(19) Евразийское патентное ведомство

(21) 201800550 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.04.30

(22) Дата подачи заявки
2018.10.08

(51) Int. Cl. ***G06F 12/08*** (2016.01)
H04L 29/08 (2006.01)
H04L 12/24 (2006.01)
H04L 12/26 (2006.01)
H04L 12/28 (2006.01)
H04L 12/66 (2006.01)
H04L 12/721 (2013.01)
H04L 12/911 (2013.01)
H04L 12/933 (2013.01)
G06F 15/173 (2006.01)
G06F 17/50 (2006.01)

(54) ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ДЛЯ НАУЧНО-ТЕХНИЧЕСКИХ РАСЧЕТОВ

(96) **2018000118 (RU) 2018.10.08**

(71) Заявитель:
ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ИНФОРМАЦИОННЫЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР" (RU)

(72) Изобретатель:
Терентьев Александр Борисович (RU)

(57) Изобретение относится к системам обработки данных, предназначенным для научно-технических расчетов, в том числе суперкомпьютерным системам. Сущность изобретения: вычислительная система для научно-технических расчётов представляет собой кластер из преимущественно однотипных по архитектуре, конфигурации и составу оборудования вычислительных узлов, объединённых высокоскоростной сетью передачи данных с помощью высокоскоростного системного коммутатора сети передачи данных, выполненный с возможностью одновременного решения по меньшей мере одной прикладной вычислительной задачи, при этом каждый вычислительный узел содержит по меньшей мере два вычислительных модуля и высокоскоростной локальный коммутатор сети передачи данных, причём локальный коммутатор сети передачи данных подключен своими портами к каждому вычислительному модулю, входящему в состав вычислительного узла, и по меньшей мере одним портом подключен к системному коммутатору сети передачи данных, причём каждый вычислительный модуль включает процессор и оперативную память, причём процессор выполнен в виде "системы на кристалле" и включает центральный процессорный элемент и непосредственно подключённые к нему блок управления оперативной памятью и контроллер сети передачи данных, причём центральный процессорный элемент содержит по меньшей мере два независимых вычислительных ядра, имеющих одинаковую либо различную архитектуру, имеющих доступ к общей оперативной памяти.

201800550 A1

201800550 A1

НАЗВАНИЕ ИЗОБРЕТЕНИЯ: **ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ДЛЯ НАУЧНО-ТЕХНИЧЕСКИХ РАСЧЕТОВ**

ОБЛАСТЬ ТЕХНИКИ

Изобретение относится к системам обработки данных, предназначенным для научно-технических расчетов, в том числе суперкомпьютерным системам.

ПРЕДШЕСТВУЮЩИЙ УРОВЕНЬ ТЕХНИКИ

Общепринятой архитектурой современных высокопроизводительных вычислительных систем для научно-технических расчётов – суперкомпьютеров - является кластер, то есть множество преимущественно однотипных по архитектуре, конфигурации и составу оборудования вычислительных узлов, объединённых высокоскоростной локальной сетью передачи данных [Лацис А. Как построить и использовать суперкомпьютер. - М., Бестселлер, 2003 г.]. По этой архитектуре построены 437 из 500 самых мощных на сегодняшний день суперкомпьютеров (актуальный список Top500 за ноябрь 2017 года [www.top500.org]). При этом доминирующей элементной базой являются универсальные высокопроизводительные и очень сложные процессоры семейства Xeon или Xeon Phi производства фирмы Intel – их используют 471 суперкомпьютера из актуального списка Top500 [www.top500.org].

Данное единообразие при всех известных достоинствах (самые известные среди них – наличие большого количества совместимого системного и прикладного программного обеспечения и единообразие обслуживания), приводит и к ряду столь же существенных проблем. Во-первых, в силу фактически монопольного положения одного производителя цены держатся на очень высоком уровне. Во-вторых, прямым следствием высокой производительности и высокой сложности продуктов является их очень высокое энергопотребление, ставшее в последние годы фактически основным ограничителем повышения производительности суперкомпьютеров. В-третьих, процессоры, ориентированные на компьютеры общего назначения, имеют избыточный для чисто вычислительных установок набор встроенного периферийного оборудования, что также приводит к избыточному энергопотреблению. В-четвёртых, универсальная архитектура этих процессоров влечёт за собой избыточную сложность сетевой подсистемы вычислительного узла, что приводит, как известно, к существенным задержкам (латентности) при обмене данными между вычислительными узлами.

В последние годы, особенно в странах Европы и в Японии, активно проводятся работы по использованию в качестве основы для перспективных суперкомпьютеров более

простых процессоров, имеющих более простую архитектуру и, как следствие, относительно невысокое быстродействие, но при этом радикально более низкое энергопотребление и радикальное более низкую цену (в первую очередь – процессоры семейства ARM, широко применяемые в бытовой, мобильной и коммуникационной технике) [https://www.hpcwire.com/2017/01/19/cea-riken-partner-arm-exascale/; https://atos.net/en/2017/press-release/general-press-releases_2017_06_19/atos-expands-range-supercomputers-include-arm-processors-new-bull-sequana-x1310]. К тому же в последнее время сформировалась новая группа полупроводниковых приборов – «системы на кристалле» (SoC, 'System-on-Chip') на базе этих процессоров, представляющих собой собранные на одном кристалле процессор, схемы управления памятью, схемы управления вводом-выводом и другие устройства и позволяющие получить законченный вычислитель простым добавлением к нему минимального количества внешних компонентов – электропитание, память и физические порты ввода-вывода. Часто такие SoC имеют в своём составе дополнительные специализированные вычислительные компоненты – графический процессор (GPU) с возможностью выполнения вычислительных задач, векторный сопроцессор либо иные. Например, широко распространённые продукты корпорации Texas Instruments включают в себя от двух до шести вычислителей различных архитектур, более десяти каналов цифрового и аналогового ввода-вывода и другое оборудование [http://www.ti.com/processors/dsp/media-processors/products.html].

Одним из существенных факторов, огранивающих на сегодняшний день широкое применение данной процессорной архитектуры в суперкомпьютерах, является относительно невысокая единичная производительность этих процессоров, что приводит к необходимости использовать бОльшее количество процессоров (по сравнению, например, с процессорами Intel) для получения той же суммарной производительности.

Другая известная проблема является следствием первой – это высокая удельная стоимость и высокое энергопотребление оборудования локальной вычислительной сети – вследствие невысокой единичной производительности данных процессоров и невозможности создания многопроцессорных плат из-за отсутствия в них штатных аппаратных средств межпроцессорного взаимодействия требуемое количество портов системного коммутатора сети передачи данных оказывается существенно больше, чем в случае процессоров Intel.

Важнейшим достоинством мультипроцессорных вычислительных систем, построенных на базе процессоров ARM, является существенно меньшее энергопотребление как отдельных вычислительных узлов, так и всего кластера в целом по сравнению с традиционными системами на процессорах Intel. Например, на

экспериментальной установке MontBlanc продемонстрировано снижение потребления электроэнергии при решении задач гидродинамики до 2-х раз по сравнению с суперкомпьютером MareNostrum, имеющим традиционную архитектуру [Oyarzun G. Efficient CFD code implementation for the ARM-based Mont-Blanc architecture [Электронный ресурс] / G. Oyarzun, R. Borrell, A. Gorobets, F. Mantovani, A. Oliva // Future Generation Computer Systems, 2018, № 79. – Режим доступа: https://doi.org/10.1016/j.future.2017.09.029, свободный]. Установка MontBlanc построена на процессоре Samsung Exynos 5 Dual, который не имеет в себе сетевого (Ethernet) интерфейса. Поэтому фактически к каждому процессору добавляется микросхема интерфейса USB-Ethernet, которая потребляет дополнительную электроэнергию и замедляет передачу данных между узлами.

РАСКРЫТИЕ ИЗОБРЕТЕНИЯ

Задачей предлагаемого изобретения является расширение арсенала технических средств, применимых для построения суперкомпьютеров.

Поставленная задача решается тем, что в вычислительной системе для научно-технических расчётов, представляющей собой кластер из преимущественно однотипных по архитектуре, конфигурации и составу оборудования вычислительных узлов, объединённых высокоскоростной сетью передачи данных с помощью высокоскоростного системного коммутатора сети передачи данных, выполненный с возможностью одновременного решения по меньшей мере одной прикладной вычислительной задачи,

каждый вычислительный узел содержит по меньшей мере два вычислительных модуля и высокоскоростной локальный коммутатор сети передачи данных,

причём локальный коммутатор сети передачи данных подключен своими портами к каждому вычислительному модулю, входящему в состав вычислительного узла, и по меньшей мере одним портом подключен к системному коммутатору сети передачи данных,

причём каждый вычислительный модуль включает процессор и оперативную память,

причём процессор выполнен в виде «системы на кристалле» и включает центральный процессорный элемент и непосредственно подключённые к нему блок управления оперативной памятью и контроллер сети передачи данных,

причём центральный процессорный элемент содержит, по меньшей мере, два независимых вычислительных ядра, имеющих одинаковую либо различную архитектуру, имеющих доступ к общей оперативной памяти.

Предпочтительно, чтобы в заявленной вычислительной системе для научно-технических расчётов по меньшей мере один вычислительный модуль дополнительно содержал в себе по меньшей мере один специализированный вычислитель.

Также предпочтительно, чтобы по меньшей мере одно вычислительное ядро каждого центрального процессорного элемента выполнено с возможностью осуществления функций управления работой соответствующего центрального процессорного элемента и выполнения вычислительных процессов, а остальные ядра выполнены с возможностью выполнения только вычислительных процессов,

причём набор функций управления включает в себя по меньшей мере

приём из сети передачи данных и загрузку системного программного обеспечения,

приём из сети передачи данных и загрузку прикладного программного обеспечения и данных для каждого из ядер этого центрального процессорного элемента,

управление вычислительными процессами, выполняемыми на других ядрах этого центрального процессорного элемента и на специализированных вычислителях,

управление оперативной памятью вычислительного модуля,

выполнение обмена данными с другими вычислительными модулями, входящими в состав любого вычислительного узла и образующими временную подгруппу вычислительных блоков, решающих совокупно одну задачу одновременно.

Также предпочтительно, чтобы по меньшей мере один вычислительный модуль дополнительно содержал в себе устройства внешней памяти или устройства ввода-вывода или устройства отображения информации или несколько дополнительных устройств.

ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

По настоящему изобретению, для достижения поставленной цели используются процессоры, выполненные в виде «системы на кристалле» и включающие в себя многоядерный центральный процессорный элемент (ЦПЭ) и непосредственно подключённые к нему блок управления оперативной памятью и контроллер сети передачи данных. Процессор может также включать в себя интегрированный графический процессор (GPU) с возможностью выполнения вычислительных задач, либо векторный сопроцессор, либо иные специализированные вычислительные компоненты.

Каждый процессор совместно с подключёнными к нему элементами оперативной памяти представляет собой вычислительный модуль. В отдельных случаях вычислительные модули могут содержать в себе дополнительно подключенные к процессору внешние специализированные вычислители, выполненные, например, на основе графических процессоров (GPU), элементов программируемой логики (FPGA) или

другие. Также некоторые модули могут дополнительно содержать в себе устройства внешней памяти или устройства ввода-вывода или устройства отображения информации или несколько дополнительных устройств.

При этом одно или более ядер многоядерного ЦПЭ могут быть выполнены с возможностью осуществления функций управления работой данного ЦПЭ, в том числе приём из сети передачи данных и загрузку системного программного обеспечения, приём из сети передачи данных и загрузку прикладного программного обеспечения и данных для каждого из ядер этого ЦПЭ, управление вычислительными процессами, выполняемыми на других ядрах этого ЦПЭ и на интегрированных и/или внешних специализированных вычислителях, управление оперативной памятью вычислительного модуля, выполнение обмена данными с другими вычислительными модулями и выполнения вычислительных процессов, а остальные ядра выполнены с возможностью выполнения только вычислительных процессов. Таким образом, вспомогательные задачи, требующие многочисленных переключений и взаимодействия с внешними устройствами, выносятся на отдельный аппаратный ресурс, а большая часть вычислительных ресурсов (вычислительных ядер) эффективно загружается непрерывным, последовательным счётом, чем достигается повышение общей вычислительной эффективности.

Несколько вычислительных модулей конструктивно объединяются в вычислительный узел, который дополнительно содержит объединяющий их высокоскоростной локальный коммутатор сети передачи данных. Локальный коммутатор содержит также один или более внешних портов, посредством которых он соединяется с высокоскоростным системным коммутатором сети передачи данных, объединяющим множество вычислительных узлов в кластер.

К данному локальному коммутатору предъявляется по сути только одно существенное требование – минимальное время задержки при передаче пакетов. При этом от него не требуется никаких возможностей по каскадированию (объединению в группы), удалённому доступу, управлению потоками данных и прочих интеллектуальных функций (при этом исключаются сложные схемы управления и память большого объёма), а количество портов и подключаемых сетевых узлов минимальны (также минимальные требования к объёму памяти). Кроме того, вычислительные модули подключаются к нему в пределах одного конструктива, что делает ненужными схемы внешних физических интерфейсов (front-end) как на самом коммутаторе (за исключением внешних портов), так и на всех вычислительных модулях. Таким образом, и локальный коммутатор, и сетевая подсистема в целом могут иметь минимальную сложность и, как следствие, минимальные энергопотребление и стоимость.

Формула изобретения

1. Вычислительная система для научно-технических расчётов, представляющая собой кластер из преимущественно однотипных по архитектуре, конфигурации и составу оборудования вычислительных узлов, объединённых высокоскоростной сетью передачи данных с помощью высокоскоростного системного коммутатора сети передачи данных, выполненный с возможностью одновременного решения по меньшей мере одной прикладной вычислительной задачи,

отличающаяся тем, что

каждый вычислительный узел содержит по меньшей мере два вычислительных модуля и высокоскоростной локальный коммутатор сети передачи данных,

причём локальный коммутатор сети передачи данных подключен своими портами к каждому вычислительному модулю, входящему в состав вычислительного узла, и по меньшей мере одним портом подключен к системному коммутатору сети передачи данных,

причём каждый вычислительный модуль включает процессор и оперативную память,

причём процессор выполнен в виде «системы на кристалле» и включает центральный процессорный элемент и непосредственно подключённые к нему блок управления оперативной памятью и контроллер сети передачи данных,

причём центральный процессорный элемент содержит, по меньшей мере, два независимых вычислительных ядра, имеющих одинаковую либо различную архитектуру, имеющих доступ к общей оперативной памяти.

2. Вычислительная система по пункту 1, отличающаяся тем, что

по меньшей мере один вычислительный модуль дополнительно содержит в себе по меньшей мере один специализированный вычислитель.

3. Вычислительная система по пункту 1, отличающаяся тем, что

по меньшей мере одно вычислительное ядро каждого центрального процессорного элемента выполнено с возможностью осуществления функций управления работой соответствующего центрального процессорного элемента и выполнения вычислительных процессов, а остальные ядра выполнены с возможностью выполнения только вычислительных процессов,

причём набор функций управления включает в себя по меньшей мере

приём из сети передачи данных и загрузку системного программного обеспечения,

приём из сети передачи данных и загрузку прикладного программного обеспечения и данных для каждого из ядер этого центрального процессорного элемента,

управление вычислительными процессами, выполняемыми на других ядрах этого центрального процессорного элемента и на специализированных вычислителях,

управление оперативной памятью вычислительного модуля,

выполнение обмена данными с другими вычислительными модулями, входящими в состав любого вычислительного узла и образующими временную подгруппу вычислительных блоков, решающих совокупно одну задачу одновременно.

4. Вычислительная система по пункту 1, отличающаяся тем, что по меньшей мере один вычислительный модуль дополнительно содержит в себе устройства внешней памяти или устройства ввода-вывода или устройства отображения информации или несколько дополнительных устройств.

ЕВРАЗИЙСКОЕ ПАТЕНТНОЕ ВЕДОМСТВО

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ

(статья 15(3) ЕАПК и правило 42 Патентной инструкции к ЕАПК)

Номер евразийской заявки: 201800550

Дата подачи: 08/10/2018 | Дата испрашиваемого приоритета:

Название изобретения: ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ДЛЯ НАУЧНО-ТЕХНИЧЕСКИХ РАСЧЕТОВ

Заявитель: ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ИНФОРМАЦИОННЫЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР"

☐ **Некоторые пункты формулы не подлежат поиску (см. раздел I дополнительного листа).**

☐ **Единство изобретения не соблюдено (см. раздел II дополнительного листа)**

А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ: см. дополнительный лист

Согласно Международной патентной классификации (МПК)

Б. ОБЛАСТЬ ПОИСКА:

Минимум просмотренной документации (система классификации и индексы МПК) H04L 12 / 721, G06F 1750, G06F 15 / 78, H04L 12/46, H04L 12/74, G06F 17/50,

Другая проверенная документация в той мере, в какой она включена в область поиска:

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
X	US 8,738,860 B1 (Cambridge, MA (US)) 27.05.2104 кол.6 (стр.60) – кол. 13 (стр. 12)	1-4
X	US 2018 / 0077228 A1 (Advanced Micro Devices, Inc., Sunnyvale, CA (US)) 15.03.2018, абз. [0014]-[0027], [0048]-[0050]	1-4
X	US 9185023 B2 (NetSpeed Systems, San Jose, CA (US)) 10.11.2015, кол.1 (стр.1) – кол. 8 (стр.3)	1-4
X, Y	US 2014/0286179 A1 (EMPIRE TECHNOLOGY DEVELOPMENT, LLC, WILMINGTON, DE (US)) 25.09.2014, абз. [0066] – [0073]	1-4
Y	US 8559307 B2 (Empire Technology Development LLC, Wilmington, DE (US)) 15.10.2013, кол. 1 (стр. 17-35), кол. 2 (стр. 11-21), кол. 2 (стр.35) – кол.5 (стр. 30), кол.7 (стр.31) – кол.8 (стр.7), кол. 9 (стр.21) – кол.11 (стр.3)	1-4

☐ последующие документы указаны в продолжении графы В ☐ данные о патентах-аналогах указаны в приложении

* Особые категории ссылочных документов:
"A" документ, определяющий общий уровень техники
"E" более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее
"O" документ, относящийся к устному раскрытию, экспонированию и т.д.
"P" документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета
"D" документ, приведенный в евразийской заявке
"T" более поздний документ, опубликованный после даты приоритета и приведенный для понимания изобретения
"X" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности
"Y" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории
"&" документ, являющийся патентом-аналогом
"L" документ, приведенный в других целях

Дата действительного завершения патентного поиска: 07/05/2019

Уполномоченное лицо:

Ведущий эксперт
Отдела механики, физики и электротехники

Д.М. Ветров

Телефон: +7(495)411-61-61*325

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ

Номер евразийской заявки:
201800550

ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ (продолжение графы В)

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
Y	US 8473818 B2 (Empire Technology Development LLC, Wilmington, DE (US)) Jun. 25, 2013 кол. 1 (стр.12-28), кол. 2 (стр. 28) – кол.4 (стр. 7), кол.4 (стр. 61) - кол.7 (стр. 25), кол.8 (стр. 51) – кол.10 (стр. 13)	1-4
Y	US 2016/0182405 A1 (Intel Corporation, Santa Clara, CA (US)), Jun. 23, 2016 абз. [0001], [0002], [0015]-[0017], [0029], [0030], [0034] ,[0035], [0037]-[0040], [0046], [0047], [0053]-[0078]	1-4

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ

Номер евразийской заявки:
201800550

КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:

G06F 12/08 *(01/01/2006)*
H04L 29/08 *(01/01/2006)*
H04L 12/24 *(01/01/2006)*
H04L 12 / 26 *(01/01/2006)*
H04L 12/28 *(01/01/2006)*
H04L 12/66 *(01/01/2006)*
H04L 12/721 *(01/01/2013)*
HL04L 12/911 *(01/01/2006)*
H04L 12/933 *(01/01/2013)*
G06F 15/173 *(01/01/2006)*
G06F 15/17312 *(01/01/2006)*
G06F 17/50 *(01/01/2006)*

Дополнительный лист