

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 395 205**

51 Int. Cl.:

G06F 17/30 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.10.2001 E 10179745 (4)**

97 Fecha y número de publicación de la solicitud europea: **22.12.2010 EP 2264617**

54 Título: **Procedimiento y aparato para replicar una base de datos**

30 Prioridad:

09.10.2000 NZ 50738600

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

11.02.2013

73 Titular/es:

**MAXIMUM AVAILABILITY LIMITED (100.0%)
46 Mulgan Way, Browns Bay
Auckland, NZ**

72 Inventor/es:

TARBELL, JAMES SCOTT

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 395 205 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento y aparato para replicar una base de datos

Campo de la invención

5 La presente invención se refiere a un procedimiento y un aparato para el procesamiento de datos. Más particularmente, pero no exclusivamente, la invención se refiere a un procedimiento y un aparato para replicar una base de datos

Antecedentes de la invención

En un número de aplicaciones de procesamiento de datos se deben procesar fragmentos de datos enviados desde un sistema de origen dentro de un formato de datos requerido sobre un sistema objetivo.

10 En muchos casos se desea replicar una base de datos sobre un sistema de ordenador objetivo a partir de una base de datos sobre un sistema de origen. Este procedimiento puede involucrar el envío de entradas del diario desde la base de datos de origen para permitir la actualización de la base de datos objetivo. Las bases de datos pueden consistir de una o más librerías, cada una de las cuales contiene uno o más ficheros, teniendo cada uno de los ficheros uno o más elementos. Cada uno de los elementos consiste de una tabla que tiene una o más filas. Una
15 entrada del diario puede contener un identificador de la librería; archivo; elemento del archivo y una fila de los datos cambiados para el elemento del fichero. Esta entrada del diario se puede usar por el sistema de ordenador objetivo para actualizar su base de datos.

Es importante que las entradas de la base de datos desde una tabla determinada se actualicen en la secuencia correcta y que se actualicen los elementos interrelacionados en la secuencia correcta. Para asegurar que las
20 entradas del diario se procesan adecuadamente el procedimiento de recepción del sistema de ordenador objetivo puede comparar un nombre de objeto (librería / fichero / elemento) con una base de datos de objetos almacenados sobre el sistema de ordenador objetivo. Cuando se localiza un objeto coincidente se puede usar el procesamiento de la información asociada con ese objeto para procesar la entrada del diario.

25 El enfoque tradicional ha sido transferir entradas del diario, almacenarlas y replicar la base de datos utilizando un dispositivo único. Este enfoque es lento y complejo.

El documento de la técnica anterior EP 0 420 425 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) del 3 de abril de 1991 desvela un sistema de procesamiento de transacciones que incluye una base de datos activa primaria y una base de datos de réplica de seguimiento. El procesamiento de cambios de la base de datos de réplica se realiza separando los registros de reproceso obtenidos del registro de transacciones de la base de datos primaria en una pluralidad de colas respectivas. Los registros de reproceso se separan de tal modo que todos los registros de transacciones para una unidad de transferencia (página) de la base de datos primaria se colocan sobre la misma cola en la secuencia de registros. Cada una de las colas se enlaza exclusivamente con uno de la pluralidad de servidores de colas en paralelo. Cada uno de los servidores de colas aplica a la base de datos de réplica los registros de reproceso en las colas que sirve exclusivamente. La base de datos de réplica se realiza por lo tanto de
30 forma consistente con los datos primarios por un mecanismo de actualización libre de bloqueos que da servicio a las páginas de la base de datos de réplica en paralelo.

Sería deseable para el sistema de réplicas de bases de datos cumplir con los siguientes requisitos:

1. Asegurar que las entradas del diario se ponen en serie por el elemento de la base de datos (en un mínimo), y por cualesquiera agrupamientos de usuarios especificados.
- 40 2. Soportar un número extremadamente grande de procesos de aplicación de la base de datos de modo que las I/O (entrada / salida) de la base de datos se puedan gestionar fácilmente.
3. Procesar las entradas del diario de modo que se minimice la cantidad de I/O del sistema (por ejemplo el paginado) entre el instante en el que se obtienen las entradas desde el diario y el instante en que se aplican a la base de datos de réplica.
- 45 4. Las funciones soportan cualquier tipo de paquetes de datos, no solo las entradas del diario, para permitir futuras extensiones a otros tipos de réplicas (por ejemplo, objetos, ficheros de flujo, etc.).
5. El sistema oculta la complejidad de las funciones de gestión de la memoria desde otros componentes.

Es un objeto de la presente invención proporcionar un procedimiento y un aparato para la réplica de información que cumple estos requisitos o al menos proporciona al público con una elección útil.

50

Revelación de la invención

De acuerdo con un primer aspecto de la invención se proporciona un procedimiento para replicar la información a partir de un sistema de origen en un sistema objetivo como se define en la reivindicación independiente 1.

5 De acuerdo con un aspecto adicional de la invención se proporciona un procedimiento como se define en las reivindicaciones dependientes.

Preferiblemente, el sistema de ordenador objetivo es un sistema de ordenador multi-procesador.

Breve descripción de los dibujos

La invención se describirá ahora a modo de ejemplo con referencia a los dibujos adjuntos en los que:

10 Figura 1: muestra un diagrama esquemático de un sistema de ordenador de origen que proporciona entradas del diario a un sistema de ordenador objetivo.

Figura 2: es un diagrama funcional que ilustra los procedimientos involucrados en la base de datos de réplica en un sistema de ordenador objetivo.

Figura 3: muestra el mapeo del espacio de almacenamiento dentro del sistema de ordenador objetivo.

15 Figura 4: muestra un diagrama de flujo que ilustra el procedimiento para la asignación de las entradas del diario a grupos de serialización.

Descripción detallada de la realización preferida

La siguiente descripción describe un procedimiento de réplica de una base de datos donde los sistemas de ordenador de origen y objetivo son ordenadores IBM AS/400 que operan bajo el sistema operativo OS/400. Se apreciará que el procedimiento es aplicable a otros sistemas con la modificación apropiada.

20 Refiriéndonos a la figura 1, el sistema de origen A contiene una base de datos primaria 1. La base de datos primaria 1 puede contener uno o más librerías. Cada una de las librerías puede contener uno o más ficheros. Cada uno de los ficheros puede contener uno o más elementos. Cada uno de los elementos comprende una tabla que tiene una o más filas. Una combinación única de librería/fichero/elemento se denomina como un objeto.

25 Cuando se modifica una fila de cualquier elemento de la base de datos primaria 1 se envía una entrada del diario que incluye el nombre del objeto y la fila modificada al receptor del diario local 2. El receptor del diario local 2 envía la entrada del diario a través del enlace de comunicaciones 3 a un receptor del diario remoto 4 de un sistema de ordenador objetivo B. El procedimiento de réplica de la base de datos 5 recibe las entradas del diario y modifica los contenidos de la base de datos de réplica 6 para mantenerla de acuerdo con la base de datos primaria 1.

30 Con referencia ahora a la figura 2 se describirán el procedimiento y el aparato para la réplica de la base de datos objetivo 6 del sistema de ordenador objetivo. Para asegurar la réplica adecuada de la base de datos de réplica 6, se actualizan los elementos de la base de datos en la base de datos de réplica 6, en el mismo orden que se modificaron en la base de datos primaria 1. Para conseguir esto se definen varios grupos de serialización 8. Las entradas del diario que tienen el mismo nombre del objeto se agrupan en un grupo de serialización común de modo que se actualicen en el orden correcto. Ciertos elementos de la base de datos pueden tener relaciones con otros elementos de la base de datos (uniones, etc.) y de este modo se pueden asignar a un grupo de serialización común para asegurar que todos los elementos interrelacionados se actualizan en la secuencia correcta. Un grupo de serialización puede contener de este modo entradas del diario para varios objetos. El uso de tales grupos de serialización posibilita dirigir la réplica de la base de datos en la secuencia apropiada así como facilitar el procesamiento eficiente en paralelo.

40 El procedimiento de recepción 7 puede bien asignar una entrada del diario recibida a un grupo de serialización, asignar una entrada del diario a un grupo de serialización por defecto o rechazar la entrada del diario. La asignación del grupo de serialización se realiza en base a una base de datos de asignación (MXSGMBAS) y un objeto temporal del índice de usuario OS/400. Las funciones de asignación de la entrada del diario se proporcionan a través de un programa del servicio ILE - que permite modificar la implementación subyacente sin recompilar/enlazar las funciones de llamada.

45 La base de datos de asignación MXSGMBAS contiene todos los objetos, su relación con otros objetos (es decir se necesita agrupar con otros objetos durante el procesamiento) y su modo requerido de procesamiento. La asignación de una entrada del diario a un grupo de serialización 8 podría dirigirse simplemente comparando el nombre del objeto de cada una de las entradas del diario recibidas con la base de datos de asignación MXSGMBAS y la asignación de la entrada del diario a un grupo de serialización en base a la información asociada. Sin embargo, la base de datos de asignación MXSGMBAS contiene muchos objetos y se requiere un tiempo considerable de procesamiento para realizar la operación de localizar la base de datos y extraer la información del procesamiento relevante. De acuerdo con la invención se usa un objeto temporal del índice de asignación de elementos (MBIX)

para almacenar la información de procesamiento para un objeto. Éste es un índice de los objetos que dan su grupo de serialización asociado y la información de procesamiento relacionada (incluyendo un enlace a sus estructuras de control asociadas).

5 Con referencia ahora a las figuras 2 y 4 se describirá la asignación del grupo de serialización. Cuando se recibe una entrada del diario en el procedimiento de recepción 7 de la etapa 11 se dirige una comparación en la etapa 12 para ver si el objeto está presente en el índice MBIX. Si es así, la operación procede a la etapa 13 y se devuelve un número del grupo de serialización y el índice del archivo de la base de datos (DBFIDX) y el procesamiento continúa dentro del grupo de serialización asignado.

10 Si el nombre del objeto no está almacenado en el índice MBIX entonces se dirige una búsqueda del nombre del objeto completo en la base de datos MXSGMBAS 9 en la etapa 14. Si la búsqueda es satisfactoria entonces se devuelve un grupo de serialización, se asigna un Índice del Fichero de la Base de datos (DBFIDX) que apuntará a la información del procesamiento almacenada en una red dinámica mantenida por el grupo de serialización asociado y se añade una entrada al índice MBIX en la etapa 15. Cada uno de los Índices de Fichero de la Base de Datos (DBFIDX) se crea simplemente aumentando un índice que es único para el grupo de serialización.

15 Si no se consigue una coincidencia en la etapa 14 a continuación se dirige una búsqueda de un nombre genérico en la etapa 16. Esto involucra una búsqueda por una librería/fichero/*todos y a continuación por una librería/*todos/*todos. Si se consigue una coincidencia genérica el nombre completo se añade a la tabla MBIX en la etapa 17 y el procesamiento continúa en las etapas 15 y 13 como antes. Si no puede conseguirse ninguna coincidencia la entrada del diario se descarta en la etapa 18.

20 Por consiguiente, en el arranque, no habrá ninguna entrada en el índice MBIX 10. A medida que se procesan las entradas del diario, los grupos de serialización y la información de procesamiento para los objetos se añadirán al índice MBIX 10. El grupo de serialización y la información de procesamiento se pueden obtener mucho más rápidamente desde la tabla de MBIX 10 que desde la base de datos MXSGMBAS 9.

Este procedimiento da los siguientes beneficios significativos de funcionamiento:

- 25 1. Los grupos de serialización no necesitan buscar una información del procesamiento relacionada con un elemento. Simplemente mantienen la información de procesamiento en una red dinámica con el Índice del Fichero de la Base de Datos como el medio de acceso.
- 30 2. Todas las operaciones relacionadas con un nombre de elemento particular pueden referirse al grupo de serialización y al valor del Índice del Fichero de la Base de Datos para identificar de forma única el elemento (un "asidero").

35 Con referencia ahora a la figura 3 se describirá el procedimiento de gestión de memoria dentro del sistema de ordenador objetivo. El espacio de objetos de almacenamiento se divide en varias unidades de almacenamiento SU1 - SU_N. Cada una de las unidades de almacenamiento tiene una cabecera de la unidad de almacenamiento 20. La cabecera de la unidad de almacenamiento 20 da el número de grupos de serialización que tienen entradas del diario en la unidad de almacenamiento. Cada uno de los segmentos de datos consiste de una cabecera de entrada de almacenamiento 21 y una entrada de almacenamiento 22. Las entradas de almacenamiento están alineadas sobre fronteras de 16 bytes con bloques de relleno 23 que rellenan cualquier espacio entre una entrada y una frontera de 16 bytes.

40 Las entradas del diario se pasan desde el procedimiento de recepción 7 para su almacenamiento en el espacio de objetos de almacenamiento 24. Las entradas del diario desde el procedimiento de recepción 7 se almacenan en el espacio de objetos de almacenamiento 24 en bloques 22. Cada entrada del diario 22 tiene una cabecera de entrada de almacenamiento asociada 21 (o asidero) que contiene información para el desplazamiento para la siguiente entrada del diario en la unidad de almacenamiento para ese grupo de serialización y un Índice de Archivo de la Base de Datos asociado (DBFIDX) que contiene la información de procesamiento para el elemento asociado con la

45 entrada del diario. La información de procesamiento se mantiene en una memoria dinámica con el Índice del Fichero de la Base de Datos como el medio de acceso.

50 Las entradas del diario de operación normal se escriben de forma consecutiva en una unidad de almacenamiento hasta que se rellenan y a continuación las entradas del diario se escriben en la siguiente unidad de almacenamiento disponible. Una vez que se ha completado la escritura en la unidad de almacenamiento se pueden leer las entradas del diario desde la unidad de almacenamiento poblada. Las unidades de almacenamiento rellenas parcialmente se pueden leer cuando los recursos del sistema no se estén utilizando de otra forma (es decir no se necesita almacenar ninguna entrada del diario).

55 Este enfoque significa que no se requiere bloquear la memoria durante la lectura y la escritura. Durante el procedimiento de escritura el procedimiento de recepción 7 tiene acceso exclusivo para escribir en la unidad de almacenamiento. No se requiere ningún bloqueo durante las operaciones de lectura y de este modo las entradas del diario se pueden leer simultáneamente de su grupo de serialización asociado. El único bloqueo requerido es disminuir el valor mantenido en la cabecera de la unidad de almacenamiento 20 cuando se lee la última entrada del diario para un grupo de serialización.

Las cola de unidades de almacenamiento disponibles (ASUQ) 25 controla el orden en el que se utilizan las unidades libres de almacenamiento. La ASUQ 25 incluye una memoria intermedia de último en entrar primero en salir (LIFO) que almacena las direcciones de las unidades libres de almacenamiento. Las entradas del diario de un grupo de serialización se leen de una unidad de almacenamiento hasta que se encuentra un valor nulo en una cabecera de entrada de almacenamiento. A medida que se lee la entrada de almacenamiento 22 se disminuye la cabecera de la unidad de almacenamiento 20. Cuando se han leído todas las entradas del diario completamente desde una unidad de almacenamiento la cabecera de la unidad de almacenamiento 20 se disminuirá hasta cero y se devolverá el número de la unidad de almacenamiento a la ASUQ y será la primera unidad de almacenamiento reasignada cuanto deban escribirse las nuevas entradas del diario en el espacio de almacenamiento. De este modo las unidades de almacenamiento usadas más recientemente se mantienen activas para reducir el conjunto en funcionamiento de las unidades de almacenamiento al mínimo.

Cuando se han leído todas las entradas del diario en una unidad de almacenamiento y se libera la unidad de almacenamiento se puede purgar todo el intervalo de direcciones de la unidad de almacenamiento sin que se requiera la escritura de datos al almacenamiento auxiliar.

Con referencia de nuevo a la Figura 2 se describirá adicionalmente el modo de procesamiento. El procedimiento de control 19 evalúa el procedimiento de réplica y controla el procesamiento en el procedimiento de recepción 7 y dentro de los grupos de serialización 8. De este modo el procesamiento se puede dirigir dentro de cada uno de los grupos de serialización sin tener en cuenta el procesamiento con otro grupo de serialización. Teniendo controlado todo el procedimiento por un procedimiento de control general 19 cada uno de los grupos de serialización puede dirigir su procesamiento aisladamente sin tener en cuenta la complejidad de la operación global.

A medida que cada uno de los grupos de serialización recibe entradas del diario para un elemento en secuencia, la actualización de ese elemento en la base de datos de réplica 6 también es secuencial. El procesamiento de elementos enlazados en un grupo de serialización particular se racionaliza.

Cuando se va a realizar una base de datos de réplica 6 se deben eliminar las transacciones parcialmente aplicadas en la base de datos primaria. En primero lugar, el procedimiento de control 19 suspende el procedimiento de recepción 7 y el procesamiento por los grupos de serialización 8. El procedimiento de control 19 identifica a continuación todos los grupos de transacciones "abiertas" (por ejemplo, ID de transacción que no han recibido aún una entrada de transacción o deshacer entrada del diario). Estas se procesan, de forma serie, desde la más reciente (es decir, el grupo de transacciones que tiene la entrada del diario más reciente) a la más antigua como sigue:

- i) un procedimiento de recepción del procedimiento de recepción 7 recibe las entradas del diario del grupo de transacciones desde el receptor del diario 26;
- ii) todas las entradas se asignan a un grupo de serialización por "defecto";
- iii) las entradas se almacenan en la unidad de almacenamiento 24 en el modo usual pero se enlazan en el orden inverso (es decir, la cabeza de la lista es la última entrada en la unidad de almacenamiento, con los enlaces moviéndose hacia atrás hasta la primera entrada en la unidad de almacenamiento);
- iv) si se rellena la unidad de almacenamiento antes de que se completen las entradas del grupo de transacciones, la unidad de almacenamiento se pone sobre la cola LIFO TLQ 27 (en lugar de liberarla para el grupo de serialización por defecto). A continuación, se asigna una nueva unidad de almacenamiento (como normal) y se continúa almacenando las entradas;
- v) cuando las entradas del diario disponibles del grupo de transacciones se han recibido completamente y se almacenan en las unidades de almacenamiento, las unidades de almacenamiento se envían a los grupos de serialización por defecto en el orden LIFO. El resultado es que el grupo de serialización recibe las entradas del diario en el orden inverso (desde la más reciente a la más antigua);
- vi) los procedimientos del grupo de serialización por defecto procesa las entradas como entradas "inversas" (las entradas incluyen un indicador para indicar que son entradas "inversas"). Esto da como resultado que se eliminan todas las inserciones que se están procesando como borrados, actualizaciones a su imagen anterior y borrados que se insertan, etc. Solo se procesan las entradas del diario que ya se han aplicado (por ejemplo, durante el procesamiento normal);
- vii) el grupo de serialización por defecto no realiza una transacción sobre las entradas "inversas" hasta que recibe la entrada del diario "grupo de transacción de datos". Esto asegura que si se encuentra un fallo durante la "limpieza" de la base de datos, está en un estado conocido. Esto posibilita arrancar de nuevo la limpieza.

Una vez que se han "eliminado" todos los grupos de transacciones "abiertas", el procedimiento de control 19 suspende los otros procedimientos y la base de datos de réplica está lista para usarse como la base de datos primaria.

ES 2 395 205 T3

Este procedimiento permite una rápida "limpieza" de las transacciones parcialmente aplicadas que no requiere usar la capacidad de procesamiento del sistema a menos que una base de datos secundaria tenga que hacerse de hecho una base de datos primaria.

El procedimiento y el aparato de la invención proporcionan varias ventajas, a saber:

- 5 1. La asignación de bloques de la unidad de almacenamiento dentro de un espacio objeto de almacenamiento y el control de la lectura / escritura evita la necesidad de bloqueos y los problemas de lecturas / escrituras concurrentes.
- 10 2. El uso de los grupos de serialización posibilita la actualización de los elementos de forma serie y la actualización de los elementos interrelacionados en una cronología correcta. Los grupos de serialización posibilitan que los flujos múltiples de las entradas del diario se procesen simultáneamente mientras que los elementos interrelacionados se procesan juntos.
- 15 3. El uso del índice MBIX reduce enormemente el tiempo de búsqueda para cada una de las entradas del diario. El uso de las cabeceras de la entrada de almacenamiento 21 (asideros) posibilita la rápida localización de siguiente entrada del diario de un grupo de serialización.
4. El uso de un procedimiento de control para evaluar la operación del procedimiento de recepción y el procesamiento dentro de los grupos de serialización posibilita a los subprocesos procesar la información de forma eficaz sin necesidad de interactuar con otros procedimientos.
5. Simple manejo de las transacciones donde las bases de datos secundarias se hacen bases de datos primarias.

- 20 Donde se hace referencia en la descripción anterior a números enteros o componentes que tienen equivalentes conocidos entonces tales equivalentes se incorporan a este documento como si se mostrasen individualmente.

Aunque esta invención se ha descrito a modo de ejemplo se apreciará que pueden realizarse mejoras y/o modificaciones a la misma sin alejarse del ámbito o el espíritu de la presente invención.

REIVINDICACIONES

1. Un procedimiento de réplica de una base de datos para replicar datos desde un sistema de origen (1) a un sistema objetivo (6) que comprende las etapas de:

- 5 i) recepción (7) de las entradas del diario desde el sistema de origen (1) en el sistema objetivo (6) donde cada una de las entradas del diario está asociada con un objeto de la base de datos que identifica una tabla de la base de datos que tiene una o más filas a actualizar, en donde el procedimiento está **caracterizado por:**
- 10 ii) extracción de la información de procesamiento desde una base de datos de asignación donde la información de procesamiento determina si los objetos de la base de datos necesitan agruparse con otros objetos de la base de datos,
- 15 iii) procesamiento de las entradas del diario de acuerdo con la información de procesamiento extraída por a) la identificación del grupo de serialización (8) asociado con el objeto de la base de datos y b) la asignación de las entradas del diario al grupo de serialización identificado (8) en base a la información de procesamiento en la base de datos de asignación, en donde las entradas del diario del mismo tipo o relacionados con otras entradas del diario se procesan en el mismo grupo de serialización (8) y
- iv) el procesamiento de los grupos de serialización (8) con las entradas del diario asignadas en paralelo para replicar los datos desde el sistema de origen (1).

20 2. El procedimiento de la reivindicación 1 en el que, cuando cada uno de los objetos de la base de datos se encuentra primero en una operación de procesamiento, se genera una entrada a un índice temporal de objetos que contiene el grupo de serialización y la información de procesamiento relacionada para ese objeto de la base de datos.

3. El procedimiento de la reivindicación 2 en el que, si ya existe una entrada al índice temporal de objetos, el grupo de serialización para un objeto de la base de datos se asigna en base a los datos en el índice temporal de objetos.

25 4. El procedimiento de la reivindicación 2 en el que, si una entrada al índice temporal de objetos no existe, se conduce una búsqueda del nombre del objeto completo en la base de datos de asignación, el grupo de serialización se obtiene a partir de la base de datos de asignación y se genera una entrada al índice temporal de objetos.

5. El procedimiento de una cualquiera de las reivindicaciones anteriores en el que la información de procesamiento para un objeto de la base de datos se mantiene en una memoria dinámica.

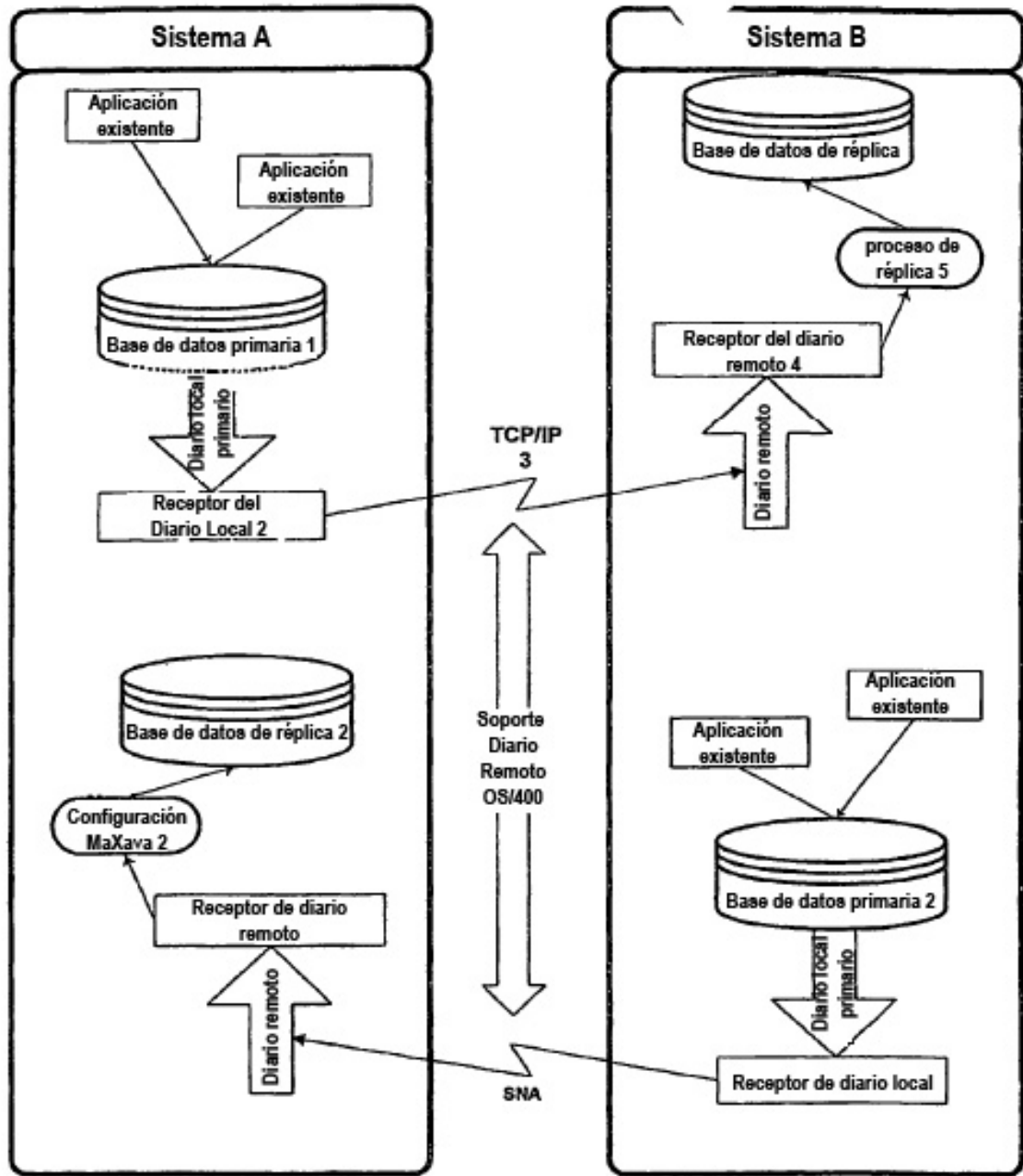


FIGURA 1

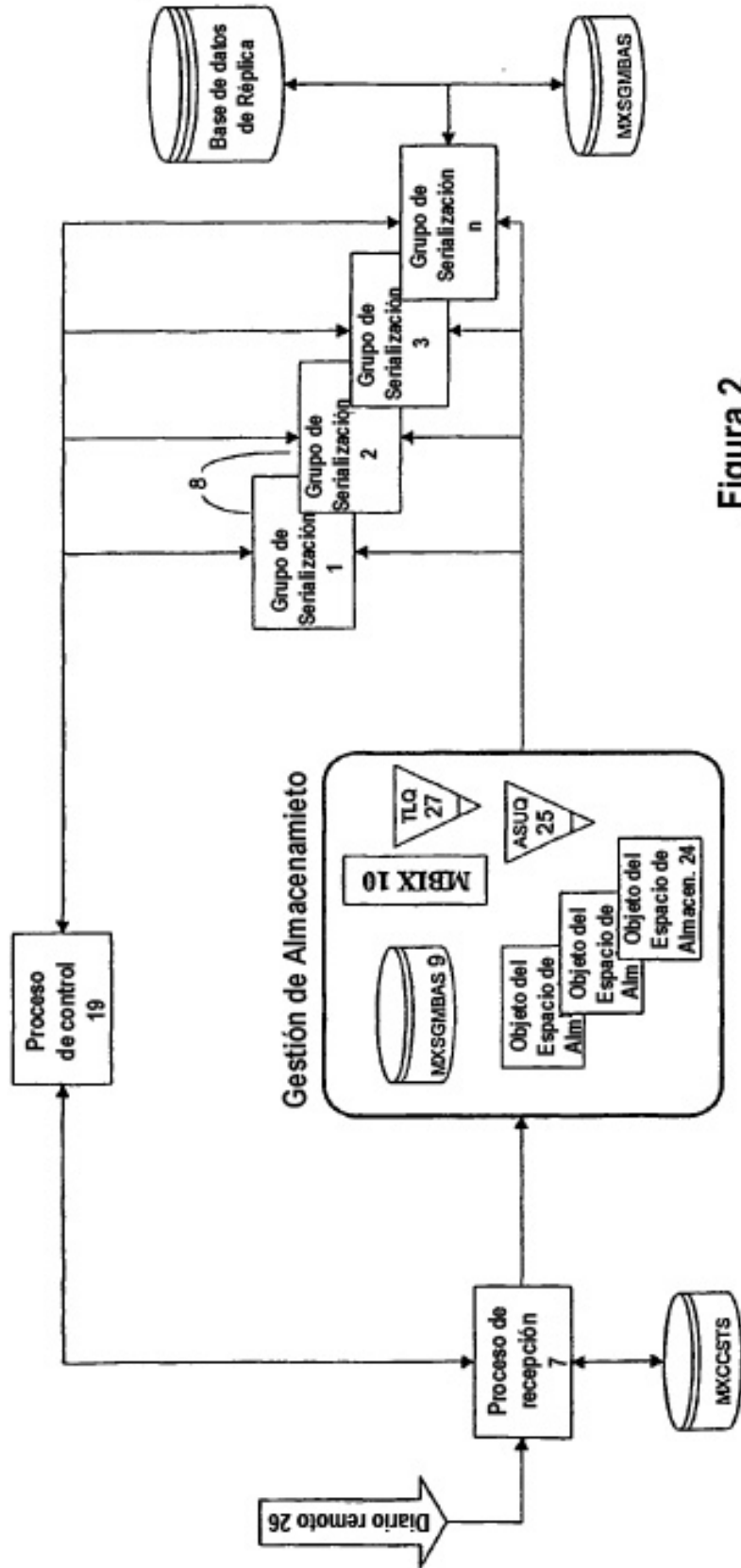


Figura 2

Objeto del Espacio de Almacenamiento

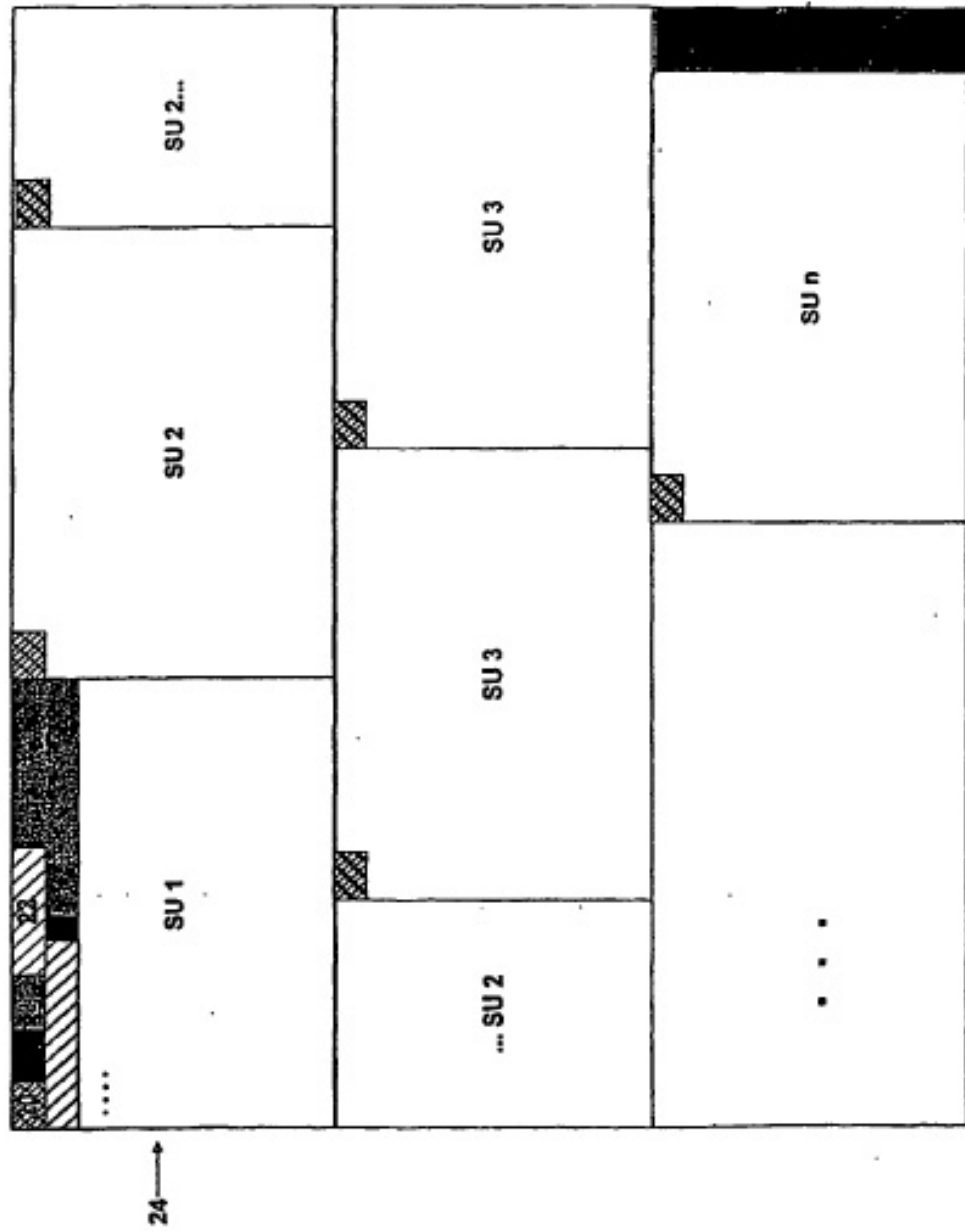


Figura 3

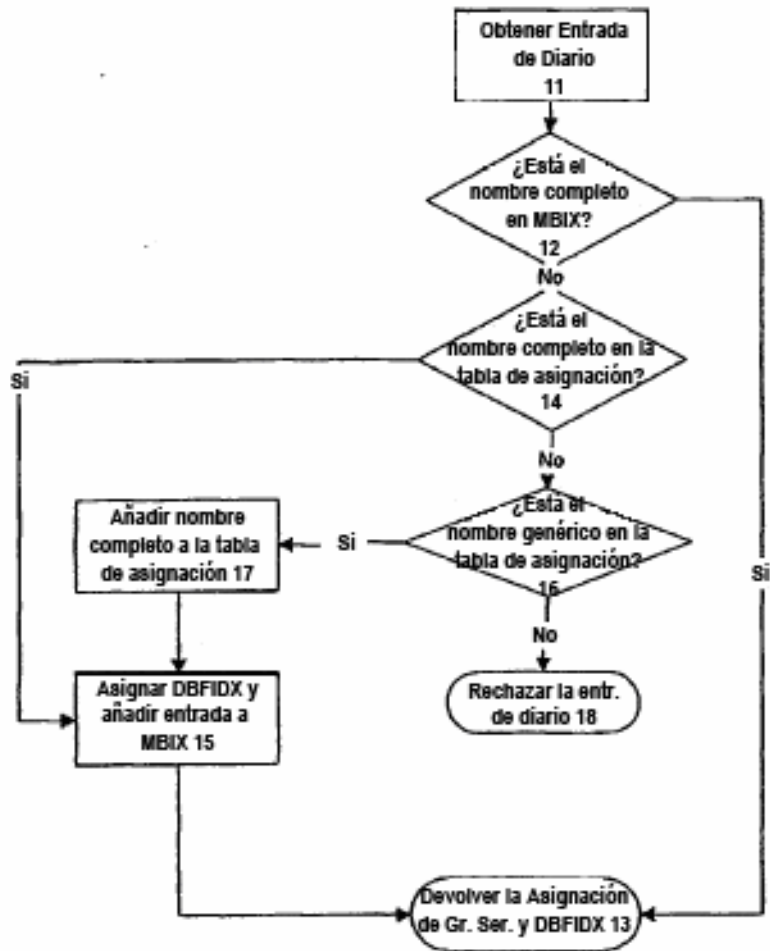


Figura 4