



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**DESARROLLO DE UN *DATAMART* DE INFORMACIÓN
ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS
Y SISTEMAS DE LA FACULTAD DE INGENIERÍA DE LA USAC**

**Mario Roberto Reyes Marroquín
Pablo Augusto Rosales Tejada**

Asesorado por el Ing. Jorge Armín Mazariegos Rabanales

Guatemala, agosto de 2007

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DESARROLLO DE UN *DATAMART* DE INFORMACIÓN
ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS
Y SISTEMAS DE LA FACULTAD DE INGENIERÍA DE LA USAC**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR:

MARIO ROBERTO REYES MARROQUÍN

PABLO AUGUSTO ROSALES TEJADA

ASESORADO POR EL ING. JORGE ARMÍN MAZARIEGOS

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, AGOSTO DE 2007

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Inga. Glenda Patricia García Soria
VOCAL II	Inga. Alba Maritza Guerrero de López
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

Mario Roberto Reyes Marroquín

DECANO	Ing. Julio González Podszueck
EXAMINADOR	Ing. Francisco Javier Guevara Castillo
EXAMINADOR	Ing. Raymond Usher Vargas
EXAMINADOR	Ing. Carlos Alfredo Azurdia Morales
SECRETARIO	Ing. Francisco Javier González López

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Inga. Glenda Patricia García Soria
VOCAL II	Inga. Alba Maritza Guerrero de López
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

Pablo Augusto Rosales Tejada

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Marlon Antonio Pérez Türk
EXAMINADOR	Ing. César Augusto Fernández Cáceres
EXAMINADOR	Ing. Edgar Estuardo Santos Sutuj
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presentamos a su consideración nuestro trabajo de graduación titulado:

DESARROLLO DE UN *DATAMART* DE INFORMACIÓN ACADÉMICA DE ESTUDIANTES DE LA ESCUELA DE CIENCIAS Y SISTEMAS DE LA FACULTAD DE INGENIERÍA DE LA USAC,

tema que nos fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, en agosto de 2006.

Mario Roberto Reyes Marroquín

Pablo Augusto Rosales Tejada

AGRADECIMIENTOS A:

Mario Roberto Reyes Marroquín

DIOS

Por darme la oportunidad de culminar esta etapa de mi vida. Al igual le doy gracias a la Virgen María y a San Juan Bosco, por ser la luz que intercede por mí ante nuestro Señor Jesús.

MIS PADRES

Por todo su amor y apoyo durante todas las etapas de mi vida, y sobre todo por el apoyo para lograr culminar esta carrera.

LA ESCUELA DE CIENCIAS Y SISTEMAS

Especialmente a los ingenieros Jorge Armín Mazariegos, Floriza Ávila y Marlón Pérez Turk, por su apoyo incondicional en la realización del presente trabajo.

AGRADECIMIENTOS A:

Pablo Augusto Rosales Tejada

MIS PADRES

Por su incondicional apoyo y estímulo, así como su valioso ejemplo.

MI FAMILIA

Por su apoyo

LA UNIVERSIDAD DE SAN CARLOS

Por su importante contribución en el desarrollo de Guatemala

EL CUERPO DOCENTE DE LA ESCUELA DE CIENCIAS Y SISTEMAS

Por compartir su conocimiento y experiencia.

LOS ASESORES

Ing. Armín Mazariegos, Inga. Floriza Ávila e Ing. Marlon Pérez, por la supervisión de este trabajo.

LOS COMPAÑEROS DE ESTUDIO

Por su amistad

DATUM

Por apoyarme en el logro de este objetivo

DEDICATORIA A:

Mario Roberto Reyes Marroquín

MIS PADRES

Por su amor, dedicación, honradez, responsabilidad, y ser siempre el buen ejemplo que deseo seguir.

MI ESPOSA

Por su inmenso amor y el apoyo incondicional en cada momento de mi vida.

MIS HIJOS

Esteban, Sebastián y Natalia, quienes son mi inspiración y orgullo.

MI HERMANO

Sergio, por su amor, ejemplo de trabajo y de ser un buen padre.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	V
GLOSARIO	VII
RESUMEN	XI
OBJETIVOS	XIII
INTRODUCCIÓN	XV
1. INTRODUCCIÓN AL PROBLEMA Y SU SOLUCIÓN	1
1.1 Definición del problema	1
1.2. Objetivos del trabajo	1
1.3. Aproximación de la solución	2
1.4. Organización del documento	3
2. INTELIGENCIA DE NEGOCIO, DEFINICIONES Y FUNDAMENTOS	5
2.1. Introducción	5
2.1.1. Definición de inteligencia de negocio	5
2.1.2. Beneficios de la inteligencia de negocio	6
2.1.3. Información estratégica	7
2.1.4. Información táctica	8
2.1.5. Información técnico-operacional	8
2.2 Objetivo de la información	9
2.2.1 Usuarios	9

2.2.2 Tipos de preguntas	9
2.2.3 Cantidad de datos	10
2.3. Introducción a <i>data warehouse</i> y <i>datamart</i>	10
2.3.1. Arquitectura <i>data warehouse</i>	15
2.3.2. Implementación del <i>data warehouse</i>	18
2.3.3. Valor del DW para la toma de decisiones	18
2.3.4 Componentes básicos	19
2.3.5 Representación del <i>data warehouse</i>	25
2.3.5.1 Esquema en estrella	25
2.3.6 Extracción, transformación y carga del <i>data warehouse</i>	26
3. MODELO DE DATOS LÓGICO	29
3.1 Especificación del modelo dimensional	29
3.1.1 Modelo estrella para asignaciones	30
3.1.2 Modelo estrella para análisis de estudiantes graduados	32
3.1.3 Modelo estrella para inscripciones	34
3.1.4 Modelo estrella para análisis de uso de salones	36
3.1.5 Dimensiones del modelo de datos lógico	37
3.1.5.1 Dimensión estudiante	37
3.1.5.2 Dimensión curso	38
3.1.5.3 Dimensión carrera	39
3.1.5.4 Dimensión ciclo académico	40
3.1.5.5 Dimensión catedrático	40
3.1.5.6 Dimensión salón	41
3.1.5.7 Dimensión sección	41
3.1.5.8 Dimensión año	41
3.1.5.9 Dimensión escuela	42
3.1.5.10 Dimensión demografía	42

3.1.5.11 Dimensión resultado	43
3.1.5.12 Dimensión horario	44
4. MODELO DE DATOS FÍSICO Y PRESENTACIÓN DE INFORMACIÓN	45
4.1 Correspondencia entre el modelo dimensional y el modelo físico de base de datos	45
4.2 Plataforma de la solución	47
4.3 Extracción, transformación y carga de datos (ETL)	50
4.3.1 Flujo de datos	51
4.3.2 Procesos de carga de las dimensiones	53
4.3.3 Registro de la actividad de un proceso de carga	54
4.3.4 Procedimiento de carga del <i>data warehouse</i>	55
4.3.4.1 Carga inicial de dimensiones estáticas	55
4.3.4.2 Carga de dimensiones y tablas de hecho	56
4.3.4.3 Rutas para los archivos de configuración y de carga de datos de DWACADÉMICO	60
4.4 Esquema de seguridad	61
4.4.1 Seguridad de la base de datos	61
4.4.2 Seguridad de <i>Discoverer Administrator</i>	62
4.4.3 Seguridad de <i>Discoverer</i> usuario final	65
4.4.4 Manejo de reportes	65
4.5 Presentación y entrega de la información contenida en el <i>data mart</i>	65
CONCLUSIONES	77
RECOMENDACIONES	79

REFERENCIAS ELECTRÓNICAS	81
BIBLIOGRAFÍA	83

ÍNDICE DE ILUSTRACIONES

FIGURAS

1. Niveles organizacionales en los que colabora la inteligencia de negocio	7
2. Diferencias entre sistemas transaccionales y <i>Data Warehouse</i> .	13
3. Estructura básica de un <i>data warehouse</i>	17
4. Ejemplo de dimensiones	20
5. Ejemplo de dimensiones	22
6. Ejemplo de dimensiones	24
7. Proceso ETL (extracción, transformación y carga)	27
8. Modelo dimensional para Asignaciones de estudiantes y cursos (esquema en estrella)	30
9. Modelo dimensional para análisis de Graduados	32
10. Modelo dimensional para análisis de inscripciones de estudiantes	34
11. Modelo dimensional para análisis de Uso de Salones	36
12. Mapeo de dimensiones a tablas relacionales del esquema de inscripciones	45
13. Mapeo de dimensiones a tablas relacionales del esquema de graduados	46
14. Mapeo de dimensiones a tablas relacionales del esquema de uso de salones	46
15. Mapeo de dimensiones a tablas relacionales del <i>data mart</i> de asignaciones	47
16. Plataforma de inteligencia de negocios	49
17. Plataforma del servidor de aplicaciones	50
18. Proceso de extracción, transformación y carga hacia el <i>data mart</i> académico de la escuela de ciencias y sistemas (ETL)	52
19. Procedimientos para la carga de dimensiones	53
20. Procedimientos para la carga de tablas de hecho	54
21. Procedimientos para la carga del <i>data warehouse</i> completo	54
22. Parámetros de la bitácora	55
23. Manejo de seguridad sobre objetos de la base de datos	62
24. Manejo de seguridad en herramienta <i>Discoverer Administrator</i>	64
25. Componentes de la herramienta <i>Oracle Discoverer</i>	66

26. Reporte de cantidad de años para graduarse	67
27. Reporte de uso de salones	68
28. Reporte de asignación de cursos	69
29. Reporte de demografía de estudiantes	70
30. Reporte de historial de cursos por estudiante	71
31. Reporte de asignaciones por catedrático	72
32. Reporte de reprobados por curso	73
33. Reporte de <i>ranking</i> de estudiantes por curso	74
34. Reporte de cantidad de graduados	75
35. Reporte de estudiantes por curso y año de inscripción	76

GLOSARIO

Ámbito	Contorno de un espacio. Espacio entre límites. En programación, alcance que tenga la definición de una variable.
Cache	En términos de la ciencia de la computación, <i>cache</i> se refiere al proceso de duplicar los datos originales que se encuentran en una ubicación difícil de acceder -usualmente en términos de tiempo- de manera que puedan ser accedidos de una forma más rápida.
Commit	En términos de la ciencia de la computación y manejo de almacenamiento de datos, se refiere a la idea de hacer permanentes un conjunto de cambios alternativos, tal es el caso de el fin de una transacción. Ejecutar un <i>commit</i> es hacer los cambios temporales permanentes. <i>Commit</i> es un acto de compromiso.
Data Warehouse	Sistema de información centralizado que contiene toda la información relevante para una organización, y que permite de una forma ágil y flexible la consulta de

información. Está orientado a convertirse en la única fuente de información para todas las áreas del negocio y ser la fuente prioritaria para todas las herramientas de inteligencia de negocio que se desee utilizar en una organización.

Data Mart

Subconjunto del *Data Warehouse* que está orientado a un área específica del negocio. Todas sus métricas y dimensiones están relacionadas con un área de negocio en particular.

Dimensión

Se refiere a la entidad que se encarga de agrupar, calificar o catalogar cada uno de los hechos contenidos dentro de un *data mart*.

ETL

Proceso de extracción, transformación y carga de datos desde una fuente determinada hacia un *data mart* o el *data warehouse* corporativo.

Herencia

Forma de obtener o extender la funcionalidad de una clase existente en una nueva clase.

Host

Ordenador/computadora conectado(a) a la *Internet*

OLTP	Sistemas de procesamiento de transacciones en línea, o sistemas transaccionales, en los cuales residen las operaciones del día a día de cada negocio y que son la fuente prioritaria de datos para cada <i>data mart</i> o el <i>data warehouse</i> corporativo.
OLAP	Sistemas de procesamiento analítico en línea, es decir, sistemas orientados al análisis de información, basados en un <i>data mart</i> o <i>data warehouse</i> .
SQL	Lenguaje de computación utilizado para crear, modificar y recuperar datos de un sistema de base de datos relacional.
Tabla de Hechos	Tabla de datos central que contiene cada una de las métricas que se desea medir dentro de un área específica del negocio.
Transacción	En el manejo de almacenamiento de datos, una transacción se refiere a una unidad de trabajo que debe ser realizada de forma atómica, consistente, aislada y durable.

RESUMEN

Inteligencia de negocio, es el proceso de obtener información sobre la organización a partir de los datos existentes. En la Era de la Información, las organizaciones tienen a su disposición vastas cantidades de datos, recolectadas en sistemas transaccionales. Dichos sistemas, son esenciales para la operación del negocio.

Disponer de datos no es lo mismo que disponer de información. Los datos se convierten en información cuando se pueden utilizar para responder a cuestiones del negocio, de tal manera que se pueda comprender mejor el funcionamiento del mismo. La Inteligencia de Negocio permite responder a tales cuestiones, por lo que los tomadores de decisiones de todos los niveles puedan responder rápidamente ante los cambios en el entorno en el que compiten.

Para responder a las preguntas del negocio por medio de la Inteligencia de Negocio, se requiere de datos históricos o de datos provenientes de distintas fuentes. Dichos datos se consolidan en lo que se denomina un *Data Warehouse*, que no es más que una base de datos diseñada para consultas y análisis. El *data warehouse* integra datos históricos derivados de los sistemas transaccionales, con información de otras fuentes que pueden ser internas o externas a la organización.

El objetivo central de este trabajo, es dar una solución al problema de las necesidades de información para la toma de decisiones de las autoridades y docentes, de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la

Universidad de San Carlos de Guatemala, en lo que respecta a información académica, construyendo un *data mart* de información académica que solvente dichas necesidades y permita a los tomadores de decisiones construir sus propios escenarios de análisis.

OBJETIVOS

GENERALES

1. Desarrollar una solución de inteligencia de negocios, que dé respuesta a las necesidades de información académica de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería, de la Universidad de San Carlos de Guatemala.
2. Mostrar los beneficios de utilizar herramientas de inteligencia de negocio para el análisis de información y el desarrollo de reportes, por parte de los tomadores de decisiones en cualquier organización.

ESPECÍFICOS

1. Determinar los requerimientos y necesidades de información de la Escuela de Sistemas, los cuales deberán verse solventados en la solución final.
2. Diseñar, desarrollar, instalar y configurar la solución de inteligencia de negocio que satisfaga los requerimientos de información recabados.
3. Capacitar al personal técnico de la Escuela de Sistemas en el uso de la solución implementada.

INTRODUCCIÓN

El presente informe final enumera y describe cada uno de los aspectos realizados dentro del programa de Ejercicio Profesional Supervisado, que se llevó a cabo en la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería, de la Universidad de San Carlos de Guatemala, con el fin de diseñar y desarrollar una herramienta de inteligencia de negocio que permita el análisis de información académica, y a la vez apoye en la toma de decisiones a las autoridades de la Escuela.

En respuesta a continuos cambios en sus ambientes, las organizaciones se han enfrentado al reto de tomar decisiones basándose en información que generalmente se encuentra dispersa y en diferentes formatos, e integrarla, generalmente les ha representado mucho esfuerzo y tiempo. La inteligencia de negocio permite integrar la información de todas las áreas de una organización para que desde una interfaz sencilla, se pueda entender cómo está funcionando y detectar áreas de oportunidad para mejorarlo. Analizando indicadores clave y utilizando herramientas para profundizar en el suficiente detalle para entender su comportamiento, se podrán tomar decisiones oportunas y acertadas.

Dentro de los aspectos realizados dentro del programa de EPS en la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería, se pueden enumerar los siguientes:

- Estudio de la situación actual y recabar necesidades de información.
- Diseño del *Data Mart* Académico basado en las necesidades de información recabadas.

- Desarrollo del *Data Mart* y procesos asociados a la extracción, transformación y carga de la información.
- Construcción de reportes para análisis y toma de decisiones.
- Capacitación a usuarios de la solución.

Enfocándonos en las tareas anteriormente enunciadas se llevó a cabo el trabajo de EPS, con el fin de contribuir a mejorar la toma de decisiones de la Escuela de Ciencias y Sistemas y de la Facultad de Ingeniería.

1. INTRODUCCIÓN AL PROBLEMA Y SU SOLUCIÓN

1.1 Definición del problema

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 de las preguntas del negocio, pero se necesita dedicar mucho 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.

1.2. Objetivos del trabajo

Con sistemas orientados a la toma de decisiones se puede formular y responder las preguntas claves sobre el funcionamiento de la escuela de sistemas accediendo directamente a los indicadores académicos, señalar cuáles son los factores que realmente inciden en el buen o mal desempeño académico,

detectar situaciones fuera de lo normal, encontrar los factores que maximicen la mejora de la organización y predecir el comportamiento futuro con un alto porcentaje de certeza.

Para esto es necesario implementar un Sistema de Apoyo Gerencial, que en el caso de nuestro EPS se limita a la creación de un primer *data mart* académico orientado a necesidades de información de la escuela de ciencias y sistemas. El sistema debe estar orientado a brindar información interrelacionada para quienes tienen responsabilidades en el ámbito estratégico y táctico de la organización.

El presente informe se limita a un área de la facultad de ingeniería de la universidad, a los requerimientos de información solicitados por dicha área como también a las herramientas disponibles.

1.3. Aproximación de la solución

Para construir el *data mart* académico se consideró como punto de partida la información existente en las bases de datos de centro de cálculo de la facultad, en lo que a estudiantes de la escuela de sistemas se refiere.

Para lograr la integración de estos tipos de sistemas se cuenta con un repositorio de datos preparado para tal fin. Este repositorio se creó bajo las características de un *data warehouse*. Para llevar adelante el desarrollo del sistema se utilizó la metodología de desarrollo y explotación de *Data warehouses*.

1.4. Organización del documento

El presente documento está organizado en diferentes capítulos, los cuáles se describen a continuación.

En el Capítulo 2 se realiza una introducción a Inteligencia de Negocios y Datawarehousing. 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.

En el Capítulo 3 se expone el diseño conceptual de la solución, presentándose los diferentes diseños realizados para las estructuras que soportarán la información para la toma de decisiones, así como una breve descripción de las consideraciones a tomarse en cuenta para la estrategia de extracción, transformación y carga de los datos hacia el *data mart*.

En el Capítulo 4 se detalla el modelo de datos físico que soportará a los diferentes objetos que constituyen el *data mart* académico. Así también se hace un descripción de los procesos de ETL (extracción, transformación y carga hacia el *data mart*), y la plataforma tecnológica utilizada en la solución.

Así también se presenta todos los aspectos referentes a la entrega de información a los usuarios finales del *data mart*. Se presentan la estructura de seguridad y también las plantillas de reportes que se han dejado creadas para que a partir de las mismas los diferentes usuarios pueden crear sus propios escenarios de análisis.

2. INTELIGENCIA DE NEGOCIO, DEFINICIONES Y FUNDAMENTOS

2.1. Introducción

2.1.1. Definición de inteligencia de negocio

Inteligencia de negocio, es el proceso de obtener información sobre el negocio a partir de los datos existentes. En la era de la información, las organizaciones tienen a su disposición vastas cantidades de datos, recolectadas en sistemas transaccionales. Dichos sistemas, son esenciales para la operación del negocio.

Disponer de datos no es lo mismo que disponer de información. Los datos se convierten en información cuando se pueden utilizar para responder a cuestiones del negocio, de tal manera que se pueda comprender mejor el funcionamiento del mismo. La inteligencia de negocio permite responder a tales cuestiones, por lo que los tomadores de decisiones de todos los niveles puedan responder rápidamente ante los cambios en el entorno de los negocios.

Una vez resuelta la etapa de la automatización de los procesos operativos y cotidianos de las organizaciones, se cuenta con una invaluable fuente de información del negocio. Esta información es la que conforma el verdadero retorno de la inversión realizada.

La inteligencia de negocio es el proceso a través del cual es posible agrupar, resumir e interpretar esta información para medir el desempeño de la empresa contra sus metas y la industria donde compete.

2.1.2. Beneficios de la inteligencia de negocio

Los principales beneficios que se puede obtener al implementar una solución de inteligencia de negocio en una organización son:

- Respuestas inmediatas a preguntas del negocio, que son básicas para la toma de decisiones.
- Integración de datos entre los diferentes sistemas de información existentes en la organización.
- Permite lograr una visión del futuro a través del análisis de datos históricos.
- Dar la libertad al tomador de decisiones para crear diferentes escenarios de análisis, sin la dependencia del área de tecnología.
- Lograr medir el desempeño de la organización en función de sus metas y la industria en donde compete.

Las soluciones de inteligencia de negocio están orientadas a apoyar a las organizaciones para resolver sus necesidades de información en los tres niveles básicos: operativo, táctico y estratégico. De esta forma se logra crear un ambiente de trabajo en el que las decisiones de cualquier nivel son tomadas en base a información y conocimiento de la realidad que vive el negocio.

Figura 1. Niveles organizacionales en los que colabora la inteligencia de negocio



2.1.3. 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.

2.1.4. 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.

2.1.5. 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.

2.2 Objetivo de la información

El objetivo del usuario operativo es que se le facilite y automatice la operación de una función específica de la empresa; el de un estratega es maximizar la función de la empresa.

2.2.1 Usuarios

El usuario es distinto incluso en la misma pirámide organizacional. Mientras los sistemas operativos tienen interfaces muy especializadas para un usuario que realiza una operación rutinaria, los usuarios estratégicos realizan consultas variadas y no previstas de la información, por lo que los sistemas deben ser sencillos y con toda la información disponible que cubra cualquier consulta requerida, de este caso el software final debe ser orientado a un usuario en particular y, por ende, deberá adecuarse al conocimiento que tenga sobre el tema.

2.2.2 Tipos de preguntas

Las preguntas que responde un sistema operacional son referentes a las transacciones que se realizan diariamente y a nivel registro o suma de registros de un solo tipo. Un usuario operativo realiza frecuentemente preguntas sobre registros como pueden ser el estado actual de una factura, movimientos de un cliente, cantidad surtida por un proveedor, fecha del último movimiento de un distribuidor, etc. Las preguntas de un ejecutivo pueden también ser específicas, pero se orientan más a agrupamientos de datos como pueden ser totales por zona, promedios de clientes, tendencias de ventas e incluso pronósticos. Toda

esta información se encuentra de alguna forma en los almacenes operativos, pero lanzar una consulta como las ventas totales del año anterior puede implicar hasta días en resolverse y otro tiempo para publicar los datos. Un sistema organizado para resolver preguntas de ambos tipos en el menor tiempo posible es lo ideal.

2.2.3 Cantidad de datos

Si un usuario procesa la información de las transacciones se mueve en el nivel registro. Si un usuario procesa información de entidades, se mueve en el nivel agrupamientos de registros, obviamente la cantidad de datos que se necesitan es distinta y debe ser un sistema diferente el que provea de esa información. Para que un director o gerente, quien necesita conocer las transacciones de toda una zona para tomar una decisión, pudiera analizar cierto comportamiento, serían necesarias muchas hojas de reportes con cientos de datos. El usuario operativo que necesita pocos registros no tiene mayor problema por recibir una hoja de reportes, pero el directivo si tendría problemas con una cantidad exagerada de papeles. Se necesitan sistemas que brinden no solo la cantidad ideal de información según el usuario, sino también que la entreguen en tiempos óptimos.

2.3. Introducción a *Data warehouse* y *Data Mart*

Hoy en día las empresas cuentan en su mayoría con sus procesos automatizados, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. 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 lo alto; y era en el estrato de la pirámide más alto 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 estratos 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 *Data warehouse*. 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 DW (de ahora en adelante los términos *Data warehouse*, Datawarehousing, Warehouse, Almacén de Datos y DW son 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 DW 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 DW es el lugar donde la gente puede acceder a sus datos.

El DW 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 (o *Data warehouse*) 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 *data warehouses* 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 *data warehouse* 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 *data warehouse* 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 OLTP - Procesamiento Transaccional en Línea (OnLine Transational Procesing, en inglés), 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.

DW 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, en inglés), usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

Para ampliar los conceptos anteriores, en la figura 2 se exponen las principales diferencias entre los sistemas Transaccionales (OLTP) y los basados en *Data warehouses*.

Figura 2. Diferencias entre sistemas transaccionales y *Data Warehouse*.

Transaccionales	Basados en Datawarehouse
Admiten el acceso simultáneo de muchos usuarios -miles- que agregan y modifican datos.	Admiten el acceso simultáneo de muchos usuarios -cientos- que consultan y no modifican datos.
Representan el estado, en cambio constante, de una organización, pero no guardan su historial.	Guardan el historial de una organización.

continúa

Transaccionales	Basados en Datawarehouse
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.

A continuación se realiza una breve explicación de una especialización de los *data warehouses*, denominada *datamart*.

Introducción 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 del *data warehouse*, 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. 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.

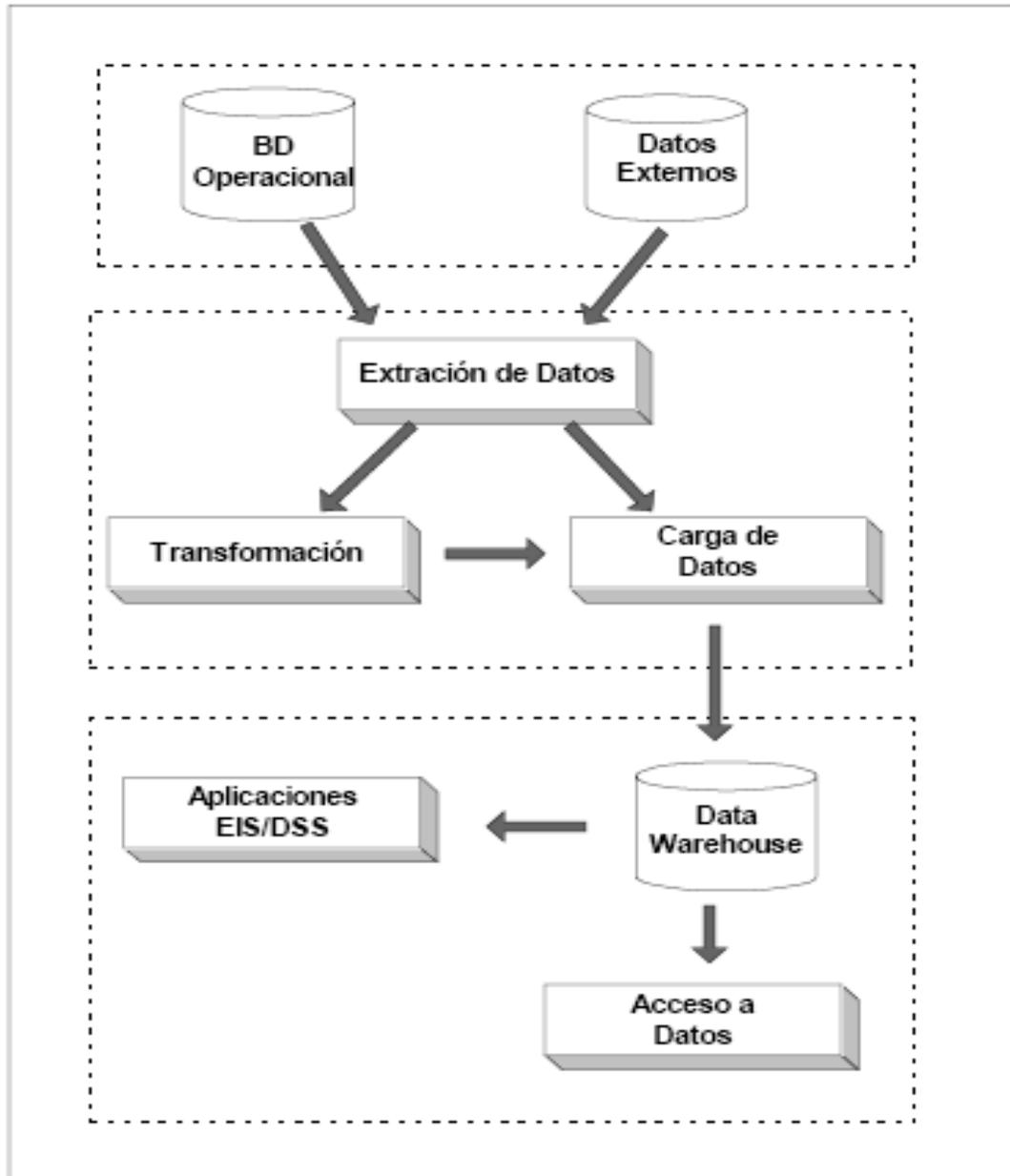
2.3.1. Arquitectura *Data Warehouse*

Antes de describir la arquitectura *data warehouse* se señala la siguiente consideración ya generalizada, presente en la literatura: el término *Data warehouse* 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. Ahora, con el propósito de facilitar el entendimiento por parte del lector, se hace especial énfasis en esta parte del capítulo sobre el contexto del cual se estará hablando al hacer referencia al término *data warehouse*.

La estructura básica de la arquitectura DW incluye:

1. Datos operacionales: un origen o fuente de datos para poblar el componente de almacenamiento físico DW. 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 DW.
3. Transformación de datos: procesos para sumarizar 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 DW.
5. *Data warehouse*: almacenamiento físico de datos de la arquitectura DW.
6. Herramientas de Acceso al componente de almacenamiento físico DW: herramientas que proveen acceso a los datos. Estas herramientas pueden ser herramientas específicas de mercado para visualización de bases multidimensionales almacenadas en *data warehouses* como también aplicaciones desarrolladas dentro de la organización del tipo EIS/DSS.

Figura 3. Estructura básica de un *data warehouse*



2.3.2. Implementación del *Data Warehouse*

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

- EL DW central: es una implementación de un solo nivel con un solo almacén para soportar los requerimientos de información de toda la empresa. En el DW central todos los usuarios de la organización acceden a la misma base de datos.
- El DW distribuido: es también una estructura de un nivel, pero particiona el almacén para distribuirlo a nivel departamental. En el DW distribuido, cada departamento, área o línea de negocio dispone de una base de datos propia con la información que solo le compete a los usuarios pertenecientes a estas áreas.
- El DW de dos niveles: combina ideas de los dos anteriores, se implementa el almacén empresarial como los departamentales. En el DW 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.

2.3.3. Valor del DW para la toma de decisiones

El valor de un DW queda descrito en tres dimensiones:

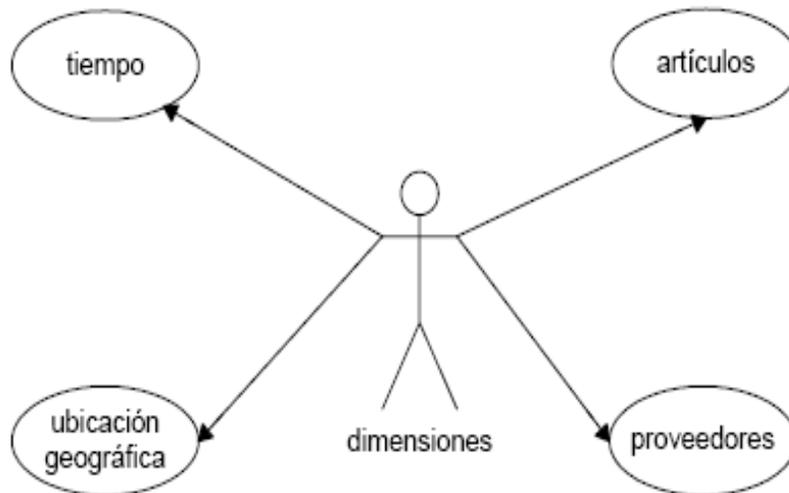
- Mejorar la Entrega de Información: información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita.
- Facilitar el Proceso de Toma de Decisiones: con un mayor soporte de información se obtienen decisiones 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.
- 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.

2.3.4 Componentes básicos

A la hora de realizar un modelo dimensional del negocio se hace necesaria la presentación de los siguientes componentes:

- Dimensiones
- Atributos
- Elementos
- Relaciones
- Jerarquías
- Tablas de Hechos o Indicadores.

Figura 4. Ejemplo de dimensiones



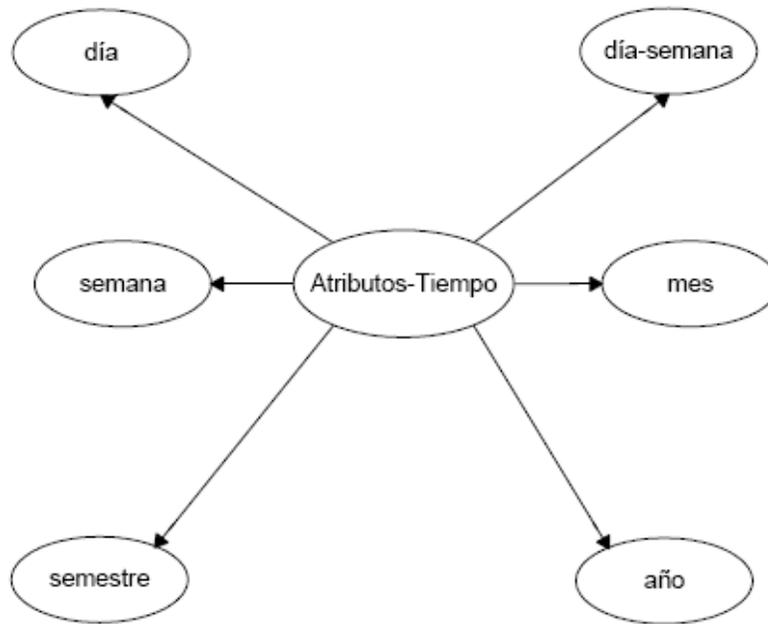
Dimensiones. Son las áreas temáticas, líneas del negocio o sujetos del negocio. Las mismas proveen un método general para organizar la información corporativa. Definidas como un grupo de uno o más atributos, separados y distintos uno de otros (es decir, que no se comparten atributos), según el esquema de implementación elegido, puede ser que no se encuentren explícitamente en el DW, sino que se presenten sólo como un recurso conceptual o lógico que ayuda a mostrar múltiples perspectivas de los datos permitiendo realizar análisis por diferentes dimensiones o incluso cruzando información entre distintas dimensiones.

En el ejemplo de dimensiones podemos ver que el tiempo, los artículos en la tienda, los proveedores y la ubicación geográfica de las tiendas, dan la dimensión básica para obtener una tabla de hechos.

Atributos. Los atributos son una agrupación de elementos o ítems dentro de una dimensión. Representan categorías o clases de elementos que tienen el mismo nivel lógico dentro de una dimensión donde todos los elementos de un atributo se relacionan con otros atributos de la dimensión de la misma forma. La finalidad de los atributos es ver la información de cada dimensión a diferentes niveles de detalle y agrupar los datos para ser analizados.

Por ejemplo, ítem de un producto; los ítems pueden tener algunos atributos característicos como ser el color y el tamaño. En un modelo dimensional, ítem, color y tamaño son todos tratados como atributos de la dimensión producto.

Figura 5. Ejemplo de dimensiones



En la figura anterior se puede ver como la dimensión tiempo puede desglosarse en varios atributos (día, semana, día de la semana, mes, año, etc.) que en conjunto conforman los calificadores de una dimensión.

Elementos. Son las instancias o valores de los atributos que, como componentes atómicos del modelo, permiten clasificar el rendimiento del negocio. Es importante aclarar que si bien no forman parte indispensable del modelado multidimensional es aconsejable su incorporación para un mayor entendimiento del modelo en etapas tempranas del ciclo de vida.

Relaciones. Son asociaciones lógicas de atributos dentro de una jerarquía definida por las instancias de los atributos y transitivas dentro de la jerarquía. Es importante aclarar que los atributos dentro de una dimensión están

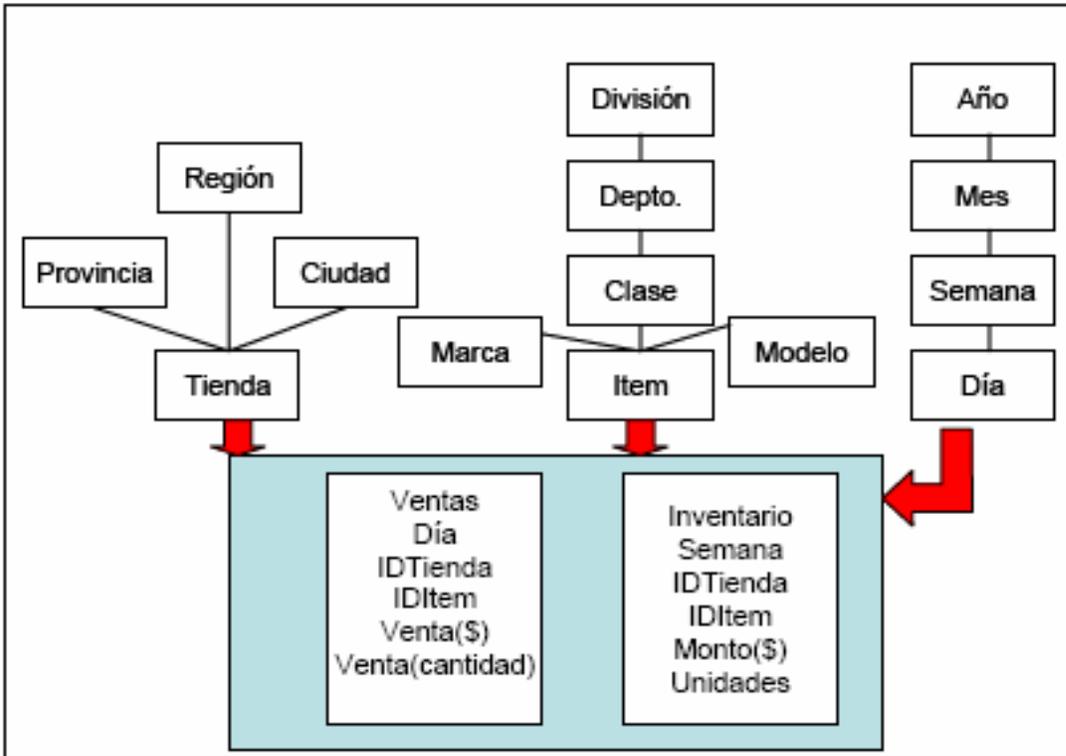
directamente relacionados unos con otros a través de los diferentes tipos de relaciones. En cambio, los atributos en diferentes dimensiones están relacionados uno con otros a través de los indicadores o variables del negocio definidas como intersección de las dimensiones.

Jerarquías. Representadas por un ordenamiento lógico dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión. Pueden existir múltiples jerarquías dentro de una dimensión pero siempre es posible identificar una jerarquía principal o columna vertebral de la dimensión y jerarquías secundarias o descriptivas compuestas por atributos característicos definidos desde la jerarquía principal.

Indicadores o tabla de hechos . Son las variables o métricas que ayudarán a medir la representación del negocio. Existen dos tipos de indicadores: básicos y derivados.

Los indicadores básicos, primitivos o crudos existen físicamente en el *data warehouse* junto a los atributos que los caracterizan, pueden venir de diferentes sistemas fuentes y tener distintos niveles de granularidad o agregación. Por ejemplo, la variable Venta(\$) es llevada diariamente mientras que el indicador Unidades en Existencias o Inventario es seguido semanalmente. Por otro lado, los indicadores derivados o métricas calculadas se construyen a partir de los indicadores base y pueden o no estar almacenados físicamente en el DW. Un ejemplo clásico de métrica derivada es Margen de Ganancia, la cual se define como la resta entre Precio y Costo (ambos indicadores básicos).

Figura 6. Ejemplo de dimensiones



En la figura anterior se puede ver que los datos derivados de las dimensiones dan origen a una tabla general que concentra los datos necesarios para poder obtener un análisis específico según los requerimientos, por ejemplo en una venta de respuestas se quiere obtener el estado actual de las ventas del negocio, obteniendo los datos necesarios de las dimensiones Ubicación, Tiempo y Artículos hacia una tabla de hechos que es la que monitorea el comportamiento de las ventas.

2.3.5 Representación del *Data Warehouse*

Existen varias formas de representación o implementación del modelo dimensional dentro de las diferentes bases de datos que existen en el mercado. En el caso de contarse con un ambiente netamente relacional, es decir, una base de datos relacional sobre la cual se hará la implementación del *data warehouse*, se suele utilizar una representación denominada "esquema en estrella", la cual proporciona simpleza y flexibilidad para el uso de cualquier herramienta de usuario final.

2.3.5.1 Esquema en estrella

El modelo de esquema en estrella puede verse como una simple estrella en la cual existe una tabla central que contiene los hechos del negocio que se desean modelar, y múltiples tablas radiales, llamadas dimensiones, conectadas a la central a través de las respectivas llaves primarias y foráneas. A diferencia de la estructura de otros esquemas de base de datos, un esquema en estrella contiene dimensiones denormalizadas.

Los esquemas en estrella presentan las siguientes características:

- El modelo es fácil de entender para los usuarios.
- La llave primaria representa a cada una de las dimensiones.
- Las columnas que no son llaves foráneas son valores.
- Las tablas de hechos están usualmente altamente normalizadas.
- Las dimensiones están completamente denormalizadas.
- Proveen una respuesta rápida a las consultas.
- El rendimiento de las consultas es mejorado reduciendo las uniones entre tablas.

- Los usuarios pueden expresar consultas complejas.
- Son ambientes ideales para proveer información a través de múltiples herramientas de presentación.

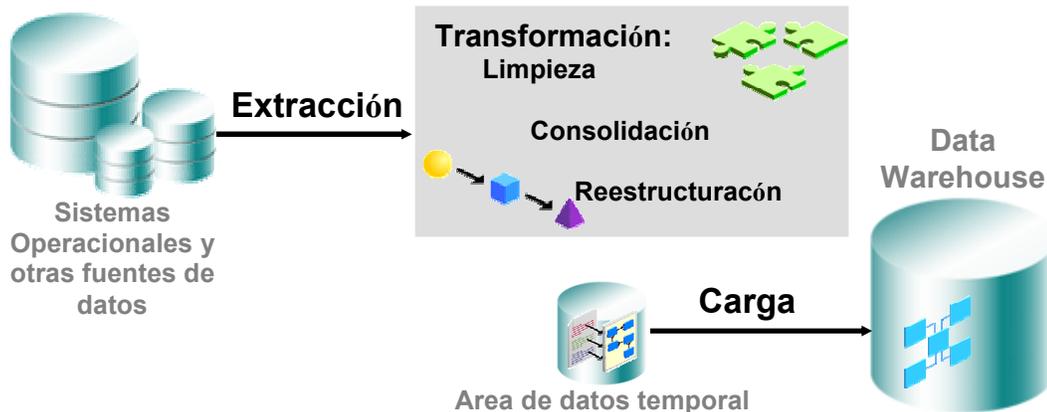
2.3.6 Extracción, transformación y carga del *Data Warehouse*

Los datos recopilados en las diferentes fuentes de almacenamiento son considerados datos "sucios" por contener muchas de las veces información innecesaria para el análisis dentro del DW y por tener desigualdad de criterio para describir cierto dato.

La limpieza de los datos "sucios" es un proceso multifacético y complejo que consiste en analizar los datos corporativos para descubrir inexactitudes, anomalías y transformar los datos para asegurar que sean precisos y coherentes. Asegurar la integridad referencial, le proporciona al DW la capacidad de identificar correctamente al instante, cada objeto del negocio, tales como un producto, un cliente o un empleado.

Los Sistemas ETL realizan las funciones de extracción de las fuentes de datos (transaccionales o externas), transformación (limpieza, consolidación, etc.) y la carga de los datos una vez que ya hayan pasado por los dos anteriores pasos, para así almacenarse en el DW.

Figura 7. Proceso ETL (extracción, transformación y carga)



En la figura anterior se muestra la secuencia que el proceso ETL ejecuta para realizar el mantenimiento del DW, en el cual el primer proceso es la extracción de los datos necesarios para el análisis de la información, una vez hecho la extracción el proceso siguiente es la transformación de los datos en datos basados según el formato estandarizado dentro del DW y por último se realiza la carga de los datos en el DW.

Los pasos para el proceso ETL consisten en:

- Validar los datos que usa la aplicación del DW para realizar las consultas de prueba.
- Producir la metadata, una descripción del tipo de datos, formato y el significado relacionado al negocio de cada campo.
- La documentación del proceso completo, para que se pueda ampliar, modificar y arreglar los datos en el futuro con más facilidad.
- La carga de los datos en el warehouse.

Existen varias razones que justifican la creación del DW para obtener la información necesaria en los procesos de gestión comercial, en lugar de obtener esa información directamente de las bases de datos de las aplicaciones operacionales:

Rendimiento: se tarda mucho menos en acceder a los datos del repositorio del DW que en hacer una consulta a varias base de datos distintas. Además hacer consultas complicadas a las base de datos de los sistemas operacionales puede empeorar el tiempo de respuesta de estos sistemas para otros usuarios.

Múltiples orígenes de datos: combinar los datos de distintas fuentes suele ser una tarea bastante complicada para las personas encargadas de tomar decisiones con esa información. Normalmente hay que homogenizar los datos de una forma u otra. Por ejemplo, es probable que no se utilicen los mismos criterios de almacenamiento (nombres de las entidades, atributos considerados, tipos, etc.) en las bases de datos de distintos departamentos.

Sin embargo, en el DW los datos se homogenizan durante el proceso de carga (ETL) todo esto para obtener datos específicos y poder realizar un adecuado análisis para la toma de decisión.

En general el DW es una herramienta útil para el análisis específico de datos y anudado a reglas de análisis pueden ofrecer apoyo a la toma de decisiones al implementarle dichas reglas.

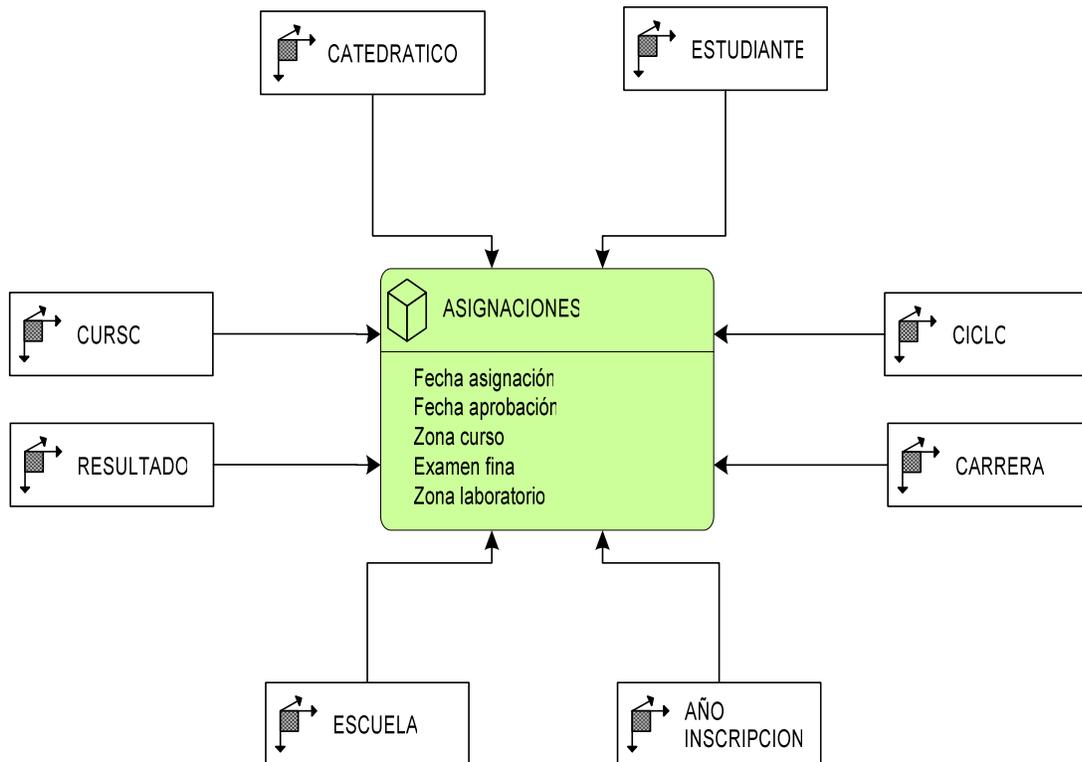
3. MODELO DE DATOS LÓGICOS

3.1 Especificación del modelo dimensional

A continuación se presenta el modelo de datos lógico para el *Data Mart* que representa los diferentes requerimientos de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos. Primero se presentan los modelos en estrella y las descripciones de las tablas de hecho y luego se describe cada dimensión del modelo.

Cabe mencionar que la carga inicial de información estuvo limitada a estudiantes de la carrera de ingeniería en ciencias de la computación y sistemas que estuvieran inscritos en la facultada de ingeniería al finalizar el primer semestre del año 2006. También se limitó la información de catedráticos, carreras y escuelas, a que se tuviese solamente aquellos relacionados con la escuela de ciencias y sistemas.

Figura 8. Modelo dimensional para Asignaciones de estudiantes y cursos (esquema en estrella)



3.1.1 Modelo estrella para asignaciones

Este esquema en estrella permite obtener información sobre cada asignación de un estudiante a un curso en un ciclo académico determinado, así como las notas del estudiante en dicho curso. Este esquema representa el corazón de la solución a nivel de base de datos, ya que presenta la información detallada de las asignaciones de los estudiante y las notas obtenidas en cada ciclo y curso asignado.

Métricas asociadas:

Zona curso: Zona del estudiante en el curso asignado.

Examen final: Nota del estudiante en el examen final.

Zona laboratorio: Zona del estudiante en el laboratorio del curso.

Dimensiones asociadas:

Estudiante, ciclo académico, curso, escuela, resultado, año inscripción, carrera, catedrático, fecha aprobación laboratorio, fecha de asignación.

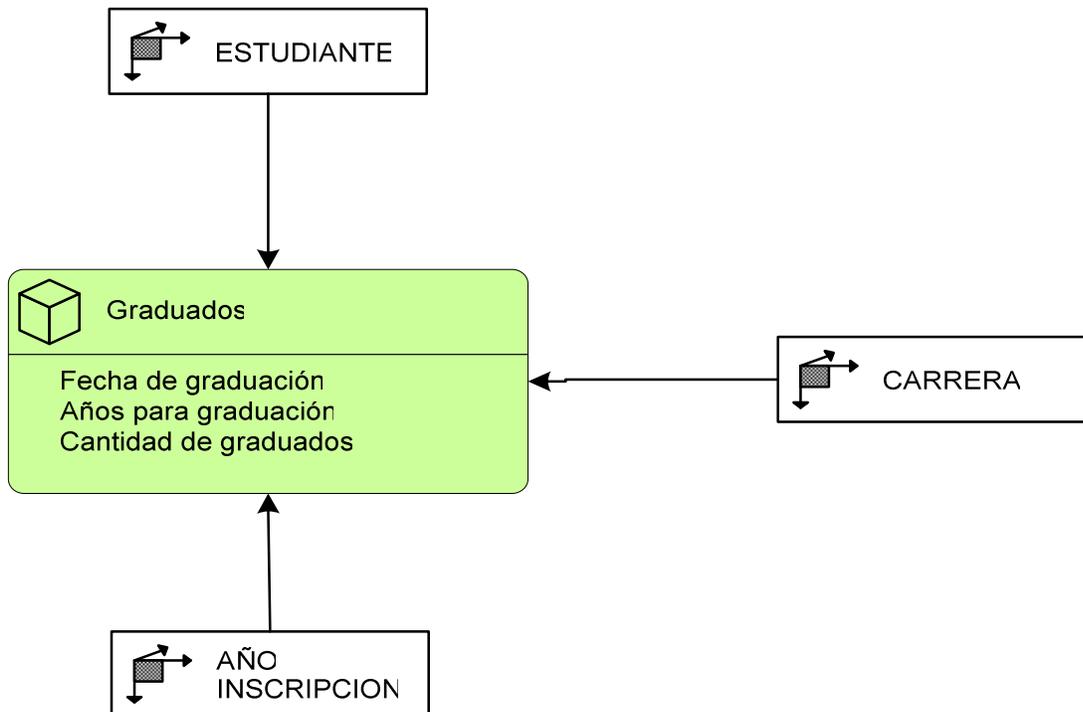
Periodicidad de carga:

Este esquema en estrella debe ser refrescada al finalizar cada ciclo académico, es decir, al final de cada periodo en el que se impartan cursos o realicen evaluaciones, como semestres, cursos de vacaciones o exámenes de retrasadas.

Granularidad de los datos:

El nivel de detalla de los datos en este esquema en estrella es el de cada asignación de un curso por parte de un estudiante en un ciclo académico. Se presenta cada curso asignado por un estudiante en cada uno de los ciclos académicos existentes, y para cada curso se presenta sus respectivas notas.

Figura 9. Modelo dimensional para análisis de Graduados



3.1.2 Modelo estrella para análisis de estudiantes graduados

Este esquema en estrella presenta información sobre los estudiantes que se han graduado en una carrera determinada, adicionalmente a datos tales como, el año de inscripción del estudiante y sus años para graduarse. Inicialmente se cuenta únicamente con información de los estudiantes graduados durante el año 2006.

Métricas asociadas:

Años para graduación: Número de años transcurridos entre el año de inscripción del estudiante y la fecha de graduación.

Cantidad de estudiantes graduados: Métricas que se obtiene del conteo de estudiantes que se han graduado por la combinación de valores de las dimensiones utilizadas en cada consulta.

Dimensiones asociadas:

Estudiante, fecha de graduación, año de inscripción, carrera.

Cálculos:

Años para graduación: es la diferencia entre la fecha de graduación y la fecha de inscripción.

Cantidad de graduados: métrica que se genera a partir de un conteo de los estudiantes que se han graduado, según la combinación de dimensiones utilizadas en la consulta.

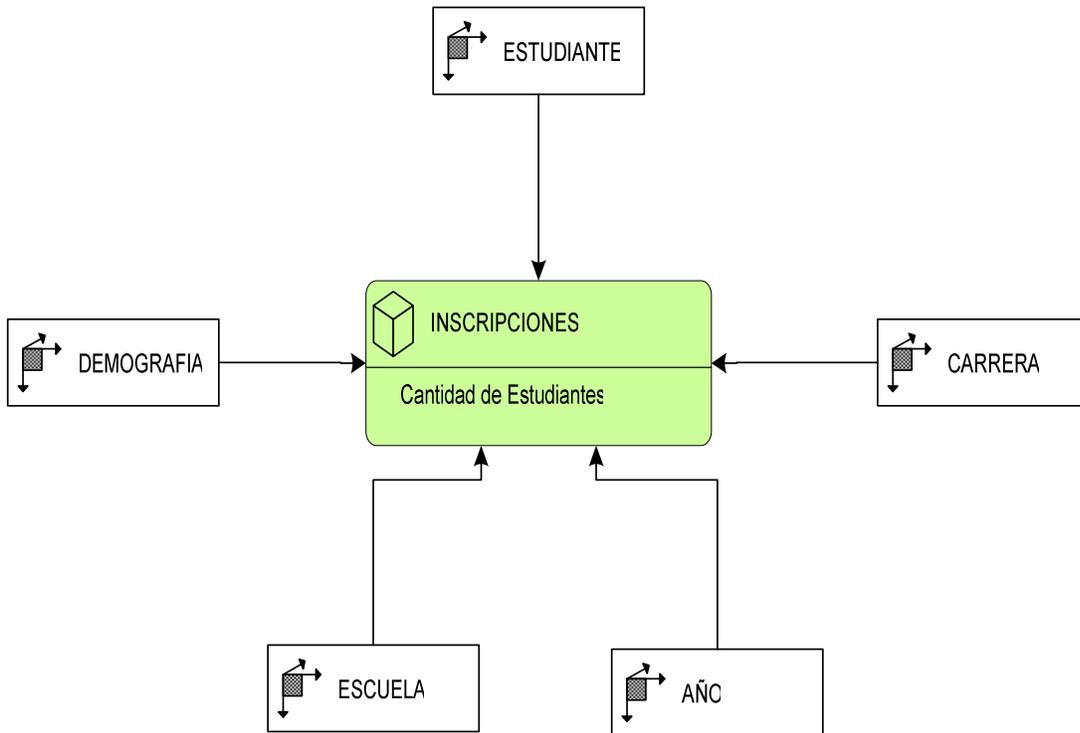
Periodicidad de carga:

Este esquema podrá ser actualizado cada vez que exista información de nuevos graduados, se recomienda que se haga cada final de semestre para tener la información actualizada.

Granularidad de los datos:

El nivel de detalle del esquema de información será de cada estudiante graduado por año de inscripción en una carrera dada.

Figura 10. Modelo dimensional para análisis de inscripciones de estudiantes



3.1.3 Modelo estrella para inscripciones

Esquema en estrella que presenta información relacionada con las inscripciones de estudiantes a la escuela de ciencias y sistemas. Este esquema toma como punto de partida la información de las asignaciones y se hace la búsqueda para reconocer a aquellos estudiantes que son nuevos en la universidad por su número de carnet. Este esquema contiene información demográfica la cual permite hacer análisis de estudiantes por datos tales como, la edad, el género, o el estado civil.

Métricas asociadas:

Cantidad de estudiante, representa el número de estudiantes que se han inscrito en un determinado año, por cada carrera, escuela y datos demográficos.

Dimensiones asociadas:

Estudiante, año de inscripción, Carrera, Escuela, Demografía de los estudiantes

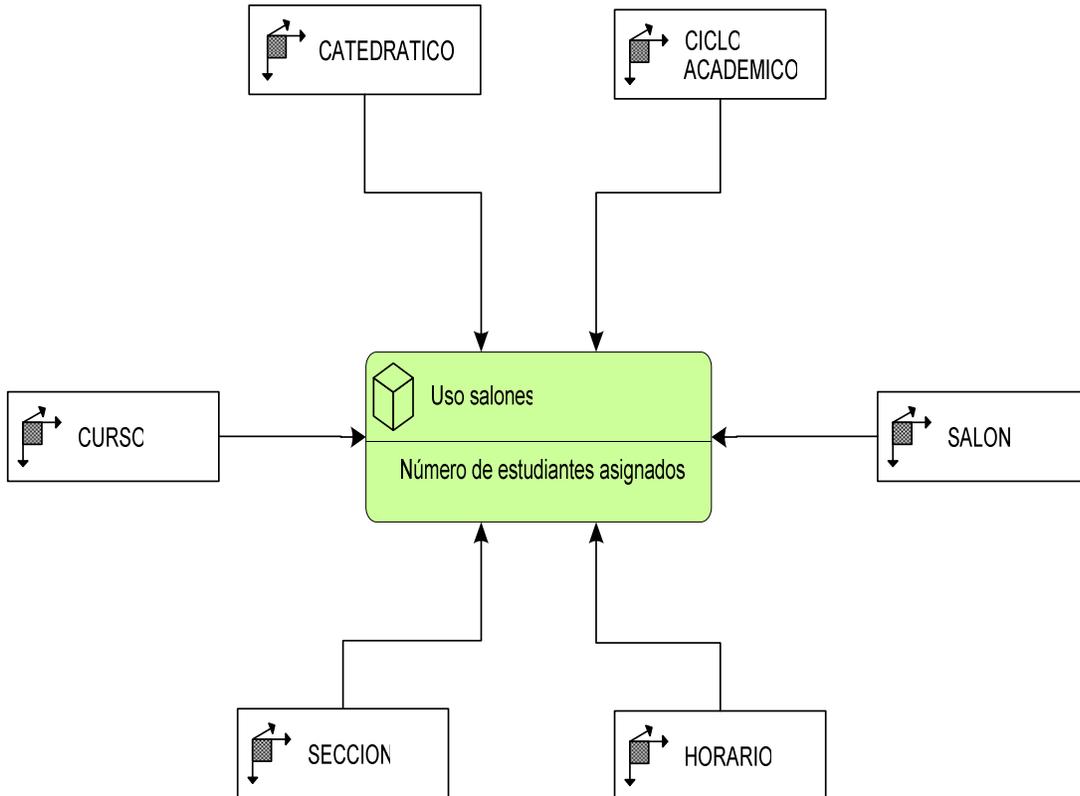
Periodicidad de carga:

Se deberá realizar, la actualización del esquema, cada semestre a fin de conocer la cantidad de estudiantes que se inscriben cada año.

Granularidad de los datos:

El nivel de detalle es de cada estudiante inscrito por año en una carrera dada.

Figura 11. Modelo dimensional para análisis de Uso de Salones



3.1.4 Modelo estrella para análisis de uso de salones

Este esquema presenta información sobre el uso de los salones que conforman cada uno de los edificios utilizados por la escuela de ciencias y sistemas, en un semestre o curso de vacaciones dado. Para cada salón se tiene información de su capacidad y de los estudiantes asignados al mismo en cada ciclo académico, por cada curso y horario del día de la semana.

Métricas asociadas:

Número de estudiantes asignados: Número de estudiantes asignados a un salón en un periodo dado.

Dimensiones asociadas:

Ciclo académico, catedrático, curso, horario, salón, sección.

Periodicidad de carga:

Se deberá refrescar la información de este esquema al inicio de cada ciclo académico en el momento que se conozca la distribución de salones dada por la facultada para cada uno de los cursos que serán impartidos.

Granularidad de los datos:

Salón en el que se imparte un curso en un ciclo académico dado a determinada hora de cada día de la semana.

3.1.5 Dimensiones del modelo de datos lógico

3.1.5.1 Dimensión estudiante

Representa a un estudiante inscrito en la Facultad de Ingeniería y que pertenece a la escuela de ciencias y sistemas. Contiene información muy general de cada estudiante con datos, tales como, su nombre y apellido, fecha de nacimiento, género y estado civil.

Atributos asociados:

Carnet, número de carnet del estudiante en la Facultad.

Nombres, nombres del estudiante.

Apellidos, apellidos del estudiante.

Fecha de nacimiento, fecha en que nació el estudiante.

Género, sexo del estudiante.

Estado civil, estado civil del estudiante

3.1.5.2 Dimensión curso

Representa a cada curso que se imparte en la pensión de estudios de la Facultad de Ingeniería, específicamente asociado a la escuela de ciencias y sistemas.

Atributos asociados:

Código departamento, código del departamento encargado de impartir el curso.

Nombre departamento, nombre del departamento encargado de impartir el curso.

Código área, código del área profesional a la que pertenece el curso.

Nombre área, nombre del área profesional a la que pertenece el curso.

Código subárea, código del subárea profesional a la que pertenece el curso.

Nombre subárea, nombre del subárea profesional a la que pertenece el curso.

Código curso, código con el que se identifica al curso.

Nombre curso, nombre del curso.

Reglas de transformación:

Se combinarán en esta dimensión el curso, el subárea, el área y el departamento al que pertenece cada curso, de tal forma que constituyan una jerarquía.

3.1.5.3 Dimensión carrera

Representa a una carrera adscrita a la Facultad de Ingeniería. Inicialmente se tiene contemplada únicamente la carrera de ingeniería en ciencias de la computación y sistemas, pero esta diseñada para poder contener a las diferentes carreras de la facultad de ingeniería.

Atributos asociados:

Código, código con el que se identifica a la carrera.

Nombre, nombre de la carrera.

Proceso de carga (mapeo de fuentes):

Los datos de la carrera se generan mediante un procedimiento manual, ya que no se tuvo información fuente del detalle de carreras, debido a la orientación del *data mart* de contener información exclusivamente de la escuela de ciencias y sistemas.

Consideraciones:

Cuando esté disponible una tabla de carreras, será necesario crear un procedimiento para cargar la dimensión carrera.

3.1.5.4 Dimensión ciclo académico

Representa a un ciclo en el que se imparten cursos o realizan exámenes de retrasada en la Facultad de Ingeniería. Cada ciclo académico debe estar registrado en la dimensión, de tal forma que se pueda obtener información de todas las asignaciones que se dan durante cada año, tales como, primero y segundo semestre, cursos de vacaciones, y exámenes de retrasada.

Atributos asociados:

Código, código del ciclo.

Nombre, nombre del ciclo.

Año, año del ciclo.

3.1.5.5 Dimensión catedrático

Representa a un catedrático que imparte cursos en la Facultad de Ingeniería. Inicialmente se cuenta con el listado de los catedráticos que imparten cursos de la carrera de ciencias de la computación y sistemas.

Atributos asociados:

Registro personal, es el número con el que se identifica a un catedrático en la Facultad..

Nombres, nombres del catedrático.

Apellidos, apellidos del catedrático.

3.1.5.6 Dimensión salón

Representa a un salón en el que imparten cursos en la Facultad de Ingeniería. Dimensión que presenta el detalle de los salones en los que se imparten cursos de la carrera de ciencias de la computación y sistemas. Los salones están agrupados por edificio en el que se encuentran.

Atributos asociados:

Nombre salón, es el nombre con el que se identifica a un salón en la Facultad.

Nombre edificio, nombre del edificio en el que se encuentra el salón.

Capacidad, capacidad de estudiantes del salón.

3.1.5.7 Dimensión sección

Representa a una sección que se puede asignar a un curso en la Facultad de Ingeniería. Sirve de información complementaria para el uso de salones, ya que permite identificar que cursos se imparten en los salones y en que horario se llevan cabo durante la semana.

Atributos asociados:

Nombre, es el nombre con el que se identifica a la sección en la Facultad.

3.1.5.8 Dimensión año

Representa a cada uno de los años en los que pudo ocurrir un evento de interés, como el de inscripción, un ciclo académico o un año de graduación.

Atributos asociados:

Año, es el año del calendario gregoriano

Proceso de carga (mapeo de fuentes):

Se generaron automáticamente 150 años consecutivos, desde 1950.

3.1.5.9 Dimensión escuela

Representa a una escuela adscrita a la Facultad de Ingeniería. Con esta dimensión se pretende dejar el diseño previsto para que se pueda agregar al *data mart* a las diferentes escuelas de la facultad. Actualmente se cuenta con información únicamente de la escuela de ciencias y sistemas.

Atributos asociados:

Nombre, es el nombre de la escuela.

Director, nombre del director de escuela.

Consideraciones:

Cuando se disponga de una tabla con los nombres de escuelas, deberá crearse un proceso que permita cargar la dimensión.

3.1.5.10 Dimensión demografía

Representa a un conjunto de características generales y clasificatorias de los estudiantes, tales como, su género, rango de edad y estado civil. Mediante esta dimensión se puede hacer análisis de los estudiantes que ingresan a la facultad, en cuanto a datos de edad, género y estado civil.

Atributos:

Género, sexo del estudiante

Estado civil, estado civil del estudiante.

Nombre rango edad, nombre del rango de edad.

Edad inicio rango, edad en la que empieza el rango.

Edad fin rango, edad en la que finaliza el rango.

3.1.5.11 Dimensión resultado

Representa el resultado que obtuvo un estudiante en un curso determinado, es decir, el calificativo de aprobado o reprobado. Esta es una dimensión derivada que resulta de la evaluación de la nota obtenida por un estudiante en un determinado curso durante un ciclo académico.

Atributos:

Nombre, es el nombre del resultado.

Consideraciones:

El proceso de ETL asociado a la tabla de hechos de asignaciones, es el que permite asociar a cada registro de notas el valor adecuado para esta dimensión. Se debe tomar en cuenta que la nota de aprobación era de 51 puntos, hasta el primer semestre del año 2005 donde cambió a ser 61.

3.1.5.12 Dimensión horario

Representa a un horario de la semana en el que se imparten cursos en la Facultad de Ingeniería. Dimensión utilizada en el uso de salones para indicar el día de la semana que se imparte un determinado curso en un salón y el horario durante el día.

Atributos:

Hora inicio, es la hora a la que inicia el horario.

Hora fin, es la hora a la que finaliza el horario.

Día semana, día de la semana en el que se encuentra el horario.

4. MODELO DE DATOS FÍSICO Y PRESENTACIÓN DE INFORMACIÓN

4.1 Correspondencia entre el modelo dimensional y el modelo físico de base de datos

En la siguiente tabla se muestra la correspondencia entre cada dimensión y tabla de hechos (del modelo de datos lógico) y las tablas de la base de datos que implementan cada una de ellas (el modelo de datos físico) y que constituyen el *data mart* académico de la escuela de ciencias y sistemas de la facultad de ingeniería de la universidad de San Carlos.

Esquema de inscripciones

Figura 12. Mapeo de dimensiones a tablas relacionales del esquema de inscripciones

Dimensión / tabla de hecho	Tabla
ESTUDIANTE	DIM_ESTUDIANTE
DEMOGRAFIA	DIM_DEMOGRAFIA
CARRERA	DIM_CARRERA
ESCUELA	DIM_ESCUELA
AÑO	DIM_ANIO
INSCRIPCIONES	FACT_INSCRIPCION

Esquema de graduados

Figura 13. Mapeo de dimensiones a tablas relacionales del esquema de graduados

Dimensión / tabla de Tabla hecho	
ESTUDIANTE	DIM_ESTUDIANTE
CARRERA	DIM_CARRERA
AÑO	DIM_ANIO
GRADUADOS	FACT_GRADUADO

Esquema de uso de salones

Figura 14. Mapeo de dimensiones a tablas relacionales del esquema de uso de salones

Dimensión / tabla de Tabla hecho	
CURSO	DIM_CURSO
CATEDRATICO	DIM_CATEDRATICO
CICLO ACADEMICO	DIM_CICLO_ACADEMICO
SALON	DIM_SALON
HORARIO	DIM_HORARIO
SECCION	DIM_SECCION
USO DE SALONES	FACT_USO_SALON

Data mart de asignaciones

Figura 15. Mapeo de dimensiones a tablas relacionales del *data mart* de asignaciones

Dimensión / tabla de Tabla hecho	
ESTUDIANTE	DIM_ESTUDIANTE
CARRERA	DIM_CARRERA
ESCUELA	DIM_ESCUELA
AÑO INSCRIPCION	DIM_ANIO
CATEDRATICO	DIM_CATEDRATICO
CICLO_ACADEMICO	DIM_CICLO_ACADEMICO
RESULTADO	DIM_RESULTADO
CURSO	DIM_CURSO
INSCRIPCIONES	FACT_INSCRIPCION

4.2 Plataforma de la solución

En la siguiente figura, se ilustra la plataforma utilizada para la solución de Inteligencia de Negocios. Los componentes principales son:

Sistemas transaccionales: Son las bases de datos en el RDBMS Informix, en donde se encuentran almacenadas las transacciones operativas de la Facultad.

Base de datos en access: Base de datos proporcionada por centro de calculo de la facultad de ingeniería, conteniendo el detalle de la información que fue convertida en archivos de texto, previo a la carga del *data mart*.

Archivos de texto: Archivos que contienen el detalle de las tablas que fueron proporcionadas por centro de calculo de la facultad de ingeniería, para tomarlos como base de carga hacia el *data mart*.

Base de datos Oracle 10g: Es la base de datos en donde se almacenan los datos del *data mart*.

Oracle Application Server 10g: Provee la infraestructura para distribuir aplicaciones a los usuarios finales a través de tecnologías de Internet.

Oracle Discoverer Plus: Permite la presentación y análisis de información por parte de los usuarios finales, sobre los datos contenidos en el *data mart*. Herramienta altamente flexible, sencilla e intuitiva para que los usuarios finales puedan crear sus propios escenarios de análisis, basados en la información contenida en el *data mart*.

Oracle Discoverer Administrador: Permite la administración de la capa de datos de usuario o End User Layer (EUL), a través de la cual se hace el mapeo entre la base de datos del *data mart* y la herramienta Oracle Discoverer, permitiéndole a los usuarios finales crear sus propios escenarios de análisis.

Figura 16. Plataforma de inteligencia de negocios

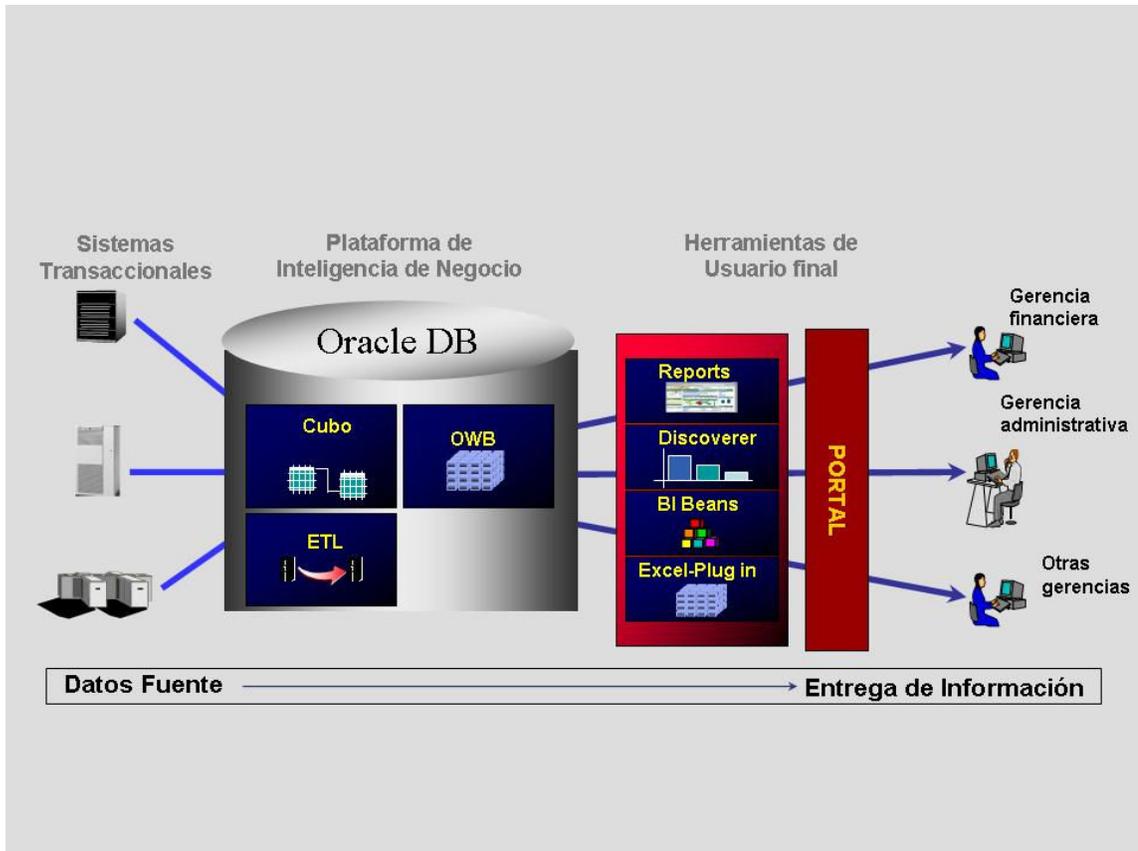
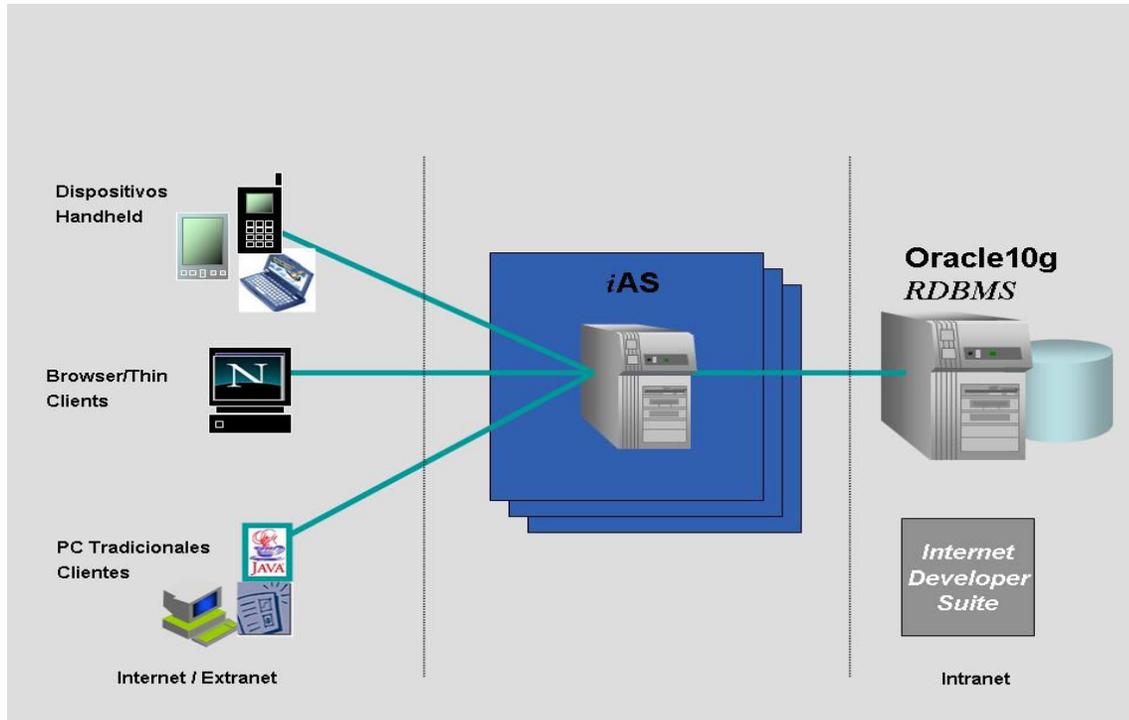


Figura 17. Plataforma del servidor de aplicaciones



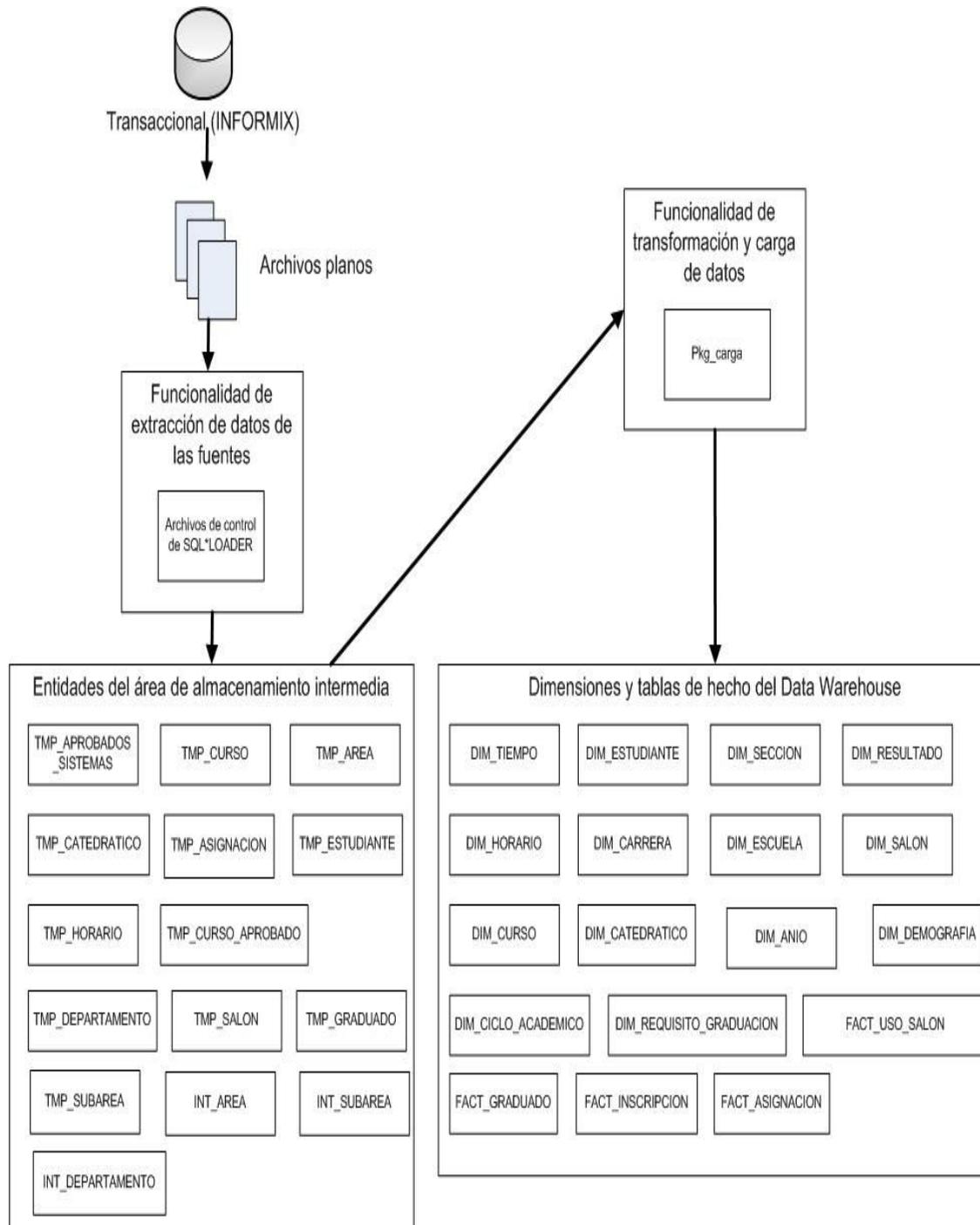
4.3 Extracción, transformación y carga de datos (ETL)

Los procesos de ETL son paquetes, procedimientos y funciones dentro de la base de datos, que tienen a su cargo la extracción de datos desde el sistema transaccional, transformación de los datos según las reglas del negocio y finalmente la carga al *Data Mart*. Los procesos ETL se dividen en dos grandes categorías, los encargados del mapeo de dimensiones y los que se encargan de poblar las tablas de hecho del *Data Mart*.

4.3.1 Flujo de datos

La siguiente figura muestra una vista panorámica del proceso de carga del *Data Warehouse*, que inicia en los sistemas transaccionales creados sobre una base de datos Informix. Luego, desde Informix se extraen los datos hacia archivos planos que son importados al área intermedia de Oracle vía el comando SQL*LOADER. Finalmente, a través del paquete PKG_CARGA, escrito en PL/SQL, se cargan los datos hacia las tablas del *Data Mart*. Los eventos importantes que suceden durante la carga se reportan en la bitácora CTL_BITACORA.

Figura 18. Proceso de extracción, transformación y carga hacia el *data mart* académico de la escuela de ciencias y sistemas (ETL)



4.3.2 Procesos de carga de las dimensiones

Los procesos que se encargan de poblar las dimensiones son procedimientos encapsulados en el paquete PKG_CARGA, que pueden ser ejecutados a través de un comando de línea desde SQL*Plus. A continuación se presentan los paquetes que cargan cada una de las dimensiones del *data mart*:

Figura 19. Procedimientos para la carga de dimensiones

Dimensión	Procedimiento de carga
DIM_RESULTADO	cargar_dim_resultado
DIM_DEMOGRAFIA	cargar_dim_demografia
DIM_ANIO	cargar_dim_anio
DIM_SECCION	cargar_dim_seccion
DIM_ESTUDIANTE	cargar_dim_estudiante
DIM_HORARIO	cargar_horario
DIM_CURSO	cargar_curso
DIM_CATEDRATICO	cargar_catedratico
DIM_CARRERA	cargar_carrera
DIM_ESCUELA	cargar_escuela
DIM_SALON	cargar_salon

Los procedimientos dentro del paquete PKG_CARGA que se encargan de poblar las tablas de hecho son:

Figura 20 . Procedimientos para la carga de tablas de hecho

Tabla de hechos	Prodimiento de carga
FACT_ASIGNACION	cargar_fact_asignacion
FACT_INSCRIPCION	cargar_fact_inscripcion
FACT_GRADUADO	cargar_fact_graduado
FACT_USO_SALON	cargar_fact_uso_salon

El siguiente procedimiento permite cargar y actualizar todas las dimensiones y tablas de hecho del *Data Mart* en una sola llamada:

Figura 21. Procedimientos para la carga del *data warehouse* completo

Funcionalidad	Prodimiento de carga
Cargar y todas las dimensiones y tablas de hecho	Cargar_data_warehouse

4.3.3 Registro de la actividad de un proceso de carga

Los procesos de carga del *Data Mart* y sus dimensiones registran su actividad en una bitácora, que es una tabla de la base de datos con los siguientes atributos:

Figura 22. Parámetros de la bitácora

Atributo	Tipo de datos	Descripción
Fecha_hora	DATE	Fecha y hora en la que se registró el evento en la bitácora.
Mensaje	VARCHAR2	Descripción del evento acontecido.
Paquete	VARCHAR2	Paquete que registra el evento.
Funcion_procedimiento	VARCHAR2	Función o procedimiento que registra el evento.

En esta tabla se registran por ejemplo, el inicio y fin de un periodo de carga del *data mart* o de las dimensiones.

4.3.4 Procedimiento de carga del *Data Warehouse*

4.3.4.1 Carga inicial de dimensiones estáticas

Algunas dimensiones no requieren ser actualizadas, sino que bastará con una carga inicial de las mismas. Estas dimensiones son:

Dimensión	Proceso de carga
DIM_RESULTADO	Pkg_carga.cargar_dim_resultado
DIM_DEMOGRAFIA	Pkg_carga.cargar_dim_demografia
DIM_ANIO	Pkg_carga.cargar_dim_anio

4.3.4.2 Carga de dimensiones y tablas de hecho

Las dimensiones no mencionadas en la sección anterior, así como las tablas de hechos requerirán ser actualizadas cada cierto tiempo, por ejemplo, semestral, anualmente o según las necesidades de los usuarios.

La siguiente tabla muestra las frecuencias de carga recomendadas para cada tabla de hechos:

Tabla de hechos	Frecuencia de carga
Inscripción	Anual
Graduados	Anual
Uso salones	Semestral
Asignaciones	Semestral

El procedimiento para cargar el *Data Mart* es el siguiente:

Ubicar en el sistema de archivos del servidor de base de datos, los archivos planos con los datos que se cargarán al *Data Mart*. Estos archivos planos son los archivos extraídos del sistema transaccional en Informix. La ruta del sistema de archivos sugerida para colocar dichos archivos es bajo el usuario oracle de Linux y su vía de acceso (path) es:

```
/home/oracle/dw_academico/archivos_fuente
```

Limpiar las tablas del área intermedia, que son aquellas tablas de la base de datos Oracle, en donde se cargarán los datos desde los archivos planos para su posterior procesamiento. Las tablas del área intermedia llevan el prefijo TMP

o INT en su nombre. Ejemplos de estas tablas son: TMP_CURSO, TMP_AREA, TMP_ESTUDIANTE, INT_AREA, etc. Esto se puede realizar invocando el procedimiento siguiente desde SQL*PLUS:

```
SQL > EXECUTE pkg_carga.limpiar_tablas_temporales;
```

Cargar todas las tablas del área intermedia desde los archivos planos, utilizando los archivos de control de SQL*LOADER. La ruta sugerida del sistema de archivos para ubicar los archivos de control es la siguiente:

```
/home/oracle/dw_academico/arch_ctl_sqlloader
```

Ejemplo:

Se desea cargar el archivo plano llamado curso.txt, hacia la tabla TMP_CURSO del área intermedia de la base de datos, por lo que se ubica dicho archivo, en la ruta arriba indicada.

Desde el shell del sistema operativo, ubicarse en el directorio bin del software de oracle. Por ejemplo así:

```
$cd /u01/app/oracle/product/10.2.0/db_1/bin
```

Ejecutar el comando de SQL*LOADER para cargar el archivo plano hacia la base de datos:

```
$ sqlldr userid=dwacademico/password  
control=/home/oracle/dw_academico/arch_ctl_sqlloader/curso.ctl  
log=/home/oracle/dw_academico/arch_ctl_sqlloader/curso.log  
data=/home/oracle/dw_academico/arch_ctl_sqlloader/curso.txt
```

Los parámetros del commando SQLLDR son los siguientes:

Parámetro	Significado
Control	Es el nombre del archivo de control de SQL*LOADER, que indica al comando SQLLDR cómo procesar el archivo de datos de entrada.
Log	Es el nombre del archivo en donde se registran los resultados de SQLLDR.
Data	Es el nombre del archivo de datos de entrada que se procesa.

Corroborar que se cargaron los datos en la base de datos, desde una sesión a la base de datos con SQL*PLUS o similar. Ejemplo:

```
SQL > SELECT * from TMP_CURSO;
```

Este proceso debería realizarse para todas las tablas del área intermedia.

Una vez cargadas todas las tablas TMP_* e INT_* se puede iniciar la carga del *Data Mart* mediante el proceso PKG_CARGA.CARGAR_DATA_WAREHOUSE, que se encarga de actualizar las dimensiones y las tablas de hecho. El proceso se invoca así:

Abrir una sesión de SQL*PLUS con el usuario DWACADEMICO. Ejecutar el proceso con los parámetros deseados:

```
SQL > EXECUTE PKG_CARGA.CARGAR_DATA_WAREHOUSE (p_anio,  
p_semestre);
```

En donde los parámetros p_año y p_semestre pueden tomar los siguientes valores:

p_semestre: '1' = primer semestre, '2' = segundo semestre. Nótese que los valores deben ir entre apóstrofes.

p_año: es el año completo como 2005, 2006 ó 2007. Nótese que *no* se utilizan apóstrofes.

Revisión de los resultados de las cargas. Una vez se han cargados los datos, se recomienda revisar el resultado de la carga consultando la bitácora (CTL_BITACORA) así:

```
SQL > SELECT fecha_hora, mensaje FROM ctl_bitacora ORDER BY 1;
```

Borrado de periodos de una tabla de hechos (fact table):

Si ocurriese un error al cargar un periodo de una o más tablas de hechos, es posible eliminar los datos del periodo equivocado mediante el siguiente procedimiento:

```
Pkg_carga.limpiar_facts_periodo(p_año VARCHAR2, p_semestre VARCHAR2);
```

En donde los parámetros p_año y p_semestre pueden tomar los siguientes valores:

p_semestre: '1' = primer semestre, '2' = segundo semestre. Nótese que los valores deben ir entre apóstrofes.

p_año: es el año completo como '2005', '2006' ó '2007'. Nótese que los valores deben ir entre apóstrofes.

4.3.4.3 Rutas para los archivos de configuración y de carga de datos de DWACADEMICO

Las siguientes son las rutas en el sistema de archivos, sugeridas para los archivos de carga y configuración del *data mart* DWACADEMICO, bajo el usuario oracle de Linux:

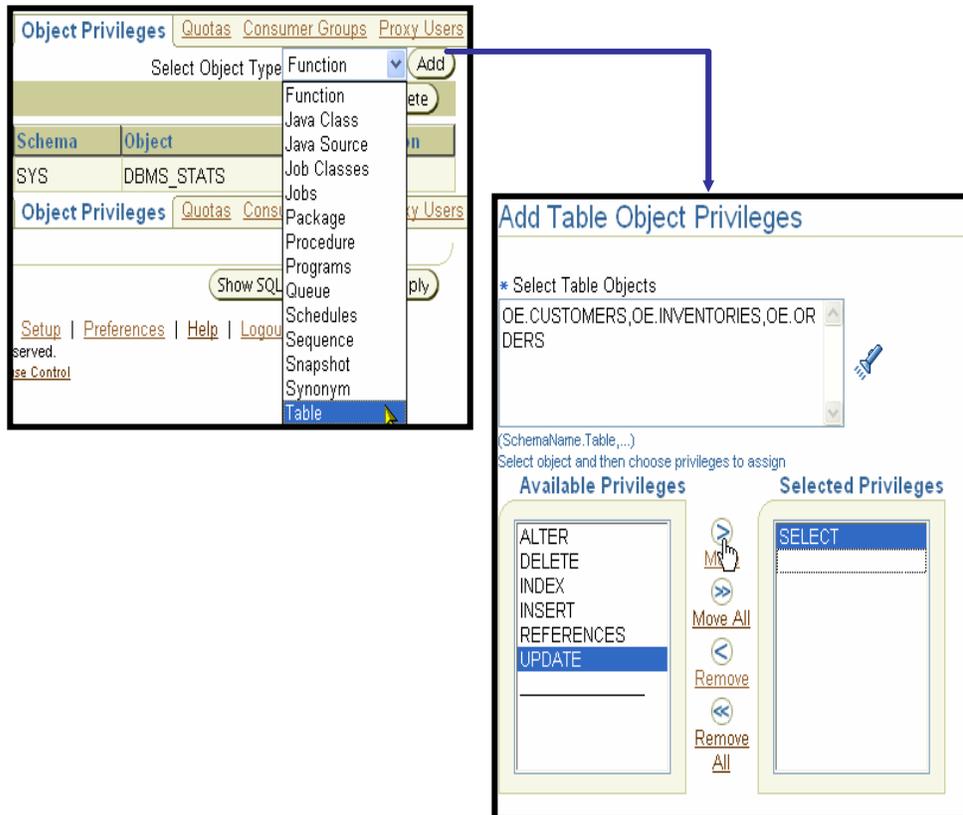
Ruta	Contenido
/home/oracle/dw_academico/archivos_fuente	Los archivos planos extraídos del sistema transaccional en Informix.
/home/oracle/dw_academico/arch_ctl_sqlloader	Archivos de control del commando SQL*LOADER para cargar las tablas del área intermedia desde los archivos planos.
/home/oracle/dw_academico/scripts_creacion	Scripts para crear los objetos a nivel de la base de datos del sistema DWACADEMICO.

4.4 Esquema de seguridad

4.4.1 Seguridad de la base de datos

Los datos del *Data Warehouse* están bajo un esquema de la base de datos al que sólo se puede acceder ingresando la contraseña apropiada. Cualquier usuarios sólo tendrá acceso a los datos si el administrador de la base de datos o de la aplicación otorga los privilegios explícitamente a los usuarios o a través de roles de la base de datos.

Figura 23. Manejo de seguridad sobre objetos de la base de datos

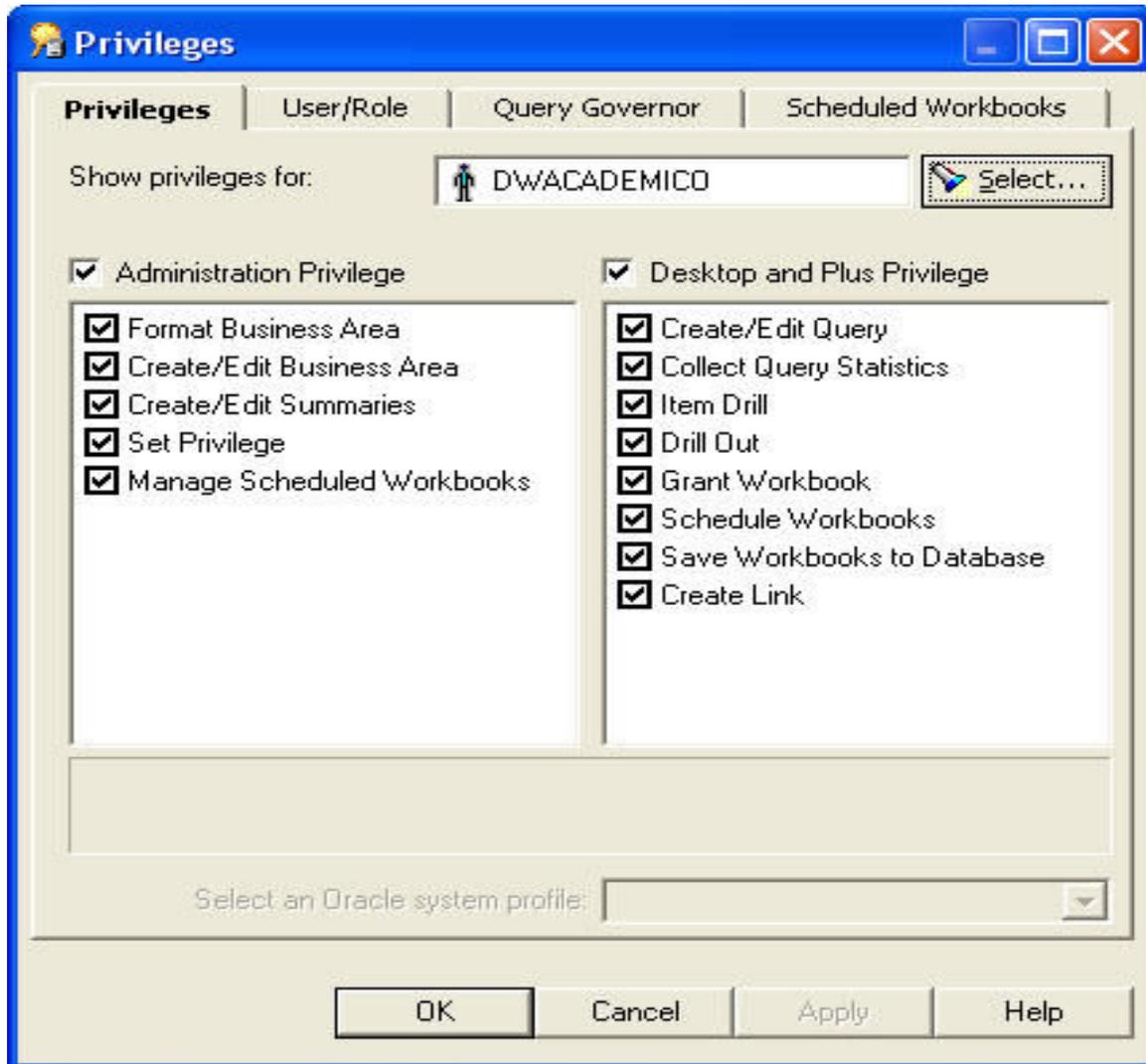


4.4.2 Seguridad de Discoverer Administrator

Además de los controles de acceso de la base de datos, Oracle Discoverer Administrator permite que el administrador de la aplicación defina qué usuarios tendrán acceso a las diferentes áreas de negocio que han sido creadas y a partir de allí pueda o no consultar la información contenida en cada *data mart*. Adicionalmente se cuenta con un esquema de seguridad para privilegiar a usuarios que puedan realizar ciertas acciones en el análisis de

información, tales como, poder modificar y salvar workbooks, programar la ejecución de los workbooks, profundizar en la información, entre otros.

Figura 24. Manejo de seguridad en herramienta Discoverer Administrator



4.4.3 Seguridad de Discoverer usuario final

La herramienta discoverer para usuarios finales, permite que los usuarios puedan dar permisos para uso de los reportes que han creado y salvado a la base de datos. Es decir, que cada usuario que tiene privilegios de salvar workboos y dar permisos sobre los mismos, podría dar acceso a otros usuarios de los workbooks que ha creado en el transcurso del tiempo.

4.4.4 Manejo de reportes

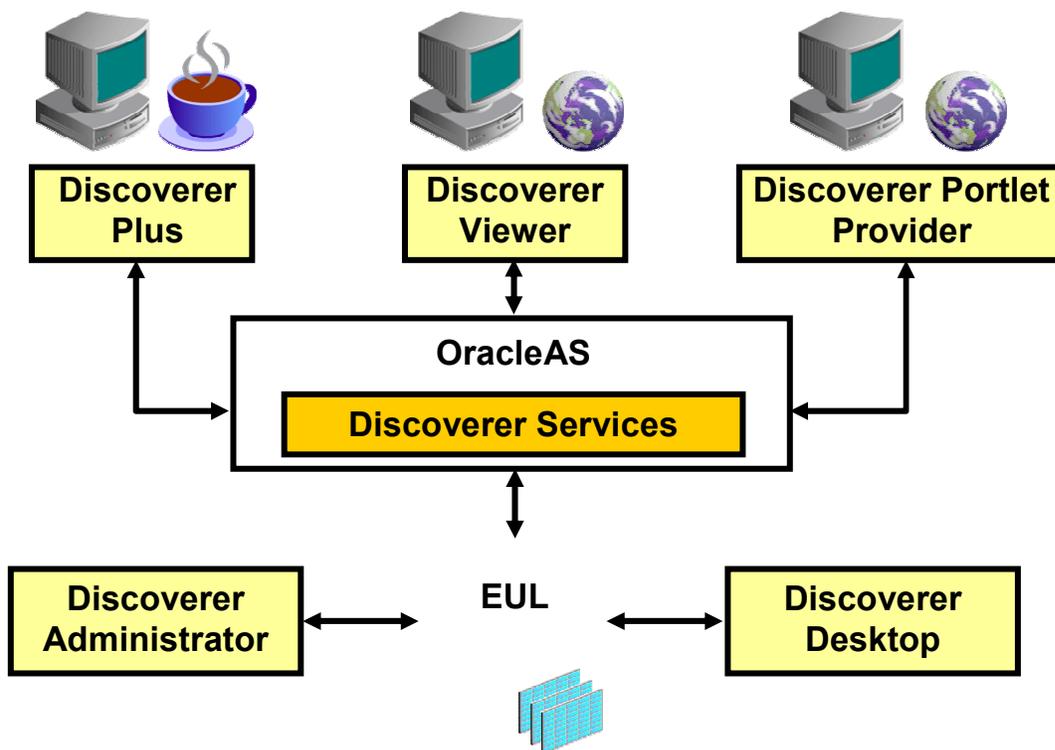
La herramienta discoverer de usuario final es la que le permite a los usuarios el poder crear sus propios escenarios de análisis. Existen una serie de reportes tipo plantialla que se han dejado preestablecidos y que fueron creado por el administrador del *data mart*. A estas plantillas tendrán inicialmente acceso los diferentes usuarios finales, y aquellos usuarios que tengan el privilegio de poder salvar workbooks, podrán modificar las plantillas predefinidas y salvarlas bajo su propio usuario en la base de datos, sin afectar las plantillas generales.

4.5 Presentación y entrega de la información contenida en el *data mart*

La herramienta de usuario final de inteligencia de negocio utilizada en el proyecto para la entrega de información es Oracle Discoverer, la cual es una herramienta muy flexible y de fácil uso para que los usuarios finales puedan crear de forma rápida y eficiente sus consultas de información. Esta

herramienta reside en el servidor de aplicaciones de Oracle, según lo descrito en la plataforma de inteligencia de negocio de la solución.

Figura 25. Componentes de la herramienta Oracle Discoverer



A continuación se presentan las diferentes plantillas que fueron creadas según las necesidades planteadas por la escuela de ciencias y sistemas en la fase inicial del proyecto. Estas plantillas no son más que una base para que los usuarios finales puedan obtener información de forma rápida, pero que al mismo tiempo sirvan de punto de partida para que los usuarios finales puedan crear sus propios escenarios de análisis.

Figura 26. Reporte de cantidad de años para graduarse

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

Arial 8 B U

Cantidad de Años para Graduarse (por Año de Inscripción)

Page Items: Carrera: INGENIERIA EN CIENCIAS Y SISTEMAS Año de Graduación: 2006

		1986	1988	1990	1992	1993	1994
▶ May	009212810	EDGAR ERNESTO	ENRIQUEZ MENA	-	14.34	-	-
	009312844	GABRIEL FERNANDO	CASTILLO CONTRERAS	-	-	13.34	-
	009416354	NEFTALI DE JESUS	CALDERON MENDEZ	-	-	-	12.34
	009516567	JULIO ANTONIO	QUINTANA GALINDO	-	-	-	-
	199811121	DANILO ESTUARDO	AC HERRERA	-	-	-	-
	199911117	JUAN RENE	SIMON ROBLES	-	-	-	-
	199911424	HECTOR ALBERTO HEBER	MENDIA ARRIOLA	-	-	-	-
	199911646	JUAN MIGUEL	INDEKEU RIVAS	-	-	-	-
	200010705	GUSTAVO ADOLFO	ALVARADO VILLATORO	-	-	-	-
▶ Jun	008612793	ODETTE CONCEPCION	ORDONNEZ ALFARO	20.43	-	-	-
	009212814	MANGLIO VINICIO R.	REYES CHAVEZ	-	14.42	-	-
	199811295	EDDY ROLANDO	VELASQUEZ CASTILLO	-	-	-	-
	199811612	LUIS RODOLFO	FARFAN MENDOZA	-	-	-	-
▶ Jul	008617497	JOSE MANUEL	GONZALEZ GOMEZ	20.51	-	-	-
	009012634	LUIS ALFREDO	GONZALEZ GOMEZ	-	16.51	-	-
	009230774	CRISTIAN EDUARDO	LAVARREDA ESTRADA	-	14.51	-	-
	009616789	WILLIAN STEVEN	MURALLES	-	-	-	-

Graduados | Uso salones | Asignaciones

Figura 27. Reporte de uso de salones

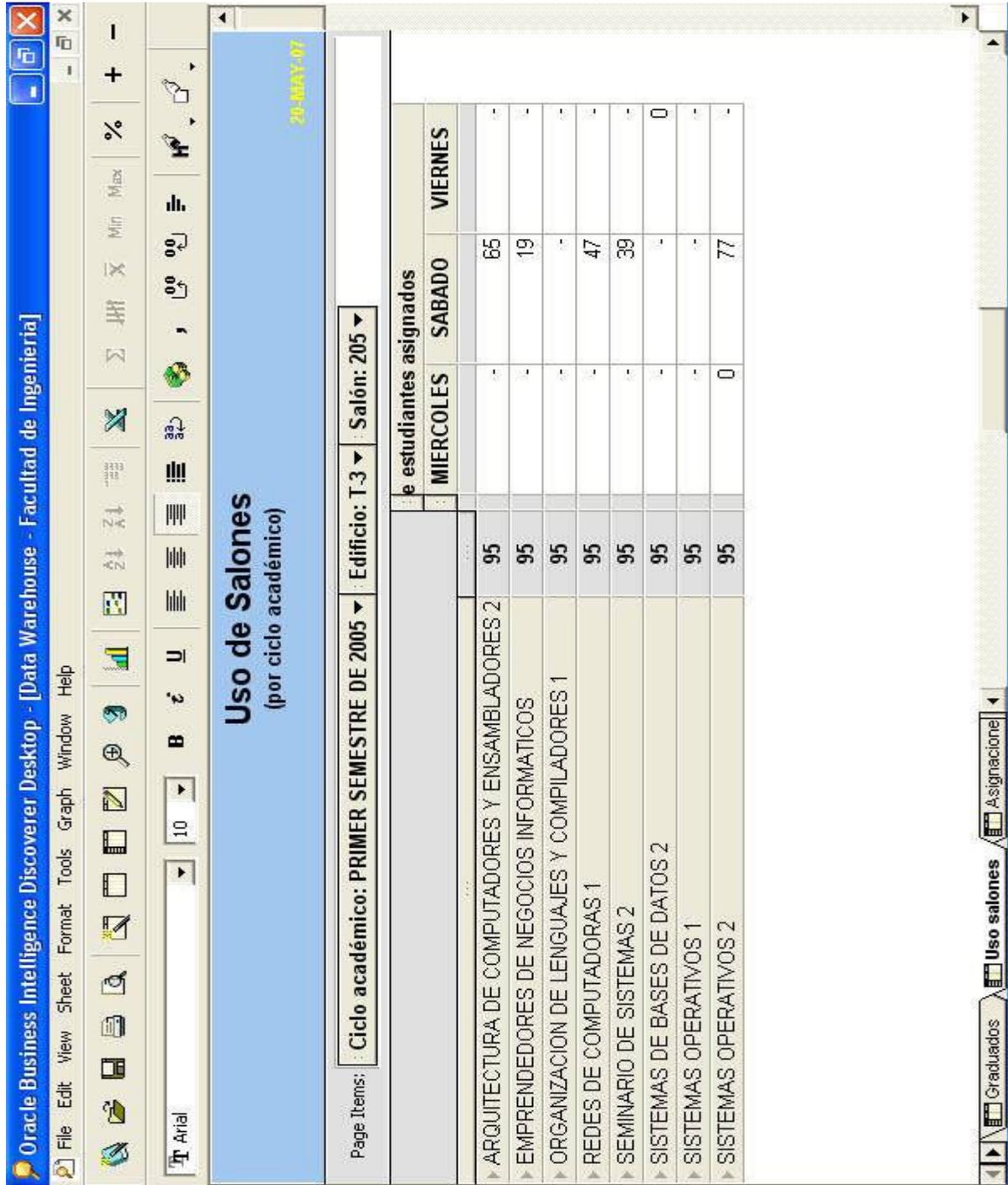


Figura 28. Reporte de asignación de cursos

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

Asignación de Cursos
(por inscripción y ciclo académico)

Page Items: Carrera: INGENIERIA EH CIENCIAS Y SISTEMAS Año inscripción: 2001 Ciclo académico: PRIMER SEMESTRE DE 2001

Catestrático	Catestrático	Carnet	Nombre	Apellidos	APROBADO			REPROBADO		
					Exam Final	Zona Curso	Nota final	Exam Final	Zona Curso	Nota final
LUIS ALBERTO	VITTORAZZI	200112537	BYRON ABEL	JAVIER RIVERA	-	-	-	0	21	21
	ESPAÑA	200112603	JOSE MIGUEL	RUIZ FINES	-	-	-	0	29	29
		200112739	PABLO ALEJANDRO	REYES OROZCO	-	-	-	9	38	47
		200112830	JUAN CARLOS	HERRERA SAN JOSE	-	-	-	0	23	23
		200112976	CLAUDIA ELIZABETH	MENDEZ ILLESCAS	-	-	-	0	33	33
		200113091	KARLA MELISSA	GARCIA BARNEOND	16	51	67	-	-	-
		200113105	RICARDO ISRAEL	MAZARIEGOS CASTILLO	13	56	69	-	-	-
		200113270	MARIA WJUALESKA	ALVAREZ BONILLA	15	52	67	-	-	-
		200113280	JOSE ALFREDO	GUDIEL MONZON	18	39	57	-	-	-
		200113369	ASTRID GUISELA	MENDEZ MEZA	17	36	53	-	-	-
		200113425	JOSE FERNANDO	OICHEITA CASTRO	-	-	-	0	6	6
		200117165	DANNY IVAN	MONTUFAR MAYORGA	-	-	-	0	29	29
		200117556	DIANA PATRICIA	MAZARIEGOS SANCHEZ	-	-	-	7	43	50
		200130550	RUDY ABRAHAM	CASTANEDA SANTIZ	-	-	-	0	28	28
WALTER ORLANDO	FIGUEROA CHAVEZ	200110203	MYNOR WILFREDO	PERALTA HERRERA	-	-	-	0	34	34
		200110612	VICTOR DANIEL	MERIDA DONIS	-	-	-	0	0	0

Graduados | Uso salones | Asignación

Figura 29. Reporte de demografía de estudiantes

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

Arial 8 B U

Demografía de Estudiantes (por Año de Inscripción)

20-03-2007

Page Items: Año reinscripción: 2006 Género: MASCULINO Rango Edad: 26 - 30

009520722	BYRON DANIEL	LOPEZ ROBLES	▶ 19-APR-1977
009613494	GERMAN ANIBAL	GIL LAROU	▶ 29-DEC-1976
009630392	EDGAR NERY	BATZ MARIN	▶ 16-JUN-1977
009630402	HUGO RENE	YAX ORDONEZ	▶ 10-SEP-1976
009712012	SERGIO	GALLARDO CONTRERAS	▶ 05-JUL-1978
009712030	CARLOS DAVID	QUEZADA LORENZANA	▶ 19-OCT-1977
009712072	HERBERT RAUL	MOLINA MORALES	▶ 02-NOV-1978
009712112	HECTOR FERNANDO	SANTOS JUAREZ	▶ 08-AUG-1978
009712169	BRYAN ALEXIS	ORELLANA SOBERANIS	▶ 28-JAN-1978
009712266	FRANCISCO VALENTIN	CASTELLANOS GUTIERREZ	▶ 05-JUL-1979
009712269	MAYKOL DAVID	LUNA MUNOZ	▶ 20-MAY-1979
009712278	ERICK ESTUARDO	RODRIGUEZ SIERRA	▶ 16-JUL-1979
009712310	JUAN ANTONIO	MARTINEZ MORALES	▶ 12-DEC-1979
009712312	CARLOS ESTUARDO	MORALES LEMUS	▶ 20-DEC-1978
009712370	JUAN CARLOS	MAEDA JUAREZ	▶ 09-JUN-1979
009712386	EDWARD EDMUNDO	VELASQUEZ RODRIGUEZ	▶ 19-APR-1978

Graduados | Uso salones | Asignaciones de cursos | Datos demográficos estudiantes | Historial de cur

Figura 30. Reporte de historial de cursos por estudiante

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

20-NOV-07

Historial de Cursos por Estudiante

(por Año de Inscripción)

Page Items: Año de inscripción: 2000 ▶ Carnet: 200011660 ▶ Nombres estudiante: EDGAR OSWALDO ▶ Apellidos estu... ▶

	Fecha Asignación	Curso	Examen Final	Zona Curso	Nota final	Resultado
▶ 1	09-NOV-2003	INTRODUCCION A LA PROGRAMACION Y COMPUTACION 1	0	11	11	REPROBADO
▶ 2	10-MAR-2004	INTRODUCCION A LA PROGRAMACION Y COMPUTACION 1	0	0	0	REPROBADO
▶ 3	29-NOV-2004	LOGICA DE SISTEMAS	0	0	0	REPROBADO
▶ 4	08-MAR-2005	LOGICA DE SISTEMAS	16	68	84	APROBADO
▶ 5	08-MAR-2005	INTRODUCCION A LA PROGRAMACION Y COMPUTACION 1	0	0	0	REPROBADO

Historial de cursos x estudiante | Asignaciones por Catedrático | Reprobados | Ranking de estudiantes

Figura 31. Reporte de asignaciones por catedrático

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

Arial 10 B U

20-05-07

Asignaciones por Catedrático

Page Items: Nombres catedrático: FRANCISCO JAVIER Apellidos catedrático: GUEVARA CASTILLO Resultado:

	PRIMER SEMESTRE DE 2004	PRIMER SEMESTRE DE 2005	PRIMER SEMESTRE DE 2006	PRIMERA RETRASADA, PRIMER SEMESTRE DEL AÑO 2003	PRIMERA RETRASADA/ PRIMER SEMESTRE D AÑO 2005
▶ SISTEMAS DE BASES DE DATOS 1	38	39	36	-	27
▶ SISTEMAS DE BASES DE DATOS 2	-	-	-	-	-
▶ SISTEMAS OPERATIVOS 1	32	17	34	4	11
TOTAL DE ESTUDIANTES	70	56	70	4	38

Historial de cursos x estudiante Asignaciones por Catedrático Reprobados Ranking de estudiantes

Figura 32. Reporte de reprobados por curso

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingenieria]

File Edit View Sheet Format Tools Graph Window Help

20.03.07

Reprobados por Curso

Page Items: Resultado: REPROBADO

	Cantidad de Estudiantes						
	1993	1994	1995	1996	1997	1998	1999
▶ LENGUAJES FORMALES Y DE PROGRAMACION	31	33	33	40	49	130	143
▶ LOGICA DE SISTEMAS	-	4	1	14	21	14	39
▶ SISTEMAS DE BASES DE DATOS 1	-	17	45	80	96	82	87
▶ ANALISIS Y DISEÑO DE SISTEMAS 1	17	12	24	36	35	52	31
▶ ANALISIS Y DISEÑO DE SISTEMAS 2	13	11	14	23	36	24	17
▶ ARQUITECTURA DE COMPUTADORES Y ENSAMBLADORES 1	9	16	23	45	43	81	57
▶ ARQUITECTURA DE COMPUTADORES Y ENSAMBLADORES 2	1	13	16	39	37	51	34
▶ ECONOMIA	24	18	25	51	48	63	47
▶ EMPRENDEDORES DE NEGOCIOS INFORMATICOS	1	2	5	1	2	6	1
▶ ESTRUCTURAS DE DATOS	16	65	43	78	46	113	118
▶ INTELIGENCIA ARTIFICIAL 1	19	11	31	40	39	29	26
▶ INTELIGENCIA ARTIFICIAL 2	1	1	1	-	3	3	-
▶ INTRODUCCION A LA PROGRAMACION Y COMPUTACION 1	16	-	27	43	60	64	83
▶ INTRODUCCION A LA PROGRAMACION Y COMPUTACION 2	25	32	35	41	37	104	92
▶ MANEJO E IMPLEMENTACION DE ARCHIVOS	7	28	33	101	74	81	49
▶ MODELACION Y SIMULACION 1	7	23	24	33	45	54	39

Historial de cursos x estudiante

Asignaciones por Catedrático

Reprobados

Ranking de estudiantes

Figura 33. Reporte de ranking de estudiantes por curso

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingenieria]

File Edit View Sheet Format Tools Graph Window Help

Arial 8 B U

Ranking de Estudiantes por Curso (por inscripción y ciclo académico)

Page Items: Año inscripción: 2005 Ciclo académico: PRIMER SEMESTRE DE 2006 Curso: INTRODUCCION A LA PROGRAMACION

Carnet	Nombre	Apellidos	Examen Final	Zona Curso	Nota final	Rank
200512045	JORGE RAUL	LU HERNANDEZ	20.00	64.00	84.00	1
200512059	PABLO SERGIO	ROMERO VELIZ	20.00	60.00	80.00	2
200511691	LUIS FERNANDO	JUAREZ AVILA	19.00	60.00	79.00	3
200512049	MARIO ALBERTO	JIMENEZ SALGUERO	20.00	59.00	79.00	3
200511636	JOSE GERARDO	GARCIA PINEDA	21.00	57.00	78.00	5
200511905	JULIO EDUARDO	MORALES TOLEDO	18.00	60.00	78.00	5
200511972	OSCAR OSWALDO	CERNA FAJARDO	20.00	57.00	77.00	7
200512079	VICTOR LEONEL	OROZCO LOPEZ	19.00	58.00	77.00	7
200511667	CARLOS ALEJANDRO	SOLORZANO ROLDAN	20.00	56.00	76.00	9
200512000	NERY MARGARITO	CHUCUY MILIAN	18.00	58.00	76.00	9
200512017	GUSTAVO WALDEMAR	GARCIA CASTELLANOS	19.00	57.00	76.00	9
200512252	LUIS EDUARDO	CORDON ALVIZURES	19.00	56.00	75.00	12
200512141	FRANCISCO ALFONSO	PINEDA JIMENEZ	17.00	57.00	74.00	13
200515877	HECTOR ADOLFO	ALVARADO VILLATORO	15.00	57.00	72.00	14
200511795	CARLOS ROBERTO	SANDOVAL REYES	19.00	52.00	71.00	15
200515984	JUAN LUIS	BLANCO DUCOUDRAY	18.00	53.00	71.00	15
200517701	LUIS PEDRO	ESTRADA CASAS	17.00	54.00	71.00	15

Historial de cursos x estudiante | Asignaciones por Catedrático | Reprobados | Ranking de estudiantes

Figura 34. Reporte de cantidad de graduados

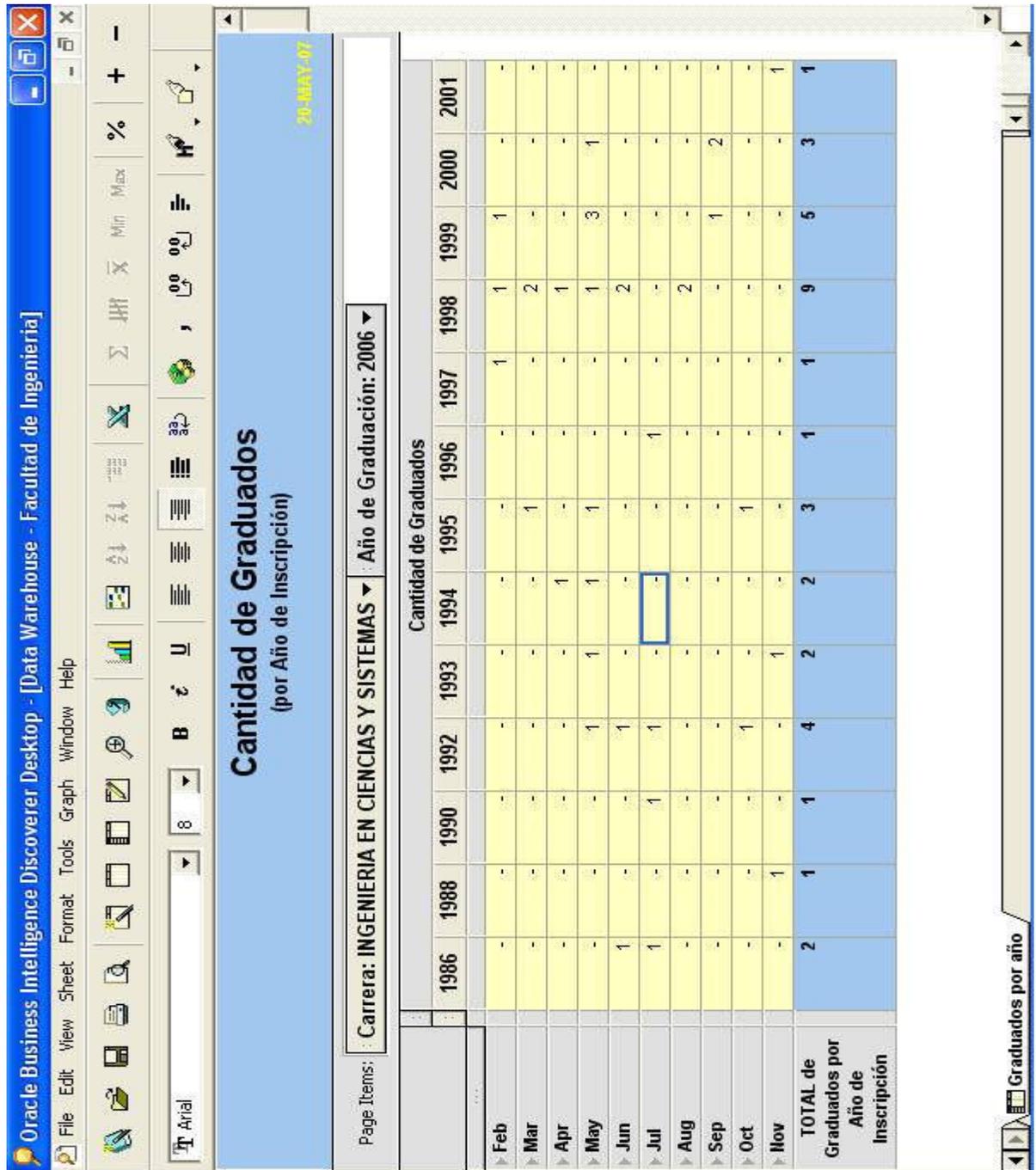


Figura 35. Reporte de estudiantes por curso y año de inscripción

Oracle Business Intelligence Discoverer Desktop - [Data Warehouse - Facultad de Ingeniería]

File Edit View Sheet Format Tools Graph Window Help

29-03-07

Análisis de Estudiantes por Curso y Año de Inscripción

Page Items: Carrera: INGENIERIA EN CIENCIAS Y SISTEMAS Año de Inscripción: 2000

	PRIMER SEMESTRE DE 2003		PRIMER SEMESTRE DE 2004		PRIMER SEMESTRE DE 2005		Cantí PRIMER S 2
	APROBADO	REPROBADO	APROBADO	REPROBADO	APROBADO	REPROBADO	
ARQUITECTURA DE COMPUTADORES Y ENSAMBLAJE	4	8	10	10	4	8	7
ARQUITECTURA DE COMPUTADORES Y ENSAMBLAJE	-	12	9	1	7	5	3
ECONOMIA	8	5	2	7	6	12	11
EMPREENDEDORES DE NEGOCIOS INFORMATICOS	-	-	-	-	-	-	3
ESTRUCTURAS DE DATOS	7	11	1	14	4	7	1
INTELIGENCIA ARTIFICIAL 1	-	-	3	-	7	1	7
INTELIGENCIA ARTIFICIAL 2	-	-	-	-	-	-	-
INTRODUCCION A LA PROGRAMACION Y COMPUTA	2	18	5	26	3	12	-
INTRODUCCION A LA PROGRAMACION Y COMPUTA	6	19	6	8	10	6	1
LENGUAJES FORMALES Y DE PROGRAMACION	9	22	4	11	2	12	4
LOGICA DE SISTEMAS	5	3	7	9	2	5	2
MANEJO E IMPLEMENTACION DE ARCHIVOS	7	3	7	11	3	9	2
MODELACION Y SIMULACION 1	-	-	-	-	-	-	4
MODELACION Y SIMULACION 2	-	-	-	2	4	8	2
Asignaciones por Catedrático 2	-	-	-	-	-	-	-

CONCLUSIONES

1. La implementación de un proceso de inteligencia de negocio en una organización, permite que la información fluya de una forma natural y controlada desde donde se producen las transacciones del día a día de la organización, hasta convertirlas en información y conocimiento que permiten a los usuarios finales tomar mejores y efectivas decisiones.
2. La creación de un *data warehouse* corporativo en la organización es una tarea ardua pero esencial cuando se desea hacer análisis de información, dado que permite centralizar toda la información relevante para la toma de decisiones, y se constituye en la fuente única de información para toda la organización.
3. En la escuela de ciencias y sistemas se obtendrán grandes beneficios al utilizarse el *data mart* académico, puesto que se podrá analizar el comportamiento de los estudiantes de la escuela y por ende se podrá tomar mejores decisiones en cuanto al uso de los recursos y el enfoque de la carrera de sistemas.
4. La creación del presente *data mart* permitirá que las diferentes áreas o unidades de la facultad cuenten con información académica de una forma autónoma, sin que exista la dependencia de centro de cálculo, siempre guardando los debidos controles de seguridad y acceso a la información.

5. La utilización del servidor de aplicaciones de *Oracle*, permite que los usuarios finales puedan conectarse desde cualquier parte de la facultad sin tener que instalar ningún *software*-cliente para poder hacer análisis de la información, así también, podrían hacerlo desde cualquier parte del mundo, siempre y cuando el servidor sea liberado para uso por medio de *Internet*, guardando los respectivos niveles de seguridad que provee la infraestructura de la facultad.

6. Contar con una herramienta de inteligencia de negocio en la Escuela de Ciencias y Sistemas, facilitará la información que muchas empresas requieren sobre referencias de los estudiantes que solicitan empleos. Así también, permitirá que la asignación de carga de estudiantes sea mejor distribuida entre los catedráticos, incluso entre otros recursos tales como los salones.

RECOMENDACIONES

1. El presente trabajo ha servido para la construcción de un primer *data mart* orientado a información académica de la Escuela de Ciencias y Sistemas, por lo que se recomienda que la solución siga creciendo inicialmente con la adición de información de todas las escuelas de la Facultad de Ingeniería, al modelo ya construído.
2. Como una segunda fase del proyecto, se debe consolidar la capacitación de los usuarios finales y tomadores de decisiones en cuanto al uso de las herramientas y al contenido del *data mart* desarrollado.
3. Se debe dar seguimiento a la carga de información hacia el *data mart*, a fin de que se mantenga actualizado y realmente sea de beneficio para los usuarios finales y tomadores de decisiones, tanto en la Escuela de Sistemas, como para las demás autoridades de la facultad.
4. Como parte del ciclo de vida de la inteligencia de negocio, es recomendable que se cree un comité que vele por el seguimiento de mejoras y adiciones al *data warehouse*, hasta lograr integrar a las diferentes áreas de negocio de la facultad, y en el largo plazo incluso tratar de integrar adiciones a nivel de toda la universidad. También es importante que este comité vele por el control de calidad de la información que se tendrá disponible dentro de las estructuras del *data warehouse*, a

fin de cerciorarse que la información allí contenida refleje la realidad de lo acontecido.

5. Se sugiere que el personal de la Escuela de Sistemas y de Centro de Cálculo que ha sido capacitado, se constituya en capacitador de los demás entes que pudiesen beneficiarse con el uso de la información contenida en el *data mart*.
6. Sería de mucho beneficio que se pudiese crear otros ejercicios de práctica supervisada orientados a acrecentar la solución que se ha dejado desarrollada por el momento, a fin de que personas con experiencia y con el deseo de graduarse, puedan aportar sus conocimientos para mejorar y engrandecer la solución de inteligencia de negocio que se está iniciando en la facultad.

REFERENCIAS ELECTRÓNICAS

1. Para más información sobre conceptos y tecnología de inteligencia de negocio, consultar <http://www.bi-spain.com>
2. Para más información sobre productos *Oracle* para inteligencia de negocio se puede consultar <http://www.oracle.com/technology/products/bi/index.html> o <http://www.oracle.com/technology/tech/bi/index.html>

BIBLIOGRAFÍA

1. Patrones de diseño, E. Gamma et al. Editorial Addison-Wesley.
2. Kimball, Ralph. "*The Data Warehouse Toolkit series*". Editorial John Wiley & Sons. 1996-2004.
3. W. H. Inmon. "*Building the Data Warehouse*". Editorial Wiley, Tercera Edición. 2002.
4. Bitam. "*Business Intelligence*". [en línea].
<http://www.bitam.com/spanish/AcercaDeBI.htm> [Consulta: Febrero de 2007].
5. Oracle Corp. "*BI Solutions*", [en línea]. <http://www.oracle.com> [Consulta: enero de 2007].
6. Poe, Vidette; Klauer, Patricia; Brobst, Stephen. "Building a Data Warehouse for Decision Support". Editorial Prentice Hall. 1998.
7. Sperley, Eric. "*The Enterprise Data Warehouse: Planning, Building, and Implementation*". Prentice Hall. 1999.
8. Hobs, Lilian; Smith, Pete. "*Oracle Database 10g Data Warehousing*". Editorial Elsevier Digital Press. Oracle Corporation. 2004.