

# BASES DE DATOS DISTRIBUIDAS

1

## Definición:

- **Consiste en una colección de sitios, conectados por medio de algún tipo de red de comunicación, en el cual**
  - Cada sitio es un sistema de BD completo por derecho propio, pero
  - Los sitios ha acordado trabajar juntos, a fin de que un usuario de cualquier sitio pueda acceder a los datos desde cualquier lugar de la red, exactamente como si los datos estuvieran guardados en el propio sitio del usuario.

2

**Bases de Datos Distribuidas****Definición:**

- Una BDD es en realidad un tipo de BD virtual cuyas partes componentes están almacenadas en varias BD “reales” distintas que se encuentran en varios sitios distintos (de hecho, es la unión lógica de esas BD reales).

3

**Bases de Datos Distribuidas****Definición:**

- En otras palabras, cada sitio local tiene
  - Sus propias BD “reales”
  - Sus propios usuarios locales
  - Su propio DBMS local
  - Software de administración de transacciones (incluyendo su propio software local para bloqueo, registro en bitácora, recuperación, etc.)
  - Así como su propio administrador de comunicación de datos local.

4

**Bases de Datos Distribuidas****Definición:**

- Es común suponer que los sitios componentes están dispersos físicamente – quizá también dispersos geográficamente -, aunque de hecho basta con que estén dispersos lógicamente.
- Dos “sitios” pueden incluso coexistir en la misma máquina física.

5

**Bases de Datos Distribuidas****Ventajas:**

- ¿Por qué son necesarias las BDD?
- La respuesta es que las empresas ya están generalmente distribuidas al menos de manera lógica (en divisiones, departamentos, grupos de trabajo, etc.)
- Y es muy probable que también lo estén de manera física (en plantas, fábricas, laboratorios, etc.);

6

**Bases de Datos Distribuidas****Ventajas:**

- De esto deducimos que por lo general también los datos ya están distribuidos
- Ya que cada unidad organizacional dentro de la empresa mantendrá los datos que son importantes para su propia operación
- Por lo tanto, el valor de la información total de la empresa está dividido en lo que a veces llamamos “islas de información”

7

**Bases de Datos Distribuidas****Ventajas:**

- Lo que hace un Sistema Distribuido es proporcionar los puentes necesarios para conectar a esas islas entre sí
- En otras palabras, permite que la estructura de la BD refleje la estructura de la empresa – los datos locales son conservados localmente en el lugar donde pertenecen de manera más lógica –
- Y al mismo tiempo, permite tener acceso a datos remotos cuando sea necesario.

8

**Bases de Datos Distribuidas****Ventajas:**

- Favorecer la naturaleza distribuidora de muchas aplicaciones, no solamente a nivel local sino incluso en diferentes lugares.
- Se consigue una compartición de los datos, sin perder un determinado control local.
- Crecimiento modular.
- Aumento de la confiabilidad y la disponibilidad.
- Economía: es más barato construir un sistema con pequeñas computadoras que uno grande si ambos dan el mismo rendimiento.
- La autonomía de estos sistemas es grande.

Universidad de Castilla-la Mancha Base de Datos Distribuidas Corbera - Delgado

9

**Bases de Datos Distribuidas****Ventajas:**

- El rendimiento se mejora. Cuando se distribuye una gran base de datos por múltiples sitios, las consultas locales y las transacciones tienen mejor rendimiento porque las bases de datos locales son más pequeñas. A parte de esta distribución, se puede conseguir lo siguiente en estos sistemas:
  - Reducir el número de transacciones ejecutándose por sitio.
  - Un paralelismo entre las consultas ejecutando varias en sitios diferentes, descomponiendo una de ellas en subconsultas que puedan ejecutarse en paralelo.

Universidad de Castilla-la Mancha Base de Datos Distribuidas Corbera - Delgado

10

**Bases de Datos Distribuidas****Desventajas:**

- Hay una menor seguridad en cuanto al control de acceso a los datos: control de replicas y errores que puedan producirse en la red.
- Mayor complejidad en el diseño e implementación del sistema. Además si la replicación de datos no se hace de forma adecuada, las ventajas se pueden transformar en desventajas.
- Excesivos costes en el intento de conseguir la transparencia mencionada anteriormente.
- Falta de estándares y de experiencia, una vez más en estos modelos avanzados de BD.
- No se puede garantizar al 100 % el rendimiento y la fiabilidad.

11

**Bases de Datos Distribuidas****Ejemplos de algunos DDBMS:****Prototipos**

- SDD-1, que fue construido en la división de investigación de Computer Corporation of America.
  - Finales de los años 70 y principios de los 80.
- R\*, una versión distribuida del prototipo System R, construida en IBM Research
  - Principios de los años 80
- Distributed Ingres, una versión distribuida del prototipo Ingres, construida en la Universidad de California en Berkeley
  - Principios de los 80

12

**Bases de Datos Distribuidas****Ejemplos de algunos DDBMS :**

Implementaciones comerciales:

- La mayoría de los productos SQL actuales proporcionan algún tipo de soporte de BDD (con diversos grados de funcionalidad).
  - Ingres/Star, el componente de BDD de Ingres
  - La opción de BDD de Oracle
  - La propiedad de datos distribuidos de DB2
  - Opciones de BDD de SQL Server

13

**Bases de Datos Distribuidas****Ejemplos de algunos DDBMS:**

Tanto los DDBMS prototipos como los productos, son relacionales (todos soportan SQL).

Además, hay varias razones por las cuales, para que un sistema distribuido sea exitoso, **debe ser** relacional.

La tecnología relacional es un requisito previo para tecnología distribuida

14

**Bases de Datos Distribuidas****Principio Fundamental**

***“Ante el usuario, un sistema distribuido debe lucir exactamente igual que un sistema que no es distribuido”***

- En otras palabras, los usuarios de un sistema distribuido deben ser capaces de comportarse exactamente como si *no* fuera distribuido.
- Todos los problemas de los sistemas distribuidos son, o deberían ser, problemas internos o en el nivel de implementación, y no externos o en el nivel de usuario.

15

**Bases de Datos Distribuidas**

**El principio fundamental nos conduce a 12 reglas u objetivos:**

- 1.- Autonomía local.
- 2.- No dependencia de un sitio central.
- 3.- Operación continua.
- 4.- Independencia de ubicación
- 5.- Independencia de fragmentación
- 6.- Independencia de replicación
- 7.- Procesamiento de consultas distribuidas
- 8.- Administración de transacciones distribuidas
- 9.- Independencia de hardware
- 10.- Independencia de sistema operativo
- 11.- Independencia de red
- 12.- Independencia de DBMS

16



**Bases de Datos Distribuidas**

**El principio fundamental nos conduce a 12 reglas u objetivos:**

**1.- Autonomía local.** Los sitios en un sistema distribuido deben ser **autónomos**.

- La autonomía local significa que todas las operaciones en un sitio dado están controladas por ese sitio; ningún sitio **X** debe depender de algún otro sitio **Y** para su operación satisfactoria.
- La seguridad, integridad y representación de almacenamiento de los datos locales permanecen bajo el control y jurisdicción del sitio local.

*M. en C. Anastacio Antolino Hernández*

17

**Bases de Datos Distribuidas**

**12 reglas u objetivos:**

**2.- No dependencia de un sitio central.** La autonomía local implica que todos los sitios deben ser tratados como iguales.

- Por lo tanto, no debe haber particularmente ninguna dependencia de un sitio “maestro” central para algún servicio central, tal que todo el sistema dependa de ese sitio central.
- Razones por las cuales no debería haber un sitio central:
  - El sitio central puede ser un cuello de botella
  - El sistema sería vulnerable; es decir, si el sitio central falla, también fallará todo el sistema

18

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**3.- Operación continua.** Una ventaja de los sistemas distribuidos es que deben proporcionar mayor *confiabilidad* y mayor *disponibilidad*.

- **Confiabilidad.** La probabilidad de que el sistema esté listo y funcionando en cualquier momento dado. Los SD pueden continuar operando cuando hay alguna falla en algún componente independiente.
- **Disponibilidad.** La probabilidad de que el sistema esté listo y funcionando continuamente a lo largo de un período especificado.

19

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**4.- Independencia de ubicación.** Conocida también como *transparencia de ubicación*.

- Los usuarios no tienen que saber dónde están almacenados físicamente los datos, sino que deben ser capaces de comportarse como si todos los datos estuvieran almacenados en su propio sitio local.
- Esto simplifica los programas de los usuarios. En particular, permite que los datos emigren de un sitio a otro sin invalidar ninguno de estos programas o actividades.

20

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**5.- Independencia de fragmentación.** Un sistema soporta la *fragmentación de datos* cuando puede ser dividida en o partes o fragmentos, para efectos de almacenamiento físico.

- La fragmentación es necesaria por razones de rendimiento: los datos pueden estar almacenados en la ubicación donde son usados más frecuentemente para que la mayoría de las operaciones sean locales y se reduzca el tráfico en la red.
- Los usuarios deben comportarse como si los datos en realidad estuvieran sin fragmentación alguna.

21

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**6.- Independencia de replicación.** El sistema soporta *replicación de datos* cuando un *fragmento* puede ser representado por muchas copias distintas, o réplicas, guardadas en muchos sitios distintos.

Las réplicas son necesarias por dos razones principales:

1. Significan un mejor *rendimiento* (las aplicaciones pueden operar sobre las copias locales en lugar de tener que comunicarse con sitios remotos)

22

**Bases de Datos Distribuidas****12 reglas u objetivos:****6.- Independencia de replicación...**

2. Pueden significar una mejor *disponibilidad* (un objeto replicado permanece disponible para su procesamiento, mientras esté disponible al menos una copia).

Por supuesto, la principal desventaja de las réplicas es que al actualizarlas es necesario actualizar todas: el ***problema de la propagación de la actualización***.

23

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**7.- Procesamiento de consultas distribuidas.** La optimización es mucho más importante en un sistema distribuido que en uno centralizado.

- El punto básico es que en una consulta que involucra a varios sitios, habrá muchas formas posibles de mover los datos en el sistema para satisfacer la solicitud, y es crucialmente importante que se encuentre una estrategia eficiente.

24

**Bases de Datos Distribuidas****12 reglas u objetivos:****8.- Administración de transacciones distribuidas.**

Existen dos aspectos principales en la administración de transacciones: *control de recuperación* y *control de la concurrencia*.

- Ambos aspectos requieren un tratamiento amplio en el ambiente distribuido.
- Ya que una sola transacción puede involucrar la ejecución de código en muchos sitios.

25

**Bases de Datos Distribuidas****12 reglas u objetivos:****8.- Administración de transacciones distribuidas...**

- Puede involucrar actualizaciones en muchos sitios y se debe de cuidar que la transacción no caiga en un bloqueo mortal (basado en el **bloqueo**).
- Para el control de la *recuperación*, es necesario asegurarse que una transacción dada sea atómica en el ambiente distribuido, el sistema debe por lo tanto asegurarse de que la transacción sea confirmada o deshecha (se puede utilizar el protocolo de **confirmación de dos fases**).

26

**Bases de Datos Distribuidas****12 reglas u objetivos:**

**9.- Independencia de hardware.** Soporte para un gran número de máquinas diferentes. Poder integrar todos los datos de todos estos sistemas y presentar al usuario una “imagen del sistema único”.

**10.- Independencia de sistema operativo.** Obviamente es necesario no sólo tener la posibilidad de ejecutar el mismo DBMS en diferentes plataformas de hardware, sino también ejecutarlo en diferentes plataformas de sistema operativo.

27

**Bases de Datos Distribuidas****12 reglas u objetivos:****11.- Independencia de red.**

Si el sistema va a tener la posibilidad de soportar muchos sitios distintos es obviamente necesario tener la posibilidad de soportar también una variedad de redes de comunicación distintas.

28

**12 reglas u objetivos:**

**12.- Independencia de DBMS.** Lo que se necesita es que *todos* los ejemplares de DBMS en sitios diferentes *soporten la misma interfaz*.

- Aunque no tienen que ser necesariamente copias del mismo software DBMS.
- En otras palabras, sería posible que el sistema distribuido fuera **heterogéneo**, al menos en cierto grado.
- Sería muy bueno si diferentes DBMS pudieran participar de alguna forma en un sistema distribuido.