

U3. DESARROLLO Y GESTIÓN DE BD DECISIONALES /DATA WAREHOUSING

Clase 3



Cátedra: BASE DE DATOS II

ESQUEMAS ESTRELLA, COPO DE NIEVE Y CONSTELACIÓN

Objetivos:

- Diseñar y comparar los principales modelos dimensionales trabajados en la unidad

BASE DE DATOS MULTIDIMENSIONAL

Un **Data Warehouse** es una base de datos con estructura multidimensional, lo cual se traduce en una forma específica de almacenamiento en la cual se definen dos estructuras principales:

- tablas de Hechos y
- tablas de Dimensiones.

La utilización de tablas de Hechos y Dimensiones, facilita la creación de estructuras de datos (Cubos Multidimensionales, Business Models, etc.) y posibilita que las consultas al SGBD sean respondidas con mucha performance.

BASE DE DATOS MULTIDIMENSIONAL

- **Hecho:** qué fenómeno se quiere analizar.
- **Medidas:** qué se quiere medir.
- **Dimensiones:** desde qué perspectivas se analiza.
- **Granularidad:** qué representa una fila de la tabla de hechos.
- **Jerarquías** dentro de las dimensiones.
- Ejemplos de dimensiones: **tiempo**, geografía, establecimiento, paciente, profesional.

BASE DE DATOS MULTIDIMENSIONAL

EJEMPLO: El Ministerio de Salud quiere analizar las consultas realizadas en los CAPS (Centro de Atención Primaria de Salud)

Quiere conocer:

- cantidad de consultas;
- consultas por CAPS;
- consultas por especialidad;
- consultas por grupo etario;
- evolución de las consultas en el tiempo.

HECHO: Consulta

MEDIDA: Cantidad de consultas

DIMENSIONES: CAPS, Especialidad, Grupo etario, Tiempo

GRANULARIDAD: Una consulta

JERARQUÍA (DE DIMENSIONES): Tiempo (año, semestre, trimestre, mes)

BASE DE DATOS MULTIDIMENSIONAL

Hasta ahora teníamos conceptos.

Ahora necesitamos construir:

TABLAS

Tenemos dos tipos principales:

TABLA DE HECHOS

Contiene el fenómeno que queremos analizar y sus medidas.

TABLAS DE DIMENSIONES

Contienen los aspectos desde los cuales analizaremos ese fenómeno.

TABLA DE HECHOS

FACT_CONSULTA

id_fecha	id_caps	id_especialidad	id_grupo_edad	cantidad	
1	10	3	2	1	
1	10	5	3	1	
1	12	3	4	1	
2	10	3	2	1	

Los primeros cuatro campos son claves que nos permiten relacionar dimensiones. Cantidad es la medida.

Las Tablas de Hechos contienen los Hechos que serán utilizados por l@s usuari@s del DW para analizar y responder preguntas de negocio.

Los Hechos (o Dato agregado): son datos cuantitativos, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de Dimensiones.

El registro del Hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de Dimensiones relacionadas a éste.

TABLAS DE DIMENSIONES

DIM_TIEMPO

id_fecha	día	mes	año
1	1	Agosto	2026
2	2	Agosto	2026

DIM_CAPS

id_caps	CAPS	localidad	departamento	zona_sanitaria
10	CAPS A	La Rioja	Capital	I
12	CAPS B	Chilecito	Chilecito	III

DIM_ESPECIALIDAD

id_especialidad	especialidad
3	Clínica
5	Pediatría

Las tablas de Dimensiones: contienen datos cualitativos y representan los aspectos de interés, mediante los cuales l@s usuari@s podrán filtrar y manipular los Hechos almacenados en las tablas de Hechos.

TABLAS DE DIMENSIONES

Las tablas de Dimensión contienen los siguientes tipos de campos:

- Clave principal.
- Claves foráneas (solo para esquemas copo de nieve y constelación).
- Datos de referencia primarios: datos que identifican la Dimensión. Por ejemplo: nombre del cliente.
- Datos de referencia secundarios: datos que complementan la descripción de la Dimensión. Por ejemplo: e-mail del cliente, celular del cliente, etc. Estos datos no son significativos para tomar decisiones, pero son potencialmente valiosos para implementarla.

Usualmente la cantidad de tablas de Dimensiones, aplicadas a un tema de interés en particular, varían entre tres y quince.

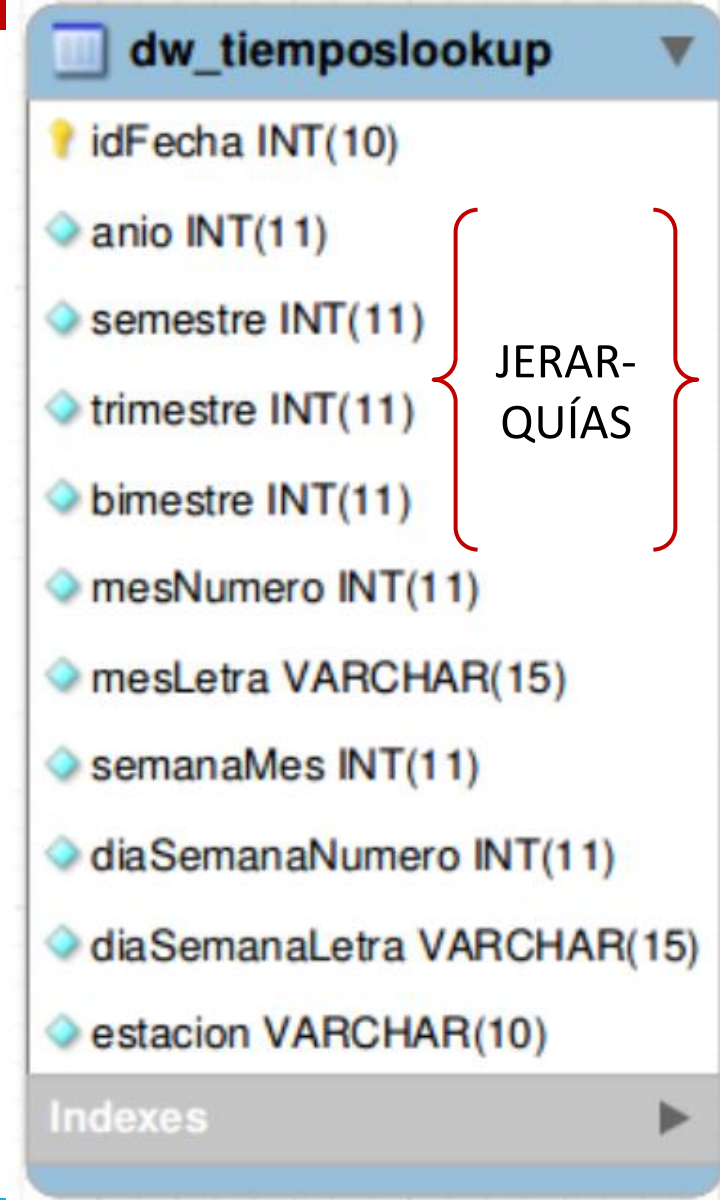
Es recomendable que la clave principal de las tablas de Dimensiones sea independiente de las claves de los Data Sources, ya que, por ejemplo, si estos últimos son recodificados, el DW quedaría inconsistente. Estas claves independientes se denominan Claves Subrogadas.

TABLA DE DIMENSIÓN TIEMPO

En un DW, la creación y el mantenimiento de una tabla de Dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se esté analizando.

La Dimensión Tiempo NO es sola una secuencia cronológica representada de forma numérica, sino que mantiene niveles jerárquicos especiales que son representativos de las actividades de la organización. Esto se debe a que l@s usuari@s podrán por ejemplo analizar las HECHOS teniendo en cuenta el día de la semana en que se produjeron, quincena, mes, trimestre, semestre, año, estación, etc.

La forma de diagramar la Dimensión Tiempo es muy sencilla, se debe tomar el campo que indique la fecha en que sucedieron los Hechos y luego analizar dicha fecha para crear los campos requeridos. Ejemplo:



dw_tiemposlookup	
idFecha	INT(10)
anio	INT(11)
semestre	INT(11)
trimestre	INT(11)
bimestre	INT(11)
mesNumero	INT(11)
mesLetra	VARCHAR(15)
semanaMes	INT(11)
diaSemanaNumero	INT(11)
diaSemanaLetra	VARCHAR(15)
estacion	VARCHAR(10)

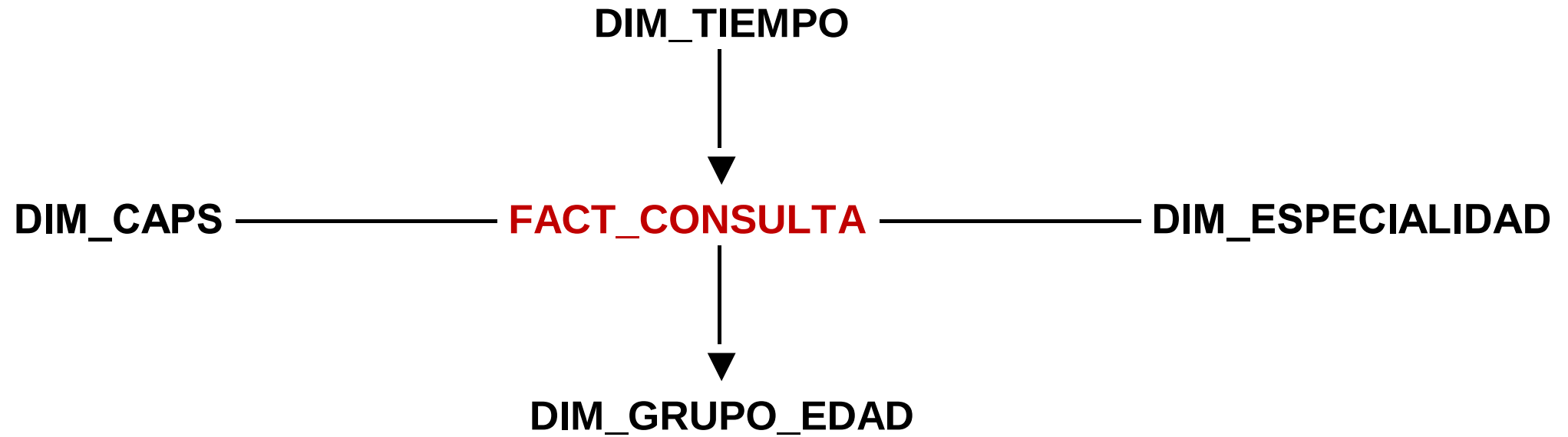
Indexes

TIPOS DE MODELOS LÓGICOS

Existen tres formas de modelar las tablas de Hechos y Dimensiones:

- **Esquema en estrella (Star Scheme).**
- **Esquema copo de nieve (Snowflake Scheme).**
- **Esquema constelación (Starflake Scheme).**

ESQUEMA EN ESTRELLA



¿Por qué estrella?

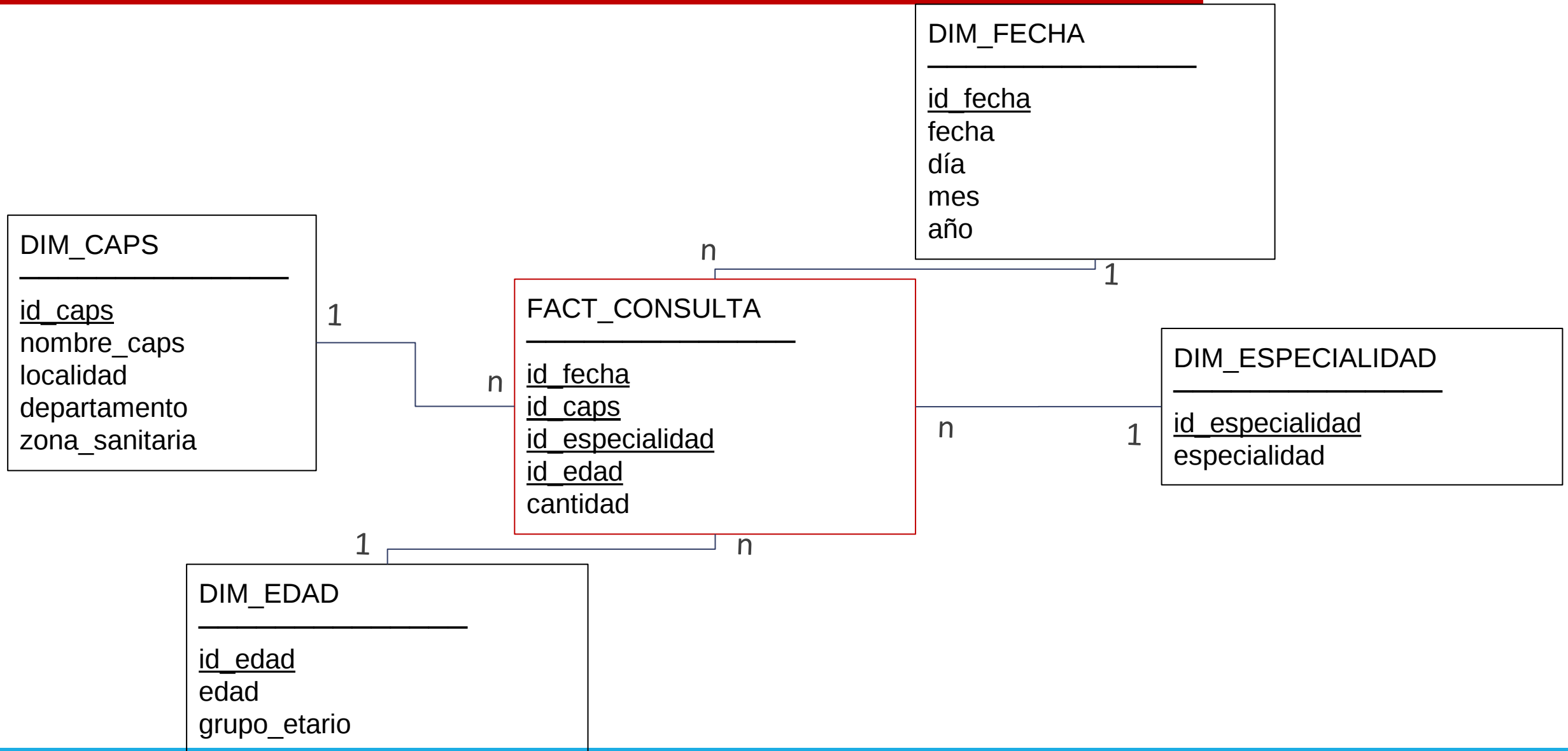
Porque tenemos:

una tabla de hechos en el centro

y

una o más dimensiones relacionadas con ella.

ESQUEMA EN ESTRELLA



ESQUEMA EN ESTRELLA

A partir de este modelo:

Podemos preguntar:

¿Cuántas consultas hubo?

¿Cuántas consultas hubo por CAPS?

¿Cuántas consultas hubo por especialidad?

¿Cuántas consultas hubo por grupo etario?

¿Cómo evolucionaron las consultas por mes?

¿Cuántas consultas de Pediatría hubo en el CAPS A durante agosto?

¿Cuántas consultas de Clínica hubo en la Zona sanitaria III?

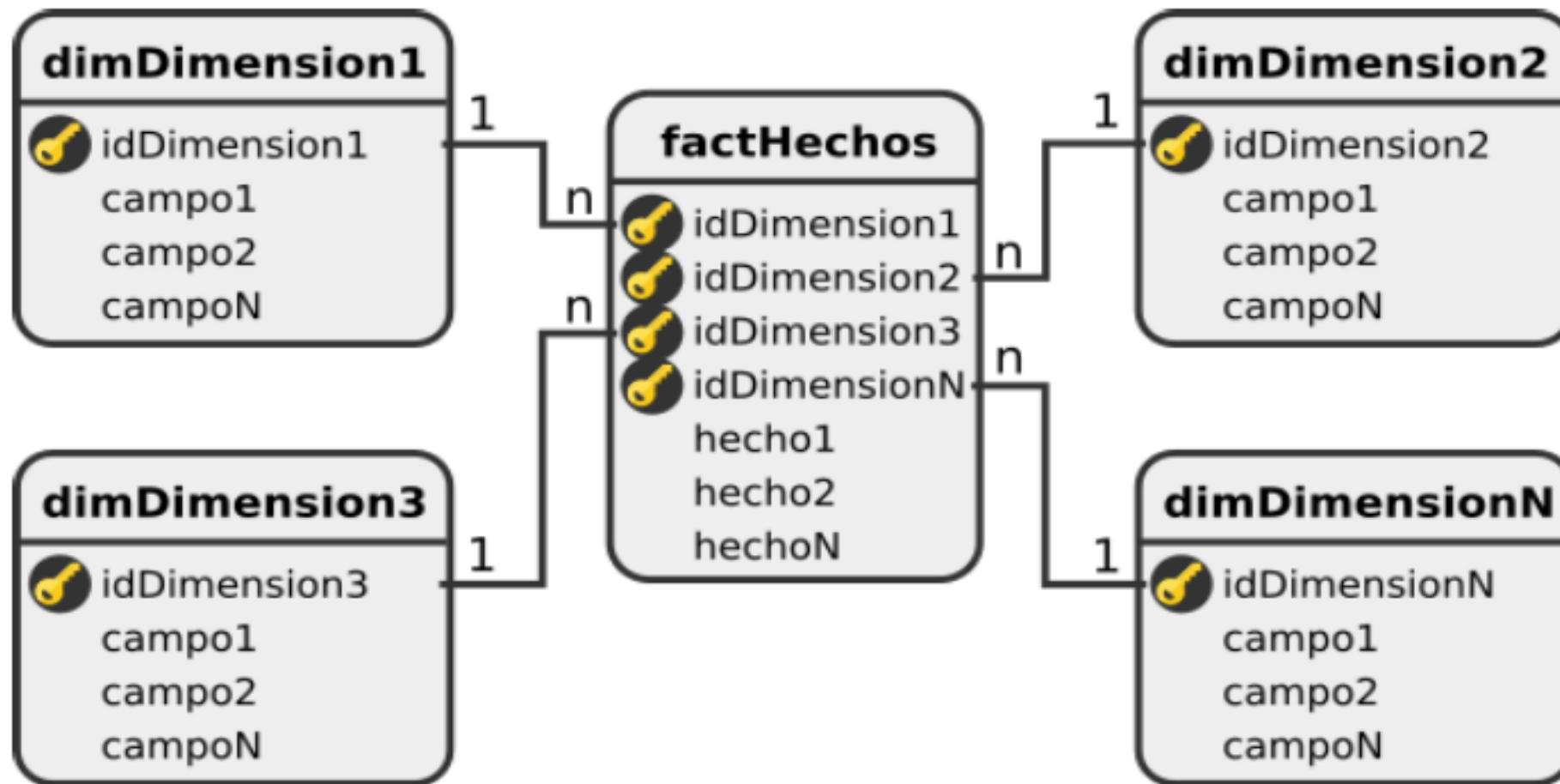
El modelo está construido para responder preguntas analíticas.

ESQUEMA EN ESTRELLA

El Esquema en Estrella (Star Scheme) está formado por:

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

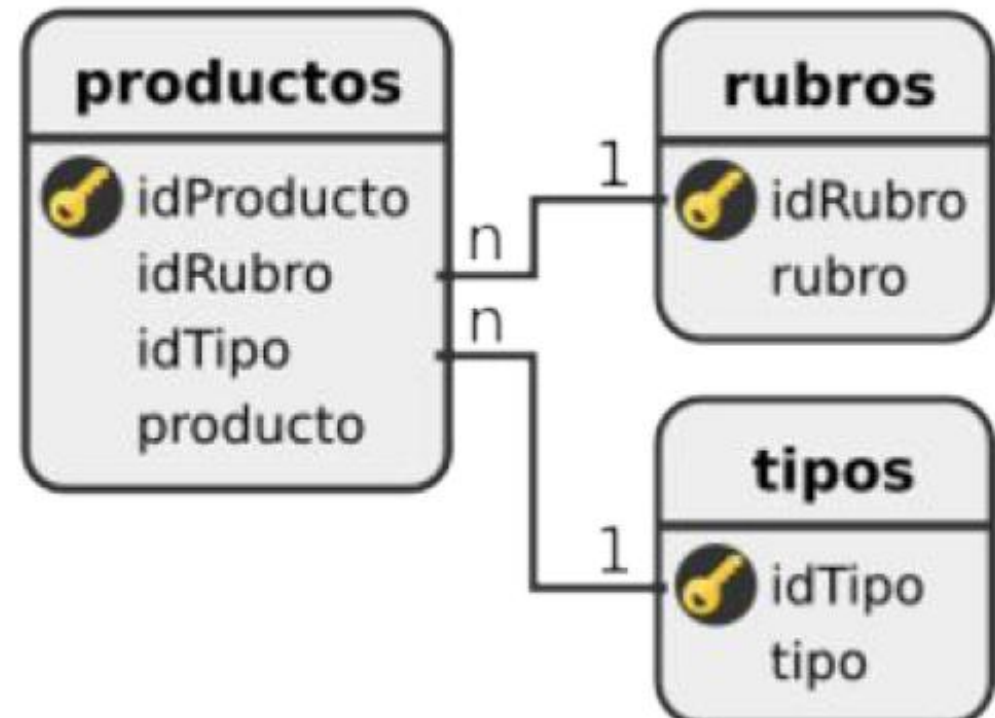
La siguiente figura representa un Esquema en Estrella estándar:



Desnormalizado



Normalizado



ESQUEMA COPO DE NIEVE – JERARQUÍAS

Jerarquías

Una Jerarquía (en este contexto) es una relación lógica del tipo padre-hijo entre las Dimensiones.

Al emplear Jerarquías podemos analizar los datos desde el nivel más general al más detallado y viceversa. Las Jerarquías manejan el nivel de agregación de los Hechos.

Por ejemplo, en Tiempo:

AÑO



MES



DÍA

En geografía:

PROVINCIA



DEPARTAMENTO

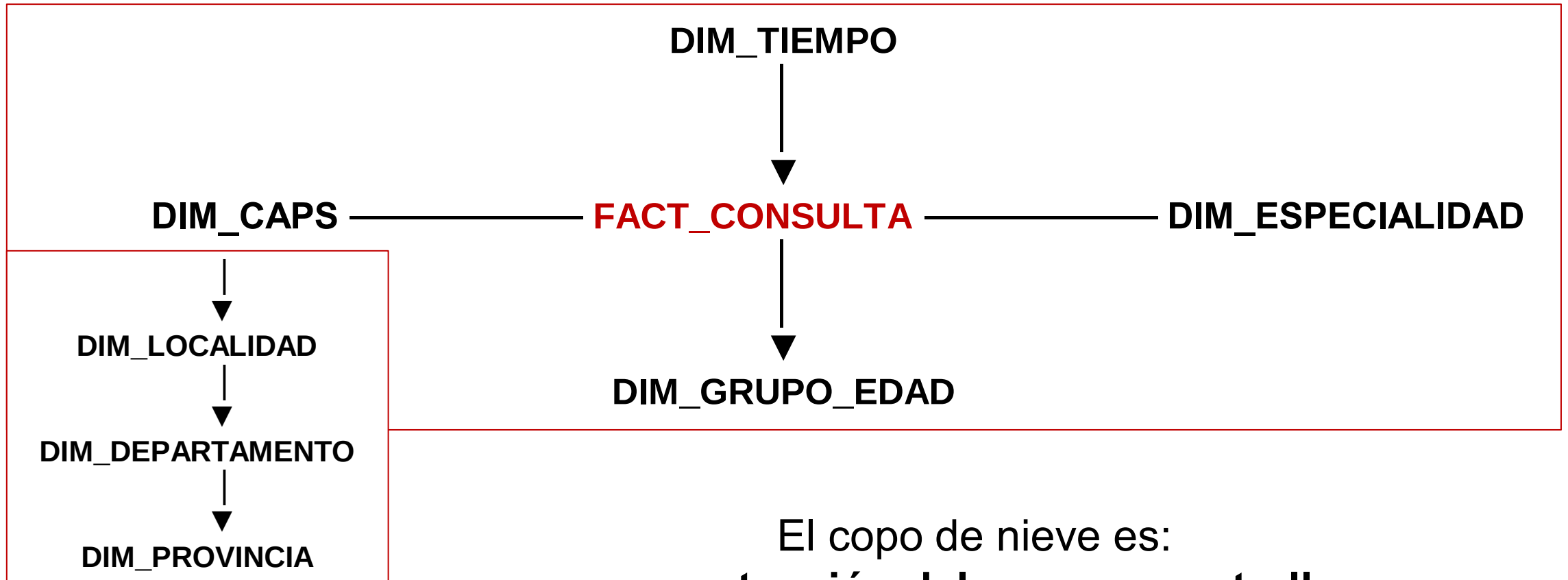


LOCALIDAD



CAPS

ESQUEMA EN COPO DE NIEVE

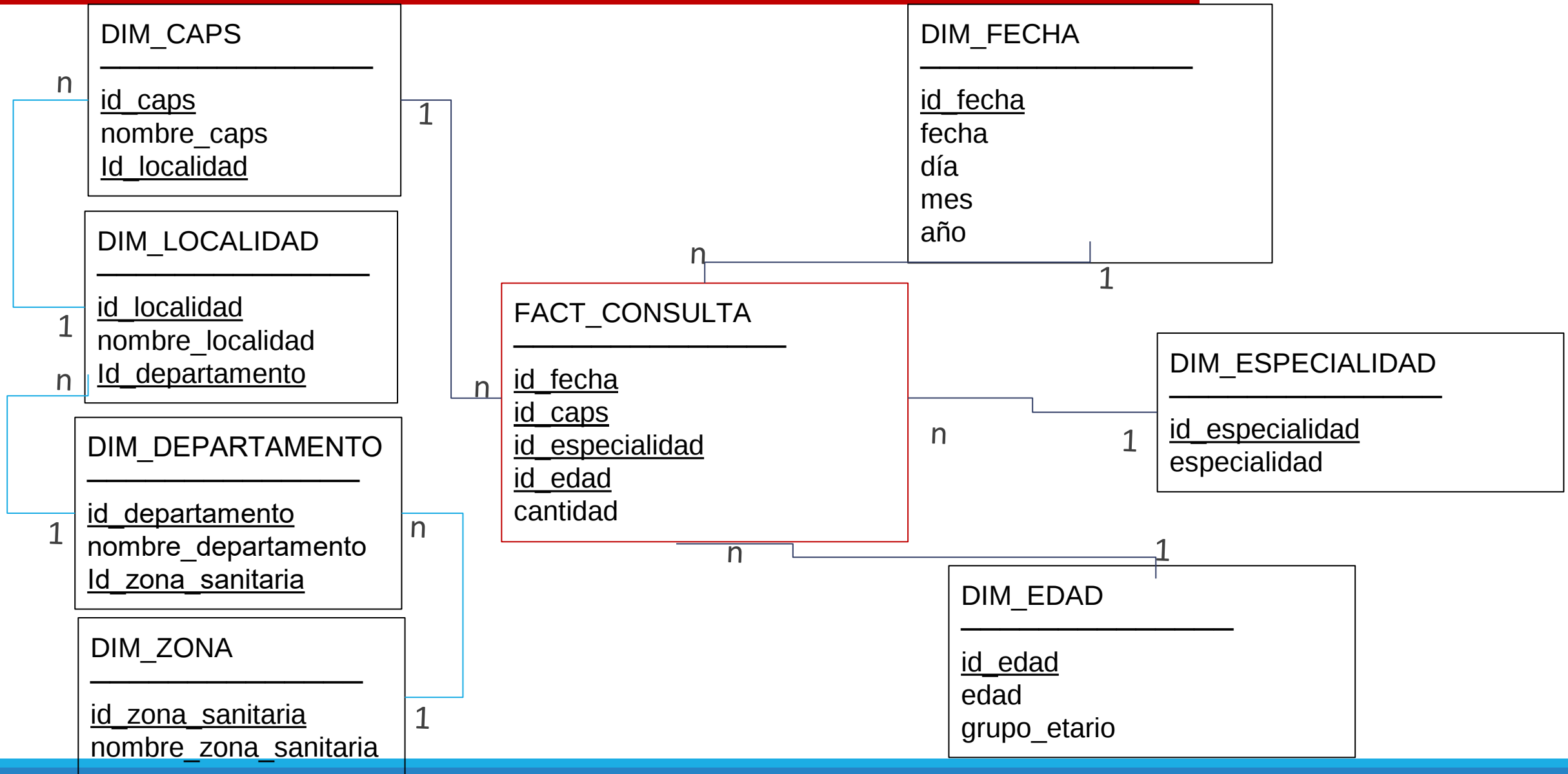


El copo de nieve es:

una extensión del esquema estrella.

La diferencia es que algunas dimensiones se separan en otras tablas relacionadas, normalizadas.

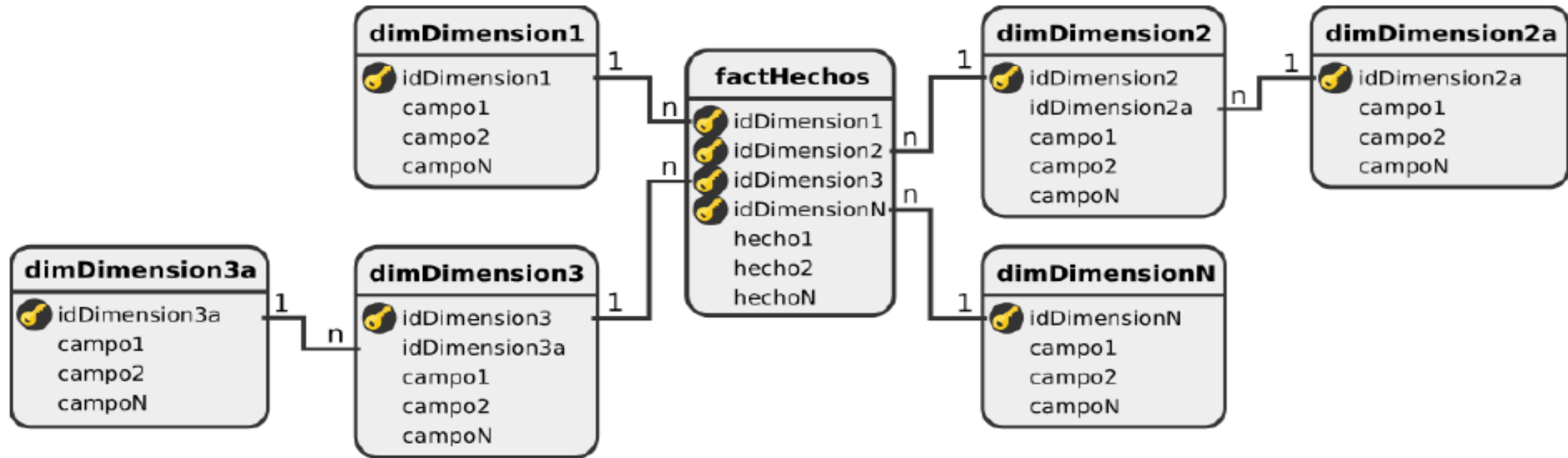
ESQUEMA EN COPO DE NIEVE



ESQUEMA COPO DE NIEVE

El Esquema Copo de Nieve (Snowflake Scheme) es:

- una extensión del modelo en Estrella, y
- se caracteriza por poseer tablas de Dimensiones organizadas en Jerarquías de Dimensiones.



Como se puede apreciar en la figura anterior:

- existe una tabla de Hechos (**factHechos**) que está relacionada con una o más tablas de Dimensiones, y
- algunas tablas de Dimensiones están relacionadas con otras tablas de Dimensiones.

ESQUEMA ESTRELLA VS. COPO DE NIEVE

★ Estrella	❄ Copo de nieve
Dimensiones más simples	Dimensiones divididas en tablas
Menor complejidad	Mayor complejidad
Más desnormalizado	Algunas dimensiones normalizadas
Consultas más sencillas	Más relaciones/JOIN
Mayor simplicidad para análisis	Representa explícitamente jerarquías

EJERCICIO 1: Construimos una Estrella

Caso

El Ministerio quiere analizar las **internaciones hospitalarias**.

Necesita conocer:

- cantidad de internaciones;
- hospital;
- diagnóstico;
- edad;
- mes.

Consigna

Identificar:

- 1) Hecho.
- 2) Medida.
- 3) Dimensiones.
- 4) Granularidad.
- 5) Dibujar el esquema estrella.

EJERCICIO 1: Construimos una Estrella (SOLUCIÓN)

1) HECHO

Internación

2) MEDIDA

Cantidad de internaciones

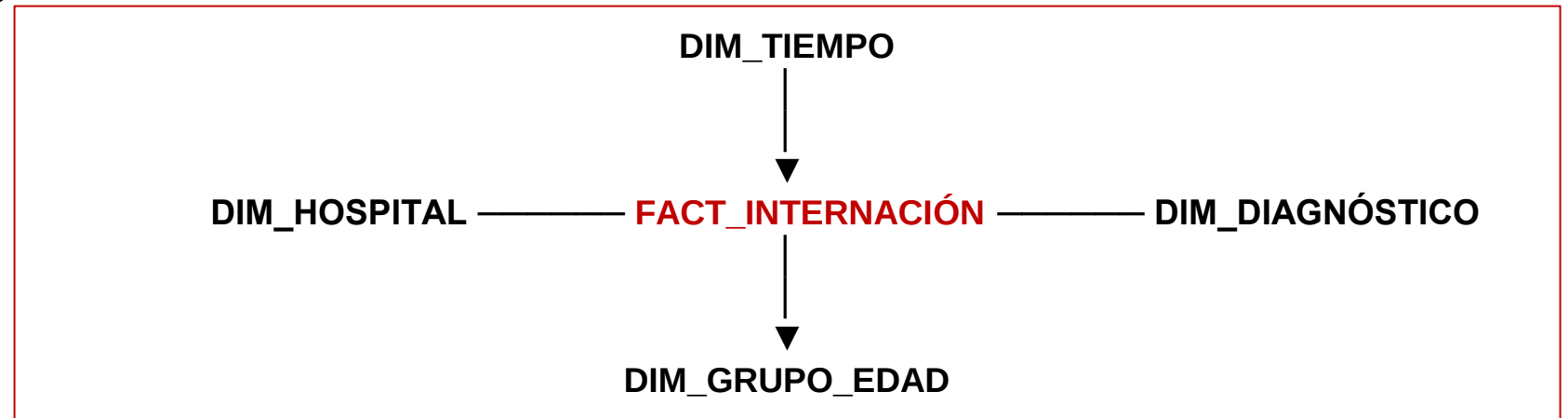
3) DIMENSIONES

- Hospital
- Diagnóstico
- Edad
- Tiempo

4) GRANULARIDAD

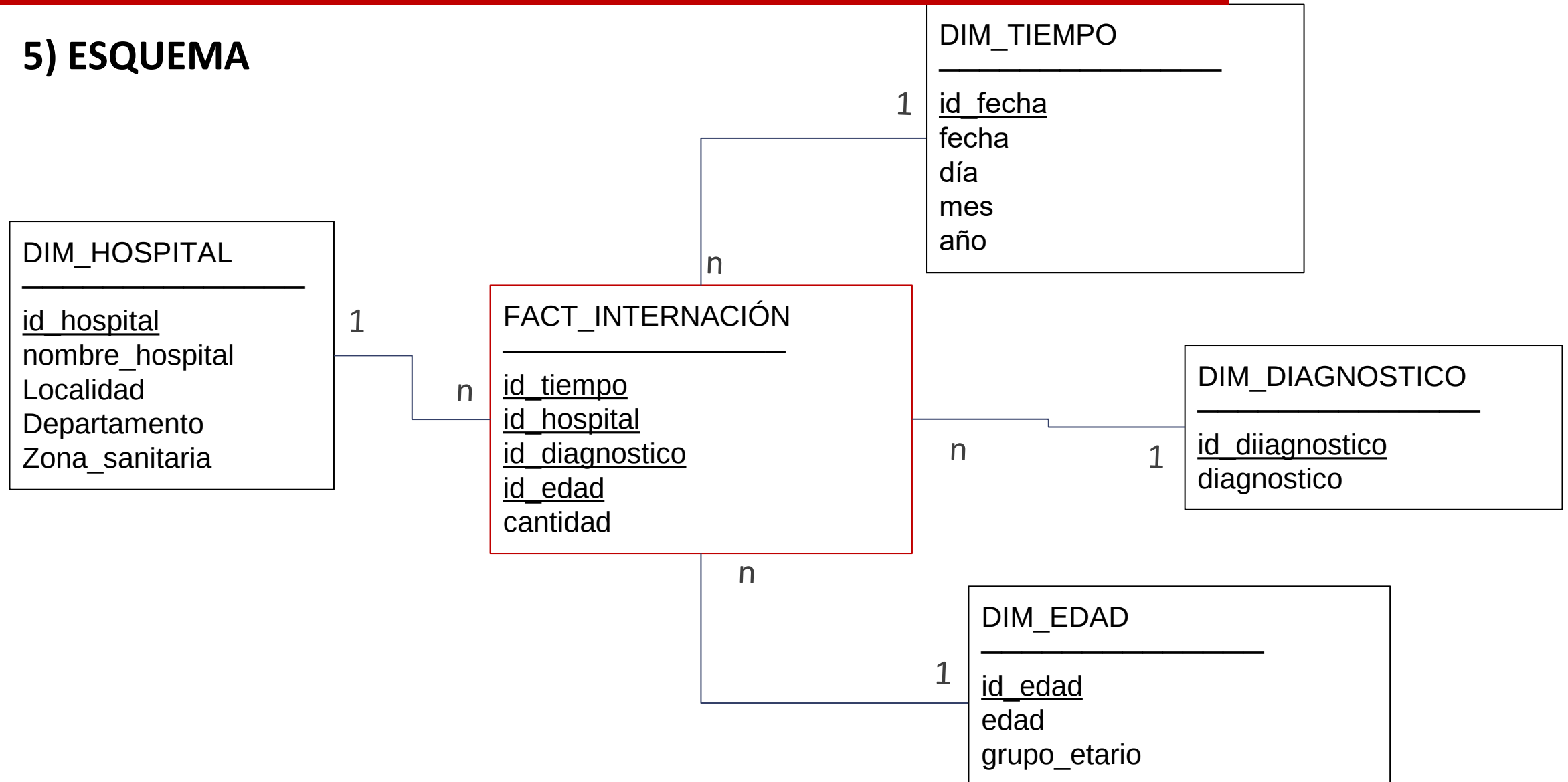
Una internación

5) ESQUEMA



EJERCICIO 1: Construimos una Estrella (SOLUCIÓN)

5) ESQUEMA



EJERCICIO 2: Convertimos Estrella en Copo de Nieve

1) HECHO

Internación

2) MEDIDA

Cantidad de internaciones

3) DIMENSIONES

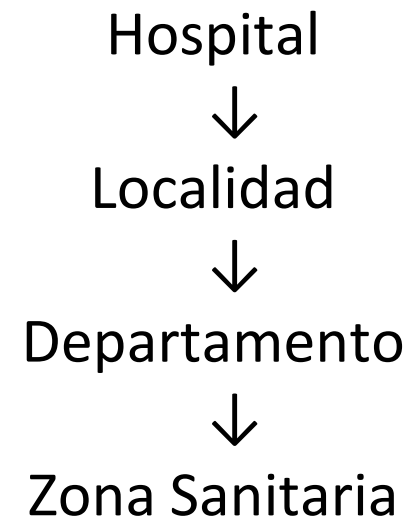
- Hospital
- Diagnóstico
- Edad
- Tiempo

4) GRANULARIDAD

Una internación

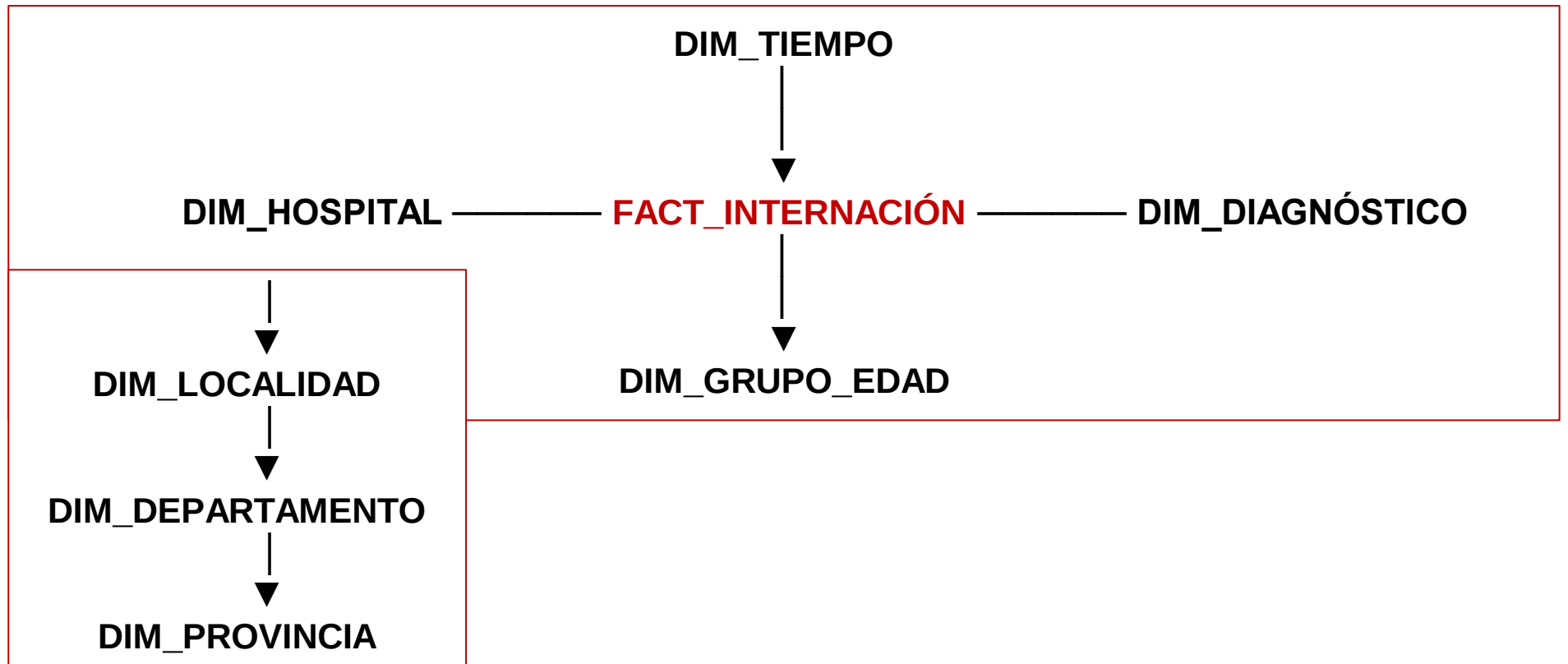
5) JERARQUÍA

La dimensión hospital tiene la siguiente jerarquía:

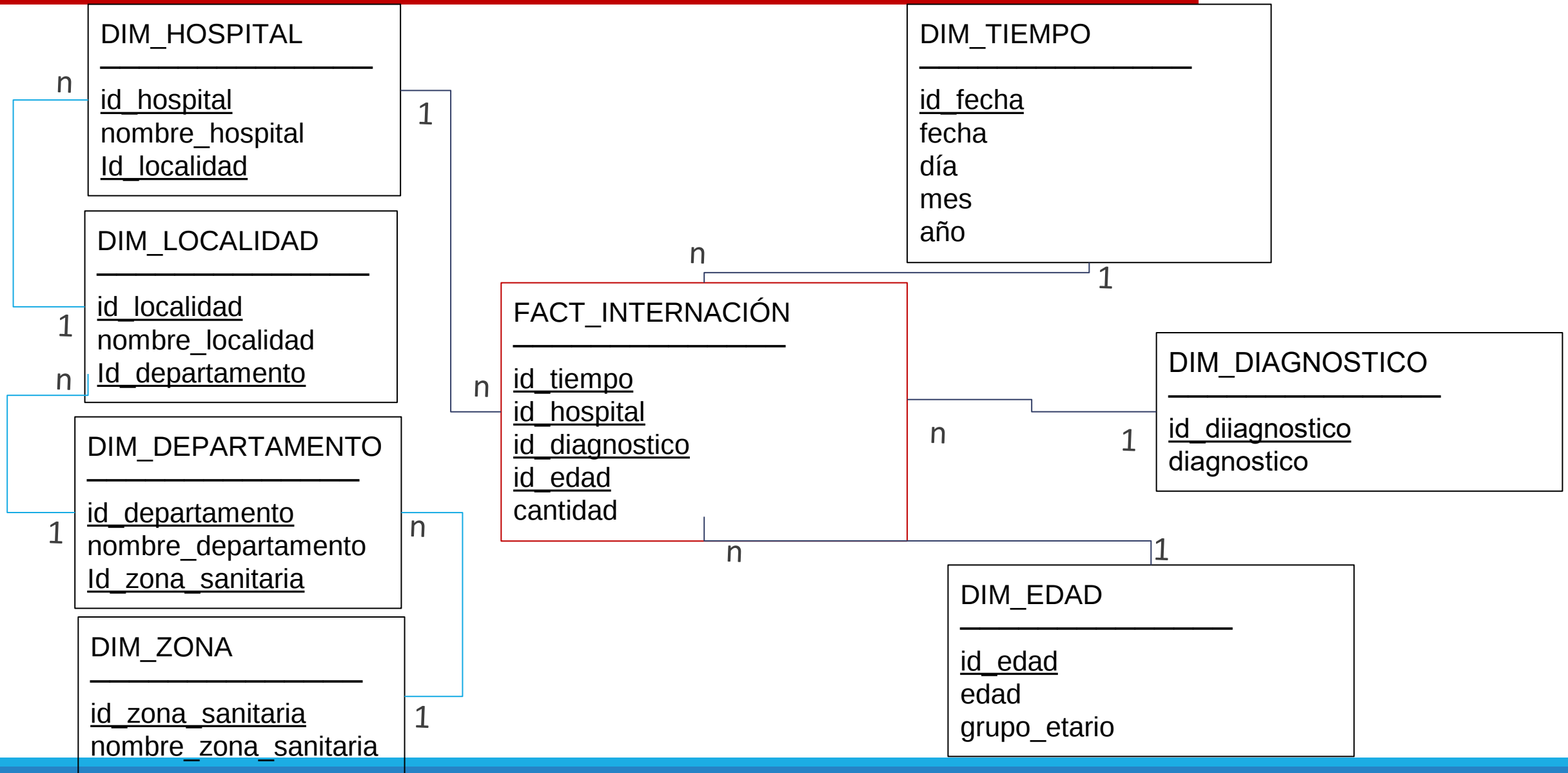


EJERCICIO 2: SOLUCIÓN

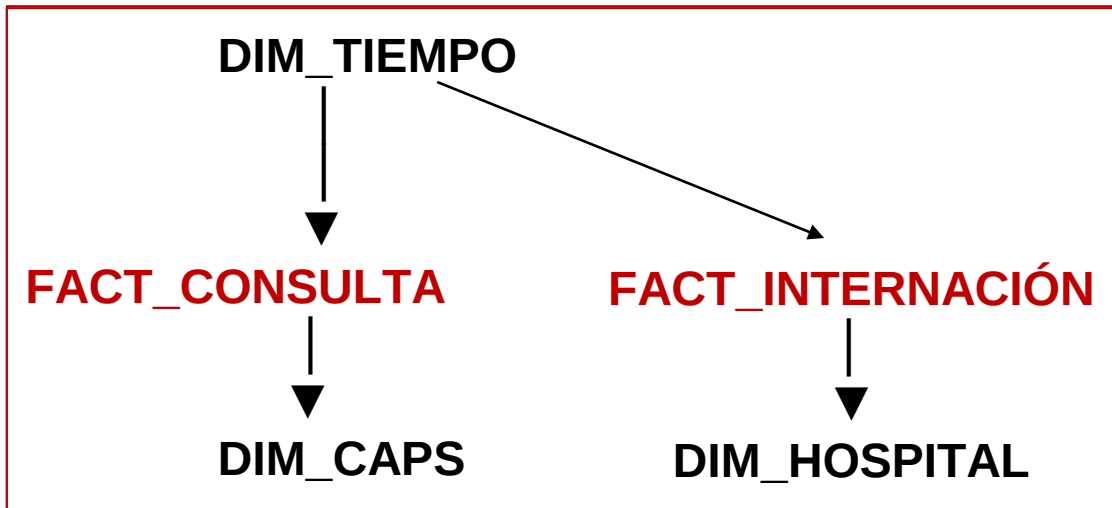
ESQUEMA



EJERCICIO 2: SOLUCIÓN



ESQUEMA EN CONSTELACIÓN



La constelación permite tener **más de una tabla de hechos**, que pueden compartir dimensiones

NO es necesario que las diferentes tablas de Hechos compartan las mismas tablas de Dimensiones.

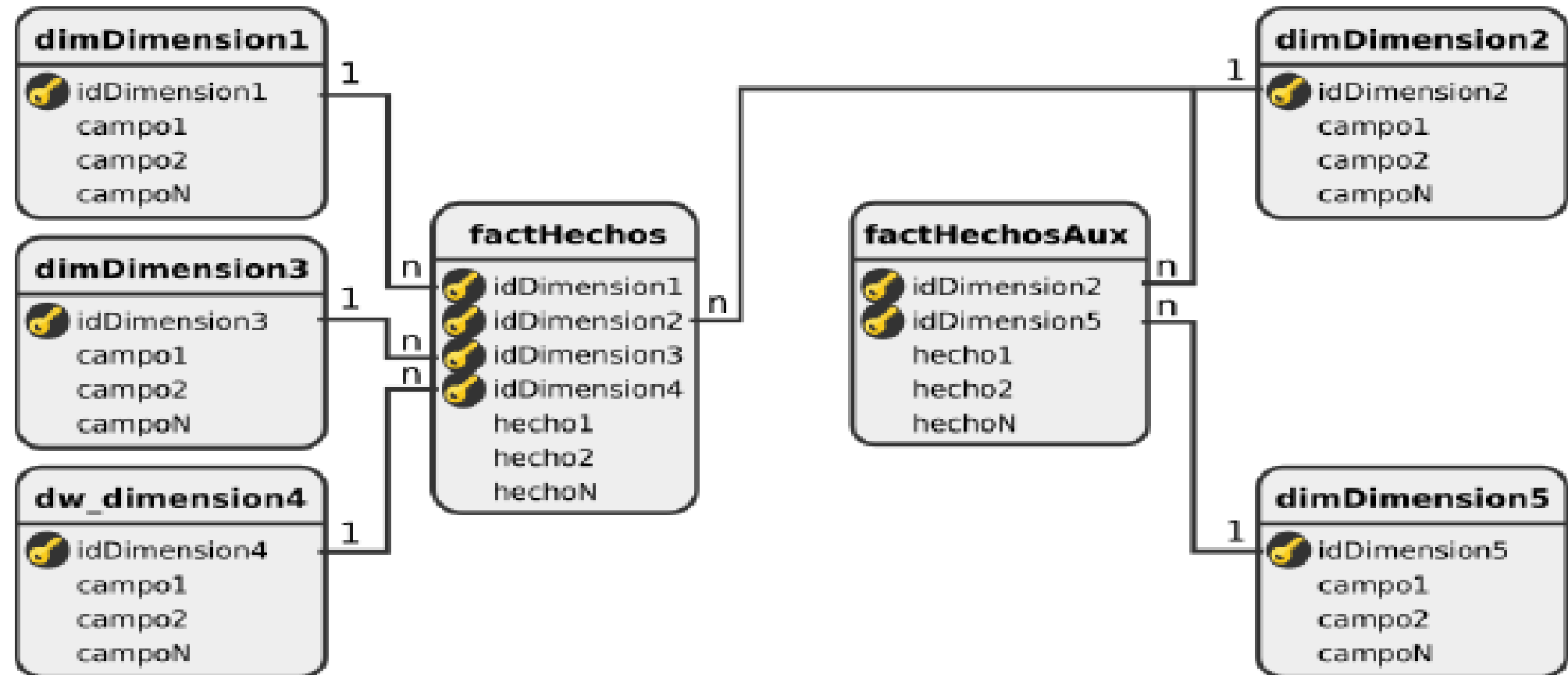
Características del Esquema en Constelación

- Permite tener más de una tabla de Hechos, por lo cual se tendrá mayor capacidad analítica.
- Contribuye a la reutilización de las tablas de Dimensiones, ya que una misma tabla de Dimensión puede utilizarse para varias tablas de Hechos.

ESQUEMA EN CONSTELACIÓN

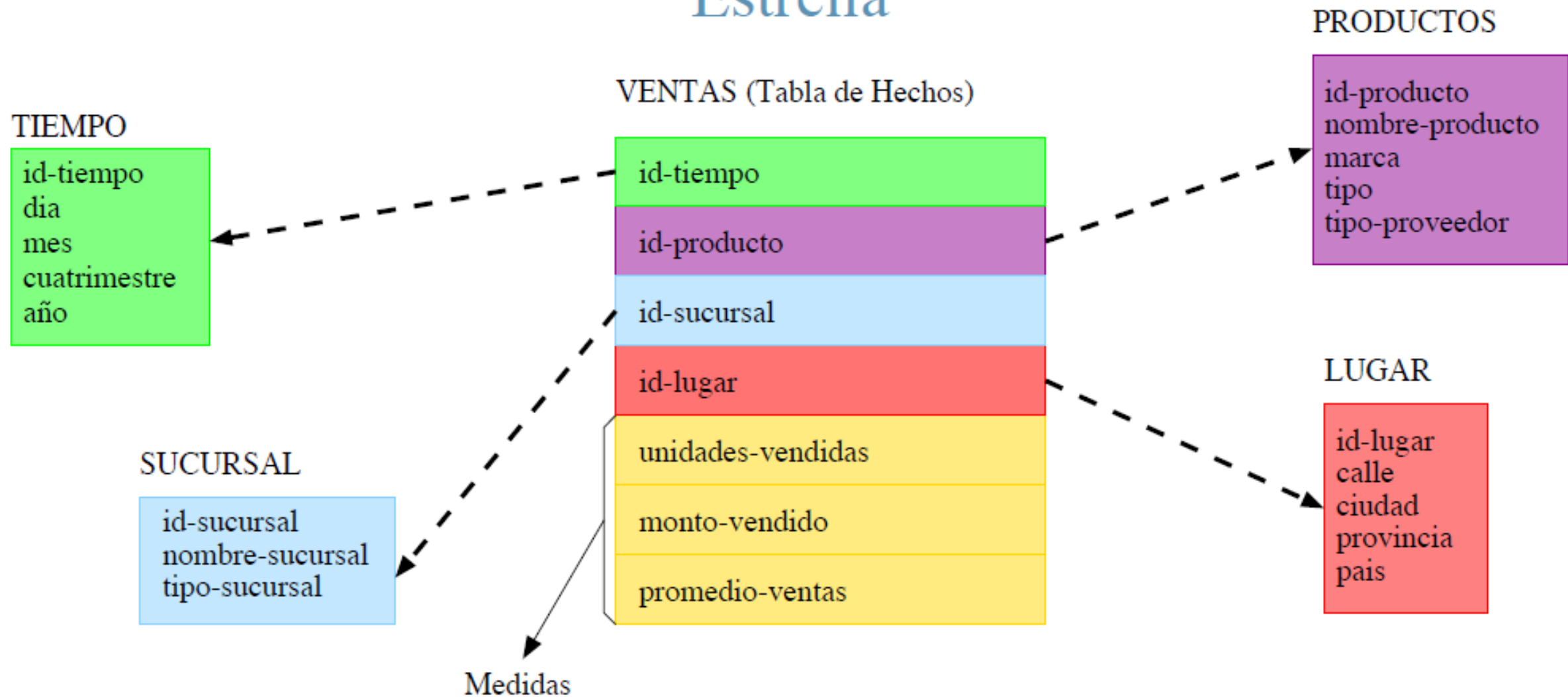
El Esquema en Constelación (Starflake Scheme):

- está compuesto por una serie de Esquemas en Estrella,
- posee una tabla de Hechos Principal (por ejemplo: **factHechos**) y
- una o más tablas de Hechos Auxiliares (por ejemplo: **factHechosAux**), las cuales pueden ser agregaciones de la Principal. Dichas tablas están relacionadas con sus respectivas tablas de Dimensiones.



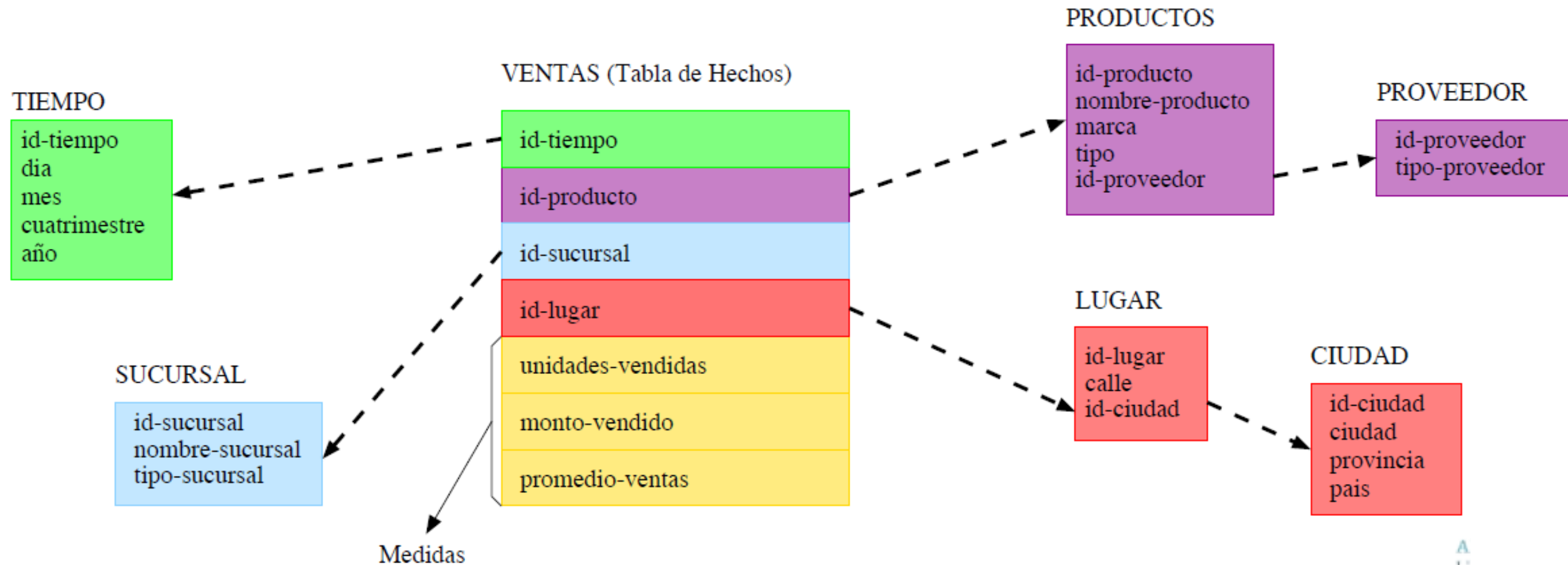
COMPARACIÓN DE ESQUEMAS - ESTRELLA

Estrella



COMPARACIÓN DE ESQUEMAS – COPO DE NIEVE

Copo de nieve



COMPARACIÓN DE ESQUEMAS – CONSTELACIÓN

Constelación

