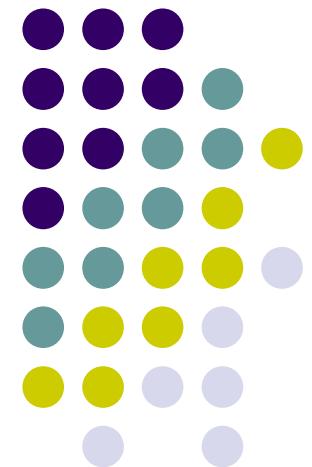
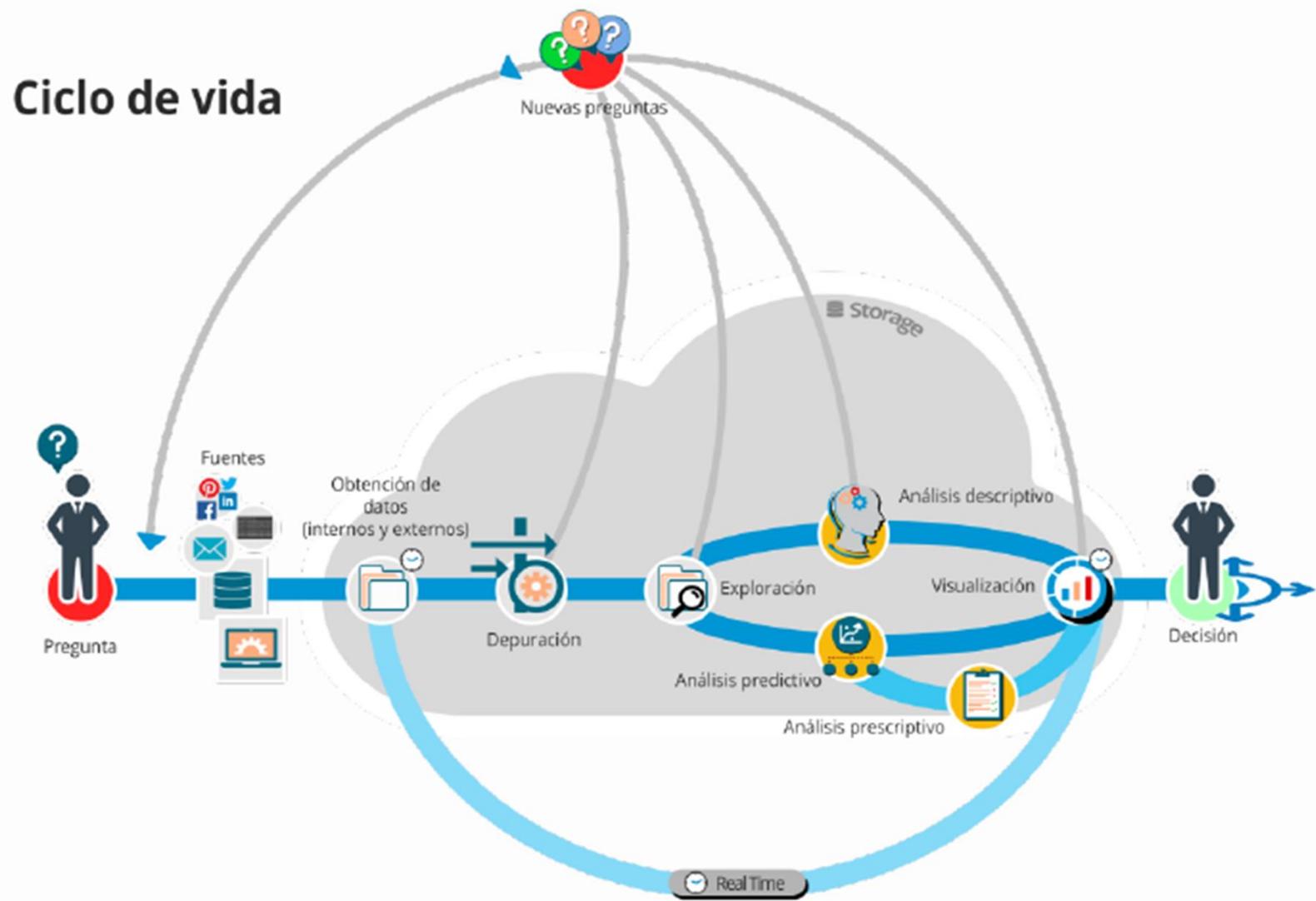


Ingeniería de datos

Tema IV:
El modelo dimensional de datos

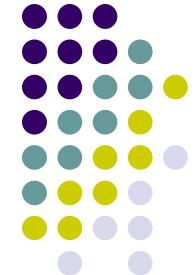


Data Warehouse - Business Intelligence Modelos dimensionales



Data Warehouse - Business Intelligence

Modelos dimensionales



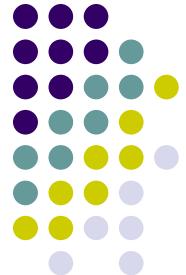
Tipos de sistemas

➤ **Online Transaction Processing (OLTP)**

- Tiempo de respuesta corto
- Pequeñas transacciones
- Operaciones de mantenimiento de datos
- Grandes poblaciones de usuarios
- Alta concurrencia
- Grandes volúmenes de datos
- Alta disponibilidad

➤ **Online Analytical Processing (OLAP)**

- Orientados a la analítica y toma de decisiones
- Se refrescan por lotes
- Se encuentran desnormalizados

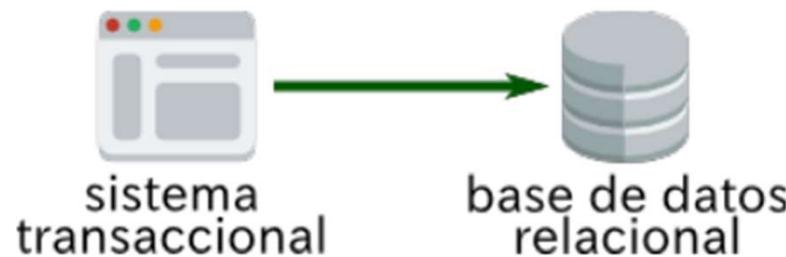


Data Warehouse - Business Intelligence

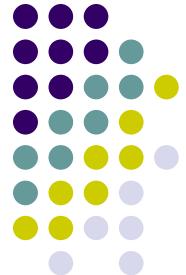
Modelos dimensionales

Business Intelligence

- Las organizaciones generan y acumulan datos constantemente para operaciones y transacciones.
- Se administran con sistemas transaccionales OLTP y se almacenan en bases de datos relacionales (no excluyente).



- La propuesta del BI es procesarlos y utilizarlos como fuente para la toma de decisiones apoyados con herramientas especializadas en presentación y estudio analítico.



Data Warehouse - Business Intelligence

Modelos dimensionales

Business Intelligence

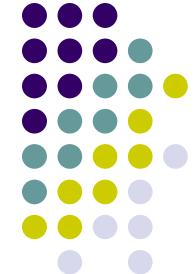
Consiste en una serie de conceptos, procesos, tecnologías y herramientas necesarias para transformar **DATO** en **INFORMACIÓN, INFORMACIÓN** en **CONOCIMIENTO** y **CONOCIMIENTO** en **SABIDURÍA** para generar **PLANES DE ACCIÓN** que sirvan de apoyo a la toma de decisiones del negocio.

Para ello **ALMACENA** y **PROCESA** grandes volúmenes de datos que analiza y explora con herramientas especializadas.



Data Warehouse - Business Intelligence

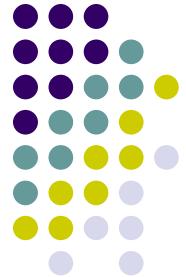
Modelos dimensionales



Business Ingelligence

Traduce los objetivos de la organización en:

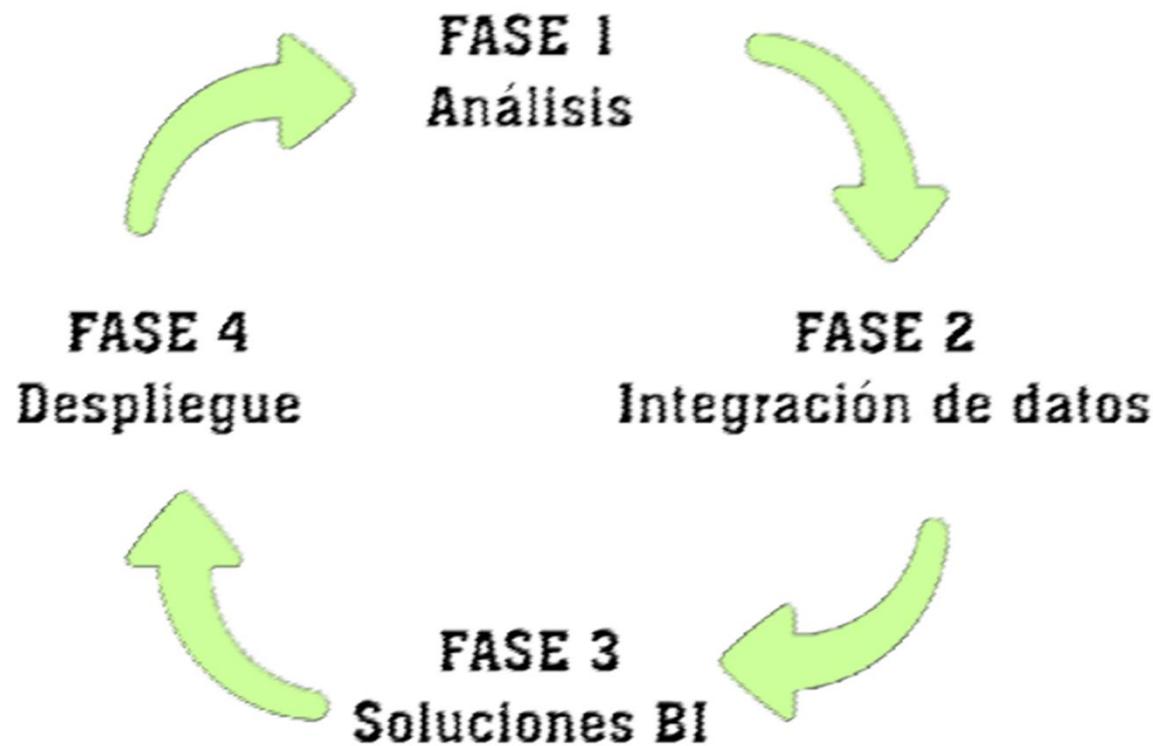
- **INDICADORES DE ESTUDIOS** (cantidad vendida, importe pagado, etc.)
- **CRITERIOS DE ANÁLISIS** (año de venta, rubro de ventas, etc.)
- **PERSPECTIVAS O ESCENARIOS DE ANÁLISIS** (conjunto de criterios)

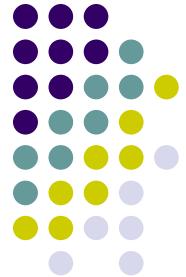


Data Warehouse - Business Intelligence

Modelos dimensionales

Proceso BI

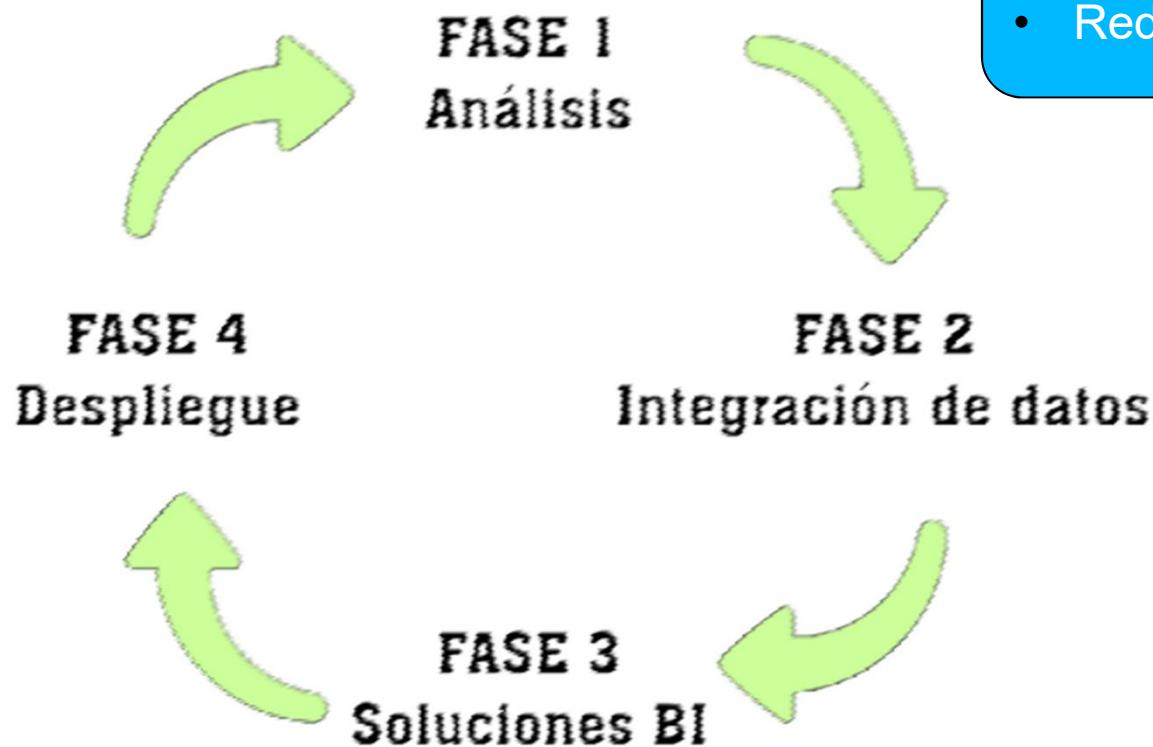


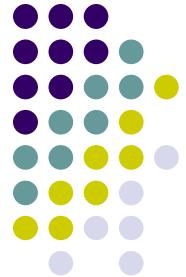


Data Warehouse - Business Intelligence

Modelos dimensionales

Proceso BI

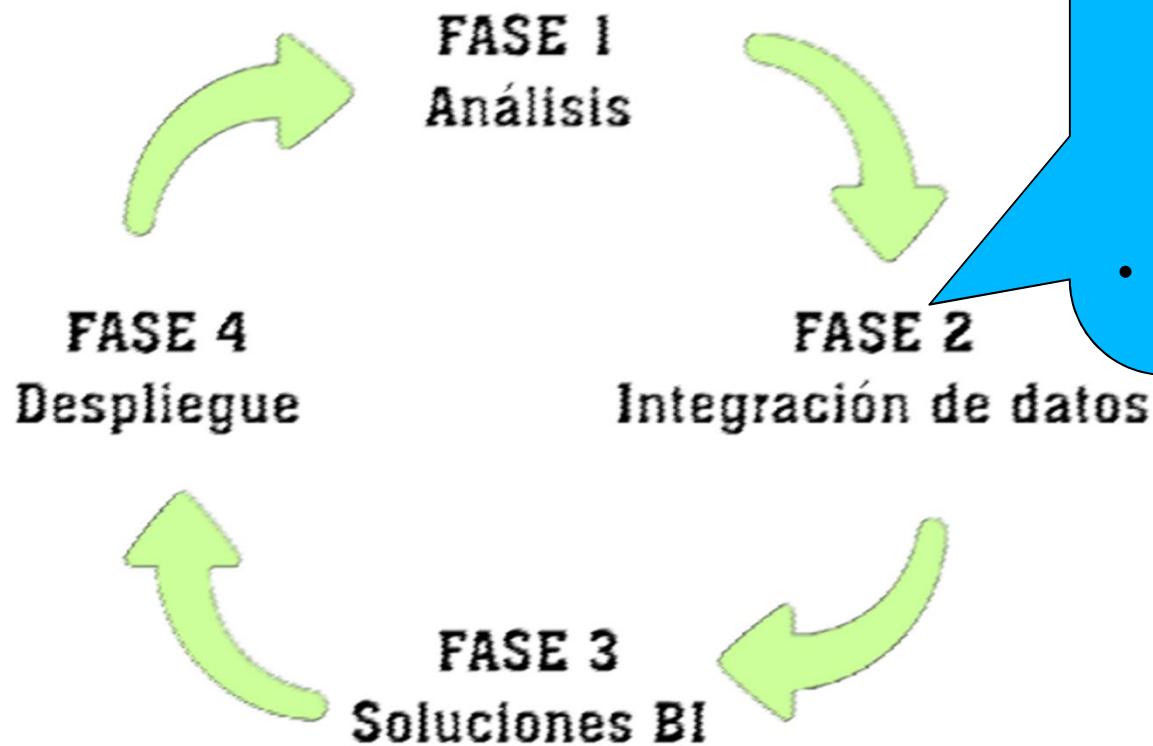




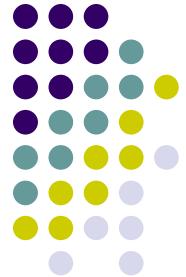
Data Warehouse - Business Intelligence

Modelos dimensionales

Proceso BI



- Integración de datos **internos** y **externos**
- Transformación de esos datos que poseen la información para dar las respuestas a las preguntas de la Fase 1.
- Carga en la base de datos para servir al BI.

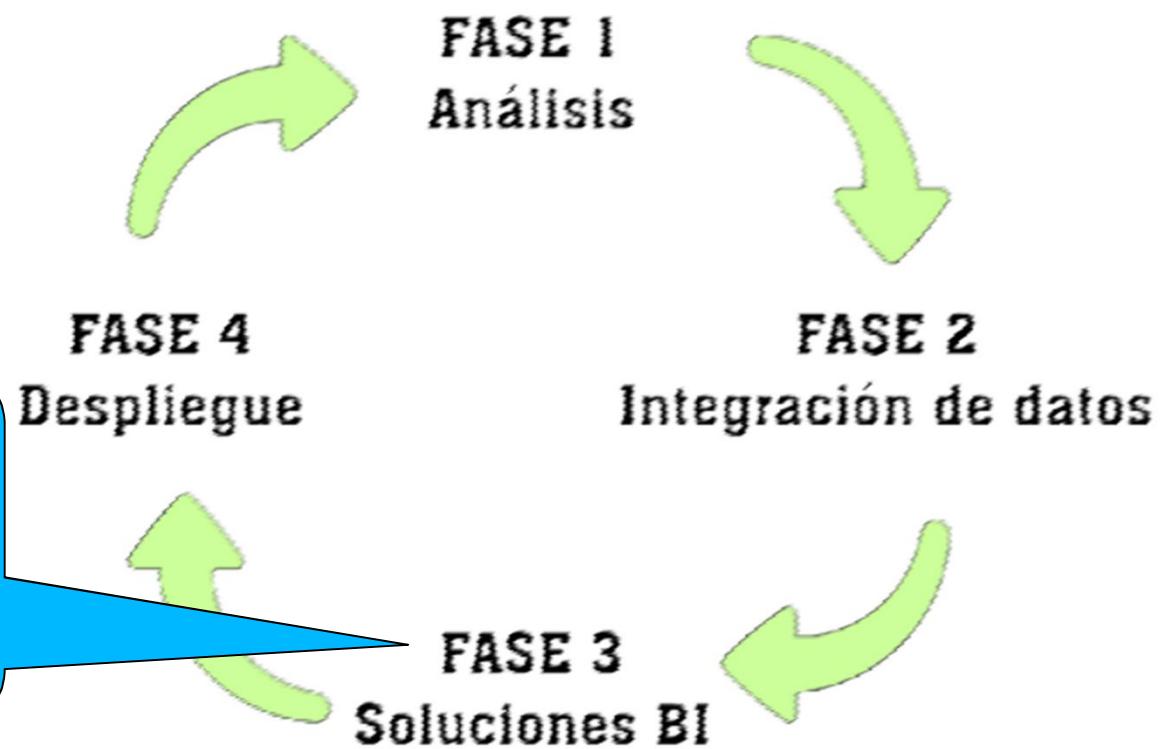


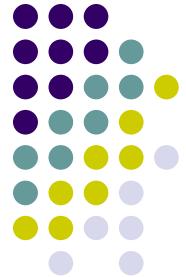
Data Warehouse - Business Intelligence

Modelos dimensionales

Proceso BI

- Uso de herramientas y técnicas de explotación.
- Reportes, tableros, indicadores, gráficos estadísticos, etc..



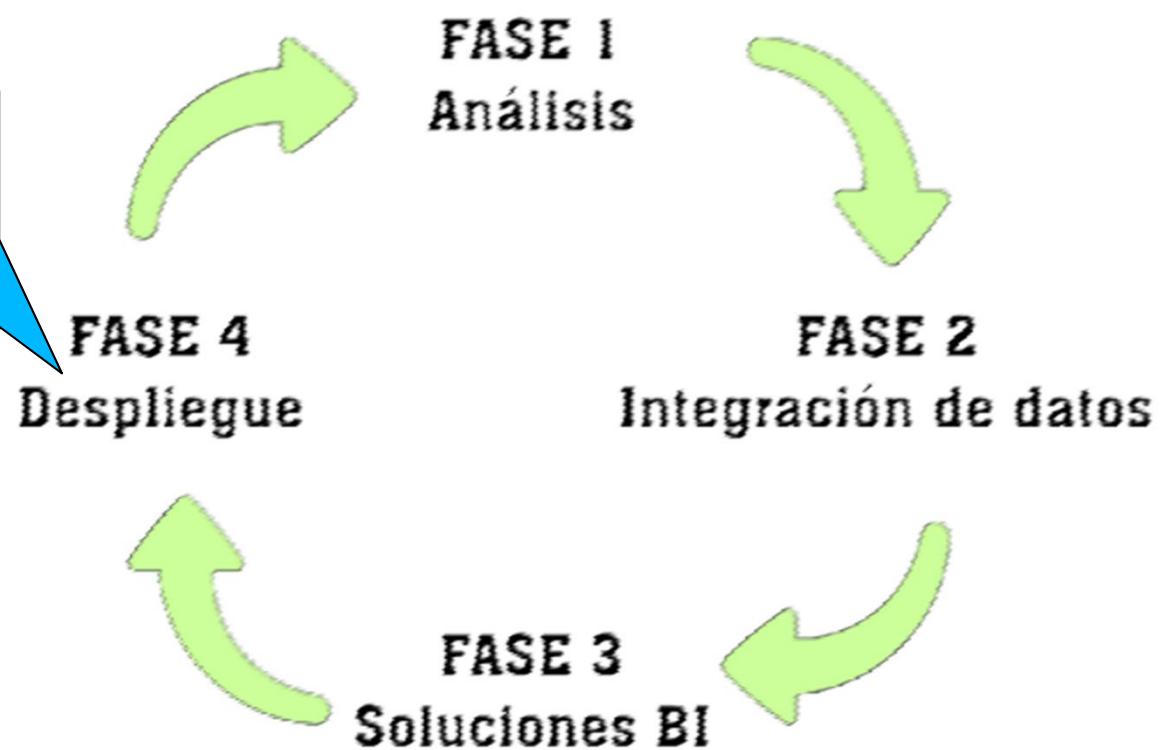


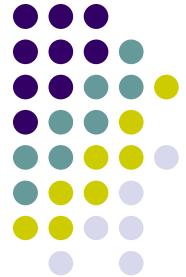
Data Warehouse - Business Intelligence

Modelos dimensionales

Proceso BI

- Entrega al usuario.
- Capacitación.
- Obtención de respuestas
- Nuevas preguntas ...





Data Warehouse - Business Intelligence

Modelos dimensionales

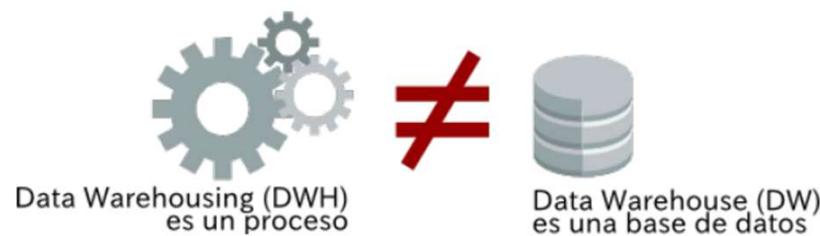
Data Warehousing & Data Warehouse

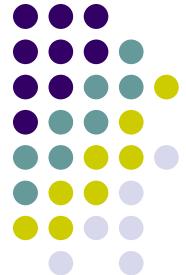
Data Warehousing (DWH):

- **Proceso** → extracción, transformación, consolidación, integración y centralización de los datos **internos y externos relacionados** en una base de datos para permitir el análisis y exploración y dar soporte al proceso de toma de decisiones.

Data Warehouse (DW):

- **Base de datos** → almacenamiento de los datos para el análisis. Es una base de datos que posee una estructura multidimensional.





Data Warehouse - Business Intelligence

Modelos dimensionales

Data Warehousing & Data Warehouse

Data Warehouse (DW):



Data
Warehouse

colección de datos
para el soporte de la toma de decisiones



orientada
al negocio



integrada



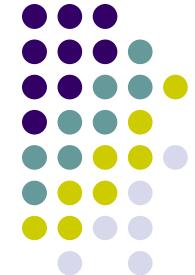
variante en
el tiempo



no volátil

Data Warehouse - Business Intelligence

Modelos dimensionales



Data Warehousing & Data Warehouse

Data Warehouse (DW):

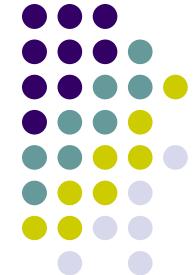


Sólo ingresarán **datos relevantes** para el análisis y toma de decisiones. Muchos datos **NO son tenidos en cuenta para el DW** por carecer de valor analítico.

Se manejan conceptos de alto nivel en una estructura multidimensional que administre los datos de **entidades tipo PERMANENTE** y de tipo **MOVIMIENTO** en tablas denominadas **DIMENSIONES** y **HECHOS** respectivamente.

Data Warehouse - Business Intelligence

Modelos dimensionales



Data Warehousing & Data Warehouse

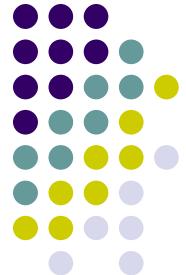
Data Warehouse (DW):



Todos los datos de orígenes heterogéneos se analizan para asegurar su calidad y pureza para consolidarse luego en el DW. Se materializa con las tareas de ETL.

Orígenes de los datos:

- Usuarios: Operacionales, Medios, Gerenciales.
- Áreas de la organización.
- Diferentes Data Sources: fuentes internas, fuentes externas.



Data Warehouse - Business Intelligence

Modelos dimensionales

Data Warehousing & Data Warehouse

Data Warehouse (DW):



- Los datos actuales son almacenados junto a los datos históricos cada uno con su **marca de tiempo**.
- Permite acceder a distintas versiones de la misma situación, de esta manera, se podrá avanzar y retroceder en el tiempo manteniendo el foco de atención sobre la **situación analizada**.

Data Warehouse - Business Intelligence

Modelos dimensionales



Data Warehousing & Data Warehouse

Data Warehouse (DW):



no volátil

La información **ES ESTABLE** → Los datos que ingresan al DW **NO CAMBIAN**.



DATA SOURCES

🔍 Consultar

✚ Insertar

✖ Eliminar

✎ Modificar



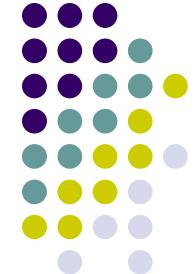
DATA WAREHOUSE

🔍 Consultar

✚ Insertar

Data Warehouse - Business Intelligence

Modelos dimensionales

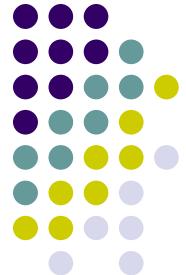


Data Warehousing & Data Warehouse

Granularidad:

Es el nivel de detalle con el que se desea almacenar la información

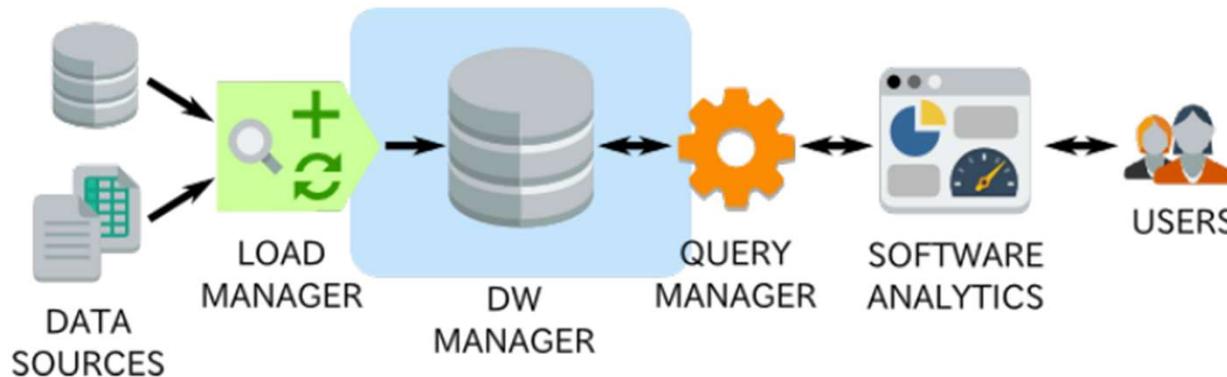
- A mayor nivel de detalle de los datos, más posibilidades analíticas,
- Los datos que posean granularidad fina (nivel de detalle) podrán ser resumidos para obtener una granularidad media o gruesa.
- Trabaja juntamente con la agregación (cálculo que de varias filas produce un único resultado) que es una situación **NO REVERSIBLE**.



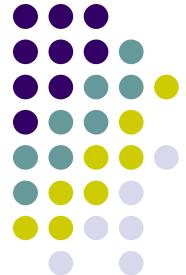
Data Warehouse - Business Intelligence

Modelos dimensionales

Arquitectura Data Warehousing



- Los datos son extraídos desde distintas fuentes.
- Los datos son integrados, limpiados, depurados y transformados y se cargan en el DW donde se construirán cubos multidimensionales y/o business models (**modelos dimensionales**).
- Los usuarios accederán a los cubos multidimensionales o business models del DW utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.



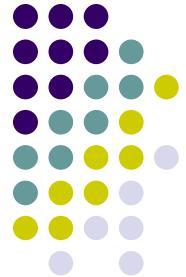
Data Warehouse - Business Intelligence

Modelos dimensionales

Arquitectura – Data Sources



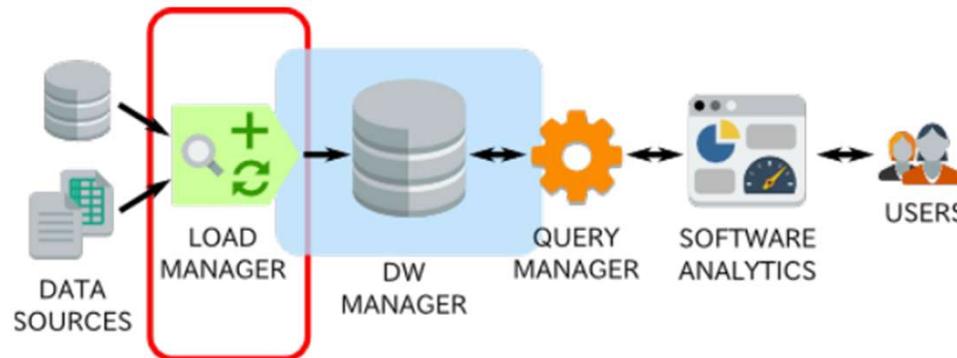
- Archivos de textos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales, SQL y NoSQL .
- Información no estructurada (páginas web, mails) Redes sociales.
- Web Services.
- Etc...



Data Warehouse - Business Intelligence

Modelos dimensionales

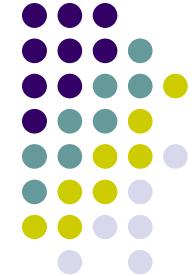
Arquitectura – Load Manager



- **INTEGRACIÓN:** procesos de extracción y transformación (manipulación, control y depuración). Luego los resultados actualizan el DW (carga).
- La **Extracción** incluirá técnicas enfocadas a obtener desde diversas fuentes **sólo los datos relevantes** y mantenerlos en una ***Staging Area***.
- La **Transformación** incluirá técnicas encargadas de compatibilizar formatos, filtrar y clasificar datos, relacionar diversas fuentes, depurar, etc.
- La **Carga** incluirá técnicas propias de la carga y actualización del DW.

Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Load Manager

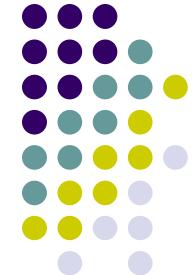
Extracción

- Seleccionar adecuadamente los datos relevantes.
- Si los Data Sources utilizan bases de datos relacionales, la extracción puede hacerse con consultas SQL o procedimientos almacenados.
- Los casos NO convencionales o NO estructurados la obtención se hará con elementos variados específicos.
- Una vez extraídos, los datos se persisten en una base de datos relacional **Staging** desconectándose de los Data Sources



Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Load Manager

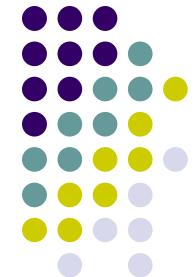
Extracción

BASE DE DATOS DEL STAGING AREA

- Pueden ser varias bases de datos, esquemas o particiones dependiendo del DBMS a utilizar.
- Permiten manipular los datos sin interrumpir ni sobrecargar los Data Sources y el DW.
- Crean una capa de abstracción entre la lectura y la carga.
- Cuando se terminan las transformaciones en éste área continua la población del DW.



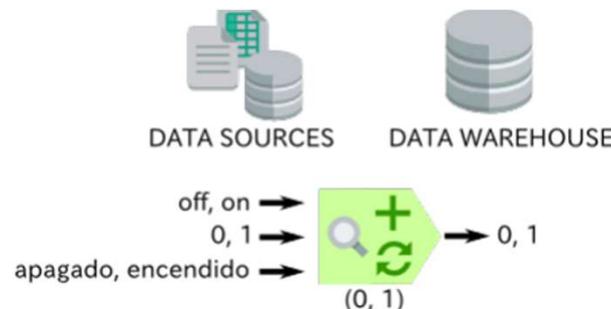
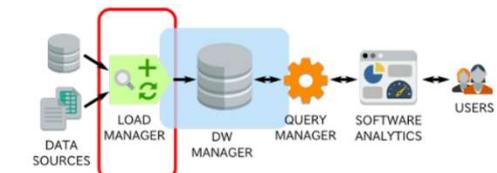
Data Warehouse - Business Intelligence Modelos dimensionales



Arquitectura – Load Manager

Transformación →

Datos Consistentes,
compatibles y congruentes



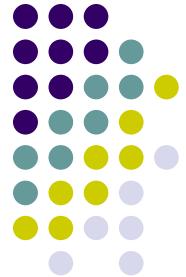
Codificación



Medida de atributos



Fuentes múltiples



Data Warehouse - Business Intelligence

Modelos dimensionales

Arquitectura – Load Manager

Transformación

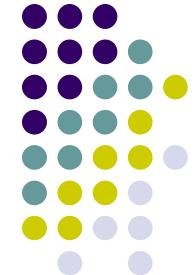
- Procesos de **Limpieza** y **Calidad** de los datos.
- Identificar la razón de la anomalía y evitar repeticiones.

Acciones ante datos anómalos o faltantes:

- Ignorarlos.
- Eliminar la columna (se eliminan el 100% de los datos).
- Filtrar la columna (se eliminan algunos de los datos).
- Filtrar la fila errónea,
- Reemplazar el valor.
- Si falta, esperar a que esté disponible.
- Discretizar los valores de las columnas. Si una columna numérica presenta un valor de 1 a 2, se utiliza el texto **Bajo**; de 3 a 7, **Medio**; de 8 a 10, **Alto**. Al ocurrir un outlier se puede reemplazar por Bajo o Alto.



Data Warehouse - Business Intelligence Modelos dimensionales



Arquitectura – Load Manager

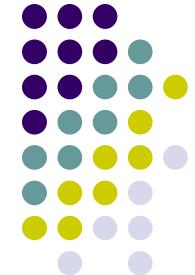
Carga

- **Carga inicial**
- **Actualización periódica:**
 - Cotejar instancias de los Data Sources involucrados.
 - Uso de *triggers*
 - Utilizar *marcas de tiempo* en los Data Sources.
 - Comparar los datos entre Data Source y DW.
 - Técnicas mixtas.
 - **FULL LOAD** (vaciado previo)
- **Mantener elementos estructurales del DW:**
 - Claves subrogadas.
 - SCD.
 - Etc.



Data Warehouse - Business Intelligence

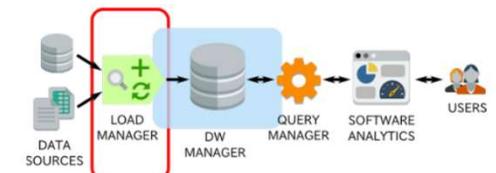
Modelos dimensionales

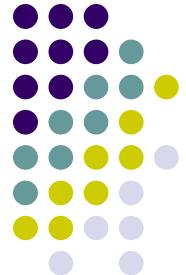


Arquitectura – Load Manager

Resumen → ETL

- Extracción de los datos desde Data Sources que se depositan en la **Staging Area**.
- Integración y transformación de los datos – tratamiento de las inconsistencias.
- Carga de los datos desde **Staging Area** al DW

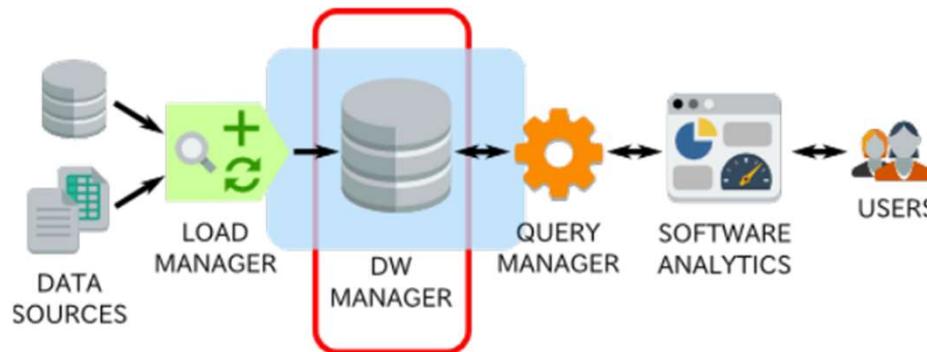




Data Warehouse - Business Intelligence

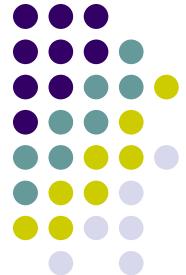
Modelos dimensionales

Arquitectura – DW Manager



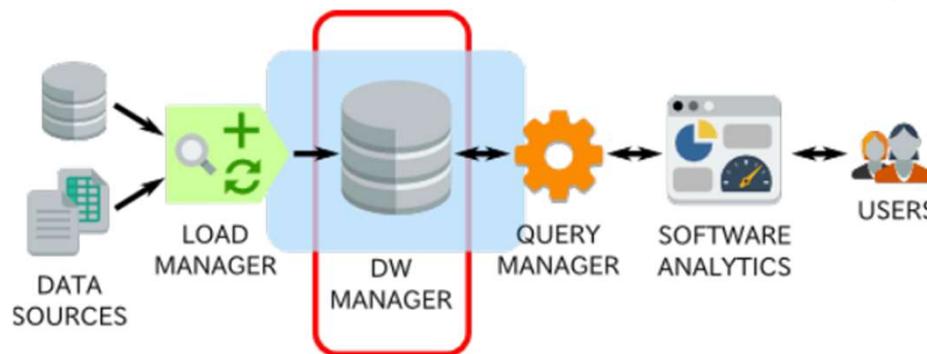
Compuesto por:

- el DW (DBMS),
- las conexiones a base de datos y otros Data Sources,
- las estructuras de datos (cubos multidimensionales, business models),
- información de autenticación y autorización (credenciales de acceso, users, roles, permisos, etc), y otros metadatos.



Data Warehouse - Business Intelligence Modelos dimensionales

Arquitectura – DW Manager

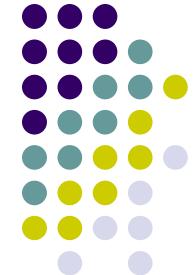


Es una base de datos con estructura multidimensional que es una forma de almacenamiento con dos elementos principales:

- ✓ **Tablas de Hechos**
- ✓ **Tablas de Dimensiones**

Data Warehouse - Business Intelligence

Modelos dimensionales



Modelos básicos



Star scheme (esquema en estrella)

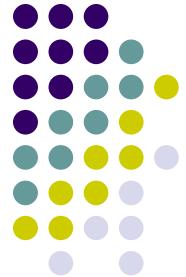


Snowflake scheme (esquema copo de nieve)



Starflake scheme (esquema constelación)

Estos modelos permiten facilitar el acceso a consultas complejas y con gran cantidad de agregaciones, es por ello que se encuentran desnormalizadas o semi desnormalizadas, reduciendo de esta manera al mínimo la cantidad de JOINs que deben emplearse para acceder a los datos requeridos.



Data Warehouse - Business Intelligence

Modelos dimensionales

Tablas de Dimensiones

- Permiten analizar los datos en el contexto del negocio.
- Datos cualitativos
- Aspectos de interés para filtrar
- Perspectivas de visión

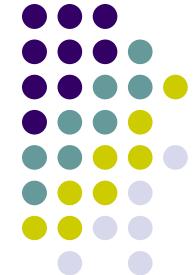
Columnas:

- Clave principal (primary key – por ejemplo **id**)
- Claves ajena (solo para esquemas copo de nieve y constelación).
- Datos **primarios**: datos que identifican la Dimensión de manera básica (por ejemplo **nombre_infraccion**).
- Datos **secundarios**: datos que complementan la descripción de la Dimensión (por ejemplo **es_grave**) Estos datos no son significativos para tomar decisiones, pero son potencialmente valiosos para implementarla incluso pueden transformarse en otras dimensiones.

d_infracciones
123 id
ABC nombre_infraccion
ABC es_grave

Data Warehouse - Business Intelligence

Modelos dimensionales

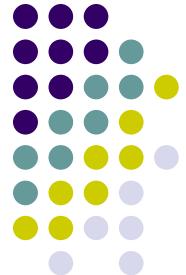


Tablas de Dimensiones

Recomendación !!!!

Las claves principales de las tablas de Dimensiones deben ser **independientes** de las claves de los Data Sources ya que, si estos últimos son recodificados, el DW quedaría inconsistente.

Estas claves serán Claves Subrogadas



Data Warehouse - Business Intelligence

Modelos dimensionales

Dimensiones especiales

La dimensión **TIEMPO**

No es solamente una secuencia cronológica representada de forma numérica, sino que mantiene **niveles jerárquicos especiales** que permiten analizar eventos acontecidos en un día de la semana, quincena, mes, trimestre, semestre, año, estación, etc.... **Y LA HORA** (o rangos de ellas).

d_fecha
123 id
fecha
123 epoch
123 dia_del_mes
ABC dia_nombre
ABC dia_nombre_corto
123 dia_semana
123 dia_trimestre
123 dia_anio
123 semana_mes
123 semana_anio
123 mes_numero
ABC mes_nombre
ABC mes_nombre_corto
123 trimestre_numero
ABC trimestre_nombre
ABC trimestre_nombre_corto
ABC bimestre_nombre
ABC bimestre_nombre_corto
ABC semestre_nombre
123 anio_actual
fecha_primerdia_semana
fecha_ultimodia_semana
fecha_comienzo_trimestre
fecha_fin_trimestre
ABC mes2_anio4
ABC mes2_dia2_anio4
<input checked="" type="checkbox"/> es_finde
<input checked="" type="checkbox"/> es_feriado
ABC nombre_feriado
ABC fechar

Data Warehouse - Business Intelligence

Modelos dimensionales



Dimensiones especiales

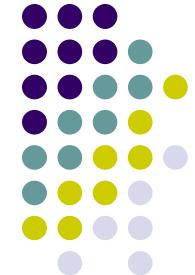
La dimensión **HORA**

Cuando se requiere la hora, se debe crear una dimensión independiente de la de fecha y conjuntamente a ésta, determinan el momento de cualquier ocurrencia.

	d_hms
	123 id
	hms
	ABC grupo1

Data Warehouse - Business Intelligence

Modelos dimensionales



Tablas de Hechos

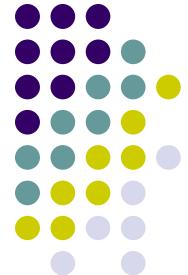
- Las tablas de Hechos contienen los **acontecimientos** o **eventos** que serán utilizados por los usuarios del DW para analizar y responder preguntas de negocio.
- Los Hechos contienen datos que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de Dimensiones.

h_alojamiento
123 id_muerto
123 id_destino
123 id_fecha
123 id_accion_alojamiento
123 id_fecha_consulta

h_liquidaciones
123 id_tipo_liquidacion
123 id_concepto_liquidacion
123 id_fecha
123 id_estado_liquidacion
123 importe
123 id_fecha_consulta

Data Warehouse - Business Intelligence

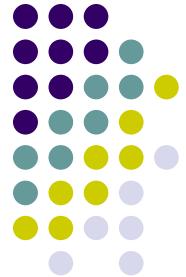
Modelos dimensionales



Tablas de Hechos

- Las tablas de Hechos contienen los **acontecimientos** o **eventos** que serán utilizados por los usuarios del DW para analizar y responder preguntas de negocio.
- Los Hechos contienen datos que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de Dimensiones.

h_actas	
ABC	id_acta
ABC	id_dominio
123	id_fecha
123	id_hms
123	id_sexo
123	id_edad
123	id_infraccion
123	id_tipo_vehiculo
123	id_estado
123	id_actor
123	id_origen
123	id_marca
123	id_modelo
ABC	tiene_titulo_ejecutivo
ABC	es_liquidacion_web
ABC	pagada
ABC	otra_jurisdiccion
ABC	fue_retenido
123	latitud
123	longitud
123	importe_teorico
123	importe_abonado
ABC	tiene_georreferencia

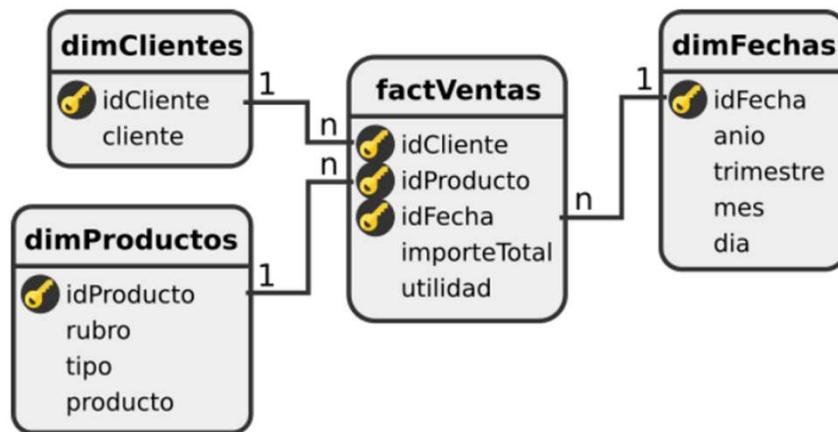


Data Warehouse - Business Intelligence

Modelos dimensionales

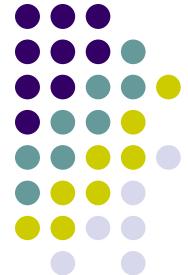
Tablas de Hechos

- La tabla de Hechos posee un identificativo compuesto por la **concatenación de las claves de las tablas de las Dimensiones relacionadas**. No requiere de la definición de una primary key.



Data Warehouse - Business Intelligence

Modelos dimensionales



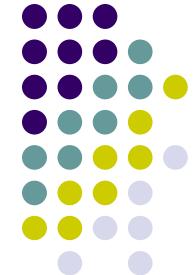
Tablas de Hechos

- Se denominan también **hecho** a los atributos que cuantifican o califican al evento u acontecimiento que representa el Hecho.
- Los **hechos básicos** son los que están como columna en la tabla como ser `importe_unitario` (del artículo), `cantidad_comprada`.
- Los **hechos derivados** son los que se calculan a partir de una expresión combinando hechos básicos como ser:

`importe_línea` = `importe_unitario * cantidad_comprada`

Data Warehouse - Business Intelligence

Modelos dimensionales

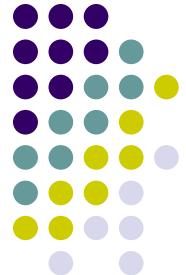


Tablas de Hechos

Indicadores/medidas

Hacen uso de los datos de los hechos para obtener **VALORES ANALIZABLES** y se definen con los siguientes metadatos:

- nombre representativo, descripción;
- tipo de agregación al momento de crear una estructura de datos (SUM, MAX, MIN, COUNT, AVG, porcentajes, fórmulas, etc.)
- agregaciones alternativas (combinación de agregaciones);
- tipo de datos (siempre numéricos).
- Se calculan dinámicamente sin necesidad de persistir en la estructura de datos.



Data Warehouse - Business Intelligence

Modelos dimensionales

Conceptos

Medidas:

son datos operativos que pueden estar agregados o detallados. Representan un hecho de la realidad; por ejemplo, una venta o descuento.

Métricas:

son valores cuantitativos que miden un desempeño. Son medidas dentro de un contexto definido. Continuando con el ejemplo de venta, podemos tener “venta año anterior” o “venta promedio”.

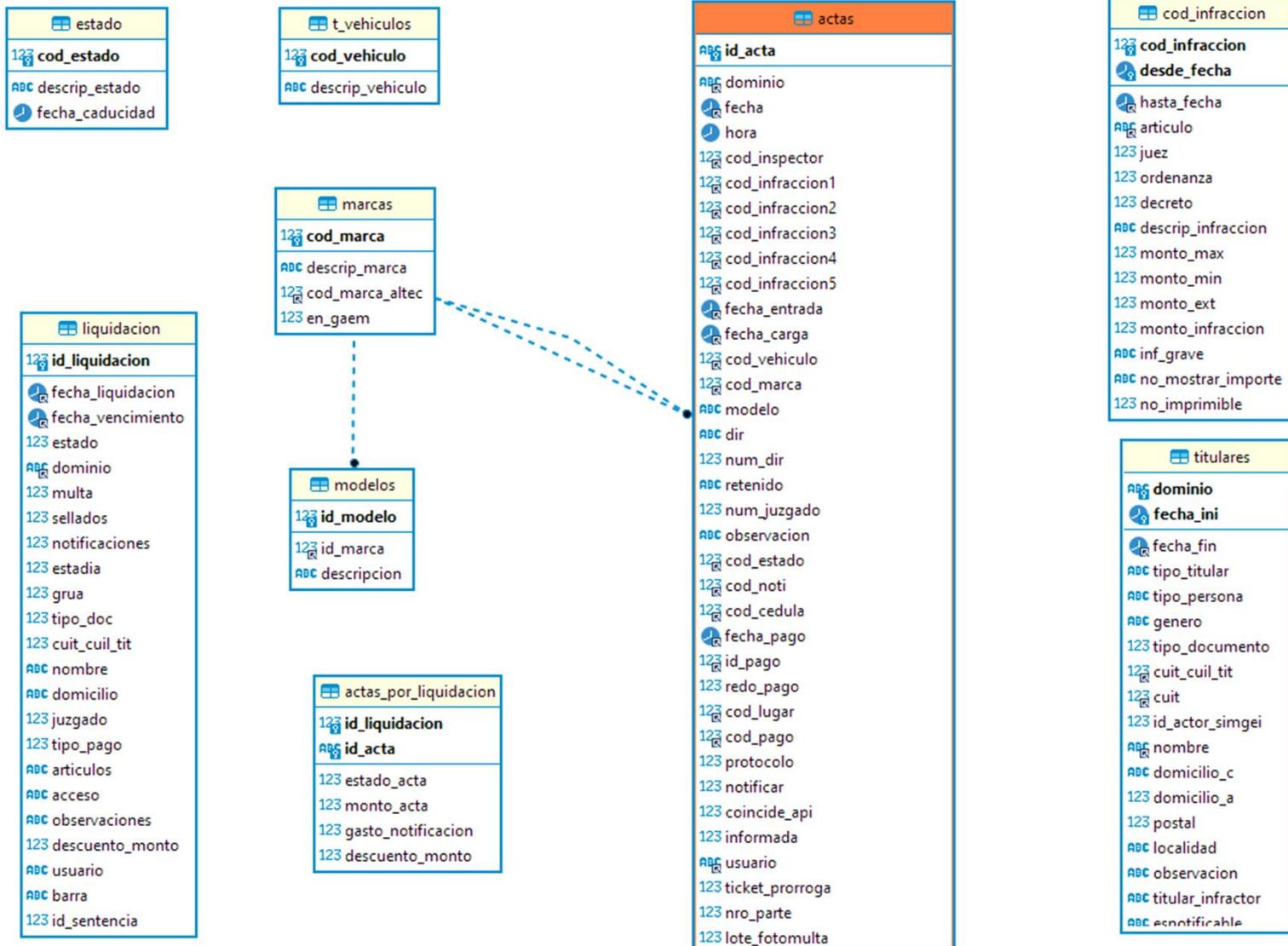
KPI (indicador):

son métricas o medidas ponderadas y evaluadas. Se necesita definir qué valor es “bueno”, cuál es “malo” y cuál es “aceptable”. Sobre la base de estas premisas podemos analizar el estado, su tendencia y su variación respecto a lo planificado o proyectado.



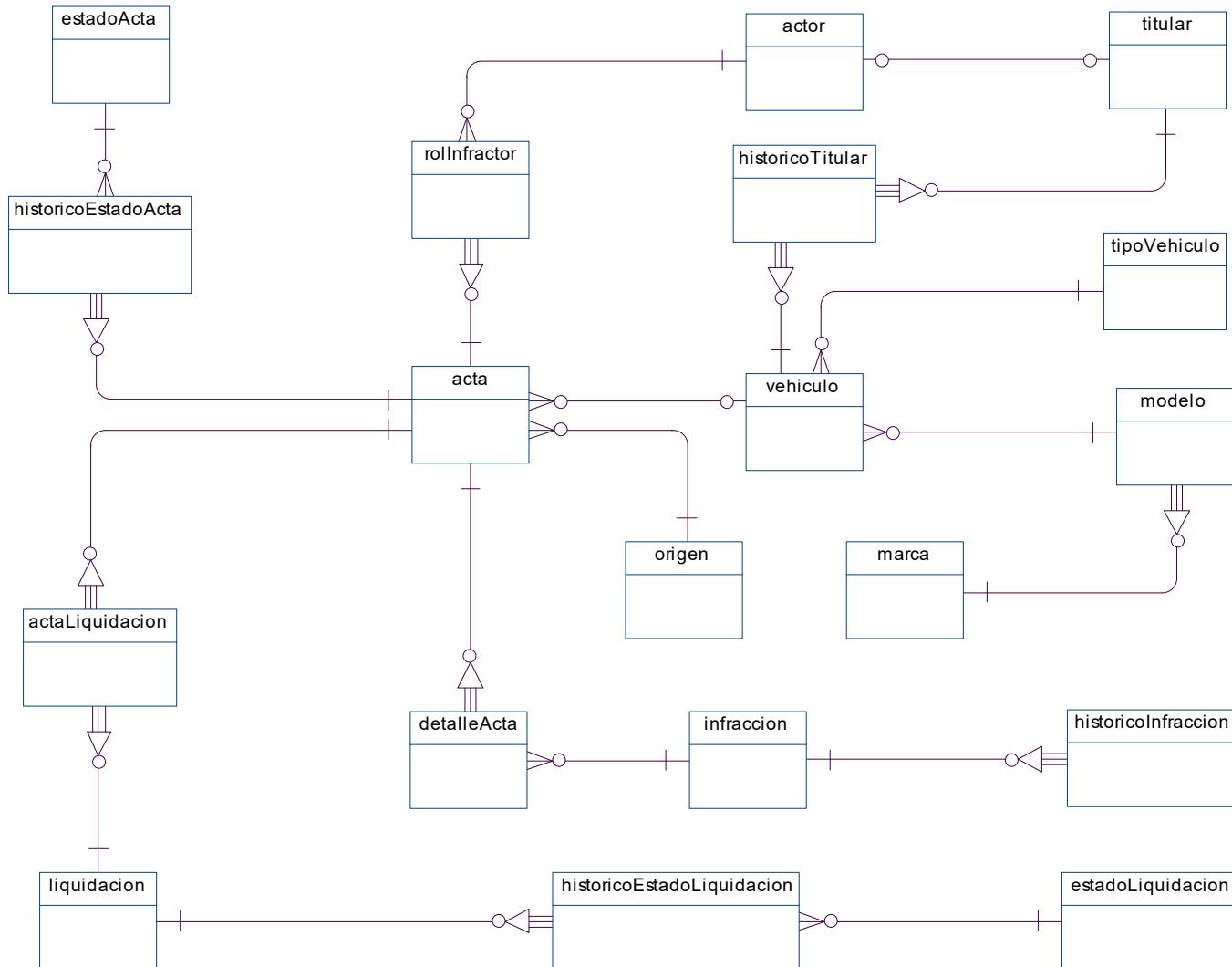
Data Warehouse - Business Intelligence

Modelos dimensionales



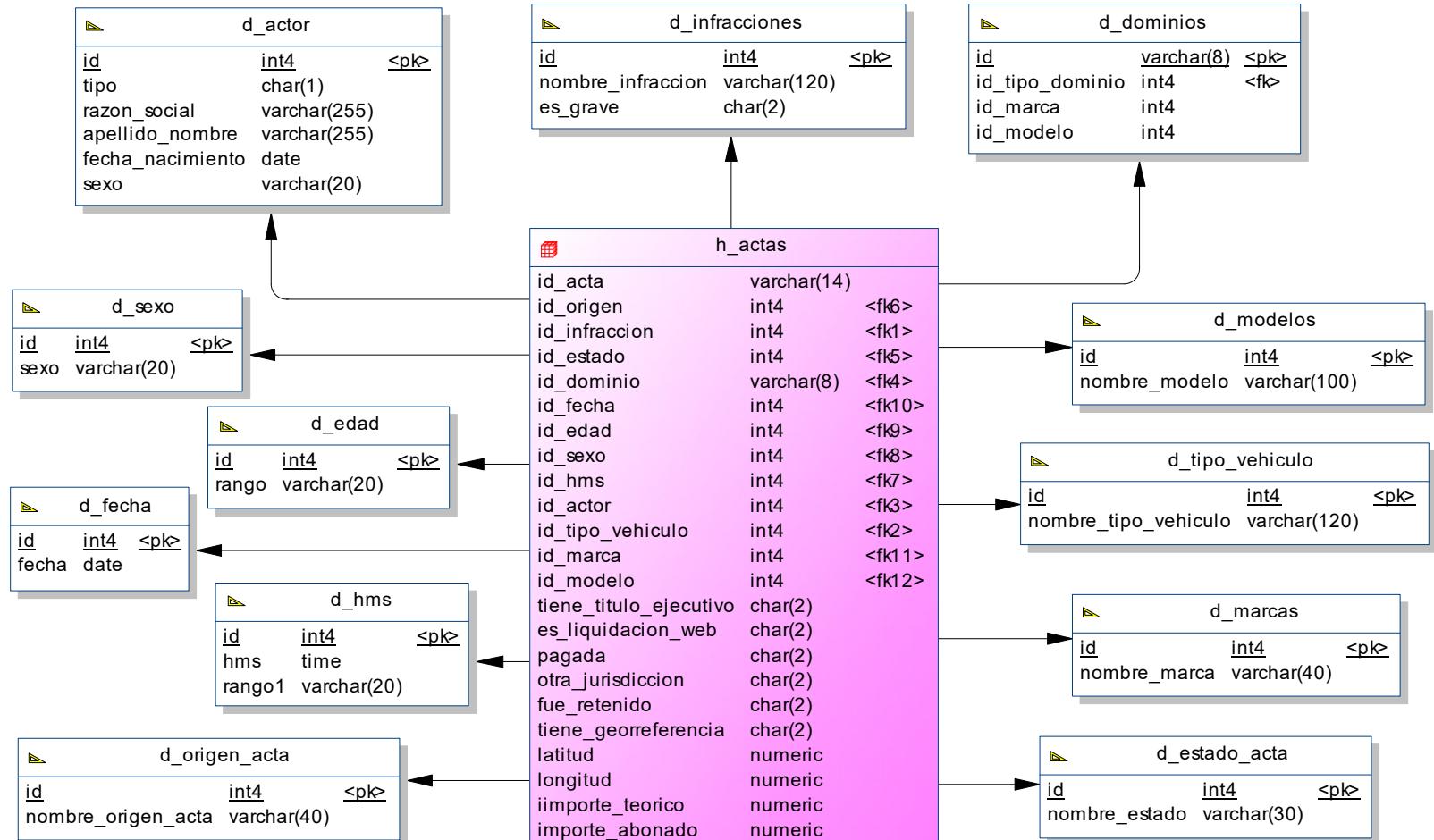
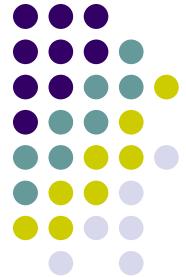
Data Warehouse - Business Intelligence

Modelos dimensionales



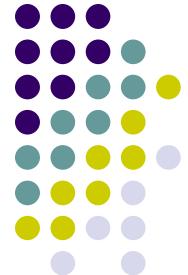
Data Warehouse - Business Intelligence

Modelos dimensionales



Data Warehouse - Business Intelligence

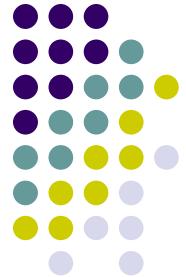
Modelos dimensionales



Tipos de tablas de Hechos

Transaction Fact Tables: Representan eventos que suceden en un determinado espacio-tiempo. Se caracterizan por permitir analizar los datos con el máximo detalle.

Factless Fact Tables/Coverage Tables: Son tablas que no tienen medidas y tiene sentido dado que representan el hecho que el evento suceda. Suelen asociarse medidas que permita facilitar las consultas requeridas.



Data Warehouse - Business Intelligence

Modelos dimensionales

Cubo dimensional

Representa los datos planos en una matriz de N-dimensional.

Se componen de:

Indicadores

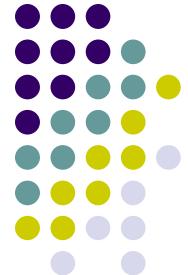
- Valores analizables calculados con operaciones sobre los Hechos.

Atributos

- Criterios de análisis mediante los cuales se analizan los Indicadores. Se obtienen de los atributos de las dimensiones.

Jerarquías

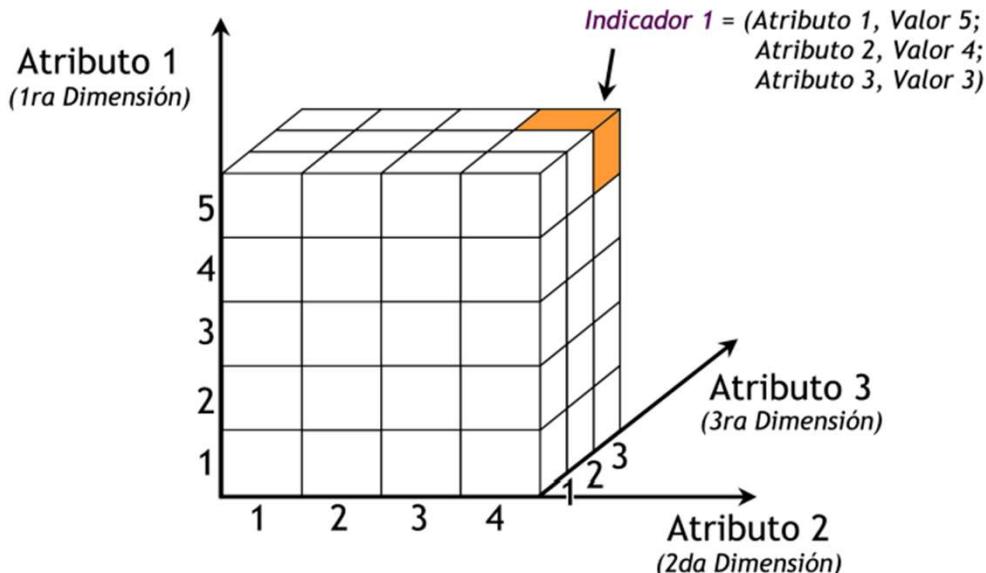
- Relación lógica del tipo padre-hijo entre los Atributos que permiten analizar los datos desde el nivel más general al más detallado y viceversa (navegación en profundidad de los Hechos). Las Jerarquías manejan el nivel de agregación de los Hechos.



Data Warehouse - Business Intelligence

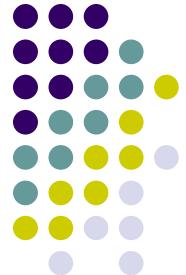
Modelos dimensionales

Cubo dimensional



El **Indicador 1**, representa el cruce del valor **5** de **Atributo 1**, el valor **4** de **Atributo 2** y el valor **3** de **Atributo 3**.

El resultado del análisis está dado por los cruces matriciales de acuerdo con los valores de las **Dimensiones** seleccionadas.

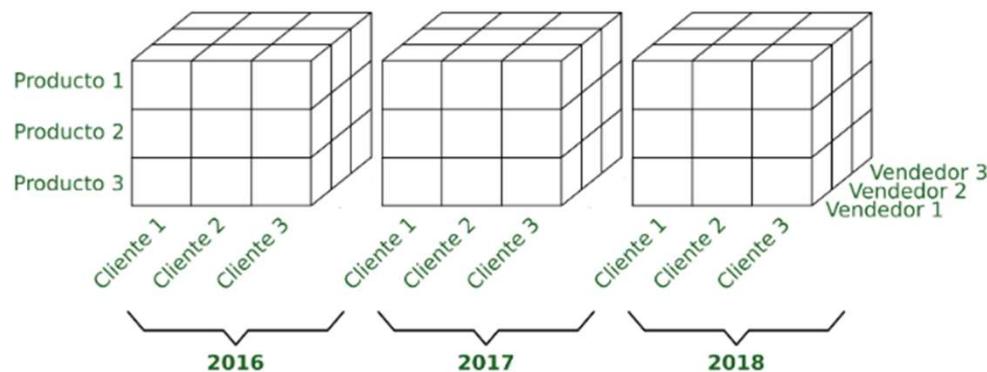


Data Warehouse - Business Intelligence

Modelos dimensionales

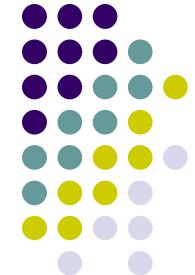
Cubo dimensional

1. El **CUBO ES EL HECHO** a analizar que corresponde a la tabla Ventas.
2. Las dimensiones se asocian a las tablas de cada una y se crean Jerarquías que utilizan alguna columna de la tabla.
3. Se crean los **Indicadores (Cantidad)** que puede estar asociado a un atributo de la tabla del CUBO.



Data Warehouse - Business Intelligence

Modelos dimensionales



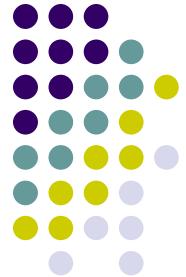
Business Models

Son una representación de los datos desde una perspectiva empresarial, viendo la información y su interrelación como entidades de ALTO NIVEL.

Se componen de:

- **Entidades**
- **Atributos**
- **Relaciones**

Son similares en su comportamiento a los cubos dimensionales.



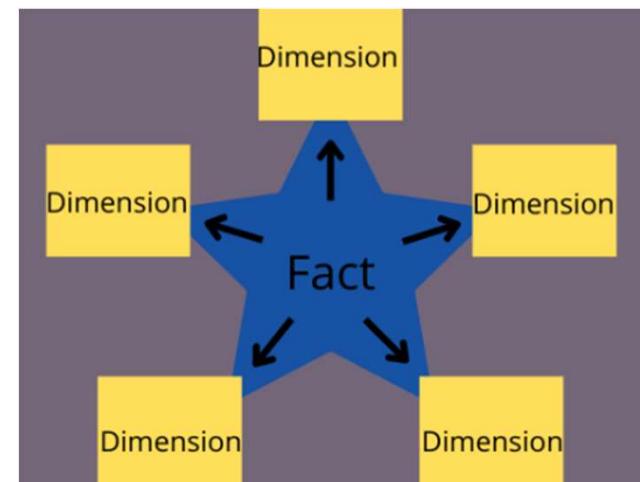
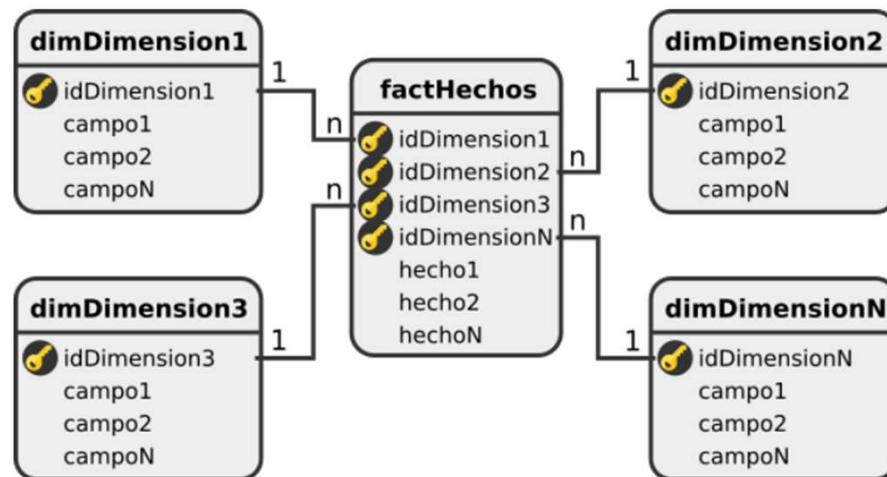
Data Warehouse - Business Intelligence

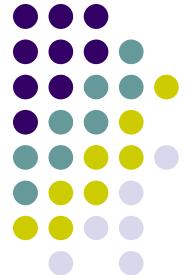
Modelos dimensionales

Modelos del DW

Esquema ESTRELLA

- UNA tabla de Hechos y
- UNA o más tablas de Dimensiones relacionadas a través de sus respectivas claves.





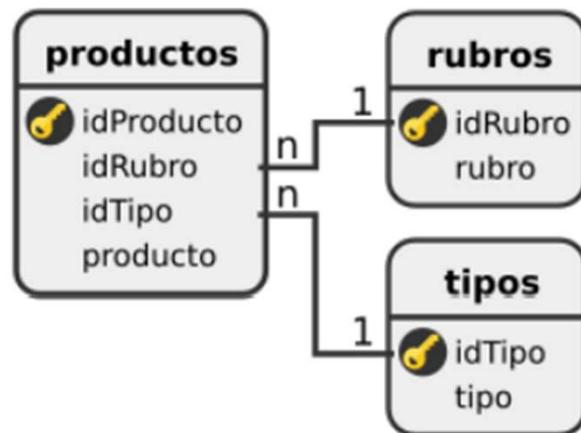
Data Warehouse - Business Intelligence

Modelos dimensionales

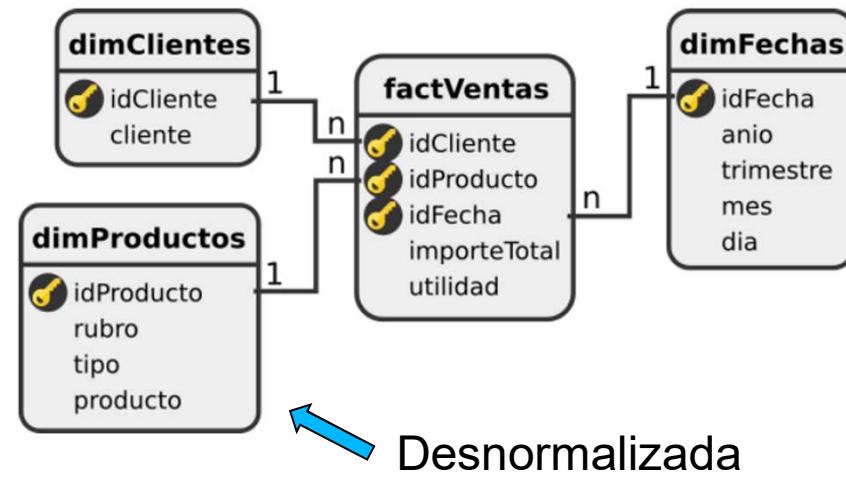
Modelos del DW

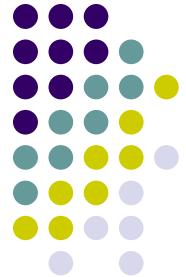
Esquema ESTRELLA

Data Source



Modelo dimensional



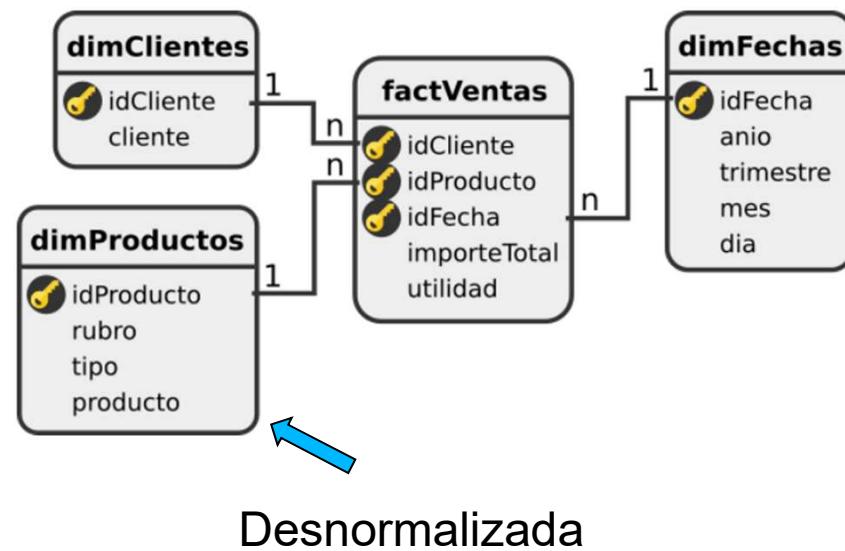


Data Warehouse - Business Intelligence

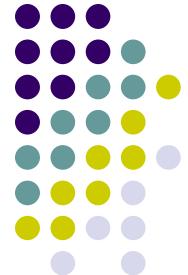
Modelos dimensionales

Modelos del DW

Esquema ESTRELLA



Los modelos multidimensionales requieren la desnormalización criteriosa para evitar uniones (JOIN) entre las tablas para las consultas.

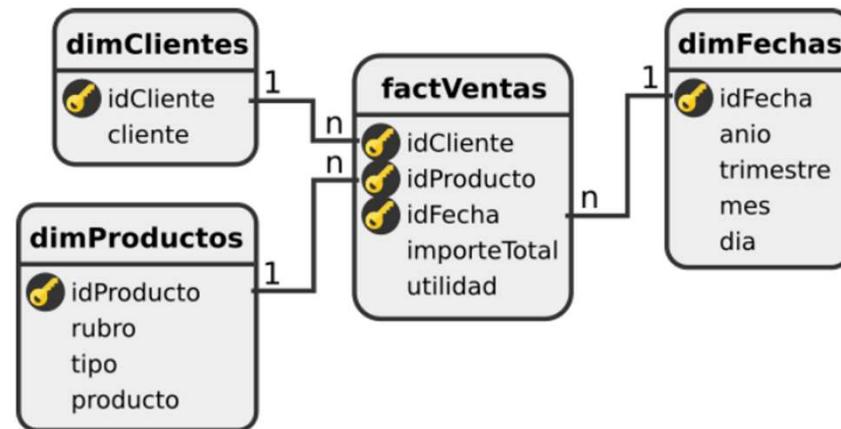


Data Warehouse - Business Intelligence

Modelos dimensionales

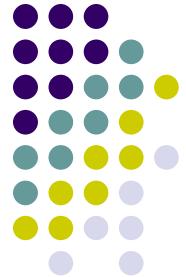
Modelos del DW

Esquema ESTRELLA



Características

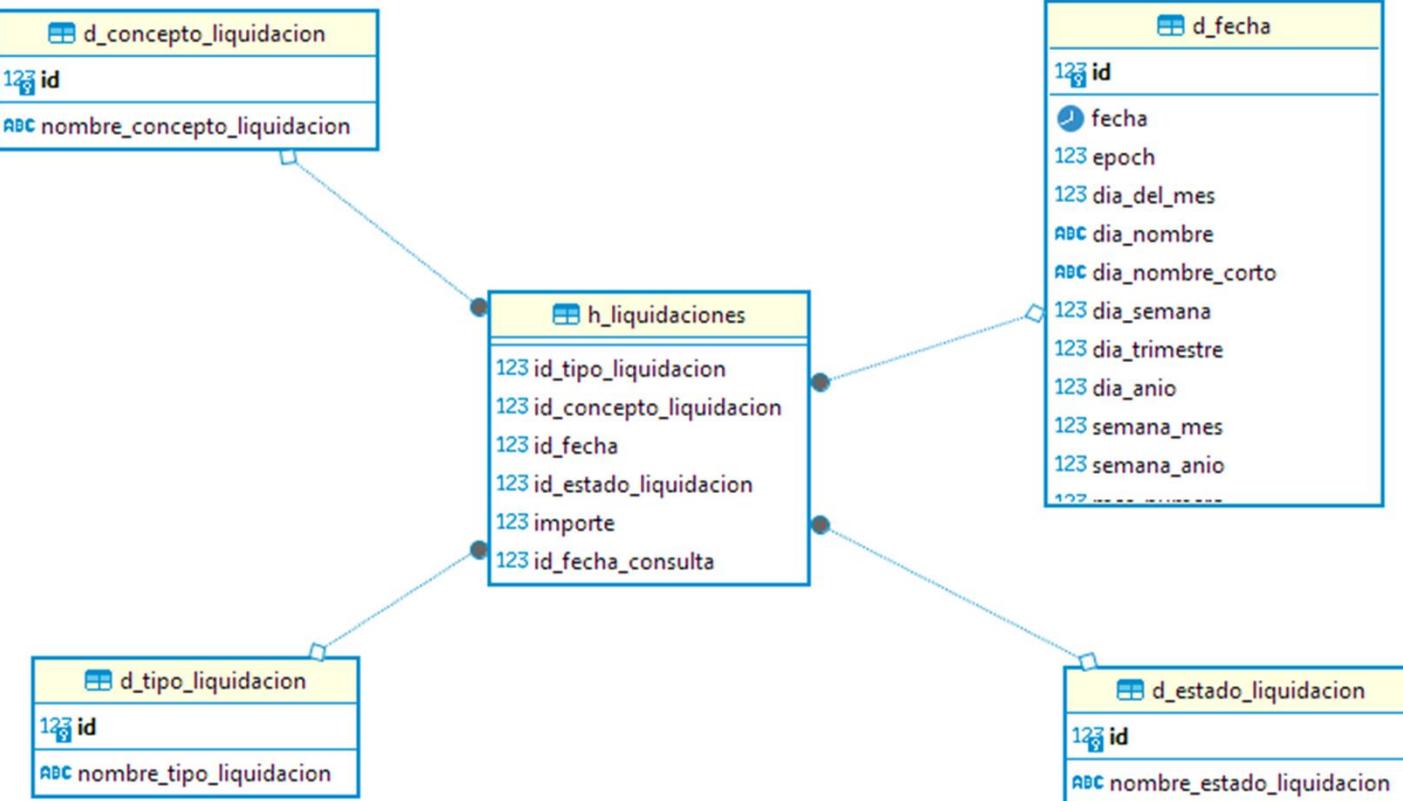
- Es el más simple de interpretar.
- Posee los mejores tiempos de respuesta.
- Es soportado por todos los visores.
- Su diseño es sencillo de mantener y actualizar.
- Existe paralelismo entre su diseño y la forma que los usuarios visualizan y manipulan los datos.



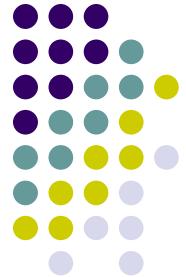
Data Warehouse - Business Intelligence

Modelos dimensionales

Modelos del DW



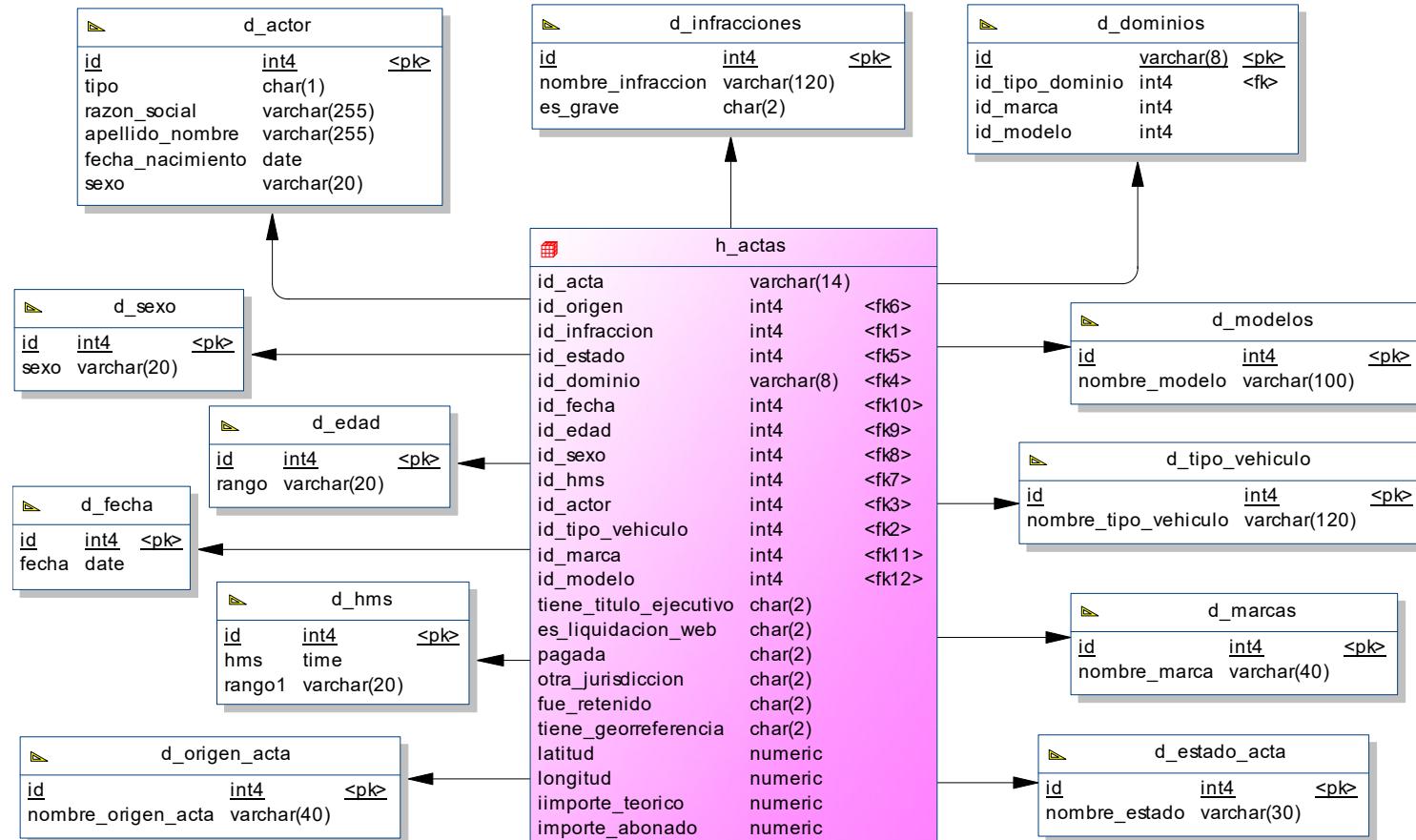
Esquema ESTRELLA

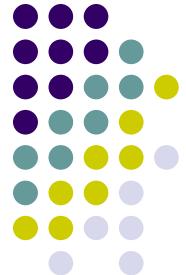


Data Warehouse - Business Intelligence

Modelos dimensionales

Modelos del DW





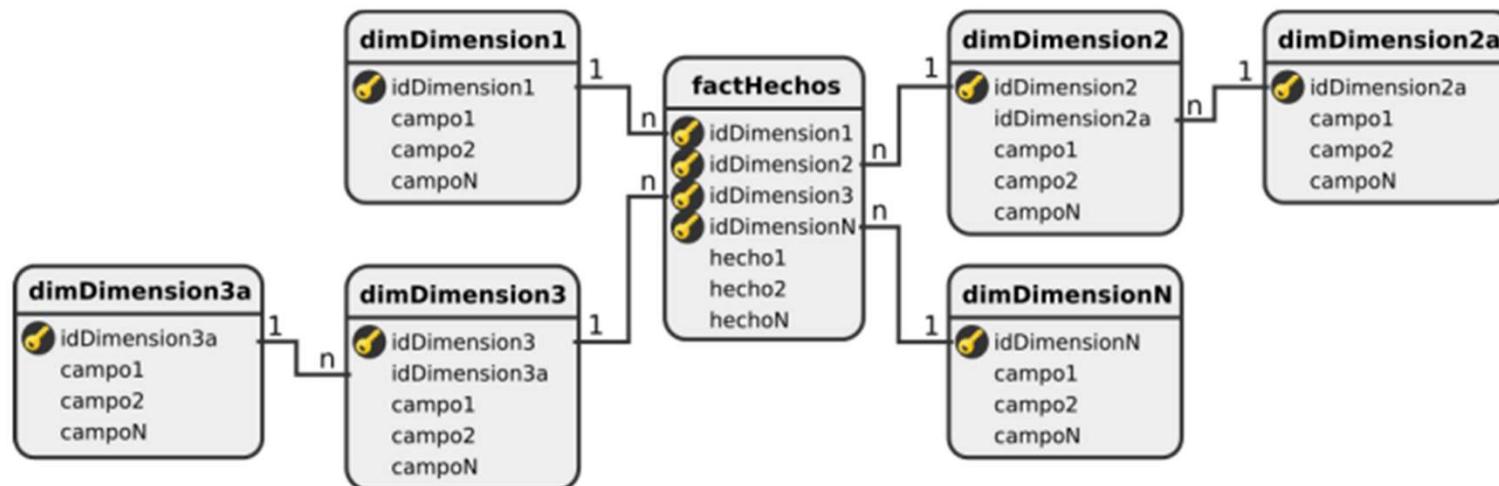
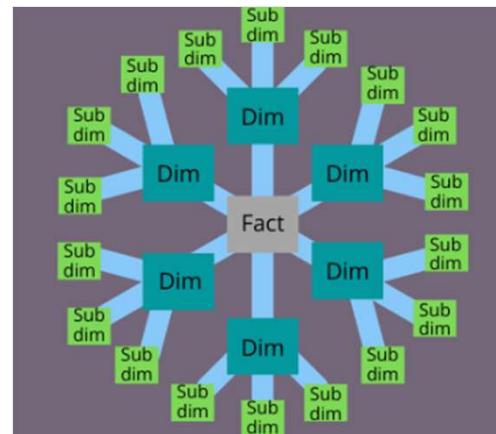
Data Warehouse - Business Intelligence

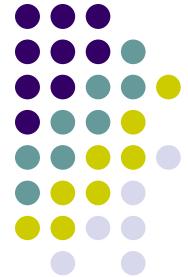
Modelos dimensionales

Modelos del DW

Esquema COPO DE NIEVE

- Extensión del esquema en ESTRELLA
- UNA tabla de Hechos
- **Más de una** tabla de Dimensión organizadas en jerarquías.





Data Warehouse - Business Intelligence

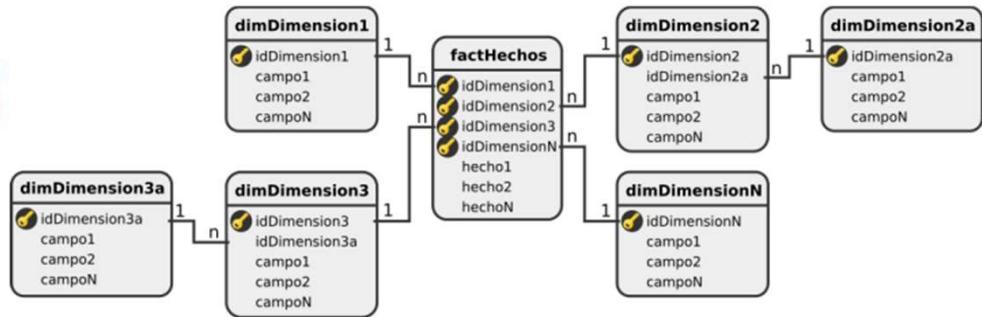
Modelos dimensionales

Modelos del DW

Esquema COPO DE NIEVE

Características

- Posibilita la segregación de los datos de las tablas de Dimensiones.
- Puede implementarse después de que se haya desarrollado un Esquema en Estrella.
- Posee mayor complejidad en su estructura.
- Utiliza menos espacio de almacenamiento.
- Es más eficiente en el caso de tablas de Dimensiones con gran cantidad de registros.
- Su semántica se ajusta a las representaciones de las diferentes Jerarquías de Dimensiones. Se deben planificar correctamente las uniones y el indexado, a fin de NO generar sobrecarga en la resolución de consultas.

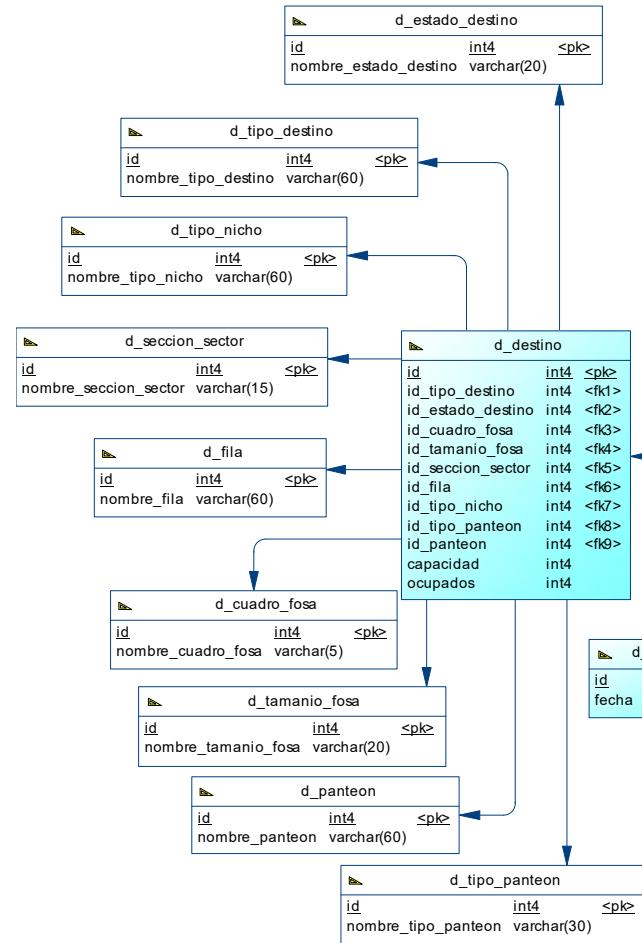




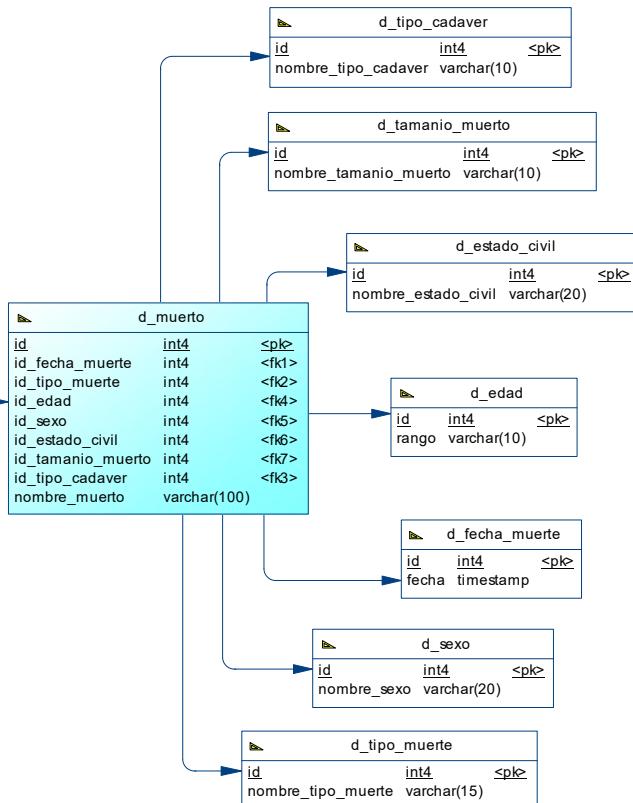
Data Warehouse - Business Intelligence

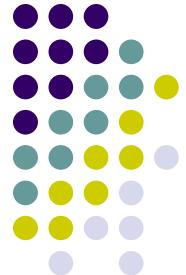
Modelos dimensionales

Modelos del DW



Esquema COPO DE NIEVE





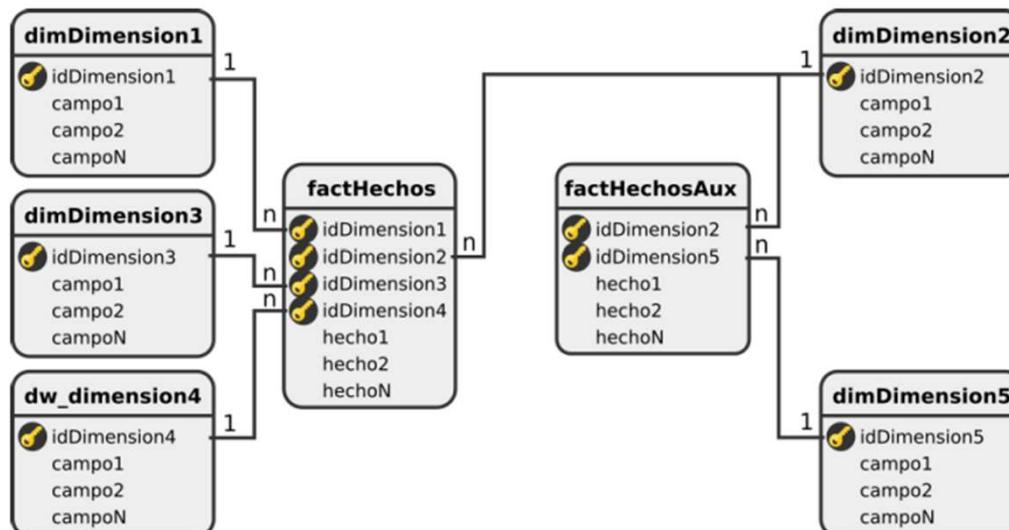
Data Warehouse - Business Intelligence

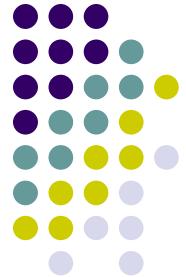
Modelos dimensionales

Modelos del DW

Esquema CONSTELACION

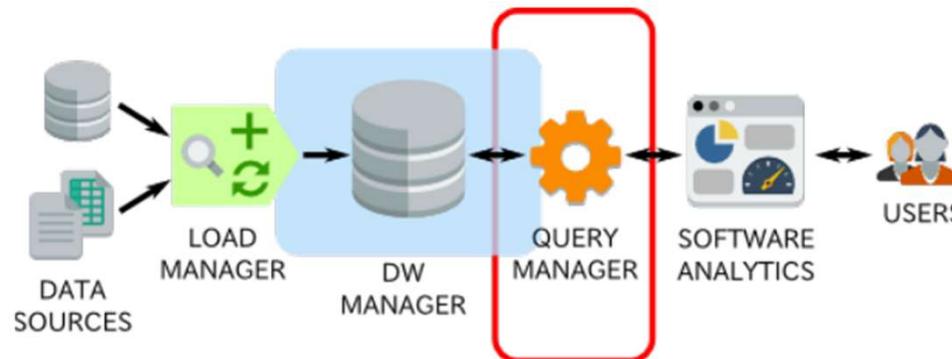
- Compuesto por varios esquemas en ESTRELLA
- **UNA** tabla de Hechos principal y **UNA o más** tablas de Hechos auxiliares (pueden ser agregaciones de la principal).
- **UNA o más** tablas de Dimensiones relacionadas a través de sus respectivas claves.





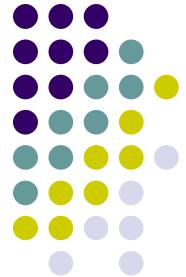
Data Warehouse - Business Intelligence Modelos dimensionales

Arquitectura – Query Manager



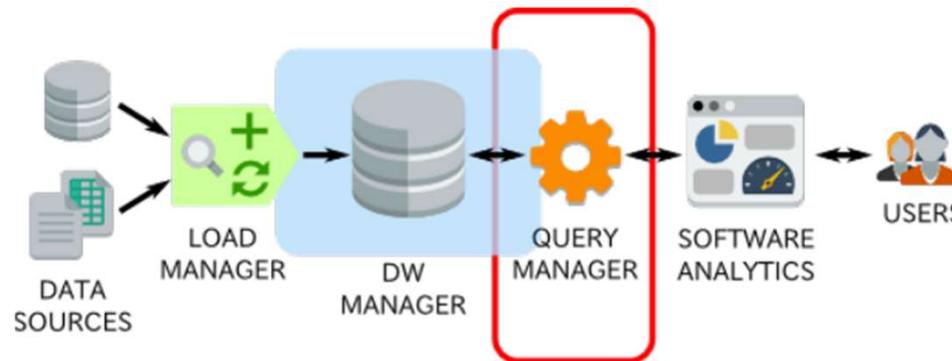
El Query Manager se encarga de resolver:

- consultas relacionales: como JOIN y agregaciones (SUM, COUNT, AVG, etc.)
- consultas propias del análisis de datos (DRILL-UP, DRILL-DOWN, DRILL-ACROSS, etc.)



Data Warehouse - Business Intelligence Modelos dimensionales

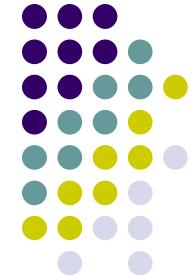
Arquitectura – Query Manager



1. Recibe consultas de los usuarios escritas en un lenguaje de alto nivel;
2. Lee los metadatos que describen los mapeos y reescribe las consultas para que sean ejecutadas en el sistema destino (por lo general SQL).
3. Obtiene los datos y utilizando, nuevamente, las estructuras de metadatos, los transforma a un formato final de alto nivel que será interpretado y renderizado por las herramientas de visualización.

Data Warehouse - Business Intelligence

Modelos dimensionales

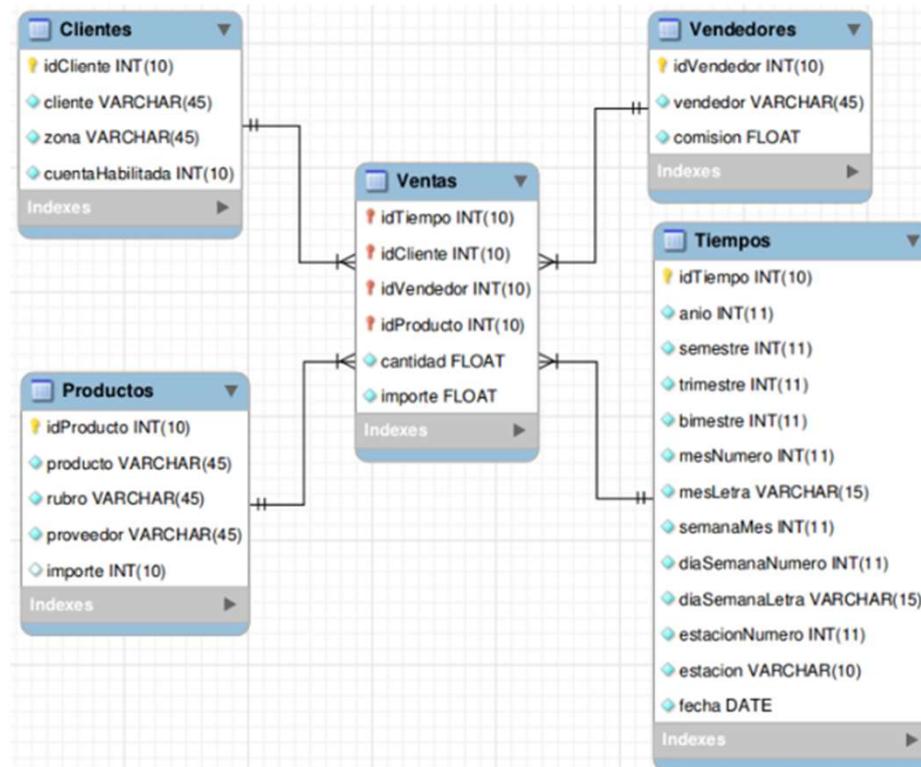


Arquitectura – Query Manager

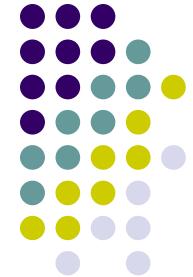


- Las principales operaciones que se realizan sobre modelos multidimensionales son:

- Drill-down, Drill-up
- Drill-across, Roll-across
- Pivot, Page
- Drill-through



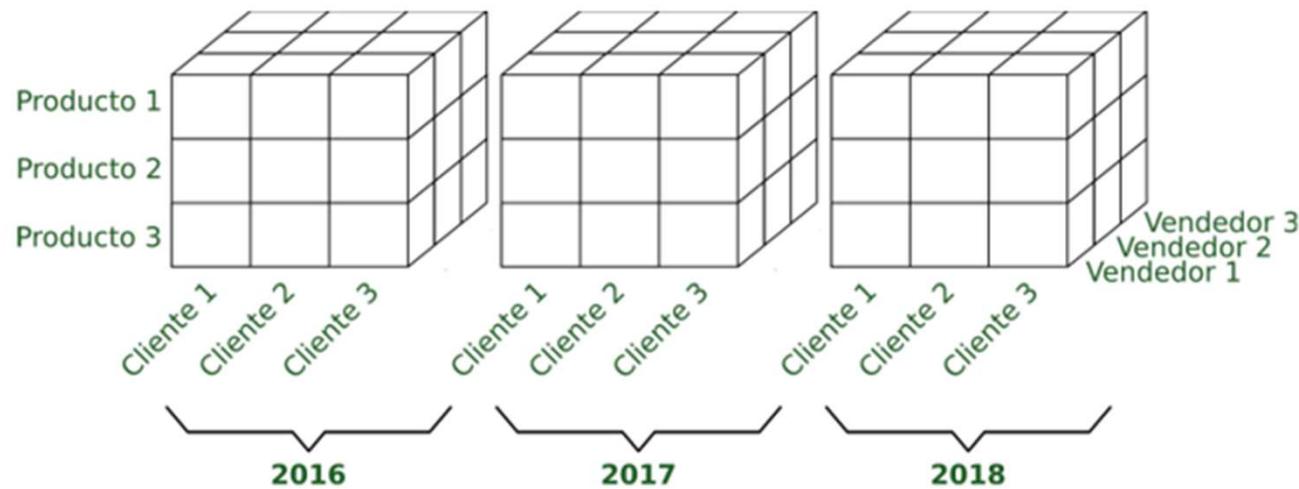
Data Warehouse - Business Intelligence Modelos dimensionales



Arquitectura – Query Manager

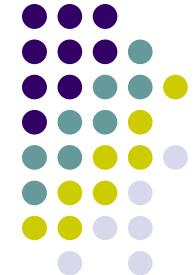


El modelo de ejemplo consta de cuatro tablas de Dimensiones (**Ci**entes, **Pro**ductos, **V**endedores y **T**iempos) y una tabla de Hechos (**V**entas). El cubo dimensional correspondiente será similar a:



Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager

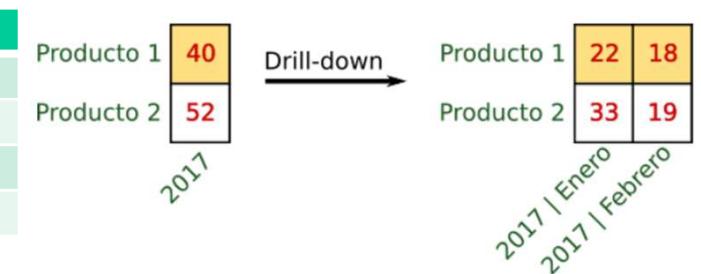


Drill-down: Va de lo general a lo específico para ver los datos con mayor nivel de detalle bajando por una **Jerarquía** definida en el cubo. Si existe la jerarquía de fechas en la dimensión **Tiempos** del tipo año→mes→día las operaciones son:

Dimensión Productos	Dimensión Fechas (año)	Hecho Ventas (cantidad)
Producto 1	2017	40
Producto 2	2017	52

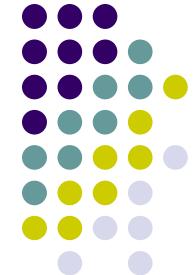
Se aplica la operación Drill-down sobre la jerarquía añadiendo un nivel de detalle:

Dimensión Productos	Dimensión Fechas (año à mes)	Hecho Ventas (cantidad)
Producto 1	2017 – Enero	22
Producto 1	2017 – Febrero	18
Producto 2	2017 – Enero	33
Producto 2	2017 – Febrero	19



Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager

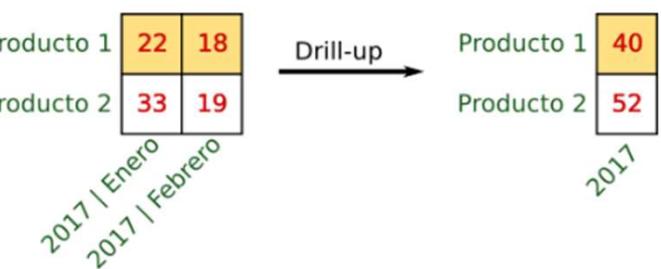


Drill-up: Va de lo específico a lo general para ver los datos con menor grado de detalle subiendo por una **Jerarquía** definida.

Dimensión Productos	Dimensión Fechas (año à mes)	Hecho Ventas (cantidad)
Producto 1	2017 – Enero	22
Producto 1	2017 – Febrero	18
Producto 2	2017 – Enero	33
Producto 2	2017 – Febrero	19

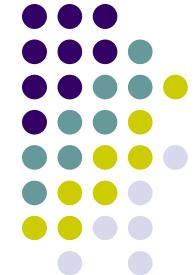
Se aplica la operación Drill-up sobre la jerarquía quitando un nivel de detalle, se obtiene la situación inicial:

Dimensión Productos	Dimensión Fechas (año)	Hecho Ventas (cantidad)
Producto 1	2017	40
Producto 2	2017	52



Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager

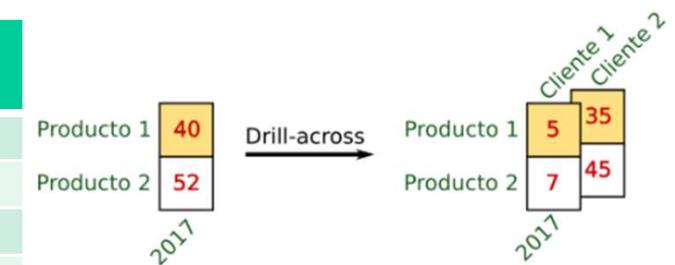


Drill-across: Va de lo general a lo específico para ver los datos con mayor nivel de detalle pero **no se aplica a una Jerarquía** sino que se agrega un atributo como nuevo criterio de análisis.

Dimensión Productos	Dimensión Fechas (año)	Hecho Ventas (cantidad)
Producto 1	2017	40
Producto 2	2017	52

Si se aplica la operación Drill-across al agregar el atributo **Cliente** quedará:

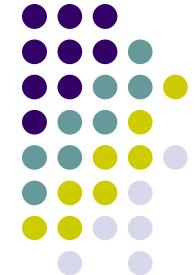
Dimensión Productos	Dimensión Fechas (año)	Dimensión Cliente	Hecho Ventas (cantidad)
Producto 1	2017	Cliente 1	5
Producto 1	2017	Cliente 2	35
Producto 2	2017	Cliente 1	7
Producto 2	2017	Cliente 2	45



Roll-across: es la operación inversa a Drill-across.

Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager



Pivot: permite cambiar el orden de visualización de **Atributos e Indicadores** para analizar la información desde distintos puntos de vista. Se tiene inicialmente el orden **Producto, Año y Cliente**:

Dimensión Productos	Dimensión Fechas (año)	Dimensión Cliente	Hecho Ventas (cantidad)
Producto 1	2017	Cliente 1	5
Producto 1	2017	Cliente 2	35
Producto 2	2017	Cliente 1	7
Producto 2	2017	Cliente 2	45

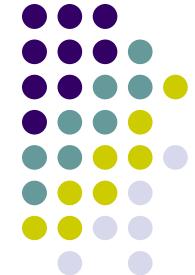
Si se aplica la operación Pivot cambiando el orden:

Dimensión Cliente	Dimensión Fechas (año)	Dimensión Productos	Hecho Ventas (cantidad)
Cliente 1	2017	Producto 1	5
Cliente 1	2017	Producto 2	7
Cliente 2	2017	Producto 1	35
Cliente 2	2017	Producto 2	45

No tiene sentido una operación inversa ya que es otro **Pivot**.

Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager



Page: presenta la información dividida en secciones tomando como criterio los valores de un **Atributo**.

Dimensión Productos	Dimensión Fechas (año à mes)	Hecho Ventas (cantidad)
Producto 1	2017 – Enero	22
Producto 1	2017 – Febrero	18
Producto 2	2017 – Enero	33
Producto 2	2017 – Febrero	19

Si se aplica Page sobre el atributo **Producto** la salida será:

Página 1 – Producto 1

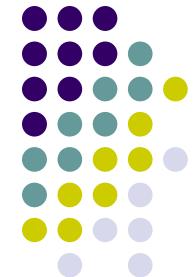
Dimensión Fechas (año à mes)	Hecho Ventas (cantidad)
2017 – Enero	22
2017 – Febrero	18

Página 2 – Producto 2

Dimensión Fechas (año à mes)	Hecho Ventas (cantidad)
2017 – Enero	33
2017 – Febrero	19

Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Query Manager

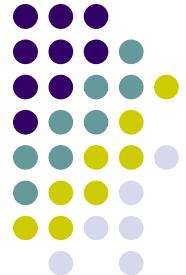


Drill-through: permite visualizar cuáles son los datos relacionados al valor de un Indicador. Los datos se mostrarán en su máximo nivel de detalle. Se tomará como referencia la siguiente tabla:

Dim Productos	Dim Fechas (año)	H Ventas - Indicador (cantidad)
Producto 1	2017	40
Producto 2	2017	52

Si se aplica la operación Drill-through sobre el Indicador de la fila seleccionada, se puede obtener el detalle completo:

Dim Fechas (Jerarquía año → mes → día)	Dim Clientes	Dim Productos	Dim Vendedores	H Ventas – Indicador (cantidad)
2017→Enero→1	Cliente 1	Producto 1	Vendedor 1	3
2017→Enero→5	Cliente 1	Producto 1	Vendedor 1	2
2017→Enero→10	Cliente 2	Producto 1	Vendedor 1	10
2017→Enero→18	Cliente 2	Producto 1	Vendedor 1	6
2017→Enero→21	Cliente 2	Producto 1	Vendedor 1	1
2017→Febrero→4	Cliente 2	Producto 1	Vendedor 1	5
2017→Febrero→12	Cliente 2	Producto 1	Vendedor 1	4
2017→Febrero→16	Cliente 2	Producto 1	Vendedor 1	5
2017→Febrero→23	Cliente 2	Producto 1	Vendedor 1	3
2017→Febrero→24	Cliente 2	Producto 1	Vendedor 1	1



Data Warehouse - Business Intelligence Modelos dimensionales

Arquitectura – Software Analytics



Estas herramientas constituyen el nexo entre el DW y los usuarios y son la parte más visible y/o tangible del proceso de Data Warehousing. Para la obtención de datos del DW, se utiliza principalmente:

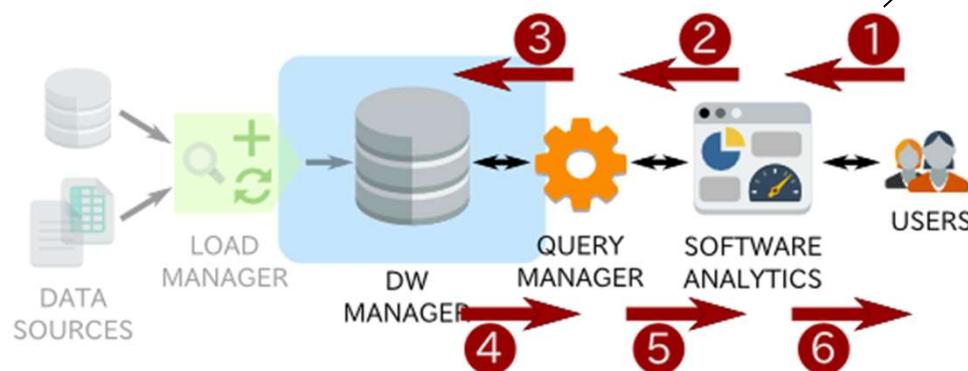
- La metadata de las estructuras de datos que han sido creadas previamente (cubos multidimensionales, business models) y
- Conexiones a bases de datos

Data Warehouse - Business Intelligence Modelos dimensionales

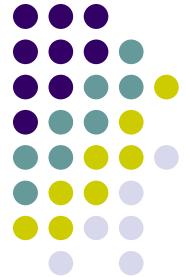


Arquitectura – Software Analytics

Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:



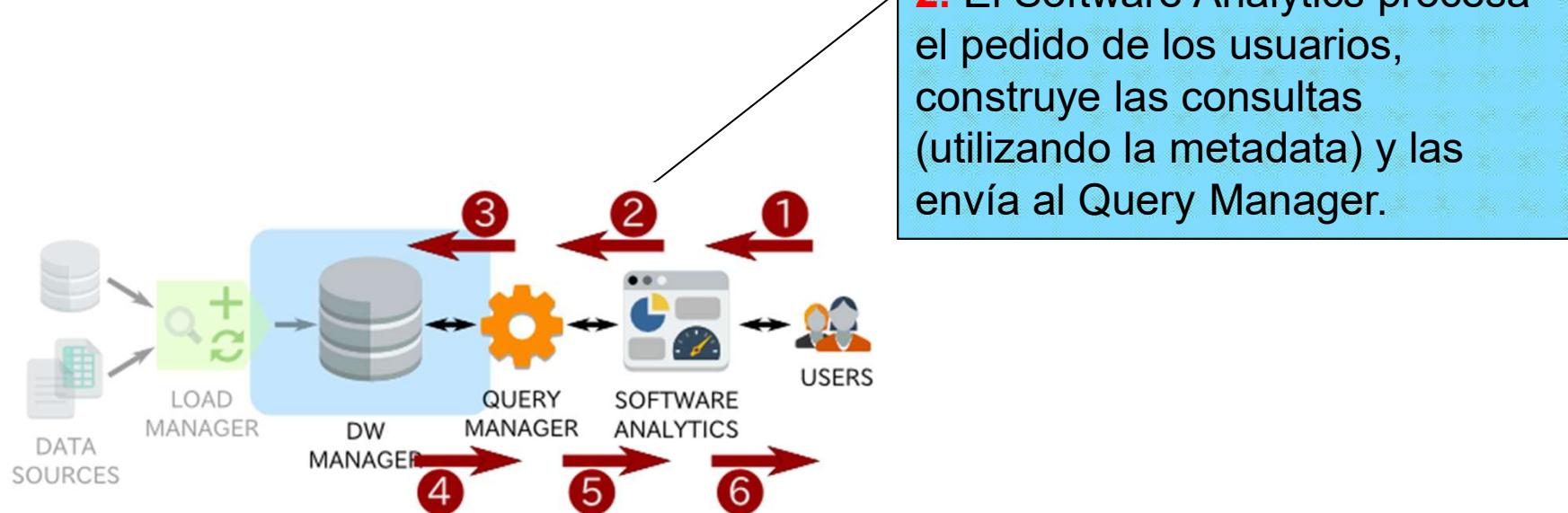
1. Los usuarios seleccionan o establecen qué datos desean obtener del DW, mediante la GUI (interfaz gráfica) del Software Analytics

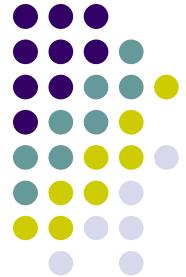


Data Warehouse - Business Intelligence Modelos dimensionales

Arquitectura – Software Analytics

Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:

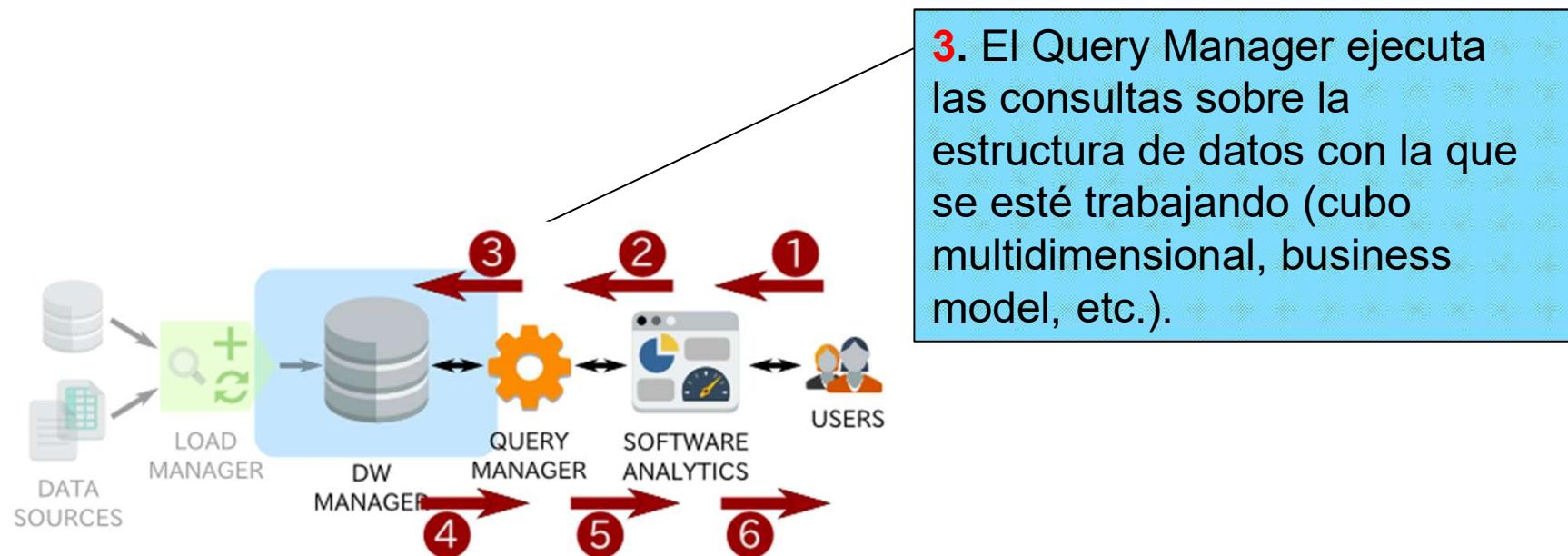




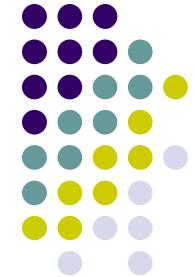
Data Warehouse - Business Intelligence Modelos dimensionales

Arquitectura – Software Analytics

Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:



Data Warehouse - Business Intelligence Modelos dimensionales

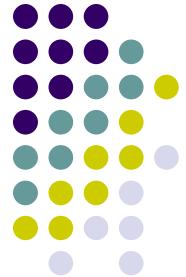


Arquitectura – Software Analytics

Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:



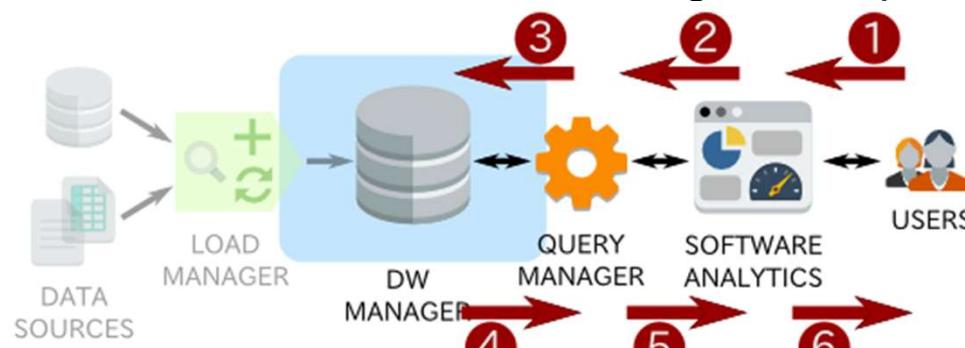
4. El Query Manager obtiene los resultados de las consultas.



Data Warehouse - Business Intelligence Modelos dimensionales

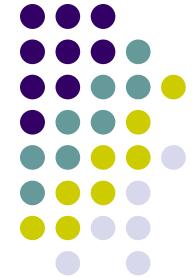
Arquitectura – Software Analytics

Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:



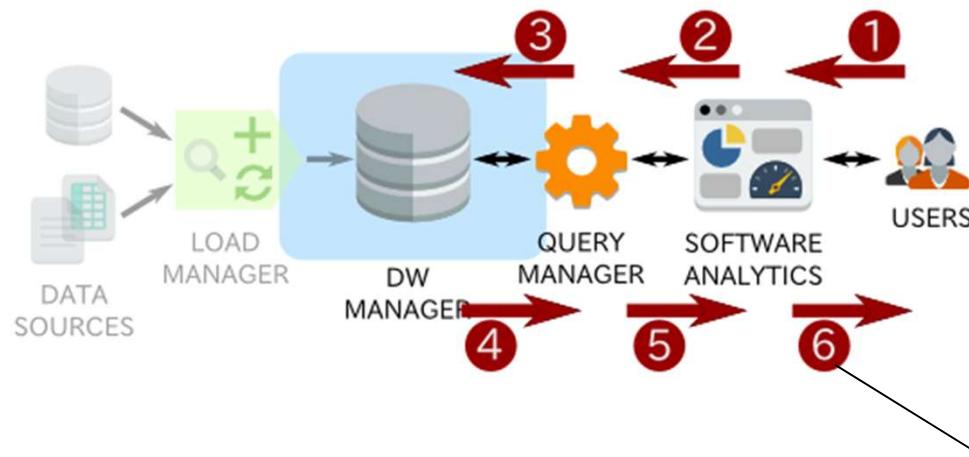
5. El Query Manager envía los datos al Software Analytics.

Data Warehouse - Business Intelligence Modelos dimensionales



Arquitectura – Software Analytics

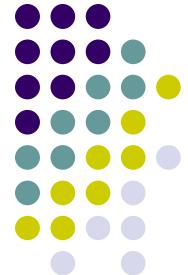
Cada vez que el usuario interactúa con el Software Analytics para explorar los datos del DW se llevan a cabo los siguientes pasos generales:



6. El Software Analytics presenta a los usuarios los datos requeridos.

Data Warehouse - Business Intelligence

Modelos dimensionales



Arquitectura – Software Analytics



Herramientas de consulta y análisis

Dashboards (reporting en general: el análisis parte de la hipótesis del usuario)

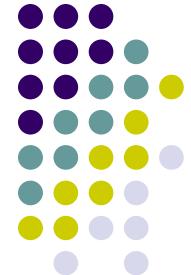
Colección de componentes gráficos de análisis como reportes, tablas, gráficos, consultas y análisis interactivos, etc. que hacen referencia a un tema en particular y que están relacionados entre sí. Presentan la información altamente resumida.

- Se componen de consultas, reportes, análisis interactivos, charts (barras, líneas, torta, áreas, etc.), semáforos, Indicadores causa-efecto, etc.
- Permiten evaluar situaciones con un solo golpe de vista.
- Poseen un formato de diseño visual muy llamativo.

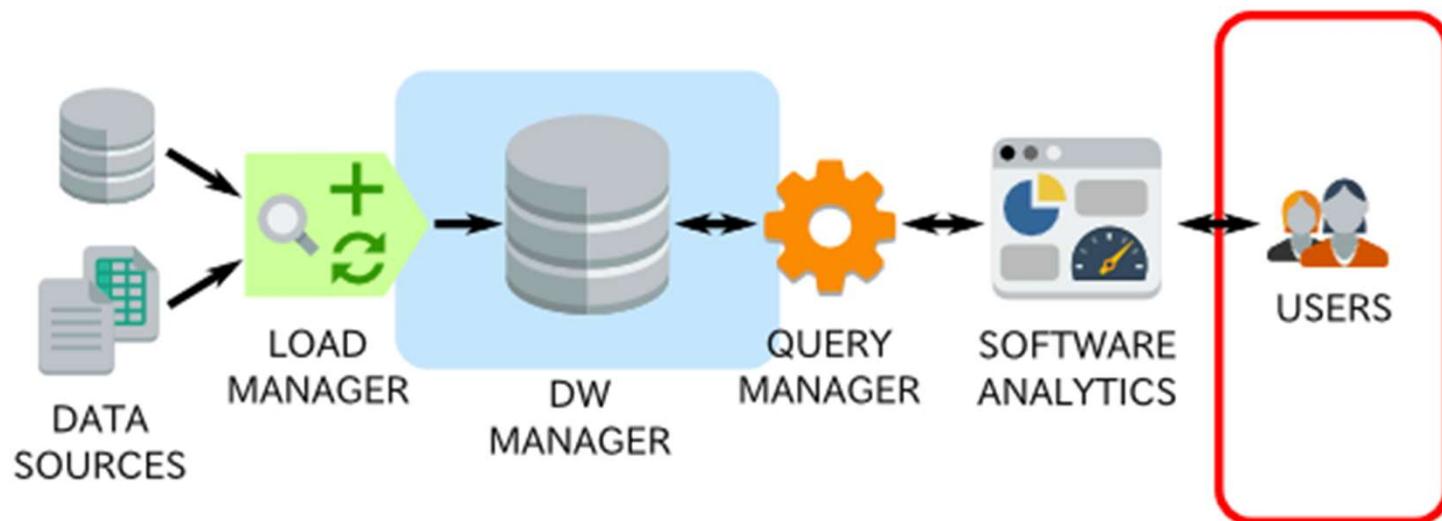
Data Mining (permite generar las hipótesis del usuario)

Se emplea para analizar factores de influencia, predecir comportamientos futuros, extraer conocimientos ocultos, agrupar o segmentar elementos similares etc. Permite inferir comportamientos y realizar predicciones. Es una técnica para descubrir patrones en grandes volúmenes de datos basados en IA y métodos matemáticos como **Redes neuronales, árboles de decisión** entre otros.

Data Warehousing Business Intelligence



Arquitectura – Users



Data Warehouse - Business Intelligence

Modelos dimensionales



Performance

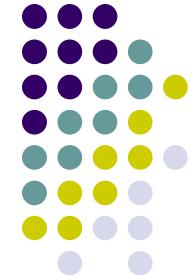
- Selección y configuración del SGBD que dará soporte al DW
- Elección de estrategias para mantener las estructuras de datos.

La Base de Datos

- Seleccionar los tipos de datos, por ejemplo, para valores enteros pequeños conviene utilizar **tinyint** o **smallint** en lugar de **int**. ya que formarán parte de las tablas de Hechos (que son las de mayor volumen) además de que toda clave primaria tiene asociado un índice que la implementa.
- Utilizar Claves Subrogadas. Utilizar técnicas de indexación.
- Utilizar técnicas de particionamiento. Crear diferentes niveles de agregación.
- Utilizar distribución de datos.
- Utilizar técnicas de multiprocesamiento distribuido, con el objetivo de agilizar la obtención de resultados, a través de la realización de procesos en forma concurrente.

Data Warehouse - Business Intelligence

Modelos dimensionales



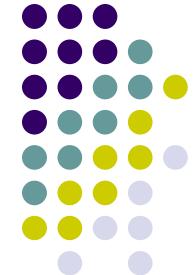
Performance

Elección de columnas

- Descartar columnas cuyos valores tengan muy poca variabilidad.
- Descartar columnas que tengan valores diferentes para cada objeto, por ejemplo, el número de documento, cuando se analizan personas.
- En los casos en exista una constelación, intentar embeberla en la dimensión más cercana a los hechos implementando una jerarquía.

Data Warehouse - Business Intelligence

Modelos dimensionales



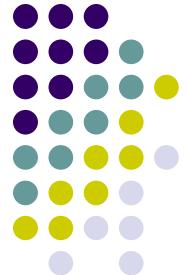
Performance

Tratamiento de las columnas

- **Factorizar**: se utiliza para descomponer un valor en dos o más componentes
- **Estandarizar**: se utiliza para ajustar valores a un tipo de formato o norma preestablecida.
- **Codificar**: se utiliza para representar valores a través de las reglas de un código preestablecido. Por ejemplo, en la columna **estado** se pueden codificar sus valores, **0** y **1**, para transformarlos en **Apagado** y **Encendido**
- **Discretizar**: se utiliza para convertir un conjunto continuo de valores en uno discreto. Por ejemplo, en el campo **intensidad** se pueden codificar los valores menores a 100 como **Baja**; los valores mayores a 100 y menores a 500 como **Media**; y los valores mayores a 500 como **Alta**.

Data Warehouse - Business Intelligence

Modelos dimensionales



Performance

Tratamiento de las relaciones

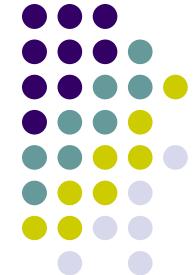
Evitar mantener en el DW tablas de Dimensiones con relaciones muchos a muchos entre ellas. Las soluciones podrían ser:

- Crear una tabla de relación (que también sería una Dimensión) que tenga las claves de ambas tablas quedando de esta manera 1,n con cada una de ellas para luego vincular la tabla de hechos con esta tabla de relación.
- Agregar las dos claves primarias de las tablas de Dimensiones en la tabla de Hechos.

La única desventaja es en cuanto a los procesos ETL, ya que se aumenta su complejidad y tiempo de proceso.

Data Warehouse - Business Intelligence

Modelos dimensionales

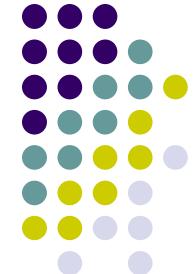


Performance

- **Claves naturales o de negocio**
- **Claves técnicas**
 - Numérico secuencial,
 - NO tienen relación directa con ningún dato ni tienen significado
 - Ocupan menos espacio y son más performantes que las claves naturales, y más aún si éstas son tipo texto.
 - Construcción y mantenimiento de índices más sencillo.
 - El Data Warehouse NO dependerá de la codificación de los Data Sources.
 - Si se modifica el valor de una clave en el Data Source, el DW lo tomará como un nuevo elemento, almacenando diferentes versiones del mismo dato.
 - Permiten la correcta aplicación de técnicas SCD.

Data Warehouse - Business Intelligence

Modelos dimensionales



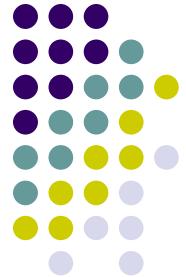
SCD (slowly changing dimensions)

Cuando ocurren cambios en las dimensiones se puede:

- Registrar el historial de cambios.
- Reemplazar los valores que sean necesarios.

Ralph Kimball planteó tres estrategias a seguir cuando se tratan las SCD: tipo 1, tipo 2 y tipo 3; pero a través de los años se profundizaron las definiciones iniciales e incluyó tipo 4 y tipo 6.

- **SCD Tipo 0:** no tener en cuenta ningún cambio.
- **SCD Tipo 1:** Sobrescribir.
- **SCD Tipo 2:** Añadir fila.
- **SCD Tipo 3:** Añadir columna.
- **SCD Tipo 4:** Historial separado.
- **SCD Tipo 6:** Híbrido.



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 1 - Sobrescribir

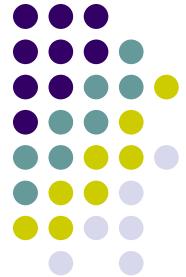
idProducto	Rubro	Tipo	Producto
1	Rubro 1	Tipo 1	Producto 1

Ahora, se supondrá que este producto ha cambiado de **rubro**, y ahora ha pasado a ser **Rubro 2**, entonces se obtendrá lo siguiente:

idProducto	Rubro	Tipo	Producto
1	Rubro 2	Tipo 1	Producto 1

Se usa donde la información histórica no sea importante de mantener (como errores de ortografía).

El ejemplo muestra ahora que todos los movimientos realizados de *Producto 1*, que antes pertenecían al *Rubro 1*, ahora pasarán a ser del **Rubro 2**, lo cual creará inconsistencias en el DW.



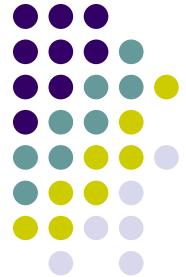
Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 2 – Añadir fila

Requiere el agregado previo de las columnas:

- **fechalinicio**: fecha desde que entró en vigencia el registro actual. Por defecto suele utilizarse una fecha muy antigua, ejemplo: **01/01/1000**.
- **fechaFin**: fecha en la cual el registro actual dejó de estar en vigencia. Por defecto suele utilizarse una fecha muy futurista, ejemplo: **01/01/9999**.
- **version**: número secuencial que se incrementa cada nuevo cambio. Por defecto suele comenzar en **1**.
- **versionActual**: especifica si el campo actual es el vigente. Este valor puede ser en caso de ser verdadero: **true** o **1**; y en caso de ser falso: **false** o **0**.



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 2 – Añadir fila

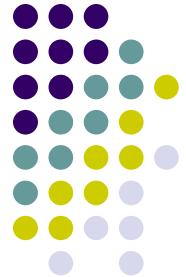
idProducto	CN_Producto	Rubro	Tipo	Producto
1	155	Rubro1	Tipo1	Producto1

Agregado de columnas:

idProducto	CN_Producto	Rubro	Tipo	Producto	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	01/01/9999	1	true

Este producto ha cambiado de Rubro, y ahora ha pasado a ser **Rubro2**, entonces se obtendrá lo siguiente:

idProducto	CN_Producto	Rubro	Tipo	Producto	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Tipo1	Producto1	09/05/2023	01/01/9999	2	true



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 2 – Añadir fila

Proceso

IdProducto	CN_Producto	Rubro	Tipo	Producto	FechaInício	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Tipo1	Producto1	09/05/2023	01/01/9999	2	true

- Se añade una nueva fila con su correspondiente clave subrogada (**idProducto = 2**) manteniéndose la clave de negocios original (**CN_Producto**)
- Se registra la modificación (**Rubro**).
- Se actualizan los valores de **Fechalnicio** y **FechaFin**, tanto de la fila nueva, como la antigua (la que presentó el cambio).
- Se incrementa en uno el valor del campo **Version** que posee la fila antigua.
- Se actualizan los valores de **VersionActual**, tanto de la fila nueva, como la antigua; dejando a la fila nueva como el registro vigente (**true**).



Data Warehouse - Business Intelligence

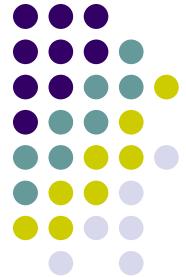
Modelos dimensionales

SCD Tipo 2 – Añadir fila

Proceso

idProducto	CN_Producto	Rubro	Tipo	Producto	FechaInício	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Tipo1	Producto1	09/05/2023	01/01/9999	2	true

- Permite mantener todos los cambios (ilimitado).
- La referencia de los hechos se hace al idProducto pero que corresponda a la fecha de vigencia y la misma CN_Producto.
- Distintos Hechos pueden corresponder al mismo producto pero en períodos distintos (mismo CN_Producto y distinto idProducto).



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 3 – Añadir columna

Requiere el agregado previo una columna adicional por cada columna cuyos valores se desean mantener en el historial:

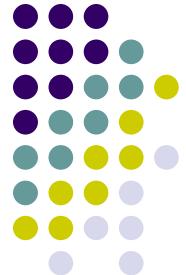
idProducto	CN_Producto	Rubro	Tipo	Producto
1	155	Rubro1	Tipo1	Producto1

Para mantener el histórico sobre los datos de la columna **Rubro**, se agregará la columna **RubroAnterior** (mismo tipo de datos). Si cambia de **Rubro1** a **Rubro2** se obtendrá:

idProducto	CN_Producto	Rubro	RubroAnterior	Tipo	Producto
1	155	Rubro2	Rubro1	Tipo1	Producto1

- En la columna **RubroAnterior** se coloca el valor antiguo (**Rubro1**).
- En la columna **Rubro** se coloca el valor vigente (**Rubro2**)

Mantiene UN SOLO CAMBIO: *El último.*



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 4 – Historial separado

- Se utiliza en combinación con alguna de las anteriores.
- Se crea una tabla adicional con los detalles de los cambios históricos.

La Dimensión original es:

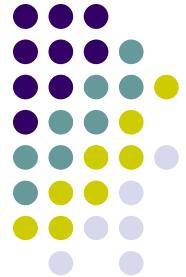
idProducto	CN_Producto	Rubro	Tipo	Producto
1	155	Rubro1	Tipo1	Producto1

Y suponiendo si se trabaja con SCD Tipo 2 y cambia **Rubro** de **Rubro1** a **Rubro2** y **Producto** de **Producto1** a **Producto50** para **CN_producto = 155**, la tabla de dimensión será ahora:

idProducto	CN_Producto	Rubro	Tipo	Producto
2	155	Rubro2	Tipo1	Producto50

Y el historial será:

idProducto	CN_Producto	Rubro	Tipo	Producto	Fechainicio	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Tipo1	Producto50	09/05/2023	01/01/9999	2	true



Data Warehouse - Business Intelligence

Modelos dimensionales

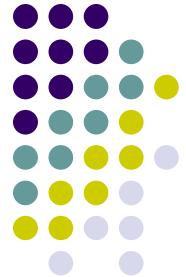
SCD Tipo 4 – Historial separado

idProducto	CN_Producto	Rubro	Tipo	Producto	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Tipo1	Producto50	09/05/2023	01/01/9999	2	true

Se asemeja a la manera que se implementan las tablas de auditoría de las bases de datos (registro anterior, registro actual).

Si se implementa SCD Tipo 3 y manteniendo el valor anterior de las columnas **Rubro** y **Producto** cuando se modifican se tendrá mayor control sobre los cambios.

idProducto	CN_Producto	Rubro	Rubro_Anterior	Tipo	Producto	Producto_Anterior	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Rubro1	Tipo1	Producto1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Rubro1	Tipo1	Producto50	Producto1	09/05/2023	01/01/9999	2	true



Data Warehouse - Business Intelligence

Modelos dimensionales

SCD Tipo 6 – Híbrido

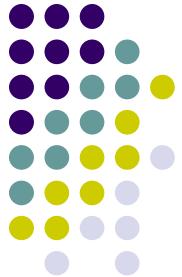
SCD Tipo **1** + SCD Tipo **2** + SCD Tipo **3** = **SCD Tipo 6**

idProducto	CN_Producto	Rubro	Rubro_Ant	Tipo	Producto	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Rubro1	Tipo1	Producto1	01/01/1000	01/01/9999	1	true

El **Rubro** y **Rubro_Ant** son iguales. El atributo **VersionActual** indica que este es el registro actual o más reciente para este producto. Cuando el *Producto1* cambia de *Rubro1* a *Rubro2* se agrega un nuevo registro por SCD Tipo 2 quedando:

idProducto	CN_Producto	Rubro	Rubro_Ant	Tipo	Producto	FechaInicio	FechaFin	Version	VersionActual
1	155	Rubro1	Rubro1	Tipo1	Producto1	01/01/1000	09/05/2023	1	false
2	155	Rubro2	Rubro1	Tipo1	Producto1	09/05/2023	01/01/9999	2	true

Se sobrescribe la información de **VersionActual** como en el SCD Tipo 1. Se crea un nuevo registro para rastrear los cambios como en el SCD Tipo 2, y se almacena el historial en la columna **Rubro_Ant** que corresponde al SCD Tipo3.



Data Warehouse - Business Intelligence

Modelos dimensionales

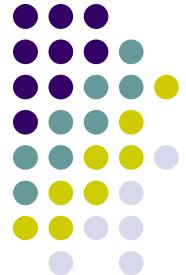
Dimensiones y Hechos especiales

1. Dimensiones realizadoras de roles
2. Dimensiones no deseadas

Dimensión tipoDocumento		Dimensión sexo	
idTipodoc	nombreTipoDoc	idSexo	nombreSexo
1	DNI	1	MASCULINO
2	CEDULA	2	FEMENINO
3	PASAPORTE		

Dimensión tipoVarios		
idtipoVarios	tipoDocu	sexo
1	DNI	MASCULINO
2	DNI	FEMENINO
3	CEDULA	MASCULINO
4	CEDULA	FEMENINO
5	PASAPORTE	MASCULINO
6	PASAPORTE	FEMENINO

3. Dimensiones degeneradas
4. Tablas de hechos sin hechos



Data Warehouse - Business Intelligence

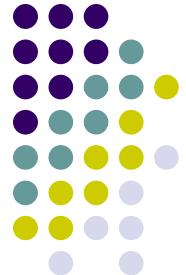
EJEMPLO PRÁCTICO

La empresa en estudio nació como un almacén de barrio, y a través de los años se convirtió en una importante cadena de ventas de productos a nivel nacional. Este crecimiento se vio acompañado con la adquisición de software para el manejo de diversos procesos de negocio, como ser: Ventas, Stock, RRHH.

Problemática

Al día de la fecha, están buscando mejorar su proceso de ventas, motivo por el cual requieren aumentar la cantidad de reportes del área y la frecuencia con la cual se generan. Hoy en día, dichos reportes se manejan en archivos Excel, que implican un excesivo costo operativo en la construcción y alta probabilidad de errores. A esto se le suma el problema de la baja capacidad de respuesta ante cambios solicitados por los usuarios finales.

Los usuarios finales son de nivel estratégico, con tendencia a realizar análisis macro de diversas aristas de la información, realizando comparaciones para evaluar las evoluciones de sus métricas. Si bien en algunas ocasiones requieren el detalle, en la mayoría de los casos parten de lo general, es decir, de información resumida y en formato gráfico.



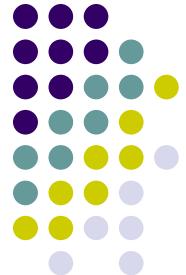
Data Warehouse - Business Intelligence

EJEMPLO PRÁCTICO

Hoy la empresa cuenta con una serie de reportes que utilizan diariamente los directivos para la toma de decisiones y que se obtienen de los datos de sus sistema OLTP. Estos reportes son:

1. Detalle de ventas de cada pedido realizado. Se genera un reporte por mes para cada sucursal:

Pedido	Nivel educativo	Nivel de renta	Método de pago	Estado civil	Edad de cliente	Artículo	Ingresos	Coste	Descuento
47836	Otro	40.000 - 50.000	Amex	Soltero/a	67	Producto 210	\$13	\$12	\$ 0
47836	Otro	40.000 - 50.000	Amex	Soltero/a	67	Producto 298	\$10	\$9	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 8	\$8	\$6	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 74	\$11	\$8	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 110	\$180	\$146	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 116	\$450	\$350	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 220	\$12	\$11	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 227	\$13	\$12	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 283	\$9	\$8	\$ 0
47844	Enseñanza secundaria	40.000 - 50.000	Visa	Casado/a	67	Producto 332	\$16	\$16	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 57	\$10	\$8	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 69	\$34	\$26	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 175	\$500	\$403	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 319	\$15	\$15	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 332	\$16	\$16	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 341	\$16	\$15	\$ 0
47848	Universidad	60.000 - 70.000	Cheque	Casado/a	36	Producto 346	\$13	\$10	\$ 0
47851	Enseñanza secundaria	40.000 - 50.000	Visa	Soltero/a	42	Producto 35	\$8	\$6	\$ 0
47851	Enseñanza secundaria	40.000 - 50.000	Visa	Soltero/a	42	Producto 162	\$230	\$188	\$ 0
47851	Enseñanza secundaria	40.000 - 50.000	Visa	Soltero/a	42	Producto 278	\$8	\$7	\$ 0
47851	Enseñanza secundaria	40.000 - 50.000	Visa	Soltero/a	42	Producto 329	\$14	\$13	\$ 0
47856	Universidad	40.000 - 50.000	Amex	Casado/a	45	Producto 219	\$15	\$13	\$ 0



Data Warehouse - Business Intelligence

EJEMPLO PRÁCTICO

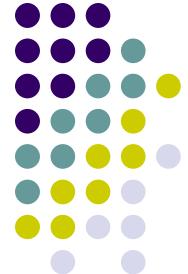
2. Detalle de ventas por artículo. Se genera mensualmente ya nivel de región, es decir, un reporte para cada una de ellas. Se distribuye manualmente por mail en formato Excel para que los gerentes de cada región puedan realizar las agrupaciones que deseen:

VENTAS DE ARTÍCULOS (REGION = SUR, MES = 201501).xlsx - Excel

J23	A	B	C	D	E	F	G	H	I	J
1										
2										
3	Subcategoría	Categoría	Artículo	Ingresos	Coste	Descuento				
4	Arte y arquitectura	Libros	Producto 1	\$184	\$131	\$ 0				
5	Arte y arquitectura	Libros	Producto 2	\$99	\$74	\$ 0				
6	Arte y arquitectura	Libros	Producto 3	\$96	\$82	\$ 0				
7	Arte y arquitectura	Libros	Producto 4	\$105	\$77	\$ 0				
8	Arte y arquitectura	Libros	Producto 5	\$81	\$59	\$ 0				
9	Arte y arquitectura	Libros	Producto 6	\$280	\$205	\$ 0				
10	Arte y arquitectura	Libros	Producto 7	\$126	\$93	\$ 0				
11	Arte y arquitectura	Libros	Producto 8	\$56	\$43	\$ 0				
12	Arte y arquitectura	Libros	Producto 9	\$55	\$41	\$ 0				
13	Arte y arquitectura	Libros	Producto 10	\$33	\$25	\$ 0				
14	Arte y arquitectura	Libros	Producto 11	\$196	\$151	\$ 0				
15	Arte y arquitectura	Libros	Producto 12	\$80	\$62	\$ 0				
16	Arte y arquitectura	Libros	Producto 13	\$280	\$205	\$ 0				
17	Arte y arquitectura	Libros	Producto 14	\$135	\$102	\$ 0				
18	Arte y arquitectura	Libros	Producto 15	\$117	\$93	\$ 0				
19	Empresa	Libros	Producto 16	\$76	\$58	\$ 0				
20	Empresa	Libros	Producto 17	\$66	\$49	\$ 0				
21	Empresa	Libros	Producto 18	\$108	\$80	\$ 0				
22	Empresa	Libros	Producto 19	\$64	\$49	\$ 0				

Data Warehouse - Business Intelligence

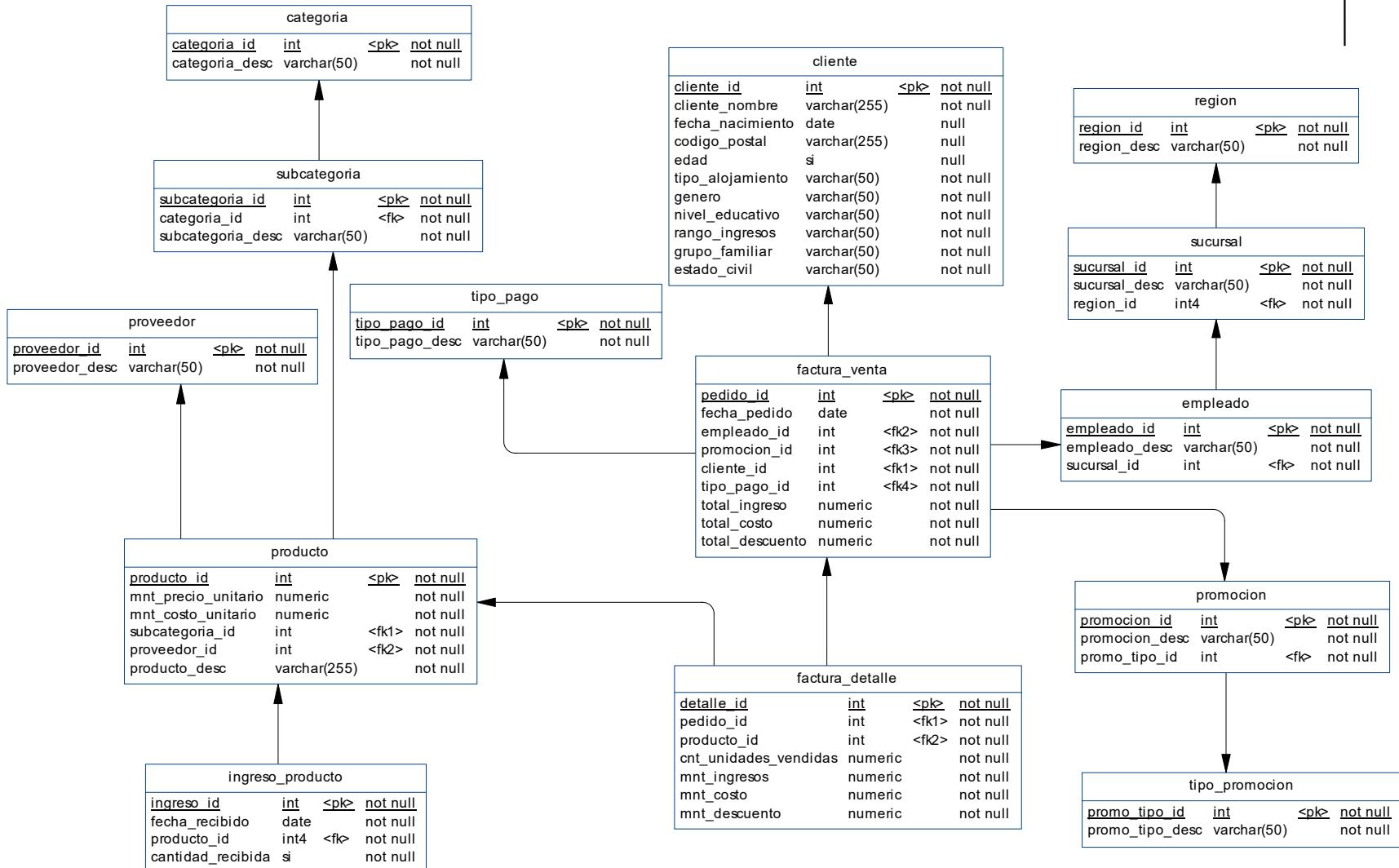
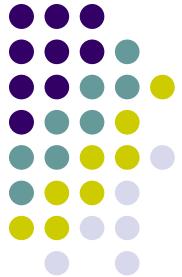
EJEMPLO PRÁCTICO



- Información demasiado detallada para el tipo de usuario que va a consumirla
- Los reportes sábana no se procesan fácilmente para poder sacar conclusiones. NO SIRVEN!!!!

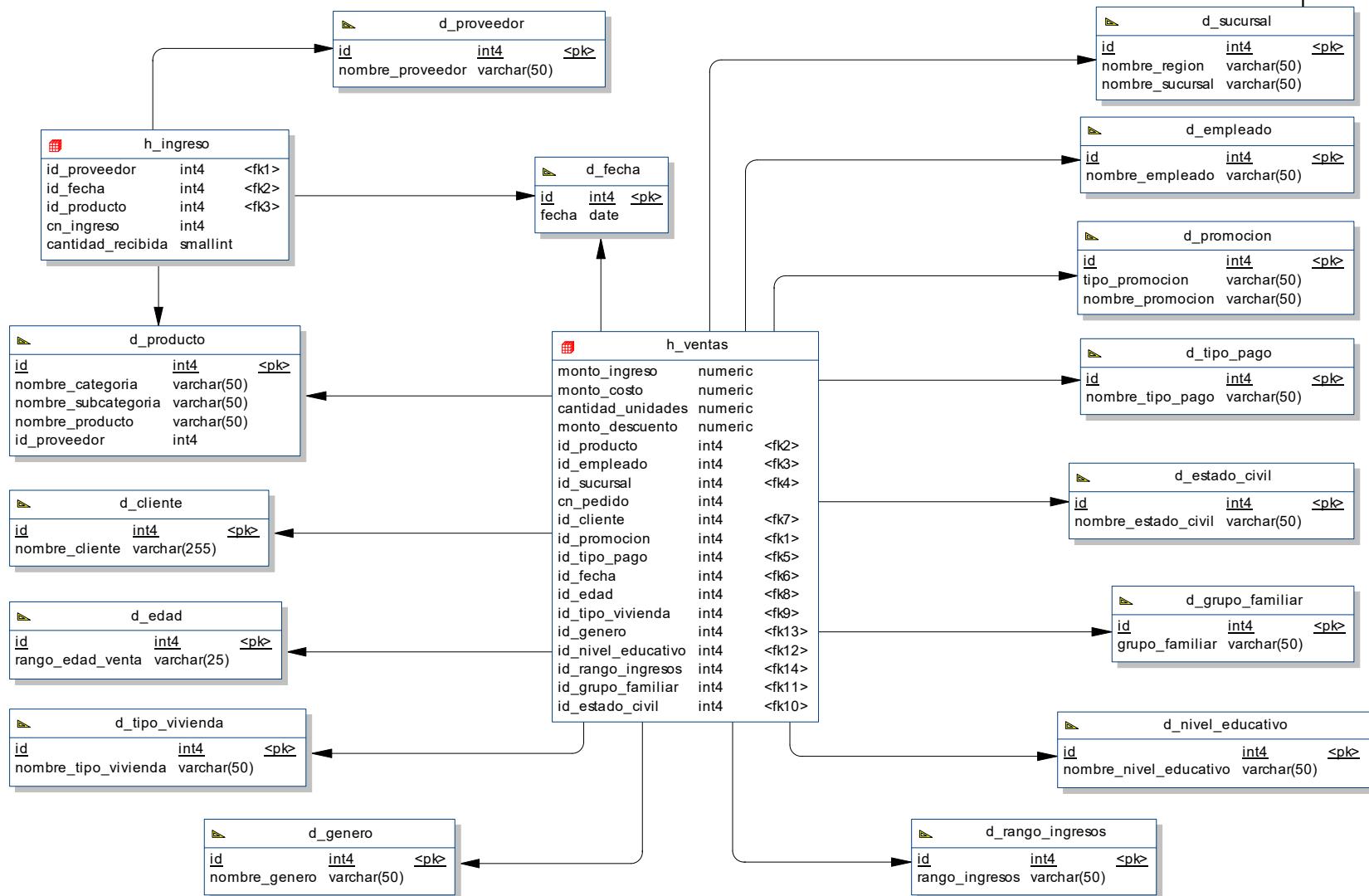
Data Warehouse - Business Intelligence

EJEMPLO PRÁCTICO





Data Warehouse - Business Intelligence EJEMPLO PRÁCTICO





Data Warehouse - Business Intelligence

EJEMPLO PRÁCTICO

Santa Fe Capital

Período

- Seleccionar todo
- 2015
- 2016
- 2017

Nivel educativo

- Seleccionar todo
- OTRO
- POSTGRADO
- SECUNDARIO

Vivienda

- Seleccionar todo
- DEPENDIENTE
- INQUILINO
- NO INFORMADO
- PROPIETARIO

Género

- Seleccionar todo
- FEMENINO
- MASCULINO
- OTRO

Estado civil

- Seleccionar todo
- CASADO
- NO INFORMADO
- PREVIAMENTE CASADO
- SOLTERO

Rango de edad

- Seleccionar todo
- 1 a 16
- 17 a 25
- 26 a 40
- 41 a 50
- 51 a 60
- 61 a 65
- 66 a 70

Región y sucursal

- Seleccionar todo
- Central Norte
- Central Sur
- Noreste
- Noroeste
- Sucursal 10

Empleado

- Seleccionar todo
- Empleado 1
- Empleado 10
- Empleado 12
- Empleado 13
- Empleado 14

Promociones

- Seleccionar todo
- Rebajas de Temporada
- Rebajas Especiales
- Sin promoción

Jerarquía de productos

- Seleccionar todo
- Electrónica
 - Cámaras
 - Equipo de sonido
 - Equipo de video
 - Ordenadores
 - Productos Electrónicos - Varios
 - Televisores

Tipo de pago

- Seleccionar todo
- Ame Express
- Cheque
- Club Card
- MC
- Visa

Restaurar

Aplicar

2.850
Cantidad de facturas

\$660.662,00
Ingreso total

Filtros Ventas Compras + 100

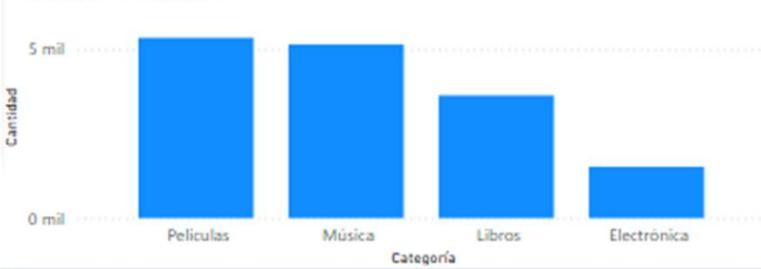


Data Warehouse - Business Intelligence

EJEMPLO PRÁCTICO

 Santa Fe Capital

Cantidades vendidas



Categoría	Cantidad
Películas	~4500
Música	~4500
Libros	~3500
Electrónica	~2000

Ingresos, costos y descuentos



Categoría	Porcentaje
Costo	44,44%
Descuento	1,58%
Ingreso	53,98%

Año Mes Ingreso acumulado Costo acumulado Descuento acumulado Margen bruto de utilidad

Año	Mes	Ingreso acumulado	Costo acumulado	Descuento acumulado	Margen bruto de utilidad
2015	enero	\$22.405,00	\$18.524,69	\$0,00	17,32 %
2016	enero	\$28.537,00	\$23.550,30	\$0,00	17,47 %
2015	febrero	\$23.980,00	\$19.733,12	\$0,00	17,71 %
2016	febrero	\$28.131,00	\$23.096,47	\$0,00	17,90 %
2015	marzo	\$9.741,00	\$8.070,41	\$0,00	17,15 %
Total		\$660.662,00	\$543.823,30	\$19.342,60	17,69 %

\$660.662,00
Ingreso total

\$543.823,30
Costo total

\$19.342,60
Descuento total

17,69 %
Margen de utilidad

Cantidad de productos por Año y Mes



Año	Mes	Cantidad de productos
2015	ene	~550
2015	abril	~600
2015	junio	~550
2015	septiembre	~1100
2015	diciembre	~400
2016	enero	~700
2016	abril	~650
2016	julio	~300
2016	octubre	~1100

Filtros Ventas Compras +



Data Warehouse - Business Intelligence

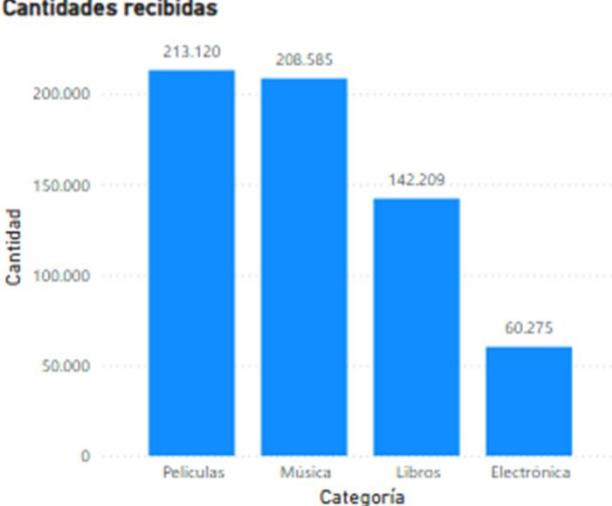
EJEMPLO PRÁCTICO

Santa Fe Capital

Proveedor

- Buscar
- Seleccionar todo
- Proveedor 1
- Proveedor 10
- Proveedor 11
- Proveedor 12

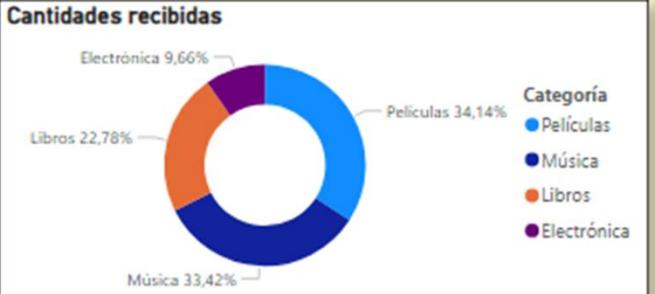
Cantidades recibidas



Categoría	Cantidad
Películas	213.120
Música	208.585
Libros	142.209
Electrónica	60.275

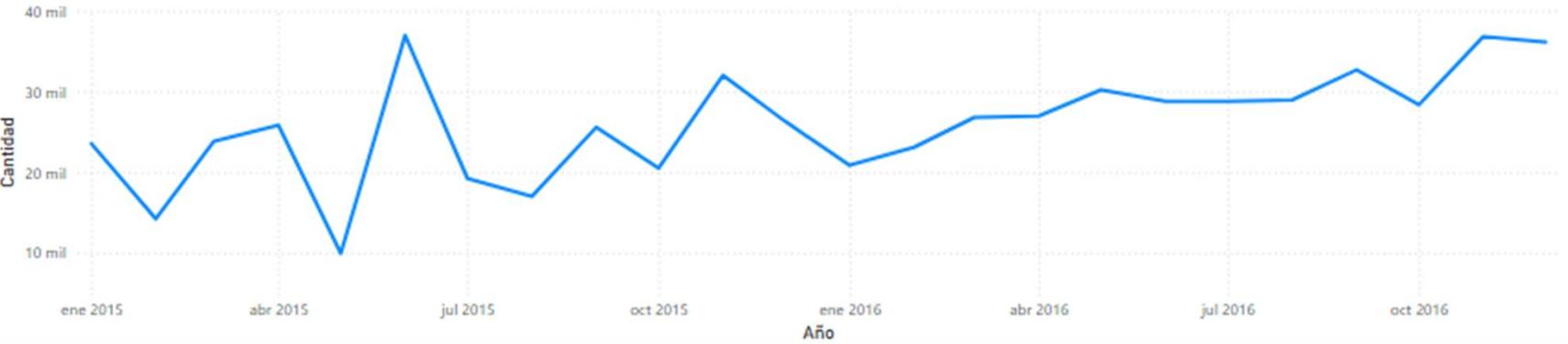
624.189
Cantidad de productos recibida

Cantidades recibidas



Categoría	Porcentaje
Películas	34,14%
Electrónica	9,66%
Libros	22,78%
Música	33,42%

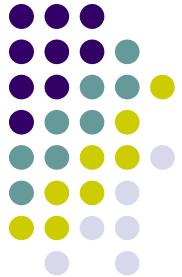
Evolución de ingresos



Año	Cantidad
ene 2015	25 mil
abr 2015	26 mil
jul 2015	18 mil
oct 2015	22 mil
ene 2016	21 mil
abr 2016	28 mil
jul 2016	29 mil
oct 2016	30 mil
diciembre 2016	37 mil

Filtros Ventas Compras x +

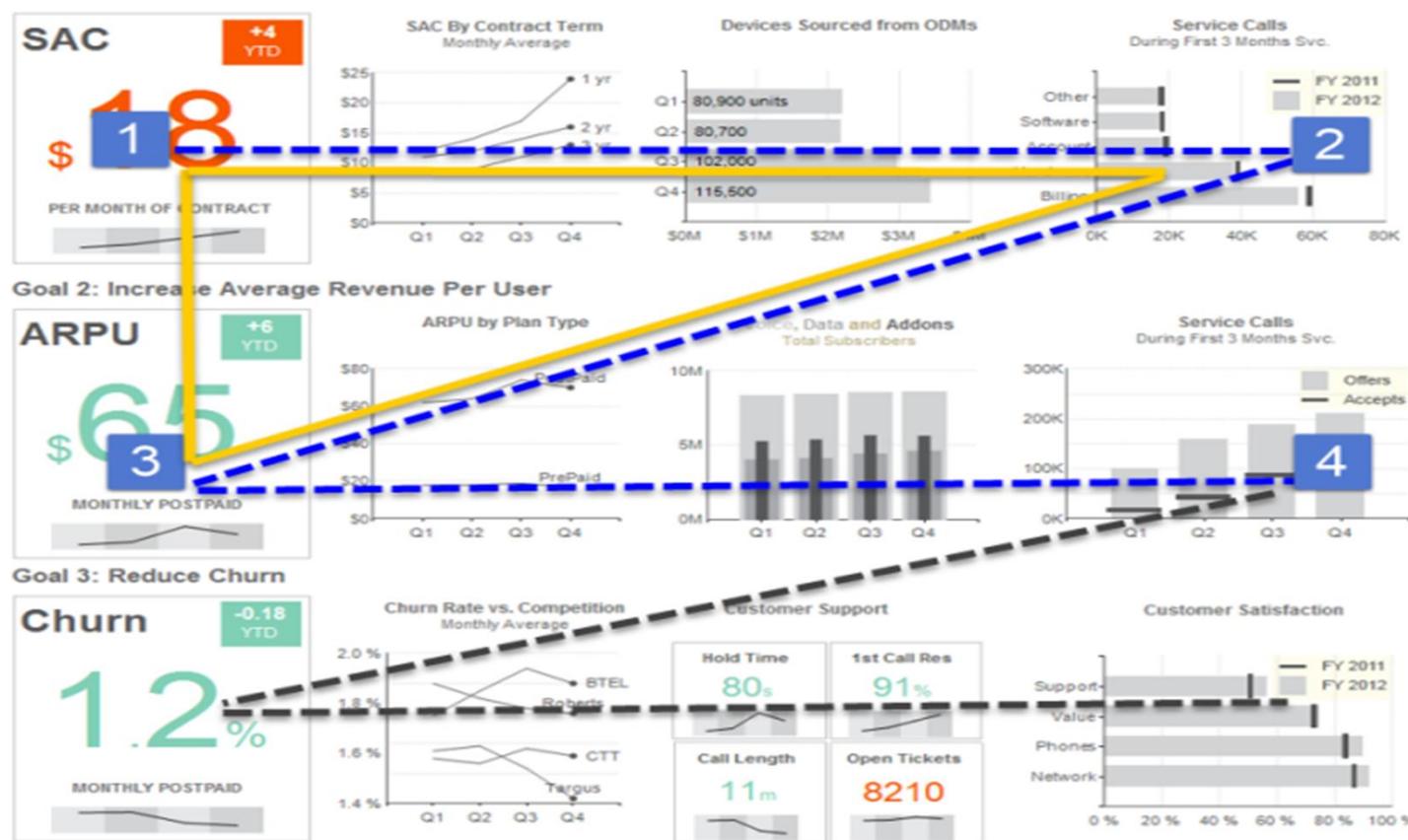
102



Data Warehouse - Business Intelligence

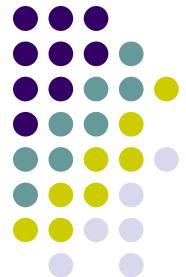
Diseño de dashboards

Disposición (layout)

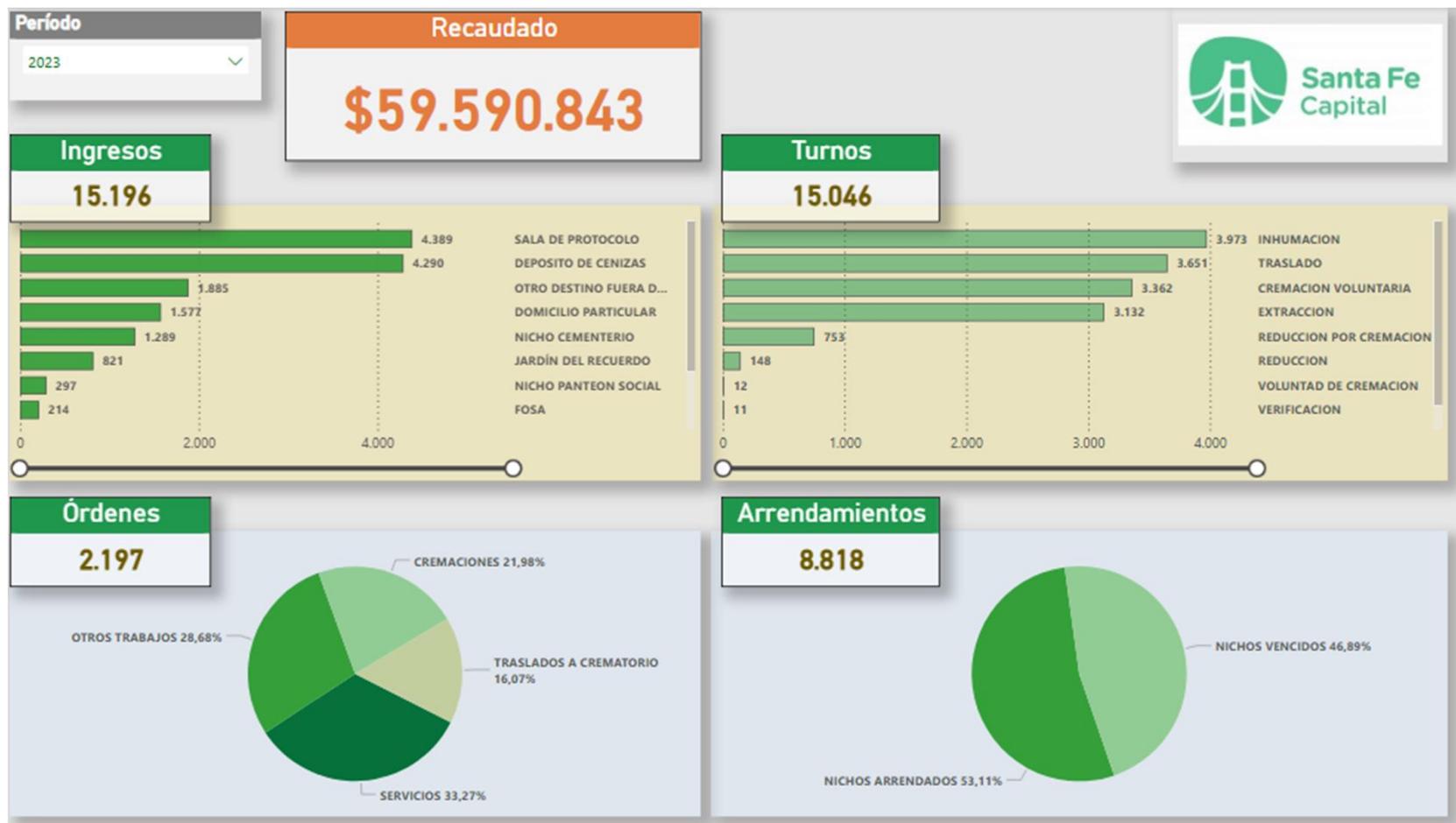


Data Warehouse - Business Intelligence

Diseño de dashboards

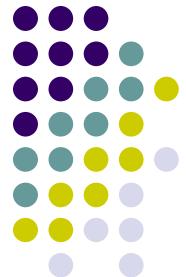


Acceso inmediato a lo relevante

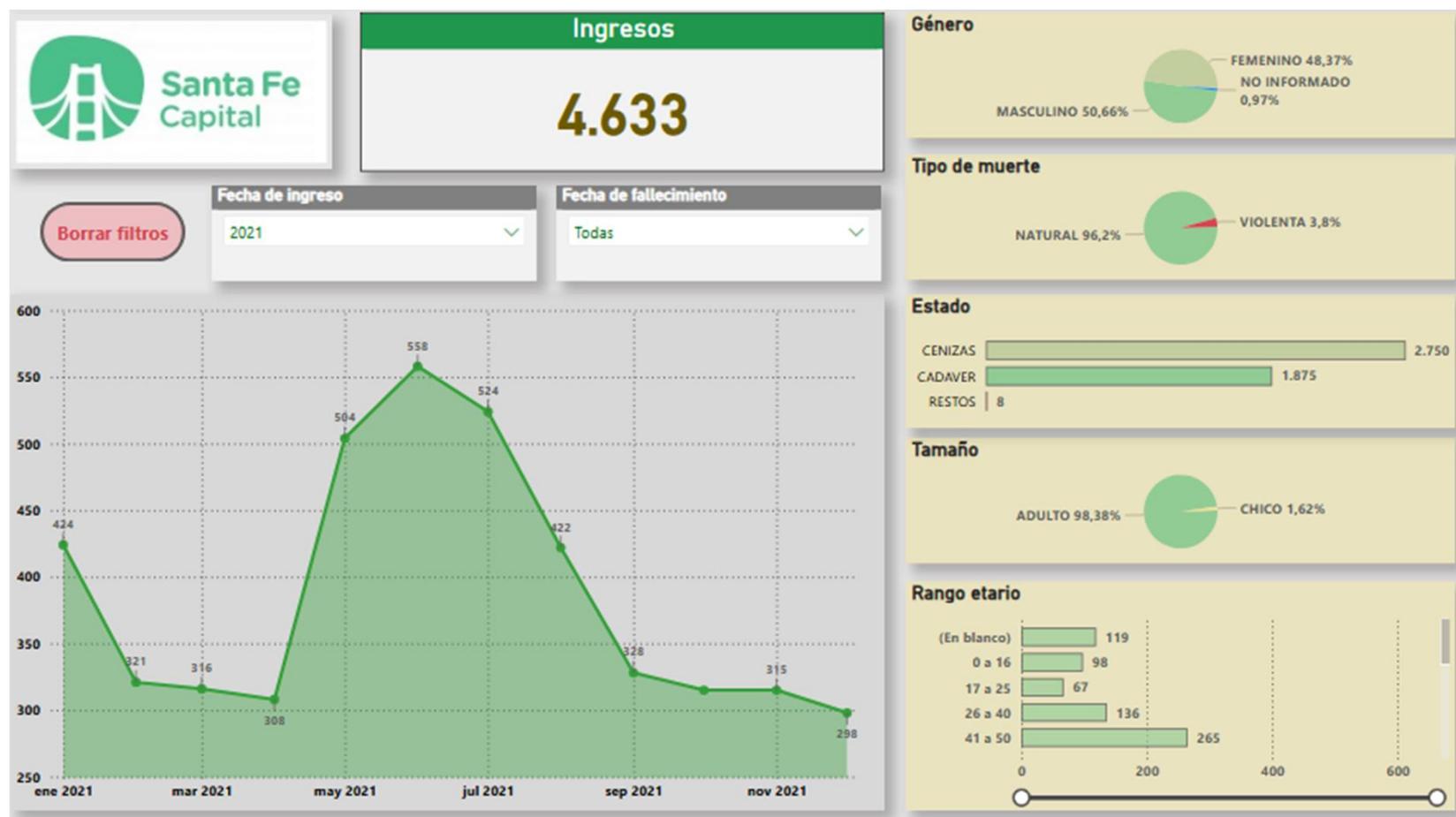


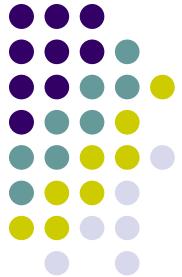
Data Warehouse - Business Intelligence

Diseño de dashboards



Acceso inmediato a lo relevante

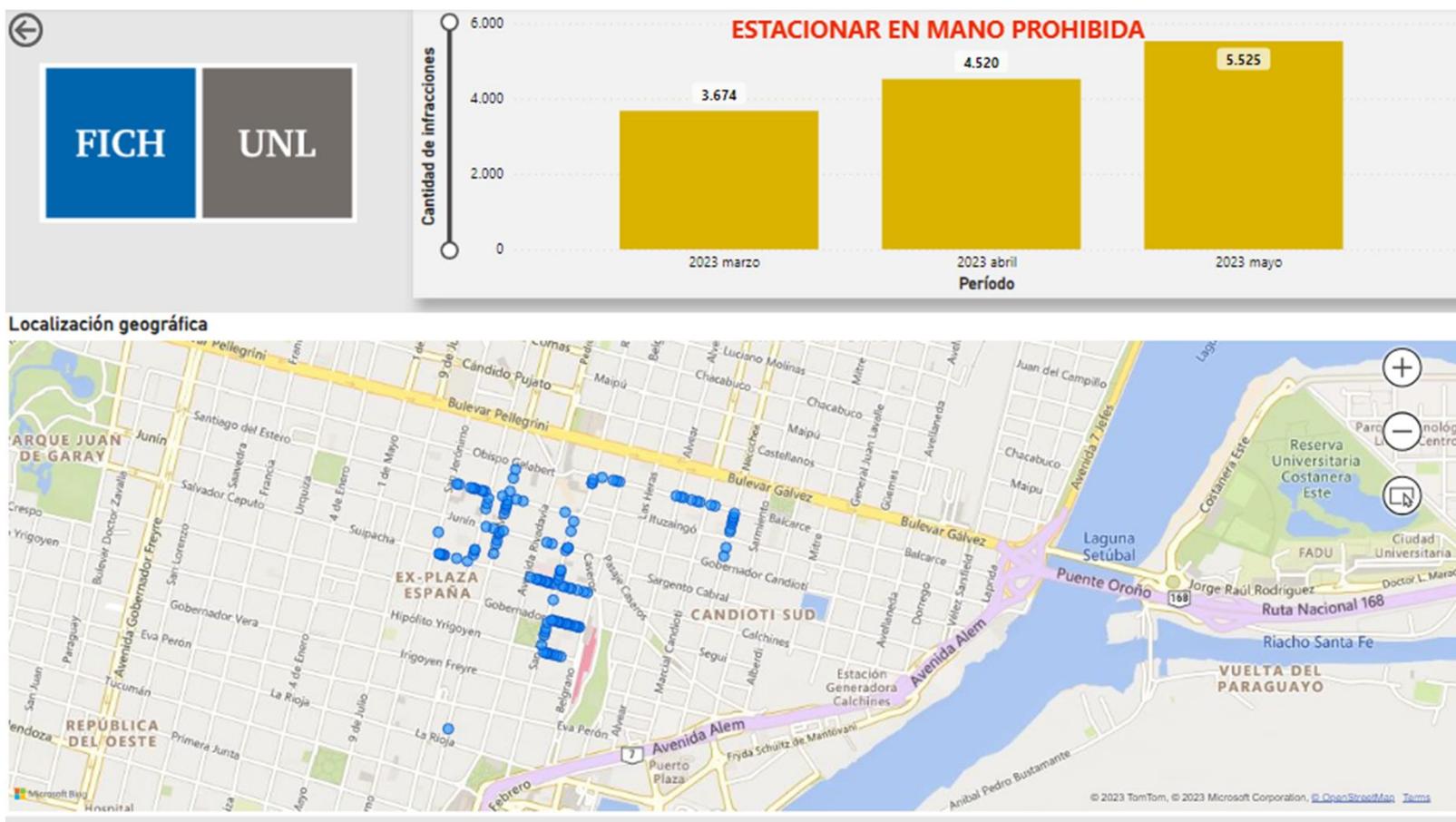


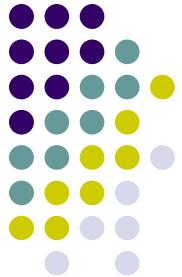


Data Warehouse - Business Intelligence

Diseño de dashboards

Menos es más





Data Warehouse - Business Intelligence

Diseño de dashboards

Errores frecuentes

The screenshot shows a web-based control interface for an Epson printer. On the left, a table lists various paper types and their properties. On the right, another table shows the selected paper type and its properties. A large circular arrow icon is overlaid on the interface, indicating a poor design choice.

Proveedor soporte	Nombre de soporte
Metamark	MDS-100-137 (WhiteGloss)
EPSON	GS Canvas Satin
ILFORD	OMNIJET_NanoSolvent_Premium_Gloss
Neschen	ERFURT Wallpaper CA DV 615 6029344
Sini	3606 Paper white 226
Sini	3607 BlueBack Satin
Maxifine	Maxifine Gloss Plus 100% cotton

#	Nombre de soporte	Tipo de soporte
1	Vinilo Brillo	Vinilo adhesivo
2	Vinilo Mate	Vinilo adhesivo
3	Canvas	Lienzo
4	Papel Fotográfico	Otros
5	Wallpaper	Otros
6	Papel White Back	Otros

A - Exceder los límites visuales de una sola pantalla

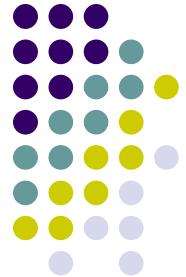
B - Mostrar excesivo detalle o precisión

C - Elegir métrica deficiente.

D - Elegir grafica inadecuada para una métrica

E - Variedad de gráficos en un solo dashboard

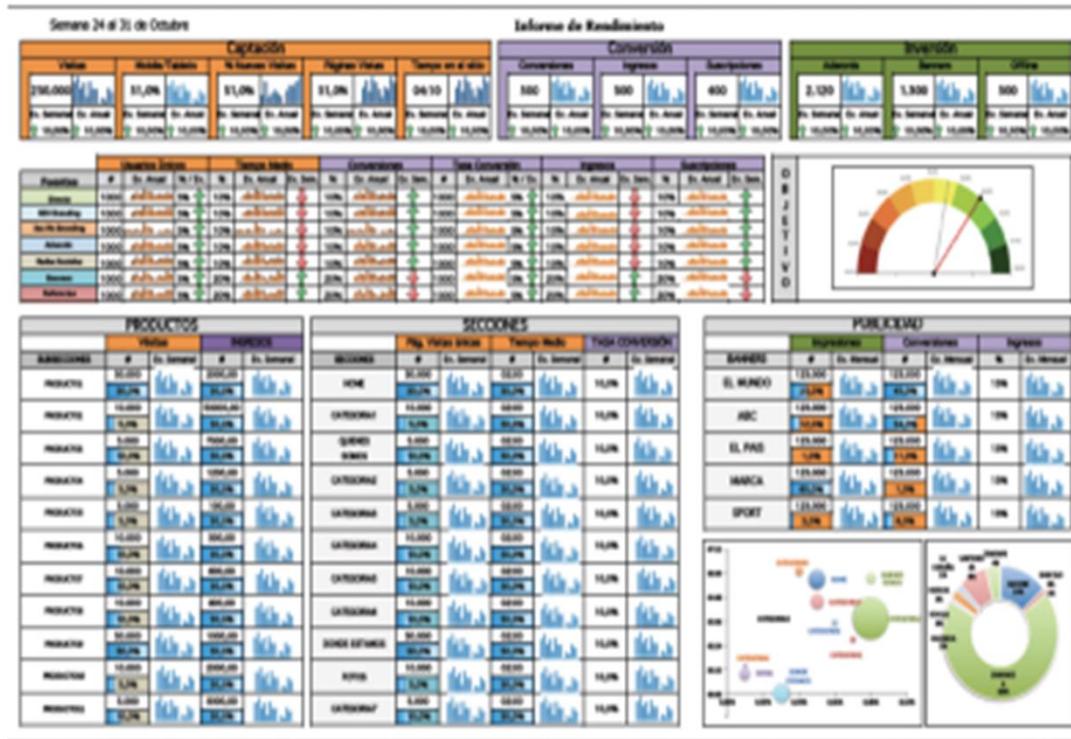
F - Usar un diseño pobre



Data Warehouse - Business Intelligence

Diseño de dashboards

Errores frecuentes



A - Exceder los límites visuales de una sola pantalla

B - Mostrar excesivo detalle o precisión

C - Elegir métrica deficiente.

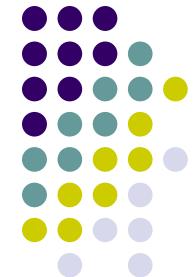
D - Elegir grafica inadecuada para una métrica

E - Variedad de gráficos en un solo dashboard

F - Usar un diseño pobre

Data Warehouse - Business Intelligence

Diseño de dashboards

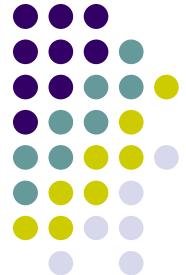


Errores frecuentes

Kg de Pan Vendidos en el Año

1.245.653 Kilos

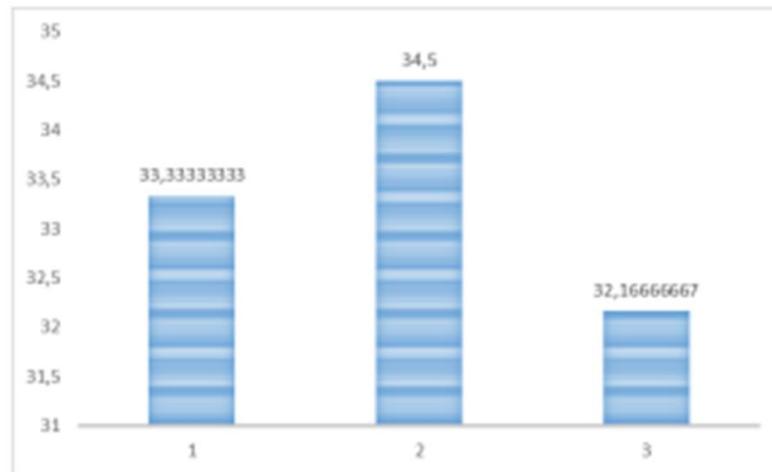
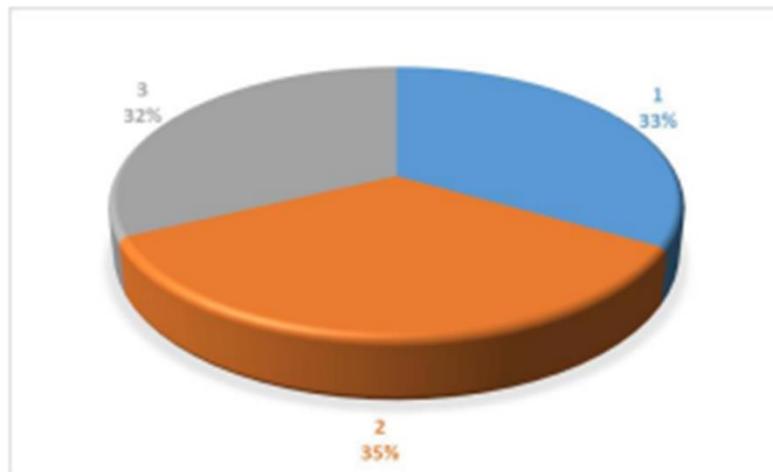
- A - Exceder los límites visuales de una sola pantalla
- B - Mostrar excesivo detalle o precisión
- C - Elegir métrica deficiente.
- D - Elegir grafica inadecuada para una métrica
- E - Variedad de gráficos en un solo dashboard
- F - Usar un diseño pobre



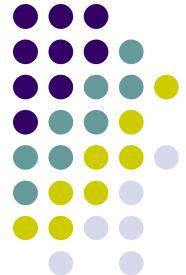
Data Warehouse - Business Intelligence

Diseño de dashboards

Errores frecuentes



- A - Exceder los límites visuales de una sola pantalla
- B - Mostrar excesivo detalle o precisión
- C - Elegir métrica deficiente.
- D - Elegir grafica inadecuada para una métrica**
- E - Variedad de gráficos en un solo dashboard
- F - Usar un diseño pobre



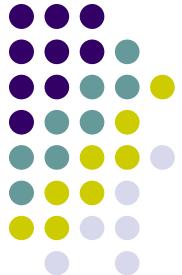
Data Warehouse - Business Intelligence

Diseño de dashboards

Errores frecuentes



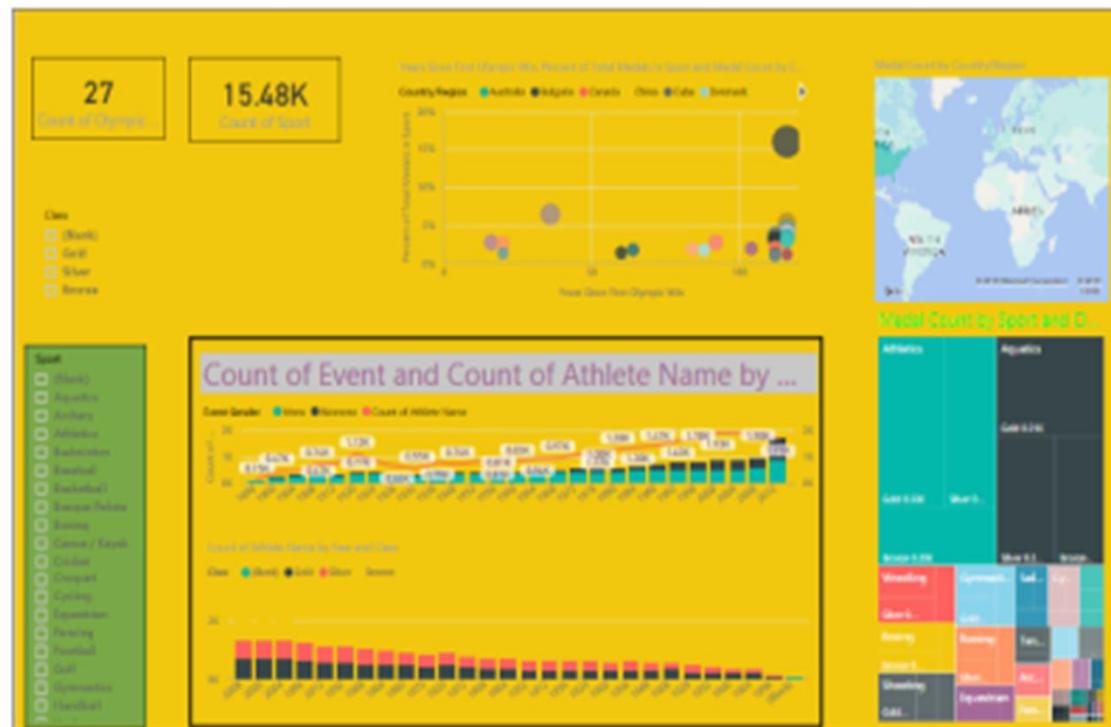
- A - Exceder los límites visuales de una sola pantalla
- B - Mostrar excesivo detalle o precisión
- C - Elegir métrica deficiente.
- D - Elegir grafica inadecuada para una métrica
- E - Variedad de gráficos en un solo dashboard**
- F - Usar un diseño pobre



Data Warehouse - Business Intelligence

Diseño de dashboards

Errores frecuentes



- A - Exceder los límites visuales de una sola pantalla
- B - Mostrar excesivo detalle o precisión
- C - Elegir métrica deficiente.
- D - Elegir grafica inadecuada para una métrica
- E - Variedad de gráficos en un solo dashboard
- F - Usar un diseño pobre