

# Data Warehouse

# AGENDA

01

## **Introducción**

Estadísticas, valor de la información, datawarehouse: definiciones e importancia.

02

## **Datamart y Datawarehouse**

Definiciones, implementación, arquitectura referencial, framework.

03

## **Arquitectura**

Capa simple, dos capas, tres capas.

04

## **Datawarehouse distribuidos**

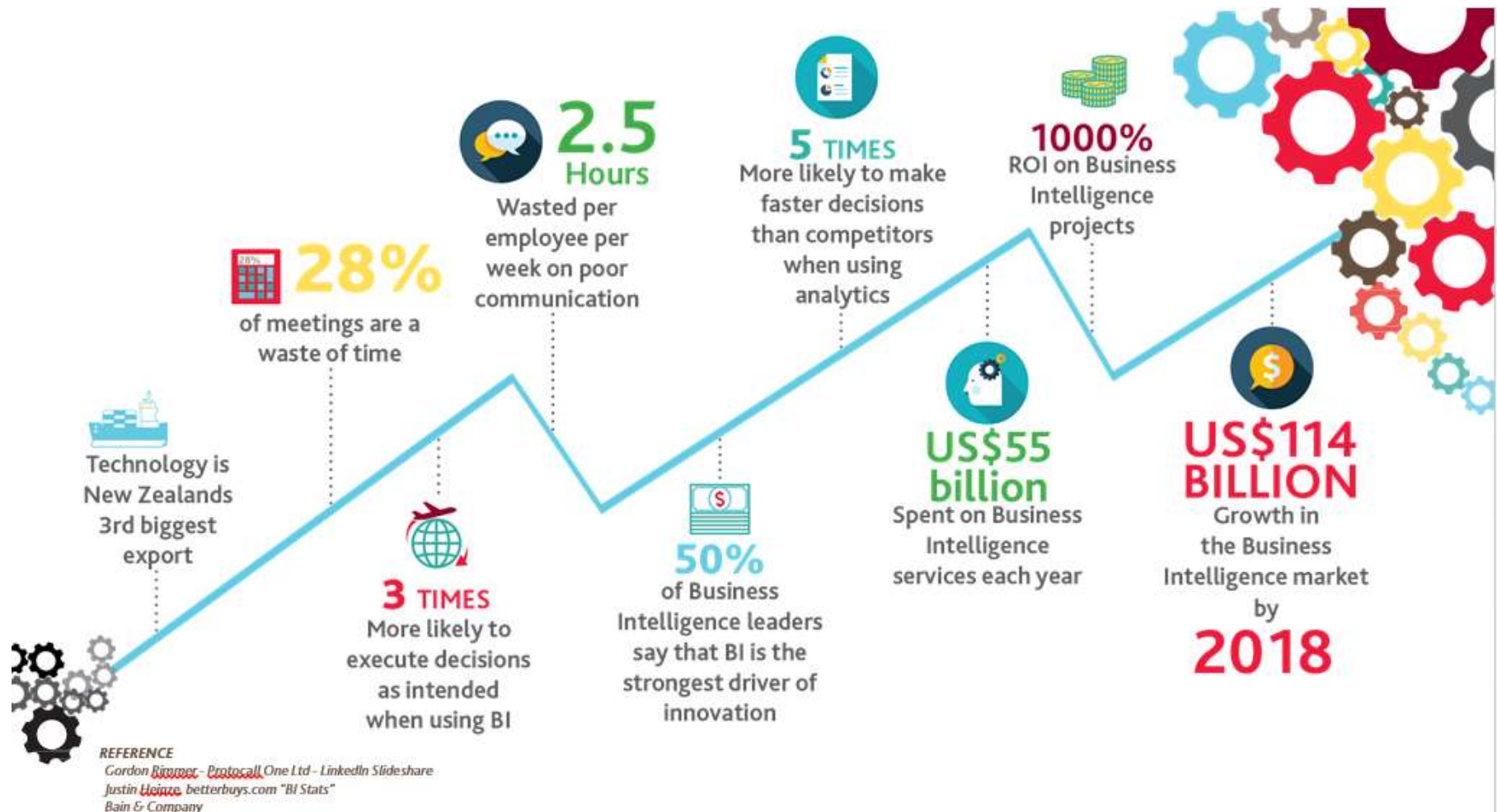
Definición, DW Local, DW Global, entorno del DW.

05

## **Otros conceptos**

Metadatos, calidad de datos, beneficios de un datawarehouse.

# Algunos números...

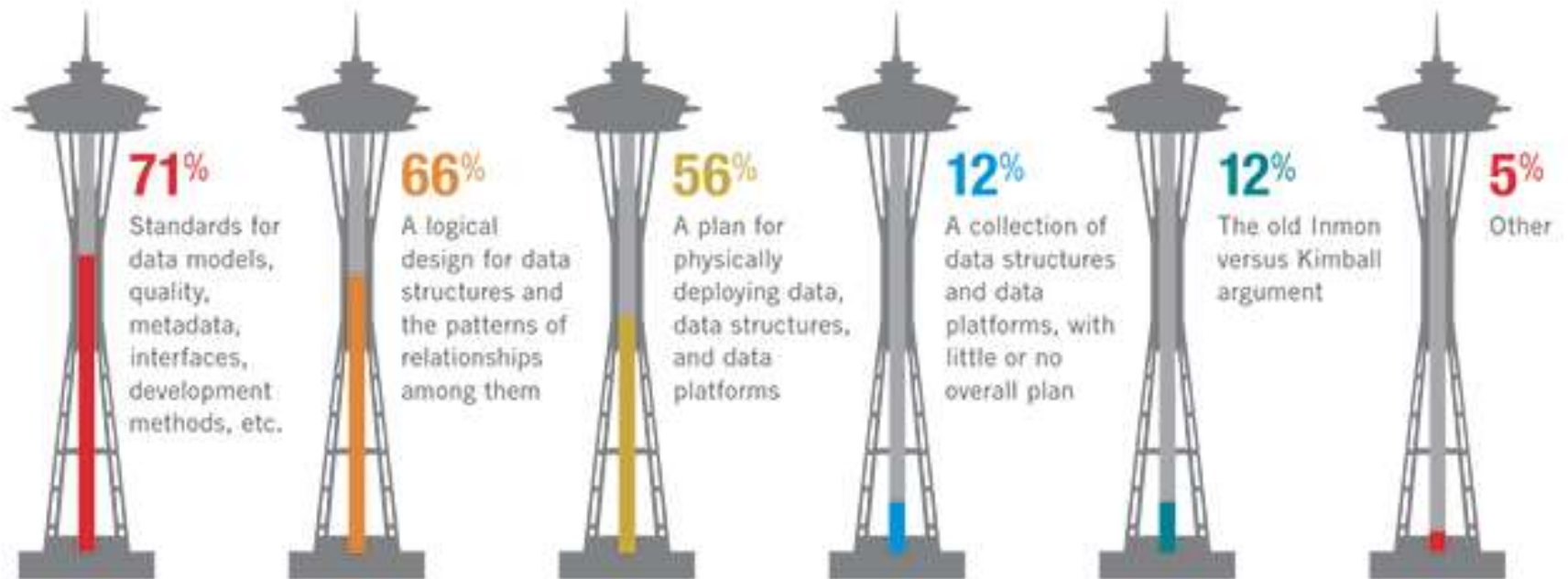


Fuente: <https://www.bdo.nz/en-nz/insights/advisory/newsletters/business-intelligence-intelligent-business-busin>

# Algunos números...

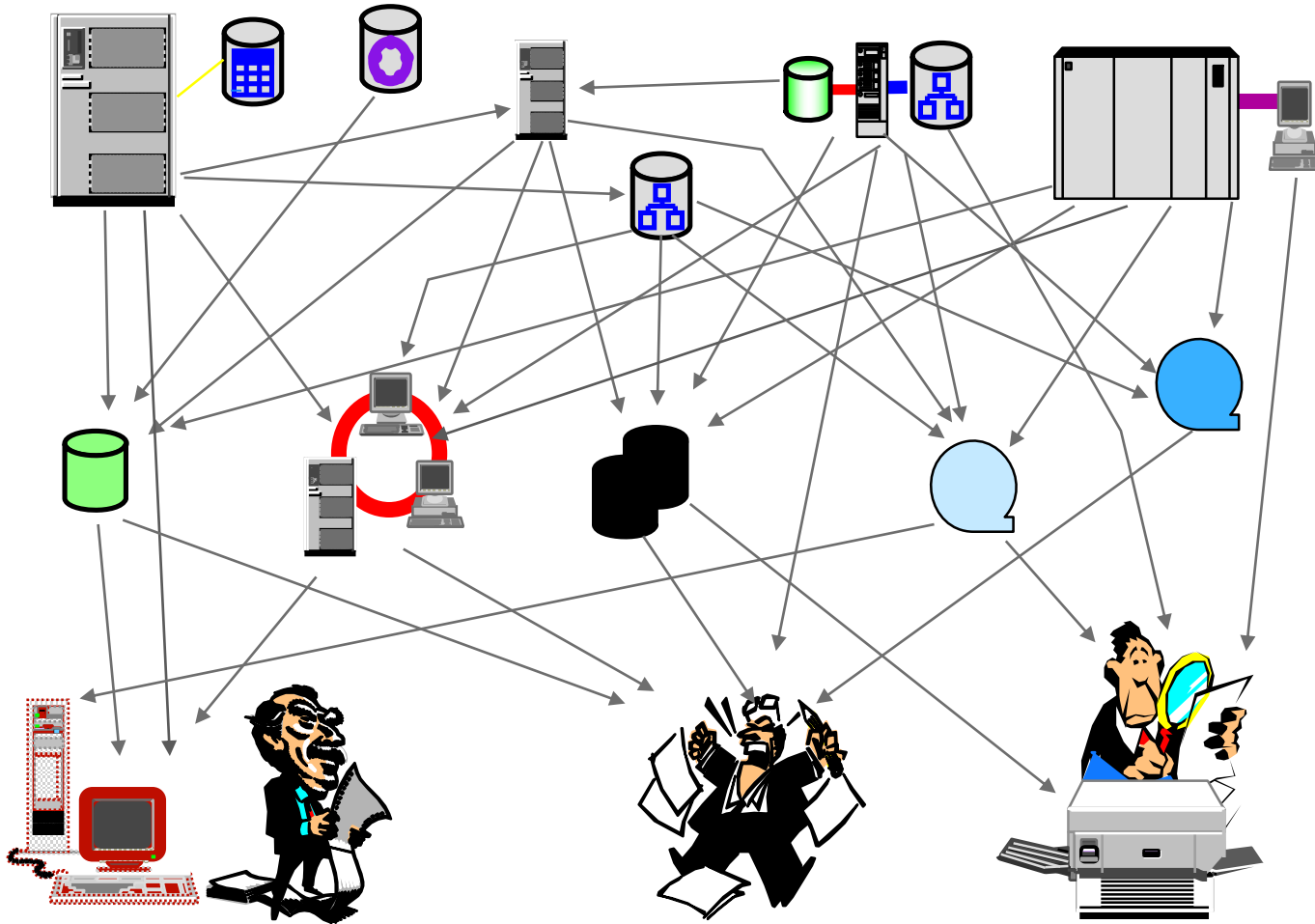
## DATA WAREHOUSE ARCHITECTURES COMBINE LOGICAL AND PHYSICAL COMPONENTS

What do you think DW architecture is? (Respondents could select multiple answers.)

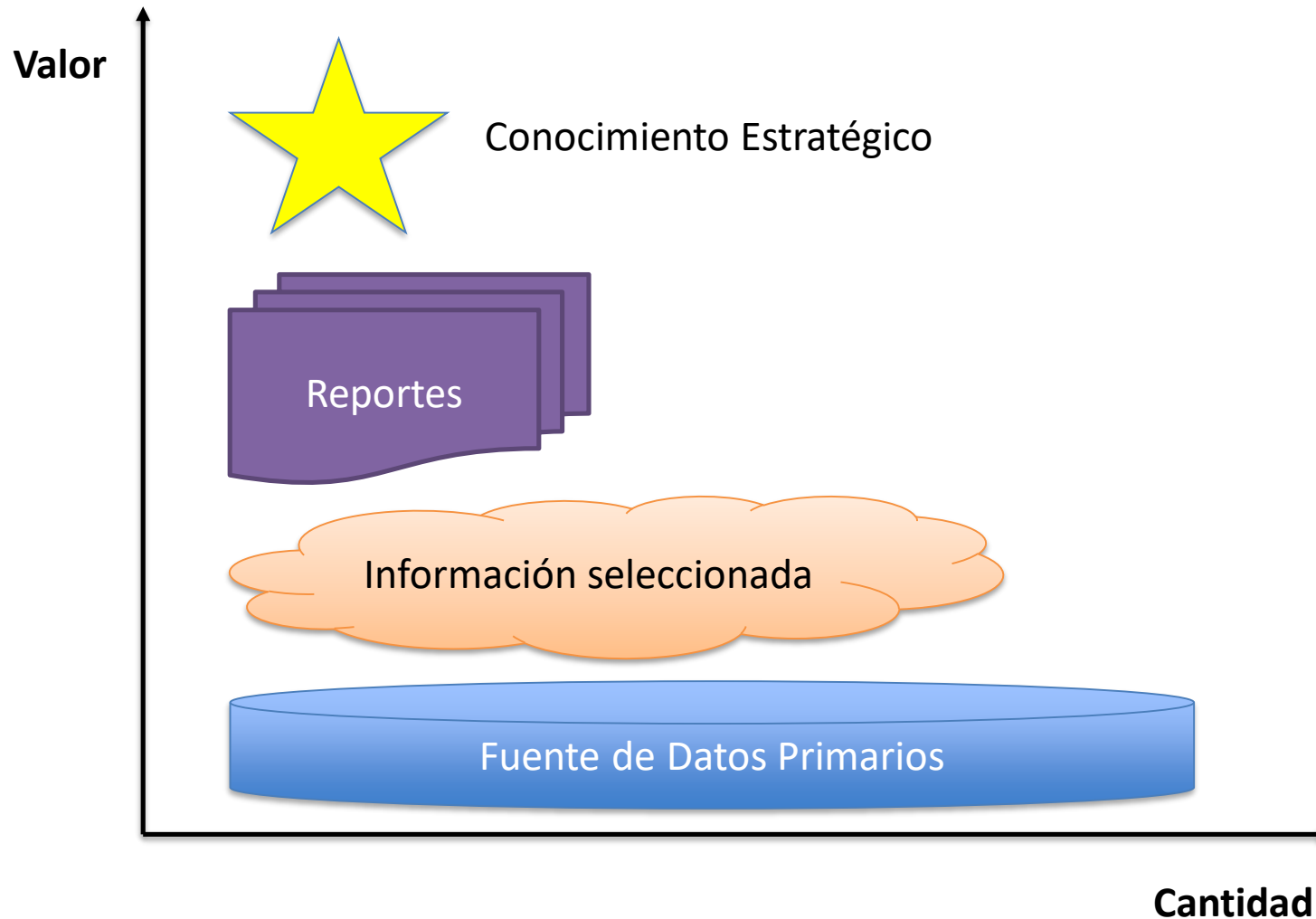


Fuente: <https://www.bdo.nz/en-nz/insights/advisory/newsletters/business-intelligence-intelligent-business-busin>

# Diversidad de Fuente de Datos



# Valor de la Información



# Conceptos



## Datawarehouse

Área de almacenamiento de información como Soporte para toma de decisiones.

- Almacena data recolectada de sistemas operacionales diversos y externos.
- Integra la data en un modelo de negocio.
- Permite el análisis de la data.
- Entrega información a las personas de toma de decisiones a lo largo de la Organización.



## Datamart

Áreas de almacenamientos de **grupos de trabajo** o **departamentales** que son pequeños en tamaño y especializados en funciones.

- Creada para mejorar la performance de acceso/análisis.
- Organizada por temas: Información del mismo evento o tema relacionado.
- Variante en el tiempo: Los cambios en la data son auditables.
- No volátil: Información permanente
- Integrada: Amplia y Consistente

# Importancia Data Warehouse

What are the most important business benefits your organization has obtained from becoming more data-driven?



**58%**

More efficient and effective operations



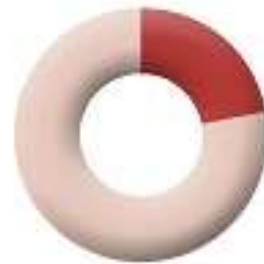
**42%**

Higher customer satisfaction and loyalty



**32%**

Better information consistency



**22%**

More successful at product and service development



**20%**

Can sense and respond faster

Fuente: <https://tdwi.org/research/2018/02/bi-all-ppm-all-best-practices-report-what-it-takes-to-be-data-driven-infographic.aspx>



# Importancia Data Warehouse

**Asegura  
consistencia**



**Toma mejores  
decisiones**



**Mejora la línea  
de Negocio**

Integración de bases de  
datos heterogéneas y  
***control de calidad de  
datos.***



Agrupamiento y des  
agrupamiento de datos  
en forma interactiva.

Análisis del problema en  
términos de dimensiones.



Ejecución de consultas  
complejas no predefinidas  
visualizando el resultado en  
forma grafica

**Funcionalidades  
de un DW**



# Conceptos asociados



# Sistemas en una empresa

## Sistemas Operacionales

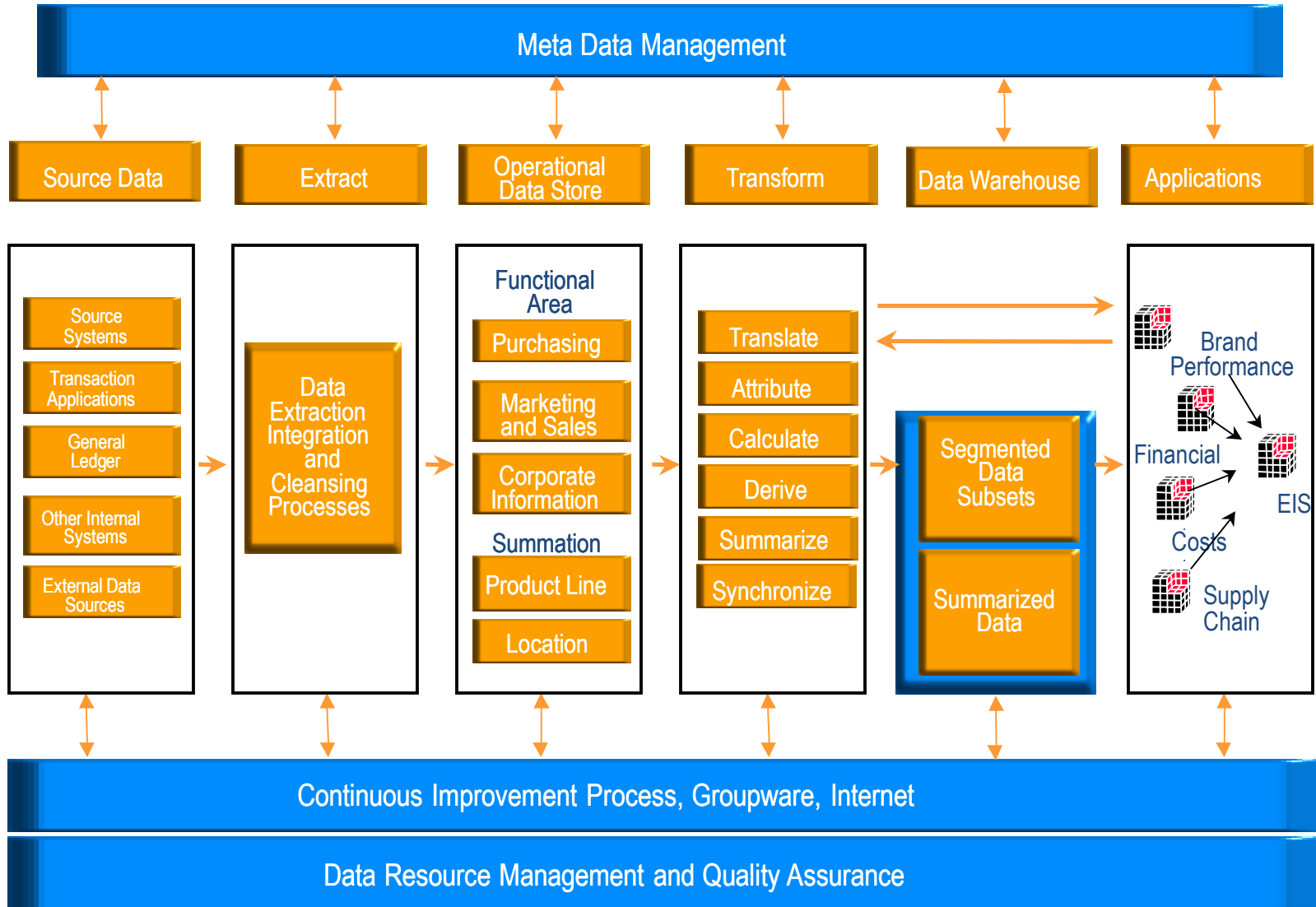
- ✓ Ejecuta el negocio
- ✓ Empleado por Oficinistas/administradores
- ✓ Data Actualizada al segundo
- ✓ Altos Volúmenes de Transacciones simples.
- ✓ Rápido tiempo de respuesta



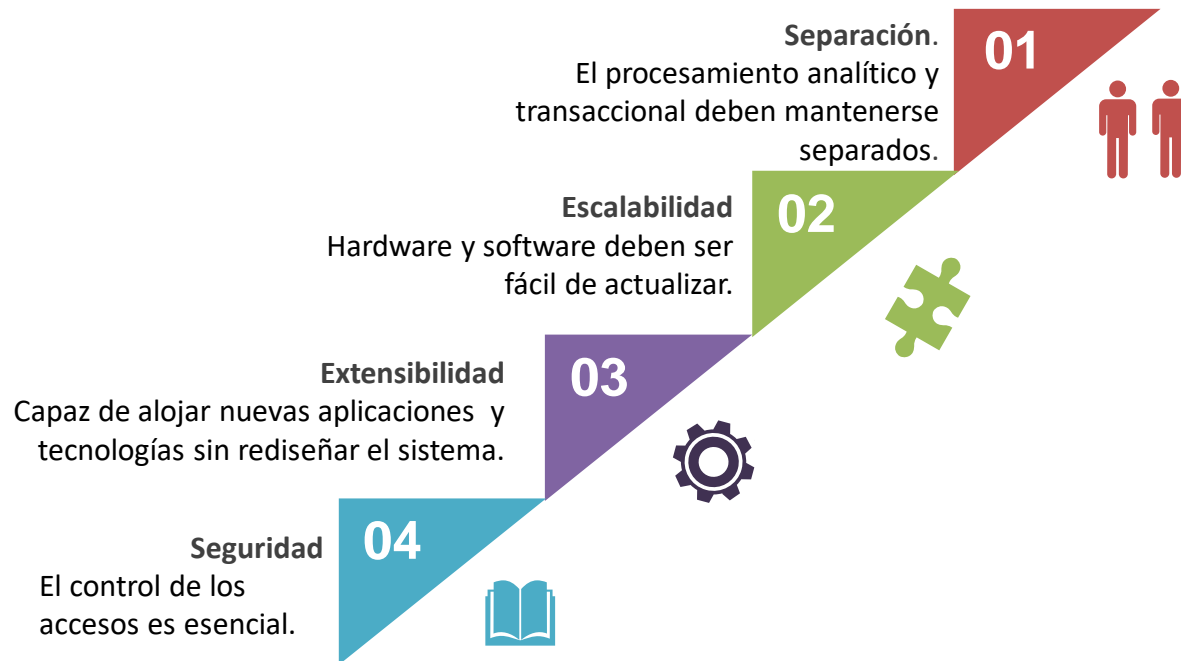
## Sistemas Analíticos

- ✓ Administra el negocio
- ✓ Empleado por Decision makers y empleados con experiencia
- ✓ Detallado y sumariado
- ✓ Integrado
- ✓ Histórico

# Arquitectura Referencial

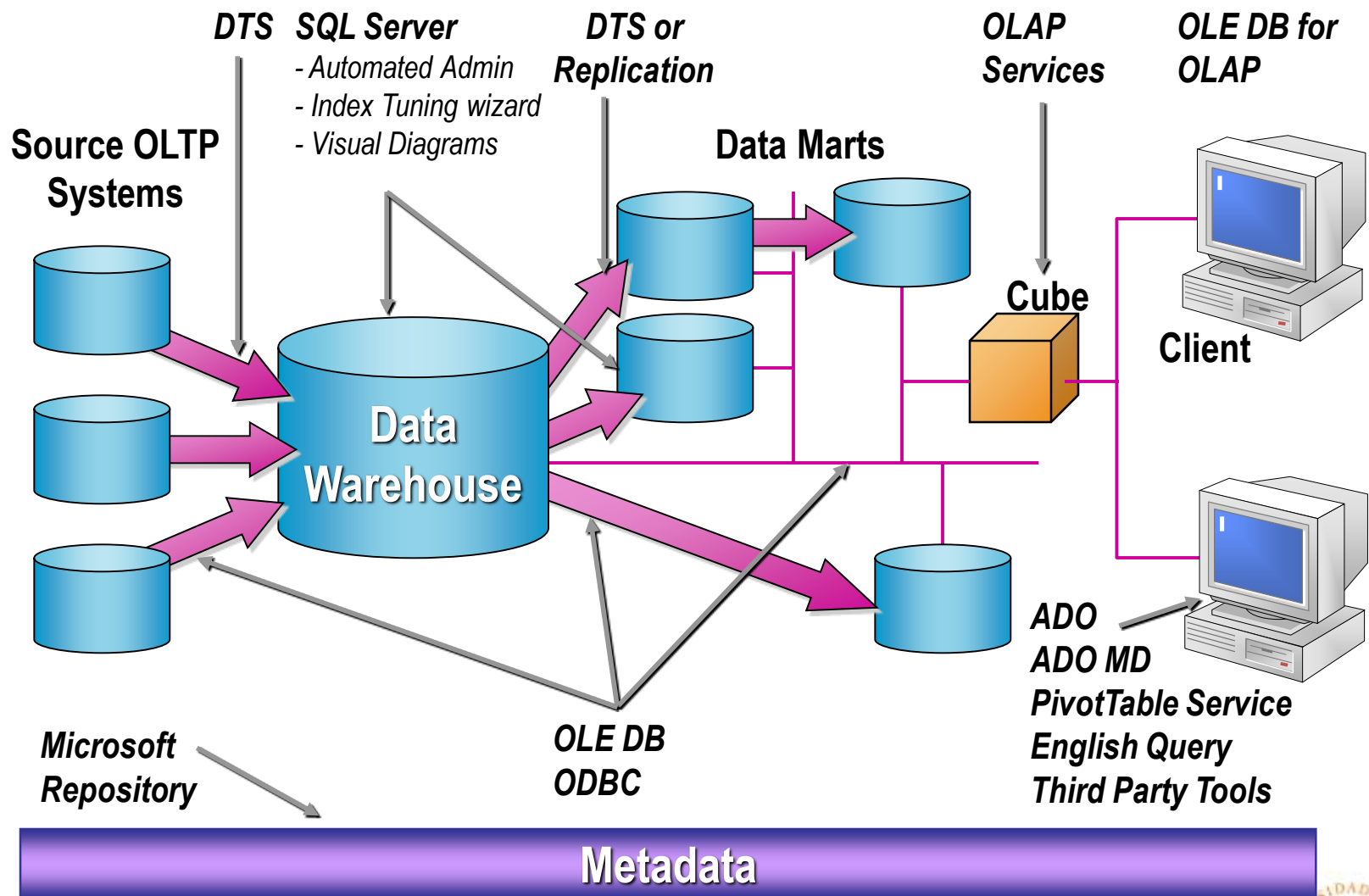


# Propiedades de Arquitectura DW



**Administración.** La gestión del almacén de datos no debe ser demasiado difícil.

# Framework

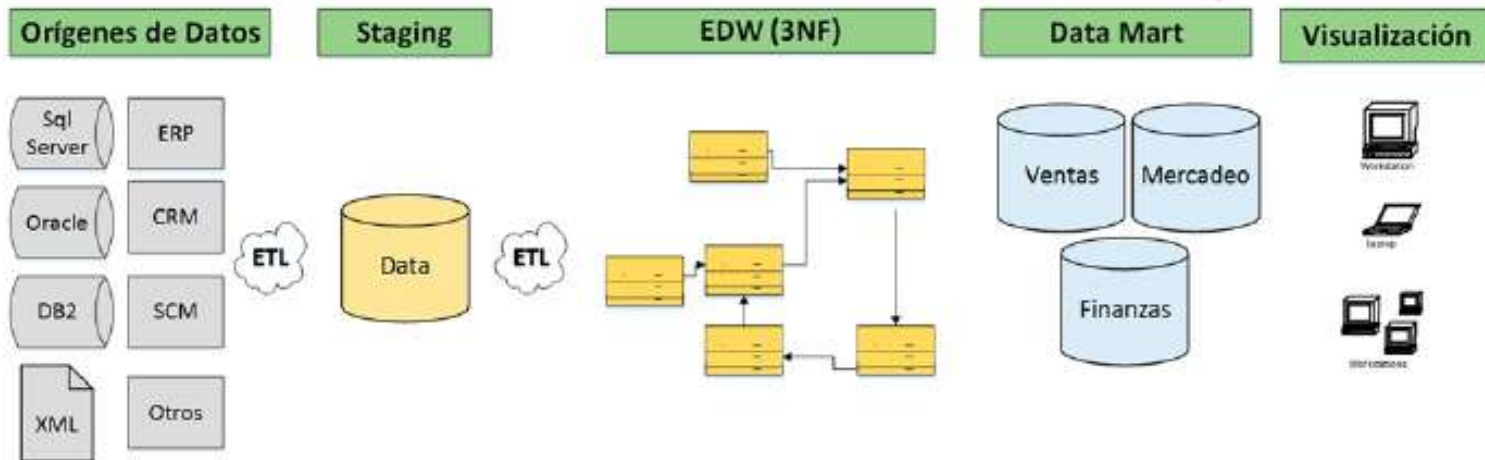


# Implementación: Top Down



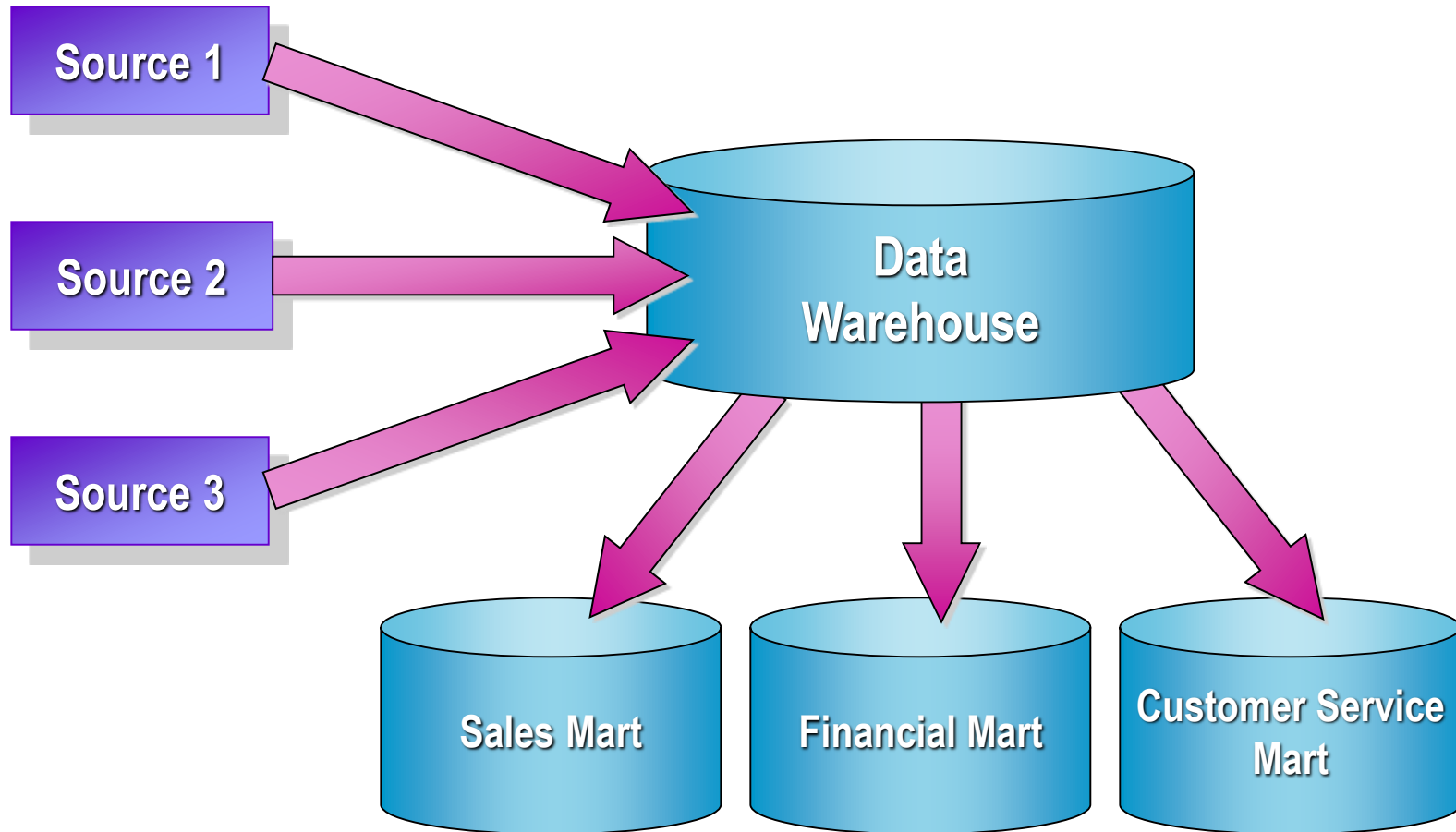
Bill Inmon  
(1945)

En este enfoque, el DW es un conjunto integrado de bases de datos, con orientación temática, diseñado para el apoyo a la toma de decisiones donde cada unidad es relevante en algún momento del tiempo. Es decir, que la información generada por varios departamentos origina el Data Warehouse.



Fuente: Moreno (2018). Data Warehouse, la nueva columna vertebral de las organizaciones

# Implementación: DW → Data Mart



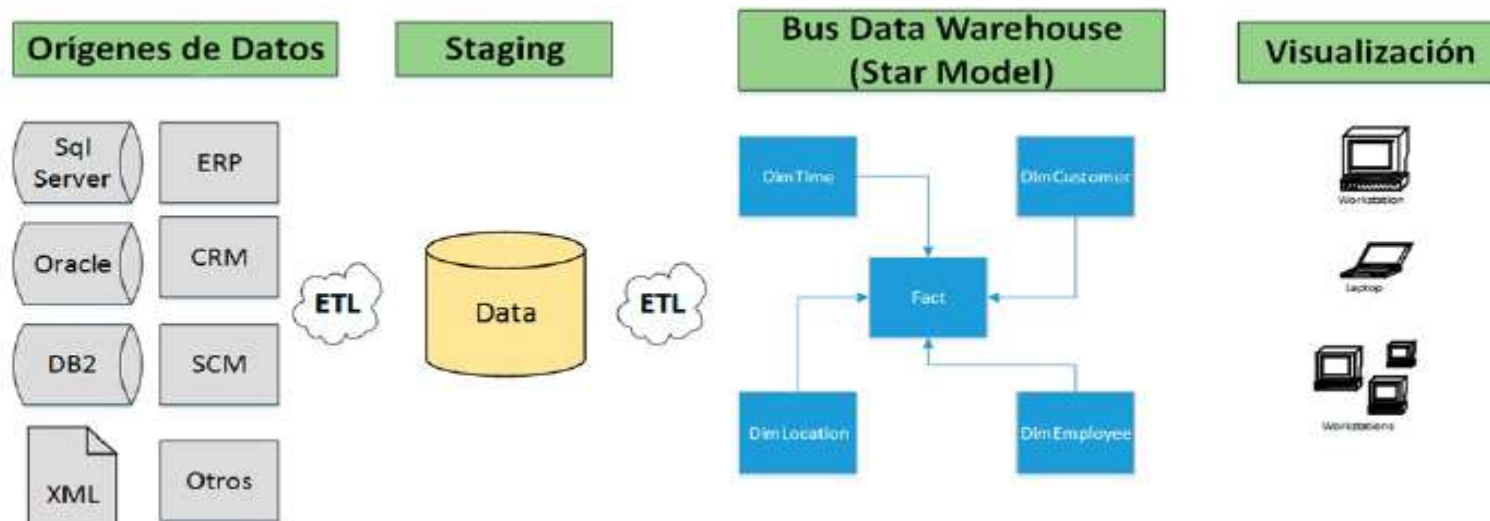


# Implementación: Bottom Up



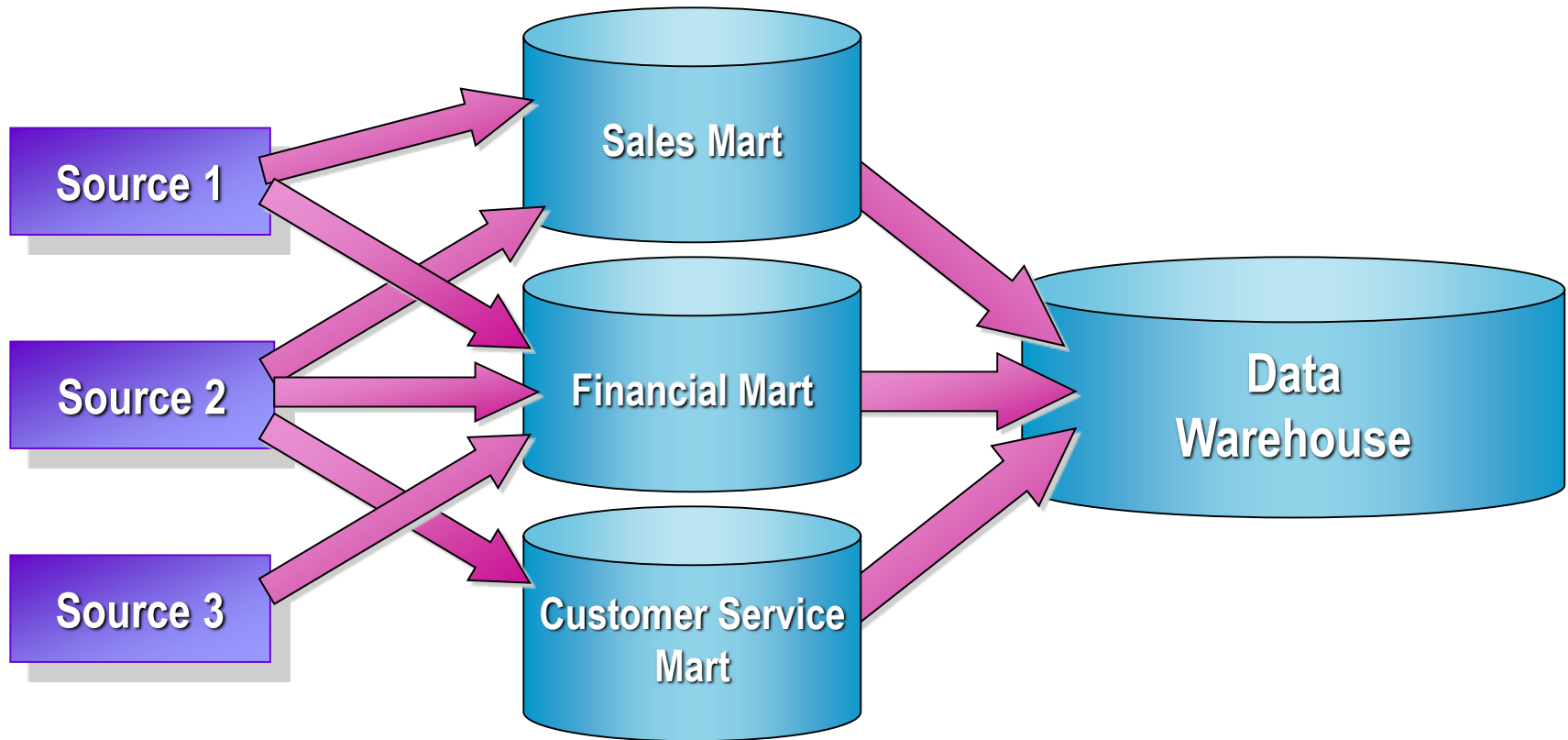
Ralph Kinball  
(1944)

El almacén de datos está compuesto por la unión de todos los Data Marts corporativos que estén relacionados entre sí, a través de dimensiones. De este modo los Data Marts se unen generando el llamado modelo en estrella.



Fuente: Moreno (2018). Data Warehouse, la nueva columna vertebral de las organizaciones

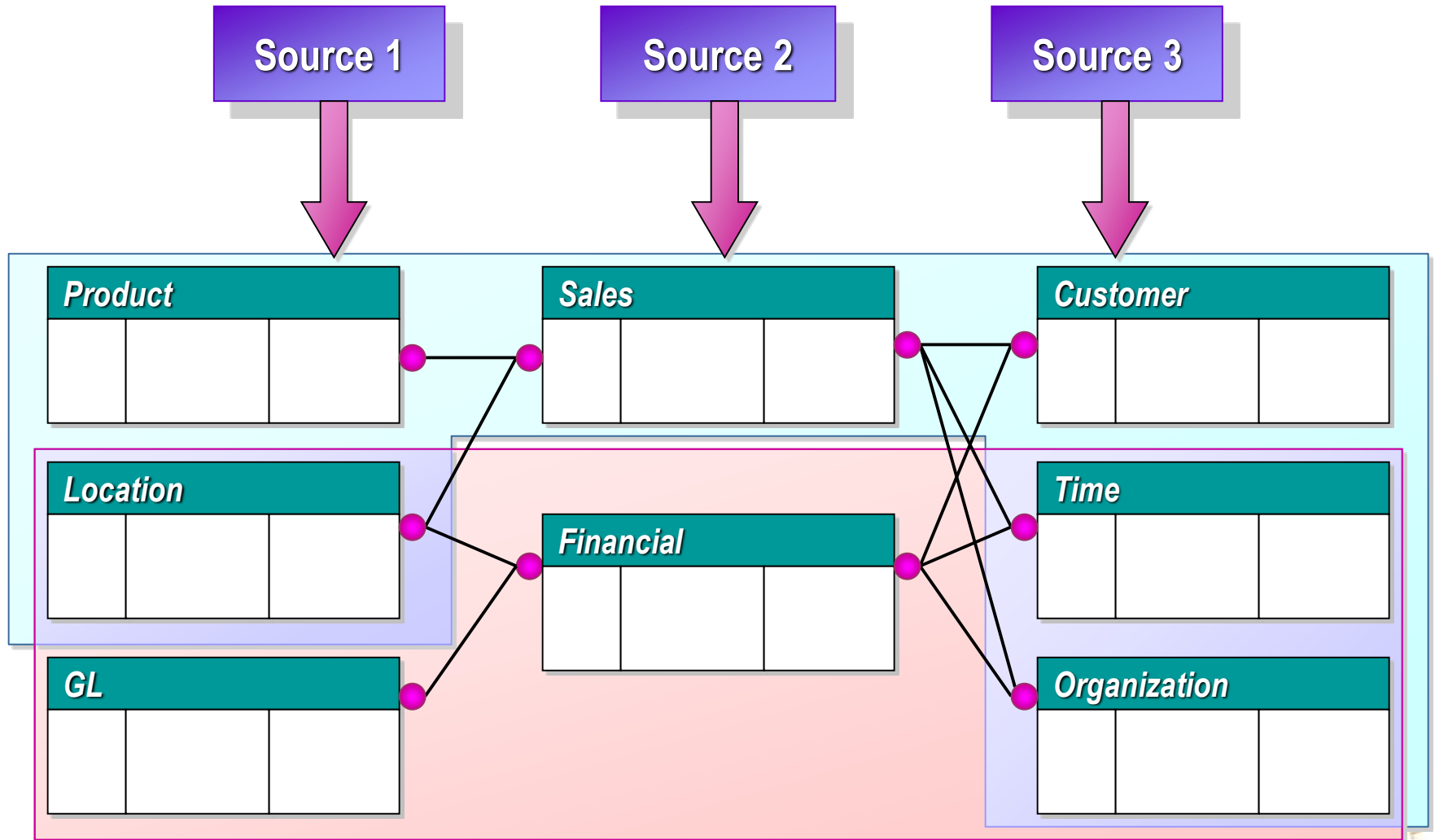
# Implementación: Data Mart → DW



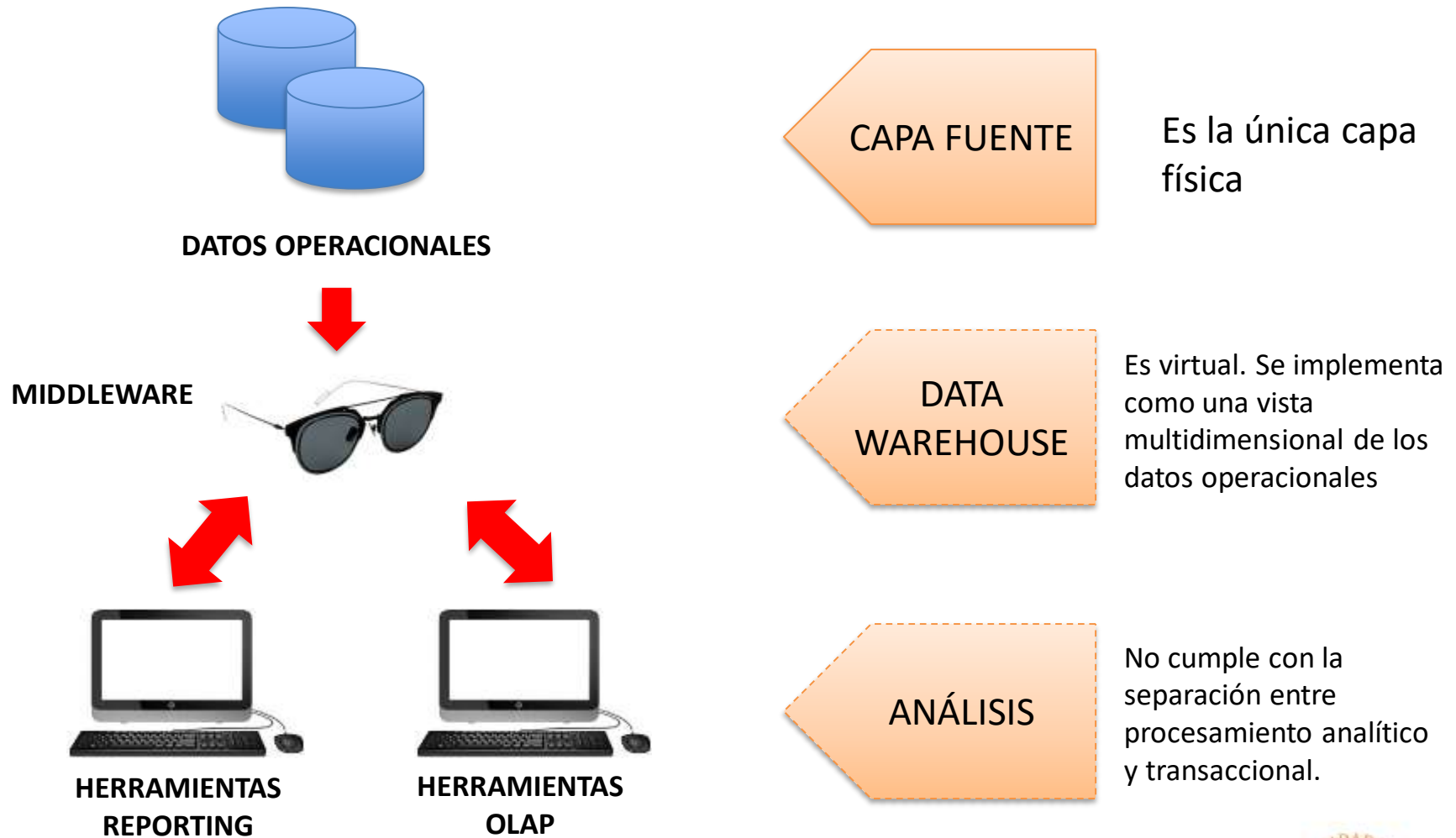
# Diferencias

Atributo	Inmon	Kinball
Construcción DW	Consume Tiempo	Toma menos tiempo
Mantenimiento	Fácil	Difícil a menudo redundante y sujeto a revisión
Costo	Alto al inicio. Menor al final.	Bajo al inicio. Cada fase siguiente costará lo mismo.
Tiempo	Largo al inicio de la configuración.	Corto al inicio de la configuración.
Habilidades	Personal capacitado.	Personal con conocimientos generales.
Requerimientos de Integración de Datos	Toda la empresa.	Áreas individuales del negocio.

# Integrando Data Marts

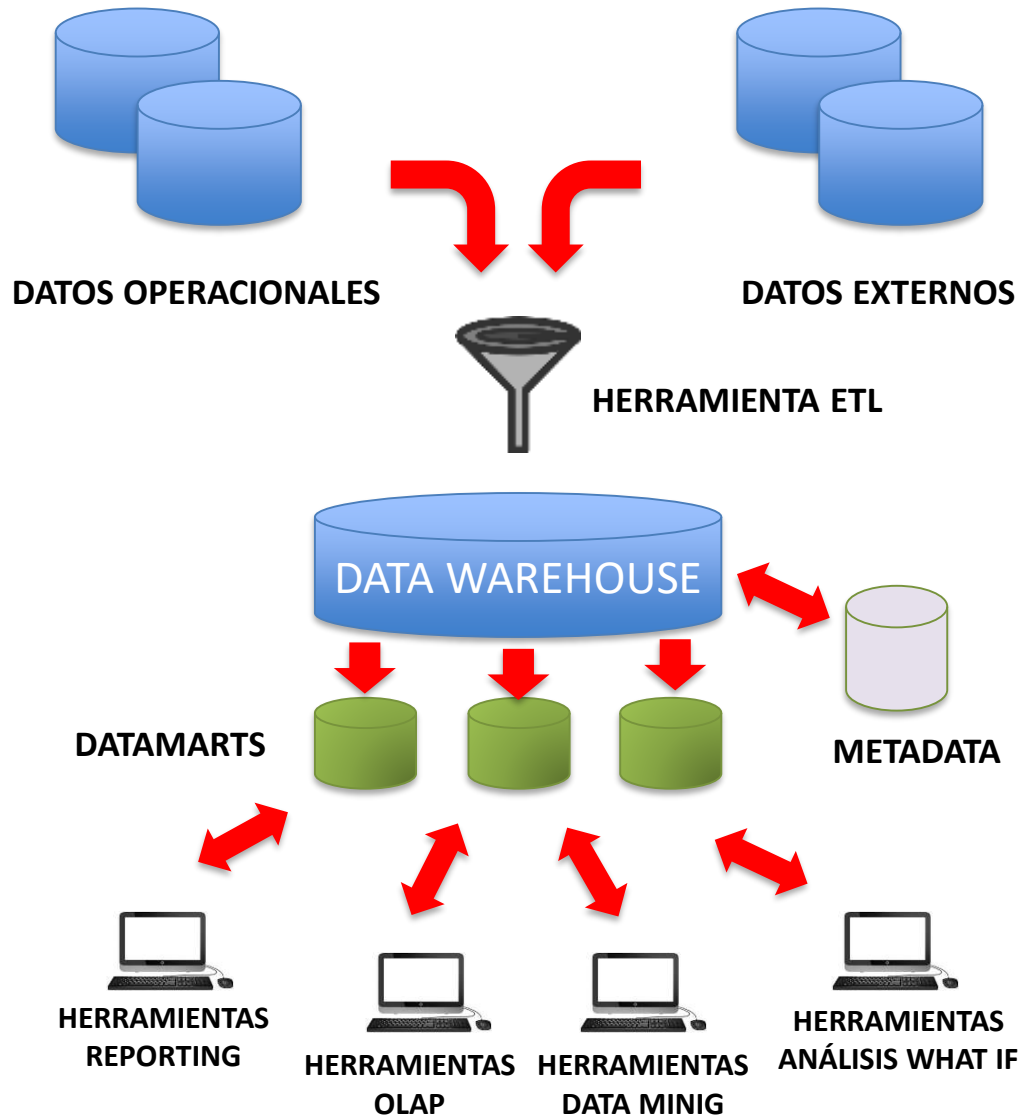


# Arquitectura de Capa Simple



*Cumple con los requisitos de integración y exactitud de los datos.*

# Arquitectura de 2 Capas



CAPA FUENTE

DATA  
STAGING

CAPA DATA  
WAREHOUSE

ANÁLISIS

# Beneficios de Arquitectura de 2 Capas

01



## Información disponible

Siempre está disponible, incluso cuando el acceso a los recursos se niega temporalmente por razones técnicas u organizativas.

02



## Las consultas de análisis del DW

No afectan la gestión de las transacciones, cuya fiabilidad es vital para que las empresas funcionen correctamente por razones operativas.

03



## Sistemas OLTP - OLAP

Falta de coincidencia en términos de tiempo y granularidad entre los sistemas OLTP, que administran datos actuales con un nivel máximo de detalle, y los sistemas OLAP, que administran datos históricos y resumidos.

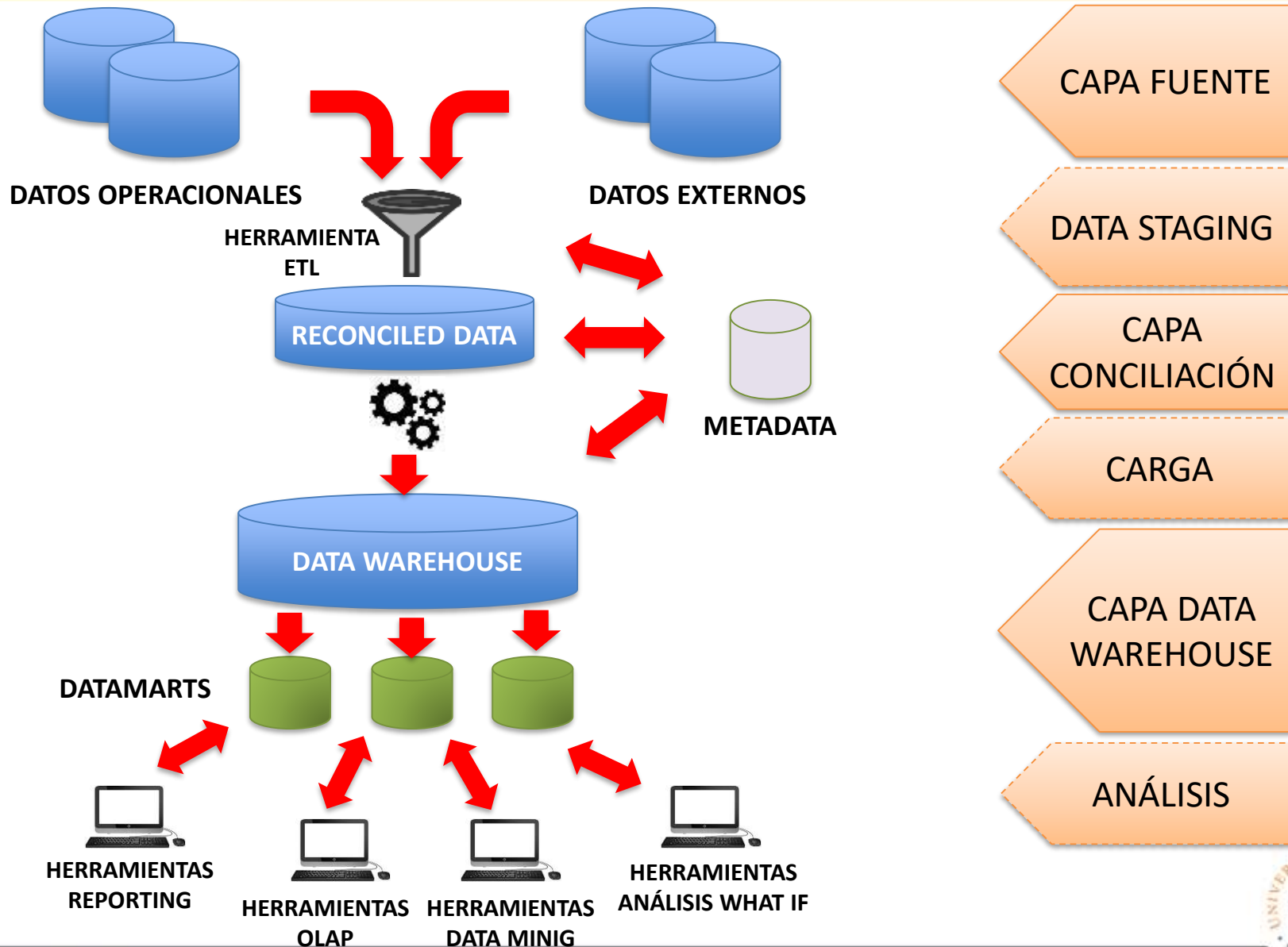
04



## Soluciones de diseño

Puede usar soluciones de diseño destinadas a la optimización del rendimiento del análisis y las aplicaciones de informes.

# Arquitectura de 3 Capas

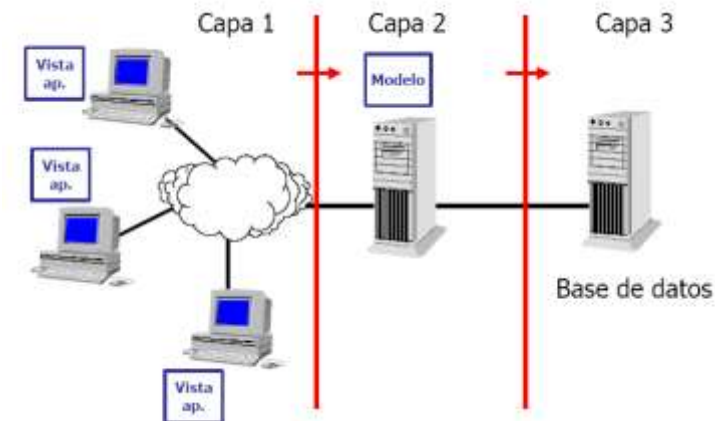




# Arquitectura de 3 Capas

La 3era Capa es la de datos reconciliados o de datos operacionales almacenados.

- ❑ Materializa los datos operacionales obtenidos después de integrar y limpiar los datos de origen.
- ❑ Mejora la realización de tareas operacionales, como generación diaria de reportes, generación de flujos de datos, etc.

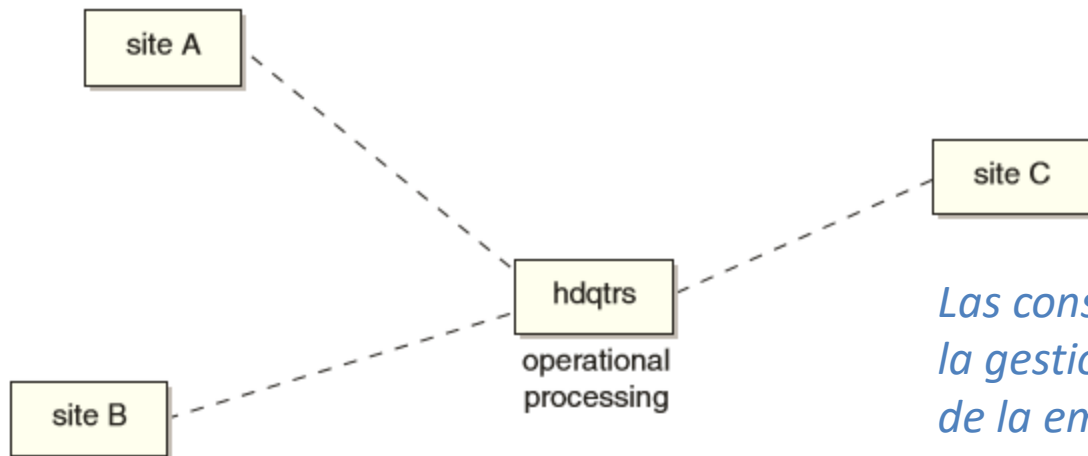


La principal ventaja de la capa de datos reconciliada es que crea un modelo de datos de referencia común para toda una empresa

# Data Warehouse distribuidos

**El negocio se distribuye geográficamente o en múltiples líneas de productos diferentes.**

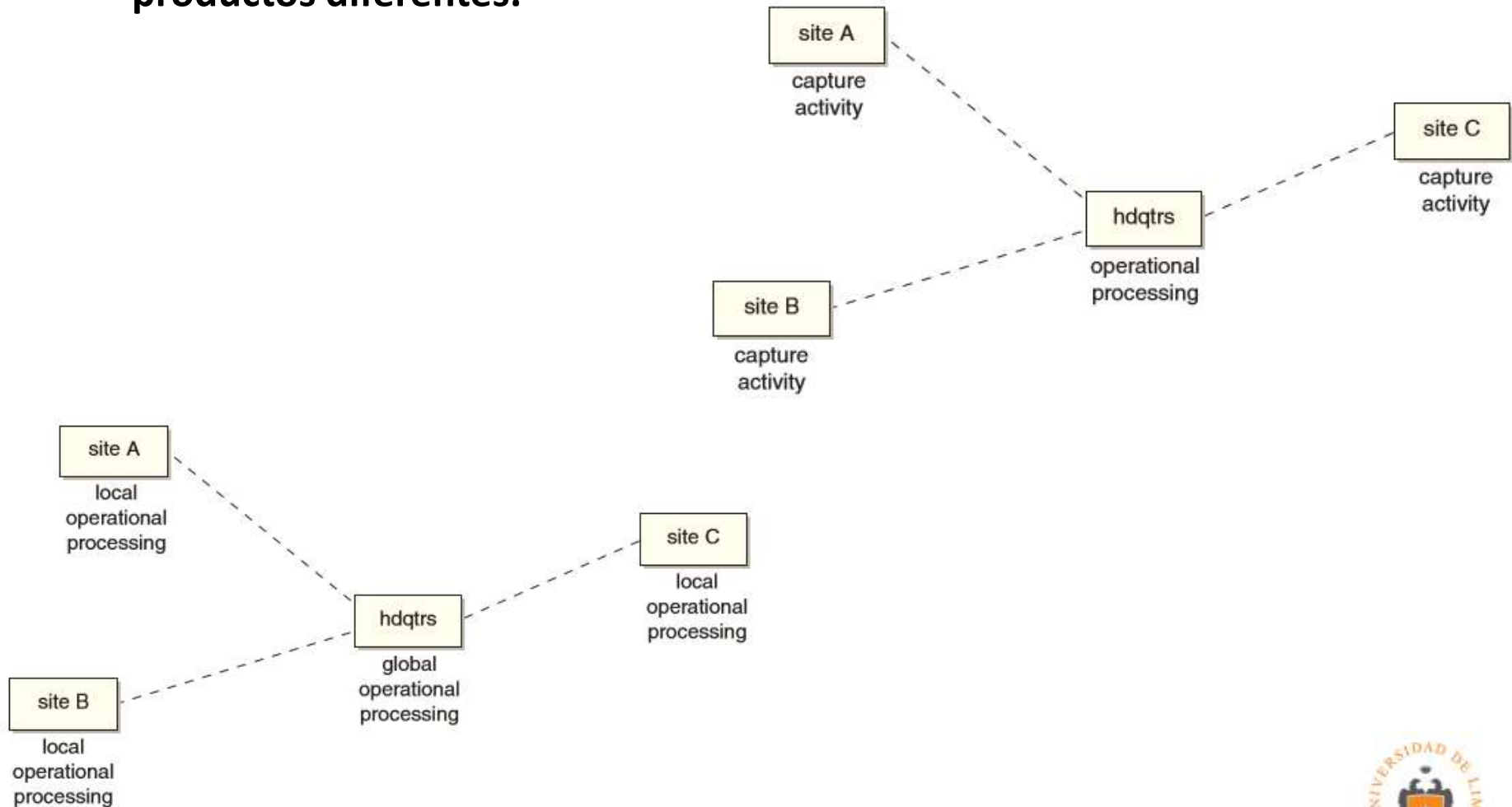
- ✓ En este caso, existe un DW local y un DW Global.
- ✓ El primero representa los datos y el procesamiento en un sitio remoto, y el segundo representa la parte del negocio que está integrada en el core negocio.



*Las consultas de análisis del DW no afectan la gestión de las transacciones (operación) de la empresa.*

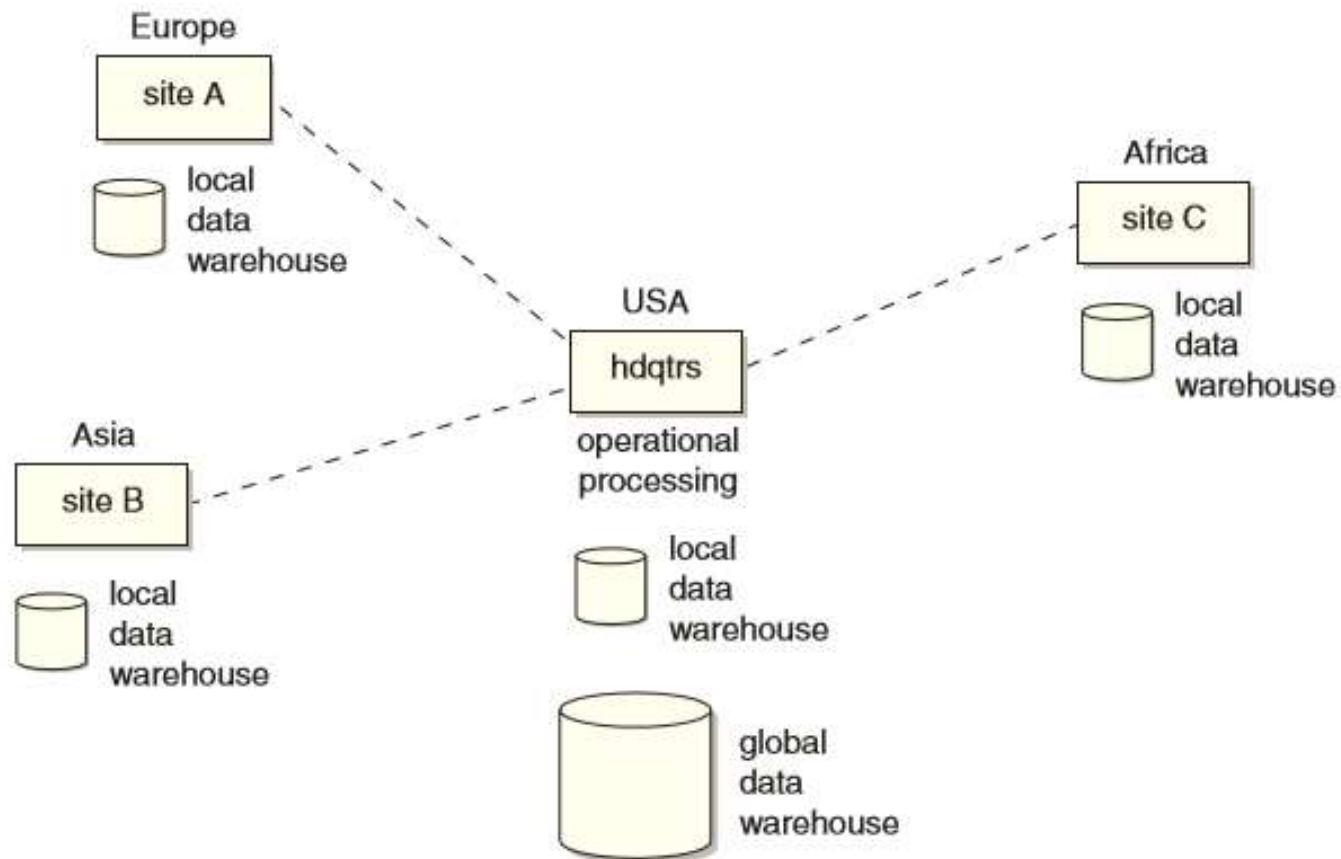
# Data Warehouse distribuidos

- ❑ El negocio se distribuye geográficamente o en múltiples líneas de productos diferentes.



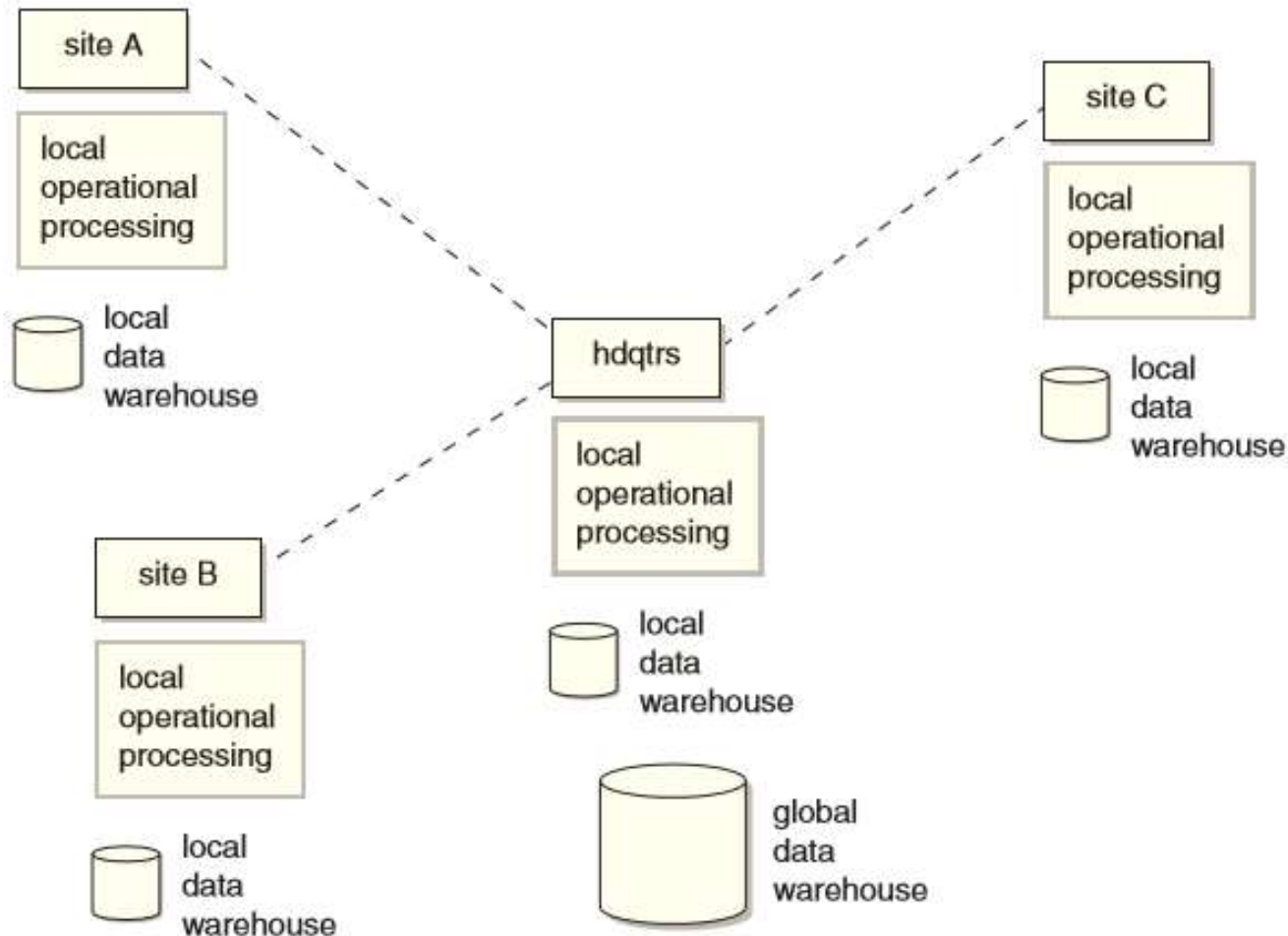
# Data Warehouse distribuidos

## ❑ DW Local



# Data Warehouse distribuidos

## ❑ DW Global



# Data Warehouse distribuidos

- ❑ **El entorno del DW contendrá una gran cantidad de datos, y el volumen de datos se distribuirá en múltiples procesadores.** Lógicamente, hay un solo DW, pero físicamente hay muchos DW que están estrechamente relacionados pero que residen en procesadores separados. Esta configuración se puede llamar almacén de datos distribuido tecnológicamente.



## El costo de entrada es barato

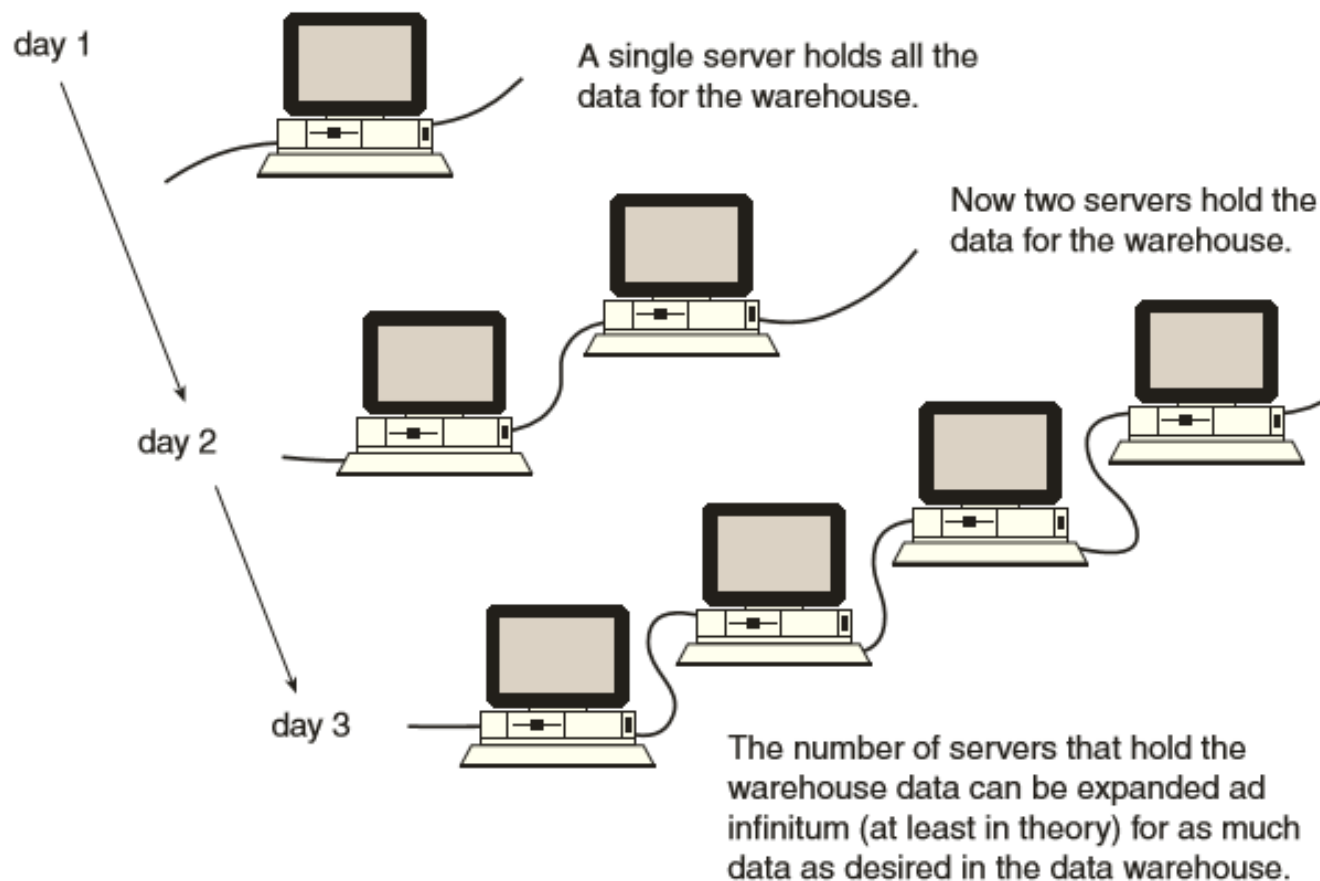
El costo de hardware y software para un DW cuando se carga inicialmente en tecnología distribuida es mucho menor que si el DW se cargara inicialmente en un hardware clásico, grande y centralizado.

## Sin límite para la cantidad de datos.

Si el volumen de datos comienza a exceder el límite de un solo procesador distribuido, entonces se puede agregar otro procesador a la red.

# Data Warehouse distribuidos

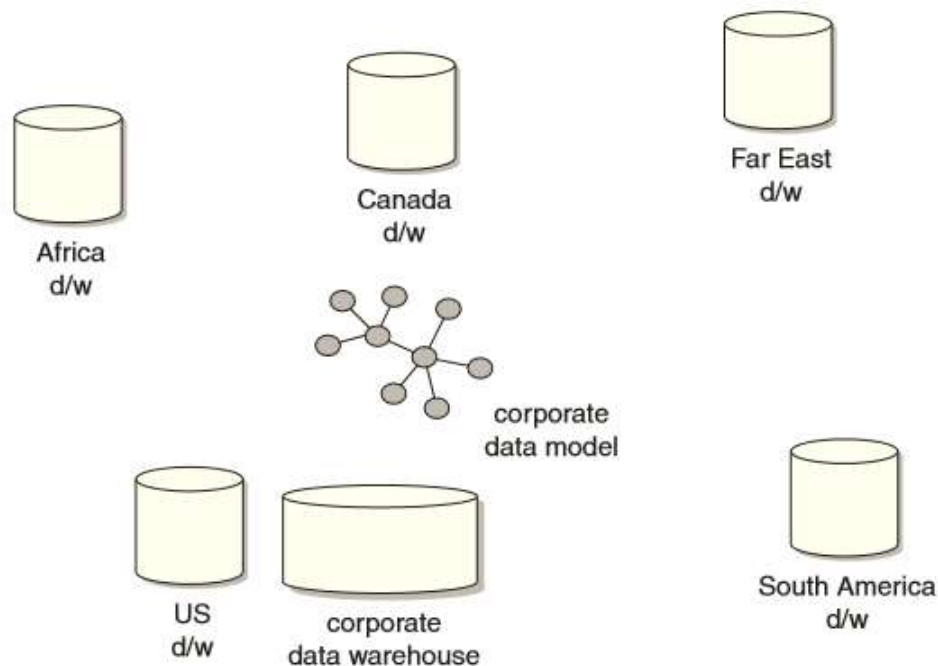
- ❑ El entorno del DW contendrá una gran cantidad de datos, y el volumen de datos se distribuirá en múltiples procesadores.



# Data Warehouse distribuidos

- ❑ **El entorno del DW crece de manera descoordinada: primero aparece un depósito de datos y luego otro.**

La falta de coordinación del crecimiento de los diferentes almacenes de datos suele ser el resultado de diferencias políticas y organizativas.

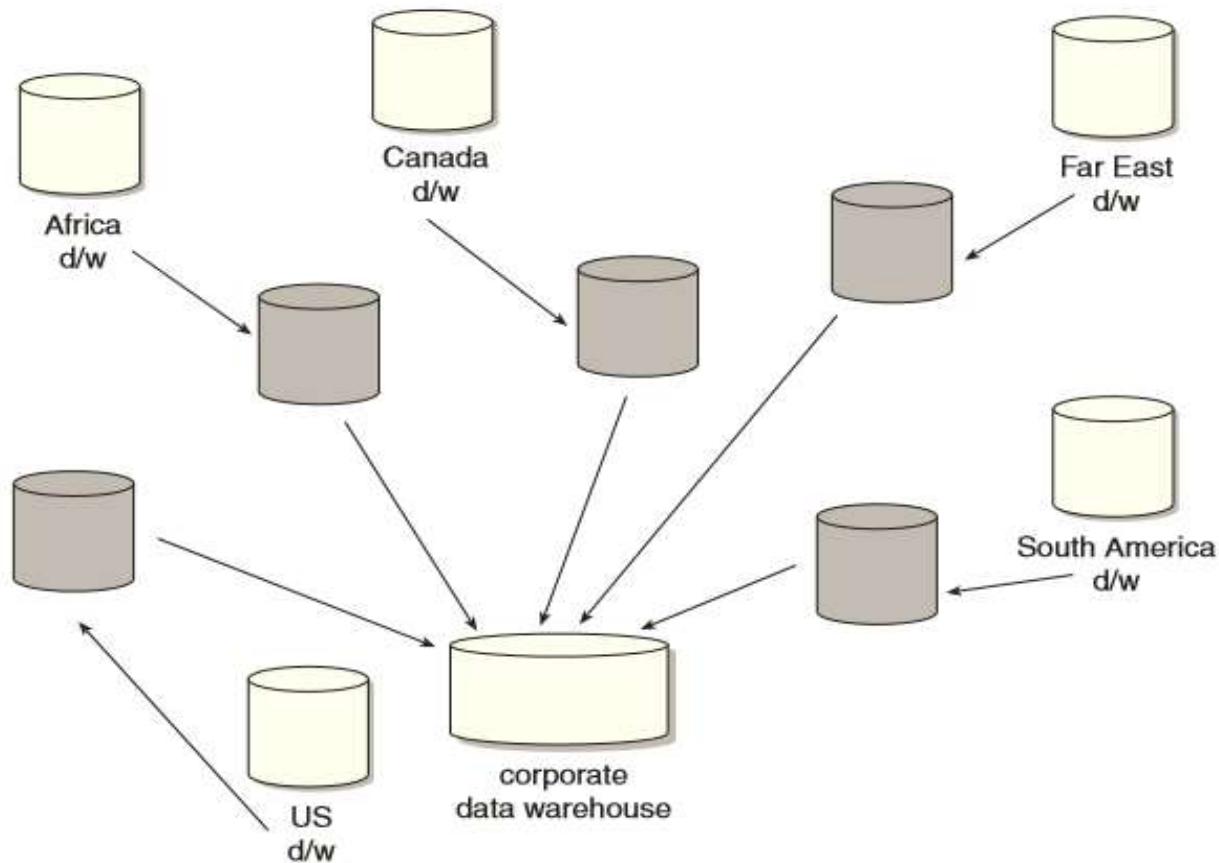


Este caso se puede llamar almacén de datos distribuidos que evoluciona independientemente.

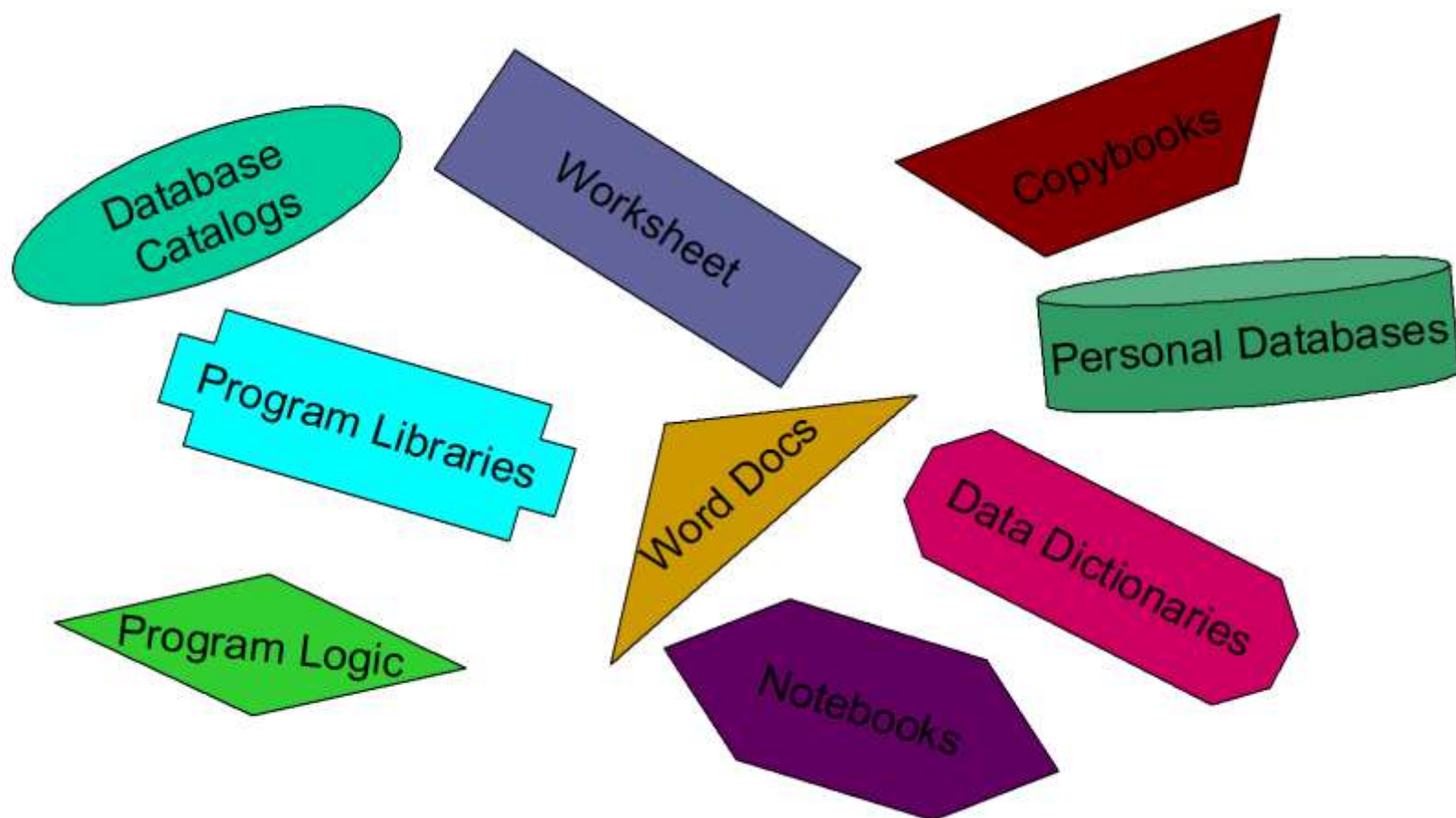


# Data Warehouse distribuidos

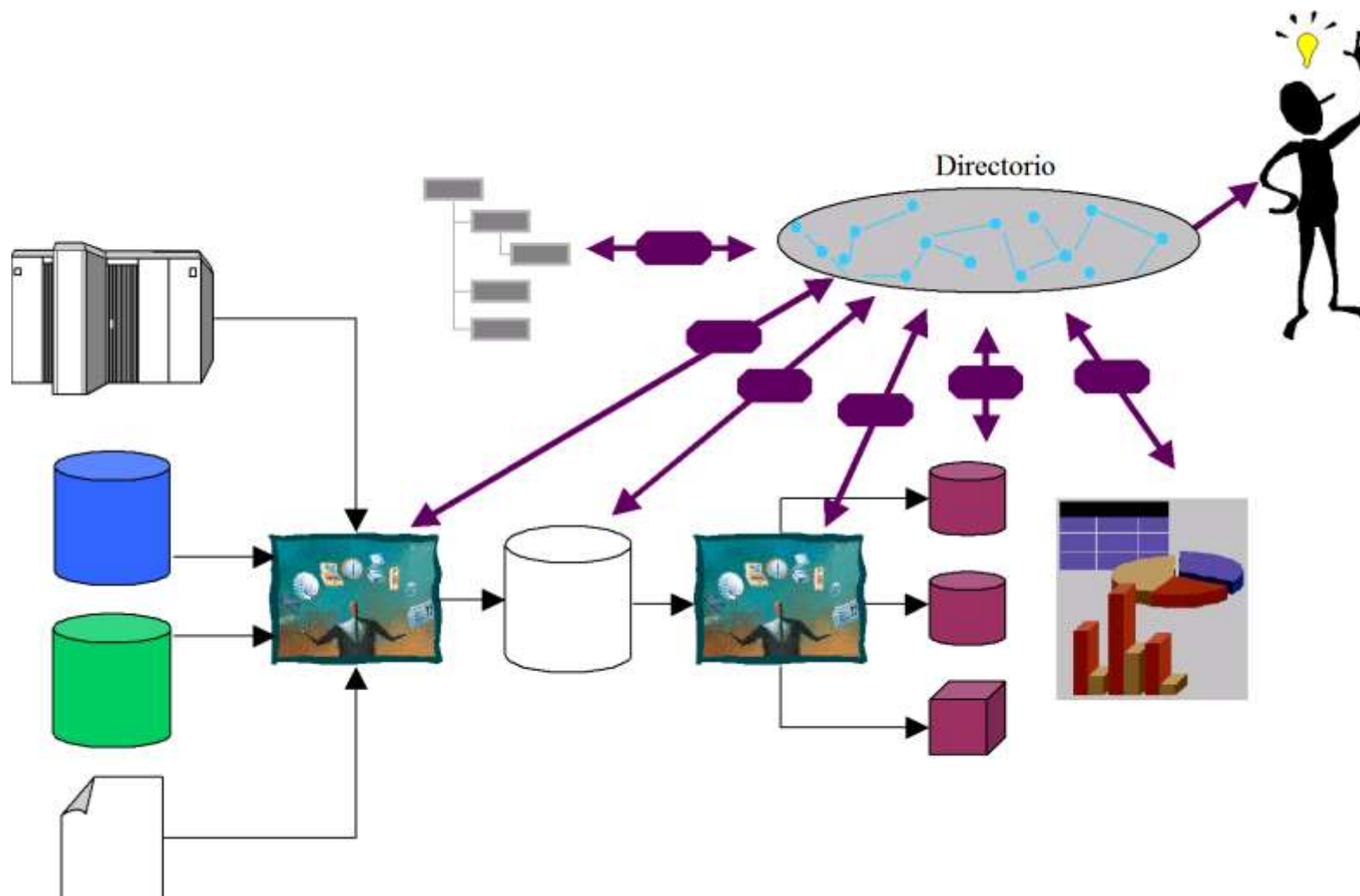
- ❑ El entorno del DW crece de manera descoordinada: primero aparece un depósito de datos y luego otro.



# Metadatos



# Arquitectura de Metadatos



# Niveles de Calidad de Datos

## Nivel 0

- Análisis de valores de dominio

## Nivel 1

- Integridad & Análisis de validez

## Nivel 2

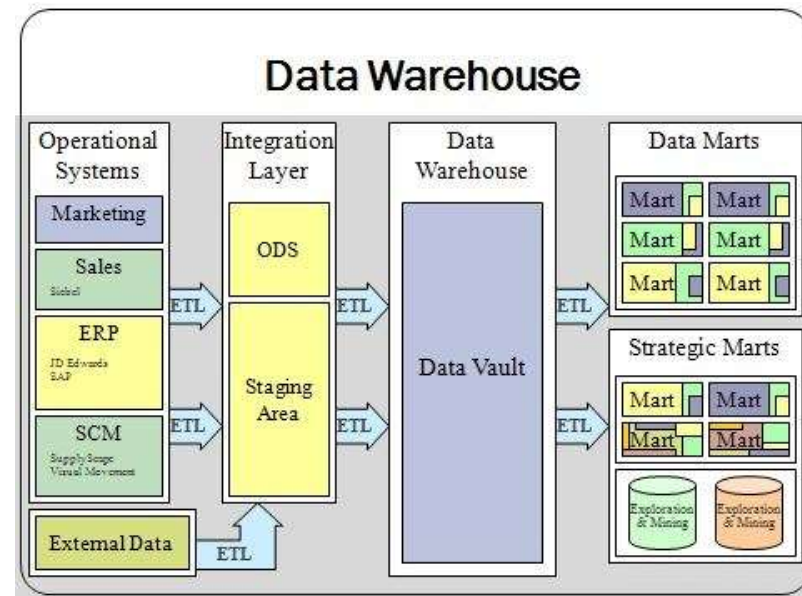
- Análisis de integridad estructural

## Nivel 3

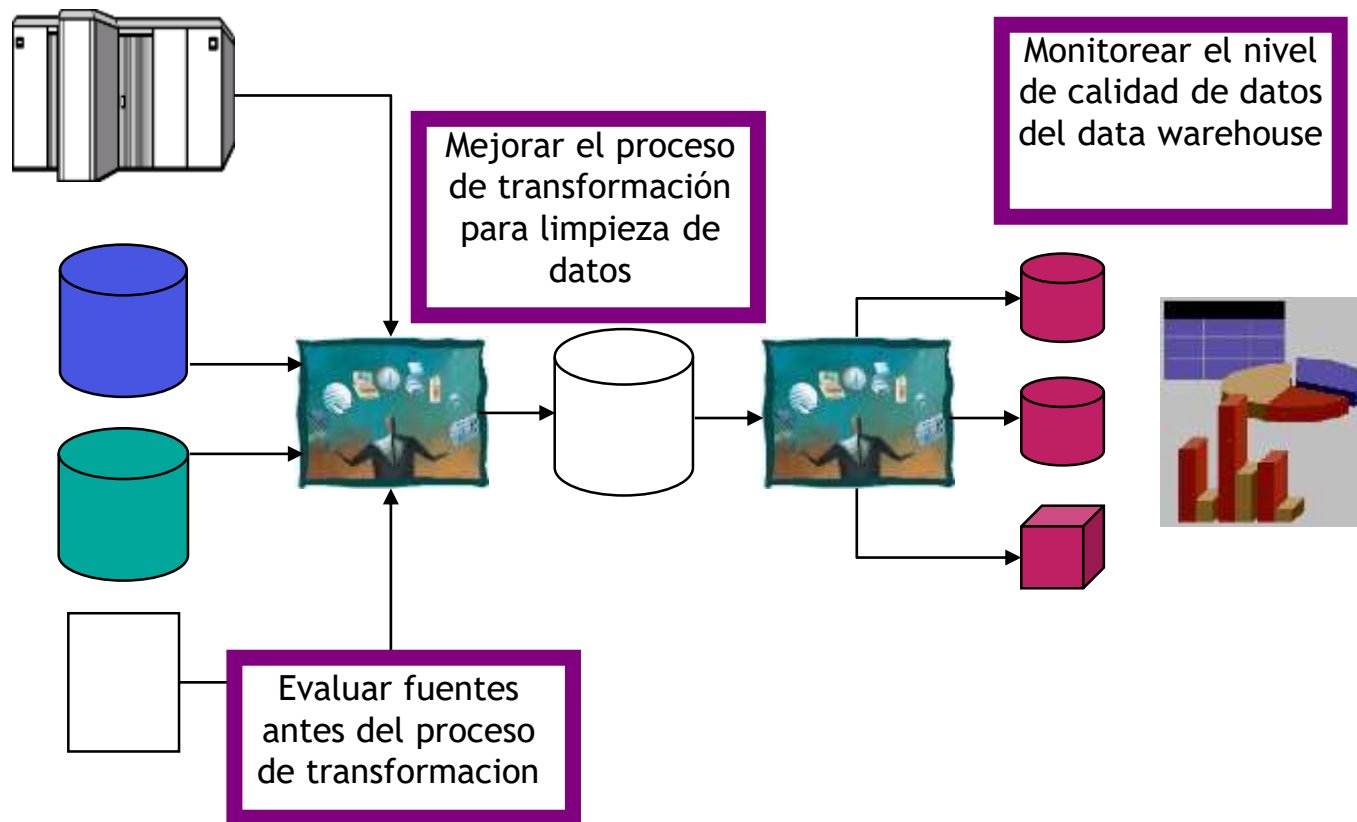
- Análisis de reglas de negocios

## Nivel 4

- Reglas de conversión de datos



# Calidad de Datos – Actividades principales



# Beneficios



## ✓ Desde perspectiva de usuarios finales

- Tener una vista global de los datos de la organización.
- Tener una herramienta fácil de usar
- Tener la capacidad de generar reportes de manera independiente
- Acceso desde internet o intranet

## ✓ Desde perspectiva IT

- Arquitectura flexible de DW que me permita cambiar cuando lo necesite.
- Presupuestos limitados
- Procesos que automáticamente actualicen los datos del DW.
- Evita el desgaste del área de IT por requerimientos de usuarios

¿Preguntas?