

## **Transacciones**

**Definición:** Es un conjunto de operaciones sobre una base de datos, donde la misma pasa de un estado consistente a otro consistente, sin conservar necesariamente la consistencia en las etapas intermedias.

**Reglas de las transacciones (ACID):**

**A Atomicidad**

**C Consistencia**

**I Isolation (Aislamiento, impedir que se ejecute otra transacción, que se entrometa, o sea se produce el bloqueo del dato)**

**D Durabilidad**

Puede existir (como en Sql7) un bit de transacción (para avisar que comienza) y puede terminar de manera exitosa (con un commit o confirmación para bajar a disco lo realizado) o puede terminar con error (y con una instrucción rollback para que se deshaga lo que se hizo)

Una transacción puede tener entorno: Monousuario

Multiusuario

Monousuario: si todo salió bien guarda físicamente los datos del log.

Multiusuario: hay acceso concurrente.

**Fallas Tipo Local:** falla en transacción corriente, en curso, overflow

**Tipo Global:** falla que afecta a todo el sistema.

- Del sistema: caída blanda, sin daño
- Del hardware: caída dura, catástrofe, daño grande.

**Protocolo de escritura de Bitácora adelantada**

**Protocolos:**

- Actualización en check point (punto de control): puede ocurrir por tiempo o por cantidad de entradas. Es rápido, se guarda todo lo hecho hasta el checkpoint.
- Actualización diferida: en un check point sólo se guarda en disco las transacciones que terminaron con un commit, no hay rollback. Guarda solo las que tiene el commit.

En caída dura: toma último back up y compara con log (que debe estar en otro disco que el de la base)

Problemas con las bases de datos distribuidas: cada sitio tiene su log, la transacción es distribuida.

Generalmente el sitio que inició la transacción hace de coordinador, hay un log para cada sitio y un log distribuido.

Se basa en un commit en 2 fases:

- Coordinador entra transacciones a hacer y los sitios le mandan un OK, si hay uno que le falla el coordinador manda orden de rollback en fase 2.
- Coordinador manda OK a cada uno de los sitios.

### **Concurrencia:**

Simultaneidad de una o varias operaciones.

Problemas:

#### **1) Modificación perdida:**

En T1 (Tiempo 1), arranca TA (Transacción A), leen dato X

En T2 (Tiempo 2), arranca TB (Transacción B), leen dato X datos = 100

En T3 (Tiempo 3), modifica TA (Transacción A), dato X (aumenta el 100%) datos = 200

En T4 (Tiempo 4), modifica TB, dato X (en base a lo que leyó en T2) (aumenta el 50%) 150 datos final.

#### **2) Dependencia no COMMITADA**

Permitir leer un dato sin esperar que una transacción que la estaba modificando haga su commit.

#### **3) Análisis consistente:**

TA (Transacción A): suma saldos

TB (Transacción B): transfiere \$10 de cuenta 1 a cuenta3

<b>CUENTA 1</b>	<b>50</b>
<b>CUENTA 2</b>	<b>40</b>
<b>CUENTA 3</b>	<b>30</b>

TA CUENTA 1 = \$50

T1 SALDO = \$50

---

TA CUENTA 2 = \$40

T2 SALDO 2 = \$90

---

TB CUENTA 1 = \$50

T3 CUENTA 1 = \$50 – \$10

CUENTA 1 = \$40

---

TB CUENTA 3 = \$30 + \$10

T4 CUENTA 3 = \$40

---

TA CUENTA 3 = \$40

T5 SALDO = \$130 (*EN REALIDAD ES \$ 120*)

Para eliminar o disminuir estos 3 problemas se utilizan protocolos:

a) Bloqueos

- **S** compartido (para lecturas)
- **X** exclusivo, este puede ser por páginas/ tabla/ registros

TB/ TA	NO BLOQUEO	S	X
NO BLOQUEO	ACCEDE A OBJETO	ACCEDE A OBJETO	NO ACCEDE
S	ACCEDE A OBJETO	ACCEDE A OBJETO	NO ACCEDE
X	NO ACCEDE	NO ACCEDE	NO ACCEDE

Problemas con los bloqueos exclusivos: puede haber un bloqueo mortal o deadlock.

TA TB

I J

J I

Recursos

Políticas para terminar con el bloqueo mortal:

- Matar los dos
- Darles un tiempo de vida
- Matar a la más vieja
- Matar al azar

**Seguridad**

**Seguridad: Que quien quiera hacer algo esté autorizado para hacerlo**

**Integridad: que lo que haga sea correcto**

- Referencial
- Constraints
- Campo fecha (Ej. : no siglo pasado)

- Triggers

#### Cuestiones a tener en cuenta:

- *Aspectos legales, éticos, etc:* que información puedo guardar y que información pueden ver los usuarios.
- *Política Interna:* Por Ej.: no de dejo que todos vean lo que ganan todos
- *Controles físicos:* Por Ej.: si se puede ingresar al centro de cómputos, y sacar un disco.
- *Política de contraseñas:* bien documentado, Por Ej.: exigir el cambio de las mismas cada 30 días, o de una longitud mínima.
- *Seguridad del sistema operativo:* es la primera barrera, pide el login.

#### Seguridad en acceso a Base de Datos

##### **Discrecional Grant**

Revoke

**Mandatorio u Obligatorio:** (muy rígido)

**Libreo naranja:** lo escribió el Dpto. de Defensa de EEUU, tiene normas y estándares de productos informáticos.

El lavanda es el de base de datos.

NADA	BAJA	MEDIA	ALTA	MAXIMO
1	2	3	4	5

Por ejemplo, el usuario 4 lee el nivel menor o igual al 4

Usuario 4 escribe solamente nivel 4.

**Ventaja:** seguridad, ya que si por ejemplo el usuario 4 modifica algo de menor nivel puede suceder que permita luego de su modificación el acceso a un usuario de menor nivel.

#### **AUDIT TRAIL**

- Login
- Día / Hora
- Lugar
- ¿Escribió?
- ¿Modifico?

Para más seguridad se pueden **encriptar** los datos.

- Permutación
- Sustitución

#### **BACKUP**

#### Cuestiones a tener en cuenta:

- ¿Dónde lo guardo?

- ¿Cada cuánto?
- ¿Qué tipo?
- **Probarlo**

Se tiene que cumplir, y verificar que se cumplan las normas. Se tiene que documentar todo y cumplir todo.

**CMM (proyecto, puntaje de 1 a 5)**

## **PLAN DE CONTINGENCIA**

Esquema de trabajo para superar extremos, de recuperación.

Métrica, equivalente al CMM, pero Español.

2 grandes etapas:

### **1) Elaboración:**

- Determinar los problemas que me pueden afectar, teniendo en cuenta la probabilidad de que sucedan.
- Escribir el plan.
- Probar el plan (ver si es factible), hacer simulacros.

### **2) Ejecución:**

- Puesta en marcha del plan
- Retorno a la actividad normal, luego pruebo el original y una vez que me aseguro que funciona bien vuelvo a trabajar con el.

## **DATAWAREHOUSE**

Un datawarehouse es un conjunto de datos integrados, orientados a una materia que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisión

- No son transitorios
- No son volátiles
- No se llevan a cabo modificaciones o eliminaciones, solo inserciones
- Guarda datos sumarios
- Orientados a una materia

### **Orientados a una materia:**

Organiza y orienta los datos en función del usuario final y sus temas de interés

Ejemplo

**Ventas, Competencias, Internaciones**

### **Datos integrados**

Los datos provienen de diferentes fuentes

La integración de datos se logra mediante la consistencia en la Convenciones de nombres, Unidades de

medida y codificación

### **Varían con el tiempo**

Mantiene tanto datos históricos como datos actuales

La información histórica es de gran importancia, permite analizar tendencias

- **Software de consultas**
- **Generadores de reportes**
- **Data mining**

**Metadatos:** Representan toda la información de administración y seguimiento necesarios para:

- **Acceso a datos**
- **Compresión y utilización**

### **Datamarts**

- Subconjuntos departamentales que focalizan objetos seleccionados
- Se caracteriza por una definición de requerimientos más rápida y fácil
- Pueden integrarse en un futuro en un DataWarehouse

**Data mining :**Extracción de información oculta y predecible de grandes bases de datos

- Predicción automatizada de tendencias y comportamientos
- Descubrimiento automatizado de modelos previamente desconocidos

### **Modelo Conceptual de un datawarehouse**

**Esquema de Hechos:** El esquema de DataWarehouse consiste en un conjunto de esquemas de hechos.  
Componentes:

**Hechos:** es un enfoque de interés para la empresa, Ej. : **ventas, competencias, internaciones**

**Dimensiones:** determina la granularidad para la determinación de los hechos. Ej. : **producto, fecha, almacén**

**Jerarquías:** Las dimensiones se asocian con sus jerarquías y especifican distintos niveles de agrupamiento

Hipercubo:

### **Operaciones:**

**Pivoting:** se rota el cubo para ver una cara en particular. Por Ej. : analizar informaciones referidas a un proveedor.

**Slicing Dicing:** se selecciona algún subconjunto del cubo. Analizar el cubo de datos restringiéndolo para algunos proveedores, productos y fechas

•

**Roll Up:** se agrupa por alguna dimensión determinada. Por Ej. : Analizar las ventas de producto a las ventas

por tipo de producto

**Drill Down:** Operación inversa: muestra información detallada de cada agrupamiento. Por Ej. Analizar las ventas de tipo de producto a las ventas por producto

### Implementaciones relacionales

**Esquema estrella:** Compuesto por una tabla central –tabla de hechos– y un conjunto de tablas mostradas en una forma radial alrededor de ésta –tablas dimensión

•

**Copo de nieve ó Pochoclo:** Extensión del esquema estrella, donde cada una de las tablas del esquema se divide en más tablas –tablas más normalizadas–

<b>Ventajas</b>	<b>Desventajas</b>
Más flexible a requerimientos	Puede agrandarse y ser inmanejable
Carga más rápida	Puede degradar la performance

### Modelo Constelación

#### Diseño conceptual de un DataWarehouse

Metodología semi automática para construir un modelo lógico de un DataWarehouse a partir de un Modelo Entidad Interrelación

Ejemplo de MER:

Transformación de una relación en entidad

Metodología :

- Definir los hechos
- Por cada hecho
- **Construir el árbol de atributos**
- **Recortar e injertar el árbol**
- **Definir dimensiones**
- **Definir atributos de hecho**
- **Definir jerarquias**
- Cada vértice corresponde a un atributo del esquema
- La raíz corresponde al identificador de F

#### Construir el árbol de atributos

Dada una porción de interés del MER y una entidad F que pertenece a él, denominamos árbol de atributos al árbol que:

#### Recortar e injertar el árbol

No todos los atributos representados en el árbol pueden ser de interés.

El árbol puede ser podado e injertado para eliminar detalles innecesarios

### Definir los hechos

Los hechos son conceptos de interés primario para realizar procesos de toma de decisiones

Un hecho puede ser representado en un MER mediante un entidad o una relación que representan archivos actualizables

### Definir dimensiones

Las dimensiones determinan cómo las instancias de hechos pueden ser agregadas para el proceso de la toma de decisiones

Deben ser elegidas entre los vértices del árbol, Ejemplo: Fecha, Producto, Almacén.

### Definir atributos de hecho

Son cantidades del número de instancias de hecho o suma/ promedio /máximo /mínimo de expresiones que involucran atributos numéricos del árbol de atributos

### Definir jerarquías

La jerarquías especifican distintos niveles de agrupamiento. El árbol ya muestra una organización jerárquica.

## **DATAMINING**

- Es la extracción de Información oculta y predecible de grandes base de datos. Las herramientas de datamining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomas decisiones proactivas. El usuario trata de obtener una relación de los datos que tengan repercusiones en su negocio.

### Los fundamentos de Dataminig:

- Recolección masiva de datos
- Potentes computadoras con multiprocesadores
- Algoritmos de datamining

### El Alcance de Dataminig

- Predicción automatizada de tendencias y comportamientos.
- Descubrimiento automatizado de modelos previamente desconocidos.

### Técnicas más comunes usadas en Dataminig

- Redes neuronales artificiales
- Árboles de decisión
- Algoritmos Genéticos



- Regla de inducción
- ANOVA (análisis de la varianza)
- Regresión
- Ji cuadrado
- Lógica Difusa
- Series Temporales

### Cómo Trabaja Datamining

Las computadoras son cargadas con mucha información donde una respuesta es conocida y luego El soft de datamining debe correr a través de los datos y distinguir Las características de los datos que llevarán al modelo. Una vez que El modelo se construyó, puede ser usado en situaciones similares donde no se conoce la respuesta.

### Introducción a OLAP

Los sistemas de soporte a la decisión usando tecnología de DW, se llaman sistemas OLAP. Estos OLAP deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos de diferentes perspectivas.
- Soportar análisis complejos contra un volumen ingente de datos.

OLAP: La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis de usuario y unas posibilidades de navegación, seleccionando la información a obtener (drill Down, Roll up, etc)

### **Arquitecturas para sistemas OLAP**

Existen dos tipos de arquitecturas:

**MOLAP** : OLAP multidimensional

**ROLAP** : OLAP relacional

**MOLAP:**

- La arquitectura MOLAP usa bases de datos multidimensionales para proporcionar el análisis.
- Un sistema MOLAP usa una base de datos multidimensión, en la que la información se almacena multidimensionalmente, para ser visualizada multidimensionalmente
- ROLAP
- La arquitectura ROLAP cree que Las capacidades OLAP están perfectamente implantadas sobre BD relacionales.
- Los usuarios finales ejecutan su análisis multidimensional a través Del motor ROLAP que transforma sus consultas a consultas de SQL. Estas consultas se ejecutan en BD relacionales y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios

## ROLAP vs. MOLAP

<b>ROLAP</b>	<b>MOLAP</b>
Muchas dimensiones	Diez o menos dimensiones.
Soportan análisis OLAP contra grandes volúmenes de datos	Se comportan razonablemente en volúmenes de datos más reducidos ( menos de 5Gb)
Herramienta flexible y general	Solución particular con volúmenes de información y número de dimensiones más modestos

## WEBHOUSING

Es tomar información proporcionada por Internet.

### **Ventajas:**

- Consistencia
- Accesibilidad
- Disponibilidad
- Bajos costos de desarrollo y mantenimiento.
- Protección de datos

### **Tipos de aplicaciones en las que se utilizan técnicas disponibles sobre DW**

ð Sistemas de marketing

ð Análisis de riesgo financiero

ð Análisis de riesgo de créditos

ð Otras áreas de aplicación

### **Áreas de aplicación**

ð Control de gestión: Presupuestación, reporting y análisis de desviaciones.

ð RRHH: Planificación de incorporaciones, gestión de carreras profesionales.

ð Logística: previsión de la demanda en infraestructura, optimización de los niveles de producción, mejora de la relación con proveedores.

### **Aspectos Técnicos en El proceso de creación y explotación de un DW**

ð Se pretende dar orientación al comprador para la preparación Del conjunto de especificaciones que definirán los requisitos que ha de cumplir la creación y explotación de un DW.

ð Análisis de las necesidades del comprador.

• Factores relevantes en el proceso de adquisición.

• Diseño Del pliego de prescripciones técnicas particulares.

### **Análisis de Las necesidades del comprador**

• Definición de los objetivos.

Se define el equipo de proyecto, alcance, funciones del DW, parámetros para evaluar el proyecto.

• Definición de los requerimientos de Información.

Entrevistas, se define estrategia y arquitectura de implementación del DW

### **Factores relevantes en el proceso de Adquisición**

• Pruebas en condiciones reales.

• Volumen y organización de datos

• Dimensionamiento de la plataforma de instalación

• Condiciones económicas y Del soporte

### **Diseño Del pliego de prestaciones Técnicas**

• Entorno Hardware:

• Host:(tipo máquina, sist.operativo y bd operacional)

• Servidor de aplicación Del DW

• Clientes:(memoria, tipo de máquina, etc.)

• Red local: (topología, protocolos, etc.)

• Entorno Software:

•

Gestor de BD para El DW.

• Volumen estimado de la BD

*Página 11 de 13*