

# Sesión 3: Almacenamiento de datos

Una Factoría de Información Corporativa (CIF) es una arquitectura de datos que permite a las organizaciones transformar grandes volúmenes de datos dispersos en información estratégica, confiable y útil para la toma de decisiones. Funciona como una "fábrica de datos", donde se recolecta, limpia, integra, almacena y distribuye la información de forma sistemática y controlada.

**R** por Kibernetum Capacitación S.A.

DATA LAKE

# ¿Qué es una Factoría de Información Corporativa (CIF)?

Una Factoría de Información Corporativa (CIF) es una arquitectura de datos que permite a las organizaciones transformar grandes volúmenes de datos dispersos en información estratégica, confiable y útil para la toma de decisiones. Funciona como una "fábrica de datos", donde se recolecta, limpia, integra, almacena y distribuye la información de forma sistemática y controlada.

Antes de profundizar en las características clave de una CIF, Observa la imagen 1, acá se ilustra un ecosistema moderno de arquitectura de datos que refleja los principios de una Factoría de Información Corporativa (CIF). A la izquierda, se observan las fuentes de datos dispersas (nube, bases de datos, sistemas ERP como SAP, archivos), que son ingestadas por herramientas como Fivetran, Stitch o Airbyte.

Luego, los datos se almacenan en un Data Lake o Data Warehouse, donde pasan por procesos de transformación (con herramientas como dbt, Spark o Dataform) y orquestación (Airflow, Dagster, etc.). Finalmente, se distribuyen para análisis, visualización (Power BI, Tableau, etc.) o integración inversa (Reverse ETL) hacia sistemas operacionales.

La imagen representa gráficamente cómo una CIF funciona como una "fábrica de datos": desde la captura hasta la entrega del dato estratégico, pasando por almacenamiento, catalogación y transformación controlada, interesante ¿Verdad?

# Características clave de una CIF

## Centralización de datos

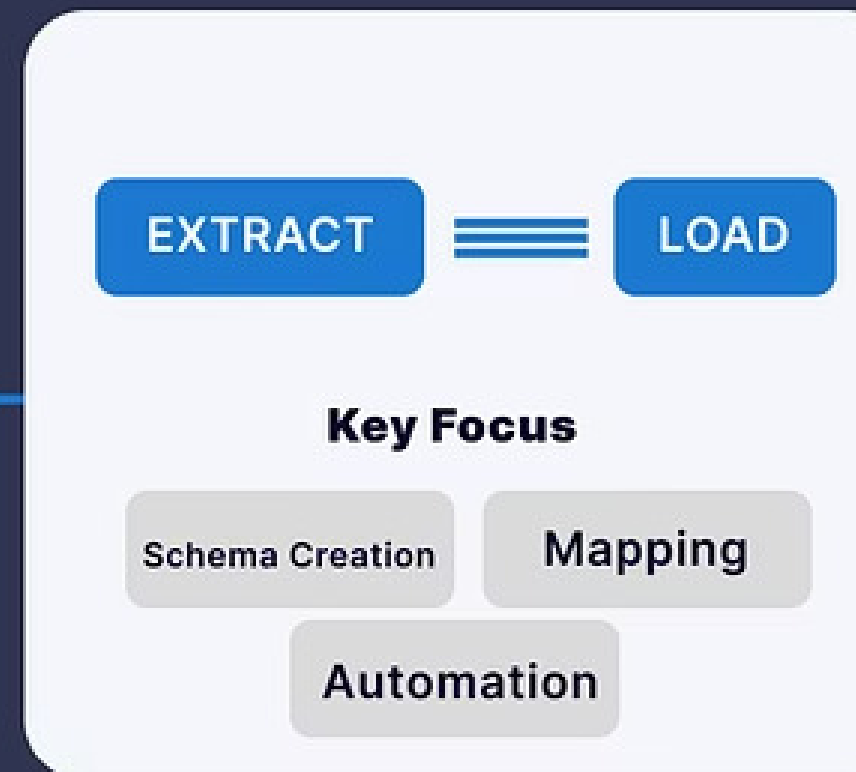
La CIF consolida información proveniente de múltiples fuentes —como sistemas operacionales, archivos externos, aplicaciones web o bases de datos heredadas— en un único entorno integrado.

## Transformación y calidad de datos

Los datos que ingresan a la CIF se transforman mediante procesos de limpieza, validación y enriquecimiento. Esto asegura que la información utilizada por los analistas y tomadores de decisiones sea precisa, consistente y completa.

## Disponibilidad para análisis

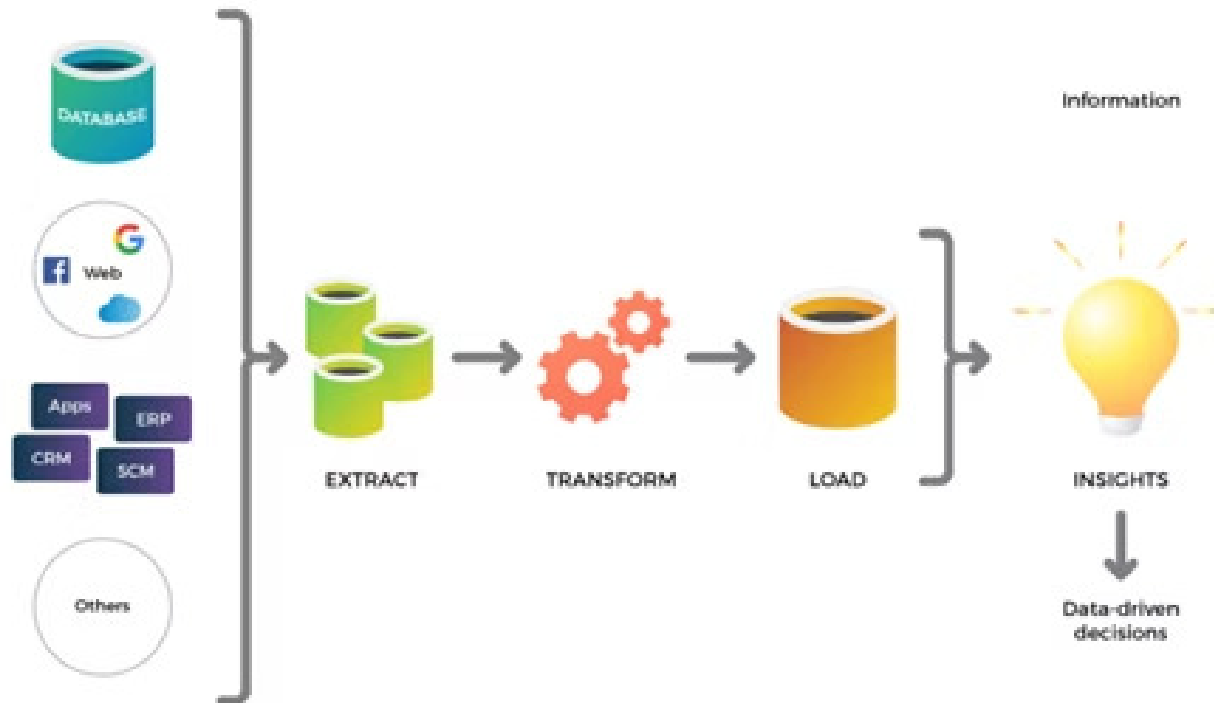
Una vez transformados y organizados, los datos se almacenan en estructuras diseñadas para facilitar el análisis, como Data Marts y Data Warehouses. Estas estructuras permiten realizar consultas eficientes, obtener reportes y alimentar herramientas de inteligencia de negocios (BI).



# Automatización de procesos

La CIF permite automatizar tareas rutinarias como la carga de datos (ETL), la generación de reportes y el monitoreo de calidad, lo que reduce errores humanos y mejora la eficiencia operativa.

Imagen 3: "Flujo automatizado de ETL en la CIF: datos extraídos de múltiples fuentes son transformados y cargados sin intervención manual, generando insights para decisiones ágiles."



# Acceso controlado y seguro

La arquitectura de la CIF contempla políticas de seguridad, control de accesos y trazabilidad, lo cual garantiza que solo usuarios autorizados puedan consultar o modificar la información.

Imagen 4



## DATA SECURITY ACCESS CONTROL



### User Permissions

John Smith

Read ☒

Write ☐

Delete ☐

### Date Access Pol

☐ Confidential Data

☐ Restricted Data

☐ Public Data

Read

Write

D

Read

Write

D

☒

☒

☐

☒

☐

☐

# Componentes principales

Una CIF moderna está compuesta por varios elementos clave:



## Fuentes de datos

Sistemas transaccionales, hojas de cálculo, aplicaciones externas, sensores IoT, etc.



## Procesos ETL

Extracción, Transformación y Carga: encargados de preparar los datos para su análisis.



## Almacenamiento estructurado

Típicamente, un Data Warehouse corporativo.



## Data Marts

Subconjuntos de datos orientados a áreas específicas del negocio (marketing, finanzas, operaciones).



## Herramientas de BI

Para exploración, visualización y reporte de datos.



## Repositorio de metadatos

Para mantener la documentación técnica y de negocio sobre los datos.



## Plataformas de acceso y distribución

Dashboards, portales analíticos, APIs, etc.

# Beneficios de implementar una CIF



**Mejor calidad y coherencia de los datos**



**Mayor capacidad para tomar decisiones informadas y basadas en evidencias**



**Reducción del tiempo y esfuerzo en la preparación de reportes**

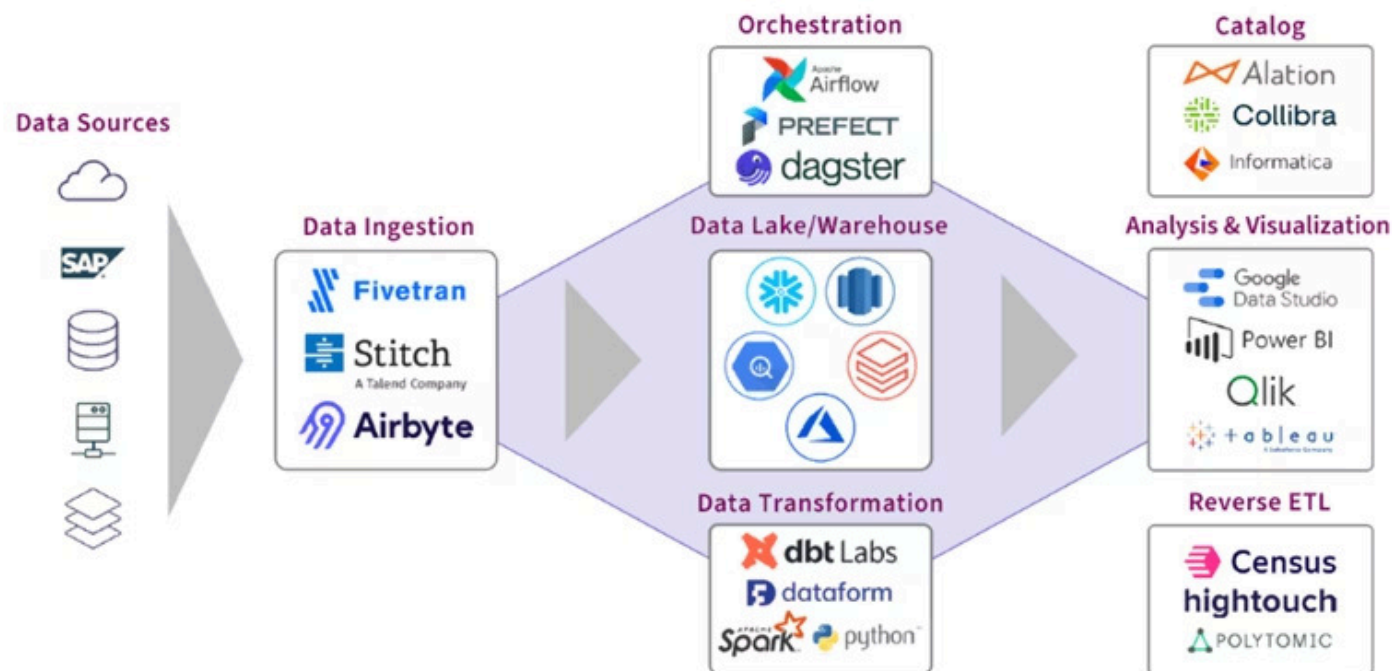


**Mayor autonomía para las áreas de negocio en el análisis de información**



**Escalabilidad y flexibilidad para adaptarse al crecimiento de la organización**

Imagen 1: Referencia: modern data stack





# Objetivos y principios de la CIF

Una CIF no solo es una arquitectura tecnológica: es una estrategia corporativa para transformar datos en activos de información. Su diseño responde a una serie de objetivos funcionales y principios orientadores que aseguran su efectividad dentro de una organización moderna.

## Objetivos de una CIF

Los principales objetivos de una CIF bien implementada incluyen:

### **Integración de fuentes de datos heterogéneas**

Reunir datos provenientes de distintas plataformas (ERP, CRM, sensores, archivos externos, etc.) en una estructura común, independientemente del formato o ubicación original.

### **Normalización y transformación de los datos**

Estandarizar y depurar los datos para garantizar uniformidad, coherencia y valor analítico. Este proceso es esencial para convertir datos "en bruto" en información confiable.

### **Almacenamiento histórico estructurado**

Conservar datos históricos y actuales en un único repositorio central (Data Warehouse), facilitando el análisis longitudinal y la trazabilidad de cambios en el tiempo.



# Más objetivos de una CIF

## Mejora de la calidad y trazabilidad

Aplicar controles de calidad para asegurar integridad, precisión, completitud y disponibilidad de los datos. Además, mantener un registro del linaje de cada dato: de dónde proviene, cómo se ha transformado y dónde se utiliza.

## Accesibilidad a través de herramientas de análisis

Permitir que los usuarios de negocio — como analistas, gerentes o equipos de BI — accedan fácilmente a la información mediante dashboards, consultas SQL, reportes automatizados o herramientas de visualización.

## Apoyo a la toma de decisiones

Transformar datos en conocimiento que sustente decisiones operativas, tácticas y estratégicas basadas en evidencia.

# Principios fundamentales de una CIF

Además de sus objetivos funcionales, toda Factoría de Información Corporativa se guía por principios clave que aseguran su estabilidad y escalabilidad:

Principio	Descripción
Centralización	Los datos se concentran en un entorno único que actúa como fuente de verdad.
Separación operativa/analítica	Se distingue entre los sistemas transaccionales (OLTP) y analíticos (OLAP).
Estandarización	Se aplican reglas uniformes de nombres, formatos y estructuras de datos.
Modularidad	La CIF se construye por capas, cada una con funciones específicas y bien definidas.
Escalabilidad	La arquitectura puede crecer con el tiempo sin perder rendimiento ni control.
Gobernanza de datos	Se aplican políticas, roles y responsabilidades claras para el manejo de datos.

# Conceptos esenciales de la CIF

Una Factoría de Información Corporativa (CIF) es mucho más que un simple repositorio de datos. Es un ecosistema completo de componentes interrelacionados que trabajan en conjunto para convertir datos dispersos en información organizada, confiable y útil para el negocio.

A continuación, se presentan los conceptos clave que definen una CIF, explicados de forma clara y didáctica.

## Datos operacionales vs. datos analíticos

Una CIF establece una separación clara entre ambos tipos de datos, permitiendo que cada uno cumpla su función sin interferir con el otro.

### Datos operacionales (OLTP)

Son los datos que generan los sistemas transaccionales (ERP, CRM, POS, etc.) y que reflejan operaciones en tiempo real como por ejemplo ventas, pagos, reservas, etc.

### Datos analíticos (OLAP)

Son los datos integrados, transformados y almacenados para análisis. Se utilizan para generar reportes, explorar tendencias o tomar decisiones estratégicas.

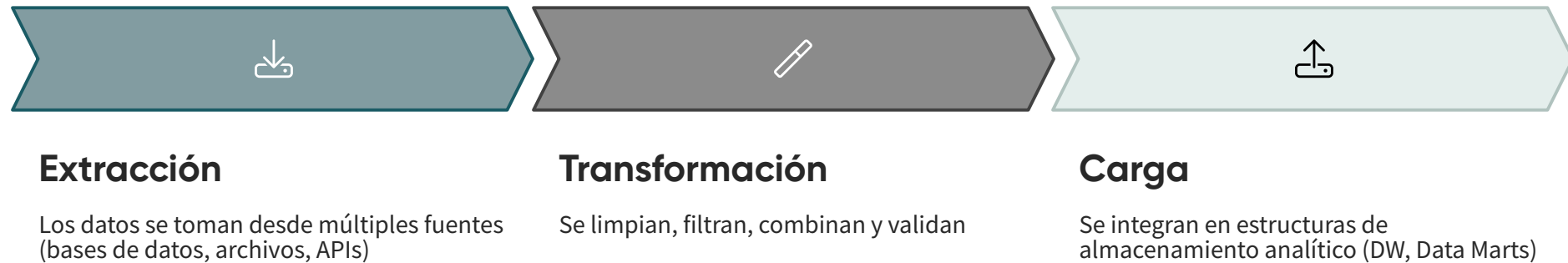
# Comparativa: Datos operacionales vs. datos analíticos

Observa esta tabla que nos ayudara a comprender mejor las diferencias:

Característica	Datos operacionales (OLTP)	Datos analíticos (OLAP)
Origen	Sistemas transaccionales (ERP, CRM, POS, e-commerce, etc.)	Almacenamiento estructurado: Data Warehouse, Data Marts
Finalidad	Registrar operaciones en tiempo real	Analizar tendencias, comportamientos, KPIs
Frecuencia de actualización	Alta (cada segundo/minuto)	Periódica (cada día, semana, mes)
Estructura de archivos	Cruda, detallada, tabla por evento o transacción	Agrupada, resumida, estructurada por dimensiones
Nombre típico de archivos	ventas_2024_03_31.csv, pagos_hoy.json, ordenes.sql	reporte_mensual_ventas.csv, kpi_financieros.xlsx, dw_ventas.parquet
Ejemplo de tabla	ordenes, clientes, pagos, facturas	hechos_ventas, dim_producto, reporte_mensual_regiones
Ejemplo de dato	{"fecha": "2025-03-31", "cliente_id": 483, "total": 24990}	Ventas marzo 2025 = \$893.500 distribuidas en 5 regiones

# Flujo de datos en la CIF

El flujo de información en una Factoría de Información sigue una secuencia lógica de etapas:



Este proceso se conoce como ETL (Extract, Transform, Load), aunque en arquitecturas modernas puede aplicarse como ELT (carga antes de transformar).

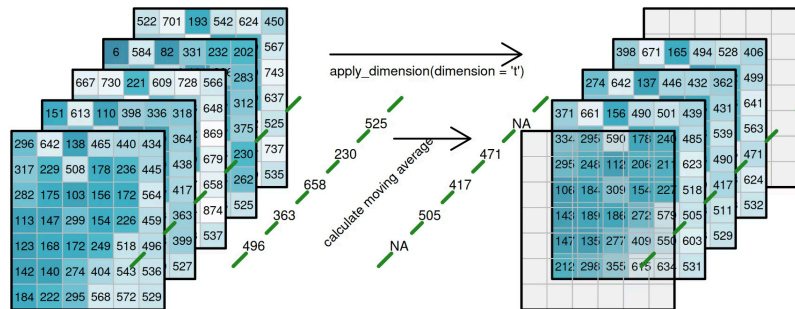
# Modelos de datos

Dentro de una CIF se utilizan modelos de datos específicos para facilitar el análisis:

## Modelo multidimensional

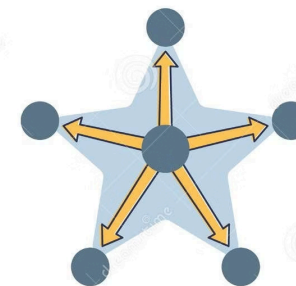
Basado en hechos (medidas numéricas) y dimensiones (categorías de análisis). Ej: ventas (hecho) por región y producto (dimensiones).

Es ideal para análisis OLAP.

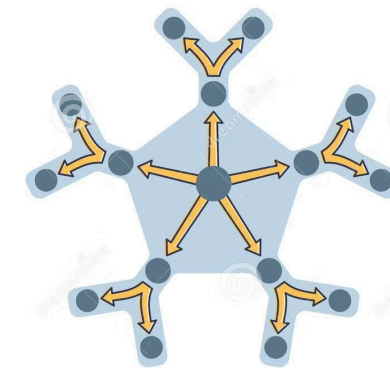


## Modelo en estrella o snowflake

Son los esquemas más comunes en el diseño de Data Warehouses.  
Permiten estructurar la información para facilitar consultas agregadas y rápidas.



## STAR SCHEMA



## SNOWFLAKE SCHEMA

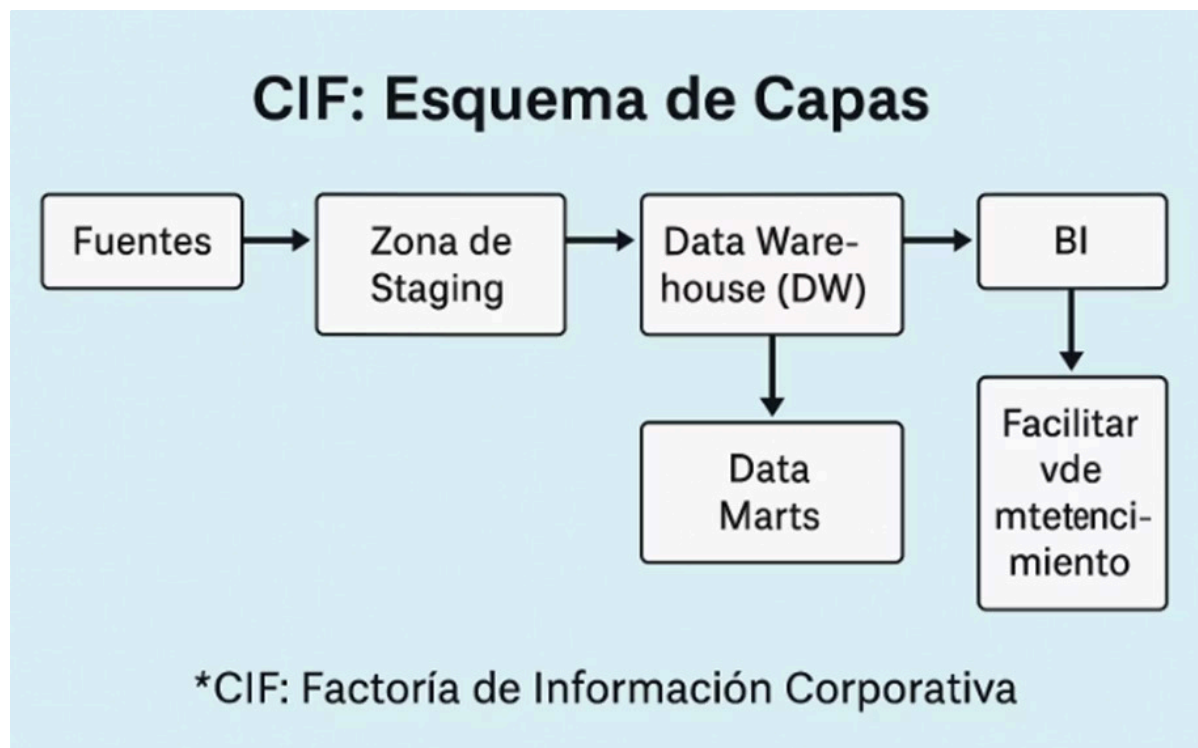
# Áreas funcionales

Una CIF contempla distintas áreas o zonas funcionales que permiten segmentar y organizar el procesamiento de datos:

Zona / área	Función principal
Staging area	Recepción y carga temporal de datos crudos
Data warehouse corporativo	Almacenamiento central de datos consolidados y confiables
Data marts	Subconjuntos de datos por unidad de negocio (ej: ventas, finanzas)
Herramientas de acceso	Reportes, dashboards, consultas ad hoc

Estas áreas no son compartimentos estancos, sino partes de una cadena integrada de procesamiento de datos.

Imagen 5: Representación visual de capas CIF





# Gobernanza y calidad

Recuerda que una CIF debe incorporar políticas de:



## Gobernanza de datos

Define roles, responsabilidades, permisos y procesos.



## Calidad de datos

Establece controles para asegurar que los datos sean útiles, precisos y confiables.



## Metadatos

Documentación que describe cada conjunto de datos, su origen, su uso y su transformación.

# Capas de una CIF

Una Factoría de Información Corporativa (CIF) está organizada en capas funcionales, cada una con un propósito específico dentro del flujo de procesamiento y aprovechamiento de los datos.

Esta estructura por capas permite:

- Separar responsabilidades técnicas.
- Mejorar el rendimiento de cada proceso.
- Aumentar la gobernabilidad de la información.
- Escalar la arquitectura a medida que la organización crece.



## Capa Operacional

Capturar los datos desde los sistemas fuente



## Capa de Staging

Recibir los datos en su forma original y prepararlos



## Capa de Data Warehouse

Almacenar datos limpios, transformados y estructurados



## Capa de Data Marts

Segmentar y personalizar los datos por áreas

5

## Capa de Herramientas BI

Permitir el acceso, visualización y análisis

# Capa Operacional

## Propósito

Capturar los datos desde los sistemas fuente.

## Fuentes

- ERP
- CRM
- Sensores IoT
- Sistemas contables
- Redes sociales

## Tipo de dato

Generalmente estructurado y en tiempo real.

## Ejemplo

Una venta registrada en el sistema POS del supermercado.



Esta capa no forma parte del sistema analítico, pero es el punto de partida del flujo de datos.

# Capa de Staging (Área de carga temporal)

## Propósito

Recibir los datos en su forma original y prepararlos para su procesamiento.

## Procesos comunes

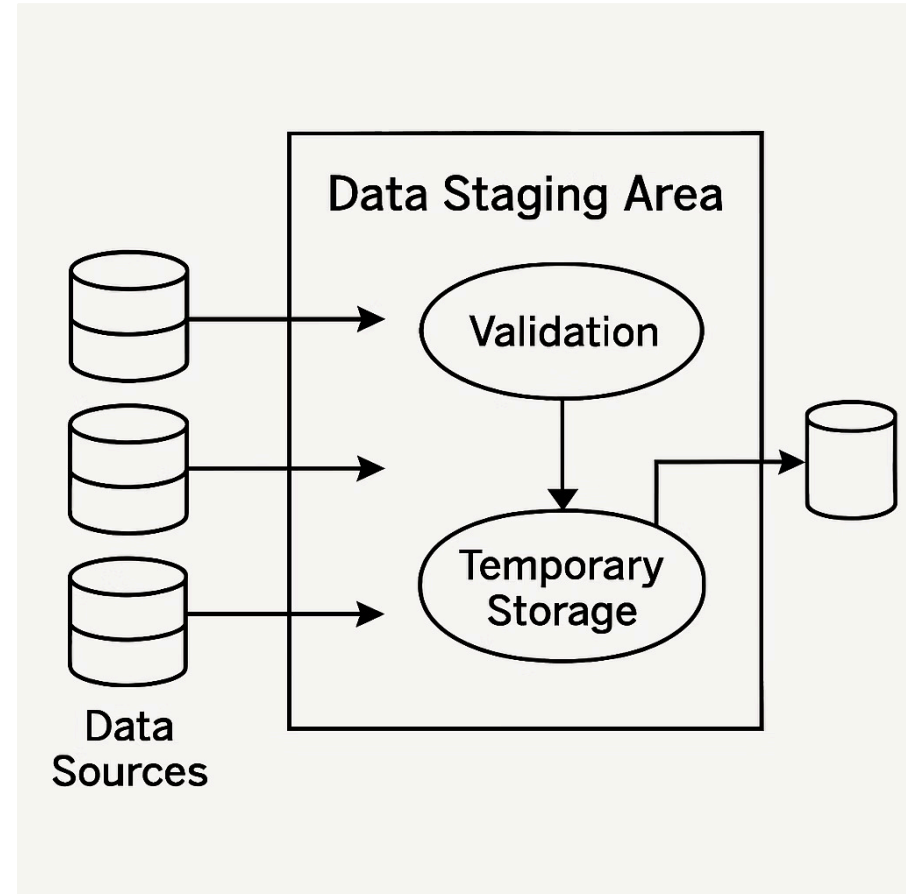
- Validación sintáctica.
- Corrección de errores.
- Normalización de formatos.

## Temporalidad

Los datos no se almacenan permanentemente aquí.

## Ejemplo

Recibir archivos CSV exportados desde un ERP.



También se conoce como área de tránsito o zona de staging.

# Capa de Data Warehouse (DW)

## Propósito

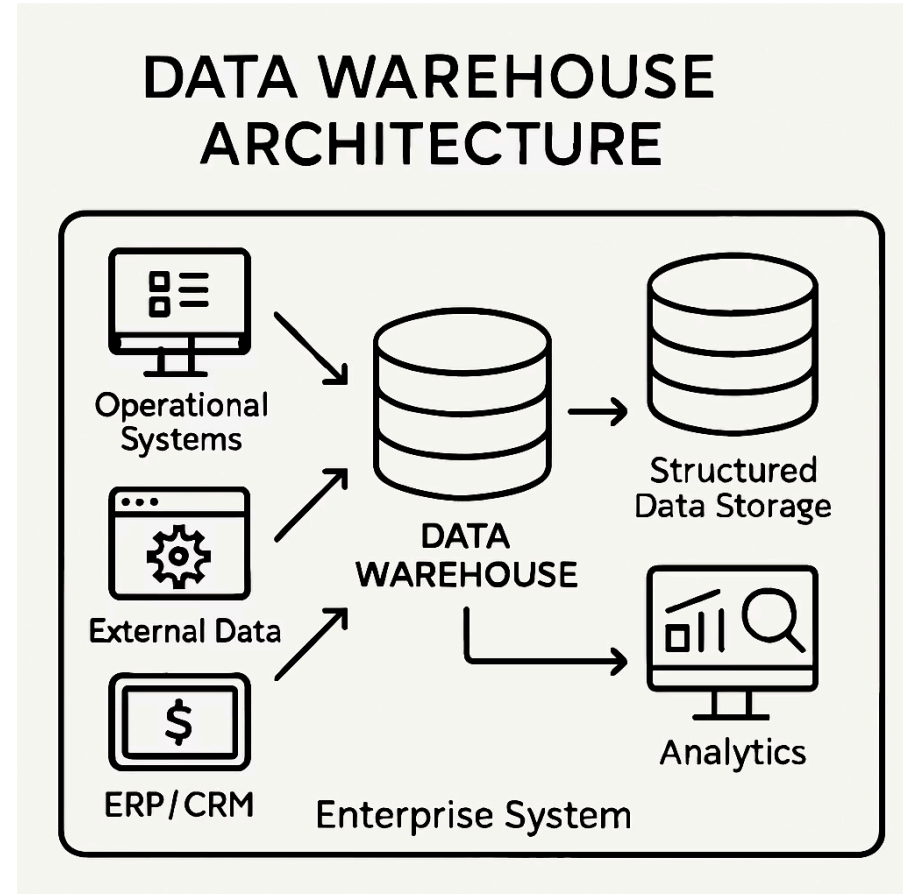
Almacenar datos limpios, transformados y estructurados para análisis históricos.

## Características

- Alta integridad.
- Datos integrados y normalizados.
- Optimizado para consultas analíticas complejas (OLAP).

## Ejemplo

Tabla de ventas históricas consolidadas con dimensiones geográficas y de producto.



Esta es la columna vertebral de la CIF.

# Capa de Data Marts

## Propósito

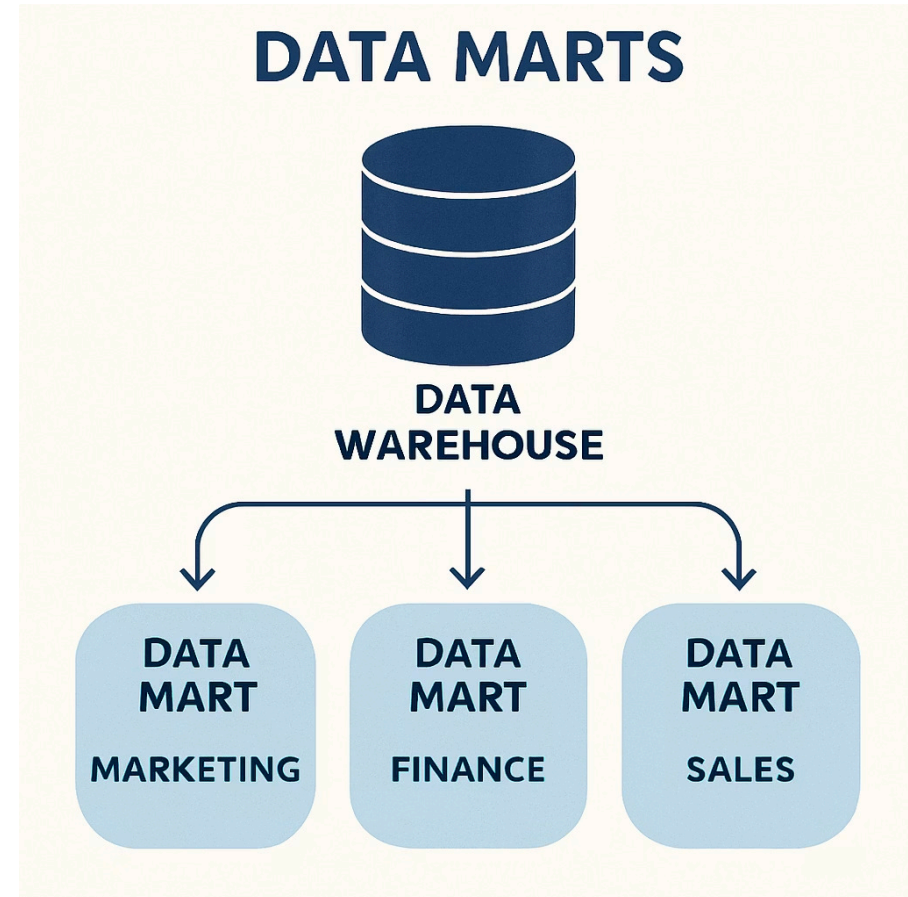
Segmentar y personalizar los datos por áreas de negocio.

## Ventajas

- Rapidez en las consultas.
- Autonomía para equipos funcionales.

## Ejemplo

Un Data Mart de Marketing con métricas de campañas digitales.



Los Data Marts pueden derivarse directamente del DW (modelo dependiente) o alimentarse desde otras fuentes (modelo independiente).

# Capa de Herramientas BI (Business Intelligence)

## Propósito

Permitir el acceso, visualización y análisis de la información.

## Herramientas comunes

- Power BI
- Tableau
- Google Data Studio
- QlikView

## Funciones

- Dashboards interactivos.
- Reportes programados.
- Consultas ad hoc.



Es la capa donde los usuarios de negocio interactúan con los datos.



# Capa de Metadatos

## Propósito

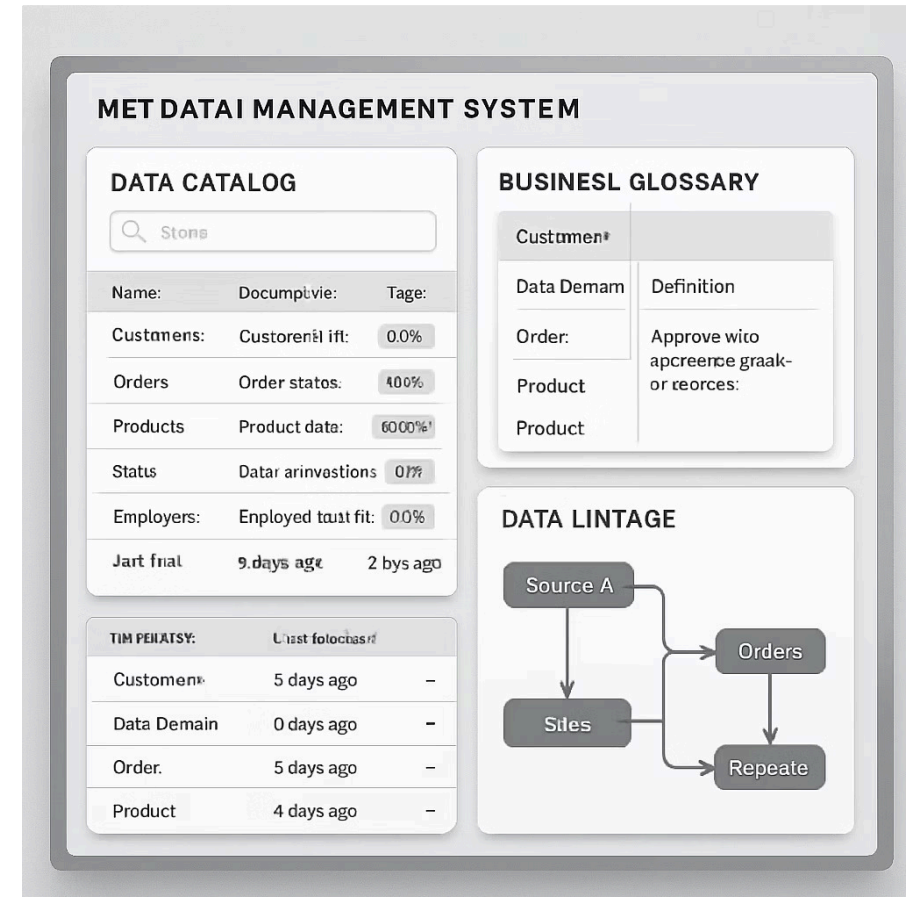
Documentar el significado, origen, transformación y uso de los datos.

## Tipos de metadatos

- Técnicos (tipos de dato, estructuras).
- De negocio (glosario, definiciones).
- Operacionales (linaje, auditoría).

## Herramientas comunes

- Catálogo de datos
- Glosarios empresariales



Esta capa mejora la gobernanza, la trazabilidad y la confianza en los datos.

# Data Lake

## Definición y propósito

Un Data Lake es un repositorio centralizado que permite almacenar grandes volúmenes de datos en su formato original: ya sean estructurados (tablas), semiestructurados (JSON, XML, CSV) o no estructurados (imágenes, audios, textos, logs, videos, etc.).



Imagen 6: "Un Data Lake es capaz de almacenar y gestionar múltiples tipos de datos: desde imágenes y audios, hasta documentos estructurados y logs de aplicaciones, sin necesidad de transformarlos previamente."

A diferencia de un Data Warehouse, que requiere definir un esquema y una estructura previa, el Data Lake acepta los datos tal como llegan, permitiendo su procesamiento posterior según su uso.

Propósito principal: servir como base flexible y escalable para el análisis exploratorio, avanzado y predictivo, especialmente en contextos de Big Data y Machine Learning.

# ¿Un Data Lake es un software?

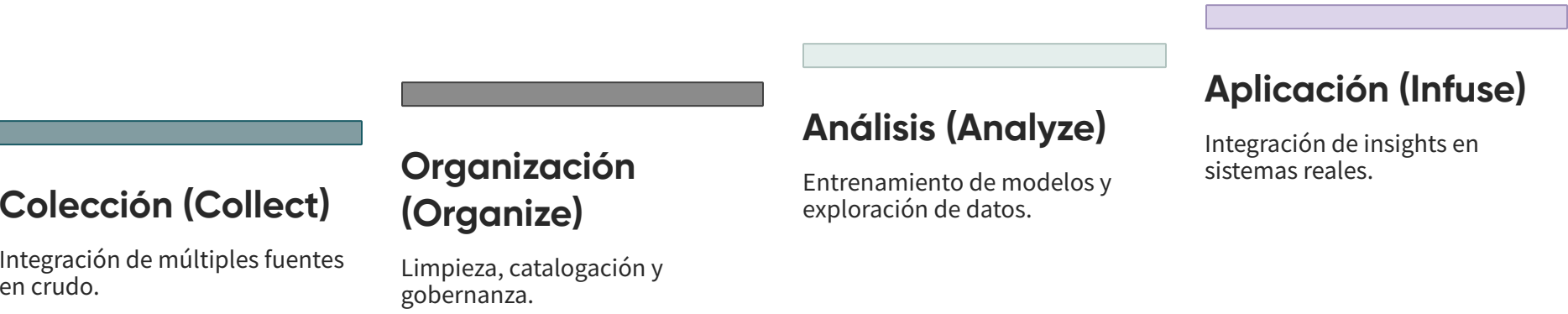
No, un Data Lake no es una herramienta única, sino una arquitectura compuesta por varias tecnologías que trabajan juntas:

Componente	Ejemplos comunes
Almacenamiento	Amazon S3, Azure Data Lake Storage, Google Cloud Storage
Procesamiento	Apache Spark, Databricks, AWS Glue, Hadoop
Consulta	Athena, Presto, Hive, Notebooks Jupyter
Gobernanza	AWS Lake Formation, Apache Ranger, Amundsen, Glue Data Catalog

Esta arquitectura modular permite construir soluciones a medida, escalables y adaptables.

## El Data Lake como habilitador del "AI Ladder" (IBM)

Según IBM, un Data Lake soporta cada paso de la escalera de madurez en IA:



El Data Lake no es solo un contenedor de datos: es una plataforma para escalar hacia soluciones de inteligencia artificial.

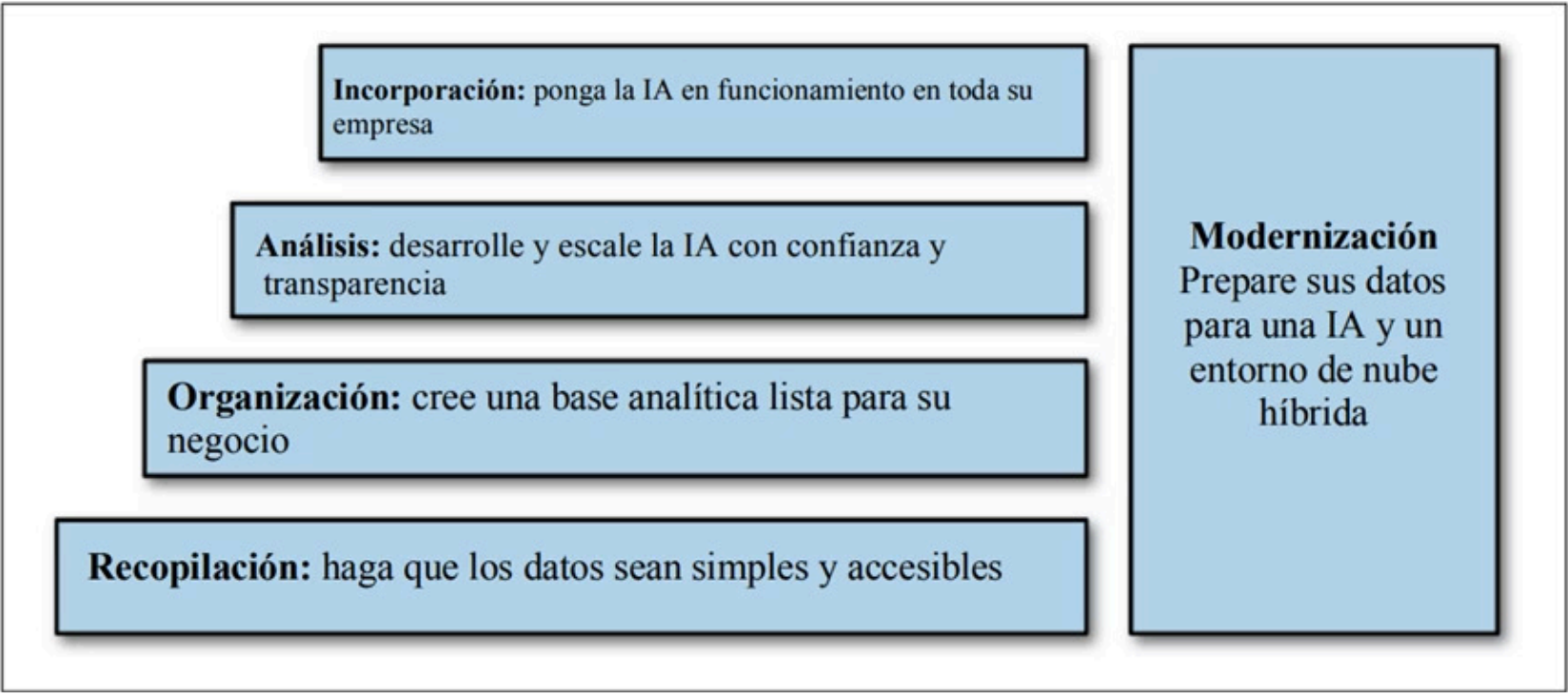


Imagen 7: Escalera de la IA de IBM (AI Ladder) - esta escalera representa los pasos fundamentales para alcanzar la madurez en el uso de Inteligencia Artificial. Cada etapa puede ser soportada por una arquitectura basada en Data Lake, desde la recopilación hasta la incorporación de IA en sistemas reales.

# Características del Data Lake

Característica	Descripción
Escalabilidad	Pensado para crecer masivamente, ideal en entornos cloud.
Flexibilidad	Acepta datos estructurados, semiestructurados y no estructurados.
Bajo costo	El almacenamiento en la nube es más económico que soluciones tradicionales.
Sin esquema previo	No requiere modelo relacional al momento de la carga.
Integración analítica	Compatible con herramientas de ciencia de datos y entrenamiento de modelos.

El Data Lake es el punto de partida para estrategias de Machine Learning, IA y minería de datos.

## Arquitectura de capas

Los Data Lakes se estructuran en zonas (layers) que representan el nivel de preparación de los datos:

Zona	Propósito
Raw Zone	Datos sin procesar, tal como llegan de la fuente.
Cleansed Zone	Datos limpios, validados, sin errores ni duplicados.
Curated Zone	Datos transformados, listos para análisis o visualización.
Analytics Zone	Datasets específicos para dashboards, IA, o reportes.

# Arquitectura de capas del Data Lake



Imagen 8: Esquema de capas en un Data Lake: desde la zona Raw hasta la zona Analytics, ilustrando el flujo de datos desde su captura hasta su uso en análisis y visualización.

## Ventajas y desventajas

### Ventajas

- Bajo costo y alta flexibilidad
- Admite múltiples formatos
- Escalable y cloud-native
- Base ideal para proyectos de ciencia de datos

### Desventajas

- Riesgo de convertirse en un Data Swamp sin control
- Requiere habilidades técnicas para su administración
- No es óptimo para reportes estructurados frecuentes
- Gobernanza y linaje complejos si no se planifican

Este es un momento para hablar de Data Swamp que no es más que un Data Lake mal gestionado y desorganizado que pierde su valor analítico.

# Casos de uso reales de Data Lakes

Como hemos visto, los Data Lakes son ampliamente utilizados en industrias que requieren gestionar volúmenes masivos de datos diversos y de múltiples orígenes. Su capacidad para almacenar información en formatos crudos y variados los convierte en una solución esencial en proyectos de transformación digital, analítica avanzada e inteligencia artificial.

A continuación, se profundiza en cuatro casos de uso reales que ilustran el valor de los Data Lakes en distintos sectores:

## Logs web y comportamiento de usuario

Los sitios web y plataformas digitales generan millones de eventos por segundo: clics, scrolls, búsquedas, rutas de navegación, errores de carga, etc. Estos logs se almacenan inicialmente en un Data Lake para luego ser utilizados en análisis de comportamiento, personalización y optimización de experiencia de usuario.

Por ejemplo, Netflix almacena logs de navegación y reproducción en un Data Lake en AWS S3. Con esta información entrena modelos de recomendación personalizados para cada usuario.

## Internet de las cosas (IoT)

Los dispositivos conectados —como sensores industriales, medidores inteligentes, dispositivos médicos o sistemas de transporte— generan grandes cantidades de datos en tiempo real. Un Data Lake permite centralizar esta información en crudo para análisis posterior.

Siemens utiliza Data Lakes en Azure para almacenar datos de sensores en plantas industriales. Estos datos se usan para predecir fallas en equipos mediante algoritmos de mantenimiento predictivo.

# Más casos de uso de Data Lakes

## Multimedia sin estructurar

Los Data Lakes permiten almacenar información no estructurada como imágenes, videos, grabaciones de audio y textos libres. Este contenido puede ser procesado posteriormente con técnicas de NLP (procesamiento de lenguaje natural), visión computacional o análisis de sentimientos.

El Hospital Mount Sinai (EE.UU.) utiliza un Data Lake en GCP para almacenar imágenes de resonancias magnéticas, registros de voz de médicos y notas clínicas. Con estos datos se entrenan modelos para la detección temprana de enfermedades neurodegenerativas.

## Modelos predictivos y Machine Learning

El entrenamiento de modelos de IA requiere grandes volúmenes de datos diversos, históricos y crudos. Los Data Lakes permiten gestionar estos datasets de entrenamiento, muchas veces enriquecidos desde múltiples fuentes internas y externas.

Airbnb consolida datos de búsqueda, reservas, opiniones y comportamiento de usuarios en un Data Lake. Luego, utiliza esa información para predecir precios dinámicos, identificar fraudes o personalizar la búsqueda.

Un Data Lake no es solo una "bodega de datos". Es el motor que permite unificar, explorar y explotar el valor de la información en cualquier organización, sentando las bases para el uso eficiente de modelos analíticos, inteligencia artificial y toma de decisiones basada en datos.



# Data Warehouse

## a. Propósito

Un **Data Warehouse (almacén de datos)** es un repositorio centralizado y estructurado donde se almacenan datos **transformados, limpios y organizados específicamente para el análisis**.

Su función principal es proveer una **única fuente de verdad** para la organización, que pueda ser consultada por analistas, gerentes y herramientas de **Business Intelligence (BI)**. Se diferencia del Data Lake porque **requiere estructura previa** y reglas estrictas de **calidad, integridad y consistencia**.



**Imagen 14:** El diagrama representa el proceso típico de integración de datos mediante ETL (Extracción, Transformación y Carga), su almacenamiento en un Data Warehouse estructurado, y su posterior explotación a través de herramientas de Business Intelligence como Power BI, Tableau o Looker.

**Desde el enfoque de IBM (2024), el Data Warehouse actúa como el corazón analítico de la empresa:** una fuente consolidada que extrae datos desde múltiples sistemas (**CRM, ERP, ventas, supply chain**), los transforma y los pone a disposición para BI, IA o análisis predictivo.

## b. Arquitectura

Un **Data Warehouse moderno** se construye sobre una arquitectura lógica y tecnológica que cumple con los siguientes elementos:

Elemento	Descripción
Integración de fuentes	Combina múltiples orígenes de datos (ERP, CRM, e-commerce, RRHH).
Almacenamiento optimizado	Utiliza estructuras de almacenamiento por columnas (columnar storage) y técnicas de compresión para acelerar consultas.
Acceso estructurado	Permite acceso por SQL, herramientas BI, y dashboards personalizados.
Modelado dimensional	Organiza la información en esquemas tipo estrella o copo de nieve.
Temporalidad histórica	Conserva versiones y datos históricos para análisis longitudinal.

Plataformas como **Snowflake, Google BigQuery, Amazon Redshift** o **IBM Db2 Warehouse** ofrecen soluciones modernas y escalables de Data Warehousing tanto **en la nube** como **on-premise**.

Según IBM, un Data Warehouse puede implementarse utilizando modelos de procesamiento como:

- **MPP (Procesamiento Paralelo Masivo):** Escala horizontalmente agregando nodos.
- **SMP (Multiprocesamiento Simétrico):** Basado en múltiples CPUs que comparten memoria.

## c. Capas de un Data Warehouse

Al igual que en la CIF, el Data Warehouse también se organiza en **capas funcionales** que facilitan su operación y mantenimiento:

Capa	Función principal
ETL Layer	Extracción, transformación y carga de datos desde fuentes externas.
Enterprise DW	Almacenamiento principal de datos integrados, normalizados y confiables.
Presentation Layer	Vistas, tablas agregadas y estructuras optimizadas para consumo analítico.

## d. Ventajas y desventajas

Ventajas	Desventajas
Datos consistentes, limpios y auditables	Mayor complejidad técnica y de implementación
Ideal para reportes repetitivos y estandarizados	Requiere inversión de tiempo y recursos
Soporte a decisiones estratégicas	Menor flexibilidad ante cambios en las fuentes
Compatible con herramientas de BI	No es ideal para datos no estructurados

Un DW es altamente confiable, pero necesita **gobernanza, roles técnicos y un diseño cuidadoso** desde el inicio.

*IBM también menciona ventajas del modelo híbrido o en la nube que vale la pena mencionar, como:*

- Escalabilidad.
- Alta disponibilidad.
- Backup
- Recuperación ante desastres en arquitecturas modernas.

## e. Casos de uso real: BRADESCO Y SU DATA WAREHOUSE CORPORATIVO

En el sector financiero, el uso de Data Warehouses se ha vuelto esencial para consolidar información proveniente de múltiples sistemas, habilitar análisis estratégicos y asegurar el cumplimiento regulatorio. Un caso representativo y bien documentado es el del **Banco Bradesco**, una de las instituciones financieras más grandes de América Latina.

**Bradesco** implementó una solución basada en **Teradata Vantage** para centralizar los datos de sus clientes, transacciones, operaciones internas y sistemas CRM. Esta arquitectura de **Data Warehouse** permitió cubrir múltiples necesidades analíticas y operativas, incluyendo:

### Monitoreo de KPIs corporativos

Seguimiento de indicadores clave como ingresos, número de transacciones y eficiencia operativa por canal.

### Análisis financiero consolidado

Generación de informes financieros, detección de riesgos y evaluación de rentabilidad por segmento de clientes.

### Cumplimiento normativo y auditoría

Consolidación de datos para reportes regulatorios requeridos por organismos financieros brasileños e internacionales.

### Cuadros de mando gerenciales

Visualización de datos estratégicos en tiempo real para la toma de decisiones por parte del directorio y gerencias de negocio.

### Tecnologías utilizadas:

- **Teradata Vantage** (plataforma de almacenamiento y procesamiento de datos)
- **Herramientas de CRM y análisis** integradas con el DWH



Imagen 15

Fuente: [Teradata/Banco Bradesco](#)

# Data Mart

a. Propósito Un Data Mart es un subconjunto de un Data Warehouse que almacena datos relevantes para un área específica del negocio, como ventas, finanzas, marketing o recursos humanos. Su propósito es facilitar el acceso a la información por parte de usuarios no técnicos, proporcionando una vista simplificada, filtrada y optimizada para sus necesidades particulares. Mientras que el Data Warehouse centraliza toda la información empresarial, los Data Marts segmentan la información para mejorar la usabilidad y el rendimiento.

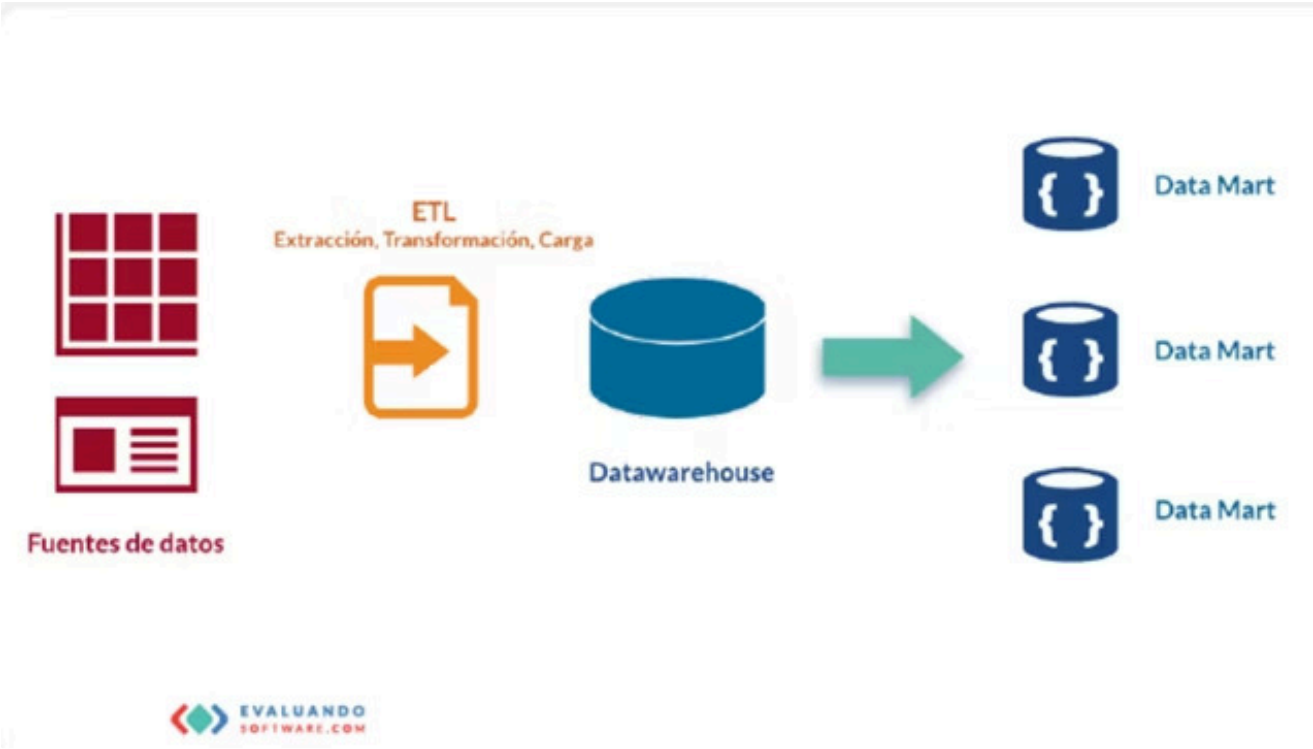


Imagen 16: Los Data Marts permiten dividir el Data Warehouse en áreas temáticas específicas, facilitando el acceso a información relevante para cada equipo del negocio. Fuente: ABC del data mart

b. Características principales

Característica	Descripción
Especialización temática	Contienen datos específicos de una unidad de negocio (ventas, RRHH, etc.).
Rapidez de consulta	Tiempos de respuesta más bajos gracias a menor volumen y complejidad.
Independencia funcional	Pueden diseñarse de forma aislada o conectados al Data Warehouse.
Fácil adopción por usuarios	Permiten a analistas o jefaturas acceder a información sin depender del equipo técnico.
Menor mantenimiento	Por su tamaño y foco limitado, son más fáciles de actualizar o migrar.

Pueden diseñarse como:

- Data Marts dependientes: se alimentan desde el Data Warehouse.
- Data Marts independientes: se alimentan directamente desde los sistemas fuente.
- Data marts híbridos: los data mart híbridos permiten combinar las fuentes de datos de un data warehouse corporativo con otras fuentes de datos, tales como sistemas transaccionales y/o operacionales.

c. Ventajas y desventajas

Ventajas	Desventajas
Consultas más rápidas y eficientes	Riesgo de duplicidad de lógica de negocio
Personalización por área o equipo	Puede generar inconsistencias entre departamentos
Facilidad de uso para usuarios no técnicos	Menor flexibilidad si el análisis requiere múltiples dominios
Rápida implementación (comparado con un DW completo)	A veces dificulta la gobernanza centralizada

Una mala coordinación entre los Data Marts puede generar lo que se llama “silos de datos”. d. Casos de uso Los Data Marts son ampliamente utilizados en organizaciones que requieren entregas rápidas de información por área funcional: Reporte de campañas de marketing Métricas como CTR, tasa de conversión, engagement por red social, ROI. Ventas por zona geográfica Comparativas entre sucursales, desempeño por región, evolución mensual. Dashboards de Recursos Humanos Rotación de personal, asistencia, evaluación de desempeño. Análisis de atención al cliente Volumen de tickets, tiempos de respuesta, satisfacción post-servicio.

1. ¿Cuándo usar cada opción?

Elegir entre un Data Lake, un Data Warehouse o un Data Mart depende del tipo de datos, del objetivo analítico y del perfil de los usuarios que consumirán la información. Estas tres soluciones no son excluyentes, sino complementarias: muchas organizaciones modernas utilizan las tres en conjunto dentro de su arquitectura de datos.

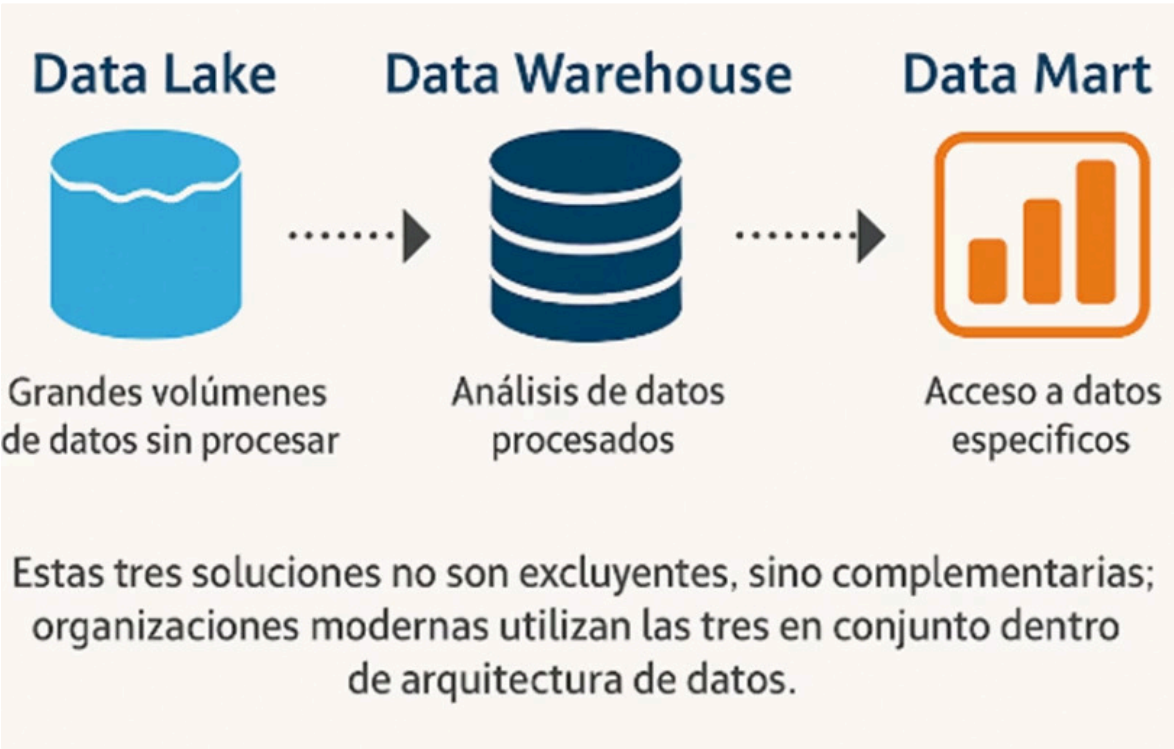


Imagen 18

a. Criterios para decidir

Criterio	Data Lake	Data Warehouse	Data Mart
Tipo de datos	Cualquier tipo (estructurado o no)	Estructurados y normalizados	Datos filtrados y específicos por área
Formato de ingreso	Crudo, sin transformación	Limpio, transformado	Derivado del DW o fuentes limitadas
Nivel de detalle	Muy alto (registro por evento)	Medio (detalle + agregado)	Enfocado (por producto, región, equipo)
Usuarios principales	Científicos de datos, ingenieros	Analistas, equipos BI, gerencia	Supervisores, áreas de negocio
Objetivo	Experimentación, exploración	Análisis histórico, KPIs, decisiones	Reportes operativos y tácticos
Velocidad de consulta	Baja sin optimización	Alta (consultas optimizadas)	Muy alta (por volumen reducido)
Flexibilidad	Muy alta	Media	Media-baja
Gobernanza y calidad	Baja si no se gestiona bien	Alta (control centralizado)	Media (riesgo de silos si no hay control)

b. Casos de decisión

A continuación, se presentan escenarios comunes y la recomendación correspondiente:

Escenario	Recomendación
Analizar logs de millones de sesiones web	Data Lake
Generar reporte mensual de ingresos consolidados	Data Warehouse
Consultar ventas por tienda para jefes zonales	Data Mart
Almacenar archivos de audio para análisis de sentimiento	Data Lake
Visualizar indicadores clave (KPI) de toda la empresa	Data Warehouse
Revisar métricas de campañas por equipo de marketing digital	Data Mart

Ejemplo combinado: arquitectura híbrida Una empresa de e-commerce implementa: Un Data Lake en AWS para almacenar logs web, archivos JSON de proveedores y reseñas de usuarios. Un Data Warehouse en Redshift para consolidar ventas, facturación y métricas clave del negocio. Varios Data Marts para las áreas de marketing, operaciones y finanzas, accedidos vía dashboards de Power BI. Este enfoque permite cubrir desde exploración de datos sin estructura hasta reportes ejecutivos consolidados. Como hemos podido ver hasta ahora, una arquitectura moderna orientada a datos, no se trata de elegir solo una opción, sino de diseñar un ecosistema donde: El Data Lake actúe como base de almacenamiento y exploración, El Data Warehouse como centro de datos consolidados y de calidad, Y los Data Marts como accesos rápidos y personalizados para usuarios finales.

1. Ejemplos reales

Para consolidar lo aprendido, te invitamos a investigar casos reales donde se apliquen soluciones basadas en Data Lake, Data Warehouse o Data Mart, ya sea de forma independiente o como parte de una arquitectura híbrida. Puedes usar como punto de partida los ejemplos revisados en las secciones anteriores: Netflix utilizando un Data Lake para logs y personalización. Bradesco centralizando indicadores clave en su Data Warehouse. Farmacias que implementan Data Marts temáticos para autonomía comercial.

Hasta este punto, no hemos configurado ni programado nada. Y eso es totalmente intencional. Antes de escribir una sola línea de código, es fundamental sostener reuniones con stakeholders, gerentes de áreas operativas y usuarios clave para entender los requerimientos reales del negocio. Solo así podremos diseñar arquitecturas de datos que sean un verdadero aporte para quienes necesitan consumir y analizar la información. Con esto cerramos nuestra exploración sobre las soluciones de almacenamiento modernas, incluyendo Data Lakes, Data Warehouses y Data Marts, comprendiendo sus diferencias, arquitecturas, beneficios y aplicaciones reales. Estas herramientas son la base sobre la cual se construyen los procesos analíticos y de inteligencia de negocio en cualquier organización.