

DATA WAREHOUSE

Sistemas de Información Organizacional

DATA WAREHOUSE = REPOSITORIO DE DATOS

SISTEMAS INTERNOS



APLICACIONES



OTROS ORÍGENES



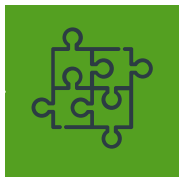
Extracción Transformación y Carga
ETL



**TOMA DE
DECISIONES**

Un datawarehouse es un repositorio de datos, integrado, no volátil, variable en el tiempo y orientado al negocio.

DATA WAREHOUSE = REPOSITORIO DE DATOS



01

INTEGRADO

Información de
sistemas
heterogéneos



02

NO VOLÁTIL

Datos perduran en
el tiempo



03

VARIABLE EN EL TIEMPO

Repositorio de
información
histórica



04

ORIENTADO AL SUJETO

Determinado
según necesidades

DATA WAREHOUSE = REPOSITORIO DE DATOS



Integrado: la información proviene de sistemas heterogéneos, como ser base de datos, sistemas transaccionales, archivos de textos, planillas de cálculos, etc.

No volátil: los datos almacenados perduran en el tiempo.

Variable en el tiempo: es un repositorio de información histórica. El tiempo en el cual son conservados los datos es mucho mayor que en sistemas transaccionales o bases de datos tradicionales.

Orientado al sujeto: los datos se organizan de manera tal que reflejan la estructura que posee el sujeto. El nivel de detalle a ser almacenado en el datawarehouse se determina según las necesidades de información que tenga el sujeto.

ARQUITECTURAS



—UN NIVEL

“Minimizar la cantidad de datos almacenados eliminando redundancia”



—DOS NIVELES

“Separar las fuentes físicas. Arquitectura no expandible y limitada a usuarios”



—TRES NIVELES

“Es la arquitectura más utilizada y está compuesta por tres niveles: inferior, medio y superior.”

Arquitectura de tres niveles

Esta es la arquitectura más utilizada. Se compone de los niveles superior, medio e inferior.

Nivel inferior: la base de datos de los servidores de Data Warehouse como nivel inferior suele ser un sistema de base de datos relacional. Los datos se limpian, transforman y cargan en esta capa utilizando herramientas de back-end.

Nivel medio: el nivel medio en el almacén de datos es un servidor OLAP que se implementa utilizando el modelo ROLAP o MOLAP.

Para un usuario, este nivel de aplicación presenta una vista abstracta de la base de datos. Esta capa también actúa como mediador entre el usuario final y la base de datos.

Nivel superior: el nivel superior es una capa de cliente front-end. El nivel superior son las herramientas y API que conecta y saca datos del almacén de datos.

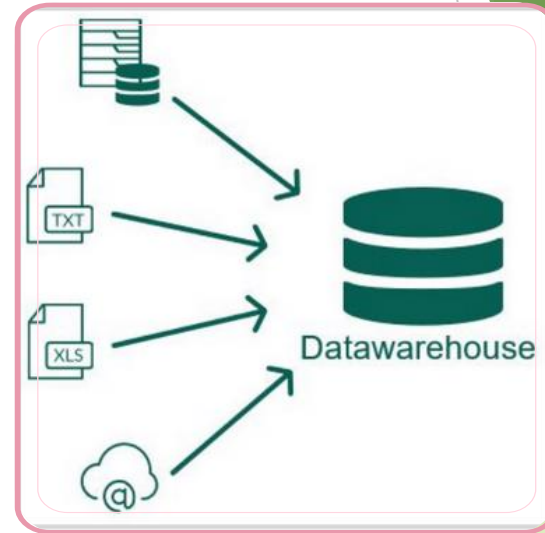
FUENTES DE DATOS

FUENTES ABIERTAS

Archivos planos, TXT, EXCEL, base de datos, servicios web.

FUENTES PROPIETARIAS

Base de datos cerrados, sistemas ERP, servicios web con conectores propietarios.



FLUJO DE DATOS

FUENTES
Sistema heterogéneo



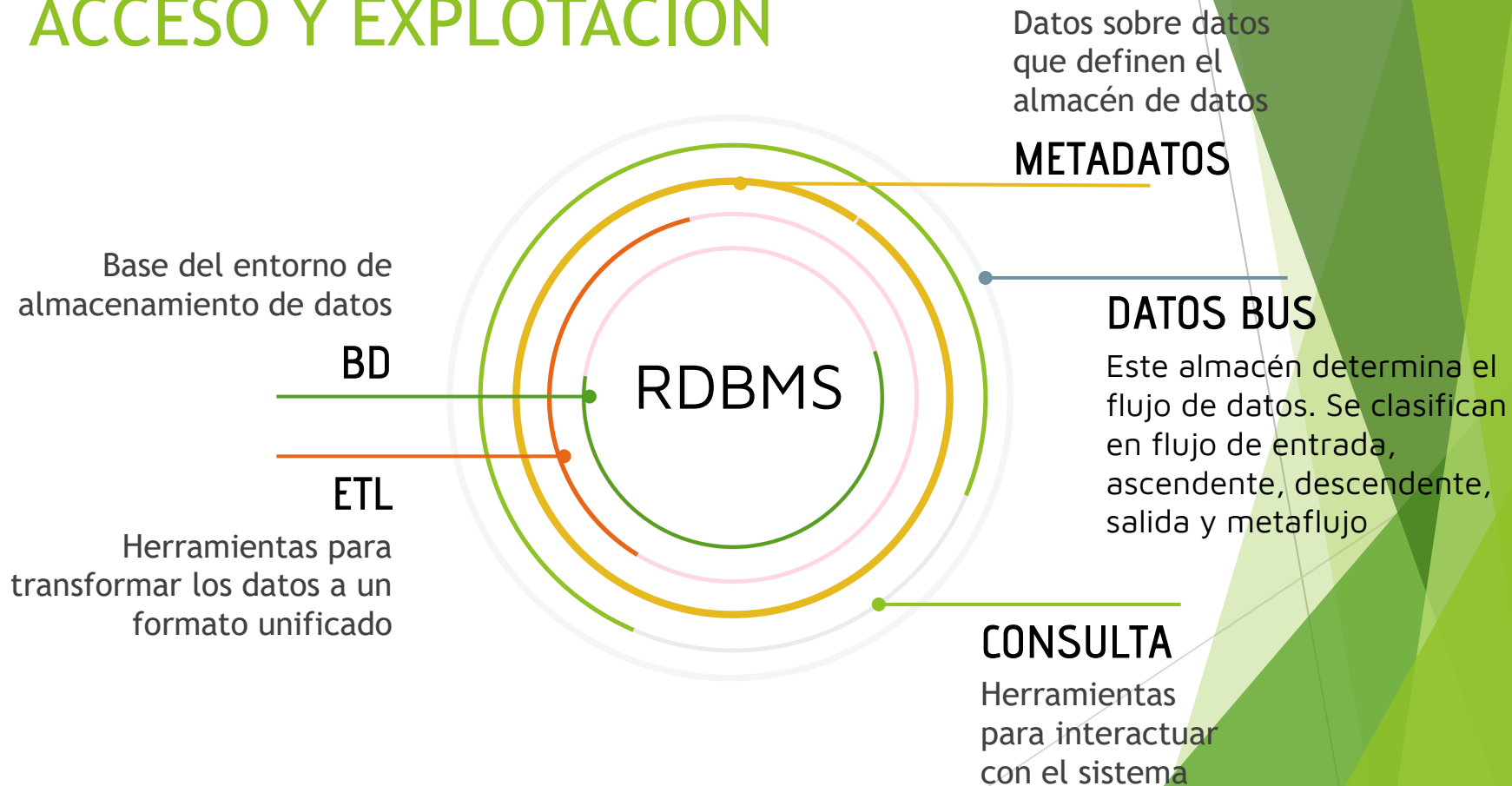
TRANSFORMACIÓN
Preparar datos para la carga

CALIDAD
Datos regidos por normas y definiciones



LIMPIEZA
Eliminación de información no útil

ACCESO Y EXPLOTACIÓN



ACCESO Y EXPLOTACIÓN

REPORTING

Técnica más antigua. Se representa en grillas o crosstab

DOCUMENTOS

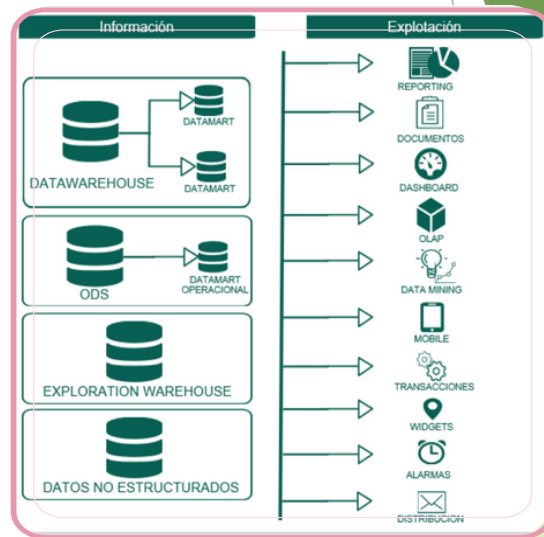
Son estáticos y no permiten demasiada interacción con el usuario

DASHBOARD

Contienen información resumida. Tienen gráficos, reportes, indicadores

ANÁLISIS OLAP

Analizar datos claves desde varios puntos de vista



OLAP

Online Analytical Processing



DATAMART OLAP

	TIPO	DESCRIPCIÓN
ROLAP	RELACIONAL	Bases de datos relacionales. Obtener fluidez y rapidez en las consultas.
MOLAP	MULTIDIM.	Bases de datos multidimensionales. Se utilizan técnicas de indexación para realizar búsquedas
HOLAP	HÍBRIDO	Bases de datos relacionales y multidim. Se utiliza una u otra técnica dependiendo lo que se necesite

OLTP

Online Transactional Processing



DATAMART OLTP

Orientadas al
procesamiento de
transacciones

TRANSACCIÓN

Se estructuran según el
nivel de aplicación

FORMATOS

BD

DATOS

Genera un proceso atómico
que involucra inserción,
modificación y borrado

No son
necesariamente
uniformes



DATA MINING

Conjunto de técnicas, procedimientos y tecnologías para explorar y analizar grandes volúmenes de información