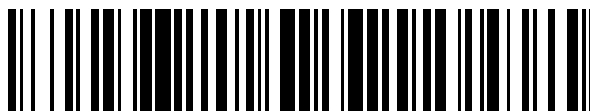


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 693 546**

51 Int. Cl.:

**H04L 29/12** (2006.01)

**H04L 12/24** (2006.01)

**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.06.2012 PCT/FR2012/051389**

87 Fecha y número de publicación internacional: **03.01.2013 WO13001208**

96 Fecha de presentación y número de la solicitud europea: **20.06.2012 E 12734972 (8)**

97 Fecha y número de publicación de la concesión europea: **01.08.2018 EP 2727319**

54 Título: **Procedimiento de asignación de direcciones lógicas a puertos de conexión de equipos de un clúster de servidores, programa de ordenador y clúster de servidores correspondientes**

30 Prioridad:

**29.06.2011 FR 1155841**

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

**12.12.2018**

73 Titular/es:

**BULL SAS (100.0%)  
Rue Jean Jaurès  
78340 Les Clayes-sous-Bois, FR**

72 Inventor/es:

**BOZGA, LIANA;  
PEYRARD, JOHANN y  
MISSIMILLY, THIERRY**

74 Agente/Representante:

**AZNÁREZ URBIETA, Pablo**

**ES 2 693 546 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**

Procedimiento de asignación de direcciones lógicas a puertos de conexión de equipos de un clúster de servidores, programa de ordenador y clúster de servidores correspondientes

5 La presente invención se refiere a un procedimiento de asignación de direcciones lógicas a puertos de conexión de equipos de un clúster de servidores. También se refiere a un programa de ordenador para la implementación de este procedimiento y un clúster de servidores.

10 Se conocen clústers de servidores que comprenden especialmente varios nodos de cálculo interconectados entre sí. Los clústers de servidores de este tipo son instalaciones informáticas que generalmente comprenden varios ordenadores en red, con un almacenamiento compartido, que aparecen desde el exterior como un calculador de muy alta potencia de cálculo, conocido como calculador de alto rendimiento de procesamiento o calculador HPC (del inglés "High Performance Computing"). Estas instalaciones optimizadas permiten la distribución de procesamientos complejos y/o cálculos paralelos en al menos una parte de los nodos de cálculo.

15 Algunos clústers de servidores, entre los más sencillos, pueden comprender equipos homogéneos que siguen un mismo protocolo de identificación, de manera que estos elementos puedan ser identificados automáticamente cuando se enciende la instalación, para una correcta inicialización y administración del clúster. Desafortunadamente, esto no es el caso de la mayoría de los clústers de servidores complejos que existen en la actualidad, con capacidades de cálculo muy altas, para los cuales es necesario generar una base de datos que recoja el conjunto de los equipos heterogéneos y de los parámetros del clúster de servidores. Esta base de datos representa entonces el referencial único de la configuración y del estado del clúster de servidores.

20 Por equipos de clúster de servidores se entiende los nodos antes mencionados, pero también los conmutadores (generalmente llamados "switch"), las bahías informáticas de almacenamiento de datos, los administradores de hardware, las alimentaciones eléctricas administrativas, o de manera más general todos los elementos que pueden conectarse entre sí utilizando puertos de conexión que requieren una dirección lógica, especialmente una dirección IP, para funcionar. Así, por ejemplo, para un clúster de servidores de 5000 nodos, se necesitan normalmente unas 15000 direcciones lógicas.

30 Una dificultad mayor es la que consiste en llenar esta base de datos con todas las informaciones necesarias para administrar el clúster de servidores, incluyendo especialmente estas informaciones direcciones lógicas para asignar a los puertos de conexión de los equipos del clúster de servidores. Las informaciones mínimas necesarias comprenden además datos estáticos de descripción lógica y de hardware de los equipos del clúster de servidores y de sus interrelaciones, tales como por ejemplo una descripción del hardware, una localización geográfica de los servidores y equipos del clúster en un centro de cálculo, un estado de las herramientas informáticas instaladas, los datos de explotación del clúster o un estado del hardware.

35 Para rellenar la base de datos, a menudo definida bajo la forma de base de datos relacional, generalmente se suelen definir solicitudes de manipulación de esta base de datos. Por defecto, se redactan manualmente en forma de líneas de código agrupadas en uno o varios archivos, que pueden alcanzar varios miles de líneas para los clústers de servidores complejos. Puede llevar varios meses revisar los documentos técnicos que definen un clúster de servidores, incluyendo la arquitectura y el cableado del clúster, y redactar estas solicitudes de manipulación de la base de datos. Además, la redacción no suele estar estructurada según un orden preestablecido, lo que la hace aún más difícil y larga. Por último, la redacción manual de las solicitudes de manipulación es una fuente de errores en la introducción y requiere múltiples comprobaciones de coherencia.

45 La patente francesa publicada con el número FR 2 931 970 B1 divulga un procedimiento que permite resolver este problema mediante la automatización de la generación de al menos un archivo de solicitudes de manipulación de la base de datos de administración de un clúster de servidores para su inicialización.

Dicho procedimiento comprende, entre otras, las siguientes etapas:

- 50 - definición de una política de direccionamiento lógico de puertos de conexión de equipos del clúster,
- a partir de una distribución lógica de los equipos de clúster en la red de transmisión de datos, de una distribución geográfica y de una definición material de los equipos del clúster, inicialización de una base de datos de administración del clúster de servidores,
- en función además de la política de direccionamiento lógico, asignación de direcciones lógicas a dichos puertos de conexión de equipos del clúster de servidores, y

- registro de las direcciones lógicas asignadas en la base de datos de administración del clúster de los servidores.

Más precisamente, según este procedimiento, la asignación de direcciones lógicas a los puertos de conexión de equipos del clúster de servidores y su registro se realiza durante la inicialización de la base de datos de administración. Por lo tanto, esta inicialización se hace no sólo sobre la base de la distribución lógica, de la distribución geográfica y de la definición material de los equipos del clúster, sino que también de acuerdo con la política de direccionamiento lógico. Este procedimiento aprovecha la definición de una política de direccionamiento de los puertos de conexión de los equipos del clúster de servidores para estructurar las etapas de generación de un conjunto de parámetros de clúster que, tras aplicar la política de direccionamiento a los puertos de conexión de los equipos del clúster de servidores, permite contemplar una generación más fácil de un archivo de solicitud de manipulación de la base de datos que permite su inicialización.

Pero esta solución no es adecuada para clústers de servidores cuyo rendimiento puede actualmente alcanzar y superar el petaflops (es decir, la capacidad de realizar  $10^{15}$  operaciones en coma flotante por segundo, las siglas flops significan la expresión inglesa "floating point operations per second"). De hecho, dado el tamaño de estos clústers, las redes lógicas pueden ser excesivamente complejas con mecanismos para aislar y asegurar las redes, ya que la definición de la política de direccionamiento lógico no puede entonces expresarse de una manera sencilla.

Además, en un procedimiento de concepción del clúster de servidores por parte de un proveedor y luego de instalar este clúster de servidores en un sitio del usuario, la inicialización de la base de datos de administración ocurre antes de que se entregue el clúster de servidores. Sin embargo, la política de direccionamiento es generalmente definida de manera precisa y ventajosa por el usuario, mientras que el procedimiento de la patente FR 2 931 970 B1 requiere que ya esté disponible antes de la etapa de inicialización de la base de datos. Para superar esta contradicción, la solución aportada impone una rigidez por la que la política de direccionamiento lógico se define de antemano y sin flexibilidad para el usuario.

Por lo tanto, puede ser deseable prever un procedimiento para asignar direcciones lógicas a puertos de conexión de equipos de un clúster de servidores que permita superar al menos algunos de los problemas y limitaciones anteriormente citadas. La invención se define en las reivindicaciones independientes 1, 8 y 9.

El objeto de la invención es, por lo tanto, un procedimiento que consta de las siguientes etapas:

- definición de una política de direccionamiento lógico de dichos puertos de conexión de equipos del clúster,
- a partir de una distribución lógica de los equipos del clúster en la red de transmisión de datos, de una distribución geográfica y de una definición material de los equipos del clúster, la inicialización de una base de datos de administración del clúster de servidores,
- además en función de la política de direccionamiento lógico, asignación de direcciones lógicas a dichos puertos de conexión de equipos del clúster de servidores, y
- registro de las direcciones lógicas asignadas en la base de datos de administración del clúster de servidores,

caracterizado porque comprende la ejecución de un asignador automático de direcciones lógicas sobre la base de parámetros de ejecución que incluye:

- un archivo de descripción estructurado de la política de direccionamiento lógico, y
- el contenido de la base de datos de administración previamente inicializada.

Así, la asignación de direcciones lógicas se realiza después de la inicialización previa de la base de datos de administración del clúster de servidores, lo que permite que el usuario final sea más libre de definir, o incluso de redefinir, el mismo, y posiblemente con total confidencialidad, su política de direccionamiento lógico. Además, al definirlo en un archivo de descripción estructurado, es posible considerar una configuración compleja en redes lógicas para la asignación de direcciones lógicas operacionales en clústers de servidores de muy alta potencia, en el rango del petaflops y más allá, que impliquen alrededor de 5000 nodos o más. Por último, dado que la base de datos de administración se inicializa antes de la asignación de direcciones lógicas, su contenido se puede explotar para una asignación automática eficiente.

Opcionalmente, el fichero de descripción estructurada de la política de direccionamiento lógico comprende una estructura de árbol de redes lógicas y rangos de direcciones lógicas en estas redes.

En particular, la estructura de árbol se puede definir según el formato XML.

Opcionalmente, la estructura de árbol también incluye:

- 5 - en un primer nivel raíz correspondiente a la capa "física" del modelo OSI de interconexión de Sistemas Abiertos, una descripción de redes físicas del clúster de servidores,
- en un segundo nivel correspondiente a las capas de "conexión de datos" y "red" del modelo OSI de interconexión de Sistemas Abiertos, una descripción de las redes lógicas del clúster de servidores,
- en un tercer nivel, una descripción del recorte de estas redes lógicas en rangos de direcciones lógicas, y
- 10 - en un cuarto nivel, una descripción de clases de los equipos del clúster de servidores y de sus rangos de direcciones lógicas asociadas.

Opcionalmente también, la ejecución del asignador automático de direcciones lógicas incluye las dos etapas siguientes:

- 15 - configuración de redes lógicas y generación de un archivo de control con instrucciones para asignar direcciones lógicas a puertos de conexión de equipos en el clúster de servidores, dichas instrucciones expresadas en el lenguaje de gestión de la base de datos, y
- ejecución del archivo de control en la base de datos de administración previamente inicializada.

También opcionalmente, la etapa de configuración de las redes lógicas en la ejecución del asignador automático de direcciones lógicas incluye una comprobación iterativa en todos los puertos de conexión a asignar de la coherencia de la política de direccionamiento definida en el archivo de descripción estructurado con una estructura física del clúster de servidores.

Opcionalmente también, las direcciones lógicas son direcciones IP.

La invención se refiere también a un programa de ordenador descargable de una red de comunicación y/o grabado en un soporte legible por ordenador y/o ejecutable por un procesador, caracterizado porque incluye instrucciones para realizar las etapas de un procedimiento de asignación de direcciones lógicas como definido anteriormente, cuando dicho programa se ejecuta en un ordenador.

La invención también tiene por objeto un clúster de servidores que comprenden varios nodos interconectados entre sí por al menos una red de transmisión de datos, incluyendo al menos un servidor de administración para los nodos del clúster, por ejemplo asociado con una bahía de almacenamiento de datos de administración, caracterizado porque comprende además una base de datos de administración inicializada y completada por un procedimiento según lo definido previamente, con los datos de administración almacenados, por ejemplo, en la bahía de almacenamiento de datos de administración y comprendiendo el servidor de administración, unos medios para gestionar esta base de datos, entre los cuales está incluido el asignador automático de direcciones lógicas.

Opcionalmente, un clúster de servidores según la invención puede incluir nodos de cálculo, al menos un nodo que realice una función de gestión de tráfico de datos explotados por los nodos de cálculo y posiblemente al menos un nodo de gestión de datos, y la red de transmisión de datos podrá incluir al menos una red de interconexión de nodos de cálculo y al menos una red de administración diferente de la red de interconexión de nodos de cálculo para conectar el servidor de administración a los nodos de cálculo, para gestionar el tráfico de datos explotados por los nodos de cálculo y de gestión de datos.

La invención será mejor comprendida utilizando la siguiente descripción, dada únicamente a modo de ejemplo y realizada con referencia a los dibujos adjuntos en los cuales:

- la figura 1 muestra esquemáticamente la estructura general de un ejemplo de clúster de servidores del tipo calculador HPC,
- 45 - la figura 2 representa esquemáticamente la configuración de una base de datos para la administración del clúster de servidores en la figura 1,
- la figura 3 ilustra los pasos sucesivos de un procedimiento de asignación de direcciones lógicas a al menos algunos de los nodos del clúster de servidores de la figura 1.

La instalación informática mostrada en la figura 1 comprende un terminal de control 10 conectado a una red dorsal 12, generalmente llamada red backbone (del inglés "backbone network"). Es también a esta red dorsal 12 a la que se conecta un clúster de servidores 14, que aparece desde el exterior, es decir, desde el punto de vista del terminal de control 10, como una única entidad de calculador HPC.

De hecho, el clúster de servidores 14 comprende varios ordenadores interconectados entre sí por varias redes, siendo todos heterogéneos.

El conjunto de ordenadores en el clúster de servidores 14 constituye el conjunto de nodos en este clúster. Más generalmente, un nodo es un calculador que puede incluir una o más unidades de cálculo.

- 5 En el clúster de servidores 14, se pueden diferenciar dos tipos de nodos: nodos de cálculo y nodos de servicio. Los nodos de cálculo son aquellos que realmente ejecutan las diferentes instrucciones de procesamientos controladas desde el terminal de control 10, bajo la supervisión de los nodos de servicios.

10 La mayoría de los nodos de servicios están duplicados por razones de seguridad. En otras palabras, cada nodo de servicio está asociado a una réplica que consta de las mismas características que aquel y está listo para reemplazarlo inmediatamente en caso de avería.

15 Cabe señalar también que en la figura 1, siempre que exista una multiplicidad de conexiones entre dos entidades (nodos y/o partes de redes), sólo se representará una conexión y se acompañará de un número que indique el número de conexiones entre las dos entidades, en aras de la claridad en esta figura. De hecho, si cada conexión tuviera que estar representada, debido a la complejidad del clúster de servidores 14, se produciría una confusión que dificultaría la comprensión de la invención.

El conjunto de los nodos de servicio del clúster de servidores 14 en la figura 1 incluye una interfaz de procesamiento 16, un servidor de administración 18, un servidor 20 de gestión de metadatos de clúster, un servidor 22 de gestión de entradas/salidas y un servidor de gestión de datos 24.

20 La interfaz del procesamiento 16, más comúnmente conocida como interfaz Login, realiza una función de interfaz de cálculo entre la red dorsal 12 y el clúster de servidores 14. Es a priori de la misma naturaleza que los nodos de cálculo, pero también está dotada con compiladores y herramientas específicas de cálculo cuya presencia en esta interfaz es necesaria para tratar las instrucciones recibidas del terminal de control 10. La interfaz de procesamiento 16 está duplicada, como se ha mencionado anteriormente por razones de seguridad, y por lo tanto está conectada, con su réplica, a la red dorsal 12 mediante dos conexiones 26.

25 El servidor de administración 18 realiza una función de administración general para el clúster 14. En particular, gestiona la distribución de las instrucciones transmitidas por la interfaz de procesamiento 16 a los distintos nodos de cálculo en función de su naturaleza y disponibilidad. También se duplica por razones de seguridad. El servidor de administración 18 y su réplica comparten una bahía de almacenamiento en disco 28 a la que están conectados por una pluralidad de fibras ópticas 29, para un acceso muy rápido a los datos almacenados.

30 Para permitir la administración del clúster de servidores 14 por un usuario del terminal de control 10, el servidor de administración 18 también suele estar conectado directamente a la red dorsal 12 con su réplica a través de dos conexiones 27. Esto también permite al usuario del terminal de control 10 tener más control sobre las estrategias y opciones de cálculo elegidas por el clúster de servidores 14. Además, en algunos modos de realización de clústers de servidores de pequeñas dimensiones que no tengan interfaz Login, este doble enlace 27 es el único enlace entre el clúster de servidores y la red dorsal.

35 El servidor 20 de gestión de metadatos, también conocido como servidor MDS (del inglés "Meta Data Server") y el servidor 22 de gestión de las entradas/salidas, también conocido como servidor OSS (del inglés "Object Storage Server"), realizan una función de gestión del tráfico de datos explotados por los nodos de cálculo del clúster de servidores 14. Para ello, gestionan un sistema de gestión de archivos distribuido, por ejemplo el sistema Lustre (marca registrada).

40 Estos dos servidores también están duplicados y cada uno está conectado a una bahía de almacenamiento por fibras ópticas. El servidor MDS 20 y su réplica comparten una bahía de almacenamiento en disco 30 a la que están conectados por una pluralidad de fibras ópticas 32. Del mismo modo, el servidor OSS 22 y su réplica comparten una bahía de almacenamiento en disco 34 a la que están conectados por una pluralidad de fibras ópticas 36.

45 Por último, el servidor de gestión de datos 24 gestiona la protección de los datos del conjunto del calculador HPC y está por ello conectado a una bahía de almacenamiento en cintas 38. Este servidor de gestión de datos 24, a diferencia de los otros nodos de servicio del clúster de servidores 14, no está duplicado en el ejemplo mostrado en la figura 1.

50

El conjunto de nodos del cálculo del calculador HPC de la figura 1 es heterogéneo e incluye varias unidades de nodos de cálculo tales como por ejemplo una primera unidad de cálculo 40 que comprende seis servidores, una segunda unidad de cómputo 42 que comprende doce servidores y una tercera unidad de cálculo 44 que comprende veinticuatro servidores.

5 Los seis servidores de la primera unidad de cálculo 40 comparten una bahía de almacenamiento específico 48 a la cual están conectados mediante un conmutador 50. Esta bahía de almacenamiento 48 proporciona acceso a volúmenes de datos que, por ejemplo, están organizados de acuerdo con su propio sistema de gestión de archivos, que puede ser diferente del gestionado por los servidores MDS 20 y OSS 22.

10 Los distintos nodos del clúster de servidores 14 antes citados están interconectados entre sí mediante varias redes.

15 Una primera red 60, llamada red de administración, generalmente del tipo Ethernet, permite conectar el servidor de administración 18 del clúster de servidores 14, a través de un puerto de administración de este servidor, a los otros nodos del clúster tales como la interfaz de procesamiento 16, el servidor MDS 20, su réplica y su bahía de almacenamiento 30, el servidor OSS 22, su réplica y su bahía de almacenamiento 34, el servidor de gestión de datos 24 y su bahía de almacenamiento en cintas 38, las unidades de cálculo primera, segunda y tercera 40, 42 y 44, la bahía de almacenamiento específica 48 de la primera unidad de cálculo 40 y la plataforma de administración 52.

20 Opcionalmente, dependiendo del hardware utilizado para los nodos de tipo servidor del calculador, la red de administración 60 puede duplicarse con una red de control primaria conectada al servidor de administración 18 a través de un puerto de control primario de este servidor diferente del puerto de administración. Esta red de control primario 62 está dedicada a encender, arrancar, detener y procesar ciertos errores primarios predeterminados, llamados errores fatales y generando archivos Core, de los servidores que administra. En el ejemplo de la figura 1, la red de control primaria 62 conecta el servidor de administración 18 con la interfaz de procesamiento 16 y su réplica, al servidor MDS 20 y su réplica, al servidor OSS 22 y a su réplica, al servidor de gestión de datos 24 y a las unidades de cálculo segunda y tercera 42 y 44.

30 Una segunda red 64, llamada red de interconexión de los nodos de cálculo, conecta entre sí, por una parte, los servidores de las unidades de cálculo primera, segunda y tercera 40, 42 y 44 y por otra parte, la interfaz del procesamiento 16, el servidor MDS 20, el servidor OSS 22 y el servidor de gestión de datos 24. La conmutación de datos que transitan entre los distintos elementos interconectados por esta red de interconexión 64 se realiza mediante un órgano de conmutación 66 de esta red, que a su vez está conectada a la red de administración 60. Esta red de interconexión 64 tiene unas características de rendimiento muy altas en comparación con las características de rendimiento de la red de administración 60. Es en efecto a través de esta red de interconexión 64 que transitan los datos de cálculo necesarios para ejecutar las instrucciones de procesamiento transmitidas por el terminal de control 10, a través de la interfaz de procesamiento 16.

35 De manera opcional, la red de interconexión 64 puede ser duplicada con una red suplementaria de interconexión 68 conectada al menos a algunos de los elementos ya conectados entre sí por la red de interconexión 64. Por ejemplo, en el clúster de servidores 14 de la figura 1, la red de interconexión suplementaria 68 conecta los servidores de la primera y la segunda unidades de cálculo 40 y 42 para duplicar su ancho de banda. Del mismo modo, la conmutación de los datos que transitan entre los distintos elementos interconectados por esta red de interconexión adicional 68 es realizada por un órgano de conmutación adicional 70 de esta red, que a su vez está conectado a la red de administración 60.

45 La estructura del clúster de servidores 14, tal como se describe anteriormente en referencia a la figura 1, es apropiada para la implementación de la invención, pero también son apropiadas otras posibles configuraciones de clústers, especialmente del tipo calculador HPC, que comprenden algunos o todos los elementos anteriormente citados, o incluso pueden ser adecuados otros elementos en el caso de una mayor complejidad. En una configuración sencilla, un clúster de servidores comprende nodos de servicios que incluyen al menos un servidor de administración, nodos de cálculo, una red de administración que conecta el nodo de administración con los otros nodos del clúster y una red de interconexión de los nodos de cálculo que tiene un mayor rendimiento que el de la red de administración lo que permite alcanzar un alto rendimiento de cálculo. De manera más general, un clúster de servidores también incluye conmutadores, bahías informáticas, administradores de hardware, alimentaciones o cualquier otro elemento que pueda conectarse entre sí mediante puertos de conexión que requieran una dirección lógica, especialmente una dirección IP, para funcionar y que formen el conjunto de sus equipos.

55 Este clúster de servidores 14, compuesto de elementos muy heterogéneos, requiere una base de datos 72 de inicialización y de administración, cuyas herramientas de administración estén por ejemplo alojadas en el

## ES 2 693 546 T3

servidor de administración 18 y cuyos datos de descripción se almacenen en la bahía de almacenamiento 28 asociada con el servidor de administración 18. Los datos, estáticos o dinámicos, de la base de datos 72 se guardan regularmente en la bahía de almacenamiento en cintas 38. Esta base de datos 72 está representada esquemáticamente en la figura 2.

- 5 Incluye un núcleo de base de datos DB, que comprende en particular sus herramientas de administración, y datos de las descripciones estructuradas (D(58), D(60,62), D(64,68), nodos, IHMI, Despliegue, @IP, Ubicación geográfica, SGF, almacenamiento) para proporcionar la información necesaria para la inicialización y la administración del clúster del servidor 14.
- 10 Estas informaciones incluyen primeramente datos D(58), D(60,62), D(64,68) referentes a las distintas redes del clúster de servidores 14: la red de administración 60, 62 y la red de interconexión 64, 68. Estos datos incluyen, por ejemplo el tipo de red, sus capacidades de transmisión, un identificador de proveedor, etc.
- 15 Las informaciones también incluyen datos "nodos" sobre los nodos de tipo servidor del clúster de servidores 14, tales como los conectados a la red de control primaria 62: la naturaleza de cada nodo (servidor de cálculo, servidor de administración, etc.), sus características técnicas (modelo, estado del hardware, capacidad de cálculo, memoria RAM y estado de las herramientas de software instaladas), un identificador de proveedor, etc.
- 20 Las informaciones también comprenden datos de "almacenamiento" de descripción sobre la infraestructura del almacenamiento, sobre el desglose lógico del volumen, sobre los modelos de despliegue, etc.
- 20 También incluyen datos "IHM" relativos a la interfaz hombre-máquina utilizada por el clúster del servidor 14, datos "SGF" relativos al sistema de gestión de archivos utilizado (por ejemplo, el sistema Lustre), datos "Despliegue" relativos a la organización del despliegue en el clúster de servidores, datos "@IP" relativos a la distribución de las direcciones IP dentro del clúster, así como datos "Localización Geográfica" relativos a la localización geográfica de los distintos elementos.
- 25 Para generar la base de datos 72, es decir, para rellenar los valores de sus datos de descripción, se utiliza ventajosamente un procedimiento cuyas etapas están ilustradas en la figura 3. Este procedimiento se desarrolla en dos etapas sucesivas: una primera etapa 100 de inicialización previa de la base de datos 72 en la que no se rellenan los datos "@IP" (de ahí su representación en gris en la figura 2), seguida de una segunda etapa 200 de asignación de direcciones lógicas a al menos una parte de los puertos de conexión de equipos del clúster de servidores 14 en la que se rellenan los datos "@IP". Por ejemplo, la primera etapa 100 es ejecutada al menos parcialmente por el diseñador y proveedor del clúster de servidores 14, mientras que la segunda fase 200 puede ser ejecutada por el usuario cuando el clúster de servidores está instalado in situ.
- 30 La primera fase 100 comienza con una etapa 102 de principio de inicialización de la base de datos 72. En esta primera etapa 102, datos estáticos, que definen una distribución lógica y geográfica de los equipos del clúster en sus diferentes redes, y que definen materialmente los equipos del clúster, son recopilados y verificados por un operador.
- 35 De manera clásica, estos datos están disponibles bajo la forma de archivos digitales, por ejemplo, tablas de datos generados usando un programa de hoja de cálculo. En efecto, estos documentos son generalmente el resultado de una fase de estudio técnico tras una licitación y tienen por objeto definir la arquitectura precisa del clúster de servidores 14.
- 40 Una primera tabla 74, que se llamará tabla de representación lógica del clúster de servidores 14, contiene una lista de hardware e interconexiones de puerto a puerto que conforman el clúster con todas las informaciones que les permite identificar de manera única como hardware y como elementos del clúster (particularmente este documento asigna números únicos para cada nodo en el clúster, sirviendo estos números de sufijos para los identificadores).
- 45 Una segunda tabla 76, que se llamará tabla de representación física del clúster de servidores 14, proporciona una información adicional sobre los equipos del clúster de servidores, especificando su ubicación en un centro de cálculo destinado a recibir el clúster del servidor, por ejemplo utilizando un sistema de coordenadas, precisando particularmente para cada cable la longitud requerida y su número de referencia de catálogo, indicando además ciertas restricciones de peso o ubicación, etc.
- 50 La verificación por el operador consiste en asegurar que los campos de las tablas 74 y 76 requeridos para inicializar la base de datos 72 estén completos.

En una etapa posterior 104, los datos de las tablas 74 y 76 se traducen automáticamente en solicitudes de manipulación de la base de datos 72 del clúster de servidores 14, que se guardan en un archivo de solicitud 80, en la etapa 110. Esta traducción de datos de archivos de tipo tabla en solicitudes es clásica y no será detallada.

- 5 En la etapa 112, un operador verifica el resultado de la traducción de datos en las solicitudes. En este punto, es posible una interacción para modificar el fichero de solicitudes 80.

10 Finalmente, en una última etapa 114, este archivo de solicitudes 80 es ejecutado por el servidor de administración 18 in situ, cuando el clúster de servidores 14 está instalado y en funcionamiento, de modo que se inicialice la base de datos 72. Sin embargo, aunque esta inicializada a partir de las tablas 74 y 76, la base de datos 72 aún no es apta para administrar el clúster de servidores 14 ya que el direccionamiento lógico de los puertos de conexión de equipos del clúster de servidores 14 todavía no se ha realizado. Las etapas 102 a 112, y posiblemente incluso la etapa 114, que contempla la implementación in situ del clúster de servidores 14, pueden ser realizadas por el diseñador/proveedor del clúster.

15 Cuando se completa la primera fase 100, la segunda fase 200 comienza con la etapa 202 de definición de una política de direccionamiento lógico de los puertos de conexión de equipos del clúster de servidores 14 y de generación de un archivo 82 de descripción estructurada de esta política de direccionamiento lógico.

20 Al utilizar el ejemplo de la versión 4 del protocolo IP de direccionamiento lógico, una dirección IP de un puerto de conexión de equipo del clúster está definida por cuatro bytes cuyos valores están separados por puntos, ordenándolos desde el byte con el peso más alto hasta el de peso más bajo. Asumiendo que esta dirección es de clase C, los primeros tres bytes definen el clúster del servidor como una red local y el último byte teóricamente permite distribuir 255 direcciones IP a los puertos de conexión de equipos del clúster de servidores. Si el clúster del servidor tiene demasiados puertos de conexión en comparación con las direcciones disponibles a priori en la clase C, entonces su dirección IP puede ser elegida en la clase B, permitiendo así una configuración de redes lógicas más complejas.

25 Una política de direccionamiento lógico consiste en predefinir unas reglas para asignar direcciones disponibles. Incluye, por ejemplo, las siguientes reglas:

- elección de la clase de direccionamiento en función del número de direcciones a distribuir en el clúster de servidores;
- reserva a priori de determinadas direcciones para los conmutadores de la red de administración,
- 30 - reserva a priori de determinadas direcciones para los conmutadores de la red de interconexión de los nodos de cálculo,
- reserva a priori de determinadas direcciones como direcciones virtuales de nodos así identificadas por un alias cuando se duplican (este es el caso, en particular, de la interfaz de procesamiento 16, el servidor de administración 18 y los nodos de gestión 20 y 22),
- 35 - reserva a priori de una zona de direcciones, para los nodos de cálculo interconectados en serie,
- asignación de una dirección de sub-red IP a cada una de las tres redes del clúster de servidores 14,
- etc.

40 Las reglas de distribución de direcciones lógicas IP están por otra parte claramente definidas en el archivo 82 de descripción estructurada de la política de direccionamiento lógico. Por ejemplo, se trata de un archivo compatible con el idioma XML (del inglés "eXtensible Markup Language) en el que una configuración lógica de red IP deseada por el usuario del clúster de servidores 14 se define como una estructura de árbol. Este formalismo estructurado para describir una configuración lógica de redes IP puede ser utilizado en particular por un experto en redes.

45 Se da un ejemplo sencillo de dicho fichero 82 en el anexo 1. Incluye una red lógica principal llamada "interconnect" compuesta por dos redes IP: una llamada "interconnect", que ocupa una subred IP identificada por las direcciones 10.5.0.0/16 y de tipo "primario", es decir, directamente asociada a una red física; la otra llamada "ib-backbone", que ocupa una subred IP identificada por las direcciones 10.2.34.0/23 y de tipo "secundario", es decir, indirectamente asociada a una o más redes físicas.

50 Cada red IP se divide entonces, a un nivel jerárquico superior en la estructura de árbol, recortada en rangos de direcciones IP.

Por ejemplo, en el archivo 82 del Anexo 1, la red IP "interconnect" incluye:



- un primer rango de primera dirección IP 10.5.31.1, reservado a los nodos de tipo "mngt" y "mws" del grupo "Service\_1";
- un segundo rango de primera dirección 10.5.32.7, reservado a los nodos de tipo "router" de interfaz "ib0" y a los nodos de tipos "mds" e "io" del grupo "Service\_1" de interfaz "ib0",
- 5 - un tercer rango de primera dirección 10.5.33.7, reservado a los nodos de tipo "router" de interfaz "ib2" y a los nodos de tipos "mds" e "io" del grupo "Service\_1" de interfaz "ib1",
- un cuarto rango de primera dirección IP 10.5.34.1, reservado para los nodos del tipo "compute" del grupo "Service\_interco",
- 10 - un quinto rango de primera dirección IP 10.5.5.36.1, reservado para los objetos "hwmanager" del tipo "bmc", y
- un sexto rango de la primera dirección IP 10.5.5.60.1, reservado para los objetos "switch" del tipo "eth".

Por ejemplo igualmente en el archivo 82 del anexo 1, la red IP "ib-backbone" incluye:

- un primer rango de la primera dirección IP 10.2.34.1, reservado para los nodos "compute" del grupo "Service\_backbone", y
- 15 - un segundo rango de primera dirección 10.2.35.231, reservado para los nodos del tipo "gw\_ip".

De manera más general, el archivo 82 está compuesto por un árbol de datos jerárquicos cuya estructura puede basarse en los principios del modelo OSI (del inglés "Open Systems Interconnection"). Así por ejemplo, la raíz del árbol de representación de configuración de redes lógicas define las redes físicas del clúster de servidores 14 que cumplen las características de la capa "física" del modelo OSI. El segundo nivel del árbol define las redes lógicas y sus características correspondiente a las características de las capas de "conexión de datos" y "red" del modelo OSI (en particular de las características que permiten considerar una división en rangos de direcciones IP y que definen parámetros de seguridad). El tercer nivel del árbol define la división de las redes lógicas en rangos de direcciones IP. El cuarto nivel del árbol define las clases de los equipos del clúster de servidores 14. Estas clases se definen gracias a un meta-lenguaje que permite vincular un equipo a sus características físicas. Un equipo se define por su modelo, su nombre, sus características y su capacidad de estar asociado a otra clase de equipo. Las informaciones sobre las clases se deducen a partir de las informaciones físicas de la base de datos 72. Este cuarto nivel del árbol asigna a una clase de equipo, que pueda clasificarse según criterios (tipo, ubicación, función), un rango de direcciones IP definido en el tercer nivel del árbol.

El paso 202 de generación del archivo 82 es seguido por la ejecución de un asignador automático de direcciones lógicas sobre la base:

- del archivo 82 de descripción estructurada de la política de direccionamiento lógico, y
- del contenido de la base de datos 72 previamente inicializada en la etapa 114.

35 Este asignador automático de direcciones lógicas es, por ejemplo, un programa de ordenador ejecutable que recibe el archivo 82 y la base de datos 72 como parámetros de ejecución.

Se da un ejemplo de archivo fuente 84 de este programa informático en el anexo 2, para una primera parte del asignador cuya función es la de aplicar la estructura lógica de una o varias redes IP definidas en el archivo 82 de descripción estructurada a cada red física definida en la base de datos 72 inicializada. A título puramente ilustrativo, el archivo fuente 84 contiene en primer lugar una definición de los objetos que se extraerán de los parámetros de ejecución: la lista de redes físicas, la lista de equipos y los grupos de equipos definidos por el cliente se extraen de la base de datos 72 inicializada; las reglas de asignación de direcciones IP se extraen del archivo de 82 de descripción estructurada. A continuación, incluye instrucciones para configurar las redes lógicas una vez que se han rellenado los parámetros. Estas instrucciones van seguidas de un recorrido de todos los equipos del clúster de servidores 14, tal y como se describe en la base de datos 45 72, para enumerar cuales son los puertos que deben asignarse o no (etapa iterativa para eliminar los puertos asignados). Finalmente, una vez identificados los puertos a asignar, el fichero fuente 84 incluye una iteración de verificación y asignación destinada a identificar cualquier posible colisión de direcciones IP o cualquier exceso de rangos predefinidos resultantes de una aplicación de la configuración lógica en las redes físicas del clúster de servidores 14 y para asignar direcciones IP a cada puerto de conexión a asignar. La ejecución del archivo fuente 84 durante una etapa 204 permite crear un archivo de control que, cuando se ejecuta en una segunda parte del asignador, durante una última etapa 206, lanza ordenes en el idioma de gestión de bases de datos, por ejemplo SQL, para enriquecer la base de datos 72 con asignaciones de direcciones IP realizadas en la primera parte.

55 Se ve claramente que un procedimiento de asignación de direcciones lógicas como el descrito anteriormente permite al usuario de un clúster de servidores definir o completar su política de direccionamiento lógico con

5 mayor flexibilidad y libertad, ya que las informaciones sobre redes lógicas a aplicar en redes físicas son generalmente específicas para cada usuario de acuerdo a sus restricciones estructurales y de seguridad. Además, durante una evolución de la red del clúster de servidores, por ejemplo, añadiendo una tarjeta de red en un nodo o una extensión del clúster de servidores, el asignador automático de direcciones lógicas puede comprobar y notificar las incoherencias que puedan aparecer y aplicar la configuración de red lógica decidida por el administrador del clúster.

10 Además, el experto en la materia verá que se pueden hacer diversas modificaciones en el modo de realización descrito anteriormente, a la luz de la enseñanza que acaba de serle revelada. En las siguientes reivindicaciones, los términos utilizados no deben interpretarse en el sentido de que limitan las reivindicaciones al modo de ejecución expuesto en la presente descripción, sino que deben interpretarse en el sentido de que incluyen todos los equivalentes que las reivindicaciones pretenden cubrir en virtud de su formulación y cuya previsión está al alcance de la persona cualificada aplicando sus conocimientos generales a la realización de la enseñanza descrita anteriormente.

ANEXO 1

Archivo 82

```

<network name="interconnect" function="ic">
  <logical_network name="interconnect" type="primary" subnet="10.5.0.0/16"
  gateway="10.5.255.254" pkey="E8FF" >
    <range first_ip="10.5.32.1" >
      <member object="node" type="mngt" group="Service_1" />
      <member object="node" type="mws" group="Service_1" />
    </range>
    <range first_ip="10.5.32.7" >
      <member object="node" type="router" interface_name="ib0" />
      <member object="node" type="mds" group="Service_1"
      interface_name="ib0" />
      <member object="node" type="io" group="Service_1"
      interface_name="ib0" />
    </range>
    <range first_ip="10.5.33.7" >
      <member object="node" type="router" interface_name="ib2" />
      <member object="node" type="mds" group="Service_1"
      interface_name="ib1" />
      <member object="node" type="io" group="Service_1"
      interface_name="ib1" />
    </range>
    <range first_ip="10.5.34.1" >
      <member object="node" type="compute" group="Service_interco"
      />
    </range>
    <range first_ip="10.5.36.1" >
      <member object="hwmanager" type="bmc"/>
    </range>
    <range first_ip="10.5.60.1" >
      <member object="switch" type="eth"/>
    </range>

```

## ES 2 693 546 T3

```
</logical_network>
<logical_network name="ib-backbone" type="secondary"
subnet="10.2.34.0/23" gateway="10.2.35.254" pkey="8132" >
  <range first_ip="10.2.34.1" >
    <member object="node" type="compute"
group="Service_backbone" />
  </range>
  <range first_ip="10.2.35.231" >
    <member object="node" type="gw_ip" />
  </range>
</logical_network>
</network>
```

ANEXO 2

ARCHIVO 84 (primera parte):

```

New NetworkList nwl           /* Listado de las redes físicas */
New EquipmentList eql        /* Listado de los equipos */
New HWGroupList hwgl        /* Grupo de equipos definidos por el cliente
New AllocationRule alloc_rule /* Reglas de asignación definidas en el archivo 82

alloc_rule=LoadRuleFromFile("/etc/networkmap.xml")

nwl= LoadNetworkListFromDB()
eql= LoadEquipmentListFromDB()
hwgl= LoadHWGroupListFromDB()

alloc_rule->add_network(nwl)
alloc_rule->add_hwgl(hwgl)
alloc_rule->add_eql(eq)

/* Supresión de los puertos asignados */
Foreach eq from eql
    Foreach netint from eq->get_interfaces()
        if ( ! netint->is_allocated())
            then
                eq->remove_from_list(netint)
            fi
    Done
Done

/* Verificación y asignación de los IP en los puertos no asignados */
Foreach nw from nwl
    Foreach eq from eql
        Foreach int from eq->get_interface
            Foreach rule from alloc_rule
                if (rule->apply(nw,eq,int)
                    then
                        rule->check_nextIP(nw)
                        rule->allocate_ip(nw,int)
                    fi
            Done
        Done
    Done
Done

```

**Reivindicaciones**

1. Procedimiento para asignar direcciones lógicas a puertos de conexión de equipos de un clúster de servidores (14), que comprende las siguientes etapas:
  - 5 - definición (202) de una política de direccionamiento lógico para dichos puertos de conexión de equipos del clúster,
  - a partir de una distribución lógica (74) de los equipos en el clúster de servidores, de una distribución geográfica (76) y de una definición de hardware (74, 76) de los equipos del clúster, inicialización (114) de una base de datos (72) de administración del clúster de servidores por:
    - 10 \* traducción (104) de los datos relativos a esta distribución lógica (74), esta distribución geográfica (76) y esta definición material (74, 76) en solicitudes de manipulación de bases de datos (72),
    - \* copia de seguridad (110.112) de estas solicitudes, y
    - \* ejecución (114) de estas solicitudes para completar la base de datos (72),
  - 15 - en función además de la política de direccionamiento lógico, de asignación (204) de direcciones lógicas a dichos puertos de conexión de equipos del clúster de servidores, y
  - registro (206) de las direcciones lógicas asignadas en la base de datos de administración del clúster de servidores,
  - 20 incluyendo la ejecución (204, 206) de un asignador automático (84) de direcciones lógicas sobre la base de parámetros de ejecución que incluyen un archivo (82) de descripción estructurada de la política de direccionamiento lógico, en el cual:
    - la inicialización (114) de la base de datos de administración (72), incluyendo la ejecución de dichas peticiones, se realiza sin asignar direcciones lógicas a dichos puertos de conexión de equipos del clúster de servidores,
    - 25 - el archivo (82) de descripción estructurada incluye una estructura de árbol de redes lógicas y rangos de direcciones lógicas en estas redes lógicas, y
    - los parámetros de ejecución del asignador automático (84) de direcciones lógicas incluyen además el contenido, obtenido mediante la ejecución de dichas solicitudes, de la base de datos de administración (72) previamente inicializada de manera a aplicarle la estructura de árbol del archivo (82) de descripción estructurada para asignar las direcciones lógicas a dichos puertos de conexión de equipo del clúster de servidores, y
    - 30 - el asignador automático (84) de direcciones lógicas se ejecuta después de ejecutar la inicialización (114) de la base de datos (72).
2. Procedimiento de asignación de direcciones lógicas según la reivindicación 1, en el que:
  - 35 - la base de datos de administración (72) previamente inicializada contiene una lista de redes físicas, y
  - la aplicación de la estructura de árbol del archivo (82) de descripción estructurada al contenido de la base de datos de administración (72) previamente inicializada consiste en aplicar la estructura lógica de una o más de las redes lógicas definidas en el archivo (82) de descripción estructurada a cada red física definida en la base de datos de administración (72) previamente inicializada.
  - 40
3. Procedimiento de asignación de direcciones lógicas según la reivindicación 2, en la que la estructura de árbol se define según el formato XML.
- 45 4. Procedimiento de asignación de direcciones lógicas según la reivindicación 2 o 3, en la que la estructura de árbol comprende:

- en un primer nivel raíz correspondiente a la capa "física" del modelo OSI de interconexión de Sistemas Abiertos, una descripción de redes físicas del clúster de servidores (14),
  - en un segundo nivel correspondiente a las capas de "conexión de datos" y "red" del modelo OSI de interconexión de Sistemas Abiertos, una descripción de las redes lógicas del clúster de servidores,
- 5
- en un tercer nivel, una descripción de la división de estas redes lógicas en rangos de direcciones lógicas, y
  - en un cuarto nivel, una descripción de clases de los equipos del clúster de los servidores y de sus rangos de direcciones lógicas asociadas.
- 10
5. Procedimiento de asignación de direcciones lógicas según una cualquiera de las reivindicaciones 1 a 4, en el cual la ejecución (204, 206) del asignador automático (84) de direcciones lógicas consta de las dos etapas siguientes:
- configurar (204) redes lógicas y generar un archivo de control que comprende instrucciones para asignar direcciones lógicas a los puertos de conexión de equipos del clúster de servidores (14); expresándose estas instrucciones en el lenguaje de gestión de la base de datos, y
- 15
- ejecutar (206) el fichero de control en la base de datos de administración (72) previamente inicializada.
- 20
6. Procedimiento de asignación de direcciones lógicas según la reivindicación 5, en el que la etapa de configuración (204) de las redes lógicas en la ejecución del asignador automático (84) de direcciones lógicas incluye una verificación iterativa de todos los puertos de conexión que deben asignarse para garantizar la coherencia de la política de direccionamiento definida en el archivo de descripción estructurada (82) con una estructura física del clúster de servidores (14).
- 25
7. Procedimiento de asignación de direcciones lógicas según una cualquiera de las reivindicaciones 1 a 6, en el que las direcciones lógicas son direcciones IP.
- 30
8. Programa de ordenador descargable desde una red de comunicación y/o grabado en un soporte legible por ordenador y/o ejecutable por un procesador, **caracterizado porque** incluye instrucciones para ejecutar etapas de un procedimiento de asignación de direcciones lógicas de acuerdo con cualquiera de las reivindicaciones 1 a 7, cuando dicho programa se ejecuta en un ordenador.
- 35
9. Clúster de servidores (14) que consta de varios nodos (16, 18, 20, 22, 24, 28, 28, 28, 30, 34, 38, 40, 42, 42, 44, 46, 48, 52) interconectados entre sí por al menos una red de transmisión de datos (58, 60, 62, 64, 68), incluido al menos un servidor (18) de administración de nodos del clúster asociado con una bahía (28) de almacenamiento de datos de administración, comprendiendo el clúster de servidores:
- equipos con puertos de conexión, incluyendo dichos nodos y todos los demás elementos que puedan acoplarse entre sí mediante puertos de conexión que requieran una dirección lógica para su funcionamiento,
- 40
- una base de datos de administración (72) inicializada mediante registro, en dicha base de datos de administración (72), de datos relativos a una distribución lógica (74) de los equipos en el clúster de los servidores (14), con una distribución geográfica (76) y con una definición de hardware (74, 76) de los equipos en el clúster de servidores (14), estando dicha base de datos de administración (72) almacenada en la bahía (28) de almacenamiento de datos de administración,
- 45
- un archivo (82) de descripción estructurada de una política de direccionamiento lógico de dichos puertos de conexión de equipos del clúster de servidores (14), y
  - un asignador automático (84) de direcciones lógicas basado en parámetros de ejecución que incluyen dicho archivo (82) de descripción estructurada de la política de direccionamiento lógico, siendo dicho asignador automático (84) alojado por el servidor de administración (18):
- 50
- un fichero de solicitudes (80) que almacena los datos relativos a la distribución lógica (74) de los equipos en el clúster de servidores (14) y relativas a la distribución geográfica (76) y a la definición

- material (74, 76), que se traduce automáticamente en solicitudes de manipulación de la base de datos (72) del clúster de servidores (74),
- 5 - un servidor de administración (18) in situ, capaz de ejecutar el archivo de solicitudes, cuando el clúster de servidores (14) está instalado y en condiciones de funcionamiento, con el fin de inicializar la base de datos (72),
- la base de datos de administración inicializada (72), incluyendo la inicialización de la base de datos de administración, la ejecución de dichas solicitudes, no incluye las direcciones lógicas asignadas a dichos puertos de conexión de equipos del clúster de servidores (14),
- 10 - el archivo (82) de descripción estructurada incluye una estructura de árbol de redes lógicas y rangos de direcciones lógicas en estas redes lógicas, y
- los parámetros de ejecución del asignador automático (84) de direcciones lógicas incluyen además el contenido de la base de datos de administración inicializada (72), de manera a aplicarle la estructura de árbol del archivo (82) de descripción estructurada para asignar las direcciones lógicas a dichos puertos de conexión de equipos del clúster de servidores (14) y a almacenarlos en la base de datos de administración inicializada (72).
- 15
- 20 **10.** Clúster de servidores (14) según la reivindicación 9, que comprenda nodos de cálculo (40, 42, 44), al menos un nodo (20, 22) que desempeñe una función de gestión del tráfico de datos gestionados por los nodos de cálculo y, opcionalmente, al menos un nodo de gestión de datos (24), y en el que la red de transmisión de datos incluya al menos una red (64, 68) de interconexión de los nodos de cálculo y al menos una red de administración (60, 62) diferente de la red de interconexión de los nodos de cálculo para conectar el servidor de administración (18) a los nodos de cálculo (40, 42, 44), de gestión del tráfico de datos gestionados por los nodos de cálculo (20, 22) y de gestión de datos (24).



Figura 1

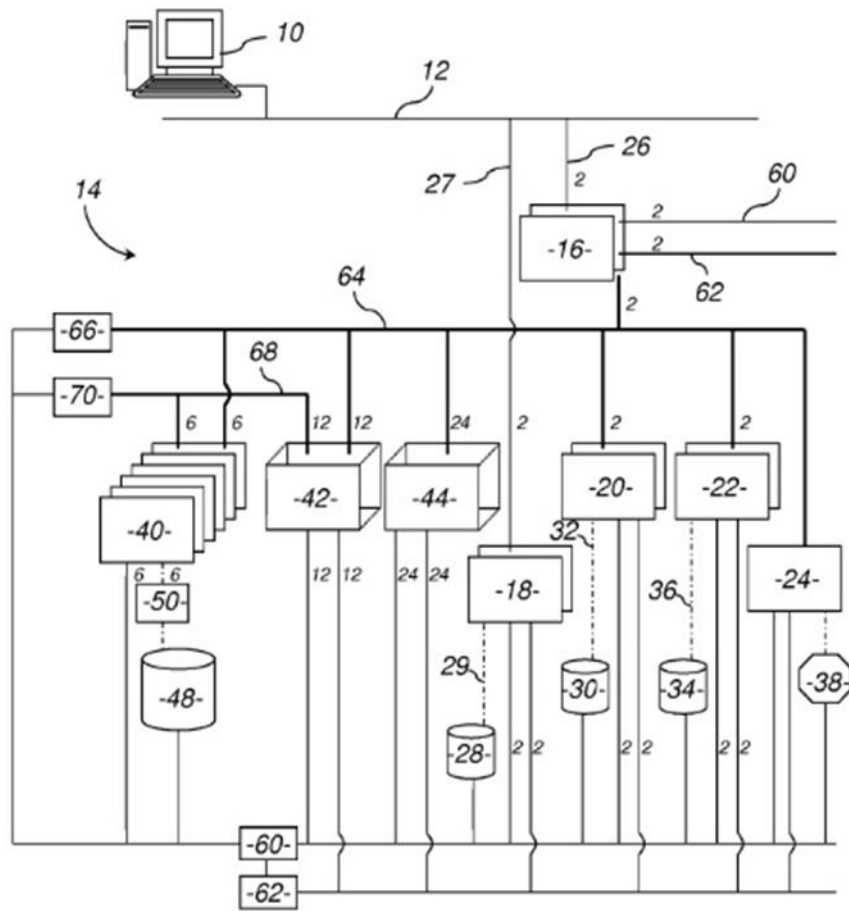


Figura 2

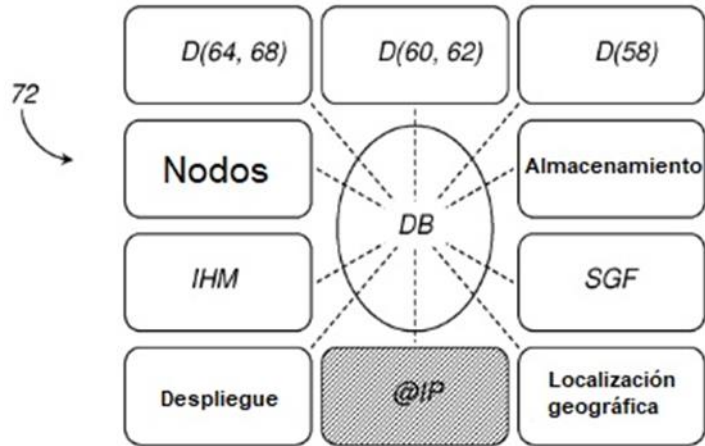


Figura 3

