

UNIVERSIDAD AUTÓNOMA GABRIEL RENE MORENO
FACULTAD DE INGENIERÍA EN CIENCIAS DE LA
COMPUTACIÓN Y TELECOMUNICACIONES
CARRERA INGENIERÍA 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: Cahuasiri Poroso David

REGISTRO: 220031665

2/2024

Santa Cruz – Bolivia

DATA WAREHOUSE

Definición: Un Data Warehouse (almacén de datos) es un sistema diseñado para almacenar, procesar y analizar grandes volúmenes de datos que provienen de múltiples fuentes dentro de una organización. A diferencia de una base de datos tradicional que se usa para el procesamiento de transacciones diarias, un Data Warehouse está optimizado para consultas complejas y análisis, permitiendo a las empresas tomar decisiones basadas en datos históricos y actuales.

Características Principales del Data Warehouse

1. Integrado:

Combina datos de múltiples fuentes (por ejemplo, bases de datos transaccionales, archivos planos, APIs) en un formato coherente y uniforme.

2. Orientado a Temas:

Los datos se organizan en torno a temas importantes para la organización, como ventas, finanzas, clientes, etc., en lugar de estar organizados por aplicaciones o procesos.

3. Histórico:

Almacena datos históricos, lo que permite a los usuarios analizar cambios y tendencias a lo largo del tiempo. Esto es diferente a las bases de datos operacionales, que generalmente almacenan solo datos actuales.

4. No Volátil:

Una vez que los datos se cargan en el Data Warehouse, no se actualizan ni se eliminan. Esto asegura que el análisis sea consistente y se pueda replicar.

Componentes del Data Warehouse

1. Fuente de Datos:

Los datos provienen de diferentes sistemas operacionales y se extraen a través de procesos de ETL (Extract, Transform, Load).

2. Proceso ETL:

Este proceso 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 diferentes esquemas como estrella o copo de nieve para optimizar las consultas.

4. Herramientas de Acceso y Análisis:

Los usuarios acceden al Data Warehouse a través de herramientas de consulta, reporting, y análisis (como BI tools). Estas herramientas permiten explorar los datos y generar informes y dashboards.

5. Metadatos:

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

Arquitectura de un Data Warehouse

1. Capa de Fuentes de Datos

Descripción: Esta capa incluye todas las fuentes de datos de donde se extraen los datos para el Data Warehouse. Puede incluir bases de datos transaccionales, aplicaciones empresariales (como ERP, CRM), archivos planos, APIs, datos en la nube, hojas de cálculo, y cualquier otra fuente relevante.

Ejemplos: SQL Server, MySQL, Salesforce, archivos CSV, APIs REST.

2. Capa de Integración (ETL/ELT)

Descripción: En esta capa se realiza el proceso de ETL (Extract, Transform, Load) o ELT (Extract, Load, Transform). Aquí es donde los datos se extraen de las fuentes, se transforman para limpiarlos, normalizarlos y adaptarlos a un formato coherente, y luego se cargan en el Data Warehouse.

Componentes:

Extracción: Recopila datos de las fuentes.

Transformación: Incluye la limpieza, deduplicación, conversión de formatos y otras operaciones para asegurar la calidad y coherencia de los datos.

Carga: Inserta los datos transformados en el almacén de datos.

Herramientas: Talend, Informatica PowerCenter, Apache NiFi, Microsoft SSIS (SQL Server Integration Services).

3. Capa de Almacenamiento de Datos

Descripción: Esta es la capa central donde se almacenan los datos dentro del Data Warehouse. Se organiza típicamente en esquemas (como esquema estrella o copo de nieve) que están optimizados para consultas y análisis. Los datos aquí son históricos y están estructurados para facilitar el acceso y análisis.

Componentes:

Data Warehouse: Almacén principal donde se guardan los datos en un formato estructurado.

Data Marts: Subconjuntos del Data Warehouse que se enfocan en áreas específicas del negocio, como finanzas, ventas, o marketing. A veces se crean para mejorar el rendimiento o para adaptar el almacén a las necesidades de un grupo específico de usuarios.

Ejemplos de Almacenamiento: Amazon Redshift, Google BigQuery, Snowflake, Microsoft Azure Synapse Analytics, Teradata.

4. Capa de Procesamiento Analítico (OLAP)

Descripción: En esta capa, los datos almacenados en el Data Warehouse se organizan en estructuras multidimensionales llamadas cubos OLAP (Online Analytical Processing), lo que facilita el análisis rápido y eficiente de grandes volúmenes de datos en diferentes dimensiones.

Funciones:

Agrupación y agregación: Sumarización de datos en varias dimensiones.

Drill-down y Roll-up: Explorar los datos desde niveles generales hasta detalles específicos, y viceversa.

Pivoting: Reorganización de los datos en diferentes perspectivas.

Herramientas OLAP: Microsoft Analysis Services, Oracle OLAP, IBM Cognos, SAP BW.

5. Capa de Presentación y Análisis

Descripción: Esta capa incluye herramientas que permiten a los usuarios finales acceder, analizar y visualizar los datos almacenados en el Data Warehouse. Aquí es donde se generan informes, dashboards, y se realizan análisis ad hoc.

Componentes:

Herramientas de BI (Business Intelligence): Se utilizan para la creación de informes y dashboards que presentan los datos de manera visual y comprensible.

Herramientas de Data Mining y Análisis Predictivo: Usadas para descubrir patrones ocultos y hacer predicciones basadas en datos históricos.

Ejemplos de Herramientas: Tableau, Power BI, QlikView, Looker, IBM Cognos, SAP BusinessObjects.

6. Capa de Metadatos

Descripción: Los metadatos son "datos sobre los datos". Esta capa almacena información que describe el contenido, estructura y contexto de los datos almacenados en el Data Warehouse. Los metadatos son esenciales para que los usuarios y aplicaciones comprendan y utilicen correctamente los datos del Data Warehouse.

Tipos de Metadatos:

Metadatos Técnicos: Describen la estructura de los datos, como tablas, columnas, relaciones, y las transformaciones realizadas.

Metadatos de Negocio: Describen el significado de los datos desde una perspectiva empresarial, como definiciones de KPI, reglas de negocio, y contexto de uso.

7. Capa de Seguridad y Gestión

Descripción: Esta capa asegura que el acceso a los datos esté controlado y monitoreado. Incluye políticas de seguridad, autenticación de usuarios, autorización, encriptación de datos, auditoría y cumplimiento normativo.

Funciones:

Control de Acceso: Definición y aplicación de permisos para acceder y manipular los datos.

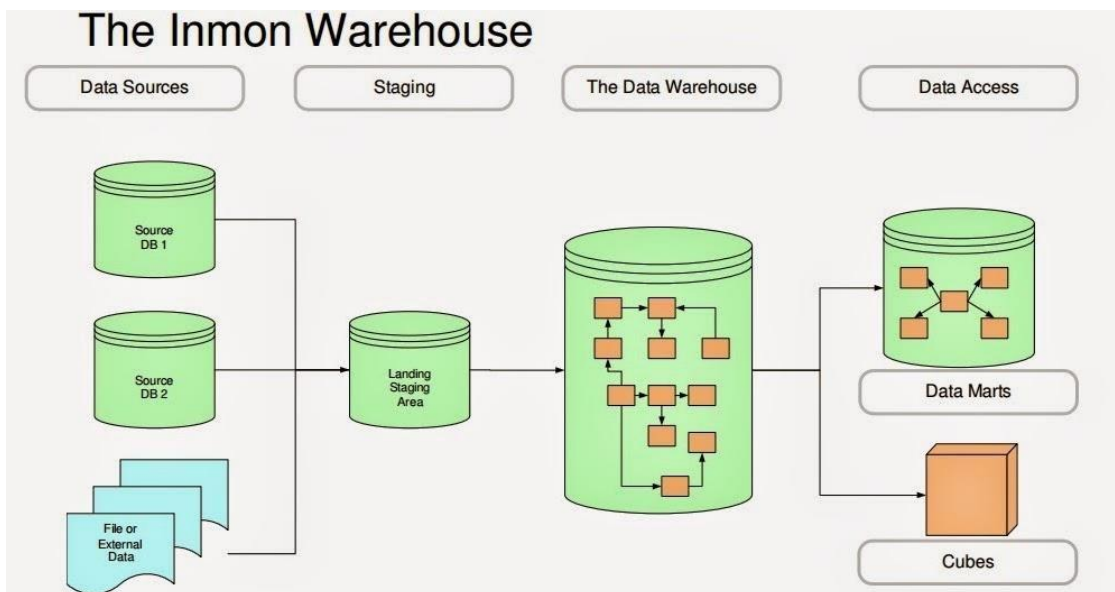
Auditoría: Registro de actividades de los usuarios para asegurar el cumplimiento de políticas y detectar posibles problemas de seguridad.

Cifrado de Datos: Protege los datos sensibles tanto en reposo como en tránsito.

METODOLOGIAS DE DATA WAREHOUSE

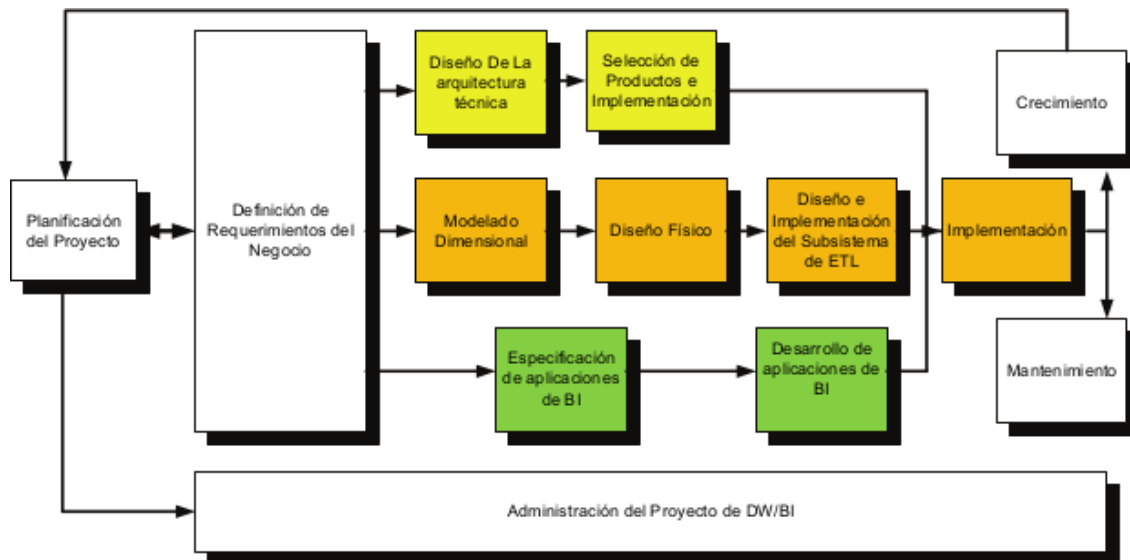
Existen varias metodologías utilizadas en el diseño e implementación de Data Warehouses. Estas metodologías proporcionan un enfoque estructurado y un conjunto de pasos para planificar, desarrollar y mantener un Data Warehouse de manera eficiente.

Metodología Inmon: Desarrollada por Bill Inmon, esta metodología se centra en la construcción de un Data Warehouse corporativo único y centralizado. Se enfoca en la integración de datos de diferentes fuentes en un único almacén de datos. Inmon aboga por la normalización de los datos y la creación de un modelo de datos dimensional para facilitar el análisis.



Metodología Kimball: Ralph Kimball desarrolló esta metodología, que se basa en la construcción de Data Marts independientes y orientados a un tema específico en lugar de un único Data Warehouse corporativo. Cada Data Mart es una colección de datos

orientada a un área de negocio particular. Kimball enfatiza la simplicidad y la rapidez de implementación.



TIPOS DE DATA WAREHOUSE

Almacén de datos empresariales (EDW): Sirve como base de datos central o clave para facilitar la toma de decisiones en toda la empresa.

ODS(ALMACÉN DE DATOS OPERATIVOS): El almacén de datos se actualiza en tiempo real.

DATAMART: Es un subconjunto de un almacén de datos destinados a brindar soporte a un departamento.