

Fundamentos de las Bases de Datos Distribuidas

M.C. Carlos Rojas Sánchez¹

¹Licenciatura en Informática
Universidad del Mar - Puerto Escodido

[Bases de Datos Distribuidas]

1 Concepto, característica de las bases de datos distribuidas

BDD

- SW que maneja BD's y que hace una distribución transparente a los usuarios
- No es una colección de archivos que estan almacenados individualmente en la red
- “logicamente interrelacionados”
- “distribuidos bajo una red de computadoras”

2 Evolución en la gestión de datos

Evolución en la gestión de datos

- Los años sesenta y setenta: sistemas centralizados
- Los años ochenta: SGBD relacionales
- Los años noventa: distribución, C/S y 4GL
- Tendencias actuales

Sistemas centralizados

- Estos sistemas trabajaban exclusivamente por lotes (batch)
- Los SGBD estaban íntimamente ligados al software de comunicaciones y de gestión de transacciones

SGBD relacionales

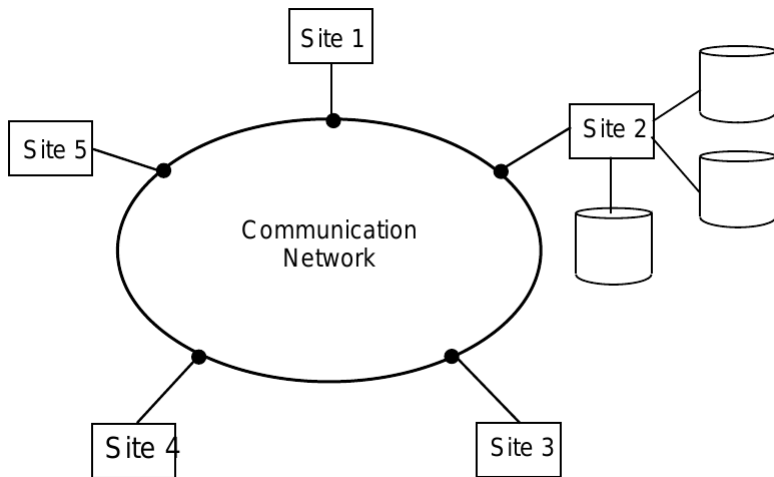
- La aparición de los SGBD relacionales supone un avance importante para facilitar la programación de aplicaciones con BD y para conseguir que los programas sean independientes de los aspectos físicos de la BD.
- Oracle aparece en el año 1980

Distribución, C/S y 4GL

- La necesidad de tener una visión global de la empresa y de interrelacionar diferentes aplicaciones que utilizan BD diferentes, junto con la facilidad que dan las redes para la intercomunicación entre ordenadores, ha conducido a los SGBD actuales, que permiten que un programa pueda trabajar con diferentes BD como si se tratase de una sola. Es lo que se conoce como base de datos distribuida.
- La tecnología que se utiliza habitualmente para distribuir datos es la que se conoce como entorno (o arquitectura) cliente/servidor (C/S). Todos los SGBD relacionales del mercado han sido adaptados a este entorno.

3 BD Distribuidas vs BD centralizadas

BD Centralizadas

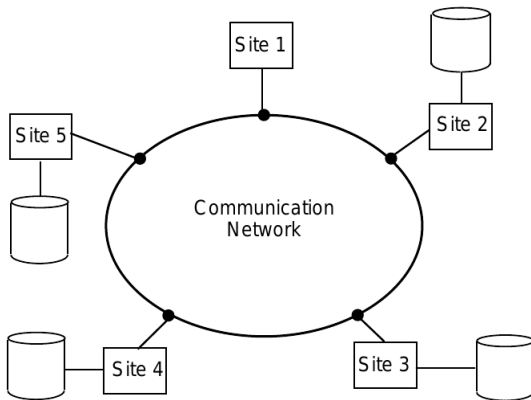


BD Centralizadas

■ Centralizado

- Control centralizado: Un solo DB
- Independencia de Datos: Organización de los datos es transparente para el programador
- Reducción de redundancia: Una sola copia de datos que se comparta
- Estructuras físicas complejas para accesos eficientes
- Seguridad

BD Distribuidas



BDD

- Una base de datos distribuida (DDB) es aquella en la que sus datos se encuentran repartidos en varias computadoras que están interconectadas por una red de comunicaciones.
- Los sitios trabajan en conjunto si es necesario con el fin de que un usuario de cualquier sitio pueda obtener acceso a los datos de cualquier punto de la red tal como si todos los datos estuvieran almacenados en el sitio propio de l usuario.

BDD

■ Aplicaciones

- Manufactura(multiplantas)
- Comando y control militar
- SIGs corporativos
- Aerolíneas
- Cadenas hoteleras
- Cualquier organización que tenga una estructura o organizacional descentralizada

BDD

■ Distribuido

- Control jerárquico: DBA global y DBA locales
- Transparencia en la Distribución: Localización de los datos es un aspecto adicional de independencia de datos
- Replicación de Datos: Copias múltiples de datos que incrementa la localidad y la disponibilidad de datos
- No hay estructuras intersitios. Uso de optimización global para reducir transferencia de datos
- Problemas de seguridad Intrínsecos

4 Las doce reglas (objetivos)

Doce reglas

Principio fundamental

Para el usuario un sistema distribuido debe ser igual que uno centralizado

Doce reglas

- Autonomía Local : Los sitios distribuidos deben ser autónomos
- No dependencia de un sitio central
- Operación Continua: Nunca debería haber necesidad de apagarse a propósito para que se pueda realizar alguna función, como añadir un nuevo sitio, o instalar una versión mejorada
- Independencia con respecto a la localización
- Independencia con respecto a la fragmentación
- Independencia de réplica

Doce reglas

- **Procesamiento Distribuido de Consultas:** El objetivo es convertir transacciones de usuario en instrucciones para manipulación de datos, y así reducir el tráfico en la red, esto implica que el proceso mismo de optimización de consultas debe ser distribuido
- **Manejo Distribuido de Transacciones:** Tiene dos aspectos principales, el control de recuperación y el control de concurrencia
- **Independencia con respecto al equipo:** El Sistema de Gestión de Base de Datos Distribuidos debe ser ejecutable en diferentes plataformas hardware
- **Independencia con respecto al Sistema Operativo**
- **Independencia con respecto a la red**
- **Independencia con respecto al DBMS**

5 Razones del proceso distribuido

Razones del proceso distribuido

- Razones económicas y organizacionales
- Fiabilidad y disponibilidad
- Distribución del control y utilización compartida de datos
- Reducción de tráfico en la comunicación de datos
- Agilización del procesamiento de consultas
- Interconexión de base de datos existentes
- Crecimiento incremental

6 Problemas de los sistemas distribuidos

Problemas de los SBDD

- Diseño de base de datos distribuidas
 - Como distribuir las Bases de datos
 - Distribución de datos replicada y no replicada
 - Problema relacionado con la administración del diccionario
- Procesamiento de consultas
 - Problema de optimización
 - $\text{Min}(\text{costo} = \text{transmisión de datos} + \text{procesamiento local})$
 - La fórmula general es un problema NP completo

Problemas de los SBDD

- Control de concurrencia

- Sincronización de acceso concurrente
- Manejo de interbloqueo

- Fiabilidad

- Como hacer que el sistema resistente a fallas?
- Atomicidad y durabilidad

- Soporte de sistema operativo

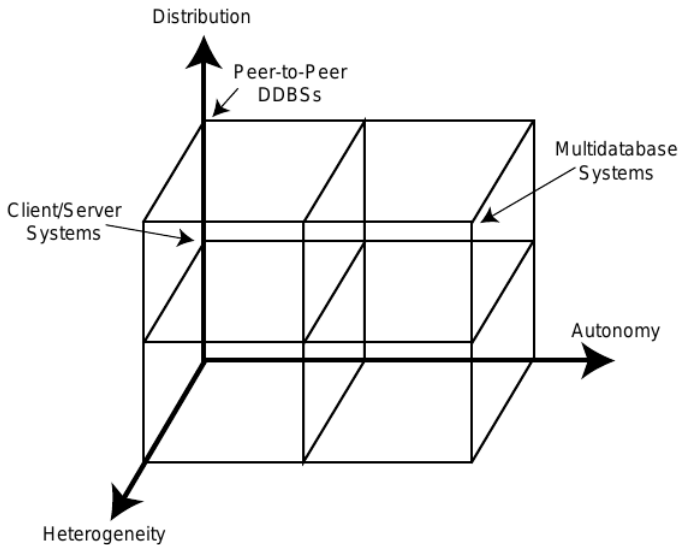
- Sistema operativo con soporte adecuado para operar bases de datos

7 Sistemas gestores de bases de datos (DBMS) distribuidos

DBMS Distribuido

- Un DBMS-D es el software que gestiona BDD y suministra mecanismos de acceso que hace la distribución transparente al usuario
- Múltiples DBMS conectados bajo una red
- Pueden ser Homogéneos o Heterogéneos

DBMS Distribuido





8 Reglas de DBMS distribuido

9 Consideraciones para distribuir la BD

Hay que decidir en qué nodos deben residir los datos y las aplicaciones que trabajan con los datos

- Si existen ya las bases de datos, hay que integrarlas para obtener el esquema global
- Si no existen, hay que definir el esquema conceptual global y fragmentar y asignar a los nodos

Bibliografía I

-  M. Tamer Özsu and Patrick Valduriez.
Principles of Distributed Database Systems.
Springer, 2011.
-  Abraham Silberschatz, Henry F. Korth and S. Sudarshan.
Fundamentos de Bases de Datos.
McGRAW-HILL, 2002.