

48

**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**



**FACULTAD DE CIENCIAS**

**CREACION DE UN ALMACEN DE DATOS  
PARA LA ADMINISTRACION DE BASE  
DE DATOS**

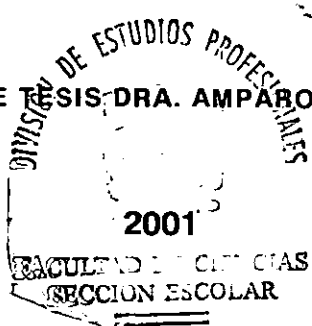
**T E S I S**

**QUE PARA OBTENER EL TITULO DE  
A C T U A R I A  
P R E S E N T A  
LOURDES LETICIA TIRADO RODRIGUEZ**



**DIRECTORA DE TESIS DRA. AMPARO LOPEZ GAONA**

290131





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**MAT. MARGARITA ELVIRA CHÁVEZ CANO**  
**Jefa de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Creación de un almacén de datos para la administración de base de datos"

realizado por TIRADO RODRIGUEZ LOURDES LETICIA

con número de cuenta 8723706-1 , pasante de la carrera de ACTUARIA

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario DRA. AMPARO LOPEZ GAONA

Propietario DRA. HANNA OKTABA

Propietario MAT. SALVADOR LOPEZ MENDOZA

Suplente M. EN I. MARIA DE LUZ GASCA SOTO

Suplente MAT. VICTOR HUGO DORANTES GONZALEZ

Consejo Departamental de  
Matemáticas

---

M. EN C. JOSE ANTONIO FLORES DIAZ

## Agradecimientos

Muchas personas han influido en mi camino para llegar hasta aquí; afortunadamente, la mayoría de ellas lo han hecho para bien. A todas ellas mi profundo agradecimiento, están siempre en mi memoria.

En primer lugar quiero manifestar mi agradecimiento a la directora de este trabajo, Dra. Amparo López Gaona, quien con su asesoramiento, sus consejos en la dirección y su confianza en mi hicieron posible la conclusión de este proyecto.

Igualmente quiero expresar mi agradecimiento a mis sinodales por dedicar tiempo a leer este trabajo, por comprender mis errores y por su sutileza para corregirme, gracias a Hanna Oktaba, Salvador López Mendoza, a Víctor Hugo Dorantes González y Maria de Luz Gasca Soto, esta última un agradecimiento especial porque adicionalmente a una profunda revisión de la tesis me proporcionó consejos y sugerencias en aspectos complementarios para la titulación.

Mi gratitud a mis profesores de todos los cursos, cada uno dejó una huella importante en mí; no puedo dejar de nombrar a mis compañeros de escuela, quienes me apoyaron en el transcurso de la carrera; en especial a Jorge Gómez.

A mis compañeros de trabajo por estar siempre al tanto de la etapa en la que se encontraba el proyecto, aportando todos su granito de arena e inyectando ánimo para continuar, especialmente a Heriberto Alejos por dedicar su tiempo en hacer la primera corrección de la redacción y estructura de la tesis.

## **Dedicatoria**

*A mis padres, por su cariño, dedicación y su deseo de que alcance metas mayores, porque siempre me visualizaron en este momento. A mi mamá por su valioso apoyo incondicional, que me permitió continuar.*

*A Miguel, Kevin y Kareem, mi esposo y mis hijos, que tantas alegrías me dan, por haber sacrificado parte de nuestro tiempo para alcanzar esta meta, porque mi deseo de impulsarlos me dio ánimo para reanudar la marcha.*

*A mis hermanas, familiares y amigos, que desean siempre lo mejor para mí.*

# ÍNDICE

## Introducción

### 1. Antecedentes

#### 1.1 Base de Datos

- 1.1.1 Evitar redundancia e inconsistencia de datos.
- 1.1.2 Acceso flexible a los datos.
- 1.1.3 Encriptación de la información.
- 1.1.4 Soportar Concurrencia.
- 1.1.5 Proporcionar Seguridad.
- 1.1.6 Facilitar Integridad.
- 1.1.7 Asegurar Disponibilidad.
- 1.1.8 Cumplir Estándares, Compatibilidad y Portabilidad.
- 1.1.9 Disponer de Conectividad y ser Distribuida.

#### 1.2 Almacén de Datos (Data warehouse)

- 1.2.1 Beneficios del Almacén de Datos
- 1.2.2 Comparación del Almacén de Datos contra Procesamiento en Línea

#### 1.3 Herramienta de Consulta

### 2. Implementación de Almacén de Datos

#### 2.1 Arquitectura de almacén de datos

- 2.1.1 Extracción de datos
- 2.1.2 Transformación
- 2.1.3 Base de Datos independiente
- 2.1.4 Acceso a los datos

#### 2.2 Ciclo de Vida del Almacén de Datos

- 2.2.1 Planeación
- 2.2.2 Requerimientos y Modelado de los Datos
- 2.2.3 Diseño físico y desarrollo de la Base de Datos
- 2.2.4 Mapeo, Carga y Transformación al Almacén de Datos
- 2.2.5 Desarrollo de la aplicación
- 2.2.6 Capacitación
- 2.2.7 Operación, mantenimiento y soporte

### 3. Administración y Mantenimiento del RDBMS de ORACLE

- 3.1 Estructura de la Base de Datos
  - 3.2 Administración de la actividad de los datos
  - 3.3 Estructura de seguridad e integridad.
  - 3.4 Procedimientos de respaldo y recuperación.
  - 3.5 Administración del Manejador de Base de Datos (DBMS)
  - 3.6 Administración del diccionario de datos
  - 3.7 Administración de espacio
  - 3.8 Documentación
- 4. Uso de la metodología de Almacén de Datos en la Administración de una Base de Datos**
- 4.1 Mercado de Datos de Administración de Espacio.
    - 4.1.1 Estructuras de espacio en la Base de Datos de un manejador Oracle
    - 4.1.2 Estándares Institucionales
    - 4.1.3 Implementación Mercado de Datos de Espacio
  - 4.2 Mercado de Datos de Sesiones
    - 4.2.1 Seguridad en el manejador de Base de Datos Oracle®
    - 4.2.2 Estándares Institucionales
    - 4.2.3 Implementación Mercado de Datos de Sesiones

## **Conclusiones**

## **Glosario de Términos**

## **Bibliografía y Referencias**

## **INTRODUCCIÓN**

La toma de decisiones no es una tarea exclusiva de las personas con funciones directivas, las resoluciones se llevan a cabo en todos los niveles y requieren estar bien fundamentadas. Es indispensable disponer de herramientas para sustentar la toma de decisiones; se requiere, durante el análisis del problema, contar con elementos que aporten fundamentos, a los comisionados de las disposiciones, para establecer las medidas adecuadas, con el fin de aumentar la confianza en el éxito de las soluciones propuestas y la reducción de incertidumbre.

Una cantidad considerable de información, útil para la toma de decisiones exitosa, se encuentra almacenada en los sistemas transaccionales de las empresas, cuyo modelado no es adecuado para proporcionar la visión dimensional requerida; adicionalmente es necesario complementar esta información con datos de otras fuentes. El uso de la tecnología es un aspecto básico en cualquier desarrollo, pero el éxito del proyecto radica en el uso adecuado de ésta. Se ha desarrollado la tecnología de Almacén de Datos (Data Warehouse) con el propósito de estructurar la metodología de los sistemas para toma de decisiones (DSS).

La tendencia es transformar la información en acción, teniendo esta circunstancia tal importancia que la información es considerada como un activo de la empresa y uno de los principales componentes de lo que se llama Comprensión del Negocio. Actualmente, la información de las empresas se encuentra almacenada en bases de datos, pequeñas o grandes, que requieren de una o más personas que las administren. A quien realiza esta tarea se le conoce como administrador de base de datos o DBA (de las siglas en inglés Database Administrator), sus actividades son extensas y de diversos tipos. Sobre el DBA recae la responsabilidad del buen funcionamiento de la Base de Datos, con este fin debe tomar decisiones importantes con respecto a los datos y consecuentemente para la empresa. Conocer a fondo el comportamiento de las bases proporciona al administrador los elementos necesarios para fundamentar sus decisiones.



Con el propósito de apoyar a los administradores de bases de datos de la institución, para conocer en forma más completa las bases de datos, y derivado de ello en tomar decisiones respecto a las bases de datos a su cargo, se desarrollaron dos sistemas utilizando la metodología de Almacén de Datos. Uno de ellos para sustentar decisiones referentes al espacio utilizado por las bases y otro que registra conexiones, este último aporta elementos para la administración de las licencias de software y para reafirmar la seguridad en las bases de datos.

Esta tesis tiene como objetivo general plantear la conveniencia del empleo de un Almacén de Datos con la finalidad de satisfacer las necesidades de soporte para la toma de decisiones.

De manera más concreta, el objetivo es mostrar el apoyo que la tecnología de Almacén de Datos brinda en la administración de una base de datos, facilitando la obtención de la información necesaria que permita soportar sus determinaciones.

La hipótesis que se plantea es que la tecnología de Almacén de Datos resuelve el manejo y uso adecuado de fuentes de datos de diversos tipos, para apoyar la toma de decisiones oportunas y fundamentadas.

Particularmente, los mercados de datos, espacio y sesiones, sustentan las decisiones del administrador de la base de datos de la institución, ya que proporcionan un conocimiento más completo del comportamiento de las bases, aportando elementos suficientes para que las resoluciones propuestas sean las más convenientes.

La tesis está estructurada en cuatro capítulos. En el primero se define lo que es un almacén de datos con las particularidades y ventajas, se describen las características de una base de datos y la herramienta utilizada para la explotación. El segundo capítulo expone las etapas de implementación de un almacén de datos. Este capítulo conduce a la aplicación de la tecnología de almacén de datos y los mercados de

datos. El tercer capítulo incluye las tareas involucradas en la administración de base de datos, y describe algunos aspectos a considerar para cada una de ellas. El cuarto capítulo explica los mercados de datos desarrollados en la institución, siguiendo el método de implementación descrito en el Capítulo dos para cubrir parte de las actividades descritas en el Capítulo tres.

# 1. Antecedentes

En este proyecto se aplica la metodología y arquitectura del Almacén de Datos para dar apoyo a la administración de una base de datos, a continuación se incluye una introducción de los tópicos involucrados.

## 1.1. *Base de Datos*

Los autores Adoración de Miguel y Piatini [DEPI9329] hacen referencia a diversas definiciones de base de datos:

“Una base de datos es un conjunto, colección o deposito de datos almacenados en un soporte informático de acceso directo. Los datos deben estar interrelacionados y estructurados de acuerdo con un modelo capaz de recoger el máximo contenido semántico.

Colección de datos interrelacionados almacenados en conjunto sin redundancias perjudiciales o innecesarios; su finalidad es servir a una aplicación o más, de la mejor manera posible; los datos se almacenan de modo que resulten independientes de los programas que los usan; se emplean métodos bien determinados para incluir nuevos datos y para modificar o extraer los datos almacenados. JAMES MARTÍN, 1975

Conjunto estructurado de datos registrados sobre soportes accesibles por ordenador para satisfacer simultáneamente a varios usuarios de forma selectiva y en tiempo oportuno. Delobel, 1982

Colección no redundante de datos compatibles entre diferentes sistemas de aplicación. Howe, 1983

Conjunto de ficheros maestros, organizados y administrados de una manera flexible de modo que los ficheros puedan ser fácilmente adaptados a nuevas tareas imprevisibles. Frank, 1988”.

Derivado de las definiciones anteriores podemos decir que una Base de Datos *es un conjunto de datos estructurados, fiables y homogéneos, organizados independientemente en archivos, accesibles en tiempo real, compartibles por usuarios concurrentes.*

Los sistemas de base de datos están diseñados para manejar grandes cantidades de información, la manipulación de los datos involucra tanto la definición de estructuras para el almacenamiento de la información como la provisión de mecanismos para la utilización de la información, además debe contar con elementos de seguridad que garanticen la integridad de la información, aun con interrupciones del sistema o intentos de accesos no autorizados.

Una base de datos debe proporcionar a los usuarios finales una visión abstracta de los datos, es decir, hacer transparente para el usuario detalles tales como la forma en que se almacenan y mantienen los datos.

Con el fin de cumplir con estos requerimientos los manejadores de Base de Datos tienen las siguientes capacidades:

#### **1.1.1. Evitar redundancia e inconsistencia de datos.**

Los archivos que mantienen almacenada la información son creados por diferentes tipos de programas de aplicación, por lo anterior, existe la posibilidad de que si no se controla detalladamente el almacenamiento se puede originar un duplicado de información. Aumentando los costos de almacenamiento y acceso a los datos, además de que se puede originar la inconsistencia de los datos, es decir, los valores para un dato almacenados más de una vez dentro de la base de datos podrían no coincidir.

### **1.1.2. Acceso flexible a los datos.**

El diseño de la base de datos debe ser cuidadosamente estudiado para ofrecer a los usuarios la facilidad de realizar sus consultas con necesidades de información distintas y no predecibles.

### **1.1.3. Encriptación de la información.**

Puesto que los datos están repartidos en varios archivos estos deben ser almacenados en cada uno de ellos con formato designado, es indispensable que el manejador cuente con un lenguaje de acceso que permita guardar y obtener los datos en forma apropiada.

### **1.1.4. Soportar Concurrencia.**

Con el fin de mejorar el funcionamiento global del sistema y obtener un tiempo de respuesta más rápido, muchos sistemas permiten que múltiples usuarios actualicen los datos simultáneamente. En estas condiciones la interacción de actualizaciones concurrentes puede dar por resultado datos inconsistentes. Para prevenir esta posibilidad deben mantenerse áreas para lectura consistente que contengan únicamente los datos que los usuarios han aceptado almacenar y de esta forma proporcionar consistencia a través de cualquier acceso que se tenga a la Base de Datos.

### **1.1.5. Proporcionar Seguridad.**

La información de toda empresa es importante, aunque unos datos lo son más que otros, por tal motivo se debe considerar el control de acceso a los mismos, no todos los usuarios pueden visualizar alguna información. Para que un sistema de Base de Datos sea confiable debe mantener un grado de seguridad que garantice la autenticación y protección de los datos.

En un banco, por ejemplo, el personal de nóminas sólo requiere acceso a la parte de la base de datos que tiene información acerca de los distintos empleados del banco y no a información financiera.

### **1.1.6. Facilitar Integridad.**

Los valores almacenados en la base de datos comúnmente deben satisfacer cierto tipo de restricciones de consistencia. El manejador debe tener la capacidad de validar dichas reglas y hacerlas cumplir a través del código apropiado.

### **1.1.7. Asegurar Disponibilidad.**

En algunas empresas se requiere tener acceso a la base sin interrupciones, incluso cuando es necesario hacer respaldo o recuperación parcial de los datos, por lo que el manejador debe contar con opciones de recuperación automáticas o manuales según la severidad del caso aun con la Base de Datos en servicio.

### **1.1.8. Cumplir Estándares, Compatibilidad y Portabilidad.**

Existen estándares definidos por organismos oficiales respecto a los lineamientos con que debe contar una base de datos, especialmente con el método de acceso, el cumplir con dichos estándares amplían la perspectiva de un manejador facilitando así la interacción con software independiente de ella así como su transportación a ambientes distintos.

### **1.1.9. Disponer de Conectividad y ser Distribuida.**

El compartir información a través de la red y entre bases de datos, permite además de la comunicación entre varios sistemas, el distribuir los recursos utilizados para un proceso entre varios nodos.

Las definiciones de las capacidades mencionadas se hicieron pensando en sistemas transaccionales, con accesos concurrentes modificando los datos, lo cual permitió a las grandes empresas automatizar sus procesos almacenando su información en forma confiable. Pero con el tiempo, la competencia entre las empresas se

intensificó, requiriendo para ello análisis detallado de sus movimientos con el fin de soportar la toma de decisiones que les permita participar en la contienda con éxito. Con el objeto de satisfacer dichas necesidades, surge el concepto de Almacén de Datos.

## **1.2. Almacén de Datos**

Almacén de Datos (Data Warehouse) no es un producto que pueda ser comprado en el mercado, sino más bien un concepto que debe ser construido, es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información a la gente de negocios. El objetivo principal es satisfacer los requerimientos de información internos de la empresa para una mejor gestión, con eficiencia y facilidad de acceso.

El Almacén de Datos convierte los datos operacionales de una organización en una herramienta competitiva, de modo que los hace disponibles a las personas que lo necesiten para el análisis y toma de decisiones. Genera bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales. Muchos Almacenes de Datos se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar. Así un Almacén de Datos resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo análisis multidimensional. De esta forma, dentro de un almacén de datos existen dos tecnologías complementarias, una relacional para consultas y una multidimensional para análisis.

Según definición de Bill Inmon [INWL93-275], el Almacén de Datos se caracteriza por ser temático, integro, histórico y no volátil. A continuación se describen las características.

**Temático:** Orientado a sujetos de la empresa, esta alineación afecta el diseño e implementación de los datos, a diferencia de un sistema transaccional, que involucra diseño de la base y diseño de los procesos, el Almacén de Datos se enfoca en el modelado de datos y diseño de la base exclusivamente.

**Integro:** Los datos deben seguir ciertas convenciones, en nombre, unidades de medida, atributos físicos. No importando de donde provengan las fuentes de datos que lo integran deberán formalizarse como un sólo tipo de datos descrito. Durante el proceso de carga de datos, debe considerarse una etapa de depuración y estandarización en los datos.

**Histórico:** Contiene fotografías de los datos en un tiempo preciso, es decir que guarda un registro del estado en que se encuentra la información a un tiempo, esta información no puede ser actualizada, lo que resulta en un constante crecimiento de la Base de Datos en el transcurso del tiempo.

**No volátil:** No debe permitir agregaciones, modificaciones, o eliminaciones de los datos que contiene, únicamente se tienen 2 tipos de operaciones en el Almacén de Datos, cargas y consultas.

### **1.2.1. Beneficios del Almacén de Datos**

A continuación se describen algunos de los beneficios del almacén de datos:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, se basa en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelación para encontrar relaciones ocultas entre los datos del almacén, obteniendo así un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones en diversos escenarios.



- Protege los sistemas transaccionales del acceso indiscriminado de usuarios que requieren de información, que podrían afectar el rendimiento de estos sistemas de misión crítica, brindando un ambiente de información independiente, protegido y bien administrado para la toma de decisiones.
- Construye un modelo de datos organizacional que permite una estandarización en el manejo de información tanto de los sistemas transaccionales como de los sistemas para el soporte a la toma de decisiones.

### **1.2.2. Comparación del Almacén de Datos contra Procesamiento en Línea**

Los sistemas tradicionales de transacciones y las aplicaciones de procesamiento analítico conocido como OLAP (de las siglas en inglés On-Line Analytical Programs), a las que pertenece el Almacén de Datos, son polos opuestos en cuanto a sus requerimientos de diseño y sus características de operación. Es de suma importancia comprender perfectamente estas diferencias para evitar caer en el diseño de un Almacén de Datos como si fuera una aplicación de transacciones en línea u OLTP (de sus siglas en inglés On-Line Transactional Programs).

Las aplicaciones de OLTP están organizadas para ejecutar las transacciones para los cuales fueron hechos. Por ejemplo, mover dinero entre cuentas, un cargo o abono, una devolución de inventario, etcétera. Por otro lado, las aplicaciones OLAP están organizadas sobre la base de conceptos; por ejemplo, el comportamiento de las ventas en un periodo específico, movimientos bancarios por región, etcétera. Carmen Gloria Wolf [WOCA00] realizó un análisis respecto a las diferencias existentes entre ambos procesamientos en distintos ámbitos.

La Tabla 1.1 muestra los contrastes entre los dos procesamientos respecto al modelado de los datos utilizado por cada uno y la naturaleza de los datos.

OLTP	OLAP
Alineación por aplicación	Alineación por dimensión
Datos organizados inherentemente por aplicación.	Datos organizados por dimensiones definidas del negocio
Enfocado en encontrar requerimientos de aplicaciones específicas para tareas concretas.	Enfocado en encontrar requerimientos de análisis empresarial
Maneja las operaciones diarias de una empresa.	Provee información utilizada para analizar un problema o situación, utiliza la comparación determinando patrones y tendencias.
Modelo de datos normalizado.	Modelo de datos dimensional
Condición de los datos cambiante.	Condición de los datos histórica y descriptiva

**Tabla 1.1 Diferencias respecto a la orientación de los datos**

OLTP	OLAP
Típicamente no integradas	Debe ser integrada
Cada tema de negocios puede tener información en diferentes sistemas	Toda información de un tema, reúne información de varios sistemas en una base
Diferentes sistemas contienen diferentes tipos de datos	Todos los tipos de datos integrados en un sistema
Diferentes convenciones de nomenclatura	Convenciones de nomenclatura estandarizados
Diferentes formatos de archivos	Formato de archivo estándar
Diferentes plataformas hardware	Un solo servidor
Los usuarios ingresan datos nuevos, abren y cierran registros, corrigen datos antiguos	Los usuarios pueden elegir los atributos más convenientes para el análisis. No pueden actualizar la información
Se ejecutan constantemente las mismas acciones	El usuario continuamente cambia el tipo de preguntas a la Base de Datos

**Tabla 1.2 Diferencias en los ámbitos de integración de la aplicación y manipulación de datos por parte de usuarios finales.**

La Tabla 1.2 muestra las diferencias respecto a las nomenclaturas, formatos y plataformas de las aplicaciones y la forma en que los usuarios finales utilizan los datos almacenados.

OLTP	OLAP
Manipulación de datos registro a registro	Carga y acceso de datos en forma masiva
Transacciones y / o rutinas de validación en el ámbito de registro	Validación realizada antes o después de cada carga
Se manejan cientos de transacciones por día	Se maneja sólo una transacción que contiene cientos de registros.
Si está bien la transacción, se asegura la consistencia de ese único conjunto de datos.	Si la carga termina bien, se tiene la consistencia asegurada de todo el conjunto de datos.
Hay una falla de soporte explícito para reconstruir la historia previa.	Similar a las capas geológicas, la base de datos dimensional puede verse como una serie de capas de datos, compuestas cada una por una imagen del OLTP, tomadas a intervalos regulares.
Datos de operaciones son altamente volátiles.	Los datos del Almacén de Datos son altamente estables, son insertados en intervalos definidos y no son modificados.
Puede haber cambios en la Base de Datos mientras se está consultando	La información que se está consultando no puede ser actualizada.

**Tabla 1.3 Diferencias en el ámbito de la administración de las transacciones y la dimensión del tiempo.**

En la tabla 1.3 se muestran las diferencias respecto a la administración de las transacciones y la relación de los datos respecto al tiempo. En el procesamiento OLTP, si se mantiene un historial, al aumentar los cambios es casi imposible reconstruir rápidamente hasta un punto pasado, si se rescribe sobre los datos, no se puede recuperar un estado anterior, a diferencia de la filosofía OLAP en la que se puede observar el estado de los datos en algún momento del pasado e inclusive proyectar hacia el futuro.

Un elemento indispensable en el concepto de Almacén de Datos es la herramienta de acceso con características especiales, existen una gran diversidad con diferentes capacidades, una de ellas es Oracle Discoverer.

### **1.3. Herramienta de Consulta**

Oracle Discoverer® es una de las herramientas para la toma de decisiones de Oracle®, es intuitiva y abierta, de consultas *ad hoc*, informes, análisis y publicación en web, que permite a los usuarios de negocios de todos los niveles en la organización tener acceso inmediato al Almacén de Datos relacional, centro de datos o sistemas de procesamiento de transacciones en línea.

El nivel de usuario final se puede crear a partir de cualquier repositorio de metadatos relacional ocultando la complejidad de las estructuras de datos y las relaciones de profundización de usuarios de empresa que necesitan centrarse en los negocios, no en los datos. Discoverer es reconocida como la herramienta de consultas *ad hoc*, informes y análisis más sencilla de utilizar, presentando la terminología de bases de datos en lenguaje comercial. Permite a los usuarios crear informes rápidamente aprovechando las condiciones y cálculos predefinidos disponibles en el nivel de usuario final.

Generar consultas con el editor de usuario es una tarea directa y automática. Se seleccionan los datos necesarios, manipulando el mouse, se modifica directamente el diseño del área de trabajo. Es posible generar filtros condicionales, porcentajes, totales y elementos calculados con la interfaz de asistentes paso por paso. No es necesario tomar en cuenta tipos de datos, paréntesis, nombres de función o valores de datos.

Cuenta con plantillas de informes estándar (tabular, tabular cruzado, clasificación de grupos y detalle maestro) para satisfacer los requisitos del negocio. Permite aplicar distintos estilos y tamaños de fuentes y colores en primer plano y de fondo a las

cabeceras y los datos. Habilita la inclusión de títulos de áreas de trabajo, numeración de páginas, condiciones y parámetros para identificar fácilmente el juego de datos al que un informe se refiere.

Permite exportar informes a un amplio número de formatos de archivo comunes, como HTML, XLS y TXT, para permitir la integración de datos de informes con otras herramientas de aplicaciones de escritorio. Accesa a bases de datos distintas de Oracle mediante ODBC (por sus siglas en inglés Open Database Connectivity).

Soporta la definición de excepciones en juegos de datos de gran tamaño, definiendo formatos diferentes en los resultados de la consulta que pueden ser importantes, para identificar fácilmente los rangos de interés y producir un efecto de semáforo al complementar los datos con colores.

Amplía las posibilidades de profundización en los datos hasta aplicaciones soportadas, como vídeo clips, archivos de sonido, archivos de texto, archivos de procesadores de textos o de hojas de cálculo. Esta extensión proporciona a los usuarios opciones abiertas para acceder a información adicional almacenada fuera de las estructuras de la base de datos. Los archivos externos se pueden almacenar en el sistema de archivos o en la base de datos Oracle.

Se integra fluidamente con productos complementarios OLAP para ofrecer soluciones de soporte de decisiones para la funcionalidad avanzada de generación de informes y análisis, para el diseño del Almacén de Datos y de las bases de datos, así como de la gestión de la generación.

Para un Almacén de Datos habilita el uso de funciones de profundización en cualquier dirección y de giro sobre un eje para explorar la información almacenada. La arquitectura de profundización permite dividir y sectorizar los datos para consultar resultados de distintas formas, lo que proporciona un cuadro completo y exacto del negocio sin tener que escribir consultas complejas en SQL (por sus siglas en inglés

Standard Query Language). En un Almacén de Datos, se resumen previamente los grandes volúmenes de datos para aumentar el rendimiento. Al solicitar información de una tabla de detalle de grandes volúmenes, Discoverer redirige las solicitudes de forma transparente a la tabla resumida.

Puede tomarse provecho de la estrecha integración de Discoverer con la base de datos Oracle para simplificar la seguridad, la escalabilidad, el acceso a los datos, inversión de la profundización hasta aplicaciones externas almacenadas en la base de datos, programación de lotes, predicción de consultas, gestión automática de resúmenes y creación de metadatos.

Como otras herramientas, Discoverer proporciona un determinador de recursos para interrumpir consultas de larga ejecución y controlar el uso de recursos. Además, Discoverer ofrece una función exclusiva de consultas que determina de antemano el tiempo que tardará el informe en ejecutarse. Esto reduce problemas y otorga al usuario control directo sobre las consultas que se ejecutan y cuándo ejecutarlas. Si una consulta va a durar mucho en línea, permite programarla para que se ejecute en otro momento mediante el cuadro de diálogo de predicción de consultas o la utilidad de programación.

La edición de administración, se utiliza para configurar el nivel de usuario final. Permite configurar operaciones en lotes, controles, por usuario o base pública, el número de informes mantenido, el tiempo de caducidad por defecto de los informes, el período en el que los usuarios pueden ejecutar informes y si los usuarios sólo pueden ejecutar informes en lote. Dispone de un asistente gráfico para crear y modificar las áreas de negocio. Dispone de una interfaz de uso sencillo para la configuración, la cual al igual que la escalabilidad y la seguridad, están basadas en el servidor.

## 2. Implementación de Almacén de Datos

Un Almacén de Datos es un rico depósito de información orientado a resolver las necesidades y especulaciones del usuario final en el ámbito de la toma de decisiones. Al brindar un acceso interactivo a la información permite resolver problemas de toma de decisiones en forma más rápida y efectiva que utilizando sistemas tradicionales. Dentro de una empresa el uso de un Sistema de Almacén de Datos se aplica no sólo en el soporte de la gerencia de alto nivel que debe tomar decisiones de objetivos y planificar estrategias, sino también en niveles más técnicos de toma de decisión.

El Almacén de Datos por sí sólo se puede decir que es una arquitectura y cierta metodología para hacerla realidad.

### 2.1. Arquitectura del Almacén de Datos

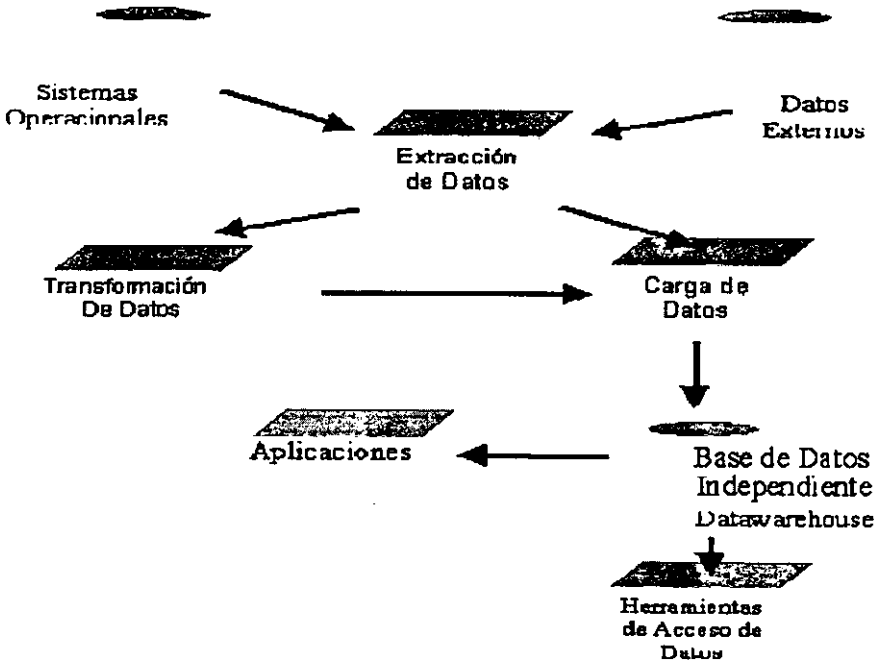


Figura 2.1 Componentes de un Almacén de Datos

Los componentes principales en la arquitectura de un Almacén de Datos, ilustrados en la Figura 2.1, se describen a continuación:

- Los datos son extraídos desde los sistemas, bases de datos y archivos fuente.
- Los datos de los sistemas fuente son integrados y transformados antes de ser cargados dentro del Almacén de Datos.
- El Almacén de Datos es una Base de Datos independiente de los sistemas fuente con acceso a los usuarios de lectura únicamente, creada específicamente para toma de decisiones.
- Los usuarios consultan el Almacén de Datos a través de herramientas o aplicaciones.

### **2.1.1. Extracción de datos**

Los datos identificados como necesarios para procesar la toma de decisiones serán extraídos desde los sistemas fuente, cuya naturaleza no necesariamente debe ser única. Otras fuentes pueden ser datos proporcionados por compañías especializadas del mercado. Estos datos podrían estar en varios formatos o medios distintos, de modo que es posible que la selectividad de datos no sea necesaria.

En general, es natural que las fuentes de datos procedan de múltiples sistemas y aplicaciones y que los campos fuente sean de diferentes bases de datos, plataformas distintas, y una extensa variedad de tipos de dato y formatos. Por lo que en este componente de la arquitectura se deben resolver problemas técnicos importantes, como serían: acceso a sistemas heterogéneos, carga sincrónica contra carga asincrónica, ejecución de consultas complejas afectando lo mínimo el rendimiento de las bases fuentes, etcétera.



### **2.1.2. Transformación**

Un componente importante de una arquitectura del Almacén de Datos es que los datos de múltiples fuentes son integrados y transformados antes de ser cargados dentro del Almacén de Datos. Esta integración se refiere a la estandarización de los datos, identificar información que tiene formato diferente o que por alguna razón se encontraba guardada en los sistemas fuente en forma duplicada y que no coinciden.

La calidad de datos en el Almacén de Datos involucra, no sólo la consistencia clásica en bases de datos sino también las nociones de pertinencia y relevancia de los datos almacenados. Por pertinencia, se entiende la utilidad de un dato para ser usado en la toma de decisiones, la relevancia de un dato es su significado como campo de información.

### **2.1.3. Base de Datos independiente**

Las Bases de Datos Operacionales son diseñadas para almacenar, actualizar y reportar día a día los datos del negocio con rapidez, mientras que los sistemas de toma de decisión son diseñados para brindar fácil acceso a información histórica, de sólo lectura y analítica. La diferencia fundamental entre estos dos sistemas es que requieren diferentes tipos de Base de Datos. Un diseño de base de datos no puede brindar ambos tipos de funcionalidad.

### **2.1.4. Acceso a los datos**

La arquitectura de Almacén de Datos incluye una herramienta o aplicación de acceso, así como el entrenamiento y soporte necesarios.

Existen herramientas de explotación de diferentes tipos. Las clásicas son: interfaces para consultas y reportes complejos, productos de análisis de datos (OLAPs) y Minería de Datos.

**Herramientas para realizar consultas y reportes complejos.** Permiten al usuario construir gráficas y reportes a partir de la información contenida en el Almacén de Datos y descrita a través del Diccionario de Datos. Algunas funcionalidades típicas de estas herramientas son: agrupamiento y desglosado dinámico de datos en reportes, cambios en el orden de los campos del reporte, visualización del resultado de las consultas en forma gráfica (barras, pastel, histogramas, etcétera). Estas herramientas generan las expresiones en el lenguaje de consulta que recupera los datos pedidos (típicamente SQL), se conectan al Almacén de Datos, recuperan el resultado y lo formatean según la especificación dada.

**OLAPs (On-Line Analytic Processing).** Permiten representar los datos del problema en términos de dimensiones. De esta manera, las consultas de análisis de datos de una dimensión en función de la otra se realizan en forma inmediata. Las herramientas OLAP pueden ser servidores, si almacenan los datos en este modelo dimensional; o clientes si se conectan a un servidor de base de datos y transforman los requerimientos dimensionales en términos relacionales. La utilización de unos u otros depende del tipo de problema a resolver. No existe un modelo que sirva para todos los problemas.

**Técnicas de Minería de Datos.** Permiten explorar el Almacén de Datos en búsqueda de relaciones desconocidas o inesperadas entre los datos. Más que analizar los resultados del negocio (OLAP), permiten modelarlo construyendo patrones estadísticos que lo identifiquen, utilizando técnicas informáticas como algoritmos matemáticos, métodos estadísticos, lógica difusa, algoritmos genéticos, inducciones de reglas, sistemas expertos (sistemas basados en el conocimiento), redes neuronales, descubrimiento de reglas y modelación predictiva para poder obtener patrones y asociaciones.

## **2.2. Ciclo de Vida del Almacén de Datos**

El ambiente de procesamiento de un Almacén de Datos es muy variado. Sin embargo la mayoría de los proyectos de Almacén de Datos requieren seguir cuidadosamente las siguientes fases de implementación:

1. Planeación
2. Obtención de requerimientos y modelado de datos
3. Diseño y desarrollo de la Base de Datos
4. Mapeo, carga y transformación de datos
5. Desarrollo de la aplicación
6. Capacitación
7. Operación, mantenimiento y soporte

### **2.2.1. Planeación**

Es una fase importante de la implementación del Almacén de Datos. Las decisiones tomadas durante la fase de planeación tienen un impacto significativo en el ámbito de implementación y en la magnitud del esfuerzo.

La fase de planeación para un Almacén de Datos, requiere, además de las tareas cotidianas de un sistema como la determinación del alcance del proyecto, establecer tareas y distribuir las, definir tiempos y establecer liberación final del proyecto, tomar las siguientes consideraciones:

- Planes de Capacitación.
- Estrategias de documentación.
- Procedimientos para respaldar.
- Estrategias de actualización de datos.
- Estrategias de operación y programación de tareas.
- Conectividad de Base de Datos.
- Utilerías de carga.
- Configuración de estaciones de trabajo.

Durante la etapa de planeación se deben llevar a cabo las siguientes actividades:

- Selección de la estrategia de implementación
- Selección de Metodología de Desarrollo
- Desarrollo de los objetivos empresariales
- Enfoque de Selección de una arquitectura
- Crear un plan de programa

Estas actividades con descritas con mayor detalle a continuación.

### **2.2.1.1. Selección de la estrategia de implementación**

Existen diversas estrategias de implementación, las de uso más común son las siguientes:

***Enfoque de arriba hacia abajo:*** Se identifican los requerimientos corporativos que debe cubrir el Almacén de Datos, útil cuando la tecnología está madura y se comprende bien, y cuando están claros y bien entendidos los problemas empresariales que deben resolverse.

***Enfoque de abajo hacia arriba:*** Generalmente comienza con experimentos y prototipos basados en tecnología. Es más rápida, útil en las primeras etapas de madurez de la tecnología.

***Combinado de los anteriores:*** Con este enfoque una organización puede explotar la naturaleza planeada y estratégica del enfoque de arriba hacia abajo, al tiempo que conserva la rápida implementación y aplicación oportuna del enfoque de abajo hacia arriba. Se adapta mejor al despliegue rápido y oportunista de la tecnología del Almacén de Datos, al tiempo que se reserva el derecho de construir una solución estratégica que tenga valor de largo plazo.

### **2.2.1.2. Selección de Metodología de Desarrollo**

Los requerimientos de la implementación de un Almacén de Datos descartan el uso de cualquier metodología que requiera una fase prolongada de acopio de requerimientos y análisis, una fase de desarrollo monolítico que tome muchos meses y una fase de despliegue que ocupe también varios meses. Dos tecnologías comunes en el desarrollo de un Almacén de Datos son el método de análisis y diseño estructurado y el método de desarrollo espiral.

#### ***Método de análisis y diseño estructurado (cascada).***

Este método consiste en reunir primero los requerimientos, se analizan y se subdividen de manera progresiva, después se construye un diseño mediante los resultados del análisis. El diseño inicia como un nivel abstracto y se divide sucesivamente en niveles más concretos.

#### ***Método de desarrollo espiral***

En este método el énfasis está en la velocidad y la culminación, con un reconocimiento de que los requerimientos no se pueden identificar con claridad o especificar al inicio, de modo que se establecen algunos requerimientos se les da una solución y vuelven nuevamente a definir requerimientos adicionales.

### **2.2.1.3. Desarrollo de los objetivos empresariales**

La definición del problema, es uno de los pasos más importantes en el desarrollo. En ella deben definirse el problema específico que resuelve, recursos disponibles y el criterio que se utilizará para medir el éxito. La meta tiene una influencia considerable en el alcance del proyecto y establece un margen de modo que el alcance no cambie y se expanda continuamente.

#### **2.2.1.4. Enfoque de Selección de una arquitectura**

Las opciones arquitectónicas existentes son las siguientes:

##### ***Almacenamiento operacional***

En esta arquitectura el metamodelo del Almacén de Datos es un esquema virtual que apunta a los metadatos de bases de datos operacionales. Es decir, no existe una base de datos donde se guarda la información seleccionada sino que se utilizan los datos existentes en el sistema transaccional, es una forma rápida de implementar un Almacén de Datos. Esta opción es utilizada cuando se desea reducir redundancia y volúmenes de información. En esencia trata de mantener el número mínimo de copias como sea posible de cada pieza de datos, como es de suponerse tiene algunas limitantes, entre ellas la imposibilidad de consultar información histórica.

##### ***Sólo Almacén de Datos***

Esta arquitectura reconoce que todos los datos fuente requieren diversas operaciones como depuración, integración, resumen y adición. Por lo anterior, tiene sentido aplicar estas operaciones una vez y traer los datos de las fuentes a un Almacén de Datos centralizado. En esta arquitectura un sólo Almacén de Datos alimenta a todos los usuarios finales con información para apoyo en las decisiones.

##### ***Mercado de Datos***

Esta arquitectura reconoce que cada departamento funcional tiene sus propias necesidades específicas y que un sólo Almacén de Datos corporativo no puede satisfacer todas estas necesidades. Estos sistemas tienen el inconveniente de que son difíciles de extender para ser utilizados por otros departamentos, primero por las limitaciones inherentes del diseño al concentrarse en un conjunto de necesidades del negocio, en adición a que cualquier expansión podría ocasionar una ruptura para los usuarios actuales.

### ***Almacén de Datos y Mercado de Datos***

Esta arquitectura es un reconocimiento de que las necesidades de Almacén de Datos específicas de un departamento se deben abordar junto con la necesidad de un Almacén de Datos corporativo.

### ***Cliente Servidor de dos capas***

Esta arquitectura emplea la separación de plataformas para las fuentes de datos, Almacén de Datos, los Metadatos y las estaciones de trabajo. Esta arquitectura permite la distribución del procesamiento en dos entidades, ya que las herramientas de acceso del usuario final se ejecutan en el cliente y las fuentes de datos de Almacén de Datos residen en servidores, reduciendo así contención, una desventaja es el alto nivel de duplicación de información y la necesidad de coordinación entre las máquinas cliente.

### ***Cliente Servidor de tres capas***

En esta opción existen tres capas: una capa cliente, una capa intermedia con base en un servidor y una tercera responsable de manejar las fuentes de datos y las transformaciones opcionales de las fuentes de datos. Los servidores manejan el software del Almacén de Datos. Las estaciones de trabajo cliente manejan las aplicaciones de consulta y reporte.

#### **2.2.1.5. Crear un plan de programa**

Un plan de programa es una visión general de la actividad del Almacén de Datos y su función en la vida diaria y semanal de la organización. Identifica los diversos departamentos y unidades empresariales que usarán el Almacén de Datos. Así como, establece prioridades sobre la implementación.

Esta actividad incluye también la definición de los recursos necesarios, internos y externos. Existen dos opciones para hacer un estimado de los costos: con base en el

historial de la organización en el desarrollo de software o con base en la arquitectura de referencia.

## **2.2.2. Requerimientos y Modelado de los Datos**

La fase de requerimientos de la implementación del Almacén de Datos es una especificación precisa de las funciones que se obtendrán del Almacén de Datos, debe considerarse la diversidad de requerimientos. La cuidadosa selección y especificación de requerimientos en esta etapa proporciona un proyecto cimentado que arroja resultados con rapidez.

Durante esta fase es indispensable comprender las necesidades del negocio y diferentes requerimientos de los participantes del Almacén de Datos. Incluye también modelado de estos requerimientos.

### **2.2.2.1. Requerimientos del propietario**

Dentro de los requerimientos es importante definir las Áreas Temáticas, las cuales son los temas de interés de diversas funciones, es decir los datos que cada área de la empresa requiere; y los que a su vez generan temas específicos para los departamentos.

Un requerimiento importante a definir es la granularidad, que se refiere al nivel de detalle de la información. La granularidad tiene una relación directa con las actividades de resumen y adición que deben realizarse sobre los datos fuente. A menor granularidad, mayor cantidad de detalle. El nivel de granularidad afecta directamente el tamaño de la base de datos y los tipos de análisis que la Base de Datos puede soportar.

Dado que el Almacén de Datos organiza un gran conjunto de datos operacionales e históricos mediante múltiples dimensiones de categorización, se deben definir dichas



dimensiones, una importante es el tiempo. Las dimensiones son datos que describen los hechos y reflejan las dimensiones del negocio.

#### **2.2.2.2. Requerimientos del arquitecto**

El arquitecto es la persona responsable de diseñar los diversos componentes del Almacén de Datos para sustentar las necesidades actuales y futuras, el debe determinar el rango de funciones y características a ofrecer, el rango de plataformas necesarias para la implementación, el uso de estándares e interfaces abiertas, así como, la flexibilidad para incorporar mejoras.

Un enfoque muy bien estructurado para el diseño de arquitectura es la planeación de arquitectura empresarial, cuya metodología ha desarrollado tres tipos de arquitectura la de datos que escribe los elementos de datos y sus relaciones, definido en general como Modelos Entidad-Relación, la de aplicación sobre la base de que un sistema es una combinación de diversas aplicaciones junto con las funciones que ofrecen y las interfaces entre ellas, y la de tecnología que describe todos los componentes de tecnología.

#### **2.2.2.3. Requerimientos del desarrollador**

Los requerimientos del desarrollador son de diferente naturaleza. Los requerimientos de tecnología incluyen fuente de datos, Almacén de Datos, Mercado de Datos, el acceso del usuario final y herramientas.

Otro tipo de requerimiento es el despliegue que está relacionado con la capacidad del Almacén de Datos de proporcionar acceso para distribuir información de manera oportuna y conveniente. Se toman en consideración: métodos de acceso, procedimientos de recuperación, herramientas de acceso, requerimientos de conectividad, requerimientos de plataforma cliente.

Respecto a la disposición para la producción del Almacén de Datos el desarrollador requiere mantener la consistencia, confiabilidad y actualidad de la información,

manejar los metadatos y el metamodelo del Almacén de Datos, asegurar que los mecanismos de transporte, Base de Datos, computadoras y mecanismos de comunicación estén listos y disponibles todo el tiempo, soporte técnico inmediato y capacidad de ayuda en línea para asistir a los usuarios cuando falle el sistema o cuando tengan preguntas operacionales, establecer políticas y procedimientos de seguridad de acceso y autenticación, administrar el tamaño de las bases de datos y mejorar el tiempo de respuesta de acceso.

Para el desarrollo y despliegue del personal y sus habilidades el desarrollador debe contar durante las distintas fases del desarrollo del Almacén de Datos con un amplio rango de habilidades. Una actividad útil de acopio de requerimientos consiste en identificar cuales habilidades se requieren para cada fase del ciclo de vida del desarrollo del Almacén de Datos.

#### **2.2.2.4. Requerimientos del Usuario Final**

Para el usuario final, el Almacén de Datos es una gran caja negra cuyo acceso principal es a través de aplicaciones y herramientas de consulta y reportes.

El usuario final necesita se ajuste la funcionalidad que ofrece el Almacén de Datos con su flujo de trabajo diario. Es decir, que en el Almacén de Datos, él pueda consultar datos útiles para su ocupación.

Dentro de sus necesidades de consulta, el usuario hace uso de escenarios como herramienta para hacer prototipos de las capacidades del Almacén de Datos y para establecer sus expectativas.

Las necesidades del usuario final consisten básicamente en tener acceso rápido a la información, que le permita la manipulación satisfactoria y excelente presentación de los reportes, tener en consideración las necesidades de una variedad de usuarios, permitir a los usuarios la creación de sus propias consultas utilizando términos

empresariales que ya conocen y ofrecer estructuras de datos consistentes, profundizar en los datos del reporte, disponer de capacitación y soporte mínimos así como manejar el hardware, software y el RDBMS (por sus siglas en inglés Relational Database Management System) del Almacén de Datos actuales.

También puede especificar los tipos de análisis de datos que desean realizar sobre los datos. A continuación se mencionan algunas de las facilidades para las consultas que el usuario final debería tener:

- Opción de dividir y seccionar la información
- Poder separar elementos de datos en varias formas
- Penetrar la información, es decir, exponer más detalle de los datos de manera progresiva
- Facilidad de buscar patrones ocultos de datos, conocido como minería de datos, navegar en el modelo de datos para examinar de una manera directa,
- Posibilidad de hacer copias y realizar modificaciones locales, que les permita construir modelos empresariales, como por ejemplo hojas de cálculo,
- Opciones de visualización de datos en dos dimensiones como las hojas de cálculo, o en múltiples dimensiones, reportes y cuadros.

Después de definir los requerimientos debe hacerse un análisis que consiste en convertir estos requerimientos en un conjunto de especificaciones que puedan apoyar el diseño, es decir al modelado de los datos.

Existen las Bases de Datos Multidimensionales, consideradas una extensión del modelo de hoja de cálculo, dichas bases almacenan los datos en arreglos, lo cual permite que la información sea actualizada sin afectar el índice. La funcionalidad de estas herramientas es alta pero limitada en escala y características. Una ventaja de este modelado es que es capaz de proveer una respuesta más eficiente que se basa principalmente en función de anticipar la manera en la cual los datos serán accedados.

### **Hipercubo:**

Es un concepto similar al cubo que no cumple con la característica de que todos sus lados son iguales, su significado es el de un objeto con más de tres dimensiones, con lados planos y cada dimensión en ángulo recto con las otras dimensiones.

Algunas técnicas de modelado que adecuan un Modelo Entidad-Relación para un sistema de Toma de Decisiones o Multidimensional serían:

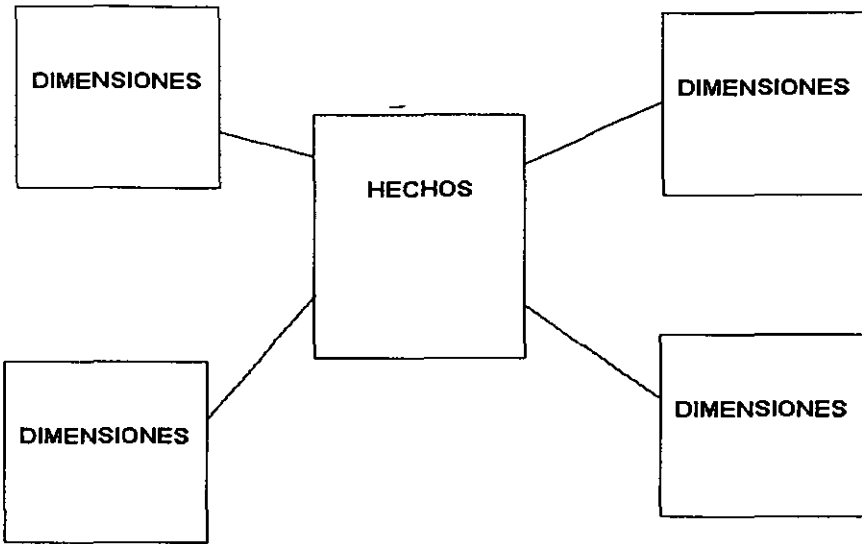
- Esquema estrella
- Esquema copo de nieve
- Esquema mixto

Estas técnicas son descritas a continuación

### **Esquema Estrella**

Como el nombre sugiere, el esquema estrella es un paradigma de modelado que tiene un sólo objeto en medio conectado con varios objetos de manera radial. El esquema estrella refleja la visión del usuario final de una consulta empresarial. El objeto en el centro de la estrella se denomina tabla de hechos y los objetos conectadas a ella se denominan tablas de dimensión. La tabla de hechos contiene la información básica a nivel transacción del negocio que es de interés para una aplicación en particular.

Un esquema estrella lógico sencillo consta de una tabla de hechos y varias tablas de dimensión. Los esquemas estrella complejos, tienen cientos de tablas de hechos y de dimensión. Una tabla de hechos contiene las mediciones básicas de los negocios y consta de millones de hileras. Las tablas de dimensión contienen atributos de negocios que se emplean como criterio de búsqueda SQL y son relativamente pequeñas.



**Fig. 2.2 Esquema estrella**

El esquema estrella brinda un rápido tiempo de respuesta, su diseño puede ser fácilmente modificado, pone en paralelo el pensamiento del usuario final y el uso de los datos, simplifica la comprensión y navegación en el metadato para usuarios y desarrolladores

### **Esquema copo de nieve**

El esquema copo de nieve es una extensión del esquema estrella en donde cada uno de los puntos de la estrella se divide en más puntos. En esta forma de esquema, las tablas de dimensión del esquema estrella contienen más normas. Las ventajas que proporciona el esquema copo de nieve son mejorar el desempeño de consultas debido a un mínimo almacenamiento en disco para los datos y mejorar el desempeño mediante la unión de tablas más pequeñas con normas, en vez de las grandes tablas sin normas. El esquema copo de nieve incrementa la flexibilidad de las aplicaciones debido a la aplicación de normas y, por lo tanto, disminuye la granularidad de las dimensiones.

El esquema copo de nieve aumenta el número de tablas con las que el usuario debe tratar e incrementa la complejidad de algunas de las consultas que debe localizar. Algunas herramientas nuevas ocultan a los usuarios el esquema físico de la base de datos y le permiten trabajar a un nivel conceptual. Las herramientas ubican las consultas del usuario en el esquema físico. Necesita de un administrador de Base de Datos para efectuar las ubicaciones.

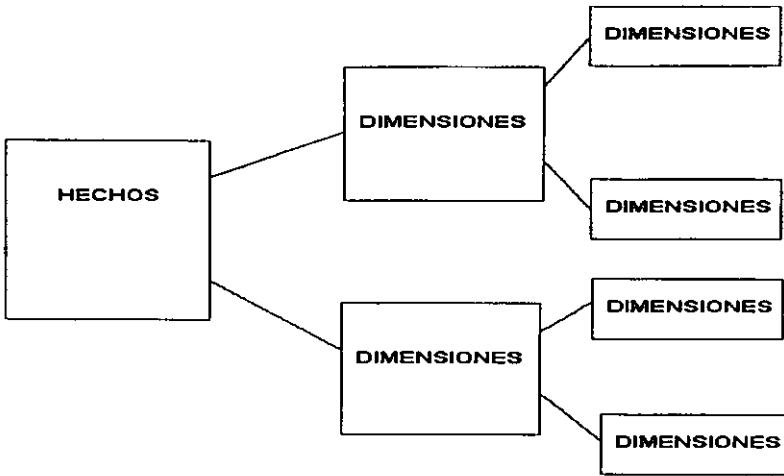


Fig. 2.3 Esquema copo de nieve

### Esquema Mixto

El esquema mixto es un ajuste entre el esquema estrella, basado en tablas de hechos y tablas de dimensión sin normas, y el esquema copo de nieve, en donde todas las tablas de dimensión tienen normas. En el esquema mixto, sólo se aplican normas a las tablas de dimensión más grandes. Por lo general, estas tablas contienen volúmenes de columnas de datos sin normas.

Los esquemas estrella y copo de nieve son adaptaciones de un modelo relacional para asemejarse al modelo multidimensional.

### **2.2.3. Diseño físico y desarrollo de la Base de Datos**

En la fase de Diseño los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Se identifican y detallan, también, los procesos que requiere, de manera interna, cada bloque de la arquitectura de referencia del Almacén de Datos. Determina la forma en que se cubrirán los requerimientos, después estas especificaciones se usan para generar extractores del Almacén de Datos y software de transformación, integración, resumen y adición.

Se distinguen dos actividades principales el diseño detallado de la Base de Datos y el correspondiente a la aplicación.

#### **Arquitectura de datos**

Dentro de la arquitectura de datos se debe considerar el desarrollo de modelos físicos de datos para las Base de Datos que deposita el Almacén de Datos y mercado de datos, la correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del Almacén de Datos o mercado de datos, el diseño de la Base de Datos, incluyendo tablas de hechos, relaciones entre tablas y tablas de descripción (dimensiones), la desnormalización de datos, identificar llaves, la estrategia de creación de índices y la creación de objetos de la base apropiados.

#### **Arquitectura de aplicación**

En la arquitectura de la aplicación se incluyen los procesos que son internos a las fuentes de datos y se relacionan con depuraciones o extracciones parciales de información y procesos que conectan las fuentes de datos con el Almacén de Datos. Los procesos que son internos al Almacén de Datos y se usan para fines de manejo interno, se incluyen también los métodos de conexión entre el Almacén de Datos con los mercados de datos o con las herramientas de usuario final. Están incluidos además los procedimientos internos del Almacén de Datos y los de los mercados de datos, las tareas internas del Almacén de Datos y los mercados de datos para configurar herramientas de análisis, los que sustentan tareas de manejo, las

actividades de administración y prácticas internas para el Almacén de Datos como sistema.

Existen diversas metodologías de diseño. Dos de ellas muy usadas son:

Prototipo rápido, recomendado para proyectos de pequeños.

Desarrollo Estructurado, especialmente para proyectos grandes o muy complejos.

### **2.2.3.1. Prototipo Rápido**

El éxito de esta metodología depende de que se cumplan con varios aspectos. Primero en formar un pequeño equipo de base de datos, programadores, técnicos de hardware, diseñadores, técnicos de control de calidad, documentación y especialistas en toma de decisiones y un sólo administrador. Segundo, definir e involucrar a un pequeño grupo de usuarios, con y sin experiencia. Estos usuarios brindarán las bases necesarias para manejar el ciclo del prototipo, escucharlos aumentará las probabilidades de éxito del prototipo. Tercero, generar un manual de usuarios y una interface de usuario como primer punto; son sorprendentes los resultados de producir un manual de usuario primero. Cuarto, utilizar herramientas específicamente diseñadas para prototipo rápido como Herramientas Visuales. Por último tener en mente que el prototipo no es la aplicación final. Una vez que el prototipo tiene éxito, entonces inicia el proceso de desarrollo.

### **2.2.3.2. Desarrollo Estructurado**

Cuando un proyecto cuenta con más de 10 personas involucradas se requiere un desarrollo estructurado. El prototipo rápido podría ser un subconjunto de un desarrollo estructurado. Este aplica a la mayoría de los desarrollos de Almacén de Datos. La documentación es extensa, el control de calidad es crítico y el número de revisiones se incrementa.



#### **2.2.4. Mapeo, Carga y Transformación al Almacén de Datos**

Durante esta fase se debe tener en consideración la definición de los sistemas fuente de los cuales se va a extraer la información, determinar los archivos de esquemas, escribir el desarrollo de las especificaciones de transformación para transformaciones sofisticadas, el mapeo de datos de fuente a Base de Datos objetivo, así como revisar planes de Capacitación.

También debe determinarse en cada fuente de datos la existencia de procesamiento anual o sumariación, la existencia de requerimientos especiales de seguridad, definir el ciclo de actualización y la frecuencia en que cambia la jerarquía.

Esta fase incluye el proceso de extracción, transformación y carga de datos en la base objeto (Almacén de Datos) como son los programas que creen y modifiquen las bases de datos para el Almacén de Datos y los Metadatos, programas que extraigan datos de fuentes relacionales y no relacionales.

En esta fase corresponde definir los procesos que realicen transformaciones de datos, tales como integración, resumen y adición así como los programas que efectúen búsquedas en bases de datos muy grandes.

La fase de transformación complementa el conjunto de funcionalidad requerida. Incluye varias funciones genéricas de transformación, como son: *la selección*, selecciona datos de acuerdo a un criterio predefinido; *separación*, divide o une la información relacionada a un área del negocio dentro de registros separados; y *normalización*, divide un simple registro de entrada en varios; todas ellas requeridas para reestructurar los diversos y dispersos datos y obtener así una estructura genérica requerida para el Almacén de Datos. *La conversión y el enriquecimiento* se requieren para limpiar y complementar los datos operacionales. Esta fase es la que más requiere de código de programación personalizado.

Para la fase de carga existen diversas opciones, el *reemplazo* donde la Base de Datos completa se sobrescribe con los nuevos datos, la de *adición* que conserva los datos actuales y adiciona los nuevos, la *fusión destructiva* la cual detecta la coincidencia entre llaves existentes y las nuevas sobrescribiendo el dato, de no ser así inserta el dato como nuevo o la de *fusión constructiva* que consiste en insertar siempre los datos nuevos en el caso en que encuentre coincidencias entre las llaves marca actualiza los existentes. La elección del método que se utilizará depende del tipo de dependencia requerida de los datos con el tiempo, las opciones son de fotografía, en el que se tiene una vista estática de los datos en un punto en el tiempo, y se utiliza una opción de reemplazo o fusión destructiva, datos transitorios muy similar al de fotografía con la diferencia de que este último es cargado en intervalos y los transitorios continuamente y datos periódicos en donde los registros no deben ser borrados nunca en este caso la fusión constructiva es la más efectiva.

Durante esta fase la automatización de la extracción, transformación y carga de los datos al Almacén de Datos, así como crear procedimientos de respaldo y recuperación y conducir una prueba completa de todos los procesos automatizados.

### **2.2.5. Desarrollo de la aplicación**

La extensión de esta fase depende del tipo de herramienta que se haya elegido para explotación, puede ir desde una herramienta comercial que el propio usuario puede utilizar hasta código en lenguajes de programación para producir reportes, debe tenerse en consideración el crear un conjunto de reportes predefinidos y el desarrollo de reportes esenciales, realizar pruebas de ellos, documentar la aplicación y el diseño de rutas de navegación

Los retos en esta fase son entender como incorporar las inversiones existentes en plataformas, tecnología y habilidades, la toma de decisiones inteligentes como "hacer en comparación con comprar", la selección y evaluación adecuada de componentes suministrados por el fabricante, la capacidad de integración de

sistemas para conjuntar los sistemas de fabricante con las fuentes de datos existentes y las herramientas de acceso de datos existentes y propuestas y la administración de metadatos.

Las características básicas de un reporteador para tener acceso a un Almacén de Datos son:

**Acceso cruzado a través de los atributos de la dimensión:** Se refiere a la capacidad de una herramienta de presentar los valores válidos de hechos sujeto a cualquier restricción sobre las dimensiones.

**Múltiples SQL:** Para calcular comparaciones o para calcular correctamente medidas no aditivas en el reporte, la herramienta debe desmenuzar la consulta dentro de múltiples consultas que son procesados separadamente por el RDMBS, la herramienta combina automáticamente los resultados de las consultas separadas de forma inteligente.

**Distinción de datos importantes:** La herramienta debe tener la capacidad de filtrar o distinguir ciertos datos importantes que pueden ser conocidos hasta después de que el RDBMS ha obtenido los resultados.

**Estudios de Comportamiento:** Consiste en tomar los resultados de reportes previos y utilizar los resultados a través del tiempo.

**Desmenuzar:** Significa dar más detalle, en el esquema estrella, los atributos en las tablas de dimensión juegan un rol crucial, son valores discretos y son la fuente de las restricciones de la aplicación y las columnas de agrupamiento.

Las tareas de validación y prueba toman lugar a través del ciclo de desarrollo, incluyendo prototipos, desarrollo, despliegue, operaciones y mejoras. Esto nunca termina, aunque no es una actividad placentera, es crítica para el desarrollo de un

sistema que trabaja y satisface los requerimientos. Es una tarea tediosa y extensa que requiere de tiempo para preparar las revisiones y analizar los resultados. Pero esta actividad puede dar sorpresas en el presupuesto y, usualmente, detecta la mayoría de los problemas antes de que se hagan grandes.

### **2.2.6. Capacitación**

Ahora que el desarrollo está completo, satisfecho el control de calidad, la documentación está lista y los productos en su sitio, se debe poner todo junto. El entrenamiento del usuario es importante para el éxito del proyecto. Se inicia por entrenar a los usuarios que actualmente utilizarán el sistema. También recordar que el entrenamiento es constante y considerar la entrada de personal al área o movimientos en los departamentos, así como la instrucción a los usuarios en las mejoras hechas al sistema.

Enfocarse en el alcance de los datos en el Almacén de Datos, la herramienta de acceso y como trabaja, la aplicación de Toma de Decisiones o el grupo de reportes iniciales, las capacidades y la forma de tener acceso a los reportes, instrucción continua y asistencia al usuario según se desarrolle el sistema.

### **2.2.7. Operación, Mantenimiento y Soporte**

Esta fase incluye las tareas necesarias para la puesta en marcha del Almacén de Datos para el usuario. Inicia con la instalación de la infraestructura física a los usuarios. Conectividad de Base de Datos, configuración de estaciones de trabajo y herramientas de acceso a los datos, verificación del acceso a la aplicación de toma de decisiones, establecimiento de una estructura de soporte para usuarios, creación de un procedimiento para agregar nuevos reportes y expandir la aplicación de Toma de Decisiones, definición de procedimientos para respaldos de la aplicación, no sólo del Almacén de Datos, creación de un procedimiento para investigar y resolver emisiones relacionadas con integridad de datos, instalación inicial -incluyendo

facilidades para las conexiones básicas de datos con las fuentes y para la actualización y sincronización de datos- la planeación y entrega de una implementación por etapas, capacitación y orientación a todo tipo de usuarios, planeación e implementación de la actualización de plataformas y el mantenimiento necesario por el Almacén de Datos cuando se requiere, administración de usuarios y sistemas, capacidad de generación de archivos permanentes y respaldos, capacidad de recuperación, confirmación de la integración dentro de la infraestructura existente, proporcionar controles de acceso y seguridad, finalmente asegurar la completa disponibilidad así como los procesos para manejar fallas de los sistemas y sus componentes de infraestructura.

Tareas tales como: respaldos, arreglo de *bugs*, actualización de software, mantenimiento y actualización de hardware, mantenimiento de cuentas, *patches* de seguridad y otras tareas administrativas deben ser realizadas continuamente. Operación y mantenimiento de tales servicios requieren un grupo de operación.

Siempre debe tenerse en consideración la aparición de requerimientos. Debido a que los usuarios se vuelven más sofisticados, querrán más capacidades. La eficiencia y rapidez con que se resuelvan éstas definirá que el sistema se mantenga en uso. La habilidad de agregar características se define desde las fases del desarrollo, en que se establece la flexibilidad del diseño.

Algunas áreas de mejoramiento que surgen al comenzarse a utilizar regularmente el Almacén de Datos son:

- Consultas empresariales que no pudieran formularse o satisfacerse debido a las limitaciones impuestas por el metamodelo del Almacén de Datos. Estas limitaciones pueden surgir debido a una falta de ciertos resúmenes o adiciones que no se realizaron en la implementación inicial.
- Consultas empresariales que comprendieran fuentes de datos externos que no formaron parte de la implementación inicial.

- Desempeño no satisfactorio de componentes clave del Almacén de Datos. Algunas de las causas son que el Almacén de Datos no se actualizó con suficiente frecuencia, las herramientas de acceso del usuario final consumieron mucho tiempo para la carga inicial de información desde el Almacén de Datos.

Otros departamentos no querían configurar sus propios mercados de datos. Ahora es necesario incrementar el ámbito del metamodelo del Almacén de Datos.

No son suficientes manuales y entrenamiento para el uso efectivo de un Almacén de Datos. Resulta indispensable la existencia de un grupo de personas con conocimiento y disponibilidad para apoyar a los usuarios, de no ser así el Almacén de Datos podría convertirse en un sistema poco utilizado y no actualizado.

Debido a las diferencias entre un Almacén de Datos y cualquier sistema de información, son requeridas las siguientes capacidades adicionales:

- Comercialización de la información: Consiste en tratar a la información en sí como si fuera un producto de mercancía. Implica también mostrar a los usuarios finales cómo desplegar la información del Almacén de Datos para su provecho, mostrando ejemplos de escenarios de uso.
- Directorio / catálogo de Información: Contiene listas de elementos de información, consultas empresariales, consultas estándar y reportes estándar disponibles a partir del Almacén de Datos, los cuales están en un lenguaje familiar para el usuario, deben ser actualizadas y entregadas con cierta regularidad.
- Examinadores de información: La información del Almacén de Datos también se presenta y distribuye a los usuarios de manera que les permita examinar y hacer copias locales de información en forma selectiva.

### **3. Administración y Mantenimiento de la base de datos**

La importancia de la actividad de la Administración de Datos es la estructura completa (empresa, institución u otro organismo), mientras que el de la Administración de Bases de Datos es particular a la Base de Datos. La Administración de la Base de Datos opera dentro de un marco proporcionado por la Administración de Datos facilitándose de esta manera el desarrollo y uso de una Base de Datos y sus aplicaciones.

En los distintos niveles y aplicaciones de Base de Datos existe la función de Administrador de Base de Datos (DBA), aunque varía en complejidad. En las Bases de Datos Organizacionales, que comúnmente permiten el acceso a centenas de usuarios, se requiere de un DBA de tiempo completo; lo anterior debido al alto volumen de procesos que debe desarrollarse, controlarse y supervisarse.

Un Administrador de Base de Datos debe poseer un conocimiento profundo de las políticas y normas de la empresa así como el criterio de la empresa para aplicarlas en un momento dado. La responsabilidad general del DBA es facilitar el desarrollo y el uso de la Base de Datos dentro de las guías de acción definidas por el Administrador de Datos.

#### **Responsabilidades del DBA**

Las responsabilidades del administrador de base de datos se concentran en los siguientes aspectos:

- Estructura de la Base de Datos .
- Administración de la actividad de los datos .
- Estructura de seguridad e integridad.
- Procedimientos de respaldo y recuperación.
- Administración del Manejador de Base de Datos, RDBMS.
- Administración del diccionario de datos.

- Administración de espacio.
- Documentación.

### **3.1. ESTRUCTURA DE LA BASE DE DATOS**

Inicia con la participación en el diseño de la Base de Datos y su puesta en práctica, así como controlar y administrar los requerimientos de la estructura, ayudando a evaluar alternativas, incluyendo los DBMS a utilizar y ayudando en el diseño general de BD. En un esquema vasto de sistemas, el DBA tiene funciones de supervisión del trabajo de personal de diseño de la BD.

Durante la creación de los datos, el administrador de la base de datos participa en el desarrollo de procedimientos y controles para asegurar la calidad e integridad de la BD.

Los requerimientos de los usuarios van modificándose, estos encuentran nuevas formas o métodos para lograr sus objetivos; la tecnología de la BD se va modificando y los fabricantes del DBMS actualizan sus productos. Todas estas consideraciones requieren la participación activa del DBA.

Las solicitudes de modificación son inevitables, una vez que el sistema ha entrado en operación, pueden aparecer solicitudes de nuevos requerimientos o estos pueden resultar de una comprensión inadecuada de los mismos. En cualquier caso, deberán efectuarse modificaciones en relación con toda la comunidad de la base de datos, ya que el impacto de tales alteraciones será resentido por más de una aplicación. En algunos casos, pueden darse modificaciones que presentan efectos negativos para algunos usuarios; en estos casos deberán tomarse en consideración los beneficios globales que serán obtenidos de tales alteraciones.

En razón del tamaño y complejidad de una base de datos y de sus aplicaciones, las modificaciones pudieran tener resultados inesperados. El DBA debe estar preparado



para restaurar la BD y reunir suficiente información para diagnosticar y corregir el problema provocado por la falla. Después de un cambio la BD es más vulnerable a fallas.

**ESQUEMA CONCEPTUAL:** es tarea del administrador de datos decidir con exactitud cual es la información que debe mantenerse en la base de datos, es decir, identificar las entidades que interesan a la empresa y la información que debe registrarse acerca de esas entidades. Este proceso por lo general se denomina diseño lógico, o conceptual, de bases de datos. Cuando el administrador de datos decide el contenido de la base de datos en un nivel abstracto, el DBA crea a continuación el esquema conceptual correspondiente.

**ESQUEMA INTERNO:** el DBA debe decidir también cómo se representará la información en la base de datos almacenada. A este proceso suele llamársele diseño físico de la base de datos. Una vez hecho esto el DBA deberá crear la definición de estructura de almacenamiento correspondiente (es decir el esquema interno) valiéndose del Lenguaje de Definición de Datos (DDL por sus siglas en inglés Data Definition Language) interno. El esquema original de la base de datos se crea escribiendo un conjunto de definiciones que son traducidas por el compilador de DDL a un conjunto de tablas que son almacenadas permanentemente en el diccionario de datos.

**MODIFICACIÓN DEL ESQUEMA Y DE LA ORGANIZACIÓN FÍSICA:** las modificaciones, tanto al esquema de la base de datos como a la descripción de la organización física de almacenamiento, se logran escribiendo un conjunto de definiciones que son usadas bien por el compilador del DDL o bien por el compilador del lenguaje de almacenamiento y definición de datos para generar modificaciones a las tablas internas apropiadas del sistema.

El esquema conceptual es el resultado de aplicar las reglas que se determinan después del establecimiento de los estándares y políticas del nivel conceptual.

Representa la visión organizacional de la BD que se obtiene al integrar los requerimientos de todos los usuarios en una empresa.

El esquema conceptual es totalmente independiente de las estructuras físicas de almacenamiento y de la representación final de los datos que los usuarios manejan.

El esquema conceptual consta básicamente de:

- a) **DEFINICIÓN DE LOS DATOS.-** Especifica las características como tipo, longitud y precisión de la información que será almacenada en la BD.
- b) **DEFINICIÓN DE LAS RELACIONES ENTRE LOS DATOS.-** Se determina los niveles de iteración que habrán de ocurrir generalmente entre múltiples archivos para obtener información compuesta y procesos complejos.

Adicionalmente a las propiedades de los datos que serán manejados y ocasionalmente en forma complementaria a estos, será necesario definir las especificaciones de:

- a) **ATRIBUTOS.-** Deberá asignarse un identificador que permita manipular en forma individual las características del objeto en cuestión (entidades).
- b) **LLAVES.-** Deberán especificarse los atributos o conjuntos de atributos mediante los cuales puede hacerse referencia a una entidad específica.
- c) **LLAVES HEREDADAS.-** Debe reconocer la factibilidad de hacer referencia a una entidad en particular dentro del conjunto de entidades, mediante una llave primaria; si ésta no se da, debe definirse el conjunto de atributos de la entidad –necesariamente débil– que será utilizado en combinación con la llave primaria de otro u otros conjuntos de entidades –necesariamente fuertes– para lograr la referencia. Este conjunto de atributos será denominado llaves heredadas.
- d) **ESPECIALIZACIÓN Y GENERALIZACIÓN.-** Debe establecerse con claridad el tipo de relación existente entre conjuntos de entidades que fueron particionadas con el objeto de optimizar el espacio de almacenamiento. Un caso de generalización provendrá en la mayoría de los casos de la fusión de tablas llevada a cabo con el objeto de reducir redundancia.

- e) **DEPENDENCIAS DE EXISTENCIAS.-** Debe especificarse con precisión si la existencia de una o más entidades –o conjuntos de entidades- están supeditadas a la existencia de otras.

El establecimiento de conexiones entre las entidades y conjuntos de entidades que conforman una BD, deben especificarse en forma precisa mediante nombre, cardinalidad y opcionalidad.

**NOMBRE DE LAS RELACIONES.-** Generalmente, una etiqueta que indica la función que la relación desempeña; a tal relación se le denomina referencia. En los modelos donde se requiere una mayor precisión en la definición de los componentes, se recomienda indicar las referencias en ambos sentidos.

**CARDINALIDAD DE LAS RELACIONES.-** Debe definirse en forma muy precisa si las entidades de cada conjunto de entidades tendrán interacción con sólo una o varias entidades del conjunto a relacionar.

Debe verificarse que la cardinalidad tenga validez para todos los casos que puedan presentarse en el manejo de la BD; es decir, si son válidas para cualquier ejemplar.

**OPCIONALIDAD DE LAS RELACIONES.-** Permiten definir con mayor claridad aquellos casos en los que una relación puede no establecerse. Las especificaciones de estas situaciones permitirán definir estructuras más precisas, consistentes y de baja redundancia.

Además de la definición de las propiedades de los datos y de las relaciones debe especificarse el formato que guardaran las siguientes estructuras:

**DICCIONARIO DE DATOS:** Los metadatos deberán precisar información que indique con claridad el tipo de datos que serán utilizados, sus ámbitos de influencia y sus limitantes de integridad.

**ÍNDICES:** Son estructuras definidas para un atributo, o conjunto de atributos asociados, que permiten simular una secuencia lógica para las entidades. La principal cualidad de un índice reside en la capacidad para acelerar el acceso a un dato específico.

**FORMATOS DE CAPTURA Y PRESENTACIÓN:** Las aplicaciones deberán proveer interfaces amigables y eficientes entre el usuario y la BD. Para esto se definirán, formatos y pantallas de captura, de consulta y de reporte. La información resultante será procesada y direccionada cada vez que se active la captura o la consulta, el formato de tal captura o consulta podrá almacenarse para su reutilización.

El modelo será el paso siguiente a la definición de los elementos que componen a los datos y que rigen a las relaciones que se dan entre estos.

Aunque existen múltiples alternativas para el desarrollo de métodos eficientes, el uso del modelo entidad relación (E-R) se ha convertido casi en un estándar para manejadores secuenciales.

La implementación del modelo E-R dará como resultado un modelo relacional; esto al convertir los elementos del diagrama a las tablas correspondientes.

Las tablas resultantes podrán ser o no relacionales según la pericia del diseñador, puede resultar redundancia – en consecuencia riesgo de inconsistencia – y otros efectos indeseables.

La normalización es el conjunto de normas que nos ayudan a diseñar una estructura de Bases de Datos óptima para su implementación, gestión y explotación desde distintas aplicaciones, consiguiendo independencia de las mismas. El creador de estas normas fue E.F. Codd.

La etapa de normalización debe refinar los detalles del modelo resultante, de tal forma que la estructura de las tablas proporcione un bajo nivel de redundancia, minimice hasta donde sea posible la inconsistencia y sea capaz de proporcionar acceso eficiente a los datos.

Ocasionalmente, será preferible llevar la normalización hasta un nivel no óptimo si se obtiene a cambio eficiencia en el acceso a los datos u otros beneficios que eleven en forma significativa el desempeño del sistema.

Un control de integridad o restricciones es aquel que nos permite definir con precisión el rango de valores validos para un elemento y / o las operaciones que serán consideraciones validas en la relación de tales elementos.

El objetivo primordial de un control de integridad es la reducción de la inconsistencia en la BD.

Las restricciones de integridad normalmente se aplican en tres niveles:

1. **UN ATRIBUTO SIMPLE.-** Se define un dominio del atributo que es totalmente independiente del resto del entorno de la Base de Datos.
2. **UN ATRIBUTO DEPENDIENTE DE OTRO.-** Se definen subconjuntos de dominios posibles para un atributo X según el valor que previamente a sido asignado al atributo W.
3. **RELACIONES ENTRE TUPLAS DE UNA O VARIAS TABLAS.-** Se especifican valores posibles para registros completos según los valores acumulados registros previos o por valores existentes en registros de otras tablas.

La implementación de la cardinalidad resultante en el modelo será una de las restricciones importantes que el sistema debe considerar.

Un diseño adecuado del nivel físico es aquel que permitirá que el registro y consulta de la información se lleve a cabo de forma eficiente.

Una especificación inadecuada de las características generales del sistema y de los procesos con alta prioridad de ejecución puede traer como consecuencia que la definición del nivel físico no se oriente hacia el aprovechamiento eficiente del medio físico de almacenamiento y los accesos que se realizan a éste.

Entre las principales actividades que deben realizarse para un buen diseño del esquema físico se encuentra la determinación de las longitudes de registro de datos que se ajusten de la manera más adecuada posible al tamaño del *buffer* que se esté utilizando.

Con lo anterior, podrán establecer procesos de alta prioridad procurando que los registros que se requieran para completar transacciones de este tipo puedan ser ubicados con la menor cantidad posible de acceso a disco.

La búsqueda de niveles eficaces en el manejo del medio de almacenamiento obedece a la gran cantidad de tiempo que es consumido por las operaciones de entrada-salida y el flujo de datos que se da entre la memoria principal y el medio de almacenamiento.

### **3.2. ADMINISTRACIÓN DE LA ACTIVIDAD DE DATOS**

El administrador de la base de datos no administra valores de datos; sino que administra actividad de datos. Dado que la BD es un recurso compartido, debe además proporcionar estándares, guías de acción, procedimientos de control y la documentación necesaria para garantizar que los usuarios trabajan en forma cooperativa y complementaria al procesar datos en la BD.

Existe gran actividad al interior de un manejador de base de datos. La concurrencia de múltiples usuarios requiere de estandarizar los procesos de operación; el DBA es responsable de tales especificaciones y de asegurarse que éstas lleguen a quienes concierne. Todo el ámbito de la BD se rige por estándares, desde la forma como se capture la información (tipo, longitud, formato), hasta como es procesada y presentada. El nivel de estandarización alcanza hasta los aspectos más internos de la BD; como se accesa a un archivo, como se determinan los índices primarios y auxiliares, la numeración de los registros, entre otros.

Debe procurarse siempre que los estándares que serán aplicados beneficien también a los usuarios, privilegiando siempre la optimización en la operación del DBMS y el apego de las políticas de la empresa.

Una administración efectiva de base de datos deberá disponer siempre de este tipo de estándares; entre las funciones del DBA se encuentra la de revisarlos periódicamente para determinar su operatividad, y en su caso ajustarlos, ampliarlos o cancelarlos. Es también su responsabilidad el que estos se cumplan.

Cuando se definen estándares sobre la estructura de la BD, estos deben registrarse en una sección del diccionario de datos a la que todos aquellos usuarios relacionados con ese tipo de proceso pueden acceder.

Otro de los aspectos que el administrador debe atender es la coordinación de las nuevas propuestas para realizar ajustes en los derechos de acceso a datos compartidos y aplicaciones específicamente propuestas serían analizadas en conjunto con los supervisores o directivos de las áreas involucradas para determinar si procede, pudieran aparecer problemas cuando dos o más grupos de usuarios quedan autorizados para notificar los mismos datos.

Es responsabilidad del DBA la publicación y mantenimiento de la documentación en relación con la actividad de los datos, incluyendo los estándares de la BD, los

derechos de recuperación y de acceso a la BD, los estándares para la recuperación en contingencias y el cumplimiento de las políticas establecidas. Los productos DBMS más populares que se encuentran en el mercado proporcionan servicios de utilerías para ayudar al DBA en la administración de los datos y su actividad. Algunos sistemas registran en forma automática los nombres de los usuarios y de las aplicaciones a las que tienen acceso así como a otros objetos de la BD. Incorpora también utilerías que permitan definir en el diccionario de datos las restricciones para que determinadas aplicaciones o módulos de ellas sólo tengan acceso a segmentos específicos de la BD.

Es responsabilidad del DBA organizar el sistema de modo que se obtenga el desempeño que sea “mejor para la empresa”, es decir el equilibrio entre tiempo de respuesta y recursos invertidos, así como realizar los ajustes apropiados cuando cambien los requerimientos.

Cuando es indispensable asegurar el acceso en tiempo real a los datos, los usuarios del sistema se interesan al extremo en el tiempo de respuesta del sistema. En los sistemas de procesamiento por lotes, el interés se traslada al caudal de transacciones que el sistema admite al tiempo que se tarda en despachar la carga de trabajo. Estos factores dependen del tiempo necesario para acceder a los datos necesarios y dependen también de la organización de éstos y su localización en las unidades de almacenamiento. La diferencia entre una organización adecuada y una organización inadecuada se reflejan como una gran diferencia en los tiempos de respuesta y en los caudales de transacciones.

Cuando el almacenamiento se prevé para un conjunto específico y bien entendido de operaciones, es posible optimizar, para esas operaciones, la organización del almacén y la localización de los datos en él. El diseñador sabe perfectamente que es lo que tiene entre manos. No siempre el proyectista es tan afortunado. En muchos casos ni siquiera sabe como se van a utilizar en realidad los archivos, cómo se consultará la base o con qué frecuencia. Resulta, por lo tanto, necesario ajustar, y hasta cambiar fundamentalmente, la organización del almacén después que el sistema ha entrado en



servicio y se han aclarado suficientemente las pautas del uso. En muchos casos, el uso de la base de datos evoluciona continuamente, a medida que más personas se van familiarizando con ella y se crean más programas de aplicación. El ajuste de la organización del almacén con el objeto de mejorar su desempeño convirtiéndose así en un proceso continuo.

Este proceso de ajuste de la base de datos se llama afinación, *tunning*. En la práctica, la afinación ha conducido a menudo a importantes economías. A veces éstas han sido tan importantes como para marcar la diferencia entre lo que es una aplicación rentable y lo que no lo es. El responsable de la afinación de la base de datos es el DBA o su grupo, y es importante que éste tenga libertad para introducir los cambios que estime necesarios, sin hacer estragos en los programas de aplicación. Sin un software apropiado, la afinación suele incurrir en costos inadmisibles en lo que se refiere al mantenimiento y la prueba de programas.

La correcta afinación tiene dos requisitos:

- a) Necesita la independencia física de los datos.
- b) Requiere medios para supervisar automáticamente el uso de la base de datos con el fin de que puedan hacerse los ajustes necesarios.

Puede necesitarse la reorganización de los archivos cuando, debido la creación de diversas cadenas, los tiempos de recuperación o de procesamiento serial se vuelven excesivos. Tal reorganización consiste en leer el archivo en forma en que se utilizaría al realizar el procesamiento en serie y escribirlo de nuevo, dejando fuera todos los registros que estén marcados como eliminados y escribiendo todos los registros restantes, nuevos y viejos, secuencialmente, en las áreas principales del archivo nuevo. Durante este procesamiento los programas de reorganización crearán nuevos índices con base en los nuevos valores.

La frecuencia de esta reorganización depende de la actividad de inserción dentro del archivo. En la práctica se encuentran intervalos de tiempo que varían de un día a un

año entre corridas de reorganización. Ya que una corrida de reorganización puede requerir mucho tiempo, generalmente se realiza antes que el archivo esté realmente lleno, para evitar problemas en las épocas de mucha actividad.

El administrador de la base de datos debe encargarse de la comunicación con los usuarios, garantizar la disponibilidad de los datos que requieren y escribir, o ayudar a los usuarios a escribir, los esquemas externos necesarios. Otros aspectos de la función de enlace con los usuarios incluyen las consultas sobre diseño de aplicaciones, la solicitud de instrucción técnica, la ayuda en la localización y resolución de problemas, y otros servicios profesionales similares relacionados con el sistema.

### **3.3. ESTRUCTURA DE SEGURIDAD E INTEGRIDAD**

Las verificaciones de seguridad y de integridad se definen durante el diseño, y al igual que la creación del esquema se integran dentro de la base a través de sentencias DDL (por sus siglas en inglés Data Definition Language). La concesión de diferentes tipos de autorización permite al administrador de la base de datos regular qué partes de la base de datos van a poder ser accedidas por los diferentes tipos de usuarios. Las restricciones de integridad se mantienen en una estructura especial del sistema que consulta el DBMS cada vez que tiene lugar una actualización en el sistema.

Existen múltiples riesgos para la seguridad de la información durante la operación, implantación y tiempos muertos en el sistema.

Cuando se está instalando o actualizando un sistema, los principales factores de riesgo son aquellos relacionados con el ajuste de formatos, dominios y otros parámetros que pueden verse afectados por la conversión del sistema.

Cuando el sistema que se implanta ha de recibir nueva información, es importante el establecimiento de códigos que permitan validar la captura para minimizar los riesgos de información no confiable.

Mientras el sistema se encuentra en uso, se dice que las operaciones se realizan en línea; es decir, la información se afecta por medio de los procedimientos definidos en el sistema.

La protección más común para reducir estos riesgos consiste en el establecimiento de claves de acceso (*Password*) tanto para acceder a la aplicación como a las diversas operaciones que ésta desempeña.

Algunos sistemas que manejan claves fijas pueden incluir controles sobre el usuario que lo obliguen a modificar su clave de acceso con cierta regularidad. Es importante que el código que mantiene la tabla de claves de usuarios en el sistema se encuentre codificada o encriptada.

Cuando el sistema no se encuentra en operación la información está expuesta a ser alterada fuera de línea; es decir, sin utilizar los programas de aplicación diseñados para este fin.

Algunas de las técnicas más utilizadas para evitar, y en algunos casos sólo para ejecutar modificaciones fuera de línea, son:

**ENCRIPAMIENTO.-** Consiste en convertir la información de la BD a un formato que resulte ilegible si no se dispone del algoritmo de conversión.

**APLICACIÓN DE TOTALES DE CONTROL.-** Consiste en generar registros ficticios que son agregados a la BD y que permitirán detectar la inserción, eliminación o modificación de datos en la gran mayoría de los casos. Los registros ficticios son creados con información que se obtiene de acumulados o valores estadísticos de los registros reales.

**DÍGITOS DE CONTROL.**- Son caracteres que se adicionan a las claves o a los datos que serán manejados con el objeto de autenticar su validez.

Su aplicación se extiende a procesos en línea y protección fuera de línea.

Precauciones adicionales a las anteriores, deben ser tomadas en cuenta para elevar el nivel de seguridad en redes de usuarios. Las más comunes son:

- Validar contraseñas no repetidas.
- Eliminar claves de acceso de usuarios deshabilitados.
- Establecer políticas y sanciones por desatender estaciones desconectadas, con acceso.
- Restringir procesos de alto riesgo a terminales con mayor nivel de seguridad.
- Establecer controles para el acceso validado a las terminales, es decir, implementar sistemas electrónicos de autenticación de terminal.
- Establecer políticas para denegar el acceso después de una cantidad determinada de intentos fallidos de un tiempo transcurrido.

Debe considerarse la posibilidad de controles alternos cuando el sistema maneja información o recursos altamente privilegiados para la organización. Las formas más comunes para autenticar la identidad del usuario son:

- a. Algo que el usuario conoce.- Contraseña, algoritmos de acceso.
- b. Algo que el usuario tiene.- Tarjetas de acceso, bandas magnéticas etcétera.
- c. Identificación de aspectos físicos del usuario.- Huellas digitales, examen de la retina, voz etcétera.

**ESTADÍSTICAS DE UTILIZACIÓN DE DISPOSITIVOS.**- Tiempos de porcentaje de actividad de procesadores, canales, controladores y discos. Es conveniente tener tanto valores promedios como valores que pertenezcan a los periodos de mayor actividad de operación del sistema.

**ESTADÍSTICAS DE ARCHIVO.-** Una matriz de procesos del usuario, comparada con actividades de archivo, es la medida básica. Las razones relativas de tipos de acceso, tales como consulta, obtención del siguiente registro y actualización, son importantes en la toma de decisiones acerca de la organización del archivo, otra medida es la densidad de archivo.

**ESTADÍSTICA DE UTILIZACIÓN DE REGISTROS.-** La frecuencia con la que se efectúan accesos a los registros para lectura o actualización proporciona una medida que puede emplearse para lograr optimización en sistemas en los que los registros están encadenados. Las fechas y los momentos del acceso así como las últimas actualizaciones son importantes para mantener la integridad, como lo es la identificación del proceso que actualizó los registros.

**ESTADÍSTICAS DE UTILIZACIÓN DE ATRIBUTOS.-** La frecuencia con la que los valores de los atributos se solicitan, actualización, utilizan como llaves o se emplean como argumentos posteriores de búsqueda, proporciona los datos para la selección óptima de acceso a datos.

### ***3.4. PROCEDIMIENTOS DE RESPALDO Y RECUPERACIÓN***

Cuando se decide utilizar un sistema de base de datos, la empresa se vuelve dependiente en grado sumo del funcionamiento correcto de ese sistema. En caso de alguna contingencia que afecte cierta porción de la base de datos, por causa de error humano, falla en el equipo o en el sistema operativo, resulta esencial poder reparar los datos implicados con un mínimo de retraso y afectando lo menos posible el resto del sistema. En teoría, la disponibilidad de los datos no dañados no debería verse afectada. El DBA debe definir y poner en práctica un plan de recuperación adecuado que incluya, una descarga o “vaciado” periódico de la base de datos en un medio de almacenamiento de respaldo, y procedimientos para cargar otra vez la base de datos a partir del vaciado más reciente cuando sea necesario.

Las técnicas de recuperación son otra función esencial del administrador de base de datos al administrar la actividad de datos. A pesar de que el DBMS lleva a cabo una parte del proceso de recuperación, los usuarios determinan en forma crítica la operatividad de esos sistemas de protección. El DBA debe anticipar fallas, y definir procedimientos estándares de operación; los usuarios deben saber que hacer cuando el sistema esté caído y que es lo primero que debe realizarse cuando el sistema esté puesto en marcha nuevamente. El personal de operación deberá saber cómo iniciar el proceso de recuperación de la BD, que copias de seguridad utilizar; cómo programar la reejecución del tiempo perdido y de las tareas pendientes. Es importante también establecer un calendario para llevar a cabo estas actividades sin afectar a otros sistemas dentro de la organización que hagan uso de los mismos recursos de cómputo. Destacan por su importancia en el proceso de recuperación y a su vez en la atención que prestan a otros sectores de la organización. Los dispositivos de comunicación remota, los sistemas de interconexión y otros accesorios de uso compartido.

En la práctica, el empleo de volver a leer la entrada de una transacción o la restauración mediante imágenes posteriores dependen de la disponibilidad de una copia de respaldo de una versión anterior de la base de datos. Si las versiones se utilizan como copias de respaldos pueden construirse volviendo a posicionar la señal actual de nivel máximo de otra manera, los respaldos se crean por copiado.

Es posible generar periódicamente copias de respaldo y conservar una serie de versiones anteriores. Cada copia de respaldo estará identificada por tiempo y fecha y por la última transacción incluida. Anteriormente, una copia de respaldo debía generarse mientras la base de datos estaba en reposo, ya que las actualizaciones durante el copiado podían provocar que la copia fuese inconsistente. Para bases de datos muy grandes puede no presentarse nunca un periodo de descanso lo suficientemente largo para realizar una copia de respaldo. Por lo anterior, en la

actualidad algunos manejadores cuentan con opciones de respaldo en línea más sofisticados.

Pasos sugeridos cuando se presente una falla en el sistema y se desee recuperar la información:

1. **Detección del error.**- el proceso de recuperación se inicia al detectar la existencia de un error. Es posible distinguir una variedad de puntos de entrada en el proceso de recuperación. Se considerarán fallas de sistemas detectadas por falta de acción del sistema o por verificaciones irrecuperables de redundancia y salida incorrecta observada por un usuario.

2. **Determinación de la fuente del error.**- para decidir cuál es la mejor acción correctora es necesario determinar la extensión del daño. Desde luego este esfuerzo es muy relacionado con la determinación del tiempo y la causa del error. Después de una caída o cuando el procesamiento sea interrumpido debido a una señal de error, es necesario determinar tanto aquellas áreas del archivo de datos que sean sospechosas como cuál fue la transacción que no se concluyó.

3. **Ubicación de errores secundarios.**- cuando se ha detectado un error que provocó una modificación inadecuada a un archivo un rastreo a través de las listas de actividad encontrará aquellas transacciones que emplearon el bloque correcto. Entonces es posible volver a introducir automáticamente el bloque correcto de las transacciones afectadas y reproducir resultados correctos. Si se actualizaron bloques mediante transacciones que leyeron bloques incorrectos antes de existir es necesario restaurar aun más el archivo.

4. **Aplicación de correcciones.** Si la extensión del daño es limitada, puede utilizarse un proceso de volver a enrollar. Las porciones dañadas del archivo se restauran aplicando primero aquellas imágenes anteriores a los bloques en error reemplazando después de las transacciones incompletas. La salida proveniente de estas transacciones se suprime de ser posible, para evitar duplicar resultados que previamente se hayan enviado a los usuarios.

### **3.5. ADMINISTRACIÓN DEL RDBMS**

Es responsabilidad del administrador de base de datos compilar y analizar estadísticas relativas al rendimiento del sistema e identificar áreas potenciales del problema. Si se requieren cambios el DBA deberá planearlos y ponerlos en práctica.

El DBA debe vigilar periódica y continuamente las actividades de los usuarios en la Base de Datos. Los productos manejadores de bases de datos incluyen tecnologías que registran estadísticas. Estos informes pudieran indicar cuales fueron los usuarios activos, que archivos y que elementos de datos han sido utilizados, e incluso el método de acceso que se ha aplicado. Pueden capturarse y reportarse las tasas de error y los tipos de errores. El DBA analizará estos datos para determinar si se necesita una modificación en el diseño de la BD para manejar su rendimiento o para facilitar las tareas de los usuarios.

El DBA deberá analizar las estadísticas de tiempo de ejecución sobre la actividad de la Base de Datos y su rendimiento. Cuando se identifique un problema de rendimiento, el DBA deberá determinar si resulta apropiada una modificación a la estructura de la Base de Datos o al sistema. Casos como la adición de nuevas llaves o su eliminación, nuevas relaciones entre los datos y otras situaciones típicas deberán ser analizadas para determinar el tipo de modificación precedente.

Cuando el fabricante del DBMS en uso anuncie una nueva versión del producto, debe realizarse un análisis de las características que ésta incorpora y sopesarlas contra las necesidades de la comunidad de usuarios. Si se decide la adquisición del producto, los usuarios deben ser notificados y capacitados en su uso. El DBA deberá administrar y controlar la migración tanto de las estructuras, como de los datos y las aplicaciones.

El software de soporte y otras características de hardware pueden implicar también modificaciones de las que el DBA es responsable ocasionalmente, estas modificaciones traen como consecuencia cambios en la configuración o en algunos parámetros de operación del DBMS.



En un principio se configura el DBMS con valores aproximados respecto a las dimensiones de los datos y los recursos poseídos, durante esta etapa se realiza un análisis de la actividad en la base y se toma en consideración realizar modificaciones. Aun en el caso en que el rendimiento parece aceptable, el DBA puede considerar modificar algunas opciones y observar su efecto sobre el sistema, siempre en búsqueda de la optimización o afinación del mismo.

Una gran cantidad de casos de inconsistencia se generan a partir de procesos que no fueron completados por el sistema; es decir, se actualizaron solamente algunos valores de todos los que participaron en una operación determinada.

Para solucionar este problema, las operaciones individuales se agrupan formando transacciones. Una transacción atómica es aquella en la que se registran las actualizaciones que derivan de todas las operaciones individuales que la conforman o de ninguna de ellas.

El control de las transacciones se realiza mediante procedimientos de bitácora. Ésta manejará las operaciones individuales hasta este momento cuando se realizará la actualización en la BD.

### **CONTROL DE CONCURRENCIA**

Cuando el sistema da soporte a múltiples usuarios en forma simultánea, la inconsistencia puede presentarse por factores adicionales:

**ACTUALIZACIÓN PERDIDA.-** Ocurre cuando la actualización de un usuario queda sobrescrita por la de otro cuando ambos accesan al sistema por un recurso común.

Un método comúnmente utilizado para evitar la actualización perdida es el bloqueo de registros o archivos. Este consiste en activar restricciones de acceso para los recursos que están siendo utilizados por un proceso en un momento determinado.

Con el objeto de no retardar la ejecución de determinados procesos que pretendan alcanzar recursos específicos, los bloqueos normalmente se establecen para los recursos en forma individual.

Esto puede traer como consecuencia interbloqueo (*dead lock*) que se presenta cuando dos procesos que compiten por recursos comunes inician su ejecución tomando alguno de ellos y bloqueándolo en este momento.

El proceso requiere para finalizar su ejecución del recurso que está bloqueado por el otro proceso presentándose de esta manera un bloqueo mutuo en el que ninguno de los procesos puede finalizar.

Una de las alternativas más comunes para minimizar el interbloqueo consiste en asignar una secuencia específica de acceso a los recursos para aquellos procesos que los requieren en forma común.

Lo anterior permitirá que otros procesos puedan participar puesto que el primer proceso no bloqueó a los recursos en forma global. Aunque la técnica anterior es eficiente en muchos casos, la mejor estrategia para tratar el interbloqueo es la conocida como bloqueo en dos fases.

Normalmente las políticas de los sistemas incluyen técnicas para detectar el interbloqueo y abortar a uno o todos los procesos involucrados para dar agilidad al sistema las formas más comunes de detecciones son:

- Control de longitud de la cola de procesos.
- Verificación de ciclos en las conexiones del proceso que tiene un recurso.

Algunos datos se usan con mucha frecuencia y otros sólo raramente. Es deseable almacenar los datos de uso frecuente de manera que resulte fácil y rápido acceder a ellos. Los datos de uso ocasional se almacenarán, en cambio, de manera más

económica. En una oficina, la información que se usa diariamente se usa en los archivadores de las secretarías; la información que sólo se consulta accidentalmente se guardará probablemente en el sótano, de modo no estorbe y su almacenamiento no cueste mucho. El equivalente del sótano en la computadora podría ser la cinta magnética, mientras que los datos de uso frecuente se hallarán en discos o tambores de modo que se los pueda leer siempre en una fracción de segundo. Toda base de datos compleja tendrá múltiples niveles de facilidad de acceso.

A medida que cambia la popularidad de un conjunto de datos, será conveniente mudarlos dentro del almacén a posiciones accesibles, de acuerdo con su actividad. En algunos casos no se mudan los datos, pero sí se modifican los índices que se utilizan para direccionarlos, de manera que puedan ser localizados rápidamente. Este proceso de ajustar el almacenamiento de los datos de acuerdo con su popularidad se llama migración de los datos. En algunos sistemas se hace automáticamente. En otros, la operación se lleva a cargo por los programadores de sistema o el administrador de datos. En ocasiones se le considera como parte del proceso de afinación de la base de datos.

### **3.6. ADMINISTRACIÓN DEL DICCIONARIO DE DATOS**

Un Diccionario de Datos es un catálogo, un depósito, de los elementos en un sistema. Como su nombre lo sugiere, estos elementos se centran alrededor de los datos y la forma en que están estructurados para satisfacer los requerimientos de los usuarios y las necesidades de la organización. En un diccionario de datos se encuentra la lista de todos los elementos que forman parte del flujo de datos en todo el sistema. Los elementos más importantes son flujos de datos, almacenes de datos y procesos. El diccionario guarda los detalles y descripciones de todos estos elementos.

Si los analistas desean conocer cuántos caracteres abarca un determinado dato o qué otros nombres recibe en distintas partes del sistema, o dónde se utiliza, encontrarán las respuestas en un diccionario de datos desarrollado en forma apropiada.

El diccionario se desarrolla durante el análisis de flujo de datos y auxilia a los analistas que participan en la determinación de los requerimientos de sistemas.

Los analistas utilizan los diccionarios de datos por cinco razones importantes:

1. Para manejar los detalles en sistemas grandes.
2. Para comunicar un significado común para todos los elementos del sistema. †
3. Para documentar las características del sistema.
4. Para facilitar el análisis de los detalles con la finalidad de evaluar las características y determinar dónde efectuar cambios en el sistema.
5. Para localizar errores y omisiones en el sistema.

**Manejo de detalles.-** Los sistemas grandes tienen enormes volúmenes de datos que fluyen por ellos en forma de documentos, reportes e incluso pláticas. De manera similar, se llevan a cabo muchas actividades que utilizan los datos existentes o que generan nuevos detalles.

**Comunicación de significados.-** Los diccionarios de datos proporcionan asistencia para asegurar significados comunes para los elementos y actividades del sistema.

**Documentación de las características del sistema.-** Documentar las características de un sistema es la tercera razón para utilizar los sistemas de diccionario de datos. Las características incluyen partes o componentes así como los aspectos que los distinguen.

**Facilidades de análisis.-** La cuarta razón para hacer uso de los diccionarios de datos es determinar si son necesarias nuevas características o si están en orden los cambios

de cualquier tipo. Los analistas de sistemas abordan las siguientes características del sistema:

- **Naturaleza de las transacciones:** Las actividades de la empresa que se llevan a cabo mientras se emplea el sistema, incluidos los datos necesarios para aceptar, autenticar y procesar cada actividad.
- **Salida y generación de reportes:** Resultados del procesamiento hecho por el sistema que son presentados a los usuarios en una forma aceptable para ellos.
- **Archivo y bases de datos:** Detalles de las transacciones y registros maestros que son de interés para la organización.
- **Capacidad del sistema:** Habilidad del sistema para aceptar, procesar y almacenar transacciones y datos.

**Localización de errores y omisiones.-** Tener información en un diccionario relacionada con las características del sistema - transacciones, consultas, datos y capacidad - dice mucho con respecto al sistema y permite evaluarlo. Pero para esto es necesario saber que la propia información es completa y exacta consiguiente, los diccionarios se emplean por una quinta razón: localizar errores en la descripción del sistema.

### **3.7. ADMINISTRACIÓN DE ESPACIO**

La medida referente al tamaño de la Base de Datos depende del hardware y de las restricciones operativas aplicables en un ambiente dado.

El término grande implica una cantidad de datos mayor que la que una sola persona puede manejar por sí misma, aunque la auxilie un sistema de computación. La cantidad real variará dependiendo de la complejidad de los datos y de sus aplicaciones.

Una base de datos muy grande es parte esencial de una empresa y será utilizada continuamente por muchos individuos. Al mismo tiempo requerirá muchos dispositivos

de almacenamiento. Una base de datos muy grande presenta problemas de manejo, ya que no puede detenerse sin afectar la operación y el bienestar de la empresa usuaria.

Se denomina fragmentación a la asignación de archivos en sectores no contiguos de un disco flexible o duro. La fragmentación ocurre debido a numerosas operaciones de eliminación y escritura de archivos.

Después de crear y borrar muchos archivos en un disco, los archivos restantes no se guardan en unidades de asignación contiguas. Así, cuando más tarde guarda un archivo o instala un nuevo programa, se guardan los datos en las unidades de asignación contiguas entre los archivos existentes. La cabeza de lectura / escritura de la unidad de discos debe recorrer grandes distancias para recuperar la información dispersa. Un proceso conocido como defragmentación, disponible en algunos manejadores, puede mejorar la eficiencia del disco hasta en un 50 por ciento, para lo cual rescribe los archivos para que queden colocados en unidades contiguas.

Defragmentación es el procedimiento en el que todos los archivos de un disco duro se rescriben en el mismo para que todas las partes de cada archivo se escriban en sectores contiguos. El resultado es un mejoramiento en la velocidad del disco en operaciones de consulta. Durante una operación normal, los archivos del disco duro quedan fragmentados de tal forma que las partes de uno se inscriben en distintos lugares del disco, lo que provoca que las operaciones de acceso a los datos sean lentas.

Llevar un control del crecimiento de la Base de Datos es una de las tareas principales del DBA en su objetivo de optimizar los recursos, específicamente el espacio físico. Tener un conocimiento lo más exacto posible de los incrementos de espacio que registrará la Base de Datos, permite conocer la situación de almacenamiento con propósitos de afinación de accesos, evitar errores durante la ejecución de procesos

exhaustivos, proyectar adquisición de disco y mantener más completa la documentación de la Base de Datos.

### **3.8. DOCUMENTACIÓN**

La tarea de documentar en la administración de la Base de Datos alcanza las tareas restantes del administrador, es decir, que el administrador debe documentar todas sus actividades. Es de suma importancia registrar que modificaciones han sido efectuadas, como fueron realizadas y cuando fueron establecidas. Una modificación sobre la estructura de la BD pudiera ocasionar un error que no apareciera a corto plazo; una vez que este surja, sin la documentación adecuada sobre las modificaciones realizadas, el diagnóstico resultaría extremadamente complicado. En estos casos, se haría necesario una secuencia de reejecuciones para intentar detectar el punto en conflicto; el riesgo de este procedimiento radica en que es posible afectar la información contenida en la BD. Para identificar un cambio es de suma importancia mantener un registro de los formatos de prueba y de las ejecuciones de las pruebas efectuadas. Si se utilizan procedimientos de prueba formatos de pruebas y métodos de registro estandarizados, el registro de los resultados de la prueba no consumirá tiempo excesivo.

La tarea de la documentación es cada vez más ligera y precisa cuando se utilizan DBMS que integran herramientas CASE para las tareas de diseño, mantenimiento y documentación. Estas mismas herramientas CASE proporcionan, en la mayoría de los casos la facilidad de generar y mantener en forma automática el Diccionario de Datos.

Una razón más para documentar consiste en la necesidad de mantener organizados datos históricos. Ocurre comúnmente que se desea realizar una consulta sobre los respaldos para conocer el estado que guardaba la información en un periodo determinado que transcurrió previamente. Los registros de modificación existentes en la documentación permitirán resolver problemas de incompatibilidad entre las estructuras que eran vigentes en el periodo de respaldo y las que lo son ahora;

permitirá también el desarrollo de módulos de ajuste que faciliten la traducción de formatos o escalas para valores almacenados.

En los casos de caídas del sistema se presenta una situación parecida; los respaldos son requeridos y habrá de verificarse su estructura; formato y escala para integrarlos a la operación del sistema.



## **4. Uso de la metodología de Almacén de Datos en la Administración de una Base de Datos**

Debido a la diversidad de responsabilidades de un DBA, es preciso contar con múltiples herramientas para apoyar sus actividades. Actualmente existen aplicaciones cuyo objetivo es satisfacer estas necesidades y donde se fusionan opciones de monitoreo y manipulación de los elementos de la Base de Datos, entre los que se incluyen productos para obtener estadísticas de utilización de recursos, distribución de espacio, monitoreo de bloqueo y esquemas de seguridad. Estas aplicaciones han representado un respaldo importante para el DBA, con el inconveniente de un costo elevado.

Para evitar un gasto excesivo en la adquisición de herramientas que apoyen al DBA, las empresas cuentan con la experiencia de las personas involucradas con el área y sus conocimientos en lenguaje estructurado de acceso a Base de Datos (SQL), es posible tomar ventaja de este recurso con el fin de desarrollar herramientas para la empresa que cubran las necesidades de la Administración de la Base de Datos. Además de ser una opción más económica, el diseño de la herramienta toma en consideración, no sólo las particularidades de las bases a administrar, incluye también los estándares de la empresa y sobre todo las necesidades específicas de la institución.

En la institución durante la implementación de nuevos sistemas, y el crecimiento de los actuales, se presentaron dos problemas concernientes a la administración de la base de datos, la necesidad de espacio físico en disco, y la insuficiencia de licencias adquiridas para acceso a la base de datos. Para resolver satisfactoriamente estos problemas se tenía la necesidad de conocer estadísticas de crecimiento en las bases de datos, con el fin de proyectar adquisición de discos o de solicitar la depuración de los sistemas y se precisaba evaluar si las licencias asignadas estaban siendo realmente utilizadas.

En lo que respecta al espacio, no se contaba con un histórico de crecimiento, un sitio donde registrar las variaciones de espacio que presentaba la base de datos, y que permitiera conocer estadísticas de requerimiento de espacio físico, como el incremento del espacio mensual de un sistema, identificación de tablas con mayor aumento y con qué regularidad se presentaban cambios en este aspecto.

Las licencias regulan el número de accesos concurrentes a la base de datos, se adquieren del proveedor de software e implican un costo. Como parte de la documentación de un sistema el área que hará uso de él debe definir el número de usuarios concurrentes dentro del sistema, con el fin de que le sea asignada dicha cantidad de licencias. El costo de este recurso es elevado y era necesario validar que las licencias asignadas estuvieran siendo realmente utilizadas.

Ante estas necesidades se desarrollaron dos Mercados de Datos (Datamart) para cubrir las necesidades del área mencionadas, uno de espacio y otro de sesiones.

La arquitectura que se utilizó fue el Mercado de Datos, la razón por la que se tomó esta decisión fue por que la información que se almacena es de utilidad únicamente para algunas áreas, el alcance de cada esquema es específico del objetivo para el que fue desarrollado, es decir, la información contenida no es de mayor relevancia para otras direcciones de la institución.

Debido a lo apremiante de la situación en la planeación de los mercados de datos, se eligió como estrategia de implementación, el enfoque de abajo hacia arriba, con el desarrollo a través de un prototipo, el cual permitía obtener resultados de forma más acelerada, no había que esperar a que se definieran los requerimientos y era posible dar inicio a la implementación de los mercados para su explotación. Por lo anterior, la metodología de desarrollo que se utilizó fue de espiral, la cual permite ampliar o modificar los requerimientos del sistema mientras se va haciendo uso de él, propiciando así la afinación, complementación y mejoramiento del sistema durante el mantenimiento.

En la planeación se definió que las áreas que tendrían acceso al Mercado de Datos serían las de administración de la Base de Datos y la de producción. Los recursos que se utilizaron son internos, se utilizaron las herramientas con las que cuenta la institución: manejador de Base de Datos Oracle® sobre sistema unix® para el almacenamiento, PL/SQL para la extracción y carga, cronitab del sistema operativo unix para programar las cargas, Oracle Discoverer® para la explotación de la información. Se determinó que estas herramientas satisfacen las necesidades de almacenamiento, carga, transformación y explotación, debido a que son herramientas institucionales que permiten acceso a las bases de datos transaccionales con características suficientes para estructurar un almacén de datos.

Los elementos que se obtienen de los mercados de datos permiten al área de administración de base de datos, conocer mejor el comportamiento de las bases respecto a los tópicos que conciernen a los mercados, espacio y sesiones, aportando fundamentos para la toma de decisiones.

Adicionalmente los Mercados de Datos facilitan, la tarea de documentación del DBA, ya que en ellos se registran algunas de las tareas administrativas que realiza.

#### ***4.1. Mercado de Datos de Administración de Espacio.***

Un recurso indispensable en todos los sistemas es el espacio físico requerido para el almacenamiento de la información. Los manejadores de base de datos asignan y distribuyen el espacio en forma específica. A continuación se describe el funcionamiento del manejo de espacio de la Base de Datos Institucional.

#### 4.1.1. Estructuras de espacio en la Base de Datos de un manejador

##### Oracle

Una Base de Datos en Oracle® es una colección de datos que es tratada como una unidad. El propósito de una Base de Datos es almacenar y recuperar información relacionada. La Base de Datos tiene estructuras lógicas y físicas.

La estructura lógica de una Base de Datos incluye *tablespace*, objetos de esquema, bloques de datos, *extents* y segmentos. Mientras que la estructura física incluye archivos de datos, de *redo* y de control.

Una Base de Datos está dividida en unidades lógicas de almacenamiento llamados *tablespaces*, las cuales agrupan estructuras lógicas relacionadas, la Base de Datos puede contener uno o más *tablespaces*.

Las unidades de utilización de espacio en una Base de Datos son bloques, *extents* y segmentos. En el nivel más fino de granularidad, se almacenan datos en bloques de datos también conocidos como bloques lógicos o bloques Oracle®. Un bloque de datos corresponde a una cantidad específica de bytes del espacio físico en disco de la Base de Datos. El manejador administra el almacenamiento de espacio en los archivos de datos en bloques de datos. El manejador de base de datos solicita la información en múltiples bloques de datos, que no son los mismos que bloques de sistema operativo. Se establece el tamaño del bloque de datos para cada Base de Datos en Oracle® cuando se crea, debe ser múltiplo de tamaño del bloque de sistema operativo dentro del máximo límite permitido por el propio sistema operativo. Los bloques de datos Oracle® son las unidades más pequeñas de almacenamiento que el manejador puede utilizar o asignar.

El siguiente nivel de espacio lógico de la Base de Datos es un *extent*. Un *extent* es un conjunto específico de bloques de datos contiguos utilizados para almacenar un tipo específico de información. Cuando se crea una tabla, el manejador asigna un segmento de datos a la tabla con un *extent* inicial de un número especificado de

bloques de datos. Aunque no se inserten datos en la tabla, los bloques de datos que corresponden al *extent* inicial, son reservados para los registros de la tabla.

Si los bloques de datos de un *extent* inicial del segmento se llenan y es requerido más espacio para contener datos nuevos, el manejador automáticamente asigna un *extent* incremental para este segmento. Un *extent* incremental es un *extent* subsecuente del mismo o mayor tamaño que el *extent* asignado previamente en este segmento.

El nivel de almacenamiento lógico de la Base de Datos por encima de un *extent* es llamado segmento. Un segmento es un conjunto de *extents*, que ha sido utilizado para una estructura de datos específica. De esta forma los datos de una tabla son almacenados en su propio segmento de datos, mientras que cada dato en un índice está almacenado en su propio segmento de índice. Si la tabla o índice es particionado cada partición es almacenado en su propio segmento.

El manejador utiliza espacio para un segmento en unidades de un *extent*. Cuando el *extent* existente de un segmento se llena, se asigna otro *extent* para este segmento. Debido a que los *extents* son asignados como son necesarios, los *extents* de un segmento pueden o no ser contiguos en disco.

Un segmento y todos sus *extents* son almacenados en un *tablespace*. Dentro de un *tablespace*, un segmento puede incluir *extents* de más de un archivo, es decir, un segmento puede distribuirse entre varios archivos de datos. Sin embargo, cada *extent* pertenece a uno y sólo un archivo de datos.

Uno o más archivos de datos son explícitamente creados por cada *tablespace* para almacenar físicamente los datos de todas las estructuras lógicas en un *tablespace*. El tamaño combinado de un archivo de datos de un *tablespace* es la capacidad de almacenamiento total de un *tablespace*.

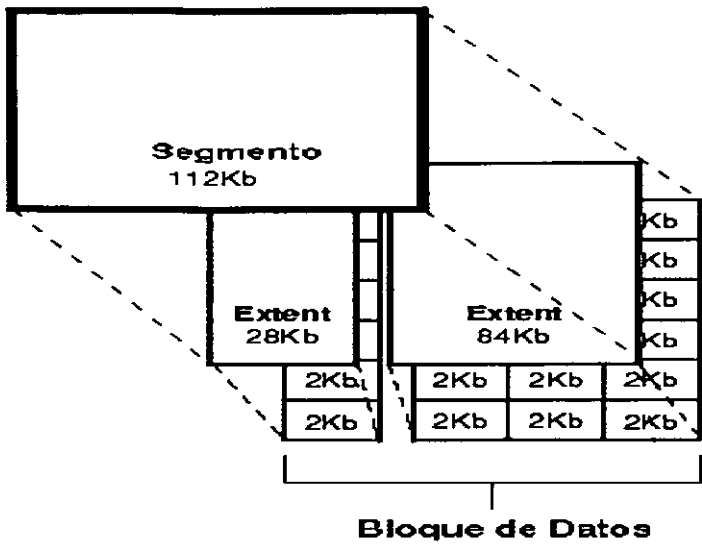


Figura4.1 Estructuras de espacio.

Cada Base de Datos tiene uno o más archivos de datos físicos, los cuales contienen todos los datos de la Base. Los datos de estructuras lógicas de la Base de Datos, tales como tablas e índices, son físicamente almacenados en los archivos de datos asignados para la Base de Datos.

Un archivo de datos puede ser asociado sólo con una Base de Datos, puede tener ciertas características establecidas para permitirles extenderse automáticamente cuando la Base de Datos necesita espacio, un *tablespace* puede tener uno o más archivos de datos.

Cada Base de Datos tiene un conjunto de dos o más archivos de *redo*. Un archivo de *redo* es un conjunto de cambios a la Base de Datos. La función primaria de los archivos de *redo* es registrar todos los cambios hechos a los datos. Lo utiliza el manejador para prevenir que un dato modificado durante una falla se escriba permanentemente a los archivos de datos, los cambios pueden ser obtenidos de los archivos de *redo* y el trabajo no se pierde.

Cada Base de Datos tiene un archivo de control, el cual contiene registros que especifican la estructura lógica de la Base de Datos. Incluye el nombre de la Base de Datos, ubicación de los archivos de datos, archivos de *redo* y registra el momento en que se creó la Base de Datos.

Un esquema es una colección de objetos de la Base de Datos. Los objetos de un esquema incluyen estructuras tales como tablas, vistas, secuencias, procedimientos almacenados, sinónimos, índices, *clusters* y ligas de base de datos.

El manejador cuenta con diversos parámetros que permiten manipular la forma en que los segmentos de la Base de Datos crecerán:

**INITIAL:** Parámetro que establece el tamaño del *extent* inicial que será reservado para el segmento. No puede ser modificado.

**NEXT:** Establece el tamaño del *extent* siguiente. Puede alterarse el valor aun después de creado.

**PCTINCREASE:** Indica el porcentaje de incremento que registrarán los *extents* con respecto al anterior a partir del tercer *extent*. Se permite modificar su valor posterior a su creación.

**MAXEXTENT:** Indica el número máximo de *extents* que pueden ser asignados al segmento. Actualizable después de creado.

Derivado de las definiciones anteriores se observa que, dentro de las estructuras físicas de la Base de Datos, los elementos contenidos en los archivos de datos son los que registran crecimiento y que pueden con mayor urgencia requerir espacio.

Basados en la necesidad de conocer la tasa de crecimiento de los segmentos, se decidió almacenar un histórico de esta información que permitiera elaborar estadísticas de crecimiento con el fin de proyectar depuración de los sistemas transaccionales, evaluar que los parámetros establecidos para los segmentos sean acordes al crecimiento que registran, la adquisición de espacio y con esto principalmente evitar errores en la Base de Datos por espacio insuficiente para continuar su función.

#### **4.1.2. Estándares Institucionales**

Respecto al manejo de espacio en las Bases de Datos existen en la institución normas de estandarización, a continuación se describen algunas de ellas:

Debe definirse un *tablespace* para datos y otro para índices por cada sistema.

Si dentro de un sistema, se identifica una tabla que se espera registre mayor crecimiento, debe definirse un *tablespace* independiente para datos e índices para contener los segmentos involucrados.

El nombre de los *tablespace* debe indicar en las primeras letras el mnemónico del sistema y el tipo de segmentos que contiene por ejemplo: MEDI\_DAT para las tablas del sistema MEDI.

Del mismo modo, los segmentos de un sistema deben contener un prefijo con el mnemónico, por ejemplo: MEDI\_MONEDAS, tabla del sistema MEDI que contiene datos de monedas.

Respecto a parámetros de almacenamiento:

INITIAL: Debe ser calculado para contener por lo menos la carga inicial, y preferentemente un año de movimiento.



NEXT: El valor de este parámetro debe al menos soportar el crecimiento de un mes.

PCTINCREASE: El valor deber ser 0 para disminuir el riesgo de fragmentación en el *tablespace*.

Cuando se libera un sistema o segmentos de un sistema, deben someterse a revisión los archivos de creación, para verificar que se toman en consideración los estándares institucionales.

#### **4.1.3. Implementación Mercado de Datos de Espacio**

El objetivo inicial del proyecto era conocer el incremento en espacio para los sistemas transaccionales de la institución, durante el transcurso del desarrollo se encontró que el mercado de datos permite evaluar los parámetros de almacenamiento de los segmentos.

Acerca de los requerimientos, el objeto del Mercado de Datos es el espacio físico utilizado por los archivos de datos de las Bases de Datos y el espacio lógico ocupado por los segmentos de los sistemas de la institución. La granularidad elegida durante el prototipo fue de *tablespace*, ya que sólo se requería conocer el crecimiento global por sistema, posteriormente se modificó este aspecto, se tuvo la necesidad profundizar en el tema para conocer que segmentos presentaban mayor cambio (aumento o disminución) en espacio, actualmente la granularidad que se utiliza es al nivel de segmento

Para el espacio lógico se registran incrementos en el espacio que utiliza un segmento, modificaciones en los parámetros del segmento, así como en el espacio físico, las modificaciones en el espacio asignado para los archivos de datos o adición de nuevos archivos de datos. Las dimensiones sobre la que se registrarán movimientos son las que describen los segmentos.

El modelo resultante de la recopilación de requerimientos, es un modelo copo de nieve, ya que la Base de Datos a utilizar es relacional y las dimensiones tienen también asociadas tablas descriptivas.

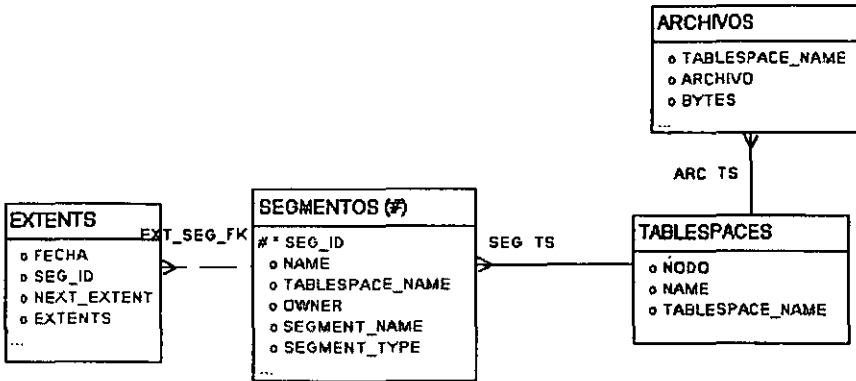


FIG. 4.2 Modelo de Mercado de Datos de Espacio

Las tablas relacionadas en el mercado de datos son:

- La tabla de *extents* es considerada la tabla de hechos y registra cambios en los parámetros de los segmentos, next, max\_extent, pct\_increase, número de *extents* y bytes.
- La tabla de segmentos es una dimensión descriptiva del segmento.
- La tabla de *tablespaces* es una dimensión descriptiva del segmento y los archivos.
- La tabla de archivos es una dimensión dinámica que registra cambios en los archivos de datos.

Para la carga y transformación, la información es extraída de bases de datos Oracle, las vistas que se utilizaron para extraer la información pertenecen al diccionario de datos DBA\_SEGMENTS y DBA\_DATAFILES.

Para la carga se utiliza el método de adición, se hizo primero una carga inicial con la información de los segmentos en ese momento, es decir, el tamaño en bytes del segmento, el número de *extents* que lo conforman y los parámetros de almacenamiento, en este método de carga no se actualizan los registros, es decir no se modifica los registros en el mercado de datos, siempre se inserta un registro nuevo cuando existe un cambio.

Las tablas fuentes registran las características actuales de los segmentos, pero en el mercado de Datos se registran incrementos, para esto, es necesario primero agrupar por segmento los datos que actualmente se encuentran en el Almacén de Datos y después comparar contra los datos que se obtuvieron de la Base de Datos origen para determinar si existió modificación, en caso de que sea así se inserta un nuevo registro.

Para automatizar las cargas se utiliza el *crontab* de *unix*<sup>®</sup>, se programa para que una vez al día, durante la mañana, realice la carga y transformación.

### **Carga y transformación para segmentos**

Durante la carga el archivo de trabajo se programa a través del *crontab* y realiza las siguientes actividades:

Se conecta a la Base de Datos origen, a través de una liga de Base de Datos, selecciona los datos de las vistas *dba\_segments* y *dba\_data\_files* y los inserta en una tabla temporal, esta conexión dura unos segundos y se desconecta para evitar degradar el tiempo de respuesta en los sistemas transaccionales

Después de llenar las tablas temporales realiza las comparaciones para registrar las diferencias que detecto entre los datos contenidos en el Almacén de Datos y los que obtiene en ese momento de las bases fuente, si existen modificaciones a los segmentos, se inserta un nuevo registro en la tabla de *extents*, cuando se trata de un segmento recientemente creado, se lleva a cabo la inserción de un registro en la

tabla de segmentos, así como en la tabla de *extents*. Para los archivos de datos, el procedimiento es similar, si encuentra diferencia de tamaño entre los datos extraídos y los que se encuentran registrados en el almacén de datos, inserta un registro en archivos con el incremento o decremento detectados en la comparación.

Este procedimiento se realiza para cada Base de Datos de sistemas transaccionales en producción.

La carga genera archivos de bitácora donde puede verificarse si existió algún problema con la carga y tomar las precauciones pertinentes.

### Consulta

La herramienta utilizada, Oracle Discoverer, permite definir consultas dinámicas y conservar las consultas que se requieren comúnmente. Diariamente se consultan el histórico de los segmentos que registran incrementos para ese día, un ejemplo de este reporte se muestra en la Figura4.3:

Consulta								
Page Items:	Name: SIRAC	Prop: SIRAC	Segmento: PR_SALDOS_PLAN_PAGO_FIN_MES					
Fecha	Tablespace	Tipo	Inicial	Next	Ext	Bytes	Max	Pct
14-JUN-1999	SIRAC_BPR_DAT	TABLE	52,428,800	20,971,520	1	52,428,800	505	00%
01-OCT-1999					0	20,971,520	0	00%
03-DEC-1999					0	41,943,040	0	00%
03-JAN-2000					0	20,971,520	0	00%
02-FEB-2000					0	20,971,520	0	00%
02-MAR-2000					0	20,971,520	0	00%
04-APR-2000					0	20,971,520	0	00%
01-MAY-2000					0	41,943,040	0	00%
01-JUN-2000					0	20,971,520	0	00%
03-JUL-2000					0	20,971,520	0	00%
01-AUG-2000					0	41,943,040	0	00%
01-SEP-2000					0	20,971,520	0	00%
01-SEP-2000				10,485,760	0	0	0	00%
03-OCT-2000					0	31,457,280	0	00%
<b>Total</b>			<b>734,003,200</b>	<b>31,457,280</b>	<b>16</b>	<b>377,487,360</b>	<b>505</b>	<b>0</b>

Figura4.3 Listado de histórico de un segmento que registro incremento el día de hoy

Consultando el histórico por segmento, es posible verificar, la cantidad de *extents* que está creciendo cuando se incrementa su espacio, para vigilar si es necesario

aumentar el parámetro de `next_extent` para satisfacer las necesidades de crecimiento mensuales del segmento.

En el ejemplo se puede observar una modificación en el parámetro `next_extent` el día 01 de septiembre, el parámetro se alteró como respuesta al crecimiento que se había estado registrando para este segmento, Figura4.3.

El mercado de datos permite detectar, también, la creación de segmentos nuevos en la base y los parámetros de almacenamiento con que fue definido.

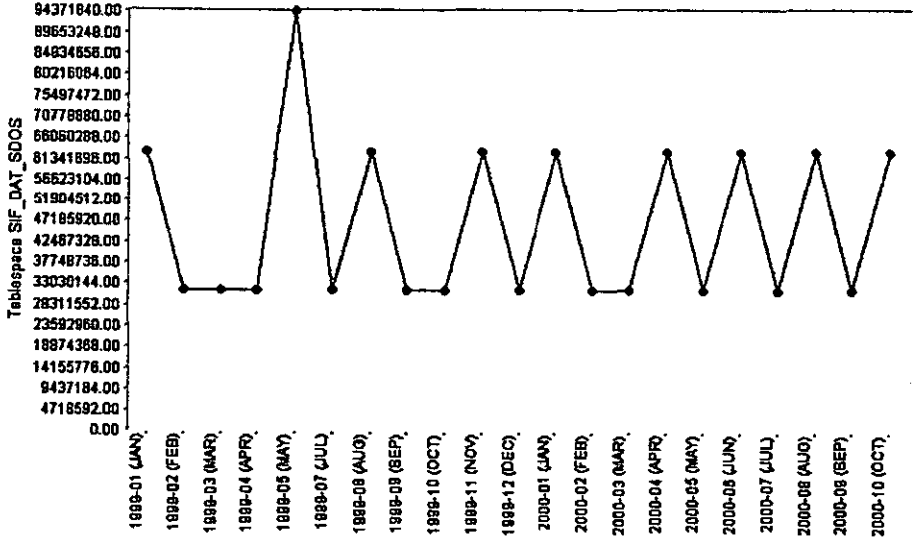


Figura4.4 Gráfico de crecimiento mensual

La consulta del crecimiento mensual muestra el incremento de espacio por `tablespace` para cada mes, proporcionando así argumentos para evaluar la

capacidad actual de la base, es decir, permite juzgar la suficiencia de espacio definido actualmente para el sistema. La Figura4.4 ilustra este hecho.

<b>Incrementos en espacio físico</b>					
<b>Base</b>	<b>Archivo</b>	<b>Fecha</b>	<b>Antes</b>	<b>Incremento</b>	<b>Despues</b>
SIRAC	/bd_sirac/inw/sirac_bpr_inx1.dbf	21-JUL-1999	629,145,600	419,430,400	1,048,576,000
		04-NOV-1999	1,048,576,000	209,715,200	1,258,291,200
		03-DEC-1999	1,258,291,200	314,572,800	1,572,864,000
		03-FEB-2000	1,572,864,000	209,715,200	1,782,579,200
		15-MAR-2000	1,782,579,200	104,957,600	1,887,436,800
		27-MAR-2000	1,887,436,800	209,715,200	2,097,152,000
		03-MAY-2000	2,097,152,000	104,957,600	2,202,009,600
		11-MAY-2000	2,202,009,600	419,430,400	2,621,440,000
	<b>Tot. Inc</b>			<b>1,992,294,400</b>	

Figura4.5 Listado de incremento en espacio físico.

El reporte de incremento en espacio físico permite consultar las variaciones de espacio en disco, por si se llevó a cabo un redimensionamiento de un archivo de datos o se adicionó uno nuevo, así como la fecha en que se realizó este cambio, como se muestra en la Figura4.5. Las modificaciones al espacio físico deben ser documentadas por el DBA, al registrar el mercado de datos estos cambios permite que sean consultados como si hubieran sido documentados en la forma tradicional.

El mercado de datos además de satisfacer los objetivos para los que fue creado, permitieron también, evaluar la suficiencia de los parámetros de almacenamiento de los segmentos, es decir, que fueron calculados adecuadamente, así como vigilar que las recomendaciones institucionales en este aspecto se estén cumpliendo, para realizar recomendaciones al área correspondiente, o llevar acabo modificaciones para los parámetros para los que es posible. Adicionalmente, sirve como auxiliar para vigilar que la creación de segmentos nuevos sea debidamente documentada.

ESTA TESIS NO SALE  
DE LA BIBLIOTECA

## **4.2. Mercado de Datos de Sesiones**

En el área de administración de la Base de Datos surge la necesidad de conocer el número de usuarios concurrentes en los sistemas, además de detectar aquellos usuarios que no registran conexión en algún tiempo, con el fin de optimizar el uso de licencias de Oracle®, así como tener un registro de los privilegios de acceso de los usuarios definidos. A continuación se describe el funcionamiento de los esquemas de seguridad de la Base de Datos institucional.

### **4.2.1. Seguridad en el manejador de Base de Datos Oracle®**

La seguridad de la Base de Datos permite habilitar o deshabilitar acciones del usuario sobre la Base de Datos y los objetos contenidos en ella. Con este fin el manejador utiliza esquemas y dominios de seguridad para controlar el acceso a los datos y restringir el uso de varios recursos de la Base de Datos.

El RDBMS habilita el control de acceso a los objetos a través de privilegios. Un privilegio es el permiso de acceso a un objeto en una forma prescrita, como por ejemplo consultar una tabla. Los privilegios son otorgados a los usuarios.

Para el manejador un usuario es un clave de acceso definido en la Base de Datos que se puede conectar y acceder a los objetos. Un esquema es una colección de objetos, como tablas, vistas cluster, procedimientos. A cada usuario se le designa un esquema con el mismo nombre que la clave de usuario.

Cuando se crea un usuario en Base de Datos o se altera uno existente, el usuario administrador debe tomar varias decisiones concernientes a la seguridad. Esto incluye:

- El método de autenticación que se utilizará, indica la forma en que validará el manejador la autenticidad del usuario.
- El *tablespace* de temporal, indica un espacio lógico donde se realizarán los ordenamientos que solicite el usuario.

- Los *tablespace* sobre los que el usuario tendrá cuota, indica el *tablespace* en que se crearán los segmentos que genere un usuario cuando no se indica en forma implícita.
- Los límites de recursos, agrupados en perfiles donde el manejador permite restringir el uso de recursos tales como, uso de cpu, número máximo de sesiones, límite de memoria que utilizará, etcétera.
- Roles y privilegios requeridos para realizar sus operaciones.

Un privilegio de sistema es el derecho a realizar un acción en particular en la base de datos, o a realizar una acción en algún objeto de esquema de cualquier tipo, por ejemplo privilegio para crear una tabla o para borrar los registros de cualquier tabla existente en la base.

Oracle® recomienda que los privilegios de sistema se otorguen exclusivamente a administradores y desarrolladores, ya que el usuario final no requiere de estas capacidades.

Un privilegio de objeto es el derecho a realizar un acción en particular sobre un objeto en específico del esquema indicado. Los privilegios de objeto varían según el tipo de objeto, es decir para una tabla, existe el privilegio que permite borrar registros, privilegio que no tienen los procedimientos, los cuales tienen el privilegio de ejecución, algunos objetos no tienen privilegios de objeto asociados.

## **Licencias**

La licencia de Oracle® es un permiso restringido para utilización de software que se adquiere con el proveedor. Existen varios tipos de licencias las más comunes son las licencias nombradas y las licencias concurrentes. En una licencia concurrente se toma en cuenta todas las sesiones que se encuentran en un tiempo, incluyendo procesos en lote, en línea, cada sesión de una conexión con múltiples sesiones. Las licencias nombradas limitan el número usuarios definidos, no existe restricción en la cantidad de sesiones por cada usuario.



Si el DBA identifica que se requiere un mayor número de licencias, debe solicitar adquisición de una cantidad apropiada. El manejador permite limitar el número de sesiones concurrentes conectadas y creados en un Base de Datos, pero la intención no era únicamente limitar sino evaluar si eran suficientes la cantidad de licencias que se habían adquirido para solicitar se aumente el monto o reasignar las que se tenían.

#### **4.2.2. Estándares Institucionales**

En cada sistema existe un propietario único, de forma que sólo él puede crear objetos y a los usuarios finales se les otorgan privilegios de acceso.

De ser posible el nombre del usuario propietario debe coincidir con el mnemónico del sistema.

El único privilegio de sistema que puede tener un usuario final es crear sesión que le permite establecer una conexión a la Base de Datos.

La creación de usuarios debe considerar el establecer un *tablespace* por omisión y un *tablespace* de temporal.

Los privilegios a los objetos se otorgan a través de roles.

El nombre de los roles para acceso a un sistema incluirá el prefijo del sistema, y una descripción del rol, por ejemplo: MEDI\_CONSULTOR, rol del sistema MEDI con privilegios de consulta únicamente.

A un usuario que requiera acceso a varios sistemas en una base de datos, se le asignará un sólo usuario de Base de Datos con los roles correspondientes para los sistemas que requiera.

Cuando se libera un sistema se somete a revisión los archivos de creación de usuarios, para verificar se cumplan los estándares y se incluyan las consideraciones especiales de cada sistema, por ejemplo un rol específico que deban tener los usuarios del sistema, acceso a la herramienta menú, etcétera.

#### **4.2.3. Implementación Mercado de Datos de Sesiones**

El objetivo del proyecto consistía en conocer el número de usuarios concurrentes en las diferentes Bases de Datos con el fin de evaluar la asignación adecuada de licencias. Recientemente, se dio un cambio en la estructura de las claves de usuarios, anteriormente se definía un usuario por cada sistema, para una misma persona aunque estuvieran en la misma Base de Datos, con el fin de optimizar las licencias se modificó este estándar y actualmente únicamente se da de alta un usuario por base y a través de los roles se le concede acceso a varios sistemas, tomando en consideración el nuevo esquema se adicionaron al mercado de datos los roles que tienen asignados los usuarios.

Existe en la base de datos una opción de auditoria, la cual permite registrar algunos detalles de los accesos de los usuarios a la base de datos, estos registros los guarda en una tabla de la misma base de datos. Se pensó en utilizar esta opción trasladando los datos de esta tabla a otra para evitar que las consultas de esta información perjudicasen el desempeño de los sistemas en producción. Pero algunos sistemas, especialmente el sistema contable presentó problemas con la ejecución de algunos procesos. También se encontró que la última versión del manejador de base de datos contaba con un control implícito de la frecuencia de accesos a la base de datos por usuario, pero el tiempo era apremiante y se requería de un periodo largo para realizar la migración de todas los sistemas a dicha versión.

La granularidad utilizada en el prototipo fue a nivel sesión y continua de esta forma ya que ha sido capaz de satisfacer las necesidades de las áreas usuarias.

Las dimensiones sobre las que se registran movimientos son nodos, instancias, sistemas, usuarios y roles con respecto a sesiones, no se definió una tabla de dimensión independiente para el tiempo, debido a que el tiempo tiene un comportamiento continuo, se encuentra incluido en la tabla de hechos.

El modelo resultante de la recopilación de requerimientos, es un modelo copo de nieve, ya que la Base de Datos a utilizar es relacional y las tablas de dimensiones tienen a su vez tablas que las describen.

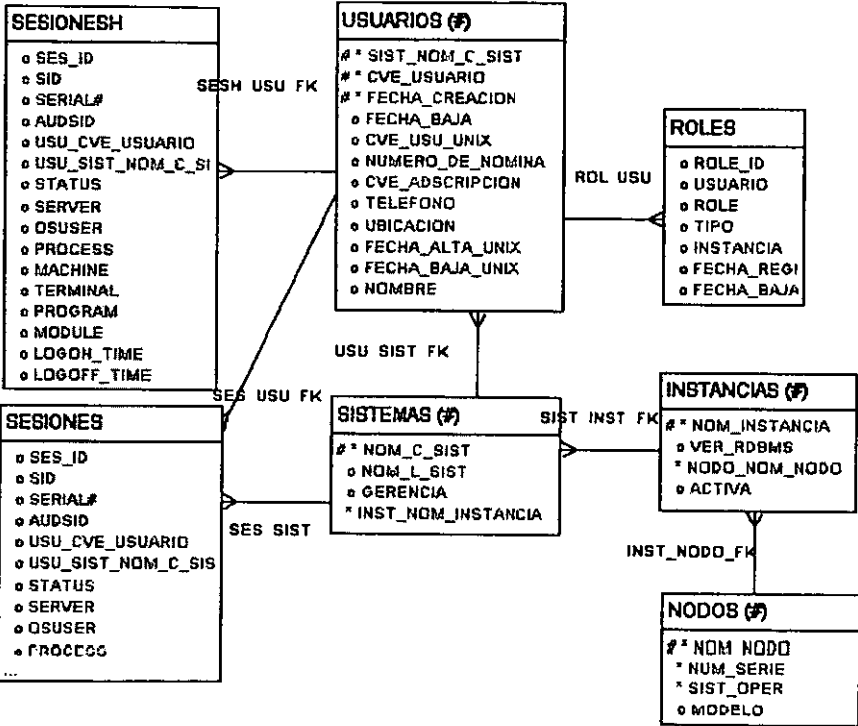


Figura4.6 Modelo Copo de Nieve del Mercado de Datos de Sesiones

Las tablas relacionadas en el Mercado de Datos de sesiones ilustradas en la Figura4.6 son:

- La tabla de **sesiones** es una tabla temporal de hechos donde se registran las conexiones que aun se encuentran establecidas.
- La tabla de **sesionesh** es considerada la tabla de hechos y registra conexiones a las Bases de Datos transaccionales, cuando se han desconectado.
- La tabla de **usuarios** es una dimensión dinámica y descriptiva de los usuarios que han sido dados de alta en las Base de Datos transaccionales ya que conserva el histórico de los que se borran registrando la fecha de baja.
- La tabla de **sistemas** es una dimensión o catálogo donde se encuentran registrados los sistemas de la institución, esta información no se encuentra en las tablas del diccionario de datos, por ello es necesario obtener la información de otras fuentes.
- La tabla de **instancias** es una dimensión descriptiva de Bases de Datos transaccionales.
- La tabla de **nodos** es una dimensión descriptiva de servidores donde se encuentran las Bases de Datos.
- La tabla de **roles**, es una dimensión dinámica que registra los roles que tienen asignados los usuarios de cada base.

Tanto las Bases de Datos de origen, como el Almacén de Datos son Oracle, las vistas de donde se extrae la información para poblar el mercado de datos son: DBA\_USERS, V\$SESSION, V\$DATABASE, DBA\_ROLE\_PRIVS y MENU\_B\_USER.

Para la carga se utiliza el método de fusión constructiva, ya que actualiza los registros existentes en la base y si no los encuentra los registra como nuevos, las tablas fuente registran situación actual, es decir usuarios que se encuentran en ese momento conectados, los datos actuales obtenidos que no se encuentran aun registrados se insertan, pero los registros que se encuentran en la tabla sesiones y no en la carga nueva, se actualizan como sesiones que se desconectaron y pasan a la tabla histórica **sesionesh**.

Para automatizar las cargas se utilizó el crontab de unix®, se programó para que cada 5 minutos, durante todo el día, realice la carga y transformación.

### **Carga y transformación**

La secuencia de carga y transformación, programada a través del crontab, es la siguiente:

Se extrae el registro actual en las vistas v\$session y dba\_users de la Base de Datos origen, a través de una liga de Base de Datos, y lo inserta en una tabla temporal, esta conexión a los sistemas transaccionales dura unos segundos y se desconecta para evitar degrado del tiempo de respuesta.

Después de llenar las tablas temporales realiza las comparaciones para registrar los cambios que detectó en la comparación, ya sea inserción de una nueva conexión o registro de salida de una conexión que ya había sido reconocida. En esta fase también valida, la creación o borrado de usuarios de la base.

La carga de los roles, se lleva a cabo dos veces el día, el procedimiento de carga se comporta de la misma manera que las cargas anteriores, realizando una comparación entre los datos actuales y los que se obtienen de extracción de los sistemas transaccionales, las vistas origen son dba\_role\_privs que registran los roles de Base de Datos que tienen asignados los usuarios y menu\_b\_user en donde se registra los roles de menú para la herramienta Menu5 de Oracle®.

### **Consulta**

Para el objetivo original del proyecto se definió una consulta por semana que indica el número de usuarios concurrentes a una hora que solicita como parámetro. El reporte despliega la cantidad de usuarios concurrentes para cada día a la hora y en la semana que se indican al momento de ejecutar, como se muestra en la Figura4.7.

	21-AGO-2000	22-AGO-2000	23-AGO-2000	24-AGO-2000	25-AGO-2000
	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES
NAFIN3	92	107	104	146	95
▶ ORAMOD	51	59	59	102	60
BNC (Buro Nacional de Crédito)	3	3	2	2	3
CORPORAT (Sistema Corporativo Institucional)	11	12	6	18	4
CPRES (Control Presupuestal)	8	13	10	11	12
FIDUCIA (Sistema Fiduciario)	20	22	26	56	30
FONAFE (Fondo Nacional de Fomento Epdal)	11	11	11	11	11
GIA (Sistema Garantía)	4	5	4	5	6
PROVEEDOR (Proveedores)	1	1	1	1	1
PUBLICA (Publicaciones)	1	1	1	1	1
SIA (Sistema Integral de Auditoría)	1	1	3	3	1
SIME (Sistema de Inventario)	2	2	5	1	3
▶ ORASIF	41	48	46	44	36
SIF (Sistema Integral Financiero)	41	48	46	44	36
NAFIN5	100	115	109	131	99
▶ ORASID	14	16	14	11	12
SIDECC (Sistema Integral de Descuento y Contc)	14	16	14	11	12
▶ SIRAC	86	99	95	120	86
SIRAC (Sistema Integral de Recuperación y Adm)	86	99	95	120	96
NAFIN6	227	249	251	298	234
▶ JORLAB	25	23	19	18	15

Figura4.7 Listado de usuarios concurrentes para la semana 34 a las 12:30 hrs.

Se definió, también, una consulta que obtuviera el máximo número de sesiones concurrentes en cada día de una semana.

Pogo Items: <b>Semana: 2000-34</b>		Maximo Sesiones Concurrentes						
		21-AUG-2000	22-AUG-2000	23-AUG-2000	24-AUG-2000	25-AUG-2000	Total	Promedio
		LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES		
NAFIN3		110	113	120	138	110	591	118
ORAMOD		55	57	62	81	59	314	63
ORASIF		55	56	58	57	51	277	55
NAFIN5		109	126	110	124	99	568	114
ORASID		16	23	18	15	17	91	18
SIRAC		91	103	92	109	82	477	95
NAFIN6		245	266	258	285	257	1311	262
JORLAB		55	57	56	56	52	276	55
ODWFIN		8	8	6	5	4	31	6
ORAFIN		80	78	73	93	78	402	80
ORAINT		40	42	48	43	42	215	43
ORASIP		62	81	75	88	81	387	77

Figura4.8 Listado de número máximo de sesiones en cada día de la semana 34

En la Figura4.8 se despliega la semana 34, es posible observar que el máximo de sesiones concurrentes en las bases, no se presentó a las 12:30. Existe otro reporte donde podemos consultar en que horario se alcanzó el máximo. El máximo de sesiones muestra la información por instancia, a diferencia de sesiones concurrentes que proporciona el detalle a nivel sistema.

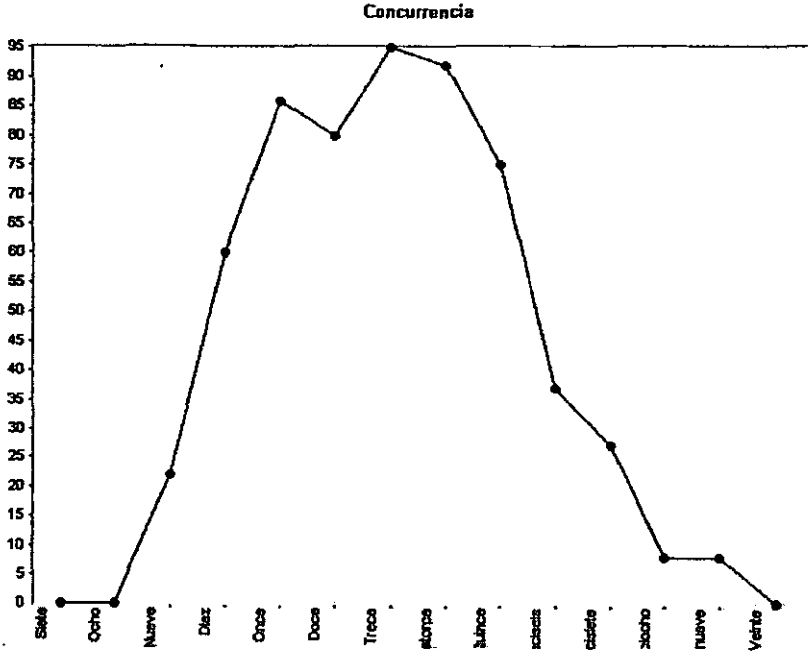


Figura4.9 Gráfico de concurrencia por hora

En el gráfico de la Figura4.9 es posible observar más fácilmente las horas de mayor concurrencia para un sistema en específico.

El mercado de datos brinda la oportunidad de consultar la fecha de último acceso de los usuarios de una Base de Datos.

### Accesos por Instancia

Instancia	Sistema	Usuario	Fecha Alta	Ultima conexion
ORASIF	SIF	AALAVEZ	21-JUN-1996	Sin Acceso
		AALCARA1	08-DEC-1998	24-OCT-2000 09:17:40
		AALFARO	26-SEP-2000	18-OCT-2000 14:59:38
		AARVIZU	21-JUN-1996	23-OCT-2000 11:46:53
		ABLANCO	12-SEP-1996	Sin Acceso
		ACEBALLO	21-JUN-1996	16-OCT-2000 14:17:28
		ADELAROS	03-FEB-2000	23-OCT-2000 17:15:06
		AFRAGOSO	21-JUN-1996	24-OCT-2000 09:55:26
		AGARCIA2	10-JUN-1999	09-OCT-2000 12:37:27
		AGRACIAN	26-SEP-2000	Sin Acceso
		AGUEVARA	21-JUN-1996	23-OCT-2000 13:26:00
		AGUZMAN1	14-JUL-1999	29-SEP-2000 10:43:46
		AHERNAND	21-JUN-1996	24-OCT-2000 12:57:14
		ALARREA	09-NOV-1999	24-OCT-2000 09:26:44
		ALIRA	21-JUN-1996	18-AUG-1999 13:55:09
		ALOREDO	12-OCT-1999	16-OCT-2000 17:13:58

**Figura4.10** Listado de último acceso de los usuarios de una Base de Datos.

En el listado mostrado en la Figura4.10 se pueden detectar las claves de usuario que no han tenido acceso a la base. También existe un reporte que despliega únicamente los usuarios que no han registrado acceso.

El mercado de datos ha cumplido su objetivo satisfactoriamente, ya que permite conocer tanto usuarios definidos para licencias nombradas como usuarios concurrentes para licencias concurrentes; se han encontrado otras funcionalidades. Por ejemplo, permite desplegar los datos de las sesiones que tienen aplicaciones graficas; auditar quien se encontraba conectado a cierta base en un momento dado; desde donde establecen conexión los usuarios, a través del mercado de datos se evidenció que los usuarios no acceden a la base desde un lugar fijo como se pensaba.



## Conclusiones

Tomando en cuenta la necesidad de disponer de elementos como respaldo en la toma de decisiones, con el propósito de aumentar la confianza en el éxito de las soluciones propuestas y la reducción de incertidumbre. Se apuntó la aportación de la metodología de Almacén de Datos para apoyar el análisis, así como se dieron a conocer detalles y consideraciones para su implementación. Se expusieron algunas de las responsabilidades del Administrador de base de datos y determinados aspectos de la Base de Datos Oracle® utilizada en la institución, para comprender la lógica de las bases de datos fuente y el modelado de los mercados de datos.

Originalmente la tecnología de Almacén de Datos se desarrolló con el objeto de dar una ventaja competitiva a las empresas del mercado, pero en el proyecto actual se mostró el apoyo que brinda en la administración de una base de datos, haciendo más fácil la obtención de la información necesaria respecto a espacio y sesiones permitiendo de esta forma tomar decisiones más acertadas.

Los mercados de datos, de espacio y sesiones, representan para el área de administración de base de datos un importante apoyo, con información oportuna y relevante para la planificación de las actividades. Los mercados de datos cumplen satisfactoriamente los objetivos por los que fueron creados, han permitido a través de su desarrollo en espiral, retroalimentarse y satisfacer nuevos propósitos y requerimientos.

En el mercado de datos de espacio es posible consultar el crecimiento que registra un sistema, permite además desmenuzar a nivel segmento las variaciones, comúnmente incrementos, en el tamaño que ocupa. Para cada sistema se puede conocer detalles de las alteraciones de espacio, físico y lógico, de considerable utilidad. Por mencionar algunos ejemplos: los días del mes en que se registra mayor crecimiento, el promedio mensual de crecimiento, el día que se borró un segmento, la frecuencia con que un segmento genera un nuevo extent, la cantidad de extents que se adiciona a un

segmento cuando registra incrementos en el espacio, etcétera. Con el paso del tiempo se ha descubierto que puede proporcionar información casi de cualquier ámbito, en cuanto a espacio se refiere, con sólo definir la consulta en forma adecuada. Existe un reporte que se consulta diariamente durante el monitoreo de las bases de datos en producción, y varios que arrojan mayor detalle respecto a las variaciones en espacio, si se requiere profundizar, ver los datos desde un ángulo distinto o agrupado de diferentes maneras.

El mercado de datos de sesiones permitió detectar las licencias malgastadas en algunos sistemas que no estaban utilizando adecuadamente éste importante recurso, cumpliendo así el objetivo original para el que fue creado. Proporcionó también datos importantes respecto a los accesos de los usuarios que permitieron auditar las conexiones a las bases de datos en producción y desarrollo, un dato histórico que ha sido de utilidad es la fecha de borrado de un usuario, dentro de la funcionalidad de la base de datos esta información no se conserva, y no queda rastro de un usuario cuando se borra de la base. Ha sido de utilidad para la implementación de proyectos de otras áreas en la gerencia, por ejemplo se deseaba saber con que usuario en sistema operativo se conectan usuarios de algunos sistemas a la base de datos con el fin de unificar cuentas, esta información se obtuvo del mercado de datos de sesiones. Otro proyecto institucional, que aun se encuentra en implementación es el de seguridad de una sola conexión (Single Sign on), para esto se requería conocer las computadoras desde las que se tiene acceso directo a las bases de datos. El mercado de datos de sesiones se consulta diariamente con fines de monitoreo, pero ha proporcionado en innumerables casos respuesta a cuestionamientos de las áreas directivas. Durante el periodo que lleva puesto en marcha ha sido necesario hacer pequeñas modificaciones para satisfacer los requerimientos, pero la respuesta a las solicitudes toma un tiempo mínimo con resultados satisfactorios, por ejemplo recientemente se adicionó información referente a roles, para uso del área de control en producción.

De éste modo, se observa, que el uso de la tecnología de Almacén de Datos apoya el manejo y uso adecuado de la información, para la toma de decisiones oportunas y fundamentadas. Los mercados de datos, de espacio y sesiones, sustentan las decisiones del administrador de la base de datos de la institución, proporcionando un conocimiento más completo del comportamiento de las bases, y aportando elementos importantes para que las resoluciones propuestas sean las más convenientes.

Debido a que se utilizaron recursos con los que cuenta la institución, no hubo un costo adicional del proyecto, aunque es posible considerar el costo individual de la licencia de acceso al manejador de base de datos Oracle<sup>®</sup>, la licencia de uso de Oracle Discoverer<sup>®</sup> y el espacio en disco utilizado por el Almacén de Datos, cerca de 300 megabytes, aproximadamente el conjunto de recursos mencionado tiene un costo de quince mil pesos.

El ahorro, que los mercados de datos han propiciado para la institución, puede contarse en millones de pesos en licencias de base de datos no utilizadas y ha permitido distribuir los recursos de espacio en forma optimizada.

Los mercados de datos, se encuentran en uso, y se continuaran explotando durante el tiempo que las bases de producción se encuentren en la versión actual, ya que un cambio en éste aspecto requeriría realizar modificaciones al modelado y considerar que algunas funcionalidades vienen implícitas en la versión siguiente. Mientras tanto la cantidad de reportes requeridos a los mercados de datos aumentan y se diversifican.

## Glosario

**Arquitectura.** Término que, restringido únicamente al ámbito del desarrollo de un sistema, indica la estructura orgánica, conceptual y técnica bajo la que se ha creado.

**Base de Datos.** Colección de datos organizada de tal modo que la computadora pueda acceder rápidamente a ella. Una base de datos relacional es aquella en la que las conexiones entre los distintos elementos que forman la base de datos están almacenadas explícitamente con el fin de ayudar a la manipulación y el acceso a éstos.

**Bug.** Término aplicado a los errores descubiertos al ejecutar un programa informático. Fue usado por primera vez en el año 1945 por Grace Murray Hooper, una de las pioneras de la programación moderna, al descubrir cómo un insecto (bug) había dañado un circuito de la computadora Mark.

**Cliente.** Un sistema o proceso que solicita a otro sistema o proceso que le preste un servicio. Una estación de trabajo que solicita el contenido de un archivo a un servidor de archivos es un cliente de éste servidor.

**Cliente-Servidor.** Se denomina así al binomio consistente en un programa cliente que consigue datos de otro llamado servidor sin tener que estar obligatoriamente ubicados en la misma computadora. Ésta técnica de consulta 'remota' se utiliza frecuentemente en redes como 'Internet'.

**Contraseña.** Conjunto de caracteres alfanuméricos que permite a un usuario el acceso a un determinado recurso o la utilización de un servicio dado.

**Data Marts, Mercado de Datos.** Almacén de Datos Especializado. Base de datos segmentada para poder atacar mejor algún ámbito de la información almacenada.

**Data Mining, Minería de Datos.** Especialización informática en el tratamiento y recuperación de información.

**Data Warehouse, Almacén de Datos.** Almacén de datos centralizado. Requiere de un software que permite conocer mejor el perfil del cliente y poder dar respuesta a sus necesidades.

**DBA, Database Administrator.** Persona encargada de la administración de base de datos.

**Dimensión.** Los almacenes de datos organizan un depósito grande de datos históricos y operacionales usando múltiples dimensiones para establecer categorías.

**Disco.** Dispositivo de almacenamiento de datos. Medio utilizado para el almacenamiento de información. Cuando se almacena información en un disco ésta se conserva.

**DSS, Decision Support System.** Tipo específico de sistemas para toma de decisiones.

**Granularidad.** Término que se utiliza en los almacenes de datos para expresar el nivel de detalle. A más alto nivel de granularidad, más bajo nivel de detalle, más alto nivel de abstracción.

**Megabyte.** Equivale a 1024 Kilobytes ó 1.048.576 bytes.

**Multitarea.** Capacidad que tiene una computadora de ejecutar varias aplicaciones al mismo tiempo.

**Multiusuario.** Capacidad de permitir el acceso a varios usuarios simultáneos a un determinado programa, aplicación, sistema o servicio en línea sin destruir la integridad del mismo.

**OLAP, On Line Analytical Program.** Tecnología de procesamiento analítico que crea información a partir de los datos existentes a través de transformaciones y cálculos numéricos.

**OLTP, On Line Transactional Program.** El término se utiliza para definir cualquier sistema de software que reúne datos usando las transacciones entre la fuente de datos y la base de datos.

**Nodo.** Dispositivo direccionable conectado a una red de computadoras mediante la utilización de los protocolos, permite a los usuarios comunicarse con otros sistemas anfitriones de una red. Tiene capacidad multiusuario y multitareas.

**Parámetro.** Valor especificado para conseguir los resultados deseados. Información que se añade al comando que inicia una determinada aplicación. Un parámetro puede ser un nombre de archivo o cualquier tipo de información.

**Password.** Ver contraseña.

**Profundización.** Dentro de un reporte o consulta, mostrar cada vez más detalle eligiendo en forma progresiva elementos en ese reporte.

**Programa.** Secuencia de instrucciones que obliga a la computadora a realizar una tarea determinada. Serie de instrucciones que sigue la computadora para llevar a cabo una tarea determinada.

**Query.** Consulta o búsqueda en una base de datos.

**RDBMS, Relational Database Management System.** Sistema de administración de base de datos relacional. Un sistema para almacenar datos construido alrededor del modelo relacional con base en tablas.

**Rebanar y picar.** Término común con el que los analistas definen el análisis de datos a lo largo de muchas dimensiones y a través de muchos subconjuntos.

**Servidor.** Computadora que ejecuta uno o más programas simultáneamente con el fin de distribuir información a las computadoras que se conecten con ella para dicho fin. Computadora que suministra espacio de disco, impresoras u otros servicios a computadoras conectadas con ella a través de una red.

**Sesión.** Conexión establecida entre programas que permite navegar por un sistema remoto o por un servicio (o aplicación) de dicho sistema.

**Software.** Programas o elementos lógicos que hacen funcionar una computadora o una red, o que se ejecutan en ellos, en contraposición con los componentes físicos de la computadora o la red.

**SQL, Structured Query Language.** Lenguaje estructurado de consultas. Es, según la Internacional Standard Organization (ISO), el lenguaje estandarizado de acceso a bases de datos.

**Unix.** Potente y complejo sistema operativo multiproceso / multitarea y multiusuario. Fue creado en 1969 por Ken Thompson y Dennis Ritchie, de la empresa norteamericana 'AT&T Laboratories'.

**Usuario.** Conjunto de caracteres alfanuméricos que sirven para identificar a un usuario para su acceso a la base de datos.

**Sistema Operativo.** Programa que administra los demás programas en una computadora.

## **Bibliografía**

- [COAB97] Corey, Michael J. and Abbey, Michael.  
Oracle Data Warehousing.  
Editorial Mc Graw Hill, 1997.
- [COSA93] Codd E.F., Codd S.B., and Salley C.T.  
Withe Paper "Providing OLAP to User-Analysts"  
Codd & Date, Inc, 1993.
- [DEBA96] Devlin, Barry  
"Data Warehouse from Architecture to Implementation"  
Addison-Wesley Pub Co., 1996.
- [DEPI93] De Miguel, Adoración y Piattini, Mario.  
Concepción y Diseño de Bases de Datos.  
Adisson Wesley Iberoamericana, 1993.
- [GIRA96] Gill, Harjinder S. y Rao, Prakash C.  
Data Warehousing, La integración de información para la mejor toma de  
decisiones.  
Prentice Hall, 1996.
- [KIRA96] Kimball, Ralph  
The Data Warehouse Toolkit  
John Wiley & Sons, 1996
- [INWL93] Inmon, William  
"Building the data warehouse"  
John Wiley & Sons; 1993.
- [SOAV96] Revista Soluciones Avanzadas.



Tecnología Data Warehouse.  
Año 4 número 34, 15 de junio 1996.

- [VIRE95] Poe, Vidette and Reeves, Laura L.  
Building a Data Warehouse for Decision Support  
Prentice Hall, 1995.
- [BUSO00] Business Solutions.  
Data Warehouse.  
[http://www.bfssystem.com/sol\\_dataware.htm](http://www.bfssystem.com/sol_dataware.htm) Consultada el [07,08,2000]
- [FERA00] Fernández Calvo, Rafael.  
Glosario básico inglés-español para usuarios de  
Internet.  
[http://www.ati.es/novatica/glosario/glosario\\_internet.html](http://www.ati.es/novatica/glosario/glosario_internet.html) Consultada el [03,10,2000]
- [GAHE00] Gallardo, Héctor E.  
Glosario, Diccionario de términos informáticos.  
<http://usuarios.amet.com.ar/gallahec/glosario.htm> Consultada el [18,10,2000]
- [HAMI00] Haisten, Michael.  
The Real-Time Data Warehouse: The Next Stage in Data  
Warehouse Evolution.  
[http://www.damanconsulting.com/solutions/data\\_org/whitepaper/realtimedw.htm](http://www.damanconsulting.com/solutions/data_org/whitepaper/realtimedw.htm) Consultada el [10,08,2000]
- [HUF00] Hurtado, Fernando  
Data Warehouse.  
<http://agamenon.uniandes.edu.co:80/sistemas/6702.htm> Consultada el [04,09,2000]
- [INCO00] Ingenieros Consultores Asociados.  
Datawarehousing.  
<http://www.ica.com.uy/dw/dw-doc.htm> Consultada el [31,08,2000]

- [REBR00] Red Brick Systems, Inc.  
The Data Warehouse: Enabling Better Decisions  
Faster.  
<http://www.redbrick.com/products/white/papers/datawha/datawh.html>  
Consultada el [17,08,2000]
- [WOCA00] Wolf, Carmen Gloria  
La tecnología Datawarehousing  
<http://www.inf.udec.cl/revista/edicion3-2/cwolff1.html>  
Consultada el [20,08,2000]