

Sistema de gestión de conocimiento corporativo mediante técnicas de Datawarehousing

CANCINO-VILLATORO, Karina*†, CASTILLO-SOLIS, Alfredo, CASTILLO-ESTRADA, Christian y DE LEÓN-VEYTIA, Magdalena

Recibido Enero 19, 2017; Aceptado Marzo 29, 2017

Resumen

A nivel directivo, el análisis de datos cobra verdadera importancia pues la competitividad de una organización depende de las decisiones tomadas, esto ha sido factor importante para el desarrollo de sistemas de información que permita la integración de datos históricos, proveniente de múltiples y posiblemente diferentes fuentes de datos con el propósito de hacer análisis y tomar decisiones. Este tipo de tecnología es de gran utilidad en inteligencia empresarial (Business Intelligence) ya que permite realizar el análisis de diferentes variables de negocio de la organización para apoyar el proceso de toma de decisiones. El presente artículo tiene como objetivo presentar los resultados obtenidos de la creación e implementación de un almacén de datos para la generación de un Sistema de Gestión de Conocimiento Corporativo de una entidad financiera en el estado de Chiapas. Se utilizó como tecnología de desarrollo SQL Server, procedimientos almacenados para el proceso ETL (extracción, transformación y carga de datos), el diseño de la interfaz gráfica de usuario fue desarrollado en primefaces, JQuery y CSS3. El Sistema Web usa la plataforma Java EE y para los métodos de recuperación y manipulación de datos, se utilizó el componente Enterprise Java Beans.

Gestión de Conocimiento, análisis de datos, almacén de datos, proceso etl

Abstract

At the management level, data analysis takes real importance for the competitiveness of an organization depends on the decisions taken, this has been important for the development of information systems that allow the integration of historical data, from multiple and possibly different data sources with the purpose of analysis and principal management decisions. This type of technology is very useful in Business Intelligence (BI) as it allows the analysis of different variables of business of the organization to support the decision-making process. The aim of this paper is to present the results obtained from the creation and implementation of a data warehouse in order to develop corporate knowledge management information for a financial corporation in Chiapas, Mexico. It was used as development technology the SQL Server Enterprise Database Management System, Abstract SQL store procedures for the ETL process (extraction, transformation and data load), the graphical user interface design of the web system was developed in Primefaces, JQuery and CSS 3 and the logical programming was developed in Java Enterprise Edition (Java EE), finally for recovery and data manipulation it was used the Enterprise Java Beans component.

Knowledge Management, Data Analysis, Data Warehouse, ETL process

Citación: CANCINO-VILLATORO, Karina, CASTILLO-SOLIS, Alfredo, CASTILLO-ESTRADA, Christian y DE LEÓN-VEYTIA, Magdalena. Sistema de gestión de conocimiento corporativo mediante técnicas de Datawarehousing. Revista de Tecnologías Computacionales 2017, 1-1: 62-69

*Correspondencia al Autor (Correo electrónico: karina.cancino@uptapachula.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

Actualmente los sistemas de información son un factor muy importante para la competitividad de una organización, sin embargo la mayor parte de las veces únicamente procesan transacciones y los directivos requieren de sistemas que proporcionen verdaderas herramientas para la toma de decisiones. Un sistema de procesamiento de transacciones es un tipo de sistema de información diseñado para recolectar, almacenar, modificar y recuperar todo tipo de información que es generada por las transacciones en una organización [1].

A nivel directivo el análisis de datos cobra verdadera importancia pues la competitividad de una organización depende de las decisiones tomadas, esto ha sido factor importante para el desarrollo de la tecnología de Data Warehouse que consiste en la integración de datos históricos, proveniente de múltiples y posiblemente diferentes fuentes de datos con el propósito de hacer análisis y a partir de este tomar decisiones [2]. Este tipo de tecnología es de gran utilidad en inteligencia empresarial (Business Intelligence) ya que permite realizar el análisis de diferentes variables de negocio de la organización para apoyar el proceso de toma de decisiones.

En la entidad financiera en la cual se desarrolló el proyecto, se cuenta con el módulo de finanzas de un Enterprise Resources Planning (ERP) que se encarga de llevar el registro y control de contabilidad, activos fijos, presupuesto, tesorería, crédito y cobranza cuentas por pagar y contabilidad electrónica y a partir los cuales se generan una variedad de informes en formato .xls los cuales son utilizados como fuentes de datos para generar reportes solicitados por la Comisión Nacional Bancaria y de Valores y por el Consejo Directivo de la entidad financiera y que son necesarios para la toma de decisiones dentro de la empresa.

Algunos de estos informes se generan en un rango de 3 a 5 días de forma manual. Debido a esta problemática se propuso el desarrollo de un almacén de datos que integre registros históricos de estados de resultados, balanzas, carteras e inversiones que permita la generación de informes para la toma de decisiones. La presencia de las tecnologías de información esta entidad financiera es indispensable, ya que permite identificar y mejorar la calidad de servicios mediante una correcta toma de decisiones.

El trabajo se organiza de la siguiente manera. La sección 2 presenta el concepto y características de un Data Warehouse para el conocimiento del lector. La sección 3 describe la metodología de desarrollo del Data Warehouse. La sección 4 se muestra la aplicación de acceso a datos. Finalmente en la sección 5 se presentan la conclusión y trabajos futuros de la investigación.

Problemática

Desde que la Entidad Financiera fue fundada en 1990 ha tenido la necesidad de realizar una considerable cantidad de informes periódicos con diferente frecuencia cada uno (mensuales, trimestrales, anuales e incluso algunos otros no tienen una periodicidad definida), algunos son necesarios como requisito por la Comisión Nacional Bancaria y de Valores, otros en cambio sirven únicamente como herramientas auxiliares para la toma de decisiones dentro de la empresa. Cada uno de los reportes es generado manualmente por empleados de la institución, lo que significa que cada cierto período de tiempo el personal tiene que trabajar horas extras a su jornada laboral ordinaria para tener en tiempo y forma la información requerida.

La Entidad Financiera cuenta con un ERP denominado Kepler que ayuda al correcto tratamiento de la información.

La base de datos de este sistema almacena cada una de las transacciones que son llevadas a cabo de manera cotidiana en la empresa pero carece de un módulo que sea capaz de generar automáticamente los reportes requeridos. Permite la exportación en formato csv de gran parte de la información y a partir de ahí los empleados deben generar los informes en forma manual lo cual les toma entre 3 y 8 días dependiendo del área y el reporte que se necesite.

Justificación

El desarrollo de una plataforma tecnológica que genere los informes periódicos de la entidad financiera permitirá que los empleados puedan concentrarse en la atención al socio y ocupaciones propias de los puestos.

Si el desarrollo del Sistema no se lleva a cabo la situación en la empresa no mejorará, lo que significa que los empleados continuarán trabajando tiempo extra para generar estos reportes; esta actividad produce un desgaste en tiempo y económicamente mayor a la institución financiera ya que tiene que cubrir los honorarios de los empleados y a largo plazo es una enorme cantidad de efectivo. En estos puntos radica la importancia del desarrollo del sistema.

Objetivos

Objetivo General

Implementar el Sistema de Gestión de Conocimiento Corporativo para la generación de informes ejecutivos de la Entidad Financiera.

Objetivos específicos

- Realizar un diagnóstico de los indicadores que se integrarán a los informes directivos.
- Diseñar la plataforma de almacén de datos financieros históricos.

- Desarrollar la aplicación de acceso a datos.

Marco Teórico

¿Qué es un Almacén de Datos (Data Warehouse)?

Un almacén de datos es un repositorio de datos que está orientado a resolver la necesidad de toma de decisiones del usuario final dentro de una organización, no solo en el soporte de la gerencia de alto nivel que debe tomar decisiones de objetivos y planificar estrategias; sino también en niveles más técnicos de toma de decisión. Los principales objetivos que persigue son: [3] hacer que la información de una organización sea de fácil acceso, presentar la información de la organización consistente, ser adaptable y resistente a los cambios, ser un bastión seguro que proteja la información activa, servir como fundamento para la toma de decisiones y ser aceptado por la comunidad del negocio si se desea estimar como exitoso.

Características

Las características principales con las que cuenta un almacén de datos son: orientado al tema, integrado, de tiempo variante y no volátil. Es orientado a temas ya que clasifica la información según los temas de interés de la organización y que a su vez estarán interrelacionados con los procesos que se llevan a cabo dentro de la organización.

Otra característica muy importante es que la información debe estar integrada, es decir la información almacenada debe coincidir con la información de entrada hacia el almacén por lo que debe haber una fase de transformación y estandarización de la misma en caso necesario.

En un almacén de datos la información que más interesa es la histórica ya que se utiliza para la identificación y evaluación de tendencias, así mismo los datos deben ser estables ya que esta pasividad en el movimiento de datos permite una perspectiva más grande del análisis y toma de decisiones.

En el ambiente operacional, la actualización (insertar, borrar y modificar), se hace regularmente sobre una base de registro por registro. En el almacén de datos hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos en el depósito, como una parte normal de procesamiento (ver Figura 1).

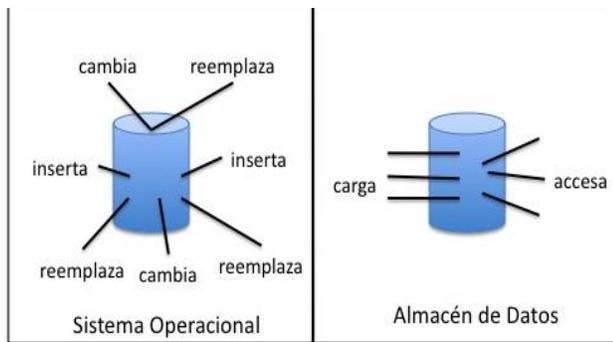


Figura 1 Operaciones sobre los datos en un Data Warehouse y un sistema operacional

Procesos ETL (Extraer, Transformar, Cargar)

Este tipo de procesos permite a las organizaciones mover datos desde diversas fuentes, transformarlos y cargarlos en otra base de datos, data mart o data warehouse para su análisis y apoyo en un proceso de negocio.

Sus principales objetivos son: controlar la extracción de datos y su automatización, acceso a diferentes tecnologías para hacer un uso adecuado de hardware, software, datos y recursos humanos existentes, utilización de arquitectura de metadatos para la definición de objetos de negocio, acceso a fuentes de datos heterogéneas, manejo de excepciones, gestión de la planificación de todos los procesos necesarios para la carga del almacén de datos, interfaz independiente de hardware y soporte en la explotación del Data Warehouse [4].

En la figura 2 se muestra el esquema generalizado de la herramienta ETL.

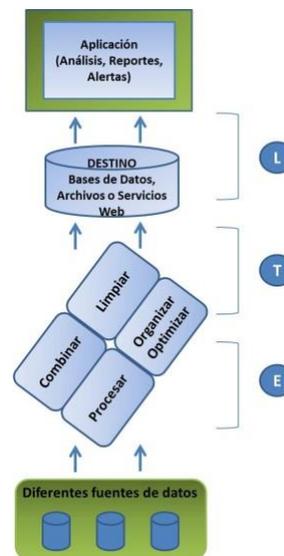


Figura 2 Esquema ETL

Hechos, Dimensiones y Granularidad

Se denomina hechos a los indicadores principales de negocio es decir las medidas numéricas que se incluyen en el sistema Business Intelligence (BI). Una tabla de hecho es la tabla central de un esquema dimensional que puede ser de estrella o copo de nieve y contiene los valores de las medidas de negocio.

Cada medida se toma mediante la intersección de las dimensiones que la definen las cuales estarán reflejadas en sus correspondientes tablas de dimensiones que rodearan la tabla de hechos y estarán relacionadas con ella [4]. En la figura 3 se muestra el ejemplo de la tabla de hechos ingresos del data warehouse estadístico con sus respectivas tablas de dimensiones.

CarteraMensual
+fecha: date
+fechaCarga: date
+indice: varchar(6)
+numero: int
+numeroCredito: varchar(30)
+tipoCredito: varchar(30)
+denominacion: varchar(100)
+apellidoPaterno: varchar(50)
+apellidoMaterno: varchar(50)
+nombre: varchar(50)
+clasContable: varchar(30)
+montoTotal: float
+fechaInicio: date
+fechaFin: date
+tasaInteres: float
+interesVigente: foat
+interesVencido: float

Figura 3 Tabla de hechos Cartera Mensual y sus tabla de dimensiones

Otra característica importante que define a una tabla de hechos es el nivel de granularidad de los datos que en ella se almacenan, es decir el nivel más atómico por el cual se definen los datos. La granularidad afecta a la cardinalidad, tanto de las dimensiones como de las tablas de hechos, a mayor granularidad (grano más fino) mayor será el número de registros final de la tabla de hechos [1], ver Fig. 4.

Metodología de Desarrollo de Software

El desarrollo de esta plataforma está dividida en dos etapas, en la primera se contempla la realización de 12 informes ejecutivos cuyas fuentes de datos provienen del sistema transaccional.

La estrategia de implementación que se utilizó fue bottom-up (abajo hacia arriba) ya que la tecnología con la que se cuenta actualmente (Sistema Kepler – ERP) se encuentra en sus primeras etapas de implementación. La metodología de desarrollo fue de análisis y diseño estructurado en cascada el cual consiste en reunir los requerimientos, analizarlos, realizar un diseño preliminar en base a los resultados del análisis [3]. El diseño inicia como un nivel abstracto y se divide sucesivamente en niveles más concretos. En la figura 4 se muestra algunas de las tablas de hechos construidas.

balanceActivo	CarteraMensual
+fecha: date	+fecha: date
+fechaCarga: date	+fechaCarga: date
+indice: varchar(2)	+indice: varchar(6)
+concepto: varchar(100)	+numero: int
+saldoMesAnterior: float	+numeroCredito: varchar(30)
+porcentajeMesAnterior: float	+tipoCredito: varchar(30)
+saldoMesActual: float	+denominacion: varchar(100)
+porcentajeMesActual: float	+apellidoPaterno: varchar(50)
+diferencias: float	+apellidoMaterno: varchar(50)
	+nombre: varchar(50)
	+clasContable: varchar(30)
	+montoTotal: float
	+fechaInicio: date
	+fechaFin: date
	+tasaInteres: float
	+interesVigente: foat
	+interesVencido: float

BalancePasivo
+fecha: date
+fechacarga: date
+indicePasivo: varchar(2)
+conceptoPasivo: varchar(100)
+saldoMesAnteriorPasivo: float
+porcentajeMesAnteriorPasivo: floa
+SaldoMesActualPasivo: float
+porcentajeMesActualPasivo: float
+diferenciasPasivo: float

Figura 4 Tablas de Hechos del Data Warehouse

La arquitectura que se utilizó es de solo almacén de datos la cual consiste en construir un Data Warehouse centralizado aplicando las operaciones de depuración, integración, resumen y adición de datos. En este tipo de arquitectura este único Data Warehouse alimenta de información a los usuarios para el apoyo a la toma de decisiones [3].

Como resultado del proceso ETL se construyeron 6 tablas de hechos: balanceActivo, CarteraMensual, BalancePasivo, carteraTrimestralVigente, carteraVencida, desgloseDeInversiones; cada una de ellas se integró a partir de diversas fuentes de datos provenientes del sistema Kepler.

Así también el nivel de granularidad es de grano fino. En la figura 5 se muestra la arquitectura lógica sobre la cual se desarrolló el Data Warehouse.

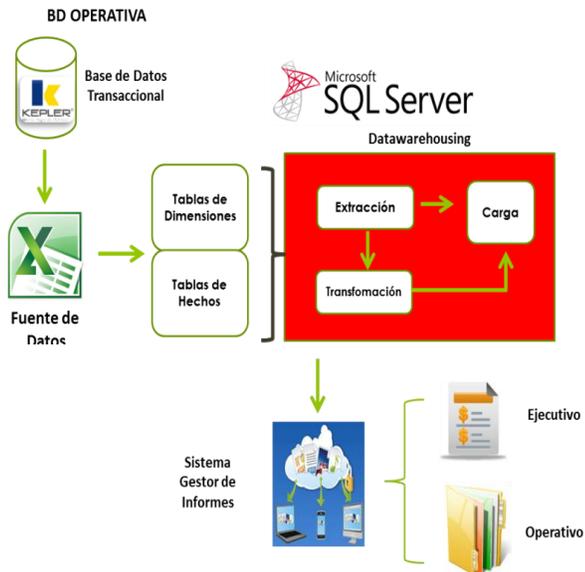


Figura 5 Arquitectura y Esquema Lógico del Data Warehouse

Se tomó como fuente de datos los archivos exportables en formato de Excel con extensión .xls del sistema transaccional Kepler – ERP.

A partir de estas fuentes de datos se realizó el proceso de extracción y transformación de datos mediante el desarrollo de procedimientos almacenados utilizando el lenguaje procedural Abstract de SQL Server Enterprise.

De estas etapas depende que los datos cargados sean consistentes, seguros y resistentes. Posteriormente se realiza la carga de datos mediante la implementación de una aplicación que permite adjuntar los archivos necesarios por cada informe que se requiere [6]. Se programaron 12 informes operativos y 6 procedimientos almacenados:

Balance Activo, Cartera Mensual, Balance Pasivo, cartera Trimestral Vigente, cartera Vencida, desglose de Inversiones. Cada procedimiento que almacena datos a las tablas de hechos carga un promedio de 600 a 1000 registros cada mes.

Para el acceso a los datos se construyó el Sistema de Gestión de Conocimiento Corporativo (SICORP) cuyo diseño de la interfaz gráfica de usuario fue desarrollado en primefaces, JQuery y CSS 3, las cuales son tecnologías de diseño web ejecutadas de lado del cliente, es decir, en el navegador web, mejorando el desempeño de la aplicación; al no requerir de peticiones constantes realizadas al servidor de aplicaciones.

Para el desarrollo del Sistema Web se utilizó la plataforma Java Edición Empresarial (Java EE); por tal motivo, es importante señalar la implementación de la arquitectura de 4 capas, quedando de la siguiente manera: La Capa Cliente o de Presentación, interactúa con el servidor enviando solicitudes para que sean procesadas; así mismo, hace referencia al diseño del entorno visual para lo cual se utilizaron las tecnologías Primefaces, JQuery y CSS 3. Por otra parte, para la Capa Web se implementaron los componentes JSP y Servlets, mismos que son ejecutados en el Servidor de aplicaciones GlassFish V3, los cuales se encargan de proveer al cliente el acceso a los datos y procesar las peticiones.

En relación a los métodos de recuperación y manipulación de datos, se utilizó el componente Enterprise Java Beans en la Capa de negocio. Finalmente, la Capa de Datos, se encarga de procesar las solicitudes de datos enviadas por la capa de negocio; además, es responsable de almacenar toda la información a través de tablas, índices, y otros objetos de datos.

Para el presente proyecto, se implementó el Sistema Gestor de Base de Datos SQL Server Empresarial. (ver Figura 6).



Figura 6 Sistema de Gestión de Conocimiento Corporativo (SICORP)

Conclusiones

En estos momentos el proyecto se encuentra en etapa de implementación y se está llevando a cabo un proceso de validación de informes por parte de la entidad financiera.

Sin embargo de acuerdo a los avances presentados se puede asegurar que la generación de los 12 informes operativos y directivos a partir de la plataforma del almacén de datos agiliza los procesos de recopilación, análisis y generación de información directiva, permite la entrega de información a tiempo a las áreas directivas y facilita la toma de decisiones ya que la consulta de los informes está disponible en tiempo real mediante la aplicación de acceso a datos.

La segunda etapa de este desarrollo tecnológico abarca la inclusión de 38 informes dentro de la aplicación de acceso a datos lo cual permitirá que mediante la carga de datos históricos se cuente con la suficiente información para en un futuro aplicar algoritmos de Minería de Datos (Datamining) para la extracción de conocimientos estadísticos o predictivos.

Agradecimiento

Los autores desean agradecer al Consejo Directivo de la Entidad Financiera por permitir el acceso a la información de las áreas de Contabilidad, Inversiones y Ejecutivos sobre la cual se basó el desarrollo de esta investigación.

Referencias

Power, D. J. (2002). *Decision Support Systems: concepts and resources for managers*. Westport, Conn., Quorum Books.

Farooq F., Mansoor S. (2010), *Real-Time Data Warehousing for Business Intelligence*, Diciembre 2010, Islamabad, Pakistán, Página 1, Consultado en ACM Digital Library.

Kimball Ralph y Ross Margy (2012). *The Data Warehouse Toolkit: the complete guide to dimensional modeling*. Segunda Edición. 2010. Editorial Wiley. United States of America.

Kimball Ralph, Caserta Joe, The Data Warehouse ETL Toolkit: practical techniques for extracting, cleaning, conforming and delivering data. (2010) Segunda Edición. Editorial Wiley. United States of America.

Harjinder G., Rao P., Data Warehousing: la integración de información para la mejor toma de decisiones. (2013). Prentice Hall.

Vieira R., Professional SQL Server Programing. (2010) Wiley Publishing, Indianapolis Indiana. Páginas 281-327, 699-735