

DESARROLLO E IMPLANTACIÓN DE UN SISTEMA ACADEMICO PARA EL ICM

Sergio Bauz Olvera¹, Washington Jama²

¹ Ingeniero en Estadística e Informática 2003

² Director de Tesis de Grado, Ing. Washington Jama.

RESUMEN

El presente trabajo es un herramienta para toma de decisiones académicas, la investigación se centrará en desarrollar un software que haga uso de conceptos de data warehousing y procesamiento analítico en línea (OLAP) orientados al Internet, para ello se diseñara un cubo de datos (Base de Datos Multidimensional) con la información histórica que se tiene de los estudiantes registrados en la carrera de Ingeniería en Estadística e Informática, permitiendo a los coordinadores de la carrera visualizar reportes académicos (Materia de mayor grado de complejidad por nivel ó termino, estimación de estudiantes que ingresarían a proceso de graduación en el próximo termino, entre otras).

La tecnología web que se investigó y aplicó al proyecto permite que los usuarios se conecten al sistema desde el Internet.

INTRODUCCIÓN

En el instituto de ciencias matemáticas los coordinadores académicos tienen la labor de planificar los términos académicos, para ello es necesario contar con toda la información académica de los estudiantes y de lo planificado en términos anteriores, debido a esto surgió la necesidad de contar con una herramienta capaz de mostrar esa información, ordenada, y de acceso rápido.

El objetivo es contar con una solución computacional que recoja conceptos de Data Warehouse y la tecnología OLAP, principales tecnologías para la construcción de sistemas de información robustos.

CONTENIDO

CAPITULO I

1. ENFOQUE DE LOS SISTEMAS DE INFORMACIÓN COMO HERRAMIENTA PARA OBTENER INDICES DE EFICIENCIA DENTRO DE UN ENTORNO ACADEMICO

1.1. QUE ES UN SISTEMA DE INFORMACION

“Un conjunto formal de procesos que, operando sobre una colección de datos estructurada según las necesidades de la empresa, recopilan, elaboran y distribuyen la información (o parte de ella) necesaria para las operaciones de dicha empresa y para las actividades de dirección y control correspondientes (decisiones) para desempeñar su actividad de acuerdo a su estrategia de negocio”. (Andreu et al., 1999).

1.2. CLASIFICACIÓN DE LOS SISTEMAS DE INFORMACIÓN.

Los sistemas de información se han dividido de acuerdo al siguiente esquema:



Figura 1. Clasificación de los Sistemas de Información.

1.2.1. SISTEMAS DE SOPORTE DE DECISIONES.

Estos sistemas están relacionados con el análisis de los datos y la toma de decisiones, frecuentemente, decisiones importantes sobre cómo operará la empresa, ahora y en el futuro. Mientras las necesidades de los datos operacionales se enfocan normalmente hacia una sola área, los datos para el soporte de decisiones, con frecuencia, toma un número de áreas diferentes y necesita cantidades grandes de datos operacionales relacionadas. Estos sistemas están basados en la tecnología data warehousing.

1.3. QUE ES UN DATA WAREHOUSE.

Tras las dificultades de los sistemas tradicionales en satisfacer los requerimientos de información, surge el concepto de Data Warehouse, como solución a las necesidades

de información globales de la empresa. Este término acuñado por Bill Inmon, se traduce literalmente como Almacén de Datos. No obstante si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

La ventaja principal de este tipo de sistemas se basa en su concepto fundamental, la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales.

1.3.1. PROCESOS DE UN DATA WAREHOUSE.

A continuación se describen dichos procesos clave en la gestión de un Data Warehouse:

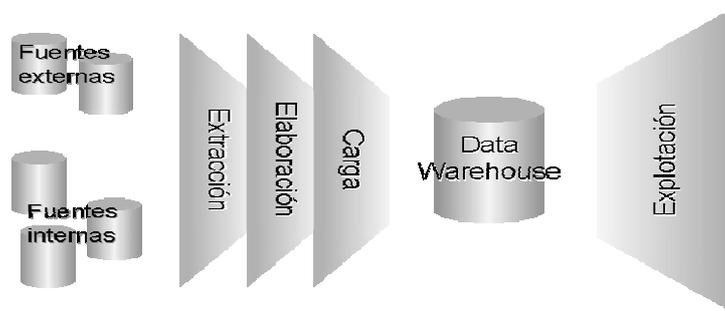


Figura 2. Gestión de un Data Warehouse

1.3.2. DATA WAREHOUSE vs. SISTEMA TRADICIONAL.

SISTEMA TRADICIONAL	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importante del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimensional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa relacionada con el negocio

1.4. OLAP (PROCESO ANALÍTICO EN LÍNEA).

OLAP pertenece al dominio de los almacenes de datos y se encuentra dentro del ámbito, mas amplio, de los sistemas de toma de decisiones (DSS, Decisión Support Systems) y de los sistemas de información ejecutiva (ELS. Executive Information). En resumen, es un conjunto de métodos para consultar inmensas cantidades de datos y encontrar patrones de interés y tendencias, para que aquellos que toman las decisiones empresariales puedan obtener ventajas competitivas u optimizar de alguna forma los procesos de negocio.

CAPITULO II

2. ANÁLISIS Y DETERMINACIÓN DE LOS REQUERIMIENTOS DE INFORMACIÓN

2.1. REQUERIMIENTOS DE INFORMACIÓN ACADÉMICA

Los requerimientos de información previa a la toma de decisiones, para lograr un mayor entendimiento del problema:

- Fase de Pre-Registro.
- Fase de coordinación Académica

2.1.1. FASE DE PRE-REGISTRO

Muchos de los directivos del ICM coinciden que esta es una de las fases de mayor utilidad de la planificación debido a que no hay en la actualidad un sistema de pre-registro orientado a la Internet, hoy en día esta actividad se la maneja a través de una aplicación instalada en los laboratorios del ICM, esta aplicación alimenta un base de datos de Pre-registro, la misma que constituye una fuente de alimentación de información para nuestro sistema en el futuro como se lo muestra en la Figura 3.

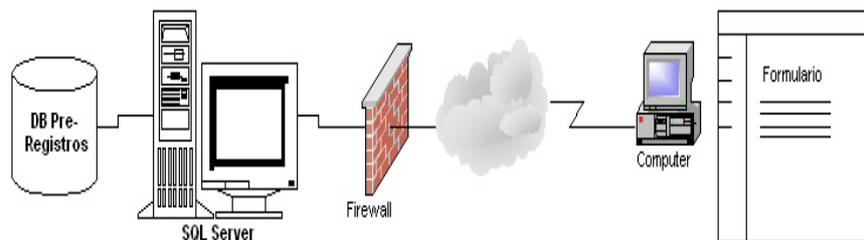


Figura 3 Esquema para Pre-registro.

2.1.2. FASE DE COORDINACIÓN ACADÉMICA.

En lo que respecta a esta fase, se procedió a entrevistar a la coordinadora de la carrera, para que ella nos diera sus sugerencias acerca de los mínimos requerimientos que debería tener nuestro sistema de información.

En esta entrevista se hablo mucho de la necesidad de tener un Data Warehouse de datos que contenga información personal y académica de los estudiantes, dicha información debería estar ordenada y clasificada de tal manera que sea de fácil acceso para las estaciones de trabajo de los coordinadores, quienes son los encargados de elaborar la planificación de los términos académicos, se mencionó que al planificar un nuevo termino académico surge la necesidad de tener un marco de referencia que nos permita determinar cual fue el balance académico del termino o los términos académicos anteriores. Esos resultados poder compararlos con los actuales para ir descubriendo nuevas oportunidades de mejoras en la planificación.

Fuente de Datos del Sistema de Información

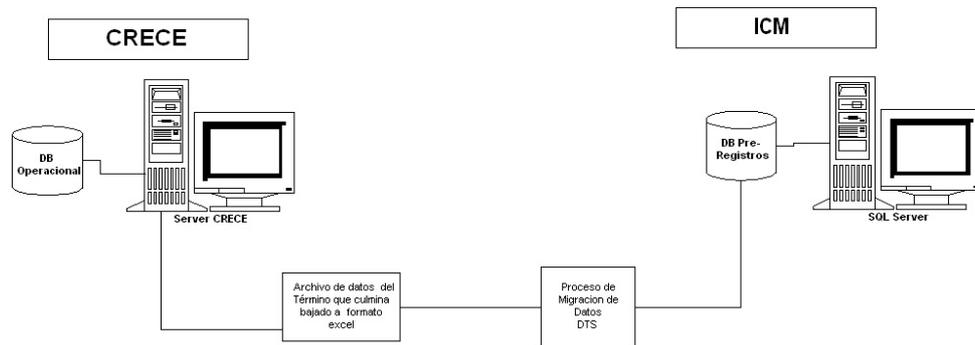


Figura 4. Esquema de carga de datos al data warehouse.

2.1.3. TIPOS DE CONSULTAS

Las consultas que debe incluir el sistema de información las organizamos en 4 grupos:

- Consultas a nivel de profesores.
- Consultas a nivel de estudiantes.
- Consultas a nivel de materias.
- Consultas de nivel de Registros.

Existen niveles de consultas en los cuales se procederá a calcular índices:

- a) Índice 1 = $\frac{\# \text{ materia aprobadas}}{\# \text{ materias tomadas}}$ b) Índice 2 = $\frac{\# \text{ materia reprobadas}}{\# \text{ materias tomadas}}$
- c) Índice 3 = $\frac{\# \text{ estudiantes que toman materias por 1era vez}}{\# \text{ estudiantes registrados en el termino}}$
- d) Índice 4 = $\frac{\# \text{ estudiantes que toman materias por 2da vez}}{\# \text{ estudiantes registrados en el termino}}$
- e) Índice 5 = $\frac{\# \text{ estudiantes que toman materias por 3era vez}}{\# \text{ estudiantes registrados en el termino}}$
- f) Índice 6 = $\frac{\# \text{ estudiantes con mas de 35 materias aprobadas}}{\# \text{ estudiantes registrados en el termino}}$

- g) Índice 7 = $\frac{\# \text{ estudiantes con menos de 35 materias aprobadas}}{\# \text{ estudiantes registrados en el termino.}}$

CAPITULO III

3. PROPUESTA INTEGRAL DEL SISTEMA DE INFORMACION

3.1. COMPONENTES DE HARDWARE Y SOFTWARE

La información histórica de los estudiantes, debemos almacenarla en mega bites de datos. Esta información debe ser extraída de alguna forma para la toma de decisiones, en este caso tenemos que hacer uso de algún software que permita capturar los datos relevantes en forma rápida y pueda verse a través de diferentes dimensiones de los datos. El software no debe limitarse únicamente al acceso de los datos, sino también al análisis significativo de los mismos.

El software o herramienta de negocios inteligentes se colocan sobre la plataforma data warehousing y proveen este servicio. Debido a que es el punto principal de contacto entre la aplicación del deposito y los usuarios, estas herramientas pueden constituir la diferencia entre el éxito y fracaso del sistema de información.

Los software usados en los sistema de información orientados a la tecnología data warehousing se clasifican en herramientas de consulta y reportes, herramientas de base de datos multidimensionales/olap (On Line Analytical processing), Sistemas de información ejecutivos, Herramientas Data Mining y los Sistemas de Gestión de Base de Datos propiamente.

3.2. COMPONENTES DE METODOLOGÍA DE USO

En el sistema de información existirá dos tipos de usuarios que podrán acceder al mismo:

Usuario Administrador del sistema. Que para nuestro caso será el responsable de la carga de información del data warehousing la misma que será enviada desde el CRECE, al finalizar el cada termino académico, esta información será sometida a un proceso de transformación de datos o DTS antes de residir en el Data Warehouse

Usuarios operadores , que para nuestro caso serán los coordinadores de las carreras del instituto de ciencias matemáticas encargados de la planificación de cada termino académico.

3.3. ESPECIFICACIONES DEL SISTEMA.

En esta parte detallamos aspectos importantes del sistema de información, concernientes a la parte funcional, a la parte de manejo, a la parte del entorno de operación, tomado en consideración que este tendrá como interfase un explorador web (Internet Explorer o Netscape). Permitiendo de esta manera una interfaz agradable para los usuarios del sistema.

ASPECTOS FUNCIONALES :

Para acceder al sistema de información se le solicita al usuario que ingrese un nombre de usuario y contraseña para su autenticación, personalizándose para cada usuario registrado en la base de datos del ICM.

ASPECTOS DE MANEJO:

La manera de usar el sistema de información, es similar al tener un sitio web, puesto que el sistema como tal tendrá una interfaz de usuario a través del Internet Explorer o Netscape, esto ayudara a que los usuarios se mantengan en un ambiente lo mas simple y amigable. Lo fundamental en este sistema de información es la capacidad para generar consultas dinámicamente.

ESPECIFICACIONES DE AMBIENTE:

El sistema de información hará uso de un servidor web con sistema operativo Windows. El sistema operativo de los clientes no es importante, debido a que solo debe tener instalado en su maquina Internet Explorer o Netscape, estos browser deben tener soporte para VbScript , javascript, flash.

CAPITULO IV

4. DESARROLLO E IMPLANTACIÓN DEL SISTEMA DE INFORMACIÓN

4.1 BASE DE DATOS MULTIDIMENSIONAL

Se basa en dos elementos fundamentales conocidos como hechos y dimensiones. La tabla de hechos contiene información que se desea medir. Las tablas de dimensiones son las ventanas en que se evalúa la información deseada. Las dimensiones ofrecen la información que se utilizara en las cláusulas where.

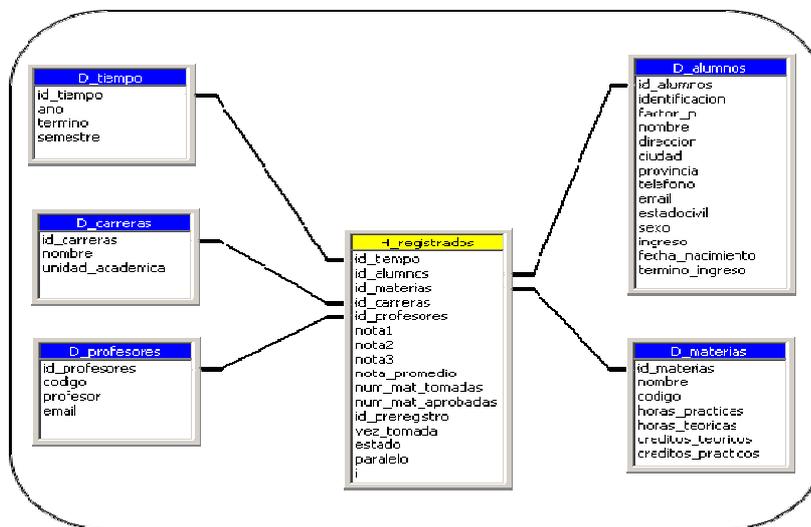


Figura 5 Esquema estrella para el Cubo de Datos

4.2. DISEÑO DE LA INTERFAZ DE USUARIO.

Para la pantalla principal se presenta dividido el área en dos secciones, la primera área ubicada en la parte superior de la página, contiene una fila con todas las opciones que presenta el sistema, la segunda área ubicada en el centro de la página, tenemos enlaces directos a las herramientas de análisis, esto se lo puede ver en la figura 6.



Figura 6. Pantalla Principal.

A continuación se muestran los estados de las materias (aprobadas, reprobadas, perdida por faltas), en un determinado año, termino, nivel que le corresponde según flujo, en las carreras de Auditoría y control de Gestión, Ingeniería en estadística e Informática, así como se muestra en la figura 7.

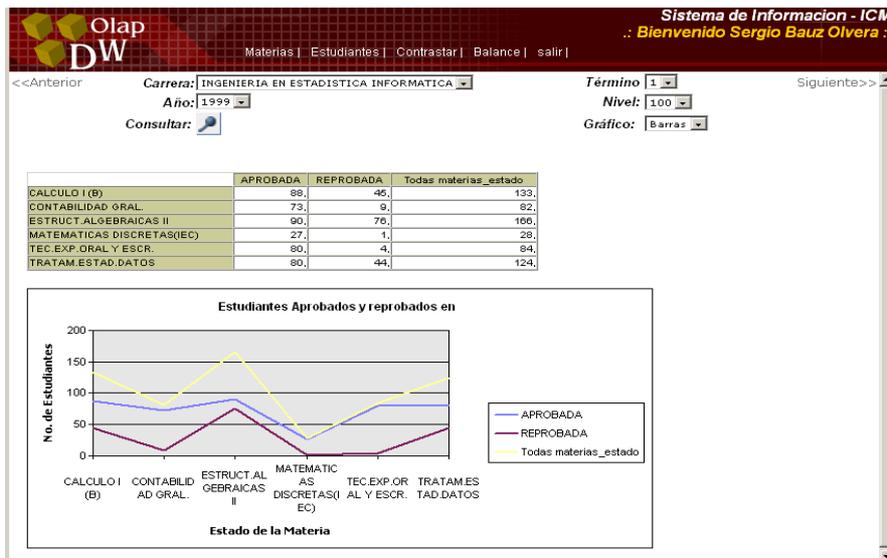


Figura 7. Estado de materias por año, termino, nivel, según flujo de la carrera de Ing. Estadística e Informática.

CONCLUSIONES

- El calculo de índices académicos haciendo uso de la tecnología OLAP (Proceso Analítico en Línea) nos ayudara a encontrar problemas en la coordinación y planificación de los términos académicos en el ICM
- Teniendo un data warehouse en el Instituto de Ciencias Matemáticas, el acceso a los datos por parte de los estudiantes seria inmediato, especialmente para quienes van a desarrollar tesis en el futuro.
- El presente proyecto de sistema de Información puede ser desarrollado con los recursos tanto en hardware como en Software que actualmente posee el Instituto de Ciencias Matemáticas.
- La principal ventaja al desarrollar una herramienta de toma de decisiones orientada al web, es en la instalación y la accesibilidad. Para la primera bastara con solo tener instalada en la estación de trabajo del usuario una versión de Internet Explorer 6.0 o superior, la accesibilidad al sistema de información será inmediata desde el Internet mediante la validación de usuario.

RECOMENDACIONES

Deberá de mejorarse en un futuro el proceso de migración de los datos de la base de datos de la ESPOL a el data warehouse del ICM, actualmente se lo hace mediante la generación de un archivo de Excel.

Para un análisis mas detallado de los resultados que presenta el sistema de información se debería hacer uso de una herramienta estadística especializada.

REFERENCIAS

- 1.- Sharon Bjeletich, Greg Mable, Microsoft SQL SERVER 7.0 al descubierto, México.
- 2.- Tobias Ratschilleo, Till Gerken, Creación de Aplicaciones web con PHP 4.0, Madrid 2001.
- 3.- BILL GATES, Los Negocios en la Era Digital, Plaza y Janes Editores, España.
- 4.- Holly Thomas, Microsoft Office Web Components and Office Server Extensions References, Washington, 1999.
- 5.- Larry Greenfield, Data Warehousing Information Center. (Web site)

6.- Data Warehousing Online. (Web site)

7.- Data Warehousing Information Center. (Web site)

8.- E.F. Codd, S.B. Codd, and C.T. Salley, Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. Technical report, 1993.

9. A. Shoshani, OLAP and Statistical Databases: Similarities and Differences, in Proc. ACM PODS '97, 185-196.