



# Inteligencia de Negocios, Datawarehousing

Jose Aguilar  
CEMISID, Escuela de Sistemas  
Facultad de Ingeniería  
Universidad de Los Andes  
Mérida, Venezuela  
Becario Prometeo, UTPL Loja, Ecuador

# Ideas introductorias



## **Según Steve Haeckel**

“El cociente de inteligencia de una empresa está determinado por la medida en que su infraestructura informática conecta la información, la comparte y le da estructura.”

# CONOCIMIENTO

## ERA INDUSTRIAL



### VALORES PREDOMINANTES

- Poder
- Control
- Disciplina
- Especialización
- Estructura jerarquizada

## ERA DEL CONOCIMIENTO



### VALORES PREDOMINANTES

- Descentralización
- Información
- Innovación
- Calidad
- Trabajo en equipo

Los trabajadores trabajan más con sus mentes que con sus manos (Knowledge Worker)



**SOCIEDAD DEL USO DE CONOCIMIENTO**

## **Dato**

Unidad mínima de información, **sin sentido en sí misma**, pero que **adquiere significado en conjunción con otros elementos** similares o diferentes.

Conjunto de símbolos que unidos de cierta forma dan un significado lógico.

## **Conocimiento**

La definición de conocimiento de Daniel Bell (1973):

“Conocimiento: una serie de **afirmaciones organizadas de hechos o ideas que presentan un juicio razonado o un resultado experimental**, que se transmite a los demás mediante algún medio de comunicación en alguna forma sistemática. Por lo tanto, distingo conocimiento de noticias y entretenimiento”.

## **Información**

Se usa en muchos sentidos diferentes. El DRAE ofrece varias definiciones entre las que podemos destacar “Acción y efecto de informar”, “Comunicación o adquisición de conocimientos” y “Conocimientos así comunicados o adquiridos”.

En otras palabras, **la información puede ser tanto la cosa que se comunica, como el acto de comunicar, como el resultado de la comunicación.**

# Embudo del Conocimiento



**“EL CONOCIMIENTO ES EL ACTIVO  
INTANGIBLE QUE MAYOR  
COMPETITIVIDAD GENERA A LAS  
NACIONES Y A LAS  
ORGANIZACIONES, EN LA  
ECONOMÍA GLOBAL”**

**“DEBE SER GERENCIADO”**

# La sociedad del Conocimiento



- **Ausencia de fronteras**, porque el conocimiento viaja aun con menos esfuerzo que el dinero
- **Movilidad ascendente**, disponible para todos en virtud de educación formal fácil de adquirir.
- La mayoría de los empleados **no son de tiempo completo en la organización ni están físicamente**
- Nacimiento de **nuevas formas de relación social**

**Nativos digitales vs. Migrantes digitales**

# Diferentes tipos de conocimiento

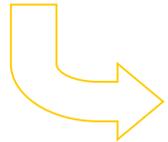


# Taxonomías del conocimiento

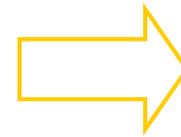
<b>TIPOS DE CONOCIMIENTO</b>	<b>DEFINICIONES</b>	<b>EJEMPLOS</b>
<b>Declarativo</b>	Saber - sobre	Qué fármaco es el apropiado para una enfermedad
<b>Procedimental</b>	Saber - cómo	Cómo administrar un fármaco en concreto
<b>Causal</b>	Saber - porqué	Entender porqué funciona o da resultado el fármaco
<b>Condicional</b>	Saber - cuándo	Saber cuando prescribir el fármaco
<b>Relacional</b>	Saber – con quién	Comprender de qué manera el fármaco interactúa con otros medicamentos
<b>Pragmático</b>	Conocimiento útil para una organización.	Conocimiento sobre clientes, productos, procesos y competidores; los cuales pueden incluir: las mejores prácticas, diseños de ingeniería, informes de mercado, esquemas de negocio, procesos de negocio, tecnología, etc...

# CONOCIMIENTOS

## EXPLÍCITO



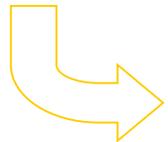
Base objetiva,  
racional (mente)



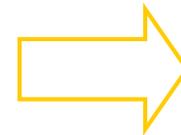
**Teoría**

**EXPLÍCITO:** Referencial, replicable, entrenable, reposa en textos, bases de datos, objetos, redes, como información.

## TÁCITO



Base subjetiva,  
experiencial (cuerpo)



**Práctica**

**TÁCITO:** Vivencial, basado en valores, creencias, actitudes, sentimientos humanos, no es fácil de replicar.

# Comparación entre las propiedades del conocimiento tácito y el explícito

Propiedades del conocimiento Tácito	Propiedades del Conociendo Explícito
Habilidad para adaptarse, y para lidiar con situaciones nuevas y excepcionales	Habilidad para diseminar, reproducir, acceder, y reaplicar por toda la organización
Pericia, el <i>saber-como</i> , el <i>saber-porque</i> , y el la necesidad por entender el <i>porque</i>	Habilidad para enseñar y para entrenar
Habilidad para colaborar, compartir una visión y transmitir una cultura	Habilidad para organizar, sistematizar; para convertir una visión en una misión y una misión en directrices operacionales
Preparar y mentoriar la transferencia de conocimiento de experiencia a nivel personal	Transferir el conocimiento a través de productos, servicios y documentos procesados.

# Organización: FÁBRICA DE EXPERIENCIA

## Conocimiento documentado

- Artículos
- Propuestas
- Presentaciones
- Políticas de la organización

## Conocimiento explícito no documentado

- Información de proyectos
- Información de clientes
- Políticas de la organización
- Experiencia de los empleados
- Procedimientos para las reuniones
- Procedimientos para la contratación
- Mejores prácticas para propuestas

## Conocimiento tácito no documentado

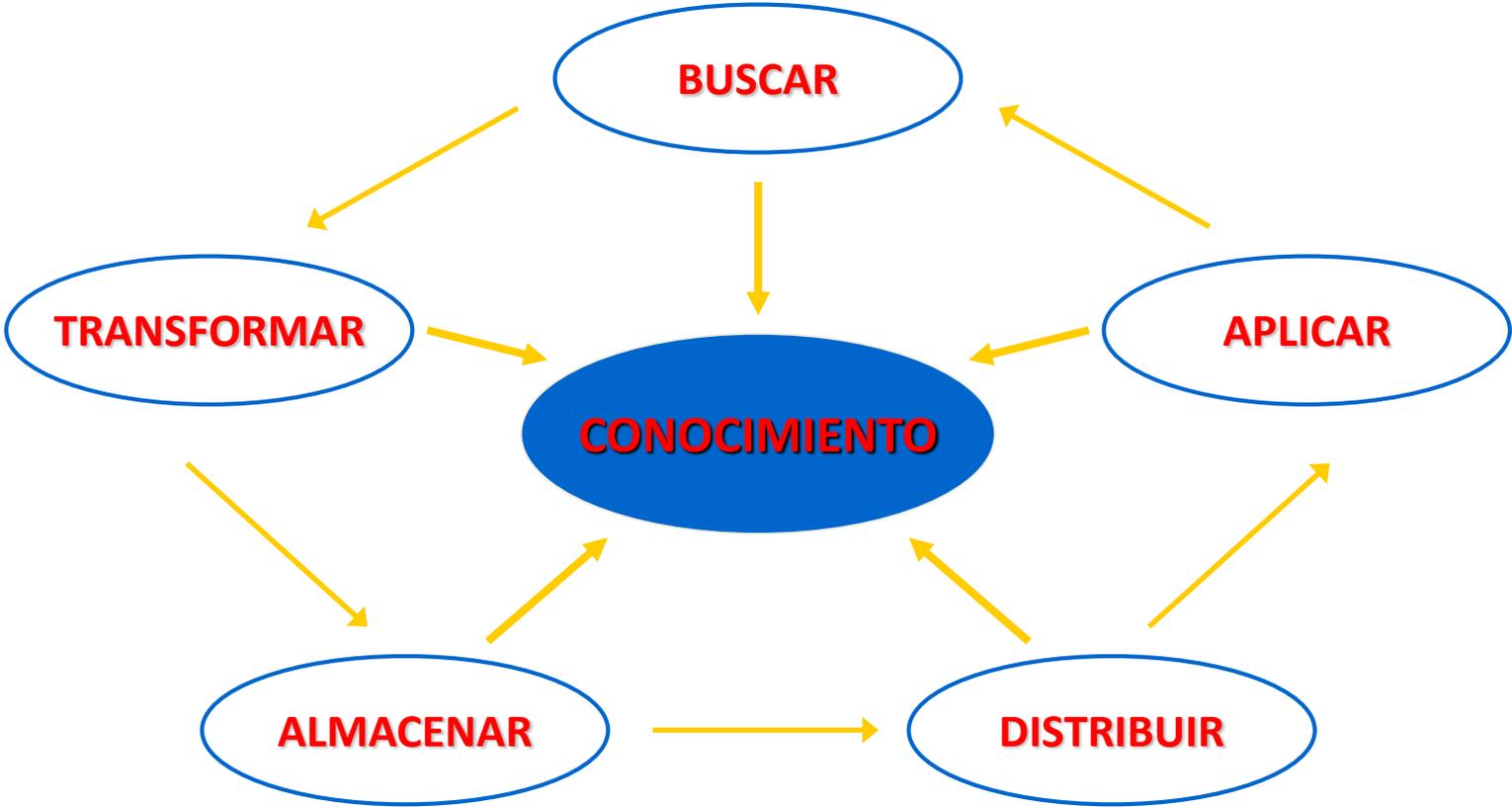
- Mejores prácticas para proyectos
- Mejores prácticas para atención a clientes

***Fácil***

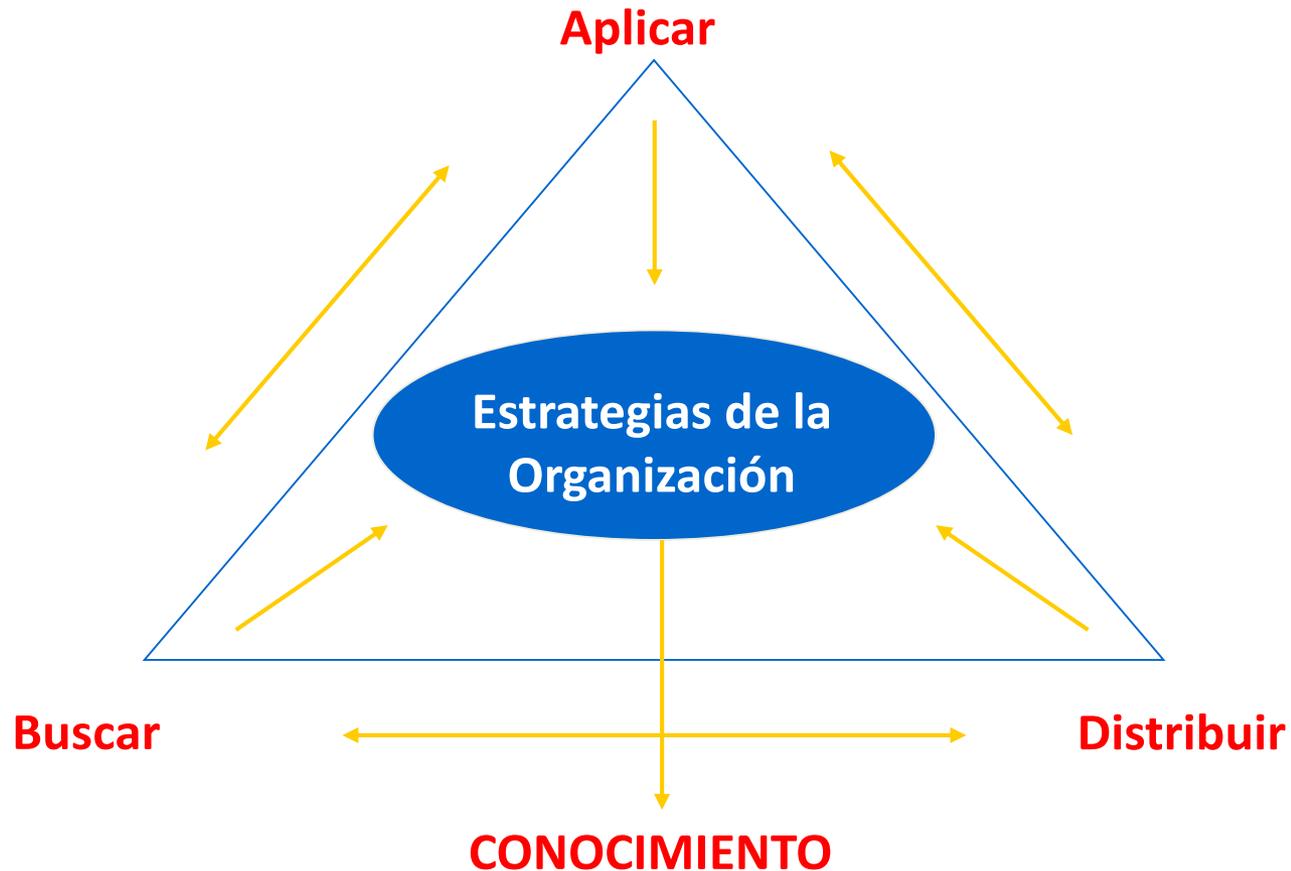
***Posible***

***Difícil***

# CONOCIMIENTO ORGANIZACIONAL



# Estrategia y Conocimiento



# Gestión del Conocimiento

La gestión del conocimiento es **crear la posibilidad para todo el mundo de entrar y ver exactamente qué está pasando, ver cuáles son las cuentas, dónde vamos bien, dónde no vamos tan bien, con una información numérica o gráfica, todo el mundo debe tener la posibilidad de entrar y ver exactamente qué está pasando y comunicarse con otras personas para mejorar todas estas cosas.**

**Bill Gates, presidente de Microsoft**

La gestión del conocimiento es la capacidad de la organización para **crear nuevos conocimientos, diseminarlos y encapsularlos en productos, servicios y sistemas.**

**Ángel L. Arbonies, presidente de Cluster del Conocimiento**

# GESTION DEL CONOCIMIENTO

Desde el punto de vista de las Organizaciones, se define el conocimiento como la información que posee valor para ella, es decir, aquella información que permite generar acciones asociadas a apoyar las nuevas oportunidades a través de la explotación de las competencias centrales de la Organización.

**Determinación de la información que la organización requiere para satisfacer sus objetivos.**

**Clasificación básica de la información de una organización:**

- Según la fuente de origen.
- Según el grado de formalización.
- Según el soporte.

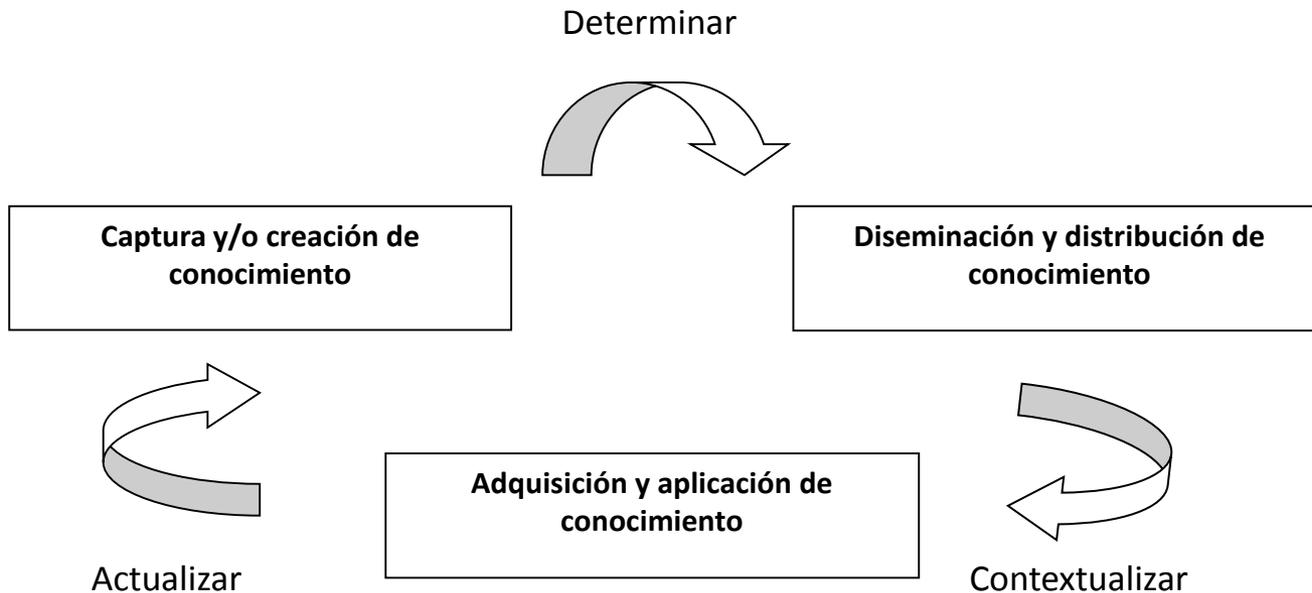
## **¿Qué es la Gestión del Conocimiento?**

Se define como el conjunto de procesos, estructuras organizativas, aplicaciones y tecnologías, a través de las cuales una empresa recoge, ordena, analiza, comparte y difunde, de manera consciente, su conocimiento entre el mayor número de empleados para aprovecharlo en beneficio de su organización.

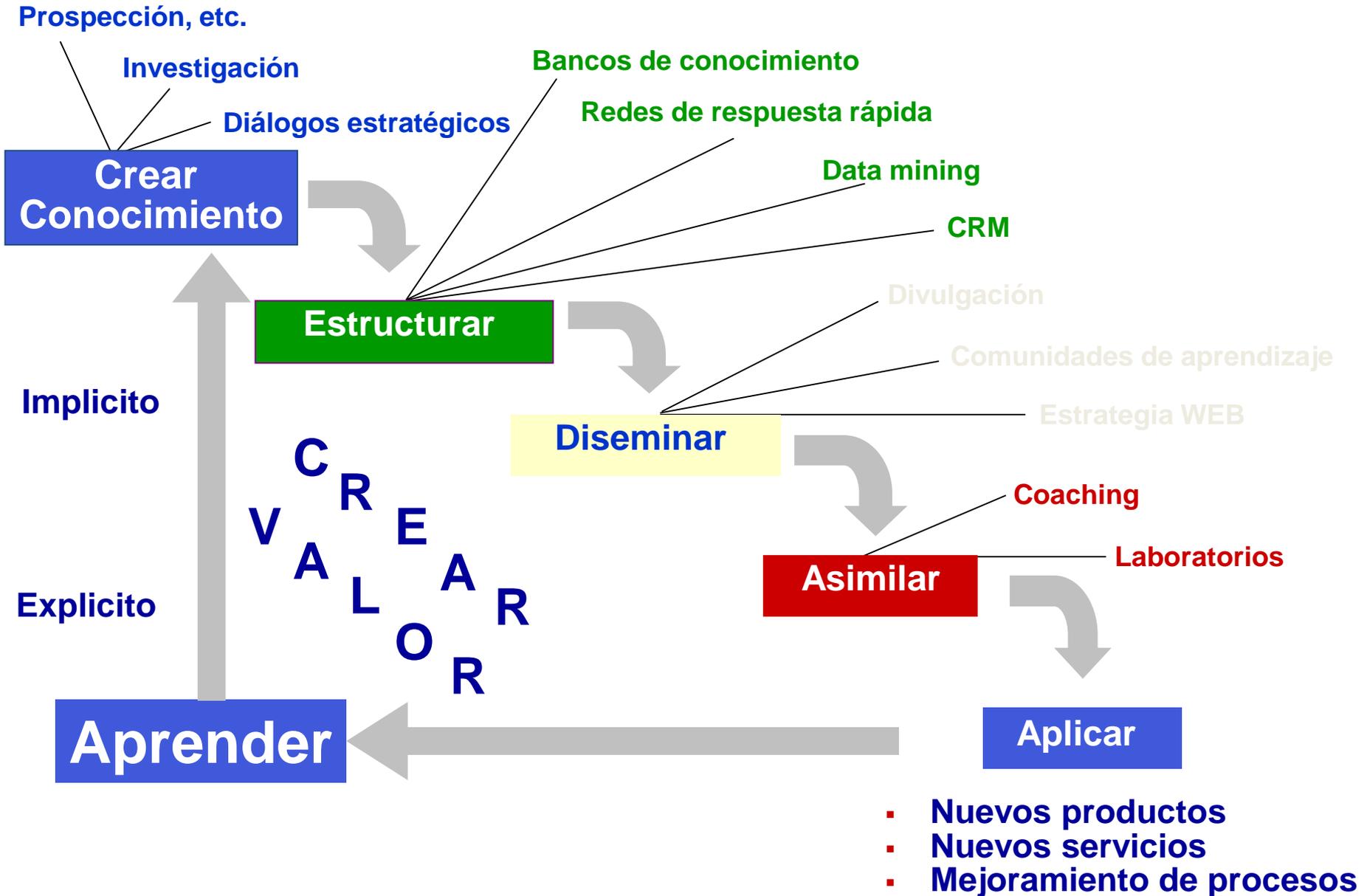
*Estamos ahogándonos de información, pero famélicos de conocimiento. John Naisbitt – Filósofo*

# Ciclo de la Gestión del Conocimiento

1. Captura y/o creación de conocimiento
2. Diseminación y distribución de conocimiento
3. Adquisición y aplicación de conocimiento



# Modelo de Gerencia del Conocimiento Skandia



# Modelo de creación de conocimiento de Nonaka y Takeuchi (1995)

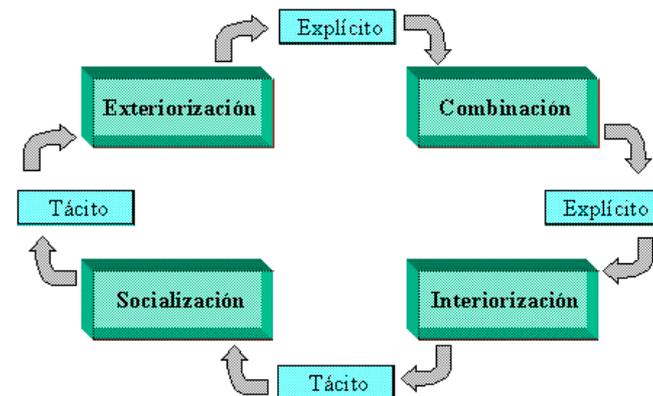
- La interacción entre estos dos tipos conocimiento, se llama **Conversión de conocimiento** y esto originó las siguientes cuatro formas:
  - **Socialización** (de Tácito a Tácito),
  - **Exteriorización** (de Tácito a Explícito),
  - **Combinación** (de Explícito a Explícito), e
  - **Interiorización** (de Explícito a Tácito).
- **La interacción de conocimiento explícito y tácito** tiene a darse sobre todo en *el nivel individual* y son unos cuantos individuos que desempeñan los cargos o el papel mas importantes.

# Modelo Espiral de Nonaka y Takeuchi



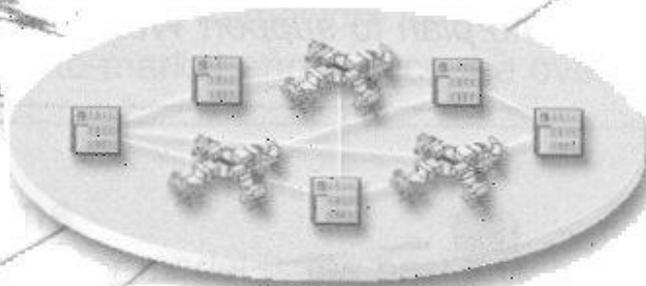
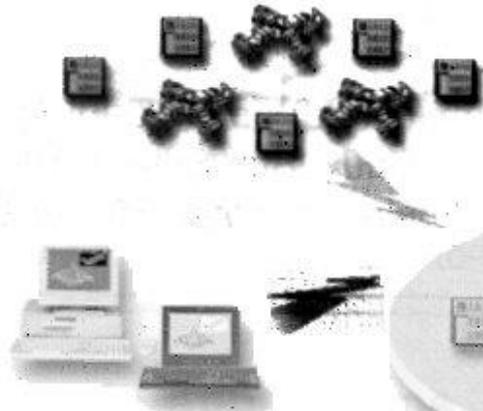
Los **4 modos de conversión** de conocimiento entre el *Conocimiento Tácito* (CT) y el *Conocimiento Explícito* (CE)

1. de CT a CT = Socialización
2. de CT a CE = Externalización
3. de CE a CE = Combinación
4. de CE a CT = Internalización



# SISTEMA DE INFORMACION MODERNO

**World Wide Web**



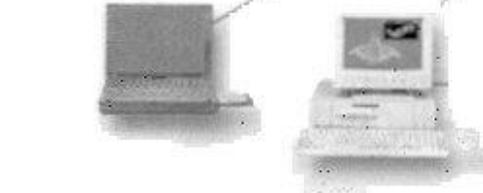
**Databases**



**Transactions**



**Web Application Server**



**Clients**

**Business Services**

**External Services**

# SISTEMA NERVIOSO DIGITAL

## Los sistemas de gestión Empresarial

**Valor de negocio**



Proporciona a sus usuarios una **profunda comprensión** y una capacidad para aprender que no podrían conseguir por otros medios



procesamiento analítico en línea

**análisis y la divulgación eficiente de datos**

# Internet - Extranet



## Era del conocimiento



Aprendizaje

Personalizar el sitio Web para cada Visitante



- Software va **adaptando dinámicamente** el site sobre la marcha de la sesión.
- Hemos pasado de un **monologo** a un **dialogo** y este a un **foro** en la Web



# Ideas introductorias

## Resemantizar a las Organizaciones

El flujo de información Digital, hace que tanto las organizaciones como los individuos, manejan instrumentos y **procesos digitales**,



**las organizaciones se deben redefinir**

# Resemantizar a las Organizaciones



Sin fronteras organizacionales

Transferir parte de los costos a los proveedores Y Clientes

Fidelidad de los clientes



Identificar las oportunidades de negocios

Estrategias de aplicación de Tecnología de Información

Análisis de la Cadena de Valor

# Tecnología Digital



El documento inteligente  
(realidad aumentada)



Los ambientes inteligentes

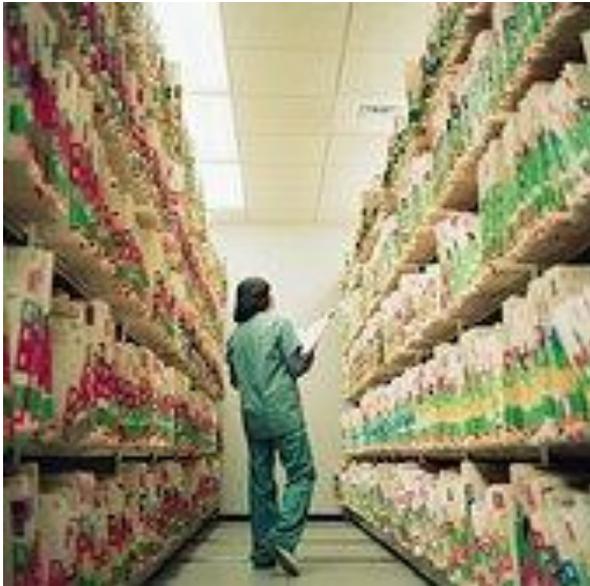


El cliente universal

# Tecnología Digital



El documento inteligente



*Integrar los documentos inteligentes a las aplicaciones empresariales, combinar el poder de la Digitalización y la lógica de negocios*

*El éxito radica en entender y administrar a fondo las interacciones entre los datos, las aplicaciones patentadas, los colaboradores internos y externos, los sistemas diversos, y los recursos de proveedores múltiples*

**realidad aumentada**

**dispositivos que añaden información virtual a la información física ya existente**

# Tecnología Digital



## Los ambientes inteligentes



Los **ambientes inteligentes** describen y manejan entornos físicos en los cuales las TICs así como los sensores, pasan mayoritariamente desapercibidos para los usuarios, puesto que se hallan discretamente integrados a objetos físicos, a infraestructuras, y al entorno cotidiano en el cual vivimos, viajamos, y trabajamos.

El objetivo de Aml es permitir que los dispositivos participen en actividades en los que nunca antes habían sido usados, posibilitando a la gente (a los usuarios) interactuar con los distintos dispositivos vía gestos, voz, movimientos,

**Además, los dispositivos pueden ser conscientes del contexto!!!**

# Administración Inteligente

No se puede administrar lo que no se puede ver.

La capacidad de **ver**  
**toda la empresa** es el aspecto más importante de la administración inteligente.

Los sistemas tienen que ser flexibles y adaptables a cambios en el proceso y modelo de los negocios, así como también a nuevas tecnologías.



# Organización y Tecnología



la **integración de personas, procesos e información**, en todos lados, en cualquier momento y desde cualquier dispositivo.

**Integración:**

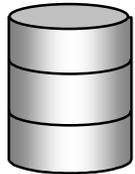
**separación de la capa de aplicaciones de negocios del entorno físico subyacente** compuesto de redes, servidores y sistemas de almacenamiento, de modo que las aplicaciones puedan instalarse con mayor flexibilidad

**Virtualización**

**automatización del entorno operativo** de acuerdo con las políticas y objetivos de negocios definidos

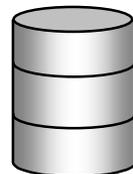
**Automatización:**

# Metodología Habitual de uso de Datos



## Sistemas Transaccionales

- Sistemas orientados a resolver los problemas de la operación diaria.
- Áreas de Sistemas Saturadas por las necesidades operacionales del día a día y requerimientos para Análisis.



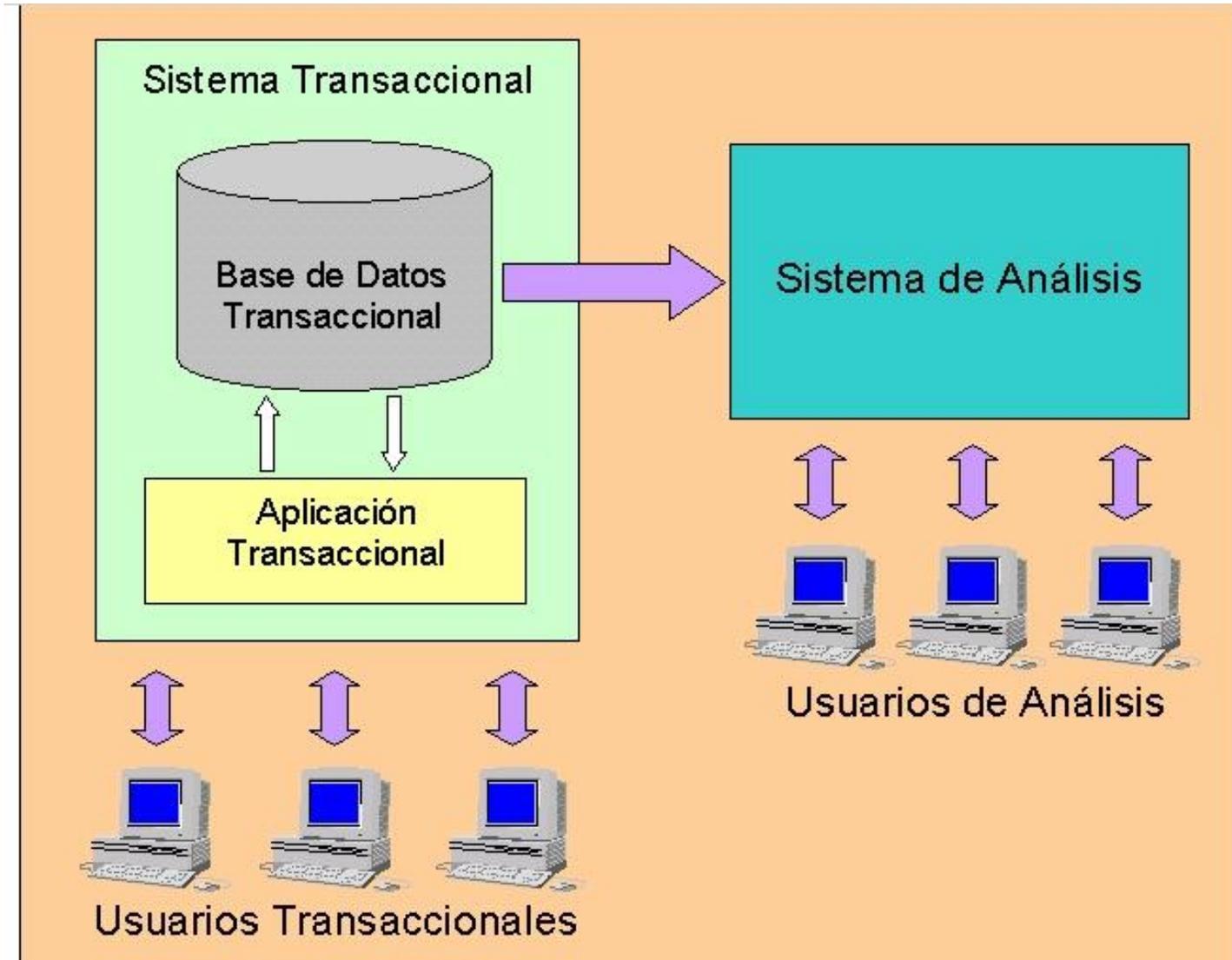
## Diversos Repositorios

- Grandes esfuerzos de recopilación, transcripción y formateo de información.
- Largos plazos de obtención.
- Gran margen de error.

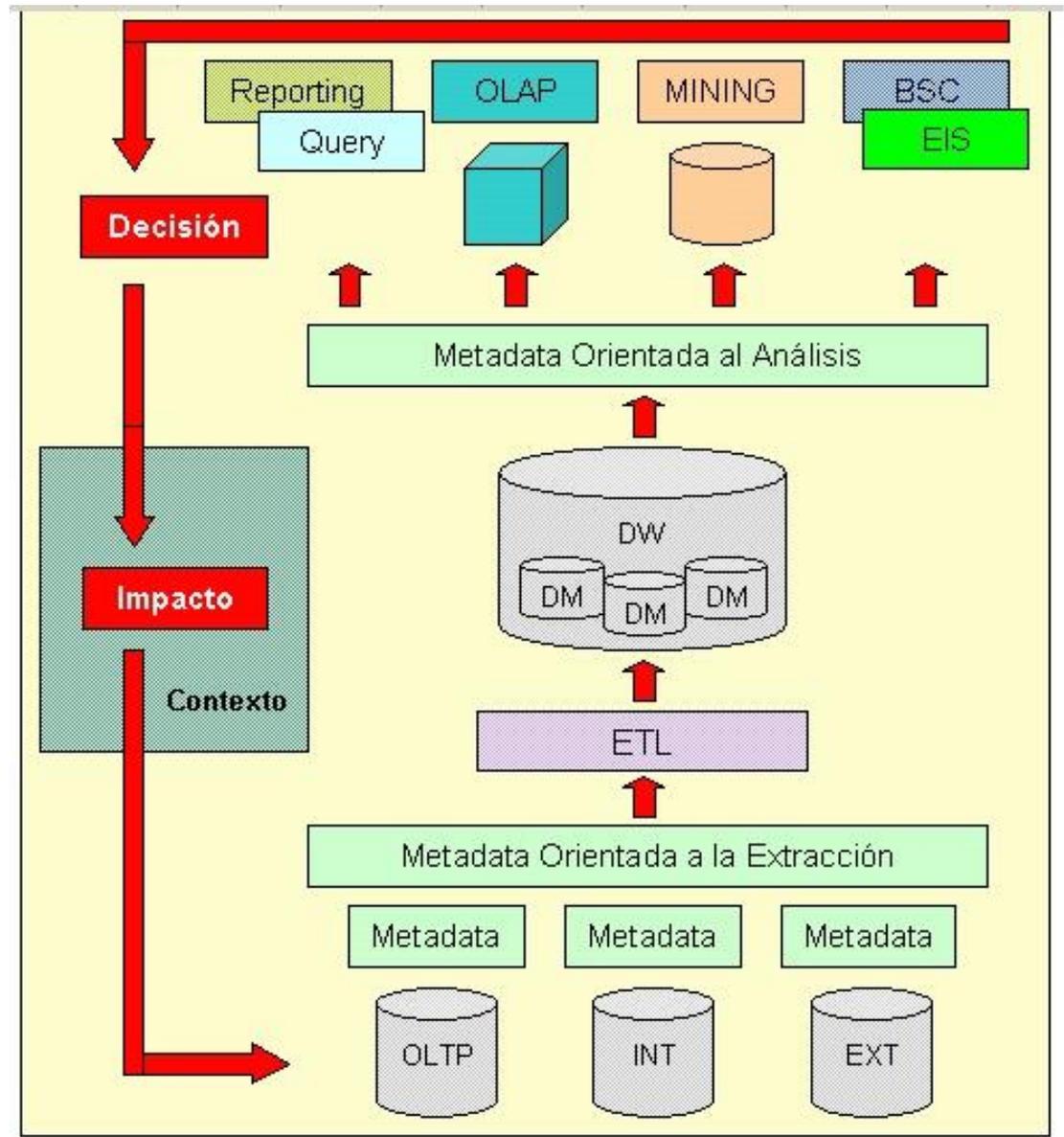
## • Información:

- Poco Oportuna
- Poco Amistosa
- Voluminosa
- Poco Relevante
- Sin Focalizar
- Poco Confiable
- Diferente entre Áreas
- Sin Cobertura Completa de Factores Críticos

# Sistemas de Análisis de Información



# Solución Inteligencia de Negocios



# Diferencias en el Diseño (OLTP vs BI)

## Sistemas Transaccionales (OLTP)

## Solución de Inteligencia de Negocios (BI)

Automatizar el proceso

Soportar la toma de decisiones

Diseñado para ser eficiente: tiempos de respuesta

Diseñado para ser efectivo:  
**información deseada**

Modela al negocio: cambia solo si lo hace el negocio

Se adapta el negocio: **para responder nuevas preguntas**

Reacciona a eventos

Se anticipa a eventos

Optimizado para transacciones

Optimizado para consultas

**Autonómico!!**

# Herramientas Open Source



- **Soluciones completas**

Pentaho, JasperReports, SpagoBI, BIRT

- **Herramientas ETL**

Clover , Enhydra Octopus

- **Desarrollos OLAP**

Mondrian, JPivot

- **Dashboards**

JetSpeed, JBoss Portal

- **Bases de Datos**

MySQL , Postgre, Greenplum

# Dato, información, conocimiento e inteligencia en la empresa

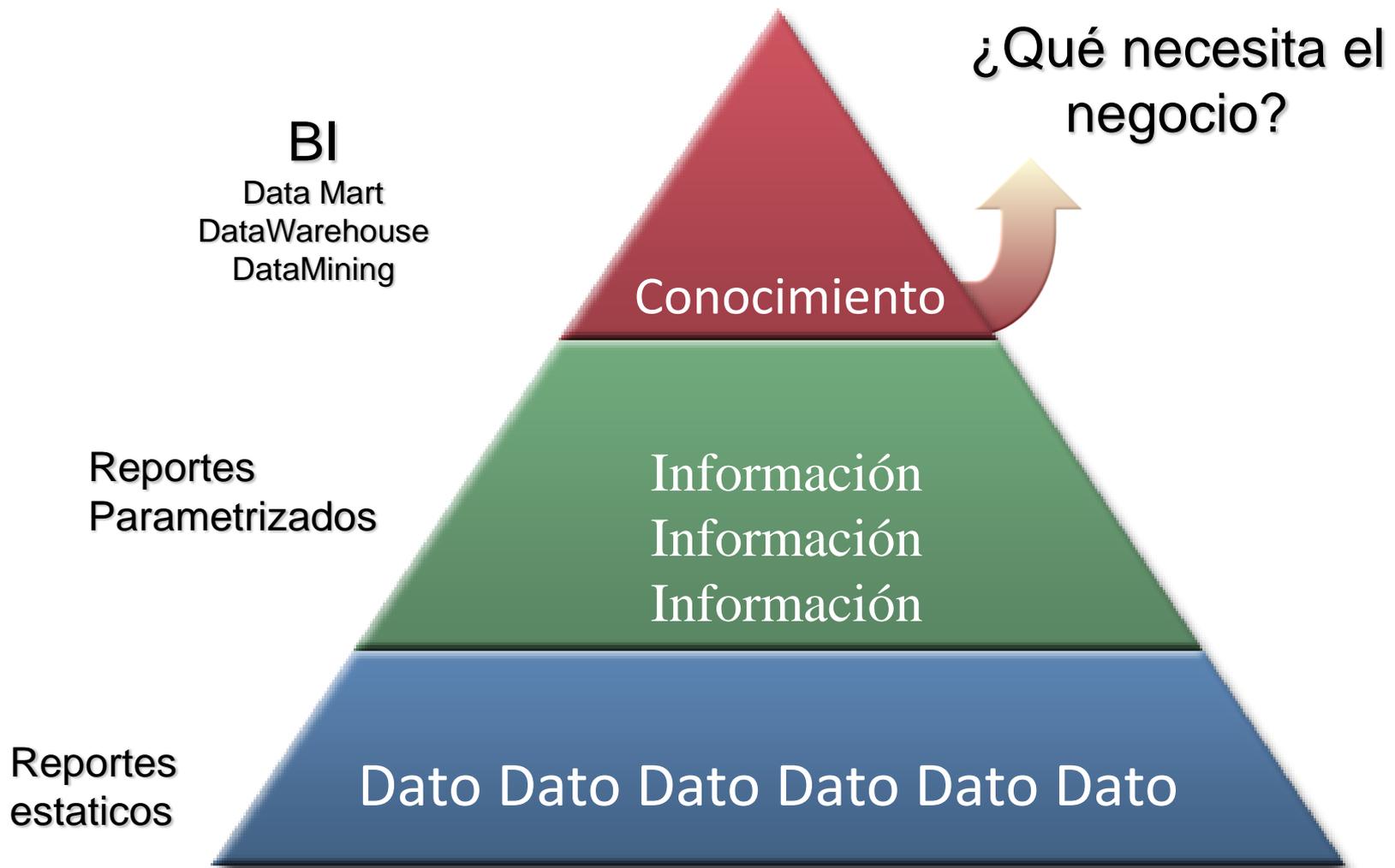


***Cadena de Valor de la Información***

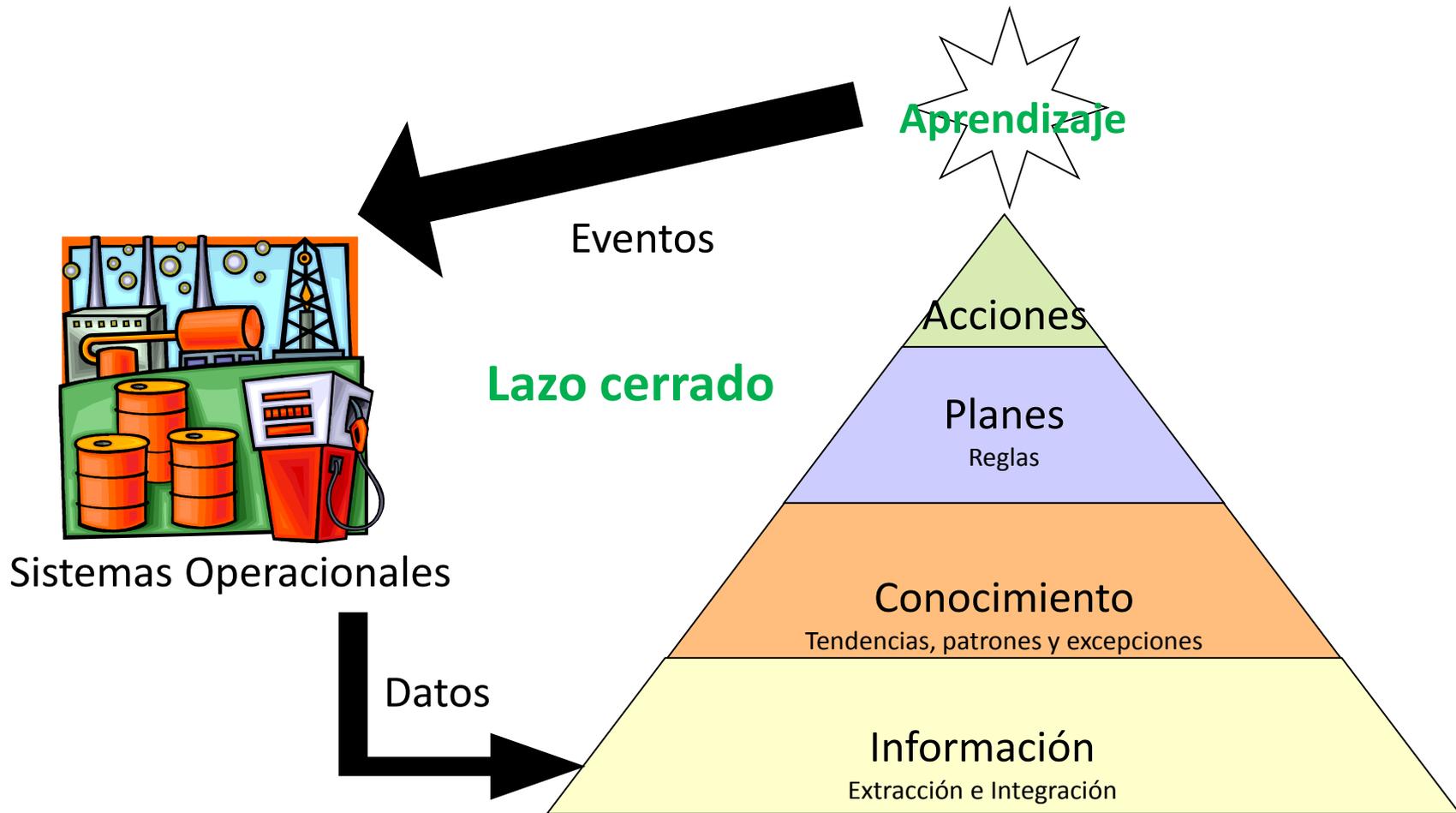
# Dato, información, conocimiento e inteligencia en la empresa



## Evolución hacia BI



# Dato, información, conocimiento e inteligencia en la empresa



# Las 10 Tecnologías tops en el 2011

1. Computación en la Nube
2. Virtualización
3. Tecnologías móviles
4. Gestión IT
5. **Business Intelligence**
6. Comunicación de datos, voz, multimedia
7. Aplicaciones empresariales
8. Tecnologías Colaborativas
9. Infraestructura
10. Web 2.0

Gartner's 2011 CIO Agenda (aka "[Reimagining IT: The 2011 CIO Agenda](#)").

# Inteligencia de Negocios (Business Intelligence - BI)



Concepto que integra por un lado el **almacenamiento** y por el otro el **procesamiento de grandes cantidades de datos**, con el principal objetivo de **transformarlos en conocimiento** para tomar decisiones, incluso en tiempo real,

- Requiere de procesos de análisis y exploración.
- El conocimiento debe ser oportuno, relevante, útil y debe estar adaptado al contexto de la organización.

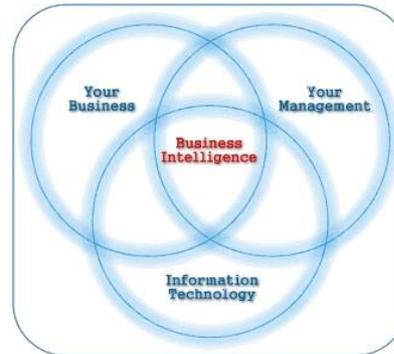
**Datos + Análisis = Conocimiento**

“Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones”

# ¿Qué es Inteligencia de Negocios?



Conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización.



Abarca la **comprensión** del funcionamiento actual de la empresa, y la **anticipación** de acontecimientos futuros, con el objetivo de **ofrecer conocimientos** para respaldar las decisiones empresariales.

# La necesidad de BI...



**Quienes necesitan un ambiente de Inteligencia de Negocios, poseen las siguientes características:**

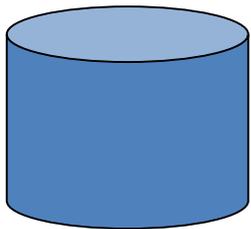
- **Los reportes provenientes de varios sistemas transaccionales, no concuerdan**
  - Los resultados financieros no concuerdan.
  - Las cantidades de inventario tampoco concuerdan.
  - Los reportes detallados no concuerdan con los reportes consolidados.
  
- **La gerencia no tiene acceso a una “imagen global corporativa” de su situación actual:**
  - ¿Cómo están nuestras finanzas?
  - ¿Quiénes son nuestros clientes?
  - ¿Qué nos han comprado?
  - ¿Cuánto inventario tenemos disponible?

# ¿Qué es Inteligencia de Negocios?



Hace a las organizaciones más inteligentes

Agregar Datos



Database, Data Mart, Data Warehouse, ETL Tools,



Presentar Datos



Dashboards,, OLAP Cubes



Enriquecer Datos



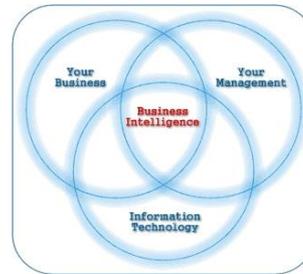
Tomar e Informar Decisiones



# ¿Qué es Inteligencia de Negocios?



Mediante ciertas herramientas y técnicas (extraer, cargar y transformar), se extraen los datos de distintas fuentes, se depuran y se preparan para luego cargarlos en un **almacén de datos**



Las herramientas inteligentes **genera información y conocimiento** relacionada con la empresa y de su entorno.

# ¿Por qué tener BI?



Responder rápidamente a los  
Retos de un entorno  
cambiante

Conocer los Clientes

**Inteligencia de Negocios:**  
**APLICACIÓN ESTRATEGICA**

Compartir la Información  
entre distintos niveles de la  
Organización

Analizar la Información para  
identificar Factores  
Críticos de la  
Organización

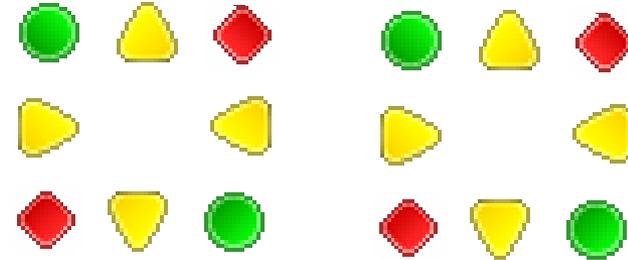
# Problemas a resolver con IN



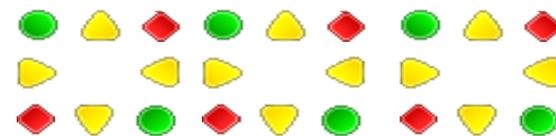
BD: Múltiples versiones no administradas de la verdad.

FROM	TO	Account Name	State	City	Postcode	Phone	Fax	Email	Website	Comments
Adam Bateson	Andrew Rogerson	Opusnet SA	Australia	Technology Park	5006	8192005 8137 SA				Third Avenue
Adam Bateson	Andrew Rogerson	Opusnet	Australia	Mawson Lakes	5020	8192005 8137 SA				Third Ave
Adam Bateson	Andrew Rogerson	Opusnet Communications	Australia	Adelaide	5005	8192005 8137 SA				Technology Park
Adam Bateson	Andrew Rogerson	Opusnet	Australia	Melbourne	5000	8192005 8103 VIC				Level 9
Adam Bateson	Andrew Rogerson	Opusnet Pty Ltd	Australia	Vest Path	5000	8192005 8103 VIC				Level 9
Adam Bateson	Andrew Rogerson	Opusnet Pty Ltd	Australia	Clare	5100	8192005 8103 VIC				Level 9/10 Moore Street
Adam Bateson	Andrew Rogerson	Opusnet	Australia	Liscombe	5241	8192005 2218 NSW				East Street
Adam Bateson	Andrew Rogerson	Cable & Wireless Opusnet	Australia	Chateauwood New 2062	2062	8192005 2218 NSW				L2, 485 Victoria Ave
Adam Bateson	Andrew Rogerson	Cable & Wireless Opusnet Pty	Australia	Sydney	2000	8192005 2218 NSW				Opus Centre Sydney
Adam Bateson	Andrew Rogerson	Opusnet Communications	Australia	North Sydney	1585	8192005 2218 NSW				51 Miller St
Adam Bateson	Andrew Rogerson	Opusnet Pty Ltd	Australia	North Sydney	1588	8192005 2218 NSW				51 Miller Street
Adam Bateson	Andrew Rogerson	Opusnet Pty Ltd	Australia	Chateauwood	2025	8192005 2218 NSW				Level 20, Tower B, 825 Park
Adam Bateson	Andrew Rogerson	Opusnet	Australia	Chateauwood	2025	8192005 2218 NSW				Tower B, Zenith Centre, 82
Adam Bateson	Andrew Rogerson	Opusnet Pty Ltd	Australia	North Sydney	1588	8192005 2218 NSW				51 Miller St
Adam Bateson	Andrew Rogerson	Opusnet	Australia	Chateauwood	2067	8192005 2218 NSW				Tower B, Zenith Centre
Adam Bateson	Andrew Rogerson	Opusnet	Australia	North Sydney	1588	8192005 2218 NSW				Level 20 Miller Street

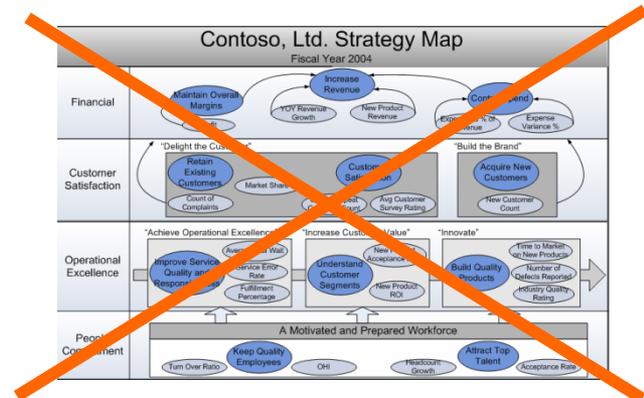
No se saben usar y desarrollar scorecards y dashboards



Análisis: El tiempo dedicado a analizar, poca información relevante



Arquitectura empresarial desalineada a las estrategias y acciones, y por tanto una baja adopción y productividad



# Ejemplos de BI reales



- Cerca de la bancarrota en el 1990s
- Invirtió en BI \$30 millones para:
  - Mejorar procesos de negocio
  - Mejorar servicios a clientes
- En 6 años recupero \$500 millones de la inversión

# Ejemplos de BI reales



- Fabricante de unidades de discos duros para computadores
- Ventas anuales sobre \$3 billones
- Usaron BI para mejorar:
  - Inventarios
  - Cadenas de proveedores
  - Ciclo de vida de Productos
  - Relaciones con los clientes.
- Redujeron costos operacionales en 50%

# Otros ejemplos

- Cadenas de Hoteles

- compilar estadísticas de ocupación promedio, para determinar precio por habitación, etc.
- también estudia el mercado, para determinar competencia. Esas tendencias pueden ser anuales, mensuales, etc.

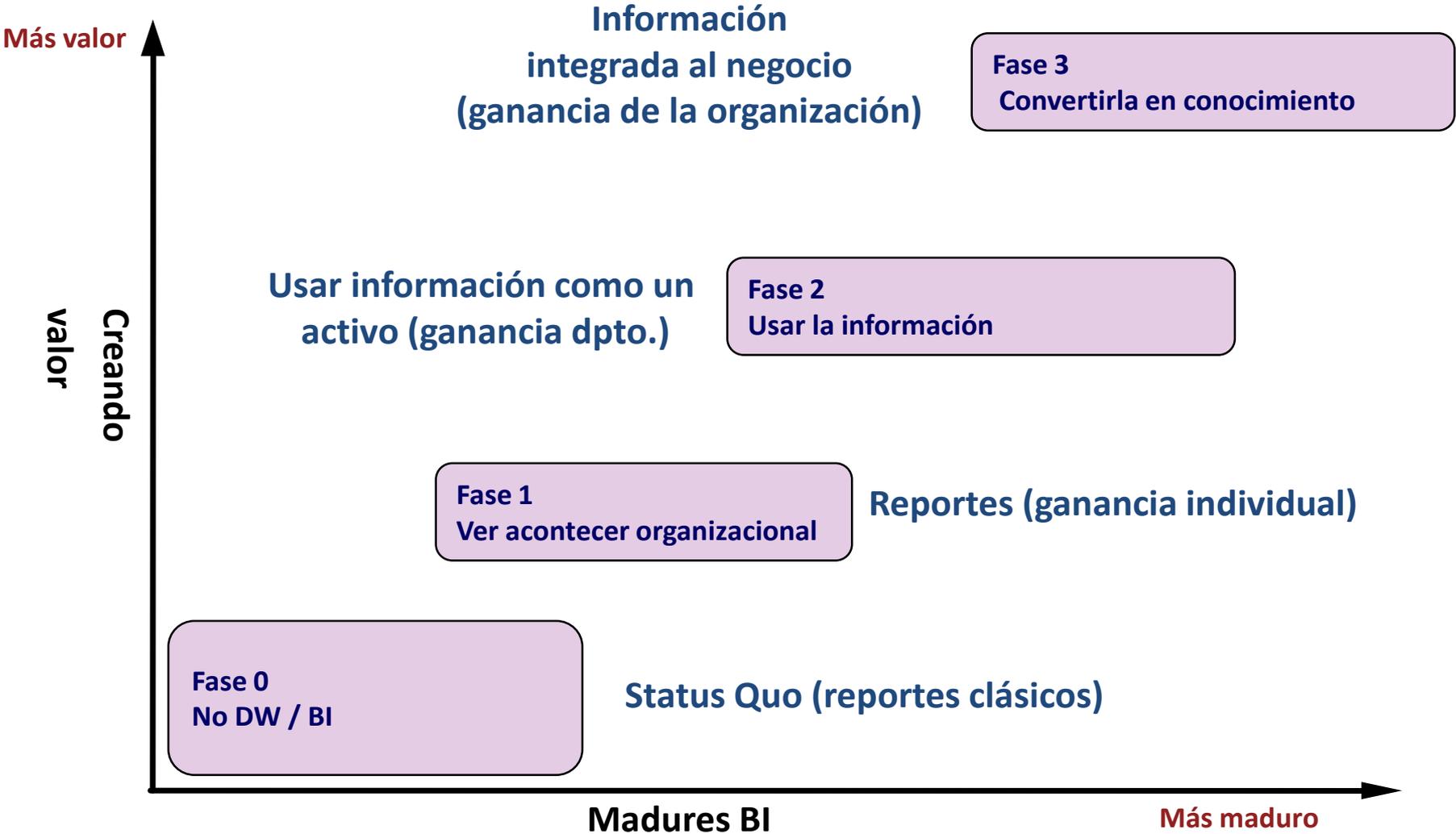
- Bancos

- Determinar a que clientes ofrecerles nuevos productos

- Compañía de telecomunicaciones

- Crear información relevante de sus usuarios

# Cambio en el uso de la Información



# Ejemplos de las Aplicaciones de la Inteligencia de Negocios



## Aplicaciones Analíticas

## Preguntas de Negocios

## Qué buscaría la IN

Segmentación de los Clientes

¿ A qué segmento del mercado pertenece mi cliente y cuales son sus características?

Relación personalizada que da mayor satisfacción y retención

Intensiones de consumo

¿ Qué tipos de clientes responderán más a mi Promoción?

Dirigirse a los clientes de acuerdo a sus necesidades

Rentabilidad del Cliente

¿Cuál es la rentabilidad de la vida útil del cliente?

para incrementar la lealtad

Tomar mejores decisiones de negocios de acuerdo a la

Detección de Fraudes

¿ Cómo saber que la posibilidad de ser fraudulenta que tiene una transacción ?

rentabilidad de los clientes

Determinar inmediatamente el fraude y tomar acciones para minimizar el costo

Evitar la Pérdida de Clientes

¿ Qué cliente tiene el riesgo de irse a la competencia?

Obtener los datos

rápidamente y tomar las

medidas para que

permanezcan en la empresa

Optimización del Canal

Escoger el mejor canal para cada segmento

Interactuar con los clientes

de acuerdo a su preferencia y

las necesidades para reducir

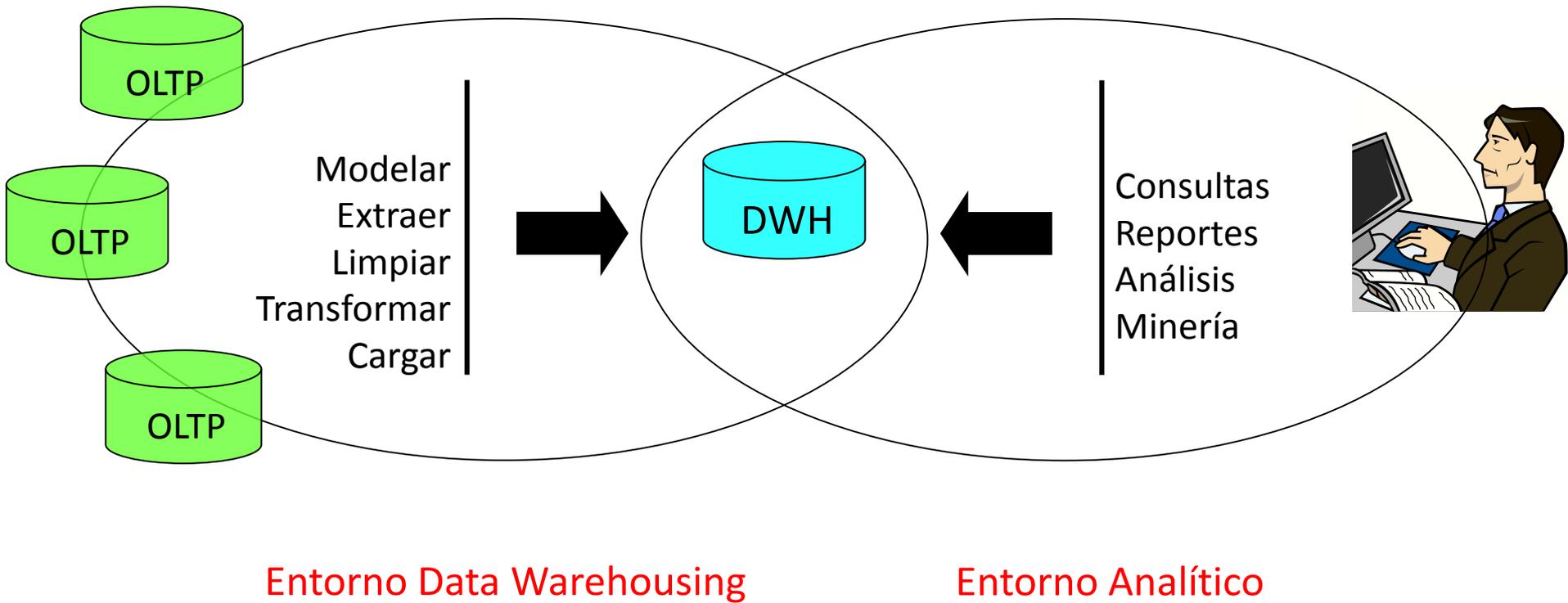
costos

# Principales Componentes de la Inteligencia de Negocios



- **Almacén de datos centralizado** (data warehouse),
- El conjunto de herramientas que utilizará el usuario final (**business analytics**),
- Las relaciones no conocidas entre las variables, que tienen que descubrirse mediante la **minería de datos** (también minería de texto y de la web),
- Metodologías complementarias como BPM (**Business Performance Management**), las cuales sirven para monitorear el desempeño y obtener ventaja competitiva

# Principales Componentes de la Inteligencia de Negocios



# Principales Componentes de la Inteligencia de Negocios



## Gestión y almacenamiento

Data warehousing

Cáalidad del dato

## Presentación de la inform.

Dashboards

Búsqueda

Visualización

Reportes

## Análisis

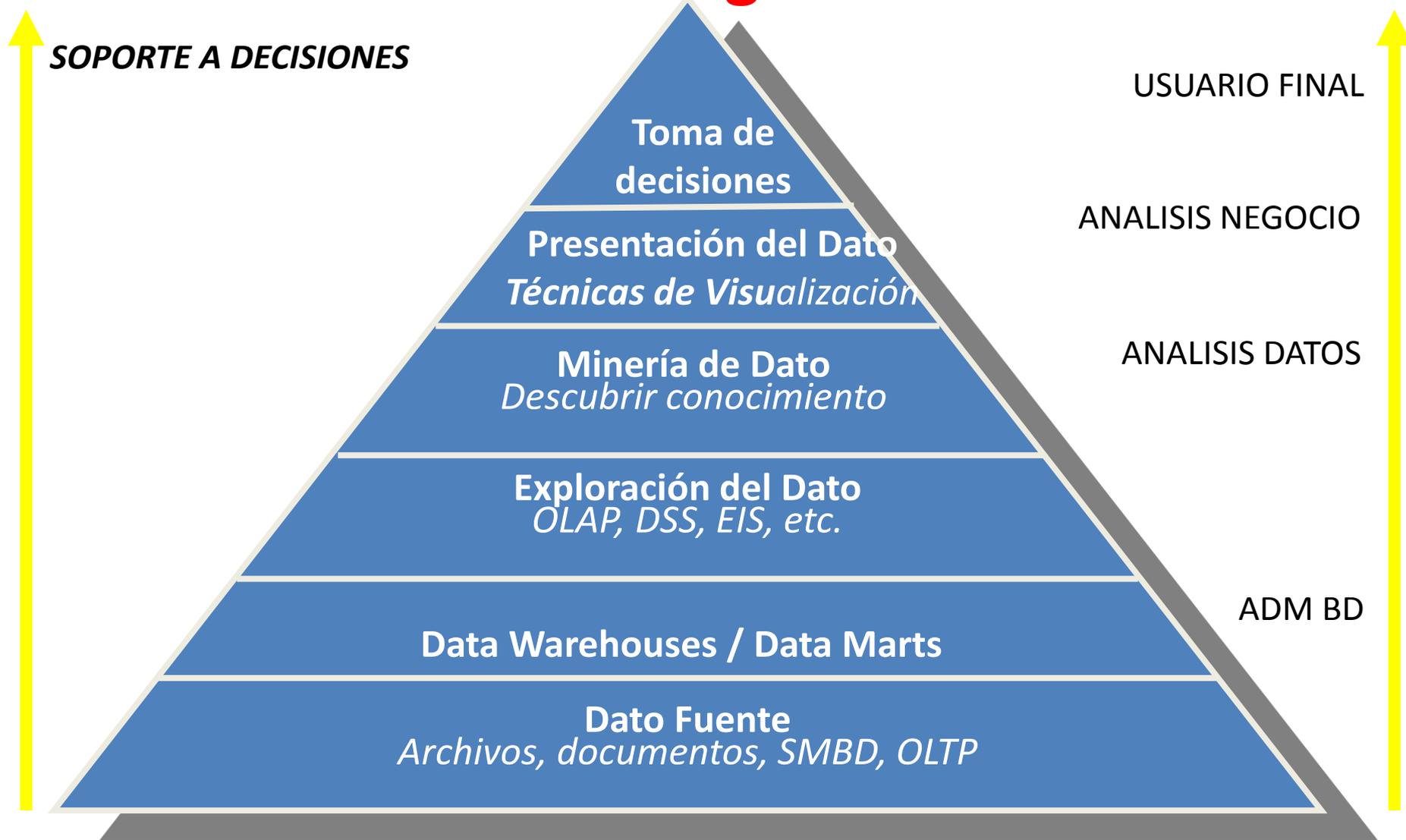
Análisis OLAP

Análisis Ad-hoc

Minería de datos

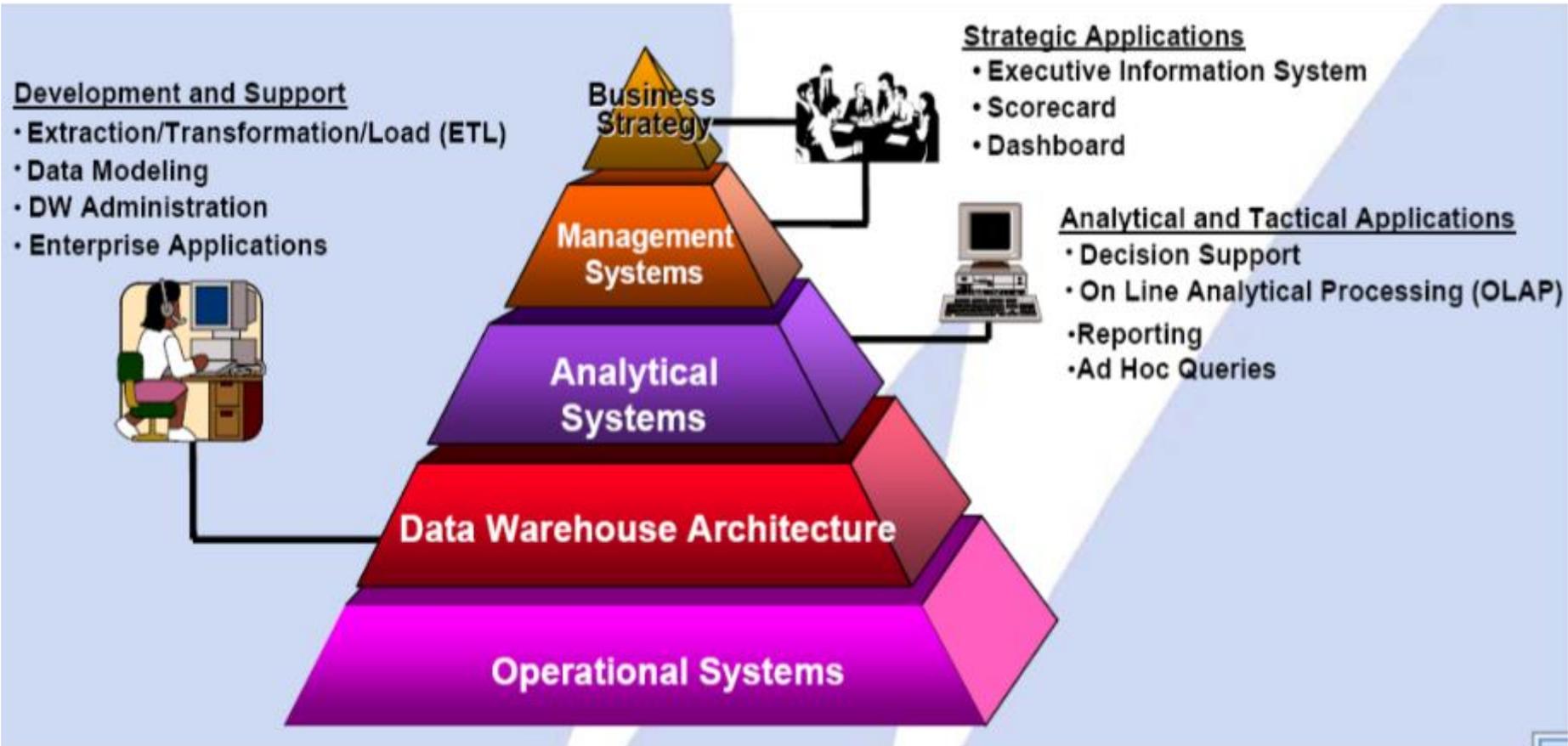
Scorecards

# Principales Componentes de la Inteligencia de Negocios





# Jerarquia de las Soluciones



(Tomado de Heinz, 2014)



# Dashboards

## Una interfaz entre las herramientas de BI y el usuario

- Se asemeja a un tablero de instrumentos del coche
- Contiene imágenes visuales para representar rápidamente métricas de negocio específicos de interés para la gestión
- Ayuda a controlar la gestión de ingresos y ventas, niveles de inventario, e identificar las tendencias y los cambios en el tiempo

# Scorecards

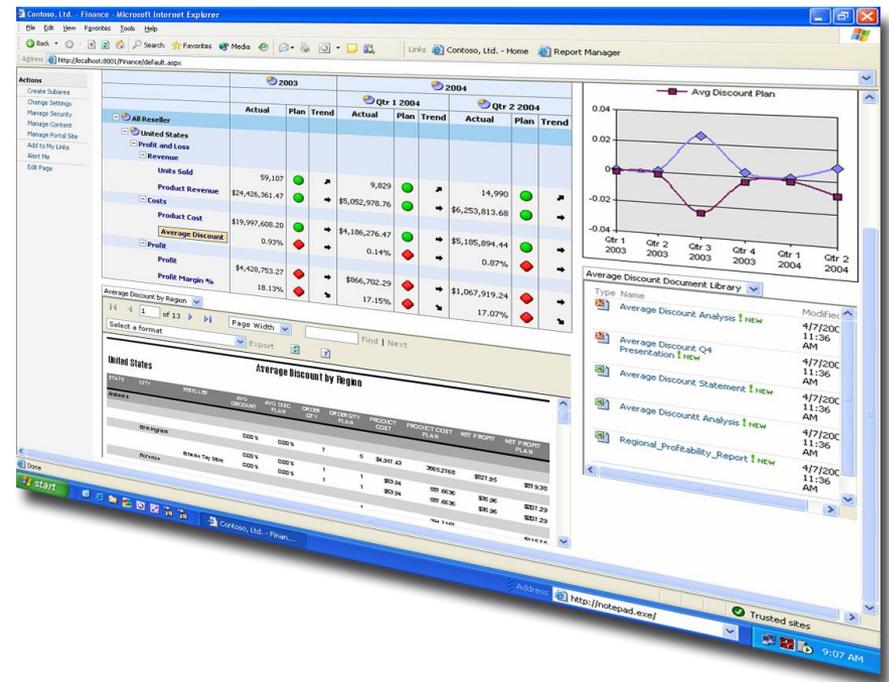


Es

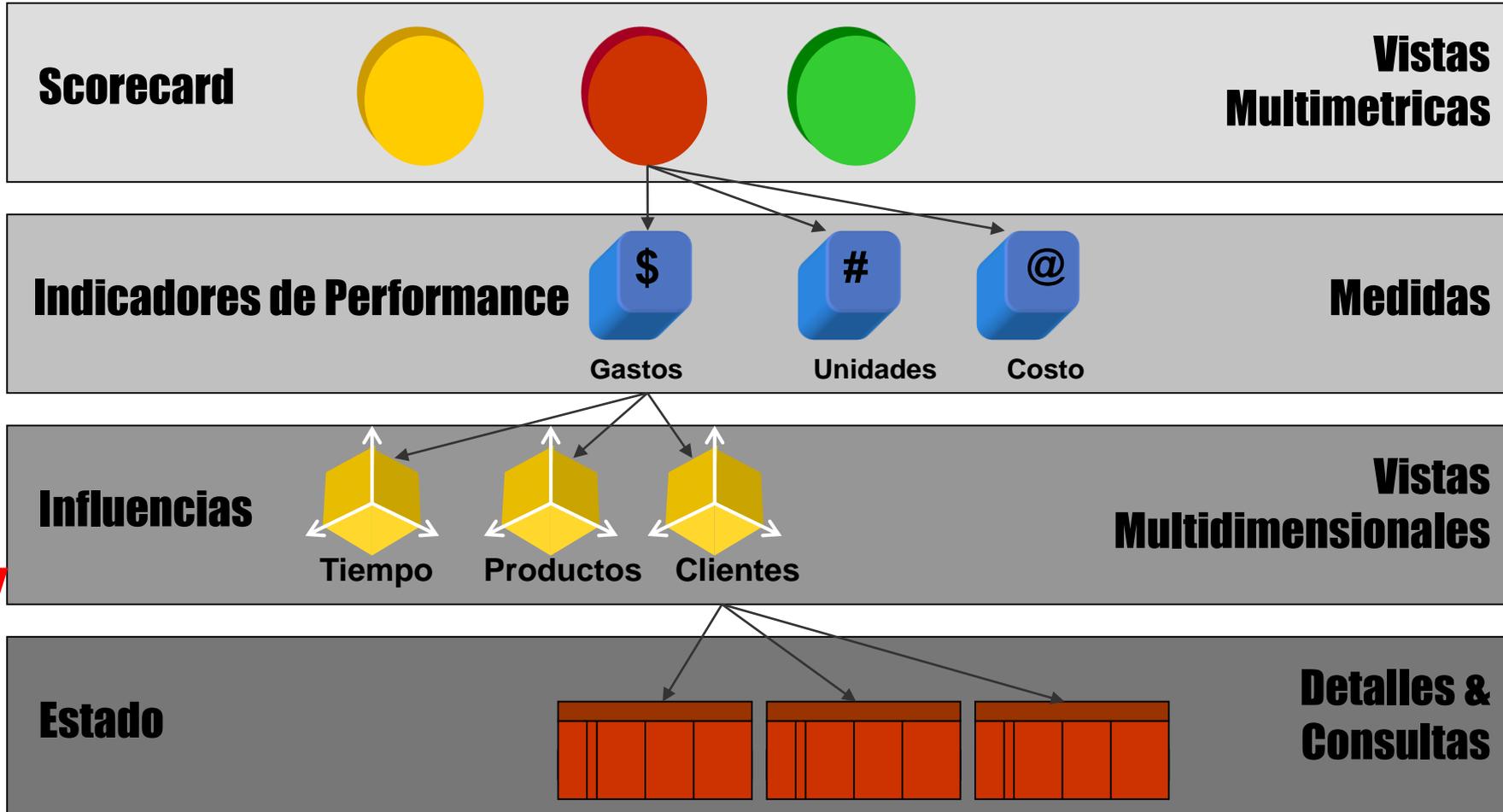
- Expresión y la visualización de una estrategia
- Punto de entrada para el análisis de la mayoría de los usuarios empresariales
- Reduce el riesgo de proporcionar datos brutos o datos en grandes volúmenes

*Debe ser:*

- ◆ Portable
- ◆ Reusable
- ◆ Flexible,
- ◆ Fácil de usar
- ◆ Integrable



# Análisis jerárquico de la información





# Proceso de Inteligencia de Negocios



# Grandes procesos



**Analizar**

*Drill down, multidimensional analysis*

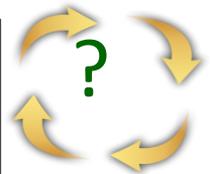
**Reportar**

Datos operacionales detallada

**Monitorear**

*Dashboards, scorecards*

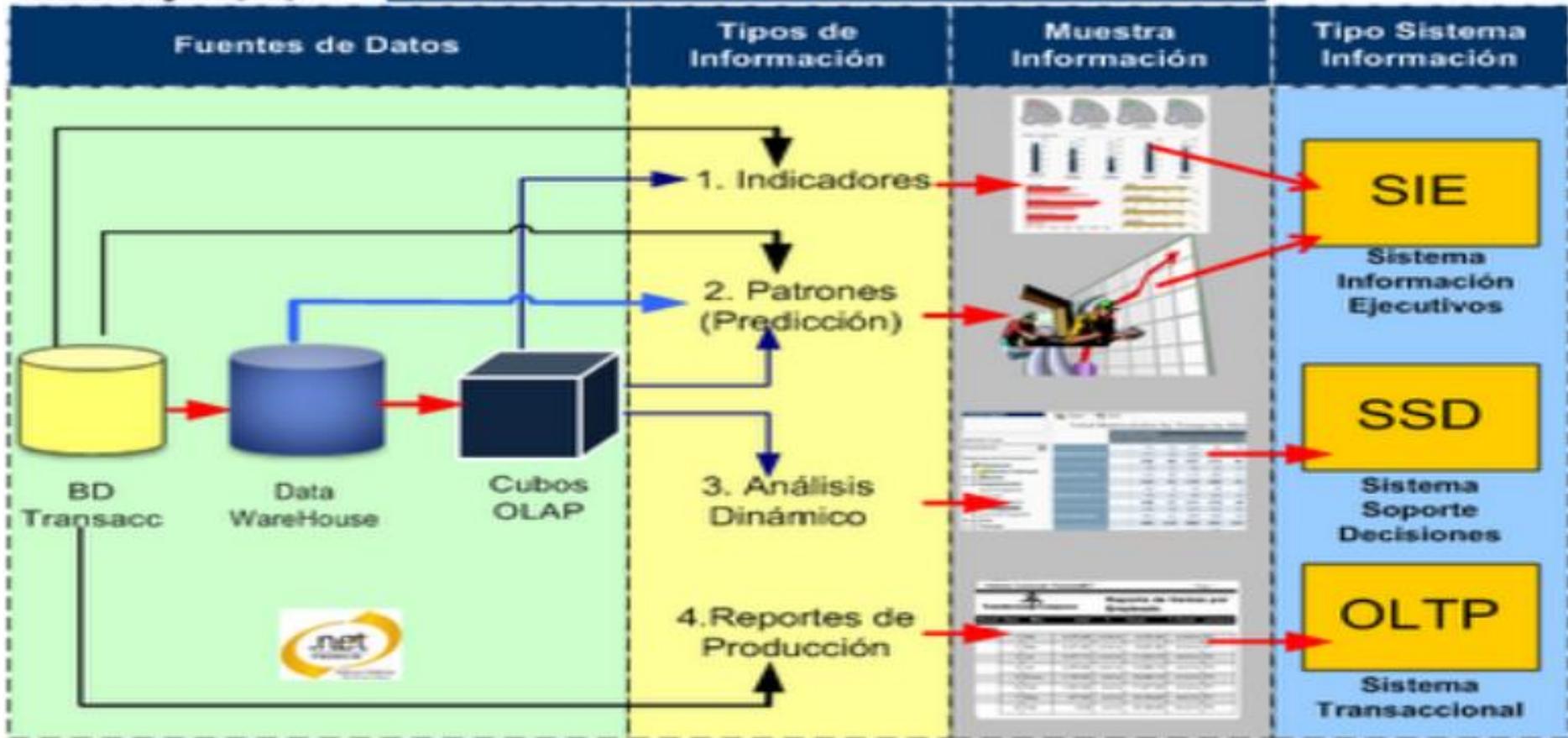
**Planificar, Tomar decisiones**



# Inteligencia de Negocios por procesos



## INTELIGENCIA DE NEGOCIOS. BI/DW-T



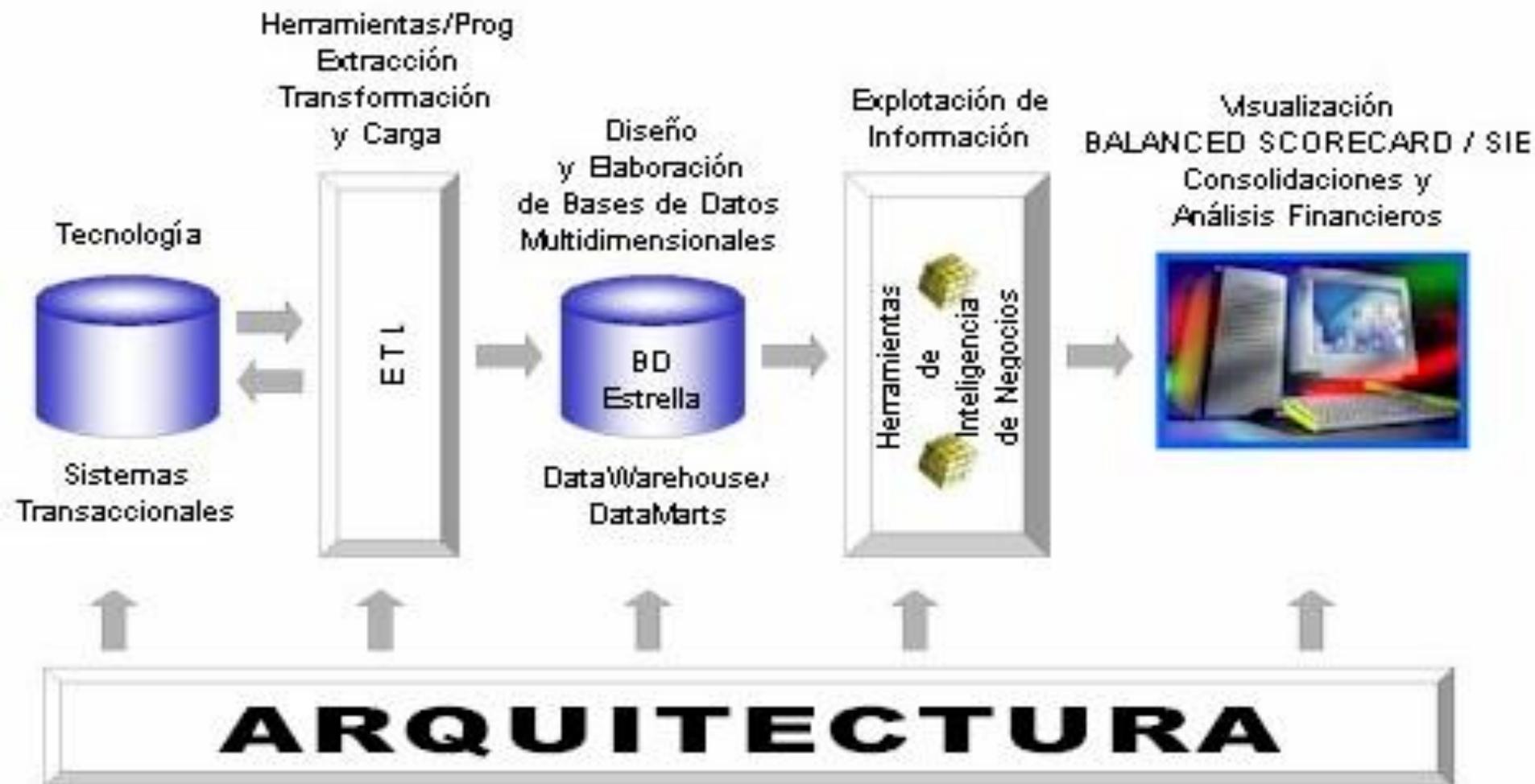
# Inteligencia de Negocios (Business Intelligence - BI)



## FASES



# PROCESO DE INTELIGENCIA DE NEGOCIOS



# Por dónde comenzar?



# Metodología para procesos de BI

***Definición de las situaciones de destino:*** definir las principales preguntas que el proyecto de BI debe responder.

- Situaciones objetivos
- Indicadores

***Definición de los Modelo de datos del proyecto de BI:*** preparar los datos a utilizar para calcular los indicadores.

- Analizar los datos operativos de la organización
- Diseñar la base de datos multidimensional
- Diseñar el proceso de ETL, que se utilizará para recuperar los datos desde las diversas fuentes de datos.
- Ejecutar el proceso ETL sobre los datos operacionales de la organización, para concebir la vista minable operativa

***Extracción de conocimiento (indicadores):*** definir diferentes tipos de modelos o métricas para la interpretación estadística y el análisis de las situaciones objetivos. En este nivel se utilizan tecnologías como OLAP, minería de datos, etc.

# Definición de las situaciones de destino: El problema

- Altos niveles de deserción en estudiantes de modalidad de estudios a distancia (MaD).



- Por que ocurre la deserción? Que factores influyen? Existe un modelo que permita predecir cuando un estudiante va a desertar? El problema es general o está focalizado en ciertas carreras?

# Definición de las situaciones de destino: El contexto



- Proceso de aprendizaje en línea centrado en los estudiantes de MaD de la UTPL
- Aprendizaje virtual apoyado en el uso herramientas LMS y VLE (sistemas SGA y EVA)
- Necesidad mejorar los niveles de permanencia de los estudiantes en las carreras que eligieron.

# Definición de las situaciones de destino: Objetivo

- Determinar los factores influyentes en la deserción estudiantil en este tipo de procesos de aprendizaje



- Buscar una respuesta al problema planteado, aplicando LA, a través del aprovechamiento de los datos históricos generados en el proceso de aprendizaje de estudiantes de MaD de UTPL.

# Definición de las situaciones de destino:

## Tipo de tareas de AdD

### Centrado en el proceso de aprendizaje:

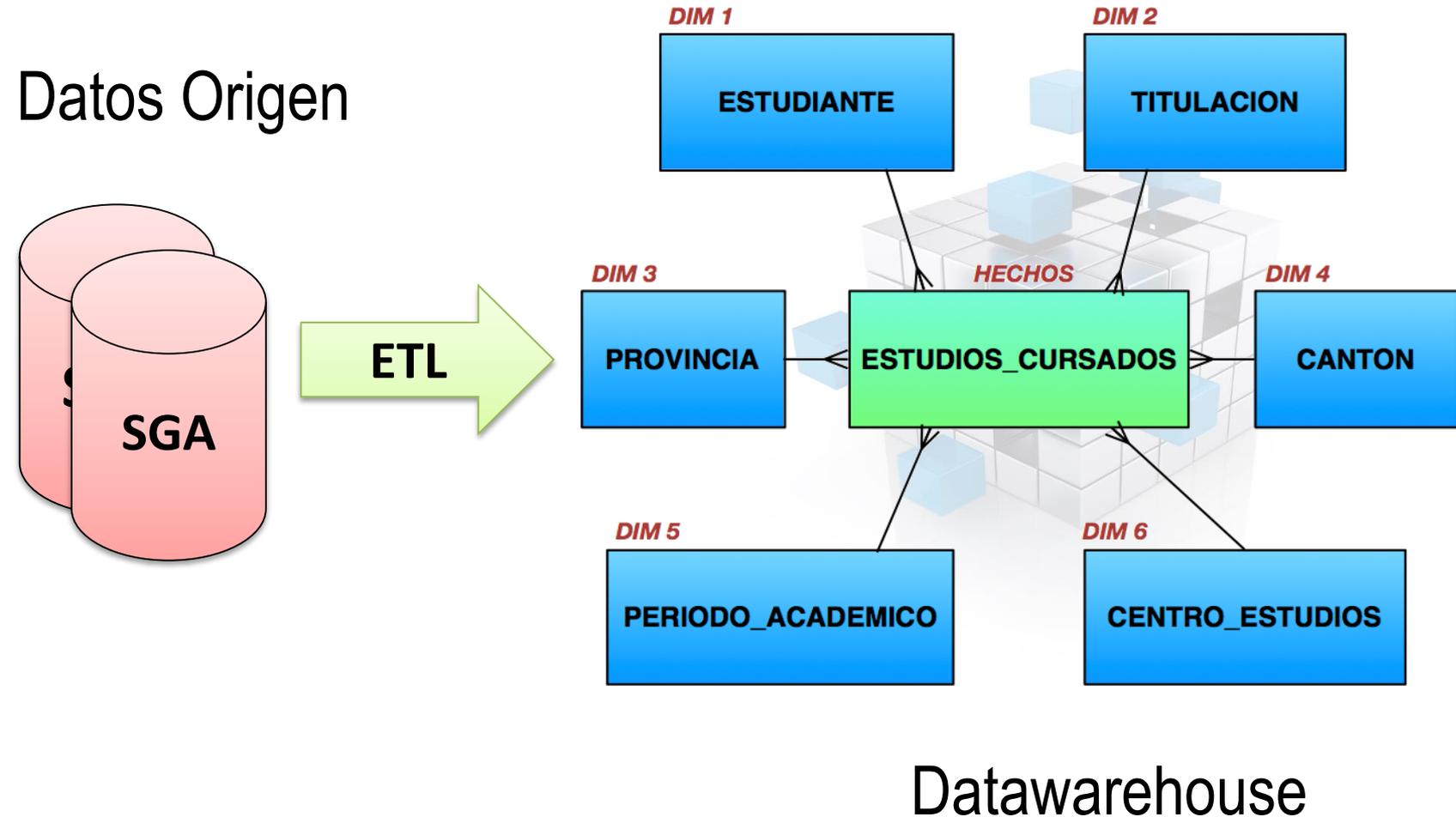
- Monitorear el progreso de los estudiantes.
- Analizar el proceso de evaluación.
- Buscar recursos y actividades a recomendar.



### Centrado en los alumnos:

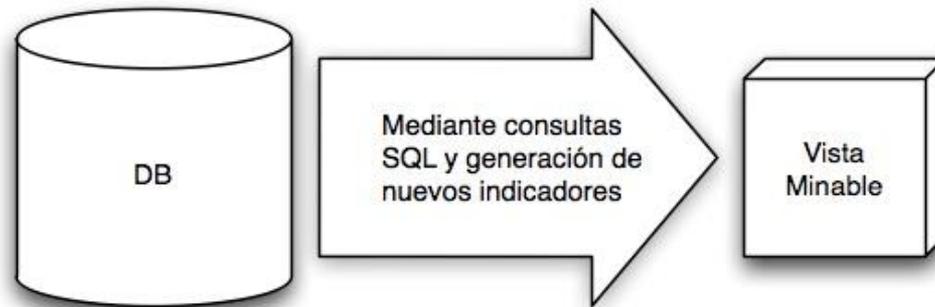
- Descubrir los estilos de aprendizaje.
- Determinar cómo el rendimiento de cada estudiante varía en función de los patrones de aprendizaje de los cursos.
- Identificar problemas y necesidades de los estudiantes

# Definición de los Modelo de datos: Diseño multidimensional



# Definición de los Modelo de datos: El proceso

## ETL



La migración de los datos se realizó mediante consultas SQL, lo cual permitió además generar nuevos indicadores tales como: Porcentaje de avance, Status carrera.

```
CREATE VIEW tmp_notas_componente_carrera
FROM
SELECT x.*, t.nota
FROM
(
SELECT DISTINCT d.ent_id, d.pra_id, d.coe_id,
IF (d.ETR_CODIGO in
('HOMOLOGADO','APROBADA',
'VALIDADO','REVALIDADO'),
'ACREDITADO','NOACREDITADO') AS "PROMO"
FROM datos d
ORDER BY 1,3
) x, tmp_notas_componente t
WHERE x.ent_id = t.ent_id
AND x.coe_id = t.coe_id
AND x.PROMO = t.PROMO
ORDER BY x.ent_id, x.pra_id, x.coe_id;
```

```
CREATE VIEW estad_promedios_notas
FROM
SELECT t.ent_id, t.pra_id,
round(ifnull(avg(IF(PROMO='ACREDITADO',
t.nota,NULL)),0),2)
AS "PROMEDIOAPROBACION",
round(avg(t.nota),2) AS "PROMEDIODIGENERAL"
FROM tmp_notas_componente_carrera t
GROUP BY 1,2;
```

```
UPDATE c_h_estudios_cursados h
SET h.promedio_aprobacion =
(SELECT e.promedioaprobacion
FROM estad_promedios_notas e
WHERE e.ent_id = h.ent_id
and e.pra_id = h.pra_id)
h.promedio_general =
(SELECT e.promediogeneral
FROM estad_promedios_notas e
WHERE e.ent_id = h.ent_id
and e.pra_id = h.pra_id);
```

# Extracción de conocimiento (indicadores): operaciones OLAP para el cálculo de indicadores

Indicador	Tipo operación OLAP
Índice de deserción por carreras	Roll-up
Centros universitarios con mayor índice de deserción	Roll-up
Índice de deserción según rango de edad de estudiantes	Roll-up
Índice acumulado de deserción según nivel de avance	Roll-up

Titulación	Indice Deserción
CIENCIAS DE LA EDUCACIÓN MENCIÓN CIENCIAS HUMANAS Y RELIGIOSAS	0.63
INFORMÁTICA	0.54
CIENCIAS DE LA EDUCACIÓN MENCIÓN QUÍMICA Y BIOLOGÍA	0.51
CIENCIAS DE LA EDUCACIÓN MENCIÓN LENGUA Y LITERATURA	0.48
INGENIERÍA EN ADMINISTRACIÓN DE EMPRESAS TURÍSTICAS Y HOTELERAS	0.48
CIENCIAS DE LA EDUCACIÓN MENCIÓN FÍSICA Y MATEMÁTICA	0.47
COMUNICACIÓN SOCIAL	0.47
ECONOMÍA	0.47
CIENCIAS DE LA EDUCACIÓN MENCIÓN INGLÉS	0.47
ADMINISTRACIÓN DE EMPRESAS	0.45

# Extracción de conocimiento (indicadores): Minería de los datos

## Predicción de deserción estudiantil

- Debido a la complejidad de la propuesta no fue posible crear un modelo predictivo universal, optamos por crear múltiples modelos aplicables a diferentes contextos: Carreras específicas, ubicaciones geográficas, etc.

N.	Condiciones	Instancias	Precisión	<u>Recall</u>
1	2;14,12,2,1,52   4;5,2   6;1,2	2529	0.6965	0.8947
2	4;1,2,4   6;2	1637	0.7108	0.8439
3	4;5,2,4   6;1,2	4490	0.6971	0.8422
4	4;5,1   6;1	1869	0.6708	0.8278

Redes bayesianas

# Características de un Proyecto de Inteligencia de Negocio

Los tipos de análisis que puede realizar la Inteligencia de Negocios son

Mejora las organizaciones al proporcionar perspectivas de negocio a todos los empleados que conducen a una mejor, más rápida, y más relevantes decisiones

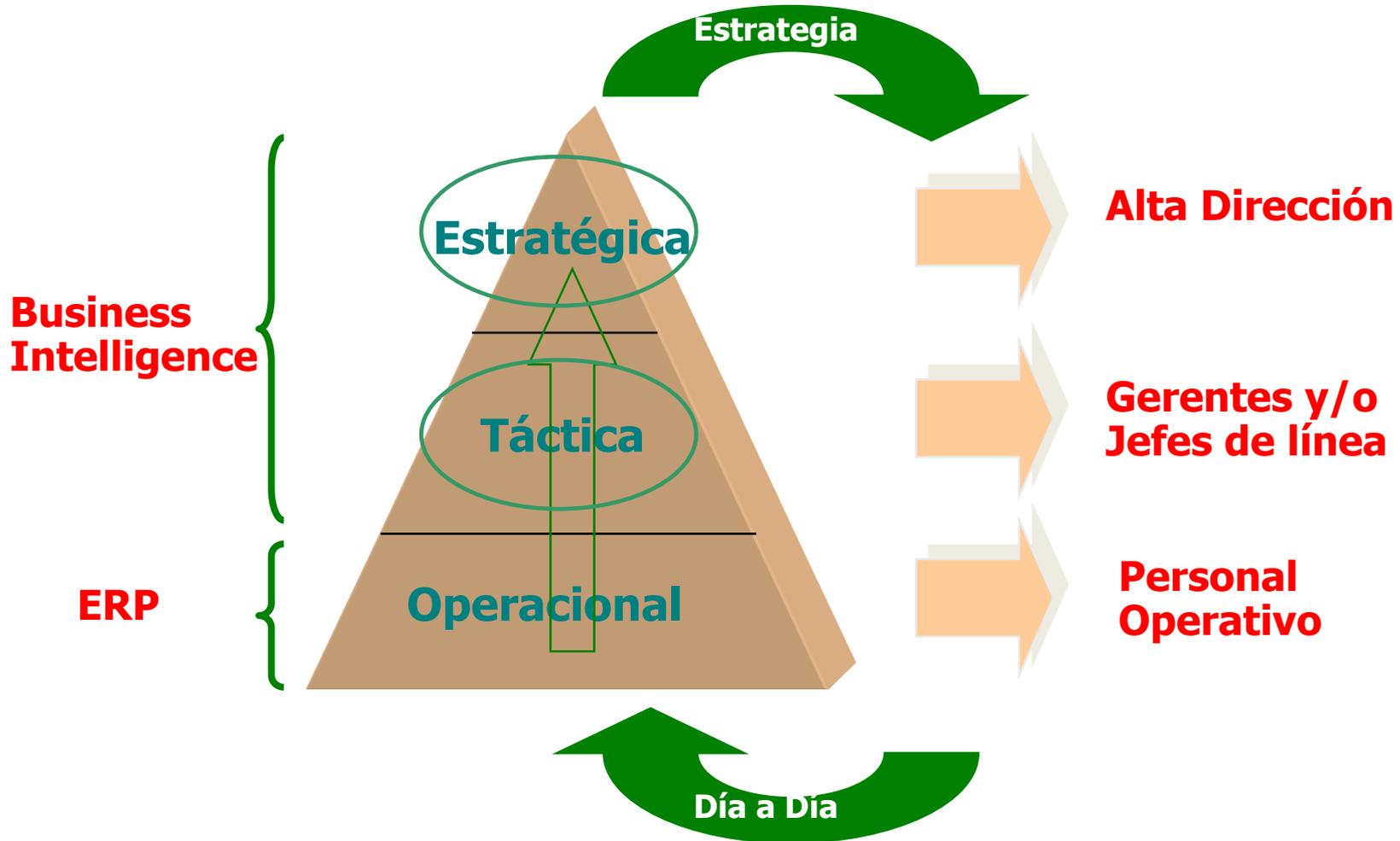


# Características de un Proyecto de Inteligencia de Negocio



	Operacional	Táctica	Estratégica
Focalizada	Operaciones diarias	Análisis de corto termino	Objetivos globales de la organización
Usuarios	Analistas, Operadores	Gerentes	Gerentes
Ventana de Tiempo	Diarios	Semanales a Meses	Mese a año
Datos	Métricas Tiempo real	Métricas históricas	Métricas históricas

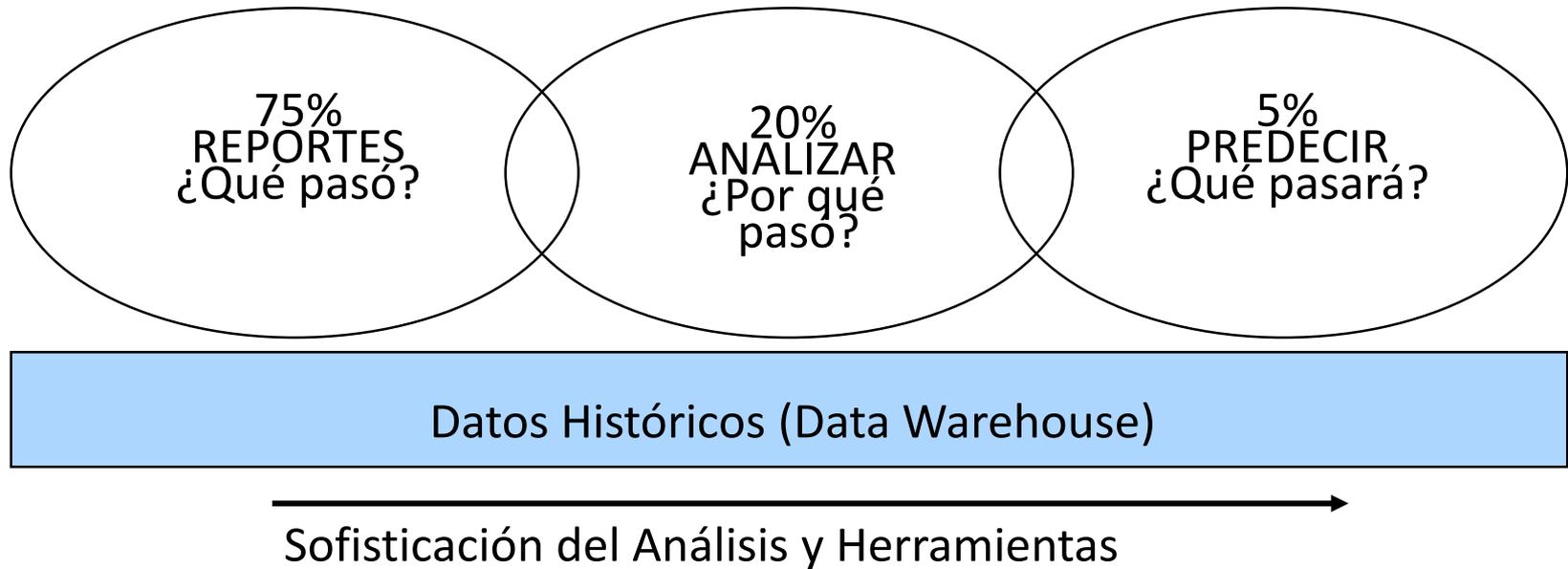
# Características de un Proyecto de Inteligencia de Negocio



# Características de un Proyecto de Inteligencia de Negocio

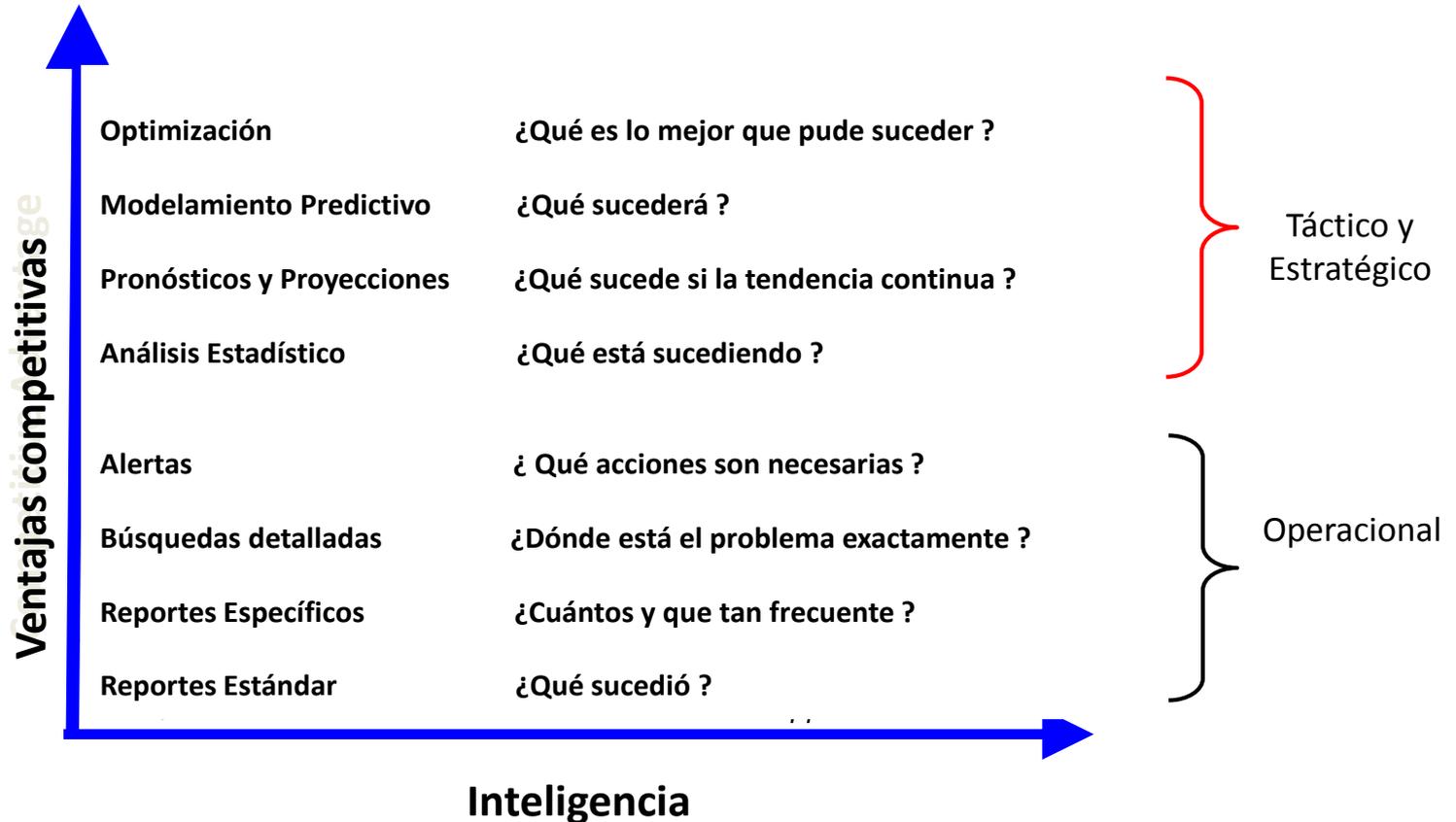


Dominios del Análisis Táctico y Estratégico

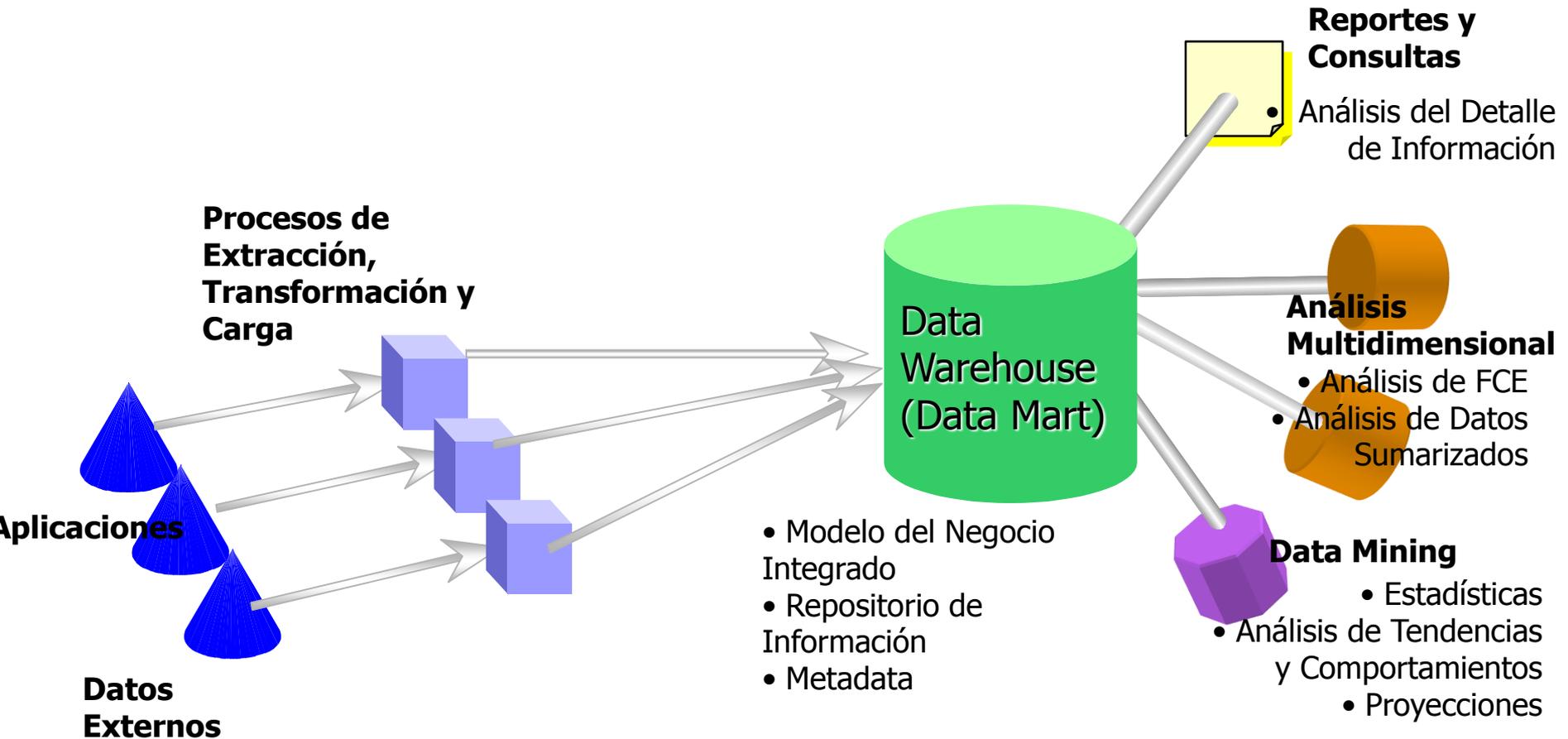


- **Decisiones Tácticas, siguiente semana o mes**
- **Decisiones Estratégicas, siguiente semestre o año**

# Características de un Proyecto de Inteligencia de Negocio



# Arquitectura IN



# Arquitectura IN



## Preparación de los Datos

## Medios de Almacenamiento

## Difusión

1. Origen de Datos

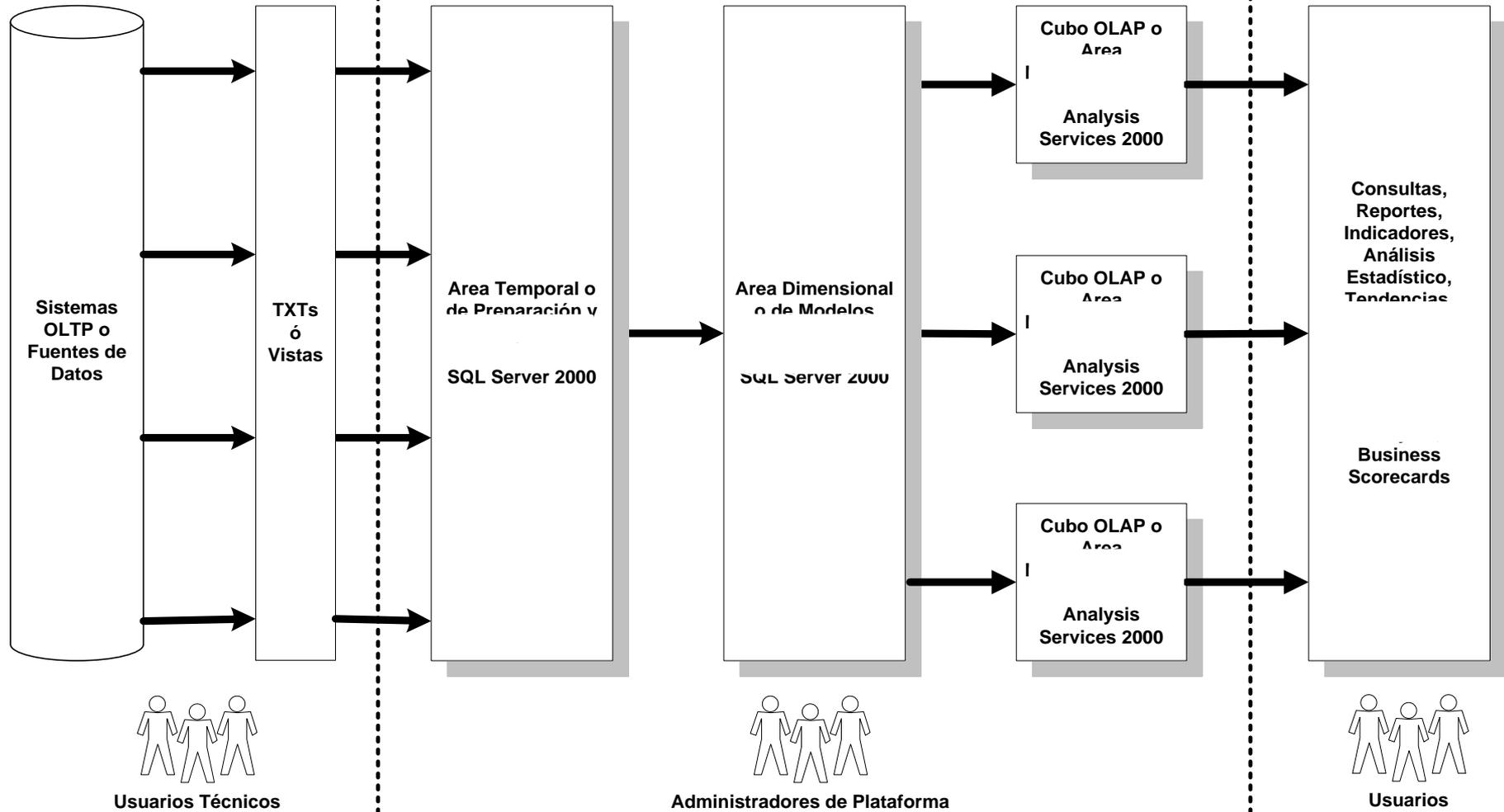
2. Extracción

3. Transformación

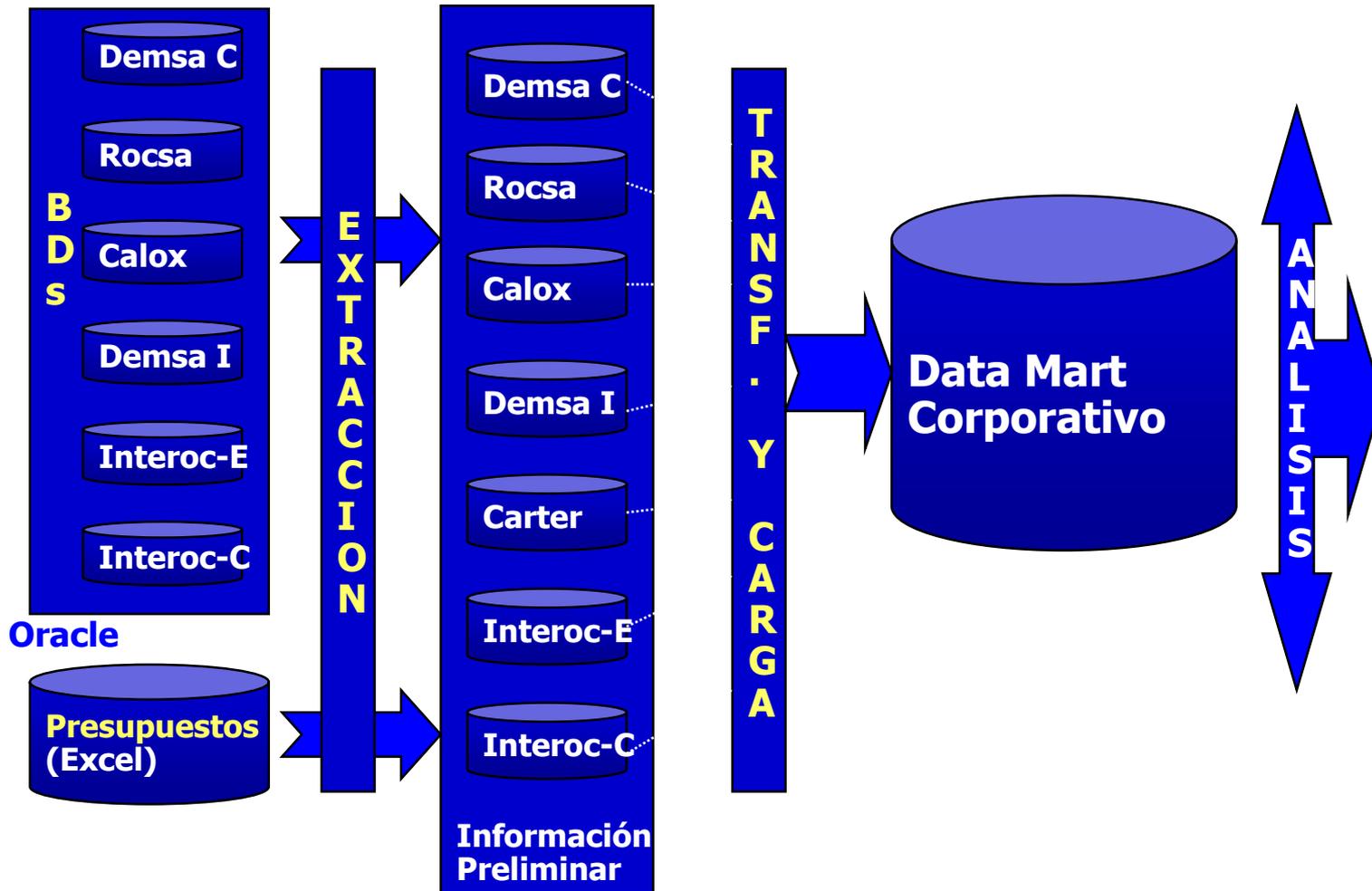
4. Carga o Alimentación

5. DataMarts

6. Visualización



# Arquitectura funcional IN



# Futuro de la IN...



- Los usuarios de IN están empezando a exigir IN en tiempo real o casi en tiempo real

**análisis en relación con su negocio, en línea.**

- Ellos querrán información fresca en la misma forma como monitorean las cotizaciones de la bolsa en línea, chequean su correo electrónico, etc..

**Análisis mensual y semanal no será suficiente**



# BI 2.0

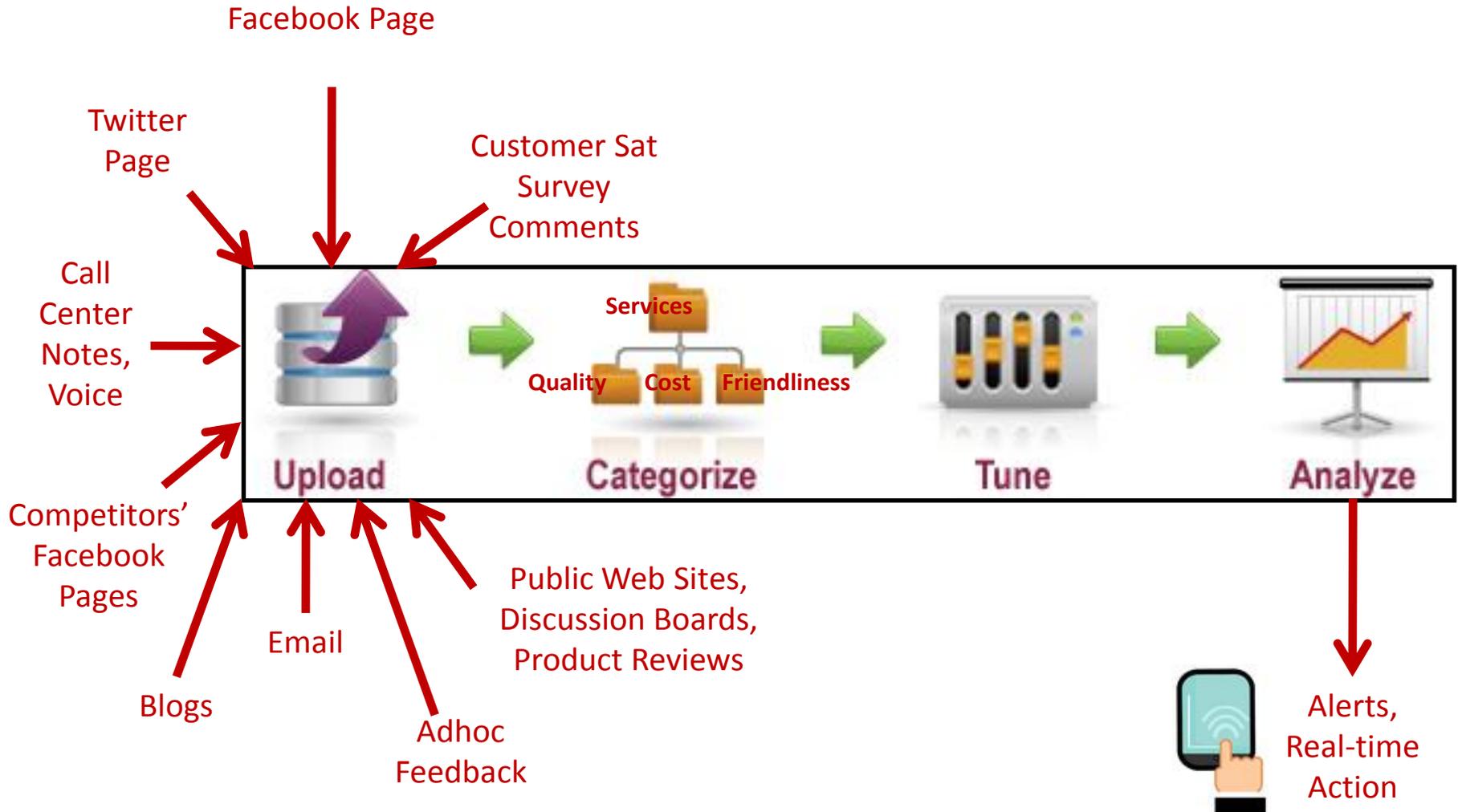
Término acuñado recientemente  
que anuncia el siguiente paso de  
la IN

"BI 2.0" describe  
la adquisición, el suministro  
y el análisis de datos en "tiempo  
real"





# Otras tendencias: Análisis de texto



## Hacer análisis de la web usando IN

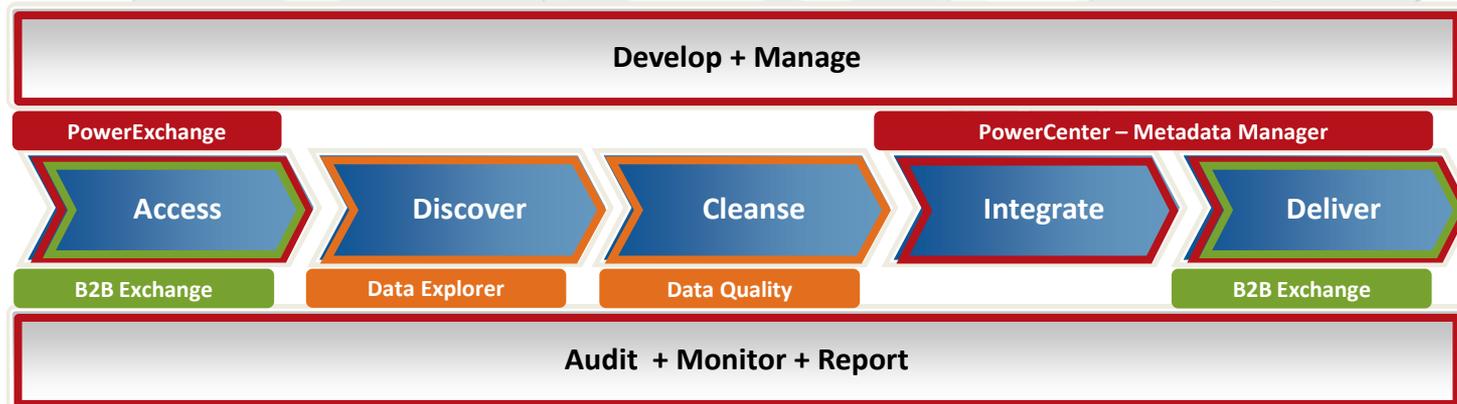


# Otras tendencias:

## Plataformas de Información (ESB)



### Informatica On Demand



On-demand Integration

Enterprise Data Integration

B2B Data Integration

Data Quality Across the Whole Enterprise

# Otras tendencias: Emotional Business Intelligence

$$EBI = EB + EI + BI$$

## Emotional Business Intelligence (EBI)

Emotional Business (EB)?

Emotional Intelligence (EI)??

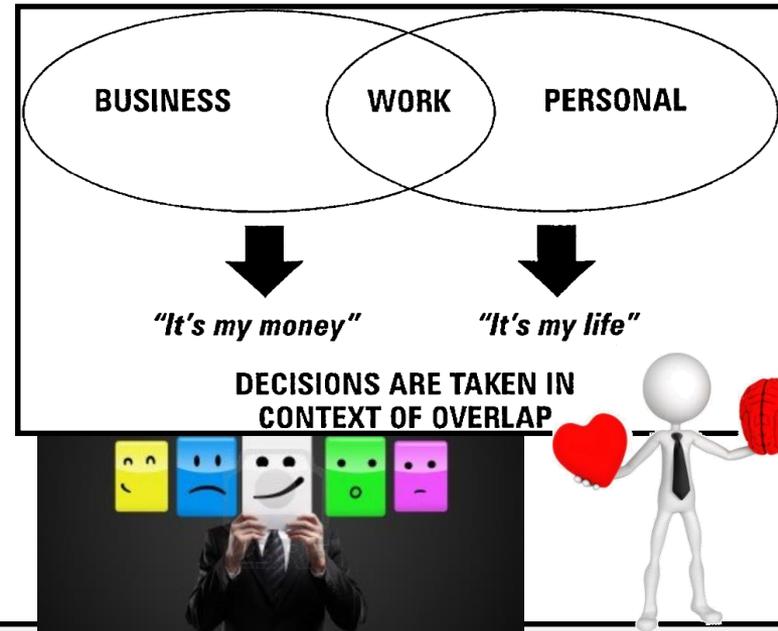
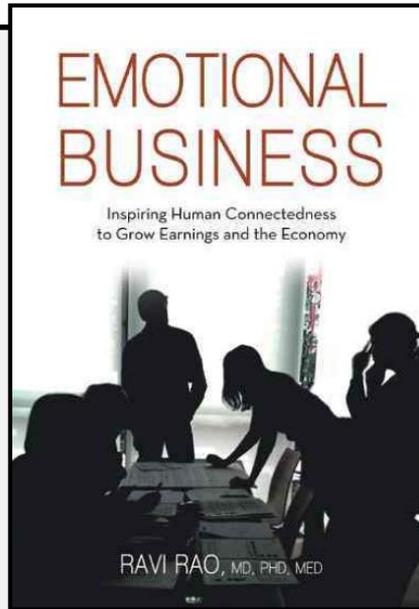
Business Intelligence (BI) ???



# Emotional Business

*“If you get the emotions right, you get the business right”... because otherwise ... “if you’ve ever felt ignored as a customer, humiliated by a teammate, ... or painfully isolated in a big company, then you’ve experienced the business effects of emotional disconnection”*

[Ravi Rao, *Emotional Business*, 2012]



*“Con EB se espera :*

- (a) Explotar la pasión intensa del cliente para aumentar las ventas; y*
- (b) Coadyuvar a los empleados hacia la eficacia .”*

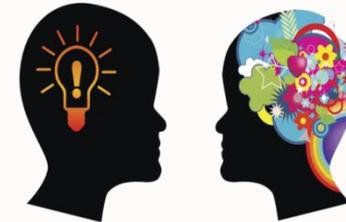
[Jay Cross, 2013]

*“Se asocia con la creación de prototipos de modelos de negocio que se centran en la **construcción emocional (vs. funcional) de las conexiones en la cadena de valor**, para permitir el crecimiento interno en las organizaciones ,y para mejorar la diferenciación de los usuarios , lo que permite descubrir oportunidades de modelos de negocio particular, ya que **vincula los aspectos funcionales (producto o servicio) del modelo de negocio a los aspectos emocionales de la cadena de valor**”* [Sam Bucolo and Cara Wrigley, 2012]

# Emotional Intelligence

*Es la capacidad de reconocer, evaluar y gestionar las propias emociones y las emociones de los otros (individuos o grupos).*

*Cociente emocional (EQ) es una medida de la capacidad de una persona para controlar las emociones (internas y externas) (incluyendo pensamientos y acciones pertinentes), para hacer frente, por ejemplo, a las presiones y demandas. Dicha medición se pretende que sea una herramienta similar al Cociente de Inteligencia (IQ), que es una medida de la inteligencia de una persona.*



**IQ vs. EQ**

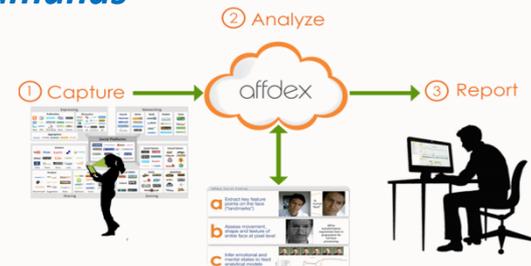


There is a strong opinion that intelligent person has to have a high IQ, however a *smart* person is also (!) requested to have high EQ (i.e., essential Emotional Intelligence skills and capabilities). The same can be also applied while talking about smart company, smart business, smart product or service, smart solution, smart decision, etc.

## **Emotional Intelligence** [Daniel Goleman]:

- (1) **conciencia de sí mismo** (la capacidad de conocer las propias emociones, fortalezas, debilidades, unidades, valores y objetivos y reconocer su impacto en los demás para guiar las decisiones);
- (2) **autorregulación** (controlar o redirigir las emociones perturbadoras y los impulsos, y la adaptación a las circunstancias cambiantes);
- (3) **habilidad social** (gestión de relaciones para mover a la gente en la dirección deseada);
- (4) **empatía** (tener en cuenta los sentimientos de otras personas, especialmente cuando se toman decisiones); y
- (5) **motivación** (impulso para lograr logro).

*Inteligencia Emocional de la máquina (o **Affective Computing**) es el ámbito de la Inteligencia Artificial que se ocupa del diseño de los sistemas y dispositivos que pueden capturar, reconocer, interpretar y procesar las emociones humanas*





UNIVERSIDAD  
DE LOS ANDES  
MÉRIDA VENEZUELA



# Datawarehousing

Jose Aguilar  
CEMISID, Escuela de Sistemas  
Facultad de Ingeniería  
Universidad de Los Andes  
Mérida, Venezuela

# Introducción a Data Warehouse

## Analizando la información de una empresa

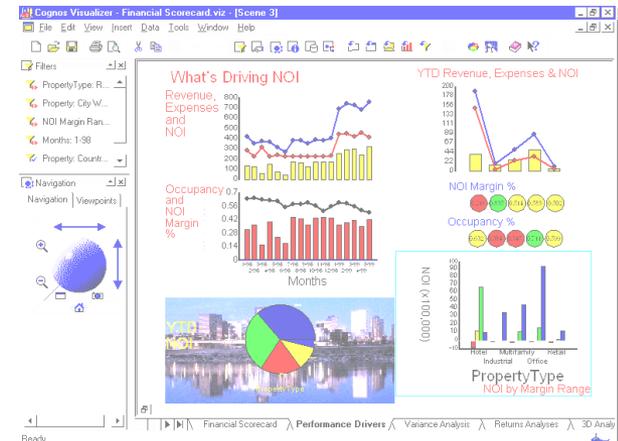
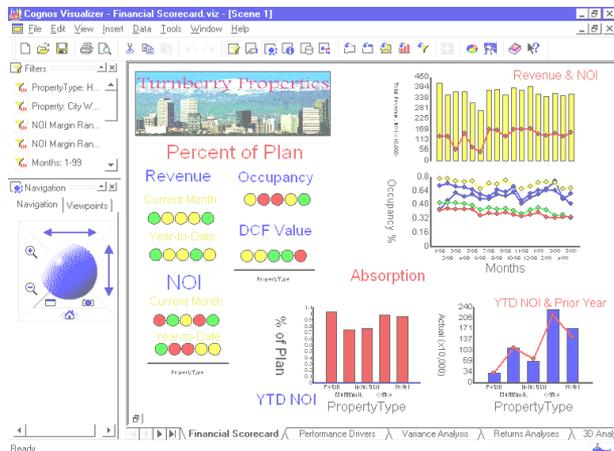
- ✓ Información periódica de las ventas
- ✓ Información del esfuerzo comercial
- ✓ Información sobre los pedidos a los proveedores

## Por qué no integrarla y cruzarla para obtener:

- ✓ En qué zonas se está vendiendo más de cada línea de productos?
- ✓ Quienes son los clientes más rentables?
- ✓Cuál es la relación entre el esfuerzo comercial y las operaciones cerradas?
- ✓ De qué proveedores se está comprando la mayor parte de los productos vendidos ?

# Introducción a Data Warehouse

- ✓ Se necesita entender no solo **QUE** está pasando, sino **CUANDO, DONDE, QUIEN, COMO Y POR QUE**.
- ✓ Requerimientos de información con **OPORTUNIDAD**.
- ✓ **ESCALAR, ENRIQUECER Y COMPARTIR** a todos los usuarios en la organización



# Introducción a Data Warehouse

## Ventas

- Número de pedidos
- Productos pedidos
- Clientes que realizaron los pedidos

- Clientes más rentables
- Pedidos más frecuentes
- Productos más rentables
- % de nuevos clientes

## Servicio al Cliente

- Datos de llamadas de nuestros clientes
- Información sobre el log de nuestra página web

- Qué clientes visitan nuestra página web ?
- % pedidos realizados por nuestros canales de ventas
- Qué consulta es más frecuente?

## Marketing

- Número de campañas realizadas y características de cada una

- % éxito de las campañas
- Qué tipo de clientes han respondido mejor a cada una de las campañas realizadas ?

## Distribución

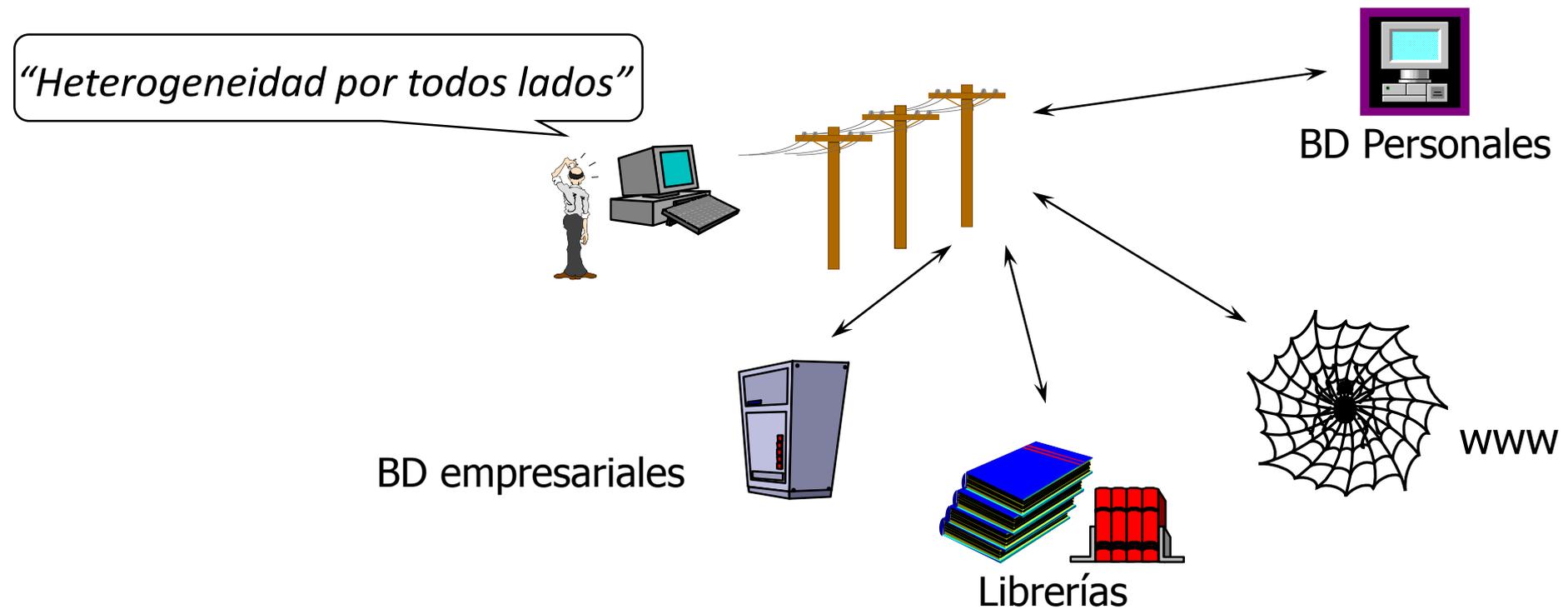
- Productos que salen diariamente del almacén
- Tiempo teórico de entrega

- Número de pedidos retrasados
- Distribuidor que tiene el mayor número de retrasos
- Tiempo medio de entrega

**DATO**

**INFORMACION**

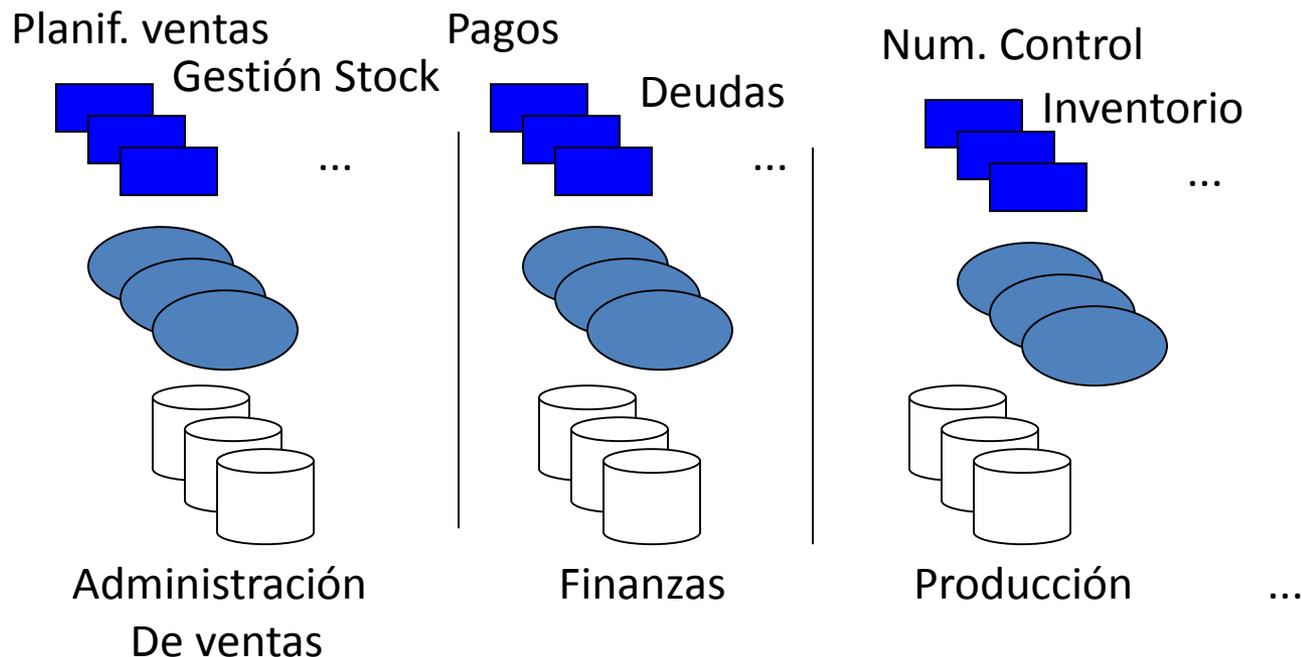
# Problemas: Heterogeneidad de las fuentes de Información



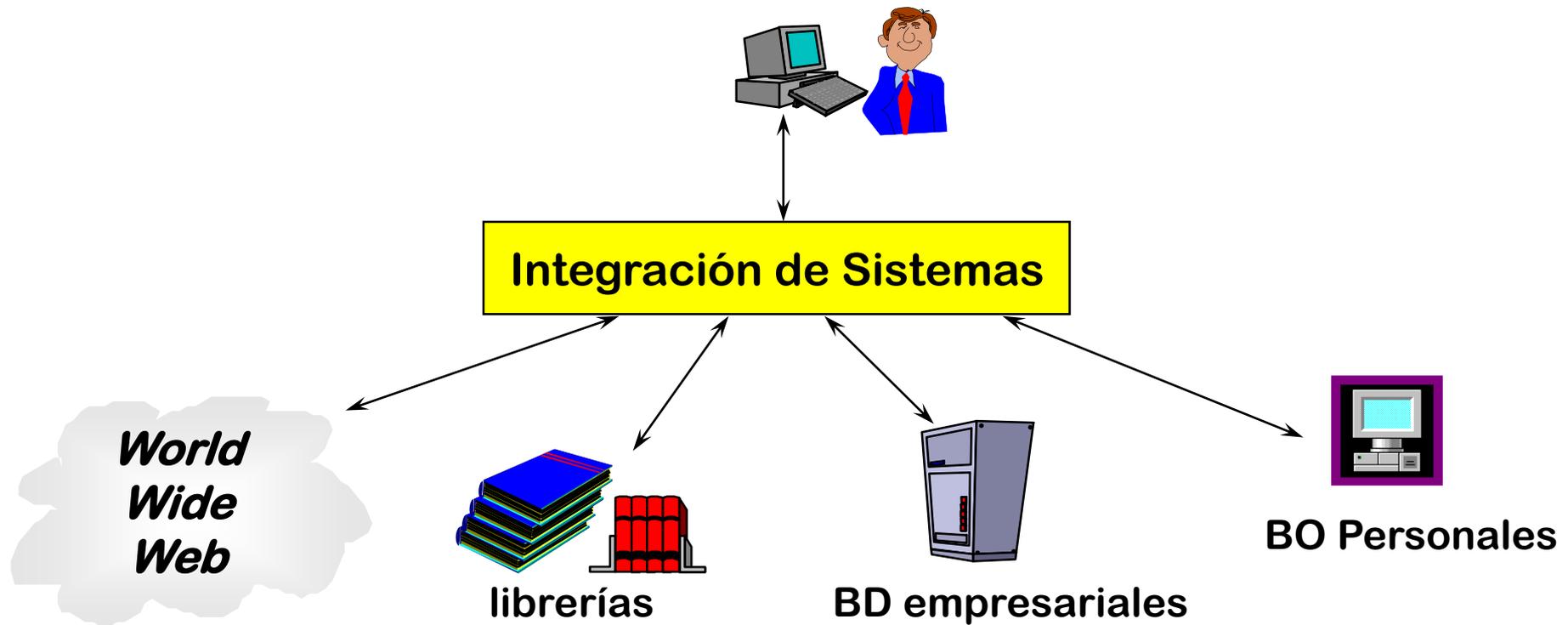
- Diferentes interfaces
- Diferentes representaciones de datos
- Información duplicada e inconsistente

# Problema: Gestión de datos en grandes empresas

- Fragmentación vertical de los sistemas de información
- Desarrollo de las aplicaciones guiadas por los sistemas operativos



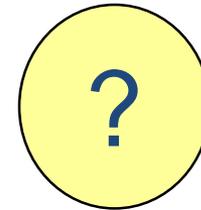
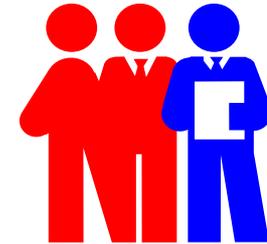
# Objetivo: Unificar Acceso a los Datos



- Recopilar y combinar la información
- Proporcionar visión integrada, en una interfaz de usuario uniforme
- Soportar el intercambio

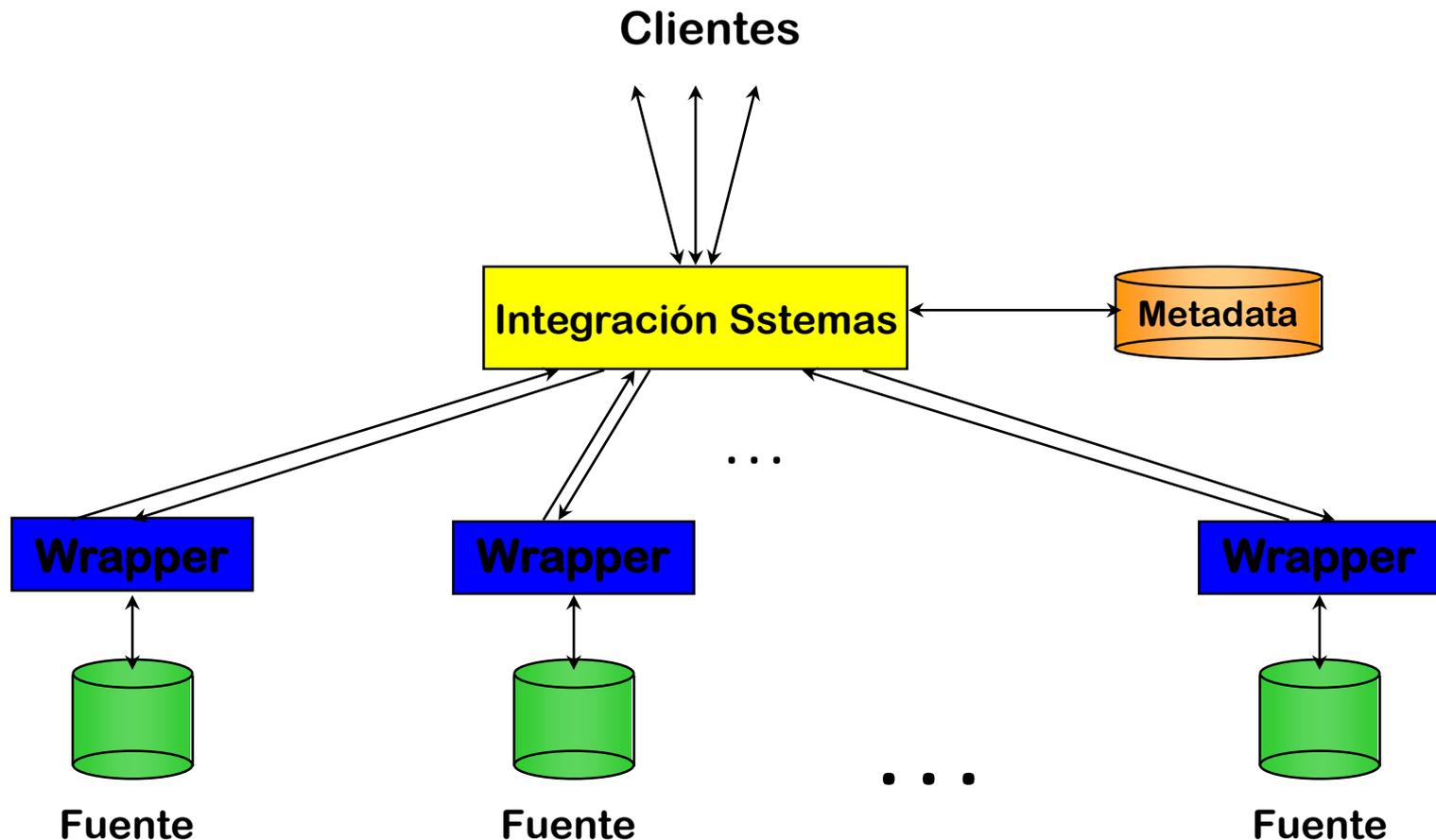
# Objetivo: Unificar Acceso a los Datos

- Dos enfoques:
  - Guiado por la consulta (perezoso)
  - Warehouse (ansioso)



# Enfoque tradicional

- Guiado por la consulta (perezoso, on-demand)

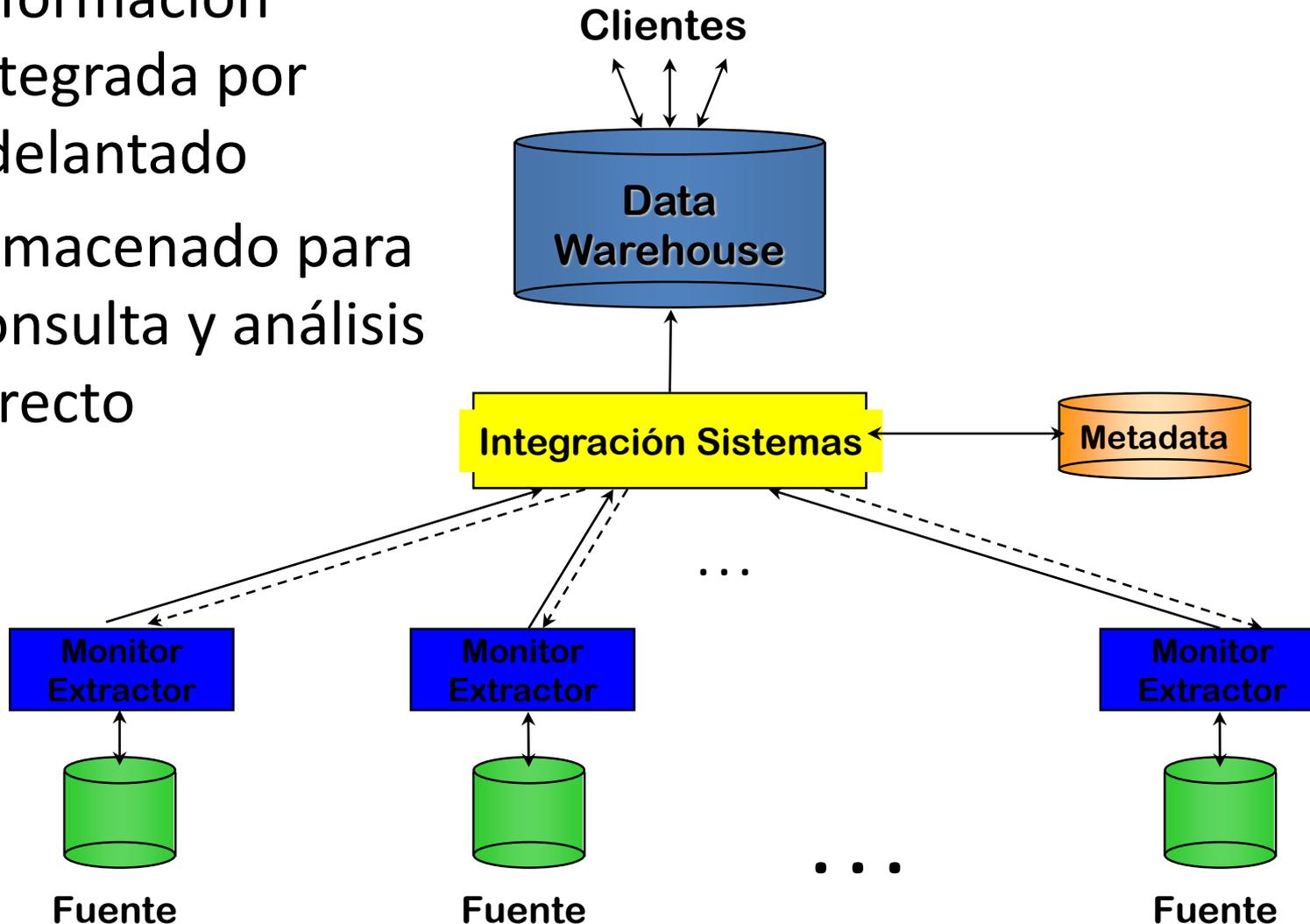


# Problema del Enfoque tradicional

- ◆ El retraso en el procesamiento de consultas
- ◆ Fuentes de información lentas o no disponibles
  - ◆ Filtrado e integración son complejas
  - ◆ Ineficiente y potencialmente costoso para las consultas frecuentes
- ◆ Compite con el procesamiento local en el sitio fuente

# Enfoque Warehousing

- Información integrada por adelantado
- Almacenado para consulta y análisis directo



# Ventaja Enfoque Warehousing

- **Alto rendimiento de la consulta**
  - Pero no necesariamente la información más actualizada
- **No interfiere con el procesamiento local en el sitio origen**
  - Las consultas complejas en warehouse
  - OLTP en las fuentes de información
- **Información copiada en el almacén**
  - Se puede modificar, anotar, resumir, reestructurar, etc.
  - Puede almacenar información histórica
  - Seguridad, sin auditoría
- **Usada en la industria**

# Concepto Data Warehousing

\*Es un **gran almacén de datos** para consultar

\*Es un **repositorio de datos** de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, toma de decisiones.

Se trata, de **un expediente completo de una organización**, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos.

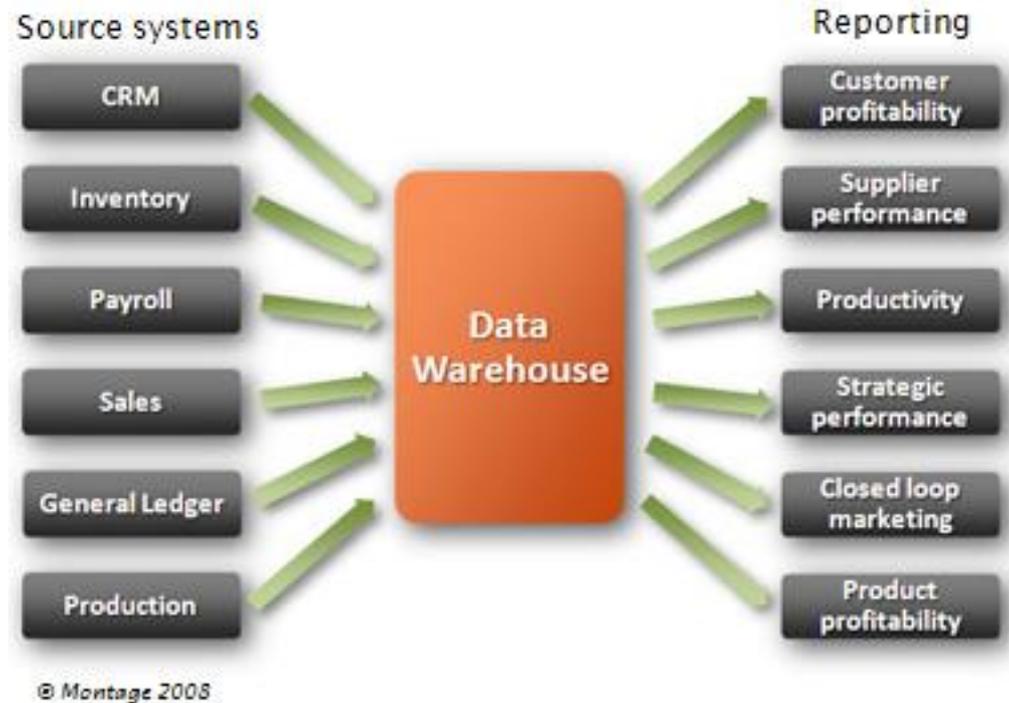
\*Tiene **gran capacidad de almacenamiento**, pues los datos pueden ser de grandes periodos de tiempo.



# Concepto Data Warehousing

Basado en el **concepto de Metadatos** (datos que describen otros datos). Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Cumple el principio de arquitectura fundamental que es **separar los sistemas transaccionales de los informacionales** en dos entornos de manera que el **análisis de los datos existentes no interfiera** con el procesamiento y registro de nuevos datos.



# Concepto Data Warehousing (Bodegas de Datos)

- Datawarehouse (Almacén de datos) se define como una colección de datos que verifican las siguientes propiedades:
  - **Está orientado a un tema**
  - **Datos integrados**
  - **No volátiles**
  - **Variante en el tiempo**

que surgieron como una herramienta de soporte para la toma de decisiones a nivel gerencial

# Orientado hacia temas



Los datos se almacenan y agrupan por temas de interés

Se agrupa por temas orientados a la organización, tales como :

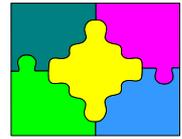
- Clientes
- Productos
- venta

en lugar de las transacciones individuales.

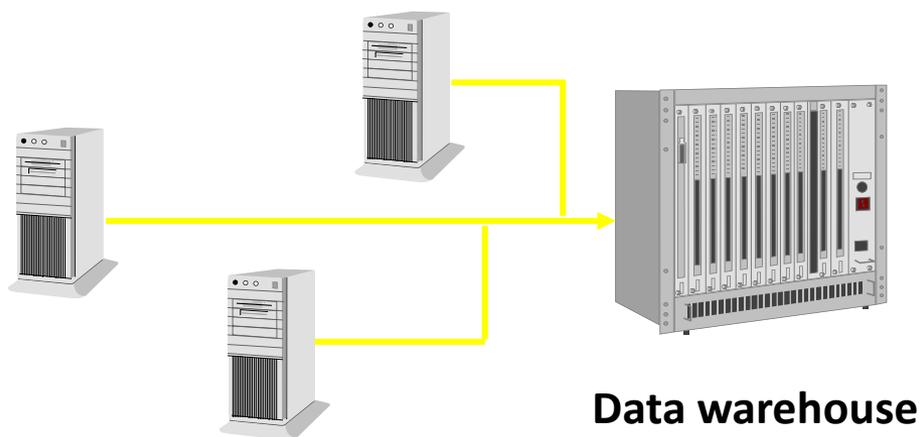
La normalización no es relevante

- **Centrándose en el modelado y análisis de datos** para la toma de decisiones, no en las operaciones diarias o el procesamiento de transacciones.
- Proporcionar una **visión sencilla y concisa en torno a cuestiones temáticas particulares** mediante la exclusión de los datos que no son útiles en el proceso de apoyo a la decisión.

# Integración de Datos



- El almacén de datos **integra datos** que provienen de varias fuentes. Parte de una base de datos operacional y mediante un proceso de **carga de datos** hace el Datawarehouse.
- El proceso de carga es lo más complicado por problemas de codificación, preparación de los datos, ...

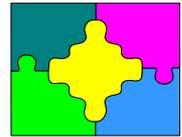


Sistemas Operacionales

Data warehouse

Los datos en el almacén deben ser:

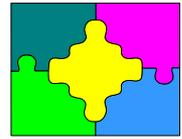
- Limpios
- Validados
- Adecuadamente integrados



# Integración de Datos

- Integración de **múltiples fuentes de datos heterogéneas**,
  - bases de datos relacionales, archivos planos, registros de transacciones en línea
- Se aplican **técnicas de limpieza y de integración de datos**.
- Garantiza la **coherencia** en los convenios de denominación, estructuras de codificación, medidas, etc., entre las diferentes fuentes de datos
  - Por ejemplo, el precio del hotel: moneda, impuestos, desayuno, etc.
- Cuando los datos se mueven a la bodega, **se convierten** al estándar usado.

# Ejemplo Integración de Datos



## Sistema de Cuenta de cheques

Jane Doe (name)  
Female (gender)  
Bounced check #145 on 1/5/95  
Opened account 1994

← *Datos operacionales*

## Sistema de cuentas de ahorro

Jane Doe  
F (gender)  
Opened account 1992

## cliente

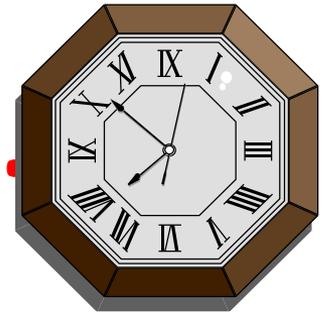
Jane Doe  
Female  
Bounced check #145  
Married  
Owns 25 Shares Exxon  
Customer since 1992

## Sistema de inversiones de clientes

Jane Doe  
Owns 25 Shares Exxon  
Opened account 1995

↑ *data warehouse*

# ... variante en el tiempo ...



- Todos los datos en el almacén de datos tienen una **marca de tiempo** en el momento de entrada en el almacén o cuando se usan en el almacén.
- Esta grabación cronológica de datos ofrece posibilidades **históricas y análisis de tendencias**.
- Normalmente, en las BD operativas **los datos se sobrescribe**, ya que los valores anteriores no son de interés.

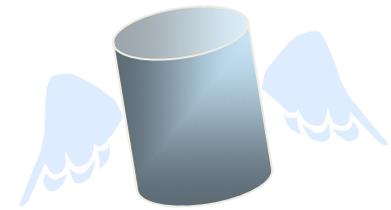
TIEMPO

# id\_tiempo

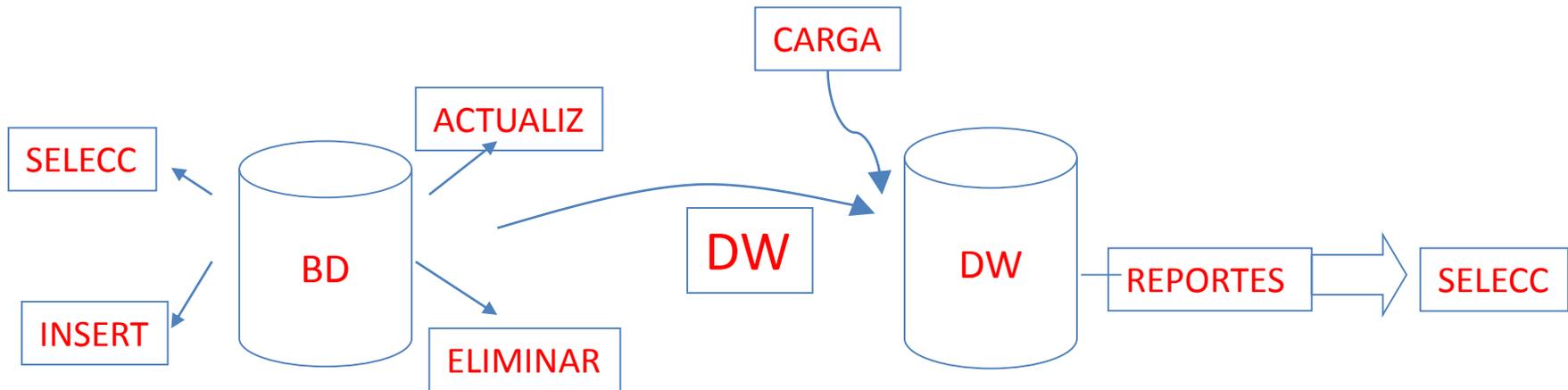
\* periodo

porque se hacen cargas de datos continuamente, cuando los datos van cambiando, se actualizan los históricos y se guardan en ficheros temporales.

# No volátiles



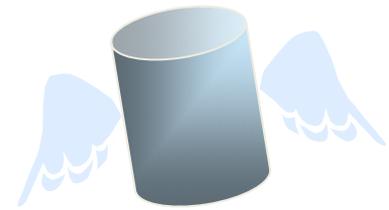
Son estables, una vez almacenados los datos **no se modifican.**



Datos actúan como un recurso estable para informes coherentes y análisis comparativo.

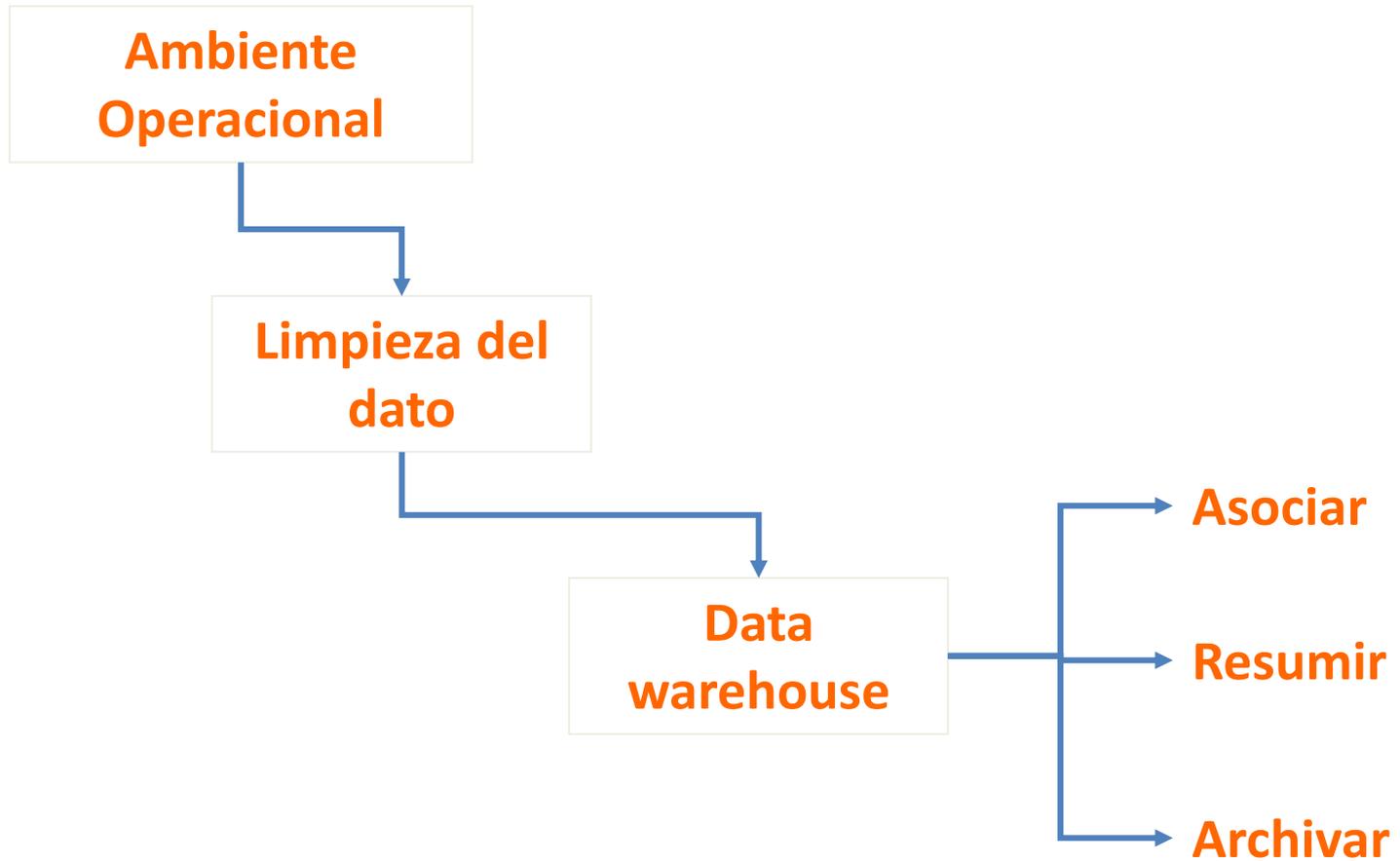
Por el contrario, los datos operativos se actualizan (insertar, eliminar, modificar, etc.)

# No volátiles



- **Físicamente separado** de los datos del entorno operacional.
- **Actualización de los datos no se produce** en el entorno de almacén de datos.
- **No requiere el procesamiento de transacciones**, ni mecanismos de recuperación y control de concurrencia
  - Requiere sólo dos operaciones en el Acceso a los datos:  
**carga inicial de datos y el acceso a los datos.**

# Flujo del Dato



# Funcionamiento de Data warehouse

## Tres funciones esenciales:

1. Recopilación de los datos desde Bases de Datos
2. Gestión de los datos en el almacenamiento
3. Análisis de datos para toma de decisiones

Sistemas de  
Soporte a  
Decisiones  
(DSS)



herramientas para  
hacer consultas o  
informes



Sistemas de  
información  
ejecutiva (EIS)



# Elementos que integran un almacén de datos

## Metadatos

"datos acerca de los datos".

Su función es recoger la descripción de la estructura del almacén de datos:

**Tablas**

**Columnas en tablas**

**Jerarquías y Dimensiones de datos**

**Relaciones entre tablas**

**Entidades y Relaciones**

# Elementos que integran un almacén de datos

## Metadato también cuenta con

- **Meta-datos de los datos Operacionales**
  - Historia de los datos (migraciones, transformaciones),
  - Estado de los datos (activo, archivados, o limpiado),
  - Información de seguimiento (estadísticas de uso de almacén, informes de errores, trazas de auditoría)
- **Los algoritmos utilizados para resumir**
- **El mapeo del entorno operativo al almacén de datos**
- **Los datos relacionados con el rendimiento del sistema**
- ...

# Elementos que integran un almacén de datos

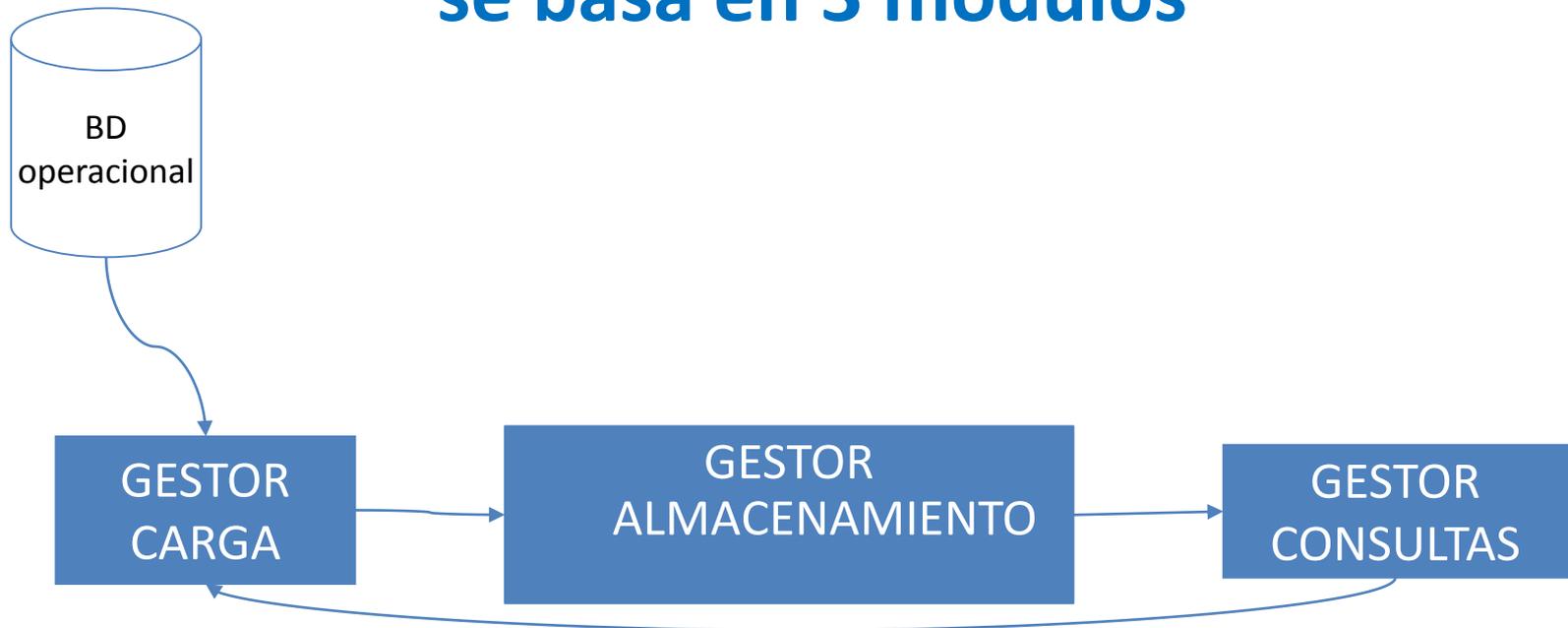
## Middleware

- Ofrece servicios a aplicaciones distribuidas que requieren usar el DW.
- Su función es asegurar la conectividad entre todos los componentes de la arquitectura de un almacén de datos (datawarehouse) para brindar servicios.

## Funciones ETL, OLAP

# ARQUITECTURA PLANA DW

se basa en 3 módulos



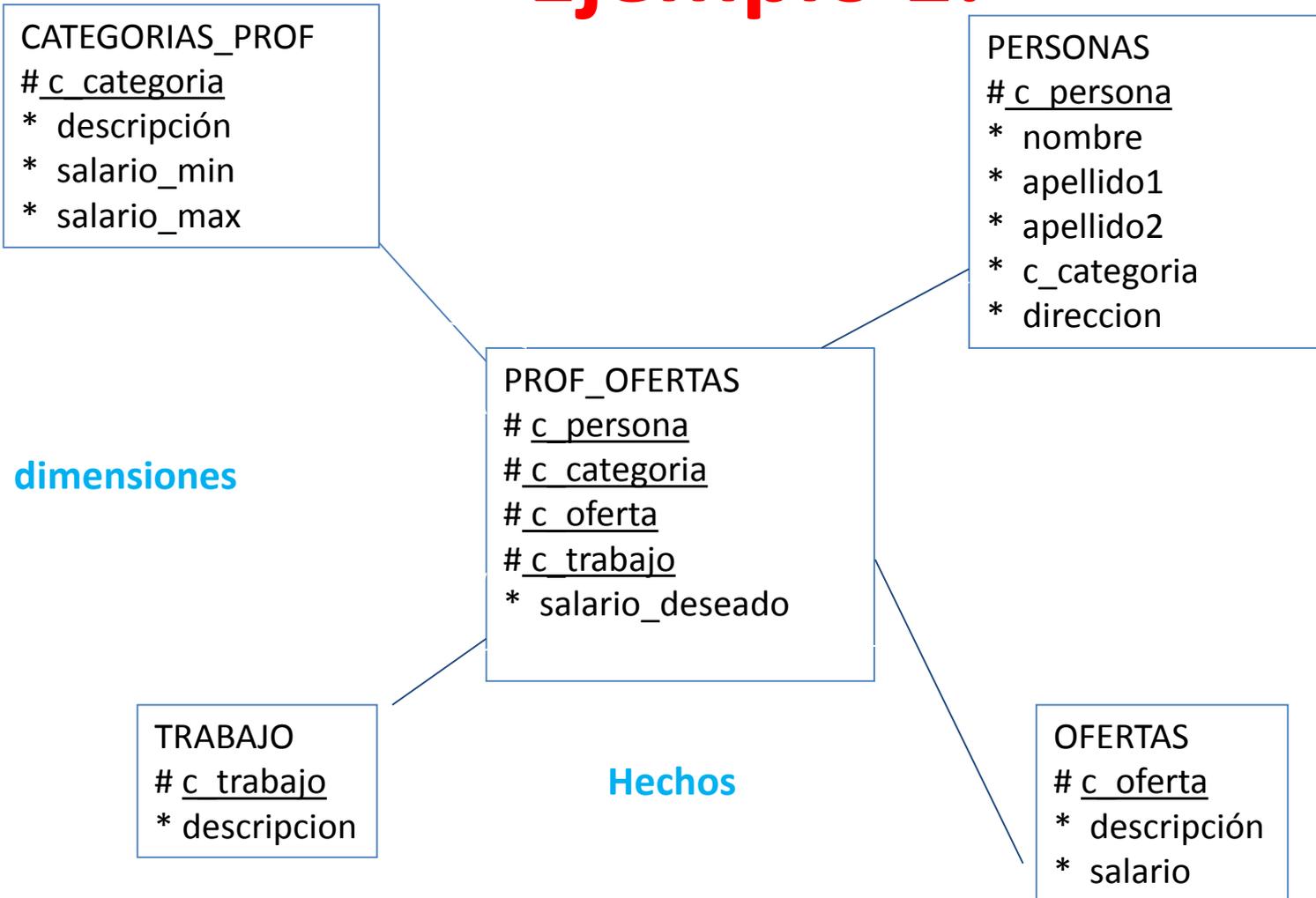
# GESTOR DE CARGA

- Permite hacer la carga.
- Dificultades:
  - La integración de los datos
  - La elección del momento de la carga
  - Minimizar el tiempo de carga
  - Buen diccionario de datos o METADATA (para evitar cometer errores en la carga)
  - Diseño de procedimientos SQL

# GESTOR DE ALMACENAMIENTO

- Se encarga del almacenamiento, de la estructura,....
- Existe una tabla llamada FACT (Hechos) y unas tablas llamadas dimensiones o tablas dimensionales.
  - Entre la tabla FACT y las tablas dimensionales suele haber relaciones 1:N
  - Este modelo tiene forma de estrella, por eso se denomina MODELO ESTRELLA

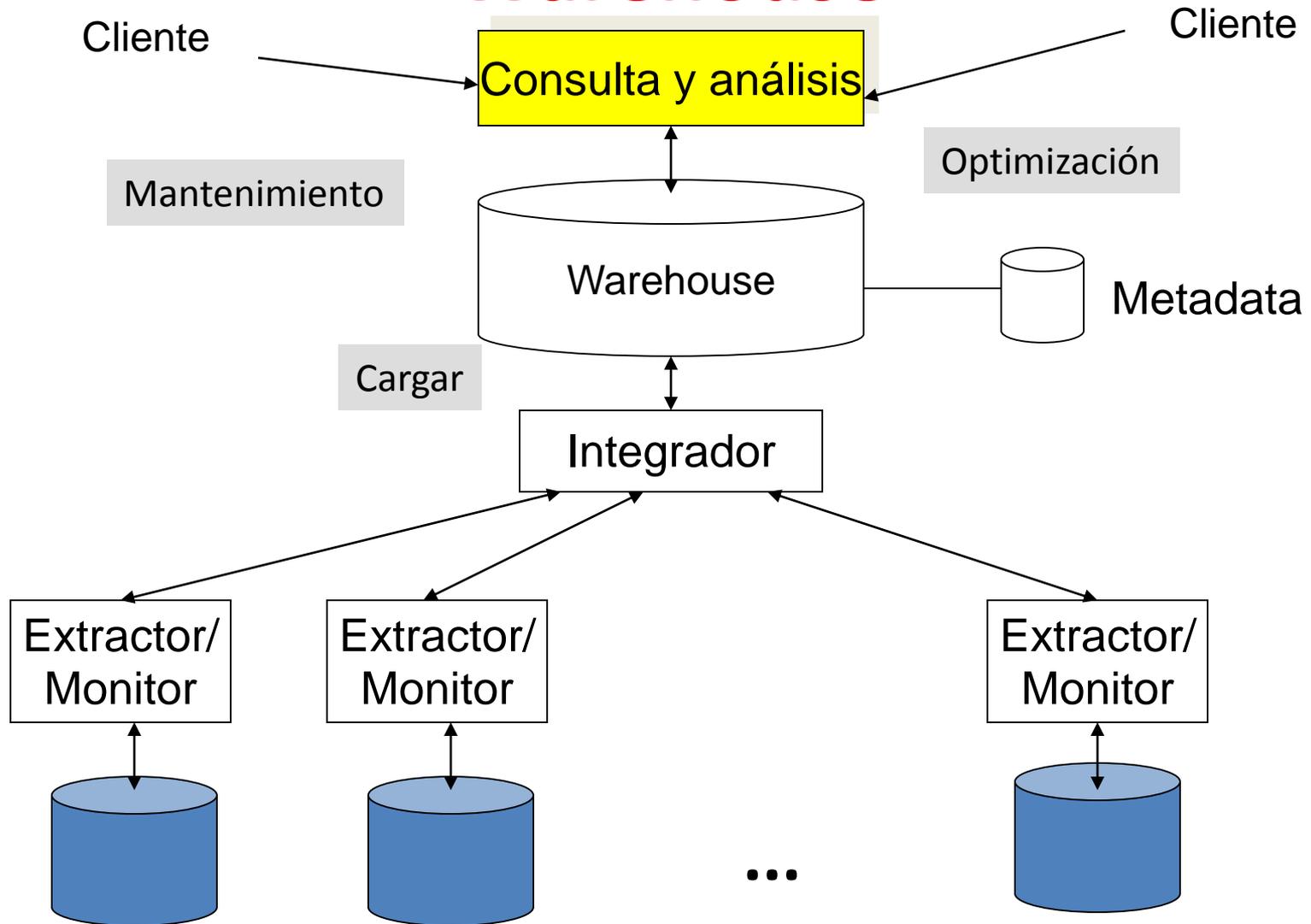
# Ejemplo 1:



# GESTOR DE CONSULTAS

- Las consultas se hacen sobre la tabla FACT.
- También define perfiles de consultas, pues ellas son diferentes dependiendo del usuario y sus necesidades.

# Arquitectura genérica Data Warehouse



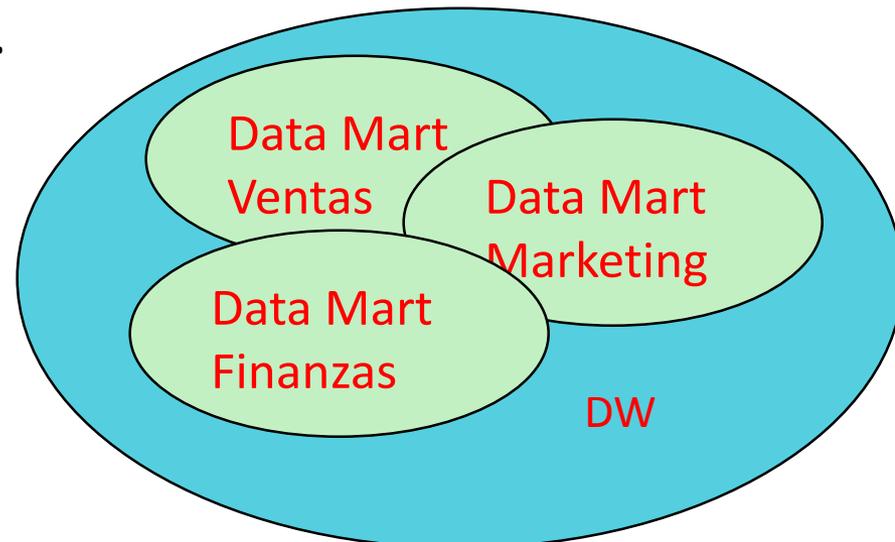
# Data Mart

Las grandes cantidades de datos en el Data Warehouse a veces se subdividen en unidades lógicas más pequeñas (**data marts**)

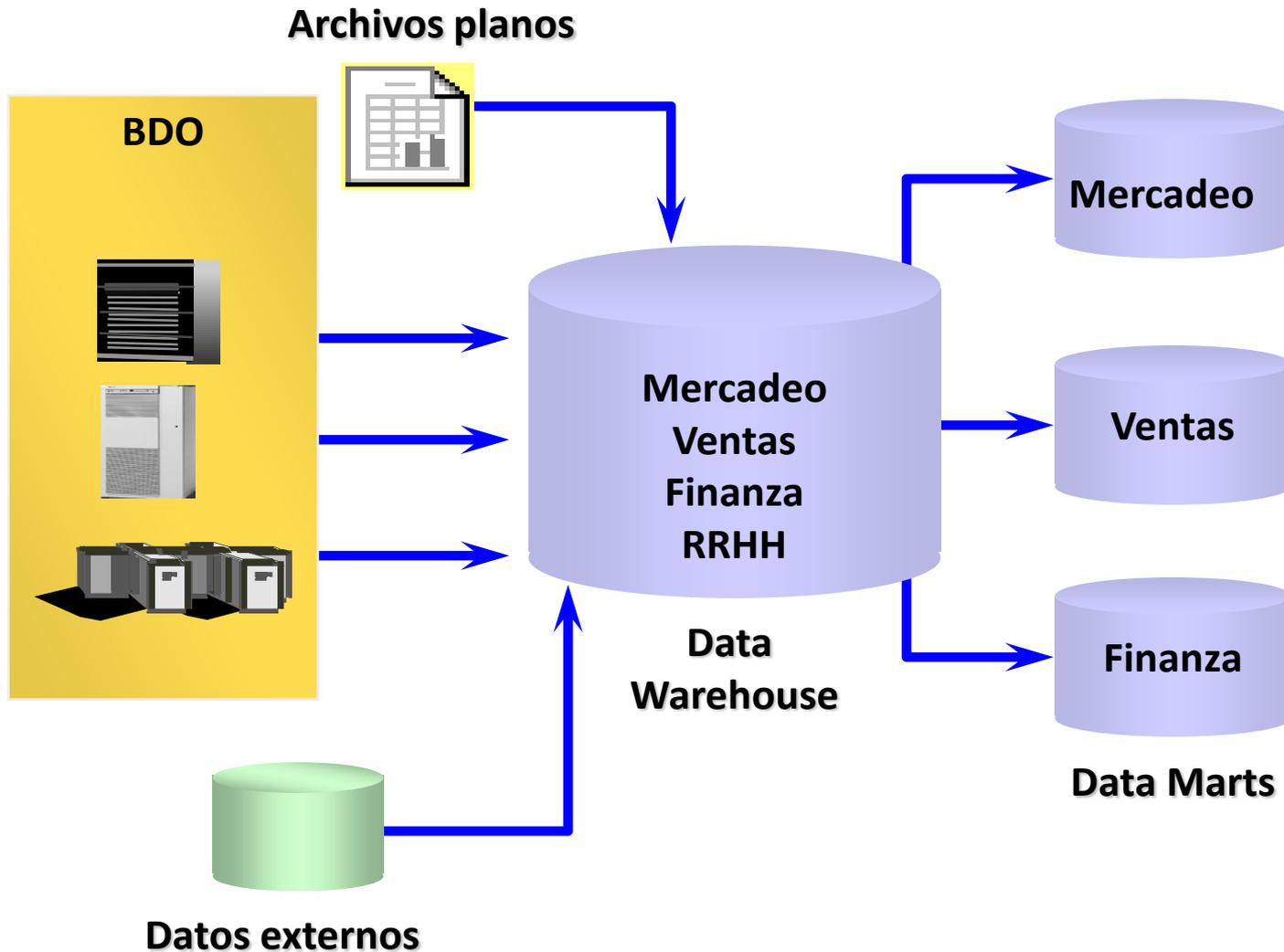


# Data Mart

- Una versión reducida de menor costo que el data warehouse, diseñado para una **unidad de negocios estratégica o departamento**.
- Entre las **características** de un data mart destacan:
  - Usuarios limitados.
  - Área específica.
  - Tiene un propósito específico.
  - Tiene una función de apoyo.

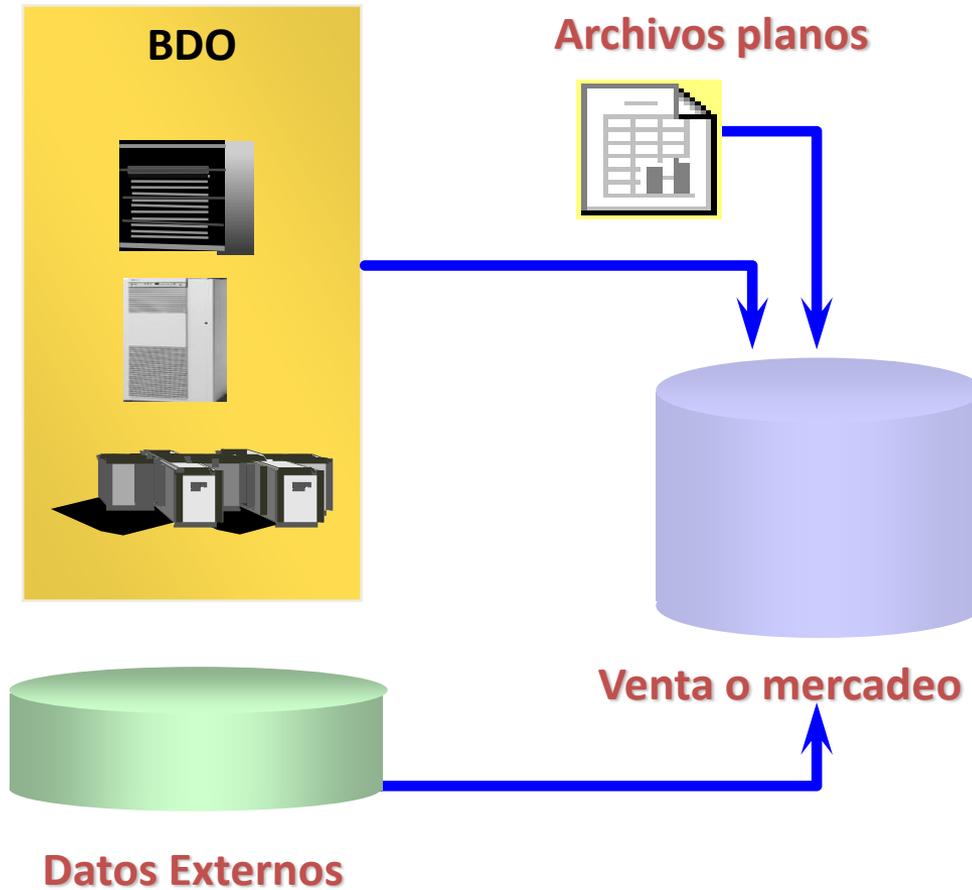


# Tipo Data Mart



Dependiente de DW

# Tipo Data Mart



Independiente de DW

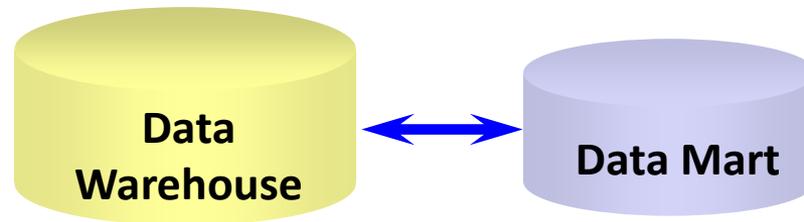
# Razones para la creación de un Data Mart

- Para dar a los usuarios un **acceso más flexible** a los datos que necesitan para analizar con más frecuencia.
- Para proporcionar datos en una forma que coincide con el **punto de vista** colectivo de un grupo de usuarios
- Para mejorar el **tiempo de respuesta** del usuario final.
- Los usuarios potenciales de un data mart están **claramente definidos**

# Razones para la creación de un Data Mart

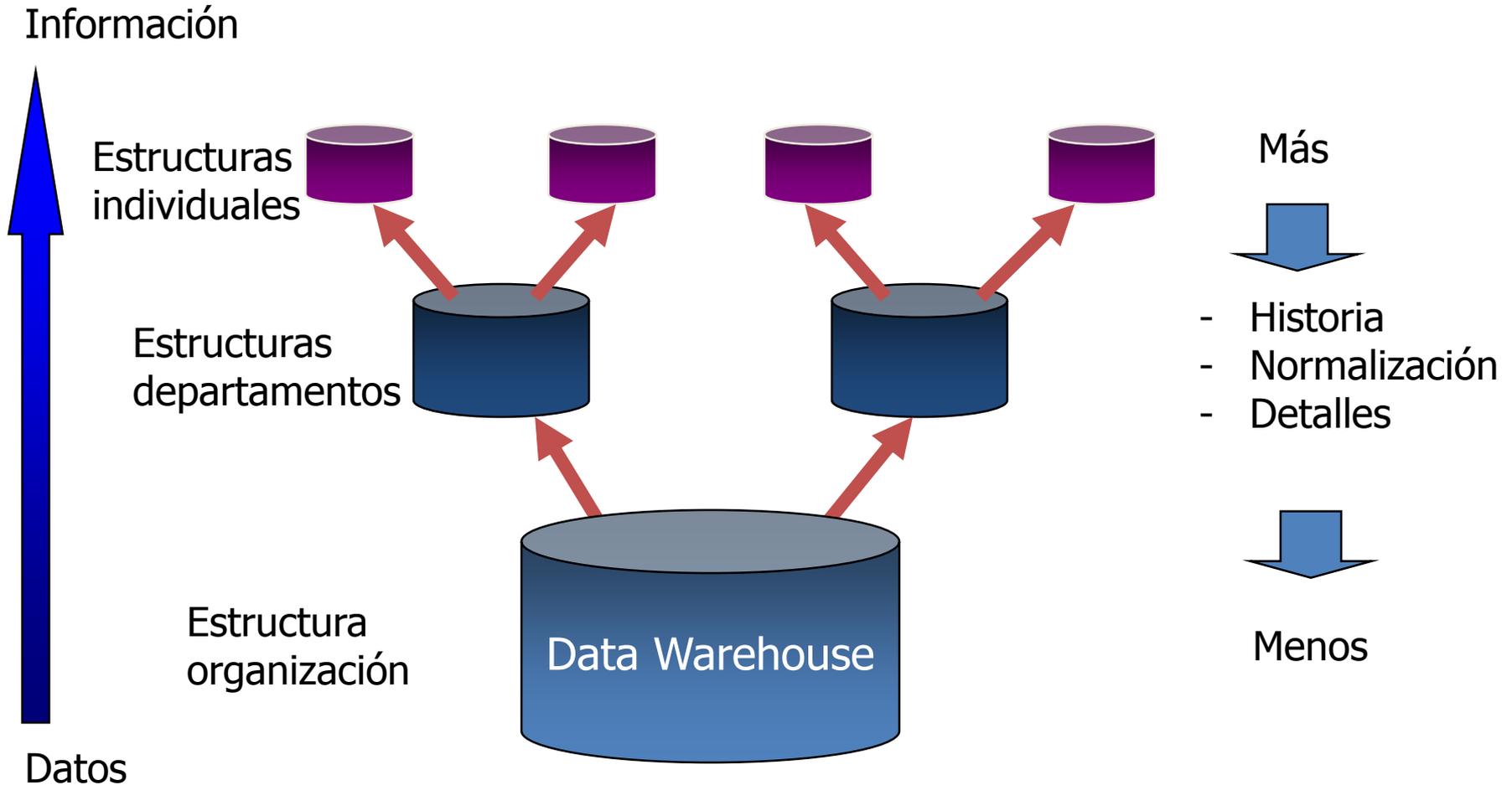
- Proporcionar datos estructurados según las **exigencias de las herramientas** de acceso de los usuarios finales.
- La construcción de un data mart **es más sencilla** en comparación con el establecimiento de un data warehouse corporativo.
- El **costo de la implementación** de data mart es **mucho menor** que el requerido para un data warehouse

# Data Warehouses Vs Data Marts

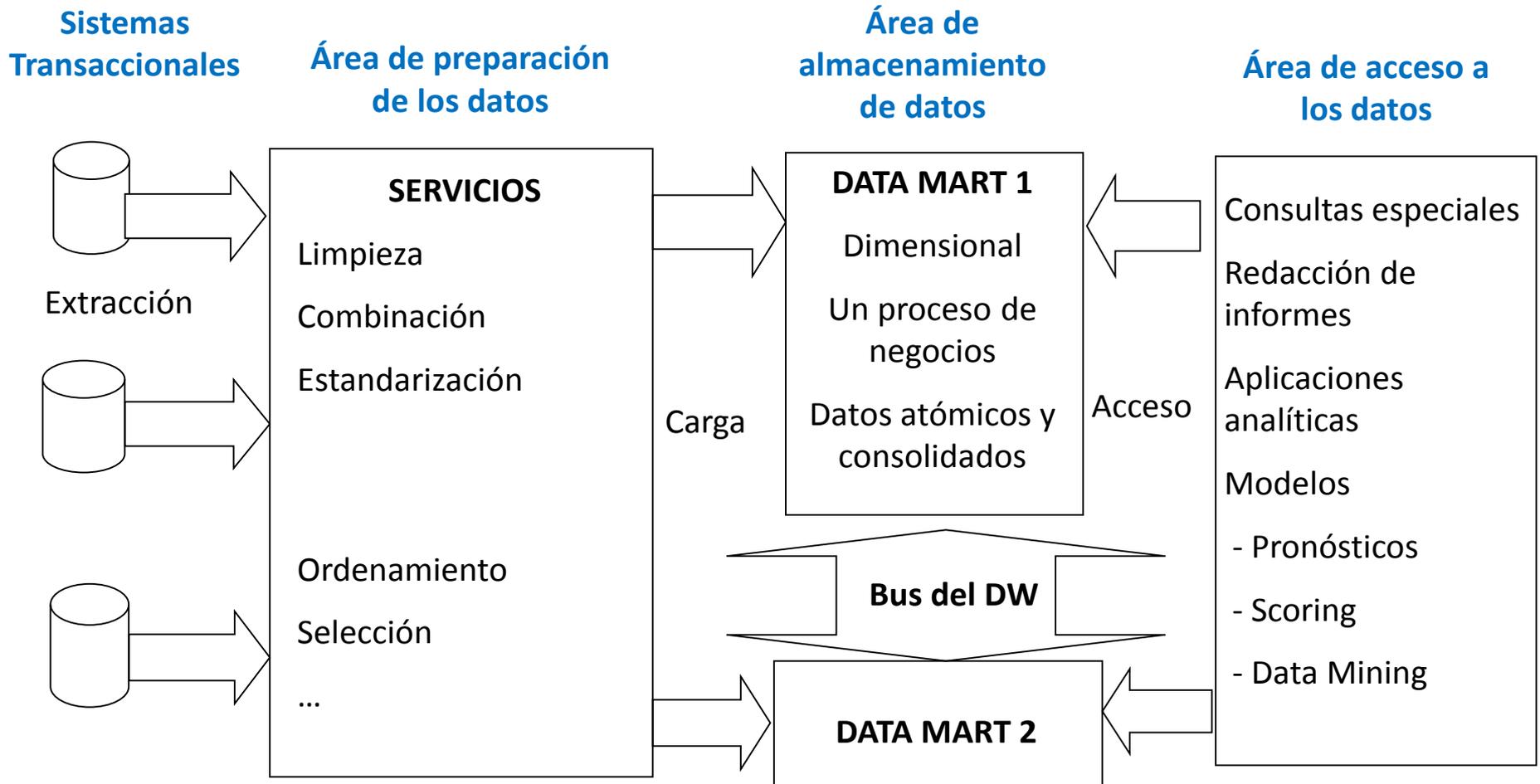


	<b>Data Warehouse</b>	<b>Data Mart</b>
<b>Alcance</b>	Empresa	Departamento
<b>Sujetos</b>	Múltiple	Grupo
<b>Fuente de datos</b>	Muchas	Pocas
<b>Tamaño (típico)</b>	100 GB a > 1 TB	<100 GB
<b>Tiempo de implementación</b>	Meses o años	Meses

# Desde Data Warehouse a Data Marts



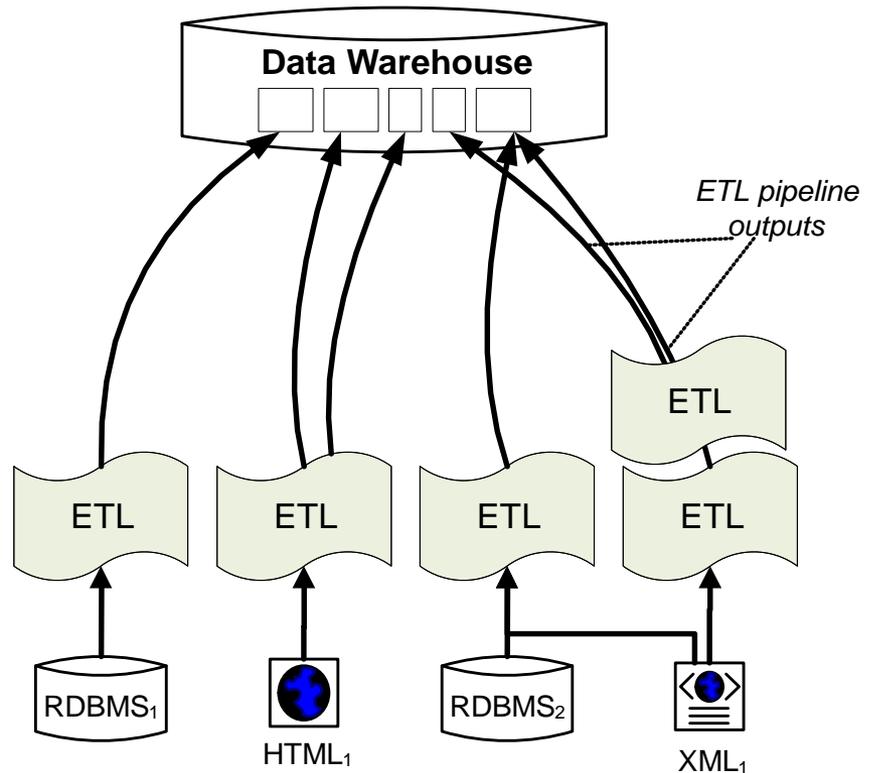
# Arquitectura genérica Data Warehouse con Data Mart



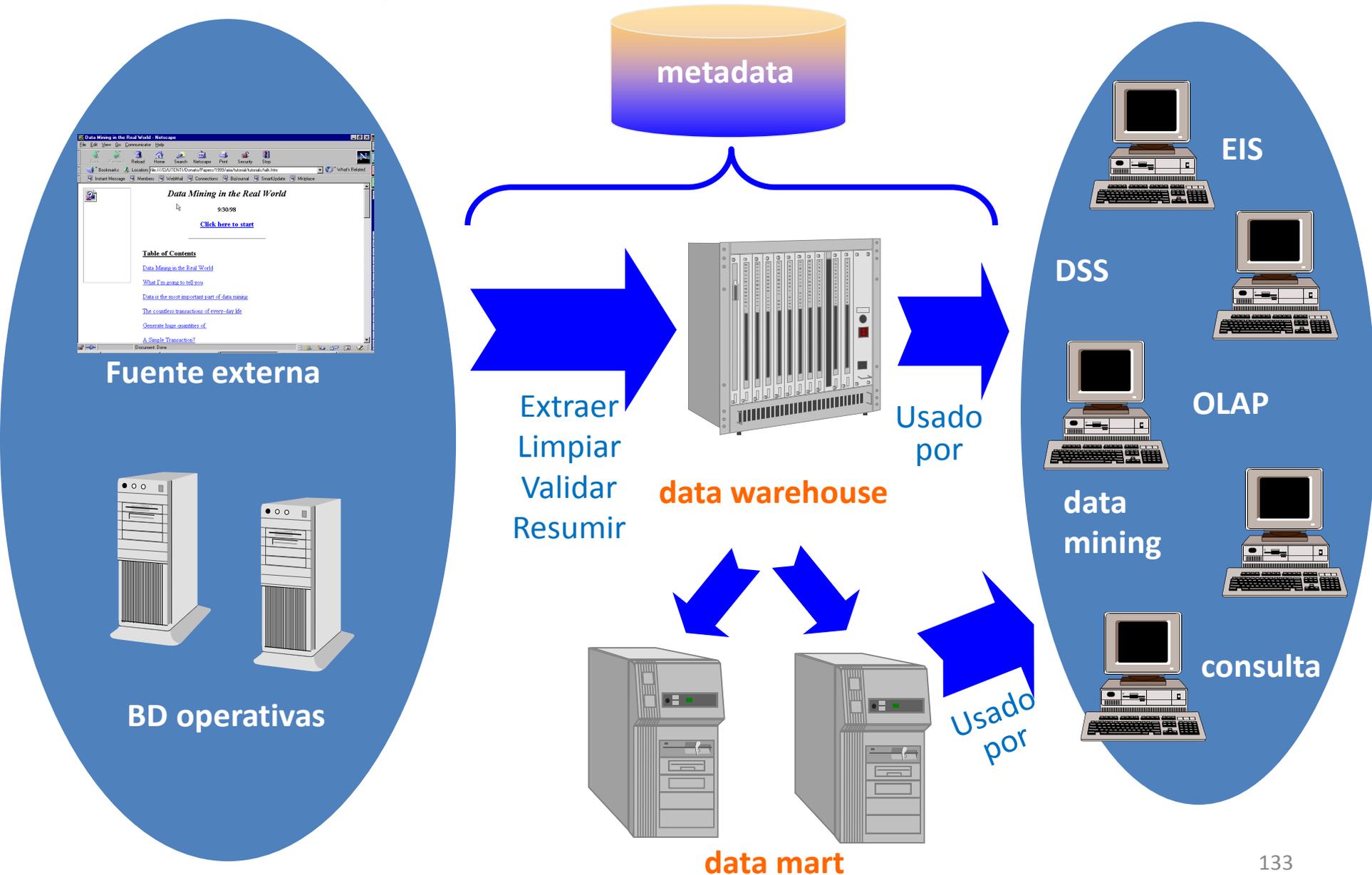
# Arquitectura Data Warehouse

## primer modelo

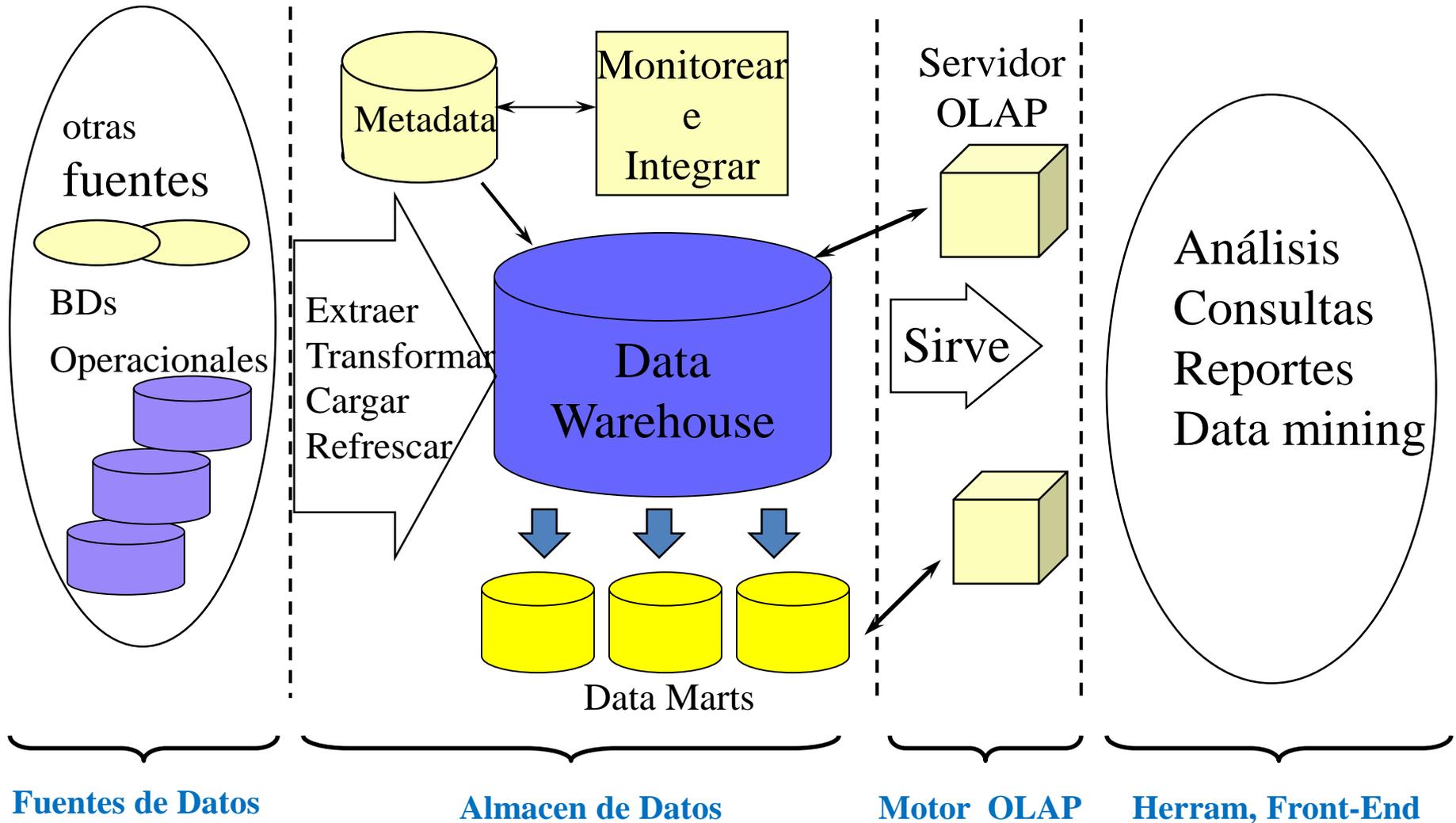
- **En la parte superior una base de datos centralizada**
  - Generalmente configurado para consultas y agregar – no transacciones
  - Muchos índices, vistas , etc.
- **Los datos se cargan y actualizan periódicamente a través de ETL (Extract / Transform / Load)**



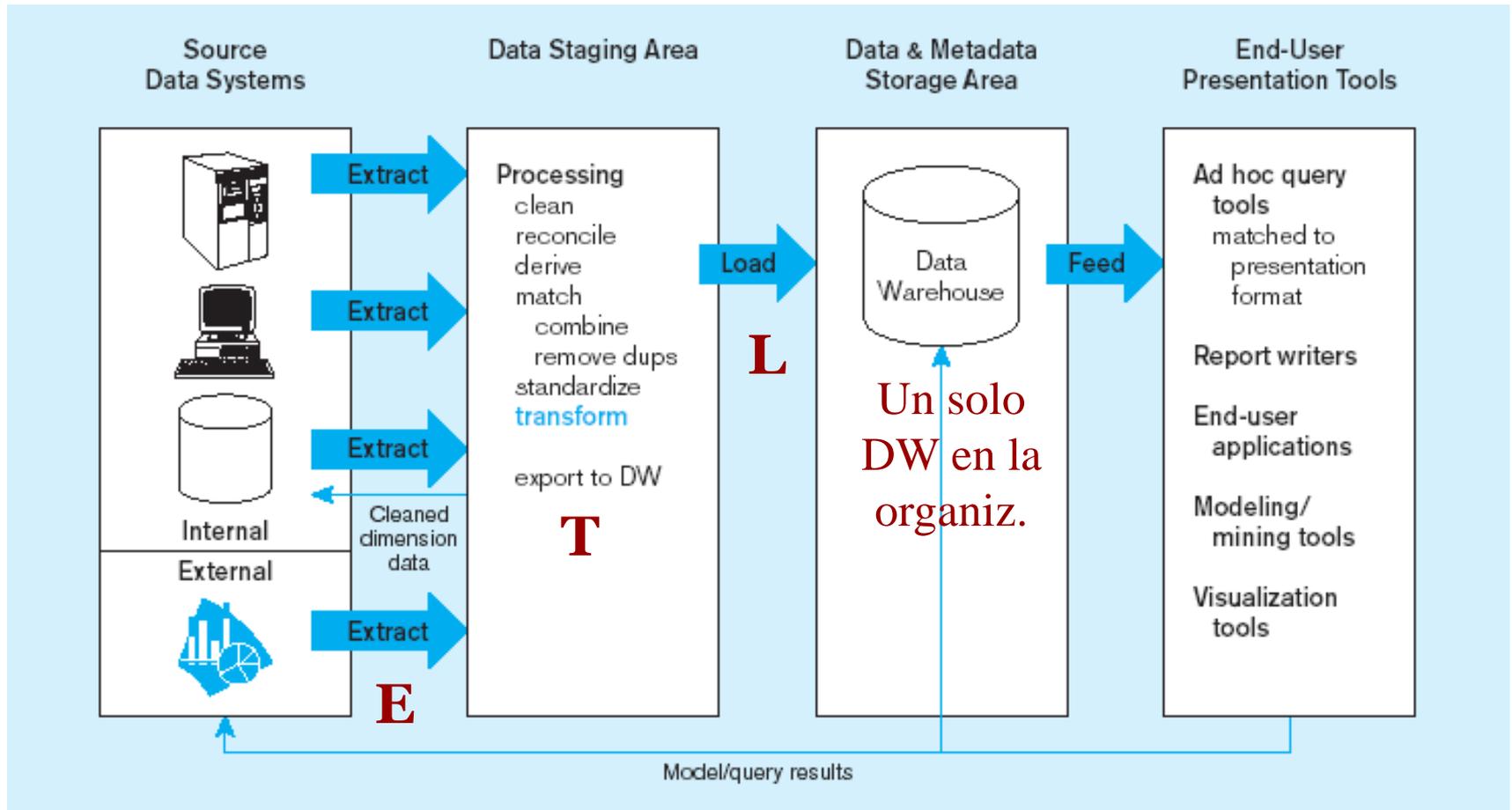
# Arquitectura varios niveles



# Arquitectura varios niveles



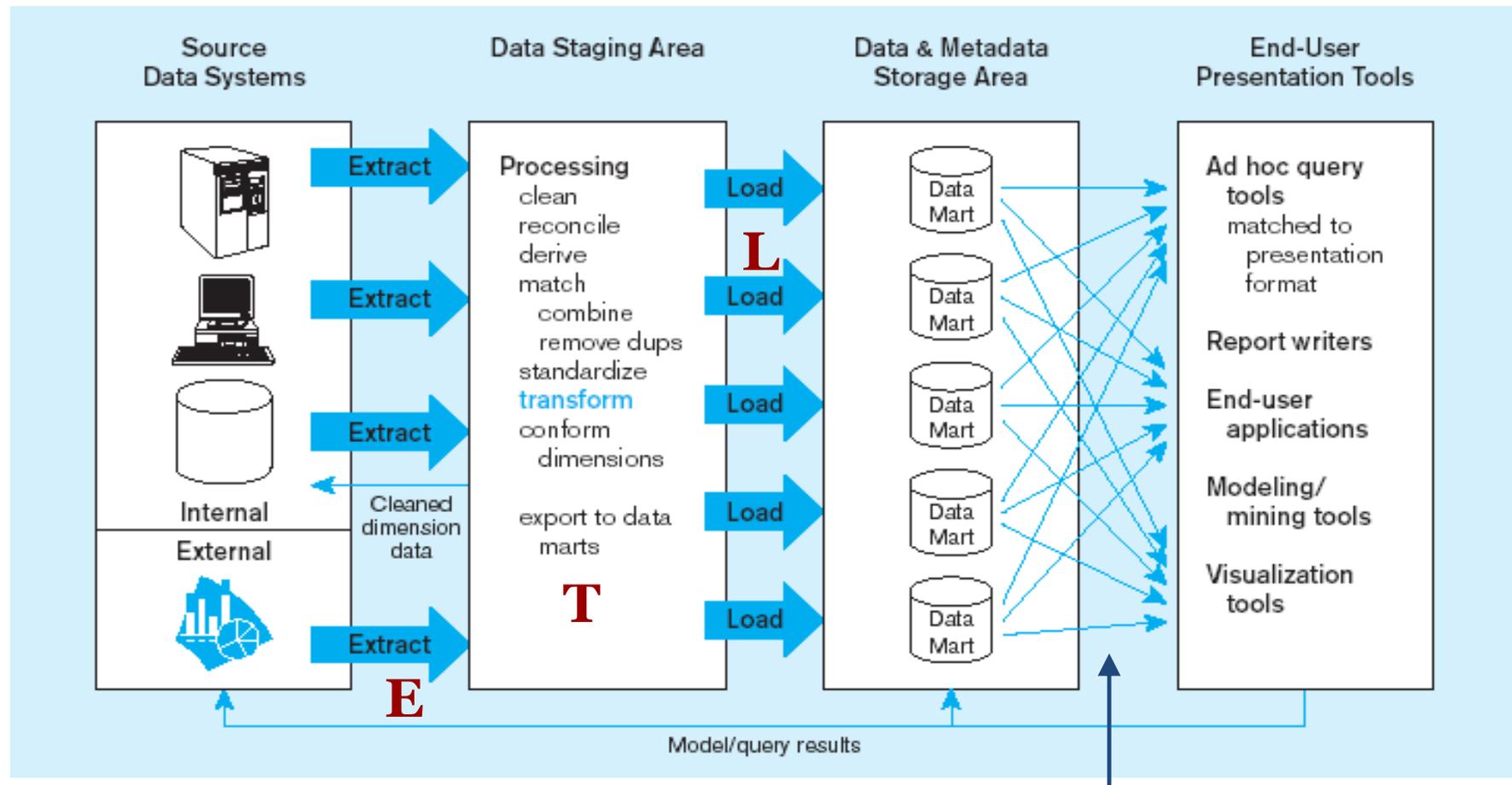
# Genérica arquitectura de niveles de DW



**Centralizado**

Extracción de datos periódica  
Datos no están completamente al día

# Arquitectura independientes Data mart

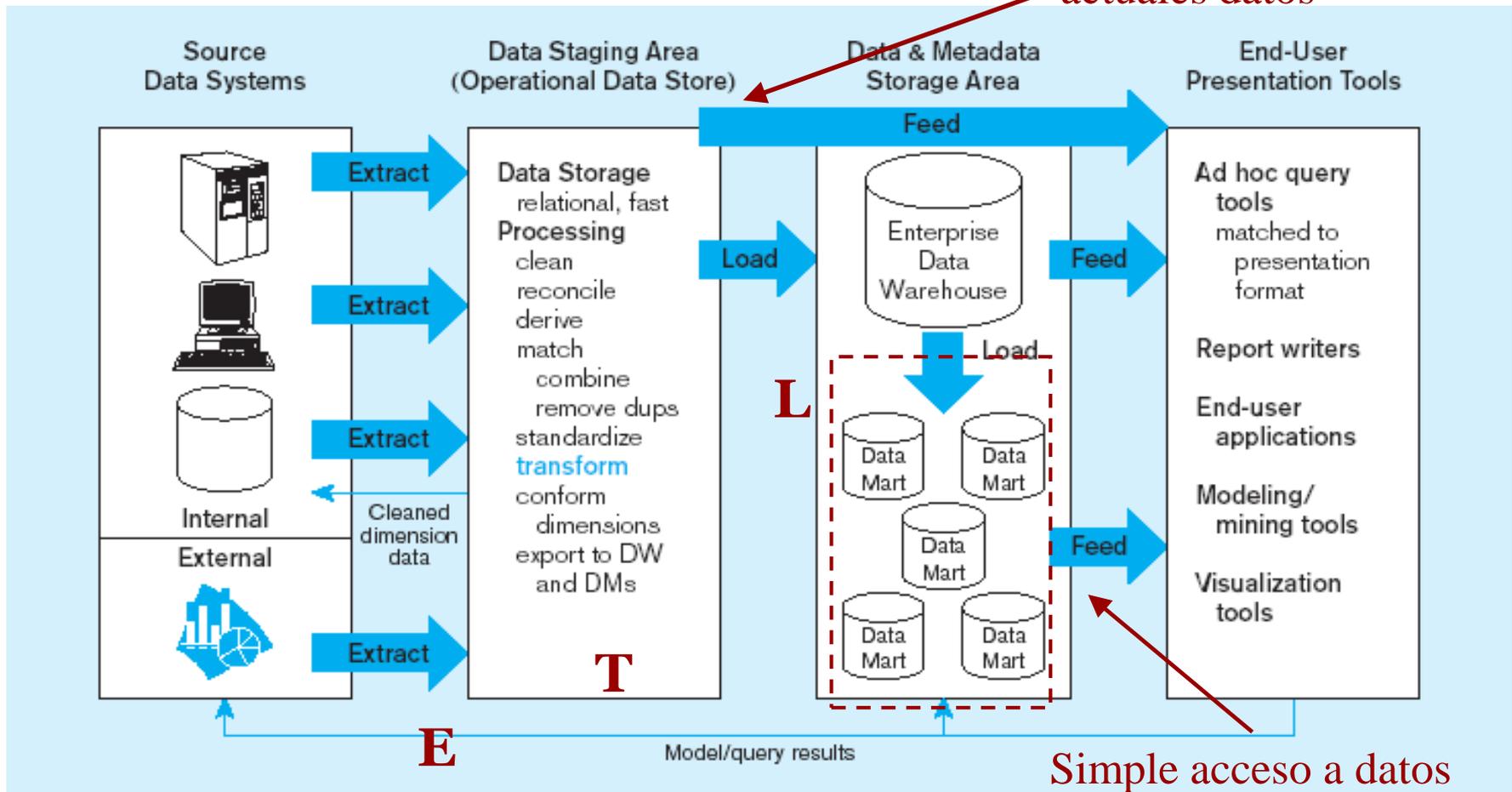


Separado ETL para c/data mart

Acceso datos por  *multiples*  data marts

# Arquitectura Híbrida

Se pueden obtener  
actuales datos

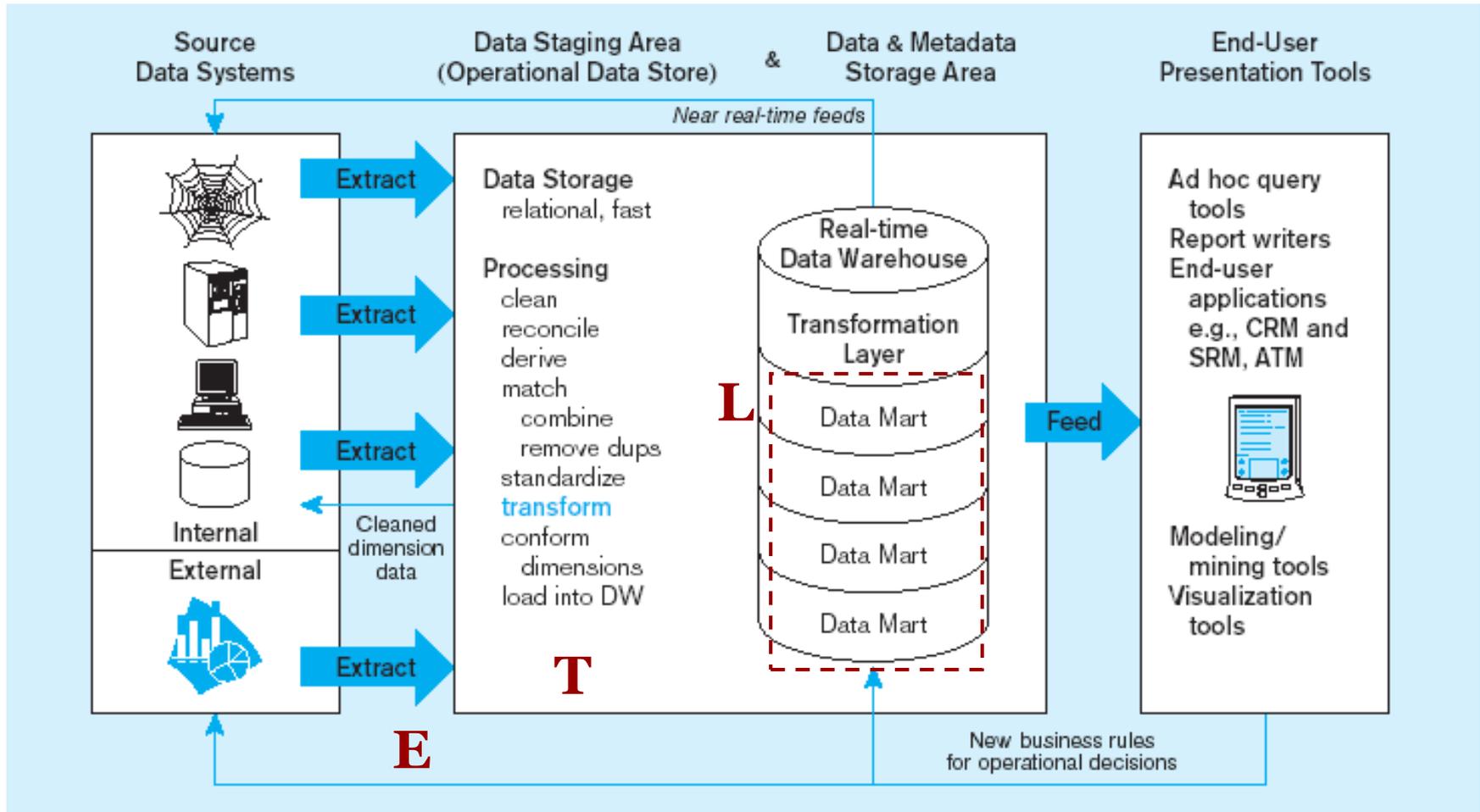


Un solo ETL para  
*enterprise data warehouse*  
(EDW)

Data marts *dependientes*  
del EDW

## Federado

# Data mart Lógico y DW-RT



ETL RT para  
*Data Warehouse*

Data marts no separado de BDs, vistas lógicas de los datos de data warehouse  
→ Fácil crear nuevos data marts

# Diferencia Data Warehousing y BD

## BD

- Una **colección estructurada** de registros o datos

## Data Warehouse

- Una **colección lógica** de información, recogida de muchos diferentes bases de datos operacionales, que apoya las actividades de análisis organizacional y las tareas de toma de decisiones

# Formato

## Modelos BD

relacional  
jerárquica  
red  
orientada a objetos  
Etc.

## Data Warehouse

- Tradicionalmente heterogénea => integración
- Se construyen envolturas/mediadores en la parte superior
- Cuando una consulta se plantea a un sitio, un **meta-diccionario** se usa para traducir la consulta en consultas apropiadas para el sitio
- Los resultados se integran en un conjunto de respuestas globales Filtradas

La información de fuentes heterogéneas se integra con antelación y se almacena en bodegas para consultas directas y análisis

## **BD**

Diseñado para  
tiempos  
operaciones en  
tiempo real

En su mayoría  
permanente  
actualizaciones

Muchas pequeñas  
transacciones

Mb - Gb de datos

## **Data Warehouse**

Diseñado para el análisis de  
la organización por  
categorías y atributos

Sobre todo se leen

Las consultas son largas y  
complejas

Gb - Tb de datos

## BD

- Instantánea
- Los datos en bruto
- Cientos de usuarios (por ejemplo, los usuarios de una organización)

## Data Warehouse

- Histórica
- Resumidos y conciliados los datos
- Decenas de usuarios (por ejemplo, los gerentes, los analistas)

# Data Warehouse vs. SMDB

- **OLTP (on-line transaction processing)**
  - Major tarea de SMDB relacionales tradicionales
  - Las operaciones del día a día: compras, inventario, financiera, nómina, contabilidad, etc.
- **OLAP (on-line analytical processing)**
  - Majores tareas de data warehouse
  - Análisis de datos y toma de decisiones

## Características diferentes (OLTP vs. OLAP):

- **Orientación de la BD:** usuario vs organización
- **Contenido de los datos:** actual, detallada vs histórica, consolidada
- **Diseño de base de datos:** aplicación + ER vs estrella + por tema
- **Vista:** actual, local vs integrada, , evolutiva
- **Patrones de acceso:** actualizar vs consultas complejas

# Las diferencias de diseño

BDs

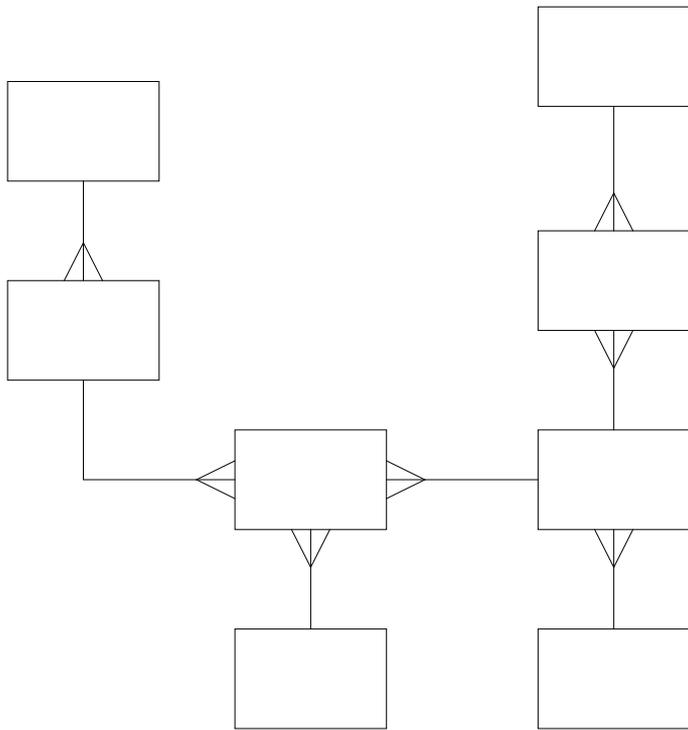
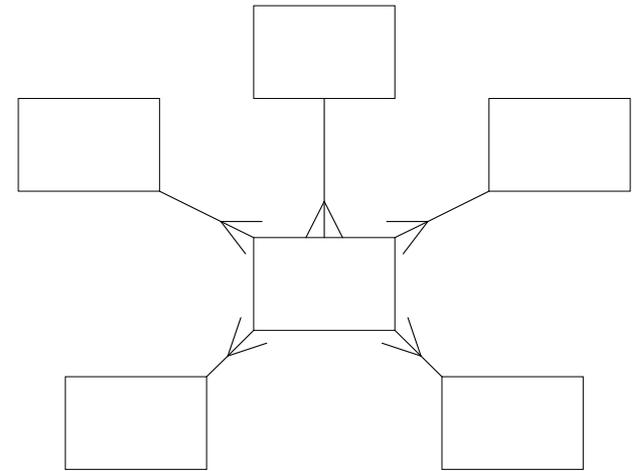


Diagrama ER

Data Warehouse



Estrella

# RESUMEN DIFERENCIAS

## BD OPERACIONAL

- Datos operacionales
- Orientado a aplicaciones
- Datos Actuales
- Datos Detallados
- Datos en continuo cambio

## DATAWAREHOUSE

- Datos de negocio
- Orientado al sujeto
- Actuales + Histórico
- Datos Resumidos
- Datos Estables

# Por qué Data Warehouse?

## Diferentes funciones y datos:

- **Datos que faltan:** apoyo a las decisiones requiere datos históricos que BDs operacionales no suelen tener
- **Consolidación de datos:** Se requiere de la consolidación (agregación, resumen) de los datos de fuentes heterogéneas
- **Calidad de los datos:** las diferentes fuentes suelen utilizar representaciones de datos inconsistentes, códigos y formatos que deben ser conciliados, etc.