Engineer

Esquema en estrella

★ Una tabla de hechos rodeada de tablas de dimensiones

Qué es

Un esquema en estrella organiza las tablas para el análisis. Una tabla central registra como números los sucesos del negocio, por ejemplo ventas, envíos o pagos. A su alrededor están las tablas que describen esos sucesos: qué producto, qué tienda, qué día y qué cliente. Si se dibuja una línea desde la tabla central hasta cada una de las demás, el diagrama parece una estrella.

La tabla central es la tabla de hechos y las de alrededor son tablas de dimensiones. Kimball Group lo describe brevemente: «Los esquemas en estrella se componen característicamente de tablas de hechos vinculadas a las tablas de dimensiones asociadas». El vínculo es una clave corriente. Cada tabla de dimensiones tiene una única clave primaria, y cada fila de hechos lleva esa clave como clave foránea, una por cada dimensión que describe el suceso.

La división responde a la distinción de Kimball Group entre medición y contexto. Un modelo dimensional se construye alrededor de los sucesos de medición de un proceso empresarial: un artículo escaneado en caja, un pedido enviado o una reclamación pagada. Los números que produce el suceso, como la cantidad y el importe, son los hechos. Todo lo que indica quién, qué, dónde, cuándo, por qué y cómo es contexto y reside en las dimensiones.

Los dos tipos de tablas tienen formas opuestas. En el nivel más detallado, una tabla de hechos contiene una fila por suceso, con las medidas numéricas, una clave foránea a cada dimensión y, a veces, una referencia como un número de factura o de recibo. Ese número es lo único que queda del documento cuando su fecha, cliente y tienda ya se han convertido en dimensiones; por eso se guarda en la fila de hechos sin una tabla de dimensiones detrás. Kimball Group lo llama dimensión degenerada. Una tabla de dimensiones, en cambio, es ancha, plana y

deliberadamente desnormalizada, con muchas columnas de texto descriptivo. Una dimensión de producto repite la marca, la categoría y el departamento en cada fila de producto, en lugar de separarlos en tablas propias.

Un esquema en estrella es la forma relacional de un modelo dimensional, construida con tablas corrientes en una base de datos SQL corriente. La otra forma que describe Kimball Group es el cubo OLAP, una base de datos multidimensional con los mismos atributos y hechos, consultada mediante lenguajes con mayor capacidad analítica que SQL, como MDX, mientras que XMLA proporciona un protocolo de acceso. Los cubos suelen construirse a partir de un esquema en estrella, como último paso antes de que los usuarios vean los datos o como capa de agregados. Las reglas de modelado son las mismas en ambos casos, y Kimball Group las considera neutrales respecto a la arquitectura: se aplican tanto a un almacén de datos de estilo Kimball como a la Corporate Information Factory de Bill Inmon, con su arquitectura de núcleo y radios.

Utilice el concepto en una decisión

Aplique «¿Qué es un esquema en estrella?» a una decisión real. Podemos ayudarle a relacionar el concepto con el sistema que está construyendo o comprando. Cifras de ventas, operaciones y finanzas reunidas para que dos departamentos no lleven dos respuestas distintas a la misma reunión.

Explore nuestros servicios Póngase en contacto →

Compruebe los supuestos detrás del término

Utilice la explicación y los trabajos enlazados para formular preguntas sobre su proyecto. Podemos examinar los supuestos con usted antes de elegir una herramienta. Flujo de datos geográficos 50 veces más rápido — 50x más rapidez en los flujos de procesamiento de datos geográficos

Flujo de datos geográficos 50 veces más rápido Póngase en contacto →

Por qué importa

La estructura hace que las preguntas analíticas se parezcan. Se eligen los hechos que se sumarán, se combina mediante JOIN la tabla de hechos con las dimensiones que los describen, se filtra por algunas columnas de dimensiones y se agrupa por otras. Kimball Group describe las tablas de hechos como el objetivo principal de los cálculos y las agregaciones dinámicas, y los atributos de dimensión como la fuente principal de restricciones de consulta, agrupaciones y etiquetas de informes. Un analista que haya aprendido ese patrón puede escribir sin ayuda la siguiente consulta sobre una tabla de hechos nueva, y una herramienta de informes puede construirla a partir de una lista de campos, porque todas las estrellas tienen la misma disposición.

La legibilidad es la segunda razón. Una base de datos operativa se diseña para almacenar cada hecho una sola vez y mantenerlo coherente, por lo que la categoría de un producto suele estar en una tabla de consulta separada, enlazada mediante un identificador. Es el diseño adecuado para el sistema que recibe pedidos y el inadecuado para quien busca las ventas por categoría, que primero debe averiguar qué tablas combinar mediante JOIN y en qué orden. Una dimensión plana reúne todas las columnas descriptivas en un lugar, expresadas con palabras. Kimball Group advierte contra la alternativa de normalizar la jerarquía de una dimensión en más tablas, porque los usuarios del negocio encuentran el resultado difícil de entender y recorrer.

La tercera razón es que una estrella bien construida puede responder a preguntas que nadie planteó al diseñarla. La granularidad de una tabla de hechos, el significado exacto de una fila, se declara antes de definir dimensiones o hechos, y Kimball Group recomienda empezar por la granularidad atómica: el nivel más bajo que registra el proceso empresarial. Una tabla con una fila por línea de recibo se puede sumar por día, tienda, marca o cualquier combinación de ellos. Una tabla que solo conserva totales diarios por tienda nunca podrá desglosarse por producto, sea cual sea la forma de preguntar. En términos de Kimball Group, los datos atómicos resisten consultas imprevisibles, mientras que los agregados presuponen las preguntas que se les harán.

El precio es una redundancia deliberada. El nombre del departamento se repite en cada fila de producto, y la carga que construye la estrella debe mantener coherentes esas repeticiones y decidir qué ocurre cuando cambia una descripción. La estrella tampoco sustituye al modelo operativo: un proceso de carga la construye a partir de él, y su actualidad depende de la última ejecución de esa carga.

Cómo se aplica en la práctica

Un comercio quiere conocer las ventas por categoría de producto y mes. El proceso empresarial es el recibo de caja y la granularidad se declara como una fila por línea de recibo. La dimensión de producto es una tabla plana; la tabla de hechos contiene una clave por dimensión, el número de recibo como dimensión degenerada y dos medidas.

```
CREATE TABLE dim_product (
    product_key  bigint PRIMARY KEY,   -- surrogate key, owned by the warehouse
    sku          text NOT NULL,        -- the code in the source system
    product_name text NOT NULL,
```

```
    brand        text NOT NULL,
    category     text NOT NULL,
    department   text NOT NULL
);

CREATE TABLE fact_sales (
    date_key       integer NOT NULL REFERENCES dim_date,
    store_key      integer NOT NULL REFERENCES dim_store,
    product_key    bigint  NOT NULL REFERENCES dim_product,
    receipt_number text    NOT NULL,   -- degenerate dimension
    quantity       integer NOT NULL,
    sales_amount   numeric(12, 2) NOT NULL
);

SELECT d.year_month, p.category, sum(f.sales_amount) AS sales
FROM fact_sales AS f
JOIN dim_date AS d USING (date_key)
JOIN dim_product AS p USING (product_key)
WHERE p.department = 'Grocery'
GROUP BY d.year_month, p.category
ORDER BY d.year_month, p.category;
```

La consulta tiene la forma de cualquier consulta a una estrella: la tabla de hechos, una operación JOIN por dimensión, filtros y agrupaciones sobre el texto de las dimensiones y una suma de un hecho. Añadir la región de la tienda al informe exige otra operación JOIN y una columna más en `GROUP BY` ; la tabla de hechos no cambia. El número de recibo permanece en la fila de hechos, de modo que las líneas de una cesta aún se pueden contar juntas, pero no recibe una tabla de dimensiones porque no hay nada más que describir sobre él.

Una estrella es una respuesta a la promesa del servicio de diseño y modelado de bases de datos: una estructura que refleje cómo funciona realmente el negocio, para poder seguir haciendo las preguntas que se necesiten después. El modelo operativo del que se construye es la otra mitad de ese trabajo. En su trabajo actual para NextMariner, la fundadora modeló un dominio marítimo en 348 tablas normalizadas distribuidas en 34 esquemas PostgreSQL, con profesionales, empresas, buques, puestos de trabajo y reseñas. Allí, las pequeñas listas estables, como estados, tipos y categorías, se convirtieron en tablas de consulta SMALLINT, lo que mantiene compactas las filas y poco costosas las operaciones JOIN, y las claves foráneas y restricciones reales hacen que la integridad sea responsabilidad de la base de datos. Esa es la estructura adecuada para un sistema que escribe filas de una en una y debe mantener la corrección. Una estrella para informes construida sobre un sistema así hace lo contrario: vuelve a convertir los códigos en palabras dentro de dimensiones planas, porque quien la lee es una persona y no la aplicación.

Dónde surgen los problemas

Se normalizan las dimensiones y la estrella se convierte en un copo de nieve. Separar marca, categoría y departamento de la dimensión de producto en tablas propias parece ordenado a quien está acostumbrado al diseño operativo. Kimball Group llama al resultado copo de nieve y lo desaconseja: los usuarios del negocio lo encuentran difícil de entender y recorrer, y las operaciones JOIN adicionales pueden ralentizar las consultas, mientras que una dimensión plana contiene exactamente la misma información. Mantenga cada dimensión como una tabla ancha y deje que la carga repita la jerarquía en cada fila.

La granularidad se ajusta al primer informe. Una tabla de hechos de ventas diarias por tienda responde al panel para el que se construyó y no permite bajar de ese nivel; la primera pregunta sobre un producto o una cesta exige volver a cargar los datos desde el origen. Una tabla que mezcla totales diarios y líneas individuales es peor, porque cualquier suma cuenta dos veces la misma venta. Declare primero la granularidad, empiece en el nivel atómico que registre el proceso y asigne una tabla de hechos propia a cada granularidad. Después se pueden añadir tablas de resumen independientes para ganar velocidad.

Los atributos de dimensión son los códigos del sistema de origen. Los sistemas de gestión operativa guardan códigos compactos, y copiarlos directamente a una dimensión produce informes con etiquetas de fila que solo entiende el sistema de origen. Kimball Group admite códigos como atributos, pero indica que los más útiles son las descripciones extensas, porque son las que acaban como etiquetas de un informe. Durante la carga, convierta cada código en palabras y consérvelo en una columna junto a la descripción para quien necesite rastrear una fila hasta su origen.

La tabla de hechos tiene forma de informe. Resulta tentador diseñarla a partir de la hoja de cálculo que ya utiliza el negocio, con una columna por mes o una fila por línea del informe de dirección. Kimball Group afirma expresamente que el diseño de una tabla de hechos se basa en la actividad física que se mide, no en los informes que se obtendrán de ella. Los diseños de los informes cambian; el recibo de caja no. Modele el suceso, con una fila por ocurrencia en la granularidad declarada, y deje que las consultas produzcan las distintas disposiciones.

Lecturas adicionales

- Star Schemas and OLAP Cubes — la definición, los hechos y las dimensiones como medición y contexto y la relación entre cubos y estrellas.
- Grain — declarar qué significa una fila de hechos y por qué la granularidad atómica es lo primero.
- Fact Table Structure — medidas numéricas y claves foráneas en una fila por suceso, diseñadas a partir de la actividad en lugar del informe.
- Dimension Table Structure — tablas anchas, planas y desnormalizadas cuyos atributos descriptivos determinan filtros, agrupaciones y etiquetas.
- Snowflaked Dimensions — por qué normalizar la jerarquía de una dimensión dificulta su uso.
- Microsoft: acceso a datos de modelos multidimensionales — MDX como lenguaje de consulta y cálculo y XMLA como interfaz a través de la cual un cliente puede enviar una sentencia MDX.

Glosario

Elija según sus restricciones reales

Traiga un conjunto de datos representativo, las preguntas que debe responder y el cuello de botella actual. Podemos acordar cómo medir la calidad, la capacidad de procesamiento y la recuperación.

Vea cuánto cuesta Póngase en contacto →

Haga práctica la terminología

Pregunte cómo se aplica «¿Qué es un esquema en estrella?» a su organización. Basta un ejemplo concreto de su sistema actual para iniciar la conversación.

Explore nuestros servicios

welcome@engineer.company +45 50 10 61 84 Copenhague, Dinamarca

Llévese esta página

Tarjeta de contacto

Creado con esmero y cariño por D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Derechos de autor © 2025 – actualidad · Engineer ApS · Todos los derechos reservados.

Impreso desde https://engineer.company/es/glossary/star-schema/