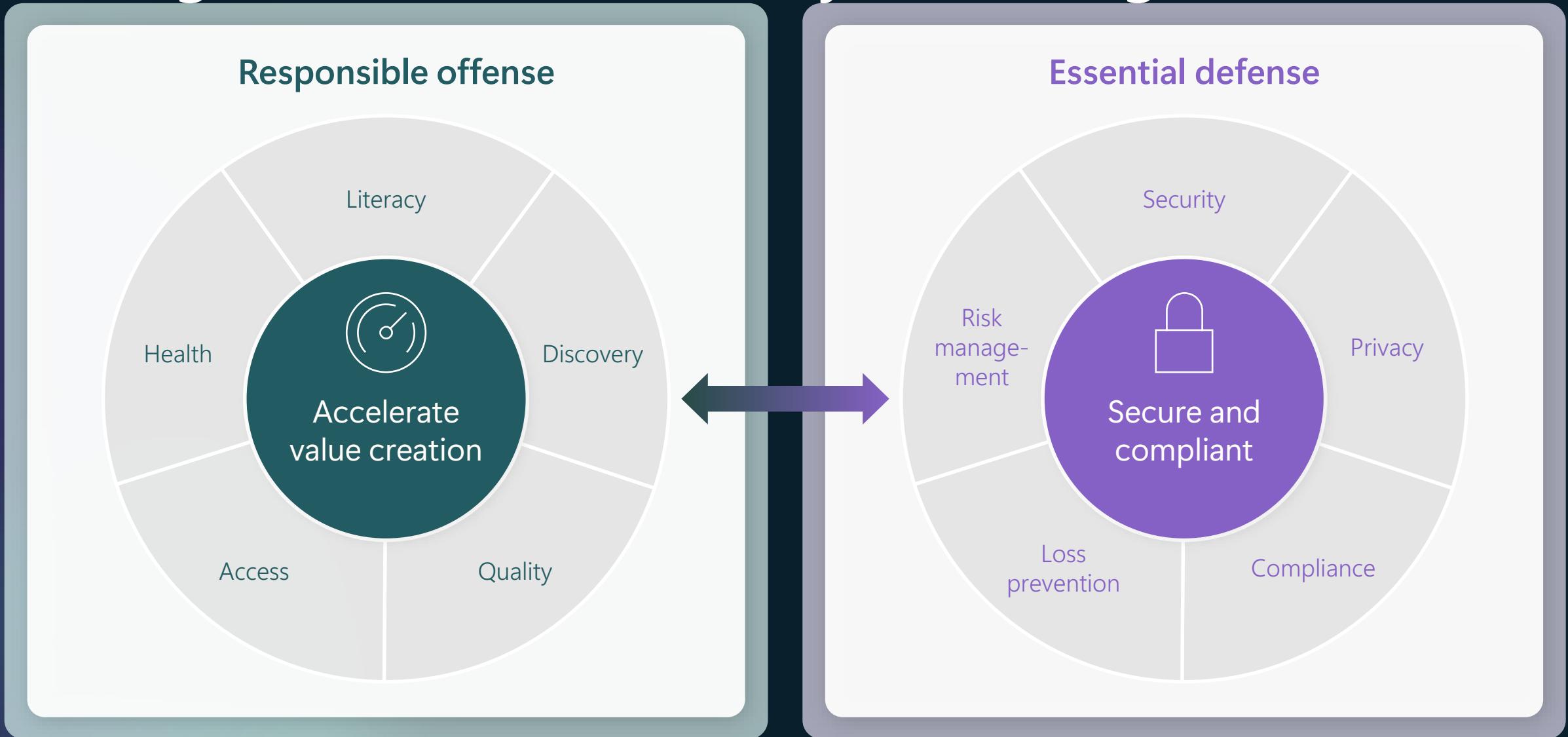
Reshape how you access, manage, and ~~keep~~ your data safe and governed with unified data and insights with Microsoft Fabric data governance, security, and compliance

Unmatched security & compliance

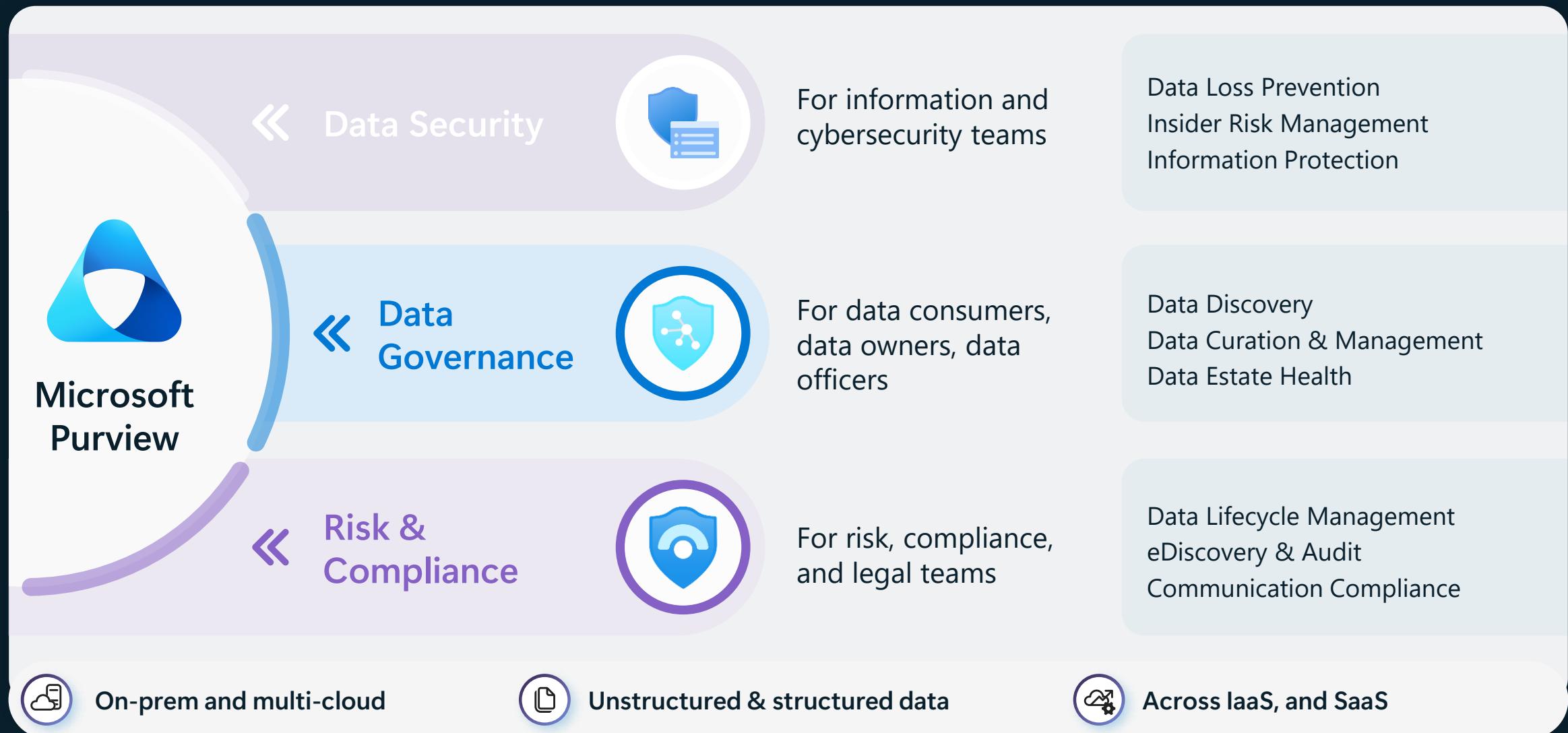
Simplified administration

Unified governance

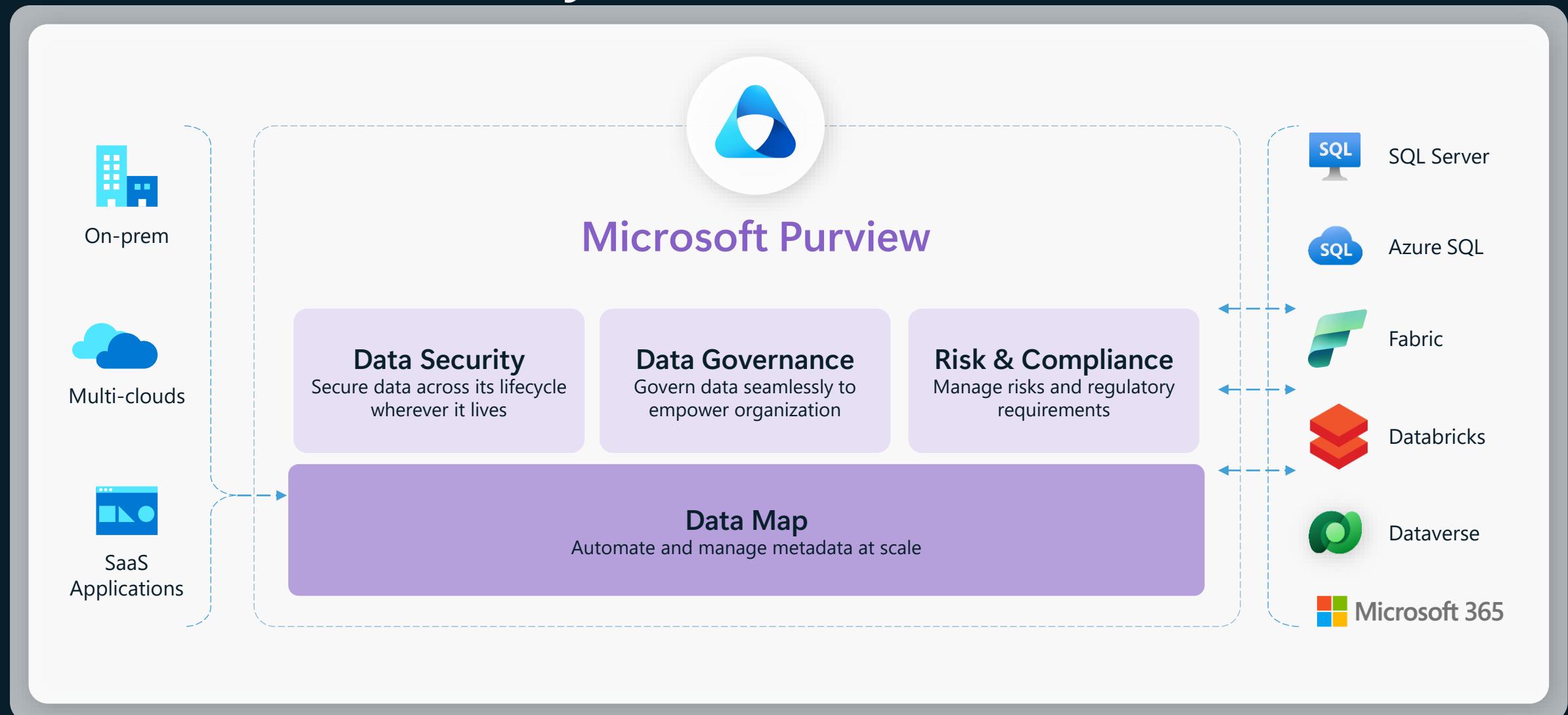
Data governance and security is evolving



Bringing data security, governance, and compliance together



Govern and secure your data estate



Data Governance for the Age of AI

The whole is greater than the sum of the parts.



Microsoft Purview
Free version





Microsoft Purview (Free Version)

- SaaS version
- Support of the following data types
 - Azure Blob Storage
 - Azure Data Lake Storage
 - Azure SQL Database
 - Subscriptions
 - Fabric

The screenshot shows the 'Browse by source type' page in the Microsoft Purview interface. At the top, there are tabs for 'All sources', 'Microsoft Azure', 'Microsoft Fabric', and 'Other', with 'All sources' being the active tab. A search bar labeled 'Filter by keyword' is also present. The main area is divided into sections for 'Microsoft Azure' and 'Microsoft Fabric'. Under 'Microsoft Azure', there are five items: 'Azure subscriptions' (3+ items), 'Azure Blob Storage' (1+ items), 'Azure Data Lake Storage Gen2' (4+ items), 'Azure SQL Database' (6+ items), 'Azure SQL Server' (4+ items), and 'Azure Storage Account' (5+ items). Under 'Microsoft Fabric', there is one item: 'Fabric workspaces'. Each item is represented by a small icon and a label with the source type and the number of items.



Microsoft Purview (Free Version)

- Annotated Assets Limit
 - If you have 50 live view data assets and add descriptions and classifications to 15 of those assets, you will have 15 annotated assets.
 - Descriptions and classifications on the same asset count only as one annotated asset.
- Data Curator Access:
 - Only data curators can access resources added to the data map using the REST APIs in the free version.
 - Compliance portal
- Account Cleanup
 - If your free version tenant remains inactive for three months, Microsoft reserves the right to clean up the account and may remove unused resources.
- Data catalog - browse and search for your data assets
- Data share - securely share data between users and organizations
- Data map history - a log of updates made to assets



Microsoft Purview (Free Version)

Free	Enterprise
Auto discovery of Azure data sources	Collections, and glossary terms to organize your data map
Role group access control to platform and apps	Fine-grained, collection-level access control to platform and apps
Definition and manual application of classifications and terms	Automated application of classifications and terms
Manual descriptions, tags, and contacts	Manual and Bulk descriptions, tags, and contacts.
<u>Manual creation of assets</u> using the <u>data map APIs</u>	Full use of <u>Microsoft Purview's REST APIs</u>
	Automated scans for the hybrid data estate
	Workflows
	Business rules
	Support for business assets and managed attributes



Microsoft Purview (Free Version)

- Upgrade?



Microsoft Purview Enterprise
Better coverage and more apps

Upgrade to Microsoft Purview Enterprise to govern, secure, and manage data in multicloud and hybrid environments. Also, get access to apps across data governance, data security, compliance, and privacy. [Learn more](#)

Upgrade account **Not now**

Upgrade to Microsoft Purview Enterprise

Country/region ⓘ
Netherlands

Azure subscription *

Resource group *

ⓘ The Microsoft Purview Enterprise tier will be provisioned with 1 capacity unit (CU) with auto scale capabilities. 1 CU = 25 operations/second and 10 GB of metadata storage. [Learn more](#)

I acknowledge that I have read, understood, and agree to the [Terms and Conditions](#)

Upgrade **Cancel**

Microsoft Purview Enterprise version



Links for the Click-Through Chapters

- Setup Business Domains -
<https://purviewdatagovernance.storylane.io/share/920vqytm1ech>
- Scan the Data Estate -
<https://purviewdatagovernance.storylane.io/share/of6j1c1q2c2l>
- Publish Data Products -
<https://purviewdatagovernance.storylane.io/share/8ubzzhh1npwl>
- Improve Data with Data Quality -
<https://purviewdatagovernance.storylane.io/share/u1046l6beqse>
- Monitor with Data Estate Health -
<https://purviewdatagovernance.storylane.io/share/wkojjbsxb90l>
- Enable Data Democratization for Consumers -
<https://purviewdatagovernance.storylane.io/share/mzzwqtfgdpre>

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

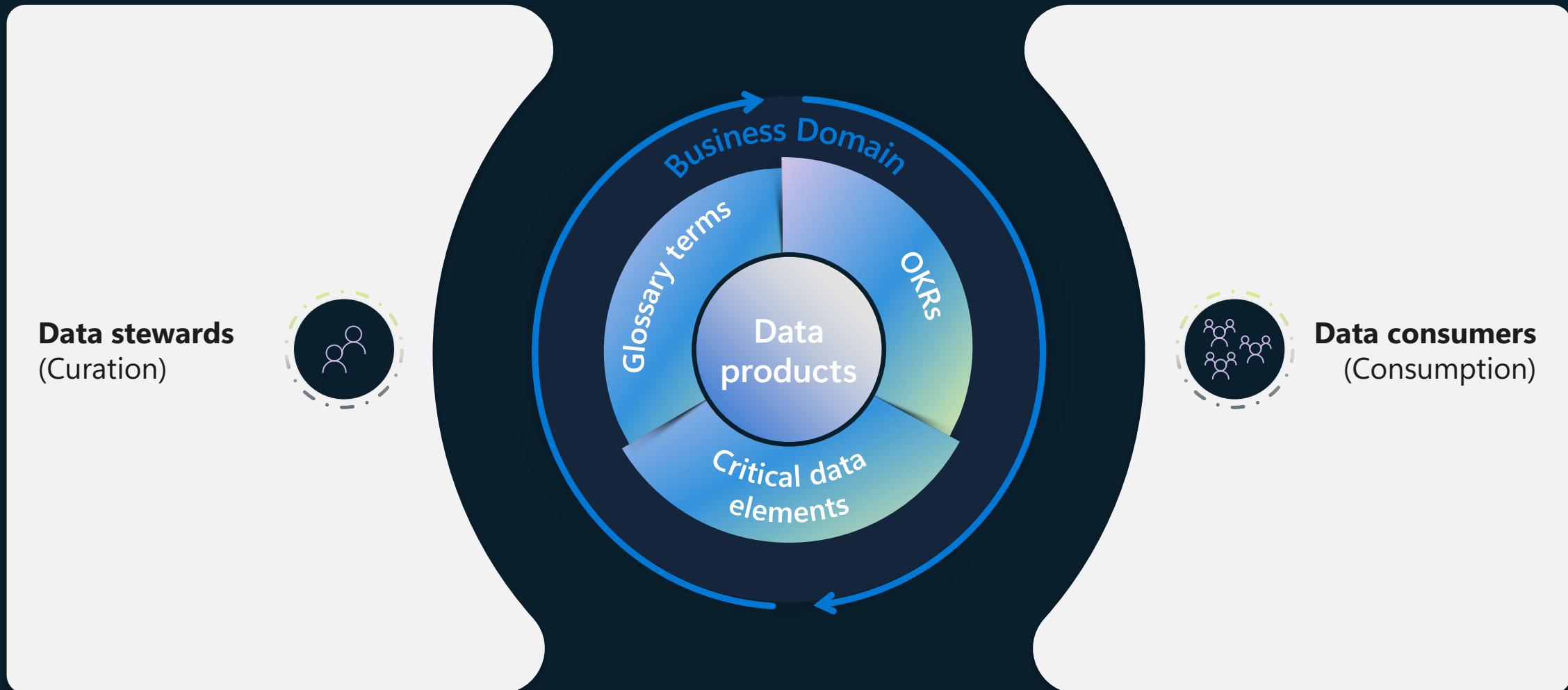
- Annotating data sources with descriptive metadata is often a lost effort. Client applications typically ignore descriptions that are stored in the data source.
- Creating documentation for data sources can be difficult and it's an ongoing responsibility to keep documentation in sync with data sources. 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.

Know your data with business concepts



Microsoft Purview Enterprise version

Demo / Whirlwind Tour



Business Domains

Business domains

Organize data products into meaningful groups and link them to business concepts. [Learn more about business domains](#)

3 items

+ New business domain

Business domain name ↑	Owner	Status
C Corporate	Wolfgang Strasser	Draft
P Personal Health	Wolfgang Strasser	Published
S Sales	Wolfgang Strasser	Published

Details **Roles** **Data estate mappings**

Sales Business domains

Type Owner Status Data quality score Health actions

Functional unit Wolfgang Strasser Published Healthy 92.8 6

This is the sales business domain.
More information will follow...

Business concepts

Data products	Glossary terms	OKRs
1	0	0
Go to data products	View all	View all

Critical data elements Custom attributes

0 0

[View all](#) [View all](#)

Details Roles Data estate mappings

Sales Business domains

Type Owner Status Data quality score Health actions

Functional unit Wolfgang Strasser Published Healthy 92.8 6

This is the sales business domain.
More information will follow...

Business concepts

Data products	Glossary terms	OKRs
1	0	0
Go to data products	View all	View all

Critical data elements Custom attributes

0 0

[View all](#) [View all](#)

What's a Business Domain?

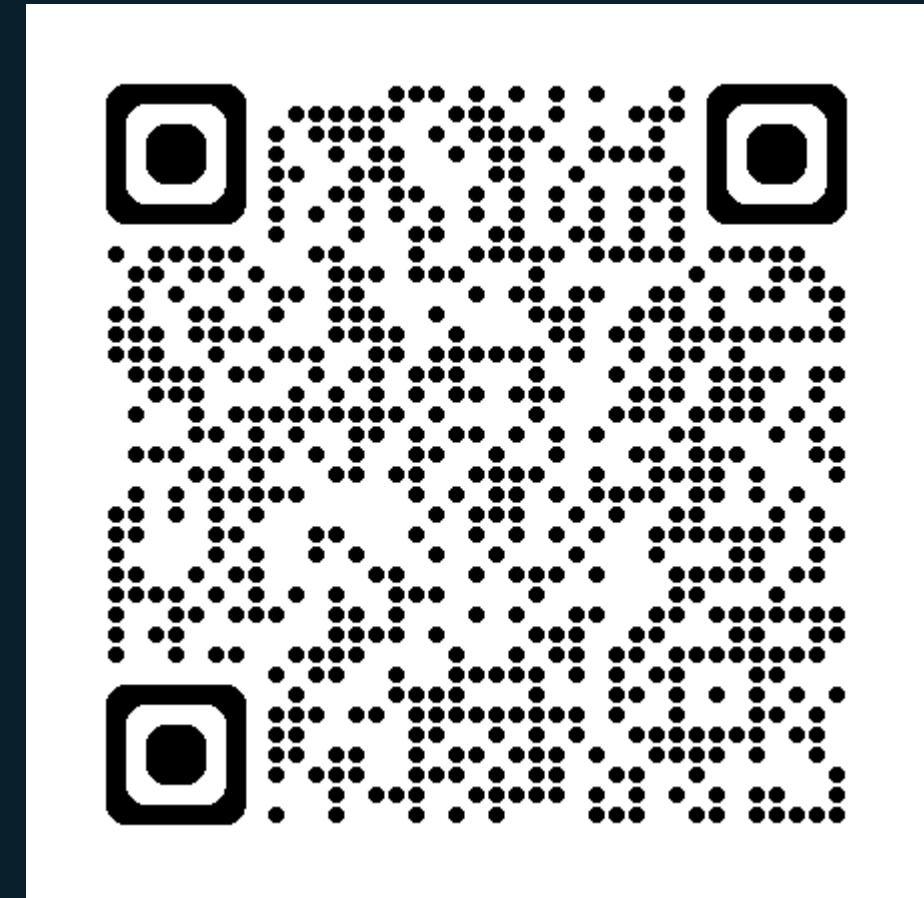
- ... boundary that enables the common governance, ownership, and discovery of data products and business concepts like glossary terms, OKRs, or critical data.

Examples of Business domains:

- **Fundamental business areas** – human resources, sales, finance, supply chain, ..
- **Subject areas** – product, parties, ..
- **Based on organizational functions** – customer experience, cloud supply chain, business intelligence

Microsoft Purview

- Business Domains
- <https://tinyurl.com/7ubush85>



Scan Data

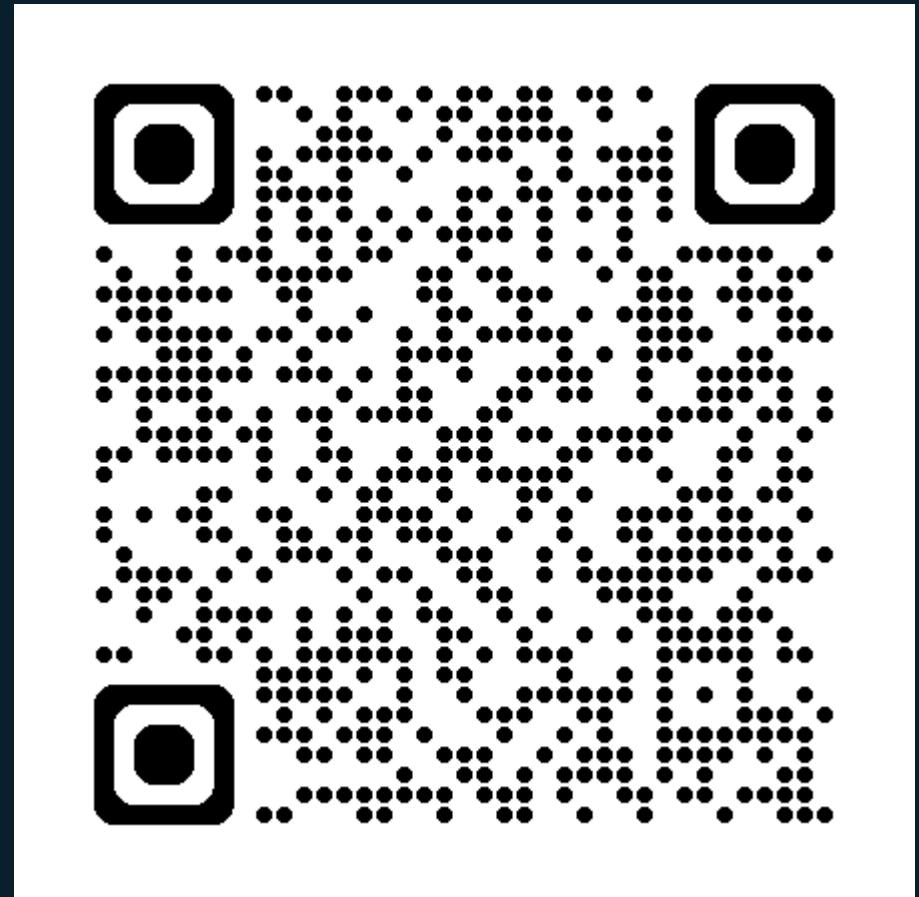
- ... boundary that enables the common governance, ownership, and discovery of data products and business concepts like glossary terms, OKRs, or critical data.

Examples of Business domains:

- **Fundamental business areas** – human resources, sales, finance, supply chain, ..
- **Subject areas** – product, parties, ..
- **Based on organizational functions** – customer experience, cloud supply chain, business intelligence

Microsoft Purview

- Scan Data Estate
- <https://tinyurl.com/2xurj6ry>



Data Products

Data Catalog > Business domains > Data products >

SalesData

Data product | Operational | Contains 8 data assets

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

Description	Business domain	Status
SalesData	Sales	Draft
Use cases	Data product owner	Active Subscribers
Sales data is one of the most important pieces of data in an enterprise.	Wolfgang Strasser	No active subscribers
	Update frequency	Data quality score
--	--	Healthy 92.8
	Health actions	Terms of use
6	No terms of use	Add
	Documentation	
No documentation	Add	

Data assets (8)

Address Azure SQL Table	Another SalesData Power BI Report	CentralCustomerList OneLake Delta Table
Customer Azure SQL Table Customer Table in Azure sQL	Product Azure SQL Table	Purview Customer Data Power BI Dataset

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

Glossary terms (0)

No glossary terms [Add](#)

Critical data elements (0)

No critical data elements

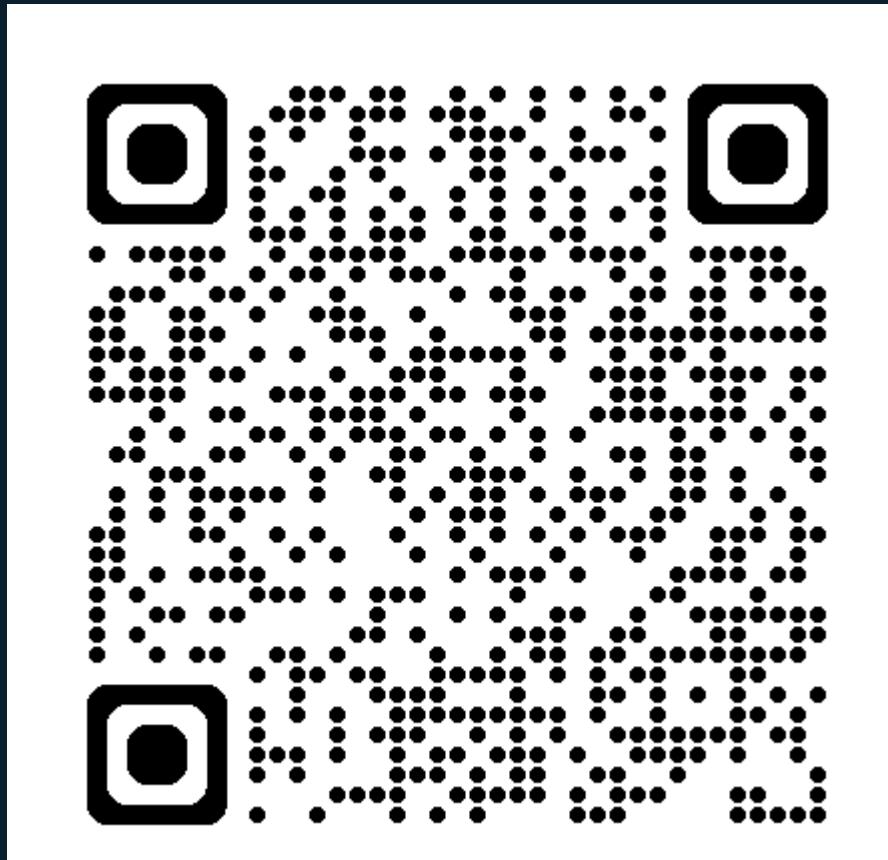
What is a Data Product?

- ... a **group of data assets** (tables, files, PBI reports, ..) – a packaged offering to an enterprise that provides assets with a use case to be shared to data consumers
- .. **Provides context to these assets** (grouping them under a use case for data consumers)
- .. involves the collection, processing, analysis and presentation of data

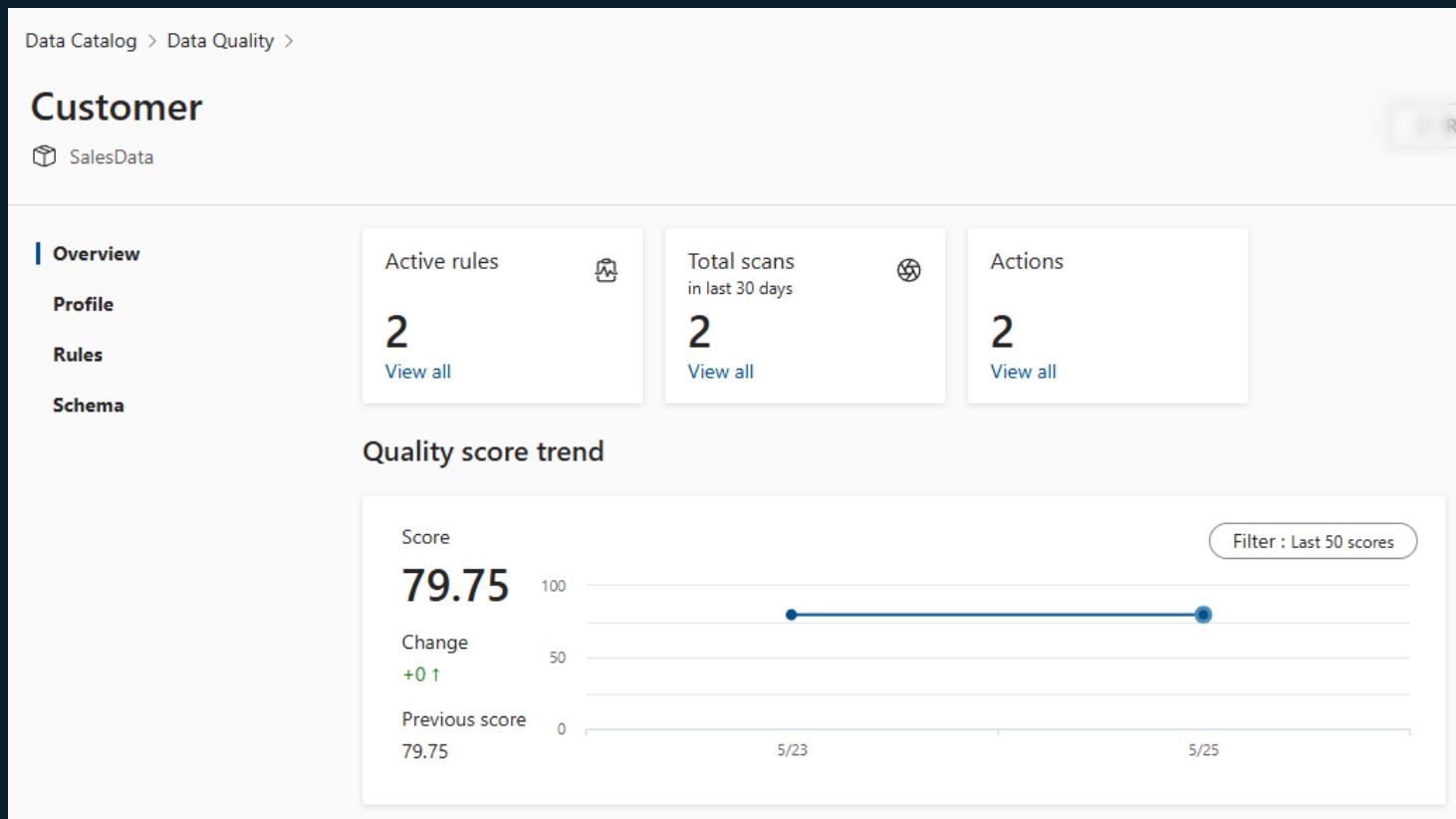
Microsoft Purview

- Publish Data Product

<https://tinyurl.com/47ea2dus>



Data Quality

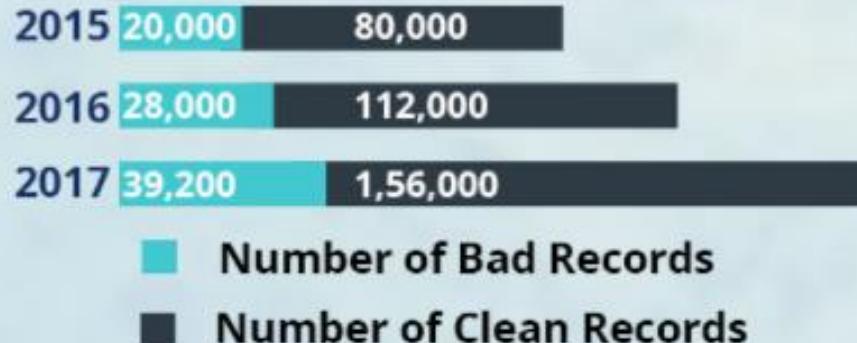


Some Risks of Bad Data (Quality)

- Poor decision-making
- Business Inefficiencies
- Mistrust
- Missed Opportunities
- Siloed Information
- Redundancy
- Lost Reputation
- Lost Revenue

Costs of Poor Data Quality

infoCleanse



Keeping the **\$100 Per Dirty Record** in mind, and using a **100,000-Record Database**, here is the astronomical cost of bad data over three years.

On average, corporate data grows at **40%** per year.

Approximately **20%** of the average database is dirty



Data Exploration (with Power BI)

Country	Units	Revenue
USA	150709	42,938,702.99
	147739	41,644,910.84
Mix	30	23,782.50
Rural	841	123,719.35
Aliqui RP-01 Productivity	284	41,869.59
Pirum RS-01 Select	557	81,849.76
Urban	2099	1,146,290.30
Total	150709	42,938,702.99

Units by Year

Category	Units
(Blank)	147739
Mix	30
Rural	841
Urban	2099
Total	150709

Units by Category

Country

Units

Revenue

Year

Category

Units

Category

Units

Data Profiling (with PowerQuery)

	ABC Country	123 ProductID	Date	123 Zip	123 Units	\$ Revenue	ABC Product Name	ABC Category
3151	USA	1085	01.02.2019	29172	1	89,20	Pirum RP-31 Productivity	null
3152	USA	1085	01.02.2019	92220	1	89,20	Pirum RP-31 Productivity	null
3153	USA	1085	01.02.2019	92592	1	89,20	Pirum RP-31 Productivity	null
3154	USA	1085	01.02.2019	95662	1	89,20	Pirum RP-31 Productivity	null
3155	USA	1085	02.02.2019	39483	1	89,20	Pirum RP-31 Productivity	null
3156	USA	1085	02.02.2019	92307	1	89,20	Pirum RP-31 Productivity	null
3157	USA	1085	03.02.2019	68352	1	86,57	Pirum RP-31 Productivity	null
3158	USA	1085	06.02.2019	91390	1	89,20	Pirum RP-31 Productivity	null
3159	USA	1085	06.02.2019	91504	1	89,20	Pirum RP-31 Productivity	null
3160	USA	1085	06.02.2019	92692	1	89,20	Pirum RP-31 Productivity	null
3161	USA	1085	06.02.2019	94268	1	89,20	Pirum RP-31 Productivity	null
3162	USA	1085	07.02.2019	38139	1	89,20	Pirum RP-31 Productivity	null
3163	USA	1085	08.02.2019	25951	1	89,20	Pirum RP-31 Productivity	null
3164	USA	1085	08.02.2019	33480	1	89,20	Pirum RP-31 Productivity	null
3165	USA	1085	08.02.2019	94061	1	89,20	Pirum RP-31 Productivity	null



= Table.Profile(sales)

	ABC Column	ABC Min	ABC Max	ABC Average	1.2 StandardDeviation	1.2 Count	ABC NullCount	ABC DistinctCount
1	Category	null	null	null	null	145330	142404	null
2	Country	USA	USA	null	null	145330	0	1
3	Date	15.01.2012	31.12.2019	03.06.2016	null	145330	0	2335
4	Product Name	Aliqui MA-02 All Season	Victoria UR-21 Regular	null	null	145330	1151	304
5	ProductID	25	2412	1712,199339	530,7132606	145330	0	457
6	Revenue	41,9475	10550,925	295,4565678	205,7228338	145330	0	879
7	Units	1	41	1,037012317	0,304103942	145330	0	20
8	Zip	1001	99926	62625,28779	28844,11006	145330	0	15827

Friday Afternoon Measurement (FAM) Method

- Poor data quality is troublesome
- Get awareness
- On a Friday afternoon..
 - Assemble the last 100 records of data (customer orders, drawings)
 - Focus on 10-15 critical data elements / attributes => spreadsheet
 - Ask 2 / 3 people for some time
 - Working record by record, mark obvious errors in color
 - Summarize the results
 - Perfect record or not column?

<https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>

<https://hbr.org/2016/07/assess-whether-you-have-a-data-quality-problem>

Friday Afternoon Measurement (FAM) Method

Friday Afternoon Measurement Spreadsheet

In this example, look at the last 100 customer orders and identify how many records are error-free.

Record	ATTRIBUTE 1 Name	ATTRIBUTE 2 Size	ATTRIBUTE 3 Amount	ATTRIBUTE 15	Perfect record?
1	Jane Doe	Null	\$472.13		No
2	John Smith	Medium	\$126.93		Yes
3	Stuart Madnick	XXXL	Null		No
4	Thoams Jones				No
100	James Olsen	24 Lockwood Road	\$76.24		No

Number of perfect records
= 67

SOURCE THOMAS C. REDMAN

© HBR.ORG

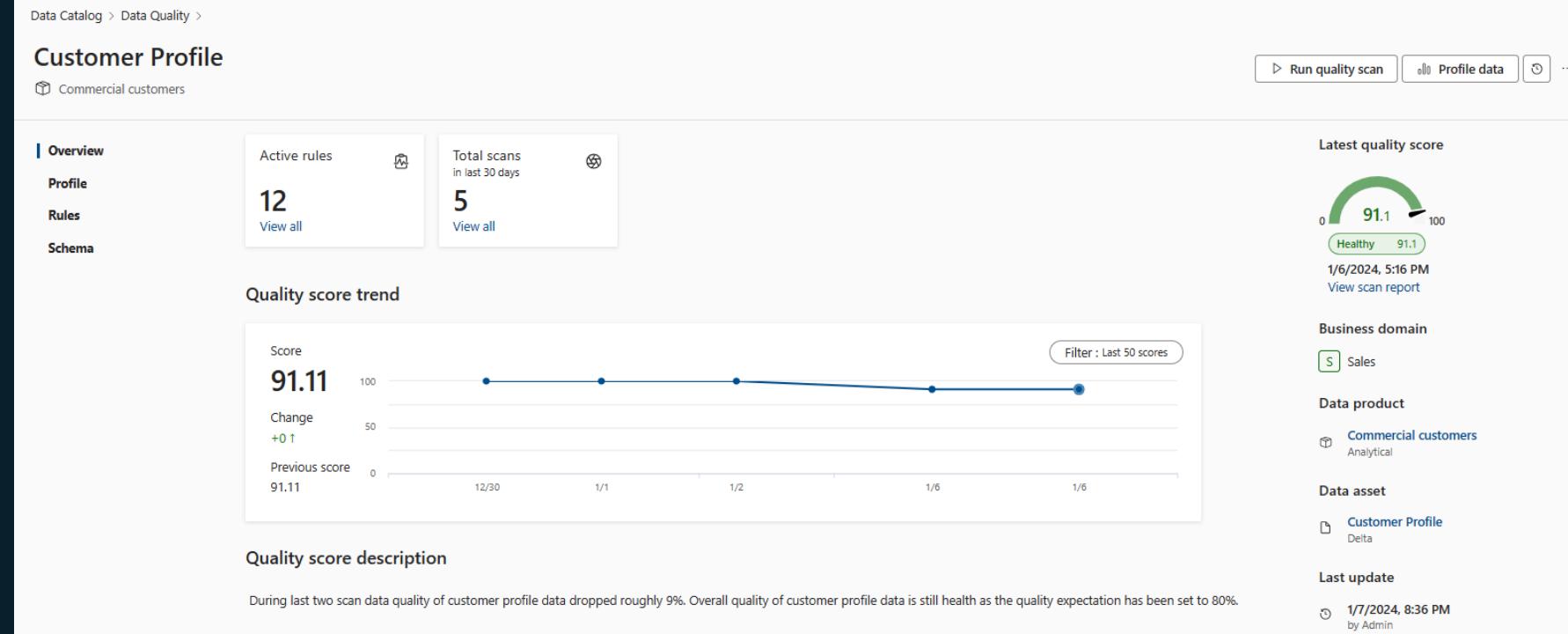
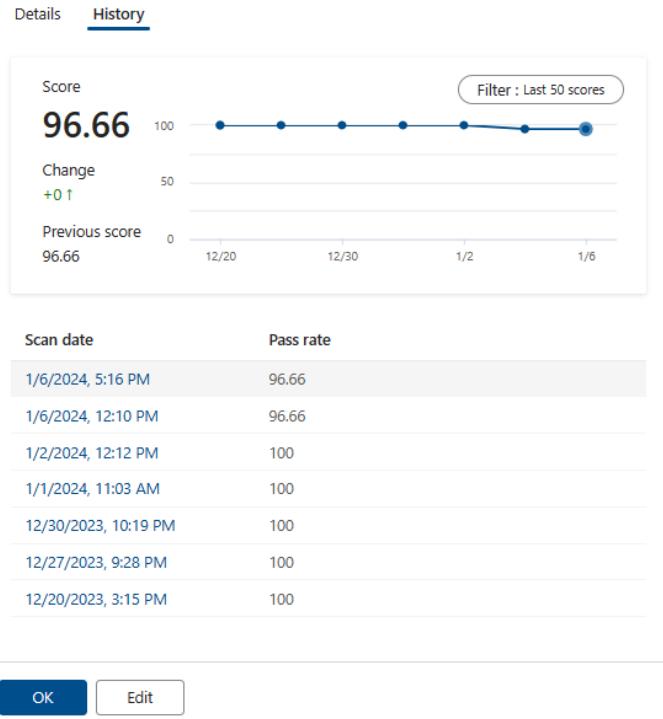
Microsoft Purview Data Quality

- Define DQ rules on Assets (in Data Products)

Rule	Definition
Freshness	Confirms that all values are up to date.
Unique values	Confirms that the values in a column are unique.
String format match	Confirms that the values in a column match a specific format or other criteria.
Data type match	Confirms that the values in a column match their data type requirements.
Duplicate rows	Checks for duplicate rows with the same values across two or more columns.
Empty/blank fields	Looks for blank and empty fields in a column where there should be values.
Table lookup	Confirms that a value in one table can be found in the specific column of another table.
Custom	Create a custom rule with the visual expression builder.

Microsoft Purview Data Quality

Empty/blank fields



Data Quality Actions

Data quality actions

Active In progress Resolved My items

Business domain: Sales X

Target entity type: Data quality ... X

Add filter

Finding name	Days active	Status	Finding type	Reason	Recommendation
Data profile abnormally high null counts de...	18 days	Not started	Data Quality Ma	Rule score for the data asset rule: Empty/blank_fields_MiddleName has fallen below minimum Low severity threshold score: 75 points. Observed value: 59.504 points.	Please run the following query in order to fetch the failed rows for this rule, after replacing the schema and table names. Query: select * from [DBSchemaName].[DBTableName] where [MiddleName] is null
Data profile abnormally high null counts de...	18 days	In progress	Data Quality Ma		
Data profile outlier values detected	18 days	Not started	Data Quality Ma	Job type	Job ID
Data profile outlier values detected	18 days	Not started	Data Quality Ma	Data quality assessment	dd22c517-3cbb-47f3-92ec-569a2491dae3
Data quality Assessment job has failed.	18 days	Not started	Data Quality Management	Data quality job unsuccessful	Product
Data asset quality rule score has fallen belo...	18 days	Not started	Data Quality Management	Data asset quality rule score	Customer

Data asset quality rule score has fallen below threshold

Status

Not started

Assigned to

Active
18 days

Data quality asset

Customer

Business domain

Sales

Business domain owner

Wolfgang Strasser

Business domain

Sales

Finding type

Data Quality Management

Action ID

8ee4a195-dba7-411e-8c83-1e7833ca6afb

Recommendation

Please run the following query in order to fetch the failed rows for this rule, after replacing the schema and table names.

Query: select * from [DBSchemaName].[DBTableName] where [MiddleName] is null

Job type

Data quality assessment

Job ID

dd22c517-3cbb-47f3-92ec-569a2491dae3

Data quality job unsuccessful

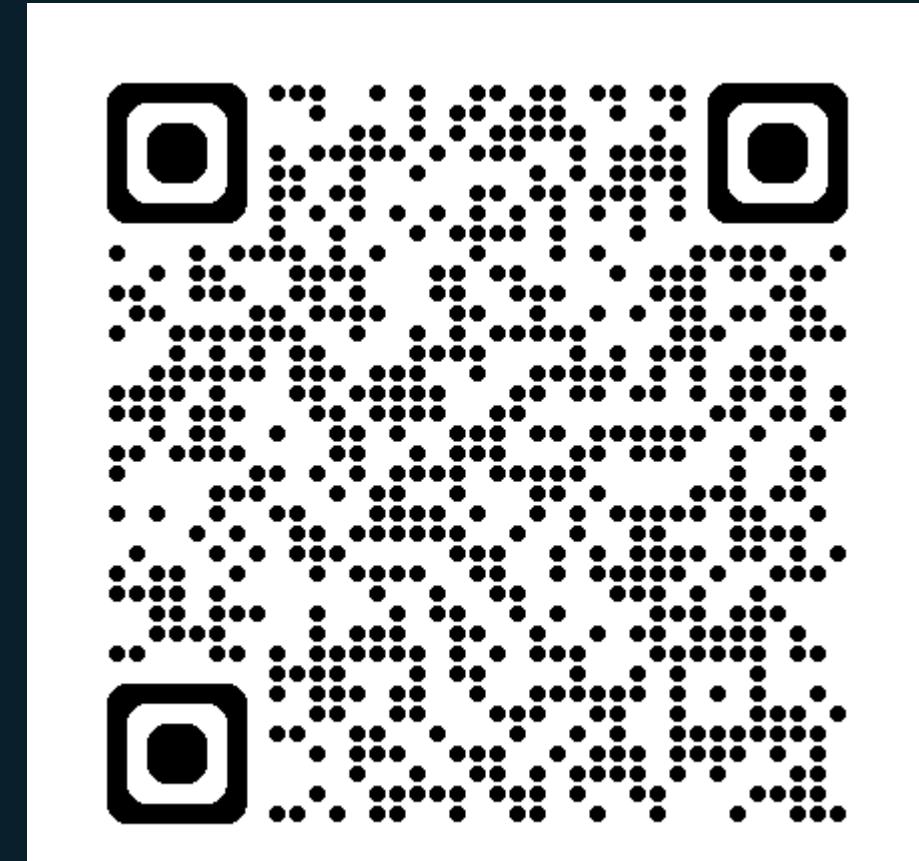
Product

Data asset quality rule score

Customer

Microsoft Purview

- Data Quality
- <https://tinyurl.com/yckdxhyh>



Data Estate Health – Health Controls

Health controls

Health actions

Metadata quality

Reports



Health controls

Controls help you analyze and track your data estate health.

Active controls
16

Not healthy
2

Fair
0

Healthy
11

Expired
0

Business domain: All

Owner: All

Status: All

26 items



80

Filter by keyword

Control	Current score	Vs. Target	Last refreshed
> Access and Use	Not healthy 0 %	-80	5/31/2024, 2:44 AM
> Discoverability	Healthy 100 %	+20	5/31/2024, 2:34 AM
> Estate Curation	Healthy 100 %	+20	5/31/2024, 2:34 AM
> Health Observability	Coming soon	--	--
> Value Creation	Coming soon	--	--
> Trusted Data	Fair 60 %	-20	5/31/2024, 2:34 AM
> Metadata Quality Management	Healthy 100 %	+20	5/31/2024, 2:44 AM
> Data Quality Management	Healthy 94 %	+14	6/11/2024, 2:15 AM

Data Estate Health – Health Controls

Health actions

C

Optimize your data health with these actions. [Learn more](#)

[Active](#) [In progress](#) [Resolved](#) [My items](#)

Total action items
9

High severity
5

Medium severity
3

Low severity
1

Assigned to: All

Business domain: All

Active since: Last 30 days X

Add filter

3 items

Filter by keyword

Group by Finding type ▾

Finding type ↑	Total findings	Severity breakdown	Finding subtype	Finding name
Access and Use	1	<div style="width: 100%; background-color: #f0a000; height: 10px;"></div>	Compliant data use	Missing terms of use on data products
Data Quality Management	6	<div style="width: 70%; background-color: #d93a3a; height: 10px;"></div> <div style="width: 30%; background-color: #ffff00; height: 10px;"></div>	Data asset quality rule score, +2 more	Data asset quality rule score has failed
Trusted Data	2	<div style="width: 100%; background-color: #f0a000; height: 10px;"></div>	Data quality enablement	Missing data quality scores on data assets

Data Estate Health – Metadata Quality

Metadata quality

Measure metadata quality by defining rules within controls

[Configure severity](#)

Controls

Search

▼ Access and Use 2

Compliant data use 1

Self-serve access enablement 1

Access request SLA

Zero-trust default access

➢ Discoverability 5

➢ Estate Curation 3

➢ Health Observability

➢ Value Creation 1

➢ Trusted Data 5

➢ Metadata Quality Management 5

Last refreshed on 05/07/2024, 11:01 PM

[Reset](#)

[Save changes](#)

Rules

The score is the average of all the rules

[+ Add rule](#)

[Combine rules](#) ▾

[Delete](#)

[Data product](#)

[Published terms of use count](#)

[Greater than](#)

0

ⓘ Note

As the metadata quality rules run, all pass check contribute to [health control scores](#), and all failed checks contribute to [health control actions](#).

Data Estate Health - Reports

Reports



Total reports
7

Active reports
7

Draft reports
0

Showing 7 of 7

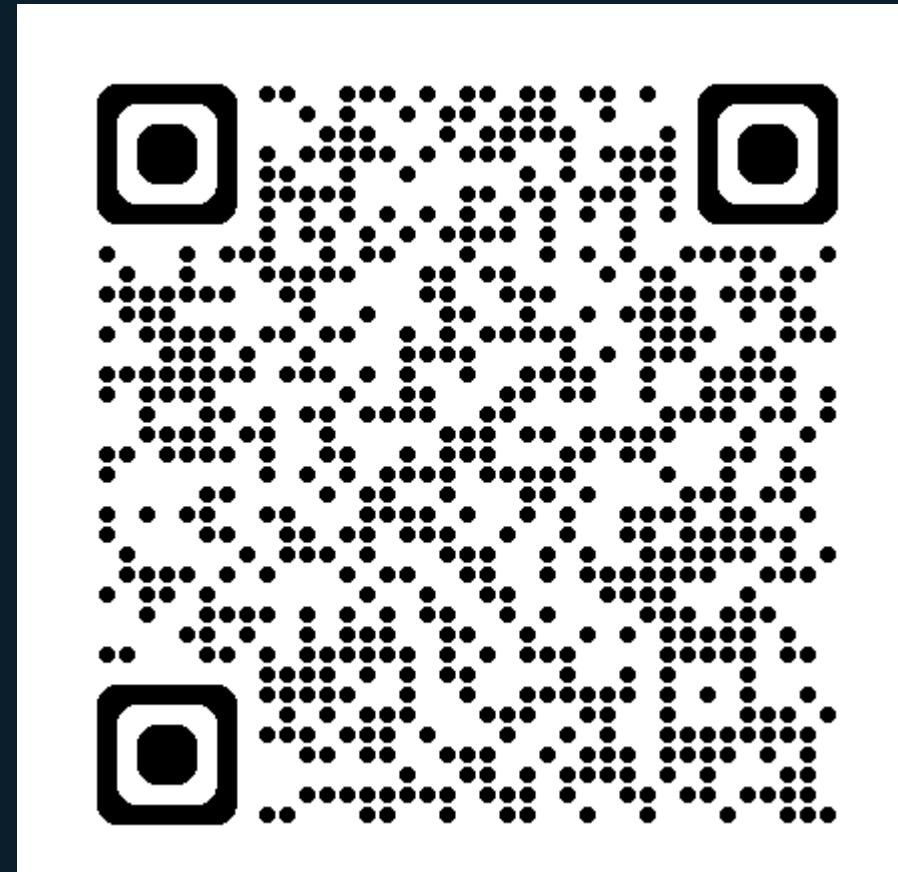
Filter by keyword

Report name ↑	Report type	Last updated	Status
Assets	System	6/11/2024, 10:45 AM	✓ Active
Catalog adoption	System	6/11/2024, 10:45 AM	✓ Active
Classifications	System	6/11/2024, 10:45 AM	✓ Active
Data governance	System	6/10/2024, 8:39 PM	✓ Active
Data stewardship	System	6/11/2024, 10:45 AM	✓ Active
Glossary	System	6/11/2024, 10:45 AM	✓ Active
Sensitivity labels	System	6/11/2024, 10:45 AM	✓ Active



Microsoft Purview

- Data Estate Health
- <https://tinyurl.com/yc2v6mk8>



Data Access / Data Policies

Request access to "Covid-19 Vaccination and Case Trending by Age"

Complete this form to request access to the data product.

User or system that needs access *

Sarah DG

Manager Approval *

Required

Purpose *

Business justification *

uptake monitoring

Terms of use *

I understand and acknowledge that data copies are not permitted

Send Cancel

Manage access policies

Policy Configuration Inherited Policies

Configure and manage access policies and a request form to control access to the data product.
[Learn more about access policies](#)

Permitted access

Data usage purpose

Authorized purposes for accessing and using the data product.

Name	Description
Assessing Fit	Use to determine if data is useful
Critical Reporting	Use to create high quality reports
Directional Insights	Use to gather directional insights

+ Add purpose

Access time limit *

Maximum time allowed for data access.

1 Years

Approval requirements

- Manager approval required
- Privacy and compliance review required

Approvers *

Approvers must manually grant access to the physical data assets in the data product.

Wolfgang Strasser

Digital Attestations

- Permit data copies
- Require acknowledgement of terms of use

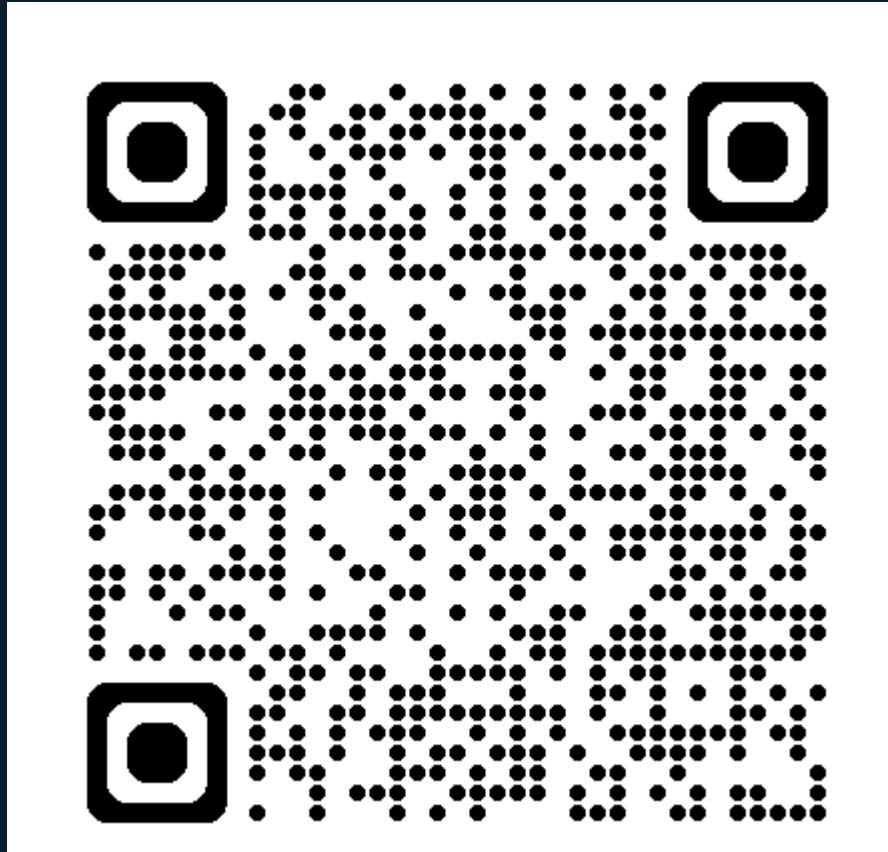
+ Add attestation

Preview request form



Democratize Data to Your Organization

<https://tinyurl.com/2t5kfxsd>



Purview and YOU

Microsoft Purview PREVIEW

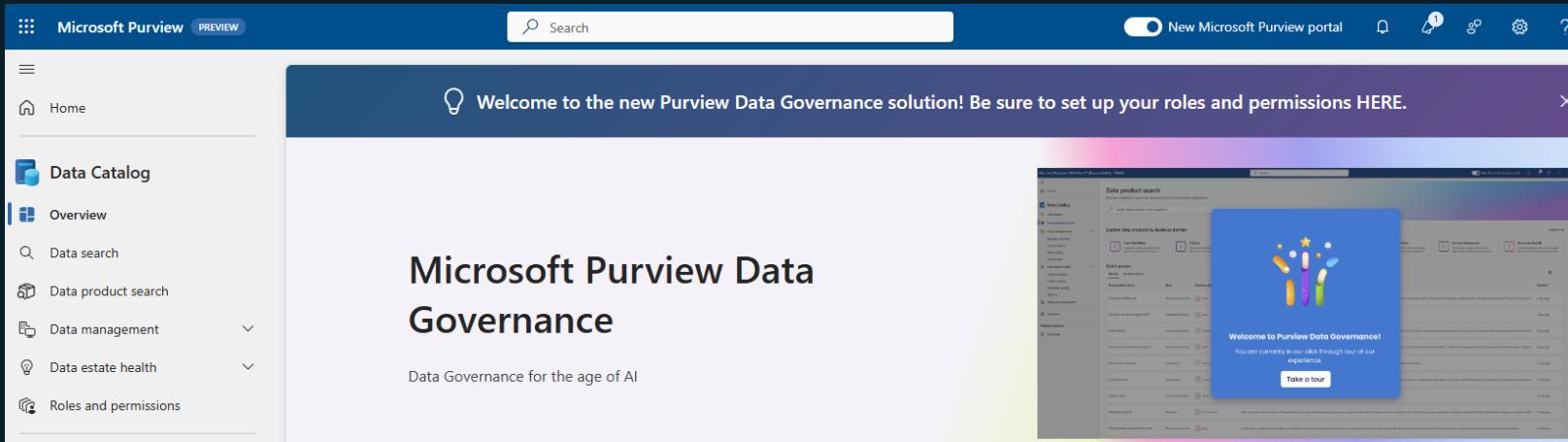
Search

New Microsoft Purview portal

Welcome to the new Purview Data Governance solution! Be sure to set up your roles and permissions [HERE](#).

Microsoft Purview Data Governance

Data Governance for the age of AI



This content is powered by a third party service.

Model your data estate for your business



Model your growing multi-cloud data estate using your durable business taxonomy, logical data concepts, and natural language.

Federate your Data Governance



Scale the management of your data estate by empowering your business domain teams to manage the quality and health of their data.

Empower your users to innovate with your data



Responsibly democratize your data by making your data estate easy to discover and safe to access by your people and your AI.

Getting started tutorial

Modeling your data estate

① Set up business domains
Create glossary terms, define CDEs, and establish OKRs.

Federating your data governance

② Set up and register your data
Build out data map, register and scan data.

purview.microsoft.com

- <https://purview.microsoft.com>
- Use your <ownUsername>
- If asked for more security options => „Ask later“
- Choose the new, not the classic portal
- Make yourself familiar with the solutions available
 - Data Map
 - Data Catalog

Data Map

≡

Home

Data Map

Domains

Metamodel

Source management

Annotation management

Solutions

Related solutions

Data Catalog

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

Collection	Actions
Sales	3 collections
Fabric	...

Lakehouses

Semantic Models

+ New collection >Edit Delete Refresh Request role

Updated on June 4, 2024, 10:29 AM by Erwin de Kreuk(MVP)

Description

No description for this collection

Collection path

```
Sales
  |
  Fabric
```

Assets: 374

Data sources: -

Scans: -

Data Catalog

- Search for „Sales“ Assets in the data catalog
- Open the details for „Power BI Wide World Importers“ Dataset
- Use the Purview UI to find out more about this data asset
 - Which properties are collected by the Purview scan?
 - To which Workspace does this data asset belong to?
 - How many items does the schema list of this asset contain?
 - Are there any experts / data owners set?

Business Domains

- Data Catalog -> Data Management -> Business Domains
- Use the „Adventureworks“ Business Domain for the next exploration steps
 - How many data products are created in this business domain?
 - How many items are assigned to each of the data products?
 - What is the status of these data products?

Data Product / Data Quality

- Exploration of the „Products“ Data Product
 - Which kind of data assets does this data product contain (source types, ..)
 - How do you rate the Data quality score of this data product?
 - For which data assets a data quality score has been calculated?
 - Check the data quality rules defined for the asset Product
 - Can you find out more about the reason for the low data quality score?
 - Does Purview provide us with some data quality actions? If yes, which ones?
 - For how many columns of this data asset are DQ-rules defined?

Purview Deep(er) Dive

What is going on behind the (Purview) scenes?

- Collections in your Purview Account
- Register Sources
- Define Scans, Scan Rule Sets
- Classifications
- Business Glossary
- Data Catalog
- Business Domains
- Data Products
- Data Quality
-

What is going on behind the (Purview) scenes?

- Collections in your Purview Account
 - <https://learn.microsoft.com/en-us/purview/how-to-create-and-manage-domains-collections>
- Register Sources
 - <https://learn.microsoft.com/en-us/purview/microsoft-purview-connector-overview>
- Define Scans, Scan Rule Sets
 - <https://learn.microsoft.com/en-us/purview/concept-scans-and-ingestion>
- Classifications
 - <https://learn.microsoft.com/en-us/purview/concept-classification>

What is going on behind the (Purview) scenes?

- Data Catalog
 - <https://learn.microsoft.com/en-us/purview/what-is-data-catalog>
- Business Domains
 - <https://learn.microsoft.com/en-us/purview/concept-business-domain>
- Data Products
 - <https://learn.microsoft.com/en-us/purview/concept-data-products>
- Data Quality
 - <https://learn.microsoft.com/en-us/purview/data-quality-overview>
-

Microsoft Purview MDM Integration

- [Home](#)

- Data Catalog**
- [Data search](#)
- [Data product search](#)
- [Data management](#)
- [Business domains](#)
- Data products**
- [Data quality](#)
- [Data activity](#)
- [Data source health](#)
- [Rules and permissions](#)

- Solutions**

- Related solutions**
- [Data Map](#)

Customer Master List

(1 item)

Data product 1 Master data and reference data Customer 3 data assets

Last updated: 10 days ago | Manage policies

Description

One master master customer list that all finance, audits and accounting activities. In addition, sales team will be using this list for CRM activities, and financial team will be using this list to calculate incentive for the sales team and customers. The data for this dataset is compiled from many sources that our sales teams can including, Salesforce and SAP. The kind of this data product is to allow the reuse of and the mastered patient records of customers. This certified data has been gathered to support everyone.

Data types

This data product will be used for accounting, forecasting, auditing, reporting, corrective comp, allocation, CRM activities, sales reporting. One master list of Microsoft customers. All data analysts, data scientists and app developers should use this data product.

The analytical master is the **CustomerID** value in the dimension table. When using this data if issues or discrepancies are discovered the source tables are included to resolve research and troubleshooting. When looking for income specific data you can join this master table with the original table to build conditions for income aligned data.

Master data

1 Sales

Table

Published to catalog

Data product views

3 owners

Active subscribers

1

Last updated

Weekly

Last activity

Master (30d)

Issues

0

New data sets

None (0)

Documentation

Read documentation

Data assets (8)

View all data assets



Customer

1 row(s) 1 col(s)

List of customers as coming from leads management tool.



Customer Address SAP

1 row(s) 1 col(s) 1 row(s) 1 col(s)

Based on the converted addresses for customers in SAP.



Customer address raw

1 row(s) 1 col(s) 1 row(s) 1 col(s)

Same data as the SAP version but in raw format.



Customer2B Golden asset

1 row(s) 1 col(s)

This data asset is a master data asset connecting profiles by using data from...



Customer profile SAP

Rows that have unique and necessary data for customer profile to be provided to the business compensation system.



Customer profile SAP

The data classified across customer tables of SAP to build a single customer profile.