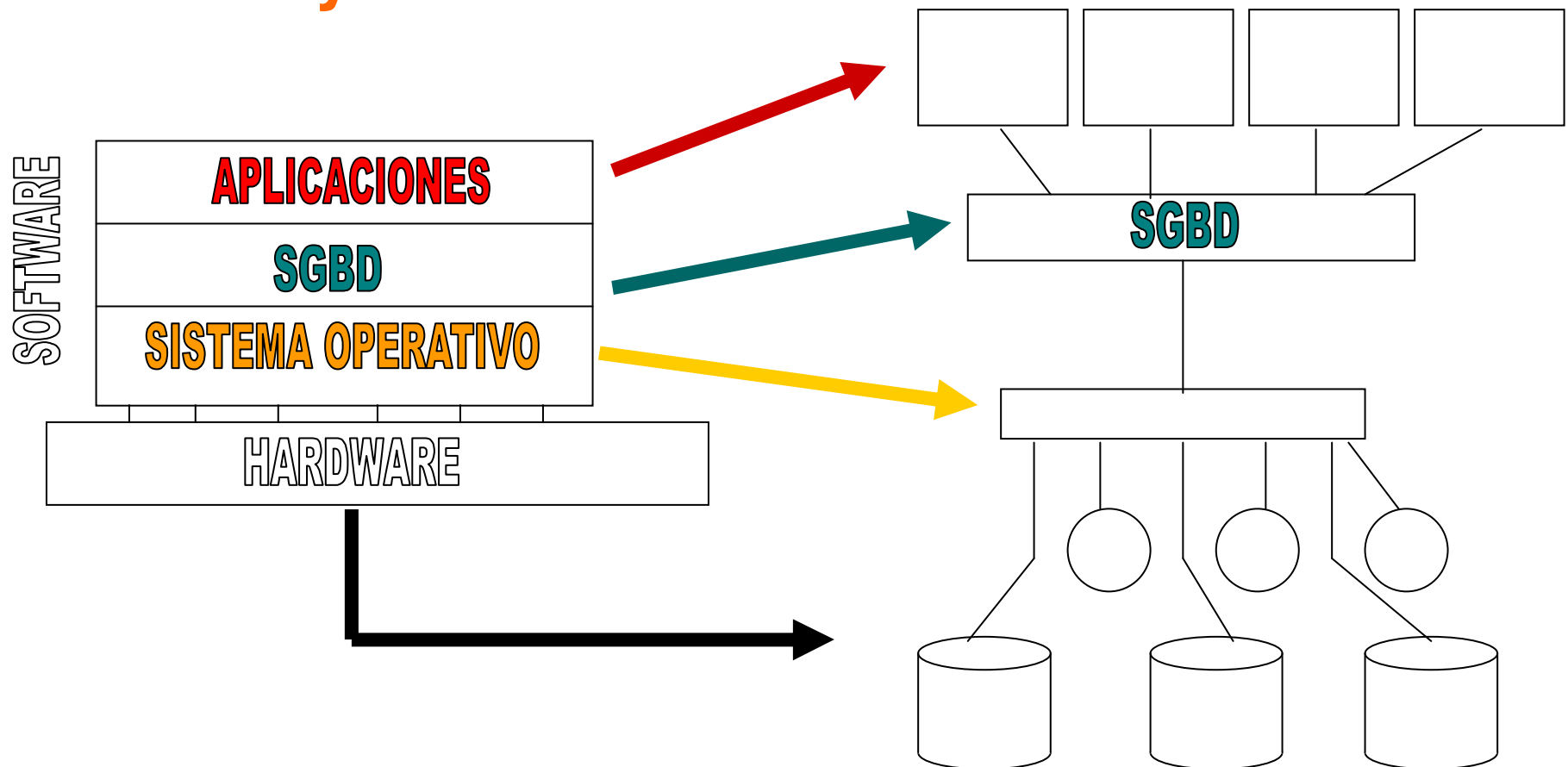


# Manejo de Datos Almacenados

### Contexto de los SGBD: Capas de *software y hardware*.

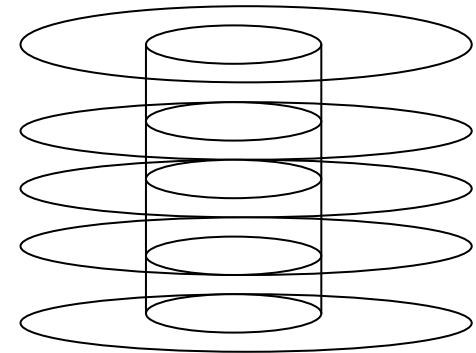


# Almacenamiento de Datos

## 1. DISCOS

Pueden recuperar cualquier página de forma aleatoria y al mismo costo (medido en operaciones de I/O):

*Pero leer páginas consecutivas es menos costoso que leerlas en forma aleatoria.*



# Almacenamiento de Datos

## 2. CINTAS

Son menos costosas que los discos.

*Pero sólo son prácticas para guardar datos históricos.*



# Archivos

- Una página o bloque es la unidad de transferencia entre memoria principal y memoria secundaria.
- Archivos: una colección de páginas, cada una conteniendo un conjunto de registros.

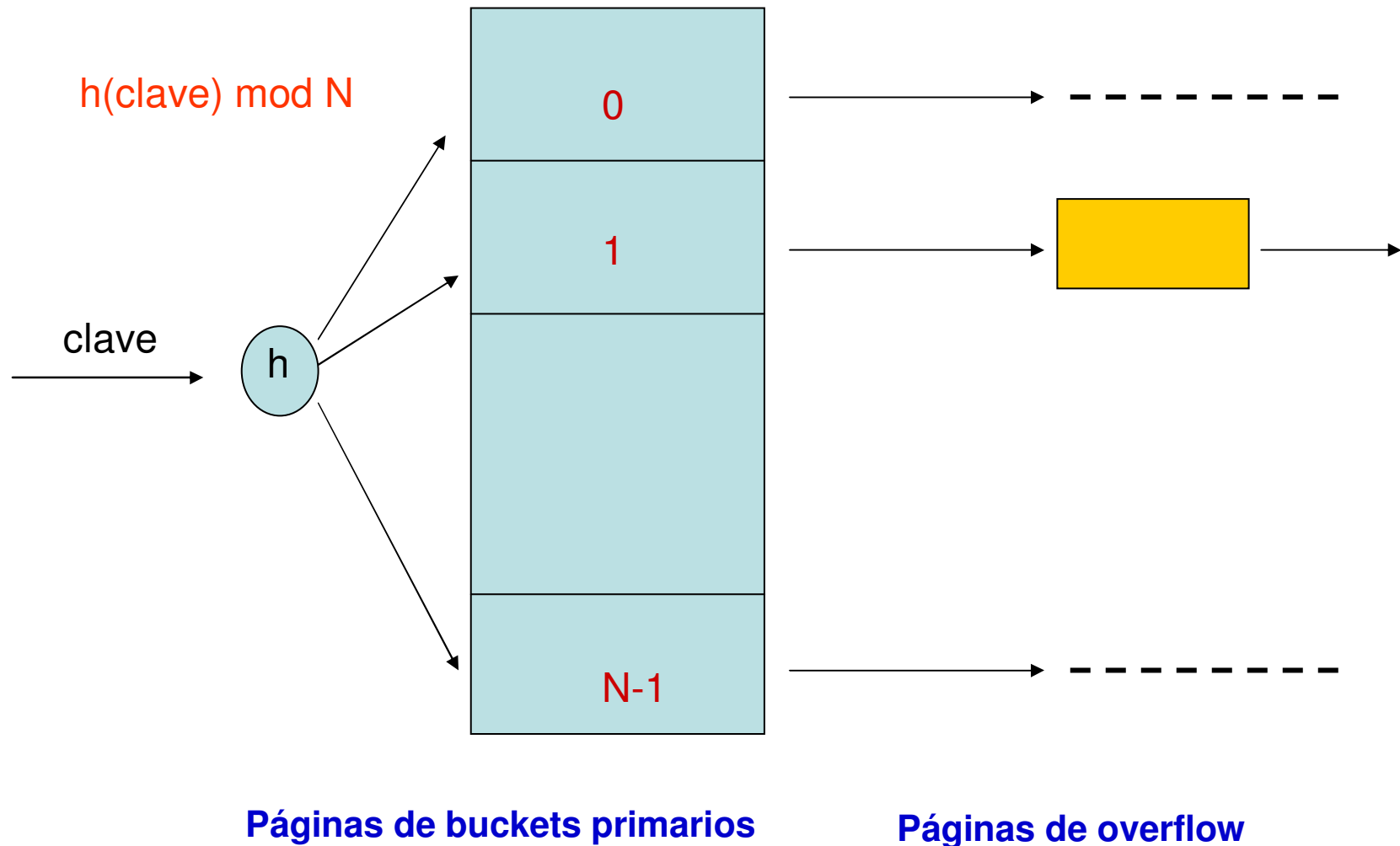
Debe permitir:

- insertar/borrar/modificar un registro
- Leer un registro particular(especificando *record id*)
- scan todos los registros.

## Tipos de Archivos

- Heap files: registros almacenados en el orden en que son insertados.
- Archivos Ordenados: los registros son ordenados haciendo uso de alguna clave.
- Archivos Hashed:
  - Colección de buckets.  
Bucket = *página primaria más cero o más páginas de overflow.*
  - *Función Hashing  $h$* :  $h(r) = b$ , donde  $b$  es el bucket donde se encuentra  $r$ .

## Hashing Estático



# Modelo de Costo

Como una buena aproximación se ignora el costo de procesamiento en memoria principal:

- **B**: Número de páginas.
- **R**: Número de registros por página.
- **D**: tiempo promedio de leer un bloque.
- Se ignora el beneficio de tener pre-cargadas las páginas en memoria.



# Operaciones para comparar las organizaciones de archivos

- **SCAN:** obtener todos los registros de un archivo. Las páginas deben ir del disco al buffer. Hay un overhead de CPU por registro para localizar el registro en la página.
- **Búsqueda por Igualdad:** obtener todos los registros que satisfacen una selección de igualdad. Los registros que cumplen con la condición son colocados en las páginas de salida.

# Operaciones para comparar las organizaciones de archivos

- **Búsqueda por Rango:** obtener los registros que cumplen la selección en un rango. “Encontrar todos los alumnos mayores de 20 años y menores de 40”.
- **Insertar:** Insertar un registro dado en el archivo. Se identifica la página donde debe insertarse el registro, se trae la página desde el disco, se modifica y se vuelve a escribir en el disco.

# Operaciones para comparar las organizaciones de archivos

- **Eliminar:** se usa el rid (record id) para eliminar un registro de un archivo . Se identifica la página donde está el registro, se lleva la página al disco, se modifica y se vuelve a escribir.

# Modelo de Costo

- **Heap Files:**
  - Las inserciones siempre se hacen al final.
- **Archivos Ordenados:**
  - El archivo se debe compactar después de eliminaciones.
  - Selecciones en el campo de ordenamiento.
- **Archivos Hashed:**
  - Se asume que no hay ningún buckets de overflow.
  - 80% de ocupación, es decir, el número de páginas requerido para almacenar el archivo es 1.25B.

# Costo de Ocupación

	Heap File	Archivo Ordenado	Archivo Hashed
Scan todos regs	<b>BD</b>	<b>BD</b>	<b>1.25 BD</b>
Búsqueda por Igualdad	<b>0.5 BD</b>	<b><math>D \log_2 B</math></b>	<b>D</b>
Búsqueda por Rango	<b>BD</b>	<b><math>D (\log_2 B + \# \text{ de } \text{páginas que hacen match})</math></b>	<b>1.25 BD</b>
Inserción	<b>2D</b>	<b>Búsqueda + BD</b>	<b>función h + 2D</b>
Eliminación	<b>Búsqueda + D</b>	<b>Búsqueda + BD</b>	<b>Búsqueda + 2D</b>

## Fuentes consultadas:

**[1] Ramakrishnan Raghu. ,**  
“Database Management Systems”.

**[2] Prof. Elsa Liliana Tovar.**  
Notas de clase compiladas entre 1997-2004.