



Universidad Nacional Autónoma de México
Programa de Posgrado en Ciencias de la Administración

**Estudio y análisis de procesamiento de datos en línea y su
incidencia en las operaciones de las organizaciones.**

T e s i s

Que para optar por el grado de:

Maestro en Administración
(organizaciones)

Presenta:
Gabino Díaz Díaz

Tutor: M.A.I. Manuel Jáuregui Renault
Facultad de Contaduría y Administración FES - Cuautitlán

México, D. F. Junio 2014.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Poder al fin dedicar esta tesis es un sueño, hecho realidad lo hago con una alegría inmensa y espero que sirva de inspiración y motivación.

A mí querida esposa Mónica

Por su paciencia y comprensión y comprensión para que yo pudiera cumplir con el mío. Por tu bondad y sacrificio me inspiraste a ser mejor , ahora puedo decir que esta tesis lleva mucho de ti, gracias por estar siempre a mi lado.

A mis hijos Vera, Emanuel, Gabriel espero que les sirva como fuente de inspiración y los motive a realizar sus sueños.

A mi Papá y mamá

Con todo mi cariño y mi amor para las personas que hicieron todo en la vida para que yo pudiera lograr mis sueños, por motivarme y darme la mano cuando sentía que el camino se terminaba, a ustedes por siempre mi corazón y mi agradecimiento

A mis hermanos y sobrinos porque son parte fundamental de mi vida

A mis amigos Gerardo, Manuel y Oscar que siempre me han llenado de consejos útiles durante una gran parte del desarrollo de la tesis

A mis maestros que en este andar por la vida, influyeron con sus lecciones y experiencias en formarme como una persona de bien y preparada para los retos que pone la vida, a todos y cada uno de ellos les dedico cada una de estas páginas de mi tesis.

A mis tutores que con paciencia revisaron el siguiente trabajo y dieron las observaciones pertinentes para la mejora del mismo.

Especial reconocimiento merece el interés mostrado por mi trabajo y las sugerencias recibidas de la Doctora Sandra Luz González, con la que me encuentro en deuda por el ánimo infundido y la confianza en mí depositada.

Un agradecimiento muy especial merece la comprensión, paciencia y el ánimo recibidos de mi familia y amigos.

A todos ellos, muchas gracias.

PLANTEAMIENTO DEL PROBLEMA

La operación diaria en las organizaciones exige la disponibilidad de datos actualizados y vigentes por lo que se han desarrollado técnica para la disposición de los mismos en línea. Este trabajo analiza la minería de datos como una de las técnicas utilizadas en la actualidad para el análisis y toma de decisiones en la empresa.

Normalmente las organizaciones utilizan información en línea para las operaciones en tiempo real esta información es de corta vigencia irrelevante en manejo de información para hacerla útil en el mediano y corto plazo es necesario transformarla, filtrarla y consolidarla con información semejante previa o simultánea. Lo que implica el empleo de algoritmos y metodologías de vanguardia para el desarrollo de información útil a mediano y largo plazo en las operaciones de decisión táctica y estratégica, Los usuarios que toman decisiones y planifican día a día, a mediano plazo o a largo plazo, la calidad, disponibilidad y presentación de la información juegan un papel categórico. Este tipo de usuarios necesitan disponer de información tanto consolidada como detallada de cómo marchan las actividades ya cumplidas, predecir tendencias y comportamientos para tomar decisiones proactivas.

Con los sistemas tradicionales se preparan reportes ad-hoc para encontrar las respuestas a algunas las preguntas, pero se necesita dedicar aproximadamente un 60% del tiempo asignado al análisis de localización y presentación de los datos, como también asignación de recursos humanos y de procesamiento del departamento de sistemas para poder responderlas, sin tener en cuenta la degradación de los sistemas transaccionales. Esta problemática se debe a que dichos sistemas transaccionales no fueron construidos con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones.

Objetivos de la investigación

Objetivo General: El presente estudio permitirá conocer y evaluar la incidencia de las herramientas OLAP (On-Line Analytical Processing) para las tomas de decisiones.

Objetivos específicos.

- Conocer las herramientas OLAP.
- Conocer sus diferentes aplicaciones.
- Analizar sus potencialidades para el análisis de datos.

Hipótesis

La técnica de minería de datos es adaptable a las necesidades de las organizaciones educativas modernas para la toma de decisiones

INTRODUCCIÓN

Actualmente en la mayor parte de las organizaciones humanas los responsables de la toma de decisiones y planificación han utilizado información proveniente de sus sistemas operacionales, sin considerar eventos similares que suceden simultáneamente lo cual, en ocasiones, permite que dos compradores efectúen la solicitud de un mismo producto, causando problemas en la distribución y entrega de los productos (conflicto de intereses).

Otras empresas extraen su información de bases de datos históricas, considerando que el mercado el día de hoy tendrá un comportamiento similar al que tuvo en el pasado, sin considerar que las condiciones operativas del momento pueden diferir sustancialmente a las anteriores, creando conflictos en el dimensionamiento del mercado de la capacidad de respuesta de la organización

Sin embargo estas bases de datos históricas de la organización debidamente analizada permitirían reconocer patrones y tendencias de comportamiento sumamente útiles para las decisiones gerenciales vinculadas con el planteamiento y la definición de estrategias en las empresas.

Es importante que la organización vaya encaminada hacia una correcta política de mantenimiento, y gestión de las bases de datos, en ese sentido y puesto que los cambios que se producen en las tecnologías y sistemas de información rápidos en demasía. En este trabajo utilizaremos técnicas de minería de datos (Data Mining) con metodología OLAP.

Esta tecnología de información representa el último avance de las bases de datos y es idóneo para la consulta y análisis de la información procedente tanto de los sistemas transnacionales internos, como de las fuentes de información externas de interés de la organización.

Esta herramienta permitirá a los directivos de las organizaciones formular preguntas, realizar consultas y analizar los datos en el momento, forma y cantidad que se precise.

Como ejemplo de la capacidad de información de esta herramienta citamos el siguiente comentario realizado por el director ejecutivo de una cadena de supermercados dirigiéndose a la asamblea general de accionistas.

“Hace 10 años pude decir cuántas bolsas de Sabritas vendimos en Acapulco. Hoy no solo puedo contestar eso mismo sino cuanto vendimos en la ciudad de Monterrey, en el municipio de Almoloya de Juárez, en la Ciudad de Toluca, en un Supermercado local, en una promoción especial al final del pasillo 4 los viernes”.

Analizando esta aseveración, las aplicaciones para soporte de decisiones basadas en esta herramienta pueden hacer más práctica y fácil la explotación de los datos, para una mayor eficacia en la toma de decisiones de las organizaciones; dándoles así, una ventaja competitiva sobre su competencia.

CAPITULO 1 MARCO TEORICO

En el Capítulo 1 se realiza una introducción a las necesidades de información de las empresas actuales, se muestra la importancia de una buena información para la toma de decisiones y el alcance de dicha información en la empresa.

Luego se presenta una introducción a los Almacenes de Datos (Datawarehousing), Análisis analítico en línea (OLAP), Minería de Datos e Inteligencia de Negocios.

En estas secciones se muestran los principios claves de estas tecnologías para un mejor entendimiento del objetivo del proyecto, bajo qué plataforma se realiza, con qué herramientas conceptuales se cuenta, cómo se organizan e integran y cómo se utilizan estas herramientas para brindar información para la toma de decisiones.

1.1. Necesidad de Información y Conocimiento en la Empresa

Las empresas actualmente caracterizan a la información como uno de los activos de la empresa [Bitam, 2002], es así, que se comienza a tratarla, especialmente aquella relacionada con datos para tomar decisiones, de una manera más metodológica. A continuación se exponen brevemente algunos conceptos relacionados con la información y su importancia estratégica para la toma de decisiones en las empresas.

- **Las Empresas en la era de la información**

Desde que las organizaciones comenzaron a guardar los datos de sus operaciones en medios de almacenamiento físico, con el fin de permitirles una mayor administración y control de la información, ha existido una necesidad de utilizarla para atender las necesidades propias del negocio.

Cuando las empresas no tienen garantizada la venta de lo que producen, realizan un cambio paulatino hacia obtener de los datos toda la información útil y estratégica para mantenerse en el mercado, dándole un lugar preponderante a las necesidades del cliente.

Actualmente, se le da un peso específico muy importante a la información como el principal conocimiento que sostiene el negocio. Existen empresas que, de modo

predominante, ofrecen servicios y giran su negocio principal sobre el manejo de la información (bancos, aseguradoras, casas de bolsa, etc.), en ellas es fácil identificar la importancia de la información, si no existiera ésta dejarían de existir. Sin embargo, hay otras en las que su giro principal es alrededor de la producción, en ellas la información debe identificarse para analizar y perfeccionar su producción (porcentajes de desecho, líneas de producción, distribución de materias, suministro, inventarios y almacenes, procesos internos, publicidad y mercadotecnia, preferencias del cliente, etc.). De hecho, en cualquier empresa se está tratando de convertir, por todos los medios posibles, esa información en conocimiento que mejore los procesos y, a su vez, se traduzca en ventajas competitivas en los mercados.

La idea de las empresas sedientas de información no surge de manera espontánea, en realidad desde que se almacenan los datos debe entenderse que tendrían un fin utilitario en algún momento, caso contrario, cualquier dato de control sería desechado instantáneamente. Lo que si surge de súbito es la imprescindible necesidad de dar respuesta rápida a los requerimientos de información para la toma de decisiones para ayudar a mejorar de alguna manera los procesos internos de negocio [Bitam, 2002].

- **El Valor de la Información**

En la época actual, que se caracteriza por un crecimiento exponencial de las nuevas tecnologías de la información y las telecomunicaciones, los activos más valiosos de una empresa ya no son activos tangibles o los depósitos en los bancos, sino los conocimientos, habilidades, valores y actitudes de las personas que forman parte de una empresa. De hecho, para generar riqueza es suficiente tener conocimiento sobre un tema determinado y explotarlo de la mejor manera posible. Los factores de la producción como capital, tierra y trabajo, han sido sustituidos por el capital intelectual, que comprende todos aquellos conocimientos tácitos o explícitos que generan valor económico para la empresa.

Prácticamente nadie cuestiona el hecho de que vivimos en la Era de la Información y que la información tiene un valor concreto en pesos, esto se evidencia porque existen empresas cuyo único negocio es alrededor de la venta de información, como por

ejemplo Gartner Group, Empresas de Internet y Amazon, entre otras. En mercadotecnia, el conocimiento es el único camino posible para sostener ventajas competitivas. Es más, en la actualidad, la información y el conocimiento son considerados como el capital intelectual que soporta la riqueza de una organización.

Para identificar el valor concreto de la información en la organización se pueden realizar dos evaluaciones [Bitam, 2002] :

a) Todas las corporaciones tienen un modelo de negocios basado en la información que poseen, por ejemplo: ¿qué influencia la compra y la demanda?, ¿en dónde hay oportunidades de negocio?, ¿qué es lo que mueve la calidad del producto y la demanda de los clientes? A medida que esta información se vuelve más exacta, la capacidad de la empresa para competir se incrementa. Visto así, la información corporativa es claramente un activo de la empresa que genera valor y su inexistencia genera pérdidas en caso de que existiera la información y ésta desapareciera, o bien, "no ganancias" en caso de que no exista.

b) Otra forma de entender la información como dinero es mediante su transformación en conocimiento tácito o explícito. El conocimiento tácito es el que tienen las personas producto de la experiencia, los estudios y la educación; los conocimientos explícitos son los que se almacenan en medios magnéticos como cintas y disquetes.

En el momento que una persona decide cambiar de empleo se está llevando consigo información, conocimientos y está vendiendo su fuerza intelectual por un mayor precio; el campo laboral nos indica que la fuerza de trabajo intelectual aumenta su costo con dos factores básicos que generan conocimiento: la experiencia y la educación.

Por su parte, si un sistema que posee información eventualmente desaparece o falla, generará pérdidas a la empresa, incluso por cada minuto que esté detenido. En la actualidad las empresas están apostando mucho por la tecnología y los individuos para que juntos tengan un conocimiento suficiente que acerque la visión interna de ambos a la realidad exterior, en la medida que esa brecha disminuye, las decisiones tomadas se acercan más a la realidad exterior, generando decisiones más correctas y en menos tiempo; si la brecha aumenta, puede ocasionar grandes pérdidas para la organización.

Es fácil entenderlo, suponiendo una situación hipotética en la cual un nuevo auto es diseñado con lujo, pero con algunos toques de un auto deportivo y, sin realizar ningún

tipo de estudio previo más que la intuición y el sentido común, se pretende lanzarlo para que sea adquirido por adultos mayores de 30 años. Para ello, una vez que se encuentra listo para la distribución, comienzan las campañas de publicidad y presentaciones orientadas precisamente a ese mercado potencial. Al cabo de cierto tiempo se dan cuenta que las campañas que lanzaron no han tenido mucho impacto en ese segmento, pero curiosamente un porcentaje similar de las ventas a la fecha se han dado en personas entre 25 y 30 años. La realidad indica que ese auto tiene un impacto mayor en un segmento distinto al que suponía.

En caso de haber tenido información suficiente sobre las preferencias de los distintos segmentos, la historia de las ventas y, sobre todo, un estudio previo de mercado se habría sabido con anticipación hacia dónde dirigir los esfuerzos de la publicidad con dos resultados benéficos: en primer lugar, la publicidad no habría sido inefectiva y el dinero utilizado en las campañas no se habría desperdiciado; y en segundo lugar, se habría atendido a los verdaderos clientes potenciales, con lo cual las ventas habrían sido mayores. El ejemplo es hipotético, pero la situación es muy similar a la cotidianeidad, muchas empresas utilizan el sentido común y la intuición para tomar decisiones, la información que se traduce en conocimientos acerca la visión interna a la realidad y esa diferencia existente es la que puede representar miles o millones de pesos. Lo que se pretende es acercar el mundo real a la visión interna para generar ganancias, para convertir la información en utilidades, para darle un valor a la información.

Si la información es un activo, debemos poder asignarle un valor en pesos. La pregunta que surge inmediatamente es cómo podemos asignarle un valor en pesos a la información. Dado un mercado libre, la primera respuesta es que el valor de la información es lo que en el mercado se pague por ella. Este recurso simple, basado en el valor percibido, muchas veces es suficiente para asignarle un valor a la información, sin embargo, no es suficiente en otros casos, por ejemplo, en el caso de una pieza de información que no vende y que es utilizada únicamente en procesos internos de toma de decisiones.

La importancia de una buena información puede ser vista como la diferencia en valor entre una decisión correcta y una decisión equivocada, en donde la decisión está

basada en esa información. Mientras más grande sea esa diferencia entre decisión correcta y errónea, mayor será la importancia de contar con una buena información [Bitam, 2002].

- **¿Por qué las Organizaciones Requieren Distintos Sistemas de Información**

Para tener completamente automatizada a la empresa es necesaria una gran infraestructura en tecnología que soporta sistemas de información. Este crecimiento tecnológico tiene distintos orígenes, que van desde la implementación, crecimiento, ampliación, integración, etc. Las condiciones actuales de los mercados han provocado, la necesidad de tecnología cada vez más avanzada para responder a las peticiones muy particulares de información.

Sistemas de Procesamiento de Datos (SPD o OLTP), Sistemas de Manufactura, Administración de Recursos Empresariales (ERP), Sistemas de Información Ejecutiva (EIS), Sistemas de Soporte a las decisiones (DSS), Sistemas Gerenciales, Manejo de Relación con Clientes (CRM), Suministro de la Cadena de Distribución (SCM), son algunos de los sistemas que surgen, se ponen de moda y luego algunos desaparecen acorde a la evolución de las empresas. Lo que es un hecho, independientemente del enfoque que esté de moda o sea más útil en el momento, es que los datos siempre serán almacenados en bases de datos y esos datos serán el soporte total a las decisiones de la empresa.

Muchos negocios requieren información de su actividad específica, por ejemplo, los ERP (Administración de Recursos Empresariales) son sistemas muy grandes y complejos en donde gran parte de su contenido se dedica a la producción, sería ilógico adquirir un sistema tan complejo y costoso si la empresa se dedica a los bienes raíces. En ese mismo sentido existen desarrollos comercializados como productos que solo son configurados en una organización en particular, pero tienen el funcionamiento mínimo necesario para cierta industria. Hay software para la industria automotriz, software para hoteles, comercio minorista, transporte, software educativo, entre otros. El motivo por el cual son distintas las herramientas utilizadas obedece a que las

actividades de misión crítica, que soportan cada una de las industrias son diferentes y como tal, también es distinto el tipo de información que puede solicitar un directivo en cada una de las industrias, motivo por el cual hay muchos productos de software dedicados a explotar la información de las bases de datos que no tienen características estándares, sino más bien son adaptables a cada situación.

Considerando las distintas necesidades en cada actividad, es fácil extrapolar la misma situación a cada empresa, incluso con actividades similares, pero lo importante es entender el último nivel en cuanto a la diferenciación de la información solicitada.

La información que fluye en una empresa está destinada a responder a diversos tipos de preguntas de sus usuarios, de ahí la necesidad que existan sistemas de información para requerimientos muy específicos que permitan la recolección y el manejo de datos.

En el interior de una empresa, los puestos son factores importantes para determinar la información que comúnmente es requerida por la gente. Los sistemas de procesamiento de datos (OLTP) hacen uso de medios de almacenamiento y técnicas para poblarlos. La mayoría de las empresas, por la cantidad de información que manejan, se basan en los OLTP para guardar muchos datos y tener tiempos de respuesta cortos a los cientos de transacciones realizadas cotidianamente, sin embargo, la eficiencia no es para la consulta masiva de grandes cantidades de información y mucho menos para el análisis de la misma.

La tecnología ha tenido que adaptar los medios de que se vale para que sean eficientes en el ámbito específico de aplicación, tanto para el diseño de estructuras de datos que ordenen la información como se desea, como en las herramientas o software que permite solucionar en tiempo y forma lo que el usuario demanda. Es importante resaltar que todos los Sistemas de Información tienen un fin muy particular, y se complementan para sostener, de la manera más eficiente, un negocio; sin embargo, no todos pueden solucionar las distintas demandas de los usuarios, precisamente porque son diseñados para alguna área de aplicación muy específica.

El motivo por el cual existen varios sistemas de información es porque los usuarios tienen preguntas muy específicas que no cualquier sistema puede resolver. De hecho, las bases de datos operacionales, que son las indispensables en cualquier organización, no están organizadas para responder a preguntas globales sino a

pequeños grupos de datos. Preguntas que involucren consultas complejas podrían resolverse en un lapso extenso, en el cual cabe la posibilidad de que la vigencia desaparezca. Lo importante es destacar que una base de datos o sistema de información no tienen la capacidad de resolver las necesidades informativas de toda la organización [Bitam, 2002].

- **Información que las Empresas Necesitan**

La tendencia de las organizaciones actuales es demandar información en los niveles donde antes la administración se basaba en la intuición y el sentido común para tomar decisiones. A pesar de que en los niveles operativos siempre se ha demandado información, históricamente no ha existido restricción alguna para brindarla al usuario. Más bien los mercados dinámicos han obligado a las empresas para que la información estratégica sea puesta en las computadoras de los directivos, este comportamiento se ha generalizado principalmente motivado por la facilidad y utilidad de la información compartida.

En estos momentos la información fluye en todos los niveles de la organización con diferentes fines (comunicación, control, administración, evaluación, etc.) independientemente de los puestos. Las empresas están entendiendo que los niveles directivos tienen una gran responsabilidad al tomar decisiones, pues el impacto que generan recae sobre toda la organización, pero también existen más personas que toman decisiones y, a pesar de que éstas no tienen un impacto global, deben ser también correctas y oportunas, pues ciertos grupos dependen de las mismas.

En una organización todos los niveles en donde toman decisiones deben tener suficiente información para apoyarse en su trabajo cotidiano, el lugar que ocupen en la pirámide organizacional se vuelve secundario cuando el enfoque es hacia el manejo de procesos y todos los puestos tienen cierta relación y dependencia entre sí.

De modo general en una pirámide organizacional, los requerimientos informativos se dividen en tres partes:

1.- Información Estratégica

Está orientada principalmente a soportar la Toma de Decisiones de las áreas directivas para alcanzar la misión empresarial. Se caracteriza porque son sistemas sin carga periódica de trabajo y sin gran cantidad de datos, sin embargo, la información que almacenan está relacionada a un aspecto cualitativo más que cuantitativo, que puede indicar como operará la empresa ahora y en el futuro, el enfoque es distinto, pero sobre todo es distinto su alcance. Se asocia este tipo de información a los ejecutivos de primer nivel de las empresas.

Un punto importante es que la información estratégica toma grandes cantidades de datos de áreas relacionadas y no se enfoca específicamente hacia una sola, de ahí que las decisiones que puedan ser tomadas impactan directamente sobre toda la organización.

2.- Información Táctica

Información que soporta la coordinación de actividades y el plano operativo de la estrategia, es decir, se plantean opciones y caminos posibles para alcanzar la estrategia indicada por la dirección de la empresa. Se facilita la gestión independiente de la información por parte de los niveles intermedios de la organización. Este tipo de información es extraída específicamente de una área o departamento de la organización, por lo que su alcance es local y se asocia a gerencias o subdirecciones.

3.- Información Técnico Operacional

Se refiere a las operaciones tradicionales que son efectuadas de modo rutinario en las empresas mediante la captura masiva de datos y Sistemas de Procesamiento Transaccional. Las tareas son cotidianas y soportan la actividad diaria de la empresa (contabilidad, facturación, almacén, presupuesto y otros sistemas administrativos). Tradicionalmente se asocian a las Jefaturas o Coordinaciones operativas o de tercer nivel.

Si consideramos factores internos y externos de una organización podríamos concluir que los requerimientos actuales se orientan a conocer y mejorar los costos de toda la cadena económica.

Estos requerimientos se reflejan en el interés por tener a la mano los diagnósticos que arrojen información específica y clave para determinada área de conocimiento, en el menor tiempo posible. La tendencia es que las áreas directivas necesitan en su escritorio la información clave de su empresa; en todos los niveles el requerimiento es similar aunque, evidentemente, tiene objetivos diferentes.

El paradigma de la información exclusiva en los niveles directivos para apoyar la toma de decisiones no es obsoleto, simplemente se debe mejorar y complementar agregando la información también en otros niveles medios y jefaturas, o sea, en cualquier persona que tenga el poder de tomar decisiones [Bitam, 2002].

1.2. Introducción a Almacenes de Datos y Datamart

Hoy en día las empresas trabajan una gran cantidad de procesos automatizados, los cuales manejan gran cantidad de datos en forma centralizada. En esta información descansa el conocimiento de la empresa, constituyendo un recurso corporativo primario y parte importante de su patrimonio.

El nivel competitivo alcanzado en las empresas les ha exigido desarrollar nuevas estrategias de gestión. En el pasado, las organizaciones fueron típicamente estructuradas en forma piramidal con información generada en su base fluyendo hacia arriba; y era en la punta de la pirámide donde se tomaban decisiones a partir de la información proporcionada por la base, con un bajo aprovechamiento del potencial de esta información. Las empresas han reestructurado y eliminado diferentes niveles de estas pirámides y han autorizado a los usuarios de todos los niveles a tomar mayores decisiones y responsabilidades. Sin embargo, sin información sólida para ayudar y apoyar las decisiones, la automatización no tiene sentido.

Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de Almacenes de Datos.

El énfasis no está sólo en llevar la información hacia lo alto sino a través de la organización, para que todos los empleados que la necesiten la tengan a su disposición [Sperley, 1999].

El AD (de ahora en adelante los términos Almacenes de Datos, Almacén de Datos y AD serán utilizados en forma indistinta) convierte entonces los datos operacionales de una organización en una herramienta competitiva, por hacerlos disponibles a los empleados que lo necesiten para el análisis y toma de decisiones.

El objetivo del AD es el de satisfacer los requerimientos de información interna de la empresa para una mejor gestión. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de los analistas y usuarios tomadores de decisiones. El AD es el lugar donde la gente puede acceder a sus datos.

El AD puede verse como una bodega donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse fácilmente según se necesiten.

Los sistemas transaccionales son dinámicos, constantemente se encuentran actualizando datos. Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos (snapshots, en inglés), para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo. Llevar a cabo un análisis complejo sobre un sistema transaccional, puede resultar en la degradación del sistema, con el consiguiente impacto en la operación del negocio.

Los almacenes de datos generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales. Muchos Almacenes de datos se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar. Así un Almacenes de datos resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo análisis multidimensional.

De esta forma, dentro de un Almacén de datos existen dos tecnologías que se pueden ver como complementarias, una relacional para consultas y una multidimensional para análisis [Sperley, 1999].

Puede considerarse que el modelo relacional en el cual se basa Procesamiento Transaccional en Línea (OLTP *OnLine Transactional Processing*), tiene como objetivo mantener la integridad de la información (relaciones entre los datos) necesaria para operar un negocio de la manera más eficiente. Sin embargo, este modelo no corresponde a la forma como el usuario percibe la operación de un negocio.

AD está basado en un procesamiento distinto al utilizado por los sistemas operacionales, es decir, este se basa en OLAP Procesos de Análisis en Línea (OnLine Analysis Process), usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

Transaccionales	Basados en Almacenes de Datos
Contienen grandes cantidades de datos, incluidos los datos extensivos utilizados para comprobar transacciones.	Contienen grandes cantidades de datos, sumariados, consolidados y transformados. También de detalle pero solo los necesarios para el análisis.
Tienen estructuras de base de datos Complejas.	Tienen estructuras de Base de datos Simples.
Se ajustan para dar respuesta a la actividad transaccional.	Se ajustan para dar respuesta a la actividad de consultas.
Proporcionan la infraestructura tecnológica necesaria para admitir las operaciones diarias de la empresa.	Proporcionan la infraestructura tecnológica necesaria para admitir análisis de los datos de la empresa.
Los analistas carecen de la experiencia técnica necesaria para crear consultas "ad hoc" contra la compleja estructura de datos.	Pueden combinar datos de orígenes heterogéneos en una única estructura homogénea y simple, facilitando la creación de informes y consultas.
Las consultas analíticas que resumen grandes volúmenes de datos afectan negativamente a la capacidad del sistema para responder a las transacciones en línea.	Organizan los datos en estructuras simplificadas buscando la eficiencia de las consultas analíticas más que del proceso de transacciones.

El rendimiento del sistema cuando está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.	Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis sin interferir en la operatoria transaccional diaria.
Los datos que se modifican con frecuencia interfieren en la coherencia de la Información analítica.	Proporcionan datos estables que representan el historial de la empresa. Se actualizan periódicamente con datos adicionales, no con transacciones Frecuentes.
La seguridad se complica cuando se combina el análisis en línea con el proceso de transacciones en línea.	Simplifican los requisitos de seguridad.

1.3 Principales diferencias entre los sistemas Transaccionales (OLTP) y los basados en Almacenes de datos

- **Introduccion a Datamarts**

El acceso a los datos de toda la empresa a veces no es conveniente (o necesario) para determinados usuarios que solo necesitan un subconjunto de estos datos, en estos casos se utilizan los Datamarts. El concepto Datamart es una especialización de los Almacenes de Datos, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando [Poe et al., 1998].

Los principales beneficios de utilizar Datamarts son:

- Acelerar las consultas reduciendo el volumen de datos a recorrer
- Estructurar los datos para su adecuado acceso por una herramienta
- Dividir los datos para imponer estrategias de control de acceso
- Segmentar los datos en diferentes plataformas hardware
- Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

- **Arquitectura Almacenes de Datos**

Antes de describir la arquitectura Almacenes de datos se señala lo siguiente el término Almacenes de datos se utiliza indistintamente para hablar de la arquitectura en sí como

también para uno de los componentes que la conforman, específicamente el que tiene relación con el almacenamiento físico de los datos [Poe et al., 1998].

- **La estructura básica de la arquitectura AD incluye:**

1. Datos operacionales: un origen o fuente de datos para poblar el componente de almacenamiento físico AD. El origen de los datos son los sistemas transaccionales internos de la organización como también datos externos a ésta.

2. Extracción de Datos: selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico AD.

3. Transformación de datos: procesos para sumar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a temas e integración principalmente.

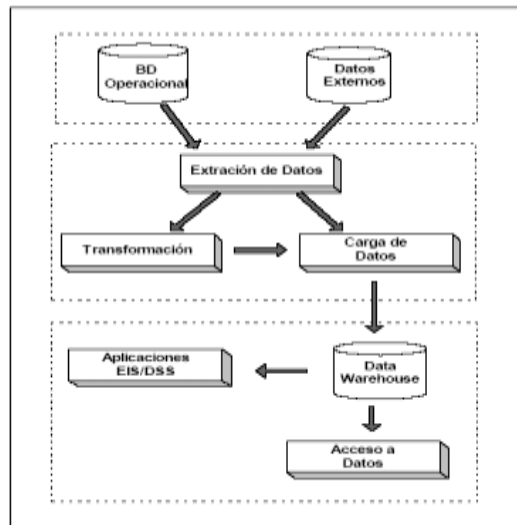
4. Carga de Datos: inserción sistemática de datos en el componente de almacenamiento físico AD.

5. Almacenes de Datos: almacenamiento físico de datos de la arquitectura AD.

6. Herramientas de Acceso al componente de almacenamiento físico AD: herramientas que proveen acceso a los datos.

Estas herramientas pueden ser herramientas específicas de mercado para visualización de bases multidimensionales almacenadas en Almacenes de datos como

también aplicaciones desarrolladas dentro de la organización del tipo EIS/DSS .



La figura 1-1 muestra la estructura básica:

Los pasos 2, 3 y 4 considerados en la figura anterior, conforman el proceso conocido como ETT (Extracción, Transformación y Transporte).

1.4 Implementación del Almacenes de datos

La forma en la cual se estructure el almacenamiento de datos AD, genera una clasificación respecto a la forma de implementar una arquitectura AD. La estructura adoptada para el Almacenes de datos se debe realizar de la manera que mejor satisfaga las necesidades empresariales, siendo entonces dicha elección factor clave en la efectividad del AD. Las implementaciones más utilizadas son [Sperley, 1999]:

- EL AD central: es una implementación de un solo nivel con un solo almacén para soportar los requerimientos de información de toda la empresa. En el AD central todos los usuarios de la organización acceden a la misma base de datos (figura 1-2).

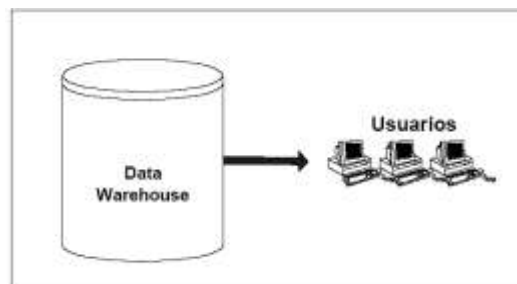


Figura 1-2

- El AD distribuido: es también una estructura de un nivel, pero particiona el almacén para distribuirlo a nivel departamental. En el AD distribuido, cada departamento, área o línea de negocio dispone de una base de datos propia con la información que solo les compete a los usuarios pertenecientes a estas áreas (figura 1-3).

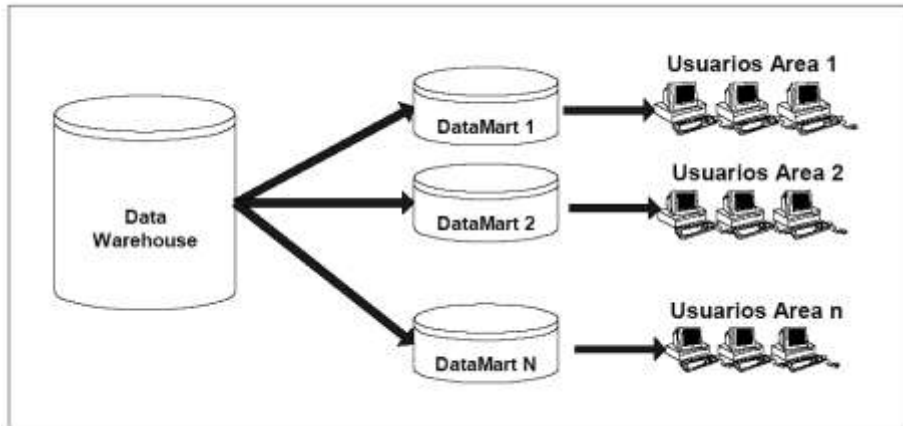


Figura 1-3

- El AD de dos niveles: combina ideas de los dos anteriores, se implementa el almacén empresarial como los departamentales. En el AD de dos niveles se dispone de una base de datos, generalmente de detalle o de información común a todos los usuarios y además cada departamento, área o línea de negocio dispone de su propia base de datos (figura 1-4)

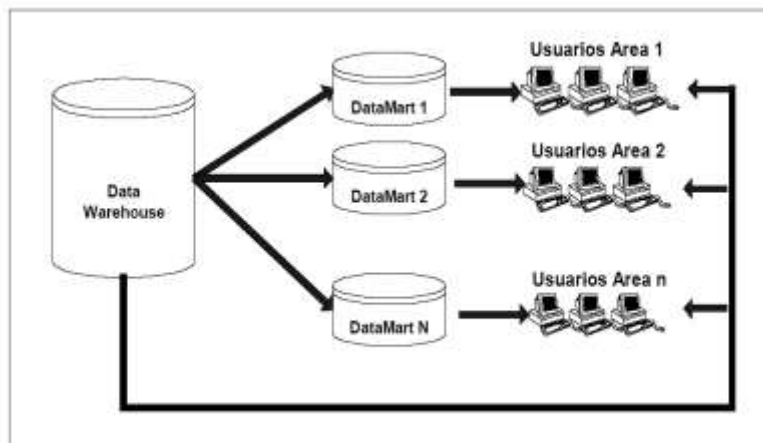


Figura 1-4

1.5 Costos e Impacto de un AD

A continuación se describen los típicos costos en que se incurre cuando se construye y se administra un Almacén de Datos También se describen los impactos que se tiene en la organización al introducir un Almacenes de Datos para la toma de decisiones.

• Costos de Construcción

Los costos de construir un AD son similares para cualquier proyecto de tecnología de información. Estos pueden ser clasificados en tres categorías [Sperley, 1999]:

- RRHH: los usuarios que participen del desarrollo deben contar con un enfoque fuerte sobre el conocimiento del área de la empresa y de los procesos empresariales. El desarrollo del AD requiere participación de la gente de negocios como de los especialistas tecnológicos; estos dos grupos deben trabajar juntos, compartiendo su conocimiento y destrezas para enfrentar los desafíos de desarrollo del AD.
- Tiempo: se debe establecer el tiempo no tan solo para la construcción y entrega de resultados del AD, sino también para el planeamiento del proyecto y la definición de la arquitectura. El planeamiento y la arquitectura, establecen un marco de referencia y un conjunto de estándares que son críticos para la eficacia del AD.
- Tecnología: muchas tecnologías nuevas son introducidas por el AD. El costo de la nueva tecnología puede ser tan sólo la inversión inicial del proyecto. Se deben tener en consideración las siguientes herramientas de soporte:
 - Soporte al Almacenes de datos
 - Soporte OLAP,
 - Soporte de Minería de datos y Soporte al EIS/DSS (deben incluir OLAP y Data Mining) y acceso a datos.

• Costos de Operación

Una vez que está construido y entregado un AD debe ser mantenido y actualizado para que tenga valor empresarial. Son justamente estas actividades de mantenimiento y

actualización, la fuente de continuos costos operacionales para un AD. Se pueden distinguir tres tipos de costos de operación

Evolutivos: ajustes continuos del AD a través del tiempo, como cambios de expectativas y cambios producto del aprendizaje de los RRHH del proyecto mediante su experiencia usando el AD.

- **Crecimiento: incrementos** en el tiempo en volúmenes de datos, del número de usuarios del AD, lo cual conllevará un incremento de los recursos necesarios como la demanda de monitoreo, la administración y la sintonización del AD.
- **Cambios: el AD:** requiere soportar cambios que ocurren tanto en el origen de datos que éste usa, como en las necesidades de la información que éste soporta. Cuando se implementa un AD, el impacto de cambios es compuesto. Existen dos orígenes primarios de cambios:
- **Cambios en el ambiente empresarial:** un cambio en el ambiente empresarial puede cambiar las necesidades de información de los usuarios. Así, el contenido del AD se puede ver afectado y las aplicaciones DSS y EIS pueden requerir cambios.
- **Cambios en la tecnología:** un cambio en la tecnología puede afectar la manera que los datos operacionales son almacenados, lo cual implicaría un ajuste en los procesos de Extracción, Transporte y Carga para adaptar las variaciones presentadas.

Un cambio de cualquiera de ellos impacta los sistemas operacionales. Un cambio en el ambiente operacional puede cambiar el formato, estructura o significado de los datos operacionales usados como origen para el AD. De esta forma serían impactados los procesos de Extracción, Transformación y Carga de datos.

Los dos primeros tipos de costos de operación, son normales en el mantenimiento de cualquier sistema de información, por lo cual no resultan ajenos; sin embargo, se debe tener especial cuidado con los costos de operación por cambios, ya que estos costos consideran el impacto producto de la relación del OLTP y del ambiente empresarial, con el AD.

1.6. Impactos del Almacenes de Datos

El éxito de AD no está en su construcción, sino en usarlo para mejorar procesos empresariales, operaciones y decisiones. Posicionar un AD para que sea usado efectivamente, requiere entender los impactos de implementación en los siguientes ámbitos [Sperley, 1999]:

- **Impactos Humanos**

La construcción del AD: a diferencia del desarrollo de aplicaciones, donde los requerimientos de la empresa logran ser relativamente bien definidos producto de la estabilidad de las reglas de negocio a través del tiempo, construir un AD depende de la realidad de la empresa como de las condiciones que en ese momento existan, las cuáles determinan qué debe contener el AD. Como se dijo anteriormente, la gente de negocios debe participar activamente durante el desarrollo del AD, desde una perspectiva de construcción y creación.

Accediendo al AD: el AD intenta proveer los datos que posibilitan a los usuarios acceder a su propia información cuando ellos la necesitan. Esta aproximación para entregar información tiene varias tareas

- a) La gente de la empresa puede necesitar aprender nuevas destrezas.
- b) Análisis extensos y demoras de programación para obtener información será eliminada. Como la información estará lista para ser utilizada, las expectativas probablemente aumentarán.
- c) Nuevas oportunidades pueden existir en la comunidad empresarial para los especialistas de información.
- d) La gran cantidad de reportes en papel serán reducidas o eliminadas.
- e) La madurez del AD dependerá del uso activo y retroalimentación de sus usuarios.

Usando aplicaciones DSS/EIS: usuarios de aplicaciones DSS y EIS necesitarán menos experiencia para construir su propia información y desarrollar nuevas destrezas. Es decir, que para los usuarios, el AD extiende el alcance de la información para que puedan acceder directamente en línea, lo que a la vez contribuye en su capacidad para

operar con mayor efectividad las tareas diarias relacionadas con la toma de decisiones. Los usuarios del AD pueden acceder a una variada información que puede ser vista de forma multidimensional, presentada como una fuente única confiable y disponible directamente por medio de sus estaciones de trabajo. Como se dijo anteriormente, los usuarios pueden usar sus herramientas familiares, hojas de cálculo, procesadores de textos y software de análisis de datos y análisis estadístico para manipular y evaluar la información obtenida desde el AD.

- **Impactos Empresariales**

Procesos Empresariales y Decisiones Empresariales.

Se deben considerar los beneficios empresariales potenciales de los siguientes impactos:

- a) Los Procesos de Toma de Decisiones pueden ser mejorados mediante la disponibilidad de información. Decisiones empresariales se hacen más rápidas por gente más informada.
- b) Los procesos empresariales pueden ser optimizados. El tiempo perdido esperando por información que finalmente es incorrecta o no encontrada, es eliminada.
- c) Conexiones y dependencias entre procesos empresariales se vuelven más claros y entendibles. Secuencias de procesos empresariales pueden ser optimizadas para ganar eficiencia y reducir costos.
- d) Procesos y datos de los sistemas operacionales, así como los datos en el AD, son usados y examinados. Cuando los datos son organizados y estructurados para tener significado empresarial, la gente aprende mucho de los sistemas de información. Pueden quedar expuestos posibles defectos en aplicaciones actuales, siendo posible entonces mejorar la calidad de nuevas aplicaciones.

Comunicación e Impactos Organizacionales.

Apenas el AD comienza a ser fuente primaria de información empresarial consistente, los siguientes impactos pueden comenzar a presentarse:

- a) La gente tiene mayor confianza en las decisiones empresariales que se toman. Ambos, quienes toman las decisiones como los afectados conocen que está basada en buena información.
- b) Las organizaciones empresariales y la gente de la cual ella se compone queda determinada por el acceso a la información. De esta manera, la gente queda mejor habilitada para entender su propio rol y responsabilidades como también los efectos de sus contribuciones; a la vez, desarrollan un mejor entendimiento y apreciación con las contribuciones de otros.
- c) La información compartida conduce a un lenguaje común, conocimiento común, y mejoramiento de la comunicación en la empresa. Se mejora la confianza y cooperación entre distintos sectores de la empresa, viéndose reducida la sectorización de funciones.
- d) Visibilidad, accesibilidad, y conocimiento de los datos producen mayor confianza en los sistemas operacionales y fomenta aún más su uso.

- **Impactos Técnicos del AD**

Considerando las etapas de construcción, soporte del AD y soporte de sistemas operacionales, se tienen los siguientes impactos técnicos:

Nuevas destrezas de desarrollo: cuando se construye el AD, el impacto más grande sobre la gente técnica está dada por la curva de aprendizaje, muchas destrezas nuevas se deben aprender, incluyendo: conceptos y estructura AD.

- a) El AD introduce muchas tecnologías nuevas (ETT, Carga, Acceso de Datos, Catálogo de Metadatos, Implementación de DSS/EIS), y cambia la manera que nosotros usamos la tecnología existente. Nuevas responsabilidades de soporte, nuevas demandas de recursos y nuevas expectativas, son los efectos de estos cambios.
- b) Destrezas de diseño y análisis donde los requerimientos empresariales no son posibles de definir de una forma estable a través del tiempo.
- c) Técnicas de desarrollo incremental y evolutivo.

d) Trabajo en equipo cooperativo con gente de negocios como participantes activos en el desarrollo del proyecto.

Nuevas responsabilidades de operación: como se expresó en el apartado anterior, los cambios sobre los sistemas y datos operacionales deben ser examinados más cuidadosamente para determinar el impacto que estos cambios tienen sobre ellos, y sobre el AD. Para la Dirección de Sistemas, el AD enriquece las capacidades del usuario autosuficiente y hace que la Dirección pueda ofrecer nuevos servicios a los usuarios, sin interferir con las aplicaciones cotidianas de producción, aunque se requiere una asignación de tiempo y personal técnico para el mantenimiento y operación del AD.

1.7 Valor del AD para la Toma de Decisiones

El valor de un AD queda descrito en tres dimensiones [Inmon & Hackathorn, 1994]:.

1- Mejorar la Entrega de Información: información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita.

2- Facilitar el Proceso de Toma de Decisiones: con un mayor soporte de información se obtienen decisiones más rápidas; así también, la gente de negocios adquiere mayor confianza en sus propias decisiones y las del resto, y logra un mayor entendimiento de los impactos de sus decisiones.

3- Impacto Positivo sobre los Procesos Empresariales: cuando a la gente accede a una mejor calidad de información, la empresa puede mejorar:

- Eliminar los retardos de los procesos empresariales que resultan de información incorrecta, inconsistente y/o no existente.
- Integrar y optimizar procesos empresariales a través del uso compartido e integrado de las fuentes de información.
- Eliminar la producción y el procesamiento de datos que no son usados ni necesarios, producto de aplicaciones mal diseñados o ya no utilizados.

1.8 Introducción al Procesamiento Analítico en Línea

La tecnología de Procesamiento Analítico en Línea OLAP (Online Analytical Processing) permite un uso más eficaz de los Almacenes de datos para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizada generalmente para sistemas de ayuda para la toma de decisiones. Primero y más importante, el OLAP presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de navegación, los usuarios finales pueden ver y entender más efectivamente la información de sus bases de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos.

En segundo lugar, el OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras de datos como cubos denominadas multidimensionales debido a que la información es vista en varias dimensiones.

Esta entrega es optimizada ya que se prepararan algunos valores calculados en los datos por adelantado, en vez de realizar el cálculo al momento de la solicitud. La combinación de navegación fácil y rápida le permite a los usuarios ver y analizar información más rápida y eficientemente que lo que es posible con tecnologías de bases de datos relacionales solamente.

El resultado final: se pasa más tiempo analizando los datos y menos tiempo analizando las bases de datos.

A pesar del proceso de almacenamiento de datos de preparar información para el consumo del usuario final se debe facilitar la búsqueda de la información.

Generalmente, las estructuras de datos de las bases tienen cierta complejidad para el usuario final, principalmente para responder a preguntas tales como: "¿Quiénes fueron los mejores vendedores de cada región durante el año pasado, mensualmente?" son complejas cuando se expresan en lenguaje SQL.

Estos retos son enfrentados con herramientas avanzadas de peticiones (queries), las cuáles ocultan al usuario final la complejidad de las base de datos.

- **Esta es la función de las herramientas OLAP.**

Todas las organizaciones tienen datos multidimensionales y la complejidad no es necesariamente una función del tamaño de la compañía. Aún a las más pequeñas compañías les gustaría poder rastrear sus ventas por producto, vendedor, geografía, cliente y tiempo. Las organizaciones han buscado durante mucho tiempo herramientas para acceder, navegar y analizar información multidimensional de una manera fácil y natural.

Las aplicaciones OLAP deberían proporcionar análisis rápidos de información multidimensional compartida. Las características principales del OLAP son [Pence & Creeth, 2002]:

- **Rápido:** proporciona la información al usuario a una velocidad constante.
- La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.
- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido “ad hoc” por el usuario.
- **Compartida:** implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.
- **Información:** acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen.

El OLAP es un componente clave en el proceso de almacenamiento de datos (data warehousing) y los servicios OLAP proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones.

- **Vistas del Usuario.**

En un modelo de datos OLAP, la información es vista como cubos, los cuáles consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios el formular peticiones complejas, arreglar datos en un reporte, cambiar de datos de resumen a datos de detalle y filtrar o seccionar los datos en subconjuntos significativos.

Por ejemplo, las dimensiones típicas de un cubo que contenga información de ventas, incluiría tiempo, geografía, producto, canal, organización y escenario (planeado o real). Las medidas típicas incluirían ventas en dólares (u otra moneda), unidades vendidas, número de personas, ingresos y gastos.

La figura 1-5 muestra un cubo con las dimensiones producto, fecha y país.

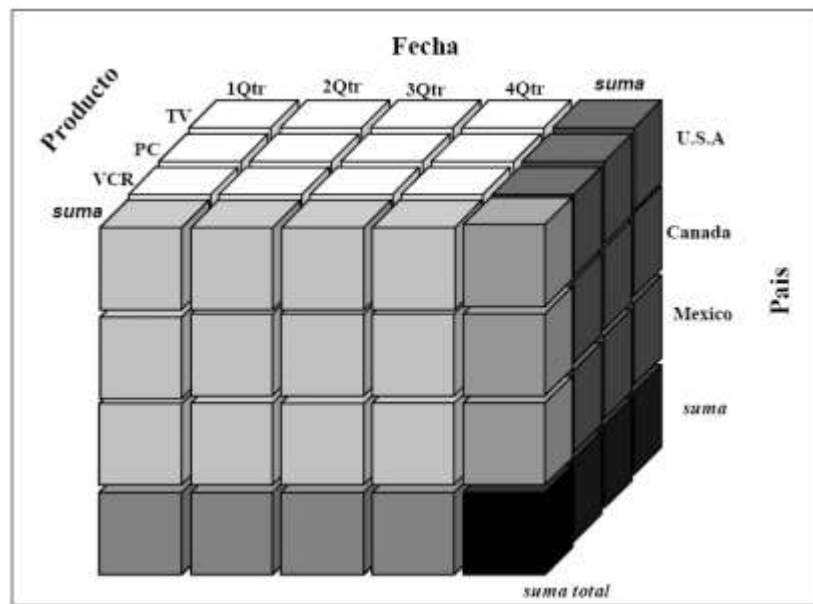


Figura 1-5. Cubo con tres dimensiones.

Dentro de cada dimensión de un modelo de datos OLAP, los datos se pueden organizar en una jerarquía que represente niveles de detalle de los datos. Por ejemplo, dentro de la dimensión de tiempo, se puede tener estos niveles: años, meses y días; de manera similar, dentro de la dimensión geografía, Se puede tener estos niveles: país, región, estado/provincia y ciudad. Una instancia particular del modelo de datos OLAP

tendrá valores para cada nivel en la jerarquía. Un usuario que vea datos OLAP se moverá entre estos niveles para ver información con mayor o menor detalle.

- **Operaciones de Usuario**

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

Normalmente este tipo de selecciones se ve reflejada en la visualización de la estructura multidimensional, en unos campos de selección que permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a una(s) dimensiones en modo tabla, pudiendo con ello realizar, entre otras las siguientes acciones [Pence & Creeth, 2002]:

- Rotar (Swap): alterar las filas por columnas (permutar dos dimensiones de análisis)
- Bajar (Down): bajar el nivel de visualización en las filas a una jerarquía inferior
- Detallar (Drilldown): informar para una fila en concreto de datos a un nivel inferior
- Expandir (Expand): ídem anterior sin perder la información a nivel superior para éste y el resto de los valores

- **Arquitectura OLAP**

Aunque en ocasiones se utilizan indistintamente, los términos Almacenes de datos y proceso analítico en línea (OLAP) se aplican a diferentes componentes de sistemas de ayuda a la toma de decisiones o sistemas de inteligencia empresarial. Los datos contenidos dentro de un Almacenes de datos se encuentran organizados para permitir el análisis por medio de herramientas OLAP.

La tecnología OLAP permite un uso más eficaz de los almacenes de datos para el análisis en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP y las técnicas de agregados de datos organizan y resumen grandes cantidades de datos

para que puedan ser evaluados con rapidez mediante el análisis en línea y las herramientas gráficas. La respuesta a una consulta realizada sobre datos históricos a menudo suele conducir a consultas posteriores en las que el analista busca respuestas más concretas o explora posibilidades.

Los sistemas OLAP proporcionan la velocidad y la flexibilidad necesarias para dar apoyo al analista en tiempo real. La figura 1-6 muestra la integración del Almacenes de datos y los procesos OLAP, que generalmente se implementan por medio de una aplicación servidora que accede al Almacenes de datos y realiza los procesos de análisis. A través de este servicio OLAP, los usuarios acceden a la información residente en las bases de datos [Microsoft Corp, 2002].

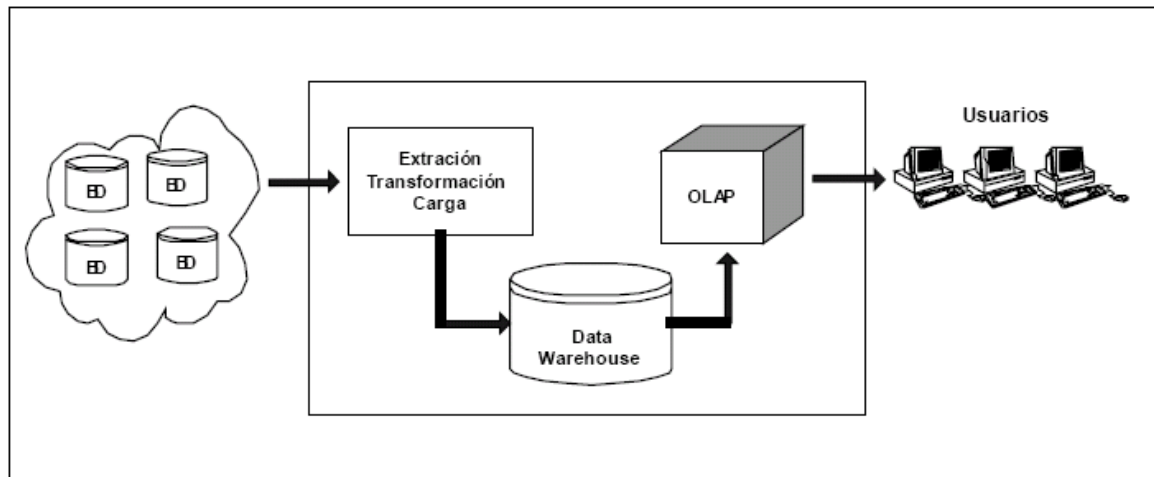


Figura 1-6. Arquitectura básica para OLAP.

A continuación se explica el modelo de datos donde se sustenta esta tecnología.

- **El modelo de datos OLAP.**

Un reto fundamental en la implementación del OLAP es mapear el esquema inicial de la base de datos a un modelo multidimensional. Esto requiere de un significativo esfuerzo de programación con muchos de los productos en el mercado hoy en día. En la evolución de los productos OLAP, el diseño de la base de datos OLAP se ha vuelto un proceso especializado y arcano, intrincadamente enlazado a la tecnología específica del OLAP que se esté implementando.

Consecuentemente, los desarrolladores de bases de datos OLAP son muy especializados, lo cual ha llevado a altos costos de desarrollo de aplicaciones concentrados en la etapa de diseño de datos.

En la mayoría de las implementaciones de OLAP, se asume que los datos han sido preparados para el análisis a través del almacenamiento de datos y que la información se ha extraído de sistemas operacionales, limpiado, validado y resumido antes de incorporarse en una aplicación OLAP.

Este es un paso vital en el proceso, que asegura que los datos que son vistos por el usuario OLAP son correctos, consistentes y que llenan las definiciones organizacionales para los datos.

Cada vez más, la información en un Almacén de Datos (Datawarehouse) se organiza en esquemas de estrella o de copo de nieve. El esquema estrella (figura 1-7) se basa en una tabla de hechos central (las medidas) que se enlaza a las tablas de dimensiones relacionadas (las categorías descriptivas de las medidas), mientras que el esquema copo de nieve (figura 1-8), una tabla de hechos central se enlaza a las tablas de dimensiones relacionadas, pero estas a su vez se enlaza a otras tablas dimensionales.

Con este tipo de esquemas simplifica el entendimiento de los datos por parte del usuario, maximiza el desempeño de las peticiones (queries) de la base de datos para aplicaciones de soporte de decisiones y requiere menor espacio de almacenamiento para bases de datos grandes.

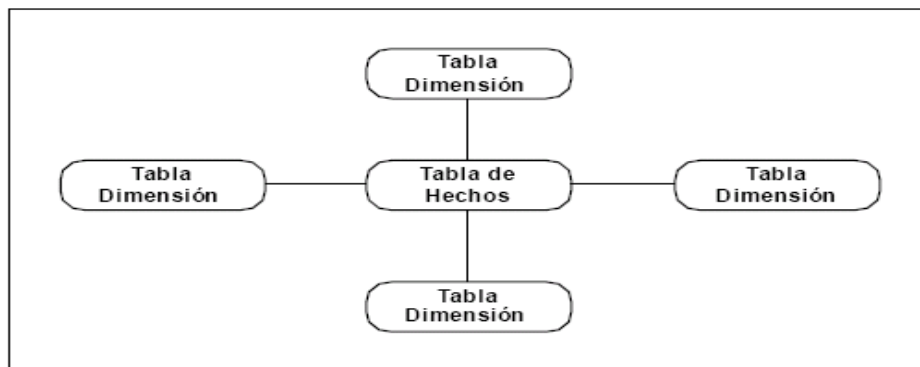


Figura 1-7. Modelo estrella.

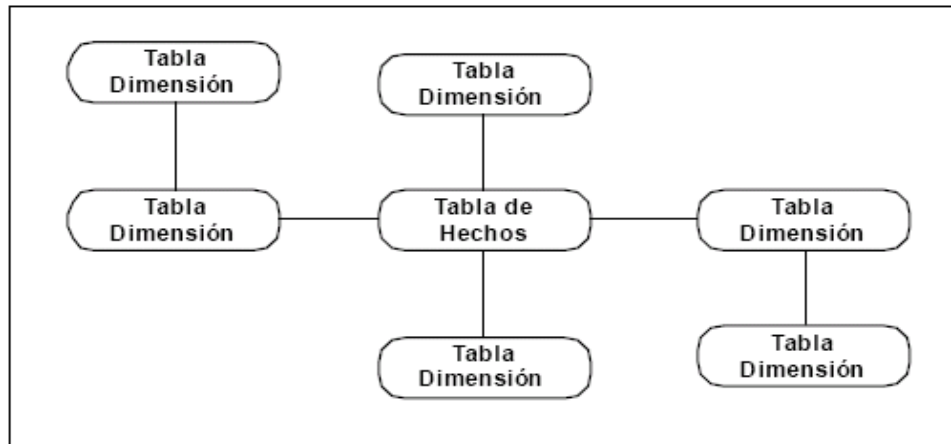


Figura 1-8. Modelo copo de nieve.

Una tabla de hechos contiene generalmente los valores o medidas que se quiere analizar, mientras las tablas de dimensiones contienen las vistas en que se quiere analizar esas medidas.

La figura 1-9 muestra un ejemplo de esquema de estrella. En este tipo de base de datos, una tabla de hechos central se enlaza a las tablas de dimensiones relacionadas.

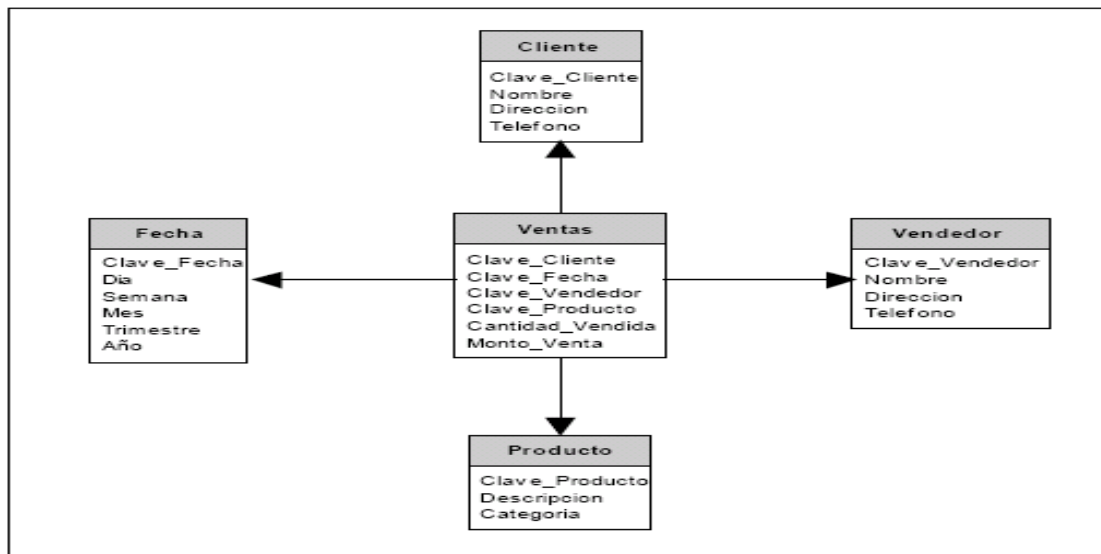


Figura 1-9. Un esquema de estrella.

La figura 1-10 muestra un ejemplo de esquema copo de nieve. Este tipo de esquema se caracteriza por tener tablas dimensionales relacionadas con otras tablas dimensionales además de vincularse a la tabla de hechos.

A continuación se enumeran algunas de las principales ventajas del esquema estrella.

- Crea una base de datos con tiempos de respuesta rápido.

- Diseño fácil de modificar.
- Simula como ven los datos los usuarios finales.
- Simplifica la navegación.
- Facilita la interacción con herramientas.

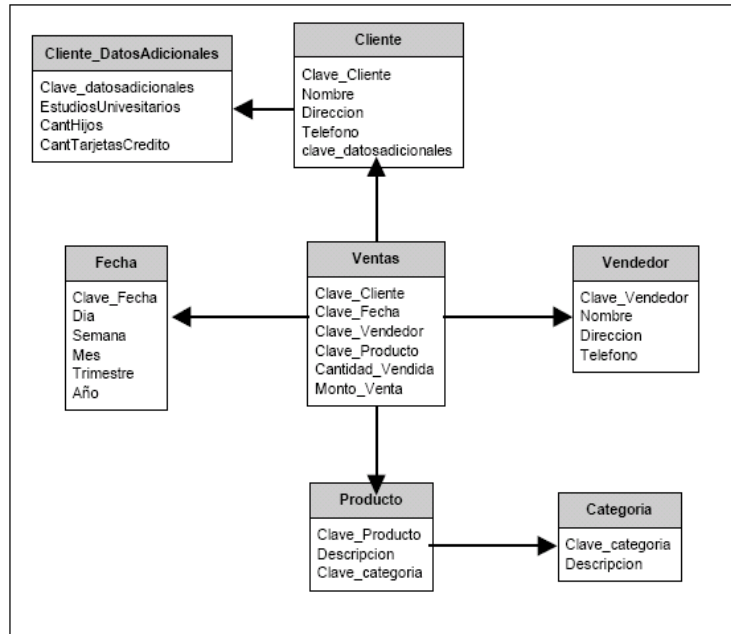


Figura 1-10 esquema copo de nieve

Los esquemas de estrella y copo de nieve son aproximaciones relacionales del modelo de datos OLAP y son un punto de partida excelente para construir definiciones de cubo OLAP. Pocos productos OLAP han tomado ventaja de este hecho. Generalmente no han provisto herramientas sencillas para mapear un esquema de estrella a un modelo OLAP y como resultado mantienen el costo de construir el modelo OLAP extremadamente alto y el tiempo de desarrollo innecesariamente largo.

Debido a la explosión de datos, las aplicaciones OLAP pueden sufrir aún más cuando los datos de detalle o fuente están distribuidos dispersamente en todo lo amplio del cubo multidimensional. Los valores faltantes o inválidos crean dispersión en el modelo de datos OLAP. En el peor caso, un producto OLAP podría almacenar un valor vacío. Por ejemplo, una compañía podría no vender todos los productos en todas las regiones, así que no aparecerían valores en la intersección donde los productos no se venden en una región particular.

La dispersión de datos, un reto para las compañías de OLAP, se ha vencido con varios grados de éxito. Las peores implementaciones resultan en bases de datos que almacenan valores vacíos, por tanto teniendo baja densidad y desperdiciando espacio y recursos. Los servicios OLAP no almacenan valores vacíos y como resultado, aún los cubos dispersamente poblados no se inflarán de tamaño. Mientras que este asunto es frecuentemente subrayado como un factor decisivo de arquitecturas OLAP, por algunos vendedores de OLAP, las diferencias entre las implementaciones de las compañías en el manejo de la dispersión son mínimas comparadas a las más importantes explosiones de datos causadas por precalcular demasiados agregados.

- **Implementación del OLAP**

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente.

Un principio clave del OLAP es que los usuarios deberían de ver tiempos de respuesta consistentes para cada vista de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados, son la base de las ganancias de desempeño del OLAP.

En los primeros días de la tecnología OLAP, la mayoría de las compañías asumía que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Después, otras compañías descubrieron que a través del uso de estructuras de base de datos (esquemas de estrella y de copo de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales (RDBMS) para el OLAP. Estos vendedores llamaron a esta tecnología OLAP relacional (ROLAP). Las primeras compañías adoptaron entonces el término OLAP multidimensional (MOLAP), estos conceptos, MOLAP y ROLAP, se explican con más detalle en los siguientes párrafos:

Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalabilidad. Por otro lado, las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes.

Un desarrollo reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

- **Sistemas MOLAP**

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales

Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico.

La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

La información procedente de los sistemas operacionales, se carga en el sistema MOLAP, mediante una serie de rutinas por lotes. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos por lotes, para calcular los datos agregados, a través de las dimensiones de negocio, rellorando la estructura MDDDB.

Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través del interface, y la lógica de aplicación de la MDDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos recompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan recalculados y almacenados previamente.

Sistemas ROLAP La arquitectura ROLAP, accede a los datos almacenados en un Almacén de Datos (Datawarehouse) para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el Almacén de Datos (Datawarehouse) se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan

mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos pre calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del Almacén de Datos (Datawarehouse), y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

- **ROLAP vs. MOLAP**

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones:

- El ROLAP delega la negociación entre tiempo de respuesta y el proceso por lotes al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se precompilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos por lotes.
- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones por lotes.
- Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones.
- Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes más reducidos (menos de 5 Gb)

Por ello, y resumiendo, el ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. El MOLAP es una solución particular,

adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones más modestos.

1.9. Introducción a la Minería de datos

La minería de datos es la extracción de información oculta y predecible de grandes bases de datos, es una tecnología para ayudar a las compañías a descubrir información relevante en sus bases de información. Las herramientas de Minería de datos predicen futuras tendencias y comportamientos. Los análisis prospectivos automatizados ofrecidos por la automatización de la Minería de datos van más allá de los eventos pasados provistos por las herramientas usuales de sistemas de soporte de decisión. Las herramientas de Minería de datos pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas.

Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar.

Muchas compañías ya colectan y refinan cantidades masivas de datos. Las técnicas de Minería de datos pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on-line). Una vez que las herramientas de Minería de datos fueron implementadas en computadoras cliente servidor de alta performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué?" y presentar los resultados en formas de tablas, con gráficos, reportes, texto o hipertexto [Groth, 1998].

- **Los Fundamentos de Minería de datos**

Las técnicas de Minería de datos son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, continuó con mejoras en el

acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real.

Minería de datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. Minería de datos está listo para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

- Recolección masiva de datos
- Potentes computadoras (algunas con multiprocesadores)
- • Algoritmos de Minería de datos utilizan técnicas que han existido por lo menos desde hace 10 años, pero que sólo han sido implementadas recientemente como herramientas maduras y confiables.

En la evolución desde los datos de negocios a información de negocios, cada nuevo paso se basa en el previo. Por ejemplo, el acceso a datos dinámicos es crítico para las aplicaciones de navegación de datos (OLAP), y la habilidad para almacenar grandes bases de datos es crítica para la Minería de datos.

Los componentes esenciales de la tecnología de Minería de datos han estado bajo desarrollo por décadas, en áreas de investigación como estadísticas, inteligencia artificial y aprendizaje de máquinas. Hoy, la madurez de estas técnicas, junto con los motores de bases de datos relacionales de alto performance, hicieron que estas tecnologías fueran prácticas para los entornos de Almacenes de datos actuales [Cabena et al., 1998].

El Alcance de Minería de datos El nombre de Minería de datos deriva de las similitudes entre buscar información de negocios en grandes bases de datos, encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de Minería de datos puede generar nueva información al proveer las siguientes capacidades:

- Predicción automatizada de tendencias y comportamientos. Minería de datos automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (targeted marketing). Minería de datos usa datos en mailing promocionales anteriores para identificar posibles objetivos para maximizar los resultados de la inversión en futuros mailing. Otros problemas predecibles incluyen pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que respondan a determinadas características.
- Descubrimiento automatizado de modelos previamente desconocidos. Las herramientas de Minería de datos barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otros problemas de descubrimiento de modelos incluye detectar transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de escritura en la carga de datos.

Cuando las herramientas de Minería de datos son implementadas en sistemas de procesamiento paralelo de alta performance, pueden analizar bases de datos masivas en minutos. Procesamiento más rápido significa que los usuarios pueden automáticamente experimentar con más modelos para entender datos complejos.

La alta velocidad de procesamiento junto a las técnicas de Minería de datos hace que sea práctico para los usuarios analizar inmensas cantidades de datos. Grandes bases de datos, a su vez, producen mejores predicciones [Groth, 1998].

Las bases de datos pueden ser grandes tanto en profundidad como en ancho: Más columnas: los analistas muchas veces deben limitar el número de variables a examinar cuando realizan análisis manuales debido a limitaciones de tiempo. Sin embargo, variables que son descartadas porque parecen sin importancia pueden proveer información acerca de modelos desconocidos. Un Minería de datos de alto rendimiento permite a los usuarios explorar toda la base de datos, sin preseleccionar un subconjunto de variables.

Más filas: muestras mayores producen menos errores de estimación y desvíos, y permite a los usuarios hacer inferencias acerca de pequeños pero importantes segmentos de población.

Las técnicas más comúnmente usadas en Minería de datos son:

- Redes neuronales artificiales: modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.
- Árboles de decisión: estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection).
- Algoritmos genéticos: técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
- Método del vecino más cercano: una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases del/de los k registro (s) más similar/es a él en un conjunto de datos históricos (donde $k > 1$). Algunas veces se llama la técnica del vecino k-más cercano.
- Regla de inducción: la extracción de reglas if-then de datos basados en significado estadístico.

1.10. Arquitectura Data Mining

Para aplicar mejor estas técnicas, deben estar totalmente integradas con los Almacenes de datos así como con herramientas flexibles e interactivas para el análisis de negocios (herramientas OLAP). Varias herramientas de Minería de datos actualmente operan fuera del AD, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación

operacional, la integración con el AD simplifica la aplicación de los resultados desde Data Mining.

El punto de inicio ideal es un Almacenes de Datos Este puede ser implementado en una variedad de sistemas de bases relacionales y debe ser optimizado para un acceso a los datos flexible y rápido.

Un server OLAP permite que el usuario analice los datos de acuerdo a como quiera mirar el negocio, resumido por línea de producto, u otras perspectivas claves para su negocio. El server de Minería de datos debe estar integrado con los Almacenes de datos y el server OLAP para insertar el análisis de negocios directamente en esta infraestructura. A medida que el Almacén de Datos (Datawarehouse) crece, la organización puede aplicar extraer la información oculta y aplicarla en futuras decisiones.

La figura 1-11 muestra una arquitectura general para Minería de datos en donde se integra al Almacén de Datos (Datawarehouse) y al servidor OLAP.

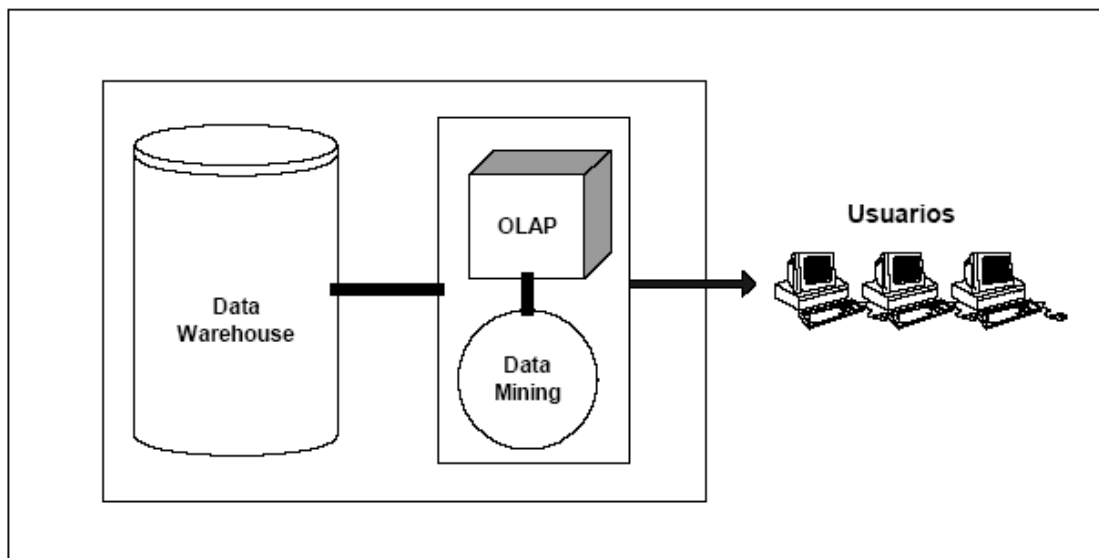


Figura 1-11. Arquitectura básica para Minería de Datos (Data Mining).

- **Implementación de Data Mining**

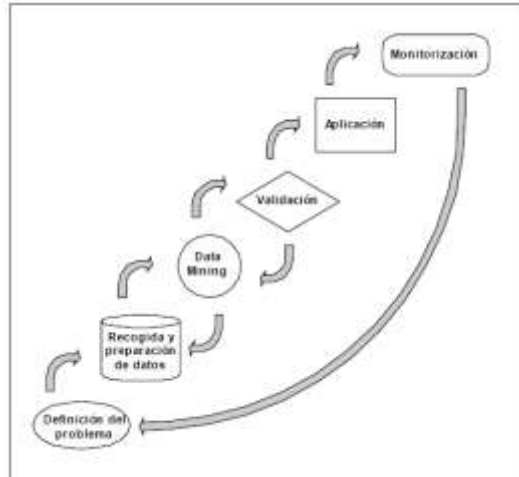
¿Cuán exactamente es capaz un Minería de datos de decir cosas importantes que se desconocen o que van a pasar? La técnica usada para realizar estas predicciones en Minería de datos se llama Modelado.

Modelado en Minería de datos es simplemente el acto de construir un modelo en una situación en donde se conoce la respuesta y luego se aplica en otra situación de la cual se desconoce la respuesta. Por ejemplo, si se busca un galeón español hundido lo primero que se podría hacer es investigar otros tesoros españoles que ya fueron encontrados en el pasado. Notaría que esos barcos frecuentemente fueron encontrados fuera de las costas de Bermuda y que hay ciertas características respecto de las corrientes oceánicas y ciertas rutas que probablemente tomara el capitán del barco en esa época. Al notar esas similitudes se arma un modelo que incluye las características comunes a todos los sitios de estos tesoros hundidos. Con estos modelos en mano se busca el tesoro donde el modelo indica que en el pasado hubo más probabilidad de darse una situación similar. Con un poco de esperanza, si tiene un buen modelo, probablemente se encontrará el tesoro.

Este acto de construcción de un modelo es algo que la gente ha estado haciendo desde hace mucho tiempo, seguramente desde antes del auge de las computadoras y de la tecnología de Data Mining. Lo que ocurre en las computadoras, no es muy diferente de la manera en que la gente construye modelos. Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta es conocida y luego el software de Minería de datos en la computadora debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo. Una vez que el modelo se construyó, puede ser usado en situaciones similares donde no conoce la respuesta.

Si alguien que tiene un modelo que puede predecir el comportamiento de los clientes, ¿cómo puede saber si es realmente un buen modelo? La primera cosa que puede probar es que aplique el modelo a su base de clientes conocidos y usuales donde ya se conoce la respuesta. Con Data Mining, la mejor manera para realizar esto es dejando de lado ciertos datos para aislarlos del proceso de Data Mining. Una vez que el proceso está completo, los resultados pueden ser testeados contra los datos excluidos para confirmar la validez del modelo. Si el modelo funciona, las observaciones deben mantenerse para los datos excluidos. Entonces, los pasos típicos para realizar Minería de datos son los siguientes [Synera, 2002]:

- **Definición del problema:** de la misma manera que en un análisis tradicional, antes de iniciar un proceso de Minería de datos debemos tener muy claro el problema que queremos resolver.
- **Recopilación y preparación de datos:** los datos originales de las BD transaccionales no están preparados para el análisis y, a veces, es necesario aplicar modificaciones, crear agregados y diseñar estructuras nuevas. Además, muchos métodos de Minería de datos necesitan los datos en un formato específico.
- **Data mining:** consiste en construir un modelo sobre los datos con capacidad predictiva y/o descriptiva, de manera que pueda utilizarse para resolver el problema planteado. Para ello, se emplean técnicas estadísticas o de Inteligencia Artificial.
- **Validación:** después de construir el modelo, éste se debe validar antes de utilizarse. La validación puede ser de carácter técnico (utilizar muestras adicionales de datos para comprobar la capacidad predictiva o descriptiva del modelo) o conceptual (ver si la interpretación es satisfactoria, si el resultado es aplicable). Si el modelo no puede validarse, es posible que sea necesario aplicar de nuevo el método de Minería de datos o modificar los datos.
- **Aplicación:** una vez validado, el modelo debe implementarse en el proceso que se desea mejorar. Dependiendo del proceso, esta implementación puede ser más o menos directa y requerir más o menos tiempo.
- **Monitorización:** debe existir un seguimiento de la implementación del modelo en el proceso que se desea mejorar para comprobar sus resultados reales. Si el resultado no es bueno, es posible que haya que redefinir los objetivos. Y puede que, aun siendo óptimo, sugiera nuevos objetivos que se pueden alcanzar.



La figura 1-12 etapas de implementación del Data mining.

1.11. Inteligencia de Negocios

En este apartado se realiza una introducción y definición de Inteligencia de Negocios, se detallan las características principales de los sistemas más utilizados para la toma de decisiones y finalmente se realiza una breve descripción de otro componente tecnológico importante para el soporte de Inteligencia de Negocios, los reportes.

- **Origen de la Inteligencia de Negocios**

Los sistemas de información son bastante recientes si los comparamos con otras áreas de conocimiento. Inicialmente, con la introducción de las computadoras en la década del 60, las aplicaciones se corrían de forma individual en archivos maestros que almacenaban los datos en medios magnéticos, con la limitante del acceso secuencial. El problema del acceso secuencial es la necesidad de recorrer todos los registros antes de encontrar el que se está buscando. Además, al almacenar operaciones individuales, pronto existió una cantidad enorme de cintas y medios de almacenamiento con redundancia, que hacía difícil su administración.

En la década del 70, cuando aparecen los dispositivos de almacenamiento de acceso directo, se da un gran paso en cuanto a la velocidad para acceder a los datos, pues con ellos las búsquedas ya no eran lineales, sino directas. Junto con ello también

aparecen sistemas de administración de bases de datos (DBMS), cuyo propósito era facilitar al programador el acceso a la información al encargarse del almacenamiento e índices.

En ese entonces se definió a la base de datos como "una fuente única de información para todo el procesamiento". Tanto la implementación de las bases de datos para el procesamiento en línea como las nuevas tecnologías y lenguajes de cuarta generación (4GL), proveyeron al usuario la facilidad de tener el control directo de los sistemas y la información, dando origen a los primeros Sistemas de Información formales. Pero también fueron el inicio del paradigma de una sola base de datos que pudiera servir tanto al procesamiento operativo como al procesamiento de alto nivel.

La información almacenada en las bases de datos mantiene el registro total de lo que sucede en la organización. Cuando un usuario operativo desea consultar transacciones unitarias o pequeños grupos de transacciones, se puede acceder directamente y extraer la información en un lapso muy breve (milisegundos); si la base de datos no es muy extensa, incluso un gerente puede también realizar una consulta (vía asistentes) que traiga información resumida sobre muchos registros e, incluso, sobre toda la base de datos, los tiempos para consultas de ese tipo son muy razonables (segundos).

¿Pero qué pasa cuando los datos sobrepasan los límites permisibles para tener la información disponible? Algunos de los factores causales de que las bases de datos se volvieran poco operativas para consultas extensas son las fusiones, la globalización, las alianzas, la diversificación de productos, el crecimiento exponencial de las empresas y, en general, todas las condiciones derivadas de la evolución natural de las empresas que trajo consigo un aumento cuantitativo de los datos que se necesitaba almacenar.

A esto hay que agregar que las herramientas o software necesario para obtener la información eran muy especializados y rara vez una persona que toma decisiones tenía el background necesario para manipular información.

La información primitiva se volvió muy extensa y poco práctica para cierto tipo de consultas, había que desarrollar nueva tecnología que permitiera derivar información calculada para satisfacer las necesidades de la administración, además, la información

primitiva representa el valor actual, es utilizada y operada en procesos repetitivos, por lo tanto, es posible su modificación.

La información derivada no puede ser actualizada porque, normalmente, contiene valores históricos, es operada y utilizada por procesos que se ejecutan aleatoriamente. La información primitiva es operacional apoyando las funciones de empleados de línea, la información derivada es para el soporte de decisiones que normalmente apoya a administradores y ejecutivos.

Teniendo tantas diferencias es complicado entender que la misma información pueda residir en una misma base de datos. De hecho así fue, la gente que toma decisiones demandó respuesta a sus necesidades en los almacenes operativos encontrando muchas limitantes en tiempo y forma, a fin de cuentas es posible satisfacer los requerimientos, pero no como los necesita quien toma decisiones, la tecnología intervino para desarrollar arquitecturas especializadas para que resida la información según su tipo y que sea operada por el software adecuado para desempeñar sus funciones.

Inteligencia de Negocios (BI) es un concepto que se asocia 100% con los niveles directivos, surge de la necesidad de contar con información para dirigir el rumbo de la empresa por los altos mandos, sin embargo, con el tiempo se ha ido ampliando el alcance de este término hasta llegar prácticamente a toda la empresa.

A pesar de relacionar completamente el término con conceptos 100% computacionales, sobre todo las herramientas utilizadas para lograr implementar un desarrollo de este tipo, la verdad es que el concepto no se construye basándose en herramientas computacionales, sino de la formulación de estrategias efectivas de negocios que respondieran a los nuevos tiempos y sus demandas.

El énfasis es en los requerimientos y de ahí se desprenden las aplicaciones, es decir, los hombres de negocio dictan las necesidades y la gente técnica investiga y adapta la tecnología para resolver favorablemente esos requerimientos con todos los medios a su alcance. BI se plantea una sinergia entre los Tomadores de Decisiones y las herramientas que emplean, la tecnología está claramente vinculada al management de las empresas, teniendo como resultado obtener ventajas competitivas, producto de decisiones mejor informadas.

En función de esta se puede entender a BI como una combinación de tecnología y desarrollo de negocios [Bitam, 2002].

- **Definición de Inteligencia de Negocios**

Como muchos otros conceptos o términos, el de Inteligencia de Negocios no escapa a la diversidad de interpretaciones. Se justifica su uso y se entiende el que sea considerado como una Tecnología de Información, pero no existe un acuerdo en cuanto a su definición. "Es el conjunto de tecnologías que permiten a las empresas utilizar la información disponible en cualquier parte de la organización para hacer mejores análisis, descubrir nuevas oportunidades y tomar decisiones más informadas." Analizando las definiciones, queda primeramente claro que BI no es una metodología, software, sistema o herramienta específica, es más bien un conjunto de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software entre otros, con un fin común para el apoyo a la toma de decisiones. A partir de elementos comunes es que podemos obtener una definición que abarca BI en cuando a su utilidad y funcionalidad en las empresas.

a) - Información

Es la esencia de BI. El fin último es proveer de información al usuario final para apoyarlo en la toma de decisiones, y esta información puede provenir tanto de los almacenes operacionales como de arquitecturas diseñadas específicamente para el análisis como DataMart y Almacenes de Datos El usuario puede necesitar información de cualquier fuente primitiva o derivada para apoyarse en su labor, para lo cual BI utiliza o construye fuentes de datos o de información interna o externa, que son la principal materia prima de esta Tecnología.

b) - Apoyo a la toma de Decisiones

Un sistema³ que exclusivamente brinde información no representa lo que se busca con BI, una segunda característica consiste en organizar y presentar los datos relevantes para que puedan verdaderamente apoyar una Toma de Decisiones.

Esto implica tecnologías, técnicas de análisis y todo aquello que sea necesario para obtener de los datos, solo aquella información relevante y útil a la labor del usuario.

Recordando el origen de BI, surge en la toma de decisiones para obtener ventajas competitivas producto de decisiones mejor informadas. Si su origen cae en el desarrollo de negocios, es lógico entender que BI sea un apoyo para tal efecto. BI abarca cualquier forma de organizar información, siempre y cuando sostenga la Toma de Decisiones.

c) - Orientación al Usuario Final

Un factor que incidió en la tecnología BI para explotar información fue que el usuario final no poseía conocimientos técnicos que le permitieran tener un acceso sencillo y directo a los datos operacionales, pues esa área está reservada para informáticos.

Por tal motivo, el usuario final no tenía de primera mano la información que necesitaba y las consultas no definidas, que son las que tradicionalmente realiza un ejecutivo, eran realizadas por terceras personas (léase secretarios, asistentes técnicos o gente de sistemas) con la dependencia consecuente. BI incluye herramientas de explotación de información orientadas a usuario final, para eliminar la dependencia de terceras personas.

Se pretende brindar las facilidades necesarias para que, con la tecnología, el usuario actúe solo. Las herramientas de BI son sencillas, intuitivas y fáciles de entender y usar; pueden tener diversos fines, como son: Informar, reportar, permitir análisis, identificar tendencias, proyectar, etc. Cualquiera que sea su función final, el común es el mismo: orientación a usuario final.

Considerando los elementos comunes en cualquier definición, puede implementarse una definición más acabada. "Inteligencia de Negocios es una combinación de tecnologías de colección de datos y manejo de información, que implementa soluciones orientadas al usuario final para apoyar la toma de decisiones, aprovechando la información estratégica disponible en cualquier parte de la organización." Para la colección de datos usa o construye almacenes de datos y los maneja con técnicas de análisis y herramientas orientadas al usuario final.

Los almacenes de datos son las fuentes operacionales (bases de datos, archivos de texto, hojas de cálculo, administradores de archivos, etc.), bases de datos operacionales, bases de datos externas, Almacenes de datos y Datamarts.

Las técnicas de análisis principales son los Sistemas de Información Ejecutiva (EIS), Sistemas de Soporte de Decisiones (DSS), Minería de datos y Herramientas de Reportes, estas últimas a veces forman parte de las anteriores. BI es un término "agrupador".

El que sea considerado como un conjunto de conceptos le da un poder enorme, pues pueden integrarse funciones que tradicionalmente estaban separadas, tales como el acceso de datos, reportes, explotación, pronóstico y análisis.

De ese modo, al menos en la actualidad en empresas grandes, BI se ha convertido en un apoyo indispensable para la Toma de Decisiones, en cualquier nivel de la organización y mucha gente está explotando el potencial estratégico de los datos operativos. Bien utilizada, BI puede ser un arma estratégica de la gente de negocios, sustentada en tecnología de sistemas [Bitam, 2002]..

- **Sistemas de Información Ejecutiva y Sistemas de Soporte a las Decisiones**

A continuación se realiza una explicación de las características principales de los Sistemas de Información Ejecutiva (Executive Informations System EIS) y de los Sistemas de Soporte a las Decisiones (Decision Support System DSS) que por medio de estos tipos de sistemas se brinda información de toma de decisiones a los usuarios finales.

- **Sistemas de Información Ejecutiva (EIS)**

En la década del 90, desde que las grandes empresas iniciaron su camino para que sus ejecutivos manipularan la información como deseaban para lograr una visión completa del negocio, hubo muchos intentos y malas acogidas motivadas por razones de tecnología, costos o simplemente cultura. El Almacén de Datos (Datawarehouse) fue un esquema de información que mantenía datos para efectos exclusivamente de análisis y estadísticos, con este avance en el almacén, los esfuerzos se centraron en utilizarlo y proveer de la información que necesitaba el ejecutivo.

Los primeros intentos se centraron en la incursión al mercado de soluciones con alarmas, instrumentos de consulta, sistemas expertos y mucho más, pero la evolución, incluso a la fecha, es hacia alcanzar a colocar en una pantalla la mayor cantidad posible de datos para realizar análisis gráficos, visuales y rápidos basados en técnicas básicamente de consolidación, agrupamiento y tendencia. Esto dio origen a los Sistemas de Información Ejecutiva [Bitam, 2002].

- **¿Qué es un EIS?**

EIS se refiere a cualquier sistema de software que muestre información ejecutiva de las diferentes áreas del negocio en un solo sistema, facilitando el monitoreo de la empresa. El EIS es una técnica de más alto nivel dentro de las herramientas de BI. Tiene como objetivo primordial proveer de toda la información necesaria a la gente que toma decisiones, de modo fácil y prácticamente con una mínima interacción con el sistema. En términos formales, "un EIS es un sistema de información que permite a los ejecutivos acceso rápido y efectivo a información compartida, crítica para el negocio, utilizando interfaces gráficas".

Las interfaces que son utilizadas en estos sistemas deben ser más sofisticadas que los sistemas transaccionales y deben incluir, en el menor número de páginas posible, la mayor cantidad de información que el usuario necesita para monitorear su empresa. La función principal de un EIS es el monitoreo de la empresa o de una área de negocio específica. Se debe entender que el monitoreo implica una visualización de lo que se quiere controlar y, por tal motivo, no debe existir mayor esfuerzo por parte del usuario que simplemente ver las consultas que cotidianamente realiza.

La información que un EIS presenta es producto de indicadores empresariales que cotidianamente son evaluados para analizar su comportamiento y lo que permite determinar su cualidad son las propiedades que le son asignadas a cada indicador.

Por ejemplo, la cuota de ventas para cada vendedor en una empresa es un indicador de su productividad dentro de la organización. Un vendedor que haya logrado una cuota de 9,000 USD el mes anterior, tiene una cantidad que, por si sola, no dice mucho, no se sabe si es buena o mala la cantidad que logró vender.

La compañía previamente diseña las "propiedades" de la métrica, que son indicadores para evaluar el comportamiento. Por información histórica se determina que menos de 8,000 USD está considerado como una cuota baja, entre 8,000 y 10,000 USD se encuentra en un rango normal, y mayor de 10,000 USD es una cuota alta que merece un bono adicional equivalente a un porcentaje sobre la diferencia. A cada uno de los calificadores se le asignan ciertas características con la intención de identificar rápidamente el rango en que se encuentran.

Con esta explicación se podría entender que en este caso, las consultas para determinar las ventas son predefinidas. Para este ejemplo, un Director de Ventas puede simplemente abrir una consulta predefinida que le presente aquellos vendedores con una cuota superior a 10,000 USD, quienes se harán acreedores a un bono adicional; de la misma forma puede visualizar fácilmente quienes no hayan logrado su cuota para tomar decisiones correctivas.

En caso de que el Director deseara ver más información relacionada con un vendedor y entrar a detalle a ver sus ventas, cantidades o clientes, estaría entrando a un Sistema de Soporte de Decisiones (DSS), el cual, a pesar de que en conjunto con un EIS, forman una poderosa solución de BI, pero merece una explicación independiente por la diferencia de su uso. Debido a la función principal de monitoreo, un EIS es considerado como un sistema muy fácil de usar y de desarrollar, pero con funcionalidades analíticas muy limitadas.

La diferencia de los EIS con otros sistemas no solo es la vistosidad y facilidad de uso. Aparte del front-end, los EIS interpretan y manipulan de forma diferente la información, pues trabajan con formatos de datos no típicos, tales como Almacenes de datos o Datamart. Prácticamente todos los EIS obtienen sus datos de matrices multidimensionales denominadas "cubos" y las herramientas en las que se desarrollan estos sistemas tienen tecnología que permite realizar consultas amplias y complejas de diversas fuentes de datos en tiempos mínimos.

Las partes importantes de un EIS son: la interfaz de usuario y la base de datos multidimensional, esto montado en una arquitectura Cliente/Servidor.

Normalmente las pantallas o escenarios que se le presentan al usuario final poseen información que se obtiene vía remota del servidor en donde existe, de forma

consolidada, proveniente de múltiples fuentes de datos. El Cliente es precisamente el usuario final, a quien se le debe presentar, de una forma amigable y sencilla, la mayor cantidad de información posible para que le apoye a visualizar sus variables de control del área específica que gestiona.

Cada escenario en un EIS se adecua a los casos particulares de cada usuario, proveyendo la información necesaria para tomar decisiones. En la mayoría de los casos, los EIS tienen en la gráfica su principal medio de representar información al usuario, pues es la forma más sencilla de realizar comparaciones entre variables.

Esto y las tablas son una forma bastante sencilla, pero útil para otorgar información al usuario, pero esos componentes deben complementarse con otras utilerías y otros componentes que permitan no solo ver, sino también analizar lo que está pasando en la empresa con solo mover el ratón.

En términos prácticos, un tomador de decisiones puede entender más el comportamiento de su empresa si ve escenarios en donde no sea necesario interactuar, pero que tenga lo que necesita para controlar su negocio y, solo en caso de ser necesario, interactuar de modo intuitivo (casi siempre con doble clic o arrastre) para obtener más detalle.

- **¿Qué debe contener un EIS**

Un sistema para ser considerado un EIS, debe reunir una serie de características adicionales a las que por defecto debe contener al formar parte de la Tecnología de Inteligencia de Negocios, es decir, brindar información y que sirva de apoyo a la toma de decisiones.

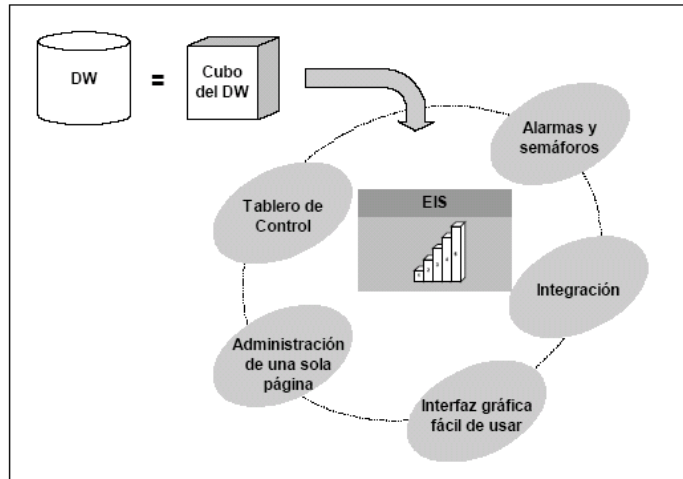


Figura 1-13 muestra las características principales de un EIS

- **Interfaz fácil de usar y ver**

La heterogeneidad en cuanto a la formación académica que un usuario final posee, es enorme. Puestos clave Pueden ser asumidos por Economistas, Administradores, Contadores, Informáticos, Actuarios, Ingenieros Industriales etc.

El único elemento común es un mínimo de conocimientos necesarios para manejar una hoja de cálculo como Excel Sin embargo, lo que no está bien definido es el límite superior de sus conocimientos.

El EIS está diseñado para un usuario final que no forzosamente domina herramientas computacionales complejas, e independientemente de que las domine, no debe invertir una parte importante de su tiempo en conocer y, posteriormente, utilizar herramientas complicadas.

Bajo esta filosofía es que se menciona como una característica de los EIS el que posean interfaces gráficas sencillas, que tengan una curva de aprendizaje corta y, además, deberán ser vistosas e intuitivas para facilitar la labor de monitoreo del tomador de decisiones.

- **Alarmas o semáforos**

Un típico EIS cuenta con funciones que le permiten al usuario notar rápidamente los errores y los valores destacables de la información. Es una de las características principales, ya que las propiedades que le son asignadas a los valores son las que permiten verdaderamente "monitorear" la información, de otro modo el ver datos planos provocaría que el usuario investigue cada uno de los valores y, posteriormente, los compare contra identificadores antes de poder determinar si son "buenos" o "malos".

La tendencia es hacia interactuar cada vez menos con los sistemas y para ellos, las alarmas se vuelven indispensables, ya que disparan indicadores para que el ejecutivo solo ponga atención donde se han sobrepasado ciertos rangos de tolerancia. Estos indicadores pueden ir desde simple colores de letra y fondos, hasta el envío de correos electrónicos o mensajes de radiolocalizador.

- **Administración de una sola página**

De la mano del Tablero de Control, la Administración de una sola página significa colocar el mayor número posible de indicadores destacables de la empresa en el menor número posible de páginas, además, permitir la navegación hacia otras páginas con más información, ya que, es prácticamente imposible colocar toda la información valiosa en un solo lugar

- **Tableros de Control**

El Tablero de Control es una herramienta que en un principio utilizaba indicadores financieros para permitir desarrollar diferentes procesos de negocio.

Un tablero de control es una herramienta en la cual el usuario puede monitorear a la empresa o área con indicadores de cualquier tipo. Diversos componentes le apoyan en esa actividad, pues cada indicador puede asociarse a un tipo de control específico. Con el tablero de control se pretende destacar los principales aspectos a monitorear, ya sea para el control o para la comunicación.

El fin último es tener la información necesaria para apoyar una decisión en la menor cantidad de pantallas. Hay que considerar también que por la misma evolución de los EIS, la función de controlar con un tablero sigue siendo la más común, pero no es la única, en algunos casos se comienza a sustituir la palabra Control por Comunicación.

- **Integración de información proveniente de los cubos**

El usuario tiene muchas veces la necesidad de tener información proveniente de múltiples sistemas o bases de datos. De hecho, para tomar una decisión estratégica es necesario poseer información de toda la empresa, no solo de una parte de ella.

Un EIS debe permitir integrar información de cualquier aplicación y presentársela al usuario final de una forma transparente para él. Esto es la base para no moverse entre aplicaciones, al integrar la información de múltiples fuentes de información es posible la Administración de una Sola Página en un Tablero de Control, el no hacerlo representa una limitante al usuario con la consecuencia de moverse hacia otras pantallas e, incluso, hacia otras aplicaciones con información dispersa.

- **Sistema de Soporte de Decisiones A continuación de define que es un DSS y se detallan las principales características de este tipo de sistemas.**

¿Qué es un DSS?

Los límites y el concepto de un DSS no han sido completamente precisados, a pesar de que la utilidad ha sido justificada en las organizaciones. Su uso indiscriminado con frecuencia lo lleva a rebasar límites de su aplicación y confundirse con términos como OLAP, Almacén de Datos (Datawarehouse) o EIS, lo cierto es que, independientemente, del término que llegue a utilizar, siempre se asocia al soporte a la toma de decisiones y, de alguna forma, todos los conceptos señalados tienen en la toma de decisiones el punto de encuentro.

"DSS se refiere a cualquier sistema de software que permite análisis de las diferentes variables del negocio para apoyar una decisión."

Puede considerarse como un sistema que se basa en un AD y crea una base de datos multidimensional, permitiéndole al usuario procesar analíticamente la información en

línea (OLAP). Como utilidad al usuario final, un DSS se valora cuando se profundiza en la información para conocer los "porqués de" los indicadores presentados, pero la infraestructura y metodologías que soporten el análisis de información son las que completan el esquema de un DSS y le permiten hacer uso de OLAP, Almacenes de datos y otros conceptos asociados.

En términos prácticos, el DSS lo vemos cuando analizamos la información, pero realmente involucra todo un proceso previo antes de poner la información en el escritorio del usuario.

Los DSS están asociados a las jerarquías dentro de los conceptos de los cubos del Almacén de Datos, por lo siguiente:

En el momento que desea conocerse el "por qué" de un dato visto en un EIS, el DSS debe permitir "profundizar" la información, o sea, conocer el detalle de la misma y de las partes que la componen, aquí se utiliza las jerarquías que fueron definidas en los cubos de información, para ver por cuál de las rutas se profundiza hacia el detalle. Un DSS se inicia cuando se analiza el detalle de una propiedad de la métrica (o indicador). Ejemplo de los anteriores cuando por medio de un EIS se observa que las ventas promedios de la compañía fueron bajas, entonces para determinar en qué zonas (la compañía vende por zonas) las ventas estuvieron por debajo y en cuáles estuvieron por encima del promedio es necesario recurrir al DSS.

Una vez conocida esta información (más toda la relacionada a las ventas de esas zonas) se podrá tomar una decisión para mejorar dichas ventas.

Los DSS se presentan como la opción viable para poder dar soporte y consistencia a la información que se diversificó enormemente. Con una metodología más estricta y más herramientas a su alcance, los DSS comenzaron a ser la contraparte de los sistemas OLTP. Más que una definición, los DSS pueden entenderse a partir de sus características de funcionalidad en una empresa al ser comparadas con los sistemas OLTP que apoyan la transacción diaria.

En un DSS la consistencia se mide globalmente. A pesar de que no importan las transacciones individuales, son precisamente estas las que, en suma, conforman una gran transacción completa y consistente. De hecho, antes de poder formar una gran

transacción, a partir de los sistemas operacionales, debe verificarse la calidad de los datos para garantizar la información correcta y coherente.

Un sistema DSS procesa una transacción por día (puede ser por cualquier periodo de tiempo, pero el común es por día), pero esa transacción es producto de miles o millones de registros que han sido procesados en el día. En vez de llamarle transacción, se le llama carga de información de producción. En este caso lo importante es el estado de consistencia del sistema antes de la carga y el estado de consistencia una vez que se ha terminado de efectuar la carga.

En cuanto a los usuarios y administradores de un DSS, no son los que ingresan cada una de las operaciones en sus PCs, sino los que solicitan una o dos hojas con resúmenes totalizados de miles de operaciones. El usuario de OLTP cambia con frecuencia el tipo de información que solicita, de ahí que sus requerimientos no sean planeados, sino heurísticos. Una pantalla de un sistema DSS deberá contener la información sobre el porqué de una tabla o gráfica, con algunos números y pantallas, al momento de pedir información sobre el porqué de determinado dato, podrá accederse a otra pantalla con la explicación y así, sucesivamente, hasta llegar al último detalle.

Una característica importante de los DSS es que la demanda puede ser muy variable, por lo que es normal implementarlo en una máquina distinta de OLTP. Al estar en una máquina distinta se convierte en un servidor que puede ser accedido por diversos usuarios, pero también puede tener información integrada de múltiples sistemas remotos de OLTP.

El tiempo es un factor importante de los DSS visto como dimensión. La inconsistencia temporal se debe evitar en un DSS. Mientras que un OLTP es complicado para explicar la historia, un DSS lo que hace es tomar fotografías instantáneas de la empresa en un momento determinado de la historia, al sumar esa serie de fotografías se conforman capas que pueden explicar cómo era la empresa en determinado período de la historia. Moviendo fotografías estáticas hacia el DSS cada cierto tiempo programado, se resuelven dos problemas de representación de tiempo que se tiene en los sistemas OLTP. Primero, el DSS permanece sin cambio en sus datos durante el día cuando los usuarios están lanzando consultas. Después, almacenando cuidadosamente la

información de cada instantánea del DSS, se puede representar cualquier momento en el tiempo correctamente.

Las fotografías instantáneas del sistema OLTP se llaman 'extracción de la información de producción', mientras que el envío hacia el DSS se llama 'carga de la información de producción', estos conceptos se denominan ETL, Extracción, transformación y Carga (Extraction, Transformation and Load).

El ciclo de vida de un DSS es una situación muy particular. Mientras que un tradicional ciclo de vida de un sistema se deben entender los requerimientos, analizar las etapas del diseño y después desarrollar; en un DSS el ciclo de vida es inverso, es decir, inicia con la información de los sistemas OLTP, una vez que se tiene, es integrada y se prueba para saber cuál es real y útil, y cual no. A continuación se escriben los programas para manejar esta información y, por último, los programas se analizan, y finalmente se entienden los requerimientos del sistema. El ciclo de vida de un DSS se orienta a la información, mientras que el de OLTP se orienta a los requerimientos.

La interpretación y manipulación de la información es muy distinta a los formatos típicos. A pesar de que la capacidad de extracción de datos puede ser de archivos de texto, tablas, etc., la forma de organizar esa información es lo que difiere, pues la mayoría de los DSS organizan la información vía matrices multidimensionales denominadas "cubos". Los cubos organizan la información de tal modo que puedan, posteriormente, 'montarse' herramientas para desarrollar sistemas complejos, que permitan realizar una gran cantidad de cálculos, consolidaciones, consultas y despliegues de información, localizadas en múltiples repositorios en un tiempo mínimo.

- **¿Qué debe contener un DSS?**

La figura 1-14 muestra las características deseables que debe tener un DSS:

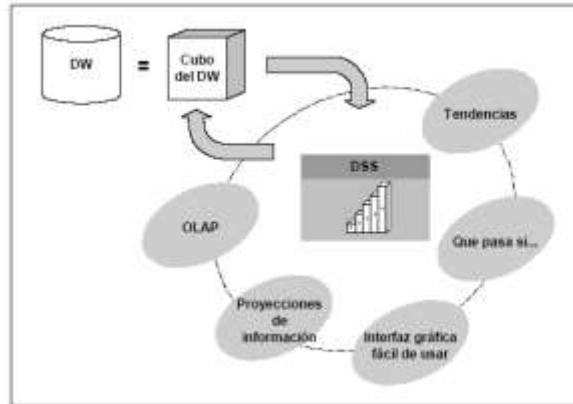


Figura 1-14. Características de un DSS

- **Análisis Multidimensional (OLAP)**

El análisis multidimensional no es privativo de arquitecturas multidimensionales, puede también llevarse a cabo en arquitectura relacional, diseñada para tal caso, lo importante para poder hacer Análisis Multidimensional no son las bases de datos, sino la estructura de la base de datos y las técnicas que se utilicen para su explotación. Las aplicaciones OLAP soportan ese tipo de análisis ya que dos de sus principales características son que permite el análisis y la multidimensionalidad.

Dentro de los niveles estratégicos organizacionales, la información se concibe como una serie de hechos multidimensionales, jerárquicos y relacionados; como ejemplo, los datos de inventarios, ventas y compras están interrelacionados y dependen entre sí. La idea del análisis multidimensional es facilitar la consulta y análisis al usuario al presentar una visión muy sencilla de los datos, muy similar a la forma como él ve la organización. La información puede ser accedida desde diferentes variables organizacionales y mostrando diferentes la perspectivas.

Los almacenes multidimensionales guardan de modo lógico sus datos en arreglos utilizando el concepto conocido como "cubo". En éste, cada una de las caras almacena una dimensión, pudiendo cruzar diferente información en una sola arista de hasta n variables. Físicamente, en la base de datos, una celda de información puede almacenar hasta n dimensiones y, con mucha facilidad, mostrar diferentes perspectivas del cubo, es decir, cambiar la consulta a otras celdas para analizar más a detalle.

Cuando un esquema similar a la realidad se guarda en medios de almacenamiento y aparte se complementa con herramientas diseñadas para obtener específicamente ese tipo de información, el resultado es poder entender a la empresa a través de la tecnología de la misma forma que lo haría en la actualidad. El Análisis Multidimensional brinda esa posibilidad al usuario y es la principal característica que debe poseer un DSS.

- **Proyecciones de Información**

Las proyecciones de negocio ofrecen al usuario un pronóstico de lo que puede ocurrir en el futuro, basándose en análisis estadístico y de regresión.

- **Tendencias**

Utiliza la información presente y pasada para evaluar el comportamiento de determinada variable en el tiempo.

- **Qué pasa si...**

Análisis prospectivo de un indicador al modificarse una o más variables que inciden en su comportamiento.

1.12. Herramientas de Reportes

A continuación se explica el estado actual de los reportes, su necesidad y las características principales que deben reunir las herramientas que los construyen y administran.

Estado Actual de los Reportes

Tradicionalmente, los reportes han sido el medio principal para disponer de información. Tanto reportes en papel como en el escritorio, el usuario depende de ellos para comunicar a la gente lo que está ocurriendo en la empresa vía los almacenes de datos.

Los reportes tienen en las bases de datos su principal fuente de alimentación y han brindado al usuario final entendido éste como cualquier persona que requiera un

reporte, la posibilidad de consultar y publicar lo que las bases de datos poseen, la limitante que siempre ha existido en este sentido es que el generar un reporte implica manejar algunas habilidades técnicas relacionadas con las bases de datos y las herramientas de software.

Los reportes o las consultas son requeridos por cualquier persona dentro de la organización, de hecho, no es una labor que esté limitada a un nivel jerárquico o puesto. Cuando un usuario con poca preparación técnica (secretarias, ejecutivos, etc.) requiere de un reporte o consulta, lo hace vía terceros, el tradicional entorno es hacer la solicitud de información a sistemas y obtenerla después de cierto tiempo, el tiempo puede ser irrelevante o puede significar la pérdida de vigencia de la información solicitada.

El usuario técnico realiza los reportes y para ello necesita conectarse a la base de datos, posteriormente, diseñar el formato requerido y, al final, obtener los datos. Tanto la conexión a la base de datos como la construcción del formato del reporte son tareas de sistemas, la obtención de datos se comparte con el usuario final, con esta lógica se entiende que el usuario final no puede manipular sus reportes en caso de necesitar modificaciones o adecuaciones al mismo, por lo tanto continúa la dependencia, incluso, para modificaciones mínimas.

Los usuarios con algunas habilidades computacionales pueden hacer uso de las herramientas de reportes orientadas a los usuarios de sistemas, pero solo para obtener reportes sencillos, no el tradicional reporte complejo que es el más útil, además, la gente que toma decisiones no debe perder su tiempo en procesar la información para después analizarla.

- **Algunas de las limitantes propias de la generación de reportes en la actualidad son:**

Integración

Los reportes complejos que necesitan de más de una fuente de información se realizan extrayendo los datos en un primer momento y, posteriormente, procesándolos en una herramienta que los integre (hojas de cálculo, archivos de texto, Access, etc.). Esto

implica preparar la información antes de que se encuentre lista para ser enviada al usuario y la preparación normalmente llevan horas o hasta días con la consecuente pérdida de tiempo en funciones secundarias improductivas.

Distribución

Los reportes se imprimen y posteriormente se distribuyen de forma manual o mediante fax. También se generan en pantalla y, luego, se envían mediante fax o correo electrónico, con lo que hay dos actividades involucradas que generan pérdida de tiempo en su ejecución: la publicación y la distribución.

Dependencias de sistemas

Las bases de datos se construyen en general para ser eficaces al almacenar datos no para consultarlos. Los lenguajes de consulta se adaptan a la estructura de las bases de datos. En función del conocimiento necesario, tanto de las bases de datos a consultar como de los lenguajes técnicos necesarios, el usuario final no está en posibilidad real de obtener su propia información, por tal motivo la dependencia de gente de sistemas es muy grande y provoca el que se tengan los reportes cuando la gente de sistemas dispone de tiempo y recursos para hacerlo, visto de otro modo, si la gente de sistemas no lo hace, el usuario final no los tendría a la mano.

Seguridad

Al dejar en manos de terceros la generación de reportes se puede incurrir en que la información valiosa para la empresa pueda ser filtrada y utilizada por personas que no son los directamente interesados en utilizarla. Un reporte de sueldos y prestaciones de los empleados, solicitado por Recursos Humanos, puede ser un detonante de conflictos en manos de inconformes.

Herramientas de consulta y reportes

Las herramientas de consulta y reportes (Query & Reporting Tools) son una categoría de herramientas de BI. Con las herramientas de reportes orientadas al usuario final se pretende mejorar la obtención de información mejorando el área de sistemas al disminuir una capa intermedia entre complejidad técnica y usuario final.

Las herramientas de reportes orientadas al usuario final son software que aísla la compleja capa técnica propia del lenguaje de sistemas, tal como lenguaje SQL, uniones

de tablas y nombres crípticos, al organizar los datos de la terminología de negocios. El resultado es que el usuario final o intermedio tiene una vista mucho más parecida a su concepción del negocio, o al menos lo suficiente como para poder generar sus propios reportes y publicación de los mismos, sin depender de los usuarios técnicos.

El usuario de sistemas continúa participando, pero básicamente en el control sobre el acceso a bases de datos, administración, mantenimiento, seguridad, impacto en la red e incluso, creación única de reportes complejos; el usuario final tiene solo un ambiente amigable en el que es posible crear sus propias consultas y reportes.

Una herramienta de reportes puede también publicar los datos que se encuentran almacenados en un AD. Desde ese punto de vista, se brinda con la tecnología, la posibilidad de que el usuario final no solo consulte y publique la información a detalle, sino información concentrada y agrupada; de ahí que también sea considerada como el soporte último para el máximo nivel de detalle de un DSS.

Cierta duda ha surgido en cuanto a la verdadera posibilidad para que el usuario final realice sus reportes y posteriormente los explote. Definitivamente depende del nivel en el cual ubiquemos al usuario final y, evidentemente, del nivel de conocimientos en materia de computación que pueda tener.

Un usuario final puede ser cualquier persona que requiera hacer una consulta o reporte y éste puede ir desde una secretaria, un asistente, un gerente o un director, cualquiera que necesite información procesada de fuentes de datos es un usuario potencial.

Sin embargo, a pesar de que el manejo de una herramienta como Excel se ha considerado un estándar que pueda utilizar cualquier administrativo de cualquier nivel, no todas las personas pueden decir que verdaderamente saben utilizar, para fines prácticos, esa herramienta.

Precisamente, para ese tipo de usuarios, las herramientas de reportes no serán útiles más que para explotar los reportes que previamente se hayan realizado, es decir, ejecutar reportes.

Otro tipo de usuario, con algunos conocimientos, si podrá construir sus propios reportes, que siguen siendo tarea de gente de sistemas, pues requieren de un procesamiento de la información antes de que sea "pintada" y, para ello, son

necesarios procedimientos externos a los datos con lenguajes de consulta altamente técnicos.

Dependiendo del nivel del usuario puede disminuirse o eliminarse la dependencia de sistemas y pueden las herramientas de reportes mantener 3 niveles de dificultad:

- Usuarios poco expertos (gerentes tácticos, directores ejecutivos) quienes solicitan la ejecución de reportes o consultas predefinidas, según parámetros predeterminados.
- Usuarios con cierta experiencia (asistentes, gerentes operativos, secretarías especializadas) pueden generar consultas o reportes flexibles, apoyándose en una interfaz gráfica intuitiva.
- Usuarios muy experimentados (sistemas) pueden crear e incluso escribir, total o parcialmente, la consulta en un lenguaje de consulta.

Características de una herramienta de reportes

Una herramienta de reportes orientada al usuario final debe también poseer algunas utilidades adicionales que faciliten la generación y publicación de reportes.

Intuitivo

Como cualquier herramienta de BI, la característica común es su facilidad de uso e intuición. Con apoyo en interfaces gráficas y visuales, un usuario con una formación estándar podrá hacer uso de una herramienta de este tipo.

Seguridad.

Deben brindar seguridad para el acceso a los reportes, tanto a nivel usuario como por grupos e, incluso, en el grado de profundidad de cada usuario a la información. Esto con la idea de que la información privada no sea accesible por cualquier persona

Publicación y distribución

Una función importante de las herramientas es eliminar la doble actividad de ejecutar y publicar primero un reporte y, posteriormente, distribuirlo a quien lo necesita. La publicación de una consulta normalmente se realiza mediante fax, correo electrónico, archivo e, incluso, de mano en mano.

El objetivo con las nuevas herramientas es facilitar estas actividades. En el momento que un usuario pueda ejecutar, desde cualquier lugar, una consulta y mandarla imprimir on-site, se está eliminando la doble función de publicación y distribución. Las herramientas de reportes deben permitir realizar consultas desde su PC en modo Cliente / Servidor, Intranet o Internet e imprimir cualquier consulta.

Navegación

La interrelación de reportes es también frecuente para generar sistemas basados en reportes y reportes auxiliares. La navegación entre ellos mediante ligas e hipervínculos es una funcionalidad más.

Programación automática

Generación de instrucciones para que los reportes se ejecuten automáticamente e incluso se distribuyan mediante correo electrónico.

Reportes dinámicos

Permitir el ingreso de parámetros de valor que hagan un reporte flexible y dinámico en el momento de su ejecución. Con esto se economizan esfuerzos al ejecutar muchas consultas a partir de un solo reporte.

Reportes por excepción

Para no consultar una cantidad muy grande de información debe existir la posibilidad de brindar reportes por excepción, es decir, lo que no cumplan con las condiciones de generalidad.

CAPITULO 2. PLAN DEL PROYECTO

El Capítulo se presenta el proyecto, se detallan las motivaciones y necesidades de contar con un sistema de Toma de Decisiones. Se expone una breve introducción de los procesos actuales de la universidad en materia de información gerencial y se realiza un esbozo de la solución.

- **Visión General del Proyecto**

En este apartado se realiza una introducción al proyecto presentando las posibles necesidades de información que puede tener una universidad para ayudar la toma de decisiones y cuáles serían los usuarios de esta información.

Finalmente se propone el desarrollo de un sistema junto con los pasos principales de su construcción de manera que ayude y facilite los procesos de toma de decisiones de la universidad.

- **Introducción al Proyecto**

La mayoría de las organizaciones cuentan con datos de los sistemas de ingreso de transacciones, vitales para registrar las operaciones que sostienen a una empresa. A pesar de la riqueza de estos datos, no se puede recurrir a ellos con facilidad cuando necesitamos encontrar respuestas sobre el funcionamiento de la organización, como por ejemplo, en el caso de una universidad:

- ¿Cuál es la proyección histórica de matriculación de alumnos en los diferentes cursos en un determinado semestre? así, puede utilizarse esta información para ampliar o reducir la infraestructura o recursos necesarios para brindar servicios acordes a la cantidad de alumnos.
- ¿Cuáles son las causas (académicas y/o financieras) de cancelaciones de matrículas de alumnos de la universidad? así, se podrían tomar acciones proactivas tanto académicas como de otra índole.
- ¿Cuáles son las empresas o actividad laboral que han incorporado pasantes según la especialidad de estos durante un determinado período? Datos que se pueden cruzar con el rendimiento académico de alumnos de dichas especialidades, obteniendo un reporte de potenciales nuevos

pasantes, a los cuáles se los puede capacitar con determinados cursos relacionados con la actividad específica.

- ¿Cuánto se sobrepasó el último trimestre el uso de las computadoras personales por alumno en los laboratorios? Es posible implementar una política de uso diferente para el siguiente semestre para evitar que los alumnos se queden sin utilizar las computadoras. Con la información correcta en tiempo y forma se evitarán inconvenientes en los cronogramas y esfuerzos innecesarios.

Si bien con los sistemas tradicionales se pueden preparar reportes ad-hoc para encontrar las respuestas a algunas de estas preguntas, se necesita mucho tiempo y recursos del departamento de sistemas para poder responderlas

Con sistemas del tipo "Inteligencia de Negocios" (*Business Intelligence*), se puede formular y responder las preguntas claves sobre el funcionamiento de la universidad accediendo directamente a los indicadores de calidad, señalar cuáles son los factores que realmente inciden en el buen o mal funcionamiento, detectar situaciones fuera de lo normal, encontrar los factores que maximicen el beneficio y crear escenarios de comportamiento futuro con un alto porcentaje de certeza.

Para esto es necesario implementar un Sistema de Apoyo Gerencial. El sistema debe estar orientado a brindar información interrelacionada para quienes tienen responsabilidades en el ámbito estratégico y táctico de la organización.

Los usuarios de este sistema no son usuarios comunes, sino usuarios que toman decisiones y planifican día a día, a mediano plazo o a largo plazo, en donde la calidad de la información juega un papel categórico. Usuarios que necesitan disponer de información tanto consolidada como detallada de cómo marchan las actividades ya cumplidas, predecir tendencias y comportamientos para tomar decisiones proactivas.

Los usuarios de este sistema pertenecen a todos los departamentos y áreas de la universidad, ya que en cada área es necesario algún tipo de información como la mencionada. Inicialmente, los usuarios potenciales de este sistema pueden pertenecer:

- Dirección

- Consejo Académico
- Secretaría General
- Secretaria Administrativa
- Secretaría de Desarrollo Institucional
- Secretaría Planeación
- Secretaría Postgrado
- Secretaría de Estudios Profesionales
- Coordinaciones

- **Descripción del Problema: El Proceso Actual**

Con los sistemas tradicionales se preparan reportes ad-hoc para encontrar las respuestas a algunas las preguntas, pero se necesita dedicar aproximadamente un 60 % del tiempo asignado al análisis de localización y presentación de los datos, como también asignación de recursos humanos y de procesamiento del departamento de sistemas para poder responderlas, sin tener en cuenta la degradación de los sistemas transaccionales.

Esta problemática se debe a que dichos sistemas transaccionales no fueron contruidos con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones.

En todos los casos se observa la necesidad de considerar como punto de partida la información existente en las bases de datos de la institución.

Según la información recibida la universidad cuenta básicamente con dos sistemas, esta información es la plataforma para diseñar y construir un Sistema de Apoyo Gerencial. Los sistemas que sustentan la principal actividad de la universidad son:

El *Sistema Administrativo Contable*, que está compuesto por:

- Contabilidad
- Sueldos
- Tesorería
- Facturación
- Cuentas a Pagar.

El *Sistema Académico*, que está compuesto por:

- Alumnos
- Profesores, Cátedras y Comisiones
- Planes de Estudio y materias
- Exámenes
- Facultades
- Relaciones Públicas y Prensa
- Encuestas
- Aulas/Laboratorios
- Graduados

Aunque los sistemas transaccionales han sido y son beneficiosos para los usuarios, nunca son diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos.

Gráficamente, los sistemas transaccionales tienen la arquitectura básica que se muestran en la figura 2-1.

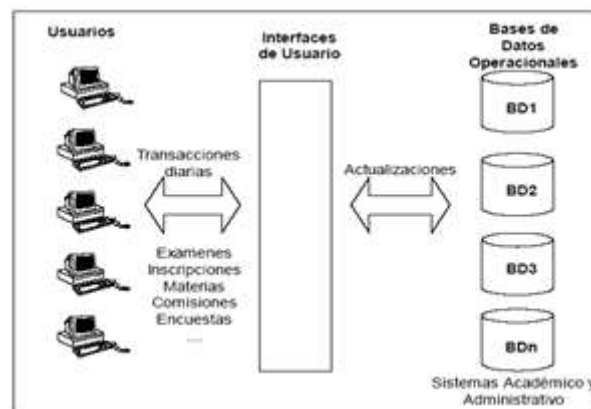


Figura 2.1 Los sistemas transaccionales de la universidad

Como se observa en la figura anterior, la característica sobresaliente es la interacción de los usuarios con las bases de datos, actualizándolas diariamente por medio de las transacciones de cada sistema en particular.

La manera usual de entregar la información es a través de emisión de reportes desde los sistemas operacionales.

Los problemas con la entrega de la información actualmente son entre otros:

- Posibles inconsistencias no detectables fácilmente
- Inflexibilidad en la forma de presentación de la información entregada

- No existe una distribución de datos adecuada a través de los distintos sectores de la universidad
- Tecnologías de datos que no soportan adecuadamente aplicaciones gerenciales y de toma de decisiones
- No se puede integrar fácilmente datos de diferentes aplicaciones.

Las autoridades necesitan mejoras en el proceso actual, particularmente en los siguientes aspectos:

- Disponibilidad de la Información: consolidada, oportuna y accesible. Información que las autoridades necesitan, en el tiempo que la necesita y en el formato que la necesitan
- Proceso de Toma de Decisiones: con un mayor soporte de información se logra que los usuarios dispongan de los datos consolidados más rápidamente para la toma de decisiones.
- Sobre los Procesos: cuando se da acceso a una mejor calidad de información, la universidad puede lograr :
 - Mejorar los retardos de los procesos que resultan de información incorrecta, inconsistente y/o no existente
 - Integrar y optimizar procesos a través del uso compartido y consolidado de las fuentes de información
 - Analizar la producción y el procesamiento de datos que no son usados, producto de aplicaciones mal diseñadas o ya no utilizadas actualmente.

Por otro lado, muchas pueden ser las razones que justifiquen el "porqué" de la situación actual, pero se puede mencionar en forma general la *normal evolución de la arquitectura de los sistemas* de cualquier organización como una de las razones de peso más importantes para sustentar la mejora continua de los propios procesos internos.

2.1. Propuesta de Desarrollo

El Sistema Universitario FES-CUAUTITLAN debe generar una variedad de productos de información, con el fin de satisfacer las necesidades cambiantes de las personas que toman decisiones a través de toda la universidad. Las personas encargadas de tomar decisiones a nivel estratégico requieren informes, pronósticos e información más resumida, ad hoc, y no programada para respaldar sus responsabilidades de elaboración de políticas y de planeación no estructuradas.

Por otra parte, es posible que las personas que toman decisiones a nivel táctico requieran informes especificados donde se haga énfasis en comparaciones detalladas de datos históricos y actuales que sustenten sus responsabilidades más estructuradas en las operaciones diarias.

Conceptualmente se requieren varios tipos de sistemas de información, los cuáles se deben integrar o fusionar en un Sistema de Apoyo Gerencial:

- Sistemas de Información Gerencial (MIS)
- Sistemas de Apoyo a las Decisiones (DSS)
- Sistemas de Información Ejecutiva (EIS).

La tabla 2-1 muestra un resumen de las principales características de cada tipo de sistema.

Características	Tipo de sistemas		
	MIS	DSS	EIS
Apoyo a las decisiones que suministra	Suministran información sobre el desempeño táctico de la organización	Suministra información y técnicas de apoyo a las decisiones para analizar problemas u oportunidades específicas	Suministran información sobre el desempeño estratégico (factores claves) de la organización
Forma y frecuencia de la Información	Respuestas e informes periódicos, de excepción, a solicitud y de entrega	Consultas y respuestas interactivas	Consultas y respuestas interactivas.
Formato de la información	Formato fijo especificado con anterioridad	Formato ad hoc, flexible y adaptable	Formato ad hoc, flexible y adaptable. Formato muy resumido y preferentemente gráfico
Metodología de procesamiento de la información	Información generada mediante extracción y manipulación de datos empresariales	Información generada mediante modelación analítica de los datos empresariales	Información generada mediante extracción y manipulación de datos empresariales como así también análisis de tendencias.

Tabla 2-1. Características de los tipos de sistemas.

Para lograr la integración de estos tipos de sistemas se deberá contar con un repositorio de datos preparado para tal fin. Este repositorio se creará bajo las características de un Almacén de datos.

Esta última definición refleja claramente el principal beneficio que el Almacén de datos aporta a la integración, eliminar aquellos datos que obstaculizan la labor de análisis de información y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Gráficamente, la figura 2-2 muestra la arquitectura básica de la solución desarrollada.

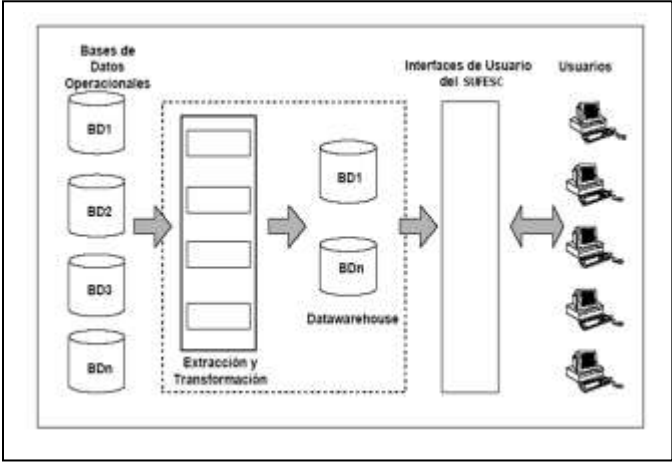


Figura 2-2 Solución propuesta

Para aplicar mejor estas técnicas, éstas deben estar totalmente integradas con el Almacén de datos así como con herramientas flexibles e interactivas para el análisis de negocios.

En la figura anterior, el Almacén de datos se alimenta de los datos residentes en las bases de datos de los sistemas transaccionales de la universidad por medio de diversos procesos (extracción, transformación y carga), este nuevo repositorio (el

Almacén de datos) está diseñado para brindar información de gestión y toma de decisiones.

Una vez definido el repositorio, a este se integran herramientas como un servidor OLAP (Procesamiento Analítico en Línea) y de Minería de datos.

Una vez definido el Almacén de datos físico y los diferentes procesos de análisis de datos, se implementan las aplicaciones de acceso a los datos (Interfaces de Usuarios), estas aplicaciones están determinadas por las características nombradas en los sistemas MIS, DSS, y EIS.

La información que brindan estos sistemas se lleva adelante de diferentes maneras, es decir, cierta información se obtiene por medio de tecnologías OLAP (que permite a los usuarios analizar el negocio desde diferentes perspectivas, entre otras funciones), otras mediante tecnologías de Minería de datos (que permite encontrar información oculta y patrones en los datos, entre otras funciones) y finalmente por medio de reportes estáticos.

- **Pasos Principales de Construcción**

El plan de trabajo es ha sido concebido para abarcar el desarrollo completo de Sistemas de Información sea cual sea su complejidad y magnitud, por lo cual su estructura responde a desarrollos máximos y es adaptable en cada momento de acuerdo a las características particulares del proyecto en cuestión.

Si bien la metodología no define explícitamente su uso para sistemas con las características mencionadas en este documento, ésta se ha adaptado en el presente plan de trabajo para incluir algunas particularidades de proyectos de explotación de Almacén de datos [Sperley, 1999] como por ejemplo:

Determinar el área del negocio a ser soportada en la primera fase del desarrollo.

- Determinar que sistemas transaccionales están asociados al área de negocio seleccionada.

Definir el origen de los datos operacionales. El modelo de datos fuente.

- Documentar las bases de datos de los sistemas transaccionales.
- Determinar las superposiciones y relaciones entre los sistemas transaccionales.
- Especificar los requerimientos para el soporte a las decisiones (MIS, DSS y EIS).
- Estudiar y analizar diferentes herramientas para OLAP y Minería de datos que se adecuen a la Universidad. Herramientas de Acceso al componente de almacenamiento físico Almacén de datos.
- Definir las diferentes personalizaciones de las herramientas seleccionadas.
- Especificar los mapeos de datos entre el Almacén de datos y los sistemas transaccionales.
- Definir el modelo de datos del Almacén de datos. El modelo de datos conceptual Almacén de datos.
- Diseñar el Almacén de datos.
- Definir la Arquitectura tecnológica del Almacén de datos.
- Especificar la extracción, transformación y carga de datos entre los sistemas transaccionales y el Almacén de datos.

En donde la extracción es la elección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico Almacén de datos; la transformación es el proceso para transformar, sumar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a los temas a explotar y finalmente la carga es la inserción sistemática de datos en el componente de almacenamiento físico Almacén de datos

- Definir las interfaces de Usuario para el acceso y explotación de los datos.

La tabla 2-2 muestra las principales tareas y subtareas, a qué metodología se corresponden, qué producto se obtiene al realizarla, qué técnica se utiliza y una breve descripción de dicha tarea.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Planificación del Sistema			El objetivo de esta fase es obtener un marco de referencia para el desarrollo del sistema de información.
Gestión de la Configuración del Sistema			
Especificar la política de gestión de la configuración para los componentes del sistema	Plan de Gestión de Configuración	N/A	Se realiza la estrategia de gestión de configuración que regirá durante todo el proyecto. Se decide que circuito se sigue para los cambios como así también que componentes serán objeto de control de configuración.
Gestión de la Calidad del Sistema			
Especificar la política de gestión de calidad para el sistema	Plan de Gestión de Calidad	N/A	Se realiza un plan de control de calidad donde se especifica en términos generales la estrategia a seguir para evaluar la calidad en el sistema. Esta estrategia apunta fundamentalmente a las pruebas que se realizan, cómo se realizan, quiénes son los participantes y responsables.
Estudio de Viabilidad del Sistema			El objetivo del Estudio de Viabilidad del Sistema es el análisis de un conjunto concreto de necesidades para proponer una solución a corto plazo, que tenga en cuenta restricciones económicas, técnicas, legales y operativas.
Establecimiento del Alcance del Sistema			En esta actividad se estudia el alcance de la necesidad planteada por el cliente o usuario, o como consecuencia de la realización de un Plan, realizando una descripción general de la misma. Se determinan los objetivos, se inicia el estudio de los requisitos y se identifican las unidades organizativas afectadas estableciendo su estructura. Se analizan las posibles restricciones, tanto generales como específicas, que puedan condicionar el estudio y la planificación de las alternativas de solución que se propongan.
Estudio de la Solicitud	Descripción General del Sistema Catálogo Objetivos EVS Catálogo de Requisitos	Secciones de Trabajo Catalogación	Se realiza una descripción general de la necesidad planteada por el usuario. Antes de iniciar el estudio de los requisitos del sistema se establecen los objetivos generales del Estudio de Viabilidad.
Identificación del Alcance del Sistema	Descripción General del Sistema Catálogo Objetivos EVS Catálogo de Requisitos	Secciones de Trabajo Catalogación Diagrama de Paquetes	Se analiza el alcance de la necesidad planteada y se identifican las posibles restricciones.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Especificación del Alcance del Estudio de Viabilidad del Sistema (EVS)	Catálogo de Objetivos del EVS Catálogo de Usuarios	Secciones de Trabajo Catalogación	En función del alcance del sistema y los objetivos del Estudio de Viabilidad del Sistema, se determinan las actividades y tareas a realizar. En particular, hay que decidir si se realiza o no el estudio de la situación actual y, en el caso de considerarlo necesario, con qué objetivo. Se identifican los usuarios participantes de las distintas unidades organizativas afectadas para la realización del Estudio de Viabilidad del Sistema
Estudio de la Situación Actual			La situación actual es el estado en el que se encuentran los sistemas de información existentes en el momento en el que se inicia su estudio. Teniendo en cuenta el objetivo del estudio de la situación actual, se realiza una valoración de la información existente acerca de los sistemas de información afectados. En función de dicha valoración, se especifica el nivel de detalle con que se debe llevar a cabo el estudio. Como resultado de esta actividad se genera un diagnóstico, estimando la eficiencia de los sistemas de información existentes e identificando los posibles problemas y las mejoras.
Determinar el área del negocio a ser soportadas en la primera fase del desarrollo.	Selección del área de comienzo.	Secciones de Trabajo	En esta tarea se debe determinar el área de negocio donde se comienza a construir el sistema. La presente tesis abarca una sola área o departamento. Luego, fuera del alcance de la tesis se podrá continuar con otras áreas.
Descripción de los Sistemas de Información Existentes (Determinar que sistemas transaccionales están asociados a las áreas de negocio seleccionadas.)	Descripción de los módulos de los sistemas actuales	Secciones de Trabajo Diagrama de Representación/Paquetes	En esta tarea se describen los sistemas de información existentes afectados, según el alcance y nivel de detalle establecido en la tarea
Descripción de los Sistemas de Información Existentes			Valoración del Estudio de la Situación Actual
Realización del Diagnóstico de la Situación Actual	Diagnóstico de la Situación Actual	Secciones de Trabajo	Con el fin de elaborar el diagnóstico de la situación actual se analiza la información de los sistemas de información existentes, obtenida en la tarea anterior y se identifican problemas, deficiencias y mejoras. Estas últimas deben tenerse en cuenta en la definición de los requisitos.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Definición de Requisitos del Sistema		Secciones de Trabajo Catalogación	Esta actividad incluye la determinación de los requisitos generales, mediante un conjunto de sesiones de trabajo con los usuarios participantes. Una vez finalizadas, se analiza la información obtenida definiendo los requisitos y sus prioridades, que se añaden al catálogo de requisitos que sirve para el estudio y valoración de las distintas alternativas de solución.
Identificación de Requisitos	Identificación de Requisitos	Secciones de Trabajo	En esta actividad se enumeran los requisitos generales, los cuáles se deberán refinar en tareas posteriores.
Catalogación de Requisitos	Catalogación de Requisitos	Catalogación	Se analiza la información obtenida en las sesiones de trabajo para la Identificación de Requisitos, definiendo y catalogando los requisitos (funcionales y no funcionales) que debe satisfacer el sistema.
Estudio de Alternativas de Solución			Este estudio se centra en proponer diversas alternativas que respondan satisfactoriamente a los requisitos planteados, considerando también los resultados obtenidos en el Estudio de la Situación Actual.
Preselección de Alternativas de Solución	Alternativas de Solución a Estudiar		Una vez definidos los requisitos a cubrir por el sistema, se estudian las diferentes opciones que hay para configurar la solución.
Descripción de alternativas de Solución (Estudiar y analizar diferentes herramientas para OLAP y Minería de datos. Herramientas de Acceso a Datos)	Descripción del Producto	Secciones de Trabajo Diagrama de Representación	Se describen las alternativas de solución consideradas.
Obtención de Requisitos	Catálogo de Requisitos Modelo de Casos de	Secciones de Trabajo Catalogación Casos de Uso/Diagrama de	En esta tarea comienza la obtención detallada de información mediante sesiones de trabajo con los usuarios.
Especificación de Casos de Uso	Catálogo de Requisitos Modelo de Casos de Uso Especificación de Casos de Uso	Secciones de Trabajo Catalogación Casos de Uso/Diagrama de Paquetes	El objetivo de esta tarea es especificar cada caso de uso identificado en la tarea anterior.
Análisis de Requisitos	Catálogo de Requisitos Modelo de Casos de Uso Especificación de Casos de Uso	Secciones de Trabajo Catalogación Casos de Uso	En esta tarea se estudia la información capturada previamente en esta actividad, para detectar inconsistencias, ambigüedades, duplicidad o escasez de información, etc. El análisis de los requisitos y de los casos de uso asociados permite identificar funcionalidades o comportamientos comunes, reestructurando la información de los casos de uso a través de las generalizaciones y relaciones entre ellos.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Validación de Requisitos	Catálogo de Requisitos Modelo de Casos de Uso Especificación de Casos de Uso	Secciones de Trabajo Catalogación Casos de Uso	Mediante esta tarea, los usuarios confirman que los requisitos especificados en el catálogo de requisitos, así como los casos de uso, son válidos, consistentes y completos.
Identificación de los subsistemas de análisis			El objetivo de esta actividad es facilitar el análisis del sistema de información llevando a cabo la descomposición del sistema en subsistemas.
Determinación de Subsistemas de Análisis	Descripción de Subsistemas de Análisis Descripción de Interfaces entre	Diagrama de Paquetes	La descomposición del sistema en subsistemas debe estar, principalmente, orientada a los procesos de negocio, aunque también es posible adoptar otros criterios lógicos.
Integración de Subsistemas de Análisis	Descripción de Subsistemas de Análisis Descripción de Interfaces entre Subsistemas	Diagrama de Paquetes	Objetivo de esta tarea es la coordinación en la elaboración de los distintos modelos de análisis de cada subsistema, asegurando la ausencia de duplicidad de elementos y la precisión en la utilización de los términos del glosario.
Elaboración del modelo de datos			El objetivo de esta actividad es identificar las necesidades de información de cada uno de los procesos que conforman el sistema de información, con el fin de obtener un modelo de datos que contemple
Elaboración del modelo de datos			Todas las entidades, relaciones, atributos y reglas de negocio necesarias para dar respuesta a dichas necesidades.
Definir el origen de los datos operacionales. El modelo de datos fuente.	Modelo de datos origen	Secciones de Trabajo Modelo Entidad/Relación	El objetivo de esta tarea es determinar el origen de los datos de los sistemas transaccionales. Se determina la o las bases de datos fuentes.
Documentar las bases de datos de los sistemas transaccionales	Modelo de datos origen	Secciones de Trabajo Modelo Entidad/Relación	En esta tarea se debe obtener el modelo de datos de las bases de datos fuente. Este modelo de datos se debe estudiar y analizar a efectos de tener conocimientos de las estructuras y objetos de BD intervinientes en los sistemas transaccionales.
Determinar las superposiciones y relaciones entre los sistemas transaccionales	Modelo de datos origen	Modelo Entidad/Relación	Se debe determinar las superposiciones o duplicidad de datos en las bases de datos fuentes, en caso de haber varias bases de datos como así también en caso de una sola base fuente.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Definir el modelo de datos del DW. El modelo de datos conceptual DW	Modelo de datos del DW/Datamarts	Modelo Entidad/Relación	A partir de los modelos de datos fuentes y según requisitos se define el modelo de datos del Datawarehouse y Datamarts.
Especificar los mapeos de datos entre el DW y los sistemas transaccionales.	Modelo de datos del DW/Datamarts	Modelo Entidad/Relación	Se realizan los mapeos entre los modelos de datos fuentes y el Almacén de datos/Datamart. Esta tarea básicamente Es de refinamiento del modelo de datos del Almacén de datos/Datamarts contemplando los procesos de extracción, transformación y carga entre las diferentes bases de datos.
Definir las interfaces de usuario			El objetivo es realizar un análisis de los procesos del sistema de información en los que se requiere una interacción del usuario, con el fin de crear una interfaz que satisfaga todos los requisitos establecidos, teniendo en cuenta los diferentes perfiles a quienes va dirigido.
Especificación de Principios Generales de la Interfaz	Principios generales de la interfaz	Secciones de Trabajo	El objetivo de esta tarea es especificar los estándares, directrices y elementos generales a tener en cuenta en la definición de la interfaz de usuario, tanto para la interfaz interactiva
Definir las diferentes personalizaciones de las herramientas seleccionadas.	Modelo de casos de Uso	Casos de Uso	Se comienza a especificar las personalizaciones de las herramientas en términos de funcionalidad.
Especificar la extracción, transformación y carga de datos entre los sistemas transaccionales y el DW.	Modelo de casos de Uso	Casos de Uso	Se especifican los procesos ETS de extracción, transformación y carga entre las bases de datos fuentes y bases de datos destino. También se definen, si es necesario procesos ETLs entre el Almacén de datos y Datamarts.
Análisis de Consistencia y Especificación de Requisitos			El objetivo de esta actividad es garantizar la calidad de los distintos modelos generados en el proceso de Análisis del Sistema de Información
Verificación de los modelos	Modelo de Casos de Uso/ Especificación de Casos de Uso Descripción de Subsistema de Análisis		
Validación de los Modelos	Modelo de Casos de Uso/ Especificación de Casos de Uso Descripción de Subsistemas de Análisis		

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Aprobación del Análisis de Información			En esta tarea se aprueba el Análisis del Sistema.
Presentación y Aprobación del Análisis del Sistema de Información	Aprobación del Análisis de Información	Presentación	En esta tarea se realiza la presentación del análisis del sistema de información para la aprobación final del mismo.
Diseño del sistema de información			El objetivo del proceso de Diseño del Sistema de Información es la definición de la arquitectura del sistema y del entorno tecnológico que le va a dar soporte, junto con la especificación detallada de los componentes del sistema de información.
Definir la arquitectura del sistema			En esta actividad se define la arquitectura general del sistema de información, especificando las distintas particiones físicas del mismo, la descomposición lógica en subsistemas de diseño y la ubicación de cada subsistema en cada partición, así como la especificación detallada de la infraestructura tecnológica necesaria
Definir la arquitectura del sistema			Para dar soporte al sistema de información.
Identificación de Subsistemas de Diseño	Diseño de la Arquitectura del Sistema Descripción de Subsistemas de Diseño	Diagrama de Paquetes Diagrama de Interacción de Objetos	En esta tarea se divide de forma lógica el sistema de información en subsistemas de diseño, con el fin de reducir la complejidad y facilitar el mantenimiento. Hay que tomar como referencia inicial los subsistemas de análisis especificados en el proceso de Análisis del Sistema de Información
Diseño de los Casos de Uso Reales			Esta actividad tiene como propósito especificar el comportamiento del sistema de información para un caso de uso, mediante objetos o subsistemas de diseño que interactúan.
Diseño de la Realización de los Casos de Uso	Diseño de la Realización de los Casos de Uso	Casos de Uso Diagrama de Interacción de Objetos	El objetivo de esta tarea es definir cómo interactúan entre sí los objetos identificados en la tarea anterior para realizar, desde un punto de vista técnico, un caso de uso del sistema de información. Para ello, se parte de los escenarios especificados en el análisis, y se detallan teniendo en cuenta que se deben llevar cabo sobre un entorno tecnológico
Diseño Físico de los Datos			En esta actividad se define la estructura física de datos que utilizará el sistema, a partir del modelo lógico de datos.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Diseño del Modelo Físico de Datos	Modelo Físico de Datos	Reglas de obtención del Modelo Físico a partir del Modelo Lógico	El objetivo de esta tarea es realizar el diseño del modelo físico de datos a partir del modelo lógico de datos o de clases/entidades.
Definir las interfaces de Usuario para el acceso y explotación de los datos.	Modelo de Casos de Uso	Casos de Uso Diagrama de Interacción de Objetos	En esta tarea se especifica las interfaces de usuario para el acceso a los datos.
Diseñar la extracción, transformación y carga de datos entre los sistemas transaccionales y el Almacén de datos.	Modelo de Casos de Uso	Casos de Uso Diagrama de Interacción de Objetos	Se especifica con detalle todos los componentes y parametrizaciones necesarias para realizar los procesos ETLs.
Verificación y Aceptación de la Arquitectura del Sistema			El objetivo de esta actividad es garantizar la calidad de las especificaciones del diseño del sistema de información y la viabilidad del mismo.
Verificación de las Especificaciones de Diseño	Diseño de la Arquitectura del Sistema Diseño Detallado de Subsistemas de Soporte Modelo Físico de Datos Diseño de la Realización de los Casos de Uso	N/A	El objetivo de esta tarea es asegurar la calidad formal de los modelos utilizados.
Aprobación del Diseño del Sistema de Información			En esta actividad se aprueba el diseño para seguir avanzando en la construcción del sistema.
Presentación y Aprobación del Diseño del Sistema de Información	Aprobación del Diseño	Presentación	En esta tarea se realiza la presentación del diseño del sistema de información para la aprobación final del mismo.
Construcción del Sistema de Información			En este proceso se genera el código de los componentes del Sistema de Información y las personalizaciones de las herramientas Para conseguir dicho objetivo, en este proceso se realizan las pruebas unitarias, las pruebas de integración de los subsistemas y componentes de acuerdo al plan de pruebas establecido.
Preparación del Entorno de Generación y Construcción			El objetivo de esta actividad es asegurar la disponibilidad de todos los medios y facilidades para que se pueda llevar a cabo la construcción del sistema de información. Entre estos medios, cabe destacar la preparación de los puestos de trabajo, equipos físicos y lógicos, gestores de
Implantación de la Base de Datos Física	Base de Datos Física	N/A	En esta tarea se debe implementar las base de datos y su gestor.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Preparación del Entorno de Construcción	Entorno de Construcción	N/A	En esta tarea se prepara el entorno en el que se construirán los componentes del sistema de información.
Generar el código de los componentes y personalización de herramientas	Código de Componentes y Personalización de Herramientas	N/A	En esta tarea se desarrollan y parametrizan en las herramientas correspondientes todas las especificaciones de diseño.
Ejecución de las pruebas unitarias			En esta actividad se realizan las pruebas unitarias de cada uno de los componentes del sistema de información, una vez codificados, con el objeto de comprobar que su estructura es correcta y que se ajustan a la funcionalidad
Preparación del entorno de las Pruebas Unitarias	Entorno de Pruebas Unitarias		En esta tarea se preparan todos los recursos necesarios para realizar las pruebas unitarias de cada uno de los componentes del sistema de información.
Realización y Evaluación de las Pruebas Unitarias	Resultado de Pruebas Unitarias	Pruebas Unitarias	El objetivo de esta tarea es comprobar el correcto funcionamiento de los componentes del sistema de información codificados o personalizaciones de las herramientas. Para cada verificación establecida, se realizan las pruebas con los casos de pruebas asociados, efectuando el correspondiente análisis y evaluación de los resultados, y generando un registro conforme a los criterios establecidos en el plan de pruebas. Seguidamente, se analizan los resultados de las pruebas unitarias, evaluándose las mismas para comprobar que los resultados son los esperados. Si los resultados no son los esperados hay que proceder a realizar las correcciones pertinentes.
Ejecutar las pruebas de integración			El objetivo de las pruebas de integración es verificar si los componentes o subsistemas interactúan correctamente a través de sus interfaces, tanto internas como externas, cubren la funcionalidad establecida, y se ajustan a los requisitos especificados en las verificaciones correspondientes.
Preparación de las Pruebas de Integración	Entorno de pruebas de Integración		En esta tarea se disponen todos los recursos necesarios para realizar las pruebas de integración de los componentes y subsistemas que conforman el sistema de información.

Nombre tarea	Producto	Técnica/práctica	Descripción tarea
Realización de las Pruebas de Integración	Resultado de pruebas de Integración	Pruebas de Integración	El objetivo de esta tarea es verificar el correcto funcionamiento de las interfaces existentes entre los distintos componentes y subsistemas, conforme a las verificaciones establecidas para el nivel de pruebas de integración.
Evaluación del Resultado de las Pruebas de Integración	Evaluación de las Pruebas de Integración		El objetivo de esta tarea es analizar los resultados de las pruebas de integración y efectuar su evaluación. Dicha evaluación recoge el grado de cumplimiento de las pruebas
Aprobación del Sistema de Información			En esta tarea se recopilan los productos del sistema de información y se presentan para su aprobación.
Presentación y Aprobación del Sistema de Información	Aprobación del Sistema de Información	Presentación	Se realiza una presentación del sistema para su aprobación.
Preparación de la Tesis de Master			
Armado del documento del trabajo de tesis	Documento de Tesis		En esta tarea se realiza toda la documentación de la presente tesis a efectos de su presentación y exposición.
Revisión y corrección del trabajo	Documento de Tesis		Se corrige todas las observaciones realizadas por las directoras de tesis.
Presentación final		Presentación	Se realiza la presentación de la tesis.

Tabla 3-2. Tareas para la realización del SUFESC

2.2 Gestión de la configuración

En el apartado se expone la estrategia de configuración que da soporte al desarrollo del sistema, fundamentalmente en lo relacionado al Control de Versiones y Control de Cambios. En este capítulo se discuten y definen aspectos como los procedimientos a seguir, los responsables y el soporte tecnológico entre otros.

- **Definición de las actividades de Gestión de Configuración**

El objetivo de la gestión de la configuración es mantener la integridad de los productos que se obtienen a lo largo del desarrollo de los sistemas de información, garantizando que no se realizan cambios incontrolados y que todos los participantes en el desarrollo del sistema disponen de la versión adecuada de los productos que manejan. Así, entre los elementos de configuración software, se

encuentran no únicamente ejecutables y código fuente, sino también los modelos de datos, modelos de procesos, especificaciones de requisitos, pruebas, documentos diversos, etc.

Si bien en este caso el desarrollo del sistema será realizado por una persona, no escapa a la gestión de configuración, ya que no se producirán inconvenientes de múltiples accesos a un producto software, pero sí ayudará a la visibilidad y orden de estos productos.

La gestión de configuración facilita el mantenimiento del sistema, aportando información precisa para valorar el impacto de los cambios solicitados y reduciendo el tiempo de implementación de un cambio, tanto evolutivo como correctivo, nos permite controlar el sistema como producto global a lo largo de su desarrollo, obtener informes sobre el estado de desarrollo en que se encuentra y reducir el número de errores de adaptación del sistema, lo que se traduce en un aumento de calidad del producto, de la satisfacción del cliente o usuarios y, en consecuencia, de mejora de la organización.

La gestión de configuración se realiza durante todas las actividades asociadas al desarrollo del sistema, y continúa registrando los cambios hasta que éste deja de utilizarse.

Las actividades de Control de Configuración que se llevan a cabo para mantener la integridad de los productos que se obtienen a lo largo del desarrollo del sistema son: Control de Versiones y Control de los Cambios.

Control de Versiones

A efectos de identificar las nuevas versiones de documentos, scripts, programas, ejecutables, y todo otro producto que resulta del desarrollo del proyecto.

Control de Cambios

El objetivo del control de cambios se centra básicamente en mantener un proceso permanente de detección de necesidad, evaluación, autorización de cambios o adaptaciones a los servicios, productos y plazos establecidos. Su principal meta es permitir el desarrollo ordenado de las actividades e interacciones entre el equipo de desarrollo y la universidad.

- **Definición del Plan de Gestión de la Configuración**

Al no contar la universidad con un Plan General de Gestión de Configuración, es necesario elaborar un plan propio para el sistema de información a desarrollar. El plan propuesto responde a la manera de gestionar tanto el control de versiones como el control de cambios. El plan se basa en estas dos directivas principales:

- Directivas para el Control de Versiones
- Directivas para el Control de Cambios.

- **Directivas para el Control de Versiones**

- **Los aspectos que se contemplan son:**

- *Identificación de todos los productos que deben ser controlados, su clasificación y relaciones entre ellos, así como el criterio o norma de identificación.*

Todos los productos que forman parte o intervienen en el desarrollo del SUFESC y que hayan sido seleccionados como elementos de configuración en el plan de gestión de la configuración, se denominan de manera que cada uno de ellos sea perfectamente identificable de forma única. La identificación de los productos se realiza cuando aparecen por primera vez en el sistema de gestión de la configuración, registrándose como la primera versión del producto en el estado que se establezca. También se define una línea base como punto de referencia en el proceso de desarrollo de software que queda marcado por la aprobación de uno o varios elementos de configuración mediante una revisión o aceptación.

Los productos de cada fase son:

- Planificación: Documentos asociados a las tareas de planificación del proyecto.
 - Documento de Plan de proyecto
 - Documento de Plan de Calidad
 - Documento de Plan de Configuración.
- Estudio de Viabilidad: Documentos asociados a las tareas relacionadas con el Estudio de Viabilidad.
 - Documento de Estudio de Viabilidad
 - Documento de Marco Conceptual.

- Análisis: Documentos asociados al análisis del sistema.
 - Documento de Análisis del Sistema
 - Archivo de Rational Rose con el Análisis del Sistema
 - Archivo de ERWin con los modelos de datos.
- Diseño: Documentos asociados al diseño del sistema.
 - Documento MS Word de Diseño del Sistema
 - Archivo de con el Diseño del Sistema Relacional
 - Archivo de E- R con los modelos de datos físicos.
- Codificación-Implementación: Programas creados, fuentes y ejecutables. Algunos productos como librerías, scripts, ejecutables tienen un nombre libre, pero siempre teniendo en cuenta que dicho nombre debe ser representativo.
 - Archivos de DTS (configuración de procesos de Extracción, Transformación y Carga)
 - Archivo de Interfaz de Usuario
- Control de Calidad: Documentos asociados a las tareas de gestión de la calidad del sistema.
 - Documento de Verificación de Calidad
 - Archivo de Base de datos del Sistema Académico (con datos de prueba).

Cada uno de los elementos creados se identifica de la siguiente forma por medio de las facilidades de la herramienta de Control de Versiones.

- Identificación del proyecto al que pertenece el elemento de configuración
- Identificación a la línea base a la que pertenece
- Identificación de la fase y subfase en que se creó
- Tipo de elemento de configuración de software
- Autor
- Estado
- Nombre o código del elemento

- N°. de versión
- N°. de entrega
- Fecha de última entrega
- Observaciones.
- **Ubicación y localización de los productos.**

Todos los productos serán gestionados por la herramienta Visual que automatizara las facilidades de Gestión de Configuración.

- *Definición del ámbito y alcance del control de la configuración, describiendo los procesos incluidos en él.*

El ámbito y alcance de la configuración es mantener la integridad y control de los productos que se obtienen a lo largo del desarrollo de los sistemas de información, garantizando que no se realizan cambios incontrolados y que todos los participantes en el desarrollo del sistema disponen de la versión adecuada de los productos que manejan. Así, entre los elementos de configuración software, se encuentran programas ejecutables y código fuente, los documentos que contienen los modelos de datos, modelos de procesos, especificaciones de requisitos y pruebas, entre otros.

La gestión de configuración se realiza durante todas las actividades asociadas al desarrollo del sistema.

- *Definición del ciclo de estados para cada tipo de producto y los criterios de trazabilidad entre los mismos.*

Durante los procesos de Análisis, Diseño, Construcción e Implantación del Sistema de Información se realizarán las actividades de identificación y registro previstas en el Plan de Gestión de Configuración, consiguiendo así mantener la consistencia entre las distintas versiones de los productos de desarrollo.

Las actividades de identificación y registro interactúan continuamente controlando y gestionando sus productos y estableciendo versiones hasta que el producto se encuentra correctamente finalizado y aceptado. Según se van generando los productos a lo largo de las actividades de un proceso, se

registran en el sistema de gestión de la configuración con el estado correspondiente.

Durante la realización de las distintas actividades, los productos obtenidos en función de su naturaleza, van pasando por distintos estados, registrándose en el sistema de gestión de la configuración. No todos los productos pasan por los mismos estados. Los estados posibles son: en elaboración, finalizado, revisado y aceptado.

Antes de ser aceptado, un producto puede sufrir numerosos cambios, e incluso después, puede ocurrir que tenga que ser modificado. Esto implica que el producto es registrado en el SGC con una nueva versión y en el estado correspondiente, de manera que entra, de nuevo, en un proceso de cambio hasta que concluya su ciclo de estados.

Otro nivel de control establecido por la gestión de configuración es el control de procesos, que facilita el conocer la situación de un sistema de información a lo largo de su desarrollo. Para establecer adecuadamente este control, las actividades de gestión de configuración, como ya se ha mencionado, registran el conjunto de productos que se obtiene al final de un proceso como un producto más, de esta forma se le pueden atribuir estados que permiten controlar el desarrollo del sistema a nivel de procesos.

- **Descripción de funciones y responsabilidades.**

Las funciones de Control de Configuración están bajo la responsabilidad del tesista. La responsabilidad principal del es llevar adelante el proyecto fundamentalmente en la parte de análisis, diseño conceptual del sistema.

- ***Generación de Informes.***

Estos informes están compuestos por la información que emite la herramienta de Control de Configuración y contienen el siguiente detalle:

- Nombre y ubicación: Nombre del elemento de configuración dentro de la herramienta de configuración
- Versión: Nº de Versión de elemento de configuración
- Identificación del proyecto a que pertenece el elemento de configuración: Nombre del proyecto
 - Identificación a la línea base a que pertenece: Nombre de la línea base.
- Identificación de la fase y subfase en que se creó: Nombre de la fase/subfase

- Tipo de elemento de configuración de software: Documento MS Word/Archivo, Excel/Archivo plano/archivo etc.
- Autor: Nombre del autor del elemento
- Estado: Creado/elaboración/finalizado/revisado y aceptado
- N° de entrega: número de entrega del elemento de configuración
- Fecha de Entrega: Fecha en que fue entregado el elemento (si aplica)
- Fecha de Actualización: Fecha de última actualización
- Observaciones: Notas y observaciones varias.

Además del informe anterior, la herramienta dispone de varios tipos de informes como por ejemplo, reporte de diferencia entre archivos y reporte de historial, entre otros.

- *Identificación de la información necesaria de control para auditoría.*
La información para el control de auditoría está basada en las características de la herramienta a utilizar para el control de versiones.

2-3. Directivas para el Control de Cambios

Está conformado por:

- La solicitud de cambio
- La evaluación de factibilidad
- La autorización y el seguimiento de los cambios.

Todo este proceso se realiza siguiendo un Procedimiento de Cambios. Se entiende por cambio a todo aquello que produzca un desfase en los productos, en los servicios, en la modalidad, en los plazos, en los requerimientos mutuos, es decir, en todo aquello que difiera de lo previsto originalmente.

Se describen brevemente los pasos que se siguen para mantener ordenado y bajo control la administración de todos los cambios que se produzcan.

- **Solicitud de Cambio**

Se entiende por "Solicitud de Cambio" a todo pedido que la universidad realice a través del personal expresamente autorizado, para que se ejecute nuevas tareas o

provisiones o modificaciones, que no hayan sido establecidas en propuesta. Asimismo, incluye cambios sugeridos por el responsable de la Tesis.

En la Solicitud de Cambio, según modelo adjunto (Formulario 2-1), se identificará, al menos:

Identificación del proyecto

- Nombre y cargo de la persona que origina el pedido de cambio
- Fecha de la solicitud del cambio
- Descripción del cambio propuesto
- Razón para el cambio propuesto.

Este documento será presentado a los responsables de la universidad. El responsable verifica el pedido, y si decide aprobarlo, le asigna:

- N° correlativo de Solicitud de cambio
- Nombre y cargo del Responsable del Proyecto
- Fecha de revisión de la Solicitud de cambio
- Categoría de la Solicitud de cambio (Crítico, Importante o Deseable).

Luego lo entrega al Responsable del Equipo de Desarrollo, es este caso, el responsable de la Tesis.

La figura 2-3 muestra el circuito del Control de Cambio para el SUFESC

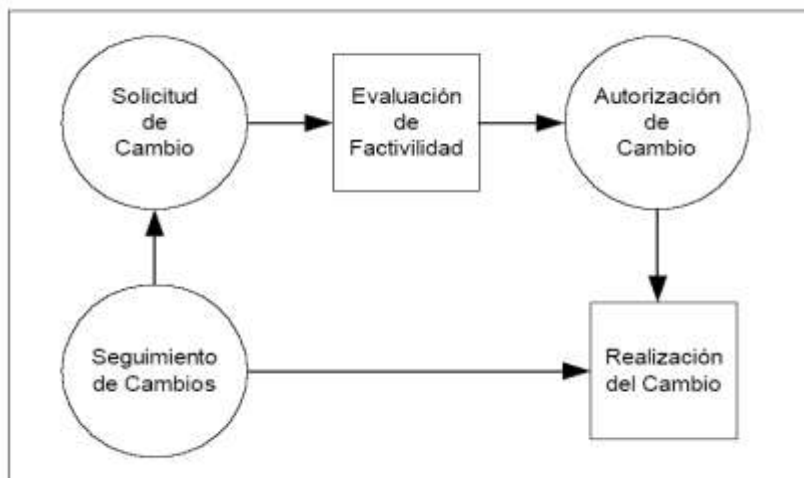


Figura 2-3. Proceso de solicitud de cambio.

- **Autorización de Cambios**

El responsable de la universidad tomará la decisión de cuándo o no implementar el cambio. La decisión será comunicada al Responsable del Desarrollo vía el formulario debidamente autorizado.

De ser aceptado el cambio y el impacto en el Proyecto o Cronograma que éste pueda producir, se tomarán las acciones necesarias para implementarlo y se realizará al finalizar la aceptación del cambio realizado.

- **Seguimiento de Cambios**

Las acciones descriptas anteriormente forman un ciclo continuo que permitirá documentar y formalizar los cambios que sean necesarios para cualquiera de los servicios, plazos y productos dentro del proyecto. Además todos los formularios, Serán gestionados por el Control de Versiones como un producto más del desarrollo del proyecto.

SOLICITUD DE CAMBIO			
Formulario N°	Proyecto SUFESC Módulo	Responsable	Fecha Emisión
Descripción			
Razón			
Prioridad: <input type="checkbox"/> Crítico <input type="checkbox"/> Importante <input type="checkbox"/> Deseable			
Origen de la solicitud:		Fecha:	
Responsable de la universidad:		Fecha:	
Impacto al Proyecto: Plan: Costo Hs:			
Aprobación			
Prioridad: <input type="checkbox"/> Crítico <input type="checkbox"/> Importante <input type="checkbox"/> Deseable			
Decisión: Realizar el cambio: No			
Por la universidad (firma, aclaración y fecha)	Responsable Desarrollo (firma, aclaración y fecha)	Otros	

Solicitud de Cambio, según modelo adjunto Formulario 2-1

CAPITULO 3 GESTIÓN DE LA CALIDAD

En el Capítulo 3 se detalla la estrategia general para el Aseguramiento de Calidad. Esta estrategia apunta fundamentalmente a las pruebas que se realizan, cómo se realizan, quiénes son los participantes y responsables.

• **Introducción a la Calidad en el SUFESC**

El objetivo del Plan de Aseguramiento de la Calidad es proporcionar un marco de referencia para el control de la calidad en el desarrollo del SUFESC.

La calidad se define como **"grado en que un conjunto de características inherentes cumple con unos requisitos"** [SCSI, 2000].

El Aseguramiento de la Calidad pretende dar confianza en que el producto reúne las características necesarias para satisfacer todos los requisitos del Sistema de Información.

Por tanto, para asegurar la calidad de los productos resultantes se realizan un conjunto de actividades que sirven para:

- Reducir, eliminar y lo más importante, prevenir las deficiencias de calidad de los productos a obtener.
- Alcanzar una razonable confianza en que las prestaciones y servicios esperados por el cliente queden satisfechas.

Para conseguir estos objetivos, es necesario desarrollar un plan de aseguramiento de calidad específico que se aplica durante la planificación del proyecto de acuerdo a la estrategia de desarrollo adoptada en la gestión del proyecto. En este plan de aseguramiento de calidad se reflejan en términos generales las actividades de calidad a realizar (normales o extraordinarias), los estándares a aplicar, los productos a revisar, los procedimientos a seguir en la obtención de los distintos productos durante el desarrollo y la normativa para informar de los defectos detectados a sus responsables y realizar el seguimiento de los mismos hasta su corrección.

El equipo de Aseguramiento de Calidad participa en la revisión de los productos seleccionados para determinar si son conformes o no a los procedimientos, normas o criterios especificados, siendo totalmente independiente del equipo de desarrollo. Las actividades realizadas por el grupo de aseguramiento de calidad vienen gobernadas por el plan. Sus funciones están dirigidas a:

- Identificar las posibles desviaciones en los estándares aplicados, así como en los requisitos y procedimientos especificados

- Comprobar que se han llevado a cabo las medidas preventivas o correctivas necesarias.

Las revisiones son una de las actividades más importantes del aseguramiento de la calidad, debido a que permiten eliminar defectos lo más pronto posible, cuando son menos costosos de corregir.

A continuación se detallan las características principales del Plan de Calidad para el desarrollo del SUFESC.

- **Constitución del Equipo de Aseguramiento de Calidad**

El equipo de desarrollo del SUFESC, al menos en el marco de la presente Tesis estará formado por los Usuarios, El Directores de Tesis y el Tesista. Dentro de este equipo existe una separación de funciones de manera que el tesista prepara los casos de prueba pero no realiza la prueba.

- **Determinación de los Productos Objeto de Aseguramiento de Calidad**

Incluye los productos software desarrollados y/o personalizados para implementar el SUFESC. Estos productos incluyen los siguientes componentes:

- Programas de Extracción de Datos
- Programas de Transformación de Datos
- Programas de Carga de Datos
- Programas de Acceso y Explotación de Datos.

También serán objeto del control de calidad todos los productos relacionados con estos programas, como por ejemplo, documentos de Planificación, de Análisis, de Diseño y de Construcción.

3.1 Alcance del Plan de Aseguramiento de Calidad

- **Objetivos.**

El objetivo del Plan de Aseguramiento de la Calidad es proporcionar un marco de referencia para el control de la calidad en el desarrollo del SUFESC. Durante el desarrollo del proyecto se pueden realizar planes de prueba más detallados, si corresponde y aplica, planes que pueden estar orientados al producto específico que se quiera controlar. Se debe tener muy en cuenta que el desarrollo del SUFESC está

compuesto por diversas aplicaciones, como las nombradas anteriormente en "*Determinación de los Productos objeto de Aseguramiento de Calidad*"

- **Identificación de las Propiedades de Calidad**

Se definen para SUFESC, las propiedades que permitan evaluar la calidad. Las propiedades principales son:

- Corrección (coincidencia con los requerimientos de especificaciones funcionales).

El grado de ajuste (Número de Especificaciones Definidas Realmente Incluidas en el Producto Entregado / Número de Especificaciones Definidas) no debe ser, en ninguno de los aspectos que se evalúe, menor al 90%.

- Fiabilidad (posibilidad de asegurar un comportamiento sin anomalías dentro de un intervalo perfectamente definido menos el tiempo medio entre fallas aceptable y comprobable).

Se expresa como Tiempo Medio Entre Fallas = Tiempo Medio de Falla + Tiempo Medio de la Reparación. En el momento de la liberación del uso a los usuarios finales, el software de aplicación entregado deberá evidenciar un Tiempo Medio Entre Fallas no menor a siete días de uso.

A los efectos enunciados están excluidas solamente las fallas claramente imputables al hardware y al software de base.

- **Robustez** (comportamiento consistente ante situaciones anormales no aclaradas en la definición de especificaciones, por ejemplo: ingreso de datos incorrectos o fallas del hardware).

Fundamentalmente se verifica mediante un control exhaustivo de la correspondencia entre cada terna "Estímulo del usuario - Proceso - Respuesta al Usuario" El manejo de inconsistencias ante el ingreso de datos erróneos deberá tener una eficiencia del 100%.

- **Performance** (no degradación de la eficiencia del hardware u otros productos de software con los que cooperará el futuro sistema).

Los tiempos de respuesta, con datos reales, a nivel LAN (redes locales), para las aplicaciones interactivas, con la mitad de los puestos de trabajo operando en forma concurrente, no deberá ser más de 5 (cinco) segundos.

- **Amigabilidad** (concepto eminentemente subjetivo: cuánto incita a su uso a un determinado perfil de usuario).

Debe responder 100% a las interfaces gráficas de usuarios utilizada corrientemente en las herramientas de oficina como por ejemplo MS Office. La navegación y uso del sistema debe ser lo suficientemente intuitivo para ser utilizado por usuarios con experiencia en interfaces gráficas Windows.

- **Verificabilidad** (economicidad con la que permite evaluar sus cualidades).

Se requiere la definición por parte de los usuarios finales de prueba para la verificación de cada resultado esperado del sistema.

- **Facilidad de mantenimiento y administración** (capacidad de ser administrado y mantenido en condiciones económicas aceptables y de poder evolucionar acompañando los cambios que se produzcan en el entorno).

Se evaluará teniendo en cuenta:

- Clara definición de la funcionalidad de cada función en la documentación de análisis y diseño.
- Correspondencia estricta entre los documentos de análisis, diseño y la aplicación desarrollada.
- Se deberá disponer de todos los manuales de Administración y Mantenimiento de las herramientas de mercado utilizadas para proveer la solución.

- **Interoperatividad** (capacidad para coexistir y cooperar con otros productos de software, es decir minimizando la utilización de recursos y mediante un código de programas compacto y conciso).

Respecto de este punto se verificará la capacidad de exportar hacia los productos de automatización de oficina de mayor difusión comercial, tales como "Excel" o "Word".

- **Oportunidad** (Capacidad para ser instalado en condiciones de aseguramiento de calidad pre-establecidas, dentro de los plazos acordados).

Se considerará cumplida esta cualidad si no se producen desviaciones mayores a un 10% respecto del cumplimiento del cronograma del proyecto.

3.2. Actividades Relacionadas con el Aseguramiento de Calidad a Realizar a lo Largo del Desarrollo del Software.

La mayor parte de las actividades y tareas están enfocadas a realizar pruebas de software para alcanzar la calidad y la funcionalidad deseada del sistema.

El control de calidad de los documentos de cada fase o etapa también están sujetos a un fuerte control de calidad por parte de los responsables de la dirección de la Tesis.

Los usuarios claves involucrados participan en actividades de verificación y control de calidad como la verificación y catalogación de los requisitos, aprobación del análisis de los requisitos, verificación y aceptación de la interfaz de usuario y finalmente verificación del sistema ya desarrollado.

- **Resumen de las Pruebas**

Las pruebas al SUFESC se realizarán para verificar:

- El correcto funcionamiento de los componentes del sistema y detección de errores
- El correcto ensamblaje entre los distintos componentes
- El funcionamiento correcto de las interfaces entre los distintos subsistemas que lo componen y con el resto de sistemas de información con los que se comunica

- El funcionamiento correcto del sistema integrado de hardware y software en el entorno de operación
- Que el sistema cumple con el funcionamiento esperado y permite al usuario de dicho sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento
- Que los cambios sobre un componente de un sistema de información, no introducen un comportamiento no deseado o errores adicionales en otros componentes no modificados.

Los tipos de pruebas son:

Pruebas Unitarias

Las pruebas unitarias tienen como objetivo verificar la funcionalidad y estructura de cada componente individualmente una vez que ha sido codificado.

Pruebas de Integración

El objetivo de las pruebas de integración es verificar el correcto ensamblaje entre los distintos componentes una vez que han sido probados unitariamente con el fin de comprobar que interactúan correctamente a través de sus interfaces, tanto internas como externas, cubren la funcionalidad establecida y se ajustan a los requisitos especificados en las verificaciones correspondientes

Pruebas del Sistema

Las pruebas del sistema tienen como objetivo ejercitar profundamente el sistema comprobando la integración del sistema de información globalmente, verificando el funcionamiento correcto de las interfaces entre los distintos subsistemas que lo componen y con el resto de sistemas de información con los que se comunica.

Son pruebas de integración del sistema de información completo, y permiten probar el sistema en su conjunto y con otros sistemas con los que se relaciona para verificar que las especificaciones funcionales y técnicas se cumplen.

En esta etapa pueden distinguirse los siguientes tipos de pruebas, cada uno con un objetivo claramente diferenciado:

Pruebas funcionales. Dirigidas a asegurar que el sistema de información realiza correctamente todas las funciones que se han detallado en las especificaciones dadas por el usuario del sistema.

Pruebas de comunicaciones. Determinan que las interfaces entre los componentes del sistema funcionan adecuadamente, tanto a través de dispositivos remotos, como locales. Asimismo, se han de probar las interfaces hombre/máquina.

Pruebas de rendimiento. Consisten en determinar que los tiempos de respuesta están dentro de los intervalos establecidos en las especificaciones del sistema.

Pruebas de volumen. Consisten en examinar el funcionamiento del sistema cuando está trabajando con grandes volúmenes de datos, simulando las cargas de trabajo esperadas.

Pruebas de sobrecarga. Consisten en comprobar el funcionamiento del sistema en el umbral límite de los recursos, sometiéndole a cargas masivas. El objetivo es establecer los puntos extremos en los cuáles el sistema empieza a operar por debajo de los requisitos establecidos.

Pruebas de disponibilidad de datos. Consisten en demostrar que el sistema puede recuperarse ante fallos, tanto de equipo físico como lógico, sin comprometer la integridad de los datos.

Pruebas de facilidad de uso. Consisten en comprobar la adaptabilidad del sistema a las necesidades de los usuarios, tanto para asegurar que se acomoda a su modo habitual de trabajo, como para determinar las facilidades que aporta al introducir datos en el sistema y obtener los resultados.

Pruebas de operación. Consisten en comprobar la correcta implementación de los procedimientos de operación, incluyendo la planificación y control de trabajos, arranque y re-arranque del sistema, etc.

Pruebas de entorno. Consisten en verificar las interacciones del sistema con otros sistemas dentro del mismo entorno.

Pruebas de seguridad. Consisten en verificar los mecanismos de control de acceso al sistema para evitar alteraciones indebidas en los datos.

Pruebas de Implantación

El objetivo de las pruebas de implantación es comprobar el funcionamiento correcto del sistema integrado de hardware y software en el entorno de operación, y permitir al usuario que, desde el punto de vista de operación, realice la aceptación del sistema una vez instalado en su entorno real y en base al cumplimiento de los requisitos no funcionales especificados.

Pruebas de Aceptación

El objetivo de las pruebas de aceptación es validar que un sistema cumple con el funcionamiento esperado y permitir al usuario de dicho sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento.

Pruebas de Regresión

El objetivo de las pruebas de regresión es eliminar el efecto onda, es decir, comprobar que los cambios sobre un componente de un sistema de información, no introducen un comportamiento no deseado o errores adicionales en otros componentes no modificados.

3.3. Procedimientos para Realizar las Pruebas

A continuación se describe el procedimiento general para formalizar las pruebas del software.

Procedimiento de Generación de Casos de Prueba

Los casos de prueba (para las pruebas funcionales) están basados en lotes de datos reales, los cuáles se han obtenido de la base actual del Sistema Académico de la universidad.

Las pruebas unitarias de cálculo se realizan en base a diversas consultas (instrucciones sql) realizadas manualmente en la base de datos del Sistema Académico. De manera de poder conseguir resultados y valores para cotejarlos con los resultados que se obtienen utilizando el sistema de información gerencial. En el capítulo de pruebas se muestran los casos de prueba completos y la integración con los formularios que se describen en los apartados siguientes.

A continuación, se muestran algunos ejemplos instrucciones Sql con los resultados obtenidos desde la base de datos del sistema académico. Las tabla 3-1 muestran resultados parciales de la ejecución de las instrucciones Sql.

Nro. de Caso de Prueba	1					
Nombre de Caso de Prueba	Antigüedad de Docentes del Dpto de Ingeniería por Materia					
Instrucción SQL	<pre>SELECT mat_doc_hist.mat_cod, mat_doc_hist.doc_id, Cint(((mat_doc_hist)[mdh_fhasta]-[mat_doc_hist][mdh_fdesde])/365) AS Antigüedad, mat_doc_hist.sit_cod, mat_doc_hist.mdh_fdesde, mat_doc_hist.mdh_fhasta, mat.dpto_cod FROM (mat_doc_hist INNER JOIN mat_doc ON mat_doc_hist.mat_cod = mat_doc.mat_cod) INNER JOIN mat ON mat_doc.mat_cod = mat.mat_cod WHERE (((mat_doc_hist.sit_cod)=0) AND ((mat_doc_hist.mdh_fhasta) Is Not Null) AND ((mat.dpto_cod)="1"));</pre>					
Resultado						
mat_cod	doc_id	Antigüedad	Sit_cod	mdh_fdesde	Mdh_fhasta	dpto_cod
11.06	16	9	0	01/03/1993	27/02/2002	1
11.21	29	9	0	01/03/1992	30/07/2000	1
11.46	591	8	0	01/08/1991	30/05/1999	1
11.02	591	8	0	01/08/1991	30/05/1999	1
11.45	546	8	0	01/03/1991	27/02/1999	1
11.44	546	8	0	01/03/1991	27/02/1999	1
11.21	666	7	0	01/08/1992	27/02/1999	1
11.23	666	7	0	01/08/1992	27/02/1999	1
11.41	626	7	0	01/03/1992	27/02/1999	1
11.41	669	6	0	01/03/1993	27/02/1999	1

Tabla 3-1. Ejemplo de caso de prueba.

.Para cada uno de los tipos de prueba se suministra la siguiente información:

- Nombre de la prueba
- Productos cubiertos por la prueba
- Objetivo de la prueba
- Si se aplica, documento contra el cual debe basarse la prueba
- Breve descripción del criterio de aceptación, el criterio de entrada y el criterio de salida.

- **Ejecución y Reporte de las Pruebas**

Después que los casos han sido desarrollados y el ambiente de pruebas ha sido preparado, se está listo para iniciar la ejecución de las mismas.

El reporte claro y preciso de problemas es crítico en el proceso de pruebas. Existen dos formularios asociados con la documentación de los problemas identificados durante las pruebas.

- El Informe de Problema
- El Log de Problemas.

El **Informe de Problema** describe el problema identificado y crea una historia para cada problema.

El **Log de Problemas** se utiliza para dar seguimiento al progreso y status de cada informe de Problema, se inicia con el primer problema identificado y se actualiza constantemente hasta terminar las pruebas.

Los informes se generan para mostrar el status actual de todos los problemas reportados.

Cada usuario testeador, debe producir al final de cada día de pruebas, un resumen de pruebas, el cual provee información sobre el status de los casos ejecutados ese día, los problemas hallados y los que quedan pendientes o deben ser ejecutados nuevamente.

- **Procedimientos de Ejecución y Reporte**

El probador analiza el caso de pruebas, siguiendo las instrucciones paso a paso especificadas en el descriptivo del caso de prueba. Se puede utilizar el formulario **"Casos de Prueba"**.

Si la ejecución de un caso de prueba produce resultados incorrectos o inesperados, el probador completa un **"Informe de Problema"**. La siguiente información debe ser registrada en el informe:

- Fecha en que se ejecutó el caso
- Número de Caso o de prueba
- Descripción del problema o error
- Nombre del probador
- Documentación de soporte que documente el problema.

Al finalizar la prueba del caso, el probador, recopila y adjunta a todos los documentos que se hayan producido durante la prueba (informes, impresiones de pantalla, impresión de mensajes de error, etc.) e imprime el contenido de los archivos de prueba requeridos para verificar los resultados esperados.

3.4. Procedimiento de Corrección de Errores

Con la documentación recopilada de las pruebas se procede a la solución de los errores y observaciones detectadas.

Una vez que el o los problemas han sido solucionados y se han entregado las correcciones, se volverán a ejecutar los casos que prueban los módulos corregidos. Este ciclo se repetirá hasta que los casos produzcan los resultados esperados.

En este proceso se irá actualizando el status de cada problema en el documento **"Registro de problemas identificados"**.

Este proceso asegura la entrega final de un producto de alta calidad y finalizará con la consecuente aceptación del mismo.

Durante la ejecución de las pruebas, se tendrán reuniones de revisión en forma regular y continua, con el fin de revisar el status de la ejecución del plan, así como el status de los problemas detectados.

Se considera que las pruebas han terminado, cuando todos los casos de prueba hayan sido ejecutados exitosamente y no queden problemas a resolver.

Formularios

La figura 3-1 muestra el formulario de **Casos de Prueba**.

Casos de Prueba			
Prueba/Caso Número:		Proyecto: SUFESC	Documento Soporte:
Versión:			
Página de			
Pre-requisitos del caso de prueba:			
Entradas requeridas:			
Objeto/Descripción de la prueba:			
Paso N°:	N° de función a probar	Acción o Instrucción	Resultado esperado

Figura 3. 1. Formulario de casos de prueba.

GUÍA PARA SU LLENADO

Propósito del formulario

Describir en forma detallada cada caso de prueba, su objetivo y los pasos a seguir por parte del "probador" para ejecutarlo.

Generado por:

Desarrollador de casos de prueba.

Para uso del: probador.

Partes de que se compone:

Encabezado

Esta página contiene la información general que se describe el caso así como sus prerequisites, nombre de documentos ampliatorios, entradas necesarias y descripción de la prueba o caso por lo que deberá ser leída y entendida por el probador antes de ejecutarla prueba.

Detalle:

Esta(s) página(s) contiene(n) los pasos específicos que el probador deberá seguir para ejecutar el caso. Incluye para cada paso, el número de identificación de la/s función o funciones que están siendo probadas, las acciones a realizar por parte del probador, y los resultados que se esperan como respuesta del sistema a cada acción.

La figura 3-2 muestra el formulario ***Informe de Problema***.

<i>Informe de Problema</i>			
Caso Número:	Proyecto: SUFESC		
Fecha:	Sistema:		
Probador:			
Descripción del problema:			
Otros Documentos :			

Figura 3-2. Formulario de informe de prueba.

GUÍA PARA SU LLENADO

Propósito del formulario

Describir en forma detallada el error detectado en determinada prueba

Para uso de:

Probador.

Partes de que se compone:

Encabezado

Contiene la información del sistema a probar, el nombre del probador y del número de caso a cual pertenece la prueba.

Detalle: Descripción del problema: Explicación detallada del error encontrado. Otros Documentos: Son documentos respaldatorios que aclaran el problema detectado y que servirán para corregir los errores.

La figura 3-3 muestra el formulario de **Registro de Problemas Identificados**.

Registro de problemas identificados

Proyecto: SUFESC..... Registro de problemas: Sistema:..... Elaborado por:..... Página:de Fecha:							
Prueba N°	Problema N°	Código de severidad	Probador	Responsable corrección	Estado	Fecha Requerida	Comentarios
Número de problemas identificados:							
Severidad	"1"	"2"	"3"	"4"	"5"		

Figura 3-3. Formulario de registros de problemas identificados.

GUÍA PARA SU LLENADO

Propósito del formulario:

Registrar todos los problemas registrados durante la ejecución de los casos de prueba, llevando registro que facilite su control y seguimiento.

Generado por:

Responsable de pruebas, tomando como entrada los informes de prueba generados por los probadores.

Para uso de:

Responsable de pruebas.

Frecuencia de generación:

Se actualiza diariamente. Cada día se inicia una página nueva con un número consecutivo.

Descripción de los campos:

- Elaborado por: Nombre del autor, normalmente el responsable.
- Fecha: Fecha de identificación de cuando fueron reportados los problemas de esa página.
- Página: Número de página del registro o LOG.
- Caso Número: Número del caso en donde se identificó el problema.
- Número de problema Número Identificador del problema asignado por el coordinador de pruebas.
- Código de severidad: Código de severidad del problema.

Las siguientes definiciones describen los códigos de severidad que serán usados para clasificar los problemas:

- **Severidad 1** -Este 'detiene todo'. Este tipo de error detiene la prueba que se está corriendo y las posteriores.
- **Severidad 2** -Este código de severidad esta normalmente asociado con un error en una rutina lógica o en una función específica. Se interrumpe la prueba para esa función o rutina lógica o en una función o rutina particular, pero otras pruebas pueden continuar.

- **Severidad 3** -Este código de severidad indica un error menor en donde una rutina lógica o función no se interrumpe, pero los resultados obtenidos pueden no ser los esperados.
- **Severidad 4**-Este código de severidad indica un error menor o problema en la documentación. Las pruebas no se interrumpen.

También puede agregarse el estado de **Observación 5**. Este estado indica que es una propuesta de mejora, no un error. Si la mejora es leve se analizará y se realizará. Si en el análisis de la observación se determina que es de cierta envergadura se deberá proponerla como un cambio siguiendo el procedimiento establecido para estos casos.

- **Probador:** Nombre del probador que identificó el problema.
- **Responsable:** Nombre de la persona a la que se asignó el problema para su solución.
- Generalmente será el tesista.
- **Status:** Status del problema Puede ser "abierto" cuando aún está en proceso de solución; "CERRADO": cuando ya se solucionó y la re-prueba fue exitosa.
- **Fecha de solución:** Si se trata de un problema ABIERTO, se pone la fecha PLANEADA de solución. Si se trata de un problema CERRADO, se actualiza este campo con la fecha real, de solución..
- **Comentario:** Este campo se puede usar para hacer anotaciones adicionales.
- **Total de problemas identificados:** Es el número total de problemas que se han identificado hasta esa página del LOG.
- **Severidad "1", "2", "3", "4":** En cada uno de estos campos se pone el total de problemas identificados con esa severidad hasta esa página del LOG.

3.5. Métodos para la Salvaguarda y Mantenimiento de la Documentación Obtenida en las Actividades de Aseguramiento de Calidad

Toda la documentación asociada al aseguramiento de calidad estará soportada por el Sistema de Control de Configuración. Los productos relacionados con la calidad van desde el presente documento, documentos específicos de análisis, diseño, implementación hasta los documentos de prueba.

CAPITULO 4. ESTUDIO DE VIABILIDAD DEL SISTEMA

En el Capítulo 4 se realiza el estudio de la viabilidad del sistema. Para llevar adelante este estudio se definen cuantitativamente los objetivos generales del sistema, las restricciones para poder cumplir con los objetivos y necesidades de información, los requisitos y alcances generales, se especifican las unidades organizacionales que participan y los usuarios del sistema. Se realiza un estudio de la situación actual de los sistemas de la Universidad desde el punto de vista de brindar información para la toma de decisiones en donde se describen brevemente los sistemas de información existentes. Se comienza con la catalogación de requisitos.

También en este capítulo se estudian las alternativas de solución teniendo en cuenta las características del sistema a construir. Se realiza una preselección de herramientas y un breve análisis de las características de las mismas. Luego se seleccionan las que ayudan a construir el SUFESC.

- **Objetivo General del Sistema**

El Sistema de Apoyo Gerencial Universitario - SUFESC - debe contemplar la siguiente premisa:

"El SUFESC debe brindar un conjunto de facilidades que permitan utilizar la información disponible para hacer mejores análisis, descubrir nuevas oportunidades y tomar decisiones más informadas".

Es decir, los objetivos principales son:

- **Facilitar los procesos de toma de decisiones.** Al incorporar un sistema de información específico, cambiará el proceso de toma de decisiones. Es decir, cambia el circuito de solicitud, búsqueda, preparación, entrega de información para finalmente tomar la decisión. Según la propia experiencia de implementación de este tipo de proyectos se facilita en más de un 40 % dicho proceso debido a la disponibilidad de la información para los usuarios, de

manera que se evita solicitarla al área de sistemas, luego sistemas debe buscar los datos correctos y preparar el informe para luego enviarlo al solicitante.

Para verificar si se alcanza este objetivo se recomienda antes de implementar el SUFESC, calcular los tiempos que insume actualmente el proceso de toma de decisiones y posteriormente luego de implementar el sistema volver a medir el tiempo del proceso en cuestión. De esta manera se tendrá valores comparativos que certificarán el alcance del objetivo planteado.

- **Mejorar la entrega de Información a diferentes áreas.**

Relacionado con el objetivo anterior, la entrega de la información con este sistema automatiza en más de un 90 %, mejorando considerablemente los tiempos de dicha entrega.

Para verificar si se alcanza este objetivo se recomienda realizar la misma práctica que en el objetivo anterior.

- **Sentar las bases para la investigación en "Inteligencia de Negocios en la Universidad.**

Este objetivo contempla la incorporación de investigaciones y preparación de personal de la universidad con habilidades en asesoramiento y consultoría en sistemas de Inteligencia de Negocios. Una medida de este objetivo es crear al menos un grupo o área de investigaciones en la universidad que se ocupe del estado actual y evolución de los sistemas de toma de decisiones.

La propuesta del SUFESC está encuadrada en la implementación de un sistemas del tipo "Inteligencia de Negocios", en donde se permita formular y responder las preguntas claves sobre el funcionamiento de la Universidad accediendo directamente a los indicadores de éxito, señalar cuáles son los factores que realmente inciden en el buen o mal funcionamiento, detectar situaciones fuera de lo normal, encontrar los factores que maximicen el beneficio y predecir el comportamiento futuro con un alto porcentaje de certeza.

Para esto es conveniente implementar un Sistema de Apoyo Gerencial. El sistema debe estar orientado a brindar información interrelacionada para quienes tienen responsabilidades en el ámbito estratégico y táctico de la organización.

Inicialmente, los usuarios de este sistema serán:

- **Responsables del Coordinación de Informática**
- **Dirección.**

La implementación del Sistema de Apoyo Gerencial abarca únicamente la implementación en el Coordinación de Informática según se ha acordado con las autoridades de la universidad. Posteriormente, fuera del alcance de la tesis se continuará la implementación en el resto de los departamentos y áreas de la universidad.

4-1. Requisitos y Alcances Generales

Como se propuso, el SUFESC está compuesto fundamentalmente por las características principales de los sistemas de información EIS y DSS.

Los requisitos principales de estos tipos de sistemas que se implementan en la presente tesis son:

Requisitos generales de un EIS:

- Interfaz gráfica fácil de usar y ver
- Tableros de Control o Indicadores
- Administración de una sola página.

Requisitos generales de un DSS:

- Análisis Multidimensional (OLAP).

Además, para dar soporte a los EIS y DSS se construye un **Datawarehouse** en donde residen los datos extraídos y transformados desde los sistemas fuentes, el alcance de la presente tesis con respecto al sistema fuente es el Sistema Académico. El Datawarehouse da soporte a las funciones OLAP de servidor.

Este Datawarehouse está construido de tal forma que se contempla una futura evolución, como se ha propuesto, ya sea en tamaño como funcional.

Para más detalle de las características funcionales y de arquitectura de los sistemas EIS y DSS como de los Datawarehouses .

- **Restricciones**

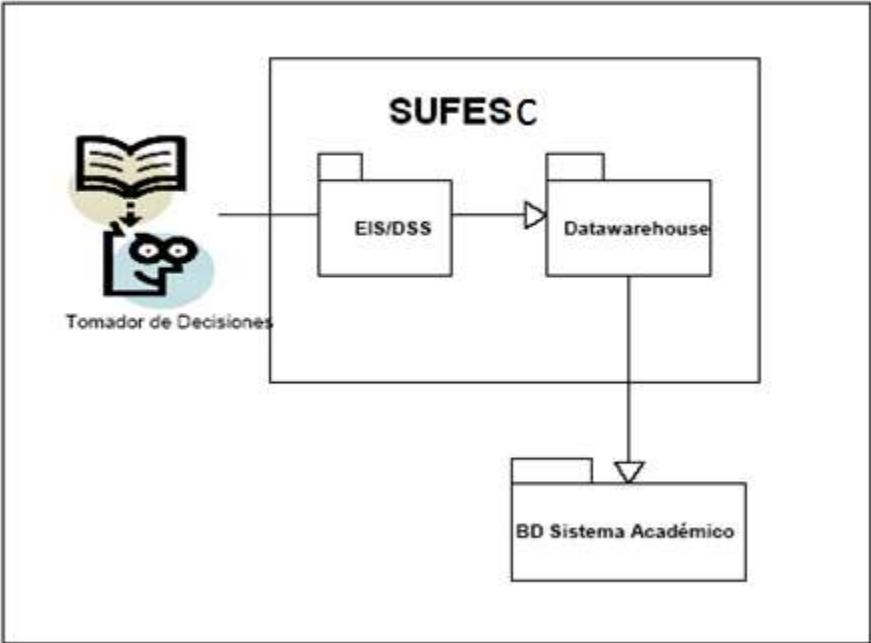
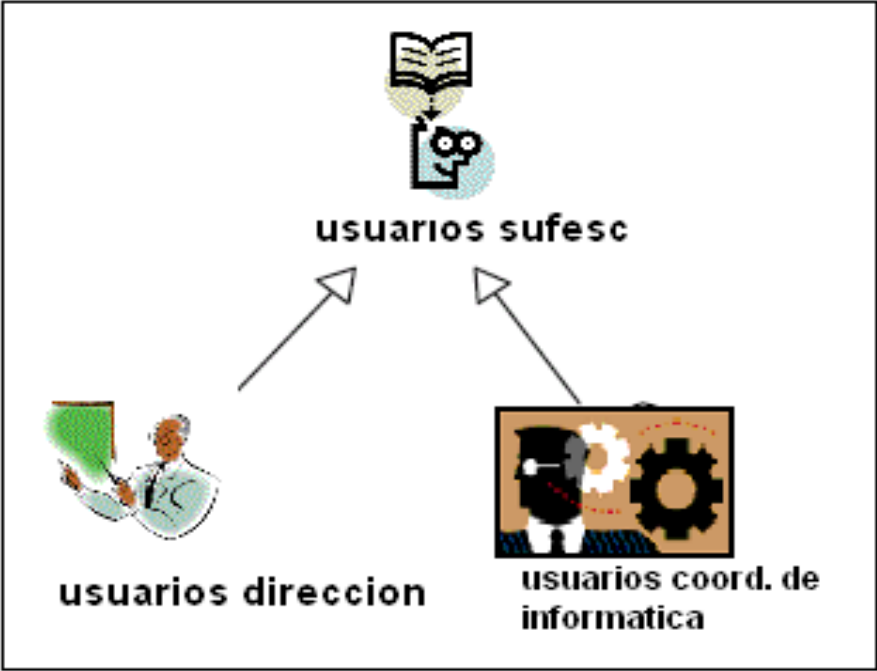
Algunas de las restricciones para llevar adelante el proyecto están dadas por:

- La limitación de elección de las herramientas para la construcción del sistema. La Coordinación de Informática dispone de una tecnología que si bien puede dar soporte y facilidades al desarrollo, implementación y mantenimiento del sistema, existen otras, las cuáles se discutirán brevemente en este capítulo que por motivos técnicos y económicos no se podrán utilizar.
- También se debe tener en cuenta la arquitectura tecnológica que dispone sus salas de Cómputo, es decir, las herramientas que se seleccionen deben ser compatibles con dicha arquitectura y sus estándares.
- Otra restricción, la más importante, es debida a la falta de información o en algunos casos, incongruencias en los datos de la base de datos del sistema fuente. Esta falta de datos e incongruencias dificulta las pruebas en incluso la realización de algunos requisitos de usuarios.

4.2. Contexto del Sistema

La figura 5-1 muestra en diagrama de contexto del sistema. Como se observa, en términos muy generales está compuesto por un Datawarehouse, en donde residen los datos a explotar, las herramientas o sistemas de análisis y uso de esos datos, extrayéndolos del Datawarehouse. Finalmente, las bases de datos fuentes del sistema

de transacciones, que es la base de datos del Sistema Académico. El Datawarehouse extrae información de esta base de datos fuente.



- **Sectores y Usuarios del Sistema**

Los sectores usuarios identificados son:

- Dirección de la Facultad
- Coordinación de Informática.
- Departamento de Matemáticas

4.3. Estudio de la Situación Actual

La situación actual es el estado en el que se encuentran los sistemas de información existentes en el momento en el que se inicia su estudio.

El objetivo de este estudio es realizar un breve resumen de los sistemas de información que dispone la Fes Cuautitlán Sólo se realiza un breve relevamiento para identificar qué información puede ser relevante y necesaria para la construcción del SUFESC.

Como resultado de esta actividad se genera un diagnóstico, estimando la eficiencia de los sistemas de información existentes e identificando las posibles fuentes de información que alimenta al nuevo sistema.

- **Descripción de los Sistemas de Información Existentes**

La FES CUAUTITLAN cuenta básicamente con dos sistemas, esta información es la plataforma por la cual se permitirá diseñar y construir un Sistema de Apoyo Gerencial.

Los tres sistemas más importantes con que cuenta la FES CUAUTITLAN son:

Sistema Administrativo Contable, en donde se encuentra información relacionada con:

- Contabilidad
- Sueldos
- Tesorería
- Facturación etc.

Sistema Académico, en donde se encuentra información relacionada con:

- Informes

- Alumnos
- Profesores, Cátedras y Comisiones
- Planes de Estudio y materias
- Exámenes
- Encuestas
- Aulas/Laboratorios
- Becas
- Pasantías/Convenios
- Graduados etc.

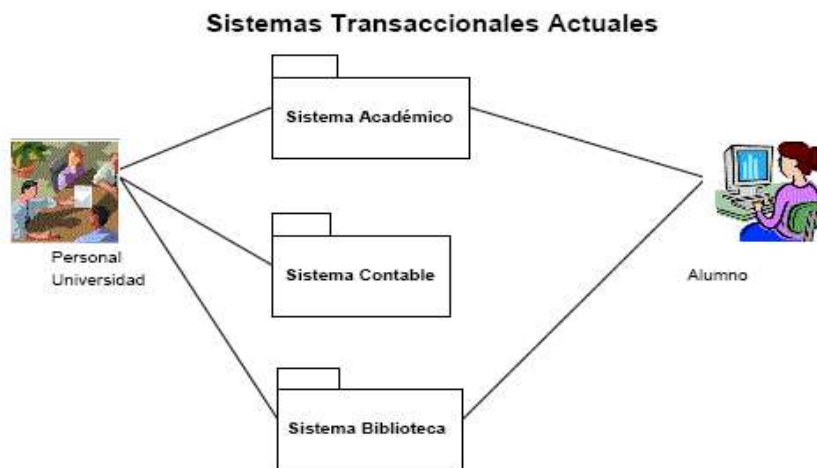
Sistema Biblioteca, en donde se encuentra información relacionada con:

- Catálogo de Libros
- Administración de Libros. ETC

A continuación se describe muy brevemente el Sistema Académico, el cual es la base de información para la primera etapa de construcción del SUFESC.

Sistema Académico

Se describen brevemente los principales módulos del sistema. La descripción y enumeración de estos es a efectos de disponer de un marco general con la funcionalidad del sistema actual.



La figura 4-3 muestra los sistemas actuales de la universidad.

Módulo de Alta de Aspirantes

Se utiliza para el ingreso inicial de datos de los aspirantes a grado, posgrado, etc.

Módulo de Datos personales

Permite mantener actualizada la información personal de los alumnos tal como domicilios, etc.

Módulo de Notas de Parciales y Finales

Módulo de actualización y consulta de notas de exámenes parciales y finales.

Módulo de Migración a Grado

Módulo que se utiliza al final de cada curso de ingreso, para crear promociones de grado.

Módulo de Impresión de Actas de Examen

Imprime las actas de examen de los alumnos de grado.

Módulo de Impresión de Diplomas

Permite tener control sobre los formatos de impresión de diplomas, fechas de

Módulo de Colegios y Universidades

Mantiene información sobre colegios secundarios, universidades y organismos de investigación con vinculación a la universidad.

Módulo de Profesores, Cátedras y Comisiones

Permite mantener información de los docentes de todos los niveles. También administra cátedras y comisiones formadas. Sirve como fuentes a los sistemas que manejan costos (sueldos).

Módulo de Planes de Estudio y Materias

Administra los planes de estudio de todos los niveles, las materias, sus correatividades y equivalencias.

Módulo de Fechas de Examen

Administra toda la información relacionada con las fechas de examen de cada materia de cada carrera.

Módulo de Sistema de Matriculación Automática

Administra las matriculaciones de cada cuatrimestre a los alumnos en las materias de su carrera en función de las materias que cursó y/o aprobó, y el plan de estudios.

Módulo de Sistema de Encuestas

Administra toda la información relacionada a las encuestas que periódicamente se realizan en las diferentes carreras de la universidad. Permite la carga, modificación y reportes.

Módulo de Asignación de Aulas y Laboratorios

Permite la administración de la información para facilitar la asignación de aulas y laboratorios de las diferentes carreras.

Módulo de Administración de Becas

Maneja información sobre becas y préstamos de honor.

Módulo de Administración de Pasantías

Permite conocer el resultado de las mismas, administra la información sobre empresas y alumnos pasantes.

Módulo de Datos de Graduados

Administra información relativa a los graduados de la institución.

Módulo de Migración a Sistema Contable

Módulo que alimenta periódicamente a los sistemas contables (Facturación y Sueldos) para cargar información sobre materias cursadas, cantidad de créditos, horas cátedra, centros de costos, etc.

Módulo de Biblioteca

Permite consultar disponibilidad de libros y publicaciones, inclusive vía Internet/Intranet además de manejar los préstamos de libros a alumnos y profesores.

Sectores usuarios del Sistema Académico

Entre los sectores usuarios del sistema Académico se encuentran:

- Ingreso

- Grado
- Posgrado
- Contaduría/Tesorería
- Directores de Departamento
- Rectoría
- Consejo Técnico Académico
- Biblioteca
- Alumnos de Grado
- Alumnos de PosGrado .

4.4. Diagnóstico de la Situación Actual

Luego del levantamiento de los sistemas actuales podemos enumerar algunas características detectadas, que hacen que dichos sistemas no faciliten la obtención de información específica para la toma de decisiones.

Hay varias particularidades que se encuentran en los sistemas transaccionales de la universidad (y en la mayoría de los sistemas orientados a transacciones) que hacen dificultosa la entrega de información de tipo gerencial, pero principalmente se puede nombrar:

- Inconsistencia de los datos
- Productividad
- Incapacidad para transformar datos en información.

A continuación se describen estas tres particularidades.

Inconsistencia de los Datos

Los diferentes departamentos tienen la posibilidad de preparar informes a partir de los datos que generan los sistemas transaccionales. ¿Qué ocurre cuando a la dirección de la universidad le llegan dos informes diciendo en uno de ellos que la actividad se redujo un 15 %, y en el otro que la actividad creció un 10 %? Evidentemente la dirección no sabe que pensar. Este es un ejemplo de la crisis de inconsistencia de información. Los motivos de estas inconsistencias son:

- No hay una base común de tiempo para los datos
- Algoritmos diferentes para tratar los datos

- Problema de los datos externos
- No hay una fuente común de los datos.

No hay una base común de tiempo para los datos: ¿hay alguna razón para creer que un análisis hecho en un determinado momento, sea el mismo en otro momento? Seguramente que no.

Los datos siempre están cambiando dentro de los sistemas transaccionales, y es muy importante poder temporizar la información de control de gestión, y así lograr la consistencia de los diferentes informes que circulan por la organización. Dicha temporalidad es imposible obtenerla dentro del ambiente transaccional. En otras palabras, se necesita un ambiente que congele los datos por un determinado tiempo, que no es el caso de los ambientes transaccionales que están cambiando continuamente.

Algoritmos diferentes para tratar los datos: todos los departamentos tendrán sin lugar a dudas diferentes algoritmos para analizar la información de negocio, pero la técnica para obtener los datos y convertirlos en información debería ser común para todos, y así asegurar la consistencia de los análisis inter-departamentales.

Problema de los datos externos: con la tecnología existente hoy en día es muy fácil obtener y poner a disposición datos cuyo origen no son solo sistemas transaccionales de la organización, sino fuentes externas de la misma. De esta forma un analista de información podría introducir datos sobre los competidores de la organización desde un periódico, y otro analista introducir información similar desde una publicación especializada. De manera que no solo no tenemos las fuentes externas unificadas, sino que además no hay coordinación entre la gente que ingresa la información externa y ninguno sabe lo que hace el otro.

No hay fuente común de los datos: si la fuente de información de control de gestión no es común a todos los departamentos entonces se corre el riesgo de perder la consistencia y sincronización en un sistema de información compartida.

Productividad

La productividad es un punto muy importante a tener en cuenta, especialmente si se poseen datos con una gran historicidad, como es el caso de la universidad. Cuando surge la necesidad de producir un informe utilizando muchos de los archivos y de los datos que la Universidad ha ido acumulado a través de los años, entonces el área de sistemas debe:

- Localizar y analizar los datos e incluir en el informe
- Consolidar los datos para el informe
- Destinar recursos de programación y análisis para la tarea.

En suma, aun cuando el programa para la generación del informe deba ser escrito y probado, y esto parezca fácil de hacer, la obtención de los datos desde el ambiente transaccional es un tarea sumamente complicada y tediosa de realizar. Además esta tarea consume tiempo, recurso que posee un costo de oportunidad muy alto si consideramos la disponibilidad tecnológica de hoy en día.

Muy probablemente los requerimientos de informes a la dirección se sucedan uno tras otros (y ninguno igual o parecido al anterior), lo cual pondrá en un verdadero aprieto al área de sistemas y repercutirá en las áreas de decisión cuando éstas no cuenten con la información solicitada en tiempo y forma.

Incapacidad para transformar datos en información

Los sistemas transaccionales no fueron construidos pensando en la integración de sus datos para la toma de decisión, sino que fueron realizados pensando en satisfacer necesidades netamente operativas de la organización. Pero la integración de los datos no es la única dificultad que el área de sistemas encontrará, el segundo obstáculo será la historicidad de los datos en línea no alcanza para satisfacer los requerimientos de información del usuario final.

Por todo lo dicho, las estructuras de datos que manejan los sistemas de datos transaccionales no son todo lo eficiente que necesitan los usuarios para nutrirse de la información necesaria para apoyar sus procesos de toma de decisión.

4.5 Definición de Requisitos del Sistema

Esta actividad incluye la determinación de los requisitos, mediante una serie de sesiones de trabajo con los usuarios participantes. Una vez finalizadas, se analizó la información obtenida definiendo los requisitos y sus prioridades, que se añaden al catálogo de requisitos que sirven además para el estudio y valoración de las distintas alternativas de solución que se propongan.

- **Catalogación de Requisitos**

Primeramente en este apartado se detallan los requisitos planteados por los usuarios como necesidades de información para su departamento, posteriormente se exponen los requisitos tecnológicos necesarios para dar soporte al sistema de información y finalmente se exponen los recursos humanos necesarios para llevar adelante el proyecto.

- **Requisitos Funcionales**

Los usuarios del Departamento de Ingeniería han formulado los siguientes requerimientos o necesidades de información. Estos requerimientos se han agrupados en diferentes módulos lógicos que son: Docentes, Alumnos, Materias, Encuestas, Relaciones Institucionales y Otras Necesidades de Información.

A) **DOCENTES**

1. Cantidad de Docentes por cargo y carga académica carga laboral (por materia y totales del departamento).

2. Cantidad de Docentes por categorías y por carga académica semanal según el siguiente rango horario:

- a) < 9hs
- b) 10 a 19 hs
- c) 20 a 29 hs
- d) 30 a 39hs
- e) y > 40 hs.

3. Cantidad de Docentes según título máximo y carga académica dedicación semanal (por materia y totales del departamento).

4. Variaciones de la composición del cuerpo docente (según cargo y carga académica).

5. Cantidad de Docentes según:

- a) Antigüedad
- b) Edad
- c) Remuneraciones
- d) Ausentismo.

6. Evolución de la estructura de personal por sectores y cátedras (incluir totales y subtotales).

B) ALUMNOS

7. Alumnos Postulantes/ Matriculados: (Evolución de varios años),

- a) N° de vacantes
- b) Cantidad de Postulantes
- c) Cantidad de Matriculados
- d) Cantidad de Egresados
- e) Tasa de Egreso
- f) Evolución ingresos y egresos por carrera.

8. Cantidad de Alumnos:

- a) Alumnos inscriptos por materia
- b) Alumnos que aprobaron la materia cursada

- c) Alumnos que reprobaron la materia cursada
 - d) Análisis de notas, según:
 - f) Tiempo para recibirse por carrera
 - g) En biblioteca (ingreso de libros, libros prestados, suscripciones)
 - h) Cantidad de Becados recibidos y enviados a otras instituciones i)
Cantidad de Alumnos Matriculados de otras universidades.
9. Calificaciones promedios por asignatura (por año).
10. Duración real de la carrera (número de graduados por año de ingreso) por rango de tiempo a :

- a) 5 años
- b) 5 a 6 años
- c) 6 a 7 años
- d) 7 a 8 años
- e) 8 a 9 años
- f) 10 o más años.

11. Información de las siguientes actividades de los graduados (por año):

- a) Cantidad de Incorporados a actividades profesionales específicas
- b) Cantidad de Incorporados a actividades profesionales no específicas
- c) Cantidad de Incorporados a actividades académicas
- d) Cantidad de Incorporados a actividades gerenciales
- e) Cantidad de Desocupados
- f) Otras actividades.

C) MATERIAS

12. Cantidad de Alumnos por materia semestral.

13. Cantidad total Alumnos Matriculados: por materia, cantidad total, con nota.

14. Materia Cursada: (por semestre).

- a) Aprobada total
- b) Reprobada

D) ENCUESTAS A LOS ALUMNOS

15. Por materia y docente: condiciones didácticas.

a) Rendimiento de la materia:

- Clases teóricas
- Trabajos de aplicación
- Organización
- Bibliografía empleada
- Opinión general.

b) Calificación General:

- Rendimiento docente
- Rendimiento de la cátedra
- Opinión general
- Rendimiento general.

F) OTRAS NECESIDADES DE INFORMACION

16.-¿Qué características tienen los Alumnos, según desempeño académico?

17.- ¿Qué características tienen los mejores Profesores, según encuestas?

• **Requisitos Tecnológicos**

Dentro de los requisitos tecnológicos se debe disponer del siguiente hardware y software, tanto en equipos servidores como estaciones clientes:

Hardware de Servidor

Computadora/Procesador: Servidor en torre PowerEdge T710

Memoria: Hasta 144 GB¹ (18 ranuras DIMM): memoria DDR3 de 1333 MHz

Disco duro: SATA de 3,5" (7.200 RPM) 1 TB

Tarjetas de interfaz de red . NIC Ethernet Gigabit de doble puerto

Broadcom® NetXtreme II™ 5709c integrada con sistema de fallas y equilibrio de carga. NIC adicionales y opcionales de 1 GBe y 10 GBe

Hardware (Conectividad) de Red: para accesos a las bases de datos fuentes y para distribución y servicios a las estaciones clientes del SUFESC.

Hardware de Estaciones Cliente

Computadora/Procesador: Desktops Dell OptiPlex Procesador de doble núcleo Intel Pentium

Memoria: Hasta 2 ranuras DIMM; SDRAM DDR2 de 800 MHz de dos canales sin ECC, admite de 1 GB a 4 GB

Disco duro: SATA II de 7200 RPM de hasta 160 GB, 3,0 GB/s; segundo disco duro disponible en mini tower

Hardware (Conectividad) de Red: para acceso a los datos en el Datawarehouse.

Tarjetas de red LAN Gigabit Broadcom® (BCM5784M) integrada

NIC Ethernet 10/100/1000* opcional con compatibilidad con PXE, reactivación remota

Software de Servidor:

Sistema Operativo Servidor: Microsoft® Windows Server® 2008 SP2, Datacenter Edition, (x64 con Hyper-V™)

RDBMS: para soporte al Datawarehouse. MS SQL Server 7.0

Herramientas de Inteligencia de Negocios: Servidor OLAP (OLAP Services de MS SQL Server 7.0).

Software de Estación Cliente:

Sistema Operativo: sistema operativo Microsoft Windows NT Workstation versión 4.0 con los Service Packs o Microsoft Windows XP o superior con los Service Packs correspondientes, Microsoft Internet Explorer 6.0 o superior

Herramienta de Cliente de Inteligencia de Negocios: Microsoft Excel 2007

Herramienta para el Control de Configuración: Microsoft Visual Source Safe 6.0 o superior. Esta herramienta es utilizada solo para desarrollo del sistema

Herramienta para el Modelado de Datos: Erwin 3.5 o superior. Esta herramienta es utilizada solo para desarrollo del sistema

Herramienta CASE: Rational Rose 2000 o superior. Esta herramienta es utilizada solo para desarrollo del sistema.

- **Recursos Humanos**

En los requisitos de personal para la construcción del sistema se destaca la necesidad de dos tipos de recursos de la universidad:

- **Recursos humanos para construir el sistema desde la perspectiva del usuario:**

Requiere la participación activa de quienes usarán el Datawarehouse. La gente de negocios debe participar activamente durante el desarrollo del Datawarehouse, desde una perspectiva de construcción y creación. En este caso deben ser los responsables del Dpto. de Ingeniería y Rectoría.

- **Recursos humanos para el soporte técnico para construir y mantener el sistema:** el Datawarehouse introduce muchas tecnologías nuevas (ETT, Carga, Acceso de Datos, Catálogo de Metadatos, Implementación de DSS/EIS) que se deberán aprender para realizar el soporte diario una vez implementado el SUFESC y fundamentalmente en la etapa inicial (debido al conocimiento del sistema transaccional) y para la prueba de usuario e implementación final en el ambiente de producción.

4.4. Estudio de Alternativas de Solución

Este estudio se centra en proponer diversas alternativas que respondan satisfactoriamente a los requisitos y objetivos planteados, considerando también los resultados obtenidos en el Estudio de la Situación Actual.

Una vez definidos los requisitos a cubrir por el sistema, se estudian las diferentes opciones que hay para configurar la solución.

En la descripción de las distintas alternativas de solución propuestas, se hace énfasis fundamentalmente en productos existentes en el mercado. Estos productos deben tener características que permitan implementar la solución propuesta o parte de ella. Los productos a considerar son aquellos que son

utilizados para brindar soluciones de Inteligencia de Negocios, entre estos productos o herramientas están las de ayuda a la construcción de Datawarehouses, motores de bases de datos, herramientas OLAP, herramientas de consultas avanzadas y visualización de información y finalmente las que aportan facilidades de minería de datos.

El objetivo del estudio de las alternativas de solución o búsquedas de herramientas que den soporte al SUFESC es realizar una breve introducción a productos software para su conocimiento básico, criterios para su comparación (ventajas y desventajas) y selección adecuada. Para saber más detalles de cada producto se deberá consultar bibliografía específica o los sitios en Internet de los productos.

- **Preselección de Alternativas de Solución**

A continuación se enumeran algunas de las herramientas y soluciones del mercado que aportan soporte tecnológico de Inteligencia de Negocios y son utilizadas por diferentes empresas y organizaciones obteniendo resultados satisfactorios a sus necesidades.

La información ha sido extraída de los sitios de Internet [Csi, 2002], [Oracle, 2002], [Microsoft, 2002], [Cognos, 2002] y [Synera, 2002]. En estos sitios se encuentra mayor detalle de los productos mencionados.

- **Construcción de Data Warehouse**

- Visual Warehouse (IBM)
- Oracle Express (Oracle Corp.)
- MS SQL SERVER (Microsoft Corp).

- **Base de Datos**

- DB2 UDB (IBM Corp.)
- Oracle (Oracle Corp.)
- Sybase SQL SERVER (Sybase Corp.)
- MS SQL SERVER (Microsoft Corp)
- Informix.

Herramientas OLAP

- DB2 OLAP Server (IBM Corp.)
- Oracle Express (Oracle Corp.)
- Power Play (Cognos Corp.)
- Business Objects (Business Objects Corp.)
- Informix-MetaCube
- MS SQL SERVER (Microsoft Corp.)
- Synera (Synera Systems Corp.)
- MicroStrategy (MicroStrategy Corp.).

Consultas y Reportes

- Lotus Approach (IBM Corp.)
- Discoverer (Oracle Corp.)
- Impromptu (Cognos Corp.)
- Business Query (Business Objects Corp.)
- Synera (Synera Systems Corp.)
- MicroStrategy (MicroStrategy Corp.)
- Excel 2000/Xp (Microsoft Corp.).

Minería de Datos

- Intelligent Miner (IBM Corp.)
- 4 thought (Cognos Corp.)
- Business Miner (Business Objects Corp.)
- MS SQL SERVER (Microsoft Corp.)
- Synera (Synera Systems Corp.)
- MicroStrategy (MicroStrategy Corp.).

Debido a las posibilidades económicas, de recursos humanos y fundamentalmente de acuerdos con empresas proveedoras como así también la plataforma tecnológica que dispone el Centro de Cómputos de la Universidad, se realiza una preselección de herramientas en donde se estudian las siguientes herramientas que son factibles de ser utilizadas para la construcción del SUFESC.

- Synera (Synera Systems Corp.)

- Suite Cognos - Impromptu, Power Play , 4 thought (Cognos Corp.)
- MS SQL SERVER (Microsoft Corp.)
- Oracle (Oracle Corp.)
- Excel 2007 (Microsoft Corp.).

- **Descripción de las Alternativas de Solución**

A continuación se realiza una breve descripción de los productos preseleccionados y que tienen mayor probabilidad de ser utilizados para dar respuesta al sistema a desarrollar. La descripción de cada herramienta está organizada en varias secciones:

- **Productos**, que enuncia los productos del proveedor que satisfacen la necesidad del proyecto
- **Características Técnicas**, que describe un resumen de las herramientas y en algunos casos se amplía con algún detalle de importancia
- **Ventajas**, en donde se enumeran algunas ventajas y características del producto que lo diferencian de otros y finalmente
- **Sobre la Empresa**, que describe fundamentalmente si el producto o herramienta tiene soporte en el país de la empresa proveedora.

Además de la descripción presentada en este documento se realizaron pruebas y demostraciones de algunas de dichas herramientas por los proveedores del mercado que las representan y dan soporte.

- **Synera SuiteTools Productos**

La compañía ofrece al mercado el producto *Synera Intelligent Exploration Suite* que brinda soluciones en Inteligencia de Negocios.

Esta suite está compuesto por varios productos donde se destaca *Synera Engine* que es un motor analítico que a diferencia de otros entornos analíticos, basados en bases de datos relacionales, permite almacenar datos sin necesidad de definir un

modelo de datos predeterminado, y se adapta fácilmente a nuevas necesidades. También permite descubrir patrones ocultos en la información.

También dentro de la suite, el otro producto destacado es *Synera Explorer*, que es una plataforma de exploración de datos desarrollada sobre *Synera Engine*, que permite un acceso fácil e inmediato a las tendencias, asociaciones y relaciones que subyacen bajo sus datos, mejorando el proceso de toma de decisiones.

Características Técnicas

El *Intelligent Exploration Suite* está compuesto de cinco componentes según muestra la figura 4-4:

- *Synera Configurator* maneja la parametrización del sistema

Synera Loader rápidamente importa grandes volúmenes de datos desde cualquier plataforma y fuente de datos

- *Synera Explorer* navegación de datos, consultas avanzadas y análisis
- *Synera Discovery* Facilidades de Data Mining. Revela patrones y reglas en los datos
- *Synera Users* maneja el acceso y seguridad
- *Synera Engine* soporta el sistema y está basado en *Synera IVM* (integrated value model).

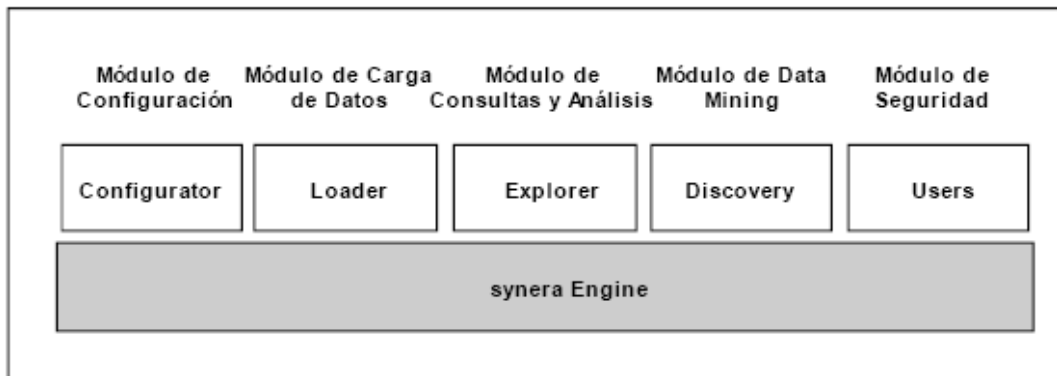


Figura 4-4. Componentes de Synera Suite.

Synera se puede configurar en las siguientes versiones:

- *Personal Synera: versión monousuario y sistema operativo Windows 95/98 para volúmenes de información no superiores a los 900 Mb.*
- *Synera (Cliente/Servidor): sistema operativo Servidor Windows NT y clientes Windows 95/98*
- *Synera (Cliente/Servidor): para UNIX con plataformas superiores a 250 Gb.*

A continuación se describe brevemente las capacidades de Synera para descubrimiento del conocimiento (Data Mining) que es una de las características más importantes del producto.

Sistema Synera de Descubrimiento de Información

Se basa en la capacidad de Synera de realizar tareas analíticas sin una dirección específica, e informar los resultados más significativos para que una persona pueda tomar una decisión apropiada y actuar en consecuencia. El sistema Synera de descubrimiento del conocimiento incluye:

- la base de conocimiento y el motor analítico.

La figura 4-5 muestra las relaciones entre la base de conocimiento, el motor analítico y los sistemas fuentes.

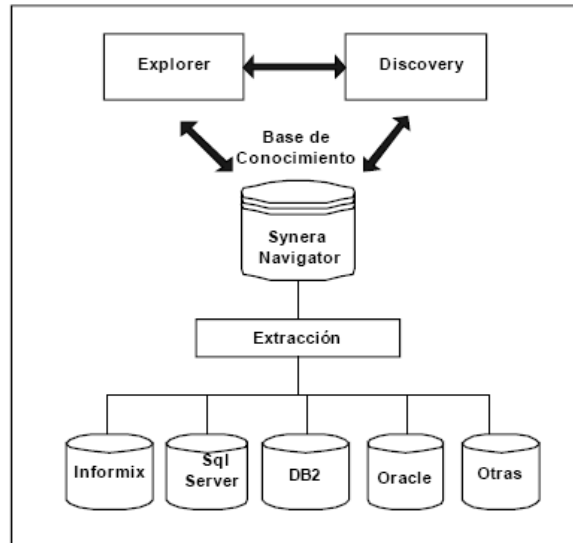


Figura 4-5. Descubrimiento de conocimiento.

La base de conocimiento

Mientras que Synera es ante todo un depósito de conocimiento, la base de conocimiento, el diccionario, y el sistema de consultas, suministran un conjunto significativo de detección de patrones y elementos, para el descubrimiento del conocimiento. El diccionario incluye información sobre los datos almacenados en la base de conocimiento, incluyendo rangos de valores, valores promedios y desviación estándar. La distribución de valores, puede mostrarse en forma gráfica y las excepciones se detectan fácilmente. Toda esta información está disponible sin necesidad de preparar consultas o realizar programación.

La capacidad irrestricta *ad-hoc* de consultas suministra funciones de descubrimiento de conocimiento adicionales extensivas. Esto incluye selección, agrupamiento, adición y muestras gráficas de los resultados. Los valores pueden asignarse a las categorías y ellas pueden entonces usarse para análisis. Las relaciones pueden explorarse sin tener conocimiento alguno de la estructura de los

datos. Las búsquedas pueden hacerse por la existencia de cualquier valor en cualquier contexto. Los datos detallados están siempre disponibles, incluso cuando se trabaja con datos agregados.

El motor analítico Synera

El motor analítico es un conjunto de herramientas integradas, que puede analizar el contenido total de la base de conocimiento, y suministrar al usuario el conocimiento encontrado en ella. Una ventaja de este sistema es que es automático. No hay necesidad de elegir y aislar conjuntos de entrenamiento y prueba, y procesos no iterativos que tardan días o incluso semanas en completarse. Incluso no es necesaria una hipótesis inicial.

Una gran parte del trabajo que normalmente se requiere para completar la comprensión de inmensas cantidades de datos es la aplicación repetitiva de algoritmos analíticos variados, a través de subconjuntos cuidadosamente elegidos entre todos los datos disponibles. Dada la cantidad de tiempo necesaria para realizar esta clase de análisis, se ha considerado poco practicable en el ambiente de negocios del mundo real.

Uno de los conceptos de diseño claves del motor analítico, es hacer esto posible y práctico, realizando este trabajo en forma eficiente y sin intervención o dirección humana.

El motor analítico usa sólo un conjunto de parámetros que limitan el campo de acción total de la búsqueda de patrón, lo que permite al usuario eliminar los patrones obvios, ahora claramente entendidos (por ejemplo, 50 % de la gente casada es mujer) y los que son tan sutiles que carecen de interés. Los algoritmos adicionales se añaden al sistema en bases en curso, para permitir que Synera de al usuario, una comprensión del sistema más profunda.

Ventajas

Algunas características y ventajas de *Intelligent Exploration Suite* son:

1. Rápida captura de datos desde las bases operacionales
2. Permite a los usuarios analizar datos y obtener resultados inmediatos luego de la carga de datos
3. Permite fácil acceso y navegación a través de todo el modelo de datos
4. No requiere experiencia técnica, permitiendo la participación de los

usuarios con un bajo uso del departamento de TI.

5. Construye y mantiene tablas de índices para soportar necesidades imprevistas de análisis.
6. Provee un conjunto de reportes "ad hoc" y capacidades de consultas en tiempo real.
7. Permite inclusión instantánea de datos adicionales sin la necesidad de reestructurar o rediseñar la base de datos.
8. Permite a los usuarios cambiar los requerimientos y definiciones en cualquier momento.
9. Muy bajo costo de operación.
10. Permite a los usuarios focalizarse en la información más que en el mantenimiento y tareas de optimización.

Sobre Synera Systems

Synera Systems es una empresa global de tecnología fundada en 1998, con organizaciones comerciales en Europa, Estados Unidos y América Latina. Barcelona (España) y Minneapolis (USA) son los dos centros corporativos en los que se concentran las actividades de marketing y desarrollo de software. Synera opera en sectores de telecomunicaciones, banca/seguros, gran consumo, retail y comercio electrónico, en los que aporta soluciones analíticas y de gestión del conocimiento. La dirección de su página web es www.Synerasystems.com. La compañía C&A Consulting ofrece el soporte para esta herramienta.

- **Cognos Enterprise Suite Tools**

Productos

Cognos ha desarrollado herramientas tecnológicas para proveer toda la información requerida para la toma de decisiones. Con los productos de Cognos se construyen los cubos multidimensionales, se generan los reportes Ad-hoc, se realiza la exploración y minería de los datos, se construye el Balanced Scorecard, se preparan las proyecciones y simulaciones financieras, se produce la información operativa del día a día, se preparan los informes gráficos unidimensionales o tridimensionales y se obtienen los mapas y semáforos para ilustración de la situación de los indicadores de desempeño. Con la solución End-to-End

de Cognos se obtienen resultados desde la construcción del primer Datamart de la organización hasta el total Data WareHouse Corporativo, comenzando desde los más elementales procesos de consulta, pasando por los procesos de extracción y transformación de datos, y llegando a la publicación y análisis de información en la Internet o la Extranet de la organización.

CognosSuite ofrece un conjunto de herramientas que se dividen en Herramientas de Análisis y Herramientas de Reportes. Juntas estas herramientas ofrecen una solución para resolver problemas de Inteligencia de Negocios. Las herramientas son:

- Cognos PowerPlay
- PowerPlay Server Web Edition (Cognos Web Server)
- Cognos Scenario
- Cognos 4Thought
- Cognos Impromptu
- Impromptu Web Query
- NovaView.

Características Técnicas

A continuación una breve descripción de estas herramientas, las cuáles están divididas en herramientas de Análisis y herramientas de Reportes.

Herramientas de Análisis

Ofrece a los tomadores de decisiones capacidades de análisis que necesitan para investigar los hechos y tendencias del negocio.

También ofrece soporte de análisis de la información de las actividades del día a día de los clientes.

- *Cognos PowerPlay*

Permite a los ejecutivos y gerentes explorar los datos corporativos desde cualquier ángulo, en cualquier combinación, identificando tendencias que no son fácilmente encontradas usando cualquier otro método de análisis.

- *PowerPlay Server Web Edition*

Provee a los usuarios la capacidad de utilizar por medio de sus web browsers otros productos de Cognos como el Web Query.

- *Cognos Scenario*

Ofrece a los usuarios analistas la habilidad de identificar y ranquear los factores que tienen un impacto significativo en las medidas claves del negocio. Automáticamente descubre los patrones y relaciones que existen en los datos del negocio.

- *Cognos 4Thought*

Ofrece a los planificadores del negocio el poder de realizar proyecciones, preguntas del tipo "Que pasa si..." y mediciones de efectividad que pueden ser usadas para determinar y definir políticas y planes que conduzcan a la empresa hacia mayores beneficios.

Herramientas de Reportes

Ofrece a los autores de reportes la habilidad de crear y distribuir el sistema de reportes que soporte los procesos de negocios de la organización. También provee información de clientes con fácil acceso a los detalles de datos por medio de reportes estándares.

- *Cognos Impromptu*

Permite a los autores de reportes extraer datos desde cualquier lugar de la organización creando reportes estándares o predefinidos.

Administración distribuida y creación de reportes combinadas con un control central hacen de Impromptu la solución para reporting (creación y distribución de información), tanto para ambientes cliente/servidor como Web. Impromptu elimina la complejidad de trabajar con la capa de base de datos, cuya estructura es difícil de entender para usuarios no técnicos, permitiendo trabajar con una representación de datos que está de acuerdo al contexto del negocio.

- *Impromptu Web Query*

Permite a los usuarios "navegar" a través de los datos corporativos usando un web browser.

La navegación con hiperlinks lleva a las personas a encontrar información de interés en cualquier momento. Impromptu Web Query también ofrece capacidades de consultas desde el web sitio empresarial.

- *Nova View*

Toma como fuente la tecnología OLAP de Microsoft. Actúa como software cliente, permitiendo facilidades como:

- Identificar tendencias y comparar resultados
- Vista de varias dimensiones del negocio
- Interactúa con información en formato gráfico, fácil de interpretar
- Facilidades de *Drill Down* para niveles de detalle, cambio de filas y columnas y cambio de formatos de gráficos que reflejan la mejor vista del negocio.

4.5. Requerimientos de HARWARE y SOFTWARE:

Sistemas Operativos: UNIX: IBM AIX, HP-UX, Sun Solaris, Compaq Unix; Windows NT, Win95, Win98, Win2000, Manejadores de Bases de Datos Soportados: Oracle, Informix, DB2, SQL Server, etc., acceso vía ODBC o conectividad directa según el motor de base de datos.

Estaciones (características mínimas): Intel Pentium o compatible, 32 Mb RAM, espacio de disco 30 Mb, Win95, Win98, WinNT, Win2000, Internet browsers (Microsoft Explorer, Netscape, etc.)

Ventajas

Las características y ventajas más importantes son:

1) Soportado para cualquier tamaño de empresa u organización.

Cognos aporta mecanismos para una Administración más sencilla y eficaz del entorno, tanto desde el punto de vista de la gestión, incorporando facilidades para la configuración de los diferentes componentes, o la posibilidad de distribuir o migrar la solución pasando de un entorno de desarrollo a otro de pruebas y de ahí a producción.

El producto tiene una arquitectura abierta que permite integrarse con el entorno tecnológico de cualquier empresa.

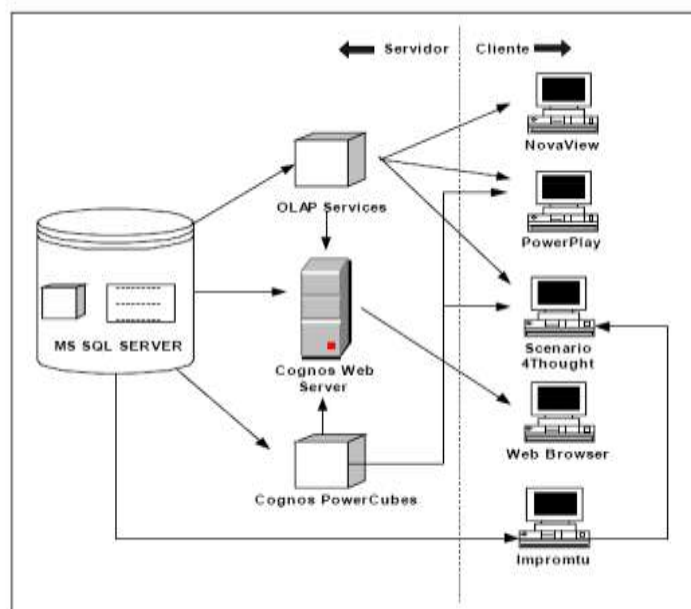


Figura 4-6.
Herramientas y arquitectura de Cognos.

Cognos puede trabajar tanto con un solo usuario con una sola base de datos como miles con un Datawarehouse. Soporta implementaciones cliente/servidor y *stand-alone*.

Los usuarios pueden acceder a los datos operacionales de cualquier fuente de datos de la empresa. Impromptu provee acceso nativo a las más populares bases de datos. Adicionalmente a su propia estructura multidimensional PowerPlay ofrece acceso nativo a Arbor Software's Essbase, Oracle's Express OLAP, Microsoft OLAP Services, y IBM's DB2 OLAP servers.

2) Entorno de usuario integrado, completo y sencillo.

El producto dispone de tecnología web. A través de un único portal web, suministrado por Cognos, se ofrece a los usuarios del sistema diferentes opciones dependiendo de sus necesidades y requerimientos funcionales, como:

- Consultar informes predefinidos
- Realizar consultas *ad-hoc*
- Analizar la información utilizando informes de gestión: se traslada la visión del negocio a un modelo multidimensional determinado por diferentes

dimensiones o perspectivas (como tipo de producto, regiones, tiempo, tipo de clientes, etc.) en el que los usuarios pueden cambiar fácil y dinámicamente el ámbito del análisis, profundizando en la información, intercambiando dimensiones (para analizar en base a otros conceptos del negocio), etc.

- Medición, Visualización y Gestión, a través de cuadros de mando de los indicadores claves de gestión del negocio

3) Mecanismos completos para el suministro de información.

Tanto a través del portal web en el que la información se organiza y se pone a disposición de los usuarios para su análisis y consulta. Cognos permite la detección y monitorización de diferentes tipos de eventos y su notificación vía e-mail de dichos eventos o alarmas con la posibilidad de envío de información, ya sea en formato HTML, MS Word, MS Excel, o el que defina el usuario.

Sobre Cognos Corp.

Cognos es una compañía canadiense fabricante de software de Business Intelligence para el análisis y explotación de información.

Los productos soportan las bases de datos más populares y los más populares servidores OLAP, como así también integración con aplicaciones corporativas o empresariales, incluyendo Oracle Financials, SAP, PeopleSoft y BaanERP. Está disponible en siete lenguajes y los productos son distribuidos por más de mil empresas consultoras alrededor del mundo. Cuenta con más de 2,5 millones de licencias en miles de empresas y organizaciones de todo el mundo. La compañía tiene 19.000 clientes en 120 países de todo el mundo. Cuenta con 2.500 empleados y sus productos BI están disponibles al público a través de una red de distribuidores constituida por 3.000 socios. La sede central está ubicada en Ottawa (Canadá) y la dirección de su página web es www.cognos.com.

- **MS SQL Server 7.0**

Productos

La compañía Microsoft ofrece la herramienta SQL SERVER 7.0 como base de datos y OLAP SERVER para dar soporte a soluciones de Inteligencia de Negocios. Ambas herramientas vienen en la misma Suite denominada MS SQL SERVER.

Características Técnicas

Se destaca que el producto será utilizado como soporte al módulo Datawarehouse (base de datos) del SUFESC y también se utilizarán las facilidades para la Extracción, Transformación y Carga (ETL) de la información dentro del Datawarehouse que se podrá realizar con los Servicios de Transformación de Datos (DTS) que también provee el MS SQL SERVER 7.0.

Además, esta versión 7.0 cuenta con un servidor OLAP y otras facilidades denominadas Servicios OLAP, los cuáles se describen brevemente. Algunas características para destacar de SQL SERVER 7.0 son:

- *Herramientas y programas:* las herramientas y los programas se ejecutan mucho más rápido y están diseñados para que produzcan menos impacto en las operaciones del servidor.
- *Copia de seguridad activa:* las copias de seguridad activas proporcionan copias de seguridad en línea de alto rendimiento con un impacto mínimo en los sistemas en funcionamiento.
- *Servicios de ayuda a la toma de decisiones:* servicios de ayuda a la toma de decisiones (DSS) es un nuevo componente de proceso analítico en línea (OLAP) con múltiples características de MS SQL Server 7.0.
- *Almacenamiento simplificado en disco:* la arquitectura de almacenamiento en disco permite la escalabilidad desde bases de datos de equipos portátiles hasta bases de datos empresariales de tamaño de terabyte.

Optimizador de consultas con múltiples fases: el optimizador de consultas con múltiples fases busca el plan óptimo para consultas para mejorar el rendimiento de consultas complejas.

- *Estadísticas automáticas:* Estadísticas automáticas extrae estadísticas mediante el análisis rápido de una muestra, habilitando el Optimizador de consultas para utilizar la información más reciente e incrementar la eficacia de las consultas.
- *Ayudante de Web:* el Asistente de Web le facilita la publicación de datos en el Web.
- *Servicios OLAP integrados:* los servicios OLAP integrados a la base de datos permiten el análisis de información con vistas multidimensionales para ayudar a tomar de decisiones.
- *DBCC:* comprueba la *coherencia* física y lógica de la base de datos. El algoritmo patentado de un solo paso agiliza el rendimiento. Las nuevas características son compatibles y pueden resolver algunos problemas. La nueva arquitectura del motor de almacenamiento reduce la necesidad de DBCC, pero supone una buena práctica.
- *Desencadenadores:* la *flexibilidad* se ha mejorado con múltiples desencadenadores por tabla y con la recursión directa de desencadenadores.
- *Memoria dinámica:* mejora de rendimiento optimizando la ubicación y el uso de la memoria. Minimiza los conflictos con otros administradores de recursos.
- *Ordenación:* la velocidad de ordenación ha mejorado ostensiblemente, especialmente cuando a la base temporal de trabajo de MS SQL SERVER tempdb está en un conjunto de discos con bandas.
- *Servicios de transformación de datos:* DTS simplifica el proceso de importación y transformación de datos de orígenes múltiples y heterogéneos, tanto de manera interactiva como automática.
- *Estadísticas automáticas:* el optimizador de consultas administra la recopilación de estadísticas y garantiza una evaluación de planes eficaz.

- *E/S mayor y más eficaz:* los bloques de E/S son cuatro veces mayores, las páginas son de 8 Kb, las extensiones de 64 Kb y los recorridos usan bloques de 64 Kb. La E/S inteligente es la tecnología clave para mejorar el rendimiento con grandes tamaños de E/S. Las lecturas anticipadas más eficientes, los recorridos en el orden físico de las filas y la E/S paralela también mejoran el rendimiento.
- *Estrategias de utilización de índices:* habilita una mejora del rendimiento. Las nuevas estrategias incluyen el uso de varios índices en una única tabla o en varias tablas, índices de cobertura múltiple y combinados, creación de índices paralelos en la misma tabla y mantenimiento automático de estadísticas de manera predeterminada.
- *Combinaciones:* las *estrategias* de combinación *hash* y mezcla mejoran el rendimiento de ciertos tipos de recuperación de datos, además de las combinaciones de bucles anidados. En una única consulta se pueden utilizar varios tipos de combinaciones. El procesador de consultas reconoce ciertos tipos de combinaciones comunes.
- *Consultas paralelas:* el *rendimiento* mejora si se utiliza la ejecución en paralelo de consultas de una única consulta a través de múltiples procesadores. Los pasos en una única consulta se ejecutan en paralelo.
- *Procesador de consultas:* el procesador de consultas se ha diseñado para admitir mejor las bases de datos grandes y las consultas complejas encontradas en la ayuda a la toma de decisiones, en el almacenamiento de datos y en las aplicaciones OLAP.
- *Bloqueo dinámico de filas:* el bloqueo completo de filas se ha implementado tanto para filas de datos como para entradas de índice. El bloqueo dinámico elige automáticamente el nivel de bloqueo óptimo (fila, página, página múltiple y tabla) para todas las operaciones de base de datos.
- *Administración dinámica del espacio:* las bases de datos pueden crecer o reducirse automáticamente dentro de límites configurables, minimizando la necesidad de la intervención del administrador de bases de datos. Ya no es necesario asignar espacio y administrar las estructuras de datos.

- *Administrador de registro:* el diseño simplificado mejora el rendimiento en truncamientos, en copias de seguridad en línea y en operaciones de recuperación.
- *Lectura anticipada:* la *lógica* de la lectura anticipada inteligente mejora el rendimiento y elimina la necesidad de la optimización manual.
- *Capacidad para utilizar cantidades mayores de memoria:* SQL Server 7.0 Enterprise admite direccionamientos de memoria superiores a 4 Gb, junto con Windows NT Server 5.0, sistemas que utilizan procesadores Alpha y otras combinaciones. Las estructuras y algoritmos simplificados de datos mejoran la concurrencia, escalabilidad y confiabilidad. Las comprobaciones en tiempo de ejecución de estructuras de datos críticos hacen que la base de datos sea más eficaz, minimizando la necesidad de comprobaciones de coherencia.
- *Almacenamiento escalable:* el formato del disco y el subsistema de almacenamiento proporcionan un almacenamiento que es escalable desde bases de datos muy pequeñas hasta bases de datos muy grandes.
- *Copia de seguridad y restauración:* las utilidades de restauración y copia de seguridad paralela escalan a las velocidades de dispositivo. Durante la copia de seguridad, totalmente sin desconexión, se mantiene un bajo impacto de los sistemas operacionales y un muy alto proceso de transacciones de servidor.
- *Carga de datos masiva:* la *velocidad* de importación y exportación de datos mejora considerablemente. Ahora utiliza OLE DB y funciona junto con el procesador de consultas para planear y optimizar las consultas.

A continuación se detalla una de las características principales de las facilidades de análisis de información para la toma de decisiones que ofrece el producto, denominadas Servicios OLAP.

Los servicios OLAP de SQL Server 7.0

Mediante el componente *MS OLAP Server* y el *MS PivotTable Service*, los servicios OLAP permiten descubrir información de las bases de datos, por ejemplo, pueden servir para respaldar las decisiones de optimización de los procesos en una empresa u organización. Con los servicios OLAP es posible que se necesite unos 20 minutos para crear un cubo de datos que analice las horas facturables de las personas del departamento, estableciendo una conexión con el sistema de horario y facturación de la compañía que se ejecuta bajo SQL Server 6.5.

Las herramientas OLAP sirven para crear, gestionar y consultar cubos de datos, y para implementar Datamarts. El OLAP Server de SQL Server 7.0 soporta MOLAP, ROLAP y HOLAP; niveles variables de agregación para optimizar el rendimiento de las consultas frente al espacio de almacenamiento; datos origen

en esquemas dimensionales o relacionales; particionado de cubos para posibilitar consultas contra orígenes de datos distribuidos y heterogéneos; análisis de utilización que permiten examinar las consultas con problemas y reconstruir las sumalizaciones para ajustar dichas consultas; posibilidad de 'deshacer' para el desarrollo de análisis de hipótesis; posibilidad de actualización incremental; y una interfaz de *OLE DB* (componente de conexión a bases de datos similar) ampliada para OLAP.

Las aplicaciones cliente pueden acceder a las funciones de OLAP Server a través del *PivotTable Service* con la especificación *OLE DB para OLAP* y el modelo de objetos *ActiveX Data Object Multi-Dimensional (ADO MD)*. Otro modelo de objetos, *Decision Support Objects (DSO)*, proporciona una interfaz de programación para gestionar el OLAP Server que es similar a la que proporciona SQL Server, *Distributed Management Objects (DMO)*.

Una base de datos OLAP Services consiste en cubos, cubos virtuales (que son uniones de cubos similares a una vista de SQL Server), orígenes de datos,

dimensiones y atribuciones (roles), que definen los privilegios de los usuarios en base a la seguridad Windows NT/2000. Después de crear el cubo y elegir la técnica de almacenamiento, OLAP Server diseña las agregaciones y procesa el cubo. Si se opta por la técnica de almacenamiento ROLAP, OLAP Server creará las tablas resumen en la base de datos origen después de procesar el cubo. De lo contrario, las agregaciones se almacenan en el formato nativo de OLAP Server. Puede escoger el grado de agregación considerando el nivel de optimización de consulta frente al espacio en disco necesario. Las agregaciones se acumulan, de modo que si elige menos rendimiento en favor de conservar espacio de disco, el motor de consultas de OLAP Server responderá a las consultas sumando las agregaciones existentes. Pueden crearse particiones en los cubos para acceder a datos procedentes de orígenes distribuidos y heterogéneos. Ésta es una característica de escalabilidad importante porque el OLAP Server puede determinar qué datos necesita de los diversos orígenes para satisfacer una consulta.

El *PivotTable Service* es el lado cliente de los OLAP Services de SQL Server. Proporciona el interfaz que permite a aplicaciones cliente conectarse con el OLAP Server de SQL Server. Puede utilizar OLE DB para OLAP o *ActiveX Data Objects* Multi-Dimensional (ADO MD) para escribir código de acceso y manipulación de cubos de datos. ADO MD permite desarrollar aplicaciones Web contra cubos OLAP Server.

El *PivotTable Service* puede almacenar en caché consultas de usuario, metadatos y datos, de modo que puede satisfacer consultas nuevas trabajando con datos residentes en el caché, en lugar de volver a realizar una consulta al OLAP Server. El OLAP Server y el *PivotTable Service* tienen mucho código en común. Este hecho proporciona al cliente características de cálculo multidimensional, almacenamiento en caché y gestión de consultas. Por ejemplo, si realiza una consulta al OLAP Server solicitando la información de ventas del primer trimestre de 1.999 y después decide comparar los resultados con los datos del primer trimestre de 1.998, el *PivotTable Service* sólo requerirá del servidor los datos de 1.998 porque los datos de 1.999 ya están en el caché. Además, el servicio *PivotTable* facilita la operación en

modalidad desconectada al permitir que las secciones o vistas del cubo permanezcan en la máquina cliente para su análisis posterior, sin requerir de una conexión con el OLAP Server.

El servicio *PivotTable* utiliza aproximadamente 2 MB de espacio en disco y 500 KB de memoria. Office 2000 incluye el servicio *PivotTable*, y la versión de Excel en Office 2000 incluye la facilidad de conectividad de incrustación entre hojas de cálculo y datos OLAP Server u otro proveedor de OLE DB para OLAP.

Algunas Ventajas

- *Facilidad de uso:* MS SQL Server facilita la tarea de los administradores de bases de datos en la construcción, manejo y despliegue de aplicaciones para negocios ya que automatizan la administración de las bases de datos mediante una interfaz gráfica fácil de usar. Además, cuenta con herramientas que simplifican el manejo de operaciones complejas como por ejemplo los servicios OLAP.
- *Autoadministración dinámica:* en SQL Server 7.0, una característica clave llamada Autoadministración Dinámica que automatiza muchas tareas de rutina. Los recursos de memoria y bloqueo se ajustan en forma dinámica; el tamaño de los archivos crece automáticamente, y las características de autorregulación garantizan un desempeño constante bajo condiciones de carga variables.
- *Desempeño, confiabilidad y escalabilidad:* las mejoras a SQL Server toman en cuenta el costo total de propiedad que representa para las organizaciones el desarrollo de aplicaciones, la capacitación en ellas y su administración. SQL Server 7.0 está diseñado para recibir mayor número de datos, transacciones y usuarios con facilidad. SQL Server 7.0 permite escalar las aplicaciones de bases de datos, conservando la estabilidad.

Integración: maximiza la inversión en la familia Windows, Microsoft Office y Microsoft BackOffice.

Microsoft Office 2000 y XP: grupo de productos con los que puede trabajar en forma integrada, entre los que se encuentran Windows NT Server y Windows 2000, Microsoft Office y BackOffice.

Sobre Microsoft Corp.

La compañía ofrece un conjunto de herramientas para desarrollar aplicaciones, administrar sistemas y redes, sistemas operativos, herramientas de escritorio y bases de datos entre otras, junto con servicios de consultoría, educación y soporte en casi todo el mundo

- **Oracle 9i Tools**

Productos

La compañía Oracle ofrece Oracle9i que es una suite de herramientas basadas en su base de datos, incluye prestaciones en las áreas de clustering, alta disponibilidad, Inteligencia de Negocios, seguridad y gestión y construcción de aplicaciones entre otras.

Las versiones que tienen las anteriores características son:

- Oracle9i Standard Edition
- Oracle9i Enterprise Edition
- Oracle 9i Personal Edition
- Oracle 9i Lite.

Características técnicas

A continuación se describe brevemente las herramientas que ofrece Oracle9i divididas en servicios de aplicaciones y servicios de desarrollo.

Servidores de Aplicaciones

Oracle9i Application Server (Oracle9iAS) cuenta con software para portal empresarial, caché de alta velocidad, Business Intelligence, rápido desarrollo de aplicaciones,

integración de aplicaciones y procesos de negocio, servicios Web, capacidades para servicios a móviles, etc., todo ello en un único servidor de aplicaciones.

Oracle ofrece un nuevo y ligero motor J2EE en Oracle9iAS, así como capacidades de cacheo ampliadas. Con soporte para *Edge Side Includes (ESI)*, una propuesta estándar para contenidos dinámicos.

Web caché Oracle9iAS permite a las aplicaciones Java dinámicas servir hasta 100 veces más usuarios que usando el mismo hardware sin la capacidad caché ESI. Esto implica que las empresas pueden reducir costos significativamente en la infraestructura de hardware necesaria para el funcionamiento de sitios Web dinámicos con alto volumen de datos. Oracle9iAS es el primer servidor que ofrece soporte para ESI.

Finalmente, Oracle9iAS permite a las compañías construir sitios y servicios web utilizando tecnologías de desarrollo de aplicaciones, incluido soporte para Java (J2EE) y XML.

4.6 .Herramientas de Desarrollo

Oracle9i Developer Suite (Oracle9iDS) complementa la infraestructura ofrecida por Oracle9i Application Server y Oracle9i Database. Oracle9i Developer Suite es una gama de herramientas que combina un entorno *Rapid Application Development (RAD)* para J2EE, XML y servicios Web en una solución completa y totalmente integrada. Oracle9iDS cuenta también con capacidades para el diseño de aplicaciones basado en modelos, trabajo en equipo y análisis de rendimiento para código entre otras características.

Para empresas interesadas en crear aplicaciones Java (J2EE) transaccionales, portales, soluciones wireless, servicios Web y proyectos de integración, Oracle9iDS está totalmente integrado con Oracle9i Application Server y con Oracle9i Database lo que reduce los costos de implementación y acelera el tiempo de terminación de los proyectos.

Las herramientas son:

- *Oracle Designer*: diseño de sus aplicaciones basándose en modelos, generación y reingeniería automática de esquemas de base de datos y aplicaciones.
- *Oracle Jdeveloper*: herramienta para desarrollo de aplicaciones Java y XML, siguiendo los últimos estándares del mercado. Desarrollo basado en componentes para arquitecturas multicapa. *Oracle Forms*: desarrollo de aplicaciones transaccionales de base de datos, tanto en entornos cliente/servidor como en Web. Además, permite la construcción de formularios multi-idioma.
- *Oracle Reports*: poderosa herramienta para la creación de reportes sofisticados y dinámicos publicados en la Web.
- *Oracle Discoverer*: conjunto de herramientas para consultas *ad-hoc* a la base de datos, le permite analizar y dar formato a los resultados de las consultas, preparar los resultados para la presentación y gestionar dicha información de manera que sea significativa para cualquier entorno o área del negocio.
- *Oracle Warehouse Builder*: herramienta para diseñar e implementar un Datawarehouse con facilidades gráficas. Accede y toma la información de interés esté donde esté, la transforma y almacena en el formato correcto para el análisis posterior.

Ventajas más destacables

De entre las nuevas prestaciones que ofrece Oracle9i, se pueden destacar las siguientes ventajas:

- Desaparece la barrera en la escalabilidad y se minimiza la capacidad informática no utilizada. Con Oracle9i, se pueden aumentar casi sin límites la capacidad de proceso utilizando clusters, sin necesidad de realizar cambios en sus aplicaciones utilizando Oracle Real Application Clusters.
- Minimización del tiempo de recuperación, ya sea por fallos de sistemas, errores humanos o de mantenimiento, utilizando los componentes del producto como *Oracle Real Application Clusters*, *Oracle Data Guard*, y

Oracle Resource Manager.

- Seguridad para la información, protección de la privacidad del usuario, con la certeza de que puede escalarse fácilmente a millones de usuarios utilizando los módulos *Oracle Advanced Security* y *Oracle Label Security*.
- Aumento de la productividad de las áreas de sistemas mediante prestaciones como auto-tuning, auto-corrección y auto-gestión del *Oracle Enterprise Manager*.
- Ofrece capacidades de soporte para aplicaciones Inteligencia de negocios con prestaciones preinstaladas de analítica avanzada y de datawarehousing.
- Gestión de contenidos y archivos de Internet fácil y eficientemente con el componente *Oracle Internet File System*.
- Rápido desarrollo de aplicaciones en el entorno integrado de Java y XML.

Sobre Oracle Corp.

La compañía ofrece bases de datos, servidores de aplicaciones, herramientas de desarrollo y aplicaciones, junto con servicios de consultoría, educación y soporte en más de 150 países de todo el mundo.

• **Excel 2007 Productos**

Excel 2007 es parte de las herramientas de oficina de Microsoft Corp. y está incluida en el paquete Office 2007. A las últimas versiones de Excel se han incorporados facilidades de visualización de datos, provenientes de cubos o bases multidimensionales. Como se ha descrito en apartados anteriores se integra muy bien con las funciones OLAP de Microsoft Sql Server 7.0.

Características Técnicas

A continuación se describen brevemente las funciones OLAP de Excel y las necesidades de software para poder brindarlas.

Funciones OLAP en Microsoft Excel

- *Recuperar y mostrar datos OLAP:* el componente OLAP puede conectarse a los orígenes de datos OLAP como lo haría a otros orígenes de datos externos. Puede trabajar con bases de datos creadas con servicios OLAP de Microsoft SQL Server, el producto servidor OLAP de Microsoft. Excel también puede trabajar con productos de otros proveedores compatibles con *OLE-DB* (interfase para comunicarse con bases de datos) para OLAP. Los datos OLAP se pueden mostrar cómo informe de tabla dinámica o de gráfico dinámico.
- *Crear archivos de cubo para utilizarlos sin conexión:* se puede utilizar el Asistente para cubos sin conexión de Excel para crear archivos con un subconjunto de los datos de una base de datos del servidor OLAP. Los archivos de cubo sin conexión permiten trabajar con datos OLAP sin estar conectado a la red. Sólo se puede crear archivos de cubo si se utiliza un proveedor OLAP.
- *Crear cubos a partir de bases de datos relacionales:* otro asistente, el Asistente para cubos OLAP, permite organizar datos consultados de bases de datos relacionales, como Microsoft SQL Server, en cubos OLAP. Un cubo permite trabajar con cantidades mayores de datos en un informe de tabla dinámica o gráfico dinámico que en otras circunstancias y, además, agiliza la recuperación de los datos.

Componentes de software necesarios para el acceso a OLAP

- *Un proveedor OLAP:* para configurar orígenes de datos OLAP es necesario uno de los siguientes proveedores OLAP:
 - *Proveedor OLAP de Microsoft:* Excel incluye el controlador de origen de datos (controlador de origen de datos: archivo de programa utilizado para conectarse a una base de datos específica. Cada programa o sistema de administración de bases de datos requiere un controlador diferente.) y el

software de cliente que necesitará para tener acceso a bases de datos creadas con el producto OLAP de Microsoft, Servicios OLAP de Microsoft SQL Server. El controlador que se incluye con Excel 2002 admite tanto la versión 7.0 como la versión 8.0 de este producto. Si tiene el controlador versión 7.0 incluido con Excel 2000, puede utilizarlo para tener acceso a bases de datos de la versión 7.0, pero para las bases de datos de la versión 8.0 deberá utilizar el controlador versión 8.0.

- *Otros proveedores OLAP:* para otros productos OLAP, deberá instalar controladores adicionales y software de cliente. Para utilizar las funciones de Excel a fin de trabajar con datos OLAP, el producto del proveedor debe adaptarse al estándar OLE-DB para OLAP y ser compatible con Microsoft Office. Para obtener información sobre cómo instalar y utilizar proveedores OLAP de otros fabricantes, póngase en contacto con el administrador del sistema o con el proveedor de los productos OLAP.
- *Bases de datos de servidor y archivos de cubo:* el software cliente OLAP de Excel admite conexiones a dos tipos de bases de datos OLAP.
 - Si en la red hay disponible una base de datos en un servidor OLAP, se puede recuperar datos de origen de la misma directamente.
 - Si se tiene un archivo de cubo sin conexión (archivo que se crea en el disco duro o en un recurso compartido de red para almacenar datos de origen OLAP) permite continuar trabajando cuando no se está conectado al servidor OLAP, ya que se pueden recuperar los datos de origen del mismo.
- *Orígenes de datos:* un origen de datos proporciona acceso a todos los datos de la base de datos o del archivo de cubo sin conexión. Una vez que se crea un origen de datos OLAP, los informes pueden basarse en él y los datos OLAP pueden devolverse a Excel como informes de tabla dinámica o gráfico

dinámico. Puede crear un origen de datos mientras utiliza el Asistente para tablas y gráficos dinámicos para crear un nuevo informe, o crearlo en Microsoft Query y utilizarlo para crear informes en Excel.

- *Microsoft Query*: es un componente opcional de Microsoft Office que se puede instalar y al que se puede tener acceso desde Excel. Puede utilizar Query para recuperar datos de una base de datos externa como Microsoft SQL o Microsoft **Access**.

Ventajas

- Es una herramienta de fácil uso para los usuarios, estos, generalmente están acostumbrados a la Interfaces gráficas muy intuitivas.
- Permite adecuar los reportes y gráficos a gusto de los usuarios.
- Permite realizar nuevos análisis de los datos que generalmente vienen pre-analizados desde servidores OLAP y servidores Datamining.
- Permite realizar las más comunes operaciones con los datos como por ejemplo *drill down*, *drill up* y *slice*, entre otras.
- Permite publicar toda la información en diferentes formatos, inclusive en formato HTML.
- Se integra fácilmente a otras herramientas de Inteligencia de Negocios.

4.7. Selección de la Solución

Las herramientas pre-seleccionadas son

- MS SQL SERVER
- ORACLE
- COGNOS SUITE
- SYNERA SUITE
- EXCEL 2000/XP.

Las dos primeras tienen como finalidad soportar físicamente el Datawarehouse. Ambas disponen de facilidades para la Carga, Extracción y

Transformación de los datos desde los sistemas transaccionales al Datawarehouse. MS SQL SERVER y ORACLE aportan aplicaciones de Inteligencia de Negocios para dar soporte a la toma de decisiones (OLAP y DataMining).

Las dos segundas tienen como objetivo dar soporte al usuario final en las tareas de Inteligencia de Negocios (EIS/DSS). Se dividen básicamente en herramientas de diseño, análisis, construcción y publicación de reportes.

Estos productos ofrecen funciones de usuario para resolver problemas de Inteligencia de Negocios facilitando el análisis y la toma de decisiones por medio de aplicaciones clientes fáciles de usar y entender.

También poseen características para facilitar la Extracción, Transformación y Carga desde los sistemas operacionales a los sistemas de Toma de Decisiones.

Finalmente Excel 2007 aporta facilidades de visualización de datos pudiéndose conectar a una variada fuente de datos multidimensionales.

La tabla 5-1 muestra un resumen y evaluación de las principales características de las herramientas preseleccionadas descritas en los apartados anteriores.

- **Evaluación de las Alternativas y Selección**

Las herramientas seleccionadas son MS SQL SERVER para dar soporte al Datawarehouse y análisis de datos y EXCEL 2000/XP para el acceso y visualización de datos.

A continuación se expone brevemente las razones de evaluación que ayudaron a seleccionar las herramientas que dan soporte al SUFESC.

MS SQL SERVER

La universidad dispone actualmente de MS SQL SERVER 7.0 que es el motor de base de datos que se utiliza para los sistemas transaccionales.

Los profesionales que dispone la Dirección de Sistemas de la universidad están muy capacitados en esta herramienta, fundamentalmente a lo referido al RDBMS

Criterios / Herramientas	MS SQL SERVER 7.0	ORACLE 9i	Synera Suite	Cognos Suite	Excel 2000/Xp
Soporte de Base de Datos/Datawarehouse	Alto	Alto	Bajo	Nulo	Nulo
Soporte a Extracción, Transformación Carga de Datos	Alto	Alto	Medio	Alto	Nulo
Soporte a OLAP	Alto	Alto	Alto	Alto	Bajo
Soporte a Datamining	Nulo	Alto	Alto	Alto	Bajo
Soporte al Acceso a Datos	Bajo	Alto	Alto	Alto	Alto
Capacidad de Escalabilidad y Crecimiento	Medio	Alto	Medio	Medio	Nulo
Soporte a la Seguridad	Alto	Alto	Alto	Alto	Bajo
Soporte a la Publicación de Información(Reporting)	Nulo	Medio	Medio	Alto	Bajo
Facilidad de Uso	Alto	Bajo	Alto	Bajo	Alto
Facilidad de Administración y Mantenimiento	Alto	Bajo	Alto	Medio	Alto
Facilidad de Integración con otras Herramientas	Alto	Alto	Bajo	Alto	Alto
Facilidad de Aprendizaje	Alto	Bajo	Medio	Medio	Alto
Recursos con Conocimientos en la Universidad	Alto	Medio	Bajo	Nulo	Alto
Adaptación a la Plataforma Técnica de la Universidad	Alto	Alto	Alto	Alto	Alto
Soporte de la Empresa Proveedora	Alto	Alto	Nulo	Alto	Alto
Costo de la Herramienta	Nulo	Alto	Nulo	Alto	Nulo
Adaptabilidad al Proyecto	Alto	Alto	Alto	Alto	Alto

Tabla 4-1. Resumen de las herramientas evaluadas.

Solamente en caso de utilizar la funcionalidad de Inteligencia de Negocios (OLAP Services y DTS como soporte a los procesos de Extracción, Transformación y Carga) los profesionales tendrán que realizar una breve capacitación o entrenamiento interno.

En el caso de ORACLE se debe realizar la compra de la versión ORACLE 9i, ya que la universidad no la posee. Además, los profesionales de la Dirección de Sistemas no poseen la capacitación necesaria para la administración de la herramienta, debiendo tomar cursos para tal fin. Como punto en contra la suite Oracle posee un gran número de aplicaciones que no se utilizarían.

EXCEL 2007

Excel 2007 dispone de un conjunto de funcionalidades que si bien no son tan potentes y versátiles como la de las herramientas de Cognos y Synera, se adaptan muy bien al proyecto (para las funciones de acceso y visualización de datos). Es una

herramienta que ya posee la universidad, está instalada en las estaciones de los usuarios y éstos ya están acostumbrados a su interfaz gráfica.

Tanto SYNERA SUITE como COGNOS SUITE ofrecen un entorno de trabajo completo de Inteligencia de Negocios. Como se observa en las demos de ambas herramientas y en la bibliografía recopilada son muy similares desde el punto de vista de la solución que ofrecen para la toma de decisiones e Inteligencia de Negocios en general.

Se debe tener en cuenta que Cognos Suite se deberá adquirir vía compra de licencias y realizar capacitación en la herramienta con el proveedor, mientras que con Synera Suite la empresa proveedora puede realizar un convenio con la Universidad.

CAPITULO 5. ANÁLISIS DEL SISTEMA

En el Capítulo se presenta el análisis del sistema. En primer término se muestra el modelo de datos del Sistema Académico, este modelo se divide en varios submodelos que agrupan de forma lógica las funcionalidades del sistema. Luego se comienza con la documentación de los requisitos de usuarios y su análisis, mediante diagramas de casos de uso, agrupándolos también según funcionalidades. Finalmente, se analiza el modelo de datos del Datawarehouse propiamente dicho, planteando las posibilidades de crecimiento y evolución.

5.1. Modelo de Datos Fuente

El modelo de datos que se describe a continuación pertenece al Sistema Académico, el cual es la fuente de datos para el SUFESC. A partir de este modelo de datos se empieza a construir el modelo del Datawarehouse. También se usa para realizar la validación y factibilidad de los requerimientos de los usuarios.

El modelo de datos fuente se divide en varios submodelos para una mejor comprensión y así facilitar su uso. Cada submodelo agrupa un conjunto de entidades que se relacionan con Alumnos, Docentes, Materias, Encuestas, Actas y Relaciones Institucionales. A efectos de la visualización de las relaciones entre los submodelos se adjunta al presente documento un archivo (Sistema Academico.ER1) con el diagrama completo en donde se pueden ver todas las relaciones entre dichos submodelos.

Entonces, las ventajas de utilizar submodelos son:

- Facilita manejar un diagrama muy grande
- Reduce el número de objetos con los que se debe trabajar
- Ayuda a focalizarse en una particular función del negocio.

En los diagramas de Entidad-Relación (ER) y en la Descripción de los Submodelos se repiten algunas entidades por los motivos explicados anteriormente.

5.1. Submodelos

Como se expuso anteriormente el modelo de datos se subdivide en seis modelos más pequeños. Estos seis modelos denominados submodelos comprenden las siguientes agrupaciones lógicas:

- Actas
 - Alumnos
 - Docentes
 - Encuestas
 - Materias
- **Submodelo Alumnos**

Diagrama ER

La figura 5-1 muestra las relaciones actuales del submodelo Alumnos:

- **Submodelo Actas Diagrama ER**

La figura 5-2 muestra las relaciones actuales del submodelo Actas:

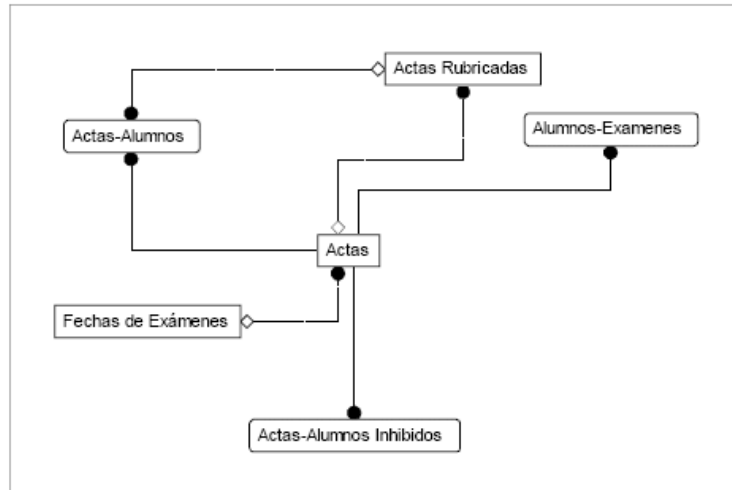


Figura 5-2. DER del Submodelo de Actas

- **Submodelo Docentes.**

Diagrama ER

La figura 5-3 muestra las relaciones actuales del submodelo Docentes:

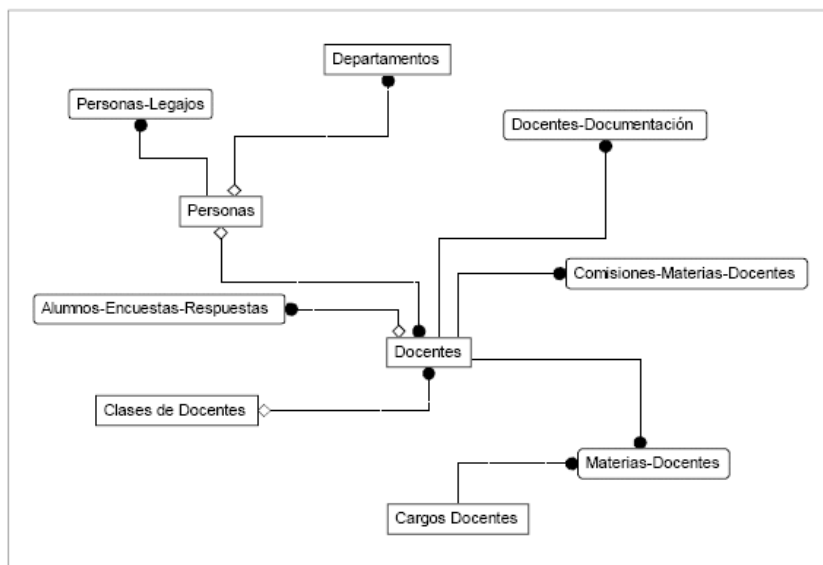


Figura 5-3. DER del Submodelo de Docentes.

- **Submodelo Encuestas**

Diagrama ER

La figura 5-4 muestra las relaciones actuales del submodelo Encuestas:

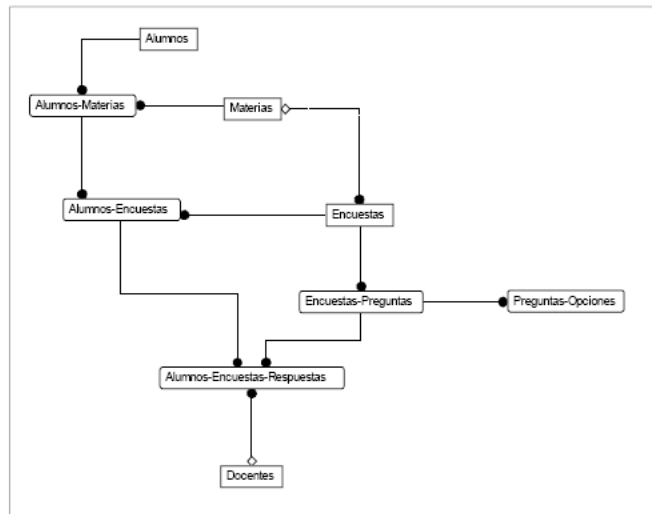


Figura 5-4. DER del Submodelo de Encuestas.

- **Submodelo Materias.**

Diagrama ER

La figura 5-5 muestra las relaciones actuales del submodelo Materias:

- **Submodelo Relaciones Institucionales**

Diagrama ER

La figura 5-6 muestra las relaciones actuales del submodelo Relaciones Institucionales:

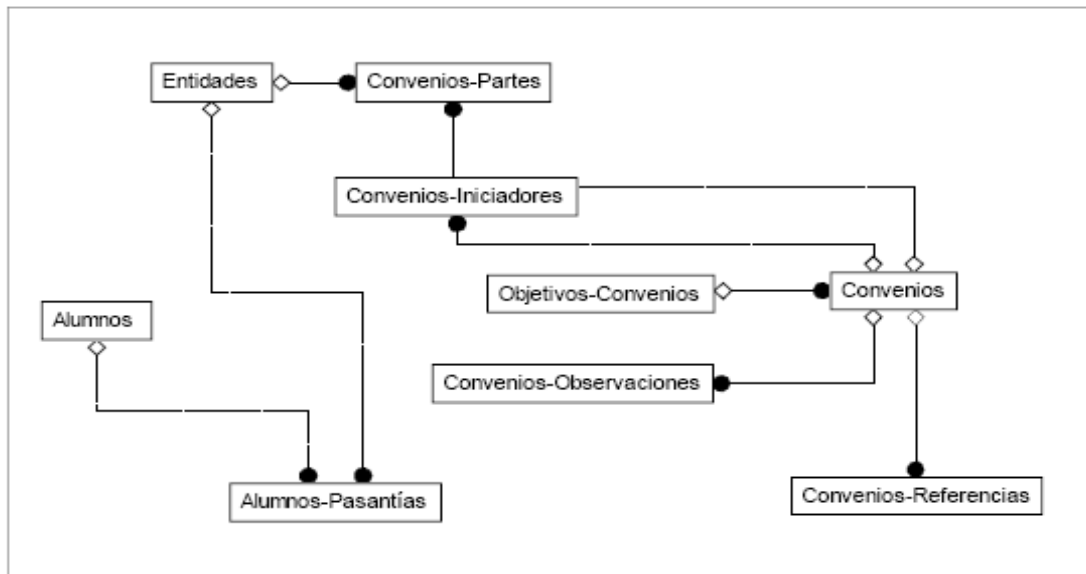


Figura 5-6. DER del Submodelo de Relaciones Institucionales.

5.2. Análisis de Requerimientos

El SUFESC está compuesto de tres componentes principales:

- Componente Datawarehouse y Datamart
- Componente Administración y Mantenimiento y
- Componente de Servicios OLAP.

A continuación se describe la modelización de dichos componentes. En primer término se describe un diagrama de contexto junto con los usuarios del sistema, luego se continúa con el análisis de los requerimientos de los usuarios, los cuáles están agrupados en paquetes lógicos, a continuación se sigue con el análisis del módulo de Administración y Mantenimiento y finalmente con el análisis del Datawarehouse y Datamart del Departamento de Ingeniería.

Los diagramas utilizados están basados en el Lenguaje Unificado de Modelado -UML- [Booch et al., 1999] y en el Anexo II se detallan las principales características.

- **Diagrama de Contexto**

La figura 6-7 muestra el diagrama de contexto del SUFESC. En el diagrama se observa a los usuarios del sistema, denominados usuarios Tomadores de Decisiones; las interfaces gráficas de usuarios, denominadas EIS/DSS, por medio de las cuáles se podrá acceder a los datos, realizar análisis y descubrimiento de información; el Datawarehouse, donde se encuentran físicamente los datos preparados para brindar servicios a la toma de decisiones; y finalmente la base de datos del sistema fuente, en este caso y en la primer etapa de la construcción del SUFESC, la base de datos del Sistema Académico.

En la figura 5-8 se observa un poco más de detalle, incorporándose al diagrama el usuario Centro de Cómputos, perteneciente a la Dirección de Sistemas y la funcionalidad Administración y Mantenimiento del Sistema. Se incorpora este detalle debido a que en los sistemas de Inteligencia de Negocios los procesos de administración, que incluyen la Extracción, Transformación y Carga de datos son procesos claves para su funcionamiento.

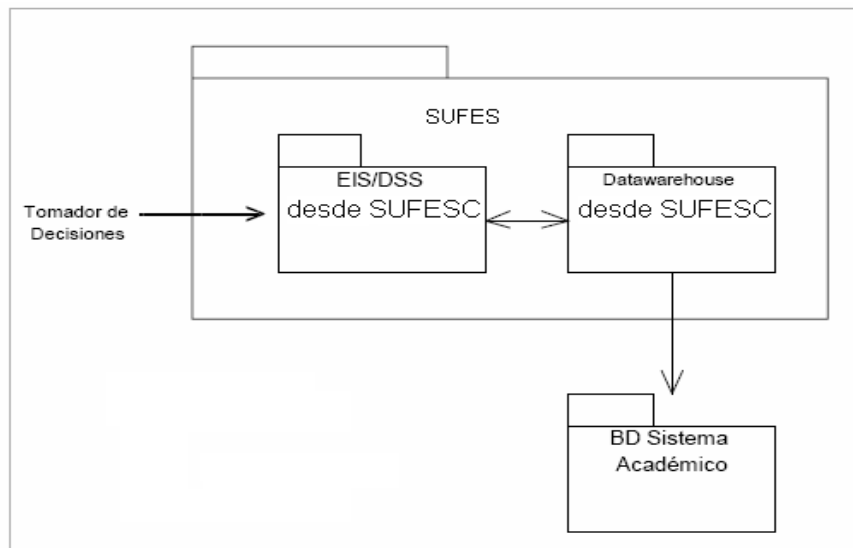


Figura 5-7. Diagrama de Contexto.

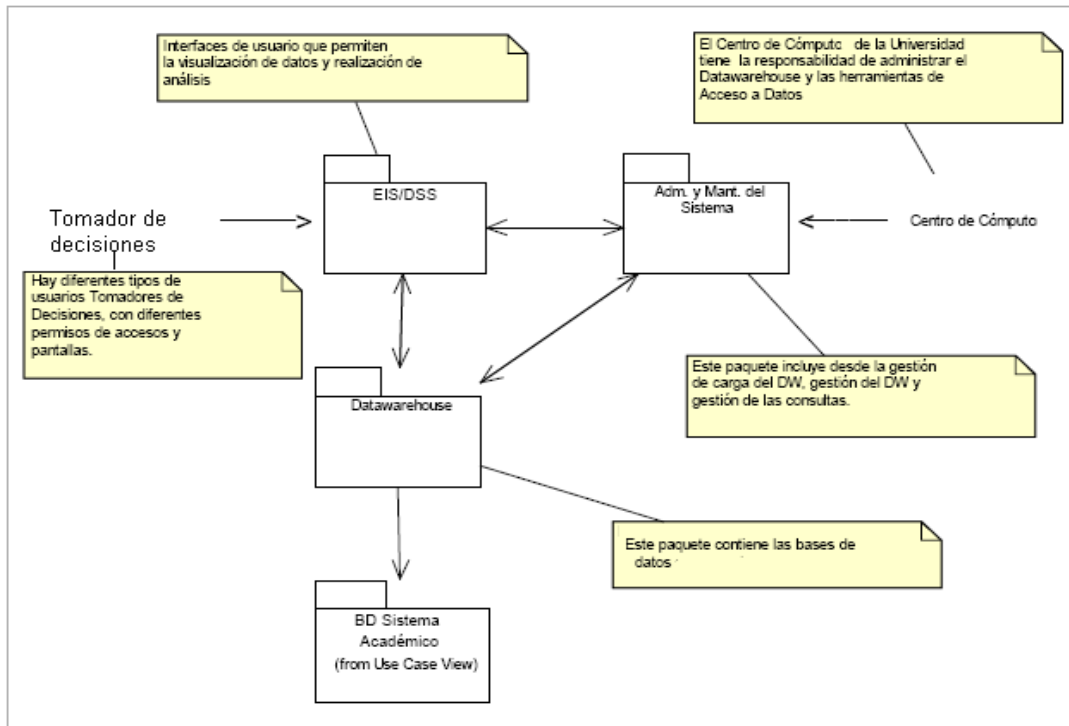


Figura 5-8. Diagrama de Contexto Ampliado.

- Usuarios**

La figura 5-9 muestra los diferentes tipos de usuarios del sistema

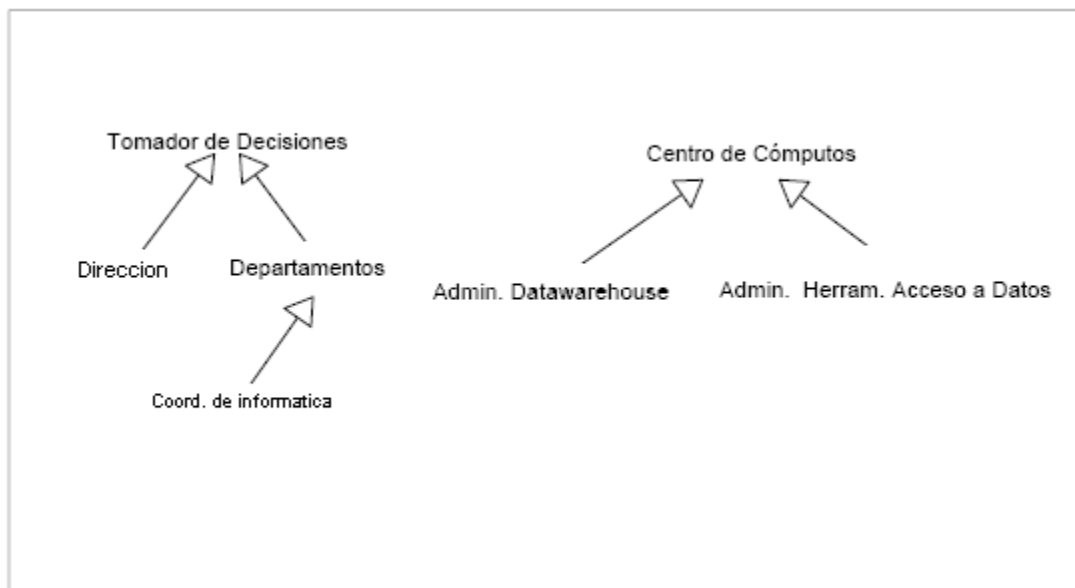


Figura 5-9. Usuarios del SUFESC.

Hay dos tipos de usuarios principales, usuarios *Tomador de Decisiones* y usuarios *Centro de Cómputos*.

El usuario *Tomador de Decisiones* es el usuario final del sistema que se divide en dos tipos: usuarios *Dirección* y usuarios *Departamentos*. Esta división es necesaria a efectos de la seguridad y accesos a determinadas funcionalidades del sistema. A su vez, el usuario *Departamentos* se divide en usuarios *la Coordinación de Informática*, ya que en un futuro habrá diferentes usuarios dependiendo de qué departamento o área de la universidad pertenezcan.

El tipo de usuario *Centro de Cómputo*, se divide en usuarios *Adm. del Datawarehouse* y *Adm. de herramienta de Acceso a Datos*. Ambos tipos de usuarios tienen funcionalidades específicas de administración y mantenimiento del sistema. El primero se ocupa de la administración de las bases de datos que conforman el Datawarehouse (*back-end* del sistema) y el segundo se ocupa de la gestión de la herramienta de consultas y análisis (*front-end* del sistema), dar soporte a los usuarios en las necesidades de información.

- **Análisis de los Requerimientos de Usuarios**

A efectos de describir los casos de uso que implementan requerimientos de usuarios se utiliza la terminología siguiente:

- Nombre del Caso de Uso

Nombre que identifica el caso de uso que implementa el requerimiento.

- Requerimiento que Implementa

Referencia al requerimiento que implementa el caso de uso. Se indica el apartado donde se enumera el requerimiento de usuario.

- Fuente de la Información

- Sistema o tipo de archivo en el que se encuentran los datos necesarios.

- Dimensiones

Visiones o ángulos en que se quiere ver la información. Estas visiones ángulos los utiliza el usuario para navegar y analizar la información desde diferentes perspectivas.

- Frecuencia de Actualización

Periodicidad con que se actualizan los datos.

- Momento de Actualización

Día en que se actualizan los datos.

- Fórmulas de cálculo.

Cálculos especiales.

- Necesidades de Seguridad

Permisos de accesos a la información.

- Cantidad / historia de datos a almacenar

Tiempo y cantidad de datos históricos a almacenar.

- Otros

Observaciones aclaratorias de cálculo.

- Observaciones

Notas aclaratorias complementarias.

Paquete EIS/DSS

El paquete EISS/DSS está dividido en subpaquetes lógicos que agrupan funcionalidades orientadas a una misma entidad conceptual. Esta agrupación

Facilita su entendimiento y mejor análisis como así también el desarrollo de cada funcionalidad.

También en el análisis se evalúa la factibilidad de brindar los requerimientos solicitados. Esta evaluación se realiza en base a la disponibilidad de los datos y de la estructura de datos en la base de datos del Sistema Académico.

El figura 5-10 muestra las diferentes agrupaciones de necesidades de usuarios

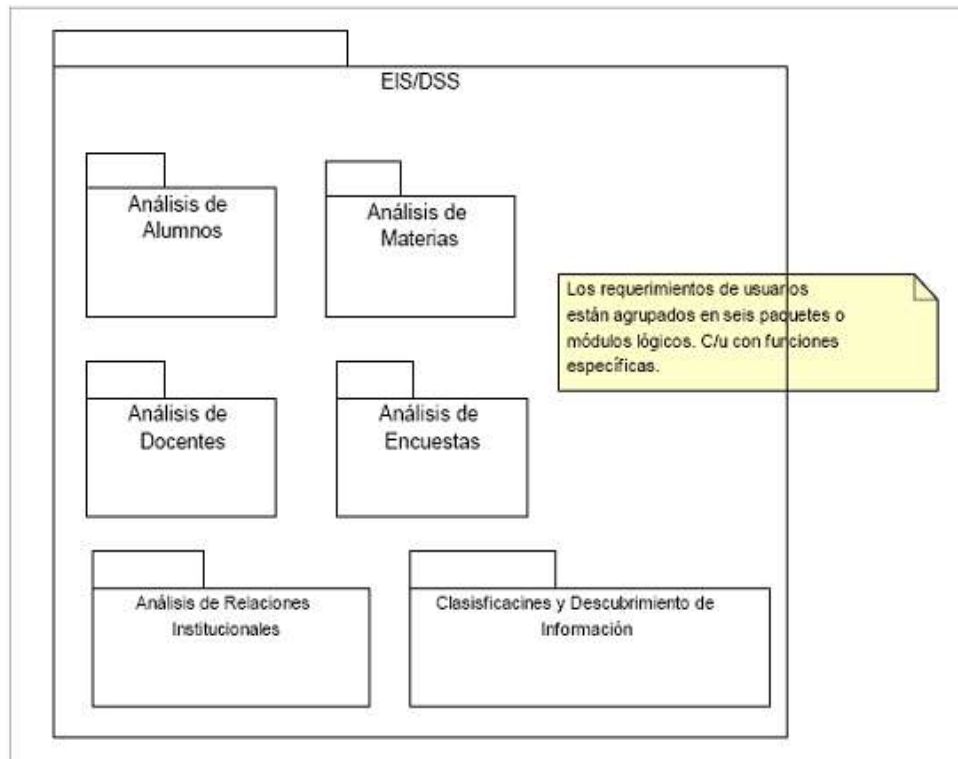
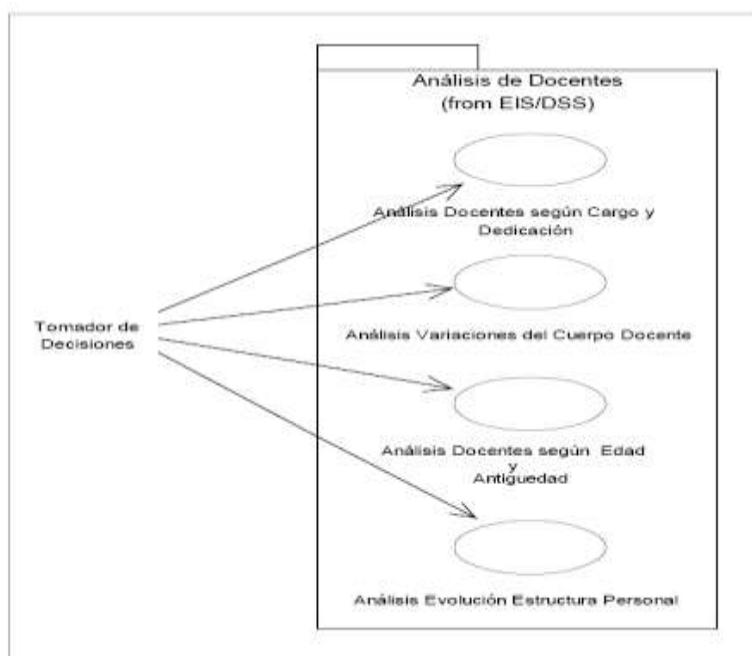


Figura 5-10. Paquete EIS/DSS.

- **Paquete de Análisis de Docentes**

El paquete de la figura 5-11 agrupa las necesidades de información en lo concerniente a los Docentes.

Las tablas 5-9 a 5-12 muestran la especificación funcional de los casos de uso que componen dicho paquete



Caso de	Análisis docentes según cargo y dedicación.
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 1, 2, y 3.
Fuente de la Información	Sistema Académico
Frecuencia de	Mensual
Momento de Actualización	Ultimo día del mes
Dimensiones de análisis de la información	Cargo - Dedicación - Planes - Materias - Títulos Docentes - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de Datos a Almacenar	Toda la historia
Otros	- La dedicación semanal debe responder según el siguiente rango horario: a) < 9hs; b) 10 a 19 hs; c) 20 a 29 hs; d) 30 a 39hs y e) > 40 hs. - Incluir totales del departamento
Observaciones	

Tabla 5-9. Caso de Uso Análisis docentes según cargo y dedicación.

Caso de uso	Análisis variaciones del cuerpo docente
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 4
Fuente de la información	Sistema Académico
Frecuencia de actualización	Semestral
Momento de actualización	Ultimo día del semestre
Dimensiones de análisis de la información	- Planes - Cargo - Dedicación
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Dedicación entre 2 años, 1996 y 2001.
Observaciones	

Tabla 5-10. Caso de Uso Análisis variaciones del cuerpo docente.

Caso de uso	Análisis de Docentes por edad y antigüedad
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 5
Fuente de la información	Sistema Académico
Frecuencia de actualización	Mensual
Momento de actualización	Ultimo día del mes
Dimensiones de análisis de la información	- Planes - Materias - Antigüedad - Edad - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Incluir totales del departamento
Observaciones	Las remuneraciones y ausentismo de los docentes no están disponibles en el modelo de datos fuente. Esta información se encuentra en el sistema Administrativo/Contable. No se puede brindar.

Tabla 6-11. Caso de Uso Análisis de Docentes por edad y antigüedad.

Caso de uso	Análisis evolución de la estructura de personal
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 6
Fuente de la información	Sistema Académico
Frecuencia de actualización	Mensual
Momento de actualización	Ultimo día del mes
Dimensiones de análisis de la información	- Planes - Materias - Cargos - Dedicación - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Incluir totales y subtotales
Observaciones	

Tabla 5-12. Caso de Uso Análisis evolución de la estructura de personal.

- **Paquete Análisis de Alumnos**

El paquete agrupa las necesidades de información en lo concerniente a los Alumnos.

La figura 5-12 muestra el paquete con los casos de uso asociados. Las tablas 5-13 a 5-19 muestran la especificación funcional de los casos de uso que componen dicho paquete.

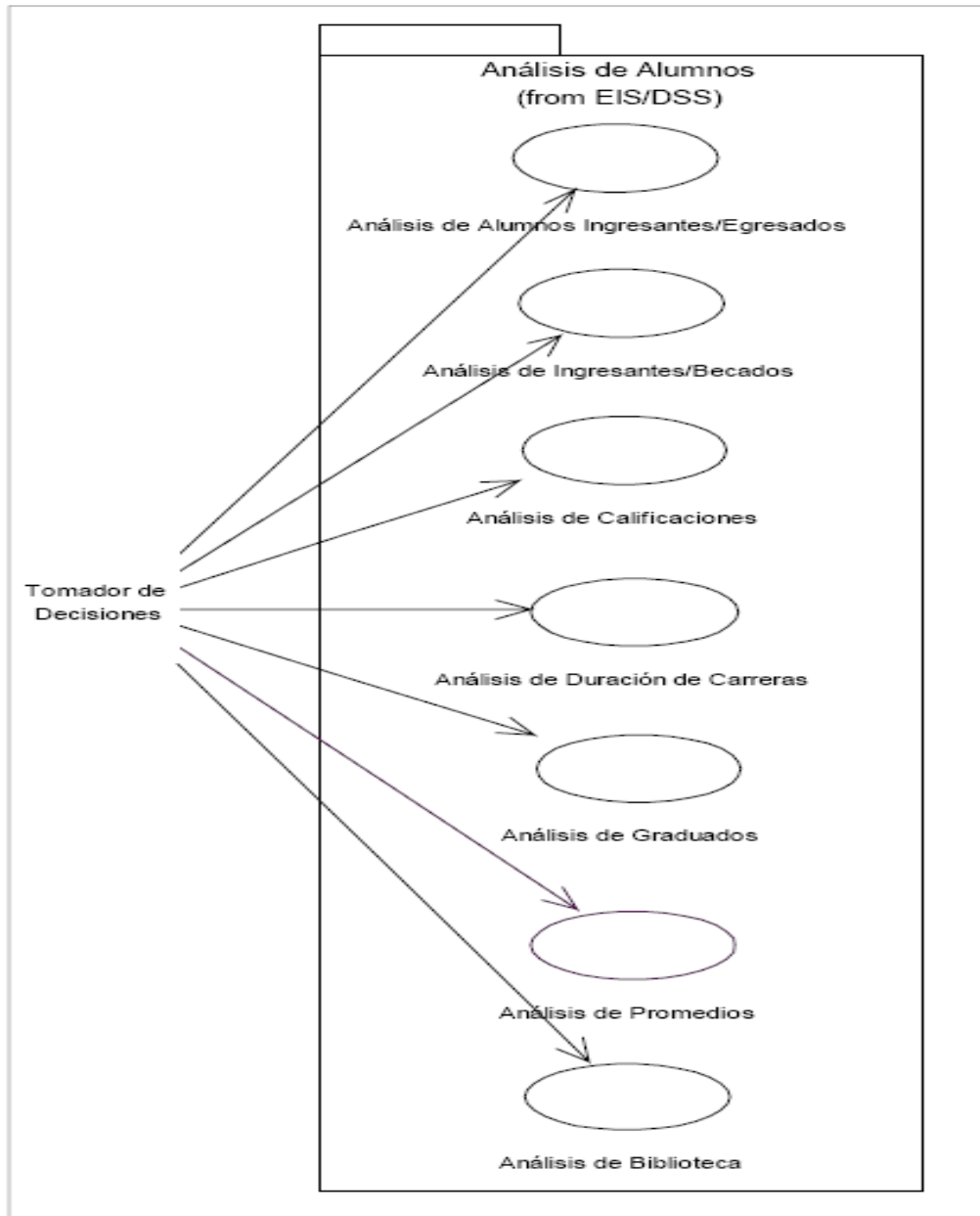


Figura 5-12. Paquete Análisis de Alumnos

Caso de uso	Análisis docentes según cargo y dedicación.
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 1, 2, y 3.
Fuente de la Información	Sistema Académico
Frecuencia de Actualización	Mensual
Momento de Actualización	Ultimo día del mes
Dimensiones de análisis de la información	- Cargo - Dedicación - Planes - Materias - Títulos Docentes - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de Datos a Almacenar	Toda la historia
Otros	- La dedicación semanal debe responder según el siguiente rango horario: a) < 9hs; b) 10 a 19 hs; c) 20 a 29 hs; d) 30 a 39hs y e) > 40 hs. - Incluir totales del departamento
Observaciones	

Tabla 5-9. Caso de Uso Análisis docentes según cargo y dedicación

Caso de uso	Análisis variaciones del cuerpo docente
Requerimiento que implementa	Ver ¿Qué es un EIS? requisito nro. 4
Fuente de la información	Sistema Académico
Frecuencia de actualización	Semestral
Momento de actualización	Ultimo día del semestre
Dimensiones de análisis de la información	- Planes - Cargo - Dedicación
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Dedicación entre 2 años, 1996 y 2001.
Observaciones	

Tabla 5-11. Caso de Uso Análisis de Docentes por edad y antigüedad.

Caso de uso	Análisis evolución de la estructura de personal
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 6
Fuente de la información	Sistema Académico
Frecuencia de actualización	Mensual
Momento de actualización	Ultimo día del mes
Dimensiones de análisis de la información	- Planes - Materias - Cargos - Dedicación - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Incluir totales y subtotales
Observaciones	

Tabla 5-12. Caso de Uso Análisis evolución de la estructura de personal.

- **Paquete Análisis de Materias**

El paquete de la figura 5-13 agrupa las necesidades de información en lo concerniente a las Materias.

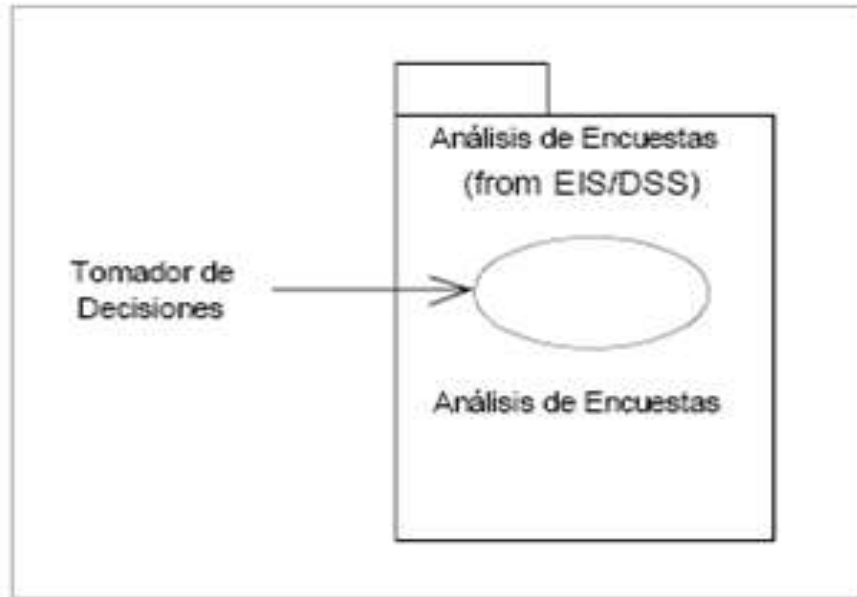


Figura 5-13. Paquete Análisis de Materias.

Caso de uso	Cantidad de Alumnos por materia cuatrimestral
Requerimiento que implementa	Ver ¿Qué es un EIS? requisito nro. 12
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Semestral
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Materia - Alumnos - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Incluir totales y subtotales.
Observaciones	

Tabla 5-20. Caso de Uso Cantidad de Alumnos por materia cuatrimestral

Las tablas 5-20 a 5-22 muestran la especificación funcional de los casos de uso que componen dicho paquete.

Caso de uso	Análisis de Alumnos Matriculados
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 13
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Semestral
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Materia - Alumnos - Notas - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	- Incluir totales y subtotales.
Observaciones	

Tabla 5-21. Caso de Uso Análisis de Alumnos Matriculados.

Caso de uso	Análisis de Notas por Materia Cursada
Requerimiento que implementa	Ver ¿Qué es un EIS? requisito nro. 14
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Seme
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Materia - Alumnos – Notas, Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	Se debe incluir: Aprobada total – Aprobada con No aprobada Promedio cursada Finales (aprobado /reprobado) totales y subtotales.
Observaciones	La información completa de notas de final no está en la base de datos del Sistema Académico. Esta información está en papel (libros de exámenes). Solo es posible obtener por medio del sistema fuente la nota final de la materia y la cantidad de veces rendida.

Tabla 5-22. Caso de Uso Análisis de Notas por Materia Cursada.

- **Paquete Análisis de Encuestas**

El paquete de la figura 5-14 agrupa las necesidades de información en lo concerniente a las Encuestas realizadas a los alumnos.

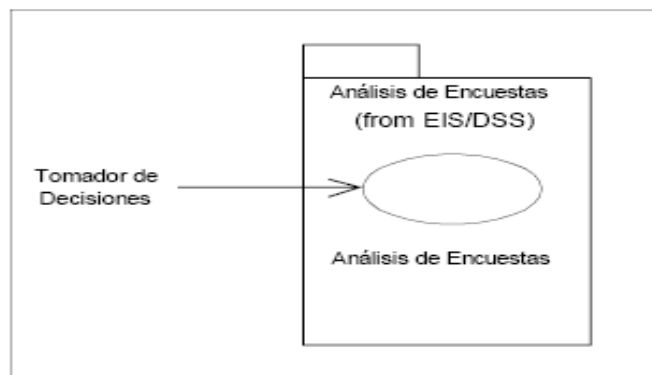


Figura 5-14. Paquete Análisis de Encuestas.

La tabla 5-23 muestra la especificación del caso de uso para el Análisis de Encuestas.

Caso de uso	Análisis de Encuestas
Requerimiento que implementa	Ver ¿Qué es un EIS? Requisito nro. 15.
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Semestral
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Materia - Docente - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	Se debe incluir: - Rendimiento de la materia: • Clases teóricas • Trabajos de aplicación • Organización • Bibliografía empleada • Opinión general - Calificación General: • Rendimiento docente • Rendimiento de la cátedra • Opinión general • Rendimiento general
Observaciones	La información de encuestas que está en la base de datos del Sistema Académico no está en formato discreto, las preguntas y respuestas están almacenadas como textos, impidiendo su cálculo numérico.

Tabla 5-23. Caso de Uso Análisis de Encuestas.

- **Paquete Análisis de Relaciones Institucionales**

El paquete de la figura 5-15 agrupa las necesidades de información en lo concerniente a las Relaciones Institucionales con otras entidades, ya sean otras universidades, colegios o empresas.

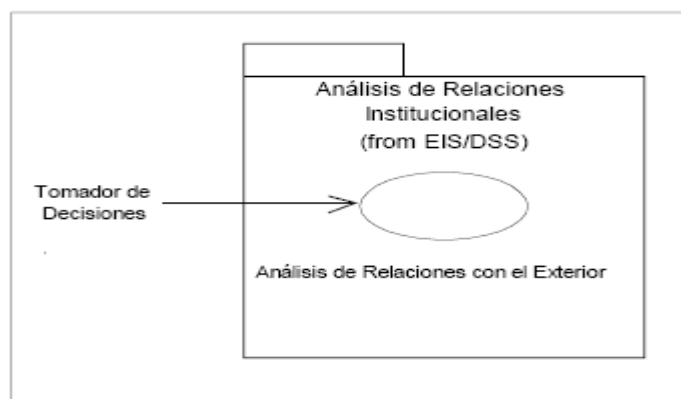


Figura 5-15. Paquete Análisis Relaciones Institucionales.

La tabla 5-24 muestra la especificación del caso de uso para el Análisis de relaciones con el exterior.

Caso de uso	Análisis de relaciones con el exterior
Requerimiento que implementa	Ver ¿Qué es un EIS? requisito nro. 16.
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Semestral
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Entidades - Convenios - Pasantes - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	Se debe incluir: - Cantidad de conferencias realizadas - Cantidad de visitantes recibidos - Cantidad de convenios realizados - Gastos en publicidad - Cantidad de pedidos de empresas por pasantes - Incluir totales y subtotales.
Observaciones	La información de Relaciones Institucionales no está en la base de datos del Sistema Académico. No se puede brindar este requerimiento.

Tabla 6-24. Caso de Uso Análisis de relaciones con el exterior.

- **Paquete de Clasificaciones y Descubrimiento de Información**

El paquete de la figura 5-16 agrupa las necesidades de información en lo concerniente a las posibilidades que debe brindar el sistema para realizar Data Mining.

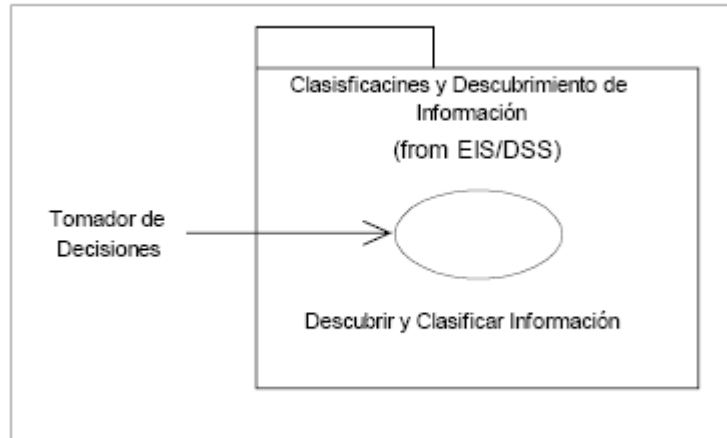


Figura 5-16. Paquete Clasificación y descubrimiento de Información.

La tabla 5-25 muestra la especificación del caso de uso Descubrir y Clasificar Información.

Caso de uso	Descubrir y Clasificar Información
Requerimiento que implementa	Ver ¿Qué es un EIS? requisito nro. 17 y 18.
Fuente de la información	Sistema Académico
Frecuencia de actualización	- Semestral
Momento de actualización	- Ultimo día del semestre
Dimensiones de análisis de la información	- Alumnos - Docentes - Fecha
Fórmulas de cálculo	N/A
Necesidades de seguridad	Director / jefe de departamento /coordinador
Cantidad / Historia de datos a almacenar	Toda la historia
Otros	Las clasificaciones a realizar son: A - Que características tienen los Alumnos, según desempeño académico? B- Que características tienen los mejores Profesores, según encuestas?
Observaciones	Los requerimientos de clasificación se obtienen mediante técnicas de Data Mining. Se debe aplicar las técnicas y pasos correspondientes para obtener las clasificaciones arriba indicadas. Además, la pregunta B- adolece de los problemas nombrados en el caso de uso "Análisis de Encuestas.

Tabla 5-25. Caso de Uso Descubrir y Clasificar Información.

5.3 Análisis del Datawarehouse

En esta sección se comienza a realizar los análisis preliminares del Datawarehouse. Se define el modelo de datos del Datawarehouse y se comienza a delinear su arquitectura, la cual se profundiza y se detalla en el diseño del sistema.

- **Paquete Datawarehouse**

El paquete Datawarehouse agrupa las bases de datos que contienen los datos provenientes de los sistemas fuentes.

Como se observa en la figura 5-17, el Datawarehouse está compuesto por dos bases de datos (que físicamente podrán estar en una misma base de datos la primera corresponde a una base denominada *Intermedia* o *Detalle* que es cargada desde la base de datos del Sistema Académico por medio de los procesos de *Administración y Mantenimiento del Sistema*.

La segunda base de datos, denominada *Datamart* o *Multidimensional* se carga también por medio de los procesos de *Administración y Mantenimiento del Sistema*, extrayendo los datos de la base *Intermedia/Detalle* y transformándolos. Esta transformación responde, como se explica más adelante, a brindar información desde una visión multidimensional, la cual facilita el análisis y toma de decisiones.

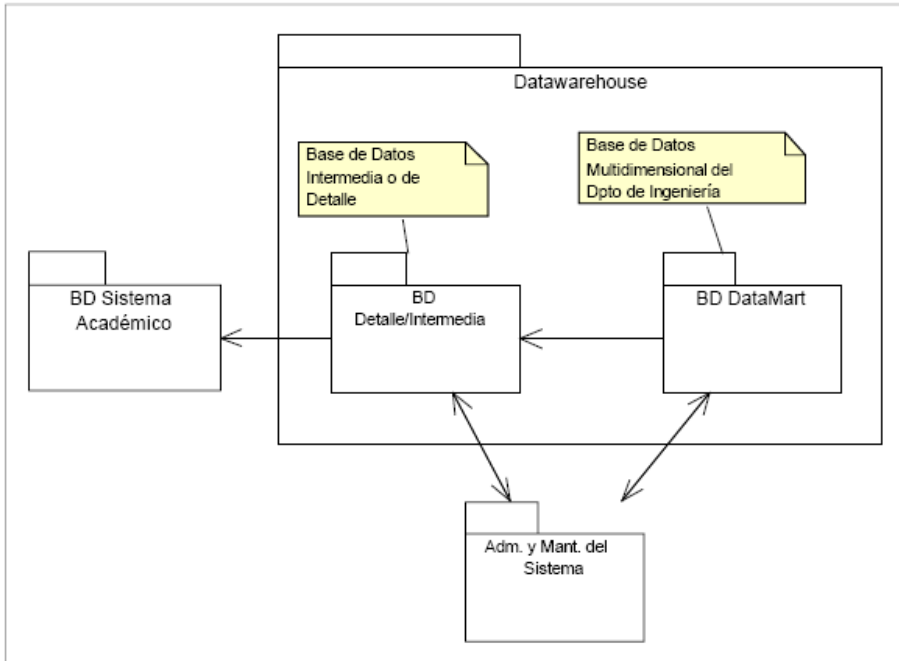


Figura 5-17. Paquete Datawarehouse.

- **Modelo de Datos de la Base de Datos Intermedia**

La estructura de datos de la base intermedia (o de detalle) es igual al modelo de datos presentado en este capítulo para la Db (base de datos) fuente. Los nombres de las entidades comienzan con el prefijo dwi_, los atributos de las entidades de la base de datos intermedia son los mismos que los de las entidades fuente. Dichos atributos mantendrán los mismos nombres a efectos de facilitar la administración y mantenimiento del Datawarehouse.

La tabla 5-26 muestra las entidades de la base intermedia.

Nombre entidad Dd fuente	Nombre entidad De	Descripción entidad Dd
Actas	Dwi_Actas	Actas de Exámenes
Actas Rubricadas	Dwi_Actas Rubricadas	Actas de Exámenes Rubricadas
Actas-Alumnos	Dwi_Actas-Alumnos	Alumnos del Acta de Examen
Actas-Alumnos Inhibidos	Dwi_Actas-Alumnos	Alumnos Inhibidos del Acta de
Alumnos	Dwi_Alumnos	Alumnos
Alumnos-Adicionales	Dwi_Alumnos-Adicionales	Información Adicional de Alumnos
Alumnos-Ayudantías	Dwi_Alumnos-Ayudantías	Ayudantías de Alumnos
Alumnos-Becas	Dwi_Alumnos-Becas	Becas del Alumno
Alumnos-Cambios de Estado	Dwi_Alumnos-Cambios de	Cambios de Estado del Alumno

Nombre entidad Dd fuente	Nombre entidad Dd	Descripción entidad De
Alumnos-Carreras	Dwi_Alumnos-Carreras	Carreras de Alumnos
Alumnos-Diplomas	Dwi_Alumnos-Diplomas	Diplomas de Alumnos
Alumnos-Encuestas	Dwi_Alumnos-Encuestas	Encuestas de Alumnos
Alumnos-Encuestas-Respuestas	Dwi_Alumnos-Encuestas-	Respuestas de Encuestas de Alumnos
Alumnos-Equivalencias	Dwi_Alumnos-Equivalencias	Equivalencias Internas/Externas de Alumnos
Alumnos-Exámenes	Dwi_Alumnos-Exámenes	Alumnos-Exámenes
Alumnos-Familiares	Dwi_Alumnos-Familiares	Familiares del Alumno
Alumnos-Inscripciones	Dwi_Alumnos-Inscripciones	Inscripciones de Alumnos
Alumnos-Materias	Dwi_Alumnos-Materias	Alumnos-Materias
Alumnos-Materias Externas	Dwi_Alumnos-Materias	Materias Externas de Alumnos
Alumnos-Pasantías	Dwi_Alumnos-Pasantías	Pasantías de Alumnos
Alumnos-Rendimiento Académico	Dwi_Alumnos-Rendimiento	Rendimiento Académico de Alumnos
Alumnos-Títulos	Dwi_Alumnos-Títulos	Títulos de Alumnos
Cargos Docentes	Dwi_Cargos Docentes	Cargos Docentes
Carreras	Dwi_Carreras	Carreras ó Especialidades
Clases de Docentes	Dwi_Clases de Docentes	Clases de Docentes
Comisiones-Materias	Dwi_Comisiones-Materias	Comisiones/Materias
Comisiones-Materias-Dictado	Dwi_Comisiones-Materias-Dictado	Dictados de Comisión/Materia
Comisiones-Materias-Docentes	Dwi_Comisiones-Materias-Docentes	Docentes de Comisión/Materia
Convenios	Dwi_Convenios	Convenios
Convenios-Iniciadores	Dwi_Convenios-Iniciadores	Personas que originan Convenio
Convenios-Observaciones	Dwi_Convenios-	Observaciones de Convenios
Convenios-Partes	Dwi_Convenios-Partes	Partes de Convenios
Convenios-Referencias	Dwi_Convenios-Referencias	Referencias de Convenios
Correlatividades	Dwi_Correlatividades	Correlatividades de Materias
Departamentos	Dwi_Departamentos	Departamentos
Docentes	Dwi_Docentes	Docentes
Docentes-Documentación	Dwi_Docentes-	Documentación de Docentes
Encuestas	Dwi_Encuestas	Encuestas de Materias
Encuestas-Preguntas	Dwi_Encuestas-Preguntas	Preguntas de Encuestas de
Entidades	Dwi_Entidades	Entidades
Equivalencias	Dwi_Equivalencias	Equivalencias de Materias
Fechas de Exámenes	Dwi_Fechas de Exámenes	Fechas de Exámenes
Materias	Dwi_Materias	Materias
Materias-Docentes	Dwi_Materias-Docentes	Docentes de la Materia
Materias-Planes de Estudio	Dwi_Materias-Planes de	Materias de Planes de Estudio
Matriculación	Dwi_Matriculación	Matriculación de Alumnos
Matriculación-Alumnos	Dwi_Matriculación-Alumnos	Alumnos Matriculados
Matriculación-Materias	Dwi_Matriculación-Materias	Materias de Alumnos Matriculados
Modalidades de Ingreso	Dwi_Modalidades de Ingreso	Modalidades de Ingreso
Modalidades de Ingreso-Materias	Dwi_Modalidades de Ingreso-Materias	Porcentajes de Ponderación de las Materias de Ingreso de acuerdo a la Modalidad para el
Niveles de Materias	Dwi_Niveles de Materias	Niveles de Materias
Numeración de Alumnos	Dwi_Numeración de	Numeración de Alumnos por Tipo
Objetivos-Convenios	Dwi_Objctivos-Convenios	Objetivos de Convenios
Orientaciones de Carreras	Dwi_Orientaciones de	Orientaciones de Carrera ó
Personas	Dwi_Personas	Personas

Nombre entidad Dd fuente	Nombre entidad Dd	Descripción entidad Dd
Personas-Legajos	Dwi_Personas-Legajos	Legajos de Personas (Docentes/Responsables de Departamento)
Planes de Estudio	Dwi_Planes de Estudio	Planes de Estudio de Carrera ó Especialidad
Preguntas-Opciones	Dwi_Preguntas-Opciones	Opciones de Preguntas Multiple
Requisitos de Materias	Dwi_Requisitos de Materias	Requisitos de Materias
Solicitudes de Excepciones	Dwi_Solicitudes de Excepciones	Solicitudes de Excepciones de
Solicitudes de Excepciones de	Dwi_Solicitudes de Excepciones de	Solicitudes de Excepciones de
Solicitudes de Excepciones de	Dwi_Solicitudes de Excepciones de Exámenes	Solicitudes de Excepciones de Exámenes
Solicitudes de Excepciones de Materias	Dwi_Solicitudes de Excepciones de Materias	Solicitudes de Excepciones de Materias
Solicitudes de Inscripción	Dwi_Solicitudes de	Solicitudes de Inscripción de Ingreso
Tipos de Alumno	Dwi_Tipos de Alumno	Tipos de Alumno
Tipos de Carreras	Dwi_Tipos de Carreras	Tipos de Carreras
Títulos	Dwi_Títulos	Títulos
Títulos Anexos	Dwi_Titulos Anexos	Títulos Anexos de Alumnos
N/A	dwi_d_antiguedad	Antigüedades por rango
N/A	dwi_d_edad	Edades por Rangos
N/A	dwi_d_estado_alumnos	Estado de los alumnos(Egresado, Regular,
N/A	dwi_d_rango_horario	Dedicación de docentes por rangos
N/A	dwi_d_rango_notas	Notas de exámenes por rango
N/A	dwi_d_rango_tiempo_graduarse	Cantidad de Años por Rangos
N/A	dwi_d_rindio_final	Descripción (valores Si, No)
N/A	dwi_d_situacion_docente	Descripción (valores Alta, Baja)

Tabla 5-26. Entidades de la base Intermedia del Datawarehouse.

• **Análisis del Enfoque de Construcción**

A continuación se describen consideraciones del uso de base de datos Intermedia en el SUFESC.

- Este Datawarehouse se construye de una manera heurística, por lo tanto se hace necesario disponer del detalle completo de la base de datos transaccional.
- En un primer momento, no se conocen todos los requisitos de usuarios necesarios.
- A medida que los usuarios van usando el Datawarehouse, van pidiéndole funcionalidades nuevas. Permite ejecutar consultas desconocidas.
- La información de detalle siempre está disponible fuera del ambiente transaccional.
- No hay procesos complejos de extracción y carga, estos se realizan dentro del Datawarehouse.
- Posibilita el uso de herramientas de Inteligencia de Negocios que acceden al detalle de los datos para crear sus propios repositorios.

- Permite la creación de Datamarts a partir de la información del Datawarehouse.

5.4. Modelo de datos de la Base de Datos Multidimensional

La base Multidimensional o Datamart está compuesta inicialmente por tablas denominadas Dimensionales y Tablas de Hechos. En esta base de datos se implementa las necesidades específicas del Dpto. de Ingeniería. También en esta base de datos se da soporte al almacenamiento de los cubos multidimensionales creados por medio de la herramienta Servicios OLAP.

Las tablas 5-27 y 5-28 describen las entidades y sus atributos de la base multidimensional o Datamart.

Nombre entidad	Descripción entidad
D_Cargos_Docentes	Entidad Dimensional que contiene la descripción de los cargos docentes.
D_Rango_Notas	Entidad Dimensional que contiene las notas de Alumnos agrupados por rangos.
D_Rango_Graduarse	Entidad Dimensional que contiene los tiempos de graduación de Alumnos agrupados por rangos.
D_Estado_Alumno	Estados en los que se encuentra el alumno (Regular, Egresado, Baja, Latente, otros)
D_Materias	Entidad Dimensional de materias.
D_Rango_Antigüedad	Entidad Dimensional que contiene la antigüedad por rangos.
D_Rango_Edad	Entidad Dimensional que contiene la edad por rangos.
D_Rango_Horario	Entidad Dimensional que contiene la dedicación horaria por rangos de los docentes
D_Situación_Docente	Entidad Dimensional que contiene descripciones de estados (Alta, Baja) de docentes
H_Alumnos_Duracion_Carrera	Entidad de Hechos que contiene la cantidad de alumnos egresados con duración de la carrera por rangos, por plan y año.
H_Alumnos_finales_aprobados	Entidad de Hechos que contiene la cantidad de alumnos con las notas por rango, plan, materia y año.
H_Alumnos_finales_desaprobados	Entidad de Hechos que contiene la cantidad de alumnos con las notas por rango, plan, materia y año.
H_Alumnos_ingresantes_egresados	Entidad de Hechos que contiene la cantidad de alumnos con las fechas de ingreso y egreso por plan y estado alumnos.
H_Alumnos_Promedios	Entidad de Hechos que contiene los promedios de notas finales de los alumnos por plan, materia y año.
H_Docentes_Edad_Antigüedad	Entidad de Hechos que contiene la cantidad de cargos Docentes por Edad y Antigüedad por materia y cargo
H_Docentes_Evolucion	Entidad de Hechos que contiene la cantidad de cargos Docentes según años de ingreso y egreso por materia y cargo
H_Docentes_por_Cargo_Dedicacio	Entidad de Hechos que contiene la cantidad de cargos Docentes por cargo y dedicación por materia.
H_Materias_Analisis_Cursada	Entidad de Hechos que contiene la cantidad de alumnos con las notas de cursada, por plan, materias y año.
H_Materias_Matriculados	Entidad de Hechos que contiene la cantidad de alumnos matriculados por año, plan y materia.

Tabla 5-27. Entidades del Datamart para el paquete EIS/DSS.

Nombre de la entidad	Clave	Descripción entidad
D_Rango_Antigüedad	claveantig	Clave de la tabla.
	descriprango	Descripción del rango de antigüedad.
	rangodesde	Cota mínima del rango al cual pertenece la antigüedad
	rangohasta	Cota máxima del rango al cual pertenece la antigüedad
D_Cargos_Docentes	c_doc_cod	Clave de la tabla
	c_doc_desc	Descripción del cargo.
H_Docentes_Edad_Antigüedad	ad Antigüedad del d	Antigüedad del docente en la materia.
	claveedad	Clave foránea de la descripción del rango de la edad del docente.
	claveantig	Clave foránea de la descripción del rango de la antigüedad del docente.
	Edad	Edad del Docente
	Per_id	Clave foránea del nombre del docente.
	Mat_cod	Clave foránea de la descripción de la materia.
H Docentes _Evolucion	claverango	Clave foránea de la descripción de la dedicación horaria del docente.
	Clavefecha	Clave foránea de la descripción de la fecha.
	clavesituacion	Clave foránea de la descripción del estado del docente.
	Per_id	Nro. de identificación de persona (docente).
	Mat_cod	Clave foránea de la descripción de la materia.
	c_doc_cod	Clave foránea de la descripción del cargo del docente.
H_Docentes_por_Cargo_Dedicacion	Mat_cod	Clave foránea de la descripción de la materia.
	c_doc_cod	Clave foránea de la descripción del cargo.
	Per_id	Nro de identificación de persona (docente).
	clavefecha	Clave foránea de la descripción de la fecha.
	claverango	Clave foránea de la descripción del rango de dedicación horaria.
	clavetitulo	Clave foránea de la descripción del título del docente.
D_Rango_Edad	claveedad	Clave de la tabla
	descripedad	Descripción del rango al cual pertenece la edad.
	rangomin	Cota mínima de rango.
	rangomax	Cota máxima de rango.
D Rango Graduarse	claverangograd	Clave de la tabla
	desc_rangograd	Descripción del rango al cual pertenece el tiempo de graduación.
	rangomin	Cota mínima de rango.
	rangomax	Cota máxima de rango.
D Materias	Mat_cod	Clave de la tabla
	Mat_tipo	"Tipo de Materia (Cuatrimestral, Anual)
	Mat_desc	Descripción de la materia.
D_Rango_Horario	claverango	Clave de la tabla
	rangoMin	Cota mínima del rango.
	rangoMax	Cota máxima del rango.
	desc_rango	Descripción del rango al cual pertenece la dedicación.
D_Situación_Docente	clavesituacion	Clave de la tabla.
	descripsituacion	"Descripción de la situación (Alta Baja)
D_Estado_Alumnos	claveestalumno	Clave de la tabla.
	descripestalumno	"Descripción del estado del alumno (Regular, Baja, Egresado, Latente, Otros)
H_Alumnos_Duracion_Carrera	ple_cod	Clave foránea de Plan del alumno
	alu id	Identificación del Alumno
	desc rangograd	Clave foránea de Tiempo de graduación por rango
	alu clase	Año de ingreso del alumno

Tabla 5-28. Atributos de las Entidades del Datamart para el paquete EIS/DSS.

Nombre de la entidad	Clave	Descripción entidad
H Alumnos finales aprobados	ple_cod	Clave foránea de Plan del alumno
	mat cod	Clave foránea de Código de la materia
	alu id	Identificación del Alumno
	a mat anio	Año en que se rindió el final
	claverangonotas	Clave foránea de Notas por rango
H_Alumnos_finales_desaprobados	ple_cod	Clave foránea de Plan del alumno
	mat cod	Clave foránea de Código de materia
	ma mat anio	Año en que se rindió el final
	alu id	Identificación del Alumno
	a mat notaf	Nota del final
H_Alumnos_ingresantes_egresado	ple_cod	Clave foránea de Plan del alumno
	F Ingreso	Fecha ingreso a la Universidad
	a car estado	Estado del alumno
	Fbaja	Fecha de baja a la Universidad
	cant alumnos	Cantidad de alumnos
H Alumnos Promedios	Ple_cod	Clave foránea de Plan del alumno
	mat cod	Clave foránea de Código de materia
	a mat anio	Año en que se rindió examen
	PromedioCursada	Promedio de curada por plan, materia y año
	PromedioFinal	Promedio de flnal por plan, materia y año
H_Materias_Analisis_Cursada	ple_cod	Clave foránea de Plan del alumno
	mat cod	Clave foránea de Código de materia
	ma mat anio	Año en que se cursó
	alu id	Identificación del Alumno
	a mat notac	Nota de cursada
	a mat notaf	Nota de final
H_Materias_Matriculados	ple_cod	Clave foránea de Plan del alumno
	mat cod	Clave foránea de Código de materia
	ma mat anio	Año de matriculación
	CuentaDealu id	Cantidad de alumnos por plan, materia y año de matriculación

Tabla 5-28. Atributos de las Entidades del Datamart para el paquete EIS/DSS.

A continuación, las figuras 5-18 a 5-27 muestran diagramas de datos pertenecientes a la modelización multidimensional para dar soporte a las necesidades de usuarios del paquete EIS/DSS

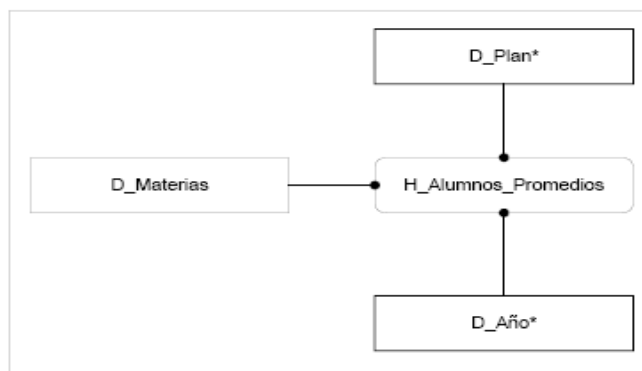


Figura 5-18. Diagrama estrella para el caso de uso Análisis de Promedios.

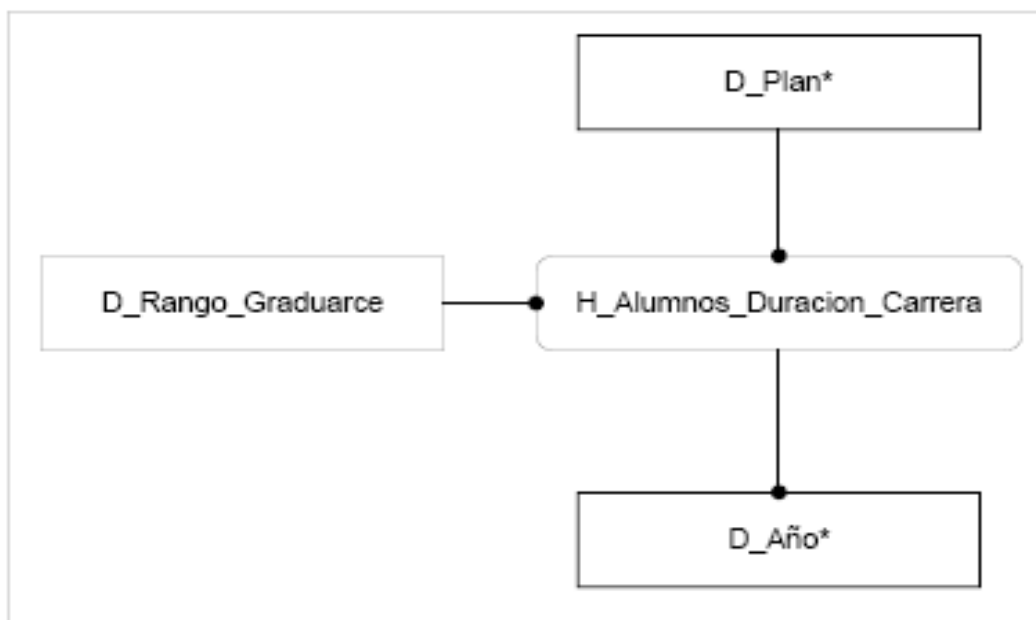


Figura 5-19. Diagrama estrella para el caso de uso Análisis Duración Carrera.

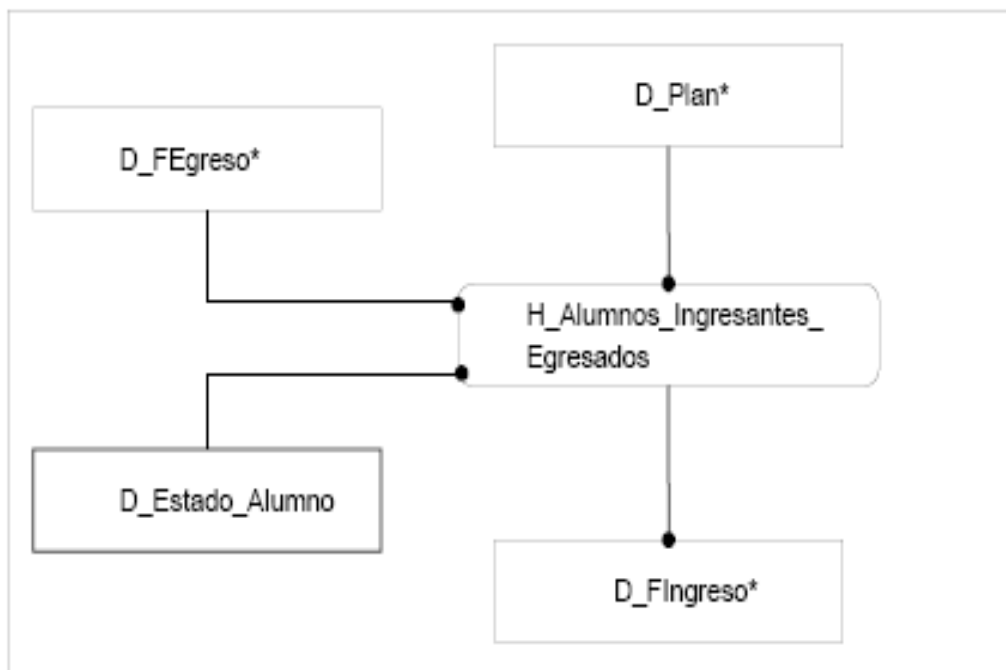


Figura 5-20. Diagrama estrella para el caso de uso Análisis Ingresantes/Egresados.

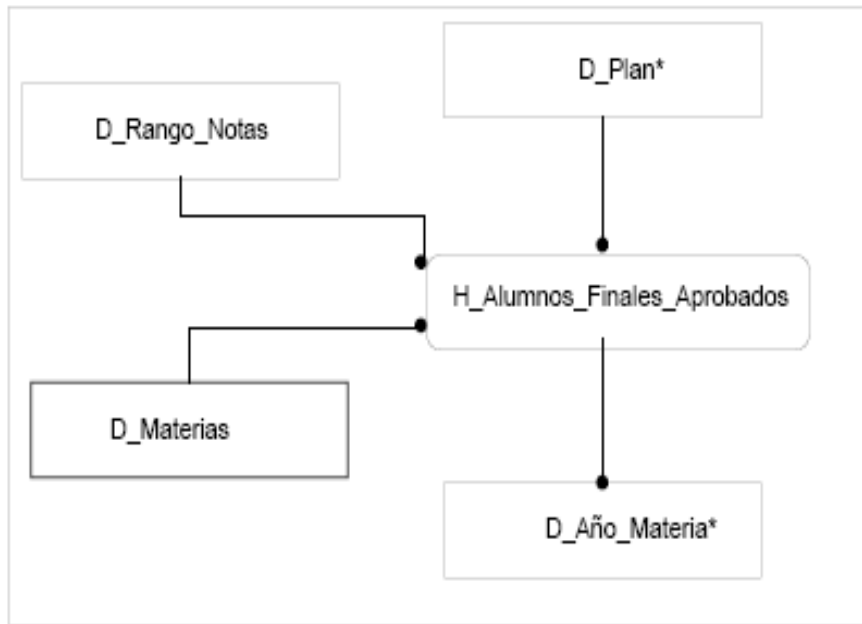


Figura 5-21. Primer diagrama estrella para el caso de uso Análisis de Calificaciones.

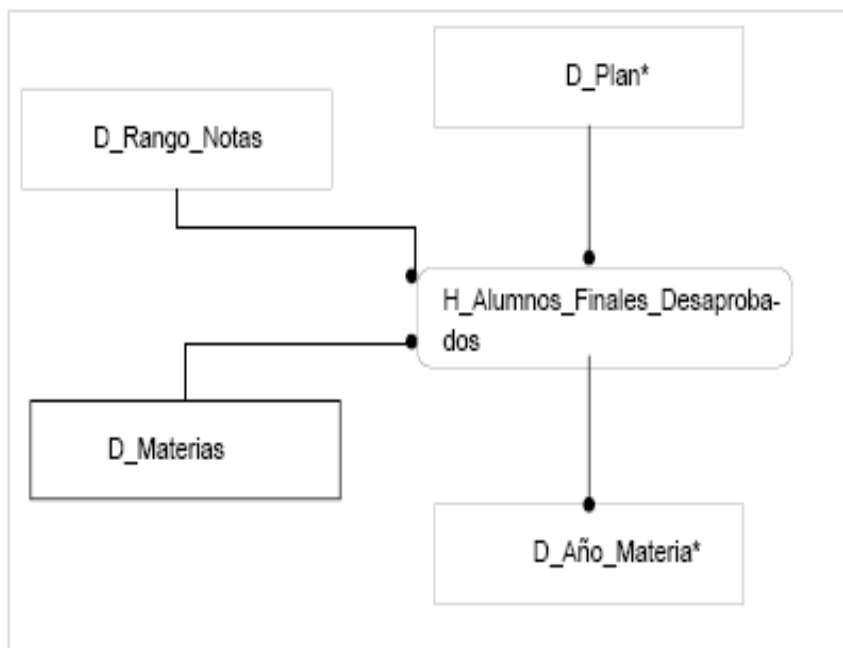


Figura 5-22. Segundo diagrama estrella para el caso de uso Análisis de Calificaciones.

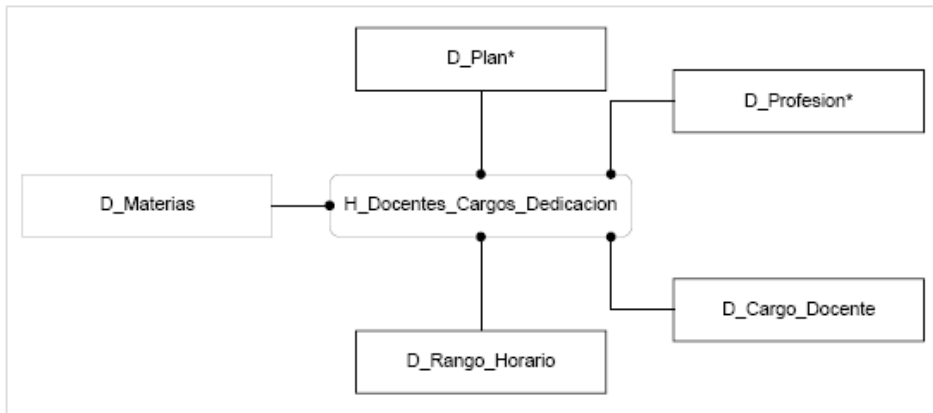


Figura 5-23. Diagrama estrella para el caso de uso Análisis de Docentes según Cargo y Dedicación.

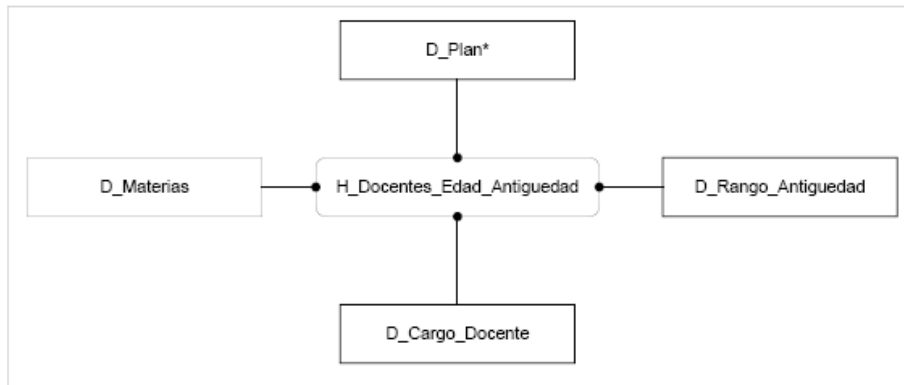


Figura 5-24. Diagrama estrella para el caso de uso Análisis de Docentes según Edad y Antigüedad.

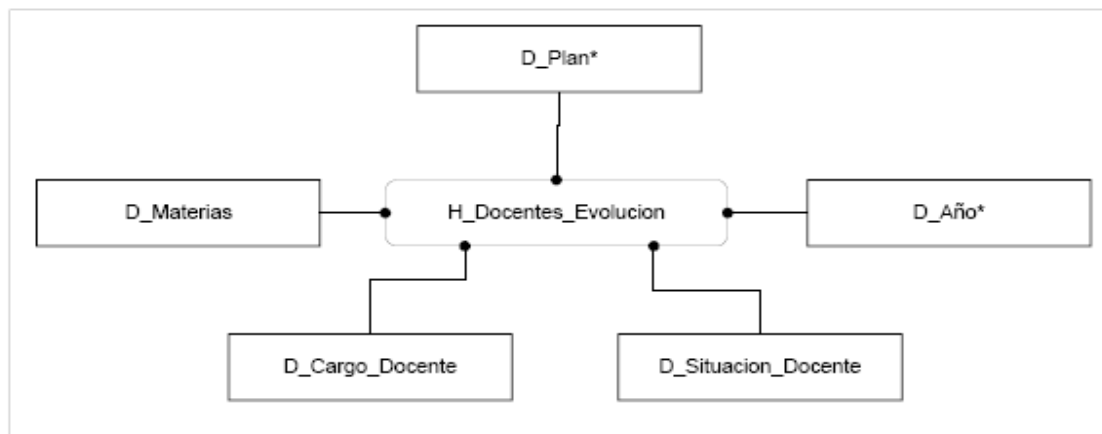


Figura 5-25. Diagrama estrella para el caso de uso Análisis de Docentes Evolución Estructura Personal/Variaciones del Cuerpo Docente.

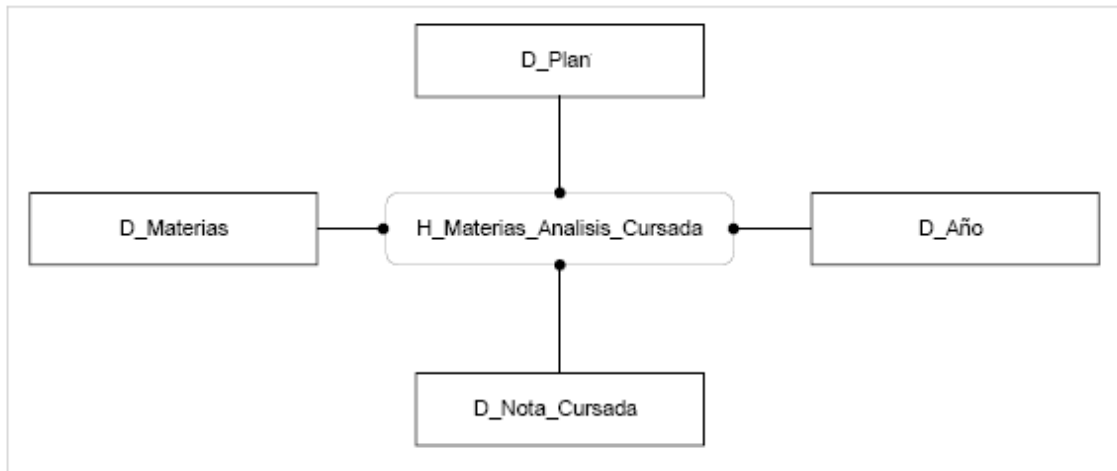


Figura 5-26. Diagrama estrella para el caso de uso Análisis de Notas por Materia Cursada.

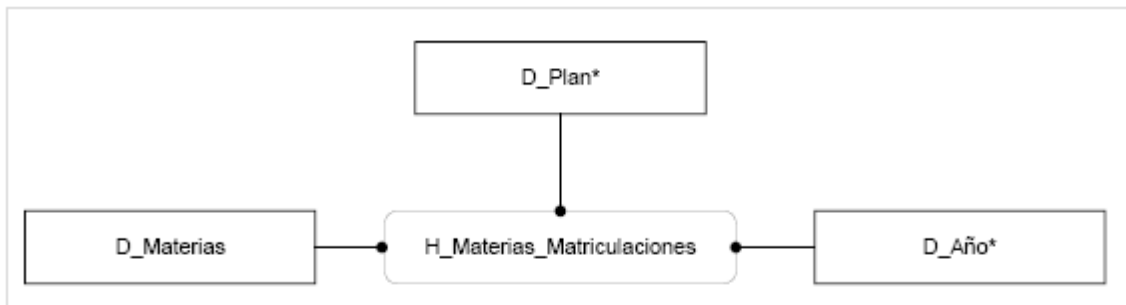


Figura 5-27. Diagrama estrella para el caso de uso Análisis de Alumnos Matriculados y Análisis de Alumnos por Materia

Las entidades con '*' no tienen tablas físicas de dimensiones, estas entidades son dimensiones virtuales, es decir, datos extraídos de las propias tablas de hechos. Sus atributos se corresponden a atributos de la tabla de hechos con la cual se relaciona.

5.4. Análisis del Enfoque de Construcción

Como el Datawarehouse está compuesto fundamentalmente por la base de datos Intermedia, se crea inicialmente el Datamart del Dpto. de Ingeniería y se propone crear varios Datamarts a partir de la base Intermedia si fuera necesario para futuras ampliaciones.

Para el enfoque de construcción se utiliza la propuesta discutida en el apartado capítulo 1, que se sustenta con una estructura de dos niveles, en donde se definen tanto el almacén empresarial (Intermedio/detalle) como los departamentales (Datamarts).

Esta decisión tiene algunas ventajas y desventajas:

- Una implementación de Datamarts sencillos permitirá al equipo de desarrollo seguir trabajando mientras los usuarios ya van realizando sus consultas. Se empieza con estos Datamarts porque minimiza el riesgo de implementaciones demasiado ambiciosas.
- Menos tiempo de desarrollo y para comenzar a utilizarlo por los usuarios.
- Menos recursos para el desarrollo y definición de requisitos.
- La integración posterior de varios Datamarts se complica, pero se minimiza esta desventaja si:
 - Se diseña la Base de Datos de cada Datamarts sin perder de vista el objetivo final de integración. Cada Datamarts se planifica por separado pero que compartan aquellas definiciones comunes a la universidad
 - Se decidió definir una estructura lógica común sobre la que se basan todos los Datamarts:
 - Se crea una arquitectura que define el marco de la organización. De manera que los Datamarts se construyen dentro de ese marco común.
 - Se definieron dimensiones conformadas. Es decir, una dimensión que significa lo mismo para cada posible tabla de hechos con la que se pueda unir.

Estas decisiones hacen posible que:

- Una única tabla de dimensión se puede usar contra múltiples tablas de hechos en el mismo espacio de bases de datos
- Las interfaces de usuario y el contenido de los datos son consistentes dondequiera que se use
- Hay una interpretación consistente de atributos

Se realizaron definiciones estándar de Hechos. Al igual que las dimensiones conformadas, se tiene en cuenta:

- Mismas unidades de medida
- Mismos periodos
- Mismas localizaciones.

Finalmente, a efectos de mantener el sistema lo más flexible posible, se decidió enfocar la construcción del Datawarehouse con una arquitectura abierta a cambios de requerimientos y cambios en las herramientas de acceso a datos.

- **Perfil de Crecimiento y Evolución del Datawarehouse**

Con el uso del SUFESC se puede implementar diferentes combinaciones de esquemas, como las siguientes:

- Esquema en Estrella: Una tabla de hechos en el centro conectada con un conjunto de tablas de dimensiones.
- Esquema Copo de Nieve: Un refinamiento del anterior donde algunas tablas se normalizan en tablas más pequeñas.
- Constelación de Hechos: Múltiples tablas de hechos comparten tablas de dimensión que se visualizan como una colección de hechos.

Además, al disponer de los datos completos de detalle en el Datawarehouse se pueden utilizar herramientas de Inteligencia de Negocios que acceden a los datos de detalle y crean sus propias bases de datos multidimensionales con estructuras y esquemas propietarios.

También se permite la creación de tablas de resumen y cálculos especiales en los casos que sea necesario para satisfacer algún tipo de requerimiento que surja cuando los usuarios empiecen a utilizar el sistema y este se extienda por el resto de las áreas o departamentos de la Universidad.

Las dimensiones conformadas y las definiciones estándar de los hechos configuran la arquitectura del Datawarehouse. Una definición de este tipo permite añadir un nuevo Datamart que puede coexistir con los ya existentes.

Se debe tener en cuenta que el sistema además de proveer información para la toma de decisiones a los usuarios puede ser utilizado para la investigación de Inteligencia de Negocios.

- **Paquete Administración y Mantenimiento**

Los módulos necesarios para la administración y mantenimiento del Datawarehouse, Datamart y su explotación son:

- Gestión de Extracción y Carga
- Gestión del Datawarehouse
- Gestión de Consultas - Acceso a los Datos.

La figura 5-28 muestra los usuarios y los diferentes módulos del sistema que son necesarios para la administración del SUFESC.

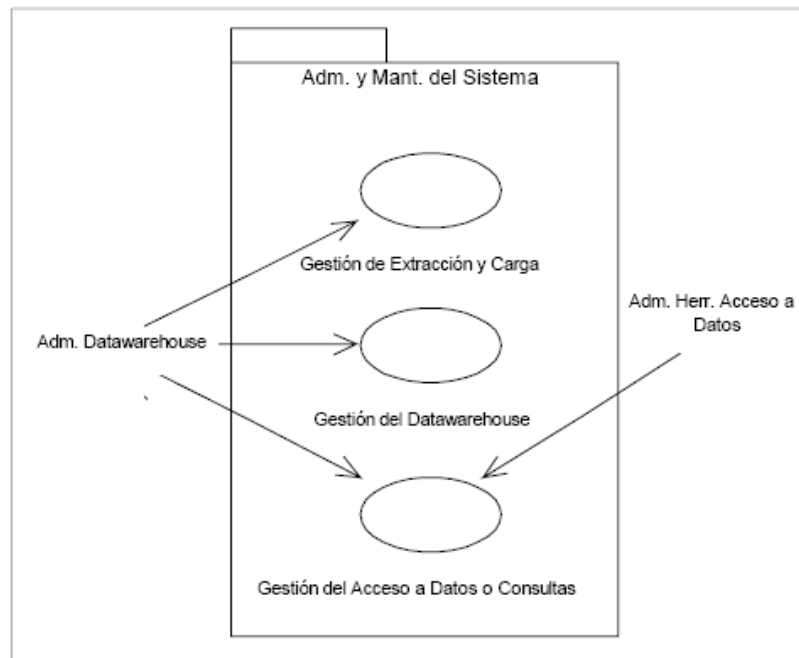


Figura 5-28. Paquete Administración y Mantenimiento.

En las tablas 5-29, 5-30, 5-31 se especifican los casos de usos que componen el paquete Administración y Mantenimiento del Sistema.

La figura 5-28 muestra un diagrama de secuencia donde se detallan los diferentes procesos de la extracción, transformación y carga de los datos desde la base de datos del sistema fuente a la base de datos del Datawarehouse/Datamart.

La descripción de la figura 5-28 es:

Condición de Inicio: El proceso comienza cuando el servicio DTS detecta que se ha cumplido la condición de inicio de los procesos. La condición de inicio se programa indicando con qué frecuencia se debe ejecutar. Diariamente se verifica que la condición se cumpla de acuerdo a su programación temporal, en caso de no cumplirse no se ejecuta el proceso siguiente.

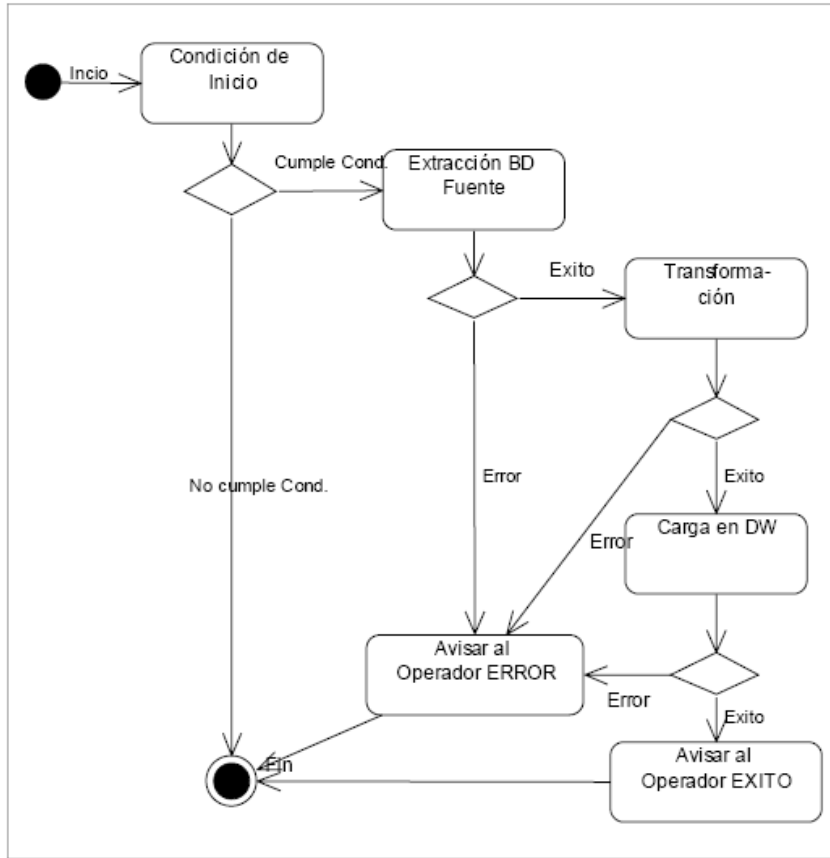
Extracción BD Fuente: Una vez que *Condición de Inicio* se cumple se ejecuta la extracción de los datos de la base de datos fuente. La extracción de los datos puede ser desde la base de sistema académico o desde la base de detalle dentro del Datawarehouse. En caso que en el proceso de extracción surja un error el proceso lo detecta e informa al operador. Este informe es vía mensaje a la consola de principal de MS SQL SERVER o vía e-mail a la casilla del correo del operador. En caso de error el proceso se detiene, en caso de éxito se ejecuta el proceso *Transformación*.

Transformación: Este proceso realiza las transformaciones necesarias de los datos (en caso que sea necesario) antes de introducirlos a la base de datos destino. En caso de error, se informa al operador de la misma manera que en el proceso anterior y se detiene la ejecución. En caso de éxito se ejecuta el proceso *Carga en DW*.

Carga en DW: Una vez extraídos y transformados, los datos son introducidos a la base de datos destino. Esta es la base de datos de Detalle y el Datamart. El manejo del error es igual que en los dos procesos anteriores. En caso de ejecución satisfactoria se informa al operador que el proceso terminó correctamente.

Avisar al Operador ÉXITO: Este proceso informa al operador por medio de un mensaje a la consola o vía e-mail a su casilla de correo electrónico que el proceso completo terminó satisfactoriamente.

Avisar al Operador ERROR: Este proceso informa al operador por medio de un mensaje a la consola o vía e-mail a su casilla de correo electrónico que en el proceso se produjo un error. Se informa el texto de error y el proceso en cuestión se detiene para que el operador resuelva el error.



Caso de uso	Gestión de Extracción y Carga
Requerimiento que implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"> • Es el módulo que contiene las facilidades necesarias para construcción, y mantenimiento de la extracción y carga de datos en el Datawarehouse y dentro del Datawarehouse. El módulo se implementa por medio de las herramientas que dan soporte el Datawarehouse. En este caso se utiliza la herramienta DTS (Servicios de Transformación de Datos), que pertenece a MS SQL SERVER. • Con DTS se importa, exporta y transforman los datos entre la base de datos fuente y el Datawarehouse. También desde la BD Detalle y el Datamart. Estos procesos se automatizan de manera que se ejecutan sin intervención de los usuarios. • Los procesos que se hacen referencia son: Condición de Inicio, Extracción de BD Fuente, Transformación, Carga en DW, Aviso al Operador de Error y Aviso al Operador de Éxito. • Los usuarios responsables de utilizar este módulo son los usuarios "Adm. Datawarehouse" y su principal tarea es la Administración y Mantenimiento de los procesos de extracción, transformación y carga. Los procesos de extracción, transformación y carga se ejecutan de forma automática por medio de un usuario/proceso del sistema. • Los usuarios del módulo son "Adm. Datawarehouse" y tienen permisos de Administrador (permisos totales sobre de las bases de datos de Detalle y Datamart) y de selección en la base
Figuras asociadas	Figura 6-18. Procesos de Extracción, Transformación y Carga
Observaciones	N/A

Tabla 5-29. Caso de Uso Gestión de Extracción y Carga.

Caso de uso	Gestión del Warehouse
Requerimiento que implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"> • Este módulo tiene funciones de administración y mantenimiento de las bases de datos que componen el Datawarehouse y el Datamart. Permite controlar el funcionamiento diario de estas bases de datos. <p>Entre sus tareas se encuentran las siguientes:</p> <ul style="list-style-type: none"> - Mantenimiento de tablas, vistas e índices. - Mantenimiento de la integridad entre los objetos de las bases de datos. - Mantenimiento de las agregaciones y actualización de las existentes. - Acumulación de los datos del Datawarehouse y Datamart para su salvaguarda. - Procesos automáticos de backups. - Administración de la seguridad y perfiles de usuarios. - Mantener el Metadatos del Datawarehouse y del Datamart. <ul style="list-style-type: none"> • La Gestión del warehouse se realiza por medio de las herramientas que provee MS SQL SERVER para realizar las tareas diarias administración y mantenimiento. • Los usuarios del este módulo son "Adm. Datawarehouse" y tienen permisos de Administrador (permisos totales sobre de las bases de datos de Detalle y Datamart). Cuando se realizan procesos automáticos de mantenimiento como por ejemplo Backups se utiliza un usuario/proceso del sistema.
Figuras asociadas	N/A
Observaciones	N/A

Tabla 5-30. Caso de Uso Gestión del Warehouse.

	Gestión de Consultas - Acceso a Datos
Requerimiento que implementa	Módulo necesario para el funcionamiento del sistema.
Descripción	<ul style="list-style-type: none"> • Permite llevar toda la lógica necesaria para apoyar el proceso de gestión de consultas y análisis de información. Implementa por medio de las herramientas OLAP Services y Excel las necesidades de los usuarios. <p>Tareas más significativas son: - Dirigir las consultas a las tablas adecuadas. - Creación y mantenimiento de Cubos multidimensionales. - Realizar los reportes y gráficos requeridos por los usuarios. - Planificación de las consultas junto con los usuarios. - Apoyo a los usuarios en modelos de análisis de la información.</p> <ul style="list-style-type: none"> • Los usuarios de este módulo son los denominados Adm. Herr. Accesos a Datos, y sus permisos de accesos en las bases de datos están restringidos a solo consultas tanto en la base de datos de Detalle como en el Datamart.
Observaciones	N/A

Tabla 5-31. Caso de Uso Gestión de Consultas - Acceso a Datos.

5.5. Verificación del Análisis

El objetivo de esta actividad es garantizar que se han tenido en cuenta todos los requerimientos especificados por el usuario como otros requerimientos necesarios para la administración y operación del Datawarehouse. También se verifica la calidad de las especificaciones de análisis y la viabilidad del mismo antes de avanzar con el diseño del sistema.

Para cumplir dicho objetivo, se llevan a cabo las siguientes tareas:

- Verificación de la calidad técnica de cada especificación
 - Aseguramiento de especificación de requisitos en el análisis.
- **Verificación de la Calidad Técnica de Cada Especificación**
El objetivo de esta tarea es asegurar la calidad formal de las distintas especificaciones conforme a la técnica seguida para su elaboración. Además, se verifica que no se haya dejado de lado ninguna especificación que pueda afectar la etapa posterior de diseño del sistema.

La tabla 5-32 muestra la verificación de las especificaciones de análisis.

Especificación de análisis	Técnica	Requerimiento	Verificado
Modelo de Datos Fuentes	Modelado de Datos	Todos los requisitos de usuario.	Si
Submodelo Actas	Modelado de Datos	Ver ¿Qué es un EIS? requisito nro. 7,8,9,10,11,12,13 y 14.	Si
Submodelo Alumnos	Modelado de Datos	Ver ¿Qué es un EIS? requisito nro. 7,8,9,10 y 11.	Si
Submodelo Docentes	Modelado de Datos	Ver ¿Qué es un EIS? requisito nro. 1, 2, 3,4,5,6	Si
Submodelo Encuestas	Modelado de Datos	Ver apartado 5.3.1.1. requisito nro. 15	Si
Submodelo Materias	Modelado de Datos	Ver ¿Qué es un EIS? requisito nro. 12,13 y 14.	Si
Submodelo Relaciones Institucionales	Modelado de Datos	Ver ¿Qué es un EIS? requisito nro. 16	Si
Diagrama de Contexto	Diagrama de Paquetes	Todos los requisitos.	Si
Usuarios	Diagrama de Caso de Uso	Todos los requisitos.	Si
Paquete EIS/DSS	Diagrama de Paquetes	Todos los requisitos de usuario.	Si
Paquete de Análisis de Docentes	Diagrama de Paquetes/Diagrama de Caso de Uso	Ver ¿Qué es un EIS? Requisito nro. 1, 2, 3,4,5,6	Si

Especificación de análisis	Técnica	Requerimiento	Verificado
Análisis docentes según cargo y dedicación.	Especificación de Caso de Uso	Ver ¿Qué es un EIS? Requisito nro. 1, 2, y 3.	Si
Análisis variaciones del cuerpo docente/ evolución de la estructura de personal	Especificación de Caso de Uso	Ver ¿Qué es un EIS? Requisito nro. 4 y 6	Si
Análisis de docentes según edad y antigüedad	Especificación de Caso de Uso	Ver apartado 5.3.1.1. requisito nro. 5	Si
Paquete de Análisis de Alumnos	Diagrama de Paquetes/Diagrama de Caso de Uso	Ver apartado 5.3.1.1. requisito nro. 7,8,9,10 y 11.	Si
Análisis de alumnos ingresantes/egresados	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 7	Si
Análisis de alumnos ingresantes/becados	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 8	Si
Análisis de calificaciones	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 8	Si
Análisis de duración de carrera	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 10	Si
Análisis de promedios	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 9.	Si
Análisis de graduados	Especificación de Caso de Uso	Ver ¿Qué es un EIS?. requisito nro. 11	Si
Análisis de biblioteca	Especificación de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 8	Si
Paquete de Análisis de Materias	Diagrama de Paquetes/Diagrama de Caso de Uso	Ver ¿Qué es un EIS? requisito nro. 12,13 y 14.	Si

Tabla 5-32. Verificación de especificaciones.

Aseguramiento de Especificación de Requisitos en el Análisis

El objetivo de esta tarea es validar los distintos modelos con los requisitos especificados para el sistema de información, tanto a través del catálogo de requisitos, mediante la traza de requisitos, como a través de la validación directa del usuario.

La tabla 5-33 muestra la verificación de los requerimientos.

Caso de uso	Requerimiento	Verificado/ Especificado
Análisis docentes según cargo y dedicación.	Ver ¿Qué es un EIS? requisito nro. 1, 2, y 3.	Si
Análisis variaciones del cuerpo docente	Ver ¿Qué es un EIS? requisito nro. 4	Si
Análisis de Docentes por edad y antigüedad	Ver ¿Qué es un EIS? requisito nro. 5	Si

Tabla 6-33. Verificación de Requerimientos

Conclusiones

El concepto de DWH está teniendo una gran aplicación en la actualidad para el desarrollo de las Instituciones, como almacén de datos.

Los objetivos son una mayor velocidad de respuesta frente a las consultas de los usuarios. Estos podrán analizar y realizar preguntas sobre años, más que sobre meses de información.

Entre las posibles ampliaciones de la funcionalidad se propone incorporar a todos los departamentos, todas las carreras y posgrados en las actuales funcionalidades del sistema.

Luego se podrá incorporar las diferentes áreas de la universidad junto con la ampliación a los diferentes sistemas transaccionales, ya que actualmente se toma como sistema fuente el Sistema Académico.

A efectos de incorporar esta ampliación será necesario tener en cuenta la integración de todos los sistemas transaccionales, ya que determinada información que disponen algunas áreas y departamentos no está integrada con los principales sistemas de la universidad.

Entre las posibles ampliaciones de la funcionalidad se propone incorporar a todos los departamentos, todas las carreras y posgrados en las actuales funcionalidades del sistema.

BIBLIOGRAFIA

[Bitam, 2002]

."Business Intelligence". [en línea]<<http://www.bitam.com/spanish/AcercaDeBI.htm>
[Consulta: Febrero de 2002].

[Booch et al., 1999]

Booch, G.; Rumbaugh, J; Jacobson, I. 1999. "El Lenguaje Unificado de Modelado". Addison Wesley. ISBN: 84-79-028-1.

[Cabena et al., 1998]

Cabena, Peter; Hadjinian, Pablo; Stadler, Rolf; Verhees, Jaap; Zanasi, Alessandro. 1998. "Discovering Data Mining: From Concept to Implementation". Prentice Hall. ISBN: 0-13-743980-6.

[Cognos, 2002]

Cognos Corp. "Cognos Enterprise Tools". [en línea]. <<http://www.cognos.com>> [Consulta: Julio de 2002].

[Csi, 2002] Csi, I. "Development Datawarehousing". [en línea]. <<http://www.datwarehousecenter.com>> [Consulta: Julio de 2002].

[Groth, 1998]

Groth, Robert. 1998. "Data Mining: A Hands On Approach for Business Professionals". Prentice Hall. ISBN: 0-13-756412-0.

[Gupta, 2002]

Gupta, V. "An introduction to Data Warehousing" [en línea]. <<http://www.system-services.com/dwintro.htm>> [Consulta: Febrero de 2002].

[Kaplan & Norton, 1996]

Kaplan, Robert S.; Norton, David P. 1996. "Translating Strategy Into Action: The Balanced Scorecard". Harvard Business School Press. ISBN: 0-87584-651-3.

[Microsoft, 2002]

Microsoft Corp. "MS SQLServer 7.0". [en línea]. <<http://www.microsoft.com>> [Consulta: Julio de 2002].

[Oracle, 2002] Oracle Corp. "BI Solutions", [en línea].<<http://www.oracle.com>> [Consulta: Julio de 2002].

[Pence & Creeth, 2002]

Pence, N, Creeth, R. "An Introduction to OLAP". [en línea].<<http://www.OLAPreport.com/>> [Consulta:Febrero de 2002].

[Poe et al., 1998]

Poe, Vidette; Klauer, Patricia; Brobst, Stephen 1998. "Building a Data Warehouse for Decision Support". Prentice Hall. ISBN: 0-13-769639-9.

[Sanchez, 2000]

Sanchez Capuchino, A. M^a M. "Estimación de Proyectos Software". [en línea]. UPM-Universidad Politécnica de Madrid. 2000.

<<http://www.ls.fi.upm.es/udis/miembros/amoreno/cocomoi.pdf>> [Consulta: Mayo de 2002].

[SCSI, 2000]

Secretaría del Consejo Superior de Informática para el Impulso de la Administración Electrónica [En línea] <www.csi.map.es/csi/metrica3/calidad.pdf> [Consulta: Julio de 2002]

[Sperley, 1999]

Sperley, Eric. 1999. "The Enterprise Data Warehouse: Planning, Building, and Implementation". Prentice Hall. ISBN: 0-13-905841-1.

[Synera, 2002]

Synera System Corp. "Synera System". [En línea] <<http://www.synerasystem.com>> [Consulta: Julio de 2002].

Fayyad, U.M.; Grinstein, G.; Wierse, A. "Information visualization in Data Mining and Knowledge Discovery" Morgan Kaufmann, Harcourt Intl., 2001.

Han, J.; Kamber, M. "Data Mining: Concepts and Techniques" Morgan Kaufmann, 2001.

Hand, D.J.; Mannila, H. and Smyth, P. "Principles of Data Mining", The MIT Press, 2000.

Hernández, J.; Ramírez, M.J.; Ferri, C. "Introducción a la Minería de Datos" Pearson Prentice Hall, 2004

Kosala, R.; Blockeel, H. "Web Mining Research: A Survey" ACM SIGKDD Explorations, Newsletter of the ACM SIG on Knowledge Discovery and Data Mining, June 2000, Vol. 2, n°1, pp. 1-15.

Mena, Jesus "Data Mining Your Website", Digital Press, July 1999 • Mitchell, T.M. "Machine Learning" McGraw-Hill 1997.

Pyle, D. "Data Preparation for Data Mining" Morgan Kaufmann, Harcourt Intl., 1999.

Thuraisingham, B. "Data Mining. Technologies, Techniques, Tools, and trends", CRC Press, 1999.

Witten, I.H.; Frank, E. "Tools for Data Mining", Morgan Kaufmann, 1999.

Wong, P. C. "Visual Data Mining", Special Issue of *IEEE Computer Graphics and Applications*, Sep/ Oct 1999, pp. 20- 46.

Corey, Michael J., Abbey, Michael (1997). Oracle DataWarehousing, Oracle Press, Osborne/Mc Graw-Hill.

Dresner, Howard (1993). "Business Inteligence: Competing Against Time". The Gartner Group, Mayo.

Dresner, Howard. "Multidimensionality: Ready or not, here it comes". The Gartner Group.

Codd, E.F.; Codd, S.B.; Salley, C.T. (1993). "Beyond Decision Support". Computerworld, N° 30. Julio.

Hurwitz, Judith S. (1996). "Modelling Complex Multi-Dimensional Information", Diciembre.

Oracle Magazine (1998). "Building the Right Data Mart". Marzo/Abril.

Pech, Iván; Estañol, Abelardo; Tirado, Leticia (1997). "Data Warehouse para Administración de Espacios de Bases de Datos". Soluciones Avanzadas, N°. 41. <http://hp.fciencias.unam.mx/revista/soluciones/SA41/dw-41.html>

Pech Escalante, Iván (ivanpech@datateam.com.mx) (1996). "Implementando el Data Warehouse ¿manejador de base de datos relacional o multidimensional?". Soluciones Avanzadas, N°. 36.

<http://hp.fciencias.unam.mx/revista/soluciones/30s/No36/dw-1.html>

The Gartner Group (1990). "The Trouble with SQL". Office Information Systems. Agosto.

Jhon Wiley W.H. Inmon and Sons. Building the Data Warehouse. USA 1996.

IBM Press 2001. IBM DAB2 UDB Business Intelligence Tutorial.

Páginas web

<http://www.oracle.com/datawarehouse/index.html>

<http://www.oracle.com/olap/html/oes.html>(OracleExpressServer)

<http://www.businessobjects.com>

<http://sie.efpol.ua.es/webolap/pagolap.htm>

<http://warehouse.chimenet.org/software/eisdss/NineRulesforEvaluatingOLAP2.htm>

<http://www.datawarehouse.com>

<http://www.sun.es/success/warehouse>

<http://www.data-warehouse.com/issues/tools.htm>

<http://www.intraware.com/ms/itwr/askjms/olap.html>

<http://www.ntu.edu/1/atmp/mc96120504.htm>

Datamining

<http://www.datamining.com>

<http://research.microsoft.com/research/datamine>

INDICE

1. PLANTEAMIENTO DEL PROBLEMA	i
Objetivos de la investigación.....	ii
Hipótesis.....	iii
INTRODUCCIÓN	1
CAPITULO 1 MARCO TEORICO.....	3
1.1. Necesidad de Información y Conocimiento en la Empresa	3
1.2. Introducción a Almacenes de Datos y Datamart.....	11
1.4 Implementación del Almacenes de datos	16
1.5 Costos e Impacto de un AD	18
1.6 Impactos del Almacenes de Datos	20
1.7 Valor del AD para la Toma de Decisiones.....	23
1.8 Introducción al Procesamiento Analítico en Línea	24
1.9. Introducción a la Minería de datos	36
1.10. Arquitectura Data Mining	39
1.11. Inteligencia de Negocios	43
1.12. Herramientas de Reportes	59
CAPITULO 2. PLAN DEL PROYECTO.....	65
2.1. Propuesta de Desarrollo.....	69
2.2. Gestión de la configuración.....	82
2.3. Directivas para el Control de Cambios	88
Capitulo 3 GESTIÓN DE LA CALIDAD.....	91
3.1 Alcance del Plan de Aseguramiento de Calidad	92
3.4. Procedimiento de Corrección de Errores	101
3.5. Métodos para la Salvaguarda y Mantenimiento de la Documentación Obtenida en las Actividades de Aseguramiento de Calidad	106
4. ESTUDIO DE VIABILIDAD DEL SISTEMA.....	107
4-1. Requisitos y Alcances Generales	109
4.2. Contexto del Sistema.....	110
4.3. Estudio de la Situación Actual.....	112
4.4. Diagnóstico de la Situación Actual	116
4.4. Estudio de Alternativas de Solución	124
4.5 Definición de Requisitos del Sistema	119
4.5. Requerimientos de HARWARE y SOFTWARE:	135
4.6 .Herramientas de Desarrollo	146
4.7. Selección de la Solución	151
Capitulo 5. ANÁLISIS DEL SISTEMA.....	155
5.1. Modelo de Datos Fuente.....	155
5.1. Submodelos	156
5.2. Análisis de Requerimientos	161
5.2.2. Descripción de los Sistemas de Información Existentes	112
5.3 Análisis del Datawarehouse.....	176
5.4. Análisis del Enfoque de Construcción.....	186
5.4. Modelo de datos de la Base de Datos Multidimensional.....	180
5.4.2.5. Excel 2007 Productos	148
5.5. Verificación del Análisis.....	193
Conclusiones	195

BIBLIOGRAFIA 196