

Tipo de artículo:

Temática:

Recibido: dd/mm/aa | Aceptado: dd/mm/aa | Publicado: dd/mm/aa

Diseño del repositorio de datos para la sala situacional de un sistema de gestión penitenciaria

Design of a data mart for a situational room of a prison management system

Malena García Izquierdo^{1*}, Manuel Macías Martínez²

¹ Isec. Centro de Informatización para la Seguridad Ciudadana. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

² Facultad 5. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

* Autor para la correspondencia: mgizquierdo@uci.cu

Resumen

La implantación del Sistema de Gestión Penitenciaria, desarrollado en la Universidad de las Ciencias Informáticas, y el correspondiente crecimiento de los datos generados trajeron como consecuencia que se viera afectado el rendimiento de la aplicación ya que no cuenta con una solución relacionada con la inteligencia del negocio. El propósito fundamental de este trabajo es el diseño de un Data Mart sobre la sala situacional del Sistema de Gestión Penitenciaria desde el punto de vista de Clasificación y Atención Integral para así dar soporte al proceso de toma de decisiones por parte del sistema penitenciario. El Data Mart está basado en la metodología *Data Warehouse Engineering Process*. Procesos de Ingeniería de Almacenes de Datos, utilizando el enfoque propuesto por Ralph Kimball así como la arquitectura de dos capas. Como gestor de base de datos se hace uso de Oracle 10g, proponiendo además, la Suite de Inteligencia de Negocio Pentaho para el proceso de extracción, transformación y carga de los datos. El equipo de desarrollo tiene entre sus metas la creación de un Data Mart para cada módulo del sistema, de manera que se puedan integrar finalmente para la construcción de un Data Warehouse central que ayude al proceso de gestión de información y una acertada toma de decisiones.

Palabras clave: almacén de datos, inteligencia del negocio, repositorio de datos.

Abstract

The implementation of the Prison Management System, developed at the University of Informatics Sciences and the corresponding growth of data generated as a result brought that was affected the performance of the application because it does not have a solution related to business intelligence. The primary purpose of this paper is the design of a Data Mart on the situation room of Prison Management System from the point of view of classification and Comprehensive Care for and support the decision making process by the prison system. The Data Mart is based on the methodology Data Warehouse Engineering Process, using the approach proposed by Ralph Kimball and two-layer architecture. As database manager makes use of Oracle 10g, it further, Suite Pentaho Business Intelligence for the extraction, transformation and loading of data. The development team has among its goals the creation of a Data

Mart for every module in the system so that eventually they can be integrated to build a Data Warehouse to assist the central information management process and accurate decision making.

Keywords: *data mart, data warehouse, business intelligence.*

Introducción

El concepto de BI, Inteligencia del Negocio, se remonta a octubre de 1958, cuando Hans Peter Luhn de IBM, escribió un artículo titulado “*A Business Intelligence System*” en el cual describía las características que debía tener un sistema de este tipo. “El objetivo básico de la *Business Intelligence* es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones” (Cano, 2007).

El progreso irreversible del mundo y las nuevas tecnologías de la informática y las comunicaciones han acentuado el papel que juegan actualmente la información y el conocimiento en el desarrollo de toda empresa, más pronunciadamente con la aplicación de la inteligencia del negocio en la Informática.

La información generada diariamente crece de manera vertiginosa y su correcta gestión es fundamental para la toma de decisiones, la entrada a nuevos mercados, análisis de las necesidades de los clientes, determinación de la rentabilidad de un producto, formular estrategias del negocio, entre otros factores imprescindibles para toda empresa. La eficiente gestión de la información de manera que ayude al proceso de toma de decisiones, es clave para la supervivencia en un mercado dinámico y competitivo como el actual. Aprender a competir utilizando esta información es primordial para el desarrollo de las empresas.

Para que la información se pueda gestionar de manera eficiente, debe utilizarse un almacén de datos, (del inglés, *Data Warehouse*, DW). Los DW proporcionan una herramienta para la toma de decisiones en cualquier área del negocio, brindando una visión global de los datos de una organización.

Podemos entender un Data Mart como un subconjunto de los datos de un DW que tiene como objetivo responder a un análisis determinado y con una población de usuarios específica.

Los DW son actualmente un punto de atención de las organizaciones y proveen un ambiente para que las empresas hagan un mejor uso de la información que es administrada por diversas aplicaciones operacionales.

“El SIGEP, concibe una solución de software cuyo objetivo general es: desarrollar e implantar un sistema informático que soporte las decisiones estratégicas del Ministerio del Interior y Justicia y de la Dirección General de Custodia y Rehabilitación del Recluso, que contribuya a garantizar el respeto a los derechos de los internos, su actividad de rehabilitación y reinserción en la sociedad” (Arias, 2007).

“Sala situacional es un subsistema que solo contaba con una definición general de su objetivo y no se encontraba estructurado por módulos dado que no se había determinado en el estudio preliminar sus funciones específicas. Durante la modelación del negocio se identificaron cuatro tipos de salidas de información para el apoyo a la gestión de la sala situacional que apuntaban a soportar la toma de decisiones operativas, tácticas y estratégicas. Estas salidas se agrupan en: información operativa, avisos, alertas e información histórica; sin embargo, no todas estas funciones son propias de la sala situacional ya que el análisis históricos de datos y el monitoreo de indicadores de riesgos (avisos y alertas) son propios del departamento de Gestión de Riesgos. La Sala Situacional fundamentalmente se dedica al monitoreo de la situación operativa y responder a situaciones de emergencia. El departamento de Gestión de Riesgos por otro lado se dedica al análisis de la información” (Arias, 2007).

La implantación del sistema y el correspondiente crecimiento de los datos generados trajeron como consecuencia que se viera afectado el rendimiento de la aplicación, a mayor escala, al incluir reportes que precisan la búsqueda en varias

tablas o varios módulos al mismo tiempo. Con el objetivo de darle una solución a este problema se crearon vistas materializadas¹ para tratar de hacer más eficiente el rendimiento de la aplicación de manera que las consultas a la base de datos se realizaran de forma más rápida. Esta variante no fue fructífera ya que las vistas se actualizaban sólo una vez al día y al sistema penitenciario le interesaba que la información fuera operativa, o sea, que a la hora de ser consultada se tuvieran los datos en tiempo real, y con las vistas no se lograba este objetivo al estar éstas en muchas ocasiones desfasadas con respecto a los datos reales. Por otra parte en el momento de actualizar las vistas disminuía el rendimiento de la aplicación al demorarse la actualización de los datos.

De igual manera al establecer los requisitos funcionales se pensaron una cantidad de reportes predefinidos lo que trajo consigo una traba para la correcta toma de decisiones por parte de la administración ya que siempre que surja la necesidad de hacer un nuevo reporte los desarrolladores son los únicos capaces de darle solución a este problema.

Se hace necesaria la creación de un Data Mart para cada módulo del sistema, de manera que se puedan integrar finalmente para la construcción de un Data Warehouse central que ayude al proceso de gestión de información y una acertada toma de decisiones.

Materiales y Métodos o Metodología Computacional

El subsistema *Observación, Clasificación y Tratamiento* permite controlar las actividades de reinserción de los privados de libertad. Dentro de este subsistema se encuentran definidos los módulos que atañen al diseño del Data Mart en cuestión. Para ello se tendrán en cuenta los siguientes conceptos:

Clasificación: Finalizado el período de observación, el individuo es diagnosticado y un equipo multidisciplinario lo clasifica en alguno de los regímenes de seguridad penal (máxima, media o mínima), para ofrecerle un tratamiento individualizado. El módulo permitirá registrar la clasificación dada a cada individuo, la fecha en que se le otorgó así como mantener un histórico de las clasificaciones que ha tenido y los motivos por los cuales ha cambiado de una a otra.

Educación: El módulo permite registrar los distintos tipos de actividades educativas (formales y no formales) que se desarrollan en los establecimientos penitenciarios. Las actividades formales se refieren a las que elevan el nivel de escolaridad de los privados de libertad (primaria, secundaria, bachillerato, universidad). Las no formales se dedican a la formación de oficios, manualidades, etc. Además el módulo gestiona la matrícula de las diferentes actividades educativas, permite mantener un registro actualizado de la asistencia de los internos a las actividades y registrar las evaluaciones finales de cada curso.

En el expediente penitenciario de cada privado de libertad queda reflejado un historial de cursos matriculados, la evaluación final obtenido y el total de horas dedicadas al estudio.

Trabajo: Este módulo permite controlar las Unidades Productivas que se conforman en los Establecimientos Penitenciarios, la incorporación de los internos a ellas, así como el registro de la asistencia diaria a las actividades laborales. En el expediente penitenciario de cada privado de libertad aparece un resumen del total de horas laboradas, así como las unidades productivas en las que ha estado incorporado.

Deporte y cultura: Con este módulo se controla la incorporación y participación de los privados de libertad en las manifestaciones culturales o disciplinas deportivas que se practican en el penal, bajo la orientación y supervisión de las coordinaciones de deporte y cultura.

¹ En un sistema de gestión de base de datos que siga el modelo relacional, una vista es una tabla virtual, que representa el resultado de una consulta. Siempre que se consulta o se actualiza una vista normal, el SGBD convierte estas operaciones en consultas o actualizaciones de las tablas usadas para definir la vista.

Para diseñar un Data Mart los dos enfoques más utilizados son el modelo relacional propuesto por Inmon y el modelo dimensional defendido por Kimball.

Se decide utilizar el enfoque de Ralph Kimball para el diseño del Data Mart, pues es menos costoso, más funcional, se pueden priorizar las áreas críticas; su estructura de datos ofrece mayor facilidad al usuario para la exploración y búsqueda de información. Igualmente se utilizó el modelo dimensional como modelo a seguir para diseñar el almacén de datos puesto que propone una forma sencilla de representar los datos y mejora el tiempo de consulta en la base de datos.

Se considera que la mejor y más completa arquitectura de datos es la de dos capas teniendo en cuenta que el almacén de datos será poblado desde un único origen de datos, con poco volumen de transformaciones. La organización física será ROLAP ya que con esta pueden hacerse un gran número de dimensiones y soporta elevados volúmenes de datos elementales, además de ser el motor OLAP de Pentaho.

ROLAP (Relational OnLine Analytic Processing): Se implementa sobre tecnología relacional, dispone de algunas facilidades para mejorar el rendimiento. Aunque el almacén de datos se organiza como una base de datos multidimensional es soportado por un SGBD Relacional, las dimensiones y tablas de hechos serán traducidas a tablas. “Los **sistemas OLAP** son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejo etc. Este sistema es típico de los Data Mart” (Sinnexus, 2010).

Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL).

Como herramienta de desarrollo para el Data Mart se seleccionó la *Suite de Business Intelligence Pentaho* en su versión 3.5 por ser de código abierto, ser actualmente la líder de código abierto en temas de depósitos de datos y ser una plataforma orientada a la solución y centrada en procesos.

Teniendo en cuenta las características de los gestores de bases de datos existentes se utilizó Oracle 10g pues es capaz de manejar eficientemente todo el volumen de información del Data Mart y proporciona todo el soporte necesario para la construcción y mantenimiento del mismo.

Resultados y Discusión

Se diseñó un Data Mart para la Clasificación y la Atención Integral de la sala situacional del SIGEP de manera que ayuda a la toma de decisiones a la dirección del sistema de justicia, enfocado a organizar y sistematizar la atención que recibirán los privados de libertad durante su permanencia en el sistema penitenciario. El mismo contiene información referente a los individuos, los centros penitenciarios, las unidades productivas que se conforman en el establecimiento penitenciario, las actividades educativas desarrolladas en el establecimiento, las disciplinas deportivas que se practican en el penal, la incorporación y participación de los privados de libertad en las manifestaciones culturales que se practican en el centro así como la clasificación de los penados en alguno de los regímenes de seguridad penal (máxima, media o mínima), para ofrecerle un atención individualizada.

El Data Mart sobre la Clasificación y la Atención Integral de la Sala Situacional del SIGEP, además de los reportes predefinidos, ofrece 5 variantes diferentes para la exploración de los datos. Los usuarios finales podrán explorar basándose en la clasificación, en este se muestran los individuos, sus expedientes y la clasificación de los individuos en cada uno de los centros penitenciarios modelados en el tiempo. De igual manera se puede explorar teniendo en

cuenta la actividad cultural; esta exploración incluye individuos, sus expedientes, el nivel y el tipo de manifestación cultural practicada en el centro por el individuo a lo largo del tiempo. También se puede enfocar la exploración centrándose en la actividad productiva, incluyendo la información de los individuos, sus expedientes, el tipo de actividad productiva que realiza en el centro a lo largo del tiempo de estancia en el penal. Otra manera de exploración es por la actividad deportiva, incluyendo individuos, sus expedientes, el nombre y tipo de disciplina deportiva en las que participa el individuo modelado en el tiempo. Por último los usuarios pueden explorar centrándose en la actividad educativa, la cual muestra a los individuos, sus expedientes y la clasificación de las actividades deportivas desarrolladas en el centro a través del tiempo.

La metodología usada para el diseño fue el *Data Warehouse Engineering Process* que es un método para llevar a cabo el diseño de un almacén de datos incluyendo las fuentes de datos operacionales, los procesos de Extracción Transformación y Carga (del inglés, *Extraction, Transformation and Load*, ETL) y el propio esquema del almacén de datos.

Para llevar a cabo la actividad de diseño existen dos estrategias de procesamiento de información: *top-down* y *bottom-up*. La primera de estas estrategias es la **top-down**, la cual define al almacén de datos según los requisitos de los usuarios finales y la otra es la **bottom-up**, en la cual se define el almacén de datos en base a los datos disponibles en las fuentes de datos. La estrategia que se siguió fue la *top down*, ya que se tuvieron en cuenta los requisitos funcionales obtenidos del SIGEP.

Para el mapeo de datos entre las diferentes fuentes de datos (SCS) y el almacén de datos (DWCS), y el mapeo de este último con el esquema conceptual del cliente, se definen los procesos de extracción, transformación y carga o ETL.

Como parte del flujo de trabajo Implementación se lleva a cabo la construcción de las estructuras físicas del Data Warehouse, definiendo el tipo de almacenamiento empleado así como creando los procesos ETL necesarios para poblar el Data Warehouse.

Se monta el Gestor de Base de Datos Oracle 10g de acuerdo al diseño propuesto, por lo que se crean las tablas con sus respectivos atributos y relaciones de manera que se define un sistema de almacenamiento adecuado para los procesos que se sucederán. Dicha base de datos multidimensional contendrá la información referente a la Clasificación y Atención Integral de la Sala Situacional del SIGEP, la cual sería actualizada de forma diaria.

Conclusiones

Un Data Warehouse da lugar a una serie importante de beneficios para la organización. Su utilización permite que la información de gestión sea: accesible, correcta, uniforme y actualizada.

El modelo dimensional brinda una forma muy sencilla de representación de los datos y mejora así el tiempo de consulta a la base de datos.

Mediante el diseño y posterior implementación de este Data Mart se le dará soporte al proceso de toma de decisiones por parte del sistema penitenciario, al brindarle una herramienta de Inteligencia de Negocio eficiente y dinámica.

El que se trate el desarrollo del Data Mart como un proceso iterativo e incremental permitirá que en un futuro se puedan incluir nuevas necesidades, procesos de negocios, así como cambios que se produzcan en el sistema penitenciario.

Referencias

CANO, JOSEP LLUÍS. *Business Intelligence: Competir con información*. s.l.: Fundación Cultural Banesto, 2007.

ARIAS, ARTURO C. *Visión del Sistema de Gestión Penitenciaria. Actualización*. La Habana: Universidad de las Ciencias Informáticas, 2007.

Sinnexus. Sinnexus. Business Intelligence + Informática Estratégica. [en línea] [Consultado el: 21 de octubre de 2010]. Disponible en: [http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx].