

TRABAJO ACADEMICO Nro. 8

Escuela Profesional:		Ingeniería de Sistemas	Asignatura:	Gestión de Datos e Información II	
Ciclo:	V	Turno:	Mañana/Tarde	Semestre Académico:	2022-1
Docentes:	Ing. Gustavo Coronel Castillo		Fecha:	Del 23/05/2022 al 28/05/2022	

Integrando Datos 1

I.	INTRODUCCION	2
	Esquema de estrella	2
	Tabla de hechos	3
	Tabla de dimensiones	4
	Esquema en copo de nieve	4
II.	OBJETIVO	6
III.	METODOLOGIA Y ACTIVIDADES	6
IV.	RECURSOS	6
V.	OBSERVACIONES.....	6
VI.	REQUERIMIENTO A RESOLVER.....	7

I. INTRODUCCION

El modelo de datos dimensional, conlleva una técnica de modelado que facilita la comprensión de la base de datos, haciéndola intuitiva para usuarios no expertos, y es comúnmente utilizada para implementar los Data Warehouse (DWH) o Data Mart (DM).

Esta técnica goza de una gran aceptación y, a menudo, es elegida como la preferida para representar datos analíticos por cumplir simultáneamente con los siguientes requerimientos:

- Dispone y estructura los datos de manera comprensibles para el usuario de negocio.
- Genera un alto rendimiento en las búsquedas desde la capa de reporting.

Dentro del modelado de datos dimensional destacan 2 conceptos clave: hechos y dimensiones.

- **Hechos:** Son las métricas, normalmente valores cuantitativos (numéricos) susceptibles de ser agregados

Ejemplo: La cantidad de ventas de coches de un concesionario, el rendimiento en euros de una empresa, el número de estudiantes de un colegio, etc.

- **Dimensiones:** Son los valores cualitativos. Proporcionan descripciones a los hechos, aportando un contexto a los mismos.

Ejemplo: Marca de coche, fecha, nombre concesionario, dirección de la empresa, nombre del colegio, etc.

Existen 2 técnicas para llevar a cabo el modelado dimensional: el esquema de estrella y el esquema de copo de nieve.

Esquema de estrella

Un esquema de estrella es un modelo de datos formado por una tabla de hechos, que contiene los datos para el análisis, rodeada de las tablas de dimensiones.



Como puedes observar en la imagen anterior la tabla de hechos es TH-Ventas y está rodeada de las dimensiones TD-Almacén, TD-Producto y TD-Cliente, almacenando el ID de cada dimensión en la tabla de hechos para, así, poder relacionar los atributos descriptivos de cada dimensión con la fila de la tabla de hechos.

El modelo estrella separa los datos del proceso de negocio en: hechos y dimensiones. Los hechos contienen datos medibles, cuantitativos, y las dimensiones los atributos que describen los datos indicados en los hechos.

Tabla de hechos

- Clave principal compuesta por los claves principales de las tablas de dimensiones.
- Registra medidas o métricas de un evento específico. Ejemplo: cliente compra un geranio de maceta de 25 cm en floristería mineral vegetal Lola a las 12:30 am del 10 de Octubre de 2027.
- Evita repetir de manera completa los atributos dimensionales. En la TH sólo irá un ID de la dimensión.
- Se diseñan según el nivel de granularidad deseado, pudiendo registrar eventos a un gran nivel de atomicidad

Tabla de dimensiones

Tienen una clave primaria simple.

- Generalmente tienen un número bajo de registros
- Cada registro puede contener un gran número de atributos
- Suelen contener una surrogate primary key, generalmente una columna de tipo entero

Las principales ventajas del esquema de estrella son:

- Queries simples. Las uniones y cruces son más sencillos, debido a su lógica, que los de un esquema normalizado.
- Lógica de reporting simplificada.
- Mejoras en el rendimiento de las consultas.
- Agregaciones más rápidas. Gracias a las queries simplificadas.

Las principales desventajas del esquema de estrella son:

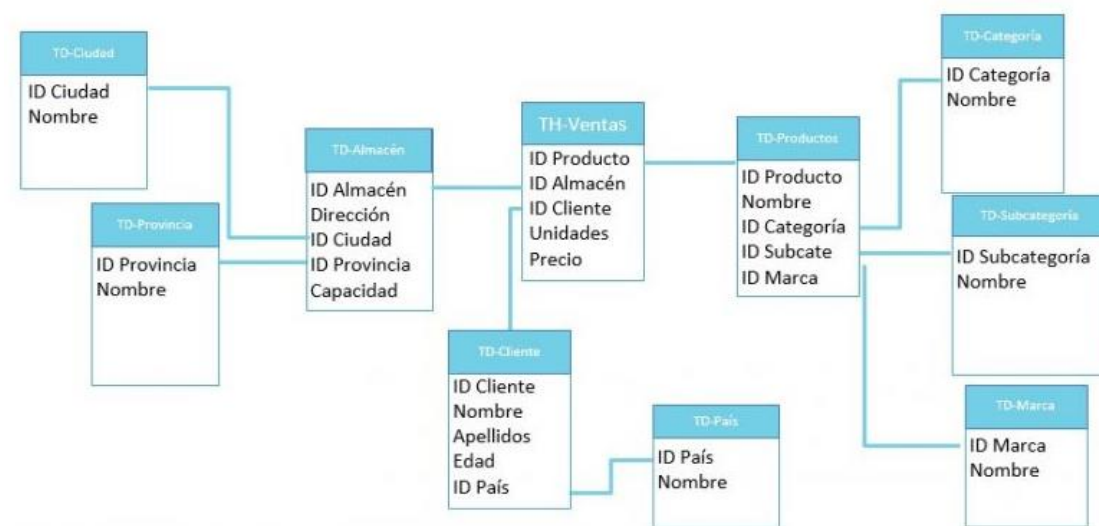
- Poco flexible. Los esquemas en estrella son contruidos para una vista de los datos en particular.

Esquema en copo de nieve

Un esquema de copo de nieve es una estructura más compleja que el esquema de estrella. Se da cuando alguna de las dimensiones se implementa con más de una tabla de datos.

El objetivo es normalizar estas tablas y reducir el espacio de almacenamiento al eliminar la redundancia.

Se representa como una tabla de hechos conectada con dimensiones anidadas. Al normalizar por completo las dimensiones el resultado parece un copo de nieve.



Como observas en la imagen se dividen las dimensiones de TD-Almacén, TD-Producto y TD-Cliente en sub-dimensiones normalizadas.

Las principales ventajas del esquema de copo de nieve son:

- Algunas herramientas de modelado de bases de datos multidimensional OLAP se optimizan.
- La normalización de los atributos reduce el almacenamiento de datos.

Las principales desventajas del esquema de copo de nieve son:

- Queries complejas debido a la normalización (implica un mayor número de cruces)
- Bajo rendimiento debido a la normalización

Después de haber descrito los esquemas de estrella y copo de nieve vamos a dejar una tabla con la comparativa entre los dos esquemas:

	Estrella	Copo de nieve
Mantenimiento	Tiene redundancia. Peor mantenimiento	No hay redundancia. Fácil mantenimiento
Facilidad de uso	Queries menos complejas. Fácil uso	Queries complejas. Difícil de entender
Rendimiento de las queries	Ejecuciones más rápidas	Más tiempo de ejecución debido a los cruces
Tipo de DWH	Data Mart	Data Warehouse
Joins	Bajo número de joins	Alto número de joins
Tablas de dimensión	Una tabla de dimensión por cada dimensión	Más de una tabla de dimensión por cada dimensión
Cuándo usarlo	Cuándo las tablas de dimensión tienen pocas filas	Cuándo las tablas de dimensión tienen un tamaño bastante elevado
Normalización / De-Normalización	Tablas de dimensiones y de hechos denormalizadas.	Tablas de dimensión normalizadas. Tablas de hechos denormalizadas.
Modelo de datos	Top-Down	Bottom-up

II. OBJETIVO

Diseñas e implementar una base de datos dimensional.

III. METODOLOGIA Y ACTIVIDADES

Desarrolle el modelo dimensional de la base de datos PUBS:

1. Desarrolle el diagrama E-R de la base de datos PUBS.
2. Realice una descripción de cada una de las tablas.
3. Realice el diseño de la base de datos dimensional, explique el objetivo de cada una de las tablas.
4. Crear el script SQL para la implementación de la base de datos dimensional.
5. Implemente la base de datos dimensional.
6. Recoger evidencias.
7. Elaborar informe que incluya conclusiones y recomendaciones.

IV. RECURSOS

1. Acceso a internet.
2. Computadora personal con Windows 10 o superior
3. Microsoft SQL Server correctamente instalado.
4. Microsoft SQL Server Management Studio
5. Plataforma Blackboard para la entrega de su solución.

V. OBSERVACIONES

1. Se debe crear una **carpeta de trabajo** con el nombre "**\GDI2\GPO##TA08**", en donde grabara todos los archivos de este trabajo académico. Donde **##** representa el número de grupo, por **01**, **02**, **03**, y así sucesivamente.
2. Para finalizar, debe empaquetar la carpeta **GPO##TA08** en un archivo **RAR** o **ZIP**, este archivo empaquetado es el que debe subir a la plataforma Blackboard en la sección correspondiente.

VI. REQUERIMIENTO A RESOLVER

Crear un Data Mart de nombre **PUBSDM** de la base de datos **PUBS** para implementar una solución de inteligencia de negocios.

Para resolver este requerimiento debe seguir todos los pasos descritos en el punto III.

El script para crear la base de datos PUBS lo encuentras en el siguiente repositorio:

<https://github.com/gcoronelc/databases>