

# ESPECIALIZACIÓN

# **Ingeniería de datos con Azure**

Curso: Microsoft Azure & ETL Fundamentals

Docente: Richard Tadeo Zenteno

## REGLAS



Se requiere **puntualidad** para un mejor desarrollo del curso.



Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.



Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.



Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.



**Identificarse** en la sala Zoom con el primer nombre y primer apellido.

## ITINERARIO

*07:00 PM – 07:30 PM      **Soporte técnico DMC***

*07:30 PM – 08:50 PM      **Agenda***

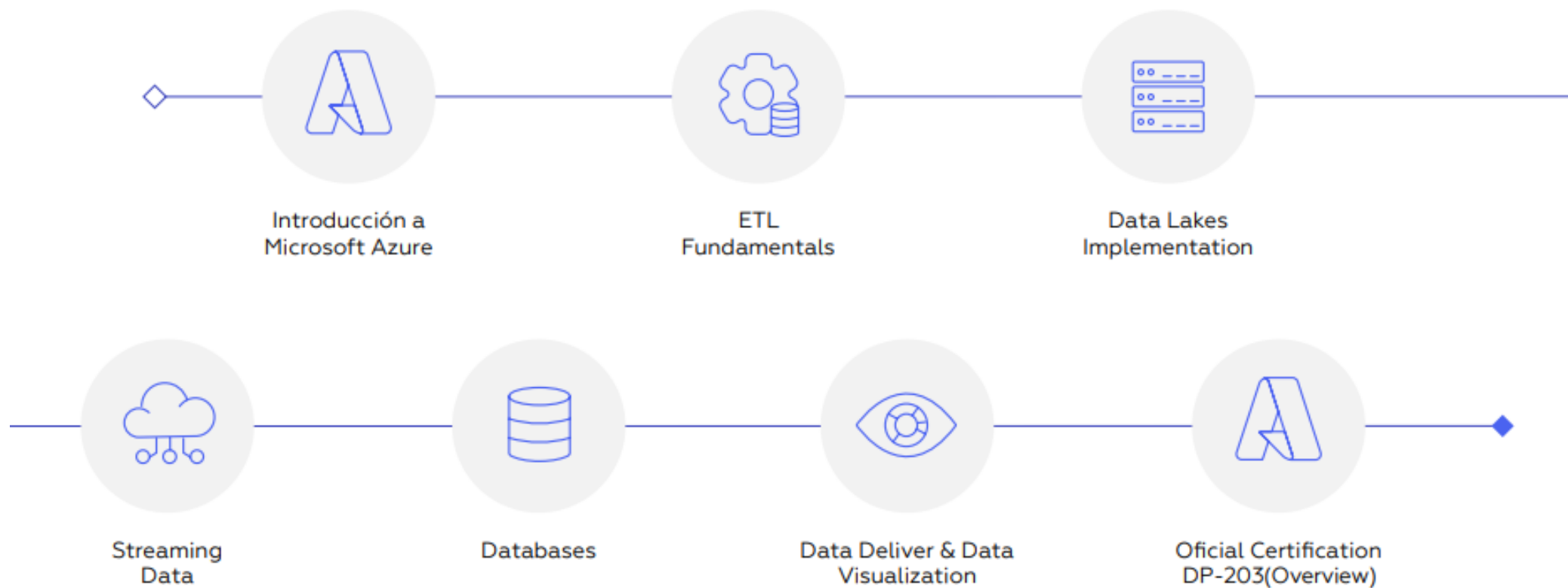
*08:50 PM – 09:00 PM      **Pausa Activa***

*09:00 PM – 10:30 PM      **Agenda***

*Horario de Atención Área Académica y Soporte*

*Lunes a Viernes 09:00 am a 10:30 pm / Sábado 09:00 am a 02:00pm*

# MALLA CURRICULAR



## CERIFICACIÓN FINAL

por **Aprobación** de la Especialización en **Ingeniería de Datos con Microsoft Azure** (48 horas académicas)

# CONTENIDO



## Introducción a Microsoft Azure

- Introducción a Cloud Computing. Proveedores de servicios Cloud, On-Premise vs. On-Cloud, principales servicios, descripción de los modelos de costos.
- Identity and Access Management (IAM). Overview de los roles principales, ejemplos de gestión de permisos.



## ETL Fundamentals

- Introducción a las soluciones ETL. Definición, descripción de sus etapas.
- Introducción a los servicios Azure Data Factory y Data Flow. Características generales, casos de uso.
- Taller: Implementación de un ETL Básico con Azure.



## Data Lakes Implementation

- Introducción a Data Lakes. Definición, arquitectura, capas (Raw, Stage, Analytics).
- Introducción a los servicios Azure Blob Storage y Storage Account.
- Taller: Implementación de un Datalake en Azure.

# CONTENIDO



## Streaming Data

- Introducción a procesamiento de datos Batch y Streaming. Diferencias Near-Real-Time y Real-Time.
- Introducción a IoT. Definición, uso de sensores, aplicaciones.
- Revisión de servicios: Azure EventHubs y IoT Hub. Características generales, ejemplos de implementación y uso.
- Taller: Manejo de Streaming al Data.



## Databases

- Introducción a las bases de datos Relacionales y No-Relacionales. Definición, características, casos de uso.
- Azure SQL Database for MariaDB. Descripción y características generales.
- Azure SQL Database for PostgreSQL. Descripción y características generales.
- Azure SQL Database for CosmosDB. Descripción y características generales.
- Taller: Diseño de una base de datos relacional y técnicas para poblarla.



## Data Deliver & Data Visualization

- Azure Synapse Analytics. Propósito del servicio, características generales.
- Fabric. Propósito del servicio, características generales.
- Taller: Conexión de Power BI a servicios de datos de Azure.

# AGENDA

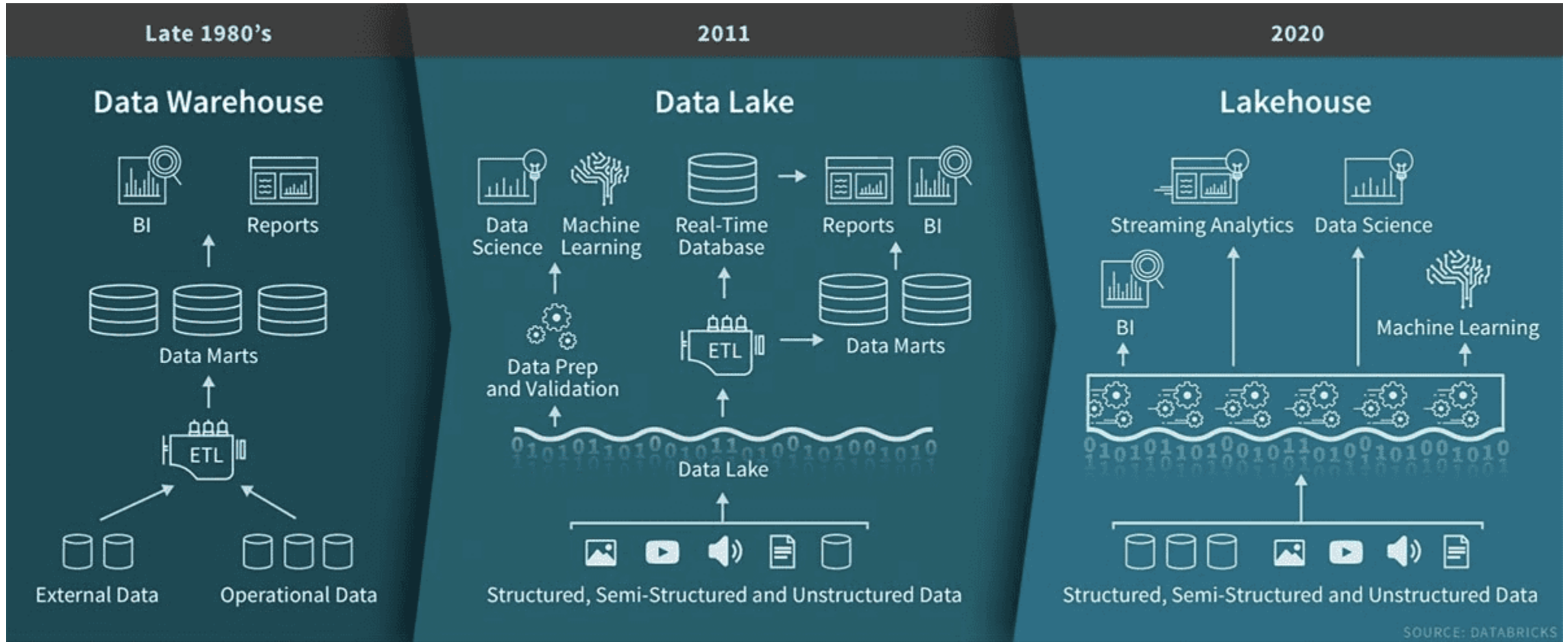
**01**

Introducción a  
Data Lakes

**02**

Características de  
Data Lakes

# Evolución Arquitectura de Datos





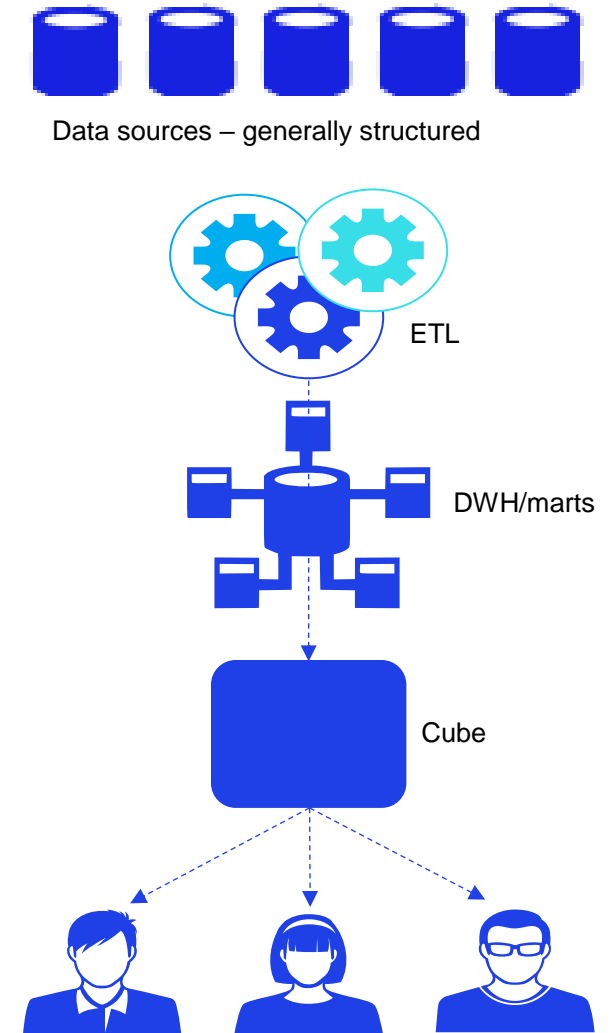
# Data Warehouse

*“Es una colección de datos diseñada para soportar la administración de la toma de decisiones, orientada por temas, integrada, perdurable y variante en el tiempo.”*

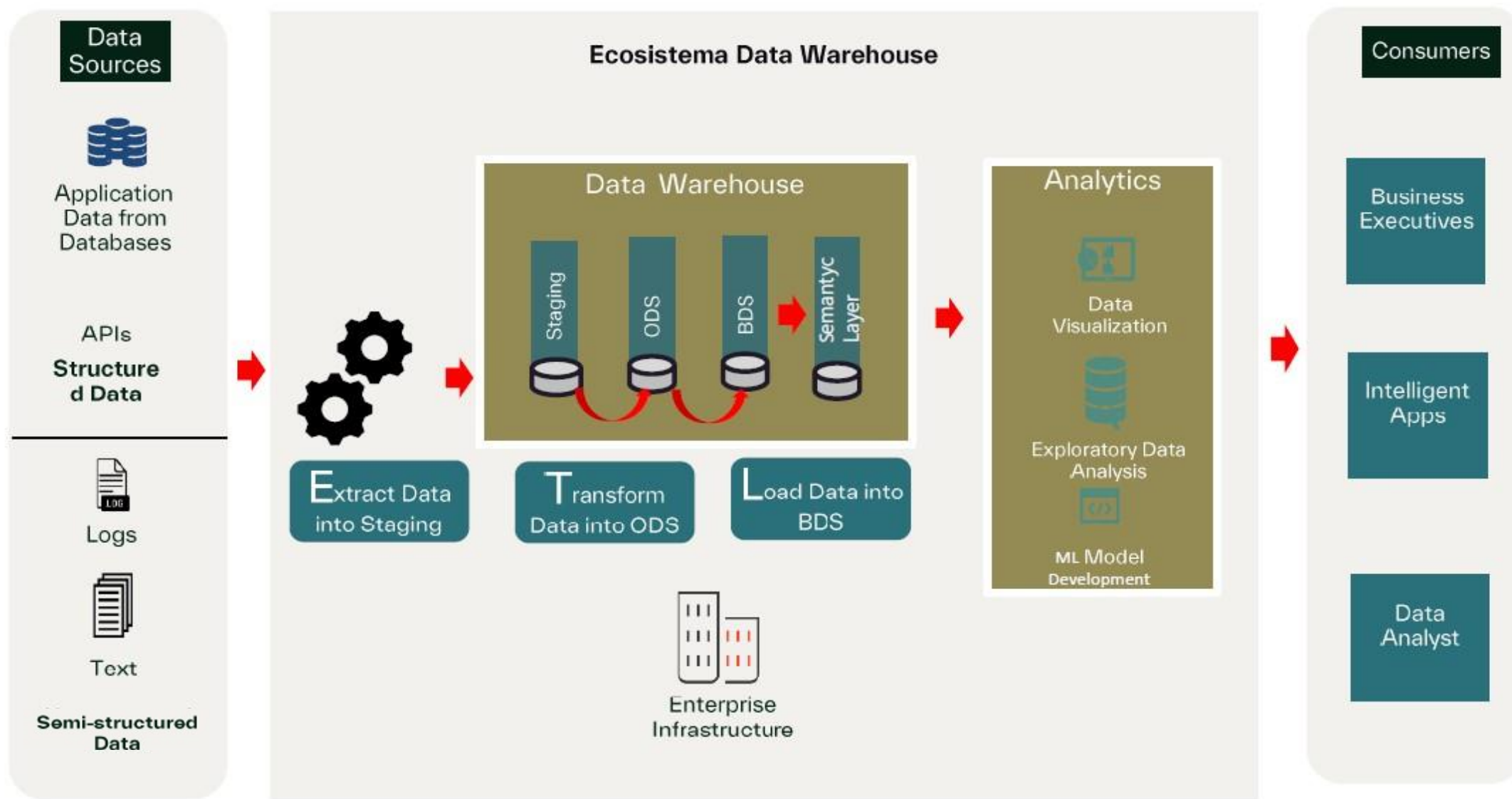
- Bill Inmon

## Características

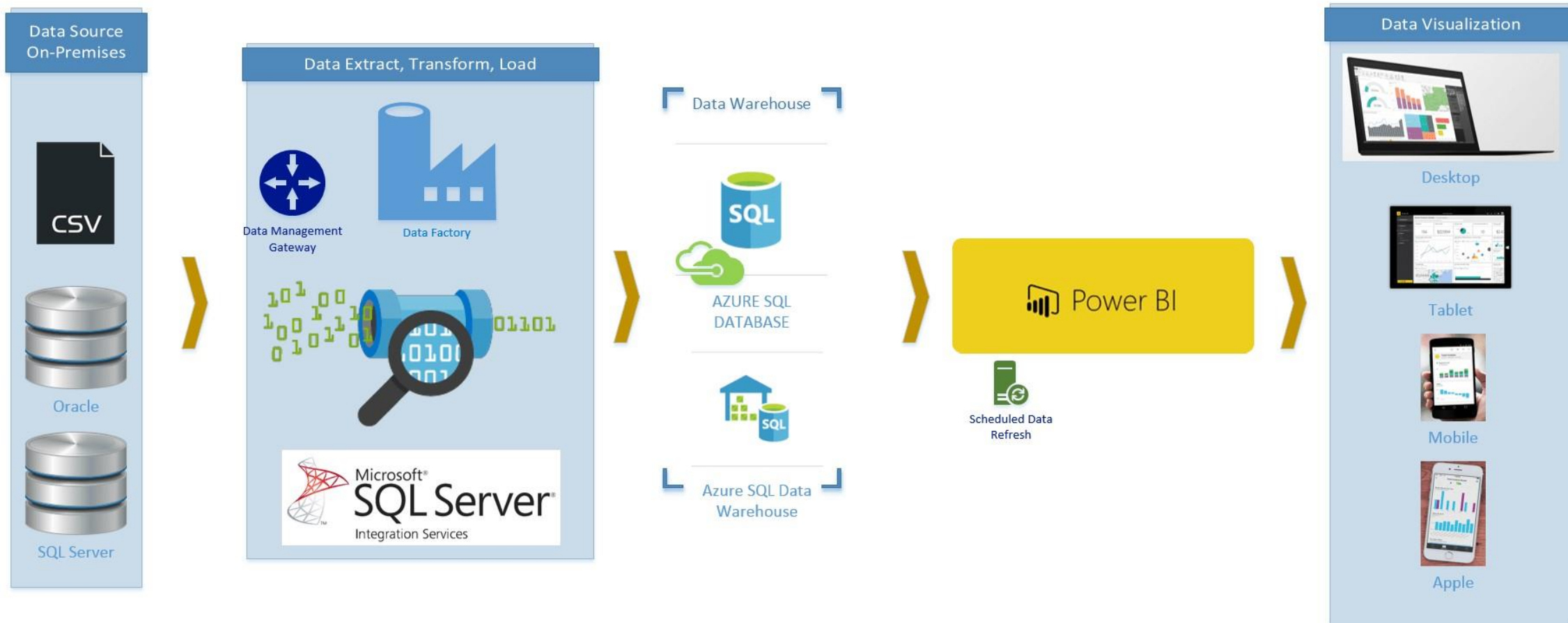
- Se simplifica el acceso a los datos.
- Las bases de datos son creadas con la información específica para ser usada en la toma de decisiones.
- Los datos se extraen desde sus sistemas de origen con un proceso de transformación e integración.
- Asegura integridad y calidad de los datos.
- Perspectiva histórica.
- **Schema on write.**



# Data Warehouse



# Data Warehouse en Azure



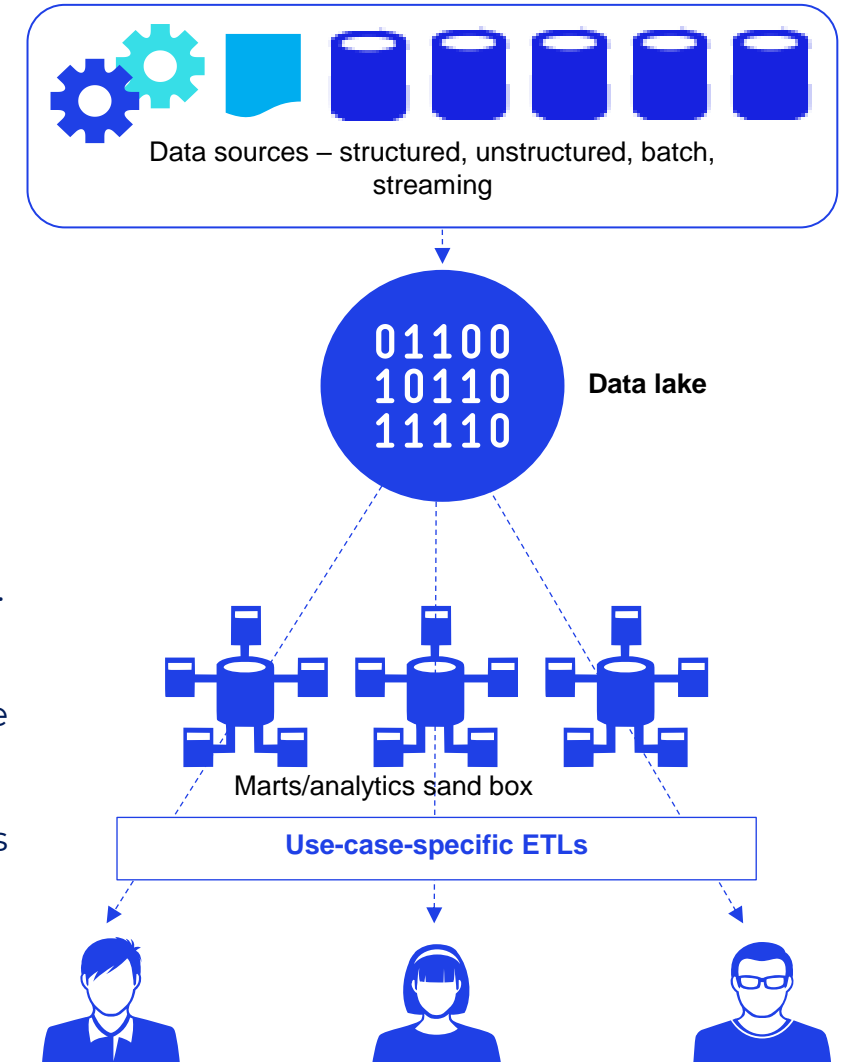
# Data Lake

*“Un data lake o lago de datos es un repositorio de almacenamiento centralizado que recoge grandes cantidades de datos de diferentes fuentes sin procesar en su formato original.”*

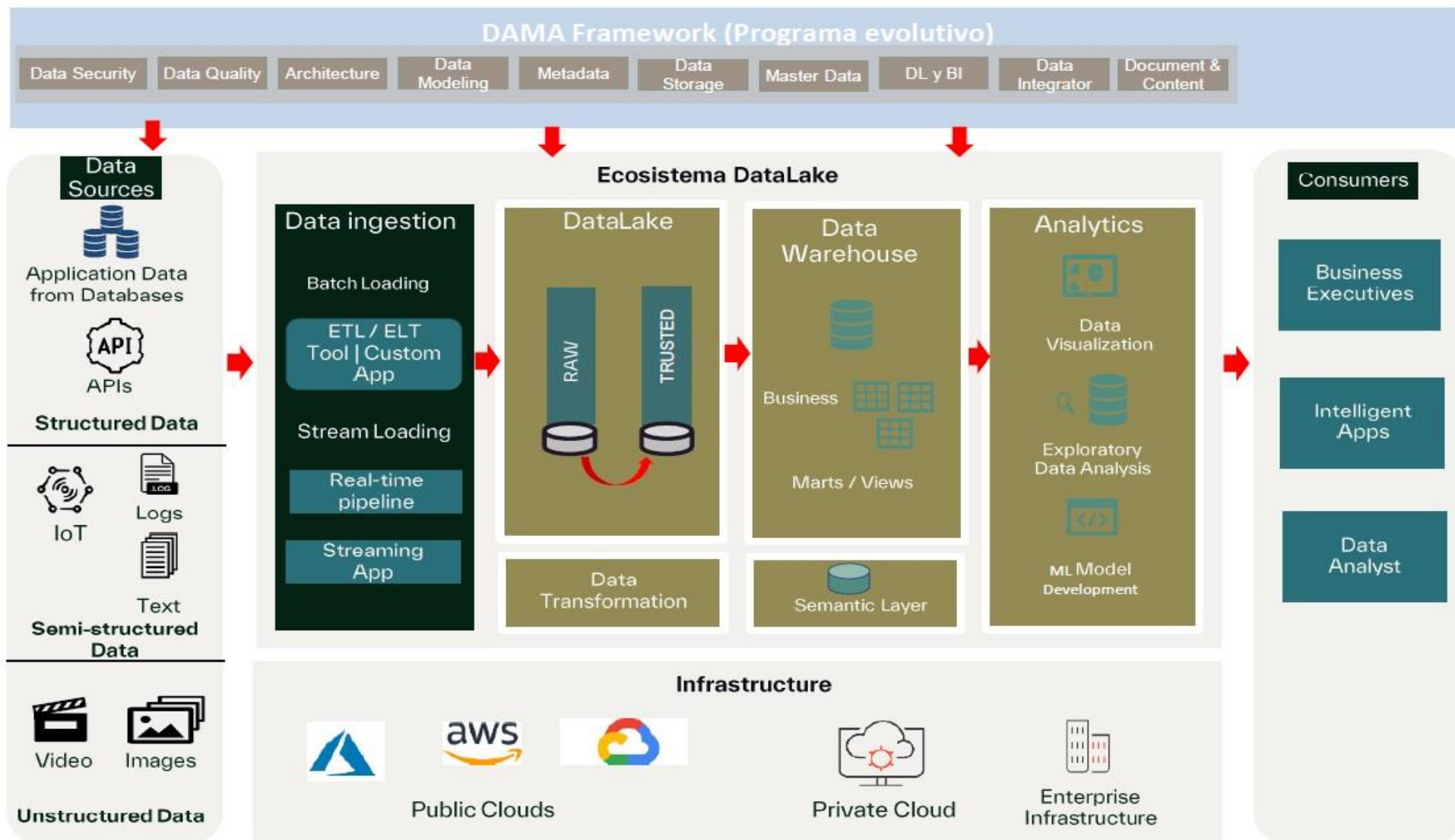
- Various Authors

## Características

- Almacenamiento de datos estructurados, semi-estructurados y no estructurados.
- Crecimiento a cualquier escala a bajo costo.
- Se almacenan datos en formatos de archivo genéricos y abiertos, como Apache Parquet, Avro y ORC.
- Da oportunidad a los científicos y analistas de datos de poder obtener insights importantes para la organización de datos que no se encuentran estructurados.
- **Schema on read.**



# Data Lake



# Data Lake en Azure



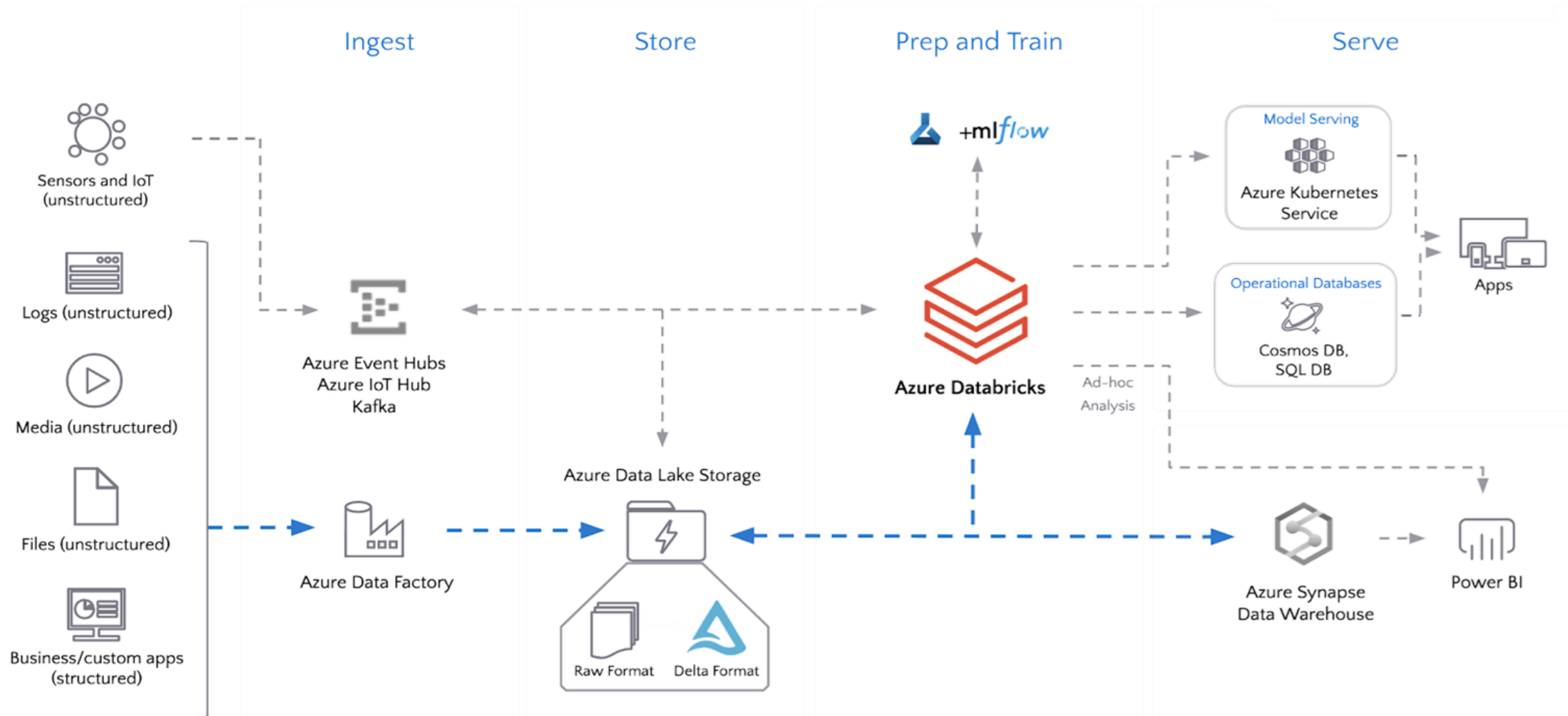
Azure Key Vault



Azure Automation

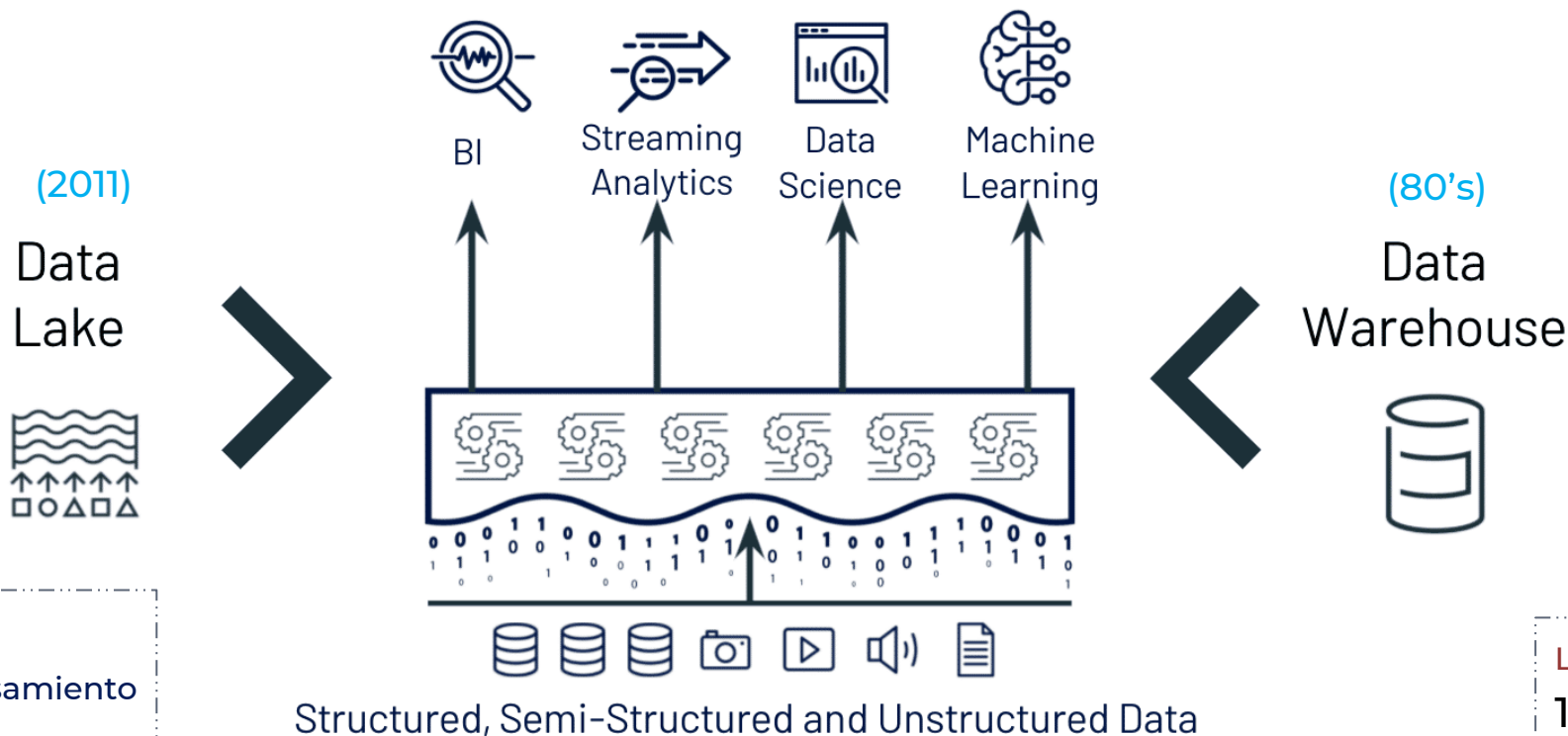


Azure DevOps



# ¿Qué es un Lakehouse?

## Lakehouse (2020)



### Limitaciones:

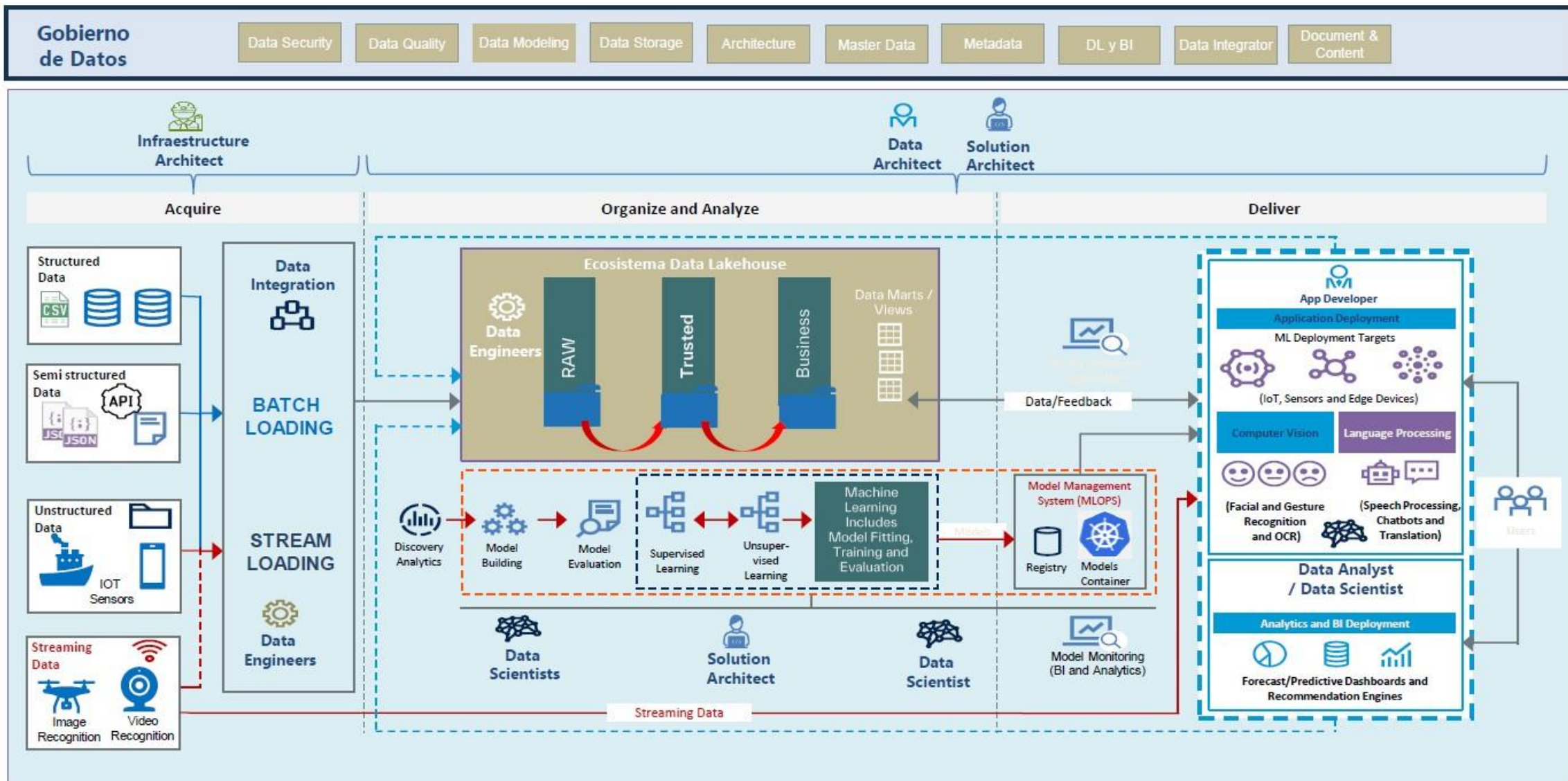
1. Dificultad para procesamiento de KB y MB.
2. Soporte para transacciones.
3. Soporte de unicidad e integridad referencial. (PK, FK)

### Limitaciones:

1. Variedad de datos.
2. ML e IA.
3. Procesamiento de cientos de GB y PB.

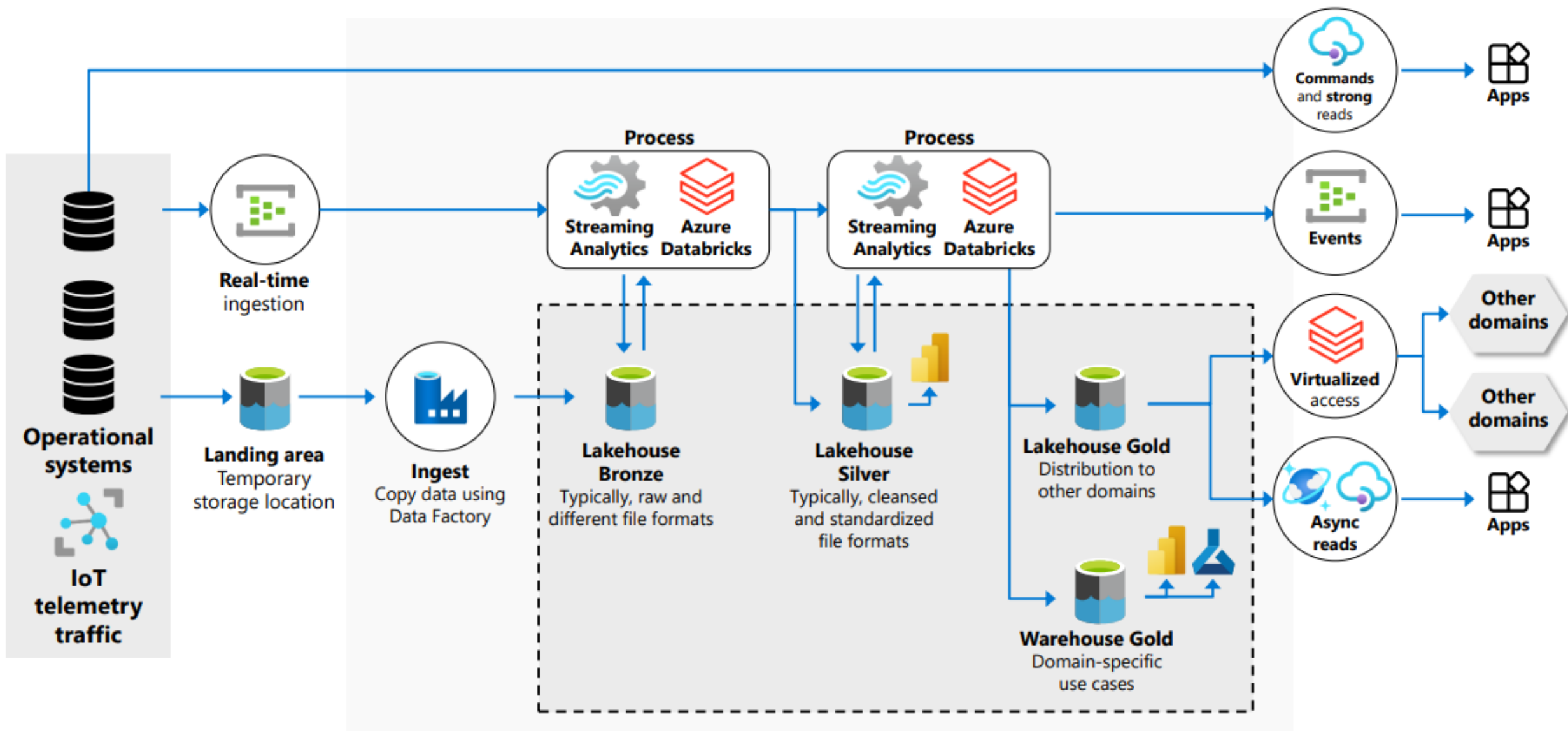


# Lakehouse

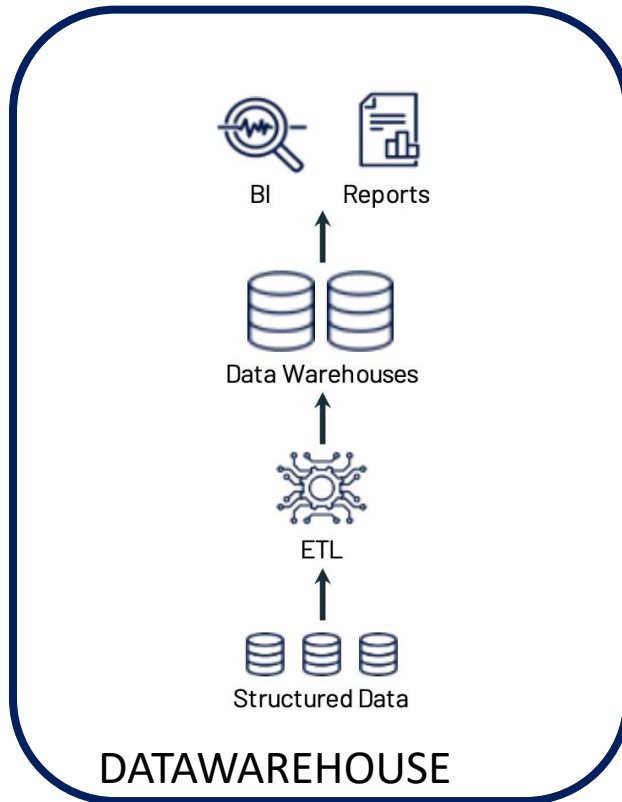




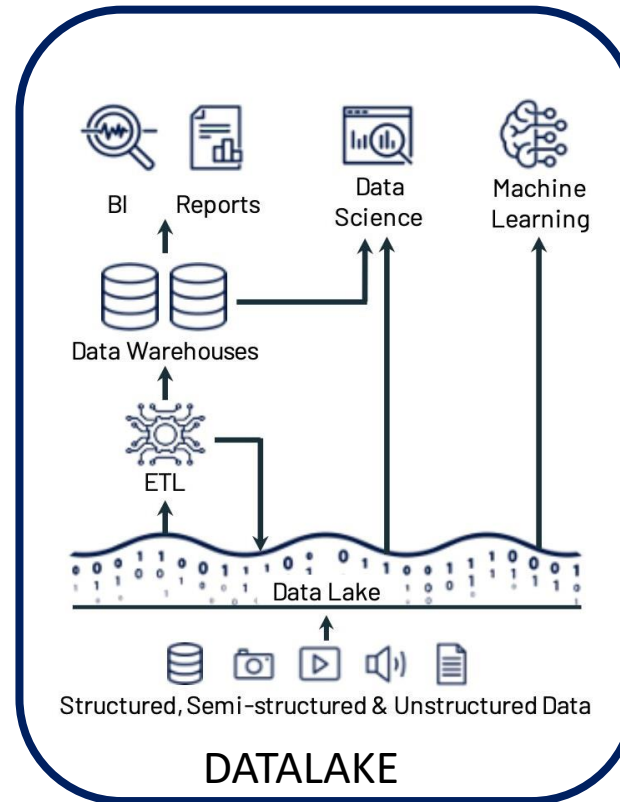
# Lakehouse en Azure



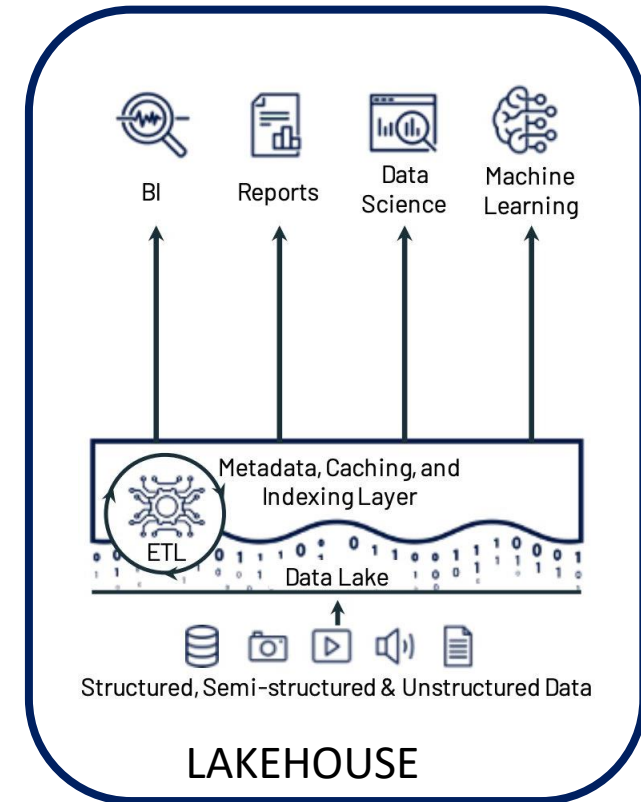
## Patrones de Adopción – DLK / DWH (Plataformas de Datos)



- Orientado al análisis de datos
- Utilizado por analistas
- Enfocado en la toma de decisiones



- Orientado al almacenamiento de datos
- Utilizado por profesionales, científicos de datos
- Enfocado en la centralización de datos para análisis avanzado



- Combinación de Datawarehouse y Datalake.
- Utilizado por profesionales, científicos de datos y analistas.
- Enfocado en la centralización de datos para análisis avanzado y toma de decisiones.

# RONDAS DE PREGUNTAS



**¡GRACIAS!**

