

Asignatura: **Administración de Bases de Datos**

## **Tema 6:** **Técnicas de Backup y Recuperación de Bases de Datos**

Pedro P. Alarcón Cavero

pedrop.alarcon@eui.upm.es

Febrero 2011

## **Contenido**

- Introducción
- Operaciones de lectura/escritura
- Técnicas de Recuperación
- Backups
- Procedimientos de recuperación

### Contenido

Introducción

Operaciones

Téc. Recup.

Backups

Proc. Recup.



Contenido

**Introducción**

Operaciones

Téc. Recup.

Backups

Proc. Recup.

## Introducción

- El ABD debe estar preparado para reaccionar ante fallos o problemas que incidan en la disponibilidad, integridad y usabilidad de la base de datos
- Para ello el ABD debe tener bien definida una estrategia de copia y recuperación de la BD
- Los procesos de copia y recuperación (*backup and recovery*) contemplan varias estrategias y procedimientos para la protección de la BD frente a la pérdida de datos y la reconstrucción de la BD después de cualquier tipo de fallo o pérdida de datos

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 3



Contenido

**Introducción**

Operaciones

Téc. Recup.

Backups

Proc. Recup.

## Tipos de fallos o incidentes

- Fallos de instancia
  - Excepción interna del SGBD o fallo del SO, que requiere reiniciar el SGBD y/o el sistema
  - Generalmente no se dañan los datos. Pérdida del contenido de la memoria volátil
- Fallos de aplicación o transacción
  - Error lógico: entrada inválida, información no localizada, etc.
  - Error del sistema: error de programación, un interbloqueo, etc.
- Fallos en medios físicos de almacenamiento
  - Daño en disco, sistemas de ficheros, daño en cinta
  - Errores humanos: borrado o modificación accidental o incorrecta de datos (filas, tablas, tablespaces, ficheros de datos)
  - Se pierden datos o quedan dañados
- Desastres
  - Guerras, terrorismo, terremotos, inundaciones, incendios, robos
  - Interrupción prolongada de suministro eléctrico

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 4

Contenido

**Introducción**

Operaciones

Téc. Recup.

Backups

Proc. Recup.

## Efectos de los incidentes

- La interrupción de la ejecución de una transacción aislada
- La interrupción de una transacción ejecutándose concurrentemente con otras
- Puede hacer que terminen anormalmente varias
- Actualización incorrecta de datos
- Puede destruir físicamente la base de datos, eliminando muchas actualizaciones

Importante detectar y subsanar los fallos lo antes posible

Tras una recuperación, la información de la BD debe ser completa y consistente

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
5

Contenido

Introducción

**Operaciones**

Téc. Recup.

Backups

Proc. Recup.

## Operaciones básicas E/S

- Tipos de almacenamiento
  - **Almacenamiento volátil**
    - No sobrevive a las caídas del sistema
  - **Almacenamiento no volátil**
    - Disco, cinta, etc
    - Se producen accidentes
  - **Almacenamiento estable frente al no estable**
    - La información no se pierde “nunca”, se repite en varios medios no volátiles (disco) con modos de fallo independientes (por ejemplo utilizando discos RAID)

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
6

Contenido

Introducción

**Operaciones**

Téc. Recup.

Backups

Proc. Recup.

## Operaciones básicas E/S

- La BD reside en almacenamiento no volátil
- Bloque o página:
  - Unidad de transferencia de datos entre disco y memoria principal
- Transferencia de bloques entre el disco y la memoria principal
  - **input (X)**
    - Transfiere el bloque físico donde se encuentra X a la memoria principal
  - **output (X)**
    - Transfiere el bloque de registro intermedio (buffer) donde está X al disco, sustituyendo el bloque físico

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
7

Contenido

Introducción

**Operaciones**

Téc. Recup.

Backups

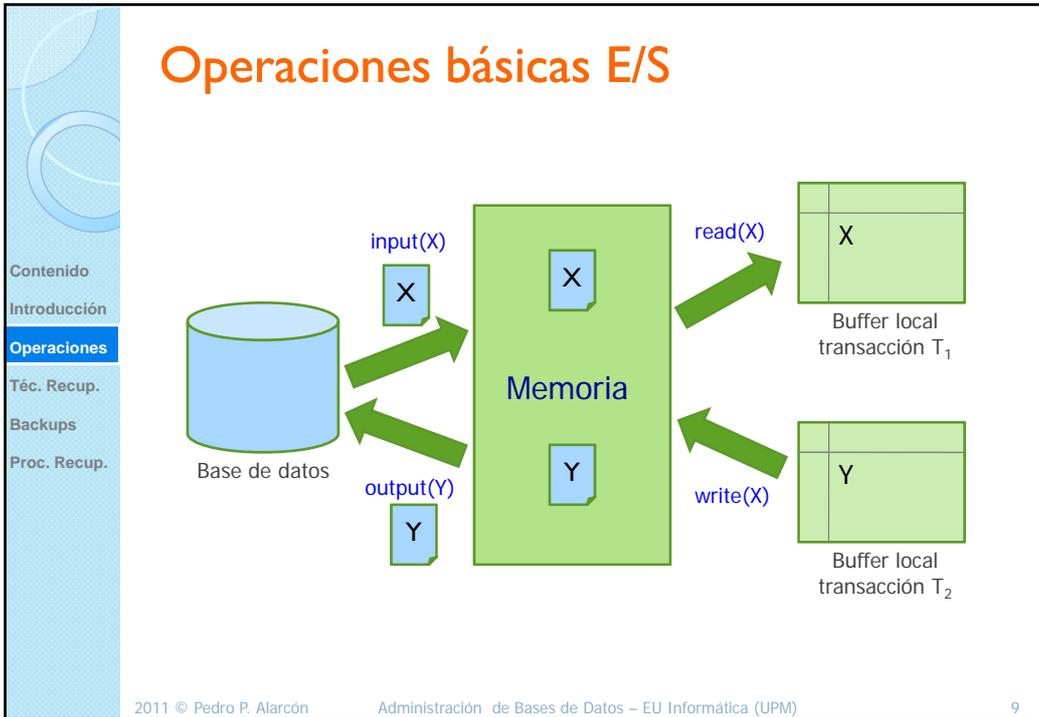
Proc. Recup.

## Operaciones básicas E/S

- Lectura/escritura entre transacciones y BD

<p><b>Read (X)</b></p> <ol style="list-style-type: none"> <li>1. Encontrar la dirección donde está X Bloque o página del disco</li> <li>2. Copiar el bloque del disco a un buffer de memoria, si no está ya en memoria</li> <li>3. Copiar X del buffer a la variable X del programa</li> </ol>	<p><b>Write (X)</b></p> <ol style="list-style-type: none"> <li>1. Encontrar la dirección donde está X Bloque o página del disco</li> <li>2. Copiar el bloque del disco a un buffer de memoria, si no está ya en memoria</li> <li>3. Copiar X de la variable del programa a su posición en el buffer</li> <li>4. Almacenar el bloque actualizado desde el buffer al disco</li> </ol>
--	---

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
8



- ## Operaciones básicas E/S
- **Gestión en memoria y en disco**
    - Un bloque del buffer se graba en disco
      - porque el gestor de buffer necesita espacio de memoria para otros propósitos
      - o porque el SGBD desea reflejar el cambio hecho a X en el disco
    - Si el sistema se cae tras ejecutar write (X, xi), pero antes de ejecutar output (X), el nuevo valor de X se pierde, no llega a escribirse en disco
- 2011 © Pedro P. Alarcón      Administración de Bases de Datos – EU Informática (UPM)      10

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Técnicas de recuperación

- **Objetivos**
  - Devolver a la BD a un estado consistente, con la menor pérdida de información y tiempo posible, incluyendo
    - Acciones durante el proceso normal de transacciones
    - Acciones después de un fallo
- **Los accesos a la BD se realizan a través de transacciones**
  - Recuperar implica poder repetir una a una las actualizaciones que se han ejecutado sobre la BD
  - Para ello es necesario que las transacciones se ejecuten según un esquema de serialización
- **Técnicas**
  - Diarios basados en actualizaciones diferidas e inmediatas
  - Doble paginación

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
11

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Diarios para recuperación

- Se utilizan también los términos *log* y *journal*
- Mantiene un registro de todas las operaciones que afectan a elementos de la base de datos
- Esta información permite recuperar la BD
- Tipos
  - Diarios basados en actualizaciones diferidas (After Image Log)
  - Diarios basados en actualizaciones inmediatas (Before Image Log)
- Se almacena en disco
- Registros posibles a reflejar
  - $T_i$ , start
  - $T_i$ , write, X, *valor\_viejo*, *valor\_nuevo*
  - $T_i$ , read, X opcional
  - $T_i$ , commit
  - $T_i$ , abort
- Operaciones sobre el diario
  - undo
  - redo

```

graph TD
    Memoria[Memoria (SGBD)] <--> Diario[(Diario)]
    Memoria <--> Base[Base de datos]
    Diario --> Base
  
```

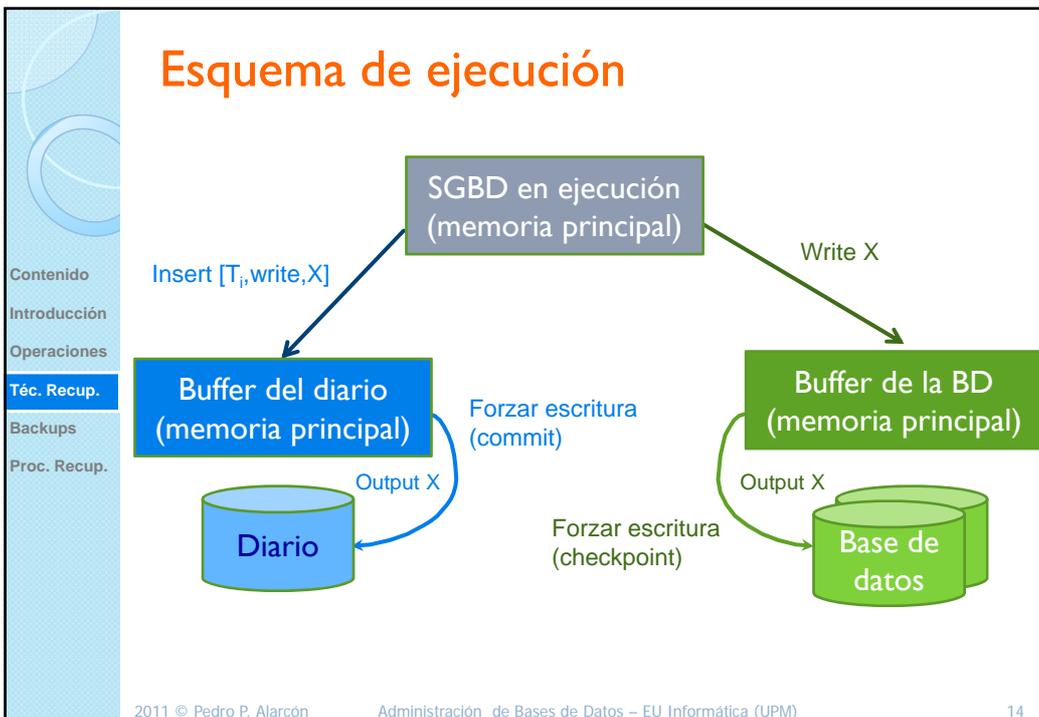
2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
12

Contenido  
Introducción  
Operaciones  
**Téc. Recup.**  
Backups  
Proc. Recup.

## Checkpoints o puntos de validación

- Los puntos de *checkpoint* en el diario garantizan que la información de las transacciones que han terminado antes de ese punto está en la base de datos
- Un *checkpoint* consiste en
  - Suspender la ejecución de las transacciones temporalmente
  - Forzar la escritura de todas las actualizaciones de buffer a disco
  - Escribir registro *checkpoint* en el diario
  - Reanudar la ejecución de las transacciones

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 13





Contenido  
Introducción  
Operaciones  
**Téc. Recup.**  
Backups  
Proc. Recup.

## Diarios con actualizaciones diferidas

- Diario con imágenes posteriores (After Image Log)
- Graba todas las actualizaciones de la BD en el diario, pero aplaza la ejecución de todas las operaciones de escritura (write) de una transacción hasta que ésta se encuentre parcialmente cometida
- Solamente requiere el nuevo valor del dato
- Si la transacción aborta (no llega a *committed*), simplemente hay que ignorar las anotaciones en el diario
- Para recuperaciones usa el procedimiento
  - **redo** (Ti), que asigna los nuevos valores a todos los datos que actualiza Ti

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 15



Contenido  
Introducción  
Operaciones  
**Téc. Recup.**  
Backups  
Proc. Recup.

## Diarios con actualizaciones diferidas

- Después de ocurrir un fallo, se consulta el diario para determinar que transacciones deben repetirse y cuales anularse
  - Ti debe anularse si el diario contiene el registro *start* pero no el *commit*
  - Ti debe repetirse si el diario contiene el registro *start* y el *commit*
- La operación *redo* debe ser idempotente, es decir, ejecutarla varias veces debe producir el mismo resultado que ejecutarla una sola vez

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 16

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Diarios con actualizaciones diferidas

- Ejemplo

Transacción T<sub>1</sub>

```
UPDATE tabla1
SET A = 5000
WHERE cond_a;

UPDATE tabla1
SET B = 25000
WHERE cond_b;
```

Transacción T<sub>2</sub>

```
UPDATE tabla2
SET C = 9000
WHERE cond_c;
```

T<sub>1</sub> y T<sub>2</sub> no comparten un mismo gránulo

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
17

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Diarios con actualizaciones diferidas

- Ejemplo de recuperación

Valores iniciales en BD: A=10000; B=20000; C=7000

Transacciones	Diario	Base de Datos
<p>T1</p> <pre style="margin: 0;">UPDATE tabla1 SET A=5000 WHERE cond_a; UPDATE tabla1 SET B=25000 WHERE cond_b; COMMIT;</pre>	<pre style="margin: 0;">&lt;T1 start&gt; &lt;T1,A, 5000&gt; &lt;T1,B, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2,C, 9000&gt; &lt;T2 commit&gt;</pre>	<pre style="margin: 0;">A = 5000 B = 25000  C = 9000</pre>
<p>T2</p> <pre style="margin: 0;">UPDATE tabla2 SET C=9000 WHERE cond_c; COMMIT;</pre>		

¡Base de datos puede significar memoria o disco!

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)

## Diarios con actualizaciones diferidas

- Ejemplo de recuperación

Valores iniciales en BD: A=10000; B=20000; C=7000

<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1,A, 5000&gt; &lt;T1, B, 25000&gt; ----- caída ----- </pre> <p style="background-color: #FFD700; padding: 2px;">A=10000;B=20000</p> <p style="text-align: center;"><b>No hay que hacer nada</b></p> <div style="border: 1px solid #003366; border-radius: 50%; width: 50px; height: 50px; margin: 0 auto; display: flex; align-items: center; justify-content: center; text-align: center;"> <p style="font-size: 8px; margin: 0;">A = 10000 B = 20000 C = 7000</p> </div>	<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1,A, 5000&gt; &lt;T1, B, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2, C, 9000&gt; ----- caída ----- </pre> <p style="background-color: #FFD700; padding: 2px;">A=5000;B=25000;C=7000</p> <p style="text-align: center;"><b>Redo (T1)</b></p> <div style="border: 1px solid #003366; border-radius: 50%; width: 50px; height: 50px; margin: 0 auto; display: flex; align-items: center; justify-content: center; text-align: center;"> <p style="font-size: 8px; margin: 0;">A = 5000 B = 25000 C = 7000</p> </div>	<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1,A, 5000&gt; &lt;T1, B, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2, C, 9000&gt; &lt;T2 commit&gt; ----- caída ----- </pre> <p style="background-color: #FFD700; padding: 2px;">A=5000;B=25000;C=9000</p> <p style="text-align: center;"><b>Redo (T1) Redo (T2)</b></p> <div style="border: 1px solid #003366; border-radius: 50%; width: 50px; height: 50px; margin: 0 auto; display: flex; align-items: center; justify-content: center; text-align: center;"> <p style="font-size: 8px; margin: 0;">A = 5000 B = 25000 C = 9000</p> </div>
---	---	---

Los cambios pueden estar en disco o en memoria

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
19

## Diarios con actualizaciones inmediatas

- Diario con imágenes anteriores (Before Image Log)
- Permite que las actualizaciones se graben en la BD mientras la transacción está todavía en estado activo (actualizaciones no cometidas)
- Antes de ejecutar un *output(X)*, deben grabarse en memoria estable los registros del diario correspondientes a X
- Los registros del diario deben contener tanto el valor antiguo como el nuevo
- El esquema de recuperación utiliza
  - **undo** (Ti): restaura los datos que Ti actualizó a los valores que tenían antes
  - **redo** (Ti): asigna los nuevos valores a todos los datos que actualizó Ti

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
20

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Diarios con actualizaciones inmediatas

- Después de ocurrir un fallo, el procedimiento de recuperación consulta el diario para determinar qué transacciones deben repetirse y cuáles deshacerse
  - Ti debe deshacerse si el diario contiene el registro *start* pero no el *commit*
  - Ti debe repetirse si el diario contiene el registro *start* y el *commit*
- Las operaciones *undo* y *redo* deben ser idempotentes para garantizar la consistencia de la BD aun cuando se produzcan fallos durante el proceso de recuperación

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
21

Contenido

Introducción

Operaciones

**Téc. Recup.**

Backups

Proc. Recup.

## Diario con actualizaciones inmediatas

- Ejemplo de recuperación

Valores iniciales en BD: A=10000; B=20000; C=7000

Transacciones	Diario	Base de Datos
<p><b>T1</b></p> <pre>UPDATE tabla1 SET A=5000 WHERE cond_a; UPDATE tabla1 SET B=25000 WHERE cond_b; COMMIT;</pre>	<pre>&lt;T1 start&gt; &lt;T1, A, 10000, 5000&gt; &lt;T1, B, 20000, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2, C, 7000, 9000&gt; &lt;T2 commit&gt;</pre>	<pre>A = 5000 B = 25000 C = 9000</pre>
<p><b>T2</b></p> <pre>UPDATE tabla2 SET C=9000 WHERE cond_c; COMMIT;</pre>		

¡Base de datos puede significar memoria o disco!

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)

## Diarios con actualizaciones inmediatas

- Ejemplo de recuperación

Valores iniciales en BD: A=10000; B=20000; C=7000

<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1, A, 10000, 5000&gt; &lt;T1, B, 20000, 25000&gt; ----- caída -----</pre> <p style="text-align: center; background-color: #FFD700; padding: 2px;">A=5000; B=25000</p> <p style="text-align: center;"><b>Undo (T1)</b></p> <div style="text-align: center; border: 1px solid black; border-radius: 50%; width: 60px; height: 60px; margin: 0 auto; display: flex; align-items: center; justify-content: center;"> <div style="text-align: left; padding: 2px;"> A = 10000  B = 20000  C = 7000 </div> </div>	<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1, A, 10000, 5000&gt; &lt;T1, B, 20000, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2, C, 7000, 9000&gt; ----- caída -----</pre> <p style="text-align: center; background-color: #FFD700; padding: 2px;">A=5000; B=25000; C=9000</p> <p style="text-align: center;"><b>Undo (T2) Redo (T1)</b></p> <div style="text-align: center; border: 1px solid black; border-radius: 50%; width: 60px; height: 60px; margin: 0 auto; display: flex; align-items: center; justify-content: center;"> <div style="text-align: left; padding: 2px;"> A = 5000  B = 25000  C = 7000 </div> </div>	<pre style="background-color: #003366; color: white; padding: 5px;"> &lt;T1 start&gt; &lt;T1, A, 10000, 5000&gt; &lt;T1, B, 20000, 25000&gt; &lt;T1 commit&gt; &lt;T2 start&gt; &lt;T2, C, 7000, 9000&gt; &lt;T2 commit&gt; ----- caída -----</pre> <p style="text-align: center; background-color: #FFD700; padding: 2px;">A=5000; B=25000; C=9000</p> <p style="text-align: center;"><b>Redo (T1) Redo (T2)</b></p> <div style="text-align: center; border: 1px solid black; border-radius: 50%; width: 60px; height: 60px; margin: 0 auto; display: flex; align-items: center; justify-content: center;"> <div style="text-align: left; padding: 2px;"> A = 5000  B = 25000  C = 9000 </div> </div>
--	--	--

Los cambios pueden estar en disco o en memoria

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
23

## Punto de validación

- Recuperación hasta un punto de validación

1. El SGBD examina el diario hacia atrás hasta localizar un registro *<checkpoint>*
2. Considera sólo los registros existentes entre este punto y el final del diario
3. Ejecuta *undo(Tj)* para las transacciones que no tengan registro *<Tj commit>*, partiendo del final del fichero
4. Ejecuta *redo(Ti)* para las transacciones que tengan su registro *<Ti commit>*, partiendo desde el punto de verificación hasta el final del diario

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
24

## Utilización de diarios

- Criterios del ABD
  - La utilización de un tipo u otro de diario depende de la instalación concreta que se tenga
  - Factores
    - Número de recuperaciones
    - Nivel de concurrencia y granularidad
  - Si abortan pocas transacciones, las imágenes anteriores son adecuadas
  - Si abortan muchas transacciones las imágenes posteriores resultan más eficientes
  - Si hay muchas actualizaciones sobre los mismos gránulos la actualización inmediata es menos eficiente y las imágenes posteriores resultarán mejor

Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

## Retrocesos en cascada

- El fallo en una transacción puede suponer retrocesos en cascada de transacciones que hayan leído datos escritos por una transacción que falló
- No son deseables, ya que llevan a deshacer una importante cantidad de trabajo
  - Se pueden evitar, bajo el bloqueo en dos fases
  - Los algoritmos de ordenamiento inicial se pueden modificar para evitar el retroceso, pero se introducen esperas (aunque no bloqueos)

Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

Contenido

Introducción

Operaciones

**Téc. Recup.**

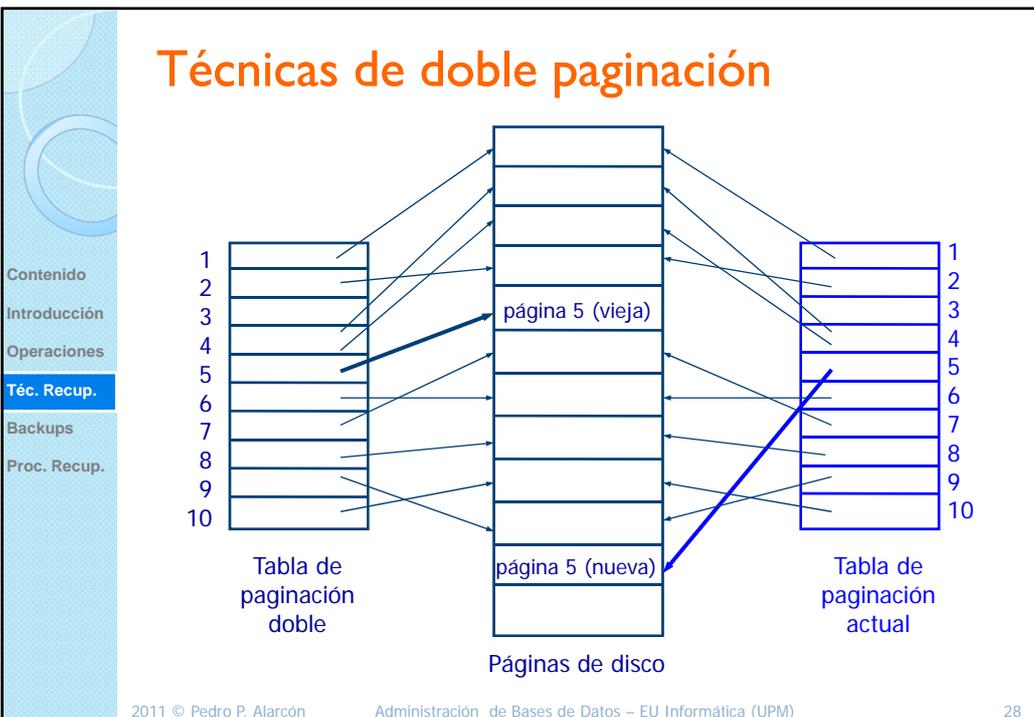
Backups

Proc. Recup.

## Técnicas de doble paginación

- Alternativa a las técnicas de recuperación basadas en diarios
- El sistema mantiene dos tablas de paginación durante la vida de una transacción, y son idénticas al comenzar la transacción
  - **Tabla de paginación actual**
    - Puede variar cuando la transacción realiza una operación *write*
    - Todas las operaciones *input* y *output* utilizan esta tabla para localizar las páginas de la BD
    - Puede almacenarse en memoria volátil
  - **Tabla de paginación doble**
    - No se modifica, y debe almacenarse en memoria no volátil

2011 © Pedro P. Alarcón
Administración de Bases de Datos – EU Informática (UPM)
27





Contenido  
Introducción  
Operaciones  
**Téc. Recup.**  
Backups  
Proc. Recup.

## Técnicas de doble paginación

- **Commitment**
  1. Comprobar que todas las páginas del buffer que haya modificado la transacción se graban en disco
  2. Grabar en disco la tabla de paginación actual
  3. Grabar la dirección en disco de la tabla de paginación actual en la posición fija de memoria estable que contenga la dirección de la tabla de paginación doble. Por tanto, la tabla de paginación actual se convierte en la tabla de paginación doble y la transacción está cometida
- **No es necesario aplicar procedimiento de recuperación**
  - Si ocurre una caída antes del paso 3 no hay que deshacer nada
  - Si ocurre una caída después del paso 3, no hay que rehacer nada

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 29



Contenido  
Introducción  
Operaciones  
**Téc. Recup.**  
Backups  
Proc. Recup.

## Técnicas de doble paginación

- **Ventajas frente a los diarios**
  - No es necesario aplicar ningún procedimiento de recuperación
  - Se elimina el tiempo para grabar registros
  - La recuperación de las caídas es más rápida
- **Desventajas**
  - Fragmentación de los datos
  - Recolección de basura
  - La doble paginación es más difícil de adaptar que un diario a los sistemas que permiten ejecución concurrente de transacciones

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 30



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backups de la base de datos

- Un backup de base de datos implica disponer de copias consistentes de los datos
- Es común en muchas organizaciones manejar de 1 a 10 terabytes en un único servidor de BD
- El tiempo de backup/recuperación puede llegar a ser muy elevado
  - Los backups se suelen almacenar en cintas
  - Ejemplo
    - Velocidad de almacenamiento en cinta DLT de 80Mb/s
    - Backup de una BD de 10Tb puede tardar casi 40 horas
- Con el incremento del volumen de datos
  - Aumenta la probabilidad de corrupción de datos físicos
  - Aumenta la frecuencia de errores humanos

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 31



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backups de la base de datos

- Nivel: base de datos, tablespace, tabla
- Índices ¿reconstruir o recuperar?
  - En tablas grandes, hacer backup de índices ralentiza la copia pero acelera la recuperación
  - En tablas pequeñas puede ser más interesante reconstruir
- Control del SGBD
  - Algunos SGBD almacenan en el catálogo información sobre los backups y recuperaciones
- Aspectos de acceso concurrente
  - *Backup off-line* o *cold backup*
    - Para realizar el backup se hace un *shutdown* de la BD (no está operativa)
  - *Backup on-line* o *hot backup*
    - El backup se realiza mientras la BD está operativa
    - Mayor complejidad y mayor sobrecarga: CPU, E/S adicional

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 32



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backup Físico

- Copia de los ficheros físicos utilizados en almacenamiento y recuperación de la BD
  - Ficheros de datos
  - Ficheros de control
  - Ficheros log
- Almacenado en otros discos o en medios de almacenamiento off-line como cintas
- Base de cualquier buen plan de copia y recuperación
- Tras realizar un backup de la BD se suelen *vaciar* los ficheros log

2011 © Pedro P. Alarcón      Administración de Bases de Datos – EU Informática (UPM)      33



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backup Lógico

- Se obtiene copia solo de los datos y no de los ficheros físicos enteros
- Utilidades
  - *export/import* : ficheros binarios
  - *unload/load*: ficheros ascii
- Se realiza durante la operación normal de la BD
- Útil en los siguientes casos
  - Recuperar objetos o filas
    - Derivado de un borrado accidental de algunas filas. Más rápido que aplicar la recuperación del backup completo
  - Actualización de *releases* del SGBD
  - Migración de datos a SGBD heterogéneos
  - Movimiento de datos en una misma organización
    - Diferentes SGBDs, hojas de cálculo, etc.
- Permiten complementar la estrategia de backups físicos

2011 © Pedro P. Alarcón      Administración de Bases de Datos – EU Informática (UPM)      34



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backup completo vs. incremental

- Backup completo vs. incremental
  - Backup completo
    - Copia completa de la BD en un instante dado
  - Backup incremental (o diferencial)
    - Solo los datos modificados desde la última copia completa o incremental de la BD
    - Más rápido y menos espacio que un backup completo
    - Recuperación más lenta, algunas tuplas han podido variar varias veces desde la copia completa
    - Algunas sistemas permiten mezclar backups incrementales
  - Elección del ABD
    - Si varían pocos datos: incremental
    - Si bloques de datos entre backups varían >30%-40%: completo
    - Factores: volatilidad, criticidad, disponibilidad

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 35



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backup de la base de datos

- El ABD debe
  - Mantener backups con copias actualizadas y exactas de los datos
  - Decidir el tipo de backup a realizar (datos, logs, completo, parcial, etc)
  - Establecer la frecuencia de realización de backups
  - Decidir el número de generaciones de backups anteriores a guardar
  - Documentar los procedimientos de copia y recuperación
  - Comprobar el índice de llenado de los ficheros log, y copiar/vaciar en su caso

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 36



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
**Backups**  
Proc. Recup.

## Backup de la base de datos

- Guías para realizar backups
  - Mantener al menos dos copias locales por backup
  - Coordinar copia local y copia externa
  - Guardar al menos dos generaciones de copia del backup
  - Crear el backup en disco y luego copiarlo a cinta, y considerar tener comprimidas las copias de cinta
  - Incluir el catálogo de la BD en el backup
  - Utilizar procesos de backup que admitan *re-start*
  - Verificar la corrección del backup generado
  - Hacer copia también de datos no almacenados en la BD pero que son utilizados por las aplicaciones
  - Una decisión prudente es tener una copia completa de cada objeto de la base de datos
  - Hacer copia tras la carga o reorganización de la BD

2011 © Pedro P. Alarcón      Administración de Bases de Datos – EU Informática (UPM)      37



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
**Proc. Recup.**

## Procedimientos de recuperación

- La recuperación puede ser una tarea compleja
- Las operaciones de copia y recuperación de la BD deben planificarse conjuntamente
- Deseable aplicar medidas preventivas para minimizar recuperaciones ante fallos
  - Sistemas de alimentación ininterrumpida (SAI/UPS)
  - Discos espejos (por ejemplo, discos RAID)
  - Tecnología *Failover*
    - un segundo equipo (replicado) preparado para entrar en funcionamiento
- Tipos
  - Recuperación normal
  - Recuperación en caliente
  - Recuperación en frío
- El tipo de recuperación a aplicar dependerá de la naturaleza del fallo

2011 © Pedro P. Alarcón      Administración de Bases de Datos – EU Informática (UPM)      38



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

## Proceso de recuperación

- Pasos habituales
  - Identificar el fallo
  - Analizar la situación
  - Determinar qué necesita ser recuperado
  - Identificar dependencias entre los objetos de la BD a ser recuperados
  - Localizar las copias de backups requeridas
  - Restaurar la copia del backup
  - Aplicar los diarios correspondientes

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 39



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

## Recuperación normal

- Tiene lugar después de una parada normal de la máquina, en la que se escribe un punto de verificación como último registro del diario
- Este procedimiento se ejecuta cuando el último registro del diario es un punto de verificación o recuperación del sistema
- Este tipo de recuperación también tiene lugar cuando aborta una transacción, debido a la razón que sea
- No requiere acción concreta por parte del ABD

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 40



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

## Recuperación en caliente

- Después de un error del sistema
- Se ejecuta cuando el último registro del diario no es un punto de verificación y el operador no indica pérdida de memoria secundaria
- El procedimiento de recuperación es el indicado en el apartado referente a los puntos de verificación en el diario
- Puede realizarse automáticamente o solicitar acción por parte del operador o ABD

2011 © Pedro P. Alarcón Administración de Bases de Datos – EU Informática (UPM) 41



Contenido  
Introducción  
Operaciones  
Téc. Recup.  
Backups  
Proc. Recup.

## Recuperación en caliente

- Factores que determinan la duración de una recuperación en caliente
  - Número de registros log que deben procesarse en la recuperación
  - Si el log está compactado o comprimido
  - Tiempo empleado por el operador en montar y desmontar las cintas requeridas
  - Tiempo empleado por el SGBD en leer la parte del log necesario para la recuperación
  - Tiempo necesario para reprocesar las páginas cambiadas por parte del SGBD

2011 © Pedro P. Alarcón Administración de Bases de Datos – EUIT Informática (UPM) 42

## Recuperación en frío

- Se realiza si se pierden datos o la BD ya no es coherente
- Utiliza
  - Copia de seguridad (backup) más reciente de la BD
  - Diario de las actividades posteriores
  - Se aplican las imágenes posteriores al respaldo
- Puede encadenar una recuperación en caliente
- Factores que determinan la duración de una recuperación en frío
  - Fallo/sustitución de hardware
  - Instalación SGBD y/o creación objetos de la BD
  - Volumen de datos del backup
  - Tiempo empleado por el operador en el proceso
  - Documentación del proceso

## Recuperación en frío

