

UNIVERSIDAD AUTÓNOMA GABRIEL RENE MORENO
FACULTAD DE INGENIERÍA EN CIENCIAS DE LA
COMPUTACIÓN Y TELECOMUNICACIONES
CARRERA INGENIERÍA EN SISTEMAS



TAREA #5

**METODOLOGÍAS PARA LA CONSTRUCCIÓN DE DATA
WAREHOUSE**

MATERIA: Sistemas para el soporte a la toma de decisiones

DOCENTE: Ing. Peinado Pereira Miguel Jesus

SIGLA: INF432 -SA

UNIVERSITARIO: Cahuasiri Poroso David

REGISTRO: 220031665

2/2024

Santa Cruz – Bolivia

La construcción de un Data Warehouse (DW) implica la aplicación de metodologías específicas que aseguren la integración, consistencia, y disponibilidad de los datos para análisis y toma de decisiones. A continuación, se describen algunas de las metodologías más utilizadas:

1. Metodología de Ralph Kimball (Enfoque Bottom-Up)

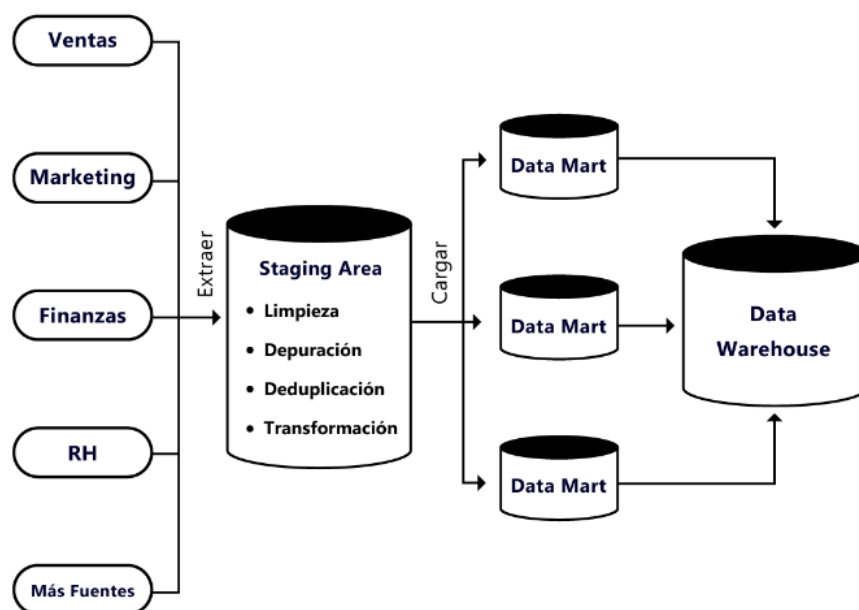
Descripción: Kimball propone un enfoque en el que se crean Data Marts específicos para áreas funcionales del negocio (por ejemplo, ventas, finanzas, etc.), y luego estos Data Marts se integran para formar un Data Warehouse.

Pasos Clave:

- Identificación de las necesidades de negocio y áreas funcionales.
- Diseño de Data Marts dimensionales (con esquemas estrella o copo de nieve).
- Integración de los Data Marts en un DW centralizado.
- Implementación de procesos ETL (Extract, Transform, Load) para la carga de datos.

Ventajas: Más rápido de implementar y enfocado en entregar valor a corto plazo. Flexible y escalable.

Desventajas: La integración de múltiples Data Marts puede ser compleja.



2. Metodología de Bill Inmon (Enfoque Top-Down)

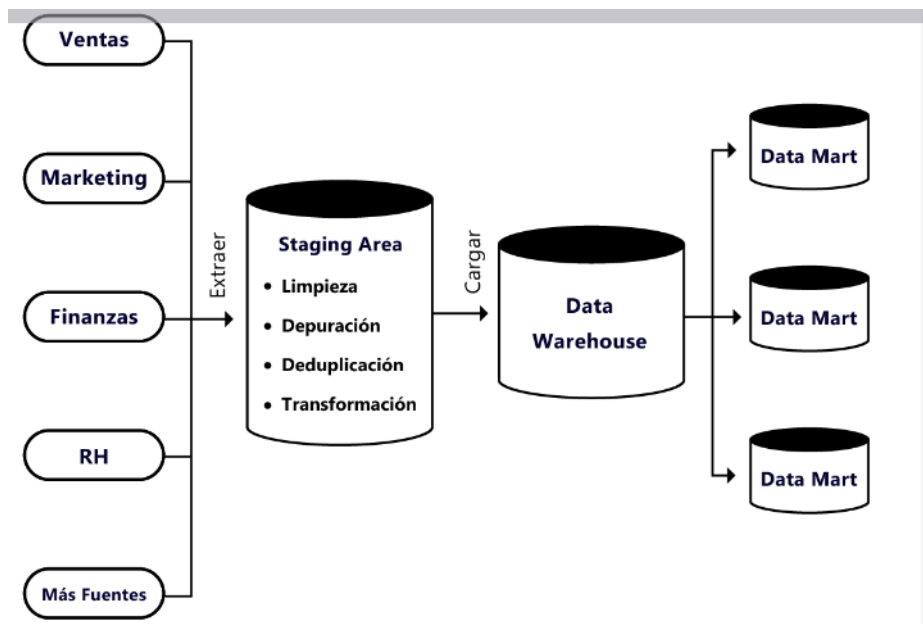
Descripción: Inmon sugiere un enfoque en el que primero se diseña un Data Warehouse centralizado y luego se derivan los Data Marts específicos según sea necesario. El DW se construye utilizando un modelo de datos corporativo con un enfoque normalizado.

Pasos Clave:

- Definición de un modelo de datos corporativo.
- Diseño del Data Warehouse en un formato normalizado (Tercera Forma Normal).
- Implementación de procesos ETL para poblar el DW.
- Creación de Data Marts derivados del DW según las necesidades del negocio.

Ventajas: Un enfoque más robusto y alineado con la visión a largo plazo de la organización. Evita redundancia de datos.

Desventajas: Tiempo de implementación más largo y mayor complejidad inicial.



3. Metodología Ágil (Agile BI)

Descripción: Adaptación de principios ágiles al desarrollo de Data Warehouses. Se enfoca en entregas incrementales y rápidas de valor.

Pasos Clave:

- Identificación de requerimientos en ciclos cortos (sprints).
- Desarrollo iterativo e incremental del Data Warehouse.
- Feedback continuo de los usuarios finales para ajustar y priorizar el desarrollo.
- Integración de procesos ETL de manera iterativa.

Ventajas: Adaptabilidad, rapidez en la entrega de resultados y mejor alineación con las necesidades cambiantes del negocio.

Desventajas: Puede ser difícil de implementar en grandes empresas con procesos bien establecidos.

4. Metodología Data Vault

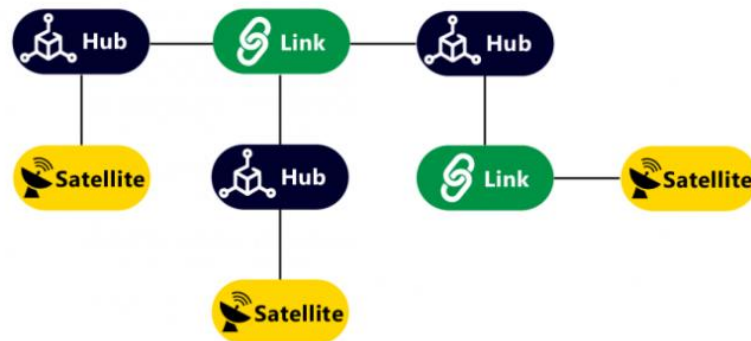
Descripción: Data Vault es una metodología que permite construir un Data Warehouse de manera ágil y escalable. Utiliza una arquitectura flexible que facilita la integración de datos de múltiples fuentes sin necesidad de modificar el modelo central.

Pasos Clave:

- Modelado del Data Warehouse utilizando componentes principales: Hub (entidades principales), Satellites (atributos de las entidades) y Links (relaciones entre entidades).
- Implementación de procesos ETL que se adaptan a la estructura modular del Data Vault.
- Creación de vistas o Data Marts específicos basados en el Data Vault.

Ventajas: Escalabilidad, flexibilidad, y adaptabilidad a cambios en las fuentes de datos.

Desventajas: Curva de aprendizaje pronunciada y puede ser excesivamente complejo para proyectos más pequeños.



5. Metodología Lean Data Warehousing

Descripción: Inspirada en principios Lean, esta metodología busca minimizar el desperdicio y maximizar la eficiencia en la construcción del Data Warehouse.

Pasos Clave:

- Identificación de valor desde la perspectiva del cliente.
- Eliminación de procesos y tareas innecesarias en el desarrollo del DW.
- Implementación de entregas continuas y mejoras incrementales.

- Monitorización y ajuste continuo basado en el feedback del negocio.

Ventajas: Enfocada en la eficiencia, reduce tiempo y costos en la construcción del DW.

Desventajas: Requiere una cultura organizacional muy enfocada en la mejora continua.

6. Metodología híbrida

Descripción: Combina aspectos de las metodologías de Kimball y Inmon, buscando un equilibrio entre la entrega rápida de valor (mediante Data Marts) y un diseño estructurado y robusto (mediante un modelo centralizado).

Pasos Clave:

- Inicia con la creación de un modelo centralizado, pero permite la implementación de Data Marts específicos a corto plazo.
- Integra procesos iterativos y retroalimentación continua del negocio.

Ventajas: Equilibrio entre agilidad y robustez.

Desventajas: Puede ser difícil de gestionar y requiere una planificación cuidadosa.

Cada metodología tiene sus propios beneficios y desafíos, y la elección de una sobre otra dependerá de factores como el tamaño de la organización, la madurez en la gestión de datos, los requerimientos de negocio, y los recursos disponibles.

Para la construcción de un Data Warehouse (DW), la planificación y el análisis son fundamentales para garantizar que el proyecto sea exitoso y cumpla con los objetivos del negocio. A continuación, se describen los pasos detallados en cada fase clave del proceso: Planificación, Análisis, Diseño, y Implementación.

Planificación

Definir los Objetivos del Proyecto:

- Identifica los objetivos estratégicos del Data Warehouse.
- Define claramente el alcance del proyecto y qué problemas del negocio se buscan resolver.

Identificar las Partes Interesadas:

- Involucra a todas las partes interesadas clave (usuarios finales, gerentes, equipos de TI) desde el inicio.
- Establece canales de comunicación y define roles y responsabilidades.

Establecer el Cronograma y el Presupuesto:

- Elabora un cronograma con hitos clave.
- Determina el presupuesto necesario, incluyendo hardware, software, personal y formación.

Análisis de Riesgos:

- Identifica posibles riesgos (tecnológicos, de integración, de recursos) y planifica medidas de mitigación.

Definir la Arquitectura General:

- Decide la arquitectura básica del DW (on-premises, cloud, híbrida).
- Evalúa las tecnologías y herramientas que se utilizarán.

Análisis

Recolección de Requisitos:

- Entrevista a los usuarios finales y otras partes interesadas para entender las necesidades de información y los reportes necesarios.
- Documenta los requisitos de datos, incluyendo las fuentes de datos y los requisitos de integridad, calidad y disponibilidad.

Análisis de Datos Existentes:

- Evalúa las fuentes de datos existentes para comprender su estructura, calidad y relevancia.
- Identifica las brechas entre los datos disponibles y los requisitos del negocio.

Mapeo de Datos:

- Relaciona las fuentes de datos con los requisitos del negocio.
- Establece cómo se transformarán y cargarán los datos en el Data Warehouse.

Definición de Indicadores Clave de Desempeño (KPIs):

- Define los KPIs que serán soportados por el DW y cómo se calcularán.

Diseño

Diseño del Modelo de Datos:

- Elige entre un diseño dimensional (esquema estrella o copo de nieve) o un diseño normalizado, dependiendo de la metodología seleccionada.
- Define las tablas de hechos y dimensiones, las relaciones entre ellas, y los índices necesarios para mejorar el rendimiento.

Diseño de Procesos ETL (Extract, Transform, Load):

- Diseña los procesos para extraer datos de las fuentes, transformarlos según las reglas de negocio, y cargarlos en el Data Warehouse.
- Considera la calidad de los datos, la transformación de formatos y la eliminación de duplicados.

Diseño de la Arquitectura Técnica:

- Especifica los componentes técnicos del DW, como servidores, almacenamiento, software de bases de datos, herramientas ETL, y herramientas de BI (Business Intelligence).

Diseño de la Seguridad:

- Establece políticas de seguridad, incluyendo acceso a datos, encriptación, y auditoría.

Diseño de la Recuperación ante Desastres:

- Planifica un sistema de respaldo y recuperación ante desastres para proteger los datos del DW.

Implementación

Construcción del Data Warehouse:

- Configura la infraestructura técnica (hardware, software, bases de datos).
- Implementa el modelo de datos en la base de datos del DW.
- Desarrolla y prueba los procesos ETL para asegurar que los datos se carguen correctamente.

Carga de Datos Inicial:

- Realiza la carga inicial de datos, asegurando que todos los datos se transformen y carguen según lo planificado.

Desarrollo de Reportes y Dashboards:

- Configura herramientas de BI para crear reportes y dashboards basados en los datos del DW.
- Asegura que los reportes cumplan con los requisitos de negocio y sean fáciles de usar.

Pruebas y Validación:

- Realiza pruebas exhaustivas para verificar la precisión de los datos, el rendimiento del sistema, y la seguridad.

- Valida que los datos cargados y los reportes generados cumplan con los requisitos definidos en la fase de análisis.

Capacitación de Usuarios:

- Capacita a los usuarios finales y al equipo de TI en el uso del Data Warehouse y las herramientas asociadas.
- Despliegue y Mantenimiento:
- Lanza el DW en un entorno de producción.
- Establece un plan de mantenimiento y monitoreo para asegurar el rendimiento continuo del sistema y la calidad de los datos.

Mantenimiento y Evolución Continua

Monitoreo Continuo:

- Monitorea el rendimiento del DW y el uso de los datos.
- Realiza ajustes en los procesos ETL y el modelo de datos según sea necesario.

Actualización y Optimización:

- Implementa mejoras basadas en el feedback de los usuarios y el análisis de rendimiento.
- Escala el DW para soportar el crecimiento del negocio y la incorporación de nuevas fuentes de datos.