Proyecto Programado I

Diseño e Implementación de una Base de Datos Distribuida

courierTEC

F. Astorga, E. Ulate, S. Víquez

***Resumen*—Las bases de datos son colecciones lógicas de datos para un propósito específico, por lo que son esenciales en la tecnología de información. Pero, además hoy día, se incluyen las tecnologías de la comunicación. Un ejemplo claro de esto son las bases de datos distribuidas, que representan bien la unión que hay entre las tecnologías de redes de computadoras y los sistemas de bases de datos. El presente paper contiene información acerca de las etapas de diseño, fragmentación y asignación de una base de datos distribuida, con su respectiva comunicación entre los sitios y la aplicación general para el respectivo uso de la empresa courierTEC y las consultas que se quieren llevar a cabo en el sistema.**

**Palabras clave:** asignación, base de datos distribuida, fragmentación, SQL Server.

**I. INTRODUCCIÓN**

La evolución de las bases de datos se remonta desde un tiempo muy antiguo, aunque claro está en ese momento no se referían con ese término. Desde siempre el ser humano ha tenido la necesidad de almacenar datos relevantes, con contenidos históricos, legales, de registros de población, entre otros. Estos datos evidentemente tienden a crecer, por lo que fue necesaria la creación de sistemas o modelos más prácticos y eficientes en el manejo de la información. [1]

Con la aparición de las computadoras y después de varios intentos por crear modelos más eficientes e intuitivos surgieron lo que hoy día conocemos como los sistemas de bases de datos, compuestos por la colección de datos, o sea la base de datos en sí (BDD), y el sistema administrador de bases de datos (SABD). [2]

En la década de los años de 1980, es cuando empieza a surgir más formalmente este término y con ello se empiezan a crear prototipos de lo que hoy conocemos como bases de datos distribuidas. En ese momento se buscaba cómo tener las características de un SABD centralizado, pero al mismo tiempo poder almacenar y acceder a los datos desde lugares diferentes, así que esto representa la unión de tecnologías de redes de computadoras y de sistemas de bases de datos. [3]

Una base de datos distribuida es una colección de múltiples bases de datos interrelacionadas distribuidas sobre una red de computadoras. [4] Es una forma de mezclar el procesamiento masivo de datos y la eficiencia en la comunicación entre sitios necesaria para que el sistema funcione como uno solo.

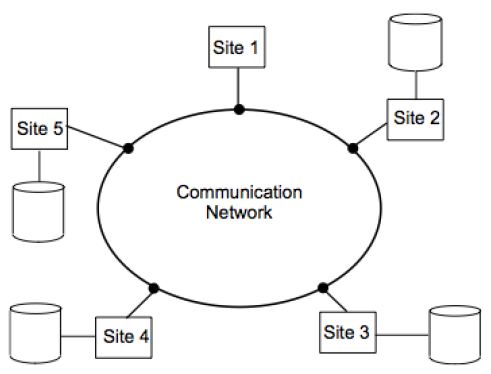


Figura 1. Estructura de un Sistema   
de Bases de Datos Distribuido [4]

**II. DESCRIPCIÓN DEL PROBLEMA**

La empresa courierTEC es una distribuidora de paquetes que ha decidido implementar un sistema distribuido que le permita registrar información de los servicios que utilizan los clientes en cada sucursal, así como monitorear y generar diversos indicadores en tiempo real. Los paquetes que despacha la empresa se clasifican en dos grandes categorías: regular, que incluye electrónicos, ropa, juguetes, elementos para el hogar, entre otros; y la especial que incluye comida, baterías, químicos, herramientas.

Esta empresa cuenta con 3 sucursales, la principal y en la que se encuentran las oficinas administrativas centrales en Heredia, otra en Cartago y la última en San José. Cada una de estas sucursales tiene un administrador que reporta al gerente en la oficina central. Se requiere una alta disponibilidad del sistema a desarrollar, de forma tal que si uno de los sitios (sucursal) sale de la red, aun así se pueden consultar sus datos desde las oficinas centrales.

Los administradores de San José y Cartago y el gerente en Heredia suelen utilizar el sistema para consultas como cantidad de dinero recaudado (paquetes retirados), listado de paquetes de los clientes, y otros valores asociados a montos promedios, tipos de paquetes y consultas que involucran un rango de fechas.

La aplicación debe tener un módulo para que los empleados de las sucursales puedan registrar la entrada y salida de paquetes, un módulo para los clientes para consultar sus paquetes y otro para las labores administrativas y gerenciales. Todo esto con validación de usuarios para permitir que solo usuarios registrados utilicen el sistema y accedan a las funcionalidades según el rol. [5]

**III. INVESTIGACIÓN**

Hay varios métodos para la replicación de bases de datos en SQL Server, entre las que se encuentran *snapshot replication*, *transactional replication* y *merge replication,* este último fue el método escogido para la implementación del sistema distribuido de entrega de paquetes courierTEC.

Merge replication es un método para copiar y distribuir datos y objetos de una BDD desde un servidor de bases de datos SQL hacia otros sitios, manteniendo sincronizados los sitios de manera consistente. Una característica clave de este tipo es que permite que tanto el servidor primario (publicador) como los suscriptores puedan modificar de manera independiente la base de datos. [6]

Esto usa un *Merge Agent* para controlar los datos, ya que es el responsable de sincronizar los cambios entre el publicador y el suscriptor, debido a que ambos pueden cambiar los datos, este agente se encarga de manejar los conflictos en caso de que hubiera. Otra característica importante y funcional para el sistema a implementar es que le permite al suscriptor una vez desconectado del publicador volver a sincronizarse cuando se reestablezca la conexión. [7]

**IV. EJEMPLO**

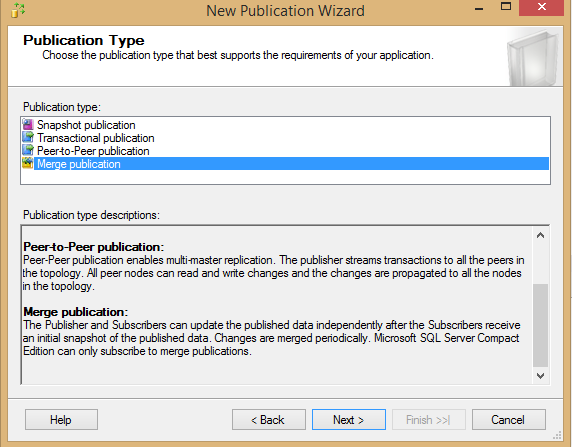


Figura 2. Posibles opciones para hacer una replicación de base de datos con el asistente de nueva publicación.

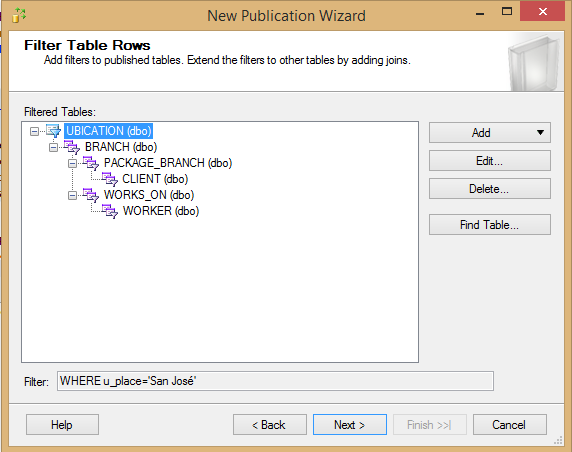


Figura 3. Fragmentación para la sucursal de San José en la Publicación

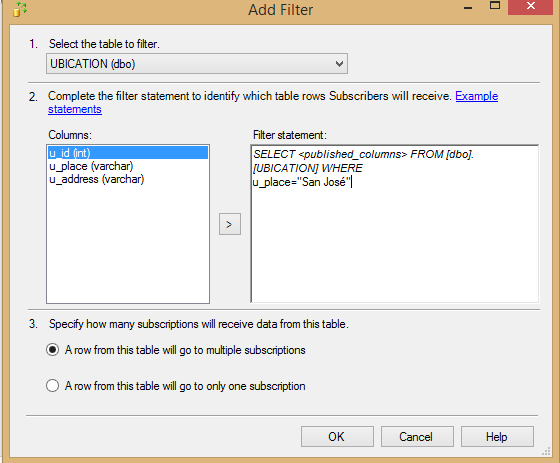


Figura 4. Fragmentación horizontal primaria   
para la sucursal de San José

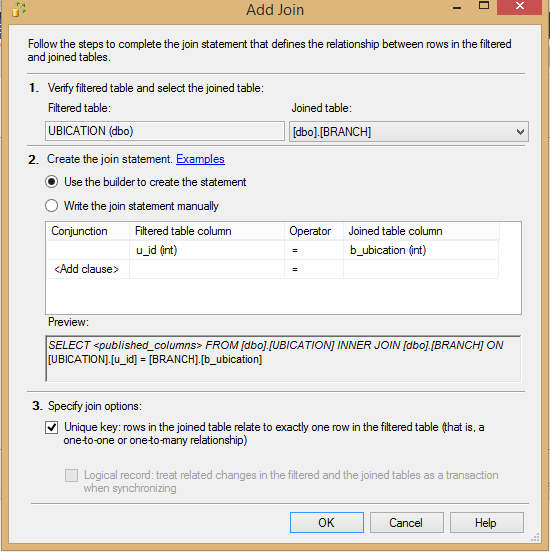


Figura 5. Fragmentación horizontal derivada para la sucursal de San José

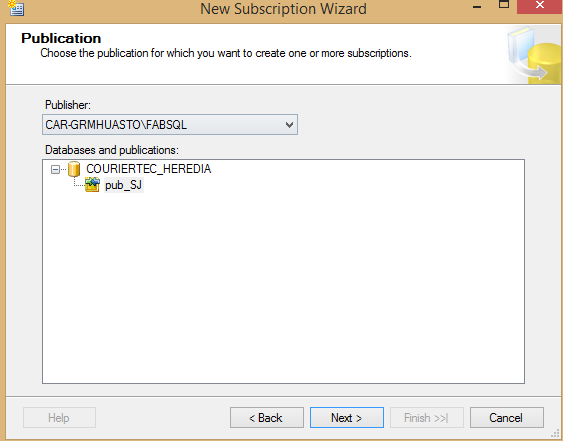


Figura 6. Se añade como Suscriptor a la sucursal de San José a courierTEC Heredia (publicador).

Para el caso de este sistema, Heredia como es la oficina principal se designa como el sitio publicador, mientras que como se muestra en la Fig. 6 el sitio San José se añade como un suscriptor, esto después de realizar las respectivas fragmentaciones horizontales, tanto primaria como derivada de los campos que se necesitan en dónde la ubicación sea “San José” (Fig. 3, 4 y 5).

Del mismo modo como se ejemplificó en dicho paper con las figuras correspondientes al nodo de San José, se necesita hacer con los campos relacionados para el sitio de Cartago, de manera que sea un suscriptor más de la replicación y así se puedan mantener en constante comunicación los tres sitios.

.

**V. CONCLUSIÓN**

Los sistemas de bases de datos distribuidos buscan darle un enfoque de utilidad al manejo de gran cantidad de información en diferentes sitios o nodos, esto sobre todo en organizaciones donde hay una división lógica y física y se usan altos volúmenes de información.

Todo esto conlleva a que se deba tener un buen manejo, fundamentado y eficiente de los datos que van a estar en el sistema y a través de la red. Mediante un buen diseño de la base de datos, con transparencia de red, fragmentación y replicación, en la que se debe tomar en cuenta factores de uso de tipo estadístico, de localización y de organización.

Y es que así los datos están donde se utilizan con mayor frecuencia, y a través de operaciones de modificación en la base de datos (escrituras) o consultas (lecturas) es posible hacer que el sistema sea eficiente en el procesamiento de dichas acciones. Claro está, sin dejar de lado la disponibilidad de los datos presentes en otros nodos y la comunicación que debe haber entre los sitios pertenecientes a la red.

**REFERENCIAS**

**1** M. Artavia Jiménez, *Tema 1: Repaso Historia Bases de Datos* Cartago, 2018, pp. 3-7.

**2** M. Artavia Jiménez, *Tema 1: Repaso SABD - Concurrencia* Cartago, 2018, pp. 4-7.

**3** M. Artavia Jiménez, *Tema 2: Introducción a Sistemas Distribuidos de Bases de Datos.* Cartago, 2018, pp. 7-76.

**4** M. Ozsu and P. Valduriez, *Principles of Distributed Database Systems, Third Edition*. New York, NY: Springer New York, 2011.

**5** M. Artavia Jiménez, *Proyecto Programado 1: Diseño e implementación de una base de datos distribuida* Cartago, 2018, pp. 1-3.

**6** “Merge Replication | Idera Glossary", *Idera*, 2018. [Online]. Available: https://www.idera.com/glossary/merge-replication. [Accessed: 22- Sep- 2018].

**7** D. Din, "SQL Server Replication - Merge Replication type step by step", *Experts Exchange*, 2018. [Online]. Available: https://www.expertsexchange.com/articles/17532/SQL-Server-Replication-Merge-Replication-type-step-by-step. [Accessed: 18- Sep- 2018].