

# Metodologías de Data Warehouse

Un Data Warehouse es un sistema que sirve para recopilar, administrar datos de diversas fuentes, analizar y conectar los datos empresariales. Siendo el núcleo del sistema de Business Intelligence para el análisis de una grandes cantidades de información, la creación de reportes para consultas y toma de decisiones.

Al diseñar una solución de Data Warehouse, las organizaciones se topan con distintas metodologías a seguir, cada una con sus ventajas. Así como ciertas características limitantes. Las cuales se deberán evaluar para seleccionar la que mejor se adapte a los requisitos del proyecto y la organización.

Las metodologías más discutidas y principales suelen ser las diseñadas por Ralph Kimball, Bill Inmon y Dan Linsted.

## **Multidimensional**

### **(BOTTOM UP/ASCENDENTE)**

La metodología multidimensional o metodología de Ralph Kimball, mantiene un diseño ascendente. Por lo que los Data Marts son los primeros en crearse y después se integran al Data Warehouse, donde se busca que el almacenamiento de datos de los usuarios se ejecute de la forma más rápida posible.

Según Kimball, un Data Warehouse es la copia de los datos transaccionales específicamente estructurados para consultas analíticas e informes, con el fin de apoyar la toma de decisiones. Con esta metodología, al crear primero los Data Marts se proporcionan capacidades analíticas de informes para procesos específicos de negocio y funcionales.

### **Principales Ventajas:**

No requiere un equipo muy grande de desarrolladores y arquitectos de datos para mantener el Data Warehouse (menor costo).

Brinda buena funcionalidad y seguimiento de las métricas o KPI's, orienta los Data Marts a informes en cuanto a procesos de departamento o de negocios.

Administración mas "simplificada" al estar concentrado en los procesos y las áreas individuales en vez de toda la organización.

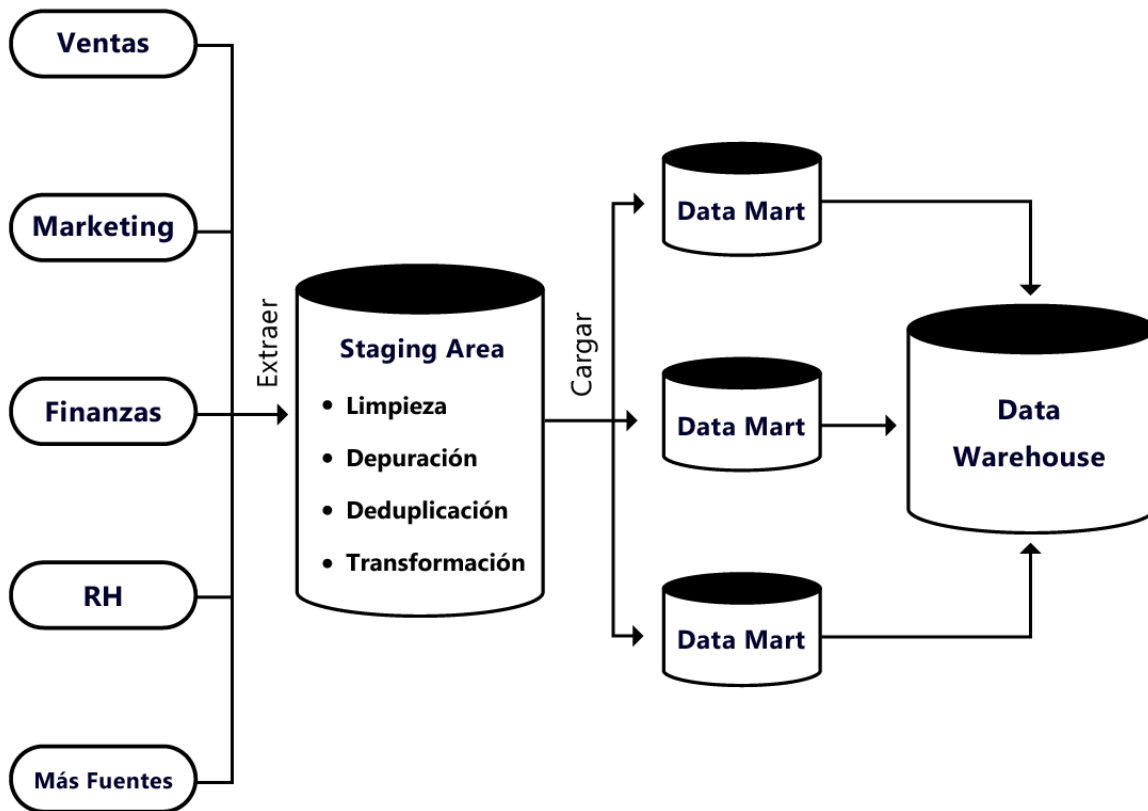
La optimización de consultas es sencilla, predecible y controlable.

### **Principales Desventajas:**

Por su enfoque en procesos y áreas, puede no llegar a cubrir o manejar todos los requisitos en los informes.

Consta de una menor flexibilidad de modificación.

## *Metodología Multidimensional de Ralph Kimball*



### **Relacional**

#### **(TOP DOWN/DESCENDENTE)**

La metodología relacional de Bill Inmon muestra un diseño descendente, donde se construye primero el Data Warehouse y posteriormente los Data Marts. Ubicando el Data Warehouse en el centro de la información corporativa lo que asegura un marco lógico en los datos.

Crea una estructura de entidades procurando que no se repitan datos. Este modelo crea una única fuente de verdad para todo el negocio.

La carga de datos se vuelve menos compleja debido a la estructura normalizada del modelo. Sin embargo, el uso de esta disposición para realizar consultas, es complicado; ya que incluye gran cantidad de tablas y vínculos.

Este modelo propone la construcción de Data Marts por separado para cada departamento. Todos los datos que entran en el Data Warehouse están integrados para garantizar la integridad y la coherencia en toda la empresa. Actuando el Data Warehouse como el único origen de datos.

### Principales Ventajas:

El Data Warehouse proporciona una única versión de la verdad, al ser el único origen de datos para los Data Marts.

Tiene una mayor facilidad de comprensión de los procesos empresariales para los usuarios, ya que el modelo lógico representa entidades empresariales detalladas.

Resulta más fácil y menos propenso al fracaso el proceso de ETL, puesto que en la actualización de los datos y las anomalías se evitan al contar con una redundancia muy baja.

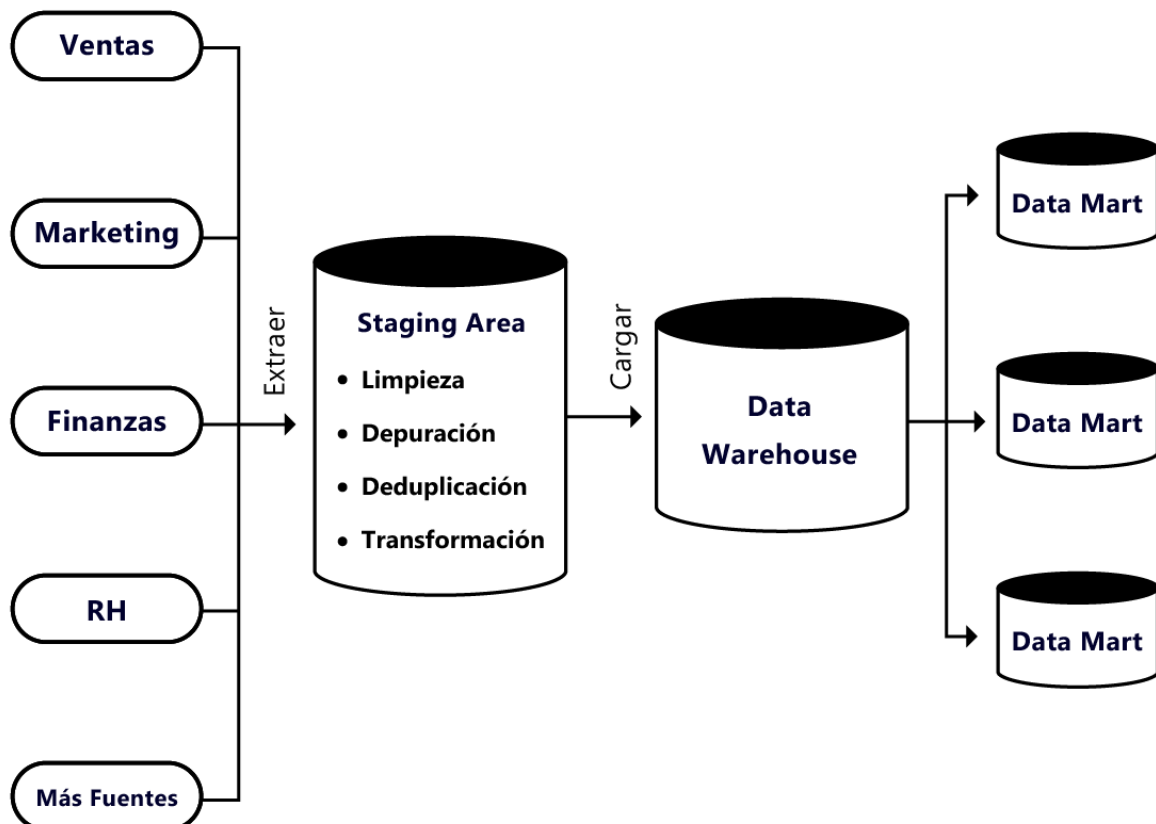
Mayor flexibilidad a cambios por necesidades analíticas, de negocio y/o por fuentes de datos.

### Principales Desventajas:

De mayor complejidad, se requieren recursos con mayor capacidad en modelado y almacenamiento de datos (generalmente de mayor costo).

Suele requerir de tiempos más largos para dado su procesos y entrega.

### *Metodología Relacional Bill Inmon*



## Data Vault

Data Vault es una metodología híbrida creada por Dan Linsted, es utilizada principalmente cuando las empresas tienen un aumento exponencial constante de datos por lo que presentan problemas de rediseño y mantenimiento.

Esta metodología permite el almacenamiento y auditoría de información histórica, carga paralela de datos y que al contar con varios Data Warehouse se pueda escalar sin tener que rediseñar por completo la solución. Proporciona además flexibilidad, lo que resulta idóneo para las organizaciones con un crecimiento exponencial constante.

Este modelo se compone de tres tablas:

HUB.- Contienen las claves únicas de un caso, tema, o empleado en concreto. (Ej. HUB\_EMPLOYEE)

LINK.- Se encargan de rastrear las relaciones de los hubs. (Ej. LINK\_EMPLOYEE\_STORE)

SÁTELITE.- Contienen los atributos relacionados con los links o hubs y los mantiene actualizados. (Ej. SAT\_EMPLOYEE)

### Principales Ventajas:

Diseñado especializado para almacenar registros (hace el proceso de registro de datos sea más sencillo).

Automatiza fácilmente los procesos ETL.

Fácil rastreo y auditoría de datos.

Permite varios sistemas de origen y relaciones con cambios frecuentes.

### Principales Desventajas:

Existe menor grado de especialización y documentación sobre esta metodología.

Puede llegar a requerir un mayor esfuerzo, adaptación y explotación de herramientas para diseñar las capas semánticas, así como ajustes y modificaciones especiales.

### *Metodología Híbrida de Dan Linsted*

