



# **Escuela Superior Politécnica del Litoral**

**Instituto de Ciencias Matemática  
Ingeniería en Estadística Informática**

**DESARROLLO E IMPLEMENTACIÓN DE UN SISTEMA  
ACADEMICO PARA EL INSTITUTO DE CIENCIAS  
MATEMÁTICAS**

**TESIS DE GRADO**

Previo a la Obtención del Título de :

**INGENIERO EN ESTADÍSTICA INFORMATICA**

Presentado por:  
Sergio Alex Bauz Olvera  
GUAYAQUIL - ECUADOR

Año: 2003



## AGRADECIMIENTO

A mi padre por su total confianza,  
a mi madre por su dedicación y  
paciencia, a mis amigos, por su apoyo  
incondicional y al Ing. Washington Jama  
por su invaluable ayuda en este trabajo



# DEDICATORIA

ING. WASHINGTON JARA  
DIRECTOR DE TESIS

A MIS PADES

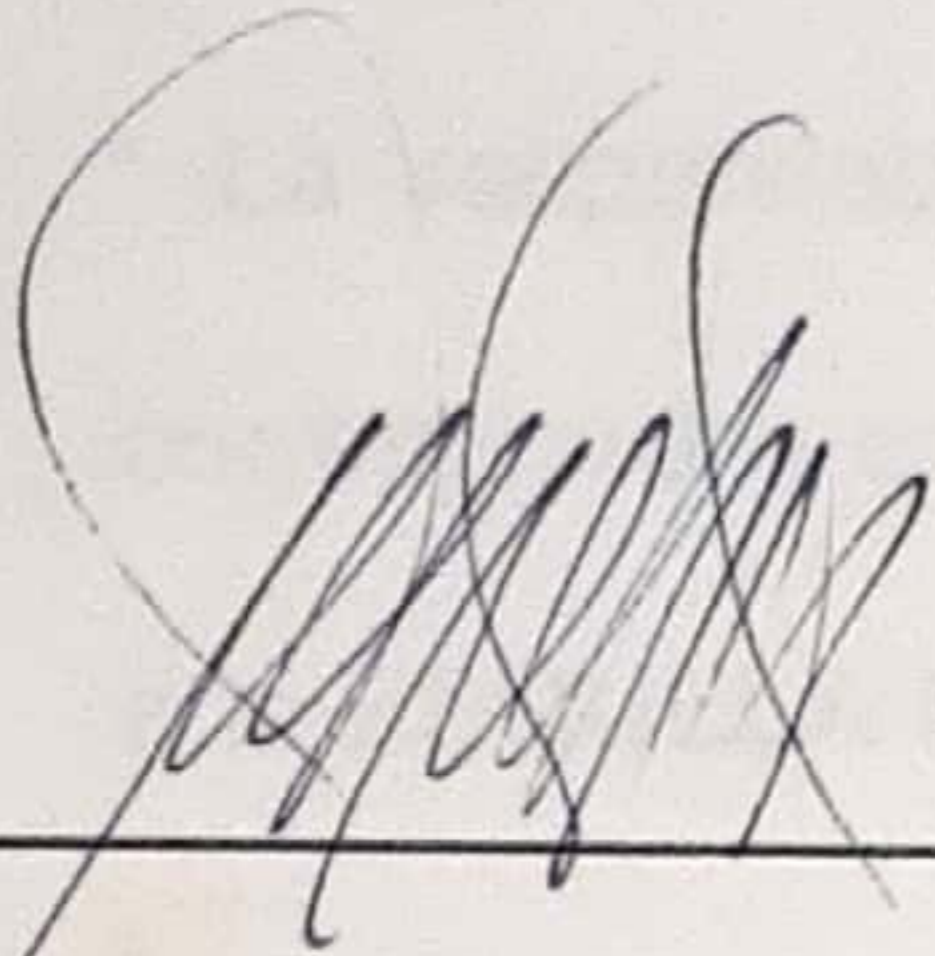
A MIS HERMANOS

A MIS AMIGOS

ING. WASHINGTON JARA  
VOCA

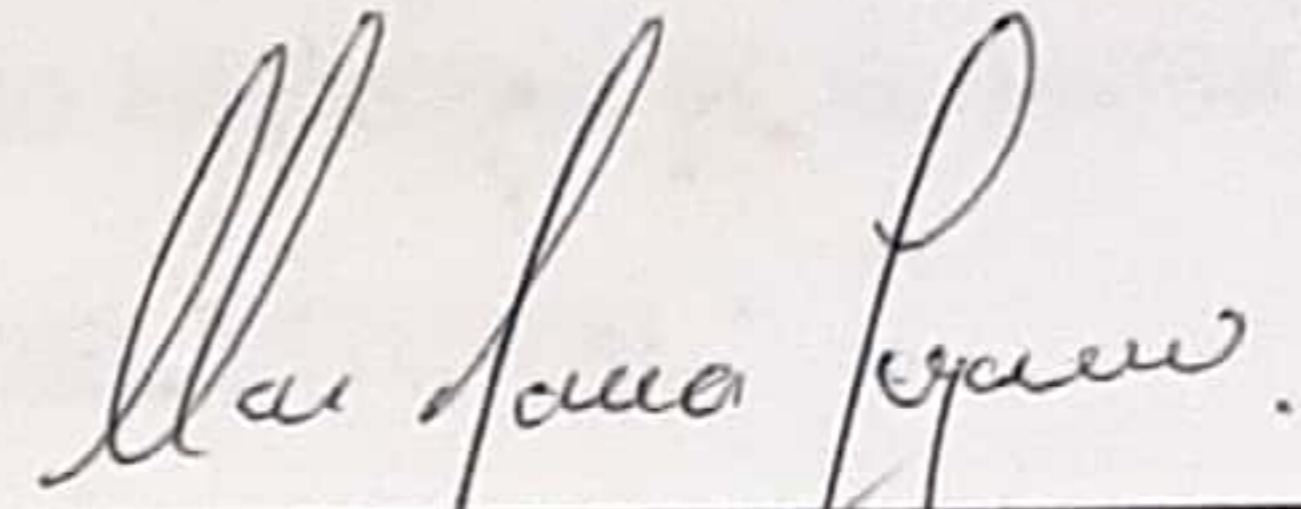


## TRIBUNAL DE GRADUACION



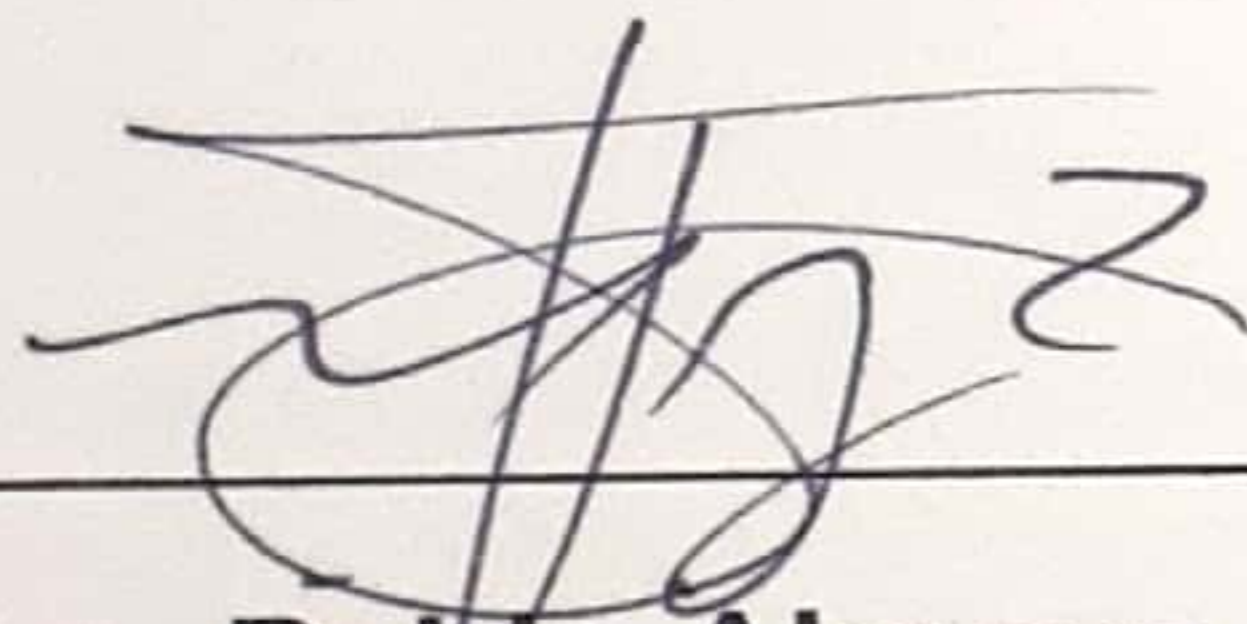
---

**Ing. Luis Rodríguez**  
**SUB-DIRECTOR ICM**



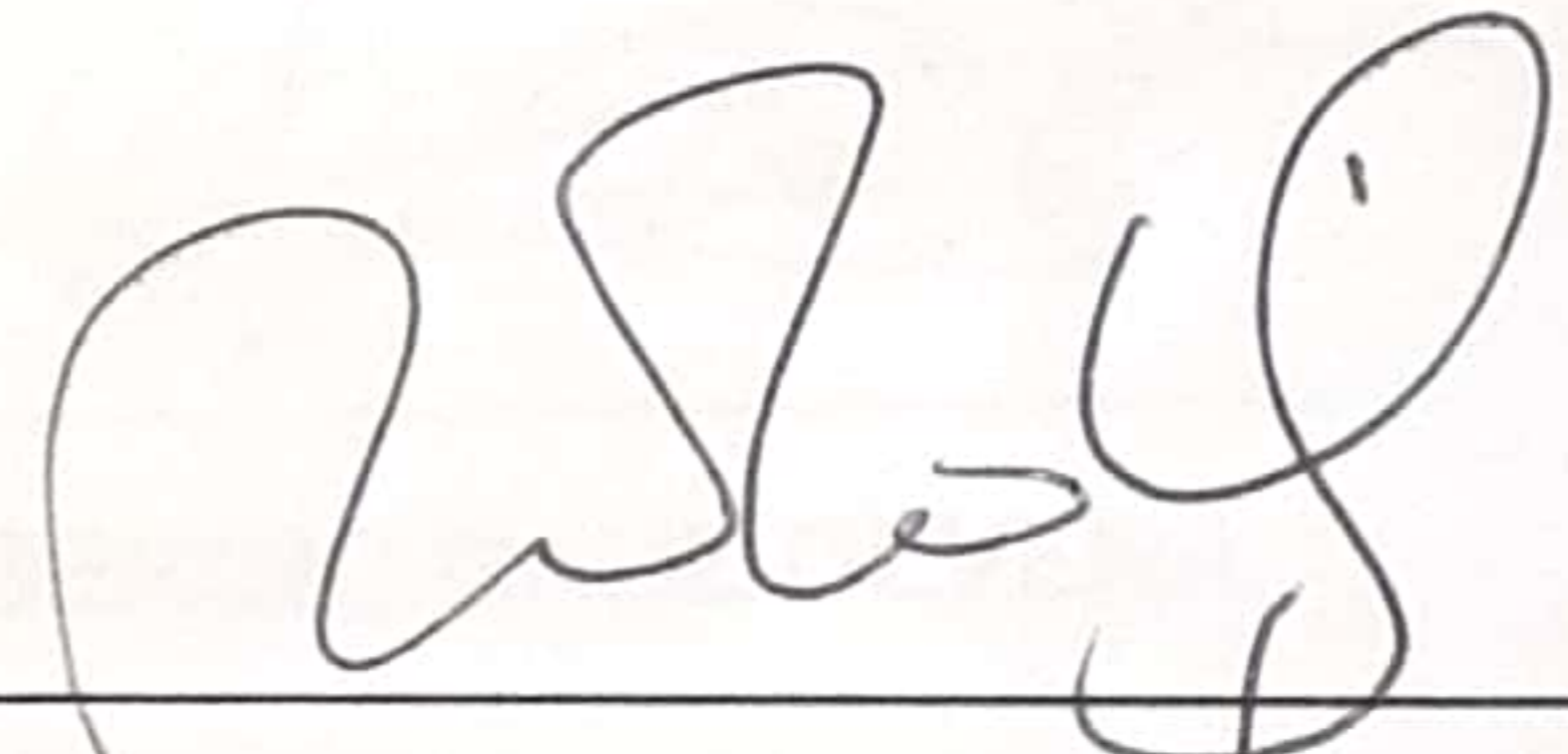
---

**Ing. Washington Jama**  
**DIRECTOR DE TESIS**



---

**Ing. Pablo Alvarez**  
**VOCAL**



---

**Ing. Washington Medina**  
**VOCAL**



## DECLARACIÓN EXPRESA

“ La responsabilidad del contenido de esta Tesis de Grado, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la **ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL** “

( Reglamento de Graduación de la **ESPOL**).

A handwritten signature in black ink, appearing to read 'Sergio Bauz Olvera', is written over a horizontal line. The signature is stylized and cursive.

**SERGIO BAUZ OLVERA**



## RESUMEN

El presente trabajo es un herramienta para toma de decisiones que consiste en desarrollar un software que haga uso de las tecnologías de Data Warehouse y Proceso Analítico en Línea (OLAP). El mismo consta de cuatro capítulos.

En el primer capítulo se ponen en consideración las definiciones, clasificación, ventajas, desventajas y características de los sistemas de sistemas de información para toma de decisiones, también se describe de manera general las definiciones de Data Warehouse y Proceso Analítico en Línea ( OLAP ) como tecnologías para el desarrollo de herramientas para toma de decisiones.

En el segundo capítulo se analizan algunas metodologías para determinar los requerimientos que debería tener el sistema de información, a continuación se habla de las técnica de recopilación de datos, además se describen las entradas y salidas básicas que debería tener el sistema según los directivos del ICM ( Usuarios del Sistema ).

En el tercer capítulo se habla de los componentes en hardware como en software para el desarrollo y ejecución del sistema, también se describen las







# ABREVIATURAS

<b>ODBC</b>	<b>Objeto de conexión a la base de datos</b>
<b>OLAP</b>	<b>Proceso Analítico en Línea</b>
<b>MIS</b>	<b>Sistema de Información Gerencial</b>
<b>EIS</b>	<b>Sistema de Información Ejecutivo</b>
<b>GIS</b>	<b>Sistema de Información Georeferencial</b>
<b>DSS</b>	<b>Sistema de Soporte de Decisiones</b>
<b>SMTP</b>	<b>Protocolo Simple de Transferencia de Correo</b>
<b>IIS</b>	<b>Internet Información Server</b>
<b>ASPMail</b>	<b>Objeto Activos para envío de correo</b>
<b>ASP</b>	<b>Paginas de Servidor Activo</b>
<b>PHP</b>	<b>Pre-Procesador de Hipertexto</b>
<b>JSP</b>	<b>Paginas de Servidor Java</b>



# INDICE GENERAL

## 1. Capitulo I

	Introducción .....	16
1.1.	Definición de Sistema .....	17
1.2.	Sistema de Información .....	17
1.2.1.	Clasificación de los Sistemas de Información .....	18
1.2.1.1.	Sistemas Estratégicos .....	19
1.2.1.2.	Sistemas Tácticos .....	19
1.2.1.3.	Sistemas Técnico Operativos .....	20
1.2.1.4.	Sistemas Interinstitucionales .....	20
1.2.1.5.	Sistemas Técnico Operacionales.....	21
1.2.1.6.	Sistemas de Soporte de Decisions .....	22
1.3.	Data Warehouse Características del Data Warehouse .....	25
1.3.1.	Integrado .....	26
1.3.2.	Temático .....	26
1.3.3.	Histórico .....	26
1.3.4.	No Volátil .....	27
1.3.5.	Procesos de un Data Warehouse.....	32
1.4.	OLAP (Proceso Analítico en Línea) .....	35

## 2 Capitulo II

	Introducción .....	38
2.1.	Técnicas de Recopilación de Datos y Elección .....	39
2.1.1.	La Entrevista .....	40
2.1.2.	Ventajas de Utilizar preguntas abiertas.....	44
2.1.3.	Desventajas de Utilizar preguntas abiertas.....	45
2.1.4.	Ventajas de Utilizar preguntas Cerradas.....	44
2.1.5.	Desventajas de Utilizar preguntas Cerradas.....	45
2.2.	Entrevistas a Directivos del I.C.M. ....	46



2.2.1.	Requerimientos de Información Académica .....	46
2.2.1.1.	Fase de Pre-registro .....	47
2.2.1.2.	Fase de Coordinación Académica .....	48
2.2.3.	Tipos de Consultas .....	50
2.2.3.1.	Consultas a nivel de profesores.....	50
2.2.3.2.	Consultas a nivel de estudiantes.....	52
2.2.3.3.	Consultas a nivel de materias.....	53
2.2.3.4.	Consultas a nivel de registros de materias.....	53
2.3.	Tareas Básicas de un Sistema de Información .....	56
2.4.	Limitaciones del Sistema de Información .....	57

### 3 Capítulo III

	Introducción .....	59
3.1.	Componentes del Sistema de Información .....	59
3.1.1	Componentes de Hardware y Software .....	60
3.1.2.	Componente de Datos .....	62
3.1.3.	Componentes de Metodología de uso .....	64
3.1.4.	Especificaciones del Sistema .....	65
3.1.4.1.	Aspectos Funcionales.....	65
3.1.4.2.	Aspectos de Manejo.....	66
3.1.4.3.	Aspectos de Ambiente.....	67
3.2.	Puntos para Implementación del Sistema .....	68
3.2.1.	Plataforma.....	68
3.2.2.	Servicios OLAP.....	69
3.3.	OLAP en el Web.....	73
3.3.1.	Cubo de Datos .....	76
3.3.2.	Servidor Web (IIS) .....	78
3.3.3.	Otras Tecnologías en Aplicaciones Web.....	78
3.3.4.	Componentes de Software.....	81
3.3.5.	Herramientas de Desarrollo.....	82

### 4 Capítulo IV

	Introducción .....	83
4.1.	Diseño del Sistema .....	83





4.1.1.	Origen de los Datos.....	85
4.1.2.	Carga de los Datos.....	86
4.1.3.	Almacenamiento de los Datos.....	89
4.1.4.	Determinación de Tabla Hechos y Dimensiones.....	89
4.1.5.	Base de Datos Multidimensional .....	91
4.1.6.	Elección del DBMS.....	93
4.1.7.	Data Warehouse en SQL Server.....	97
4.1.8.	Tablas Principales.....	98
4.1.9.	Diseño de la Interfaz de Usuario.....	101
4.2.	Desarrollo del Sistema .....	105
4.2.1.	Programación del Lado de la Aplicación.....	105
4.2.1.1.	Objeto AspMAIL.....	106
4.2.1.2.	Método SendMail.....	107
4.2.1.3.	Manejo de Errores.....	108
4.2.1.4.	Elección del Lenguaje VBScript.....	108
4.2.1.5.	Utilización de VBScript.....	110
4.2.1.6.	ActiveX Scripting.....	111
4.2.1.7.	Declarar Variables.....	112
4.2.1.8.	Constantes.....	113
4.2.1.9.	Comentarios.....	113
4.2.1.10.	JavaScript.....	113
4.2.1.11.	Active Server Pages.....	115
4.2.1.12.	Conección a la Base de Datos.....	116
4.2.1.13.	Abriendo la Base de Datos.....	117
4.2.1.14.	Tablas en HTML.....	118
4.2.2.	Programación del lado de la Base de Datos.....	118
4.2.2.1.	Procedimientos Almacenados.....	119
4.2.2.2.	Disparadores.....	121
4.2.2.3.	Visualizando Resultados.....	122
4.3.	Pruebas del Sistema.....	124
4.3.1.	Prototipos del Sistema.....	124
4.3.2.	Análisis del Prototipo.....	125
4.3.3.	Necesidades de Usar un Prototipo.....	127
4.3.4.	Determinación del Sistema a Implementar.....	128
4.3.5.	Pruebas del Sistema.....	130
4.3.6.	Planificación de las Pruebas.....	132
4.3.7.	Herramientas para Pruebas.....	135



4.3.8.	Realización de Pruebas.....	136
4.3.9.	Tipos de Pruebas .....	136
4.4.	Uso y Mantenimiento .....	137
	<b>Conclusiones y Recomendaciones.....</b>	<b>139</b>
	<b>Anexos .....</b>	<b>140</b>
	<b>Bibliografía .....</b>	<b>142</b>

# 1. ENFOQUE DE LOS SISTEMAS DE INFORMACIÓN COMO HERRAMIENTA PARA OBTENER BIENES DE EFICIENCIA DENTRO DE UN ENTORNO DE APRENDIZAJE

El presente trabajo se centra en el análisis de los sistemas de información como herramienta para obtener bienes de eficiencia dentro de un entorno de aprendizaje. Se exploran los conceptos clave y se discuten las implicaciones de su uso en el contexto educativo. El objetivo principal es comprender cómo estos sistemas pueden mejorar el proceso de enseñanza y aprendizaje, optimizando recursos y mejorando los resultados académicos. Se examina el papel de la tecnología en la creación de entornos de aprendizaje personalizados y adaptados a las necesidades de cada estudiante. Se discuten los desafíos asociados con la implementación de estos sistemas, como la falta de recursos humanos y técnicos, y se ofrecen recomendaciones para superarlos. El estudio concluye que los sistemas de información son una herramienta poderosa que, cuando se utiliza correctamente, puede transformar el entorno de aprendizaje y mejorar significativamente la eficiencia y la calidad de la educación.



# **CAPÍTULO I**

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

En esta sección se pondrán en consideración diferentes conceptos que tienen que ver con utilización de sistemas de información para obtener oportunidades de mejoras en las instituciones.

### **De que manera las organizaciones usan la información.**

“Esta pregunta es mucho mas difícil de contestar de lo que parece. La información es un componente intrínstico de casi todo lo que hace una organización, tanto que su función se ha vuelto transparente”. Sin embargo, la pregunta no es inoportuna. Sin una cabal comprensión de los procesos humanos y de la organización a través de las cuales la



información se transforma es discernimiento, conocimiento y acción, una organización es incapaz de aprovechar el verdadero valor de sus recursos informativos y tecnologías de información. Trataremos de presentar una visión preliminar del marco conceptual que une las principales formas en que una organización usa la información de manera estratégica, e indicamos cómo esos procesos están interrelacionados íntimamente y podrían controlarse para crear una organización inteligente”.

Bajo este concepto de organización inteligente y en su afán por conseguir una herramienta que permita recopilar, analizar y ayudar a tomar decisiones, ponemos en consideración de los directivos del Instituto de Ciencias Matemáticas de la ESPO, la utilización de un sistema de información que de soporte a toma de decisiones permitiendo calcular índices académicos, a través del uso de consultas multidimensionales a una base de datos que contiene información histórica de los estudiantes registrados en la carrera de ingeniería en estadística Informática, Auditoría y Control de Gestión, esta base de datos se encuentra optimizada y depurada.

Estas grandes bases de datos se las conocen como Repositorios de datos o Data warehouse, junto con la tecnología OLAP (On Line



analytical Processing – proceso analítico en línea) conforman, una poderosa herramienta para quienes tienen a su cargo la responsabilidad dirigir los destinos del instituto de Ciencias Matemáticas

Esta herramienta tendrá como front-end un browser (Internet Explorer o netscape) haciendo de esta manera la navegación mas amigable para quienes interactúen con el sistema.

### **1.1 Definición de Sistema.**

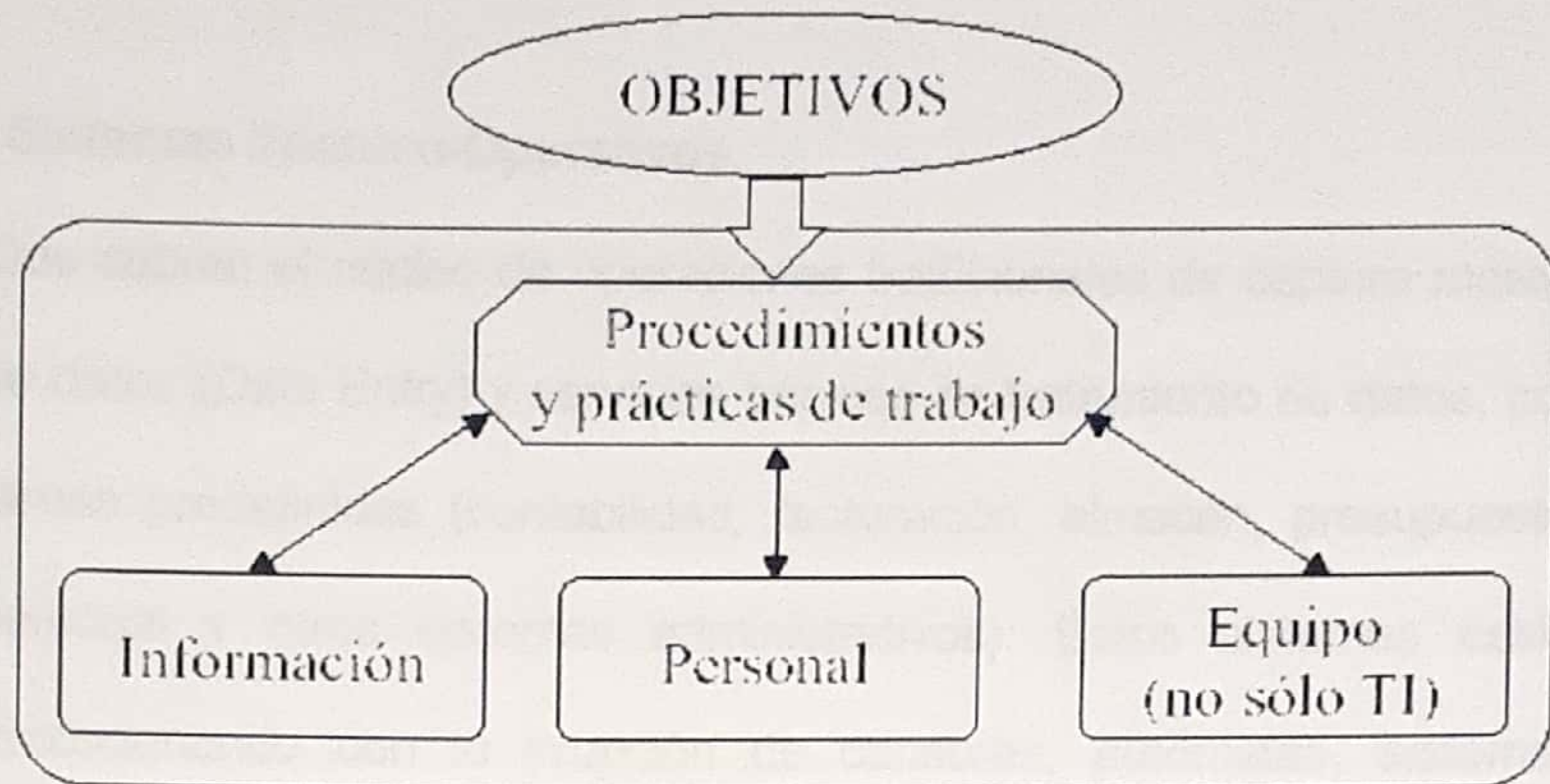
“Un conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objetivo”.

### **1.2 Que es un Sistema de Información.**

“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).



En la figura 1 se muestran los elementos que conforman un sistema de información.



**Figura 1.** Elementos del Sistema de Información y sus relaciones

### 1.2.1 Clasificación de los Sistemas de Información.

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



**Figura 2.** Clasificación de los Sistemas de Información.



### **1.2.1.1 Sistemas Estratégicos.**

Orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionándole un soporte básico, en forma de mejor información, para la toma de decisiones. Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible, al contrario de los casos anteriores, cuya utilización es periódica.

Destacan entre estos sistemas: los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial-AI).

### **1.2.1.2 Sistemas Tácticos.**

Diseñados para soportar las actividades de coordinación de actividades y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema, proporcionar informes y, en resumen, facilitar la gestión independiente de la información por parte de los niveles intermedios de la organización.

Destacan entre ellos: los Sistemas Ofimáticos (OA), Sistemas de Transmisión de Mensajería (E-mail y Fax Server), coordinación y



control de tareas (Work Flow) y tratamiento de documentos (Imagen, Trámite y Bases de Datos Documentarios).

#### **1.2.1.3 Sistemas Técnico-Operativos.**

Que cubren el núcleo de operaciones tradicionales de captura masiva de datos (Data Entry) y servicios básicos de tratamiento de datos, con tareas predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos). Estos sistemas están evolucionando con la irrupción de sensores, autómatas, sistemas multimedia, bases de datos relacionales más avanzadas y data warehousing.

#### **1.2.1.4 Sistemas Interinstitucionales.**

Este último nivel de sistemas de información recién está surgiendo, es consecuencia del desarrollo organizacional orientado a un mercado de carácter global, el cual obliga a pensar e implementar estructuras de comunicación más estrechas entre la organización y el mercado (Empresa Extendida, Organización Inteligente e Integración Organizacional), todo esto a partir de la generalización de las redes informáticas de alcance nacional y global (INTERNET), que se convierten en vehículo de comunicación entre la organización y el



mercado, no importa dónde esté la organización (INTRANET), el mercado de la institución (EXTRANET) y el mercado (Red Global).

Sin embargo, la tecnología data warehousing basa sus conceptos y diferencias entre dos tipos fundamentales de sistemas de información en todas las organizaciones: los sistemas técnico-operacionales y los sistemas de soporte de decisiones. Este último es la base de un data warehouse.

#### **1.2.1.5 Sistemas Técnico-Operacionales.**

Como indica su nombre, son los sistemas que ayudan a manejar la empresa con sus operaciones cotidianas. Estos son los sistemas que operan sobre el "backbone" (columna vertebral) de cualquier empresa o institución, entre las que se tiene sistemas de ingreso de órdenes, inventario, fabricación, planilla y contabilidad, entre otros.

Debido a su volumen e importancia en la organización, los sistemas operacionales siempre han sido las primeras partes de la empresa a ser computarizados. A través de los años, estos sistemas operacionales se han extendido, revisado, mejorado y mantenido al punto que hoy, ellos son completamente integrados en la organización.



Desde luego, la mayoría de las organizaciones grandes de todo el mundo, actualmente no podrían operar sin sus sistemas operacionales y los datos que estos sistemas mantienen.

#### **1.2.1.6 Sistema de Soporte de Decisiones.**

Por otra parte, hay otras funciones dentro de la empresa que tienen que ver con el planeamiento, previsión y administración de la organización. Estas funciones son también críticas para la supervivencia de la organización, especialmente en nuestro mundo de rápidos cambios.

Las funciones como "planificación de marketing", "planeamiento de ingeniería" y "análisis financiero", requieren, además, de sistemas de información que los soporte. Pero estas funciones son diferentes de las operacionales y los tipos de sistemas y la información requerida son también diferentes. Las funciones basadas en el conocimiento son los 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. Estos sistemas no sólo tienen un enfoque diferente al de los operacionales, sino que, por lo general, tienen un alcance diferente.



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.

Son estos sistemas sobre los que se basa la tecnología data warehousing.

El data warehouse, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Un data warehouse es una colección de datos en la cual se encuentra integrada la información de la Institución y que se usa como soporte para el proceso de toma de decisiones gerenciales. Aunque diversas organizaciones y personas individuales logran comprender el enfoque de un Warehouse, la experiencia ha demostrado que existen muchas dificultades potenciales.

Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, consultas y el menor tiempo de uso de la información.



Las aplicaciones para soporte de decisiones basadas en un data warehousing, pueden hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio, que no se logra cuando se usan sólo los datos que provienen de las aplicaciones operacionales (que ayudan en la operación de la empresa en sus operaciones cotidianas), en los que la información se obtiene realizando procesos independientes y muchas veces complejos.

Un data warehouse se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego, cargadas en el data warehouse. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.

La innovación de la Tecnología de Información dentro de un ambiente data warehousing, puede permitir a cualquier organización hacer un uso más óptimo de los datos, como un ingrediente clave para un

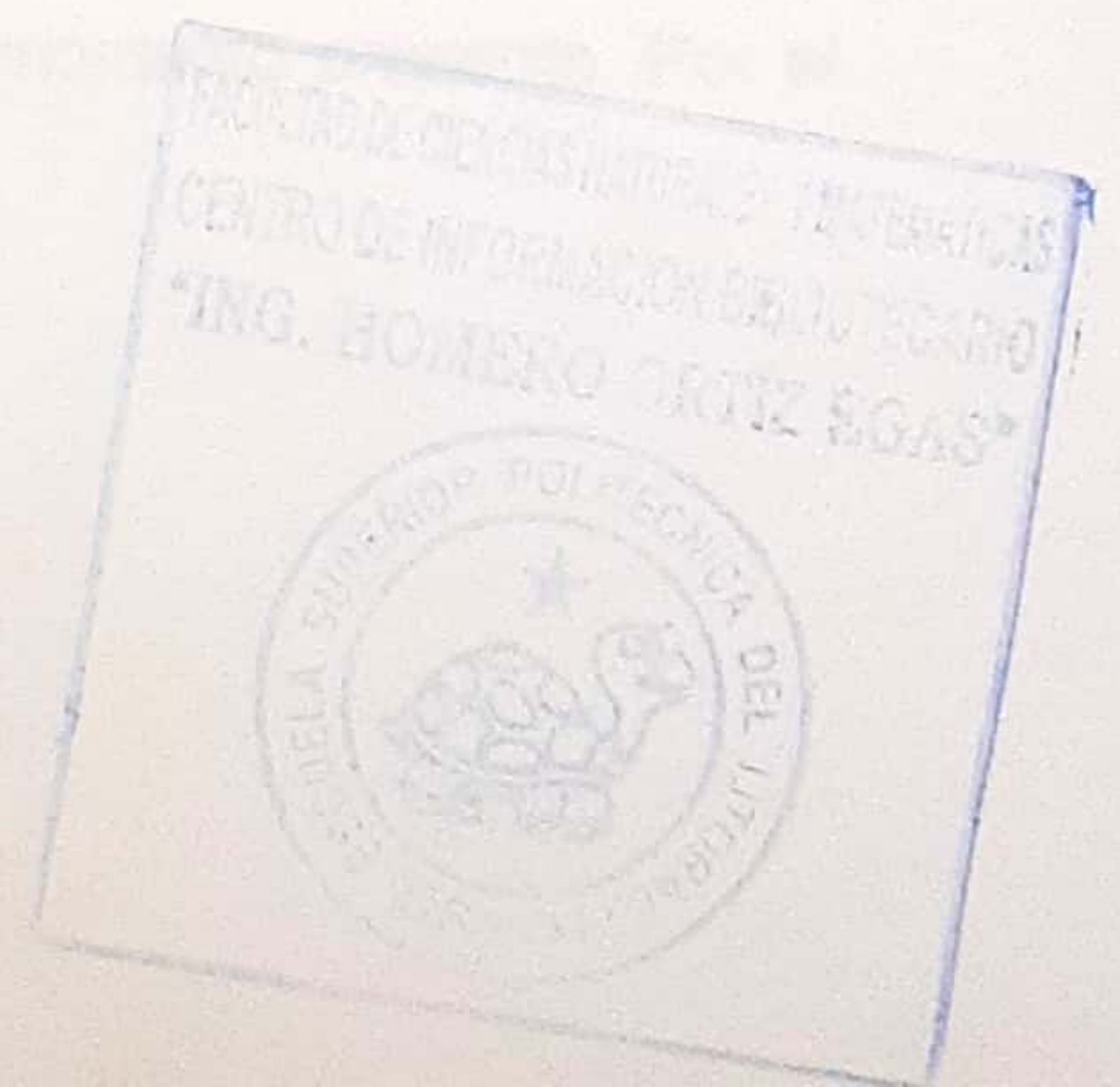


proceso de toma de decisiones más efectivo. Las organizaciones tienen que aprovechar sus recursos de información para crear la información de la operación del negocio, pero deben considerarse las estrategias tecnológicas necesarias para la implementación de una arquitectura completa de data warehouse.

### 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. Según definió Bill Inmon, el Data Warehouse se caracteriza por ser:





### **1.3.1 Integrado.**

Los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

### **1.3.2 Temático.**

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

### **1.3.3 Histórico.**

El tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el



contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

#### **1.3.4 No Volátil.**

El almacén de información de un Data Warehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

E.F. Codd, considerado como el padre de las bases de datos relacionales, ha venido insistiendo desde principio de los noventa, que disponer de un sistema de bases de datos relacionales, no significa disponer de un soporte directo para la toma de decisiones. Muchas de estas decisiones se basan en un análisis de naturaleza multidimensional, que se intentan resolver con la tecnología no orientada para esta naturaleza. Este análisis multidimensional, parte de una visión de la información como dimensiones de negocio.



Tiempo de resolución. La visión general de la información de ventas para estas dimensiones definidas, la representaremos, gráficamente como el cubo de abajo:



Un gerente de una zona estaría interesado en visualizar la información para su zona en el tiempo para todos los productos que distribuye. Un director de producto, sin embargo querría examinar la distribución geográfica de sus productos, para toda la información histórica almacenada en el Data Warehouse.

O se podría también examinar los datos en un determinado momento o una visión particularizada.





A su vez estas dimensiones tienen una jerarquía, interpretándose en el cubo como que cada cubo elemental es un dato elemental, del que se puede extraer información agregada. En el ejemplo anterior podría ser:

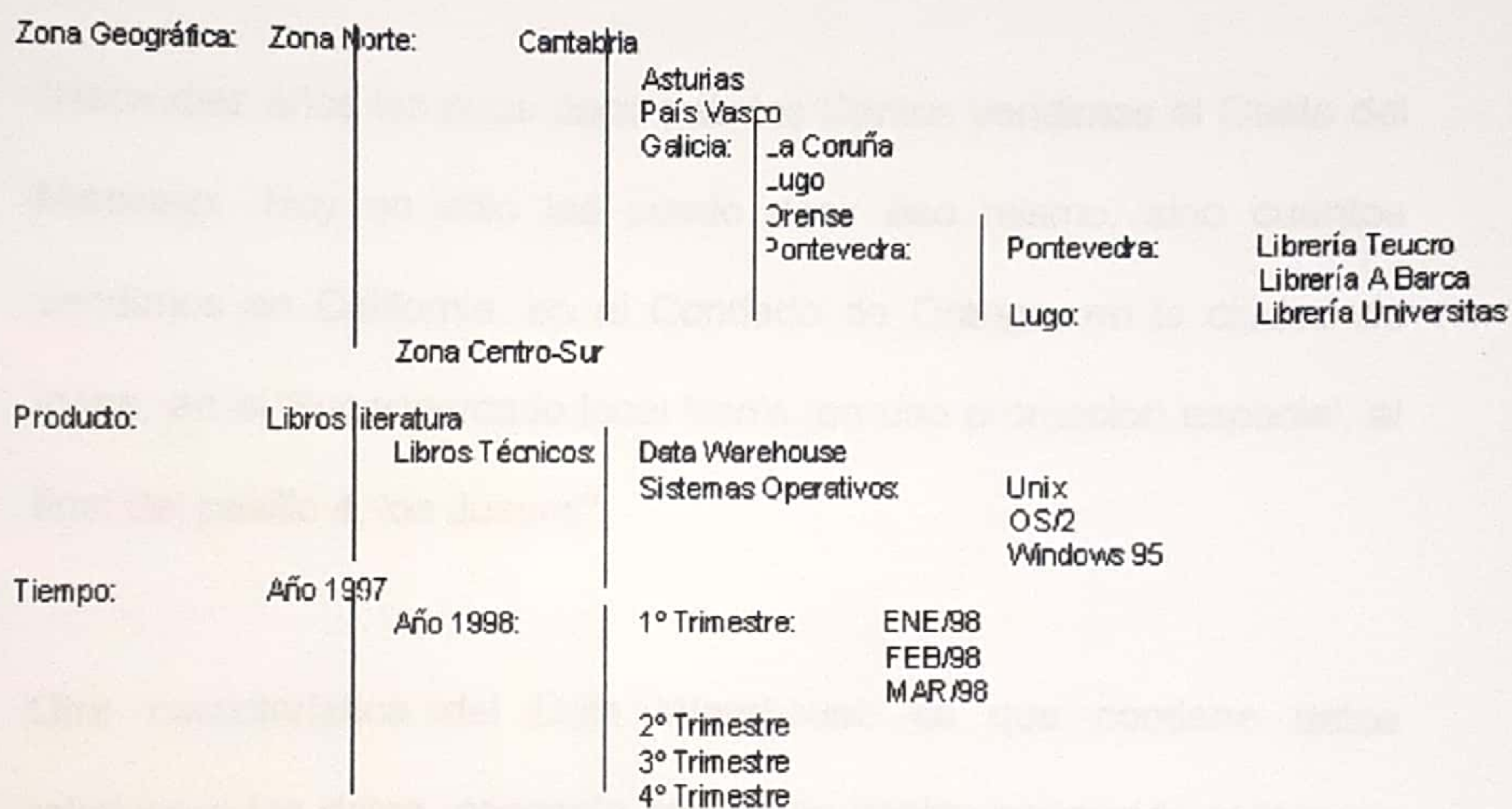


Figura 3. Ejemplo de Cubo de Información



Y así por ejemplo se podría querer analizar la evolución de las ventas en Galicia de libros de Física por meses desde Febrero del 1996 hasta Marzo del 1997. Ello es fácil de obtener (incluso a "golpe de ratón") si la información de ventas se ha almacenado en un Data Warehouse, definiendo estas jerarquías y estas dimensiones de negocio.

En este sentido citamos las palabras de D. Wayne Calloway Director Ejecutivo de Operaciones de Pepsico en una asamblea general de accionistas:

"Hace diez años les pude decir cuántos Doritos vendimos al Oeste del Mississippi. Hoy no sólo les puedo decir eso mismo, sino cuántos vendimos en California, en el Condado de Orange, en la ciudad de Irvine, en el Supermercado local Von's, en una promoción especial, al final del pasillo 4, los Jueves".

Otra característica del Data Warehouse es que contiene datos relativos a los datos, concepto que se ha venido asociando al término de metadatos. Los metadatos permiten mantener información de la procedencia de la información, la periodicidad de refresco, su fiabilidad, forma de cálculo, etc., relativa a los datos de nuestro almacén. Estos metadatos serán los que permitan simplificar y



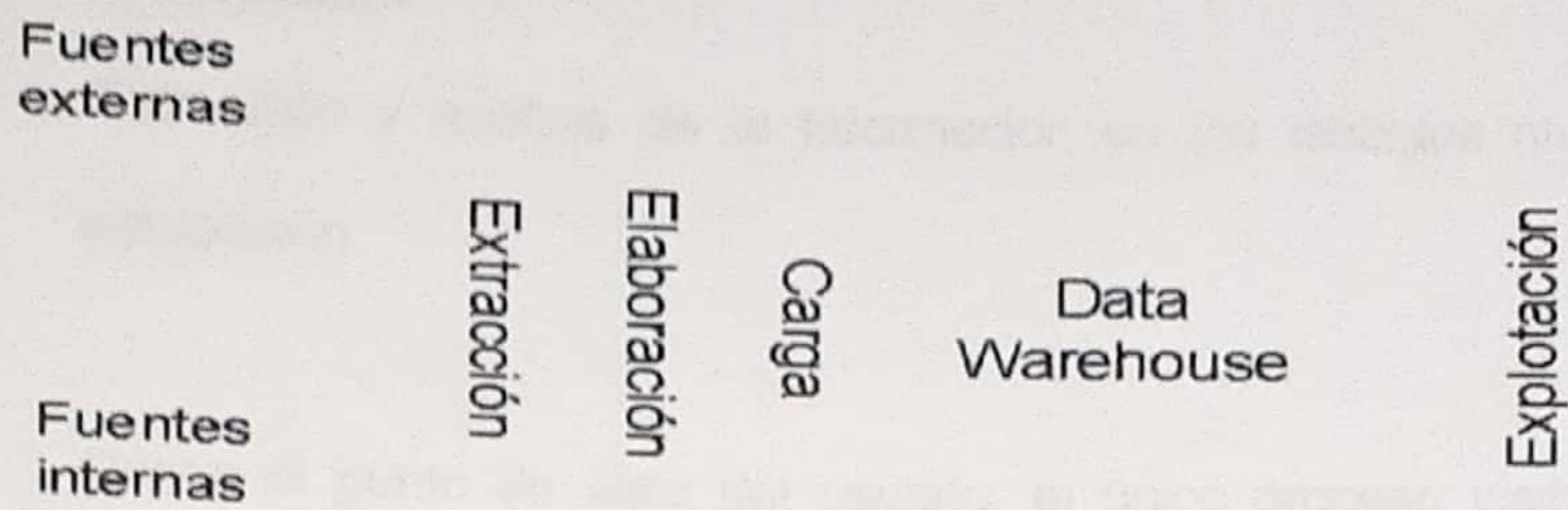
automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Los objetivos que deben cumplir los metadatos, según el colectivo al que va dirigido, serían:

- Soportar al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de navegación.
- Soportar a los responsables técnicos del Data Warehouse en aspectos de auditoría, gestión de la información histórica, administración del Data Warehouse, elaboración de programas de extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos, etc.

Para comprender el concepto de Data Warehouse, es importante considerar los procesos que lo conforman. A continuación se describen dichos procesos clave en la gestión de un Data Warehouse:





**Figura 4.** Gestión de un Data Warehouse

### 1.3.5. PROCESOS DE UN DATAWARE HOUSE.

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

- **Extracción:**  
Obtención de información de las distintas fuentes tanto internas como externas.
- **Elaboración:**  
Filtrado, limpieza, depuración, homogeneización y agrupación de la información.
- **Carga:**  
Organización y actualización de los datos y los metadatos en la base de datos.



- **Explotación:**

Extracción y análisis de la información en los distintos niveles de agrupación.

Desde el punto de vista del usuario, el único proceso visible es la explotación del almacén de datos, aunque el éxito del Data Warehouse radica en los tres procesos iniciales que alimentan la información del mismo y suponen el mayor porcentaje de esfuerzo (en torno a un 80%) a la hora de desarrollar el almacén.

Las diferencias de un Data Warehouse con un sistema tradicional las podríamos resumir en el siguiente esquema:

SISTEMA TRADICIONAL	DATA WAREHOUSE
<ul style="list-style-type: none"> <li>• Predomina la actualización</li> </ul>	<ul style="list-style-type: none"> <li>• Predomina la consulta</li> </ul>
<ul style="list-style-type: none"> <li>• La actividad más importante es de tipo operativo (día a día)</li> </ul>	<ul style="list-style-type: none"> <li>• La actividad más importante es el análisis y la decisión estratégica</li> </ul>
<ul style="list-style-type: none"> <li>• Predomina el proceso puntual</li> </ul>	<ul style="list-style-type: none"> <li>• Predomina el proceso masivo</li> </ul>
<ul style="list-style-type: none"> <li>• Mayor importancia a la estabilidad</li> </ul>	<ul style="list-style-type: none"> <li>• Mayor importancia al dinamismo</li> </ul>
<ul style="list-style-type: none"> <li>• Datos en general desagregados</li> </ul>	<ul style="list-style-type: none"> <li>• Datos en distintos niveles de detalle y agregación</li> </ul>



<ul style="list-style-type: none"> <li>• Importancia del dato actual</li> </ul>	<ul style="list-style-type: none"> <li>• Importancia del dato histórico</li> </ul>
<ul style="list-style-type: none"> <li>• Importante del tiempo de respuesta de la transacción instantánea</li> </ul>	<ul style="list-style-type: none"> <li>• Importancia de la respuesta masiva</li> </ul>
<ul style="list-style-type: none"> <li>• Estructura relacional</li> </ul>	<ul style="list-style-type: none"> <li>• Visión multidimensional</li> </ul>
<ul style="list-style-type: none"> <li>• Usuarios de perfiles medios o bajos</li> </ul>	<ul style="list-style-type: none"> <li>• Usuarios de perfiles altos</li> </ul>
<ul style="list-style-type: none"> <li>• Explotación de la información relacionada con la operativa de cada aplicación</li> </ul>	<ul style="list-style-type: none"> <li>• Explotación de toda la información interna y externa relacionada con el negocio</li> </ul>

del mismo y suponen el mayor porcentaje de esfuerzo (en torno a un 80%) a la hora de desarrollar el almacén una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando a un departamento usuario como piloto y expandiendo progresivamente el almacén de datos a los demás usuarios. Por ello es importante elegir este usuario inicial o piloto, siendo importante que sea un departamento con pocos usuarios, en el que la necesidad de este tipo de sistemas es muy alta y se puedan obtener y medir resultados a corto plazo.



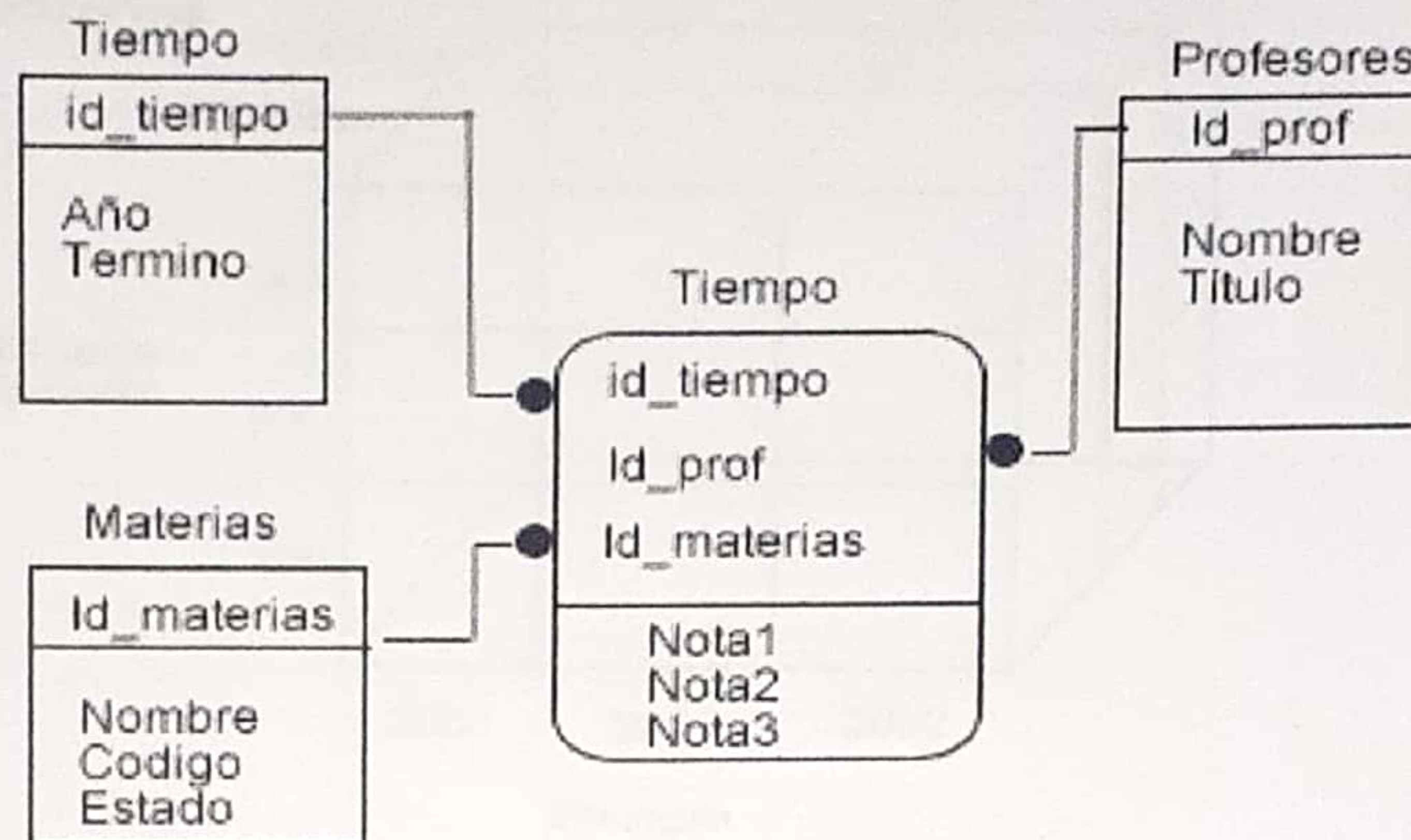
#### 1.4 Qué es OLAP( Proceso Analítico en Línea ) .

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 (EIS. *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.

Los productos OLAP utilizan diversas aproximaciones para alcanzar sus metas DSS-EIS, pero por debajo de estas distintas implementaciones yace una base de datos dimensional, que se utiliza para construir cubos de datos, los cubos son bloques lógicos de datos, que las personas pueden utilizar para observar la actividad histórica empresarial. El esquema dimensional básico contiene una tabla de hechos con sucesos detallados, que se identifican mediante dimensiones. los hechos son medidas de las actividades; el dinero y las cantidades constituyen normalmente las medidas de una base de datos dimensional, Las dimensiones colocan a los hechos en su contexto e incluyen atributos, como unidades de tiempo, información de productos e información geográfica, por nombrar unos cuantos. figura 1.2 representa un modelo



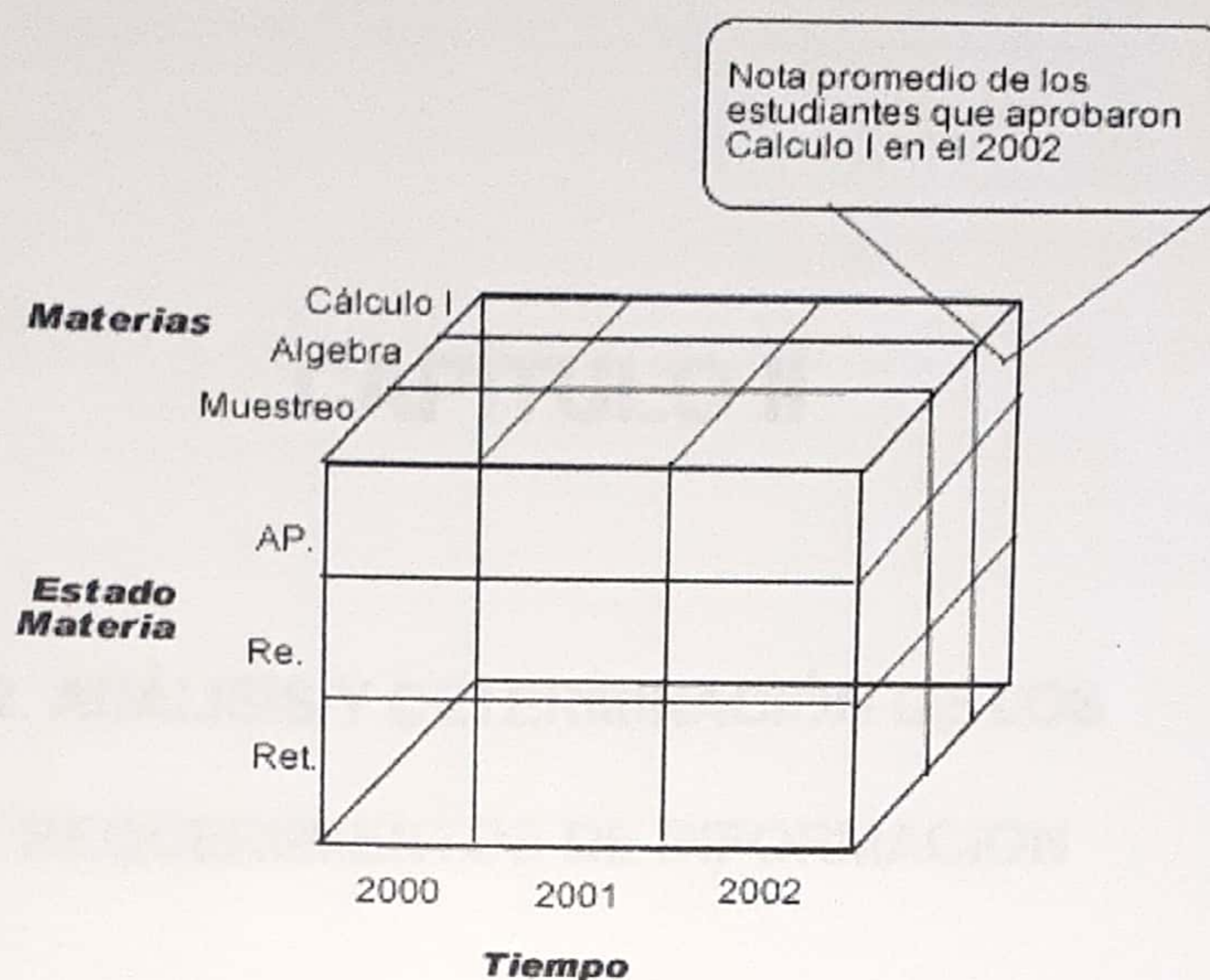
simplificado de una base de datos dimensional y es un ejemplo de un esquema en «estrella» básico.



**Figura 1.5.** Esquema estrella.

Las dimensiones constituyen la ruta de acceso a los hechos interesantes; los miembros de una dimensión deben ser textuales y se emplean como criterios para las consultas y como encabezados de fila y columna en los resultados de las consultas. Utilizando las dimensiones, las medidas seleccionadas se pueden visualizar como cubos, como se muestra en la Figura 1.6





**Figura 1.6** Base de datos multidimensional Vista en forma de un cubo

Cada celda del cubo contiene un hecho. El criterio de filtrado y los ejes del cubo provienen de las tablas de dimensiones. Si la medida seleccionada en la tabla Materias es nota, entonces la celda en la que se interceptan la dimensión tiempo 2002, la dimensión de tipo de materia Cálculo I, y la dimensión estado de la materia contendrá la nota correspondientes a los estudiantes que aprobaron Cálculo I en 2002. Los hechos tienen que tener tres características (ser numéricos, tener un rango continuo de valores y ser aditivos), para producir resultados significativos sobre distintos períodos de tiempo.

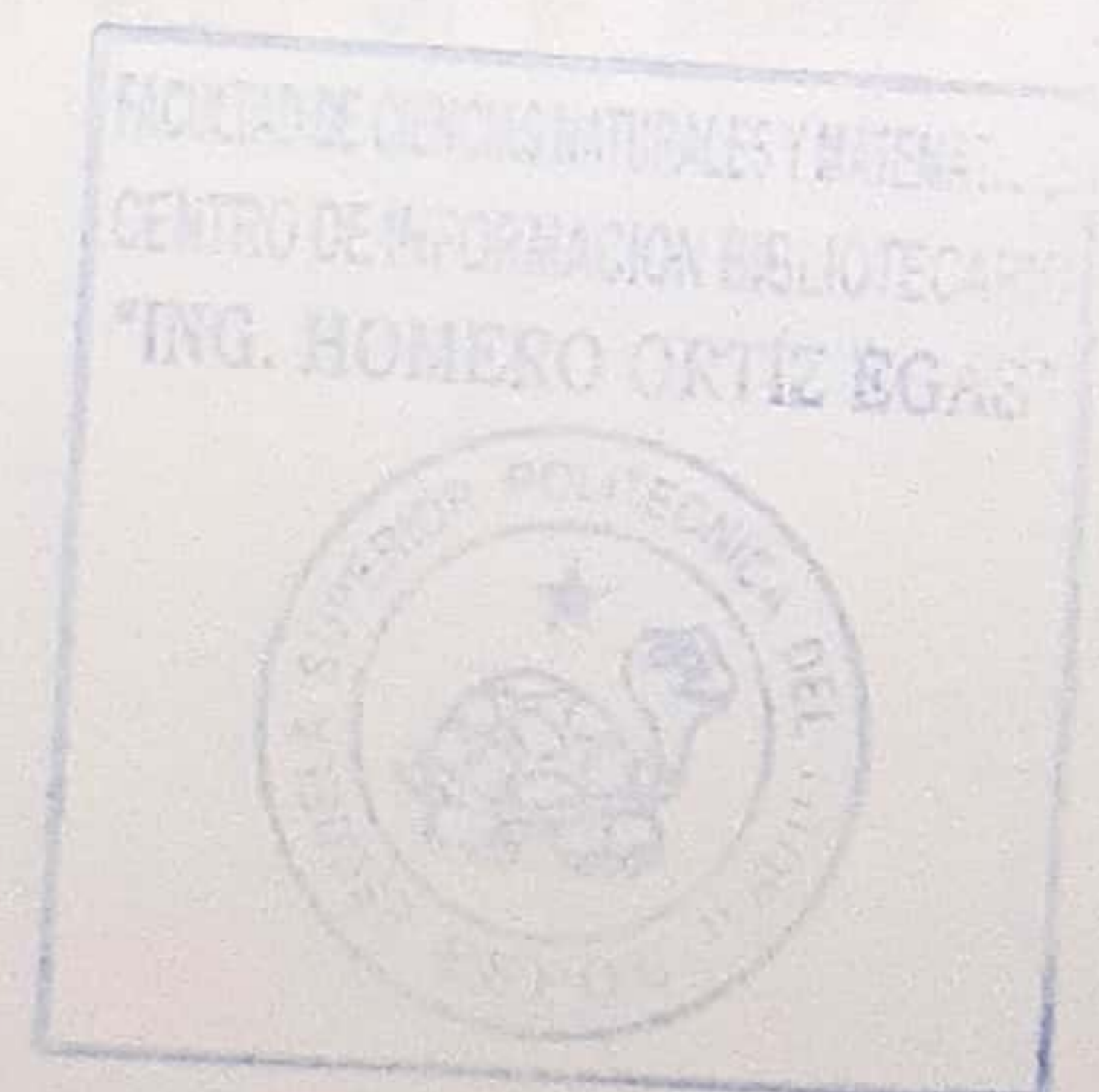


# CAPÍTULO II

## 2. ANÁLISIS Y DETERMINACIÓN DE LOS REQUERIMIENTOS DE INFORMACION

En esta sección se pondrán en consideración diferentes metodologías para la determinación de los requerimientos de información, para lo cual abordaremos la que se ajuste a nuestro proyecto.

- En este punto nos concentraremos en comprender que información necesitan los usuarios, cuales deberían ser las salidas de información que cumplan con los requisitos mínimos para que los usuarios del sistema información hagan valido el uso de esta herramienta como medio para desarrollar su trabajo de toma de decisiones, cuales son las fuentes de información actuales, cuales es el alcance del sistema de información, entre otras.





- Las técnicas de recopilación de datos permiten la interacción directa con los usuarios, con su ambiente de trabajo, con el tiempo que destinan para sus actividades de planificación de tareas, con las formas de reportes preparados por ellos, permitiendo identificar cuales son las entradas y salidas de datos que ellos están acostumbrados, con la interfaz de usuario que mejor se acomoda al entorno de trabajo, entre otras.

## 2.1 Técnicas de Recopilación de datos.

Las principales metodologías para recopilar datos dentro del contexto del análisis de sistemas son:

- El muestreo e investigación de los datos relevantes.
- Las entrevistas.
- Cuestionarios.
- El comportamiento de los tomadores de decisiones y su ambiente de oficina y hasta la elaboración de los prototipos.

De las varias técnicas de recopilación de datos mencionadas anteriormente nosotros aremos uso de la entrevista por ser flexible en cuanto a factores como tiempo disponible por directivos, discusión de consultas predeterminadas que se debería incluir en el sistema, esquemas gráficos de presentación, entre otros.



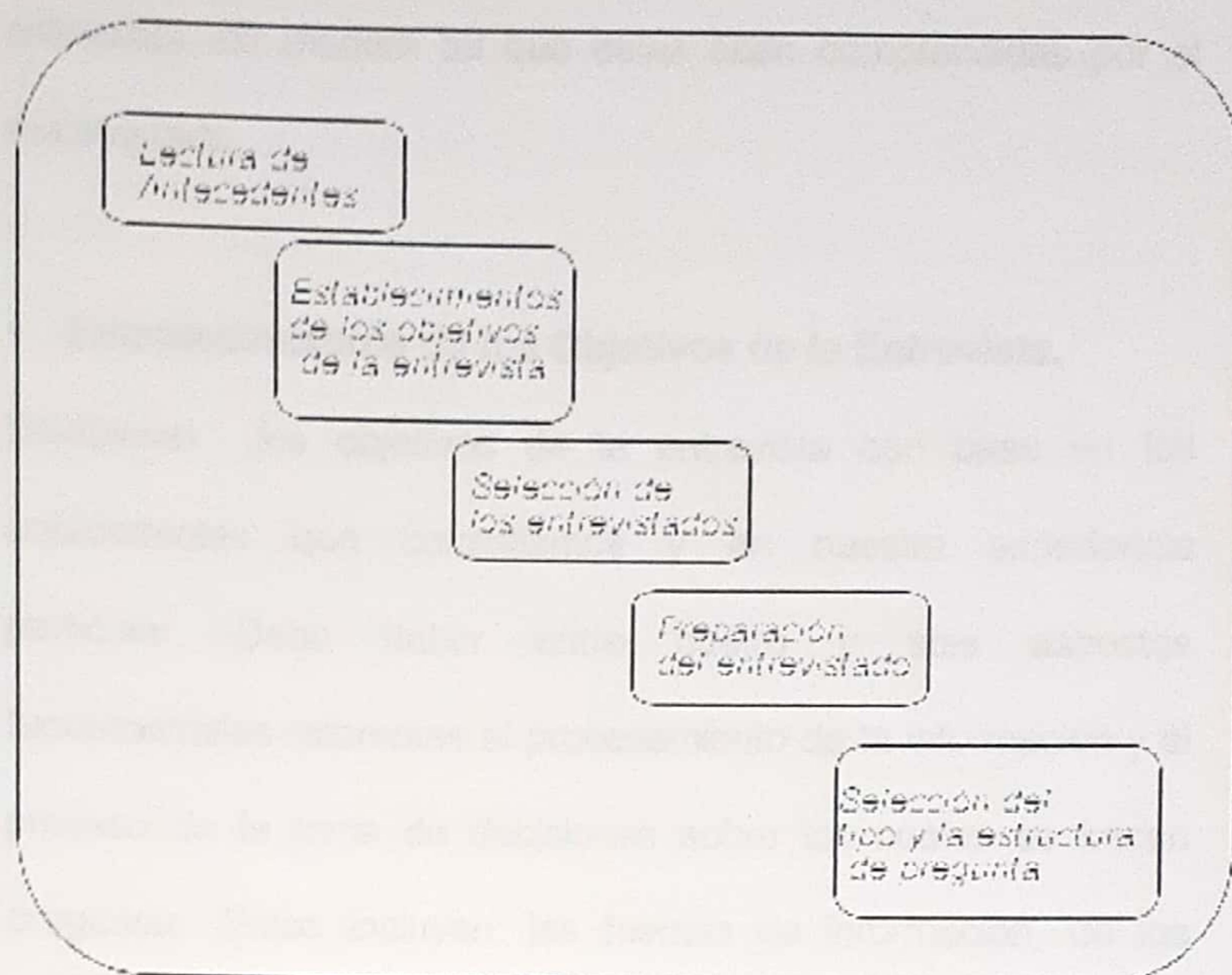
La técnica permite interactuar directamente con los directivos del ICM, mediante la preparación de preguntas básicas que nos ayuden a transmitir el pensamiento de los usuarios a nuestro sistema de información. La metodología de trabajo es concertar reuniones con los involucrados de la planificación del término académico en el Instituto de Ciencias Matemáticas.

A continuación hablaremos con mayor detalle de la técnica seleccionada para este proyecto.

### 2.1.1 La Entrevista.

- **Definición.-** Una entrevista para la recopilación de información es una conversación dirigida con un propósito específico, que se basa en un formato de preguntas y respuestas. En la entrevista desea conocer tanto las opiniones como los sentimientos del entrevistado, acerca del estado actual de los sistemas, sus metas personales, de la organización y de los procedimientos informales, sobre todo de busca la opinión de las persona entrevistada. Las opiniones pueden ser mas importantes y reveladoras que los mismos hechos.
- **Planeación de la Entrevista.-** Para visualizar mejor este punto mostramos un esquema grafico de sus pasos que deben seguirse figura 2.1.





**Figura 2.1.** Pasos para planeación de la entrevista.

- **Lectura de Antecedentes.**

Consultar y comprender el mayor número posible de antecedentes de los entrevistados y de la organización. Estos se obtienen frecuentemente de manera sencilla por medio de un informe que haya hecho nuestro entrevistado, Esto nos ayudará a fijarnos en el tipo de lenguaje que los miembros de la organización utilizan para describirse y describir a la organización. Esto permitirá elaborar un vocabulario, que eventualmente, nos capacitará para redactar las preguntas de la



entrevista, de manera tal que estas sean comprendidas por el entrevistado.

- **Establecimientos de los Objetivos de la Entrevista.**

Establecer los objetivos de la entrevista con base en los antecedentes que consultamos y en nuestra experiencia particular. Debe haber entre cuatro y seis aspectos fundamentales referentes al procesamiento de la información y al proceso de la toma de decisiones sobre los cuales se hacen preguntas. Estos incluyen: las fuentes de información, de los formatos de la información, la frecuencia de la toma de decisiones, la calidad de la información y el estilo de la toma de decisiones.

- **Selección de los Entrevistados.**

Cuando se decide a quien entrevistar, esto debe incluir a gente clave de todos los niveles del sistema. Es importante mantener un equilibrio de tal manera que se cubran tantas necesidades de los usuarios como sea posible.

- **Preparación del Entrevistado.**

Preparar a la persona que entrevistaremos, avisándole con tiempo para que pueda pensar acerca de la entrevista. Reserve



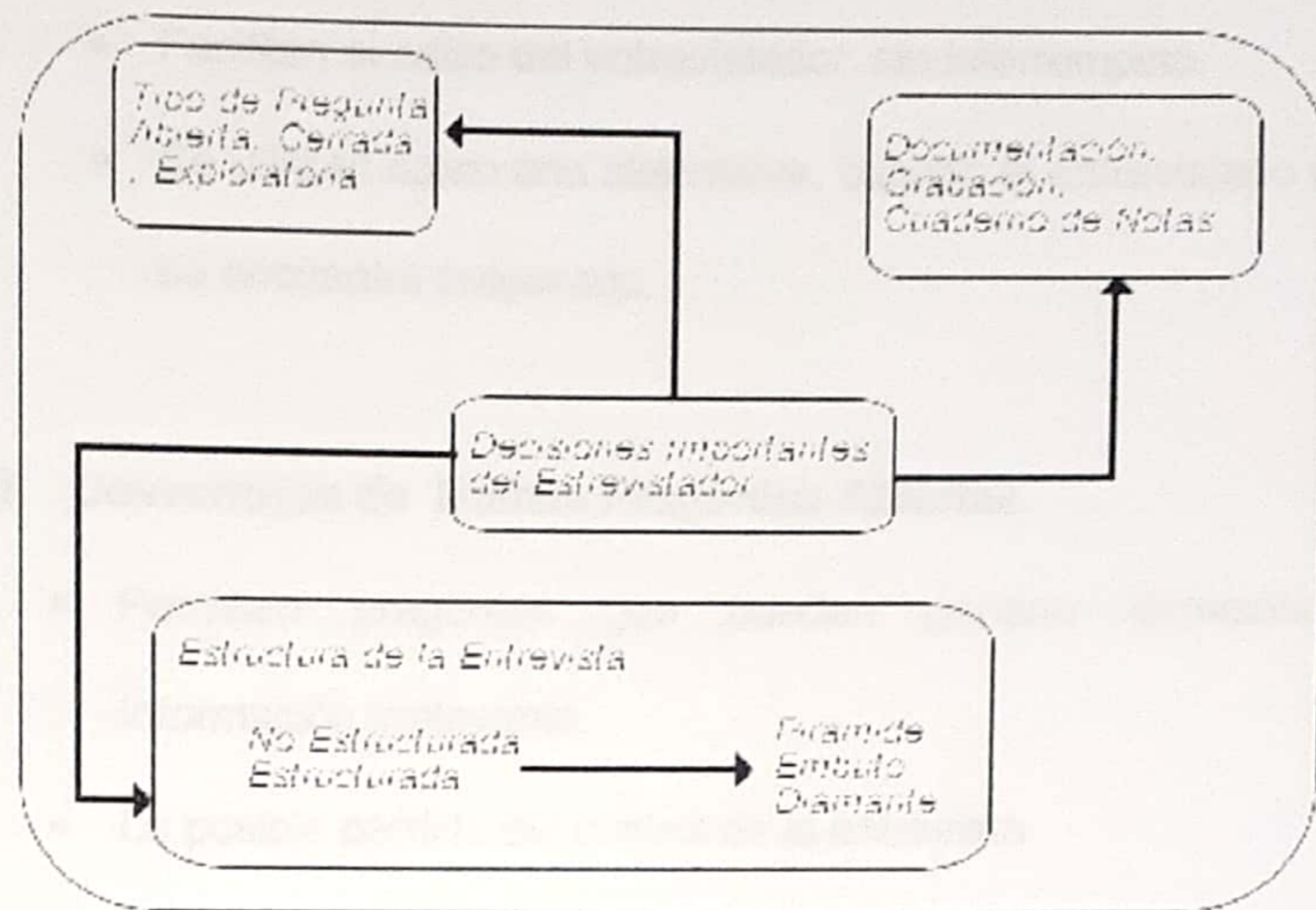
un tiempo especial para sus llamadas y reuniones. Las entrevistas fluctuaran entre 45 minutos y una hora debido al poco tiempo disponible por los directivos. Si la entrevista se extendiera probablemente el entrevistado lamentara nuestra visita.

- **Selección del Tipo y Estructura de Preguntas.**

Redactar las preguntas que cubran los aspectos fundamentales de la toma de decisiones, detectados al plantear los objetivos de la entrevista. Existen dos tipos básicos de preguntas son las abiertas y las cerradas. Cada tipo de pregunta puede lograr algo diferente de las otras y cada una tiene sus ventajas y sus inconvenientes.

Es posible estructurar su entrevista bajo tres patrones diferentes: el de estructura pirámide, de estructura de embudo y la estructura de diamante. Cada uno de ellos es adecuado para condiciones particulares.





**Figura 2.2** Decisiones que hace el analista de sistemas antes de la entrevista.

### 2.1.2. Ventajas de Utilizar Preguntas Abiertas.

- Simplifican las cosas para el entrevistado.
- Permiten al entrevistador, seleccionar el vocabulario del entrevistado, lo que refleja su educación, valores y creencias.
- Proporcionan una gran riqueza de detalle.
- Revelan nuevas alternativas sobre preguntas no consideradas.
- Hacen mas interesante la entrevista.
- Permiten una mayor espontaneidad.



- Facilitan el estilo del entrevistador, sin interrumpirlo.
- Se utilizan como una alternativa, cuando el entrevistado no se encuentra preparado.

### **2.1.3. Desventajas de Utilizar Preguntas Abiertas.**

- Permiten preguntas que pueden generar demasiada información irrelevante.
- La posible pérdida del control de la entrevista.
- Permiten respuestas que pueden llevar demasiado tiempo para la cantidad de información que aportan.
- Pueden dar la apariencia de que el entrevistador no se preparó.
- La posible apariencia de que el entrevistador se encuentra en una expedición, sin objetivos reales de la entrevista.

### **2.1.4. Ventajas de Utilizar Preguntas Cerradas.**

- Ahorran tiempo.
- Facilitan la comparación entre entrevistas.
- Llegan al punto de interés.
- Mantienen el control de la entrevista.
- Cubren rápidamente diversos aspectos.
- Obtienen datos de relevancia.



### **2.1.5. Desventajas de Utilizar Preguntas Cerradas.**

- Aburren al entrevistado.
- Pierden riqueza del detalle (al plantear al entrevistado un marco de referencia).
- Se pueden perder ideas centrales por el punto anterior.
- No favorecen un clima de armonía entre el entrevistado y el entrevistador.

## **2.2 Entrevistas a Directivos del ICM.**

Esta actividad se la desarrollo en los lugares de trabajo de los directivos, lo fundamental es tener una idea de cuales son las actividades que se desarrollan al inicio del termino académico y al final del mismo.

### **2.2.1 Requerimientos de Información Académica.**

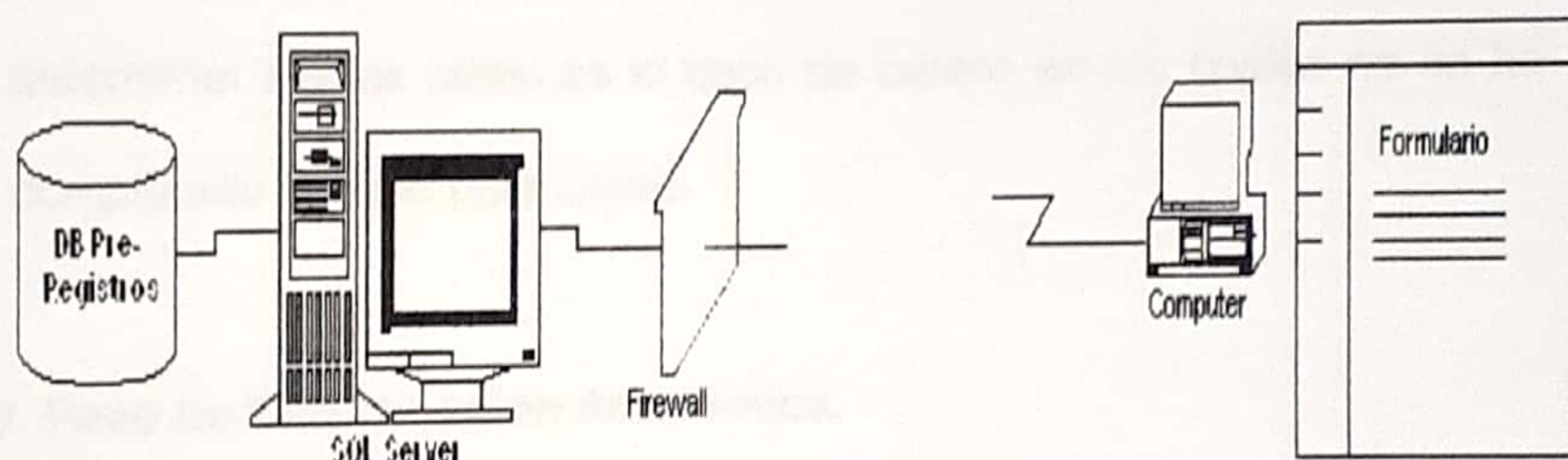
Los requerimientos de información previa a la toma de decisiones, para lograr un mayor entendimiento del problema:

- Fase de Pre-Registro
- Fase de coordinación Académica



### 2.2.1.1 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 2.3.



**Figura 2.3** Esquema para Pre-registro.

Lo que se desea es tener una forma ( formulario de ingreso ) para que este sirva de fuente de información para la base de datos de Pre-registros, la misma que se la utilizaría en el proceso de planificación del termino académico, ya sea para hacer las estimaciones correspondientes al número de cursos disponibles, número de estudiantes en el semestre, preferencia de profesores, entre otros.



Los campos para este formulario se muestran en la tabla 1.

Campo	Descripción
Nombres	Nombres del estudiante
Apellidos	Apellidos del estudiante
Dirección	Lugar donde reside actualmente
Teléfonos	Teléfonos del estudiante
E-mail	Dirección electrónica del estudiante
Numero de matricula	Matricula del estudiante
Factor P	Factor socioeconómico
Código de carrera	Código de la carrera
Materias a tomar	Posibles materias a tomar por el estudiante
Horario a elegir	Posibles horario de clases
Profesor	Profesor con quien desearía tomar la materia

**Tabla 1.**

Con esta información el sistema tomaría esta base de datos para determinar alertas como es el caso de cursos en los cuales no se ha completado el cupo para abrirlo.

#### **2.2.1.2. 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.

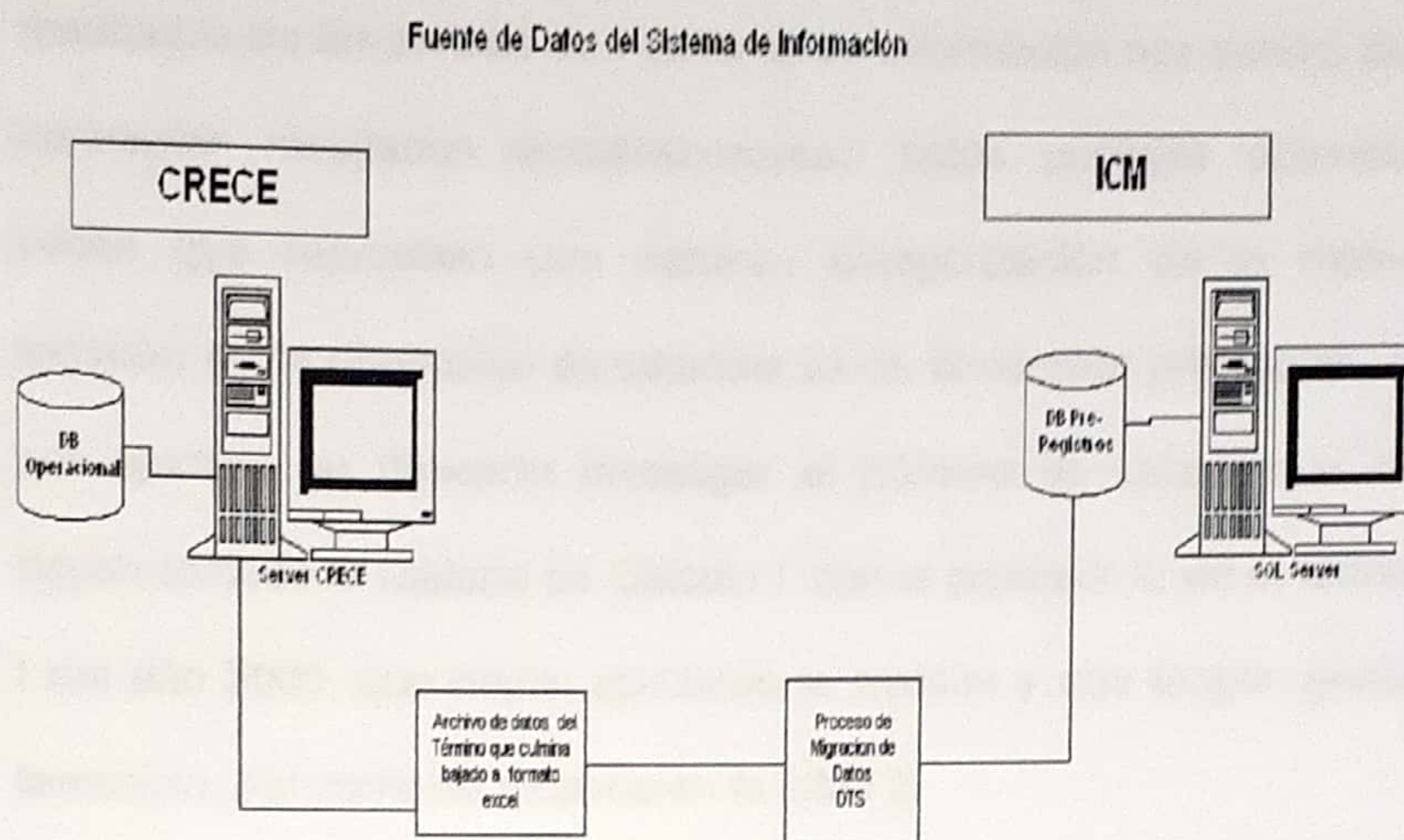
En esta entrevista se hablo mucho de la necesidad de tener un repositorio de datos que contenga información personal y académica de los estudiantes, dicha información debería estar ordenada y



clasificada de tal manera que sea de fácil acceso para las estaciones de trabajo de los directivos, 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.

Los resultados se mostraran a los usuarios en un formato amigable, de fácil comprensión. Las salidas del sistema de información serán básicamente consultas a nuestro Data Warehouse, el mismo que tendrá como fuente directa de alimentación los datos que proporcione el CRECE (Centro de Registros Calificaciones y Estadísticas) , Esta entidad deberá bajar a formato (por ejemplo Excel) información personal y académica de los estudiantes, esto se lo hará un vez que culmine el termino académico, dicha información será cargada al data warehouse a través de un proceso de migración a las bases de datos histórica de nuestro sistema de información, como se muestra en la figura 2.4





**Figura 2.6.** Esquema de carga de datos al Data Warehouse.

### 2.2.3. 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.

#### 2.2.3.1. Consulta a nivel de Profesores.

En este nivel de consulta se requiere una serie de resultados en donde se tenga como parámetro el profesor que dicta la materia, los



resultados de las salidas de l sistema de información nos servirá para contrastar resultados académicos(sean estos puntajes obtenidos, veces que reprobaban una materia, categorización de la materia tomada) de la dimensión estudiantes vs. la dimensión profesores, así por ejemplo se desearía investigar el número de estudiantes que hayan tomado la materia de Calculo I, con el profesor X, en el término I del año 2000, que hayan aprobado la materia y que tengan género femenino. Así como se muestra en la tabla 2.

	Materias	Tiempo	Estado Materia	Género	Medida
Profesor X	Calculo I	2000-Term. I	Aprobada	Femenino	# de Estudiantes
Profesor Y	Calculo II	2000-Term. II	Aprobada	Femenino	# de Estudiantes
Profesor Z	Calculo III	2000-Term. III	Aprobada	Femenino	# de Estudiantes

**Tabla 2.**

Del cuadro podemos observar una nueva propuesta de consulta de base datos en donde se mezclan mas de dos dimensiones con la finalidad de generar nuevo conocimiento, que nos ayude a visualizar de manera rápida y confiable algún dato de interés para las futuras planificaciones de términos académicos.



### 2.2.3.2. Consulta a nivel de Estudiantes.

Para el caso de los estudiantes registrados en cada término académico, se preguntara al sistema de información por los estudiantes que se han registrado en una determinada materia, cuantas veces ha tomada dicha materia, factor p del estudiante, materias aprobadas hasta el momento, lugar de residencia del estudiante, año de ingreso a la ESPOL.

Para esquematizar este tipo de consultas con muchas variables hemos procedido de igual manera que el nivel anterior. Así por ejemplo se desearía investigar el número de estudiantes que reprobaron una materia determinada en el término 2000 con el profesor X y que sea de género masculino como se muestra en la tabla 3.

	Profesor	Tiempo	Estado Materia	Género	Medida
Materia X	profesor X	2000-Term. I	Aprobada	Masculino	# de Estudiantes
Materia Y	profesor Y	2000-Term. II	Aprobada	Masculino	# de Estudiantes
Materia Z	profesor Z	2000-Term. III	Aprobada	Masculino	# de Estudiantes

**Tabla 3.**



Del cuadro podemos observar la intersección de las variables materias, profesores, tiempo, género, filtradas a través del número de estudiantes registrados en el término académico consultado.

Así mismo debe existir la posibilidad de seleccionar nuevas variables y nuevas formas de medición, esto debe ser en forma dinámica e interactiva.

### 2.2.3.3. Consulta a nivel de Materias.

Este nivel es considerado uno de los críticos del sistema pues ahí se encuentran registrados los puntajes de aprobación, estado de la materia, las veces que se ha tomado una materia, entre otras. Lo que se quiere es determinar si existe algún dato importante que resulte de combinar la variable materias, con las variables género del estudiante y profesor.

Por ejemplo se desearía investigar el numero de estudiantes que hayan tomado la materia X, con el profesor X, en el término I del año 2000, que hayan aprobado la materia y que tengan genero femenino.

Así como se muestra en la tabla 4.

	Estado	Tiempo	Vez Tomada	Género	Medida	Profesor
Materia X	Aprobada	2000-Term. I	1era. Vez	Femenino	# de Estudiantes	ProfesorX
Materia Y	Reprobada	2000-Term. II	2era. Vez	Femenino	# de Estudiantes	ProfesorY
Materia Z	Retirado	2000-Term. III	3era. Vez	Femenino	# de Estudiantes	ProfesorZ

**Tabla 4.**



La tabla de anterior presenta un esquema que se ajusta a las necesidades de información para este nivel, donde podemos observar que rápidamente se identifican factores críticos como las veces que se reprueba una materia, veces tomada una materia, profesor que dicto la materia. Esta intersección de resultados nos conducirá a generar reportes de rendimiento estudiantil por términos académicos, por año. Todo en un ambiente visual agradable y de fácil comprensión.

#### **2.2.3.4. Consulta a nivel de Registros de Materias.**

A Nivel de planificación académica se requiere básicamente cuantificar históricamente cual ha sido el comportamiento de los registros de materias a través de los términos académicos, a continuación se detalla las consultas para este nivel:

- Número de estudiantes registrados en fechas ordinarias.
- Número de estudiantes registrados en fechas extraordinarias.
- Número de estudiantes registrados en cada término.
- Número de estudiantes que tienen menos de 35 materias.
- Número de estudiantes registrados en cada una de las materias (previo al inicio del semestre).



Existen niveles de consultas en los cuales se procederá a calcular índices que nos indiquen nuevas oportunidades de mejora en la planificación, entre estos índices tenemos los siguiente:

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



### 2.3. Tareas Básicas de un Sistema de Información.

- Consulta exploratoria de datos históricos en las carreras de Ingeniería en Estadística Informática, Auditoría y Control de Gestión.
- Centralizar la información en un gran repositorio de datos, debidamente clasificado y desagregado en las dimensiones.
- Que tenga un nivel de acceso a la información fácil para los usuarios, es decir que el usuario del sistema de información pueda llevar los resultados de la consulta a herramientas de escritorio como Excel para un análisis mas detallado.
- Obtener índices estudiantiles que midan el desempeño estudiantil.
  - Visualización de consultas en un ambiente grafico y de fácil entendimiento.
  - Impresión y emisión de reporte reducido.
  - Decisiones con mayor información y toma de decisiones más rápida.
  - Un sistema de información que puedan acceder a una riqueza de información multidimensional, presentado coherentemente como una fuente única confiable y disponible a ellos por medio de sus estaciones de trabajo.



- Un sistema de información que enriquezca las capacidades del usuario autosuficiente y hace lo factible para ofrecer nuevos servicios a los usuarios, sin interferir con las aplicaciones cotidianas de producción.
- Que sirva como fuente de información histórica y punto de partida para futuras investigaciones o proyecto de análisis de datos mas avanzados como puede ser la minería de datos (“Data Mining”).

#### **2.4. Limitaciones del Sistema de Información.**

Como en todo sistema de información los usuarios juzgan al mismo por sus resultados o las salidas de información. A continuación se detallamos algunas limitaciones que tendría nuestro sistema de información.

- La fuente de información directa para nuestro data warehouse es un archivo de Excel que se genera al finalizar los términos



académicos en el CRECE, la manipulación de este archivo podría ocasionar que se tenga una data sucia.

- El sistema de información solo mostrará salidas a manera de consulta, no hará investigaciones exploratorias mas complejas.
- El ambiente que correrá el sistema de información es web, debido a que su desarrollo se hará mediante el empleo de herramientas de Microsoft solo estará disponible para las versiones de Internet Explorer 6.0 o superior.



# CAPÍTULO III

## 3. PROPUESTA INTEGRAL DEL SISTEMA DE INFORMACION

En esta sección se pondrán en consideración diferentes tecnologías de información, que se deben tomar en cuenta al momento de diseñar el sistema de información.

### 3.1 COMPONENTES DEL SISTEMA DE INFORMACIÓN

En esta parte se tomará en cuenta los componentes básicos de un sistema de información:

- Componente de Hardware
- Componente de Software
- Componente de datos
- Componente de Metodología de uso.



A continuación hablaremos con mayor detalle de cada componente así como su contribución al proyecto:

### 3.1.1 Componente de Hardware y Software

La información histórica de los estudiantes, debemos almacenarla en giga 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.

Para la parte de hardware se hizo una revisión de los servidores que cuenta el instituto de matemáticas, actualmente estos están equipados con la tecnología que se necesita para desarrollar el sistema de información. Por lo cual los requerimientos de hardware justificado con lo que se tiene al momento.

### 3.1.2 Componente de Datos

Dentro del esquema elegido ( Data Warehousing ) para el sistema de información, existen niveles diferentes de esquemmatización y detalle con respecto a los datos.

- Detalle de Datos Actuales.

En gran parte, el interés mas importante radica en el detalle de los datos actuales, debido a que refleja las



ocurrencias mas recientes, las cuales son de gran interés, es voluminoso, ya que se almacena al mas bajo nivel de granularidad (grado de dependencia en un recurso), casi siempre se almacena en el disco duro, el cuales de facil acceso aunque su administración sea costosa y compleja

- Detalle de datos antiguos.

La data antigua que se encuentra almacenada en los servidores operacionales de la institución. No es frecuentemente accesada y se almacena a un nivel de detalle.

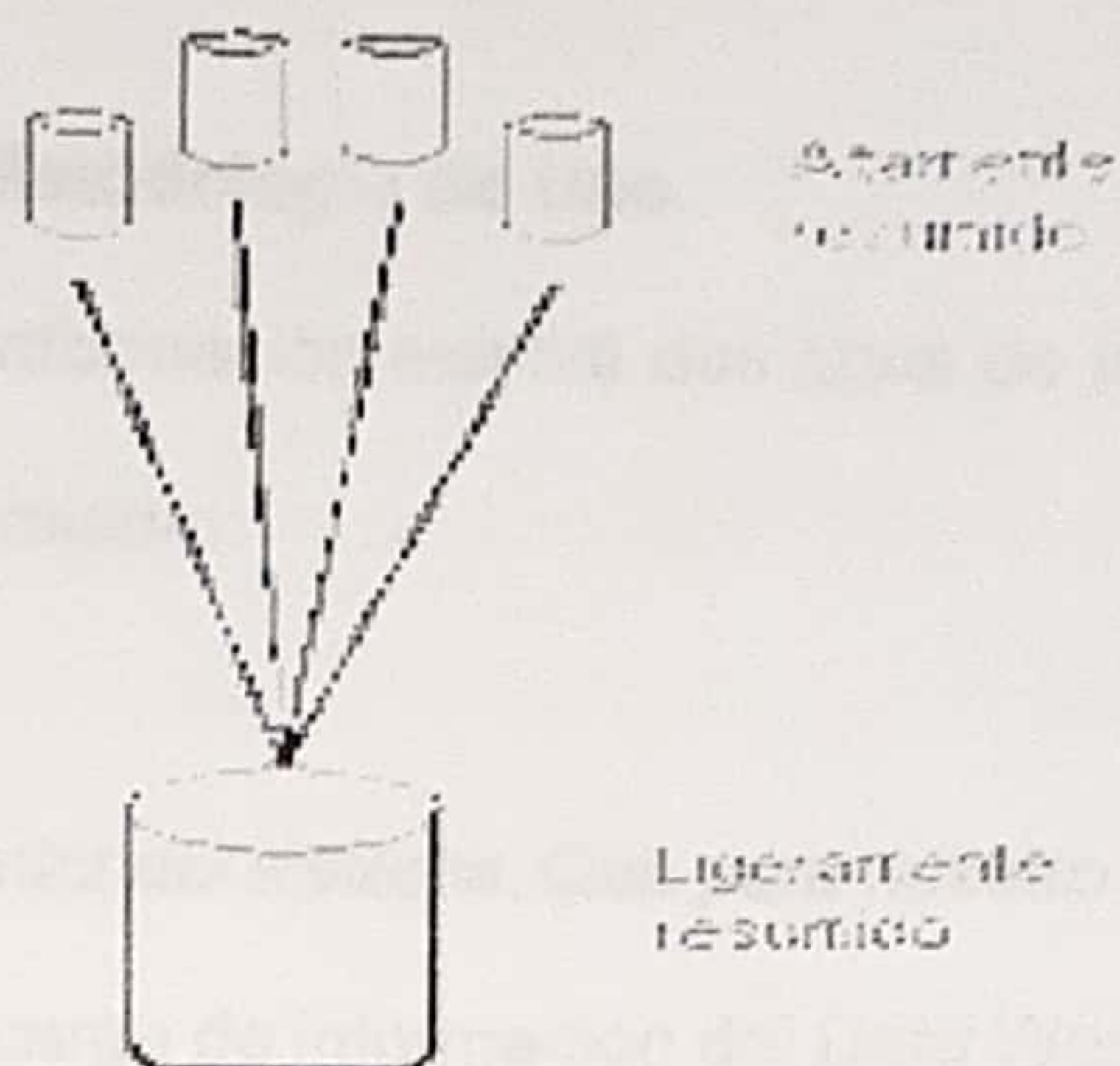
- La data ligeramente resumida.

es aquella que proviene desde un bajo nivel de detalle actual.

- La data completamente resumida.

El siguiente nivel de datos encontrado en un sistema de información orientado a un estructura de Data Warehousing es el de los datos completamente resumidos. Estos datos son compactos y fácilmente accesibles.





- **Metadato.**

El componente final de un sistema de información orientado a un data warehousing es el de metadato, de muchas maneras la metadata se sitúa en una dimension diferente al de los otros datos, debido a que su contenido no es tomado directamente desde el ambiente operacional. La metadata juego un rol especial muy importante en el sistema de información y es usado como:

Un directorio para ayudar al analista a ubicar los contenidos del sistema de información.

Una guía para el mapeo de datos de cómo se transforma, del ambiente operacional al de data warehousing.



### 3.1.3 Componente 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 cada termino académico, esta información será sometida a un proceso de transformación de datos o DTS antes de residir en el Data Warehouse, además será el responsable de la creación actualización o eliminación de nuevas cuentas de usuarios operadores del sistema.

Usuarios operadores, para nuestro caso serán los directivos del instituto de ciencias matemáticas encargados de la planificación de cada termino académico, dichos usuarios tendrán acceso a la visualización de consultas generadas, a partir de la selección de atributos ( dimensiones ).

Limitándose a consultas y no podrán modificar la información, debido a que el sistema es una herramienta de apoyo a la toma de decisiones.



### **3.1.4 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.

#### **3.1.4.1. Aspectos Funcionales :**

- Para acceder al sistema de información se le solicita al usuario que ingrese un nombre de usuario y contraseña para su autenticación, permitiendo de esta manera restringir el uso de esta herramienta solo a los usuarios autorizados (administrador o directivos del ICM). En caso de ser un directivo del ICM y no tiene un nombre de usuario y clave se le indicara la dirección de e-mail del administrador del sistema para que se contacte con el.
- Esta aplicación debe ser capaz de personalizarse para cada usuario registrado en la base datos del ICM, Si es un usuario operador debe de cargársele la distintas dimensiones o atributos a consultar en el Data Warehousing, el usuario seleccionara de



manera interactiva las dimensiones que posteriormente serán los parámetros de filtración de la consulta multidimensional.

#### **3.1.4.2. 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, existen elementos adicionales que también se tomo en cuenta para la construcción de la interfase, aspectos gráficos como animaciones, texturas, son incorporados en el sistema. El factor peso de las páginas en kb. en este caso no juego un papel preponderante debido a que el sistema de información trabajara en la Intranet del instituto de ciencias matemáticas.
- El uso de una interfase de usuario por medio de un Browser se debe a:  
Facilidad de visualización de resultados especialmente los gráficos.



No se necesita instalar componentes adicionales en la maquina del cliente (usuarios del sistema). Bastara que cada usuario tenga un Browser ( Internet explorer o Netscape, en sus 2 versiones anteriores).

#### **3.1.4.3. 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.
- El sistema de información será implementado de acuerdo al siguiente esquema:

Los datos corresponde a un base de datos y el Cubo de datos esta diseñada en Microsoft SQL Server 2000 y los Servicios OLAP respectivamente, utilizando ADO, ADO MD, MDX como objeto de datos y lenguaje de consulta multidimensional, en la capa de servicios web utilizará ASP sobre el servidor IIS para generar las páginas HTML y ASP.



- Para la elaboración de los gráficos que acompañaran a las salidas de consultas, se hará uso de los Objetos Office web componentes instalados en el servidor web.

### **3.2 PUNTOS IMPORTANTES PARA LA IMPLEMENTACION DEL SISTEMA.**

En la presente sección mencionaremos todas las herramientas que se utilizaron para desarrollar el sistema de información empezando por la plataforma, tecnología usada hasta llegar a las herramientas de desarrollo.

#### **3.2.1 Plataforma.**

En lo que respecta a este punto se tomo en consideración la plataforma desarrollada por Microsoft. Haremos uso de los sistemas operativos Windows junto con las herramientas de Gestión de Base de datos, Servicios OLAP, Servicios de IIS (Internet Information Server ). Esta plataforma provee de los recursos y capacidades de software para desarrollar sin problema nuestro sistema de información, permitiendo una integración modular favoreciendo la escalabilidad y robustez del sistema.



### 3.2.2 Servicios OLAP y Almacén de Datos.

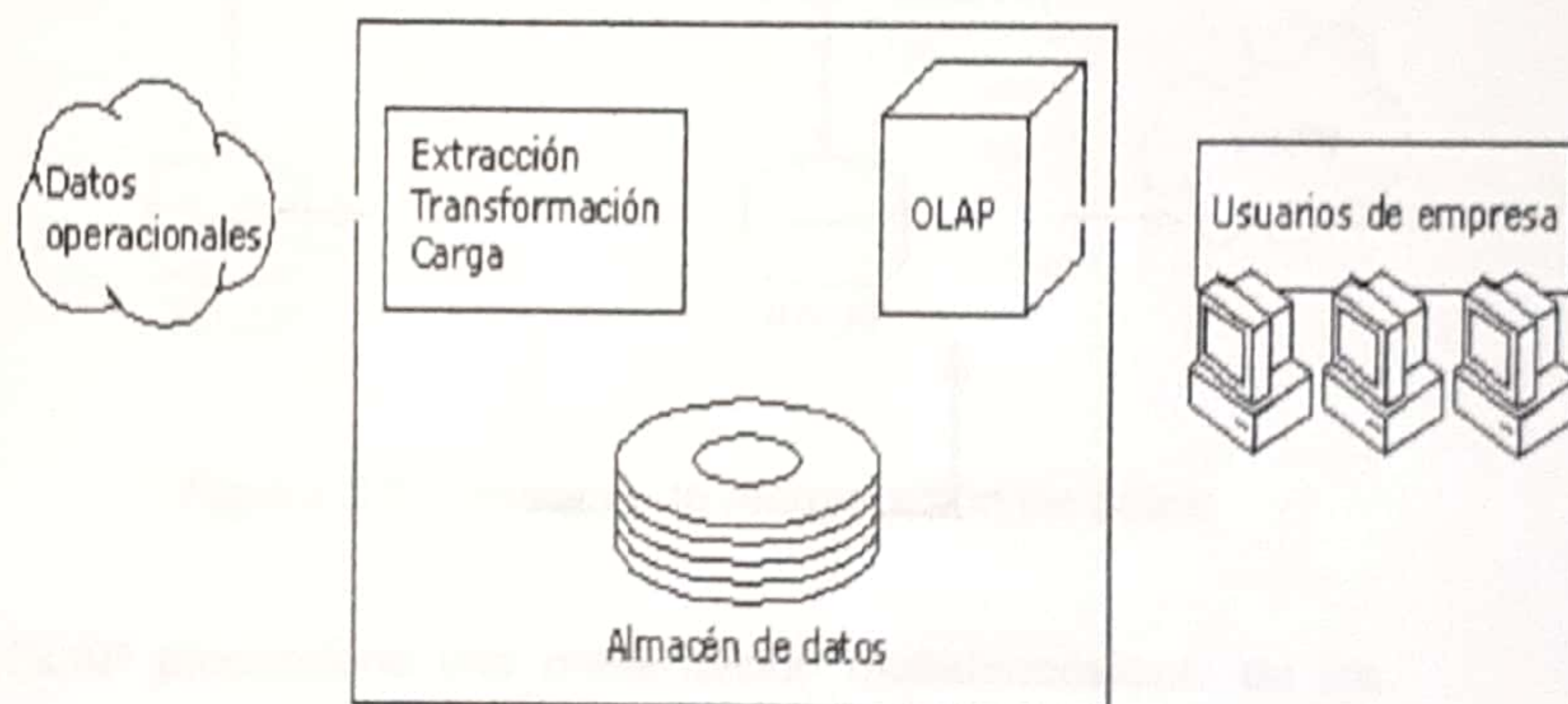
Aunque en ocasiones se utilizan indistintamente, los términos almacén de datos y proceso analítico en línea ( OLAP ) se aplican a diferentes componentes los sistemas conocidos como sistemas de ayuda a la toma de decisiones o sistemas de inteligencia empresarial. Los componentes de estos tipos de sistemas incluyen bases de datos y aplicaciones que proporcionan las herramientas que necesitan los analistas para tomar decisiones en relación con el soporte técnico de la organización .

Estas bases de datos contiene la información que normalmente, representa el historial empresarial de una organización. Estos datos históricos se utilizan para realizar el análisis que apoyen las decisiones empresariales a diferentes niveles de la organización, desde el diseño estratégico a la evaluación del rendimiento de una unidad determinada de la organización.

La tecnología OLAP permite un uso eficaz de los Data Warehouse para el análisis en línea, lo que proporciona respuesta rápida a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP y las técnicas de agregado de datos organizan y resumen grandes



cantidades de datos para que puedan ser evaluados con rapidez mediante el análisis en línea y las herramientas graficas. La respuesta a una consulta realizada sobre datos históricos a menudo suele conducir a consultas posteriores en las que el analista busca respuestas mas concretas o explorara posibilidades. Los sistemas OLAP proporcionan la velocidad y la flexibilidad necesarias para dar apoyo al análisis en tiempo real.

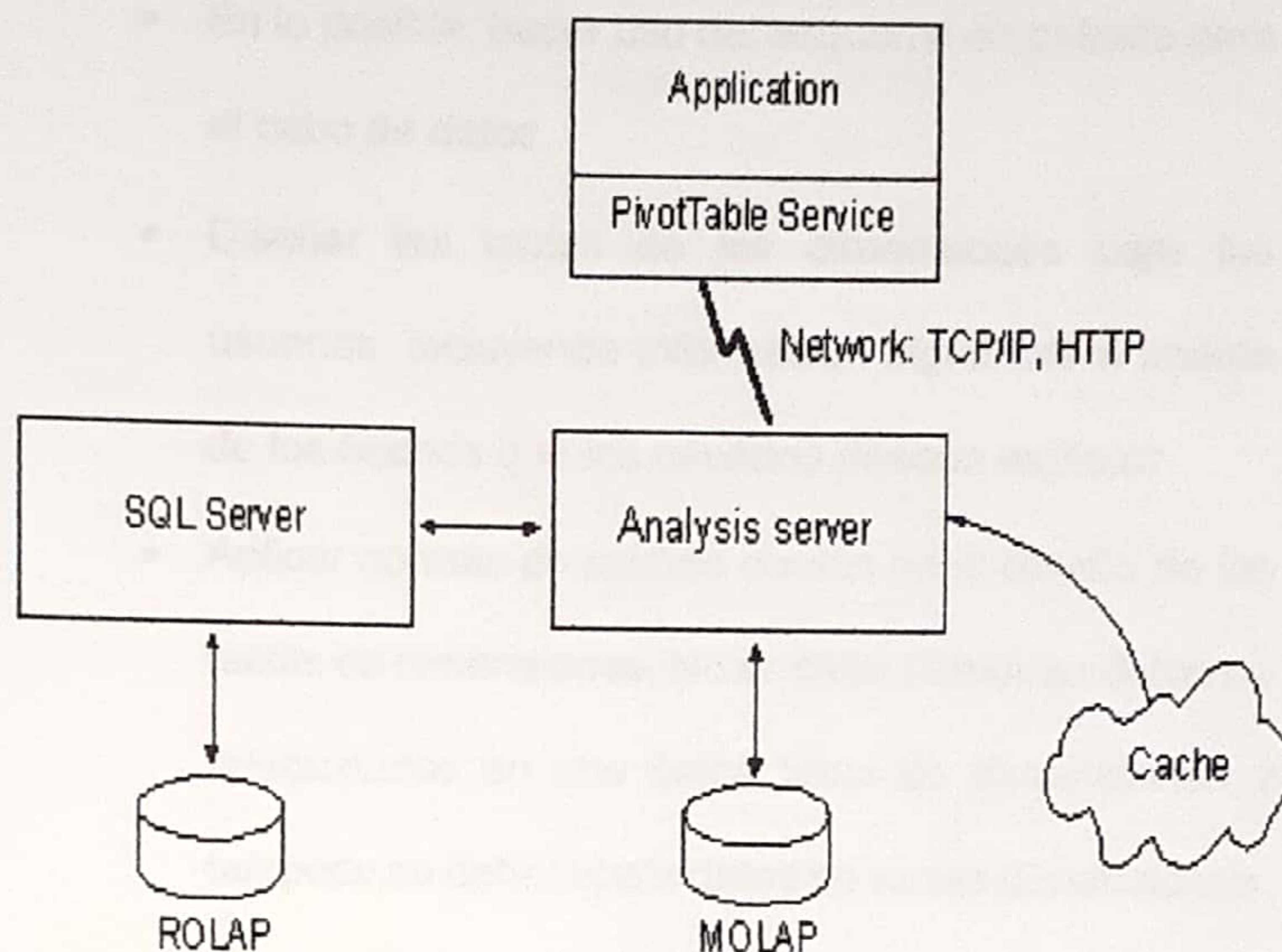


**Figura 3.1.** Almacén de datos y OLAP.

La recuperación de datos de los Servicios OLAP de Microsoft información, puede ser a través de una conexión de red o a través de compartir recursos en el lado del cliente (memoria), los protocolos que se manejan pueden ser TCP/IP o http (Para el presente proyecto se elige el segundo protocolo de red )



continuación se muestra en la figura 3.2 un esquema de recuperación de datos.



**Figura 3.2** Esquema de recuperación de datos

OLAP proporciona una presentación multidimensional de los datos de un almacén de datos, mediante la creación de cubos de datos. El diseño de la estructura del almacén de datos puede afectar a la facilidad con la que se podrán diseñar y construir estos cubos.

Los servicios OLAP se basan en que los datos proporcionados por el almacén de datos han pasado por un proceso de limpieza (es decir son exactos, estables y tienen integridad referencial).



El uso de esta tecnología requiere que se tenga en cuenta los siguientes factores de diseño.

- En lo posible, hacer uso del esquema en estrella para el cubo de datos.
- Diseñar las tablas de las dimensiones para los usuarios, incluyendo información significativa acerca de los hechos que los usuarios desean explorar.
- Aplicar normas de sentido común en el diseño de las tablas de dimensiones, No se debe combinar datos no relacionados en una única tabla de dimensiones y tampoco se debe repetir datos en varias dimensiones.
- No resumir en exceso la tablas de hechos, conservando el nivel de granularidad mas fino que necesite el usuario para tener acceso y almacenar todos los registros de la tabla de hechos en el mismo nivel de detalle.
- No se debe crear tablas auxiliares de datos resumidos.
- Se debe crear índices en los campos de claves, para cada tabla de dimensiones, cree un índice en su columna de claves. Para cada tabla de hechos, cree un único índice en la combinación de columnas que



contienen las claves externas de las tablas de dimensiones asociadas con la tabla de hechos.

- Se debe asegurar la integridad referencial de los datos, es importante que todos los hechos se encuentren representados en todas las tablas de dimensiones.
- Se debe diseñar una estrategia de actualización de datos, cuando se agreguen o modifiquen los datos de un almacén, se deberán actualizar los cubos que se hayan creado a partir de los datos anteriores para que los usuarios puedan disponer de estos nuevos datos.

### **3.3 OLAP en el Web.**

Las aplicaciones clientes pueden comunicarse con los servicios OLAP a través de Microsoft IIS, mediante los protocolos http o https. Esto trae consigo el uso de paginas que accedan a bases de datos multidimensionales, a continuación mostramos algunas tecnologías que están disponible en este campo.

CGI (Common Gateway Interface) es un mecanismo que permite a un servidor Web ejecutar un programa o script sobre el servidor y enviarle como salida a un Web browser. Esta fue la primera solución al desarrollo de aplicaciones Web, que proveía



resultados generados dinámicamente. La desventaja de éstos, es que cada requerimiento inicia un nuevo proceso de lado del servidor, consumiendo los recursos del mismo y disminuyendo notablemente el desempeño.

ISAPI (Internet Server Application Programming Interface) es un conjunto de funciones para Servidores de Internet. El modelo ISAPI fue desarrollado para ser una alternativa de alto rendimiento al CGI. Este modelo provee un número de ventajas sobre el modelo CGI, incluyendo la disminución de contenido innecesario (overhead), rápida carga y mejor escalabilidad.

Una alternativa a CGI e ISAPI, es la utilización de tecnología que implementa scripts para crear páginas interactivas como parte de una aplicación Web. Esta tecnología permite separar la lógica de programación del diseño de la página a través del uso de componentes que son llamados por la propia página, lo cual hace que el desarrollo, la implementación y el mantenimiento resulten mas rápidos y fáciles.

Dentro de esta tecnología se encuentran:

- Páginas Active Server (ASP) de Microsoft es un entorno de secuencia de comandos del servidor que puede utilizar para crear y ejecutar dinámicas e interactivos de servidor



web. Con ASP puede combinar paginas HTML, comando de secuencias y componentes COM para crear paginas web interactivos o eficaces aplicaciones web, que son fáciles de desarrollar y modificar, los lenguajes de programación que se usan son VBScript, Jscript.

- Hypertext Preprocessor (PHP) es una mezcla de varios lenguajes, principalmente utiliza la sintaxis que la mayoría de los programadores conocen de C, pero es muy diferente. PHP es interpretado. PHP reconoce diferentes tipos de variables, pero no comprueba un tipo fijo. PHP reconoce las clases pero no los tipos estructurados.
- La tecnología JSP (Java Server Pages ) esta basada en la tecnología de servlets Java; de hecho, cuando se compilan los script JSP, se crean los servlets Java. Las capacidades de presentación de los servlets son limitadas, de modo que JSP utiliza etiquetas de tipo XML y scriptlets escritos en Java para separar la recuperación de datos de la lógica de presentación. y que generan código HTML en el lado del cliente, los scripts de JSP tienen como lenguaje de programación a Java.



La ventaja de utilizar scripts que se ejecuten del lado del servidor es que disminuye el tiempo de desarrollo de aplicaciones web complejas, lo que representa una gran ventaja en el mundo moderno, ya que las aplicaciones Web necesitan ser actualizadas rápidamente para permanecer competitivas.

### **3.3.1 Cubo de Datos.**

Los cubos son los principales objetos del proceso analítico en línea ( OLAP ), un cubo es un conjunto de datos que normalmente se construye a partir de un subconjunto de un almacén de datos y se organiza y resume en una estructura multidimensional definida por un conjunto de dimensiones y medidas.

Un cubo proporciona un mecanismo fácil de usar para consultar datos con tiempos de respuesta rápidos y uniformes. Los usuarios finales usan aplicaciones cliente para conectarse con los Servicios Olap y consultar los cubos del servidor. En la mayoría de las aplicaciones cliente, los usuarios finales ejecutan una consulta en un cubo mediante la manipulación de los controles de la interfaz de usuario, que determinan el contenido de la consulta.



Esto evita la necesidad de escribir consultas basadas en lenguaje. Los datos de resúmenes precalculados denominados agregaciones proporcionan el mecanismo para obtener tiempos de respuesta rápidos y uniformes a las consultas. Las agregaciones se crean para un cubo antes de que los usuarios finales puedan obtener acceso al mismo. Los resultados de una consulta se recuperan de las agregaciones, de los datos de origen del cubo del almacén de datos, de una copia de estos datos en los Servicios OLAP, de la caché del cliente o de una combinación de estos orígenes.

Cada cubo tiene un esquema, que es el conjunto de tablas combinadas del almacén de datos del que el cubo obtiene sus datos de origen. La tabla central del esquema es la tabla de hechos, el origen de las medidas del cubo. Las demás tablas son las tablas de dimensiones, los orígenes de las dimensiones del cubo.



### **3.3.2 Servidor Web (IIS)**

Internet Information Server (IIS) es un conector (gateway) de alto desempeño para la capa de servicios de presentación que hace posible desarrollar aplicaciones basadas en el Web.

Las Páginas Activas de Servidor (ASP), forman el ambiente scripting de código interpretable utilizado por el Internet Information Server para crear y ejecutar aplicaciones interactivas y dinámicas que al combinarse con DHTML y componentes de código compilado, permiten crear poderosas aplicaciones basadas en Web.

Haciendo uso de este ambiente, IIS permite adoptar el paradigma transaccional para las aplicaciones basadas en Internet. Las transacciones son la parte de la infraestructura que hace posible ejecutar verdaderas aplicaciones de negocios con rápido desarrollo, escalabilidad sencilla y confiabilidad.

### **3.3.3 Otras tecnologías en aplicaciones Web**

Además de las tecnologías presentadas anteriormente, en la actualidad están a disposición otras tecnologías para el



desarrollo de sitios Web de contenido dinámico, entre las cuales podemos citar las siguientes:

El **PHP** (Hypertext Preprocessor) es una tecnología basada en scripts embebidos en páginas HTML y ejecutados en el servidor. La mayor parte de la sintaxis de los scripts ha sido tomada de C, Java y Perl con algunas características propias.

Al igual que ASP, la meta de esta tecnología es facilitar el desarrollo de aplicaciones para la Internet, diferenciándose en que utiliza C como lenguaje de programación y que está diseñado principalmente para ejecutarse con UNIX.

Una de las características más potentes de PHP, es el soporte a una gran cantidad de bases de datos, entre las que pueden mencionarse: InterBase, MSQL, MySQL, Oracle, Informix, PostgreSQL, entre otras.

PHP también ofrece la integración con varias bibliotecas externas, desde generar documentos PDF hasta analizar código XML (eXtensible Markup Language).

Los servlets representan una nueva tecnología que surge a la limitación existente en los applets desarrollo para el Web en dos vertientes distintas: la imposibilidad de acceder a otro



servidor que no sea el mismo en el que el applet se este ejecutando (generalmente la máquina cliente) y la limitación en el acceso a los servicios de dicha máquina.

Un servlet es una aplicación que se ejecuta en un servidor Web esperando por resolver peticiones efectuadas por los clientes. Mediante los servlets se puede tener acceso a otros servidores y acceder a la información que en ellos haya contenida, por ejemplo una base de datos.

Según todo lo expuesto anteriormente los servlets deben verse como la evolución lógica de los CGI, y que estos últimos no son más que aplicaciones cuyo cometido es resolver peticiones hechas por clientes que adolecían de un número de inconvenientes considerables como el hecho del lenguaje de programación utilizado (generalmente lenguajes interpretados), un rendimiento muy bajo, baja portabilidad, dificultad de comunicación entre CGIs, etc. que se han visto superado por la aparición de los servlets.

La tecnología **JSP** (JavaServer Pages) proporciona especificación y servicios de documentos que combinan elementos del lenguaje de etiquetas estáticas y elementos creados dinámicamente mediante objetos del lenguaje Java.



Las páginas JSP se ejecutan como Servlets Java normalmente en el contexto de un servidor Web, donde generan documentos de respuesta dependientes del contenido usando datos almacenados en bases de datos y en otros objetos de aplicación basados en el servidor.

### **3.3.4 Componentes de Software**

Entre los componentes de software utilizados para llevar a cabo este proyecto, tenemos:

El servidor Web que se contará es Internet Information Server ( IIS ) y los servicios OLAP de Microsoft SQL Server, ambos servidores estarán conectados para responder las consultas de los usuarios, que serán enviadas a través de paginas ASP.

En lo referente al acceso a datos se utilizará el componente ActiveX ADO ( ActiveX Data Object ), ADO MD, MDX como lenguaje de consulta multidimensional, que permite abstraer la manera de conexión con diferentes repositorios de datos, además de formar parte del modelo de programación para IIS, y que han sido diseñados para el desarrollo de aplicaciones Web.



En cuanto a la generación de gráficos que resultan de las consultas multidimensionales, haremos uso de los objetos Office Web Components (OWC) para la visualización de resultados.

### 3.3.5. Herramientas de Desarrollo

Para la funcionalidad del sistema, se utilizarán las siguientes herramientas de desarrollo:

- Para la generación de scripts ASP (VBScript y JavaScript) que contienen los módulos de entrada, interpretación y salida de información, se utilizará Macromedia Dreamweaver MX porque actúa como integrador de los servicios de presentación con los diferentes componentes de los servicios del negocio.
- Para la preparación y presentación de imágenes y animaciones se utilizarán Macromedia Fireworks MX y Flash MX respectivamente.
- Para el almacenamiento de datos se utilizará el DBMS Microsoft SQL Server 2000.
- Para la creación del cubo de datos se utilizará los Servicios OLAP de Microsoft SQL Server 2000





# **CAPÍTULO IV**

## **4. DESARROLLO E IMPLANTACIÓN DEL SISTEMA DE INFORMACIÓN**

En este capítulo abordaremos en detalle el desarrollo del sistema de información con las herramientas que mencionamos en el capítulo anterior.

### **4.1. DISEÑO DEL SISTEMA.**

En esta sección hablaremos en términos generales de la lógica para la identificación de los usuarios y personalización del sistema, de la manera en que el usuario debe seleccionar las filas y columnas para las consultas multidimensionales que son importantes para la toma de decisiones, visualización de estas en el browser , así como las visualización de las tablas de datos con sus respectivos gráficos de barras, líneas.

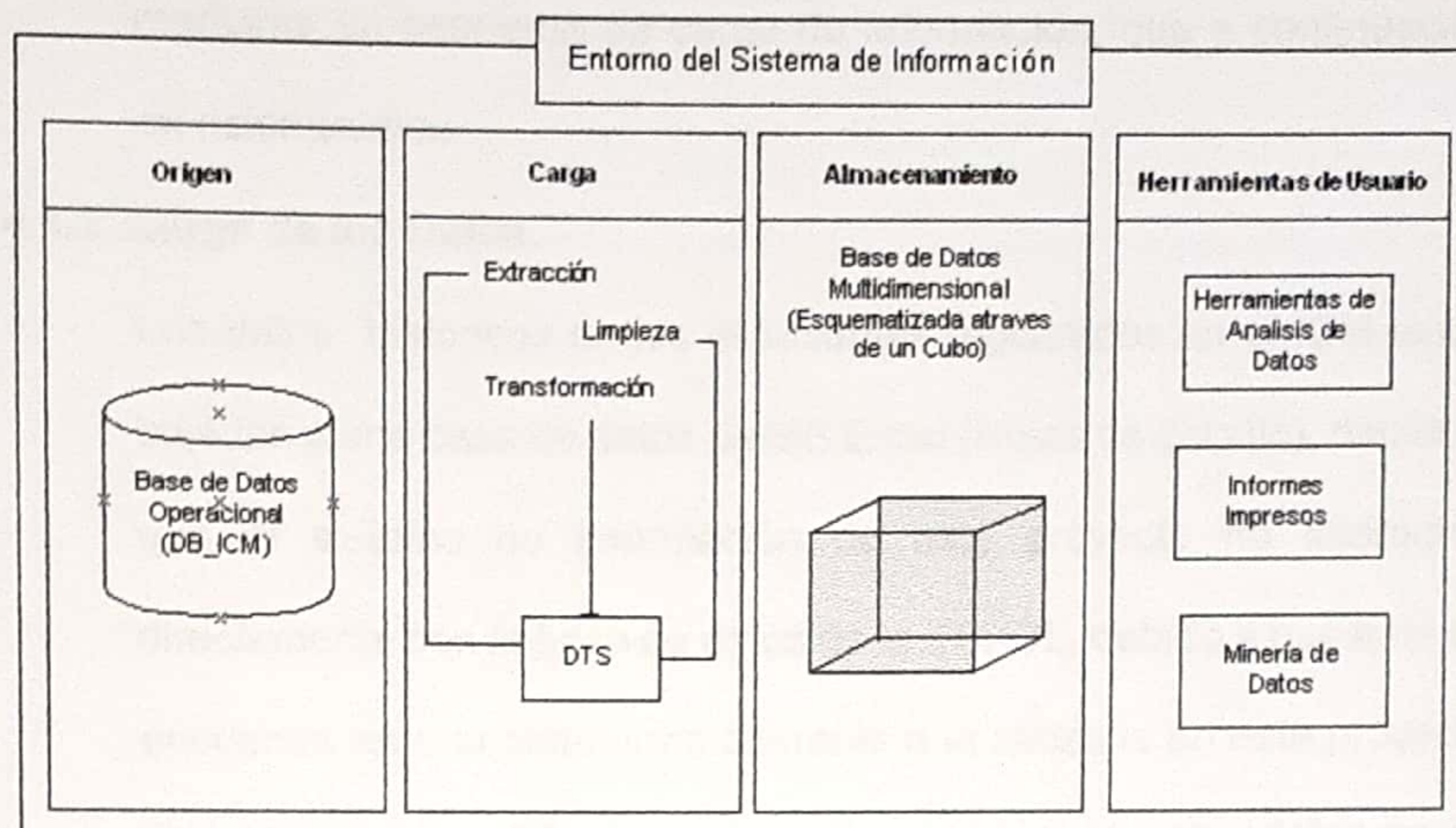


Teniendo claro las necesidades de información por parte de los usuarios, realizaremos la selección de las tablas del repositorio de datos (Data Warehouse), que mas adelante serán las tablas de dimensiones y hechos en nuestro cubo de datos, finalizado esto desarrollaremos los módulos de generación de gráficos para visualización de resultados multidimensionales y diseño de la interfaz de usuario.

Los sistema de soporte a la toma de decisiones ( Decision Support Systems DSS ) junto con los servicios OLAP ( Proceso analítico en Línea ) constituye una herramienta informática que nos permiten generar nuevo conocimiento a través de información histórica que se encuentra almacenada en las bases de datos operaciones de las instituciones. El uso de estas herramientas ayudaran a los directivos del ICM a identificar y resolver problemas en la planificación de términos académicos.

A continuación mostramos un esquema grafico de los componentes de un almacén de datos y los servicios OLAP de Microsoft, arquitectura elegida para este proyecto.





**Figura 4.1.** Entorno del Sistema de Información.

#### 4.1.1. Origen de los Datos.

El origen de los datos del sistema de información se encuentran en una base de datos relacional en MS SQL Server 2000, cuyo esquema se tuvo que diseñar debido a que el instituto de ciencias matemáticas no cuenta con repositorio de datos ( Data Warehouse ) para el proyecto, este repositorio estará almacenado en el servidor del ICM ( Instituto de Ciencias Matemáticas ).

En esta base de datos se tendrá información histórica de los alumnos de las carreras de Ingeniería en Estadística Informática, Auditoría y Control de Gestión, esta información es ingresada a la base de datos



mediante un esquema de carga de información, que a continuación mencionaremos.

#### **4.1.2. Carga de los Datos.**

Los datos históricos de los estudiantes registrados en el ICM serán bajados a una base de datos de MS Excel (Hojas de cálculo), debido a que el sistema de información de este proyecto no interactúa directamente con la base de datos de la ESPOL, debido a que esta se encuentra en una plataforma diferente a la utilizada en este proyecto de investigación. Al finalizar el termino académico estos datos serán cargados al repositorio del ICM mediante un DTS (Proceso de Transformación de Datos) el mismo que brevemente explicaremos.

El proceso de carga de los datos hacia el repositorio del ICM, se lo hará mediante un proceso de transformación de datos (DTS), esto consiste en migrar los datos de la base de datos de la ESPOL ( archivo Excel ) hacia el repositorio del ICM, para llevar a cabo este proceso se desarrollo una aplicación en Visual Basic 6.0 ( dts\_ICM.exe ), que permite hacer las siguientes tareas:

- Extraer del archivo de Excel información personal y académica (Nombres, apellidos, dirección, año y termino de ingreso, sexo, estado civil, teléfono, año y termino de registro, numero de matricula, código y nombre de materias tomadas, paralelo, horas teóricas y practicas, estado de la materia, nota de



examen 1, nota de examen 2, nota de examen 3, promedio, vez tomada la materia, entre otros campos ).

- Limpieza de datos consiste en registrar en la base del ICM información que sea consistente, puesto que si se agregan datos que faltan o están redundantes las agregaciones al repositorio podrían ser imprecisos.
- La transformación de los datos consiste en codificar los valores que toman cada columna del archivo de Excel un código único que identificara a ese registro en las tablas de hechos y dimensiones que mas adelante indicaremos. Por ejemplo un estudiante de ingeniería en Estadística Informática, que haya tomado las materias de microeconomía y análisis de variable real, en el año 1999 termino II. Este registro se ingresara a la base de datos de la siguiente manera:

Id\_tiempo=1; indica que se registro en el año 1999 termino 2.

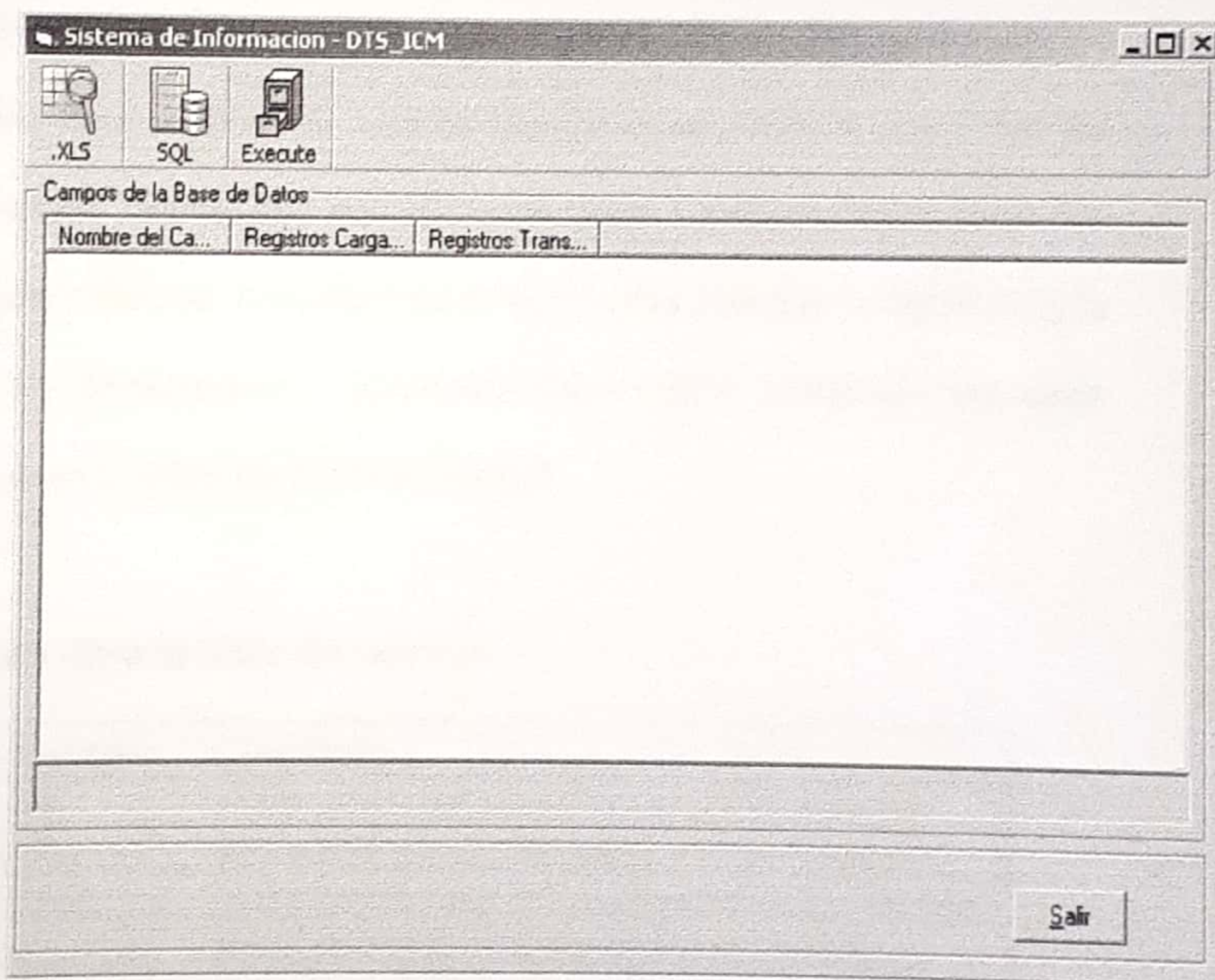
Id\_materias=2; indica que tomo la materia de microeconomía.

Id\_materias=3; indica que tomo la materia de variable real.

Id\_carreras=1; indica que es un estudiante de Ingeniería en Estadística Informática.



El proceso de transformación de datos culmina con la inserción de estos códigos a cada una de las tablas del repositorio de datos, este proceso se debe llevar a cabo cuando el término académico este finalizado por completo debido a que la información académica de los estudiante una vez que ingreso al repositorio del ICM esta ya no puede ser modificada a continuación se muestra la pantalla principal del programa de carga de datos al repositorio del ICM.



**Figura 4.2** Pantalla Principal del programa de migración de datos



#### 4.1.3. Almacenamiento de los Datos.

Los datos históricos de los estudiantes una vez ingresados al repositorio del ICM, estos servirán de fuente de información para la base de datos multidimensional que será representada por un cubo de información aplicando los servicios OLAP de Microsoft SQL Server.

#### 4.1.4. Determinación de las Tablas de Hechos y Dimensiones.

Para iniciar este proceso debemos primero determinar cuales son los campos para la tabla de hechos (contiene las notas 1, 2 y 3 de los estudiantes registrados en cada termino así como las claves foráneas para identificar las dimensiones relacionadas con estos registros) y la tabla de dimensiones (Contiene información detallada de cada registro en la tabla de hechos) tabla 5.

#### Campos para la tabla de hechos:

Campo ó Medidas	Descripción
Id_alumnos	índice para la tabla alumnos (dimensión alumnos)
Id_materias	índice para la tabla materias (dimensión materias)
Id_profesores	índice para la tabla profesores (dimensión profesores)
Id_tiempo	índice para la tabla tiempo (dimensión tiempo)
Id_carreras	índice para la tabla carreras (dimensión carreras)
Nota1	Indica la nota sobre 100 puntos del primer examen.
Nota2	Indica la nota sobre 100 puntos del segundo examen.
Nota3	Indica la nota sobre 100 puntos del tercer examen.

**Tabla 5.** Campos para tabla de hechos



**Campos para la tabla alumnos (dimensión alumnos):**

Campo	Descripción
Id_alumnos	índice para la tabla de hechos
estud	Nombre del estudiante
residencia	Dirección domiciliaria del estudiante
sexo	Género masculino o femenino
Est_civil	Estado civil del estudiante
Ter_ingreso	Término que ingreso a la ESPOL
Anio_ingreso	Año que ingreso a la ESPOL
Fact_p	Factor socio económico
Procedencia	Ciudad de origen del estudiante
Tip_colegio	Tipo de colegio (fiscal, particular, fisco misional)

**Tabla 6.** Campos para dimensión alumnos.**Campos para la tabla materias (dimensión materias):**

Campo	Descripción
Id_materias	índice para la tabla de hechos
nombre	Nombre de la materia
codigo	Código de la materia
Hor_teoricas	Horas teóricas asignada a la materia
Hor_practicas	Horas practicas asignada a la materia
Nivel	Nivel según flujo de carrera
Área	Campo al que pertenece según flujo de carrera

**Tabla 7.** Campos para dimensión materias.**Campos para la tabla profesores (dimensión profesores):**

Campo	Descripción
Id_profesores	índice para la tabla de hechos
identificación	Cedula, pasaporte
Nom_profesor	Nombre del profesor
email	Dirección electrónica

**Tabla 8.** Campos para dimensión profesores.



**Campos para la tabla tiempo (dimensión tiempo):**

Campo	Descripción
Id_tiempo	Índice para la tabla de hechos
anio	Año de registro
termino	Término de registro

**Tabla 9.** Campos para dimensión tiempo.

**Campos para la tabla carreras (dimensión carreras):**

Campo	Descripción
Id_carreras	Índice para la tabla de hechos
nombre	Nombre de la carrera
U_academica	Unidad académica que pertenece

**Tabla 10.** Campos para dimensión carreras.

**4.1.5. Base de Datos Multidimensional.**

Actualmente, la metodología de diseño favorita se conoce como almacén de datos dimensional. Se basa en dos elementos fundamentales conocidos como hechos y dimensiones. La tabla de hechos contiene información que se desea medir. Se basan en las tareas que se desean dar soporte.

Por ejemplo, se tiene una tabla de hechos que contenga las notas de los tres parciales en la carrera de Ingeniería en Estadística Informática o el número de veces en que los estudiantes han tomado una materia. Las tablas de

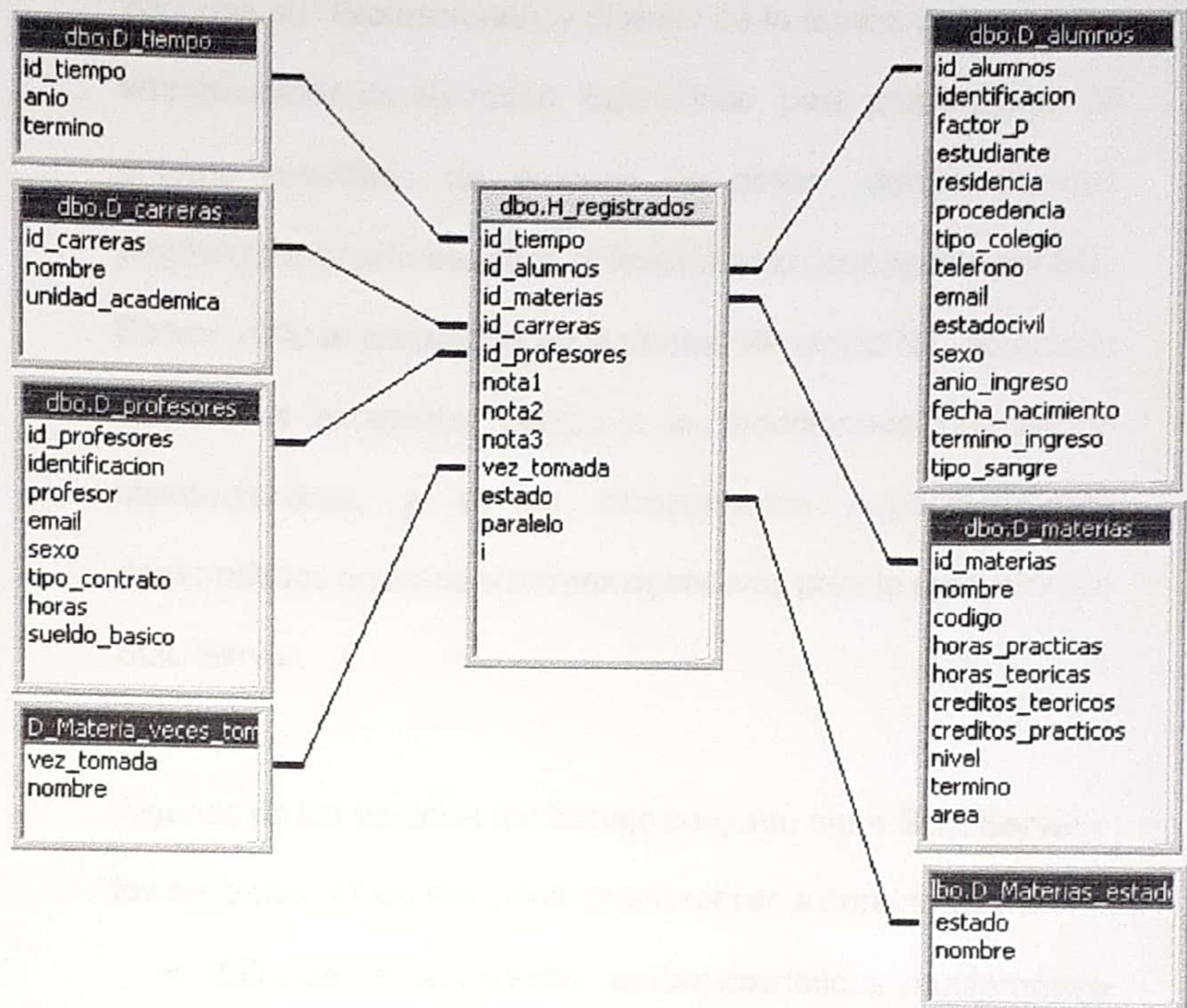


dimensiones son las ventanas en que se evalúa la información deseada. Las dimensiones ofrecen la información que se utilizará en las cláusulas **where**.

Las tablas de dimensión incluyen tablas sobre alumnos, materias, profesores, tiempo. En su conjunto la tabla de hechos y dimensiones se conocen como esquema estrella. Un esquema es sencillamente, una estructura de datos que contiene una serie de tablas. Estas tablas se suelen organizar utilizando el método normalizado de diseño.

Las tablas de datos normalizadas se ajustan para almacenar datos en forma muy eficaz. En la Figura 4.3 se representa el esquema estrella .





**Figura 4.3** Esquema estrella para el Cubo de Datos

#### 4.1.6. Elección del DBMS

Este proyecto hace uso del DBMS "Microsoft SQL Server 2000" y los Servicios OLAP, el cual es un manejador de base de datos relacionales y bases de datos OLAP, compuesto por un conjunto de elementos, que se integran con los sistemas operativos NT de Windows (Windows NT 4.0 Workstation y Server, Windows 2000 Profesional, Server y Advanced Server y



Windows XP Professional), y el resto de la familia de productos empresariales de Microsoft, BackOffice, para proporcionar un entorno avanzado de proceso de datos., dentro de una arquitectura cliente-servidor, el rendimiento conseguido por SQL Server 2000 al ejecutarse en sistemas Windows NT, ediciones Server, es excelente, debido a la mencionada orientación cliente-servidor, y a los componentes específicamente desarrollados en estos sistemas operativos para la ejecución de SQL Server.

Algunas de las ventajas del trabajo conjunto entre SQL Server y los sistemas NT de Windows se enumeran a continuación:

- SQL Server aprovecha las características multiproceso de Windows NT, utilizando todos los procesadores instalados para optimizar el manejo de datos.
- El sistema de seguridad de SQL Server 2000 está conformado por un usuario creado en la base de datos. De esta forma, el usuario del sistema de información sólo debe iniciar su sesión en el sistema y la aplicación automáticamente se identificará al comenzar su sesión de trabajo, puesto que al conectarse con SQL Server, se establece una relación de confianza en la que SQL



Server verifica si el usuario y la clave de acceso es correcto.

- Para las labores de supervisión del funcionamiento, SQL Server aprovecha el Visor de sucesos del sistema operativo para insertar sus propios mensajes, unificando en un sólo lugar el sistema de avisos. De igual modo, utiliza el Monitor del sistema de Windows NT para aspectos relacionados con el rendimiento de las base de datos.
- Sobre la disponibilidad inmediata de los datos en casos de fallo del servidor, SQL Server aprovecha las capacidades de clustering de que dispone Windows NT Server, de forma que si en un sistema se han instalado dos servidores en clúster, SQL Server realizará el cambio al servidor de respaldo en el caso de que se produzca una caída del principal.

SQL Server puede ejecutarse en una amplia variedad de sistemas operativos. Dependiendo del sistema, podrá actuar



como cliente o servidor. Los sistemas a los que se proporciona capacidad de servidor son:

- Windows NT en cualquiera de sus ediciones (Server,y Workstation) y Windows 9x. En cuanto a los sistemas para los que SQL Server dispone de elementos de cliente están los antes mencionados en el aspecto de servidor más Windows 3.x, MS-DOS, Macintosh y UNIX.
- Finalmente, fuera del ámbito de sistemas, también puede ejecutarse como cliente dentro de los navegadores de Internet.
- En cuanto a la escalabilidad del motor de datos, puede manejar desde pequeñas base de datos en modo local a grandes base de datos con conexiones de miles de usuarios y más de un terabyte de capacidad de almacenamiento.



#### 4.1.7. Data Warehouse en SQL Server 2000.

El componente encargado de guardar la información histórica desde el primer semestre del año 1999 hasta el segundo semestre del año 2002 es el repositorio de datos (Data warehouse ), mediante una base de datos relacional en SQL Server 2000.

Una Repositorio de datos, es una base de datos que está formada por una serie de elementos, también denominados objetos de la base de datos, que permiten organizar la información, relacionarla con otros objetos de la base de datos, mantener su integridad, etc.

Entre los objetos más importantes podemos destacar los siguientes:

- Tabla.
- Índice.
- Vista.
- Procedimiento almacenado.
- Función o rol.
- Desencadenador.



El sistema desarrollado para esta tesis necesita de un repositorio, para lo cual se diseña una base de datos cuya estructura permita realizar consultas, análisis y actualización de datos.

La base de datos para el repositorio del ICM incluye los siguiente:

#### **4.1.8. Tablas Principales:**

- *Estudiantes.*- Tabla donde se almacenan información de los estudiantes registrados en las carreras de Ingeniería en Estadística Informática y Auditoria y control de Gestión.
- *Materias.*- Tabla donde se encontrará la información detallada de las materias dictadas en cada termino académico.
- *Profesores.*- Tabla donde se encontrará la información detallada de los profesores que dictaron las materias en los términos académicos comprendidos entre el primer semestre del año 1999 al segundo semestre del año 2003.



- *Registros.*- Tabla donde se almacenan los registros de los estudiantes en los periodos mencionados anteriormente.
- *Carreras.*- Tabla donde se encontrará la información de las carreras que tiene el Instituto de Ciencias Matemáticas.
- *Tiempo.*- Tabla donde se encontrará la información correspondiente a las fechas de los periodos y términos académicos.
- *Materias Estado.*- Tabla donde se encontrará información de los estados que toma una materia (Aprobada, Reprobada, Perdida por faltas ).



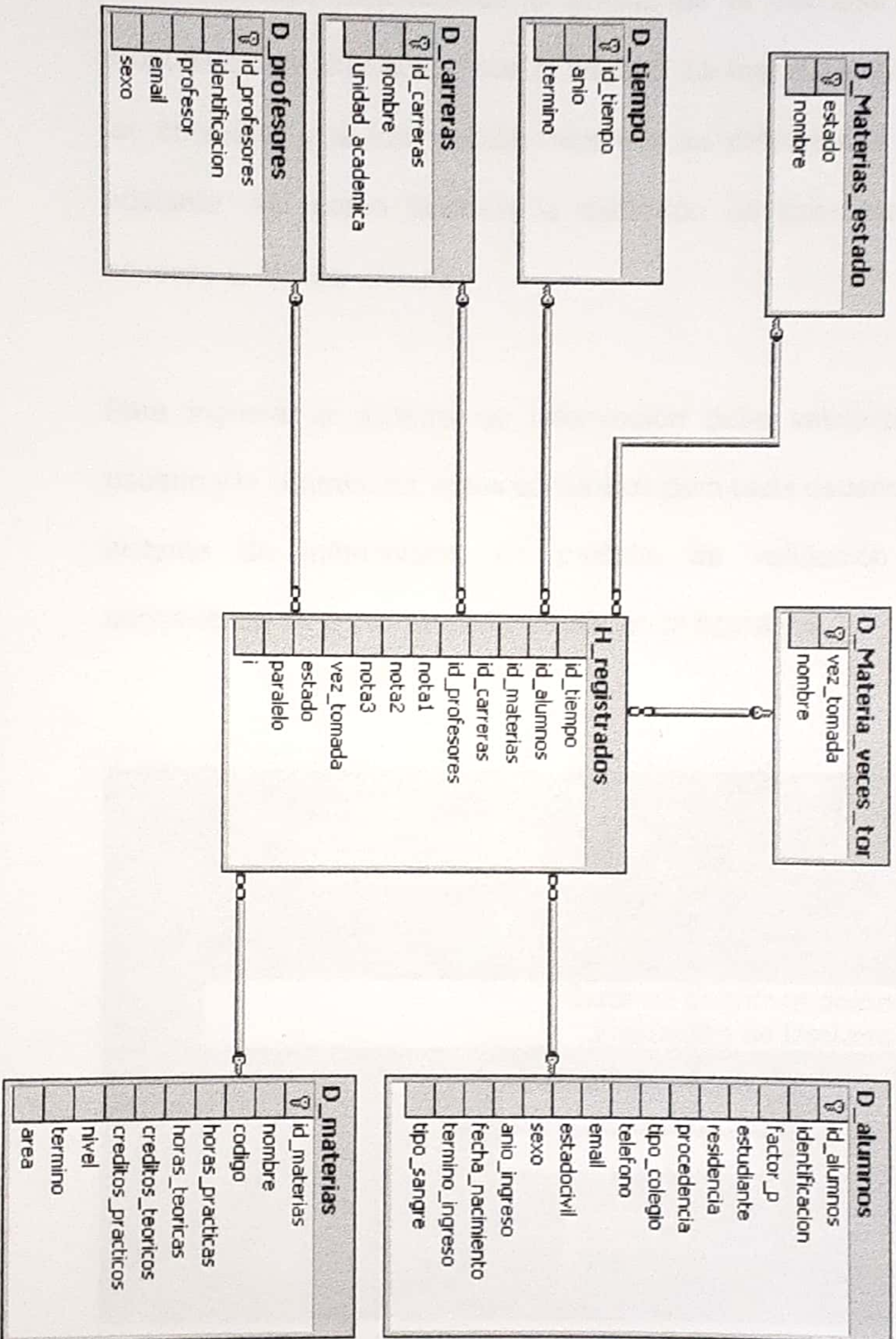


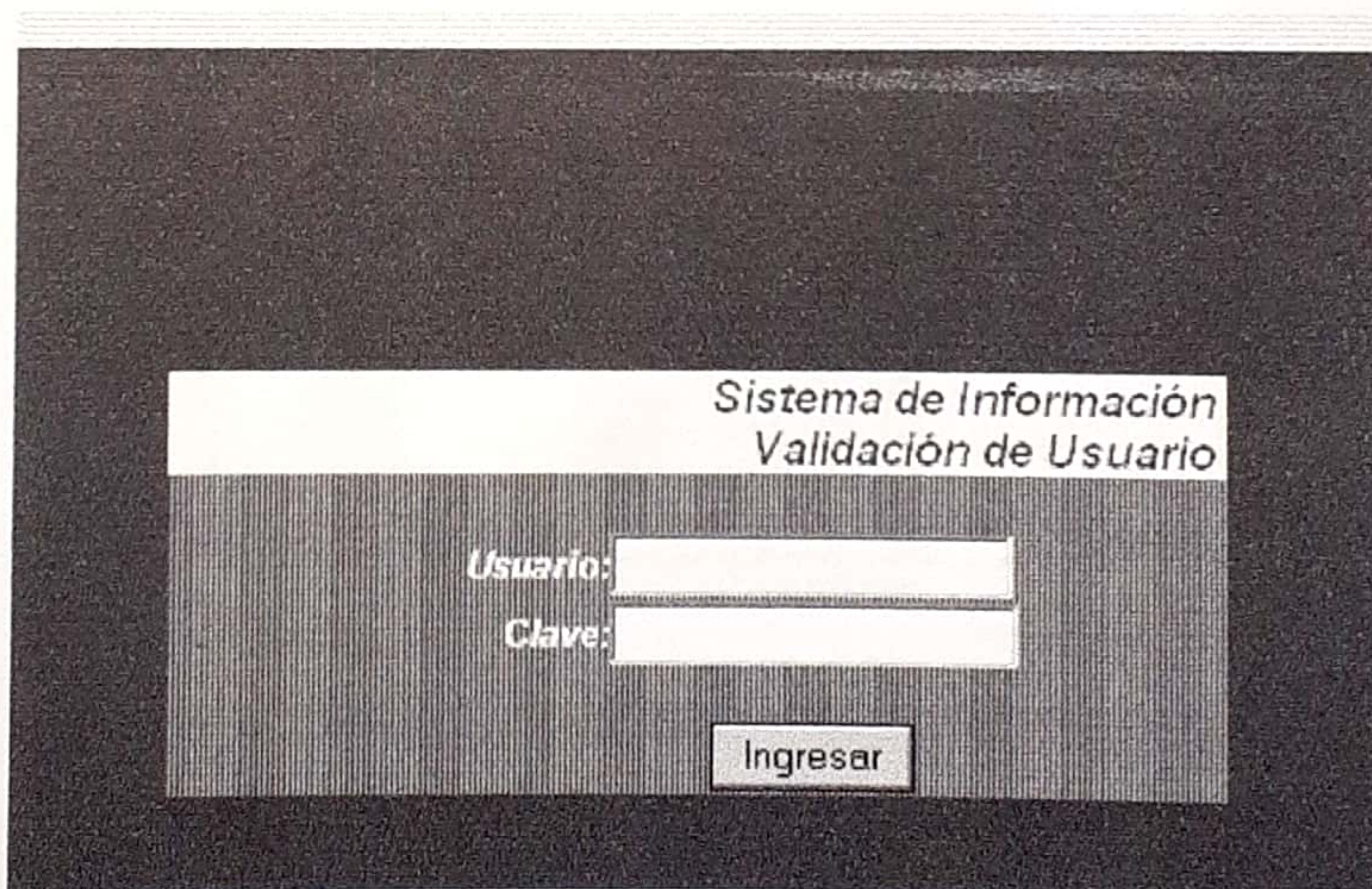
Figura 4.4. Estructura del Data Warehouse.



#### 4.1.9. Diseño de la interfaz del usuario

A continuación presentamos el diseño de la interfase web propuesto para interactuar con el usuario. La forma de realizar las consultas y la información mostrada se definiremos mas adelante, así como también la distinción de opciones de acuerdo al tipo de usuario.

Para ingresar al sistema de información debe validarse el usuario y la contraseña, estos son únicos para cada usuario del sistema de información. La pantalla de validación de contraseñas se presenta continuación en la figura 4.5.

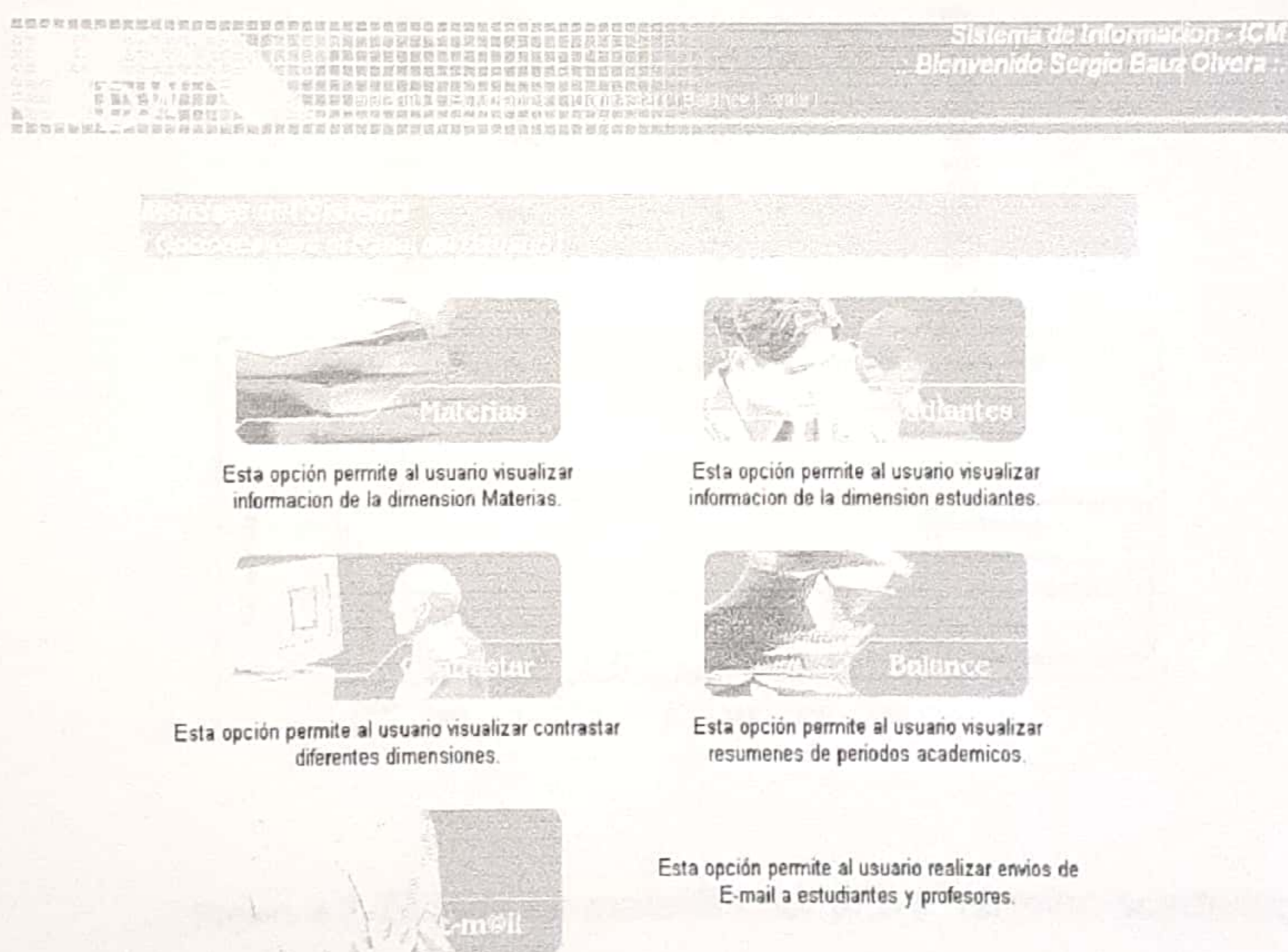


The image shows a screenshot of a web interface for user validation. The background is dark. At the top, there is a white rectangular box containing the text "Sistema de Información" and "Validación de Usuario" stacked vertically. Below this box, there are two input fields. The first is labeled "Usuario:" and the second is labeled "Clave:". Below the "Clave:" field, there is a button labeled "Ingresar".

**Figura 4.5.** Pantalla de Validación de Usuarios



Para la pantalla principal se presenta dividido el área en dos secciones ( Frames), la primera área ubicada en la parte superior de la pagina, contiene un mensaje de bienvenida al usuario así como una fila con todas las opciones que presenta el sistema, la segunda área ubicada en el centro de la pagina , tenemos enlaces directos a las herramientas de análisis. También el centro de la pagina se destinará para mostrar las tablas y gráficos resultantes de las consultas elegidas por los usuarios esto se lo puede ver en la figura 4.6.



**Figura 4.6. Pantalla Principal.**



En cuanto a la información de utilidad para el usuario (coordinadores de carrera), se presentan consultas del número de materias aprobadas, reprobadas, reprobadas por falta, desde el primer semestre de 1999 hasta el segundo semestre del 2002 desagregada por términos académicos, así como se muestra Figura 4.7.

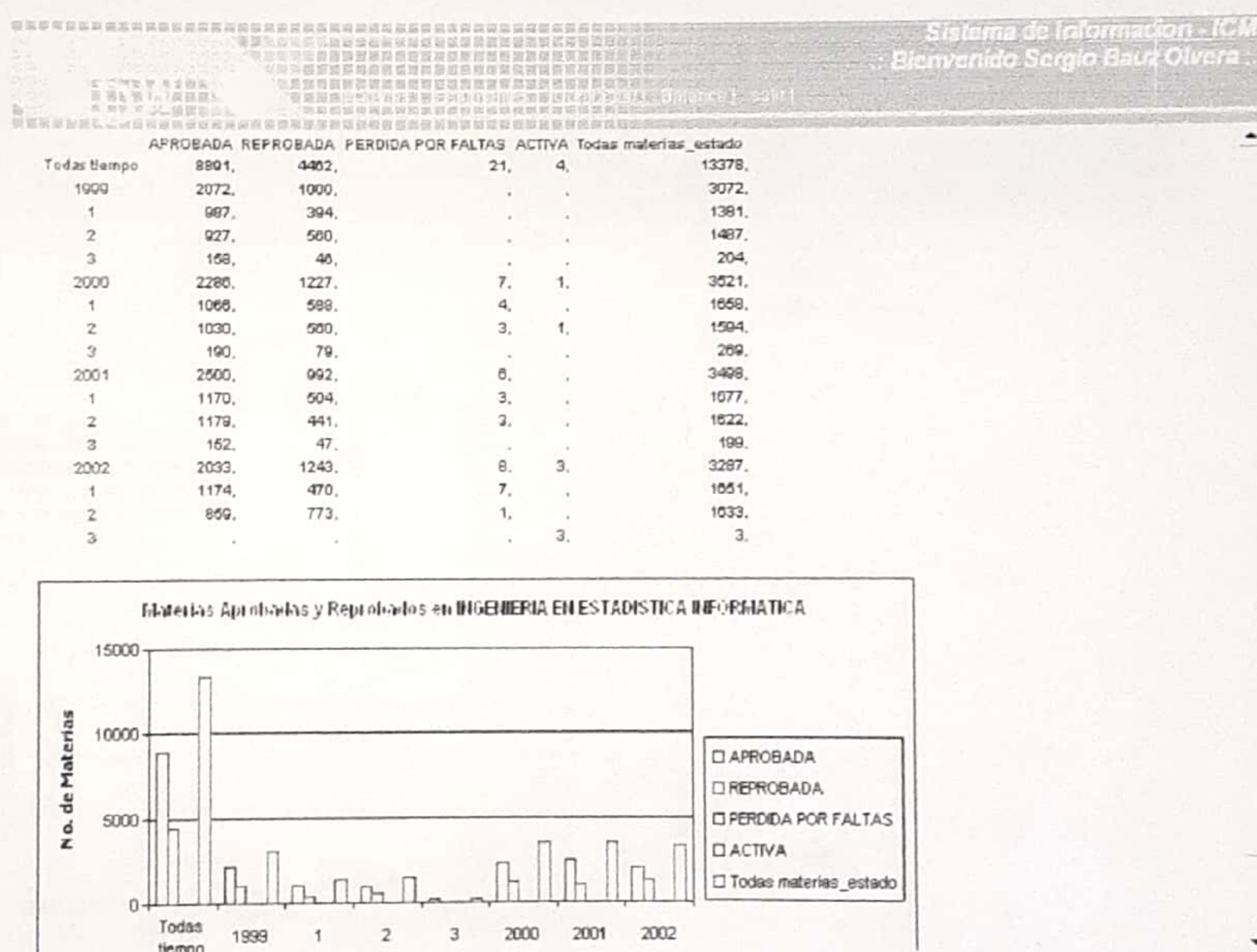
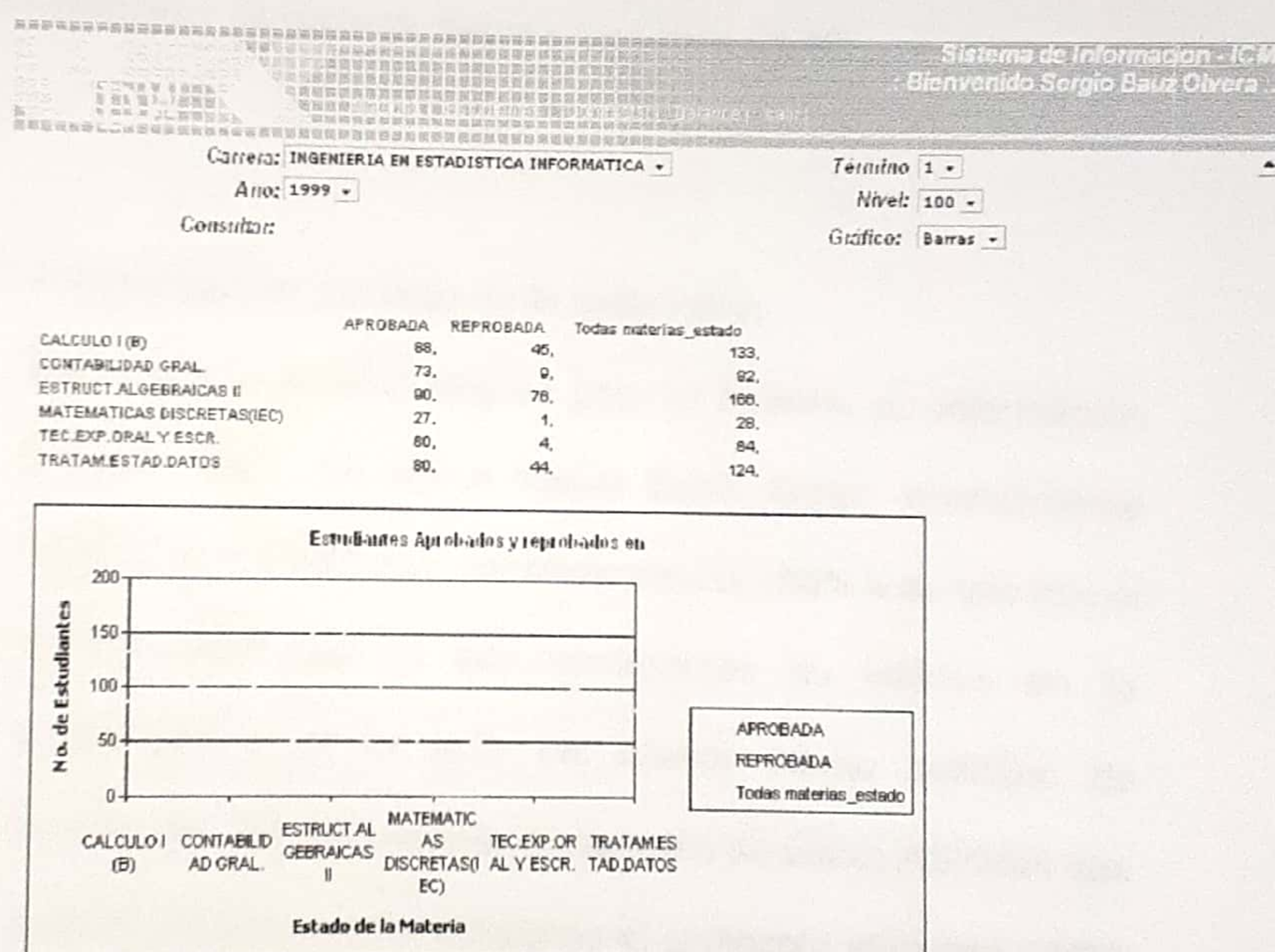


Figura 4.7. Estado de materias por año y término académico



A continuación se muestra opciones de selección para los periodos anteriormente analizados, visualizando 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 Informática, así como se muestra en la figura 4.8.



**Figura 4.8** Estado de materias por año, termino, nivel, según flujo de la carrera de Ing. Estadística e Informática.



Todo en detalle se mostrará en los anexos.

## **4.2. DESARROLLO DEL SISTEMA**

En esta sección se va a explicar los detalles concernientes a la implementación del sistema, la cual incluye lo referente a los lenguajes de programación utilizados tanto en las páginas de la aplicación como en el manejador de base de datos y Servicios OLAP.

### **4.2.1. Programación del lado de la aplicación**

Dado que la interfaz elegida para el sistema de información (Pagina Web). Se utilizo Visual Basic Script, normalmente referido como VBScript, orientado casi al 100% a su uso con el entorno ASP por lo que obviaremos su utilidad en la programación en el lado del cliente (como sustituto de JavaScript). En esta aplicación también se utiliza ASPMail que permite enviar e-mails utilizando el protocolo standard SMTP desde cualquier programa que pueda utilizar componentes ActiveX/OLE , incluyendo ASP.

Esto incluye:



- Mensajes SMTP (envío).
- Múltiples archivos adjuntos.
- Soporte para MIME y UUEncoding para múltiples archivos adjuntos.
- Mapa de caracteres US Ascii e ISO-8859-1.
- Codificación de la línea del tema para temas de mensajes de 8 bits.
- Servidores SMTP secundarios (Si el servidor SMTP primario se cae, el servidor secundario es utilizado).
- Soporte especial para la cabecera (Standard X-Priority headers, MS Mail (including Exchange) priority headers, Urgent header, ConfirmReading and ReturnReceipt Headers).

#### **4.2.1.1. Objeto ASPMail.**

El primer paso para usar el ASPMail es el de crear el objeto.

Esto puede realizarse con el siguiente código:

```
<% Set Mailer = Server.CreateObject ("SMTPsvg.Mailer") %>
```

#### **Configurar algunas propiedades**

El primer paso para usar el ASPMail es el de crear el objeto.

Esto puede realizarse con el siguiente código:



Las propiedades que se deben configurar son:

*FromName* El nombre que debe ser utilizado para el Sender del mensaje.

*FromAddress* La dirección e-mail que debe ser utilizada como el sender del mensaje.

*Subject* Este será utilizado como el Subject del mensaje.

*BodyText* Este será el mensaje propiamente dicho.

*RemoteHost* El servidor SMTP a ser utilizado para enviar el mensaje.

*AddRecipient name, email addr* Este método agrega un nombre y una dirección email al campo "To:" del mensaje

#### **4.2.1.2. Método SendMail**

El método SendMail no toma ninguna propiedad, pero devuelve un True / False para indicar Success ó Failure en el envío del mensaje email.

#### **4.2.1.3. Manejo de Errores**

Una vez que llamamos al SendMail y hemos controlado una falla de envío podríamos querer saber cual fue la condición del



error. El ASPMail ofrece una Property llamada "Response" que contiene una descripción textual del porqué el mensaje no fue enviado. Se puede utilizar este mensaje para informar al usuario ó archivarlo en un archivo.

Ejemplo:

```
if Mailer.SendMail then
' Message sent sucessfully
response.write ("Your message was sent")
else
' Message send failure
response.write ("Your message was not sent.")
response.write ("The error was: " & Mailer.Response)
end if
```

#### 4.2.1.4. Elección del lenguaje VBScript

Las razones por las cuales se utilizo en este proyecto VBScript son principalmente:

- Similitud con Visual Basic. Es cierto que la gran mayoría de desarrolladores ASP, especialmente los primeros, no venían del entorno Web sino del de programación. Era más normal que un programador Visual Basic hiciese la



programación de un sitio Web, utilizando por tanto VBScript por ser un subconjunto de lo que ya sabía.

- Sencillez, tanto en sintaxis como en funcionamiento. VBScript (y cualquiera de los lenguajes de la familia Visual Basic) son probablemente los lenguajes de programación más cercanos al lenguaje natural (inglés, claro) que existen. Esto hace más sencillo su aprendizaje: si se quiere hacer algo mientras ocurra tal cosa, se utilizará la instrucción `while`, por ejemplo. Todas las instrucciones terminan con alguna de cierre, como `if..end if`, lo que hace que el código sea mucho más sencillo de leer que utilizando, por ejemplo, la sintaxis de corchetes de C o Javascript.
- Es imposible no pensar en una trampa de Microsoft para extender su lenguaje: puesto que ASP es una tecnología de Microsoft, parece lógico que la propia empresa favoreciese este lenguaje en detrimento de Javascript, probablemente utilizándolo desde el principio en los primeros ejemplos, tutoriales y libros sobre ASP.

No es en lo absoluto obligatorio utilizar VBScript para las páginas ASP. De hecho, si se está familiarizado con Javascript



o se es programador de C y no de Visual Basic, es mucho mejor utilizar JavaScript. Tiene prácticamente las mismas posibilidades y su sintaxis es muy cercana a la de estos dos lenguajes.

#### **4.2.1.5. Utilización de VBScript**

VBScript es un lenguaje de script. Estos lenguajes, que en castellano se denominan en teoría lenguajes de secuencias de comandos o también de automatización, se caracterizan por depender de una aplicación anfitriona: VBScript no puede crear un programa que envíe texto al usuario; depende de la aplicación anfitriona. En el caso de utilizarlo para escribir páginas ASP, dependerá de los objetos intrínsecos de ASP para recibir y enviar datos. En el caso de utilizarlo en el cliente, Internet Explorer se encargará de "entenderle" y suministrarle los medios que necesita.

#### **4.2.1.6. ActiveX Scripting**

La aplicación anfitriona (vamos a hablar a partir de ahora del servidor Web) utiliza ActiveX Scripting para ejecutar código VBScript. ActiveX Scripting (o Active Scripting) es un motor de script reutilizable que se puede usar en cualquier aplicación, en



este caso en ASP (o Internet Explorer). Por defecto, viene con dos lenguajes integrados: VBScript y JScript.

ASP funciona extrayendo todo el código VBScript de la página, instanciando ActiveX Scripting y pidiéndole que ejecute el código. Además, ASP le pasa al motor de script el contexto, que no son más que referencias a los objetos ASP (Response, Request...) para que el código pueda ejecutarse: cuando pone Response.Write, tiene que escribir el texto para que sea enviado como HTML, por ejemplo. Eso no es problema de VBScript, sino de ASP.

Por defecto, el lenguaje de script de las páginas ASP es VBScript, así que si no definimos nada, ActiveX Scripting pensará que el código es VBScript e intentará ejecutarlo como tal. Si queremos que ejecute código JScript, tendremos que indicarlo en la página ASP así:

```
<%@LANGUAGE = "JScript"%>
```

O establecerlo por defecto en la configuración del servidor Web.



#### 4.2.1.7. Declarar variables

No es obligatorio declarar las variables que utilicemos en VBScript. Además, como es un lenguaje sin tipo, declararla se limita a "avisar" de que vamos a utilizar tal variable con tal nombre: <<< Dim strsql >>>

Si no se declara la variable, se podrá utilizar de todas formas en el primer momento en que se haga referencia a ella. Esto, sin embargo, no es muy aconsejable, puesto que se podría escribir incorrectamente y obtener errores difíciles de encontrar.

De hecho, es muy recomendable incluir la instrucción Option Explicit al principio del script, que obliga a tener todas las variables que utilicemos declaradas antes de su uso, devolviendo un error en caso contrario. Así se sabrá si en lugar de nombre se ha puesto por error "nimbre" por ejemplo y por esto no funcionaba el script.

#### 4.2.1.8. Constantes

Las constantes son como las variables en cuanto que se referencian mediante el nombre pero, a diferencia de éstas, su valor nunca cambia. VBScript además define sus constantes intrínsecas o constantes de VBScript.





#### 4.2.1.9. Comentarios

Los comentarios son fragmentos de texto que podemos incluir en el script y que no se ejecuta, de forma que sólo son visibles cuando leemos el código del script. Son muy útiles y recomendables, puesto que pueden ayudar a clarificar ciertas partes de código que no es muy clara.

Para escribir comentarios utilizamos la apóstrofe (') o comilla simple al comienzo de la línea. El resto de la línea es un comentario:

```
Request.form("userID")
```

```
' Esto es un comentario
```

#### 4.2.1.10. Java Script

Para poder utilizar imágenes que no sean de excesivo peso (en Kilobytes) de tal forma que se pueda interactuar con ellas de manera dinámica procedemos a colocar las mismas en el diseñador y manejador de imágenes Macromedia Fireworks MX. Para esta implementación se dividió en slides la imagen de la pantalla principal del sistema, para que al momento de



cargarse la página en el explorador esta no sea demasiado lenta.

Luego se importa la imagen desde Macromedia Dreamweaver MX para empezar las conexiones del mismo a las consultas y gráficos del sistema.

El script utilizado para efectuar esta operación es JavaScript y se lo codifica de la siguiente manera:

```
<script language="JavaScript">  
    <!-- function MM_swapImage() {  
        function MM_swapImgRestore() {  
            function MM_preloadImages() {  
                function MM_findObj(n, d) {  
                    }  
                }  
            }  
        }  
    }  
    // End --> </script>
```

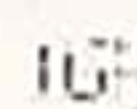
Estas cuatro funciones son las que controlan el despliegue de las imágenes del index uniendo los slides y formando la página principal del Sitio.



#### 4.2.1.11. Active Server Pages (ASP)

El lenguaje de programación utilizado como ya hemos dicho en muchas ocasiones es ASP (Active Server Pages) con aplicaciones VBScript y JavaScript, apoyados en lenguaje HTML. A continuación se mostrará las características generales del sistema y la programación usada a lo largo del sistema.

Los símbolos `<% función %>` indican que el comando que se va a ejecutar está implementado en Lenguaje ASP el cual se muestra esta figura si observamos la página sin activar.



Dentro de este comando ASP se especifica en la aplicación a utilizar, en este caso es VBScript.

```
<%@LANGUAGE="VBSCRIPT"%>
```

#### 4.2.1.12. Conexión a la base de datos

Puede utilizar la opción ODBC del Panel de control (o el Administrador de ODBC) para agregar, modificar y eliminar controladores y orígenes de datos ODBC en cualquier momento



posterior a la instalación de los archivos ODBC mediante el programa Instalar.

Para iniciar la opción ODBC del Panel de control:

1. En el grupo Principal del Administrador de programas de Microsoft Windows, haga doble clic en el icono "Panel de control".
2. En la ventana Panel de control, haga doble clic en el icono "ODBC".

Cuando inicie la opción ODBC del Panel de control (o el Administrador de ODBC), puede hacer clic en el botón "Ayuda" que aparece en cada pantalla para ver Ayuda sobre la forma de introducir la información en dicha pantalla. La Ayuda para la opción ODBC del Panel de control muestra información detallada sobre la forma de utilizar este programa.

Habiendo especificado con anterioridad el Data Source Name en el Administrador ODBC, se establece la conexión del sistema a la base de datos SQL Server, en una página, de la siguiente manera.



```
<!--#include file="Connections/conn.asp" -->
```

Este archivo (conn.asp) contiene todos los parámetros de conexión entre la aplicación y la base de datos, tales como: el driver de conexión, el usuario y password para conectarse a través de una variable llamada MM\_conn\_STRING.

#### 4.1.2.13. Abriendo la base de datos

Se creará un objeto *recordset*, este no es mas que una forma de obtener registros de la fuente de datos, para poder identificar el origen de la información, especificando que tabla es la que vamos a utilizar. Aquí se abre la base de datos y se definen las sentencias DML (Insert, Select, Update, Delete) que se van a realizar.

```
<%  
  
set rs = Server.CreateObject("ADODB.Recordset")  
  
rs.ActiveConnection = MM_conn_STRING  
  
rs.Source = "Sentencias DML"  
  
rs.CursorType = 0  
  
rs.CursorLocation = 2  
  
rs.LockType = 3  
  
rs.Open()
```



%>

#### **4.2.1.14. Tablas en HTML**

Para implementar las consultas a los usuarios se van a utilizar tablas, donde se mostrarán los datos en forma alfabética. Se crean tablas con varias filas, las cuales se incrementan según la cantidad de datos que se van a mostrar por materias, por notas, etc. El código para crear una tabla es el siguiente:

```
<table width="100" border="0" cellspacing="0"
cellpadding="0" height="25">
```

#### **4.2.2. Programación del lado de la base de datos**

La programación del lado de la base de datos se la realiza en el motor de la base de datos (MS SQL Server), desarrollando los llamados procedimientos almacenados (Stored procedure).

##### **4.2.2.1. Procedimientos almacenados**

Los procedimientos almacenados son conjuntos de instrucciones del lenguaje SQL, pudiendo almacenarse en el propio servidor de base de datos, dando con esto un aumento en el desempeño de las consultas, en el motor de la base de



datos se encuentra todo un conjunto de rutinas para desarrollo de aplicaciones potentes y robustas, teniendo capacidad para compilación y depuración de estos subprogramas.

Por ejemplo, el procedimiento para distinguir que usuario está navegando en el sistema es el siguiente:

```
CREATE PROCEDURE ver_usuario
```

```
  @user varchar(40) AS
```

```
  Select * from usuarios
```

```
  where usuario = @usuario
```

Con esta sentencia, se crea un procedimiento almacenado, de nombre "ver\_usuario", al que se le pasa un parámetro, llamado @usuario, de tipo varchar(20), se realiza una consulta para obtener todos los datos del usuario, cuyo nombre de usuario coincida con el parámetro. Para la ejecutar los procedimientos almacenados se debe escribir el comando EXEC ó EXECUTE.

```
EXEC ver_usuario 'user'
```

Las instrucciones SQL que contienen los procedimientos almacenados, consisten en sentencias muy próximas al inglés



diseñadas para seleccionar registros de una o varias tablas de acuerdo con el criterio utilizado. Las sentencias de solicitud SQL pueden ser usadas en el momento del diseño en la propiedad recordsource de un control de datos esto permitiría crear mediante programación ASP asociados a controles de datos los criterios SQL utilizan la palabra clave SELECT, seguida de una de estas palabras claves WHERE, FROM HAVING, GROUP-BY, ORDER BY.

EL criterio SELECT selecciona los campos de la tabla, GROUP BY se utiliza para agrupar datos, ORDER BY el campo de referencia que ordena la información que se muestre en las tablas de forma ascendente y FROM es el nombre de la tabla

```
CREATE PROCEDURE estudiantes as  
SELECT * FROM estudiantes  
ORDER BY nombre ASC
```

Devuelve todos los campos de la tabla usuarios ordenados ascendentemente por nombre.

La cláusula WHERE nos permite especificar un criterio para recuperar los datos.

```
CREATE PROCEDURE tipo_user as
```



```
SELECT * FROM usuarios  
WHERE tipo= 'administrador' ORDER BY id_user DSC
```

Devuelve todos los campos de la tabla usuarios donde el campo TIPO sea igual a administrador ordenados ascendentemente por id de usuario.

#### 4.2.2.2. Disparadores

Otra de las herramientas que tiene el motor de la base de datos son los llamados disparadores, conocidos también como "Triggers", es una clase de procedimientos almacenados, que son invocados cuando hay una operación que afecte una determinada tabla en la base de datos, las operaciones comunes donde se usa los Triggers son:

- Insertar
- Eliminar
- Actualizar

Para la escritura de un Trigger se usa la siguiente sintaxis:

```
CREATE TRIGGER nombre ON tabla FOR acción AS código
```



#### 4.2.2.3. Visualizando Resultados

Para llamar a un procedimiento almacenado desde ASP propiamente desde Dreamweaver MX, a través de un objeto Recordset, se utilizan las siguientes sentencias:

```
Set rs = Server.CreateObject("ADODB.Recordset")
    rs.ActiveConnection = MM_conn_STRING
    rs.Source = "{exec user}"
    rs.CursorType = 0
    rs.CursorLocation = 2
    rs.LockType = 1
    rs.Open()
```

Claramente se muestra en la tercera línea, la forma en como se invocan a los procedimientos almacenados en el servidor utilizando la instrucción EXEC reemplazando como argumento, el parámetro USER, que es una variable que contiene el nombre de usuario para ser comparado, lo demás es sintaxis propia del lenguaje.

La información se mostrara en una tabla con un grafico que se generara en el momento. Para graficar en el browser se hace uso de los objetos Office Web Components de Microsoft Office ,



Este modulo viene en el momento que se instala Microsoft Office, en caso de que el servidor web no tenga instalado esta herramienta por motivos de seguridad, se deberá solo instalar el archivo OWC.dll Este archivo tiene la capacidad para generar cualquiera de los gráficos mostrados en los paquetes de Office.

Con el comando `rs.Fields.Item.Campo.value` que despliega los registros seleccionados anteriormente en el recordset.

Ejemplo:

```
<p align="left">  
  <font size="1" face="Verdana,">  
    <%  
      Response.Write(rs("nombre"))  
      Response.Write(rs("apellido"))  
    %>  
  </font>  
</p>
```



### **4.3. PRUEBAS DEL SISTEMA**

Las pruebas para este sistema se llevo a cabo en el laboratorio Omega del Instituto de Ciencias Matemáticas, esto nos dió una pauta acerca de los inconvenientes que se podrían presentar en el momento de la implantación del sistema de información, los requerimientos de hardware y software se deben de tomar muy en cuenta, ates de poner en marcha el sistema.

#### **4.3.1. Prototipos del sistema**

Antes del desarrollo del sistema de información se elaboro un prototipo para encontrar los posibles inconvenientes que se tendrían en aspectos como migración de datos, visualización de consultas y configuración del sistema en las estaciones de trabajo de los usuarios. Esto se lo hizo con la finalidad de hacer pruebas al sistema antes de que este entre en ejecución.

La confiabilidad es un aspecto del diseño, por lo que debe estar dentro del sistema, más adelante se indican las estrategias específicas que se utilizaron en el prototipo del sistema.



### 4.3.2. Análisis del Prototipo

A pesar de los mejores esfuerzos para la implementación del sistema, las necesidades de información no siempre se establecen correctamente. Esto puede ocurrir porque los usuarios pueden saber sólo que necesitan mejorar el sistema en ciertas áreas, o que deben de modificar los procedimientos existentes, o simplemente no están muy seguros del tipo de información que necesitan.

Por lo tanto, se pueden considerar los prototipos para proporcionar la información necesaria y formular el diseño o desarrollar el sistema, debido a que un prototipo es un sistema de trabajo que se desarrolla con rapidez para probar las ideas y el entendimiento sobre el nuevo sistema, en otras palabras no sólo es un diseño en papel, sino un software que produzca información impresa o en pantalla.

Los objetivos de esta técnica del prototipo son:

- Validar la funcionalidad del sistema.
- Probar los principios y verificar que el sistema funciona



- Incrementar el nivel de productividad del desarrollo de sistemas.
- Busca un acercamiento (interacción) entre el usuario y el sistema.
- Presentar en forma jerárquica las pantallas y el diálogo con el usuario.
- Obtener una realimentación (evolución del prototipo).

Temas complejos de procesamiento de datos tienen la necesidad de prototipos, mediante éstos se busca obtener una retroalimentación del análisis y del diseño antes de realizar su construcción.

Para tal efecto, se trabaja con el usuario final creando diálogos (pantallas) para entrada de datos, consultas, menús, efectuar demostraciones de ellos y realiza generación de informes, manipulación de información por pantalla.

Al construir el prototipo se facilitó el desarrollo del sistema, éste se realizó de las tres formas siguientes; un prototipo en papel que describa la interacción hombre-máquina, de forma que facilite al usuario la comprensión de como se producir tal interacción, un prototipo que funcione y por ultimo uno que



implemente algunos subconjuntos de la función requerida del sistema requerido.

#### **4.3.3. Necesidad de utilizar un prototipo**

##### No se conocen los requerimientos

La naturaleza de esta aplicación es tal, que la información disponible en la base de datos creada en SQL Server no está actualizada con respecto a las características que debe tener el sistema para satisfacer los requerimientos de los usuarios.

Este problema se origina porque se dispone de los datos de los posibles usuarios del sistema.

##### Nueva tecnología

El deseo de instalar nueva tecnología ya sea en los campos de la informática y comunicación de datos u otras áreas relacionadas, abre nuevas fronteras.

Después de la evaluación del sistema se determinó que su programación y diseño, debería implementarse vía Internet a través de una aplicación Web codificada en Active Server Pages (lenguaje de programación ASP) y demás herramientas descritas en el capítulo tres.



#### 4.3.4. Determinación del Sistema a implementar

El desarrollo de este prototipo y su evaluación final se llevó a cabo en una forma ordenada, sin importar las herramientas utilizadas. Los pasos seguidos fueron los siguientes:

##### Identificación de requerimientos conocidos

La determinación de los requerimientos de esta aplicación es tan importante para el método y ciclo del desarrollo del mismo. Por consiguiente se determinaron los requerimientos de hardware, software e interfaz de usuario y así establecer un estándar del sistema, además de conocer si se contaba con las herramientas necesarias.

##### Desarrollo de un modelo de trabajo

Es útil comenzar el proceso de construcción del prototipo con el desarrollo de un plan general que permita a los usuarios conocer lo que se espera de ellos y del proceso de desarrollo. Un cronograma para el inicio y el fin de la implementación es de gran ayuda, al igual que la utilización de un bosquejo.



Para el desarrollo del prototipo se prepara los siguientes componentes:

- La interfaz del usuario. El lenguaje para el diálogo entre el usuario y el sistema, pantallas y formatos para la entrada de datos.
- Módulos esenciales de procesamiento. Modulo de gráficos estadísticos (Líneas, Barras, Pasteles , entre otros).
- Salida del sistema. Diseño de la consulta y búsqueda de información.

Para construir este prototipo se tuvo que seguir estándares para que los datos que fueron empleados estuvieran ordenados y organizados (longitud de datos, características, tipo, entre otros).

#### Revisión del prototipo

Durante la evaluación del sistema se capturó información sobre la satisfacción del usuario (lo que le agrada y desagrada). Los cambios al prototipo son planificados con los usuarios antes de llevar a cabo tales modificaciones.



Se realizaron revisiones periódicas de la implementación y diseño del sistema; por parte del Director del Proyecto. De esta manera se deberá asegurar que las especificaciones del diseño y arquitectura del sistema se ajusten a las especificaciones funcionales obtenidas en la fase de análisis de sistemas. Y por último, se obtendrá la aprobación formal del presente diseño.

#### **4.3.5. Pruebas del sistema**

La prueba del equipo lógico es el método más usado para determinar si el sistema funciona como debe. El proceso de pruebas es uno de los componentes de un conjunto de actividades que permiten asegurar la calidad del producto realizado.

Uno de los principios básicos en la realización de pruebas es que estas han de ser llevadas a cabo por personas distintas a los diseñadores de los programas, tanto para evitar una simple verificación de que el programa funcione correctamente, como para probar que ese programa ha sido concebido e interpretado correctamente.

A continuación se detallan dos tipos de pruebas:



**a) Pruebas Globales.** Una vez que se han probado los componentes individuales y se han integrado, se ha de probar el sistema global. En esta etapa pueden distinguirse los siguientes tipos de prueba, cada uno con un objetivo claramente definido:

- Pruebas funcionales
- Pruebas de comunicación
- Pruebas de rendimiento
- Pruebas de volumen
- Pruebas de sobrecarga
- Pruebas de disponibilidad de datos
- Pruebas de facilidad de uso
- Pruebas de operación
- Pruebas de entorno
- Pruebas de seguridad

**b) Pruebas de aceptación.** Son aquellas pruebas que realiza el usuario con el objeto de comprobar si el sistema es aceptable. Estas pruebas son del mismo tipo que las mencionadas anteriormente, pero son determinadas por el usuario, en lugar de serlo por el equipo de desarrollo.



Un lugar especial de estas pruebas, es el de la ejecución en paralelo con el viejo sistema (en nuestro caso no se dio esto por ser primera vez que se desarrolla uno de estos sistemas ), para comparar los resultados producidos por ambas ejecuciones.

#### **4.3.6. Planificación de las pruebas.**

La fase de pruebas, por su envergadura e importancia necesita una organización seria y fiable. Ante una fase de pruebas, se debe tomar como axioma que se van a encontrar errores.

Los componentes de una planificación ser:

- **Objetivos:** Definir los objetivos de cada fase de las pruebas.
- **Criterios de terminación:** Especificar cuando se deben acabar las pruebas.
- **Cronología:** Fijar los tiempos necesarios para cada fase (diseño, escritura, ejecución).



- Responsabilidades: Especificar los responsables de cada fase, así como quién corregir los errores detectados.
- Bibliotecas de caso de prueba y normas: Crear técnica de identificación, escritura y almacenamiento de casos de prueba.
- Herramientas: Establecer cuáles pueden ser las herramientas de pruebas que se van a utilizar
- Tiempo de máquina: Determinar el tiempo que se necesita en cada fase del proyecto de prueba.
- Configuración de equipo: Detallar la necesidad de hardware y software especiales de equipo o de un período concreto.
- Integración: Describir el plan de integración del sistema.
- Métodos de seguimiento: Especificar los métodos que se han de utilizar en las pruebas.
- Depuración: Definir un mecanismo para informar sobre los errores detectados, para seguir el proceso de las correcciones y para incorporar éstas al sistema.
- Terminación de las pruebas. Es difícil que el último error detectado, era el único que quedaba. Sin embargo,



existen métodos para mostrar cuando esta próximo el final, los dos más comunes son:

- Terminar la prueba cuando el tiempo establecido para la misma ha expirado.
- Terminar la prueba cuando todos los casos de prueba se ejecutan sin detectar errores.

Otros métodos mas complicados de aplicar pero más efectivos son:

- Estimación de número total de errores del programa.
- Estimación del porcentaje de estos errores que pueden encontrarse fácilmente.
- Estimación de qué fracción de errores se originan en procesos particulares de diseño.

Para realizar cualquiera de estas estimaciones es necesario contar con una historia o experiencia previa que permita definir dichas estimaciones.

#### **4.3.7. Herramientas para pruebas**



Son aquellas que apoyan la fase de la evaluación de un Sistema o de partes del mismo contra las especificaciones.

Incluyen facilidades para examinar la correcta operación del sistema así como el grado de perfección alcanzado en comparación con las expectativas.

La revolución del procesamiento de datos de manera computarizada, junto con las practicas de Diseño sofisticadas están cambiando de forma dramática la manera en que se trasladan las especificaciones de Diseño de Sistemas de Información funcionales.

#### **4.3.5. Realización de pruebas**

Para cada sistema se realizarán diferentes clases de pruebas:

Pruebas Unitarias: Todos los componentes del sistema que se desarrollen individualmente para comprobar su correcto funcionamiento.

Pruebas de integración: Se prueba la integración entre los componentes del sistema para demostrar que se puede encajar correctamente.

Pruebas de sistemas: Se prueba el sistema globalmente.



#### 4.3.6. Tipos de pruebas

Existen dos tipos de pruebas :

- Pruebas del tipo CAJA BLANCA, que permite examinar la estructura interna del programa.
- Pruebas del tipo CAJA NEGRA, donde los casos de prueba se diseñan considerando exclusivamente las entradas y salidas del sistema, sin preocuparse por la estructura interna del mismo.

#### 4.4. USO Y MANTENIMIENTO DEL SISTEMA

Este sistema de información constituye una herramienta de ayuda al análisis y toma de decisiones, teniendo como interfaz de usuario un browser en las estaciones de trabajo de los directivos del ICM, la navegación se la hará por la intranet del Instituto de Ciencias Matemáticas, a través de una aplicación de fácil manejo para los usuarios.

La navegación es personalizada (esto es en el momento que el usuario inicia una sesión el sistema se personaliza tanto en menú e



información de Bienvenida) pues estos son generados por llamadas a la base de datos.

Cada usuario podrá generar consultas multidimensionales de manera dinámica, mediante la selección de los parámetros para la búsqueda en el repositorio de datos. Las variables (Dimensiones y Medidas) que el usuario seleccione constituirán los parámetros para las filas y columnas que se mostraran en la consulta. Cada visualización de las tablas serán acompañados por sus respectivos gráficos de barras o líneas según sea el caso.

Existe otras funcionalidades que tendrá el sistema como migrar los datos a hojas de Excel para después llevarlo a algún software estadístico y poder hacer un análisis mas detallados de los datos.



# 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 al 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.





# **ANEXOS**



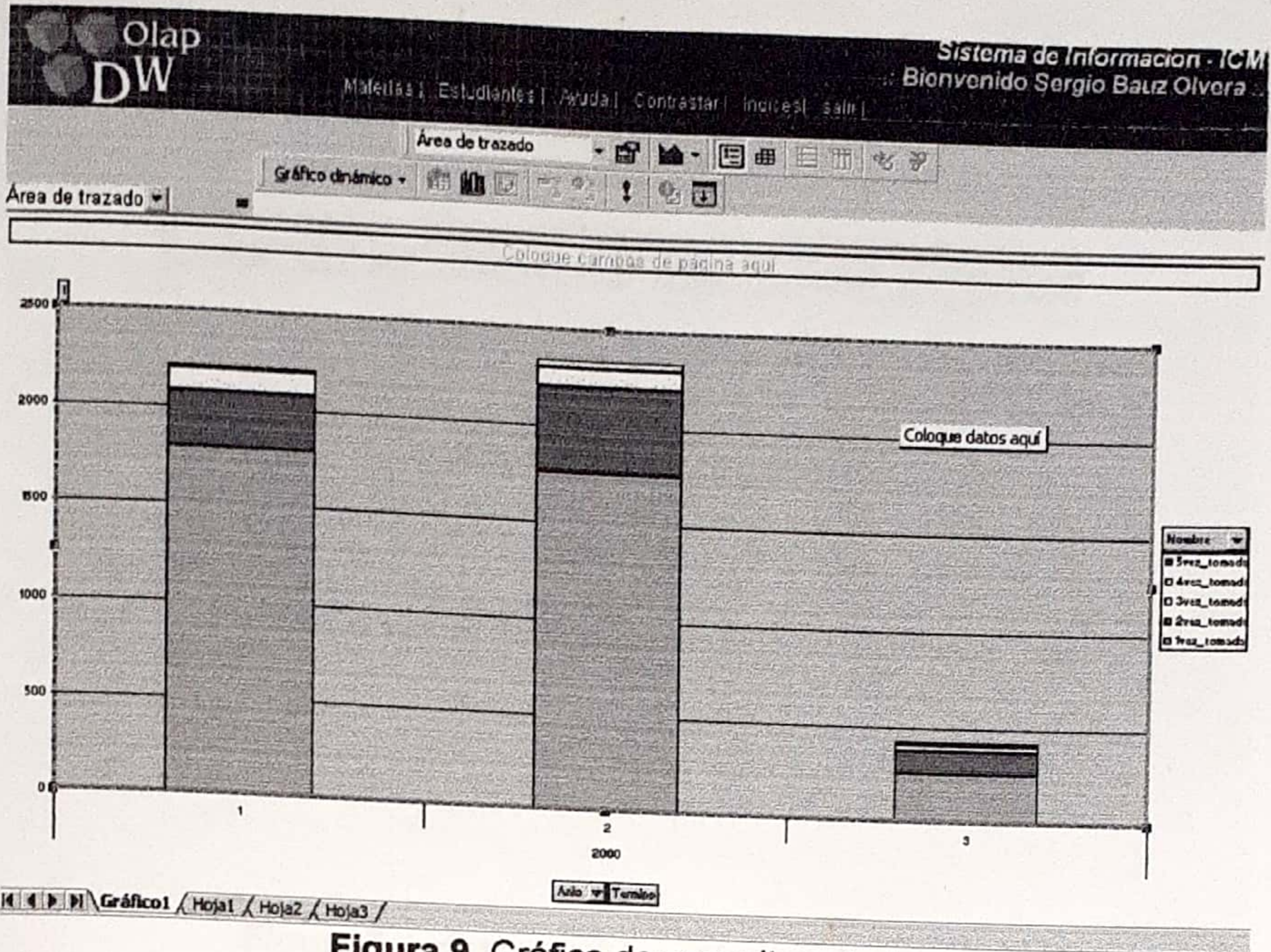


Figura 9. Gráfico de consulta anterior



Carrera: INGENIERIA EN ESTADISTICA INFORMATICA Período: 1999 Tér. 2

**Mensaje del Sistema**  
 [ Cálculo de Indices Académicos ]

Estado	Indice	Año	Término
APROBADA	0,623	1999	2
REPROBADA	0,377	1999	2

Vez Tomada	Indice	Año	Término
1vez_tomada	0,761	1999	2
2vez_tomada	0,172	1999	2
3vez_tomada	0,055	1999	2
4vez_tomada	0,012	1999	2

Figura 7. Índices por Año y Término Académico

Olap DW Sistema de Informacion - ICM  
 Materias | Estudiantes | Ayuda | Contratar | Indices | salir

Tabela dinámica

G15

	A	B	C	D	E	F	G	H	I
1									
2	Ver...								
3									
4									
5									
6	Anio	Termino	Nombre	1vez tomada	2vez tomada	3vez tomada	4vez tomada	5vez tomada	Total general *
7	2000	1		1793	290	120	10		2213
8		2		1763	445	96	38		2342
9		3		259	120	32	18		429
10	Total 2000 *			3815	855	248	66		4984
11	Total general *			14746	3426	1008	217	1	19398
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									

Hoja1 / Hoja2 / Hoja3

Figura 8. Consulta multidimensional por tiempo, estado de la materia , medido por materias registradas



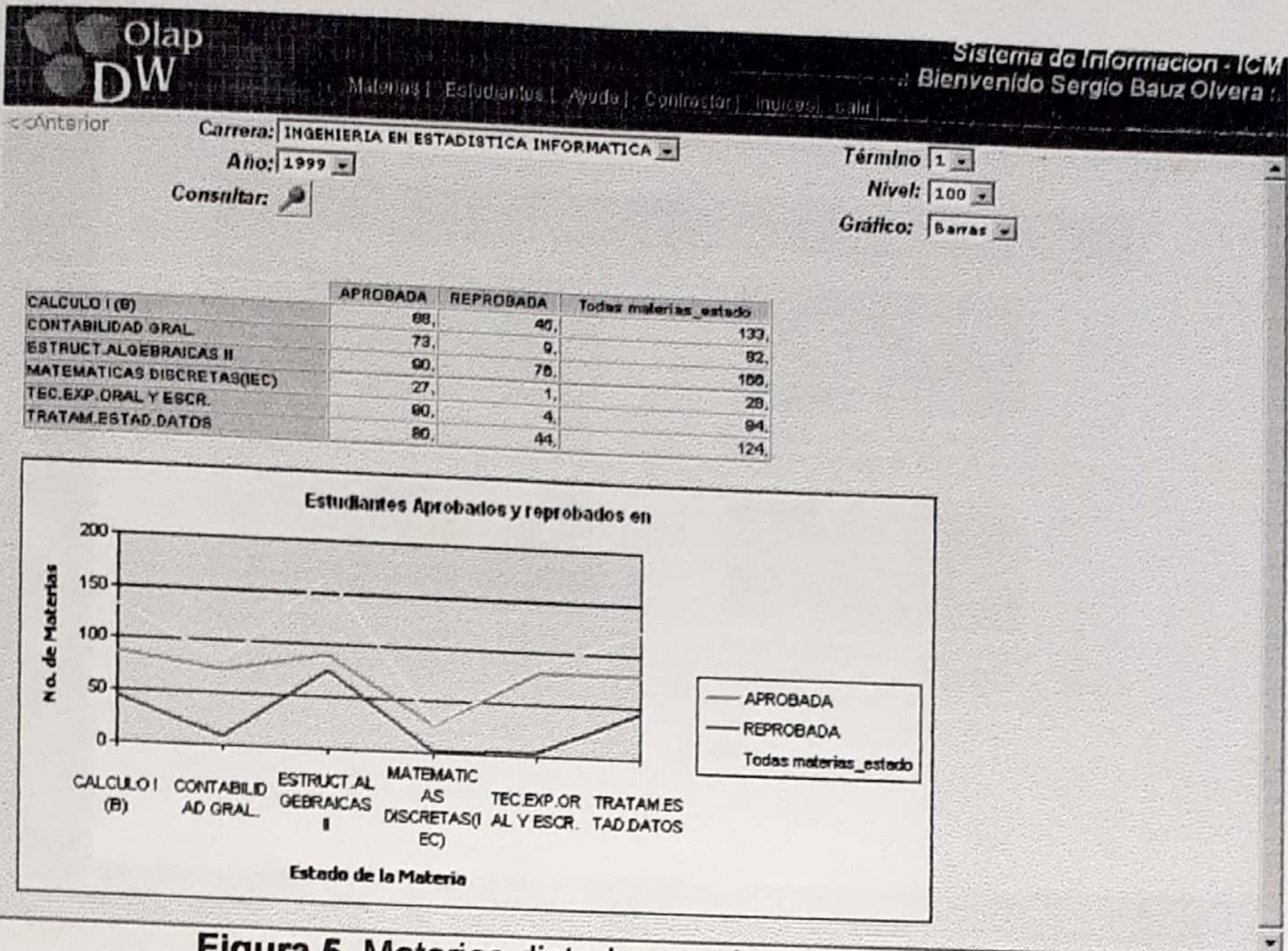


Figura 5. Materias dictadas según Período Específico

Olap DW Sistema de Informacion - ICM Bienvenido Sergio Bauz Olvera

Materias | Estudiantes | Ayuda | Contratar | Indices | salir

# Mat. Aprobadas Nota Prom. Alto Por Año de Ingreso

Mensaje del Sistema [Listado de Estudiantes] 07/07/2004 20:20:37

Registros encontrados: 1 a 15 de 479 INGENIERIA EN ESTADISTICA INFORMATICA

No. Matricula	Estudiante	Proc.	Factor P	Tip./Coleg	Sexo	# Mat.
1.- 199904269	ZAMBRANO LOOR KAREN JAHAIRA	CALCETA		8	Fiscal Femenino	44
2.- 199918426	PLUA MORAN RAQUEL PATRICIA	GUAYAQUIL		3	Fiscal Femenino	43
3.- 199914193	ANDRADE ROSERO ALLISON VANESSA	GUAYAQUIL		4	Fiscal Femenino	43
4.- 199914342	DURAN VINUEZA WALTER FRANCISCO	GUAYAQUIL		29	Fiscal Masculino	42
5.- 199918442	SUAREZ HERNANDEZ CARLOS ANIBAL	GUAYAQUIL		8	Fiscal Masculino	42
6.- 199914235	ARMIJOS TORO LIVINO MANUEL	GUAYAQUIL		10	Fiscal Masculino	42
7.- 199914433	LOOR QUEVEDO JORGE SIMON	GUAYAQUIL		18	Fiscal Masculino	41
8.- 199914631	VELIZ QUINTERO GRACIELA MARIA	GUAYAQUIL		13	Fiscal Femenino	41
9.- 199914268	BAQUERIZO BENITEZ SHULLY MARILYN	GUAYAQUIL		5	Fiscal Femenino	41
10.- 199914466	MONTANEZ MOSCOSO CAROLINA LUCIA	VENEZUELA		8	Fiscal Femenino	40
11.- 199914615	VASQUEZ VILLOW VANESSA VIVIANA	GUAYAQUIL		3	Fiscal Femenino	39
12.- 199904194	SANCHEZ CERON JOFFRE ERNESTO	GUAYAQUIL		3	Fiscal Masculino	39
13.- 199912181	ARAUJO HERRERA CARLOS ALBERTO	GUAYAQUIL		6	Fiscal Masculino	38
14.- 199914657	RODRIGUEZ AROCA MARIA DE LOS ANGELE	GUAYAQUIL		14	Fiscal Femenino	38
15.- 199914573	SALAZAR GARCES DANNY RICHARD	GUAYAQUIL		11	Fiscal Masculino	38

email

Figura 6. Estudiantes con alta nota promedio por Materia y Carrera



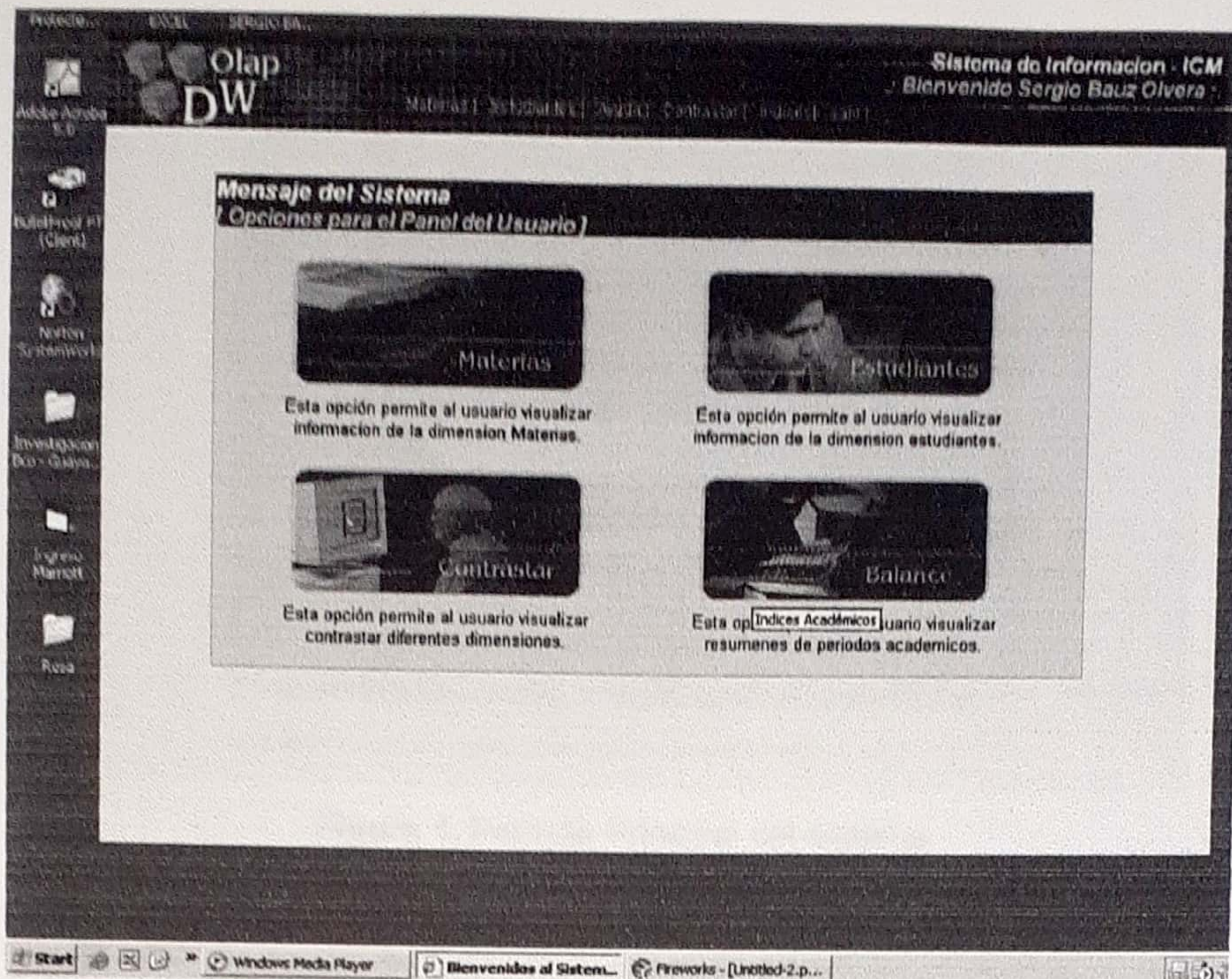


Figura 3. Pantalla de Bienvenida y Opciones Principales

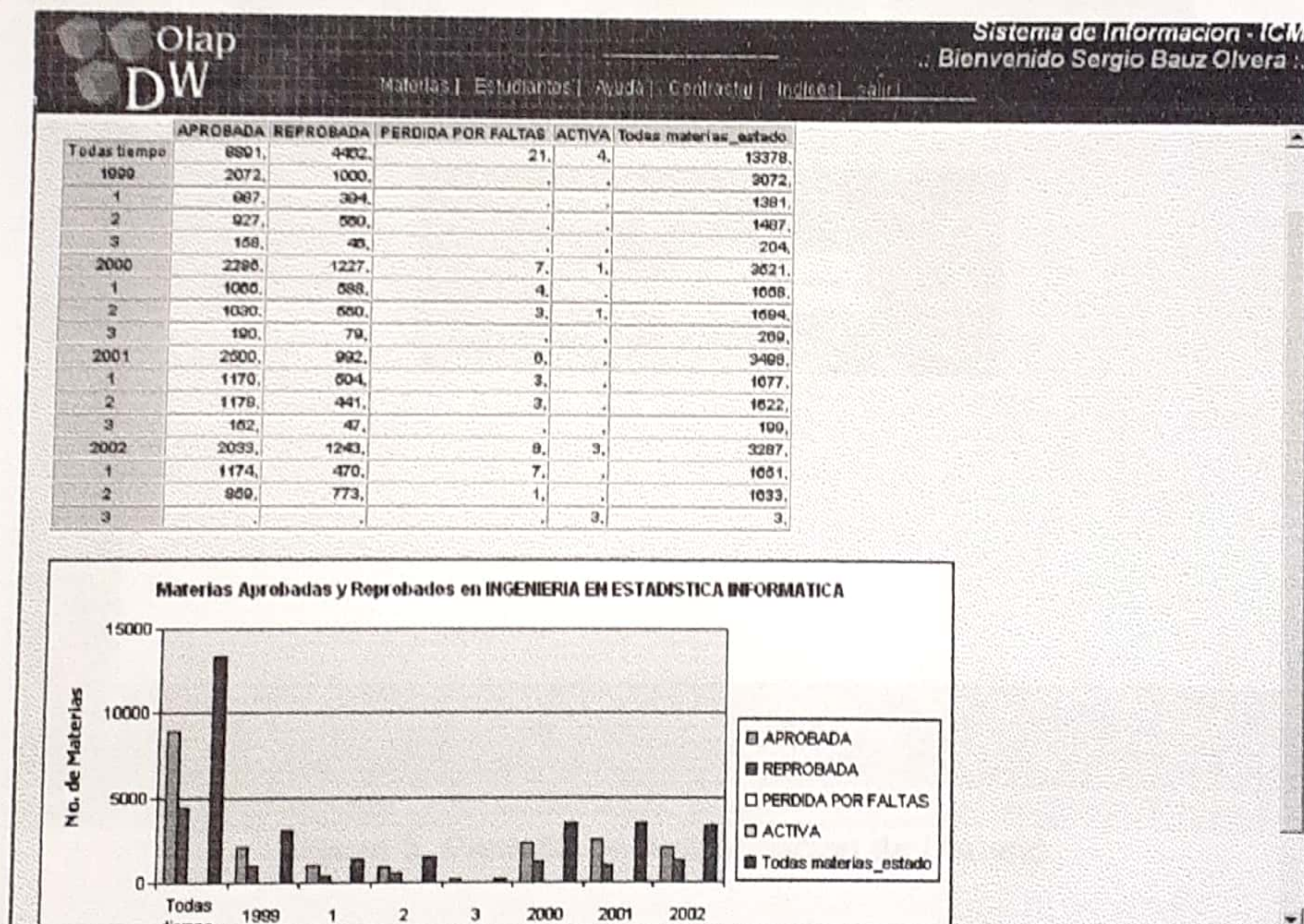


Figura 4. Estado de las Materias por Período



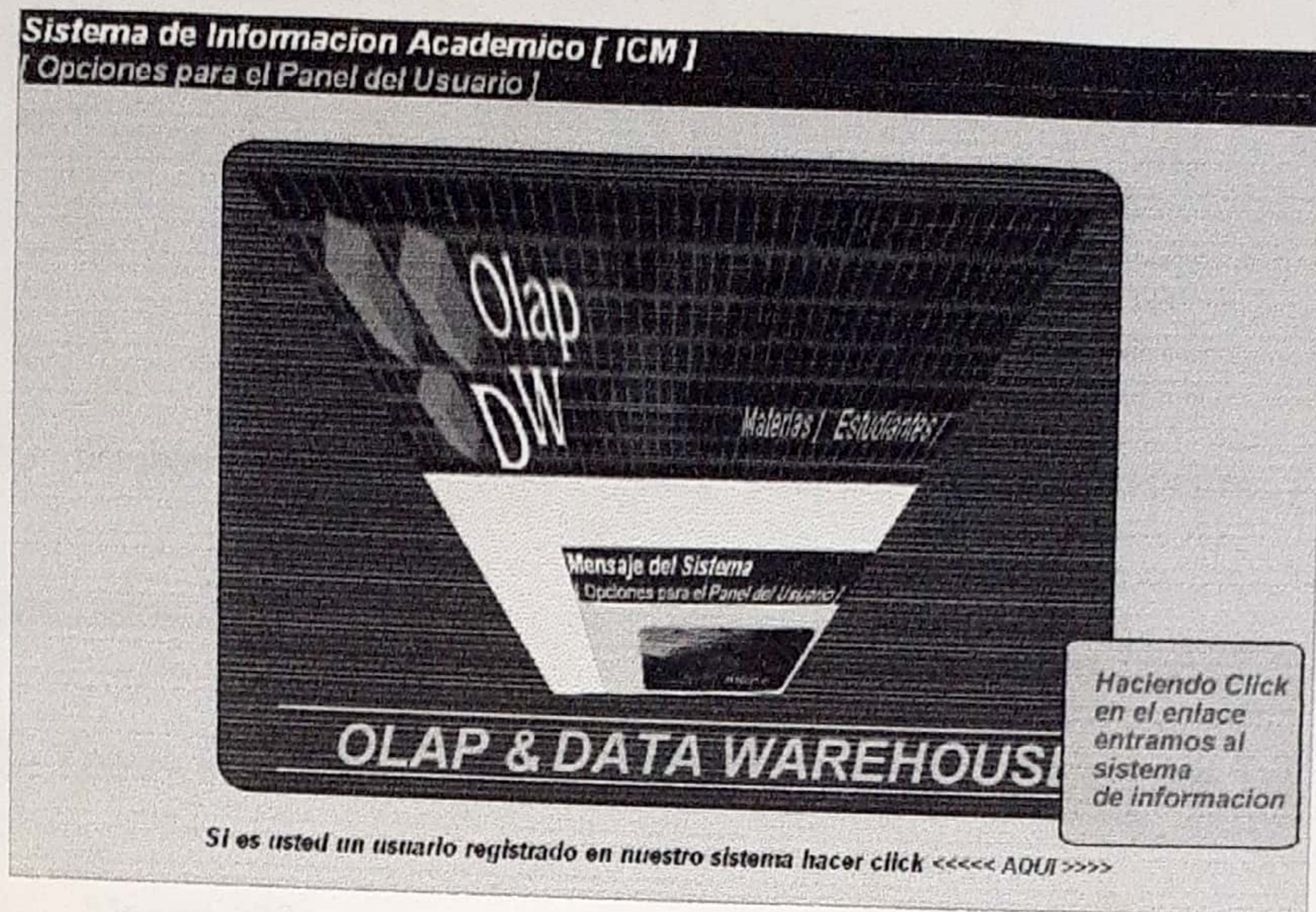


Figura 1. Pantalla Principal del sistema

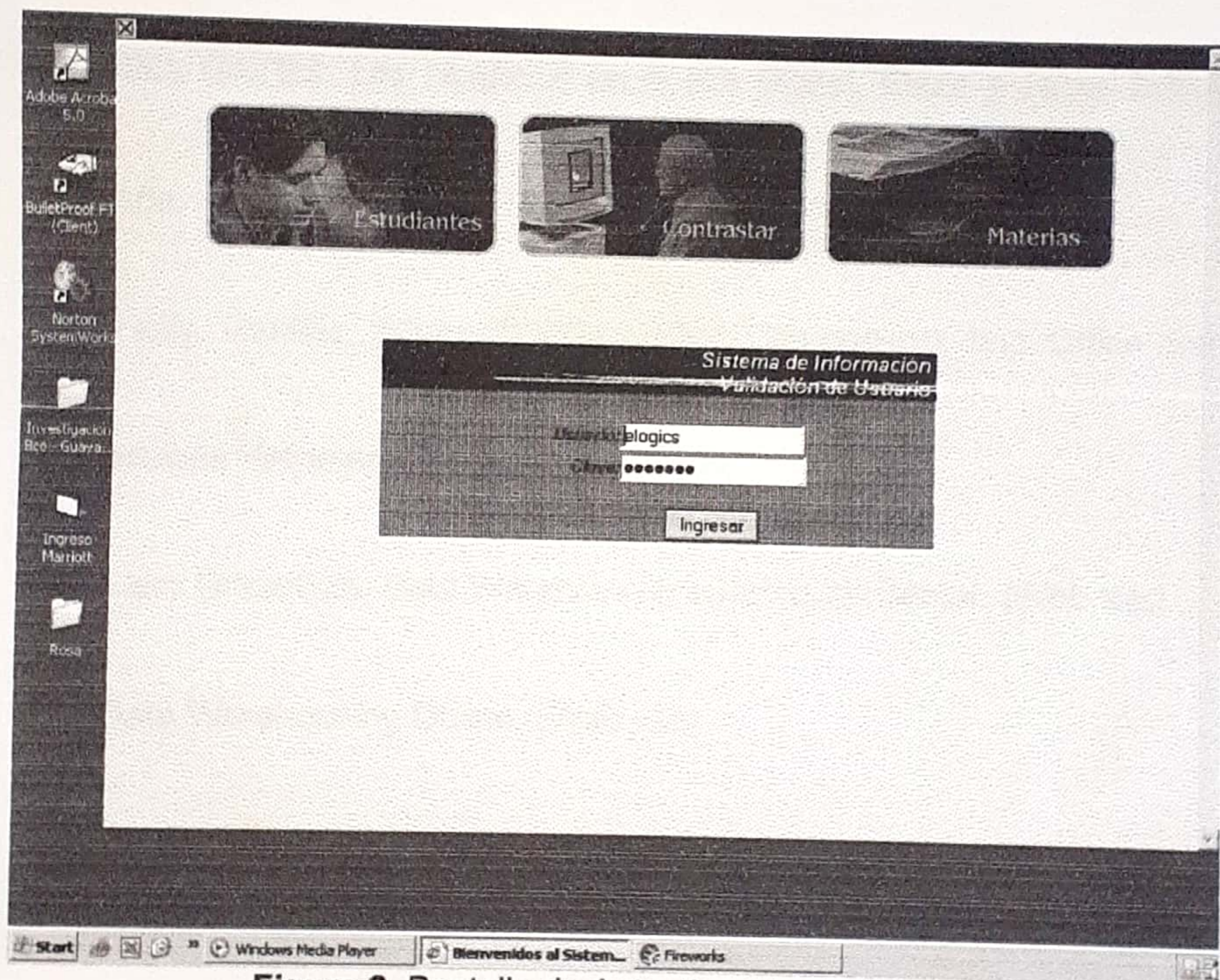


Figura 2. Pantalla de Autenticación de Usuario



## BIBLIOGRAFIA

- 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.