



Diseño de un Datamart orientado al proceso de ventas usando la herramienta de Inteligencia de Negocios SQL Server 2014

Autor-Ana Mercedes MONTENEGRO RIVERA

Universidad Técnica del Norte, Av. 17 de Julio, Ibarra, Imbabura
montenegroany@gmail.com

Resumen. *El presente artículo tiene como finalidad dar a conocer un resumen del proyecto de tesis titulado “DISEÑO DE UN DATAMART ORIENTADO AL PROCESO DE VENTAS USANDO LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS SQL SERVER 2014”, en el cual se pone en práctica los conceptos y metodologías de Inteligencia de Negocio para la construcción de un datamart.*

En dicho proyecto se hace uso de las herramientas contenidas en SQL Server 2014 para el proceso de extracción transformación y carga, así como también las herramientas contenidas en Microsoft Excel para análisis de datos y la presentación de informes interactivos.

Palabras Claves

Inteligencia de Negocios, DataMart, SQL Server, Microsoft Excel.

Abstract. *This article aims to provide an overview of the thesis project entitled “DISEÑO DE UN DATAMART ORIENTADO AL PROCESO DE VENTAS USANDO LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS SQL SERVER 2014”, in which it implements the concepts and methodologies of Business Intelligence for building a datamart.*

In this project using the tools in SQL Server 2014 to the process transformation extraction and loading, as well as the tools in Microsoft Excel for data analysis and presentation of interactive reports it is made.

Keywords

Business Intelligence, DataMart, SQL Server, Microsoft Excel.

1. Introducción

Las empresas están compuestas por varios procesos, que generan gran cantidad de información diariamente, misma que se almacena en sistemas transaccionales que se encargan de administrarla. Dicha información es lo más valioso de una empresa, ya que de ella se puede obtener conocimientos que permiten plantear estrategias competitivas.

Actualmente en muchas empresas se presenta un problema a la hora de la toma de decisiones estratégicas, ya que toda la información que generan las empresas como resultado de sus procesos, se encuentra dispersa y desorganizada por lo que su análisis se vuelve lento, y de bajo rendimiento, provocando así que las decisiones se tomen sin toda la información relevante o se tomen malas decisiones.

Las soluciones de Inteligencia de Negocios permiten tener toda la información, estructurada y consolidada en un solo repositorio, permitiendo explotarla, para la generación de reportes que permitan plantear estrategias de ventas eficaces, con menor tiempo y esfuerzo, por lo cual el objetivo principal en este proyecto se centra en el diseño un DATAMART orientado al proceso de ventas utilizando las herramientas que Microsoft pone a disposición para Inteligencia de Negocios como SQL Server 2014 y varios componentes de Microsoft Excel.

2. Materiales y Métodos

Partiendo de la arquitectura de Inteligencia de Negocios se ha desarrollado el presente proyecto, usando como guía la metodología propuesta por Ralph Kimball.

Haciendo uso de los conceptos básicos y las herramientas de Inteligencia de Negocios se ha construido un modelo de datos para analizarlos y

obtener reportes de ventas, tomando como información fuente la que genera una empresa comercial en producción.

2.1 Arquitectura de Inteligencia de Negocios

La Inteligencia de negocios posee una arquitectura global muy flexible, capaz de proporcionar al diseñador varias opciones para crear una solución de BI, en general se basa en el siguiente proceso:



Ilustración 1: Arquitectura Inteligencia de Negocios

2.1.1 Fuentes De Información

Son todas las fuentes de información, sobre las cuales se trabaja para sacar los datos que alimentara al repositorio de datos, se puede usar múltiples fuentes, en esta fase el objetivo principal implica la realización de un estudio minucioso de la información que se tiene disponible acerca de su formato, disponibilidad y calidad, para así poder capturarla, procesarla y dejarla lista para usarla.

2.1.2 Procesos ETL

Este conjunto de procesos permiten obtener los datos de las fuentes de información para transformarlos y adaptarlos al modelo definido y luego depositarlos en las bases de datos de destino o repositorios.

Esta fase es una de las más importantes de los proyectos de Inteligencia de Negocios, e incluso uno de los más largos ya que consume por lo menos el 50% de tiempo del ciclo de vida del proyecto, además que debe estar bien diseñado para tener mejores resultados a la hora de la toma de decisiones.

2.1.3 Repositorio de Información

En esta parte de la arquitectura de Inteligencia de negocios se encuentra a los **DATAWAREHOUSE** como almacén de datos, en este almacén es en donde se depositara toda la información que se haya tratado con los procesos ETL (Extracción, transformación y carga), aquí la información estará lista para ser analizada para la toma de decisiones. De este concepto se deriva también los **DATAMART** que consiste en un repositorio de datos orientado a un tema específico.

2.1.4 Herramientas de Analisis

Dentro de estas herramientas hay varios conceptos que se debe conocer por ejemplo:

Herramientas OLAP (OnLine Analytical Processing) permiten analizar la información que reside en los Datawarehouse o los Datamart. Su principal objetivo es agilizar las consultas de grandes volúmenes de información para lo cual utiliza estructuras multidimensionales conocidas como Cubos.

Bases de datos multidimensionales tienen la capacidad de almacenar datos con varias dimensiones, es decir en vez de un valor se encuentran varios dependiendo de los ejes definidos, está orientada a consultas complejas y de alto rendimiento, con grandes volúmenes de datos. Esta base de datos está compuesta por varias estructuras dimensionales llamadas cubos. Para el almacenamiento de los procesos de negocio utiliza tablas de hechos y tablas de dimensiones.

Cubos: La representación gráfica de OLAP es lo que se conoce como “**Cubos**”. Un cubo OLAP es un vector multidimensional de N dimensiones en donde se almacena la información de manera ordenada facilitando un análisis rápido de sus contenido. Una base de datos multidimensional puede estar formada por varios cubos OLAP, dependiendo de sus requerimientos a analizar.

2.1.5 Herramientas de Visualización

Existen varias herramientas que permiten trabajar sobre la información contenida en los Datawarehouse y los Datamart, su objetivo principal es permitir acceder, analizar y visualizar dicha información facilitando la elaboración de informes tanto en detalle como sobre información agregada. Principalmente estas herramientas proporcionan una interfaz de usuario para explorar la información, como por ejemplo Dashboards, Cuadros de mando, etc

2.2 Metodología Kimball

En cuanto a la metodología usada para la construcción del modelo de datos se usó la que más se ajusta a las necesidades, como lo es la que propone Ralph Kimball, en donde se construye en primera instancia un DATAMART.

Para la optimización de consultas y mejorar la facilidad de uso de datamarts, Kimball propone el modelo de datos como esquema estrella. Además sugiere una visión Bottom-Up, es decir una vez extraído los datos de las fuentes se construye varios datamart departamentales e independientes, modelados en forma dimensional; que

posteriormente pueden evolucionar para formar un Datawarehouse, que se relacionan entre sí a través de sus dimensiones.

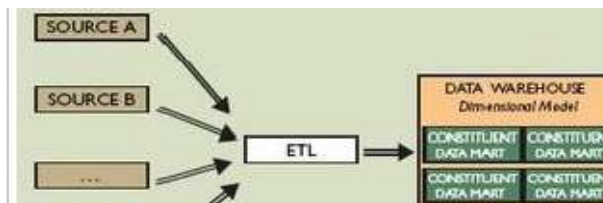


Ilustración 2: Esquema Metodología Kimball

2.2.1 Fases de la metodología Kimball

En resumen prácticamente esta metodología propone 4 fases para la construcción del DWH, las cuales se detalla a continuación:

Fase 1. Selección del proceso de negocio

- En esta etapa lo primero es identificar que proceso o procesos de negocio se va a modelar para combinar y entender los requerimientos con los datos disponibles. Estos procesos son las actividades que desarrolla la empresa como rutina diaria y que están soportados por sistemas transaccionales.

Fase 2. Declarar la granularidad de la información

- Una vez identificado los procesos es necesario definir la granularidad, es decir: ¿Qué nivel de detalle debería estar disponible en el modelo dimensional?. Este paso es sumamente importante ya que de él dependerá el modelo de datos y la implementación del datamart.

Fase 3. Elección de las dimensiones de análisis

- Se deben elegir las dimensiones que se aplicarán para cada tabla de hechos, las dimensiones deben responder a la pregunta: ¿Cómo los usuarios describen los datos que resultan del proceso de negocios?.

Fase 4. Identificar los hechos o métricas

- En esta parte se debe definir los hechos numéricos o métricas que formarán parte de la tabla de hechos. Los hechos son determinados por esta pregunta: “¿Qué estamos midiendo?”. Los usuarios están muy interesados en el análisis de estas medidas de la ejecución de procesos de negocio.

Tabla 1: Fases de la metodología Kimball

En base a todas las sugerencias de esta metodología se ha desarrollado el presente proyecto en cuatro fases:

2.2.2 Planificación

El objetivo de esta fase es planear las actividades a ser ejecutadas, definir los recursos a emplear e identificar los requerimientos.

| | |
|--|---|
| | Identificación y análisis de requerimientos de la funcionalidad del área de ventas basándose en la información de análisis existente en los datos fuente. |
| | Diseño y Construcción de un almacén de datos o Data Mart que resida en una base de datos de SQL Server 2014. |
| | Brindar la posibilidad de generar análisis de información y reportes de ventas desde la bodega de datos. |
| | Desarrollar un proceso ETL (Extracción, Transformación y Carga) para la carga de información desde la base fuente hasta la nueva bodega de datos. |
| | Obtención de conocimiento en manejo de herramientas de desarrollo utilizadas. |
| | Construcción de reportes a través de Excel. |

Tabla 2: Planificación - Actividades

Los requerimientos del departamento de ventas sujeto del presente análisis abarca temas con respecto a:

Análisis de clientes: ¿Quiénes son los mejores clientes? ¿Cómo ha sido su movimiento durante el último año?, por ejemplo para dependiendo de esto dar un trato especial a los mejores clientes.

Análisis de territorio: ¿Dónde están localizados los mejores clientes? ¿En qué zona están los clientes?

Análisis de artículos: ¿cuáles fueron los artículos más vendidos? ¿Qué artículos tienen menos acogida?

Análisis de vendedores: ¿Cuáles fueron los mejores vendedores?

Todos los análisis que se realizan son en función del **Tiempo**, por lo cual es una dimensión obligatoria.

2.2.3 Análisis y Diseño

El objetivo de esta fase es definir el alcance de cada una de las áreas de análisis en sus medidas, dimensiones y atributos. Así mismo la creación del modelo de datos lógico y físico, que soporta el área de análisis.

Una vez definidas e identificadas las tablas de dimensiones con sus respectivos atributos y la tabla de hechos con sus medidas, se pasa a la construcción del modelo de datos en donde se define como está estructurada la tabla de hechos con las dimensiones.

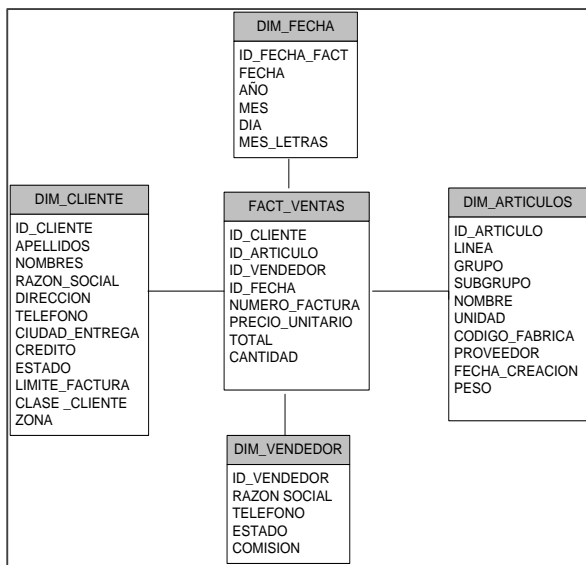


Ilustración 3: Modelo de Datos

2.2.4 Construcción y Desarrollo

En esta fase de construcción se diseña todos los procesos necesarios para obtener los reportes finales, empezando con la construcción de la base de datos multidimensional a partir de los datos de origen analizados anteriormente, para lo cual se realiza los procesos ETL mediante el SQL Server Integration Services para poblar el repositorio de datos destino.



Ilustración 4: Procesos ETL

Posteriormente se construyó el modelo tabular usando la herramienta Power Pivot para Excel, tomando como origen la base de datos multidimensional poblada con los procesos ETL.

Power Pivot permite el acceso al modelo de datos desde Excel, para ser analizados y visualizados por el usuario final, a través del uso de tablas dinámicas que muestran la información y permiten interactuar con ella.

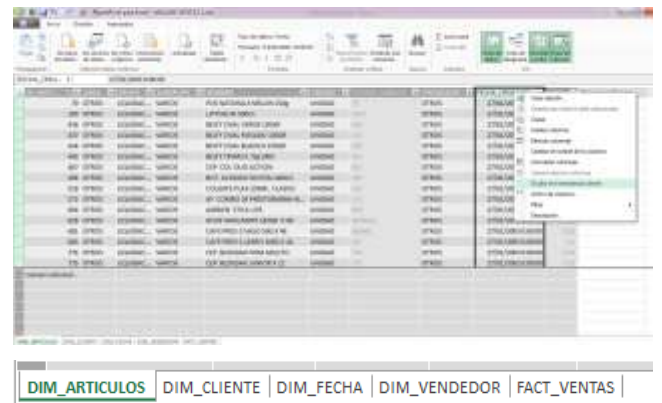


Ilustración 5: Visualización de la Información en Excel

2.2.5 Pruebas

El objetivo en esta fase es probar el modelo y la información que contiene, por medio de los reportes definidos en el alcance del proyecto, para lo cual se interactúa con la información exportada al Excel

| AÑO | Suma de TOTAL | Suma de UTILIDAD |
|----------------------|--------------------|-----------------------|
| 2011 | 5948240,61 | \$417.520,21 |
| 2012 | 8196733,59 | \$557.408,93 |
| 2013 | 7702623,13 | \$472.898,93 |
| 2014 | 7204904,08 | \$511.372,29 |
| Total general | 29052501,41 | \$1.959.200,36 |



Ilustración 6: Ambiente Excel para reportes

3. Resultados

Para el análisis de datos y la presentación de reportes, se usó los componentes de Excel como son, Power Pivot y Power View.

Varios de los reportes realizados permitieron encontrar varios datos importantes para la toma de decisiones, a continuación se describen varios de ellos:

Para el siguiente reporte, se analiza a través de las medidas, las Ventas y Utilidad de los artículos, en este reporte se puede navegar de acuerdo a la granularidad de la dimensión artículo, es decir se puede observar las ventas y utilidad por línea, grupo y subgrupo; además se puede agregar también la zona en la que se comercializa dicho producto, y todo esto en base al tiempo y su respectiva granularidad; dando así un reporte bastante interesante en cuanto a los productos.

En este reporte se pudo observar una desventaja en cuanto a la clasificación de los productos, ya que existe muchos productos que no están clasificados por línea, grupo o subgrupo, por lo cual se almacenan en la variable Otros, provocando que la información resultante no sea tan específica, pero sin embargo de los productos clasificados se puede también observar que la Línea de productos Isabel es la más vendida y la que genera mayor utilidad en general. En este reporte igual se puede obtener la información de acuerdo a todos los criterios disponibles en el panel de filtros.



Ilustración 7: Reporte Ventas y Utilidad

Se ha elaborado el siguiente reporte para demostración, en el que se puede analizar la información, en tres secciones diferentes, por **ZONA** y por **CLIENTE** que se encuentran en formato de barras y por **LÍNEA** que se encuentra en formato de tabla tipo tarjeta, y también se puede analizar la información con el filtro de tiempo.

En este reporte se puede también navegar por la información de manera específica, es decir si se da clic sobre una barra que la información se

actualice, **por ejemplo** si se selecciona la barra de la zona Tulcán el reporte se actualiza con los datos correspondientes a esa Zona y se puede observar por ejemplo que: en Tulcán el mejor Cliente es el Sr. Chilinguina y que el producto que más compra es de la línea ISABEL seguido de ILENSA.

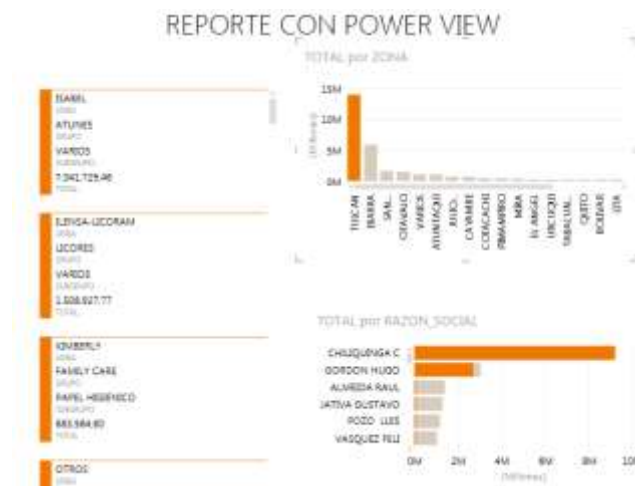


Ilustración 8: Reporte Ventas

4. Conclusiones

De la experiencia adquirida como resultado de la construcción del presente proyecto se puede concluir varios aspectos:

La herramienta utilizada para el desarrollo del presente proyecto como se propuso en el objetivo, fue SQL Server 2014, la cual permitió la construcción adecuada del DataMart del área de Ventas, constatando así que es una herramienta eficiente, y que tiene muchos recursos para seguir explotando en el tema de Inteligencia de Negocios, como por ejemplo la minería de datos, entre otros.

Las soluciones de Inteligencia de Negocios aplicadas a toda clase de empresas u organizaciones proveen mayor competitividad y rendimiento ya que impulsan de manera adecuada a las decisiones estratégicas, permitiendo al usuario profundizar en los datos y manejarlos desde diferentes perspectivas.

El usar metodologías para el desarrollo y construcción de un modelo de datos para Inteligencia de negocios, es muy beneficioso, ya que aparte de ser una guía, permite que las actividades a realizar estén organizadas y planificadas ahorrando así recursos y tiempo.

Agradecimientos

Agradezco de manera muy especial a todas las personas que han contribuido de alguna manera a concluir este proyecto, como lo son, Ing, Mauricio Rea Director de Tesis al Ing. Omar Lara, Ing. Diego Trejo, Ing José Luis Rodríguez, que conformaron el tribunal de defensa y con sus consejos puede hacer las correcciones y mejoras al proyecto, también quiero agradecer a Ing. Jorge Quiguango y Ing. Martin Ruiz por su apoyo al compartir sus conocimientos sobre el tema, que fueron indispensables para la construcción del proyecto.

Referencias Bibliográficas

- [1] Roebuck, K. (2012). DATA WAREHOUSING. Emereo Publishing.
- [2] Cientificos,L. (2015). MODELADO PREDICTIVO PARA LA INTELIGENCIA DE NEGOCIOS. Createspace.
- [3] Kimball, R. (2013) THE DATA WAREHOUSE TOOLKIT (3RA Ed.). CANADA: Jhon Wiley & Sons INC.
- [4] Barbieri, C. (2011). *Business Intelligence*. Rio de Janeiro Brasil: Elsevier Editorial Ltda.
- [5] Cesar, P. L. (2014). Técnicas de minería de datos e inteligencia de negocios : IBM SPSS Modeler. Garceta Grupo Ed. 2014.
- [6] Diaz, L. E. (2014). Inteligencia de Negocios: ¿la base para la solución de sus problemas? T21.

Sobre el Autor

Ana MONTENEGRO Nació en Tulcán - Carchi - Ecuador, el 01 de Enero de 1982. Realizó sus estudios en el “Instituto Técnico Superior Tulcán” donde finalizó en el año 2000, obtuvo el título de Bachiller en Ciencias Especialización Informática. Actualmente, es egresado de la Facultad de Ingeniería en Ciencias Aplicadas, carrera de Ingeniería en Sistemas de la Universidad Técnica del Norte.

