
DESARROLLO DE LA APLICACIÓN PARA EL MANEJO DE OBJETOS DE BASES DE DATOS EN SITIOS DISTRIBUIDOS.

200011006 – Adolfo Francisco Lopez Cuzco

Resumen

En la actualidad el uso de bases de datos en diferentes sitios es extensa y por consiguiente el costo para el procesamiento de toda la información es elevada por lo que se busca minimizar el procesamiento de datos y con ello minimizar el costo total de dicho procesamiento de datos, para ello se busca de sistemas eficientes que controlen el flujo de datos, los data center, las redes y los servicios a los usuarios.

Para resolver el problema se desarrolla una aplicación utilizando un TDA o tipo de dato abstracto, este tipo de dato es una lista circular simplemente enlazada.

Palabras clave

Bases de datos
Tuplas
Atributos
Entidad
HP-Hard

Abstract

At present the use of databases in different sites is extensive and therefore the cost for the processing of all the information is high, so it is sought to minimize data processing and thereby minimize the total cost of such data processing. For this, efficient systems are sought to control the flow of data, data centers, networks and services to users.

To solve the problem, an application is developed using an ADT or abstract data type, this data type is a simply linked circular list.

Keywords

Databases
Tuples
Attributes
Entity

Introducción

Para el manejo de alojamiento de objetos de bases es necesario optimizar los recursos para poder reducir los costos por el procesamiento de datos.

Para el estudio se debe conocer sobre la forma en que están estructuradas, modeladas o diseñadas las distribuciones de datos.

Existen modelos que se pueden aplicar para encontrar un patrón de uso de las bases de datos

Para el desarrollo de la aplicación se utilizará el método de agrupamiento el cual esta basado en representar el patrón en una matriz de frecuencias y luego representar dicha matriz por su equivalente matriz binaria luego reducirla para obtener una matriz optimizada.

Desarrollo del tema

Existen buenas razones técnicas para distribuir datos. La mas obvia es la referente a la sobrecarga de canales de entrada y salida a los discos en donde se almacena finalmente la información. Es mucho mejor distribuir los accesos a la información sobre diferentes canales que concentrarlos en uno solo. Otra razón de peso es que las redes de computadoras empezaron a trabajar a velocidades razonable abriendo la puerta a la distribución del trabajo y la información, todo esto ha llevado a que se debe tener un mejor esquema en el alojamiento de información a través de bases de datos y el manejo de procesamiento de dichos datos es un tema muy importante en la industria de la tecnología del procesamiento de datos se derivan diversos varios temas importantes como la capacidad de

procesamiento todo esto conlleva un costo por tal efecto se necesita optimizar este costo.

Para el desarrollo de la solución del problema es necesario saber como funciona el procesamiento de datos y el manejo de objetos de datos al igual que el esquema.

a. Bases de datos

Un objeto de base de datos es una entidad de una base de datos, esta entidad puede ser un atributo, un set de tuplas, una relación o un archivo.

Los objetos de base de datos son unidades independientes que deben ser alojadas en los sitios de una red

La definición formal para los objetos, sitios y consultas son.

$$O = \{O1, O2, \dots, Ono\},$$

Donde O es el set de objetos

$$S = \{S1, S2, \dots, Sns\},$$

Donde S es el set de sitios

$$Q = \{q1, q2, \dots, qnq\},$$

Donde Q es el set de consultas.

b. Esquema de alojamiento

El esquema de alojamiento es importante ya que nos ayuda a como distribuir datos en los nodos de ambiente distribuido de manera optima que es lo que se busca.

El problema de diseño de distribución consiste en determinar el alojamiento de datos de forma que los costos de acceso y comunicación son minimizados. Como muchos otros problemas reales, es un problema combinatorio NP-Hard.

c. NP- Hard

En teoría de la complejidad computacional, la clase de complejidad NP-hard (o NP-complejo, o NP-difícil) es el conjunto de los problemas de decisión que contiene los problemas H tales que todo problema L en NP puede ser transformado polinomialmente en H. Esta clase puede ser descrita como aquella que contiene a los problemas de decisión que son como mínimo tan difíciles como un problema de NP. Esta afirmación se justifica porque si podemos encontrar un algoritmo A que resuelve uno de los problemas H de NP-hard en tiempo polinómico, entonces es posible construir un algoritmo que trabaje en tiempo polinómico para cualquier problema de NP ejecutando primero la reducción de este problema en H y luego ejecutando el algoritmo A.

Un método propuesto para resolver este tipo de problemas consiste en aplicar una metodología de agrupamiento.

Para “nt” tuplas y “ns” sitios, el método consiste en tener la matriz de frecuencia de acceso en los sitios $F[nt][ns]$ de la instancia objetivo, transformarla en una matriz de patrones de acceso y agrupar las tuplas con el mismo patrón.

El patrón de acceso para una tupla es el vector binario indicando desde cuál sitio la tupla es accedida.

Por ejemplo si se tiene la siguiente matriz.

Par poder resolver este problema se desarrollará una aplicación que determine esta matriz binaria

y luego la matriz reducida. Esto se hará aplicando un tipo de dato abstracto para el manejo de la estructura de datos.

Este tipo de dato es una lista circular simplemente enlazada.

d. TDA

Un Tipo de dato abstracto (en adelante TDA) es un conjunto de datos u objetos al cual se le asocian operaciones. El TDA provee de una interfaz con la cual es posible realizar las operaciones permitidas, abstrayéndose de la manera en cómo estén implementadas dichas operaciones. Esto quiere decir que un mismo TDA puede ser implementado utilizando distintas estructuras de datos y proveer la misma funcionalidad.

El paradigma de orientación a objetos permite el encapsulamiento de los datos y las operaciones mediante la definición de clases e interfaces, lo cual permite ocultar la manera en cómo ha sido implementado el TDA y solo permite el acceso a los datos a través de las operaciones provistas por la interfaz.

En este capítulo se estudiarán TDA básicos como lo son las listas, pilas y colas, y se mostrarán algunos usos prácticos de estos TDA.

e. TDA lista

Una lista se define como una serie de N elementos E_1, E_2, \dots, E_N , ordenados de manera consecutiva, es decir, el elemento E_k (que se denomina elemento k-ésimo) es previo al

elemento E_{k+1} . Si la lista contiene 0 elementos se denomina como lista vacía.

Las operaciones que se pueden realizar en la lista son: insertar un elemento en la posición k , borrar el k -ésimo elemento, buscar un elemento dentro de la lista y preguntar si la lista esta vacía.

f. Lista circular simplemente enlazada:

No existe ningún nodo que apunte a null. La lista no tiene fin ya que al llegar al último nodo empieza de nuevo la lista. Se accede a la lista mediante el primer nodo o también llamado inicio de la lista

Conclusiones

El manejo del alojamiento de información de objetos de bases de datos en sitios distribuidos genera gran consumo de recursos y por ende un costo elevado.

Para optimizar el costo es necesario optimizar también el modelo o el esquema de alojamiento de datos.

Referencias bibliográficas

C. J. Date, (1991). *An introduction to Database Systems*. Addison-Wesley Publishing Company, Inc.