
Data Tolosa 2023

Stockage :
Comment stocker ses données

Jean-Pierre Riehl
Christophe Laporte

DATA TOLOSA 2023

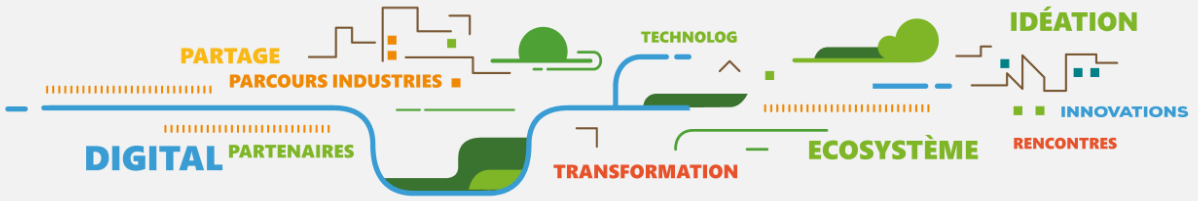
Merci à nos sponsors

SRA
LA PERFORMANCE INTÉGRÉE

SRA Sud-Ouest

Conseil IT

Microsoft
Experiences Lab



Heuu, non !

@conseilit

/christophelaporte

Christophe Laporte

Consultant - formateur

A survécu au changement de millénaire

Eleveur de bases de données relationnelles

SQL Server Certified Master

La tête dans les nuages

Azure Solutions Architect, Azure Database Administrator

Jean-Pierre Riehl

Technology Innovation Lead

avanade

@djeepy1

http://blog.djeepy1.net

MVP Data Platform
Since 2008

Microsoft® Most Valuable Professional

French Data Community **Leader**
Since 2011

GUSS

Club Power BI

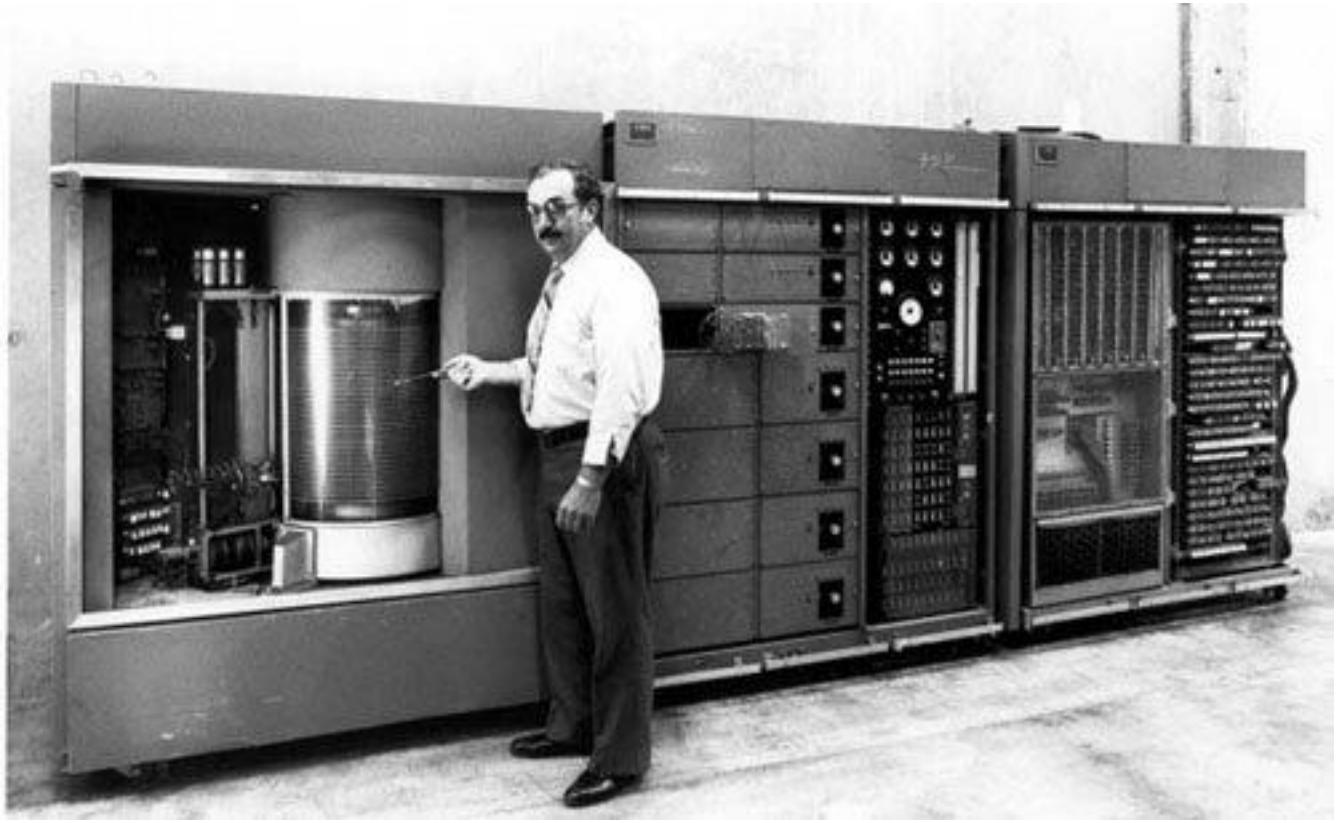
Evolution du stockage

Séquence « vieux con »

1900 -> 1950 – Cartes perforées

1951 – UNIVAC & bandes magnétiques

1956 - RAMAC 305 – 5 Mo



1971 & 1976 : disquettes souples

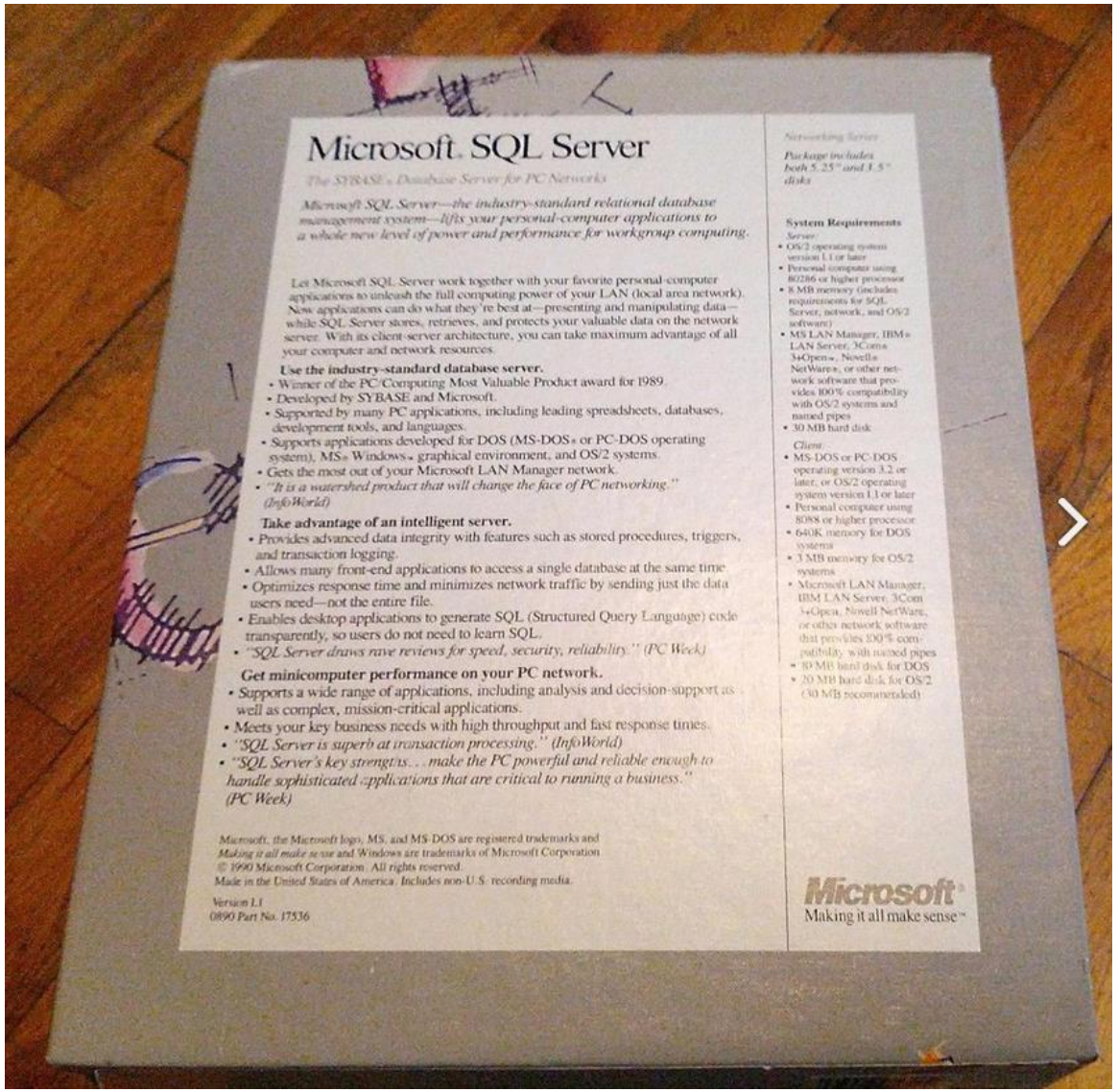
1982 & 1994 : disquettes rigides

Archéologie SQL Server

```
ISQL.EXE
OS/2      Ctrl+Esc = Window List       Type HELP = help
[C:\SQL\BINP]isql /U sa
Password:
1> select @@version
2> GO

 ----------------------------------------
 SQL Server 1.11
        Mon Jul 15 15:14:47 1991

(1 row affected)
1> _
```



Microsoft® SQL Server

The SYBASE-based Database Server for PC Networks

Microsoft SQL Server—the industry-standard relational database management system—lifts your personal-computer applications to a whole new level of power and performance for workgroup computing.

Let Microsoft SQL Server work together with your favorite personal-computer applications to unleash the full computing power of your LAN (local area network). Now applications can do what they're best at—presenting and manipulating data—while SQL Server stores, retrieves, and protects your valuable data on the network server. With its client-server architecture, you can take maximum advantage of all your computer and network resources.

Use the industry-standard database server.

- Winner of the PC Computing Most Valuable Product award for 1989.
- Developed by SYBASE and Microsoft.
- Supported by many PC applications, including leading spreadsheets, databases, development tools, and languages.
- Supports applications developed for DOS (MS-DOS® or PC-DOS operating system), MS® Windows®, graphical environment, and OS/2 systems.
- Gets the most out of your Microsoft LAN Manager network.
- *"It is a watershed product that will change the face of PC networking." (InfoWorld)*

Take advantage of an intelligent server.

- Provides advanced data integrity with features such as stored procedures, triggers, and transaction logging.
- Allows many front-end applications to access a single database at the same time.
- Optimizes response time and minimizes network traffic by sending just the data users need—not the entire file.
- Enables desktop applications to generate SQL (Structured Query Language) code transparently, so users do not need to learn SQL.
- *"SQL Server draws rave reviews for speed, security, reliability." (PC Week)*

Get minicomputer performance on your PC network.

- Supports a wide range of applications, including analysis and decision-support as well as complex, mission-critical applications.
- Meets your key business needs with high throughput and fast response times.
- *"SQL Server is superb at transaction processing." (InfoWorld)*
- *"SQL Server's key strengths . . . make the PC powerful and reliable enough to handle sophisticated applications that are critical to running a business." (PC Week)*

Version 1.1
0890 Part No. 17536

Networking Series

Package includes both 5.25" and 3.5" disks

System Requirements

Server:

- OS/2 operating system version 1.1 or later
- Personal computer using 80286 or higher processor
- 8 MB memory (includes requirements for SQL Server, network, and OS/2 software)
- MS LAN Manager, IBM® LAN Server, 3Com® 3+Open®, Novell® NetWare®, or other network software that provides 100% compatibility with OS/2 systems and named pipes
- 30 MB hard disk

Client:

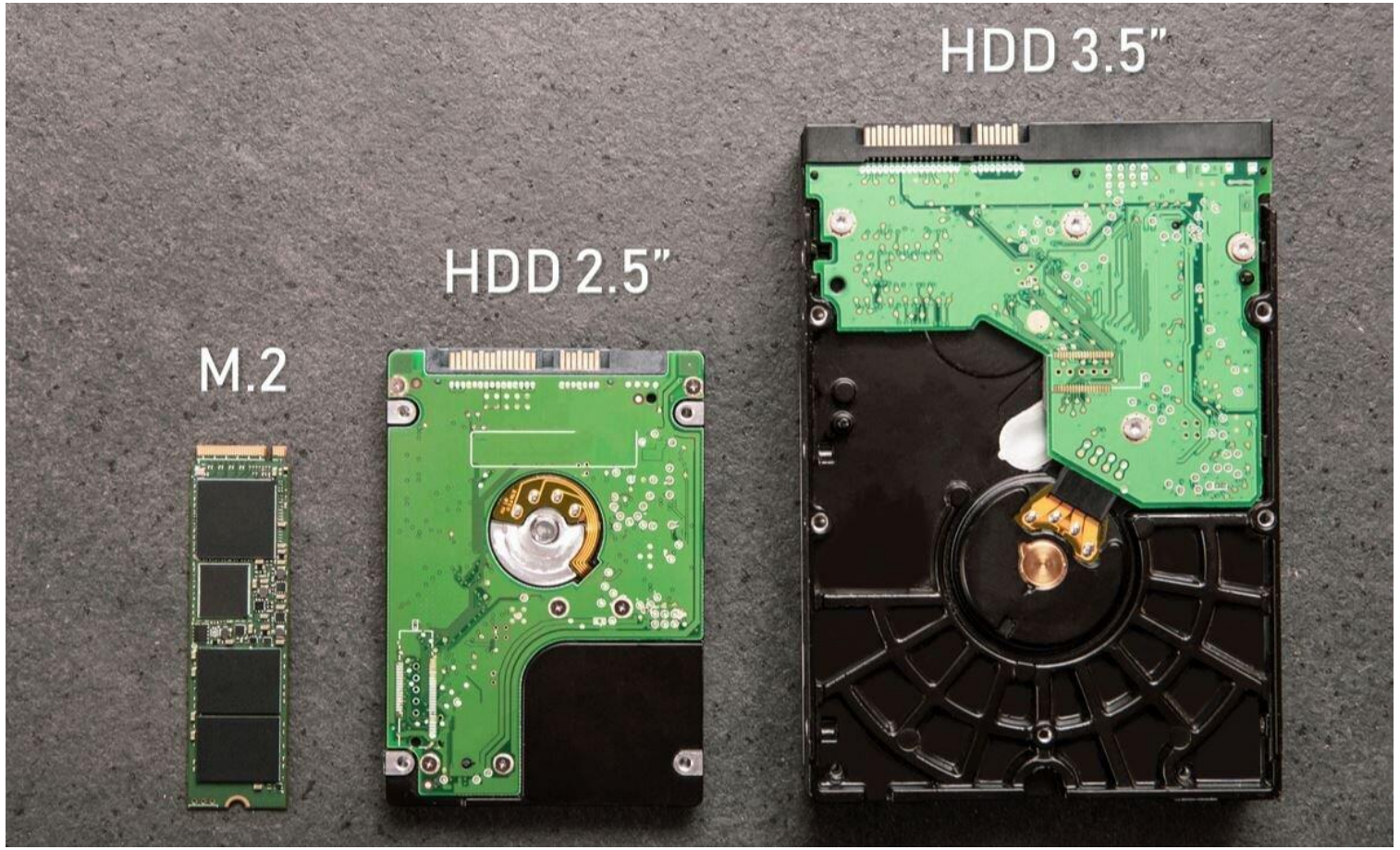
- MS-DOS or PC-DOS operating version 3.2 or later, or OS/2 operating system version 1.1 or later
- Personal computer using 8088 or higher processor
- 640K memory for DOS systems
- 3 MB memory for OS/2 systems
- Microsoft LAN Manager, IBM LAN Server, 3Com 3+Open, Novell NetWare, or other network software that provides 100% compatibility with named pipes
- 10 MB hard disk for DOS
- 20 MB hard disk for OS/2 (30 MB recommended)

Microsoft®
Making it all make sense™

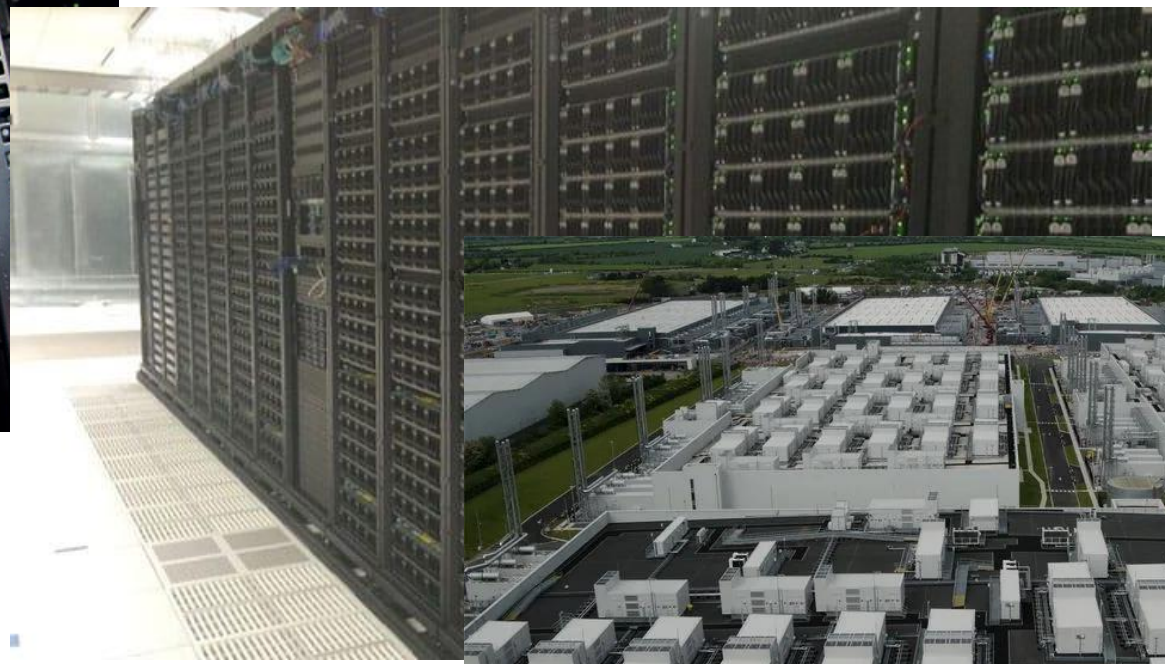
1995 & 1999 & 2001 : mémoire flash

1985 & 1995 & 2006 : disques optiques

Disques durs actuels



Au-delà du disque dur



Au-delà du disque dur

Stockage des données : un challenge

Répondre à plusieurs problématiques

- Limites
  - Besoins en volumétrie …. 3To pour un cerveau humain !
  - 5Mo en 1956 … 18 To aujourd'hui
- Couts
  - Plusieurs solutions pour un même besoin
  - 26 M€/Go en 1956 … 0,02€/Go aujourd'hui
- Gestion
  - Service managé
  - Portabilité
  - Stratégie de sortie

Répondre à plusieurs problématiques

- Disponibilité
  - SLA
  - Haute disponibilité
  - Répartition de charge (géographique ?)
- Latence
  - Ecriture
  - Lecture
- Sauvegarde / Restauration
  - PITR

Répondre à plusieurs problématiques

- Gestion des transactions
  - Atomicité
  - Consistance
  - Isolation
  - Durabilité
- Sécurité
  - Authentification
  - Autorisation
  - Chiffrement
  - Audit

Typologie des données

Données structurées	• Bases de données relationnelles • Bases de données décisionnelles • Table Storage (Key/Value) ?
Données semi structurées	• Fichiers XML • Fichiers JSON • Fichiers YAML
Données non structurées	• Fichiers média • Fichiers Ms Office • Fichiers texte

Les magasins de données

Les magasins de données

Les magasins de données

Les magasins de données

Les magasins de données

Les magasins de données

Data Warehouse

Datalake

LAKEHOUSE VS WAREHOUSE

LAKEHOUSE		WAREHOUSE
Unlimited	Data Volume	Unlimited
Unstructured, semi-structured, structured	Type of data	Structured (Tables, Columns)
Data engineer, Data scientist	Who is using it?	Data warehouse developer, SQL engineer
Spark (Scala, PySpark, Spark SQL, R) Spark notebooks, spark job definitions	Primary skillset	SQL SQL Scripts
Spark & T-SQL (SQL Endpoint for Lakehouse, Limited T-SQL coverage)	How to read data?	Spark & T-SQL (great T-SQL coverage)
✗ No	Multi-Table transactions?	Yes ✓
Yes, files and tables can be sources	Shortcuts support	Yes – access data through lakehouse, Tables can be sources
Across lakehouse and warehouse tables & across lakehouses	Query across items	Across lakehouse and warehouse tables

@cmaneu

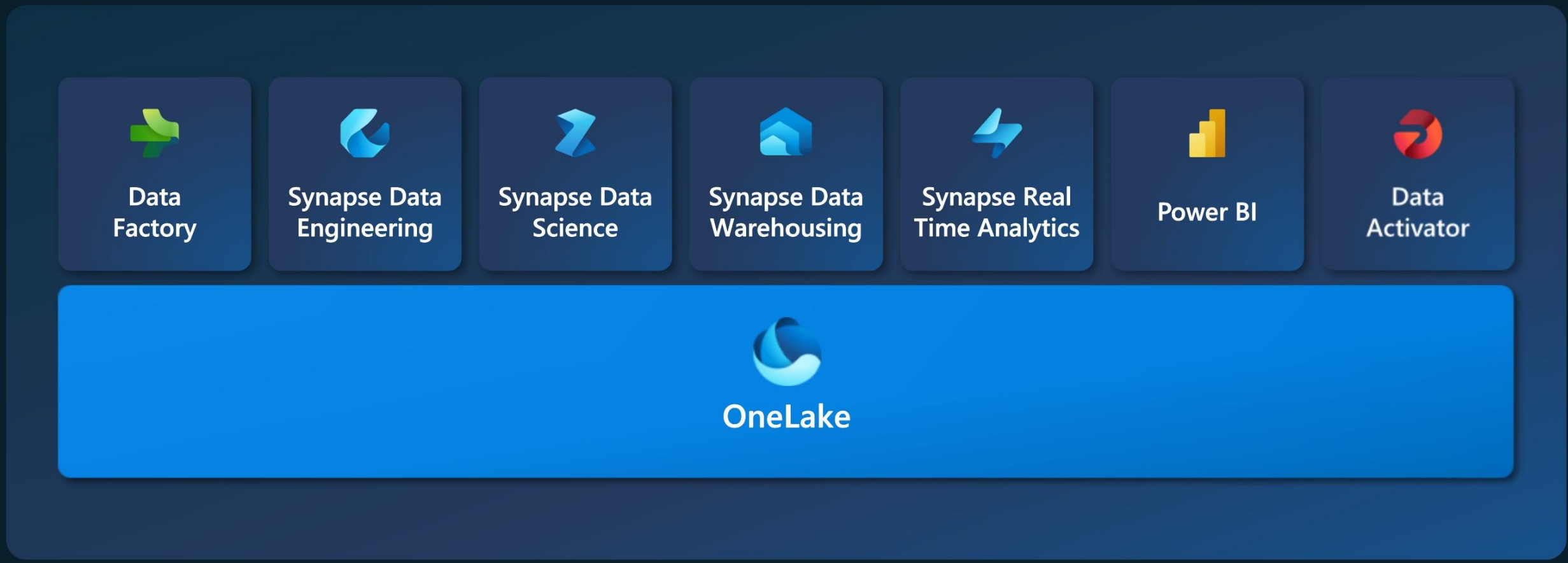
Microsoft Fabric

"It's perhaps the biggest launch of a data product from Microsoft since the launch of SQL Server."

Satya Nadella

Microsoft Fabric

The data platform for the era of AI



Microsoft Fabric

Data analytics for the era of AI

Complete Analytics Platform

Lake Centric and Open

Empower Every Business User

AI Powered

Microsoft Fabric

Data analytics 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

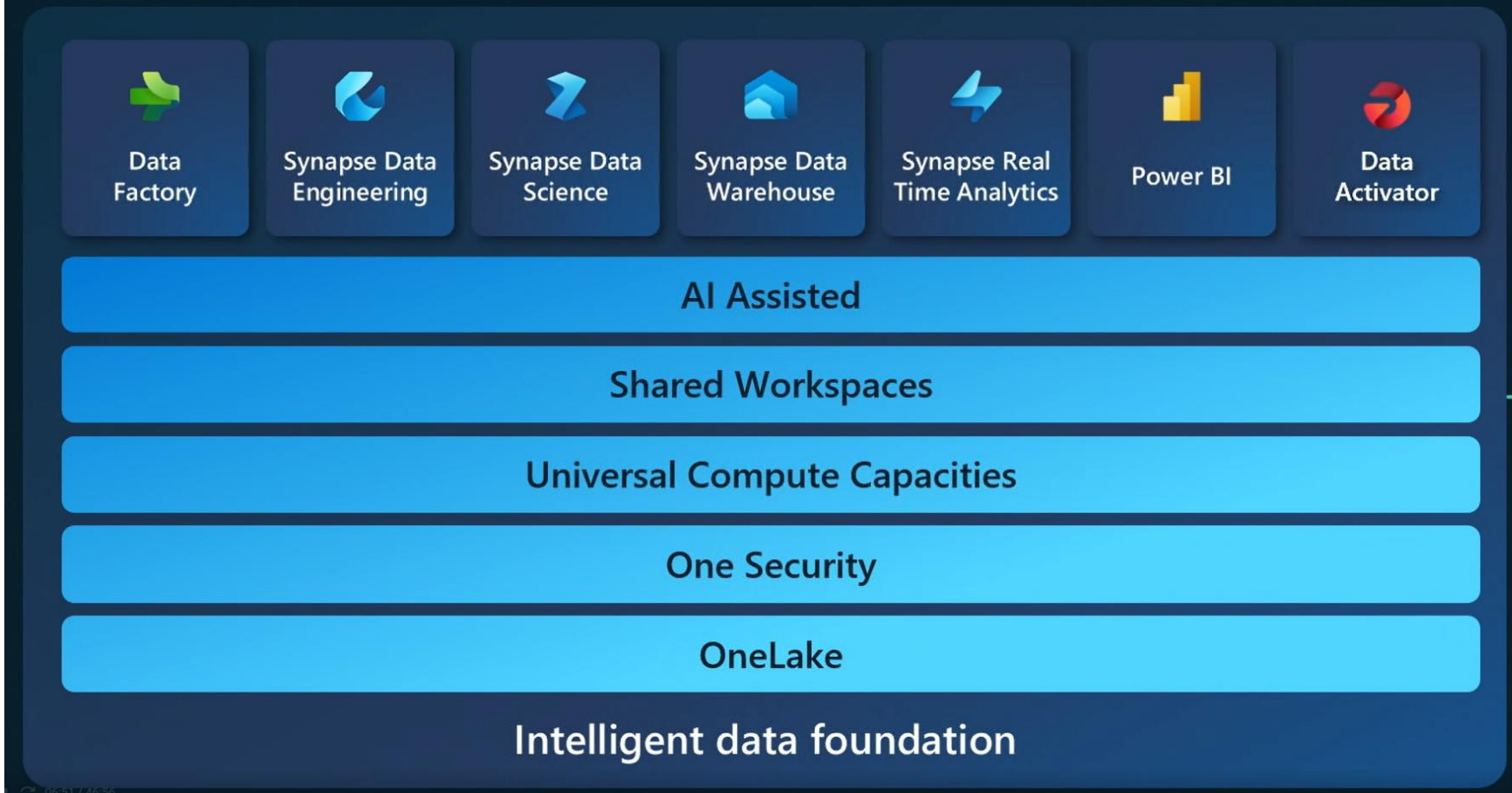
Empower Every Business User

- Familiar and intuitive
- Built into Microsoft 365
- Insight to action

AI Powered

- Copilot accelerated
- GPT on your data
- AI-driven insights

Microsoft Fabric

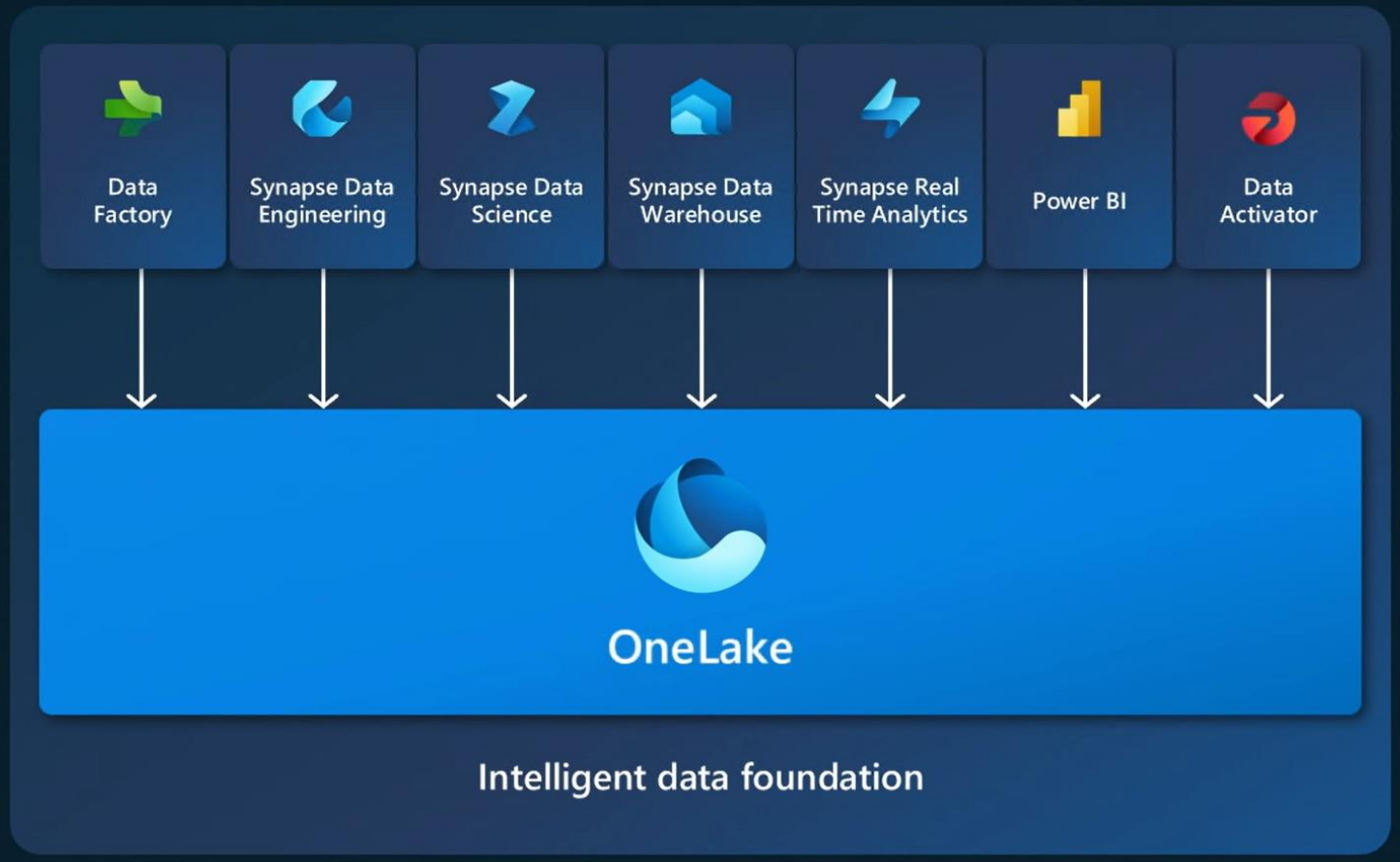


Single…

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

OneLake for all Data

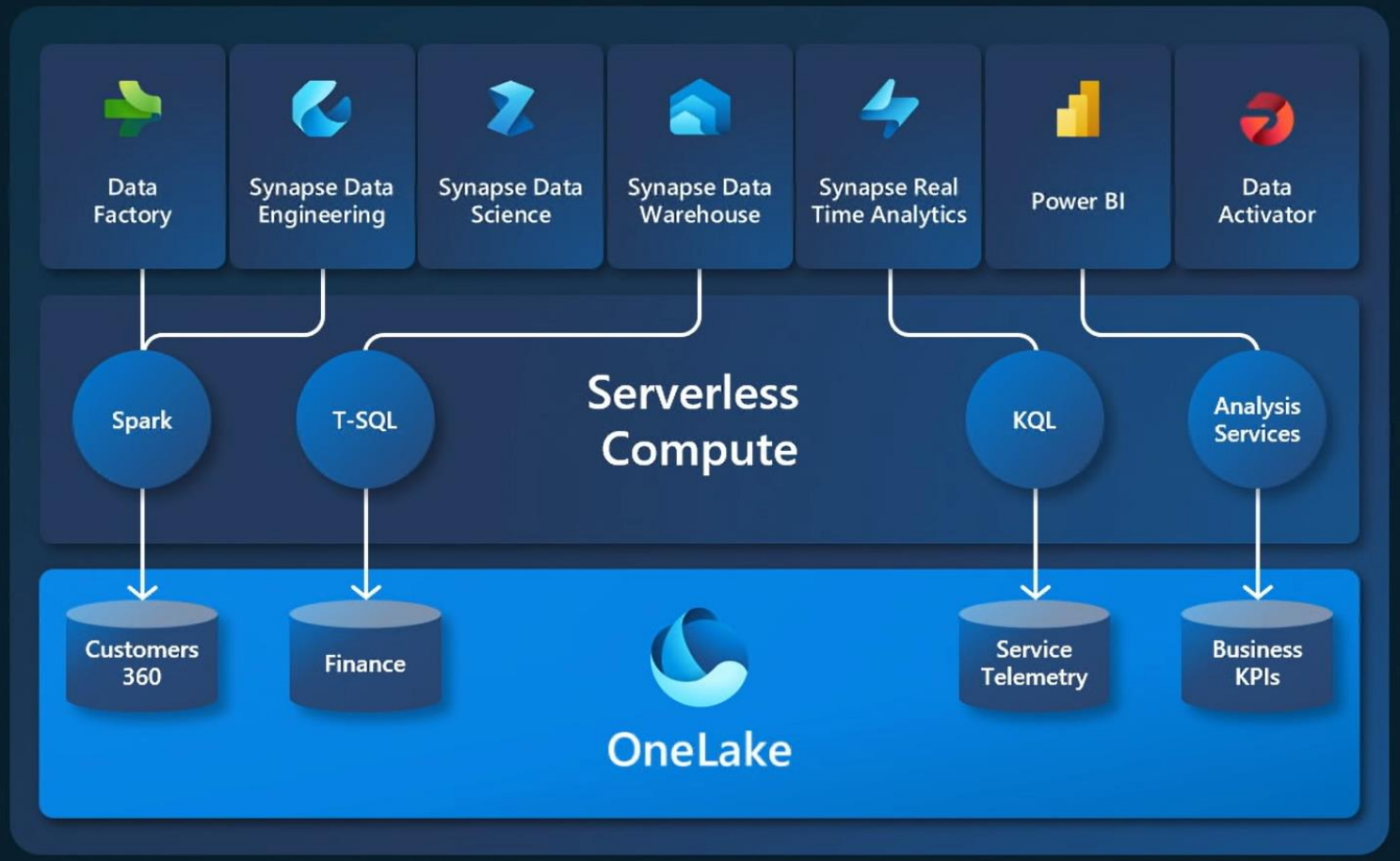
"The OneDrive for Data"



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

One Copy for all computes

Universal security makes it real

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

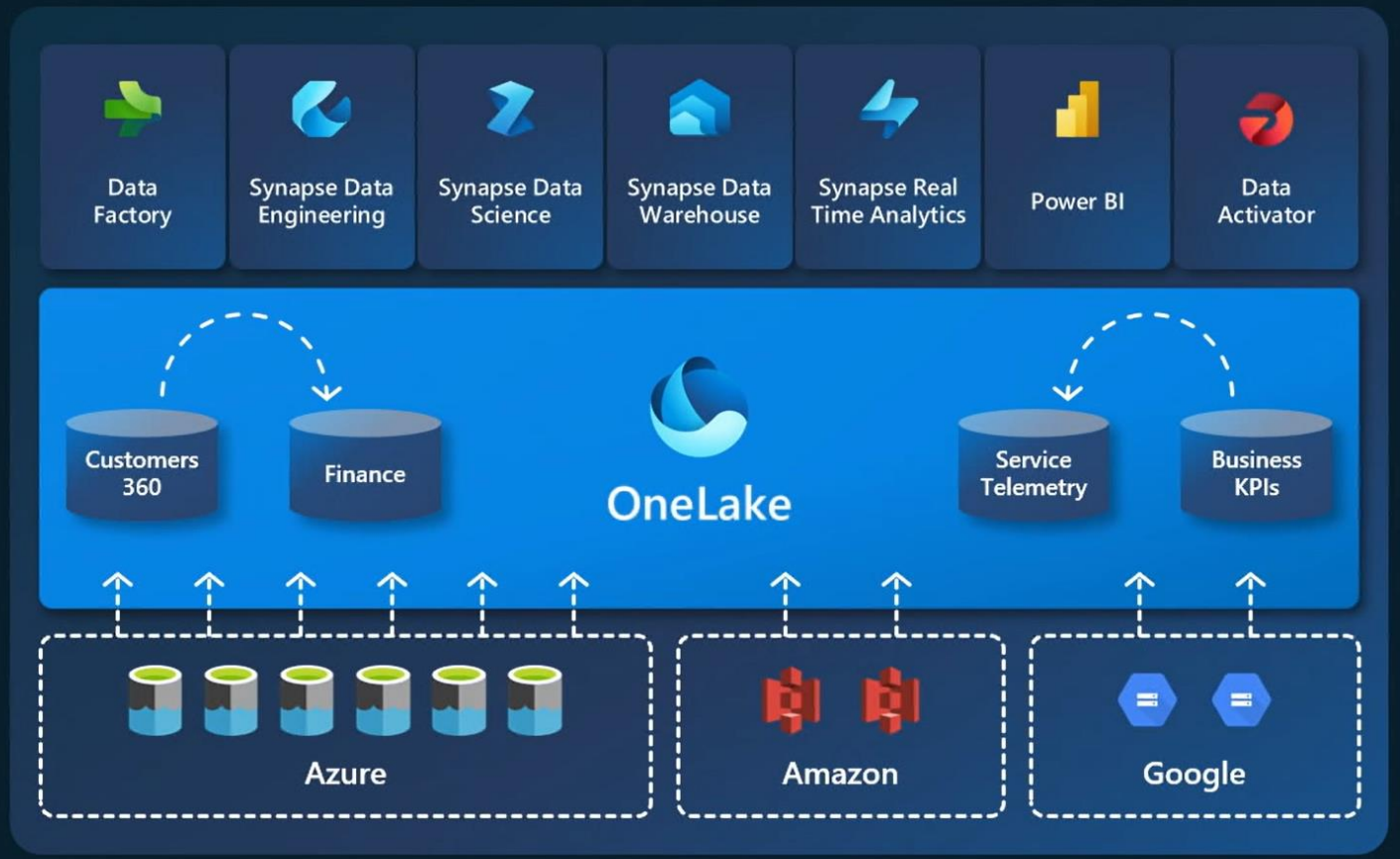
Once data is stored in the lake, it is directly accessible by all the engines without needing any import/export

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

Shared universal security model is enforced across all the engines

Taking One Copy to the Next Level

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 a multi-cloud data lake

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

Introducing Data Activator