

Introducción a la Inteligencia de Negocio

Tema 1

Profesores:

Juan C. Trujillo,
LUCENTIA Research Group



Universitat d'Alacant
Universidad de Alicante



Departamento de
Lenguajes y Sistemas
Informáticos

INGP. 2021

Indice

- 1. Introducción asignatura y planteamiento de proyectos
- 2. Introducción a las fuentes de datos y Open Data
- 3. Teoría sobre BI - Modelo conceptual (4h)
 - Ejercicio práctico de fútbol utilizando Mondrian Schema Workbench
- 4. Teoría sobre BI - Modelo lógico (4h)
 - Ejercicio modelo lógico de fútbol e implementar sobre MySQL

Indice

☐ 5. ETLs

- ☐ Ejercicio con Pentaho Data Integration para transformación de datos

☐ 6. Cuadros de mando, Visualizaciones y análisis OLAP sobre Pentaho

- ☐ Ejercicio análisis OLAP del ejercicio de fútbol

Indice

☐ 7. Otras herramientas disponibles para el análisis:

☐ Primera parte de la clase

- ☐ Power BI

- ☐ Librerías de charting

- ☐ Spark

- ☐ Apache Zeppelin

- ☐ Qlink

☐ Segunda parte

- ☐ Planificación de proyecto con fuentes de datos y objetivos de análisis

Indice

- 8, 9, 10, 11, 12 Clases de trabajo sobre los proyectos y consultas al profesorado sobre su evolución (3 horas)
- 8. Fundamentos de análisis (Tipos de análisis y para que se utilizan: predictivo, descriptivo, prescriptivo) → 1 hora
- 9. Técnicas de Inteligencia Artificial (tipos de técnicas, métodos de aplicación, frameworks y herramientas) → 1 hora
- 10. Estrategias de negocio, Indicadores y toma de decisiones (1h) → 1 hora
- 11. Casos de éxito (1h)
- 12. Tendencias actuales y futuras (1h)

Indice

- Preguntas, dudas, problemas sobre la práctica antes de la exposición (4h)
- 1 4. Exposición de los proyectos (4h)

Indice

7

- Motivación y nuevas necesidades de gestión de datos
- El almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y analíticos
- Arquitecturas de ADs
- Introducción NoSQL

Nuevas necesidades de gestión de datos

8

- Entornos económicos altamente competitivos
 - ▣ Típica pregunta a SGBDR
 - ¿Cuántos zapatos vendimos el último mes?
 - ▣ Empresas necesitan adoptar decisiones estratégicas
 - ¿Cuántos zapatos del 41 de color rojo se vendieron el último mes en la zona norte, este y sur; comparados con las ventas del mismo mes el año pasado ?
 - ¿Qué tipo de cliente me ha estado comprando el BMW 320i durante los últimos 10 años ?
 - ¿Directivo, profesor, trabajador escala básica,...?

Nuevas necesidades de gestión de datos

9

- ▣ Req. 1. Gran volumen de datos (varios años, clientes, productos, almacenes, etc.)
 - Históricos y procedentes de distintas fuentes

- ▣ Req. 2. Tienen que ser presentados en un entorno amigable y fácil de usar
 - Entender el tipo de preguntas “estratégicas”

Nuevas necesidades de gestión de datos

10

□ ¿Son válidos los sistemas OLTP para tales decisiones?

▣ Problemas

- Datos históricos no disponibles en sistemas diarios OLTP
- Normalmente en distintas fuentes de datos
 - Proveedores, Clientes, componentes, productos defectuosos, etc.
- Los directivos no saben manejar tales sistemas y,
- Rendimiento, errores, etc....

Nuevas necesidades de gestión de datos

11

- El uso de OLTP requeriría:
 - ▣ Integrar datos → Consumir tiempo (req. 1)
 - ▣ ¿Podría un analista manejar herramientas transaccionales ? (req. 2)

- Vamos a hacer breve historia sobre las soluciones planteadas para manejar datos históricos

Nuevas necesidades de gestión de datos

Aproximaciones tradicionales a datos históricos

12

□ Datos procedentes de sistemas heredados

- 1970's se hace uso de IBM mainframes

 - Cobol, CICS, IMS, DB2, etc

- 1980's plataformas AS/400 y VAX/VMS

- Hoy en día muchas aplicaciones de negocio “corren” sobre estos sistemas

 - Muchos años recogiendo datos y reglas de negocio → dificultad para llevarlos a otro sistema

 - Los datos se vuelcan en librerías donde accederán otras aplicaciones de negocio

- Coste de aplicaciones de negocio es grande

Nuevas necesidades de gestión de datos

Aproximaciones tradicionales a datos históricos

13

- Datos extraídos en el escritorio (1990's)
 - ▣ Se reduce la distancia entre usuario final y programador
 - PC con hojas de cálculo, herramientas de análisis, etc.
 - ▣ Herramientas de análisis que acceden a datos producidos por sistemas heredados
 - ▣ Problema: datos permanecen **fragmentados** y están orientados a necesidades específicas de usuarios final
 - Soluciones parciales
 - No todos usuarios tienen la misma destreza

Nuevas necesidades de gestión de datos

Aproximaciones tradicionales a datos históricos

14

- Sistemas de apoyo a la decisión (SAD) y Sistemas de Información de Ejecutivos (SIE):
 - ▣ SAD: Datos en **detalle**. Ejecutivos medios y bajos
 - ▣ EIS: Datos **consolidados**. Altos ejecutivos
 - Más orientados a vista multidimensional de los datos
 - ▣ Son similares y solapan funcionalidades
 - ▣ Son los precursores de los Almacenes de datos
 - ▣ Alto precio y de nuevo descoordinación sobre los datos necesarios para efectuar el análisis

Nuevas necesidades de gestión de datos

Aproximaciones tradicionales a datos históricos

15

□ Resumen características SAD y EIS:

- Datos descritos en **términos de negocio**, en lugar de términos técnicos como tupla, fila o tabla.

 - Sistemas enfocados a usuarios no técnicos

- Datos **preprocesados** siguiendo patrones de reglas de negocio

 - Beneficios por la venta de productos en distintos almacenes

- Vistas **consolidadas** de datos

 - Aunque permiten ver datos en detalles, rara vez pueden acceder a todos los datos en detalle

Nuevas necesidades de gestión de datos

16

- Las técnicas de almacenes de datos proporcionan las herramientas analíticas de sus precursores
- Proporcionan soluciones **globales** a la organización en lugar de soluciones particulares
 - ▣ Datos orientados a satisfacer a la organización entera

Indice

17

- Motivación y nuevas necesidades de gestión de datos
- El almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y analíticos
- Arquitecturas de ADs

El almacén de datos: primera aproximación

18

□ Evolución de los sistemas de almacenes de datos

▣ Desde los primeros sistemas de gestión de ficheros (a. 1970) hasta los Sistemas de Gestion de Bases de Datos actuales (SGBD)

- SGBD eficientes
- SGBD robustos
- Múltiples herramientas de alto nivel que facilitan su manejo
 - Servidor
 - Cliente

El almacén de datos: primera aproximación

19

- Codd en los 90 establece 12 reglas OLAP
 - ▣ Investigación enfocada al tratamiento de datos relacionales y haciendo uso de tecnología multidimensional
- A finales de 90 se han demandado herramientas OLAP potentes de este tipo
 - ▣ Unas utilizan tecnología relacional como base (ROLAP)
 - ▣ Otras utilizan datos directamente en forma multidimensional (MOLAP)

El almacén de datos: primera aproximación

20

- Empresas en la actualidad
 - ▣ Aproximadamente el 90% de SGBD son relacionales
 - Integridad de los datos
 - Independencia de los datos, etc.
 - ▣ Orientados a los procesos diarios de la empresa
 - Sistemas de Procesamiento Transaccional en Línea (*On-Line Transactional Processing, OLTP*)
 - Compras de productos, ventas, pedidos, gestión de clientes, etc.
 - ▣ Datos históricos → almacenamientos externos (cintas, discos,...)

El almacén de datos: primera aproximación

21

- El Almacén de datos (Data Warehouse, DW)
 - ▣ Sistemas que almacenan datos históricos para ser utilizados por los SAD
 - ▣ Son sistemas eminentemente de consulta enfocados a extraer conocimiento de los datos históricos almacenados
 - ▣ El análisis de los datos → On-Line Analytical Processing (OLAP)
 - Utilizan el modelado multidimensional (cubos, hipercubos, etc)

El almacén de datos: primera aproximación

22

- Definición según W. Inmon (uno de los “padres”) (1992)
- “Una colección de datos **orientados por tema**, **variables en el tiempo** y **no volátiles** que se emplea como apoyo a la toma de decisiones estratégicas”

El almacén de datos: primera aproximación

23

□ Orientados por tema

▣ El diseño enfocado a responder eficientemente a consultas estratégicas

- Actividades de interés: compra, ventas, alquileres,...
- Contexto de análisis: clientes, vendedores, productos, etc...

▣ El modelado Multidimensional (primera aproximación)

- Hechos → actividades de interés
- Dimensiones → contexto de análisis

El almacén de datos: primera aproximación

24

- Integrados
 - ▣ Datos integrados de distintas fuentes de datos operacionales
- Variables en el tiempo
 - ▣ Datos relativos a un periodo de tiempo y se incrementan periódicamente
- No volátiles
 - ▣ Los datos almacenados no se modifican ni actualizan nunca, sólo se añaden nuevos datos

El almacén de datos: primera aproximación

25

- Objetivo técnicas de almacenamiento de datos (Data Warehousing)
 - ▣ Ofrecer información histórica para la toma de decisiones integrando la información procedente de distintas fuentes de datos operacionales

El almacén de datos: primera aproximación

26

- Las tecnologías de almacenamiento de datos integran eficientemente tecnología de bases de datos con análisis de datos
 - Bases de Datos: Un SGBD que soporte el almacén
 - Análisis de datos: herramientas que permitan llevar a cabo un análisis sencillo de los datos
 - Las más extendidas: herramientas OLAP
 - Análisis Multidimensional basado en el modelo multidimensional

El almacén de datos: primera aproximación

27

- Ejemplo: conocer los **ventas de productos** con respecto a los **productos** vendidos, los **almacenes** donde se vendieron y el **tiempo**

Ventas			Producto.Grupo = "Supermercado"			
			Comida		Bebida	
			Cong	Fresco	Refresco	Alcohol
Almacén. comunidad = "Comunidad Valenciana"	Alicante	Albatera	100	200	300	400
		Elche	500	600	700	800
	Valencia	Sagunto	900	1000	1100	1200
		Cullera	1300	1400	1500	1600

El almacén de datos: primera aproximación

28

□ Ventajas para las empresas

- ▣ Decisiones soportadas por datos fiables
 - ▣ Rentabilidad de inversiones
 - ▣ Aumentar la competitividad en los nuevos entornos hostiles
 - ▣ Entornos amigables → los directivos analizan por ellos mismos los datos
- Por fin se entienden con los ordenadores → lo hemos conseguido !!!

El almacén de datos: primera aproximación

29

□ Inconvenientes

- ▣ Infravalorar los recursos necesarios para alimentar el DW a partir de datos operacionales
- ▣ No planificar el alto esfuerzo para lograr un buen diseño
- ▣ Nunca está acabado → incremento continuo de requerimientos ad-hoc
- ▣ ¡¡¡¡ Cuidado !!!! A más cantidad de datos → Más seguridad

Indice

30

- Motivación y nuevas necesidades de gestión de datos
- El almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y analíticos
- Arquitecturas de ADs

Sistemas operacionales vs. Almacenes de datos

31

	OLTP	OLAP
Usuario	■ Profesional de TI	Analista de Información
Función	■ Operaciones diarias	Apoyo a la decisión
Diseño de BD	■ Orientada a aplicación (Basado en EE-R)	Orientado al tema (esquema estrella, copos)
Datos	■ Actuales, Aislados	Históricos, Consolidados
Vistas	■ Detallados, Planos, Relac.	Agregados
Destino/utilización	■ Estructuradas, repetitivas	Ad-Hoc
Unidades de trabajo	■ Transacciones simples	Consultas complejas
Acceso	■ Lectura/escritura	Lectura principalmente
# Registros accedidos	■ Decenas	Millones
# Usuarios	■ "Miles"	"Centenares"
Tamaño de la BD	■ 100 MB-GB	100 GB-TB
Medidas de rendimiento	■ N° transacciones	N° consultas, Respuesta

Indice

32

- Motivación y nuevas necesidades de gestión de datos
- El almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y analíticos
- Arquitecturas de ADs

Indice

33

- Arquitecturas de DW
- Esquema tradicional de un DW
- Aplicaciones de DW en el mercado
- Aproximaciones de diseño

Arquitecturas de DW

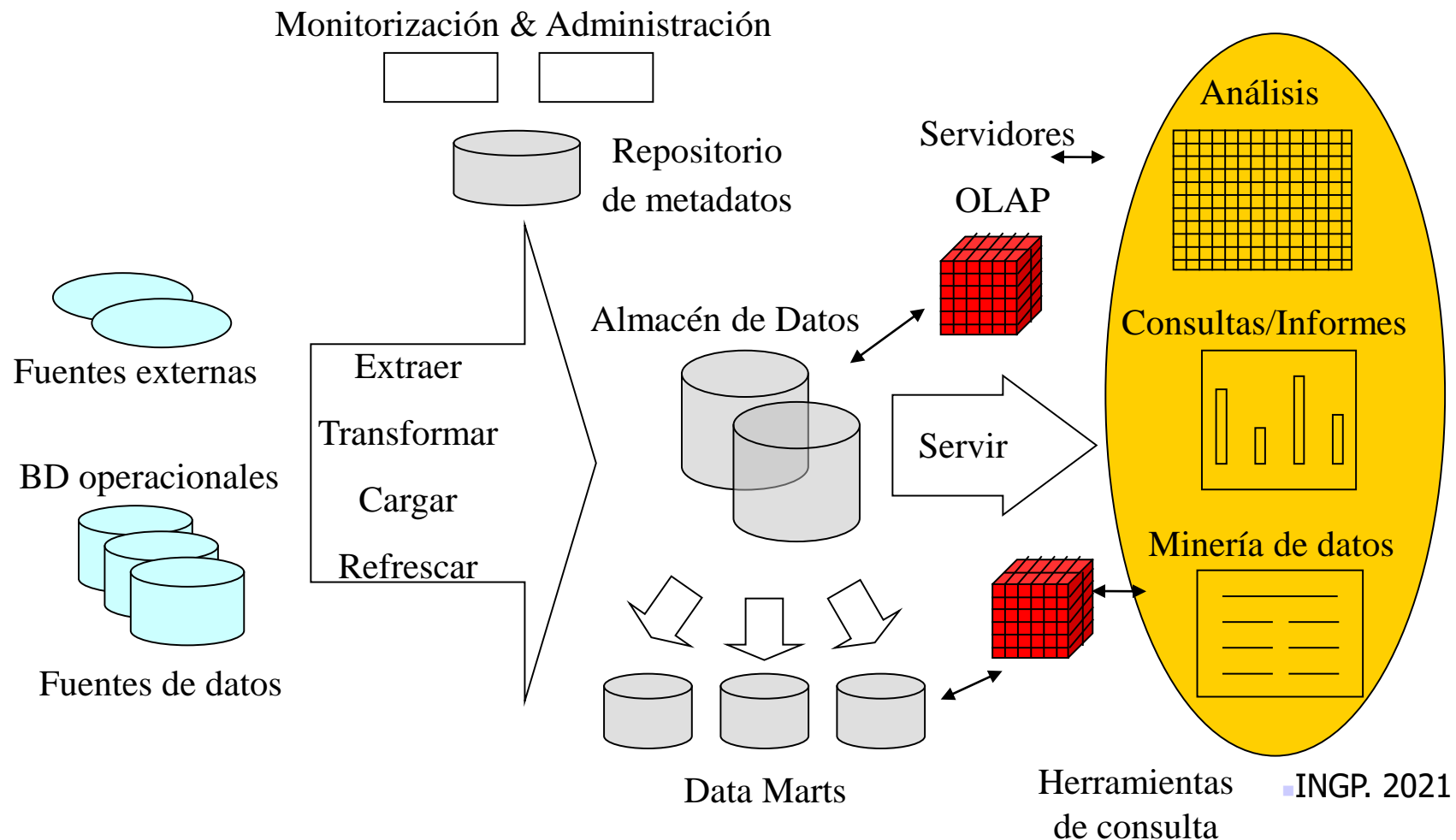
34

- Existen distintas **variantes** de la arquitectura de un DW según las necesidades del negocio
- Algunas de ellas:
 - Arquitectura **tradicional**: Diseñada para análisis de grandes cantidades de datos estructurados
 - En **tiempo real**: Cuando las decisiones requieren de datos tan pronto como se generan
 - Para “**Big Data**”: Cuando la información a tratar incluye información no estructurada (redes sociales!)

Arquitecturas de DW

35

□ Esquema general:



Arquitecturas de DW

36

- Arquitectura de tres capas
 - ▣ Servidor del repositorio o base de datos del almacén de datos
 - Casi siempre un SGBD Relacional
 - ▣ Servidores OLAP
 - Relational OLAP (ROLAP)
 - Extiende SGBD relacionales para permitir operaciones MD
 - Multidimensional OLAP (MOLAP)
 - Directamente implementa el modelo MD en vectores

Arquitecturas de DW

37

- Arquitectura de tres capas (II)
 - ▣ Clientes → Herramientas
 - Informes y consultas
 - OLAP (On-Line Analytical Processing)
 - Data Mining

Arquitecturas de DW

38

□ ¿Por qué Data Warehouse separado?

▣ Rendimiento (Performance)

- Consultas OLAP complejas → Ralentización del servidor
- Métodos de implementación, accesos, etc. distintos

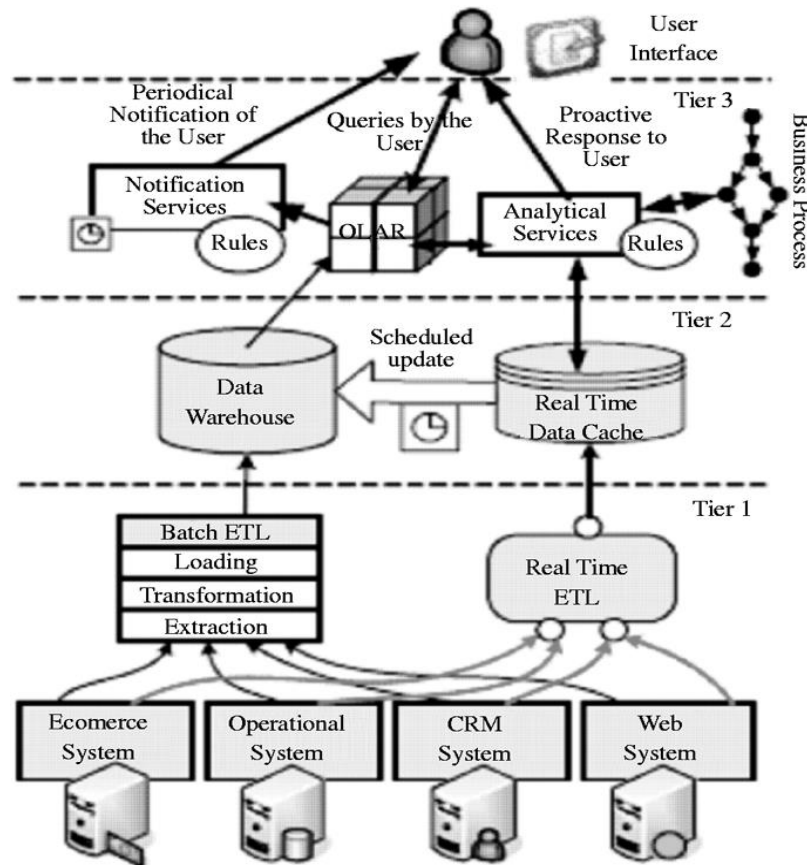
▣ Funcionalidad

- Datos no existentes → históricos
- Datos consolidados (agregados, sumados, resumidos, etc.) de distintas fuentes
- Calidad de datos
 - Diferentes fuentes → representaciones distintas, etc...

Arquitecturas de DW

39

- Arquitectura tradicional: más en detalle en la [siguiente sección](#)
- En tiempo real:



Source: Nguyen Manh *et al.* (2005)

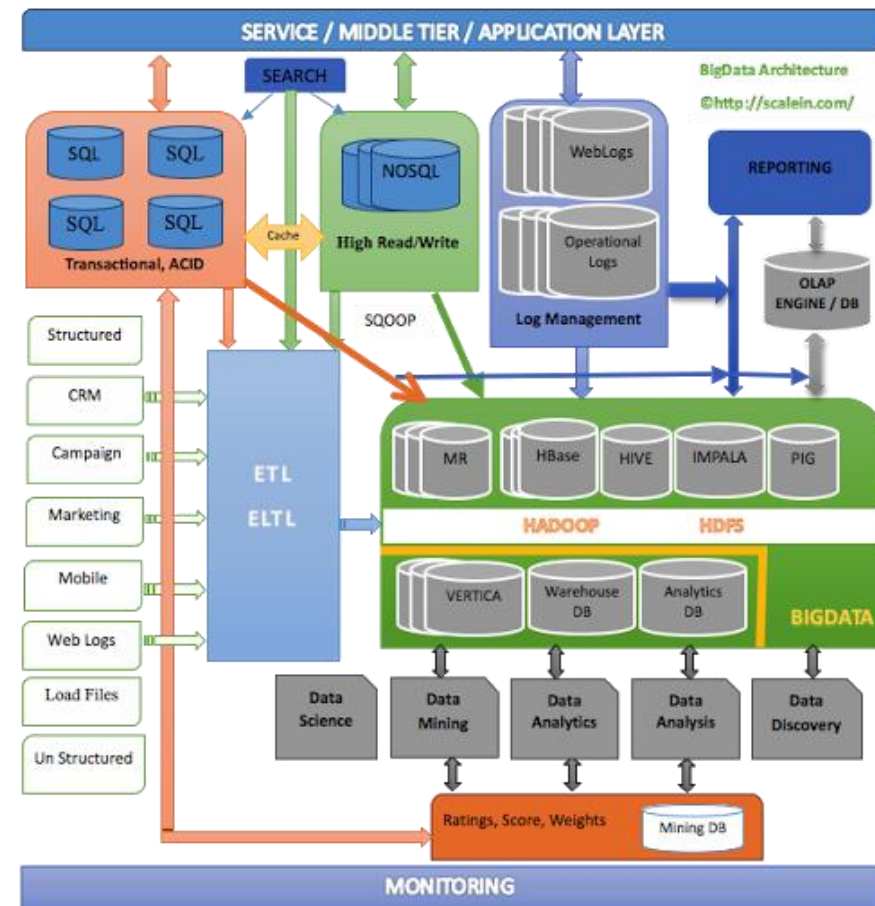
Arquitecturas de DW

Arquitectura BigData típica

Características:

- ☐ Almacenamiento de diferentes tipos de datos
 - Semi-estructurados (Marketing/ campañas/ móvil/ web logs)
 - Estructurados
 - Ficheros de log
- ☐ Carga de datos desde diferentes bases de datos (MySQL, Oracle, PostgreSQL, MongoDB, etc)
- ☐ Minería de datos
- ☐ Analíticas
- ☐ Almacenes de datos para reporting
- ☐ Análisis por lotes (Hadoop)
- ☐ Web caching
- ☐ Search

Imagen via (<http://scalein.com/>)



Indice

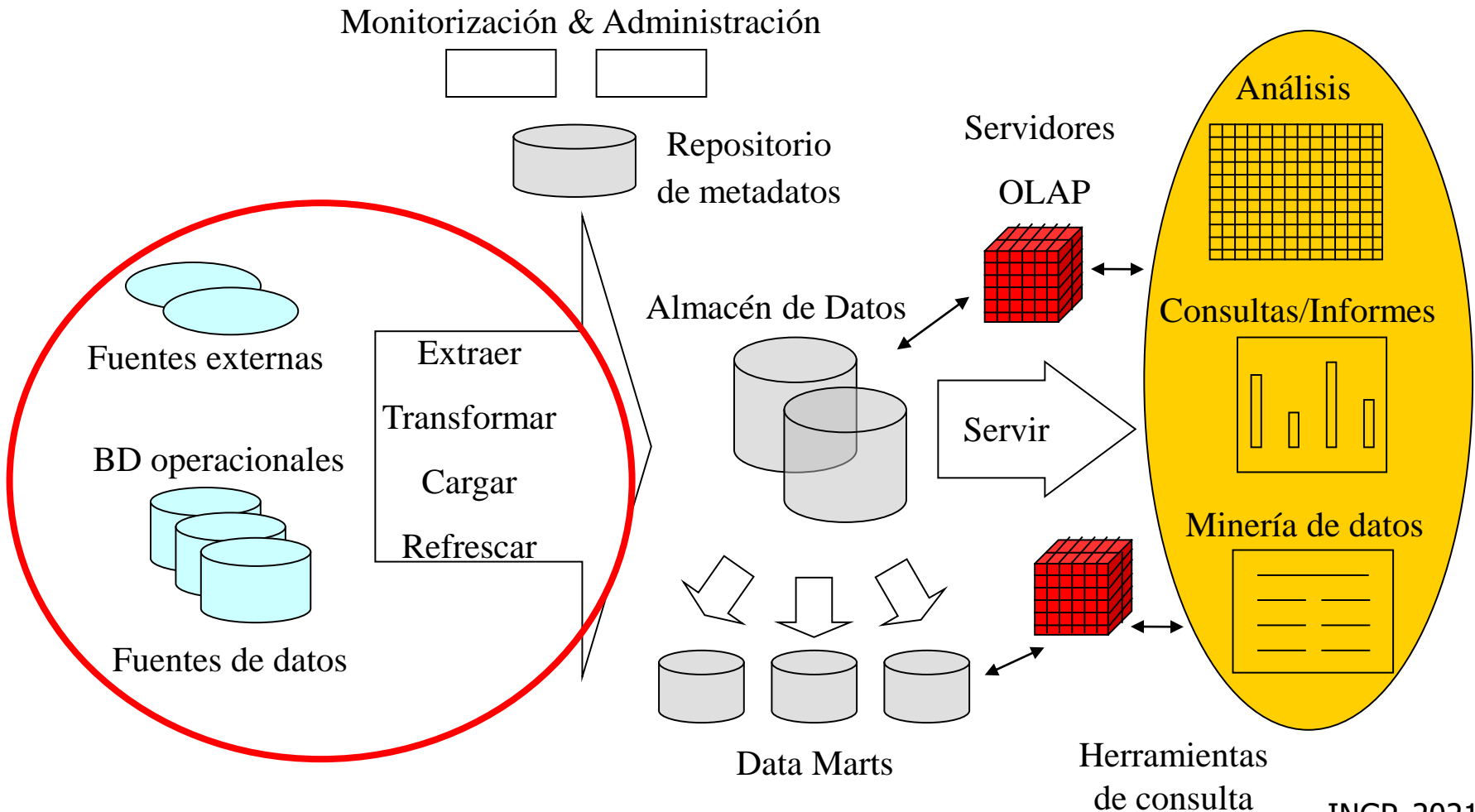
41

- Arquitecturas de DW
- Esquema tradicional de un DW
- Aplicaciones de DW en el mercado
- Aproximaciones de diseño

Esquema de una arquitectura de DW

Fuentes de datos

42



Esquema de una arquitectura de DW

Fuentes de datos

43

□ Fuentes de datos

- ▣ Fuentes de datos operacionales de la empresa
- ▣ Bases de datos externas (públicas o privadas)
- ▣ Ficheros planos
- ▣ Datos en formato tradicional:
 - documentos, facturas, albaranes, etc.
- ▣ Internet → Cantidad ingente de datos

Esquema de una arquitectura de DW

Fuentes de datos

44

- Procesos para alimentar de datos el almacén (ETL)
 - ▣ Extracción (Extraction)
 - ▣ Limpieza (Cleaning) y Transformación (Transformation)
 - ▣ Carga (Loading)
 - ▣ Refresco

- Extracción
 - ▣ Procesos que recogen los datos necesarios del almacén

Esquema de una arquitectura de DW

Fuentes de datos

45

□ Limpieza (Cleaning)

▣ Fundamental que los datos del almacén sean correctos

- Decisiones estratégicas

▣ Muchas fuentes de datos → alta probabilidad de error y anomalías

- Longitud de campos inconsistentes
- Descripción inconsistente (¿¿¿¿¿ Qué es dirección ????)
- Valores incoherentes (“U. Alicante” vs “UA”)
- Valores nulos

Esquema de una arquitectura de DW

Fuentes de datos

46

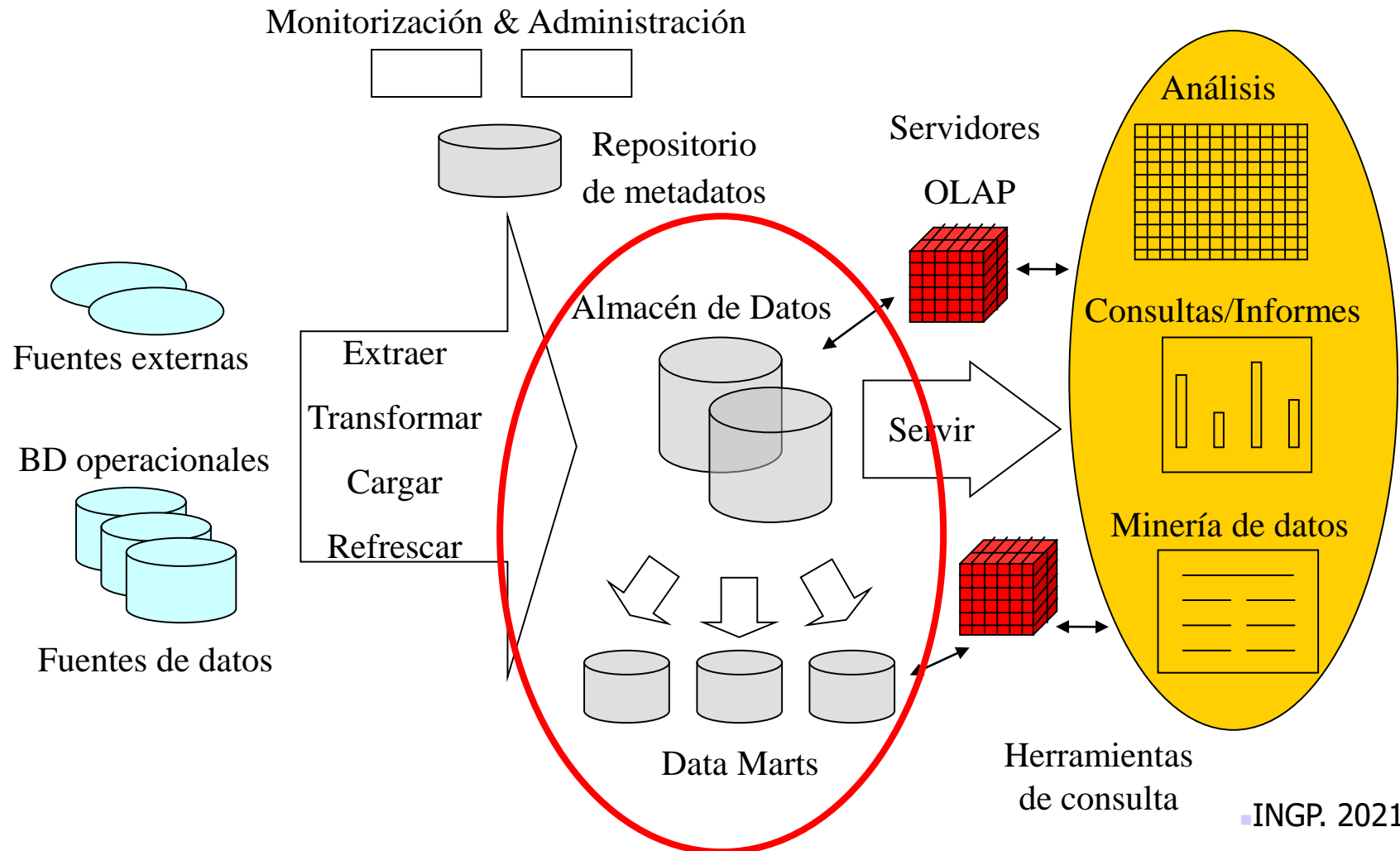
□ Carga (Loading)

- ▣ Una vez que los datos se extraen, limpian y transforman → CARGAR
- ▣ Se puede necesitar más pre-proceso antes de carga
 - Comprobar reglas de integridad de nuevo
 - Calcular datos agregados
 - Construir Tablas derivadas y virtuales e intermedias
 - Construir índices
 - Calcular tiempo → Muchas veces por la noche
 - ¿¿ Y si multinacional ??

Esquema de una arquitectura de DW

El almacén o repositorio

47



Esquema de una arquitectura de DW

El almacén o repositorio

48

□ Datos orientados por tema

▣ Los sistemas OLTP están optimizados para las transacciones

■ NORMALIZAR

- Muchas transacciones con pocos datos

- Ej. datos de clientes en varias tablas

▣ ¿Almacén de datos normalizado?: PROBLEMAS

- ¿Directivo es capaz de tener visión de todas las tablas y relaciones?

- Pocas transacciones que incluyen muchos datos

- Operación MAS cara en BD: unión de tablas

Esquema de una arquitectura de DW

El almacén o repositorio

49

- ▣ Los datos están orientados por tema
 - En un solo lugar (digamos tabla) datos referentes a un concepto que es el objeto de estudio
 - Ej. Tabla para clientes
 - Ventas
 - Compras
 - Vehículos, etc.

Esquema de una arquitectura de DW

El almacén o repositorio

50

□ Integrados

- ▣ Están coherentemente agrupados a partir de datos de las fuentes de datos
- ▣ También hay datos derivados
- ▣ Para ello: procesos de limpieza y transformación
- ▣ Hay errores difíciles de detectar: ¡¡¡¡ Cuidado !!!!!
 - Ej. Código producto válido → Exhaustivo análisis de datos

Esquema de una arquitectura de DW

El almacén o repositorio

51

□ Integrados,.....

▣ Problemas de incoherencia: resumir en 4 tipos

■ Descripción

- J. A. Rodríguez vs. Jose A. Rodriguez

■ Codificación

- Varón “V”, Hembra “H”; en otra BD Varón “H”, Hembra “M”

■ Unidades

- Estatura: 1,70 mts; 170 cm

■ Formato

- Número de teléfono como cadena de caracteres (965- 90 34 00) vs.como entero (965903400)

Esquema de una arquitectura de DW

El almacén o repositorio

52

□ No volátiles

- ▣ En sistemas OLTP se pueden modificar datos (ej. tuplas)
 - Unidades de pedido 200; si cliente modifica, se cambia.

- ▣ En DW nunca se modifican (salvo excepciones: ver dimensiones que cambian lentamente en tema IV), se añaden nuevos datos para el análisis
 - Un pedido con una fecha de 200 uds.
 - Un pedido con otra fecha de 150 uds.

Esquema de una arquitectura de DW

El almacén o repositorio

53

□ Variables en el tiempo

▣ No volatilidad → Dimensión básica: El TIEMPO

- Datos analizados en función del tiempo

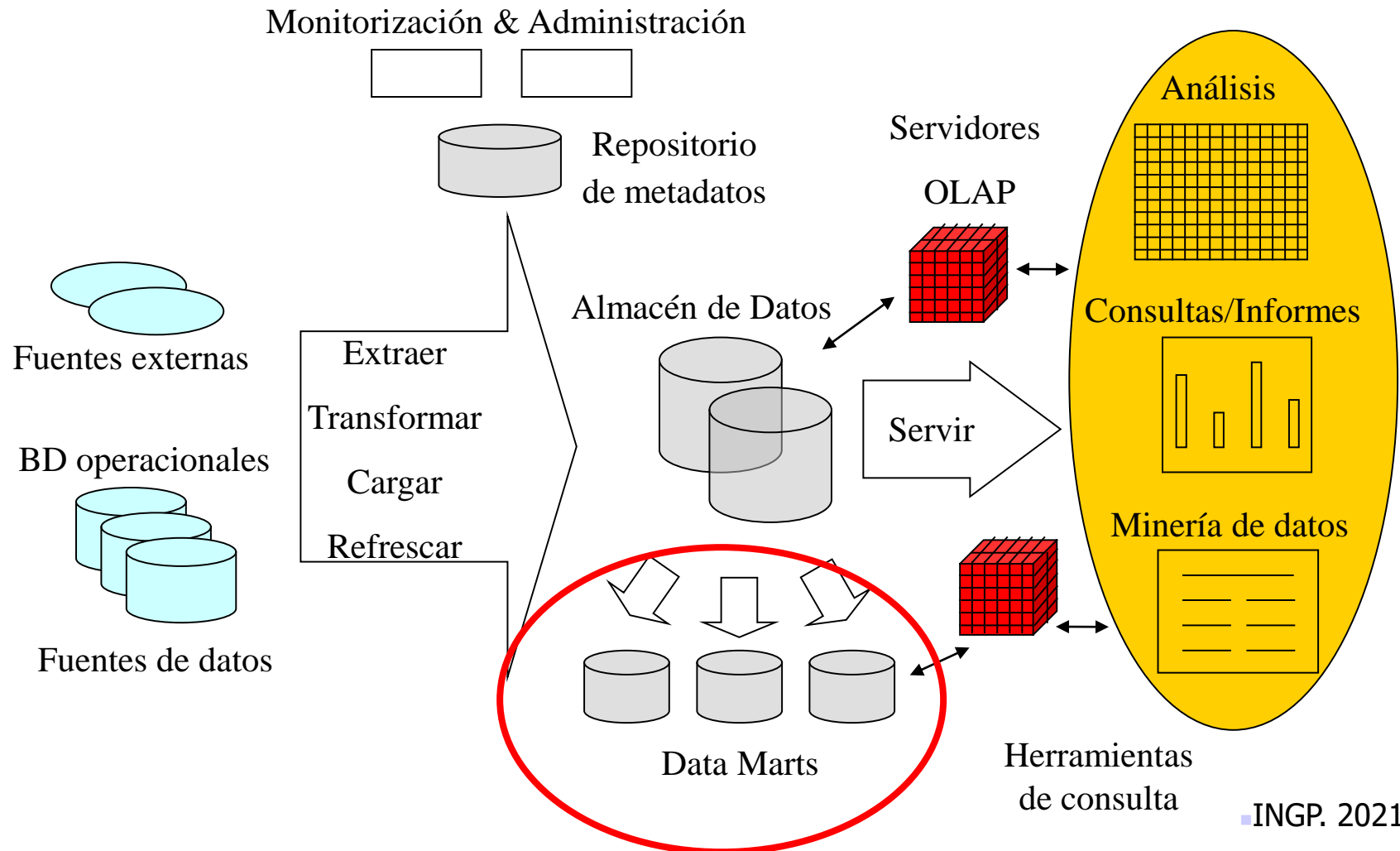
- Ej. Anterior del pedido

- ¿ Por qué un cliente ha variado la cantidad de su pedido en una semana ?
- ¿ Por qué han cambiado los gustos de un cliente en varios años ?
- Etc.

Esquema de una arquitectura de DW

El almacén o repositorio

54



Esquema de una arquitectura de DW

El almacén o repositorio

55

□ Data Marts

- Es como una vista del almacén de datos
- Se definen para satisfacer las necesidades de un departamento o sección dentro de una empresa
- Normalmente, en la práctica, suelen contener más cantidad de información agrupada que en detalle

Esquema de una arquitectura de DW

El almacén o repositorio

56

□ Data Marts.....

- Para su construcción se pueden seguir dos aproximaciones:
 - (I) Definir primero el almacén de datos y, a partir de él, definir los data marts, ó
 - (II) Definir primero los data marts departamentales y, posteriormente integrarlos en un almacén de datos global para la organización
- Nota: Si la envergadura de la empresa es considerable y, la experiencia en construir DW poca, es aconsejable seguir la aproximación 2

Esquema de una arquitectura de DW

El almacén o repositorio

57

□ Resumen: Data warehouse vs. Data Marts

▣ Enterprise DW

- Información sobre “temas” de toda organización
- Requiere complejo modelado de negocio
- Puede llevar AÑOS para construir e implementar

▣ Data Mart

- Departamental → sub-temas
 - Ej. Marketing data mart, Clientes, productos, ventas !!!!!
- Más rápido agregar
- OJO !!! Integración con DW puede ser compleja

Esquema de una arquitectura de DW

El almacén o repositorio

58

□ Virtual Data Warehouse

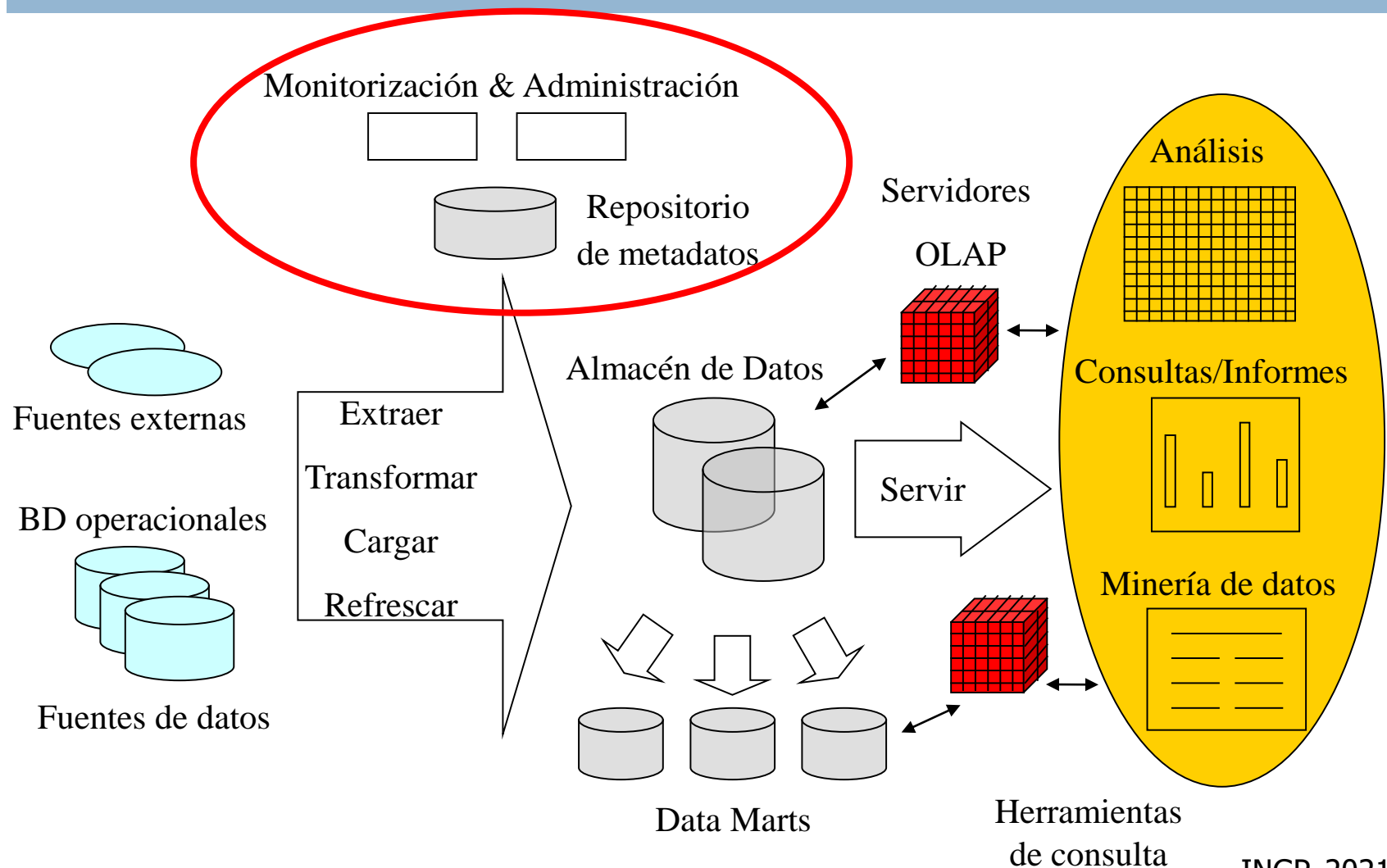
▣ Vistas sobre bases de datos operacionales

- Materializan algunas vistas agregadas → consultas eficientes
- Mayor facilidad en la construcción
- Requieren exceso de capacidad del servidor operacional
- Muchas empresas dicen tener DW cuando el administrador ha creado un VIRTUAL DW

Esquema de una arquitectura de DW

Los metadatos

59



Esquema de una arquitectura de DW

Los metadatos

60

□ Los metadatos

▣ Son datos sobre datos

- Qué dato se guarda (ej. clientes)
- Dónde se guarda (tabla clientes)
- Campos de la tabla
- Con qué datos de las fuentes se corresponden
- Niveles de agregación
- Procesos de carga → ¿ Cuándo se actualizan ?
 - ¿ Cuándo fue la última actualización?
- Patrón de dato válido (Ej. *Apellido1 Apellido2, Nombre*)

Esquema de una arquitectura de DW

Los metadatos

61

□ Los metadatos ...

▣ Son datos sobre datos ...

- Reglas de transformación
- ¿ Cuándo se incorporan al almacén de datos ?
- Y muchos más...

Esquema de una arquitectura de DW

Los metadatos

62

□ Tipos

- Administrative metadata (Toda la información necesaria para el DW)
 - Fuentes de datos y contenidos
 - Esquema del data warehouse, vistas y datos agregados
 - Dimensiones de análisis con sus jerarquías
 - Consultas e informes predefinidos
 - Localización y contenido de los Data Marts
 - Diseño físico → particionamiento de datos

Esquema de una arquitectura de DW

Los metadatos

63

□ Tipos...

▣ Business metadata

- Información y términos de negocio
- Políticas de posesión de datos
- Políticas de permiso de datos por usuarios (seguridad)

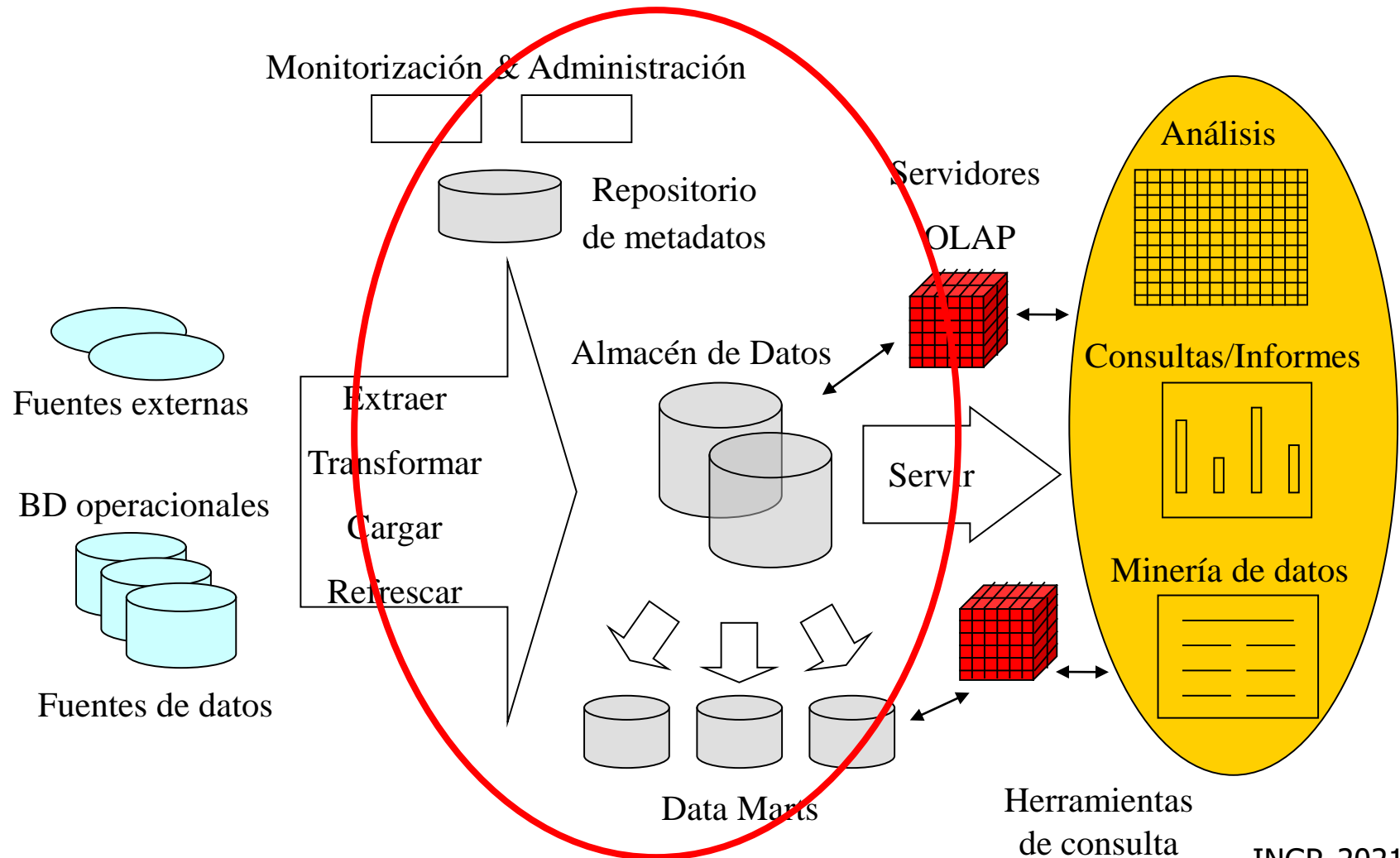
▣ Operational metadata

- Obtiene información recogida durante el proceso del “almacén de datos”
 - Datos migrados y secuencia de transformaciones aplicadas
 - Auditoría
 - Informes de error

Esquema de una arquitectura de DW

Servidor del almacén de datos

64



Esquema de una arquitectura de DW

Servidor del almacén de datos

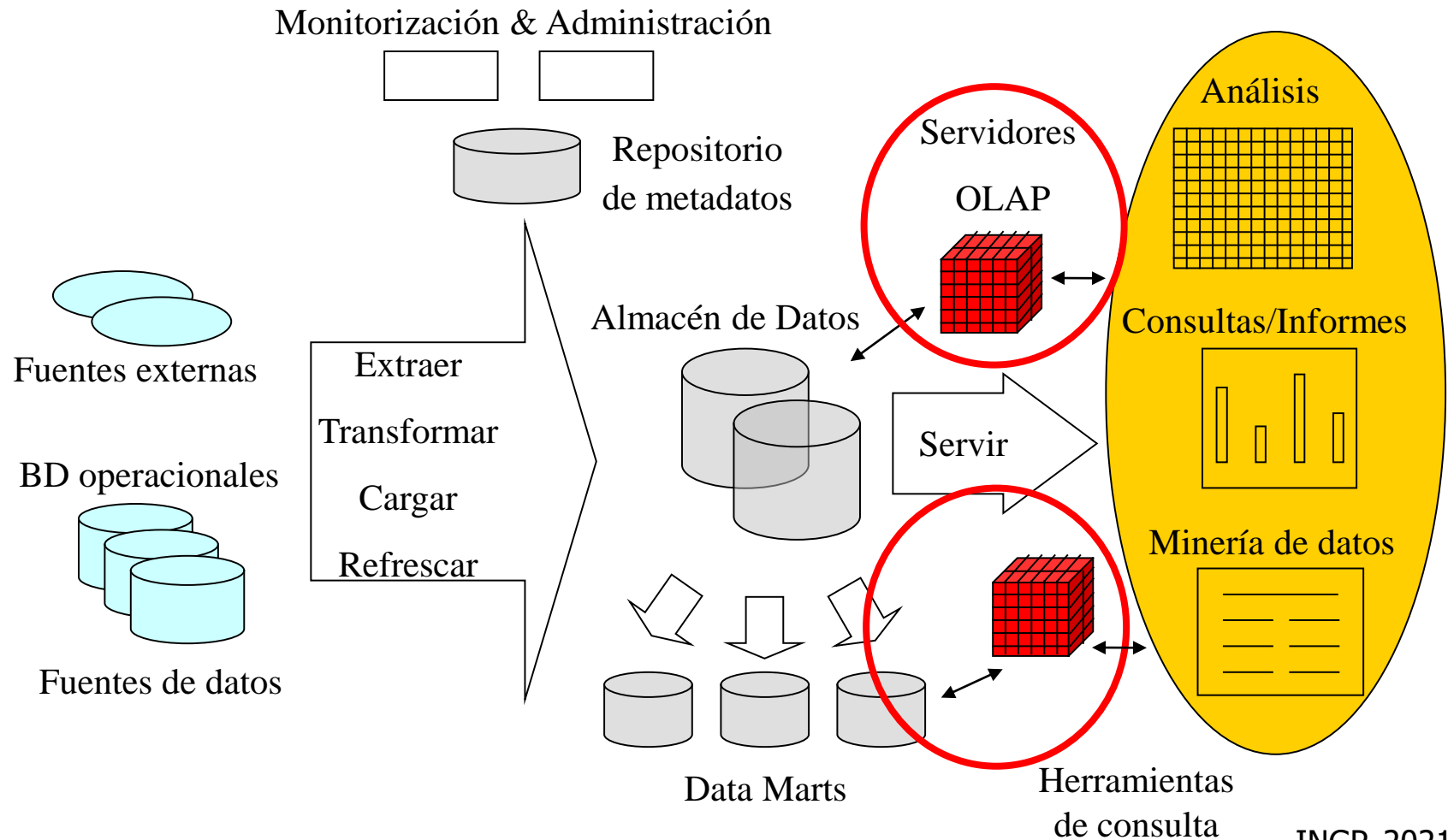
65

- El servidor es un SGBD que se encarga de
 - ▣ Gestionar el repositorio propio del almacén de datos
 - ▣ Coordinar los procesos ETL que alimentan el DW
 - ▣ Procesan las consultas lanzadas sobre el almacén y devuelven los datos
 - ▣ Generalmente son servidores relacionales

Esquema de una arquitectura de DW

Servidor del almacén de datos

66



Esquema de una arquitectura de DW

Servidor de consultas

67

□ El servidor de consultas

- ▣ En la mayoría de las arquitecturas se utiliza un servidor distinto al del almacén de datos
 - Rendimiento y mantenimiento
- ▣ La mayoría de herramientas funcionan con esta arquitectura
 - Ejemplo. MicroStrategy
- ▣ Dos tecnologías ampliamente utilizadas
 - ROLAP
 - MOLAP

Esquema de una arquitectura de DW

Servidor de consultas

68

□ Servidores ROLAP

- ▣ Utilizan tecnología Relacional (Relational OLAP)
- ▣ Utilizan extensiones del SQL estándar para soportar el acceso multidimensional a los datos (tema 5)
- ▣ Métodos de implementación adecuados para representar los datos multidimensionales en tecnología relacional (tema 5)
- ▣ Ventaja: Basado en un estándar

Esquema de una arquitectura de DW

Servidor de consultas

69

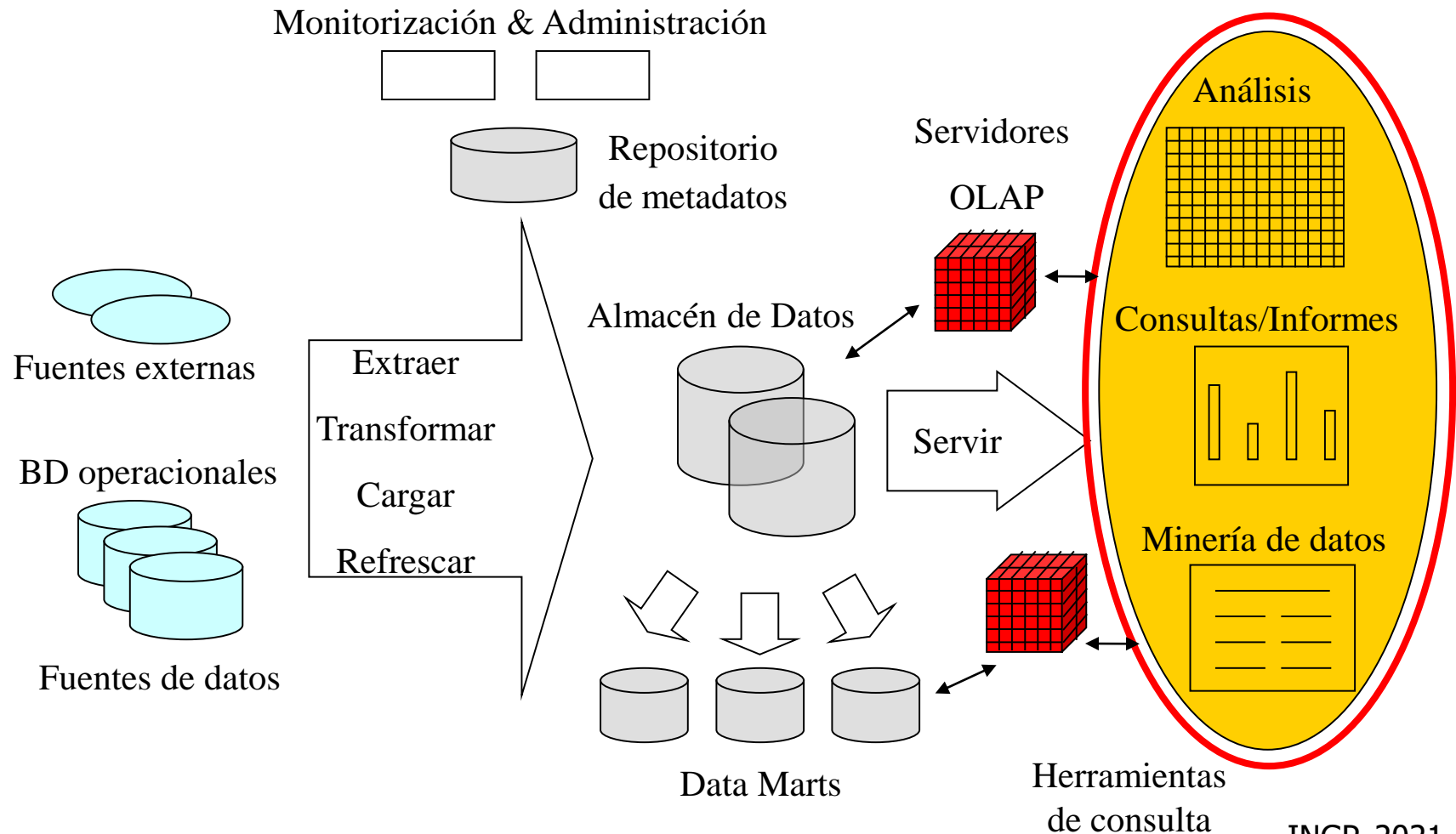
□ Servidores MOLAP

- ▣ Utilizan tecnología Multidimensional (Multidimensional OLAP)
- ▣ Los datos están almacenados directamente en matrices
- ▣ Operaciones de consulta están implementadas directamente sobre estas matrices
 - No están basados en SQL estándar
- ▣ Ventaja: Suelen ser más rápidos que los ROLAP
- ▣ Inconveniente: no basados en un estándar
- ▣ Algunos de los más famosos
 - Essabse (Arbor), Accumate (Kenan)

Esquema de una arquitectura de DW

Herramientas de consultas

70



Esquema de una arquitectura de DW

Herramientas de consultas

71

□ Generación de informes

▣ Consultas ad-hoc e informes

- Permiten acceso a los datos base
- Informe se construye con “point-and-click”

▣ Entornos consulta/informes de directivos

- Muestra datos base en términos de negocio
- Utiliza consultas predefinidas/almacenadas
- Soporte limitado para consultas ad-hoc
- Informes se pueden presentar como documentos

Esquema de una arquitectura de DW

Herramientas de consultas

72

□ OLAP Query Tools

- ▣ Objetivo: Dar soporte a consultas ad-hoc para el analista del negocio
 - Analistas del negocio conocen hojas de cálculo

- ▣ Extensión de hojas de cálculo para análisis del DW
 - Enriquecido semánticamente con términos del negocio
 - Análisis multidimensional de los datos
 - Hechos
 - Dimensiones

Esquema de una arquitectura de DW

Herramientas de consultas

73

- Data Mining (Minería de datos)
 - ▣ Descubre tendencias y patrones (minería interpretativa)
 - ▣ Crean modelos y hacen predicciones (minería predictiva)
 - Utilizan los datos existentes para detectar tendencias y crear modelos
 - Aplican los patrones y modelos a datos nuevos
 - Importante → Adaptación de modelos según nuevos datos

Indice

74

- Arquitecturas de DW
- Esquema tradicional de un DW
- Aplicaciones de DW en el mercado
- Aproximaciones de diseño

Esquema de una arquitectura de DW

Aplicaciones de DW en el mercado

75

- ▣ Vendedores y plataformas de IN
 - Oracle
 - Microsoft Analysis Services
 - SAP
 - SAS
 - Microstrategy
 - IBM Cognos
 - Pentaho (Community y Enterprise)
 - BIRST
 - Information Builders

Esquema de una arquitectura de DW

Aproximaciones de diseño

76

- Primera aproximación metodológica para el diseño de almacenes de datos
 - ▣ Arriba a abajo (Top-down)
 - Diseñar y modelar el DW en función de los requerimientos
 - Obtener los datos para “poblar” el DW a partir de las fuentes de datos operacionales encontradas
 - Diseñar los procesos ETL
 - Normalmente la más utilizada
 - Sólo aplicable en sistemas sencillos

Indice

77

- Arquitecturas de DW
- Esquema tradicional de un DW
- Aplicaciones de DW en el mercado
- Aproximaciones de diseño

Esquema de una arquitectura de DW

Aproximaciones de diseño

78

□ Primera aproximación metodológica para el diseño de almacenes de datos

▣ Abajo a arriba (bottom-up)

- Diseñar y modelar el almacén de datos en función de los datos que estén presentes en las fuentes de datos operacionales
- Diseñar los procesos ETL
- El usuario final ha de adaptarse al diseño en lugar de dirigir el diseño del DW

▣ Híbrida

- Combinación de las dos (Tema 2)

El almacén de datos

Bibliografía

79

- Juan Trujillo, Jose Norberto Mazón, Jesús Pardillo. *Diseño y explotación de almacenes de datos: Conceptos Básicos de Modelado Multidimensional*. Ed. Club Universitario. 2011.
 - ▣ Tema 1
- Giovinnazo (2000). *Object-Oriented Data Warehouse Design: Building a star schema*
 - ▣ Tema 1 (introducción)
- Inmon (2002). *Building the Data Warehouse* (3ª ed.)
 - ▣ Tema 1 (introducción)
- Kimball (2002). *The Data Warehouse Toolkit* (3ª ed.)
 - ▣ Tema 1 (introducción)
- Thomsen (2000). *OLAP solutions: Building Multidimensional Information Systems*
 - ▣ Tema 1 (introducción)

Introducción a la Inteligencia de Negocio

Tema 1

Profesores:

Juan C. Trujillo
LUCENTIA Research Group



Universitat d'Alacant
Universidad de Alicante



Departamento de
Lenguajes y Sistemas
Informáticos

INGP. 2021