Microsoft

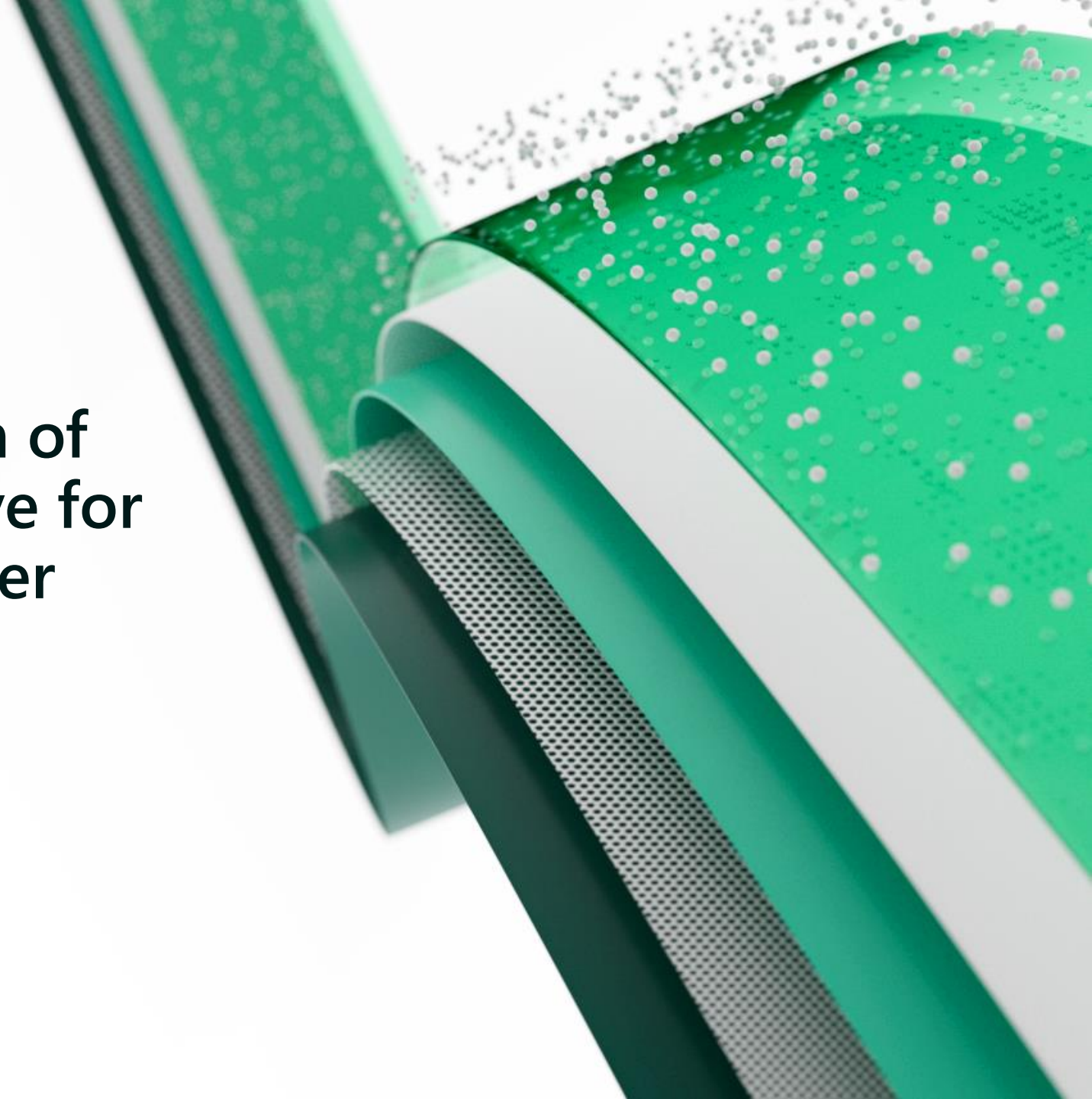
OneLake - the foundation of Microsoft Fabric, OneDrive for data and the data silo killer

Pawel Potasinski

Senior Program Manager

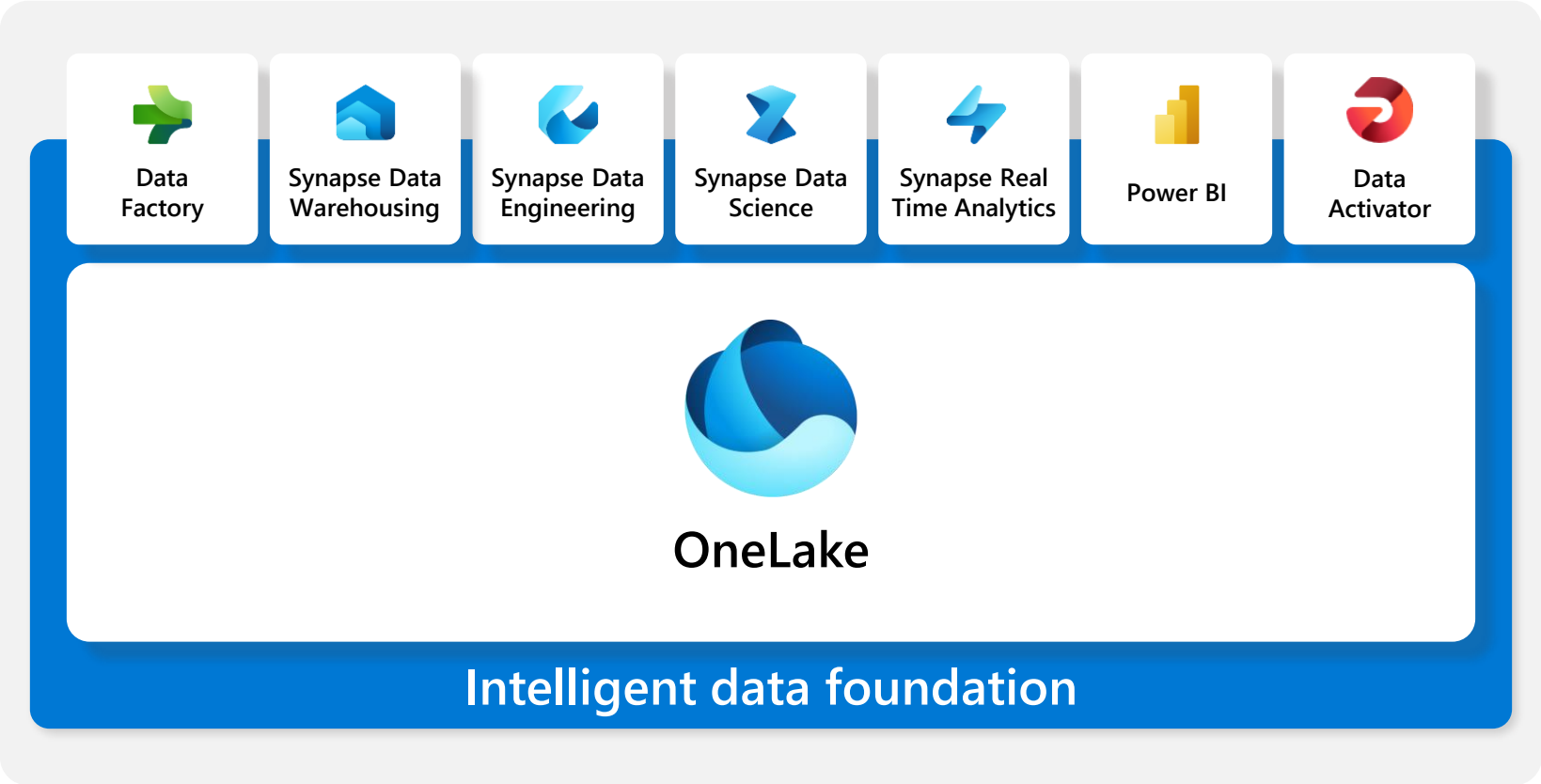
@pawelpotasinski

/in/pawelpotasinski



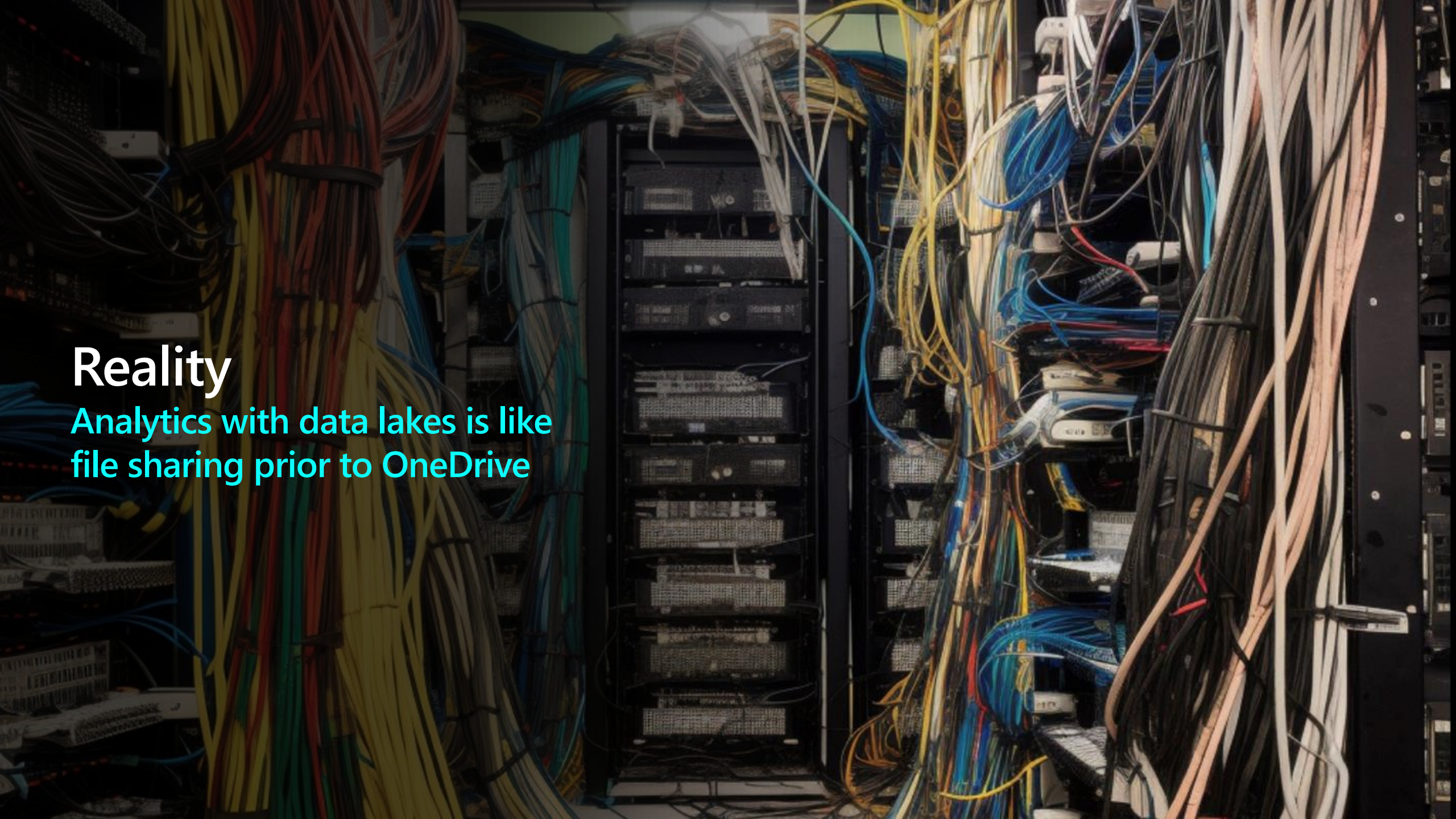
Microsoft Fabric

The data platform for the era of AI



Expectations of data lakes

- One place for an organization to land all data, structured and unstructured.
- Break down data silos, making it easier to blend and analyze data together.
- Simplify security, governance and data discovery enabling all user and applications to access the data they need.



Reality

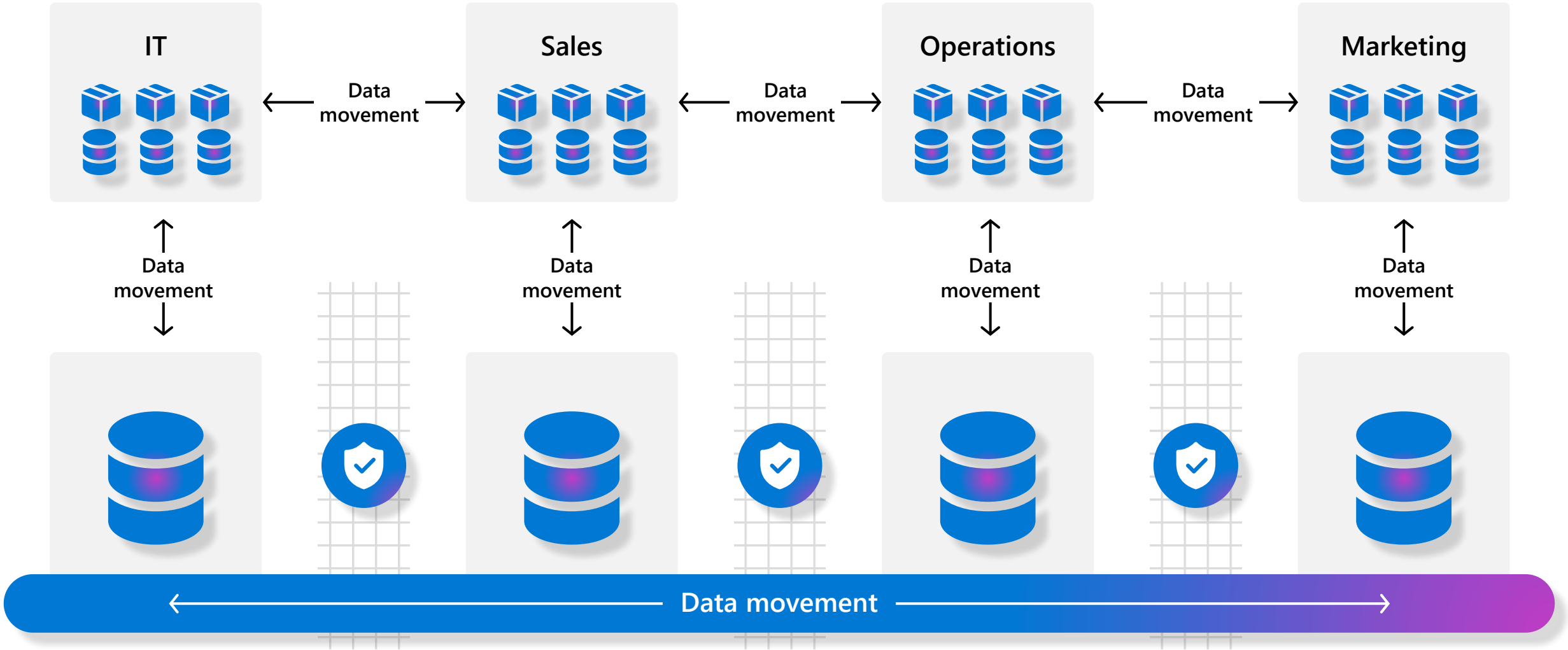
Analytics with data lakes is like file sharing prior to OneDrive

Reality: DIY data lakes using storage

You buy storage and build a complicated data lake solution on top of it.

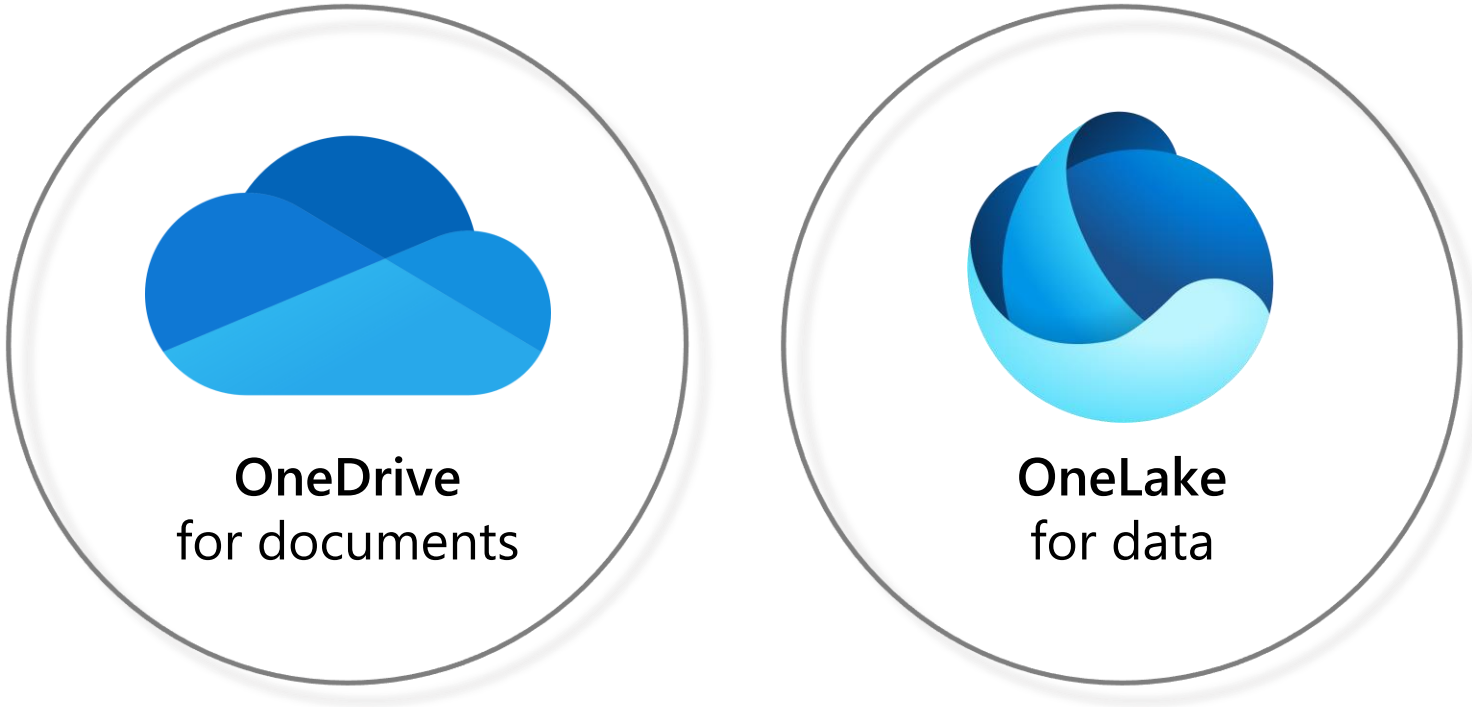
- Getting the desired value from one data lake is hard.
- Business groups get frustrated with the pace of change when having to coordinate through a central team.
- Data mesh pattern can enable business groups to work independently with multiple business domain driven lakes.
- However, Data Mesh requires each team to manage their own lake.

Today: Multiple siloed lakes with lots of duplication



OneLake for all data

"The OneDrive for data"



OneLake provides a data lake as a service without you needing to build it

OneLake for all data

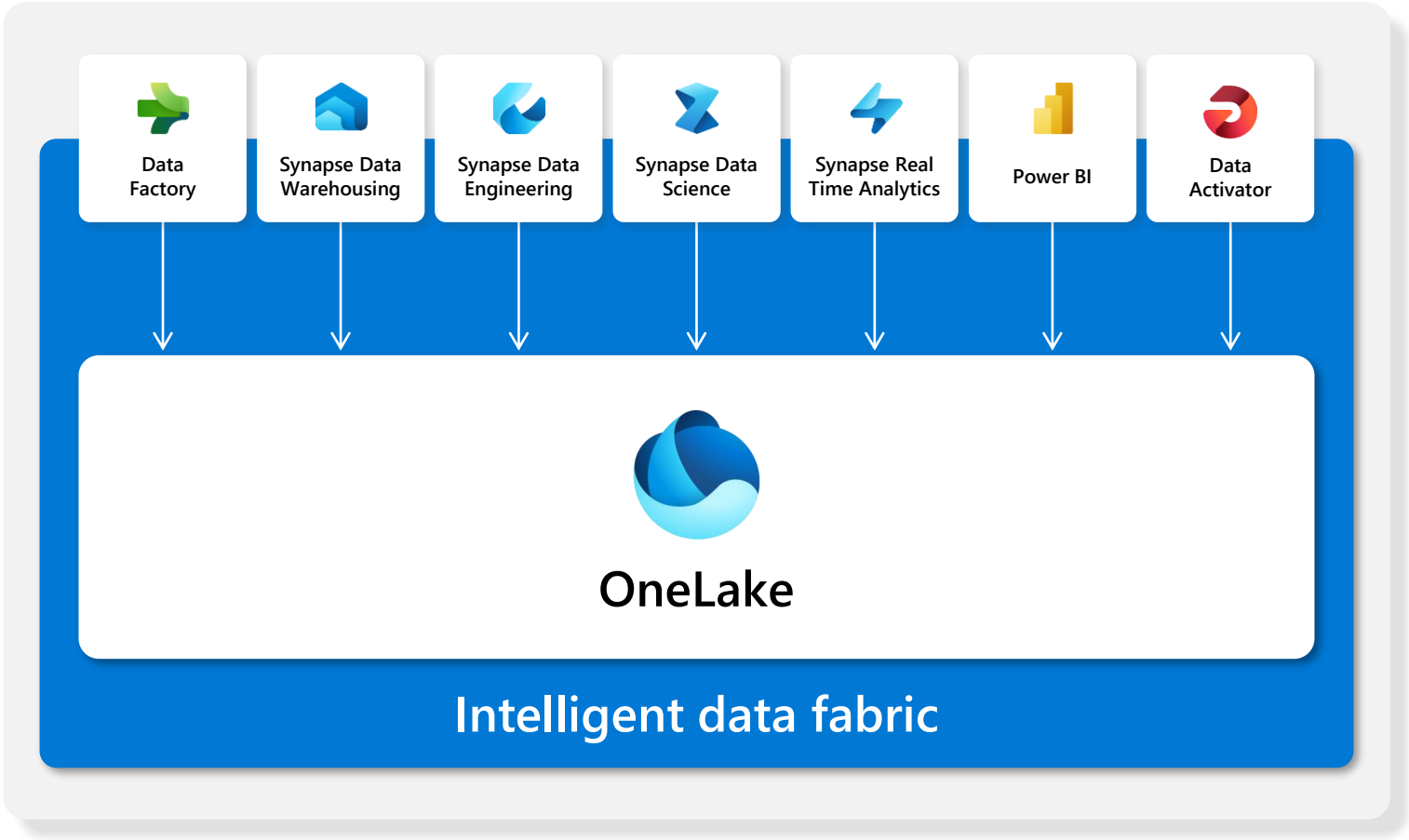
"The OneDrive for data"

OneLake

One Copy

One Security

OneLake Data Hub



OneLake for all data

"The OneDrive for data"

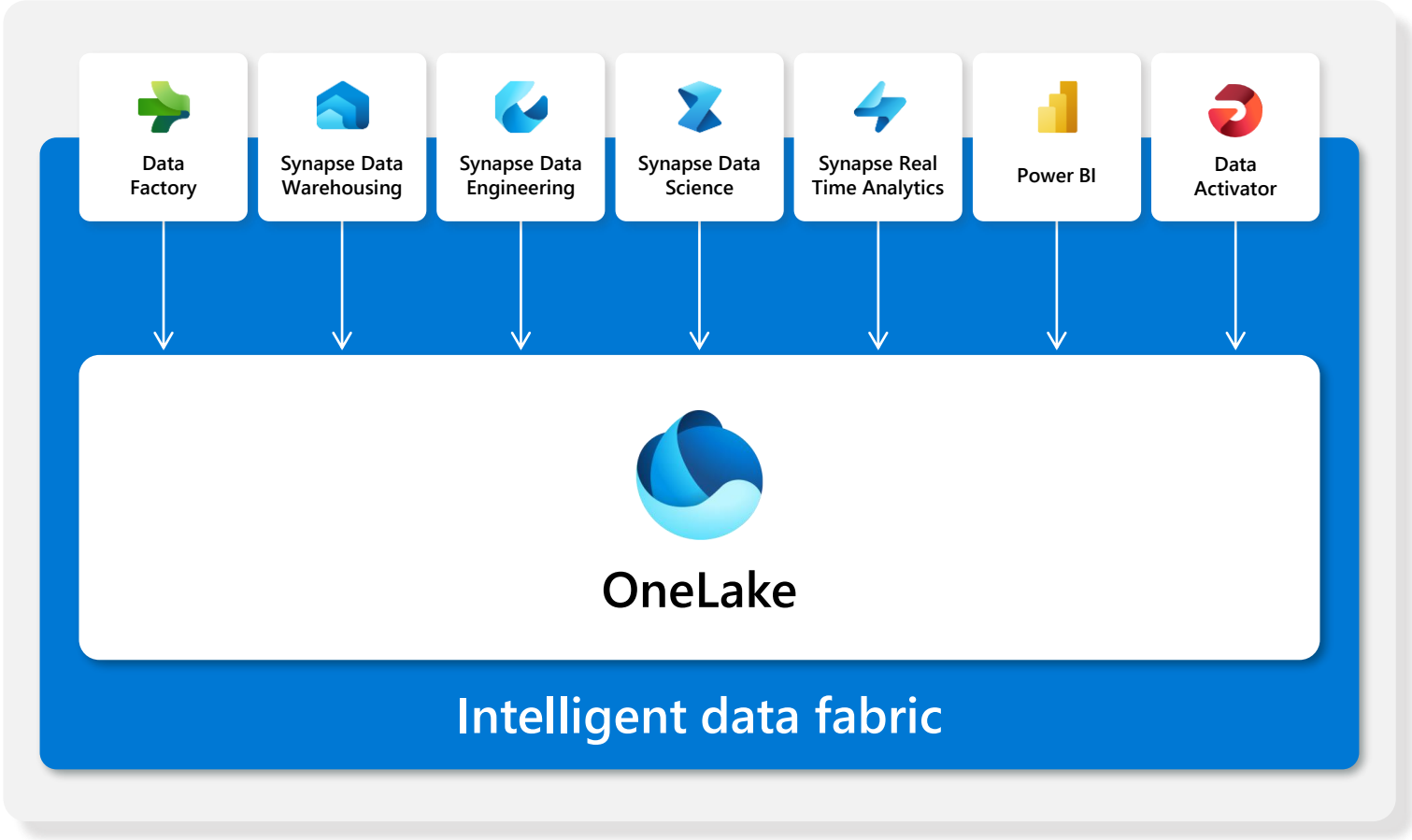
OneLake

- A single unified logical SaaS data lake for the whole organization (no silos)
- Organize data into domains
- Foundation for all Fabric data items
- Provides full and open access through industry standard APIs and formats to any application (no lock-in)

One Copy

One Security

OneLake Data Hub



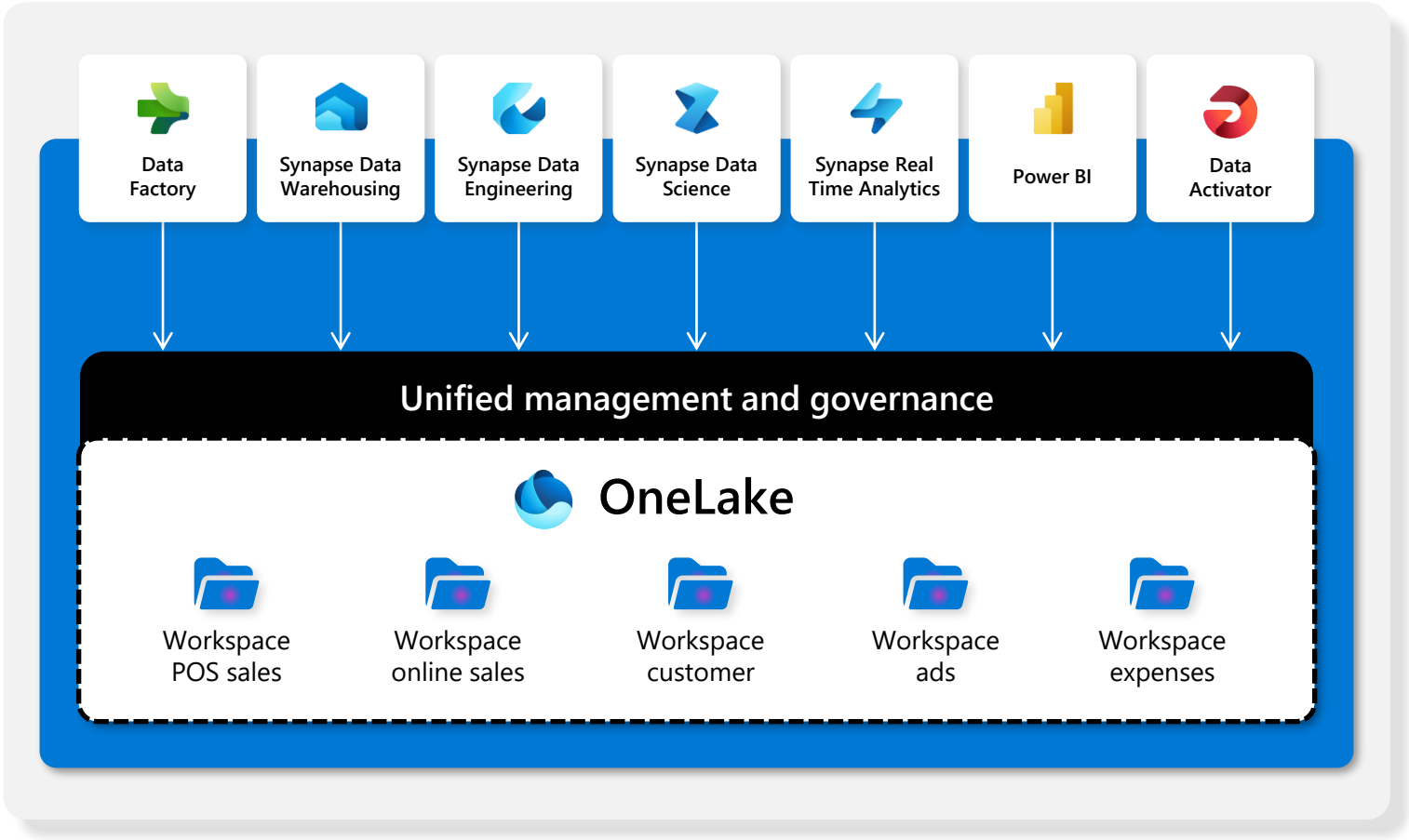
A single unified SaaS data lake

"No silos"

OneLake comes **automatically provisioned** with every Fabric tenant with no infrastructure to manage.

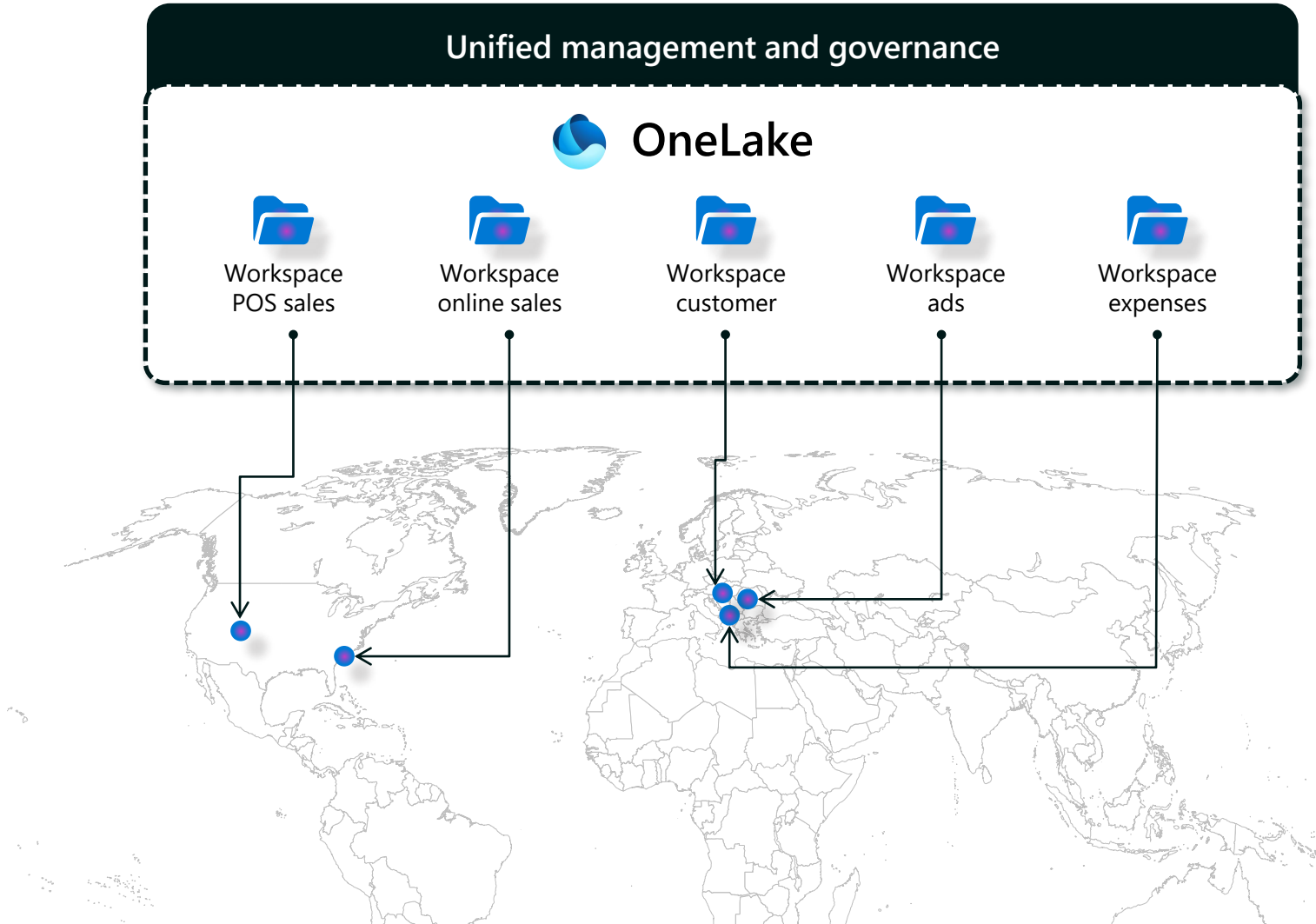
Any data in OneLake works with **out-of-the-box governance** such as data lineage, data protection, certification, catalog integration, etc. All data is ultimately under the control of a tenant admin.

OneLake enables **distributed ownership**. Different workspaces allow different parts of the organization to work independently while still contributing to the same data lake. Each workspace can have its own administrator, access control, region and capacity for billing.



OneLake which logically spans the world

Workspaces can reside in **different regions** around the world while still being part of the same data lake.



The foundation of all Fabric data items

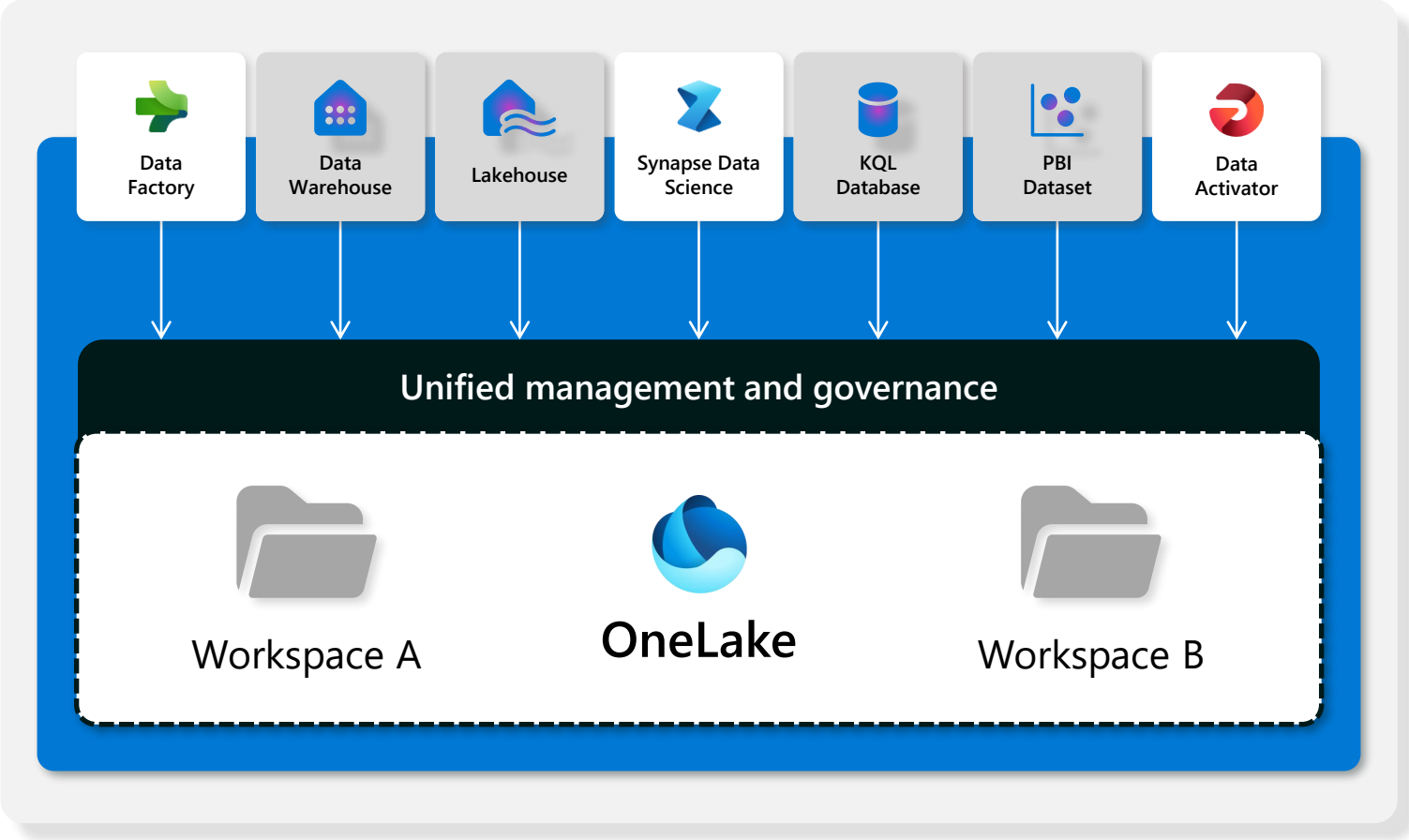
Fabric data items are prewired to store data in OneLake in open formats

All data is stored in a Fabric **data item**.

In Power BI users interact with data via items like datasets.

In Fabric, different items give tailored experiences for each persona and scenario when accessing OneLake.

Like how Office stores Word, Excel, and PowerPoint documents in OneDrive.



The foundation of all Fabric data items

Fabric data items are prewired to store data in OneLake in open formats

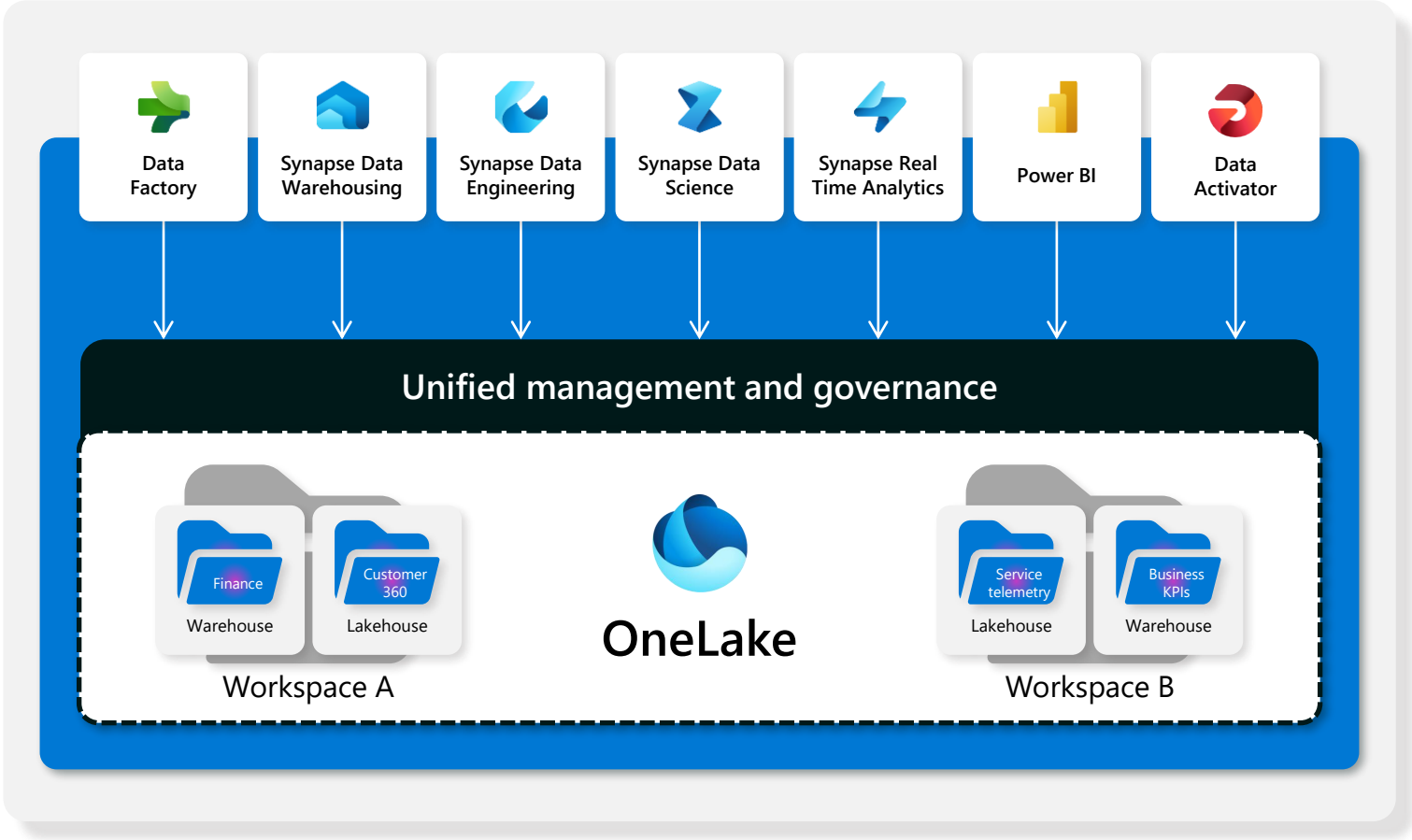
All data is stored in a Fabric **data item**.

In Power BI users interact with data via items like datasets.

In Fabric, different items give tailored experiences for each persona and scenario when accessing OneLake.

Like how Office stores Word, Excel, and PowerPoint documents in OneDrive.

All data is stored on **open file formats**. All tabular data will be written in **delta format**.



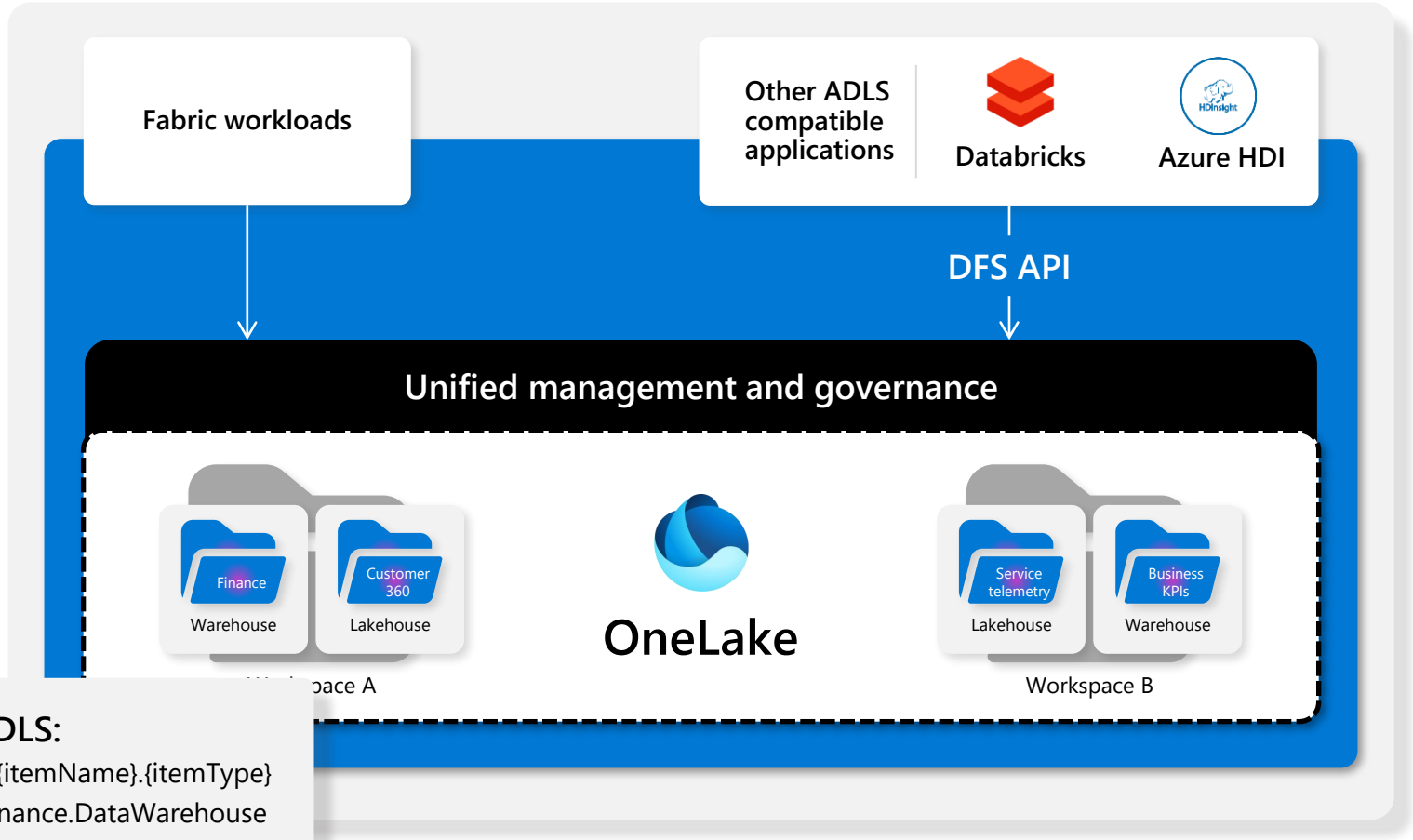
Open access to data in OneLake

No lock-in with industry standard APIs and open file formats

By supporting the ADLS Gen2 DFS APIs and SDKs, OneLake is **compatible with existing ADLS applications**

Tenants will appear as one big storage account with different workspaces appearing as different containers with **data organized into folders**

Underlying physical **storage is virtualized** away. OneLake ensures proper scale and performance.



Addressing Fabric workspaces and items as ADLS:

https://onelake.dfs.fabric.microsoft.com/{workspaceName}/{itemName}.{itemType}

Ex. https://onelake.dfs.fabric.microsoft.com/Workspace A/Finance.DataWarehouse

OneLake for all domains

OneLake gives a true data mesh as a service

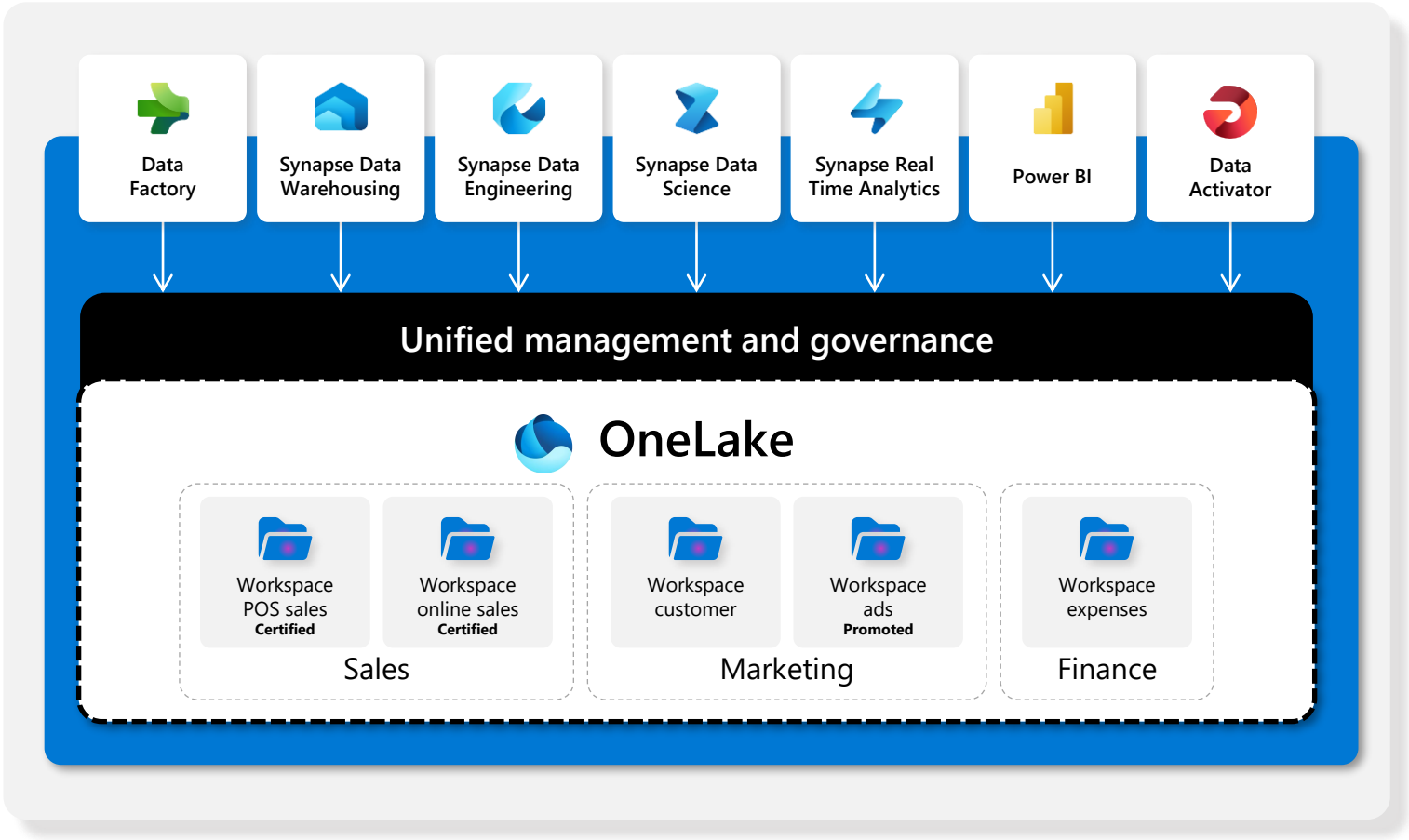
Introducing **domains** as an integral part of Fabric: A domain is a way to logically group together all the data in an organization relevant to an area or field, according to business needs

Domains are defined with **domain admins** and **contributors** who can **associate** workspaces and group them together under a relevant domain

Federated governance can be achieved by delegating settings to domain admins, thus allowing them to achieve more **granular control** over their business area

Domains simplify **discovery** and **consumption** of data across the organization, thus allowing business optimized consumption

Avoid data swamps by endorsing certain data as **certified** or **promoted**, **thus encouraging reuse**.



DEMO

OneLake in action

OneLake for all data

"The OneDrive for data"

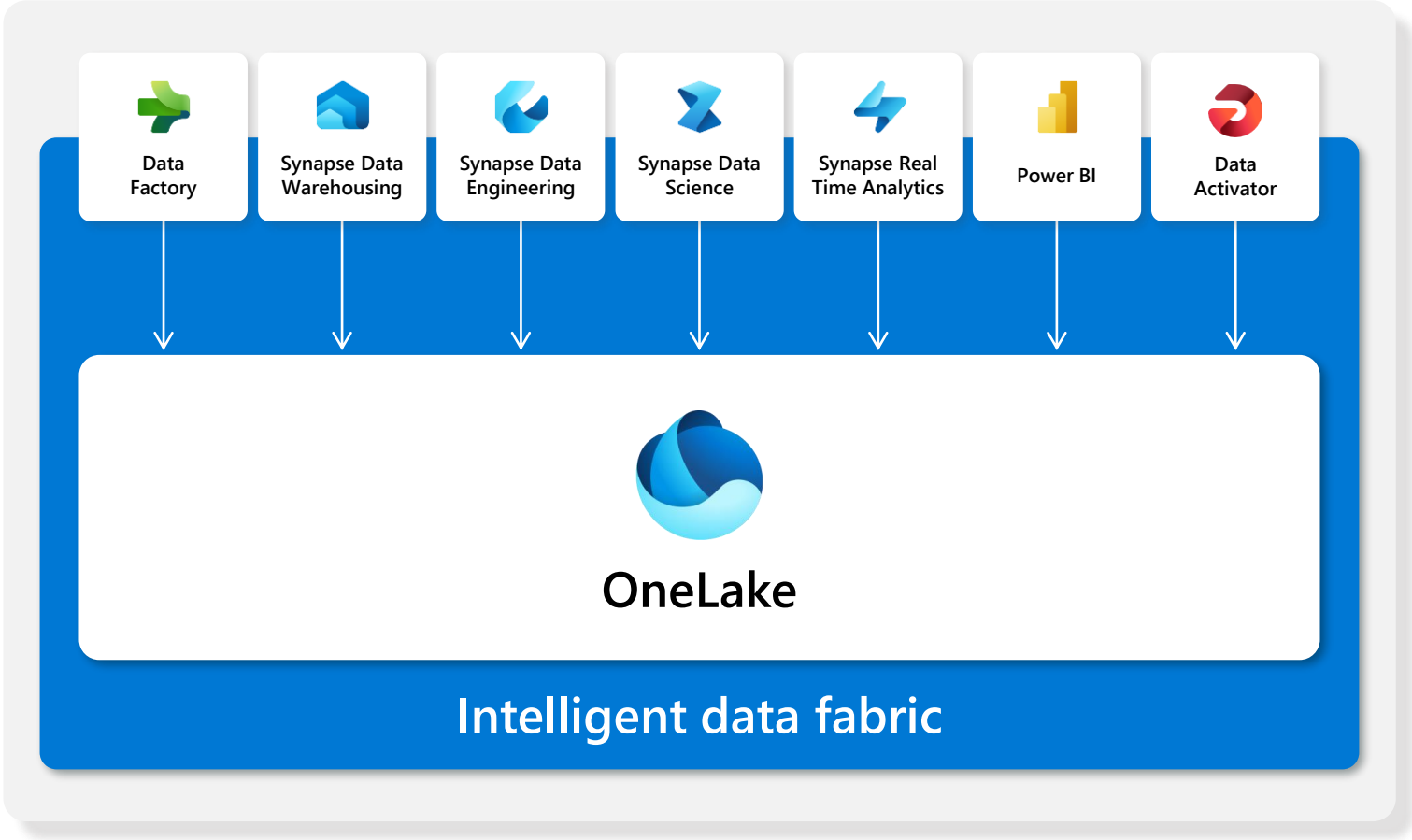
OneLake

- A single unified logical SaaS data lake for the whole organization (no silos)
- Organize data into domains
- Foundation for all Fabric data items
- Provides full and open access through industry standard APIs and formats to any application (no lock-in)

One Copy

One Security

OneLake Data Hub



OneLake for all data

"The OneDrive for data"

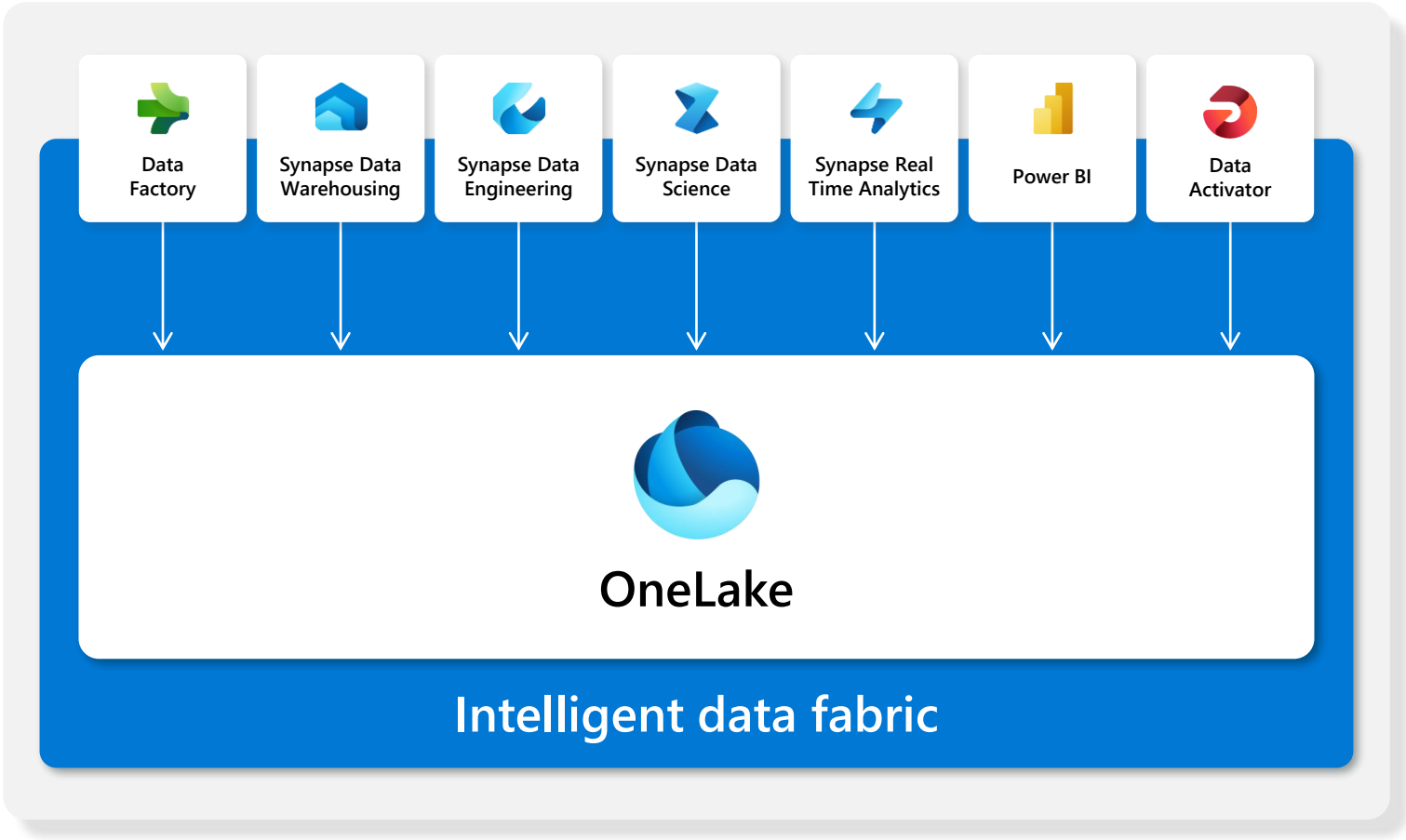
OneLake

One Copy

- Virtualize data across domains and clouds into a single logical lake with shortcuts
- The One Copy of data for all the analytical engines of Fabric without moving or duplicating data

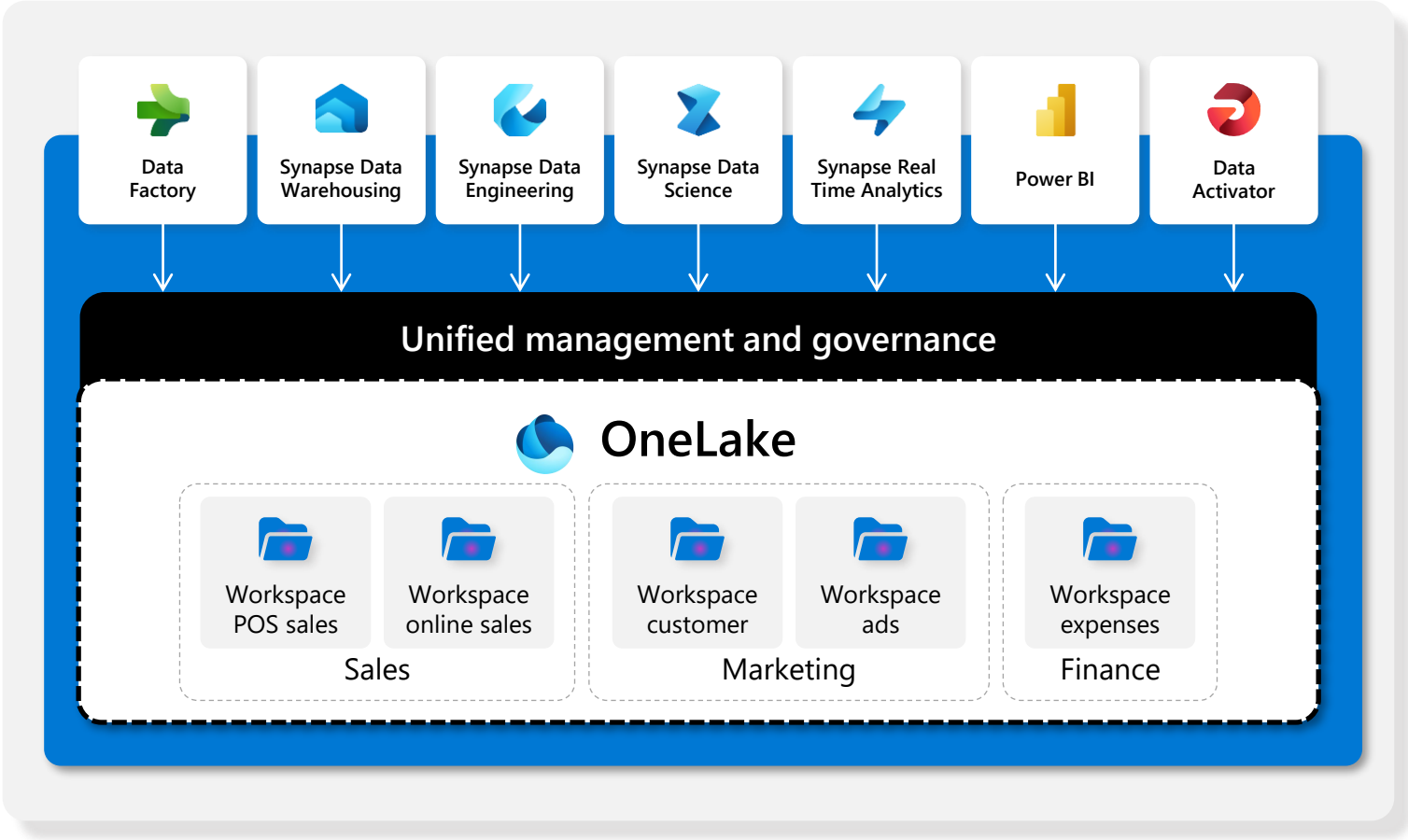
One Security

OneLake Data Hub



OneLake gives a true data mesh as a service

One Copy enables data to be used across domains, clouds and engines

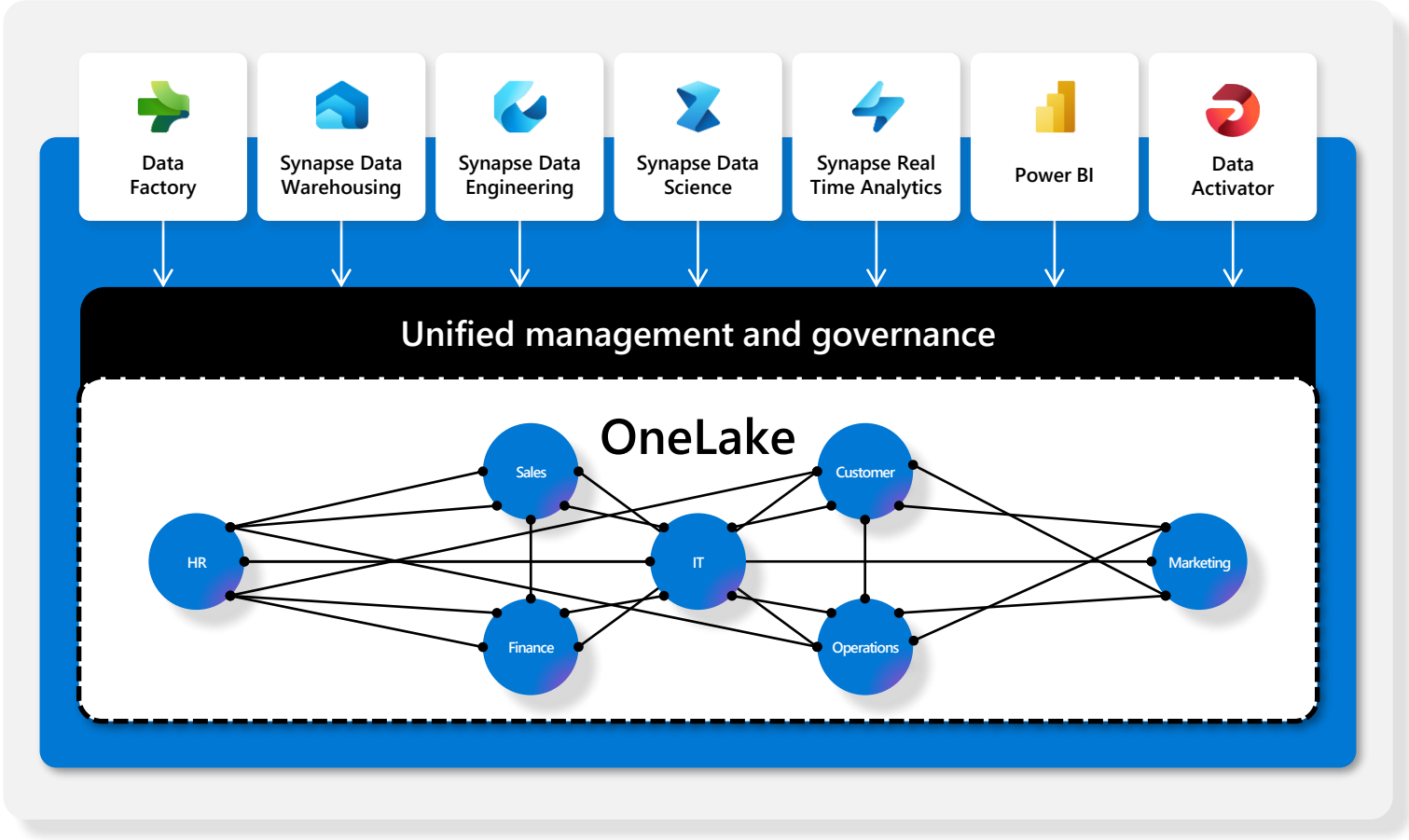


OneLake gives a true data mesh as a service

One Copy enables data to be used across domains, clouds and engines

An organization will have **many data domains** with many workspaces with different data owners. However, a single data product can span multiple domains.

Shortcuts provide the connections between domains so that data can be virtualized into a single data product without data movement, data duplication or changing the ownership of the data.



Shortcuts virtualize data across domains and clouds

No data movements or duplication

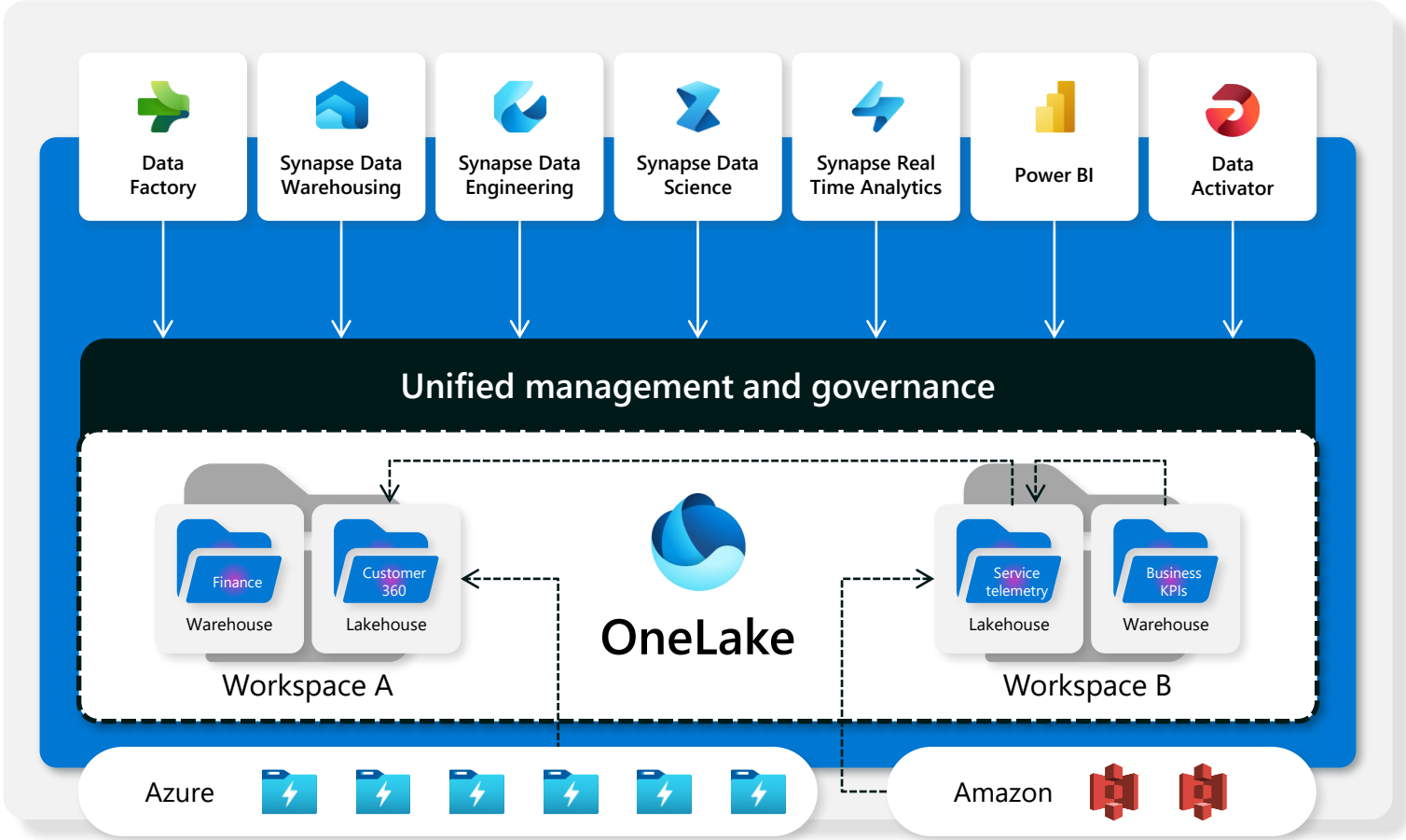
A shortcut is a **symbolic link** which points from one data location to another

Create a shortcut to make data from a warehouse part of your lakehouse

Create a shortcut within Fabric to **consolidate data** across items or workspaces without changing the ownership of the data. Data can be reused multiple times **without data duplication**.

Existing **ADLS Gen2** storage accounts and **Amazon S3 buckets** can be managed externally to Fabric and Microsoft while still being virtualized into OneLake with shortcuts

All data is mapped to a **unified namespace** and can be accessed using the same APIs including the ADLS Gen2 DFS APIs



Shortcuts virtualize data across domains and clouds

No data movements or duplication

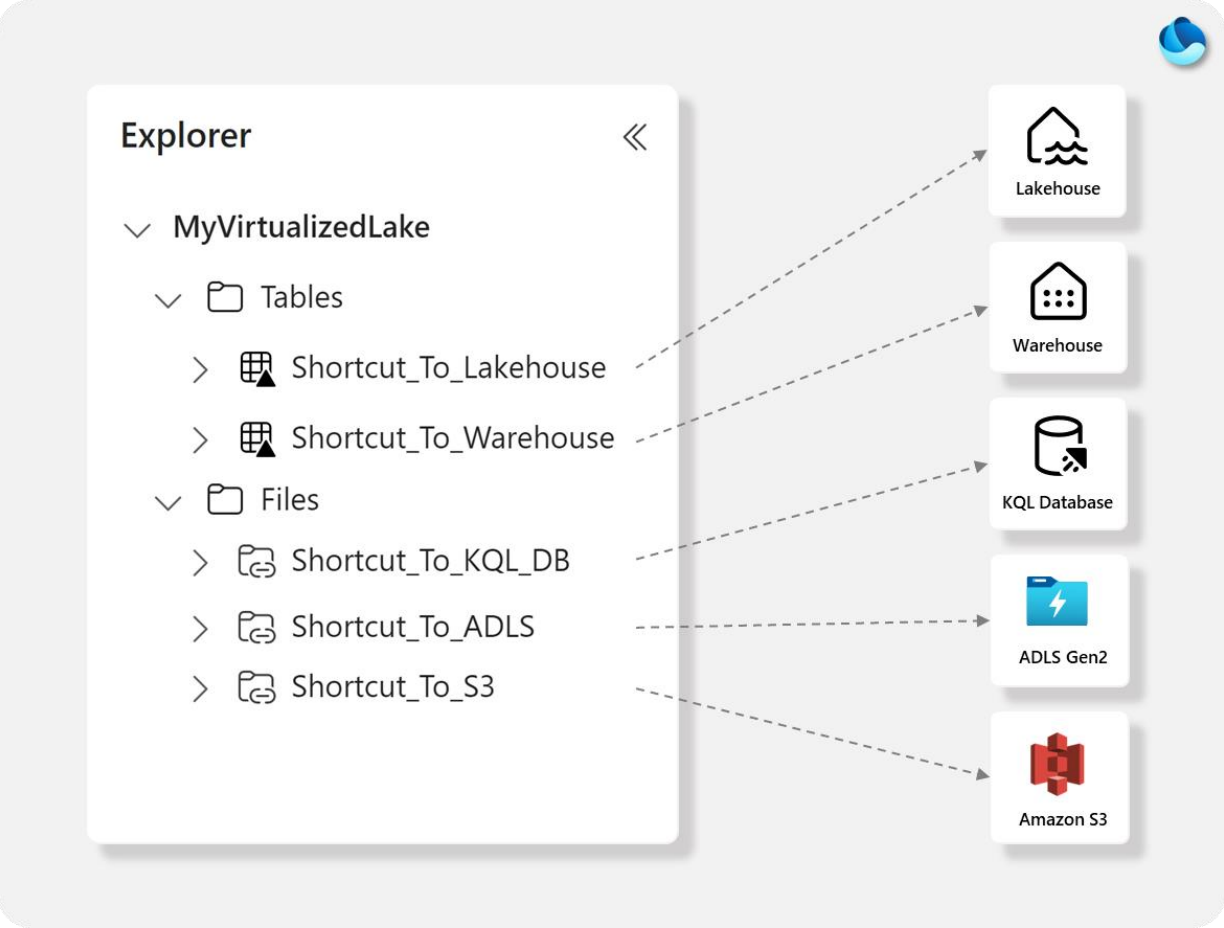
A shortcut is a **symbolic link** which points from one data location to another

Create a shortcut to make data from a warehouse part of your lakehouse

Create a shortcut within Fabric to **consolidate data** across items or workspaces without changing the ownership of the data. Data can be reused multiple times **without data duplication**.

Existing **ADLS Gen2** storage accounts and **Amazon S3 buckets** can be managed externally to Fabric and Microsoft while still being virtualized into OneLake with shortcuts

All data is mapped to a **unified namespace** and can be accessed using the same APIs including the ADLS Gen2 DFS APIs



DEMO

One Copy with shortcuts

OneLake for all data

"The OneDrive for data"

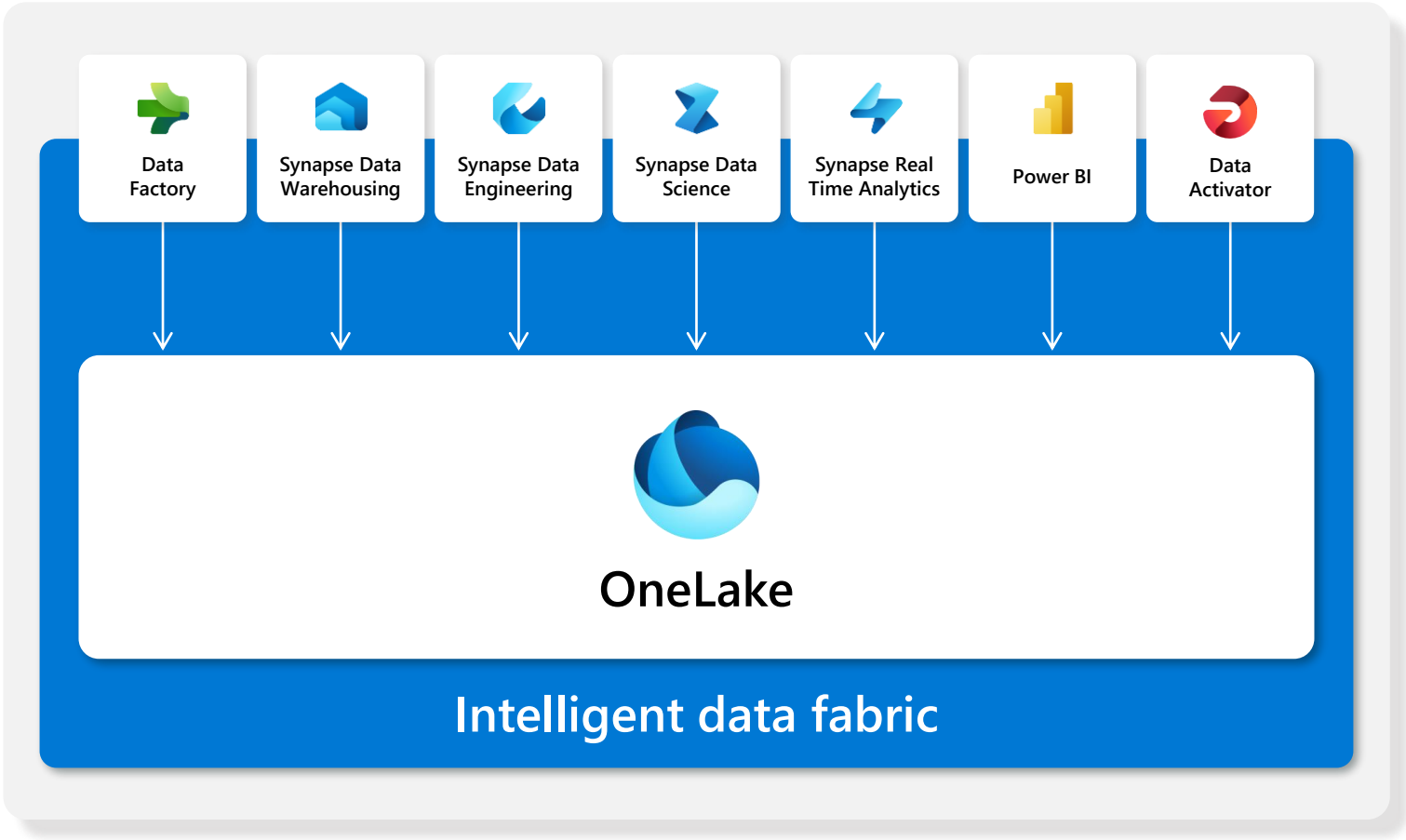
OneLake

One Copy

- Virtualize data across domains and clouds into a single logical lake with shortcuts
- The One Copy of data for all the analytical engines of Fabric without moving or duplicating data

One Security

OneLake Data Hub



OneLake for all data

"The OneDrive for data"

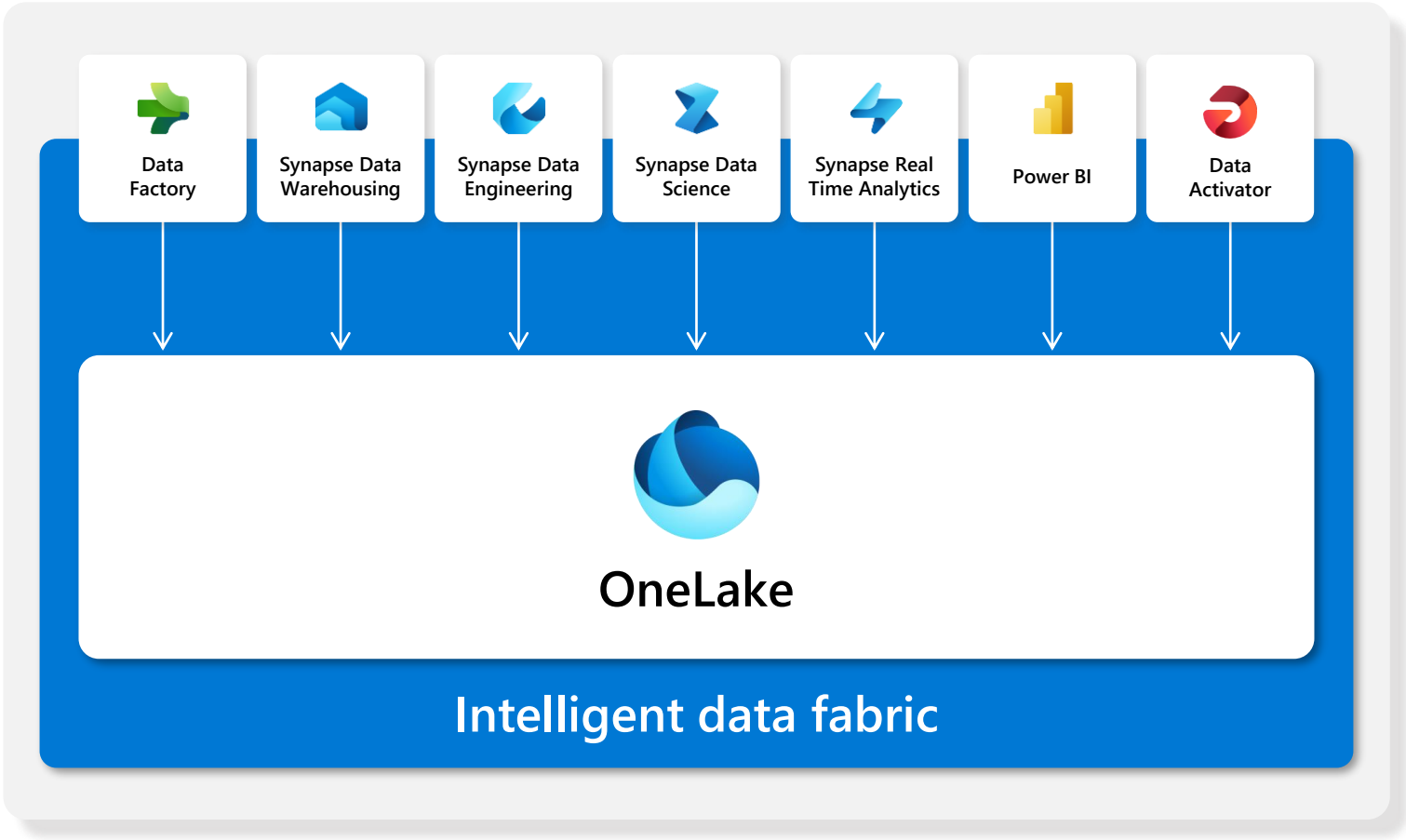
OneLake

One Copy

One Security

- Secure once and use anywhere

OneLake Data Hub



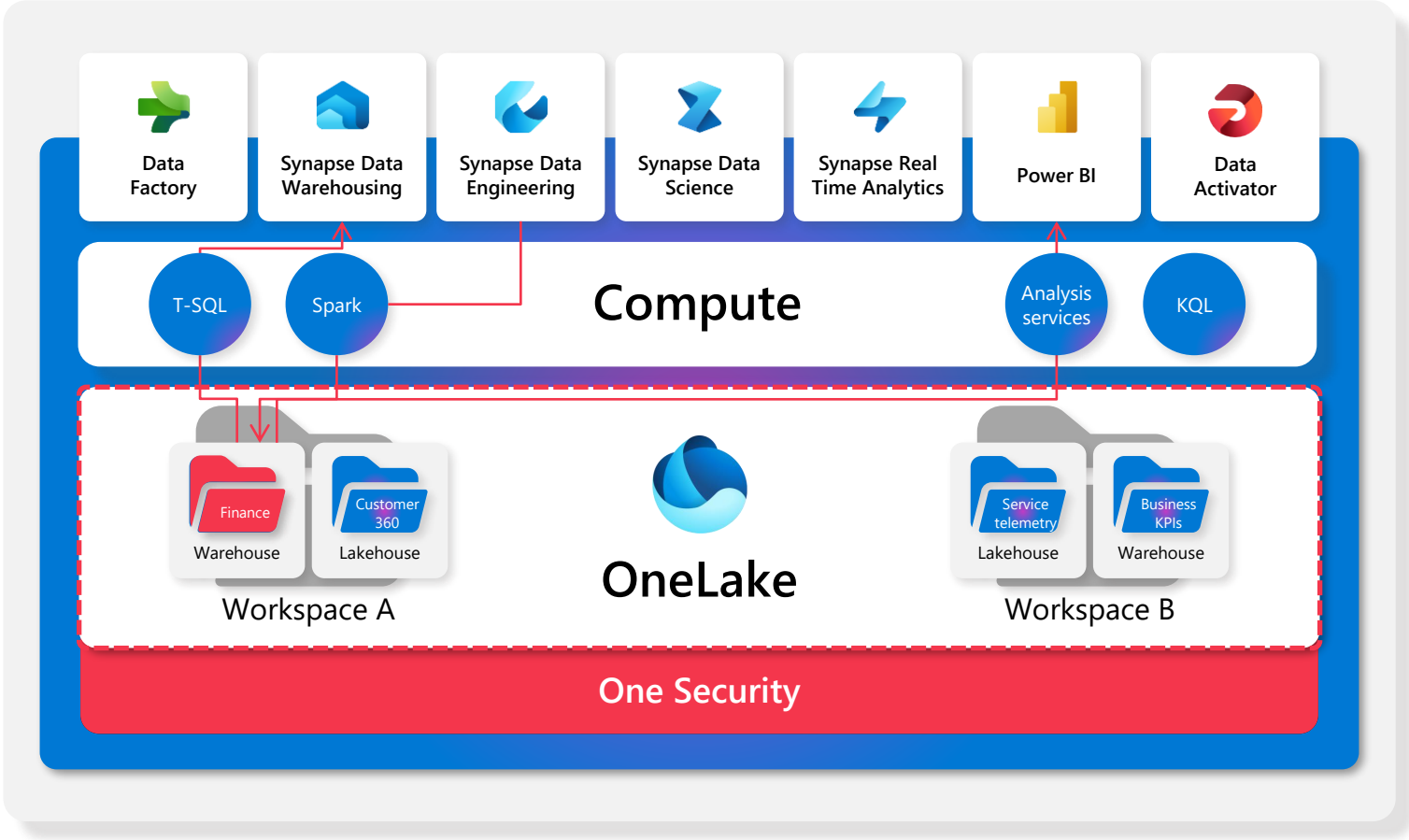
One Security (roadmap)

Secure once and use anywhere

One Security: a shared **universal security model** defined in OneLake alongside the data providing the necessary security capabilities so that data does not need to be copied into another place to be secured

More granular security can be defined once in OneLake. This includes column level security, row level security, data masking, etc.

Security definitions will flow across **all shortcuts** and will be enforced in **all engines** including non-Fabric engines



OneLake for all data

"The OneDrive for data"

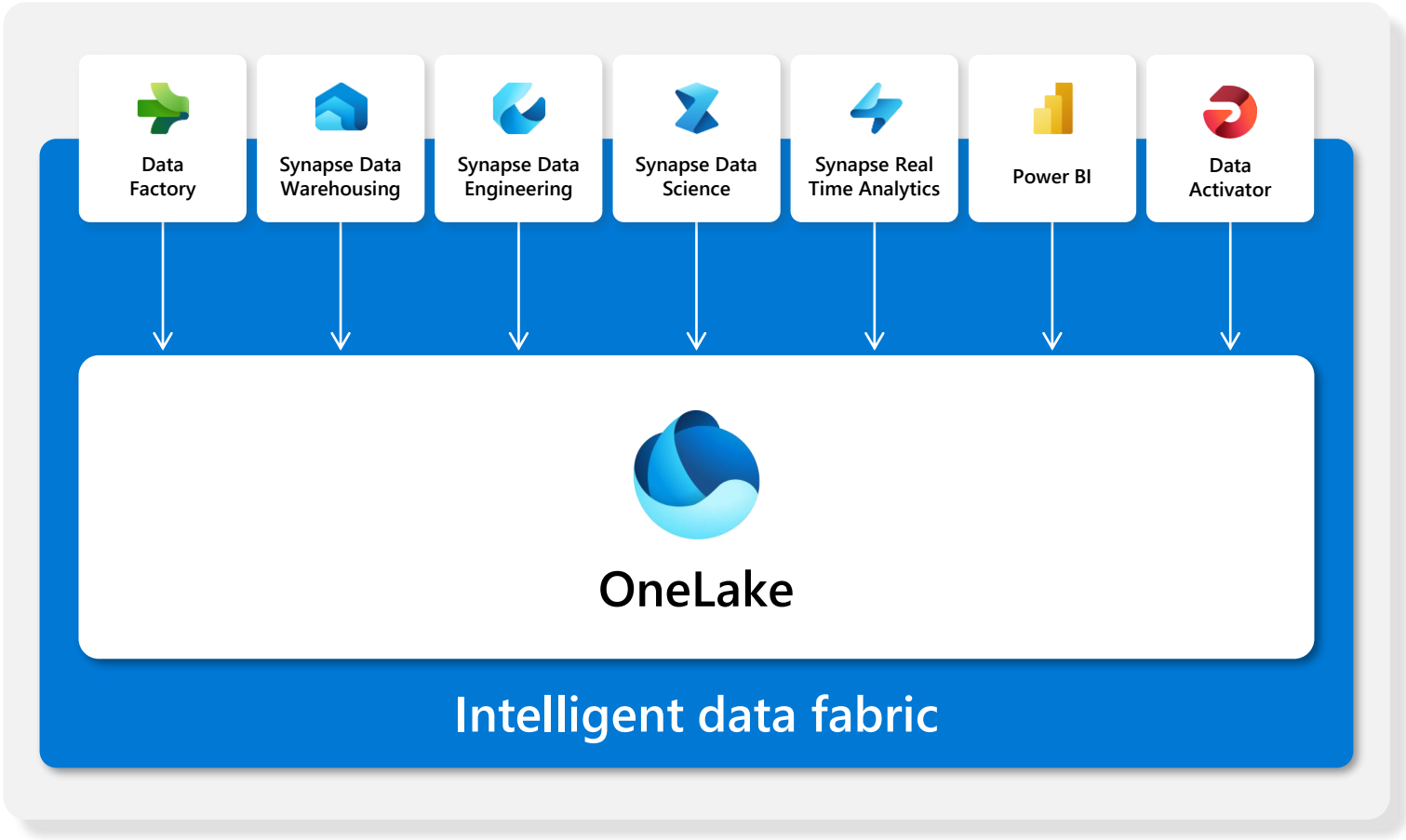
OneLake

One Copy

One Security

- Secure once and use anywhere

OneLake Data Hub



OneLake for all data

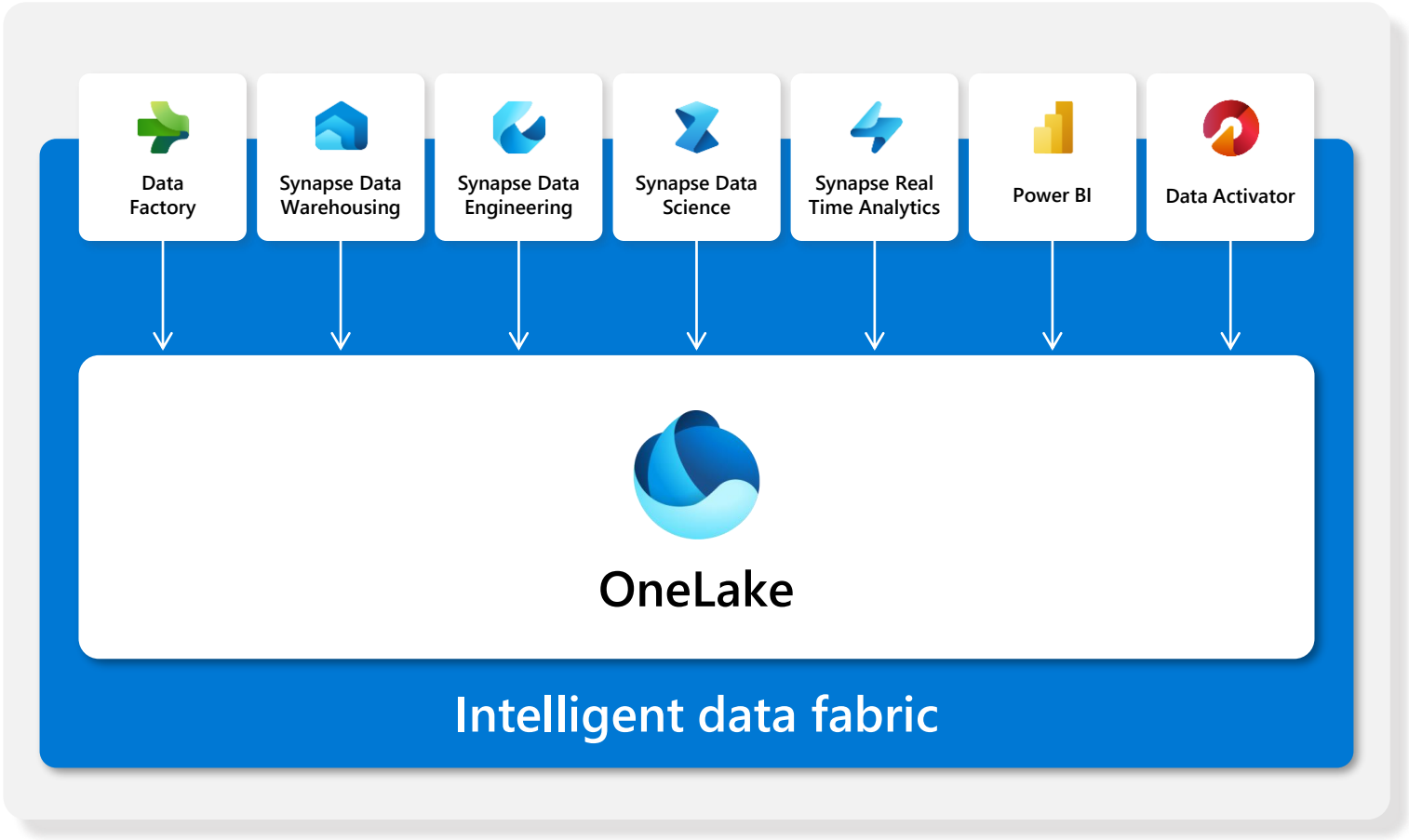
"The OneDrive for data"

OneLake

One Copy

One Security

OneLake Data Hub



OneLake Data Hub

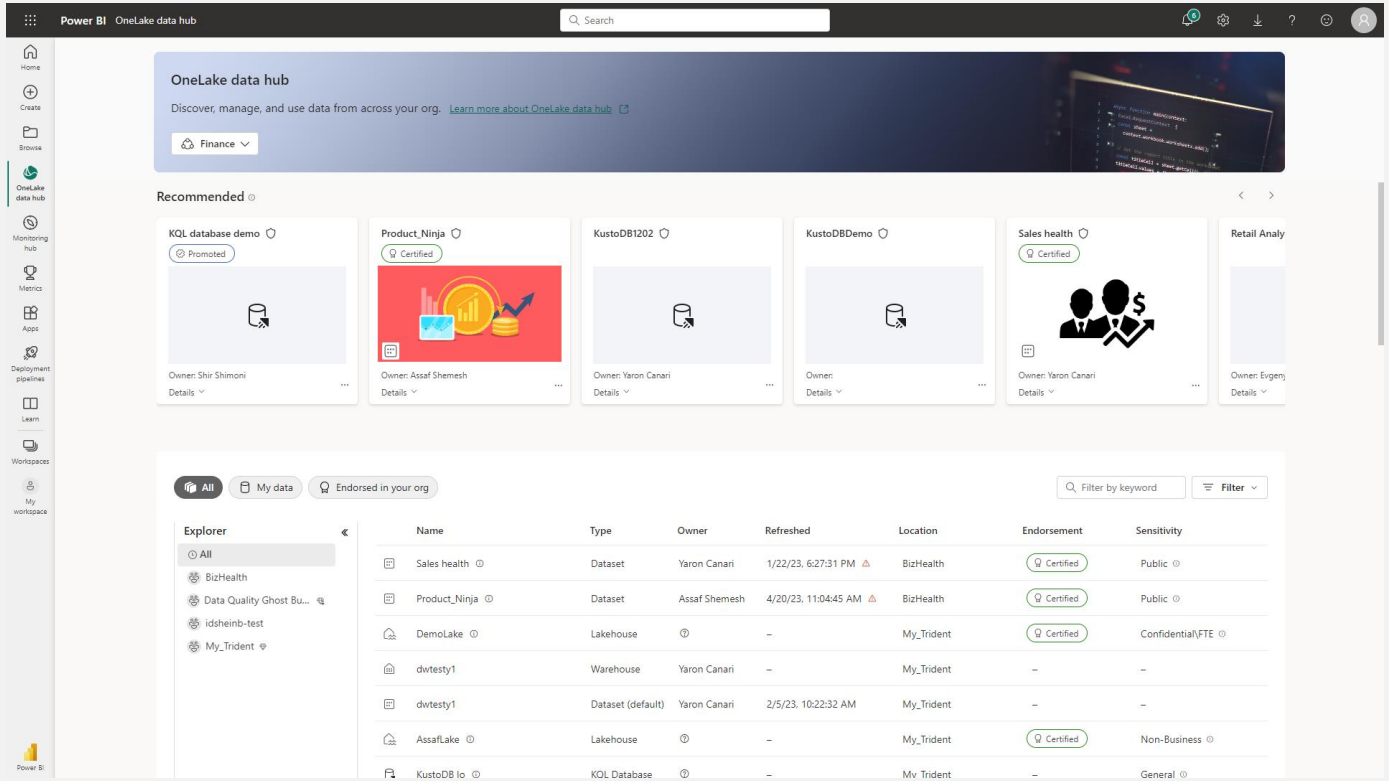
Discover, manage and use data in one place

Central location within Fabric to discover, manage, and reuse data

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

Efficient data discovery using **search, filter and sort**

Explorer capability to easily browse and find data by its folder (workspace) hierarchy



DEMO

OneLake Data Hub

OneLake for all data

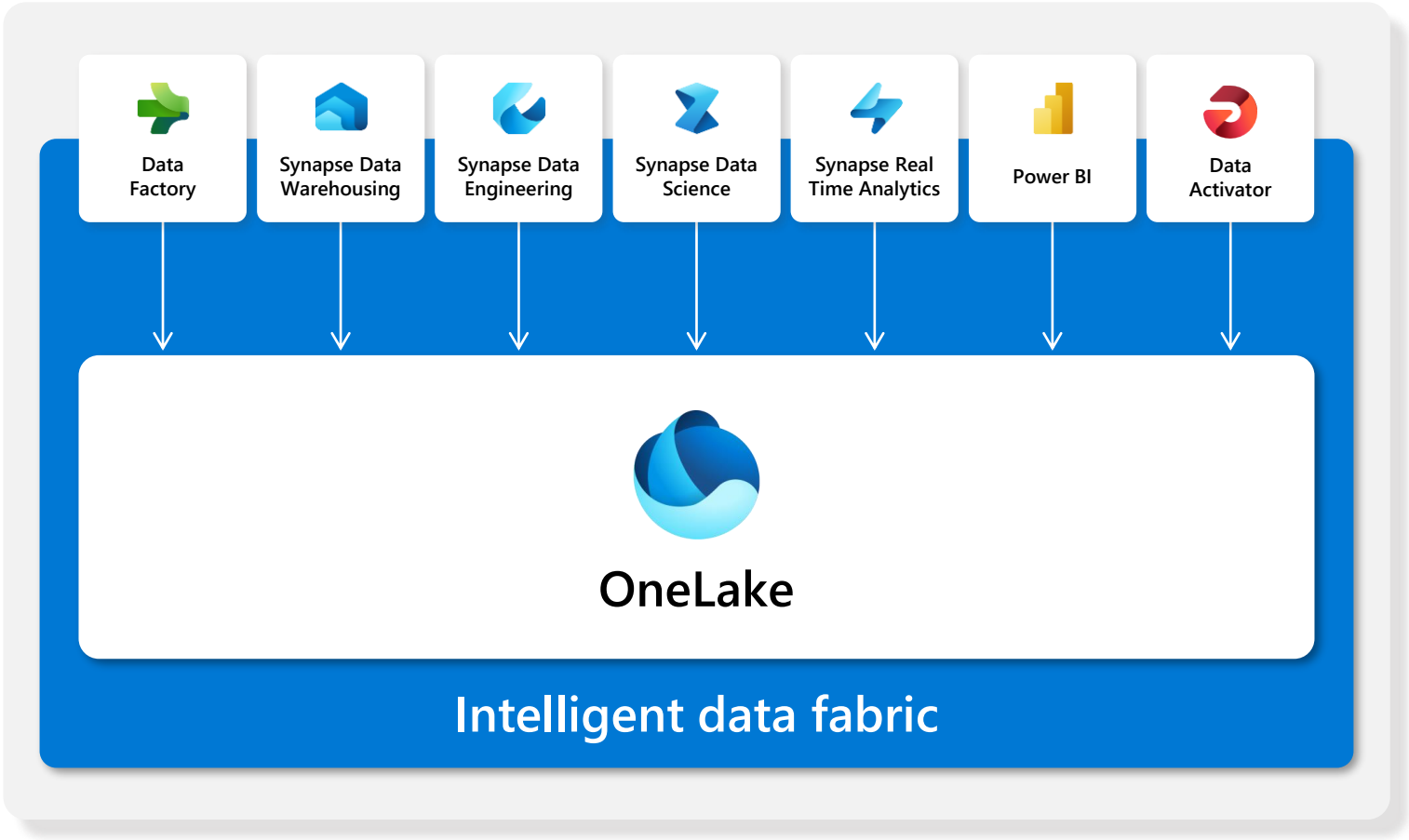
"The OneDrive for data"

OneLake

One Copy

One Security

OneLake Data Hub



Microsoft Fabric Community Resources

Community Call to Action

- ✓ **Try Microsoft Fabric for free: https://aka.ms/try-fabric**
- ✓ **Join the Fabric community: https://aka.ms/fabriccommunity**
- ✓ **Share and vote for ideas to improve Fabric: https://aka.ms/fabricideas**
- ✓ Read and comment our blog: https://aka.ms/fabricblog

Learn More about Microsoft Fabric

- Product announcement: https://aka.ms/fabric
- Digital Event at Build (videos): https://aka.ms/build-with-analytics
- Product website: https://aka.ms/microsoft-fabric
- Documentation: https://aka.ms/fabric-docs
- **Microsoft Learn: https://aka.ms/learn-fabric**
- **End-to-end scenario tutorials: https://aka.ms/fabric-tutorials**
- Fabric e-book: https://aka.ms/fabric-get-started-ebook
- Fabric Notes: https://aka.ms/fabric-notes