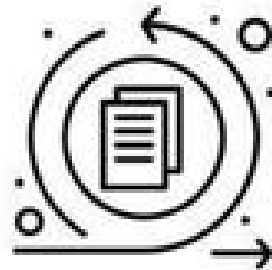
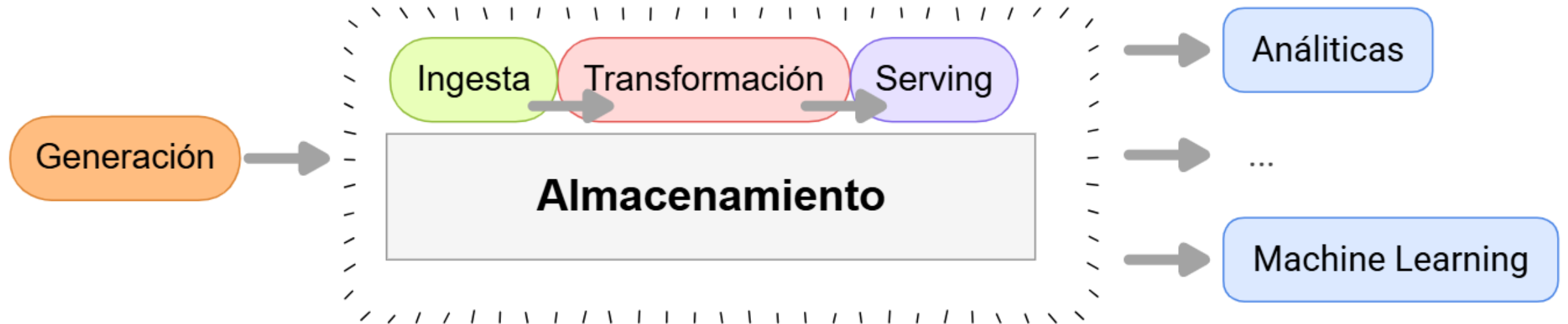


# El Ciclo de Vida de la Ingeniería de Datos (Data Engineering Lifecycle)



# Ciclo de Vida de la Ingeniería de Datos



---

Seguridad | Gestión de Datos | DataOps | Arquitectura de Datos | Orquestación | Ingeniería de Software

# Generación

Un ingeniero de datos consume datos de un sistema de origen, pero normalmente no posee ni controla el sistema de origen en sí.

Se necesita un entendimiento de la forma, la frecuencia y la velocidad a la que se generan los datos, y la variedad de datos que se generan.

Línea de comunicación abierta con los propietarios del sistema de origen.

# Generación

## **Ejemplos de sistemas de origen:**

- Dispositivos IoT
- Terminales de tarjetas de crédito
- Operaciones de la bolsa
- Sensores de telescopios
- Spreadsheets

# Generación

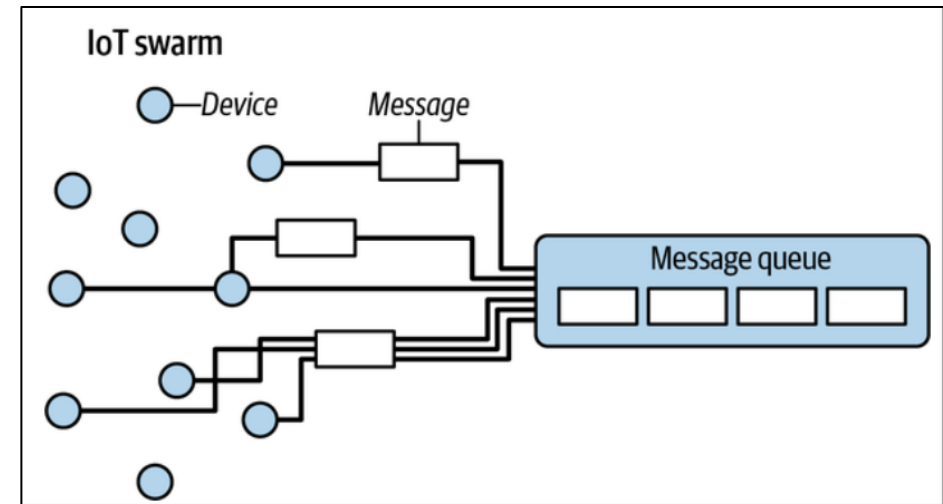
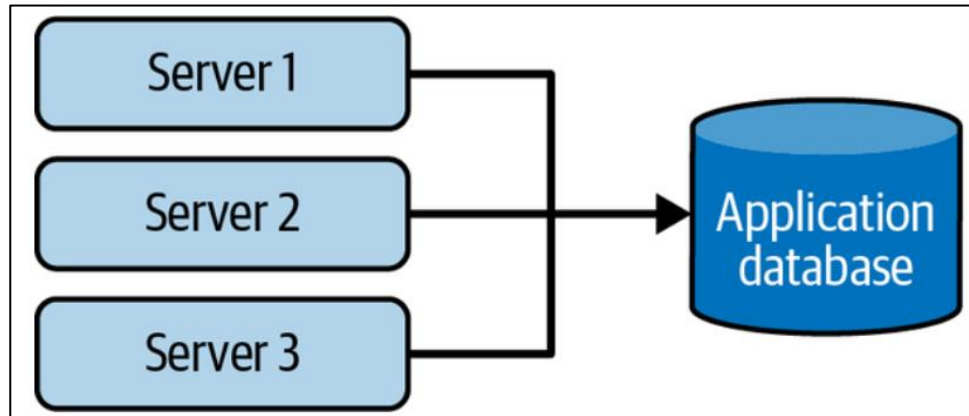
## **Ejemplos de sistemas de origen:**

- Dispositivos IoT
- Terminales de tarjetas de crédito
- Operaciones de la bolsa
- Sensores de telescopios
- Spreadsheets

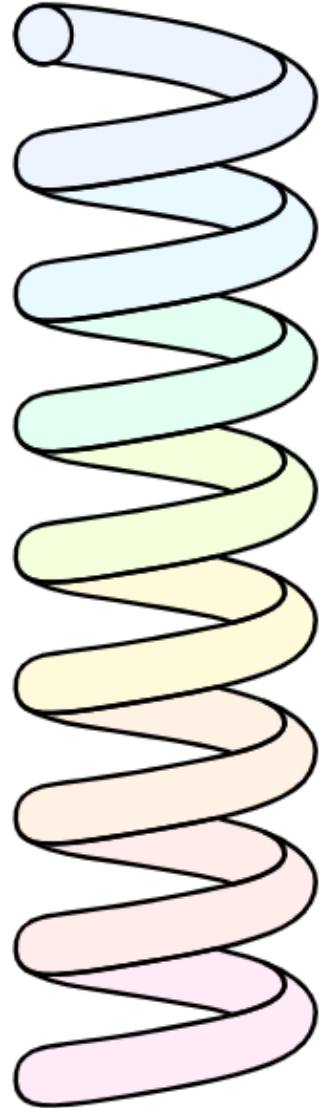
## **Tipos de sistemas de origen:**

- Archivos y datos no estructurados (Excel, CSV, JSON, XML, TXT)
- APIs
- Bases de Datos
- Application Databases
- Fuentes de Datos de Terceros
- Colas de Mensajes y Plataformas de Streaming de Eventos

# Generación



# Generación



Identificar Características del Sistema de Origen



Comprender la Persistencia de Datos



Evaluar la Velocidad de Generación de Datos



Evaluar la Consistencia de los Datos



Evaluar la Frecuencia de Errores



Evaluar la Presencia de Duplicados

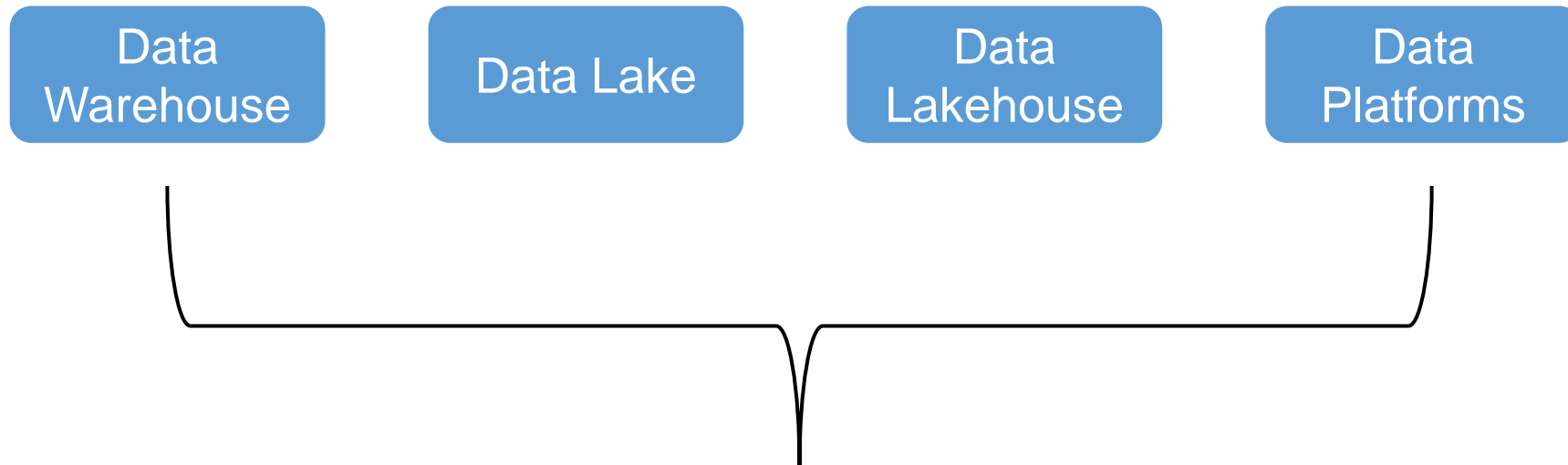


Evaluar la Puntualidad de los Datos



Evaluar el Esquema de Datos

# Almacenamiento



Sistema de almacenamiento:  
almacenamiento de **archivos**, de **bloques**, de  
**objetos** o basados en **cache y RAM**



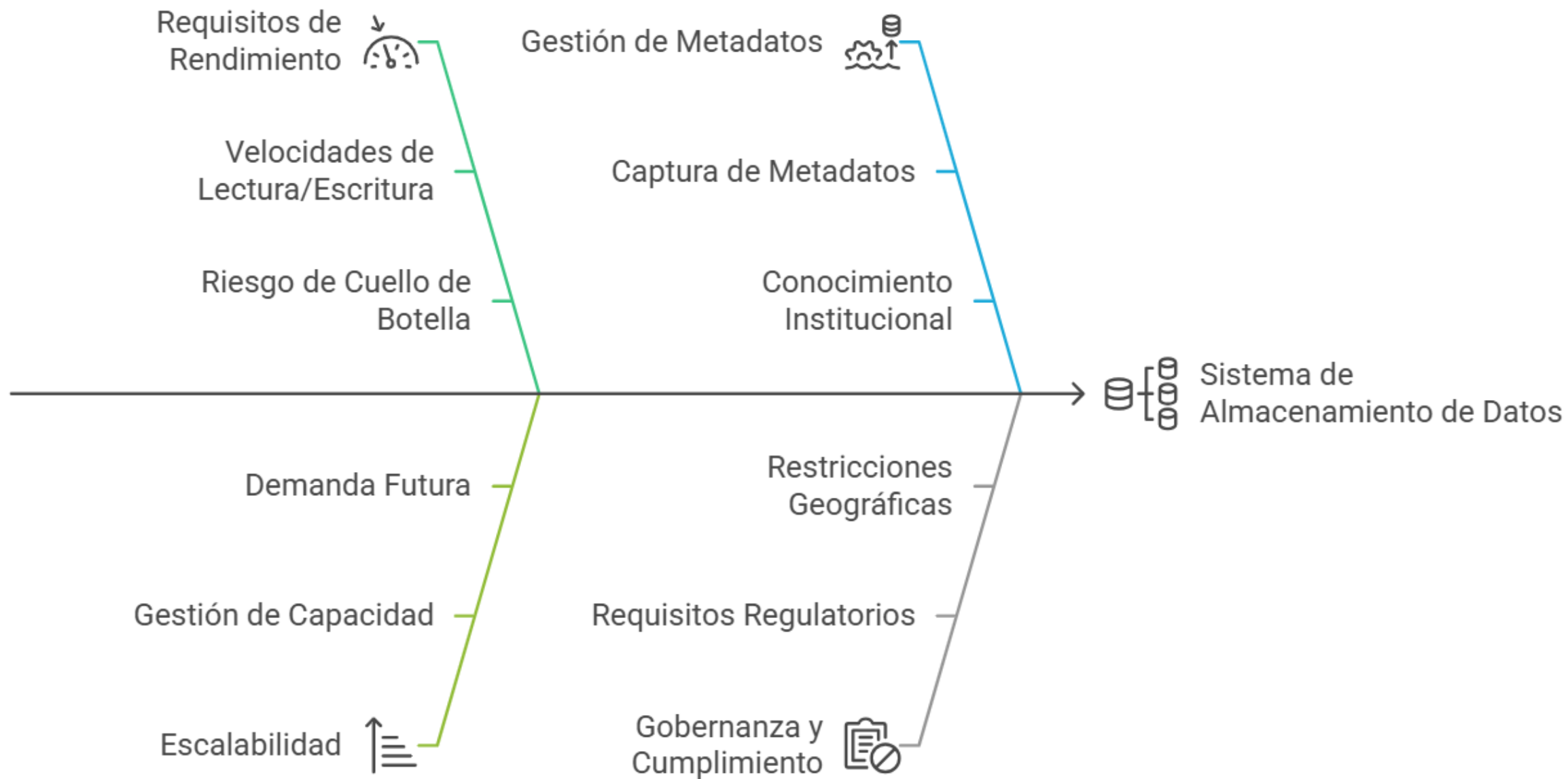
# Almacenamiento

Data  
Warehouse

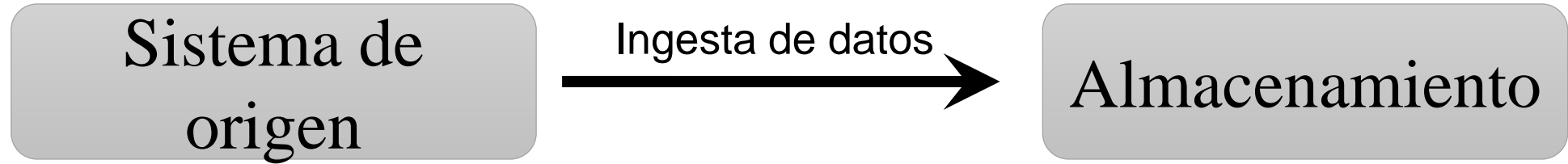
Data Lake

Data  
Lakehouse

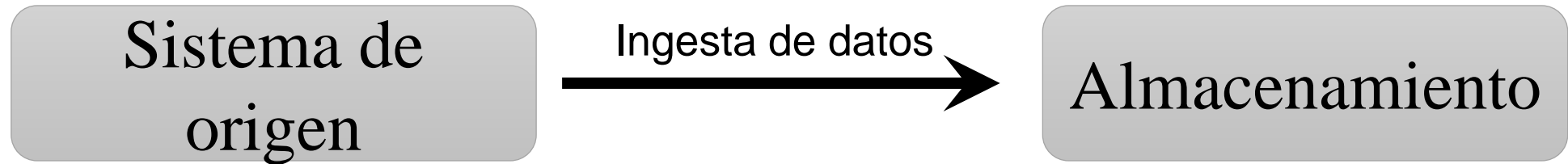
Data  
Platforms



# Ingesta



# Ingesta



## **Factores a considerar:**

- Frecuencia
- Sincronía
- Serialización
- Escalabilidad
- Carga
- Estabilidad

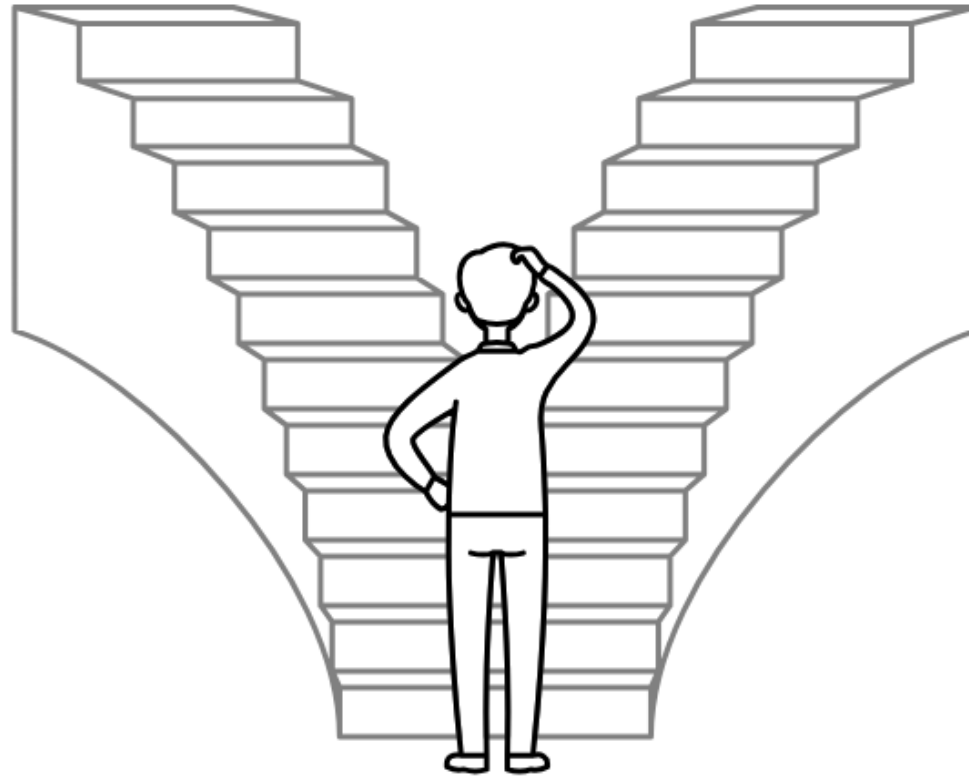
# Lotes (Batch) vs. Streaming

## Ingesta por Lotes

Adecuado para manejar grandes volúmenes de datos a la vez; menos complejo para sistemas de almacenamiento.

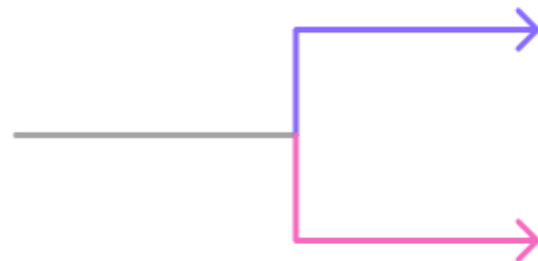
## Ingesta por Streaming

Ideal para datos en tiempo real; permite acciones inmediatas y análisis.





**Push o Pull**



### **Modelo Push**

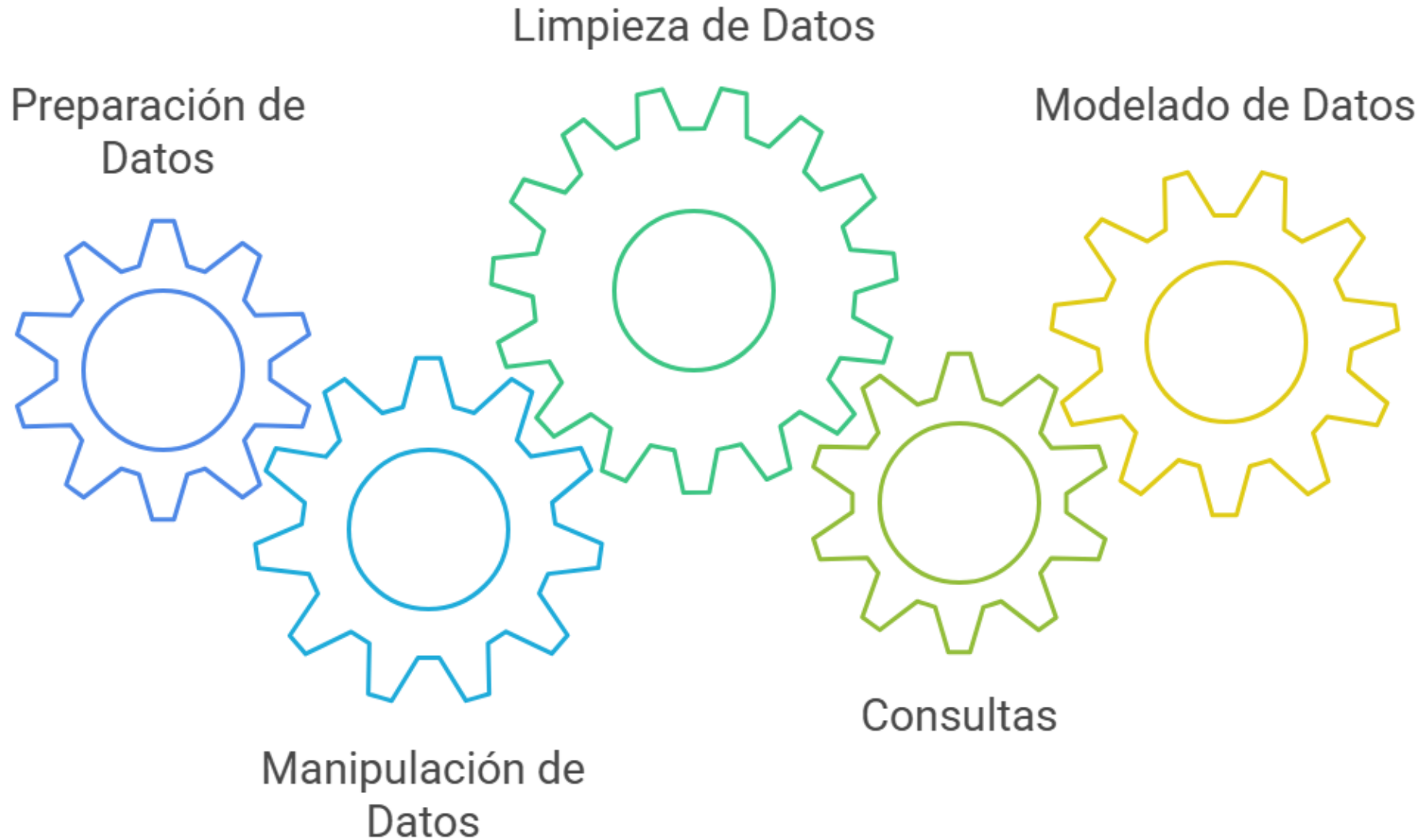
Sistema de origen envía datos hacia un destino, ejemplo: streaming, sensores IoT



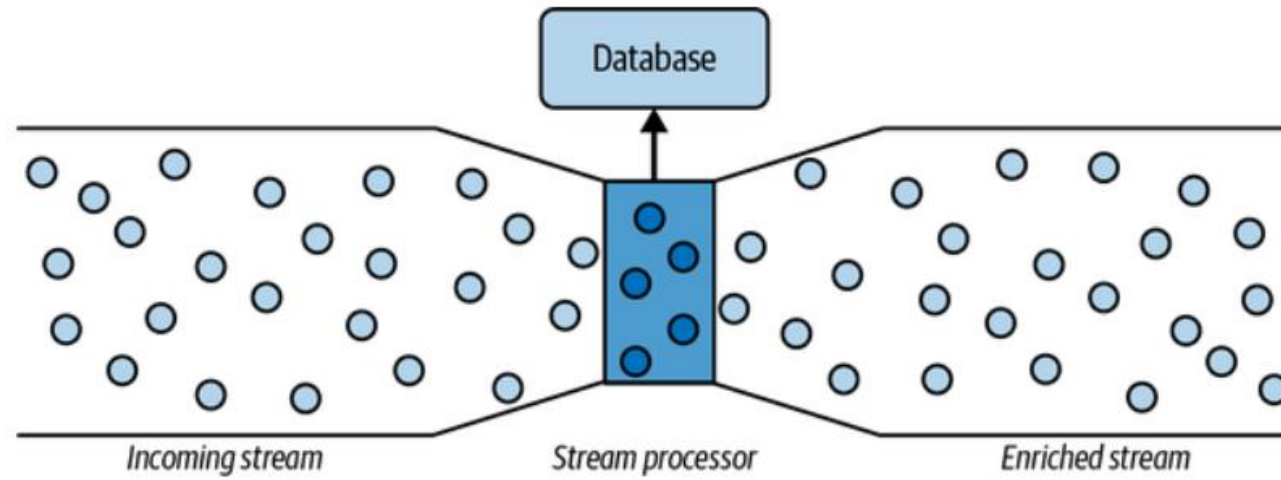
### **Modelo Pull**

Los datos se recuperan del sistema de origen, ejemplo: ETL.

# Transformación

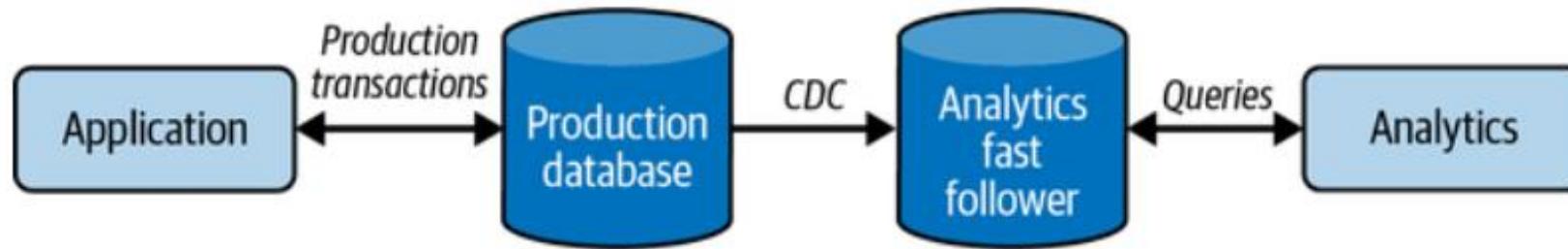


# Transformación



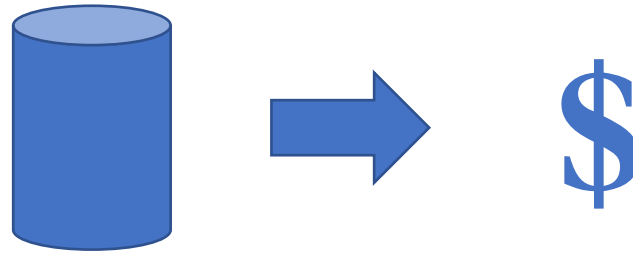
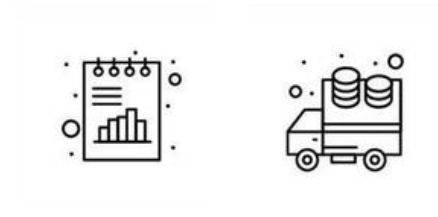


# Transformación



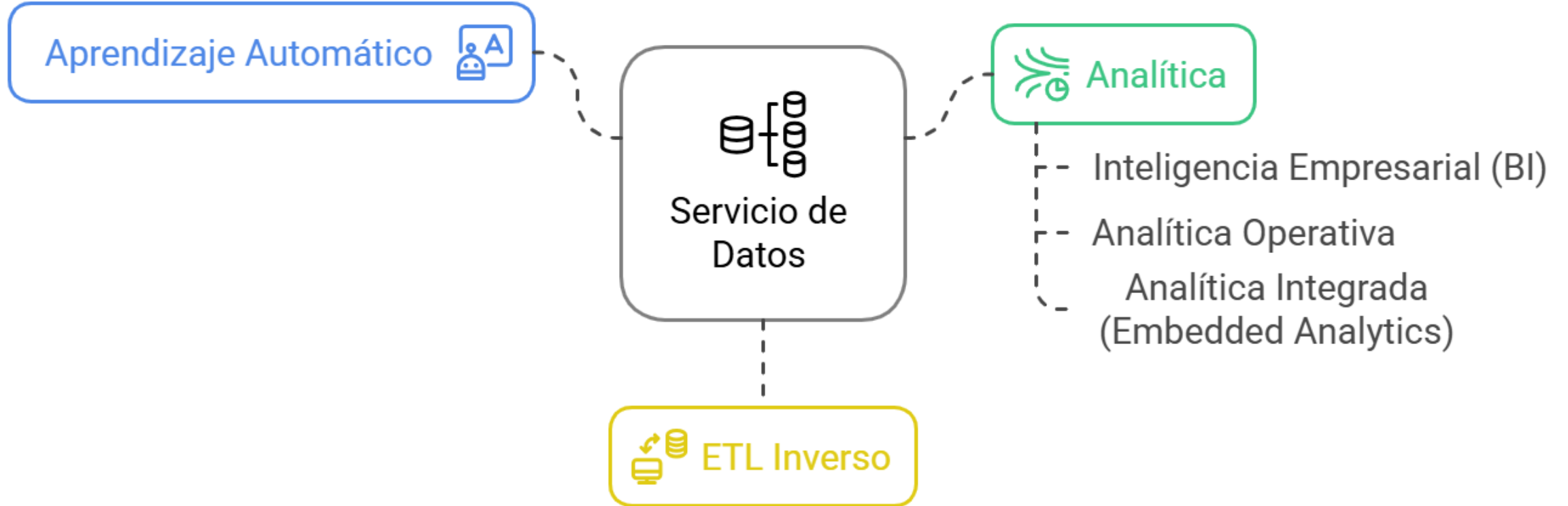
# Serving

Es hora de obtener valor de los datos

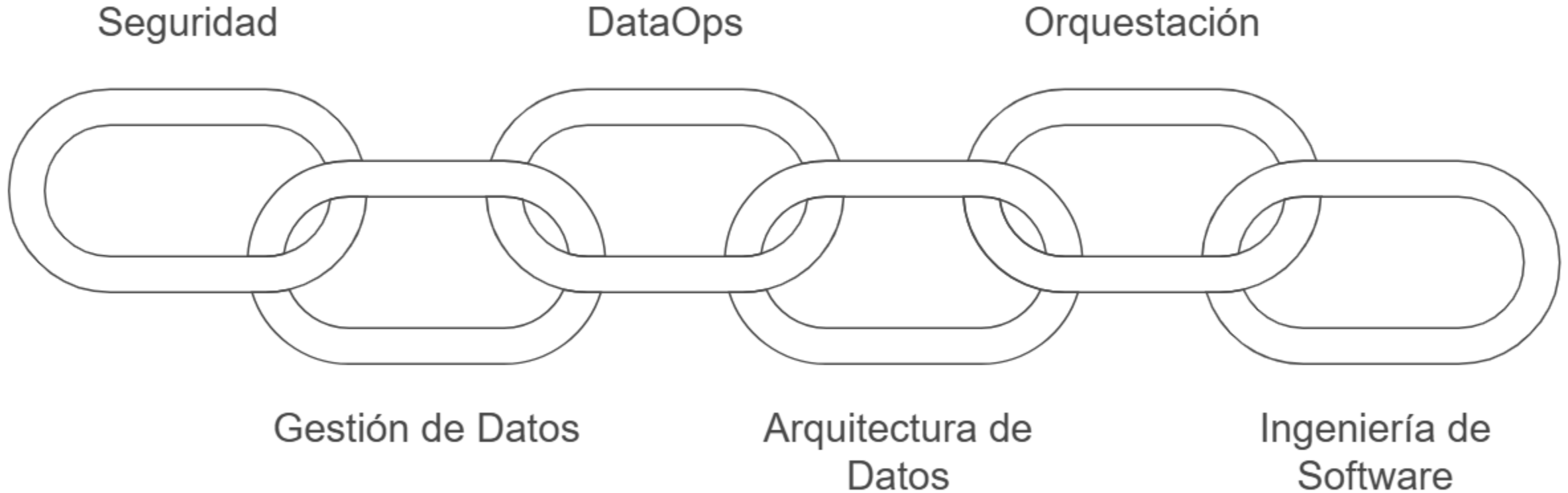


# Serving

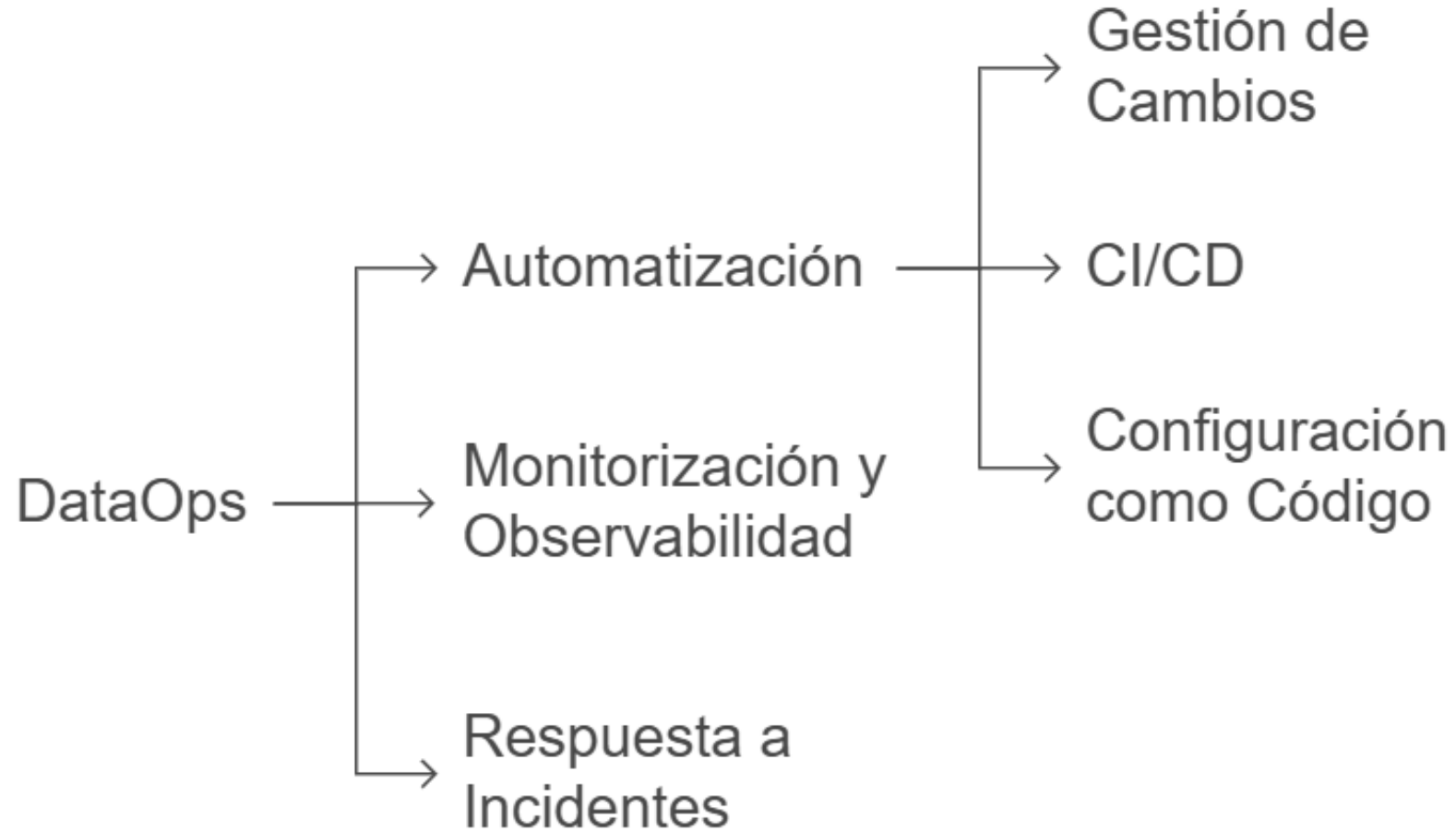
## Ejemplos de Servicio de Datos



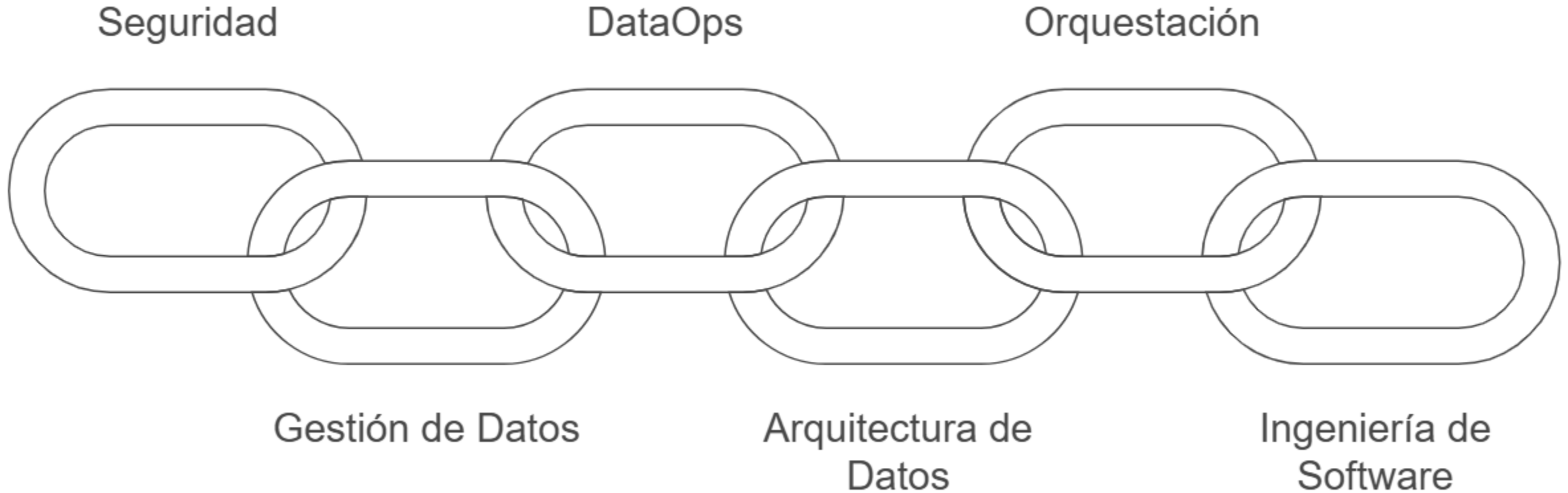
# Corrientes Subyacentes en el Ciclo de Vida de la Ingeniería de Datos



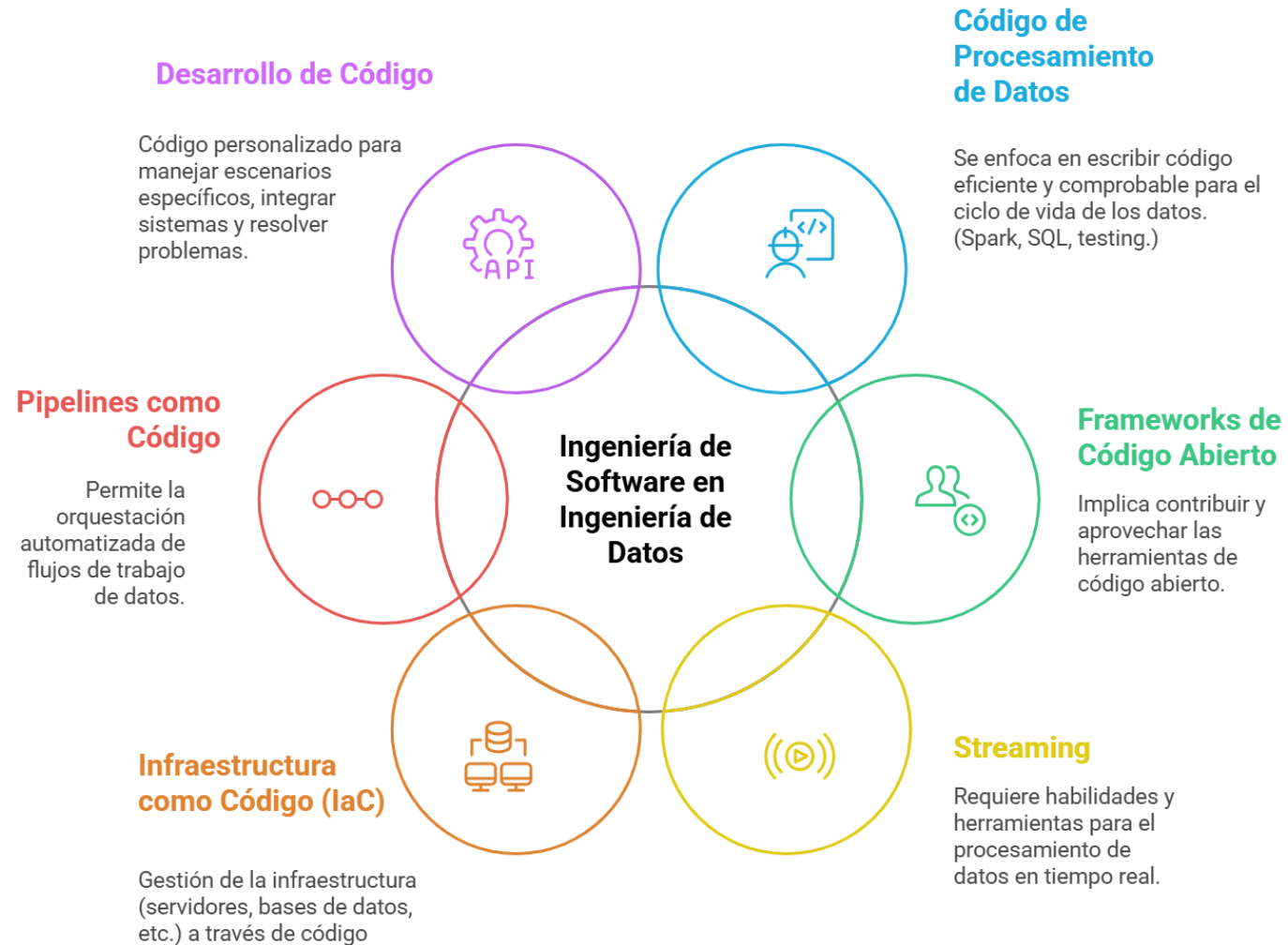
# Prácticas de DataOps

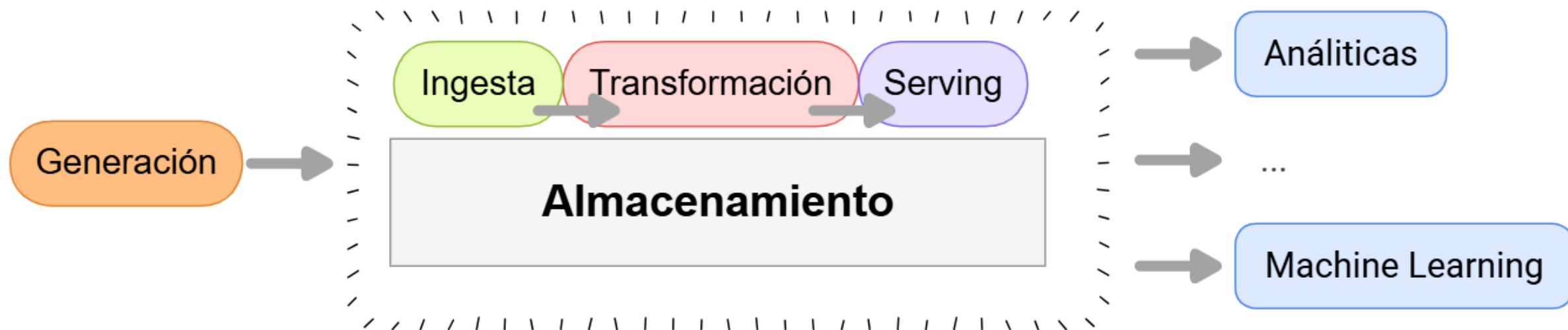


# Corrientes Subyacentes en el Ciclo de Vida de la Ingeniería de Datos



# Áreas comunes con la Ingeniería de Software





---

Seguridad | Gestión de Datos | DataOps | Arquitectura de Datos | Orquestación | Ingeniería de Software