

**UNIVERSIDAD AUTÓNOMA “GABRIEL RENÉ MORENO” FACULTAD DE CIENCIAS
DE LA COMPUTACIÓN Y TELECOMUNICACIONES**

CARRERA INGENIERÍA EN SISTEMAS E INFORMÁTICA



METODOLOGIAS PARA LA CONSTRUCCION DATA WAREHOUSE

MATERIA: Sistemas para el Soporte a la toma de decisiones

DOCENTE: Ing. Peinado Pereira Miguel Jesús

GRUPO: SA

INTEGRANTES:

REGISTRO:

- Mamani Rodriguez Kasandra

216164737

2-2024

SANTA CRUZ – BOLIVIA

METODOLOGÍA DE RALPH KIMBALL (Enfoque Bottom-Up)

La Metodología de Ralph Kimball, también conocida como el enfoque Bottom-Up, es una estrategia ampliamente utilizada para la implementación de Data Warehouses en organizaciones. Este enfoque se centra en la creación inicial de Data Marts específicos para áreas funcionales del negocio, como ventas, finanzas, marketing, entre otros, que se van integrando progresivamente para formar un Data Warehouse centralizado y cohesionado.

Descripción Enfoque Bottom-Up de Kimball:

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:

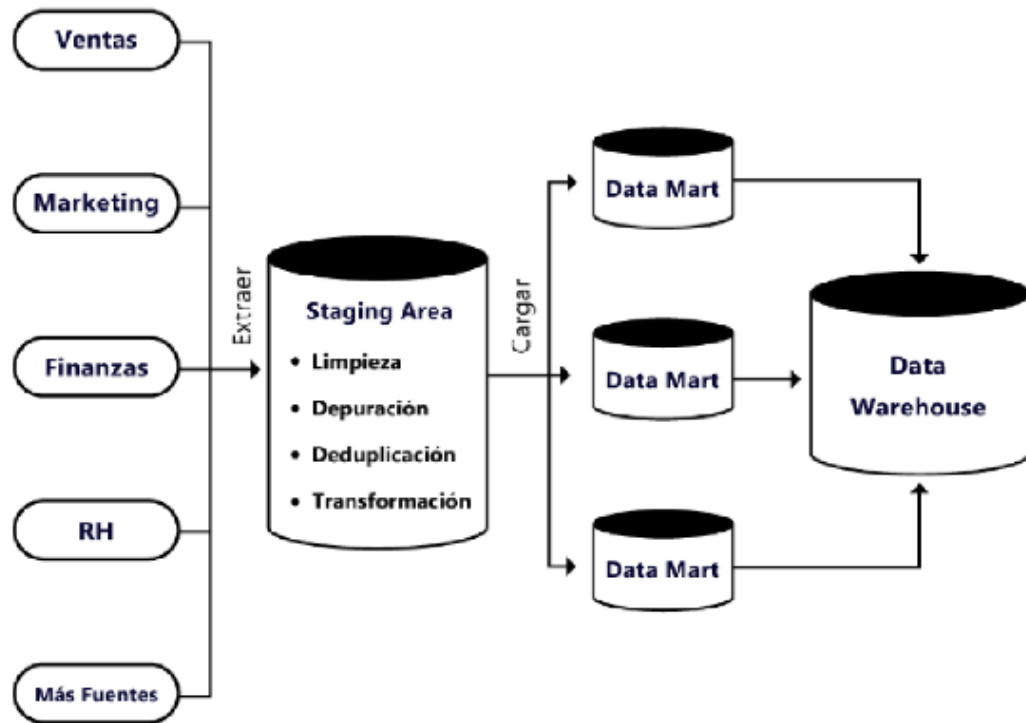
- Identificación de las necesidades de negocio y áreas funcionales.
- Diseño de Data Marts dimensionales (con esquemas estrella o copo de nieve).
- Integración de los Data Marts en un DW centralizado.
- Implementación de procesos ETL (Extract, Transform, Load) para la carga de datos.

Ventajas del Enfoque Bottom-Up de Kimball:

- **Implementación Rápida:** Permite a las organizaciones comenzar con áreas funcionales críticas, obteniendo valor a corto plazo.
- **Flexibilidad:** Es adaptable y puede expandirse a medida que las necesidades del negocio evolucionan.
- **Escalabilidad:** A medida que se integran más Data Marts, el Data Warehouse puede crecer y adaptarse a mayores volúmenes y complejidades de datos.

Desventajas del Enfoque Bottom-Up de Kimball:

- **Complejidad en la Integración:** La unificación de múltiples Data Marts en un Data Warehouse centralizado puede ser compleja, especialmente si no se gestionan bien las dimensiones comunes desde el principio.
- **Duplicación de Esfuerzos:** La falta de una visión centralizada inicial puede llevar a la duplicación de esfuerzos en la creación y mantenimiento de dimensiones comunes.



METODOLOGÍA DE BILL INMON (ENFOQUE TOP-DOWN)

La Metodología de Bill Inmon, conocida como el enfoque Top-Down, es una estrategia integral y metódica para la construcción de Data Warehouses. Este enfoque parte de la creación de un Data Warehouse (DW) centralizado y normalizado a nivel corporativo, y luego se derivan Data Marts específicos según las necesidades del negocio.

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.

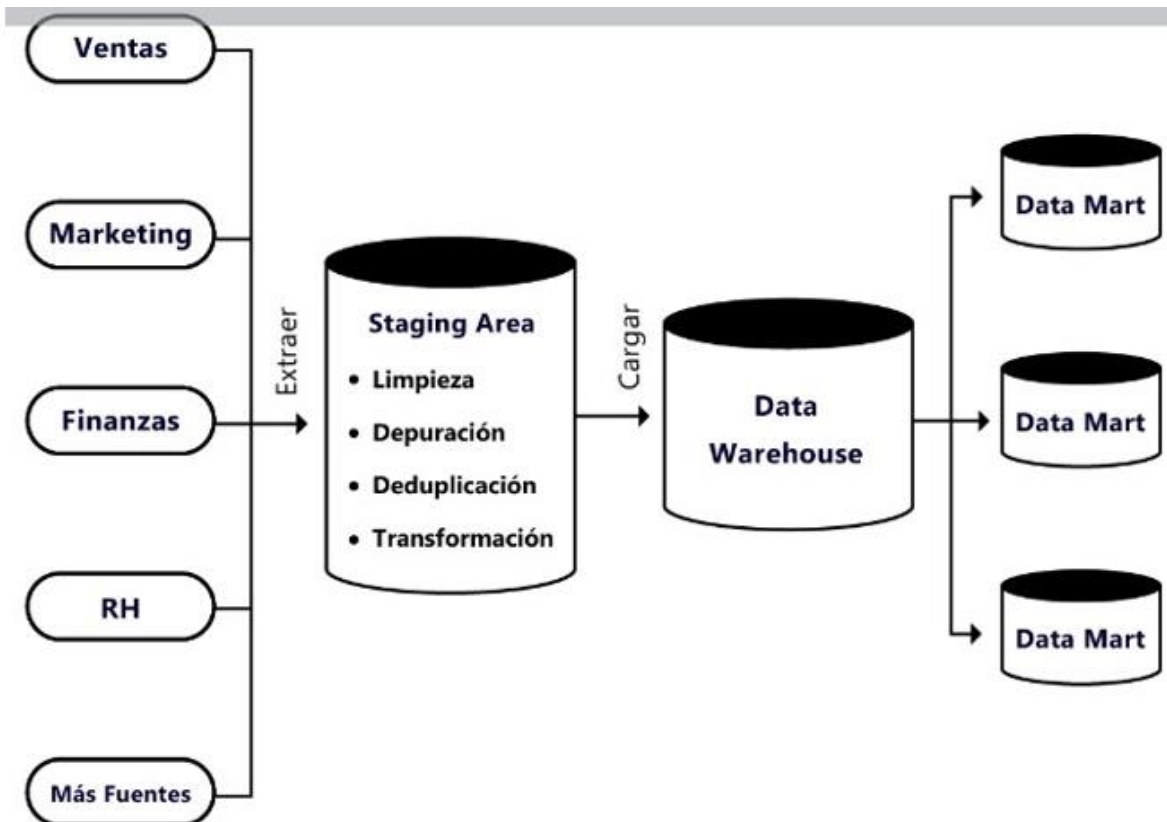
Ventajas del Enfoque Top-Down de Inmon:

- **Visión a Largo Plazo:** Este enfoque permite la creación de un DW robusto y alineado con la visión estratégica de la organización, facilitando la evolución y expansión del sistema.
- **Minimización de Redundancia:** El diseño normalizado evita la duplicación de datos, lo que mejora la integridad y calidad de los mismos.

- **Consistencia y Coherencia:** Al centralizar los datos en un DW antes de derivar Data Marts, se asegura una coherencia total en la información utilizada por diferentes áreas funcionales.

Desventajas del Enfoque Top-Down de Inmon:

- **Tiempo de Implementación:** La planificación y desarrollo de un DW centralizado y normalizado es un proceso largo y puede retrasar la entrega de valor inicial.
- **Mayor Complejidad Inicial:** El enfoque requiere un análisis exhaustivo y un diseño complejo desde el principio, lo que puede ser difícil de gestionar y ejecutar.



METODOLOGÍA ÁGIL (AGILE BI)

La Metodología Ágil (Agile BI) es una adaptación de los principios ágiles al desarrollo de Data Warehouses y soluciones de Business Intelligence (BI). Este enfoque se centra en la entrega rápida y continua de valor a través de desarrollos incrementales e iterativos, permitiendo una mayor adaptabilidad a las necesidades cambiantes del negocio.

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 de Agile BI:

- **Adaptabilidad:** El enfoque iterativo permite adaptarse rápidamente a los cambios en los requisitos del negocio y a nuevas oportunidades de mejora.
- **Rapidez en la Entrega de Resultados:** Al entregar incrementos funcionales del Data Warehouse en cada sprint, los usuarios pueden empezar a obtener valor rápidamente sin tener que esperar a la finalización del proyecto completo.
- **Mejor Alineación con las Necesidades del Negocio:** El feedback continuo y la colaboración estrecha con los usuarios aseguran que el proyecto esté siempre alineado con lo que el negocio realmente necesita.

Desventajas de Agile BI:

- **Dificultad en Grandes Empresas:** La implementación de Agile BI puede ser desafiante en organizaciones grandes con procesos bien establecidos y jerarquías rígidas, donde el cambio hacia una metodología ágil puede encontrar resistencia.
- **Posible Falta de Planificación a Largo Plazo:** Si no se gestiona correctamente, el enfoque ágil puede llevar a una falta de visión a largo plazo y una acumulación de deuda técnica, debido a la entrega continua de pequeños incrementos.

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:

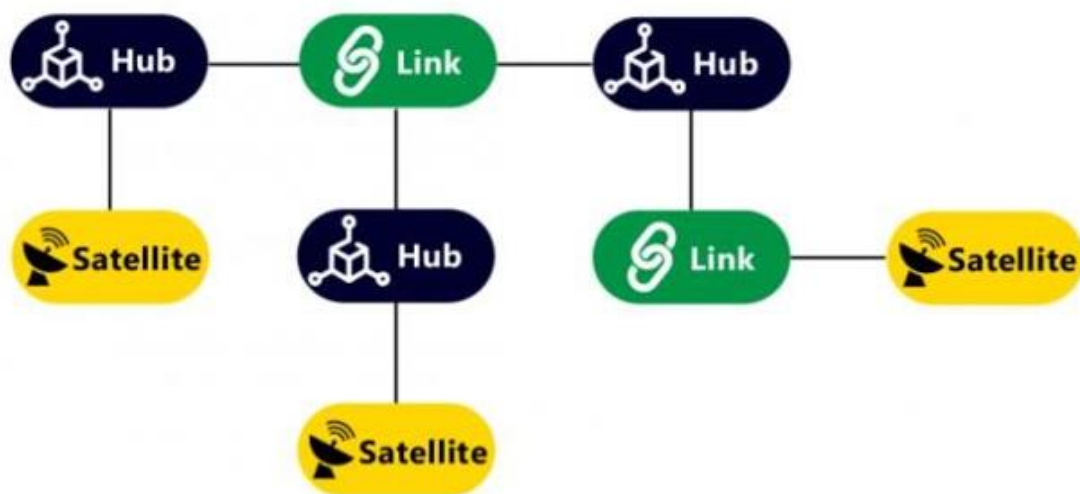
- Modelado del Data Warehouse utilizando componentes principales: Hub (entidades principales), Satellites (atributos de las entidades) y Links (relaciones entre entidades).
- Implementación de procesos ETL que se adaptan a la estructura modular del Data Vault.
- Creación de vistas o Data Marts específicos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

Ventajas de la Metodología Data Vault:

- **Escalabilidad:** Data Vault está diseñado para manejar grandes volúmenes de datos y se adapta bien a la integración de nuevas fuentes de datos a medida que la organización crece.
- **Flexibilidad:** Permite cambios en las fuentes de datos sin afectar el modelo central, lo que facilita la evolución y adaptación del Data Warehouse a lo largo del tiempo.
- **Adaptabilidad:** El enfoque modular facilita la adaptación a cambios en el negocio y la incorporación de nuevas funcionalidades sin necesidad de reestructurar el Data Warehouse.

Desventajas de la Metodología Data Vault:

- **Curva de Aprendizaje Pronunciada:** La metodología Data Vault puede ser compleja y requiere un entendimiento profundo del modelado de datos, lo que puede ser un desafío para equipos nuevos en la metodología.
- **Complejidad en Proyectos Pequeños:** Para proyectos de menor escala, la implementación de Data Vault puede ser excesivamente compleja y no proporcionar suficientes beneficios en comparación con otras metodologías más simples.



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 de la Metodología Lean Data Warehousing:

- **Enfoque en la Eficiencia:** Al eliminar tareas y procesos innecesarios, la metodología Lean reduce el tiempo y los costos asociados con la construcción del Data Warehouse, permitiendo un desarrollo más rápido y eficiente.
- **Reducción de Desperdicios:** Se enfoca en optimizar recursos y minimizar el trabajo desperdiciado, asegurando que cada actividad realizada aporte valor directo al cliente.
- **Adaptabilidad:** La entrega continua y las mejoras incrementales permiten que el Data Warehouse se ajuste rápidamente a las necesidades cambiantes del negocio.

Desventajas de la Metodología Lean Data Warehousing:

- **Requiere una Cultura Organizacional Enfocada en la Mejora Continua:** Para que la metodología Lean funcione, la organización debe estar comprometida con los principios de mejora continua, lo que puede ser un desafío en entornos más tradicionales o jerárquicos.
- **Potencial de Sub-optimización:** Al enfocarse tanto en la eficiencia y la eliminación de desperdicios, existe el riesgo de sub-optimizar algunas áreas o de no prestar suficiente atención a aspectos críticos que no son inmediatamente visibles como fuentes de valor.

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 de la Metodología Híbrida:

- **Equilibrio entre agilidad y robustez:** Combina la capacidad de responder rápidamente a las necesidades del negocio con la estabilidad y consistencia de un modelo de datos centralizado.
- **Flexibilidad:** Permite que las organizaciones se beneficien de los enfoques de Kimball y Inmon, adaptándose a diferentes circunstancias y requisitos a lo largo del tiempo.
- **Escalabilidad:** Facilita la evolución del Data Warehouse sin sacrificar la integridad de los datos o la coherencia a largo plazo.

Desventajas de la Metodología Híbrida:

- **Gestión Compleja:** La implementación de una metodología híbrida puede ser difícil de gestionar debido a la necesidad de equilibrar y coordinar múltiples enfoques y procesos.

- **Requiere Planificación Cuidadosa:** Para que la metodología híbrida sea efectiva, es esencial una planificación meticulosa y una comprensión clara de las necesidades a corto y largo plazo del negocio.

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.