(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 699 260**

(51) Int. CI.:

G06F 17/30 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **29.08.2012 PCT/EP2012/066739**

(87) Fecha y número de publicación internacional: **07.03.2013 WO13030217**

(96) Fecha de presentación y número de la solicitud europea: **29.08.2012 E 12753954 (2)**

(97) Fecha y número de publicación de la concesión europea: **26.09.2018 EP 2751716**

(54) Título: **Método para el mantenimiento de datos**

(30) Prioridad:

02.09.2011 US 201113224415

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
08.02.2019

(73) Titular/es:

COMPUVERDE AB (100.0%)
Östra Vittusgatan 36
371 33 Karlskrona, SE

(72) Inventor/es:

BERNBO, STEFAN;
MELANDER, CHRISTIAN;
PERSSON, ROGER y
PETERSSON, GUSTAV

(74) Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

DESCRIPCIÓN

Método para el mantenimiento de datos

Campo técnico

La presente descripción se refiere a un aparato y a un método para acceder, escribir y eliminar datos en un sistema de almacenamiento de datos que comprende una pluralidad de nodos de almacenamiento de datos, los métodos pueden emplearse en un servidor y/o en un nodo de almacenamiento en el sistema de almacenamiento de datos. La divulgación se refiere además a los nodos de almacenamiento o servidores que pueden ser capaces de realizar tales métodos.

Antecedentes

Un método de este tipo se describe en la publicación de patente de Estados Unidos N.º 2005/0246393 A1. Este método se desvela para un sistema que puede usar una pluralidad de centros de almacenamiento en localizaciones geográficamente dispares. Pueden incluirse gestores de almacenamiento de objetos distribuidos para mantener la información relacionada con los datos almacenados. Un problema asociado a un sistema de este tipo es cómo lograr un mantenimiento de datos simple, aunque robusto y confiable.

Por otra parte, el documento WO 2010/046393 A2 desvela un sistema de almacenamiento distribuido que comprende una pluralidad de nodos de almacenamiento. Usando una transmisión unidifusión y multidifusión, una aplicación de servidor puede leer y escribir datos en el sistema de almacenamiento.

Sumario de la invención

La invención se define por las reivindicaciones.

En general, todos los términos utilizados en las reivindicaciones se han de interpretar de acuerdo con su significado ordinario en el campo técnico, a menos que explícitamente se defina de otro modo en el presente documento. Todas las referencias a “un/una/el/la [elemento, dispositivo, componente, medio, etapa, etc.]” deben interpretarse abiertamente como que se refieren a al menos una instancia de dicho elemento, dispositivo, componente, medio, etapa, etc., a menos que explícitamente se indique lo contrario. Las etapas de cualquier método desvelado en el presente documento no tienen que realizarse en el orden exacto desvelado, a menos que se indique explícitamente.

Breve descripción de los dibujos

Lo anterior, así como los objetivos, las características y las ventajas adicionales de los ejemplos desvelados pueden entenderse mejor a través de la siguiente descripción detallada, ilustrativa y no limitativa, haciendo referencia a los dibujos adjuntos, en los que los mismos números de referencia pueden usarse para elementos similares.

La figura 1 es una vista esquemática de un sistema de almacenamiento de ejemplo.
La figura 2 es un diagrama de bloques esquemático de ejemplo de una serie de elementos de datos almacenados en el sistema de almacenamiento.
La figura 3 es un diagrama de bloques esquemático de un identificador de elemento de datos de ejemplo.
La figura 4 es un diagrama de bloques esquemático de un método de ejemplo para recuperar datos.
Las figuras 5a-c son ilustraciones de comunicaciones de ejemplo entre diferentes entidades en el sistema de almacenamiento.
La figura 6 es un diagrama de bloques esquemático de un método de ejemplo para almacenar datos.
La figura 7 es un diagrama de bloques esquemático de un método de ejemplo para eliminar datos.

Descripción detallada

Ejemplos detallados de los métodos y sistemas desvelados pueden describirse haciendo referencia a los dibujos. La presente divulgación está relacionada con un sistema de almacenamiento de datos distribuidos que comprende una pluralidad de nodos de almacenamiento. En la figura 1, se describe una estructura de ejemplo del sistema y el contexto en el que puede usarse.

Un ordenador de usuario 1 puede acceder, por ejemplo a través de Internet 3, a una aplicación 5 que se ejecuta en un servidor 7. El contexto de usuario, como se ilustra en este caso, puede ser por lo tanto una configuración cliente-servidor. Sin embargo, debería observarse que el sistema de almacenamiento de datos a desvelar puede ser útil también en otras configuraciones, por ejemplo, usando otros métodos de comunicación.

En el caso ilustrado, dos aplicaciones 5, 9 pueden ejecutarse en el servidor 7. Por supuesto, sin embargo, cualquier número de aplicaciones puede ejecutarse en el servidor 7. Cada aplicación puede tener una API (interfaz de programación de aplicaciones) 11 que puede proporcionar una interfaz en relación con el sistema de

almacenamiento de datos distribuidos 13 y puede soportar solicitudes, normalmente solicitudes de escritura y lectura, procedentes de las aplicaciones que se ejecutan en el servidor. Los datos pueden leerse y escribirse en el sistema de almacenamiento usando los métodos descritos en detalle en la solicitud de patente de Estados Unidos N.º 13/125,524, presentada el 21 de abril de 2011, cuyo contenido se incorpora en este caso como referencia en el presente documento. Los métodos de lectura y escritura de datos, por lo tanto, pueden no elaborarse más en detalle en el presente documento. Desde el punto de vista de una aplicación, la información de lectura o escritura desde/hacia el sistema de almacenamiento de datos 13 puede parecer lo mismo que usar cualquier otro tipo de solución de almacenamiento, por ejemplo, un servidor de archivos o un disco duro.

Cada API 11 pueden comunicarse con los nodos de almacenamiento 15 en el sistema de almacenamiento de datos 13, y los nodos de almacenamiento pueden comunicarse entre sí. Como alternativa, o adicionalmente, uno o más de los nodos de almacenamiento 15 pueden incluir una API 23 para soportar solicitudes como se ha descrito anteriormente. Estas comunicaciones pueden basarse en TCP (protocolo de control de transmisión) y UDP (protocolo de datagramas de usuario). También pueden usarse otros protocolos de comunicación.

Los componentes del sistema de almacenamiento de datos distribuidos pueden ser los nodos de almacenamiento 15 y las API 11 en el servidor 7 que pueden acceder a los nodos de almacenamiento 15. La presente divulgación pueda describirse en relación con los métodos realizados en el servidor 7 y en los nodos de almacenamiento 15. Esos métodos pueden implementarse principalmente como una combinación de implementaciones de software/hardware que se ejecutan en el servidor y en los nodos de almacenamiento, respectivamente. Las operaciones del servidor y/o de los nodos de almacenamiento pueden determinar conjuntamente, en general, la operación y las propiedades del sistema de almacenamiento de datos distribuidos.

Aunque en la figura 1, el servidor 7 se ilustra como un miembro del sistema de almacenamiento 13 que está separado de los nodos de almacenamiento 15 debería observarse que el servidor 7 puede ser un nodo de almacenamiento que incluya la funcionalidad de servidor.

El nodo de almacenamiento 15 puede implementarse normalmente por un servidor de archivos que está provisto de una serie de bloques funcionales. Por lo tanto, el nodo de almacenamiento puede incluir un medio de almacenamiento 17, que, por ejemplo, puede incluir una cantidad de unidades de disco duro interiores (por ejemplo, conectadas a través de electrónica de dispositivos integrada (IDE), accesorio de tecnología avanzada en serie SATA) y/o similares) o discos duros exteriores (por ejemplo, conectados a través del bus serie universal (USB), Firewire, Bluetooth y/o similares), opcionalmente configurados como un sistema RAID (matriz redundante de disco independiente). Sin embargo, también son concebibles otros tipos de medios de almacenamiento.

Cada nodo de almacenamiento 15 puede contener una lista de nodos que incluye las direcciones IP de todos los nodos de almacenamiento en su conjunto o grupo de nodos de almacenamiento. El número de nodos de almacenamiento en un grupo puede variar de unos pocos a cientos o miles de nodos de almacenamiento.

El medio de almacenamiento 17 puede almacenar uno o más elementos de datos 19, 21 en la forma de objetos recopilados 19 o datos de carga útil en la forma de archivos de datos 21. Un objeto de recopilación 19 puede incluir un conjunto de referencias. Una referencia puede ser una referencia a uno o más archivos de datos almacenados en el sistema de almacenamiento, por ejemplo, los archivos de datos 21. Una referencia también puede ser una referencia a otro objeto de recopilación 19 almacenado en el sistema de almacenamiento. Una referencia puede incluir un puntero (por ejemplo, una dirección de memoria) a una localización de almacenamiento de un nodo de almacenamiento 15. Una referencia puede incluir un identificador del objeto de recopilación o del archivo de datos al que se hace referencia.

Como se desvelará en más detalle a continuación, el objeto de recopilación 19 puede usarse para implementar una capa estructurada en el sistema de almacenamiento. Los archivos de datos 21 a los que se hace referencia en el objeto de recopilación 19 pueden representar una implementación de archivos de datos almacenados de este tipo en la estructura. Los objetos recopilados adicionales 19 a los que se hace referencia en el objeto de recopilación 19 pueden representar una implementación de subdirectorios almacenados de este tipo en el directorio.

Un objeto de recopilación 19 puede implementarse como un objeto de datos que tiene un formato predeterminado. El objeto de datos puede ser un archivo especial en el sistema de archivos del medio de almacenamiento 17 en el sentido de que puede ser un archivo binario para interpretarse por la API. En un ejemplo, el objeto de datos puede ser un archivo de datos estándar en el sistema de archivos del medio de almacenamiento 17; el objeto de datos puede ser, por ejemplo, un archivo de texto sin cifrar que indica los objetos recopilados referenciados 19 y/o los archivos de datos 21. Un objeto de datos puede ser legible usando las mismas rutinas del sistema de archivos que los archivos de datos 21.

La figura 2 ilustra esquemáticamente un objeto de recopilación 19a de acuerdo con un ejemplo. El objeto de recopilación 19a puede tener un identificador de objetos recopilados asociado 20a. El identificador 20a puede ser, por ejemplo, un identificador único universal (UUID). El identificador de objetos recopilados 20a puede incluirse en una cabecera del objeto de recopilación 19a. Sin embargo, el identificador de objetos recopilados 20a puede

almacenarse en un registro mantenido en el nodo de almacenamiento 15, por ejemplo, en lugar de estar incluido en el objeto de recopilación 19a. En un ejemplo, el UUID y/o el registro mantenido en el nodo de almacenamiento 15 pueden asociar el objeto de recopilación 19a al identificador de objetos recopilados 20a, por ejemplo, señalando la dirección de memoria donde se encuentra el objeto de recopilación 19a. Por lo tanto, el objeto de recopilación 19a puede formar un primer elemento de datos que se identifica mediante una primera clave única.

El objeto de recopilación 19a puede incluir un campo 22a con un identificador 20b de otro objeto de recopilación 19b, por ejemplo en la forma de una cadena. El objeto de recopilación 19a puede incluir una referencia al objeto de recopilación 19b. El objeto de recopilación 19b puede almacenarse en el mismo nodo de almacenamiento que el objeto de recopilación 19a o en otro nodo de almacenamiento distinto del objeto de recopilación 19a. El sistema de almacenamiento puede usar el identificador 20b en el campo 22a para localizar y acceder al objeto de recopilación 19b. Por lo tanto, el objeto de recopilación 19b puede formar un segundo elemento de datos que se identifica mediante una segunda clave única.

En un ejemplo con el fin de implementar sistemas de almacenamiento de gran tamaño que abarquen múltiples redes, los identificadores de elementos de datos 20a-d pueden incluir dos elementos de datos. Haciendo referencia a la figura 3, el primer elemento de datos 30 puede ser un ID de clúster 31 que puede identificar el clúster donde se localiza el elemento de datos (el objeto de recopilación 19a-c o el archivo de datos 21a). La dirección del clúster puede ser una dirección de multidifusión 32. La API puede usar la dirección de multidifusión 32 para enviar una solicitud de un elemento de datos a un clúster específico. El segundo elemento de datos 33 puede ser un ID de elemento de datos 34 formado por un número único 35 que identifica el elemento de datos 19a-d dentro del clúster. El número único 35 puede ser un número con una longitud definida, por ejemplo, 128 bits, o la longitud puede variar. El número único 35 puede incluir un gran número de bits, lo que permite que un gran número de elementos de datos se identifiquen de manera única dentro del clúster. Mediante esta disposición, un elemento de recopilación en un clúster puede hacer referencia a otro elemento de recopilación o archivo de datos en otro clúster. En otras palabras, la clave única primera y segunda pueden incluir una dirección de grupo que señala a un subconjunto de los nodos de almacenamiento dentro del sistema, y un identificador de elemento de datos que identifica un elemento de datos dentro del subconjunto de nodos de almacenamiento.

Haciendo referencia de nuevo a las figuras 1 y 2, el servidor 7 puede, por ejemplo, incluir un registro que indica un nodo de almacenamiento 15 que almacena el objeto de recopilación (por ejemplo, el objeto de recopilación 19a) asociado a un identificador específico (por ejemplo, el identificador 20a). En otro ejemplo, el objeto de recopilación 19a puede localizarse usando el método de lectura descrito en la solicitud de patente de Estados Unidos N.º 13/125.524. Brevemente, de acuerdo con este método de lectura, el servidor 7 o un nodo de almacenamiento 15 pueden enviar un mensaje de multidifusión a la pluralidad de nodos de almacenamiento 15. El mensaje de multidifusión puede incluir el identificador 20a del objeto de recopilación deseado 19a. Cada nodo de almacenamiento 15, en respuesta a la recepción del mensaje de multidifusión, puede escanear su medio de almacenamiento 17 en busca de un objeto de recopilación que tenga dicho identificador. Si lo encuentra, el nodo de almacenamiento 15 puede responder e indicar que almacena el objeto buscado al originador del mensaje de multidifusión. A continuación, puede accederse al objeto de recopilación 19a por medio de una solicitud de unidifusión enviada a un nodo de almacenamiento de respuesta 15 que almacena el objeto de recopilación 19a.

De acuerdo con la presente comunicación de multidifusión de ejemplo puede usarse para comunicarse simultáneamente con una pluralidad de nodos de almacenamiento. Por multidifusión o multidifusión IP se entiende en este caso una comunicación punto a multipunto que puede realizarse enviando un mensaje a una dirección IP que puede estar reservada para aplicaciones de multidifusión. Por ejemplo, un mensaje, por ejemplo una solicitud, puede enviarse a dicha dirección IP (por ejemplo, 244.0.0.1), y diversos servidores receptores pueden registrarse como suscriptores a esa dirección IP. Cada uno de los servidores receptores puede tener su propia dirección IP. Cuando un switch en la red recibe el mensaje dirigido a 244.0.0.1, el switch puede reenviar el mensaje a las direcciones IP de cada servidor registrado como suscriptor.

En principio, un único servidor puede estar registrado como suscriptor a una dirección de multidifusión, en cuyo caso puede lograrse una comunicación punto a punto. Sin embargo, en el contexto de esta divulgación, una comunicación de este tipo puede, no obstante, considerarse como una comunicación de multidifusión ya que se emplea un esquema de multidifusión.

De acuerdo con la presente comunicación de multidifusión de ejemplo puede referirse a una comunicación con un único receptor. Una tercera parte de la red puede iniciar una comunicación de unidifusión y puede dirigirse a un único receptor específico.

Además del objeto de recopilación 19a, el objeto de recopilación 19b puede incluir un campo 22b con un identificador 20c de un tercer objeto de recopilación 19c. El objeto de recopilación 19c puede incluir un campo 22c con un identificador 20d de un archivo de datos 21a. En otras palabras, uno cualquiera de los objetos recopilados 19a-c (o, por ejemplo, cada uno de los objetos recopilados 19a-c) puede representar un segundo elemento de datos que incluye una referencia al tercer elemento de datos, y el archivo de datos 21a puede representar un segundo elemento de datos que incluye datos de carga útil, por ejemplo, una imagen.

Al designar el objeto de recopilación 19a como un objeto de recopilación raíz, el objeto de recopilación 19a puede representar un directorio raíz 19a del sistema de almacenamiento. Análogamente, el objeto de recopilación 19b puede representar un subdirectorio 19b del directorio raíz 19a. El objeto de recopilación 19c puede representar un subdirectorio del subdirectorio 19b. El archivo de datos 21a puede representar un archivo de datos almacenado en el subdirectorio 19c. Los objetos recopilados 19a-c pueden definir de este modo una estructura de almacenamiento jerárquica. La estructura puede denominarse como un árbol de directorio.

Haciendo referencia a las figuras 4 y 5a-c, puede desvelarse un método para analizar una estructura de directorios con el fin de acceder a un archivo 19, 21 almacenado en un nodo de almacenamiento 15.

El punto de partida de la estructura de directorios puede ser una clave raíz predefinida. Por ejemplo, cualquiera de los nodos de almacenamiento 15 puede incluir una clave raíz. Esta clave puede almacenarse fuera del clúster de almacenamiento y puede usarse para identificar el primer elemento de datos (por ejemplo, el objeto de recopilación 19a) en la estructura de directorios. Un clúster de almacenamiento puede tener múltiples claves raíz que permiten al usuario tener múltiples estructuras de directorio individuales almacenadas dentro del mismo clúster de almacenamiento. Las estructuras de directorio pueden abarcar diversos clústeres de almacenamiento. La clave raíz puede almacenarse junto con la información exterior que describe la estructura de directorios almacenada dentro del clúster.

En el bloque 40, el servidor 7 puede recibir la clave raíz, que puede identificar el primer elemento de datos 19, 21 y puede pasar el identificador único para identificar el archivo dentro del sistema de almacenamiento a la API 11. En un ejemplo, la API 23 puede implementarse en un nodo de almacenamiento 15, en el que la clave raíz puede recibirse en el nodo de almacenamiento 15 en lugar de en el servidor 7.

En el bloque 41, la API 11 en el servidor 7 puede hacer multidifusión de una solicitud del elemento de datos (por ejemplo, el objeto de recopilación 19a) identificado por la clave raíz a los nodos de almacenamiento 15a-e en el sistema de almacenamiento, o a un subconjunto de los nodos Por ejemplo, el mensaje de multidifusión puede enviarse a un clúster específico, por ejemplo, usando la configuración de identificador de elemento de datos desvelada en relación con la figura 3. De acuerdo con un ejemplo, el elemento de datos (por ejemplo, el objeto de recopilación 19a) identificado por la clave raíz puede ser un elemento de datos especial en el sentido de que puede incluir metadatos adicionales que pueden usarse por el sistema. Los ejemplos de dichos datos pueden ser información sobre los permisos de acceso a los elementos en la estructura de directorios, información sobre dónde almacenar ciertos elementos de datos (por ejemplo, en un nodo de almacenamiento con acceso rápido, tal como una unidad de estado sólido (SSD)), y similares.

En el bloque 42, el nodo de almacenamiento 15a-e, en respuesta a recibir el mensaje de multidifusión, puede escanear sus respectivos medios de almacenamiento 17 en un intento de localizar el elemento de datos identificado por el ID de elemento de datos 34 en la clave raíz.

Para fines de ilustración, puede suponerse en este ejemplo que los nodos 15b y 15e localizan el elemento de datos identificado por el ID de elemento de datos 34. En el bloque 43, los nodos 15b, 15e que encuentran el elemento de datos pueden responder con información sobre qué otros nodos 15b, 15d, 15e pueden contener el elemento de datos y la carga de ejecución actual (por ejemplo, cuán ocupados están los nodos, cuántas solicitudes recibieron los nodos, cuánto espacio libre hay en el nodo, etc.) en el nodo 15b, 15e. El elemento de datos solicitado puede almacenarse en una pluralidad de nodos de almacenamiento 15b, 15d, 15e, en los que la API puede recopilar la información recibida desde los nodos 15b, 15d, 15e y puede esperar hasta que se hayan recibido las respuestas de más del 50 % de los nodos de almacenamiento enumerados 15b, 15d, 15e que contienen el elemento de datos antes de que pueda tomar una decisión sobre cuál seleccionar para la recuperación del elemento de datos. La decisión puede basarse en qué nodo tiene la carga de ejecución más baja.

En el bloque 44, la API 11 puede enviar una solicitud de unidifusión del archivo específico al nodo de almacenamiento elegido. En este ejemplo, para fines de ilustración, puede asumirse que se elige el nodo de almacenamiento 15b. La API 11 puede recuperar el elemento de datos del nodo de almacenamiento 15b. La API 11 puede mantener una lista de todos los nodos de almacenamiento 15b, 15d, 15e que almacenan copias del archivo en el caso de un fallo de lectura o comunicación con el nodo seleccionado 15b. Si se produce un error, la API 11 puede seleccionar de manera transparente el siguiente mejor nodo de la lista y continuar la operación de lectura.

En el bloque 45 la API puede interpretar el contenido del elemento de datos recuperado. Si la estructura de directorios comprende niveles adicionales, el elemento de datos recuperado puede ser un objeto de recopilación 19b. Si es así, la API 11 puede leer el campo 22b que puede incluir un identificador 20b que se refiere a otro objeto de recopilación 19c en la estructura de directorios. Por ejemplo, la API puede recuperar la clave única, es decir, el identificador 20b, que identifica el segundo elemento de datos, por ejemplo, el objeto de recopilación 19b, del primer elemento de datos recibido, por ejemplo, el objeto de recopilación 19a. A continuación, el proceso puede regresar al bloque 41 y puede continuar analizando la estructura de directorios. Por lo tanto, tanto la primera como la segunda solicitud pueden enviarse desde una interfaz de programación de aplicaciones, API. El proceso puede continuar hasta que el último objeto en la estructura de directorios se haya identificado y recuperado, por ejemplo, el archivo

de datos 21a, con lo cual el proceso puede terminar en 46. En otro ejemplo, la API 11 puede enviar una solicitud de actualización al objeto identificado, por ejemplo, una orden para alterar o concatenar datos en el elemento de datos correspondiente al objeto en la estructura de directorios.

A modo de ejemplo, puede ser que el archivo de datos 21 se encuentre en la raíz de la estructura de directorios. En tal caso, el proceso puede repetirse una sola vez, ya que el primer objeto de recopilación recuperado 19a puede contener una referencia al archivo de datos 21a. Se enfatiza que el objeto de recopilación recuperado además de incluir la referencia al archivo de datos 21a también puede incluir referencias a otros elementos de datos, tal como el objeto de recopilación 19b.

Por lo tanto, de acuerdo con lo anterior, puede implementarse un método en un sistema de almacenamiento de datos que incluya los nodos de almacenamiento de datos interconectados por medio de una red de comunicaciones para acceder a archivos. El método puede incluir enviar una solicitud de un primer elemento de datos 19, 21 (por ejemplo, el objeto de recopilación 19a) a una pluralidad de nodos de almacenamiento 15a-e. El primer elemento de datos puede incluir una referencia a un segundo elemento de datos (por ejemplo, el archivo de datos 21a o el objeto de recopilación 19b), almacenado en el sistema de almacenamiento. El método puede incluir recibir el primer elemento de datos desde al menos un nodo de almacenamiento 15b, y enviar una solicitud del segundo elemento de datos a la pluralidad de nodos de almacenamiento 15a-e basándose en la referencia incluida en el primer elemento de datos.

Como un ejemplo ilustrativo, haciendo referencia a la figura 2, la API puede leer de manera recursiva e interpretar la envolvente referenciada para resolver una ruta en una estructura de directorios. Por ejemplo, la API puede identificar una clave no estructurada que representa un archivo en la ruta estructurada. Por ejemplo, un usuario que accede al sistema de almacenamiento puede querer resolver la ruta:
“/Documents/Sample_Pictures/Blue_Hills.jpg”.

En la figura 2, el objeto de recopilación 19a pueden representar la clave raíz “/” (identificado por la clave única 20a) y el identificador 22a puede incluir una referencia al objeto de recopilación 19b que representa la carpeta “Documentos/” (identificada por la clave única 20b). El identificador 22b en el objeto de recopilación 19b puede incluir una referencia al objeto de recopilación 19c que representa la carpeta “Sample_Pictures/”. Finalmente, el identificador 22c en el objeto de recopilación 19c puede incluir una referencia al archivo de datos 21a que comprende los datos de carga útil para el archivo “Blue_Hills.jpg”. Por lo tanto, al leer recursivamente las referencias en los objetos de recopilación, puede crearse una estructura de archivos virtual en un sistema de almacenamiento no estructurado.

Haciendo referencia a las figuras 6 y 5a-c, se desvela un método para analizar una estructura de directorios con el fin de crear un archivo 19, 21 en un nodo de almacenamiento 15.

De manera similar al sistema desvelado en la figura 4, el punto de partida de la estructura de directorios es una clave raíz predefinida. La clave raíz puede ser una clave arbitraria y puede haber muchas claves raíz en todo el sistema. Esta clave puede almacenarse fuera del clúster de almacenamiento y puede usarse para identificar el primer elemento de datos (por ejemplo, el objeto de recopilación 19a) en la estructura de directorios.

En el bloque 60, el servidor 7 puede recibir la clave raíz, y puede pasar el identificador único para identificar el archivo dentro del sistema de almacenamiento a la API.

En el bloque 61, la API 11 puede resolver la ruta al elemento de datos deseado de acuerdo con el método anterior.

En el bloque 62, la API 11 en el servidor 7 puede hacer multidifusión de una solicitud para almacenar el elemento de datos (por ejemplo, el objeto de recopilación 19c) que incluye el identificador para todos los nodos de almacenamiento 15a-e en el sistema de almacenamiento, o para un subconjunto de los nodos, por ejemplo, dentro de un clúster específico, por ejemplo, usando la configuración de identificador de elemento de datos desvelada en relación con la figura 3.

En el bloque 63, los nodos de almacenamiento 15a-e, en respuesta a recibir el mensaje de multidifusión, pueden verificar que el ID de elemento de datos 34 no está ya en uso.

En el bloque 64, un nodo de almacenamiento 15a-e que falla al encontrar un archivo existente con ese identificador específico puede responder con un reconocimiento que puede indicar: espacio de almacenamiento libre en el nodo de almacenamiento, una indicación de la edad del hardware que el nodo de almacenamiento está ejecutando, la carga actual de la CPU y/o la posición geográfica del nodo de almacenamiento 15a-e en la forma de latitud, longitud y altitud o similares.

En el bloque 65, la API 11 puede seleccionar tres nodos de almacenamiento (por ejemplo, los nodos de almacenamiento 15a, 15b y 15e) basándose en los datos devueltos desde los nodos de almacenamiento que respondieron a la solicitud de multidifusión. Cuando se han seleccionado los tres nodos más adecuados, la API 11

puede enviar una solicitud a los tres nodos simultáneamente para almacenar el elemento de datos. Si se produce un error durante la transferencia del elemento de datos a uno de los nodos seleccionados 15a, 15b, 15e, la operación continúa, por ejemplo, mientras más del 50 % de los nodos seleccionados estén operativos.

En el bloque 66, el campo de identificador 22b en el elemento de datos de un nivel superior en la estructura de directorios (por ejemplo, el primer elemento de datos - el objeto de recopilación 19b) puede actualizarse con una referencia al elemento de datos almacenado (por ejemplo, el objeto de recopilación 19c), ya sea recuperando el primer elemento de datos de acuerdo con el método de lectura de acuerdo con lo anterior o accediendo directamente al primer elemento de datos, por ejemplo, si el servidor ha guardado en caché el identificador del primer elemento de datos.

Con el fin de aumentar la integridad de los datos en el sistema, el método anterior puede complementarse con la acción de, antes de almacenar el elemento de datos, recuperar el primer elemento de datos en el caso de que la comunicación con todos los nodos de almacenamiento se pierda después de que el elemento de datos se haya almacenado pero antes de que se actualice el primer elemento de datos. Mediante este procedimiento, la API puede reanudar el proceso de actualización una vez que se reanude la comunicación con los nodos de almacenamiento.

Por lo tanto, de acuerdo con lo anterior, puede implementarse un método en diversos dispositivos dentro de un sistema de almacenamiento de datos, incluyendo los nodos de almacenamiento de datos interconectados por medio de una red de comunicaciones. El método puede incluir almacenar un primer elemento de datos en al menos un nodo de almacenamiento y actualizar un segundo elemento de datos, almacenado en al menos un nodo de almacenamiento, agregando una referencia al primer elemento de datos en el segundo elemento de datos.

Haciendo referencia a las figuras 7 y 5a-c, se desvela un método para analizar una estructura de directorios con el fin de eliminar un archivo 19, 21 en un nodo de almacenamiento 15.

En similitud con la divulgación en relación con la figura 4, el punto de partida de la estructura de directorios puede ser una clave raíz predefinida pero arbitraria. Esta clave puede almacenarse fuera del clúster de almacenamiento y puede usarse para identificar el primer elemento de datos (por ejemplo, el objeto de recopilación 19a) en la estructura de directorios.

En el bloque 70, el servidor 7 puede recibir la clave raíz y puede pasar el identificador único para identificar el archivo en el sistema de almacenamiento a la API.

En el bloque 71, la API 11 puede resolver la ruta al elemento de datos deseado de acuerdo con el método anterior.

En el bloque 72, la API 11 en el servidor 7 puede hacer multidifusión de una solicitud con respecto a la localización del elemento de datos (por ejemplo, el objeto de recopilación 19c) que incluye el identificador de nodos de almacenamiento 15a-e en el sistema de almacenamiento, o a un subconjunto de los nodos, por ejemplo, dentro de un clúster específico, por ejemplo, usando la configuración de identificador de elemento de datos desvelada en relación con la figura 3.

En el bloque 73, los nodos de almacenamiento 15a-e, en respuesta a recibir el mensaje de multidifusión, pueden escanear su medio de almacenamiento respectivo 17 para localizar el elemento de datos identificado por el ID de elemento de datos 34.

En el bloque 74, los nodos que localizan el elemento de datos pueden responder con la información con respecto a otros nodos que puede almacenar el elemento de datos y la carga de ejecución actual en el nodo. El elemento de datos solicitado puede almacenarse en una pluralidad de nodos de almacenamiento. La API puede recopilar la información recibida desde los nodos y puede esperar hasta que haya recibido las respuestas de más del 50 % de los nodos de almacenamiento enumerados que contienen el elemento de datos antes de tomar una decisión sobre qué nodos seleccionar para la eliminación del elemento de datos.

En el bloque 75, la API 11 puede enviar una solicitud de unidifusión para eliminar el archivo específico (por ejemplo, el objeto de recopilación 19c) a los nodos de almacenamiento elegidos.

En el bloque 76, el campo de identificador 22b en un elemento de datos de un nivel superior en la estructura de directorios (por ejemplo, el objeto de recopilación 19b) puede actualizarse eliminando la referencia al elemento de datos eliminado (por ejemplo, el objeto de recopilación 19c). La actualización puede producirse al recuperar el primer elemento de datos de acuerdo con el método de lectura descrito anteriormente y/o al acceder directamente al primer elemento de datos, por ejemplo, si el servidor almacenó en caché el identificador del primer elemento de datos. En el caso donde el elemento de datos que se va a eliminar se localiza a varios niveles por debajo en la estructura de directorios, la operación de eliminación puede expresarse como el método desvelado en relación con la figura 4 con la adición de i) eliminar el primer elemento de datos, y ii) actualizar el segundo elemento de datos eliminando la referencia al primer elemento de datos.

Por lo tanto, puede implementarse un método de eliminación de datos en un sistema de almacenamiento de datos que incluye unos nodos de almacenamiento de datos interconectados por medio de una red de comunicaciones. El método puede incluir eliminar un primer elemento de datos almacenado en al menos un nodo de almacenamiento. El método también puede incluir actualizar un segundo elemento de datos, almacenado en al menos un nodo de almacenamiento, eliminando una referencia al primer elemento de datos en el segundo elemento de datos.

Los objetos de recopilación 19 pueden manejarse y mantenerse de manera similar a los archivos de datos. Esto puede permitir que los datos se almacenen en una estructura de almacenamiento plana, por ejemplo, sin subdirectorios o dentro de un único directorio. Puede crearse una estructura de almacenamiento jerárquica virtual agregando los objetos recopilados 19 que incluyen referencias a otros objetos recopilados 19 y/o a los archivos de datos 21. Incluso permite que los mismos datos se organicen en varias estructuras de almacenamiento jerárquicas virtuales diferentes usando diferentes conjuntos de objetos recopilados 19.

Por razones de seguridad de datos, parte o toda la información almacenada en el sistema de almacenamiento (por ejemplo, los objetos recopilados 19 y/o los archivos de datos 21) pueden almacenarse de manera redundante en el sistema de almacenamiento. Los objetos recopilados 19a-c y el archivo de datos 21a pueden almacenarse en dos o más nodos de almacenamiento 15. Cada instancia de un objeto de recopilación o archivo de datos puede estar asociada al mismo identificador. En tal caso, el método de lectura descrito anteriormente puede resultar en una respuesta de cada nodo de almacenamiento que almacena el objeto de recopilación. Por lo tanto, un objeto de recopilación almacenado de manera redundante puede recuperarse de uno o de todos los nodos de almacenamiento que almacenan el objeto de recopilación.

REIVINDICACIONES

1. Un método para mantener los datos en un sistema de almacenamiento de datos, incluyendo el sistema de almacenamiento de datos unos nodos de almacenamiento de datos (15) interconectados por medio de una red de comunicaciones, estando el método **caracterizado por**:

enviar (41) una solicitud por multidifusión de un primer elemento de datos a al menos uno de la pluralidad de nodos de almacenamiento, incluyendo dicho primer elemento de datos una referencia a un segundo elemento de datos almacenado en el sistema de almacenamiento, comprendiendo la referencia una segunda clave única asociada al segundo elemento de datos;
recibir (43) el primer elemento de datos desde al menos un nodo de almacenamiento de la pluralidad de nodos de almacenamiento; y
enviar (41) una solicitud por multidifusión del segundo elemento de datos a un subconjunto de la pluralidad de nodos de almacenamiento, en donde el subconjunto se determina basándose en la referencia incluida en el primer elemento de datos,
en donde el primer elemento de datos está asociado a una primera clave única, y comprendiendo cada una de las claves únicas primera y segunda una dirección de clúster y un identificador de elemento de datos, en donde la dirección de clúster identifica un subconjunto de dichos nodos de almacenamiento dentro del sistema, el identificador de elemento de datos, que es un identificador único universal, UUID, identifica el elemento de datos asociado almacenado en el subconjunto de nodos de almacenamiento, y el segundo elemento de datos se identifica basándose en la segunda clave única del primer elemento de datos recibido.

2. El método de acuerdo con la reivindicación 1, que comprende además enviar las solicitudes primera y segunda desde una interfaz de programación de aplicaciones, API.

3. El método de acuerdo con la reivindicación 2, en el que la API se implementa en un servidor en comunicación con el sistema de almacenamiento.

4. El método de acuerdo con la reivindicación 2, en el que la API se implementa en un nodo de almacenamiento dentro del sistema de almacenamiento.

5. El método de acuerdo con la reivindicación 2, en el que el primer elemento de datos se recibe en la API y la primera clave única se identifica en la API.

6. El método de acuerdo con la reivindicación 1, en el que el segundo elemento de datos incluye una referencia a un tercer elemento de datos.

7. El método de acuerdo con la reivindicación 1, en el que el segundo elemento de datos incluye datos de carga útil.

8. El método de acuerdo con la reivindicación 7, en el que los datos de carga útil son un archivo de imagen.

9. Un nodo de almacenamiento de datos en un sistema de almacenamiento de datos que incluye otros nodos de almacenamiento de datos interconectados a través de una red de comunicaciones, estando el nodo de almacenamiento de datos **caracterizado por**:

una interfaz de comunicaciones configurada para comunicarse a través de la red de comunicaciones; y
una interfaz de comunicaciones configurada para comunicarse a través de la red de comunicaciones;
una interfaz de programación de aplicaciones, API, (23) configurada para:

enviar una solicitud por multidifusión de un primer elemento de datos a una pluralidad de nodos de almacenamiento a través de la interfaz de comunicaciones,
recibir el primer elemento de datos desde al menos un nodo de almacenamiento de la pluralidad de nodos de almacenamiento a través de la interfaz de comunicaciones, incluyendo dicho primer elemento de datos una referencia a un segundo elemento de datos almacenado en el sistema de almacenamiento, comprendiendo la referencia una segunda clave única asociada al segundo elemento de datos,
determinar un subconjunto de la pluralidad de nodos de almacenamiento para enviar una solicitud del segundo elemento de datos basándose en la referencia incluida en el primer elemento de datos, y
enviar la solicitud por multidifusión del segundo elemento de datos al subconjunto de la pluralidad de nodos de almacenamiento,

en donde el primer elemento de datos está asociado a una primera clave única, comprendiendo cada una de las claves únicas primera y segunda una dirección de clúster y un identificador de elemento de datos, en donde la dirección de clúster identifica un subconjunto de dichos nodos de almacenamiento dentro del sistema, el identificador de elemento de datos, que es un identificador único universal, UUID, identifica el elemento de datos asociado almacenado en el subconjunto de nodos de almacenamiento, y el segundo elemento de datos se identifica basándose en la segunda clave única del primer elemento de datos recibido.

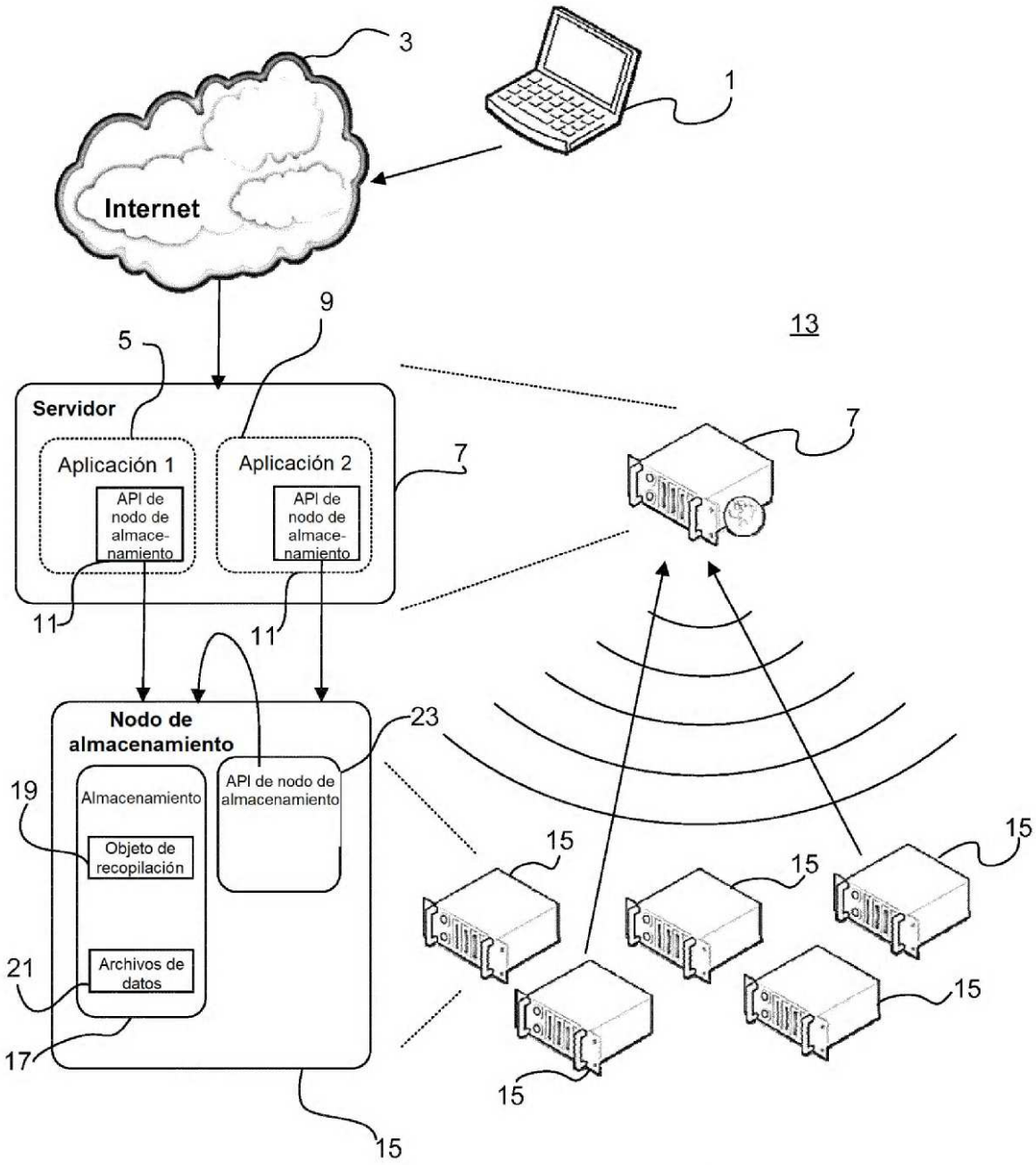
3
1
Internet
9
5
13
Servidor
Aplicación 1
API de nodo de almace-namiento
Aplicación 2
API de nodo de almace-namiento
7
7
11
11
Nodo de almacenamiento
23
19
Almacenamiento
Objeto de recopilación
API de nodo de almacenamiento
21
Archivos de datos
17
15
15
15
15
15
15

Fig 1

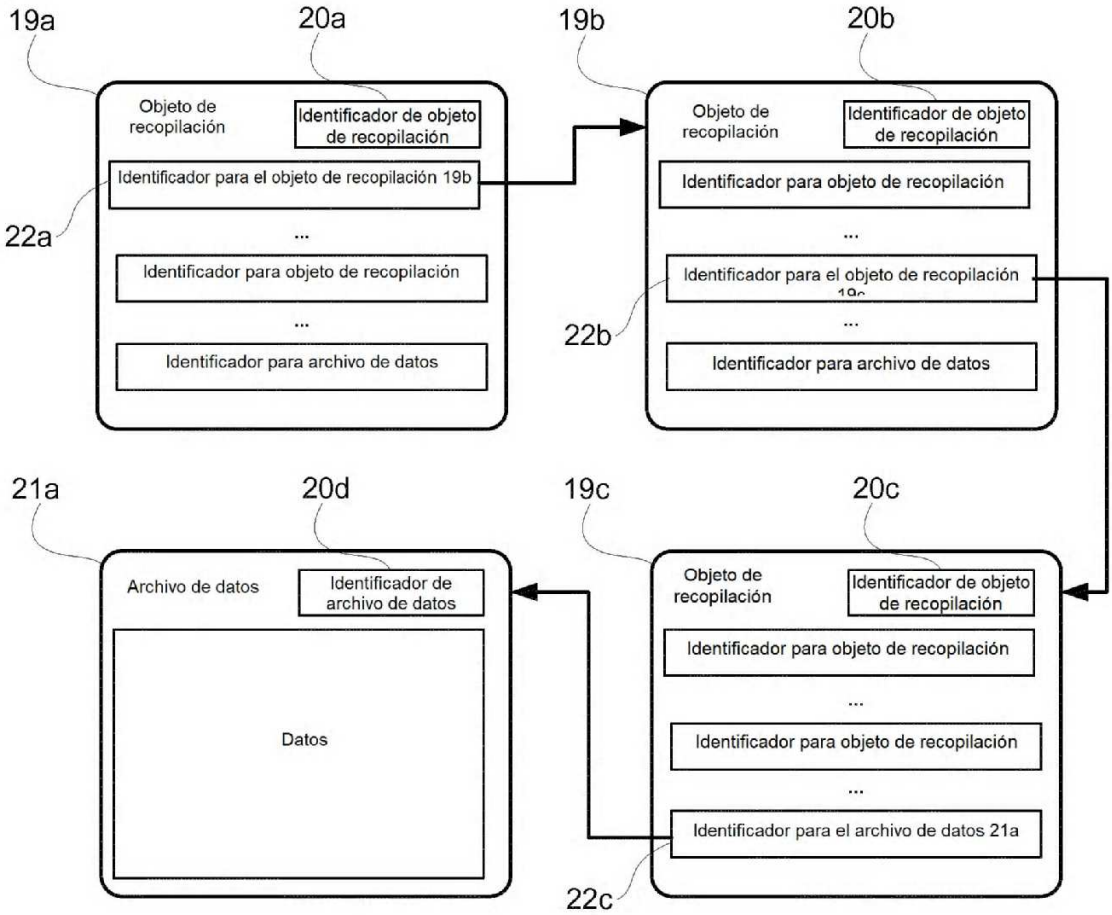
19a
20a
19b
20b
Objeto de recopilación
Identificador de objeto de recopilación
Identificador para el objeto de recopilación 19b
22a
...
Identificador para objeto de recopilación
...
Identificador para archivo de datos
Objeto de recopilación
Identificador de objeto de recopilación
Identificador para objeto de recopilación
...
Identificador para el objeto de recopilación
22b
...
Identificador para archivo de datos
21a
20d
19c
20c
Archivo de datos
Identificador de archivo de datos
Datos
Objeto de recopilación
Identificador de objeto de recopilación
Identificador para objeto de recopilación
...
Identificador para objeto de recopilación
...
Identificador para el archivo de datos 21a
22c

Fig 2

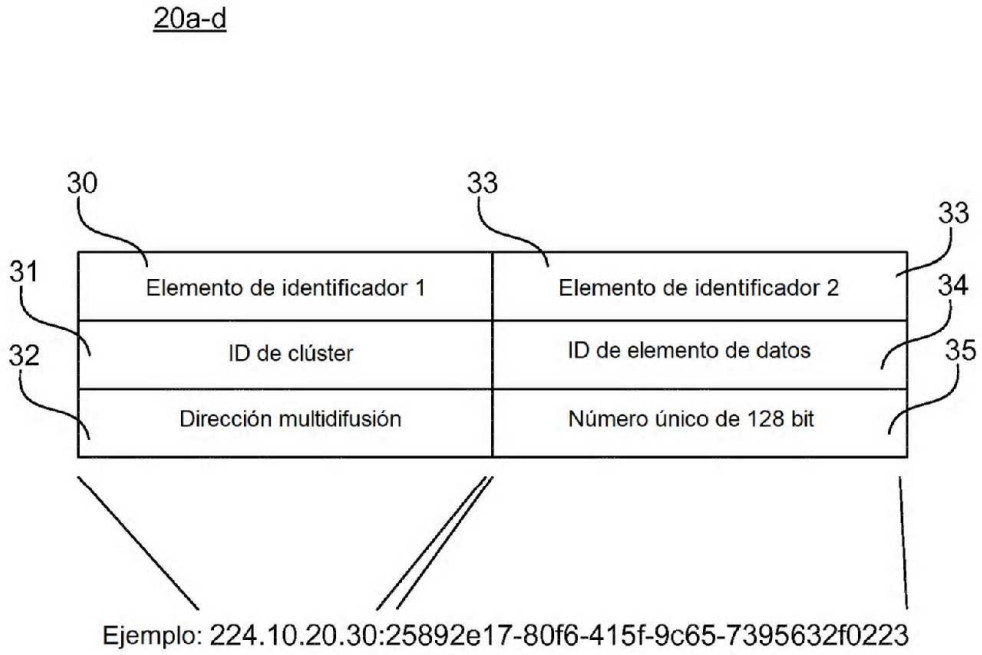
20a-d
30
33
33
31
Elemento de identificador 1
Elemento de identificador 2
34
32
ID de clúster
ID de elemento de datos
35
Dirección multidifusión
Número único de 128 bit
Ejemplo: 224.10.20.30:25892e17-80f6-415f-9c65-7395632f0223

Fig 3

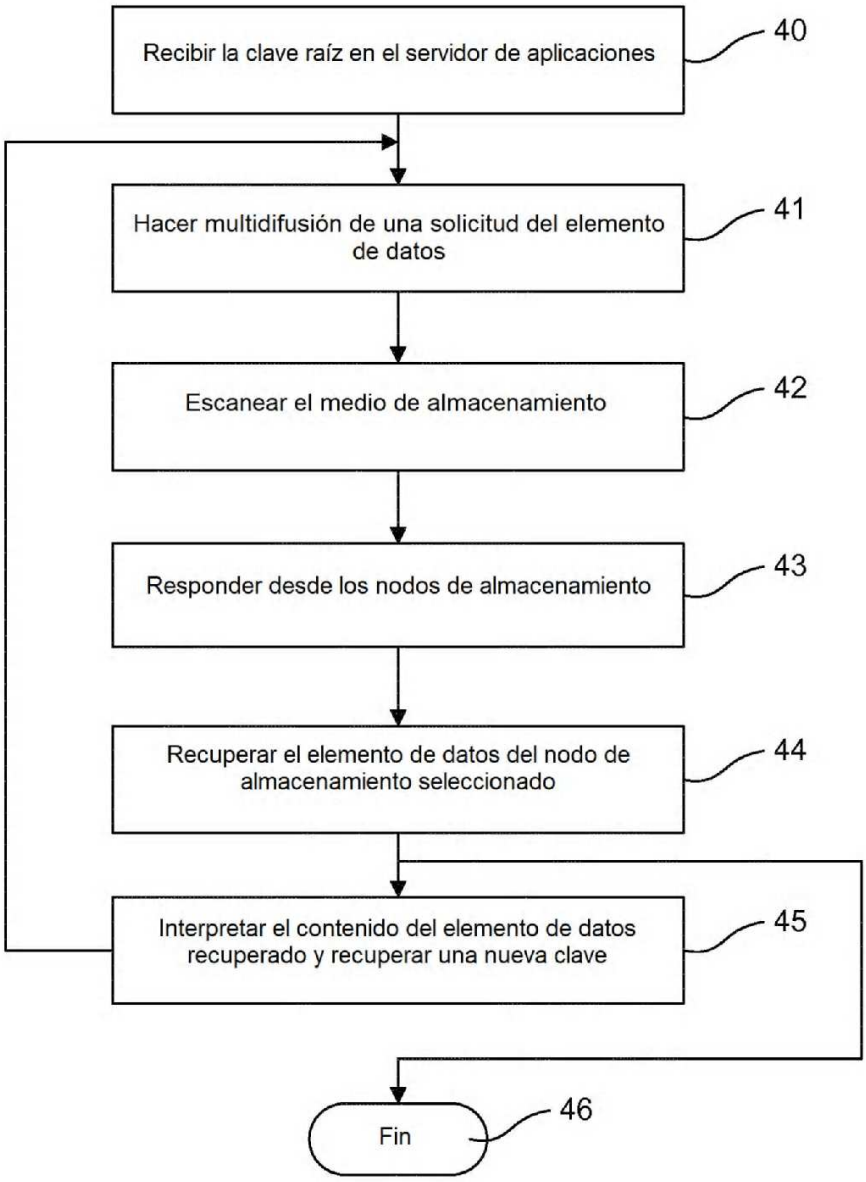
Recibir la clave raíz en el servidor de aplicaciones
40
Hacer multidifusión de una solicitud del elemento de datos
41
Escanear el medio de almacenamiento
42
Responder desde los nodos de almacenamiento
43
Recuperar el elemento de datos del nodo de almacenamiento seleccionado
44
Interpretar el contenido del elemento de datos recuperado y recuperar una nueva clave
45
Fin
46

Fig 4

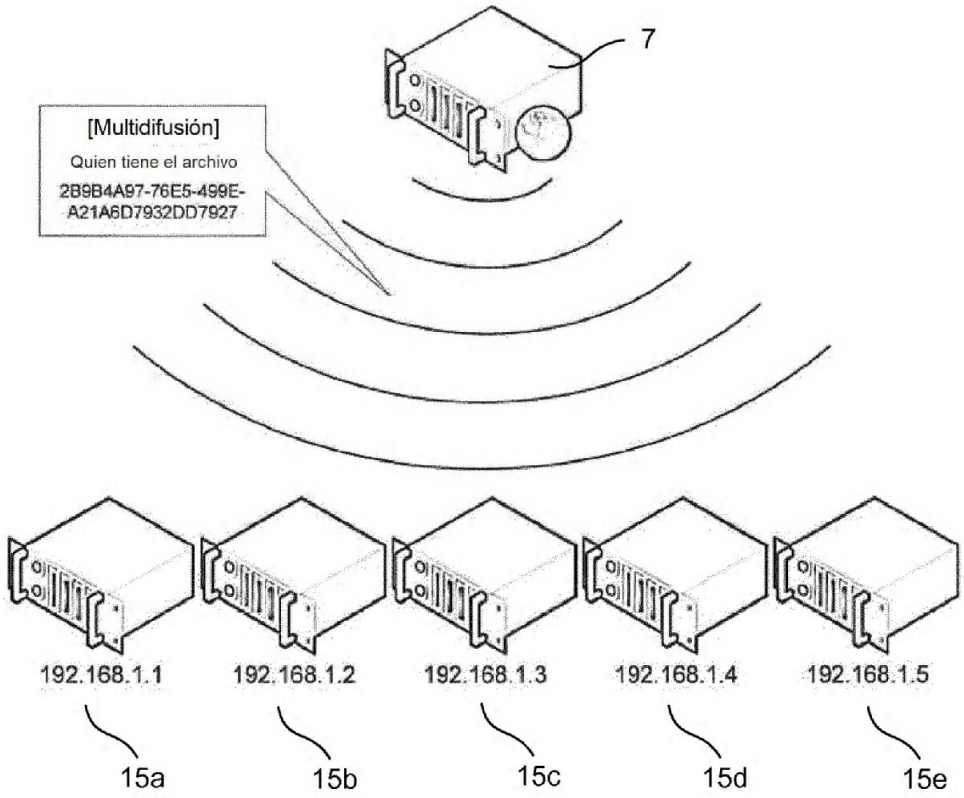
7
[Multidifusión]
Quien tiene el archivo
2B9B4A97-76E5-499E-A21A6D7932DD7927
192.168.1.1
192.168.1.2
192.168.1.3
192.168.1.4
192.168.1.5
15a
15b
15c
15d
15e

Fig 5a

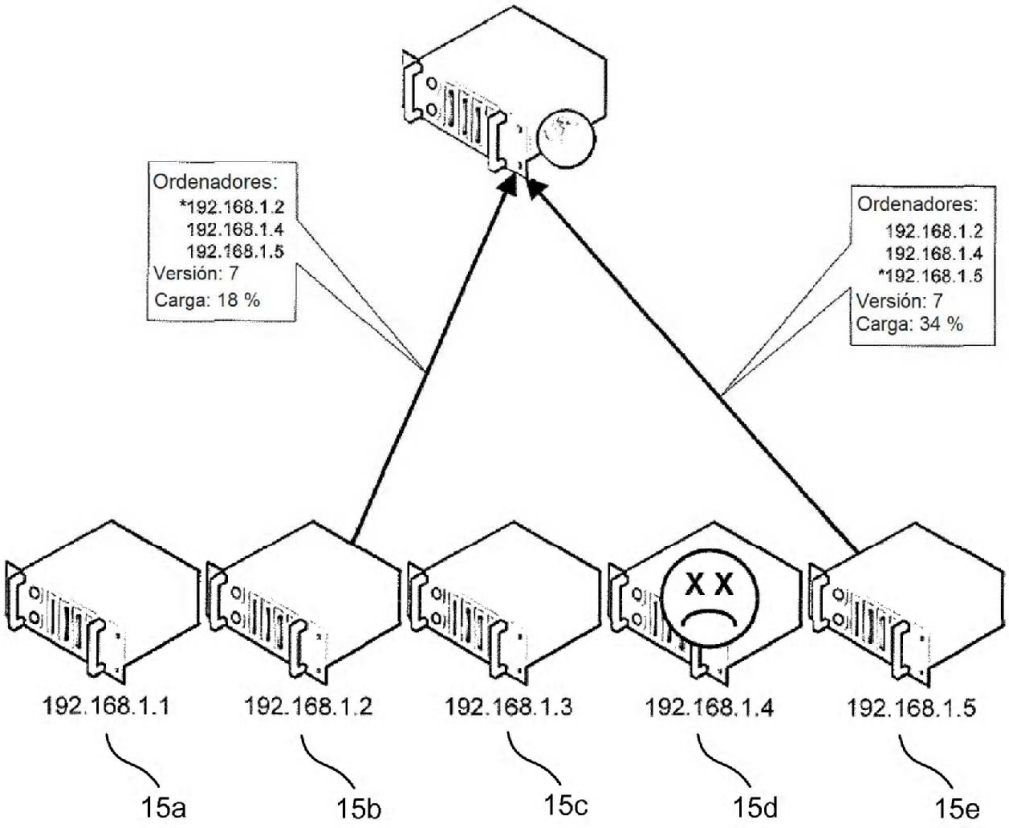
Ordenadores:
*192.168.1.2
192.168.1.4
192.168.1.5
Versión: 7
Carga: 18 %
Ordenadores:
192.168.1.2
192.168.1.4
*192.168.1.5
Versión: 7
Carga: 34 %
X X
192.168.1.1
192.168.1.2
192.168.1.3
192.168.1.4
192.168.1.5
15a
15b
15c
15d
15e

Fig 5b

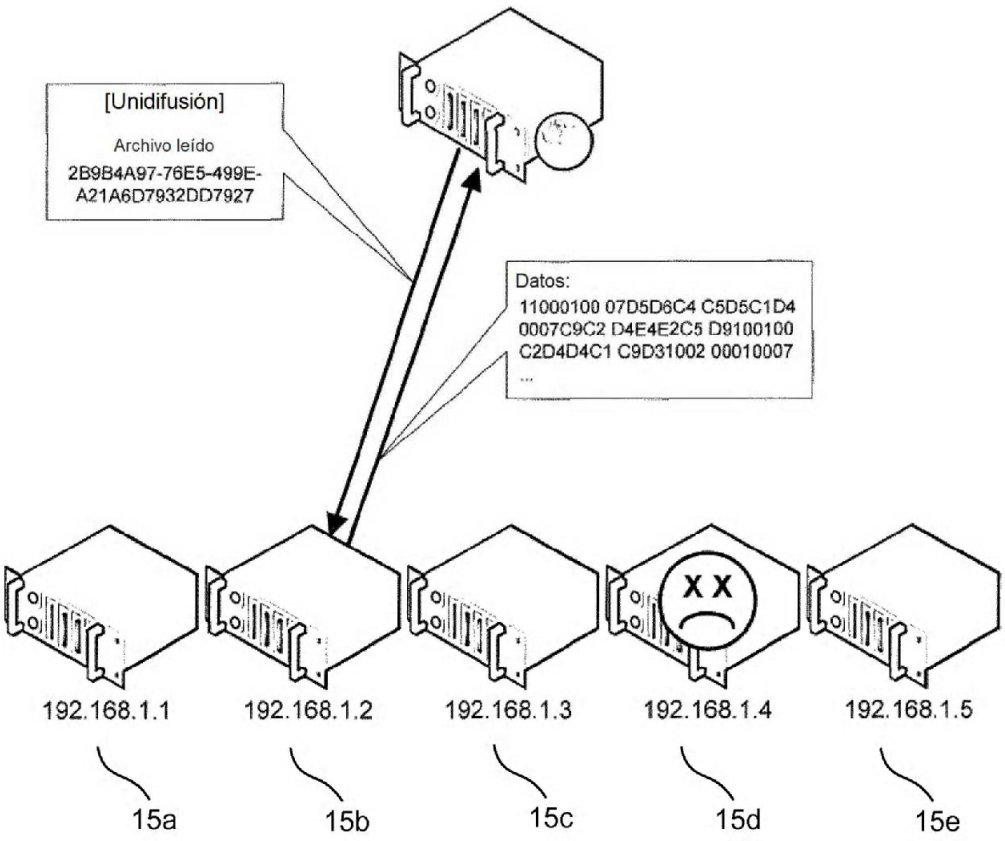
[Unidifusión]
Archivo leído
2B9B4A97-76E5-499E-A21A6D7932DD7927
Datos:
11000100 07D5D6C4 C5D5C1D4
0007C9C2 D4E4E2C5 D9100100
C2D4D4C1 C9D31002 00010007
...
X X
192.168.1.1
192.168.1.2
192.168.1.3
192.168.1.4
192.168.1.5
15a
15b
15c
15d
15e

Fig 5c

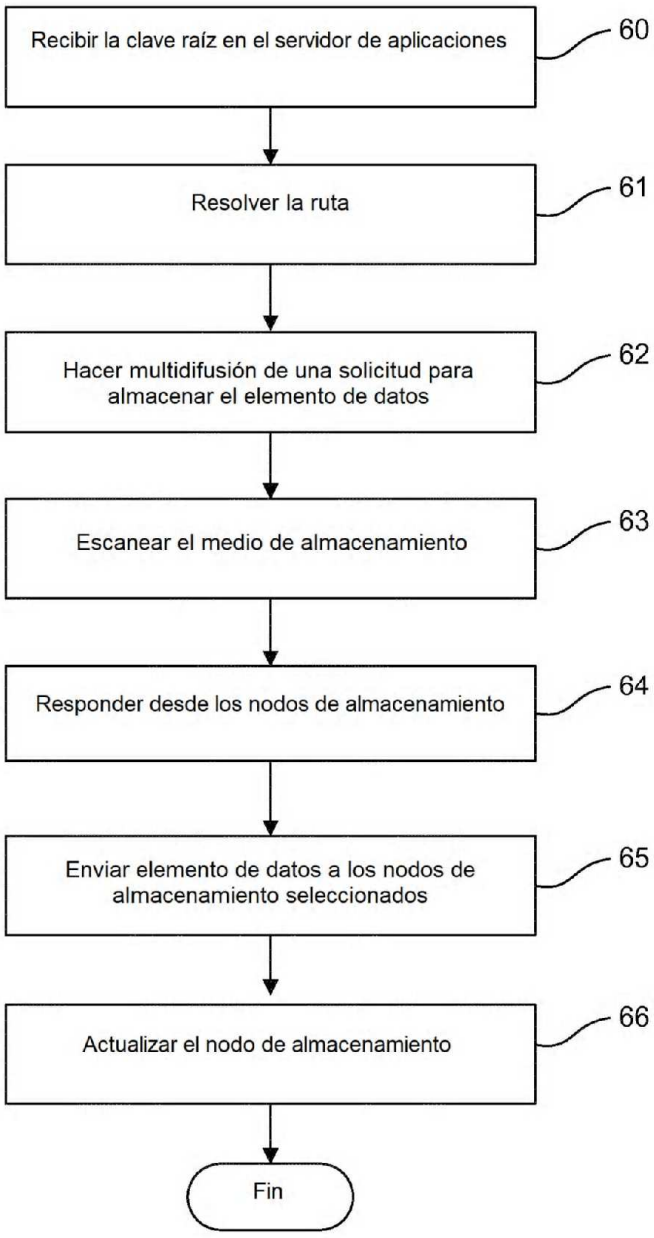
Recibir la clave raíz en el servidor de aplicaciones
60
Resolver la ruta
61
Hacer multidifusión de una solicitud para almacenar el elemento de datos
62
Escanear el medio de almacenamiento
63
Responder desde los nodos de almacenamiento
64
Enviar elemento de datos a los nodos de almacenamiento seleccionados
65
Actualizar el nodo de almacenamiento
66
Fin

Fig 6

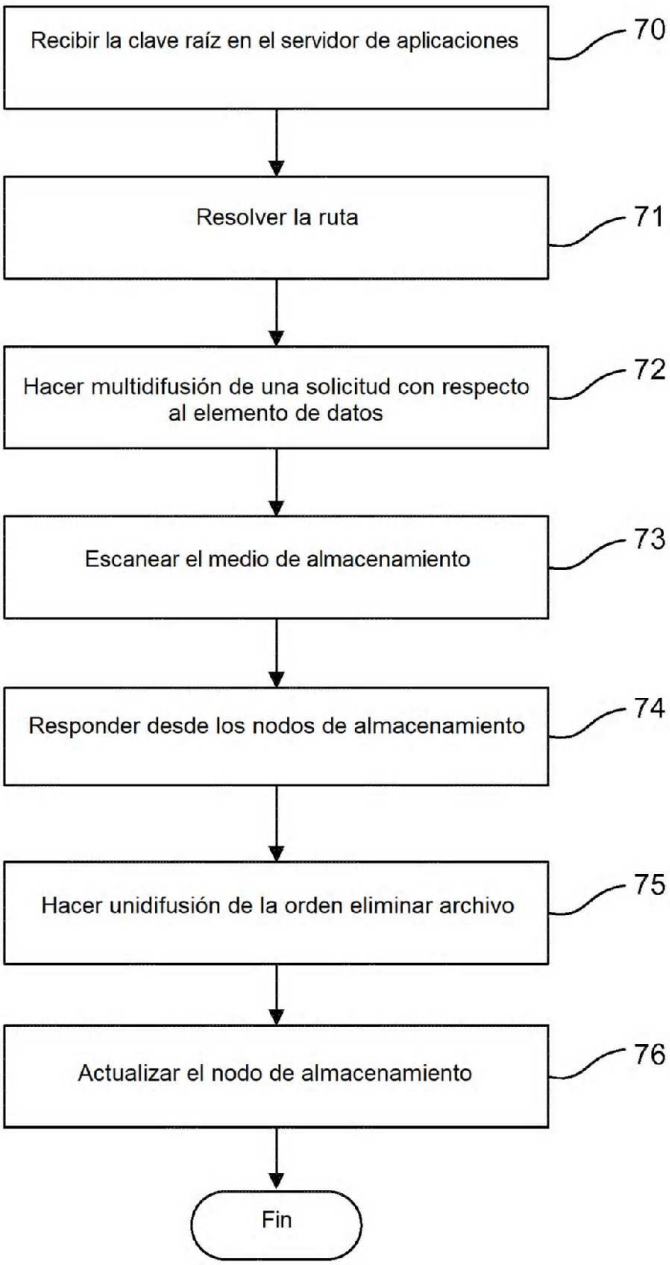
Recibir la clave raíz en el servidor de aplicaciones
70
Resolver la ruta
71
Hacer multidifusión de una solicitud con respecto al elemento de datos
72
Escanear el medio de almacenamiento
73
Responder desde los nodos de almacenamiento
74
Hacer unidifusión de la orden eliminar archivo
75
Actualizar el nodo de almacenamiento
76
Fin

Fig 7