

UNIVERSIDAD MAYOR DE SAN ANDRES
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMATICA



TESIS DE GRADO

“MODELO INTEGRADOR DE DATOS ACTIVOS EN INTELIGENCIA EMPRESARIAL PARA LA TOMA DE DECISIONES”

PARA OPTAR AL TITULO DE LICENCIATURA EN INFORMATICA
MENCION: INGENIERIA DE SISTEMAS INFORMATICOS

POSTULANTE: DAVID IVAN DAVALOS CENTELLAS
DOCENTE TUTOR: Lic. Freddy Miguel Toledo Paz
DOCENTE REVISOR: Lic. Grover Alex Rodríguez Ramírez

La Paz – Bolivia
2009

DEDICATORIA

Mi presente investigación de tesis que me permite acceder al título de Licenciado en Informática está dedicado exclusivamente a los seres mas queridos y cercanos a mi vida, que es mi familia, mi mamá Dora de Dávalos, mi papá Antonio Dávalos, mis hermanos Dossy Dávalos, Marco A. Dávalos mi cuñada Pamela Oressana mis sobrinas Sheril y Samira, y a mi novia con mucho amor Giovanna Vargas.

Gracias por su apoyo incondicional de todos los días.

Agradecimientos

Un agradecimiento muy especial con mucho significado en esta etapa de mi vida al Lic. Grover Alex Rodríguez Ramírez quien con su paciencia y su tiempo supo corregir mis errores cometidos durante el proceso de investigación en la elaboración de la Tesis de Grado como mi directo revisor, de todo corazón muchas gracias Licenciado por apoyarme y guiarme en todo momento.

Lic. Freddy Miguel Toledo Paz muchas gracias por brindar su tiempo y comprensión y lo mas importante guiarme como docente tutor y aconsejarme en las interrogantes que presentaba en mi investigación.

Por la colaboración de información y de respuesta con la experiencia y trabajo directo en este tipo de investigación al Lic. Wilson Cuellar

A la Universidad Mayor de San Andrés quien nos brinda la posibilidad de realizar estudios superiores y en especial a la carrera de Informática por permitirme conseguir un título profesional que me ayuda en el desarrollo de mi persona.

RESUMEN

El presente documento presenta una investigación sobre la integración de diferentes gestores de base de datos, particularmente el momento de la heterogeneidad que se presentan el momento de compartir datos de diferentes fuentes.

Ya que esta heterogeneidad dificulta el proceso de integración de base de datos heterogéneas, por lo cual se propone un modelo de integración de datos activos en inteligencia empresarial para la toma de decisiones que permite ampliarse en futuras investigaciones.

De este caso la propuesta esta orientada a tratar el caso específico de dos bases de datos administradas por dos sistemas gestores de base de datos diferentes, en las cuales se realizan transacciones, manteniendo la consistencia de las base de datos en todo momento y accediendo a un entorno amigable para su fácil administración de los datos y la toma de decisiones.

INDICE ESPECIFICO

DEDICATORIA	i
AGRADECIMIENTOS	ii
RESUMEN	II

INDICE

Pagina

Capitulo I

PERFIL DE TESIS

1. Introducción	1
1.2 Antecedentes	1
1.3 Planteamiento del problema	2
1.3.1 Problemas Específicos	3
1.4 Objetivo	3
1.4.1 Objetivo General	3
1.4.2 Objetivos Específicos	3
1.5 HIPOTESIS	4
1.6 Justificación	4
1.6.1 Justificación Teórica	4
1.6.2 Justificación Tecnológica	4
1.6.3 Justificación Social	5
1.6.4 Justificación Económica	5
1.7 Alcances y Limites	5
1.7.1 Alcances	6
1.7.2 Limites	6
1.8 Aporte	6
1.9 Metodología científica	6
1.10 Herramienta (Lenguaje Unificado de Modelado) UML	7

INDICE

Pagina

Capitulo II

MARCO TEORICO

MARCO TEORICO	8
2. MARCO TEORICO	9
2.1. Sistema de Base de Datos	9
2.1.1 Base de Datos	9

2.2.2 Sistema Gestor de Base de Datos (SGBD)	10
2.2.3 Modelo Relacional	11
2.3 Integración de Datos	11
2.4. Extracción y conversión de datos	12
2.5 Sistema Multi Base (Datawarehouse)	12
2.5.1 Los objetivos fundamentales de un Sistema multibase	12
2.5.2 Los elementos básicos de un sistema Multibase	13
2.5.3 Los procesos básicos del sistema multibase	14
2.5 Tipo de Conexiones a Gestores de Base de Datos	15
2.6.1JDBC	15
2.6.2 ODBC	15
2.5 METODOLOGÍA UML	16
2.7.1 Modelo de Cosos de Uso	16
2.7.2 Diagrama de Casos de Uso	16
2.7.3Elementos	16
2.7.4 Actores	17
2.7.5 Modelo de Análisis	17
2.7.6 Diagrama de Secuencia	17
2.7.7 Diagrama de Clases	18
2.7.8 Diagrama de Colaboración	19
2.7.9 Diagrama de Estados	19
2.5 HERRAMIENTAS DE INVESTIGACIÓN	20
2.5.1 SLQ Server	20

INDICE

Página

Capitulo III

MARCO APLICATIVO

3 MARCO APLICATIVO	23
3.1 DESCRIPCION INFORMAL	23
3.2 DESCRIPCION FORMAL	28
3.2.1 Gestor de Base de Datos	28
3.3 Extracción, Transformación y Cargado	30
3.3.1 Generación de los Archivos (o Transformación del archivo)	31
3.3.2 Cargado de Archivos	33
3.4 Tipo de Almacenamiento	33
3.5 Vista Usuario	34
3.6 DEFINICION DE LOS CASOS DE ESTUDIO	34
3.6.1 PLICACION DEL PROCESO DE INTEGRACION	37
3.6.2PREINTEGRACION	37

3.6.3 COMPARACION	38
3.6.4 CONFORMACION	40
3.6.5 RESTRUCTURACION Y ASOCIACION	41
3.6.6 APLICACIÓN DEL ESQUEMA INTEGRADO	42
3.7 DEMOSTRACION DE HIPOTESIS	43

INDICE		Pagina
Capitulo IV		
CONCLUSIONES Y RECOMENDACIONES		
CONCLUSIONES Y RECOMENDACIONES		44
4.1 CONCLUSIONES		44
4.2 RECOMENDACIONES		45
REFERENCIAS BIBLIOGRAFICAS		46
ANEXOS		48
DOCUMENTACION		49

INDICE DE GRAFICAS		
GRAFICO		PAGINA
Figura 2.1 Base de Datos		10
Figura 2.2 principales componentes de un gestor de base de datos.		10
Figura 2.3 Modelo Relacional Según C. J. DATE.		11
Figura 2.9 Casos de USO		16
Figura 2.10 Diagrama de Secuencia		18
Figura 2.11 Diagrama de Clases		18
Figura 2.12 Diagrama de Colaboración		19
Figura 2.13 Diagrama de Estados		20
Figura 2.14 SLQ Server		20
Fig. 3.1.1 Figura de extracción y conversión de nueva relación		24
Fig. 3.1.2 Modelo Integrador de Datos Activos en Inteligencia Empresarial para la Toma de Decisiones		27
Fig. 3.2 Modelo Integrador de Datos Activos en		

Inteligencia Empresaria para la Toma de Decisiones formal	28
Fig. 3.2.1 Arquitectura de la Base de Datos	28
Fig. 3.2.1 Esquema de Gestor de Base de Datos	29
Fig. 3.2.2 Esquema Heterogéneo de Base de Datos	30
Fig. 3.3 Extracción Transformación y Cargado	30
Fig. 3.3 .1 Vista formada según el uso de operadores del algebra relacional	31
Fig. 3.3 .2 Transformación	31
Fig. 3.3.3 Cargado de Archivo de Vista Final	32
Fig. 3.3.4 Cargado de Archivo	33



1. Introducción

Actualmente se observan nuevas tendencias de evolución en los modelos de integración de sistemas. El Modelo de Integración de Datos activos en Inteligencia Empresarial explícitamente permite a las empresas o instituciones públicas una oportuna toma de decisiones mediante el manejo de información.

El impacto en las empresas o instituciones públicas ven que la información compartida y la destrucción de las barreras entre las diferentes áreas que manejan diferentes fuentes de datos generan ventajas competitivas.

Las empresas o instituciones públicas conformadas por sus diferentes áreas trabajan independientemente y para su único beneficio, cada área se encierra en su propio objetivo sin permitir la comunicación y el compartimiento de información con las otras áreas ya que este es un comportamiento aceptable, pero ocasiona que la empresa o institución no sea muy competitiva, es por eso que las áreas necesitan asociarse o integrarse en el compartimiento de datos para la toma de decisiones oportunas, es así que se ve la necesidad de proponer un modelo que permita mejorar el manejo de la información proveniente de las diferentes fuentes de datos aplicadas en la integración de diferentes sistemas, esto como una estrategia para la toma de decisiones.

1.2 Antecedentes

La extracción de datos, el manejo de información de las diferentes fuentes de base de datos con los que cuenta una empresa o institución pública aun sigue siendo de forma tradicional, trabajan independientemente y se realiza de forma manual

Provocando pérdida de tiempo, inseguridad, retardo de trabajo, pérdida de información y elevando los gastos dentro de las mismas, estos problemas con los que atraviesan las instituciones públicas y privadas no les permiten ser muy competitivas.

En la actualidad nos encontramos en un ambiente cada vez más competitivo, donde las empresas e instituciones públicas requieren de herramientas sólidas

que las asistan a la toma de decisiones que puedan traer beneficios y mejoras en sus procesos.

El modelo Integrador de Datos Activos permite otorgar apoyo a los clientes del negocio, tiempos rápidos y respuestas a sus problemas con eficiente manejo de información que permita la oportuna toma de decisiones y disminución de los costos totales de operaciones.

De estas consideraciones surge la necesidad de construir un modelo Integrador de Datos Activos en Inteligencia Empresarial para la Toma de Decisiones que permita la extracción y el agrupamiento de los datos.

1.3 Planteamiento del problema

El problema se origina debido a que las empresas e instituciones públicas en el país, no analizan los beneficios y ventajas estratégicas que proporciona la tecnología para el apoyo a la toma de decisiones.

Provocando pérdida de tiempo, inseguridad, retardo de trabajo, pérdida de información y elevando los gastos dentro de las mismas, este tipo de estructuras organizacionales de las empresas o instituciones públicas no les permiten ser muy competitivas.

De ahí la importancia de crear un modelo Integrador de Datos Activos eficiente para el manejo de información que permita la toma oportuna de decisiones dentro de una empresa o una institución pública.

¿De que manera se puede lograr que los procesos de extracción de datos que ocasiona un movimiento rápido eviten las inconsistencias y pérdida de información de forma segura entre origen y destino de los diferentes sistemas de base de datos de manera que pueda ayudar a la toma de decisiones?

1.3.1 Problemas Específicos

- a) El manejo de datos inconsistentes ocasiona una mala toma de decisiones.
- b) La información generada no se actualiza a tiempo causando una ralentización en la toma de decisiones.
- c) Dificultad de manipular los datos con respecto a la heterogeneidad de los mismos.
- d) Los reportes generados son lentos y tediosos en su elaboración
- e) Poca seguridad de información ocasiona dudas y pérdida de datos en la toma de decisiones.
- f) No se conoce los diferentes sistemas de bases de datos y la importancia de las mismas.

1.4 Objetivo

1.4.1 Objetivo General

Desarrollar un modelo Integrador de Datos Activos en Inteligencia Empresarial que sea eficiente para el manejo de información logrando procesos de extracción de datos que ocasione un movimiento rápido y seguro, evitando inconsistencias y pérdida de información entre origen y destino de los diferentes sistemas de base de datos de manera que pueda ayudar a la toma de decisiones.

1.4.2 Objetivos Específicos

- Plantear un esquema homogéneo de base de datos heterogéneas
- Diseñar un modelo integrador de base de datos heterogéneas que soporte el proceso de extracción de datos
- Diseñar un modelo de control de información estable para las variables que afectan el manejo de datos entre los diferentes sistemas de base de datos.
- Construcción de un prototipo de manejo de información entre diferentes sistemas de base de datos
- Diseñar un modelo de comunicación para la integración de base de datos heterogéneos.

1.5 HIPOTESIS

El modelo Integrador de Datos Activos en Inteligencia Empresarial para la toma de Decisiones basado en el manejo de información relacionada con variables que afectan a la extracción de datos de un origen a un destino de diferentes sistemas de base de datos sera capaz de eliminar inconsistencias y pérdida de información de forma rápida y segura.

1.6 Justificación

La perspectiva de las diferentes empresas especializadas en el desarrollo de herramientas de extracción y conversión de datos integradoras de sistemas se agranda y pasamos de ver, estas herramientas como un medio utilizable dentro de Inteligencia Empresarial como una importante herramienta de gestión empresarial global para la toma de decisiones, que nos acerca a la realidad del trabajo de la integración general de los diferentes fuentes de información dentro una empresa o una institución publica.

El presente trabajo proporcionara facilidades para mejorar la toma de decisiones que lleven al éxito a empresas o instituciones públicas.

1.6.1 Justificación Teórica

EL modelo de datos activos requiere de la extracción de las diferentes fuentes de información que es la base teórica para la construcción del sistema multibase como también la conversión a un lenguaje estándar que implica en este caso al mejo de archivos planos de texto

1.6.2 Justificación Tecnológica

La aplicación y conmutación de paquetes, la utilización de computadoras con los respectivos lenguajes instalados para su desarrollo y Licencias Autorizadas son las principales herramientas para el presente trabajo.

Una investigación técnica dentro del modelo integrador de datos activos para la ayuda a la toma de decisiones nos permite conocer la extracción de datos y conexión entre diferentes sistemas de base datos, conversión de los datos a un lenguaje estándar para la construcción de un datawarehouse, que permite un manejo fácil y seguro de información.

1.6.3 Justificación Social

El trascendido social que tendrá este trabajo de investigación básicamente será reflejado en el beneficio directo a las empresas o instituciones públicas. Con la tendencia al manejo rápido, seguro y eficaz de los datos con los que se trabaja.

Desarrollar un modelo de datos activos para la oportuna toma de decisiones es un aporte a las estructuras organizacionales de las empresas que necesitan trabajar en equipo para que puedan obtener mejores resultados y una buena toma de decisiones.

1.6.4 Justificación Económica

Al plantear el modelo Integrador de Datos Activos pretendemos crear un marco de construcción que proporcione lineamientos generales a fin de que los diferentes sistemas dentro una empresa o institución pública puedan utilizar como base este modelo que permita optimizar la oportuna toma de decisiones.

Los beneficios que se obtendrán con el desarrollo del Modelo serán mayores a los costos y gastos de la elaboración o de producción del mismo, ya que el impacto económico en la implementación del Sistema reducirá costos y tiempo en la manipulación en la información justificándose su desarrollo económico.

1.7 Alcances y Limites.

El modelo de Integración de Datos Activos, apoya a la oportuna toma de decisiones permitiendo obtener información entre los diferentes sistemas de base de datos que participan dentro de una empresa, solucionara problemas específicos de manejo de información de forma rápida y segura donde se adaptara fácilmente a los requerimientos del personal.

El alcance del presente trabajo será la implementación de un prototipo integrador de datos activos que contara con los siguientes procesos:

1.7.1 Alcances

- ❖ Extracción de los datos de las diferentes fuentes de información.
- ❖ Transformación de los datos extraídos a un lenguaje estándar.
- ❖ Cargado de los datos convertidos al almacenamiento multi base.
- ❖ Intercambio de información entre los diferentes sistemas.
- ❖ Consultas y reportes de forma fácil y sencilla.
- ❖ Disminución de gastos materiales y tiempos de trabajo en el compartimiento de Información.

1.7.2 Limites

Le modelo de integrador de datos activos permite el movimiento de información entre origen y destino de diferentes sistemas de base de datos de forma rápida y segura limitándose a conexiones externas de fuentes de datos masivas que ocasione inconsistencias de manejo de datos.

1.8 Aporte

El aporte principal del modelo integrador de datos activos es el movimiento rápido y seguro de información entre origen y destino de diferentes fuentes de datos.

Los resultados proporcionados permitirán a los programadores tener mayor conocimiento sobre herramientas que se puedan utilizar para la integración de los datos, para que mantenga la consistencia de las diferentes fuentes de datos cuando realicen transacciones.

El modelo integrador de datos activos una vez implementado será una herramienta de mucha ayuda para la toma de decisiones.

1.9 Metodología científica

El método de investigación que se adoptara para realizar el presente trabajo será el método científico ya que nos ayudara a poner a prueba la solución de la hipótesis planteada.

El método científico consiste en la realización de una serie de procesos específicos que utiliza la Ciencia para adquirir conocimientos. Estos procesos específicos son una serie de reglas o paso, bien definidos, que permiten que al final de su realización se obtengan unos resultados fiables y seguros. Permitiendo contribuir a la ciencia ya que este trabajo de investigación se ve como es posible desarrollar una herramienta que procese transacciones de bases de datos heterogéneas.

1.10 Herramienta (Lenguaje Unificado de Modelado) UML

El UML (Lenguaje Unificado de Modelado) es una herramienta que permite el desarrollo de sistemas. Permitiendo a los creadores de sistemas generar diseños que capturen sus ideas en forma convencional y fácil de comprender para comunicarlas a otras personas.





CAPITULO II

MARCO TEORICO

Sistema de Base de Datos

- Definición de Sistema de Base de Datos
- Integración de Datos
- Extracción y conversión de datos
 - Archivos Planos
- Sistema Multibase (DataWarehouse)
- Tipo de Conexiones a Gestores de Base de Datos
 - Conexión JDBC
 - Conexión ODBC
- METODOLOGÍA UML
- Herramientas de Investigación
 - SQL Server

CAPITULO II

2. MARCO TEORICO

En el presente capítulo se hace referencia a los conceptos generales del Integrador de Datos Activos identificándolo como componente principal para una estructura de Inteligencia Empresarial.

Posteriormente se describen las principales características del Integrador de Datos Activos.

Así mismo se hace una relación de las tecnologías inmersas que forman parte importante durante y después de la implementación del proyecto de desarrollo del Integrador de Datos Activos. Finalmente se destaca la importancia de la arquitectura del Integrador de Datos Activos.

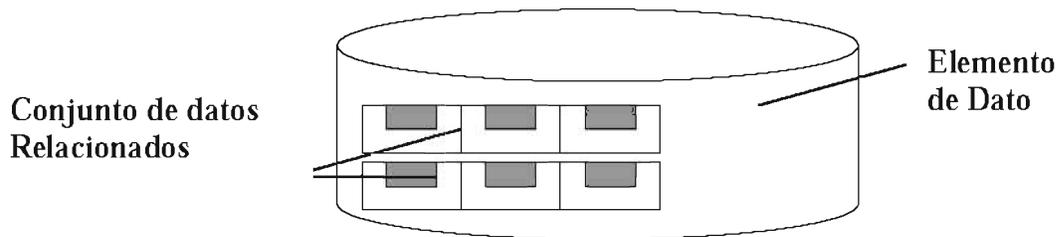
2.1. Sistema de Base de Datos

Para tener una definición completa de sistema de base de datos, inicialmente se repasara algunos conceptos importantes. En primer lugar un sistema de base de datos tiene elementos como una base de datos, un sistema gestor de datos consta también del hardware y el personal apropiados

2.1.1 Base de Datos

Se considera como base de datos como un conjunto de datos relacionados entre si, donde por datos se entiende como hecho conocido que pueden registrarse y tiene un significado implícito. La base de datos debe tener una definición que contenga los tipos de datos, las estructuras y restricciones de los datos que almacenaran en ella

Figura 2.1 Base de Datos



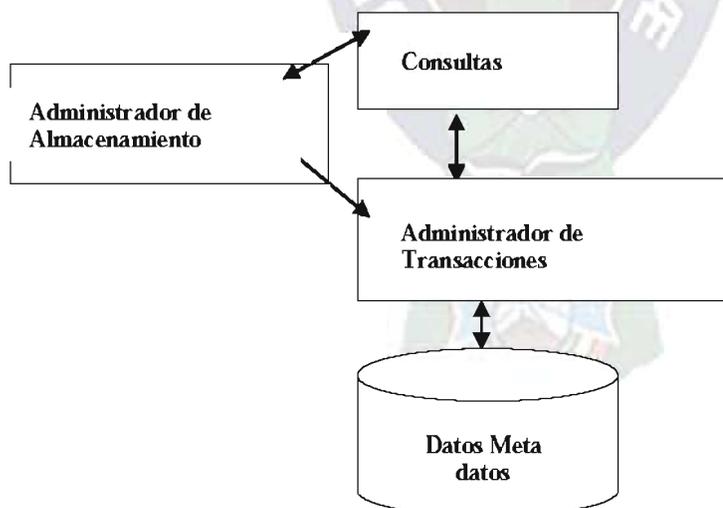
2.2.2 Sistema Gestor de Base de Datos (SGBD)

Un SGBD es considerado como el medio con el cual un programa de aplicación o un usuario final examina y manipula los datos almacenados en la base de datos.

El SGBD es un conjunto de programas que brinda un conjunto de servicios a los usuarios finales, los programadores y otros, por lo tanto, SGBD es un software de propósito general.

Componentes de un SGBD autor (Ullman J.)

Figura 2.2 principales componentes de un gestor de base de datos.



Según Ullman J. en su libro "Introducción a los Sistemas de Base de Datos", indica que un SGBD permita:

- A los usuarios crear otra base de datos y especificar su esquema por medio de un lenguaje especializado denominado Lenguaje de Definición de Datos.
- Ofrezca la capacidad de consultar los datos y modificarlo.
- Soporte el almacena de cantidades muy voluminosas de datos durante un largo periodo, protegiendo contra utilización no autorizada.

2.2.3 Modelo Relacional

El modelo relacional tiene tres partes principales que tiene que ver con la estructura de datos, la integridad de los datos y la manipulación de datos, respectivamente.

Figura 2.3 Modelo Relacional Según C. J. DATE.



2.3 Integración de Datos

La información contenida en las diferentes áreas funcionales es el recurso principal de una organización, en este sentido la información obtenida de estas son empleadas en distintas formas en beneficio de la organización.

Permitiendo una buena toma de decisiones.

2.4. Extracción y conversión de datos

La extracción de datos es la conexión a las diferentes fuentes de datos permitiendo manipular los datos y realizar el proceso de conversión a un lenguaje estándar en este caso a un archivo plano (Texto).

Velocidad de proceso y transformación evalúan registro por registro. Uso de recursos, es que una base de datos está preparada para la optimización de recursos ya sea de disco, memoria y proceso, lo cual hace que el rendimiento del proceso sea administrado por la configuración de la base de datos.

2.5 Sistema Multi Base (Datawarehouse)

Es un repositorio de datos de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, reportes y decisiones.

Estos valores normalmente son guardados por la organización de dos formas:

- Los sistemas operacionales de registros
- Y el Data Warehouse

Crudamente hablando, los sistemas operacionales de registros es donde los datos son depositados y el Data Warehouse es de donde se extraen esos datos.

2.5.1 Los objetivos fundamentales de un Sistema multibase son:

- Hace que la información de la organización sea accesible: los contenidos del Data Warehouse son entendibles y navegables, y el acceso a ellos son caracterizado por el rápido desempeño. Estos requerimientos no tienen

fronteras y tampoco límites fijos. Cuando hablamos de entendible significa, que los niveles de la información sean correctos y obvios. Y Navegables significa el reconocer el destino en la pantalla y llegar a donde queramos con solo un clic. Rápido desempeño significa, cero tiempo de espera. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.

- Hacer que la información de la organización sea consistente: la información de una parte de la organización puede hacerse coincidir con la información de la otra parte de la organización. Si dos medidas de la organización tienen el mismo nombre, entonces deben significar la misma cosa. Y a la inversa, si dos medidas no significan la misma cosa, entonces son etiquetados diferentes. Información consistente significa, información de alta calidad. Significa que toda la información es contabilizada y completada. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.
- Es información adaptable y elástica: el Sistema Multibase esta diseñado para cambios continuos. Cuando se le hacen nuevas preguntas al Sistemamultibase, los datos existentes y las tecnologías no cambian ni se corrompen. El diseño de Data Marts separados que hacen al multibase, deben ser distribuidos e incrementados. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.
- Es la fundación de la toma de decisiones: del sistema multibase tiene los datos correctos para soportar la toma de decisiones.

2.5.2 Los elementos básicos de un sistema Multibase

- Área de tráfico de datos: es un área de almacenamiento y grupo de procesos, que limpian transforman, combinan, remover los duplicados, guardan, archivan y preparan los datos fuente para ser usados en el sistema multibase.
- Servidor de presentación: la maquina física objetivo en donde los datos del sistema multibase son organizados y almacenados para consultas directos por los usuarios finales, reportes y otras aplicaciones.

- Modelo dimensional: una disciplina específica para el modelado de datos que es una alternativa para los modelos de entidad – relación.
- Data Mart: un subgrupo lógico del sistema multibase completo.
- Data sistema multibase: búsquedas fuentes de datos de la empresa. Y es la unión de todos los data marts que la constituyen.
- Almacenamiento operacional de datos: es el punto de integración por los sistemas operacionales. Es el acceso al soporte de decisiones por los ejecutivos.
- OLAP: actividad general de búsquedas para presentación de texto y números del sistema multibase, también un estilo dimensional específico de búsquedas y presentación de información y que es ejemplificada por vendedores de OLAP.
- ROLAP: un grupo de interfases de usuarios y aplicaciones que le dan a la base de datos relacional un estilo dimensional.
- MOLAP: un grupo de interfases de usuarios, aplicaciones y propietarios de tecnología de bases de datos que tienen un fuerte estilo dimensional.
- Aplicaciones para usuarios finales: una colección de herramientas que hacen las consultas, analizan y presentan la información objetivo para el soporte de las necesidades del negocio.

2.5.3 Los procesos básicos del sistema multibase

- Extracción: este es el primer paso de obtener la información hacia el ambiente del Sistema multibase
- Transformación: una vez que la información es extraída hacia el área de tráfico de datos, hay posibles pasos de transformación como; limpieza de la información, tirar la basura que no nos sirve, seleccionar únicamente los campos necesarios para el Sistema multibase, combinar fuentes de datos, haciéndolas coincidir por los valores de las llaves, creando nuevas llaves para cada registro de una dimensión.
- Carga: al final del proceso de transformación, los datos están en forma para ser cargados.

2.6 Tipo de Conexiones a Gestores de Base de Datos

2.6.1 JDBC

Es una interfase de acceso a bases de datos estándar SQL que proporciona un acceso uniforme a una gran variedad de bases de datos relacionales. JDBC también proporciona una base común para la construcción de herramientas y utilidades de alto nivel. Consiste en un conjunto de clases e interfaces escritas en el lenguaje de programación.

Las siguientes tareas que realiza un JDBC son:

- Establece una conexión con una base de datos.
- Envía sentencias SQL.
- Manipula datos.
- Procesa datos.

2.6.2 ODBC

Cadena de conexión que permite a los usuarios a conocer los elementos de una base de datos sus interfaces son más sencillas que las de un JDBC como también para las consultas SQL. Proporciona una instalación fácil con los sistemas operativos XP, Windows Server 2003.

Las siguientes tareas que realiza un ODBC son:

- Establece una conexión con una base de datos.
- Envía sentencias SQL.
- Manipula datos.
- Procesa datos.

2.7 METODOLOGÍA UML

El UML (Lenguaje Unificado de Modelado) es una herramienta que permite el desarrollo de sistemas. Permitiendo a los creadores de sistemas generar diseños que capturen sus ideas en forma convencional y fácil de comprender para comunicarlas a otras personas.

2.7.1 Modelo de Casos de Uso

Un caso de uso es una descripción de la secuencia de interacciones que se producen entre un actor y el sistema, cuando el actor usa el sistema para llevar a cabo una tarea específica. Expresa una unidad coherente de funcionalidad, y se representa en el Diagrama de Casos de Uso mediante una elipse con el nombre del caso de uso en su interior. El nombre del caso de uso debe reflejar la tarea específica que el actor desea llevar a cabo usando el sistema.

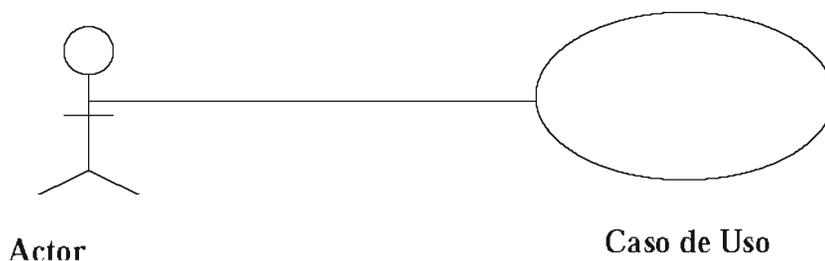
2.7.2 Diagrama de Casos de Uso

Un Diagrama de Casos de Uso muestra la relación entre los actores y los casos de uso del sistema. Representa la funcionalidad que ofrece el sistema en lo que se refiere a su interacción externa.

2.7.3 Elementos

Los elementos que pueden aparecer en un Diagrama de Casos de Uso son: actores, casos de uso y relaciones entre casos de uso.

Figura 2.9 Casos de USO



2.7.4 Actores

Un actor es una entidad externa al sistema que realiza algún tipo de interacción con el mismo.

Se representa mediante una figura humana dibujada con palotes. Esta representación sirve tanto para actores que son personas como para otro tipo de actores (otros sistemas, sensores, etc.).

2.7.5 Modelo de Análisis

El modelo de análisis es una vista interna del sistema, estructurado por clases y paquetes, para comprender como debería darse forma al sistema, es decir como debería ser diseñado e implementado.

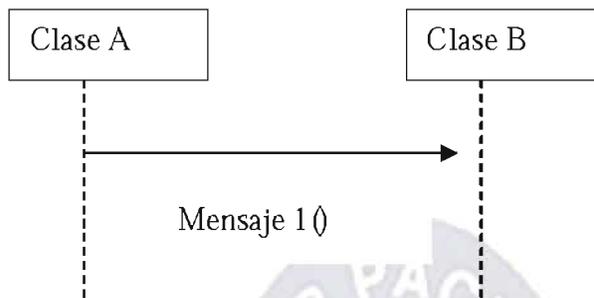
2.7.6 Diagrama de Secuencia

Un diagrama de Secuencia muestra una interacción ordenada según la secuencia temporal de eventos. En particular, muestra los objetos participantes en la interacción y los mensajes que intercambian ordenados según su secuencia en el tiempo.

El eje vertical representa el tiempo, y en el eje horizontal se colocan los objetos y actores participantes en la interacción, sin un orden prefijado. Cada objeto o actor tiene una línea vertical, y los mensajes se representan mediante flechas entre los distintos objetos. El tiempo fluye de arriba abajo.

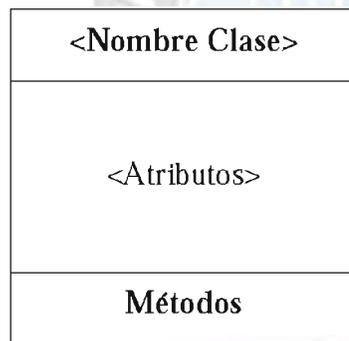
Se pueden colocar etiquetas (como restricciones de tiempo, descripciones de acciones, etc.) bien en el margen izquierdo o bien junto a las transiciones o activaciones a las que se refieren.

Figura 2.10 Diagrama de Secuencia



2.7.7 Diagrama de Clases

Figura 2.11 Diagrama de Clases



Un diagrama de clases esta constituido por los siguientes elementos:

- Clase: atributos, métodos y visibilidad.
- Relaciones: Herencia, composición, agregación, asociación y uso

Una clase se representa mediante una caja subdividida en tres partes: En la superior se muestra el nombre de la clase, en la media los atributos y en la inferior las operaciones. Una clase puede representarse de forma esquemática (plegada), con los detalles como atributos y operaciones suprimidos, siendo entonces tan solo un rectángulo con el nombre de la clase. En la Figura se ve cómo una misma clase puede representarse a distinto nivel de detalle según interese, y según la fase en la que se esté.

2.7.8 Diagrama de Colaboración

Un Diagrama de Colaboración muestra una interacción organizada basándose en los objetos que toman parte en la interacción y los enlaces entre los mismos (en cuanto a la interacción se refiere). A diferencia de los Diagramas de Secuencia, los Diagramas de Colaboración muestran las relaciones entre los roles de los objetos. La secuencia de los mensajes y los flujos de ejecución concurrentes deben determinarse explícitamente mediante números de secuencia.

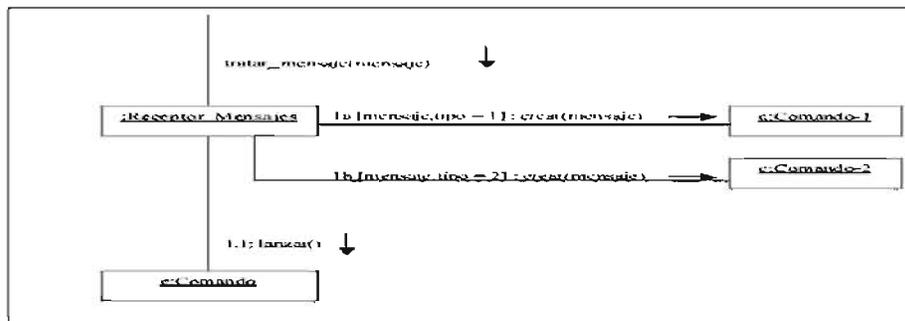
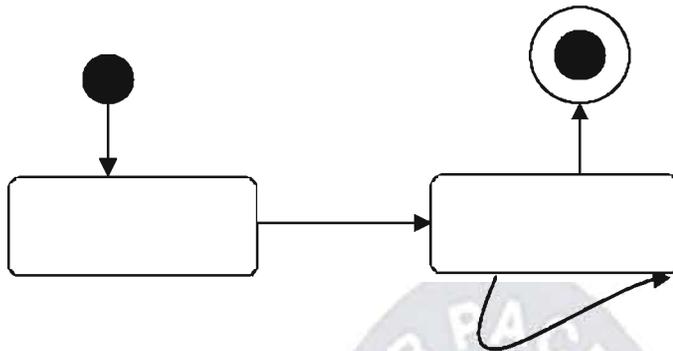


Figura 2.12 Diagrama de Colaboración

2.7.9 Diagrama de Estados

Un diagrama de estados puede representar ciclos continuos o bien una vida finita, en la que hay un estado inicial de creación y un estado final de destrucción (del caso de uso o del objeto). El estado inicial se muestra como un círculo sólido y el estado final como un círculo sólido rodeado de otro círculo. En realidad, los estados inicial y final son pseudos estados, pues un objeto no puede “estar” en esos estados, pero nos sirven para saber cuáles son las transiciones iniciales y finales(es).

Figura 2.13 Diagrama de Estados



Incluiremos todos los modelos de diagramas que contiene el UML, que forman parte del sistema de software.

2.8 HERRAMIENTAS DE INVESTIGACIÓN

2.8.1 SQL Server

SQL Server es un sistema administrador para Bases de Datos relacionales basadas en la arquitectura Cliente / Servidor (RDBMS) que usa Transact-SQL para mandar peticiones entre un cliente y el SQL Server.

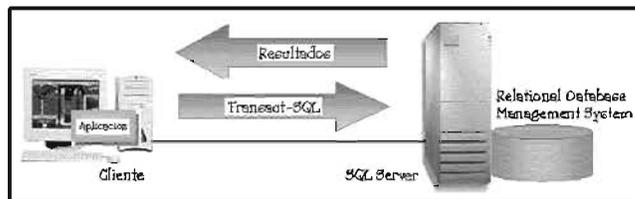


Figura 2.14

ARQUITECTURA CLIENTE / SERVIDOR:

SQL Server usa la arquitectura Cliente / Servidor para separar la carga de trabajo en tareas que corran en computadoras tipo Servidor y tareas que corran en computadoras tipo Cliente:

El Cliente es responsable de la parte lógica y de presentar la información al usuario. Generalmente, el cliente corre en una o más computadoras Cliente, aunque también puede correr en una computadora Servidor con SQL Server.

SQL Server administra Bases de Datos y distribuye los recursos disponibles del servidor (tales como memoria, operaciones de disco, etc.) entre las múltiples peticiones.

La arquitectura Cliente /Servidor permite desarrollar aplicaciones para realizar en una variedad de ambientes.





CAPITULO III
MARCO APLICATIVO

MARCO APLICATIVO

- DESCRIPCION INFORMAL
 - Limpieza
 - Integración
 - Transformación
 - Carga de Datos
- DESCRIPCION FORMAL
 - Gestor de Base de Datos
- Extracción, Transformación y Cargado
 - Generación de los Archivos (Transformación del archivo)
 - Cargado de Archivos
- Tipo de Almacenamiento
- Vista Usuario
- DEFINICION DE LOS CASOS DE ESTUDIO
 - APLICACION DEL PROCESO DE INTEGRACION
 - PREINTEGRACION
 - COMPARACION
 - CONFORMACION
 - RESTRUCTURACION Y ASOCIACION
 - ESQUEMA INTEGRADO
 - APLICACIÓN DEL ESQUEMA INTEGRADO
 - DEMOSTRACION DE HIPOTESIS

CAPITULO III

3 MARCO APLICATIVO

3.1 DESCRIPCION INFORMAL

El Modelo de Integración de Datos activos en Inteligencia Empresarial para la Toma de Decisiones proporciona un almacenamiento de datos, funcionalidad y receptividad a las consultas que van mas allá de las posibilidades de las base de datos destinadas a transacciones. De forma que sea eficiente para el manejo de información logrando procesos de extracción de datos que ocasione un movimiento rápido y seguro, evitando inconsistencias y pérdida de información entre origen y destino de los diferentes sistemas de base de datos de manera que pueda ayudar a la toma de decisiones. Dando una solución de recoger datos, transformarlos en información útil, cargarlos en una estructura multibase para luego utilizar herramientas de consulta o de reportes para el análisis de información.

Para evitar las inconsistencias de datos, dentro del modelo integrador, se debe cumplir con las principales tareas de procesamiento de manejo de datos referente a la extracción y transformación, ya que cada una de estas tareas es fundamental para una buena respuesta a las consultas o reportes deseados por el usuario.

Extracción:

Es la recolección de campos con sus respectivos atributos de las diferentes tablas de cada base de datos, los cuales están dentro de los gestores de base de datos, permitiendo formar una nueva relación que sirva para la toma de decisiones.

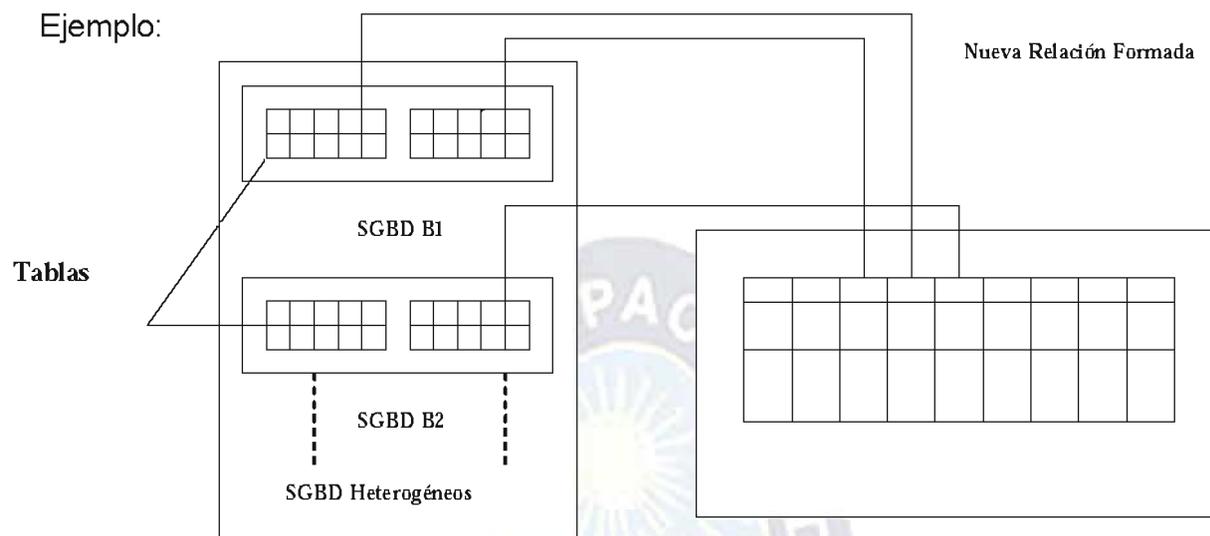


Fig. 3.1.1
Figura de extracción y conversión de nueva relación

Limpieza:

Detectar y eliminar errores rellenar atributos vacíos, resolver inconsistencias.

La limpieza de los datos es importante por la existencia de datos:

Incompletos: atributos sin valor, falta de atributos o el valor del atributo

Eje. Salarios=" -10".

Inconsistentes: contiene discrepancias

Eje. Edad="42" fecha de nacimiento="03/07/1998"

¿Por que nos encontramos con ese tipo de problemas?

Incompletos por que: no es necesario el dato que se registra, consideraciones diferentes cuando el dato es registrado y cuando es analizado ó problemas humanos, hardware, software.

Incorrectos debido a error humano o del programa al introducir el dato.

Inconsistentes debido a que provienen de diferentes fuentes de datos, modelo de datos no normalizados.

Todos estos problemas hacen un análisis no fiables e inexactos mostrándonos registros duplicados o valores no asignados que llevan a obtener una estadística incorrecta.

Lo que busca nuestro sistema es tener datos de características de calidad, cumpliendo con los siguientes puntos.

Un dato debe ser:

- ✓ Preciso
- ✓ Completo
- ✓ Consistente
- ✓ Creíble
- ✓ Con valor añadido
- ✓ Interpretable
- ✓ Accesible
- ✓ Riguroso en el tiempo

Para que cumpla con esta norma es necesaria la corrección de datos inconsistentes eje. (Fecha_de_nacimiento > hoy). Rellenar valores perdidos a causa de un mal funcionamiento del equipo, los datos nunca fueron rellenos o no existía en aquel momento o se eliminaron con otra información registrada. Además podemos darle la solución como:

Etiquetarlas con una constante global como: "Desconocido"

Integración:

Identificar la mejor fuente de datos para cada registro

En la integración de los datos, presentara una redundancia a causa de la integración de las base de datos. Eje.

Esquema BD1= {001, Alejandra, Medrano, Díaz, Contadora}

Esquema BD2= {001, Alejandra, Medrano, Díaz, 3000Bs}

Permitiendo manejar consultas se logra una nueva relación

Integración = {0011, 001, Alejandra, Medrano, Díaz, Contadora, 3000Bs}

Identificación de objetos: el mismo atributo o nombre puede tener diferentes nombres en cada base de datos, el atributo puede aparecer como derivado en otra tabla.

Para no encontrarnos con este tipo de problema es necesario identificar la fuente de datos más fiable para cada dato, ya que una integración cuidadosa ayuda a reducir las redundancias y las inconsistencias.

Transformación:

Estandarizar códigos y formatos de representación, definir dominios, usar conversiones y combinaciones para generar nuevos campos, corregir datos.

En esta acción es necesario estandarizar códigos y formatos de representación como:

Convertir números cardinales a ordinales

Separar fecha y hora

Poner textos descriptivos

Unificar código (hombre-mujer, 0-1)

Unificar estándares: unidades de medida, tiempo, moneda.

Carga de Datos:

Definir procesos de carga y sincronización

Puede requerir bastante tiempo, aconsejable realizar en horas de baja carga de los sistemas, su sincronización debe ser eficientemente computacional.

Para la revisión del cambio es necesario mantener una imagen del antes y la otra del después de la extracción de la información, comparar ambas y detectar los cambios.

La extracción, transformación y el cargado de los datos permite realizar un nuevo archivo que conlleva a realizar una nueva estructura según las vistas de los usuarios para luego ser cargada al Sistema Multibase, cada vista de usuario esta formada mediante un lenguaje de manipulación de datos, dentro de lo que es las instrucciones SQL.

Luego que tenemos la información cargada a la nueva estructura multibase, la información es manipulada mediante lenguaje MDX (expresiones multidimensionales) con sentencias SQL que permite expresar problemas olab de forma fácil, para las consultas o reportes requeridos.

Ejemplo

```
SELECT
  {[Medidas].Ventas} on columns,
  {[Fecha].2002}, [Fecha].[2003]} on filas
From Ventas
Where {[tiendas].[USA].[CA]}
```

La representación grafica de nuestro modelo integrador de datos activos la representamos de la siguiente manera. (Ver figura 3.1.2)

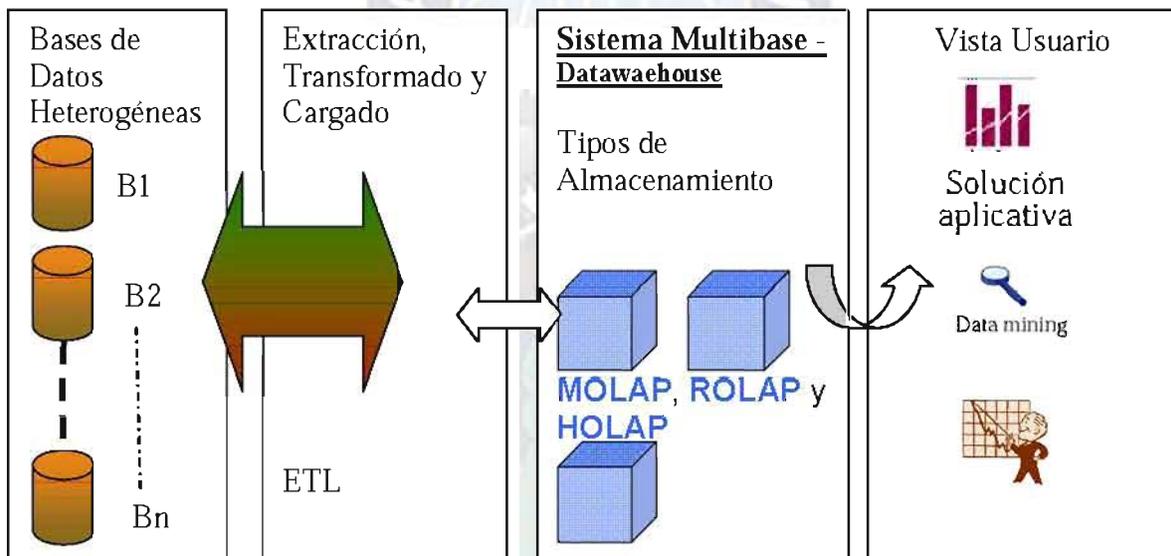


Fig. 3.1.2
Modelo Integrador de Datos Activos en Inteligencia
Empresarial para la Toma de Decisiones

3.2 DESCRIPCION FORMAL

La representación del sistema integrador de datos activos en la descripción formal se representara de la siguiente manera.

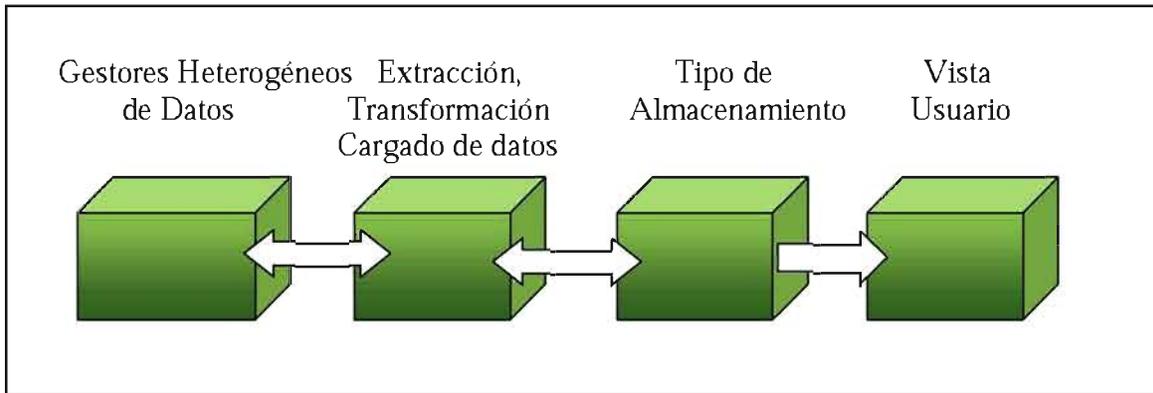


Fig. 3.2
Modelo Integrador de Datos Activos en Inteligencia Empresarial para la Toma de Decisiones

3.2.1 Gestor de Base de Datos

La representación grafica para el SGBD

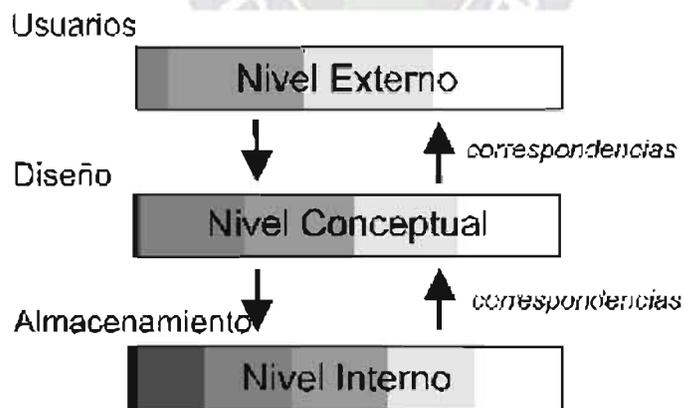


Fig. 3.2.1
Arquitectura de la Base de Datos

El nivel clave en este modelo integrador es el conceptual y el interno. Éste contiene la descripción de las entidades, relaciones y propiedades de interés para

la toma de decisiones, y constituye una plataforma estable desde la que podemos proyectar los distintos esquemas, sobre el esquema interno, que describe los datos según el sistema físico. Las posibles proyecciones de datos quedan resumidas en la grafica. (Ver Fig. 3.2.1)

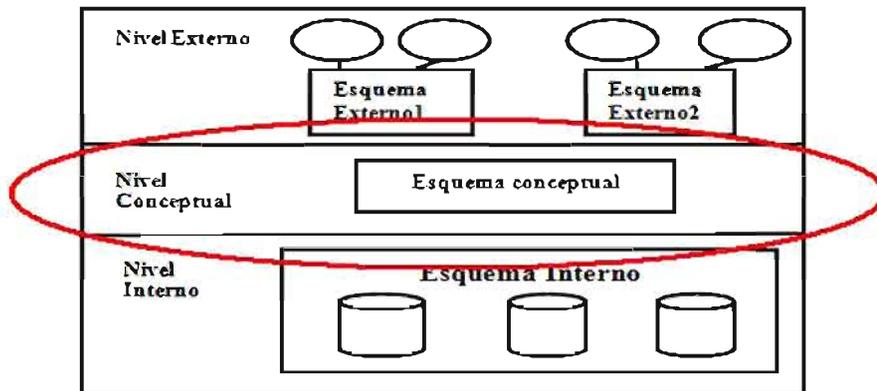


Fig. 3.2.1 Esquema de Gestor de Base de Datos

La cual nos permite hacer una representación de cada una de las bases de datos utilizando la variable B = Base de Datos que se encuentran dentro del nivel interno del cada uno de los gestores de base datos (Ver fig. 3.2.1)

En nuestro modelo integrador esta representado como:

B1: Base de Datos1

B2: Base de Datos2

⋮

Bi Base de Datos i

Donde Bi es la variable representativa de una base de datos cualquiera que es además parte de su correspondiente Sistema Gestor de Base de Datos.

Con lo que se quiere llegar es a la integración de las bases de datos heterogéneos (Ver Fig. 3.2.2)

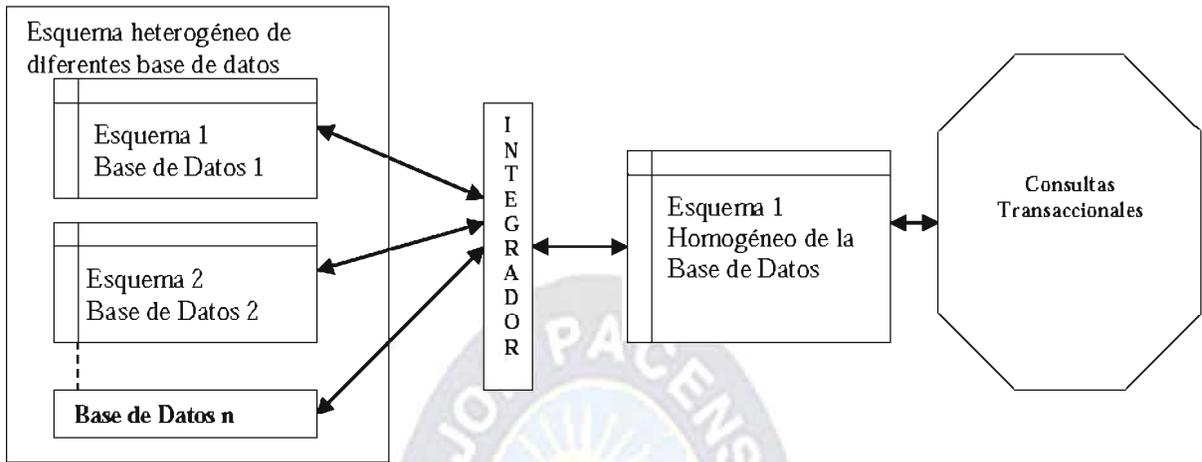
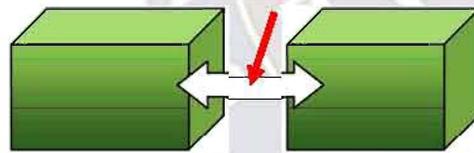


Fig. 3.2.2
Esquema Heterogéneo de Base de Datos

3.3 Extracción, Transformación y Cargado

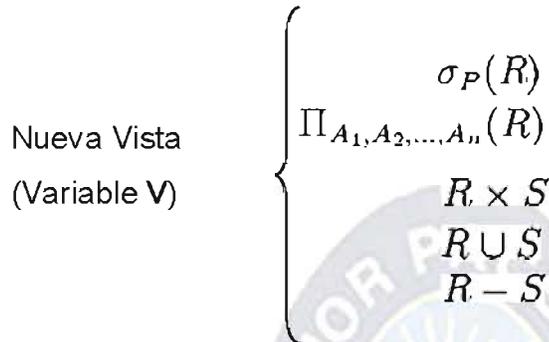
Fig. 3.3
Extracción Transformación y Cargado



Sistemas Heterogéneos de Datos

La extracción de la información la podemos representar mediante formulas de base de datos que las realizamos con la aplicación del algebra relacional, la cual nos permite realizar vistas para la nueva relacion que utilizara en la toma de decisiones.

Fig. 3.3 .1
Vista formada según el uso de operadores del algebra relacional



Esta nueva vista esta formada según el uso los operadores del algebra relacional dentro de las diferentes bases de datos del nivel interno, según los pasos que se deben respetar para su debida combinación, correspondiente a los diferentes sistemas gestores de base de datos.

Una representación de combinación podría ser:

$$\text{Resultado} = \Pi_{A_1, A_2} \sigma_P(R)$$

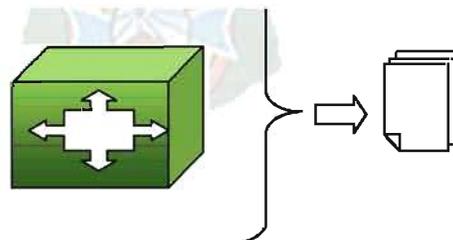
Donde:

Su representación de la nueva vista según la base de datos es la siguiente:

(VB1, VB2,.....VB3.....VBi)

3.3.1 Generación de los Archivos (o Transformación del archivo)

Fig. 3.3 .2 Transformación



Para la transformación del nuevo archivo de la VB1, VB2,.....VB3,....VBi. nos permite emplear la función (G_A) "Generación nuevo Archivo" que representa la transformación, la cual es realizada mediante diferentes funciones que utiliza el prototipo o sistema, como la limpieza, validación de campos, control de variable, carácter de separación, controlador de combinación de caracteres.

$$(G_A) = \left\{ \begin{array}{l} \text{Limpieza} \\ \text{Validación de campos} \\ \text{Control de Variables} \\ \text{Carácter de separación} \\ \text{Controlador de combinación} \\ \text{de caracteres} \end{array} \right.$$

Donde nos da como resultado una vista final la cual también puede ser presentada como la función (Archivo_V_F), a esta función la representamos de la siguiente manera:

$$(\text{Archivo_V_F}) = G_A = \{VB1, VB2, \dots, VB3, \dots, VBn\}$$

La cual se carga al sistema multibase para la toma de decisiones.

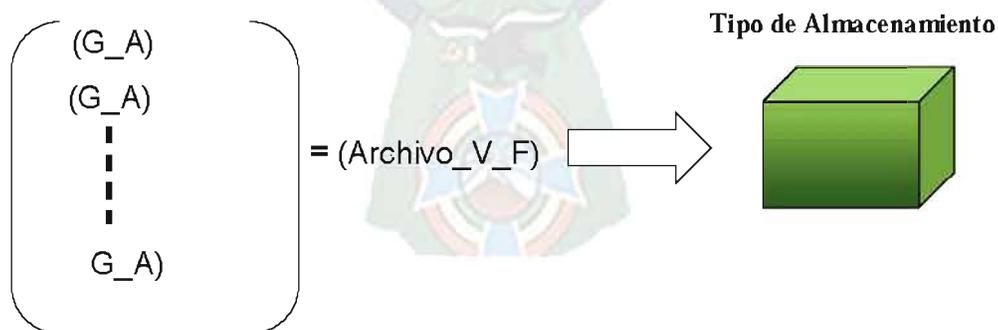
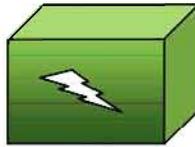


Fig. 3.3.3
Cargado de Archivo de Vista Final

3.3.2 Cargado de Archivos

Fig. 3.3.4
Cargado de Archivo



El cargado del archivo final (Archivo_V_F) mejora el performance en las consultas dentro de la estructura multibase, cuando se tiene un aumento de cantidad de espacio de almacenamiento. Permitiendo tener un conocimiento de un porcentaje aceptable de cantidad de datos almacenados, para luego pasar a ser procesados según los requerimientos del administrador multibase. (Ver Figura 3.3.3)

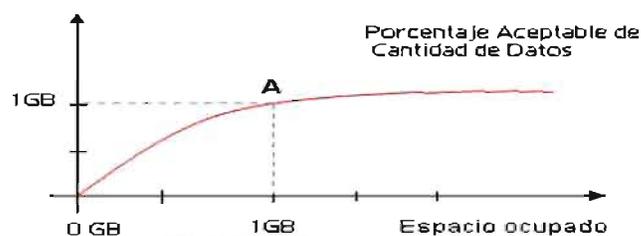


Fig. 3.3.3

Se puede sugerir el siguiente proceso dentro de la estructura multibase para no perjudicar el performance:

Hacer un vaciado de los archivos ya utilizados de una fecha determinada a otra llevando a otro almacén de datos o (CD, discos duro, cintas, etc.), esto para mejorar el performance para luego poder reutilizar la información dentro de lo que sería del Data mining u o otra aplicación para su uso.

3.4 Tipo de Almacenamiento

Se toma en cuenta los tipos de almacenamiento que son el MOLAP, ROLAP, OLAP (Ver Marco Teórico Sistema Multibase Datawarehouse), que permite el

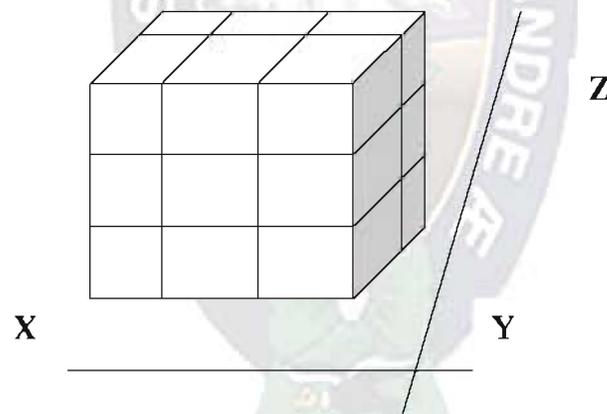
manejo de estructuras multidimensionales y es representada de la siguiente manera:

En teoría de bases de datos, un cubo OLAP es una representación abstracta de la proyección de una relación de un Sistema administrador de bases de datos relacionales. Dada una relación de orden N , se considera la posibilidad de una proyección que dispone de los campos X, Y, Z como clave de la relación y de W como atributo residual. Categorizando esto como una función se tiene que:

$$W : (X,Y,Z) \rightarrow W$$

Los atributos X, Y, Z se corresponden con los ejes del cubo, mientras que el valor de W devuelto por cada tripleta (X, Y, Z) se corresponde con el dato o elemento que se rellena en cada celda del cubo.

Cubo Dimensional



3.5 Vista Usuario

Las vistas de los usuarios pueden ser realizadas mediante exportaciones a hojas Excel, Graficas ó consultas directas a los archivos finales, ó en el caso de realizar una minería de datos.

3.6 DEFINICION DE LOS CASOS DE ESTUDIO

En la investigación se tomara el caso particular que consta de dos SGBD, sobre las que se pretende realizar la integración.

La compañía de ingeniería tiene la base de datos anteriormente descrita, en dos bases de datos que fueron desarrolladas sobre distintos SGBD pero comparten la estructura básica señalada. Para el caso de estudio se tiene la siguiente ejemplificación.

Esquema 1 (SGBD 1)

ING (CODING, NOMBREING, TITULO)

PROY (CODPROY, NOMBREPROY, PRESUP, LUGAR)

TRABE (CODING, CODPROY, RESP, DURA)

SUE (TITULO, SALARIO)

Esquema 2 (SGBD 2)

INGENIERO (IDING, NOMING, TITULO, SUELDO)

PROYECTO (IDPROY, NOMPROY, PRES, LOC, NOMCLI)

TRABAJAEN (IDING, IDPROY, RESP, DURAC)

Fig. Esquema 3.6.1

ING

CODING	NOMBREING	TITULO
E09	Juan Rodríguez	Ing Eléctrico
E02	Miguel Sánchez	Analista de Sist
E15	Armando Larrea	Ing Telecom
E05	Beatriz Molina	Programador
E06	Jorge Castañón	Analista de Sist
2	Luis Chávez	Ing Eléctrico
E13	Julia Jiménez	Ing Telecom

TRABE

CODPROY	NOMBREPROY	PRESUP	LUGAR
P1	Instrumentación	150000	La Paz
P2	Desarrollo de base de datos	135000	Oruro
P3	Instalación de Red	250000	Cbba
P4	Mantenimiento	310000	La Paz
P5	Instalación de Red	500000	Oruro

SUE

TITULO	SALARIO
Ing Eléctrico	5000
Analista de Sist	3000
Ing Telecom	3700
Programador	2500

PROY

CODIGN	CODPROY	RESP	DURA
E09	P1	Administrador	12
E02	P1	Analista	24
E02	P2	Analista	6
E15	P3	Consultor	10
E15	P4	Ingeniero	48
E05	P2	Programador	18
E06	P2	Administrador	24
E12	P4	Administrador	48
E13	P3	Ingeniero	36
E13	P5	Ingeniero	23

Fig. 3.6.2 Esquema2

INGENIERO

IDING	NOMING	TITULO	SUELDO
E01	Roberto Dávila	Programador	2400
E03	Carlos Tapia	Analista de Sist	3400
E04	Alberto Cortez	Ing Eléctrico	4000
E07	Javier Molina	Programador	2400
E09	Juan Rodríguez	Ing Electrico	4000
E10	Marcelo Silva	Ing Telecom	2700
E11	Daniel Pérez	Ing Eletnco	4000
E15	Armando Larrea	Ing Telecom	2700

TRABAJAEN

IDING	IDPROY	RESP	DURAC
E01	P1	Programador	12
E03	P1	Analista	24
E04	P1	Consultor	6
E04	P3	Diseñador	10
E07	P3	Programador	48
E07	P5	Programador	18
E09	P5	Consultor	24
E10	P3	Administrador	48
E11	P5	Administrador	36
E11	P3	Ingeniero	23
E15	P5	Ingeniero	

PROYECTO

IDPROY	NOMPROY	PRES	LOC	NOMCLI
P1	Instrumentación	150000	La Paz	Carlos Romero
P2	Desarrollo de Base de Datos	135000	Oruro	Jorge Soria
P3	Instalación de Red	250000	Cbba	José Méndez
P4	Mantenimiento	310000	La Paz	Antonio Soliz
P5	Instalación de Red	500000	Oruro	Aldo Castro

Puede apreciarse que en ambas instancias de las bases de datos los proyectos son los mismos, así también existen algunos ingenieros que están presentes en las dos bases de datos. La asignación de proyectos es diferente en cada base de datos aunque incluye en algún caso a los mismos trabajadores.

3.6.1 PLICACION DEL PROCESO DE INTEGRACION

Identificaremos a los esquemas que participaran en la integración en este sentido se tiene:

$$S1 = \{R1^1, R2^1, R3^1, R4^1\} = \{ING, PROY, TRABE, SUE\}$$

$$S2 = \{R1^2, R2^2, R3^2\} = \{INGENIERO, TRABAJAEN, PROYECTO\}$$

Para cada relación de los esquemas se tiene los siguientes atributos

$$R1^1 = \{CODING, NOMBREING, TITULO\}$$

$$R2^1 = \{CODPROY, NOMBREPROY, TITULO\}$$

$$R3^1 = \{CODING, CODPROY, RESP, DURA\}$$

$$R4^1 = \{TITULO, SALARIO\}$$

$$R1^2 = \{IDING, NOMING, TITULO, SUELDO\}$$

$$R2^2 = \{IDPROY, NOMPROY, PRES, LOC, NOMCLI\}$$

R3²= {IDING, IDPROY, RESP, DURAC}

Ahora si podemos realizar la integración de los esquemas.

3.6.2PREINTEGRACION

Ahora utilizamos la metodología de integración. En este caso se representamos dos esquemas a integrar (S1, S2) se ve por conveniente adoptar la metodología de escalera binaria, las razones son las siguientes.

- Con esta metodología, si existiera la incorporación de otra base de datos, la integración se la hará con el esquema conceptual global obtenido.
- El trabajo realizado contemplara el tratamiento de dos esquemas de base de datos, los correspondientes al caso de estudio.

Como segundo caso se realizara una inspección general sobre los esquemas a integrar, llegando a tener una primera percepción de los posibles problemas que presentaran.

Se identifica los aspectos más llamativos, como las llaves de cada relación:

S1

Relación ING: llave CODING, TITULO

Relación PROY: llave CODPROY

Relación TRABE: llave CODING, CODPROY

Relación SUE: llave TITULO

S2

Relación INGENIERO: llave IDING

Relación PROYECTO: llave IDPROY

Relación TRABAJAEN: llave IDING, IDPROY

Hasta este punto ya se puede apreciar que existe una correspondencia entre relaciones, por ejemplo entre la relación PROY y PROYECTO, se advierte que estas tienen en común los proyectos de la compañía; también ya puede observarse algunos conflictos, estos se analizan en la siguiente fase:

3.6.3 COMPARACION

Conflicto 1. El conflicto de sinónimos se presenta a diferentes niveles por ejemplo existe sinónimos entre relaciones (Por Eje. ING e INGENIERO) los dos nombres

de relaciones representan el mismo concepto, el de ingenieros. Dentro de las anteriores relaciones también se observa que entre los atributos (Por Eje. NOMBREING e NOMING) se presenta el conflicto de sinónimos.

Conflicto 1.1. Este conflicto también se puede presentar. Ocurre cuando dos nombres representan distintas cosas (Por Eje. Un atributo DIRECCION que en una base de datos puede representar una dirección indicando la calle, avenida, zona, mientras que en otra base de datos puede indicar una dirección a tomar como norte, sur, este, oeste)

Conflicto 2. Las dependencias se pueden presentar en cada esquema individualmente, en el caso de estudio el conflicto ocurre en el esquema 2 obsérvese que existe una dependencia entre los atributos TITULO y SUELDO. En el caso de que existiera una modificación en el título de un ingeniero el atributo SUELDO deberá actualizarse también.

Conflicto 2.1 un conflicto detectado en la relación S en el esquema 1 el atributo SUELDO en el esquema 2 pues ambos representan el sueldo de un ingeniero, pero son diferentes elementos de modelado.

Realizamos la comparación de los esquemas

Fig.3.6.3 Comparación de Esquemas

S1	CONFLICTOS	S2
ING	SINONIMO	INGENIERO
CODING	SINONIMO	IDING
NOMBREING	SINONIMO	NOMING
TITULO	CONDUCTA	TITULO
	TIPO, ATRIBUTO FALTANTE, DEPENDENCIA	
PROY	SINONIMO	PROYECTO
CODPROY	SINONIMO	IDPROY
NOMBREPROY	SINONIMO	NOMPROY
PRESUP	SINONIMO	PRES
LUGAR	SINONIMO	LOC
	ATRIBUTO FALTANTE	NOMCLI
TRABE	SINONIMO	TRABAJAEN
CODING	SINONIMO, CONDUCTA	IDING
CODPROY	SINONIMO, CONDUCTA	IDPROY
RESP	NINGUNO	RESP
DURA	SINONIMO	DURAC
SUE	TIPO	
TITULO	ATRIBUTO FALTANTE	
SALARIO	ATRIBUTO FALTANTE	

Como se observa se hace una comparación entre los atributos comunes en ambas relaciones de los esquemas donde se identifican los conflictos presentados puede haber mas de un conflicto para una relación o atributo.

3.6.4 CONFORMACION

En todos los casos de sinónimos se escogerá el nombre mas representativo para el esquema integrado por ejemplo entre los nombres de las relaciones ING (R1¹) e INGENIERO (R1²) se escogerá el nombre de **INGENIERO** pues el mas representativo.



En el caso de los atributos faltantes en ING (R1¹), el atributo puede ser encontrado siguiendo la ruta de la relación ING (R1¹) hacia INGENIERO (R1²) de donde se captura el atributo SALARIO en el caso práctico es mejor contar con otra relación en el esquema integrado, de esta forma se incluye en la nueva relación lo siguiente:



En el caso de los atributos los nombres incompletos como PRESUP y PRES, en los esquemas S1, S2 respectivamente se escogerá un nombre representativo pudiendo ser PRESUPUESTO

Figura 3.6.4 Esquema Integrado

S1	ESQUEMA INTEGRADO	S2
ING	INGENIERO	INGENIERO
CODING	CODING	IDING
NOMBREING	NOMBREING	NOMING
TITULO	TITULO	TOTULO
		SUELDO
PROY	PROYECTO	PROYECTO
CODPROY	CODPROY	IDPROY
NOMBREPROY	NOMBREPROY	NOMPROY
PRESUP	PRESUPUESTO	PRES
LUGAR	LUGAR	LOC
	NOMBRECLI	NOMCLI
TRABE	TRABAJAEN	TRABAJAEN
CODING	CODING	IDING
CODPROY	CODPROY	IDPROY

RESP	CARGO	RESP
DURA	DURACION	DURAC
SUE	PAGA	
TITULO	TITULO	INGENIERO. TITULO
SALARIO	SUELDO	INGENIERO. SUELDO



Otro atributo faltante se representa entre PROY y PORYECTO la primera relación no presenta el el atributo NOMCLI, en consecuencia este atributo sera tomado en cuenta y se especificara que pertenece solo a la relación PROYECTO en el segundo esquema.

3.6.5 RESTRUCTURACION Y ASOCIACION

La reestructuración fue realizada con éxito **Figura 3.12** en un esquema integrado en el que se retuvo la información contenida en los dos esquemas. Esta reestructuración nos permite que se pueda conseguir cualidades deseables del esquema integrado.

ESQUEMA INTEGRADO

INGENIERO (CODING, NOMBREING, TITULO)

PROYECTO (CODPROY, NOMBREPROY, PRESUPUESTO, LUGAR, NOMBRECLI)

TRABAJAEN (CODING, CODPROY, CARGO, DURACION)

PAGA (TITULO, SUELDO)

Este esquema permitirá realizar transacciones en ambas bases de datos componentes, ya que será un esquema homogéneo.

3.6.6 APLICACIÓN DEL ESQUEMA INTEGRADO

Una transacción global como una local deberá mantener a las bases de datos componentes en estados consistentes. La forma en que una transacción global se lleva a cabo comienza en el usuario de la multibase de datos que define los cambios que realizara. La gran ayuda será que contara con el esquema conceptual global que le guiara el momento de realizar los cambios.

Para su verificación se sigue el siguiente ejemplo.

Actualizar el titulo de Armando Larrea de Ingeniero de Telecomunicaciones por Analista de Sistemas

Begin

Read (INGENIERO. TITULO)

Read (PAGA. SUELDO)

Where INGENIERO. NOMBREING="Armando Larrea"

INGENIERO. TITULO ← "Analista de Sistemas"

PAGA. SUELDO ← "3400"

Write (INGENIERO. TITULO)

Write (INGENIERO. SUELDO)

END

3.7 DEMOSTRACION DE HIPOTESIS

La comprobación de la hipótesis planteada se obtuvo a partir de la aplicación.

Esta transacción se realizo mediante diferentes bases de datos, con sus respectivos SGBD.

La técnica utilizada para la experimentación fue de la siguiente manera: para realizar modificaciones sobre las bases de datos se debe tener una consistencia entre las bases de datos en todo momento.



CAPITULO IV
CONCLUSIONES
RECOMENDACIONES

CONCLUSIONES Y RECOMENDACIONES

- CONCLUSIONES
- RECOMENDACIONES

CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

Al terminar el presente trabajo se puede hacer algunas conjeturas sobre el mismo:

El algebra Relacional es el lenguaje de consulta de Base de Datos que nos permite especificar las operaciones a realizar según su orden la cual la asemeja a un lenguaje de programación de tipo ingenioso.

Se logró conseguir un modelo de integración formal para consultas hacia base de datos multidimensional. Con lo cual cumplimos con uno de los objetivos trazados en el modelo integrador de datos activos en inteligencia empresarial para la toma de decisiones.

Respecto a la prueba de hipótesis se ha mostrado y demostrado que es posible desarrollar una herramienta que permita integrar diferentes Gestores de Base de Datos en una sola para luego ser aprovechada para el usuario en cualquier tipo de negocio.

Se puede concluir finalmente que el presente trabajo ha logrado conseguir parcialmente el objetivo general trazado para la investigación, esto debido a que se pudo manipular Gestores de Base de Datos Heterogéneas, sin afectar en su estructura de cada una de ellas y logrando una integración de todos los datos almacenados y distribuidos en sus Bases de Datos respectivos.

4.2 RECOMENDACIONES

Las recomendaciones con respecto al presente trabajo son las siguientes:

El Modelo Integrador de Datos Activos en Inteligencia Empresarial para la Toma de Decisiones debe ser aplicado sobre algoritmos que hagan uso del algebra relacional para representar las consultas hacia la Base multidimensional para luego ser utilizados en la toma de decisiones.

Con el modelo integrador se recomienda realizar funciones de verificación de perdida de datos, ya que con este modelo no debe existir perdida de ningún dato de cualquier base de datos que este interviniendo en la integración.

Se debe llegar a tener un control absoluto del performace en el cargado, en las consultas y los reportes que se realicen dentro del sistema.

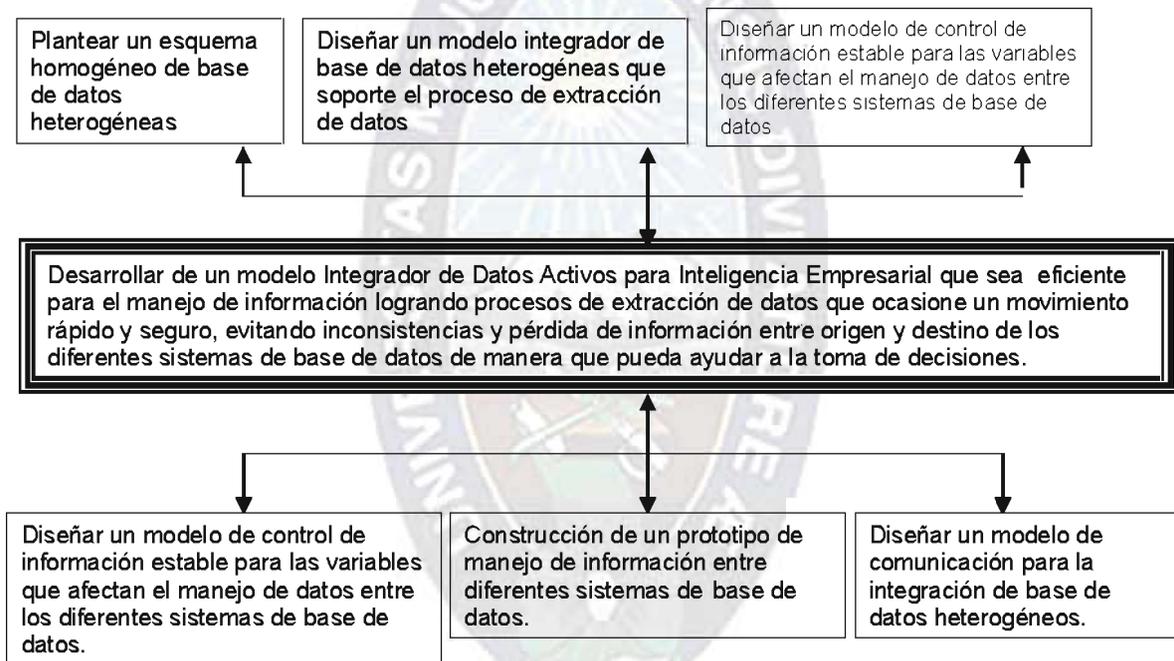
El aprovechamiento del performace con respecto al hardware utilizado en el modelo

Se debe seguir haciendo pruebas dentro del sistema para aumentar las principales tareas (Ver 3.1) que permitan que el manejo de datos sea mas claro y de mayor fiabilidad como también la aplicación de métodos de seguridad.

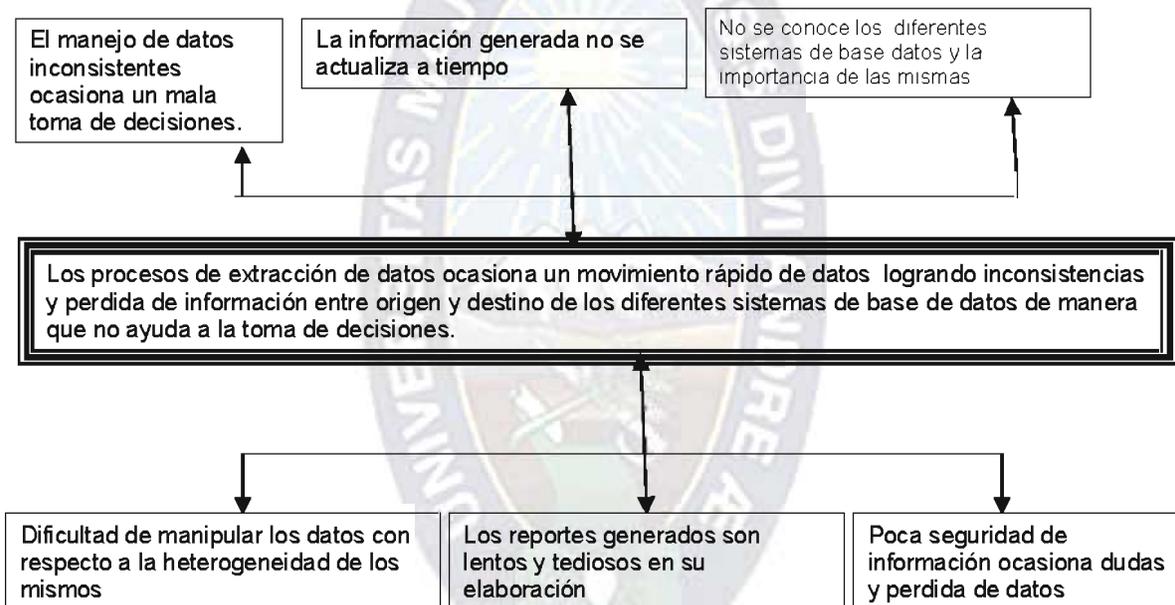


Anexos

ARBOL DE OBJETIVOS



ARBOL DE PROBLEMAS



**Cronograma
Modelo Integrador de Datos Activos**

GESTIÓN 2008	Inlcio y Final	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE	
ACTIVIDAD SEGÚN OBJETIVOS ESPECIFICOS	Fecha	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	
Selección de Tema de Tesis	Ago 18 Ago 29	■					
Introducción, Antecedentes, Planteamiento del problema, Problemas Específicos, Objetivo, Objetivo General, Objetivos Específicos, Hipótesis	Sep 1 Oct 17	■					
Justificación, Justificación Teórica, Justificación Tecnológica, Justificación Social, Justificación Económica, Alcances y Límites, Aporte, Metodología científica, Modelo de Sistema, Metodología	Oct 22 Nov 26			■			
PREDEFENSA	Dic 8 Dic 15					■	

Cronograma de GANT
Modelo Integrador de Datos Activos (Inteligencia Empresarial)

GESTIÓN 2009			FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO										
ACTIVIDAD SEGÚN OBJETIVOS ESPECÍFICOS	INICIO	FIN	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
			1ro Semanas															
CAPITULO III Diseño Arquitectonico Diseño de la Interfaz	Feb 2	Mar 16	[Gantt bar from Feb 2 to Mar 16]															
			6															
Herramientas de Investigacion (SQL Server, JDBC, ODBC, etc.)	Mar 18	Abril 8					[Gantt bar from Mar 18 to Apr 8]											
			4															
Definicion de Casos de Estudio	Abril 6	Abril 17								[Gantt bar from Apr 6 to Apr 17]								
			2															
Aplicación de Integración	Abril 17	May 8								[Gantt bar from Apr 17 to May 8]								
			4															
Aplicación de esquema Integrado, DataWarehouse	May 11	May 22									[Gantt bar from May 11 to May 22]							
			2															
Analisis de Costos	May 25	May 29												[Gantt bar from May 25 to May 29]				
			1															
DEFENSA	Juni 22	Juni 25												[Gantt bar from Jun 22 to Jun 25]				
			4															

