

DATA PLATFORM WITH MICROSOFT SYNAPSE



What we believe should be the goal?

Develop a data architecture that is robust, powerful, scalable, and developable enough to meet future expectations without the need to "redevelop" it again in the short/medium term.



Instant data access

We need a Data Platform that supports instant access to data, agile, with adequate response to the claims of analytical users in order to promote its use, extend the data culture in the organization and respond to the needs of analysis and decision making.



Unique and shared information

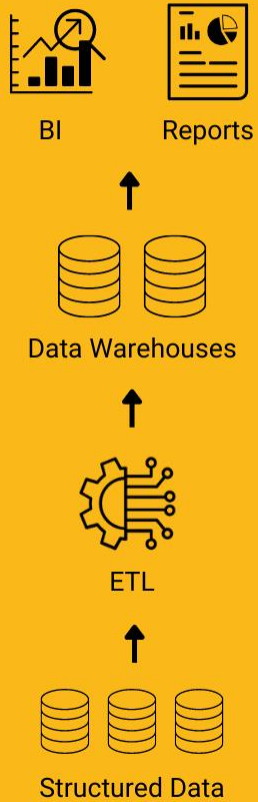
Sufficiently broad to be extensible to the entire organization and to all current and potential future information needs. The information, of course, must be secure, governable and governed as well as transversal to all the processes of the organization.



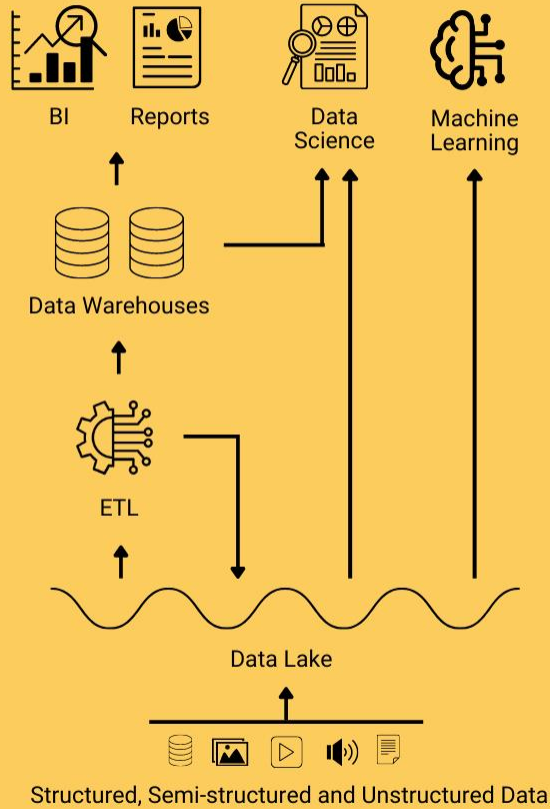
Scalable and developable without rework

Modern, with the possibility of building piece by piece, in a phased way so as not to spend unnecessary iterations, covering all the needs of analysis, AI and Machine Learning not only current but also future.

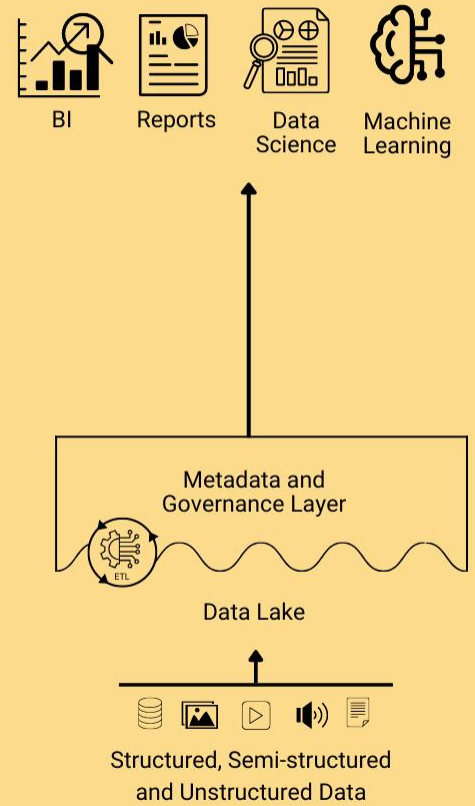
Data Warehouse



Data Lake

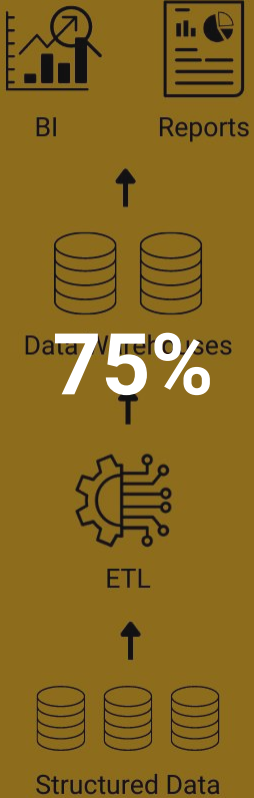


Data Lakehouse



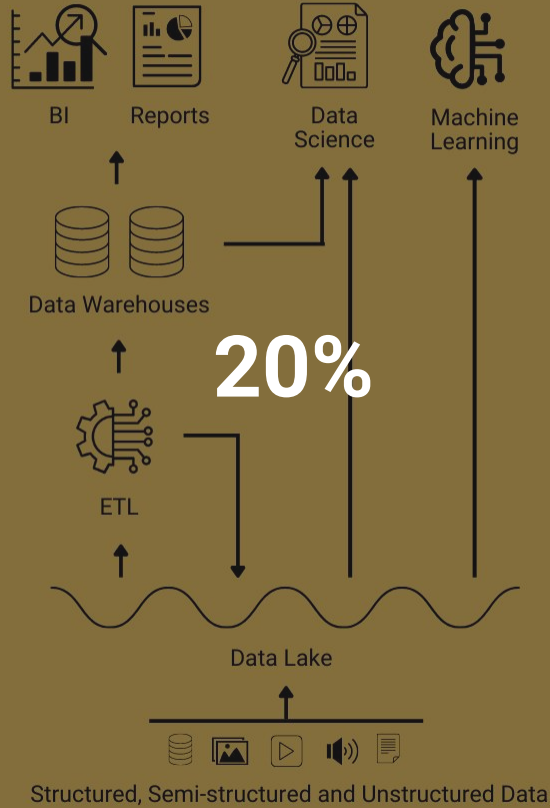
¿What is the current situation?

Data Warehouse



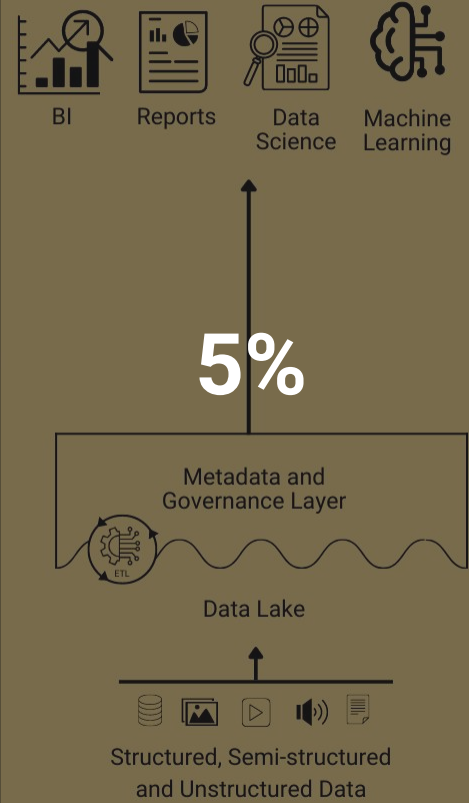
75%

Data Lake



20%

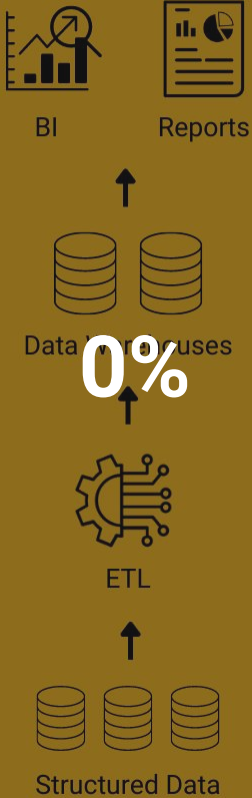
Data Lakehouse



5%

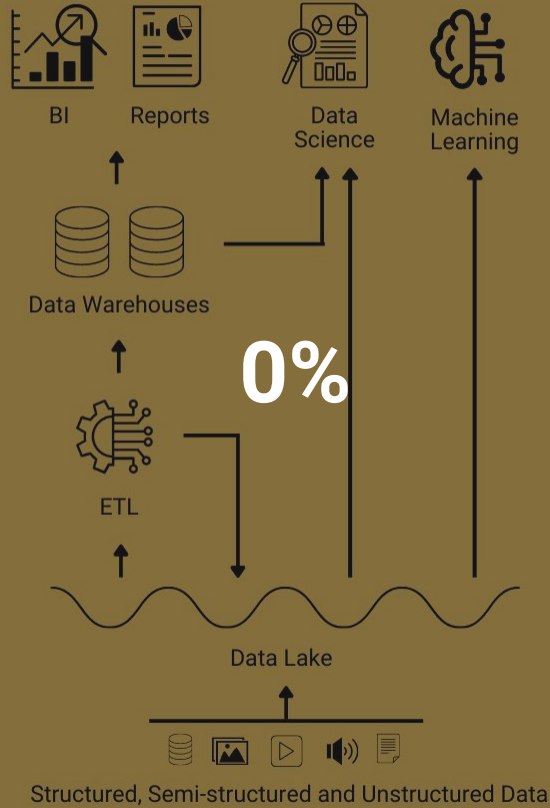
¿ What we want to achieve?

Data Warehouse



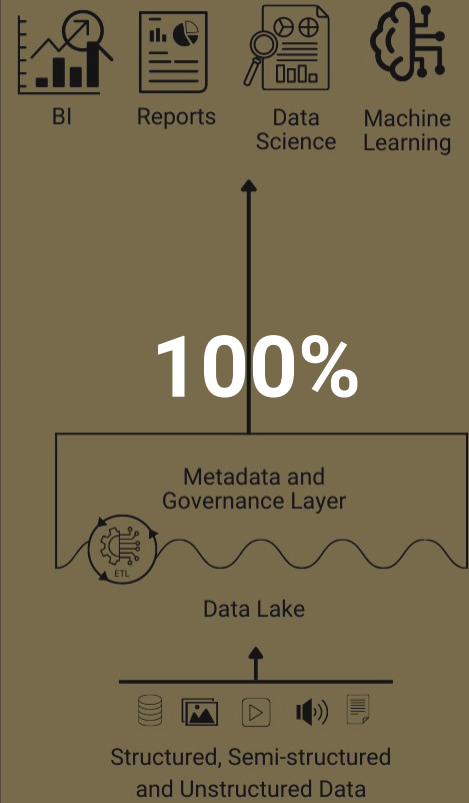
0%

Data Lake



0%

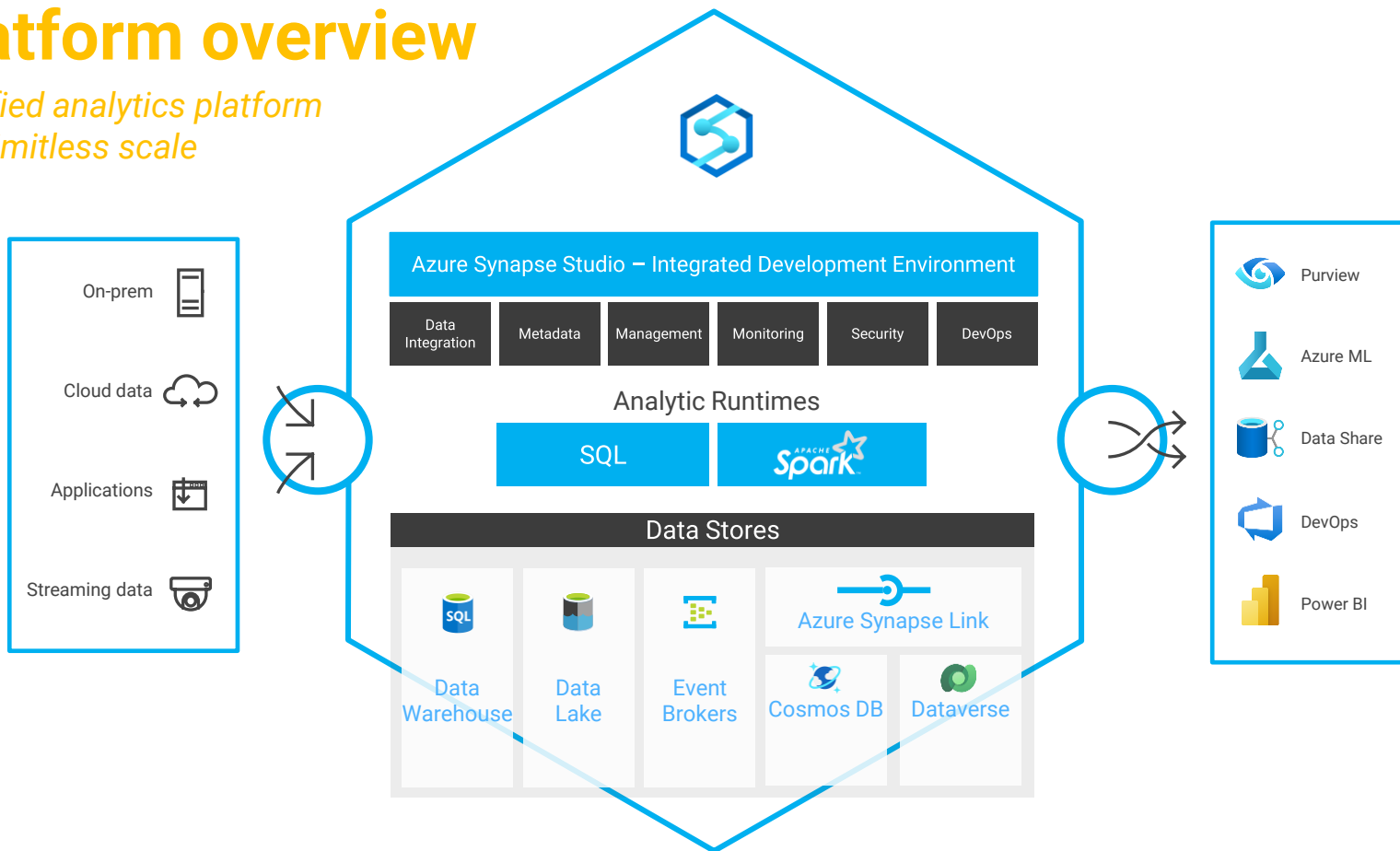
Data Lakehouse



100%

Platform overview

*A unified analytics platform
with limitless scale*



Synapse - Integration

Services and applications:

 Dynamics 365	 Dynamics AX	 Dynamics CRM	 PayPal (vista previa)	 QuickBooks (vista previa)	 REST	 Servicio web de Amazon Marketplace	 Shopify (vista previa)	 Snowflake
 GitHub	 HubSpot	 Jira	 SAP ECC	 Salesforce	 ServiceNow	 Tabla web	 Xero	 Zoho (vista previa)

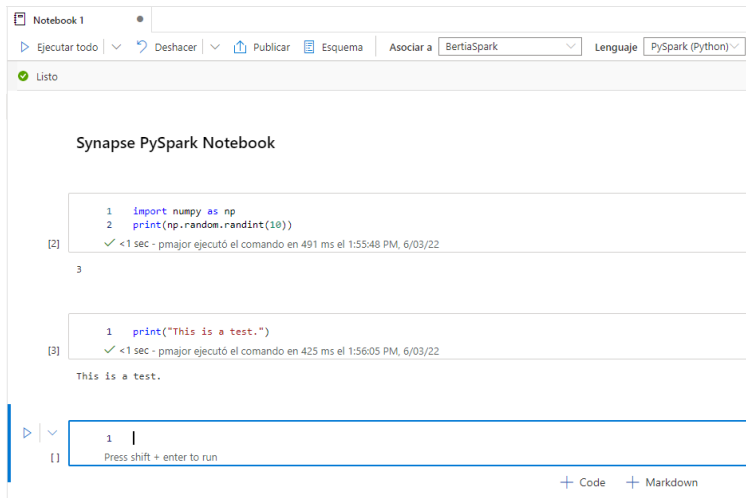
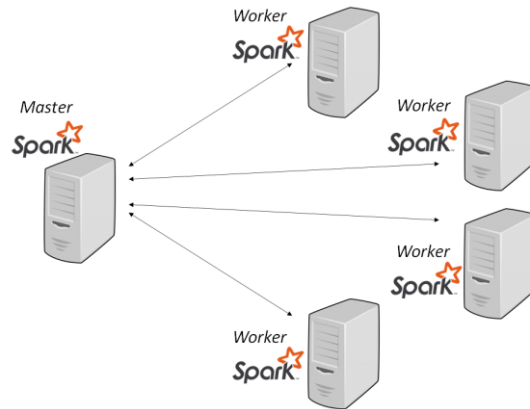
Storage systems:

 Almacenamiento de blobs de Azure	 Almacenamiento de tablas de Azure	 Azure Cognitive Services	 Amazon RDS para Oracle	 Amazon RDS para SQL Server	 Amazon Redshift	 Hive	 Informix	 Instancia administrada de Azure SQL Database	 Presto	 SAP HANA	 SAP BW mediante MDX
 Azure Cosmos DB (API de MongoDB)	 Azure Cosmos DB (API de SQL)	 Azure Data Explorer (Kusto)	 Apache Impala	 Centro abierto de SAP BW	 DB2	 MariaDB	 Microsoft Access	 MySQL	 Spark	 Sybase	 Tabla de SAP
 Azure Database for MariaDB	 Azure Database for MySQL	 Azure Database for PostgreSQL	 Explorator	 Fénix	 Google AdWords	 Netezza	 Oracle	 PostgreSQL	 Teradata	 Vertica	 servidor SQL Server

Synapse – Apache Spark

What it is and how it works?

- An Apache Spark cluster is a group of virtual machines working together.
- We can add or remove cores (machines) from the cluster on demand, according to the current need.
- It is pay-as-you-go, so we will never overpay.
- Apache Spark for Synapse allows programming in multiple languages:
 - Python
 - Scala
 - .Net (C#)
 - SQL
- It is programmed in Notebooks within the Synapse Workspace.



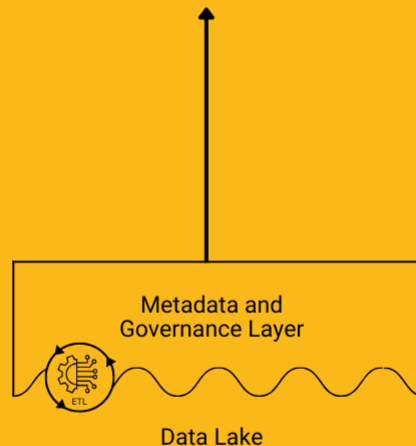
Synapse – DELTA LAKE

The **Delta** format is a key piece when it comes to the **Data Lakehouse**.

It offers several advantages :

- **Concurrency control** through ACID transactions (Atomicity, Consistency, Isolation, Durability).
- **History of changes.** A versioning of the data is saved that even allows you to revert changes or read previous versions of the data.
- **Deletes y Upserts.** Allows you to delete and update data just as you would in a traditional DWH.
- Possibility of **incremental data loading**, avoiding unnecessary data reprocessing and saving a lot of costs.

Data Lakehouse

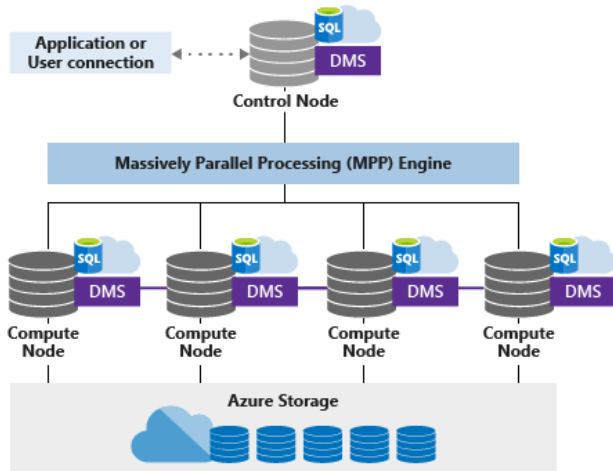


Structured, Semi-structured
and Unstructured Data

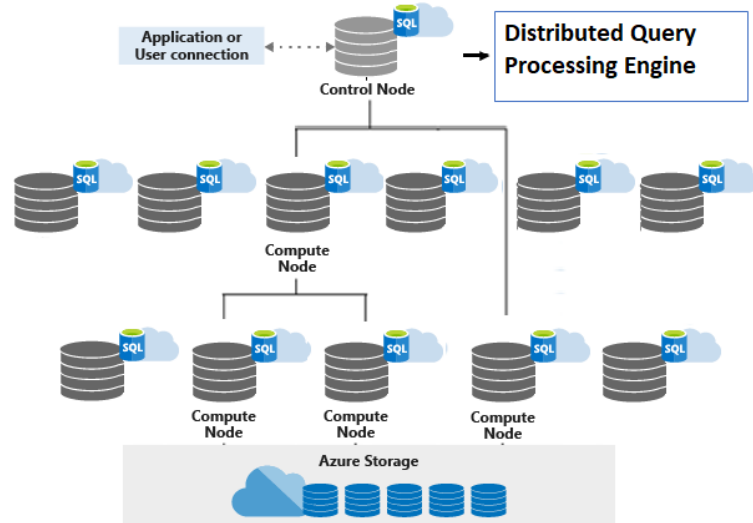
Synapse – SQL Pools

- Previously known as Azure SQL Data Warehouse.
 - Payment per defined DWU (computing units).
 - Allows to ingest and consult data.
- It scales automatically and is paid per amount of data consulted (in TB). Only allows querying.

Dedicated SQL pool



Serverless SQL pool



Synapse – Power BI Integration



Synapse is fully integrated with Power BI.

For example, we can read data directly from Datalake by querying views with SQL Serverless.

**Thank you for
your attention**

