

# Introducción

La creación de un modelo de datos excelente es una de las tareas más importantes que un analista de datos puede realizar en Microsoft Power BI. Al hacer este trabajo de forma correcta, ayudará a los usuarios a comprender mejor los datos, lo que facilitará la creación de informes de Power BI útiles para ellos y para usted.

Un buen modelo de datos ofrece las ventajas siguientes:

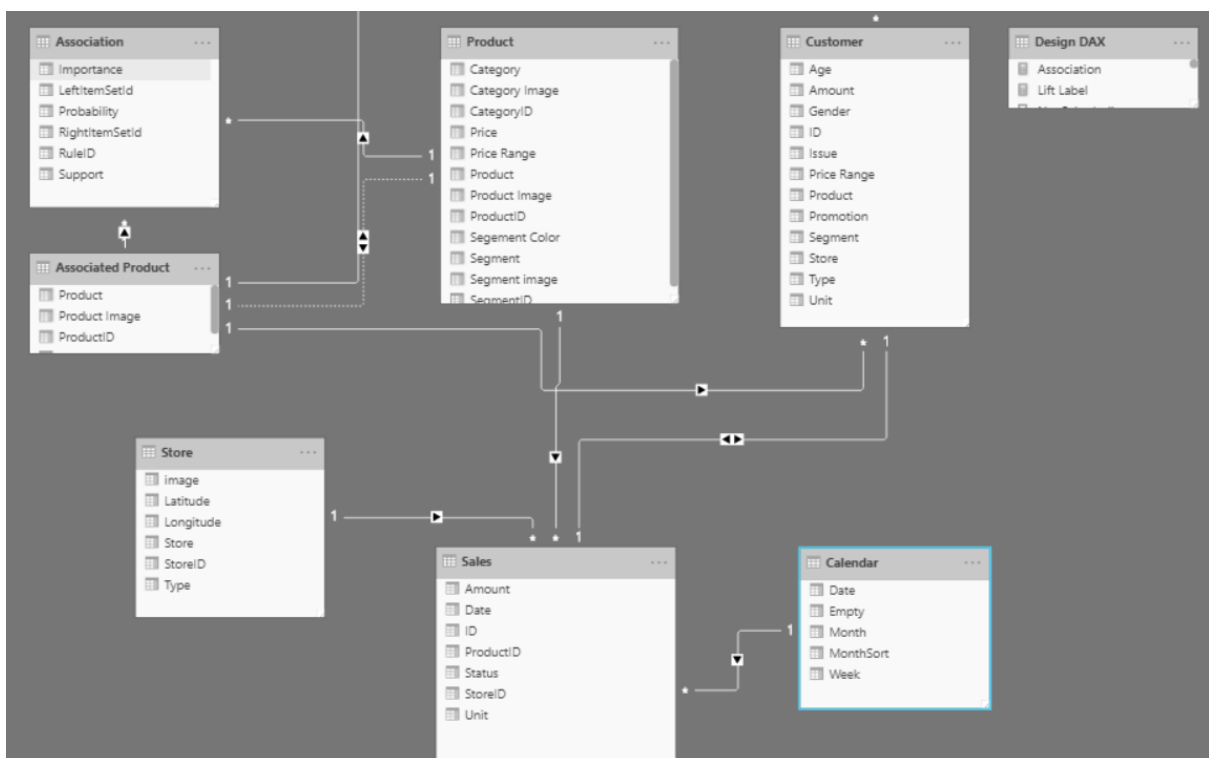
- La exploración de datos es más rápida.
- Las agregaciones son más fáciles de crear.
- Los informes son más precisos.
- La creación de informes tarda menos tiempo.
- Los informes son más fáciles de mantener en el futuro.

Proporcionar reglas fijas sobre qué convierte a un modelo en un buen modelo de datos es difícil, porque todos los datos son diferentes y su uso variable. Por lo general, un modelo de datos más pequeño es mejor porque funcionará más rápido y será más sencillo de usar. Pero la definición de lo que engloba un modelo de datos más pequeño es igualmente problemática, ya que se trata de un concepto heurístico y subjetivo.

Normalmente, un modelo de datos más pequeño se compone de menos tablas y menos columnas en cada tabla que el usuario puede ver. Si importa todas las tablas necesarias de una base de datos de ventas, pero el número total de tablas es de 30, el usuario no lo encontrará intuitivo. La contracción de todas esas tablas en cinco hará que el modelo de datos sea más intuitivo para el usuario, mientras que si el usuario abre una tabla y encuentra 100 columnas, puede que le resulte abrumadora. Si se quitan las columnas innecesarias para proporcionar un número más fácil de administrar, aumentará la probabilidad de que el usuario lea todos los nombres de

columna. En resumen, al diseñar los modelos de datos, el objetivo debe ser la simplicidad.

La siguiente imagen es un modelo de datos de ejemplo. Los cuadros contienen tablas de datos, donde cada elemento de línea del cuadro es una columna. Las líneas que conectan los cuadros representan las relaciones entre las tablas. Estas relaciones pueden ser complejas, incluso en un modelo tan sencillo. El modelo de datos se puede desorganizar fácilmente y el número total de tablas del modelo puede aumentar de manera gradual. Para que el modelo de datos sea simple, completo y preciso se requiere un esfuerzo constante.



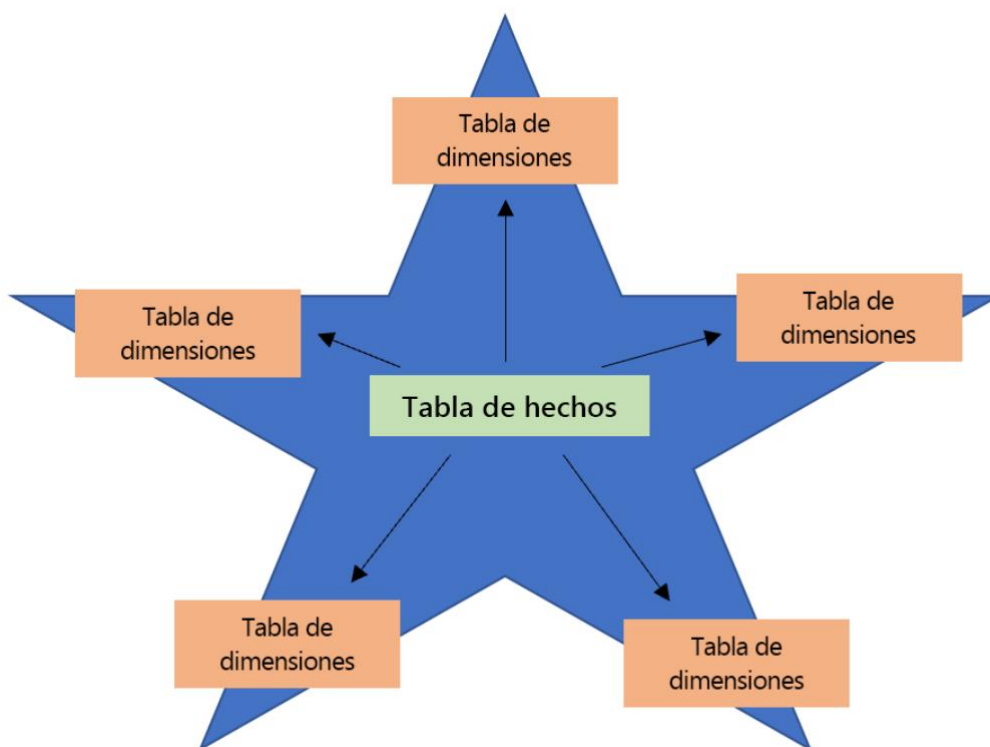
Las relaciones se definen entre las tablas a través de claves principales y externas. Las claves principales son columnas que identifican cada fila de datos única que no sea NULL. Por ejemplo, si tiene una tabla Clientes, puede tener un índice que identifique a cada cliente único. La primera fila tendrá un identificador de 1, la segunda fila un identificador de 2, etc. A cada fila se le asigna un valor único, al que se puede hacer referencia mediante este valor simple: la clave principal. Este proceso es importante cuando se hace referencia a las filas de otra tabla, que es lo que hacen las claves externas. Las relaciones entre las tablas se establecen cuando tiene claves principales y externas en común entre tablas distintas.

Power BI permite crear relaciones a partir de tablas con orígenes de datos diferentes, una función eficaz que permite extraer una tabla de Microsoft Excel y otra de una base de datos relacional. Después, tendrá que crear la relación entre esas dos tablas y las tratará como un conjunto de datos unificado.

Ahora que ha obtenido información sobre las relaciones que componen el esquema de datos, podrá explorar un tipo específico de diseño de esquema, el esquema de estrella, que está optimizado para alto rendimiento y facilidad de uso.

## Esquemas de estrella

Puede diseñar un esquema de estrella para simplificar los datos. No es la única manera de simplificarlos, pero es un método popular; por tanto, todos los analistas de datos de Power BI deben comprenderlo. En un esquema de estrella, cada tabla del conjunto de datos se define como una tabla de hechos o de dimensiones, como se muestra en el siguiente objeto visual.



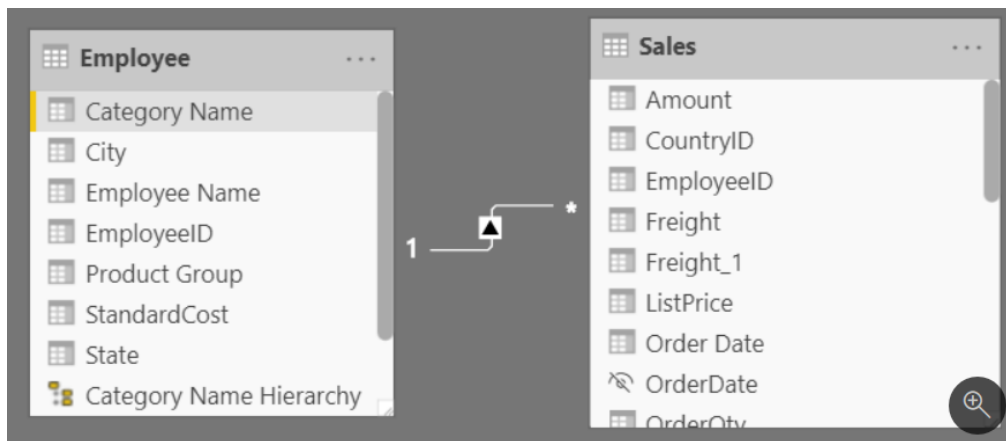
Las tablas de hechos contienen valores de datos de eventos o de observación: pedidos de ventas, recuentos de productos, precios, fechas y horas de transacciones, y cantidades. Las tablas de hechos pueden contener varios valores repetidos. Por ejemplo, un producto puede aparecer varias veces en varias filas para diferentes clientes en fechas distintas. Estos valores se pueden sumar para crear objetos visuales. Por ejemplo, un objeto visual del total de pedidos de ventas es una suma de todos los pedidos de ventas en la tabla de hechos. Con las tablas de hechos, es habitual ver columnas rellenas con números y fechas. Los números pueden ser unidades de medida, como el importe de venta, o pueden ser claves, como un identificador de cliente. Las fechas representan el tiempo que se registra, como la fecha del pedido o la del envío.

Las tablas de dimensiones contienen los detalles sobre los datos de las tablas de hechos: productos, ubicaciones, empleados y tipos de pedido. Estas tablas están conectadas a la tabla de hechos a través de columnas de clave. Las tablas de dimensiones se usan para filtrar y agrupar los datos de las tablas de hechos. Por el contrario, las tablas de dimensiones contienen valores únicos, por ejemplo, una fila por cada producto de la tabla Productos y una fila por cada cliente de la tabla Cliente. En el objeto visual de pedidos de ventas totales, puede agrupar los datos para ver los pedidos de ventas totales por producto, donde el producto son datos de la tabla de dimensiones.

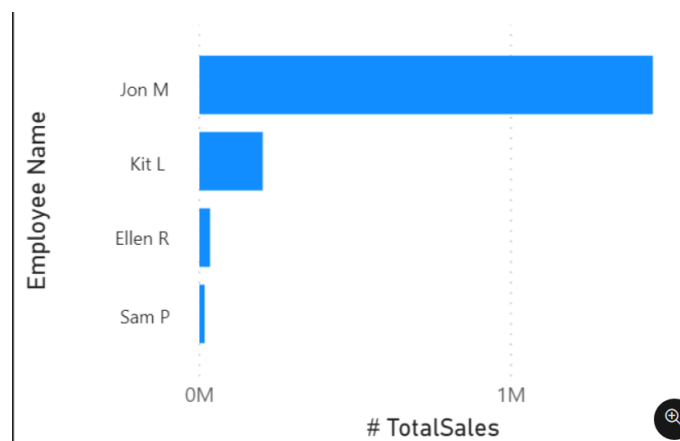
Las tablas de hechos suelen ser mucho mayores que las de dimensiones porque en ellas se producen numerosos eventos, como ventas individuales. Las tablas de dimensiones suelen ser más pequeñas porque está limitado al número de elementos que puede filtrar y agrupar. Por ejemplo, un año solo contiene un número finito de meses y los Estados Unidos solo se componen de un número concreto de estados.

Teniendo en cuenta esta información sobre las tablas de hechos y las de dimensiones, se preguntará cómo puede crear este objeto visual en Power BI.

Los datos pertinentes se encuentran en dos tablas, Employee (Empleado) y Sales (Ventas), como se muestra en el siguiente modelo de datos. Como la tabla Sales contiene los valores de pedidos de venta, que se pueden agregar, se considera una tabla de hechos. La tabla Employee contiene el nombre de empleado específico, que filtra los pedidos de ventas, por lo que sería una tabla de dimensiones. La columna común entre las dos tablas, que es la clave principal de la tabla Employee, es EmployeeID, por lo que puede establecer una relación entre las dos tablas en función de esta columna.



Al crear esta relación, puede diseñar el objeto visual según los requisitos, como se muestra en la ilustración siguiente. Si no hubiera establecido esta relación, pero sin olvidar las similitudes entre las dos tablas, habría tenido más dificultades para crear el objeto visual.



Los esquemas de estrella y el modelo de datos subyacente son la base de los informes organizados; cuanto más tiempo dedique a crear estas conexiones y el diseño, más fácil será crear y mantener informes.