

MODELO DE DATA WAREHOUSE CON APLICACION DE INTELIGENCIA DE NEGOCIOS PARA LAS PYMES

DATA WAREHOUSE MODEL WITH BUSINESS INTELLIGENCE APPLICATION FOR PYMES

¹Javier Concepción Sánchez Espinoza, ²César Aldo Canelo Sotelo

RESUMEN

La propuesta consistió en crear un modelo para diseño de una base de datos multidimensional aplicando Inteligencia de Negocios para que las PYMES la utilicen, de manera que cuando se hagan las consultas, estas sean eficientes, es decir menor tiempo de respuesta con optimización de almacenamiento en pocos pasos. Por naturaleza en las PYMES no han tenido referencia de uso en el Perú. Se construyó primero un modelo de entidad relación se pasó al modelo multidimensional, con la herramienta MS-SQL Server 2012, se implementó el modelo usando su suite Business Intelligence. Se contó con distintos niveles de agregación donde se recurrió a la jerarquía, que nos permitió agregar la información a distintos niveles. El nivel de detalle con el que se construyó el modelo dependió de las interrogantes que se respondió a partir del mismo. Se construyó normalmente al máximo nivel de detalle, por la buena explotación de la información. Los resultados han sido mostrados con un ejemplo de consolidado de la información de varias tablas, donde ha sido hecho para que los responsables tomen decisiones inmediatas. Con este modelo se dio a conocer la ventaja que tiene con otros productos desarrollados para el uso de los usuarios finales y no pueden ser modificados por los analistas. El estudio concluyó con un prototipo que pueden ser usado en las Pymes en general por los procesos comunes que tienen.

Palabras clave: Almacén de datos, cubo, dimensiones, inteligencia de negocios, tabla de hechos.

ABSTRACT

The proposal was to create a model for designing a multidimensional database by applying Business Intelligence for PYMES to use it, so that when queries are made, they are efficient, it means less response time with storage optimization in few steps. However, in Peru, there is no reference to its use in PYMES. A relationship entity model was first constructed, later a multidimensional model, with the MS-SQL Server 2012 tool and the model was implemented using its Business Intelligence suite, as well as different levels of aggregation where the hierarchy was used, this allowed adding the information at different levels. The level of detail with which the model was built depended on the questions that were answered from it. It was built normally to the maximum level of detail, by the good exploitation of the information. The results have been shown with an example of consolidating the information from several tables so those responsible make immediate decisions. With this model, the advantage it has with other products developed for the use of end users was made known and can not be modified by analysts. The study concluded with a prototype that can be applied in PYMES in general by the common processes they have.

Keywords: Data warehouse, cube, dimensions, business intelligence, table of facts.

¹Facultad de Ingeniería Industrial y de Sistemas, Departamento Académico de Sistemas y Telemática, Universidad Nacional de Ingeniería. Lima-Perú.
E-mail: : jsanchez@uni.edu.pe

²Facultad de Ingeniería Industrial y de Sistemas, Departamento Académico de Sistemas y Telemática, Universidad Nacional de Ingeniería. Lima-Perú.
E-mail: ccanelo@uni.edu.pe

INTRODUCCIÓN

En la actualidad *las pequeñas y medianas empresas PYMES*, requieren aplicar las tecnologías de información para su desarrollo. El problema es que hay una gran diversidad de rubros en las empresas, por ejemplo, las que venden al consumidor final, las que transforman, las que venden materias primas o productos sofisticados, para cualquier tipo de empresa, existen muchos beneficios. Lo más importante es primero identificarlos para estar conscientes de lo que son las tecnologías de información y ver la manera que éstas puedan ser utilizadas. La Competitividad está actualmente en las PYMES, por eso necesitan el uso de las *Tecnologías de Información y Comunicaciones TICs*. Existen esfuerzos en colocar a las PYMES en un lugar diferente a la informalidad a nivel internacional.

Organizaciones internacionales han realizado estudios que las identifican como parte importante de la economía, por la gran cantidad de estas empresas. En el Perú el 97% de empresas son Pymes. Si la Base de Datos es tradicional, es engorroso hacer varias aplicaciones para la toma de decisiones; sin embargo, actualmente hay softwares para hacer Inteligencia de Negocios, por lo tanto, se propone un modelo del *almacén de datos data warehouse* de manera que la información sea clave para la toma de decisiones. La data es la de Microsoft que es apropiada para crear el modelo propuesto como un prototipo. Las TICs pueden ayudar a la eficiencia de los tiempos de respuesta de los procesos que se involucran en la elaboración de algún bien o servicio.

Desde que se seleccionan las materias primas, pasando por el proceso de transformación, inspección de calidad, ventas, logística y pago, el sistema puede dar información sobre los procesos. En el caso de las pequeñas y medianas empresas, estos sistemas crean la posibilidad de agilizar los tiempos de espera, de entrega y los de respuesta por parte del cliente. Entre más pronto el cliente o la empresa se enteren de información notoria, será mejor la respuesta. Sin embargo, con la información registrada se pueden mostrar para hacer notar los indicadores esenciales o hacer análisis con curvas, con la consecuencia de tomar decisiones para la mejora de la organización, así incrementa su competitividad. Cuando se tienen tecnologías ya integradas, la compañía tiene que hacer más eficientes la transacción y la comunicación.

Con respecto a la bibliografía se considera prioritariamente el artículo de ESAN, también la definición de Inteligencia de Negocios por la consultora Gartner, una de las grandes existentes en Norteamérica, finalmente se cuenta con los aportes de dos grandes investigadores europeos sobre el tema, un americano y un israelita

El objetivo es presentar un modelo de diseño de Base de Datos en opción multidimensional que muestre un uso eficiente y contemple conceptos que pueden estar

al alcance de los diseñadores y usuarios aplicando Inteligencia de Negocios en las PYMES.

MATERIAL Y MÉTODOS

Se emplea el uso del software propietario MS-SQL Server. El tipo y diseño del estudio es el de investigación exploratoria, explotando al software empleado porque tiene la opción de crear la técnica de Business Intelligence, el método se centra en preparar ampliación al data warehouse a partir de la Base de Datos relacional existente.

El diseño de estudio es experimental, como ensayo de campo que al final, el resultado del estudio servirá para que las Pymes puedan entender de esta nueva opción informatizada.

La muestra utilizada consiste de la información de la Base de Datos ejemplo de Microsoft *Northwind*, consistente de exportación de productos alimenticios a diversos países. Las tablas principales son la de productos con 77 tipos, categorías de los productos que son 8, empleados de la compañía que son 9, los clientes son 91, los pedidos son 830, los proveedores son 29 y los despachadores son 3.

La mayoría de las Pymes tienen procesos parecidos, donde interesa más las ventas, además la mayoría tiene informatizadas sus procesos con proveedores, compras, productos y ventas, útiles para su contabilidad y las tributaciones. Se adecua esta data que tiene contenido similar para que las PYMES puedan utilizar este modelo propuesto. Todas estas entidades de información de las PYMES, las mayorías se manejan con software desarrollados en Base de Datos tipo Entidad Relación. Justo con el software MS SQL Server se tiene la opción de hacer Inteligencia de Negocios desde inicios de la década del 2000, Se utiliza está opción donde se convierte el modelo Relacional a Multidimensional con las opciones del software y finalmente, se muestran con los diagramas principales de la aplicación de la Inteligencia de Negocios los resultados deseados.

Data Warehouse (Curto, 2011): Es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización, independientemente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes).

(Gartner, 2014): Business Intelligence (BI) es un término general que incluye las aplicaciones, la infraestructura y las herramientas y las mejores prácticas que permiten el acceso y el análisis de la

información para mejorar y optimizar las decisiones y el rendimiento.

Cano (2007): Los componentes del Business Intelligence a utilizar son: fuentes de información que alimentarán a la data warehouse, el proceso *extraer* Extract, *transformar* Transform and *cargar* Load, ETL y la solución *On-Line Analytical Processing* OLAP, que nos debe proveer capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y

análisis de escenarios en grandes volúmenes de datos, donde la parte principal se centra en el data warehouse (figura 1).

Limitaciones: Se usa la Base de Datos Northwind del ejemplo de Microsoft, para crear el prototipo. No hay una Pyme específica de aplicación ni muestra estadística, la propuesta es de una aproximación general, dado que en nuestro medio es casi nulo en estadísticas de Business Intelligence.

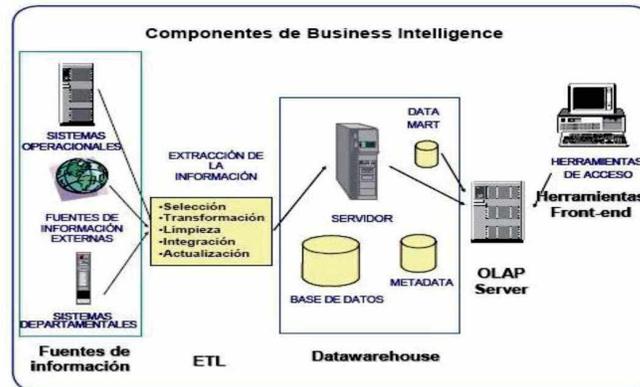


Figura 1. Componentes de Business Intelligence
 Fuente: Josep Lluís Cano (2007)

En la figura 2 se muestran los procesos de integración de las aplicaciones, Relaciones del cliente, Cadena de Suministros, Patrimonio etc.

Los elementos de una base de datos multidimensional (Curto, 2011) son:

Tabla de hechos: Es la representación en el data warehouse de los procesos de negocio de la organización. Por ejemplo, una venta puede identificarse como un proceso de negocio de manera que es factible, si corresponde en la organización, considerar la tabla de hecho ventas.

Dimensión: Es la representación en el data warehouse de una vista para un cierto proceso de negocio. Si se regresa al ejemplo de una venta, para la misma tenemos el cliente que ha comprado, la fecha en la que se ha realizado. Estos conceptos pueden ser

considerados como vistas para este proceso de negocio. Puede ser interesante recuperar todas las compras realizadas por un cliente. Ello hace entender por qué se identifica como una dimensión. En la figura 3 se ve un cubo con tres dimensiones: Categorías de Producto, Medidas y año 2007.

La figura 4 muestra el Esquema Estrella, donde la tabla de hechos está en el centro y alrededor están las dimensiones con sus datos de un caso de datos de filiación de personas en una entidad. La figura 5 muestra el Esquema Copo de Nieve, que es la variante del modelo estrella donde algunas dimensiones pueden jerarquizar su información, como se observa la información puede desagregarse.

Las dimensiones apuntan a la tabla de hechos, si son más de tres el concepto de cubo es imaginario, se conserva siempre la palabra cubo.

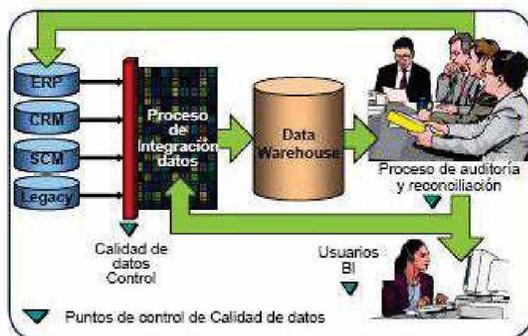


Figura 2. Integración de procesos del BI
 Fuente: Joseph Curto (2011)

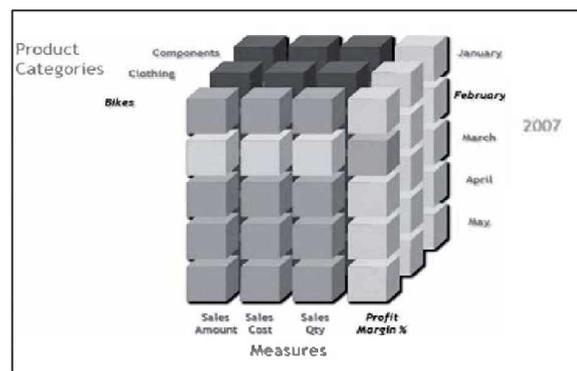


Figura 3. Ejemplo de Cubo
 Fuente: Joseph Curto (2011)

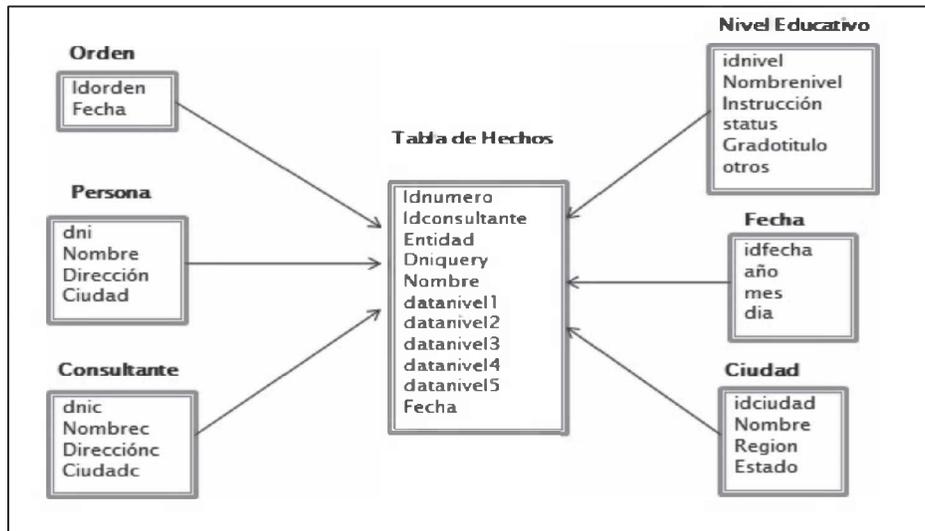


Figura 4. Esquema Estrella
 Fuente: Elaboración propia

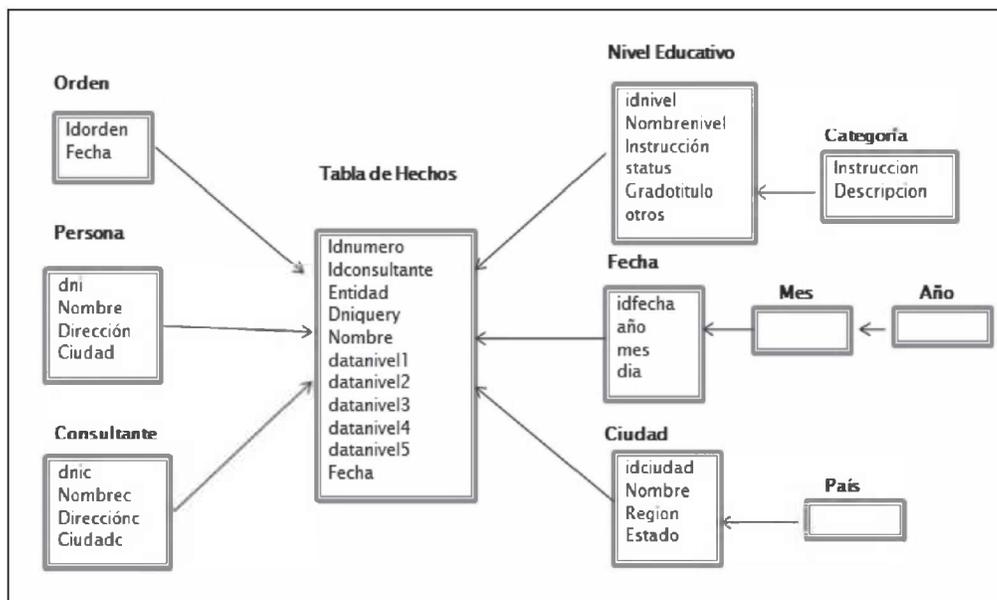


Figura 5. Esquema Copo de Nieve
 Fuente: Elaboración propia

RESULTADOS

Se aplicó Inteligencia de negocios con SQL Server, se utilizó como base de datos inicial *Northwind*, El diagrama entidad relación de esta base de datos se muestra en la figura 6.

La base de datos *Northwind* está conformada por las siguientes tablas:

Customers (clientes): La tabla registra la información relacionada con la cartera de clientes con la que trabaja *Northwind*. Dicha tabla contempla, el nombre de la

compañía, el nombre del contacto, la dirección, la región, ciudad, país, entre otros campos.

Employees (empleados): La tabla registra la información relacionada con los empleados que atienden las órdenes en *Northwind*. Dicha tabla contempla los apellidos, nombre, fecha de contratación, fecha de nacimiento, así como el superior inmediato del empleado, entre otros campos.

Categorías (categorías): La tabla registra la información relacionada con las categorías que agrupan varios productos.

Products (productos): La tabla registra la información de los productos que son solicitados en las órdenes, como nombre del producto, categoría asociada, precio unitario, proveedor, unidad de medida, estado del producto, entre otros campos.

Orders (órdenes): La tabla registra los movimientos de cada orden generada, en la cual se consigna el cliente solicitante, el empleado que la atiende, la fecha, entre otros campos.

Order Details (detalle de órdenes): La tabla registra el detalle de cada línea de producto de cada orden generada, es decir, el producto, la orden, la cantidad solicitada, el precio unitario y el descuento.

Luego se crean las vistas en la ventana como se ve en la figura 7 en un diagrama Entidad Relación.

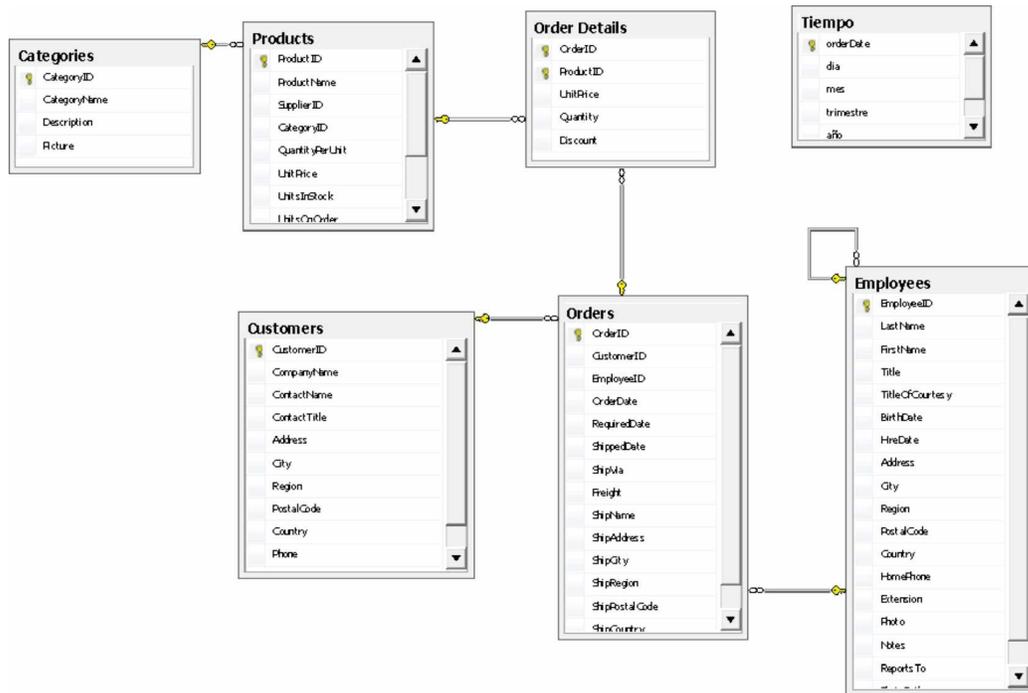


Figura 6. Diagrama Relacional Base de Datos Northwind
Fuente: Elaboración propia

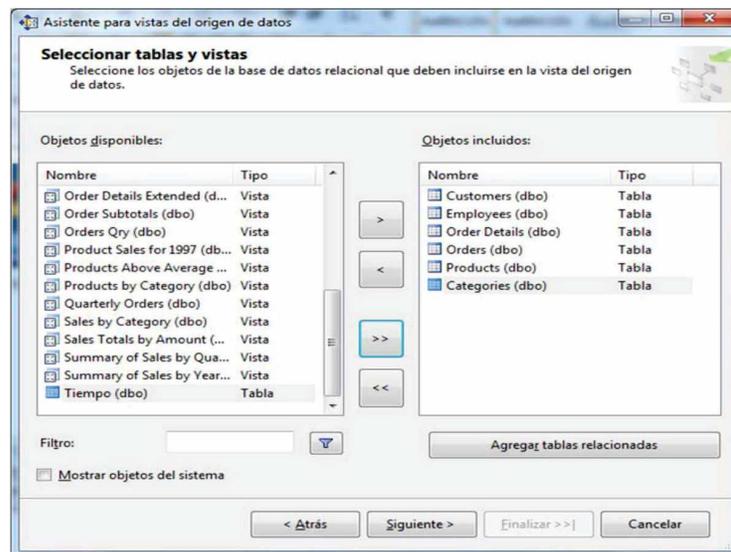


Figura 7. Creación de Vistas del Modelo
Fuente: Elaboración propia

Después se agregaron tablas a la Vista de origen de datos (figura 8). Al agregar la tabla *Tiempo*, formó parte del diseño de la vista, se relacionó con la tabla *Tiempo* y con la tabla *Orders*. Se creó la relación antes mencionada seleccionando la columna *OrderDate* de la tabla *Orders* y con el arrastre del mouse hacia la tabla

Tiempo. En la ventana se especificó la relación siguiente: Tabla de origen *Orders*, Columnas de origen: *OrderDate*, Tabla destino: *Tiempo*, Columnas de destino: *OrderDate*. Al finalizar el proceso de creación de vistas dio resultado una pantalla (figura 9).

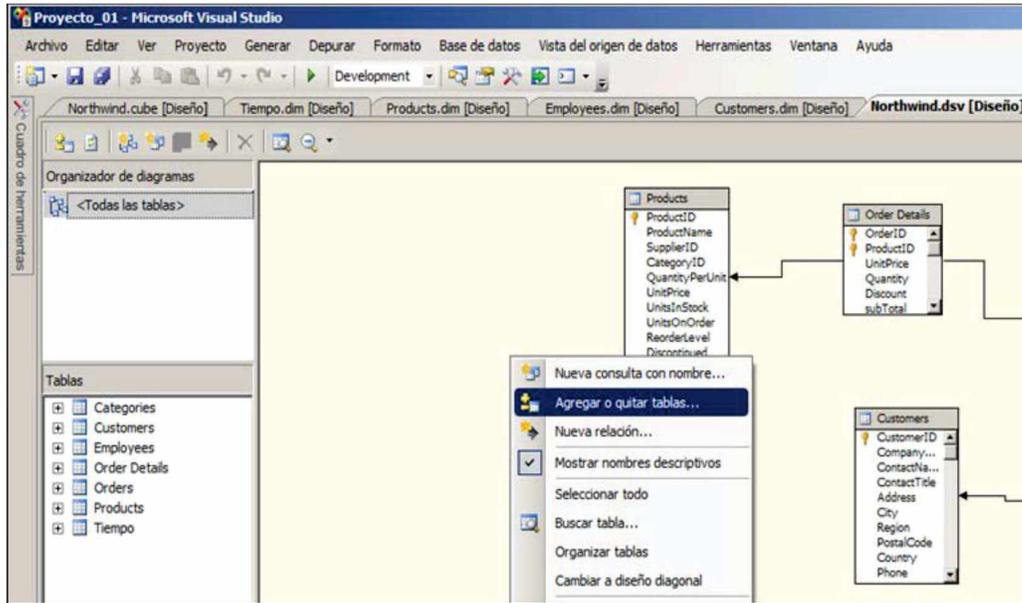


Figura 8. Agregación de Tablas a la Vista
Fuente: Elaboración propia

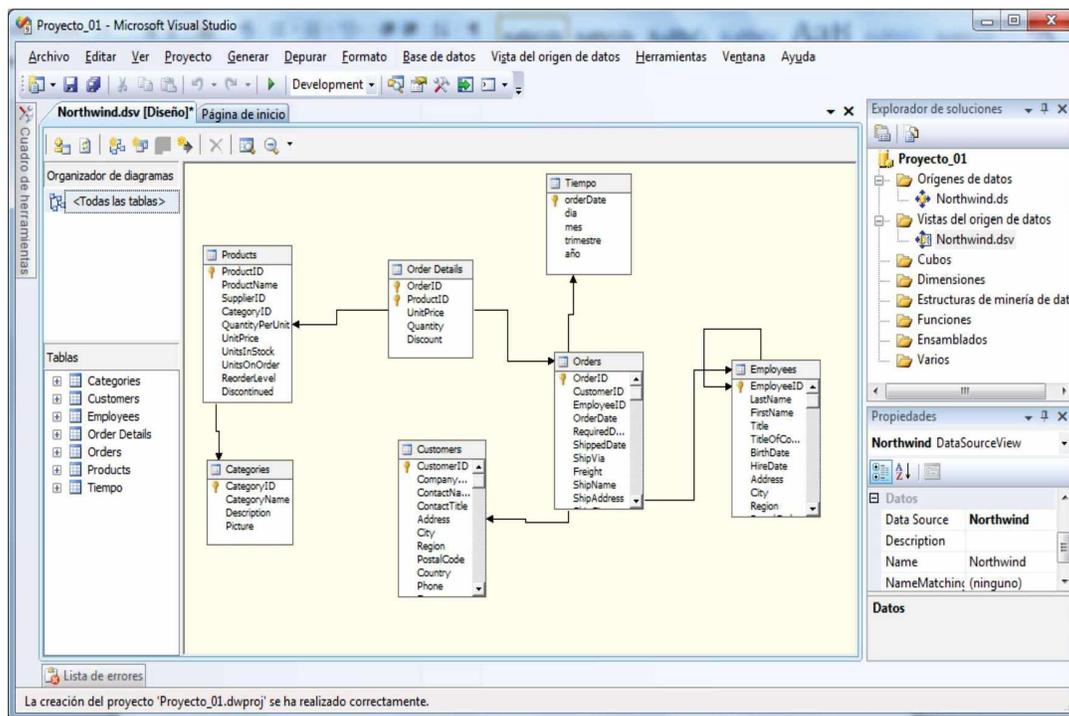


Figura 9. Vistas creadas
Fuente: Elaboración propia

Luego se crearon las dimensiones. Una dimensión de base de datos es una colección de objetos relacionados, denominados atributos, que se usaron para proporcionar información sobre los datos de hechos de uno o varios cubos. Los atributos típicos de una dimensión de producto pueden ser el nombre, la categoría, la línea, el tamaño y el precio del producto. Estos objetos están enlazados a una o varias columnas de una o varias tablas de una vista del origen de datos. De manera predeterminada, estos atributos están visibles como jerarquías de atributo y se pueden utilizar para comprender los datos de hechos en un cubo. Los atributos se pueden organizar en jerarquías definidas por el usuario que proporcionan rutas de navegación para ayudar a los usuarios al examinar los datos de un cubo.

Los cubos contuvieron todas las dimensiones en las que los usuarios basan sus análisis de los datos de hechos. Una instancia de una dimensión de base de datos en un cubo se denomina dimensión de cubo y se relacionó con uno o más grupos de medida en el cubo. Una dimensión de base de datos se utilizó varias veces en un cubo. Por ejemplo, una tabla de hechos puede tener varios hechos relacionados con el tiempo y se puede definir una dimensión de cubo independiente que sirva de ayuda para analizar cada uno de ellos. Sin embargo, solo es necesario que haya una dimensión de base de datos relacionada con el tiempo, lo que significa también que solo es necesario que haya una tabla de base de datos relacional que esté

relacionada con el tiempo para admitir varias dimensiones de cubo basadas en el tiempo.

Se creó la dimensión *Customers*. Para crear una dimensión en la ventana Explorador de soluciones se seleccionó *Dimensiones* y se dio clic con derecha, luego se seleccionó *Nueva Dimensión* (figura 10).

Al finalizar se mostró la vista diseño de la dimension *Customers*. Se puede observar tres zonas importantes: Atributos, Jerarquias y Vista de origen de datos. Luego se creó el Cubo. Un cubo es una estructura multidimensional que contiene información con fines analíticos, sus componentes principales son las dimensiones y las medidas. Las dimensiones definen la estructura del cubo que se utiliza para segmentar y dividir los datos, y las medidas proporcionan valores numéricos agregados importantes para el usuario final. Como estructura lógica, un cubo permitió a una aplicación *cliente* recuperar valores, de medidas, como si estuvieran almacenados en las celdas del cubo; las celdas se definen para cada posible valor resumido. Las celdas del cubo se definieron por la intersección de miembros de dimensión y contienen los valores agregados de las medidas en esa intersección concreta.

Se mostró el cubo *Northwind.cube*, se presentaron dos zonas principales: *Medidas* y *Vista de origen de datos* (figura 11).

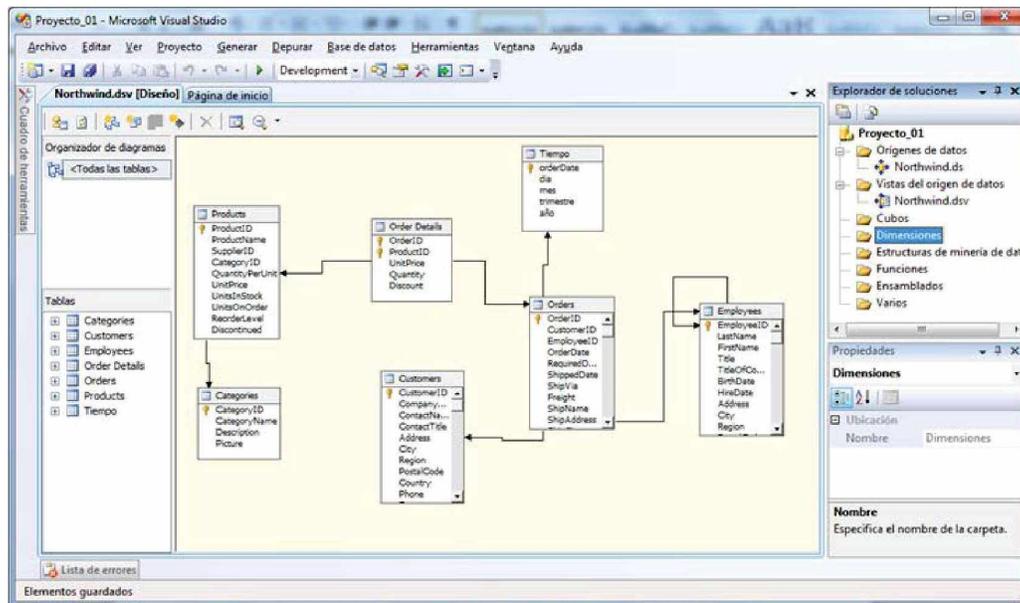


Figura 10. Creación de dimensión *Customers*
Fuente: Elaboración propia

Se seleccionó la ficha *Uso de dimensiones*. Se tuvieron que relacionar las dimensiones: *Customers*, *Employees* y *Tiempo*. Esta relación es referenciada se utilizó la dimensión *Orders*. Finalmente se mostró el

diseño del cubo *Customers.cube* concluido. Se grabó el proyecto (figura 12).

Luego se procesó la carga de datos. Para esta carga de los datos a la estructura Dimensional se debió procesar

para lograr este objetivo. De la ventana Explorador de Soluciones se seleccionó Proyecto 01, con un clic con derecha y del menú desplegable Procesar. En la ventana Procesar Cubo *Northwind* se dio un click en el botón *Ejecutar*.

En la ventana *Progreso del proceso* se da un clic en el botón *Cerrar*. Se obtuvo los resultados deseados con la *selección de la ficha Examinador*. Se expandió la dimensión *Tiempo*, seleccionando *Año* (figura 13), se construyó el modelo que es del tipo copo de nieve.

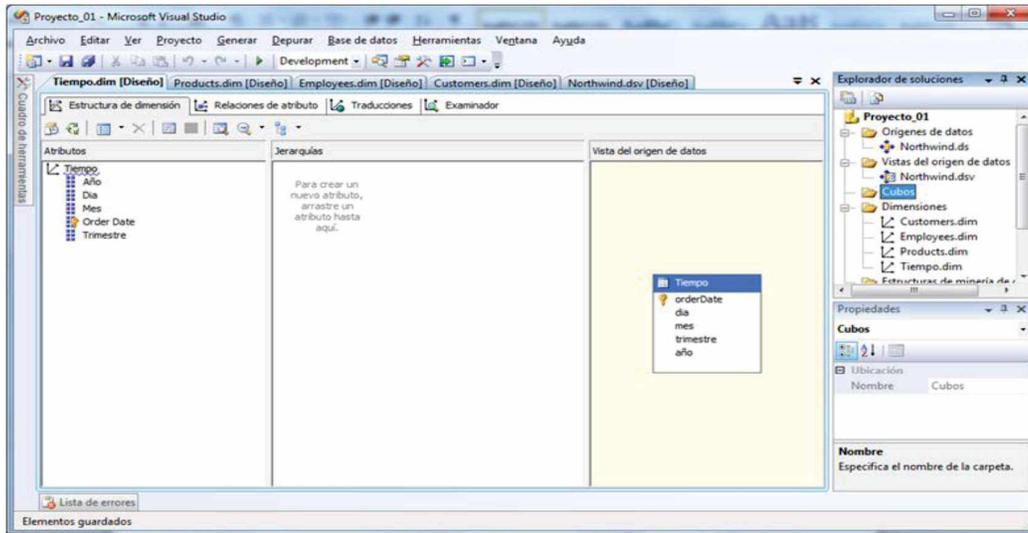


Figura 11. Creación de cubo
Fuente: Elaboración propia

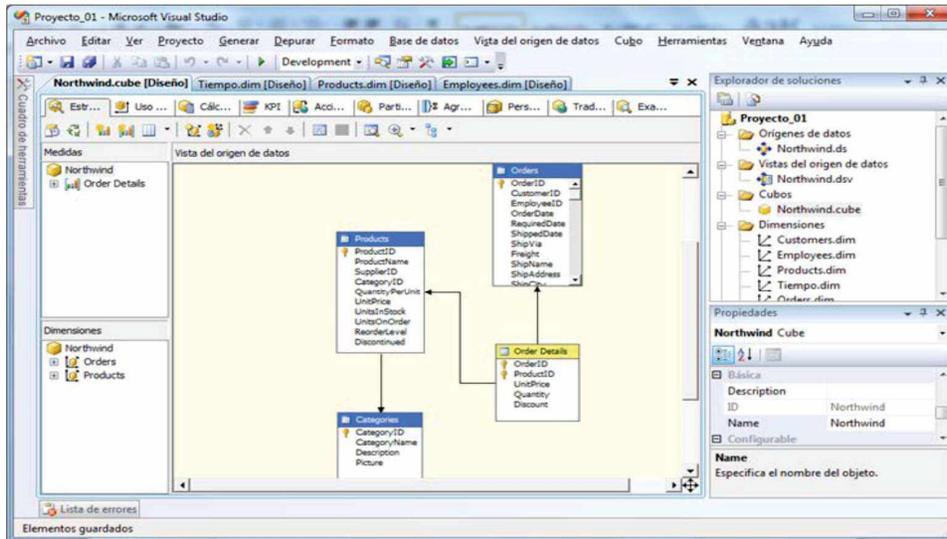


Figura 12. Cubo de la Base de Datos *Northwind*
Fuente: Elaboración propia

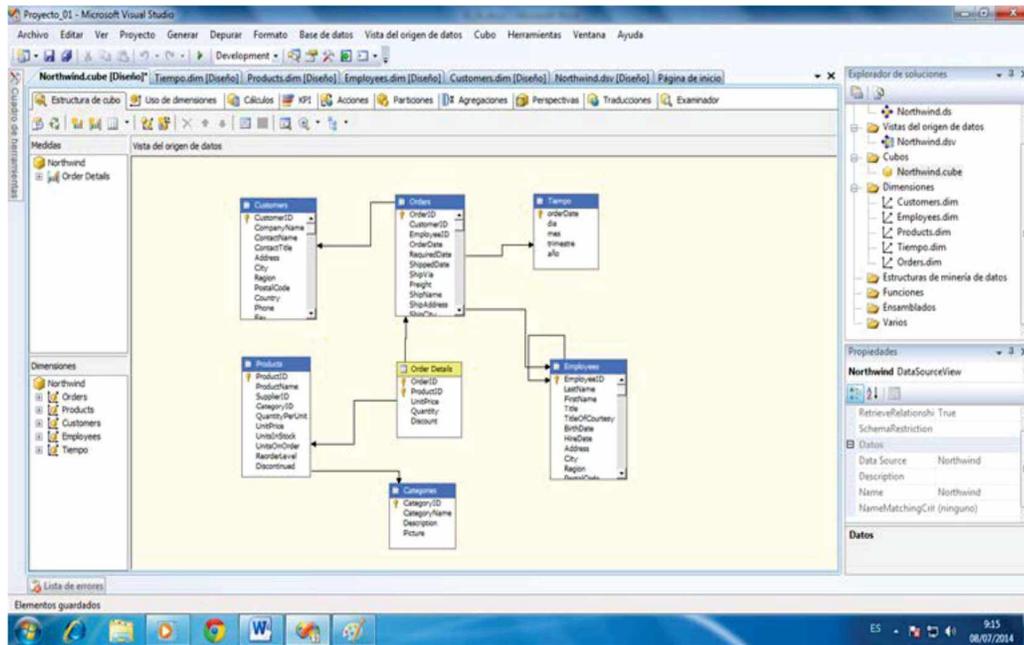


Figura 13. Cubo concluido
 Fuente: Elaboración propia

Se expandió *Customers*, se seleccionó *Country* y se arrastró hasta colocar campos de fila aquí. Se expandió *Measures*, se seleccionó *Quantity* y se arrastró hacia *Coloque campos de totales o campos detallados*. En el cuadro final (figura 14) se vió la información resumida por las filas a los países de la exportación y en la columna a los años requeridos, donde arrojaron

la información compactada que sirvió para análisis y la toma de decisiones posterior. Con la data empleada, los resultados se ven en la tabla 1, donde se demostró con la data de ejemplo cómo se puede resumir directamente las exportaciones de los clientes.

Country	1996	1997	1998	Total general
Argentina	94	245	339	
Austria	949	2347	1871	5167
Belgium	185	516	691	1392
Brazil	768	2057	1422	4247
Canada	297	1249	438	1984
Denmark	232	766	172	1170
Finland	157	604	124	885
France	658	1807	789	3254
Germany	1910	4756	2547	9213
Ireland	490	799	395	1684
Italy	81	424	317	822
Mexico	229	551	245	1025
Norway	48	21	92	161
Poland	45	59	101	205
Portugal	135	342	56	533
Spain	229	140	349	718
Sweden	375	949	911	2235
Switzerland	232	628	415	1275

Figura 14. Resultado Final, Información de años con dimensiones elegidas
 Fuente: Elaboración propia

Tabla 1. Resultados Finales

Country	Año			
	1996	1997	1998	Total general
Argentina		94	245	339
Austria	949	2347	1871	5167
Belgium	185	516	691	1392
Brazil	768	2057	1422	4247
Canada	297	1249	438	1984
Denmark	232	766	172	1170
Finland	157	604	124	885
France	658	1807	789	3254
Germany	1910	4756	2547	9213
Ireland	490	799	395	1684
Italy	81	424	317	822
Mexico	229	551	245	1025
Norway	48	21	92	161
Poland	45	59	101	205
Portugal	135	342	56	533
Spain	229	140	349	718
Sweden	375	949	911	2235
Switzerland	232	628	415	1275

Fuente: Elaboración propia

DISCUSIÓN

Matute (2013) sostiene que en el Perú hay algunas empresas que ya usan la herramienta, pero en focos específicos, pues empezar un proceso de inteligencia de negocios requiere de gente bastante especializada y de herramientas de software para los procesos de evaluación. Pero se ve que se puede aplicar en los supermercados, tiendas por departamentos, bancos e incluso en los vendedores de autos. Una Pyme sí tendría mayor dificultad de utilizar el recurso por el volumen de conocimientos y habilidades que se requieren, pero existe en la nube empresas en Estados Unidos que dan este tipo de servicios para pequeños y medianos negocios. En esta investigación se demuestra que hay que utilizar la herramienta y que sería de gran utilidad para las Pymes.

Curto (2011) menciona además el software Pentaho, el cual tiene un diseño fijo, es decir fue elaborado como otros actuales como el *Tableau* que funcionan directamente para cargar información de tablas, se construye el cubo automáticamente y se ven los resultados con el apoyo de la analítica. La propuesta es mejor que estos softwares que son destinados al usuario final, en cambio con la opción demostrada con el SQL Server es usada a nivel del analista en el *back-end* (interno de la PC) y también en el usuario (*front-end*), además se cuenta con el analista para interpretaciones dudosas.

En el uso de la Data Warehouse y Business Intelligence, Davenport y Harris (2007), indican que las principales ventajas son: mejoras en los reportes de información rutinarios, mejoras en los reportes de información especiales, en búsquedas para análisis detallados, obtención de alertas y análisis estadísticos.

Con esta investigación se ha aproximado bastante con estas ventajas.

Turban *et al.* (2007) indican que las áreas que reciben mayor beneficio del Data Warehouse son marketing y ventas, clientes y canales, precios y contratos, pronósticos, desempeño la Data Warehouse es en segmentación de clientes, propensión a la compra, rentabilidad del cliente, detección de fraudes, evitar la pérdida de clientes y optimización de canales de comercialización en ventas, finanzas, cadena de abastecimientos, servicio al cliente y procesos. Con esta investigación el resultado final es para actuar de inmediato para no perder a clientes, además hacer ajustes en procesos del negocio.

En los países desarrollados ya están usando, en cambio en los países en vía de desarrollo hay intentos o nada, en comparación, debemos divulgarlo inmediatamente, así es nuestra realidad.

CONCLUSIONES

El uso de *Inteligencia de Negocios* permite simplificar el proceso de extracción de información en un formato apropiado para la toma de decisiones.

Los resultados pueden exportarse en hojas Excel o a cualquier interface prediseñada para una mejor vista al usuario final.

El tiempo de respuesta si se hace varias consultas en SQL comparado con la ejecución del modelo dimensional creado, esta última demora menos tiempo.

Las Pymes en general de nuestro medio, con los procesos parecidos indicados en este estudio y con el prototipo presentado en esta investigación, ahorrarían

tiempo, a la vez costos, en vez de usar sus sistemas relacionales y también tiene ventajas frente al software propietario y/o libre, preparados para el usuario final.

REFERENCIAS BIBLIOGRÁFICAS

- Cano, J. (2007). *Business Intelligence: Competir con Información*. España: Ed. Banesto ESADE.
- Curto, J. (2011), *Introducción al Business Intelligence*. España: Editorial UOC.
- Davenport, T. y Harris, J. (2007). *Competing on Analytics, The New Science of Winning*. Boston: Harvard Business School Press.
- Gartner. (2014), *Glosario de términos*. Recuperado de: <http://www.gartner.com/it-glossary/business-intelligence-bi/>.
- Matute, G. (2013). *¿Porque es Importante Aplicar la Inteligencia de Negocios?*. Conexión Esan Perú. Recuperado de:
<http://www.esan.edu.pe/conexion/actualidad/2013/04/12/inteligencia-negocios-empresa/>.
- Turban, E., Aronson, J., Liang, T. y Sharda, R. (2007). *Decision Support Business Intelligence Systems*. Upper Saddle River: Pearson Prentice Hall.