**Implementación en Google Cloud ☁️**

Google implementa el concepto de una **Factoría de Datos Corporativos (CIF)** a través de su plataforma **Google Cloud**. Su enfoque se basa en una arquitectura unificada, inteligente y cloud que combina lo mejor de los Data Warehouses y los Data Lakes, a menudo denominada "Lake House".

* **Data Warehouse:** El servicio principal es **Google BigQuery**. Es un data warehouse cloud y altamente escalable que permite a las empresas almacenar y analizar enormes volúmenes de **datos estructurados**. Su principal ventaja es la separación entre almacenamiento y cómputo, lo que permite escalar ambos de forma independiente y pagar solo por las consultas que se ejecutan.
* **Data Lake:** La base para un data lake en Google Cloud es **Google Cloud Storage (GCS)**. Es un servicio de almacenamiento de objetos de bajo costo, diseñado para guardar cualquier tipo de dato en su formato original (estructurado, semi-estructurado y no estructurado).
* **Implementación del CIF (La Arquitectura "Lake House"):** Google integra estos dos mundos de forma nativa. La tecnología clave es **BigQuery**, que no solo funciona como un data warehouse tradicional, sino que también puede consultar directamente los datos que residen en el data lake mediante una función llamada **BigLake Tables**. De esta manera, Google crea una factoría de datos unificada donde un único motor de análisis puede trabajar tanto con datos crudos y no estructurados del lago como con datos procesados y estructurados del almacén, eliminando los silos de información. Servicios como **Cloud Data Fusion** para la ingesta (ETL/ELT) y **Looker** para la visualización (BI) completan la factoría.

|  |  |  |  |
| --- | --- | --- | --- |
| Eje de Comparación | Data Warehouse (DW) | Data Lake (DL) | Corporate Information Factory (CIF) |
| **Tipo de datos** | Principalmente **estructurados**. Datos procesados, validados y curados para el análisis. | **Todos los tipos:** estructurados, semi-estructurados y no estructurados. Almacena datos en su formato crudo y original. | **Todos los tipos.** Es una arquitectura integral que incluye y gestiona tanto el Data Lake como el Data Warehouse. |
| **Aplicación del esquema** | **Schema-on-Write.** La estructura de los datos se define *antes* de escribirlos. Los datos se limpian y transforman para encajar en el esquema. | **Schema-on-Read.** Los datos se almacenan sin un esquema predefinido. La estructura se aplica en el momento de la lectura o consulta. | **Ambos.** Aplica schema-on-write para su componente de Data Warehouse y schema-on-read para su componente de Data Lake. |
| **Casos de uso ideales** | Business Intelligence (BI), reportes corporativos, dashboards y análisis de métricas de negocio con datos históricos. | Data science, exploración de datos, machine learning, análisis de big data. Ideal cuando las preguntas futuras son desconocidas. | Estrategia de datos a nivel empresarial. Sirve a todos los usuarios, desde analistas de negocio hasta científicos de datos. |
| **Nivel de madurez técnica necesario** | **Intermedio.** Requiere conocimientos sólidos de SQL, modelado de datos (ej. esquemas de estrella) y procesos ETL. | **Alto.** Requiere habilidades en tecnologías de big data (ej. Spark, Hadoop), programación y sistemas distribuidos. | **Muy Alto.** Requiere una visión de arquitectura de datos completa, gobernanza de datos y la integración de múltiples sistemas complejos. |