

Alternativas para la selección del diseño de las bases de datos operativas y gerenciales en unidades de información
Alternatives for selecting the database design and management of operational data into libraries

Mynor Fernández

Escuela de Bibliotecología y Ciencias de la Información
Universidad de Costa Rica
mynor.fernandez@ucr.ac.cr

Resumen

Una de las principales tareas de cualquier proyecto de automatización de unidades de información es el relativo a la selección del diseño de las bases de datos; sin embargo, es importante señalar que frecuentemente cuando se está seleccionando una aplicación para automatización generalmente se hace un análisis superficial de las bases de datos asociadas, prestando poca importancia a la completitud de las bases de datos y sus posibilidades de procesamiento para atender las necesidades de información requeridas, con la consecuencia de que esta forma de proceder probablemente conducirá a un fracaso del proyecto de automatización.

Este artículo trata sobre las diferentes alternativas que se le presentan al equipo responsable del proyecto de automatización, para seleccionar el diseño de las bases de datos que soportarán el proyecto de automatización, con la finalidad de señalar cuáles son los aspectos más importantes que se deben considerar para cada una de las opciones disponibles. Se debe insistir siempre en cualquiera de la alternativas disponibles requiere de un análisis meticuloso y crítico de las bases de datos asociadas y su capacidad para atender las necesidades de información identificadas, tanto operativas como gerenciales de la Unidad.

Palabras claves: Proyecto de automatización, diseño de bases de datos, bases de datos operativas, bases de datos gerenciales.

Abstract

One of the main tasks of any automation project developed in units of information is concerned with the selection of the design of the database, where it is important to note that often when it is selecting an application for automation, is usually done a superficial analysis of the associated databases, giving less importance to the completeness of the databases and their possibilities of processing for the needs of information required, with the likely result that this way of proceeding probably will lead to a failure of the automation project.

Therefore, this article discusses the different alternatives that are presented to the team responsible for the automation project, to select the design of databases that will support the automation project, in order to point out what the most important aspects to consider are for each of the available options. It should always insist on any of the available alternatives requires a thorough and critical databases associated analysis and its ability to meet the information needs identified, both operational and management level

Keywords: Automation project, databases design, operational databases, management database.

Introducción

Uno de los grandes retos que se le presenta al equipo interdisciplinario que dirige un proceso de automatización de unidades de información es el concerniente a la selección del diseño de la base de datos, tanto operativa como gerencial, que se utilizará para soportar las aplicaciones que atenderán las necesidades de información que presenta la Unidad.

Este trabajo tiene varias alternativas, la primera es si se utilizará un diseño a la medida que satisfaga todas las necesidades de información el usuario; para esta opción el trabajo es más laborioso ya que se deben de crear las bases de datos a la medida del usuario y también implicará la programación de aplicaciones personalizadas que utilicen estas bases de datos creadas según las necesidades del usuario.

Esta alternativa, reviste especial importancia, ya que aunque presentaremos otras, el diseño mixto y el diseño empaquetado, las actividades de identificación de datos, que se realizan en el diseño a la medida para identificar los requerimientos del usuario y diseñar las correspondientes bases de datos, también deben ser realizadas en las otras dos alternativas, esto con la finalidad de verificar que las bases de datos prediseñadas cumplan con los requerimientos del usuario, en el caso del diseño empaquetado, o bien que se pueda determinar la magnitud de los cambios que deben realizarse, en el caso del diseño mixto.

Otra alternativa es si se decide utilizar un diseño empaquetado junto con una aplicación denominada frecuentemente como paquete de software. En este caso, el equipo interdisciplinario tiene un margen limitado de decisión sobre el diseño de la base de datos que utilizará, ya que este viene prediseñado. Lo que procede aquí es evaluar si la base de datos prediseñada cumple con las necesidades de información que se deben atender, tanto presentes como futuras. Es importante señalar que la subvaloración de la importancia que representa esta actividad en el proyecto de automatización puede provocar el fracaso de este, cuando llegue el momento y se determine que el diseño de las bases de datos no satisface las necesidades de información que se deben atender en la Unidad.

Una tercera alternativa que se tendría sería la utilización de diseño mixto que se da cuando se utiliza una aplicación que permita obtener y modificar el código fuente de esta, así como la modificación del diseño de la base de datos prediseñado que viene junto a ella. De hecho esta alternativa que se podría definir como la combinación de las dos anteriores, requiere un mayor esfuerzo que la segunda alternativa que es la utilización de

un software empaquetado con su propia base de datos prediseñada y un menor esfuerzo que la primera alternativa que es un desarrollo completo a la medida tanto de la base de datos como de la aplicación que se utilizará.

En suma, el diseño o selección de la base de datos que se utilizará en el proyecto de automatización de unidades de información es una decisión técnica y estratégica que debe ser tomada luego de un cuidadoso análisis de las necesidades de información que deberá atender el proyecto de automatización en el nivel operativo, lo cual consiste en el estudio detallado del nivel transaccional del día a día que se debe atender en la unidad de información, como el nivel gerencial que debe ser satisfecho con las consultas a nivel gerencial que se requieran satisfacer a partir de las bases de datos diseñadas.

Es oportuno señalar que cualquiera de la alternativas para la selección del diseño de la base de datos que se elija no existe posibilidad de tener un diseño completo y perfecto en primera instancia, ya que como nos dice Johnson, M., & Kasangian, S. (2010), este es un proceso continuo y en constante mejora, de hecho existen gran cantidad de bases de datos reales que incluyen datos incompletos.

Diseño a la medida

El diseño a la medida implica un trabajo de análisis y diseño de sistemas realizado por equipos interdisciplinarios de ingenieros de sistemas y bibliotecólogos que analizan tanto el entorno como el interior de una unidad de información, para identificar todos los flujos de datos que circulan dentro de la unidad así como los flujos de datos que ingresan y salen de la misma hacia otras unidades externas.

A partir de estos análisis de flujos de datos se esbozan las entidades de información que agrupan los datos en estructuras conceptuales de información que consisten en unidades lógicas de información que agrupan un conjunto de atributos para describir un objeto, persona, lugar u organización.

En una primera parte del análisis se detectan todos los requisitos de datos que tienen relación con las transacciones que se realizan en día a día en la organización para obtener así el diseño de las entidades de información y sus atributos que satisfacen las necesidades operacionales de la unidad de información. Luego como parte de este trabajo creativo de diseño se visualizan cuáles son los atributos que se les deberán agregar a estas entidades diseñadas para satisfacer las necesidades de información que se deberán compensar en la unidad de información.

Completadas tanto las necesidades operativas como las necesidades gerenciales se tienen conformadas las entidades de información y sus atributos que formarán el diseño del modelo conceptual de la base de datos a la medida que soportará el proyecto de automatización. El siguiente paso será establecer cuáles serán los atributos de cada entidad que serán elegidos como la llave primaria para el acceso inequívoco de la información albergada por cada una de las entidades. Luego se deberán establecer cuáles atributos serán elegidos como llaves secundarias para el acceso múltiple de la información almacenadas en cada una de las entidades.

En resumen, luego de la identificación de las entidades de información, se deberá realizar un proceso de diseño conceptual que permita definir cuáles son las llaves primarias y secundarias de cada una de las entidades, así como las relaciones que se dan entre cada una de ellas. Luego se pasa a un diseño lógico donde las entidades identificadas con sus llaves primarias y secundarias, así como las relaciones entre ellas, se transforman en un modelo relacional compuesto de tablas, claves primarias y secundarias, así como las relaciones entre las tablas. Aquí los diseñadores deben tomar en cuenta lo que señalan Johnson, M., & Kasangian, S. (2010), que uno de los más notables desajustes entre la teoría y en la práctica es la distinción entre la teoría de la modelo de datos relacional y las bases de datos del mundo real, por tanto el proceso de diseño es un proceso continuo y en constante mejora.

Este aspecto también lo confirma Wilson M. (1981) quien dice que los requisitos de datos, una vez determinados, deben ser mapeados en el modelo de datos lógico y a las estructuras de almacenamiento físico soportadas por sistema administrador de base de datos (DBMS). Aunque aspectos técnicos de diseño pueden variar de un modelo de datos y DBMS a otro; los criterios generales de diseño se mantienen, como son: redundancia de datos al mínimo, evitan anomalías de actualización, búsqueda de un rendimiento adecuado, y facilidad de cambio de diseño. Finalmente, después de que la base de datos se ha implementado, el diseño lógico y físico puede ser ajustado.

Debido a este arduo y laborioso trabajo de análisis y construcción es importante mencionar que el diseño a la medida es una solución que se utiliza cada vez con menor frecuencia, a causa del alto costo que representa tener un equipo interdisciplinario enfocado en visualizar y construir las entidades de información requeridas en una unidad de información que finalmente será la materia prima para construir las bases de datos operativas y gerenciales que deben satisfacer las necesidades de información que presenta la Unidad.

Como un ejemplo simple de esto, supongamos que tenemos la entidad libro que estará conformada por los siguientes atributos: sigla del libro, ISBN del libro, título del libro, editorial, país, localización física, número de páginas, entre otros atributos que describen las necesidades de información de una Unidad. El campo ISBN del libro podría ser la llave primaria de esta entidad, ya que es un campo único, mientras que el título podría ser una llave secundaria que es un campo múltiple, ya que varios libros podrían tener el mismo título.

Luego podríamos identificar la entidad usuario que estaría formada por los siguientes atributos: número de identificación de usuario, nombre de usuario, teléfono de usuario, dirección email, dirección residencia, género del usuario, profesión del usuario, entre otros. Donde el atributo número de identificación podría ser la llave primaria, por ser un campo único y el nombre del usuario podría ser una clave secundaria, ya que varios de los usuarios podrían tener el mismo nombre.

Continuando identificamos la entidad préstamo que está conformada por los siguientes campos: ISBN del libro, identificación de usuario, fecha de préstamo, fecha de devolución, días de atraso, monto de la multa. Donde la llave primaria sería la unión de los tres atributos ISBN del libro, identificación del usuario y fecha de préstamo, para tener un acceso único a un préstamo, suponiendo que en una fecha solo se hace un préstamo

de un determinado libro a un usuario. Mientras que ISBN del libro o identificación de usuario podrían convertirse en llaves secundarias de esta entidad.

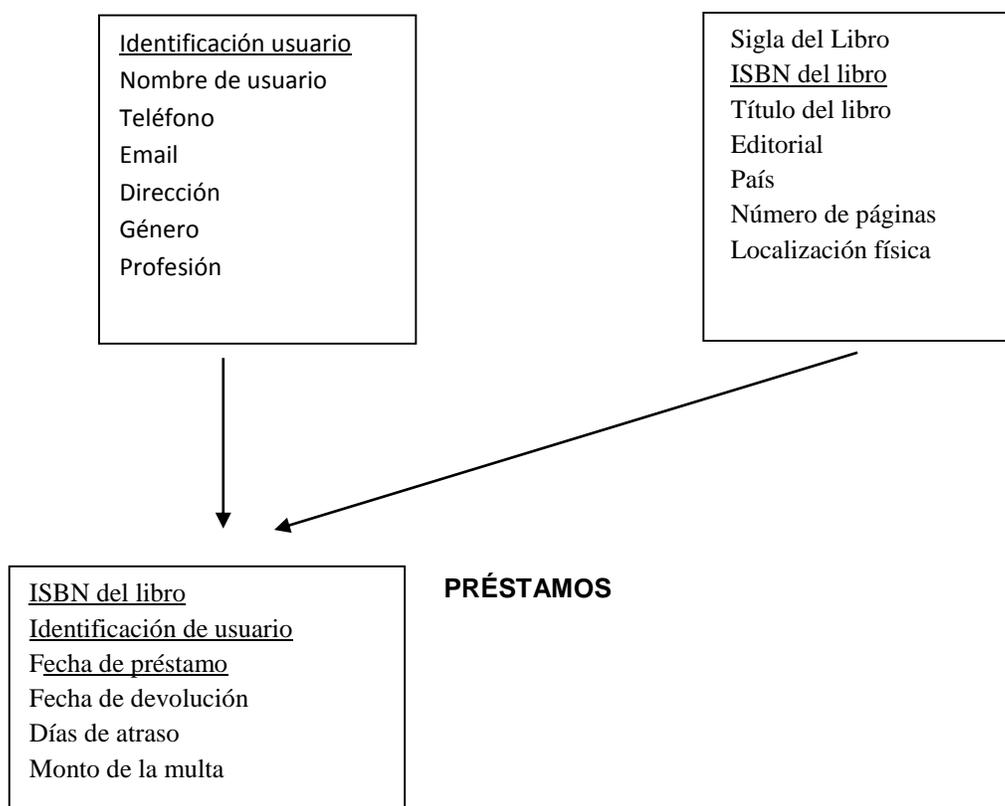
Este es un simple ejemplo, de tres entidades con pocos atributos cada una que se establece con el objetivo de ilustración de este trabajo de crear un diseño a la medida, donde claramente se puede visualizar que entre mayor número de entidades surjan con múltiples atributos y definición de claves primarias y secundarias, así como la mayor identificación de múltiples relaciones entre ellas. La complejidad de este trabajo diseño a la medida aumenta y la posibilidad de fallo se incrementa, por la posible no identificación de entidades, atributos o relaciones existentes que producirán un diseño a la medida incompleto que no podrá satisfacer las necesidades de información de la unidad en estudio.

El siguiente paso en el diseño a la medida sería conformar un diseño entidad relación de nuestra base de datos que estaría integrado por las tres entidades identificadas a este momento con sus respectivas relaciones. Donde es importante señalar lo que dice Cedeño Trujillo, A (2006).

El modelo entidad-relación (MER) es una técnica poderosa para el diseño de sistemas transaccionales en el entorno de las bases de datos relacionales. Permite la normalización de la estructura de datos física, obteniéndose un diseño sin redundancias en lo datos y ocupándose el menor espacio de almacenamiento.

El modelo entidad relación respectiva se visualiza en la figura 1:

Figura 1. Diseño conceptual a la medida



Hasta aquí podríamos señalar que la cantidad de atributos que describe cada una de las entidades identificadas va a depender de las necesidades de información que tenga la unidad de información donde se está realizando el análisis, lo que definirá cuál será la extensión total de cada entidad identificada.

Luego de tener el diseño conceptual de la base de datos se deberá someter el diseño a un proceso de normalización del diseño para obtener el modelo relacional de la base de datos, donde las entidades se convierten en tablas cuyos campos de acceso serán las llaves primarias y secundarias definidas. A través de este proceso de normalización se eliminarán los problemas de redundancia de la información, ya que se erradica la duplicidad de la información, para garantizar una mayor integridad de la bases de datos al reducir al máximo los problemas de actualización que se dan como producto de información duplicada.

Así vemos que este es un proceso complejo y laborioso que implica la identificación de todas las entidades de información que satisfacen tanto las necesidades operativas como las necesidades gerenciales de la unidad de información y ahí la importancia que se realice con equipos interdisciplinarios conformados por ingenieros de sistemas especialistas en diseño y bibliotecólogos especialistas en necesidades de información en una Unidad. Pero al ser un diseño a la medida, existe la posibilidad de crear todas las entidades con sus atributos requeridos para soportar las necesidades de información, tanto operativas como gerenciales de la unidad de información.

Diseño empaquetado

Por diseño empaquetado entendemos un diseño de base de datos, como un conjunto de entidades, sus atributos y relaciones ya preestablecido, o sea que ya viene incorporado con una aplicación de paquete para resolver una necesidad de automatización en una unidad de información. Los responsables del proyecto de automatización deben ser muy cuidadosos en el análisis de estas bases de datos y sus campos que las describen, por cuanto son formatos preestablecidos que se crearon para una aplicación genérica que resuelve las necesidades de múltiples unidades de información.

Este diseño es utilizado cuando se selecciona una aplicación, denominada frecuente paquete, para automatizar una unidad de información. Existen dos posibilidades, una aplicación privativa donde la única forma posible de hacerle cambios es contratar al proveedor de la aplicación o un tercero autorizado por este para hacerle cambios a las bases de datos y a las aplicaciones asociadas, donde en la mayoría de los casos son costos prohibitivos para una unidad de información de bajo presupuesto. La otra posibilidad es utilizar una aplicación de software libre que permita hacer cambios tanto en los programas fuentes de la aplicación, como en sus bases de datos. A esta opción le llamaremos diseño mixto y se cubrirá más adelante en este mismo artículo.

Bajo cualquiera de las dos opciones se debe enfatizar que el equipo interdisciplinario de automatización deberá ser meticuloso en el análisis de las tablas y atributos que vienen prediseñados con la aplicación. Ya que como se afirmó anteriormente, el diseño empaquetado tiene como objetivo suplir los requerimientos de distintas unidades de información, siendo así un diseño genérico, donde la unidad de información deberá adaptarse a este, a través de cambios en los procedimientos que se realizan en la Unidad,

esto debido al costo prohibitivo que representa la realización de cambios en estas aplicaciones y bases de datos asociadas; aspecto de costo que se maximiza si se trata de una aplicación privativa.

Cuando no se presta atención a la necesidad de realizar un análisis detallado de estas aplicaciones genéricas que obliga a utilizar un diseño empaquetado, se condena a la unidad a realizar cambios en los procedimientos de trabajo con la finalidad de adaptarse a los requerimientos técnicos que presentan los paquetes de software orientados a automatizar una unidad de información. En suma, la unidad debe adaptarse a la aplicación y no la aplicación a la unidad.

Diseño mixto

Por diseño mixto se entenderá cuando se selecciona una aplicación con base de datos preestablecidas, las cuales pueden ser modificadas por ingenieros de sistemas especialmente contratados para adaptarlas a las necesidades de información de la Unidad, para esta opción se presentan dos opciones; la primera, la utilización de una aplicación privativa, con el pago de derechos de licencia por su uso y la contratación del proveedor o de un tercero autorizado a modificar la aplicación; la segunda, la utilización de una aplicación de software libre con posibilidad de modificación de los programas fuentes.

Para este propósito existen múltiples aplicaciones de software libre con licencias GNU o GPL, que permiten la modificación del código fuente y de las bases de datos asociadas. Davis D. & Jabeen I. (2011) definen el sistema operativo GNU/Linux como uno de los más conocidos ejemplos de Software Libre y de Código Abierto (FOSS). El Tratado de Libre y Open Source Software es un paradigma que fomenta un sentido de y participación comunitaria, basándose en el conocimiento compartido.

Con el diseño mixto se puede alcanzar la versatilidad de diseño que nos permite el diseño a la medida, explicado en la sección anterior; sin embargo, es importante señalar que dependiendo del grado de cambios que deban realizarse en las aplicaciones y las bases de datos para su adaptación requerida por la unidad de información, podría ser que el esfuerzo en tiempo y costo producto de estas actividades de ajuste del software, iguale o supere el tiempo y costo que implica un diseño a la medida.

Igualmente como el diseño empaquetado se debe hacer un análisis meticuloso de las bases de datos asociadas para determinar el nivel de cambios que deberán realizarse tanto en las bases de datos como en las aplicaciones correspondientes.

Aunque no es el objetivo de este artículo queda claro que para tomar la decisión de realizar un diseño a la medida versus la opción de adoptar una aplicación privativa versus la alternativa de utilizar una aplicación de software libre con posibilidad de modificar la aplicación y sus bases de datos se deberá realizar un estudio de factibilidad que abarque tanto la factibilidad económica, la factibilidad técnica, la factibilidad operativa y la factibilidad de plazo, donde se muestren los costos y beneficios de cada una de las opciones.

Debe quedar claro a este nivel que el software libre lo que representa es el no pago de costos por licencia de uso de las aplicaciones y de las bases de datos asociadas, pero sí involucra el pago por servicios profesionales al personal técnico responsable de realizar los cambios en las aplicaciones y bases de datos, así como por el proceso de implantación de las mismas en la unidad de información.

Finalmente es importante destacar que generalmente se pone poca atención en la selección del diseño de base de datos en un proyecto de automatización, por lo que generalmente se selecciona una aplicación con sus correspondientes base de datos preestablecidas, sin prestar suficiente atención a las necesidades de información que se deberán atender en el futuro; de esta manera, cuando llega el momento de satisfacerlas, se dará una alta probabilidad de que sea imposible atenderlas por la inexistencia de datos y de ahí se puede dar el fracaso del proyecto de automatización.

Así, son múltiples los proyectos de automatización que ponen poca atención en la calidad y completitud del diseño de las base de datos, tanto a nivel operativo como a nivel gerencial, ya que se enfocan primordialmente en la evaluación de si el nivel de ejecución y exactitud de las transacciones del día a día se realiza satisfactoriamente con la aplicación elegida.

Base de datos operativos

Por bases de datos operativas entenderemos todas las tablas, atributos y relaciones que soportan todas las transacciones del día a día que se realizan en la unidad de información. Esto es independiente de la alternativa de diseño que se seleccione para disponer de las bases de datos en la Unidad.

Así estas transacciones son de suma importancia para el usuario final de la Unidad así como para los bibliotecarios que trabajan en ella. Dentro de estas transacciones podemos distinguir, a manera de ejemplo, las siguientes:

- Préstamo de un libro
- Devolución de un libro
- Reserva de un libro
- Alerta de un libro
- Cálculo de multa por retraso
- Consulta del catálogo por autor
- Consulta del catálogo por título
- Consulta del catálogo por materia
- Inclusión en el catálogo de un nuevo libro
- Eliminación del catálogo de un libro
- Modificación de los atributos que caracterizan a un libro en el catálogo
- Reporte de usuarios ordenados por nombre
- Reporte de libros prestados con fecha de vencimiento
- Reportes del material catalogado clasificados por título, autor u otra variable

De esta forma, podríamos seguir listando las diferentes transacciones que se realizan en el día a día en una unidad de información hasta tener una lista exhaustiva de todas ellas, y de las transacciones que están asociadas a una o más tablas que contienen los atributos requeridos hasta obtener las bases de datos operativas que soportan a la aplicación o aplicaciones que implementan cada una de las transacciones identificadas en la Unidad.

Herramientas para extraer información operativa

La principal herramienta para extraer información de las bases de datos operativas las constituyen las propias aplicaciones que implementan las transacciones del día a día identificadas. Los datos que se extraen para su presentación se realizan con la ejecución de cada una de las transacciones que satisfacen las necesidades de información que soporta la aplicación y las bases de datos operativas asociadas.

Es claro que teniendo las bases de datos de la aplicación se pueden utilizar múltiples herramientas independientes a esta, que permiten utilizar bases de datos relacionales para incrementar las salidas de información con diferentes niveles de detalle, según las características técnicas de la herramienta y de las necesidades de información identificadas. Así entre otras herramientas tenemos las más comunes como son generadores de reportes y generadores de gráficos que facilitan la producción de información con múltiples perspectivas de visión a partir de la tenencia de estas bases de datos.

Es importante señalar que uno de los factores claves de éxito de un proyecto de automatización es tener un adecuado y completo diseño de las bases de datos, ya que a partir de este existen múltiples herramientas de terceros, tanto libres como privativas para extraer información operativa con diferentes niveles de detalle para satisfacer la variedad de las múltiples salidas de información que se requiere en la Unidad.

A continuación se presentan algunos enlaces de posibles aplicaciones en software libre que podríamos utilizar para este propósito, dejando claro que existen una gran variedad de herramientas para este fin:

- Jasper Reportes: <http://www.nan-tic.com/es/empresa/rd/jasper-reports/>
- Report Management: <http://reportman.sourceforge.net/>
- Birt: <http://www.eclipse.org/birt/>
- iReports: <http://ireport.sourceforge.net/cap3.html>

Bases de datos gerenciales

Por bases de datos gerenciales entendemos las que conforman todas las tablas, atributos y relaciones que soportan todas las transacciones de nivel gerencial que se determinan en un estudio de requerimientos en la unidad de información, donde se debe considerar lo que dicen Cravero, A., Sepúlveda S., Mazón J & Trujillo, J (2013).

En este contexto, el objetivo primordial de la ingeniería de requerimientos para un AD (almacén de datos) es representar a los usuarios, sus objetivos y las relaciones entre los mismos, con el fin de alcanzar los objetivos estratégicos del negocio. Esta etapa es crucial en el desarrollo del AD, ya que generalmente las partes interesadas presentan necesidades que están relacionadas con sus propios objetivos y no a los organizacionales y, por tanto, el AD final puede no estar alineado a la estrategia del negocio.

Así por nivel gerencial se entiende todas las necesidades de información, cuyo objetivo principal es la toma de decisiones a mediano y largo plazo. Así las bases de datos gerenciales son almacenes de datos producto de resúmenes, procesamiento o proyecciones que se realizan a partir de las bases de datos operativas, con la finalidad de facilitar la generación de las consultas gerenciales requeridas. Lo primordial de estas bases de datos es su contribución con la gestión estratégica de la Unidad, donde se considera lo que dice Fernández M. (2014)

Se favorecerá la gestión estratégica a través del establecimiento de perspectivas de gestión, que son en realidad diferentes vistas conceptuales de la organización para las que se definirán objetivos estratégicos asociados. Estos permitirán el desarrollo organizacional de cada vista, relacionándolas con indicadores estratégicos que permitan visualizar el grado de logro en el tiempo de los objetivos planteados.

De esta forma, se pueden almacenar estos indicadores asociados con objetivos estratégicos en las bases de datos gerenciales para facilitar que las bases de datos gerenciales estén alineadas con la estrategia de la organización y así favorecer la toma de decisiones de alto nivel.

Aquí aparecen varios tipos de almacenes de datos donde, en primera instancia, se encuentra el data warehouse que es una estructura de base de datos utilizada para el procesamiento gerencial de la información de toda la organización. En un data warehouse organizacional se integran tanto fuentes internas como externas de diversa índole, donde se da la posibilidad de que existan fuentes dispares y heterogéneas, pero el objetivo de esta estructura es crear un repositorio central de datos homogéneo que frecuentemente es llamado data warehouse.

En un data warehouse se almacenan tanto datos actuales como históricos de diversas fuentes internas y externas de la organización y estos son utilizados para la generación de informes y la realización de consultas de alto nivel, tales como proyecciones y análisis de gestión. Así lo afirman Neil, C, De Vincenzi, M. & Pons, C, (2014), cuando dicen que las empresas utilizan los datos operacionales acumulados en los años y almacenados en las estructuras que se denominan data warehouse para ayudar a entender y manejar sus actividades.

Mientras que se mantiene una base de datos operacional con los datos actuales, el data warehouse mantiene un histórico datos de la empresa, como resultado, estas estructuras de datos, tienden a crecer constantemente con el tiempo por lo que requieren una alta capacidad de almacenamiento.

En la segunda instancia de este nivel gerencial aparece el almacén de datos que se llama data mart, que son estructuras que se especializan en un tema o un área específica de una organización. El data mart cubre una parte de la organización, mientras que el data warehouse cubre toda la organización. En otras palabras, también podemos decir que los data marts son subconjuntos de datos orientados a un área funcional o un área administrativa dentro de una organización, que se crean con la finalidad de facilitarle la toma de decisiones a esta área, donde un data mart se podría ver como un almacén de datos que es subconjunto de un data warehouse, También se puede decir que la unión de todos los data mart de la organización generan el data warehouse organizacional. Según Wolf (citado en Chinchilla Arley, R (2011) un data mart es definido como:

(...) un conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un áreas o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo solo sentido para el personal de ese departamento y sus datos no tienen por qué tener las mismas fuentes que otro datamart.

Es importante enfatizar que gran parte de los datos almacenados en los data mart como en los data warehouse se cargan desde los sistemas transaccionales de la organización con sus correspondientes bases de datos operativas asociadas. Tal como lo señala Poe (citado en Chinchilla Arley, R (2011)) sobre los data warehouse: “una base de datos de solo lectura, donde la información extraída de los sistemas operacionales corrientes de la empresa es transformada, integrada y resumida para luego ser usada con efectividad en el soporte de decisiones”.

Por eso, cuando se trata con data warehouse, realmente se está hablando de bases de datos operativas procesadas. En otras palabras, la materia prima de los data warehouse la constituyen las bases de datos operativas, ya que estas almacenan el resultado de las transacciones del día a día, mostrando un nivel de detalle de estas, mientras que las base de datos gerenciales producen resúmenes o agrupamiento de estas transacciones, que muestran resultados globales tanto de comportamiento como de proyección, que facilitan realizar análisis de alto nivel de la actividad de la organización para facilitar la toma de decisiones. Como lo dice en resumen Cedeño Trujillo, A. (2006).

Data Warehouse, es una tecnología para el manejo de la información, que soporta el procesamiento informático y provee una plataforma sólida que permite realizar análisis a partir de datos históricos y actuales. Su función esencial es ser la base de un sistema de información. Facilita la integración de sistemas de aplicación no integrados proveniente de fuentes de datos heterogéneas (bases corporativas, bases propias, de sistemas externos, ficheros, etc.), brinda una visión integrada de dicha información, especialmente enfocada hacia la toma de decisiones por parte del personal de la organización.

Herramientas para extraer información gerencial

Aquí se debe indicar que lo esencial e importante es que el diseño de la base de datos contemple la solución de las necesidades gerenciales identificadas en el estudio de requerimientos. Para este punto es importante considerar lo que nos dicen Cravero, A., Sepúlveda S., Mazón J & Trujillo, J (2013).

Más que una simple recopilación de datos, el AD surge de un proceso definido en tres etapas: (1) extracción de datos de distintas fuentes de datos, (2) la transformación y carga de datos de forma coherente en el AD [1] y (3) el acceso a los datos integrados de una manera eficiente y flexible. Las dos primeras etapas forman parte del proceso conocido como Extraction-Transformation-Load (ETL).

Donde sí se tiene el adecuado diseño de los almacenes data marts y data warehouse, existen múltiples herramientas de terceros que se pueden utilizar para extraer la información gerencial requerida por la Unidad. Para este fin se dispone de las herramientas ETL que, por su siglas en inglés, son herramientas orientadas a la extracción, transformación y carga, probablemente en data warehouse o data marts. Una lista de estas herramientas se puede encontrar en el siguiente enlace: <http://www.etltool.com/list-of-etl-tools>

Por otra parte, también debemos tomar en cuenta lo que nos dice Timarán-Pereira, R. (2013) que los modelos relacionales de sistemas de bases de datos actuales están diseñados principalmente para aplicaciones de negocios al nivel operativo. Así el éxito del lenguaje SQL está ligado a la posibilidad de que primitivas de este lenguaje apoyan la vasta mayoría de estas aplicaciones operativas, no las gerenciales, pero para consultas de tipo gerencial se debe tomar en cuenta lo que dice Cedeño Trujillo, A. (2006).

Una técnica mucho más poderosa para la interrogación de los datos, es el modelo dimensional o multidimensional. El modelo multidimensional, es mucho menos riguroso en cuanto a organización; le permite a analistas y diseñadores más flexibilidad en el diseño, para lograr un mayor desempeño y optimizar la recuperación de la información, desde un punto de vista más cercano al usuario final.

Para soportar la familia emergente de nuevas necesidades para el procesamiento de los data marts y data warehouse, requerirán de herramientas OLAP (procesamiento analítico en línea que trabajan sobre modelos multidimensionales).

Tanto los data warehouse como los data mart reúnen información gerencial, almacenada en bases de datos centrales, diseñadas especialmente para su utilización en la atención de consultas gerenciales de alto nivel, para lo que se utilizan estas herramientas tipo OLAP. Según nos dice Chinchilla Arley, R (2011) que define el OLAP como:

El OLAP incluye, principalmente, la incorporación de grandes cantidades de datos dispersos y puede abarcar millones de datos con relaciones complejas. Su objetivo es contribuir al análisis de estas relaciones mediante la facilitación de búsqueda de patrones, tendencias y excepciones. Debe soportar procesamiento lógico y estadístico de los resultados, sin que el usuario tenga que programar, además de incluir los requerimientos de seguridad para la confidencialidad y las actualizaciones recurrentes.

Infraestructura tecnológica requerida

Frecuentemente cuando se utilizan los términos bases de datos gerenciales, data mart, data warehouse, herramientas OLAP, se nos viene a la mente grandes instalaciones de equipo de cómputo y costos elevados que están vedados para bibliotecas pequeñas o bibliotecas medianas. Sin embargo, hoy en día, con el alto desarrollo que existe en computación en la nube, donde se tiene la posibilidad de utilizar múltiples servidores virtuales debidamente enlazados y sincronizados para almacenar tanto las bases de datos operativas, así como los almacenes data marts y data warehouse que almacenan la información gerencial de una Unidad. Esto con costos asociados relativamente bajos. Donde además, como ya se explicó anteriormente se abre la posibilidad de poder utilizar herramientas de terceros para el procesamiento de la información almacenada en estos almacenes de datos.

Bajo este panorama, de una amplia utilización de la nube y sus posibilidades de procesamiento y almacenamiento de bajo costo, se hace una realidad totalmente factible que tanto una biblioteca pequeña o mediana con bajo nivel de presupuesto pueda disponer de estas capacidades de procesamiento y almacenamiento de la información, tanto operativa como gerencial; aspecto que antes solo era posible para unidades grandes y sofisticadas con altos niveles de presupuesto, ya que se requerían grandes y costos instalaciones de cómputo.

Tal como lo señala Collins (citado en Fernández, M. (2011) que nunca antes una serie de iniciativas, tanto de *hardware* como de *software*, había tratado de reducir el costo de la computación de forma que las comunicaciones y el acceso a la información, dados por un hecho en el mundo desarrollado, se encuentren más disponibles en el resto del planeta.

En resumen, una unidad de información que quiera disponer de estas posibilidades de almacenamiento y procesamiento operativo y gerencial de la información, lo que necesita son equipos de bajo costo con capacidad de conectarse internet, además de contar con un enlace de banda ancha de acceso a internet que le permita la posibilidad de utilizar múltiples servidores de bases de datos debidamente enlazados y sincronizados a través de internet.

Estos equipos pueden ser notebooks o tablets que se encuentran en una alta diversidad de precios en el mercado, pero para este propósito se requieren de equipos con precios bajos. Debido a que si se tiene la información almacenada en la nube y se procesa en servidores virtuales asociados, a través de un enlace de banda ancha a internet; enlace que se caracteriza por presentar cada vez un costo relativo menor, ya que como transcurre el tiempo, los proveedores de estos enlaces los ofrecen cada vez con mayor capacidad, tanto de megabytes transmitidos como de megabytes almacenados, y a un precio menor. De esta forma, no se requiere que los equipos tengan sofisticadas configuraciones con un alto nivel de procesamiento o una alta capacidad de almacenamiento que son los factores técnicos que elevan el costo de los mismos.

Conclusiones

La selección de cualquiera de las alternativas para diseño de las bases de datos operativas o gerenciales que soportarán la satisfacción de las necesidades de información de una biblioteca es primordial y se convierte en un aspecto central, que debe ser realizado con un estricto análisis que cubra todas las demandas de información que surgirán a corto, mediano y largo plazo en la unidad de información. Para esto el analista debe tener una alta capacidad de análisis, producto de una alta formación académica y una alta experiencia en unidades de información, que le permita visualizar tanto de las necesidades actuales como de las posibilidades que se darán en el futuro. Queda claro, que el descuido o la subvaloración de la importancia de esta tarea, de seguro provocará el fracaso del proyecto de automatización.

La denominación de bases de datos operativas o bases de datos gerenciales se refiere a dos visiones sobre el tipo de transacciones que soporta la base de datos, por un lado las transacciones del día a día y por otro las transacciones de alto nivel orientadas a satisfacer las necesidades gerenciales. Pero es importante tener claro esta división, ya que independientemente de la alternativa de diseño que se escoja, las bases de datos asociadas al proyecto deberán satisfacer las necesidades de estas dos visiones para el éxito del proyecto.

Un proyecto automatización que involucre la selección de un diseño de bases de datos para una unidad de información que abarque tanto las necesidades operativas como gerenciales de información, así como el uso de múltiples almacenes de datos, data marts y data warehouse, debidamente enlazados y sincronizados, que necesite mantener costos razonables dentro de los límites presupuestarios de la unidad, debe ser apoyado con almacenamiento y procesamiento en la nube a partir de la utilización de enlaces de internet de banda ancha de alta capacidad, que faciliten la realización del proyecto, así como la reducción de costos de este.

El establecimiento de las herramientas que se utilizarán para extraer la información de las bases de datos es un aspecto secundario y de menor importancia que el concerniente al diseño adecuado y completo de las bases de datos, tanto operativos como gerenciales. Ya que si se dispone de estas bases de datos debidamente diseñadas se pueden utilizar múltiples herramientas de terceros, tanto privativas como de software libre, para extraer y presentar la información que satisfaga las necesidades identificadas de la Unidad.

Referencias bibliográficas

- Cedeño T., A. (2006). Modelo multidimensional. *Ingeniería Industrial*, 27(1), pp.15-18.
- Chinchilla A., R (2011). Mercado de datos: conceptos y metodología de desarrollo. *Tecnología en Marcha*, 24(3), pp. 55-56.
- Cravero, A., Sepúlveda S., Mazón J y Trujillo, J (2013). Un enfoque de ingeniería de requerimientos basada en el alineamiento de almacenes de datos y la estrategia del negocio. *Ingeniare- Revista de Ciencias de Ingeniería*, 21(3), pp. 314-327.
- Davis D. y Jabeen I. (2011). Learning in the GNU/Linux community. *Proceedings of the 2011 conference on Information technology education (SIGITE '11)*. ACM, New York, NY, USA, pp. 21-26. DOI=10.1145/2047594.2047600. Disponible en <http://doi.acm.org/10.1145/2047594.2047600>

- Fernández, M. (2012) Computación en la nube para automatizar unidades de información. *Bibliotecas : Revista de la Escuela de Bibliotecología, Documentación e Información*, 30 (1). Disponible en: <http://www.revistas.una.ac.cr/index.php/bibliotecas/article/view/3894/3738>.
- Fernández- M. (2014) Control estratégico de gestión en unidades de información. *Bibliotecas : Revista de la Escuela de Bibliotecología, Documentación e Información*, 4(1). Disponible en: <http://revistas.ucr.ac.cr/index.php/eciencias/article/view/12864>
- Johnson, M., y Kasangian, S. (2010). A relational model of incomplete data without nulls. *Proceedings of the Sixteenth Symposium on Computing: the Australasian Theory*, 109, pp. 89-94. Disponible en: <http://dl.acm.org/citation.cfm?id=1862317.1862329&coll=DL&dl=ACM&CFID=344168486&CFTOKEN=93029307>
- Neil, C.; De Vincenzi, M. y Pons, C, (2014). Design method for a Historial Data Warehouse, explicit valid time in multidimensional models. *Ingeniare- Revista de Ciencias de Ingeniería*, 22(2), pp. 218-231.
- Timarán-Pereira, R. (2013). Three-Step method to tightly integrate data mining tasks into a relational database system". *Ingeniería y Competitividad*, 15(2), pp. 125-136.
- Wilson M. (1981). A requirements and design aid for relational data bases. *Proceedings of the 5th international conference on Software engineering (ICSE '81)*. IEEE Press, Piscataway, NJ, USA, pp. 283-293.