

Data Management Fundamentals

Módulo 13 – Data Warehousing y Business Intelligence



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas



Índice

- 1. INTRODUCCIÓN**
- 2. FRAMEWORK DE DAMA**
- 3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS**
- 4. ENTREGABLES**
- 5. ACTIVIDADES**
- 6. HERRAMIENTAS**
- 7. TÉCNICAS**
- 8. GUÍAS DE IMPLEMENTACIÓN**
- 9. MÉTRICAS**

Índice

- 1. INTRODUCCIÓN**
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

¿Qué es Business Intelligence?



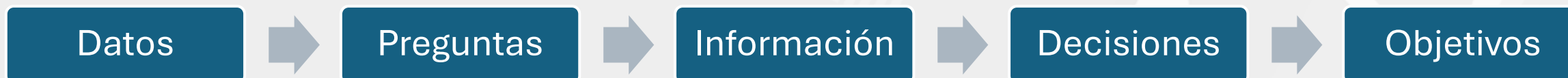
“Tiene dos significados:

- En primer lugar, se refiere a un tipo de **análisis de datos** destinado a comprender las **actividades** y **oportunidades** de la organización. Los resultados de este análisis se utilizan para **impulsar el éxito** de la organización. [...]
- En segundo lugar, BI se refiere a un **conjunto de tecnologías** que soportan este tipo de análisis de datos.”

Es un tipo de análisis de datos destinado a comprender las **actividades** y **oportunidades** de la organización.

Pasado y presente

Futuro



Ejemplo: en una cadena de restaurantes podemos **analizar las ventas** de cada sucursal para identificar:

- **Identificar los platos más populares** para hacer **promociones**
- **Ajustar el inventario** para **reducir costes** y **evitar desperdicios**
- **Ajustar la plantilla** para dar **mejor servicio**
- **Detectar cambios** de comportamiento de clientes para **adaptarnos**

Características generales del BI

- **Visualización de datos:** Presenta la información a través de gráficos, tablas y dashboards interactivos.
- **Análisis avanzado:** Utiliza técnicas estadísticas y algoritmos para identificar patrones y tendencias.
- **Accesibilidad:** Permite a los usuarios acceder a la información en tiempo real desde cualquier dispositivo.

Objetivos del BI

- **Mejora de la toma de decisiones:** Proporciona información precisa y oportuna para apoyar las decisiones empresariales.
- **Optimización de procesos:** Identifica oportunidades de mejora y eficiencia dentro de la organización.
- **Aumento de la competitividad:** Ayuda a responder rápidamente a los cambios del entorno competitivo (mercado y competencia).
- **Reducción de costos:** Permite identificar y eliminar ineficiencias operativas.

¿Qué es un DataWareHouse (DWH)?

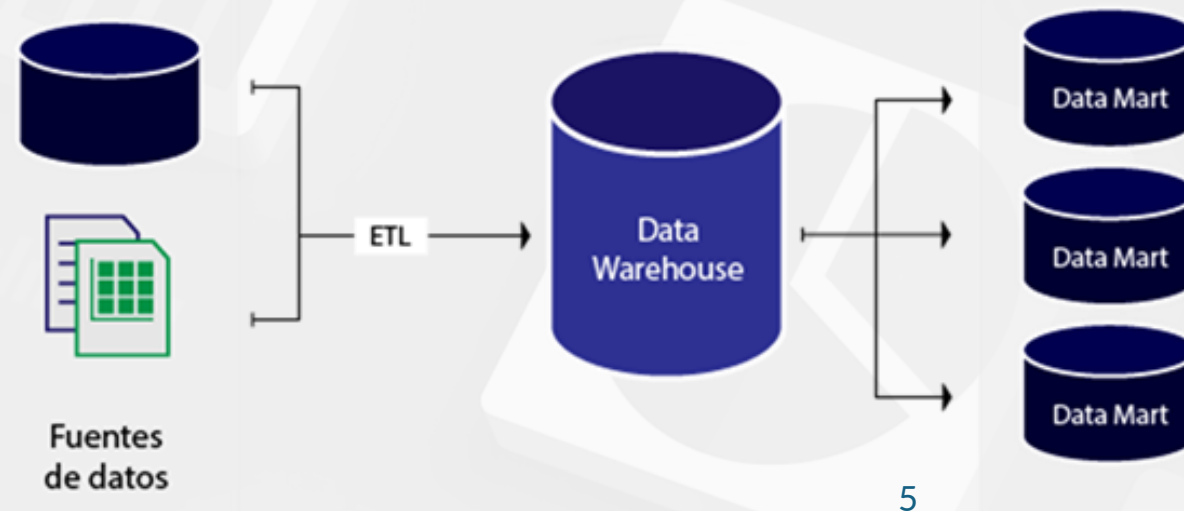


El concepto **DWH** tiene dos componentes principales:

- Un **almacén de datos centralizado** que da soporte a la toma de decisiones
- El **software asociado** para recopilar, limpiar, transformar y almacenar datos desde una variedad de fuentes, operacionales o externas.



Un **DataMart** es una **copia de un subconjunto** de datos del DWH diseñada para **soportar tipos de análisis concretos** y que proveen una solución específica y acotada a un área de la empresa .



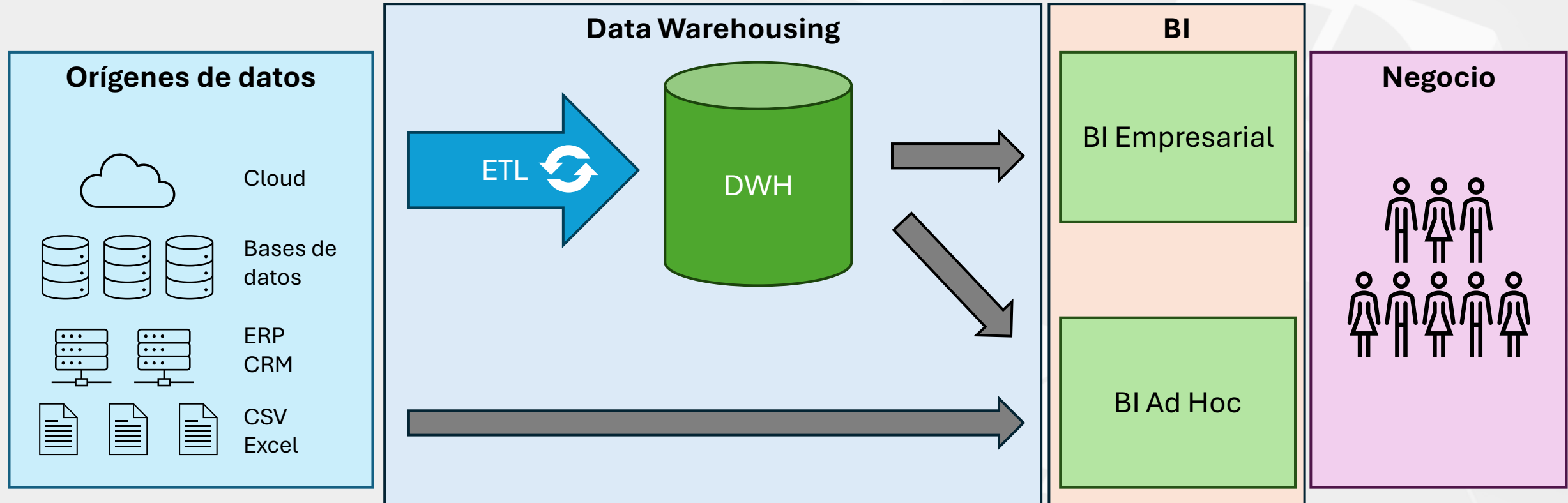
Características generales DWH

- Enfoque global y transversal a la empresa.
- Almacena datos de otros sistemas.
- Organiza los datos para aumentar su valor.
- Facilita el acceso y uso de los datos para el análisis.
- Contiene datos integrados y fiables.
- Contiene datos históricos.
- Variante en el tiempo (fotos de cada instante)
- No volátil (no se modifica)

Objetivos del DWH

- Dar soporte a la actividad de BI
- Disponibilizar el análisis de negocio y la toma de decisiones de forma efectiva
- Encontrar formas de innovar basadas en la información de sus datos

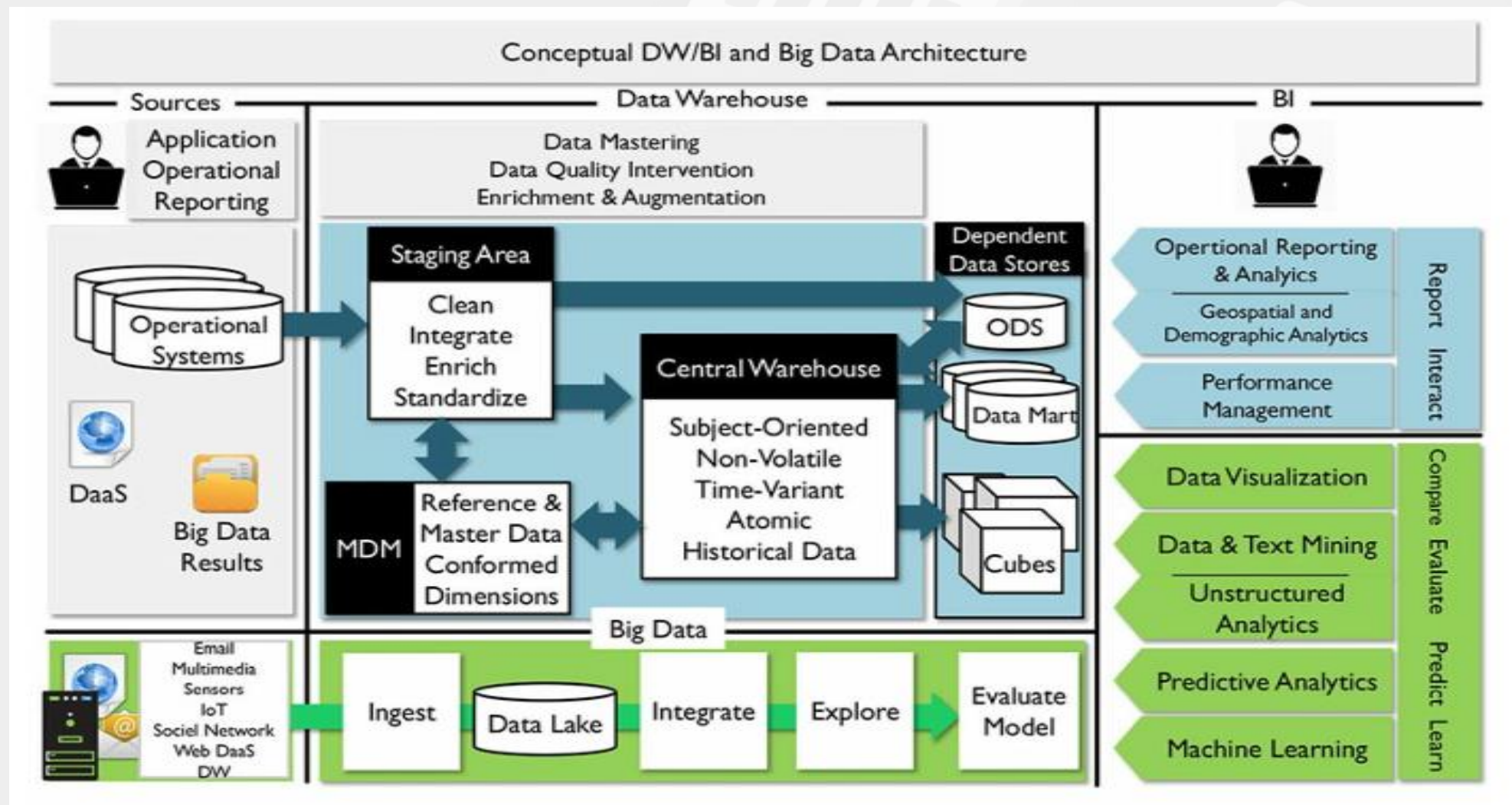
Data Warehousing y BI



PREGUNTA

- ✗ Business Intelligence (BI) se enfoca únicamente en la recolección de datos, sin analizarlos. **BI no solo recopila datos, sino que los analiza para apoyar la toma de decisiones**
- ✓ El proceso de BI implica la extracción, transformación y carga (ETL) de datos antes de su análisis.
- ✓ BI ayuda a las empresas a identificar tendencias, patrones y oportunidades de negocio a partir de los datos.
- ✓ Los dashboards e informes generados por BI permiten visualizar información clave para la toma de decisiones.
- ✗ El BI se centra principalmente en el análisis del pasado y no tiene aplicaciones en la predicción de tendencias futuras. **BI incluye análisis descriptivo, pero también puede incorporar modelos predictivos y de machine learning**

Componentes de arquitectura (DMBoK)



PREGUNTA

- ✗ Un Data Warehouse es un sistema diseñado para el procesamiento transaccional en tiempo real (OLTP). **Un DWH está diseñado para análisis y toma de decisiones (OLAP)**
- ✓ Los datos en un Data Warehouse tienen las siguientes características (entre otras): son integrados, históricos y no volátiles.
- ✓ Un Data Warehouse almacena datos de múltiples fuentes para proporcionar una visión consolidada de la organización.
- ✗ Los datos en un Data Warehouse se actualizan en tiempo real para reflejar cambios inmediatos. **Generalmente, los datos se cargan en lotes o en períodos definidos, no en tiempo real**
- ✓ El propósito principal de un Data Warehouse es facilitar la toma de decisiones estratégicas.

Enfoques de diseño DWH: Inmon y Kimball

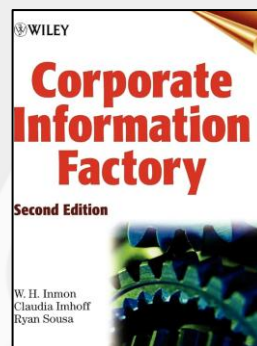
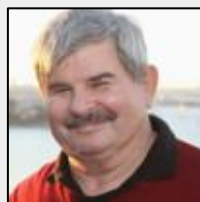
- Las dos grandes corrientes de diseño de DWH fueron diseñadas por **Bill Inmon** y **Ralph Kimball** respectivamente en los 90
- Comparten características comunes, pero tienen un enfoque distinto.

Bill Inmon: enfoque Top-Down

CIF (Corporate Information Factory).

Se basa en la creación un DWH corporativo con el modelo lógico completo de cada entidad, los datos se exponen en DataMarts.

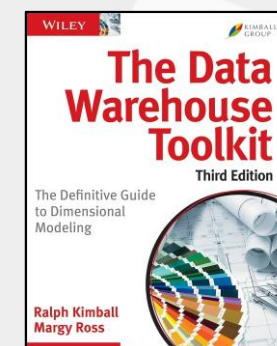
- Basado en el modelo en copo de nieve
- Datos normalizados
- Enfoque de arriba hacia abajo (top-down)



Ralph Kimball: enfoque Bottom-Up

Se basa en el diseño modular de DataMarts (estrellas) que se van uniendo para formar el DWH.

- Basado en el modelo en estrella
- Datos desnormalizados
- Enfoque de abajo hacia arriba (bottom-up)
- El DWH es una conjunción de DataMarts
- Coherencia e interoperabilidad a través de dimensiones conformadas



Enfoque Inmon: Características del DWH

Inmon identifica las siguientes diferencias de un DWH respecto a los sistemas operacionales:



Orientado a un tema: Organizado en base a entidades, no a procesos.



Integrado: Los datos están unificados y cohesionados.



Variante de tiempo: Los datos se almacenan como en una fotografía instantánea. Consultar datos basados en un periodo de tiempo específico siempre producirá el mismo resultado.



No volátil: Los registros normalmente no se actualizan. Un conjunto de registros puede representar diferentes estados de una misma transacción.



Datos agregados y detallados: Los datos se almacenan al máximo nivel de granularidad posible, pero también se crean agregaciones de datos.



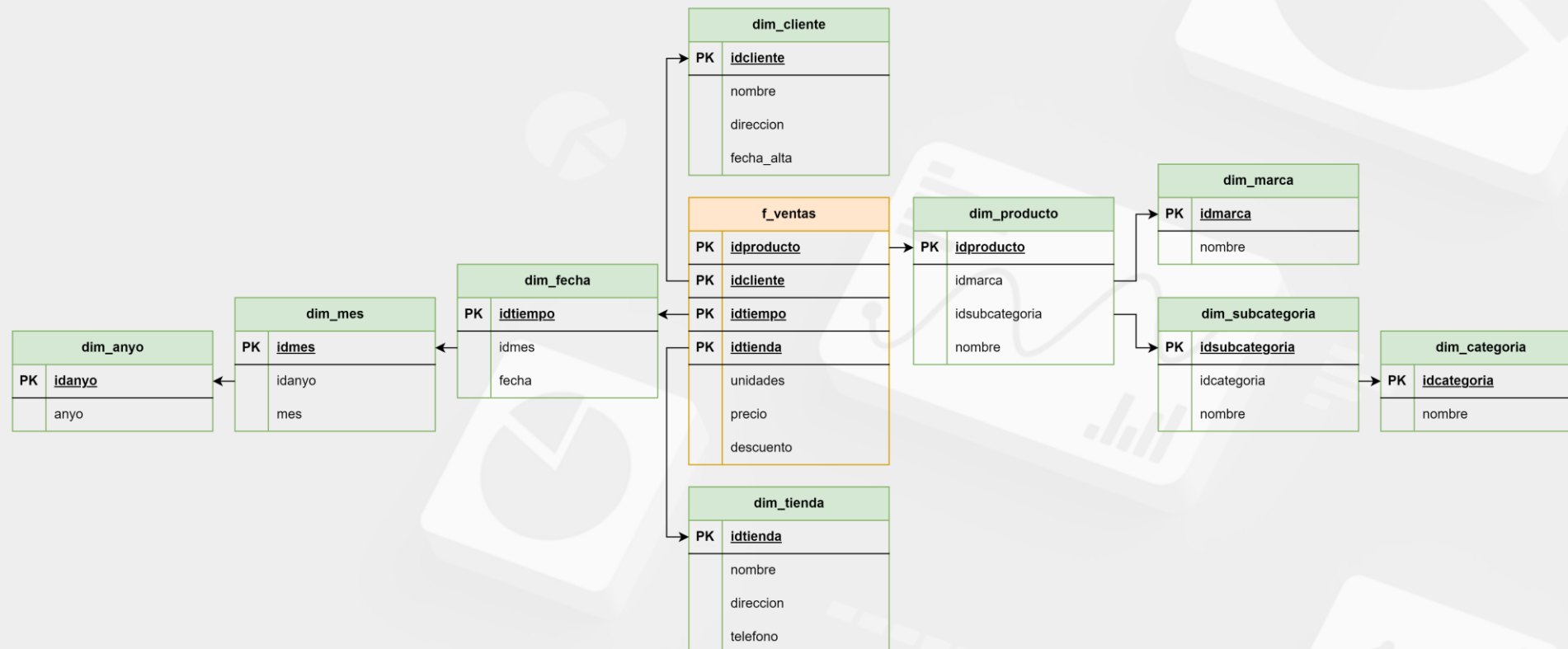
Histórico: Almacena gran cantidad de información histórica además de la actual.

Enfoque Inmon: Modelo en copo de nieve

Se basa en la normalización de los datos (3NF). Tiene tablas de hechos y dimensiones como el modelo en estrella.

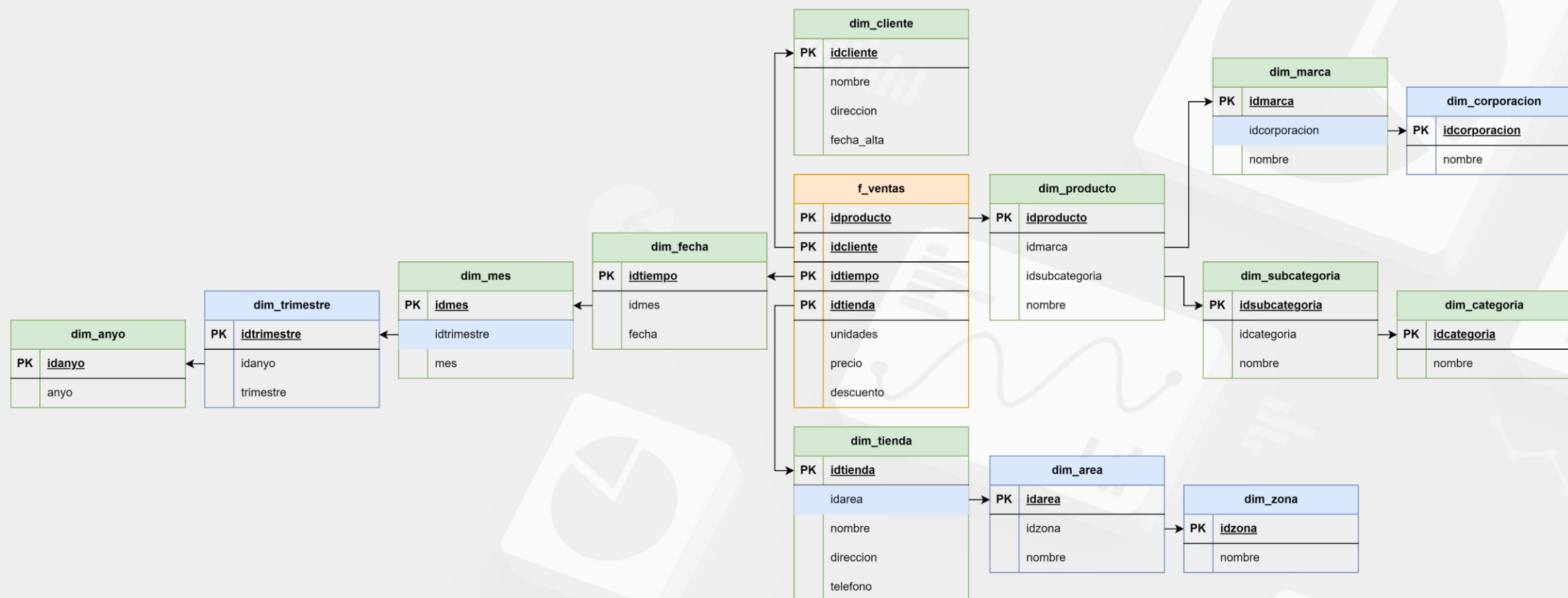
Negocio integrado en la estructura: La propia estructura de las tablas y sus relaciones permite identificar las relaciones de las entidades de negocio y sus dependencias.

Cambios de negocio simples: Aplicar cambios de negocio tales como estructuras o nuevas agrupaciones es sencillo.



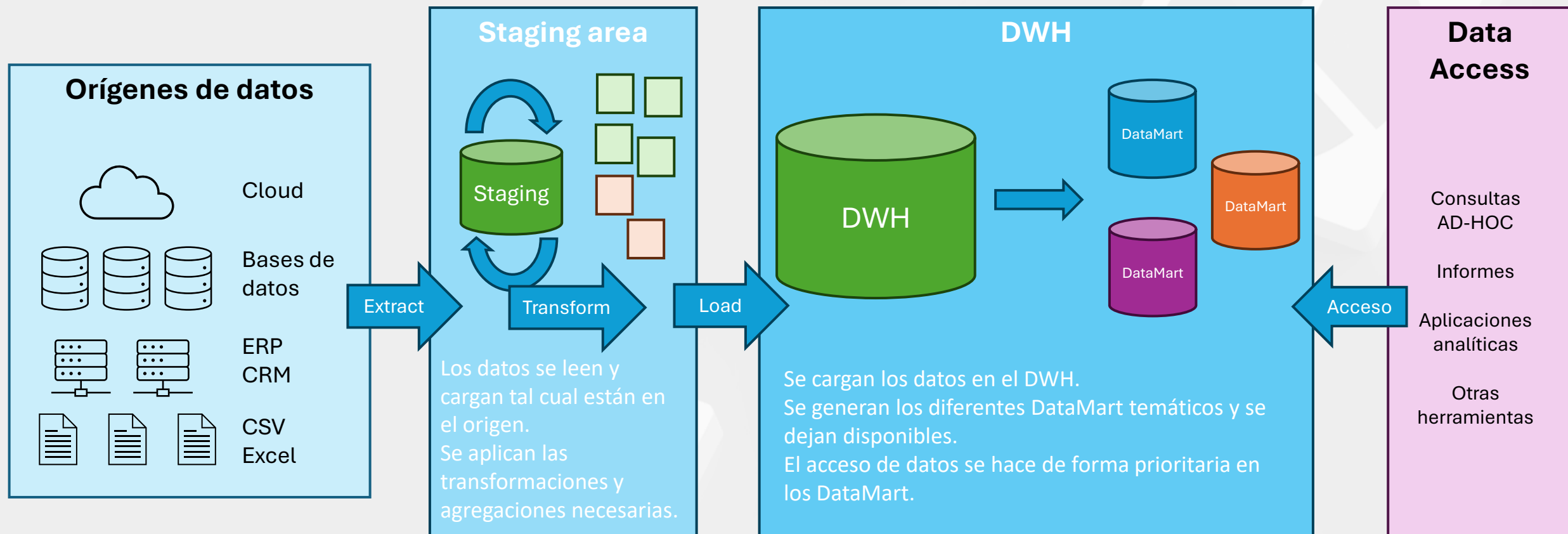
Enfoque Inmon: Ejemplo de cambios

Se pueden incluir nuevos niveles en las jerarquías sin afectar los desarrollos existentes y cumpliendo con los requerimientos del negocio.



Enfoque Inmon: Arquitectura

La transformación se realiza en el área de Staging. Una vez se han generado las tablas y objetos necesarios se cargan en el DWH. A continuación, se generan los DataMart:



Enfoque Inmon: Arquitectura en detalle

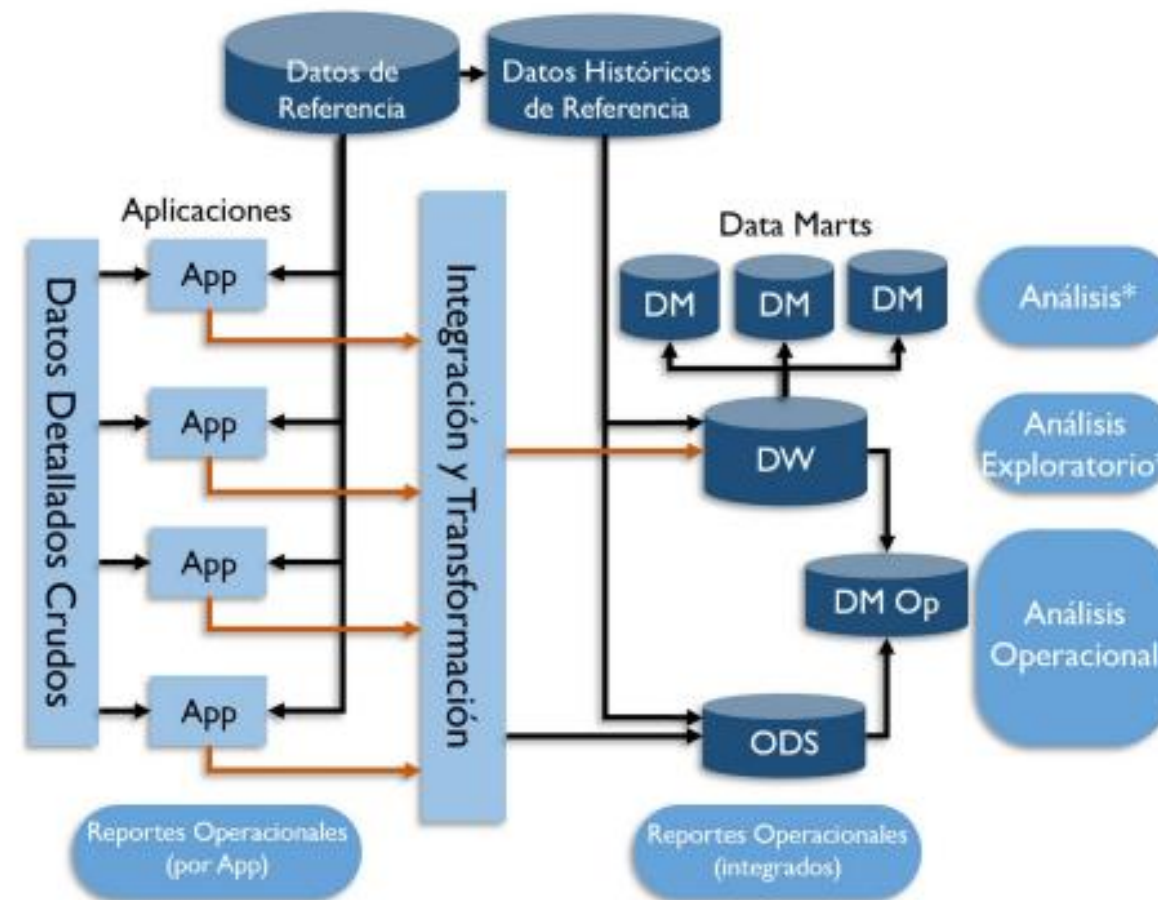


Figura 80 La Fábrica de Información Corporativa

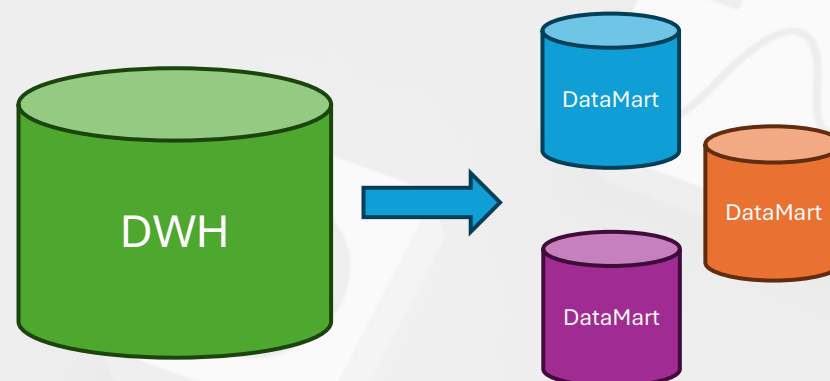
Enfoque Inmon: Ventajas e inconvenientes

Ventajas

- El DWH integrado sirve como fuente única de la verdad a nivel corporativo.
- Poca redundancia de datos al estar normalizado.
- Flexible y capaz de adoptar los cambios de negocio de forma simple.

Inconvenientes

- La puesta en marcha inicial es más lenta que en el enfoque Kimball.
- El modelado requiere de una abstracción y conocimiento de negocio importante.
- Los procesos de ETL son más complejos y tienen pasos adicionales (creación de DataMarts).



PREGUNTA

- ✓ El enfoque de Inmon propone un Data Warehouse centralizado y normalizado.
- ✗ Inmon sugiere que los Data Marts deben diseñarse primero y luego integrarse en un Data Warehouse. **Inmon propone primero el Data Warehouse corporativo y luego los Data Marts**
- ✗ En el modelo de Inmon, los datos se almacenan en estructuras altamente desnormalizadas. **Se almacenan en un modelo normalizado, generalmente en 3FN**
- ✓ Inmon enfatiza la integridad de los datos y la reducción de la redundancia mediante un modelo relacional en copo de nieve.
- ✗ El enfoque de Inmon facilita consultas rápidas y análisis directo en sus bases de datos.
Al estar normalizado, no está optimizado para consultas rápidas; se diseñan Data Marts para ello

Enfoque Kimball: Premisas

Las premisas de este enfoque son:



Basado en el modelo en estrella: con menos tablas que el copo de nieve.



Datos desnormalizados: se repiten datos.



Enfoque de abajo hacia arriba (bottom-up).



El DWH es una conjunción de DataMarts. Se diseñan los DataMarts por separado.



Coherencia e interoperabilidad a través de dimensiones conformadas

Enfoque Kimball: Características del DWH



Diseño por capas

El enfoque Kimball se basa en un diseño por capas en el que se crean primero los data marts, facilitando la organización de datos.



Integración en Data Warehouse

Después de construir los data marts, estos se integran en un Data Warehouse, lo que permite una visión unificada de los datos.



Metodología Ágil

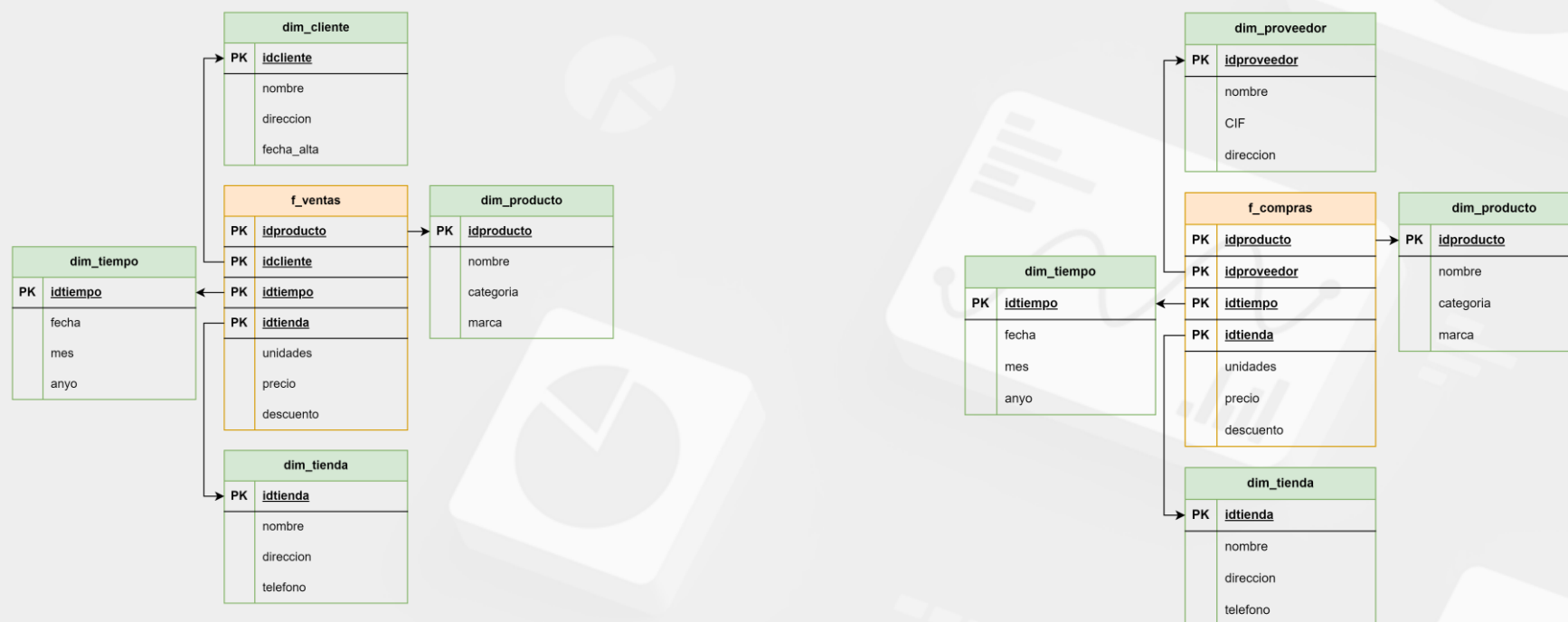
El enfoque Kimball es ágil y permite realizar análisis de manera más sencilla, adaptándose rápidamente a los cambios en los requisitos de negocio.

Enfoque Kimball: Modelo en estrella

Se basa en tablas de hechos y dimensiones, una o varias estrellas componen un DataMart:

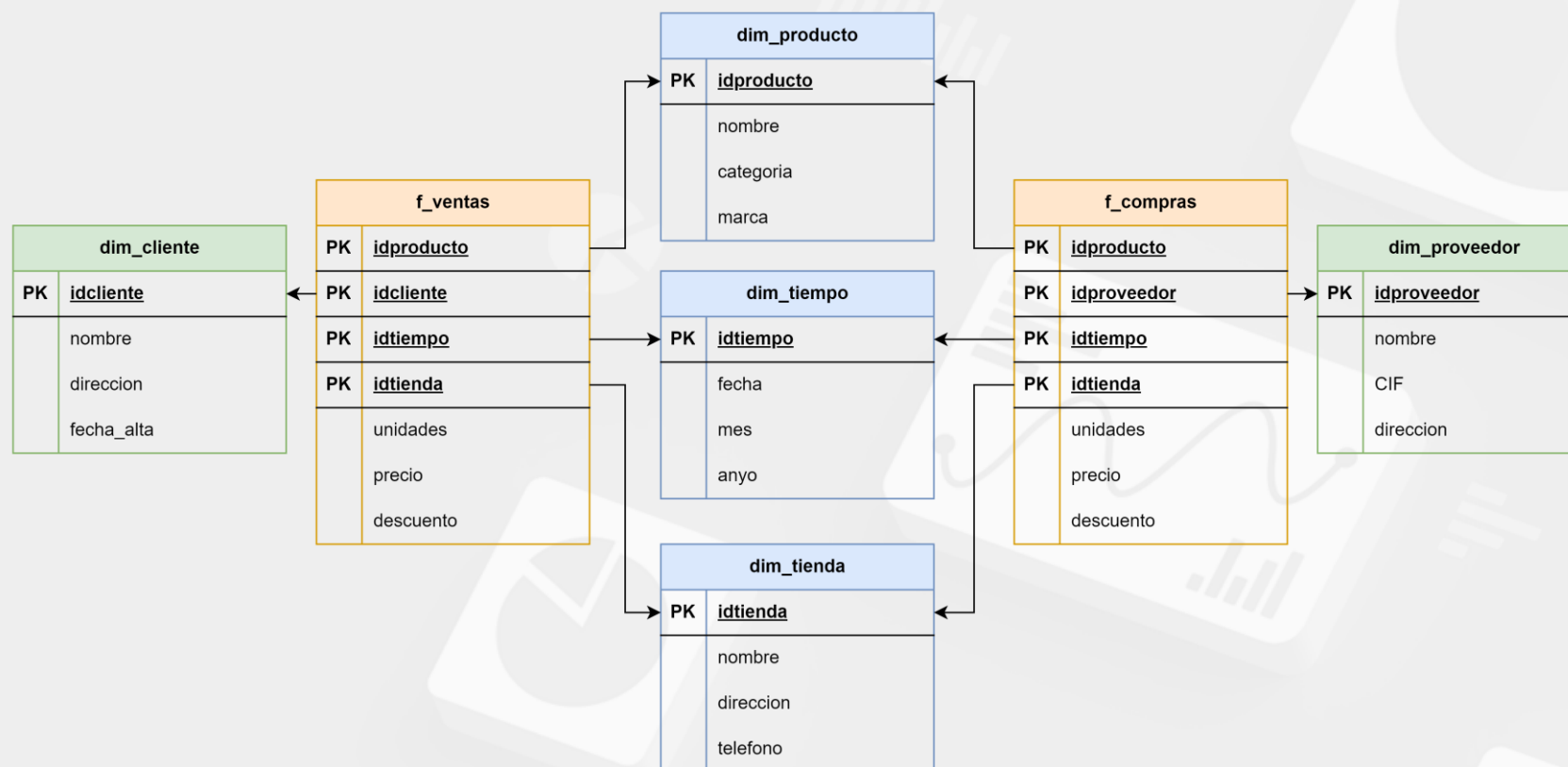
Tabla de hechos: Contiene los eventos cuantificables y las medidas, datos transaccionales numéricos.

Tabla de dimensión: Da forma a un eje de análisis, permitiendo agregar o desglosar los datos de las tablas de hechos.



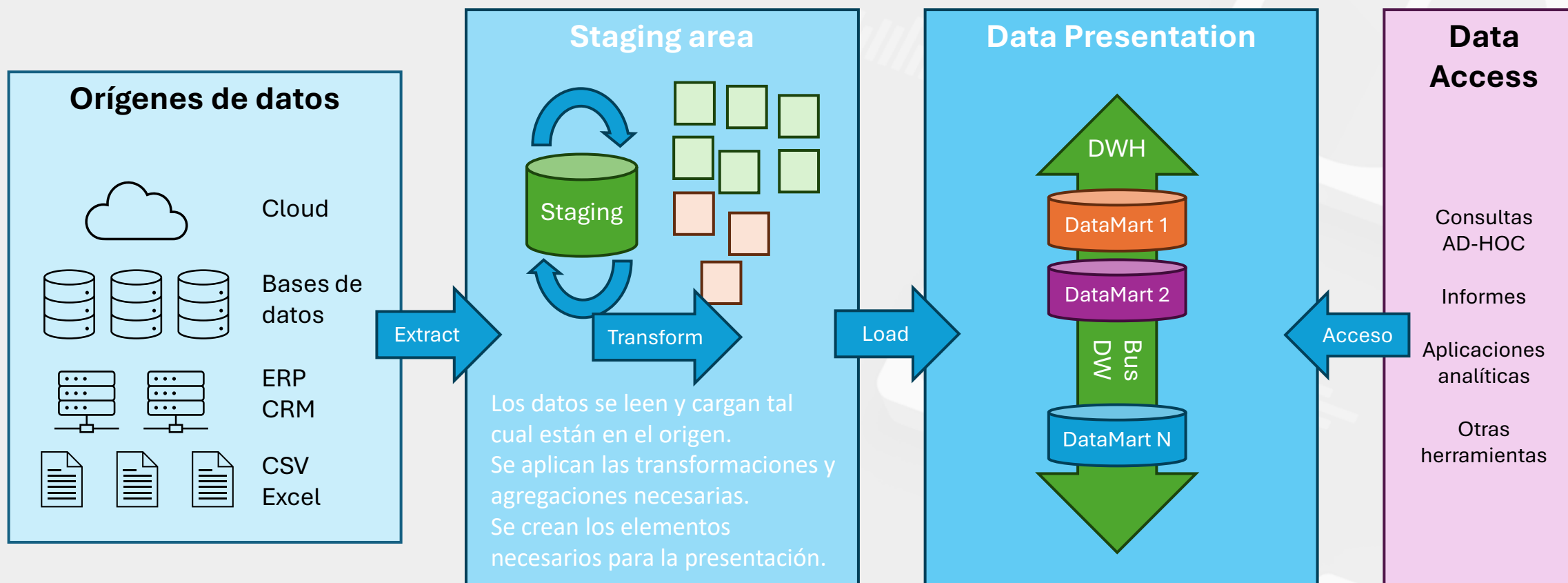
Enfoque Kimball: Dimensiones conformadas

Cuando diferentes estrellas comparten dimensiones se denominan **dimensiones conformadas**. Estas dimensiones permiten **cruzar los datos de múltiples estrellas** proporcionando coherencia e interoperabilidad a lo largo del DWH. Se denomina también **BUS DW** al conjunto de dimensiones conformadas.

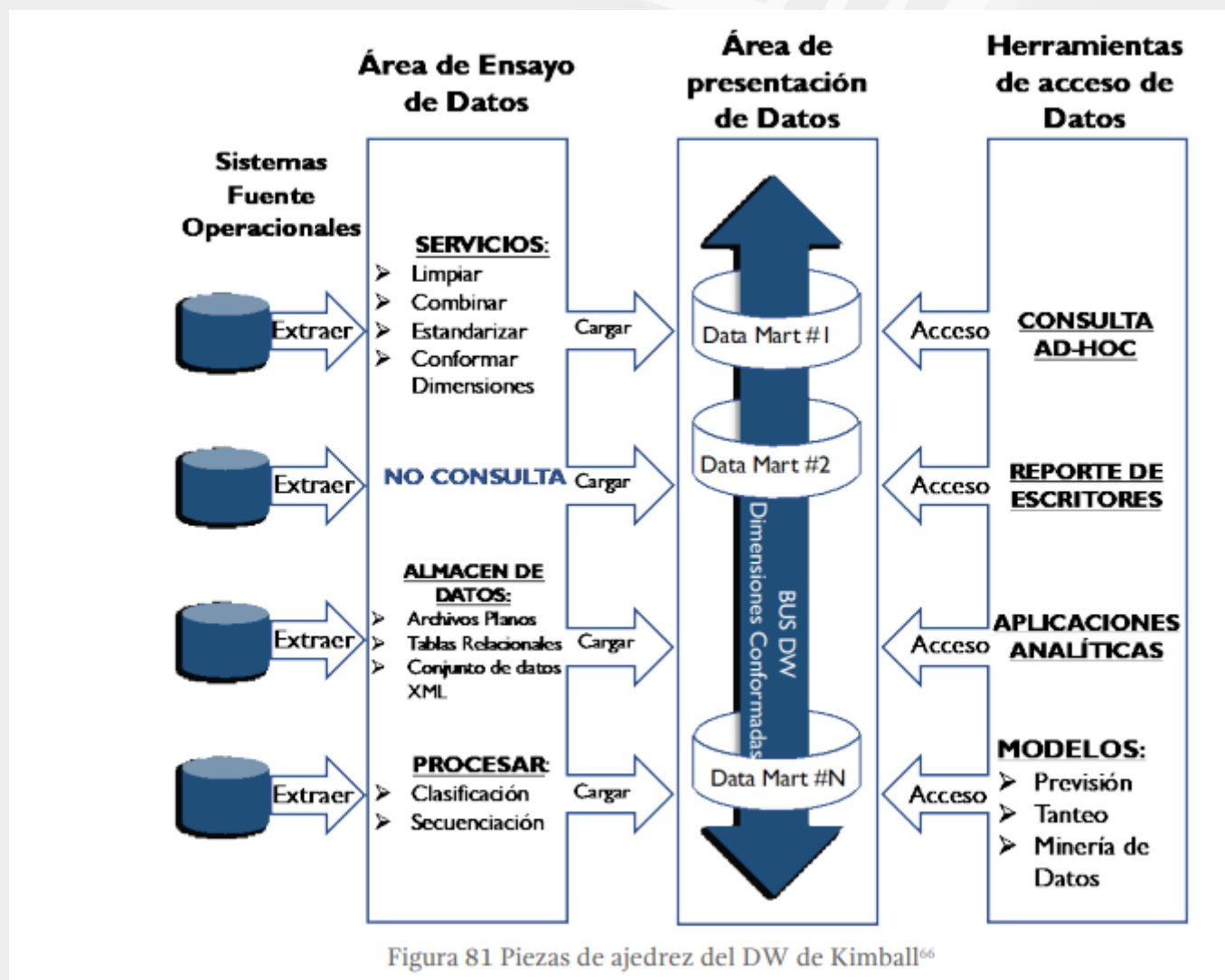


Enfoque Kimball: Arquitectura

La transformación se realiza en el área de Staging. Una vez se han generado las tablas y objetos necesarios se cargan en los DataMart correspondientes. El conjunto de DataMarts relacionados por las dimensiones conformadas componen el DWH:



Enfoque Kimball: Arquitectura en detalle



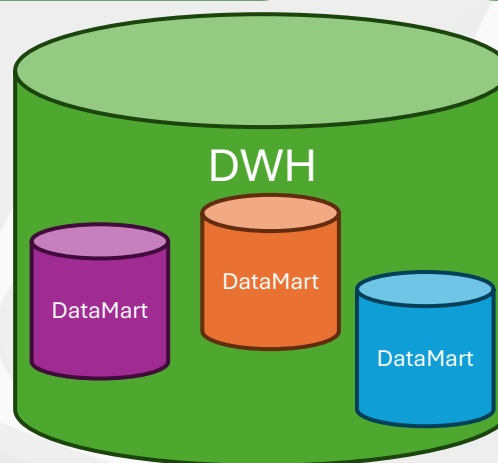
Enfoque Kimball: Ventajas e inconvenientes

Ventajas

- El modelo dimensional es rápido de construir.
- El esquema en estrella es sencillo de comprender y simplifica las consultas.
- Gracias a las dimensiones conformadas se pueden utilizar para realizar consultas a lo largo de múltiples estrellas.

Inconvenientes

- La actualización de datos es problemática y puede generar incoherencias.
- Al estar los datos desnormalizados se pierde el concepto de única fuente de la verdad.
- Los cambios por necesidades de negocio pueden requerir de actualizaciones complejas del modelo.



PREGUNTA

-  El enfoque de Kimball sugiere construir primero un Data Warehouse corporativo y luego crear Data Marts derivados. **Kimball propone construir Data Marts primero, que luego forman un DWH**
-  El modelo de Kimball se basa en esquemas estrella para la modelización de datos.
-  Kimball enfatiza la desnormalización de los datos para mejorar la velocidad de consulta.
-  El enfoque de Kimball no permite integrar fácilmente diferentes áreas de negocio en un mismo DWH. **Se hace a través de las dimensiones conformadas**
-  En el modelo de Kimball, los Data Marts se diseñan primero y se integran en un Data Warehouse lógico.

Índice

1. INTRODUCCIÓN
2. **FRAMEWORK DE DAMA**
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

DWH y BI en el Framework de Dama

Capítulo 11

El concepto surge en los años 80, cuando la tecnología empezó a permitir a las organizaciones integrar datos en un modelo común.

Ventajas

- Proporcionar información sobre la operación: visión unificada de las operaciones.
- Soporte a la toma de decisiones.
- Reducir la redundancia: evitamos llevar datos de un sistema a otro.
- Mejorar la consistencia: se integra una sola vez, sin duplicados.

Comienzan a construirse en los años 90, conjuntamente con los primeros sistemas de BI.

Es un concepto en evolución:

DataWarehouse



DataLake



DataLakeHouse



Logical DataWareHouse

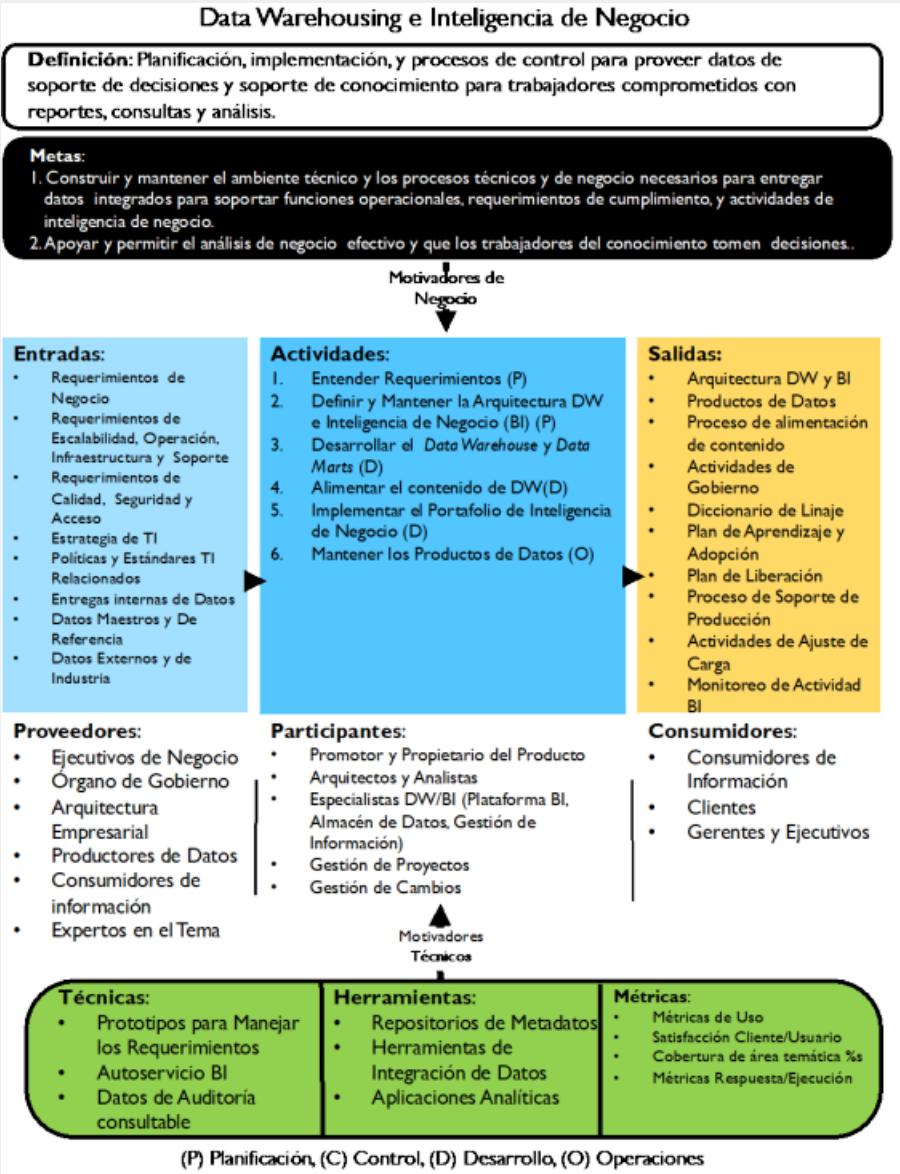


DAMA-DMBOK

DATA MANAGEMENT BODY OF KNOWLEDGE

2ND EDITION

DataWarehousing y BI en el Framework de Dama



Metas:

1. Construir y mantener el ambiente técnico y los procesos técnicos y de negocio necesarios para entregar datos integrados para soportar funciones operacionales, requerimientos de cumplimiento, y actividades de inteligencia de negocio.
2. Apoyar y permitir el análisis de negocio efectivo y que los trabajadores del conocimiento tomen decisiones.

(P) Planificación, (C) Control, (D) Desarrollo, (O) Operaciones

- a. **(P) Las Actividades de Planificación** establecen el curso estratégico y táctico para cumplir los objetivos de Gestión de Datos. Las actividades de planificación ocurren de manera recurrente.
- b. **(D) Las Actividades de Desarrollo** se organizan en torno al ciclo de vida de desarrollo de *software* (SDLC) (análisis, diseño, construcción, prueba, preparación y despliegue).
- c. **(C) Las Actividades de Control** garantizan la continuidad de la calidad de los datos y la integridad, fiabilidad y seguridad de los sistemas a través de los cuales se accesan y se utilizan los datos.
- d. **(O) Las Actividades de Operación** apoyan el uso, mantenimiento y mejora de sistemas y procesos a través de los cuales los datos son accesados y utilizados.

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. **MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS**
4. ENTREGABLES
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

Metas y Motivadores del negocio

Metas:

1. Construir y mantener el ambiente técnico y los procesos técnicos y de negocio necesarios para entregar datos integrados para soportar funciones operacionales, requerimientos de cumplimiento, y actividades de inteligencia de negocio.
2. Apoyar y permitir el análisis de negocio efectivo y que los trabajadores del conocimiento tomen decisiones.

Objetivos del DWH

Dar soporte a la actividad de BI

Disponibilizar el análisis de negocio y la toma de decisiones de forma efectiva

Encontrar formas de innovar basadas en la información de sus datos

Objetivos del BI

Mejora de la toma de decisiones: Proporciona información precisa y oportuna para apoyar las decisiones empresariales.

Optimización de procesos: Identifica oportunidades de mejora y eficiencia dentro de la organización.

Aumento de la competitividad: Ayuda a las empresas a responder rápidamente a los cambios del entorno competitivo (mercado y competencia).

Reducción de costos: Permite identificar y eliminar ineficiencias operativas.

Principios



Enfoque en los objetivos empresariales



Objetivo final en mente



Piense y diseñe globalmente, actúe y construya localmente



Sumarice y optimice al final



Promover transparencia y autoservicio



Construya metadatos



Colabore con otras iniciativas de datos



El mismo tamaño no se ajusta a todo

PREGUNTA

CUÁL DE LOS SIGUIENTES EJEMPLOS NO ES UN PRINCIPIO DE IMPLEMENTACIÓN DE UN DATA WAREHOUSE

- a) Enfoque en los objetivos empresariales
- b) Promover transparencia y autoservicio
- c) Construir Metadatos
-  **Un único tamaño de DWH se ajusta a todas las necesidades**
- e) Piense y diseñe globalmente, actúe y construya localmente

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. **ENTREGABLES**
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. **ACTIVIDADES**
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

Actividades

- 01 Entender los requerimientos**
 ¿Cuáles son los objetivos principales?
- 02 Definir y Mantener la Arquitectura DWH/BI**
 ¿Qué infraestructura y herramientas necesito?
- 03 Desarrollar el DWH y los Data Marts**
 Desplegar la tecnología y construir los modelos y procesos
- 04 Poblar el DWH**
 Cargar los datos
- 05 Implementar el Portafolio de BI**
 Necesidades y herramientas de BI
- 06 Mantenimiento de los Producto de Datos**
 Gestionar los incrementos y asegurar la operación

P

P

D

D

D

O

P

PLANIFICACIÓN

C

CONTROL

D

DESARROLLO

O

OPERACIONES

Actividades

01

Entender los requerimientos

Metas y estrategia de negocio

Toma de decisiones / Regulación / Riesgos
Financiero / Comercial / Clientes

Identificar áreas involucradas

Entrevistar a las personas adecuadas

¿Qué hacen?
¿Por qué?

Capturar preguntas específicas

Yo, como ...[rol]... quiero ...[necesidad]...
para ...[objetivo]...

Identificar Kpi's y Métricas

Identificar fuentes de datos

Catalogar requerimientos

Priorizar

Mejoras simples y valiosas
Mejoras evangelizadoras
Quick Wins

Actividades

02 Definir y Mantener la Arquitectura de BI

Orígenes

- De donde vienen los datos
- Identificar las tecnologías de origen

Destinos

- DataWarehouse
- DataLake
- Data Lake House

Frecuencia

- Batch
- Real Time

Cómo

- Hardware
- Software

Número de ambientes

- **Entorno de Desarrollo:** Para hacer **pruebas iniciales** sin afectar a datos reales. Es donde ingenieros y analistas trabajan en nuevas funcionalidades y modelos de datos.
- **Entono de Test:** Para **validar** que lo desarrollado funciona correctamente antes de llevarlo a producción.
- **Entorno productivo:** Es donde se ejecuta el Data Warehouse real, con datos en vivo que se usan en reportes y análisis del negocio. Aquí todo debe ser **estable y seguro**.

Plan de liberación

- **Demanda actual:** backlog priorizado de necesidades a desarrollar.
- **Planificación de recursos:** número de hilos de desarrollo necesarios en cada momento.
- **Programación de entregas:** Roadmap detallado de puestas en producción.

Actividades

03 Desarrollar el DWH y los DataMarts

Datos

- Ingestar: ¿Cuáles son las mejores fuentes? ¿Hay duplicidades? Mapear las fuentes a los destinos.
- Limpiar: ¿Qué datos se han de cargar en el DWH? ¿todo el nivel de detalle? ¿todo el histórico? ¿hay datos sin calidad suficiente?
- Remediar: corregir datos erróneos antes de las cargas iniciales.
- Integrar: ¿Cómo unificamos datos de diversas fuentes?
- Transformar: crear reglas que documentan los procesos de negocio. Requiere participación directa de los SME's.
- Creación de taxonomías: identificación de principales dominios de datos.

Tecnología

- Desplegar herramientas de BackEnd
- Desplegar sistemas de almacenamiento: diferentes niveles (capa raw), optimizaciones, etc.
- Desplegar herramientas de movimientos de datos: ETL, ELT, CDC, virtualización, etc.

Herramientas de BI

- Identificar el tipo de herramientas necesarias
- Realizar estudios y Pruebas de concepto
- Desplegar herramientas

Actividades

04 Poblar el DWH

La mayor parte del trabajo en cualquier esfuerzo de DW/BI es la preparación y procesamiento de los datos.

Para diseñar el procesamiento de los datos hay que tener en cuenta lo siguiente:

Latencia

¿Cuál es la diferencia de tiempo aceptable entre el dato en origen y en el DataMart?

Disponibilidad de fuentes

¿A qué hora puedo empezar a extraer los datos?

Ventanas de carga

¿De que tiempo dispongo para hacer transformaciones?

Tiempos de procesado

¿Cuántas transformaciones he de hacer y qué tardo?

Estrategia de cargas

¿Es mejor cargarlo todo cada día o solo el incremento?

Captura de cambios

¿Cómo identificamos los cambios?

Herramientas

¿de que herramientas dispongo?

Actividades

04

Poblar el DWH

Tipos de procesamiento de carga

Tipo de carga	Método	Características
Histórica	Carga completa	Suele hacerse durante el primer despliegue del DWH
	Cambios en marcas de tiempo	Carga parcial sencilla basada en marcas de tiempo del origen
Carga por lotes	Cambios en tablas de registro	Los datos a cargar se identifican en el origen en unas tablas de registro
	Transacciones de base de datos	Se basa en replicar los cambios del registro transaccional de la base de datos
	Cambios en mensajes	Los cambios se publican como mensajes
	Carga completa	Carga simple pero lenta al tener que identificar los cambios
	Goteo	Los cambios se guardan en lotes en origen y se procesan al superar un umbral de mensajes o tamaño
Tiempo casi real	Mensajería	Útil para paquetes pequeños de datos, los datos se publican en un bus y los procesos suscritos van procesando según lo requieren
	Streaming	Los datos se acumulan en destino y son procesados en orden

Actividades

05

Implementar el Portafolio de BI

¿Necesidades de los usuarios?



Reporting estático

Informes fijos sin interacción del usuario (por ejemplo, en pdf)



Cuadro de mandos

Panel visual con indicadores clave (KPIs) en tiempo real para tomar decisiones (ejemplo: dashboard de ingresos y costos).



Aplicación analítica

Herramienta avanzada que permite explorar datos y descubrir tendencias (ejemplo: analizar el comportamiento de clientes).



Autoservicio

Sistema donde los usuarios pueden crear sus propios reportes sin depender del equipo técnico

¿Herramientas?



Sistemas de reporting y análisis tradicionales

Soluciones de nicho



Aplicaciones transaccionales

Incorporan cierta capacidad de análisis a través de cubos agregados



Virtualización / Capa semántica

Permite democratizar los datos



Plataformas de BI

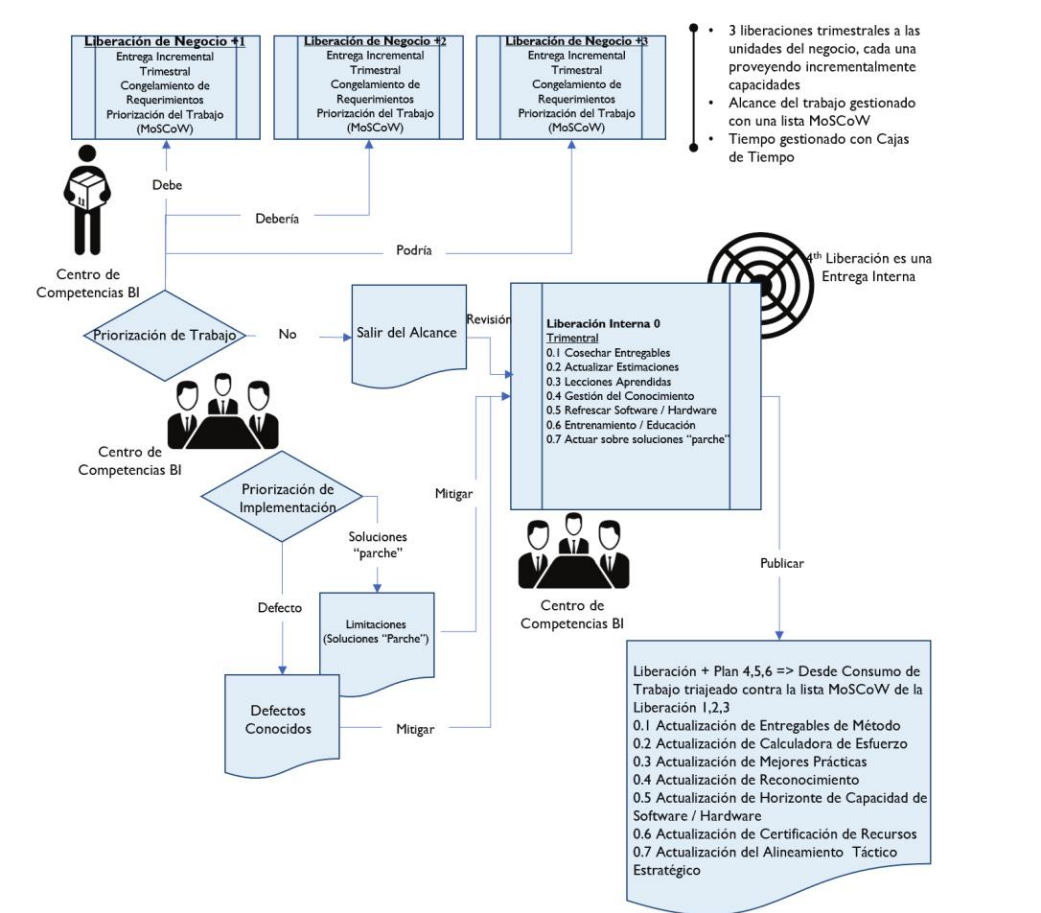
Incluyen acceso, limpieza, transformación, reporting, BI, linaje, etc.

Actividades

06

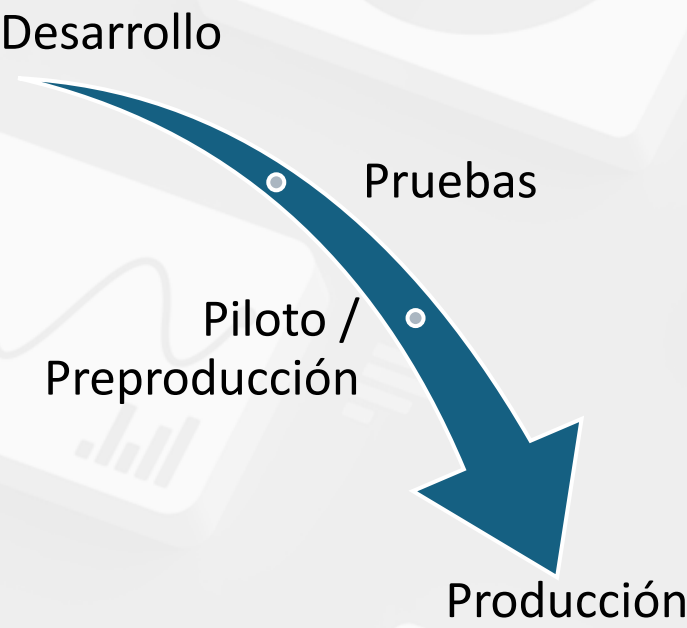
Mantenimiento de los Productos de Datos

Gestión de liberaciones



Ciclo de vida de desarrollo

Ejecutar las iteraciones que incrementan el valor



Actividades

06

Mantenimiento de los Productos de Datos

Balancear entre Agilidad y Robustez



Actividades

06

Mantenimiento de los Productos de Datos

Data Ops (Monitorización y Optimización)

Es importante tener en cuenta los siguientes conceptos:

SLA de servicio

¿Lo estoy cumpliendo?

Incidentes

¿Como gestionar los errores de las cargas?
¿Puedo minimizar el impacto?

Optimización

¿Puedo mejorar mis procesos?
¿Puedo mejorar mi infraestructura?

Tiempos de cargas

¿Están aumentando?
¿Puedo anticipar problemas?

Cuellos de botella

¿Cuál es el camino crítico?
¿Puedo optimizarlo?

Procesos obsoletos

¿Hay procesos innecesarios?
¿Qué hacemos con ellos?

Rendimiento

¿Hay problemas de performance?
¿Estoy llegando al límite de recursos?

Utilización

¿Que productos da datos no se usan?
¿Por qué? ¿Qué hacemos con ellos?

Escalados

¿Puedo anticipar y planificar los escalados?

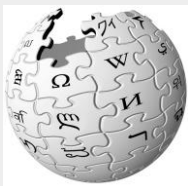
PREGUNTA

-  Al priorizar es mejor empezar por aquellas actividades mas lentas y costosas y dejar las más fáciles para el final (V o F) **Es mejor empezar por mejoras simples y valiosas**
-  Desarrollo, Preproducción y Producción son entornos necesarios para liberar mejoras de forma ordenada (V o F)
-  Es necesario monitorizar la operación para detectar y eliminar cuellos de botella (V o F)
-  Poblar el Data WareHouse es relativamente fácil y rápido siempre que hayamos hecho un buen diseño (V o F) **La mayor parte del trabajo en cualquier esfuerzo de DW/BI es la preparación y procesamiento de los datos.**

Índice

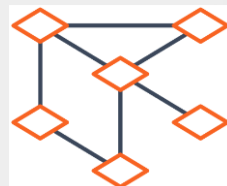
1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. ACTIVIDADES
6. **HERRAMIENTAS**
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

Herramientas



GLOSARIO DE NEGOCIO

- Términos de negocio
- Definiciones
- Taxonomías
- Ejemplos y Reglas
- Responsables



LINAJE DE DATOS

- Investigación de errores
- Análisis de impacto
- Origen de los datos
- Datos no usados



INTEGRACIÓN DE DATOS

- Scripting
- ETL /ELT
- CDC
- Virtualización
- Schedule - Monitorización



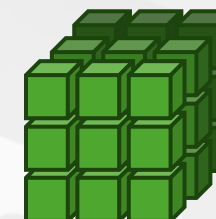
BUSINESS INTELLIGENCE

- Reporting estático
- Análisis dinámico
- Autoservicio
- Detección de patrones



REPORTING OPERACIONAL

- Sistemas Operacionales
- Usuario final
- Crear Reports AdHoc



ANALISIS OLAP

- Online Analytical Processing
- Cubos pre-calculados
- Slice/Dice/Pivot/Roll-Up/ Drill-Down/ Drill-Up
- MOLAP, ROLAP, HOLAP

Herramientas

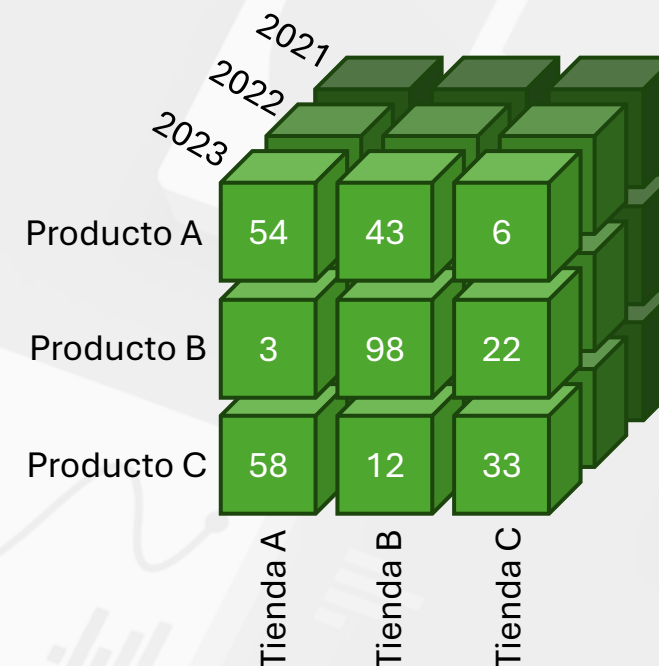
Análisis multidimensional , OLAP

- OnLine Analytical Processing (OLAP)
- Se basa en consultas contra dimensiones y medidas contenidas en una base de datos multidimensional, un cubo.
- Cada cruce de dimensiones está calculado y almacenado para mejorar el rendimiento
- Las dimensiones tienen jerarquías
- Suelen ser soluciones propietarias de cada fabricante.

¿Cómo han evolucionado las ventas del producto C en España por trimestre durante los últimos 3 años?"

Ventajas

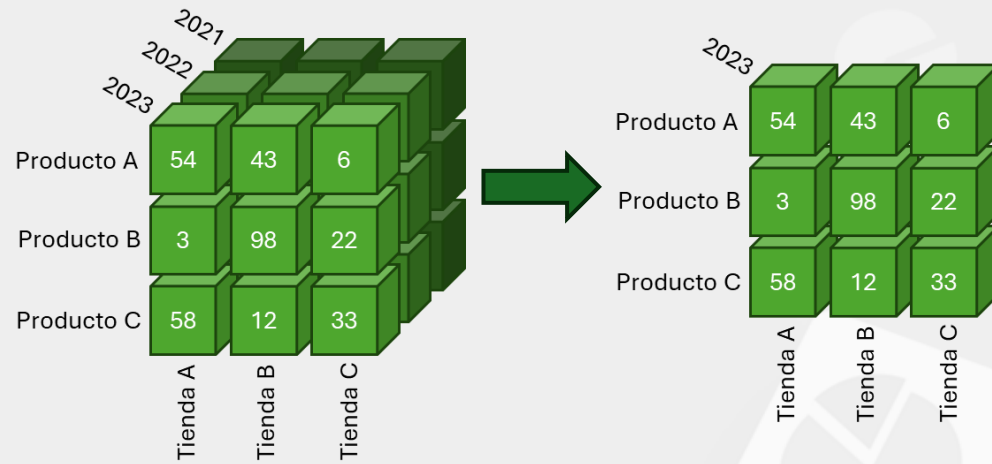
- Consultas muy rápidas
- Análisis de grandes volúmenes históricos
- Agregación dinámica
- Alta capacidad de exploración (drill-down, roll-up...)



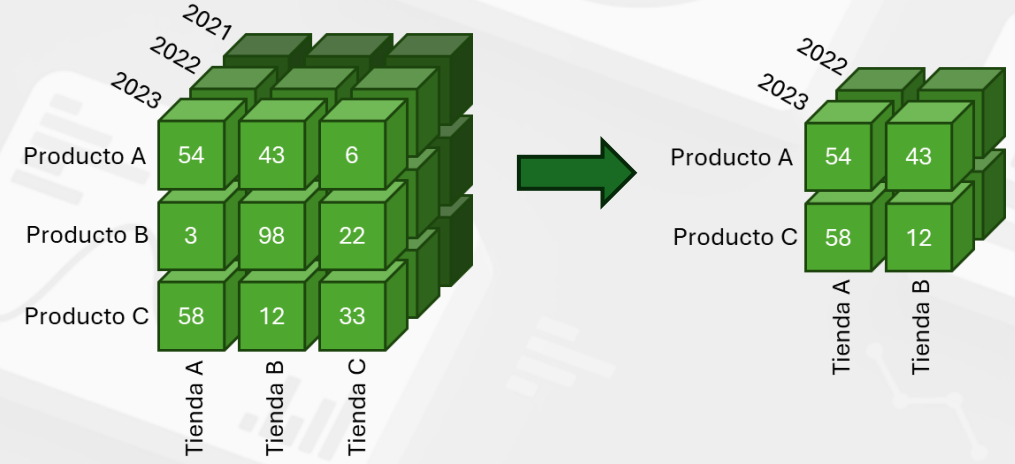
Herramientas

OLAP Operaciones, Selección

Slice: Selección para un valor de una de las dimensiones, filtrar una columna



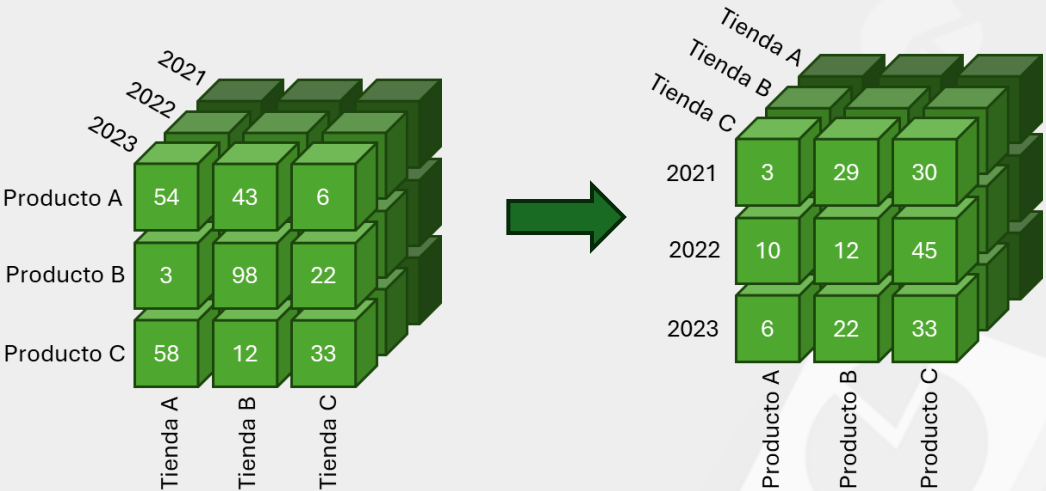
Dice: Selección de valores de más de una de las dimensiones, aplicar varios filtros a la vez



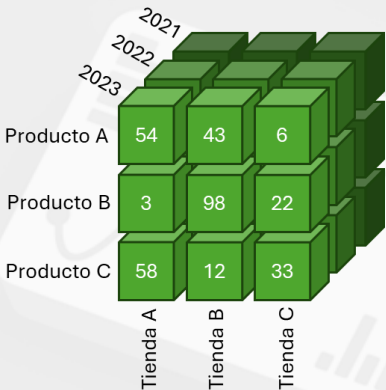
Herramientas

OLAP Operaciones, Pivot (columnas a filas) y Roll-Up (agregar)

Pivot: Cambia la orientación dimensional



Roll-up: Calcula los subtotales y totales para la selección de dimensiones. Requiere de una definición de fórmula de agregación



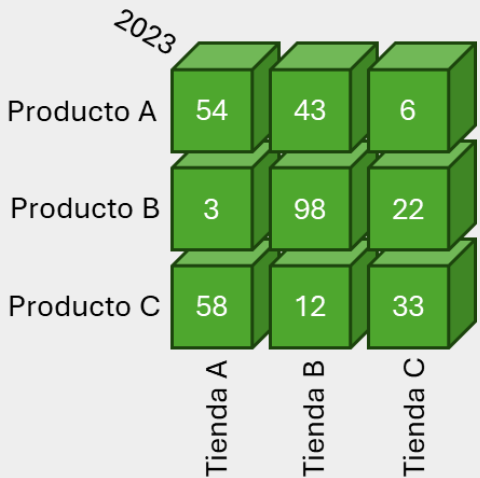
Producto A	54	43	6	103
Producto B	3	98	22	123
Producto C	58	12	33	103
	115	153	61	329
	Tienda A	Tienda B	Tienda C	

Herramientas

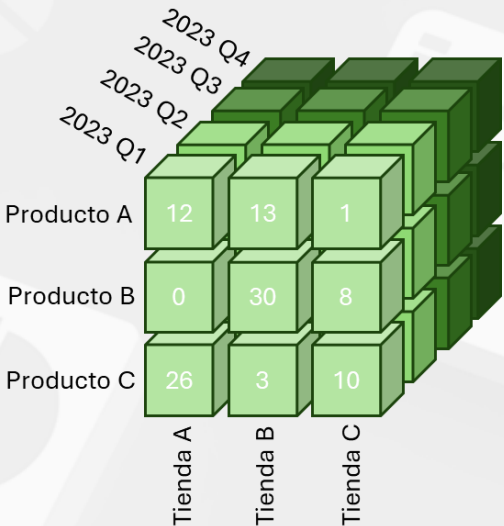
OLAP Operaciones, Drill-down (bajar de nivel) y Drill-Up (subir de nivel)

Permite aumentar el desglose (drill-down) o disminuirlo (drill-up) en las dimensiones con jerarquías

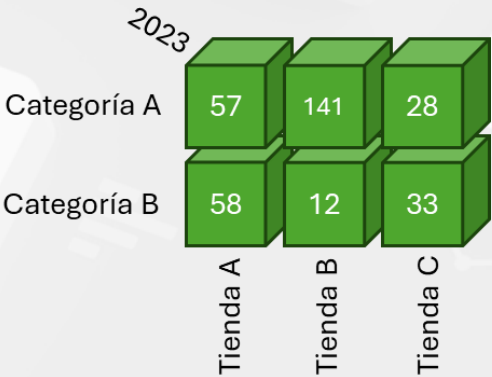
Escenario base, slice de 2023



Drill-down de 2023



Drill-up de producto a categoría

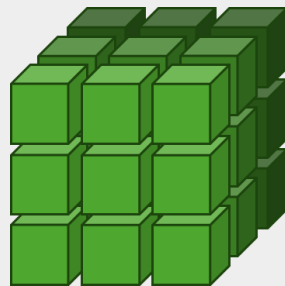


Herramientas

OLAP Tipos de cubo

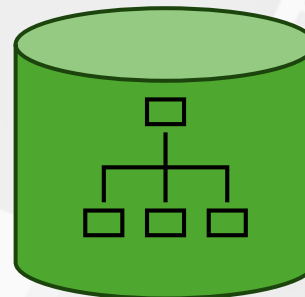
MOLAP, Multidimensional OLAP

Basado en un formato propietario, suele ser un archivo



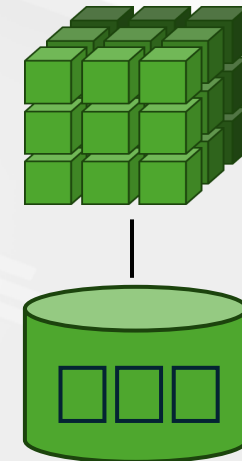
ROLAP, Relational OLAP

Montado sobre una estructura de tablas en BBDD relacional. Se suele utilizar cuando los datos desglosados son muchos



HOLAP, Hybrid OLAP

Híbrido entre los dos anteriores, suele usar MOLAP para los agregados y ROLAP para obtener los detalles



PREGUNTA

-  CDC es el acrónimo de Centro de Control, y se utiliza para monitorizar las cargas nocturnas en el DWH (V o F) **CDC significa Change Data Capture y es una tecnología de replicado de datos**
-  El linaje de los datos permite identificar los problemas derivados a la hora de hacer cambios en las fuentes de datos (V o F)
-  Un buen glosario de datos debe contener definiciones claras y entendibles (V o F)
-  Slice, Dice y Pivot son operaciones que podemos ejecutar en una herramienta de integración de datos (V o F) **Son operaciones de tecnología OLAP**

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. ACTIVIDADES
6. HERRAMIENTAS
7. **TÉCNICAS**
8. GUÍAS DE IMPLEMENTACIÓN
9. MÉTRICAS

Técnicas

Prototipos



- El DWH es el primer lugar donde se detectan problemas de calidad relacionados con la integración de datos.
- Crear juegos de datos de prueba antes de las tareas de implementación permite reducir riesgos por datos inesperados.
- Permite evaluar el esfuerzo y el alcance.

BI Autoservicio



- Canal de entrega fundamental.
- Democratiza el acceso a los datos, y permite el uso de técnicas avanzadas antes sólo disponibles en el entorno de TI.
- Mensajes, alertas, visualización, interacción, desarrollo de aplicaciones analíticas y generación de cuadros de mando.

Auditoría



- Mantener linaje e información de auditoría en todas las estructuras y procesos.
- El linaje ayuda a la resolución de problemas y cuando se comparte mejora la confianza de los usuarios

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. **GUÍAS DE IMPLEMENTACIÓN**
9. MÉTRICAS

Guías de Implementación

Para el éxito de la implantación se necesita una **arquitectura adaptable** y un **equipo capaz** de afrontar tanto la carga diaria como la evolución.

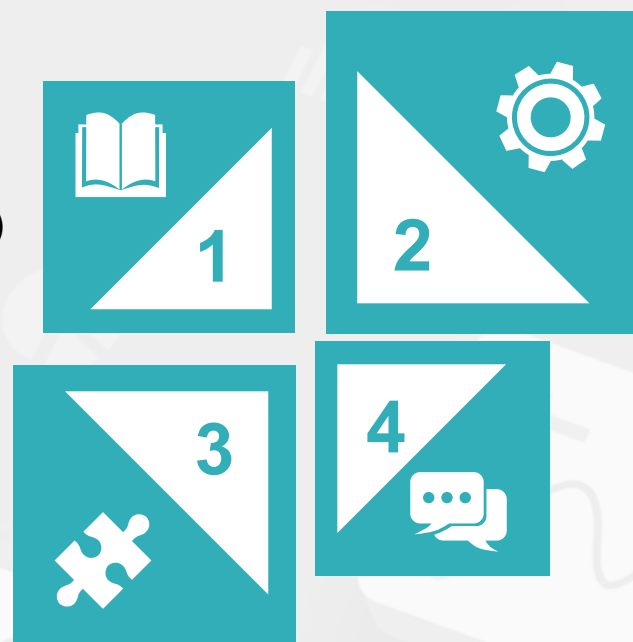
Evaluación de la preparación/riesgo

Antes de comenzar la implementación:

- Listado de verificación de requisitos
- Restricciones de seguridad y sensibilidad de los datos
- Selección de herramientas (limitaciones de seguridad)
- Asegurar recursos
- Evaluar los datos de origen

Gestión de la configuración

- **Control de versiones:** permite gestionar el histórico de cambios de código, volver a versiones anteriores y trabajar en equipo sin perder información.
- **DevOps:** Automatización, colaboración y mejora continua para hacer que el software se construya, pruebe y despliegue más rápido y con menos errores.



Hoja de ruta de liberación

- Desarrollo incremental **planificado**
- Priorización alineada con negocio
- Hay que mantener un Backlog actualizado
- Gestionar dependencias es parte de la priorización

Organización y cambio cultural

Factores de éxito

- Esponsorización del negocio
- Identificar objetivos de negocio
- Implicación del negocio (SME's)
- Formación de negocio (alfabetización)
- Alineamiento con IT
- Centro de excelencia (CoE) de datos

Índice

1. INTRODUCCIÓN
2. FRAMEWORK DE DAMA
3. MOTIVADORES DE NEGOCIO, METAS Y PRINCIPIOS
4. ENTREGABLES Y PARTICIPANTES
5. ACTIVIDADES
6. HERRAMIENTAS
7. TÉCNICAS
8. GUÍAS DE IMPLEMENTACIÓN
9. **MÉTRICAS**

Métricas

Métricas de uso	Indican el nivel de uso del sistema en la organización	N° usuarios registrados N° usuarios conectados N° conexiones concurrentes
Porcentaje de cobertura de áreas temáticas	Describe el uso del DWH por parte de los departamentos (a nivel de topología de datos). Permite identificar solapamientos y candidatos a ser compartidos	Estadísticas de uso por parte de usuario o informe y departamento
Métricas de respuesta y rendimiento	Miden cómo está funcionando el sistema, así como cuál es la respuesta desde el ámbito de la operación. Pueden afectar a los SLA definidos	Tiempo de carga (por proceso) % de cargas correctas Tiempo de respuesta ante fallos Listado de consultas / informes más utilizados

RESUMEN

A Recordar

- BI es un tipo de análisis de datos destinado a comprender las actividades y oportunidades de la organización.
- DWH es un almacén de datos centralizado que da soporte a la toma de decisiones.
- Las dos grandes corrientes de diseño de DWH fueron diseñadas por Bill Inmon y Ralph Kimball respectivamente en los 90.
- Inmon se basa en la creación un DWH corporativo con el modelo lógico completo, los datos se exponen en DataMarts.
- Kimball Se basa en el diseño modular de DataMarts (estrellas) que se van uniendo para formar el DWH.
- Debemos entender bien los requerimientos del negocio antes de empezar a diseñar el DWH.
- La mayor parte del trabajo en cualquier esfuerzo de DW/BI es la preparación y procesamiento de los datos.
- Debemos ser capaces de balancear entre agilidad y robustez.
- El control de la operación (data Ops) es crucial para mantener los SLA de servicio.

¡Gracias!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas

