



Seminario de Sistemas 2 [A]

Jose Fernando Alvarez Morales

Día, Fecha	Martes, 05/08/2024
Hora de inicio:	17:20

Laboratorio

Seminario de Sistemas 2 Sección “A”



Agenda



- Avisos
- Clase 3
 - Datawarehouse vs Datamart
 - Modelos de Datos
 - SSIS
- Corto 1
- Tarea 2
- Practica 1
- Ejemplo Practico



Avisos Generales

Clase 3



Datamart

- Un datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica.



Datamart

- Se caracteriza por disponer la estructura óptima de datos para analizar la información a detalle desde todas las perspectivas que afectan a los procesos de dicho departamentos.

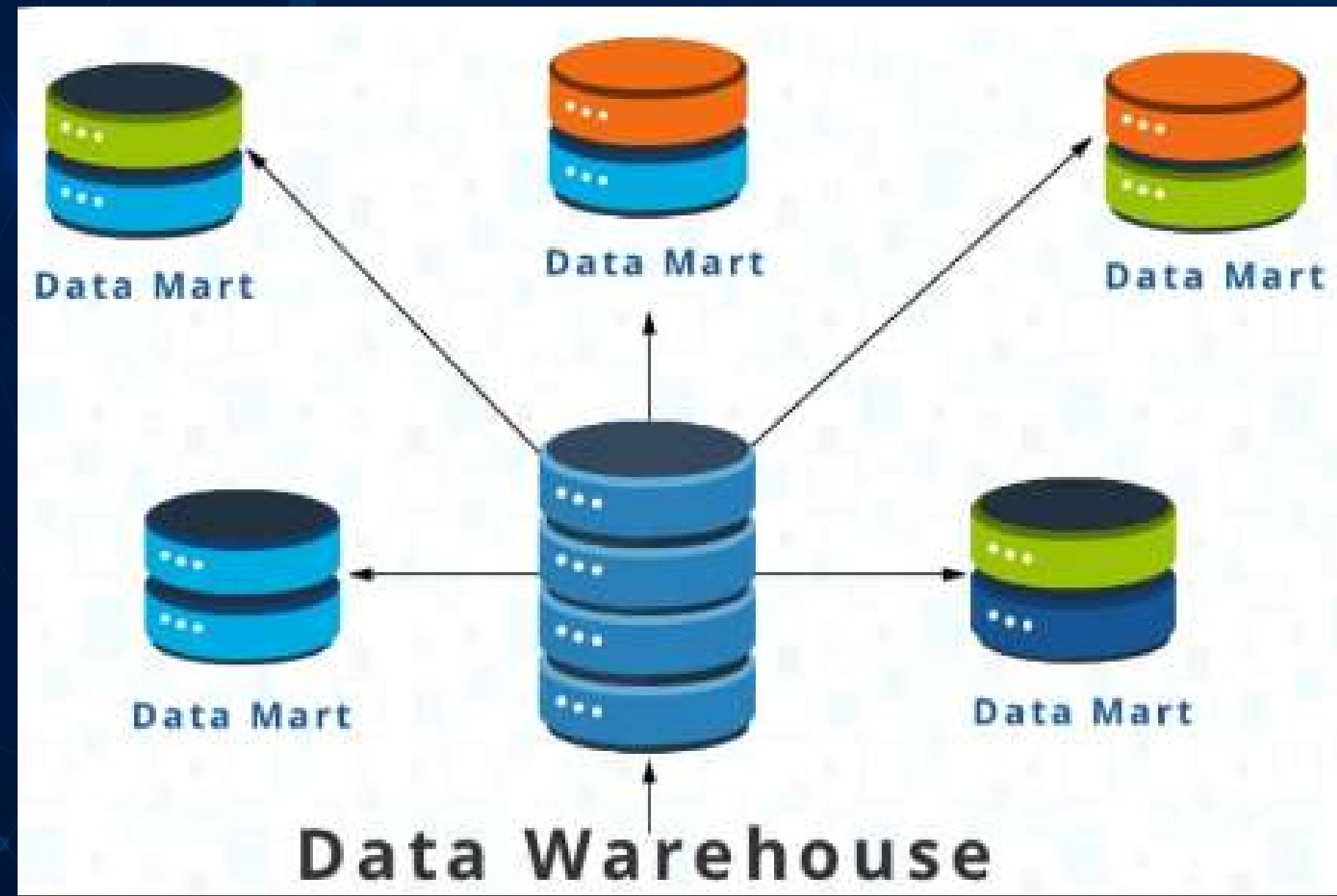


Datamart

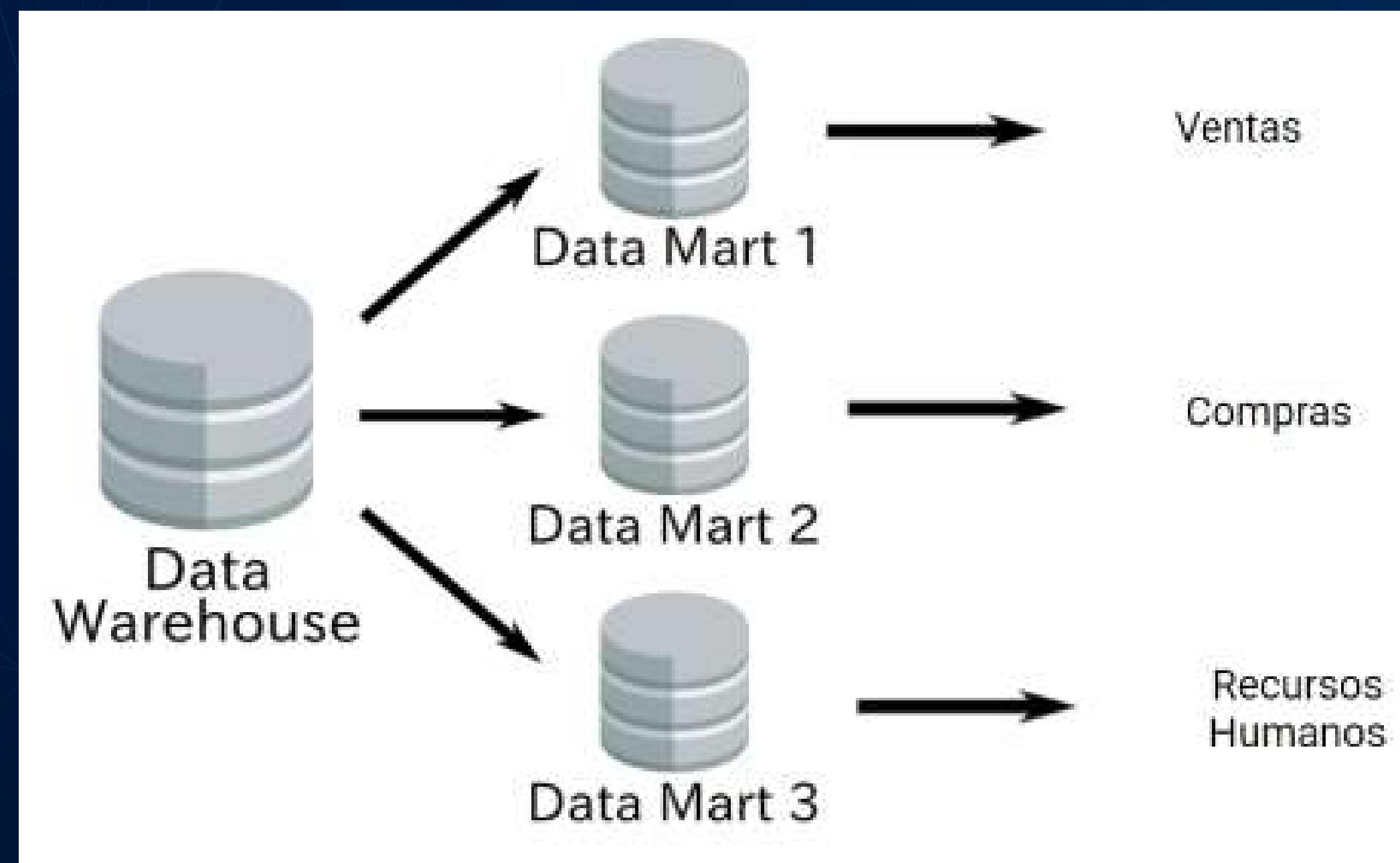
- Los datamarts son subconjuntos de los datos de Datawarehouse con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.



Representación gráfica de un Datamart



Representación gráfica de un Datamart



Datawarehouse y Datamart



Al hablar de un Datawarehouse, en este se contiene **todos** los datos de una organización

Un Datamart solamente obtiene un subconjunto de los datos de una organización, lo que hace centrarlo en un área específica dentro de una organización

Datawarehouse y Datamart



Un problema que surge es cuando el datawarehouse llega a crecer y a tornarse muy complejo. Debido a esto el rendimiento de las consultas decae y el modelo deja de ser óptimo.

En estos casos la solución es la creación de datamarts especializados por áreas como Ventas, Compras, etc.



Ventajas y Desventajas del Datamart

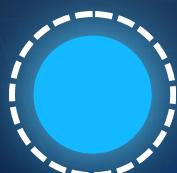
Ventajas

- Consulta mas rápidas debido al poco volumen de datos a recorrer.
- Fácil acceso a los datos que se utilizan con frecuencia.
- Su costo de construcción es relativamente menor a la de un datawarehouse.

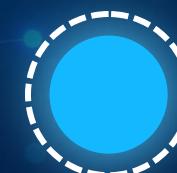
Desventajas

- No maneja grandes volúmenes de información.
- No considera otras fuentes de datos de la empresa.

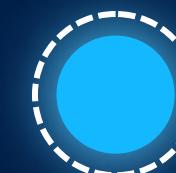
Tipos de Datamart



Dependientes



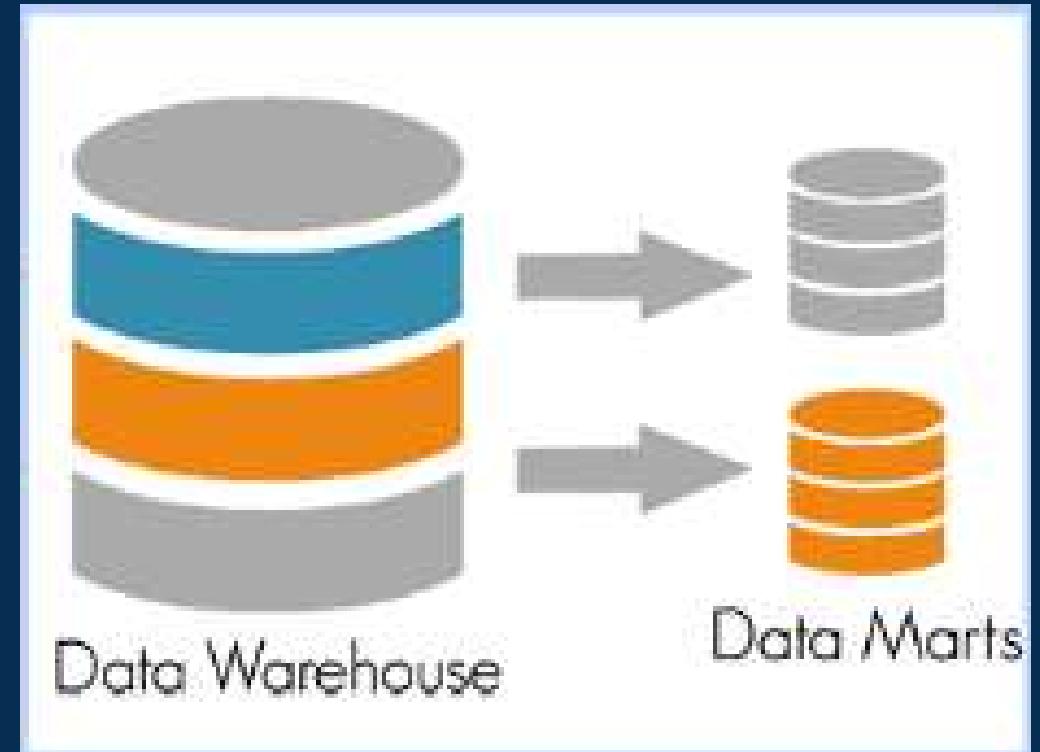
Independientes



Híbridos

Dependiente:

Este tipo de data mart se basa en un data warehouse centralizado. Obtiene sus datos del data warehouse y los organiza de una manera específica para las necesidades del grupo de usuarios objetivo.



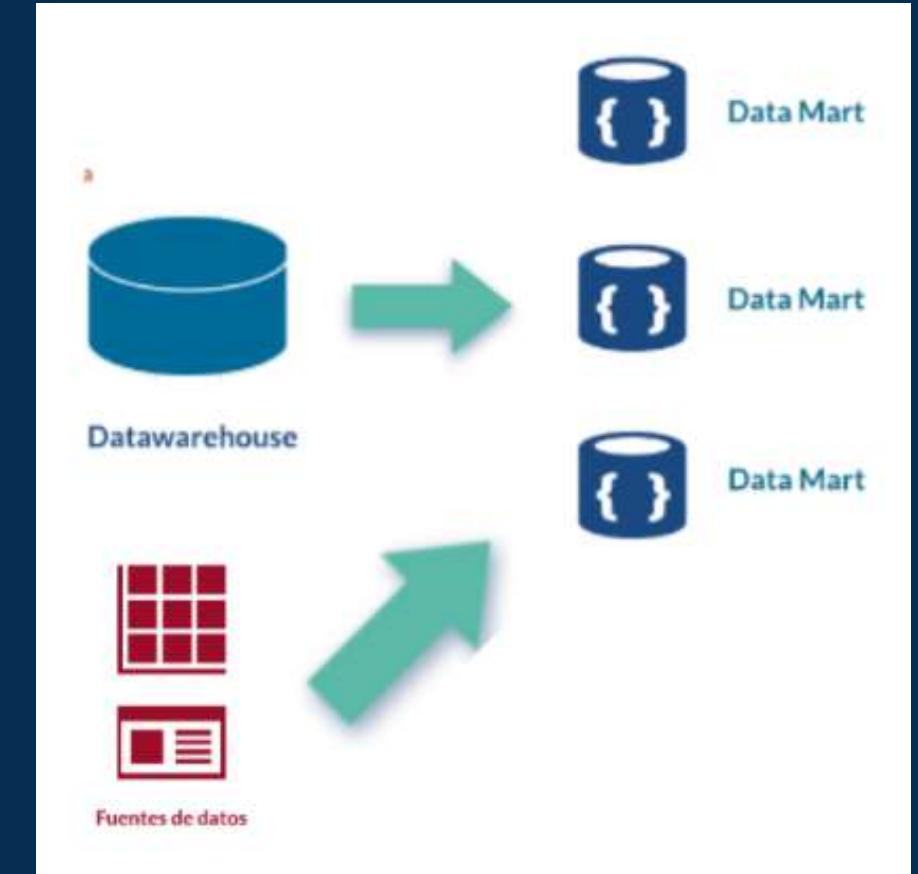
Independiente:

Este tipo de data mart no depende de un data warehouse centralizado. Se crea y administra de forma independiente, con sus propios datos y estructuras.



Híbrido:

En este tipo combina datos de un datawarehouse con otros sistemas de fuentes de datos.



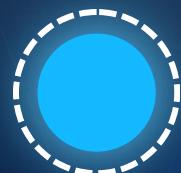
Datawarehouse vs Datamart

	Datawarehouse	Datamart
Alcance	Almacena información relacionada con todo el sistema.	Se limita a almacenar información de un área de la organización en específico.
Tamaño e Integración	Almacena grandes cantidades de datos provenientes de muchas fuentes de datos, por lo que suele ser más grande.	Se concentra en resúmenes de datos totalizados por lo que suele ser más pequeña.
Creación	La creación es más complicada ya que debe contemplar todos los datos del sistema.	La creación es más simple ya que tiene menos relaciones y están enfocados a sólo un tema.

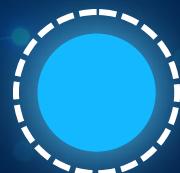
Datawarehouse vs Datamart

	Datawarehouse	Datamart
Costo de Manejo	Más costoso, porque requiere más recursos físicos para manejar grandes cantidades de datos.	Es menos costoso ya que requiere menos recurso físicos para manejar los datos requeridos.
Objetivo	Optimizar la obtención de datos, integrando y optimizando los datos fuente.	Es diseñado para entregar de manera óptima la información para el soporte de decisiones de negocio.

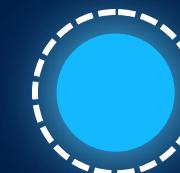
Modelos de Datos



Modelos de Estrella o
Star Scheme



Modelo de Copo de
nieve o Snowflake
Scheme



Modelo Constelación
(Copo de
Constelaciones) o
StarFlake Scheme

Modelo Estrella



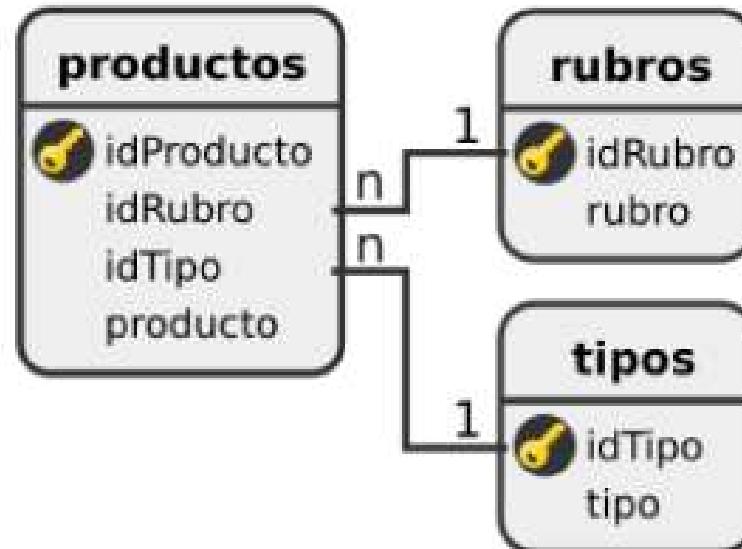
- Es el más sencillo en su estructura, consta con:
- Una tabla central de Hechos.
- Varias tablas de dimensión.
- Lo característico de este modelo es que la única tabla que tiene relación con otras tablas es la de hecho.

Modelo Estrella

Desnormalizado

dimProductos
idProducto
rubro
tipo
producto

Normalizado



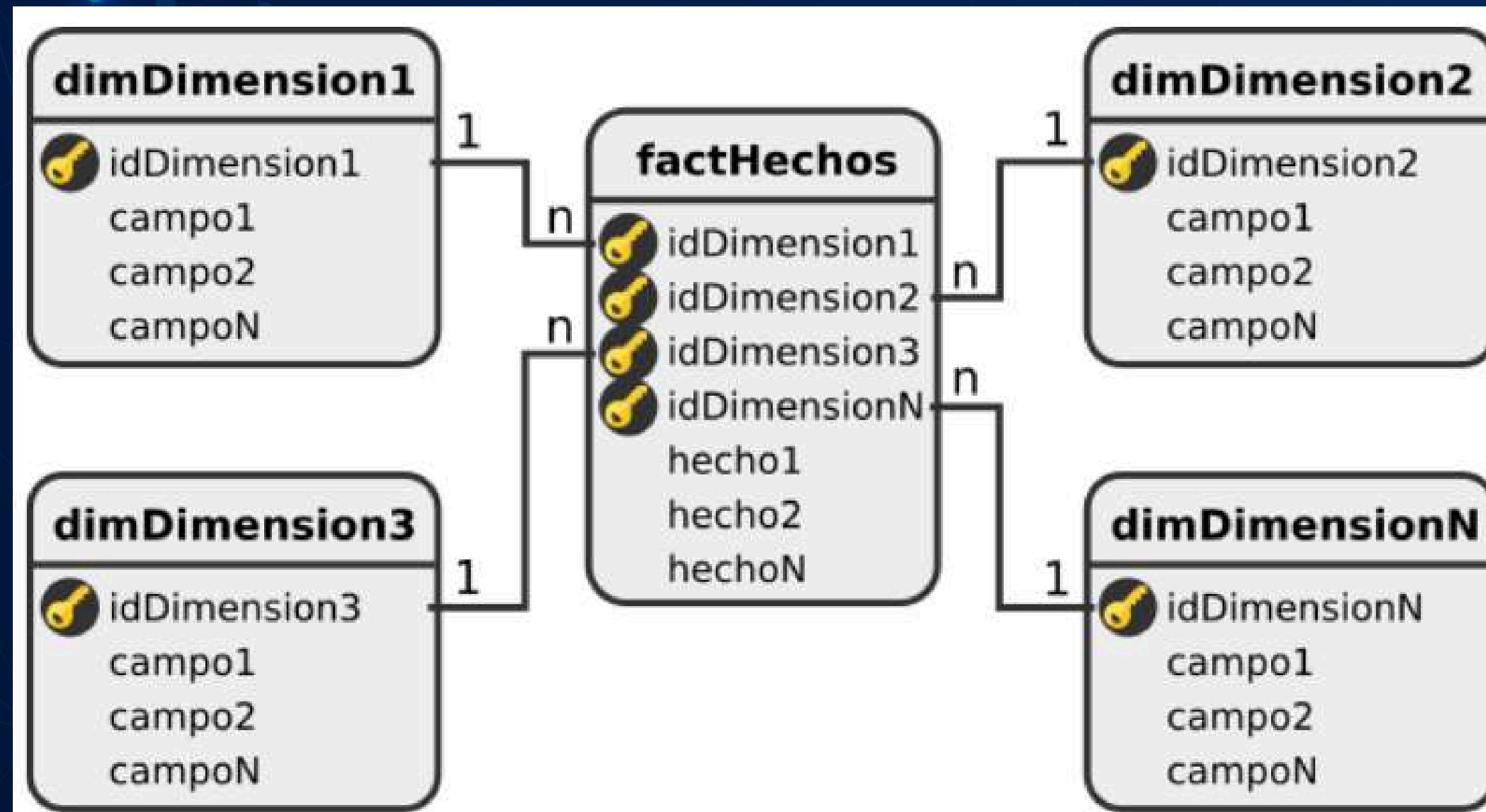
- Las tablas de dimensión sólo están relacionadas con la tabla de hechos.
- Las tablas de dimensión se encuentran desnormalizadas.

Modelo Estrella

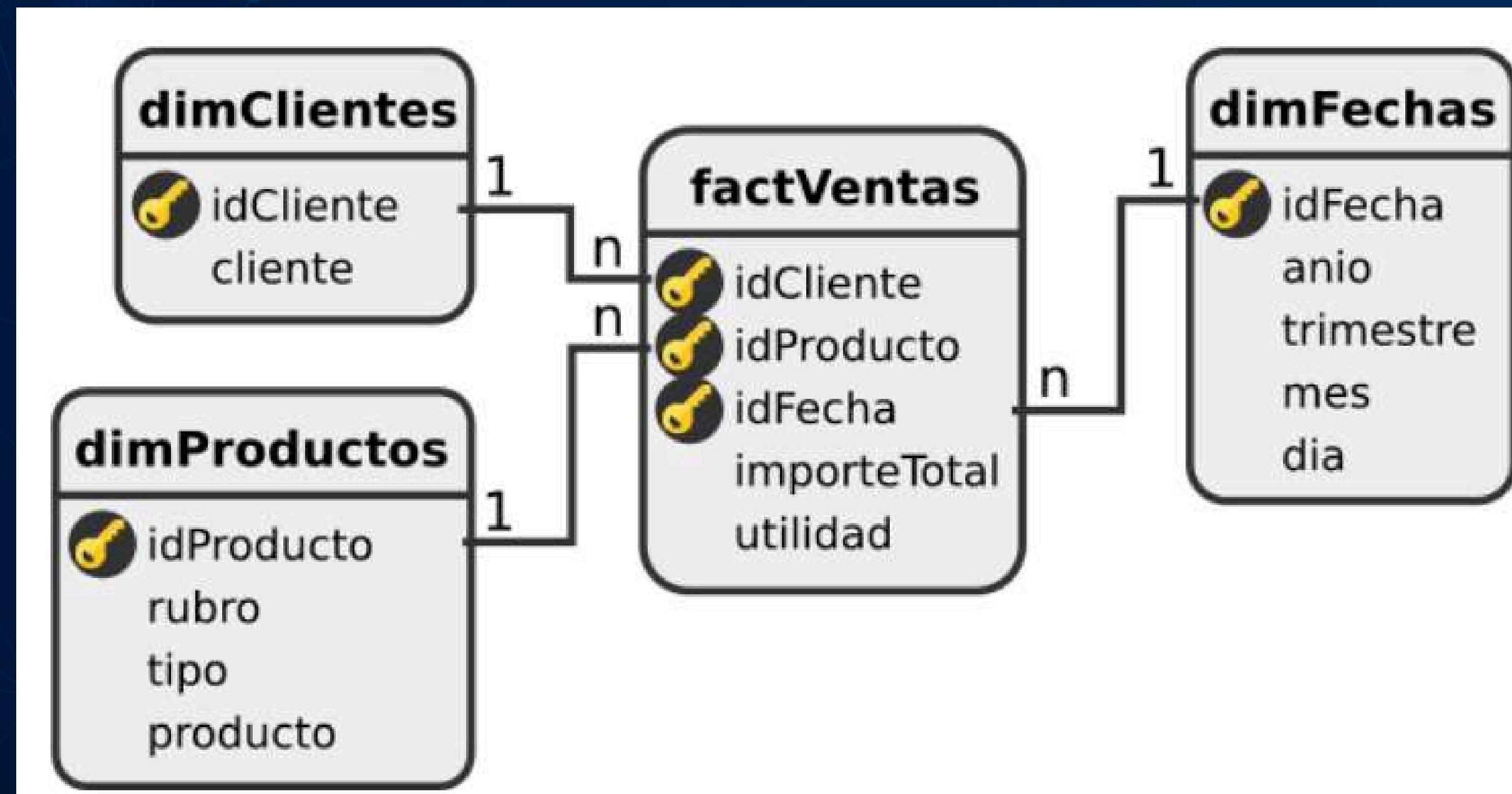


- El esquema estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios.
- Este modelo es soportado por casi todas las herramientas de consulta y análisis.

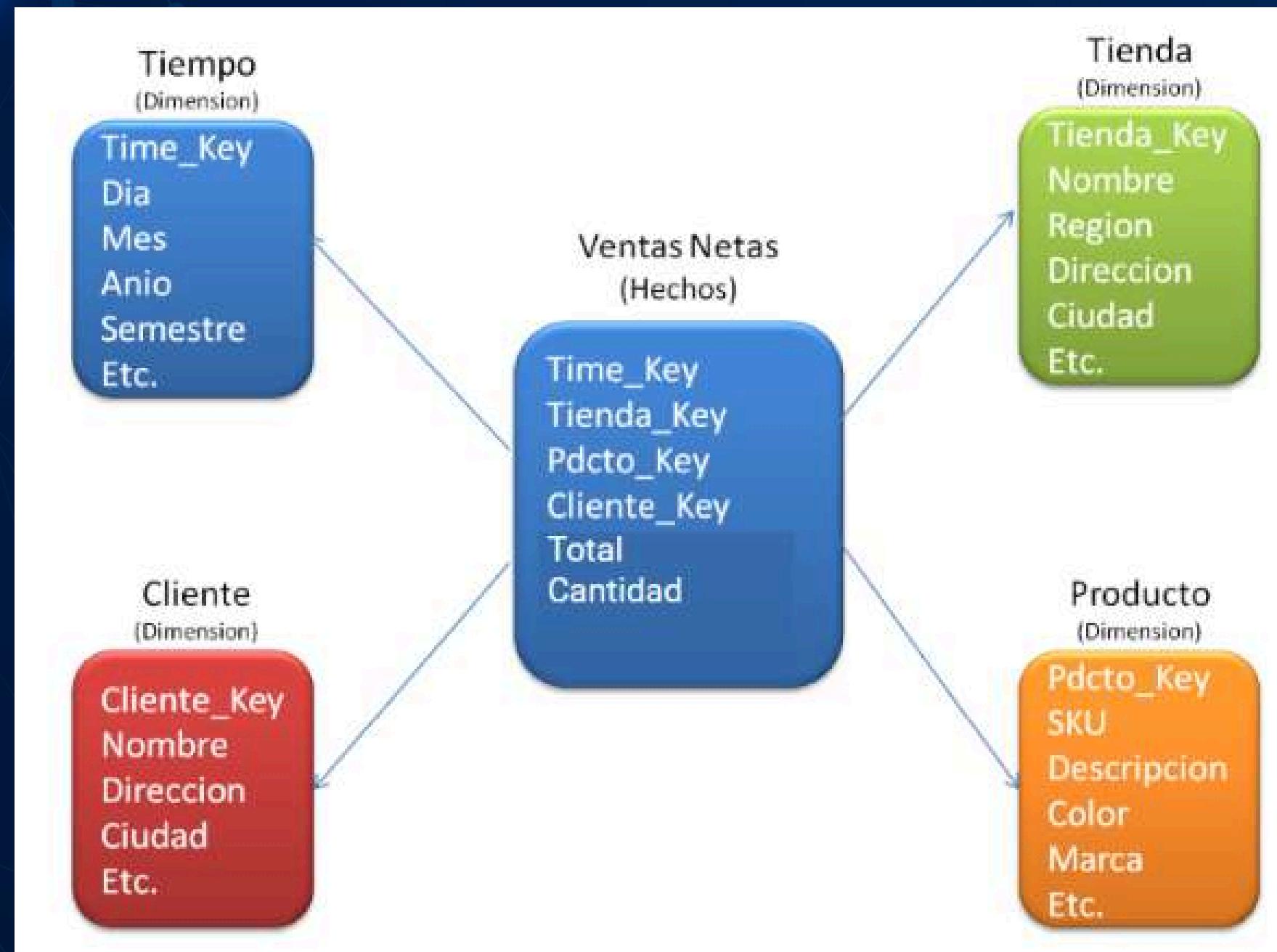
Representación gráfica del modelo Estrella



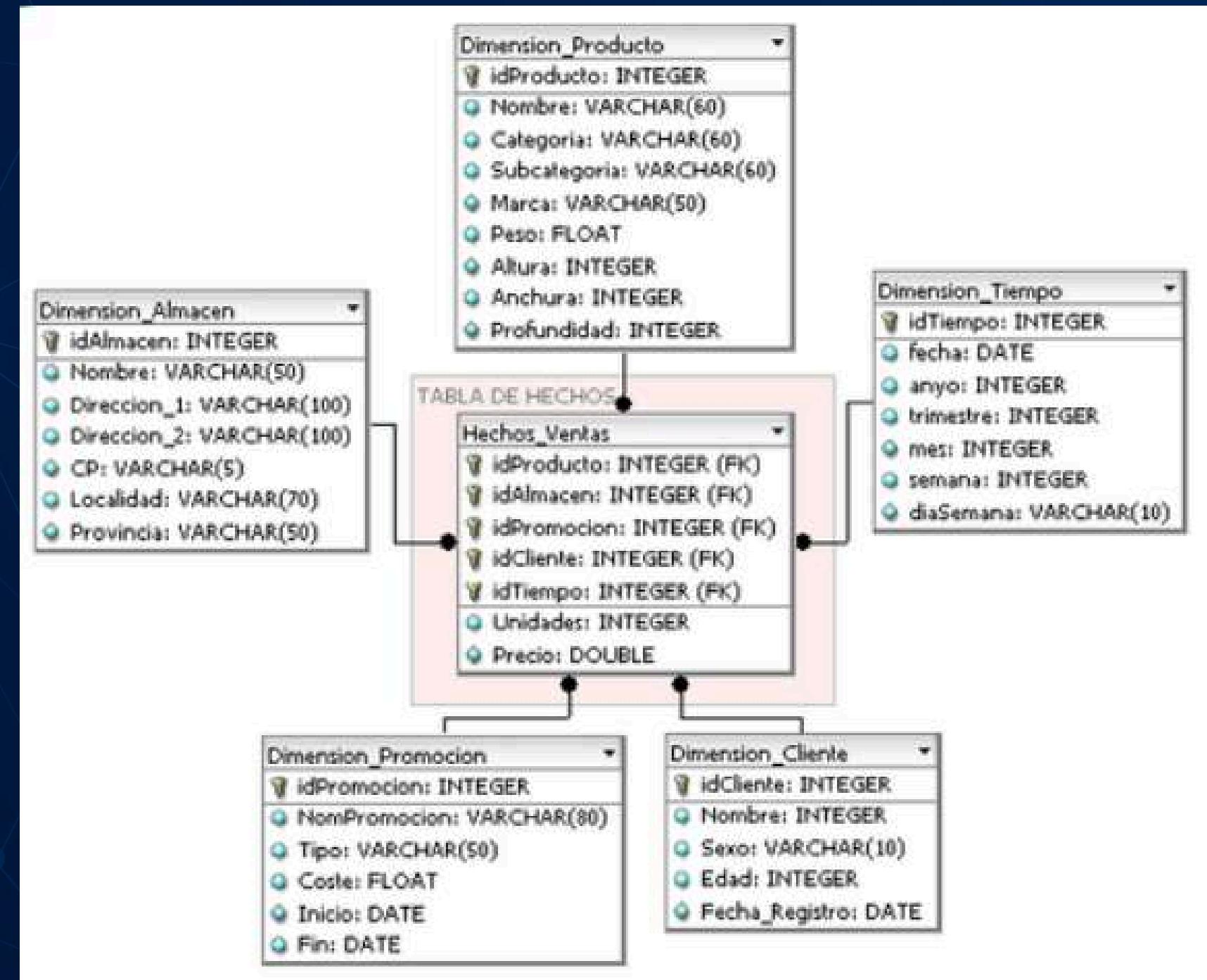
Representación gráfica del modelo Estrella



Representación gráfica del modelo Estrella



Representación gráfica del modelo Estrella



Modelo Copo de Nieve

- Es una variación o desviación de un modelo estrella.
- En este modelo la tabla de hechos deja de ser la única relacionada con otras tablas ya que existen otras tablas que se relacionan con las dimensiones.
- Puede implementarse luego de haber desarrollado un Modelo Estrella.



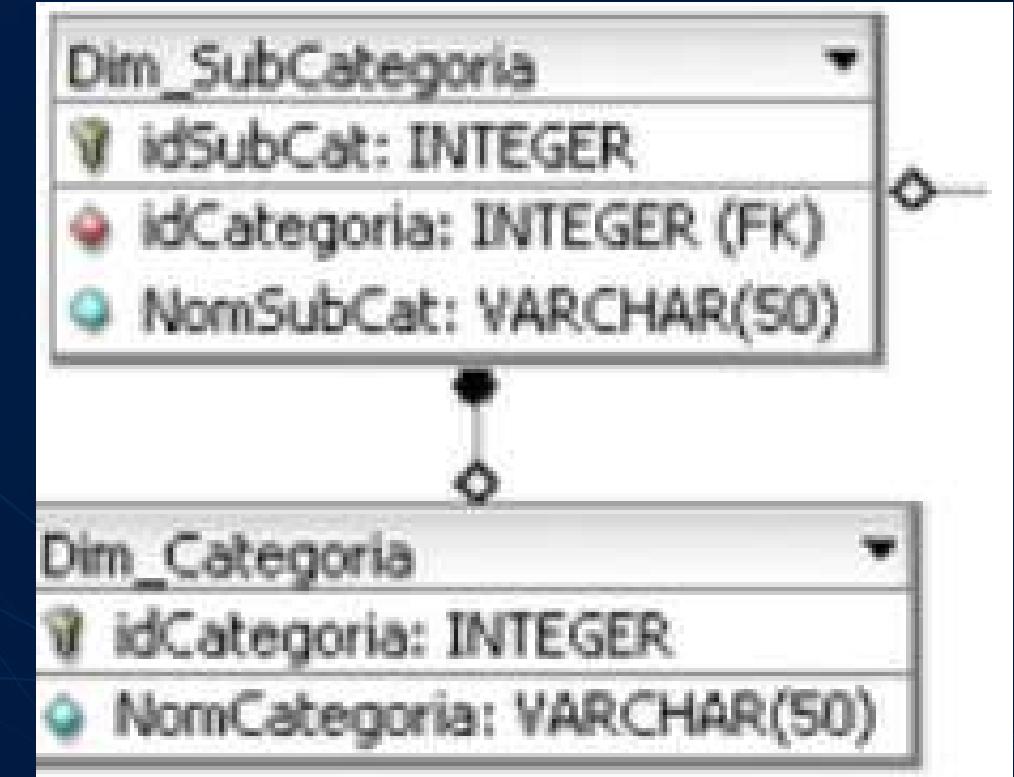
Modelo Copo de Nieve

- Existen tablas que no tienen relación directa con la tabla de hechos.
- Este modelo fue creado para facilitar el mantenimiento de las dimensiones.
- La extracción de datos es más difícil y vuelve la tarea de mantener el modelo un poco más compleja.



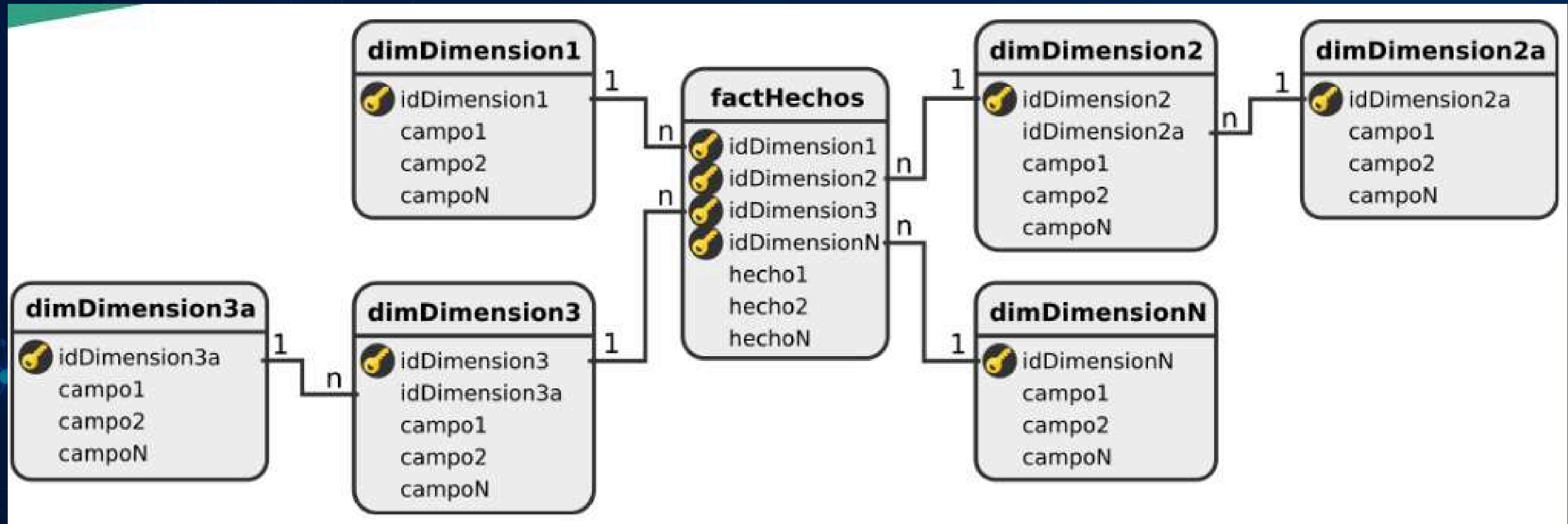
Modelo Copo de Nieve

- Su finalidad es normalizar las tablas y así reducir el espacio de almacenamiento al eliminar la redundancia de datos.
- Este modelo puede poseer tablas de dimensiones organizadas en jerarquía.

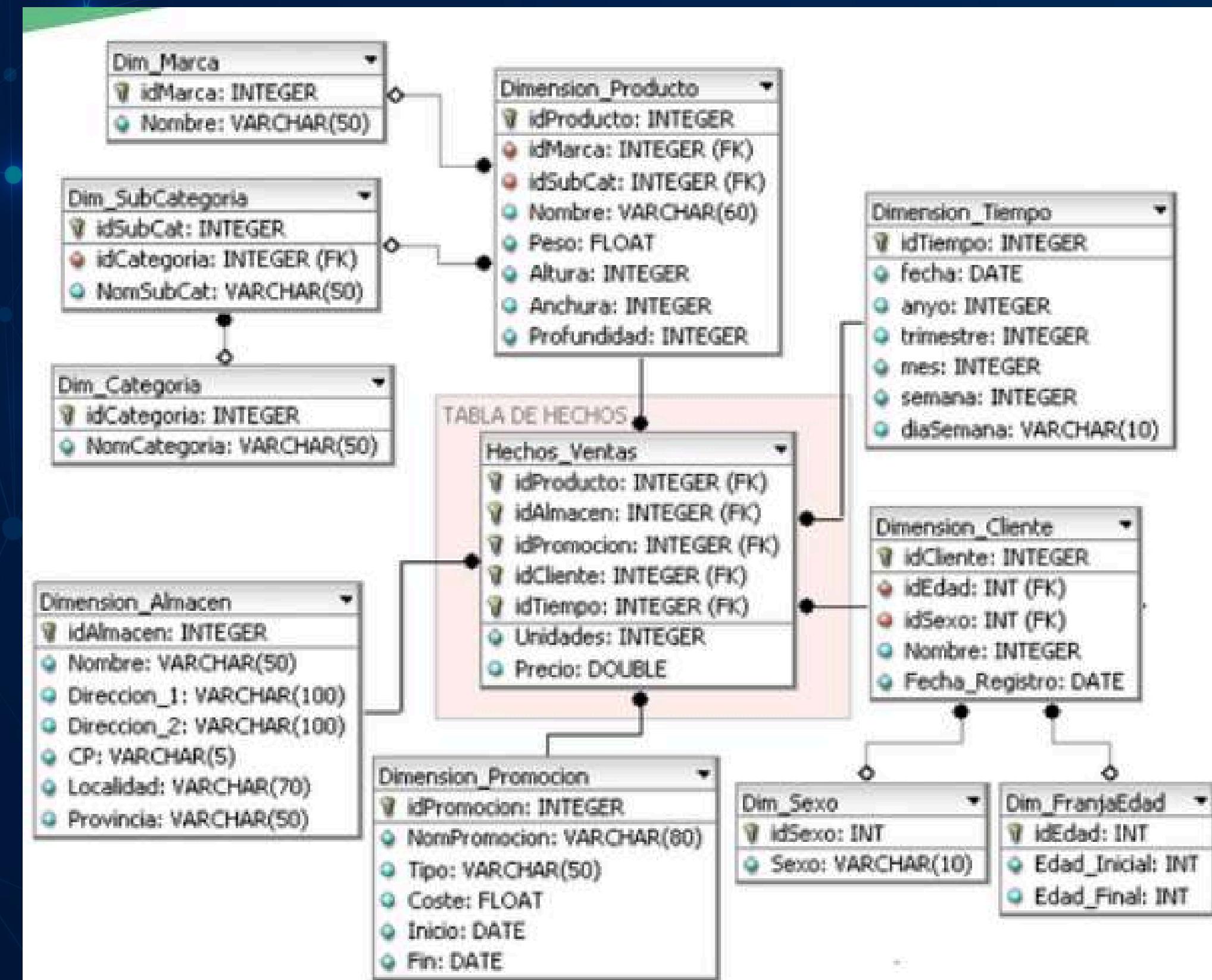


Representación gráfica del modelo

Copo de Nieve



Representación gráfica del modelo Copo de Nieve





Modelo Constelación

- Está compuesto por una serie de Esquemas en Estrella.
- Posee lo siguiente:
- Una tabla de Hechos **principal**.
- Una o más tabla de Hechos **Auxiliares**, dichas tablas están relacionadas con sus respectivas tablas de Dimensiones.



Modelo Constelación

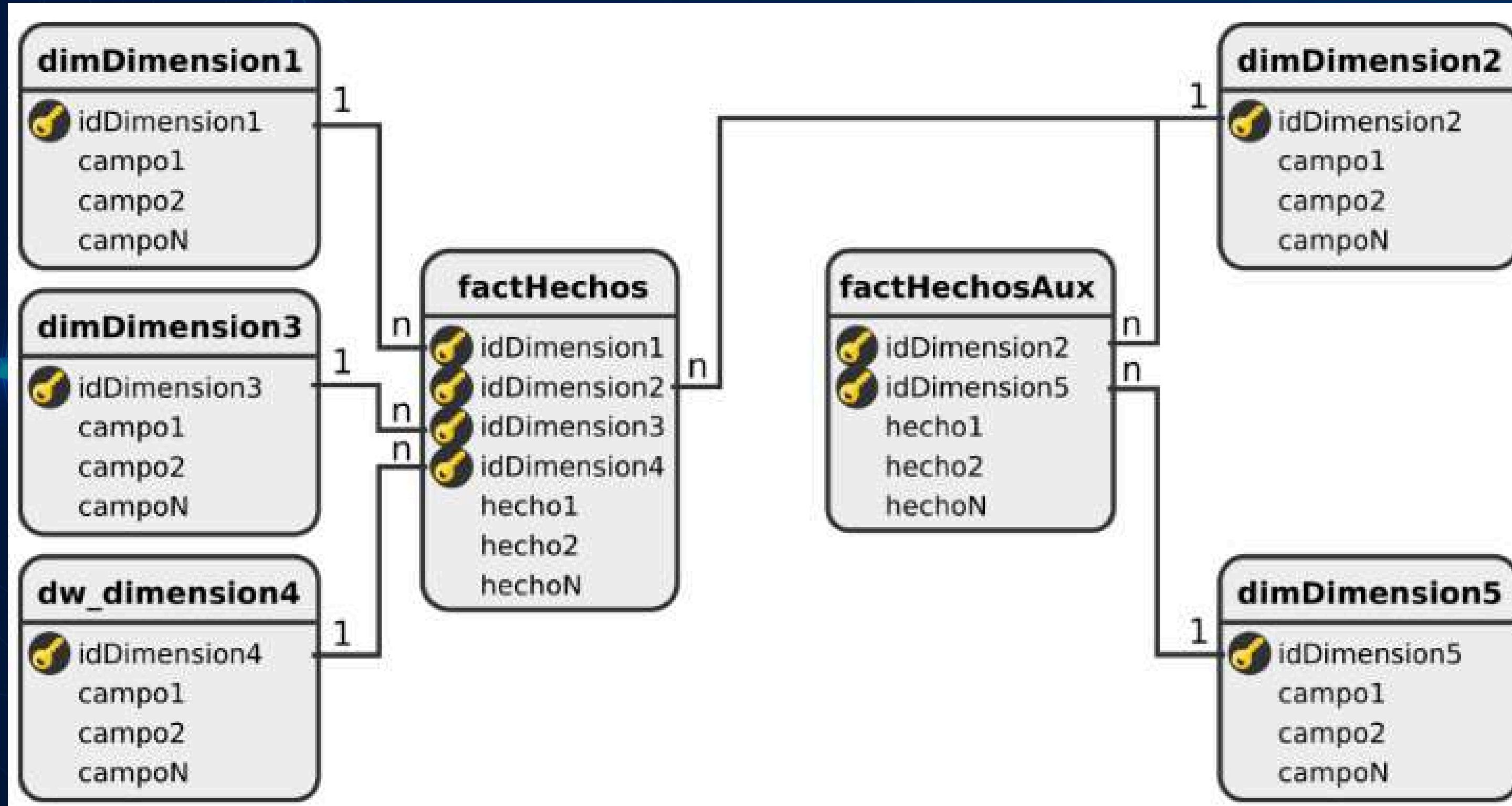
- Las tablas de Hechos Auxiliares pueden vincularse con solo algunas de las tablas de Dimensiones asignadas a la tabla de Hechos Principal, y también pueden hacerlo con nuevas tablas de Dimensiones que se necesiten.



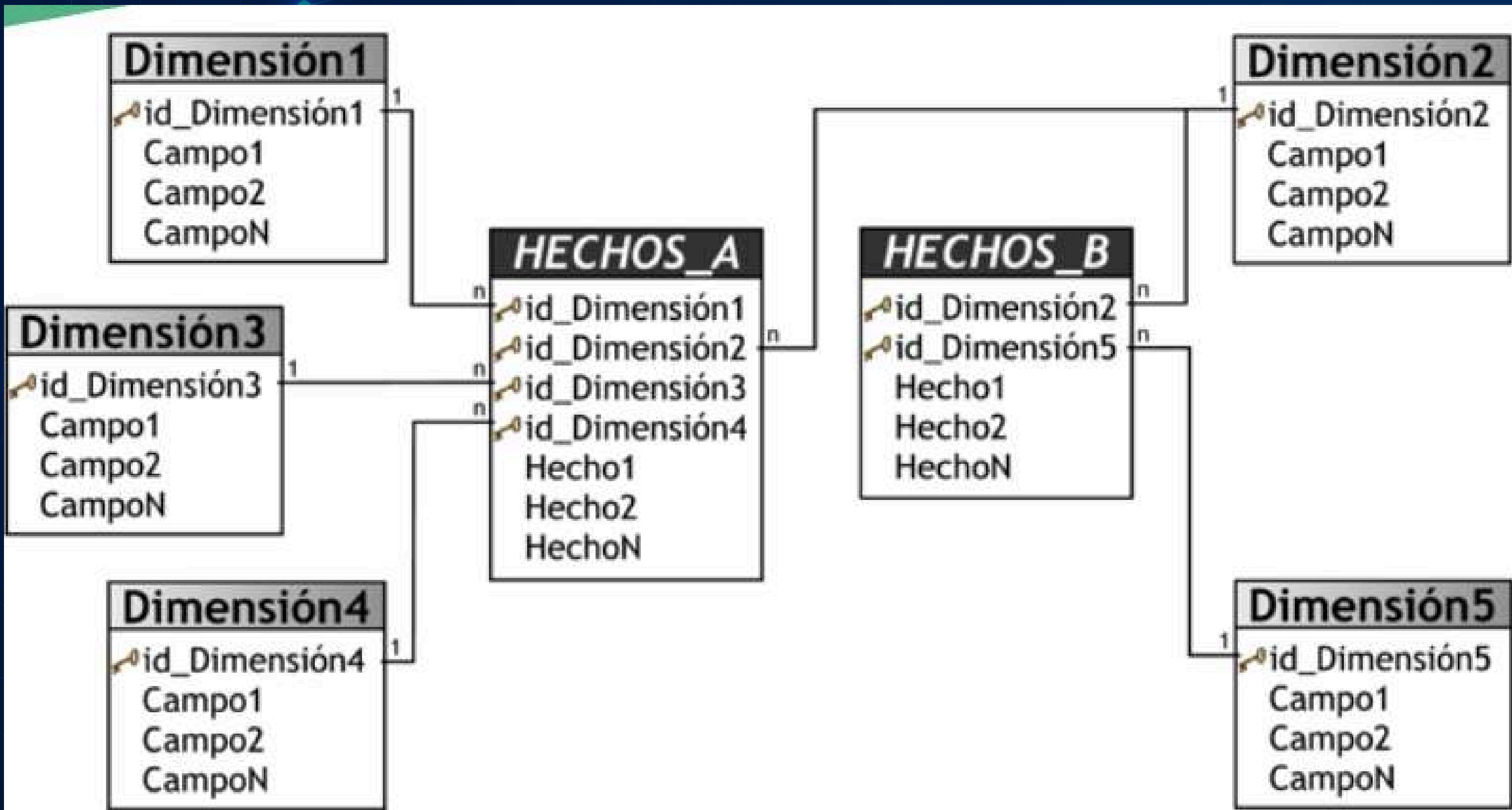
Modelo Constelación

- No es necesario pero se puede dar el caso que las diferentes tablas de Hechos comparten las mismas tablas de Dimensiones.
- Su capacidad analitica es mayor debido a que permite tener más de una tabla de hechos.
- Contribuye a reutilizar tablas de Dimensiones, ya que una misma tabla de Dimensión puede utilizarse para varias tablas de Hechos.

Representación gráfica del modelo Constelación



Representación gráfica del modelo Constelación



Modelo Estrella vs Modelo Copo de Nieve

	Estrella	Copo de nieve
Tablas	Contiene una tabla de hechos rodeada de varias dimensiones.	Contiene una tabla de hechos rodeada de varias dimensiones, que a su vez pueden estar rodeadas de varias dimensiones.
Redundancia	Alta redundancia	Poca redundancia
Facilidad de Uso	Fácil uso.	Difícil de entender, uso mas complicado.

Modelo Estrella vs Modelo Copo de Nieve

	Estrella	Copo de nieve
Joins	Con un solo join es posible relacionar la tabla de hechos y la de dimensiones.	Requiere múltiples joins para hacer los análisis.
Rendimiento de Query's	Ejecuciones más rápidas	Ejecuciones más complejas, debido a cruces.
Estructura	Descentralizada.	Normalizada

Modelo Estrella vs Modelo Copo de Nieve

	Estrella	Copo de nieve
Diseño de bases de datos	Simple	Complejo
Cuando Usarlo	Cuando las tablas de dimensión tiene pocas filas	Cuando las tablas de dimensión tienen un tamaño bastante elevado.

Resumiendo



Características - Modelo Estrella



- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

Características - Modelo Copo de Nieve



- Posee mayor complejidad en su estructura.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.

Características - Modelo Copo de Nieve



- Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

Características - Modelo Constelación



- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.
- No es soportado por todas las herramientas de consulta y análisis.

**Thank You
Dudas?**

Tarea 2

- Instalar SQL Server 2014 o Superior
- Adjuntar un Screenshot con la vista de SQL Server ya instalado en su computadora con Su nombre y carne en Pantalla

Entrega: 10/08/2024 23:59

Practica 1

Laboratorio

Seminario de Sistemas 2 Sección “A”



SSIS

SSIS (SQL Server Integration Services) es una plataforma de integración de datos desarrollada por Microsoft como parte de su suite de SQL Server. Fue introducido en 2005 como reemplazo de Data Transformation Services (DTS), la herramienta de integración de datos anterior de SQL Server.



SSIS

- Fue diseñado para realizar extracción, transformación y carga de datos (ETL) entre distintas fuentes de datos como bases de datos relacionales, archivos planos, servicios web, etc.
- Simplifica y automatiza tareas complejas de integración de datos, como limpieza de datos, agregación, combinación de datos de múltiples orígenes, etc.
- Permite embeber código .NET personalizado para extender funcionalidades.

Componentes



Apoyo

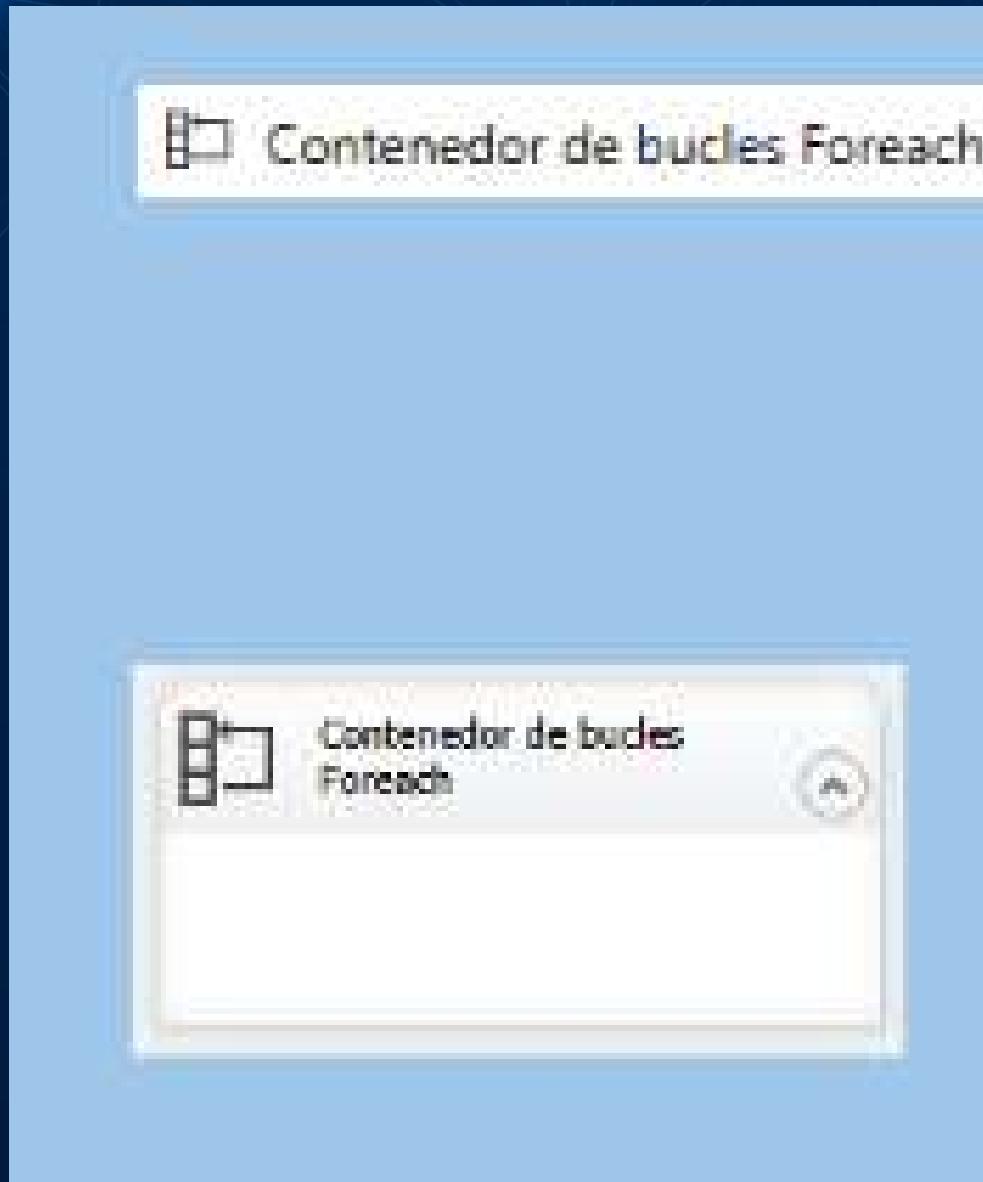


Permite ejecutar consultas y comandos SQL contra bases de datos dentro de un paquete SSIS.



Es uno de los componentes centrales en SSIS que permite mover y transformar datos dentro de un paquete

Apoyo



Permite iterar y repetir la ejecución de un flujo de control sobre una enumeración especificada



Permite incorporar código personalizado en lenguajes .NET para extender las funcionalidades en un paquete SSIS

Extracción

Origen de ADO.NET

Origen de ADO.NET

ADO.NET permite de manera flexible y optimizada extraer datos desde bases de datos relacionales en SSIS utilizando consultas SQL parametrizadas.

Origen de OLE DB

Origen de OLE DB

El origen OLE DB permite de forma sencilla conectarse y extraer datos de diversos orígenes en SSIS

Extracción



Origen de Excel



Origen de Excel

El origen de Excel es un componente de flujo de datos en SSIS que permite extraer datos desde archivos de Excel.



Origen de archivo plano



Origen de archivo plano

El origen de Archivo Plano es el componente indicado en SSIS para leer y extraer datos desde archivos de texto plano de diferentes formatos.



Transformación

Columna Derivada

Este crea una nueva columna dependiendo de la expresión que se evalúe, puede ser desde una concatenación, una suma, hasta una expresión condicional, apoyándonos con las opciones que nos dan en la parte superior derecha.

The screenshot shows the 'Derived Column' transformation configuration window. At the top, there's a title bar with the text 'Editor de transformación Columna derivada'. Below it, a message says: 'Especifique las expresiones utilizadas para crear nuevos valores de columna e indique si los valores actualizan las columnas existentes o bien llenan columnas nuevas.' On the left, there's a tree view with categories like 'Variables y parámetros' and 'Columnas'. On the right, there's a large list of mathematical and logical operators. In the center, there's a table with one row:

Nombre de columna d...	Columna derivada	Expresión	Tipo de datos
NuevaColumna	<agregar como colum...	col1+col2	

At the bottom, a yellow warning message states: 'El componente no tiene ninguna columna de entrada disponible y no se puede configurar correctamente. Para corregir el problema, conecte una salida con columnas disponibles de otro componente a la entrada de este componente.' There are buttons at the bottom: 'Configurar la salida de errores...', 'Aceptar', 'Cancelar', and 'Ayuda'.



Transformación

División condicional

Este componente tendrá dos salidas una que cumpla con la condición que nosotros configuremos y otra en la que no se cumpla. Normalmente se usa para la verificación de datos.

The screenshot shows the 'Editor de transformación División condicional' (Conditional Split Transformation Editor) window. At the top, there's a title bar with the component icon and the text 'División condicional'. Below it is a toolbar with icons for 'Nuevo', 'Abrir', 'Guardar', 'Copiar', 'Pegar', 'Borrar', 'Copiar formato', 'Formato', 'Ayuda', and 'Ayuda en línea'. The main area is divided into several sections:

- Panel izquierdo:** Includes sections for 'Variables y parámetros' and 'Columnas'.
- Panel central:** A large tree view showing various functions categorized under 'Funciones matemáticas', 'Funciones de cadena', 'Funciones de fecha y hora', and 'Funciones NULL'. One function, 'ISNULL', is highlighted with a tooltip: 'Devuelve un valor nulo de un tipo de datos solicitado.'
- Table section:** A table titled 'Especifique las condiciones utilizadas para dirigir filas de entrada a salidas específicas. Si una fila de entrada no coincide con ninguna condición, se dirigirá a una salida predeterminada.' It has columns 'Orden', 'Nombre de salida', and 'Condición'. It contains one row:

Orden	Nombre de salida	Condición
1	CarneNull	ISNULL(Der_Carne)
- Buttons:** At the bottom right are buttons for 'Aceptar' (Accept), 'Cancelar' (Cancel), and 'Ayuda' (Help).



Ordenar

Este componente ordenara el conjunto de datos que tengamos de la forma que le indiquemos, en el podremos quitar los elementos repetidos eligiendo las columnas que se quieren verificar.

Transformación

The screenshot shows the 'Editor de transformación Ordenar' (Sort Transformation Editor). The title bar says 'Editor de transformación Ordenar'. The main area has a message: 'Especifique las columnas que se ordenarán y establezca el tipo y el criterio de ordenación. Las columnas no seleccionadas se copiarán sin ninguna modificación.' (Specify the columns to be sorted and set the type and sorting criteria. Unselected columns will be copied without any modification.)

On the left, there's a list titled 'Columnas de entrada disponibles' (Available Input Columns) with the following data:

	Nombre	Paso a través
<input type="checkbox"/>	Carne	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Nombre	<input checked="" type="checkbox"/>
<input type="checkbox"/>	LlevaLab	<input checked="" type="checkbox"/>
<input type="checkbox"/>	PossibleNota	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Der_Carne	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	CARNE	<input type="checkbox"/>
<input type="checkbox"/>	NOTA	<input checked="" type="checkbox"/>

Below this is a table for defining the sort operation:

Columna de entrada	Alias de salida	Tipo de orden	Criterio de or...	Marcadores de comparación
CARNE	CARNE	ascendente	1	

Transformación



Mezclar

Este componente nos permitirá combinar dos orígenes de datos, convirtiendo estos en un solo flujo de datos.

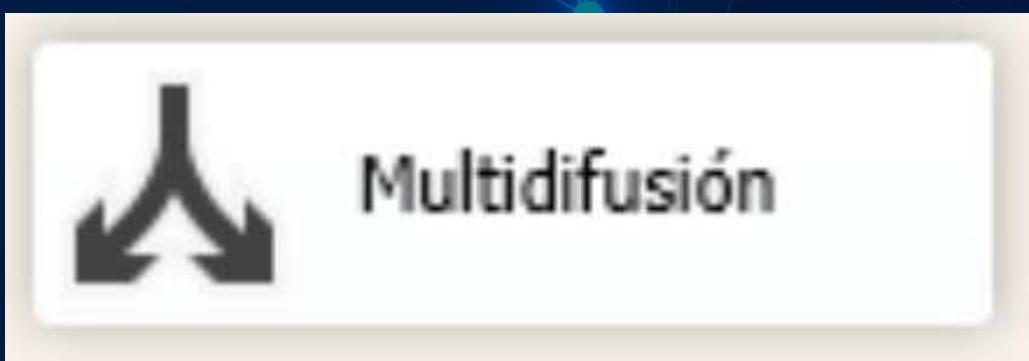


Transformación



Multidifusión

Este componente tiene varias salidas a partir de un conjunto de datos distribuye a estas N salidas creando “copias” de estos datos.



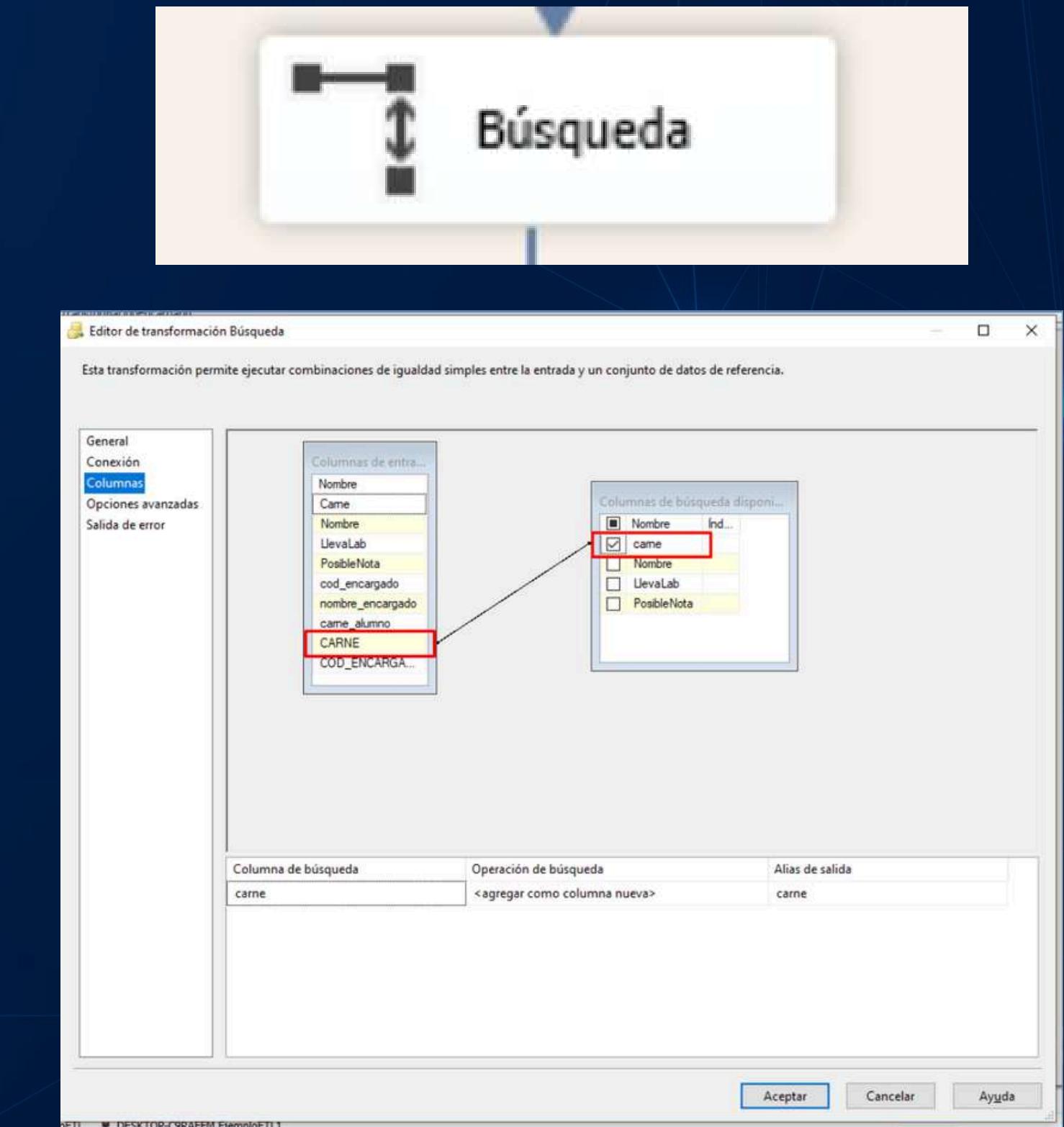
Transformación



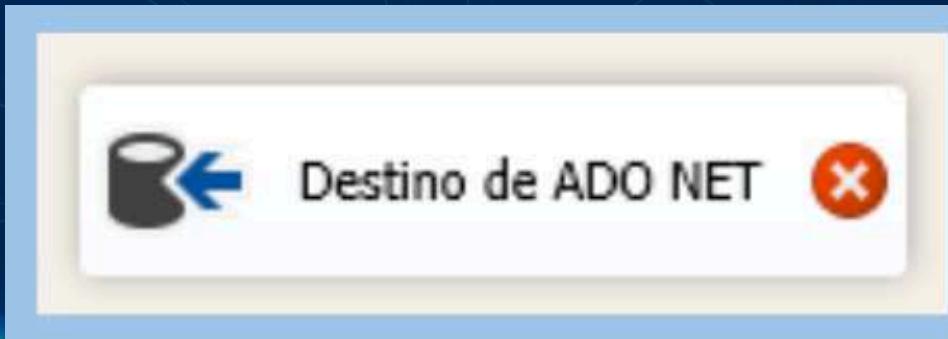
Búsqueda

Componente que nos ayudará a buscar registros desde una tabla para verificar si existen o hacer otro tipo de validaciones con ellos.

* Se recomienda que se use con orígenes y destinos OLE DB.

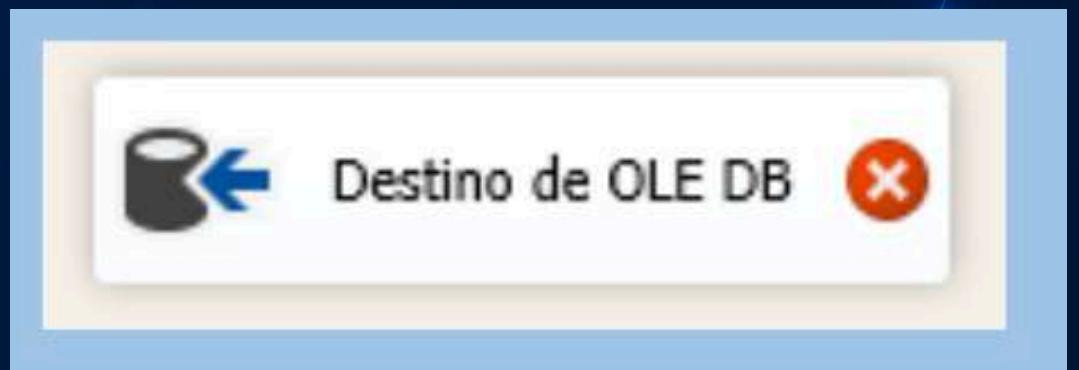


Carga

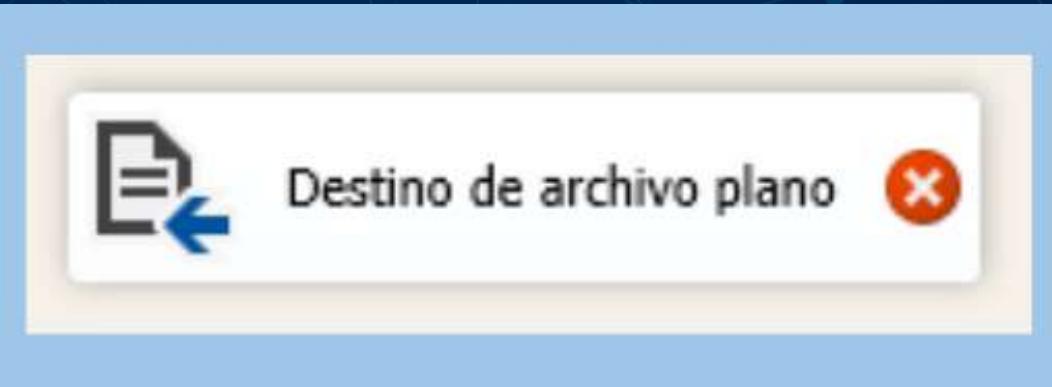


ADO.NET es un componente de flujo de datos en SSIS que permite cargar datos en orígenes de datos relacionales usando ADO.NET.

El componente Destino de OLE DB en SSIS es un destino de datos que permite cargar información en orígenes de datos OLE DB como SQL Server, Excel, Oracle, DB2, etc.

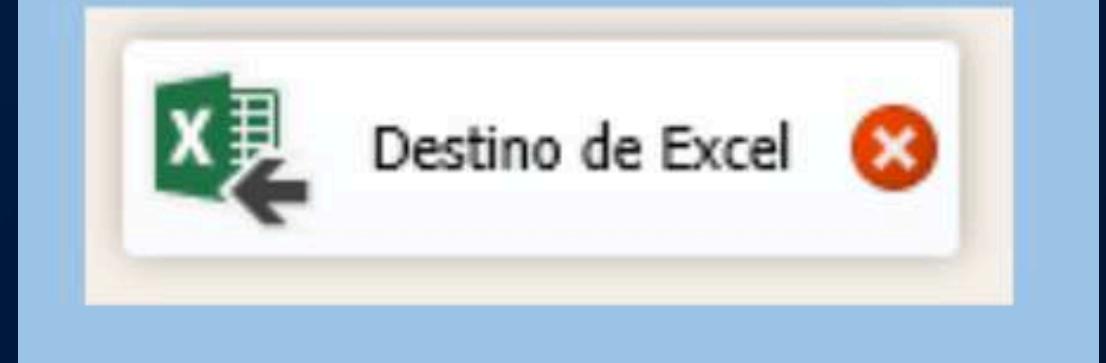


Carga



El componente Destino de Excel en SSIS permite escribir datos desde un flujo de datos a un archivo de Excel.

el destino de Archivo Plano brinda un método sencillo en SSIS para escribir datos a archivos planos de diferentes formatos.



**Thank You
Dudas?**

