

El modelo estrella



Universidad
Tecnológica
del Perú

Contenido:

1. Introducción
2. Logro
3. Transformación
4. Cierre

- # 1. Introducción:
- La creación de un Datamart



Universidad
Tecnológica
del Perú

2. Logro de la sesión

Al finalizar la sesión, el estudiante podrá entender los conceptos básicos del modelo estrella para un Data mart

3. Transformación



Universidad
Tecnológica
del Perú

Ejemplos relacionados al modelo estrella

Relevancia del esquema de estrella para los modelos de Power BI

El diseño de esquema de estrella y muchos conceptos relacionados, son muy importantes para desarrollar modelos de Power BI optimizados para el rendimiento y la facilidad de uso.

Tener en cuenta que cada objeto visual de informe de Power BI genera una consulta que se envía al modelo de Power BI (lo que el servicio Power BI denomina un conjunto de datos). Estas consultas se usan para filtrar, agrupar y resumir los datos del modelo. Por tanto, un modelo bien diseñado es aquel que proporciona tablas para filtrar y agrupar y tablas para resumir. Este diseño se ajusta bien a los principios de los esquemas de estrella:

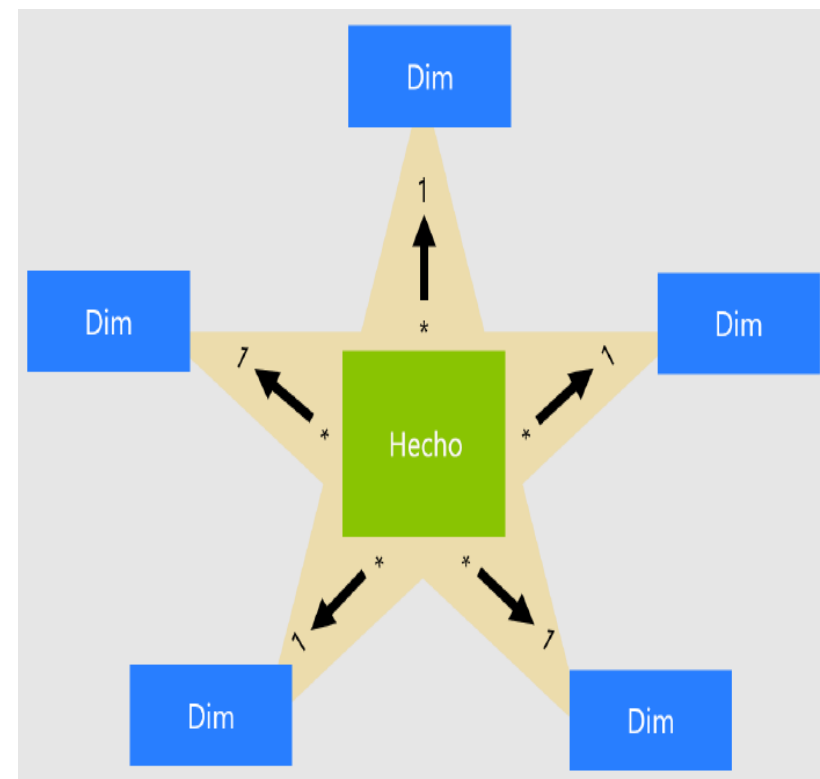
- Las tablas de dimensiones admiten el *filtrado* y la *agrupación*
- Las tablas de hechos admiten el *resumen*

Desaprende lo que te limita

3. Transformación

Ejemplos relacionados al modelo estrella

Los modeladores no establecen ninguna propiedad de tabla para configurar el tipo de tabla como dimensión o hecho. Es un hecho que las relaciones de modelo determinan. Una relación de modelo establece una ruta de propagación de filtros entre dos tablas, mientras que la propiedad **Cardinalidad** de la relación es la que determina el tipo de tabla. Una cardinalidad de relación común es *uno a varios* o su inversa *varios a uno*. El "uno" es siempre una tabla de tipo de dimensiones, mientras que el "varios" siempre es una tabla de tipo de hechos.



Desaprende lo que te limita

3. Transformación



Universidad
Tecnológica
del Perú

Ejemplos relacionados al modelo estrella

Un diseño de modelo bien estructurado debe incluir tablas de tipo de dimensiones o de tipo de hechos. Evitar mezclar los dos tipos en una sola tabla. También se recomienda intentar ofrecer el número correcto de tablas con las relaciones adecuadas aplicadas. Además, es importante que las tablas de tipo de hechos siempre carguen datos en un nivel de detalle coherente.

Por último, es importante entender que un diseño óptimo del modelo es en parte ciencia y en parte arte. A veces, puede interrumpir con buena orientación cuando tenga sentido hacerlo.

Desaprende lo que te limita

3. Transformación



Universidad
Tecnológica
del Perú

Ejemplos relacionados al modelo estrella

Hay muchos conceptos adicionales relacionados con el diseño de esquemas de estrella que se pueden aplicar a un modelo de Power BI. Entre estos conceptos se incluyen los siguientes:

1. Medidas
2. Claves suplentes
3. Dimensiones de copo de nieve
4. Dimensiones realizadoras de roles
5. Dimensiones de variación lenta
6. Dimensiones no deseadas
7. Dimensiones degeneradas
8. Tablas de hechos sin hechos

Desaprende lo que te limita

3. Transformación



Universidad
Tecnológica
del Perú

Ejemplos relacionados al modelo estrella

1. Medidas:

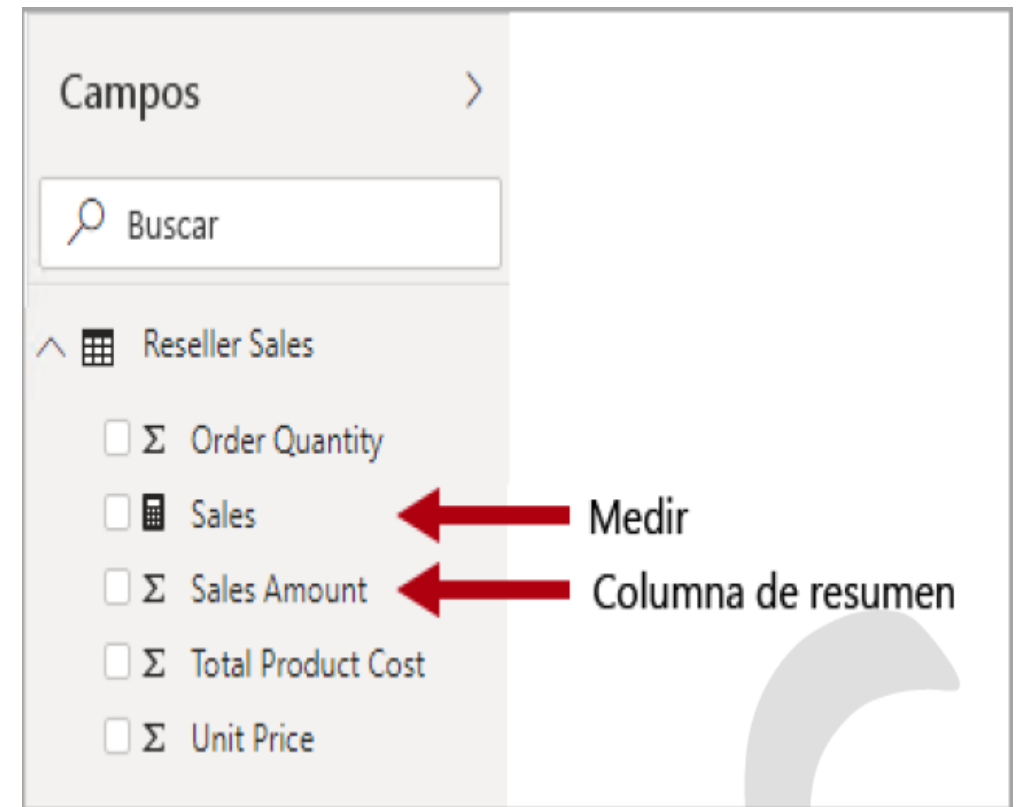
En el diseño de esquemas de estrella, una **medida** es una columna de tabla de hechos que almacena valores que se van a resumir.

En un modelo de Power BI, una **medida** tiene otra definición, aunque similar. Es una fórmula escrita en Expresiones de análisis de datos (DAX) que permite resumir. Las expresiones de medida suelen aprovechar funciones de agregación de DAX como SUM, MIN, MAX, AVERAGE, etc. para generar un resultado de valor escalar en tiempo de consulta (los valores nunca se almacenan en el modelo). La expresión de medida puede abarcar desde agregaciones de columnas simples hasta fórmulas más sofisticadas que invalidan las propagaciones de contexto o de relación de filtrado.

3. Transformación

Ejemplos relacionados al modelo estrella

Los modelos de Power BI admiten un segundo método para lograr el resumen. Cualquier columna (y, normalmente, las columnas numéricas) se puede resumir mediante un informe visual o Preguntas y respuestas. Estas columnas se conocen como *medidas implícitas*. Resultan cómodas para el desarrollador de modelos, ya que en muchos casos no es necesario crear medidas. Por ejemplo, la columna de ventas **Importe de venta** del distribuidor Adventure Works podría resumirse de varias maneras (suma, recuento, media, mediana, mín., máx., etc.), sin necesidad de crear una medida para cada tipo de agregación posible.



Desaprende lo que te limita

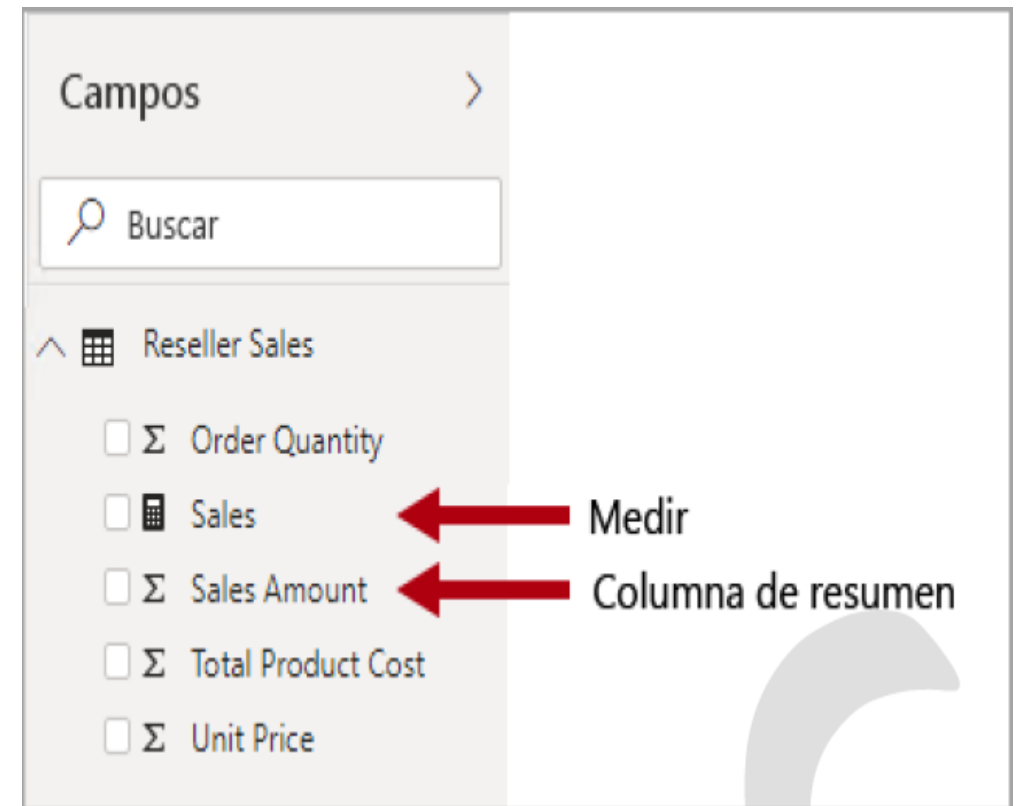
3. Transformación

Ejemplos relacionados al modelo estrella

2. Claves suplentes

Una **clave suplente** es un identificador único que se agrega a una tabla para admitir el modelado de esquemas de estrella. Por definición, no se define ni se almacena en los datos de origen.

Normalmente, las claves suplentes se agregan a las tablas de dimensiones del almacén de datos relacionales para proporcionar un identificador único para cada fila de la tabla de dimensiones.



Desaprende lo que te limita

3. Transformación

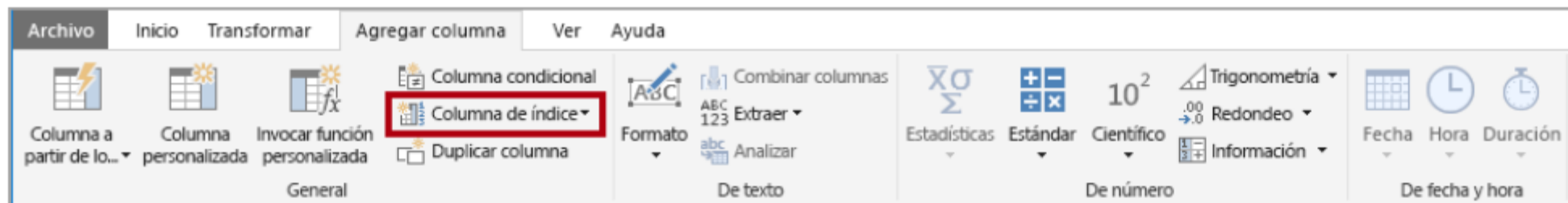
Ejemplos relacionados al modelo estrella

2. Claves suplentes

Una **clave suplente** es un identificador único que se agrega a una tabla para admitir el modelado de esquemas de estrella. Por definición, no se define ni se almacena en los datos de origen.

Normalmente, las claves suplentes se agregan a las tablas de dimensiones del almacén de datos relacionales para proporcionar un identificador único para cada fila de la tabla de dimensiones.

Las relaciones del modelo de Power BI se basan en una sola columna única de una tabla, que propaga los filtros a una sola columna de otra tabla. Cuando una tabla de tipo de dimensiones del modelo no incluye una sola columna única, debe agregar un identificador único para que se convierta en el "uno" de una relación. En Power BI Desktop, puede lograr fácilmente este requisito si crea una columna de índice de Power Query.



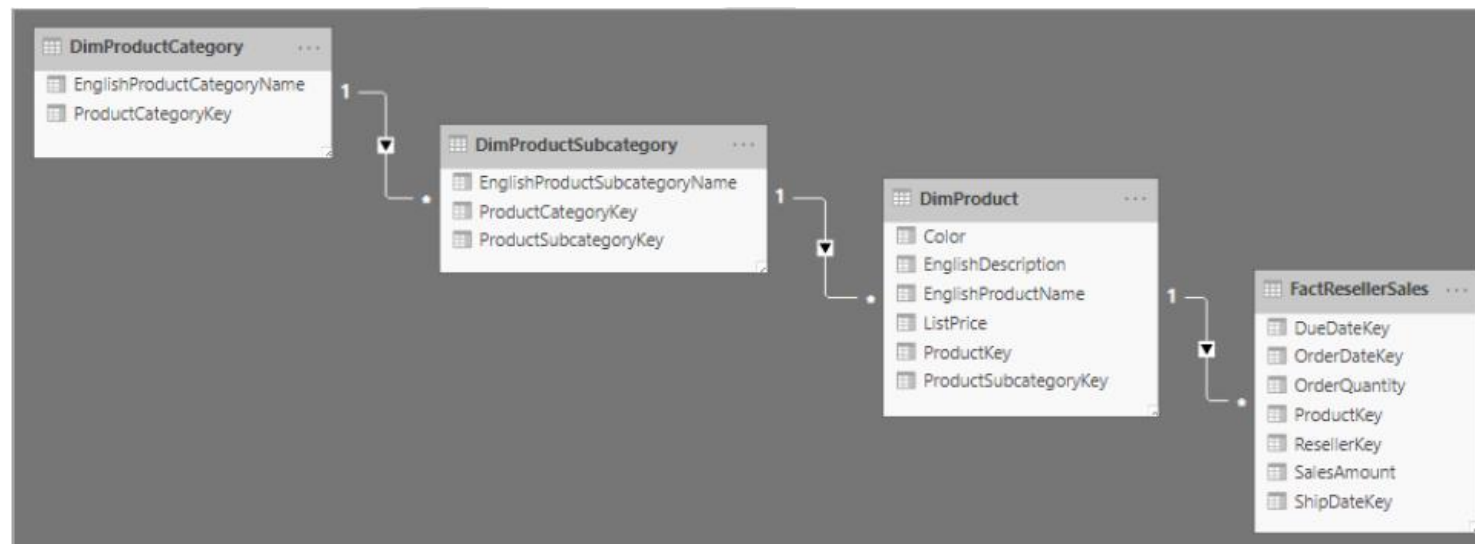
3. Transformación

Ejemplos relacionados al modelo estrella

3. Dimensiones de copo de nieve

Una **dimensión de copo de nieve** es un conjunto de tablas normalizadas para una sola entidad de negocio. Por ejemplo, Adventure Works clasifica los productos por categoría y subcategoría. Las categorías se asignan a subcategorías y, a su vez, los productos se asignan a subcategorías. En el almacén de datos relacionales de Adventure Works, la dimensión de producto se normaliza y se almacena en tres tablas relacionadas: **DimProductCategory**, **DimProductSubcategory** y **DimProduct**.

Podemos imaginar las tablas normalizadas colocadas hacia fuera de la tabla de hechos, formando un diseño de copo de nieve.



lo que te limita

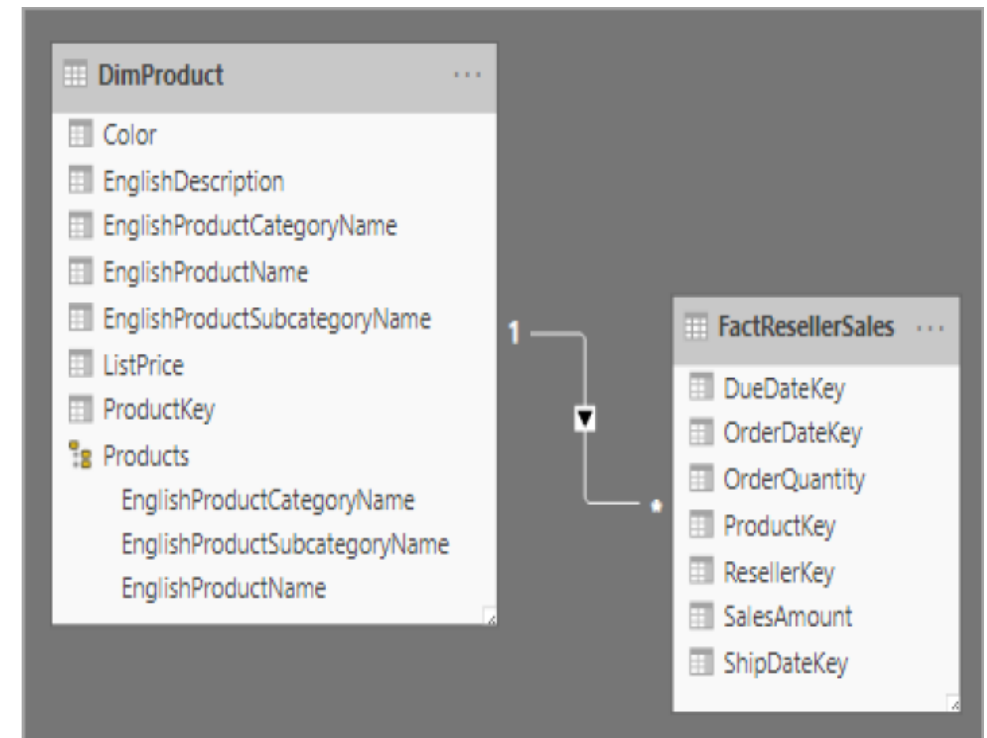
3. Transformación

Ejemplos relacionados al modelo estrella

3. Dimensiones de copo de nieve

En Power BI Desktop, puede optar por imitar un diseño de dimensiones de copo de nieve (quizás porque los datos de origen lo hacen) o integrar (desnormalizar) las tablas de origen en una sola tabla del modelo. Por lo general, las ventajas de una sola tabla del modelo compensan las ventajas de varias tablas del modelo. La decisión óptima puede depender de los volúmenes de datos y de los requisitos de facilidad de uso del modelo.

Si elige integrar en una sola tabla del modelo, además puede definir una jerarquía que abarque el mayor y menor nivel de detalle de la dimensión. Posiblemente, el almacenamiento de datos desnormalizados redundantes requiera para el modelo, mayor almacenamiento, especialmente en el caso de tablas de dimensiones muy grandes.



3. Transformación

Ejemplos relacionados al modelo estrella

4. Dimensiones de variación lenta

Una **dimensión de variación lenta** (SCD) es aquella que administra correctamente el cambio de los miembros de la dimensión a lo largo del tiempo. Se aplica cuando los valores de la entidad empresarial cambian con el tiempo y de forma ad hoc. Un buen ejemplo de dimensión de variación *lenta* es una dimensión de cliente, concretamente sus columnas de detalles de contacto, como la dirección de correo electrónico y el número de teléfono. Por el contrario, algunas dimensiones se consideran de variación rápida si un atributo de dimensión cambia con frecuencia, como el precio de mercado de un artículo. El enfoque de diseño común en estas instancias es almacenar los valores de atributo de variación rápida en una medida de tabla de hechos.

La teoría de diseño de esquemas de estrella hace referencia a dos tipos de SCD comunes: Tipo 1 y tipo 2. Una tabla de tipo de dimensiones puede ser de tipo 1 o de tipo 2, o admitir ambos tipos simultáneamente para columnas diferentes.

3. Transformación

Ejemplos relacionados al modelo estrella

4. Dimensiones de variación lenta

SCD de tipo 1

Una **SCD** de **tipo 1** siempre refleja los valores más recientes y, cuando se detectan cambios en los datos de origen, los datos de la tabla de dimensiones se sobrescriben. Este enfoque de diseño es común para las columnas que almacenan valores auxiliares, como la dirección de correo electrónico o el número de teléfono de un cliente. Cuando cambia la dirección de correo electrónico o el número de teléfono de un cliente, la tabla de dimensiones actualiza la fila del cliente con los nuevos valores. Es como si el cliente tuviera siempre esta información de contacto.

Una actualización no incremental de una tabla de tipo de dimensiones del modelo de Power BI logra el resultado de una SCD de tipo 1. Actualiza los datos de la tabla para garantizar que se carguen los valores más recientes.

3. Transformación

Ejemplos relacionados al modelo estrella

4. Dimensiones de variación lenta

SCD de tipo 2

Una **SCD** de **tipo 2** admite el control de versiones de los miembros de la dimensión. Si el sistema de origen no almacena versiones, suele ser el proceso de carga de almacenamiento de datos el que detecta los cambios y administra de forma adecuada el cambio en una tabla de dimensiones. En este caso, la tabla de dimensiones debe usar una clave suplente para proporcionar una referencia única a una *versión* del miembro de la dimensión. También incluye columnas que definen la validez del intervalo de fechas de la versión (por ejemplo, **StartDate** y **EndDate**) y, posiblemente, una columna de marca (por ejemplo, **IsCurrent**) para filtrar fácilmente por miembros de la dimensión actuales.

3. Transformación

Test



Explique los tipos de Dimensiones de variación lenta (SCD)

Desaprende lo que te limita



**Universidad
Tecnológica
del Perú**