

Elementos de Bases de Datos

Dpto. Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur

Lic. María Mercedes Vitturini
[mvitturi@cs.uns.edu.ar]

Clase 20

1er. Cuatrimestre de 2004



Arquitecturas de Sistemas de Bases de Datos

- La arquitectura de un sistema de bases de datos está influenciada por el sistema informático subyacente en el que se ejecuta el sistema de bases de datos.
- En la arquitectura de un sistema de bases de datos interesan aspectos como:
 - La conexión en red
 - Capacidades de procesamiento en paralelo
 - Capacidades de distribución de datos.

Elementos de Bases de Datos
Clase 20

2

Diferentes Arquitecturas

Estudiaremos básicamente los siguientes sistemas:

- Sistemas Centralizados.
- Sistemas Cliente/Servidor.
- Sistemas Paralelos.
- Sistemas Distribuidos.

Elementos de Bases de Datos
Clase 20

3

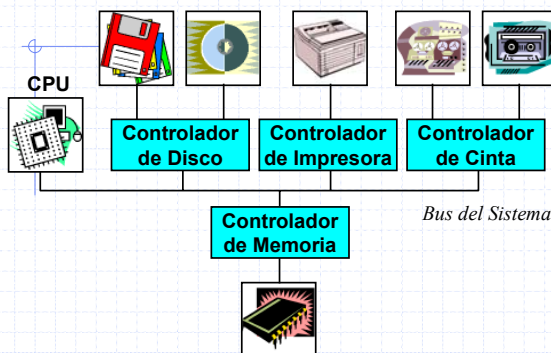
Sistemas Centralizados- Características

- No interactúan con otros sistemas.
- Abarcan desde sistemas para un único usuario a mainframes de alta performance.
- Contienen una o (en algunos casos unas pocas) CPU, una memoria principal y una serie de controladores encargados de funciones específicas.
- La CPU y los controladores pueden ejecutar acciones de manera concurrente, compitiendo por el acceso a memoria.
- El uso de memorias caché reduce el número de accesos de la CPU a la memoria principal.

Elementos de Bases de Datos
Clase 20

4

Sistemas Centralizados



Elementos de Bases de Datos
Clase 20

5

Sistemas Centralizados

- Dentro de la arquitectura Sistemas Centralizados se distinguen:
 - Sistemas monousuarios.
 - Sistemas multiusuarios.

Elementos de Bases de Datos
Clase 20

6

Sistemas Centralizados

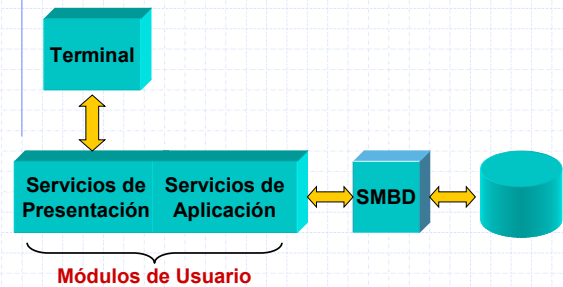
Sistemas Monousuario

- Son principalmente usados o administrados por una sola persona.
- Cuentan con una CPU, uno o dos discos rígidos, y un sistema operativo monousuario.
 - Computadoras Personales (PC's).
 - Estaciones de Trabajo (Workstations).
- En general no proporcionan mecanismos de control de concurrencia, facilidades de recuperación, etc.

Elementos de Bases de Datos
Clase 20

7

Procesamiento de Transacciones: Sistema Centralizado Monousuario



Elementos de Bases de Datos
Clase 20

8

Sistemas Centralizados

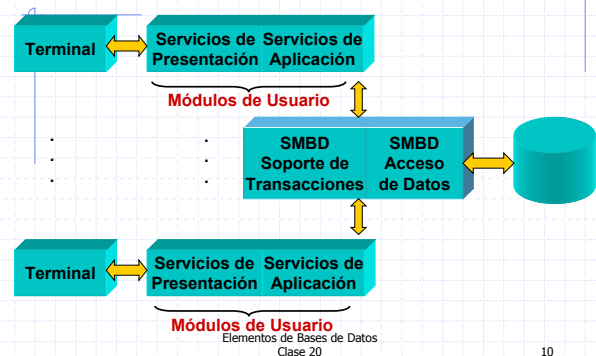
Sistemas Multiusuario:

- Generalmente cuentan con varios usuarios operando el sistema al mismo tiempo, más de una CPU's, memoria y un sistema operativo multiusuario.
- Generalmente disponen de facilidades para multitareas (concurrencia).
- Puede atender a varios usuarios que están conectados al sistema por medio de terminales.

Elementos de Bases de Datos
Clase 20

9

Procesamiento de Transacciones Sistema Centralizado Multiusuario



Elementos de Bases de Datos
Clase 20

10

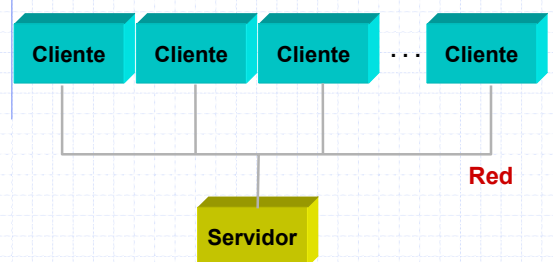
Sistemas Cliente-Servidor

- Con el desarrollo de la tecnología para computadoras personales, actualmente, los sistemas centralizados actúan como *sistemas servidores* que satisfacen los requerimientos generados por *sistemas clientes*.
- Las funciones de la base de datos se dividen en:
 - Aplicaciones Back-End (manejan estructuras de acceso, evaluación y optimización de consultas, control de concurrencia y recuperación ante fallos).
 - Aplicaciones Front-End (abarcan generadores de reportes y facilidades para el diseño de interfaces gráficas para el usuario).

Elementos de Bases de Datos
Clase 20

11

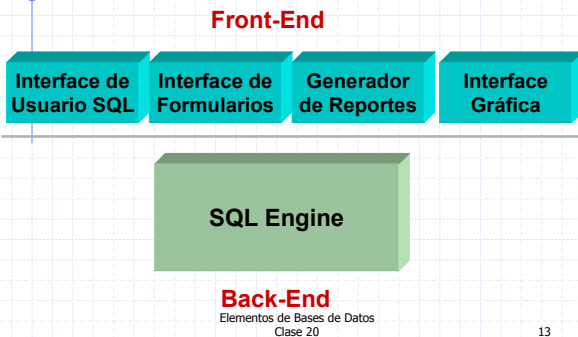
Sistemas Clientes-Servidor



Elementos de Bases de Datos
Clase 20

12

Funcionalidades Front-End y Back-End



13

Categorías de los sistemas Cliente-Servidor

- Los sistemas servidores se clasifican en:
 - Servidores de Transacciones.
 - Servidores de Datos.

Elementos de Bases de Datos
Clase 20

14

Categorías de sistemas cliente-servidor

Servidores de Transacciones (de Consultas):

- Proveen una interface mediante la cual los clientes pueden enviar **requerimientos para ejecutar una acción**.
- En respuesta el servidor ejecuta la acción y envían el resultado al cliente.
- Los usuarios especifican los pedidos directamente en SQL o desde la aplicación usando llamadas a procedimientos remotos.
- Respetan la división funcional entre aplicaciones back-end (en el servidor) y aplicaciones front-end (en el cliente).

Elementos de Bases de Datos
Clase 20

15

Categorías de sistemas cliente-servidor

Servidores de Datos

- Los clientes interactúan con los servidores haciendo **requerimientos para lectura o actualización** de datos.
- Permiten manejar diferentes unidades de datos (páginas, tuplas, objetos)
- Proporcionan facilidades de indexación y manejo de transacciones que protegen a los datos de posibles inconsistencias.
- Trabajan con redes locales con alta tasa de transferencia.
- Los clientes son comparables al servidor y tienen capacidades para realizar las aplicaciones front y back end.

Elementos de Bases de Datos
Clase 20

16

Sistemas Paralelos

- Mejoran el procesamiento y las velocidades de E/S usando más de una CPU y más de un medio de almacenamiento (disco) en paralelo.
- En estos sistemas, muchas operaciones se desarrollan **simultáneamente**, en contraposición a los sistemas seriales y concurrentes.
- Surgen junto con la demanda de aplicaciones que manejan bases de datos muy grandes (del orden de terabytes) y de miles de transacciones por minuto.

Elementos de Bases de Datos
Clase 20

17

Sistemas Paralelos

Categorías:

- Coarse-Grain (granularidad gruesa):** máquinas con un pequeño número de procesadores potentes.
- Massively parallel o Fine-Grain (granularidad fina):** máquinas con cientos de procesadores pequeños.

Elementos de Bases de Datos
Clase 20

18

Medidas de Performance

- Un SMDB considera dos parámetros:
 - Productividad (Throughput):** el número de tareas que pueden terminarse en un período de tiempo.
 - Tiempo de Retorno:** el tiempo que toma para completar una única tarea.
- En cambio, un SO también considera:
 - Utilización de CPU:** se busca maximizar el uso del (o los) procesador(es) central(es).
 - Tiempo de Espera:** el tiempo que espera un proceso en realizar una operación de entrada/salida.
 - Tiempo de Respuesta:** el tiempo que toma en dar una primera respuesta a un proceso.

Elementos de Bases de Datos
Clase 20

19

Performance de Sistemas Paralelos

- Los sistemas paralelos aumentan el rendimiento procesando muchas transacciones en paralelo.
- El tiempo de retorno también es mejor ya que pueden desarrollar subtareas de una transacción en paralelo.
- Cuestiones de estudio en sistemas paralelos:
 - Speedup:** se refiere a ejecutar una tarea en menos tiempo incrementando el grado de paralelismo.
 - Scaleup:** se refiere a tratar tareas más complejas incrementando el grado de paralelismo.

Elementos de Bases de Datos
Clase 20

20

Arquitecturas Paralelas de Bases de Datos

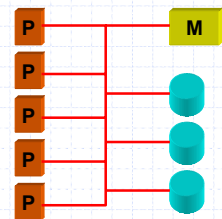
- Existen distintos modelos de arquitectura para máquinas paralelas:
 - Memoria compartida:** todos los procesadores comparten una memoria común.
 - Disco compartido:** todos los procesadores comparten un disco común.
 - Nada compartido.**
 - Jerárquico:** una combinación de los anteriores.

Elementos de Bases de Datos
Clase 20

21

Arquitecturas Paralelas de Bases de Datos

- Memoria Compartida:** todos los procesadores y discos comparten una memoria en común.
- '+' Es eficiente la comunicación entre procesadores.
- '-' El inconveniente es que la arquitectura no admite más de 32 a 64 procesadores ya que el bus de interconexión resulta un cuello de botella.

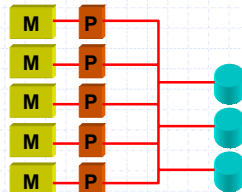


Elementos de Bases de Datos
Clase 20

22

Arquitecturas Paralelas

- Disco Compartido:** Todos los procesadores comparten todos los medios de almacenamiento.
- Cada procesador cuenta con su memoria principal.
 - '+' el bus no es cuello de botella
 - '+' proporcionan tolerancia a fallas de procesador.
 - '-' el disco es el cuello de botella. Esto limita el crecimiento.

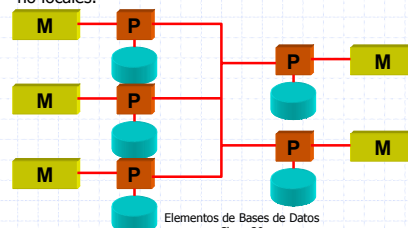


Elementos de Bases de Datos
Clase 20

23

Arquitecturas Paralelas

- Nada Compartido:** Cada procesador tiene su propia memoria y su propio disco.
 - '+' Son ampliables y pueden soportar gran número de procesadores.
 - '-' La desventaja está en el costo de comunicación y acceso a disco no locales.

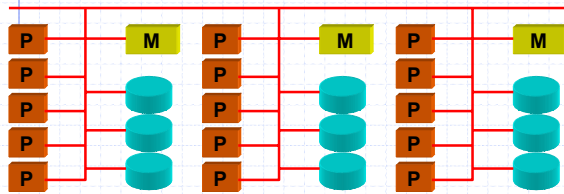


Elementos de Bases de Datos
Clase 20

24

Arquitecturas Paralelas

- **Jerárquico:** Es un sistema híbrido entre memoria compartida, disco compartido y nada compartido.
- Las arquitecturas de memoria virtual distribuida han surgido de la idea de arquitecturas jerárquicas.



Elementos de Bases de Datos
Clase 20

25

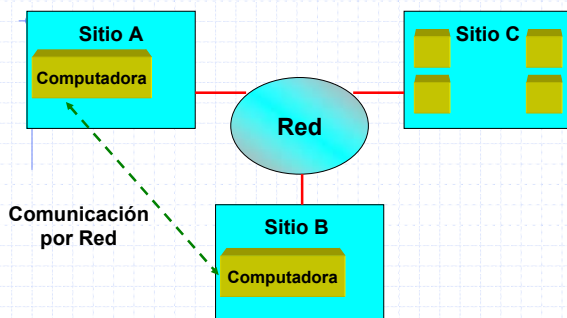
Sistemas Distribuidos

- La base de datos se almacena en **varias computadoras** de un sistema distribuido comunicadas entre si mediante diferentes dispositivos de comunicación.
- Las computadoras no comparten memoria ni disco.
- En un sistema distribuido, las computadoras pueden variar en tamaño y función, desde workstations hasta mainframes.
- Cada computadora en un sistema distribuido se denomina **sitio** o **nodo**, dependiendo del contexto en el que se la menciona.

Elementos de Bases de Datos
Clase 20

26

Sistemas Distribuidos



Elementos de Bases de Datos
Clase 20

27

Sistemas Distribuidos

Características:

- Cada sitio conoce la existencia de los otros.
- Los sitios comparten un esquema global, aunque algunas relaciones (conjuntos de tuplas) pueden ser almacenados sólo en algunos sitios.
- Cada sitio provee un entorno para ejecutar tanto transacciones locales como globales.
- Cada sitio ejecuta el mismo software de manejo de bases de datos distribuidas.
- Comparte muchas características de los sistemas cliente/servidor y de los sistemas paralelos.

Elementos de Bases de Datos
Clase 20

28

Sistemas Distribuidos vs. Paralelos

- Los **sistemas distribuidos** se asemejan a los **sistemas paralelos con estructura nada compartido**.
- Sin embargo, generalmente están en **lugares diferentes** (geográficamente separados), son administradas **independientemente** y tienen una **interconexión más lenta**.
- Otra diferencia es el tipo de transacciones que admiten. Las **transacciones locales** acceden a datos de un único sitio mientras que las **transacciones globales** pueden acceder a datos que están en sitios diferentes a donde se inicia la transacción.

Elementos de Bases de Datos
Clase 20

29

Sistemas Distribuidos - Ventajas

- **Compartimiento de Datos:** los usuarios de un sitio pueden acceder a datos de otros sitios.
- **Autonomía:** cada sitio es capaz de mantener el control de los datos que están almacenados localmente. Existe un administrador global del sistema y cada sitio cuenta con un administrador local.
- **Disponibilidad:** si un sitio falla, el resto de los sitios puede continuar operando. Si la información se replica, el fallo de un sitio puede no afectar a todo el sistema.

Elementos de Bases de Datos
Clase 20

30

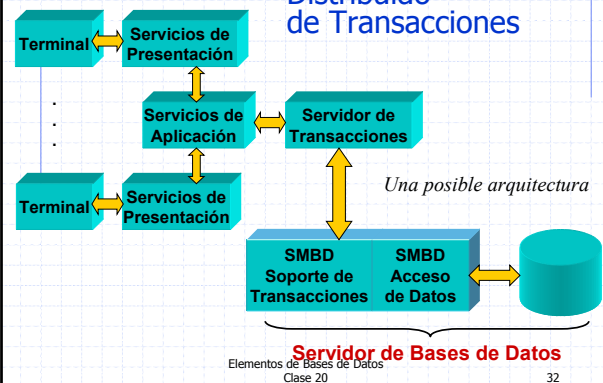
Sistemas Distribuidos - Desventajas

- **Costo del Desarrollo de Software:** es más difícil y costoso implementar un sistema de base de datos distribuido.
- **Mayor Probabilidad de Errores:** Puesto que los sitios que constituyen el sistema distribuido operan en paralelo, es más difícil asegurar la correctitud de los algoritmos.
- **Incremento en la Sobrecarga de Procesamiento:** el intercambio de mensajes y los cómputos adicionales para coordinar los sitios constituyen una sobrecarga importante.

Elementos de Bases de Datos
Clase 20

31

Procesamiento Distribuido de Transacciones



Elementos de Bases de Datos
Clase 20

32

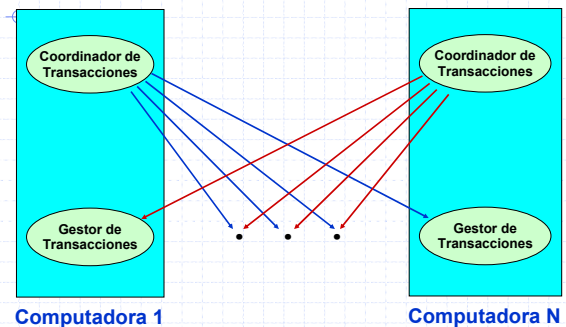
Modelo Transaccional Distribuido

- Cada sitio de un sistema contiene dos subsistemas:
 - El **gestor de transacciones**: gestiona la ejecución de aquellas transacciones (o subtransacciones) que acceden a los datos almacenados en el sitio local.
 - Mantener la bitácora
 - Participar en el esquema de control de concurrencia.
 - El **coordinador de transacciones**: encargado de la ejecución de las transacciones (locales y globales) iniciadas en el sitio local.
 - Iniciar las transacciones.
 - Dividir las transacciones en subtransacciones.
 - Coordinar la ejecución de subtransacciones.

Elementos de Bases de Datos
Clase 20

33

Modelo Transaccional Distribuido



Elementos de Bases de Datos
Clase 20

34

Tipos de Redes

- Las bases de datos distribuidas y los sistemas cliente/servidor se construyen sobre la base de redes de comunicación
- Existen dos tipos de redes:
 - Redes de área local (LAN).
 - Redes de área amplia (WAN).
- Entre ellas varían en velocidad y confiabilidad.

Elementos de Bases de Datos
Clase 20

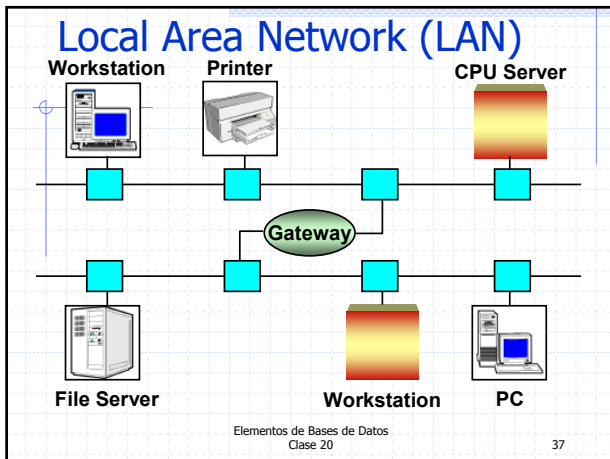
35

Tipos de Redes

- **LANs - Local Area Networks (1970s).**
 - Cubren áreas geográficas pequeñas (por ejemplo, un edificio o varios edificios cercanos).
 - Conexiones por cable coaxial o fibra óptica.
 - Ancho de banda: 1MB/seg a 1GB/seg (por ejemplo, Ethernet: 10 megabit/seg, Fast Ethernet: 100–144 megabits/seg).

Elementos de Bases de Datos
Clase 20

36



Tipos de Redes

- **WANs - Wide Area Networks (1960s).**
 - Cubren áreas geográficas extensas (ciudades, países, continentes).
 - Conexiones típicamente por teléfono, soportadas por líneas de fibra óptica o por satélite.
 - Las conexiones pueden ser discontinuas (UUCP de Unix) o continuas (Internet).

Elementos de Bases de Datos
Clase 20 38

Diferentes Topologías de Redes

- **Red Totalmente Conectada:** existe un vínculo entre cada par de nodos.
- **Red Parcialmente Conectada:** algunos pares de nodos no están conectados.
- **Red Tipo Arbol:** existe una estructura arbórea con un nodo raíz, sus hijos. Cada nodo puede tener 0 o más nodos hijos pero sólo un nodo padre.
- **Red Estrella:** existe un nodo al cual todos están conectados.
- **Red Anillo:** los nodos están conectados circularmente.

Elementos de Bases de Datos
Clase 20 39

Red Totalmente Conectada

The diagram shows a fully connected network with six nodes labeled A, B, C, D, E, and F arranged in a hexagonal pattern. Every node is connected to every other node, forming a complete graph. A red-bordered box on the right contains the following text: '+ Gran tolerancia a fallas de comunicación.' and '- Muy alto costo de instalación.'

Elementos de Bases de Datos
Clase 20 40

Red Parcialmente Conectada

The diagram shows a partially connected network with six nodes labeled A, B, C, D, E, and F. Node A is at the top and is connected to nodes B, E, and F. Node B is connected to A and C. Node C is connected to B and D. Node D is connected to C. Node E is connected to A. Node F is connected to A. A red-bordered box on the right contains the following text: '+ Aceptable costo de instalación.' and '- Si un mensaje falla por un camino, puede ser redirigido (routed) por otro camino.'

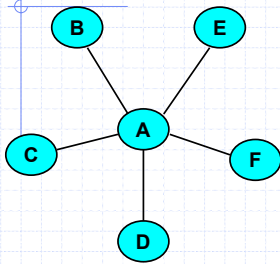
Elementos de Bases de Datos
Clase 20 41

Red Tipo Arbol

The diagram shows a tree network with a root node A at the top. Node A has two children: B and E. Node B has two children: C and D. Node E has one child: F. A red-bordered box on the right contains the following text: '+ Bajos costos de instalación y comunicación.' and '- Si se cae un nodo, sus hijos quedan aislados.'

Elementos de Bases de Datos
Clase 20 42

Red Tipo Estrella

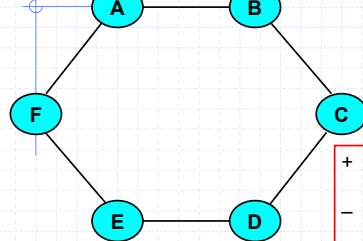


- + El costo de comunicación es bajo.
- + Si falla un nodo lateral queda aislado pero el resto continúa normalmente.
- Si falla el nodo central, el sistema queda totalmente desconectado.

Elementos de Bases de Datos
Clase 20

43

Red Tipo Anillo



- + El grado de disponibilidad es alto.
- El costo de comunicación suele ser alto porque debe atravesar muchos vínculos.

Elementos de Bases de Datos
Clase 20

44

Temas de la clase de hoy

- Arquitecturas de Sistemas de Bases de datos
 - Sistemas Centralizados.
 - Sistemas Cliente - Servidor.
 - Sistemas Paralelos.
 - Sistemas Distribuidos.
- Redes
 - Tipos de Redes
 - Topologías
- Bibliografía
 - "Fundamentos de Bases de Datos" – A. Silberschatz. Capítulo 16 y 18.
 - "Databases and Transaction Processing" – P. Lewis. Capítulo 22

Elementos de Bases de Datos
Clase 20

45