# Modelo Dimensional vs Modelo Tabular

Nelia Escalante Marón
2014049551
UPT – Ingeniería de Sistemas
EPIS
Tacna, Perú

Christian Cespedes Medina
2010036256
UPT – Ingeniería de Sistemas
EPIS
Tacna, Perú

Yerson Coaquira Calizaya
2015053225
UPT – Ingeniería de Sistemas
EPIS
Tacna, Perú

Javier Octavio Arteaga Ramos
2007028981
UPT – Ingeniería de Sistemas
EPIS
Tacna, Perú

Flor Condori Gutierrez
2015053227
UPT – Ingeniería de Sistemas
EPIS
Tacna, Perú

***Resumen***— Desde el siglo pasado se ha investigado en aras de incrementar la eficiencia en el almacenamiento y el acceso a las bases de datos analíticas, sobre cuyos resultados las grandes compañías han introducido productos comerciales. En este escenario, Microsoft SQL Server 2012 ofrece dos opciones independientes para la creación de los modelos analíticos, el modelo multidimensional y el reciente modelo tabular. En este informe se profundiza en las características y potencialidades de cada uno, proponiendo los criterios más importantes que, a juicio de los autores, se deben tener en cuenta al emprender un nuevo proyecto. Se propone además una solución computacional que brinda a los especialistas y ejecutivos tanto visiones particulares como integradoras del estado del negocio, aprovechándose las facilidades recientes que proporciona la plataforma de Inteligencia de Negocios de Microsoft para la implementación de ambos modelos, dimensional y tabular.

***Abstract***— *Since the last century, research has been carried out in order to increase efficiency in storage and access to analytical databases, on the results of which large companies have introduced commercial products. In this scenario, Microsoft SQL Server 2012 offers two independent options for the creation of analytical models, the multidimensional model and the recent tabular model. This report delves into the characteristics and potential of each one, proposing the most important criteria that, in the opinion of the authors, should be taken into account when undertaking a new project. It is also proposed a computational solution that provides specialists and executives with both particular views and integrating the state of the business, taking advantage of the recent facilities provided by the Microsoft Business Intelligence platform for the implementation of both dimensional and tabular models.*

***Keybwords***— Modelos, Inteligencia de Negocios, Dimensional, Tabular.

---

# 1. INTRODUCCION

El siguiente trabajo fue hecho en consecuencia de un trabajo encargado para el curso de Inteligencia de Negocios.

Consta de 2 partes: EL marco teórico y las referencias, la investigacion se ha hecho para la comparativa de dos tipos de modelados de tablas de base de datos en base a SQL: Modelado Dimensional y el Modelado Tabular.

Unas de las principales caracteristicas que distiguen al modelo Tabular es que a nivel de consultas es muchisimo más veloz, como tambien que no necesita generar Agregaciones lo que simplifica el tiempo de procesamiento.

En caso del modelo Dimensional su uso es necesario en caso que se quiera "jerarquizar"las tablas de un base de datos. En lo que destaca este modelo es su modo óptimo de organizar datos en los sistemas de Bussiness Intelligence y lo mas destacable es que lo puede hacer mediante base de datos relacionales (ROLAP) o Base de Datos Dimensional (MOLAP). Se representa graficamente creando .estrellas.o çubos".

# 2. MARCO TEÓRICO

## 2.1. Modelo Tabular

Los modelos tabulares son bases de datos "en memoria" de Analysis Services. Gracias a los algoritmos de compresión avanzados y al procesador de consultas multiproceso, el motor analítico en memoria xVelocity (VertiPaq) ofrece un acceso rápido a los objetos y los datos de los modelos tabulares para aplicaciones cliente de reportes como Microsoft Excel y Microsoft Power View.

Los modelos tabulares admiten el acceso a los datos mediante dos modos: modo de almacenamiento en caché y modo DirectQuery. En el modo de almacenamiento en caché, puede integrar datos de varios orígenes como bases de datos relacionales, fuentes de distribución de datos y archivos de texto planos. En el modo DirectQuery, puede omitir el modelo en memoria, lo que permite a las aplicaciones cliente consultar los datos directamente en el origen relacional (SQL Server). Analysis Services proporciona funciones de procesamiento analítico en línea (OLAP) y minería de datos para aplicaciones de Business Intelligence.

Los proyectos multidimensionales si bien les falta mucho para poder ser tan estables como las bases de datos transaccionales están en una etapa más avanzada de desarrollo y grandes empresas ya lo utilizan.[Sánchez et al. (2015)]

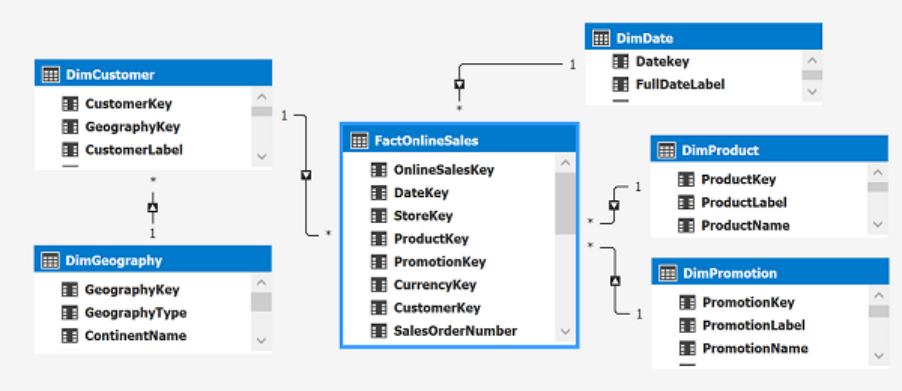


Figura 1: Modelo Tabular en SQL Server

### 2.1.1. Ventajas del Modelo Tabular

- Mucho más veloz en consultas.
- No requiere generar Aggregations (agregaciones) por lo que se simplifica el tiempo de procesamiento.
- Gracias al DAX (el lenguaje para acceder a los datos equivalente al MDX), tiene mayor flexibilidad para obtener información.
- Es intuitivo por lo que es mucho más rápido y fácil de entender e implementar.
- Se basa en modelos relacionales.

### 2.1.2. Problemas del Modelo Tabular

- Las particiones no se procesaban en paralelo si no secuencialmente, lo que hace que sea más lento el procesamiento.
- No se pueden usar múltiples idiomas.
- Si son muchos datos tarda bastante en manejar configuraciones de diferentes particiones.
- El modelo tabular acapara demasiada memoria RAM y a su vez es dependiente de tal que afectará a otras aplicaciones

### 2.1.3. Sugerencias

- Primeramente, si ya se tiene una base de datos multidimensional, no se recomienda moverse a base de datos tabulares.
- El hardware requerido para un proyecto tabular es muy diferente al requerido por un proyecto multidimensional. Por la compresión de datos, requiere menos disco una modelo tabular, pero requiere mucha más memoria RAM porque todo lo usa en memoria. En general, se necesita un buen CPU y memoria.

- Los modelos tabulares consumen muchos recursos, por lo que se recomienda hacer pruebas del funcionamiento en un servidor de desarrollo y no en producción.
- Se puede tener un modelo tabular y uno multidimensional instalados en la misma máquina, pero no es recomendable hacerlo en producción.

### 2.1.4. Jerarquías

Las jerarquías, en los modelos tabulares, son metadatos que definen las relaciones entre dos o más columnas de una tabla. Las jerarquías facilitan la navegación de los usuarios del cliente y su inclusión en un informe.

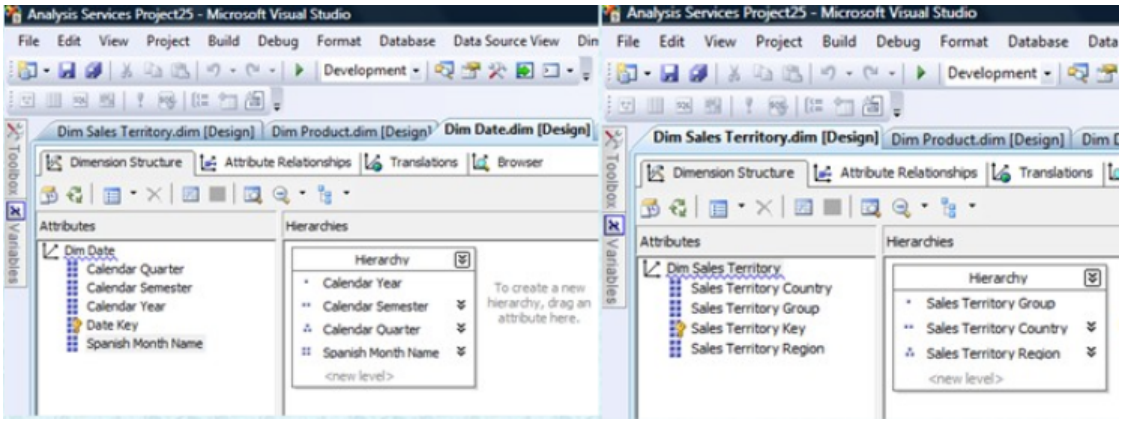


Figura 2: Jerarquías en el Modelo Tabular

Es una colección de niveles basados en atributos. Por ejemplo, una jerarquía de tiempo puede contener los niveles año, trimestre, mes, semana y día (ver imagen). Los usuarios finales pueden utilizar una jerarquía para examinar los datos del cubo. Se pueden revisar las propiedades de la Jerarquía con el menú contextual.


Figura 3: Método para mostrar las propiedades

En la siguiente tabla se describen las propiedades de una jerarquía definida por el usuario.

| Propiedad | Descripción |
|---|---|
| AllMemberName | Contiene el título en el idioma predeterminado para el miembro All de la jerarquía. |
| AllowDuplicateNames | Determina si se permiten nombres duplicados en la jerarquía. Los valores son True y False. El valor predeterminado es True. |
| Descripción | Contiene la descripción de la jerarquía. |
| DisplayFolder | Especifica la carpeta en la que se muestra la jerarquía a los usuarios. |
| ID | Contiene el identificador único (Id.) de la jerarquía. |
| MemberNamesUnique | Determina si los nombres de miembro de la jerarquía deben ser únicos. Los valores son True y False. El valor predeterminado es False. |
| Nombre | Contiene el nombre de la jerarquía. |

Figura 4: Propiedades de una jerarquía

### 2.1.5. Tablas y relaciones en el modelo tabular en Power Bi

Cuando estamos desarrollando un proyecto de Business Intelligence con Power BI el punto fuerte de nuestro trabajo es garantizar que el modelo de datos esté debidamente diseñado para que funcione correctamente. Hay que mirar las tablas del modelo y comprobar la correcta definición de las relaciones entre ellas. Todo el tiempo que dediques a comprobar todas y cada una de las relaciones jugará a tu favor a la hora de crear las visualizaciones y los informes analíticos de tu proyecto de BI. Fíjate en la posición de cada tabla, en el lado que ocupan de la relación, Uno vs Muchos, en las columnas que intentas utilizar y si falla, comprueba que es la columna adecuada y que tiene el tipo adecuado.[Cuevas et al. (2016)]

A. Si la tabla está del lado Uno de la relación, todos los valores de la columna que utilizas son distintos y no hay nulos.

B. Una vez que estés en la fase de visualización si obtienes resultados inesperados, como que el dato no se segmenta para cada filtro o que tienes valores nulos que no deben existir, regresa a comprobar la calidad de las relaciones definidas en tu modelo tabular. Por lo general allí encontrarás el problema y podrás darle solución.

C. Hay dos aspectos que se admiten en la definición del modelo, las relaciones **Muchos a Muchos** y la Dirección de filtro cruzado. No te lo recomiendo, puede ser muy problemático, a menos que ya seas un experto y estés muy seguro de todas las implicaciones de utilizar este tipo de configuración.

D. No te compliques, no le hagas a Power BI más difícil el trabajo. Casi siempre, por no ser absoluta, es posible evitarlo. El modelo tabular es suficientemente rico y flexible como para permitirnos dar solución a los escenarios más complejos. Y si no queda otra, hay que informarse muy bien de los posibles inconvenientes de este tipo de diseño.

E. Un elemento interesante es que si el modelo es complejo, podemos crear esquemas que segmenten la complejidad y muestren partes del modelo, lo que es bastante más cómodo de trabajar.

## 2.2. Modelo Dimensional

Hay un amplio acuerdo entre los usuarios de DataWarehouse de que el modelamiento dimensional es la mejor forma de presentar la información, porque es la mejor manera de reunir las principales metas de diseño:

- Presentar la información a los usuarios en la forma más simple posible.
- Retornar los resultados a los usuarios lo más rápido posible.
- Proveer información relevante que guarde pistas de los procesos subyacentes.

El modelo dimensional es mucho más fácil de entender para los usuarios que un sistema basado en un modelo normalizado de un sistema fuente típico, aunque un modelo dimensional típicamente contiene exactamente la misma información que un modelo normalizado.

Un modelo dimensional tiene menos tablas y la información es agrupada en categorías de negocio coherentes que tienen sentido para los usuarios. Estas categorías ayudan a los usuarios a navegar por el modelo ya que categorías enteras pueden ser pasadas por alto si no son útiles para un determinado análisis

***Según (Musso, 2012)***, el Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento.

***Según (Musso, 2012)***, cada Modelo Dimensional está compuesto por una tabla con una llave combinada, llamada tabla de hechos, y con un conjunto de tablas más pequeñas llamadas tablas de dimensiones, como se muestra en la Figura 2. Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Hechos:** es una colección de piezas de datos y datos de contexto. Cada hecho representa una parte del negocio, una transacción o un evento.
- **Dimensiones:** es una colección de miembros, unidades o individuos del mismo tipo.
- **Medidas:** son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.

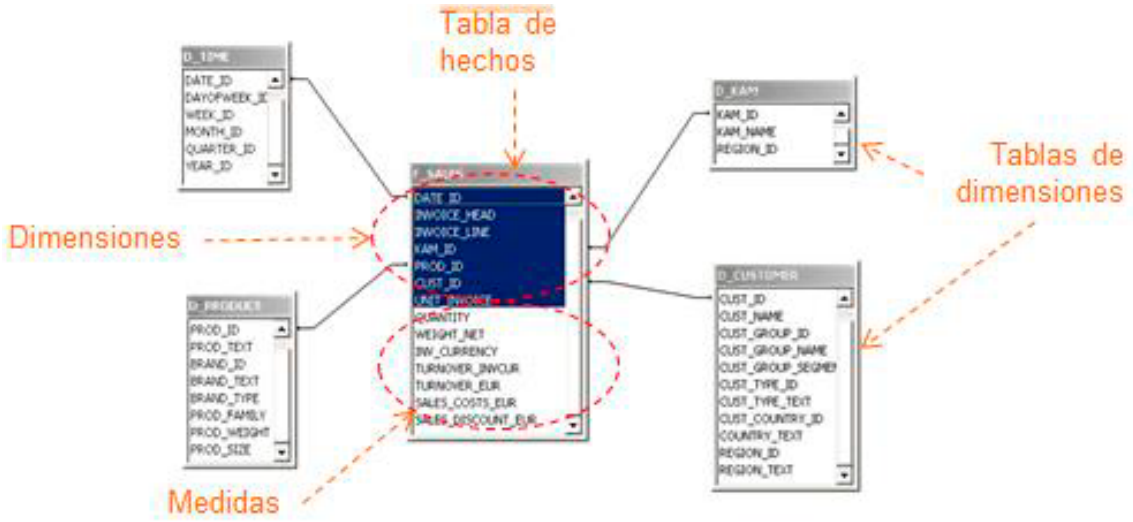


Figura 5: Deficion del modelo Dimensional

Cada punto de entrada a la tabla de hechos está conectado a una dimensión, lo que permite determinar el contexto de los hechos.

***Por lo que (Musso, 2012)*** señala que dado que es muy común representar a un modelo dimensional como una tabla de hechos rodeada por las tablas de dimensiones, frecuentemente se le denomina también modelo estrella o esquema de estrella-unión.

### 2.2.1. Características del modelo de dimensional

- Contiene métricas numéricas del negocio.
- Puede contener grandes volúmenes de datos.
- Puede crecer rápidamente.
- Puede contener datos base, derivados y resumidos.
- Son típicamente aditivos.

### 2.2.2. Modelado dimensional

Modelado dimensional (DM en inglés) nombra a un conjunto de técnicas y conceptos utilizados en el diseño de almacenes de datos. Se considera que es diferente del Modelo entidad-relación. El modelado de dimensiones no implica necesariamente una base de datos relacional, el mismo enfoque de modelado, a nivel lógico, se puede utilizar para cualquier forma física, tal como archivos de base de datos multidimensional o planas.

Según el consultor de almacenamiento de datos Ralph Kimball,1 el modelado dimensional es una técnica de diseño de bases de datos destinadas a apoyar a las consultas de los usuarios finales en un almacén de datos. Se orienta en torno a la comprensibilidad y rendimiento. Según él, aunque el modelo entidad relacional orientado a transacciones es muy útil para la captura de transacción, se debe evitar en la entrega al usuario final.

El modelado dimensional siempre utiliza los conceptos de hechos (medidas) y dimensiones (contexto). Los hechos son normalmente (pero no siempre) los valores numéricos que se pueden agregar, y las dimensiones son grupos de jerarquías y descriptores que definen los hechos. Por ejemplo, la cantidad de ventas es un hecho; marca de tiempo, producto, NoRegistro, NoTienda, etc., son elementos de dimensiones. Los modelos dimensionales son construidos por el área de proceso de negocio, por ejemplo, las ventas en tiendas, inventarios, reclamaciones, etc. Debido a que las diferentes áreas de proceso de negocio comparten algunas pero no todas las dimensiones, la eficiencia en el diseño, la operación y la coherencia, se logra usando tablas de dimensión, es decir, utilizando una copia de la dimensión compartida. El término "tablas de dimensión"se originó por Ralph Kimball.[Ramos (2016)]

### 2.2.3. Proceso de modelado dimensional

El modelo tridimensional se construye sobre uncon dimensiones de la tabla de hechospara construir el esquema, el siguiente modelo de diseño se utiliza:

A. Escoger el proceso de negocio:
   El proceso de modelización dimensional se basa en un método de diseño de 4 pasos que ayuda a asegurar la facilidad de uso del modelo dimensional y el uso del almacén de datos. Los fundamentos del diseño se basan en el proceso de negocio real que debe cubrir el almacén de datos. Por lo tanto, el primer paso en el modelo es describir el proceso de negocio en él se basa el modelo. Esto podría ser por ejemplo una situación de ventas en una tienda al por menor. Para describir el proceso de negocio, se puede optar por hacer esto en texto plano o utilizar Notación de Modelado de Procesos de Negocio (BPMN en inglés) u otras guías de diseño, como el Lenguaje Unificado de Modelado (UML en inglés).

B. Declarar el "grain":
Después de describir el Proceso de Negocio, el siguiente paso en el diseño es declarar el "grain"del modelo. El "grain"del modelo es la descripción exacta de lo que el modelo dimensional debería concentrarse. Para aclarar lo que significa el "grain", usted debe escoger el proceso central y describirlo con una sola oración. Además el "grain"(oración) es a lo que se le va a construir sus dimensiones y tabla de hechos. Puede que le resulte necesario volver a este paso para alterar el "grain"debido a nueva información obtenida en lo que su modelo supone entregar.

C. Identificar las dimensiones:
El tercer paso en el proceso de diseño es definir las dimensiones del modelo. Las dimensiones deben ser definidas dentro del "grain"de la segunda etapa del proceso de modelado dimensional de 4 pasos.

Las dimensiones son la base de la tabla de hechos, y es donde se recogen los datos de la tabla de hechos. Normalmente las dimensiones son sustantivos, como fecha, tienda, inventario, etc. Estas dimensiones son donde se almacenan todos los datos. Por ejemplo, la dimensión fecha podría contener datos tales como año, mes y día de la semana.

D. Identificar los hechos:
Después de definir las dimensiones, el siguiente paso en el proceso es crear las llaves de la tabla de hechos. Este paso es identificar los hechos numéricos que poblarán cada fila de la tabla de hechos. Este paso está estrechamente relacionado con los usuarios de negocio del sistema, ya que es donde consiguen el acceso a los datos almacenados en el almacén de datos. Por lo tanto la mayor parte de las filas de la tabla de hecho son cifras numéricas, aditivos tales como cantidad o costo por unidad, etc.[Cedeño Trujillo (2006)]

### 2.2.4. Normalización de Dimensión

La Normalización de Dimensión elimina atributos redundantes. Las dimensiones están estrictamente unidas en las sub-dimensiones.Tiene una influencia en la estructura de datos que difiere de muchas filosofías de almacenes de datos.
Los desarrolladores a menudo no normalizan las dimensiones debido a varias razones:

- La normalización hace la estructura de datos más compleja.
- El tiempo de ejecución puede ser más lento, debido a las muchas uniones entre tablas.
- El ahorro de espacio es mínimo.
- Los Mapeados de bits de índices no pueden utilizarse.
- Los Mapeados de bits de índices no pueden utilizarse.
- Los Mapeados de bits de índices no pueden utilizarse.

El rendimiento de consultas, bases de datos 3NF sufren de problemas de rendimiento cuando se agregan o recuperar muchos valores dimensionales que pueden requerir análisis.

Hay algunos argumentos sobre por qué la normalización puede ser útil. Puede ser una ventaja cuando una parte de la jerarquía es común a más de una dimensión. Por ejemplo, una dimensión geográfica puede ser reutilizable porque tanto las dimensiones de los clientes y los proveedores la utilizan.

### 2.2.5. Beneficios del modelado dimensional

Los beneficios del modelado dimensional son los siguientes:

A. Comprensibilidad - En comparación con el modelo normalizado, el modelo dimensional es más fácil de entender y más intuitivo. En los modelos dimensionales, la información se agrupa en dimensiones coherentes, por lo que es más fácil de leer e interpretar. La simplicidad también permite al software navegar las bases de datos de manera eficiente. En los modelos normalizados, los datos se divide en muchas entidades discretas e incluso un proceso de negocio simple podría resultar en docenas de tablas unidas entre sí de una manera compleja.

B. El rendimiento de consultas - Los modelos dimensionales están más desnormalizados y optimizados para las consultas de datos, mientras que los modelos normalizados buscan eliminar redundancias de datos y están optimizados para la carga de transacciones y actualización. El marco predecible de un modelo dimensional permite a la base de datos hacer fuertes supuestos sobre los datos, por lo que puede tener un impacto positivo en el rendimiento. Cada dimensión es el equivalente a un punto de entrada en la tabla de hechos, y esta estructura simétrica permite un manejo eficaz de consultas complejas. La optimización de consulta se hace simple, predecible y controlable.

C. Extensibilidad - Los modelos dimensionales son escalables y fácilmente acomodados a nuevos datos inesperados. Las tablas existentes pueden ser cambiados, ya sea por la simple adición de nuevas filas de datos en la tabla o ejecutar en SQL algún comando de tipo .Alter Table". Las consultas o aplicaciones montadas sobre el almacén de datos no necesitan ser reprogramadas para acomodarse a los nuevos cambios. Las consultas y aplicaciones antiguas continúan funcionando sin producir resultados diferentes. Pero en los modelos normalizados cada modificación se debe considerar cuidadosamente, debido a las complejas dependencias entre las tablas de bases de datos.

## Referencias


Cedeño Trujillo, A. (2006). Modelo multidimensional. *Ingeniería Industrial*, 27(1).

Cuevas, A. S., Sanchez, M. T., Hernandez, L. G., Cuevas, A. S., and Suarez, R. R. (2016). Comparing tabular and multidimentional model in a real bi solution. *IEEE Latin America Transactions*, 14(7):3393–3399.

Ramos, S. (2016). Data warehouse, data marts y modelos dimensionales. *SolidQ Global*.

Sánchez, M. T., Cervantes, Y. E., Cuevas, A. S., Hernández, L. G., and Cuevas, A. J. S. (2015). Modelación tabular: una alternativa sugerente para el análisis de los datos. *Ciencias de la Información*, 46(1):3–10.