novática
CEPIS UPGRADE Revista de la Asociación de Técnicos de Informática

Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software). ***Novática*** co-edita asimismo **UP**GRADE, revista digital de **CEPIS** (*Council of European Professional Informatics Societies*), en lengua inglesa, y es miembro fundador de **UP**ENET (***UP****GRADE* **E***uropean* **NET***work*).

<http://www.ati.es/novatica/>
<http://www.ati.es/reicis/>
<http://www.cepis.org/upgrade>

ATI es miembro fundador de **CEPIS** (*Council of European Professional Informatics Societies*) y es representante de España en **IFIP** (*International Federation for Information Processing*); tiene un acuerdo de colaboración con **ACM** (*Association for Computing Machinery*), así como acuerdos de vinculación o colaboración con **AdaSpain**, **AI2**, **ASTIC**, **RITSI** e **Hispalinux**, junto a la que participa en **ProInnova**.

Coordinación Editorial, Redacción Central y Redacción ATI Madrid
Padilla 66, 3º, dcha., 28006 Madrid
Tlfn.914029391; fax.913093685 <novatica@ati.es>
Composición, Edición y Redacción ATI Valencia
Av. del Reino de Valencia 23, 46005 Valencia
Tlfn./fax 963330392 <secreval@ati.es>
Administración y Redacción ATI Cataluña
Via Laietana 46, ppal. 1ª, 08003 Barcelona
Tlfn.934125235; fax 934127713 <secregen@ati.es>
Redacción ATI Aragón
Lagasca 9, 3-B, 50006 Zaragoza.
Tlfn./fax 976235181 <secreara@ati.es>
Redacción ATI Andalucía <secreand@ati.es>
Redacción ATI Galicia <secregal@ati.es>
Suscripción y Ventas <http://www.ati.es/novatica/interes.html>, ATI Cataluña, ATI Madrid
Publicidad
Padilla 66, 3º, dcha., 28006 Madrid
Tlnf.914029391; fax.913093685 <novatica@ati.es>
Imprenta: Derra S.A., Juan de Austria 66, 08005 Barcelona.
Depósito legal: B 15.154-1975 -- ISSN: 0211-2124; CODEN NOVAEC
Portada: Resolución en marcha - Concha Arias Pérez / © ATI
Diseño: Fernando Agresta / © ATI 2003

Nº 211, mayo-junio 2011, año XXXVII

sumario

Tema del próximo número: "*Innovación y emprendimiento en Informática*"

Modelos de construcción de *Data Warehouses*

José María Arce Argos
Gerente de Business Intelligence & CRM en Oesía

<chema.arce.argos@gmail.com>

Resumen: *A través del presente artículo nos adentraremos en las técnicas, modelos y consejos más básicos para crear valor en nuestras organizaciones. Todo ello de la mano del Data Warehouse y de sus posibles modelos de implementación. El artículo profundiza, comparando algunos de los tipos de modelos posibles y sus consecuencias. Prestando especial atención en la necesidad de las empresas de poder medir, como única salida hacia el éxito empresarial.*

Palabras clave: *Datawarehouse, modelo en estrella, modelo E-R, modelos multidimensionales.*

Autor

José María Arce Argos es Gerente de *Business Intelligence & CRM* en Oesía**.** Ha desarrollado su carrera profesional en empresas tan importantes como Leinsa, IBM, Ernst & Young, SAS Institute, Bull España, Altran SDB e Inad. Cuenta con 22 años de experiencia en consultoría, de los cuales 16 años dedicados al *Business Intelligence*. Redactor y colaborador en la revista "Gestión del Rendimiento", colaborador en TodoBI, profesor durante 10 años en el Máster de "Sistemas de Información e Investigación de Mercados" de la *Business Marketing School* (ESIC), ponente en diversos eventos en IIR, SAS, CUORE Oracle, Univ. de Zaragoza, etc. A lo largo de su trayectoria profesional ha diseñado, coordinado, participado o dirigido iniciativas de BI en clientes como Publiespaña, British American Tobacco, Halifax, ONT, REE, Continente (Carrefour), Grupo Abengoa, ABN Anro Bank, B. Santander, Open Bank, Ing Direct, Dirección Gral. de Estadística (JCYL), TeleMadrid, Hospital San Cecilio de Granada, DFA, RED.ES, Egailan, INEM, INSS, Policía Municipal de Madrid, Canal de Isabel II, etc. Es autor del Blog BIB: <http://josemariaarce.blogspot.com/>.

1. Introducción

Es complejo intentar desvelar, a través de unas líneas todo lo que puede incluir un modelo de *Data Warehouse*[1] (DW) e incluso llegar a un acuerdo en su significado y tipo de estructuración física.

En cualquier caso, a estas alturas de la madurez de *Business Intelligence* (BI) en España, todos coincidimos en que el único pilar básico para el desarrollo de soluciones de negocio es sin lugar a dudas el DW. Almacenamiento pensado, diseñado y construido por y para unas necesidades agresivas de análisis, análisis completamente impredecibles.

En España llevamos más de 16 años hablando de DW y de BI, aunque lamentablemente tras las siglas DW se esconde un gran desconocido, obviando su objetivo, sus técnicas y sus posibles tipos de modelos. Desconocido al haberse puesto de moda desde 1998 otros "tecnicismos" informáticos (*Customer Intelligence*, BSC, CRM, etc.).

Las nuevas herramientas, más sofisticadas, han derivado erróneamente en profesionales más técnicos y con menos visión del negocio, obviando temas muy relacionados con los modelos de datos de DW y su representación formal en un modelo especifico. En cualquier caso, empecemos por el principio.

Los modelos de un *Data Warehouse* mantienen necesariamente una relación directa con los tipos de almacenamiento que se determinen e incluso deben y pueden ser optimizados según decenas de criterios. Los cuales van desde el tipo de consulta más demandado, optimización de acceso por claves e incluso tablas agregadas, hasta la optimización según la herramienta de explotación utilizada. Por todo ello, no es sencillo establecer un único criterio o patrón de cara a la correcta construcción de un DW. El presente artículo únicamente pretender sentar unas bases mínimas para ayudar a la selección del almacenamiento más adecuado, así como algunos consejos de cara a realizar un modelo de DW que permita un éxito en su iniciativa de BI, o por lo menos, adquirir algún conocimiento para poder comprender o entender a los consultores o responsables de desarrollarlo.

Además, es importante reflejar alguna idea previa sobre todo ello. Un buen sistema de BI requiere de un DW, de hecho casi todas las iniciativas de BI tienen por debajo o se apoyan mayoritariamente en un DW, aunque no se diga e incluso se oculte. También es cierto que las herramientas de explotación cada día son más poderosas e incluso permiten realizar agrupaciones, desgloses, etc., incluso contra un modelo, digamos cortésmente, poco evolucionado. Lo cual posiblemente denota un error en el planteamiento, pues siempre será mejor que ciertos conceptos figuren en modelo que generados al vuelo por una herramienta, pues podríamos tener el mismo concepto "N" veces y calculado de formas diferentes. En el caso de producirse esta situación implicaría la muerte del sistema BI, básicamente por la desconfianza sobre la calidad de los datos e incluso por posibles resultados contradictorios. En este ejemplo no está fallando el BI, está fallando el diseño del modelo de DW.

Para acabar esta introducción, comentar que en decenas de clientes hemos escuchando quejarse de importantes problemas de rendimiento en sus DW, de la necesidad de poner más maquina, cambiar de herramienta, meter nuevas estructuras más optimizadas, etc. En la mayoría de los casos los problemas no se solucionan así. El 90% de los problemas de los sistemas de BI residen en un mal modelo de datos (DW).

2. Almacenamientos y alternativas

Sin entrar en las características que debe cumplir un DW, pues creo que ya está todo escrito, mencionar que el DW tiene como única finalidad la capacidad de analizar la información de interés desde diversos puntos de vista, a lo largo del tiempo, entre otras muchas cosas. Ello obliga a que esté diseñado y construido de una forma específica, siendo pues necesario cumplir con el término OLAP (*OnLine Analytical Process*), en contra de los diseños tradicionales, los cuales ahora los denominamos OLTP (*OnLine Transaction Processing*).

Años antes del DW ya se intentaron montar este tipo de sistemas, a base de duplicar el mundo operacional. El tiempo demostró que dicho camino no era el más efectivo y que tenía serias limitaciones, tomando fuerza la necesidad de que el DW esté diseñado con características OLAP.

Recordemos que los sistemas OLAP cumplen con más de 50 posibilidades de "navegación" por la información, como rotar, bajar, profundizar, expandir, colapsar, etc. Todo esto y mucho más es simplemente imposible en los sistemas tradicionales (OLTP), motivo extra para olvidarse de replicar sistemas, solución ya probada (años 70) y que no dio los resultados esperados.

Existen diversas alternativas para la arquitectura de un DW. En cuanto a tipos de almacenamiento, mencionaremos las más habituales:

- Solución **M**OLAP (*Multidimensional OLAP*): Montar el almacenamiento bajo una base de datos multidimensional propietaria (MDDB de SAS, ESSABE, etc.).
- Solución **R**OLAP (*Relacional OLAP*): Montar el almacenamiento bajo una base de datos relacional (SQLServer, Oracle, DB2, etc.).
- Soluciones **H**ibridas, denominadas HOLAP. Las cuales combinan ambas alternativas en una única solución, interesante y sabia alternativa según las necesidades.

Debemos considerar que detrás de unas siglas se esconden unas claras diferencias tecnológicas y unas ventajas e inconvenientes que deben conocerse a priori. Las diferencias más llamativas o significativas las podemos clasificar en cuatro puntos:

Necesidades de procesamiento y tiempo de respuesta: Como no existen soluciones ideales cada una ofrece diversas capacidades: Mientras los modelos ROLAP deben calcular y buscar los datos al vuelo, lo cual consume tiempo y retarda la respuesta, los modelos MOLAP tienen todo precocinado y sus tiempos son mejores. Lógicamente las necesidades *batch* para crear los MOLAP son claramente superiores a los ROLAP. También mencionar que los volúmenes de datos gestionados en ROLAP son muy, pero que muy superiores a los MOLAP, a más datos más tiempo. En algún cliente he intentado crear MOLAP con importantes volúmenes y simplemente fue imposible. Cada tecnología vale para una cosa.

Gestión e incremento en las visiones de negocio: La visiones de negocio o dimensiones, utilizando términos de BI, son casi infinitos en el mundo ROLAP, pero no por el contrario en el mundo MOLAP. Los modelos MOLAP empiezan a flaquear en la frontera de las 10 dimensiones, pues al crear el MOLAP se deben de calcular todos los cruces o combinaciones posibles, por lo tanto ojo al número de combinaciones usadas.

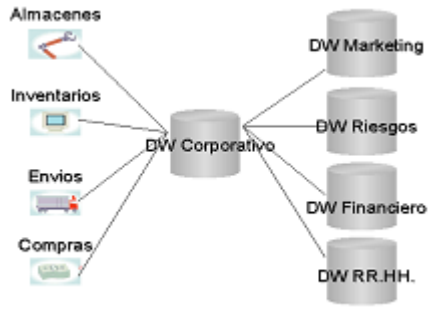

Figura 1. Arquitectura ROLAP para Data Warehouse.

Necesidades cambiantes de agrupaciones o consolidaciones: En este caso los modelos ROLAP tienen una capacidad más camaleónica, pues todo se calcula al vuelo, siendo más flexibles y potentes. Por contra, los modelos MOLAP requieren de una precocina en *batch*. En general, incluir cualquier concepto olvidado en un DW es algo laborioso, especialmente en función de donde se deba incluir. Incluir una nueva métrica en una tabla de hechos (*tabla fact*)[2] es lo más costoso, pues no vale solo con ponerla, será necesario recargar esa métrica y todo su historial asociado. Es casi mejor pecar por exceso, un olvido en una tabla de hechos se paga muy caro.

Volúmenes y capacidades: Posiblemente el éxito de los modelos ROLAP viene especialmente por esta capacidad. Son los únicos que pueden enfrentarse a volúmenes de Gb e incluso Teras de información, algo simplemente imposible en el mundo MOLAP. Lo cual lógicamente hace que sus respuestas no ofrezcan unos tiempos óptimos, pero en otras soluciones simplemente no se pueden hacer. La arquitectura ROLAP es la clara vencedora por capacidad y flexibilidad. Dejando el mundo MOLAP a soluciones o necesidades muy controladas, pequeñas, departamentales o, hablando en cristiano, de "andar por casa". Es normal encontrarse con evoluciones en base a *Data Marts*[3], decisión muy extendida e incluso razonable años atrás. Especialmente usadas en las primeras iniciativas de BI, pues resultan rápidas de desarrollar y nos permiten que en nuestras compañías, "los mayores", entiendan las capacidades y bondades de dichos sistemas. Les gustarán mucho y obtendremos financiación para continuar. Pero debemos ser conscientes de las limitaciones que tendremos a medio o largo plazo, pues tendremos silos de información y limitadas capacidades de cruces, lo cual terminará obligando a desarrollar más y más *Data Marts*.

Posiblemente habremos salido de una pesadilla operacional para entrar en una pesadilla informacional, gracias a la proliferación de *Data Marts* en nuestra organización.

La otra, o mejor dicho, la única alternativa eficaz, flexible y que siempre nos permitirá un crecimiento y mantenimiento acorde a nuestras necesidades, si hablamos de DW, entendiéndolo como una solución integrada, consolidada y de amplia cobertura empresarial (no me atrevo a decir corporativo, pero lo pienso), es montar el DW bajo arquitectura ROLAP, como se ilustra la **figura 1**.

Bajo esta figura podemos ofrecer lo mejor de ambos mundos. Dado la limitación física de los MOLAP, dejemos los cimientos del BI bajo ROLAP, y diseñemos estructuras MOLAP para accesos controlados, rápidos y eficaces a los departamentos o necesidades especificas. Como ya hemos comentado, una solución Hibrida (HOLAP) es una sabia salida que ofrece lo mejor de cada arquitectura.

Una vez tratado ligeramente el tema de los almacenamientos posibles y considerando que nos decantamos por su gestión bajo un motor relacional, debemos considerar que los modelos tradicionales no están pensados para las necesidades que debe de soportar un DW. Lo cual requiere un diseño específico, como ya hemos comentado, el cual permita su explotación OLAP. Tradicionalmente nos encontramos con dos tendencias: los modelos en estrellas y los modelos copo de nieve.

3. Tipos de modelos en DW

Tal vez antes de entrar en faena, me gustaría comentar que en definitiva hablamos de lo mismo, pues de un modelo en estrella llego a un copo de nieve y viceversa, por lo tanto, no merece la pena pegarse en qué es mejor (guerrilla promovida por los fabricantes, a mediados de los 90, y sus intereses particulares), pues finalmente son más de lo mismo.

Aquellos que han sido alumnos míos durante mis 10 años siendo profesor en un Máster de prestigio sobre "Sistemas de Información e Investigación de Mercados", bien lo saben y ha quedado demostrado. Sin embargo existen importantes limitaciones e incluso costes ocultos, como los asociados a los mantenimientos, a las mejoras, a la dificultad en los nuevos procesos, etc.

Con independencia de un tipo de modelo u otro, es conveniente quedarnos con algunas ideas que nos van a permitir buscar la mejor solución y evitar algunas trampas en el diseño de un DW. Con el ánimo de ser lo más claro posible evitaré usar tecnicismos. Entre las muchas coletillas adquiridas en estos años y relacionadas con el DW os comentaría:

- "***Divide y vencerás***": Se puede empezar a diseñar un DW sin conocer al 100% las necesidades de toda la organización, desglosa el gran proyecto por dominios de información y "ataca" uno a uno, sin perder la visión del gran sistema. Recuerda: "*Piensa en grande, haz en pequeño*". También aplicable al "*divide y vencerás*" a la forma de diseñar las dimensiones. Considerando la forma de acceso del usuario final y la "navegación" deseada, considera que cuanto más "normalizado" se diseñe tendrás más flexibilidad y un mantenimiento más sencillo que será más comprensible por los DBA más tradicionales.
- "***Diseñar cabezas de ciervo***": Existen trampas en un DW como son las *"trampas de abanico"* y *"trampas de abismo"*, las cuales exclusivamente ocurren bajo un modelo mal diseño. Un modelo simple, es como la cabeza de un ciervo. La cabeza es la tabla que contiene los valores numéricos a analizar, las métricas o indicadores, la *tabla Fact*. Mientras los

cuernos son sus dimensiones. Las dimensiones jamás se tocan o cruzan entre sí. El único punto en común que tienen son la cabeza del ciervo. Los cuernos solamente tienen un punto de contacto con la cabeza… con este simple y claro ejemplo, nunca tendréis problemas

■ "***Si metes basura, sacas basura***": Expresión asociada a la importancia de los procesos ETL (*Extract, Transform and Load*) para la carga de los datos, a sus controles de calidad del dato, etc. Todos los esfuerzos, todos los diseños y cualquier otra actividad no tendrá ningún valor, sin la credibilidad y calidad de los datos.

■ "***El exceso de análisis conduce a la parálisis***": Es más que discutible como se abordan algunos proyectos de DW, especialmente los grandes (corporativos). Como ya hemos comentado se debe fragmentar por dominios e irlo abordando o desarrollando por iteraciones, en caso contrario estamos muertos, no ofreceremos resultados nunca y tendremos otro bonito fracaso.

4. ¿Qué tipo de modelo diseño?

Como hemos comentado, aunque ambos modelos pueden derivar el uno en el otro, en función de aplicar técnicas de normalización o desnormalización, el resultado final sí tiene su importancia. Hoy por hoy casi todas las herramientas de explotación pueden "atacar" a cualquier modelo, pero solamente algunas son capaces de aprovechar al máximo las capacidades de los modelos copo de nieve, en general, todas se mueven por el mundo de las estrellas.

Tradicionalmente los fabricantes de herramientas, basadas en modelos estrellas, han sido especialmente críticos con su competencia, pues se posicionaba con otro tipo de modelo. Ello provocó y todavía se nota, mensajes muy erróneos lanzados al mercado. Del mismo modo, actualmente nos quieren vender ideas como: hacer un DW en 10 minutos, todo ello sin técnicos, o de la vital importancia de disponer de análisis en un iPad, etc., cuando la mayoría de los sistemas funcionan por los pelos y no son adecuadamente explotados. En general, las necesidades de los usuarios finales no están alineadas con las estrategias comerciales de algunos fabricantes, mal que les pese.

En función de sus necesidades, conocimientos, herramientas y estrategias pueden optar por un tipo de modelo u otro. La **figura 2** visualiza un modelo en estrella, en el centro la tabla de hechos donde se encuentran los indicadores numéricos a analizar, a su alrededor una tabla por dimensión o visión del negocio. Dentro de cada una se encuentran todos los conceptos asociados a dicha dimensión.

No soy enemigo de los modelos en estrella, pues mi maestro y casi padrino en BI fue Ralph Kimball. También es cierto que por aquellos años, aquella forma de diseñar fue una revolución en sí misma y tampoco había muchas más alternativas.

Puntualizar que en un proyecto de BI, no existe una sola estrella, de hecho, algunos hablan de constelaciones… En una base de datos de DW, podríamos encontrar decenas de *tablas fact* y sus dimensiones asociadas, comunes y no comunes entre diversas *tablas fact*. Pues en caso contrario, si piensan que con una estrella lo tienen todo resuelto, permítanme dos comentarios:

■ No malgaste su tiempo y dinero, pues algo no he sabido explicarle o para usted un fichero texto ya sirve para tomar decisiones.

■ Hágalo usted mismo, con una tabla dinámica de Excel.

Queda claro que de cara a intentar explicar temas de modelos, nos apoyamos con la expresión mínima (una sola tabla de hechos y sus dimensiones), las cual son ilustradas en las figuras, pero la realidad no es tan simple. La otra opción es el modelo copo de nieve, el cual tiene más o menos la apariencia que muestra la **figura 3**.

Los defensores de los modelos en estrella, modelos completamente desnormalizados, han basado su defensa en algunas afirmaciones que vamos a repasar, aunque el tiempo ha demostrado que "no es oro todo lo que reluce":

■ Fácilmente entendible, consultas sencillas (por usar pocas tablas). Lo comparto, pues casi cualquiera puede entender a simple vista este modelo, digamos que no asusta al verlo. Supongo que su expansión se encuentra relacionada con su relativa sencillez al diseñarlo, pues deja todo el "marrón" a la herramienta, sobre la cual se debe definir la lógica de navegación. Como ya he indicado lo considero un error potencialmente alto. Las estrellas simples sobre BBDD relacionales son lo más parecido conceptualmente a las BBDD multidimensionales, a los castigados cubos, e incluso a las tablas dinámicas del Excel, etc. Supongo que por ello, algunos defienden el Excel como una herramienta de BI, sin más comentarios.

■ Son más rápidos (por usar pocas tablas, pocos *joins*). Directamente, es una verdad a medias. Solamente podría ser así con poco volumen de información. Además, las políticas de claves, bajo este tipo de estructuras, no favorecen las búsquedas, aparte de ser francamente difícil planificar a priori la clave o claves más adecuadas, pues nunca podemos saber por dónde nos vendrá la consulta.

Otro problema más serio, y que solamente por ello **debería hacernos replantear toda la solución**, es que no es posible implementar integridad referencial. Por ello nuestro gestor nunca será capaz de garantizar

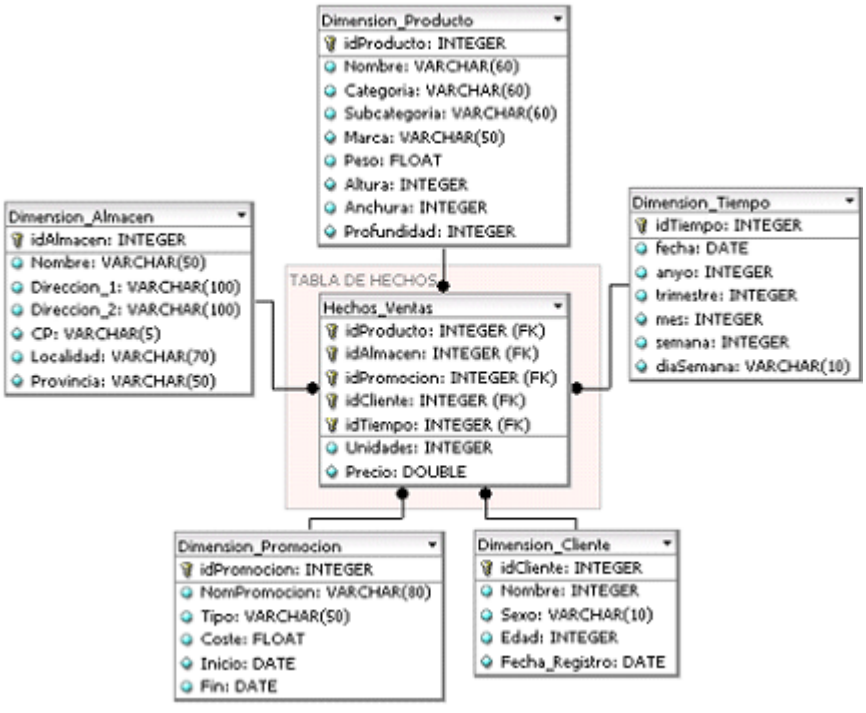

Figura 2. Modelo en estrella para un *Data Warehouse*.

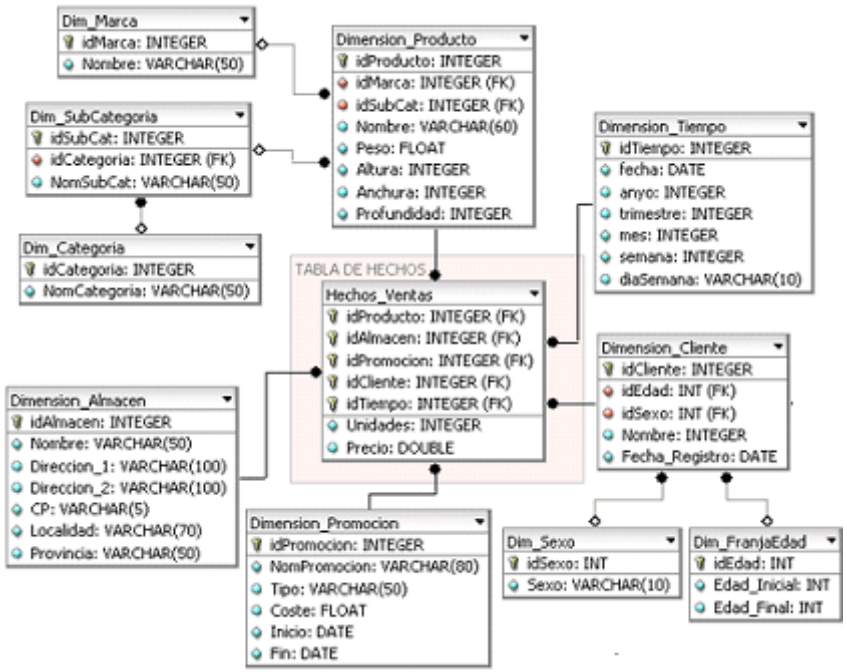

Figura 3. Modelo copo de nieve para un *Data Warehouse*.

la consistencia y la calidad del dato. Obligando a desarrollar unos procesos ETL muy estudiados y minuciosos.

En resumen, por ahorrar tiempo en el análisis, en el diseño y no complicarse la vida, algunas veces "te entregan unas estrellitas muy monas". Nadie te explicará y te pondrá sobre aviso de que:

- Ese tipo de diseño puede implicar perder calidad y consistencia en los datos, pues no puedes definir integridad referencial entre campos de una misma tabla.
- Ello ocasiona pasar el control de la consistencia a los procesos ETL, lo cual implica un esfuerzo extra y un coste elevado.
- Dicha forma de diseñar delega la lógica de la navegación a la herramienta final y no está implícita en el modelo, esto es igual a problemas tarde o temprano.
- Dificultad para determinar índices para mejorar respuesta.
- Relaciones padres e hijos no identificados.
- La velocidad que supuestamente ganamos por no hacer *joins* por diversas tablas, la podemos perder por el gran volumen de la dimensión (en única tabla) y por la posible ausencia de índices adecuados, etc.

Creo que ello, unido a otros puntos difíciles de explicar sobre papel nos debería por lo menos cuestionar qué modelo establecer y avisarnos sobre todo lo que leemos por Internet. Lo más sensato: Haga una prueba con el escenario más parecido posible a la realidad, también respecto al volumen de información.

5. Conclusión

Los negocios están cambiando constantemente debido a cambios económicos, evoluciones tecnológicas, alteraciones en el mercado, impactados por diversos cambios culturales y sociales e incluso por fenómenos meteorológicos.

Todo ello obliga a replantearse las estrategias actuales y debería provocar una transformación en nuestro propio negocio. Así, un factor clave de éxito, e incluso de supervivencia, viene derivado de la capacidad de las organizaciones de gestionar de forma eficiente sus datos, y transformarlos en información útil y disponible para acertar en las decisiones. Esto y solo esto, es *Business Intelligence.*

***Business Intelligence* no es tecnología, es negocio y es estrategia**. BI implica muchas cosas, pasando por la **vocación de medir para actuar en consecuencia,** gran problema pendiente en las organizaciones.

Actuar no es hacer un informe. Es la capacidad de controlar y gestionar las organizaciones, basada en datos e informaciones veraces y no en hipótesis. Es la capacidad de alinear la estrategia con las operaciones, es la capacidad de orientarse realmente hacia el cliente, es la capacidad de entender, es comprender y transmitir los objetivos empresariales y su desempeño, es la capacidad de crear consenso en la organización, derivando todo ello en un **cambio cultural**.

En caso contrario, nos quedaremos con unas decenas de informes, tras unas inversiones muy importantes.

BI
=
Cambio Cultural
=
Capacidad + agilidad + decisiones

Es necesaria la redefinición de las competencias del BI dentro de las organizaciones. Pues siendo equipos de vital importancia y estratégicos, tan importantes o más que departamentos de marketing, financieros, recursos humanos, etc. suelen brillar por su ausencia. Delegando sus competencias en un equipo técnico más o menos formado pero con una visión alejada del verdadero sentido del BI.
No podemos montar un cambio cultural donde no existe estrategia, aquellas empresas que han dado el salto y tiene una infraestructura real, tanto en recursos humanos como materiales, se están posicionando. Ustedes deciden…

Para terminar les dejo una frase que leí por Internet, la cual tomo prestada (gracias a su autor, aunque no recuerde quien fue): ***"En la nueva económica, el grande ya no devorará al chico, sino el ágil le ganará al lento".***

Bienvenido al apasionante mundo de los negocios, bienvenido a *Business Intelligence.*

Bibliografía

Ralph Kimball. *The data warehouse toolkit: Practical techniques for building dimensional data warehouse*, 1996.
Bill Inmon. *Building the Data Warehouse*. 1st Edition. Wiley and Sons, 1992.

Notas

[1] *Un **Data Warehouse** (DW) es una base de datos usada para generación de informes. Los datos son cargados desde los sistemas operacionales para su consulta. Pueden pasar a través de un almacén de datos operacional para operaciones adicionales antes de que sean usados en el DW para la generación de informes* (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <http://en.wikipedia.org/wiki/Data_warehouse>).

[2] *En las bases de datos, y más concretamente en un data warehouse, una **tabla de hechos** (o tabla fact) es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio.* <http://es.wikipedia.org/wiki/Tabla_de_hechos>

[3] *Un **Data mart** es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.* <http://es.wikipedia.org/wiki/Data_mart>.