

Proceso de transformación de una base de datos relacional a modelo multidimensional haciendo uso de la metodología Hefesto para la elaboración de un Datamart

Casas Huamanta, Edwin Roi¹; Huamán Camas, Danher²
Recibido 20 de mayo de 2016, Aceptado 4 de junio de 2016
Received: May 20, 2016 Accepted: June 4, 2016

RESUMEN

Hoy en día, en las diferentes actividades que las organizaciones realizan, se generan datos, como producto secundario, que son los resultados de todas las operaciones que se realizan, estos datos se convierten en información valiosa para la organización, con el pasar el tiempo éstas sirven para tomar decisiones y generar valor agregado a los productos o servicios que se ofrecen. Para tomar decisiones en base a esta información, se debe aplicar la inteligencia de negocios, la cual se encarga de recolectar, procesar y presentar la información organizada, la cual toma como una de las principales actividades, al proceso de transformación y modelamiento multidimensional de las bases de datos operacionales que la organización maneja. En esta investigación se hace uso de la metodología Hefesto para poder realizar el modelo multidimensional, partiendo de una base de datos operacional, esta metodología está dividida en cinco fases: dirigir y planear, recolección de información, procesamiento de datos, análisis y producción y por último está la difusión de la información procesada; cabe mencionar que nos hemos enfocado en las tres primeras fases de la metodología mencionada para así poder llegar a la construcción de un DataMart, que posteriormente se procederá a la migración de los datos de la base de datos operacional, a la base de datos multidimensional, también realizamos una comparación del tiempo de ejecución de consultas en los dos modelos de base de datos (MER y MMD), mostrando un tiempo menor en la ejecución de las mismas consultas en la base de datos multidimensional, garantizando así la rapidez en la que esta brinda la información.

¹ Estudiante de Ingeniería de Sistemas. Facultad de Ingeniería y Arquitectura. Universidad Peruana Unión. E-mail: edwin.casas@upeu.edu.pe

² Estudiante de Ingeniería de Sistemas. Facultad de Ingeniería y Arquitectura. Universidad Peruana Unión.

Palabras clave: Modelado, entidad, DataMart, hefesto, multidimensional.

ABSTRACT

Nowadays in the various activities that organizations conduct, data is generated, as a secondary product, which is the results of all operations conducted. These data become valuable information for the organization, with the passing of time they serve to make decisions and generate added value to the products and / or services that are offered. To make decisions based on this information, it should be apply business intelligence, which is responsible for collecting, processing and presenting organized information, which takes as one of the main activities the process of transformation and multidimensional modeling of operational databases that the organization manages. In this research is use the methodology of Hephaestus to perform the multidimensional model, from an operational database, this methodology is divided into five phases: directing and planning, data collection, data processing, analysis and production, and finally there is the dissemination of processed information. Be mentioned that we have focused on the first three phases in the above methodology in order to reach the construction of a DataMart, which will later proceed to the migration of the data of the operational base to the multidimensional database, also we make a comparison of the execution time of consultations on the two database model (MER and MMD), showing a shorter time in the execution of the same consultation in the multidimensional database. Thus, ensuring the speed at which this provides information.

Keywords: Modelling, organization, DataMart, Hephaestus, multidimensional.

INTRODUCCIÓN

En la actualidad, casi todas las organizaciones necesitan de la automatización de sus procesos mediante la implantación de sistemas de información transaccionales que den soporte a sus procesos operacionales, con el fin de reducir los costos y obtener una ventaja competitiva, a fin de dar un valor agregado a sus productos o servicios.

Con el transcurrir del tiempo, estos sistemas de información van almacenando datos en sus bases de datos que han sido diseñadas bajo la metodología del modelado entidad relación y el proceso de normalización, resultando en estructuras muy robustas que permiten almacenar información de todos los flujos de datos de los procesos involucrados en el proceso de automatización. Sin embargo, estas estructuras no están diseñadas adecuadamente para almacenar, procesar y presentar información que permita entre otras cosas, descubrir patrones de comportamiento de los clientes o de las diferentes dimensiones del negocio que se desea analizar,

perdiendo en consecuencia una excelente oportunidad para mejorar el apoyo en la toma de decisiones.

Es por ello que al diseñar una solución de inteligencia de negocios, basado por ejemplo en un datawarehouse, es necesario transformar el modelo de base de datos relacional del sistema operacional existente en una estructura que soporta ciertos niveles de redundancia y que además está diseñada para hacer un análisis del negocio en función a las diferentes dimensiones, perspectivas y características de sus procesos involucrados, denominando un modelo multidimensional.

En esta investigación se aplicó la metodología Hefesto, de Darío, R. (2010), para construir un modelo multidimensional en estrella, cuya estructura está diseñada para responder a necesidades de información de nivel táctico y estratégico de la organización, evitando con ello acceder a los datos de la base de datos de operaciones, reduciendo los inconvenientes de consultar sobre la base de datos en producción, afirmamos así que el tiempo y los recursos utilizados en los resultados de las consultas en el modelo multidimensional difiere significativamente respecto del modelo entidad relación.

Yalan y Palomino (2012), en uno de sus estudios realizados, muestran que a través de la elaboración de un DataMart se permitió apoyar al área que fue objeto de estudio haciendo entrega de oportuna y relevante de información, de antemano también se redujo el tiempo en la elaboración de los reportes y así apoyar con ello a la toma de decisiones.

Así mismo también Rojas (2014), al realizar su investigación sobre la implementación de un DataMart, concluye que se logró identificar los procesos que permiten llevar a cabo la toma de decisiones para así realizar un análisis de requerimientos de la empresa en estudio.

Según Manzano de Armas, J; Molina Perez, D. y Castro S, (2015) la metodología que se utiliza para la construcción e implementación del DataMart es HEFESTO, la cual puede adaptarse a cualquier ciclo de vida de desarrollo de un sistema. Tiene como objetivo entregar una primera implementación del DataMart que satisfaga parte de las necesidades y expectativas de los usuarios. Esta metodología se caracteriza por la facilidad en la que se distinguen los objetivos y resultados esperados en cada fase. Está basada en los requerimientos del usuario y es capaz de adaptarse rápidamente y con facilidad a cualquier cambio en el negocio.

Así mismo Manzano de Armas, J; Molina Perez, D. y Castro S, (2015) indican que la metodología HEFESTO está orientada a la construcción de almacenes de datos para Análisis Dimensional (OLAP) y está formada por las siguientes fases: 1). Análisis de requerimientos. En esta fase se identifica los requerimientos del usuario a través de una serie de preguntas. 2). Análisis de los sistemas OLTP. Es donde se procede a examinar el sistema el cual tiene toda la información requerida para el diseño del DataMart. 3). Modelo lógico del DataMart. Es donde se confecciona el modelo

lógico de la estructura del DataMart; teniéndola como base conceptual, que ya ha sido creada anteriormente, 4). Procesos de extracción, transformación y carga de datos (ETL). En esta fase se debe realizar el análisis, la definición y el desarrollo de los procesos.

Un DataMart, según IBM (1999) es un conjunto del Datawarehouse, con un alcance de contenido limitado. Este se usa para un solo departamento de una organización o un problema particular de análisis dentro de una organización. Una DataMart por sí sola, no es un Datawarehouse (DW), ya que un DW tiene más usuarios y más temas que un DataMart.

Según Inmon (2005) un DataMart se clasifica en:

DataMart dependiente

Los DataMart dependientes son aquellos que reciben los datos desde una Datawarehouse. En este tipo de DataMart la fuente de los datos es única.

DataMart independiente

Son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros Datawarehouse. Este tipo de DataMart se alimenta generalmente de las organizaciones.

DataMart híbrido

Los DataMart híbridos permiten combinar las fuentes de datos de un Datawarehouse corporativo con otras fuentes de datos tales como sistemas transaccionales u operacionales.

MATERIALES Y MÉTODOS

Participantes

Para la realización de esta investigación, el trabajo se inicia con el estudio de una base de datos relacional de un sistema comercial, diseñado por los estudiantes del IX Ciclo de la carrera de ingeniería de sistemas de la Universidad Peruana Unión filial Tarapoto, que abarca 6 procesos principales: Compras, Ventas, Inventario, movimientos, Cuentas por Cobrar, Cuentas por Pagar.

A fin de revisar la congruencia del modelo relacional, se diseñaron varias consultas para mostrar información sobre los procesos mencionados.

Luego de evaluada y puesto en prueba esta base de datos operacional, se obtuvieron resultados satisfactorios sobre su funcionamiento, por lo que se decidió iniciar

el proceso de población (llenado de datos), mediante algoritmos repetitivos y aleatorios para simular una base de datos con información histórica con cinco años de antigüedad.

Instrumentos

Se aplicó el modelado multidimensional, mediante la metodología Hefesto, para lo cual se realizaron los siguientes pasos de acuerdo a la metodología:

1. Reconocimiento del modelo entidad relación.
2. Identificación de procesos de negocios soportados por el modelo.
3. Identificación de las dimensiones de análisis del negocio.
4. Identificación de las tablas de acontecimientos en donde se almacena la información transaccional de los procesos identificados en el punto n° 2, dividiendo el procesos en:
 - a. Identificación de los atributos dimensionales.
 - b. Identificación de los atributos medibles.
 - c. Identificación de los atributos calculables.
 - d. Identificación de la dimensión tiempo.
5. Modelado multidimensional, aplicando el modelo estrella.
6. Migración de datos desde la base de datos operacional hacia la base de datos multidimensional.
7. Ejecución de consultas SQL sobre el modelo multidimensional a fin de verificar que se obtienen los mismos resultados tabulados, garantizando así que la información sea confiable.

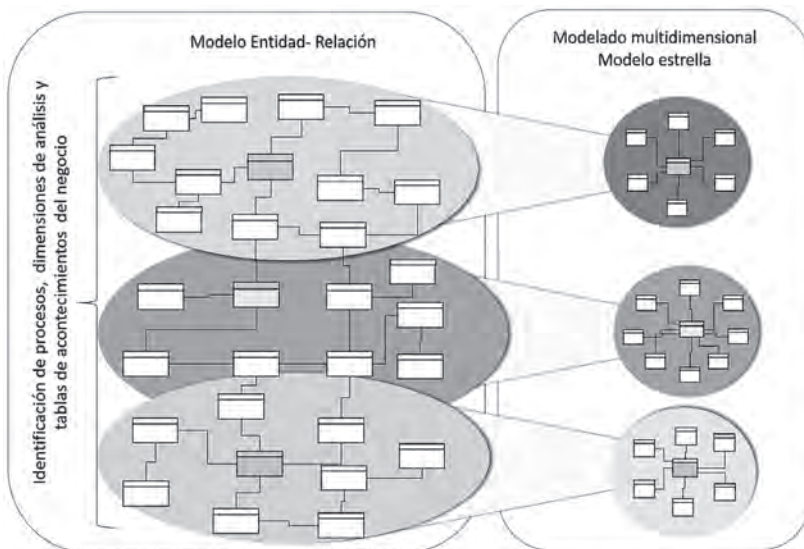


Figura 1. Diseño de la investigación.

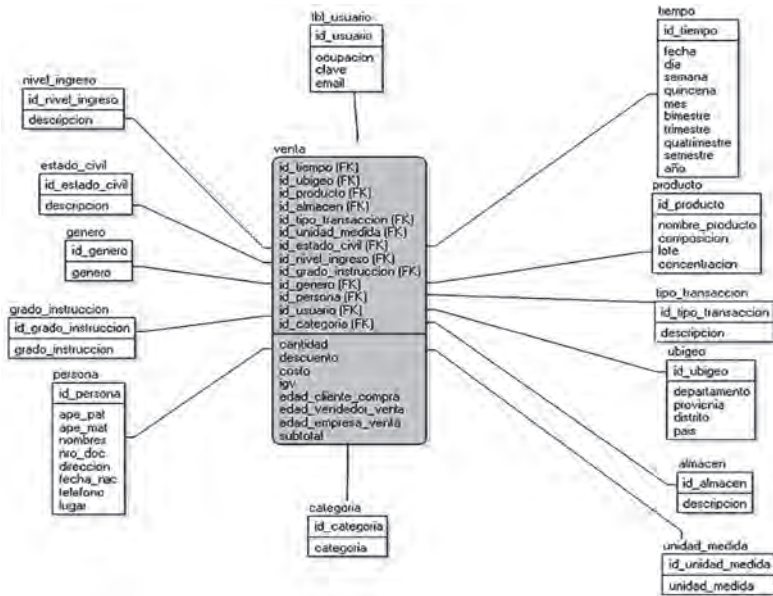


Figura 4. Modelo estrella de ventas.

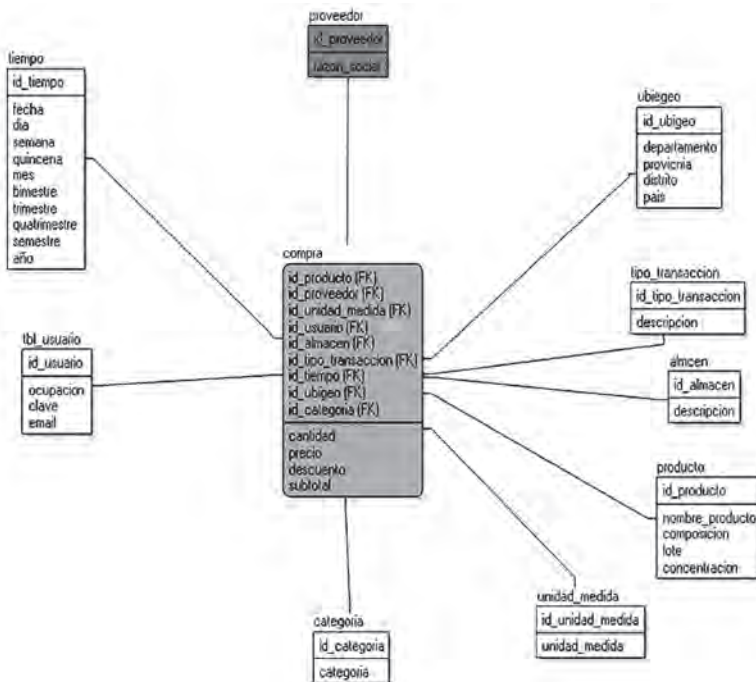


Figura 5. Modelo estrella de compras

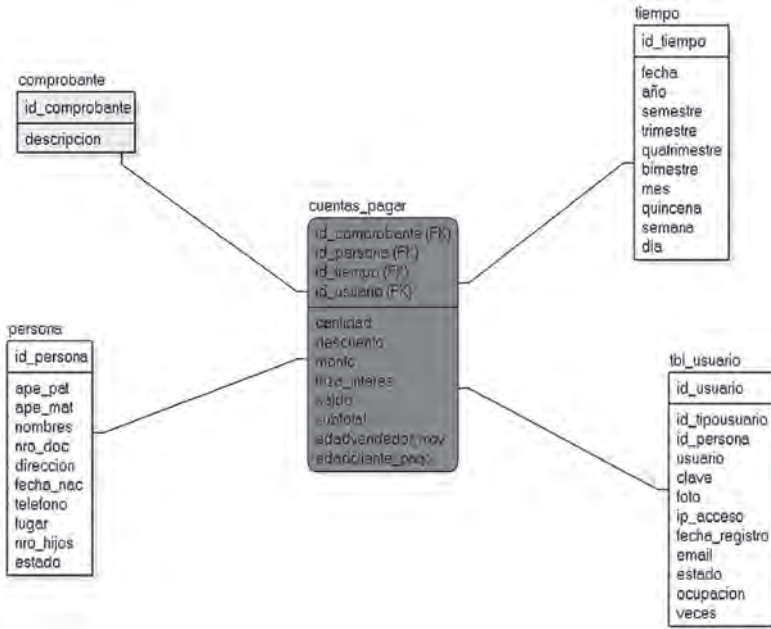


Figura 6. Modelo estrella de cuentas por pagar.

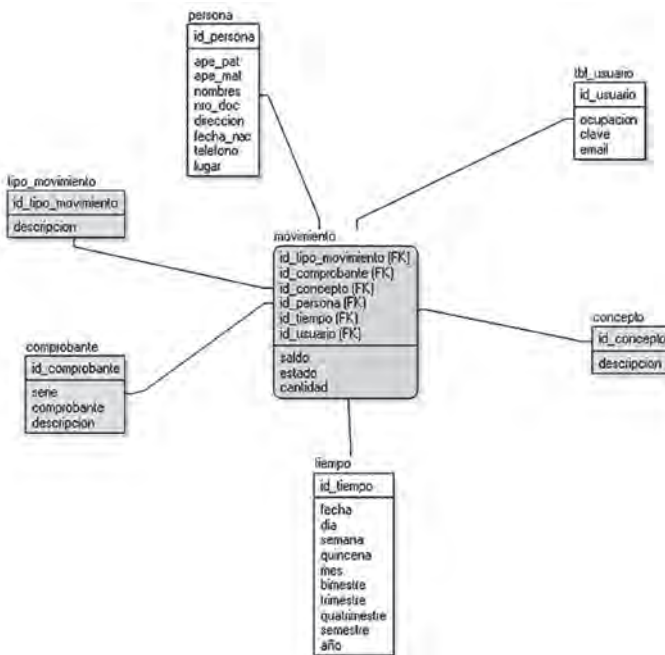


Figura 7. Modelo estrella de movimientos.

Resultados 2

También se muestra como segundo resultado una tabla en la que se compara el rendimiento de los tiempos de ejecución de las consultas sobre ambos modelos de base de datos.

Tabla 1

Comparación del rendimiento de los tiempos de ejecución de consultas SQL por modelo.

Consulta	Ejecución 1	Ejecución 2	Ejecución 3	Ejecución 4	Promedio
MER	0.203s	0.208s	0.0215s	0.160s	0.148125s
MMD	0.016s	0.015s	0.009s	0.0016s	0.0104s

De las consultas ejecutadas en 4 tiempos distintos se muestra que en el modelo entidad relación se obtiene un promedio de 0.148125s, mientras que para el modelo multidimensional se obtuvo un valor menor, el cual es 0.0104s.

CONCLUSIONES

De acuerdo a los resultados obtenidos en la investigación, podemos concluir lo siguiente:

- Se logró identificar los procesos de negocio, las dimensiones que el modelo entidad relación soporta, como también se logró identificar las tablas de acontecimientos en donde se almacena la información transaccional, esto tiene relación con la investigación de Rojas (2014), por qué se logró identificar todos los procesos involucrados que permiten llevar a cabo la toma de decisiones y se realizó el modelado en estrella de la base de datos multidimensional.
- El rendimiento de los tiempos de ejecución de las consultas diseñadas para cada modelo, muestran mejores resultados para el modelo multidimensional, esto entonces confirma el estudio realizado por Yalan y Palomino (2012), que se redujo el tiempo en la elaboración de los reportes para la toma de decisiones con la elaboración del DataMart.
- La metodología Hefesto, por sus características y el propio método es muy fácil de utilizar y aplicar para realizar el modelado multidimensional a cualquier base de datos que tenga una estructura relacional.

RECOMENDACIONES

Se recomienda hacer énfasis en las ventajas competitivas que ofrece este tipo de herramientas para la organización de manera que se pueda estimular la innovación y la integración de los procesos y la información.

AGRADECIMIENTOS

Agradecemos al Mg. Miguel Ángel Valles Coral, Docente del curso Administración de Base de Datos III, el cual contribuyo a través de su asesoramiento en el desarrollo de esta investigación.

Referencias

- Yalan, J & Palomino, L.(2012). *Implementación de un Datamart como una solución de Inteligencia de Negocios para el área de logística de T-Impulso*. Revista de Investigación de Sistemas de Información. Volumen(10) p 53-63.
- Rojas, A (2014). *Implementación de un data mart como solución de inteligencia de negocios, bajo la metodología de ralph kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría general de la república* (Tesis de pregrado) Universidad de San Martín de Porres.
- Manzano de Armas, J, Molina, D. y Castro S. (2015) *Implementación de un Data Mart para un servicio de Dosimetría Externa*. Argentina
- MySQL. (2007) Enterprise Data Warehousing with MySQL. MySQL Business White.
- IBM (1999) IBM - Fundamentals of Data Warehouse and Business Intelligence for Knowledge Management – Instructor Guide. Diciembre 1999. Página 48
- W. H. Inmon. Building the DataWarehouse. Editorial Wiley,