

# TEMA 4

## INTEGRACIÓN DE SISTEMAS

Natalia Padilla Zea, Eladio GarvÍ, José Samos

# Contenido

2

4.1. Integración de sistemas transaccionales y multidimensionales

4.2. Componente ETL

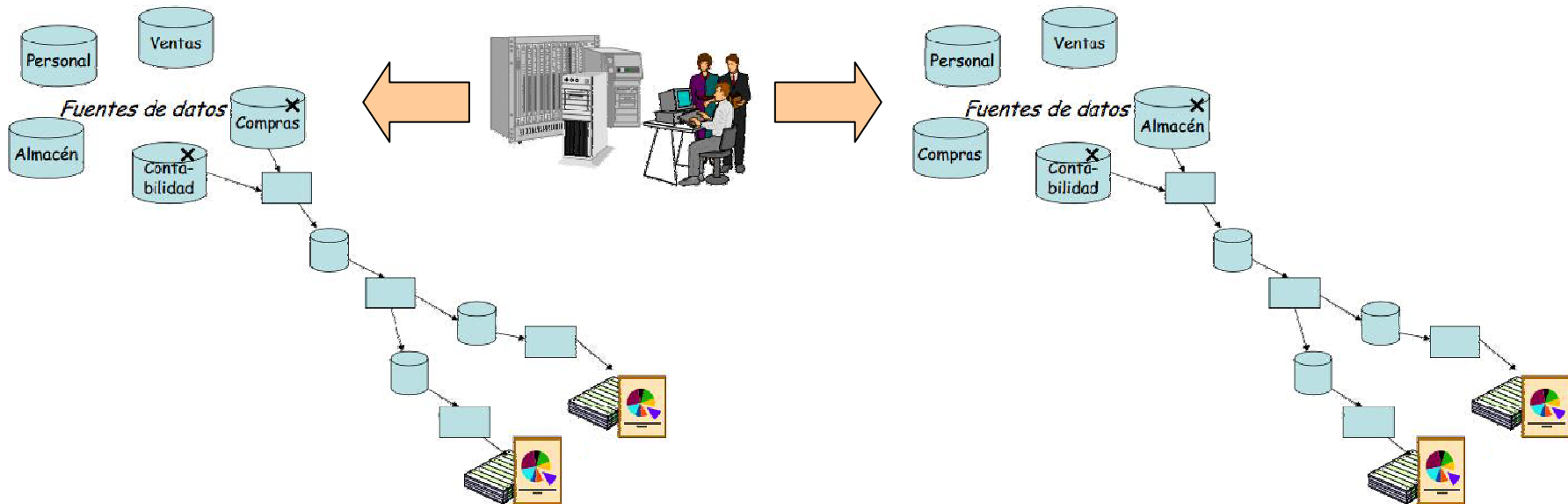
4.3. Metadatos

4.4. Definición de proyectos de integración de sistemas

# 4.1. Integración de sistemas

3

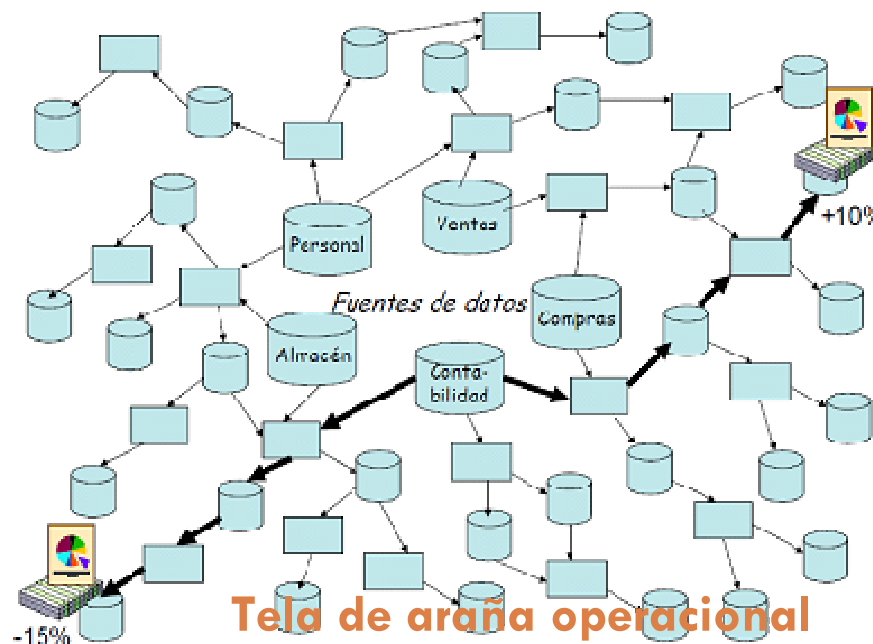
- Pasos para la obtención de un informe
  - ▣ Se usan varias fuentes de datos para generar distintos informes: localizar, extraer, transformar e integrar, cargar



# 4.1. Integración de sistemas

4

- La “tela de araña operacional” (W. Inmon, 1992)
  - ▣ A veces, se usan distintos orígenes para generar informes iguales. Y, en ocasiones, no coinciden.



## ■ Problemas:

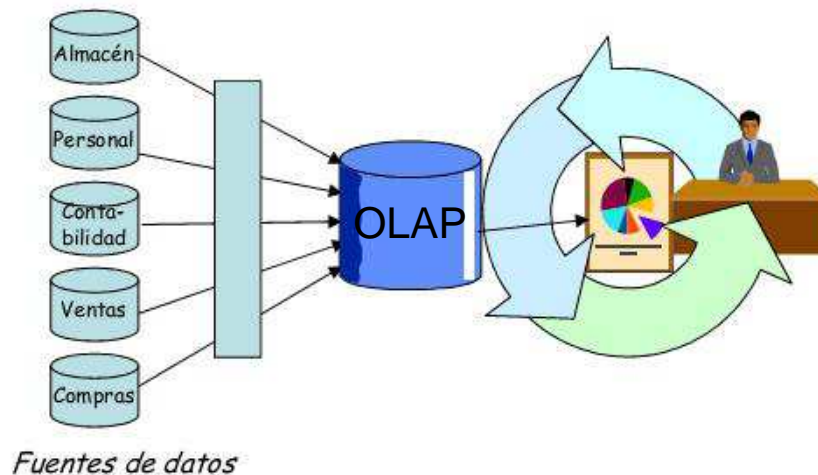
- Aumento de coste
- Tiempo requerido
- Falta de fiabilidad en los informes

## ■ Posible solución...

# 4.1. Integración de sistemas

5

- Construcción de un almacén de datos
  - ▣ Se integran las fuentes en un sistema multidimensional común para la empresa centrado en el foco de atención

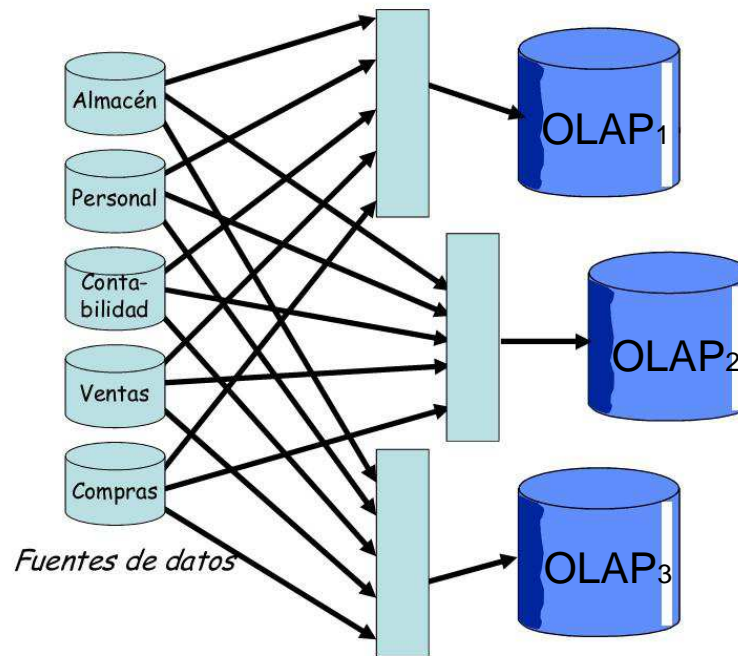


- Problema:
  - Existencia de varios focos de atención
- Posible solución...

# 4.1. Integración de sistemas

6

- Construcción de varios almacenes de datos
  - ▣ Si se tienen varios focos de atención a partir de un mismo conjunto de fuentes de datos



## ■ Problema:

- Proliferación de sistemas OLAP

## ■ Solución...

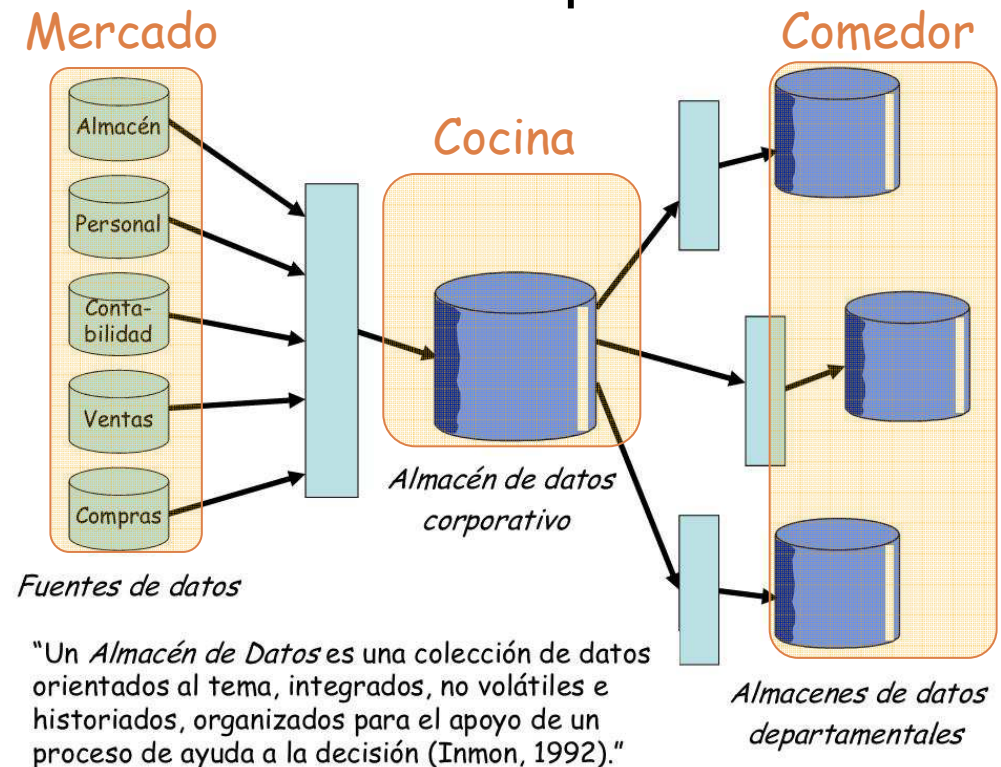
# 4.1. Integración de sistemas

7

- La fábrica de información corporativa (FIC)
  - Si hay un programa de integración para cada fuente de datos; se parece a la tela de araña operacional

- Se crea un almacén de datos corporativo para montar la FIC
- Entran datos y sale información

## □ Denominación de Kimball



## 4.2. Componente ETL

8

- ETL: Extraction, Transformation, Load
- Proceso global de ETL
  - ▣ Tareas a realizar una vez:
    - Identificar la fuente de datos
    - Identificar los datos objetivo
    - Crear una correspondencia entre la fuente y el objetivo
    - Definir el modo de replicación de datos
    - Programar la replicación



## 4.2. Componente ETL

9

- Proceso global de ETL:

- ▣ Tareas a realizar frecuentemente:

- Capturar los datos necesarios de la fuente
    - Transferir los datos entre fuente y objetivo
    - Transformar los datos capturados
    - Aplicar los datos capturados al objetivo
    - Confirmar el éxito o no de la replicación
    - Documentar el resultado en los metadatos
    - Mantener las definiciones de fuentes, objetivos y correspondencias (cuando se requiera)

## 4.2. Componente ETL

10

- Localización de datos
  - ▣ En diferentes sistemas
    - BD locales
    - Web
    - Catálogos
    - ...
  - ▣ Características de los datos
    - Precisión
    - Completitud
    - Atemporales
    - Compatibilidad estructural
    - Proximidad a la fuente original (¡Muy importante!)

## 4.2. Componente ETL

11

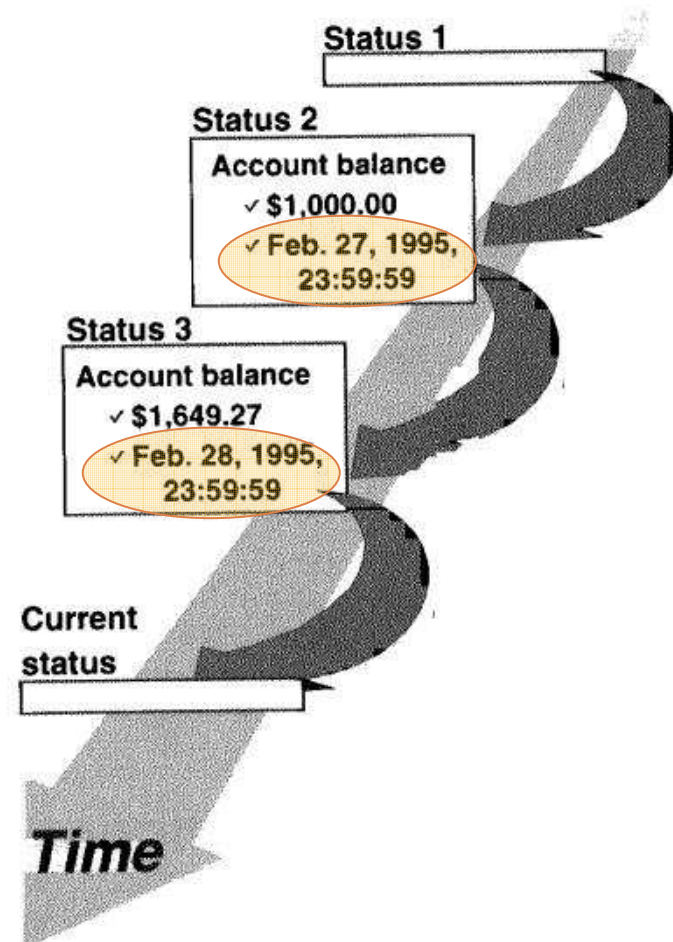
- Extracción de datos (E)
  - ▣ Extraer los datos requiere:
    - Carga inicial (una vez)
    - Actualizaciones (tantas como se necesiten)
  - ▣ Enfoques para la extracción:
    - Diferido
    - Inmediato

## 4.2. Componente ETL

12

### □ Modo de extracción de datos “Diferido”

- Sólo se miran los datos cada cierto tiempo y se recoge ese valor
- Se toma un resumen al final de un periodo
- Problema: se pierde detalle

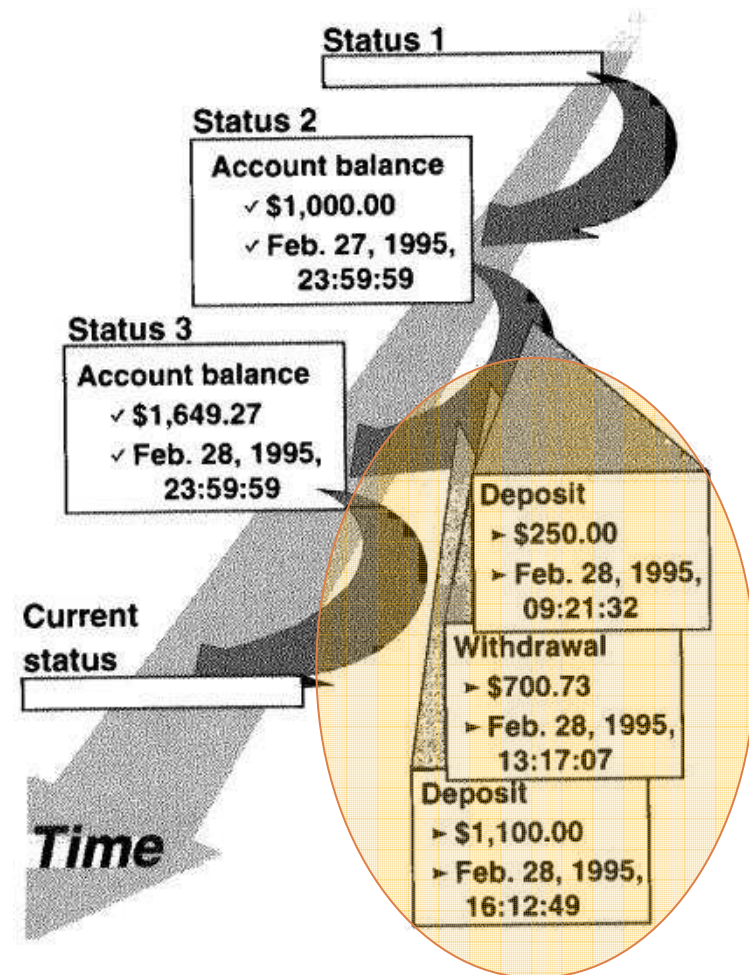


## 4.2. Componente ETL

13

### □ Modo de extracción de datos “Inmediato”

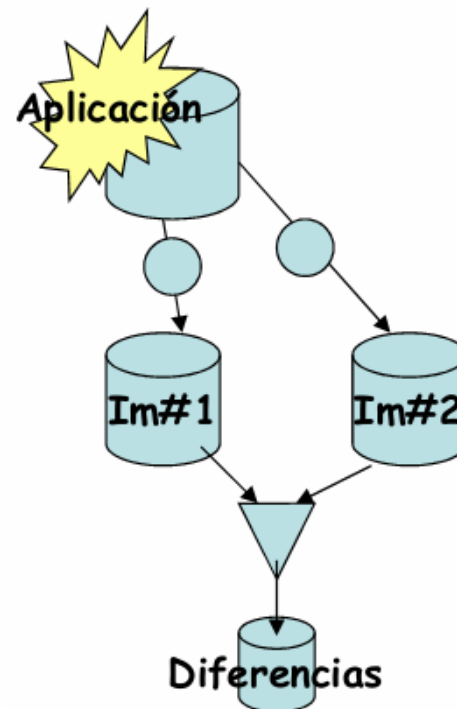
- Se registra cada cambio que se produzca en los datos
- Se ven todos los movimientos (no pierde detalle)



## 4.2. Componente ETL

14

- Métodos de extracción de datos
  - ▣ Caso 1: comparación de imágenes
    - Se mantiene copia de la imagen anterior y se compara con la actual, cada cierto tiempo, para ver las diferencias
    - Es un enfoque diferido



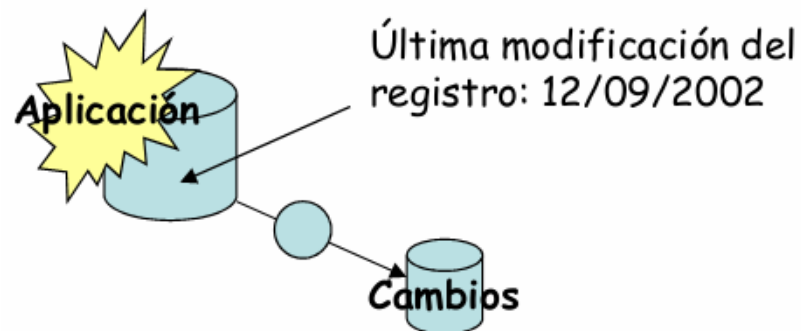
## 4.2. Componente ETL

15

### □ Métodos de extracción de datos

#### ▣ Caso 2: huella de tiempo

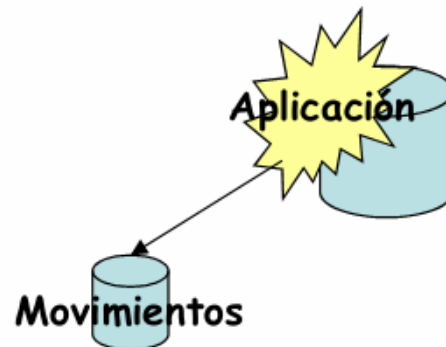
- Se anota la fecha de última modificación y se registran los cambios encontrados en la fecha actual respecto a la otra
- Es un enfoque diferido porque se pierden cambios intermedios



## 4.2. Componente ETL

16

- Métodos de extracción de datos
  - ▣ Caso 3: registrar movimientos
    - La aplicación que maneja los datos también guarda todos los movimientos cuando se producen
    - Es un enfoque inmediato

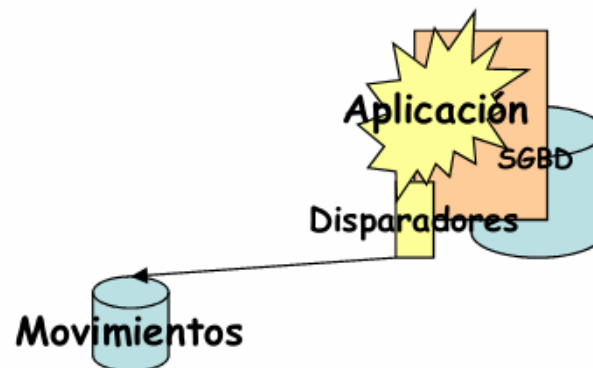




## 4.2. Componente ETL

17

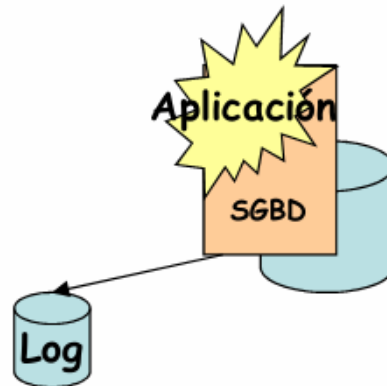
- Métodos de extracción de datos
  - ▣ Caso 4: registrar movimientos mediante disparadores
    - Teniendo un SGBD para los datos, incluir disparadores que guarden todos los movimientos cuando se producen
    - Es un enfoque inmediato



## 4.2. Componente ETL

18

- Métodos de extracción de datos
  - ▣ Caso 5: registrar movimientos mediante archivos “log”
    - Teniendo un SGBD para los datos, usar los archivos de log (o bitácora) para obtener todos los movimientos producidos
    - Es un enfoque inmediato, pero los movimientos tienen mucho ruido



## 4.2. Componente ETL

19

### □ Métodos de extracción de datos

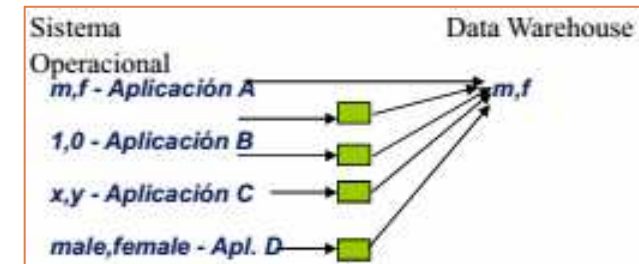
#### ▣ ¿Cuál es mejor?

- Considerar el enfoque
- Tener en cuenta la fuente de datos
- Analizar las posibilidades del software disponible

## 4.2. Componente ETL

20

- Transformación e Integración (T)
  - ▣ Adaptar los datos al modelo del Data Warehouse
    - Hay que volver a codificar, unificar formatos
  - ▣ Comprobar:
    - Datos incompletos
    - Datos duplicados (no necesariamente iguales)
    - Datos erróneos o inconsistentes
    - Datos vacíos o ilegibles
    - Diferencias de codificación
    - Agregaciones necesarias...
  - ▣ Notificar los errores a las fuentes



## 4.2. Componente ETL

21

- Carga (L)
  - ▣ Una vez transformados los datos convenientemente se incorporan al sistema
  - ▣ ¿Cuándo se hace?
    - Inicial: todos los datos de las fuentes para cargar el SMD
    - Actualizaciones periódicas: con los datos modificados
    - Agregaciones, resúmenes: actualizar los agregados

## 4.2. Componente ETL

22

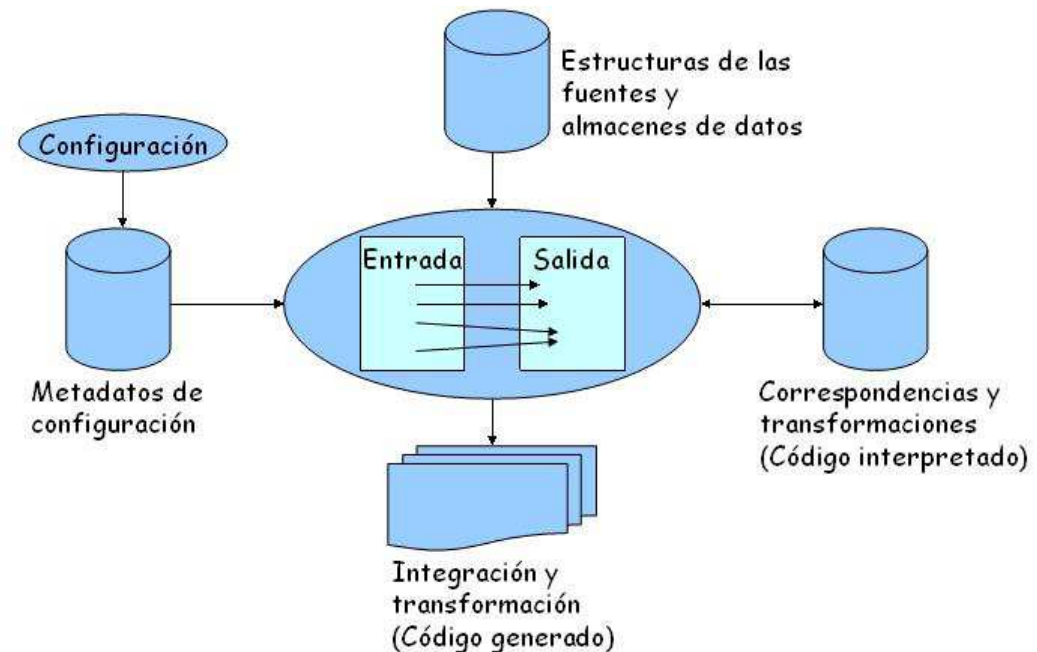
### □ Soporte al proceso de ETL

#### ▣ Herramientas:

- Integration Services (en MS SQL Server)
- Kettle (o Pentaho Data Integration -PDI-, en Pentaho Analysis)

#### ▣ Opciones de código:

- Generado
- Interpretado



## 4.3. Metadatos

23

### □ Características:

- ▣ Son datos sobre los datos
- ▣ Facilitan la tarea de ETL
- ▣ Se encuentran en las fuentes de datos, en el programa de ETL, en el SMD...

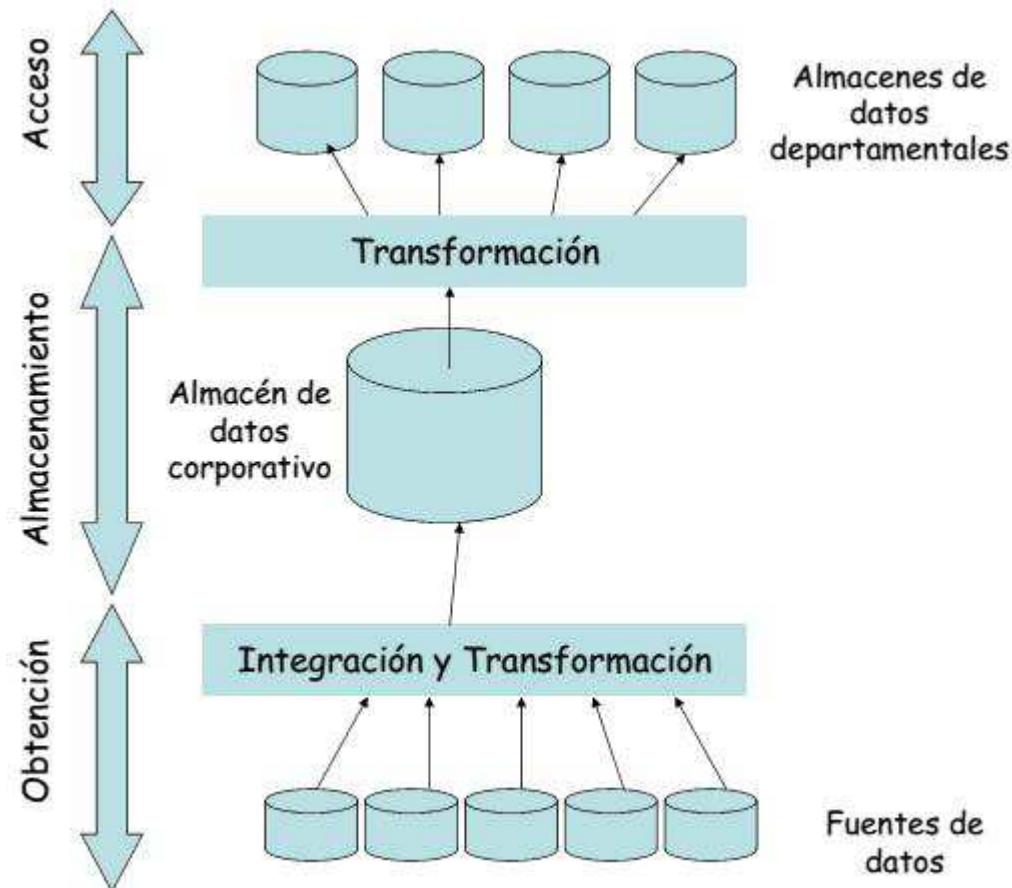
### □ Tipos:

- ▣ Construcción: de la estructura de los datos
  - Tipo, longitud, ...
- ▣ Funcionamiento:
  - Resultados de cargas, tiempo de consulta, registros de error...
  - Uso: sobre los informes elaborados

## 4.4. Definición de proyectos

24

- Componentes de los proyectos de integración de sistemas

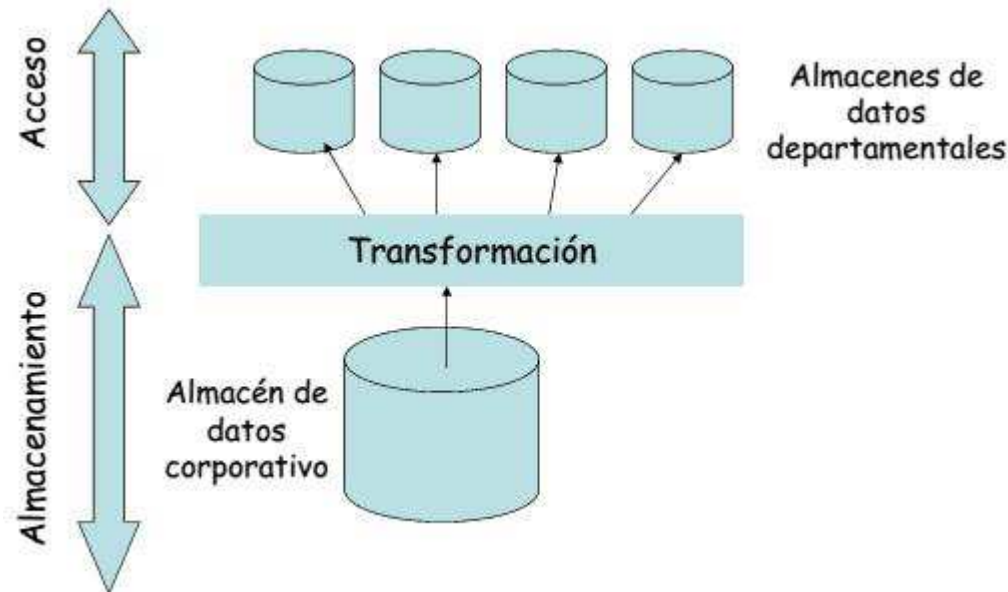




## 4.4. Definición de proyectos

25

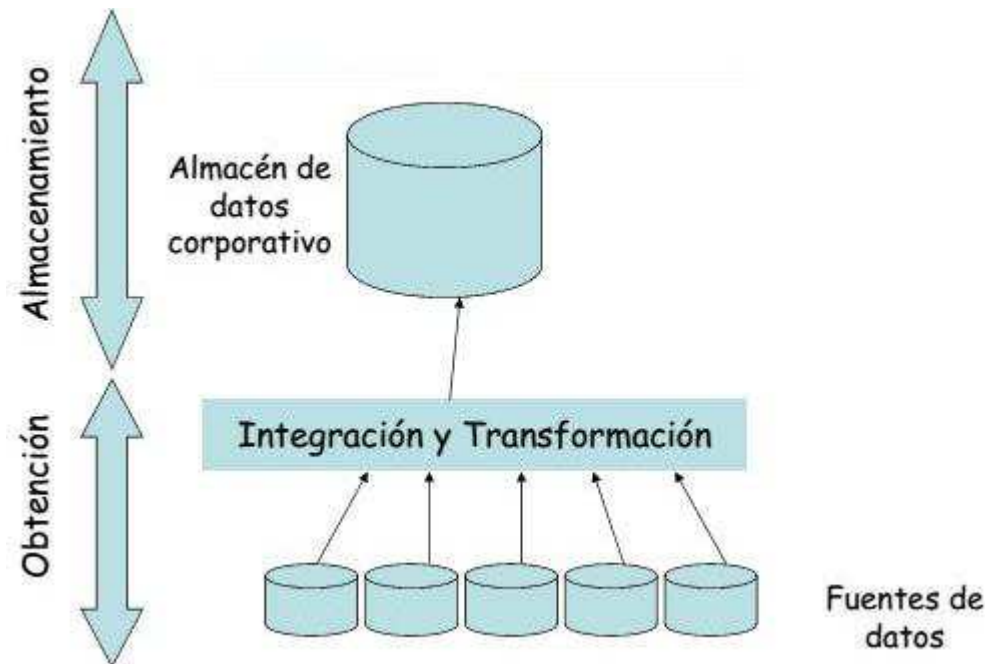
- Enfoque orientado por requisitos
  - ▣ Sólo se miran los requisitos de los usuarios
  - ▣ Se hace el diseño y la implementación
  - ▣ Se simulan los datos
  - ▣ Problema: en el sistema real no funciona



## 4.4. Definición de proyectos

26

- Enfoque orientado por datos
  - ▣ Se cogen los datos de las fuentes
  - ▣ Se vuelcan en el SMD
  - ▣ Problema: no tiene en cuenta las necesidades del usuario



## 4.4. Definición de proyectos

27

### □ Enfoque mixto

#### ▣ Características de las operaciones a realizar:

- Continuas: tratan los aspectos de ETL de forma iterativa
- Basadas en valor: teniendo en cuenta el beneficio que se va a obtener de cada operación; es difícil de estimar
- Autónomas: cada una trabaja sobre aspectos diferentes y construye una parte, teniendo en cuenta desde los datos hasta el decisor
- Deben contar con la infraestructura operacional y la física
- No es importante el orden en que se tratan las partes

## 4.4. Definición de proyectos

28

### □ Enfoque mixto

