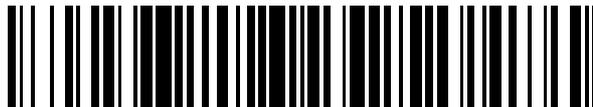


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 690 454**

51 Int. Cl.:

H04L 29/08 (2006.01)

G06F 12/02 (2006.01)

G06F 12/0862 (2006.01)

G06F 8/41 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.06.2011 E 16157884 (4)**

97 Fecha y número de publicación de la concesión europea: **15.08.2018 EP 3131265**

54 Título: **Método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT, nodo y sistema**

30 Prioridad:

13.01.2011 CN 201110006813

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

21.11.2018

73 Titular/es:

**HUAWEI TECHNOLOGIES CO. LTD. (100.0%)
Huawei Administration Building Bantian
Longgang District
Shenzhen, Guangdong 518129, CN**

72 Inventor/es:

**BAO, DONG y
YANG, DEPING**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 690 454 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

Método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT, nodo y sistema

Campo de la invención

- 5 La presente invención está relacionada con el campo de las tecnologías de la información y, en particular, con un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT, un nodo y un sistema.

Antecedentes de la invención

- 10 A medida que se desarrolla la red la información experimenta un aumento explosivo. Un sistema de almacenamiento inter pares (Peer to Peer; P2P) basado en la tecnología de tabla hash (aleatorizada) distribuida (Distributed Hash Table, DHT) es altamente extensible y puede soportar un almacenamiento masivo de datos. Por otra parte, en un sistema de almacenamiento P2P, un segmento de datos consecutivos se divide en varios bloques, y se almacena en varios nodos del sistema, de modo que un cliente no puede leer los datos ininterrumpidamente del sistema, lo que da lugar a una baja eficiencia en la lectura de datos.

- 15 En un método existente de obtención anticipada de datos para un sistema de almacenamiento DHT, un cliente determina el contenido y la ubicación de almacenamiento de los datos que se deberán utilizar a continuación de acuerdo con una regla de acceso de un programa de aplicación y, a continuación, el cliente lee los datos de cada unidad de almacenamiento apropiada y almacena los datos en el propio cliente, de modo que cuando una aplicación de una capa superior necesita los datos, la aplicación de la capa superior puede obtener directamente los datos desde el cliente sin necesidad de transmisión a través de la red, mejorando de este modo la eficiencia en la lectura. Sin embargo, la ejecución de la operación de obtención anticipada de datos por parte del sistema de almacenamiento DHT debe depender del cliente, por lo que el método de obtención anticipada tan solo mejora la eficiencia de la comunicación desde el cliente a cada nodo, pero la eficiencia de la lectura de datos del sistema aún no ha variado, esto es, la eficiencia de la lectura de datos del sistema sigue siendo baja.

- 25 El documento de ZUONING YIN Y OTROS "DHT Based Collaborative Multimedia Streaming and Caching Service (Servicio colaborativo de streaming y caching multimedia basado en DHT)", TENCON 2005 2005 REGION 10 del IEEE, IEEE, PI, (20051101), ISBN 978-0-7803-9311-0, páginas 1 - 6, presenta un modelo que crea un servicio multimedia de streaming (transmisión continua) y caching (almacenamiento anticipado) en un overlay (recubrimiento) basado en una tabla hash distribuida. Utilizan DHT para establecer un servicio de caching descentralizado. La información sobre la ubicación de los segmentos almacenados en caché se distribuye entre diferentes nodos pares y la información sobre los segmentos del mismo vídeo se organiza en un grupo de multidifusión a nivel de aplicación.

- 35 El documento US 2004/117398 A1 proporciona un sistema en el que se crea una BD en un entorno de virtualización, un servidor de gestión obtiene información de procesamiento de la BD como, por ejemplo, el plan de ejecución de procesamiento de la BD y el grado de prioridad de procesamiento del SGBD (sistema de gestión de la BD), predice a qué datos se va a acceder en un futuro próximo y el orden de los accesos basado en dicha información, y solicita que se transfieran a las cachés de los dispositivos de almacenamiento los datos a los que se va a acceder en el futuro cercano en función de los resultados de la predicción.

- 40 El documento WO 2005/066767 A1 divulga técnicas para la gestión de la memoria caché. Se recibe una nueva solicitud para acceder a un primer bloque de datos. Se determina la probabilidad de que se pueda acceder a un segundo bloque de datos en función del acceso al primer bloque de datos. Se determina si dicha probabilidad excede el umbral de lectura anticipada. Si la probabilidad excede el umbral de lectura anticipada se obtiene con antelación el segundo bloque de datos.

- 45 El documento de ZHENHUA LI Y OTROS "ContinuStreaming: Achieving high playback continuity of Gossip-based Peer-to-Peer streaming (ContinuStreaming: lograr una alta continuidad de reproducción de streaming inter pares basada en Gossip)", SIMPOSIO INTERNACIONAL DEL IEEE 2008 SOBRE PROCESAMIENTO PARALELO Y DISTRIBUIDO: [IPDPS 2008]; MIAMI, FL, 14 - 18 DE ABRIL DE 2008, IEEE, PISCATAWAY, NJ, EE. UU., (20080414), ISBN 978-1-4244-1693-6, páginas 1 - 12, divulga que el streaming inter pares (P2P) basado en gossip (protocolo de comunicación ordenador-ordenador) ha demostrado ser un método efectivo y resiliente para el streaming de contenidos de medios cualificados en entornos de red dinámicos y heterogéneos. Debido a la aleatoriedad intrínseca del gossiping, algunos segmentos de datos no se pueden diseminar a cada nodo a tiempo, lo que afecta seriamente a la continuidad de reproducción de los medios. En este documento describimos el diseño de ContinuStreaming, un sistema de streaming P2P basado en gossip que puede mantener una alta resiliencia y una baja sobrecarga, a la vez que aporta una propiedad novedosa y crítica: una cobertura completa de la diseminación de los datos. Con la ayuda de la DHT, los segmentos de datos que posiblemente se pierdan debido a la planificación de datos basada en gossip se pueden extraer previamente de forma rápida mediante recuperación de datos bajo demanda con el fin de garantizar la reproducción continua. Analizamos los resultados tanto de análisis teóricos como de simulaciones completas en diversas topologías de superposición de traza real para demostrar la efectividad de

nuestro sistema. Los resultados de la simulación muestran que ContinuStreaming supera a los sistemas representativos actuales de streaming P2P basados en gossip incrementando la continuidad de reproducción muy cerca de 1,0 con solo un 4% o menos de sobrecarga extra.

5 El documento WO2010/076536 A2 está relacionado con un sistema para procesar una solicitud emitida por un cliente para obtener datos, que incluye: una pluralidad de nodos (201, 202, 203, 204, 205) organizados en una red P2P, apropiados para almacenar un conjunto de datos distribuido en varios nodos y almacenar una tabla de identificación (200) apropiada para identificar entre la pluralidad de nodos el nodo capaz de suministrar los datos en respuesta a la solicitud, caracterizado por que dicho sistema incluye: un servidor proxy (30) que almacena dicha tabla, a la que se pretende dirigir la solicitud, siendo dicho servidor apropiado únicamente para, al recibir la solicitud:
10 identificar, de acuerdo con la tabla, el nodo capaz de suministrar los datos en respuesta a la solicitud, y retransmitirle la solicitud al nodo identificado.

15 El documento US 2010/146569 A1 divulga un sistema o método de vídeo a la carta (VoD) inter pares (P2P) que utiliza una pluralidad de nodos pares descodificadores (STB) de VoD con configuración IP en un grupo para proporcionar datos de contenidos de vídeo a un nodo par cliente solicitante del mismo grupo. Se puede designar a otro nodo par del grupo como nodo par padre para la descarga de contenidos de reproducción de alta prioridad que se requieren de forma sustancialmente inmediata para su visualización en reproducción. El contenido de baja prioridad almacenado anticipadamente se le puede descargar al nodo par cliente solicitante desde otro nodo par del grupo distinto del nodo par padre. El ancho de banda del nodo par cliente solicitante no utilizado por el módulo de reproducción para la descarga de los datos del contenido de vídeo de reproducción de alta prioridad se puede
20 asignar para descargar el contenido a almacenar anticipadamente. Utilizando el nodo par cliente solicitante, se pueden servir datos de contenido de vídeo para su visualización en reproducción en un monitor de vídeo, en donde dicho contenido de datos de vídeo servido puede comprender el contenido de la reproducción y el contenido almacenado anticipadamente.

Resumen de la invención

25 Los modos de realización de la presente invención proporcionan un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT, un nodo y un sistema, con el fin de resolver el problema de la baja eficiencia de lectura para el sistema de almacenamiento DHT en la técnica anterior.

La presente invención se define en las reivindicaciones adjuntas independientes, a las cuales se debería hacer referencia. Las características ventajosas se explican en las reivindicaciones adjuntas dependientes.

30 Breve descripción de los dibujos

Con el fin de ilustrar de forma más clara las soluciones técnicas de acuerdo con los modos de realización de la presente invención o de la técnica anterior, a continuación se presentan brevemente los dibujos adjuntos requeridos para describir los modos de realización o la técnica anterior.

35 La FIG. 1 es un diagrama de flujo de un primer modo de realización de un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención;

la FIG. 2 es un diagrama esquemático de una estructura de almacenamiento de datos en un sistema de almacenamiento de tabla hash distribuida DHT;

40 la FIG. 3 es un diagrama de flujo de interacción de la aplicación del método de obtención anticipada por parte de cada uno de los nodos de almacenamiento participantes en el método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT que se ilustra en la FIG. 1;

la FIG. 4 es un diagrama de flujo de un segundo modo de realización de un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención;

45 la FIG. 5 es un diagrama esquemático de la estructura de un primer modo de realización de un nodo de almacenamiento de un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención;

la FIG. 6 es un diagrama esquemático de la estructura de un segundo modo de realización de un nodo de almacenamiento de un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención;

50 la FIG. 7 muestra el proceso principal para el procesamiento de una solicitud de lectura por parte de un nodo de almacenamiento de un sistema de almacenamiento de tabla hash distribuida DHT, proporcionado en un modo de realización de la presente invención; y

la FIG. 8 muestra el proceso principal realizado por un nodo de almacenamiento de un sistema de almacenamiento de tabla hash distribuida DHT después de recibir una solicitud de obtención anticipada, proporcionado en un modo de realización de la presente invención.

Descripción detallada de los modos de realización

5 Con el fin de hacer más comprensibles los objetivos, las soluciones técnicas y las ventajas de los modos de realización de la presente invención, a continuación se describen de forma clara y completa las soluciones técnicas de acuerdo con los modos de realización de la presente invención haciendo referencia a los dibujos adjuntos en los modos de realización de la presente invención.

10 La FIG. 1 es un diagrama de flujo de un primer modo de realización de un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención. Tal como se ilustra en la FIG. 1, el método incluye:

S101: Un primer nodo de almacenamiento recibe una solicitud de lectura enviada por un cliente.

15 S102: El primer nodo de almacenamiento determina un bloque de datos a obtener con antelación y un segundo nodo de almacenamiento donde debe residir el bloque de datos a almacenar anticipadamente en función del bloque de datos leído y un umbral establecido de bloques de datos a obtener con antelación, y le envía una solicitud de obtención anticipada al segundo nodo de almacenamiento, en donde el primer nodo de almacenamiento y el segundo nodo de almacenamiento son los mismos o diferentes nodos de almacenamiento, la solicitud de obtención anticipada incluye información de identificación del bloque de datos a obtener con antelación, y la información de identificación se utiliza para identificar de manera unívoca el bloque de datos a obtener con antelación.

20 S103: El segundo nodo de almacenamiento lee de un disco el bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada, y almacena el bloque de datos a obtener con antelación en un buffer (memoria intermedia de almacenamiento temporal) local, de modo que un cliente lea el bloque de datos almacenado anticipadamente del buffer local del segundo nodo de almacenamiento.

25 El método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención es aplicable a un sistema de almacenamiento de tabla hash distribuida DHT. Este tipo de sistema de almacenamiento almacena los datos en forma de DHT, y durante el almacenamiento de los datos, el sistema de almacenamiento generalmente divide un segmento de datos consecutivos (es decir, datos brutos) en varios bloques y almacena los bloques de datos resultantes en varios nodos de almacenamiento en el sistema. Por ejemplo, en el sistema de almacenamiento DHT que se ilustra en la FIG. 2, los bloques de datos resultantes D1, D2, ... D8 corresponden a un segmento de datos consecutivos, y los datos brutos están almacenados en los discos de los nodos de almacenamiento N1, N2, N3 y N4. Un método de acceso a los datos para cada uno de los bloques de datos en el sistema de almacenamiento consiste generalmente en get (key, & value) (obtener (clave, & valor)), donde key es un identificador único de un bloque de datos en el sistema de almacenamiento, y value es el contenido de datos del bloque de datos. El disco para almacenar los datos puede ser, por ejemplo, un disco duro de unidad con electrónica integrada (Integrated Drive Electronics, IDE) corriente, un disco duro de tecnología avanzada de conexión en serie (Serial Advanced Technology Attachment, SATA) o un disco duro de estado sólido.

35 El primer nodo de almacenamiento (el encargado de ejecutar los pasos S101 y S102) puede ser cualquier nodo de almacenamiento en el sistema de almacenamiento DHT. Después de recibir una solicitud de lectura para leer un bloque de datos D1 determinado enviada por el cliente, el primer nodo de almacenamiento realiza una operación de obtención anticipada. En particular, como los bloques de datos que constituyen los datos brutos están numerados secuencialmente, lo que se puede reflejar específicamente mediante valores clave de los bloques de datos, el primer nodo de almacenamiento puede determinar el valor clave del bloque de datos que hay que leer después de D1 en función del valor clave del bloque de datos D1 leído, y también puede realizar, además, una operación de DHT con el valor clave del bloque de datos que hay que leer, para obtener un segundo nodo de almacenamiento en el que se encuentra el bloque de datos, y enviarle una solicitud de obtención anticipada al segundo nodo de almacenamiento, en donde la solicitud de obtención anticipada incluye información de identificación (el valor clave) del bloque de datos a obtener con antelación, de modo que el segundo nodo de almacenamiento realiza una operación de obtención anticipada del bloque de datos que hay que leer después de D1 de acuerdo con la información de identificación.

50 Se debe tener en cuenta que el segundo nodo de almacenamiento no se refiere a un nodo del sistema, sino que pueden ser múltiples nodos de almacenamiento del sistema. Esto se debe a que un bloque de datos determinado que hay que leer anticipadamente puede ser múltiple. Los bloques de datos a obtener con antelación se pueden almacenar en segundos nodos de almacenamiento diferentes. En consecuencia, después de determinar los bloques de datos a obtener con antelación, cada segundo nodo de almacenamiento puede leer desde un disco un bloque de datos que se va a leer con antelación y, a continuación, almacenar el bloque de datos en un buffer local, esto es, el segundo nodo de almacenamiento copia el bloque de datos a obtener con antelación del disco al buffer local, de tal modo que el cliente puede leer directamente el bloque de datos del buffer local. Además, el segundo nodo de almacenamiento y el primer nodo de almacenamiento pueden ser el mismo nodo de almacenamiento, esto es, el al

menos un bloque de datos a obtener con antelación determinado por el primer nodo de almacenamiento puede encontrarse en una unidad de almacenamiento local. En este caso, el primer nodo de almacenamiento lee el bloque de datos a obtener con antelación desde un disco local de acuerdo con la información de identificación del bloque de datos a obtener con antelación contenida en la solicitud de obtención anticipada, y almacena el bloque de datos a obtener con antelación en un buffer local.

Como la ejecución de la operación de obtención anticipada tiene en realidad como objetivo reducir el tiempo de transmisión de red de cada bloque de datos y la capacidad del buffer en cada uno de los nodos de almacenamiento es limitada, y se consideran factores como, por ejemplo, el factor de que si en el buffer hay más bloques de datos leídos anticipadamente se reduce la velocidad de lectura, se establece generalmente un umbral de bloques de datos a obtener con antelación, y el contador de bloques de datos a leer anticipadamente está limitado por el umbral de bloques de datos a obtener con antelación. El umbral de bloques de datos a obtener con antelación es generalmente mayor o igual que la relación entre el tiempo de lectura de un solo bloque de datos y el tiempo de transmisión en la red de un solo bloque de datos. Además, el umbral de bloques de datos a obtener con antelación también se puede determinar en función de la capacidad del buffer del nodo de almacenamiento y las características de otros dispositivos de hardware en cada nodo de almacenamiento, por ejemplo, un disco, una unidad central de procesamiento (Central Processing Unit, CPU), una placa maestra y una tarjeta de red.

Después de recibir la solicitud de lectura enviada por el cliente, cada uno de los nodos de almacenamiento del sistema comprueba si el bloque de datos a leer existe en el buffer, esto es, si el bloque de datos a leer se ha obtenido previamente. Si el bloque de datos a leer ya se encuentra en el buffer, el nodo de almacenamiento puede leer directamente el bloque de datos del buffer y devolverle el bloque de datos al cliente; y si el bloque de datos a leer no existe en el buffer, ello indica que el bloque de datos no se ha obtenido previamente, y el nodo de almacenamiento puede leer el bloque de datos desde el disco. Durante la lectura del bloque de datos o después de la lectura del bloque de datos, cada uno de los nodos de almacenamiento puede iniciar además un nuevo proceso de obtención anticipada como el que se ha descrito más arriba, por lo que no se vuelve a describir en la presente solicitud.

En el método de obtención anticipada de datos para el sistema de almacenamiento de tabla hash distribuida DHT proporcionado en este modo de realización, se puede determinar el bloque de datos a obtener con antelación en función del bloque de datos leído y el umbral establecido de bloques de datos a obtener con antelación y, a continuación, el bloque de datos a obtener con antelación del disco se copia al buffer, de modo que después de recibir la solicitud de lectura del cliente, el bloque de datos a obtener con antelación pueda ser leído directamente del buffer, mejorándose de este modo la eficiencia del acceso a los datos para cada uno de los nodos de almacenamiento del sistema.

Para ilustrar con mayor claridad el primer modo de realización del método de obtención anticipada de datos proporcionado en la presente invención, la FIG. 3 muestra un diagrama de flujo de interacción de la ejecución de un método de obtención anticipada, por parte de cada uno de los nodos de almacenamiento de un sistema de almacenamiento. Tal como se ilustra en la FIG. 3, en este modo de realización se realiza la descripción haciendo referencia a un ejemplo en el que un cliente le envía a un nodo de almacenamiento A una solicitud de lectura para leer un bloque de datos D_n , y el nodo de almacenamiento A inicia una operación de obtención anticipada. El método incluye:

S201: Un cliente localiza un nodo de almacenamiento A a partir de la clave del D_n y le envía una solicitud de lectura al nodo de almacenamiento A.

El método para localizar por parte del cliente el nodo de almacenamiento a partir de la clave del bloque de datos puede ser un método en el que se utilice un algoritmo DHT general, el cual no se describe en detalle en la presente solicitud.

El cliente puede iniciar la operación de lectura del bloque de datos en el nodo de almacenamiento A de varias formas, por ejemplo, el cliente puede iniciar la operación de lectura de datos en el nodo A mediante una interfaz de almacenamiento get (key, & value) proporcionada por el sistema de almacenamiento DHT, o iniciar la operación de lectura de datos en función de un rango de la clave, o de otras formas que no se enumeran en la presente solicitud.

S202: El nodo de almacenamiento A busca el bloque de datos D_n en el buffer; si el bloque de datos D_n existe en el buffer, el nodo de almacenamiento A recupera directamente el bloque D_n del buffer y le devuelve el bloque D_n al cliente.

S203: Si el bloque D_n no se encuentra en el buffer, lee el bloque D_n de un disco y le devuelve el bloque D_n al cliente.

S204: El nodo de almacenamiento A inicia un proceso de operación de obtención anticipada. Con el fin de simplificar la descripción, en este modo de realización se supone que el nodo de almacenamiento A determina, en función del bloque D_n y el umbral establecido de bloques de datos a obtener con antelación, que es necesario volver a leer con antelación dos bloques de datos D_{n+1} y D_{n+2} (los dos bloques de datos siguientes al bloque D_n en los datos brutos cuyos números son consecutivos al del bloque D_n en los datos brutos), respectivamente, y, de acuerdo con un

algoritmo de DHT determina que D_{n+1} y D_{n+2} se encuentran almacenados en un nodo de almacenamiento X y un nodo de almacenamiento Y, respectivamente. A continuación se ejecutan los pasos S205 y S206.

S205: El nodo de almacenamiento A le envía una solicitud de obtención anticipada del bloque de datos D_{n+1} al nodo de almacenamiento X.

- 5 S206: El nodo de almacenamiento A le envía una solicitud de obtención anticipada del bloque de datos D_{n+2} al nodo de almacenamiento Y.

Los pasos S205 y S206 no tienen un orden particular, y se pueden ejecutar de forma síncrona o asíncrona.

S207: El nodo de almacenamiento X recibe la solicitud de obtención anticipada del bloque de datos D_{n+1} y, a continuación, ejecuta un proceso de copia del bloque de datos desde un disco a un buffer.

- 10 S208: El nodo de almacenamiento Y recibe la solicitud de obtención anticipada del bloque de datos D_{n+2} y, a continuación, ejecuta un proceso de copia del bloque de datos desde un disco a un buffer.

En los pasos S207 y S208, después de recibir las solicitudes de obtención anticipada enviadas por el nodo de almacenamiento A, el nodo de almacenamiento X y el nodo de almacenamiento Y pueden copiar directamente los bloques de datos correspondientes de los discos a los buffer.

- 15 No obstante, con el fin de mejorar la efectividad de la obtención anticipada del sistema, el nodo de almacenamiento que recibe la solicitud de obtención anticipada puede comprobar además si la obtención anticipada es necesaria, y en un modo de realización que se describe más abajo se describe la operación de comprobación de si es necesaria la obtención anticipada.

- 20 S209: Cuando el cliente necesita leer el bloque de datos D_{n+1} , el cliente localiza el nodo de almacenamiento X en función de la clave del bloque D_{n+1} e inicia una operación de lectura del bloque de datos en el nodo de almacenamiento X mediante la interfaz de almacenamiento get (key, & value) proporcionada por el sistema de almacenamiento DHT.

S210: Después de recibir la solicitud de lectura del bloque de datos D_{n+1} , el nodo de almacenamiento X comprueba que el bloque D_{n+1} se encuentra en el buffer, y le devuelve directamente al cliente el bloque D_{n+1} del buffer.

- 25 S211: Cuando el cliente necesita leer el bloque de datos D_{n+2} , el cliente localiza el nodo de almacenamiento Y en función de la clave del bloque D_{n+2} e inicia una operación de lectura de bloque de datos en el nodo de almacenamiento Y mediante la interfaz de almacenamiento get (key, & value) proporcionada por el sistema de almacenamiento DHT.

- 30 S212: Después de recibir la solicitud de lectura del bloque de datos D_{n+2} , el nodo de almacenamiento Y comprueba que el bloque D_{n+2} se encuentra en el buffer, y le devuelve directamente al cliente el bloque D_{n+2} del buffer.

La FIG. 4 es un diagrama de flujo de un segundo modo de realización de un método de obtención anticipada de datos para el sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención. Tal como se ilustra en la FIG. 4, en el método de obtención anticipada de datos para el sistema DHT proporcionado en este modo de realización, después de recibir una solicitud de obtención anticipada, el nodo de almacenamiento debe comprobar además si es necesaria la operación de obtención anticipada, con el fin de mejorar la efectividad de la obtención anticipada del sistema. Concretamente, después de realizar una operación de lectura de un bloque de datos, cada uno de los nodos de almacenamiento registra el bloque de datos leído junto con la marca de tiempo correspondiente en una tabla de registro de lecturas, en donde la marca de tiempo sirve para identificar el instante en el que se ha leído el bloque de datos. Al recibir la solicitud de obtención anticipada, el nodo de almacenamiento determina si debe realizar una operación de obtención anticipada en función del contador del bloque de datos leído en la tabla de registro de lecturas y la marca de tiempo correspondiente al último tiempo de lectura. Si el contador del bloque de datos leído es mayor o igual que un umbral establecido de bloques de datos a obtener con antelación, y la diferencia entre el instante de obtención anticipada actual y el último tiempo de obtención anticipada que indica la marca de tiempo es menor o igual a un tiempo establecido, se lleva a cabo la operación de obtención anticipada.

- 45 La razón para establecer el umbral de bloques de datos leídos en el registro de lecturas es que si un cliente solo lee unos pocos bloques de datos, ello no quiere decir necesariamente que el cliente lee indudablemente los bloques de datos en los datos brutos de forma secuencial, y ello puede dar lugar a una determinación inexacta de lectura secuencial, y en consecuencia hacer que se produzca innecesariamente una obtención anticipada. Sin embargo, el umbral de bloques de datos leídos no se puede configurar con un valor muy grande, ya que esto puede reducir la eficiencia de la obtención anticipada en una etapa anterior e incrementar el consumo del sistema durante la copia y almacenamiento del registro de lecturas. En consecuencia se puede establecer el umbral de bloques de datos ya leídos en función de, por ejemplo, el contador de bloques de datos en los datos brutos y las configuraciones de hardware del sistema.

Además, la diferencia entre el instante de obtención anticipada actual y el último tiempo de lectura que indica la marca de tiempo debe ser menor o igual que un cierto tiempo, lo cual indica que el bloque de datos a obtener con antelación y los bloques de datos ya leídos corresponden todos a una operación actual de lectura de datos brutos del cliente, y no a una operación previa de lectura de datos brutos del cliente.

5 Este modo de realización proporciona un proceso de lectura de los bloques D1-D4 originado por el cliente, en el que se supone que la operación de obtención anticipada se realiza cuando el umbral establecido de bloques de datos leídos en el registro de lecturas es 3. El método incluye específicamente:

S301: Un cliente le solicita a un nodo de almacenamiento A que lea el bloque de datos D₁.

10 S302: Se supone que el bloque D₁ es el primer bloque de datos leído por el cliente, el nodo de almacenamiento A lee el bloque de datos D₁ de un disco y le devuelve el bloque de datos D₁ al cliente.

S303: El nodo de almacenamiento A anota <D₁> (incluida la marca de tiempo de D₁) en un registro de lecturas, en donde la tabla de registro de lecturas corresponde al mismo segmento de datos brutos.

15 S304: El nodo de almacenamiento A analiza y determina que el bloque D₂ es posterior al bloque de datos D₁, calcula un nodo de almacenamiento B que almacena el bloque D₂, le envía una solicitud de obtención anticipada del bloque D₂ al nodo de almacenamiento B, y anota <D₁> en el registro de bloques leídos.

20 El proceso en el que el nodo de almacenamiento A analiza el bloque D₂ puede consistir en: determinar el contador de bloques de datos a obtener con antelación de acuerdo con un umbral establecido de bloques de datos a obtener con antelación. Como el bloque de datos leído y el bloque de datos a obtener con antelación pertenecen a los mismos datos brutos, el bloque de datos leído y el bloque de datos a obtener con antelación tienen el mismo prefijo, y el bloque de datos leído y el bloque de datos a obtener con antelación están numerados secuencialmente en los datos brutos, en donde el prefijo es un identificador único de los datos brutos a los que pertenece un bloque de datos, y la posición de los datos en los datos brutos se identifica mediante un número. En consecuencia, el nodo de almacenamiento A puede determinar el prefijo y el número del bloque de datos a obtener con antelación en función del prefijo y el número del bloque de datos D₁ leído y, a continuación, realizar una operación hash (aleatorización) con el bloque de datos a obtener con antelación en función del prefijo y el número del bloque de datos a obtener con antelación, con el fin de determinar un nodo de almacenamiento para el bloque de datos a obtener con antelación.

25

30 S305: Después de recibir la solicitud de obtención anticipada desde el nodo de almacenamiento A, el nodo de almacenamiento B comprueba que en el registro de lecturas solo se encuentra el bloque de datos D₁ leído, y el nodo de almacenamiento B tan solo almacena el registro de lecturas, pero no ejecuta una operación de obtención anticipada.

S306: El cliente le solicita al nodo de almacenamiento B que lea el bloque de datos D₂.

S307: Como el nodo de almacenamiento B no ha leído anticipadamente el bloque D₂, el nodo de almacenamiento B lee el bloque de datos D₂ de un disco y le devuelve el bloque de datos D₂ al cliente.

35 S308: El nodo de almacenamiento B actualiza el registro de lecturas a <D₁, D₂> (incluida la marca de tiempo del bloque D₂).

S309: El nodo de almacenamiento B analiza el bloque D₃ y determina que el bloque D₃ es posterior al bloque de datos D₂, calcula la ubicación del bloque D₃, le envía una solicitud de obtención anticipada del bloque D₃ a un nodo de almacenamiento C en el que se encuentra el bloque D₃, y anota <D₁, D₂> en el registro histórico de bloques.

40 S310: Después de recibir la solicitud de obtención anticipada del nodo de almacenamiento B, el nodo de almacenamiento C comprueba que el registro de lecturas incluye 2 bloques de datos leídos, lo cual es menor que el umbral 3 de bloques de datos leídos, y el nodo de almacenamiento C tan solo almacena el registro de lecturas, pero no ejecuta una operación de obtención anticipada.

S311: El cliente le solicita al nodo de almacenamiento C que lea el bloque de datos D₃.

45 S312: Como el nodo de almacenamiento C no ha leído anticipadamente el bloque D₃, el nodo de almacenamiento C lee el bloque de datos D₃ de un disco y le devuelve el bloque de datos D₃ al cliente.

S313: El nodo de almacenamiento C actualiza el registro de lecturas a <D₁, D₂, D₃> (incluida la marca de tiempo del bloque D₃).

50 S314: El nodo de almacenamiento C analiza el bloque D₄ y determina que el bloque D₄ es posterior al bloque de datos D₃, calcula la ubicación del bloque D₄, le envía una solicitud de obtención anticipada del bloque D₄ a un nodo de almacenamiento D en el que se encuentra el bloque D₄, y anota <D₁, D₂, D₃> en el registro de bloques.

S315: Después de recibir la solicitud de obtención anticipada del nodo de almacenamiento C, el nodo de almacenamiento D comprueba que el registro de lecturas incluye 3 bloques de datos leídos, lo cual es igual al

umbral 3 de bloques de datos leídos, y el nodo de almacenamiento D comprueba, además, si la diferencia entre el instante de lectura anticipada actual y el último tiempo de lectura indicado por la marca de tiempo es menor o igual que el tiempo establecido, y en la presente solicitud se asume que la diferencia es menor o igual que el tiempo establecido, por lo que el nodo de almacenamiento almacena el registro de lecturas, lee el bloque D₄ de un disco, y almacena el bloque D₄ en un buffer.

S316: El cliente le solicita al nodo de almacenamiento D que lea el bloque de datos D₄.

S317: Como el nodo de almacenamiento D ha leído el bloque D₄ anticipadamente, el nodo de almacenamiento D lee directamente el bloque de datos D₄ del buffer y le devuelve el bloque de datos D₄ al cliente.

A continuación, el nodo de almacenamiento D continúa realizando una operación de obtención anticipada de un bloque de datos que no se vuelve a describir en la presente solicitud.

Además, como una vez que el contador de bloques de datos ya leídos en el registro de lecturas es igual al umbral establecido de bloques de datos leídos, se puede realizar la lectura, con el fin de mejorar el tiempo para realizar la determinación de la obtención anticipada por parte del nodo de almacenamiento, y economizar espacio de almacenamiento de la tabla de registro de lecturas, si el contador del bloque de datos de la tabla de registro de lecturas es mayor que el umbral establecido de bloques de datos leídos se borra un bloque de datos leído previamente en función de la marca de tiempo, de tal modo que el contador del bloque de datos de la tabla de registro es igual al umbral establecido de bloques de datos leídos.

Se debe observar que, con el fin de asegurar la concordancia de un contenido con un bloque de datos en un disco, si se vuelve a grabar en un buffer un bloque de datos a obtener con antelación, es necesario volver a grabar de forma sincronizada en el disco el mismo bloque de datos; análogamente, si se graba de nuevo un bloque de datos en el disco, es necesario volver a grabar de forma sincronizada en el buffer el mismo bloque de datos a obtener con antelación. Para registrar la relación entre el bloque de datos a obtener con antelación en el buffer y el bloque de datos en el disco se puede utilizar una cierta estructura de datos, de modo que cuando se modifique el bloque de datos en el bloque de datos del disco se pueda modificar de forma sincronizada el bloque de datos a obtener con antelación.

Las personas experimentadas en la técnica deberían entender que la totalidad o una parte de los procesos del método de acuerdo con los modos de realización se pueden realizar mediante un hardware apropiado bajo el control de las instrucciones de un programa para ordenador. El programa puede estar almacenado en un medio de almacenamiento legible por un ordenador. Al ejecutarse el programa se lleva a cabo el proceso del método de acuerdo con los modos de realización. El medio de almacenamiento puede ser un disco, una memoria de solo lectura (Read-Only Memory, ROM), una memoria de acceso aleatorio (Random Access Memory, RAM), etc.

La FIG. 5 es un diagrama esquemático de la estructura de un primer modo de realización de un nodo de almacenamiento del sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención. Tal como se ilustra en la FIG. 5, el equipo incluye un módulo 11 de análisis, un módulo 12 de notificación y un módulo 13 de obtención anticipada.

El módulo 11 de análisis está configurado para recibir una solicitud de lectura enviada por un cliente y determinar un bloque de datos a obtener con antelación y un nodo de almacenamiento en el que se encuentra el bloque de datos a obtener con antelación en función de un bloque de datos leído y un umbral establecido de bloques de datos a obtener con antelación.

El módulo 12 de notificación está configurado para enviarle una solicitud de obtención anticipada al nodo de almacenamiento, en donde la solicitud de obtención anticipada incluye información de identificación del bloque de datos a obtener con antelación, la información de identificación se utiliza para identificar de forma unívoca el bloque de datos a obtener con antelación, y hay al menos un bloque de datos que es el bloque de datos a obtener con antelación.

El módulo 13 de obtención anticipada está configurado para leer de un disco el bloque de datos a obtener con antelación determinado por el módulo 11 de análisis de acuerdo con la solicitud de obtención anticipada, y almacenar el bloque de datos a obtener con antelación en un buffer local, de modo que el cliente lea el bloque de datos a obtener con antelación del buffer local.

De acuerdo con el nodo de almacenamiento del sistema de almacenamiento de tabla hash distribuida DHT proporcionado en este modo de realización, el bloque de datos a obtener con antelación se puede determinar en función del bloque de datos leído y el umbral establecido de bloques de datos a obtener con antelación y, a continuación, el bloque de datos a obtener con antelación en el disco se copia al buffer, de modo que cuando se recibe la solicitud de lectura desde el cliente, el bloque de datos a obtener con antelación se puede leer directamente del buffer, mejorándose de este modo la eficiencia del acceso a los datos para cada uno de los nodos de almacenamiento del sistema y economizando los recursos del sistema.

La FIG. 6 es un diagrama esquemático de la estructura de un segundo modo de realización de un nodo de almacenamiento del sistema de almacenamiento de tabla hash distribuida DHT proporcionado en la presente invención. Tal como se ilustra en la FIG. 6, el equipo incluye un módulo 11 de análisis, un módulo 12 de notificación y un módulo 13 de obtención anticipada.

5 A partir del modo de realización anterior, el nodo de almacenamiento puede incluir, además:

un módulo 14 de registro, configurado para registrar el bloque de datos leído junto con una marca de tiempo del mismo en una tabla de registro de lecturas, en donde la marca de tiempo sirve para identificar el instante en el que se ha leído el bloque de datos. La solicitud de obtención anticipada enviada por el módulo 12 de notificación incluye, además, la tabla de registro de lecturas. Asimismo, el nodo de almacenamiento puede incluir además:

10 un módulo 15 de determinación, configurado para comprobar si el bloque de datos a obtener con antelación se debe adquirir con antelación en función del contador del bloque de datos leído y la marca de tiempo con el último tiempo de lectura en la tabla de registro de lecturas registrados por el módulo 14 de registro.

15 Si el módulo 15 de determinación comprueba que el contador de bloques de datos leídos es mayor o igual que el umbral establecido de bloques de datos leídos, y la diferencia entre el instante de obtención anticipada actual y el último tiempo de lectura indicado por la marca de tiempo es menor o igual que un tiempo establecido, el módulo 13 de obtención anticipada lee del disco el bloque de datos a obtener con antelación, y almacena en el buffer local el bloque de datos a obtener con antelación.

Asimismo, el nodo de almacenamiento puede incluir, además:

20 un módulo 16 de actualización, configurado para borrar un bloque de datos leído previamente en función de la marca de tiempo si el contador del bloque de datos registrado en la tabla de registro de lecturas por el módulo 14 de registro es mayor que el umbral establecido de bloques de datos leídos, de modo que el contador del bloque de datos registrado en la tabla de registro es igual al umbral establecido de bloques de datos leídos.

Además, el módulo 11 de análisis se puede configurar específicamente para

25 determinar el contador del bloque de datos a obtener con antelación de acuerdo con el umbral establecido de bloques de datos a obtener con antelación;

determinar la información de identificación del bloque de datos a obtener con antelación a partir de la información de identificación del bloque de datos leído, en donde la información de identificación consta de un prefijo y un número; el prefijo es un identificador único de los datos brutos a los que pertenece el bloque de datos, y el número identifica la posición del bloque de datos en los datos brutos; y

30 realizar una operación hash con el bloque de datos a obtener con antelación en función del prefijo y el número del bloque de datos a obtener con antelación, con el fin de determinar el nodo de almacenamiento del bloque de datos a obtener con antelación.

Asimismo, el módulo 13 de obtención anticipada se puede configurar, además, específicamente para

35 comprobar si el bloque de datos a obtener con antelación existe en el buffer local de acuerdo con la información de identificación; y

si el bloque de datos a obtener con antelación no se encuentra en el buffer local, leer del disco el bloque de datos a obtener con antelación de acuerdo con la información de identificación, y almacenar en el buffer local el bloque de datos a obtener con antelación, de tal modo que el cliente lee del buffer local el bloque de datos a obtener con antelación.

40 El equipo incluye además:

un módulo 17 de regrabación, configurado para volver a grabar en el buffer local el bloque de datos a obtener con antelación, y volver a grabar de forma sincronizada en el disco el mismo bloque de datos; y configurado, además, para volver a grabar en el disco el bloque de datos, y volver a grabar de forma sincronizada en el buffer local el mismo bloque de datos a obtener con antelación.

45 El nodo de almacenamiento del sistema de almacenamiento de tabla hash distribuida DHT que se ilustra en la FIG. 5 y la FIG. 6 puede ser cualquier nodo de almacenamiento en un sistema de almacenamiento DHT, y el hardware de almacenamiento que se ilustra en la FIG. 5 y la FIG. 6 incluye, por ejemplo, el buffer, el disco, una CPU, una placa maestra y una tarjeta de red. Para las operaciones específicas realizadas por el nodo de almacenamiento al ejecutar el método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT se puede hacer referencia a las descripciones en los modos de realización anteriores. La FIG. 7 ilustra el proceso principal en el que un nodo de almacenamiento de un sistema de almacenamiento de tabla hash distribuida DHT procesa una solicitud de lectura proporcionado en un modo de realización de la presente invención. La FIG. 7 no muestra todos los módulos que hay en el nodo de almacenamiento ni las operaciones relevantes de los mismos, y

para los detalles del proceso de ejecución por parte del nodo de almacenamiento de una operación de obtención anticipada se puede hacer referencia a los métodos descritos en los modos de realización anteriores. Haciendo referencia a la FIG. 7, el proceso incluye específicamente:

S401: Un cliente le solicita a un nodo de almacenamiento que lea el bloque de datos D_1 .

5 Se debe observar que, como lo que se proporciona en la presente invención es un método de obtención anticipada de bloques de datos para un sistema de almacenamiento DHT, el módulo funcional relevante para ejecutar una operación de obtención anticipada se describe esencialmente en los modos de realización de los nodos de almacenamiento del sistema de almacenamiento DHT que se proporcionan en la presente invención, y el módulo funcional relevante para ejecutar una solicitud de lectura no se describe de forma específica. Sin embargo, el módulo para ejecutar la operación de obtención anticipada y el módulo para ejecutar la operación de lectura pueden ser en realidad la misma unidad funcional en el nodo de almacenamiento. Para simplificar la descripción, el módulo 13 de obtención anticipada que se proporciona en el modo de realización de la presente invención puede servir como módulo de ejecución para ejecutar la operación de obtención anticipada y la operación de lectura.

10 Por consiguiente, en el paso S401, el módulo que recibe la solicitud para leer el bloque de datos D_1 enviada por el cliente es el módulo 13 de obtención anticipada.

S402: El módulo 13 de obtención anticipada del nodo de almacenamiento busca el bloque D_1 en un buffer, y comprueba si el bloque de datos D_1 ya se ha obtenido previamente.

S403: Si el bloque de datos D_1 ya se ha obtenido, el módulo 13 de obtención anticipada obtiene directamente el bloque de datos D_1 del buffer y le devuelve el bloque de datos D_1 al cliente.

20 S404: Si el bloque de datos D_1 no se ha obtenido previamente, el módulo 13 de obtención anticipada lee el bloque de datos D_1 de un disco.

S405: Devolverle el bloque de datos D_1 al cliente.

S406: El módulo 14 de registro anota $\langle D_1 \rangle$ (incluida la marca de tiempo de D_1) al registro de lecturas.

S407: El módulo 13 de obtención anticipada le notifica al módulo 11 de análisis el bloque de datos leído.

25 S408: El módulo 11 de análisis analiza el contador de bloques de datos que se deben adquirir a continuación anticipadamente y un nodo de almacenamiento; si el bloque de datos a obtener con antelación se encuentra en el nodo de almacenamiento local se ejecutan los pasos S407 y S408; y si el bloque de datos a obtener con antelación se encuentra en otro nodo de almacenamiento, se ejecuta el paso S409.

30 S409: El módulo 12 de notificación le envía una solicitud de obtención anticipada al módulo 13 de obtención anticipada.

S410: El módulo 15 de determinación determina si hay que leer anticipadamente el bloque de datos a obtener con antelación; y, si es así, se ejecuta el paso S411.

S411: El módulo 13 de obtención anticipada lee el bloque de datos del disco local y lo almacena en el buffer.

35 S412: El módulo 12 de notificación puede enviarle una solicitud de obtención anticipada a otro nodo de almacenamiento.

40 En este modo de realización únicamente se describen la operación de lectura realizada después de que cualquier nodo de almacenamiento del sistema de almacenamiento DHT haya recibido la solicitud de lectura enviada por el cliente y el proceso principal para activar la operación de obtención anticipada, y para los detalles del proceso de ejecución por parte del nodo de almacenamiento de la operación de obtención anticipada se puede hacer referencia a los métodos descritos en los modos de realización anteriores, por lo que no se vuelven a describir en la presente solicitud.

45 La FIG. 8 ilustra el proceso principal que realiza un nodo de almacenamiento del sistema de almacenamiento de tabla hash distribuida DHT proporcionado en un modo de realización de la presente invención después de recibir una solicitud de obtención anticipada. La FIG. 8 no muestra todos los módulos que hay en el nodo de almacenamiento ni las operaciones relevantes de los mismos, y para los detalles del proceso de ejecución por parte del nodo de almacenamiento de una operación de obtención anticipada se puede hacer referencia a los métodos descritos en los modos de realización anteriores. Tal como se ilustra en la FIG. 8, el proceso incluye específicamente:

50 S501: Un módulo 13 de obtención anticipada del nodo de almacenamiento recibe una solicitud de obtención anticipada de un nodo D_n enviada por otro nodo de almacenamiento, en donde la solicitud incluye el registro de lecturas de los datos brutos.

S502: El módulo 13 de obtención anticipada comprueba si el bloque de datos D_n existe en un buffer; si el bloque de datos D_n se encuentra en el buffer, ello indica que se ha realizado una operación de obtención anticipada del bloque de datos, y la operación de obtención anticipada ya no se lleva a cabo; en caso contrario, se ejecuta el paso S503.

5 S503: El módulo 13 de obtención anticipada le entrega el registro de lecturas de los datos brutos al módulo 15 de determinación.

S504: El módulo 15 de determinación determina si hay que copiar el bloque D_n del disco local a un buffer; si es así, se ejecuta el paso S505.

S505: El módulo 15 de determinación le envía una instrucción de ejecución de obtención anticipada al módulo 13 de obtención anticipada.

10 S506: El módulo 13 de obtención anticipada lee el bloque D_n del disco local.

S507: El módulo 13 de obtención anticipada almacena el bloque D_n leído en el buffer local.

15 La presente invención proporciona, además, un modo de realización de un sistema de almacenamiento de tabla hash distribuida DHT, y el sistema incluye múltiples nodos de almacenamiento proporcionados en los modos de realización de la presente invención. Para los diagramas estructurales del sistema se puede hacer referencia a la FIG. 5 y la FIG. 6.

En un modo de realización de ejemplo, el sistema de almacenamiento DHT puede incluir un primer nodo de almacenamiento y un segundo nodo de almacenamiento.

20 El primer nodo de almacenamiento está configurado para recibir una solicitud de lectura enviada por un cliente, determinar un bloque de datos a obtener con antelación y un segundo nodo de almacenamiento en el que se encuentra el bloque de datos a obtener con antelación, en función del bloque de datos leído y un umbral establecido de bloques de datos a obtener con antelación, y enviarle una solicitud de obtención anticipada al segundo nodo de almacenamiento, en donde el primer nodo de almacenamiento y el segundo nodo de almacenamiento son el mismo nodo o nodos de almacenamiento diferentes, la solicitud de obtención anticipada incluye información de identificación del bloque de datos a obtener con antelación, y la información de identificación se utiliza para
25 identificar de forma unívoca el bloque de datos a obtener con antelación.

El segundo nodo de almacenamiento está configurado para leer de un disco el bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada enviada por el primer nodo de almacenamiento, y almacenar en un buffer local el bloque de datos a obtener con antelación, de modo que el cliente lee el bloque de datos a obtener con antelación del buffer local del segundo nodo de almacenamiento.

30 Por otra parte, el primer nodo de almacenamiento se puede configurar, además, para registrar el bloque de datos leído y la marca de tiempo del mismo en una tabla de registro de lecturas, en donde la marca de tiempo sirve para identificar el instante en el que se ha leído el bloque de datos, y la tabla de registro de lecturas corresponde al mismo segmento de datos brutos.

35 La solicitud de obtención anticipada enviada por el primer nodo de almacenamiento al segundo nodo de almacenamiento incluye, además, la tabla de registro de lecturas.

Por otra parte, el segundo nodo de almacenamiento se puede configurar, además, para determinar si es necesario adquirir anticipadamente el bloque de datos a obtener con antelación en función del contador del bloque de datos leído y la marca de tiempo con el último tiempo de lectura en la tabla de registro de lectura.

40 Si el contador del bloque de datos leído es mayor o igual que un umbral establecido de bloques de datos leídos, y la diferencia entre el instante de obtención anticipada actual y el último tiempo de lectura que se indica en la marca de tiempo es menor o igual que el tiempo establecido, el bloque de datos a obtener con antelación se lee del disco, y el bloque de datos a obtener con antelación se almacena en el buffer local.

45 Por otra parte, el primer nodo de almacenamiento se puede configurar específicamente, además, para determinar el contador del bloque de datos a obtener con antelación en función del umbral establecido de bloques de datos a obtener con antelación.

La información de identificación del bloque de datos a obtener con antelación se determina a partir de la información de identificación del bloque de datos leído, en donde la información de identificación consta de un prefijo y un número, el prefijo es un identificador único de los datos brutos a los que pertenece el bloque de datos, y el número identifica la posición del bloque de datos en los datos brutos.

50 La operación hash se realiza con el bloque de datos a obtener con antelación en función del prefijo y el número del bloque de datos a obtener con antelación, con el fin de determinar el bloque de datos a obtener con antelación y el segundo nodo de almacenamiento en el que se encuentra el bloque de datos a obtener con antelación, y la solicitud de obtención anticipada se le envía al segundo nodo de almacenamiento.

En el sistema de almacenamiento de tabla hash distribuida DHT que se proporciona en este modo de realización, cada uno de los nodos de almacenamiento interactúa ejecutando un proceso de una operación de obtención anticipada de datos, para el que se puede hacer referencia al método de obtención anticipada de datos para la tabla hash distribuida DHT que se proporciona en la presente invención, por lo que no se vuelve a describir en la presente solicitud.

5

REIVINDICACIONES

5 1. Un método de obtención anticipada de datos para un sistema de almacenamiento de tabla hash distribuida DHT que comprende un primer nodo de almacenamiento y un segundo nodo de almacenamiento, comprendiendo dicho método de obtención anticipada de datos:

recibir (S101), por parte de un primer nodo de almacenamiento, una solicitud de lectura para leer un bloque de datos enviada por un cliente que ha localizado el primer nodo de almacenamiento mediante un algoritmo DHT;

10 determinar (S102), por parte del primer nodo de almacenamiento, un bloque de datos a obtener con antelación en función de la lectura de un bloque de datos leído de acuerdo con la solicitud de lectura y un umbral establecido de bloques de datos a obtener con antelación;

determinar (S102), por parte del primer nodo de almacenamiento, un segundo nodo de almacenamiento en el que se encuentra el bloque de datos a obtener con antelación en función del bloque de datos a obtener con antelación;

15 enviar (S102), por parte del primer nodo de almacenamiento, una solicitud de obtención anticipada al segundo nodo de almacenamiento, en donde la solicitud de obtención anticipada comprende información de identificación del bloque de datos a obtener con antelación, y la información de identificación se utiliza para identificar de forma unívoca el bloque de datos a obtener con antelación; y

leer (S103) de un disco, por parte del segundo nodo de almacenamiento, el bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada, y

20 almacenar (S103), por parte del segundo nodo de almacenamiento, el bloque de datos a obtener con antelación en un buffer (memoria intermedia de almacenamiento temporal) local del segundo nodo de almacenamiento, de modo que el cliente lee el bloque de datos a obtener con antelación del buffer local del segundo nodo de almacenamiento.

2. El método de acuerdo con la reivindicación 1, que comprende además:

25 registrar, por parte del primer nodo de almacenamiento, el bloque de datos leído y una marca de tiempo del bloque de datos leído en una tabla de registro de lecturas, en donde la marca de tiempo se utiliza para identificar el instante en el que se ha leído el bloque de datos, y la tabla de registro de lecturas corresponde a un segmento de datos brutos consecutivos;

30 en donde la solicitud de obtención anticipada enviada por el primer nodo de almacenamiento al segundo nodo de almacenamiento incluye, además, la tabla de registro de lecturas.

3. El método de acuerdo con la reivindicación 2, en donde antes de leer del disco, por parte del segundo nodo de almacenamiento, el bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada, el método comprende además:

35 determinar, por parte del segundo nodo de almacenamiento, si es necesario anticipar la lectura del bloque de datos a obtener con antelación en función de un contador del bloque de datos leído y la marca de tiempo con el último tiempo de lectura en la tabla de registro de lectura; y

40 leer del disco, por parte del segundo nodo de almacenamiento, el bloque de datos a obtener con antelación si el contador del bloque de datos leído es mayor o igual que un umbral de bloques de datos leídos y la diferencia entre el instante de obtención anticipada actual y el último tiempo de lectura que indica la marca de tiempo es menor o igual que un tiempo establecido, y almacenar el bloque de datos a obtener con antelación en el buffer local del segundo nodo de almacenamiento.

4. El método de acuerdo con la reivindicación 3, que comprende además:

45 borrar, por parte del segundo nodo de almacenamiento, un bloque de datos leído previamente en función de una marca de tiempo si el contador del bloque de datos en la tabla de registro de lecturas es mayor que el umbral establecido de bloques de datos leídos, de tal modo que el contador del bloque de datos en la tabla de registro de lecturas sea igual al umbral de bloques de datos leídos.

5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en donde la determinación, por parte del primer nodo de almacenamiento, del bloque de datos a obtener con antelación en función del bloque de datos leído y el umbral establecido de bloques de datos a obtener con antelación comprende:

50 determinar, por parte del primer nodo de almacenamiento, un contador del bloque de datos a obtener con antelación de acuerdo con el umbral establecido de bloques de datos a obtener con antelación;

- 5 determinar, por parte del primer nodo de almacenamiento, la información de identificación del bloque de datos a obtener con antelación a partir de la información de identificación del bloque de datos leídos, en donde la información de identificación consta de un prefijo y un número, el prefijo es un identificador único de los datos brutos datos a los que pertenece el bloque de datos leído, y el número identifica la posición del bloque de datos leído en los datos brutos; y
- realizar una operación hash (aleatorización) con el bloque de datos a obtener con antelación en función del prefijo y el número del bloque de datos a obtener con antelación, con el fin de determinar el bloque de datos a obtener con antelación y el segundo nodo de almacenamiento en el que se encuentra bloque de datos a obtener con antelación.
- 10 6. El método de acuerdo con la reivindicación 5, en donde la lectura del disco, por parte del segundo nodo de almacenamiento, del bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada consiste específicamente en:
- comprobar, por parte del segundo nodo de almacenamiento, si el bloque de datos a obtener con antelación existe en el buffer local de acuerdo con la información de identificación; y
- 15 si el bloque de datos a obtener con antelación no se encuentra en el buffer local, leer del disco, por parte del segundo nodo de almacenamiento, el bloque de datos a obtener con antelación de acuerdo con la información de identificación.
7. Un sistema de almacenamiento de tabla hash distribuida DHT, que comprende un primer nodo de almacenamiento y un segundo nodo de almacenamiento, en donde
- 20 el primer nodo de almacenamiento está configurado para recibir una solicitud de lectura para leer un bloque de datos enviada por un cliente que ha localizado el primer nodo de almacenamiento mediante un algoritmo DHT, para determinar un bloque de datos a obtener con antelación en función de la lectura de un bloque de datos leído de acuerdo con la solicitud de lectura de datos y un umbral establecido de bloques de datos a obtener con antelación, para determinar un segundo nodo de almacenamiento donde se encuentra el bloque de datos a obtener con antelación de acuerdo con el bloque de datos a obtener con antelación y para
- 25 enviarle una solicitud de obtención anticipada al segundo nodo de almacenamiento, en donde la solicitud de obtención anticipada comprende información de identificación del bloque de datos a obtener con antelación, y la información de identificación se utiliza para identificar de forma unívoca el bloque de datos a obtener con antelación; y
- 30 el segundo nodo de almacenamiento está configurado para leer de un disco el bloque de datos a obtener con antelación de acuerdo con la solicitud de obtención anticipada enviada por el primer nodo de almacenamiento, y para
- almacenar el bloque de datos a obtener con antelación en un buffer local del segundo nodo de almacenamiento, de tal modo que el cliente lee el bloque de datos a obtener con antelación del buffer local del segundo nodo de almacenamiento;
- 35 en donde el cliente, el primer nodo y el segundo nodo son entidades distintas interconectadas entre sí mediante el algoritmo DHT.
8. El sistema de almacenamiento de tabla hash distribuida DHT de acuerdo con la reivindicación 7, en donde el primer nodo de almacenamiento está configurado, además, para registrar el bloque de datos leído y una marca de tiempo del bloque de datos leído en una tabla de registro de lecturas, en donde la marca de tiempo se utiliza para identificar el instante en el que se ha leído el bloque de datos, y la tabla de registro de lecturas corresponde a un segmento de los datos brutos;
- 40 en donde que la solicitud de obtención anticipada enviada por el primer nodo de almacenamiento al segundo nodo de almacenamiento comprende, además, la tabla de registro de lecturas.
9. El sistema de almacenamiento de tabla hash distribuida DHT de acuerdo con la reivindicación 8, en donde el segundo nodo de almacenamiento está configurado, además, para determinar si es necesario leer anticipadamente el bloque de datos a obtener con antelación de acuerdo con un contador del bloque de datos leído y una marca de tiempo con el último tiempo de lectura en la tabla de registro de lecturas; y
- 45 leer del disco el bloque de datos a obtener con antelación si el contador del bloque de datos leído es mayor o igual que un umbral establecido de bloques de datos leídos, y una diferencia entre el instante de obtención anticipada actual y el último tiempo de lectura indicado por la marca de tiempo es menor o igual que un tiempo establecido, y almacenar el bloque de datos a obtener con antelación en el buffer local del segundo nodo de almacenamiento.
- 50

10. El sistema de almacenamiento de tabla hash distribuida DHT de acuerdo con una cualquiera de las reivindicaciones 7 a 9, en donde el primer nodo de almacenamiento está configurado específicamente para determinar un contador del bloque de datos a obtener con antelación de acuerdo con el umbral establecido de bloques de datos a obtener con antelación;

- 5 determinar la información de identificación del bloque de datos a obtener con antelación a partir de la información de identificación del bloque de datos leído, en donde la información de identificación consta de un prefijo y un número, el prefijo es un identificador único de los datos brutos a los que pertenece el bloque de datos, y el número identifica la posición del bloque de datos leído en los datos brutos.

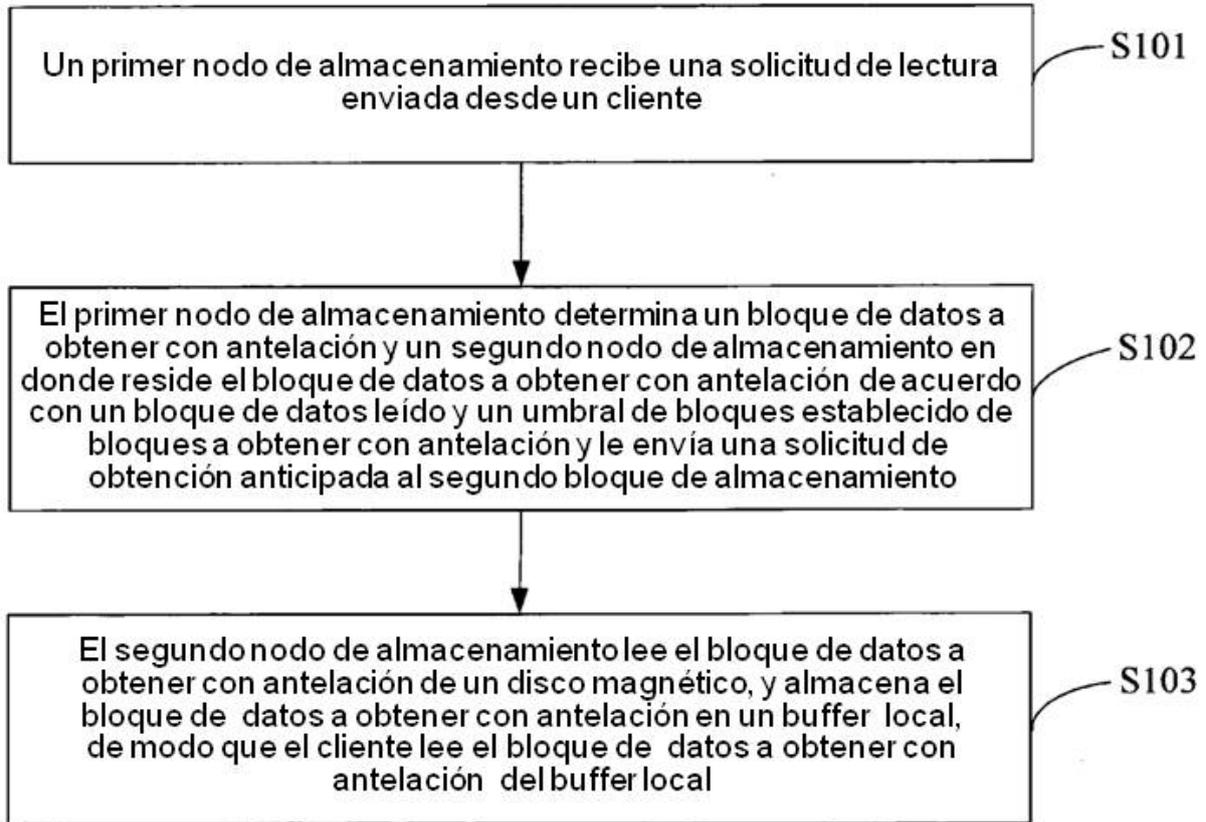


FIG. 1

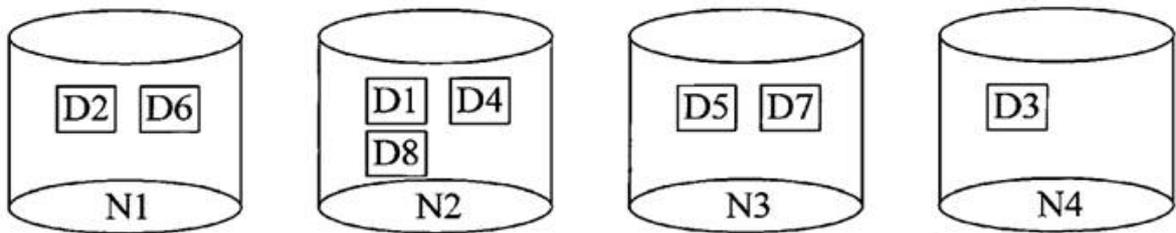


FIG. 2

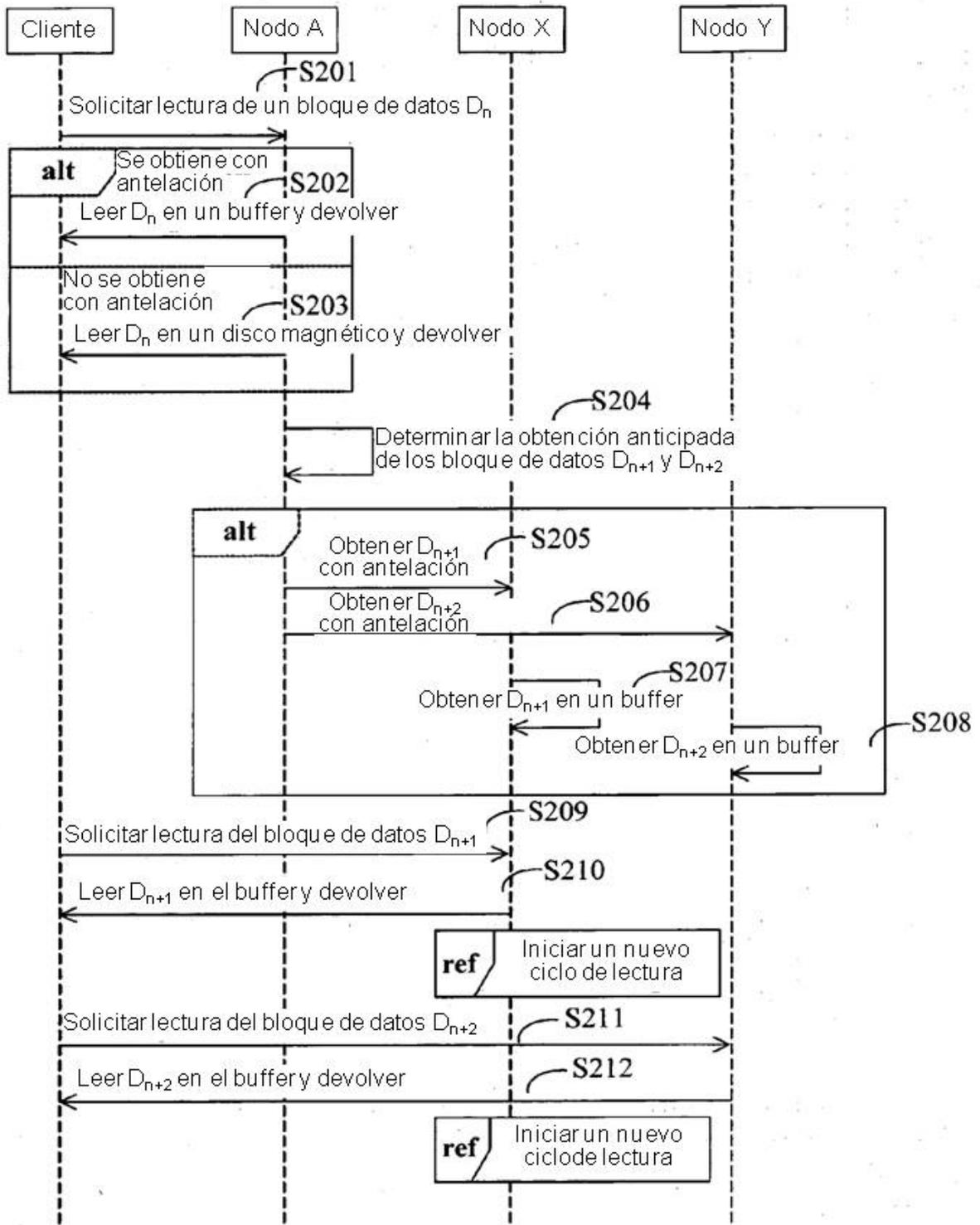


FIG. 3

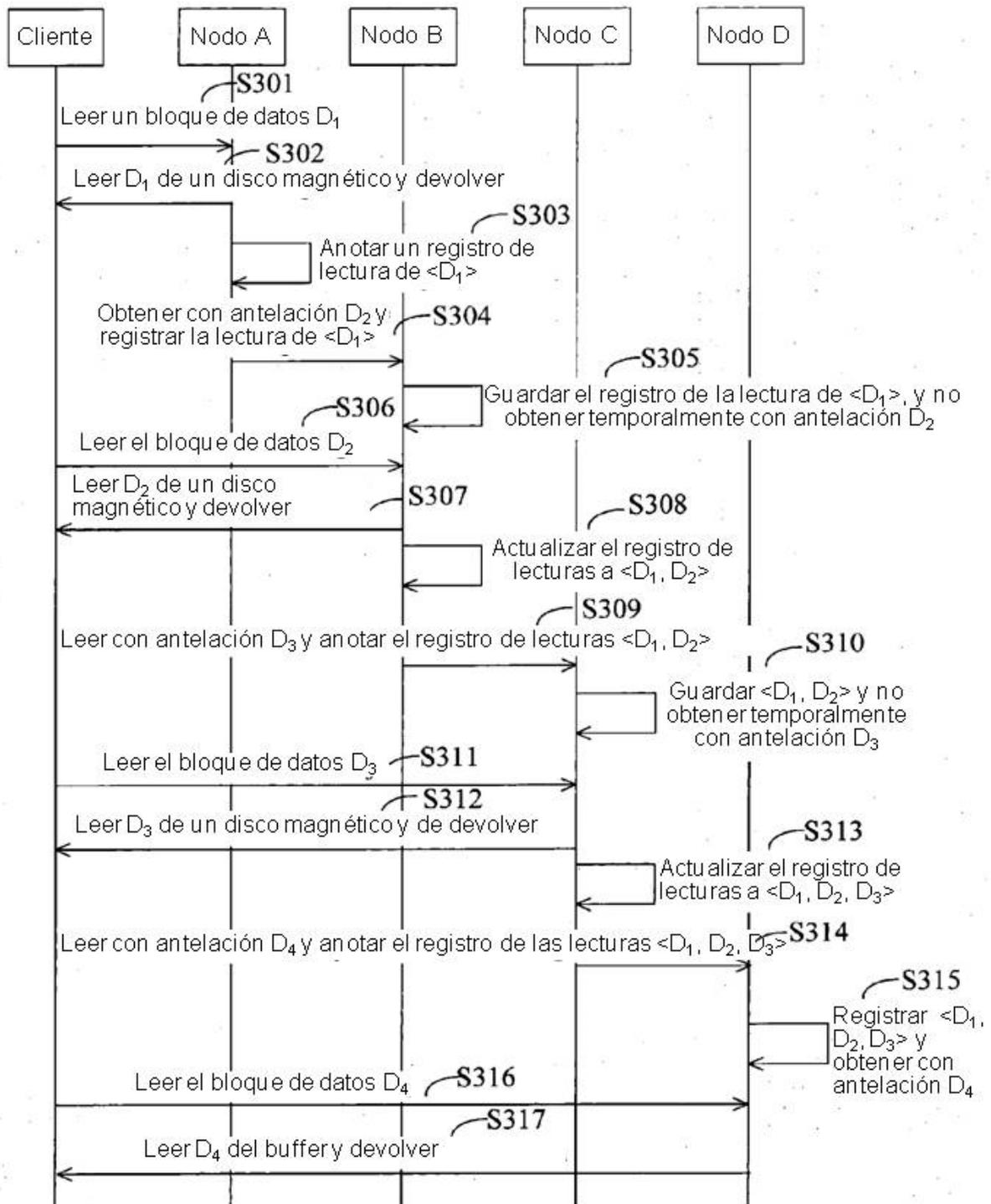


FIG. 4

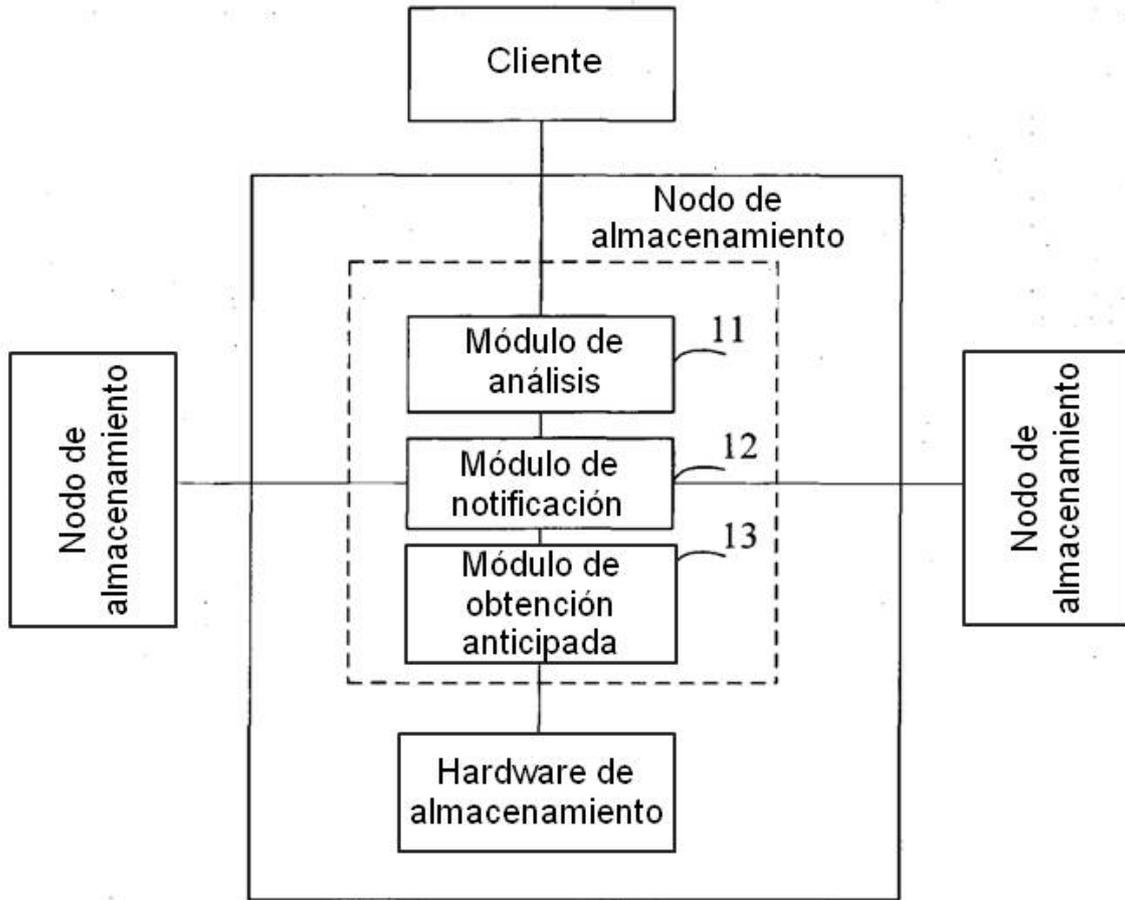


FIG. 5

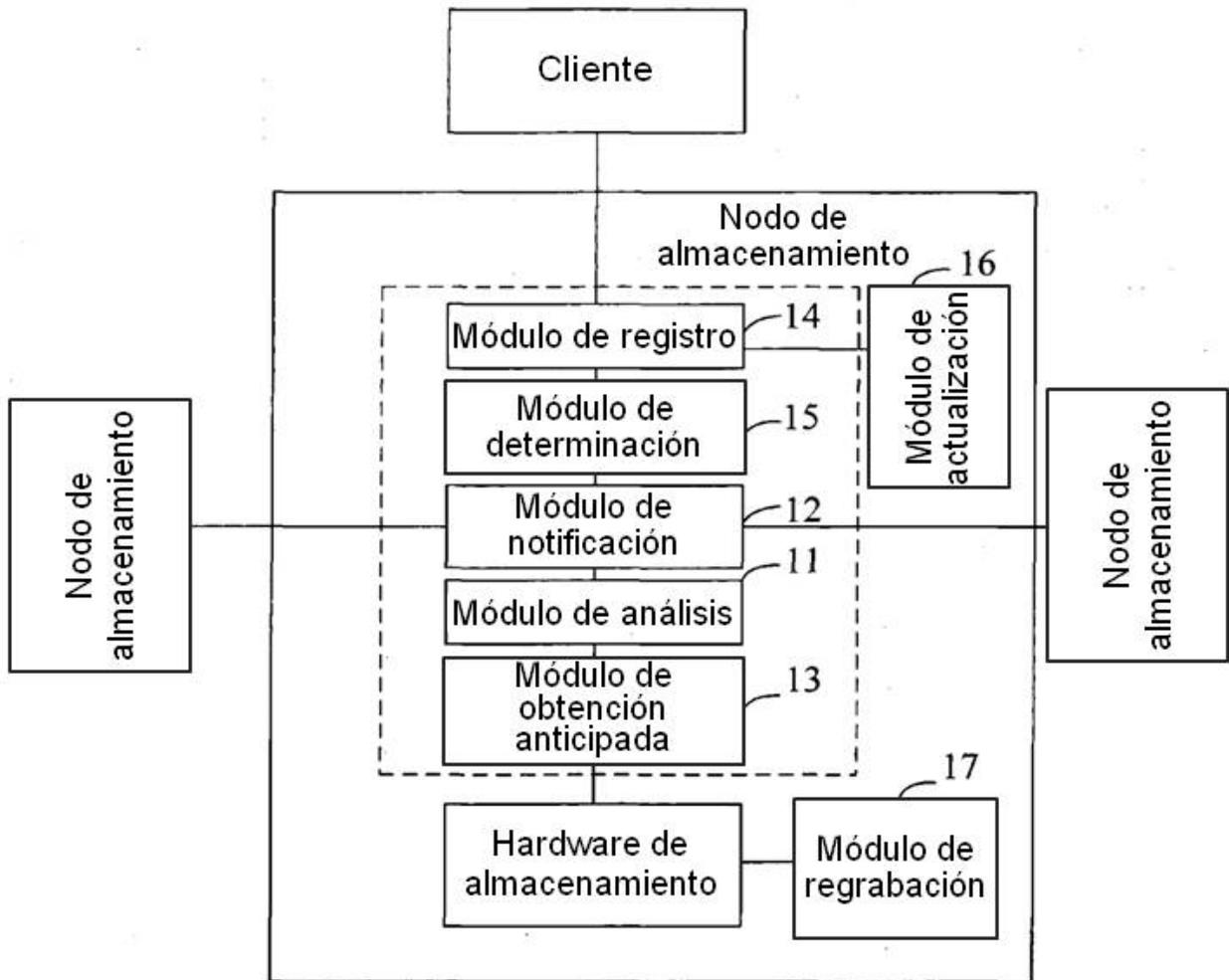


FIG. 6

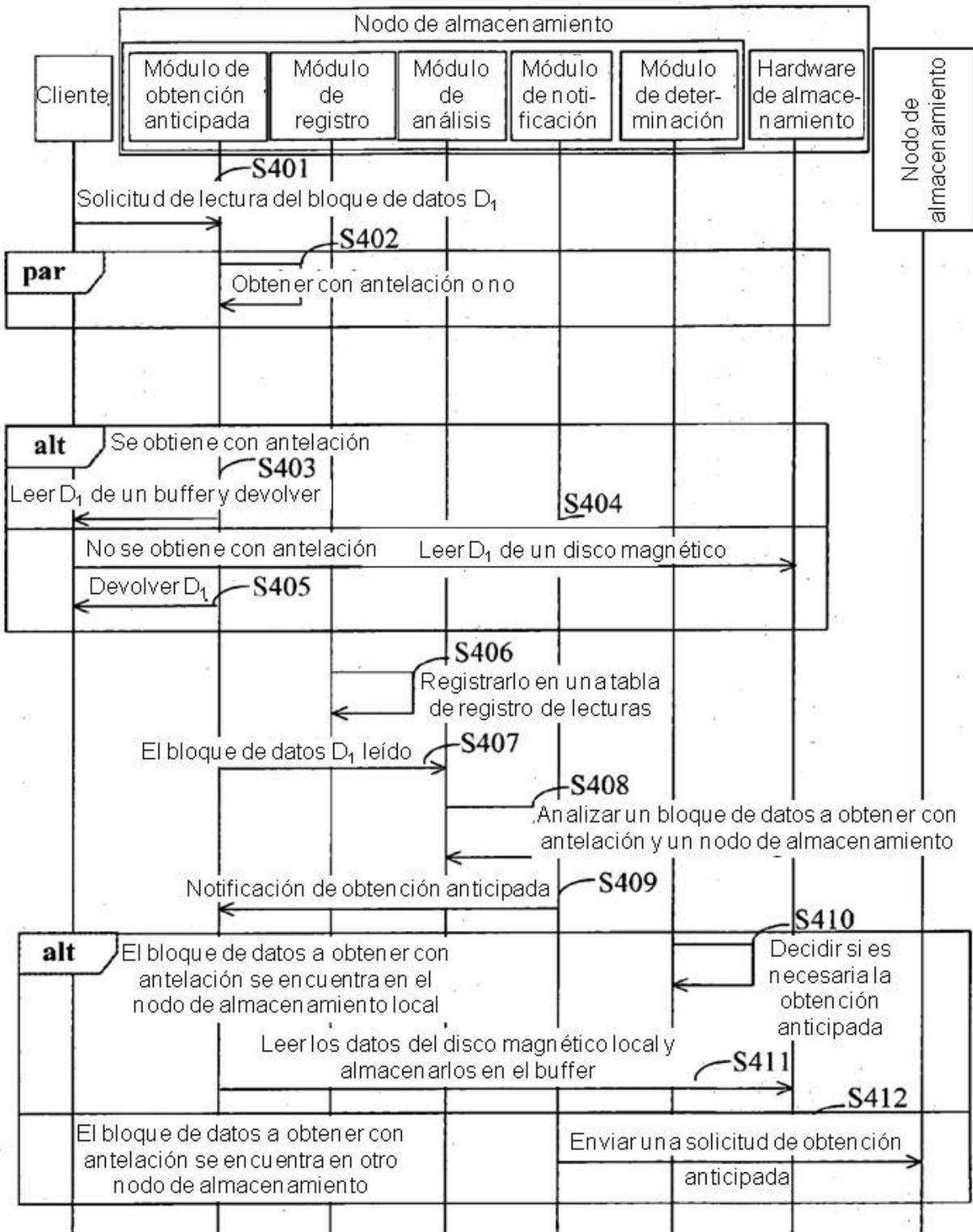


FIG. 7

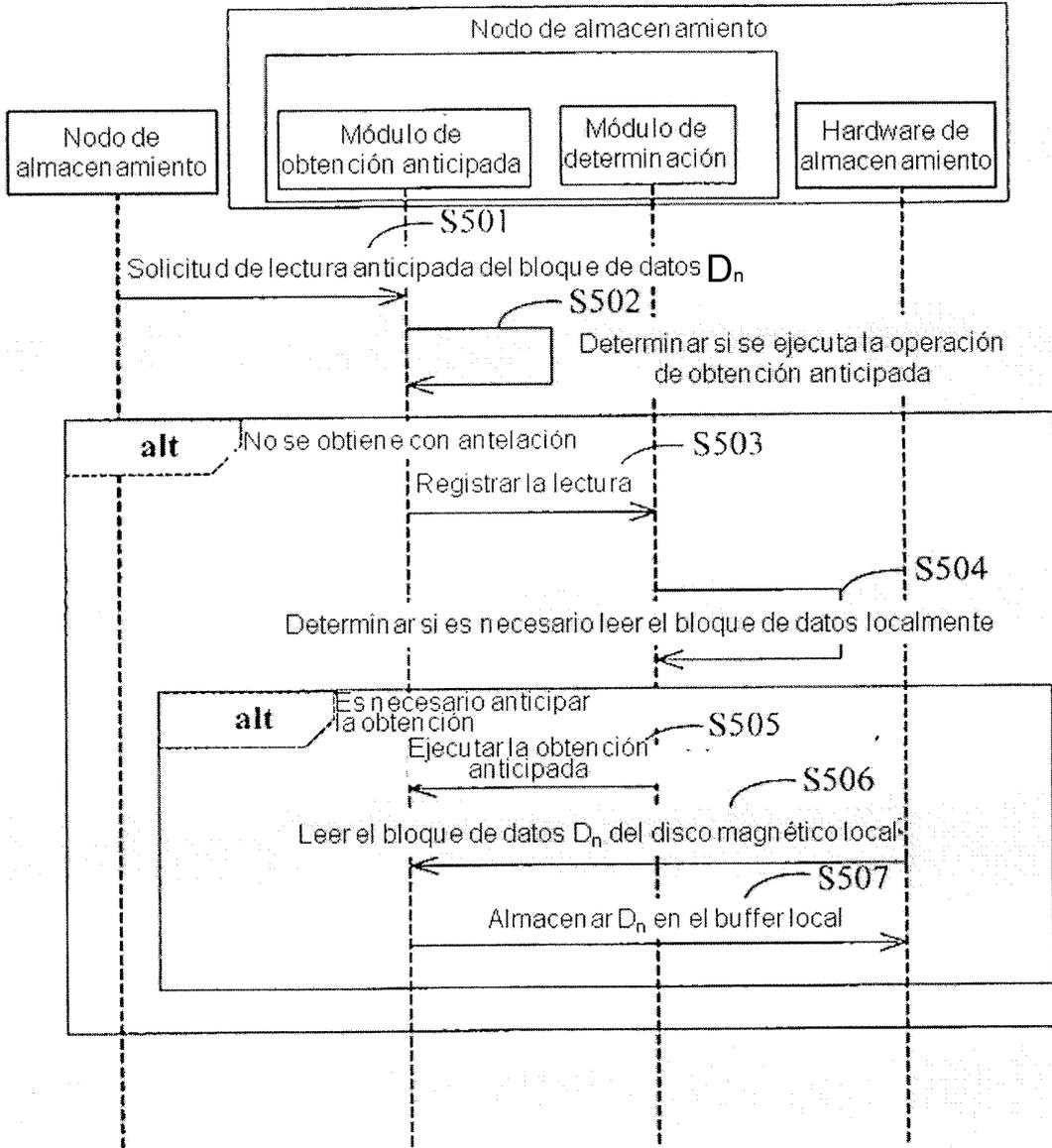


FIG. 8