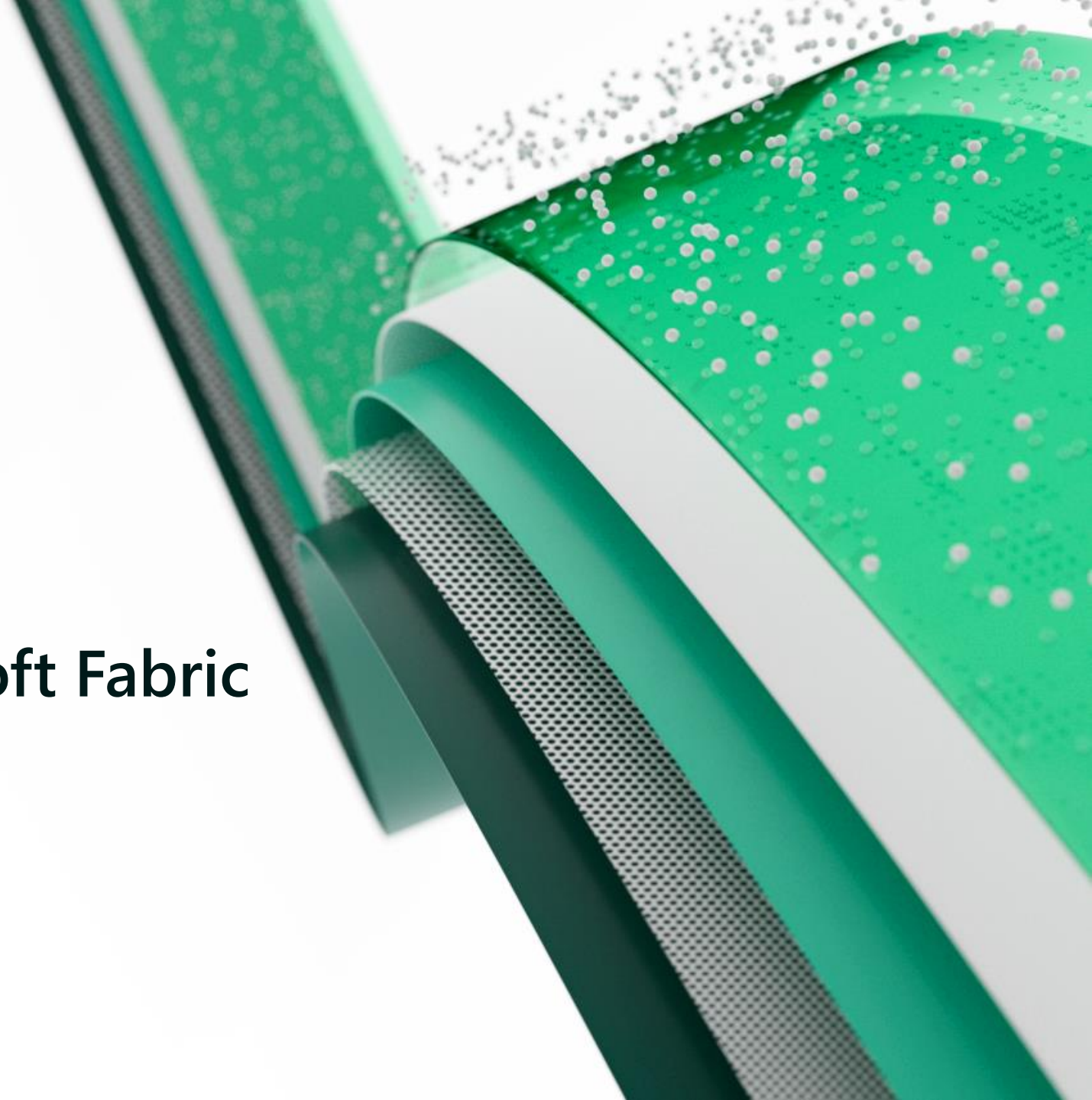
Microsoft | Microsoft Fabric



A Guided Tour of Microsoft Fabric

Benni De Jagere
Senior Program Manager
Fabric Customer Advisory Team

Slides

DATA:Scotland

Scotland's Data Community Conference

datascotland.org
@datascotland

TIMEXTENDER

ADVANCING ANALYTICS

hippo data

quorum

Tabular Editor

DATAmasterminds

Waterstons
we're with you

Benni De Jagere

Senior Program Manager | Fabric Customer Advisory Team (FabricCAT)

Fabric CAT

.be Member

@BenniDeJagere

/bennidejagere

/bennidejagere

/bennidejagere

#SayNoToPieCharts

Introducing Microsoft Fabric

OneLake

Data Factory

Synapse
Data Engineering

Synapse
Data Warehouse

Synapse
Data Science

Synapse
Real-Time Analytics

Power BI

Lake-centric and open

Role-specific tools

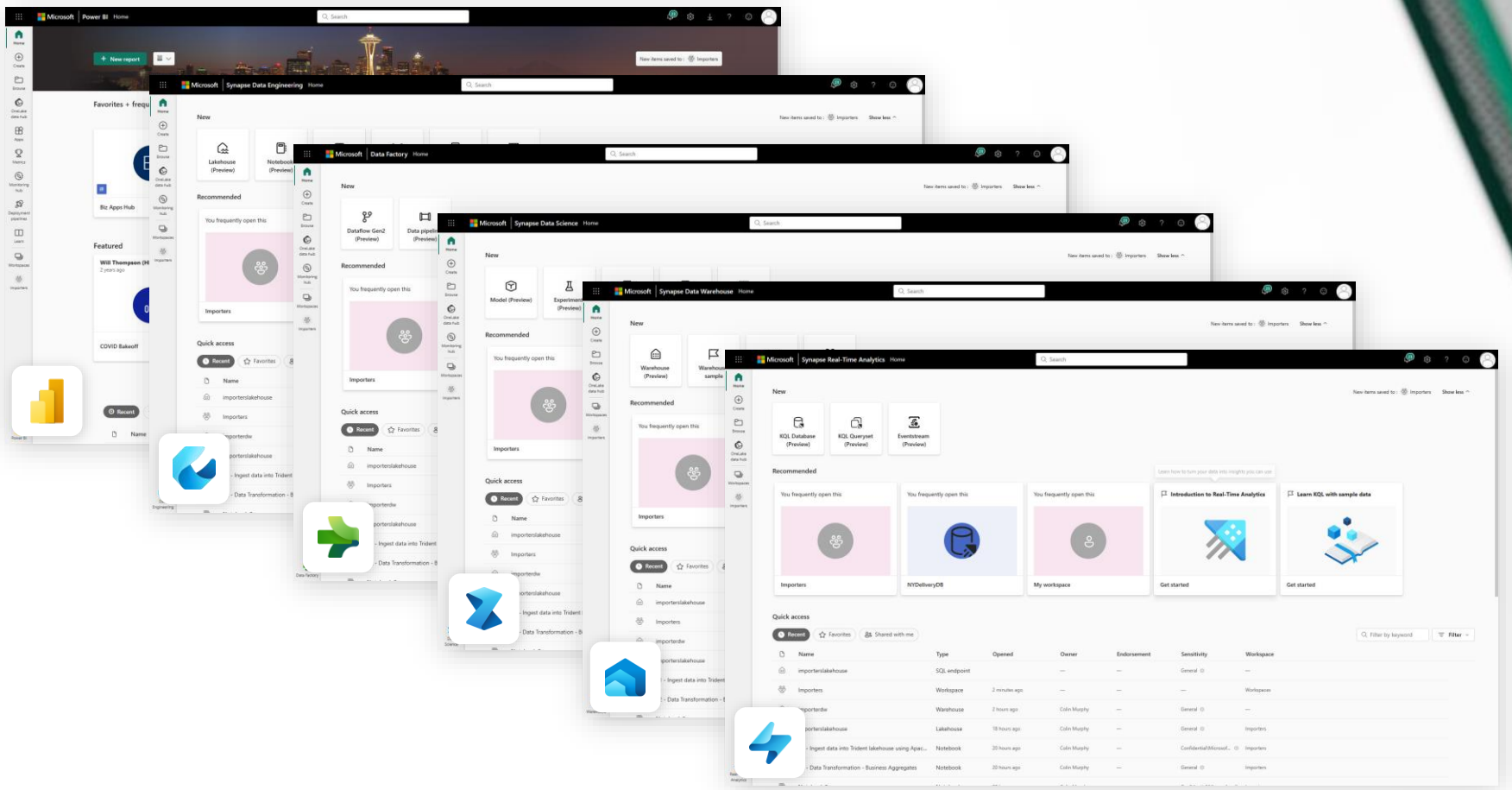
Empower every Microsoft 365 user

Persistent governance and security

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

Persona optimized experiences



Data integration experience

Bring together low-code, AI-enabled experiences, multi-cloud connectivity, and persistent data security and governance to help solve complex scenarios for all developers

- 200+ native data source connectors
- Cloud-scale data movement with Data Factory
- Low-code interface for ingesting data from hundreds of data sources using Dataflows Gen2
- Petabyte scale transformation
- Powerful, enterprise-grade Data Factory experience with the best of ADF and Power Query together

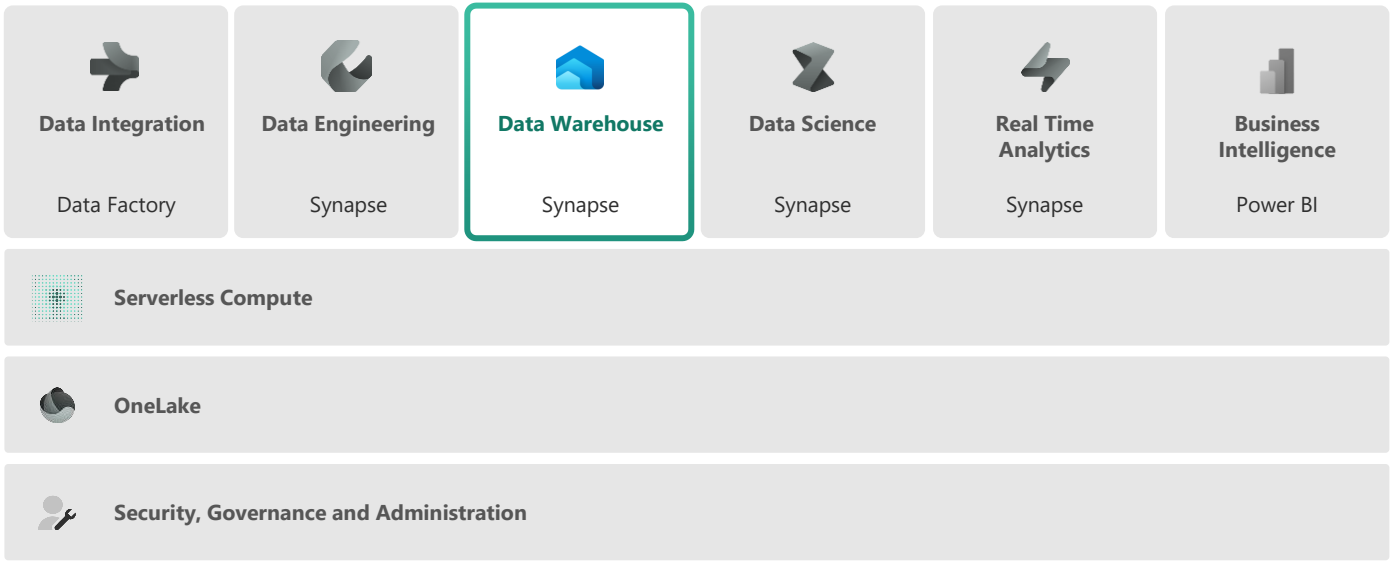
Data engineering experience

Build you data estate and empower data engineers with a world-class Spark platform, fully integrated with Data Factory, to transform and democratize data at scale

- World-class Spark serverless compute
- Pro and low-code authoring experience
- Schedule and orchestrate data transformations with notebooks and Spark jobs
- Launch clusters on demand and dynamically scale in, scale out, pause, and resume
- Perform code-free interactive data exploration and add to your data pipeline

Data warehouse experience

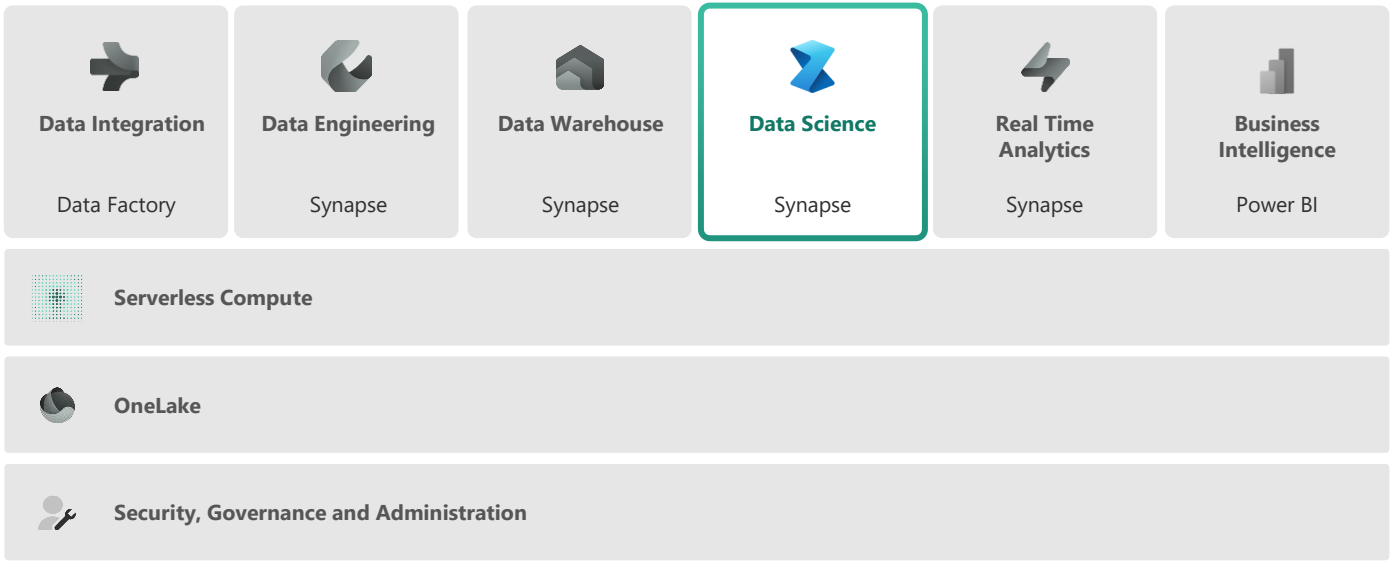
Achieve data platform goals with east and cost efficiency, while empowering your developers and engineers with accelerated reporting and insights



- Industry-leading SQL performance and scale, including reporting tools using T-SQL and TDS end-points
- Fully separate compute storage that can independently scale
- Build relational layers on top of physical data in lakehouse and natively store data in open Parquet/Delta Lake to remove need for data duplicity
- Consumer data with Power BI for reporting and visualization
- TLS 1.2 encryption protects all connections for granular security across your data platform

Data science experience

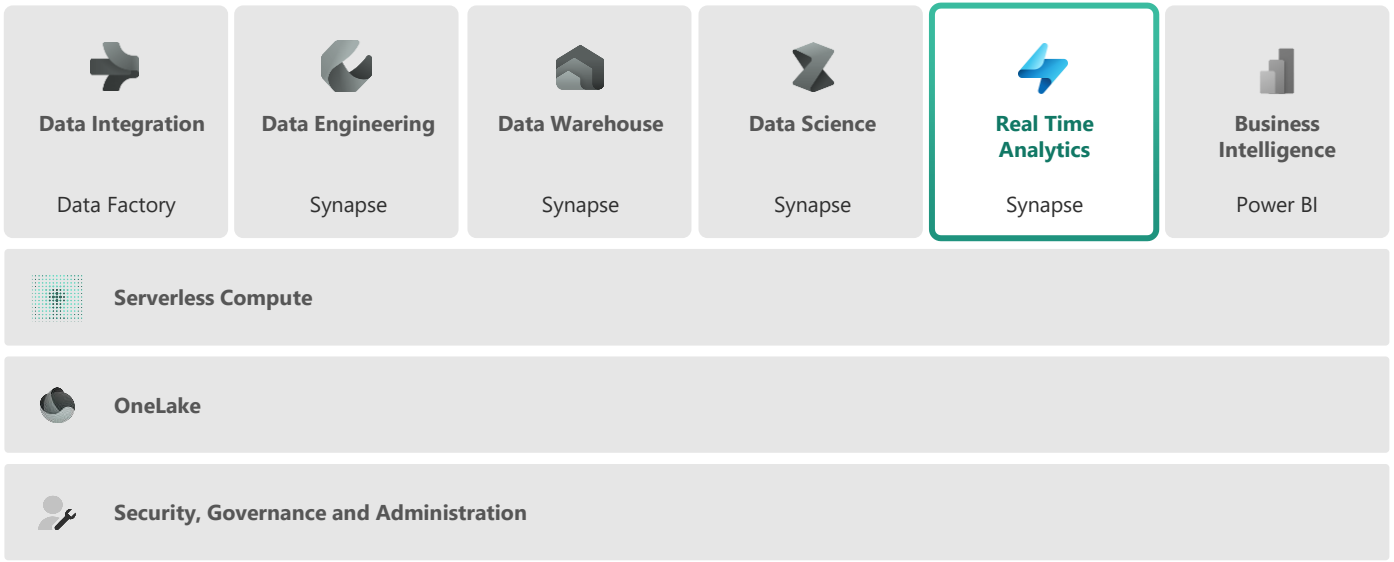
Build, deploy, and operationalize machine learning models with boundaryless collaboration and seamless integration with Azure ML



- Access data from multiple sources and store data and insights in lakehouse(s)
- Leverage data science capabilities for model prediction at scale to gain and share business insights
- Iterate, build, and track Machine Learning experiments using ML flow
- Perform transformation, exploration, and featurization by leveraging built-in experiences
- Collaborate with others via Notebook, Power BI, and lakehouses in real-time

Real-time analytics experience

Unlock value and turn insights into actions with real-time analysis across telemetry data to better predict, optimize, and improve data applications



- High velocity, low latency data analysis capable of indexing diverse data formats and structures up to several petabytes large
- Leverage Kusto Query Language (KQL) to explore data and discover patterns, identify anomalies and outliers, and create statistical modeling
- Easy ingestion of data from any source, in any data format, like applications, websites, and IoT devices
- Democratize data responsibly with Microsoft Purview and use observability with Data Activator to activate timely actions

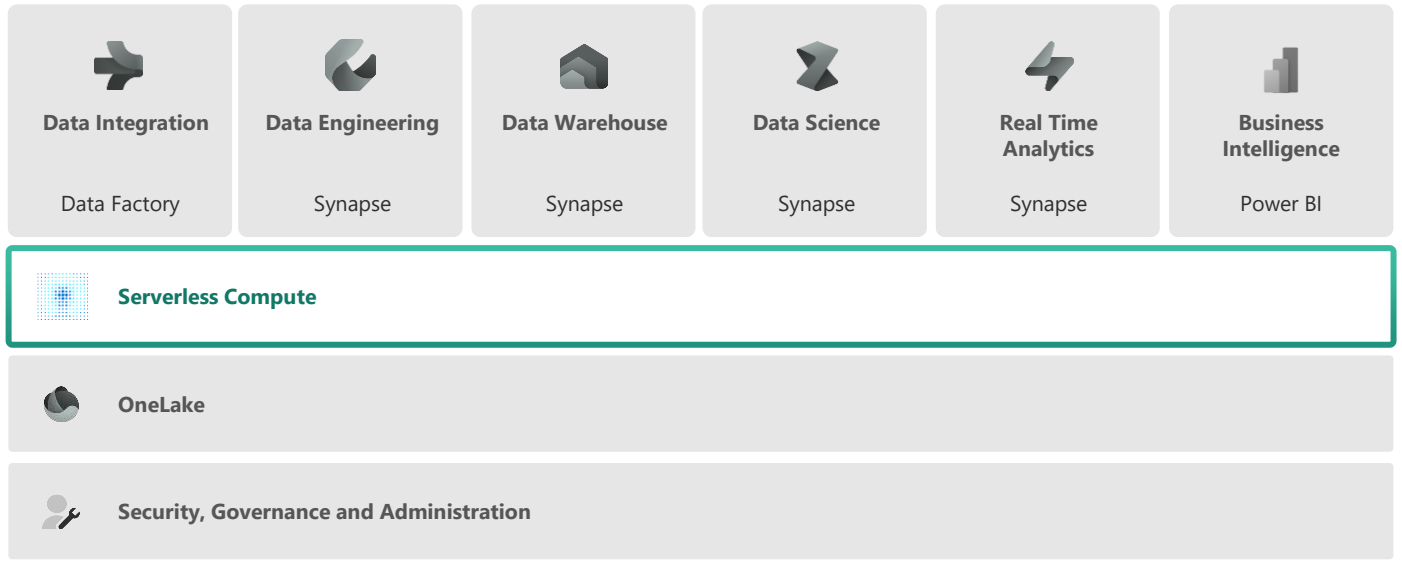
Business intelligence experience

Uncover powerful insights with intelligence visuals, leverage data quickly and intuitively, and help achieve faster and better data-based decisions with the industry leading Power BI platform

- Native Integration with Microsoft 365
- Powerful built-in AI capabilities and visuals illuminate hidden patterns, opportunities and anomalies in data with the click of a button
- Connect to, index, and certify datasets in the Power BI data hub
- Build governed databases, like data models or data marts, in a trusted and secure hub

Serverless compute

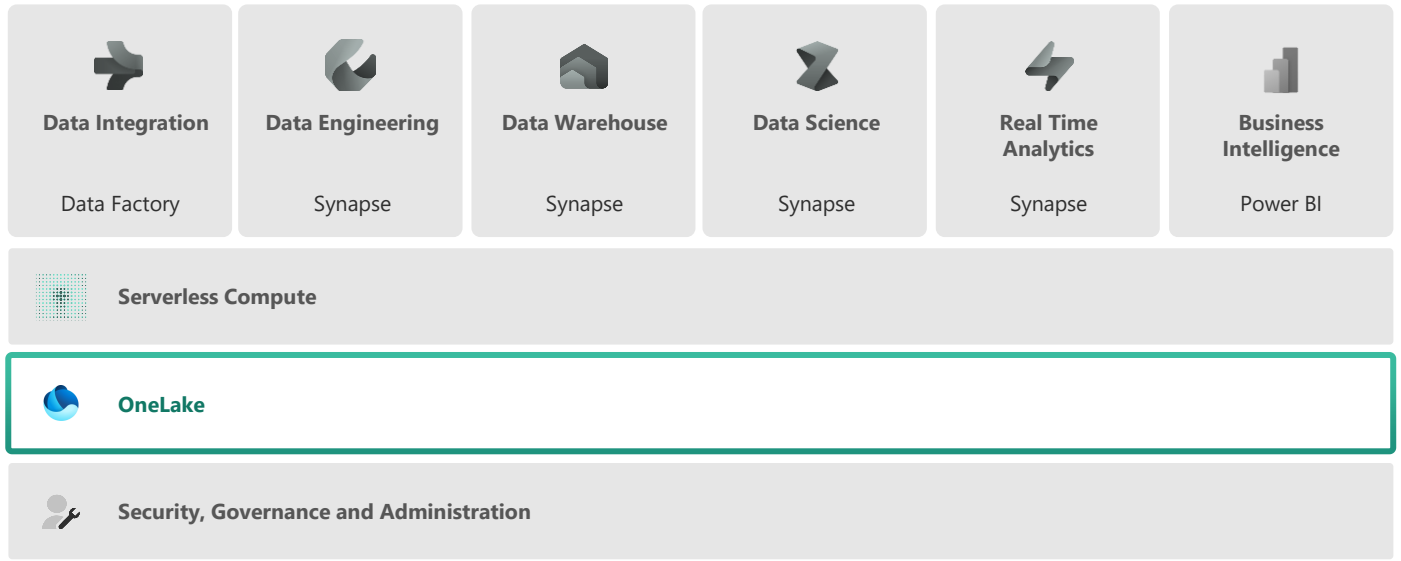
Run workloads without having to provision compute. Live compute pools are already provisioned with a workspace resulting in instant productivity.



- Limitless scale for the most demanding jobs
- Automatically create with each workspace
- With support for high concurrency mode there's no need to start a new session to run a notebook
- Memory optimized

Unified data foundation with OneLake

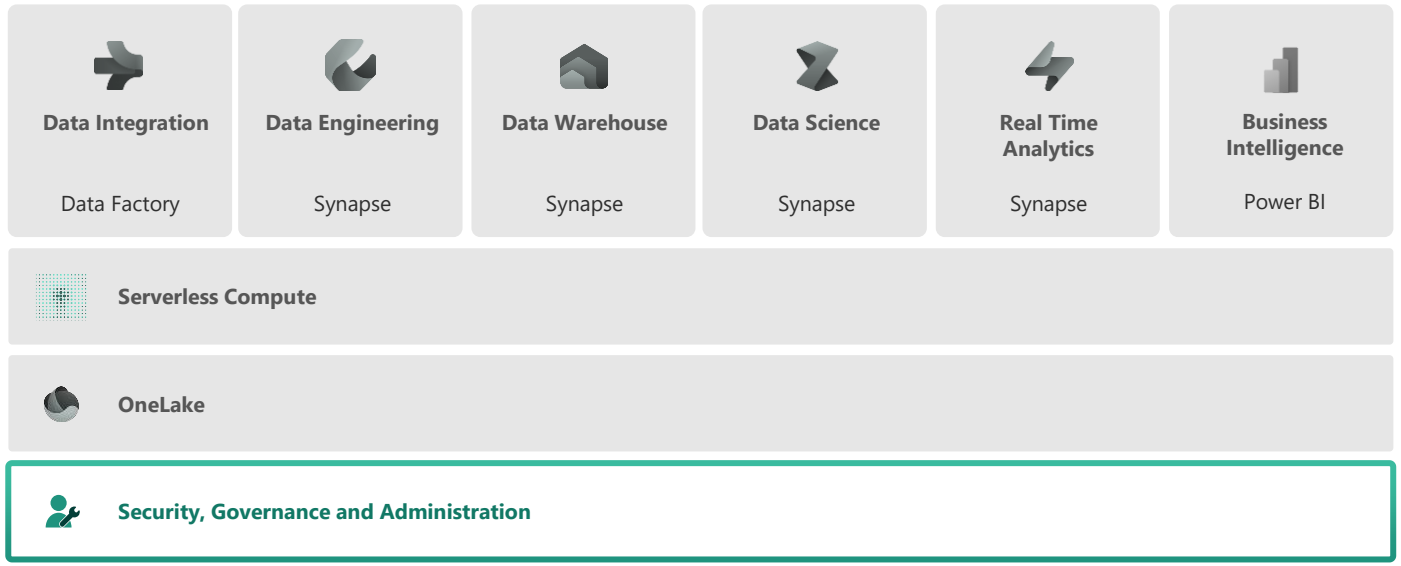
Eliminate pervasive and chaotic data siloes to provide a unified, secure, and centralized storage system for developers with OneLake—the "OneDrive" for data



- A single SaaS lake for the whole organization
- Abstraction layer over Azure Storage and third-party storage services, provisioned automatically with the tenant
- All workloads automatically store their data in OneLake workspace folders
- All data is organized in an intuitive hierarchical namespace
- Data in OneLake is automatically indexed for discovery, sharing, governance, and compliance

Persistent data governance and security

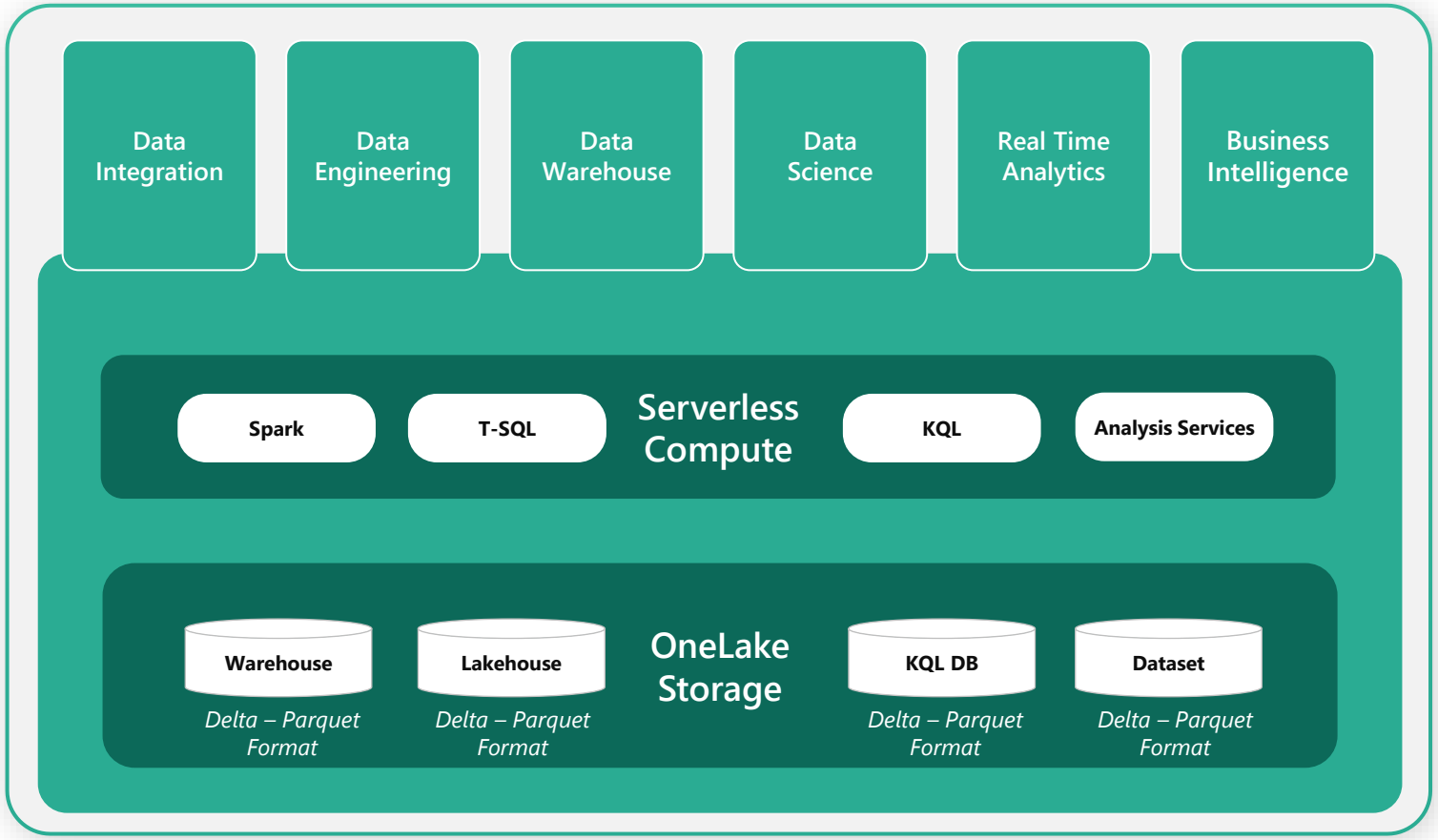
Built-in security and reliability features secure your data, and a set of governance and compliance capabilities help you manage, protect, and monitor sensitive information



- Centralized Administration via the Admin Center
- Governance powered natively by Microsoft Purview
- Insights via Admin Monitoring, Capacity Insights and more, enabling effective administration and governance
- Security at Data, Workspace and Network layers
- Key enterprise promises like data residency, auditability and much more

OneLake for all data

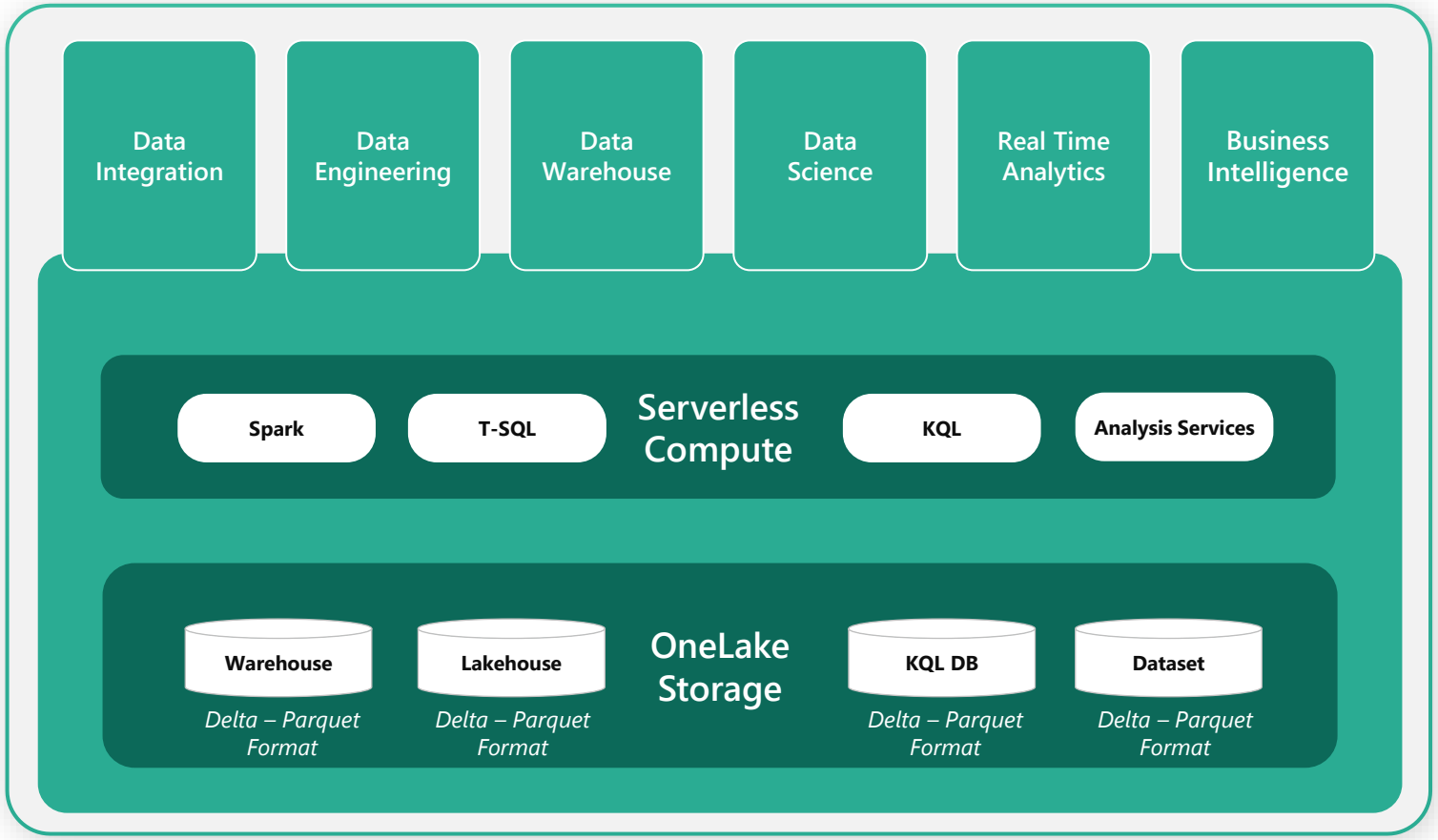
"The OneDrive for data"



- 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

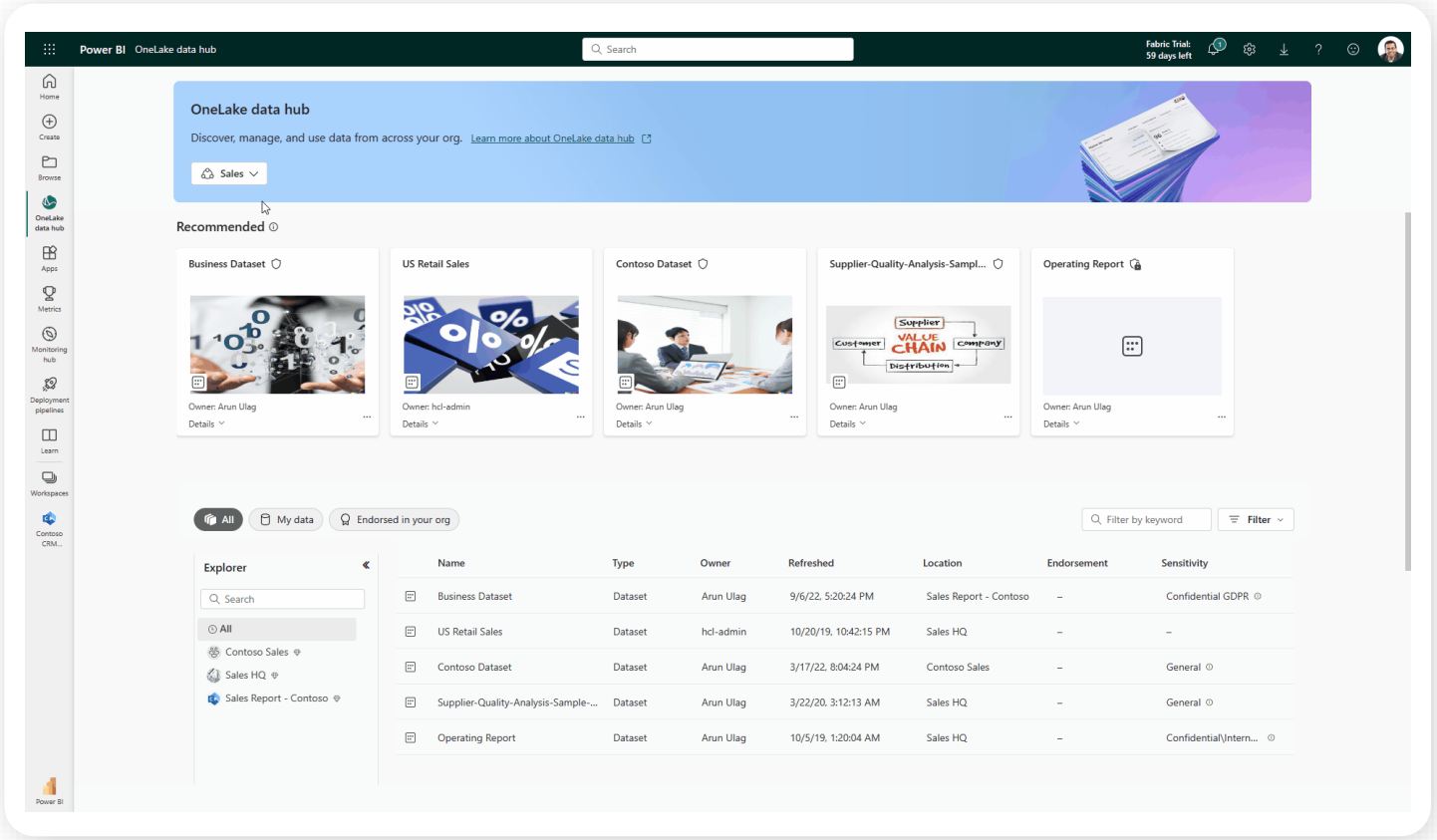
Real separation of compute and storage



- 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
- 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
- Shared universal security model is enforced across all engines

OneLake data hub

Discover, manage, and reuse data in one place



Central location within Fabric to discover, manage, and reuse data. Pervasive experience – available **everywhere** users discover data (create shortcuts, get data in dataflow, attach a notebook, create a dataset, and more).

Data can be easily discovered by its **domain** (e.g., Finance) so users can see what matters to them.

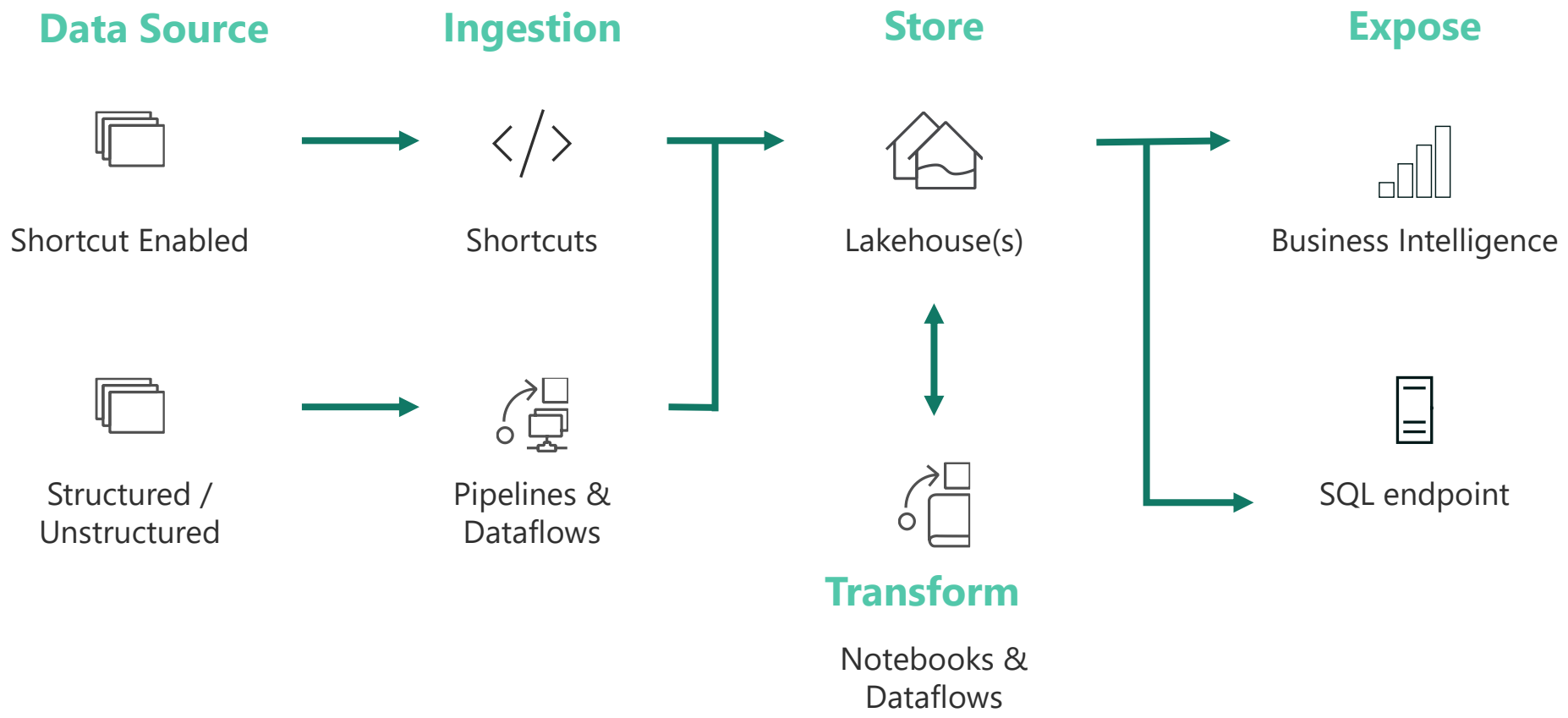
Efficient data discovery using search, filter and sort, or via browsing by folder (workspace) hierarchy.

Data **actions** such as preview, exploration, tracking lineage, report creation, analyze in Excel, and export can be easily done, even by **non-technical users**.

Available in **Microsoft Teams**, thus bringing OneLake Data hub to **Office,** enabling both technical and non-technical users with discovery, reuse and exploration.

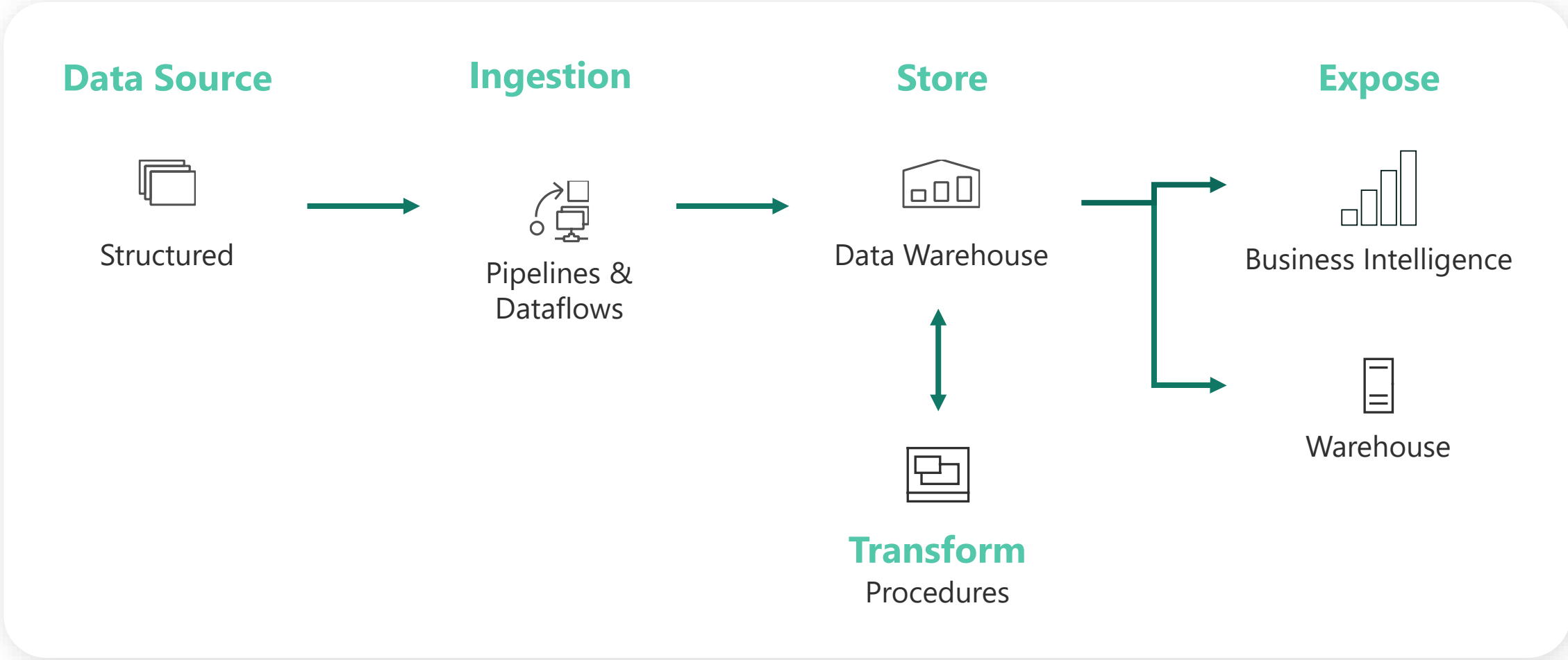
Lakehouse scenario

End-to-end analytics scenario

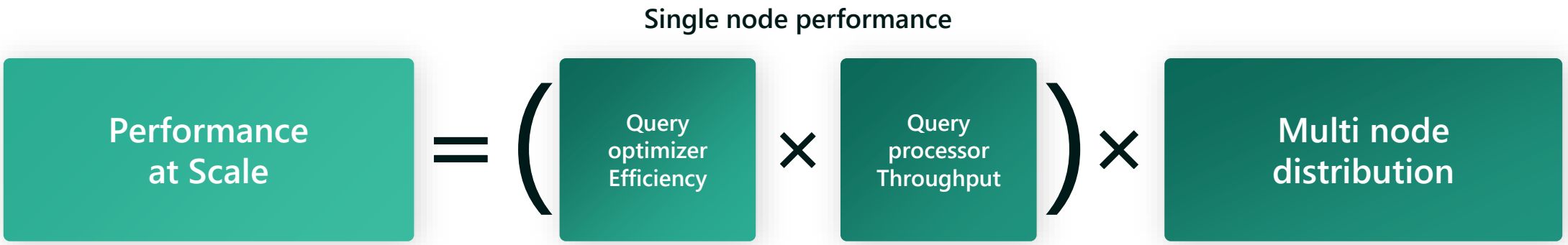


Data warehouse scenario

End-to-end analytics scenario



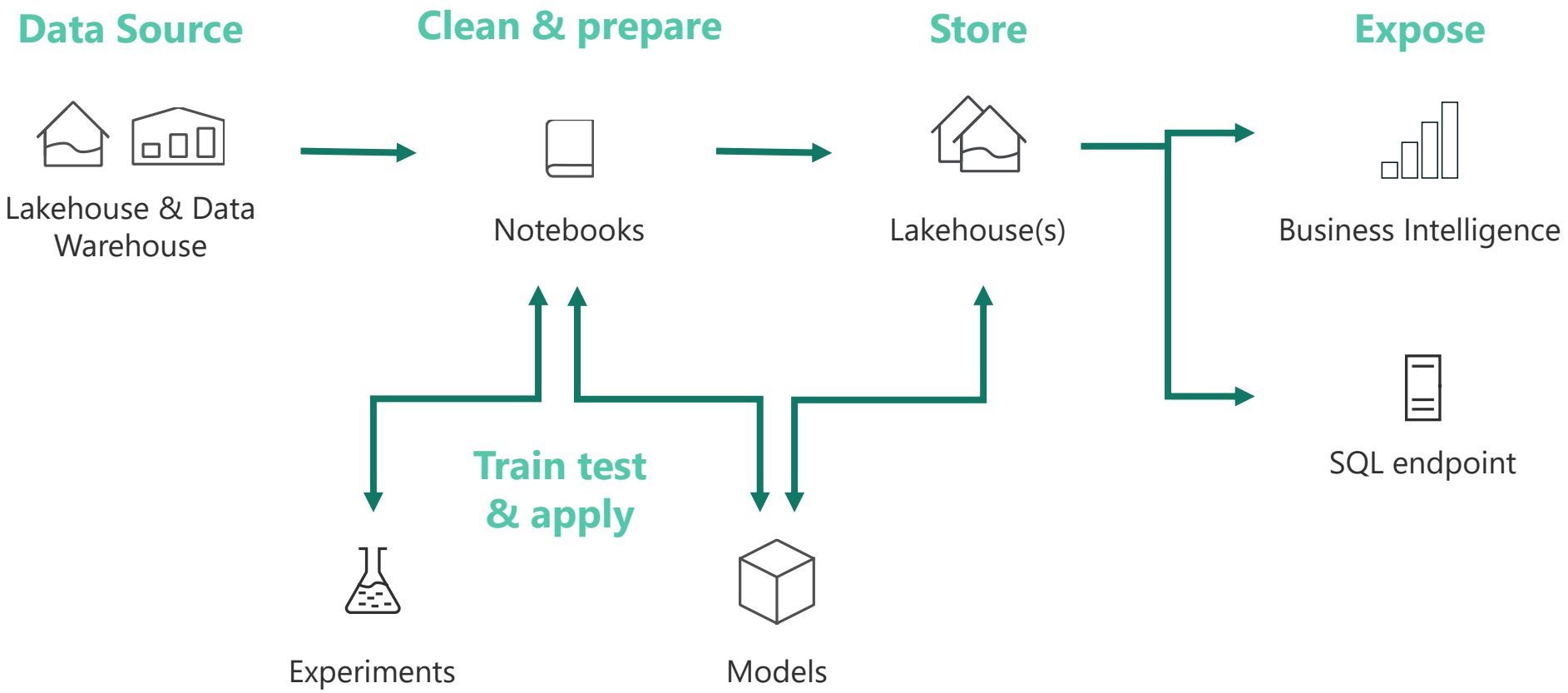
No Knobs Performance at scale with open data format



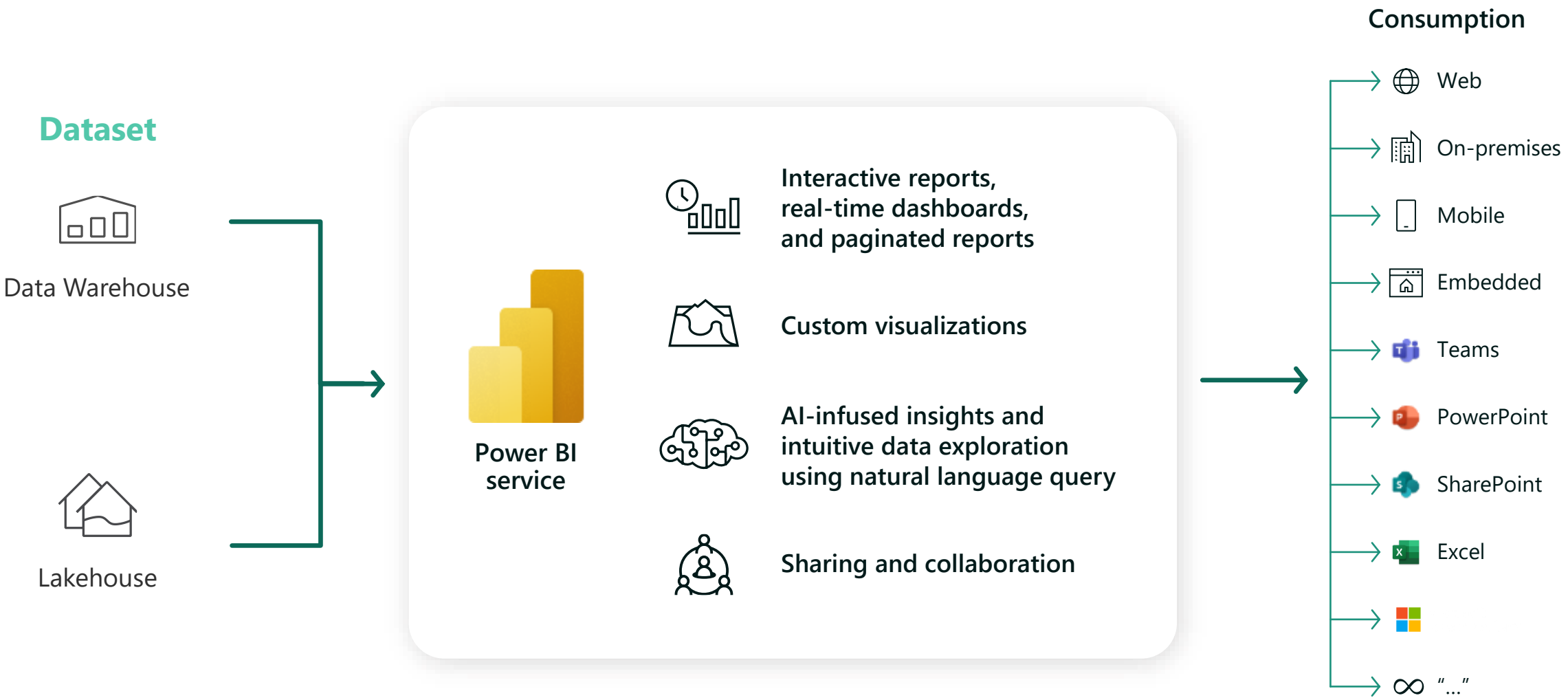
SQL Server QO	The best query optimizer in the industry
VertiPaq	The fastest columnar query processor in the world (originated from Power BI)
Polaris	The most scalable distributed query processor in the world with the only published **Petabyte** scale benchmark (VLDB 13, August 2020)

Data science scenario

End-to-end analytics scenario



Power BI: The bridge between data and decisions



Fully integrated Power BI semantic layer

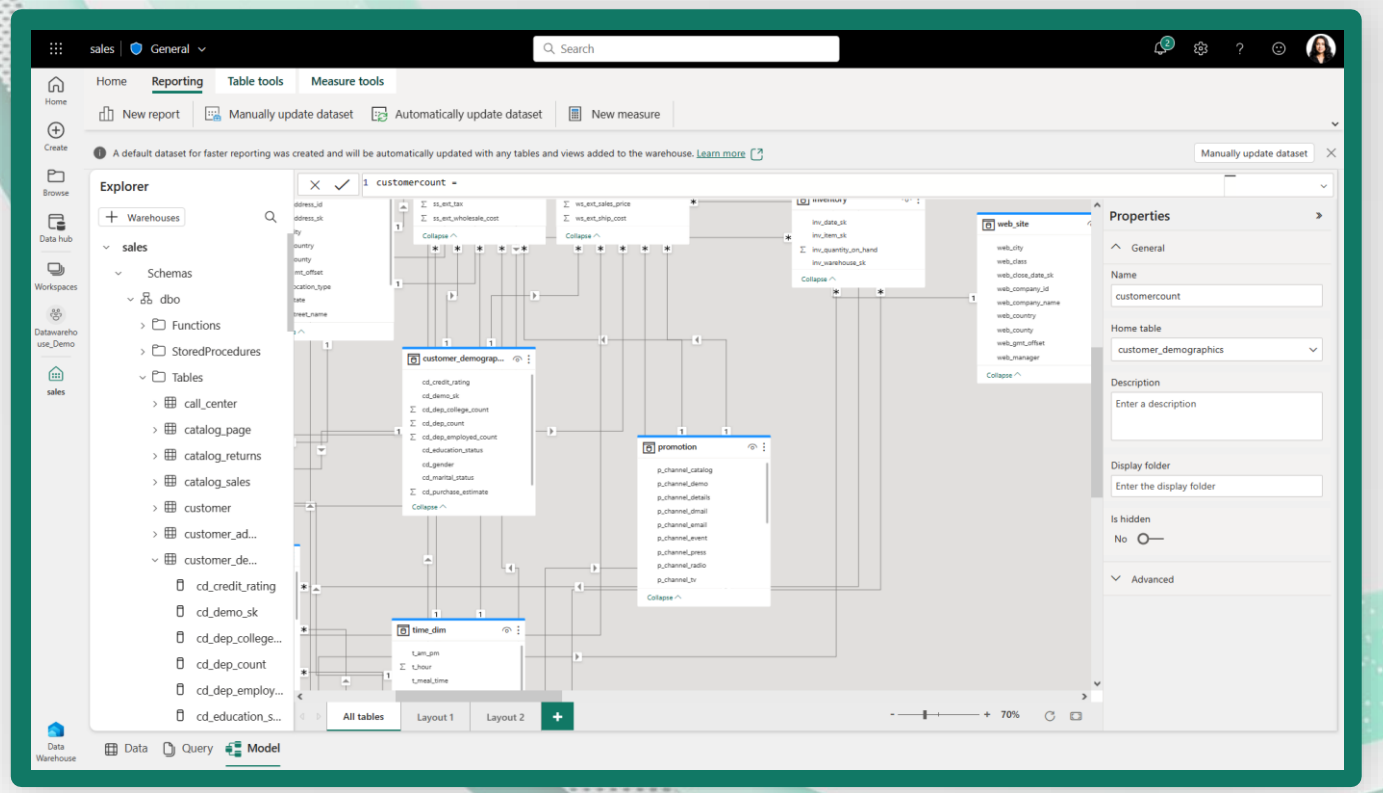
Reduce integration and gain insight from your data in seconds

Built-in Power BI enables everyone to visualize their data in seconds

---

Key Capabilities:

- Auto-generated dataset creating a Lakehouse/Warehouse
- Always in sync
- Default dataset in Direct Lake mode but automatically switches to Direct Query or Import mode as security or performance needs change
- Flexibility to add/remove tables to dataset
- Full web authoring experience for creating measures



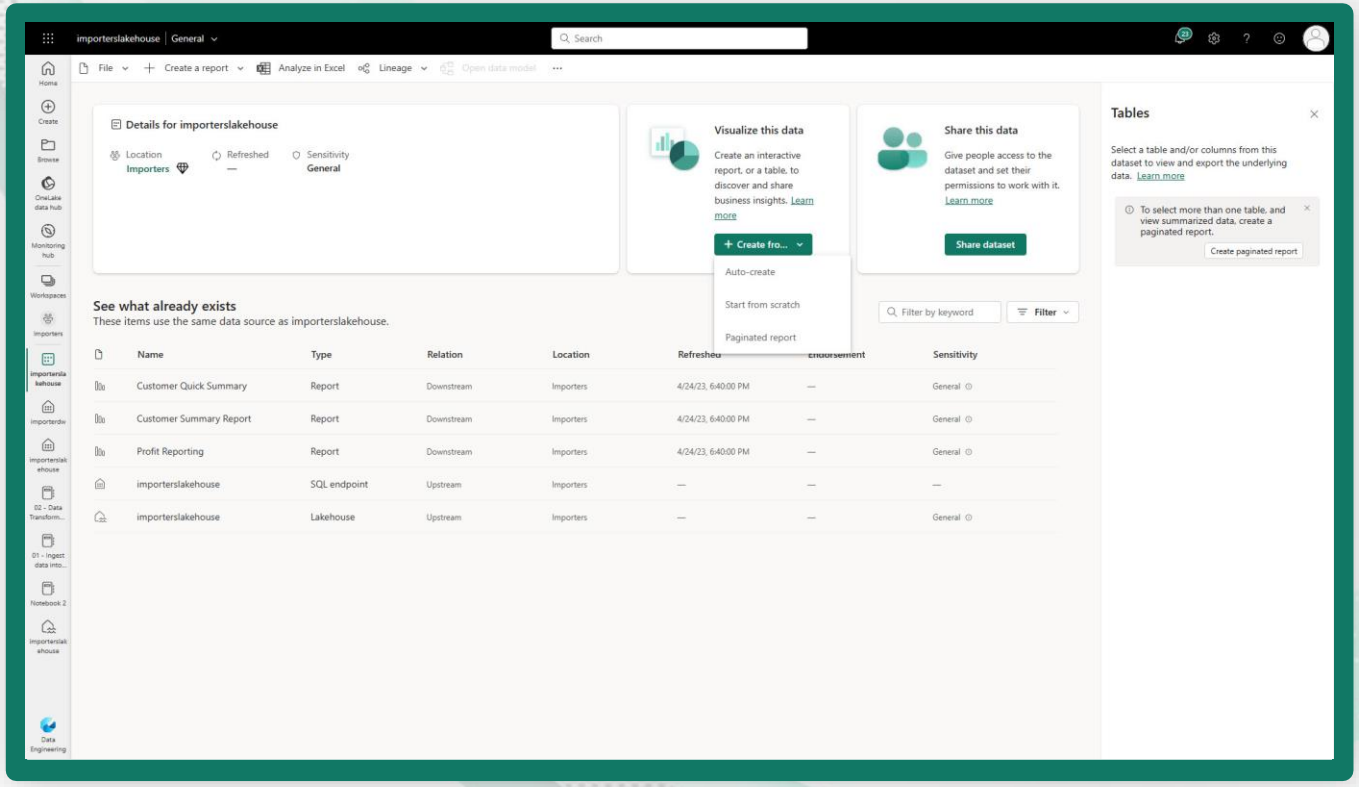
Power BI | Datasets

Auto create Power BI reports from your Lakehouse dataset

Blazing fast performance with Direct Lake Mode. Reduce time and complexity.

Key Capabilities:

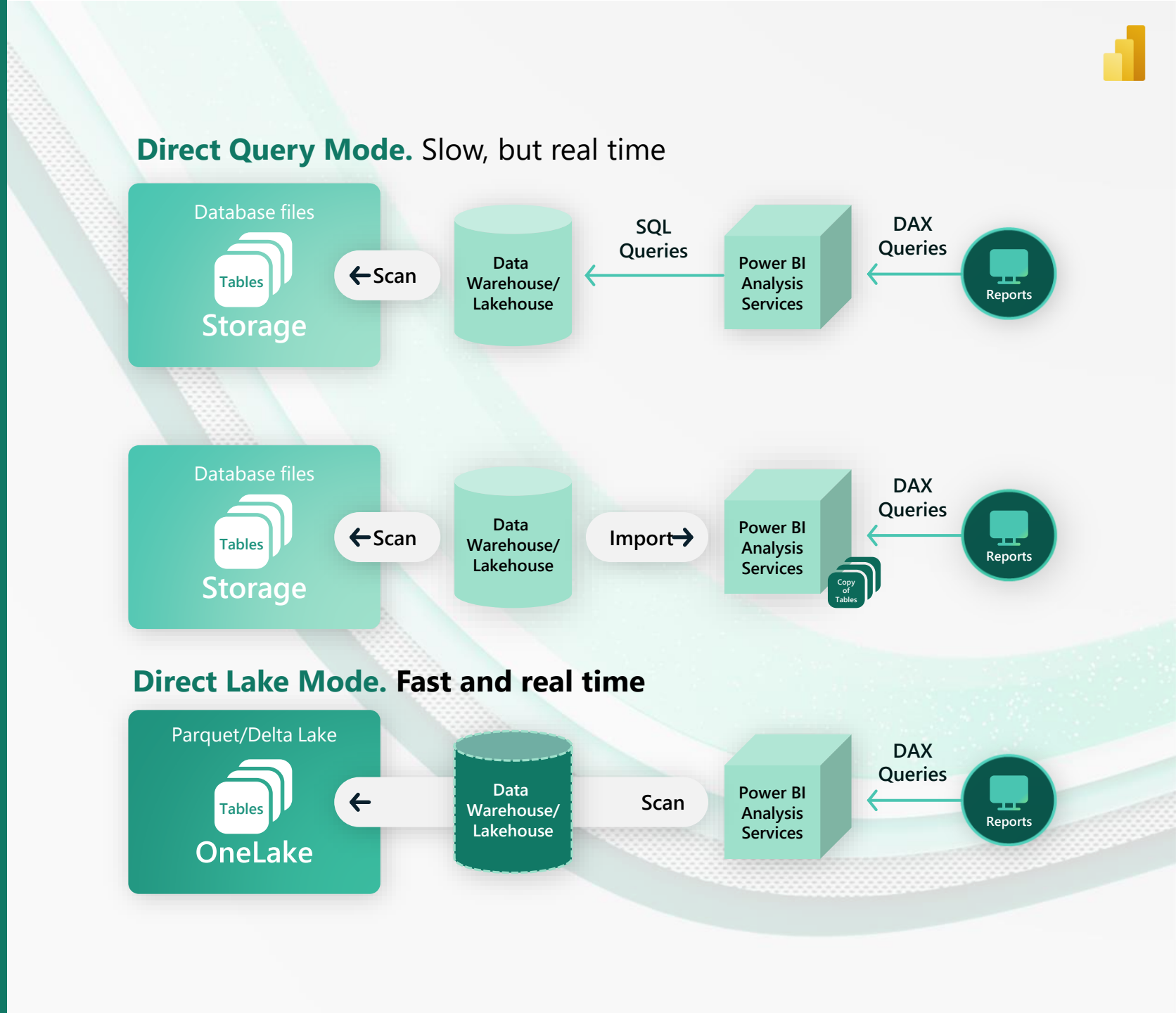
- Create a Power BI dataset with specific tables and views from the Lakehouse
- Build reports directly on top of a Warehouse
- Create a Power BI dataset with specific tables and views from the warehouse
- Create an interactive report, or table to discover and share business insights
- Share data and give people access to the dataset and set their permissions to work with it



Power BI | Direct Lake Mode

Direct Lake is a fast-path to load the data from the lake straight into the Power BI engine, ready for analysis

Direct Lake is based on loading parquet-formatted files directly from a data lake without having to query a Lakehouse endpoint, and without having to import or duplicate data into a Power BI dataset



Security, Governance, and Administration

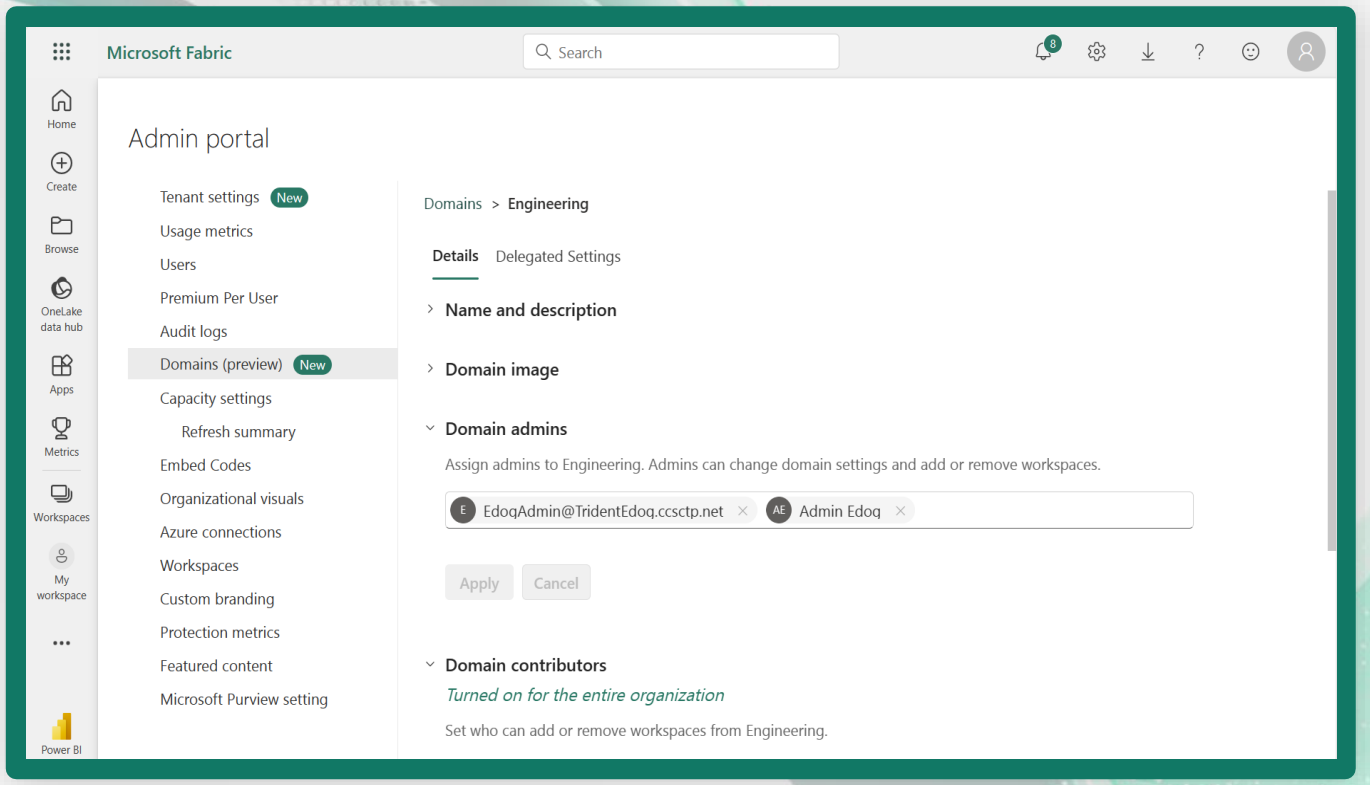
Fabric domains

A domain is a way of logically grouping together all the data in an organization that is relevant to a particular area or field

Commonly, this is used to group data by business departments. To group data into domains, workspaces are associated with domain and a domain attribute becomes part of their metadata. This supports management and governance of data according to specific department regulations, restrictions, and needs

Key Capabilities:

- Create a domain and associate workspaces with domains
- Assign domain roles and define domain admins/contributors



Workspace roles

Workspace roles let you manage who can do what in a Fabric workspace

These roles can be assigned to individuals or security groups, Microsoft 365 groups, and/or distribution lists

Key Capabilities:

- Workspace roles in Fabric extend the Power BI workspace roles by associating new capabilities such as data integration and data exploration with existing Power BI roles
- Define users or user groups as Admin, Member, Contributor, or Viewer

Microsoft Fabric workspace roles

Capability	Admin	Member	Contributor	Viewer
Update and delete the workspace.	✓			
Add or remove people, including other admins.	✓			
Add members or others with lower permissions.	✓	✓		
Allow others to reshare items.[1]	✓	✓		
View and read content of data pipelines, notebooks, Spark job definitions, ML models and experiments, and Event streams.	✓	✓	✓	✓
View and read content of KQL databases, KQL query-sets, and real-time dashboards.	✓	✓	✓	✓
Connect to SQL endpoints of Lakehouse and Data warehouse.	✓	✓	✓	✓
Read Lakehouse and Data warehouse data and shortcuts[2] through SQL endpoints.	✓	✓	✓	-[3]
Read Lakehouse and Data warehouse data and shortcuts[2] through OneLake APIs and Spark.	✓	✓	✓	-
Read Lakehouse data through Lakehouse explorer.	✓	✓	✓	-
Write or delete data pipelines, notebooks, Spark job definitions, ML models and experiments, and Event streams.	✓	✓	✓	-
Write or delete KQL query-sets, real-time dashboards, and schema and data of KQL databases, Lakehouses, data warehouses, and shortcuts.	✓	✓	✓	-
Execute or cancel execution of notebooks, Spark job definitions, ML models and experiments.	✓	✓	✓	-
Execute or cancel execution of data pipelines.	✓	✓	✓	✓
View execution output of data pipelines, notebooks, ML models and experiments.	✓	✓	✓	✓
Schedule data refreshes via the on-premises gateway.[4]	✓	✓	✓	
Modify gateway connection settings.[4]	✓	✓	✓	

[1] Contributors and Viewers can also share items in a workspace, if they have Reshare permissions.

[2] Additional permissions are needed to read data from shortcut destination. Learn more about shortcut security model.

[3] Admins, Members, and Contributors can grant viewers granular SQL permissions to read Lakehouse and Data warehouse data through SQL endpoints.

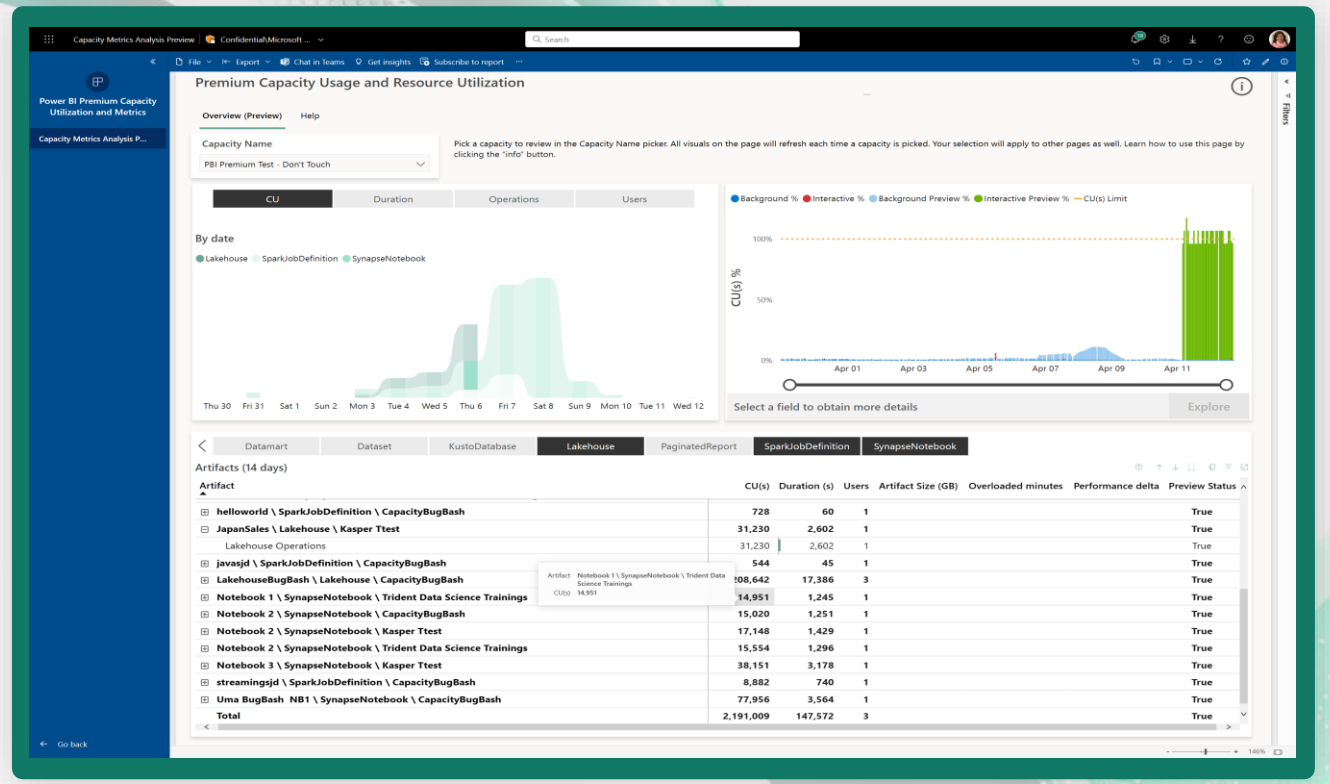
[4] Keep in mind that you also need permissions on the gateway. Those permissions are managed elsewhere, independent of workspace roles and permissions.

Capacity metrics reporting

Tenant wide visibility into capacity usage for all Fabric workloads in one place

Key Capabilities:

- Identify resource usage trends across Lakehouse, Notebook and Spark Job Definition and other Fabric artifacts operations
- View preview usage in parallel with other production workloads to make data-driven capacity sizing decisions
- Monitor the impact of throttling to user experience to make scale-up decisions

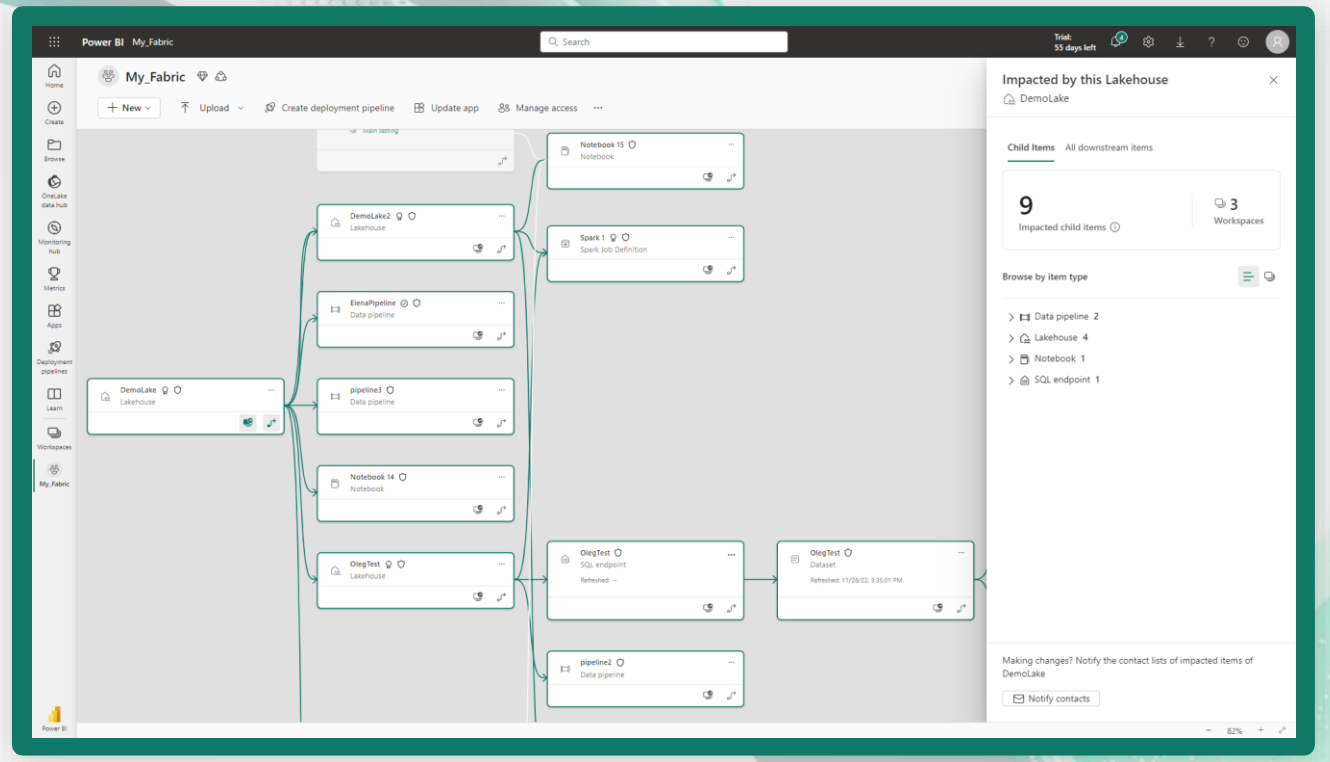


Track data lineage and impact analysis

Data owners can gain insight and track data journey and impact analysis across their analytics projects

Key Capabilities:

- See lineage view of complex analytical projects
- Understand how the data flows from the source to the destination 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
- Gain trust in your data while understanding the source of the data
- Perform impact analysis to assess impact for upcoming changes



Endorse data for visibility and reuse

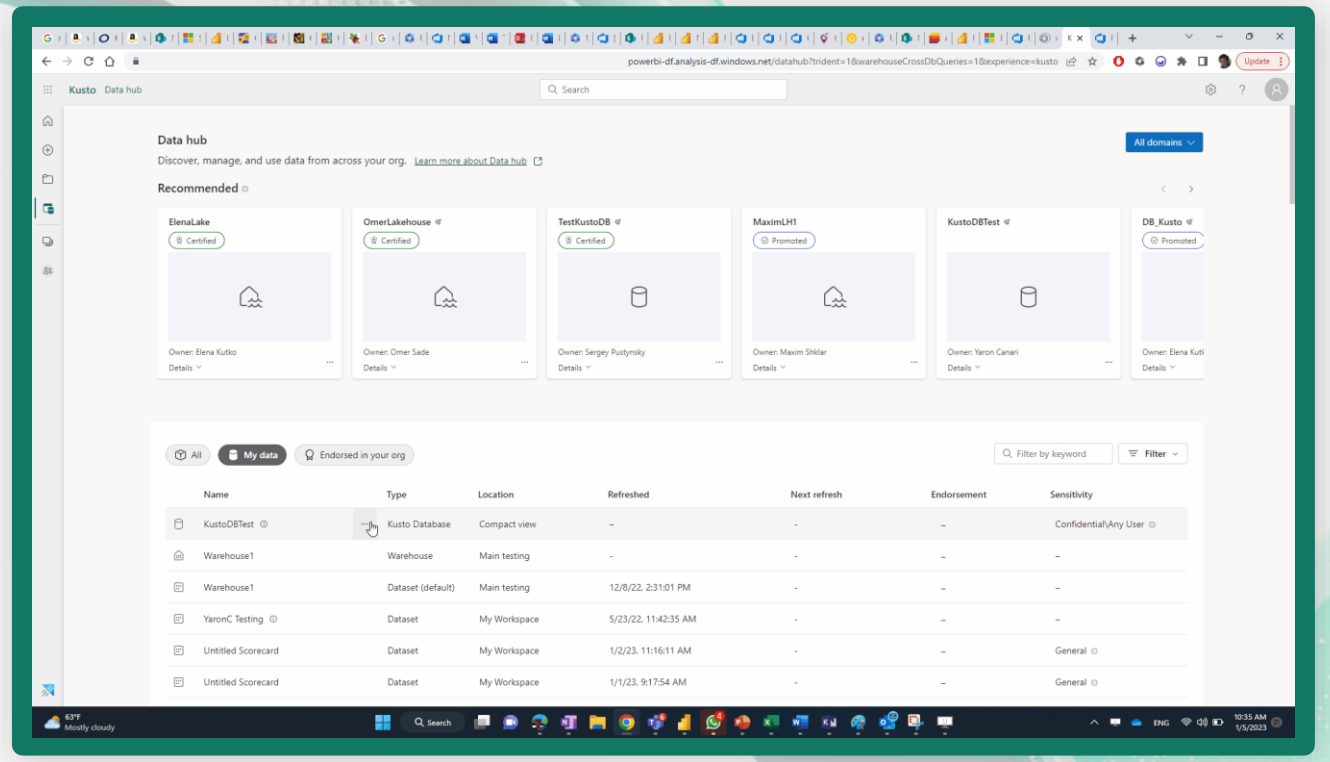
Data owners & Admin can endorse valuable, high-quality content in Fabric to create **curated sources of truth**, increase their **discoverability**, and encourage **reuse**

Key Capabilities:

Certification: authorized users/groups can certify data that meets the organization's quality standards

Promotion: content owners can promote data as a means to highlight data they find valuable and recommend for others to use

Endorsed items get higher visibility in Fabric data discovery experiences



Tenant-wide metadata visibility

Admins can gain **tenant wide visibility** of their data estate via Fabric's Scanner APIs

---

Key Capabilities:

- 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, incremental scans
- Supports 1st and 3rd party catalog tool integrations. Purview data catalog, Informatica, Collibra, Alation are integrated with Power BI, in various stages of integrating with Fabric

Sample Response

Status code: 200

JSON Copy


```
{
  "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",
            "certifiedBy": "john@contoso.com"
          },
          "sensitivityLabel": {
            "labelId": "85b38049-4259-43a2-8feb-244e222d96c0"
          },
          "users": [
            {
              "displayName": "John Nick",
              "emailAddress": "john@contoso.com",
              "appUserAccessRight": "ReadExplore",
              "identifier": "john@contoso.com",
              "graphId": "3fadb6e4-130c-4a8f-aeac-416e38b66756",
              "principalType": "User"
            }
          ]
        }
      ],
      "dashboards": [
        {
          "id": "80814ece-9302-49e3-b6bc-bad2f7a86c1a",
          "displayName": "CompositeModelParamsDashboard",
          "isReadOnly": false,
          "tiles": [
            {
              "id": "e687cc21-5b32-48f5-8c5e-4b844f190579",
              "title": "CompositeModelParamsDashboard",
              "reportId": "c6d072d1-ed20-4b60-8329-16c4b934203b",
              "datasetId": "132593c4-bf8d-4548-8f25-1ebb16a1613c"
            }
          ],
          "sensitivityLabel": {
            "labelId": "d9b9581a-0594-4c39-81c5-91ddf40baeda"
          },
          "users": [
            {
              "displayName": "John Nick",
              "emailAddress": "john@contoso.com",
              "appUserAccessRight": "ReadExplore",
              "identifier": "john@contoso.com",
              "graphId": "3fadb6e4-130c-4a8f-aeac-416e38b66756",
              "principalType": "User"
            }
          ]
        }
      ],
      "datasets": [
        {
          "id": "e7e8a355-e77b-4418-a7b8-ae5972fdaa03",
          "name": "ExportB",
          "tables": [
            {
              "name": "DW_Revenues",
              "columns": [
                {
                  "name": "RowNumber-2662979B-1795-4F74-8F37-6A1BA8059B61",
                  "dataType": "Int64",
                  "isHidden": true
                }
              ],
              "measures": [
                {
                  "name": "MyMeasure",
                  "expression": "CALCULATE(SELECTEDVALUE('DW_Revenues DW_RevenuesTestToBeDeleted'[Nu
                  "description": "My measure",
                  "isHidden": false
                }
              ],
              "isHidden": false,
              "description": "My table",
              "source": [
                {
```

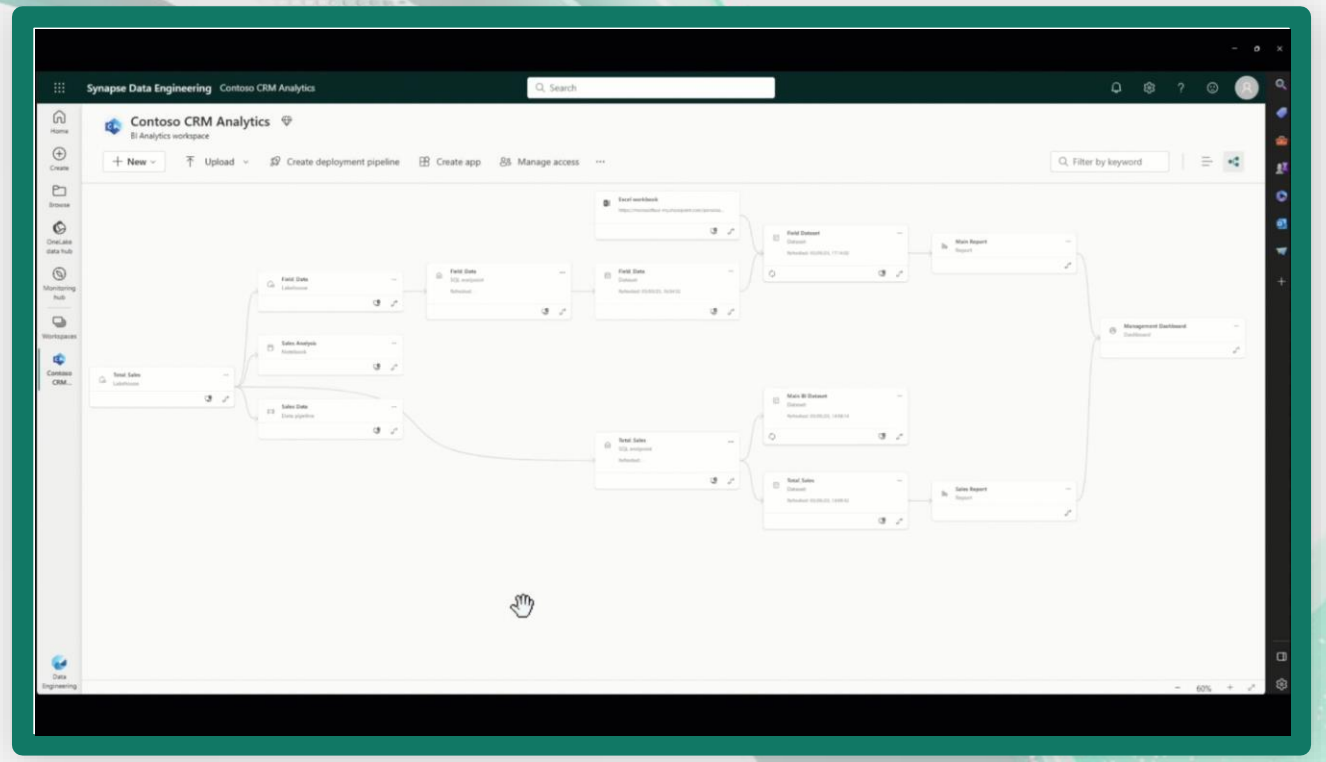
Purview information protection for Fabric

Admins and Data owners can ensure data remains protected and classified across Fabric leveraging sensitivity labels

---

Key Capabilities:

- Label Fabric items with organization's sensitivity labels, used in Microsoft 365 apps
- Sensitivity labels follow the data automatically as it flows from data lake to business users within Fabric and to Office

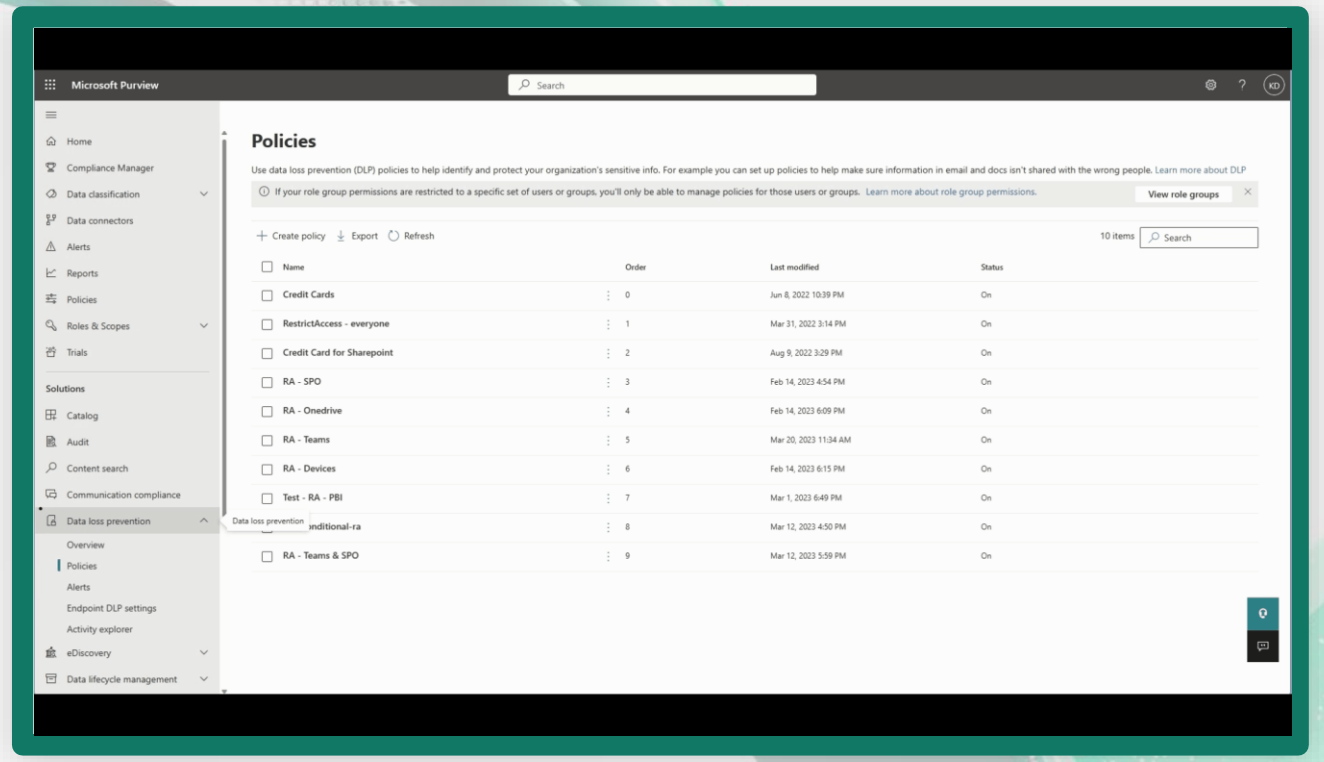


Purview loss prevention policies

Compliance admins can configure data loss prevention policies to detect upload of sensitive data to Power BI models

Key Capabilities:

- Automatic detection of upload of sensitive data, to Power BI datasets, supporting trigger alerts for admins and custom policy tip to data owners
- Data owners can provide feedback using report a false positive action
- Expansion to the full set of Fabric's is in progress



End-to-end tutorials

Lakehouse tutorial
https://learn.microsoft.com/en-us/fabric/data-engineering/tutorial-lakehouse-introduction

Data Science tutorial
https://learn.microsoft.com/en-us/fabric/data-science/tutorial-data-science-introduction

Real-Time Analytics tutorial
https://learn.microsoft.com/en-us/fabric/real-time-analytics/tutorial-introduction

Data warehouse tutorial
https://learn.microsoft.com/en-us/fabric/data-warehouse/tutorial-introduction

Power BI tutorial
https://learn.microsoft.com/en-us/power-bi/fundamentals/fabric-get-started

Data Factory tutorial
https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-end-to-end-introduction

Session Feedback

Event Feedback

DATA:Scotland

Slides can be found at :

https://github.com/BenniDeJagere/Presentations/{Year}/{Date}_{Event}