

SISTEMA DE TRANSMISION DE INFORMACION ENTRE BASES DE DATOS DISTRIBUIDAS

José Luis Torre Hernández

Departamento de Lenguajes y Sistemas Informáticos
Universidad del País Vasco

1. Sistema de transmisión de información entre bases de datos distribuidas

En este artículo describiremos cómo se ha realizado un Sistema para controlar la Información generada en una Base de Datos distribuida en lugares alejados geográficamente. La diferencia frente a otros sistemas más sofisticados radica en la economía, y flexibilidad de la solución propuesta. El Sistema de Comunicaciones se instala sobre un soporte físico de extrema sencillez; con ello se consigue una instalación muy económica.

Con el fin de simplificar la construcción del Sistema se han utilizado herramientas del entorno de desarrollo UNIX. El trabajo puede considerarse como una capa que engloba al Sistema UUCP, éste fue concebido y desarrollado por Mike Lesk en los Laboratorios Bell. Concretamente utilizaremos la versión denominada HoneyDanBer, desarrollada por Honeyman, Nowitz y Redman, que se ha utilizado hasta la fecha y ha sido la base la utilizada en la versión 4 del UNIX.

La estructura geométrica de la red es una estrella, en la que el nodo principal es el centro y las sucursales son ramificaciones de la misma. Esta estructura se presenta en muchísimas de las empresas de nuestro entorno, tanto si funcionan como centrales o como delegaciones de otros nodos más importantes.

En cada uno de los nodos del Sistema de Comunicaciones, se utiliza el Sistema de Archivos para almacenar las informaciones que van a circular a través de la red. Fundamentalmente existirán dos directorios, el directorio de envío, en el

que se almacenarán los datos que deben enviarse a las delegaciones y el directorio de recogida en el que se recibirán las informaciones generadas en los otros nodos de la red. En cada uno de estos directorios, deben existir tantos directorios como el resto de nodos existentes en el Sistema.

Para conseguir un sistema más fácil de utilizar, los nombres de las sucursales se utilizarán para representar los nombres de cada una de sus correspondientes máquinas. Es decir, en cada uno de los directorios de envío y de recogida deben existir los nombres de todas la máquinas a las que debe enviarse y recibir información.

Con esta estructura, la construcción de los nombres de los archivos a transmitir es relativamente sencilla. En primer lugar se utiliza el nombre de la máquina como lugar de origen de la información, a continuación el nombre del directorio que indica el nombre de la máquina destino, y por último el nombre del archivo a transmitir.

Todo el Sistema se sustenta sobre líneas de transmisión asíncrona [1], conectando ordenadores directamente o a través de modems, para utilizar la red telefónica conmutada.

Con el fin de evitar errores, todas las transmisiones deben estar automatizadas, por ello hemos evitado los modems de respuesta manual. Con ésto no queremos decir que no puedan utilizarse los modems manuales, sino que en este caso el sistema no podría estar totalmente automatizado y el usuario tendría que intervenir en cada conexión, cosa que consideramos inaceptable.

En relación al modem indicaremos que debe utilizarse uno que funcione bien en las líneas de nuestra red telefónica conmutada, y para ello debe reunir las siguientes características:

- 1) Eficiente detección de señal portadora.
- 2) Desconexión automática de la línea cuando se desconectan los modems.
- 3) Detección correcta de línea ocupada y tono de marcar.
- 4) Compatible con las normas Europeas de la C.C.I.T.T.
- 5) Programable mediante órdenes estándares (Hayes).

2. Fases de la comunicación

El envío constará de varias fases. En primer lugar la que denominaremos Descarga de la Base de Datos, después la Copia de Seguridad y por último la propia Comunicación.

La fase de Comunicación debería realizarse por la noche, cuando todos los puntos de la Red hayan terminado las dos fases anteriores.

Por la mañana, una vez que se han recibido los datos correctamente, debería realizarse la fase de Carga.

Por último, se debería comenzar a trabajar en los procesos diarios.

Fase de Descarga

El proceso de descarga, se encargará de obtener los datos de la Base de Datos y generar en cada uno de los directorios correspondientes a los nodos destino, un conjunto de ficheros ASCII a partir de los ficheros maestros de la Base de Datos. También se descargarán aquellos pedidos, albaranes y facturas pendientes de transmitir en el momento de realizar la descarga.

Durante esta fase, los usuarios podrán seguir trabajando en la aplicación, ya que la fase de descarga debe leer el archivo de transacciones registro a registro, seleccionando exclusivamente aquellos que sean anteriores al instante de la descarga, grabándolos en sus correspondientes directorios y eliminándolos, o marcándolos, del archivo de transacciones. De este modo, si durante la fase de descarga se añaden registros no serán descargados hasta la próxima descarga. Resumiendo, se pueden seguir introduciendo pedidos, albaranes y facturas durante la descarga.

Durante esta fase, se creará el indicador DESCARGA.I, para identificar que el proceso se está ejecutando. Al final de esta fase se elimina automáticamente. Por tanto, si no existe DESCARGA.I, se podrá comenzar la siguiente fase. Además no se permite lanzar una nueva descarga, a no ser que expresamente se indique lo contrario (siempre hay imprevistos de última hora), en cuyo caso la nueva información se añadirá a los ficheros ASCII.

Al finalizar correctamente se establece el indicador ENVIAR.F. Su finalidad es muy amplia, ya que se mantiene mientras no se ha realizado la transmisión a todas las sucursales, y en ellas, una vez recibida la información, y comprobado que se han recibido los ficheros transmitidos, y que su tamaño es el adecuado, puede realizarse la carga en cada una de las sucursales. Por supuesto, también deberá realizarse la fase de carga en la central, y cuando todo haya terminado podrá borrarse el ENVIAR.F.

Cuándo exista ENVIAR.F no debe ejecutarse la fase de descarga.

Fase de Copia y Compresión

Antes de la comunicación propiamente dicha, se realiza automáticamente una copia de seguridad de todos los ficheros que deben transferirse, con el fin de poder recuperarlos en caso necesario. Si la comunicación ha sido correcta se borrarán (en origen) los ficheros transmitidos, pero se mantendrán las copias realizadas. Esto es necesario porque los ficheros tienen el mismo nombre en todas las comunicaciones (esto hace más fácil su identificación).

Al realizar la fase de copia, se preceden las tres letras que identifican el nombre del fichero por el mes, día, hora y minuto del momento de la copia, quedando nombres de fichero similares a los siguientes:

07020845.alb 07020845.fac 07020845.ped 07030840.alb 07030840.fac 07030840.ped

Cuando los Sistemas Operativos de las máquinas origen y destino lo permitan, comprimiaremos los ficheros antes de su envío. Hemos previsto indicar esta circunstancia creando un fichero denominado COMPRESS, de tamaño cero en el directorio de envío, de este modo, el compresor primero busca este fichero, y en caso de encontrarlo sabe que debe realizar la compresión. Es tarea del Administrador del Sistema ocuparse de crear dicho fichero, ya que de este modo conseguirá un ahorro sustancial (aproximadamente el 50 %) en el costo de las comunicaciones.

Una vez que la copia y la posible compresión han finalizado, se realiza la generación del fichero NOMBRES, que contiene el verdadero nombre de cada uno de los ficheros a transmitir (si se encuentran comprimidos acabarán en .Z o en .ZIP).

A partir de este momento, podría comenzarse la fase de comunicación, pero es conveniente esperar a que en el resto de los nodos se realice el mismo proceso. Por lo tanto, todo queda a la espera de que se lance la comunicación en el momento adecuado; normalmente por la noche.

Al final se crea el indicador TRANSFER.I, para significar que todo está preparado para las comunicaciones.

Fase de Comunicación

Su funcionamiento en esencia, es el siguiente:

En primer lugar, se solicita de la máquina remota el fichero NOMBRES, que se encuentra en el directorio de envío y contiene los nombres de los ficheros que la máquina remota debe enviar a la máquina local.

Posteriormente, se solicita la transmisión de un fichero auxiliar, denominado DESCANSO, para permitir que el proceso pueda determinar cuando se recibe el fichero NOMBRES, sin cortar la comunicación.

Cuando se comprueba que se ha recibido el fichero NOMBRES, se comienzan a solicitar, uno a uno, los ficheros contenidos en NOMBRES.

A continuación, se realizan las peticiones de transferencias de ficheros de la máquina local a la remota (esto es más sencillo porque en la máquina se sabe que tiene que transmitirse).

En esta fase, se deben generar automáticamente los nombres completos de los ficheros. La construcción del nombre de los ficheros de la máquina destino es algo complicado, ya que la máquina que solicita la petición de transferencia debe conocer el directorio de la máquina remota.

La solución más sencilla consiste en establecer un directorio común para todas las máquinas. Cuando esto no sea posible deben utilizarse nombres de directorios diferentes para la central y para las sucursales. Es decir utilizaremos las variables

`D_RED_C=\usr\spool\uucppublic`, y

`D_RED_S=\var\spool\uucppublic`.

siendo `D_RED_C` para la central y `D_RED_S` para las sucursales.

En algunos programas hará falta utilizar ambas variables, pero cuando no sea necesario se utilizará la variable `D_RED`.

Evidentemente en la sucursal deberá cumplirse que `D_RED = D_RED_S`, y en la central `D_RED = D_RED_C`.

Tras las peticiones de los ficheros éstos se irán recibiendo en las máquinas remotas, normalmente por la noche. Una vez que se han enviado y recibido todos los ficheros, podemos considerar que se ha finalizado la transmisión.

Por la mañana deberían haberse recibido los ficheros en todas las máquinas.

Fase del Test de Recepción

Esta es una fase importante, aunque sólo sea porque nos ha servido para detectar errores de funcionamiento en las primeras instalaciones.

Se basa en un proceso, recorre, que recorre todos los directorios del directorio de RECOGIDA, comprobando si se han recibido todos los ficheros y si su tamaño es el adecuado.

Hemos de saber que siempre se recibe el fichero NR_CAR, que contiene el tamaño de los ficheros que se han transferido. Si se comprueba que el tamaño de los ficheros, de ese directorio, no coincide con el contenido del fichero NR_CAR, es que todavía no se ha terminado la transmisión o que hemos tenido algún problema.

El proceso recorre, indicará al nodo origen que la recepción ha sido correcta solicitando el envío del indicador ENVIADO.I, al remitente correspondiente.

Este proceso debe ejecutarse tanto en el nodo central como en los periféricos, recorriendo todos los directorios en los que exista la posibilidad de recibir información. Además, cuando la recepción ha sido correcta (en un determinado directorio), se establece el indicador RECIBIDO.I, lo cual evita inspeccionar de nuevo en ese directorio.

Fase de Transmisión de Pares

Hemos previsto el proceso conectar, con el fin de poner en contacto las máquinas origen y destino, para que intercambien los indicadores ENVIADO.I, generados en la fase anterior en el supuesto de que las transmisiones hayan sido correctas.

Fase de Borrado

Cuando el proceso revisar, comprueba que se han enviado todos los ficheros, éstos se borran automáticamente. Esto es así cuando en los directorios de envío se comprueba que existe el fichero ENVIADO.I.

Fase de Copia

Antes de la fase de carga se realiza una copia de seguridad de los ficheros recibidos y que van a utilizarse durante la fase de carga.

Tal como se hace con las copias de seguridad antes del envío, MMDDHHMM.XXX, representará la fecha de la copia y el nombre del fichero.

Fase de Carga de la Base de Datos

La fase de carga de la Base de Datos se realiza tanto en la sucursal como en la central mediante el programa denominado carga. Su funcionamiento se basa en seleccionar la sucursal (o la central) y realizar los procesos de carga de la Base de Datos.

Durante la fase de carga existe el indicador CARGA.I, para que ningún usuario intente la realización de una carga manual. Al finalizar el proceso carga, CARGA.I se elimina automáticamente.

Hemos escrito un test para comprobar que la recepción de los archivos ha sido correcta, tanto en las sucursales como en la central. Se denomina RECIBIDO.T, y se encarga de crear el indicador RECIBIDO.I, cuando comprueba que la recepción ha sido correcta. Antes de efectuar la carga de los archivos recibidos, se pone en funcionamiento este test, y en el caso de que el resultado sea correcto se realizará automáticamente la carga de los ficheros de cada una de las sucursales escogidas.

Se seleccionan automáticamente todas las sucursales que ejecuten correctamente el test RECIBIDO.T, y se presentan en pantalla; entonces se hace una carga selectiva por parte del usuario. Sin embargo, cuando se desea que la carga sea totalmente automática, debe utilizarse una variable de entorno y asociarle el nivel máximo de automatización, es decir NIVEL = 99. Como puede imaginarse, hemos previsto esta variable para permitir diversos niveles de automatización en función de las necesidades del usuario.

En las sucursales la carga se realizará cuando existan movimientos procedentes de la central.

Cuando las cargas producen errores, éstos se listan para su posterior análisis, generando al mismo tiempo el fichero correspondiente que podrá ser revisado en caso necesario. Los ficheros de errores se borran al cabo de un mes.

Borrado de Ficheros Cargados

Si todo ha funcionado correctamente y no existen ficheros de errores, se hace una llamada automática al programa fin que se encarga de borrar los ficheros cargados, y el indicador ENVIAR.I, de manera que pueda realizarse un nuevo descarga y una nueva transferencia.

Fase de Trabajo Diario

En esta fase, los usuarios trabajarán actualizando la Base de Datos con los movimientos diarios debidos al funcionamiento normal de la empresa hasta la hora del descarga, en cuyo momento se comenzará el Ciclo completo.

3. Conclusiones

Hemos explicado el Ciclo completo de Transmisión de la Información generada en una Base de Datos Distribuida, analizando las principales fases implicadas en el transcurso del tiempo desde que se generan los datos en la aplicación hasta que se almacenan en sus correspondientes destinos, pasando por los procesos de transmisión mediante la línea telefónica o línea directa. Hemos descrito los principales indicadores que deben tenerse en cuenta para que no exista posibilidad de que algunos procesos interfieran en el funcionamiento de otros, con los que no pueden ejecutarse simultáneamente. También hemos indicado alguno de los test utilizados para comprobar que la transmisión ha sido correcta. Hemos realizado escritura de los programas que implantan esta solución en el lenguaje shell desarrollado por Bourne, y funcionan en ksh prácticamente sin ninguna modificación, consiguiendo con ello gran portabilidad y facilidad de mantenimiento. En la actualidad se está desarrollando una versión en C++ con el fin de obtener mejores rendimientos, a la vez que se mantiene la portabilidad.

Bibliografía:

[1] RS-232 ISO-2110 American National Standards Institute 1430 Broadway NY 10018, USA.

[2] La Communication sous UNIX. Jean-Marie Rifflet. McGraw-Hill, Paris.

[3] X.25 Made Easy. Nicolas M. Thorpe & Derek Ross. Prentice Hall 1992.

[4] Computers Networks. Protocols, Standards and Interfaces. Uyles Black. Prentice Hall 1993.

[5] FDDI. Technology and Applications. Sonu Mirchandani and Raman Khanna. Wiley 1992.