

# Cátedra Bases de Datos

## Revisión de Primera Clase

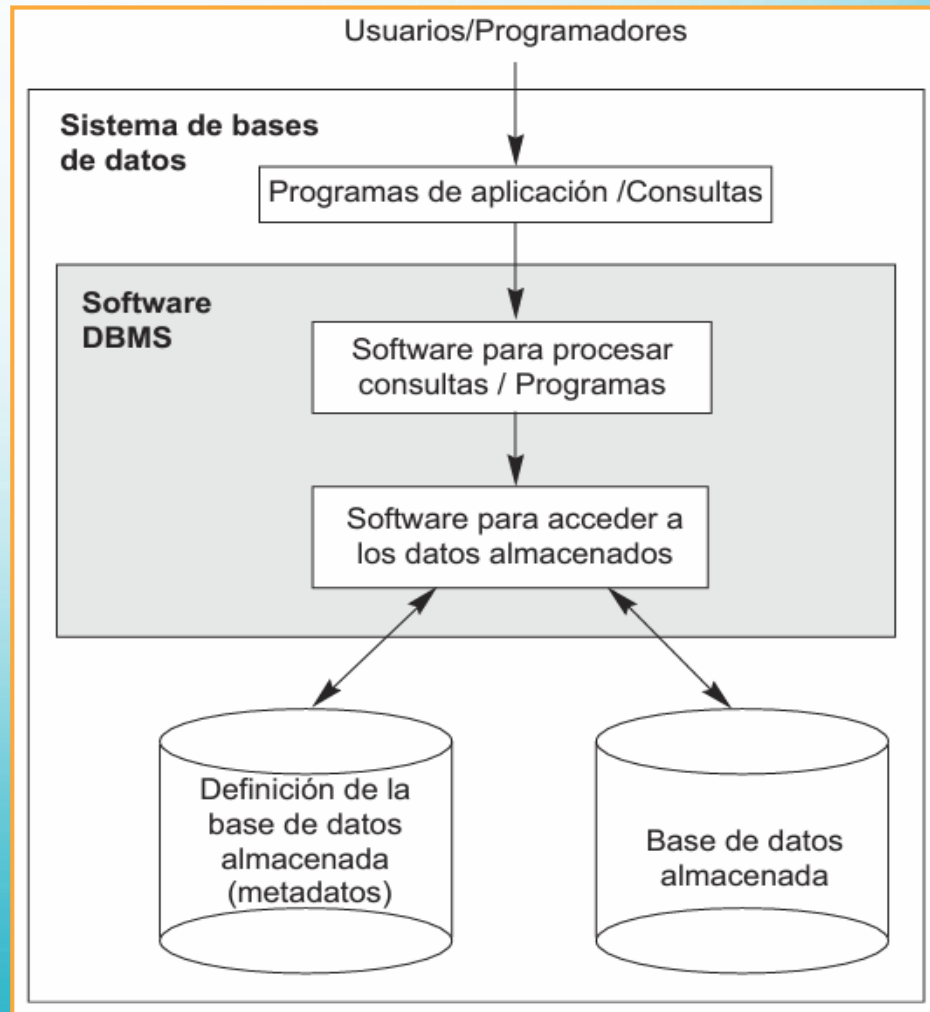
### Concepto BD

**DBMS** permite:

1. DDL
2. DML
3. DCL
4. Integridad
5. Respaldo y recuperación
6. Transacciones

### Componentes del Sistema de BD

#### Usuarios



# Cátedra Bas

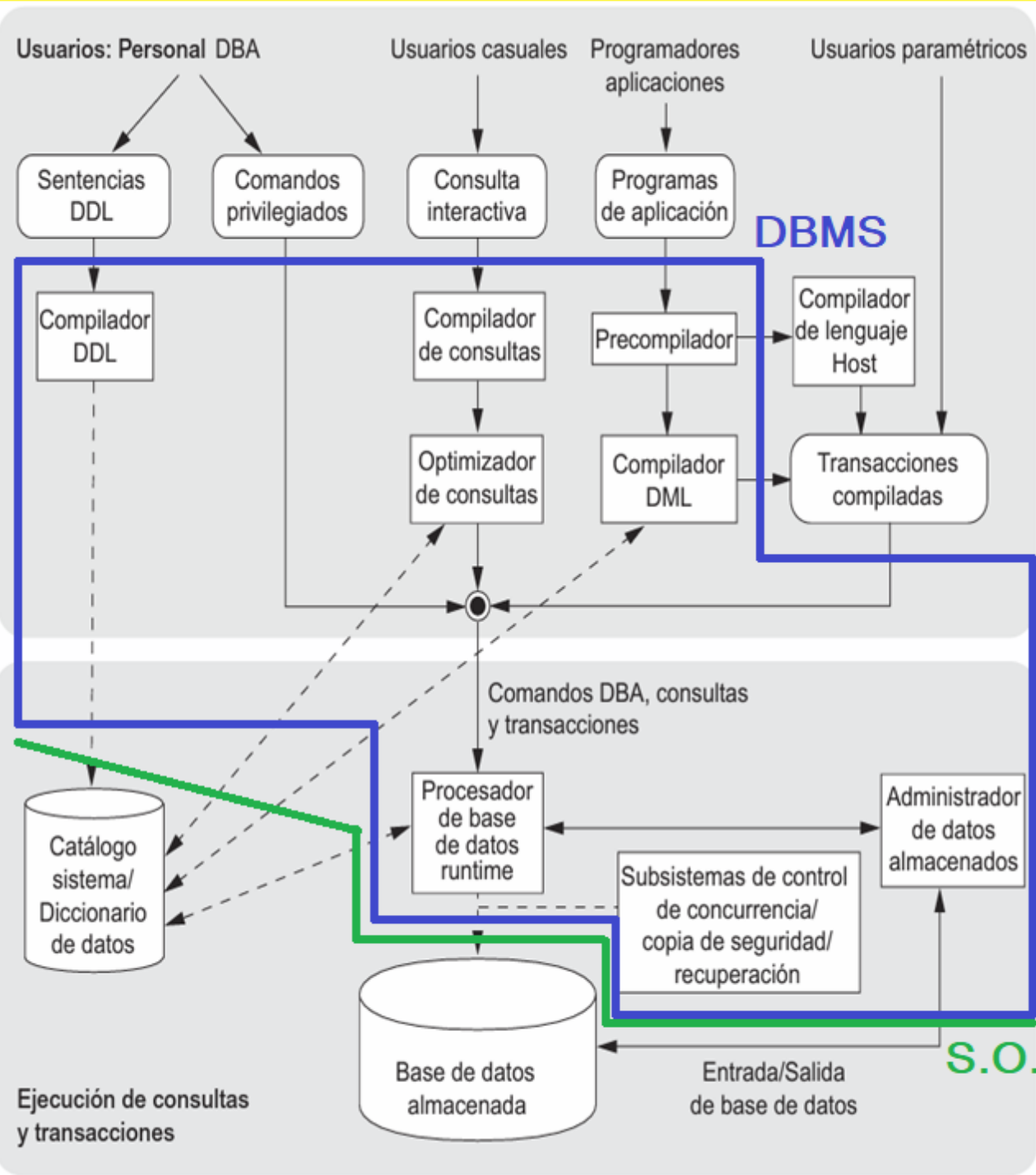
## Componentes del DBMS

**Compilador DDL:** procesa las definiciones de los esquemas y almacena las descripciones de los esquemas en el catálogo

**Compilador de consultas:** analiza las consultas sintácticamente, nombres de elementos y objetos. Traduce a formato interno.

**Optimizador de consultas:** reconfigura, reordena, elimina redundancias y usa algoritmos e índices para mejorar la ejecución de la consulta. Genera código ejecutable.

**Precompilador DML:** extrae los comandos DML y los envía al compilador DML. El resto del programa al compilador del lenguaje Host



# Cátedra Bas

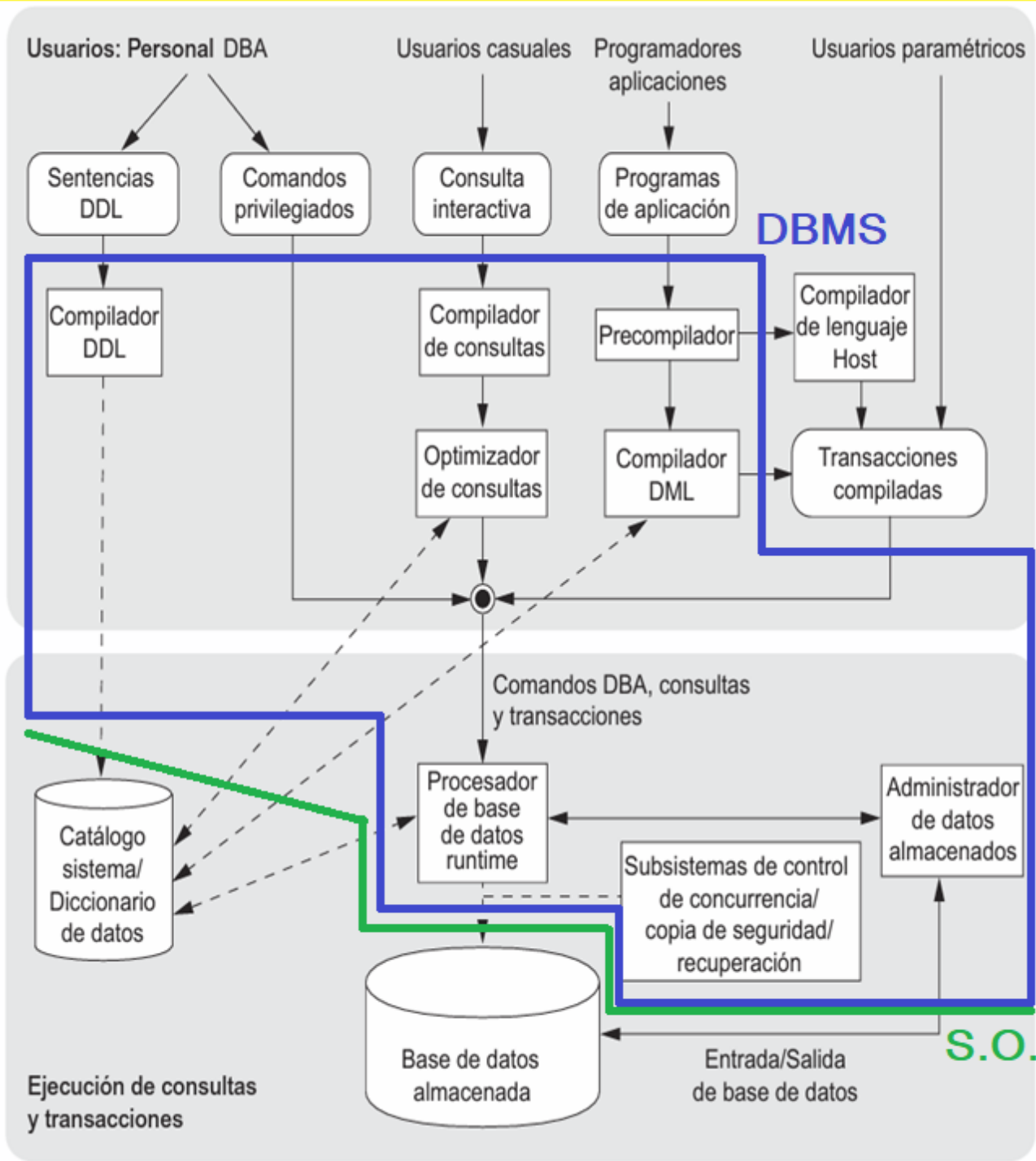
## Componentes del DBMS

**Compilador DML:** compila y genera código objeto que le pasa al Procesador de DB runtime

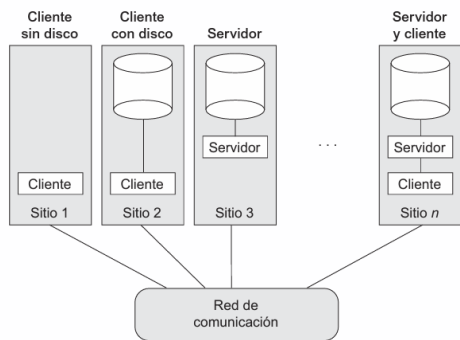
**Procesador de BD runtime:** recibe comandos y los ejecuta, accede al diccionario de datos y se comunica con el S.O. para E/S. Administra el buffer en la memoria de procesamiento, compartida o no con S.O.

**Subsist control concurrencia:** subfunción del Procesador de BD para colaborar con control de concurrencia.

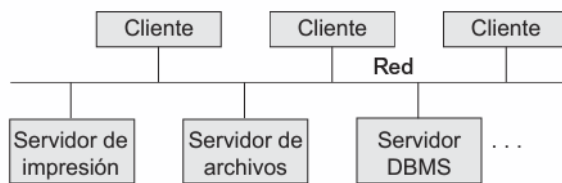
**Adm. de datos almacenados:** funciona con el Procesador de BD para poder comunicarse con el S.O. y realizar E/S.



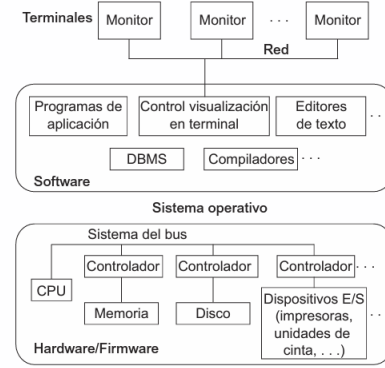
Arquitectura cliente/servidor física de dos capas.



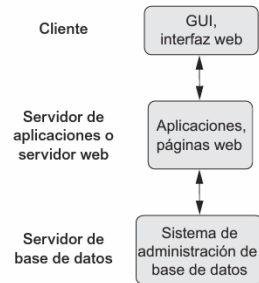
Arquitectura cliente/servidor



Arquitectura centralizada física.



Arquitectura cliente/servidor lógica de tres capas



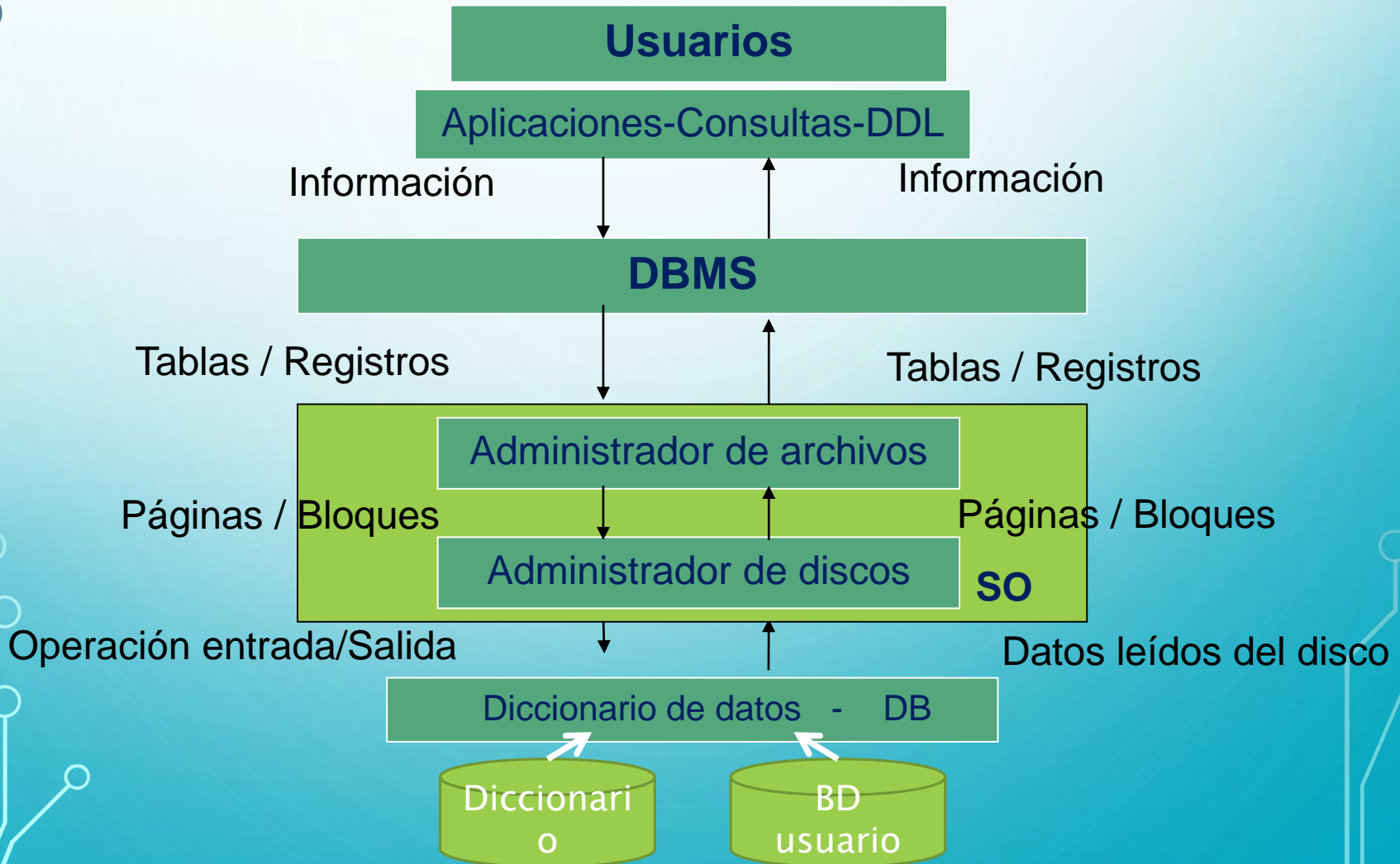
## Arquitecturas al implementar:

1. Centralizada
2. Servidor de Archivos (servidores especializados)
3. Cliente/Servidor de dos capas: componentes de software entre los sistemas clientes y servidor.
4. Arquitectura de tres capas y n capas: por aparición de la web.
  - **Clientes:** contienen interfaz (GUI) para mostrar y pedir datos-reglas comerciales
  - **Servidor intermedio:** procesa solicitud, envía comandos al servidor de BD y hacia el cliente
  - **Servidor de BD:** administración de los datos (puede dividir en más servidores)



# Cátedra Bases de Datos

## Nivel Interno: Acceso a datos



# Cátedra Bases de Datos

- **Almacenamiento Principal o Primario:** memorias, acceso rápido, capacidad limitada y volátil.
  - **Almacenamiento Secundario:** disco fijo -on line-
  - **Almacenamiento Terciario:** discos ópticos y cintas-off line- [IBM](#) [SUN](#)
- 
- **Proceso de Diseño Físico de la BD:**
    - Importante para: diseñador BD, DBA, implementador
    - Influye en performance
    - Implica colocar a los registros en el medio de almacenamiento (desordenado-ordenado-dispersión o hashing)



# Cátedra Bases de Datos

## Nivel Interno: Almacenamiento de BD

### Conceptos:

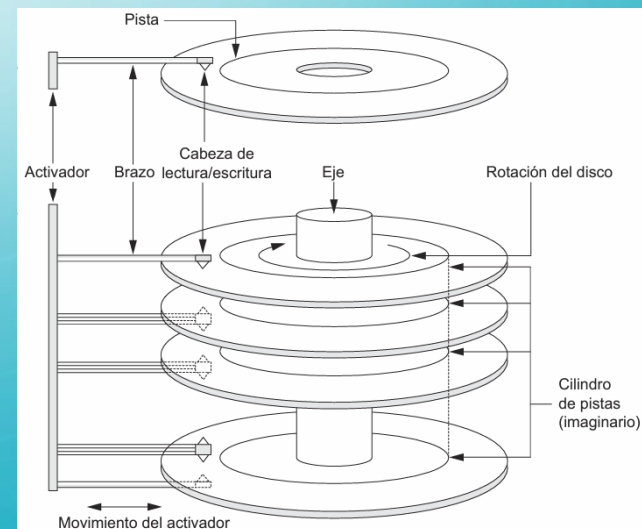
- **Bloque o página:** tamaño establecido por el SO al inicializar –no dinámico–

Es la unidad de transferencia entre disco y memoria.

- **Cluster:** bloques contiguos en el disco

- **Dirección del bloque:** para hacer E/S

Nro.cilindro + Nro.pista + Nro.bloque



# Cátedra Bases de Datos

- **Registros:** colección de valores –campos- BLOB (Binary Large Objects) se almacenan independientes con punteros.
- **Ficheros:** secuencia de registros - longitud fija y variable
- **Organización extendida y no extendida:** bloque/tamaño de registro
- **Asignación de bloques:** continua – enlazada – combinada – indexada
- **Cabecera de fichero:** para determinar longitud/orden de campos-separadores de registros

Tipos de organización de registros. (a) No extendida. (b) Extendida.





# Cátedra Bases de Datos

## Operaciones sobre ficheros

- Recuperación: localizar para examinar o procesar con condición—no modifican fichero-
- Actualización: localizar con condición para modificar fichero

**Organización de ficheros** —como se colocan e interconectan registros y bloques-

### 1- Ficheros de registros desordenados (Heap-Pila-Secuencial)

- Más sencillo: orden de entrada
- Inserción de registros es muy eficaz: lee último bloque del fichero, agrega, reescribe bloque
- Búsqueda puede resultar lenta: lineal, bloque a bloque
- Eliminación produce pérdida de espacio o se hace marca en un byte
- Reorganización periódica

# Cátedra Bases de Datos

## Organización de ficheros (continuación)

### 2- Ficheros de registros ordenados-ordenación física-

- De acuerdo a los valores de uno de los campos (clave)
- Inserción: costosa, hacer espacio entre registros
  - conviene reservar % del bloque –pero también se completa-
  - usar fichero de desbordamiento y luego se mezcla con el maestro/principal
- Eliminación: costosa
- Búsqueda por el campo de ordenación es eficaz (<,>,<=,>=,...) –binaria-

Algunos bloques de un fichero ordenado (secuencial) de registros EMPLEADO con Nombre como campo clave de ordenación.

**Bloque 1**

Nombre	Dni	FechaNac	Trabajo	Sueldo	Sexo
Aaron, Ed					
Abbott, Diane					
⋮					
Acosta, Marc					

**Bloque 2**

Adams, John					
Adams, Robin					
⋮					
Akers, Jan					

# Cátedra Bases de Datos

## Organización de ficheros (continuación)

### 3- Técnicas de dispersión –hash–

- Utiliza función de dispersión y se aplica al valor del campo hash
- Si campo hash es la PK se denomina Clave Hash
- Obtiene la dirección del bloque de disco
- Búsqueda por igualdad de un solo campo
- Función  $h(K) = K \bmod M$
- Valor del Campo K y M el número de slots (0 a M-1) con  $K=100$  y  $M=5 \Rightarrow h(K)=0$
- Si K es char se puede tomar ASCII
- Colisiones: dirección contiene registro
- Ubicarlos en posiciones subsiguientes / desbordamiento / aplicar otra función

# Cátedra Bases de Datos

## Nivel Interno

### Indexación

- Estructura auxiliar más pequeña
- Aceleran el acceso a datos
- Al crearlos ya existe un fichero con organización –desordenado, ordenado o disperso-
- Se puede utilizar cualquier campo para indexar, incluso compuestos
- Varios índices por fichero de datos
- Ofrecen rutas de acceso que no afectan la organización física
- Considerar la actualización de los índices generados
- Actualmente los DBMS generan índices (PK, FK)



# Cátedra Bases de Datos

**Indexación:**

## **Clasificación de los índices**

- Tipos de índices ordenados de un nivel
  - Principal
  - Agrupado
  - Secundario
- Índices multinivel



# Cátedra Bases de Datos

## Índice Principal:

- Precondición: Fichero de datos ordenado (físico) por campo clave
- No contiene valores repetidos
- El fichero índice tiene 2 columnas:
  1. Contiene valor del campo clave del primer registro de cada bloque (registro ancla)
  2. Puntero al bloque que lo contiene
- Es un índice escaso (no denso)
- Ocupa menos espacio que el fichero de datos, registro de longitud fija
- Más registros del fichero índice en cada bloque
- Búsqueda binaria en el fichero índice y lectura del bloque
- Actualización:
  - Inserción es un problema: pueden cambiar registros anclas o usar desbordamiento
  - Borrado: se usa borrado con marcador

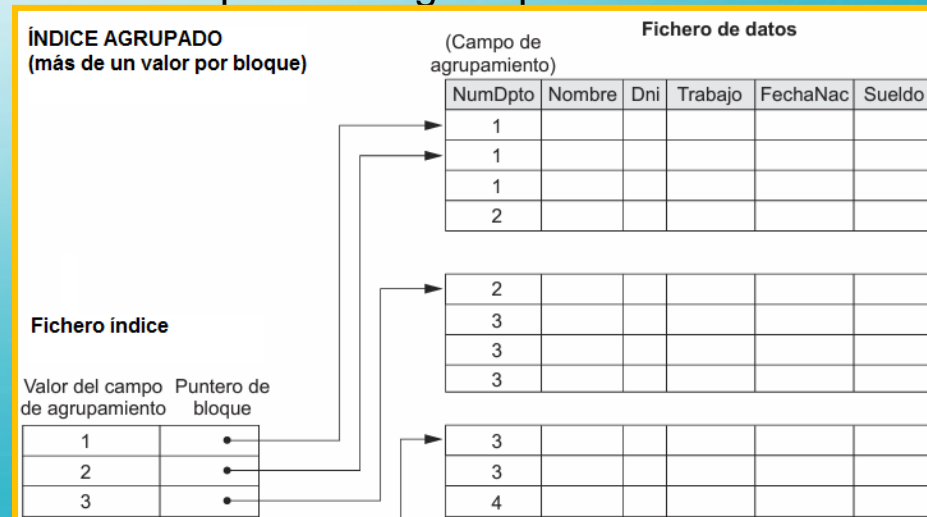
Bloques del fichero Estudiantes ordenados físicamente por el campo clave de ordenación Legajo.



# Cátedra Bases de Datos

## Índice Agrupado:

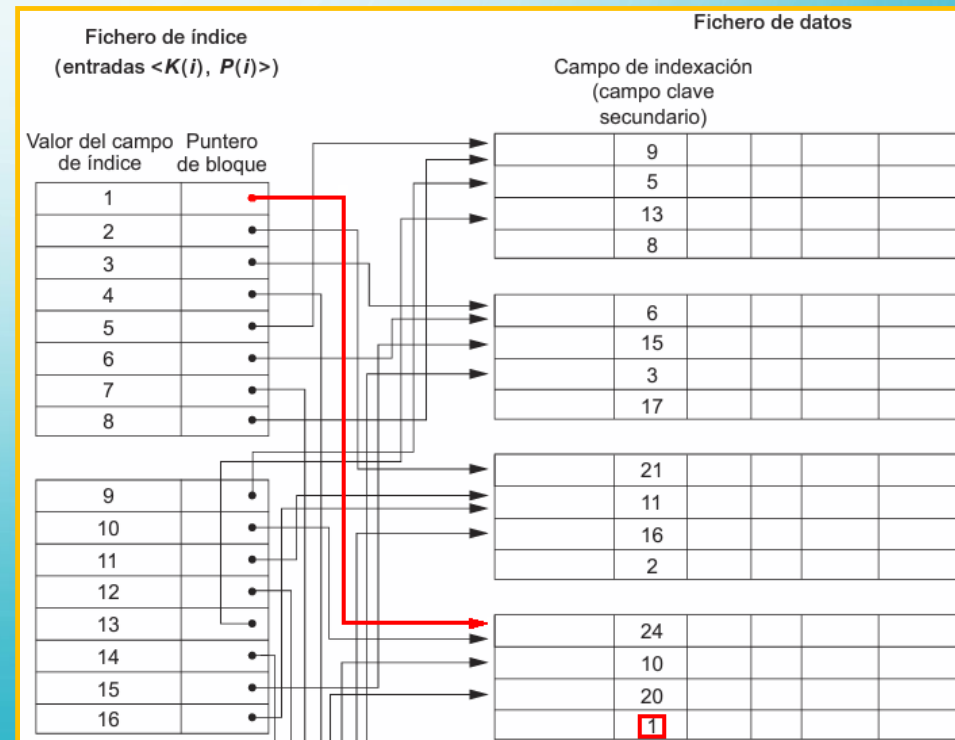
- Precondición: Fichero de datos ordenado (físico) por campo no clave.
- El fichero índice tiene 2 columnas:
  1. Contiene un registro por cada valor distinto del campo agrupado
  2. Puntero al primer bloque que contiene un registro con ese valor
- Es un índice escaso (no denso)
- Ocupa menos espacio que el fichero de datos, registro de longitud fija
- Más registros del fichero índice en cada bloque
- Búsqueda binaria en el fichero índice y lectura del bloque
- Actualización:
  - Inserción y borrado son un problema: reservar un bloque o contiguos por cada valor distinto del campo de indexación



# Cátedra Bases de Datos

## Índice Secundario:

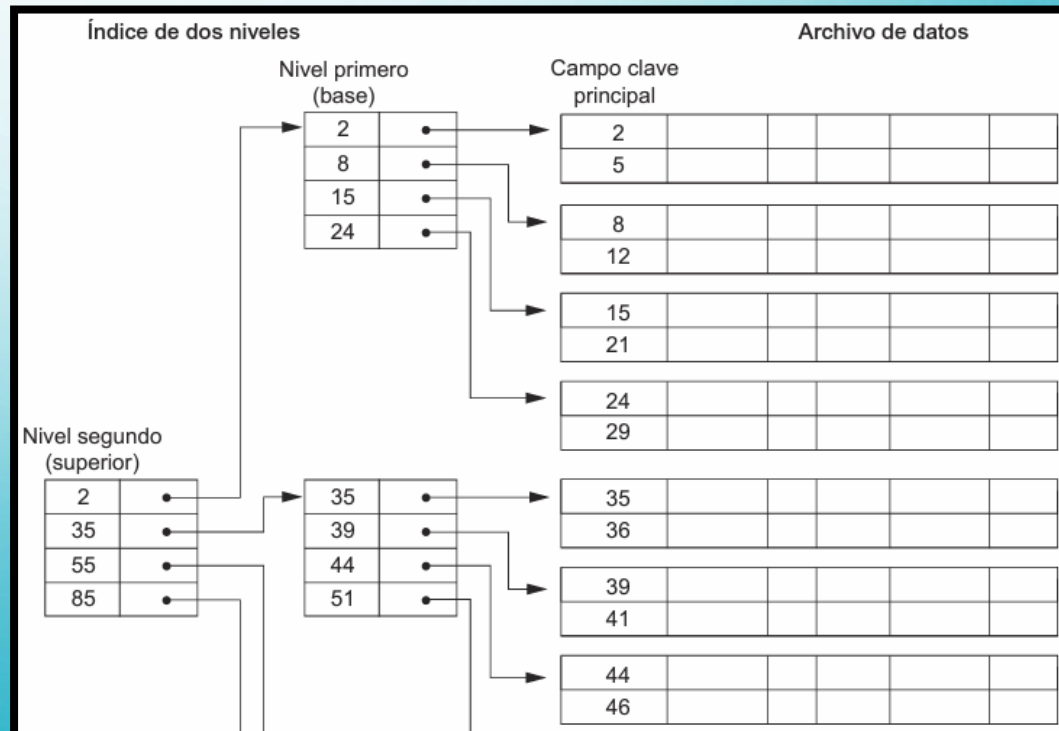
- Precondición: fichero ya tiene índice
- Generado por clave candidata (condiciones?) o no clave con valores duplicados
- Puede haber muchos índices secundarios por fichero de datos
- Proporcionan orden “lógico” a los registros almacenados
- El fichero índice tiene 2 columnas:
  1. Campo de indexación
  2. Puntero a bloque o a registro



# Cátedra Bases de Datos

## Índice Multinivel:

- Se aplican después de generar índice (principal, agrupado o secundario)
- Ese índice de 1° nivel es sobre campo clave (valores no repetidos) y entradas de longitud fija
- Genera 2° nivel-principal: Toma al 1° nivel como fichero ordenado con valor distinto en clave
- 2° nivel posee una entrada por bloque
- Genera 3° nivel (si el 2° nivel ocupa más de un bloque)
- Así sucesivamente hasta que las entradas de un mismo nivel se almacenan en un bloque



# Cátedra Bases de Datos

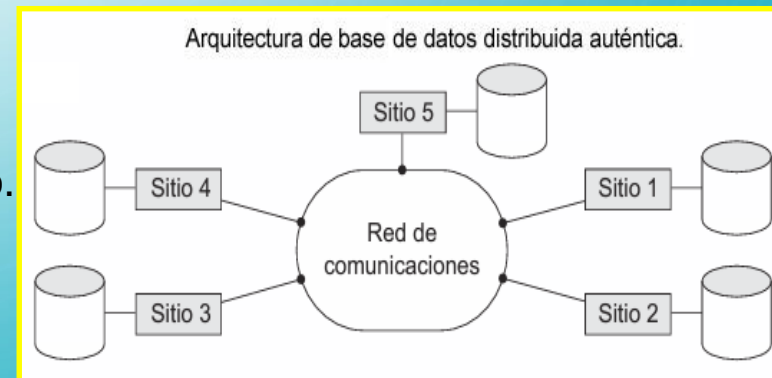
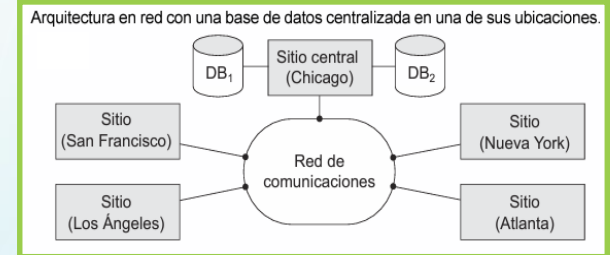
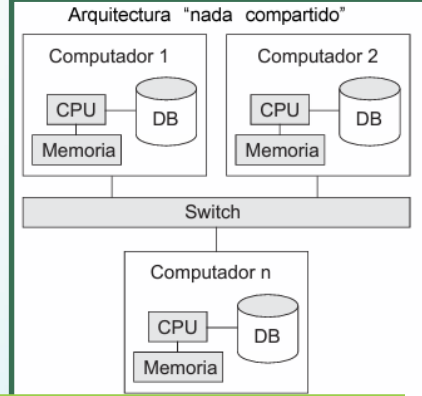
## Bases de Datos Distribuidas (BDD)

Surge por la unión de dos tecnologías:

- Base de Datos
- Comunicación de datos y redes.

### Conceptos:

- **DDB**: objeto virtual cuyas partes se almacenan en diferentes sitios.
- **Sistema de DDB**: conjunto de sitios conectados mediante red de comunicación.
- **Sitio**: es un sistema de BD en sí mismo (hard-soft-usuarios-datos).
- **DDBMS**:
  - Software para administrar la BDD.
  - Hace que la situación sea transparente al usuario.
- **Transacciones**: locales / globales





# Cátedra Bases de Datos

## Técnicas de diseño de BDD

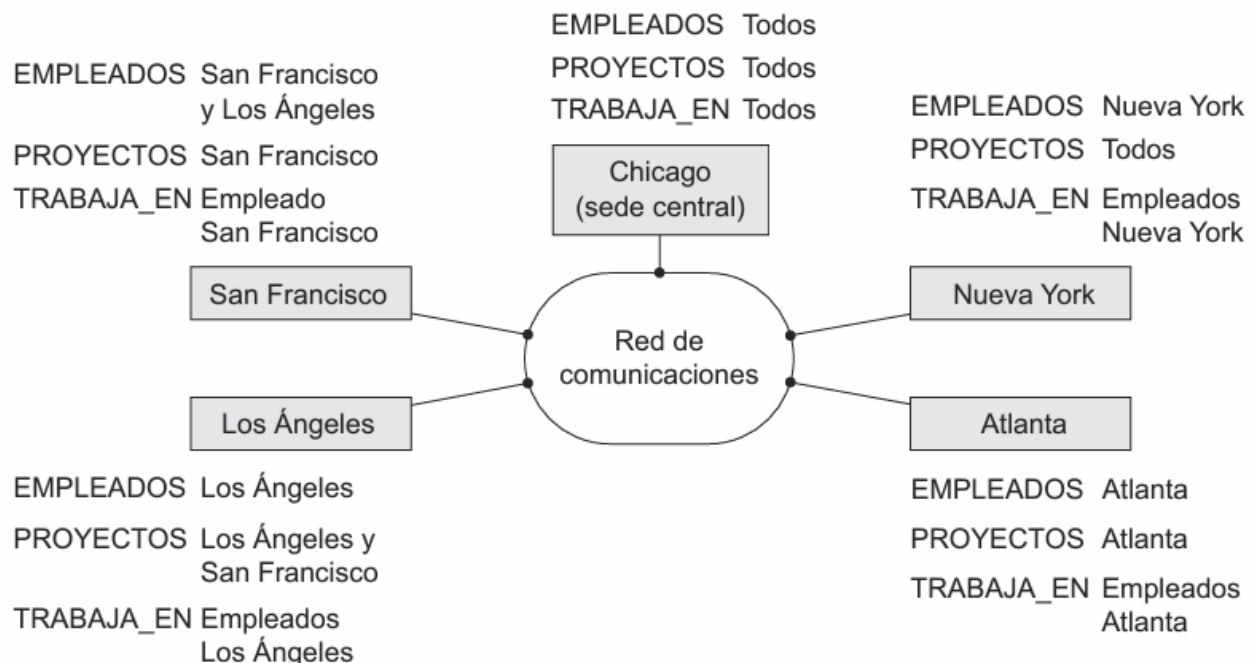
### • Fragmentación de datos

- Horizontal
- Vertical
- Mixta o híbrida

### • Replicación

- Total
- Parcial

Distribución de datos y replicación entre bases de datos distribuidas.



# Cátedra Bases de Datos

## Bases de Datos Distribuidas (BDD)

### Ventajas de las BDD

1. Administración de datos distribuida con distintos niveles de transparencia.
  - a. Transparencia de red o de distribución
    - . localización
    - . denominación
  - b. Transparencia de replicación
  - c. Transparencia de fragmentación
  - d. Transparencia de diseño y ejecución
2. Incremento de la fiabilidad y la disponibilidad
3. Rendimiento mejorado: uso de datos del sitio y ejecución paralela de consultas
4. Expansión más sencilla: en datos y procesadores

# Cátedra Bases de Datos

## Funciones Adicionales del DDBMS

1. Seguimiento de los datos
2. Procesamiento de consultas distribuidas
3. Administración de transacciones distribuidas
4. Administración de datos replicados
5. Recuperación de una BDD
6. Seguridad
7. Administración del directorio (catálogo) distribuido

**Aumentan la complejidad respecto a los DBMS**

# Cátedra Bases de Datos

## Diferencias a nivel de Hardware entre BDD y Centralizada

- . Existen múltiples equipos llamados sitios o nodos
- . Los sitios deben estar conectados por algún tipo de red

## La conexión puede ser a través de:

- Red de área local
- Red de área expandida

Se pueden usar diferentes topologías

## Tipos de Sistemas de BDD:

- Grado de homogeneidad del DDBMS: homogéneo / heterogéneo
- Con o sin autonomía local



# Cátedra Bases de Datos

## Cuestiones que cambian entre BD centralizada y BDD:

- Costo de transferencia de datos
- Acceso concurrente y consistencia de datos
- Fallo de sitios
- Fallo de enlaces
- Confirmar actualización en datos de distintos sitios
- Interbloqueo (deadlock)

