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.

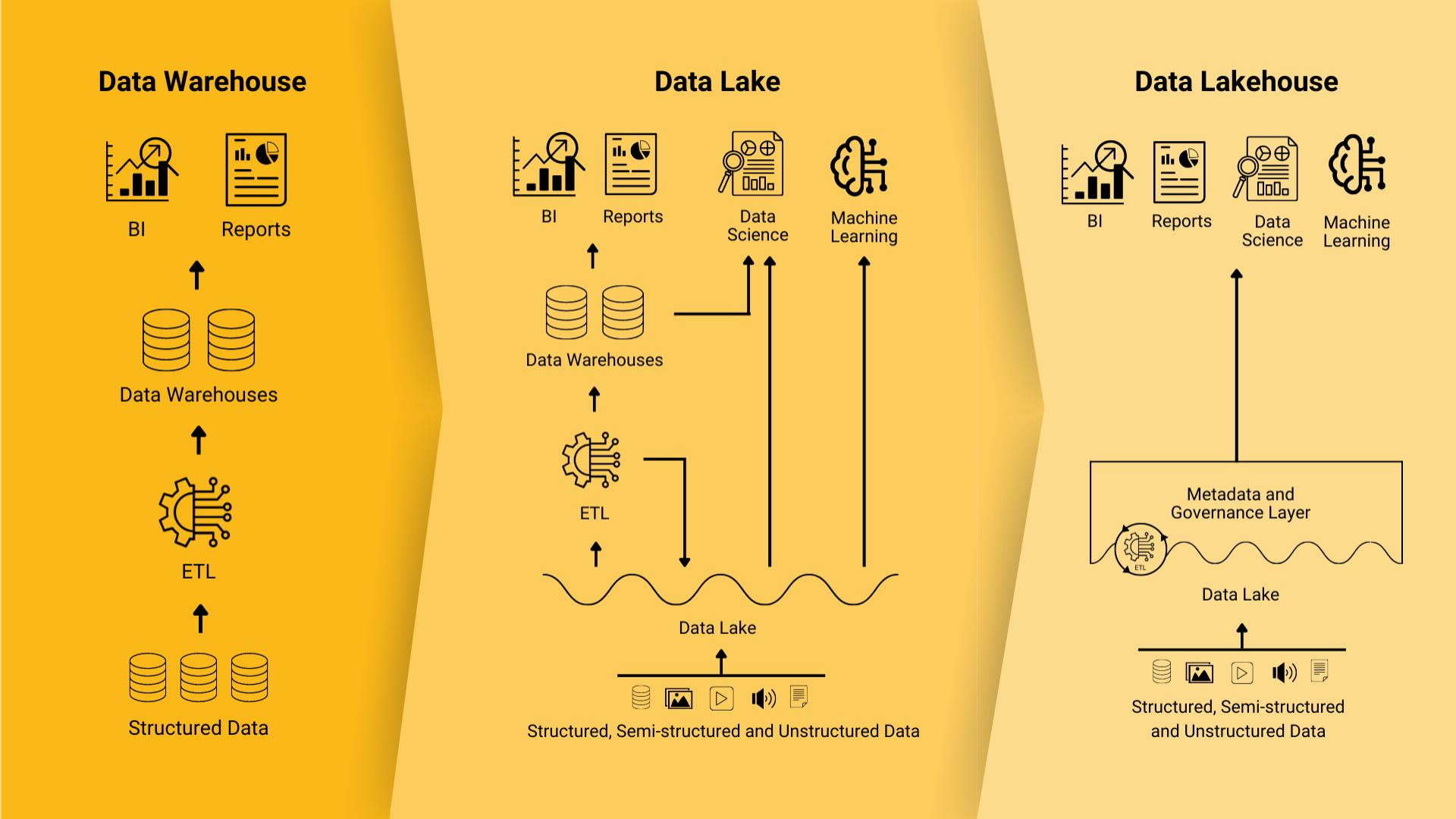
Unique and shared information

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

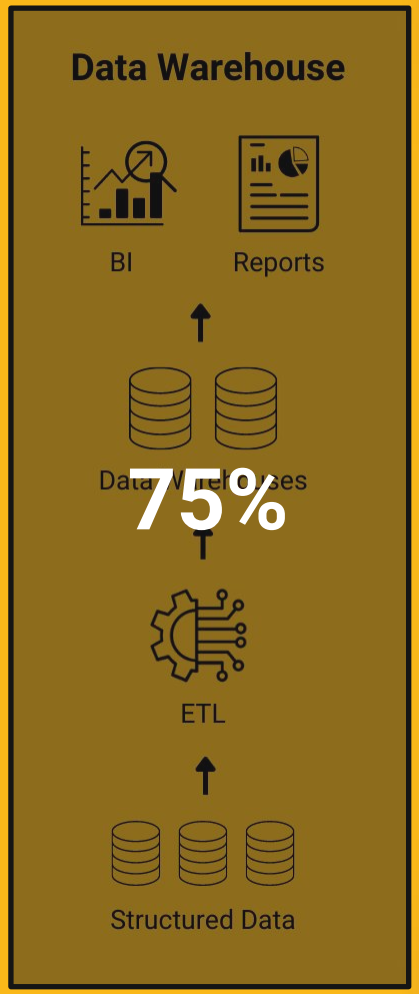
Modern, with the possibility of building piece by piece, in a phased way so as not to spend unnecessary iterations, covering all the needs of analysis, AI and Machine Learning not only current but also future.

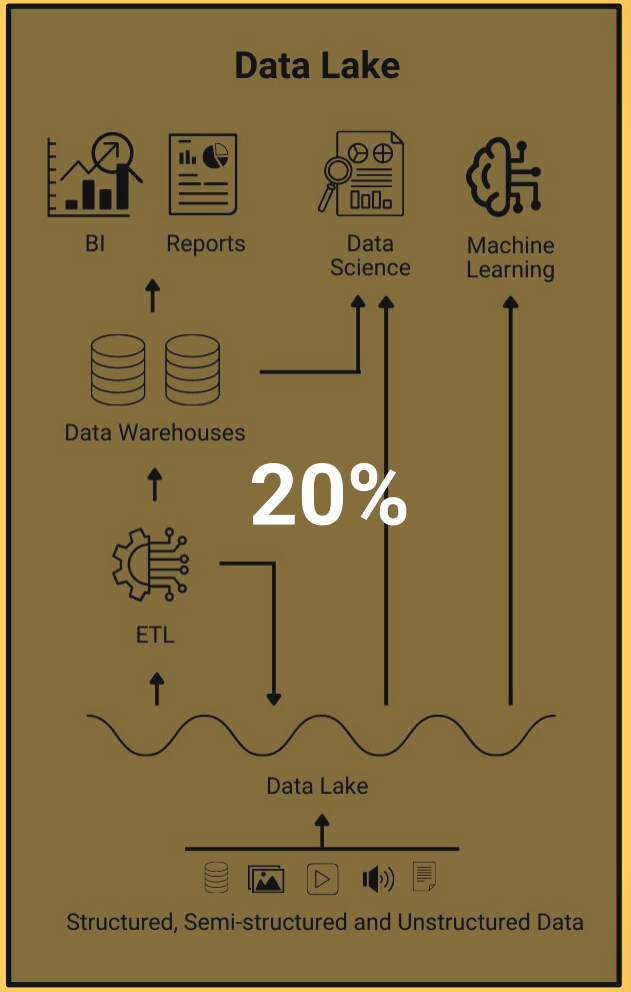
bertia
SMART SOLUTIONS

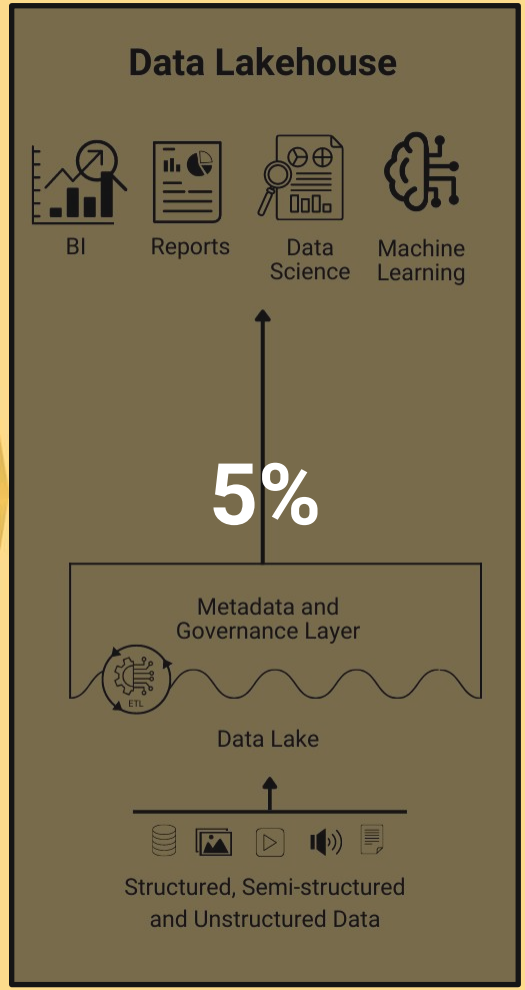

Data Warehouse
BI
Reports
Data Warehouses
ETL
Structured Data
Data Lake
BI
Reports
Data Science
Machine Learning
Data Warehouses
ETL
Data Lake
Structured, Semi-structured and Unstructured Data
Data Lakehouse
BI
Reports
Data Science
Machine Learning
Metadata and Governance Layer
ETL
Data Lake
Structured, Semi-structured and Unstructured Data

¿What is the current situation?

bertia
SMART SOLUTIONS

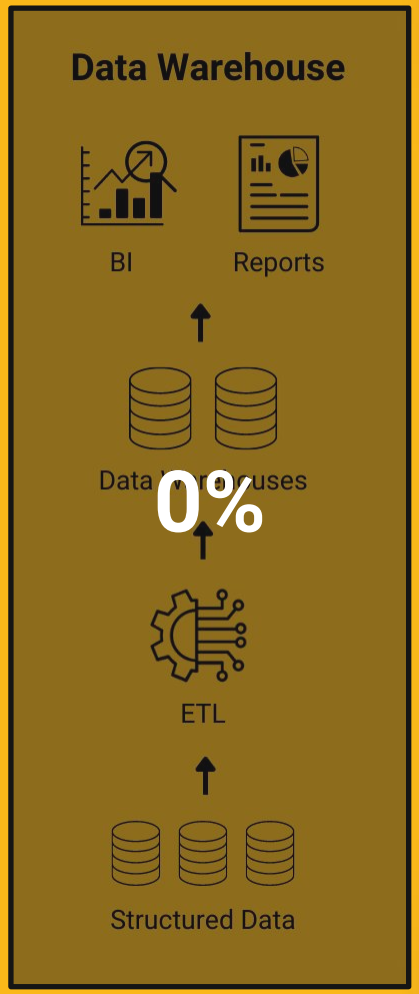

Data Warehouse
BI
Reports
Data Warehouses
75%
ETL
Structured Data

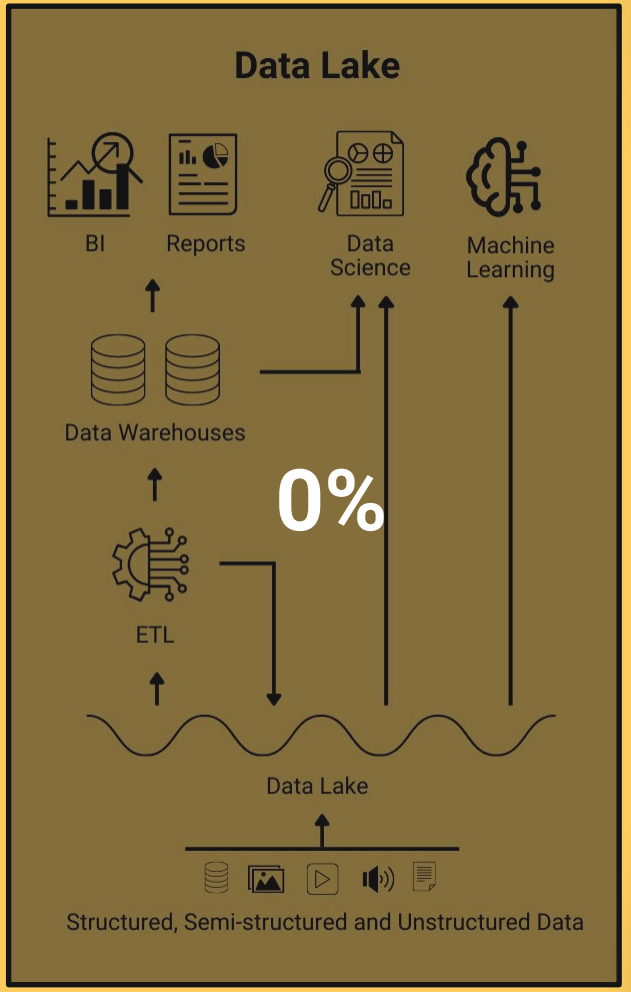

Data Lake
BI
Reports
Data Science
Machine Learning
Data Warehouses
20%
ETL
Data Lake
Structured, Semi-structured and Unstructured Data

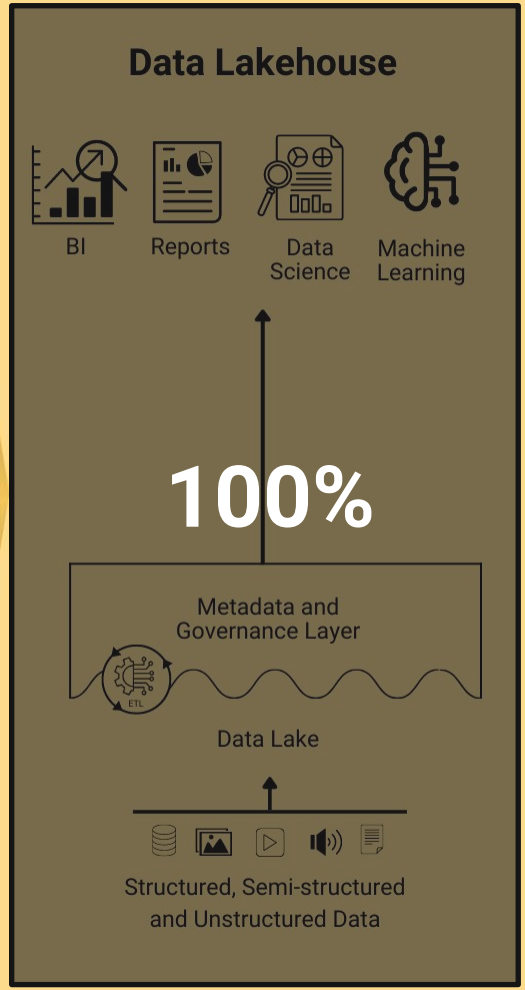

Data Lakehouse
BI
Reports
Data Science
Machine Learning
5%
Metadata and Governance Layer
ETL
Data Lake
Structured, Semi-structured and Unstructured Data

¿ What we want to achieve?

bertia
SMART SOLUTIONS

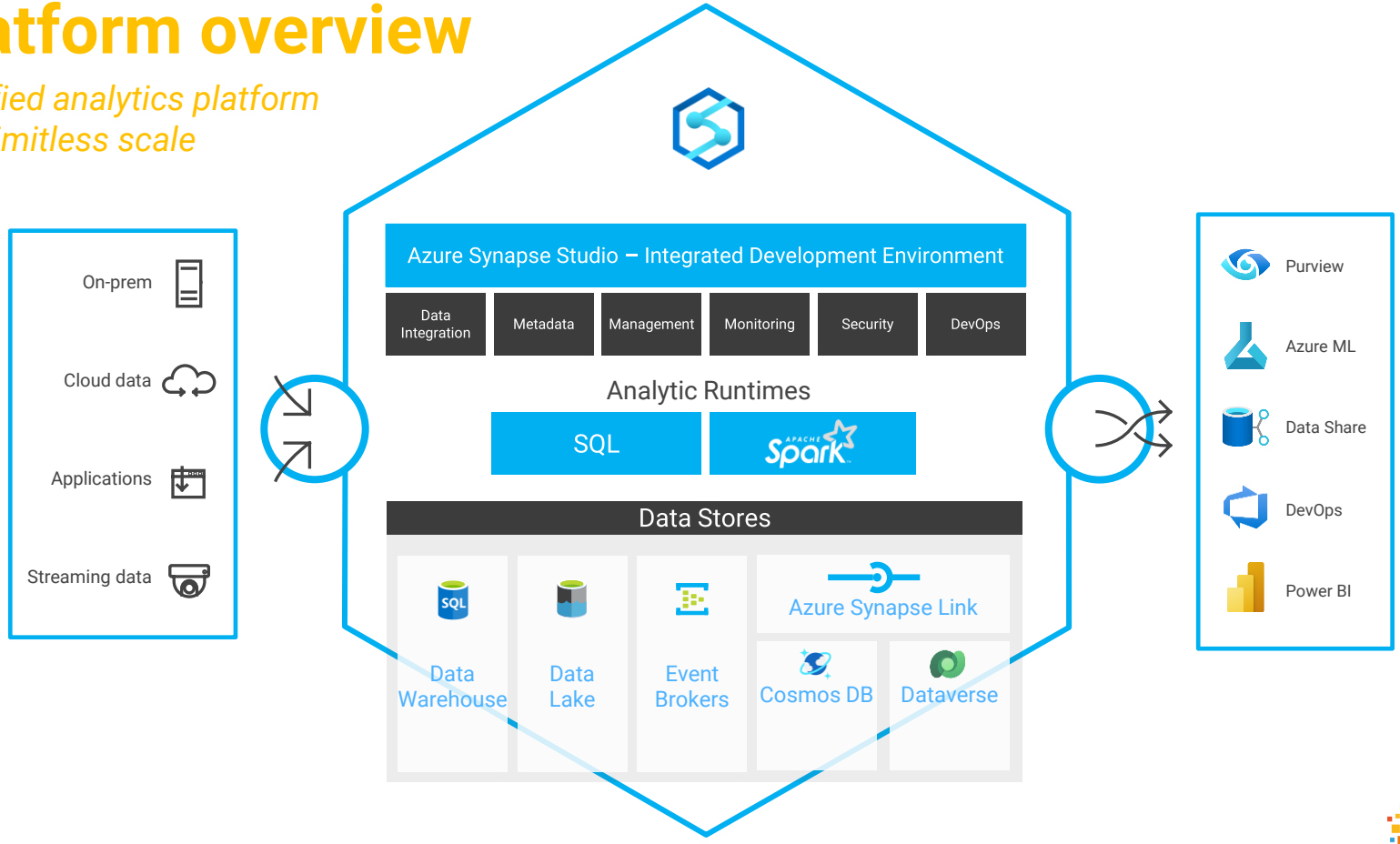

Data Warehouse
BI
Reports
0%
ETL
Structured Data


Data Lake
BI
Reports
Data Science
Machine Learning
Data Warehouses
0%
ETL
Data Lake
Structured, Semi-structured and Unstructured Data


Data Lakehouse
BI
Reports
Data Science
Machine Learning
100%
Metadata and Governance Layer
ETL
Data Lake
Structured, Semi-structured and Unstructured Data

Platform overview

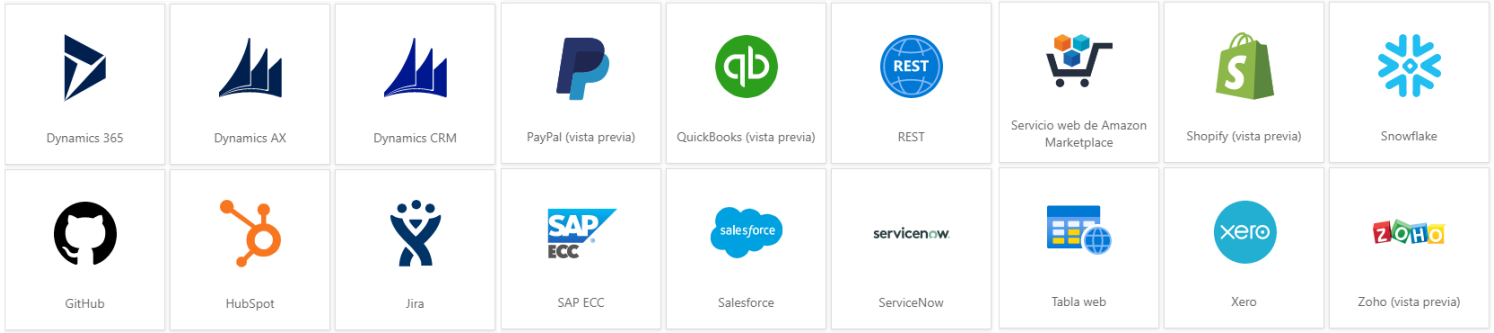
A unified analytics platform with limitless scale



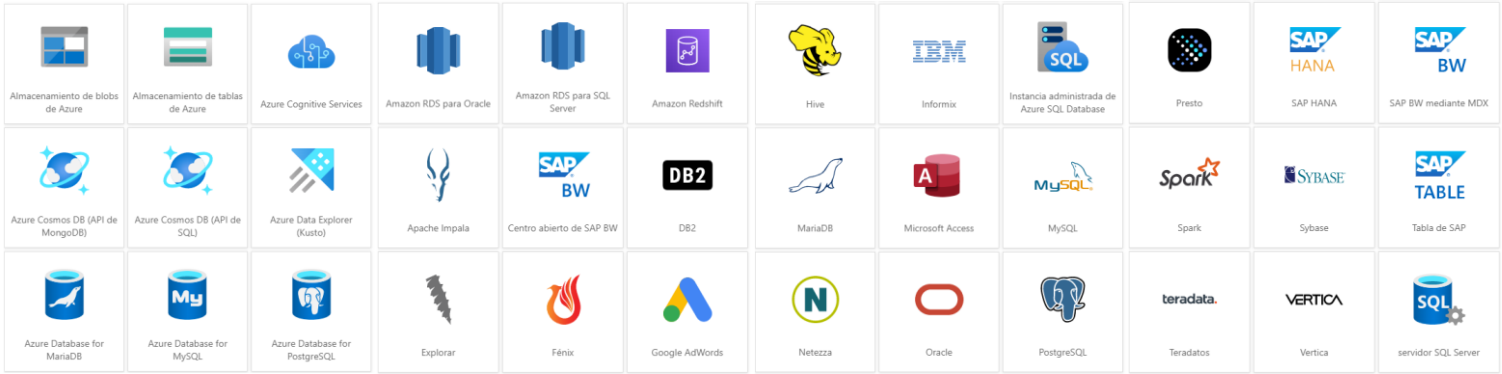
bertia SMART SOLUTIONS

Synapse - Integration

Services and applications:



Storage systems:

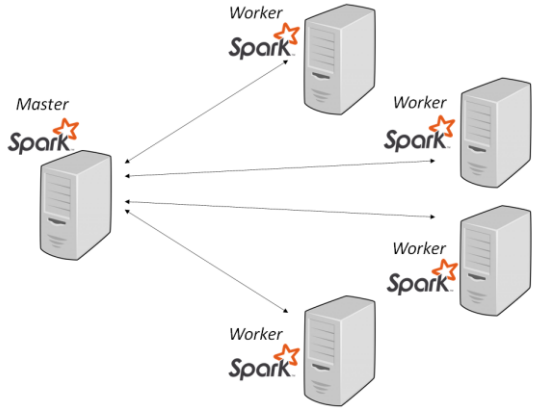


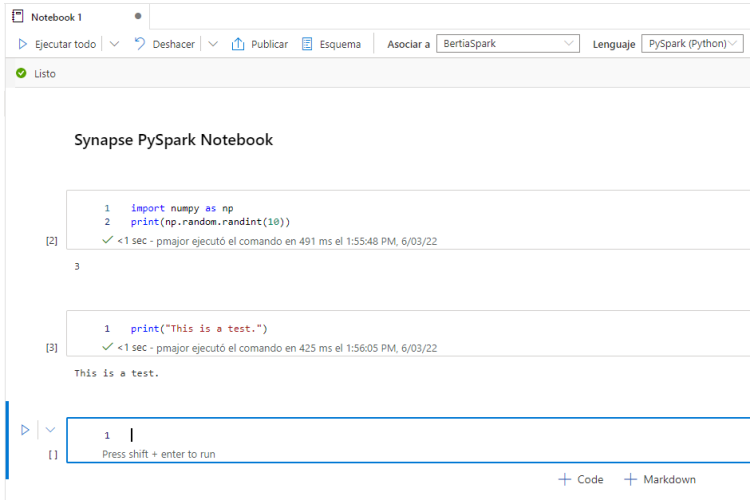
bertia SMART SOLUTIONS

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.





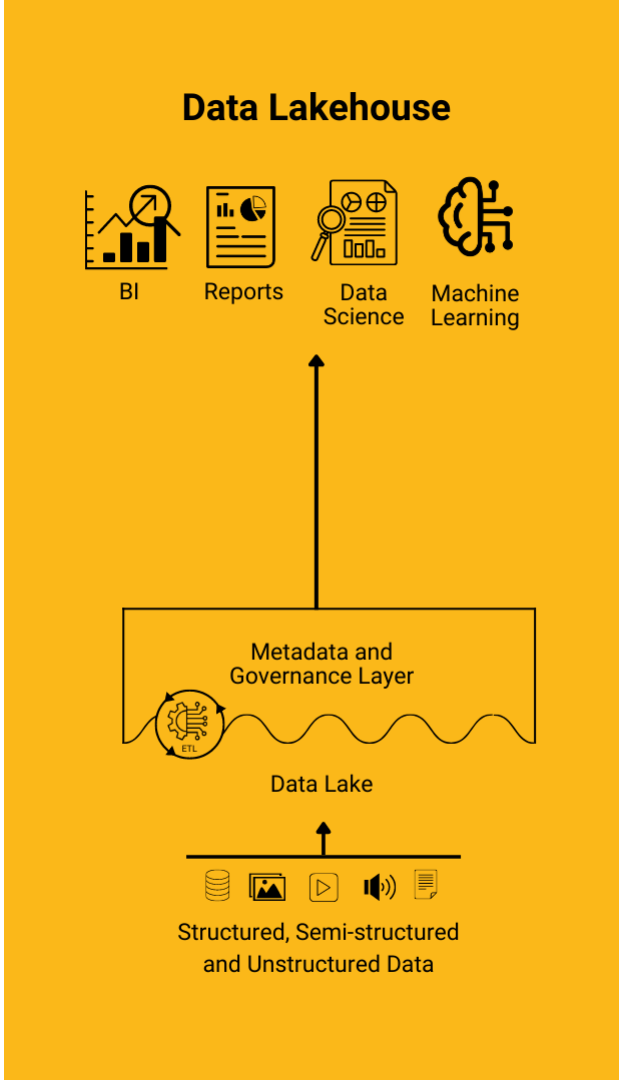
bertia SMART SOLUTIONS

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.



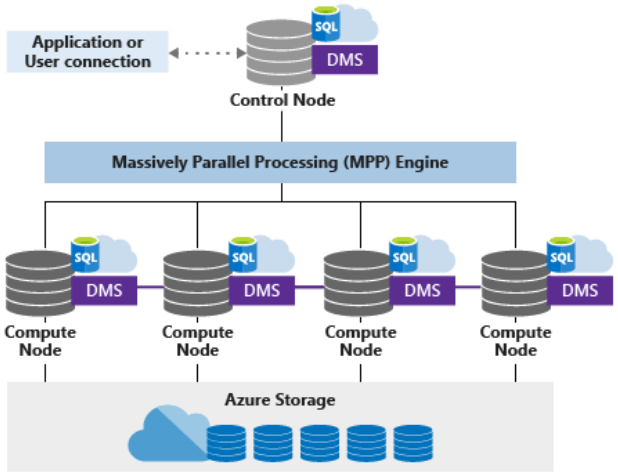
bertia SMART SOLUTIONS

Synapse – SQL Pools

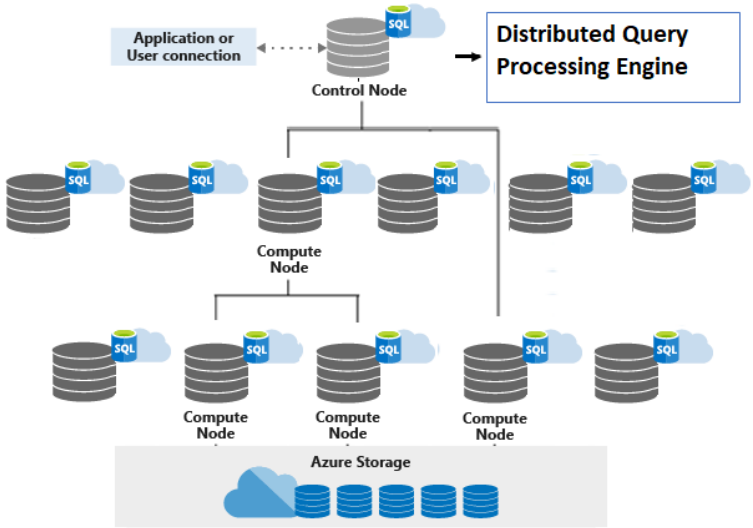
- Previously known as Azure SQL Data Warehouse.
- Payment per defined DWU (computing units).
- Allows to ingest and consult data.

- It scales automatically and is paid per amount of data consulted (in TB). Only allows querying.

Dedicated SQL pool



Serverless SQL pool

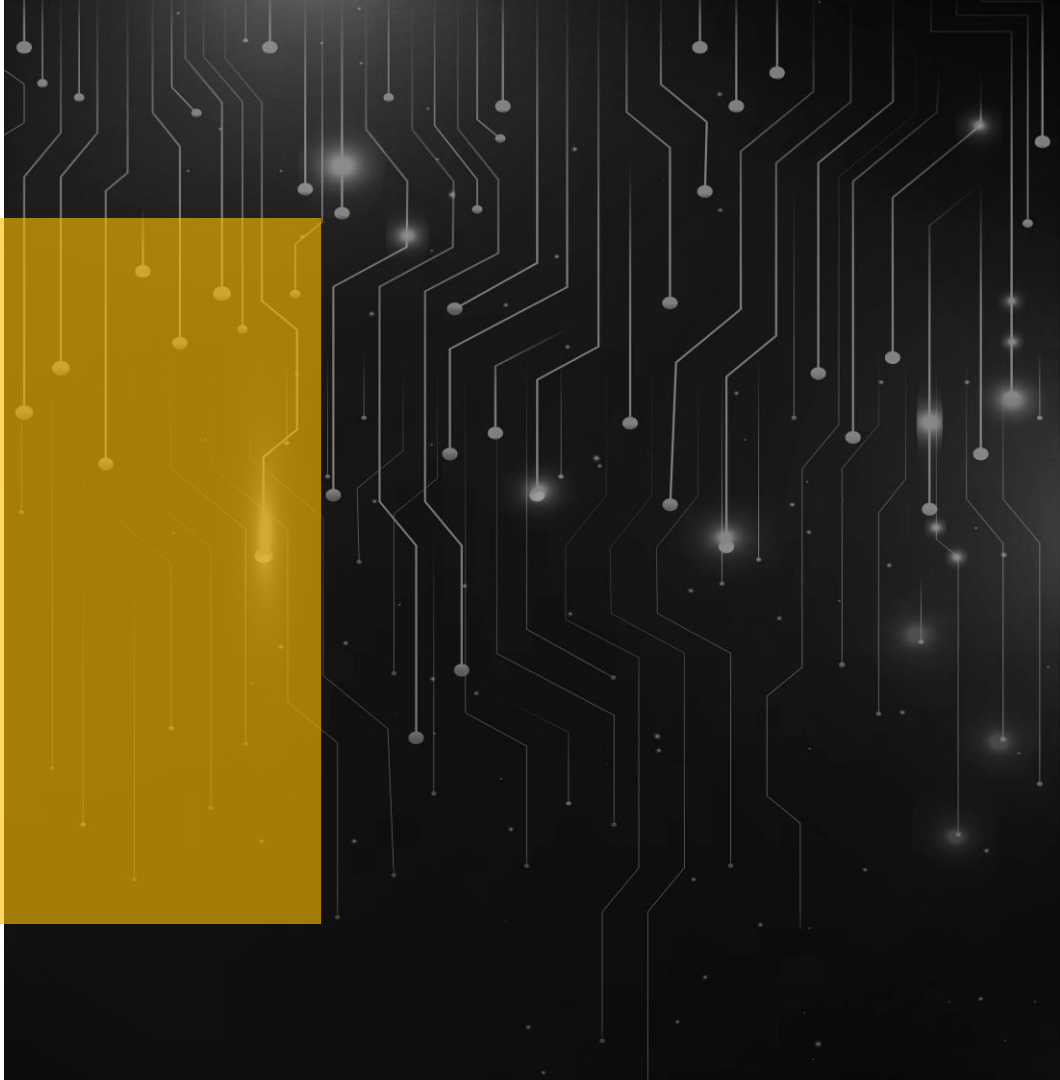


bertia SMART SOLUTIONS

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.

bertia SMART SOLUTIONS

Thank you for your attention