

UNIVERSIDAD AUTONOMA GABRIEL RENE MORENO
FACULTAD DE INGENIERIA EN CIENCIAS DE LA COMPUTACION Y
TELECOMUNICAICONES

CARRERA DE INGENIERIA EN SISTEMAS



TAREA# 3

DATA WAREHOUSE

MATERIA: Sistemas para el soporte a la toma de decisiones
DOCENTE: Ing. Peinado Pereira Miguel Jesus
SIGLA: INF432 -SA
UNIVERSITARIO: Milan Limachi Villa
REGISTRO: 218151871

2/2024

Santa Cruz - Bolivia

DATA WAREHOUSE

Un **Data Warehouse** o almacén de datos es un sistema diseñado para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos que provienen de diversas fuentes dentro de una organización. A diferencia de las bases de datos transaccionales que se utilizan para la operación diaria, el Data Warehouse está optimizado para realizar consultas complejas y análisis de datos históricos y actuales, ayudando a las empresas en la toma de decisiones estratégicas.

2. Características Principales de un Data Warehouse:

1. Integrado:

- Combina datos de múltiples fuentes, como bases de datos transaccionales, aplicaciones empresariales (CRM, ERP) y archivos planos, en un formato coherente y unificado.

2. Orientado a Temas:

- Organiza los datos en torno a temas relevantes para el negocio, como ventas, finanzas o clientes, en lugar de basarse en los procesos o aplicaciones.

3. Histórico:

- Almacena datos históricos para analizar tendencias y patrones a lo largo del tiempo, a diferencia de las bases de datos operacionales que generalmente solo contienen datos actuales.

4. No Volátil:

- Una vez que los datos se cargan en el Data Warehouse, no se modifican ni eliminan, lo que asegura que el análisis sea replicable y consistente.

3. Componentes del Data Warehouse:

1. Fuente de Datos:

- Los datos provienen de diversas fuentes operativas, como bases de datos transaccionales, sistemas ERP, hojas de cálculo, APIs, entre otros.

2. Proceso ETL (Extract, Transform, Load):

- Extrae los datos de las fuentes, los transforma para asegurar consistencia y calidad, y finalmente los carga en el Data Warehouse.

3. Almacén de Datos:

- Es la base de datos central donde se almacenan los datos. Puede estar organizado en esquemas de estrella o copo de nieve para optimizar las consultas.

4. Herramientas de Acceso y Análisis:

- Herramientas de consulta y análisis que permiten a los usuarios interactuar con los datos, como herramientas de Business Intelligence (BI) para generar informes y dashboards.

5. Metadatos:

- Son datos que describen otros datos, ayudando a los usuarios a entender la estructura, contenido y uso de los datos almacenados.

4. Arquitectura de un Data Warehouse:

1. Capa de Fuentes de Datos:

- Incluye las diversas fuentes de datos que se extraen para poblar el Data Warehouse. Estas fuentes pueden ser bases de datos transaccionales, archivos planos, APIs, etc.

2. Capa de Integración (ETL/ELT):

- El proceso de **ETL** o **ELT** se encarga de extraer los datos, transformarlos para asegurar su calidad y formato coherente, y cargarlos en el almacén de datos.

3. Capa de Almacenamiento de Datos:

- Es el núcleo del Data Warehouse, donde se almacena toda la información estructurada. Aquí se organizan en esquemas optimizados para consultas.

4. Capa de Procesamiento Analítico (OLAP):

- Los datos se organizan en cubos multidimensionales para realizar análisis rápidos y eficientes. Permite explorar los datos desde diferentes dimensiones o perspectivas.

5. Capa de Presentación y Análisis:

- Los usuarios acceden al Data Warehouse a través de herramientas de análisis, generación de informes, y dashboards que facilitan la interpretación de los datos.

6. Capa de Metadatos:

- Contiene información sobre la estructura de los datos (metadatos técnicos) y su contexto de negocio (metadatos de negocio), esenciales para un uso eficiente.

7. Capa de Seguridad y Gestión:

- Asegura el control de acceso y la protección de los datos, con funciones de autenticación, autorización, auditoría y encriptación.

5. Metodologías de Data Warehouse:

1. Metodología Inmon:

- Desarrollada por Bill Inmon, esta metodología se enfoca en la creación de un **Data Warehouse corporativo centralizado** que integre todas las fuentes de datos de la empresa. Los datos se normalizan y luego se crean modelos de datos dimensionales para facilitar el análisis.

2. Metodología Kimball:

- Desarrollada por Ralph Kimball, esta metodología se basa en la creación de **Data Marts** específicos para cada área del negocio, que luego se integran para formar un Data Warehouse global. Esta metodología es más rápida de implementar y se enfoca en la simplicidad.

6. Tipos de Data Warehouse:

1. Almacén de Datos Empresariales (EDW - Enterprise Data Warehouse):

- Es el almacén de datos centralizado de toda la organización, diseñado para soportar la toma de decisiones a nivel empresarial.

2. ODS (Operational Data Store):

- Es un tipo de almacén de datos que se actualiza en tiempo real y se utiliza para gestionar operaciones diarias de la organización, en lugar de análisis históricos.

3. Data Mart:

- Es un subconjunto de un Data Warehouse que se centra en un área de negocio específica, como marketing o finanzas. Los Data Marts permiten un acceso más rápido a los datos relevantes para grupos específicos de usuarios.