



Onelake with Microsoft Fabric, The Data Lake-as-a-Service Platform



Tuesday 01-10-2024

Room 4

Time 11:15-12:15





Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark



@erwindekrek



linkedin.com/in/erwindekrek



erwindekrek.com



github.com/edkreuk



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



We Are InSpark

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



Featured Partner Microsoft Fabric

accenture

Capgemini

EY
Building a better
working world

LTIMindtree

pwc

KPMG

tcs
TATA
CONSULTANCY
SERVICES

avanade

SONATA
SONATA SOFTWARE

MAQ Software

iLink Digital

LANTERN

slalom

3Cloud

bakertilly

READYMIND
NEXT GENERATION SERVICES

Fellowwind

NEODESIC
an IBM Company

OBUNGI

bouvet

INSPARI

preludeSYS
IMPLICIT KNOWLEDGE

ONYX
DATA

Simpson Associates
The Data Analytics Company

INSPARK

QUADRANT

ADASTRA

macaw

twoday capacity

CELEBAL
TECHNOLOGIES

MANDELBULB
TECHNOLOGIES

Cloud Services

SDK

Nortal

ELASTACLOUD

Objectives of today



- Importance of Data
- Onelake
- Shortcuts
- External Data Sharing
- Recap





KONFERENZ

2024

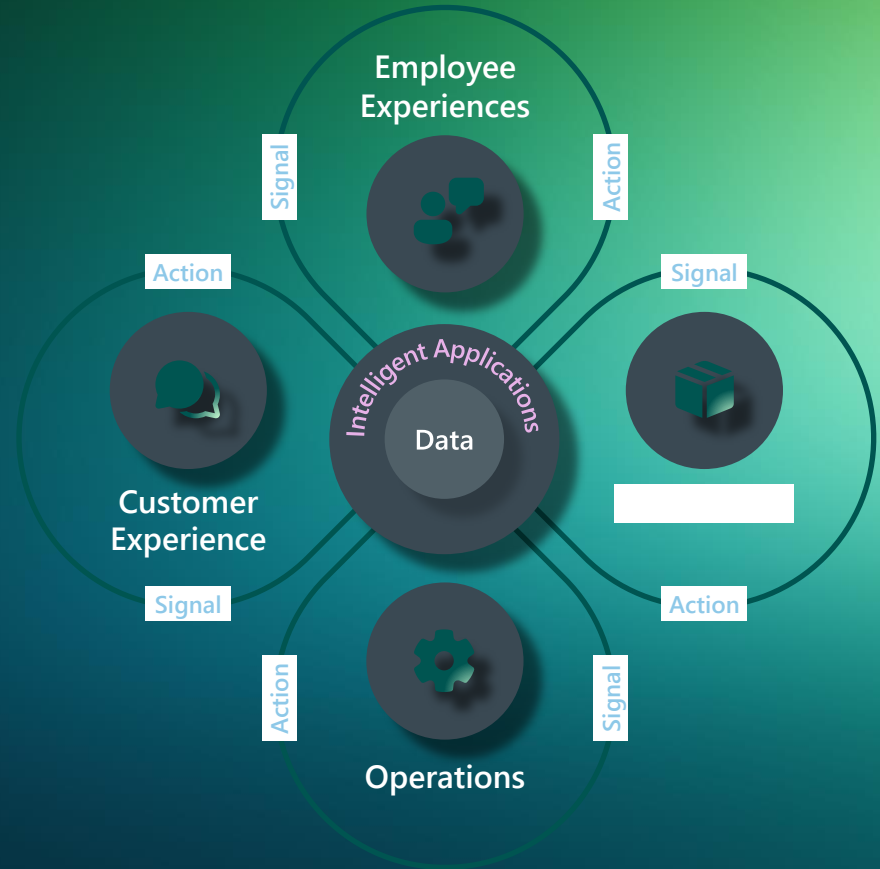
A world awash with data...



Data is the oxygen of digital transformation

"A new kind of company — we call them insights driven businesses — has formed. They are growing at an average of more than 30% annually"

Forrester Analytics Business Technographics Global Data & Analytics Survey





Today's data value creation challenges



Siloed systems and data



Fragmented toolset, Platforms
requiring advanced skills



Costly integration and
ongoing maintenance



Proprietary Platform, Vendor Lock in



Infrastructure not scalable
as business grows

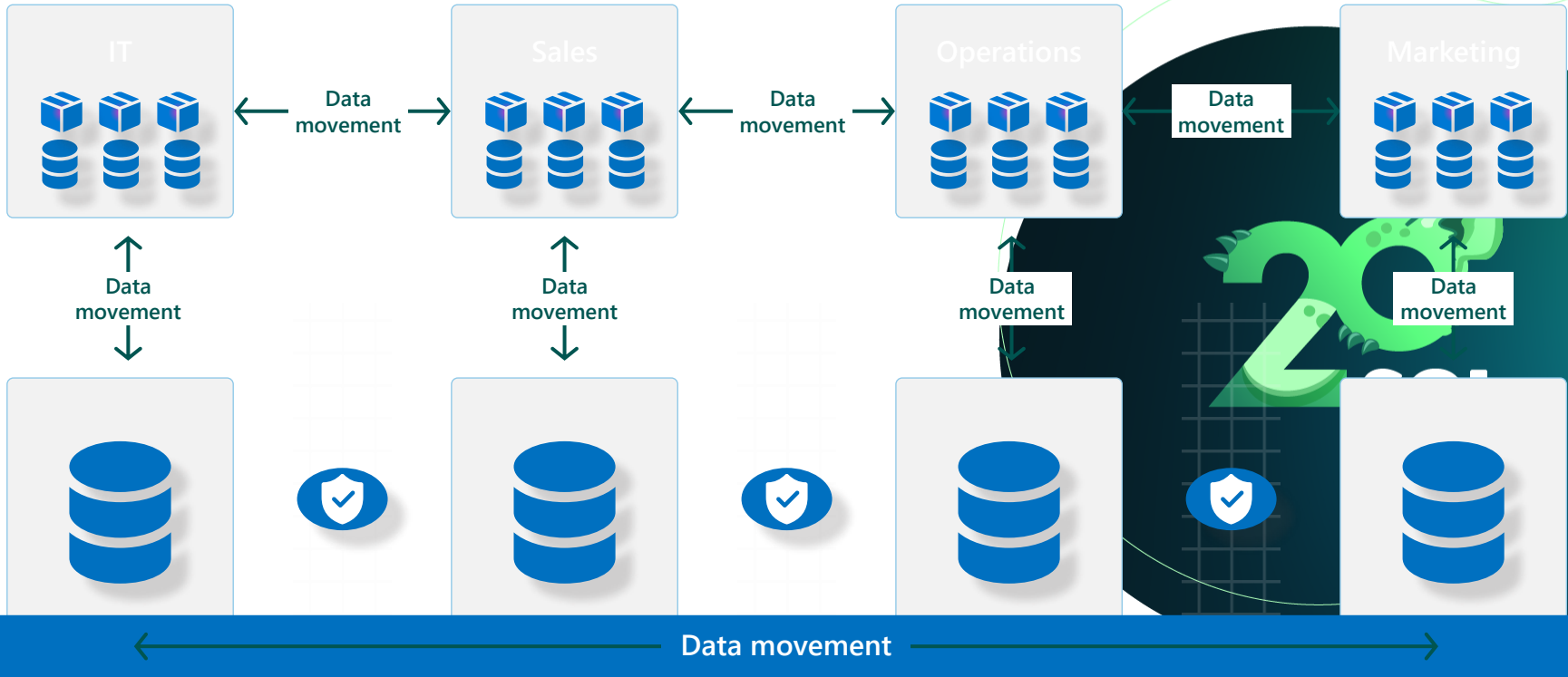


Gaining BI adoption to
streamline data sharing, collaboration



Data and security risks,
Governance & Compliance

Starting line: Multiple siloed lakes with lots of duplication



Typical sentiment about data-backed decision processes

*"Too difficult
to get to the
data"*

*"Data is
highly
fragmented"*

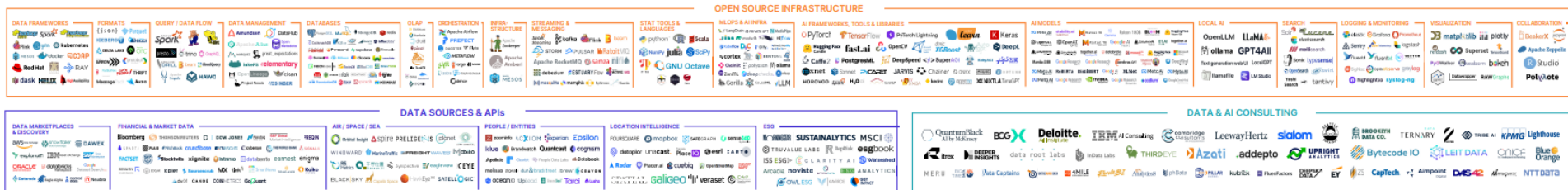
*"We cannot
re-use what
we have"*

*"We do not
have that
expertise"*



How do you translate data into
competitive advantage?

THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE





Unify,

I am the Chief Information Officer
and don't want to be the Chief
Integration Officer.

Help me unleash AI in my data
estate."

Every CIO, Every Enterprise



How many are using Microsoft Fabric?





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

AI
Powered



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

AI Powered

Copilot accelerated

ChatGPT on your data

AI driven insights



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

AI Powered

Copilot accelerated

ChatGPT on your data

AI driven insights



Microsoft Fabric

The unified data platform for the era of AI



Data
Factory



Synapse Data
Engineering



Synapse Data
Science



Synapse Data
Warehousing



Real Time
Intelligence



Power BI



Data
Activator



AI



OneLake



Purview

Unified
architecture

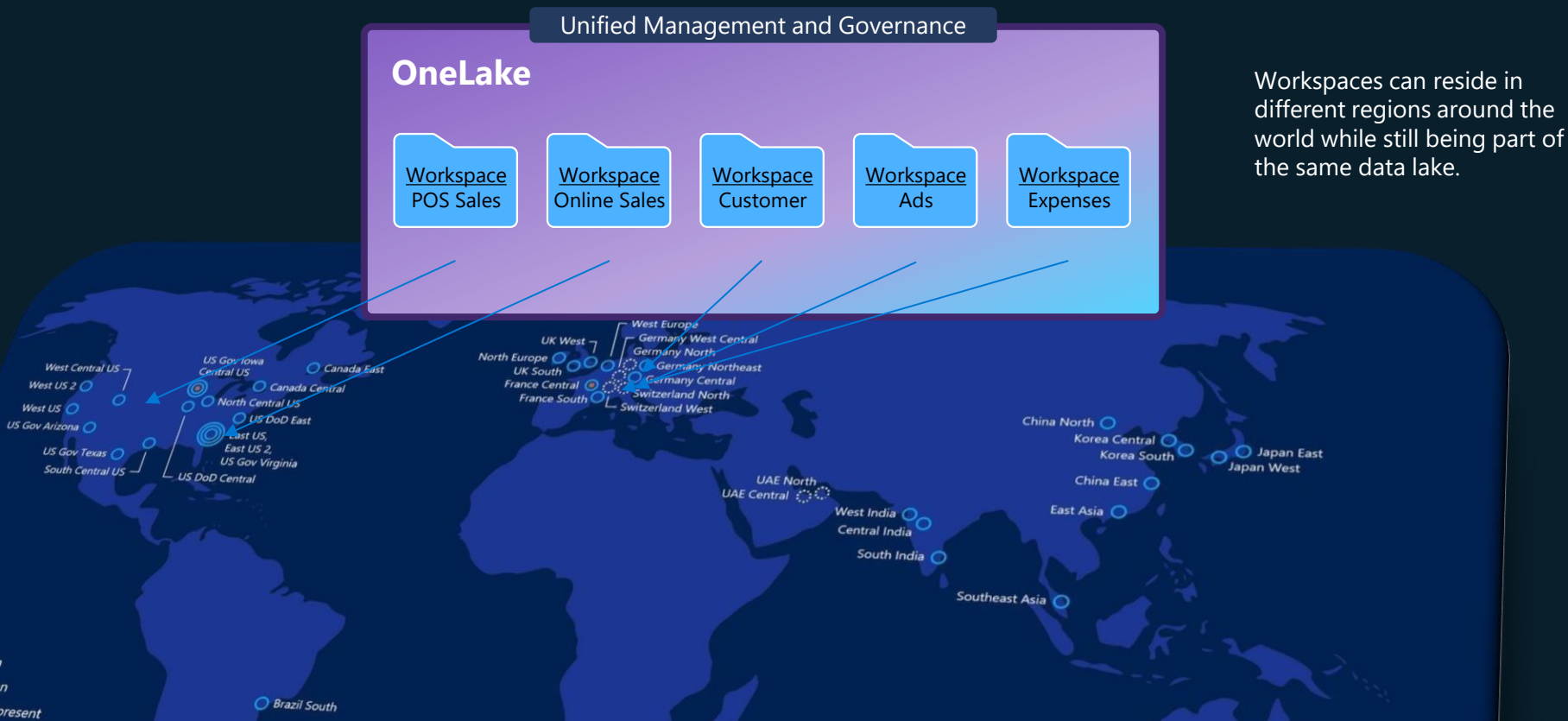
Unified
experience

Unified
governance

Unified
business model

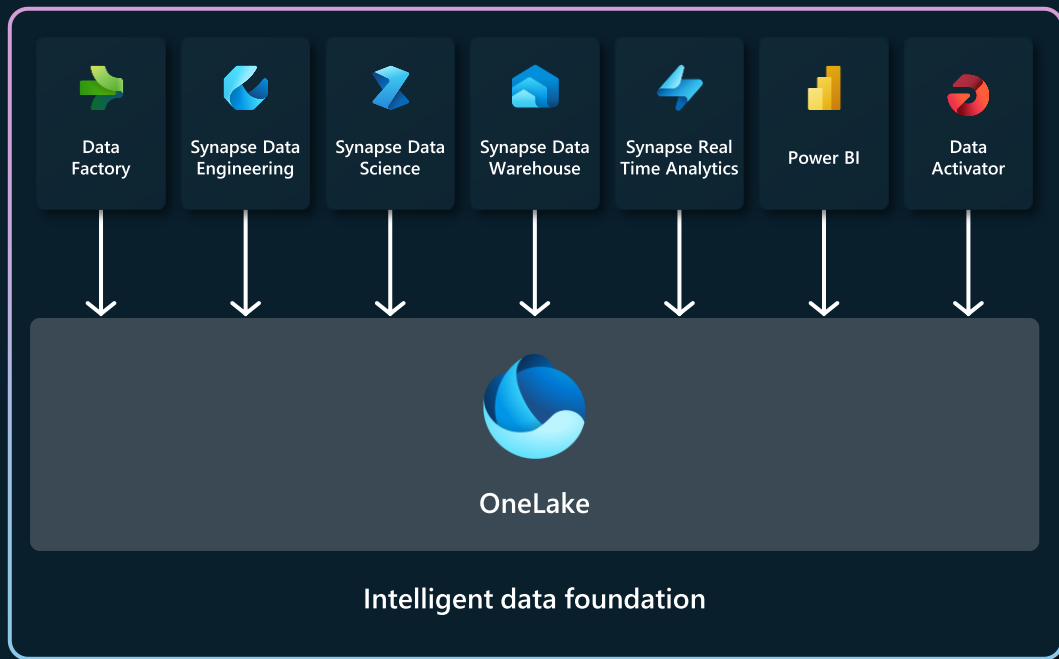
OneLake which logically spans the world

"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

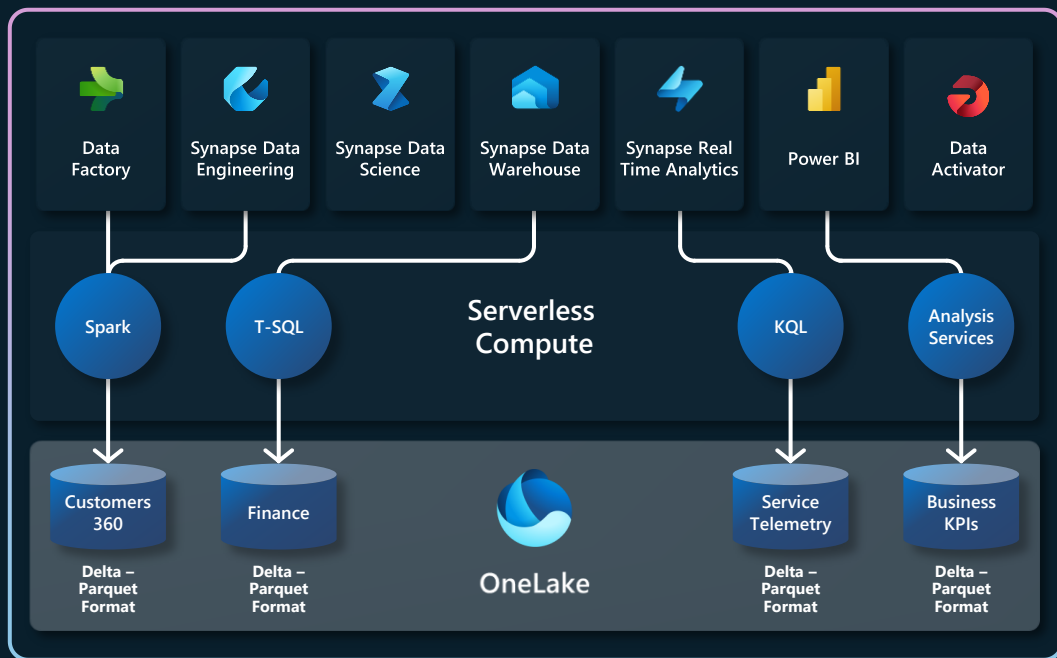
All workloads automatically store their data in the OneLake workspace folders

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 computes

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

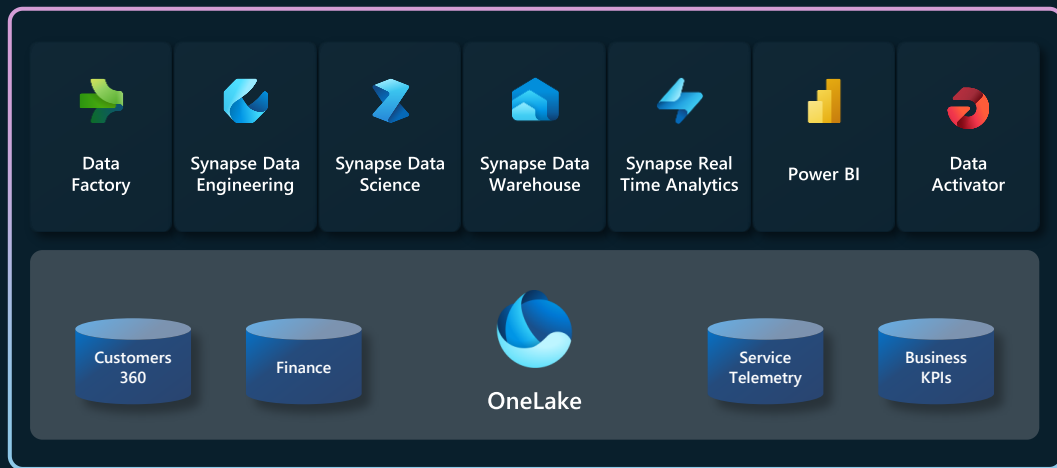
Delta – Parquet, an open standards format, is the storage format for all tabular data in Analytics vNext

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

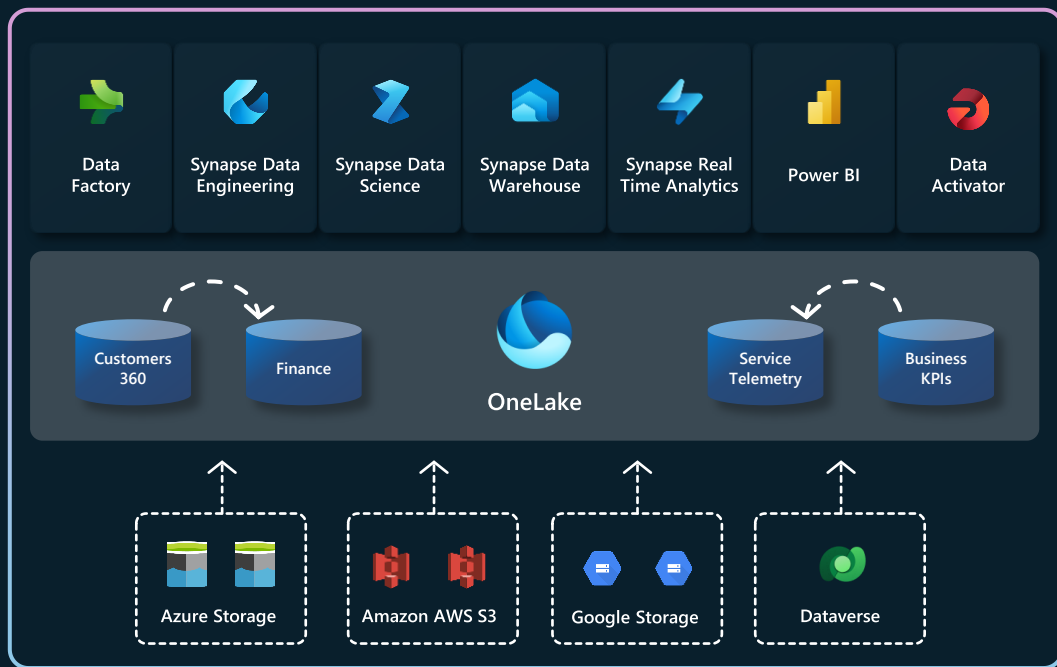
With [shortcuts](#), data throughout OneLake can be composed together without any data movement

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

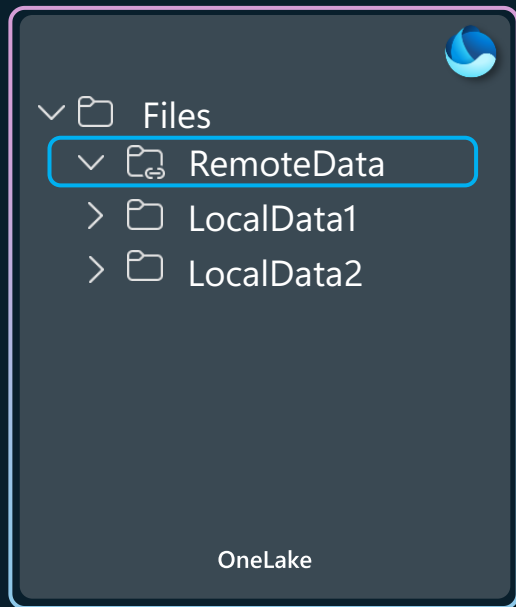
With [shortcuts](#), data throughout OneLake can be composed together without any data movement

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

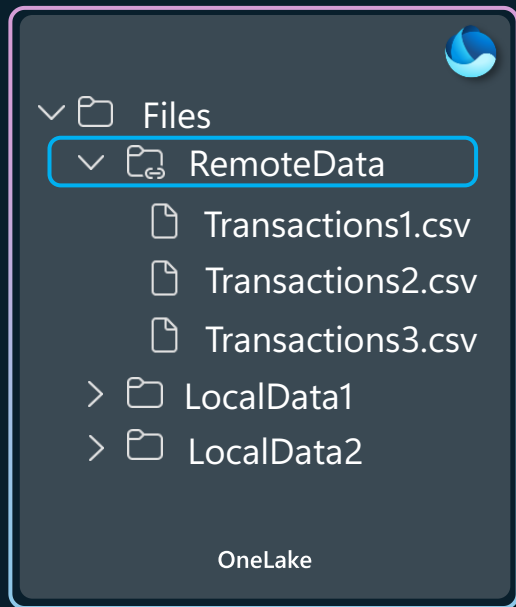
What is a shortcut?

Shortcut Location



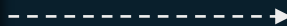
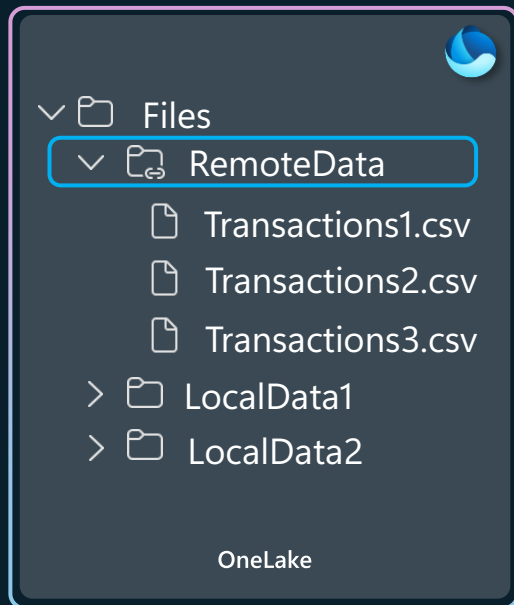
What is a shortcut?

Shortcut Location

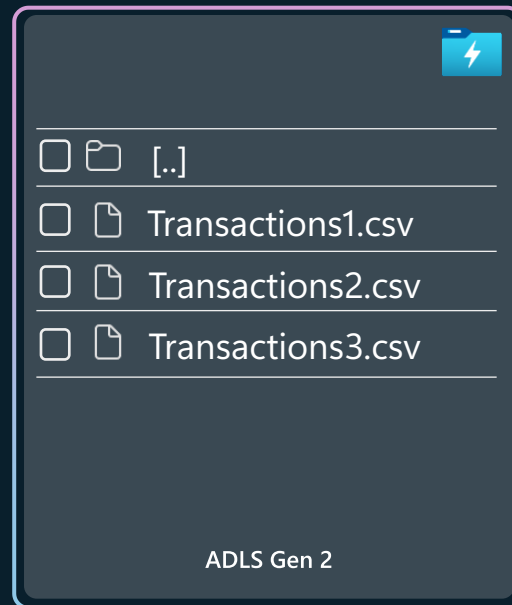


What is a shortcut?

Shortcut Location



Shortcut Location



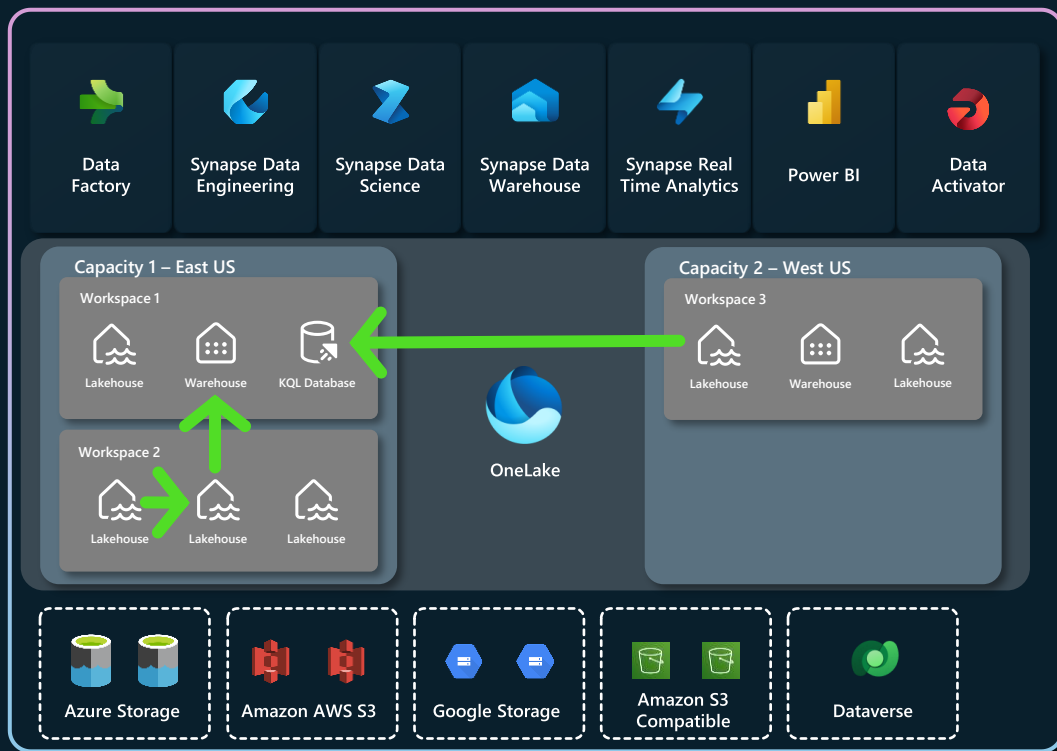


Any questions so far?



Unifying data in OneLake

Shortcuts



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

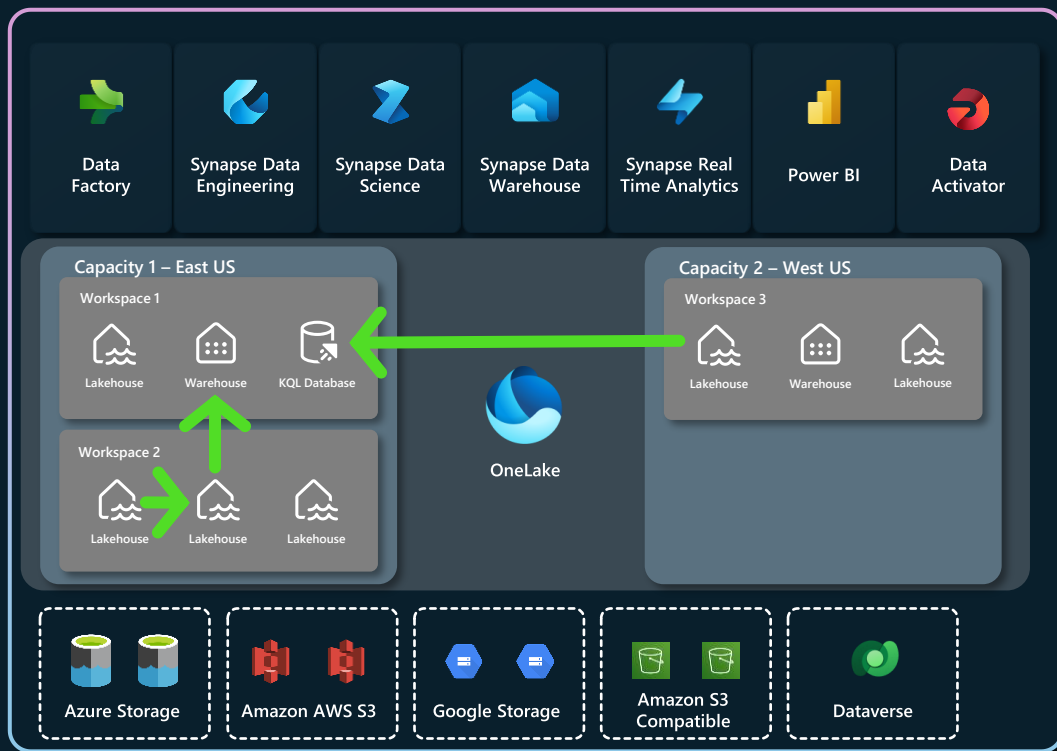
With [shortcuts](#), data throughout OneLake can be composed together without any data movement

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



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

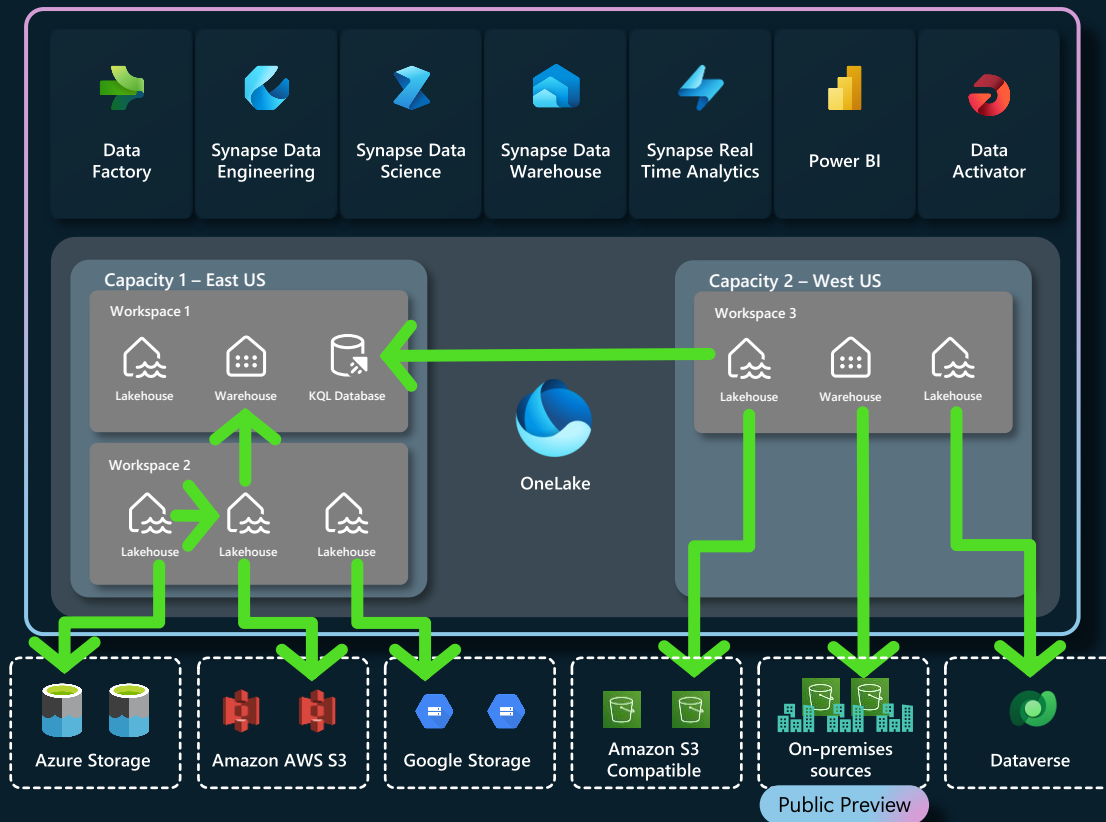
With [shortcuts](#), data throughout OneLake can be composed together without any data movement

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

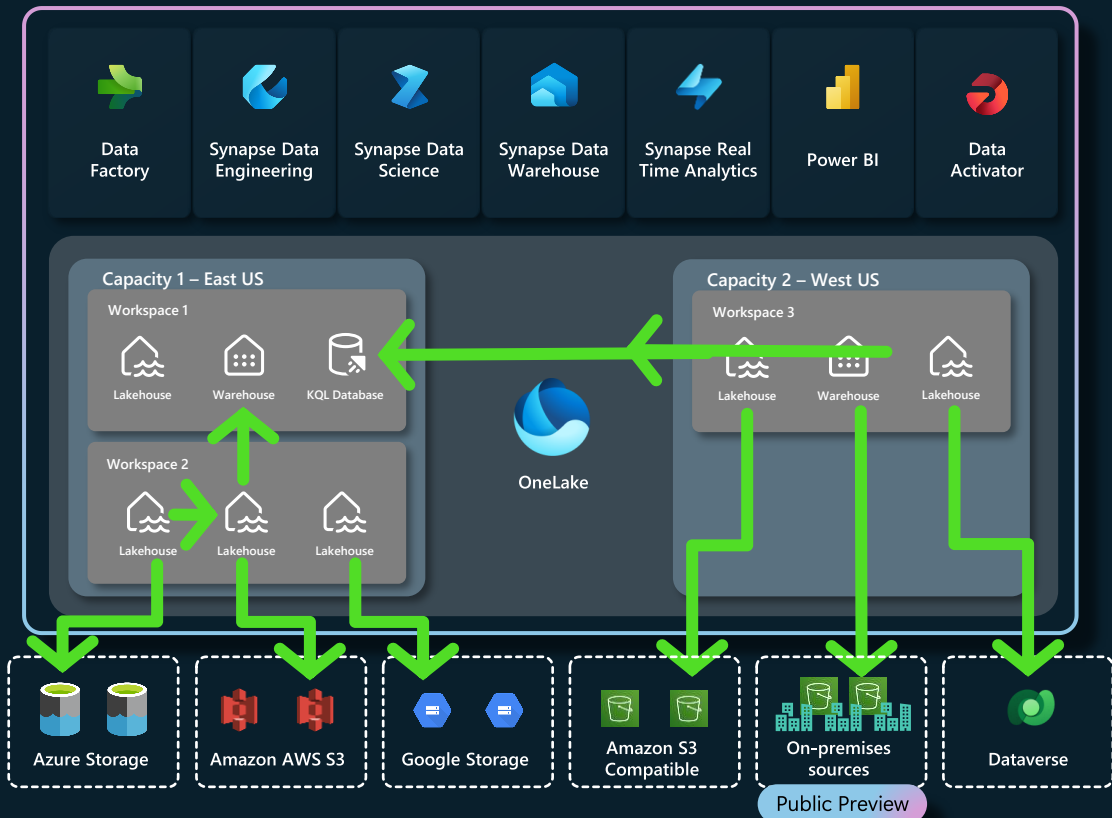
Unifying data in OneLake

Shortcuts



Unifying data in OneLake

External Data Sharing (Preview)



Unifying data in OneLake

External Data Sharing (Preview)

External data sharing (preview)

Enabled for the entire organization

Users can share a read-only link to data stored in OneLake with collaborators outside your organization. When you grant them permission to do so, users can share a link to data in lakehouses and additional Fabric items. Collaborators who receive the link can view, build on, and share the data both within and beyond their own Fabric tenants, using their organization's licenses and capacities. [Learn More](#)

 Enabled

⚠ Usage of shared data within a recipient tenant is not restricted by policies within your tenant, such as Conditional Access, Information Protection, Microsoft Cloud App Security, etc. Shared data may also be processed outside the region where it's stored.

Users can accept external data shares (preview)

Enabled for the entire organization

Users can accept a read-only link to data from another organization's Fabric tenant. Users who accept an external share link can view, build on, and share the data, both inside and outside of your organization's tenant. For more information about the security limitations of this preview feature, view the feature documentation. [Learn More](#)

 Enabled

⚠ Your organization is solely responsible for use of this feature. Potential risks include malicious intent on the part of the data provider, changes in the data at source, and legal liability arising from the accessing of regulated, proprietary, or sensitive data, responsibility for licensing enforcement, or the processing of data across regions.

Mirroring in Fabric

Connect your databases and data warehouses to OneLake seamlessly with Zero ETL (Preview)



Azure Cosmos DB



Mongo DB



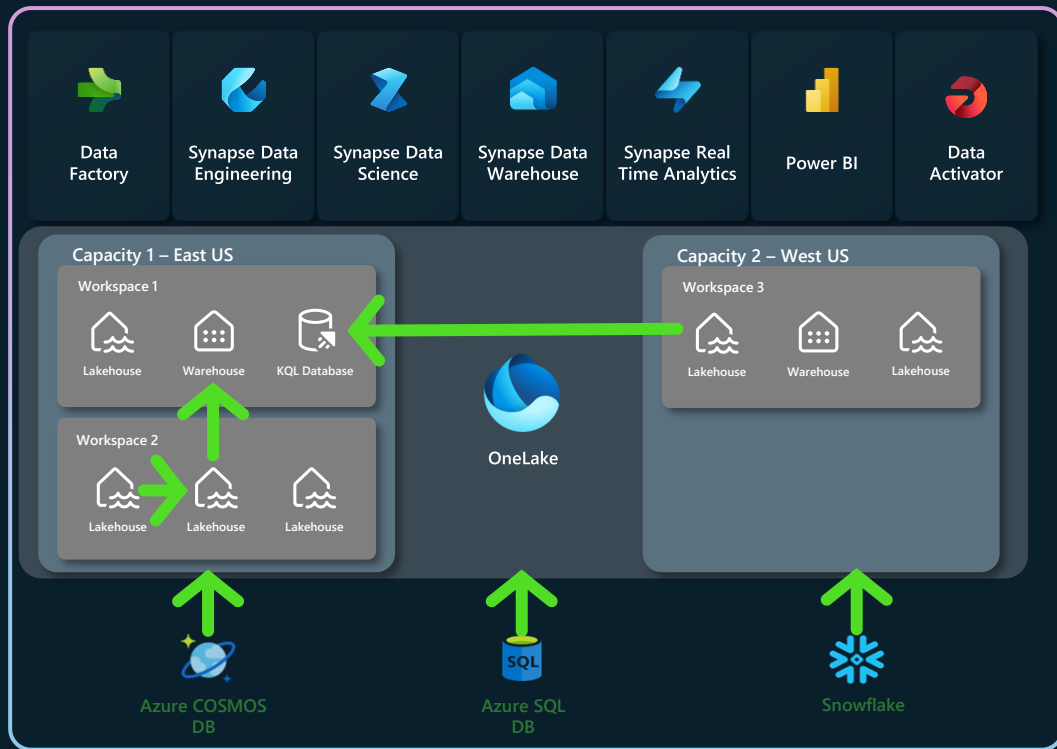
Azure SQL DB



Snowflake

Mirroring in Fabric

Connect your databases and data warehouses to OneLake seamlessly with Zero ETL (Preview)



Fabric Mirroring allows adding existing databases and data warehouses to Fabric without any ETL with a new item type **Mirrored database**

A full live querying experience of the source database is available for the Mirrored database

Data is replicated into OneLake in Delta format and kept up-to-date in near-real-time via change data feed technology

All the Fabric experiences instantly work with the OneLake replica, including data warehousing, Spark, BI experiences

Analysts and Data Scientists can work with real-time data

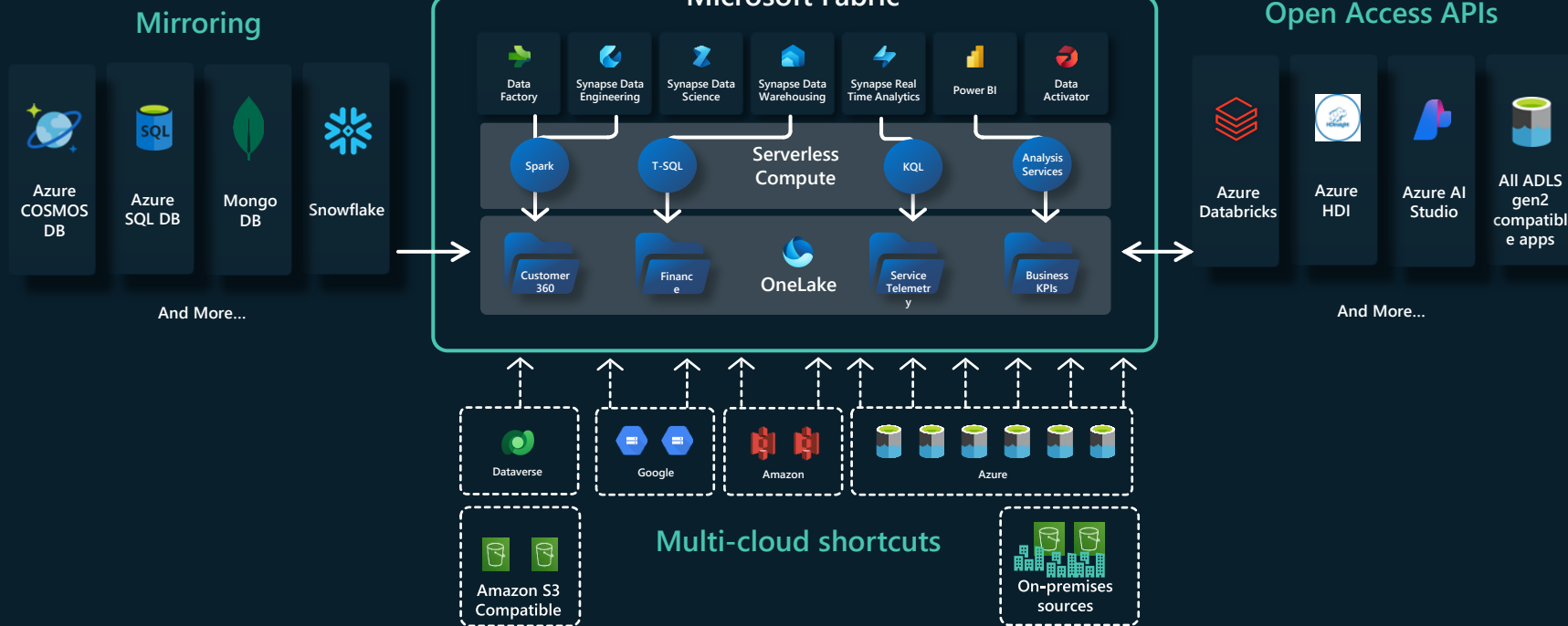
The replica protects operational databases from analytical queries

All roads lead to OneLake

Creating Data Gravity in OneLake

Fabric compute engines

Microsoft Fabric



Public Preview



Innovate to accelerate

DEMO



Any questions left?



FEEDBACK FEEDBACK





Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark



@erwindekrek



linkedin.com/in/erwindekrek



erwindekrek.com



github.com/edkreuk



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

