

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 696 528**

51 Int. Cl.:

G06F 11/16 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.05.2016** E 16172317 (6)

97 Fecha y número de publicación de la concesión europea: **15.08.2018** EP 3252606

54 Título: **Métodos para la sincronización del sistema de manejo de datos independiente**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
16.01.2019

73 Titular/es:

**IG KNOWHOW LIMITED (100.0%)
Cannon Bridge House 25 Dowgate Hill
London EC4R 2YA, GB**

72 Inventor/es:

**LEVINE, BORIS;
SIMMONDS, DAVID JOHN y
CLOKE, JAMIE ROLAND**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 696 528 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

Métodos para la sincronización del sistema de manejo de datos independiente

Campo técnico

5 Esta invención se refiere a sistemas y métodos para la sincronización de sistemas o disposiciones de manejo o procesamiento de datos independientes, en particular sistemas o disposiciones de sincronización que manejan, procesan y/o almacenan grandes cantidades de datos desde flujos de datos, y en los que los sistemas independientes procesan datos casi idénticos.

Antecedentes

10 La generación, procesamiento, transmisión y recepción de flujos de datos e información son bien conocidos. Diversos sistemas y procesos utilizan grandes flujos de datos o señales de datos para transmitir datos a través de un enlace de datos, bus u otra capa física de una ubicación a otra. Por ejemplo, los sistemas de telecomunicaciones, los sistemas de medios de difusión y sistemas de ordenador en red utilizan dichas técnicas.

15 La redundancia y capacidad de recuperación en estos sistemas y procesos es un problema considerado anteriormente. Por ejemplo, pueden estar duplicadas las partes de un sistema de manejo y procesamiento de datos. Se sabe que el contenido de un servidor se puede duplicar al copiar su contenido en otro servidor, normalmente en otra ubicación física.

20 Un sistema (100) de duplicación básico se ilustra en la Figura 1. Los datos se proporcionan por las fuentes de datos; el dibujo ilustra diversas fuentes 1-3 (102, 104, 106) a 'n' (108). Los datos de estas fuentes de datos se recopilan, compilan, multiplexan o se juntan para formar un flujo (110) de datos. El flujo de datos se proporciona luego al sistema (112) principal en una primera ubicación (Ubicación 1), para el último uso de datos normal o la operación de procesamiento que se requiera. Con el fin de proporcionar redundancia para el sistema principal (por ejemplo, un servidor de datos), el flujo de datos o los datos procesados por el sistema principal se pueden pasar a un sistema (116) de duplicación. El sistema de duplicación, por ejemplo, un servidor de duplicación, se encuentra comúnmente en una ubicación diferente (Ubicación 2), por lo que la transferencia se realiza a través de una red (114) de comunicaciones. Por lo tanto, el sistema de duplicación puede alojar los mismos datos (o datos procesados) que los existentes en el sistema principal. Esto proporciona la ventaja de que, si el sistema principal funciona mal, el otro sistema puede proporcionar el contenido de datos requerido en su lugar.

30 Una desventaja de dichas disposiciones de duplicación es que el contenido del sistema de duplicación nunca es inherentemente exactamente igual al del sistema principal, debido a la latencia en la transferencia del sistema principal al sistema de duplicación a través de la red de comunicaciones; los datos en el sistema de duplicación son, por lo tanto, una versión demorada o incompleta de aquellos en el sistema principal. Este tipo de disposición también requiere que todo el conjunto de datos se transfiera o copie desde un sistema hasta otro; para los sistemas que procesan grandes cantidades de datos, esto requiere un gran ancho de banda en el canal que transfiere los datos y/o largos periodos para que se transfiera una determinada cantidad de datos. Estas desventajas surgen en parte porque el sistema de duplicación simplemente contiene una copia del contenido del sistema principal, que normalmente se envía al sistema de duplicación desde el sistema principal como se ilustra en este documento; no hay un enlace directo entre las fuentes 102-108 (o el flujo 110 de datos) y el sistema de duplicación.

40 La Figura 2 ilustra otro sistema (200) considerado anteriormente, en el que los flujos 1 y 2 (210, 226) de datos separados se compilan a partir de cargas o enlaces desde las mismas fuentes (202-208) de datos. Normalmente, los datos de las fuentes de datos se procesan para generar los flujos de datos. Los flujos de datos separados se generan en las dos ubicaciones separadas, Ubicación 1 y Ubicación 2, y finalmente se almacenan (220, 222, 236, 238) en esas ubicaciones separadas. Por lo tanto, se intenta proporcionar la redundancia en las dos ubicaciones separadas en una etapa anterior.

45 Dado que los flujos de datos ahora están en ubicaciones separadas, ahora es posible determinar en una etapa temprana si hay algún problema con el flujo de datos; por ejemplo, puede haber un problema durante la generación del flujo, o puede haber un retardo en la carga desde la fuente 1 (202) al flujo 1 (210) de datos, que no aparece en la carga desde la fuente 1 hasta el flujo 2 (226) de datos. Esto contrasta con el sistema en la Figura 1, en el que cualquier problema con el flujo (110) de datos será inherente a los datos en el sistema principal o en el sistema de duplicación; el sistema de duplicación simplemente proporciona una copia de seguridad en caso de que el sistema principal falle.

55 Cada flujo de datos se pasa a un detector (212, 230) de problemas y a un sistema de almacenamiento que archiva el flujo (214) para uso posterior si es necesario. Si no se detecta ningún problema con el flujo, los datos simplemente omiten el siguiente bloque (216, 232) y se almacenan (220), para el uso o proceso posterior para el cual se requieren los datos del flujo de datos. Sin embargo, si hay un problema con el flujo de datos, entonces el sistema de procesamiento se debe cerrar para que se pueda solucionar el problema; la producción y el almacenamiento continuado del flujo son de poco uso si el flujo es ahora problemático (por ejemplo, contenido incorrectamente asignado o retrasado, datos corruptos o incompletos).

- 5 Por lo tanto, los datos se pasan a un generador (218) de flujo de reemplazo. Este toma el contenido (214, 228) archivado del otro flujo de datos, que se ha transferido desde la otra Ubicación a través de una red (224) de comunicaciones, y construye un flujo de datos corregido, que luego puede almacenarse como una versión alternativa o corregida del flujo (222) de datos. Esta versión, por supuesto, se retrasa necesariamente al tener que
- 10 detener la generación/procesamiento del flujo de datos, esperar la transferencia de los datos archivados desde la otra Ubicación y reconstruir el flujo de datos. Normalmente, la cantidad de datos reemplazados es grande, ya que puede ser significativa la demora entre el problema que se detecta, el flujo que se apaga, el contenido archivado que se recibe desde la otra ubicación y la reconstrucción del flujo hasta la hora actual (o lo más cerca posible). Por lo tanto, este sistema nuevamente requiere una gran transferencia de datos a través de la red de comunicaciones.
- 15 Otra desventaja de dicho sistema es que es altamente complejo y, por lo tanto, requiere una gran cantidad de infraestructura y/o capacidad de procesamiento computacional. Una desventaja adicional, comparada, por ejemplo, con el sistema de duplicación, es que este sistema necesita determinar si o no existe un problema específico con el flujo de datos, con el fin de proporcionar la redundancia.
- Una desventaja adicional es que puede haber casos en que ambos flujos de datos sean parcialmente inválidos, por ejemplo, cuando solo un subconjunto de uno sea inválido o inexacto, pero un subconjunto diferente del otro también sea inválido. Esto significa que incluso si el flujo 1 se reconstruye con el flujo 2, la combinación resultante también será necesariamente parcialmente inválida.
- 20 Adicionalmente, sin embargo, por muy precisa que sea la reconstrucción de los flujos en este caso o por el buen tiempo en que estén cronometrados, e incluso si los flujos no experimentan problemas y rara vez se reconstruyen, este sistema no obstante tendrá dificultades para sincronizar los dos (a cuatro) sistemas (220, 222 y 236, 238) de datos almacenados entre sí. Es decir, será difícil mantener la paridad entre los conjuntos de datos en las diferentes Ubicaciones y garantizar que sean precisos para los mismos datos más recientes. Esto será especialmente evidente en situaciones en las que las cargas de datos son de alta velocidad o alta densidad, por lo que se esperan grandes cantidades de elementos de datos cada segundo.
- 25 Dado que los flujos se procesan a partir de la entrada de las mismas fuentes, serán muy similares, pero generalmente no son idénticos. Puede ser, por ejemplo, que una diferencia entre los dos flujos de datos no sea detectable por un detector (212, 230) de problemas de flujo de datos; por ejemplo, puede ser que los datos recibidos de las fuentes de datos para procesarlos en los flujos de datos se reciban en diferentes órdenes, debido a la separación de las Ubicaciones 1 y 2. Esto significa que puede que no haya un problema detectable per se en
- 30 ninguno flujo, pero no obstante, que los flujos serán diferentes, creando un problema de sincronización potencial.
- Por ejemplo, un evento puede ser grabado o monitorizado por ambas fuentes 1 y 2 (202, 204) de datos, siendo la primera de las fuentes de datos la que se va a recibir en un flujo de datos que es grabado en ese flujo. Durante la monitorización, la fuente 1 de datos cambia su informe del evento desde un primer resultado o tipo de datos hasta un segundo resultado de datos, y luego hasta un tercero. En contraste, la fuente 2 de datos cambia inmediatamente desde el primer resultado hasta el tercer resultado. Si, en el flujo 1 de datos, la carga de la fuente 1 se recibe momentáneamente antes que la fuente 2, los datos se anotarán en el flujo 1 de datos, que cambiará del primer resultado al segundo y del segundo al tercer resultado. Sin embargo, si, en el flujo 2 de datos, la fuente 2 se recibe primero, se puede observar en el flujo 2 de datos que los datos cambiaron del primer resultado al tercero, sin encontrar nunca el segundo resultado de datos. Por lo tanto, los flujos diferirán sustancialmente, sin que haya ningún
- 35 problema identificable inherente dentro de ninguno de ellos.
- Una desventaja adicional con esta disposición es que probablemente se requeriría una implementación práctica del sistema para combinar o copiar entre los datos resultantes en el almacenamiento de datos en cada ubicación, es decir, entre el almacenamiento habitual del flujo (220) y el almacenamiento del flujo (222) reconstruido y retardado. Esto sería para que los dos sistemas de almacenamiento estén sincronizados, para que no entren en conflicto, o para que se pueda generar un único conjunto de datos a partir de ellos. Esta es entonces una gran carga computacional adicional que implica una gran transferencia de datos, requiere una gran cantidad de almacenamiento de datos adicional, y también retrasa aún más el uso del conjunto final de datos.
- 40
- 45 El documento US 2002/0077782 describe una unidad de microcontrolador (MCU) que tiene una unidad de procesamiento primaria o principal, una unidad de procesamiento secundaria acoplada a la unidad de procesamiento primaria y una memoria común acoplada a las unidades de procesamiento primaria y secundaria. Se proporciona un método para detectar un fallo en la MCU que incluye las etapas de leer un algoritmo de control almacenado en la memoria común por la unidad de procesamiento primaria, leer el algoritmo de control almacenado en la memoria común por la unidad de procesamiento secundaria, comparar la salida primaria y la salida secundaria y detectar sensiblemente un fallo, si la salida primaria no coincide con la segunda salida.
- 50
- 55 El documento WO 01/80009 describe un sistema de ordenador tolerante a fallos que incluye una primera y una segunda unidades de procesamiento central (CPU) que producen flujos de salida de datos esencialmente idénticos, un búfer de retardo de dispositivo de toma de decisiones que tiene un primer búfer FIFO y un segundo búfer FIFO, y un módulo I/O conectado a las CPU. El módulo I/O incluye un comparador para la comparación de bits de los flujos de salida de datos de la CPU. El primer flujo de salida de datos de la CPU se transmite a los dispositivos periféricos

si ambas salidas de la CPU siguen siendo sustancialmente idénticas. De lo contrario, si el comparador indica diferencias, los datos de la primera y segunda CPU en cola se enrutan a los primero y segundo FIFO, respectivamente, y los datos subsiguientes se conservan en los respectivos búferes de la CPU. Mientras las CPU continúan procesándose, los procedimientos de diagnóstico en curso intentan identificar a una u otra CPU que funciona mal y que la CPU restante funciona correctamente. Si el diagnóstico resultante no es concluyente, la CPU que tiene la menor tasa de corrección de errores se identifica como que funciona correctamente. En cualquier caso, la salida en búfer y el flujo de salida de datos procesados posteriormente desde la CPU que funciona correctamente después de esto se transmiten a los dispositivos periféricos.

El documento GB 2122393 describe un sistema de procesamiento de datos múltiple que incluye primero y segundo microprocesadores para ejecutar el mismo procesamiento de datos de forma paralela. Un selector suministra selectivamente uno de los datos de salida de los microprocesadores a un dispositivo de I/O externo. Un comparador detecta un estado de procesamiento de datos fuera de lo normal al comparar los datos de salida del primer microprocesador con los datos de salida del segundo microprocesador. Un monitor de operación evalúa el estado de operación de los microprocesadores en respuesta a una señal del comparador e indica al selector a que suministre datos de salida enviados desde un canal normal hasta dispositivo I/O externo.

La presente invención pretende abordar o por lo menos reducir sustancialmente los problemas descritos anteriormente con los sistemas de la técnica anterior.

Resumen de la invención

Los aspectos y realizaciones de la invención se establecen en las reivindicaciones adjuntas.

De acuerdo con un aspecto de la presente invención, que se define en detalle en las reivindicaciones 1 y 12 independientes adjuntas, se proporciona un sistema para sincronización de primer y segundo sistemas de manejo de datos independientes, cada uno de los primero y segundo sistemas de manejo de datos que recibe datos desde la misma pluralidad de proveedores de datos, el primer y segundo sistemas de manejo de datos que procesan primer y segundo flujos de datos generados desde los datos recibidos desde la pluralidad de proveedores de datos, en la que el primer y segundo sistemas de manejo de datos se disponen en diferentes ubicaciones, y en el que el primer y segundo flujos de datos no son exactamente idénticos, pero son lo suficientemente idénticos para soportar el propósito de la redundancia de proporcionar los dos flujos de datos separados en las dos ubicaciones, el sistema comprende: a) un primer sistema de manejo de datos en una primera ubicación que comprende: un primer módulo de generación de flujo de datos, que comprende un primer compilador para combinar datos recibidos desde la pluralidad de proveedores de datos en a primer flujo de datos; y un primer sistema de procesamiento y almacenamiento, que comprende: un primer almacén de datos para almacenar el primer flujo de datos; un primer módulo duplicador para duplicar datos desde el primer flujo de datos; un primer módulo combinador para combinar datos copiados desde el segundo flujo de datos con el primer flujo de datos; y por lo menos un primer puerto de comunicaciones para transmitir información y/o datos a y recibir información y/o datos desde una red de comunicaciones; b) un segundo sistema de manejo de datos en una segunda ubicación remota desde la primera ubicación, que comprende: un segundo módulo de generación de flujo de datos, que comprende un segundo compilador para combinar datos recibidos desde la pluralidad de proveedores de datos en un segundo flujo de datos; y un segundo sistema de procesamiento y almacenamiento, que comprende: un segundo almacén de datos para almacenar el segundo flujo de datos; un segundo módulo duplicador para duplicar datos desde el segundo flujo de datos; un segundo módulo combinador para combinar datos copiados desde el primer flujo de datos con el segundo flujo de datos; y por lo menos un segundo puerto de comunicaciones para transmitir información y/o datos a y recibir información y/o datos desde una red de comunicaciones; y c) por lo menos un módulo compactador para comparar la información desde el primer y segundo flujos de datos, en el que el primer y segundo sistemas de procesamiento y almacenamiento se configuran para, a través de por lo menos un primer y segundo puertos de comunicaciones, transmitir información a y desde por lo menos un módulo compactador, y para transmitir datos entre el primer y segundo sistemas de procesamiento y almacenamiento, a través de la red de comunicaciones, y en el que: i) el primer y segundo módulos de generación de flujo de datos se configuran para generar primer y segundo flujos de datos para procesamiento y almacenamiento en los respectivos primer y segundo sistemas de procesamiento y almacenamiento; ii) luego de recepción del primer y segundo flujos de datos en el primer y segundo sistemas de procesamiento y almacenamiento, el primer sistema de procesamiento y almacenamiento se configura para transmitir primera información desde una sección más reciente del primer flujo de datos hasta el módulo compactador, y el segundo sistema de procesamiento y almacenamiento se configura para transmitir segunda información desde una sección más reciente correspondiente del segundo flujo de datos hasta el módulo compactador; iii) el módulo compactador se configura para: comparar la primera y segunda información; medir un valor para discrepancia entre la primera y segunda información; probar el valor de discrepancia contra un umbral; y, en el que el valor de discrepancia excede el umbral, generar: desde la primera y segunda información un indicador del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contiene una mayor calidad de datos; y una instrucción de duplicación para el módulo duplicador del respectivo sistema de procesamiento y almacenamiento cuyo flujo de datos se ha indicado por el indicador para contener la mayor calidad de datos; iv) el módulo duplicador para el flujo de datos indicado se configura para, si se indica, duplicar los datos desde dicha sección más reciente del flujo de datos indicado; v) el sistema de procesamiento y almacenamiento para el flujo de datos indicado se configura para transmitir, a través del puerto de

comunicaciones, los datos duplicados al módulo combinador del sistema de procesamiento y almacenamiento para el otro flujo; y vi) el módulo combinador para el otro flujo se configura para reemplazar dicha sección más reciente correspondiente del otro flujo con los datos duplicados desde dicha sección más reciente del flujo indicado.

- 5 En una realización, el módulo compactador se configura para, luego de las etapas de comparación, prueba, medición y cualquier etapa de generación, insertar en cada uno del primer y segundo flujos de datos un marcador (606) de punto de tiempo que indica el fin de la sección actual de los flujos que se comparan, y el primer y segundo sistemas de procesamiento y almacenamiento se configuran para definir dicha sección más reciente del respectivo flujo de datos entre un marcador de punto de tiempo más reciente en el respectivo flujo de datos y un punto de tiempo actual.
- 10 De forma adecuada, la primera información desde la sección más reciente del primer flujo de datos, y la segunda información desde la sección más reciente correspondiente del segundo flujo de datos, transmitida hasta el módulo compactador, comprende un subconjunto de los datos contenidos en aquellas secciones de los respectivos flujos de datos.
- 15 En realizaciones, el primer y segundo flujos de datos se subdividen cada uno en una pluralidad de canales (602) de datos, en la que la subdivisión en el primer flujo de datos corresponde a la subdivisión en el segundo flujo de datos, y dicha sección más reciente del primer flujo de datos es una sección más reciente de un primer canal de datos de la pluralidad de canales del primer flujo de datos, y dicha sección más reciente correspondiente del segundo flujo de datos es una sección más reciente correspondiente de un primer canal de datos correspondiente de la correspondiente pluralidad de canales del segundo flujo de datos.
- 20 Preferiblemente, después de las etapas (ii) a (vi), la siguiente sección más reciente del primer flujo de datos para comparación es la sección más reciente de un segundo canal de datos de la pluralidad de canales de datos del primer flujo de datos, y la siguiente sección más reciente correspondiente del segundo flujo de datos para comparación es la sección más reciente correspondiente de un segundo canal de datos correspondiente de la correspondiente pluralidad de canales del segundo flujo de datos.
- 25 En realizaciones, el módulo compactador se configura para comparar la primera y segunda información, medir el valor para discrepancia, y probar el valor contra el umbral, utilizando una jerarquía de secuencias de filtros de discrepancia, la jerarquía de secuencias de filtros de discrepancia que procede desde una comparación de discrepancia gruesa hasta una comparación de discrepancia fina.
- 30 De forma adecuada, el módulo compactador se configura, para generación del indicador (508) del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contiene la mayor calidad de datos: para comparar dicha primera y segunda información utilizando cada una de una pluralidad de factores de comparación cualitativos; y utilizar los resultados de la pluralidad de comparaciones para generar el indicador.
- 35 Preferiblemente, la pluralidad de factores de comparación cualitativos comprende una jerarquía de factores de comparación.
- En realizaciones, el módulo compactador es operable para comparar la primera y segunda información en dos o más resoluciones de la primera y segunda información, con el fin de medir la discrepancia y/o generar el indicador de calidad.
- 40 Preferiblemente, las dos o más resoluciones de la primera y segunda información son resoluciones diferentes en tiempo de la primera y segunda información. En una realización, el módulo compactador es operable para: comparar un primer periodo de tiempo de la primera información, desde dicha sección más reciente del primer flujo de datos, y un primer periodo de tiempo correspondiente de la segunda información, desde dicha sección más reciente correspondiente del segundo flujo de datos; y después de la primera comparación del periodo de tiempo, comparar un primer subperiodo del primer periodo de tiempo de la primera información con un primer subperiodo correspondiente del primer periodo de tiempo correspondiente de la segunda información.
- 45 De acuerdo con otro aspecto de la presente invención, se proporciona un método para sincronizar primer (302, 304) y segundo (310, 312) sistemas de manejo de datos independientes, cada uno de los primero y segundo sistemas de manejo de datos que recibe datos desde la misma pluralidad de proveedores (301) de datos, el primer y segundo sistemas de manejo de datos que procesan primer (302) y segundo (310) flujos (600) de datos generados desde los datos recibidos desde la pluralidad de proveedores de datos, en la que el primer y segundo sistemas de manejo de datos se disponen en diferentes ubicaciones, y en la que el primer y segundo flujos de datos no son exactamente idénticos, pero son lo suficientemente idénticos para soportar el propósito de la redundancia de proporcionar los dos flujos de datos separados en las dos ubicaciones, el método comprende: a) en un primer sistema (302, 304) de manejo de datos en una primera ubicación: combinar datos recibidos desde la pluralidad de proveedores de datos en un primer flujo (600) de datos, mediante un primer módulo (302) de generación de flujo de datos que comprende un primer compilador; y en un primer sistema (304) de procesamiento y almacenamiento: almacenar el primer flujo de datos en un primer almacén (402) de datos; duplicar datos desde el primer flujo de datos en un primer módulo (404) duplicador; combinar datos copiados desde el segundo flujo de datos con el primer flujo de datos en un primer

módulo (406) combinador; y transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones mediante por lo menos un primer puerto de comunicaciones; b) en un segundo sistema (310, 312) de manejo de datos en una segunda ubicación remota desde la primera ubicación: combinar datos recibidos desde la pluralidad de proveedores de datos en un segundo flujo (600) de datos mediante un segundo módulo de generación de flujo de datos (310), que comprende un segundo compilador; y en un segundo sistema (312) de procesamiento y almacenamiento: almacenar el segundo flujo de datos en un segundo almacén (402) de datos; duplicar datos desde el segundo flujo de datos en un segundo módulo (404) duplicador; combinar datos copiados desde el primer flujo de datos con el segundo flujo de datos en un segundo módulo (406) combinador; y transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones mediante por lo menos un segundo puerto de comunicaciones; y c) mediante por lo menos un módulo (306, 314) compactador, comparar la información desde el primer y segundo flujos de datos, el método comprende adicionalmente, mediante el primer y segundo sistemas de procesamiento y almacenamiento, a través de por lo menos un primer y segundo puertos de comunicaciones, transmitir información a y desde por lo menos un módulo compactador, y transmitir datos entre el primer y segundo sistemas de procesamiento y almacenamiento, a través de la red de comunicaciones, el método aún comprende adicionalmente: i) mediante el primer y segundo módulos de generación de flujo de datos, generar primer y segundo flujos de datos para procesamiento y almacenamiento en los respectivos primer y segundo sistemas de procesamiento y almacenamiento; ii) luego de recepción del primer y segundo flujos de datos en el primer y segundo sistemas de procesamiento y almacenamiento, mediante el primer sistema de procesamiento y almacenamiento, transmitir (502) primera información desde una sección (608) más reciente del primer flujo de datos hasta el módulo compactador, y mediante el segundo sistema de procesamiento y almacenamiento, transmitir segunda información desde una sección más reciente correspondiente del segundo flujo de datos hasta el módulo compactador; iii) mediante el módulo compactador: comparar (504) la primera y segunda información; medir un valor para discrepancia entre la primera y segunda información; probar (506) el valor de discrepancia contra un umbral; y, en el que el valor de discrepancia excede el umbral, generar: desde la primera y segunda información un indicador (508) del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contiene una mayor calidad de datos; y una instrucción (422) de duplicación para el módulo duplicador del respectivo sistema de procesamiento y almacenamiento cuyo flujo de datos se ha indicado por el indicador para contener la mayor calidad de datos; iv) mediante el módulo duplicador para el flujo de datos indicado, si se indica, duplicar (422) los datos desde dicha sección más reciente del flujo de datos indicado; v) mediante el sistema de procesamiento y almacenamiento para el flujo de datos indicado, transmitir, a través del puerto de comunicaciones, los datos duplicados al módulo combinador del sistema de procesamiento y almacenamiento para el otro flujo; y vi) mediante el módulo combinador para el otro flujo, reemplazar (424, 707) dicha sección más reciente correspondiente del otro flujo con los datos duplicados desde dicha sección (710) más reciente del flujo indicado.

35 Breve descripción de los dibujos

Ahora se describirán una o más realizaciones de la invención, solo por vía de ejemplo, con referencia a los dibujos acompañantes en los que:

La Figura 1 es un diagrama que ilustra un sistema de duplicación considerado previamente;

40 La Figura 2 es un diagrama que ilustra un sistema considerado previamente alternativo para redundancia en un sistema de procesamiento de datos;

La Figura 3 es un diagrama que ilustra un sistema para proporcionar redundancia en un sistema de procesamiento de datos de acuerdo con una realización de la invención;

La Figura 4a es un diagrama que ilustra las características de un componente de la Figura 3, de acuerdo con una realización de la invención;

45 La Figura 4b es un diagrama que ilustra la funcionalidad de los componentes de las Figuras 3 y 4a, de acuerdo con una realización de la invención;

La Figura 5 es un diagrama de flujo que ilustra las etapas de un método de acuerdo con una realización de la invención;

50 La Figura 6 es un diagrama que ilustra el procesamiento de una porción de un flujo de datos de acuerdo con una realización de la invención;

La Figura 7 es un diagrama que ilustra el procesamiento de un único canal de un flujo de datos de acuerdo con una realización de la invención; y

La Figura 8 es un diagrama que ilustra parte de un flujo de datos procesado de acuerdo con una realización de la invención.

55 Descripción detallada

La siguiente descripción establece realizaciones no limitantes específicas de la presente invención.

Las realizaciones de la invención proporcionan métodos y sistemas para la comparación en tiempo real de datos en un par (o más) de flujos de datos para detectar disparidades, y cuando sea necesario para reemplazar secciones de un flujo de datos con la sección correspondiente del otro. Estas realizaciones permiten que dos sistemas independientes (por ejemplo, en diferentes ubicaciones) se mantengan como redundancia entre sí para sincronizarse con un grado suficiente de precisión y puntualidad. Las realizaciones de la invención también permiten una transferencia de datos mucho más eficiente entre los dos sistemas o flujos de datos que los métodos considerados anteriormente, ya que la comparación en curso puede señalar exactamente las partes del flujo o flujos que se deben reemplazar, reduciendo de esta manera la cantidad de datos que necesitan ser transferidos.

5 Dado que la comparación es en curso o continua, no es necesario que el sistema se apague para que se pueda reemplazar cualquiera del flujo dado. En los sistemas en los que se requiere que los datos almacenados finales estén disponibles inmediatamente para servir a los usuarios o sistemas en otra parte de un sistema más amplio, esto evitará demoras en el acceso a esos datos.

15 Adicionalmente, dado que existe una comparación de los flujos resultantes entre las dos ubicaciones, se pueden identificar cualesquier discrepancias que no fueron evidentes en los sistemas considerados anteriormente (por ejemplo, debido a cualquier latencia entre las fuentes de datos).

Más aún, en realidad no existe necesidad de que los sistemas de la invención sean capaces de determinar cuál es el problema o problema potencial con un flujo de datos o carga, o incluso si hay uno; la comparación simplemente identifica una disparidad entre los flujos de datos e indica una acción de reemplazo limitada si se identifica una disparidad. Especialmente en sistemas en los que están presentes vastas cantidades de datos (por ejemplo, conjuntos de datos que contienen millones de puntos de datos), generalmente no tendrá sentido intentar determinar si los datos en una ubicación son mejores que los otros; un conjunto puede ser más preciso y menos preciso que el otro en varios puntos del conjunto. En sistemas considerados anteriormente, tales como el que se describe con referencia a la Figura 2, esta pregunta no solo es absurda, sino que, normalmente, ni siquiera se considera; todo el conjunto de datos se copia desde una ubicación hasta otra cuando se identifica un problema con un flujo. En realizaciones de la invención, los conjuntos de datos se comparan en secciones muy pequeñas (en comparación con el tamaño del flujo), y en cualquier caso se comparan si puede existir un problema con uno u otro flujo; es cualquier disparidad entre los conjuntos de datos que se determina.

30 Adicionalmente, las realizaciones de la invención son capaces de determinar las disparidades entre los datos en las dos ubicaciones con altos grados de granularidad; por ejemplo, puede aparecer en la comparación que los datos de una ubicación son similares a los de la otra, pero la diferencia real y su significado o, de lo contrario, solo se pueden determinar con un nivel de detalle inferior a aquel de una comparación inicial. Por ejemplo, las realizaciones de la invención permiten la sustitución de subconjuntos de datos, en lugar del flujo completo, lo que permite que se construya un conjunto de datos válido a partir de los dos flujos, incluso cuando ambos son parcialmente inválidos.

35 Adicionalmente, el tratamiento por realizaciones de la invención de subconjuntos separados de los datos, y la delimitación por marcadores de puntos de tiempo en secciones de estos subconjuntos, significa que los procesos de comparación que se van a realizar se pueden parcelar o cuantificar de manera efectiva, lo que significa que los recursos del sistema se pueden utilizar para realizar las comparaciones en paralelo, cuando esto puede dar un beneficio de rendimiento.

40 Una ventaja adicional de las realizaciones de la invención es que la comparación de los datos de las dos ubicaciones puede ser de aspectos cualitativos de los datos, o solo de partes de los flujos de datos. En contraste con los sistemas de la técnica anterior en los que los conjuntos de datos en sí mismos se transfieren en su totalidad (o incluso a algún otro tipo de sistema en el que podrían compararse los conjuntos de datos completos), la carga computacional por lo tanto se reduce enormemente. En general, la comparación de datos o indicadores de esos datos es computacionalmente barata, mientras que el intercambio de los datos en sí es computacionalmente costoso. En realizaciones de la invención, el intercambio de datos solo se realiza siempre que sea necesario, y solo en pequeñas secciones.

50 Una ventaja adicional es que, en realizaciones en las que los flujos de datos se almacenan y se ponen a disposición de los usuarios o sistemas posteriores de inmediato, y luego se hacen las comparaciones para mejorar estos datos almacenados, los datos pueden estar disponibles de esta manera, instantáneamente, y las otras características ventajosas de la invención se pueden implementar tan pronto como se completen las comparaciones y cualquier copia. Por lo tanto, en dichas realizaciones puede haber un retardo cercano a cero en la provisión de datos a los usuarios; en algunos sistemas de procesamiento de datos, esto puede ser extremadamente importante.

55 La Figura 3 es un diagrama que ilustra un sistema (300) para proporcionar redundancia en un sistema de procesamiento de datos de acuerdo con una realización de la invención. Los flujos 1 y 2 (302 y 310) de datos están compuestos o compilados de la misma fuente o una a más fuentes de datos separadas (no mostradas en la Figura 3) de manera similar a aquella mostrada para el sistema considerado anteriormente de la Figura 2. Los flujos de

datos reciben los datos de las fuentes a través de enlaces o cargas, indicados en 301. Puede ser que la misma fuente de datos única suministre los datos para ambos flujos de datos.

Los medios para transferir los datos desde las fuentes de datos hasta los flujos para la compilación pueden ser mediante algún tipo de comunicación local, tal como un enlace de datos o un bus. En otras realizaciones de la invención descritas en el presente documento, las Ubicaciones 1 y 2 están separadas geográficamente, y por lo tanto, estas realizaciones resuelven el problema encontrado por el sistema considerado anteriormente de sincronización precisa de los sistemas separados. En dichas realizaciones, la comunicación entre las fuentes de datos y los flujos en las ubicaciones separadas puede ser por redes de comunicaciones, tales como redes inalámbricas o de área extensa cableadas, comunicaciones por satélite, redes de telecomunicaciones o canales cableados (líneas terrestres).

El flujo de datos generado a partir de los datos de las fuentes puede ser de cualquier tipo conocido en la técnica en el que los datos de fuentes únicas o múltiples se pueden componer en un flujo para transferencia o transmisión. Por ejemplo, el flujo de datos puede estar empaquetado, o en bloques o en marcos, y puede multiplexar o combinar de otro modo elementos de datos separados en el mismo flujo, que puede separarse en una etapa posterior. En realizaciones de la invención, el flujo de datos toma la forma de un flujo multiplexado en paquetes, en el que los paquetes de datos llevan cada uno una carga útil de datos. Se pueden multiplexar múltiples series de paquetes de datos para proporcionar un flujo de transporte o combinado, y luego se multiplexan para devolver las series de datos separadas. En realizaciones, las diferentes fuentes de datos pueden proporcionar diferentes series de datos para ser (o pueden estar) empaquetadas de esta manera, que luego pueden ser multiplexadas. Adicionalmente, desde cada fuente de datos, los datos proporcionados pueden estar en elementos, bandas o canales separados de esa fuente, que pueden combinarse o multiplexarse en la misma carga.

En las realizaciones de la invención, un flujo de datos se compone de muchas bandas o canales de datos discretos, cada canal de datos se compone de una serie de paquetes de datos encabezados, cada uno de los cuales lleva una carga útil de datos. Estas características se describen adicionalmente con referencia a las Figuras 6 a 8.

El flujo de datos es esencialmente una serie temporal de datos, que permite las características de los métodos de la realización de la invención. Por lo tanto, los flujos pueden compararse con referencia a la misma dirección en el tiempo, y por ejemplo (como se describe con más detalle adelante) se pueden utilizar marcadores de puntos de tiempo para delinear el punto en el último flujo comparado.

En cada una de las Ubicaciones 1 y 2, se ubica un sistema similar o idéntico. El flujo de datos generado se carga a un sistema (304, 312) de procesamiento y almacenamiento de datos, que proporciona muchas de las funciones de las realizaciones de la invención. De este modo, el flujo 1 de datos se carga a un primer sistema (304) de procesamiento/almacenamiento, y el flujo 2 de datos a un segundo (312). Se debe tener en cuenta que, una vez que se ha recibido (y procesado) un flujo de datos en los respectivos sistemas de almacenamiento y procesamiento de datos, esos datos normalmente se ponen a disposición o se entregan (320, 322) para uso por un sistema o usuario posterior en un sistema o disposición más amplio que requiera acceso a esos datos. Por ejemplo, esto puede ser por una relación cliente-servidor común, con el sistema de procesamiento/almacenamiento que actúa como servidor.

Cada uno de dichos sistemas de procesamiento y almacenamiento se comunica con un comparador (306, 314) local. El primer comparador local para el primer sistema de procesamiento/almacenamiento también está en comunicación, a través de una red 308 de comunicaciones, con el segundo sistema de procesamiento/almacenamiento correspondiente en la Ubicación 2, y el segundo comparador local (en la Ubicación 2) está en comunicación con el primer sistema de procesamiento/almacenamiento, a través de la red de comunicaciones. El primer y el segundo sistema de procesamiento/almacenamiento también se comunican directamente (316) entre sí, a través de la red.

Como se describirá con más detalle a continuación, los comparadores 306 y 314 locales permiten la comparación, en tiempo real, de los datos de los flujos 1 y 2 de datos entre sí a través de los sistemas de procesamiento/almacenamiento en las ubicaciones separadas (y a través de la red de comunicaciones). Se puede observar cualquier disparidad entre los datos de los flujos y el contenido se puede copiar de un flujo a otro si se requiere, por ejemplo, a través del enlace 316 del sistema directo. La provisión de dos comparadores proporciona redundancia adicional en el sistema, de tal manera que, en el caso de mal funcionamiento de un comparador, las comparaciones de flujo (y los reemplazos cuando sea necesario) aún se pueden realizar. Adicionalmente, si uno de los comparadores tiene un bajo rendimiento, aún se puede confiar en el sistema para que produzca una sincronización confiable de los flujos, ya que por lo menos uno de los comparadores debe proporcionar en todo momento el reemplazo necesario de las secciones del flujo.

En una realización alternativa, los comparadores 306 y 314 se pueden reemplazar por un único comparador, en una tercera ubicación accesible desde/hacia la Ubicación 1 y la Ubicación 2 a través de la red (314) de comunicaciones. Esto se puede proporcionar en los casos en que la potencia de cómputo local es limitada, o cuando se determina que es más eficiente utilizar un solo comparador, en lugar de dos comparadores locales. Por ejemplo, en ciertas realizaciones, el uso de una única ubicación para el comparador puede proporcionar una ventaja de tener un solo

árbitro de la comparación entre los folios. También puede ser que se necesite menos poder de cómputo en general para un comparador alojado en una única ubicación.

La Figura 4a es un diagrama que ilustra las características del sistema (304) de procesamiento y almacenamiento de datos en la Ubicación 1 de la Figura 3, de acuerdo con una realización de la invención. El procesamiento/almacenamiento 312 en la ubicación 2 tiene las mismas características correspondientes (en comunicación con su propio comparador local).

Como se muestra en la Figura 3 y la Figura 4a, el flujo 1 de datos se recibe en el sistema 304 de procesamiento/almacenamiento. El flujo de datos se puede pasar al comparador (306 de la Figura 3); en realizaciones, los mismos medios (402) de almacenamiento, tales como medios de almacenamiento físicos, por ejemplo, medios de memoria volátiles, se pueden utilizar para el almacenamiento inicial y para el almacenamiento (408) final del flujo de datos. La memoria se puede manejar utilizando los protocolos adecuados para evitar la sobrescritura de cualquiera de los datos almacenados finalmente por los nuevos datos de flujo que se reciben. En realizaciones, el flujo de datos puede de hecho estar disponible o servir desde el sistema de procesamiento/almacenamiento tan pronto como se haya recibido; incluso si los datos aún no se han comparado con el otro flujo de datos para superar inconsistencias o disparidades, puede ser que los datos en ese estado aún no refinado puedan no obstante ser útiles para el último usuario o sistema. En este sentido, la comparación y las posibles modificaciones de los datos del flujo 1 de datos pueden verse como una mejora de los datos almacenados para el flujo 1 de datos, sin interrumpir el servicio de los datos para los usuarios o sistemas posteriores. De hecho, es una ventaja de las realizaciones de la invención que no se interrumpa la entrega de los datos desde el flujo de datos de los sistemas 304, 312 de procesamiento/almacenamiento, aunque el proceso de comparación y modificación se realice en tiempo real.

Una vez que se ha recibido el flujo de datos, sus datos se pueden dirigir (por ejemplo, desde el almacenamiento 402 de datos) hacia el comparador local como se muestra en la Figura 4a. El flujo 1 de datos también se proporciona al comparador en la segunda ubicación. En la segunda ubicación, de manera similar, el segundo flujo de datos se proporciona al comparador local, y también al comparador de la Ubicación 1. El comparador local, al determinar una disparidad o diferencia entre los dos flujos de datos, puede proporcionar instrucciones al sistema 304 de procesamiento/almacenamiento para copiar o duplicar una sección del flujo 1 de datos que se va a utilizar para reparar o conformar el flujo 2 de datos, al reemplazar una sección del flujo 2 de datos. En la realización en la Figura 4a, las instrucciones se reciben desde el comparador en la copiadora/duplicadora (404). El contenido de datos copiados se proporciona luego al sistema de procesamiento/almacenamiento en la Ubicación 2. En el caso en el que se determine que el flujo 2 de datos debe proporcionar una sección copiada para la sustitución en el flujo 1, los datos copiados se reciben desde el sistema de procesamiento/almacenamiento de la Ubicación 2 en un combinador (406) del sistema (304) de procesamiento/almacenamiento del flujo 1. El combinador luego reemplaza la sección correspondiente del flujo 1 con la sección copiada del flujo 2.

En cualquier caso, ya sea que una sección determinada del flujo 1 de datos se ha pasado sin modificar, o copiado para transferir al flujo 2 de datos, o modificar por el combinador, el flujo 1 de datos no modificado o modificado resultante se puede almacenar en el almacenamiento 402 de datos del sistema 304 de procesamiento/almacenamiento. De esta manera, los contenidos del almacenamiento de datos (final) en cada una de las ubicaciones debe ser idéntico, o lo más cerca posible: siempre que haya una disparidad entre los dos flujos, los flujos se han conformado entre sí al copiar una sección desde una a la otra; y el procesamiento se ha realizado sobre la marcha, en lugar de pausar o apagar para esperar un proceso de detección de problemas, o para que arribe una gran cantidad de datos.

En realizaciones alternativas, el almacenamiento de los datos del flujo de datos en cada sistema de procesamiento/almacenamiento se puede realizar mediante algún almacenamiento inicial o búfer, que puede dirigir los datos desde el flujo hacia el comparador local como se muestra en la Figura 4a. En dichas realizaciones, este almacenamiento inicial puede ser parte del bloque 402 de almacenamiento de datos mostrado. La Figura 4b es un diagrama que ilustra la funcionalidad del sistema de procesamiento/almacenamiento, el comparador y los enlaces de comunicaciones para el flujo 1 de datos/Ubicación 1 de las Figuras 3 y 4a, de acuerdo con una realización de la invención.

Como se describió anteriormente, en la Ubicación 1, el flujo 1 de datos se ingresa al sistema de procesamiento/almacenamiento. Luego se compara (420) en tiempo real con los datos del flujo 2 de datos, que se recibe desde la Ubicación 2 (desde el sistema de procesamiento/almacenamiento del mismo). La comparación indica la copia de los datos del flujo 1 si es necesario, es decir, si se encuentra una disparidad o diferencia. Esta instrucción da como resultado la copia (422) de la sección relevante del flujo 1. Los datos copiados desde el flujo 1 se pasan luego al sistema de procesamiento/almacenamiento para el flujo 2, para su incorporación al flujo 2 de datos. Si no existe dicha instrucción desde el comparador, no se requiere copiar el flujo 1 para esta sección.

Si no ha habido instrucciones para copiar (422) desde el flujo 1, no obstante, puede ser que una sección del flujo 1 de datos se ha determinado por el comparador para requerir el reemplazo mediante datos desde el flujo 2. En tal caso, la comparación (420) da como resultado una instrucción para incorporar (424) datos copiados del flujo 2. Los datos se reciben desde el sistema de procesamiento/almacenamiento de la Ubicación 2, y se incorporan en los

datos del flujo 1 (424). Una vez que esto ha tenido lugar, los datos del flujo 1 de datos se pueden devolver al almacén de datos (por ejemplo, para actualizar el registro de almacenamiento de datos desde un formulario sin modificar anterior).

5 Si no existe instrucciones para incorporar datos desde el flujo 2, entonces el flujo 1 de datos pasa por alto la etapa de incorporación y se envía para el almacenamiento de datos.

10 En las realizaciones descritas con referencia a las Figuras 3, 4 y 4a anteriores, existen dos comparadores (306, 314) locales, uno para cada sistema de procesamiento/almacenamiento local. Por supuesto, normalmente será el caso que ambos comparadores decidan lo mismo para la sección dada del flujo de datos que se compara en cualquier momento, y por lo tanto, la copiadora/duplicadora (404) y combinadores (406) se pueden enviar instrucciones desde
 15 ambos comparadores para copiar desde un flujo hasta el otro para la sección dada. Estas instrucciones se pueden manejar de manera adecuada para que no compitan, por ejemplo, al simplemente filtrar las instrucciones recibidas en los sistemas de procesamiento/almacenamiento; si se reciben las mismas instrucciones desde ambos comparadores, el sistema 304 de almacenamiento de procesamiento se puede programar para descartar una de las (mismas) instrucciones. En realizaciones alternativas, el sistema puede simplemente permitir que ambos comparadores instruyan la misma acción en esos casos, de modo que, en aquellos casos, la copiadora/duplicadora y el combinador en una ubicación pueden ser indicados para realizar la acción de copia dos veces; esto simplemente dará como resultado un proceso de copia/transferencia adicional, ya que la sección a copiar será necesariamente la misma.

20 En una realización alternativa, el tiempo de los comparadores se puede compensar ligeramente, de tal manera que dichos conflictos son más raros. Si el primer comparador ya ha realizado una copia/transferencia para una sección dada de los flujos, los datos con los que trabajará el segundo comparador serán ahora la sección modificada desde ese flujo, y en ese caso no será necesario indicar la copia/transferencia. En los casos en que el primer comparador, por alguna razón, haya perdido una disparidad, o esté funcionando mal, el segundo comparador verá ahora la disparidad y llevará a cabo el proceso de copia/transferencia.

25 La Figura 5 es un diagrama de flujo que ilustra las etapas de un método de acuerdo con una realización de la invención. Estas etapas utilizan la infraestructura y los componentes del sistema, por ejemplo, como se describió anteriormente con referencia a las Figuras 3, 4a y 4b, y se pueden implementar durante las etapas de comparación e incorporación descritas, por ejemplo, utilizando los comparadores (306, 314), copiadora/duplicadoras (404) y combinadores (406) descritos. La Figura 5 también se refiere a las características de la Figura 6, un diagrama que
 30 ilustra el procesamiento de una porción de un flujo de datos de acuerdo con una realización de la invención.

En primer lugar, los datos del flujo 1 de datos y el flujo 2 de datos se recuperan para comparación. En las realizaciones, los marcadores se insertan en los flujos para indicar dónde terminó la última sección del flujo de datos comparado. Por lo tanto, en la etapa 502, todos los datos desde ambos flujos ya que el último marcador en el flujo se recopilan para comparación.

35 En la figura 6 se muestra una representación de un flujo (600) de datos de acuerdo con realizaciones de la invención. En realizaciones, el flujo de datos está formado por segmentos de datos separados o canales 602, que representan diferentes tipos de datos combinados en los flujos de datos. Como se muestra en este documento, este flujo de datos incluye los canales A, B y C, y no se muestran canales D adicionales a n. El número de canales puede ser extremadamente grande, siempre que las cantidades de datos sean lo suficientemente concisas para que los
 40 comparadores puedan recorrer los canales. Por ejemplo, en un entorno manejado por los inventores, 'n' era del orden de 100.000.

45 Por ejemplo, estos diferentes segmentos, tipos de datos o bandas/canales pueden ser diferentes tipos de elementos de medios en un flujo de transporte de medios, o diferentes tipos de audio en un flujo de datos de audio combinado. En otras realizaciones de la invención, los canales son para diferentes tipos de datos que son producidos por las fuentes de datos; cada tipo de datos se actualiza regularmente con nuevos eventos de datos, y estos eventos y valores de datos asociados son rastreables dentro de cada canal de datos. Los canales de datos son discretos, ya que tratan con diferentes tipos de datos o, a veces, con representaciones separadas de los mismos datos.

50 Dado que los canales de datos son discretos, cada uno se pueden tratar por separado y, por lo tanto, comparar por separado. Por lo tanto, cuando se comparan el flujo 1 de datos y el flujo 2 de datos, las comparaciones se realizan en realizaciones reales entre canales individuales de esos flujos, por ejemplo, comparando el canal A del flujo 1 con el canal A del flujo 2, y así sucesivamente. Esto permite el reemplazo de secciones en estos canales, en lugar de en todo el flujo, lo que permite que se construya un conjunto de datos válido a partir de los dos flujos, incluso cuando ambos son parcialmente inválidos, debido a que una sección de diferentes canales en cada uno tiene una discrepancia con el otro flujo

55 Cada canal de datos en estas realizaciones está compuesto por bloques, cuadros o paquetes de datos 604. Por ejemplo, cuando el flujo de datos está empaquetado, cada canal comprende una serie de paquetes de datos, cada uno con un encabezado y una carga útil de datos, de la manera conocida en el art.

Para cada canal, la comparación como se describe anteriormente con referencia a la Figura 5 se realiza entre todos los datos en un canal desde el último marcador insertado. Los marcadores 606 se insertan en el conjunto de datos de cada canal al final del proceso de comparación, como se describirá a continuación; la siguiente comparación siempre comienza desde ese marcador y recopila todos los datos desde entonces, para que nunca se pierda o no se compare ningún dato.

En la Figura 6, la dirección de 'viaje' del flujo de datos o canal es de izquierda a derecha. Por lo tanto, dado que el último marcador 606 en el canal A, existe una cantidad de datos en una sección de datos (608) de ese canal abarcado en la Figura por la flecha 610. Esta sección 608 es, por lo tanto, el conjunto de datos que se comparará con su contraparte del flujo 2 del canal A. Como puede verse, la sección 608 está formada por una pluralidad de paquetes de datos 604. Todos los datos anteriores al marcador 606 ya se han comparado en operaciones de comparación anteriores.

La sección correspondiente del flujo 2 de datos también se identificará, en paralelo, exactamente de la misma manera. Es decir, el canal A del flujo 2 tendrá un marcador 606 insertado en el punto correspondiente del flujo de datos del canal, por ejemplo, en el mismo paquete numerado o sello de tiempo. La comparación se indicará para el mismo comparador (ya sea 306 o 314) y, por lo tanto, el punto final, de corte o de inicio será el mismo para ambas corrientes; por ejemplo, se puede estipular que se recoja el mismo número de paquetes para cada canal desde el último marcador, o que transcurra un tiempo determinado desde que se insertó el último marcador antes del corte para la comparación. Por lo tanto, las secciones de los dos flujos de datos que se van a comparar son lo más parecidas posible a la comparación que se realizará.

Este proceso de marcado, selección e identificación de la sección que se va a comparar con su contraparte se puede llevar a cabo en los comparadores (306, 314) en sí mismos, y los flujos 1 y 2 de datos se cargan constantemente de ellos. Sin embargo, en una realización alternativa, si se requiere que los comparadores procesen solo la comparación para conservar la potencia de procesamiento, el marcado, la selección y la identificación de las secciones para la comparación se pueden realizar en los sistemas de procesamiento/almacenamiento (304, 312) antes de pasando a los comparadores. Por ejemplo, estos pasos pueden llevarse a cabo en o junto con el almacenamiento inicial/búfer 402.

Se puede observar que el hecho de que los canales se pueden tratar por separado, y que los marcadores de puntos de tiempo se utilizan para delinear secciones de esos canales, significa que la función de los comparadores puede ser totalmente independiente de los flujos de datos y, de hecho, de los demás, y puede verse simplemente como una mejora de los datos que se almacenan para cada flujo. Cada acción del comparador en las secciones correspondientes de los flujos se realiza efectivamente para que lo haga un comparador, lo que significa que cada comparador puede estar en cualquier ubicación y, de hecho, puede haber cualquier número de ellos; si se debe acelerar el procesamiento del flujo, se pueden utilizar comparadores adicionales en paralelo.

Volviendo a la Figura 5, ahora que se ha identificado la sección del flujo 1 para la comparación, y la sección correspondiente del flujo 2 se habrá identificado exactamente de la misma forma, estos dos conjuntos de datos se recuperan (502) para la comparación. Los conjuntos de datos se comparan debidamente (504) y se calcula una discrepancia entre los conjuntos de datos. La discrepancia entre los conjuntos de datos, y cuál de ellos puede ser preferible copiar desde, se puede calcular de varias maneras, en función de uno o más de una pluralidad de factores asociados con las respectivas secciones de flujo de canales recibidas.

Estas secciones, como se señaló anteriormente, serán en gran medida idénticas. Las mismas fuentes de datos se utilizan para los flujos 1 y 2 de datos; se compilan de la misma manera; el procesamiento de los flujos se realiza de la misma manera en todos los componentes descritos anteriormente. Sin embargo, especialmente en ciertos tipos de entornos de procesamiento de datos donde las cantidades de datos son vastas, y la cantidad de nuevos eventos de datos o elementos de datos emitidos por segundo también es muy grande, inevitablemente habrá discrepancias entre dos de tales flujos de datos, incluso si los medios para producirlos se construyen cuidadosamente para evitar cualquier discrepancia. En un entorno observado por los inventores, por ejemplo, el almacenamiento es de petabytes de datos, y se registran cinco mil millones de elementos de datos por segundo.

Como ejemplo, la sincronización comparativa de los flujos de datos en tales entornos inevitablemente se desviará de vez en cuando. Una de las fuentes de datos que se está utilizando puede, por algún motivo, eliminar un dato o paquete al transmitir a una ubicación, pero no a la otra. Estos ejemplos son previos a la consideración de cualquier problema inherente al tener los flujos de datos alojados en diferentes ubicaciones geográficas; los enlaces o cargas desde las fuentes a las diferentes ubicaciones pueden comportarse de manera diferente, por ejemplo. El ejemplo dado anteriormente con respecto al sistema considerado anteriormente de la Figura 2 es uno; si los datos de diferentes fuentes se reciben en momentos diferentes, los resultados de los datos anotados en los dos flujos pueden diferir sustancialmente, sin que haya ningún problema inherente con ninguno de los dos.

Sin embargo, los métodos y sistemas de realización de la invención aprovechan la noción de que los conjuntos de datos almacenados finales no necesitan ser completamente idénticos (ya que esto podría implicar siempre copiar uno u otro flujo al otro), pero simplemente que los conjuntos finales deben ser lo más similares posible para que la redundancia entre los dos conjuntos finales sea lo suficientemente útil, al tiempo que se eliminan las discrepancias

más grandes entre los flujos. Por lo tanto, el uso de un umbral permite que pasen algunas discrepancias muy pequeñas, mientras que se tratan las más significativas.

5 Los métodos y sistemas de realizaciones de la invención, por lo tanto, buscan determinar cualquier diferencia entre las últimas secciones de flujo de canal correspondientes. La determinación de la diferencia se puede utilizar junto con una medida de una diferencia en la calidad de las corrientes; sin embargo, puede haber factores tanto cualitativos como cuantitativos que se incluyen en la evaluación de qué flujo/canal es de mejor calidad.

10 La diferencia(s) puede estar simplemente en el contenido de los conjuntos de datos; por ejemplo, puede ser que el contenido de uno (o más) de los paquetes de datos contenidos en la sección del canal A del flujo 1 difiera ligeramente del de la sección del canal A del flujo 1. Por ejemplo, un parámetro simple o datos el valor puede diferir, quizás solo en un paquete de datos o en un dato dentro de ese paquete. Por ejemplo, puede ser que un valor acumulado de un parámetro en un canal de datos haya alcanzado un valor más alto en la sección actual de ese canal desde el flujo 2, que en la sección correspondiente de ese canal desde el flujo 1. En otro ejemplo, un (o una pluralidad de) parámetros de segundo orden relacionados con los datos en una sección o paquete pueden compararse, como una relación señal/ruido para los dos conjuntos de datos que se comparan.

15 Estos factores se pueden determinar mediante una comparación simple de valores en casos simples, y para datos más complejos se podría utilizar una medida de similitud tan común en la técnica para comparar datos y establecer un nivel de diferencia o discrepancia. Una medida simple de este tipo de característica sería una marca de tiempo en un encabezado de paquete de datos dado. Por ejemplo, puede ser por algún motivo que se haya determinado que cierto evento de datos ocurrió al mismo tiempo de acuerdo con los datos compilados en el flujo 1 de datos, pero se determinó que ocurrió en un momento un poco más tarde de acuerdo con el flujo 2 de datos.

20 Otras diferencias detectables entre los conjuntos de datos se asocian con el nivel estructural del flujo de datos. Por ejemplo, las secciones del (canal del) flujo de datos pueden diferir en el número de paquetes, bloques o cuadros de datos que se han recopilado en el período desde el último marcador, o por el contrario en la longitud del período desde el último marcador en el que arriba el último de un número dado de paquetes de datos.

25 Las diferencias de potencial también se pueden rastrear a lo largo de una sesión determinada de procesamiento, para indicar con más detalle si una diferencia es significativa. Por ejemplo, si un determinado parámetro de datos difiere entre los conjuntos de datos en una cantidad, si esta cantidad de diferencia es una ocurrencia común para este canal de datos, como lo indica el seguimiento, puede ser que la diferencia se pueda determinar como intrascendente.

30 Otras disparidades entre los flujos de datos serán más obvias que los cambios sutiles en el contenido de los datos o la duración o temporización del flujo/canal. Por ejemplo, si un flujo de datos falla por completo, los comparadores no tendrán nada que comparar de una carga. Otros problemas más importantes, que se pueden observar con poco procesamiento, pueden incluir que el flujo de datos esté corrompido de alguna manera, que el contenido de los datos esté confuso de alguna manera, o que esté completamente equivocado en lugar de un ligero retardo que se podría esperar.

35 En realizaciones de la invención, los conjuntos de datos de los dos flujos se pueden comparar utilizando una combinación de factores tales como aquellos descritos anteriormente, con el fin de proporcionar una evaluación más precisa de la comparación. Por ejemplo, puede ser que una pequeña disparidad de tiempo no sea significativa, pero cuando se combina con diferencias en otros lugares, la calidad es lo suficientemente más baja en un flujo que en el otro para justificar el reemplazo de esa sección con los datos desde el otro flujo. En realizaciones, cada tipo de factor se puede comparar y tener su propio umbral diferente; una disparidad de tiempo puede tener un umbral más bajo que un cierto tipo de disparidad de datos, por ejemplo.

40 En realizaciones, los factores evaluados pueden estar en capas u ordenados; por ejemplo, una evaluación de si el flujo está presente puede descartar rápidamente (en) si hay una interrupción grave. La evaluación puede entonces, si es necesario, proceder a comparaciones más sutiles.

45 En las realizaciones de la invención, las comparaciones (para calcular la discrepancia o para determinar la calidad comparativa) también se pueden realizar en diferentes niveles de flujos de datos. Por ejemplo, en lugar de limitarse a evaluar cada canal durante un período similar, y al ingresar marcadores en el último punto de evaluación como se ilustra en la Figura 6, se pueden realizar niveles de comparación adicionales comparando períodos más largos o más cortos de los canales de flujo de datos. Por ejemplo, si el período (610) indicado desde el último marcador en este caso es de una hora, se pueden hacer otras comparaciones (durante esta misma sección, por ejemplo) en períodos de diez minutos, un minuto o incluso un día. Las realizaciones particulares que utilizan esta característica se describirán con más detalle a continuación, con referencia a las Figuras 5 y 6.

50 También puede ser que se utilicen diferentes tipos de discrepancia para incitar verificaciones en otros tipos de discrepancia, o en tipos de diferencia de calidad. Por ejemplo, si hay un problema obvio con un flujo, el proceso de copia/transferencia se indicará de manera directa, pero si no existe un problema obvio, el sistema puede solicitar una verificación de la discrepancia de un tipo dado y luego de la temporización. De manera similar, si se ha identificado la discrepancia, la verificación de la comparación de calidad puede comenzar con un indicador de

diferencia de calidad más obvio, y si se pasa, luego verificar si existen indicadores más sutiles, o con resoluciones de tiempo más altas (ver más abajo).

5 Volviendo a la Figura 5, una vez realizada la comparación (504), se determina si la discrepancia o disparidad entre los conjuntos de datos es mayor que un umbral (506). Por ejemplo, si se ha determinado que un flujo de datos indica que un evento de datos ocurrió en el tiempo t , y el otro indica que sucedió en el tiempo $t+10$ segundos, esto no se puede considerar lo suficientemente significativo como para justificar el incurrir en el coste de transferir el conjunto de datos desde el otro flujo para su incorporación. Si en un conjunto de datos extremadamente grande existe una diferencia de n en un parámetro en un canal entre los flujos, es posible que esto se pueda ignorar. Sin embargo, si la diferencia es $500n$, se puede actuar.

10 El umbral también se puede evaluar en vista del historial de diferencias en un determinado flujo o canal. Por ejemplo, si se puede tolerar una cantidad general de diferencia en un tipo de datos, ya que el seguimiento indica que esta diferencia es común, el umbral se puede establecer con esto en mente.

15 Utilizando este método, los sistemas de la invención pueden reducir la transferencia innecesaria de datos cuando los flujos no son exactamente idénticos, pero son suficientemente idénticos para soportar el propósito de la redundancia de proporcionar los dos flujos de datos separados en las dos ubicaciones. Por ejemplo, como se señaló anteriormente en contraste con los sistemas considerados anteriormente, no es necesario transferir flujos completos de datos. Por supuesto, en ciertos entornos, la redundancia requerirá que los flujos de datos sean exactamente más idénticos que en otros; en dichos casos, el umbral y la cantidad asociada de potencia de cómputo requerida pueden ajustarse al resultado apropiado para ese entorno.

20 Este método también implica que en realidad no existe necesidad particular de saber si hay o no un problema con cualquiera de los flujos de datos. En un caso en el que los flujos de datos no concuerdan ligeramente con la temporización de un evento de datos, si los flujos se compilan desde la misma fuente, puede que ni siquiera se pueda determinar en un sentido objetivo cuál de los flujos es 'correcto'. Es simplemente suficiente para comparar los flujos y determinar si hay una disparidad, y si la disparidad es lo suficientemente significativa como para determinar el flujo y la copia de mejor calidad de ese flujo al otro.

Si se determina en 506 que la discrepancia no es mayor que el umbral, el proceso procede a insertar (516) un nuevo marcador (606) en el canal de datos para indicar el final de la sección que se ha comparado, y donde debe comenzar la siguiente comparación. El proceso se puede repetir, comenzando en la primera etapa 502 para la siguiente sección del canal.

30 Si se determina que la discrepancia es mayor que el umbral, el proceso continúa con el bloque 508, para determinar cuál de los flujos 1 y 2 tiene datos de mejor calidad en esta sección del canal de datos que se está comparando. Puede ser como se señaló anteriormente que las medidas de calidad (o cualitativas) ya se han determinado en la etapa de comparación. Por ejemplo, se puede haber determinado que uno de los flujos contiene datos corruptos; esto se puede utilizar fácilmente como una medida de qué flujo es de mejor calidad y, por lo tanto, se debe utilizar para reemplazar los datos en el otro.

35 A menudo puede ser claro qué flujo es de mejor calidad, ya que el flujo que tiene datos que es obviamente más disparejo, por ejemplo, una tendencia rastreada en esos datos se habrá recogido en la etapa de comparación, y este flujo, por lo tanto, tendrá sus datos reemplazados. Sin embargo, puede haber varias situaciones en las que la comparación sea concluyente en cuanto a si hay una disparidad, pero no concluyente en cuanto a qué flujo se debe copiar para reemplazar la sección en el otro flujo. Un ejemplo simple podría ser que ambos flujos tienen un valor diferente para un parámetro, pero ambos divergen desde una tendencia o medir en una cantidad similar.

40 En las realizaciones de la invención, los conjuntos de datos de los dos flujos se comparan utilizando una combinación de indicadores de calidad, que pueden incluir factores tales como aquellos descritos anteriormente al comparar las corrientes, de tal manera que se pueda determinar una comprensión general de la calidad relativa del flujo. Por ejemplo, se puede haber determinado en comparación que un parámetro difiere entre los conjuntos de datos del canal que se procesa, pero puede requerir evaluaciones comparativas adicionales para descubrir cuál es el mejor flujo para reemplazar el otro. Por ejemplo, puede ser que el primer parámetro sea significativamente diferente en un flujo desde los valores anteriores; esto puede indicar que se debe reemplazar esta sección de flujo.

45 En realizaciones, la determinación de la calidad comparativa entre los flujos se llevará a cabo normalmente en base a factores prima facie relativos a los datos en los propios flujos. Por ejemplo, si un resultado de datos en un flujo se indica como 'X', y la cantidad de eventos de datos acumulados en ese flujo es alta, pero el resultado en el otro flujo es 'Y', y la cantidad de eventos de datos acumulados en ese flujo es muy bajo, se puede suponer que el resultado 'X' es el correcto y, por lo tanto, que el primer flujo es de mejor calidad. Dichas comparaciones pueden indicar directamente la calidad, o se puede utilizar una puntuación acumulativa de dichas comparaciones para las secciones que se comparan.

50 Un factor de uso común es si los datos en un flujo no tienen sentido o son evidentemente erróneos por alguna razón. Por ejemplo, un flujo puede indicar que se ha producido un resultado de evento de datos en el que se desprende claramente del resto de la sección que se compara que este resultado aún no se debe. Un flujo puede indicar un

valor 'alto' cuyo valor absoluto es inferior al valor 'bajo' indicado. Un flujo empaquetado puede incluir un cuadro que no se refiere de la manera habitual a su contenido, o a los cuadros siguientes y anteriores.

5 Una vez hecha la determinación (508) de qué el flujo es de mejor calidad, se puede elegir la acción de copiado respectiva. Cuando se determina que el flujo 1 es de mejor calidad para esta sección del canal actual, los datos desde el último marcador del flujo 1 en ese canal actual se copian (510) desde el flujo 1 y reemplazan la sección correspondiente del canal actual en el flujo 2, desde el último marcador. De manera similar, si se determina que el flujo 2 es superior, ocurre la transacción opuesta (512).

10 En algunos casos, puede ser indeterminable qué flujo es el de mejor calidad, una vez que se identifica una disparidad. En dicho caso, puede ser que, en el interés de la eficiencia y el progreso en el almacenamiento de los flujos, el sistema omite (514) la etapa de copia/reemplazo para esta sección, e inserte el nuevo marcador (516) listo para comenzar en la próxima sección. En una alternativa, se utiliza un árbitro (514) externo para evaluar el flujo de mejor calidad. Esto podría ser un componente o aplicación dedicada separada con mayor poder de cómputo, que solo se puede utilizar para estos casos raros, o puede ser un usuario o supervisor del sistema al que se le proporcionen los datos para revisarlos y se le solicite devolver una decisión.

15 En otra realización, si la longitud de la sección para el comparador es un período dado del canal(es), se puede solicitar al sistema que no inserte un marcador si no se realiza la determinación de la calidad. Esto significaría que la próxima ronda de evaluación de la calidad se haría de manera efectiva en dos secciones o períodos del canal; esto puede proporcionar un resultado cuando no es el período más corto.

20 En otra realización, puede ser que, en casos indeterminados, se lleve a cabo una opción predeterminada para copiar un flujo (por ejemplo, el flujo 1) al otro.

En una realización, en dichos casos indeterminables, puede ser que se pueda utilizar un nivel o resolución diferente de los flujos para la comparación. Esto se indica en la Figura 5 como una salida (518 alternativa) de la evaluación (508) de calidad, que vuelve a la etapa de comparación de los dos flujos para calcular la discrepancia (504), pero esta vez en el nivel/resolución diferente.

25 Por ejemplo, puede ser que la sección que se está comparando con los flujos sea de un período de una hora. Puede ser indeterminable cuál es el mejor flujo, por lo que el proceso se puede repetir a intervalos de diez minutos dentro de esa sección de una hora; cada sección de diez minutos se puede comparar con su sección de contraparte de diez minutos en el otro flujo; esto se puede repetir a lo largo de la sección de la sección original de una hora hasta que se identifique la sección de diez minutos con calidad diferente. De hecho, si la comparación de cada sección de diez minutos también es indeterminable, puede ser que se intente una resolución mayor/nivel menor adicional, por ejemplo, secciones de cinco minutos o de un minuto.

30 Una ventaja adicional de esta característica es que, si se identifica una diferencia de calidad en un nivel inferior, entonces el proceso de copia y transferencia se puede implementar solo en esta sección (por ejemplo, diez minutos) en lugar de en el nivel superior (una hora). Por supuesto, esto requerirá mucho menos intercambio de datos y menos carga computacional.

Esta característica puede ser particularmente útil cuando la discrepancia identificada (506) es significativamente mayor que el umbral, pero la determinación de la calidad es indeterminable. En realizaciones, solo se le puede pedir al sistema que investigue (518) en niveles más bajos en los que este es el caso (tal vez a una tasa porcentual por encima del umbral), y para omitir en otros (514).

40 También puede ser que la comparación inicial de las secciones iniciales (una hora, por ejemplo) indique que no existe una diferencia de calidad apreciable en el nivel de una hora, pero que existe un indicador de que los niveles más bajos pueden dar una diferencia de calidad; en dichos casos, se le pedirá al sistema que investigue (518).

45 Adicionalmente, estas características, por supuesto, evitan la necesidad de comparar todos los datos disponibles de toda la sección de datos; por ejemplo, dado que la comparación en el nivel de una hora tiene una resolución más baja en los datos, no existe una comparación de todos los datos que estarían disponibles en la resolución más alta (por ejemplo, un minuto o un segundo). Se puede utilizar la resolución más alta, pero solo cuando se requiera como anteriormente. Esta disposición, que podría considerarse similar a la de un árbol de Fenwick, será mucho más eficiente que la comparación de todos los datos.

50 También se puede observar que, si es solo una porción de resolución más alta (minutos) de una sección de un flujo que se copia sobre el otro flujo, esto puede dar lugar a ciertos tipos de flujo de datos automáticamente en el recálculo en el flujo de datos que cubren la sección más amplia (hora); por lo tanto, el cambio en el nivel más alto/resolución más baja se puede efectuar simplemente al copiar una pequeña parte del flujo.

55 Ciertos conjuntos de datos pueden tener parámetros que, en una resolución de tiempo dada puede indicar características cualitativas o cuantitativas de capas de mayor resolución de tiempo en los datos. Por ejemplo, en una resolución de una hora, un resultado de datos puede tener límites superiores e inferiores, y estos límites pueden parecer iguales en los dos flujos. Sin embargo, puede ser que en un flujo haya un número mucho mayor de eventos

de datos en el primer período de diez minutos que en el otro flujo. Un indicador de esta resolución más alta puede incluirse en los datos de resolución más baja; por ejemplo, los datos de una hora pueden incluir los límites de datos superior e inferior, pero también la cantidad total de eventos de datos en el período de una hora, o incluso la cantidad total de eventos de datos por período de diez minutos. Por lo tanto, una comparación inicial del período de una hora que parece similar puede encontrar un indicador de una diferencia en una resolución más alta, que luego se puede investigar para encontrar que un período de diez minutos tenía límites de datos más altos y más bajos que en su contraparte en el otro flujo.

El uso apropiado de esta característica de comparación de resolución múltiple puede ser determinable sobre la base del período de la sección que se evaluó más recientemente; si la última sección (610) que se comparó tenía una hora de duración, al sistema se le pedirá que continúe en ese nivel de resolución para la siguiente comparación de la sección, y que solo proceda a una resolución más alta si se indica arriba.

En ciertas circunstancias, incluso se puede determinar que los indicadores de diferencias en las resoluciones de mayor tiempo no son necesarios o relevantes. Por ejemplo, si las secciones actuales que se comparan son un día entero, y existe un indicador de que puede haber una diferencia de calidad en el nivel de un minuto, puede ser que dicha diferencia para el sistema dado en ejecución sea tolerable, es decir que los dos conjuntos de datos/flujo independientes en las dos ubicaciones no necesitan ser redundantes a ese nivel para el período actual.

Una vez que se ha realizado la copia e incorporación, o se ha determinado la discrepancia por debajo del umbral, o se ha decidido el árbitro externo o se ha omitido la evaluación de la calidad, se inserta un nuevo marcador (516) al final de la sección actual, y el proceso comienza de nuevo, evaluando las siguientes secciones correspondientes.

En realizaciones adicionales de la invención, puede ser que los datos en los sistemas 304 y 312 de procesamiento/almacenamiento, más que simplemente estar disponibles en su estado de series de tiempo para usuarios o sistemas posteriores, también se puedan empaquetar o volver a emitir en un flujo de datos (ahora modificado) para enviar o transmitir a otro sistema o ubicación. En dichas realizaciones, se puede observar que, si la diferencia identificada entre los flujos en la etapa de comparación es que están fuera de tiempo entre sí (o que se ha recibido un número diferente de paquetes), incluso si se hace un reemplazo, entonces el tiempo/numeración todavía puede ser dispar. Por lo tanto, en el caso de transmitir el flujo modificado, en algunos casos, esta situación se resolverá de forma algo automática, ya que la sección que se copia del flujo que se determina es la de mejor calidad, se puede extender el flujo de menor calidad a una longitud adecuada, por lo que la temporización/numeración ahora coincide. Por ejemplo, si el flujo 2 es corto por un paquete, la sección utilizada como reemplazo del flujo 1 será un paquete más grande que la sección identificada en el flujo 2, por lo que el flujo 2 ahora tendrá la longitud correcta. En situaciones donde esto no sucede automáticamente, el flujo que tiene una sección reemplazada se puede volver a temporizar, o tener cualquier espacio tapado con paquetes nulos, para corregir la disparidad adicional.

Se debe tener en cuenta que el sistema puede, en todas las etapas del proceso mostrado en la Figura 5, registrar estadísticas relacionadas con las comparaciones, las determinaciones de calidad y las acciones de copiado tomadas. Estas estadísticas pueden ayudar a una nueva iteración de los algoritmos utilizados para la comparación y/o la determinación de la calidad.

Como se puede ver en el diagrama de flujo, el proceso ilustrado en la Figura 5 es repetitivo y continuo. En las realizaciones, los flujos 1 y 2 de datos se comparan constantemente, y ninguna parte del flujo se deja sin comparación en ningún punto, ya que cada sección de cada canal de ambos flujos de datos sigue lo anterior a la etapa de comparación en el comparador(es).

Otra ventaja de este método es que las secciones que se comparan entre flujos son cortas. Esto significa que la carga computacional es baja, por lo que los comparadores pueden trabajar en tiempo real, en lugar de tener que pausar o apagar el sistema. Otra ventaja es que las evaluaciones de comparación se pueden mantener simples, como solo verificar la disparidad de tiempo. Esto también permite una carga computacional baja y, por lo tanto, una aplicación en tiempo real.

Volviendo ahora a la Figura 6, en realizaciones de la invención una vez que se ha realizado la evaluación de la sección 608 en el canal A, y se ha completado cualquier reemplazo de o por esa sección, se habrá insertado un nuevo marcador 606 en el canal A, desde el cual se realizará la siguiente comparación en el canal A. El proceso de la Figura 5 luego procede a la evaluación del siguiente canal en el flujo, el canal B. De la misma manera, todos los datos del último marcador 606 en el canal B se comparan ahora con la sección del canal B correspondiente en el flujo 2. Luego de esto, la evaluación se realiza en la última sección del canal C, y así sucesivamente.

Se puede observar que las secciones 608 para los canales A, B y C parecen tener diferentes longitudes, es decir, que la flecha 610 indicadora desde el último marcador hasta "presente" es diferente para cada canal. Esto se debe a que, en las realizaciones de la invención, el comparador toma un período de tiempo estándar (o lo más cercano posible para que se ajuste a un número entero de paquetes de datos en el canal actual) desde el último marcador para cada canal. Dado que la densidad de datos variará dentro de cada canal, y de hecho entre canales, es decir, la carga útil del paquete de datos en los canales puede diferir significativamente en tamaño de uno al siguiente, variará la cantidad de carga útil de datos procesable en el período de tiempo estándar. Tenga en cuenta que las diferencias

que se muestran en la Figura 6 son exageradas para facilitar la referencia; en la práctica, las diferencias entre los tamaños de las cargas útiles de datos serán generalmente pequeñas.

La Figura 7 es un diagrama que ilustra el procesamiento (700) de un solo canal 702 del flujo 1 de datos, y el canal 703 único correspondiente desde el flujo 2 de datos, de acuerdo con una realización de la invención. Como puede verse, esto implica que esencialmente los canales son idénticos; la tasa de paquetes es la misma, y los marcadores (708) están en los mismos lugares. Aquí, los datos del flujo 1 están coloreados en un tono más claro que los datos del flujo 2, para facilitar solo la comprensión. La dirección del 'viaje' de los canales de datos se indica mediante la flecha 704. Cada canal de datos es como se describió anteriormente, está constituido por paquetes de datos y comprende secciones de datos (706, 710, 712) que se comparan entre sí, como se describió anteriormente.

Como se puede ver en la Figura, se comparan las primeras secciones de los canales, sin que se necesario el copiado o reemplazo. En la sección 710 en los datos del canal del flujo 2, se compararon las secciones, se identificó una disparidad y los datos del flujo 2 se identificaron como superiores, por lo que esta sección de los datos del canal del flujo 2 se copia (707) y se incorpora a los canales de datos del flujo 1. De manera similar, más tarde, una sección 712 de los datos del flujo 1 se copia e incorpora en los datos del canal del flujo 2. Tenga en cuenta que es probable que los reemplazos de secciones de canales en flujos de datos típicos ocurran con mucha menos frecuencia que lo indicado en este documento.

La Figura 8 es un diagrama que ilustra la parte (800) del flujo 1 de datos procesado de acuerdo con realizaciones de la invención. El flujo de datos, nuevamente compuesto por los canales A, B, C y D a n (802) y secciones dentro de esos canales (804) subdivididos por los marcadores 806, se ha procesado como se describió anteriormente. Por lo tanto, ciertas secciones (808), nuevamente indicadas por el sombreado más oscuro, han sido reemplazadas por las secciones correspondientes del flujo 2 de datos.

Una ventaja adicional de la invención es que, dado que los canales en cada flujo de datos se tratan por separado, Se puede realizar una granularidad más fina de comparación y reemplazo de secciones problemáticas. Esto significa que en un flujo de datos que de otro modo podría ser de muy alta calidad, se puede reemplazar una pequeña sección de un canal de datos del flujo, mientras que para el resto del flujo este flujo se utilizará generalmente para proporcionar reemplazos al otro flujo. Esto contrasta con los métodos considerados anteriormente en los cuales se reconstruirían flujos completos utilizando todo el contenido de otro flujo.

Las realizaciones de la invención como se describió anteriormente comprenderán componentes de hardware computacional y dispositivos procesadores. Estos pueden incluir software y/o hardware para proporcionar la funcionalidad y características descritas en este documento. Por lo tanto, los dispositivos pueden incluir uno o más matrices lógicas, memorias, circuitos analógicos, circuitos digitales, software, firmware y procesadores. Los componentes de hardware y firmware de los dispositivos pueden incluir varias unidades especializadas, circuitos, software e interfaces para proporcionar la funcionalidad y características descritas en este documento. Los procesadores pueden ser o incluir uno o más microprocesadores, circuitos integrados de aplicación específica (ASIC), dispositivos lógicos programables (PLD) y matrices lógicas programables (PLA).

El almacenamiento de datos se puede proporcionar por la memoria del ordenador, que puede ser o incluir RAM, ROM, DRAM, SRAM y MRAM, y puede incluir firmware, tal como datos estáticos o instrucciones fijas, BIOS, funciones del sistema, datos de configuración y otras rutinas utilizadas durante el funcionamiento de los dispositivos y procesadores. El almacenamiento de datos de gran volumen o a largo plazo puede tomar la forma de un disco, cinta, DVD u otro medio de almacenamiento en serie direccionable o de capacidad razonablemente alta. Se pueden proporcionar múltiples dispositivos de almacenamiento o estar disponibles para los dispositivos.

Los procesadores y/o controladores pueden comprender uno o más procesadores computacionales, y/o elementos de control que tienen uno o más procesadores electrónicos. Por lo tanto, se debe considerar que los usos del término "procesador" o "controlador" en el presente documento se refieren a un solo procesador, controlador o elemento de control, o a una pluralidad del mismo; cuyas pluralidades pueden operar en concierto para proporcionar las funciones descritas. Adicionalmente, las funciones individuales y/o separadas de los procesadores o controladores pueden ser alojadas o realizadas en diferentes unidades de control, procesadores o controladores.

Para configurar un procesador o controlador, se puede proporcionar un conjunto adecuado de instrucciones que, cuando se ejecutan, hacen que dicha unidad de control o dispositivo computacional implementen las técnicas especificadas en este documento. El conjunto de instrucciones se puede incrustar adecuadamente en dicho uno o más procesadores electrónicos. Alternativamente, el conjunto de instrucciones se puede proporcionar como software para ejecutarse sobre el dispositivo.

Se puede observar que, como los componentes del comparador son tan eficientes, al procesar pequeñas secciones de datos repetidamente y al ciclizar a través de los canales de datos en los flujos de datos, estos componentes (o aplicaciones que ejecutan estos componentes) pueden ser comparativamente pequeños, y ocupan pequeñas cantidades de recursos del sistema. Adicionalmente, las comparaciones indicadas anteriormente se pueden ejecutar al aplicar instancias separadas de un componente comparador a las secciones de un flujo de datos; por ejemplo, tres comparadores pueden trabajar cada uno en un tercio del flujo de datos. De esta manera, la duración del ciclo

necesaria para procesar todo el flujo se puede reducir aplicando más comparadores. Esto, a su vez, puede reducir el período (610) para cada canal de datos desde el último marcador, ya que este período generalmente se determina por el tiempo que toma el comparador(es) para realizar un ciclo a través de la comparación de cada uno de los canales de datos con sus contrapartes en el otro flujo.

- 5 Se pueden hacer muchas modificaciones a los ejemplos anteriores sin apartarse del alcance de la presente invención como se define en las reivindicaciones adjuntas.

Dentro del alcance de esta solicitud, se pretende expresamente que los diversos aspectos, realizaciones, ejemplos y alternativas establecidos en los párrafos anteriores, en las reivindicaciones y/o en la siguiente descripción y dibujos, y en particular las características individuales de los mismos, se pueden tomar independientemente o en cualquier combinación. Es decir, todas las realizaciones y/o características de cualquier realización pueden combinarse de cualquier manera y/o combinación, a menos que dichas características sean incompatibles. El solicitante se reserva el derecho de cambiar cualquier reivindicación presentada originalmente o presentar cualquier reivindicación nueva de acuerdo con lo anterior, que incluye el derecho de enmendar cualquier reivindicación presentada originalmente para que dependa y/o incorpore cualquier característica de cualquier otra reivindicación, aunque no haya sido reivindicada originalmente de esa manera.

10

15

REIVINDICACIONES

1. Un sistema (300) para sincronización de primer (302, 304) y segundo (310, 312) sistemas de manejo de datos independientes, cada uno de los primero y segundo sistemas de manejo de datos que recibe datos desde la misma pluralidad de proveedores (301) de datos, el primer y segundo sistemas de manejo de datos que procesan primer (302) y segundo (310) flujos (600) de datos generados desde los datos recibidos desde la pluralidad de proveedores de datos, en la que se disponen el primer y segundo sistemas de manejo de datos en diferentes ubicaciones, y en la que el primer y segundo flujos de datos no son exactamente idénticos, pero son lo suficientemente idénticos para soportar el propósito de la redundancia de proporcionar los dos flujos de datos separados en las dos ubicaciones, el sistema comprende:
- 5 a) un primer sistema (302, 304) de manejo de datos en una primera ubicación que comprende:
- un primer módulo (302) de generación de flujo de datos, que comprende un primer compilador para combinar datos recibidos desde la pluralidad de proveedores de datos en un primer flujo (600) de datos; y un primer sistema (304) de procesamiento y almacenamiento, que comprende:
- un primer almacén (402) de datos para almacenar el primer flujo de datos;
- 15 un primer módulo (404) duplicador para duplicar datos desde el primer flujo de datos;
- un primer módulo (406) combinador para combinar datos copiados desde el segundo flujo de datos con el primer flujo de datos; y por lo menos un primer puerto de comunicaciones para transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones;
- 20 b) un segundo sistema (310, 312) de manejo de datos en una segunda ubicación remota desde la primera ubicación, que comprende:
- un segundo módulo de generación de flujo de datos (310), que comprende un segundo compilador para combinar datos recibidos desde la pluralidad de proveedores de datos en un segundo flujo (600) de datos; y
- un segundo sistema (312) de procesamiento y almacenamiento, que comprende:
- un segundo almacén (402) de datos para almacenar el segundo flujo de datos;
- 25 un segundo módulo (404) duplicador para duplicar datos desde el segundo flujo de datos;
- un segundo módulo (406) combinador para combinar datos copiados desde el primer flujo de datos con el segundo flujo de datos; y
- por lo menos un segundo puerto de comunicaciones para transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones;
- 30 y
- c) por lo menos un módulo (306, 314) compactador para comparar la información desde el primer y segundo flujos de datos,
- en el que el primer y segundo sistemas de procesamiento y almacenamiento se configuran para, mediante por lo menos un primer y segundo puertos de comunicaciones, transmitir información a y desde por lo menos un módulo compactador, y para transmitir datos entre el primer y segundo sistemas de procesamiento y almacenamiento, a través de la red de comunicaciones,
- 35 y en la que:
- i) el primer y segundo módulos de generación de flujo de datos se configuran para generar primer y segundo flujos de datos para procesamiento y almacenamiento en los respectivos primer y segundo sistemas de procesamiento y almacenamiento;
- 40 ii) luego de recepción del primer y segundo flujos de datos en el primer y segundo sistemas de procesamiento y almacenamiento, el primer sistema de procesamiento y almacenamiento se configura para transmitir (502) primera información desde una sección (608) más reciente del primer flujo de datos hasta el módulo compactador, y el segundo sistema de procesamiento y almacenamiento se configura para transmitir segunda información desde una sección más reciente correspondiente del segundo flujo de datos hasta el módulo compactador;
- 45 iii) el módulo compactador se configura para:
- comparar (504) la primera y segunda información; medir un valor para discrepancia entre la primera y segunda información; probar (506) el valor de discrepancia contra un umbral; y, en el que el valor de discrepancia excede el umbral, generar:

- desde la primera y segunda información un indicador (508) del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contiene una mayor calidad de datos; y
- una instrucción (422) de duplicación para el módulo duplicador del respectivo sistema de procesamiento y almacenamiento cuyo flujo de datos se ha indicado por el indicador para contener la mayor calidad de datos;
- 5 iv) el módulo duplicador para el flujo de datos indicado se configura para, si se indica, duplicar (422) los datos desde dicha sección más reciente del flujo de datos indicado;
- v) el sistema de procesamiento y almacenamiento para el flujo de datos indicado se configura para transmitir, a través del puerto de comunicaciones, los datos duplicados al módulo combinador del sistema de procesamiento y almacenamiento para el otro flujo; y
- 10 vi) el módulo combinador para el otro flujo se configura para reemplazar (424, 707) dicha sección más reciente correspondiente del otro flujo con los datos duplicados desde dicha sección (710) más reciente del flujo indicado.
2. El sistema de la Reivindicación 1, en el que el módulo compactador se configura para, luego de las etapas de comparación, prueba, medición y cualquier etapa de generación, insertar en cada uno del primer y segundo flujos de datos un marcador (606) de punto de tiempo que indica el fin de la sección actual de los flujos que se comparan,
- 15 y en el que el primer y segundo sistemas de procesamiento y almacenamiento se configuran para definir dicha sección más reciente del respectivo flujo de datos entre un marcador de punto de tiempo más reciente en el respectivo flujo de datos y un punto de tiempo actual.
3. El sistema de la Reivindicación 1 o Reivindicación 2, en el que la primera información desde la sección más reciente del primer flujo de datos, y la segunda información desde la sección más reciente correspondiente del segundo flujo de datos, transmitida al módulo compactador, comprende un subconjunto de los datos contenidos en aquellas secciones de los respectivos flujos de datos.
- 20 4. El sistema de cualquier reivindicación precedente, en el que:
- el primer y segundo flujos de datos se subdividen cada uno en una pluralidad de canales (602) de datos, en la que la subdivisión en el primer flujo de datos corresponde a la subdivisión en el segundo flujo de datos,
- 25 y en la que dicha sección más reciente del primer flujo de datos es una sección más reciente de un primer canal de datos de la pluralidad de canales del primer flujo de datos, y dicha sección más reciente correspondiente del segundo flujo de datos es una sección más reciente correspondiente de un primer canal de datos correspondiente de la correspondiente pluralidad de canales del segundo flujo de datos.
- 30 5. Los sistemas de la Reivindicación 4, en los que después de las etapas (ii) a (vi), la siguiente sección más reciente del primer flujo de datos para comparación es la sección más reciente de un segundo canal de datos de la pluralidad de canales de datos del primer flujo de datos, y la siguiente sección más reciente correspondiente del segundo flujo de datos para comparación es la sección más reciente correspondiente de un segundo canal de datos correspondiente de la correspondiente pluralidad de canales del segundo flujo de datos.
- 35 6. El sistema de cualquier reivindicación precedente, en el que el módulo compactador se configura para comparar la primera y segunda información, medir el valor para discrepancia, y probar el valor contra el umbral, utilizando una jerarquía de secuencias de filtros de discrepancia, la jerarquía de secuencias de los filtros de discrepancia que procede desde una comparación de discrepancia gruesa hasta una comparación de discrepancia fina.
- 40 7. El sistema de cualquier reivindicación precedente, en el que el módulo compactador se configura, para generación del indicador (508) del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contiene la mayor calidad de datos, para:
- comparar dicha primera y segunda información utilizando cada una de una pluralidad de factores de comparación cualitativos; y
- utilizar los resultados de la pluralidad de comparaciones para generar el indicador.
- 45 8. El sistema de la Reivindicación 7, en el que la pluralidad de factores de comparación cualitativos comprende una jerarquía de factores de comparación.
9. El sistema de cualquier reivindicación precedente, en el que el módulo compactador es operable para comparar la primera y segunda información en dos o más resoluciones de la primera y segunda información, con el fin de medir la discrepancia y/o generar el indicador de calidad.
- 50 10. El sistema de la Reivindicación 9, en el que las dos o más resoluciones de la primera y segunda información son resoluciones diferentes en tiempo de la primera y segunda información.

11. El sistema de la Reivindicación 10, en el que el módulo compactador es operable para:

comparar un primer periodo de tiempo de la primera información, desde dicha sección más reciente del primer flujo de datos, y un primer periodo de tiempo correspondiente de la segunda información, desde dicha sección más reciente correspondiente del segundo flujo de datos; y

5 después de la primera comparación del periodo de tiempo, comparar un primer subperiodo del primer periodo de tiempo de la primera información con un primer subperiodo correspondiente del primer periodo de tiempo correspondiente de la segunda información.

10 12. Un método para sincronizar el primer (302, 304) y segundo (310, 312) sistemas de manejo de datos independientes, cada uno de los primero y segundo sistemas de manejo de datos que recibe datos desde la misma pluralidad de proveedores (301) de datos, el primer y segundo sistemas de manejo de datos que procesan primer (302) y segundo (310) flujos (600) de datos generados desde los datos recibidos desde la pluralidad de proveedores de datos, en la que el primer y segundo sistemas de manejo de datos se disponen en diferentes ubicaciones, y en el que el primer y segundo flujos de datos no son exactamente idénticos, pero son lo suficientemente idénticos para soportar el propósito de la redundancia de proporcionar los dos flujos de datos separados en las dos ubicaciones, el método comprende:

15 a) en un primer sistema (302, 304) de manejo de datos en una primera ubicación:

combinar datos recibidos desde la pluralidad de proveedores de datos en un primer flujo (600) de datos, mediante un primer módulo (302) de generación de flujo de datos que comprende un primer compilador; y

en un primer sistema (304) de procesamiento y almacenamiento:

20 almacenar el primer flujo de datos en un primer almacén (402) de datos;

duplicar datos desde el primer flujo de datos en un primer módulo (404) duplicador;

combinar datos copiados desde el segundo flujo de datos con el primer flujo de datos en un primer módulo (406) combinador; y

25 transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones mediante por lo menos un primer puerto de comunicaciones;

b) en un segundo sistema (310, 312) de manejo de datos en una segunda ubicación remota desde la primera ubicación:

combinar datos recibidos desde la pluralidad de proveedores de datos en un segundo flujo (600) de datos mediante un segundo módulo (310) de generación de flujo de datos, que comprende un segundo compilador; y

30 en un segundo sistema (312) de procesamiento y almacenamiento:

almacenar el segundo flujo de datos en un segundo almacén (402) de datos;

duplicar datos desde el segundo flujo de datos en un segundo módulo (404) duplicador;

combinar datos copiados desde el primer flujo de datos con el segundo flujo de datos en un segundo módulo (406) combinador; y

35 transmitir información y/o datos a y recibir información y/o datos desde una red (308) de comunicaciones mediante por lo menos un segundo puerto de comunicaciones;

y

c) mediante por lo menos un módulo (306, 314) compactador, comparar la información desde el primer y segundo flujos de datos,

40 el método comprende adicionalmente, mediante el primer y segundo sistemas de procesamiento y almacenamiento, a través de por lo menos un primer y segundo puertos de comunicaciones, transmitir información a y desde por lo menos un módulo compactador, y transmitir datos entre el primer y segundo sistemas de procesamiento y almacenamiento, a través de la red de comunicaciones,

el método aún comprende adicionalmente:

45 i) mediante el primer y segundo módulos de generación de flujo de datos, generar primer y segundo flujos de datos para procesamiento y almacenamiento en los respectivos primer y segundo sistemas de procesamiento y almacenamiento;

- 5 ii) luego de recepción del primer y segundo flujos de datos en el primer y segundo sistemas de procesamiento y almacenamiento, mediante el primer sistema de procesamiento y almacenamiento, transmitir (502) primera información desde una sección (608) más reciente del primer flujo de datos hasta el módulo compactador, y mediante el segundo sistema de procesamiento y almacenamiento, transmitir segunda información desde una sección más reciente correspondiente del segundo flujo de datos hasta el módulo compactador;
- 10 iii) mediante el módulo compactador: comparar (504) la primera y segunda información; medir un valor para discrepancia entre la primera y segunda información; probar (506) el valor de discrepancia contra un umbral; y, en el que el valor de discrepancia excede el umbral, generar:
desde la primera y segunda información un indicador (508) del cual dicho primer flujo de datos sección más reciente y dicho segundo flujo de datos sección más reciente correspondiente contienen una mayor calidad de datos; y
una instrucción (422) de duplicación para el módulo duplicador del respectivo sistema de procesamiento y almacenamiento cuyo flujo de datos se ha indicado por el indicador para contener la mayor calidad de datos;
- 15 iv) mediante el módulo duplicador para el flujo de datos indicado, si se indica, duplicar (422) los datos desde dicha sección más reciente del flujo de datos indicado;
- v) mediante el sistema de procesamiento y almacenamiento para el flujo de datos indicado, transmitir, a través del puerto de comunicaciones, los datos duplicados al módulo combinador del sistema de procesamiento y almacenamiento para el otro flujo; y
- 20 vi) mediante el módulo combinador para el otro flujo, reemplazar (424, 707) dicha sección más reciente correspondiente del otro flujo con los datos duplicados desde dicha sección (710) más reciente del flujo indicado.

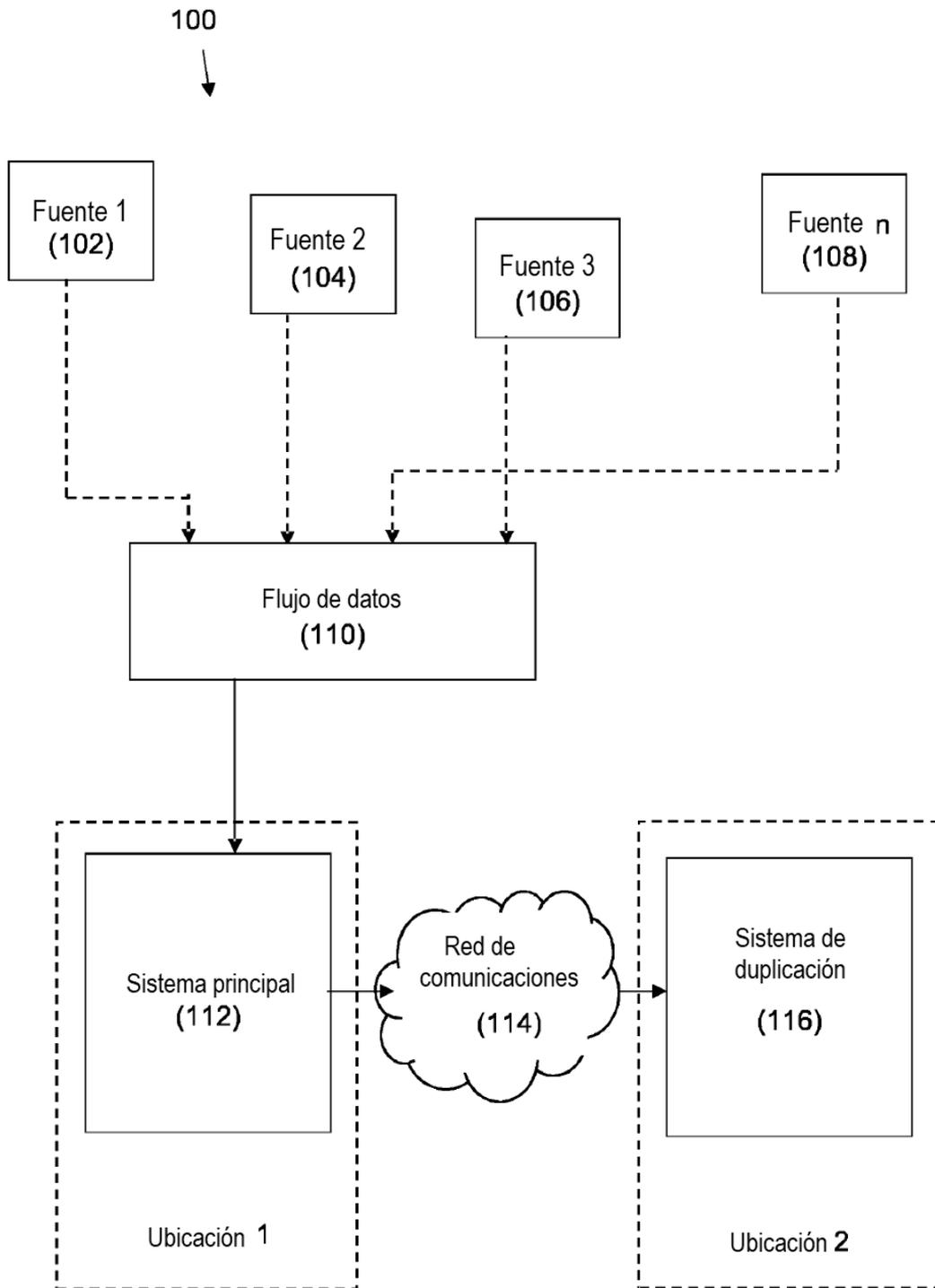
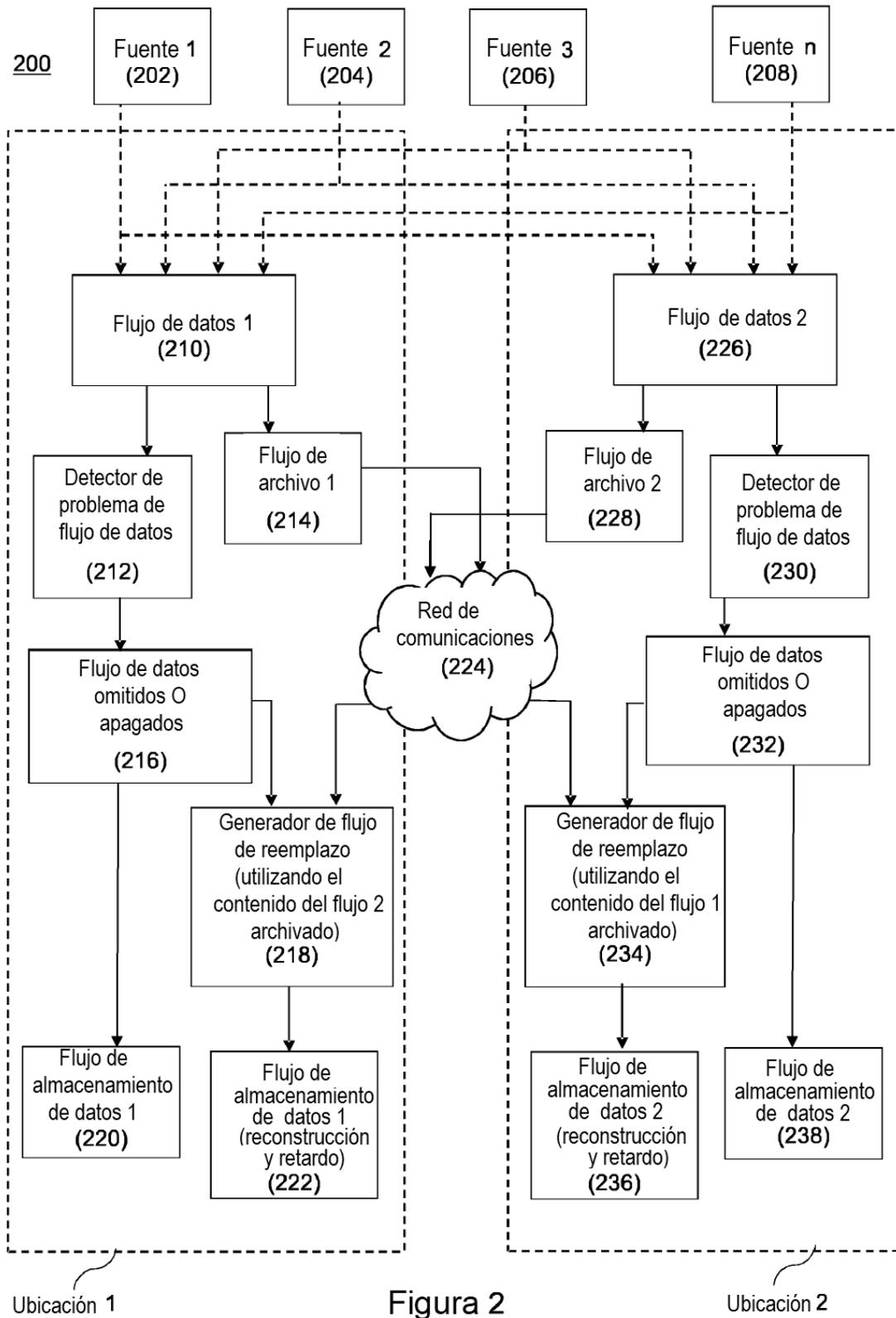


Figura 1



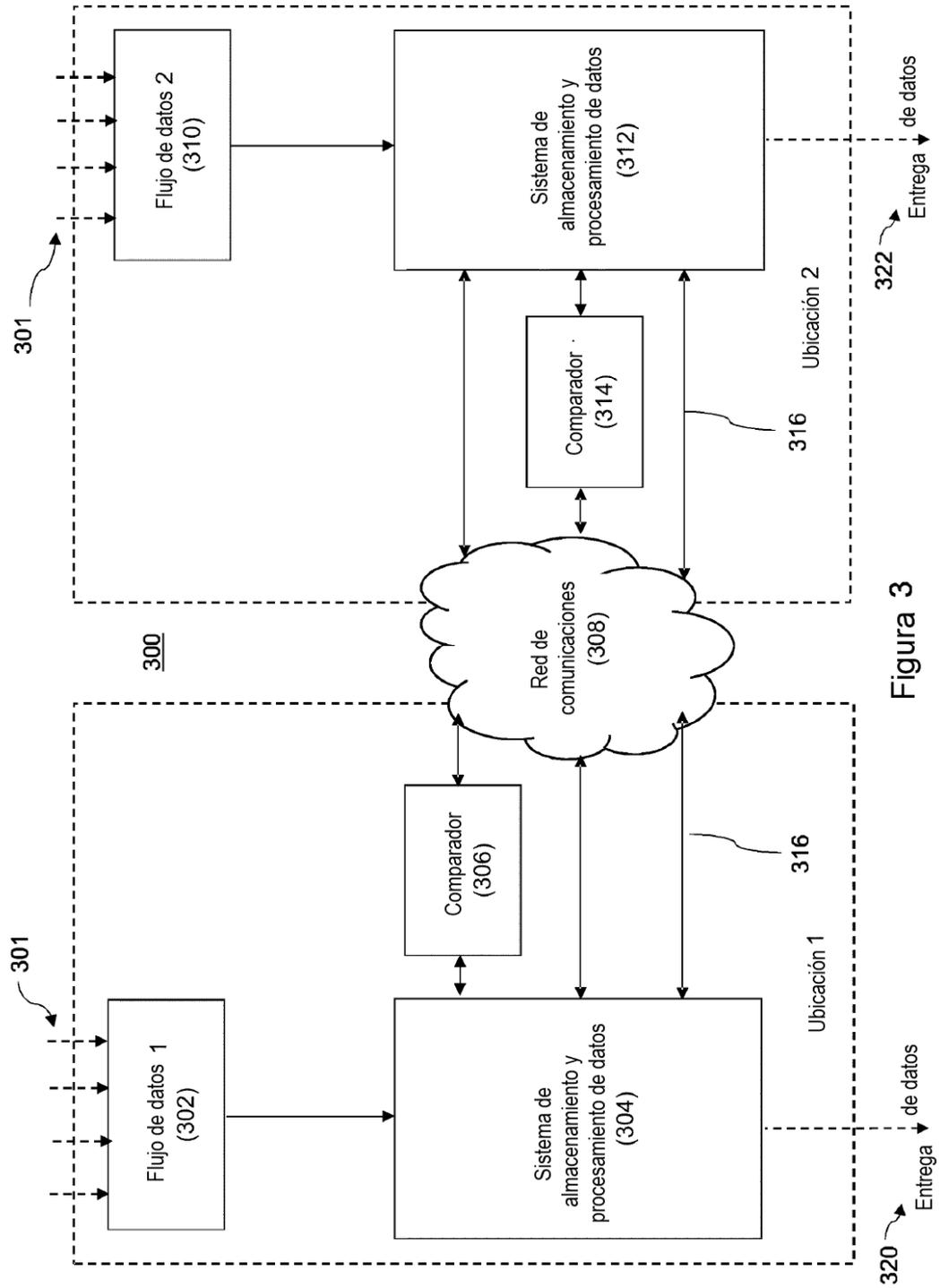


Figura 3

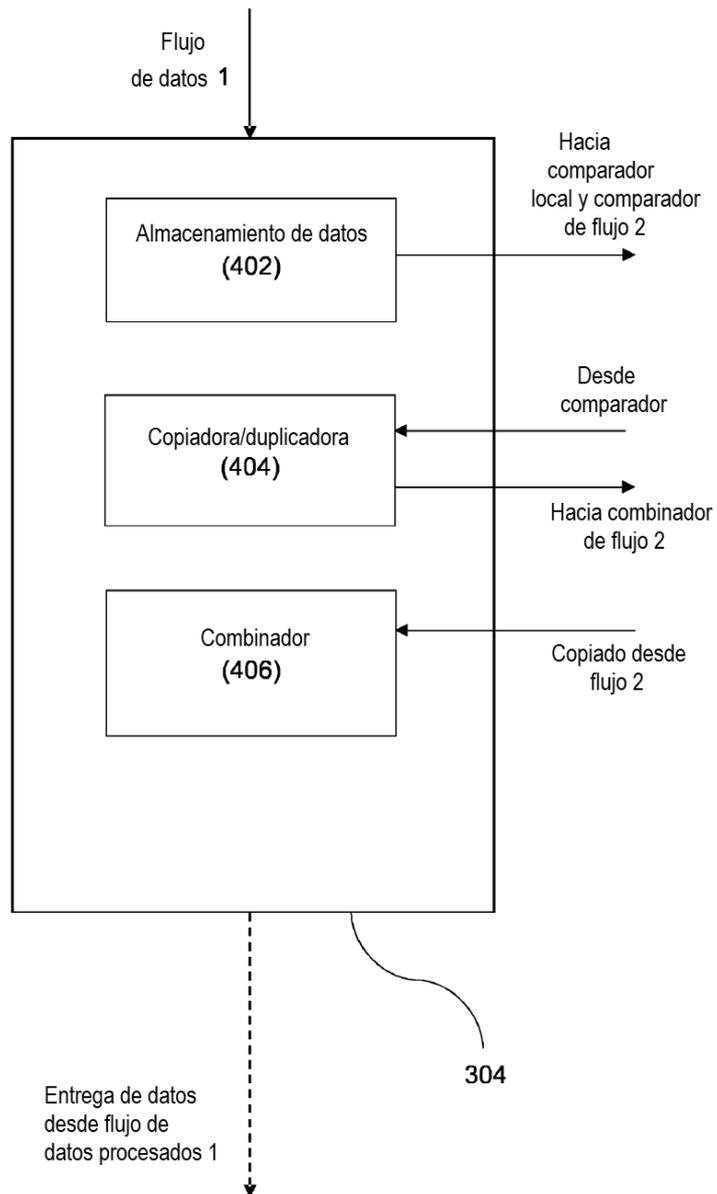


Figura 4a

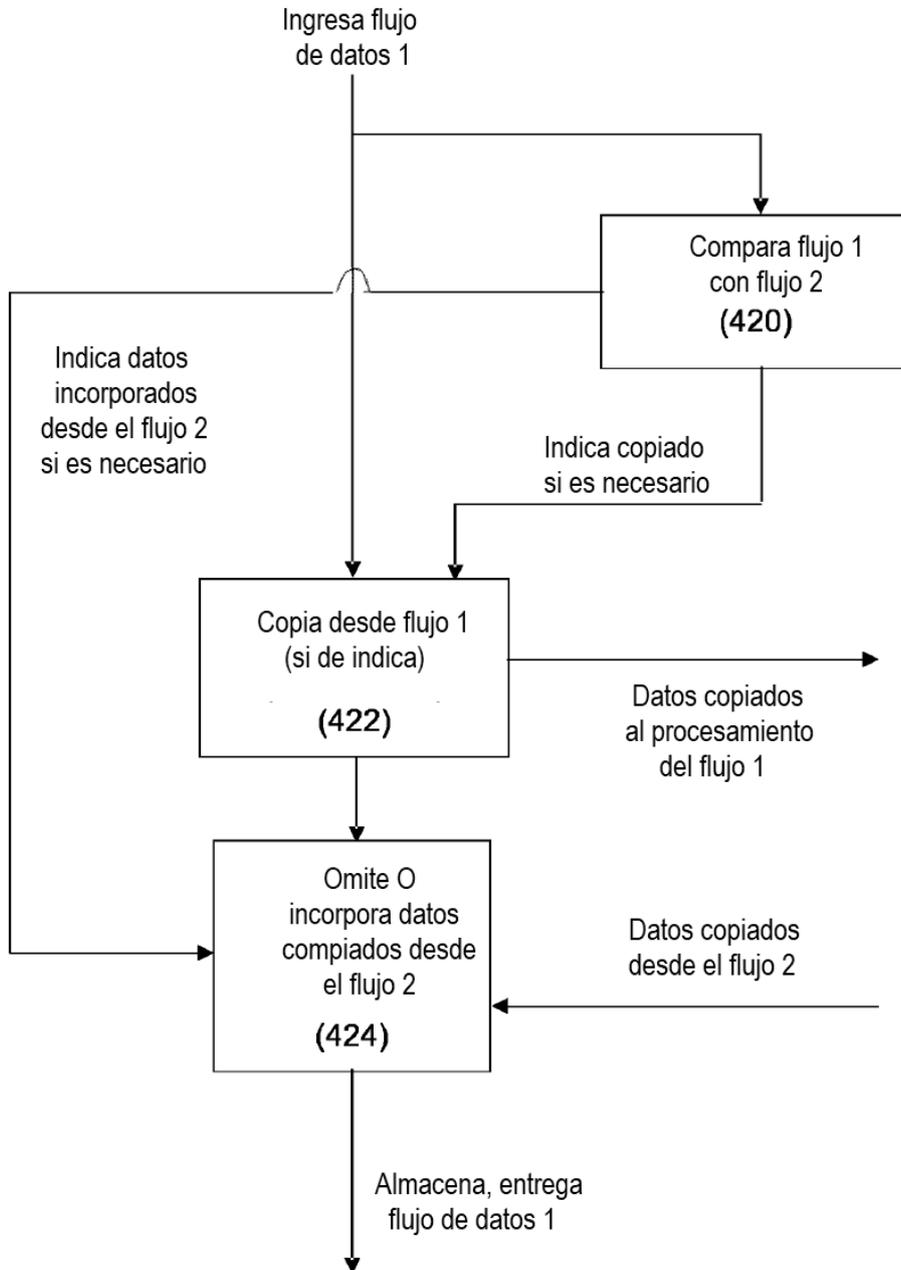


Figura 4b

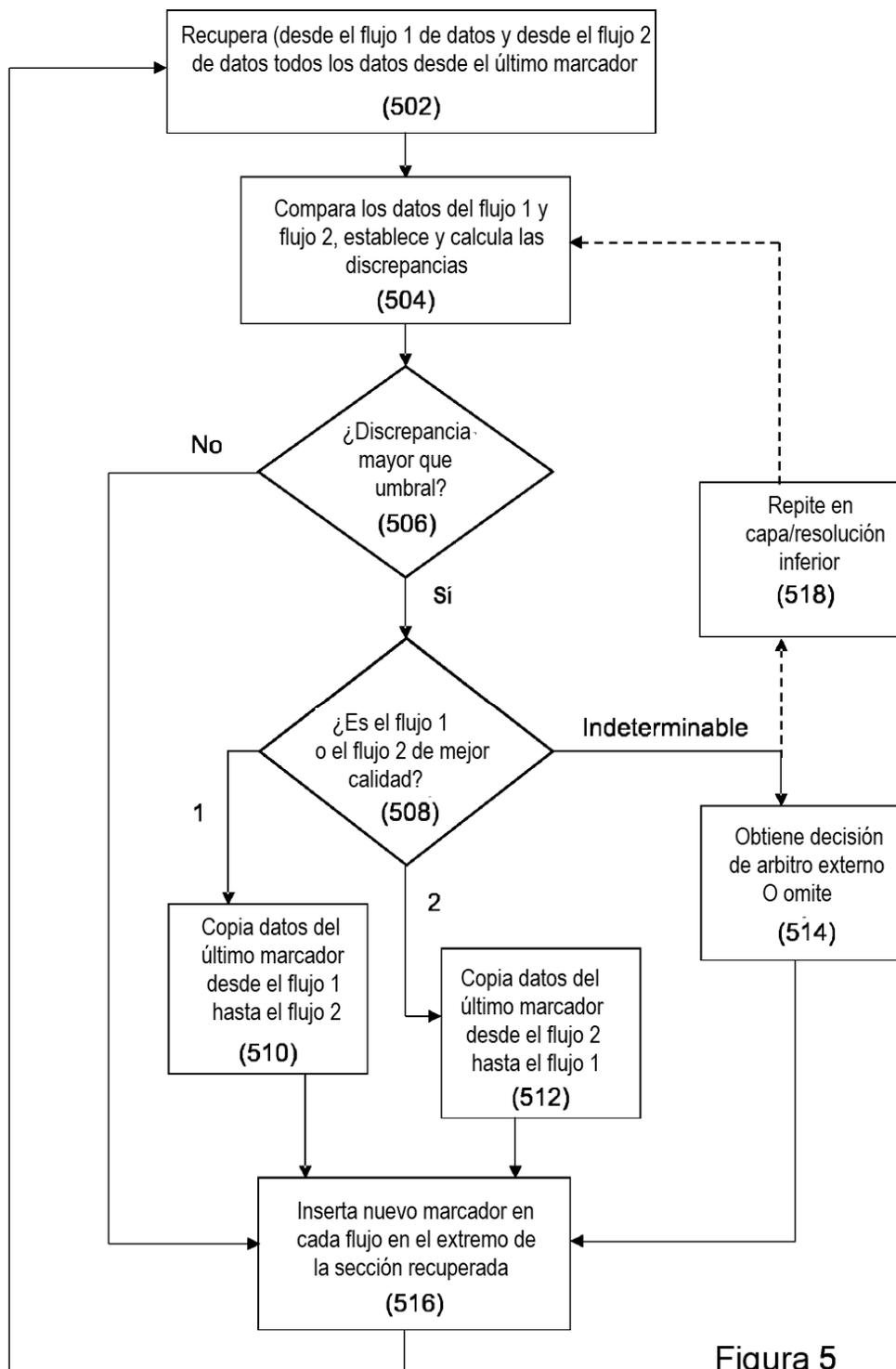


Figura 5

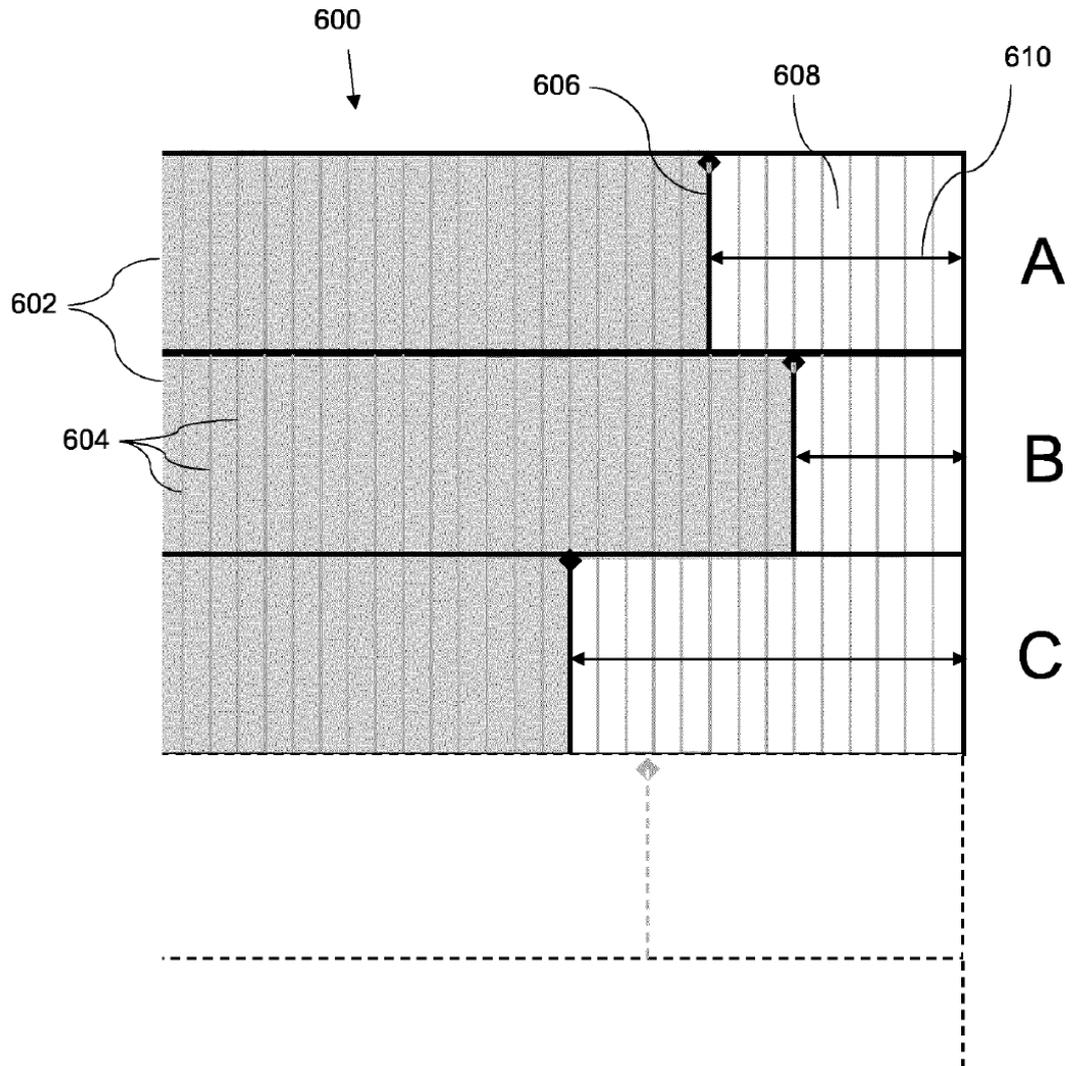


Figura 6

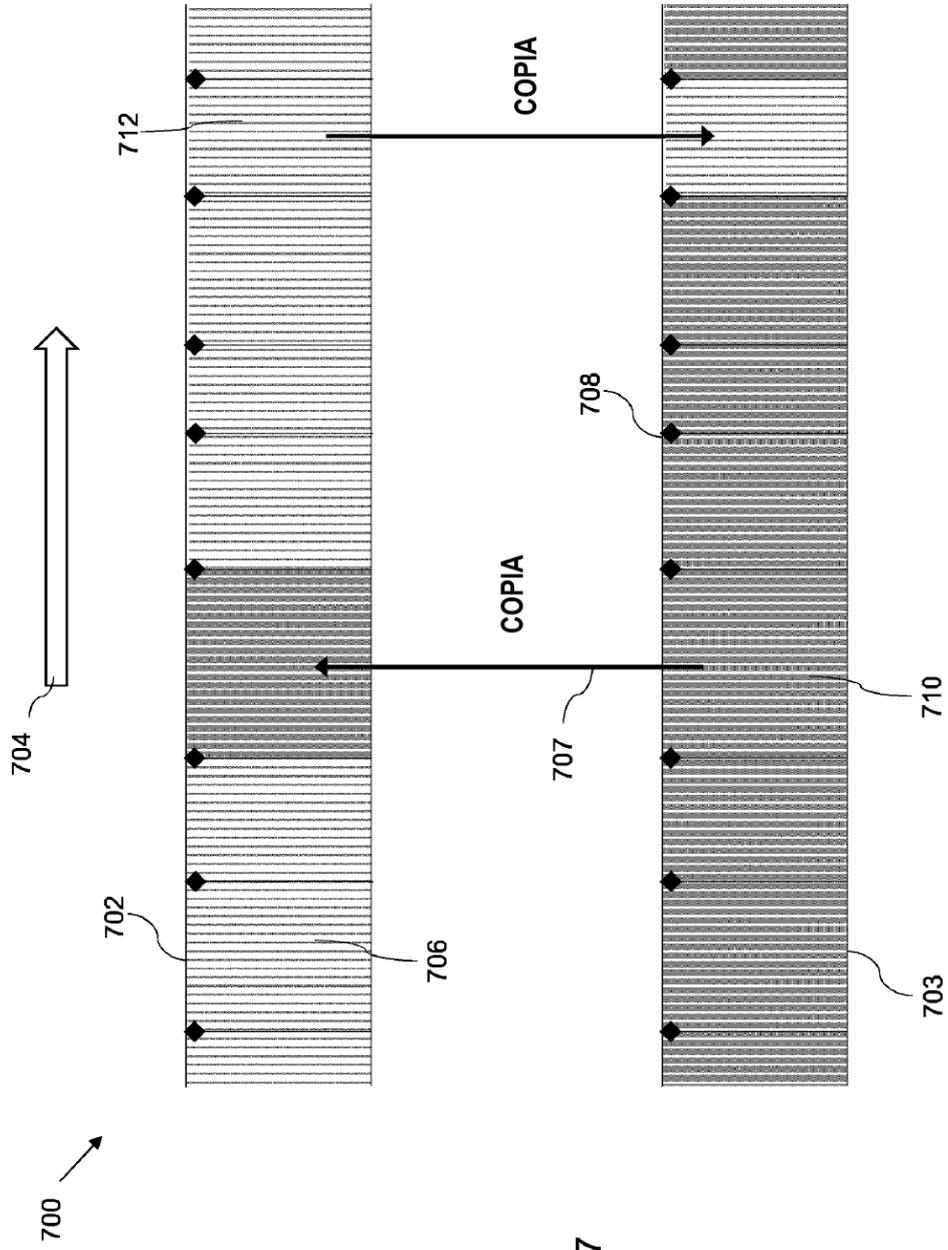


Figura 7

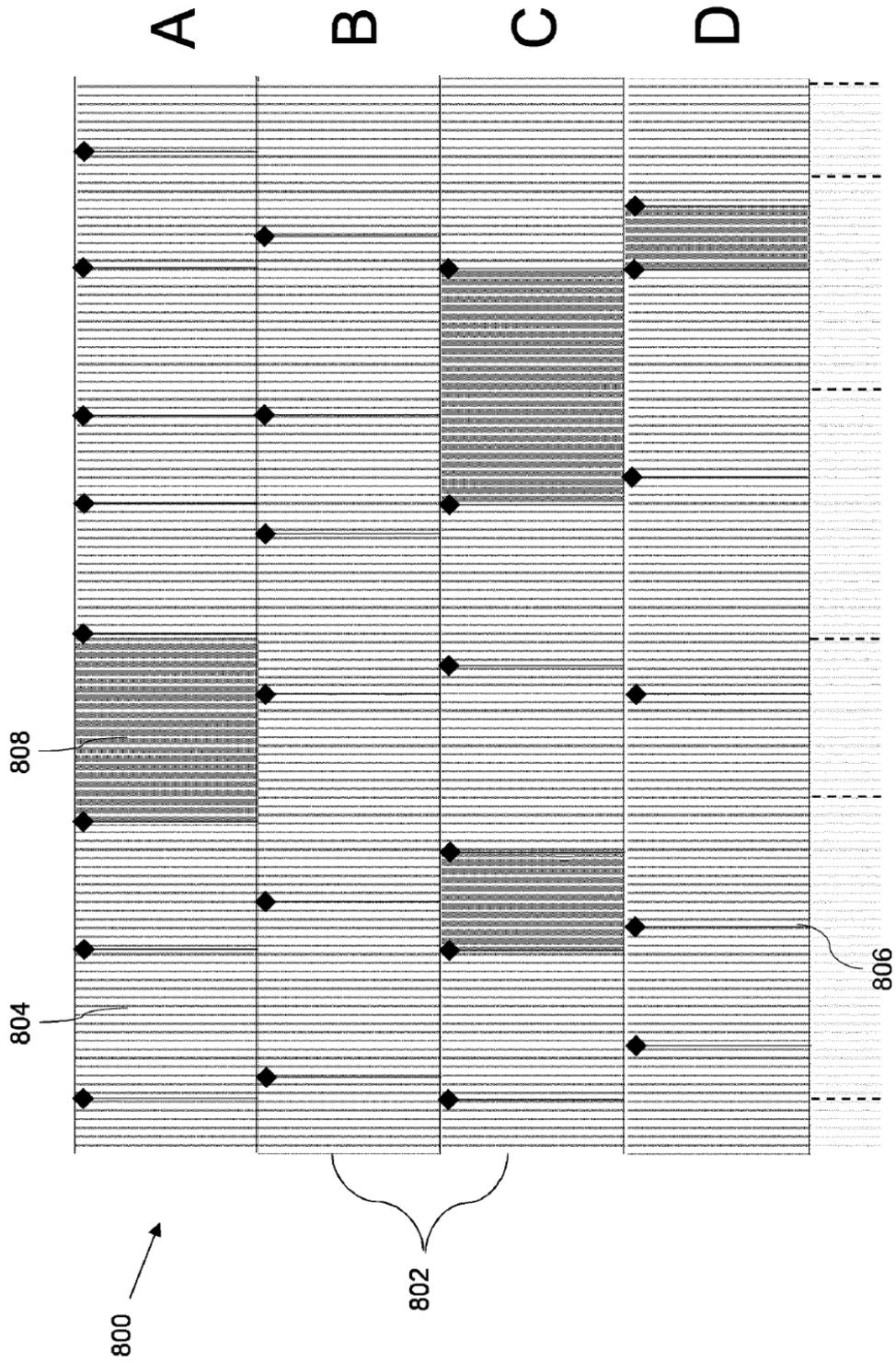


Figure 8