
Minimización de Costos en la Transmisión de Información en Bases de Datos Distribuidas

202300539 – José Emanuel Monzón Lémus

Resumen

El presente ensayo aborda la optimización en el alojamiento de objetos de bases de datos en redes distribuidas, siendo este un tema muy relevante en la computación en la nube y la gestión de grandes volúmenes de datos. Este tipo de problemas afecta directamente en la eficiencia y costo de operaciones de las empresas e instituciones que dependen del acceso rápido y seguro a los datos.

Las para poder implementar soluciones al problema planteado se utilizó un enfoque sobre la replicación de datos y estrategias de minimización de costos a través de técnicas de agrupamiento de datos. Lo anteriormente mencionado fue logrado a través de la creación de estructuras de almacenamiento de datos por medio de TDA's y apuntadores, con los cuales se manejará la información de la mejor forma posible y con ello se consiguiera reducir el uso de recursos utilizados para procesar esta información al momento de ser enviada. Todo ello por medio de procesamiento de los datos desde archivos de entrada y salida en los cuales la optimización fue realizada.

Palabras clave

1. Optimización
2. Almacenamiento de datos
3. Redes distribuidas
4. Replicación de datos
5. Agrupamiento

Abstract

This essay addresses the optimization of object storage in distributed database networks, a topic highly relevant to cloud computing and the management of large data volumes. Such issues directly impact the efficiency and cost of operations for companies and institutions that rely on quick and secure data access.

To implement solutions to the problem, an approach focused on data replication and cost minimization strategies through data clustering techniques was utilized. This was achieved by creating data storage structures using Abstract Data

Types (ADTs) and pointers, allowing for optimal data management and reducing the resources required for processing this information during transmission. All this was accomplished through the processing of data obtained in input files and returned in output files where the optimization was performed.

Keywords

1. Optimization
2. Data Storage
3. Distributed Networks
4. Data Replication
5. Clustering

Introducción

El problema propuesto para este proyecto se basa en la búsqueda de una solución con la cual se logre obtener una optimización de envío y recepción de información en una red distribuida de máquinas a través de redes, esto con el objetivo de obtener una disminución en los gastos y con ello lograr el mínimo posible, esto siendo muy importante en distintas áreas, ya que cualquier empresa busca producir más, mejor y con el menor gasto posible.

Por ello se planteó una solución por medio de la utilización de estructuras propias con las cuales se pudiera almacenar y posteriormente procesar la información para que posteriormente se pueda obtener datos mas compactos que puedan y por consiguiente ser utilizados de una forma eficiente y rápida.

Desarrollo del tema

En primer lugar, como se mencionó anteriormente la solución para el planteamiento propuesto se basa en

el almacenamiento de la información en formato de matrices, siendo estas estructuras con las que se realizaran todos los procesos para poder reducir el uso de recursos en el procesamiento de archivos con formato XML.

a. Lectura de archivos XML

La lectura de este tipo de archivos se basa en la agrupación de información por medio de etiquetas, las cuales pueden tener el nombre y orden que sea más apropiado según el uso para el que se requiera. Los archivos XML al igual que HTML es un lenguaje de marcado de hipertexto, esto quiere decir que un lenguaje formado por etiquetas, pero en este caso como se mencionó anteriormente puede ser asignado cualquier nombre y atributos a cada una de sus etiquetas. Los archivos XML cuentan con una estructura que puede ser considerada como un árbol donde la etiqueta principal es el troco del árbol y las sub etiquetas son las ramas las cuales pueden contener otra sub etiqueta u hoja que es la etiqueta donde se almacena la información que puede ser utilizada o podría ser procesada como en este caso para optimizarla, para este caso las ramas y subramas fueron utilizadas para poder formar estructuras de matrices con el objetivo de ser procesadas posteriormente, todas estas matrices fueron almacenadas dentro de una lista circular para poder acceder a todas las matrices en cualquier momento.

b. Generación de matrices

La generación de las matrices fue basada en el recorrido de del árbol (etiqueta principal)

de donde se obtuvieron cada una de las ramas (etiquetas contenedoras) de donde se obtuvieron el número de filas y columnas correspondientes a la matriz además de su nombre para identificar cada matriz, posterior se realizó un recorrido dentro de las ramas para poder acceder a las hojas ósea el contenido o dato almacenado, siendo este el dato que formara un nodo de la matriz, además de sus atributos se obtuvo la posición correspondiente del dato dentro de la matriz, una vez recorridas todas las ramas, subramas y accedidas las hojas estas mismas se almacenaron dentro de una lista circular con el objetivo de tener acceso inmediato para su utilización y procesamiento en el momento que sea requerido, estas estructuras anteriormente mencionadas son TDA's que fueron desarrollados con la utilización de punteros, con ello permitiendo que los datos sean almacenados de manera correcta con toda la información obtenida desde el archivo de entrada.

c. Generación de matrices binarias

Las matrices binarias son matrices donde solo existen los elementos 1 y 0 en cada uno de sus nodos, estas matrices fueron construidas en base los datos contenidos en los nodo de las matrices originales, esto se logró identificando cada uno de los nodos de la matriz, verificando entonces la posición (x, y), una vez identificado el dato contenido se compara si este mismo es un número diferente de cero, si esto es así, entonces en su correspondiente matriz binaria agregará en la misma posición del nodo "1", en caso esto

no se cumpliera y existiese un cero dentro del nodo, simplemente se escribiría un "0" nuevamente. Por último, estas matrices son almacenadas al igual que las matrices originales ingresadas dentro de una lista circular diferente de la anteriormente creada para las matrices originales.

d. Generación de matrices reducidas

La matrices reducidas, como su nombre lo indica se trata de reducir la matriz original, esto es logrado a través de comparaciones dentro de las filas almacenadas en las matrices binarias, las cuales son comparadas una a una cada fila, esto con el objetivo de encontrar filas que sean iguales, donde una vez identificadas de cada una de las filas son obtenidas sus ubicaciones para luego compararlas con las ubicaciones de las matrices originales y así obtener las filas que serán sumadas para luego ser almacenadas dentro de nuevas filas que posteriormente formaran matrices para almacenarse dentro de una lista circular con el objetivo de un acceso rápido estas misma en cualquier momento de la ejecución del programa. Todo este proceso para lo obtención de una matriz reducida además fue apoyado por la utilización de un TDA tipo cola, el cual fue utilizado para poder almacenar cada una de las filas con el objetivo de siempre tener los datos almacenados dentro de una estructura que permita realizarlo en un orden específico, en este caso el primero en entrar, primero en salir, además de conservar los datos que fuesen desencolados dentro de colas auxiliares para que no exista una pérdida de

información durante el procesamiento de los mismo.

e. Generación de gráficas

La generación de las gráficas para los reportes es en base a los datos contenidos en los archivos de entrada y los datos que serán escrito dentro del archivo de salida, estos gráficos fueron realizados con las librerías Graphviz, dichos gráficos muestran cual es la estructura contenida de la matriz que sea seleccionada y permiten con ello la visualización de los cambios que fueron realizados para poder obtener las matrices reducidas que son mostradas en conjunto con la matriz original para una mejor visualización de la optimización de datos obtenida después del procesamiento de la información.

f. Escritura de archivos XML

La escritura de los resultados obtenidos dentro de archivos del formato XML se realiza con el objetivo de obtener un acceso para su utilización posterior dentro de distintas redes distribuidas, donde los archivos ya no contarán con filas de datos duplicados binariamente, como se propuso en la solución obteniendo una reducción de los recursos para su utilización dentro del ámbito de las redes distribuidas. Este proceso se realiza de manera inversa a la escritura, ya que se asigna cada nodo de la matriz a una hoja con sus datos de posición correspondientes, posterior a ello se crearon las ramas donde se colocan las hojas

correspondientes, además de asignarle sus nuevas dimensionales y atributos correspondientes, por último, todo lo anterior se inserta dentro de la etiqueta árbol para que posterior fuera escrito dentro de un archivo con el formato correspondiente, para este caso formato XML.

Todos los procesos realizados anteriormente fueron agregados a una clase principal desde la cual se puede acceder a todas están seleccionándolas a través de una entrada de texto, por medio de una interfaz de consola.

Conclusiones

Los datos que son procesados dentro de grandes redes de información pueden requerir de la utilización de grandes cantidades de recursos, por ello es por lo que este tipo de soluciones a través del procesamiento de los datos puede ayudar a que estos se reduzcan de una manera significativa y por ello se reduzca también los gastos realizados. Siendo el último punto el objetivo buscado en la realización de este proyecto.

La utilización de apuntadores y TDA's dentro de este tipo de cuestionamientos nos puede ayudar a poder desarrollar una mejor lógica con la cual podamos desarrollar posteriormente mejores soluciones para otros problemas que puedan existir dentro de ámbitos empresariales.

El uso de TDA's permitió que el almacenamiento y procesamiento de los datos que serán ingresados en los archivos XML, ya que todos los datos fueron organizados y asignados de manera ordenada a sus

espacios correspondientes según los atributos en el archivo de entrada, todo lo anterior, facilito la obtención de las matrices binarias, reducidas, además de las gráficas correspondientes a cada una de las matrices que fueron almacenadas.

Anexos

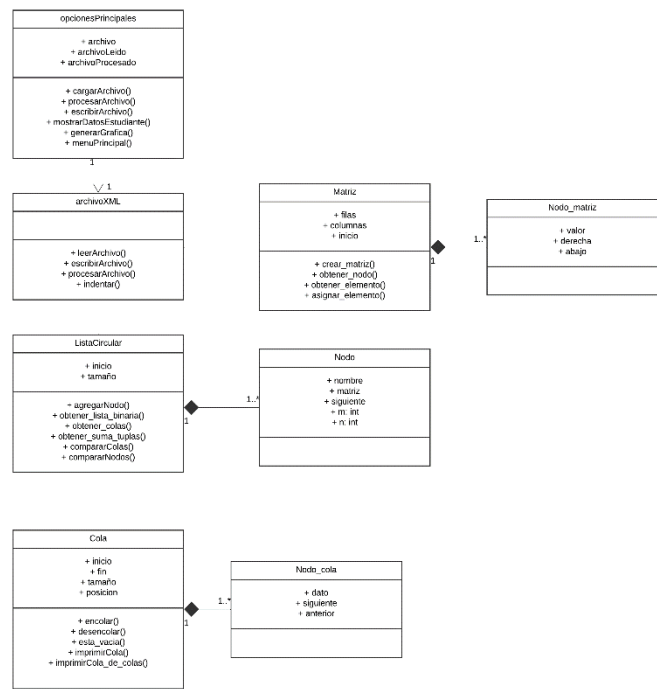


Figura 1. Diagrama de clases.

Fuente: elaboración propia.

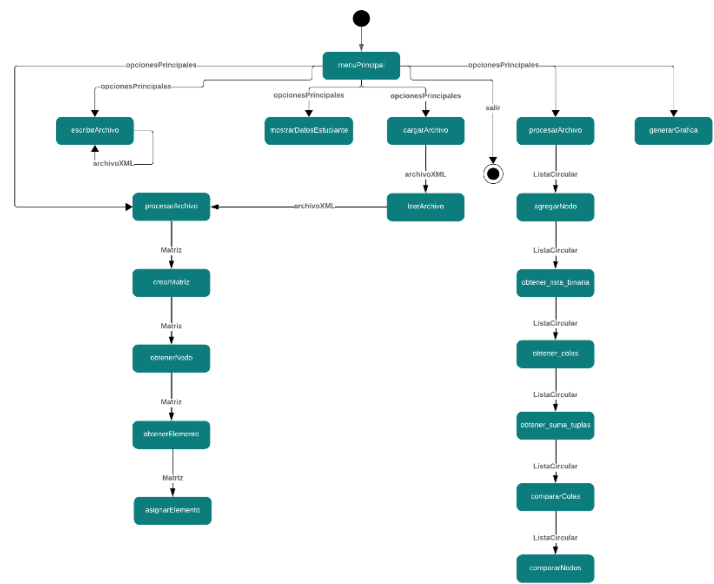


Figura 2. Diagrama actividades.

Fuente: elaboración propia.

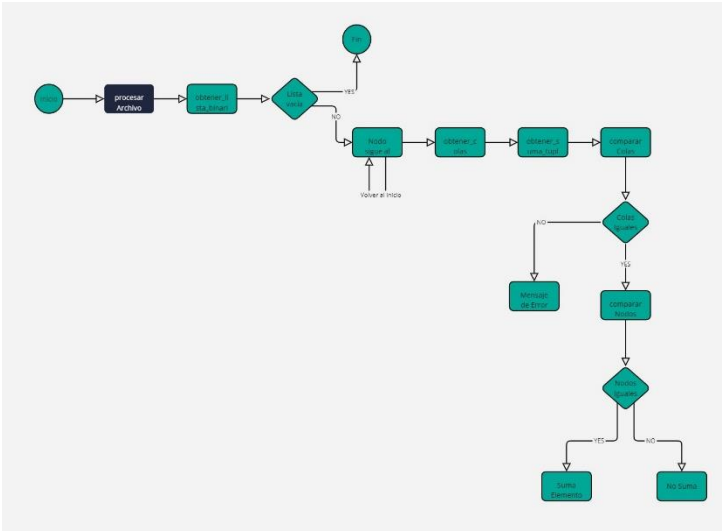


Figura 3. Diagrama actividades procesamiento de datos.

Fuente: elaboración propia.

Referencias bibliográficas

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