

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 743 469**

51 Int. Cl.:

G06F 9/50 (2006.01)

H04L 29/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.10.2011 PCT/EP2011/067888**

87 Fecha y número de publicación internacional: **19.04.2012 WO12049247**

96 Fecha de presentación y número de la solicitud europea: **13.10.2011 E 11768015 (7)**

97 Fecha y número de publicación de la concesión europea: **12.06.2019 EP 2628080**

54 Título: **Disposición de ordenadores en racimo para el tratamiento de una tarea de cálculo y procedimiento correspondiente**

30 Prioridad:

13.10.2010 EP 10187436

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

19.02.2020

73 Titular/es:

**PARTEC CLUSTER COMPETENCE CENTER
GMBH (100.0%)
Possartstrasse 20
81679 München, DE**

72 Inventor/es:

LIPPERT, THOMAS

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 743 469 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

Disposición de ordenadores en racimo para el tratamiento de una tarea de cálculo y procedimiento correspondiente

La presente invención se refiere a una disposición de ordenadores en racimo. En concreto, se refiere a una disposición de ordenadores en racimo con una gestión de recursos mejorada por lo que respecta a la aplicación de nodos de cálculo para el tratamiento de tareas de cálculo escalables así como de tareas de cálculo complejas. Especialmente, se refiere a una disposición de ordenadores en racimo para el tratamiento de una tarea de cálculo y a un procedimiento para operar la disposición de ordenadores en racimo. La disposición de ordenadores en racimo de acuerdo con la presente invención hace uso de una funcionalidad de aceleración, la cual ayuda a los nodos de cálculo a llevar a cabo una tarea de cálculo determinada. La presente invención, así mismo, se refiere a un producto de programa informático configurado para llevar a cabo el procedimiento así como a un medio legible por ordenador para almacenar el producto de programa informático.

En las técnicas son conocidas las disposiciones de ordenadores en racimo que comprenden unos nodos de cálculo que incluyen al menos un procesador así como unos aceleradores estrechamente acoplados a los nodos de cálculo para externalizar cálculos de necesidades de elevados recursos. Un fuerte acoplamiento de aceleradores con los nodos de cálculo se traduce en una asignación estática y conduce a una sobre o subsuscripción de aceleradores. Ello puede conducir a una carencia de recursos o puede conducir a un suministro excesivo de recursos. Dicha asignación estática de aceleradores a los nodos de cálculo tampoco proporcionan una tolerancia a fallos en el caso de que el acelerador se averíe.

El documento US 2009/0213127 A1 muestra un procedimiento de fijación guiada de aceleradores hardware a las ranuras de un sistema de cálculo que incluye la división de un primer grupo de aceleradores hardware en una pluralidad de clases prioritarias, la división de un primer grupo de ranuras del sistema de cálculo en una pluralidad de niveles jerárquicos y la fijación de cada acelerador hardware del primer grupo de aceleradores hardware a una ranura en correspondencia con los aceleradores hardware en base a la comparación de una clase prioritaria del acelerador hardware y un nivel jerárquico de la ranura.

Un ordenador en racimo (cluster) se describe por C.S.R.: Prabhu en "Cálculo de Red y en Racimo", PHI Learning Pvt. Ltd., 2008 como sigue "Un racimo es una categoría específica de sistema de tratamiento distribuido paralelo consistente en una agrupación de ordenadores autónomos interconectados que funcionan de manera conjunta como un único recurso de cálculo integrado. Cada uno de estos ordenadores autónomos puede ser considerados como un nodo. Un nodo, en general, puede ser un único procesador o un sistema de multiprocesadores, un PC, una estación de trabajo o un SMP con memoria, funcionalidades de E / S y un sistema operativo. Un racimo incorporará dos o más nodos conjuntamente. Estos nodos pueden estar alojados en una misma carcasa o pueden estar situados como nodos autónomos dentro de una LAN, todos interconectados por una red de interconexión de alta velocidad. La arquitectura de un racimo incorporará también un middleware de racimo que proporcione una única visión del sistema al usuario. Este middleware es el componente esencial de la LAN o de la carcasa de racimo, lo que permite una única visión sistemática del racimo". El mismo trabajo sigue manifestando: "Los racimos son agregaciones de procesadores en configuraciones paralelas. En los racimos, la asignación de recursos se lleva a cabo mediante un gestor de recursos y un sistema de programación centralizados. Todos los nodos de un racimo funcionan de modo cooperante de manera conjunta, como una única fuente unificada. En el caso de una red de distribución eléctrica, cada nodo incorpora su propio gestor de recursos y no tiene por objeto la provisión de una única visión del sistema. Un racimo comprende múltiples nodos independientes interconectados que funcionan cooperativamente entre sí como una única fuente unificada. Todas las interacciones de usuario con un racimo pasan por un sistema centralizado que gestiona la asignación de recursos a las labores de aplicación. Algunas redes de distribución eléctrica son agrupaciones de racimos, por ejemplo la red NSF Tera y la red eléctrica mundial que incorpora varios racimos como nodos situados en diferentes países (Canadá, Japón y Australia). El sistema de gestión en racimos (como por ejemplo el Sun grid engine) no es realmente un sistema de gestión de distribución eléctrica. Dichos sistemas presentan un control centralizado, un control completo sobre los componentes individuales y un conocimiento completo de las necesidades de los usuarios y un estado del sistema - dichos sistemas no son redes de distribución eléctrica".

La publicación "rCUDA" : reducción del número de aceleradores en base a la GPU en racimos de alto rendimiento" de José Duato, Rafael Mayo et al., Conferencia Internacional sobre Cálculo y Simulación de Alto Rendimiento (HPCS); Fecha de Impresión: 28 de junio de 2010 - 2 de julio de 2010, en página(s): 224 - 231, describe un marco que posibilita la aceleración GPU remota en racimos de alto rendimiento, posibilitando con ello una reducción del número de aceleradores instalados en el racimo. Ello puede conducir a ahorros de energía, adquisición, mantenimiento y espacio.

La publicación "Un paquete para un CL abierto en base a cálculos heterogéneos sobre racimos con muchos dispositivos GPU" de Amnon Barak, et al., del Departamento de Ciencias Informáticas de la Universidad Hebrea de Jerusalén describe un paquete para ejecutar un MPabierto, C++ en aplicaciones de un OpenCL no modificado sobre racimos con muchos dispositivos GPU. Así mismo, se proporciona una implementación de las especificaciones OpenCL y extensiones de API de OpenCL que permiten aplicaciones sobre un nodo hosting para utilizar de manera transparente dispositivos en racimo de amplio aspecto.

La Fig. 1 muestra una disposición de ordenadores en racimo de acuerdo con el estado de la técnica. La disposición de ordenadores en racimo comprende varios nodos de cálculo, CN, que están interconectados y que llevan a cabo de manera conjunta una tarea de cálculo. Cada nodo de cálculo, CN, está firmemente acoplado con un acelerador, Acc. Como se puede apreciar en la Fig. 1, un nodo de cálculo, CN, comprende una unidad de acelerador, ACC, que está virtualmente integrada en el nodo de cálculo, CN, junto con un microprocesador, por ejemplo una unidad de procesamiento central, UPC. Como se introdujo con anterioridad, el acoplamiento fijo de aceleradores, Acc, a los nodos de cálculo, CN, conduce a una sobre - o subscripción de aceleradores, Acc, dependiendo de la tarea de cálculo. Así mismo, no se dispone ninguna tolerancia a fallos en el caso de un fracaso de uno de los aceleradores, Acc. En la disposición de ordenadores en racimo conocida de acuerdo con la Fig. 1, los nodos de cálculo, CN, comunican entre sí sobre una infraestructura, de manera que los aceleradores, Acc, no intercambian información directamente, sino que requieren un nodo de cálculo, CN, que sitúa en interfaz la infraestructura, IN, para el intercambio de datos.

Por tanto, es un objeto de la presente invención proporcionar una disposición de ordenadores en racimo que permita una flexibilidad de comunicaciones por lo que respecta a los intercambios de datos entre el acelerador y los nodos de cálculo así como un acceso directo de los nodos de cálculo a todos y cada uno de los aceleradores. Así mismo, es un objetivo de la presente invención proporcionar un acoplamiento dinámico de aceleradores con los nodos de cálculo en el tiempo de ejecución.

Estos objetos se solventan mediante una disposición de ordenadores en racimo con las características de acuerdo con la reivindicación de patente 1. El alcance de la presente invención se define por las reivindicaciones adjuntas. Formas de realización que no se incluyen en el alcance de las reivindicaciones no describen parte de la invención.

Por consiguiente, se proporciona una disposición de ordenadores en racimo para el tratamiento de una tarea de cálculo, comprendiendo la disposición de ordenadores en racimo:

- una pluralidad de nodos de cálculo, cada uno de los cuales se sitúa en interfaz con una infraestructura de comunicación, estando al menos dos de los cuales dispuestos para de manera conjunta calcular al menos una primera parte de la tarea de cálculo;
- al menos un potenciador que está dispuesto para calcular al menos una segunda parte de la tarea de cálculo, situándose cada potenciador en interfaz con la infraestructura de comunicación; y
- un gestor de recursos que está dispuesto para asignar al menos un potenciador a al menos uno entre la pluralidad de nodos de cálculo para el cálculo de la segunda parte de la tarea de cálculo, llevándose a cabo la asignación en función de una métrica de asignación predeterminada.

En esta disposición de ordenadores en racimo la funcionalidad de aceleración se suministra mediante potenciadores independientes. La disposición de ordenadores en racimo descrita permite un acoplamiento débil de esos potenciadores con los nodos de cálculo, los cuales pueden también ser designados como módulos de cómputo. Por tanto, en este punto, es factible una compartición de aceleradores en forma de potenciadores mediante los nodos de cálculo. Para una asignación de un potenciador a un nodo de cálculo, se puede disponer un gestor de recursos, en forma de un módulo o de gestor de recursos o de un nodo de gestor de recursos. El gestor de recursos puede establecer una asignación estática al comienzo de un tratamiento de una tarea de cálculo. Como alternativa o adicionalmente, se puede establecer una asignación dinámica en tiempo de ejecución, lo que significa durante el tratamiento de la tarea de cálculo.

El gestor de recursos está dispuesto para proporcionar informaciones de asignación a los nodos de cálculo para externalizar partes de las tareas de cálculo a partir de al menos un nodo de cálculo hasta al menos un potenciador. El gestor de recursos puede ser implementado como una unidad de hardware específica, una unidad virtual o puede ser un compuesto de cualquiera de ellas. Especialmente el gestor de recursos puede formarse mediante un elemento cualquiera entre: un microprocesador, un componente de hardware, un componente virtualizado o un demonio. Así mismo, partes del gestor de recursos pueden ser distribuidas a través del sistema y comunicar a través de una infraestructura de comunicación.

La comunicación entre potenciadores se lleva a cabo mediante un protocolo de red. Por tanto, la asignación de potenciadores se lleva a cabo en función de las necesidades de la aplicación, lo que significa en dependencia del tratamiento de una tarea de cálculo específica. La tolerancia a fallos en el caso de una avería de los potenciadores se suministra así como se estimula la escalabilidad. Se hace posible la escalabilidad mediante el soporte del desarrollo del sistema incremental, cuando los potenciadores se incorporan de manera independiente respecto de los nodos de cálculo. Por tanto, el puede diferir el número de nodos de cálculo y el número de los potenciadores incorporados. Así, se establece una máxima flexibilidad en la provisión de los recursos de hardware. Por otro lado, todos los nodos de cálculo comparten la misma capacidad de crecimiento.

Una tarea de cálculo se puede definir por medio de un algoritmo, un código de fuente, un código binario y también mediante una combinación de cualquiera de ellos. Una tarea de cálculo puede, por ejemplo, ser una simulación que tenga que ser calculada por la disposición de ordenadores en racimo. Así mismo, la tarea de cálculo puede comprender varios problemas implicados, también designados como subtareas, las cuales en su totalidad describan

la tarea de cálculo global. Es posible dividir la tarea de cálculo en varias partes, por ejemplo al menos una primera parte de la tarea de cálculo y al menos una segunda parte de la tarea de cálculo. Así mismo, es posible que la disposición de ordenadores en racimo solventa las partes de la tarea de cálculo en paralelo o en sucesión.

5 Cada nodo de cálculo se sitúa en interfaz con una infraestructura de comunicación, también designada como interconexión. De modo análogo, cada potenciador se sitúa en interfaz con la infraestructura de comunicación. Por tanto, los nodos de cálculo así como los potenciadores interactúan por medio de la infraestructura de comunicación. En consecuencia, el nodo de cálculo comunica con cada potenciador a través de la infraestructura de comunicación sin la necesidad de implicar un nodo de comunicación adicional al tiempo que se intercambian datos desde un nodo de cálculo hasta un potenciador. De esta manera, se establece una asignación dinámica de nodos de cálculo con los
10 potenciadores, de manera que los nodos de cálculo procesan al menos una parte de la tarea de cálculo y no son requeridos para su pase a través de las informaciones desde un nodo de cálculo a un potenciador. Por consiguiente, es posible acoplar directamente los potenciadores a la infraestructura de comunicación sin la necesidad de un nodo de cálculo intermedio como típicamente se implementa en el estado de la técnica.

15 Para llevar a cabo la asignación entre potenciadores y los nodos de cálculo se requiere un conjunto específico de normas. Por tanto, se proporciona una métrica de asignación, la cual sirve como base para decidir cuál sea el potenciador que se acople con el nodo de cálculo específico. La métrica de asignación puede ser gestionada por un gestor de recursos. La gestión de la métrica de asignación se refiere al establecimiento y a la actualización de normas entre las cuales se incluya al menos un potenciador, el cual sea asignado a al menos otro nodo de cálculo nominado. Por tanto, es posible actualizar la métrica de asignación en el tiempo de ejecución. Dichas normas de
20 asignación pueden ser creadas como una función de una distribución compensada de carga, la cual detecte la carga de trabajo de la disposición de ordenadores en racimo, especialmente de los potenciadores. Así mismo, es posible detectar las capacidades de cálculo de los potenciadores y también detectar las exigencias de las tareas de cálculo y asignar un potenciador seleccionado, lo cual proporcione las capacidades requeridas al nodo de cálculo. Para determinar una asignación inicial de potenciadores a los nodos de cálculo, se predetermina la métrica de asignación pero puede ser alterada en el tiempo de ejecución. Por tanto, la asignación estática se dispone al inicio del
25 tratamiento de la tarea de cálculo y la asignación dinámica se proporciona en el tiempo de ejecución.

En una forma de realización de la presente invención, se forma la métrica de asignación determinada de acuerdo con al menos un grupo de técnicas de especificación métricas, comprendiendo el grupo: una lógica temporal, una matriz de asignación, una tabla de asignación, una función de probabilidad y una función de coste. Por tanto, pueden
30 tenerse en cuenta dependencias temporales para asignar los potenciadores. Puede resultar que se defina un orden temporal en los potenciadores, el cual asegure que un potenciador específico siempre es asignado a un nodo de cálculo en el caso de que un potenciador adicional no consiga resolver al menos una parte de la tarea de cálculo. Por tanto, se puede considerar una jerarquía entre potenciadores para su asignación. Una métrica de asignación puede nominar una identificación de un nodo de cálculo y puede, así mismo, definir identificaciones de
35 potenciadores compatibles que puedan ser asignados. Una función de probabilidad puede, por ejemplo, describir que en el caso de que falle un potenciador específico a la hora de calcular una determinada tarea de cálculo un potenciador adicional pueda solventar la misma tarea de cálculo en una probabilidad específica. Así mismo, las funciones de coste pueden ser aplicadas para la evaluación de las capacidades de recursos requeridas y, así mismo, para la evaluación de las capacidades de cálculo suministradas de los potenciadores. Por tanto, las tareas de cálculo de determinadas exigencias pueden ser transferidas a los potenciadores apropiados.
40

Un historial de cálculo, también designado como registro de diarios de cálculo, puede también ser aplicado para una asignación dinámica. Por tanto, las tareas de cálculo pueden ser empíricamente evaluadas mediante el cálculo de al menos un primer potenciador y el registro de los tiempos de respuesta y, así mismo, mediante el tratamiento de la misma tarea de cálculo sobre al menos un potenciador adicional y el registro de los tiempos de respuesta. Por tanto,
45 las capacidades de los potenciadores pueden ser registradas, empíricamente evaluadas y por tanto ser asignadas a los nodos de cálculo en función de las capacidades requeridas y de sus capacidades previstas. Las tareas de cálculo específicas pueden comprender informaciones de prioridad, las cuales indiquen hasta qué punto es urgente deban ser calculada esta tarea de cálculo específica. También puede resultar que los nodos de cálculo específicos proporcionen una prioridad, la cual indique hasta qué punto es urgente un tratamiento de una tarea de cálculo, o al menos una parte de una tarea de cálculo, y su comparación con otras partes de las tareas de cálculo que se originen procedentes de otros nodos de cálculo. Por tanto, es posible proporcionar informaciones de prioridad por lo que
50 respecta a partes aisladas de la tarea de cálculo así como informaciones de prioridad que se refieran a nodos de cálculo.

Una vez que un potenciador es asignado a un nodo de cálculo, el potenciador procesa partes específicas de una
55 tarea de cálculo. Esto puede llevarse a cabo mediante una llamada de procedimiento distante, una transferencia de parámetros o una transmisión de datos. La complejidad de la parte de la tarea de cálculo puede ser evaluada en función de una transferencia de parámetros. En el caso de que un parámetro contenga una matriz, la complejidad de la transferencia de parámetros puede ser evaluada por el número de las dimensiones de la matriz.

Para situar el interfaz la infraestructura de comunicación se dispone una unidad de colocación en interfaz, la cual
60 esté dispuesta entre un nodo de cálculo y la infraestructura de comunicación. Siendo una unidad de posición en interfaz diferente de la primera unidad de posición en interfaz, y ello puede disponerse entre el potenciador y la

infraestructura de comunicación. La unidad de comunicación de interfaz puede ser diferente del nodo de cálculo y también diferente del potenciador. La unidad de posición en interfaz proporciona una funcionalidad de red, sin que se disponga para procesar partes de la tarea de cálculo. La unidad de posición en interfaz simplemente proporciona la funcionalidad por lo que respecta a los problemas de administración y comunicación de las tareas de cálculo. Por ejemplo, dicha unidad puede proporcionar una funcionalidad por lo que respecta a un encaminamiento y a una transmisión de los datos referidos a la tarea de cálculo.

Así mismo, la aceleración puede también llevarse a cabo de manera inversa mediante la externalización de al menos una parte de la tarea de cálculo desde al menos un potenciador hasta al menos un nodo de cálculo. Por tanto, el flujo de control y de informaciones se invierte por lo que respecta a las cuestiones anteriormente introducidas de la invención.

De acuerdo con un aspecto de la presente invención, la asignación predeterminada puede formarse de acuerdo con al menos un grupo de técnicas de especificación de matriz, comprendiendo el grupo: una lógica temporal, una matriz de asignación, una tabla de asignación, una función de probabilidad y una función de coste. Esto puede proporcionar la ventaja de que se pueda formar la métrica de asignación predeterminada mediante el uso de un modelo formal o semiformal o un tipo de datos.

De acuerdo con otro aspecto de la presente invención, la métrica de asignación predeterminada se especifica en función de al menos un elemento entre un grupo de parámetros de asignación, comprendiendo el grupo: informaciones de recursos, informaciones de coste, informaciones de complejidad, informaciones de escalabilidad, un registro de diarios de cálculo, informaciones de compilación, informaciones de prioridad y una estampilla de tiempo. Esto puede proporcionar la ventaja de que la asignación puede llevarse a cabo de manera dinámica en el tiempo de ejecución con arreglo a la consideración de los diferentes parámetros de tiempo de ejecución y en respuesta a características de las tareas de cálculo específicas.

De acuerdo con otro aspecto de la presente invención, la asignación de al menos un potenciador a uno de la pluralidad de nodos de cálculo desencadena al menos una señal entre un grupo de estas, comprendiendo el grupo: una llamada de procedimiento distante, una transferencia de parámetros y una transmisión de datos. Esto puede proporcionar la ventaja de que al menos una parte de las tareas de cálculo sean transferidas desde un nodo de cálculo hasta al menos un potenciador.

De acuerdo con otro aspecto de la presente invención, cada nodo de cálculo y cada potenciador se sitúa en interfaz con la infraestructura de comunicación respectivamente a través de la unidad de posición en interfaz. Esto puede proporcionar la ventaja de que los datos pueden ser comunicados por medio de la infraestructura de comunicación sin la necesidad de un nodo de cálculo intermedio. Por tanto, no se requiere acoplar un potenciador con un nodo de cálculo directamente, si no que se consigue una asignación dinámica.

De acuerdo con otro aspecto de la presente invención, la unidad de posición en interfaz comprende al menos un grupo de componentes, comprendiendo el grupo: una interfaz virtual, un sustituto temporal, un receptáculo, un controlador de red y un dispositivo de red. Ello puede proporcionar la ventaja de que los nodos de cálculo como los potenciadores pueden también ser virtualmente conectados a la comunicación y a la infraestructura. Así mismo, puede fácilmente accederse a las infraestructuras de comunicación existentes.

De acuerdo con otro aspecto de la presente invención, la comunicación y la infraestructura comprende al menos un grupo de componentes, comprendiendo el grupo: un bus, un enlace de comunicación, una unidad de conmutación, un encaminador y una red de alta velocidad. Esto puede proporcionar la ventaja de que las infraestructuras de comunicación existentes puedan ser utilizadas y puedan crearse nuevas infraestructuras de comunicación mediante dispositivos en red habitualmente disponibles.

De acuerdo con otro aspecto de la presente invención, cada nodo de cálculo puede al menos un elemento entre un grupo de componentes, comprendiendo el grupo: un procesador de múltiples conductores, un racimo, un ordenador, una estación de trabajo y un procesador de usos múltiples. Esto puede proporcionar la ventaja de que los nodos de cálculo sean escalables en gran medida.

De acuerdo con otro aspecto de la presente invención, el al menos un potenciador comprende al menos un grupo de componentes, comprendiendo el grupo: un procesador de múltiples conductores, un procesador escalable, un coprocesador, una unidad de procesamiento gráfico, un racimo de muchos procesadores multiconductores, y un procesador monolítico. Esto puede proporcionar la ventaja de que los potenciadores sean implementados para procesar problemas específicos a gran velocidad.

Los nodos de cálculo típicamente aplican procesadores que comprenden una unidad de control extensiva en cuanto diversas tareas de cálculo tienen que ser procesadas de manera simultánea. Los procesadores que son aplicados en los potenciadores típicamente comprenden una unidad lógica aritmética extensiva y una simple estructura de control cuando está siendo comparada con los procesadores de nodos de cálculo. Por ejemplo, SIMD, también referido a ordenadores individuales de múltiples datos de instrucciones, pueden encontrar aplicación en los potenciadores. Por tanto, los procesadores que son aplicados en los nodos de cálculo difieren en su diseño de procesador en comparación con los procesadores que se aplican en los potenciadores.

De acuerdo con otro aspecto de la presente invención, el gestor de recursos está dispuesto para actualizar dicha métrica de asignación determinada durante el cálculo de al menos una parte de dicha tarea de cálculo. Esto puede proporcionar la ventaja de que la asignación de los potenciadores a los nodos de cálculo puede llevarse a cabo de forma dinámica en el tiempo de ejecución.

- 5 El objeto se resuelve también mediante un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con las características de la reivindicación 11 de la patente.

Por consiguiente, se proporciona un procedimiento para operar una disposición de ordenadores en racimo para tratar una tarea de cálculo, comprendiendo el procedimiento:

- 10 - el cálculo de al menos una primera parte de la tarea de cálculo por al menos dos de la pluralidad de nodos de cálculo, situando en interfaz cada nodo de cálculo una infraestructura de comunicación;
- el cálculo de al menos una segunda parte de la tarea de cálculo mediante al menos un potenciador, situándose cada potenciador en interfaz con la infraestructura de comunicación; y
- 15 - la asignación de al menos un potenciador a uno entre la pluralidad de nodos de cálculo mediante un gestor de recursos para el cálculo de la segunda parte de la tarea de cálculo, llevándose la asignación a cabo como una función de una métrica de asignación predeterminada.

Así mismo, se proporcionan un programa informático configurado para llevar a cabo el procedimiento introducido así como un medio legible por ordenador para almacenar el producto de programa informático.

A continuación se describirá la invención simplemente de forma ilustrativa con referencia a los dibujos que se acompañan:

- 20 La Fig. 1 muestra una disposición de ordenadores en racimo de acuerdo con el estado de la técnica
- La Fig. 2 muestra una ilustración esquemática de una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención.
- La Fig. 3 muestra una ilustración esquemática de una disposición de ordenadores en racimo de acuerdo con un aspecto adicional de la presente invención.
- 25 La Fig. 4 muestra una ilustración esquemática de un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención.
- La Fig. 5 muestra una ilustración esquemática de un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con otro aspecto de la presente invención.
- 30 La Fig. 6 muestra una ilustración esquemática de flujo de control de una disposición de ordenadores en racimo de acuerdo con un aspecto más de la presente invención.
- La Fig. 7 muestra una ilustración esquemática de un flujo de control que implementa una aceleración inversa de una disposición de ordenadores en racimo de acuerdo con otro aspecto de la presente invención.
- 35 La Fig. 8 muestra una ilustración esquemática de flujo de control de una disposición de ordenadores en racimo de acuerdo con un aspecto más de la presente invención.
- La Fig. 9 muestra una ilustración esquemática de una topología de red de una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención.

En las líneas que siguen, se designarán los mismos conceptos con los mismos signos de referencia si no se indica lo contrario.

- 40 La Fig. 2 muestra una disposición de ordenadores en racimo que comprende un racimo C de potenciadores, BG. El racimo comprende, en la presente forma de realización, cuatro nodos de cálculo, también designados como CN, así como tres potenciadores, también designados como B. Un acoplamiento flexible de potenciadores con los nodos de cálculo se establece mediante una infraestructura de comunicación, IN, como una llamada de interconexión. Este tipo de infraestructura de comunicación, IN, puede ser implementado, por ejemplo, utilizando InfiniBand. Por tanto,
- 45 cada uno de los potenciadores B puede ser compartido por cualquiera de los nodos de cálculo, CN. Así mismo, se puede llevar a cabo una virtualización a nivel de racimo. Cada potenciador, o al menos una parte de los potenciadores puede ser virtualizado y disponer de ellos en los nodos de cálculo de manera virtual.

- En la presente forma de realización, las tareas de cálculo son procesadas por al menos uno de los nodos de cálculo, CN, y al menos una parte de las tareas de cálculo puede ser transferida hacia al menos uno de los potenciadores B. Los potenciadores B están dispuestos para calcular problemas específicos y proporcionar una potencia de
- 50

- tratamiento o procesamiento específica. Por tanto, los problemas pueden ser externalizados desde uno de los nodos de cálculo, CN, hasta los potenciadores B, ser calculados por el potenciador y el resultado puede ser redistribuido hacia el nodo de cálculo. La asignación de potenciadores ESB a los nodos de cálculo, CN, puede llevarse a cabo por un gestor de recursos, también designado como RN. El gestor de recursos inicializa una primera asignación y, a continuación, establece una asignación dinámica de potenciadores B a los nodos de cálculo, CN.
- Para la comunicación entre potenciadores y los nodos de cálculo se proporciona una interfaz de programación de aplicaciones, también designada como API. Los potenciadores B pueden ser controlados de manera transparente por los nodos de cálculo mediante las respectivas llamadas de las funciones API. La API abstrae y potencia los modelos de programación finales efectivos de los potenciadores. Así mismo, la API puede proporcionar unos medios para la tolerancia a fallos en el caso del fracaso de un potenciador. Un protocolo de comunicación implicado en las llamadas API puede ser dispuesto en capa sobre la parte superior de una capa de comunicación. A continuación se proporciona una descripción breve de un conjunto de llamadas API de acuerdo con un aspecto de la presente invención, en el que el parámetro "acelerador" puede especificar el potenciador contemplado:
- 5 - `aanInit (acelerador)`
inicializa el potenciador antes del uso
 - 15 - `aanFinalize (acelerador)`
Libera las informaciones de contabilidad sobre el potenciador después del uso
 - `aanMemAlloc (dirección, tamaño, acelerador)`
Asigna los tamaños de Bytes de memoria sobre el potenciador contemplado
 - 20 - `aanMemFree (dirección, acelerador)`
Devuelve la dirección de memoria del dispositivo asignado
 - `aanMemFree (dirección, acelerador)`
Libera el inicio de memoria en la dirección sobre el potenciador referenciado
 - `aanMemCpy (dst, src, tamaño, dirección, acelerador)`
Copia la dirección de memoria desde el tamaño de los Bytes del src a dst
 - 25 - La dirección de la operación de la copia puede ser:
 - (i) Potenciador a host,
 - (ii) Host a potenciador
 - `aanKernelCreate (file_name, funct_name, kernel, acelerador)`
Crea un núcleo definido por el nombre del archive
 - 30 - `(file_name)` y el nombre de la función `(funct_name)` para la ejecución del potenciador contemplado
Retorna el manejo hacia el núcleo
 - `aanKernelSetArg (Núcleo, índice, tamaño, alineación, valor)`
Define el argumento para la ejecución del núcleo por su índice en la lista de argumentos, el tamaño, las exigencias de alineación (*align*), y valor
 - 35 - `aanKernelRun (kernel, grid_dim, block_dim)`
Comienza la ejecución del núcleo sobre el potenciador asociado con el núcleo en una llamada previa al `acKernelCreate()`.
El número de hebras se determina por el número de hebras por bloque (*block_dim*) y el número de bloques de la red (*grid_dim*)
 - 40 - `aanKernelFree (kernel)`
Libera los recursos asociados con el núcleo

La Fig. 3 muestra otra disposición en racimo de acuerdo con un aspecto de la presente invención, la disposición de ordenadores en racimo representada está dispuesta para calcular tareas de cálculo específicas, especialmente en el contexto de una tecnología de racimos de alto rendimiento. Un análisis más intenso de las características del

portafolio de los códigos de aplicación de los racimos de alto rendimiento científicos revela que muchos códigos con Exascale necesita incluir, por un lado, unos bloques de código que estén adecuadamente indicados para el Exascaling, y, por otro lado, dichos bloques de código que son demasiado complejos para ser escalables. A continuación se efectúa la distinción entre altamente escalable y complejo al nivel de los bloques de código y se introducen las nociones de Exascale Code Blocks (ECB) y Code Blocks complejos (CCB).

Evidentemente, no hay ningún código puramente escalable en gran medida y no hay tampoco ningún código estrictamente complejo. Cada código presenta unos elementos complejos altamente escalables y menos escalables. De hecho, hay un continuo entre ambos extremos. Es destacar que muchos elementos de un código no requieren una elevada escalabilidad sino que, por el contrario, requieren una gran memoria local. Así mismo, es evidente que todos los elementos de comunicación presentan una gran ventaja con arreglo a un pequeño paralelismo.

Para estos problemas, cuando se produce un correcto equilibrio entre los ECBs y los CCBs en términos de las relativas cantidades de memoria (esto es, los grados de libertad transferidos de las cantidades relativas de memoria, esto es, los grados de libertad transferidos del ECB con respecto al CCB), los tiempos de ejecución y de datos que deben ser intercambiados, sugiere por sí mismo adaptarse a esta situación por medio de una solución arquitectónica específica. Consistiendo esta solución en un esquema de ordenadores en racimo tradicional junto con un potenciador Exascale con unos potenciadores firmemente conectados y que estén conectados con un racimo por medio de la red del racimo. Este esquema dualístico presenta la ventaja de ensanchar sustancialmente el campo de aplicación estrecha anticipada de los puros sistemas Exascale.

Surge un modelo arquitectónico de naturaleza aproximada en el que las partes altamente escalables o ECBs de un código de aplicación son ejecutadas sobre una arquitectura multinúcleo paralela, a la que se accede dinámicamente, mientras que los CCBs son ejecutados sobre un sistema de racimo tradicional adecuadamente dimensionado, que incluye la conectividad junto con un sistema dinámico refinado de asignaciones de recursos.

Los racimos en la Exascale requieren unos elementos de virtualización para garantizar su resiliencia y fiabilidad. Aunque los aceleradores locales, en principio, posibilitan una visión sencilla del entero sistema y en particular pueden utilizar el ancho de banda local extremadamente elevado, son elementos hardware absolutamente estáticos, adecuadamente indicados para su recolección o paraletización maestro - esclavo. Por tanto, sería difícil incluirlos en una capa de software de virtualización. Así mismo, no habría tolerancia a fallos si un acelerador fallara, y no hubiera ninguna tolerancia respecto de una sobre o subscripción.

Los nodos de cálculo, CN, del racimo están internamente acoplados por una interconexión de racimo estándar, por ejemplo, Mellanox InfiniBand. Esta red está extendida para incluir también el potenciador ESB. En la figura se han dibujado tres de dichos potenciadores. Cada uno de los ESBs consisten en una multitud de aceleradores multinúcleo conectados por una red de baja latencia rápida específica.

Esta conexión de los CNs con los ESBs es muy flexible. Resulta posible una compartición de la capacidad de los aceleradores entre los nodos de cálculo. La visualización del nivel de racimo no resulta obstaculizada por el modelo y se puede explotar el completo paralelismo de los ESBs. La asignación de ESB a CN avanza a través de un gestor de recursos, RM, dinámico. Una asignación estática en tiempo de inicio puede convertirse en dinámica en tiempo de ejecución. Toda la comunicación CN - ESB avanza por medio del protocolo de red de racimo. La comunicación intraAC requerirán nuevas soluciones. La asignación de ESBs puede seguir la aplicación que necesite y la tolerancia a fallos se garantiza en el caso de fallos de acelerador mientras todos los nodos de cálculo compartan la misma capacidad de crecimiento.

Como elemento de cálculo del potenciador se puede aplicar el procesador multinúcleo de Intel Knight's Corner (KC). El chip KC consistirá en más de 50 núcleos y se espera contar con una capacidad de cálculo de DP de por encima de 1 Teraflop/s por chip. Con 10,000 elementos se estaría al alcance de un rendimiento total de 10 de Petaflop/s. El predecesor del KC, el procesador Knight's Ferry (KF) se utilizará en el proyecto para crear un sistema piloto a base de PCIe para estudiar el concepto racimo - potenciador (CN - ESB).

Cuando la velocidad de cálculo del KF sobrepase los procesadores comerciales actuales en un factor de aproximadamente 10, el sistema de comunicación intraESB tiene que ser dimensionado en la medida correspondiente. El sistema de comunicación del ESB requiere al menos 1 Terabit/s por tarjeta utilizado como una implementación de un sistema bus, que proporcione una tasa de comunicación de 1.44 Terabit/s por tarjeta. Ello se concreta en una topología de 3d que proporciona 6 enlaces por tarjeta. En relación con su sencillez esta topología resulta aplicable para un potenciador en base a los aceleradores multinúcleo. Incluso con dos direcciones reservadas para un encaminamiento anticipado, EXTOLL, puede saturar el rendimiento del PCI Express por lo que se refiere a la tasa de transmisión de datos. La latencia puede alcanzar los 0.3 μ s, cuando se base en una realización ASIC. Actualmente, EXTOLL, se realiza por medio de FPGAs.

La Fig. 4 muestra un diagrama de flujo para ilustrar un aspecto de un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con la presente invención. En una primera etapa 100 al menos una primera parte de una tarea de cálculo es calculada por al menos dos de la pluralidad de nodos de cálculo, CN, situándose en interfaz cada nodo de cálculo, CN, en una infraestructura de comunicación, IN. Así mismo, se lleva a cabo un cálculo

- de al menos una segunda parte de la tarea de cálculo en la etapa 101 por al menos un potenciador B, situándose cada potenciador B en interfaz con la infraestructura de comunicación, IN. Así mismo, se lleva a cabo la asignación de al menos un potenciador B a uno de la pluralidad de nodos de cálculo, CN, en la etapa 102 mediante un gestor de recursos, RM, para el cálculo de la segunda parte de la tarea de cálculo. Como la flecha a la derecha en la Fig. 4 indica, el flujo de control puede retroapuntar a la etapa 100. Después de la asignación de un potenciador B a al menos uno de la pluralidad de nodos de cálculo, CN, en la etapa 102, la asignación puede ser comunicada a un nodo de cálculo, CN, que utilice la asignación transmitida en posteriores etapas de externalización. Por tanto, el cálculo de al menos una segunda parte de la tarea de cálculo se efectúa en la etapa 101 en función de la etapa 102 de asignación.
- La Fig. 5 muestra un diagrama de flujo que ilustra un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención. En la presente forma de realización, después de la asignación en la etapa 202 del al menos un potenciador B a uno de la pluralidad de nodos de cálculo, CN, se lleva a cabo la etapa de cálculo 201 al menos de una segunda parte de la tarea de cálculo. Por tanto, es posible seleccionar un potenciador B específico y en base a la asignación que se establezca en la etapa 202, un potenciador B calcule la al menos segunda parte de la tarea de cálculo. Esto puede ser ventajoso en el caso de que la al menos segunda parte de la tarea de cálculo se transfiera al gestor de recursos, RM, el cual asigne un potenciador B a la segunda parte de la tarea de cálculo. El gestor de recursos, RM, puede entonces transmitir la segunda parte de la tarea de cálculo al potenciador B, sin necesidad de que el nodo de cálculo, CN, contacte directamente con el potenciador B.
- Con referencia a las Figs. 4 y 5, el experto en la materia apreciará que cualquiera de las etapas puede efectuarse de modo reiterado, en un orden diferente y pueden comprender también otras subetapas. Por ejemplo, la etapa 102 puede llevarse a cabo antes de la etapa 101, lo que se traduce en un cálculo de una primera parte de la tarea de cálculo, en una asignación de un potenciador a un nodo de cálculo y, finalmente al cálculo de la segunda parte de la tarea de cálculo. La etapa 102 puede comprender unas subetapas tales como el retorno de la calculada al menos segunda parte de la tarea de cálculo de nuevo al nodo de cálculo, CN. De ahí, el potenciador B retorna el resultado calculado de nuevo a los nodos de cálculo, CN. Los nodos de cálculo, CN, pueden utilizar el valor retornado para el cálculo de otras tareas de cálculo y pueden de nuevo transferir al menos una parte adicional de una tarea de cálculo a al menos uno de los potenciadores B.
- La Fig. 6 muestra un diagrama de bloques de un flujo de control de una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención. En la presente forma de realización, un nodo de cálculo, CN, recibe una tarea de cálculo y solicita un potenciador B para externalizar al menos una parte de la tarea de cálculo recibida. Por tanto, se accede a un gestor de recursos, RM, el cual transfiere la parte de la tarea de cálculo a un potenciador seleccionado B. El potenciador B calcula la parte de la tarea de cálculo y devuelve un resultado, que es indicado por la flecha más a la derecha. De acuerdo con otro aspecto de la presente forma de realización, el valor de retorno puede ser retrotraído al nodo de cálculo, CN.
- La Fig. 7 muestra un diagrama de bloques de un flujo de control, que implementa una aceleración inversa, de una disposición de ordenadores en racimo de acuerdo con aspecto de la presente invención. En la presente forma de realización, se lleva a cabo una aceleración de cálculo de tareas de cálculo que son calculadas por al menos un potenciador B mediante la asignación por al menos un nodo de cálculo CN, a al menos un potenciador B. De aquí, el controlador y el flujo de información es invertido por lo que respecta a la forma de realización mostrada en la Fig. 6.
- El cálculo de las tareas puede, por tanto, ser acelerado mediante las tareas de cálculo de externalización a partir de los potenciadores B hasta al menos un nodo de cálculo, CN.
- La Fig. 8 muestra un diagrama de bloques de un flujo de control de una disposición de ordenadores en racimo de acuerdo con otro aspecto de la presente invención. En la presente forma de realización, el gestor de recursos, RM, no pasa al menos una parte de la tarea de cálculo al potenciador B, sino que el nodo de cálculo, CN, solicita una dirección u otra identificación de un potenciador, B, el cual está dispuesto para calcular la al menos una parte específica de la tarea de cálculo. El gestor de recursos, RM, devuelve la dirección requerida al nodo de cálculo, CN. El nodo de cálculo, CN, no es capaz de acceder directamente al potenciador B por medio de la infraestructura de comunicación, IN. En la presente forma de realización, se accede por parte de la infraestructura de comunicación, IN, a través de las unidades de disposición en interfaz. Los nodos de cálculo, CN, acceden a la infraestructura de comunicación, IN, disponiendo en interfaz la unidad, IU1 y el potenciador B se dispone en interfaz con la infraestructura de comunicación, IN, mediante la unidad de disposición en interfaz IU2.
- Así mismo, el gestor de recursos, RM, está dispuesto para evaluar las capacidades de recursos del potenciador B y lleva a cabo la asignación, lo que significa la selección del potenciador B en función de las capacidades de recursos evaluadas de cada uno de los potenciadores B. Para llevar esto a cabo, el gestor de recursos, RM, puede acceder a la métrica de asignación, la cual puede estar almacenada en una base de datos, DB, o cualquier tipo de fuente de datos. El gestor de recursos, RM, está dispuesto para actualizar la métrica de asignación, la cual puede llevarse a cabo con arreglo a la utilización de un sistema de gestión de bases de datos. La base de datos, DB, puede ser implementada como cualquier tipo de almacenamiento. Por ejemplo, puede ser implementada como una tabla, un registro o un caché.

La Fig. 9 muestra una ilustración esquemática de una topología de red de una disposición de ordenadores en racimo de acuerdo con un aspecto de la presente invención.

5 En una forma de realización, los nodos de cálculo comparten una primera infraestructura de comunicación común, por ejemplo una topología en estrella con una unidad de conmutación central, S. Una adicional, segunda infraestructura de comunicación es suministrada para la comunicación de los nodos de cálculo, CN, con los nodos de potenciador, BN. Una tercera infraestructura de comunicación se suministra para la comunicación entre nodos potenciadores, BN. A continuación, una interfaz de red de alta velocidad para la comunicación entre nodos potenciadores, BN, puede estar provista de una interfaz de comunicación específica BN - BN. La infraestructura de comunicación BN - BN puede ser implementada como una topología en 3d.

10 En otra forma de realización, se disponen dos infraestructuras de comunicación una para la comunicación entre nodos de cálculo, CN, y otra infraestructura de comunicación para la comunicación entre los nodos potenciadores, BN. Ambas infraestructuras de comunicación pueden estar acopladas por al menos un enlace de comunicación desde la primera red hasta la segunda red o desde la segunda red hasta la primera red. A continuación, un nodo de cálculo, CN, seleccionado o un nodo potenciador, BN, seleccionado es conectado con respectivamente, otra red. En
15 la presente Fig. 9, un nodo potenciador, BN, está conectado con la infraestructura de comunicación de los nodos de cálculo, CN, con arreglo a la utilización de una unidad de conmutación, S.

En otra forma de realización, el propio grupo de potenciadores, BG, puede ser conectado a la infraestructura de comunicación de los nodos de cálculo, CN, o a una infraestructura de comunicación intermedia.

20 Las infraestructuras de comunicación pueden en términos generales diferir entre otras características en cuanto a su topología, ancho de banda, protocolos de comunicación, tasa de transferencia e intercambio de mensajes. Un potenciador B, puede, por ejemplo, comprender de 1 a 10,000 nodos potenciadores, BN, pero no está restringido a este límite. El gestor de recursos, RM, puede en términos generales gestionar partes de los nodos potenciadores, BN, y puede, por tanto, tabicar el número total de nodos potenciadores, BN, y formar dinámicamente unos potenciadores B a partir de dicho número de nodos potenciadores, BN. La unidad de conmutación S puede ser
25 implementada por un conmutador, un encaminador o cualquier dispositivo de red.

El experto en la materia apreciará la posibilidad de incluir otras disposiciones de los componentes de la disposición de ordenadores en racimo. Por ejemplo, se puede acceder a la base de datos, DB, mediante otros componentes, respectivamente nodos de la disposición de ordenadores en racimo. Los nodos de cálculo, CN, ilustrados así como el grupo de potenciadores, BG, ilustrado pueden ser uno entre muchos otros nodos de cálculo, CN, así como uno
30 entre muchos grupos potenciadores, BG, respectivamente, que accedan al gestor de recursos, RM, y / o a la infraestructura de comunicación, IN. Así mismo, la aceleración también puede llevarse a cabo de manera inversa externalizando al menos una parte de la tarea de cálculo a partir de al menos un potenciador B hasta al menos un nodo de cálculo.

35

REIVINDICACIONES

- 5 1.- Una disposición de ordenadores en racimo para el tratamiento de una tarea de cálculo, comprendiendo la disposición de ordenadores en racimo una pluralidad de nodos de cálculo (CN) cada uno de los cuales está dispuesto en interfaz con una infraestructura de comunicación (IN), estando al menos dos de los nodos dispuestos para calcular de manera conjunta al menos una primera parte de dicha tarea de cálculo; **caracterizada porque** la disposición de ordenadores en racimo comprende además:
- 10 una pluralidad de potenciadores (B), estando al menos un potenciador (B) de la pluralidad de potenciadores dispuesto para calcular al menos una segunda parte específica de dicha tarea de cálculo después de haber sido asignada a un nodo de cálculo, disponiéndose cada potenciador (B) en interfaz con dicha infraestructura de comunicación (IN) y en la que los potenciadores presentan un diseño de procesador que incorpora una unidad lógica aritmética relativamente extensa y una estructura de control relativamente simple en comparación con un diseño de procesador de los nodos de cálculo; y
- 15 un gestor de recursos (RM) que está dispuesto para llevar a cabo la asignación del al menos un potenciador (B) al nodo de cálculo (CN) para el cálculo de dicha segunda parte de dicha tarea de cálculo, llevándose a cabo la asignación en función de una métrica de asignación predeterminada,
- en la que la pluralidad de nodos de cálculo y la pluralidad de potenciadores están dispuestos de manera que, durante el tratamiento de dichas asignaciones de tareas de cálculo de los nodos de cálculo y de los potenciadores puedan disponerse de manera que al menos
- 20 (i) uno o más de los nodos de cálculo de la pluralidad de nodos de cálculo está dispuesto para comunicar con uno o más potenciadores de la pluralidad de potenciadores,
- (ii) uno o más de los potenciadores puede ser compartido por más de un nodo de cálculo de la pluralidad de nodos de cálculo, y
- (iii) cada uno de los potenciadores puede ser asignado a cada uno de los nodos de cálculo,
- 25 y en la que dicho gestor de recursos está dispuesto para llevar a cabo la asignación mediante el uso de la métrica predeterminada en un principio de tratamiento de dicha tarea de cálculo y para actualizar la métrica de asignación de dicho tratamiento para de esta manera llevar a cabo una asignación dinámica de potenciadores a los nodos de cálculo durante el tiempo de ejecución mediante
- (i) la inicialización de la asignación al principio del tratamiento mediante la utilización de la métrica de asignación predeterminada,
- 30 (ii) la alteración de la métrica de asignación, y
- (iii) la ejecución de una reasignación utilizando la métrica de asignación alterada durante dicho tratamiento de la tarea de cálculo.
- 35 2.- La disposición de ordenadores en racimo de acuerdo con la reivindicación 1, en la que dicha métrica de asignación predeterminada se forma de acuerdo con al menos una técnica entre un grupo de técnicas de especificación de métrica, comprendiendo dicho grupo: una lógica temporal, una matriz de asignación, una tabla de asignación, una función de probabilidad y una función de coste.
- 40 3.- La disposición de ordenadores en racimo de acuerdo con una de la reivindicaciones precedentes, en la que dicha métrica de asignación predeterminada se especifica en función de al menos un parámetro entre un grupo de parámetros de asignación, comprendiendo dicho grupo: informaciones de recursos, informaciones de coste, informaciones de complejidad, informaciones de escalabilidad, un registro de diarios de cálculo, informaciones de compilador, informaciones de prioridad y una estampilla temporal.
- 45 4.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que dicha asignación de al menos un potenciador (B) a uno de dicha pluralidad de nodos de cálculo (CN) desencadena al menos una señal entre un grupo de señales, comprendiendo dicho grupo: una llamada de procedimiento distante, una transferencia de parámetros, una transmisión de datos.
- 50 5.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que el nodo de cálculo (CN) y cada potenciador (B) se disponen en interfaz con dicha estructura de comunicación (IN) respectivamente por medio de una unidad de disposición en interfaz (IU1; IU2).
- 6.- La disposición de ordenadores en racimo de acuerdo con la reivindicación 5, en la que dicha unidad de disposición en interfaz (IU1; IU2) comprende al menos un componente entre un grupo de componentes, comprendiendo dicho grupo: una interfaz virtual, un sustituto temporal, un receptáculo, un controlador de red y un dispositivo de red.

- 7.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que dicha infraestructura de comunicación (IN) comprende al menos un componente entre un grupo de componentes, comprendiendo dicho grupo: un bus, un enlace de comunicación, una unidad de conmutación, un encaminador y una red de alta velocidad.
- 5 8.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que cada nodo de cálculo (CN) comprende al menos un componente entre un grupo de componentes, comprendiendo dicho grupo: un procesador multinúcleo, un racimo, un ordenador, una estación de trabajo y un procesador multipropósito.
- 10 9.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que dicha el menos un potenciador (B) comprende al menos un componente entre un grupo de componentes comprendiendo dicho grupo: un procesador multinúcleo, un procesador escalar, un coprocesador, una unidad de tratamiento gráfico, un racimo de procesadores multinúcleo y un procesador monolítico.
- 15 10.- La disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones precedentes, en la que dicho gestor de recursos (RM) está dispuesta para actualizar dicha métrica de asignación predeterminada durante el cálculo de al menos una parte de dicha tarea de cálculo.
- 20 11.- Un procedimiento para operar una disposición de ordenadores en racimo de acuerdo con una de las reivindicaciones 1 a 10 para el tratamiento de una tarea de cálculo, comprendiendo el procedimiento:
- el cálculo de al menos una parte de dicha tarea de cálculo por al menos dos de una pluralidad de nodos de cálculo (CN), disponiéndose en interfaz cada nodo de cálculo (CN) con una infraestructura de comunicación (IN); y
 - el cálculo de al menos una segunda parte de dicha tarea de cálculo por al menos un potenciador (B) de una pluralidad de potenciadores (B), disponiéndose cada potenciador (B) en interfaz con dicha infraestructura de comunicación (IN) después de asignar el al menos un potenciador (B) a uno de dicha pluralidad de nodos de cálculo (CN) por un gestor de recursos (RM), para el cálculo de dicha segunda parte de dicha tarea de cálculo, siendo efectuada dicha asignación en función de una métrica de asignación predeterminada,
- 25 en la que durante el tarea de dichas asignaciones de tarea de cálculo de nodos de cálculo y de potenciadores se disponen de manera que al menos
- (i) uno o más de los nodos de cálculo de la pluralidad de nodos de cálculo comunica con uno o más potenciadores de la pluralidad de potenciadores,
 - (ii) uno o más de los potenciadores puede ser compartido por más de un nodo de cálculo de la pluralidad de nodos de cálculo, y
 - (iii) cada uno de los potenciadores puede ser asignado a cada uno de los nodos de cálculo,
- 30 y en la que dicho gestor de recursos lleva a cabo la asignación utilizando la métrica predeterminada en un inicio de tratamiento de dicha tarea de cálculo y actualiza la métrica de asignación durante dicho tratamiento para de esta forma llevar a cabo una asignación dinámica de potenciadores a los nodos de cálculo durante el tiempo de ejecución mediante
- (i) la inicialización de la asignación al principio del tratamiento utilizando la métrica de asignación predeterminada,
 - (ii) la actualización de la métrica de asignación, y
 - (iii) la ejecución de una reasignación utilizando la métrica de asignación actualizada durante dicho tratamiento de la tarea de cálculo.
- 35 12.- Un producto de programa informático que está configurado para llevar a cabo el procedimiento de acuerdo con la reivindicación 11.
- 45 13.- Un medio legible por ordenador para almacenar dicho producto de programa informático de acuerdo con la reivindicación 12.

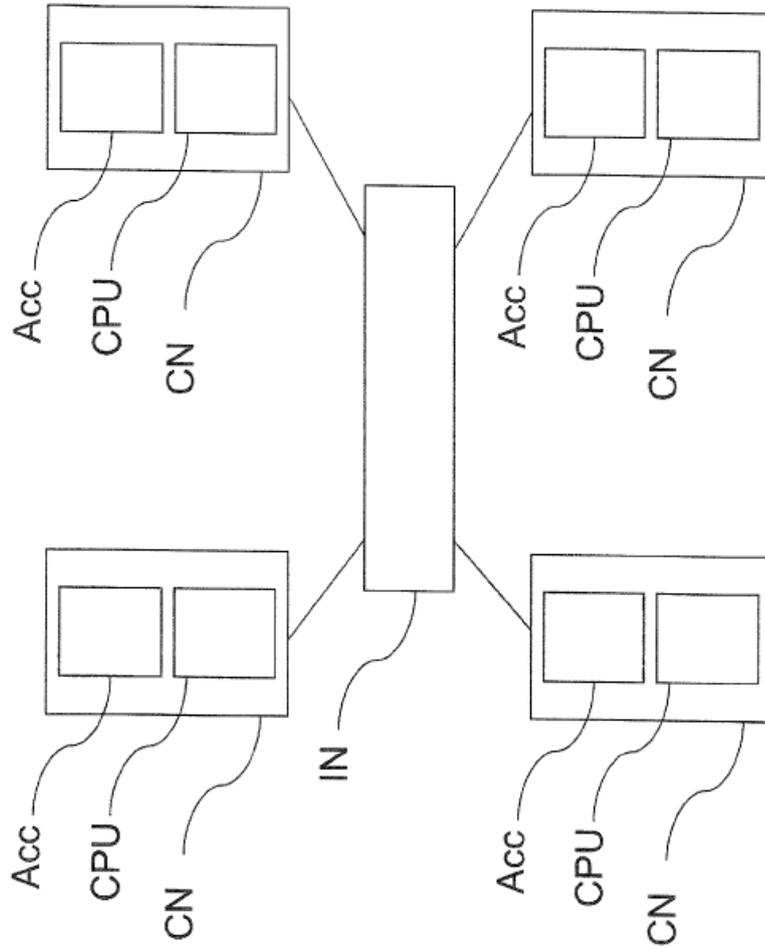


Fig.: 1

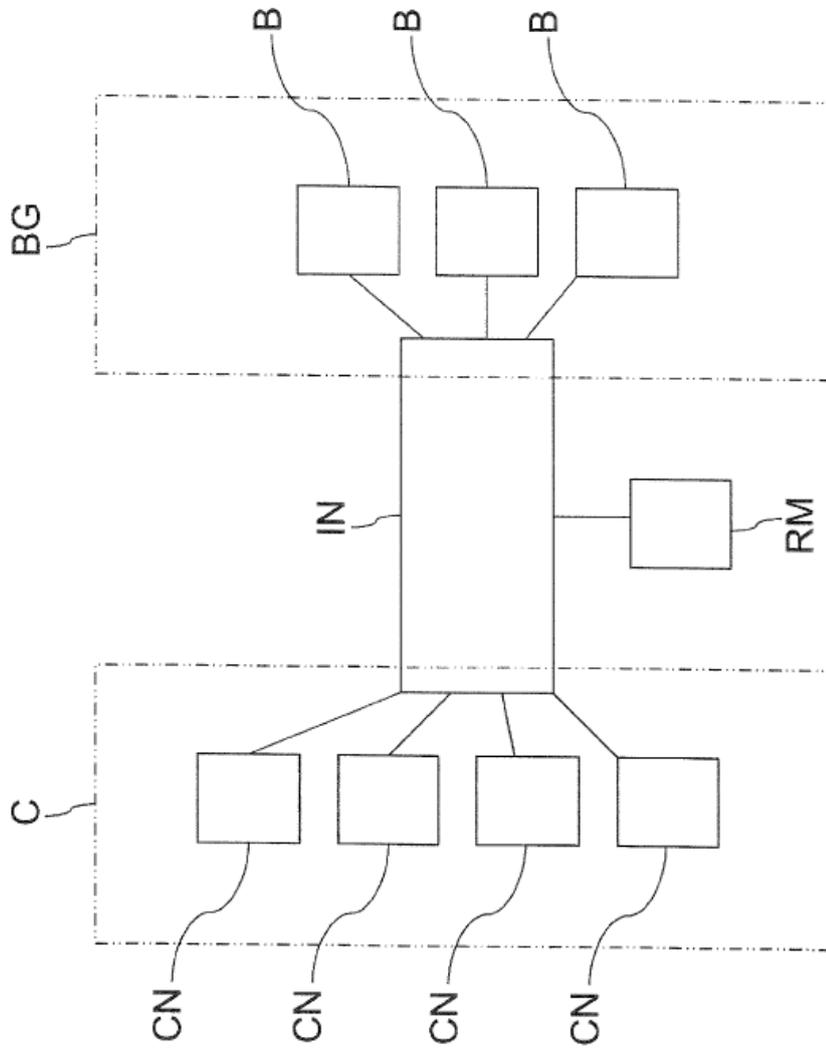


Fig.: 2

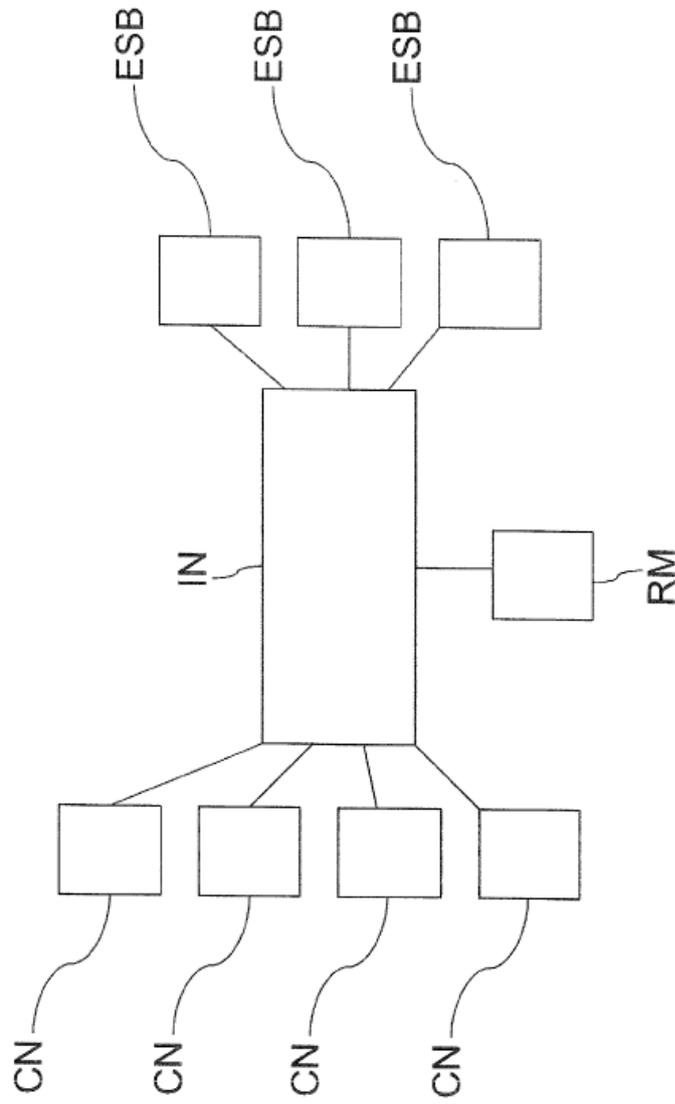


Fig.: 3

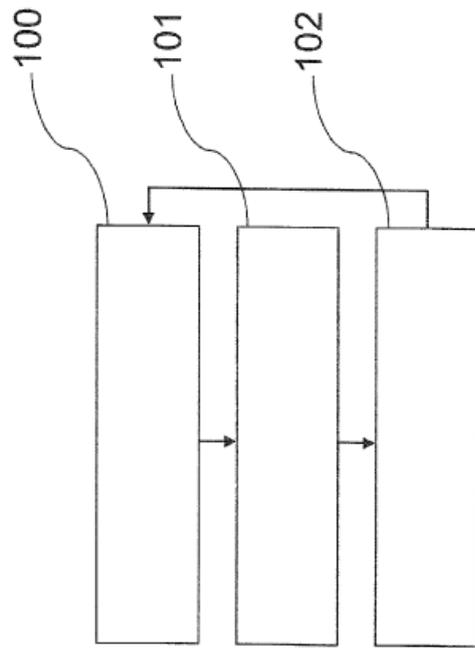


Fig.: 4

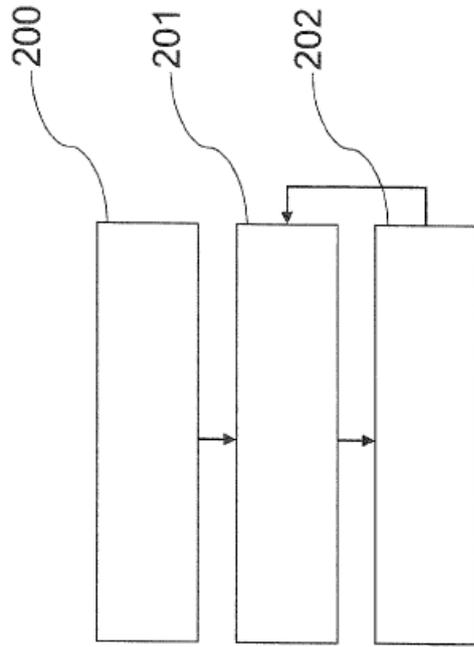


Fig.: 5

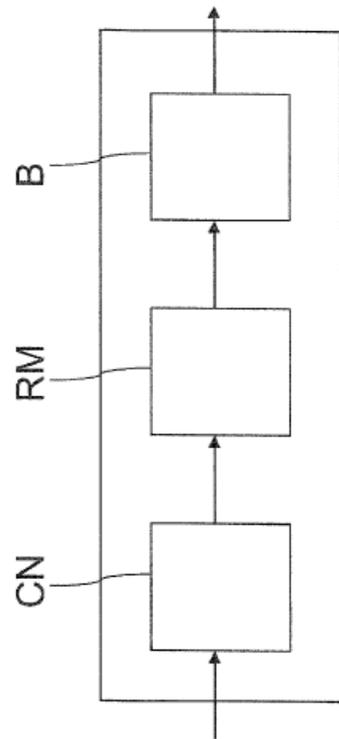


Fig.: 6

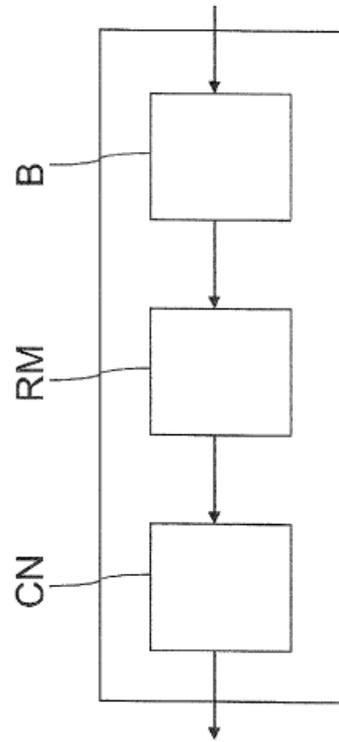


Fig.: 7

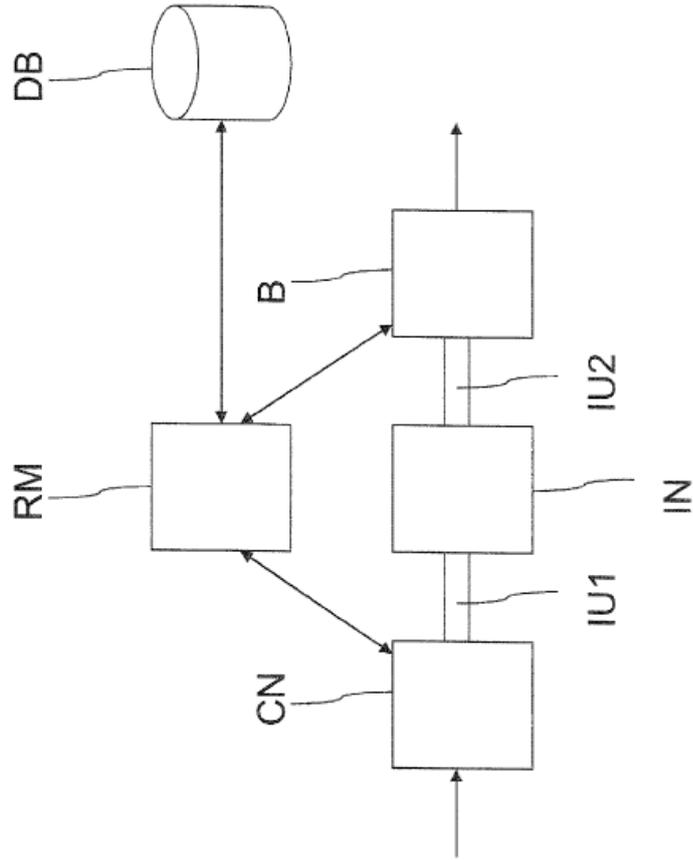


Fig.: 8

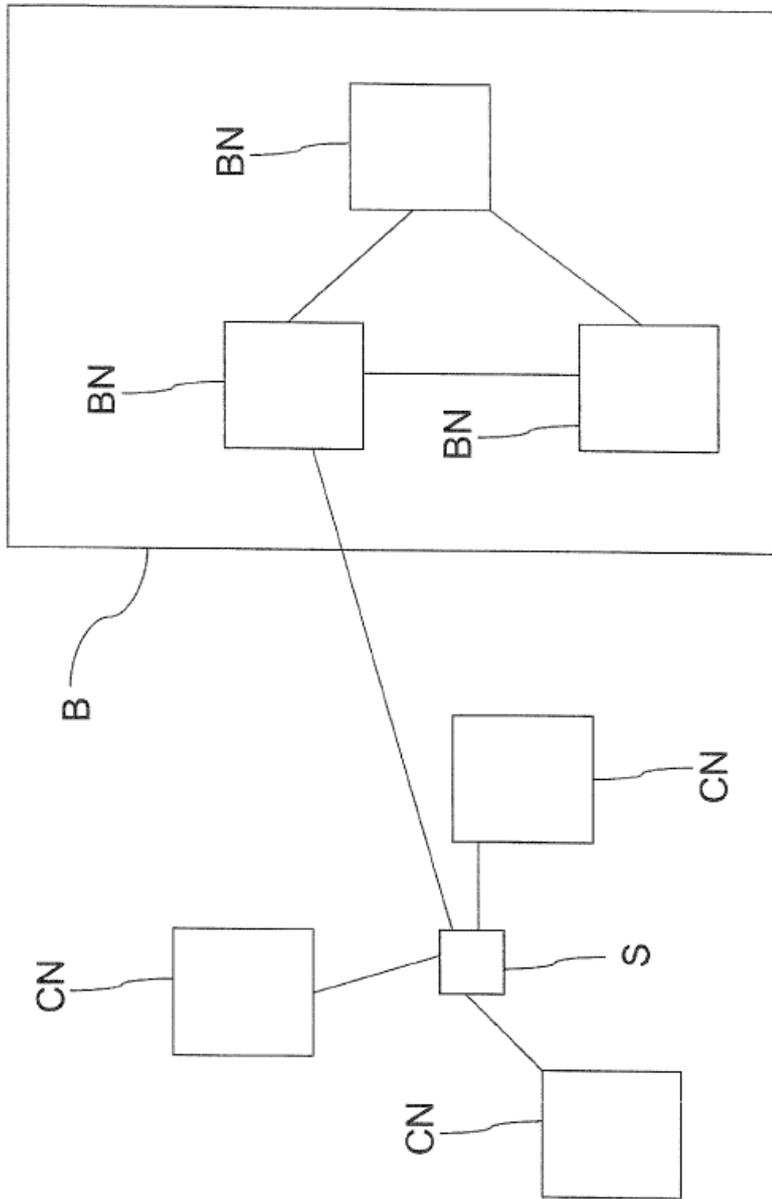


Fig.: 9