

Optativo de Especialidad I

Arquitectura Inteligencia de Negocios.

Ingeniería Informática

Primavera

DEFINICIÓN

Se le llama inteligencia de negocios o Business Intelligence a un conjunto de técnicas que permiten **facilitar la toma de decisiones** en una organización, a través del uso de tecnologías de información.

Se trata básicamente de analizar información obtenida de distintas fuentes: sistemas de producción, financieros, otros sistemas internos y externos

¿Por qué tener BI?

Responder rápidamente a los
Retos de un entorno
económico cambiante



Conocer la
Rentabilidad de los
Clientes

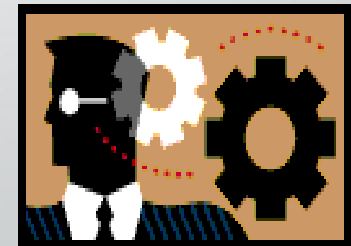


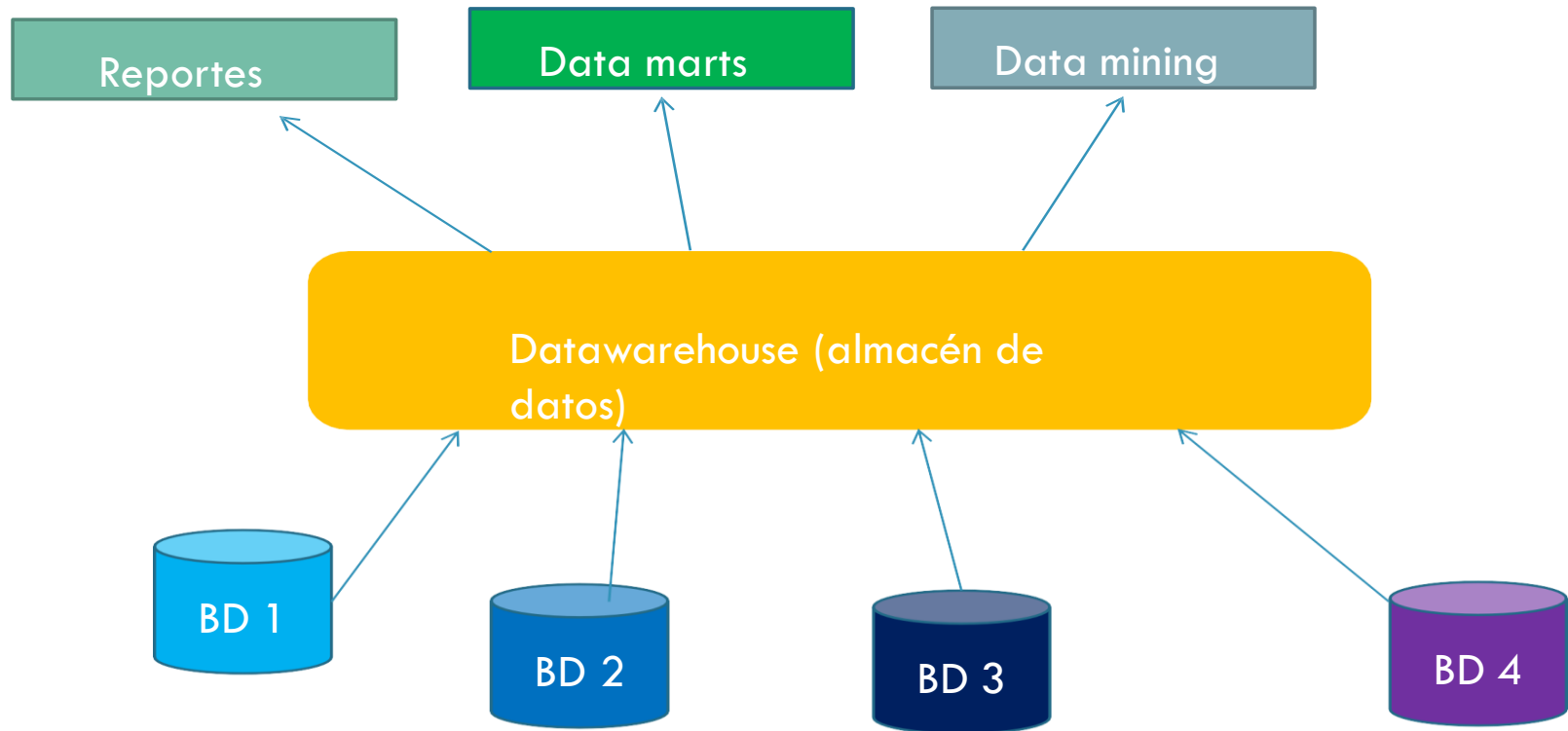
BUSINESS
INTELLIGENCE
APLICACIÓN ESTRATEGICA

Compartir la Información
entre distintos niveles de la
Organización



Analizar la Información
para identificar Factores
Críticos del negocio





EL CENTRO DE LA BI: DATAWAREHOUSES

Colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Diseñada para favorecer el análisis y la divulgación eficiente de datos.

Los DW surgen para **resolver consultas** que no pueden ser manejadas por Bases de Datos Operacionales (BDO)

TIPO DE CONSULTAS

¿Cuáles fueron los ingresos totales para Chile en el tercer trimestre del 2008?

¿Cuáles fueron las tres áreas más populares de cada ciudad para el arriendo de inmuebles el 2008?

¿Cuál es el ingreso mensual por venta de inmuebles en cada sucursal,
comparado con el promedio de los 12 meses anteriores?

Consultas hipotéticas “Que sucedería si...”

Análisis de escenarios alternativos, exploratorios, etc.

OBJETIVOS DE LAS BD DE CONSULTA

Resolver la problemática del acceso a la información para la toma de decisiones

Factores Críticos

- Tiempo de acceso
- Integración y Calidad de la Información

Se necesita:

- Conocimiento: Materia prima para toma de decisiones, es lo que se desea construir.
- Información.
- Datos.

LOS DATOS

Materia prima de la información.

Generados por procesos que no necesariamente los explotan.

Existen

No siempre son fáciles de acceder

No siempre son explotados

LA INFORMACIÓN

Materia prima para conocer los fenómenos reales.

Un ítem de datos es información según el contexto de toma de decisiones

- Dependiendo del tipo de decisiones a tomar.
- Dependiendo de la persona encargada.
- Dependiendo de su calidad.

La información suele ser difícil de obtener:

- Deben obtenerse los datos:
 - A partir de los cuales se construye la información.
 - Que definen el contexto del mismo.

OLTP Y OLAP

OLTP: Procesamiento de Transacciones On Line – Sistemas orientados a la Producción

- Prioridad: tiempo de respuesta a transacciones read-write.
- Se manejan datos actuales muy detallados.
- Estables y de larga vida útil.

OLAP: Procesamiento Analítico On Line – Sistemas Orientados a la Toma de Decisiones

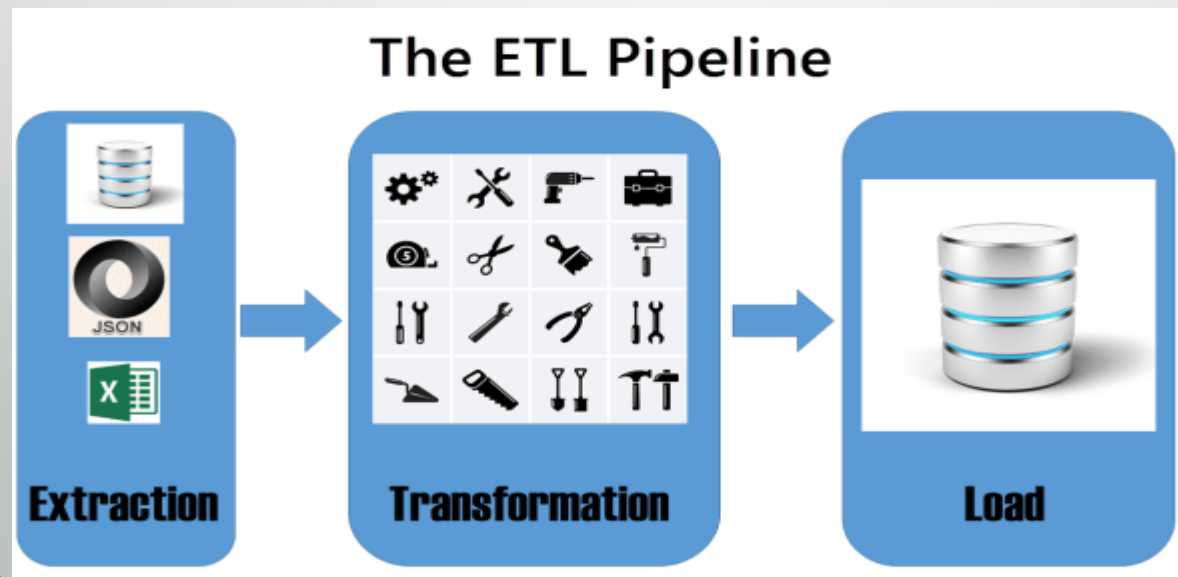
- Prioridad: expresividad y eficiencia en consultas complejas
- Datos actuales e históricos, resumidos
- En constante evolución

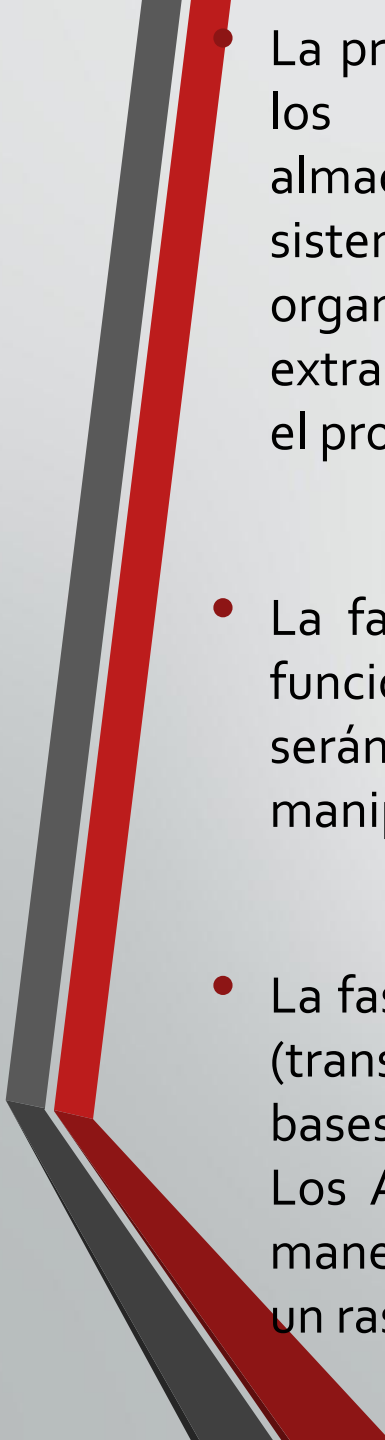
¿Qué es un Sistema OLTP?

- El OLTP por su sigla en inglés Procesamiento de Transacciones En Línea (OnLine Transaction) es un programa que facilita, gestiona y administra aplicaciones transaccionales por lo general para entradas de datos, recuperación, y procesamiento de transacciones que responde a las peticiones del usuario.
- Una transacción es una secuencia de operaciones que se lleva a cabo por medio de una base de datos de manera atómica. Las operaciones pueden ser de cuatro tipos diferentes: SELECT, INSERT, DELETE y UPDATE

ETL

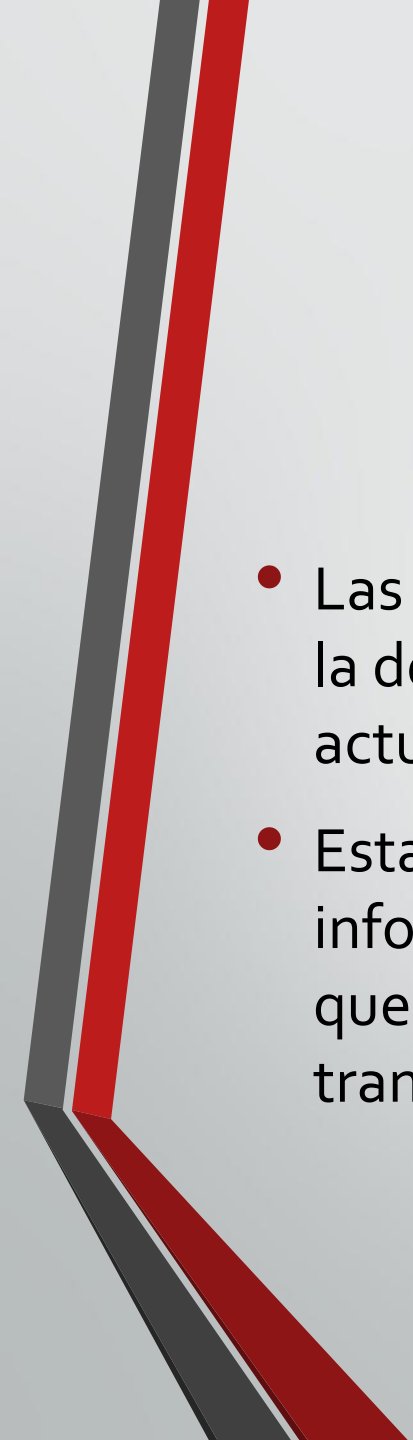
- ETL son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, transformarlos y cargarlos en otra base de datos, datamart, o datawarehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.



- 
- La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos y con la extracción se convierten los datos a un formato preparado para iniciar el proceso de transformación.
 - La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.
 - La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los Almacenes de Datos mantienen un historial de los registros de manera que se pueda hacer una auditoria de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

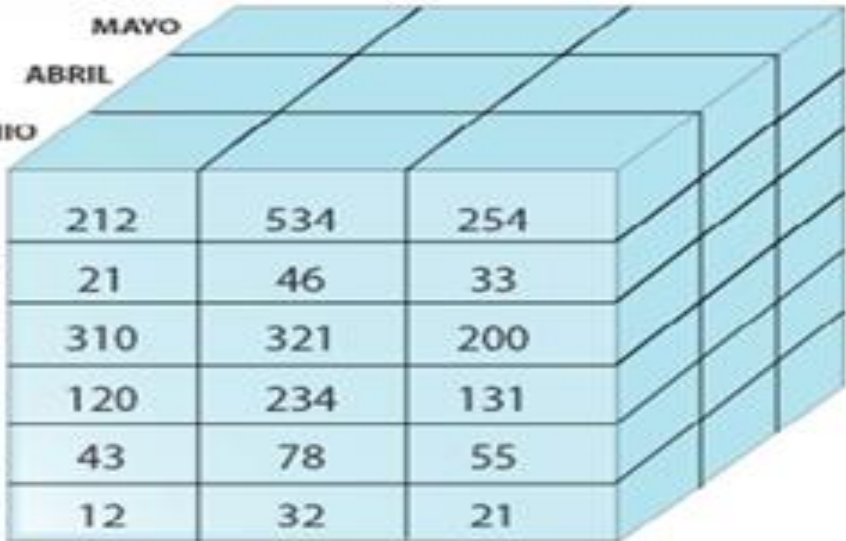
¿Qué se entiende por Sistemas OLAP (procesamiento analítico en línea (On-Line Analytical Processing))?

- Los sistemas OLAP son bases de datos que están enfocadas al procesamiento analítico, este análisis implica por lo general la lectura de grandes cantidades de datos con el fin de extraer información útil por ejemplo, tendencia de ventas, patrones de comportamiento de clientes, consumidores.
- OLAP es una solución utilizada en el campo de la Inteligencia cuyo objetivo es agilizar la consulta de grandes cantidades de datos, para lo cual utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP).

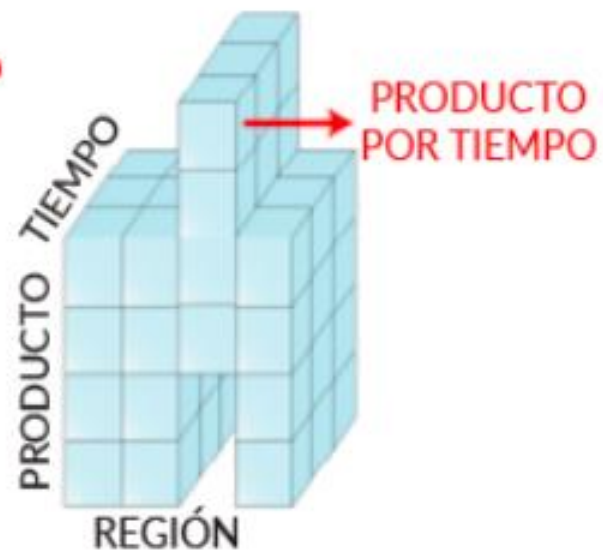
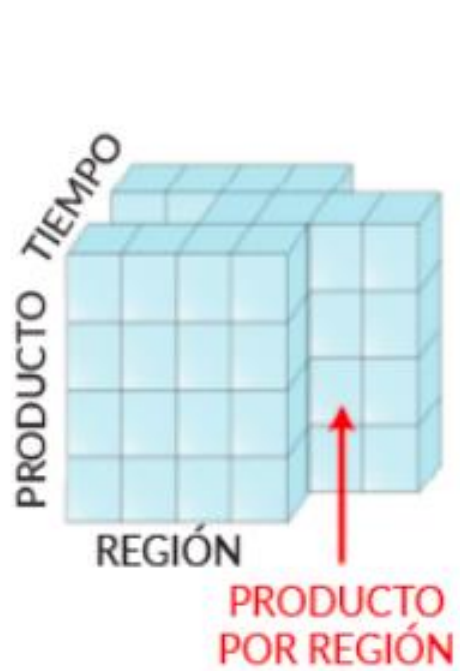
- 
- Las funciones más usadas en la Base de Datos OLAP es la de consulta de información, muy rara vez se realizan actualizaciones, y eliminaciones de información.
 - Estas bases de datos por lo general se alimentan de información proveniente de los sistemas operacionales que ya existen por medio de un proceso de extracción, transformación y carga. (ETL).

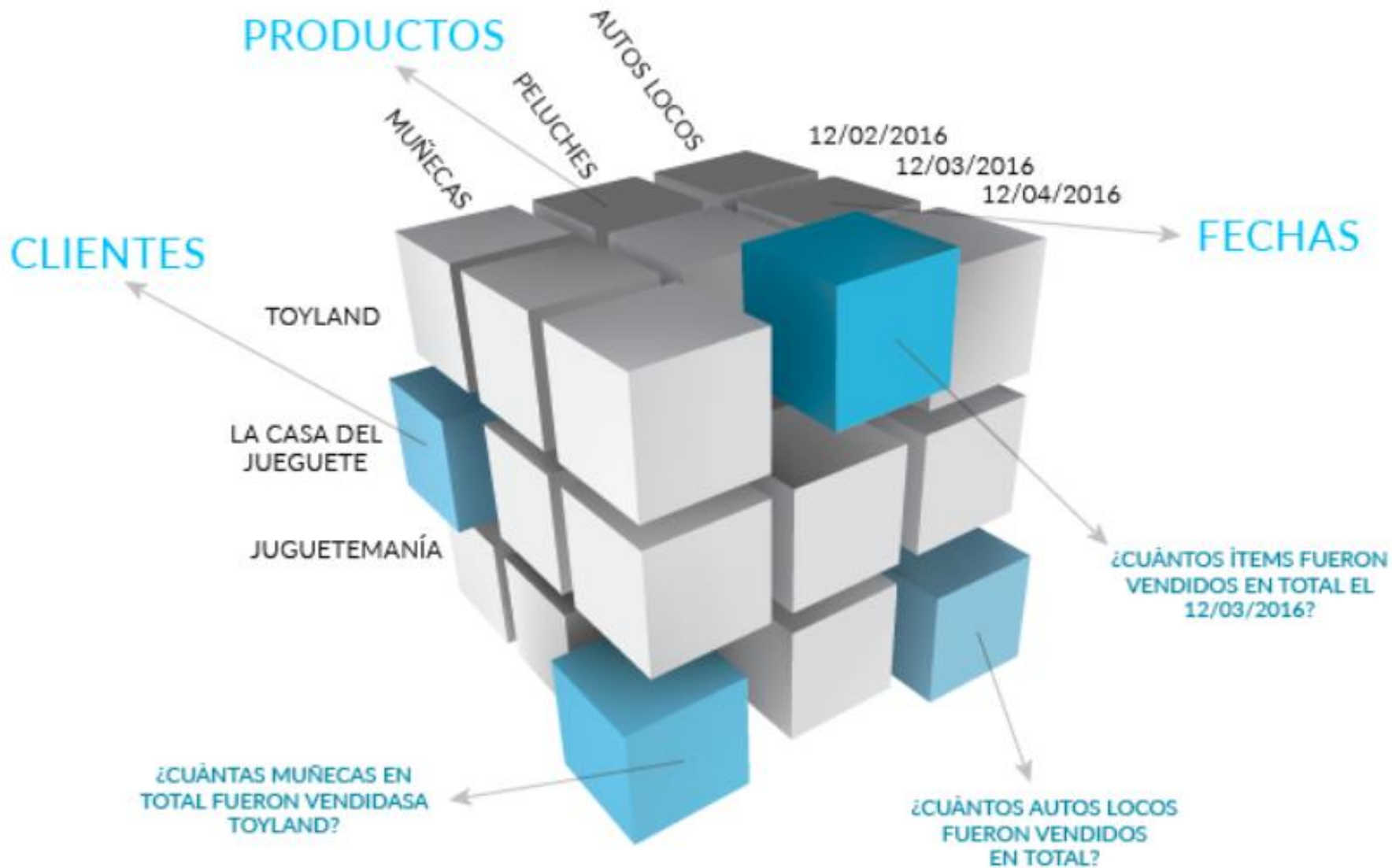
Cubos OLAP

- Las bases de datos OLAP se dividen en cubos de información, cada uno de los cubos lo organiza y diseña un administrador de cubos con el fin de que logre adaptarse a la forma en la que recupera y analiza los datos, esto con el fin de que sea más rápido de crear y de usar los informes de las tablas y gráficos dinámicos que se requieran.
- El usuario realiza su consulta a la herramienta OLAP seleccionando los atributos que considere pertinentes o necesarios de este esquema multidimensional sin tener conocimiento de la estructura interna o esquema físico del almacenamiento de datos.
- Estas herramientas generan la consulta correspondiente y la envían al gestor de consultas del sistema, mediante la sentencia SELECT.



	JUNIO	ABRIL	MAYO
Producto 1	212	534	254
Producto 2	21	46	33
Producto 3	310	321	200
Producto 4	120	234	131
Producto 5	43	78	55
Producto 6	12	32	21
	Argentina	Brasil	Chile





ALMACENES DE DATOS O DW

Bases de datos Operacional

- Datos Operacionales
- Orientado a la aplicación
- Actual
- Detallada
- Cambia continuamente

Data warehouse

- Datos del negocio para Información
- Orientado al sujeto
- Actual + histórico
- Detallada + más resumida
- Estable

DW: CARACTERÍSTICAS

Datos Orientados a Temas

- Los datos se organizan en torno a los temas principales de la organización

Datos integrados

- Heterogeneidad de datos:
 - Diferentes áreas de la organización.
 - Diferentes tipos (tradicionales, geográfico, documentos).
 - Diferentes fuentes
- Aspectos a resolver en la integración:
 - Unificación de conceptos.
 - Construcción del dato integrado a partir de los fuentes.
 - Disimilaridad esquemática y semántica

DW: CARACTERÍSTICAS

Datos históricos:

- Se deben manejar los datos con su referencia temporal.
- El tiempo siempre es una dimensión relevante

Datos no volátiles:

- Una vez que los datos son estables, se pueden ingresar al DW
- Los datos no se modifican, sólo se incorporan nuevos
- Los datos deben ser lo suficientemente estables como para permitir análisis “largos” sin que cambien durante el mismo.
- Esto se obtiene como consecuencia de:
 - La historización.
 - La planificación de la carga.

DATA MARTS

Son aplicaciones de análisis de datos en áreas precisas de negocios. Por ejemplo: Adquisiciones, Registro Curricular, Marketing, Producción.

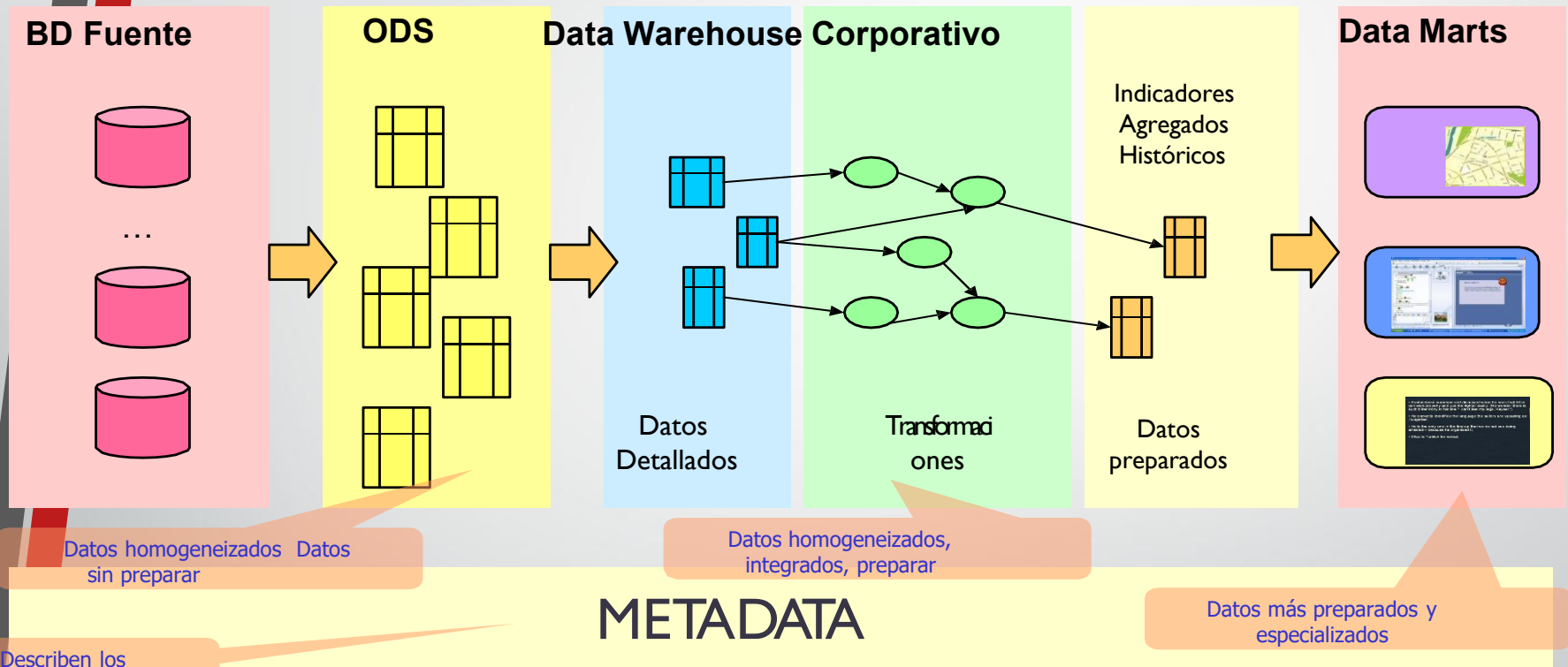
Toman sus datos del Data Warehouse.

Priorizan la funciones de análisis de datos:

- Interfaces a usuario.
- Indicadores específicos al área de negocio.

Normalmente basados en OLAP.

COMPONENTES DE UN DW



DW: TIPOS DE DATOS

Datos Detallados

- Corresponden a los datos insertados desde los sistemas de producción.
- Diferencia con datos de BD-Fuentes:
 - Están en formato homogéneo (vía ODS: sistema de organización de datos).
 - Pasaron un primer control de calidad.
 - Pueden ya haber pasado un primer proceso de agregación.
- Generalmente se trata de volúmenes importantes de datos.

DW: TIPOS DE DATOS

Datos Agregados

- Resultantes de aplicar funciones de totalización sobre datos detallados. Por ejemplo: total mensual de ventas por producto.
- Información significativa para analizar.
- Permiten reducir volúmenes de datos.
- Cálculo interactivo plantea problemas de rendimiento

DW: TIPOS DE DATOS

Datos Históricos

- Datos (base o agregados) a los cuales se les agrega una marca de tiempo.
- Generan volúmenes importantes de datos

DW: TIPOS DE DATOS

Metadatos

- Consiste en información sobre los datos del DW.
- Incluye información sobre:
 - Semántica de los datos y su localización en el DW.
 - Localización de los datos en los sistemas de producción y reglas de transformación.
 - Especificación de fórmulas de cálculo de agregados.
 - Información sobre frecuencias de carga, mecanismo de historización, etc.
- Constituye una pieza clave para el control de calidad de los datos y la explotación eficaz del DW.

DW: TIPOS DE OPERACIONES/ TRANSFORMACIONES

Extracción de datos.

- Consiste en extraer los datos de la BD fuente y cargarlo en DW.

Filtrado.

- Consiste en filtrar datos no admisibles en el DW.

Modificación de formato o valores.

- Consiste en adaptar formatos o valores para que cumpla pautas definidas en el DW.

Integración.

- Consiste en integrar datos provenientes de dos fuentes.

DW: TIPOS DE OPERACIONES/ TRANSFORMACIONES

Cálculos y Consolidaciones (Agregaciones).

- Consiste en calcular indicadores a partir de datos base. Pueden implicar consolidaciones.

Generación de datos históricos (históricos).

- Consiste en agregar marcas de tiempo a datos.

Generación de versiones.

- Consiste en agregar atributos diferenciadores de diferentes versiones de un objeto base.
- La histórico permite hacer esto marcando la versión con un valor temporal.

¿Qué es Data Warehouse?

- DW es un conjunto de tecnologías ,NO ES UN PRODUCTO.
- Es una arquitectura que debe construirse de acuerdo a las necesidades y entorno específico de los clientes, y debe construirse de manera *iterativa* para consolidar y administrar datos de varias fuentes con el propósito de conseguir en un periodo de tiempo aceptable:
 - *Ayudar a la toma de decisiones(DSS).*
 - *Descubrir conocimiento(Data Mining->minería de datos).*
 - *Responder preguntas de negocio(OLAP->análisis de datos).*

Almacén de Datos



Colección de datos diseñada para dar apoyo a los procesos en la toma de decisiones

características

Orientada hacia la información relevante de la organización.

Integrada

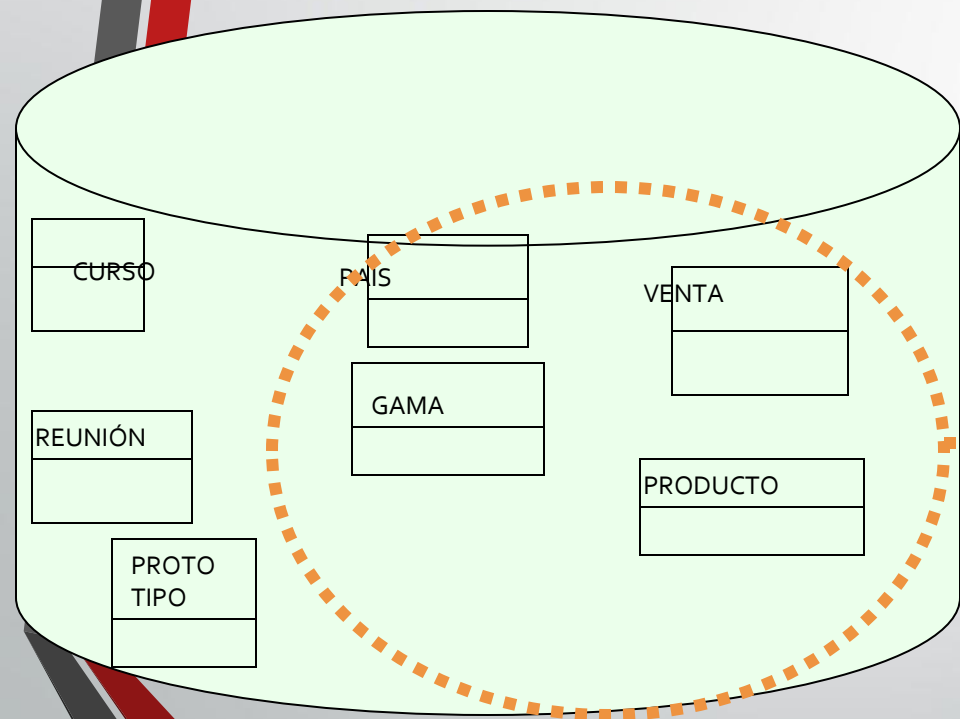
Variable en el tiempo

No volátil

- AD: Orientada hacia la información relevante en el tiempo.



Se diseña para consultar eficientemente información relativa a las actividades (ventas, compras, producción...) básicas de la organización, no para soportar los procesos que se realizan en ella, gestión de pedidos, facturación, etc...

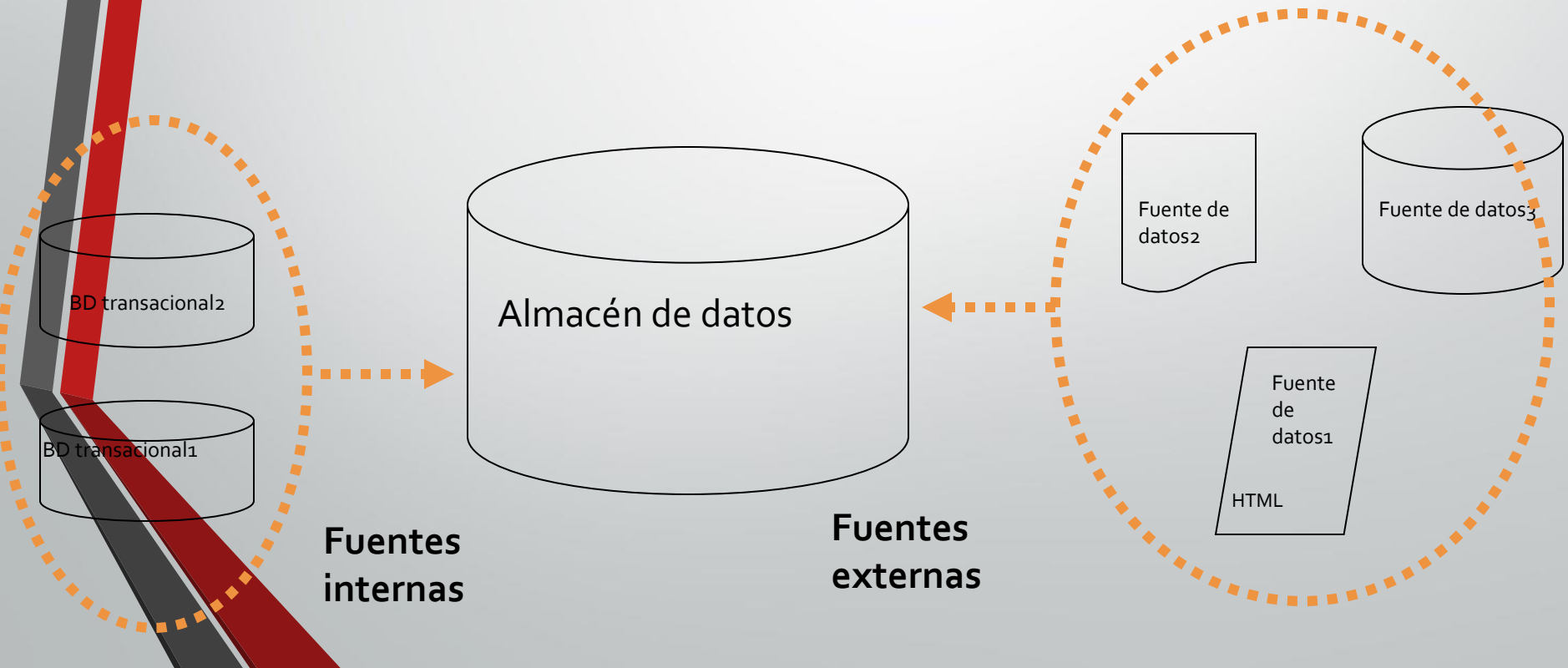


Información necesaria

- AD: Integrada



Integra datos recogidos de diferentes sistemas operacionales de la organización(y/o fuentes externas)



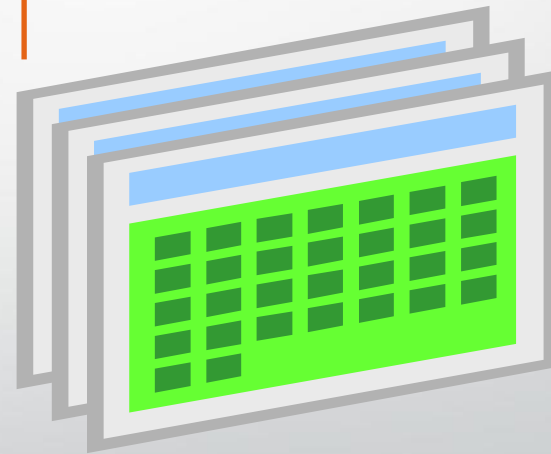
- AD: Variable en el tiempo.



Los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.

Los datos son almacenados como fotos (snapshots) correspondientes a periodos de tiempo.

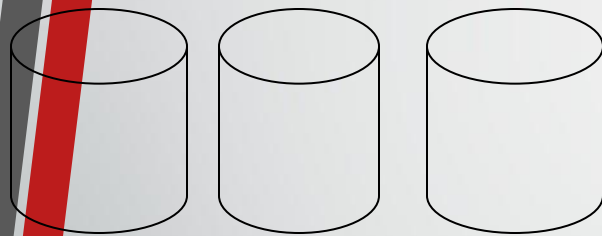
TiempoDatos	
01/2003	Datos de Enero
02/2003	Datos de Febrero
03/2003	Datos de Marzo



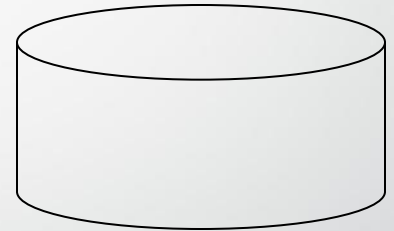
- AD: No volátil



Los datos almacenados no son actualizados
, solo son incrementados



CARGA



BD operacionales

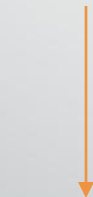
Almacén de datos



INSERT

DELETE

UPDATE



READ

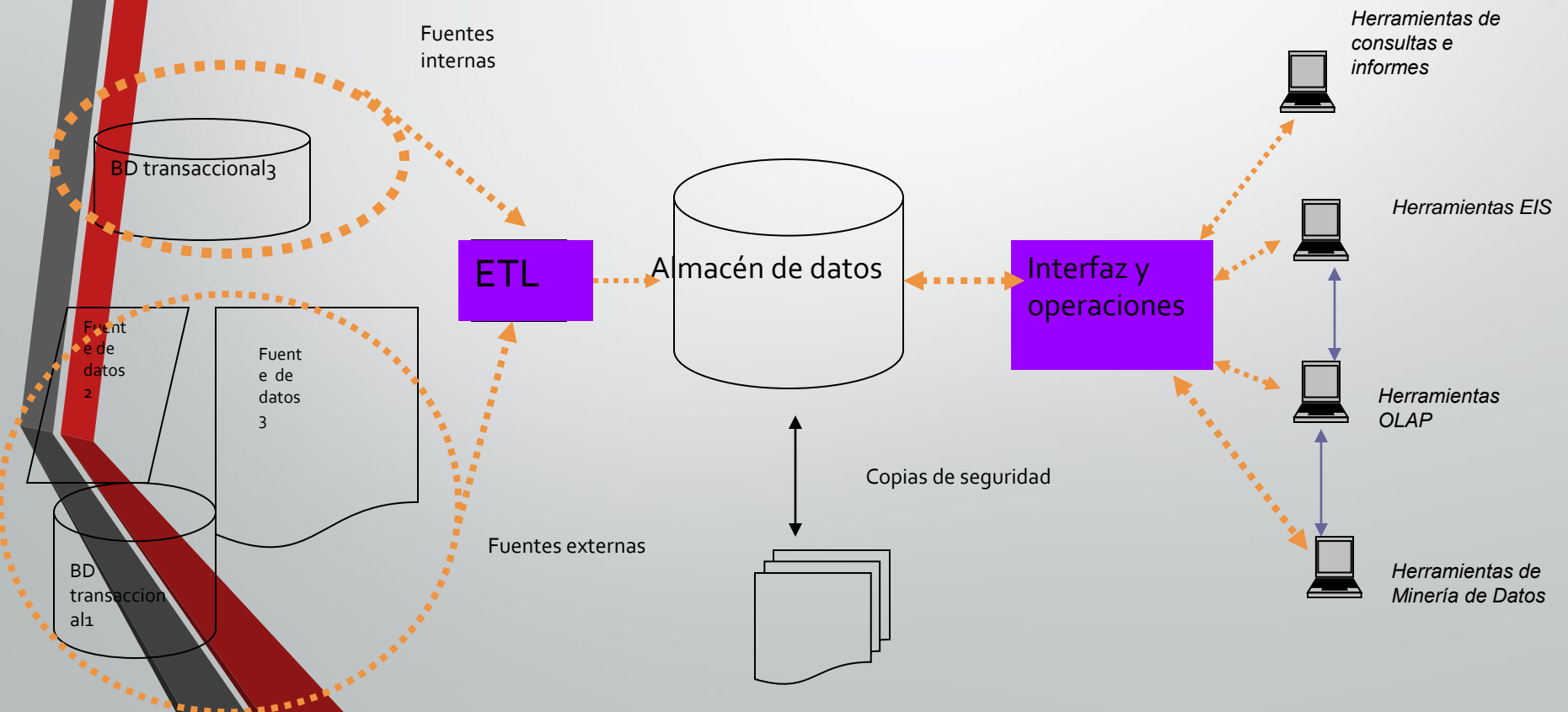



READ

El periodo de tiempo cubierto por un AD varía
entre 2 y 10 años.

Arquitectura

- La arquitectura de un AD viene determinada por su situación central como fuente de información para las herramientas de análisis.





Ventajas	Inconvenientes
Menos carga de trabajo	Lento y muy costoso
Facilita la estrategia de empresa	Privacidad de los datos
Rentabiliza su inversión	Recuperación ante fallos en carga
Mejora la productividad y competitividad en el mercado	Optimización de los recursos

KPI

- Indicadores de desempeño clave: es un valor tangible que permite medir y cuantificar si estamos en el camino correcto para lograr nuestro objetivo.
 - **Alineados al objetivo:** Nos aseguraremos que los indicadores están alineados con los objetivos estratégicos y de la organización.
 - **Alcanzables:** Los KPI deben tener datos que sean fáciles de conseguir, que sean alcanzables por la empresa.
 - **Agudos:** Los KPI deben mantener a todos en la misma página y moverse en la misma dirección.
 - **Preciso:** Los datos que fluyen en los indicadores deben ser exactos, nunca aproximaciones o hipótesis.
 - **Útiles:** Debemos preguntarnos: ¿Los indicadores nos dan la información sobre el negocio que necesitamos y son útiles? Si la respuesta es No, cambiamos los indicadores.
 - **Vivo:** Su negocio siempre está creciendo y cambiando. Sus KPI deben evolucionar también.