

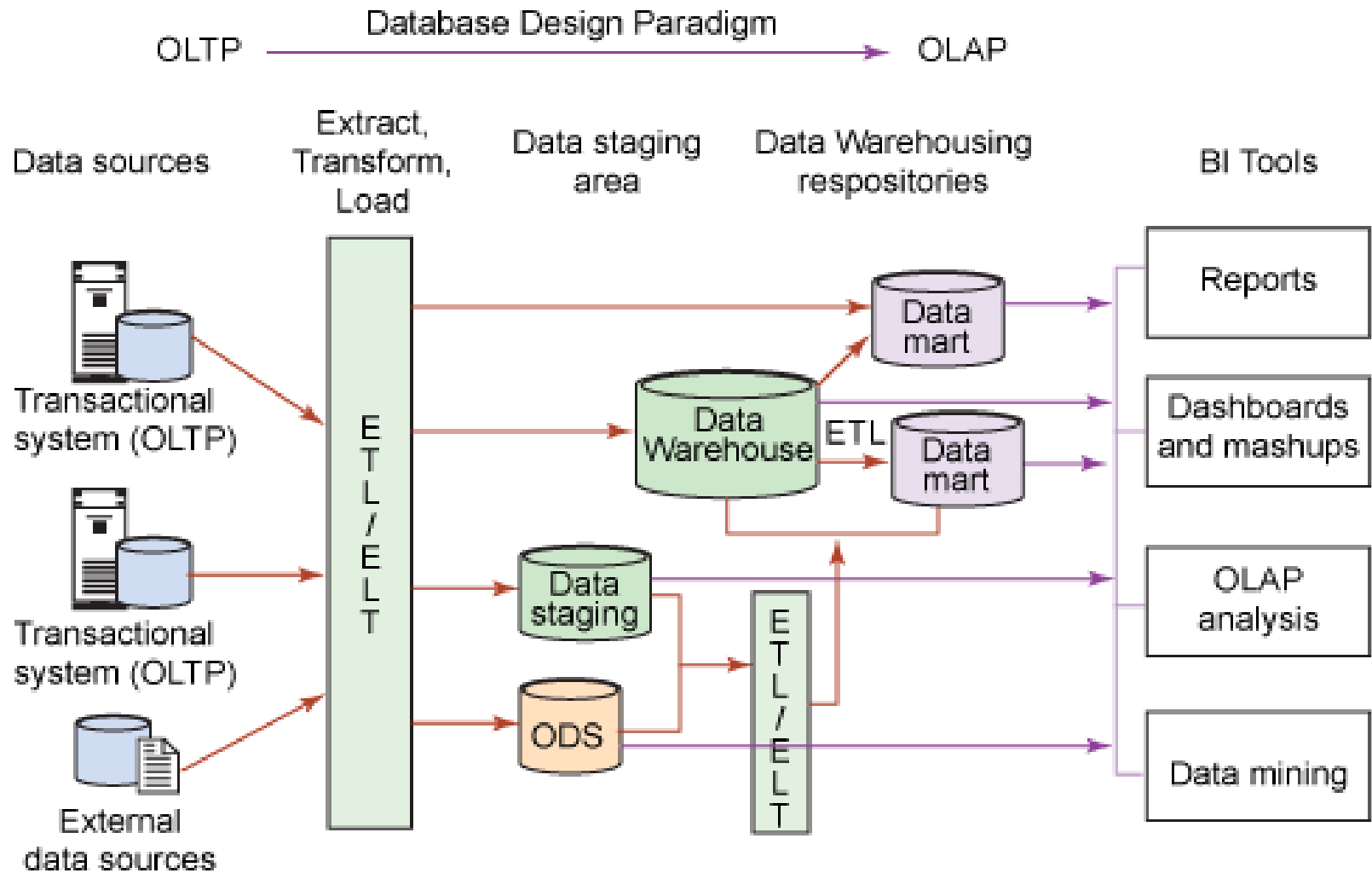


Tecnológico
de Monterrey

BASE DE DATOS AVANZADAS DATA WAREHOUSE

Enrique Alfonso Calderón Balderas

Infraestructura típica de un DWH





Operaciones a realizar para construir un DWH

- Los datos se extraen de Bases de Datos operacionales y de fuentes externas.
- Se hace un proceso de limpieza para minimizar posibles errores, llenando datos faltantes cuando sea posible, consolidando datos equivalentes y realizando cálculos en caso necesario.
- Se hace un proceso de transformación buscando reconciliar diferencias semánticas entre los datos obtenidos (típicamente a través de vistas).
 - Ejemplo: misma persona con diferentes nombres; múltiples formas de referirse a la misma compañía; diferentes # de cuenta generados por diferentes aplicaciones para un mismo cliente; códigos de producto inválidos que se capturaron manualmente; etc.



Operaciones a realizar para construir un DWH

- Se procede con la carga de datos en el DWH vía la materialización de las vistas definidas. En esta fase también se ordena la información y se sumariza según sea necesario.
- Los datos se pueden particionar e indexar para efectos de eficiencia de acceso.
- Periódicamente el DWH se actualiza para reflejar de manera acorde los cambios en las fuentes de datos.

Modelo multidimensional

El objetivo del modelo multidimensional es hacer accesible y sencilla de consultar la información proveniente de múltiples fuentes de datos transaccionales



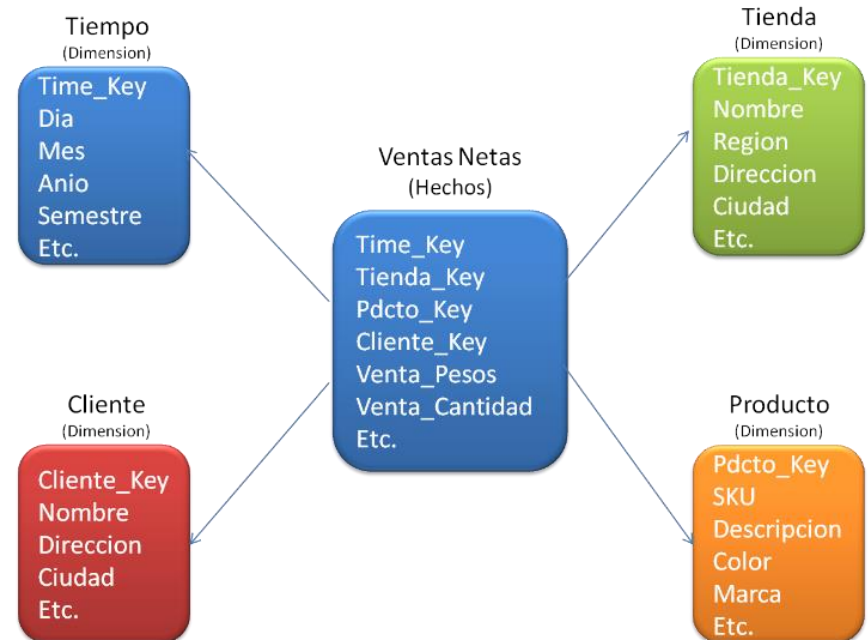
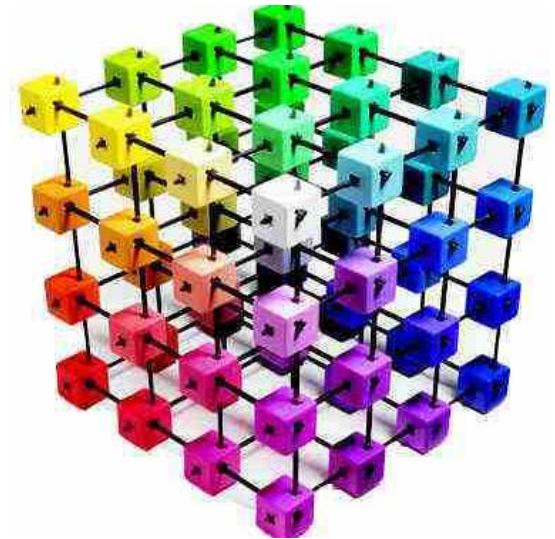
Modelo multidimensional

El modelo multidimensional (lógico)

- Cubos

Se mapea al modelo de Estrella o copos de nieve (físico)

- Tabla de hechos
- Tabla de dimensiones





Multidimensionalidad

PRODUCTO

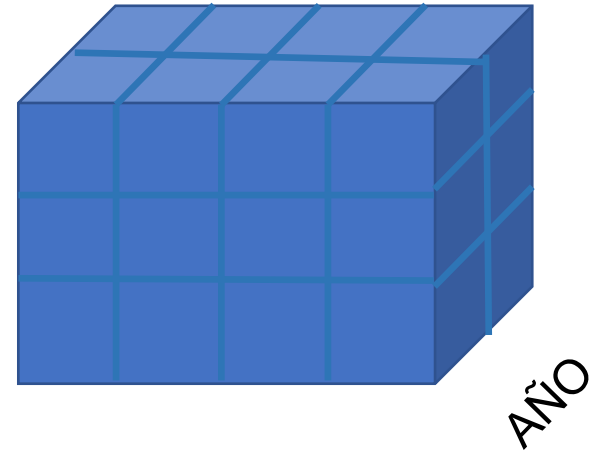
REGION

	R1	R2	R3
P1			
P2			
P3			

DOS DIMENSIONES

PRODUCTO

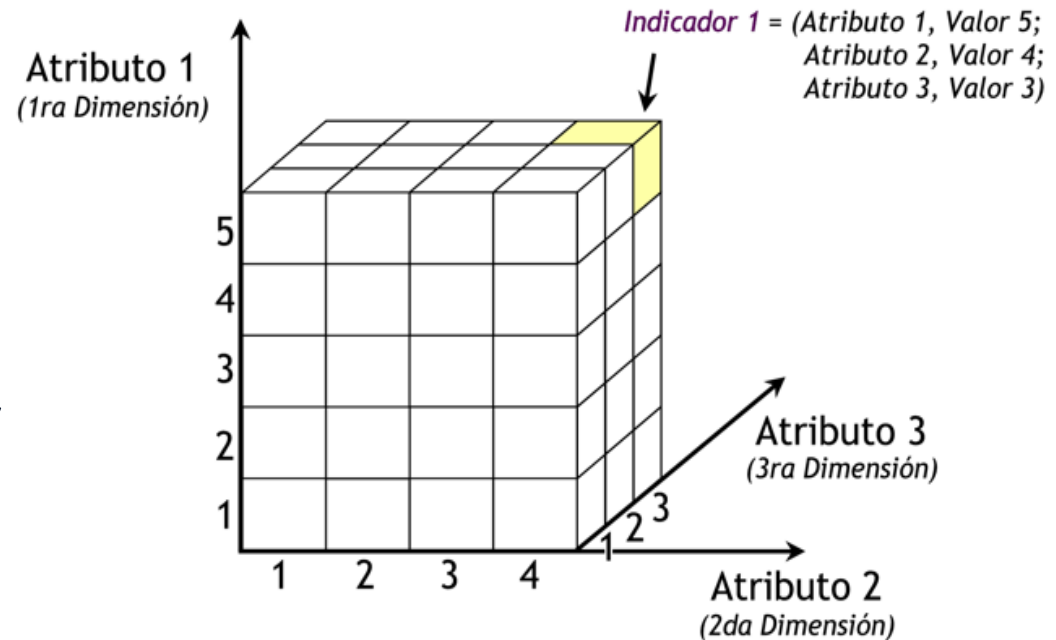
REGION



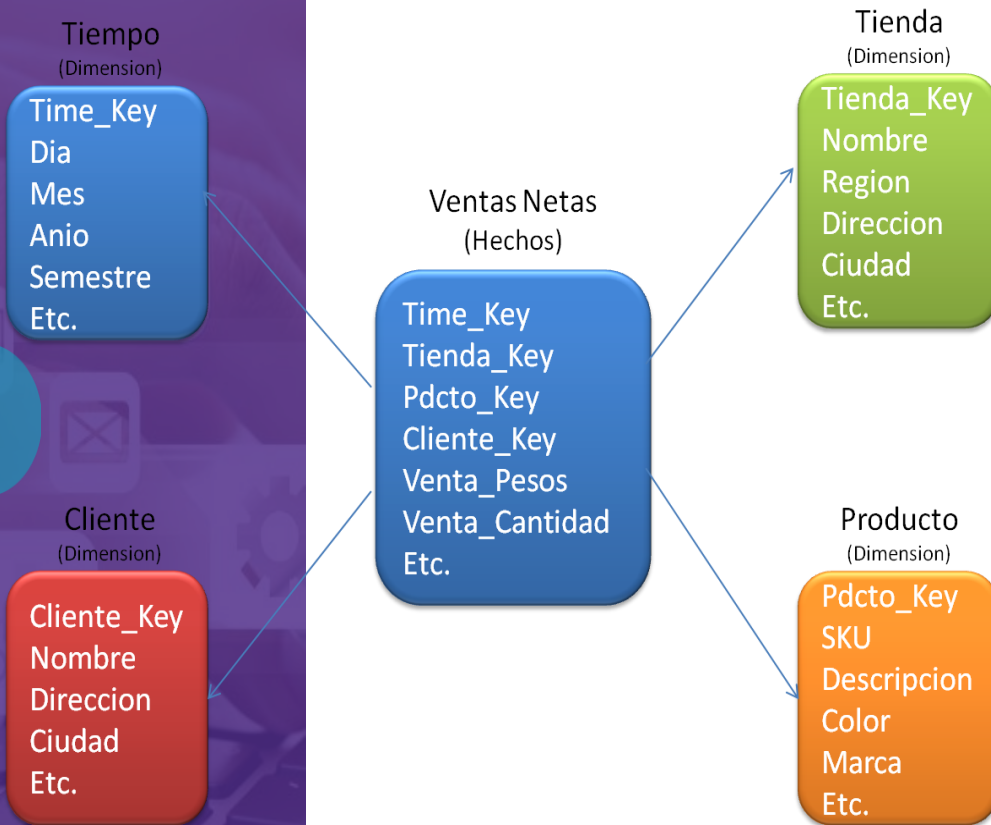
TRES DIMENSIONES

Multidimensionalidad

- “Cubos” de n dimensiones
- Se busca ver los indicadores del negocio a la luz de múltiples dimensiones.
- Debe ser sencillo particionar en cualesquiera de las dimensiones para aislar el conjunto de indicadores de interés y aplicarles el análisis.



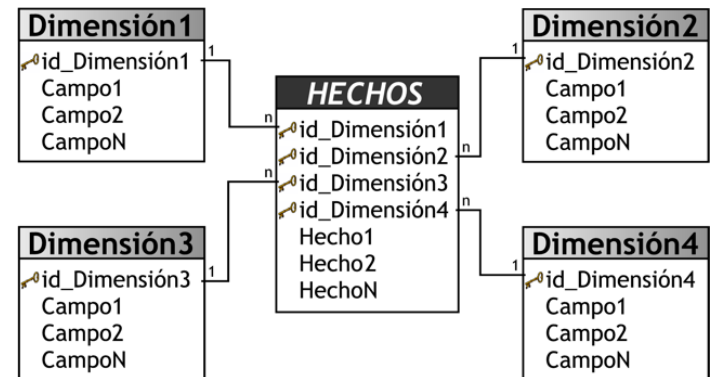
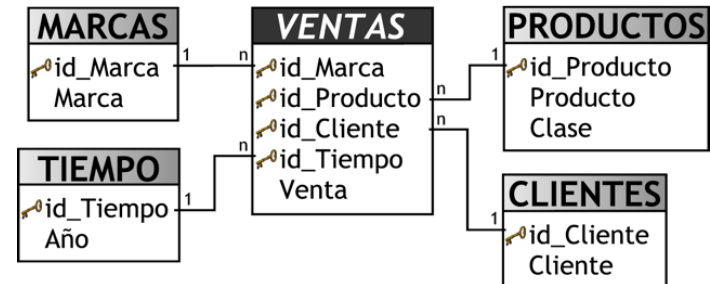
Modelación de un DWH – modelo estrella



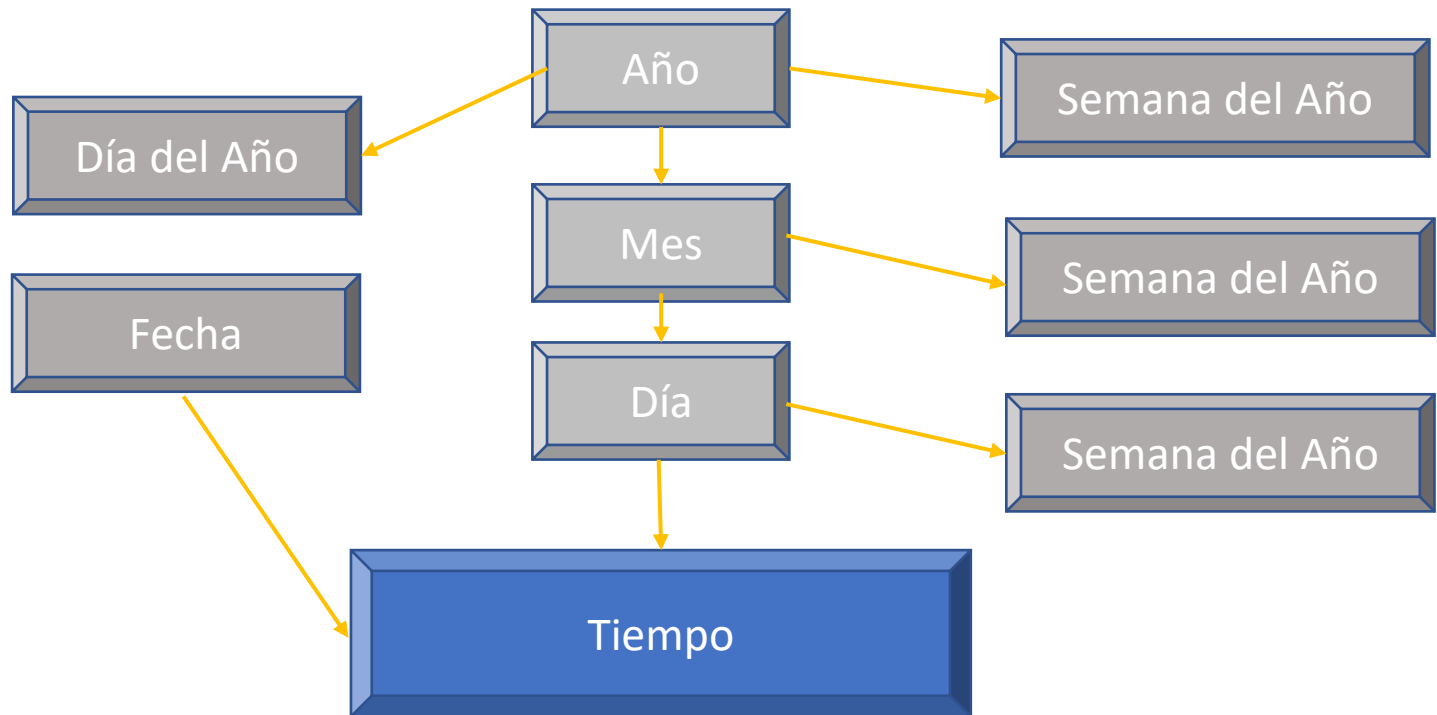
- Compuesto por una tabla llamada “de hechos” y por un conjunto de tablas llamadas “de dimensiones”.
- La tabla de hechos contiene llaves foráneas de TODAS las dimensiones con las que se relaciona. (para garantizar la consistencia de los datos)
- HECHOS – métricas
- DIMENSIONES – descripción

Modelación de un DWH – modelo estrella

- La información contenida en una tabla de hechos por si sola no nos dice nada, la mayoría de las veces contiene únicamente números.
- Por otro lado las dimensiones comúnmente contienen información descriptiva. Los atributos de las dimensiones muchas veces serán los encabezados de las columnas en los reportes.
- En un DW, la creación y el mantenimiento de una tabla de dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se este analizando



Detalle de la dimensión TIEMPO



- Existen muchas maneras de diseñar esta tabla
- Una buena práctica es evaluar con cuidado la temporalidad de los datos
- Separar en Fecha (año, mes, día, etc.) y Hora (hora, minuto, segundo, etc.)



Integración de un DWH

- El modelo de un DWH se integra de la suma de todos los modelos estrella de los cubos requeridos por la organización.
- En el modelo se deben identificar las dimensiones comunes entre los cubos y se busca definirlas una sola vez.
- Un DWH comúnmente esta compuesto por varios Data Marts.



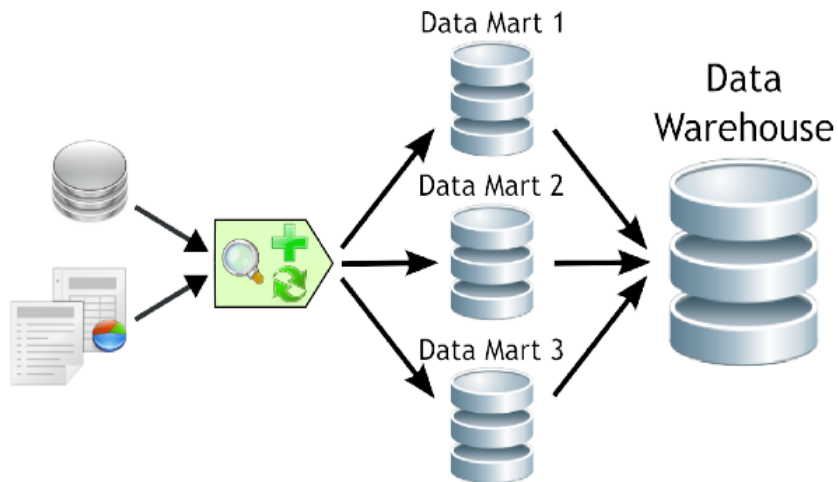
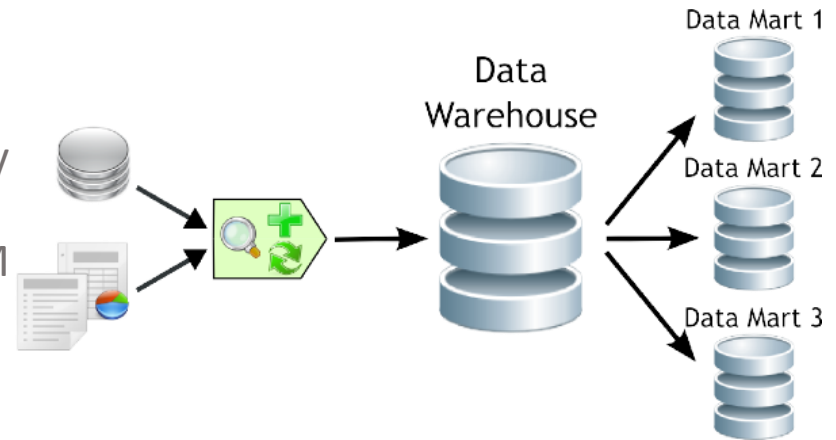
Data Mart

- Un Data Mart es un subconjunto de uno o mas cubos del DWH total, típicamente al servicio de un proceso de negocio particular.
- Implementar Data Marts críticos y gradualmente sumarlos hasta integrar el DWH total es una buena practica.
- De la misma manera que el diseño de una Base de Datos Distribuida podemos usar una de dos arquitecturas.
 - Top-Down
 - Bottom-up

Data Mart

- Top-Down

primero se define el DWH y luego se desarrollan, construyen y cargan los DM a partir del mismo.



- Bottom-Up

En esta arquitectura, se definen previamente los DM y luego se integran en un DWH centralizado.



Traslado de un modelo estrella al modelo relacional

- Cada Cubo se instrumenta con una tabla de hechos y varias tablas de dimensiones.
- Una tabla de dimensión tiene todos los atributos que la describen y el identificador de la misma (llave primaria comúnmente generada por el sistema).
- La tabla de hechos tiene atributos para representar los indicadores y los identificadores que hereda de cada una de las dimensiones del Cubo (llaves foráneas).



Integridad relacional del modelo.

- Se define un constraint de llave primaria o un índice único para el identificador de cada tabla de dimensión
- En la tabla de hechos se define como llave primaria la concatenación de todos los identificadores de las dimensiones.
 - En algunas ocasiones es mas practico usara una llave primaria –autogenerada– y usar las llaves foráneas como índices únicos



Derechos Reservados 2020 Tecnológico de Monterrey
Prohibida la reproducción total o parcial de esta obra sin expresa
autorización del Tecnológico de Monterrey.

Vicerrectoría Académica y de Innovación Educativa