



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. Introduction

Nowadays many companies presents a problem when it comes to making strategic decisions, since information generating companies as a result of their processes, is scattered and disorganized so its analysis becomes slow and underperforming, provoking so decisions are taken without all the relevant information, or bad decisions.

Business intelligence solutions enable you to have all the information, structured and consolidated in a single repository, allowing to exploit it, for the generation of reports allowing to consider strategies of effective sales, with less time and effort, so the main objective in this project focuses on design a DATAMART oriented sales process using tools that Microsoft provides to business intelligence as SQL Server 2014 and various components of Microsoft Excel.

2. Materials and Methods

Based on the architecture of Business Intelligence has developed this project, using as a guide the methodology proposed by Ralph Kimball.

Using the basic concepts and business intelligence tools has built a data model to analyze and get sales reports, based information source which generates a trading company in production.

2.1 Business Intelligence Architecture

Business intelligence has a global very flexible, able to provide the designer with several options for creating a BI solution in general is based on the following process:

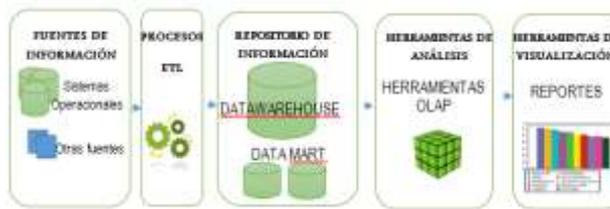


Illustration 1: Business Intelligence Architecture

2.1.1 Sources of Information

They are all sources of information, which is working to get the data fed to the data repository, you can use multiple sources, at this stage the main objective involves making a careful study of the information that is available about its format, availability and quality so as to capture it, process it and have it ready for use.

2.1.2 Processes ETL

This set of processes allow to obtain data from data sources to transform and adapt to the defined model and then deposit them in the target databases or repositories.

This phase is one of the most important projects Business Intelligence, and even one of the longest as it consumes at least 50% of cycle time project life also should be well designed to have better results when making decisions.

2.1.3 Information Repository

In this part of the architecture of Business Intelligence is the **DATA WAREHOUSE** as a data store, this store is where all the information that has been treated with ETL processes (Extraction, transformation and loading) is deposited, here the information will be ready for analysis for decision-making. This concept is also derived the **DATA MART** consisting of a data repository oriented to a specific topic.

2.1.4 Analysis Tools

Among these tools there are several concepts that should be known as:

Tools OLAP (OnLine Analytical Processing) allow to analyze the information that resides in the Datawarehouse or Datamart. Its main objective is to streamline the consultation process large volumes of information which uses multidimensional structures called cubes.

Multidimensional databases have the ability to store data with several dimensions, ie instead of a value are various depending on the defined axis is oriented to complex queries and high performance with large volumes of data. This database is

composed of several dimensional structures called buckets. For the storage of business processes using fact tables and dimension tables.

Cubes: The graphical representation of OLAP is what is known as "cubes". An OLAP cube is a multidimensional vector of N dimensions where information is stored in an orderly manner facilitating rapid analysis of its content. A multidimensional database can consist of several OLAP cubes, depending on their requirements to analyze.

2.1.5 Display Tools

There are several tools that allow work on the information contained in the Datawarehouse and the Datamart, its main purpose is to allow access, analyze and visualize this information to facilitate reporting in detail on both aggregate information. Mainly these tools provides a user interface to explore information, such as Dashboards, Scorecards, etc.

2.2 Methodology Kimball

As for the methodology used to construct the data model was used which best fits the needs, as is proposing Ralph Kimball, where it is built primarily a DataMart.

To query optimization and improve the usability of datamarts, Kimball proposes the data model as a star schema. In addition suggests a Bottom-Up view, ie once extracted data from various sources and independent departmental data mart is built, modeled dimensionally; which can then evolve to form a data warehouse, which are interrelated through their dimensions.

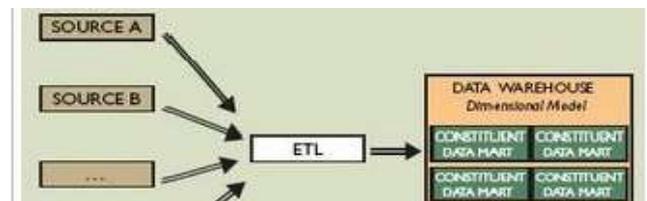


Illustration 2: Methodology scheme Kimball

2.2.1 Phases Kimball methodology

In summary this methodology practically proposes four phases for the construction of DWH, which is detailed below:

Phase 1. Selection of the business process

- At this stage the first to identify that process or business processes be modeled to combine and understand the requirements with available data. These processes are the activities undertaken by the company as a daily routine and are supported by transactional systems.

Phase 2. Declare the granularity of information

- Once identified processes is necessary to define the granularity, that is to say ¿What level of detail should be available in the dimensional model ?. This step is extremely important because it will depend on the data model and implementation of the datamart.

Phase 3. Choice of the dimensions of analysis

- Dimensions should be chosen to be applied for each fact table , the dimensions must answer the question : ¿How do users describe the data resulting business process ?.

Phase 4. Identify the facts or metrics

- This part must define the numerical facts or metrics that will be part of the fact table . The facts are determined by this question : "What are we measuring? " . Users are very interested in the analysis of these performance measures of business processes.

Table 1: Phases Kimball Methodology

Based on all the suggestions of this methodology, this project has been developed in four phases:

2.2.2 Planning

The objective of this phase is to plan the activities to be carried out, to define the resources to use and identify the requirements.

	Identification and analysis of requirements from the functionality of the sales area based on existing source data analysis information.
	Design and construction of a data warehouse or Data Mart that resides in a database of SQL Server 2014.

	Provide the ability to generate sales from the data warehouse reports and information analysis.
	Developing an ETL (extraction, transformation and loading) process to load information from the source base until the new winery of data.
	Obtaining knowledge in management of development tools used.
	Construction of reports through Excel.

Table 2: Planning - Activities

The requirements of the sales department subject of this analysis covers topics regarding:

Customer analysis: who are the best customers? How was his movement last year?, for example to depending on this give special treatment to the best customers.

Analysis of territory: where are the best customers located? Are customers in which zone?

Analysis of articles: what were the best-selling items? What items are less host?

Analysis of vendors: what were the best sellers?

All the analyses carried out are in function of time, which makes it a mandatory dimension.

2.2.3 Analysis and Desing

The objective of this phase is to define the scope of each of the areas of analysis in their measures, dimensions, and attributes. Likewise the creation of logical and physical data model that supports the analysis area.

Once defined and identified the dimension tables with their respective attributes and table of facts with your measurements, is passed to the construction of the model of data where is defined as the fact with the dimensions table is structured.

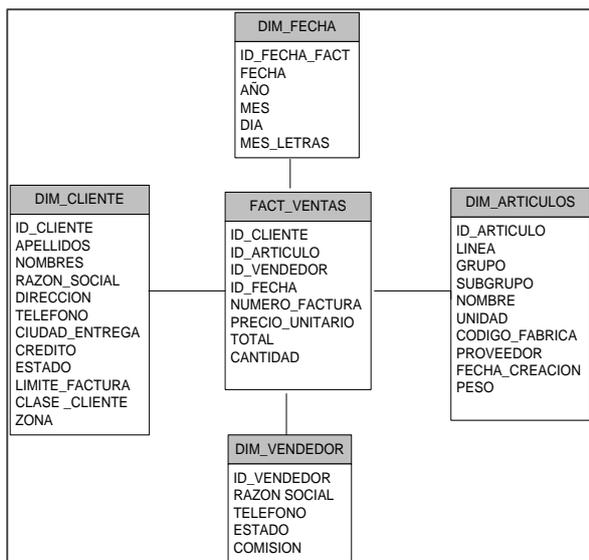


Illustration 3: Data Model

2.2.4 Construction and Development

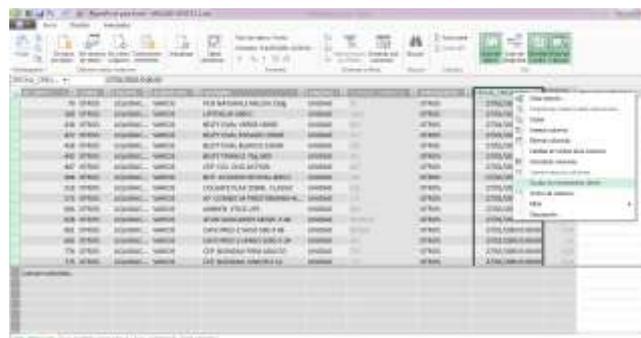
In this phase of construction is designed all the processes necessary to obtain the reports end, beginning with the construction of the multidimensional database from the source data analyzed previously, which is performed by using the SQL Server Integration Services ETL processes to populate the data destination repository.



Illustration 4: Process ETL

Subsequently built the tabular model using the tool Power Pivot for Excel, taking as origin the multidimensional database populated with ETL processes.

Power Pivot allows access to the data model from Excel, to be analyzed and visualized by the end user, through the use of dynamic tables that display information and allow you to interact with it.



DIM_ARTICULOS	DIM_CLIENTE	DIM_FECHA	DIM_VENDEDOR	FACT_VENTAS
Registro: 1 de 5911				

Illustration 5: Viewing the information in Excel

2.2.5 Testing

The objective in this phase is to test model and the information contained therein, by means of the reports defined in the scope of the project, for which it interacts with the information exported to Excel

CLASIF. ARTICULOS	AÑO	Suma de TOTAL	Suma de UTILIDAD
All	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



Illustration 6: Excel environment for reports

3. Results

Excel components were used for data analysis and the submission of reports, such as, Power Pivot and Power View.

Several of the reports carried out allowed to find several important data for decision-making, below are several of them:

For the following report analysed through measurements, sales and utility of the articles, in this report you can browse according to the granularity of the dimension article i.e. you can see sales and

utility line, group and subgroup; also can be added also the area in which is marketed that product, and all this on the basis of the time and their respective granularity; thus giving a very interesting report regarding products.

In this report could be observed a disadvantage regarding the classification of products, because there is many products that are not classified by line, group or subgroup, which is stored in the variable others, causing the resulting information to not be so specific, but nevertheless classified products you can also see that Isabel product line is the most sold and which generates more useful in general. In this same report get the information in accordance with all the criteria available in the Filter panel.



Illustration 7: Report Sales and Earnings

Has been prepared the following report for demonstration, in which information, in three different sections, can be analyzed by **ZONA** and for **CLIENTE** you are in format of bars and for **LÍNEA** that is tabular type card, and can also scan information with the filter of time.

In this report you can also browse by information specifically, IE if you click on a bar that the information is updated, for example if you select the zone bar Tulcán the report is updated with the data corresponding to that area and you can see for example that: in Tulcán the best customers is he Sr. Chilinguiga and that the product you purchase is more line ISABEL followed by ILENSA.

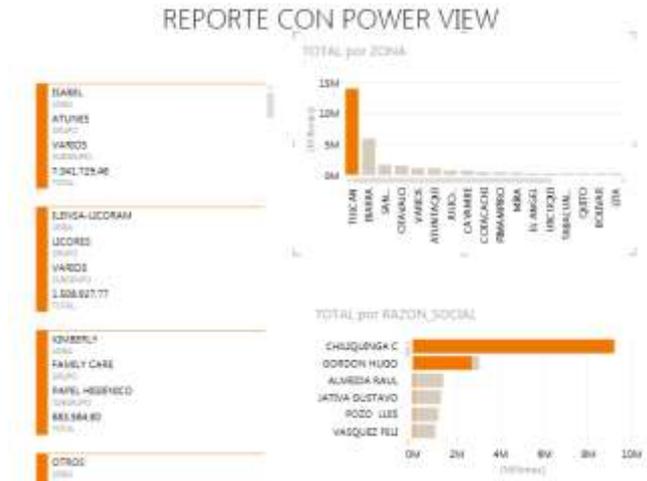


Illustration 8: Report Sales

4. Conclusions

The experience acquired as a result of the construction of this project several aspects can be concluded:

The tool used for the development of this project as proposed in the objective was SQL Server 2014, which allowed the proper construction of area sales DataMart, noting so it is an efficient tool, and has many resources to continue operating in the field of business intelligence, such as for example the data mining, among others.

Business intelligence solutions applied to all kinds of companies and organizations provide greater competitiveness and performance since they drive properly to strategic decisions, allowing the user to drill down into the data and manage them from different perspectives.

The use of methodologies for the development and construction of a data model for business intelligence, is very beneficial, because apart from being a guide, allows that the activities are organized and planned, thus saving time and resources.

Thanks

Thank a very special way to all those who have contributed in any way to conclude this project, as they are, Ing, Mauricio Rea Director de Tesis, Ing. Omar Lara, Ing. Diego Trejo, Ing José Luis Rodríguez, who formed the court defense and with their advice can make corrections and improvements to the project, I also want to thank Ing. Jorge Quiguango and Ing. Martin Ruiz for their support by



sharing their knowledge on the subject, which were indispensable for the construction of the project.

Bibliographic References

- [1] Roebuck, K. (2012). DATA WAREHOUSING. Emereo Publishing.
- [2] Científicos, 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*.

About the Author

Ana MONTENEGRO born in Tulcán - Carchi - Ecuador, the 01 January 1982. He studied at the “Instituto Técnico Superior Tulcán” where he finished in the year 2000, he obtained the title of Ciencias Especialización Informática. Currently, he is a graduate of the Facultad de Ingeniería en Ciencias Aplicadas, career of Ingeniería en Sistemas of the Universidad Técnica del Norte.

