

DISEÑO DE DATA WAREHOUSE

BASE DE DATOS II

OBJETIVOS

- IDENTIFICAR LOS PRINCIPALES CONCEPTOS Y LAS PRINCIPALES VENTAJAS DE UN DATA WAREHOUSE (ALMACÉN DE DATOS)
- CONOCER LAS DIFERENCIAS ENTRE LOS SISTEMAS DE PROCESAMIENTO DE TRANSACCIONES EN LÍNEA (OLTP) Y LOS ALMACENES DE DATOS
- IDENTIFICAR EL CICLO DE VIDA DE UN DATA WAREHOUSE

INTRODUCCIÓN

- Las empresas crecen mas al entender mejor su entorno de negocios.
- Los sistemas deben de proveer información actualizada y todo momento a las altas gerencias de las empresas.
- Alinear los objetivos de datos a los objetivos de la empresa.
- Estrategias de negocios: ventaja competitiva, bajo costo, diferenciación y economía de escala
- Vida del producto

CONCEPTOS

- **Inteligencia de negocio (Business Intelligence):** es una estrategia empresarial que persigue incrementar el rendimiento de la empresa o la competitividad del negocio, a través de la organización inteligente de sus datos históricos (transacciones u operaciones diarias).
- **Minería de datos (Data Mining):** es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto o área de la empresa.

DATA WAREHOUSE

- Las empresas de todo el mundo constantemente adquieren datos de negocios y presentación de informes
- Un Data Warehouse ayuda a cualquier negocio a adquirir datos, mientras que al mismo tiempo aseguran de que los procesos normales del día a día se cumplan

DW & BI



DATA WAREHOUSE

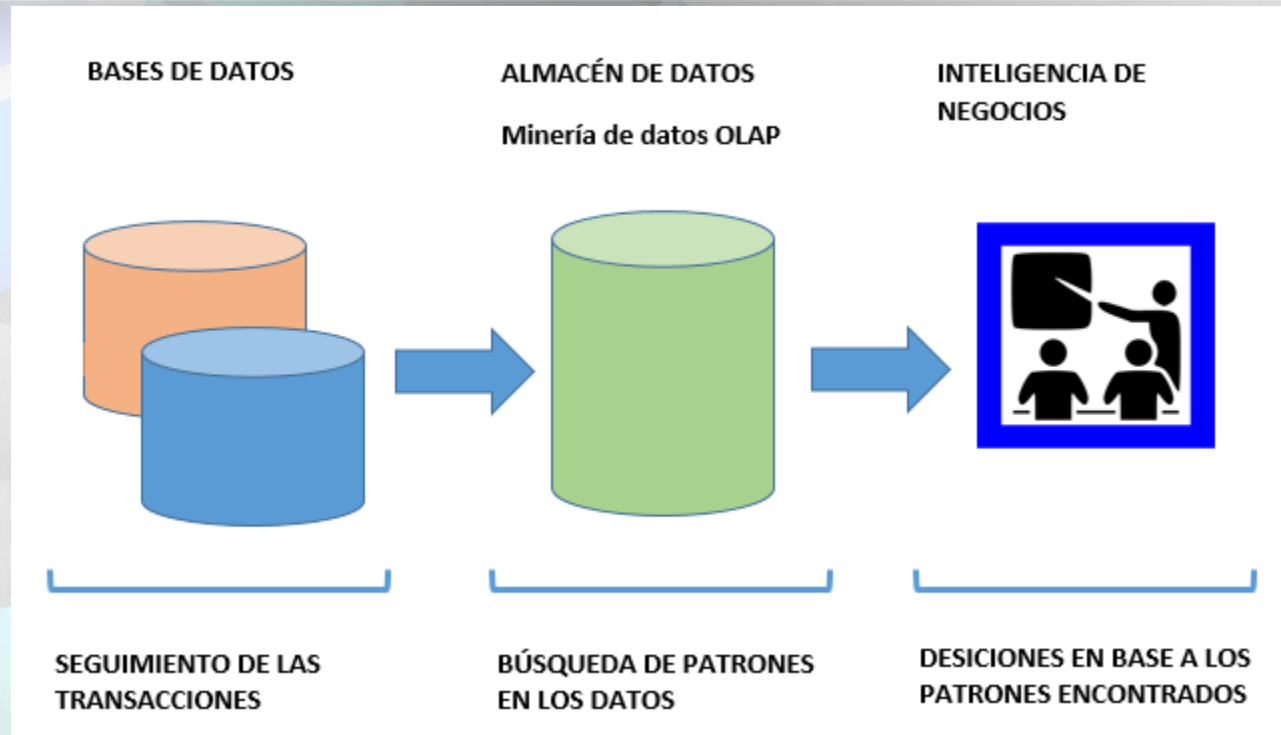
- Un Data Warehouse es una colección de datos clasificada por temas, integrada, variable en el tiempo y no volátil que se utiliza como ayuda al proceso de toma de decisiones por parte de quienes dirigen una organización (definición según W.H Inmon)



DATA WAREHOUSE

Definición	Datos
Clasificados por temas	Son datos organizados con orientación hacia entidades (producto, clientes) en lugar de estar orientados hacia el proceso (facturación, pedido)
Integrados	Se crean y diseñan fuera de la base de datos operacionales, también los datos pueden proceder de diferentes sistemas de aplicaciones utilizados dentro de la organización.
Variables en el tiempo	Una vez que los datos son almacenados, estos no cambian y el almacén de datos puede tener un tiempo de vida de cinco a diez años
No volátiles	Normalmente las únicas operaciones sobre la base de datos de datos se reducen a captura de datos y acceso a los mismos

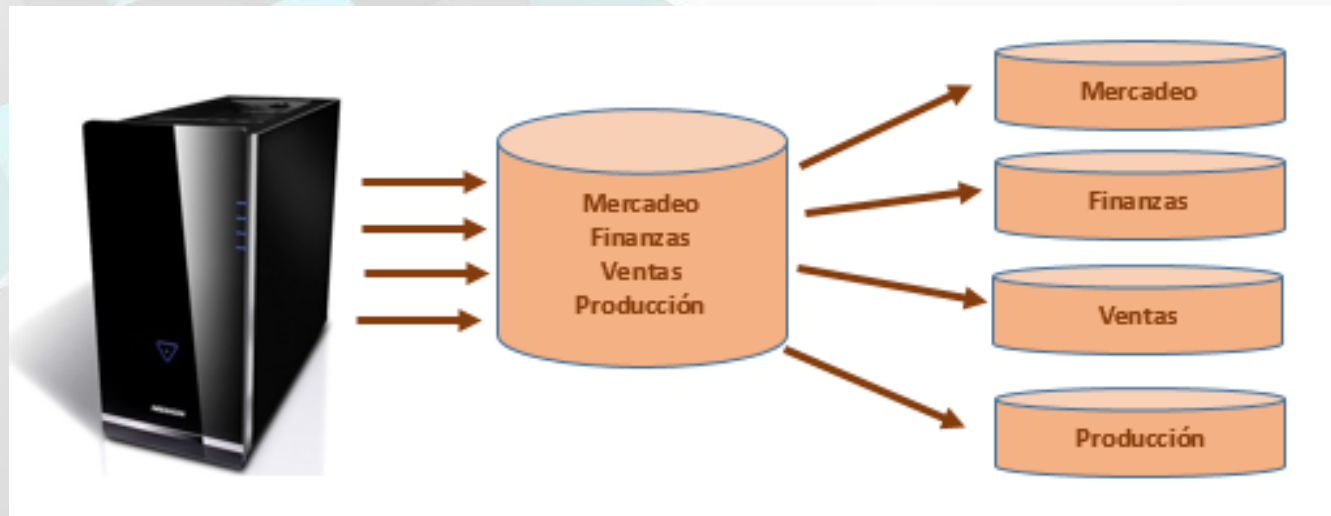
EJEMPLO DE LA FUNCION DE UN DATA WAREHOUSE



- Las aplicaciones típicas de data warehouse se utilizan a diario en el sector bancario y financiero, así como algunas otras en las áreas de administración de riesgos de créditos. También son muy utilizadas en empresas de artículos de consumo, para comprender los patrones de compra y hábitos del consumidor.

DATA MART

- Es un pequeño datawarehouse, creado para una determinada cantidad de usuarios de un área de la empresa.
- Tiene como función apoyar a los sistemas de toma de decisiones.



DATA MART

- Algunos de los cuestionamientos que se deben solucionar cuando la empresa cuenta con tecnología de apoyo, como son data warehouse y el data mart, son los siguientes:
 - ¿Quiénes son mis mejores clientes?, ¿Dónde se ubica?, ¿Qué compran?, ¿Qué los caracteriza?, ¿Qué líneas de productos son los mas rentables?, ¿Qué productos son mis 10 mejores?

SISTEMAS OLTP Y DATAWAREHOUSE

- Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones.
- Una transacción genera un proceso atómico (que debe ser validado con un commit, o invalidado con un rollback), y que puede involucrar operaciones de inserción, modificación y borrado de datos.
- El proceso transaccional es típico de las bases de datos operacionales.

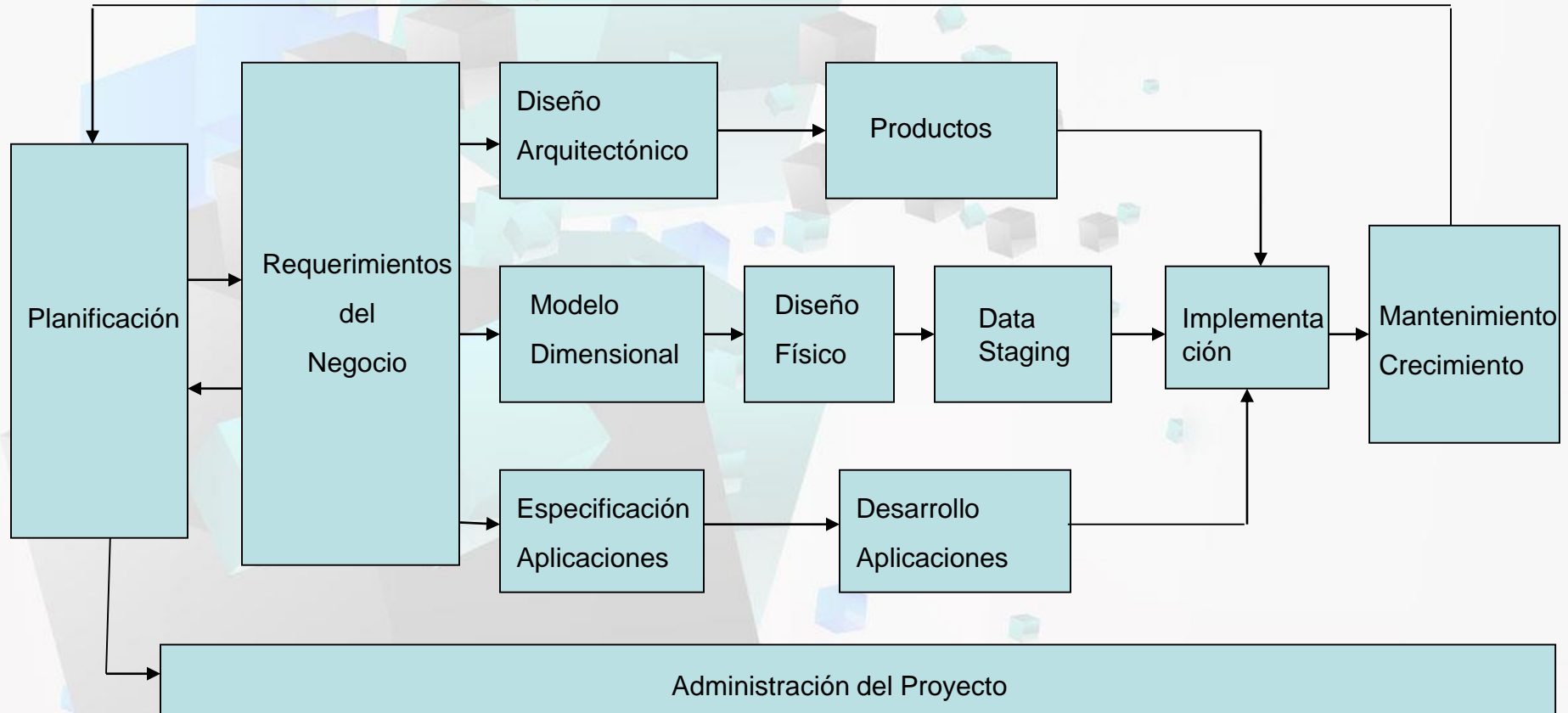
COMPARACIÓN DE LOS SISTEMAS OLTP Y DATAWAREHOUSE

SISTEMAS OLTP	DATAWAREHOUSE
Almacén de datos actuales	Almacén de datos históricos
Almacena datos detallados	Almacena datos resumidos en poca o gran medida
Los datos son dinámicos	Los datos son principalmente estáticos
Procesamiento repetitivo	Procesamiento ad-hoc, no estructurado y heurístico
Alta tasa de transacciones	Tasa media o baja de transacciones
Dirigido por transacciones	Dirigido por análisis
Orientado a la aplicación	Orientado a los temas
Soporta las decisiones cotidianas	Soporta las decisiones estratégicas
Sirve a un gran número de usuarios administrativos / operacionales	Sirve a un número relativamente bajo de usuarios de tipo gerencial

IMPLEMENTACIÓN DE UN DATA WAREHOUSE

- Un proyecto de Data Warehouse debe basarse en satisfacer necesidades del negocio
- Los datos se deben presentar de acuerdo con las variables del negocio
- Un Data Warehouse está en permanente evolución
- Cada proyecto de Data Warehouse debe tener un principio y un fin

CICLO DE VIDA DE UN DATAWAREHOUSE



PLANIFICACIÓN

- Predisposición de la organización (apoyo de la gerencia, motivación de la gerencia, factibilidad)
- Alcance (valor del negocio, prioridades, plazos, riesgos)
- Justificación de negocio (estimación de costo y beneficio, costos de HW y SW, ampliación de requerimientos)
- Aspectos humanos (funciones técnicas, funciones de administración, transferencia de conocimiento)
- Plan del proyecto (identificación de tareas necesarias, participación de personas claves)
- Puntos clave (Desarrollar cuidadosamente el plan del proyecto, ser un director de proyecto con capacidad de motivar, administrar y comunicar a todos los niveles)

REQUERIMIENTOS DEL NEGOCIO

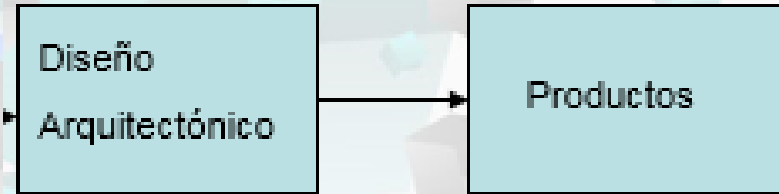
- Relevamiento de requerimientos (entrevista a los altos directivos, gerentes intermedios y profesionales del negocio)
- Entrevistas sobre los datos (conocedores de las fuentes de datos, conocedores del negocio)
- Actividades post-entrevista (resumen ejecutivo, detalle de cada proceso de negocios)

DESARROLLO DEL PROYECTO

El desarrollo del proyecto se realiza en tres líneas

- Línea de tecnología
- Línea de aplicaciones
- Línea de datos

Línea de tecnología



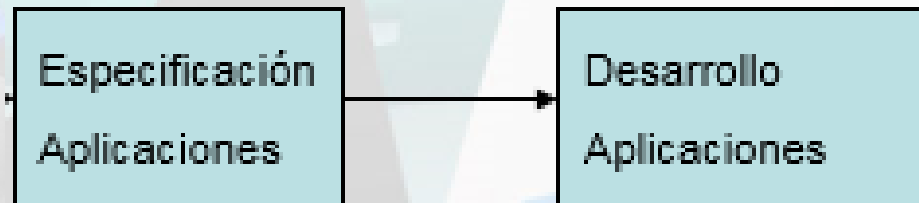
- Arquitectura
 - Componentes
 - Fuentes de datos
 - Organización de datos (Data Staging)
 - Presentación de datos
 - Acceso a los datos
 - Influencia de los requerimientos (alta disponibilidad, ubicación de los usuarios)
- Selección de productos (política de compras, requerimientos técnicos y funcionales)

Línea de datos



- Modelo dimensional (Proceso de negocios, Granularidad (nivel de detalle), Tablas de hechos ,Tablas de dimensión)
- Diseño físico de la base de datos (Nombres de columnas,Tipos de datos, Claves, Particiones)
- Data Staging (es un área intermedia de almacenamiento de datos utilizada para el procesamiento de los mismos durante procesos de extracción, transformación y carga (ETL))

Línea de aplicaciones



- Especificación y desarrollo de aplicaciones (Diferentes niveles de necesidad de información para diferentes usuarios, Documentación de relevamiento de requerimientos)

IMPLEMENTACIÓN

- Convergencia de las tres líneas
- La línea de datos es la más impredecible
- Fechas comprometidas
- Asegurarse de que el Data Warehouse esté en condiciones
- Educación (entender cual es la audiencia)