

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 698 326**

51 Int. Cl.:

G06F 17/30 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.12.2013 PCT/GB2013/053308**

87 Fecha y número de publicación internacional: **26.06.2014 WO14096796**

96 Fecha de presentación y número de la solicitud europea: **16.12.2013 E 13806052 (0)**

97 Fecha y número de publicación de la concesión europea: **17.10.2018 EP 2936344**

54 Título: **Archivo de datos utilizable en búsquedas**

30 Prioridad:

**20.12.2012 GB 201223060
13.02.2013 EP 13275027**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
04.02.2019

73 Titular/es:

**BAE SYSTEMS PLC (100.0%)
6 Carlton Gardens
London SW1Y 5AD, GB**

72 Inventor/es:

**REID, IAIN NORMAN NICOL;
JARVIS, RICHARD THOMAS;
WINFIELD, DAFYDD HUW LEWIS y
GARDINER, PETER STUART**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 698 326 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

Archivo de datos utilizable en búsquedas

5 Esta invención se refiere al almacenamiento de registros de transacciones y en particular, aunque no de manera exclusiva, a técnicas para un almacenamiento y posterior recuperación eficientes de registros de transacciones basados en el tiempo en un AOParchivo de datos, por ejemplo, para el almacenamiento a largo plazo de registros de llamadas de una red de telecomunicaciones de línea fija o móvil.

10 La legislación reciente ha aumentado la responsabilidad sobre los operadores de telecomunicaciones, en particular, para mantener los registros de llamadas de forma que se puedan buscar y recuperar, bajo petición, en cualquier momento durante un período de hasta dos años desde la fecha en la que se generaron los registros de llamada, por ejemplo, para apoyar una investigación policial. Dado el enorme número de llamadas realizadas en las redes de telecomunicaciones móviles y fijas cada día, los grandes requisitos de almacenamiento de datos para los registros de llamadas resultantes conllevan una inversión financiera significativa por parte de los operadores respectivos, en particular, a la vista de la necesidad de que cualquier método seleccionado de almacenamiento debe ser capaz de soportar las peticiones posteriores de recuperación. Dichas peticiones, que con frecuencia se realizan con escasa anticipación y que requieren una respuesta muy rápida, pueden requerir una capacidad de búsqueda y recuperación de todos los registros de transacciones que coincidan con los criterios de búsqueda en base a cualquier combinación de parámetros almacenados.

El documento US 2008/104149 A1 (VISHNIAC EPHRAIM MERIWETHER [US] ET AL) del 1 de mayo de 2008 expone un método de gestión del almacenamiento de unidades de datos accesibles individualmente.

20 Desde un primer aspecto, la presente invención reside en un método para generar y buscar un archivo de registros de transacciones, en el que cada registro de transacciones comprende valores para cada uno de una pluralidad de campos de datos, de los cuales al menos uno se refiere al tiempo, donde el método comprende los pasos de:

(i) recibir una pluralidad de registros de transacciones;

25 (ii) recopilar los registros de transacciones recibidos en una pluralidad de grupos de registros de transacciones de acuerdo con un criterio de agrupación predeterminado;

(iii) generar uno o más índices de primer nivel para la pluralidad de grupos de registros de transacciones, en el que cada uno del o de los índices de primer nivel se basa en al menos un campo de datos asociado con los registros de transacciones, y en el que el o los índices de primer nivel incluyen datos suficientes únicamente para identificar uno o más grupos de registros de transacciones;

30 (iv) aplicar un algoritmo de compresión seleccionado para comprimir cada uno de la pluralidad de grupos recopilados en el paso (ii); y

(v) almacenar el o los índices de primer nivel y los grupos de registros de transacciones comprimidos,

35 en el que dicho criterio de agrupación predeterminado se define de acuerdo con el algoritmo de compresión seleccionado y con el contenido de los registros de transacciones recibidos, de modo que se recopila al menos una cantidad predeterminada de datos para formar un grupo de registros, con el fin de lograr de ese modo un nivel predeterminado de eficacia de la compresión en el paso (iv);

40 (vi) recibir los criterios de búsqueda para la recuperación de los registros de transacciones dentro de los grupos comprimidos, donde los criterios de búsqueda definen un valor o rango de valores para uno o más campos de datos representados en los registros de transacciones, lo que incluye al menos un campo de datos representado en el o los índices de primer nivel;

(vii) utilizar uno del o de los índices de primer nivel almacenados generados con respecto a los grupos de registros de transacciones comprimidos y almacenados para seleccionar uno o más grupos de registros de transacciones comprimidos relacionados con los criterios de búsqueda recibidos;

(viii) descomprimir el o los grupos de registros de transacciones seleccionados;

45 (ix) identificar uno o más registros de transacciones individuales, dentro del o de los grupos de registros de transacciones descomprimidos, cuyo contenido coincida con los criterios de búsqueda recibidos,

50 en el que (ix) comprende generar uno o más índices de segundo nivel para los registros de transacciones individuales contenidos dentro de los registros de transacciones descomprimidos y utilizar el o los índices de segundo nivel para identificar y recuperar el o los registros de transacciones individuales que coinciden con los criterios de búsqueda recibidos.

Las realizaciones preferidas de la presente invención pueden reducir el volumen de almacenamiento de datos necesario para capturar un, a menudo, vasto número de registros de transacciones y además almacenarlos de tal manera que estos se puedan recuperar y buscar rápidamente, con cualquier nivel de detalle que sea necesario, en una investigación posterior.

5 Preferentemente, en el paso (ii), los criterios de agrupación predeterminados definen un número mínimo de registros de transacciones o un volumen mínimo de datos que deben estar contenidos dentro de un grupo de registros de transacciones que se recopilan para la compresión. Como alternativa, los criterios de agrupación predeterminados definen una medida de la diversidad en el contenido de datos que se debe lograr cuando se recopila un grupo de registros de transacciones para la compresión. No obstante, aunque dichos criterios pueden ayudar al aumento de la
10 eficacia de la compresión de un algoritmo de compresión elegido, algunos criterios de agrupación alternativos adicionales pueden definir un rango de valores de uno o más tipos de datos relacionados con los registros en un grupo de registros de transacciones que se recopilan para la compresión.

15 En una aplicación con un caudal de datos elevado así como también un volumen de datos elevado, en el paso (ii), los criterios de agrupación predeterminados pueden comprender además los criterios para asignar uno o más grupos recopilados de registros de transacciones a uno de una pluralidad de fragmentos de procesamiento y almacenamiento, y en el que, en el paso (iii), cada uno del o de los índices de primer nivel se generan mediante, y con respecto a los registros a los que se asignan, uno diferente de dicha pluralidad de fragmentos.

20 En una realización preferida, el nivel de eficacia de la compresión predeterminado comprende al menos una reducción de un 50% en el almacenamiento de datos requerido para un grupo de registros de transacciones comprimido en comparación con el requerido para un grupo no comprimido respectivo. En una realización preferida adicional, el nivel de eficacia de la compresión predeterminado comprende al menos una reducción de un 70% en el almacenamiento de datos requerido para un grupo de registros de transacciones comprimido en comparación con el requerido para un grupo no comprimido respectivo. En una realización preferida adicional más, el nivel de eficacia de la compresión predeterminado comprende al menos una reducción de un 90% en el almacenamiento de datos
25 requerido para un grupo de registros de transacciones comprimido en comparación con el requerido para un grupo no comprimido respectivo.

Entre las aplicaciones preferidas de la presente invención, los registros de transacciones pueden comprender registros de llamadas de telecomunicaciones o registros de transacciones financieras.

30 De manera conveniente, se puede utilizar la misma funcionalidad de indexación para crear índices en el primer y segundo nivel.

En una realización preferida, para aumentar la velocidad con la que se pueden recuperar y buscar los datos almacenados, el paso (ix) comprende además almacenar de manera permanente el o los índices de segundo nivel.

35 Desde un segundo aspecto, la presente invención reside en un producto de programa informático que comprende un portador de datos que tiene almacenado en este, o que comprende unos medios para acceder y facilitar la ejecución de, unos medios de código de programa informático, que cuando se cargan y ejecutan en un ordenador hacen que el ordenador implemente el método de acuerdo con el primer o segundo aspecto de la presente invención.

Ahora se describirán con más detalle, únicamente a modo de ejemplo, las realizaciones preferidas de la presente invención, haciendo referencia a los dibujos anexos en los cuales:

40 la figura 1 es un diagrama de flujo que muestra, de manera esquemática, los pasos en un método de almacenamiento de datos de transacciones, de acuerdo con una realización preferida de la presente invención;

la figura 2 es un diagrama de flujo que muestra, de manera esquemática, los pasos en un método de búsqueda y recuperación preferido para acceder a los datos de transacciones almacenados, de acuerdo con el método de la figura 1;

45 la figura 3 es un diagrama que muestra los módulos funcionales clave en un aparato preferido para implementar el método de almacenamiento de datos esquematizados en la figura 1, de acuerdo con una realización preferida de la presente invención; y

la figura 4 es un diagrama que muestra los módulos funcionales clave en un aparato preferido para implementar el método de búsqueda y recuperación esquematizado en la figura 2, de acuerdo con una realización preferida de la presente invención.

50 La presente invención se refiere a la captura, el almacenamiento y, según se requiera, la posterior recuperación de datos de transacciones. Los datos de transacciones se generan habitualmente en forma de registros de transacciones que registran los detalles de un tipo particular de evento cada vez que se produce dicho evento. Los datos capturados en un registro de transacciones incluirán habitualmente un registro de la fecha y hora del evento

junto con uno o más parámetros suficientes para definir el evento. Los datos específicos registrados se determinarán usualmente mediante los objetivos para la captura de los datos de transacciones. Por ejemplo, dichos datos pueden formar la base de un proceso de facturación para una tarificación de servicios basada en las transacciones, p. ej., en servicios de telecomunicaciones donde se solicita el pago a los usuarios con respecto a las llamadas que estos inician. En otros sistemas, los datos capturados pueden proporcionar estadísticas de utilización con fines de planificación y gestión de la capacidad.

La presente invención se puede aplicar a la captura y almacenamiento de cualquier forma de datos de transacciones. Las fuentes habituales incluyen, además de los operadores de telecomunicaciones: instituciones financieras que registran las transacciones financieras; compañías de transporte público en las que se activa el pago por viajar mediante la denominada "tarjeta magnética"; y cualquier tipo de instalación de acceso seguro en la que se utilizan cerraduras de puertas controladas por tarjetas magnéticas y se generan registros de transacciones que capturan los instantes de entrada y los datos de la tarjeta magnética que identifican los individuos que acceden.

Los datos de transacciones pueden servir para otros fines además de para su necesidad original. Por ejemplo, se puede proporcionar una instalación de búsqueda y recuperación basada en parámetros para facilitar la identificación y recuperación de los registros de transacciones relacionados con unos valores o rangos de parámetros particulares. No obstante, es necesario equilibrar dicho requisito con los problemas prácticos de almacenamiento de datos y su accesibilidad, en particular si dichos fines requieren que los registros se mantengan durante un período de tiempo apreciable después de su generación. Este es el caso con los registros de llamadas de telecomunicaciones que pueden proporcionar información útil a la policía del paradero de individuos particulares sujetos a investigaciones penales. La legislación en vigor en la actualidad en el Reino Unido impone un período mínimo de mantenimiento de los registros de llamadas de telecomunicaciones. Dado que se generan aproximadamente 700 millones de registros de llamadas al día en el Reino Unido, a partir de 2011, el volumen de datos que se debe capturar y almacenar, de forma que se puedan buscar y recuperar rápidamente, es muy significativo e impone unos costes de sobrecarga adicionales a los operadores de telecomunicaciones para conservar dichos datos de una forma fácilmente recuperable. La presente invención tiene como objetivo aliviar los problemas asociados con la captura y recuperación de registros de transacciones que se generan en dichos volúmenes.

Las realizaciones preferidas de la presente invención implementan un proceso de captura y recuperación de datos que opera habitualmente como dos subprocesos separados en el tiempo: un primer subproceso relacionado con la captura y almacenamiento continuo de los registros de transacciones como un proceso de manipulación masiva de datos; un segundo subproceso relacionado con la búsqueda y recuperación posterior e intermitente de los registros de transacciones específicos desde los datos almacenados. A modo de compendio, ahora se proporcionará una descripción esquematizada del primer y segundo subproceso, el primero haciendo referencia a la figura 1 y el segundo haciendo referencia a la figura 2.

Haciendo referencia en primer lugar a la figura 1, un diagrama de flujo resume los pasos esenciales en un subproceso de captura y almacenamiento de datos preferido, que opera habitualmente de manera continua y como diferentes procesos en paralelo, comenzando en el PASO 10 con la captura de los registros de transacciones, habitualmente mediante transmisión continua a través de una red o mediante otro método de transferencia desde una fuente de registros de transacciones. En el PASO 15, los registros de transacciones capturados se importan en un formato de datos predeterminado, si es diferente del formato de los registros capturados, incluyendo este paso cualquier identificación necesaria de tipos de datos esenciales dentro de una estructura de campos de los registros de transacciones capturados y una comprobación de la calidad de los datos. En el PASO 20, los registros importados se recopilan en grupos basándose en criterios de agrupamiento predeterminados tal como se analizará a continuación. En el PASO 25, se crea un índice de primer nivel, o se actualiza, basándose en uno o más de los campos de datos identificados en el PASO 15, y el índice se almacena de manera permanente. La elección del tipo o tipos de datos indexados tiene como objetivo facilitar la selección posterior de al menos uno de los grupos de registros recopilados, del PASO 20, relacionados con un período de tiempo específico o con un rango de valores específico de un tipo de datos indexados. En el PASO 30, cada uno de los grupos de registros recopilados se comprimen utilizando un algoritmo de compresión de datos predeterminado, y en el PASO 35 los grupos de registros de transacciones comprimidos se almacenan de forma recuperable.

Habiendo capturado, recopilado, indexado y almacenado conjuntos de registros de transacciones, puede ser que estos permanezcan en el almacenamiento durante un período de tiempo legal, después del cual se borran. No obstante, en el caso de que se necesiten identificar y recuperar esos registros de transacciones particulares, habitualmente con muy poco tiempo de preaviso, se implementa el siguiente subproceso de búsqueda y recuperación preferido en la presente invención.

Haciendo referencia a la figura 2, un diagrama de flujo resume el subproceso de búsqueda y recuperación preferido que comienza en el PASO 50 con la recepción de los criterios de búsqueda definidos por el usuario. En el PASO 55, los criterios de búsqueda se utilizan inicialmente para identificar uno o más grupos de registros de transacciones almacenados y comprimidos sobre la base del o de los tipos de datos esenciales representados en el índice de primer nivel. La indexación, y por lo tanto la identificación de uno o más grupos de registros, se hará habitualmente

sobre la base de los períodos de tiempo de las transacciones y el grupo o los grupos de registros identificados serán aquellos que potencialmente contienen los registros de transacciones relacionados con los momentos dentro de un período definido. No obstante, se puede proporcionar una indexación de primer nivel y una recuperación sobre la base de cualquier tipo de datos representado en un registro de transacciones.

5 En el PASO 60, se recuperan y descomprimen el grupo o los grupos de registros identificados a su formato importado o convertido. A continuación, se indexan los registros descomprimidos, en el PASO 65, para crear un índice de segundo nivel con un mayor nivel de detalle utilizando cualquier selección de los tipos de datos identificados representados dentro de la estructura de campos de datos de los registros de transacciones, que incluya preferentemente y además de aquel o aquellos tipos de datos representados en el índice de primer nivel.
 10 Preferentemente, se utiliza la misma técnica de indexación para crear el índice de segundo nivel a la que se utilizó para crear el índice de primer nivel. El índice de segundo nivel está pensado para permitir la identificación de registros de transacciones particulares, no solo sobre la base de intervalos de tiempo más específicos de los que se requerirían para identificar grupos de registros almacenados, sino también sobre la base de otros datos almacenados en los registros de transacciones, tales como el identificador del número o la ubicación del teléfono al que se llama en el caso de registros de llamadas de telecomunicaciones. Por lo tanto, en el PASO 70, los registros de transacciones se identifican y extraen sobre la base de los criterios de búsqueda definidos por el usuario utilizando el índice de segundo nivel.

Al generar únicamente un índice de primer nivel antes del almacenamiento de los grupos de registros de transacciones, suficiente únicamente para identificar grupos de registros particulares almacenados, el proceso de
 20 captura y almacenamiento de registros es muy rápido y se puede minimizar la capacidad de almacenamiento de datos requerida para los grupos de datos comprimidos y el índice de primer nivel. Una característica particularmente conveniente de la presente invención se refiere a la recopilación de los registros en grupos sobre la base de unos criterios de agrupamiento predeterminados. Los criterios de agrupamiento requieren que se recopile al menos una cantidad de datos predeterminada para formar un grupo de registros antes de la compresión, con el fin de garantizar que la eficacia del algoritmo de compresión de datos excede un nivel predeterminado. Esto puede suponer la
 25 captura de datos a lo largo de un período de tiempo variable que depende de la tasa de generación de los registros de transacciones por parte de la fuente, de modo que afecta al nivel de resolución de los datos proporcionados por el índice de primer nivel. No obstante, la eficacia de la compresión se puede definir no solo mediante el grado de compresión logrado, de modo que se reduzcan los requisitos de almacenamiento de datos, sino también, cuando es necesario capturar y almacenar volúmenes de datos muy grandes, por la velocidad del proceso global de compresión y almacenamiento. Los niveles de compresión habituales pueden exceder un 90% de reducción del volumen de datos almacenado si la composición del grupo se elige correctamente. Con frecuencia, los datos de transacciones son muy variables con poca repetición dentro y en los distintos registros. Esto limita la eficacia de los algoritmos de compresión de datos conocidos en el caso de que estuvieran limitados a comprimir únicamente unas
 30 cantidades relativamente pequeñas de registros.

Ahora se describirán con más detalle los pasos particulares dentro de cada uno de los subprocesos descritos anteriormente de manera esquemática.

Haciendo referencia de nuevo a la figura 1, el PASO 10, capturar los registros de transacciones, comprende recibir los registros de transacciones desde una fuente conocida tanto mediante la transmisión de los registros a través de
 40 una red, a medida que se crearon o a intervalos regulares en forma de lotes de registros descargados a través de la red, como por medio de otra clase de portador de datos. El proceso de captura también garantiza que cualesquiera datos externos esenciales, tal como un identificador de la fuente o el período de tiempo con el que están relacionados los registros, si no se registran dentro de los propios registros, también son capturados y asociados con los registros de transacciones capturados respectivos en esta etapa.

El PASO 15, importar los registros de transacciones capturados, comprende el procesamiento inicial de los registros de transacciones capturados, por ejemplo, importar esos registros de datos en un formato común, según se requiera, de acuerdo con una plantilla de estructura de datos asociada con su fuente, o mediante otros medios de análisis. Si se requiere, también se puede llevar a cabo una comprobación de calidad de los datos como parte del paso de importar, lo que facilita que los registros que están incompletos, o no tienen un valor posterior debido a otras
 50 razones tales como la corrupción de los datos, sean rechazados en esta etapa en lugar de consumir capacidad de procesamiento y almacenamiento adicional.

En el PASO 20, recopilar los registros importados, se lleva a cabo un agrupamiento de los registros de transacciones importados de acuerdo con unos criterios de agrupamiento predeterminados. Los criterios de agrupamiento se diseñan no solo para mejorar la eficacia del algoritmo de compresión de datos implementado en el PASO 30, sino también para optimizar los pasos de indexación y almacenamiento de los registros recopilados, tal como se analizará con más detalle a continuación, lo que facilita la asignación de uno o más grupos de registros para su procesamiento mediante uno de múltiples hilos de procesamiento paralelo, donde cada hilo se inicia de acuerdo con la demanda. El criterio de agrupamiento también puede tener en cuenta los requisitos para la búsqueda y recuperación de los registros comprimidos almacenados, por ejemplo, garantizando que se alcanza un equilibrio

entre el tamaño de los grupos de registros recopilados y la velocidad potencial de identificación de los grupos de registros relevantes y de registros individuales dentro de estos.

Los criterios de agrupamiento pueden definir cualquiera de un número de diferentes parámetros de los que se conoce que afectan a la eficacia del algoritmo de compresión de datos. Dichos parámetros pueden incluir: volumen de datos; medidas de la diversidad de los datos dentro y en los distintos registros de transacciones; y los rangos de valores para los campos de datos de los registros de transacciones seleccionados. La recopilación puede dar como resultado unos agrupamientos de registros que varían en su volumen de datos y en los intervalos de tiempo u otros rangos de datos que representen, con el fin de satisfacer los criterios de agrupamiento. La recopilación también puede tener en cuenta los datos externos capturados en el PASO 10.

En el PASO 25, crear un índice de primer nivel, se utiliza una técnica de indexación conocida para generar o extender un índice de primer nivel basándose en uno o más campos de datos en los registros de transacciones o en los datos externos capturados en el PASO 10. El índice de primer nivel está pensado para registrar un enlace entre los valores de un campo de datos indexado y uno o más grupos de registros de transacciones recopilados que probablemente incluyan registros que contienen, o se relacionan con, dichos valores. El índice de primer nivel representa únicamente los datos suficientes para identificar uno o más grupos de registros de transacciones y no pretende hacer referencia a registros individuales dentro de un grupo. De hecho, un grupo de registros identificado puede no contener ningún registro que tenga un valor particular definido para uno de los campos indexados, aunque contendría registros dentro de un rango de valores que incluye el valor definido. De esta manera, el índice de primer nivel tiene el beneficio de ser simple y rápido de crear y de imponer una sobrecarga de almacenamiento de datos pequeña.

Cuando, tal como se hace referencia anteriormente con relación al PASO 20, los criterios de agrupamiento asignen uno o más grupos de registros a un hilo de procesamiento de datos diferenciado, se crea el índice de primer nivel para indexar aquellos grupos de registros de transacciones que se procesan dentro de ese hilo, creándose un índice de primer nivel diferente para cada hilo de procesamiento de ocurrencia reciente. También se puede generar y mantener en el primer nivel un índice adicional para enlazar el campo o los campos de datos indexados con los supergrupos de registros, un supergrupo de grupos que comprende aquellos grupos procesados mediante un hilo de procesamiento diferenciado.

En una realización alternativa, se puede crear un único índice de primer nivel para indexar todos los grupos de registros recopilados. En ese caso, la recopilación de registros importados recientemente en un nuevo grupo activa una extensión del índice de primer nivel.

El PASO 30, comprimir los registros de transacciones importados, comprende la ejecución de un algoritmo de compresión de datos conocido para comprimir cada grupo de registros de transacciones recopilados, con el fin de generar un archivo de datos comprimidos recuperables de manera independiente. Preferentemente, se utiliza un algoritmo de compresión, tal como el ampliamente conocido PZip, aunque se pueden utilizar algoritmos de compresión alternativos tal como le es conocido a un experto en el campo de la compresión de datos. La eficacia de los algoritmos de compresión, tales como el PZip, está influenciada por, entre otros factores, la estructura de los datos a comprimir, el nivel de repetición en los datos u otras medidas de la diversidad de los datos. Los criterios de agrupamiento predeterminados, mencionados anteriormente, están pensados para dar como resultado grupos de registros de transacciones para los cuales el algoritmo de compresión elegido puede lograr un nivel preferido de eficacia de la compresión. Dada la naturaleza de los registros de transacciones, que con frecuencia contienen poca repetición de los datos dentro de, y entre, los registros, un criterio de agrupamiento puede definir un número mínimo de registros de transacciones de un tipo particular para garantizar un nivel de confianza dado en la redundancia de datos. Preferentemente, se puede lograr un nivel de compresión de un 90% o más, es decir, la generación de un archivo de datos comprimidos que ocupa un 10% o menos de los requisitos de almacenamiento de datos descomprimidos, con el algoritmo PZip con un conjunto preferido de criterios de agrupamiento.

Habiendo alcanzado un nivel preferido de compresión de los grupos de registros de transacciones importados, los archivos de datos comprimidos resultantes se almacenan en el PASO 35 de una forma recuperable en una instalación de almacenamiento masivo de datos o en portadores de datos adecuados. Cuando sea apropiado, el índice de primer nivel puede registrar la información que identifica la ubicación física de los archivos de datos almacenados, por ejemplo, cuando los archivos de datos comprimidos se almacenan en medios extraíbles.

Preferentemente, tal como se analizó anteriormente, cuando los supergrupos de registros se procesan en hilos de procesamiento de datos diferentes, los archivos de datos comprimidos resultantes y el índice de primer nivel resultante se pueden almacenar de manera permanente en un área de almacenamiento de datos diferenciada para cada hilo.

En el caso de que sea necesario recuperar los registros de transacciones almacenados, la presente invención proporciona una funcionalidad de búsqueda y recuperación para facilitar la identificación y recuperación de cualesquiera registros de transacciones relacionados, o que contienen datos que coinciden, con los criterios de

búsqueda definidos.

5 Se proporciona un motor de búsqueda y recuperación, que se describe con detalle a continuación, para implementar la funcionalidad de búsqueda y recuperación esquematizada anteriormente haciendo referencia a la figura 2, que facilita a un usuario iniciar una consulta estructurada o una búsqueda de texto libre de datos almacenados. La consulta estructurada puede comprender unos valores o rangos de valores definidos por el usuario de campos de datos particulares que se conoce que existen dentro de la estructura de campos de un registro de transacciones. Estos campos de datos pueden incluir cualesquiera campos de datos que forman la base del índice de primer nivel, de modo que se pueda realizar una selección inicial de uno o más grupos de registros sobre los cuales llevar a cabo una búsqueda de segundo nivel. De manera similar, se puede iniciar una búsqueda de texto libre siguiendo una selección de primer nivel de grupos de registros a buscar, aunque en principio se puede llevar a cabo una búsqueda de texto libre de todo el almacenamiento de datos dado un período de tiempo lo suficientemente prolongado. El proceso esquematizado anteriormente haciendo referencia a la figura 2 se describirá a continuación con más detalle.

10 Haciendo referencia de nuevo a la figura 2, el PASO 50, recibir los criterios de búsqueda definidos por el usuario, hace referencia a la introducción por parte de un usuario de los valores o rangos de valores de uno o más campos de datos, de los que se conoce que están contenidos en los registros de transacciones almacenados, como la base para identificar y extraer los registros coincidentes. Los criterios de búsqueda definen al menos un valor o rango de valores de un campo de datos que forma la base del índice de primer nivel.

15 El PASO 55, identificar los grupos de registros a buscar, comprende tomar un valor o rango de valores introducidos por parte de un usuario de al menos un tipo de datos representado en el índice de primer nivel y utilizar estos datos introducidos, a través del índice de primer nivel, para identificar uno o más grupos de registros almacenados que se relacionan con el valor o rango de valores introducido. A continuación, el grupo o grupos de registros identificados se recuperan del almacenamiento, en el PASO 60, y se descomprimen utilizando un algoritmo de descompresión correspondiente al algoritmo de compresión utilizado en el PASO 25, al que se hace referencia anteriormente. Habiendo restaurado el grupo o grupos de registros recuperados a su formato importado, el motor de búsqueda y recuperación genera un índice de segundo nivel de todos los registros recuperados, en el PASO 65, utilizando una técnica de indexación conocida y basándose en uno o más tipos de datos de los que se conoce que están comprendidos en la estructura de campos de los registros de transacciones recuperados. Los criterios de búsqueda definidos por el usuario pueden definir un valor o rango de valores para cualquier combinación de los tipos de datos representados en el índice de segundo nivel, de modo que se puedan identificar y presentar los registros relevantes.

20 La generación de un índice de segundo nivel del grupo o grupos de registros recuperados facilita, en el PASO 70, una identificación y extracción particularmente rápida de los registros de transacciones que coinciden con los criterios de búsqueda introducidos por un usuario. También existe la ventaja de que el tiempo de procesamiento requerido en el PASO 65 para la generación del índice de segundo nivel más complejo y el almacenamiento de datos consiguiente se consume únicamente en el caso de que sea necesario buscar y recuperar registros, en lugar de durante el proceso de captura y almacenamiento de datos representado en la figura 1. El índice de segundo nivel se restringe al grupo o grupos de registros recuperados utilizando el índice de primer nivel, y por lo tanto es mucho más pequeño y se crea más rápidamente que un índice a todos, o a un subconjunto más grande, de los registros almacenados. Preferentemente, el índice de segundo nivel se crea utilizando la misma técnica de indexación a la utilizada para crear el índice de primer nivel.

25 En un planteamiento alternativo, se puede crear un único índice para el segundo nivel en el PASO 25, y se puede almacenar de modo que no se requiera la generación del índice de segundo nivel en el PASO 65. Si las restricciones temporales son más una limitación en la etapa de captura y almacenamiento de datos, el subproceso de la figura 1, que, en la etapa de búsqueda y recuperación, el subproceso de la figura 2, en ese caso se puede preferir la generación independiente de dos índices con niveles de detalle diferentes en el PASO 25 y el PASO 65.

30 A continuación, haciendo referencia a la figura 3 y la figura 4 se describirá un aparato de almacenamiento y recuperación de datos de transacciones preferido, para implementar los subprocesos descritos anteriormente haciendo referencia a la figura 1 y la figura 2, respectivamente.

35 Haciendo referencia inicialmente a la figura 3, y adicionalmente a la figura 1, un aparato 100 comprende dos subsistemas principales para implementar el proceso de la figura 1 para la captura y el almacenamiento de registros de transacciones. El primero, un motor de adquisición 105, se proporciona para capturar los datos de transacciones descargados desde los dispositivos registro de transacciones 110 a través de una red 115 (o mediante otro medio tal como los dispositivos de lectura de medios fijos, no se muestran en la figura 3) y para llevar a cabo el procesamiento inicial de los datos capturados de acuerdo con el PASO 10 y el PASO 15 de la figura 1. El segundo, un motor de indexación y almacenamiento 120, se proporciona para llevar a cabo la indexación y el almacenamiento de los datos de transacciones procesados inicialmente de acuerdo con el PASO 20, 25, 30 y 35 de la figura 1, utilizando técnicas preferidas, que se describirán con más detalle a continuación, que facilitan un almacenamiento de datos rápido y eficiente y una recuperación posterior efectiva.

También se puede proporcionar un tercer subsistema, un motor de búsqueda y recuperación 125, mencionado anteriormente, dispuesto de modo que implemente la funcionalidad de búsqueda y recuperación de la figura 2 de los datos almacenados por el aparato 100. El motor de búsqueda y recuperación 125 se puede integrar más o menos con el aparato 100, de acuerdo con las aplicaciones particulares a las cuales se debe aplicar la presente invención.

Por ejemplo, la búsqueda y recuperación de los datos capturados y almacenados por el aparato 100 se puede llevar a cabo mediante un módulo de software independiente, integrado de manera opcional con software de terceros, que implementa la funcionalidad de búsqueda y recuperación preferida de la presente invención en el contexto de una aplicación específica. Para otras aplicaciones, puede ser conveniente integrar la funcionalidad de búsqueda y recuperación con los otros dos subsistemas 105, 120 del aparato 100, de modo que todas las funcionalidades de captura, almacenamiento, búsqueda y recuperación de datos se puedan integrar dentro de un único producto.

El motor de adquisición 105 está provisto de una interfaz de red 130 para facilitar el acceso a las fuentes de datos, p. ej., a los dispositivos de registro de transacciones 110, a través de una red 115. Los datos descargados desde los dispositivos de registro de transacciones 110 pueden existir en un número de formatos diferentes. Se proporciona un conversor de formatos 135 con el objetivo tanto de reestructurar los datos recibidos en un formato predeterminado que tiene unos campos de datos identificables, tal como se define en un almacenamiento de referencia 140 de estructuras de datos o en las plantillas de registros esperadas desde cada fuente 110 de registros de transacciones, como de determinar el formato de los datos de transacciones recibidos al menos en la medida de que se pueda identificar el campo o campos de datos esenciales requeridos para formar un índice de primer nivel. También se proporciona un módulo de análisis de calidad de los datos 145 para llevar a cabo el preprocesamiento de los datos recibidos, con el fin de garantizar que estos cumplen con los niveles de calidad predeterminados, en particular con respecto a su integridad, legibilidad, precisión, formato, latencia (¿están los datos actualizados?), etc. con relación a otros registros procedentes de la misma fuente o haciendo referencia a la estructura y contenido de los registros esperados de esa fuente, tal como se define en el almacenamiento de estructuras de datos 140 o en cualquier otro sitio. Esto proporciona una oportunidad para descartar los registros de transacciones que no tendrán uso futuro debido, por ejemplo, a datos perdidos o corruptos.

El motor de adquisición 105 envía los datos procesados inicialmente al motor de indexación y almacenamiento de datos 120. Se proporciona un módulo de recopilación de registros y fragmentación 150 para implementar los criterios de agrupamiento mencionados anteriormente, de acuerdo con el PASO 20 de la figura 1, de modo que la funcionalidad de compresión de datos posterior, en particular, pueda rendir a los niveles requeridos. El módulo 150 también se dispone de modo que implemente lo que se denomina un algoritmo de "fragmentación". El algoritmo de fragmentación se dispone de modo que asigne uno o más grupos, un "supergrupo", de registros recopilados a un "fragmento" de procesamiento y almacenamiento de datos que comprende un hilo de procesamiento de datos de ejecución independiente y un área de almacenamiento de datos 155 diferenciada real o virtualmente en una instalación de almacenamiento masivo de datos 160. El área de almacenamiento 155 para cada fragmento comprende un almacén 165 para los datos comprimidos y un almacén 170 para el índice de primer nivel creado con respecto a los registros procesados en ese fragmento. Se pueden ejecutar múltiples hilos de procesamiento de datos en el motor de indexación y almacenamiento 120 para facilitar un procesamiento y almacenamiento rápido en paralelo de los datos capturados. Por tanto, los pasos de la indexación de primer nivel (PASO 25), la compresión de datos (PASO 30) y el almacenamiento de datos (PASO 35) se pueden implementar dentro de un hilo de procesamiento de datos dedicado a la indexación, compresión y almacenamiento del supergrupo de registros asignados a un fragmento particular, uno de distintos hilos de procesamiento de software en paralelo instanciados de acuerdo con los criterios de agrupamiento en el PASO 20. Cada hilo de procesamiento termina tras la finalización del procesamiento y almacenamiento del supergrupo de registros asignados al fragmento respectivo.

La funcionalidad de indexación se representa en la figura 3 mediante un módulo de indexación 175, dispuesto de modo que implemente una técnica de indexación conocida para generar o actualizar un índice de primer nivel con respecto a los grupos de registros asignados a un fragmento particular. El campo o los campos de datos particulares en los cuales se debe basar el índice de primer nivel se pueden identificar, si es necesario, haciendo referencia al almacén de estructuras de datos 140, de acuerdo con la fuente 110 de los registros que se procesan. Habitualmente, para los datos de transacciones, el índice de primer nivel se generará utilizando los campos de datos y datos temporales, aunque se pueden establecer otras bases para el índice de primer nivel, por ejemplo, la ubicación geográfica de un grupo en un evento tramitado si se captura en, o está asociado con, los registros. El módulo de indexación 175 también se dispone de modo que mantenga un índice de fragmentación (180) que comprende enlaces a los supergrupos de registros que se almacenan (165) con cada fragmento. El índice de fragmentación (180) se puede basar en el mismo campo o campos de datos, o equivalentes, que el índice de primer nivel, de modo que, en respuesta a una consulta de búsqueda y recuperación, se puedan identificar uno o más supergrupos de registros almacenados (165) de potencial relevancia a la consulta y acceder y recuperar sus índices de primer nivel (170) almacenados respectivos. El índice de fragmentación se puede almacenar de manera permanente en un almacén del índice de fragmentación 180 y actualizar con la creación de cada nuevo fragmento, o se puede mantener en una memoria volátil para un acceso más rápido.

El módulo de indexación 175 se puede disponer de modo que aplique la misma técnica de indexación para generar los índices de segundo nivel (PASO 60 en el proceso de la figura 2), bajo demanda, relacionados con los registros

de transacciones recuperados como soporte de la funcionalidad del motor de búsqueda y recuperación 125, tal como se describirá a continuación. De manera opcional, dichos índices de segundo nivel se pueden crear al mismo tiempo que los índices de primer nivel, y se pueden combinar con los índices de primer nivel para formar un único índice que se almacenará (170) junto con los registros comprimidos (165).

5 La funcionalidad de compresión de datos se representa en la figura 3 mediante un módulo de compresión de datos 185, dispuesto de modo que implemente un algoritmo de compresión de datos predeterminado para comprimir los grupos de registros de transacciones asignados a un fragmento particular. Los criterios de agrupamiento implementados por el módulo de recopilación y fragmentación de registros 150 están diseñados para garantizar que la eficacia de la compresión del algoritmo de compresión de datos alcanza un nivel predeterminado. Si se
10 implementa más de un tipo de algoritmo de compresión de datos mediante el módulo de compresión de datos 185, el módulo de recopilación de registros 150 puede implementar diferentes criterios de agrupamiento, siendo seleccionados los criterios de agrupamiento relevantes de acuerdo con el algoritmo de compresión de datos a utilizar para comprimir un grupo de registros particular, o los registros de una fuente 110 particular.

15 Los registros de transacciones comprimidos se almacenan (165) de manera permanente en la instalación de almacenamiento masivo de datos 160, la cual puede estar ubicada de manera remota con relación al aparato 100, si se requiere.

El motor de búsqueda y recuperación 125 mencionado anteriormente se dispone de modo que implemente la funcionalidad de búsqueda y recuperación preferida de la figura 2 sobre los datos almacenados por el aparato 100 en el almacén masivo de datos 160. El motor de búsqueda y recuperación 125 puede estar más o menos integrado
20 con el aparato 100 de acuerdo con las aplicaciones particulares a las cuales se debe aplicar la presente invención. Ahora se describirá, haciendo referencia a la figura 4, un aparato preferido para implementar la funcionalidad de la figura 2.

Haciendo referencia a la figura 4, se proporciona el motor de búsqueda y recuperación 125 con un módulo de interfaz de usuario 200 dispuesto de modo que se comunique con un equipo terminal de usuario 205, por ejemplo, a través de la red 115, con el fin de facilitar (PASO 50 en la figura 2) a un usuario la definición de los criterios de búsqueda para la identificación y recuperación de los registros almacenados. Se proporciona un módulo de búsqueda y recuperación 210 para recibir los criterios de búsqueda definidos por el usuario e implementar inicialmente el PASO 55 en el proceso de la figura 2. En primer lugar, esto conlleva acceder al índice de fragmentación (180) almacenado, creado y actualizado por el módulo de indexación 175, y utilizarlo para identificar
25 una o más áreas de almacenamiento de datos 155 relevantes que es probable que incluyan los registros (165) relevantes para los criterios de búsqueda definidos, al menos sobre la base del campo o campos de datos del índice de primer nivel. A continuación, se recupera el índice de primer nivel (170) almacenado para cada área de almacenamiento de datos 155 identificada y el módulo de búsqueda y recuperación 210 lo utiliza para identificar uno o más grupos de registros (165) comprimidos almacenados específicos, de las áreas de almacenamiento de datos 155 seleccionadas de potencial relevancia para los criterios de búsqueda definidos. A continuación, se recuperan los grupos de registros comprimidos identificados desde sus almacenes 165 respectivos (PASO 60 en la figura 2) y se pasan a un módulo de descompresión de datos 215 para ser restaurados a forma descomprimida, invirtiendo la compresión de datos aplicada previamente a esos registros por el módulo de compresión de datos 185.

30 A continuación, se puede aplicar el módulo de indexación 175, o un módulo de indexación independiente dedicado al motor de búsqueda y recuperación 125, para utilizar la misma técnica de indexación subyacente a la que se utilizó para generar el índice de primer nivel, con el fin de generar un índice de segundo nivel de los registros descomprimidos (PASO 65 en la figura 2) utilizando una selección predeterminada de los campos de datos de los que se conoce que existen en la estructura de datos de registros aplicable. Si es necesario, se puede hacer referencia al almacén de estructuras de datos 140 para determinar el formato de los campos de datos disponibles para los registros particulares recuperados, por ejemplo, sobre la base de su fuente 110.

35 Obviamente, se puede utilizar una técnica de indexado diferente a la utilizada para el índice de primer nivel con el fin de generar el índice de segundo nivel, siempre que esta logre el mismo objetivo de facilitar llevar a cabo una búsqueda en los registros descomprimidos para el nivel requerido mediante los criterios de búsqueda definidos.

40 A continuación, se dispone el módulo de búsqueda y recuperación 210 de modo que utilice el índice de segundo nivel para llevar a cabo una búsqueda a nivel de registro de los registros descomprimidos, de acuerdo con los criterios de búsqueda definidos, para identificar y recuperar los registros de transacciones relevantes (PASO 70 en la figura 2). Los registros recuperados se pueden presentar al usuario a través de la interfaz de usuario 200.

45 En una aplicación típica, la presente invención se puede aplicar a la captura, almacenamiento y recuperación posterior de registros de llamadas de telecomunicaciones en nombre de los proveedores de servicios de telecomunicaciones fijas y móviles. Los registros de llamadas de telecomunicaciones para llamadas de voz de línea fija contienen habitualmente, para cada llamada realizada, al menos una fecha y hora de inicio de la llamada, una fecha y hora de finalización, un número de línea llamante y un número de línea llamado. Los registros de llamadas

para llamadas de teléfono móvil contienen habitualmente estos cuatro elementos de información junto con otros datos, por ejemplo, datos indicativos de la ubicación geográfica del suscriptor del móvil llamante y del suscriptor del móvil llamado respectivo. El motor de adquisición 105 recibe los registros de llamadas, habitualmente en lotes, desde los proveedores de servicios respectivos y, tras la conversión de formato (135) necesaria y la comprobación de calidad de los datos (145), que incluye, si se requiere, la ordenación de los registros por fecha y hora de inicio, pasa los registros al motor de almacenamiento e indexación 120. Allí, se recopilan los registros de llamadas de modo que se logre los niveles de compresión de datos requeridos y el algoritmo de fragmentación asigna los grupos de registros particulares a un fragmento dedicado de procesamiento y almacenamiento de datos, donde los grupos de registros asignados a cada fragmento se indexan a un primer nivel sobre la base del campo de fecha y hora de inicio de las llamadas. También se actualiza un índice de fragmentación 120 sobre la base del intervalo de fecha y hora representado por el supergrupo de registros de llamadas asignado a cada fragmento. Después de la compresión (185), cada grupo de registros recopilado se almacena (165) en el área de almacenamiento 155 asignada para el fragmento, junto con el índice de primer nivel (170) de esos grupos de registros.

En el caso de que sea necesario extraer registros de llamadas particulares, un usuario puede definir unos criterios de búsqueda por medio de la interfaz de usuario 200. Las consultas habituales soportadas por el motor de búsqueda y recuperación 125 incluyen las siguientes:

(1) Proporcionar los registros de todas las llamadas que se originan desde un número del suscriptor llamante dado dentro de un intervalo de tiempo definido en una fecha particular;

(2) Proporcionar los registros de todas las llamadas a un número del suscriptor dado dentro de un intervalo de tiempo definido en una fecha dada;

(3) Proporcionar los registros de todas las llamadas a un número móvil de destino dado, cuando está dentro de un área geográfica dada, dentro de un intervalo de tiempo definido en una fecha dada;

(4) Proporcionar los registros de todas las llamadas realizadas por un suscriptor móvil particular, cuando está dentro de un área geográfica dada, en una fecha dada.

(5) Proporcionar los registros de todas las llamadas a un número de destino dado que duren menos de 10 segundos, iniciadas entre unas fechas dadas.

En general, las consultas relacionadas con los datos transaccionales es probable que estén relacionadas con al menos un período de tiempo definido. Por lo tanto, en respuesta a una consulta habitual, el módulo de búsqueda y recuperación 210 accede al índice de fragmentación (180) para seleccionar una o más áreas 155 de almacenamiento de datos que es probable que incluyan grupos de registros relevantes a cada período de tiempo definido. A continuación, se recupera el índice de primer nivel (170) almacenado en cada área de almacenamiento 155 seleccionada y se utiliza para identificar el grupo o los grupos de registros particulares que es probable que contengan los registros de llamadas dentro del período de tiempo definido. A continuación, se descomprimen (215) el grupo o los grupos de registros identificados y se genera un índice de segundo nivel (175) de los registros descomprimidos basándose en al menos los campos de datos necesarios para satisfacer el resto de la consulta. A continuación, se utiliza el índice de segundo nivel para identificar los registros particulares de los registros descomprimidos cuyos contenidos coinciden con los términos de la consulta y para presentarlos en la interfaz de usuario 200.

En el caso de que se generara (175) un único índice antes de la compresión de datos (185) y se almacenara (170) basándose en todos los campos de datos que es probable que estén involucrados en una consulta posterior, a continuación, no sería necesario generar de manera independiente un índice de segundo nivel en esta etapa y se podría satisfacer la consulta e identificar los registros relevantes sobre la base del índice recuperado para cada área de almacenamiento 155 seleccionada.

Los registros de transacciones identificados por el motor de búsqueda y recuperación 125 se pueden suministrar a los usuarios a través de un número de diferentes tipos de interfaz, que incluyen dispositivos de presentación visual que muestran la salida del módulo de interfaz de usuario 200, o de transferencia electrónica de datos a través de la red 115 o a través de otros tipos de red, de acuerdo con la urgencia de la necesidad y la manera en la que se utilizarán los datos.

Una aplicación preferida adicional de la presente invención es la captura y almacenamiento de registros de IP del tráfico de internet, que conllevan volúmenes de datos y tasas de generación de datos de potencialmente un orden de magnitud mayor que para registros de llamadas de telecomunicaciones.

Se pretende que las variaciones en el montaje de los módulos funcionales de la presente invención, tal como sería evidente para alguien experto en la técnica relevante, estén dentro del alcance de la presente invención tal como se reivindica ahora.

REIVINDICACIONES

- 5 1. Un método implementado por ordenador para generar y buscar un archivo de registros de transacciones, en el que cada registro de transacciones comprende valores para cada uno de una pluralidad de campos de datos, de los cuales al menos uno se refiere al tiempo, donde el método comprende los pasos de:
- (i) recibir una pluralidad de registros de transacciones;
- (ii) recopilar los registros de transacciones recibidos en una pluralidad de grupos de registros de transacciones de acuerdo con un criterio de agrupación predeterminado;
- 10 (iii) generar uno o más índices de primer nivel para la pluralidad de grupos de registros de transacciones, en el que cada uno del o de los índices de primer nivel se basa en uno de los campos de datos asociados con los registros de transacciones, y en el que el o los índices de primer nivel incluyen datos suficientes únicamente para identificar uno o más grupos de registros de transacciones;
- (iv) aplicar un algoritmo de compresión seleccionado para comprimir cada uno de la pluralidad de grupos recopilados en el paso (ii); y
- 15 (v) almacenar el o los índices de primer nivel y los grupos de registros de transacciones comprimidos, en el que dicho criterio de agrupación predeterminado se define de acuerdo con el algoritmo de compresión seleccionado y con el contenido de los registros de transacciones recibidos, de modo que se recopila al menos una cantidad predeterminada de datos para formar un grupo de registros, con el fin de lograr de ese modo un nivel predeterminado de eficacia de la compresión en el paso (iv);
- 20 (vi) recibir los criterios de búsqueda para la recuperación de los registros de transacciones dentro de los grupos comprimidos, donde los criterios de búsqueda definen un valor o rango de valores para uno o más campos de datos representados en los registros de transacciones, lo que incluye al menos un campo de datos representado en el o los índices de primer nivel;
- 25 (vii) utilizar uno del o de los índices de primer nivel almacenados generados con respecto a los grupos de registros almacenados de transacciones comprimidos para seleccionar uno o más grupos de registros de transacciones comprimidos relacionados con los criterios de búsqueda recibidos;
- (viii) descomprimir el o los grupos de registros de transacciones seleccionados;
- (ix) identificar uno o más registros de transacciones individuales, dentro del o de los grupos de registros de transacciones descomprimidos, cuyo contenido coincida con los criterios de búsqueda recibidos,
- 30 en el que el paso (ix) comprende generar uno o más índices de segundo nivel para los registros de transacciones individuales contenidos dentro de los registros de transacciones descomprimidos y utilizar el o los índices de segundo nivel para identificar y recuperar el o los registros de transacciones individuales que coinciden con los criterios de búsqueda recibidos.
- 35 2. El método de acuerdo con la reivindicación 1, en el que en el paso (ii), los criterios de agrupamiento predeterminados definen un número mínimo de registros de transacciones o un volumen mínimo de datos que deben estar contenidos dentro de un grupo de registros de transacciones que se recopilan para su compresión.
3. El método de acuerdo con la reivindicación 1, en el que en el paso (ii), los criterios de agrupamiento predeterminados definen una medida de la diversidad que se debe lograr en el contenido de los datos, cuando se recopila un grupo de registros de transacciones para su compresión.
- 40 4. El método de acuerdo con la reivindicación 1, en el que en el paso (ii), los criterios de agrupamiento predeterminados definen un rango de valores para uno o más de los tipos de datos relacionados con los registros en un grupo de registros de transacciones que se recopilan para su compresión.
- 45 5. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que en el paso (ii), los criterios de agrupamiento predeterminados comprenden además los criterios para asignar uno o más grupos de registros de transacciones recopilados a uno de una pluralidad de fragmentos de procesamiento y almacenamiento, y en el que en el paso (iii), cada uno del o de los índices de primer nivel se generan mediante, y con respecto a los registros a los que se asigna, uno diferente de dicha pluralidad de fragmentos.
6. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que el nivel de eficacia de la compresión predeterminado comprende al menos una reducción de un 50% en el almacenamiento de datos

requerido para un grupo comprimido de registros de transacciones, en comparación con el requerido para un grupo no comprimido respectivo.

5 7. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que el nivel de eficacia de la compresión predeterminado comprende al menos una reducción de un 70% en el almacenamiento de datos requerido para un grupo comprimido de registros de transacciones, en comparación con el requerido para un grupo no comprimido respectivo.

10 8. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que el nivel de eficacia de la compresión predeterminado comprende una reducción de al menos un 90% en el almacenamiento de datos requerido para un grupo de registros de transacciones comprimidos, en comparación con el requerido para un grupo no comprimido respectivo.

9. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que los registros de transacciones comprenden los registros de llamadas de telecomunicaciones.

10. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que los registros de transacciones comprenden registros financieros.

15 11. El método de acuerdo con cualquier reivindicación anterior, en el que se utiliza la misma funcionalidad de indexación para crear los índices tanto en el primer como en el segundo nivel.

12. El método de acuerdo con cualquier reivindicación anterior, en el que el paso (ix) comprende además almacenar de manera permanente el o los índices de segundo nivel.

20 13. Un producto de programa informático que comprende un portador de datos que tiene almacenados en este, o que comprende unos medios para acceder y facilitar la ejecución de, unos medios de código de programa informático, que cuando se cargan y ejecutan en un ordenador hacen que el ordenador implemente el método de acuerdo con una cualquiera de las reivindicaciones 1 a 12.

Fig. 1

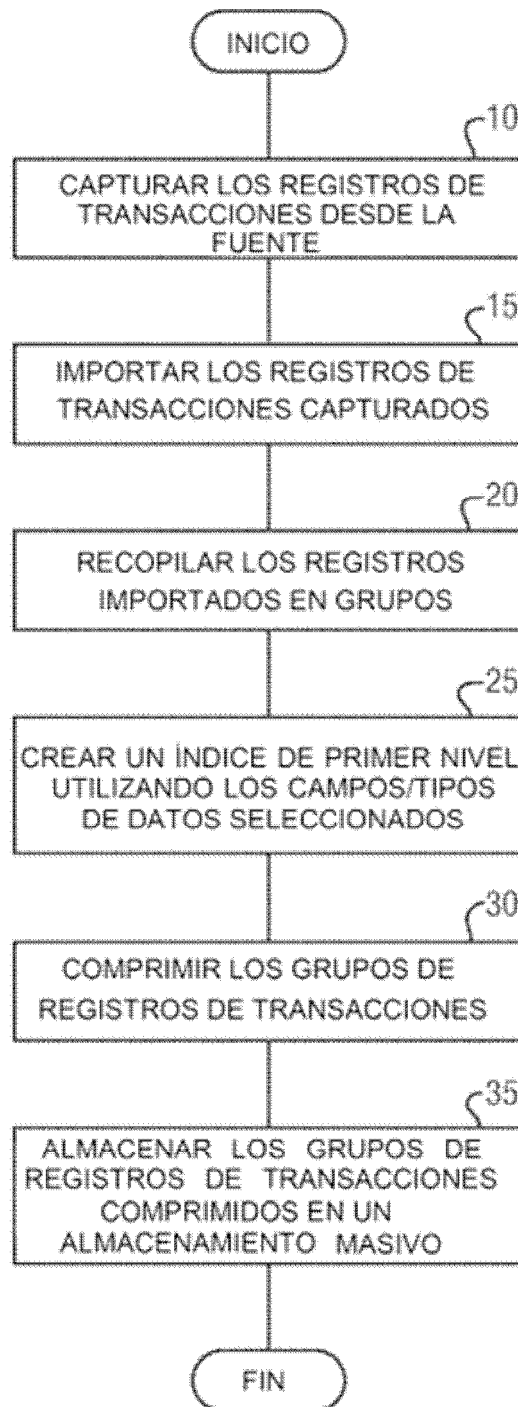


Fig. 2



Fig. 3

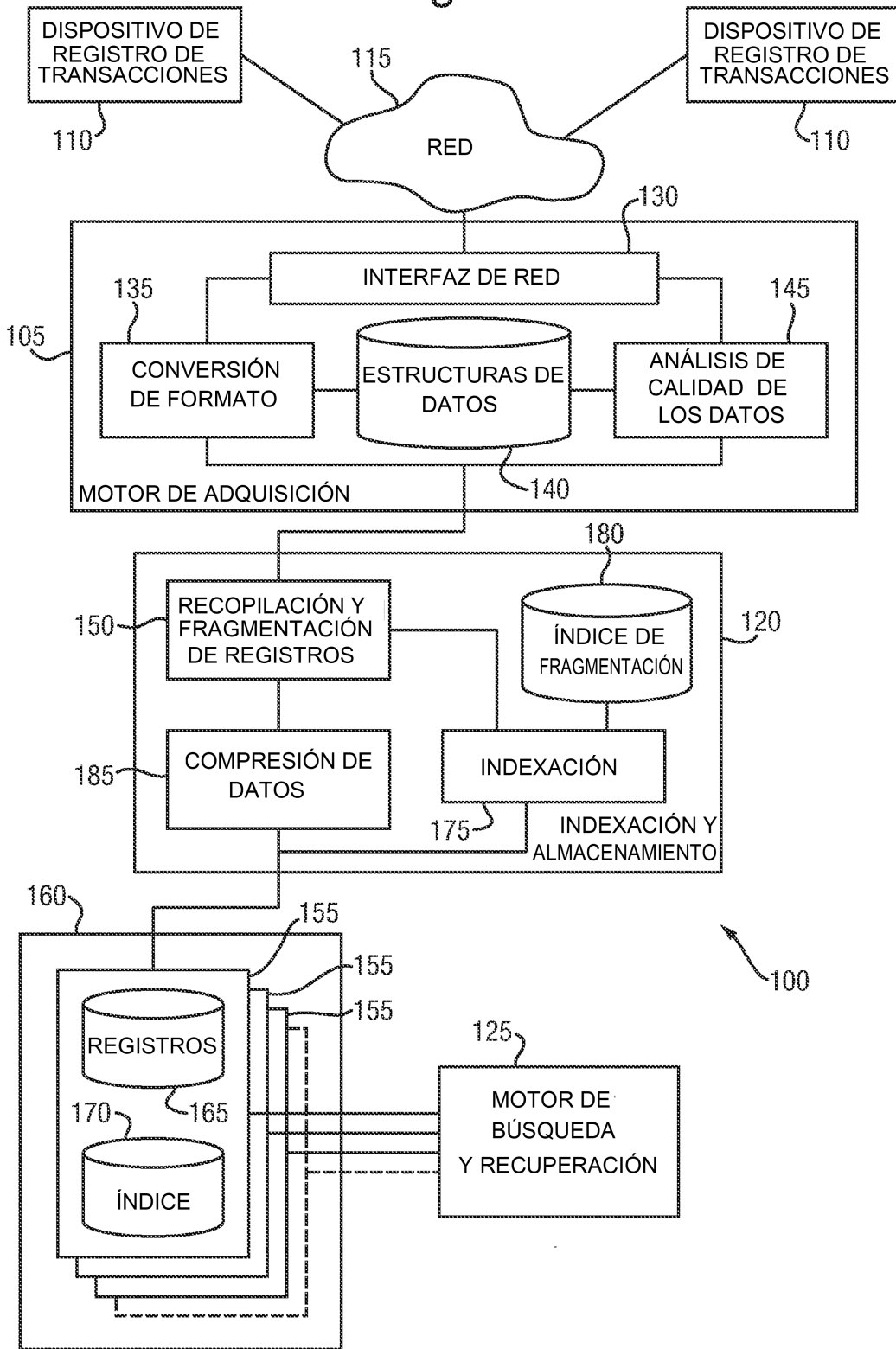


Fig. 4

