(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 638 288**

(51) Int. CI.:

G06F 13/16 (2006.01)
G06F 13/10 (2006.01)
G06F 15/167 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **25.03.2011 PCT/US2011/030046**

(87) Fecha y número de publicación internacional: **06.10.2011 WO11123361**

(96) Fecha de presentación y número de la solicitud europea: **25.03.2011 E 11763271 (1)**

(97) Fecha y número de publicación de la concesión europea: **31.05.2017 EP 2553587**

(54) Título: **Asignación de la semántica RDMA con un dispositivo de almacenamiento de alta velocidad**

(30) Prioridad:

18.06.2010 US 818952
02.04.2010 US 320596 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
19.10.2017

(73) Titular/es:

MICROSOFT TECHNOLOGY LICENSING, LLC (100.0%)
One Microsoft Way
Redmond, Washington 98052, US

(72) Inventor/es:

PINKERTON, JAMES, T. y
TALPEY, THOMAS, M.

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

DESCRIPCIÓN

Asignación de la semántica RDMA con un dispositivo de almacenamiento de alta velocidad

Antecedentes

Los ordenadores se han convertido en dispositivos altamente integrados en el trabajo, la vivienda, dispositivos móviles y muchos otros lugares. Los ordenadores pueden procesar cantidades masivas de información de manera rápida y eficiente. Algunas aplicaciones diseñadas para ejecutar sistemas informáticos, permiten que los usuarios lleven a cabo una amplia variedad de funciones incluyendo aplicaciones de negocios, tratamiento escolar, esparcimiento y otras más. Las aplicaciones software a menudo son diseñadas para llevar a cabo tareas específicas, como por ejemplo aplicaciones de procesador de texto para redactar documentos, programas de correos electrónicos para enviar, recibir y organizar correos electrónicos.

En algunos casos, los sistemas informáticos comunican entre sí utilizando diferentes tipos de aplicaciones software. Sin embargo, dicha comunicación interaplicaciones típicamente requiere un procesador para procesar cada salida enviada y cada entrada recibida. Esta carga de trabajo incrementada puede imponer una carga de trabajo excesiva sobre un procesador, especialmente cuando muchos miles o millones de entradas y salidas están siendo enviadas entre diferentes aplicaciones software.

Para evitar dicha implicación de dicho procesador, aunque se sigue manteniendo la comunicación entre las aplicaciones (o entre una aplicación y un almacenamiento de datos distante) puede ser implementado el acceso directo a la memoria distante (RDMA). El RDMA permite leer y escribir solicitudes procedentes de aplicaciones para que fluyan directamente desde la aplicación hasta un dispositivo de almacenamiento de red o hacia un servidor de almacenamiento. La solicitud de escritura o lectura puede ser satisfecha y transferida entre dos ordenadores sin que se requiera la aplicación de uno u otro procesadores del sistema.

El documento US 2010/0083247 A1 introduce un sistema de procesado que incluye una pluralidad de máquinas virtuales que han compartido el acceso a un subsistema de memoria de estado sólido no volátil (NVSSM), utilizando el acceso directo a la memoria distante (RDMA). El subsistema NVSSM puede incluir una memoria flash y otros tipos de memoria de estado sólido no volátil. El sistema de procesado utiliza unas listas de dispersar - agrupar para especificar las operaciones de lectura y escritura del RDMA. Múltiples lecturas o escrituras pueden ser combinadas en una simple lectura y escritura RDMA, respectivamente, las cuales a continuación pueden ser descompuestas y ejecutadas como múltiples lecturas y escrituras, respectivamente, en el sistema NVSSM. Los accesos a la memoria generados por una simple lectura o escritura RDMA pueden ser dirigido hacia dispositivos de memoria diferentes del subsistema NVSSM.

Breve sumario

Es un objeto de la invención mejorar el procesado y la realización de las solicitudes de lectura y escritura en un almacenamiento local de datos.

Este objeto se resuelve mediante la materia objeto de las reivindicaciones independientes.

Formas de realización preferentes se definen en las reivindicaciones dependientes.

Formas de realización descritas en la presente memoria están dirigidas a extender las semánticas del acceso directo de memoria distante (RDMA) para habilitar la implementación de un sistema de almacenamiento local y para proporcionar una interfaz de gestión para inicializar un almacenamiento local de datos. En una forma de realización, un sistema informático extiende las semánticas RDMA para proporcionar el acceso al almacenamiento local utilizando el RDMA, en el que la extensión de las semánticas RDMA incluye lo siguiente: la correspondencia de los verbos RDMA de una interfaz de verbos RDMA con un almacenamiento local de datos y la alteración de las semánticas de orden del RDMA para permitir el procesado fuera de orden y / o las terminaciones fuera de orden. El sistema informático también accede a diversas porciones del almacenamiento local de datos utilizando las semánticas RDMA extendidas.

En otra forma de realización, un sistema informático instancia una interfaz de gestión que está configurada para inicializar un almacenamiento local de datos persistentes. El sistema informático recibe una entrada indicativa de diversas configuraciones que deben ser aplicadas al inicializar el almacenamiento local de datos persistentes, donde al menos una de las configuraciones identifica la forma en que el almacenamiento local de datos persistentes debe ser particionado. El sistema informático particiona el almacenamiento local de datos persistentes en particiones que son cada una accesible utilizando la interfaz de gestión. El sistema informático también inicializa el almacenamiento local de datos persistentes particionados como un punto terminal local del RDMA utilizando la interfaz de gestión.

Este Sumario está previsto para introducir una selección de conceptos de forma simplificada que se describen con mayor amplitud a continuación en la Descripción Detallada. Este Sumario no pretende identificar características clave o características esenciales de la materia objeto reivindicada, ni está concebido para ser utilizado como ayuda en la determinación del alcance de la materia objeto reivindicada.

Características y ventajas adicionales se expondrán en la descripción que sigue, y una parte será evidente a partir de la descripción, o puede ser comprendida mediante la práctica de las enseñanzas que incorpora. Características y ventajas de la invención pueden llevarse a cabo y obtenerse por medio de los instrumentos y combinaciones especialmente señalados en las reivindicaciones adjuntas. Características de la presente invención se pondrán de manifiesto de forma más acabada a partir de la descripción subsecuente o de las reivindicaciones adjuntas, o pueden ser conocidas por la práctica de la invención tal como se define en las líneas que siguen de la presente memoria.

Breve descripción de los dibujos

Para clarificar aún más las ventajas y características expuestas y otras de las formas de realización de la presente invención, se ofrecerá una descripción más concreta de la presente con referencia a los dibujos adjuntos, debe apreciarse que estos dibujos representan solo formas de realización típicas de la invención y, por tanto, no deben ser considerados como limitativos de su alcance. La invención se describirá y analizará con concreción y detalle adicionales mediante el uso de los dibujos que se acompañan, en los cuales:

La Figura 1 ilustra una arquitectura informática en la que formas de realización de la presente invención pueden operar incluyendo la extensión de las semánticas de acceso directo a la memoria distante (RDMA) para habilitar la implantación de un sistema local de almacenamiento.

La Figura 2 ilustra un diagrama de flujo de un procedimiento ejemplar para extender las semánticas del acceso directo a la memoria distante (RDMA) para habilitar la implementación de un sistema local de almacenamiento.

La Figura 3 ilustra un diagrama de flujo de un procedimiento ejemplar para proporcionar una interfaz de gestión para inicializar un almacenamiento local de datos.

La Figura 4 ilustra una forma de realización en la cual un usuario interactúa con un almacenamiento de datos utilizando una interfaz de gestión.

Descripción detallada

Las formas de realización descritas en la presente memoria están dirigidas a la extensión de las semánticas del acceso directo a la memoria distante (RDMA) para habilitar la implementación en un sistema de almacenamiento local y para proporcionar una interfaz de gestión para inicializar un almacenamiento local de datos. En una forma de realización, un sistema informático extiende las semánticas RDMA para proporcionar el acceso de almacenamiento local utilizando el RDMA, de forma que la extensión de las semánticas RDMA incluye lo siguiente: la asignación de los verbos RDMA de una interfaz de verbos RDMA con un almacenamiento local de datos y la alteración de las semánticas de ordenación del RDMA para permitir el procesado fuera de orden y / o las terminaciones fuera de orden. El sistema informático también accede a diversas porciones del almacenamiento local de datos utilizando las semánticas RDMA extendidas.

En otra forma de realización, un sistema informático instancia una interfaz de gestión que está configurada para inicializar un almacenamiento local de datos persistentes. El sistema informático recibe una entrada indicativa de diversas configuraciones que deben ser aplicadas al iniciar el almacenamiento local de datos persistentes, de forma que al menos una de las configuraciones identifique la forma en que debe ser particionado el almacenamiento local de datos persistentes. El sistema informático particiona el almacenamiento local de datos persistentes en particiones cada una de las cuales es accesible utilizando la interfaz de gestión. El sistema informático también inicializa el almacenamiento local de datos persistentes particionados como un punto terminal local del RDMA utilizando la interfaz de gestión.

El análisis siguiente se refiere ahora a la pluralidad de procedimientos y a los actos de procedimiento que pueden ser llevados a cabo. Debe destacarse que, aunque los actos del procedimiento pueden ser analizados en un orden determinado o ilustrado en un diagrama de flujo como sucedidos en un orden concreto, no es necesaria ninguna ordenación concreta que, al menos específicamente, se establezca, o que se requiera porque un acto dependa de otro acto para ser completado antes de que el acto se lleve a cabo.

Formas de realización de la presente invención pueden comprender o utilizar un ordenador de propósito especial o general que incluya un hardware informático como por ejemplo uno o más procesadores y una memoria de sistema, como se analiza con mayor detalle más adelante. Formas de realización dentro del alcance de la presente invención incluyen también medios físicos y otros legibles por ordenador para soportar o almacenar instrucciones y / o estructuras de datos ejecutables por ordenador. Dichos medios legibles por ordenador pueden ser cualquier medio disponible al que se puede acceder mediante un sistema informático de propósito general o especial. Los medios legibles por ordenador que almacenan instrucciones ejecutables por ordenador son medios de almacenamiento informáticos. Los medios legibles por ordenador que portan instrucciones ejecutables por ordenador son medios de transmisión. Así, a modo de ejemplo, y sin limitación, formas de realización de la invención pueden comprender al menos dos tipos claramente diferenciados de medios legibles por ordenador: medios de almacenamiento informático y los medios de transmisión.

Los medios de almacenamiento informático incluyen RAM, ROM, EEPROM, CD-ROM u otro almacenamiento por discos ópticos, almacenamiento por discos magnéticos u otros dispositivos de almacenamiento magnéticos o cualquier otro medio para almacenar medios de código de programa deseados en forma de instrucciones ejecutables por ordenador o de estructuras de datos y a los cuales se pueda acceder mediante un ordenador de propósito general o especial.

Una "red" se define como uno o más enlaces de datos que habilitan el transporte de datos electrónicos entre sistemas informáticos y / o módulos y / o dispositivos electrónicos. Cuando la información es transferida o suministrada a través de una red u otra conexión de comunicaciones (ya sea por cable, inalámbrica o una combinación de por cable o inalámbrica) hacia un ordenador, el ordenador adecuadamente contempla la conexión como un medio de transmisión. Los medios de transmisión pueden incluir una red y / o unos enlaces de datos que pueden ser utilizados para acarrear o bien medios de código de programa deseados en forma de instrucciones ejecutables por ordenador o bien estructuras de datos y a las que se pueda acceder mediante un ordenador de propósito general o especial. También se podrían incluir combinaciones de los anteriores dentro del alcance de los medios legibles por ordenador.

Así mismo, tras alcanzar los diversos componentes del sistema informático, unos medios de código de programa en forma de instrucciones o estructuras de datos ejecutables por ordenador pueden ser transferidos automáticamente desde los medios de transmisión hasta los medios de almacenamiento informáticos (o viceversa). Por ejemplo, las instrucciones o estructuras de datos ejecutables por ordenador recibidas a través de una red o enlace de datos pueden ser almacenadas en memoria intermedia en una RAM dentro de un módulo de interfaz de red (por ejemplo, un "NIC"), y a continuación, en último término, ser transferidos al sistema informático RAM y / o a unos medios de almacenamiento informáticos menos volátiles en un sistema informático. Así, se debe entender que los medios de almacenamiento informáticos pueden ser incluidos en componentes de un sistema informático que también (o incluso fundamentalmente) utilicen medios de transmisión.

Las instrucciones ejecutables por ordenador comprenden, por ejemplo, instrucciones y datos que provocan que un ordenador de propósito general, un ordenador de propósito especial o un dispositivo de procesado de propósito especial lleven a cabo una determinada función o grupo de funciones. Las instrucciones ejecutables por ordenador pueden ser, por ejemplo, instrucciones de formato intermedio, binarias, como por ejemplo lenguaje de ensamblaje o incluso código de fuente. Aunque la materia objeto ha sido descrita en lenguaje específico respecto de características estructurales y / o actos metodológicos, se debe entender que la materia objeto definida en las reivindicaciones adjuntas no están necesariamente limitada a las características o actos descritos anteriormente. Por el contrario, las características y los actos descritos se divulgan como formas ejemplares de implementación de las reivindicaciones.

Los expertos en la materia apreciarán que la invención puede llevarse a la práctica en entornos informáticos de red con muchos tipos de configuraciones de sistemas informáticos incluyendo, ordenadores personales, ordenadores de sobremesa, ordenadores portátiles, procesadores de mensajes, dispositivos de mano, sistemas multiprocesadores, sistemas electrónicos de consumidor programables o basados en microprocesador, PCs de red, miniordenadores, grandes ordenadores, teléfonos móviles, PDAs, dispositivos de teleaviso, encaminadores, conmutadores, y similares. La invención puede también llevarse a la práctica en entornos de sistemas distribuidos en los que los sistemas informáticos distantes y locales, que están enlazados (ya sea por enlaces de datos cableados, enlaces de datos inalámbricos, o bien una combinación de enlaces de datos cableados e inalámbricos) a través de una red, ambos llevan a cabo tareas. En un entorno de sistema distribuido los módulos de programa pueden ser situados tanto en dispositivos de almacenamiento de memoria locales como distantes.

La Figura 1 ilustra una arquitectura 100 informática en la que pueden emplearse los principios de la presente invención. La arquitectura 100 informática incluye una interfaz 110 de gestión. La interfaz 110 de gestión puede ser cualquier tipo de interfaz software que permita que un usuario interactúe con o de cualquier otra forma utilice el almacenamiento 140 local de datos. El almacenamiento 140 local de datos (o simplemente "almacenamiento de datos" en la presente memoria) puede ser cualquier tipo de sistema de almacenamiento de datos que incluya una unidad magnética u óptica, una colección de unidades, una base de datos, una red de área de almacenamiento u otro sistema de almacenamiento. Un usuario 105 puede desear interactuar con el almacenamiento de datos. Dicha interacción puede incluir el almacenamiento o la modificación de los datos almacenados. El usuario puede enviar una solicitud 106 de datos hacia la interfaz 110 de gestión la cual, a continuación, procesa esa solicitud utilizando uno o más de sus módulos internos.

Por ejemplo, la interfaz 110 puede utilizar el módulo de extensión de las semánticas de acceso directo de la memoria distante (RDMA) para extender las semánticas 116 RDMA de manera que estas semánticas puedan ser utilizadas para acceder a los datos dispuestos sobre el almacenamiento 140 de datos. Las semánticas RDMA pueden incluir aquellos comandos, consultas, verbos u otras semánticas que permiten que un dispositivo o programa hable con otro dispositivo o programa utilizando el RDMA. En algunos casos, el módulo 120 de asignación puede ser utilizado para asignar diversos verbos RDMA de una interfaz de verbos RDMA con el almacenamiento local de datos. El módulo 125 de alteración de las semánticas puede ser utilizado para alterar diversas semánticas de ordenación del RDMA para posibilitar que o bien uno u otro o ambos del procesado fuera de orden y de las terminaciones fuera de orden. Dichas asignaciones y alteraciones (extensiones RDMA) pueden permitir que un

usuario o aplicación software acceda al almacenamiento 140 local de datos utilizando las semánticas RDMA extendidas. Estos conceptos se analizarán con mayor detalle más adelante con respecto al procedimiento 200 de la Figura 2.

A la vista de los sistemas y arquitecturas anteriormente descritos, serán mejor apreciadas las metodologías que pueden ser implementadas de acuerdo con la materia objeto divulgada con referencia a los diagramas de flujos de las FIGS. 2 y 3. Para simplificar el análisis, las metodologías se muestran y describen como una serie de bloques. Sin embargo se debe entender y apreciar que la materia objeto reivindicada no está limitada por el orden de los bloques, en cuanto algunos bloques pueden aparecer en órdenes diferentes y / o de manera concurrente con otros bloques a partir de lo que se representa y describe en la presente memoria. Además, no todos los bloques ilustrados pueden ser requeridos para implementar las metodologías descritas a continuación en la presente memoria.

La Figura 2 ilustra un diagrama de flujo de un procedimiento 200 para extender las semánticas de acceso directo de memoria distante (RDMA) para habilitar su implementación en un sistema de almacenamiento local. El procedimiento 200 se describirá a continuación con referencia frecuente a los componentes y datos del entorno 100.

El procedimiento 200 incluye un acto de extensión de una o más semánticas RDMA para proporcionar un acceso de almacenamiento local utilizando el RDMA (acto 210). Por ejemplo, el módulo 115 de extensión de las semánticas RDMA puede extender las semánticas 116 RDMA para permitir el acceso al almacenamiento 140 local de datos utilizando el RDMA. Las semánticas RDMA pueden ser extendidas añadiendo comandos adicionales u otras propiedades que permitan que el RDMA sea utilizado en combinación con el almacenamiento local. En algunos casos, las semánticas RDMA extendidas pueden incluir al menos una semántica que informe de cuándo cada operación se ha completado. Así, cuando, por ejemplo, se ha completado una operación de lectura o escritura, la semántica informa de que la lectura o escritura se ha completado. Pueden utilizarse muchas otras semánticas de este tipo, solas o en combinación con las semánticas anteriormente mencionadas.

Como se indicó anteriormente, el almacenamiento 140 local de datos puede incluir diversos tipos de soluciones de almacenamiento de datos. En algunos casos el almacenamiento 140 local de datos es un dispositivo blanco basado en una red de memoria. En otros casos, el almacenamiento local de datos es algún tipo de almacenamiento basado en una memoria flash. El almacenamiento local de datos puede ser configurado para almacenar en caché diversas porciones de datos para un periodo de tiempo configurable. En algunos casos, las semánticas RDMA extendidas son configuradas para proporcionar acceso distante al almacenamiento local de datos. Así, un usuario 105 puede ser capaz de acceder al almacenamiento 140 local de datos a distancia utilizando las semánticas 116 RDMA extendidas mediante el módulo 115.

En algunos casos, como se indicó anteriormente, las semánticas RDMA extendidas pueden incluir semánticas de interfaz de programación aplicativa (API) de red RDMA (o verbos RDMA) que se han utilizado para que los datos RDMA accedan a través de una red. La API de red RDMA puede extenderse y utilizarse para acceder a los datos almacenados en un almacenamiento local de datos, así como los datos almacenados en un almacenamiento de datos distante. Por consiguiente, se puede utilizar la misma API de red RDMA extendida para acceder tanto a los datos locales como distantes procedentes de los almacenamientos locales y distantes de datos. Dicho de otra forma, se puede acceder a los datos distantes almacenados en un almacenamiento de datos distante utilizando las mismas semánticas RDMA extendidas utilizadas para acceder a los datos locales almacenados en el almacenamiento local de datos. De esta manera, puede ser utilizada una sola API de red RDMA extendida para acceder tanto a los almacenamientos de datos locales como distantes.

El procedimiento 200 incluye un acto de asignación de uno o más verbos RDMA de la interfaz de verbos RDMA con un almacenamiento local de datos (acto 220). Por ejemplo, el módulo 120 de asignación puede asignar uno o más verbos RDMA de una interfaz de verbos RDMA con el almacenamiento 140 local de datos. Los verbos RDMA pueden ser asignados con el almacenamiento local de datos de una forma que permita que un usuario o aplicación software (por ejemplo, 105 o 107, respectivamente) acceda a los datos dispuestos sobre el almacenamiento de datos utilizando los verbos RDMA. La asignación puede también incluir la nominación del almacenamiento local de datos para que el almacenamiento local de datos sea reconocido como un punto terminal por el RDMA. Por consiguiente, cuando el almacenamiento local de datos es reconocido como un punto terminal, el usuario u otras varias aplicaciones software puede acceder y utilizar el almacenamiento de datos como un punto terminal RDMA típico.

El procedimiento 200 incluye un acto de alteración de una o más semánticas de ordenación del RDMA para permitir que al menos un elemento entre el procesado fuera de orden y las terminaciones fuera de orden (acto 230). Por ejemplo, el módulo 125 de alteración de las semánticas puede alterar las semánticas 116 RDMA para posibilitar el procesado fuera de orden y / o las terminaciones fuera de orden. El procesado fuera de orden permite que la interfaz 110 de gestión envíe solicitudes al almacenamiento 140 local de datos y las solicitudes son recibidas (por ejemplo, la solicitud 106 de datos de usuario y la solicitud 108 de datos de aplicación), tras lo cual el almacenamiento local de datos puede responder a las solicitudes fuera de orden. Dado que las solicitudes no tienen que ser respondidas por orden, se permite una mayor flexibilidad al responder a las solicitudes de datos. Las terminaciones fuera de orden también proporcionan mayor flexibilidad en el sentido de que cada operación se puede completar (y la finalización de la señal) fuera de orden.

En algunos casos un usuario o programa de software puede querer (o necesitar) asegurar que se ha completado una lectura, escritura u otra operación. En algunos casos, al utilizar el RDMA, las solicitudes pendientes pueden requerir ser evacuadas por orden para ejecutar una operación (por ejemplo, una operación de lectura). Dichas operaciones pueden ser referidas como operaciones barrera. Así, las operaciones barrera aseguran que una transferencia de datos determinado se ha producido mediante la evacuación de las transferencias de datos pendientes y mediante la afirmación de que las operaciones se han completado. Estrictamente hablando, las operaciones barreras pueden ser utilizadas para completar rápidamente un elemento en lugar de cancelar ese elemento.

En algunas formas de realización, las capacidades de barrera existentes del RDMA se mantienen aunque permitiendo al mismo tiempo el procesado fuera de orden y las terminaciones fuera de orden. Por ejemplo, una fila de espera enviada puede ser tratada fuera de orden aunque manteniendo al tiempo el soporte para una operación de barrera (la operación de finalización de una operación barrera). Una operación de barrera verifica que se han producido uno o más episodios anteriores, mientras que la operación barrera asegura que se ha producido una transferencia de datos determinada. Así, un usuario puede insertar algo dentro de la fila en espera enviada que dependa de todas las operaciones delante de ella que completen, y la operación de barrera asegura que las operaciones de delante han sido completadas.

De forma adicional o alternativa, las operaciones barrera pueden ser tratadas para evacuar las transferencias de datos pendientes y asegurar la finalización de al menos una operación de datos. Además, las garantías de red suministradas por el RDMA pueden ser selectivamente relajadas para su uso en el almacenamiento de datos locales hasta el grado indicado por un usuario. Mediante la relajación de estas garantías, las implementaciones de almacenamiento local que no necesitan necesariamente estas garantías pueden ser implementadas de una manera más eficiente. La relajación de estas garantías también permite que la transferencia de datos se lleven a cabo en paralelo y se completen en paralelo. Estrictamente hablando, se puede conseguir un acceso a datos más rápido a los datos almacenados en el almacenamiento 140 de datos.

El procedimiento 200 incluye un acto de acceso a una o más porciones del almacenamiento local de datos utilizando las semánticas RDMA extendidas (acto 240). Por ejemplo, la interfaz 110 de gestión puede acceder al almacenamiento 140 local de datos utilizando las semánticas RDMA extendidas. Por consiguiente, una vez que han sido implementadas las extensiones semánticas RDMA, el usuario 105 y / o la aplicación 107 pueden acceder al almacenamiento 140 local de datos utilizando el RDMA. En algunos casos, las semánticas RDMA extendidas pueden ser implementadas en una capa de gestión de almacenamiento que incluya un sistema de ficheros y / o una base de datos.

En algunas formas de realización, la aplicación 107 software puede establecer un enlace entre la aplicación y el almacenamiento 140 local de datos utilizando las semánticas RDMA extendidas. La aplicación software puede entonces acceder a al menos una porción de la aplicación de los datos almacenados en el almacenamiento local de datos utilizando el enlace establecido. Después de que el enlace se ha establecido, los datos pueden ser transferidos entre la aplicación software y el almacenamiento local de datos sin una implicación de un procesador. Por consiguiente, el enlace establecido permite la transferencia de datos hacia y desde la aplicación 107 y el almacenamiento 140 local de datos. Esta transferencia de datos puede producirse sin ninguna implicación procedente de un procesador central en el sistema informático sobre el cual la aplicación está siendo ejecutada. Así, los datos pueden ser transferidos directamente desde la memoria de aplicación hasta la memoria de datos (y viceversa) sin implicación del procesador.

La Figura 3 ilustra un diagrama de flujo del procedimiento 300 para proporcionar una interfaz de gestión para inicializar un almacenamiento local de datos. El procedimiento 300 se describirá a continuación con referencia frecuencia a los componentes y datos del entorno 100 de la Figura 1 y al entorno 400 de la Figura 4.

El procedimiento 300 incluye un acto de iniciación de una interfaz de gestión que está configurado para inicializar un almacenamiento local de datos persistentes (acto 310). Por ejemplo, un sistema informático puede instanciar la interfaz 410 de gestión que esté configurada para inicializar el almacenamiento 440 local de datos persistentes. La interfaz de gestión puede ser cualquier tipo de aplicación software que permita el acceso al almacenamiento 440 de datos. En algunos casos, la interfaz de gestión puede ser una interfaz de gestión de una tercera parte suministrada por alguien distinto del proveedor del almacenamiento de datos.

El procedimiento 300 incluye un acto de recepción de una entrada indicativa de una o más configuraciones que deben ser aplicadas al inicializar el almacenamiento local de datos persistentes, en el que al menos una de las configuraciones identifica la manera en que el almacenamiento local de datos persistentes debe ser particionado (acto 320). Por ejemplo, la interfaz 410 de gestión puede recibir una entrada 406 de un usuario 405 que incluya unas configuraciones 130 / 430 de inicialización de almacenamiento de datos. Estas configuraciones deben ser aplicadas al inicializar el almacenamiento 440 de datos. Por ejemplo, las configuraciones 431 de partición pueden indicar en cuántas particiones debe ser particionado el almacenamiento 440 de datos, cómo debe ser el tamaño de cada partición, cómo deben ser configurados los desplazamientos 442 de memoria para cada partición, así como otras configuraciones. En algunas formas de realización, el usuario 405 puede controlar cualquiera o todas las configuraciones 430 de inicialización de almacenamiento de datos. La interfaz 410 de gestión puede también permitir

que un usuario configure las configuraciones 432 de acceso a usuario para cada partición. Así, un usuario puede limitar (asegurar) el acceso a determinadas particiones, garantizando al tiempo un acceso total a otras particiones.

El procedimiento 300 incluye un acto de partición del almacenamiento local de datos persistentes en una o más particiones que pueden ser cada una accesible utilizando la interfaz de gestión (acto 330). Por ejemplo, la interfaz 410 de partición puede particionar (o iniciar con la partición) el almacenamiento 440 local de datos en particiones que sean accesibles utilizando la interfaz 410. La accesibilidad de las configuraciones 432 puede ser aplicada en combinación con la partición para que el acceso a las particiones nuevamente creadas queden limitadas de acuerdo con las configuraciones de accesibilidad.

El procedimiento 300 incluye un acto de inicialización del almacenamiento de datos persistentes particionados como un punto terminal del RDMA local utilizando la interfaz de gestión (acto 340). Por ejemplo, la interfaz 410 de gestión puede inicializar el almacenamiento 440 local de datos persistentes particionados como un punto terminal del RDMA local. Cuando el almacenamiento de datos ha sido inicializado como un punto terminal del RDMA, el usuario puede acceder a los datos del almacenamiento de datos por medio de la interfaz de gestión a través del RDMA. Dicho acceso al punto terminal del RDMA permite que una aplicación transfiera datos directamente desde la memoria de aplicación hasta una o más particiones sobre el almacenamiento de datos sin implicación de un procesador. Además, como se indicó anteriormente, debido a que, al menos en algunas formas de realización, las garantías tradicionales de red RDMA pueden relajarse cuando el RDMA sea utilizado con un almacenamiento local, la transferencia de datos y otras operaciones pueden llevarse a cabo y completarse en paralelo.

Por consiguiente, se han proporcionado procedimientos, sistemas y productos de programa informáticos que extienden las semánticas RMDA para habilitar el acceso a datos y el almacenamiento de datos en un sistema de almacenamiento local utilizando las semánticas RDMA extendidas. Además, se puede disponer una interfaz de gestión que permita que un usuario o aplicación software inicialice un almacenamiento local de datos. El almacenamiento de datos puede ser inicializado de una manera personalizada de acuerdo con cuáles sean las configuraciones de inicialización que el usuario haya establecido.

REIVINDICACIONES

1.- Un procedimiento puesto en práctica por ordenador en un sistema informático que incluye un procesador y una memoria, estando destinado el procedimiento para extender unas semánticas de acceso directo de memoria distante, RDMA, para habilitar la implementación de un sistema de almacenamiento local, comprendiendo el procedimiento:

un acto (210) de extensión de una o más semánticas RDMA para proporcionar acceso al almacenamiento local utilizando el RDMA, en el que la extensión de la semántica RDMA comprende lo siguiente:

un acto (220) de asignación de uno o más verbos RDMA de una interfaz de verbos RDMA con un almacenamiento local de datos; y

un acto (230) de alteración de una o más semánticas de ordenación RDMA para permitir al menos un procesado fuera de orden y unas terminaciones fuera de orden; y

un acto (240) de acceso a una o más porciones del almacenamiento local de datos utilizando las semánticas RDMA extendidas,

comprendiendo además el procedimiento:

un acto de establecimiento por una aplicación software de un enlace entre la aplicación y el almacenamiento local de datos utilizando las semánticas RDMA extendidas, en el que después de que el enlace se ha establecido, los datos son transferidos entre la aplicación software y el almacenamiento local de datos sin ninguna implicación del procesador.

2.- El procedimiento de la reivindicación 1, en el que la asignación incluye la nominación de los almacenamientos locales de datos, de manera que los almacenamientos locales de datos son reconocidos como puntos terminales por el RDMA.

3.- El procedimiento de la reivindicación 1, en el que las capacidades de barrera existentes del RDMA son mantenidas permitiendo al tiempo el procesado fuera de orden y las terminaciones fuera de orden.

4.- El procedimiento de la reivindicación 1, en el que las semánticas RDMA extendidas son implementadas en una capa de gestión de almacenamiento que incluye al menos uno de entre un sistema de ficheros y una base de datos, o en el que las semánticas RDMA extendidas están configuradas para proporcionar un acceso distante al almacenamiento local de datos.

5.- El procedimiento de la reivindicación 1, en el que las semánticas RDMA extendidas incluyen al menos una semántica que informa acerca de cuándo cada operación se ha completado.

6.- El procedimiento de la reivindicación 1, en el que la aplicación software accede al menos a una porción de los datos de aplicación almacenados en el almacenamiento local de datos utilizando el enlace establecido.

7.- El procedimiento de la reivindicación 1, que comprende además un acto de procesado de una o más operaciones barrera para evacuar las transferencias de datos pendientes y asegurar la terminación de al menos una operación de datos.

8.- El procedimiento de la reivindicación 1, en el que el almacenamiento local de datos comprende un dispositivo blanco basado en una red de memoria o en un almacenamiento basado en una memoria flash.

9.- El procedimiento de la reivindicación 1, en el que una o más porciones de datos son almacenadas en caché en el almacenamiento local de datos.

10.- El procedimiento de la reivindicación 1, en el que una o más garantías de red provistas por el RDMA son selectivamente relejadas para su uso en el almacenamiento local de datos hasta el grado indicado por un usuario.

11.- Un producto de programa informático para implementar un procedimiento para proporcionar una interfaz de gestión para inicializar un almacenamiento local de datos, comprendiendo el producto de programa informático uno o más medios de almacenamiento legibles por ordenador en los cuales están almacenadas las instrucciones ejecutadas por ordenador que, cuando son ejecutadas por uno o más procesadores del sistema informático, provocan que el sistema informático lleve a cabo el procedimiento, comprendiendo el procedimiento:

un acto de instanciación de una interfaz de gestión que está configurada para inicializar un almacenamiento local de datos persistentes;

un acto de recepción de una entrada que indica una o varias configuraciones que deben ser aplicadas al inicializar el almacenamiento local de datos persistentes, en el que al menos una de las configuraciones identifica cómo el almacenamiento local de datos persistentes debe ser particionado;

un acto de partición del almacenamiento local de datos persistentes en una o más particiones cada una de las cuales es accesible utilizando la interfaz de gestión; y

un acto de inicialización del almacenamiento local de datos persistentes como un punto terminal local del RDMA utilizando la interfaz de gestión, en el que después de que el almacenamiento local de datos persistentes haya sido inicializado como punto terminal del RDMA, los datos son transferidos entre una aplicación y una o más particiones del almacenamiento local de datos sin ninguna implicación del procesador.

12.- El producto de programa informático de la reivindicación 11, en el que la interfaz de gestión comprende una interfaz de gestión de una tercera parte.

13.- El producto de programa informático de la reivindicación 11, en el que la interfaz de gestión permite que un usuario configure un elemento entre:

una pluralidad de particiones en las que debe particionarse el almacenamiento local de datos persistentes;

unas configuraciones de acceso a usuario para cada partición; y

unos desplazamientos de memoria para cada partición.

14.- Un sistema (100) informático que comprende lo siguiente:

uno o más procesadores;

una memoria del sistema;

un almacenamiento (140) local de datos;

uno o más medios de almacenamiento legibles por ordenador sobre los cuales están almacenadas unas instrucciones ejecutables por ordenador que, cuando son ejecutadas por uno o más procesadores, provoca que el sistema informático lleve a cabo un procedimiento para la extensión de la semánticas de acceso directo a memoria distante, RDMA, para habilitar la implementación en el almacenamiento local de datos, comprendiendo el procedimiento lo siguiente:

un acto de extensión de una o más semánticas RDMA para proporcionar un acceso al almacenamiento local de datos utilizando el RDMA, en el que la extensión de la semánticas RDMA, comprende lo siguiente:

un acto de asignación de una o varias semánticas de interfaz de programación aplicativa, API, de red RDMA de una interfaz de semánticas de API de red RDMA en el almacenamiento local de datos; y

un acto de alteración de una o varias semánticas de ordenación RDMA - para permitir al menos un procesado fuera de orden y unas terminaciones fuera de orden;

un acto de acceso de una o más porciones del almacenamiento local de datos desde el sistema informático distante utilizando las semánticas API de red de RDMA extendidas; y

un acto de acceso a una o más porciones del almacenamiento local de datos desde el sistema informático local utilizando las mismas semánticas API de red RDMA extendidas,

comprendiendo además el procedimiento:

un acto de una aplicación (107) software que establece un enlace entre la aplicación y el almacenamiento local de datos utilizando las semánticas RDMA extendidas, en el que después de que el enlace se ha establecido, los datos son transferidos entre la aplicación software y el almacenamiento local de datos sin ninguna implicación del procesador.

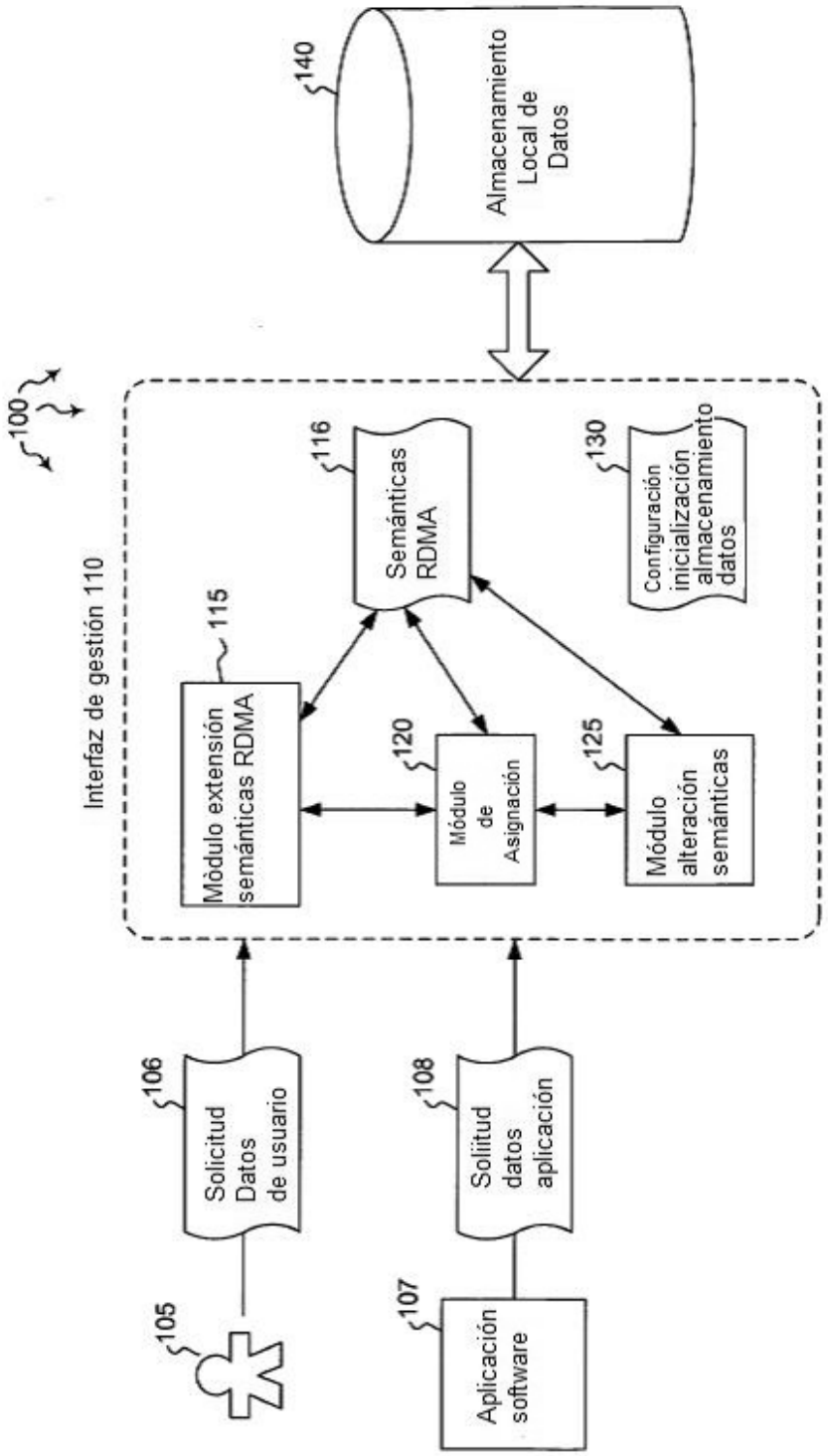

140
Almacenamiento Local de Datos
100
Interfaz de gestión 110
116
Semánticas RDMA
130
Configuración inicialización almacenamiento datos
115
Módulo extensión semánticas RDMA
120
Módulo de Asignación
125
Módulo alteración semánticas
106
Solicitud Datos de usuario
108
Solilitud datos aplicación
105
107
Aplicación software

Figura 1

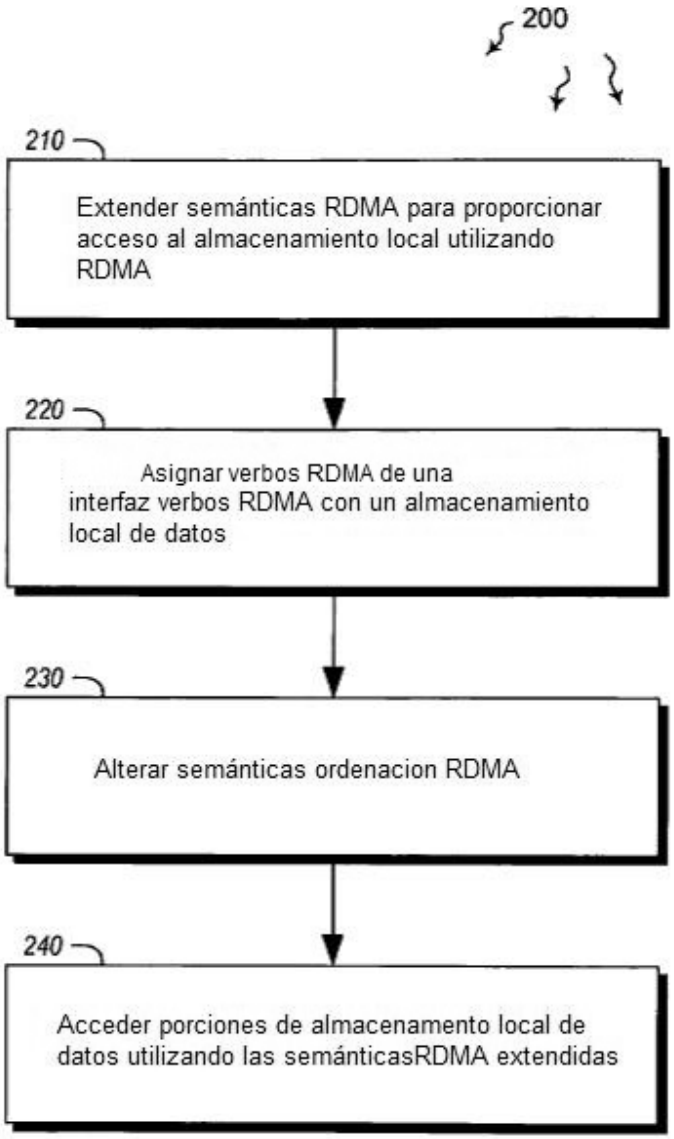
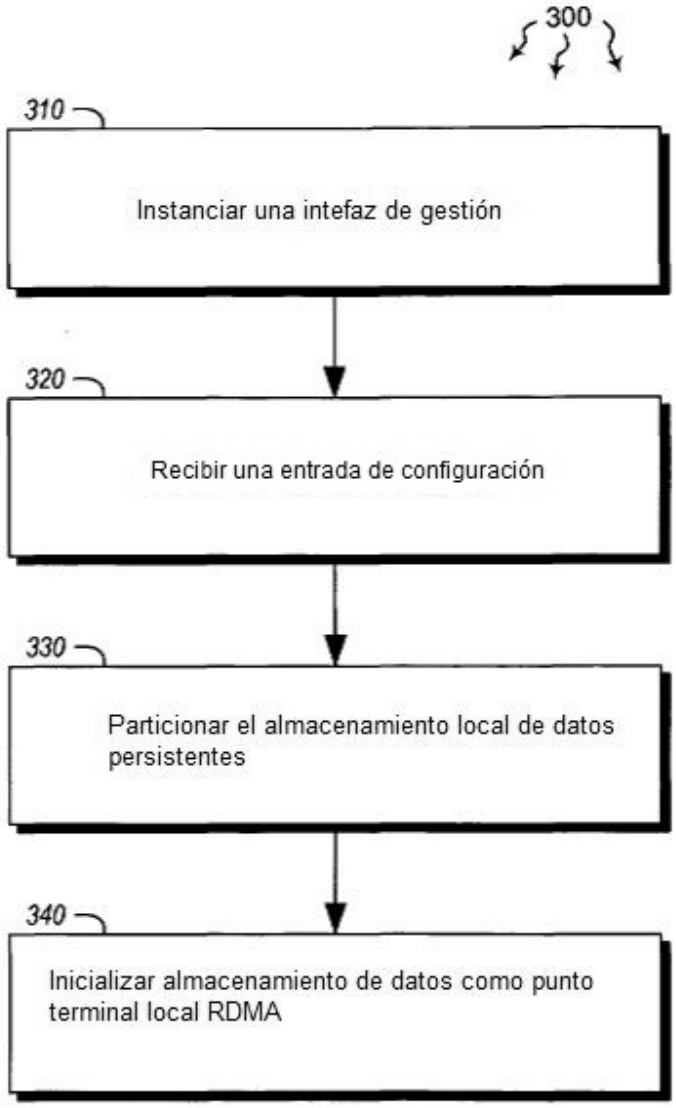

200
210
Extender semánticas RDMA para proporcionar acceso al almacenamiento local utilizando RDMA
220
Asignar verbos RDMA de una interfaz verbos RDMA con un almacenamiento local de datos
230
Alterar semánticas ordenacion RDMA
240
Acceder porciones de almacenamento local de datos utilizando las semánticasRDMA extendidas

Figura 2


300
310
Instanciar una intefaz de gestión
320
Recibir una entrada de configuración
330
Particionar el almacenamiento local de datos persistentes
340
Inicializar almacenamiento de datos como punto terminal local RDMA

Fgura 3

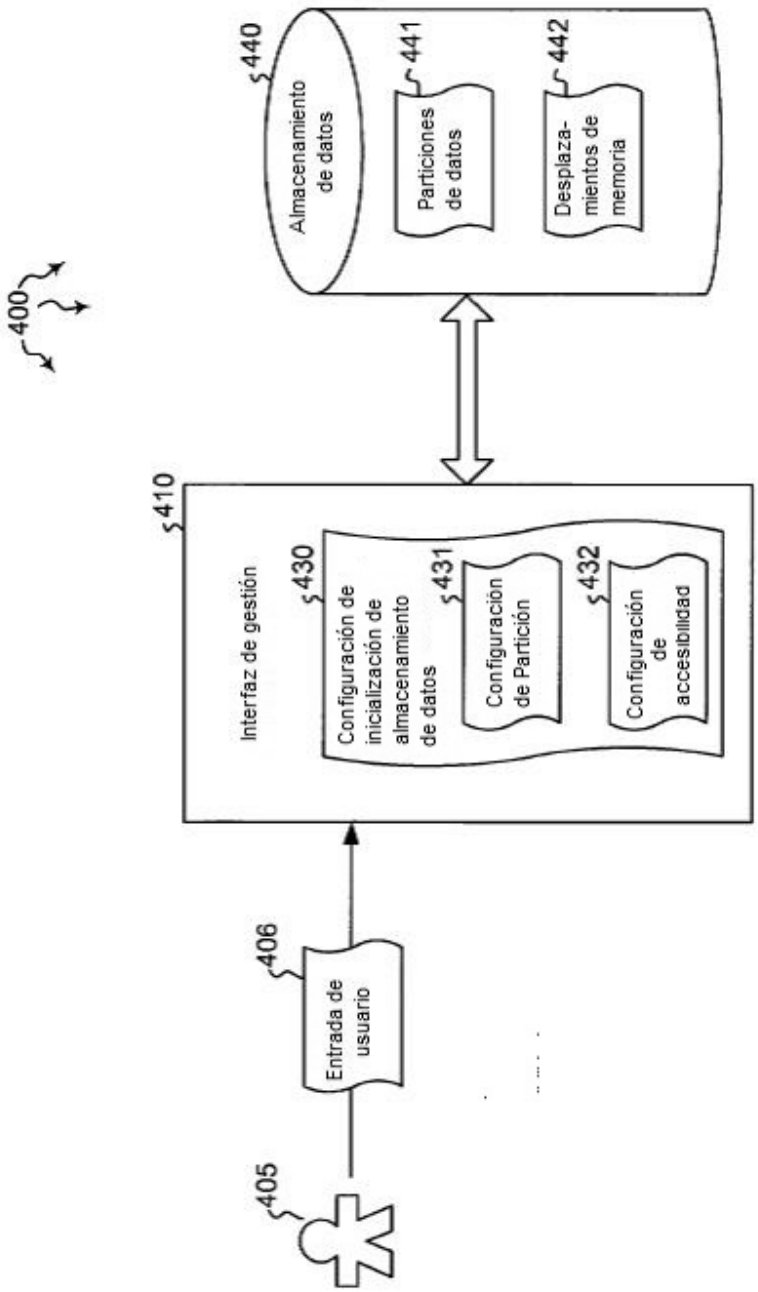

400
440
Almacenamiento de datos
441
Particiones de datos
442
Desplaza-mientos de memoria
410
Interfaz de gestión
430
Configuración de inicialización de almacenamiento de datos
431
Configuración de Partición
432
Configuración de accesibilidad
406
Entrada de usuario
405

Figura 4