

Hechos y Dimensiones: Modelado Dimensional (12)

por Salvador Ramos | Jul 28, 2017 | BI & Analytics | 0 Comentarios

El Modelado Dimensional es utilizado hoy en día en la mayoría de las soluciones de BI. Es una mezcla correcta de **normalización y desnormalización**, comúnmente llamada **Normalización Dimensional**. Se utiliza tanto para el diseño de Data Marts como de Data Warehouses. Y necesita que conozcas perfectamente los conceptos de **Hechos y Dimensiones**.

Básicamente hay dos tipos de tablas:

- Tablas de Dimensión (*Dimension Tables*)
- Tablas de Hechos (*Fact Tables*)

Hechos y Dimensiones: Trabajando con Hechos

Los Hechos están compuestos por los detalles del proceso de negocio a analizar, contienen datos numéricos y medidas (métricas) de Negocio a analizar. Contienen también elementos (claves externas) para contextualizar dichas medidas, como por ejemplo el producto, la fecha, el cliente, la cuenta contable, etc.

Las Tablas de Hechos (*Fact Tables*), son tablas que representan dicho proceso de negocio, por ejemplo, las ventas, las compras, las incidencias recibidas, los pagos, los apuntes contables, los clics sobre nuestro sitio web, etc. Veamos con más detalle los elementos que las componen:

- Clave principal: identifica de forma única cada fila. Al igual que en los sistemas transaccionales toda tabla debe tener una clave principal, en una tabla de hechos puede tenerla o no, y esto tiene sus pros y sus contras, pero ambas posturas son defendibles.
- Claves externas (*Foreign Keys*): apuntan hacia las claves principales (claves subrogadas) de cada una de las dimensiones que tienen relación con dicha tabla de hechos.
- Medidas (*Measures*): representan columnas que contienen datos cuantificables, numéricos, que se pueden agregar. Por ejemplo, cantidad, importe, precio, margen, número de operaciones, etc.
- Metadatos y linaje: nos permite obtener información adicional sobre la fila, como por ejemplo, que día se incorporó al Data Warehouse, de qué origen proviene (si tenemos varias fuentes), etc. No es necesario para el usuario de negocio, pero es interesante analizar en cada tabla de hechos.

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok

Read more

PK¿?Si	FK1	...	FKn	M1	...	Mn
1	1200		30	10		100257,3
2	1001		30	2		2105,05
3	12		28	5		
4	17		12	14		
5	7800		5	25		
6	14103		51	78		
...
100.000.000	254		28	65		

Las tablas de hechos, habitualmente son muy estrechas, tienen pocas columnas, además éstas son en su mayoría numéricas y de una longitud corta, de muy pocos bytes. Aunque sí que suelen ser muy largas, tienen un gran número de filas.

ProductoFK	FechaVentaFK	FechaEnvioFK	FechaCobroFK	TiendaSK	ComercialSK	Albaran	Cantidad	PrecioUnita...	Descue...	PrecioCosto	Impuestos	OrigenFila	FechaOper
218	20050701	20050713	20050708	676	285	SO43659	6	5,70	0	3,3963	2,736	3	20050702
223	20050701	20050713	20050708	676	285	SO43659	2	5,1865	0	5,7052	0,8298	3	20050702
220	20050701	20050713	20050708	676	285	SO43659	4	20,1865	0	12,0278	6,4597	3	20050702
326	20050701	20050713	20050708	117	285	SO43660	1	419,4589	0	413,1463	33,5567	3	20050702
319	20050701	20050713	20050708	117	285	SO43660	1	874,794	0	884,7083	69,9835	3	20050702
300	20050701	20050713	20050708	442	288	SO43661	1	809,76	0	699,0928	64,7808	3	20050702
296	20050701	20050713	20050708	442	288	SO43661	1	714,7043	0	617,0281	57,1763	3	20050702
304	20050701	20050713	20050708	442	288	SO43661	2	714,7043	0	617,0281	114,3527	3	20050702
223	20050701	20050713	20050708	442	288	SO43661	4	5,1865	0	5,7052	1,6597	3	20050702

En la figura anterior se muestra la tabla de hechos de ventas, en la que podemos apreciar que tiene una serie de claves externas a las dimensiones: Producto, Fecha (hay varias fechas), Tienda y Comercial. Adicionalmente tiene el número de venta; luego una serie de medidas: Cantidad, Precio Unitario, Descuento, PrecioCosto e Impuestos. Y por último un par de columnas de metadatos: OrigenFila (identifica el sistema de origen desde el que se obtuvo dicha fila) y FechaOper (fecha en la que la fila entró en la tabla de hechos).

Es importante a la hora de diseñar una tabla de hechos, tener en cuenta el nivel de granularidad que va a tener, es decir, el nivel de detalle más atómico que vamos a encontrar de los datos. No es lo mismo tener una fila por cada venta, que una fila donde se indiquen las ventas del día para cada artículo y tienda. A mayor granularidad, mayor será el número de filas de nuestra tabla de hechos, y dado que el espacio en disco y rendimiento no se ven notablemente afectados en los sistemas actuales, debemos llegar siempre al máximo nivel de granularidad que resulte útil a los usuarios.

Uno de los errores más frecuentes y que debemos evitar en nuestros diseños es añadir dimensiones en una tabla de hechos antes de definir su granularidad.

La agregación es el proceso por el cual se resumen los datos a partir de las filas de detalle de máxima granularidad. Hoy en día disponemos de sistemas OLAP y sistemas In-Memory que se encargan de agregar dichos datos y ofrecerlos al usuario con una gran rapidez y eficacia.

Hechos y Dimensiones: Trabajando con Dimensiones

Las **Dimensiones** nos permiten contextualizar los hechos, agregando diferentes perspectivas de análisis a ellos. Si una agregación de una medida dos devuelve el valor 17.538 unidades, por sí sólo no nos dice nada, en cambio, si le agregamos las perspectivas tiempo, tienda y cliente, podríamos

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok

Read more

Las **Tablas de Dimensiones** son las almacenan la información de las dimensiones. Una dimensión contiene una serie de atributos o características, por las cuales podemos agrupar, rebanar o filtrar la información. A veces estos atributos están organizados en **jerarquías** que permiten analizar los datos de forma agrupada, dicha agrupación se realiza mediante relaciones uno a muchos (1:N). Por ejemplo, en una dimensión Fecha es fácil que encontremos una jerarquía formada por los atributos Año, Mes y Día, otra por Año, Semana y Día; en una dimensión Producto podemos encontrarnos una jerarquía formada por los atributos Categoría, Subcategoría y Producto. Como se ha podido comprobar en los ejemplos, se puede dar el caso de que exista más de una jerarquía para una misma dimensión.

Las tablas de dimensiones, por lo general son muy anchas (contienen muchos atributos, y éstos pueden tener bastantes caracteres cada uno) y cortas (suelen tener pocas filas). Por ejemplo en una dimensión Producto, podemos encontrarnos con que tiene varias decenas de atributos, y que éstos están desnormalizados. No es extraño encontrarnos aquí valores ya en texto, como claves a otras tablas, por ejemplo categoría ('alimentación', 'textil', etc.), subcategoría ('congelados', 'frescos', 'bebidas', etc.), colores ('rojo', 'verde', 'amarillo', 'azul', etc.), tamaño ('pequeño', 'mediano', 'grande'), etc.

ProductoSK	ProductoBK	Categoría	Subcategoría	Producto	Peso	Tamaño	Color	Coste	FechaInicio	FechaFin
350	BK-M82B-44	Bicicleta	Bicicleta de montaña	Montaña: 100, negra, 44	21,13	44	Black	1898,0944	2005-07-01 ...	2006-06-30
606	BK-M82B-44	Bicicleta	Bicicleta de paseo	Montaña: 100, negra, 44	21,13	44	Black	1898,0944	2007-06-30 ...	NULL
352	BK-M68S-38	Bicicleta	Bicicleta de montaña	Montaña: 200, plateada, 38	23,35	38	Silver	1117,8559	2006-07-01 ...	2007-06-30
581	BK-M68S-38	Bicicleta	Bicicleta de paseo	Montaña: 200, plateada, 38	23,35	38	Silver	1117,8559	2007-06-30 ...	NULL
353	BK-R79Y-42	Bicicleta	Bicicleta de montaña	Montaña: 200, plateada, 38	23,35	38	Silver	1265,6195	2007-06-30 ...	NULL

En una tabla de dimensiones, habitualmente no es posible utilizar la clave de negocio (*business key*) como clave principal (*primary key*), e incluso en el caso de que sea posible no es aconsejable. Una clave de negocio como clave principal no es aconsejable en muchos casos por temas de rendimiento, ya que desde este punto de vista es recomendable utilizar números enteros de pocos bytes (en el caso concreto de SQL Server es muy recomendable usar el tipo de datos INT que ocupa 4 bytes). Si en el sistema transaccional tenemos, por ejemplo, una clave principal que es un char(10) siempre será más óptimo utilizar un tipo de datos numérico de menos bytes como clave principal en nuestra tabla de dimensiones. Adicionalmente hay casos en los que ya no es tan sólo un tema de rendimiento, sino que simplemente no es viable utilizar como clave principal de la tabla de dimensiones la clave principal de la tabla del transaccional. Por ejemplo, si tenemos varios orígenes de datos que queremos consolidar en una misma tabla de dimensiones, y cada uno utiliza un tipo de datos o longitud diferente, o simplemente para un mismo elemento en cada tabla de origen tiene un valor diferente. Por otro lado cuando queremos almacenar el historial de cambios, cumpliendo las características básicas que debe cumplir un Data Warehouse según *Inmon*, necesitamos tener, en muchos casos, varias filas con las diferentes versiones de los atributos y el periodo durante el que han estado vigentes, lo que implica que en la tabla de dimensiones habrá duplicidades en la clave de negocio, lo que impide que ésta pueda ser la clave principal.

Por tanto, lo habitual es que optemos por tener una clave principal diferente, esta clave es lo que se conoce con el nombre de clave subrogada.

Una **Clave subrogada** (*subrogate key*) es un identificador único que es asignado a cada fila de la tabla de dimensiones, en definitiva, será su clave principal. Esta clave no tiene ningún sentido a nivel de negocio, pero la necesitamos para identificar de forma única cada una de las filas. Son siempre de tipo numérico, y habitualmente también son autoincrementales. En el caso de SQL Server recomendamos que sean de tipo INT con la propiedad *identity* activada (es una recomendación genérica, a la que siempre habrá excepciones).

Una **Clave de negocio** (*business key*) es una clave que actúa como *primary key* en nuestro origen de datos, y es con la que el usuario está

adelante al explicar el concepto de *Slowly Changing Dimensions*.

Hasta aquí la entrada de hoy, hemos definido las tablas de hechos y dimensiones, en la próxima entrada veremos el concepto de Slowly Changing Dimensions.

[About](#)[Latest Posts](#)**Salvador Ramos**

Consultor, Formador y Mentor en Business Intelligence. SQL Server MVP.

Director de Formación en SolidQ.

Microsoft MCSE 2012: Business Intelligence.



Máster en BI & Analytics con tecnologías Microsoft

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok

Read more