

UNIVERSIDAD, TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

ESCUELA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

ANTEPROYECTO DE TESIS

TEMA:

**“FRAGMENTACIÓN DE TABLAS EN BASES DE DATOS
DISTRIBUIDAS”**

APLICATIVO:

**“IMPLEMENTACIÓN DE UNA BASE DE DATOS DISTRIBUIDA EN EL
SISTEMA ACADÉMICO DE LA FICA”**

AUTORA:

VERÓNICA ALEXANDRA LÓPEZ T.

DIRECTOR:

ING. MIGUEL ORQUERA

Ibarra, Diciembre 2007



DISEÑO DEL ANTEPROYECTO DE TESIS

TEMA:

**“FRAGMENTACIÓN DE TABLAS EN BASES DE
DATOS DISTRIBUIDAS”**

APLICATIVO:

**“IMPLEMENTACIÓN DE UNA BASE DE DATOS
DISTRIBUIDA EN EL SISTEMA ACADÉMICO DE LA
FICA”**



EL PROBLEMA:

En la actualidad las bases de datos nos han facilitado la vida. Hoy en día en cualquier institución o empresa se ha vuelto indispensable conectarse a la Bases de Datos que contenga información útil e intercambiar, aumentar o modificar datos entre ellas, como por ejemplo la tabla de notas de un estudiante específico, la cual se la obtendría fácilmente realizando una consulta de las tablas implicadas.

El problema radica en la sobrecarga de base de datos en un solo servidor, ya que la FICA posee un sistema centralizado, y sus datos se encuentran aislados totalmente, debido a que cada escuela posee su propia base de datos, dificultando así las consultas entre bases de datos. Lo ideal sería tener una Base de Datos Distribuida que pertenezcan lógicamente a un sólo sistema, pero se encuentre físicamente esparcido en varios sitios de la red.

Cada sitio es un sistema de base de datos en sí mismo, con el fin de que un usuario de cualquier sitio de la red pueda obtener acceso a los datos, tal como si todos los datos estuvieran almacenados en el sitio propio del usuario, es en realidad una especie de objeto virtual, cuyos datos se almacenan físicamente en varias bases de datos reales pero ubicadas en diferentes sitios, es decir la unión lógica de esas bases de datos.

Por lo tanto el principal problema que se plantea es la fragmentación de tablas, el mayor reto es identificar la fragmentación ideal para los datos en si, además de resolver los grados de fragmentación.



La fragmentación es viable debido a que los datos pueden almacenarse en localidades donde son utilizados con mayor frecuencia, de tal manera que la mayor parte de las operaciones sean sólo locales lo cual reducirá el tráfico en la red.

JUSTIFICACIÓN:

Este tema trata de simplificar las consultas de una base de datos, para poder acceder desde cualquier sitio, en cualquier punto de la red tal como si todos los datos estuvieran almacenados en el sitio mismo.

Para esto debemos construir sistemas distribuidos de bases de datos, ya que se están utilizando cada vez más al igual que se usan las arquitecturas de cliente-servidor. Para lo cual debemos tener claro lo que es una Base de Datos Distribuida, la estructura de ésta se basa en las comunicación entre el conjunto de localidades.

Hay que construir el prototipo del sistema distribuido de bases de datos que incluyen compartir la información, fiabilidad, disponibilidad y agilizar el procesamiento de las consultas.

Se Propone una solución al aislamiento de los datos y al tráfico de la red, con un rediseño del sistema académico de la FICA, con una Base de Datos Distribuida, la cual cada carrera tendrá en su servidor los datos solo de ésta, tanto de estudiantes como profesores, materias y las que sean necesarias. La Administración de la Base de Datos será Centralizada, ya que se trata de una sola Base de Datos, cuyas tablas están particionadas, se investigará también las diferentes opciones de distribución de datos y se obtendrá la mejor, según las necesidades de la Facultad.



OBJETIVOS

General

- Análisis de la fragmentación de tablas en Bases de Datos Distribuidas y su influencia en el incremento de rendimiento y accesibilidad de los datos.

Específicos

- Analizar las diferentes Arquitecturas de Bases de Datos Distribuidas
- Investigar sobre los diferentes tipos de fragmentación de tablas y sus beneficios en el rendimiento y accesibilidad.
- Determinar los principales aspectos de la Administración de la Bases de Datos Distribuidas.
- Analizar las diferentes formas de replicación utilizadas en Bases de Datos Distribuidas.
- Determinar una arquitectura adecuada de Base de Datos Distribuida de acuerdo a los requerimientos del sistema académico de la FICA.
- Implementar la base de datos distribuida en el sistema académico de la FICA.



MARCO TEÓRICO

Las Bases de Datos en la actualidad son muy importantes debido a que en ellas se guarda gran información de suma importancia

Ventajas de las bases de datos

- Independencia de datos y tratamiento.
- Coherencia de resultados.
- Cumplimiento de ciertas normas.

BASES DE DATOS DISTRIBUIDAS.

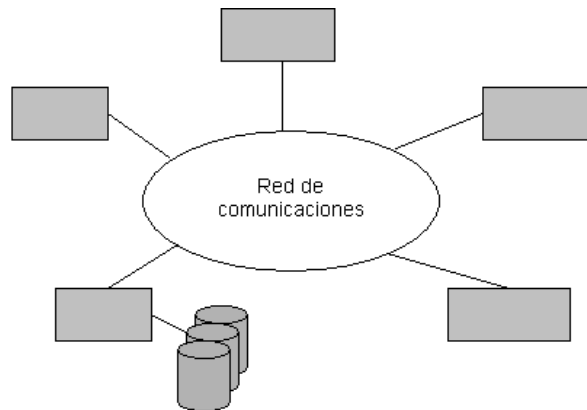
Es una colección de datos que pertenecen lógicamente a un sólo sistema, pero se encuentra físicamente esparcido en varios "sitios" de la red. Un sistema de base de datos distribuidos se compone de un conjunto de sitios, conectados entre sí mediante algún tipo de red de comunicaciones, en el cual:

- Cada sitio es un sistema de base de datos en sí mismo.
- Los sitios han convenido en trabajar juntos 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 del usuario.

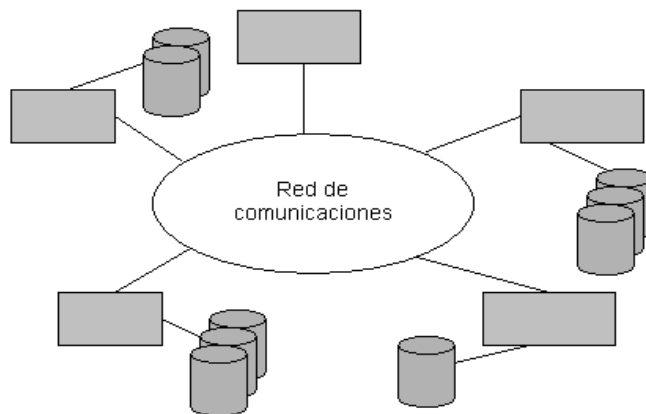


BASES DE DATOS DISTRIBUIDAS Vrs. CENTRALIZADA

Los sistemas de bases de datos centralizados, residen en una sola localidad.



Los sistemas de bases de datos distribuidos se encuentran en varias localidades.



Ventajas Con respecto a Sistemas Centralizados

- Una de las ventajas de los sistemas distribuidos es la economía, pues es mucho más barato, añadir servidores y clientes cuando se requiere aumentar la potencia de procesamiento.
- El trabajo en conjunto.



- Tienen una mayor confiabilidad. Al estar Distribuidas la carga de trabajo, la falla de una de ellas no afecta a las demás.
- Capacidad de crecimiento incremental. Se puede añadir procesadores al sistema en forma gradual según sus necesidades.

SISTEMAS DE BASES DE DATOS DISTRIBUIDAS

- **Una base de datos Distribuidas (BDD)** es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes sitios por una red de comunicaciones.
- **Un sistema de bases de datos Distribuidas (SBDD)** es un sistema en el cual múltiples sitios de bases de datos están ligados por un sistema de comunicaciones, de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si los datos estuvieran almacenados en su sitio propio.
- **Un sistema de manejo de bases de datos Distribuidas (SMBDD)** es aquel que se encarga del manejo de la BDD y proporciona un mecanismo de acceso que hace que la distribución sea transparente a los usuarios. El término transparente significa que la aplicación trabajaría, desde un punto de vista lógico, como si un solo SMBD ejecutado en una sola máquina.
- **Un sistema de base de datos Distribuidas (SBDD)** es el resultado de la integración de una base de datos Distribuidas con un sistema para su manejo.



Diseño de la base de dato Distribuidas.

Cuando se busca eficiencia en el acceso de la información se deben considerar:

- Como fragmentar la información.
- Como asignar cada fragmentación entre los diferentes sitios de la red.
- Si la información esta repetida.

FRAGMENTACIÓN

Una relación que corresponde a una tabla, consiste en dividirla en fragmentos menores, para ello hay las siguientes alternativas lógicas, como son: fragmentación horizontal, fragmentación vertical y fragmentación mixta.

TIPOS DE FRAGMENTACIÓN

Fragmentación Horizontal

La división o fragmentación horizontal trabaja sobre las tuplas, dividiendo la relación en subrelaciones que contienen un subconjunto de las tuplas que alberga la primera.

Fragmentación Vertical

Se basa en los atributos de la relación para efectuar la división. Estos dos tipos de partición podrían considerarse los fundamentales y básicos. Sin embargo, existen otras alternativas. Fundamentalmente, se habla de fragmentación mixta o híbrida cuando el proceso de partición hace uso de los dos tipos anteriores.



Fragmentación Mixta.

La fragmentación mixta puede llevarse a cabo de tres formas diferentes: desarrollando primero la fragmentación vertical y, posteriormente, aplicando la partición horizontal sobre los fragmentos verticales (denominada partición VH), o aplicando primero una división horizontal para luego, sobre los fragmentos generados, desarrollar una fragmentación vertical (llamada partición HV), o bien, de forma directa considerando la semántica de las transacciones.

Réplica: El sistema conserva varias copias o réplicas idénticas de una tabla. Cada réplica se almacena en un nodo diferente.

Ventajas:

Disponibilidad: el sistema sigue funcionando aún en caso de caída de uno de los nodos.

Aumento del paralelismo: Varios nodos pueden realizar consultas en paralelo sobre la misma tabla. Cuantas más réplicas existan de la tabla, mayor será la posibilidad de que el dato buscado se encuentre en el nodo desde el que se realiza la consulta, minimizando con ello el tráfico de datos entre nodos.

Réplica y fragmentación de datos

Las técnicas de réplica y fragmentación se pueden aplicar sucesivamente a la misma relación de partida. Un fragmento se puede replicar y a su vez esa réplica ser fragmentada, para luego replicar alguno de esos fragmentos.

Administrador de la Base de Datos. Es la persona encargada de definir y controlar las bases de datos, además proporciona asesoría a los usuarios. Es la persona o equipo de personas profesionales responsables del control y manejo del sistema de base de datos.



HIPÓTESIS

La implementación de una Base de Datos Distribuidas mejora el rendimiento y la accesibilidad de los datos, debido a una disminución de tráfico de la red y a un trabajo colaborativo de varios servidores como si fuera un solo.

METODOLOGÍA

- Como primera fase se inicia el proyecto con una investigación referente a Base de Datos Distribuidas, la cual será el pilar fundamental en el desarrollo de la misma. Para esto utilizaré Internet, fuentes bibliográficas y otras.
- Como segunda fase se realiza una investigación breve a cerca de los Sistemas de Base de Datos Centralizada, para luego compararlo con los Sistemas de Base de Datos Distribuidas, la que nos dará una idea clara del porque utilizar Base de Datos Distribuidas.
- Como tercera fase se realiza un estudio de los diferentes tipos de fragmentación, como es: fragmentación horizontal y fragmentación vertical, las que nos ayudaran a optar por la mejor alternativa, para la fragmentación de tablas.
- Como cuarta fase estudiaremos la transparencia de los datos, como es de localización, de fragmentación y de réplica, esta última se refiere a la recuperación y la tolerancia a fallos.



- Como quinta fase se realizará una investigación acerca de la Administración de la Base de datos, como es: definir y controlar las bases de datos, además proporcionar asesoría a los usuarios.
- Como sexta fase será el diseño e “Implementación de una Base de Datos Distribuidas en el Sistema Académico de la FICA”, la cual se demostrará realizando este presente aplicativo con los conocimientos anteriormente investigados, a lo que será sometido a diferentes pruebas y a sus debidas correcciones.
- Como séptima etapa se puede derivar las diferentes conclusiones y recomendaciones del presente proyecto desarrollado.

Una vez concluido con estas etapas podemos realizar toda la documentación necesaria la que apoyará los conocimientos teórico - prácticos de nuestra investigación.

ALCANCE

La mayoría de empresas, organizaciones, e instituciones se basan en la información que se encuentra en las bases de datos y otras fuentes muy diversas, tanto internas como externas.

En el presente trabajo vamos a investigar las diferentes funciones de datos, como trabajan y se integran las tablas, las réplicas entre los diferentes servidores. Los tipos de fragmentación, las que nos ayudarán a optar por la mejor alternativa, de igual manera los grados de fragmentación.



Se investigará la Administración y se obtendrá cual es la mejor manera de Administrar una Base de Datos Distribuidas.

La siguiente propuesta será demostrada en la ejecución del aplicativo planteado como es: “IMPLEMENTACIÓN DE UNA BASE DE DATOS DISTRIBUIDAS EN EL SISTEMA ACADÉMICO DE LA FICA”. En lo que utilizaremos el Sistema de Base de Datos Oracle 11, ya que se ha adoptado en la Universidad para soportar el Sistema Académico, debido a que tiene una amplia gama de funcionalidades como es: mecanismos de seguridad, realización de copias de seguridad y recuperación, control de espacio, conectividad de carácter abierto y herramientas de desarrollo.

TABLA DE CONTENIDOS

1. CAPITULO I

Base de Datos Distribuidos

- 1.1. Introducción
 - 1.1.12 Antecedentes
 - 1.1.13 Almacenamiento Distribuido
- 1.2. Sistemas de gestión de bases de datos distribuidos
 - 1.2.1 Funciones y Arquitectura de un SGBDD
 - 1.2.2 Funciones de un SGBDD
- 1.3. Las Doce Reglas
- 1.4. Ventajas e Inconvenientes
- 1.5. Base de Datos Distribuidas Vrs Centralizada
 - 1.6.1 Base de Datos Centralizada.
 - 1.6.2 Ventajas e Inconvenientes.



2. CAPITULO II

Fragmentación de Tablas

- 2.1. Introducción
 - 2.1.1 Definición
- 2.2. Tipos de Fragmentación
 - 2.2.1 Fragmentación Horizontal.
 - 2.2.2 Fragmentación Vertical.
 - 2.2.3 Fragmentación Mixta.
 - 2.2.4 Grados de fragmentación

3. CAPITULO III

Arquitectura de Distribución de Datos.

- 3.1. Transparencia
 - 3.1.1 Transparencia de localización
 - 3.1.2 Transparencia de fragmentación
 - 3.1.1. Transparencia de réplica
- 3.2. Formas de Réplica

4. CAPITULO IV

Administración de Bases de Datos Distribuidas

- 4.1 Conceptos Generales
 - 4.1.1 Administración la Estructura de la Base de Datos
 - 4.1.2 Administración de la Actividad de Datos
 - 4.1.3 Administrar el Sistema Manejador de la Base de Datos
 - 4.1.4 Confirmar la Seguridad de la Base de Datos
- 4.2 Objetivos del Administrador de la Base de Datos
 - 4.2.1. Integridad de Datos
 - 4.2.2. Seguridad y Disponibilidad de Datos
- 4.3 Funciones del Administrador.
 - 4.3.1. Creación de Usuarios
 - 4.3.2. Creación de Bases de Datos y Tablas



- 4.3.3. Restricciones de Integridad
- 4.3.4. Respaldos y recuperaciones

5. CAPITULO V

Desarrollo de la Aplicación

- 5.1. Modelado
- 5.2. Desarrollo.
- 5.3. Implementación.
- 5.4. Documentación

6. CAPITULO VI

Conclusiones y Recomendaciones

- 6.1. Glosario de términos.
- 6.2. Conclusiones.
- 6.3. Recomendaciones.
- 6.4. Posibles Temas de Tesis.
- 6.5. Bibliografía.



ANEXOS

COSTOS

Tabla de Costos para el desarrollo del Proyecto.

Materiales	Valor del Material	Valor a Gastar
HARDWARE		
Computador	1120	0
Intel Pentium 4 2.4 Ghz		
1 G de Memoria RAM		
Disco de 80 Gb.		
Monitor 17"		
Impresora HP deskjet 656c		
Teclado y Mouse PS/2		
SOFTWARE		
Windows XP Profesional	400	0
Office 2003	250	0
Oracle 11	460	0
Software Adicional	100	0
Capacitación	80	80
BIBLIOGRAFÍA		
Libros	30	30
Internet	150	150
Revistas y Folletos	30	30
Copias	20	20
OTROS MATERIALES		
Cartuchos de Impresión	50	50
Utiles de Oficina	50	50
SUBTOTAL	2740	410
Imprevistos 20%	200	200
TOTAL	2940	610



BIBLIOGRAFÍA

INTERNET.

- http://www.lafacu.com/apuntes/informatica/base_datos/default.htm#Introducción
- <http://arraquis.dif.um.es/~rafa/bd1.htm>
- <http://sacbeob.8m.com/tutoriales/bddistribuidas/sacbeob.8m.com>
- http://tecnomaestros.awardspace.com/bases_datos_distribuidas.php
- <http://dmi.uib.es/~bbuades/sistdistr/sld007.htm>
- <http://members.fortunecity.es/lrmdl/SO7.htm#VSDRC>
- <http://sacbeob.8m.com/tutoriales/bddistribuidas/>
- <http://pdf.rincondelvago.com/bases-de-datos-distribuidas.html>
- http://www.dia.eui.upm.es/cgi-bin/asigfram.pl?cual=sis_dis&nombre=Sistemas-Distribu%EDdos

Libros

- Fundamentos de Bases de Datos, 4ª edición
Silberschatz, A., Korth, H.F. y Sudarshan, S. McGraw-Hill. 2002
- Sistemas de bases de datos. Conceptos fundamentales 2ª edición
Elmasri, R. y Navathe, S.B. Addison-Wesley. 1997
- George Coulouris. Sistemas Distribuidos. Tercera Edición. Addison Wesley. Madrid. 2001.



ÍNDICE

TEMA	PAGINAS
Tema del Anteproyecto	1
El Problema	2
Justificación	3
Objetivos.....	4
Marco Teórico.....	5
Hipótesis.....	10
Metodología.....	10
Alcance.....	11
Tabla de Contenidos	12
Costos.....	15
Bibliografía	16