



Universidad
Ricardo Palma

FACULTAD Y ESCUELA
INGENIERIA INFORMATICA



Curso: IF1018 Tópicos Especiales de Ingeniería de Software

Unidad: 01

Tema(s): Introducción a Data Warehouse

Semestre Académico: 2013-I

Profesor(a): Carlos Armas Romero



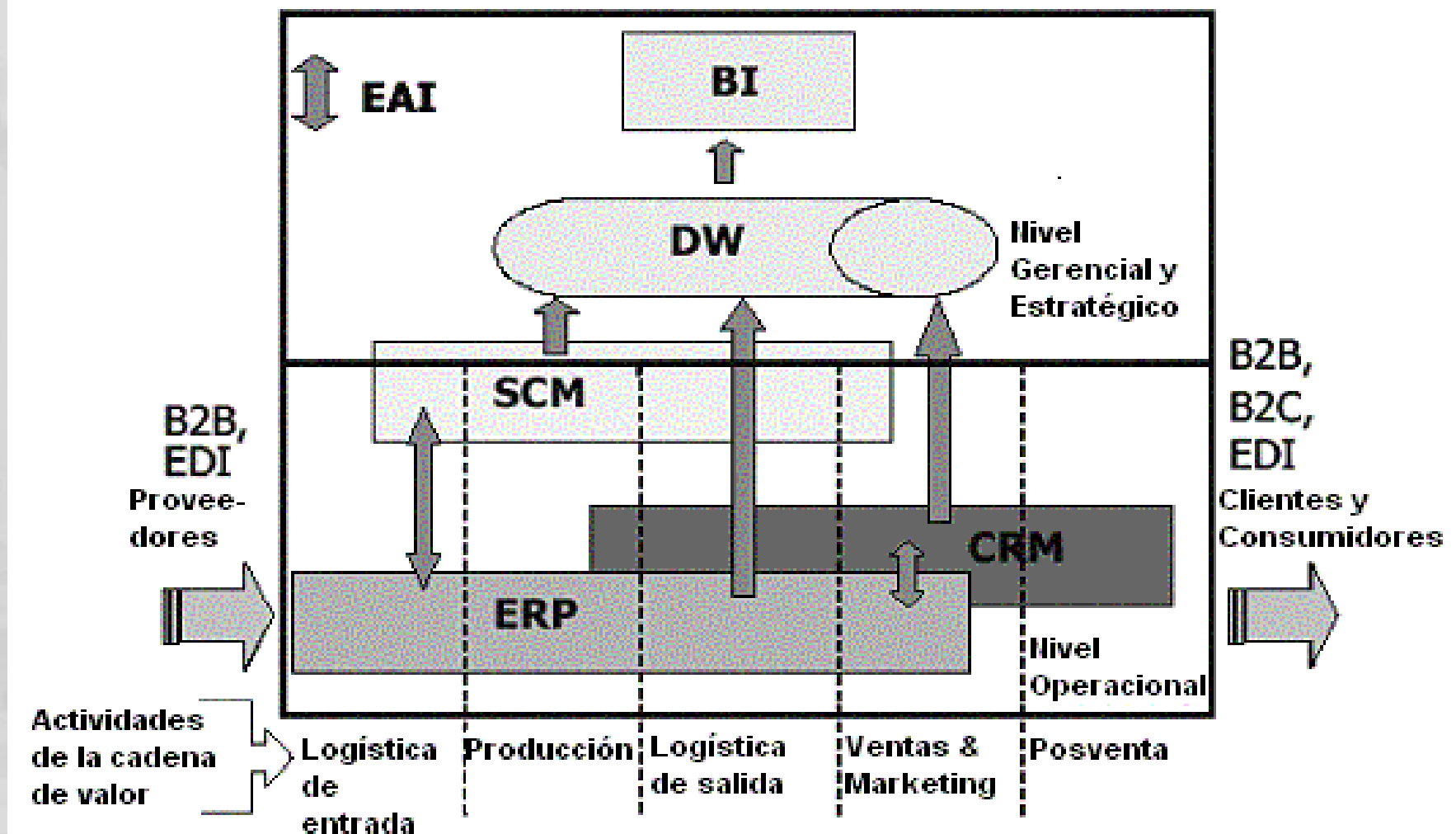
Los materiales expuestos aquí son responsabilidad de sus autores y tienen fines exclusivamente didácticos, en actividades educativas con miembros de la comunidad estudiantil, sin fines de lucro directo o indirecto. Ley N° 822 Artículo ° 41 inciso C.



Plantilla desarrollada por: Centro de Educación Virtual 2011

www.urp.edu.pe

Data Warehousing

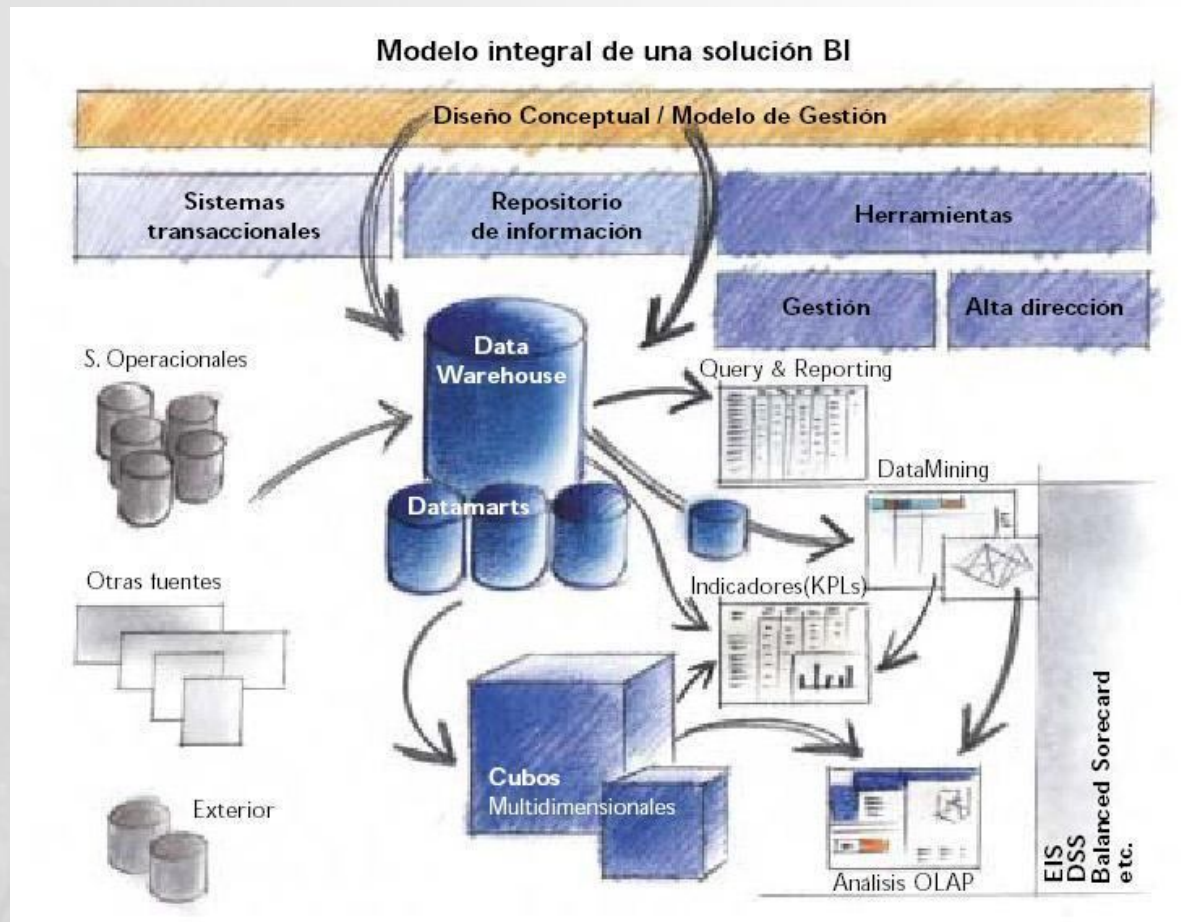


B.I. : Business Intelligence

Software que le posibilita a los usuarios la obtención de informaciones corporativas mas fácilmente.

El software BI debe permitir que sean derivadas las transacciones y sumarios que el usuario necesite, sin que se precise conocer cuales son las fuentes de esos datos

B.I. : Business Intelligence



B.I. : Business Intelligence



B.I.: Recursos y herramientas

- Fuentes de datos : warehouses, data marts, etc.
- Herramientas de administración de datos.
- Herramientas de extracción y consulta.
- Herramientas de modelización (Data Mining)

Evolución: Business Data to Business Information

Etapa	Pregunta de Negocio	Tecnología disponible	Proveedores	Características
Data Collection (1960)	¿Cuál fue el total de ventas en Capital Federal y GBA?	Computadoras, cintas, discos	IBM, NCR, etc	Retrospectivo Estático
Data Access (1980)	¿Cuáles fueron las ventas por sucursal en Capital Federal y GBA?	RDBMS SQL	Oracle, Informix, Sybase, etc	Retrospectivo Dinámico

Evolución: Business Data to Business Information

Etapa	Pregunta de Negocio	Tecnología disponible	Proveedores	Características
Data Navigation (1990)	¿Cuál fue el total de ventas en Capital Federal? Drill down a GBA	OLAP DW	Pilot, Discoverer, Arbor, etc	Retrospectivo Dinámico Niveles múltiples
Data Mining (2000)	¿Cómo evolucionarán las ventas en el próximo año?	Algoritmos avanzados Multiprocesadores	Intelligent Miner (IBM) SGI SAS, etc	Prospectivo. Proactivo

Definición de Datawarehouse

“El Data Warehouse es una colección de datos orientados al tema, integrados, no volátiles e historizados, organizados para el apoyo de un proceso de ayuda a la decisión”.

“El objetivo de los DWs (almacenamiento de datos) es almacenar los datos oriundos de los diversos sistemas de información transaccionales en varios grados de relaciones y sumarios, con el fin de facilitar y agilizar los procesos de toma de decisión por diferentes niveles gerenciales.”

Características de un DW

Orientados al usuario :

Los datos almacenados en el Data Warehouse proveen información sobre **un tema** en particular en vez de atender la operatoria de gestión de la compañía.

Características de un DW

Integridad :

Esos datos deben ser tratados (identificados, catalogados, recolectados, a disponibilidad, transformados en informaciones) e integrados permitiendo diversas formas de consultas, a través de mecanismos amistosos de las herramientas de los usuarios.

Características de un DW

No volátil :

El almacén de información de un Data Warehouse existe para ser leído, y no modificado.

La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Características de un DW

Histórico :

Los DWs son normalmente actualizados en Batch (actualizados una o algunas veces por día y no conteniendo los datos de las transacciones de la empresa de manera on-line) y pueden contener grandes cantidades de datos.

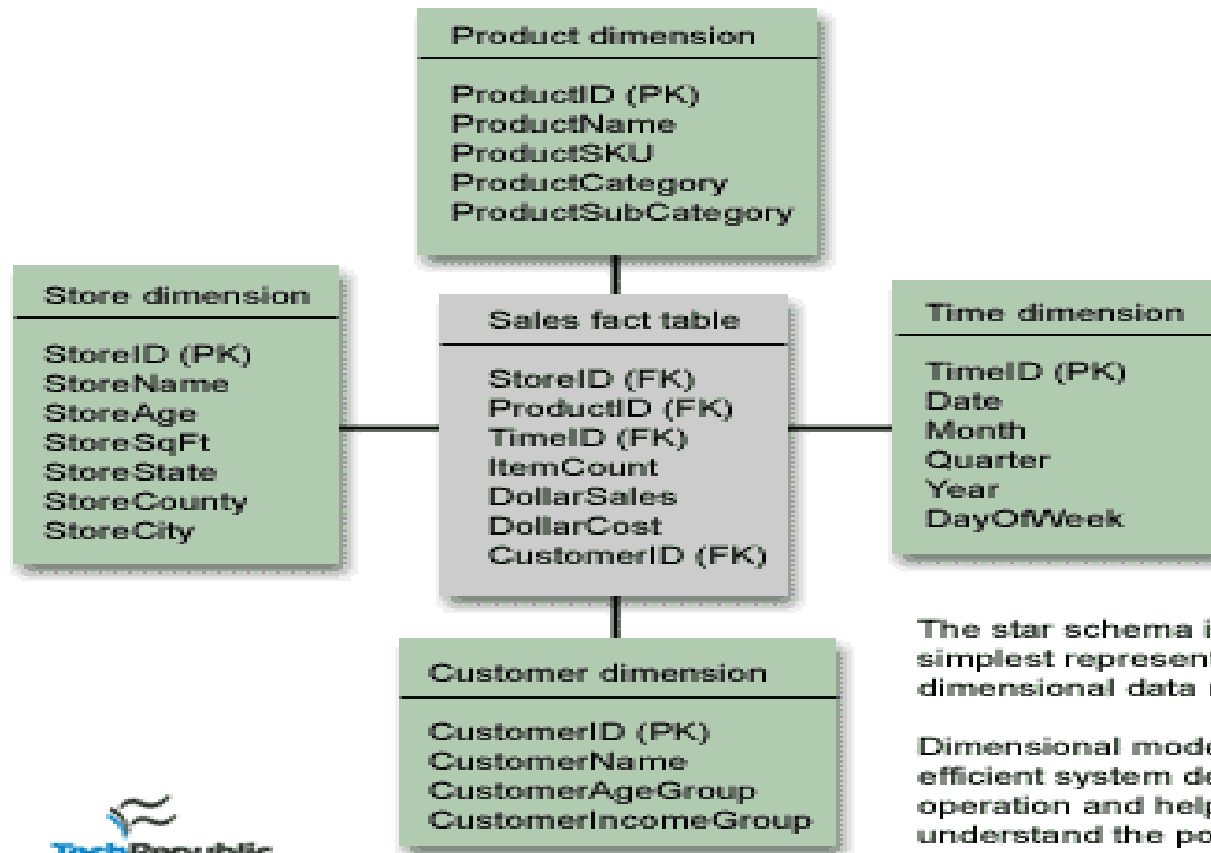
Análisis multidimensional

Un sistema de bases de datos relacionales, no significa disponer de un soporte directo para la toma de decisiones.

Muchas de estas decisiones se basan en un análisis de naturaleza multidimensional, que se intentan resolver con la tecnología no orientada para esta naturaleza.

Este análisis multidimensional, parte de una visión de la información como dimensiones de negocio.

Estructura Multidimensional



The star schema is the simplest representation of the dimensional data model.

Dimensional models support efficient system design and operation and help end users understand the possibilities of the system.



Metadatos

Los metadatos permiten mantener información de la procedencia de la información, la periodicidad de refresco, su fiabilidad, forma de cálculo, etc., relativa a los datos de nuestro almacén de datos.

Otra característica del Data Warehouse es que contiene datos relativos a los datos, concepto que se ha venido asociando al término de metadatos.

Estos metadatos serán los que permitan simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Metadatos

Soportan al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de navegación

Datos Operacionales y Datos Warehouse

	Datos operacionales	Data Warehouse
Contenido	Valores elementales	Datos sumarios, derivados
Organización	Por aplicación	Por tema
Estabilidad	Dinámicos	Estáticos hasta su actualización

Datos Operacionales y Datos Warehouse

	Datos operacionales	Data Warehouse
Estructura	Optimizada para uso transaccional (NORMALIZADA)	Optimizada para queries complejos (DESNORMALIZADA)
Frecuencia de acceso	Alta	Media y baja
Tipo de acceso	Lectura / escritura Actualización campo por campo	Lectura Sumarización

Datos Operacionales y Datos Warehouse

	Datos operacionales	Data Warehouse
Uso	Predecible Repetitivo	Ad hoc Heurístico
Tiempo de respuesta	Segundos	Segundos a minutos

Explotación del Warehouse

