

Propuesta de repositorio de datos para el béisbol cubano

Autor: Yoamel Acosta Morales - 24-02-2015

<https://vinculando.org/beta/repositorio-datos-para-beisbol-revolucionario-cubano.html>



Resumen

La larga y fructífera trayectoria del béisbol cubano en eventos tanto nacionales como internacionales en todas las categorías ha generado un importante cúmulo de datos. Almacenada en las más diversas fuentes, la información referente al deporte nacional cubano pudiera ser de mucha utilidad, no solo para su consulta de altísimo valor histórico, sino para la realización de análisis de tendencias en determinados momentos ante la influencia de los más diversos fenómenos económicos, sociales y deportivos.

El propósito fundamental de este trabajo es el diseño de un Data Mart sobre el beisbol revolucionario cubano, que permita no solo la consulta de datos históricos normalizados, sino a la toma de decisiones a partir de tendencias observadas en el rendimiento nacional e internacional de las diferentes categorías.

El Data Mart estaría 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 propone el PostgreSQL, proponiendo la Suite de Inteligencia de Negocio Pentaho para el proceso de extracción, transformación y carga de los datos.

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

Introducción

Con su génesis a fines de los años 50 del siglo pasado el concepto de BI, Inteligencia del Negocio surge del trabajo de uno de los especialistas de la compañía norteamericana IBM, quien 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” (1).

En cada esfera de la vida se generan hoy en día grandes volúmenes de información a una velocidad vertiginosa y su correcta gestión es fundamental para la toma de decisiones, la entrada a nuevos mercados y formular estrategias, entre otros factores imprescindibles para toda esfera de la sociedad actual.

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 cualquier esfera que aspire a la competitividad en la nueva era de la información.

Los repositorios de datos (DWH- Datawarehouse en inglés) proporcionan una herramienta para la toma de decisiones en cualquier área del negocio, brindando una visión global de los datos manejados.

Podemos entender un Mercado de datos (DM- Data Mart en inglés) como un subconjunto de los datos de un DWH que tiene como objetivo responder a un análisis determinado y con una población de usuarios específica.

Existen hoy en día numerosos sistemas que manejan información relacionada con el béisbol revolucionario cubano, se pueden mencionar algunos como los servicios brindados por el sitio de la COCO, el sitio del béisbol cubano, el sitio inder.uci.cu, otros de carácter local como el sitio beisbol.uci.cu, los sitios de emisoras y periódicos provinciales y algunos de carácter foráneo

como es el caso de beisbolencuba.com.

La dispersión de la información, su alto volumen y tasa de crecimiento acelerado han traído como consecuencia que se vea afectado el rendimiento de las aplicaciones. Algunas presentan problemas al realizar búsquedas complejas en cuanto al tiempo de respuesta y en su mayoría no se pueden realizar muchas consultas de cierta complejidad que involucren varias tablas al mismo tiempo.

Estos sistemas se pensaron con una cantidad de reportes predefinidos lo que trajo consigo una traba para la correcta toma de decisiones por parte de la dirección del deporte en las diferentes instancias, ya que siempre que surja la necesidad de hacer un nuevo reporte los que desarrollaron el sistema son los únicos capaces de darle solución a este problema.

Se propone la creación de un Data Mart para cada tipo de evento asociado al béisbol cubano, 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.

Desarrollo

Se propone el desarrollo de los subsistemas Series Nacionales y Eventos Internacionales.

Series Nacionales: El subsistema permitiría registrar los resultados de juegos, jugadores, autoridades (árbitros, directores, cuerpos de dirección y autoridades administrativas) que se desarrollan en las diferentes series nacionales. El subsistema se encargará de controlar el rendimiento de todos los jugadores y equipos, a partir de las estadísticas generadas en cada desafío.

Eventos Internacionales: Este subsistema permitiría controlar el tipo de evento, así como los juegos efectuados, incluyendo contrarios y resultados del equipo e individuales de los jugadores en cada partido. Además de incluir las autoridades (árbitros, directores, cuerpos de dirección y autoridades administrativas) que se ocuparon de la dirección del partido en cuestión.

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”. (2)

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 propone la Suite de Business Intelligence Pentaho 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 propone utilizar PostgreSQL, 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. Además de ser el más potente de los disponibles con código abierto.

Conclusiones

Un Data Warehouse da lugar a una serie importante de beneficios para la dirección del béisbol en Cuba, que pueden favorecer la mejora de los resultados internacionales, actualmente en franca depresión. Su utilización permite que la información de gestión sea: accesible, correcta, uniforme y actualizada. Las decisiones arbitrarias por falta de información clara quedarían en la historia.

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 de la dirección del béisbol cubano en todas las categorías y niveles, al brindarle una herramienta de Inteligencia de Negocio eficiente y dinámica.

Autores: Ing. Annia Verdecia Boza¹, Ing. Yoamel Acosta Morales²

1. Centro de Informatización de Entidades. Departamento de Aplicaciones de Gestión Empresarial. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370
2. Departamento de Programación y Sistemas Digitales, Facultad 6, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

Referencias:

- (1) CANO, JOSEP LLUÍS. Business Intelligence: Competir con información. s.l.: Fundación Cultural Banesto, 2007.
- (2) 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].