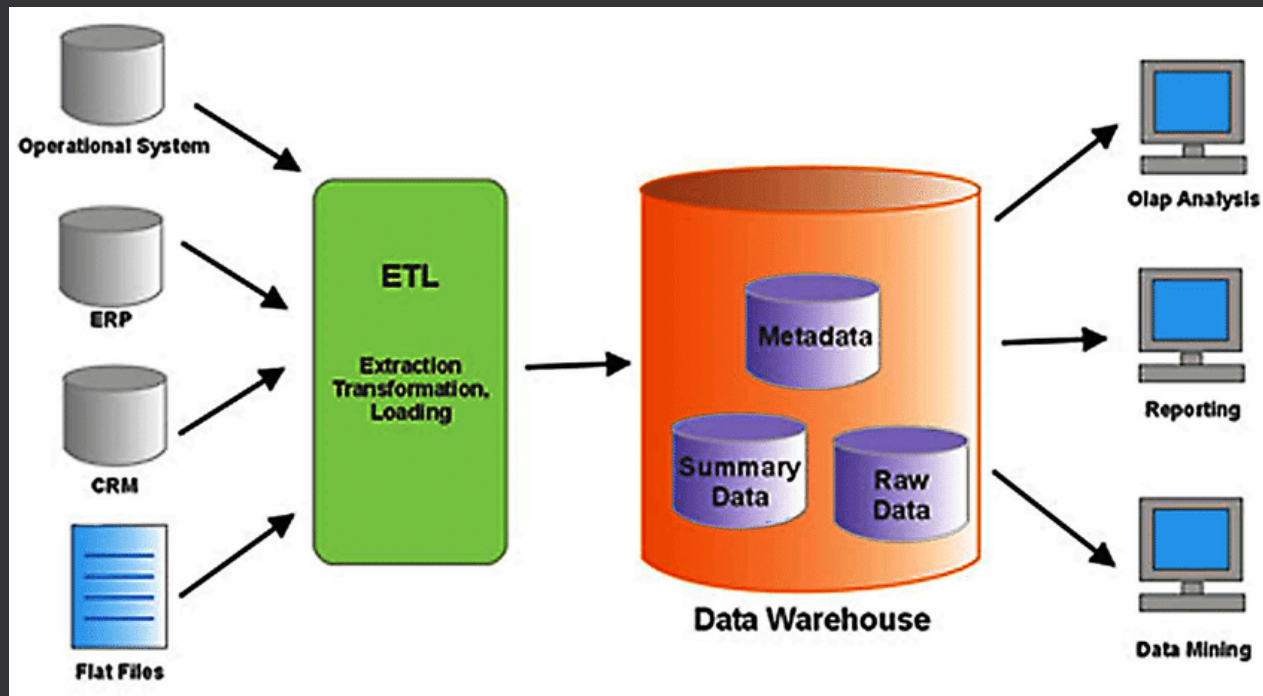


DATA WAREHOUSE



Clase 2

Ciclo 2-2020

ÍNDICE

1. Introducción.
2. ¿Qué es Data Warehouse?
3. Arquitectura.
4. Ventajas/Inconvenientes.
5. Diferencias entre una base de datos transaccional y un Data Warehouse

ÍNDICE

1. **Introducción.**
2. ¿Qué es Data Warehouse?.
3. Arquitectura.
4. Ventajas/Inconvenientes.
5. Diferencias entre una base de datos transaccional y un Data Warehouse

INTRODUCCIÓN

- **Objetivo:**

Análisis de datos para la toma de decisión.

INTRODUCCIÓN

- El aumento espectacular del volumen de datos hace evidente la necesidad de una infraestructura para la lógica de información.
- Surge como respuesta a la problemática de extraer información sintética a partir de datos atómicos almacenados en bd de producción.

INTRODUCCIÓN

ALMACÉN DE DATOS(AD)

Motivación

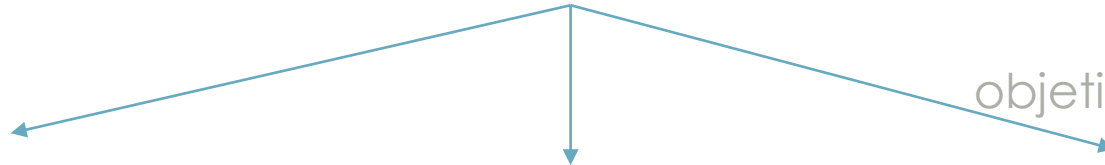


Disponer de Sistemas de Información de apoyo a la toma de decisiones



Disponer de DB que permitan extraer conocimiento de la información histórica almacenada en la organización.

objetivos



Análisis de la organización.

Previsiones de evolución.

Diseño de estrategias

INTRODUCCIÓN

- **Ejemplo**

Organización: Cadena de supermercados

Actividad objeto de análisis: ventas de productos

Objetivo: aumentar ventas con publicidad adecuada

- Problema 1: Necesitamos sólo datos necesarios de la BD
- Problema 2: Fuentes de datos diversas (BDs diferentes, ficheros de texto, ficheros XML...)
- Problema 3: Fuentes de datos externas
- Problema 4: Demasiados datos
- Problema 5: Análisis en tiempo real

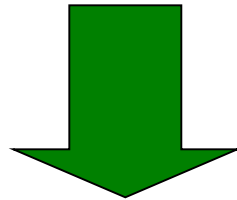
INTRODUCCIÓN

- CONCLUSIÓN:

LA BD NO BASTA!!!!

NECESITAMOS OTRA COSA

DATA WAREHOUSE



ÍNDICE

1. Introducción.
2. **¿Qué es Data Warehouse?.**
3. Arquitectura.
4. Ventajas/Inconvenientes.
5. Diferencias entre una base de datos transaccional y un Data Warehouse

¿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).

¿Qué es Data Warehouse?

ALMACEN DE DATOS(AD)



Bases de Datos diseñada para el objetivo de exploración distinto que al de las BD`s de los sistemas operacionales

Sistema
Operacional



BD orientada al
proceso

Sistema de
almacén de
datos(DW)



BD orientada al análisis

¿Qué es Data Warehouse?

ALMACEN DE DATOS(AD)



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

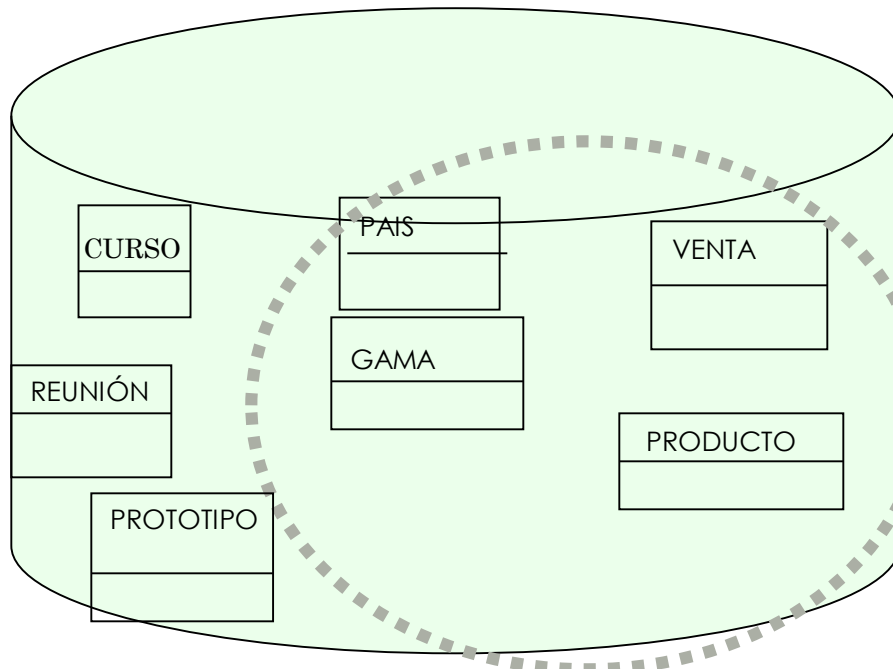
No volátil

¿Qué es Data Warehouse?

- 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...



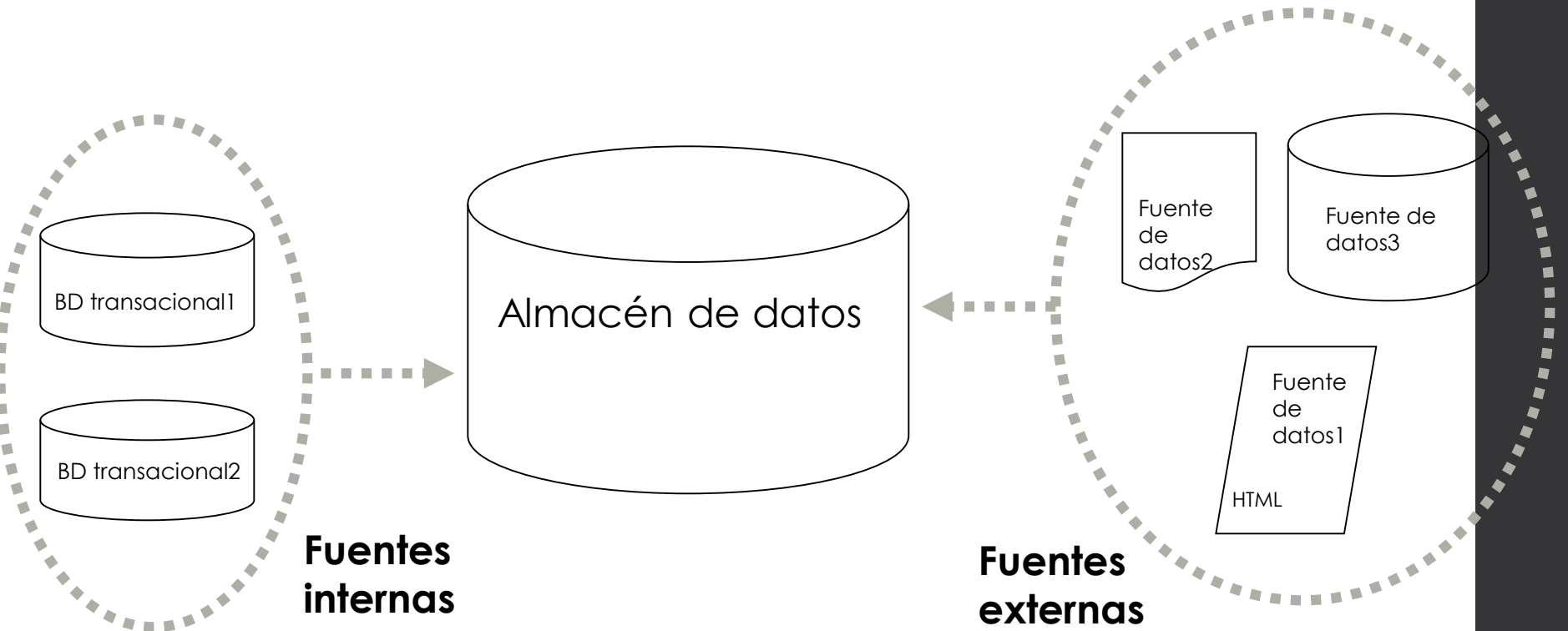
Información necesaria

¿Qué es Data Warehouse?

- AD: Integrada



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




¿Qué es Data Warehouse?

► AD: Variable en el tiempo.

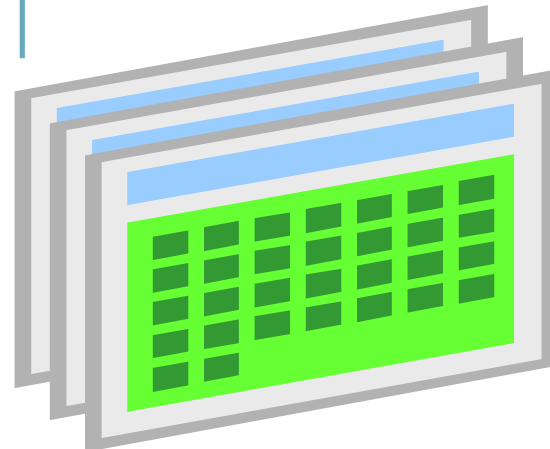


Los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.

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

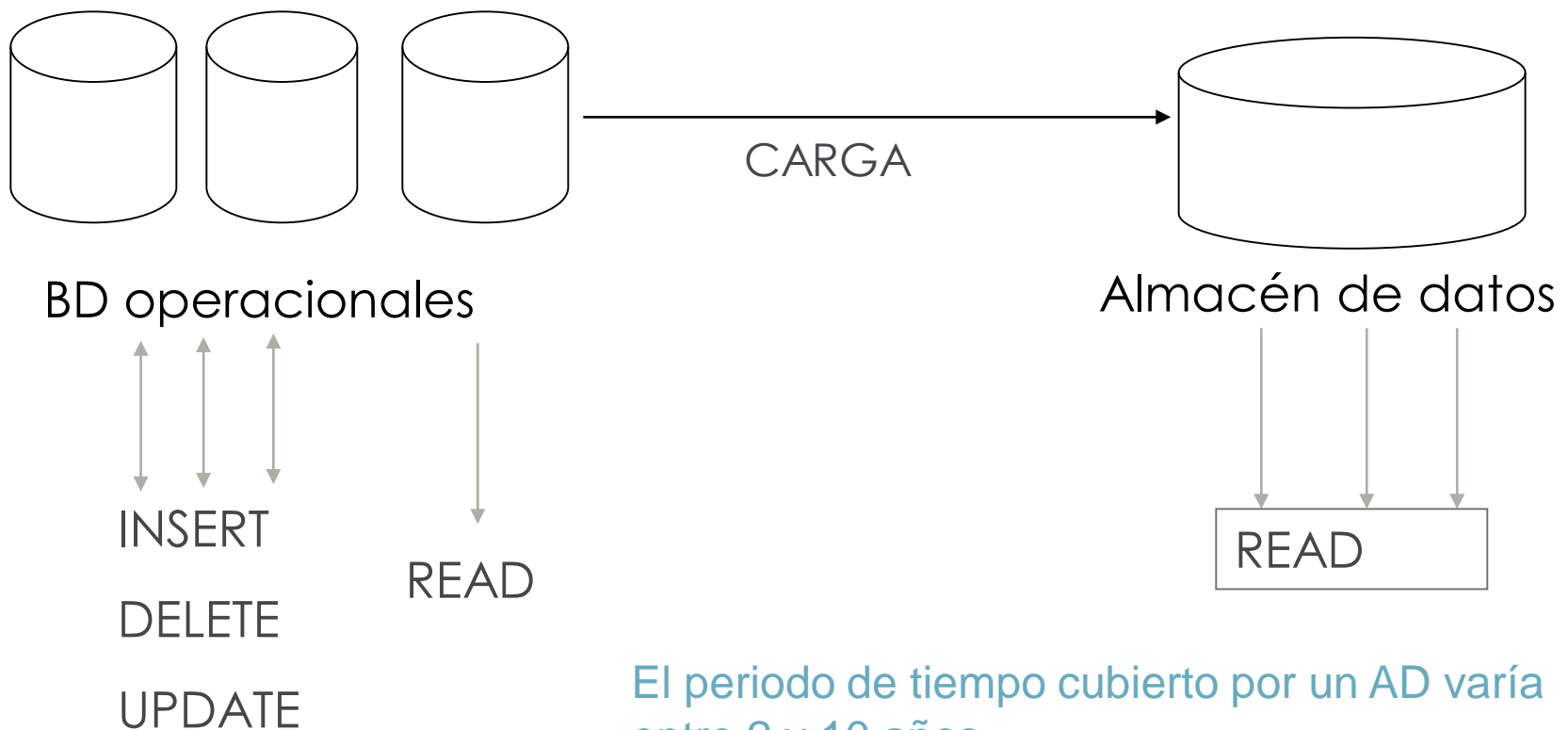


TiempoDatos	
01/2003	Datos de Enero
02/2003	Datos de Febrero
03/2003	Datos de Marzo



¿Qué es Data Warehouse?

- **AD: No volátil** ➡ Los datos almacenados no son actualizados, solo son incrementados

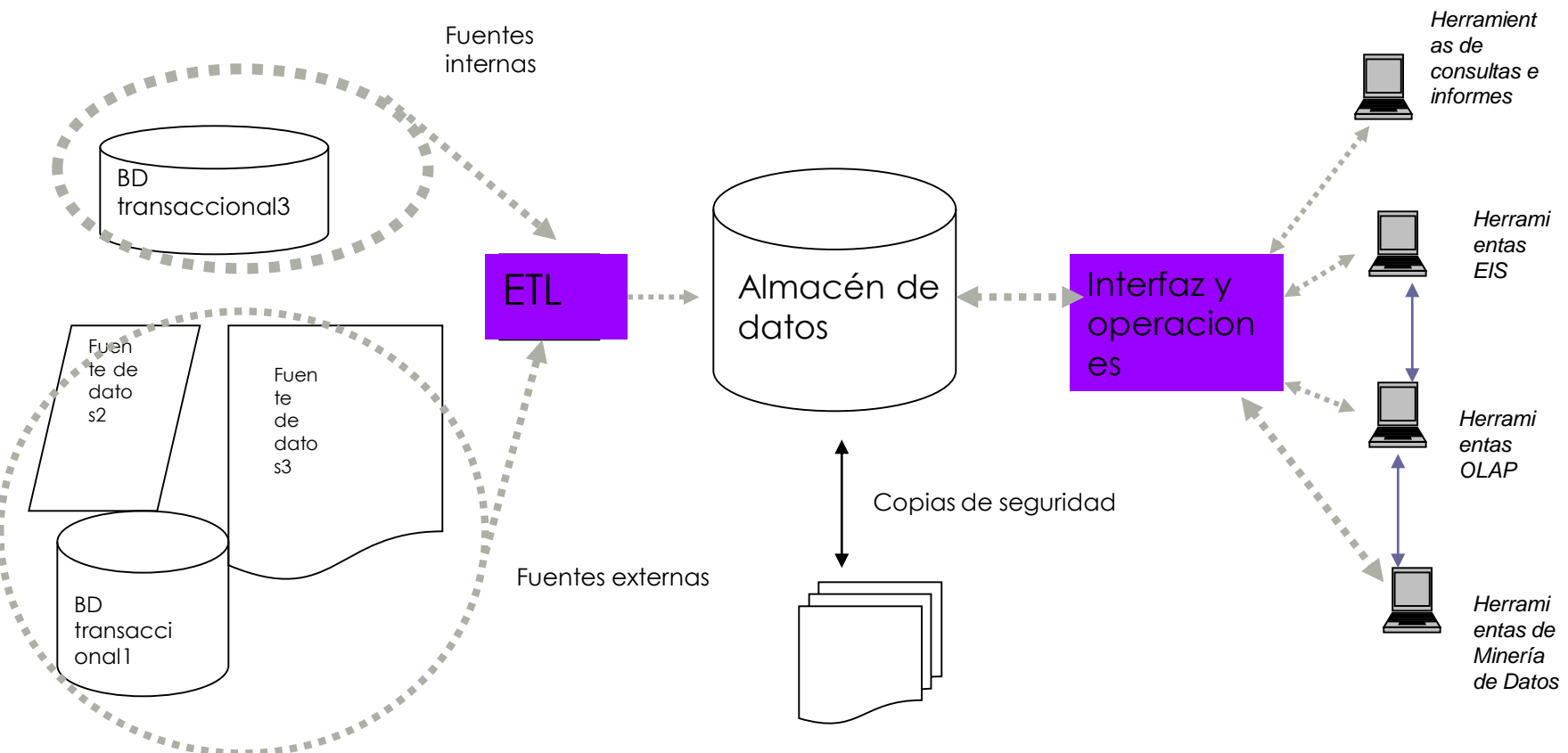


ÍNDICE

1. Introducción.
2. ¿Qué es Data Warehouse?.
3. **Arquitectura.**
4. Ventajas/Inconvenientes.
5. Diferencias entre una base de datos transaccional y un Data Warehouse

ARQUITECTURA

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



ARQUITECTURA

Sistema ETL:

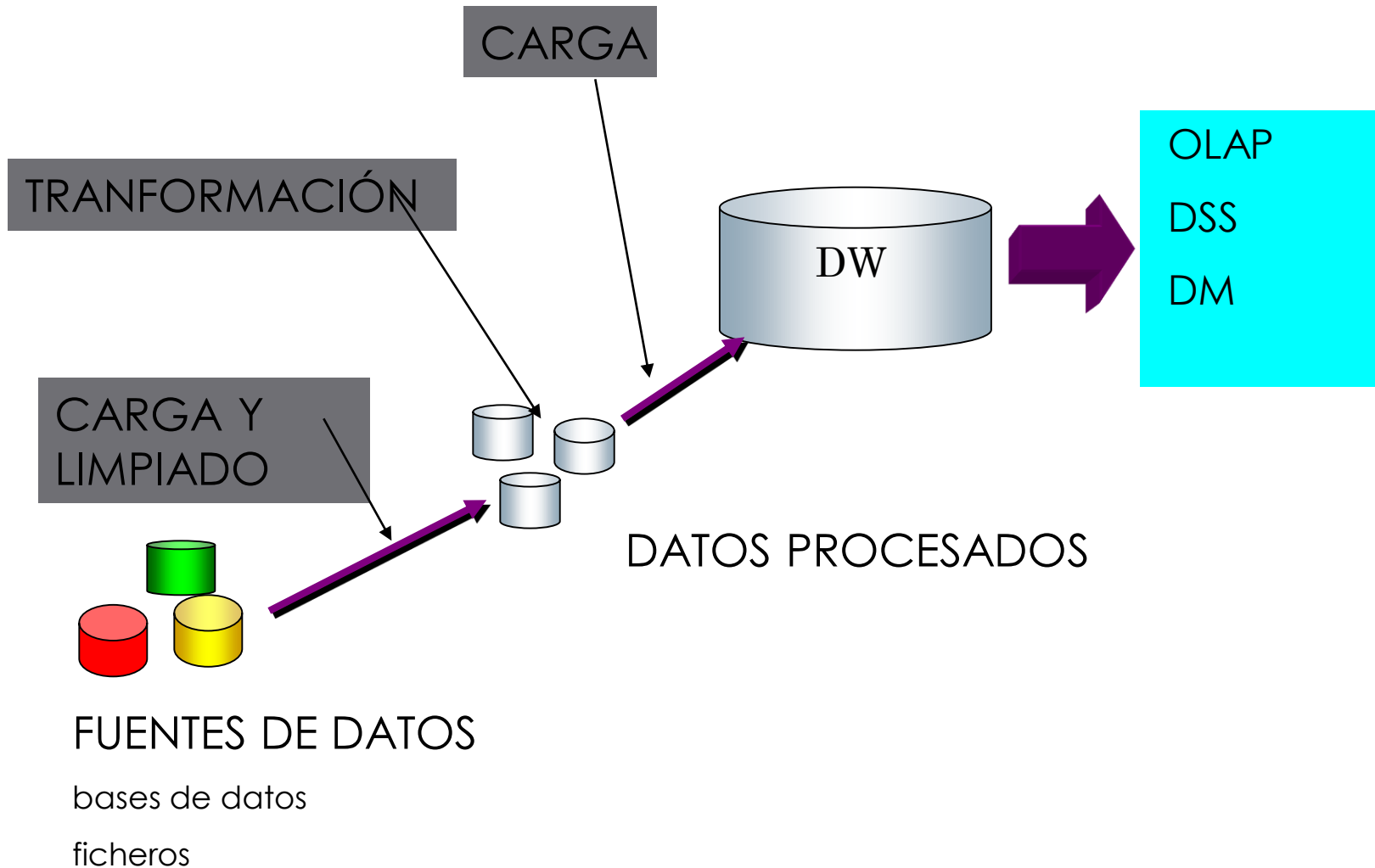
Realiza las funciones de *extracción* de las fuentes de datos (transaccionales o externas), *transformación* (limpieza, consolidación..) y *carga* del AD.

ARQUITECTURA

Interfaces y Operaciones de Consulta:

Permiten acceder a los datos y sobre ellos se conectan herramientas más sofisticadas (OLAP, EIS, minería de datos).

ARQUITECTURA



ÍNDICE

1. Introducción.
2. ¿Qué es Data Warehouse?.
3. Arquitectura.
4. **Ventajas/Inconvenientes.**
5. Diferencias entre una base de datos transaccional y un Data Warehouse

Ventajas e inconvenientes

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

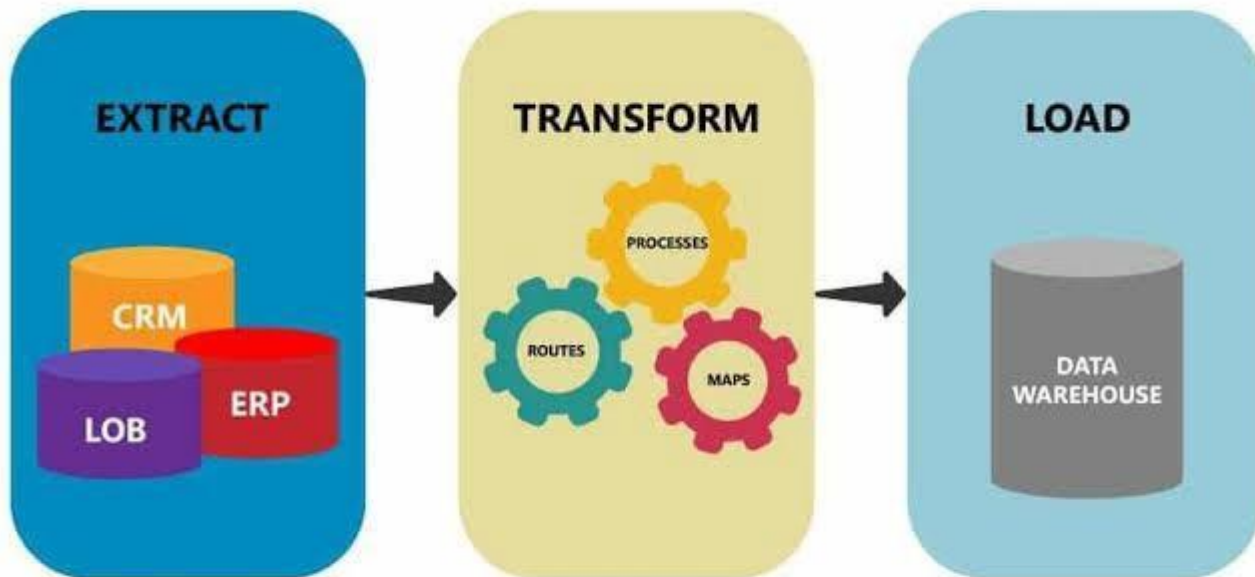
ÍNDICE

1. Introducción.
2. ¿Qué es Data Warehouse?.
3. Arquitectura.
4. Ventajas/Inconvenientes.
5. **Diferencias entre una base de datos transaccional y un Data Warehouse**

Diferencias entre una base de datos transaccional (OLTP) y un Data Warehouse (OLA)

	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica y táctica.
Tipo de datos	Operacionales.	Para la toma de decisiones.
Modelo de datos	Normalizado.	Desnormalizado.
Consulta	SQL.	SQL más extensiones.
Datos consultados	Actuales.	Actuales e históricos.
Horizonte de tiempo	60 - 90 días.	5 - 10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumarización.
Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
Número de transacciones	Elevado	Medio o bajo
Tamaño	Pequeño - Mediano.	Grande.
Tiempo de respuesta	Pequeño (segundos - minutos).	Variable (minutos - horas).
Orientación	Orientado a las aplicaciones.	Orientado al negocio.
Sello de tiempo	La clave puede o no tener un elemento de tiempo.	La clave tiene un elemento de tiempo.
Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y utilización.

Principios de diseño para procesos de ETL



ETL - Extract, Transform, Load

Índice

1. ¿Qué tan sencillos son los procesos ETL?
2. Para empezar: Requerimientos y arquitectura

Índice

1. **¿Qué tan sencillos son los procesos ETL?**
2. Para empezar: Requerimientos y arquitectura

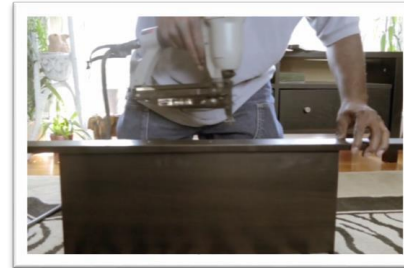
ETL? Si, Eso es Fácil!



Entender los
requerimientos



Identificar los datos
y herramientas



Construirlo



...y listo!

A Veces la Realidad es Otra



...y listo!

ETL Nunca Ha Sido Fácil

InformationWeek Reports, 2012 BI and Information Management Trends

Cual son los mayores impedimentos para el éxito del manejo de información en su empresa?

- 59% Acceso a información confiable
- 46% Integración de datos (ETL)
- 41% Saneamiento y de-duplicación de datos

The ETLData Warehouse toolkit, 2004. Joe Caserta / Ralph Kimball

70% a 80% del presupuesto y tiempo en proyectos de Data Warehouse lo consume el ETL

“ETL hace o deshace el Data Warehouse”

Índice

1. ¿Qué tan sencillos son los procesos ETL?
- 2. Para empezar: Requerimientos y arquitectura**

Requerimientos

Funcional

Contenido, información

Regulatorios

Regulaciones legales

Calidad de información

“Aptitud para su uso”

Seguridad

Backups, datos en transito, staging, código, etc.

Archivado, retención y linaje

Transparencia, auditoría, monitoreo

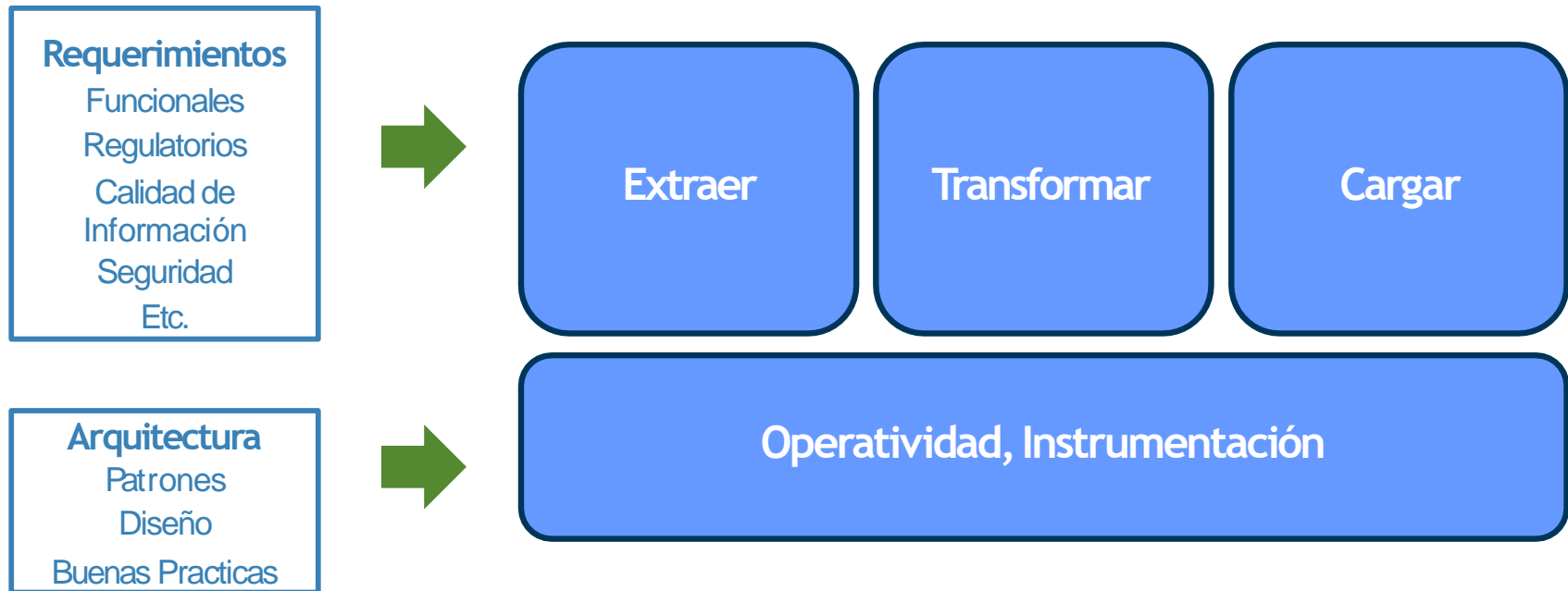
Interfaces de consumo

“Aptitud para su uso”

Habilidades técnicas disponibles

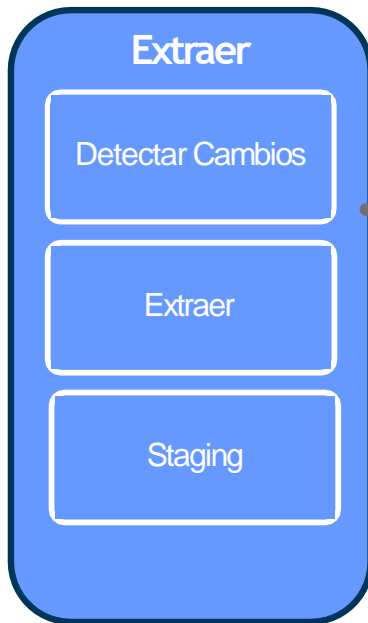
A veces la mejor herramienta es aquella que sabemos como usar

¿De que Depende un Buen Proceso ETL?



- **Fase 1: Extracción**

- El objetivo de un proceso ETL es producir datos limpios y accesibles que puedan utilizarse para analíticas u operaciones comerciales.



- Los datos en bruto deben extraerse de una variedad de fuentes, por ejemplo:
 - Bases de datos existentes
 - Registros de actividad como el tráfico de red, informes de errores, etc.
 - Rendimiento y anomalías de aplicaciones
 - Incidencias de seguridad
 - Otras actividades transaccionales que deben comunicarse para dar cumplimiento normativo



- **Fase 2: Transformación**

- La fase de transformación de los procesos de ETL aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.
- Estas directrices pueden ser declarativas, pueden basarse en excepciones o restricciones pero, para potenciar su pragmatismo y eficacia, hay que asegurarse de que sean:
 - Declarativas.
 - Independientes.
 - Claras.
 - Con una finalidad útil para el negocio.



- **Fase 3: Carga**

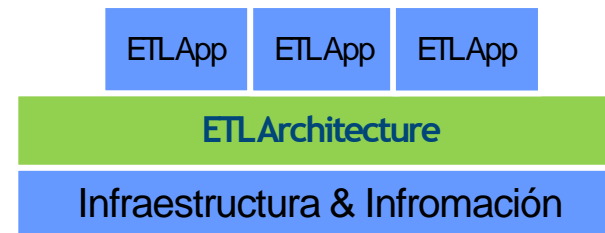
- La última fase de un proceso de ETL típico es la carga de esos datos extraídos y transformados a su nuevo destino.
- Existen dos vías habituales de cargar los datos a un almacén de datos: la carga completa y la carga incremental.

Beneficios De Una Buena Arquitectura

Consistencia y estabilidad

Reducción de costos: Mantenimiento, capacitación

Re-usabilidad



Actividad

- Estudiar el siguiente contenido de la pagina Web
- <https://www.dataprix.com/es/data-warehousing-y-metodologia-hefesto/34-datawarehouse-manager>