СУПЕР КОМПЬЮТЕРЫ
ЗИМА-2014

Председатель редакционного совета
Владимир ВОЕВОДИН
Vladimir.voevodin@supercomputers.ru

Над номером работали:
Выпускающий редактор
Игорь ЛЁВШИН
Igor.levshin@supercomputers.ru
Арт-директор
Виктория ИВАШКОВА
Victoria.ivashkova@supercomputers.ru
Корректор
Юлия НИКУЛИНА

Тексты:
К. АМЕЛИН
Н. АМЕЛИНА
С АНДРЕЕВ
Д. БУДАЕВ
Олег ГРАНИЧИН
С. ДБАР
Г. КАШЕВАРОВА
А. КЛИМЕНТОВ
Александра КЛИМОВА
Кирилл КОЛГАНОВ
А. ЛАЦИС
Е. ЛЕВИН
Игорь ЛЁВШИН
И. МАЙОРОВ
Александр МУРАШОВ
А. ПЕПЕЛЯЕВ
Е. ПЛОТКИН
Николай СЕМЁНОВ
Сергей СЕРЁЖИН
П. СКОБЕЛЕВ
Михаил ТЮТЛЯЕВ
Е. ТЮТЛЯЕВА
Леонид ЧЕРНЯК

Иллюстрации:
Владимир КАМАЕВ
Обложка: Олег ПАЩЕНКО

Учредитель
Даниэль ОРЛОВ

Издатель
ООО «Издательство СКР-Медиа»

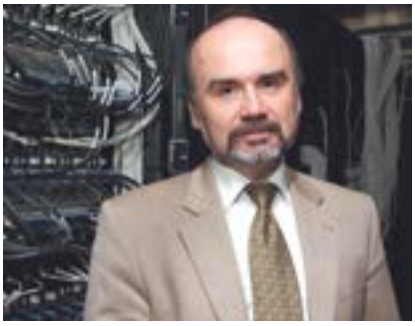
Генеральный директор
Даниэль ОРЛОВ
Daniel.orlov@supercomputers.ru

Адрес редакции и Издателя:
117342, Москва, ул. Бутлерова, 17Б
www.supercomputers.ru

Издание «СУПЕРКОМПЬЮТЕРЫ» зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации СМИ ПИ №ФС77-38346 от 10.12.2009

Тираж 5000 экз.

Отпечатано
Типография ООО «Вива-Стар»
107023, Россия, Москва,
ул. Электрозаводская, д. 20, стр. 3
www.vivastar.ru



От редакции

Мы снова с вами, с нашей уважаемой аудиторией, хотя и из-за чудес с ценами и валютами задержали выпуск зимнего номера почти на месяц. Благодаря поддержке коллег мы справились с трудностями, чем доказали времени и себе, что правильное и нужное имеет право на существование. А то, что существует, обязано иногда проходить такую проверку. В новом году у нас много планов. Редакционный портфель как никогда быстро пополняется, так что нам хочется выпускать журналы чаще, нежели четыре раза в год. Удерживает пока только осторожность: мир опять меняется, меняются правила. Но неизменным остается наше внимание к научной и внедренческой деятельности коллег. Надеемся, что многие из вас будут готовы поделиться опытом решения задач на суперкомпьютерах. Нам кажется, что публикация в научно-популярном журнале должна стать обязательной формой представления результатов работы помимо научной публикации, стендовых докладов и выступлений на конференциях. Аудитория журнала – не только и не столько специалисты, сколько люди, заинтересованные в развитии отечественной науки и промышленности. Демонстрируя свои достижения, вы рождаете цепочки аналогий, способных вытянуть против течения даже самый тяжелый корабль. Пусть это тяжело, но тяжело не значит плохо.

А во все времена, будь то тучные годы или нервное время перемен, главным всегда является сохранение школы и темпа движения научной мысли. То, чем мы занимаемся, немного больше слова «наука», это так или иначе постижение мира, осмысление его и, возможно, что устройство заново. Это не происходит без вдохновения.

И вдохновение передается от учителей к ученикам, оно чувствуется в результатах работ, в физических параметрах установок, в том азарте, с которым решается самая трудная задача.

Но сколь бы не увеличивалась мощность вычислителей, в мире всегда останется вдоволь задач, которые нужно решать и для решения которых нужен поистине космический масштаб разума. Так устроен мир.

Владимир В. Воеводин,
Председатель редакционного совета
журнала «Суперкомпьютеры»

В номере:

35

Синергия применения суперкомпьютерных современных технологий производства

Новости

Новые системы Oracle

Компания Oracle в январе представит новое, 6-е поколение своих устройств.
Oracle Virtual Compute Appliance X5 предполагает работу совместно с Oracle FS1 Series Flash Storage System и позиционируется вместе с ней как полное инфраструктурное решение. Оно может быть развернуто за считанные часы.
Oracle Database Appliance X5 идеально для распределенных офисов и для отделений. В системе есть все для вычислений, хранения, она поставляется со всем необходимым для поддержки работы ДБМС и приложений ПО. В решение входит InfiniBand.
Устройство Oracle Big Data Appliance X5 оптимизировано под Hadoop и NoSQL. Пользователям доступна и Oracle Big Data SQL, которая благодаря расширениям Oracle SQL дает возможность работать с Hadoop и NoSQL, не переписывая приложения с SQL-запросами.
Zero Data Loss Recovery Appliance X5 создано для тех, чьи данные наиболее критичны. Защита данных происходит на всех базах данных Oracle, почти не уменьшая производительность систем.

Население против Amazon

Планы по строительству ЦОД для AWS – облачных сервисов Amazon – вызвали протесты местного населения. Центр, строительство которого намечено в пригородах Вашингтона, потребует высоковольтных линий, которые пройдут через жилые районы. На запросы энергетическая компания ответила, что мачты ЛЭП будут высотой около 30 метров. Местные политики предложили закопать линии за счет хозяев ЦОД, что существенно удорожит проект. Компромиссный вариант, при котором ЛЭП сохранятся, но в сочетании с подземными силовыми коммуникациями, обойдется в $140 млн вместо $62.5. Мощности ЦОД предполагается использовать в том числе под государственные проекты.

A-Class

Новая система T-Platforms A-Class — это энергоэффективное модульное суперкомпьютерное решение с максимальной вычислительной плотностью и масштабируемостью до мультипетафлопсного уровня.

ПРОИЗВОДИТЕЛЬНОСТЬ И МАСШТАБИРУЕМОСТЬ

- 256 узлов с пиковой производительностью 535 Тфлопс.
- Масштабируемость до 192 систем (102 Пфлопс).

СИСТЕМА ОХЛАЖДЕНИЯ И ЭНЕРГОЭФФЕКТИВНОСТЬ

- Пиковая производительность в расчёте на ватт потребляемой энергии – 3570 Мфлопс/Вт*.
- Высокая энергоэффективность за счёт повторного использования нагретой воды и режима «свободного охлаждения» системы.
- Теплоизолированное стоечное шасси 52U с охлаждением компонентов горячей водой.
- Низкий уровень шума за счёт применения водяного охлаждения.

МОДУЛЬНАЯ АРХИТЕКТУРА

- Система уровня стойки с интегрированной сигнально-силовой объединительной платой и подсистемами электропитания и охлаждения.
- Архитектура шасси поддерживает различные конфигурации вычислительных модулей на базе однопроцессорных и двухпроцессорных узлов или узлов с несколькими ускорителями.

ГИБКАЯ СЕТЕВАЯ АРХИТЕКТУРА

- Поддержка топологии 3D и 4D Torus, Flattened Buttefly, Hypercube.
- Встроенная двухуровневая коммутация двух независимых сетей Gigabit/10Gigabit Ethernet.
- Встроенная коммутация двух независимых сетей FDR Infiniband.

ОТКАЗОУСТОЙЧИВОСТЬ

- Два независимых модуля управления с выделенными сетевыми фабриками Ethernet поддерживают горячую замену, обеспечивая функции отказоустойчивого управления и мониторинга системы.
- Независимые серверы управления A-Class позволяют отслеживать состояние компонентов системы и управлять нагрузкой и конфигурациями установленного ПО.
- Отказоустойчивая архитектура электроснабжения системы реализована на основе 8 независимых групп высокоэффективных блоков питания с горячей заменой блоков питания и режимом резервирования N+1.
- Мониторинг на уровнях шасси, секции и узла. В случае аварийной ситуации система управления автоматически отключает подачу воды и электропитания к шасси.
- Система охлаждения и электрическая и сетевая инфраструктуры разнесены по разным зонам стоечного шасси для безопасной эксплуатации и упрощения обслуживания A-Class в режиме «онлайн».

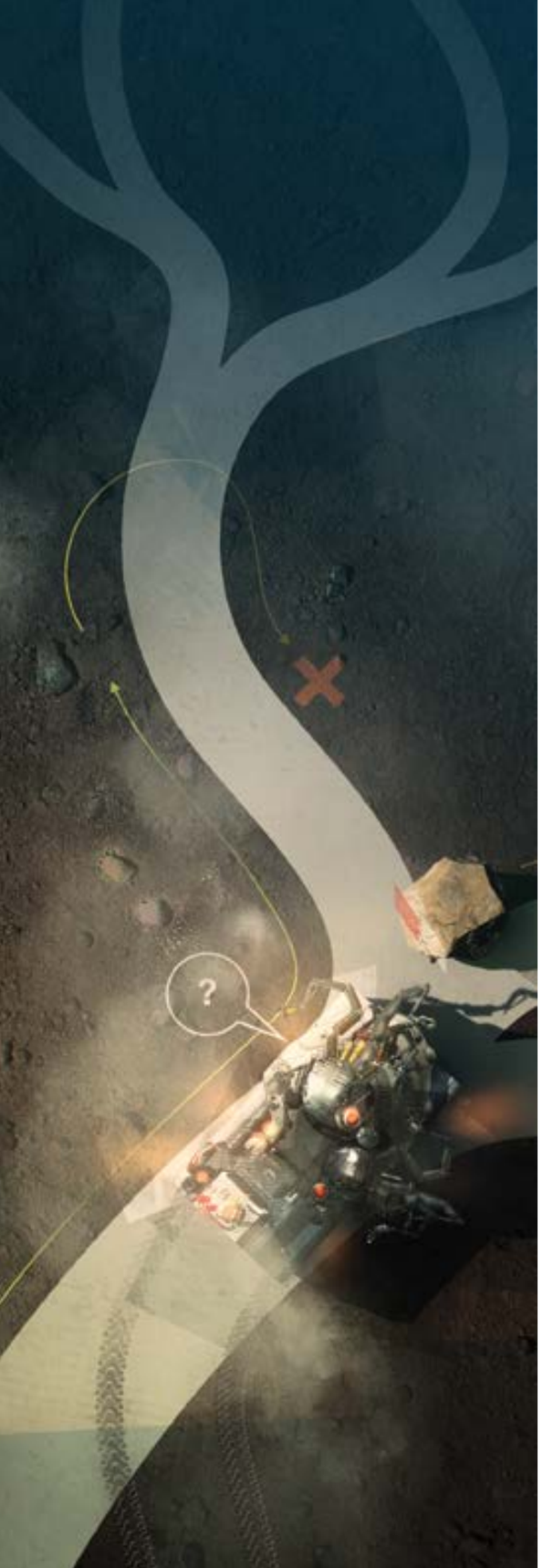
СИСТЕМНОЕ ПО

- Система управления кластером ClustrX: управление пользователями, менеджер ресурсов, система мониторинга, управление оборудованием, включая предсказание отказов, ПО автоматического отключения оборудования ЦОД ClustrX.Safe (САОО).
- ПО для поддержки новых топологий: маршрутизация в сети Dragonfly и Flattened Butterfly, оптимизированные библиотеки MPI для поддержки новых топологий и адаптивной маршрутизации.

все основания **считать**

Т-ПЛАТФОРМЫ

www.t-platforms.ru/a-class



Пейзаж с оленем и вертолетом

Текст Игорь Лёвшин

Иногда обсуждение ситуации в HPC-отрасли вызывает странные ассоциации. Мне, например, представился вдруг поселок на Крайнем Севере. Добраться до него можно двумя способами:
1. Олень.
2. Вертолет.
Если мы говорим о нашем, земном, то есть кремнии, то разговор обычно не покидает пределов тем многоядерных процессоров, ускорителей и неуклонного увядания закона Мура. Олени – неторопливые, надежные.
Если мы говорим об альтернативах, будущем, в разговор стремительно врываются квантовые компьютеры, занимая собой обычно все пространство беседы. Это вертолеты. Летают быстро, но когда еще будут?
Не видно на горизонте ни автобуса, ни снегохода, ни парохода, ни даже подводы, впряженной во владимирского тяжеловоза.
Но даже если они не видны, их следовало бы выдумать – из эстетических соображений, из законов композиции Большой Картины. Они все же имеются, но про них часто забывают.

Под водой – кораллы

Уже не первый год глобальное суперкомпьютерное сообщество пребывает в некотором нетерпении. Близкие к нулю изменения в верхней части списка TOP500 сочетаются с какими-то глубинными течениями, иногда довольно бурными.
Два крупнейших проекта Министерство энергетики США доверило триумвирату IBM–NVIDIA–Melanox. Как известно, IBM и так прекрасно чувствует себя в контексте списков TOP500. Новость в том, что будущие системы не будут использовать процессоры Intel и вообще архитектуру x86. Вычисления будут производить процессоры IBM POWER9, усиленные графическими ускорителями NVIDIA нового поколения – Volta, поддерживающими быстрые интерфейсы NVLink вместо привычных PCIe. Интерконнект обеспечит Mellanox.
Речь идет не просто о паре суперкомпьютеров в ряду других, а о двух системах, каждая из которых превысит по производительности 100 ПФлопс. Министерство выделило $300 млн на разработки и реализацию их в рамках программы Coral (Collaboration of Oak Ridge, Argonne, and Livermore – то есть охватывающей три важнейшие национальные лаборатории Министерства). Одну из них, в Окриджской лаборатории, назвали Summit, другую, в Ливерморской, – Sierra. Они должны быть запущены в 2017 году. В Аргоннской лаборатории еще не определились с архитектурой будущей системы.
Понятно, что такой разворот серьезнейшего заказчика в сторону архитектуры POWER может породить волны подражателей и проторить колею для будущих инсталляций. Но этим его значение не исчерпывается. Дело в том, что машины эти будут не просто мощные. Они позиционируются как data centric, то есть ориентированные на данные, а не на вычисления. Именно этим, как утверждается, был определен выбор. Мало того, что у них будет скорость передачи данных до 17 ПБ/с, главная цель таких архитектур – обработка возможно больших объемов данных непосредственно там, где данные расположены, вместо того чтобы идти по традиционному пути передачи данных в вычислительные узлы, где они обрабатываются, после чего возвращаются обратно в место хранения. Но насколько новые машины будут действительно data centric, еще предстоит понять.

Ребристые базы данных

Без упоминания Больших Данных сейчас вообще обходится мало какой большой проект, но в самом мире Big Data, ранее зачастую просто-напросто подразумевавшем Hadoop и MapReduce, начинают оформляться направления, названия которых будут на слуху уже в ближайшем, может быть, будущем. Например: то, что казалось частным случаем СУБД – графовые СУБД, – очень скоро вырастет из частного случая в нечто большее. В духе того, что происходило с реляционными СУБД, которые появились как частный случай, а стали почти синонимом СУБД. Во всяком случае, похоже, что каждая уважающая себя корпорация будет заводить у себя

Многообразие суперкомпьютерных миров и тест HPCG

Текст С.С. Андреев, С.А. Дбар, А.О. Лацис, Е.А. Плоткин,
Институт прикладной математики имени М.В. Келдыша РАН

Потенциал роста производительности суперкомпьютеров традиционной архитектуры практически исчерпан. Производительность процессорного ядра почти не растет уже более 10 лет. Безудержное наращивание количества процессорных ядер в суперкомпьютере приводит к целому ряду трудноразрешимых проблем, важнейшая (но не единственная) из которых – проблема энергопотребления. Поиск и освоение новых архитектур становится неприятной, но все более насущной необходимостью.
В последние 4–5 лет путь этого поиска представлялся непростым, но более или менее обозримым и понятным. Графические процессоры позволяют наращивать производительность и экономить энергию, но очень тяжелы в программировании. Спешить их осваивать не следует – на подходе Xeon Phi, который позволит решать все те же проблемы, но без сложности программирования, характерной для графических процессоров. Надо просто дождаться его появления...
Сегодня мы знаем, что графические процессоры не оправдали связанных с ними надежд во многом, а идущий им на смену долгожданный Xeon Phi – практически ни в чем. Чуда не произошло, волшебного способа ускорить работу программ, написанных полвека назад (или тем же способом, каким писали полвека назад), так и не появилось. Проблема создания и освоения новых архитектур снова стоит перед нами во всей своей полноте.
Утверждение, что эта проблема стоит именно перед нами, разработчи-

хотя бы одну графовую базу данных. Если графовые базы данных наводнят большой бизнес, изменят не только софт, но и сами машины. Вспомним, что когда-то важнейшей частью бизнеса IBM было направление AS/400 – машин, у которых была своя, довольно оригинальная операционная система. На AS/400 никто не стал бы считать форму крыла автомобиля. Там было все оптимизировано под работу с реляционной СУБД DB2, так как реляционная СУБД была (и, в общем, так и осталась пока) ключевой составляющей бизнес-ПО. А это значит, что надо было поддержать в железе быстрые и надежные транзакции, определенные типы шифрования. Если сейчас делать машину под работу с коммерческими графовыми БД огромных объемов, придется использовать даже другие интерконнекты: эффективное перемещение по ребрам графа требует совсем других способов передачи сообщений между вычислительными узлами.

Как будет по-русски IoT?

Если взглянуть на Большие Данные не с того края, где их обрабатывают, а с того, где их производят, то станут видны крупным планом такие ландшафты, как Интернет Вещей. Понятие вовсе не новое, но не слишком привычное: пока даже в профессиональных изданиях аббревиатуру IoT расшифровывают – для непонятливых читателей. В то же время никому не придет в голову расшифровывать «IT» в непрофильных изданиях, а «HPC» – в профильных. Между тем от IoT ждут больших новостей уже в 2015 году. С Интернетом Вещей, возможно, приключилось нечто вроде фальстарта. Еще во времена бурного подъема Java на каждом углу можно было слышать, что и лифт в офисе, и электрический чайник скоро получат свои IP-адреса. Чайники до сих пор прекрасно обходятся без IP, а в сознании массы IT-обывателей прочно поселилось представление, что Интернет Вещей будет состоять из чайников, общающихся с холодильниками. На самом деле будущее IoT связывают с Интернетом Вещей, делающих бизнес – в офисе и вне его. Маркетологи-аналитики тоже рассчитывают поживиться с трафика обменивающихся информацией устройств, участвующих в купле-продаже. А вот сокращение V2V во многих источниках даже не расшифровывают как понятное. Это обмен данными между движущимися средствами (Vehicle to Vehicle).

Возможно, очень скоро обмен телеметрической информацией между автомобилями станет в каких-то странах обязательным стандартом. Как бы там ни было, есть прогнозы, что в ближайшее время данные, связанные с IoT, могут достигнуть объемов зетабайтов, то есть порядка 10^{21}.

Микросхемы соберут себя сами

Но пока речь шла о «возвышенных» изменениях, не затрагивающих «низменные» уровни – попросту говоря, сам кремний или то, что должно прийти ему на смену. Архитектуры-то новые, а кремний старый. Технологические процессы – подновленные буквально дедовские методы.
Но выходит, что изменения грядут и там. Например, кремниевые транзисторы довольно скоро, возможно, будут собирать себя сами.
Речь идет о технологиях сополимеров. По сути это новые нанометоды, которые использовались для разных целей, с их помощью собирали структуры из наностержней, например. Их пробуют использовать и для наращивания плотности хранения дисков. Теперь самособираться будут критические элементы 3D-транзисторов.
До этого топология микросхемы задавалась фотолитографией: грубо говоря, на пластину наносится фоторезист, который разрушается светом. Светят сквозь фотошаблон, на котором и запечатлена вся топология будущего чипа. Вопрос в том, каким светом облучать, чтобы не попасться на физические ограничения, связанные с длиной волны и другими эффектами. Ну, или можно облучать не светом, а рентгеновским излучением или потоком электронов. Можно «подложить» под поток света прослоечку жидкости, которая за счет собственного коэффициента отражения соответственно улучшит разрешение (приблизительно так и делают, получая современные 14-нм чипы). Свет «портит» фоторезист, те части, которые были на свету, а не в тени, удаляются, те области, которые фоторезист перестал защищать, вытравливаются.
Чем же недавно удивили журналистов ученые из IBM? На традиционной кремниевой подложке традиционном методом вытравливаются определенной формы канавки. Они подобраны так, что осождающиеся на них блокирующие сополимеры вырастают в определенной формы объемные структуры, у которых характерные размеры могут оказаться

ками и пользователями суперкомпьютеров, а не перед флагманами компьютерной индустрии – производителями микропроцессоров массового выпуска, – в данном случае не является ни оговоркой, ни преувеличением, заставляющим заподозрить авторов этой статьи в мании величия. При сегодняшнем уровне технологий программируемой логики физическое изготовление сколь угодно нетрадиционных вычислителей собственной разработки в условиях небольшой исследовательской лаборатории – задача вполне разрешимая и даже почти рутинная. Конечно, оснащение вычислительных узлов даже небольшого экспериментального кластера модулями FPGA-сопроцессоров – универсальными «заготовками» для вновь создаваемых вычислителей – требует определенных финансовых затрат. Все же рискнем предположить, что в настоящее время тех, кто мог бы себе это позволить, на порядок больше, чем было, например, всех занимавшихся параллельными вычислениями в докластерную эпоху (примерно с 1990 по 2000 год).
Словом, изготовить вычислитель новой архитектуры может если и не каждый, то многие. Было бы что изготавливать. Проблема действительно стоит именно перед нами, и формулируется убийственно просто: кто и как мог бы заняться проектированием и испытанием новых вычислительных архитектур?
Вновь создаваемые архитектуры будут проблемно-ориентированными, а значит – многочисленными. Их придется оценивать и сравнивать между собой, а для этого нужен измерительный инструмент – тест (или система тестов) производительности, являющийся в каком-то смысле общеизвестным и общепринятым. Требования к такой системе тестирования предъявляются довольно специфические. Во-первых, она должна «сравнивать несравнимое» – сопоставлять разные архитектуры, каждая из которых, скорее всего, хорошо справляется с одними задачами, и гораздо хуже – с другими. При этом производительность должна демонстрироваться вместе с инженерной методикой ее достижения: реализация теста должна быть примером и руководством к действию, а не просто констатировать, что производительность достигнута. Во-вторых, она должна быть руководством по налаживанию межцехового диалога между системными инженерами, системными программистами и прикладными программистами.
В самом деле, хорошие проблемно-ориентированные архитектуры не могут быть созданы ни силами одних системных инженеров (инженеры не знают методов), ни силами одних прикладных программистов (прикладные программисты не знают железа).
Словом, тест должен олицетворять собой конструктивный, систематический и междисциплинарный подход к сопоставлению архитектур.

Новый тест производительности HPCG совершенно не случайно появился в это очень интересное и очень тяжелое для суперкомпьютерной отрасли время. Отвечает ли он поставленным требованиям? С одной стороны, сам по себе если и отвечает, то далеко не полностью. С другой стороны, как сам тест, так и приводимая его авторами мотивация дают нам много ценных идей и для построения более полезной системы тестов, и для налаживания на ее основе межцехового диалога.
Чтобы понять весь драматизм интриги, связанной с появлением HPCG, напомним некоторые основные ее моменты, скорее всего, известные многим читателям.
Старый тест производительности HPL был предложен в то время, когда только появились процессоры с заметными объемами кэш-памяти, впервые позволившими значительно ускорять вычисления за счет локализации обработки на малых объемах данных. Требовалось подчеркнуть исключительную важность такой локализации, и был предложен тест, позволявший это сделать. Причем сделать настолько хорошо, что никакая реальная программа по возможности мелкоблочной локализации с предложенным искусственным тестом и близко сравниться не могла. За прошедшие с тех пор 20 с лишним лет усовершенствование процессоров общего назначения оставило системным архитекторам единственный резерв архитектурного повышения производительности – ту самую локализацию (все остальные резервы давно вычерпаны). Теперь никому не требуется объяснять, как важна локализация – задача в том, чтобы, с одной стороны, действительно обеспечить ее на уровне метода, с другой – «достойно вознаградить» за это выросшим быстродействием на уровне архитектуры. Старый тест с его искусственными, неправдоподобно хорошими возможностями локализации обработки из стимула прогресса стал его тормозом – теперь он провоцирует создание архитектур, способных эффективно выполнять, грубо говоря, сам этот тест, и ничего больше.
Ошибку требовалось исправить, и она была исправлена: новый, основанный на реальном численном методе тест HPCG вообще не позволяет локализовать обработку данных в мелких блоках. И это – в то время, когда никаким другим путем ускориться уже в принципе нельзя! Виток диалектической спирали пройден и завел нас в тупик.
Зачем же нам такой тест? Неужели только для демонстрации любимого некоторыми математиками (и, кстати, не вполне верного) утверждения, гласящего, что «распараллеливаются только самые плохие методы»?
Присмотримся подробнее не к букве, но к духу нового теста. В отличие от старого, он является «действующей моделью» реально интересного с практической точки зрения численного метода. Авторы теста предлагают для определения производительности хрономе-

меньше типичных для литографии. Два полимера взаимодействуют, выстраивают нечто совместное, не давая излишне разрастаться. Потом ученые уничтожают обоих – печальный удел. Но в одном из них содержится как составная часть кремний, а в другом – нет. Кремний осаждается в нужных местах как память об объемных структурах из полимеров. Ожидается, что они заменят традиционные технологии в наиболее критичных местах – там, где «плавники» современных 3D-транзисторов, например.

Оптики обещают цифры

Прежде чем вернуться к той части пейзажа, где в начале статьи был помещен вертолет, хочется обратиться опять к Big Data: в любом случае в будущем без них не удастся ступить и шага.

Одна из проблем в мире Больших Данных – то, что основная часть технологий Hadoop, прежде всего, не родились естественным образом в недрах HPC-сообщества, не были разработаны учеными для ученых. Алгоритмы и механизмы, которые там используются, были выбраны из соображений каких угодно, но только не максимально возможной быстроты работы и использования последних достижений параллельного программирования и быстрых топологий интерконнектов. Но ведь можно направить силы на то, чтобы переписать наиболее критичные куски кода Hadoop под существующий или будущий уровень HPC-систем. Это будет делаться и уже делается, но возникает вопрос, пока риторический (и с отрицательным ответом): не лучше ли сразу направить усилия на принципиально другие разработки, способные справляться с Большими Данными?

Например, взять в помощь оптические вычислители. Британская компания Optalysys объявила, что прототипы оптических процессоров будут готовы через считанные месяцы, а коммерческие продукты появятся в 2017 году, то есть примерно в одно время с двумя суперкомпьютерами на POWER9, о которых писали в начале статьи.

Всего через несколько лет, утверждают в компании, экзафлопсной производительности можно будет достигнуть на устройствах размером с ПК.

В начале 2015 года должен быть готов прототип с производительностью 340 ГФлопс, показывающий работоспособность концепции. Вычислительный элемент этих оптических процессоров – жидкие кристаллы.

Электрооптический эффект позволяет вводить данные в кристалл, а соответствующие изменения оптических свойств дают возможность получать результат на выходе со скоростью прохождения света через кристаллы. Дифракционные свойства и Фурье-оптика позволит использовать такие вычислители для решения задач с преобразованиями Фурье, гидродинамики и распознавания образов. А вот и Big Data: в 2017 году появится два вида устройств. Первое из них будет приспособлено для задач анализа Больших Данных и работать в паре с традиционным суперкомпьютером. Первый рабочий образец должен будет развить производительность 1.32 ПФлопс. К 2020 году она будет доведена до 300 ПФлопс. Заявка относительно скромная на фоне того, что в 2019–2020 годах уже ожидаются традиционные суперкомпьютеры экзафлопсной производительности. Зато вариант для решения уравнений и моделирования – Optalysys Optical Solver Supercomputer – начнет сразу с 9 ПФлопс, а в 2020-м достигнет уже 17.1 Экзафлопс.

Надо сказать, что разработчики квантовых компьютеров пока не рискуют прогнозировать столь подробно. Что свидетельствует о нескромности оптиков, мягко говоря. Или, с другой стороны глядя, о то ли показной, то ли органичной уверенности в себе и в перспективности своих разработок.

В пейзаже, который мы попытались нарисовать, выбор объектов может показаться случайным.

Многое не попало в поле нашего зрения: новейший проект по сверхпроводящим суперкомпьютерам, целая россыпь разработок искусственного интеллекта с разных сторон, подступающих к естественному, да мало ли что?

Но это уже другая история, все они подождут до следующего раза, если стремительно возникающие в этом пейзаже фигуры не заслонят их – на время, конечно, – от читателя.

трировать не метод, а алгоритм (конкретную реализацию метода), причем измерять предлагается время выполнения фиксированного числа итераций. С точки зрения конструктивности (использования теста как руководства к действию) интереснее было бы попробовать модифицировать алгоритм таким образом, чтобы он лучше сходился для тяжелых, сильно заполненных и плохо обусловленных матриц и при этом допускал эффективную реализацию на новых архитектурах. Измерять при этом следует не время выполнения фиксированного числа итераций, а время (или затраты энергии) получения сходимости с заданной относительной погрешностью. Такая метрика для результатов тестирования свидетельствовала бы о том, что способ модификации алгоритма выбран правильный.

Рамки настоящей статьи не позволяют рассказать о том, как именно нам удалось построить такой модифицированный вариант метода HPCG. Скажем лишь, что нам это удалось. Потребовались многочисленные вычислительные эксперименты и, главное, действительно глубокая перестройка алгоритма, не сводимая к формально тождественным преобразованиям программного кода. В результате был получен пример инженерной методики, показывающий, какого рода работу приходится выполнять для взаимной подгонки метода и создаваемого для его эффективной реализации оборудования. Излишне было бы говорить о том, что львиная доля работы пришлась не на проектирование сопроцессора в FPGA, а на работу с методом и в терминах метода, а также на вычислительный эксперимент.

Метод, придуманный как остроумный способ «поставить на место» зазнавшихся любителей бить рекорды в реализации HPL, был не менее остроумным способом модифицирован, с некоторыми шансами на успех. Что нам это дает? Здесь мы подходим к следующему требованию к системе тестов для новых архитектур, а именно – к собственно систематичности. Какое место, хотя бы приблизительно, занимает метод HPCG в суперкомпьютерной вселенной, в общей картине мира? Актуален ли он, или стоит особняком и применяется редко?

Забегая вперед, отметим, что место это – более чем почетное, и это легко показать.

Роль задач математической физики, в частности – механики сплошной среды, среди всего множества задач, решаемых на суперкомпьютерах, очевидна и в особом представлении не нуждается. В разностном приближении эти задачи превращаются в системы линейных алгебраических уравнений, причем с матрицами очень большого размера, очень разреженными и весьма специального вида. Итерационным методам решения таких СЛАУ посвящена специальная область численных методов – вычислительная линейная алгебра. Метод HPCG – один из типичных, хорошо известных методов вычислительной линейной алгебры. Но всего таких методов, с учетом вариантов и модификаций, десятки. Что мы узнаем об их реализуемости на новых архитектурах, если даже и справимся с HPCG?

Несмотря на исключительное разнообразие и многочисленность, методы вычислительной линейной алгебры удивительно просто и понятно классифицируются по технике параллельной реализации. С некоторыми оговорками, методы эти распадаются, во-первых, по способу представления матрицы СЛАУ (собственно матрица или разностный шаблон), а во-вторых – по способу сбора и распространения информации об изменениях решения в ходе итераций (методы типа простой итерации или крыловские). Классификация по двум независимым признакам дает нам четыре класса, и ни один из них не пуст. Принцип классификации прост и равно понятен системному инженеру, системному и прикладному программисту. Поскольку классы выделены именно по способу параллельной реализации, тестов в искомом тестовом наборе должно быть всего четыре – по числу классов, причем в каждом классе можно взять самый простой численный метод.

Из самых общих соображений ясно, что с матрицей общего вида, скорее всего, работать труднее, чем с разностным шаблоном, а с методом, на каждой итерации собирающим информацию со всего решения и использующим ее в каждой точке – труднее, чем с «локальным» методом типа простой итерации. Таким образом, если мы совладаем с крыловскими методами для матриц общего вида, то практически вся вычислительная линейная алгебра у нас в руках, а ведь это – солидная часть механики сплошной среды и не только.

Самым простым представителем этого самого сложного класса методов вычислительной линейной алгебры как раз и является HPCG – номер 4 в искомом тестовом наборе.

Из сказанного уже можно сделать некоторые выводы.

Задача создания и освоения новых проблемно-ориентированных архитектур беспрецедентна как по актуальности, так и по сложности. Ее решение объективно требует высокого уровня межцехового взаимодействия инженеров, системных и прикладных программистов. Центральным элементом системы такого взаимодействия могла бы быть система конструктивных тестов производительности. Мы показали, что такая система легко получается из основных методов вычислительной линейной алгебры, поскольку последние очень легко и понятно классифицируются в отношении техники параллельной реализации. Наиболее сложные из этих методов не могут быть адаптированы к новым архитектурам путем формально эквивалентного преобразования кода и поэтому должны глубоко перерабатываться на смысловом уровне. Оценка оправданности и корректности такого рода преобразований могла бы учитываться при тестировании, если в условиях тестирования заменить измерение времени фиксированного числа итераций на измерение времени получения сходимости с заданной точностью.

NESUS в массы

Хесус Карретеро возглавляет проект NESUS – Network for Sustainable Ultrascale Computing Systems (Сети Устойчивых Вычислительных Систем Экзамасштабов). NESUS входит в число так называемых Акций COST – (European) Cooperation in Science and Technology – Кооперацию по Науке и Технике Евросоюза. Хесус преподает в Мадридском университете Карлоса III.

Игорь Лёвшин: Какие задачи стоят перед NESUS? Что, на ваш взгляд, было бы лучшим результатом проекта?

Хесус Карретеро: Цель акции NESUS – поддержать исследования устойчивости вычислительных систем сверхбольших масштабов (ULTRASCALE COMPUTING SYSTEMS – UCS). Речь идет об огромных и сложных системах, в которых сочетаются параллельные и распределенные вычисления. Они дадут возможность доступа к вычислительным ресурсам беспрецедентному количеству пользователей. А при такой сложности и при таких масштабах, как в UCS, устойчивость систем станет важнейшей проблемой.

В данный момент цель NESUS – исследования вопросов устойчивости, связанных в основном с энергоэффективностью, но вообще устойчивость связана и с решением проблем программируемости, управления данными, отказоустойчивости, энергоэффективности и масштабируемости. Чтобы добиться устойчивости, надо понимать, как влияют все эти факторы на устойчивость всей экосистемы, а не отдельных компонентов. Наша задача – скоординировать совместные европейские разработки, направленные на поиск реалистичного решения проблемы устойчивости UCS. Мы стараемся объединить различные сообщества разработчиков, чтобы они вместе исследовали стек системного ПО, в который входят парадигмы программирования, среда выполнения, ПО промежуточного слоя, отказоустойчивость, управление данными, модели энергоэффективного управления и их собственные приложения, чтобы улучшить устойчивость UCS, а также возможности переписывания и программирования заново приложений, которые будут эффективно использовать устойчиво работающие UCS. Результатом может стать и сама по себе кооперация, и, конечно, сами по себе исследования. Для кооперации NESUS создаст мультидисциплинарный форум, чтобы представители различных сообществ разработчиков смогли обмениваться идеями. Такой форум мог бы стать базой, точкой отсчета для всего европейского мира устойчивых UCS – я имею в виду и исследования, и образование, и промышленные применения, и участие в инвестировании. По части исследований Акция поддержит и оригинальные инициативы, создаст критическую массу исследователей. Говоря более конкретно, Акция предложит каталог приложений и инструментов для устойчивых UCS. Будет создана дорожная карта, синхронизирующая европейские усилия в этом направлении.

И. Л.: Почему в проекте участвует так много стран? Правда ли, что NESUS – самый большой по числу стран-участников проект в Европе из тех, что финансируется Советом Европы? Какова организационная структура проекта, обеспечивающая взаимодействие такого количества стран и налаживающая их эффективную совместную работу?

Х. К.: В предварительном варианте проекта были 19 стран. На собрании по поводу старта проекта были представители уже 29 стран. Сейчас, после 9 месяцев работы у нас 45 стран, включая 33 страны, входящие в COST, 6 ассоциированных членов (включая Россию) и 6 партнеров из США, Австралии, Колумбии, Канады и Мексики. Вы правы, это самый широкий проект, финансируемый Советом Европы.

Организационная структура работает довольно эффективно. У нас есть председатель правления, заместитель председателя, председатели по главным направлениям и лидеры рабочих групп (их шесть). Все они образуют Ядро Акции, которое координирует повседневную работу. Стратегические решения и направления деятельности, бюджет утверждаются Управляющим Комитетом, в который входят по два представителя от каждой страны.

Мы используем прежде всего электронные коммуникации. Портал **nesus.eu** – это инструмент и для распространения (в Интернете), и для организации (наш Интранет). Но самое важное – это энтузиазм самих людей и их желание работать вместе. В рабочих группах они занимаются следующими темами:

- *Систематическое обучение и последние достижения в UCS.*
- *Модели программирования и среды времени выполнения.*
- *Отказоустойчивость приложений и окружение среды выполнения.*
- *Устойчивое управление данными.*
- *Энергоэффективность.*
- *Приложения.*

Группа «Научные задачи ближайшего времени» (Short Term Scientific Missions – STSM) обеспечивает молодым ученым возможность работать в других странах, обычно как следствие научной кооперации.

Другое направление – Школы. Они проводятся в течение одной недели и представляют собой комбинацию теории и практики. Докладчики Школы – ключевые фигуры Акции и представители промышленности. Комитет Акции определяет, какие темы проходить студентам в течение года.

Еще существует Аспирантский симпозиум. Он должен помочь молодым ученым и соискателям степени «доктор философии» представить свои идеи на самом раннем этапе. Им помогают советами члены Верхнего комитета Акции, которые сменяются каждый год.

Проводятся и семинары Акции. На них представляются в виде открытых презентаций и книги материалов все работы, выполненные в рамках Акции за год.

Также устраиваются и мероприятия по распространению идей Акции – совместные публикации, презентации и конференции, новостные ленты, веб-портал, публикации не только в специализированной, но и в общей прессе.

Осуществляется и совместная деятельность по кооперации с промышленностью, комитетами по стандартам, ассоциациям и проектам Евросоюза. Мы проводим семинары для более тесного сотрудничества с нашими партнерами из промышленности и продвигаем открытие новых возможностей в промышленности для аспирантов и сотрудников Акции.

Этой деятельностью мы не ограничиваемся – мы открыты для других проектов, если под них удается получить бюджет, как записано в документах COST. Это относится, например, к образовательным программам для старшекурсников.

И. Л.: Нельзя ли подробней рассказать об образовательных проектах NESUS?

Х. К.: Будучи акцией COST, NESUS больше ориентирован на научную деятельность. Тем не менее у нас есть и образовательные инициативы – ежегодная Школа. Обучение может быть организовано как деятельность внутри рамок Акции, а может и как кооперация с Акцией. Мы, например, можем рассматривать проведение курсов, предлагая IT-компаниям помогать аспирантам получать ученую степень.

Главное ограничение – бюджет: для Акции такого огромного масштаба он не так уж велик. Вообще, для всей деятельности NESUS характерны приоритеты для молодых: мы делаем акцент на гендерном равноправии и на ученых степенях для молодых исследователей. Школы, симпозиумы для аспирантов и STSM, о которых я говорил, придуманы специально для них.

Для старшекурсников у нас нет какой-либо ориентированной специфически на них деятельности, поскольку NESUS – сообщество исследователей, а студенты не могут по уставу быть членами Акции.

И. Л.: Существуют ли связи между NESUS и ETP4HPC (European Technological Platform for HPC)? Между NESUS и большими международными инициативами – EESI (European Exascale Software Initiative) и BDEC (Big Data and Extreme Computing)?

Х. К.: Сейчас мы как раз налаживаем связи с другими инициативами Евросоюза. У нас есть контакты с ETP4HPC, мы обмениваемся с ними информацией. Жан-Франсуа Лавиньон, президент ETP4HPC, присутствовал на нашем последнем собрании в Париже, и, надеюсь, теперь наши связи станут более прочными. Что касается EESI, мы пока в фазе первых контактов, но, разумеется, мы будем работать с ними. Некоторые члены NESUS принадлежат к обоим сообществам, так что связи – вообще не проблема.

И. Л.: В мире два главных суперкомпьютерных события (конференции + выставки): SC – Supercomputing Conference в ноябре в США и ISC – International Supercomputing Conference в июле, в Германии. Планирует ли столь крупный международный проект, как NESUS, участие в них в какой-либо форме?

Х. К.: В 2014 году NESUS проводил свои презентации на CCGRID, EUROPAR, CloudNet, EuroMPI/Asia и SBAC-PAD.

Мы уже участвовали в SC в ноябре этого года в США. У нас был плакат и презентация на стенде INRIA. Презентацию проводила профессор Анна Элстер из Норвежского университета науки и техники (NTNU). А в 2015 году мы запланировали презентацию на ISC.

Apollo на разных широтах

Текст Игорь Лёвшин

Системы Apollo 6000 и Apollo 8000 были представлены в июне этого года компанией HP с изрядной торжественностью – и в Лас-Вегасе, и на ISC 2014 в Лейпциге. Показывали космические ролики, сдергивали ткань, произнесли речи. Было ясно, что службы маркетинга компании придают событию большое значение.

Нельзя сказать, что Apollo 6000 и Apollo 8000 внимание HPC-сообщества поделили поровну. Старшая система – в фокусе прожекторов, вокруг младшей ажиотажа меньше. И это неудивительно, ведь они настолько разные, что многим даже кажется странным похожесть названия (кстати, «старожилы» IT вспоминали не только космические корабли, но и рабочие станции Apollo, популярные в 80-е). Скромник Apollo 6000 работает на воздушном охлаждении, герой дня Apollo 8000 охлаждается теплой водой.

Эти машины интересны своими конструктивными особенностями, но они интересны еще и тем, где установлены первые работающие системы или где заработают будущие. И еще они замечательны некоторыми необычными функциями. Они, например, отапливают помещения. Можно заподозрить, что в этих словах спрятана ирония, но ее в них нет. Это реальность, приятная, например, для жителей университетского городка заполярного города Тромсё.

Появление у HP систем с водяным охлаждением само по себе, конечно, не удивляет. У IBM есть Blue Gene/Q, у отечественных производителей тоже есть охлаждаемые жидкостью системы: у компаний РСК и «Т-Платформы». У IMMERS вычислители вообще полностью погружены в жидкость. Вопрос был в том, какой тип водяного охлаждения выберет HP.

Выбрано было нестандартное решение: тепло отводится двумя контурами разного уровня. Сверху донизу стойку пронизывает «водяная стена», которую также называют «водонапорной башней». Стена отводит тепло от лезвий, которые соприкасаются со стеной. Внутри самих лезвий вода отсутствует. Тепло от нагревающихся компонентов, запечатанных медными тепловыми трубками, передаётся «водяной стене» через тепловую шину. Внутри трубок давление понижено (субатмосферное), чтобы даже в случае повреждения самой трубки (о котором тут же оповестит система) исключить вытекание жидкости из трубки. Это уникальная особенность позволяет заменять отдельные лезвия так, как если бы они охлаждались не водой, а воздухом – все соединения «сухие». Такой возможности пока больше не предлагает никто. Та часть лезвия, к которой подведены трубки, прижимается рычажком к полированной – для лучшего теплового контакта – поверхности «стены». Таким образом охлаждаются процессоры, память и модули с ускорителями Xeon Phi или GPU NVIDIA, если они есть в конфигурации. Коммуникационные устройства охлаждаются воздухом. В Apollo 8000 на стойку приходится 8 коммутаторов InfiniBand.

Управляющий блок внутри стойки контролирует температуры по всему охлаждающему тракту. Вода, охлаждающая процессоры, имеет температуру до 30°C.

В стойке 288 процессоров Intel последнего поколения (Haswell) (72 двухузловых лезвия). К ней подводится кабель 480 В при 80 КВт мощности.

Кроме вычислительных стоек имеется инфраструктурная стойка. Половину ее занимает блок системы охлаждения (CDU), который обслуживает «водяные стены» вычислительных стоек. Кстати, вода, которая циркулирует в системе, практически из бытового водопровода (должна удовлетворять не слишком требовательным стандартам ASHRAE). CDU обслуживает до 4 стоек, позволяя отводить в сумме 320 КВт тепла.

Одна из целей, которую ставили перед собой инженеры HP, разрабатывая Apollo, – это возможность инсталлировать системы даже петафлопсной производительности в считанные дни, а не недели и месяцы. Для этого было упрощено и оптимизировано все, что связано с монтажом охлаждающего оборудования. Все шланги и соединения оснащены зажимами, которые способен подключить даже любитель. Все заранее собрано и протестировано. И, по свидетельствам представителей Арктического Университета (Тромсё), монтажники действительно уложились в 3 дня (одна вычислительная и одна инфраструктурная стойка).

Системы Apollo 6000 тоже годятся для инсталляций петафлопсного уровня, но в целом они особенно хорошо подходят для обычных корпоративных систем, вплоть до уровня лаборатории или небольшой рабочей группы. В стойке можно разместить до 80 серверов (узлов). Но конструктивный элемент здесь шасси HP Apollo a6000 в формфакторе 5U, в которое укладывается 10 серверов. Шасси имеет общую для своих серверов систему воздушного охлаждения. Шасси Apollo a6000 может обслуживать до десяти серверных модулей. Система Apollo 6000 может быть укомплектована серверами HP ProLiant XL220a Gen8 v2 или верхней моделью линейки – HP ProLiant XL230a Gen9. Объявлена и поддержка будущих серверов серий XL240 и XL250. Таким образом, будет полный ряд серверов для построения в том числе и гибридных систем с графическими ускорителями и сопроцессорами. Используемый в шасси сетевой модуль поддерживает различные конфигурации FlexibleLOM – как гигабитные, так и 10-гигабитные порты.

Как известно, в суперкомпьютере Tsubame 2.5 работают серверы HP ProLiant SL390s G7. Со своими без малого 3 ПФлопс он сейчас отодвинут на 15-е место списка TOP500, но Tsubame, безусловно, в высшей лиге. Маркетологи HP старательно подчеркивают, что Apollo прекрасно подойдут для построения сверхмощных суперкомпьютеров для научных институций и национальных лабораторий. Но ими целевая аудитория отнюдь не исчерпывается. Для инженеров или финансистов подойдут не только Apollo 6000, которые компактны (HP утверждает, что их вычислительная плотность на 60% больше, чем у типичных стоек блейд-серверов с воздушным охлаждением), но и более дорогие и мощные Apollo 8000. Одно из важных преимуществ, которое ценят на корпоративном рынке, – это малый уровень шума, вообще свойственный системам с жидкостным охлаждением.

Говоря об инсталляциях своих новых систем, HP чаще и охотнее всего упоминает NREL – Национальный Центр Возобновляемой Энергетики. Еще бы – разработки Apollo были связаны с заказом этого центра. HP и сам NREL сочли решение очень удачным, и HP решила его тиражировать. Многие узлы системы Apollo 8000, особенно те, что связаны с охлаждением, разрабатывались совместно с NREL или соответствовали сформулированным ими требованиям. Центр хотел иметь у себя систему, соответствующую духу и букве организации: расходуемая энергия не должна пропадать. В результате этого сотрудничества родилась система, которая не только вобрала в себя самые современные технологии, но и учла опыт эксплуатации и практику технического сопровождения крупной высокопроизводительной вычислительной установки.

В результате система, которая называется Peregrine («Сапсан»), не ограничивается своими прямыми функциями, а еще и отапливает офисный корпус лаборатории. Начальство центра собирается экономить по $1 млн в год на сбереженной электроэнергии. NREL находится в штате Колорадо, гористом и ветреном, где отопление нелишне. Сейчас пиковая производительность системы уже больше петафлопса, и она может быть увеличена почти в два раза. 18 стоек уместились приблизительно на 100 кв. м. PUE системы – около 1.06.

Возможно, еще более интересный случай для нас – система Apollo 8000, установленная в Арктическом Университете Тромсё в Норвегии, в 300 с лишним километрах севернее Полярного круга. В этом городе средняя температура в июле 12°C, так что выбор их ясен: машина отапливает университетский кампус. «Мы хотим научиться использовать теплую воду от системы круглый год, – говорят представители университета. – Сейчас наша задача, чтобы у нашего нового ВЦ общей мощностью 2 МВт использовалось 80% тепла. В этом случае мы сэкономим огромные суммы денег, и это, в свою очередь, даст нам возможность перейти в петафлопсную эру, нарастив наши скромные 200 ТФлопс». Мне кажется, было бы приятно подумать о подобной системе где-нибудь в Мурманске, Архангельске или Норильске.

Сразу две системы Apollo 8000 должны появиться в Польше. Россия периодически вырывается вперед перед Польшей в списке TOP500, а Польша периодически приближается к России или даже догоняет нас по числу инсталляций. Но в Польше есть интересная и полезная, как мне кажется, особенность суперкомпьютинга: децентрализация. Крупные системы более или менее равномерно распределены по основным 4–5 университетским городам. Сейчас в городе Гданьске, в местном Технологическом университете появится система HP Apollo 8000, за 1.2 ПФлопс которой университету предстоит заплатить $9 млн. В проекте участвуют компании Megatel и Action. Гданьский суперкомпьютер должен попасть в первые 50 систем из списка TOP500. В 30 стойках общим весом 15 тонн будет 2600 процессоров.

На противоположном конце Польши, на юге, в Кракове, будет установлен Apollo 8000 с пиковой производительностью 1.7 ПФлопс. Он будет называться Prometheus и в 4 раза превзойдет по мощности своего предшественника – «Зевса» (Zeus). Владелец его – Cyfronet AGH. В системе будет 1728 серверов HP Apollo 8000 XL730fGen9 – около 41000 процессоров Intel E5-2680v3 (Haswell-EP), память 215 ТБDDR4 и сеть Mellanox FDR InfiniBand. К ним будут подсоединены файловые системы объемом 10 ПБ и пропускной способностью 150 ГБ/с. Система будет установлена в созданном под нее ВЦ в Cyfronet и отдана в коллективное пользование ученых-физиков, химиков, биологов, нанотехнологов и энергетиков.

Кроме того, «Прометей» войдет в польский грид PLGrid, в грид ЦЕРНа (проект LHC) и в грид геофизиков EPOS.

Нет сомнения, что география Apollo 8000 в ближайшее время будет расширяться.

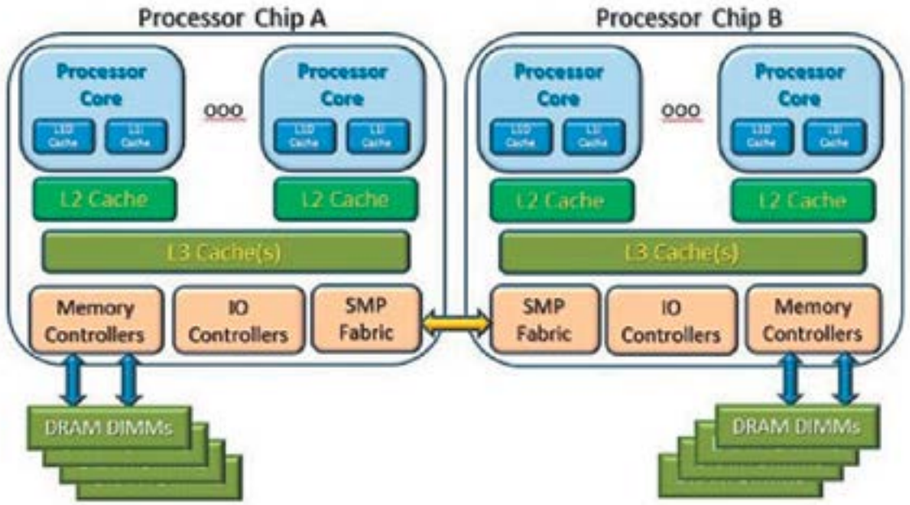
Кластеры для Больших Данных

Текст Леонид Черняк

Менее чем за 20 лет кластеры Beowolf прошли, казалось бы, немыслимый путь от простейших сборок на деревянных стеллажах до суперкомпьютеров, занимающих верхние позиции в TOP 500, сохранив при этом свой врожденный дефект – узел такого кластера может работать только с собственной памятью, из чего следует – большие массивы таким кластерам недоступны. Тем не менее идея сборки больших машин из массовых компонентов остается привлекательной, что дает неплохой шанс новому классу компьютеров – кластерам с общей памятью, называемым SMC (Shared Memory Clusters).

Появление SMC является закономерным следствием эволюции от первых многопроцессорных систем до современных MPP. Родина SMC – стартап Numascale, созданный при Университете города Осло, сохранивший и приумноживший традиции известного в свое время производителя миникомпьютеров Norsk Data. Основным достижением Numascale стал коммутатор NumaConnect, вместе с ним она сделала еще один шаг в развитии архитектуры NUMA.

Распределенный доступ к общей памяти называют симметричным мультипроцессированием (Symmetric Multiprocessing, SMP), решения класса SMP можно разделить на два типа – UMA и NUMA. UMA (Uniform Memory Access) обеспечивает всем процессорам равные права, здесь время доступа к памяти одинаково для всех. UMA был реализован в 1962 году в Burroughs D825, это был первый в мире многопроцессорный компьютер, позже UMA применили в старших моделях System/360 и VAX, сегодня UMA можно найти в современных системах, обычно с числом процессоров не более 8, а также в серверах малой и средней производительности, рабочих станциях и даже в мобильных устройствах. Дальнейшему расширению UMA препятствует ограниченный потенциал масштабирования, в этой схеме есть непреодолимый барьер в виде бутылочного горла между процессорами и памятью. Поэтому в качестве выхода из сложившегося положения была предложена схема NUMA (Non-Uniform Memory Access), родоначальником этого направления в 1990-е годы стала та же компания Burroughs, позже NUMA развивали практически все основные участники рынка. От UMA эта схема отличается поддержкой разделения всей памяти на отдельные банки, физически распределенные между процессорами, в NUMA процессор в первую очередь получает доступ к собственной памяти и только во вторую – ко всем остальным



Архитектура кластера с разделяемой памятью:
ProcessorCore – ядро процессора; L1 Cache–кэш L1; L2 Cache–кэш L2; L3 Cache– Кэш L3; ooo– …; MemoryControllers – контролеры памяти; IO Controllers – Контролеры ввода/вывода; SMP Fabric–контролер SMP; DRAM DIMMs–память DRAM

банкам. Последующему развитию NUMA помешало то, что современные процессоры работают существенно быстрее, чем память и, чтобы как-то согласовать скорости, используют кэш-память нескольких уровней (L1, L3, L3), появление кэша затрудняет использование NUMA в чистом виде. В кэше находится временная копия данных из основной памяти, и, если процессор – владелец кэша внес в него изменения, то он должен каким-то образом информировать об этом другие процессоры, использующие те же данные, посредством соответствующего механизма. Этот процесс называют когеренцией, ccNUMA (cache coherent Non-Uniform Memory Architecture) есть ни что иное, как вариация на тему NUMA, реализующая этот механизм.

В современных условиях возникает очередная дилемма: SMP-компьютеры прекрасно справляются с большими массивами, но при том они чрезвычайно дороги, а кластеры во много раз дешевле, но не приспособлены к актуальным задачам работы с Большими Данными. Как же совместить ценовые преимущества кластеров с функциональными достоинствами SMP? На этот вопрос есть два ответа: обратная виртуализация и кластеры с разделяемой памятью.

Можно собрать из множества машин, входящих в кластер, одну виртуальную SMP-машину– такое действие называют «виртуализацией для сборки» (Virtualization for aggregation). Агрегация открывает возможность для создания высокопроизводительных конфигураций из дешевых серверов, которые при переносе в облака будут в принципе способны решать задачи, требующие больших объемов памяти. При всей своей внешней привлекательности у технологий агрегации посредством виртуализации есть существенный и непреодолимый недостаток – низкая скорость работы с виртуально разделяемой памятью и, как следствие, значительные задержки. То и другое с неизбежностью следует из сохранения традиционных для кластеров систем интерконнекта.

Более эффективному решению дилеммы способен помочь такой аппаратный или программно-аппаратный подход, который бы позволил узлам кластера напрямую работать с памятью, находящейся в других узлах, тем самым обеспечив создание кластеров с общей памятью – SMC, о которых шла речь выше. Таким образом, предлагается новый этап в развитии существующих методов прямого доступа в память DMA (Direct Memory Access). Чаще всего DMA применяют для ускорения подсистем ввода-вывода с 1960-х годов, из них наибольшую известность приобрели шины ISA и PCI, а позже DMA стали использовать для подключения процессора к памяти сопроцессоров - в IBM Cell, DirectGMA GPU в AMD и NVLink в новых GPU NVIDIA. NUMA и ccNUMA тоже можно представить как вариации на тему DMA, позволяющие множеству процессоров работать с общей памятью. SMC дает возможность каждому процессору получать доступ к памяти в любом узле кластера, в этом смысле SMC служит продолжением NUMA, но реализацию DMA в кластерном исполнении справедливо будет назвать не прямой, а многоуровневой.

Технология NumaConnect восходит своими корнями к Scalable Coherent Interconnect (SCI) – стандарту, задуманному в качестве замены господствовавшей в начале 1990-х шинной архитектуры в многопроцессорных системах, который к тому же он включал поддержку удаленного прямого доступа к памяти (Remote Direct Memory Access – RDMA). В своих старших моделях SCI использовали компании Convex, DataGeneral, Sequent, Cray и SunMicrosystems. В силу разных причин стандарт оставался нишевым. Среди немногих сохранивших верность SCI была и остается норвежская компания Dolphin Interconnect Solutions. Несколько лет назад от нее отпочковалась Numascale с целью объединения имеющихся наработок в части SCI с известной технологией HyperTransport.

NumaConnect родственник межпроцессорным коммуникациям, что отличает его от сетевого стандарта InfiniBand, появившегося в результате слияния двух разработок: NextGeneration I/O (NGIO) и Future I/O (FIO). NGIO разрабатывался в Intel, имея в виду две цели: во-первых, построить последовательный интерфейс, отделенный от связки «процессор – оперативная память», во-вторых – обеспечить посредством этого интерфейса взаимодействие процессов, принадлежащих приложениям, работающим на разных серверах. Параллельно консорциум IBM, Compaq, Adaptec, 3Com, Cisco и Hewlett-Packard разрабатывал аналогичный стандарт FIO. Если NGIO задумывался для стандартных серверов, то FIO предназначался для платформ корпоративного класса. В начале 1999 года обе спецификации были обнародованы, а уже осенью произошло слияние.

Разная исходная позиция привела к существенным различиям между двумя интерфейсами. В данном контексте существенно то, что в InfiniBand реализован пакетный режим с широкой полосой пропускания, а NumaConnect был задуман именно для работы с разделяемой памятью.

Основным продуктом NumaConnect является микросхема NumaChip, изготавливаемая IBM, – в ней материализован стандарт SCI. Она сочетает в себе логику, осуществляющую управление когерентным кэшем распределенной памяти, с 7-портовым коммутатором. NumaChip подключается к Opteron (cHT) по HTX и в соответствии с SCI поддерживает адресацию к 4096 узлам и 256 Тбайт памяти на узел.

В каждом из узлов кластера устанавливается по одному NumaChip, соединение между ними выполняется по одной из двух возможных топологий – по двумерному или трехмерному тору. Первая топология используется в системах малого и среднего размера, до сотен узлов, в больших же системах предпочтительнее вторая. При такой архитектуре не требуется никакого специального централизованного управления, тем самым обеспечивается высокая надежность. К числу достоинств систем, построенных на базе NumaChip, относится отсутствие каких-либо специальных требований к системному ПО – они могут работать под управлением любой ОС, поддерживающей SMP– например, Linux, Solaris и Windows Server. В данный момент Numascale поставляет программу первоначальной загрузки, служащую для инициализации системы, – этот загрузчик прошел тестирование на Linux. После ее исполнения для пользователя кластер с распределяемой памятью представляется как обычная SMP-машина. Вполне естественно, что в силу меньшей в несколько раз скорости доступа к удаленной памяти по сравнению с локальной при равном числе ядер SMC будет уступать настоящей SMP-машине, но следует учесть, что в пересчете на одно посадочное место для процессора удельная стоимость находится в пределах 2 тыс. долларов, а в крупных SMP, стоящих сотни тысяч и миллионы долларов, она составляет как минимум 40–50 тыс. долларов, то есть кластерное решение на порядок и более дешевле.

SMC-кластеры с технологией Numascale были установлены в Технологическом университете имени короля Мангкута (Таиланд) и в Центре HPC в Шеньжене (Китай). На конференции SC14 в Новом Орлеане было объявлено, что Numascale совместно с Supermicro и AMD собрали самую большую в мире конфигурацию с разделяемой памятью в одном из американских ЦОД, оставшемся анонимным. Эта конфигурация состоит из 108 серверов 1U Supermicro на процессорах AMD Opteron 6386, объединенных в трехмерный тор и размещенных в трех стойках, суммарный размер памяти составляет 20.7 Тбайт. Здесь же, на SC14, Numascale и совместно с тайваньской 1degreenorth представила готовое аналитическое решение NumaQ, которое масштабируется начиная от 128 ядер и 1 Тбайта памяти и работает под управлением Red Hat Enterprise Linux. В данный момент на NumaQ устанавливается статистический пакет R, поддерживаемый средствами ПО NumaManager. Внешне для пользователя все выглядит так, как если бы он работал на обычном ПК.

IBM до продажи своего бизнеса Lenovo выпускала сервер модели x3755, укомплектованный адаптером NumaConnect. Сервер имеет модульную конструкцию, в модуле устанавливается два 16-ядерных процессора. Модули собираются в узлы, минимальная конфигурация состоит из 8 узлов (256 ядер) с общей памятью 896 Гбайт, максимальная состоит из 48 узлов, в ней размер прямо адресуемой памяти более 5 Тбайт.

«Невидимая революция»

Текст К.С. Амелин, Н.О. Амелина, Д.С. Будаев, О.Н. Граничин, Е.С. Левин, И.В. Майоров, П.О. Скобелев

Облачные технологии прочно вошли в нашу жизнь. Многие уже не представляют себе мир без таких сервисов, как Dropbox, Instagram, Facebook, поисковых сервисов – скажем, Google или Yandex.

Не многие догадываются, насколько серьезные вычислительные ресурсы на самом деле задействуются при очередном поиске в Google. Инфраструктуру поисковых сервисов можно сравнить с мощнейшими суперкомпьютерами, способными в моменты пиковых загрузок задействовать порядка 600 000 процессорных ядер. По некоторым оценкам, в 2012 году Google для своей вычислительной фермы мог использовать около 13 миллионов (!) процессорных ядер по всему миру. Многие люди пользуются облачными сервисами и даже не подозревают, какая сложная информационная и телекоммуникационная инфраструктура скрывается за такими казалось бы понятными и близкими им социальными сервисами. В ближайшее время IT-директоры крупных компаний, пожалуй, уже не смогут нарисовать схему инфраструктуры собственной компании по причине ее огромной сложности. Заметную роль в данном случае играют облачные технологии, возможности использования платформы как услуги, предоставление доступа к приложениям, работающим в облаке и полностью независимым от оборудования конечного пользователя.

По данным Forbes Magazine на октябрь 2014 года, около 75% опрошенных пользователей применяют сервисы облачных вычислений в том или ином виде, будь то хранение изображений или потоковое видео, или сервисы облачных ежедневников и напоминаний. Более того, около 86% компаний по всему миру используют более одного сервиса облачных вычислений в своей работе. Исследование Forbes Magazine прогнозирует, что уже в течение следующих 5–10 лет более 59% всех информационных технологий будет реализовано в облачной инфраструктуре. Ожидается, что уже к 2016 году облачные сервисы Азиатско-Тихоокеанского региона будут оперировать и хранить порядка 1.5 зетабайта данных!

Напомним, что 1 зетабайт – это 10^{21} байт. Для сравнения, нагрузка на облачные сервисы Североамериканского региона ожидается на уровне 1,1 зетабайта. Так какие же сюрпризы готовит для пользователей отрасль облачных вычислений?

Распределенные вычисления и «суперэластичные» приложения

Есть мнение, что в течение следующих 5 лет многие крупные компании «накопят» значительные вычислительные мощности, которые можно будет перепродавать по аналогии с крупными домохозяйствами, продающими излишек электроэнергии обратно в сеть. В частности, концепция Cisco's Intercloud не противоречит такому подходу. Действительно, подключение к сервису облачных вычислений очень напоминает подключение к электросети. Полагаем, что в скором времени будет разработано еще больше «стандартизированных под облако» приложений, которые можно будет подключать к облаку для масштабирования и повышения их быстродействия. «Сырые» вычислительные мощности становятся все более доступными, а это означает, что цены на услугу «инфраструктура как сервис» будут только снижаться. Вместе со снижением цен и распространением распределенных архитектур одно и то же приложение сможет использовать сразу несколько распределенных вычислительных подсистем. Потенциально возможен вариант, когда ваше приложение задействует вычислительные ресурсы некоторого региона или даже нескольких регионов. Уже сейчас Amazon Web Services работают по такому принципу (http://www.cloudcomputing-news.net/news/2014/oct/30/cloud-computing-in-2020-looking-into-that-crystal-ball/).

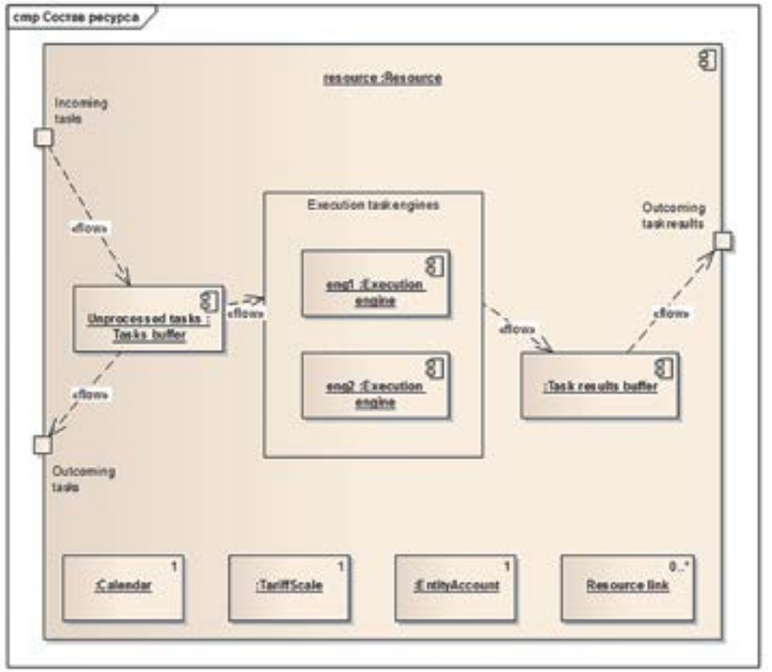


Рис. 1. Общая схема устройства ресурса

Интернет Вещей и взаимодействие уровня «machine-to-machine»

Давайте представим себе такой сценарий. Одно устройство (например, смартфон) в какой-то момент понимает, что ему требуются дополнительные вычислительные мощности, чтобы завершить некоторую работу. Такой работой может быть, скажем, обработка графического изображения. А причиной, по которой смартфон принял решение подключиться к облаку, была необходимость экономии заряда батареи. Смартфон (или программный агент смартфона) заключает сделку на подключение к среде облачных вычислений, успешно выполняет необходимый расчет и экономит заряд батареи. Пользователь добирается до места, где можно подзарядить телефон и есть безлимитный интернет-доступ. В этих новых условиях агент смартфона способен уже продавать вычислительные ресурсы смартфона. Считаете, это фантастика? Уже сейчас приложение ROCCATPOWER-GRID превращает смартфон в мощное игровое устройство, задействующее для сложных вычислений удаленные вычислительные ресурсы. Проще говоря, приложение, зная свой SLA (Service Level Agreement, соглашение об уровне предоставления услуги), может искать таких поставщиков ресурсов (интернет-провайдера, провайдеры облачных вычислений), которые в наилучшей мере могут удовлетворить потребности пользователя с наименьшими денежными издержками. Возможно, это кому-то не понравится, но голосовой помощник Siri от Apple уже отчасти реализует

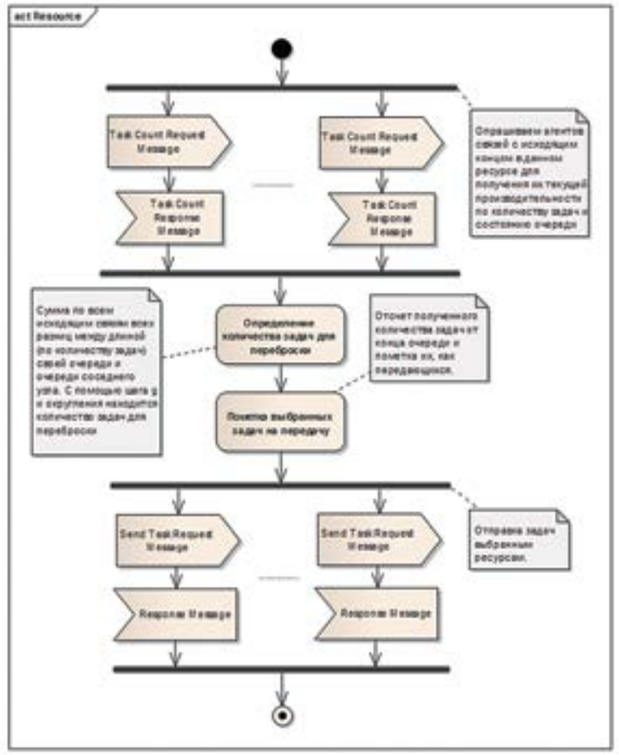


Рис. 2. Общий протокол балансировки

данную концепцию, когда обращается в дата-центры компании Apple не только для того, чтобы получить результаты запросов, но и для того, чтобы повысить точность самих запросов всех пользователей сервиса Siri (http://www.cloudcomputing-news.net/news/2014/apr/23/explosive-growth-cloud-computing-infographic/). Другая интересная сторона мира Интернета Вещей – получение новых свойств привычных устройств без модернизации самого устройства (hardware) посредством простого обновления программного обеспечения.

Новые возможности

Все вышеупомянутые технологические наработки накладывают определенные условия на работу и порядок мышления при проектировании и поддержании функционирования подобных сложных инфраструктур и облачных систем. Чтобы справиться с быстро меняющимися требованиями электронного бизнеса и уметь быстро масштабировать современные вычислительные системы, нужно перейти от статических моделей управления к динамическим. Требуется новый подход, когда правила, модели и код собираются динамически и настраивают все элементы инфраструктуры, необходимые для приложения в данный момент времени в данной ситуации. Уже сейчас с уверенностью можно утверждать, что без систем автоматического отслеживания состояния оборудования и приложений, использующих данное оборудование в облаке, невозможно будет себе представить надежные облачные системы будущего. И в этом контексте очень перспективным видится использование мультиагентных технологий и в целом мультиагентный подход к созданию и организации подобных систем мониторинга статуса вычислительных ресурсов, их диагностики, конфигурации и реконфигурации в зависимости от ситуации. То есть в данном случае можно говорить о мультиагентном подходе к управлению распределенными динамическими системами, когда каждый узел инфраструктуры управляется программной сущностью, агентом, способным анализировать информацию, непосредственно доступную от «подконтрольного» устройства, а также доступную посредством коммуникации с агентами соседних устройств. «Облачные дата-центры станут похожи на живые организмы», – говорит Джозеф Перер (Joseph Reger, Chief Technology Officer of Fujitsu Technology Solutions) (http://www.zdnet.com/cloud-computing-10-ways-it-will-change-by-2020_p2-7000001808/). Центры обработки данных все больше будут напоминать экосистемы, разрастаясь в моменты пиковых нагрузок и сокращаясь в моменты снижения активности используемых сервисов. Эти экосистемы по сути будут управляться и регулироваться программным обеспечением с функциями контроля и конфигурации оборудования. В рамках такого подхода возможна постановка различных целей мультиагентного управления, одной из которых может быть задача балансировки нагрузки на вычислительные ресурсы. Исследованию таких возможностей посвящен проект «Разработка мультиагентной технологии управления распределенными гетерогенными вычислительными ресурсами для адаптивной балансировки загрузки устройств в реальном времени», выполняемый авторами в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы».

Балансировка загрузки узлов распределенной вычислительной сети

Динамическое распределение задач по вычислительным ресурсам в гетерогенной сети предлагается проводить с помощью мультиагентного подхода. Основным сущностям системы ставятся в соответствие программные агенты: агент задачи, агент ресурса, агент канала связи, агент продукта (решенной задачи) и штабной агент.
Агенты описываются атрибутами, включающими наиболее существенные поля, свойства и правила взаимодействия агентов. Каждый агент имеет свою цель и стремится к ее достижению, отправляя и получая сообщения от других агентов. При решении задач балансировки нагрузки в вычислительной среде происходят изменения значений атрибутов агентов, поэтому в каждый момент времени существует некоторое состояние мультиагентной системы, которое итерационно стремится к равновесному состоянию. Данное равновесное состояние мультиагентной системы представлено в сцене – специальном контейнере, выделенной самостоятельной конструкции внутри ядра системы, которая позволяет хранить сущности, объектные связи между ними и другие необходимые для решения поставленных задач данные. Взаимодействие агентов осуществляется при помощи обмена сообщениями, в результате которых изменяются внутренние состояния агентов. При этом каждый агент стремится к локальному улучшению своих показателей, достигая консенсуса с остальными агентами. Состояние агентов по своим показателям оценивается при помощи «функций удовлетворенности», показывающих, насколько показатели отличаются от «идеальных». Таким образом, происходит улучшение заданных показателей системы в целом.
В рамках предлагаемой архитектуры системы ресурс представляет из себя некий объект (в частности, устройство), который обладает способностью принимать задачи, исполнять их (проводить некие вычисления, а также другие типы работ), отдавать результаты их выполнения, а также передавать задачи на выполнение связанным ресурсам (рис. 1). Агент ресурса предназначен для получения структурной информации из базы знаний о наличии и параметрах соединений с ресурсами, получения данных от агентов каналов, поддержки расписания работы, взаимодействия со штабным агентом, планирования задач для выполнения на данном ресурсе.
При своей инициализации агент ресурса получает всю необходимую информацию о том ресурсе, с которым он будет дальше работать. После активизации агент осуществляет первый раунд балансировки нагрузки.
В дальнейшем агент повторяет балансировку при поступлении таких событий, как получение новой задачи, изменение расписания работы, стоимостей выполнения задач и др.

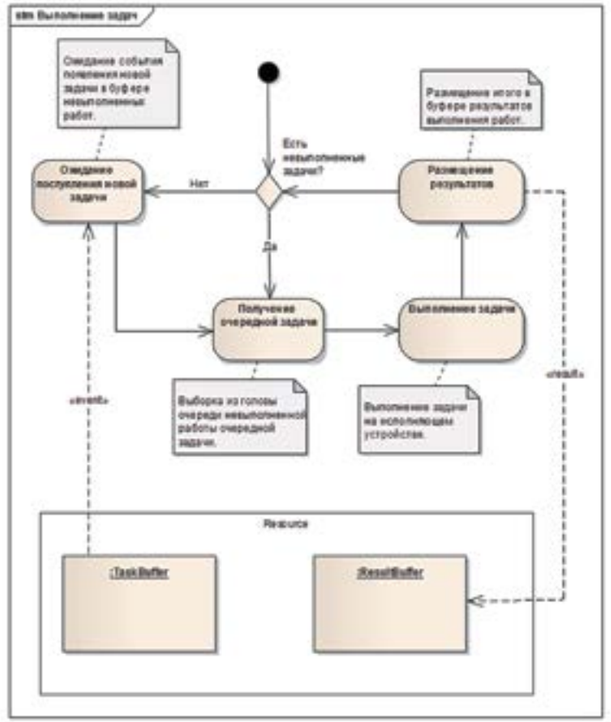


Рис. 3. Выполнение имеющихся работ

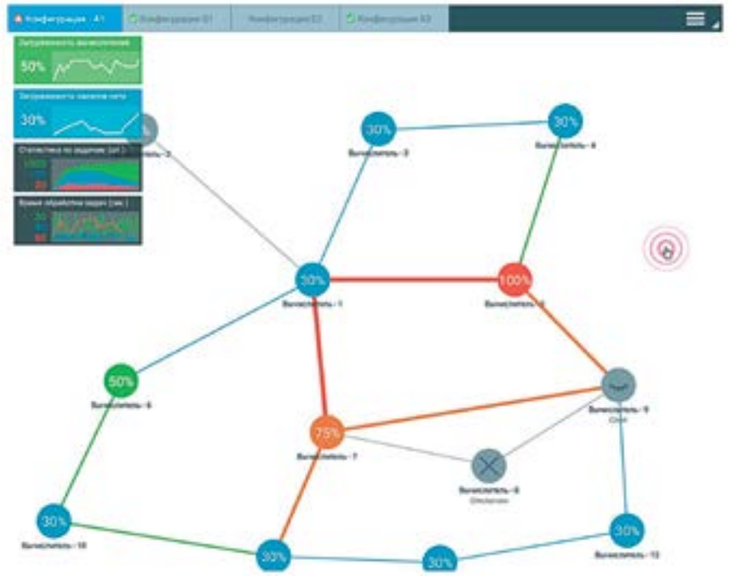


Рис. 4. Общий вид рабочего окна при запуске Системы

Агент ресурса осуществляет управление задачами для этого ресурса. Основной задачей агента ресурса является отслеживание состояния загруженности своего ресурса с целью оптимального расположения задач по ресурсам во всей системе. Оптимальность в данном случае подразумевает максимально быстрое выполнение всех поступающих задач в систему вычислителей с минимальными затратами. Под затратами понимается, например, время занятости или стоимость выполнения.

Агент задачи выполняет такие функции, как агрегация данных о задаче, взаимодействие с агентами ресурсов с целью согласования оплаты стоимости, проводит переговоры другими агентами по перемещению в очередях ресурсов.

При поступлении запроса о потенциальной удовлетворенности предлагаемым перемещением агент определяет, как предлагаемое перемещение повлияет на его характеристики, затем вычисляет удовлетворенность на основе определенного виртуального состояния, имеющейся информации о компонентах целевых функций и их весах. После определения значения формируется KPI и отправляется ответ на запрос текущей удовлетворенности отправителю исходного сообщения.

В системе также присутствуют агенты каналов, обеспечивающие создание и изменение топологии сети ресурсов, определение алгоритма разделения канала и приоритета передаваемых задач и стремящиеся максимально использовать свою пропускную способность.

Агент продукта формирует уже решенные задачи и распределяет их по адресатам.

Штабной агент собирает текущую статистику во фрагментах сети и участвует в разрешении конфликтов между агентами задач при затруднениях в выработке консенсуса. Цель достигается путем общения агентов между собой для перераспределения задач. Переговоры агентов – основной механизм для решения задачи по планированию размещения вычислительных задач на ресурсах.

Мультиагентный протокол переговоров для балансировки очереди задач

При запуске очередного прохода по балансировке, который инициируется поступившими сообщениями, агент ресурса определяет, какие есть каналы связи с другими узлами. Выбрав соответствующие связи, агент текущего обслуживаемого ресурса формирует запрос для получения их текущей производительности по количеству задач и состоянию очереди. Получив ответ с запрошенной информацией, находит сумму по всем исходящим связям всех разниц между длиной (по количеству задач) своей очереди и очереди соседнего узла. Затем находится количество задач для переброски, после чего выбранные задачи пересылаются соответствующим ресурсам.

При планировании пересылки учитывается, как изменится общий ключевой показатель эффективности (KPI) группы «Исходный ресурс – Целевой ресурс – Пересылаемая задача». Перемещение осуществляется только при увеличении суммы KPI данных трех участников.

Удовлетворенности агентов определяются на основе заданных для каждого компонента собственных целевых функций.

Общий вид протокола представлен на рис. 2. Помимо задач балансировки и ответа на поступающие сообщения одной из основных потребностей агента ресурса является управление процессом выполнения имеющихся задач.

В рамках первого этапа работ по разработке мультиагентной технологии управления распределенными вычислительными ресурсами также разработаны макеты визуального представления компонент системы.

Для решения основной задачи проекта были разработаны программные модели и алгоритмы взаимодействия для исследования производительности систем обработки и хранения данных.

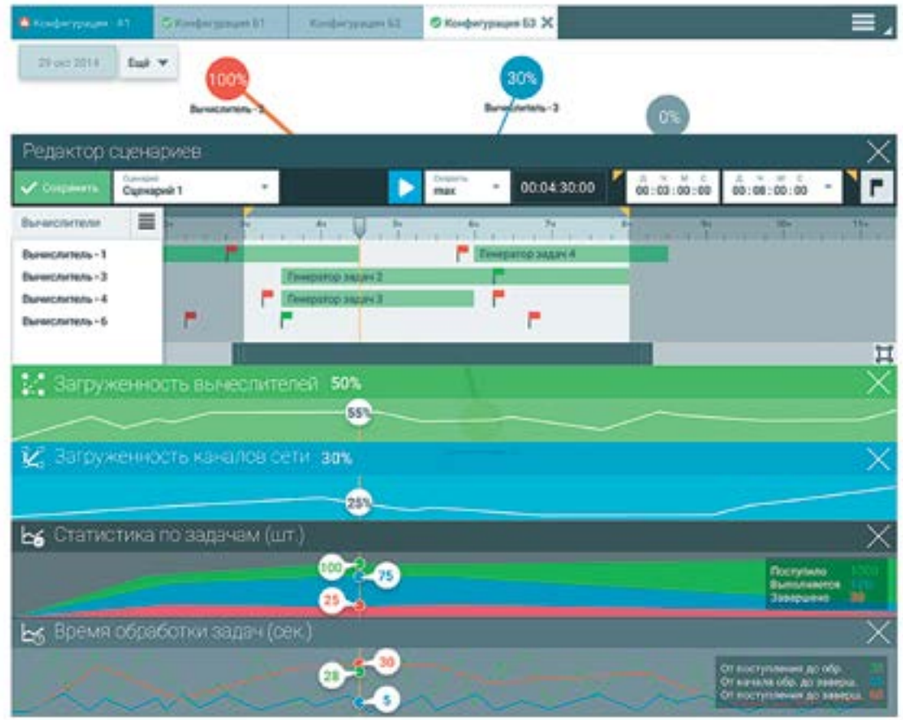


Рис. 5. Вид рабочего окна Системы с отображаемыми элементами «События Системы» и «Сценарии»

Для исследования характеристик системы был разработан макет прототипа системы, реализующий базовые классы агентов и сущностей, протоколы взаимодействия и алгоритмы мультиагентного перераспределения поступающих задач. Результаты исследований показали, что созданный в результате выполнения первого этапа работ макет системы и заложенные в него технологические решения обеспечивают возможность адаптивного перераспределения задач в сетях вычислительных ресурсов и возможность обработки задач.

Макет прототипа системы, созданный в рамках первого этапа работ, обеспечивает возможность моделирования динамического перераспределения задач посредством взаимодействия агентов ресурсов, что позволяет гибко реагировать на непредвиденные события, моделируемые или возникающие в условиях работы сетевых систем обработки и хранения данных, что позволяет планировать и оптимизировать распределение ресурсов вычислительной системы в режиме реального времени. Моделирование позволяет проводить выполнение сценариев, анализировать реакции и упреждающие воздействия системы в ответ на непредвиденные события.

Макет прототипа системы может применяться для дальнейшего развития технологии, при анализе и проектировании новых алгоритмов управления сетевыми вычислительными ресурсами, процесс работы которых характеризуется постоянными перестроениями конфигураций, динамикой изменений в среде, а также неопределенностью в характеристиках поступающих вычислительных задач, что требует адаптивности при реакции на непредсказуемые события в режиме реального времени.

Суммируя все вышесказанное, можно с уверенностью утверждать, что мы находимся в начале эры потрясающих возможностей в облачных вычислениях. И очень важно вовремя заметить открывающиеся возможности «невидимой революции»

К вопросу о федеративной организации распределенной ЦЕРН

Текст А. Климентов
(НИЦ «Курчатовский институт», БНЛ)

Ландшафт современных компьютерных ресурсов огромен и разнообразен, сложен и разнороден, способен к решению многих задач, и скорее выглядит как большой архипелаг, чем как группа материков.

Наряду с центрами коллективного пользования, которые подразумевают использование ресурса группами ученых, часто работающих в разных областях наук, узковедомственными и/или узкоспециализированными ВЦ, существуют суперкомпьютеры, грид-консорциумы, коммерческие и академические облачные ресурсы, университетские вычислительные кластеры. Такова тенденция организации вычислительных мощностей (киберинфраструктуры) во всем мире. В настоящее время распределенная киберинфраструктура и ее составляющие используются в лучшем случае индивидуально, а чаще как изолированные ресурсы. Аристотель утверждал, что «целое больше, чем сумма его частей», поэтому федеративная организация РКИ позволит использовать компьютерные ресурсы более эффективно, что будет выгодно как «владельцам» ресурса, так и пользователям.
Создать федерацию компьютерных ресурсов совсем не просто, потому что упомянутые выше мощности часто принадлежат разным проектам и управляются разными структурами, имеющими свои особенности и приоритеты. Несмотря на социологические и организационные сложности, мы обсудим возможность создания федерации распределенной киберинфраструктуры (ФРКИ) в рамках академических, университетских и исследовательских центров, и покажем, что это мотивировано экономически и технически.
Существующие проблемы создания ФРКИ характеризуются:
а) недостатком блоков, из которых можно построить федерацию;
б) точечными решениями, не выходящими за пределы немедленного использования и конкретной задачи (и/или центра), другими словами, это прикладные решения с низким уровнем абстракции для интерфейсов и модулей программного обеспечения;
в) вопросы интеграции РКИ, как правило, рассматриваются только после создания инфраструктуры, а не на этапе разработки архитектуры распределенной системы (например, при создании грид-инфраструктуры и программного обеспечения для экспериментов на Большом Адронном Коллайдере не была предусмотрена возможность использовать ресурс университетских и суперкомпьютерных центров).
Таким образом, необходимо решить все три проблемы, и одновременно необходимо следовать следующим принципиальным подходам при создании ФРКИ: единый метод и уровень абстракции управления ресурсами; общая система запуска задач и передачи данных в гетерогенной компьютерной среде; интегрируемые и развиваемые средства для разработки и управления программным обеспечением (ПО) ФРКИ.
Все вместе это составляет инженерную проблему фундаментального характера, а отсутствие адекватного ее решения приводит к экономическим и функциональным потерям.
Участвуя в международном научном сотрудничестве ATLAS и, в частности, в разработке и создании компьютерной модели эксперимента и системе распределенной обработки данных, мы пришли к выводу о необходимости федеративного устройства вычислительных ресурсов. И хотя начальная мотивация связана с экспериментами на Большом Адронном Коллайдере и количественные требования имеют специфическую особенность экспериментов в области физики высоких энергий, качественные требования являются типичными для научных приложений в областях наук, требующих анализа и обработки данных в пета- и экзабайтном диапазоне.

Мотивация в создании федеративной киберинфраструктуры для ATLAS и экспериментов на Большом Адронном Коллайдере

Самый большой в мире прибор для исследований, Большой Адронный Коллайдер (БАК), работает в Международной лаборатории ЦЕРН в Женеве, Швейцария. Эксперименты на БАК изучают фундаментальную природу материи и основных сил, которые формируют нашу Вселенную. ATLAS, одна из самых больших научных коллабораций, когда-либо созданная в фундаментальной науке, находится в центре исследований на БАК. Чтобы решить беспрецедентную проблему обработки экзабайтных данных, ATLAS, как и другие эксперименты БАК, использует вычислительную инфраструктуру грид, развернутую в рамках проекта Worldwide LHC Computing Grid (WLCG). WLCG – самая большая академическая распределенная вычислительная среда в мире, состоящая из около 150 вычислительных центров, в которых более 8000 ученых анализируют данные БАК в поисках новых явлений в физике (на рис. 1 показана карта вычислительных центров, входящих в проект WLCG).
Научный прорыв 2012–2013 годов – открытие бозона Хиггса – был триумфом научного мегапроекта Большой Адронный Коллайдер, когда эксперименты ATLAS и CMS обработали беспрецедентные объемы информации и подтвердили открытие новой элементарной частицы.

Рис. 1

С точки зрения потребностей в обработке и анализе данных, ATLAS является ярким примером того, почему необходима федеративная организация вычислительных мощностей.
Типичные вычислительные потребности эксперимента характеризуются следующими величинами: два миллиона выполненных заданий, три миллиона используемых ЦПУ часов в сутки.

- *Наиболее интенсивное ATLAS-«задание» – это приложение для обработки данных, использующее до 2 Гбайт памяти на одном ядре в течение примерно 12 часов, обрабатывающее максимально несколько гигабайтов входных данных и производящее выходные данные такого же объема.*
- *Непрерывная глобальная передача данных на уровне 100 Гбит/с суммарно.*
- *Требуется шесть миллионов ядро-часов (core-hours) для первого шага обработки одного петабайта данных ATLAS, полученных от одного миллиарда актов столкновений на БАК.*
- *Совокупный поток заданий ATLAS использует более 120 000 ЦПУ-ядер с пиковой производительностью приблизительно 0.24 Петафлопса (такую производительность обеспечивает суперкомпьютер N27 из списка TOP100).*

Есть по крайней мере два дополнительных параметра, заслуживающих упоминания:
а) загрузка мощностей не является статичной во времени и зависит от многих причин. При хорошо определенном среднем значении в 2 миллиона задач в день существуют сильные временные флуктуации (рис. 3), когда потребность в ресурсах возрастает в 10 и более раз в течение короткого (дни) промежутка времени, что соответствует классическим примерам систем с ограничением в передаче данных и доступа к распределенному вычислительному ресурсу;
б) стабильная во времени потребность в компьютерных мощностях (т. н. steady state demand).
Даже в отсутствии пиковых нагрузок ресурс, предоставляемый консорциумом WLCG, часто недостаточен. Пример ATLAS'а не единичен, другие эксперименты БАК: ALICE, CMS и LHCb, сталкиваются со схожей проблемой.
Необходимо также отметить, что следующее поколение программного обеспечения экспериментов в области физики высоких энергий и ядерной физики (ФВЭиЯФ) будет гораздо более комплексным, сложным и неоднородным, поэтому пост-Хиггс эра (исследование свойств новой частицы и, возможно, поиск второй и третьей частиц а la Хиггс) потребуют в будущем гораздо большего вычислительного ресурса (по разным оценкам, экспериментам БАК будет необходимо

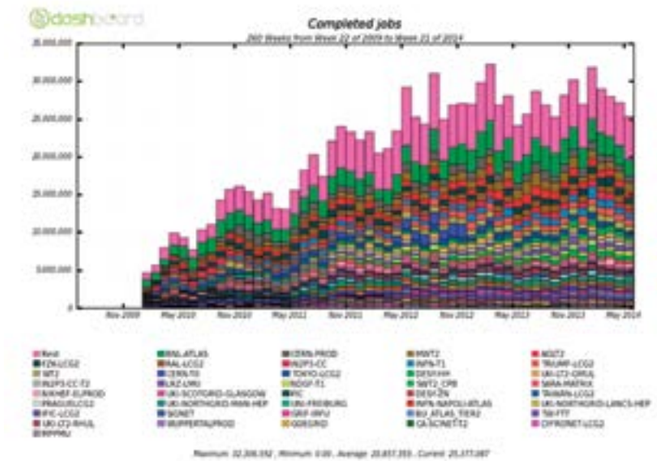


Рис. 2

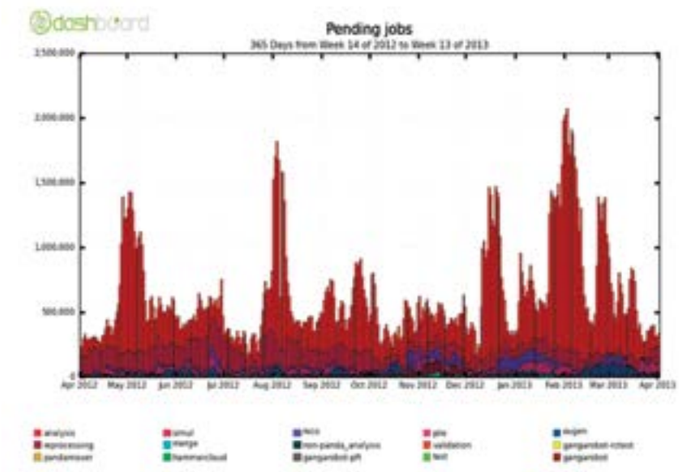


Рис. 3

40–50-кратное увеличение вычислительной мощности в течение следующих 3–5 лет). Что в первую очередь позволяет сделать предположение о возможной роли суперкомпьютеров, до последнего времени не входящих в список традиционно используемых в ФВЭиЯФ мощностей.

Более того, «загрузка» суперкомпьютеров приложениями ФВЭиЯФ не только может быть потребностью экспериментов, но и будет иметь сильный экономический аргумент для «владельцев» подобных машин. Хотя точная цифра загрузки суперкомпьютерных центров широко не афишируется, будет возможным предположить, что она не превышает 90% (авторы использовали информацию последней суперкомпьютерной конференции (SC14, Новый Орлеан, ноябрь 2014 г.) и технические данные некоторых центров в России, США и Европе). Таким образом, около 10% ресурса могут быть использованы в режиме backfill, без изменения существующего портфеля задач, что повысит «загрузку» машин и процент их утилизации. Т. е. более гибкое и эффективное использование суперкомпьютерного ресурса возможно за счет приложений ФВЭиЯФ, когда такой ресурс доступен в суперкомпьютерном центре и не используется для специальных приложений.

Важно понимать, что подобный подход представляет интерес для многих научных приложений (за пределами экспериментов на БАК) и для многих научных дисциплин.

Как организовать федерацию в неоднородной компьютерной среде?

Как обсуждалось ранее, есть много нерешенных проблем и препятствий для создания ФРКИ, часть из них – это функциональные трудности (запуск заданий, передача данных, создание информационной системы для описания гетерогенного ресурса), другие относятся к области идентификации пользователей, возможным протоколам обмена информацией и политики использования ресурса. В последнее время есть успешные попытки синхронизировать и гармонизировать вторую группу проблем, но в этом есть смысл, если существуют все узловые функциональные блоки для создания ФРКИ. Возможное решение может быть основано на использовании существующей системы управления заданий PanDA.

PanDA – одна из самых успешных систем, разработанных в области физики высоких энергий. PanDA (акроним для Production and Distributed Analysis – производственный и распределенный анализ), используемый тысячами физиков в эксперименте ATLAS, управляет упорядоченным потоком задач (например, задачи обработки данных и Монте-Карло моделирования) и хаотическим потоком задач (задачи анализа данных, запускаемые участниками эксперимента). PanDA уже была процитирована в качестве примера «отказоустойчивого программного обеспечения высокой производительности для быстрого, масштабируемого доступа к репозиториям данных многих видов» при объявлении об «Инициативе по развитию и исследованию технологий больших данных». На рис. 4 (Т. Маено, БНЛ) схематически показана архитектура системы.

В настоящее время PanDA обобщается и пакетируется как программный комплекс, уже доказавший свою применимость при экстремальных масштабах, для более широкого использования в области обработки Больших Данных.

В рамках проекта, поддерживаемого Правительством Российской Федерации, – megaPanDA, реализуемого на базе НИЦ «Курчатовский институт» (ГК № 14.Z50.31.0024), ведется разработка и интеграция системы управления данными и через единый портал происходит запуск задач на грид и суперкомпьютерный комплекс в «Курчатовском институте». Первоначальный класс задач включает в себя приложения ФВЭиЯФ и вычислительной биологии.

Создание ФРКИ с использованием системы PanDA как основного ПО должно следовать базовым принципам, описанным ранее: единый подход и уровень абстракции управления ресурсами; общая система запуска задач и передачи данных в гетерогенной компьютерной среде; интегрируемые и развиваемые средства программного обеспечения.

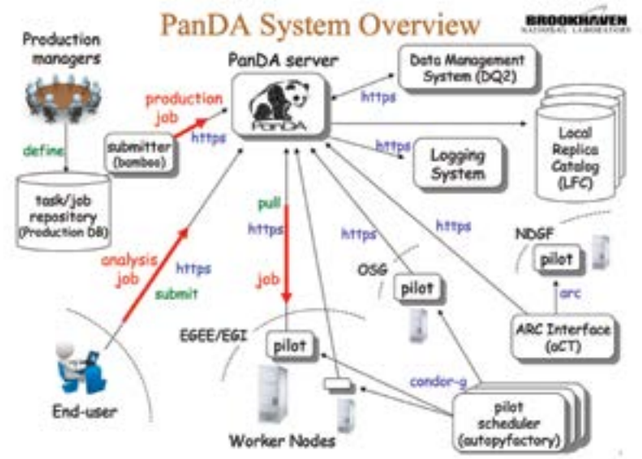


Рис. 4

В первую очередь необходимо:

- *определить класс научных приложений, которым необходима федерация вычислительных ресурсов, поддерживаемых WLCG (грид) и суперкомпьютерными центрами;*
- *создать архитектуру ФРКИ, основанную на принципах федеративного устройства вычислительных мощностей и системы PanDA;*
- *выбрать решение проблемы, позволяющее гибкую настройку системы и применимую для экспериментов ФВЭиЯФ (БАК, НИКА, FAIR), астрофизики (эксперименты AMS, LSST), экспериментов на установке XFEL (DESY, Гамбург, ФРГ), вычислительной химии и биологии.*
- *создать единую информационную систему для ресурсов с технологиями «грид – суперкомпьютерный центр – облачные вычисления» с учетом интеграции в будущем университетских кластеров и специализированных ВЦ;*
- *разработать систему управления данными, интегрированную с системой запуска заданий (PanDA).*

Подход на основе системы megaPanDA адресует общие требования научных приложений и предлагает исследовать, как компьютерные мощности могут быть использованы совместно, при этом сохраняя свою независимость и особенности, а также для различных научных приложений.

В настоящий момент работа начата в Научном исследовательском центре «Курчатовский Институт» и Объединенном институте ядерных исследований (Дубна) в сотрудничестве с ЦЕРН, Брукхэвенской национальной лабораторией (США), университетами Штата Техас и Rutgers, экспериментами ATLAS и ALICE, а также с суперкомпьютерными центрами России, Европы и Америки.

На рис. 4 показан пример интеграции PanDA, работающий на суперкомпьютере #2 – Titan.

Возможный эффект такой интеграции может быть выше и послужить началом для организации ресурсов для совершенно иных масштабов и типов задач.

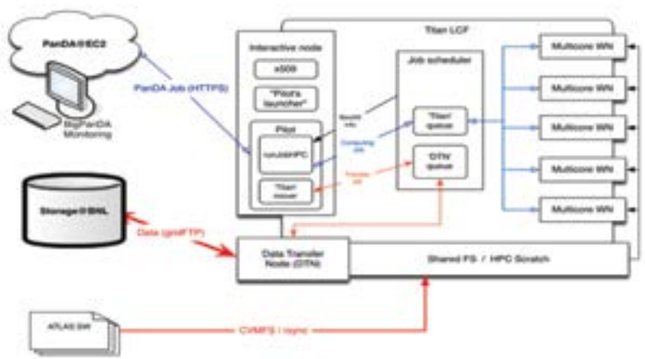
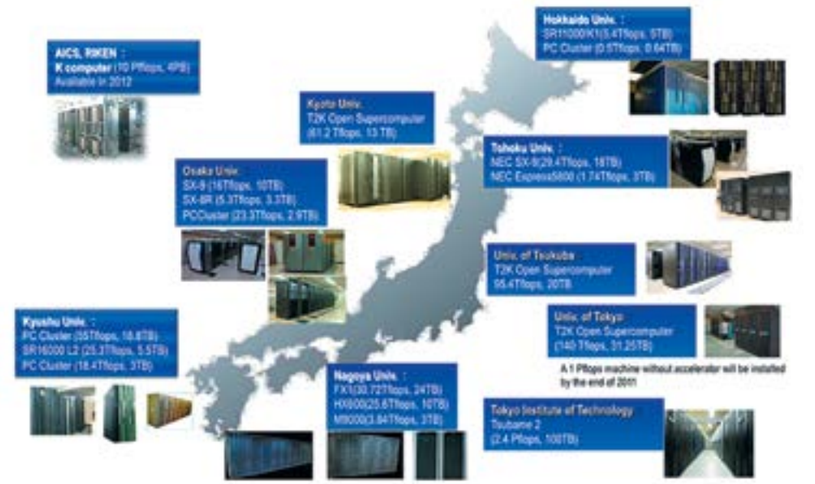


Рис. 5

Киберинфраструктура

Текст Екатерина Тютляева, Михаил Тютляев



Стоит отметить еще один критический аспект развития науки в эпоху «Четвертой парадигмы» – воспроизводимость. Общеизвестно, что воспроизводимость является одной из основных характеристик научного знания. Под этим термином понимается возможность повторить методы и результаты научного исследования. Этот вопрос, относящийся к философии науки, неожиданно и непосредственно влияет на область разработки систем хранения. Очевидно, что, если в одной организации удается получить научный результат путем анализа большого объема данных, научное сообщество должно иметь доступ к начальным данным и результатам исследования. Наиболее оптимальным решением является создание специализированной национальной киберинфраструктуры, которая позволяет обеспечить доступ к большим массивам специализированных данных всем заинтересованным группам специалистов. Такая система действительно способствует получению нового научного знания в эпоху Больших Данных. В мировой практике существует несколько успешных примеров.

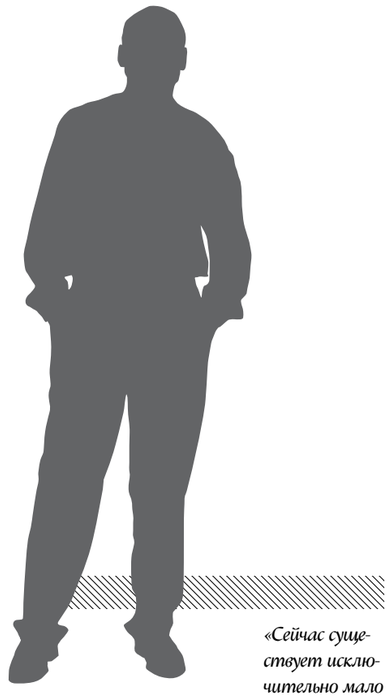
Наиболее впечатляющим примером является японская киберинфраструктура HPCI, базирующаяся на системе Gfarm (GRDI Data farm), которая объединяет все ведущие суперкомпьютерные центры страны (см. рис.).

Всего в ней распределено 20 PB пространства с производительностью копирования в 1 ГБ/с, при этом поддерживается параллельная репликация файлов, настройки приватности. Для того чтобы получить доступ к киберинфраструктуре, достаточно быть зарегистрированным на любом из ресурсов, входящих в ее состав.

В США существует несколько специализированных научных сетей, ориентированных на различные области науки, к примеру, упомянутое выше облако для хранения научных данных, в котором операции ввода-вывода данных производятся бесплатно, в отличие от коммерческих решений. Другим примером может являться инициатива iPlant, базирующаяся на системе управления данными на основе правил iRODS, использующей вычислительные узлы с глобальной системой управления данных. Ключевой особенностью iRODS является возможность управления различными цифровыми объектами, сохраненными на множестве систем (NFS, HDFS, HPSS и т.д.). При этом сохраняется единое и целостное пространство имен, централизованная система метаданных, множество клиентов и API и возможность установки гибких настроек безопасности и управления данными (запросы через Web, SQL, Hadoop, использование специализированных рабочих потоков и правил для обработки данных). Кроме инициативы iPlant iRODS используется в ряде академических организаций и научных репозиториев по всему миру. Статистика использования iRODS [Irodsstat] позволяет оценить данную систему как одну из наиболее популярных на сегодняшний день систем для организации академических data-грид киберинфраструктур.

Системы хранения данных лидирующих супер-компьютеров

Текст Екатерина Тютляева, Михаил Тютляев



«Сейчас существует исключительно мало практически значимых высокомасштабируемых приложений, которые НЕ работают интенсивно с данными».

Alok Choudhary, IESP

Начинать статью с разговора о том, что Большие Данные накапливаются повсюду, – это не самое оригинальное вступление. Но зато правдивое. На конференциях, проходивших в текущем году, прозвучало несколько любопытных фактов на эту тему.

Например, уже к 2015 или 2016 году количество электронных устройств, имеющих выход в Сеть, будет превышать численность населения Земли в два раза. Как следствие, утверждается, что к 2015 году потребуется пять лет, чтобы посмотреть все видео, которое проходит через сети IP за одну секунду. По материалам журнала «Гарвард», за последние два года человечество собрало больший объем данных, чем за всю предыдущую историю. Наконец, исследования утверждают, что к 2015 году количество вакантных мест для специалистов по работе с данными и аналитиков достигнет 4.4 миллиона, и только треть этих мест будет занята.

Мы живем в эпоху новой, четвертой парадигмы получения научного знания, основанной на технологиях сбора, анализа, визуализации и поиска закономерностей в больших массивах данных. Исследования, связанные с обработкой Больших Данных, проводятся в самых различных областях:

- *Помощник профессора управления «Гарварда» провел следующий эксперимент: 87 профессорам предложили предсказать решение Верховного Суда по рассмотренным делам за последний год. Профессора отлично разбирались в юриспруденции и знали решение Верховного Суда в предшествующих аналогичных случаях. С ними соревновалась статистическая модель, анализирующая доступные данные. Эксперимент продемонстрировал, что статистическая модель однозначно побеждает в производительности и точности решения как одного профессора, так и небольшие группы.*
- *В маркетинге анализ Больших Данных позволяет разработать*

Таблица 1. Июнь, 2006

	Имя	Объем	Пропускная способность	Файловая система
1	BlueGene/L			
2	BGW	60 ТБ		GPFS
3	ASC Purple	1.6 ПБ (2 ПБ на 2007 г.)	102 ГБ/с	
4	Columbia	650 ТБ		
5	Tera-10	1 ПБ	100 ГБ/с	Lustre
6	Thunderbird	120 ТБ	6.0 ГБ/с	Lustre
		50 ТБ	4.0 ГБ/с	PANASAS
7	TSUBAME Grid Cluster	1 ПБ (2007)	8 ГБ/с	Lustre
8	JUBL	-	-	-
9	Red Storm	240 ТБизначально, 1753 ТБ к 2008 г.	50 ГБ/с	Lustre
10	Earth-Simulator	240 ТБ		

Таблица 2. Июнь, 2011

	Имя	Объем	Пропускная способность	Файловая система
1	K Computer	(от 100 ПБ до 1 ЕВ) ожидается	(~ГБ/с) (~ТБ/с).	FEFS
2	Tianhe-1A	1 ПБ		Lustre
3	Jaguar	10 ПБ	240 ГБ/с	Spider
4	Nebulae	-	-	-
5	TSUBAME2.0	15 ПБ, 7+8 на магнитных д.		Lustre, NFS Home
6	Cielo	10 ПБ	160 ГБ/с	PANASAS
7	Pleiades	6.9 ПБ		7 Lustre
8	Hopper	2 ПБ+ FS NERSC	35 ГБ/с	Lustre
9	Tera-100	20 ПБ	500 ГБ/с	Lustre
10	Roadrunner	2 ПБ	~60 ГБ/с	PANASAS

Таблица 3. Ноябрь, 2014

	Имя	Объем	Пропускная способность	Файловая система
1	Tianhe-2	12.4 ПБ	~750 ГБ/с	Lustre/H2FS
2	Titan	10.5 ПБ	240 ГБ/с	Lustre
3	Sequoia	55 ПБ	850 ГБ/с	Lustre
4	K Computer	40 ПБ	965 ГБ/с	Lustre
5	Mira	7.6 ПБ	88 ГБ/с	GPFS
6	Piz Daint	2.5 ПБ	138 ГБ/с	Lustre
7	Stampede	14 ПБ	150 ГБ/с	Lustre/H2FS
8	JUQUEEN	5.6 ПБ	33 ГБ/с	GPFS
9	Vulcan	55 ПБ	850 ГБ/с	Lustre
10	Cray CS Storm	?	?	Lustre

эффективные рекомендательные сервисы, такие как у корпораций Netflix и Amazon. Более интересными примерами могут служить исследования, которые определяют, что женщина-клиент беременна, если она заинтересована в лосьонах для тела без отдушек, или исследование компаний, предоставляющих кредитные карты, что показателем благонадежности клиента является приобретение им противоскользящих ковриков под мебель.

• В социальной сфере проводится огромное количество исследований, выявляющих взаимосвязь между условиями окружающей среды и здоровьем населения, предсказывающих увеличение преступности в определенном районе и т.п.

Согласно докладу департамента энергии США в области научных данных, сейчас можно назвать такие объемы:

• Данные БАК – 15 ПБ/год.

• Данные проекта, посвященного исследованию генома человека, – 10 ПБ.

• Работы по изучению источников света – 300 ТБ/день.

• Климатические данные – 100 ЕБ(ожидается).

Рассмотренные примеры доказывают, что организация процесса работы с данными становится первоочередным вопросом в области высокопроизводительных вычислений.

Итак, немного о тенденциях в области организации систем хранения данных для ведущих суперкомпьютеров мира.

Анализ рейтинга TOP500

Признанным средством анализа профиля области является рейтинг TOP500, позволяющий оценить-системы, которые используются лидерами в области высокопроизводительных вычислений. Посмотрим для сравнения первую десятку ведущих суперкомпьютеров за 2006, 2011и 2014 годы.

Итак, в 2006 году в первой десятке суперкомпьютеров лидирующей была СХД суперкомпьютера ASC Purple, обладающая объемом в 1.6 ПБ и пропускной способностью в 102 ГБ/с. В свое время это была знаковая система, которая позволила преодолеть так называемый «гигабайтовый барьер», выражающийся в неспособности интерконнекта большого суперкомпьютера «насытить» процессор данными.

В 2011 году ведущая тройка по объему составляла 20 ПБ (Tera-100), 15 ПБ (Tsubame 2.0) и 10 ПБ (Cielo, Jaguar).

В 2014 году лидирующая тройка составляет 55ПБ (Sequoia, Vulcan) и 40ПБ (K Computer). 10 ПБ в 2011 году является третьим по значимости показателем, в то время как в 2014 году как минимум 6 систем из 10 уверенно превышают этот показатель. Максимальный показатель пропускной способности (965 ГБ/с, K Computer) увеличился почти в два раза (по сравнению с максимальным за 2011 год, 500 ГБ/с, Tera-100) и уверенно приближается к 1 ТБ/с.

Интересно также проследить эволюцию системы хранения на K Computer. В 2011 году было заявлено, что ожидается расширение хранилища до объема в экзабайт данных и пропускной способности в 1 ТБ/с. В настоящее время пропускная способность (965 ГБ/с) соответствует заявленной, а доступный объем (40 ПБ) ниже ожидаемого на два порядка.

Впрочем, с объемом систем хранения все не так однозначно. Во-первых, большинство установок имеет доступ к системам хранения данных на магнитных лентах.

К примеру, с суперкомпьютера Stampede доступна 60 ПБархивная система TACC.

Во-вторых, ряд суперкомпьютеров из первой десятки разделяет систему хранения как инфраструктуру на различных уровнях – от организационного до государственного. Так, K Computer играет ключевую роль в японской инфраструктуре HPCI.

Рассмотрим подробнее наиболее интересные технологии.

Системы управления данными

Ведущую роль в верхней части рейтинга играет файловая система Lustre, которая также установлена на системе хранения суперкомпьютера NCSA Bluewaters,с показателем емкости в 24 ПБ и рекордной пропускной способностью в 1100 ГБ/с. Также в первую десятку рейтинга вошли системы, использующие GPFS, и хотя по средним показателям они проигрывают системам, базирующимся на Lustre, они демонстрируют перспективные подходы к организации системы хранения. Например, на суперкомпьютере Mira организован промежуточный уровень хранения по типу Burst Buffer.

Окончательно вытеснена из первой десятки кластерная файловая система PANASAS. Как отмечает в своем докладе Торбен Клинг Петерсен, PANASAS исчерпала свои возможности. Впрочем, в июне 2014 года была представлена новая версия – PanFS6.0, базирующаяся на инновационных принципах, которые могут повысить ее конкурентоспособность.

Но на уровне широкого круга пользователей высокопроизводительных кластеров Lustre и GPFS обладают сравнимой популярностью. Согласно результатам переписи Intersect 360 за 2014 год, прогресс в области адаптации параллельных файловых систем проходит чрезвычайно медленно.

Наиболее решительные шаги наблюдаются в области академических и правительственных исследований. Большинство ПО, обеспечивающего управление системой хранения (36%), предоставляется производителем систем хранения. Lustre и GPFS являются наиболее часто упоминаемыми системами с показателями в 19% и 17 %соответственно.

Объектное хранилище

По-прежнему лидирующую позицию сохраняет концепция объектного хранения цифровых объектов, которая должна в ближайшем времени вытеснить файловые системы, наиболее ярким представителем которой является ФС Lustre. Каждый файл и директория в Lustre рассматривается как отдельный объект с определенными атрибутами. Метаданные, включающие информацию о распределении файла в системе хранения, содержатся на сервере метаданных, в то время как сам файл хранится в объектном хранилище. Такой подход обладает рядом преимуществ для больших систем хранения и предоставляет практически неограниченные возможности масштабирования емкости хранилища, и увеличения скорости доступа к данным.

Интенсивные операции с метаданными, впрочем, являются узким местом современных систем, и для решения этой проблемы предлагаются подходы, основанные на:

- «ленивой» синхронизации пространства имен и оптимистичной верификации метаданных.
- использовании модели графа отношения (применяется для анализа социальных сетей и других приложений, имеющих дело с Big Data) для сохранения «богатых метаданных» в объектном хранилище.

Burst Buffer

Одним из решений, которое позволит минимизировать время ввода-вывода для приложений и обеспечить надлежащую пропускную способность, является так называемый Burst Buffer. Под этим термином понимается включение промежуточного уровня хранения, состоящего из устройств хранения, обеспечивающих высокую про-

пускную способность и сравнительно небольшой объем хранения. Эти устройства должны действовать как участок подготовки необходимых данных или как кэш с обратной записью для высокопроизводительных систем хранения. Рядом исследователей утверждается, что к 2020 году уровень Burst Buffer станет обязательным компонентом лидирующих высокопроизводительных установок.

Коммерческие разработки

Все более перспективной выглядит архитектура многоуровневого хранилища, которая позволяет интегрировать в единую систему уровень SSD, различные традиционные диски и зачастую хранилище на магнитных лентах. Основная проблема заключается в разработке слоя абстракции, который позволит предоставить единообразный доступ ко всем уровням полученной системы и обеспечить автоматическое перемещение данных между этими уровнями, представленными различными производителями, программными и аппаратными решениями.
Ряд ведущих коммерческих компаний уже представляет соответствующие разработки.

- *Компания Data Direct Networks представила на конференции Supercomputing-2014 новую версию решения IME (Infinite Memory Engine). IME позволяет объединить присутствующие в системе высокопроизводительные флэш-накопители в уровень Burst Buffer, который повышает производительность подсистемы ввода-вывода за счет интеграции небольших операций ввода-вывода в более крупные и удаления замедляющих коммуникации ограничений на уровне POSIX, таких как блокировка доступа к файлу. Предложенный менеджер рассчитан на тысячи узлов ввода-вывода, петабайты данных и работает с любыми модификациями Lustre и GPFS. При этом не требуется никаких дополнительных изменений в приложениях.*
- *Компания Cray предлагает решение Tiered Storage и систему управления Versity, позволяющую определять и настраивать различные стратегии хранения, обеспечивает миграцию данных между уровнями и создание резервных копий при достижении специфических указанных условий.*
- *IBM предлагает решение Easy Tier, которое автоматически перемещает часто используемые данные с традиционных дисков на SSD-диски согласно запатентованным алгоритмам, которые вычисляют частоту доступа к данным и перемещают более «горячие» данные, к которым наиболее частый доступ, на SSD-диски, а «остывшие» данные, частота доступа к которым снизилась, – на традиционный уровень хранения.*

SSD

Тем не менее, согласно исследованию InterSect360, использование SSD-накопителей в высокопроизводительных установках все еще находится на ранней стадии развития. Только 14% систем, модифицированных после 2012 года, используют SSD как минимум на нескольких узлах. При этом очень мало систем используют только SSD-накопители. Чаще всего SSD применяется для организации дополнительного слоя между памятью и традиционными жесткими дисками.

Хранилища на магнитных лентах

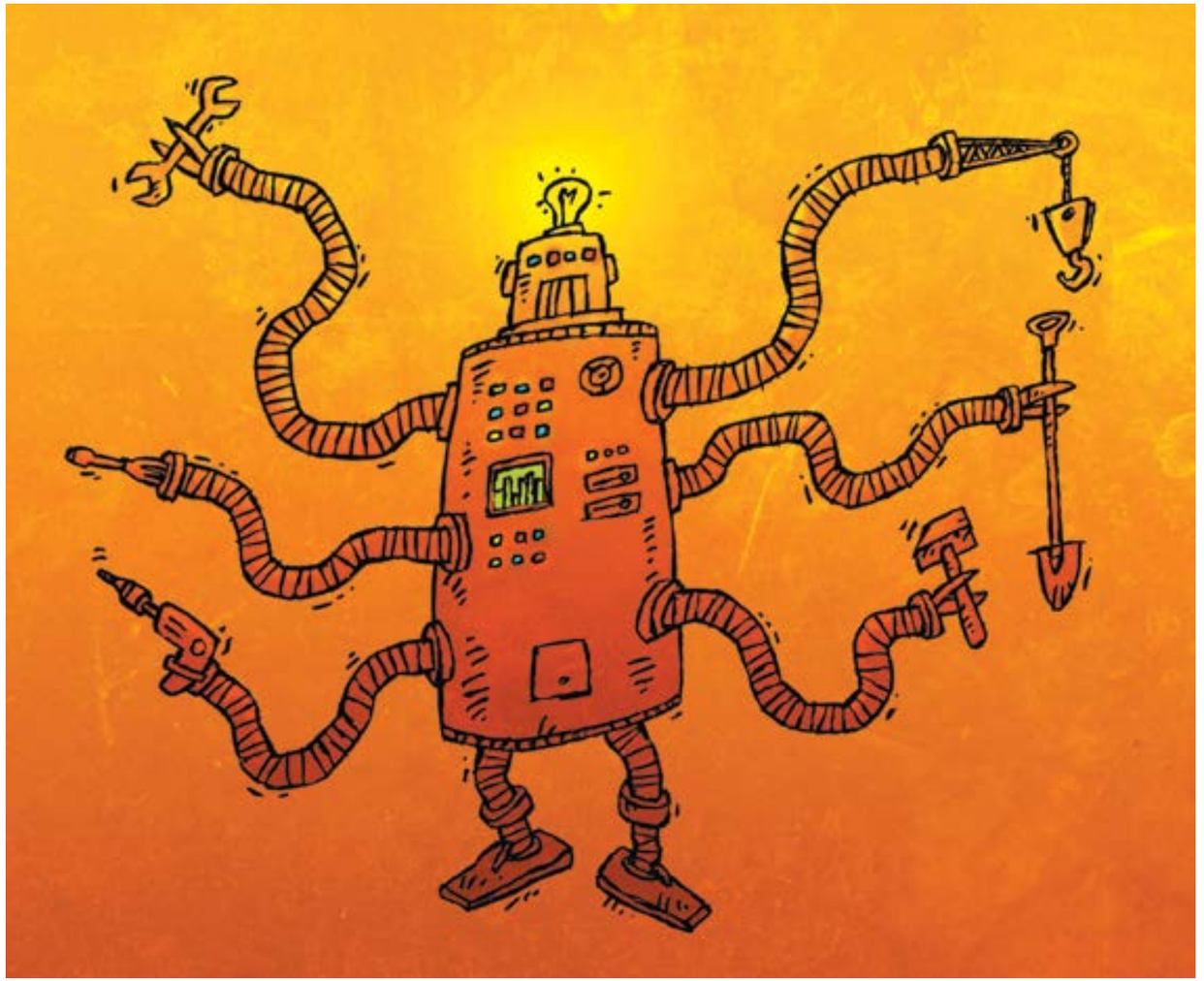
Стоит отметить, что, несмотря на развитие технологий хранения, магнитные носители по-прежнему предлагают лучшее соотношение цена/объем. Кроме того, они обладают высокой надежностью и выгодны в обслуживании. Согласно практическому наблюдению, современные технологии магнитных дисков предоставляют работоспособность примерно на 30 лет, если их хранить при 17 градусах по Фаренгейту и 30% относительной влажности.
По указанным причинам магнитные носители все еще популярны, особенно для хранения редко или никогда не используемых данных. Согласно исследованию InterSect 360 за 2013 год, 30% пользователей крупнейших высокопроизводительных установок используют магнитные хранилища для хранения архивных данных.
Впечатляющим примером может служить попадание черного ящика от Colubia Space Shuttle на дерево в Техасе вследствие катастрофы. Поиск занял несколько недель, в течение которых вода попала в устройство, но данные на магнитных дисках удалось восстановить.

RAID

Другой проблемой, связанной с организацией надежного хранилища, является устаревание технологии RAID, которая в настоящее время является узким местом в производительности дисковых массивов. Самым актуальным решением является так называемый jBOD. Под этим термином понимают дисковый массив, в котором единое логическое пространство распределено по дискам последовательно, или же когда каждый диск виден как отдельное устройство в операционной системе. jBOD все активнее проникают в HPC. Свежим примером может быть облако для научных данных, запущенное в октябре сего года суперкомпьютерным центром Сан Диего. Утверждается, что это самая большая академическая облачная система с производительностью 10 ТБ/с и объемом 5.5 ПБ.
Впрочем, есть и альтернативные подходы, например, компания PANASAS предлагает технологию RAID6+, которая базируется на троекратном копировании данных, и распределенный пофайловый RAID как часть файловой системы PanFS6.0.

Синергия применения суперкомпьютерных современных технологий производства

Текст Николай Семёнов, Сергей Серёжин, Кирилл Колганов, Александр Мурашов
Иллюстрация Владимир Камаев



На протяжении множества лет основной технологией производства изделий было удаление «лишнего» материала. Технология была актуальной до тех пор, пока изделия производились из недорогих материалов, а экономика стран не была выраженно рыночной. Современные технологии определяются актуальными государственными задачами, которые даже для промышленности подчас являются инновационными. Зачастую актуальные задачи государства также связаны с применением новых материалов с характерными свойствами. Такие материалы довольно дороги в цене, а некоторые выпускаются только под ряд конкретных изделий — по этой причине объекты, создаваемые по старой технологии, за счет утилизации материала при резании стоят чрезвычайно дорого или не могут быть созданы в принципе без применения новых подходов.

С освоением технологии литья, штамповки, сварки и т.п. появилась возможность за меньшее количество технологических операций приблизиться к спроектированной форме изделия.

На определенном этапе развития промышленности, несмотря на содержание небольшого объема лишнего материала в таких заготовках, эти технологии все же позволили осуществлять технические прорывы в серийном производстве в течение многих лет. Даже сейчас литье и штамповка занимают одну из основных технологических ниш промышленного производства несмотря на наличие значительных ограничений в отношении допустимой геометрии конечных изделий. Несмотря на достаточно высокий уровень актуальности этих технологий, эволюция производства продолжает стремиться к усовершенствованию способов получения конечных изделий. Сегодня открывается возможность совершить качественный промышленный скачок благодаря сочетанию передовых вычислительных и производственных технологий.

Аддитивные технологии

В 1984 году Чак Халл (Chuck Hull) из корпорации 3D Systems изобрел стереолитографию — процесс послойного отверждения жидких фотополимеров ультрафиолетовым облучением. В патенте 1986 года Халл дал такое определение: «система для построения трехмерных объектов с помощью создания образов поперечных срезов объекта». С тех пор было изобретено и модернизировано много других технологий, основывающихся на данном принципе. Возникло понятие «аддитивные технологии» (additive technology), означающее комплекс подходов к изготовлению изделия путем «добавления» (addition) материала, в отличие от традиционных технологий механической обработки, в основе которых лежит принцип удаления «лишнего» материала из заготовки. Также получил широкое распространение термин «3D-печать».

Если на заре применения аддитивных технологий трехмерная печать использовалась главным образом для изготовления функциональных и демонстрационных прототипов, то в настоящее время ведутся успешные экспериментальные работы по производству подобным способом уже непосредственно конечных изделий.

В настоящее время материалы, пригодные для применения в аддитивных технологиях, не ограничиваются фотополимерами. Сегодня существует промышленное оборудование для трехмерной печати из большого числа металлических сплавов и разнообразных прочных и жаростойких пластиков. Современные литейные формы могут быть отпечатаны из подготовленного песка, зерна которого обработаны специальной смолой. Применяемые материалы позволяют создавать даже конечные функционирующие системы, а с точки зрения возможности реализации конструктивных особенностей с помощью послойного создания объекта можно сформировать внутри него криволинейные полости практически произвольной формы, как имеющие выходы на внешнюю поверхность, так и обладающие полностью замкнутым объемом.

Известно, что корпорация «Боинг» производит некоторые воздуховоды сложной формы целиком с помощью трехмерной печати из пластика. Компания BAE Systems разработала и изготовила с помощью аддитивных технологий ряд деталей для истребителей Tornado, в том числе валы отбора мощности с двигателей. Самолеты Королевских ВВС Великобритании, оснащенные этими деталями, совершают полеты без каких-либо известных осложнений. Также в прессе сообщалось, что американская компания SpaceX установила отпечатанный на 3D-принтере главный клапан подачи окислителя в один из девяти двигателей Merlin 1D ракеты Falcon 9, успешно запущенной 6 января 2014 года. При этом изготовление детали новым способом заняло менее двух дней, тогда как типичное время производства с помощью литья составляло месяцы.

Аддитивные технологии имеют много преимуществ. Для создания изделия на 3D-принтере не требуется разработка дорогостоящих литейных форм, матриц и пуансонов или какой-либо другой оснастки — поэтому штучные изделия получаются дешевле и появляются на свет намного быстрее, чем при изготовлении традиционными способами. Эта же особенность 3D-принтеров делает их эффективным инструментом для изготовления прототипов изделия в процессе его разработки.

Изделие производится с высокой точностью — точность большинства современных промышленных 3D-принтеров находится на уровне нескольких десятков микрометров.

Другой важной возможностью, которую предоставляют некоторые из аддитивных методов, является создание изделий с градиентным переходом материала. В 2014 году НАСА изготовила опору для монтирования на спутнике зеркала с помощью послойного нанесения металлического порошка на вращающийся стержень, порошок спекался лазерным лучом. При этом состав подаваемого материала постепенно заменялся от оси создаваемой детали к ее краям. Такие методы позволяют создавать монолитные детали с вариацией механических, термических и магнитных свойств в разных частях изделия.

Аддитивные технологии, их инфраструктура и высокопроизводительные вычисления

Модернизация всей производственной цепи позволяет достичь той цели, которую ставит перед собой государство — инновационное развитие. Усовершенствование всех производственных этапов во всех рыночных отраслях при поддержке государства позволяет вывести на рынок множество высококачественных конкурентных продуктов и достичь замещения импортных аналогов. На практике для сокращения сроков достижения целей оказалось недостаточным развить один процесс из технологической цепочки. Например, за последние годы в конструкторские бюро РФ глубоко и эффективно проникли методики численного моделирования на этапе проектирования изделий. Актуальным становится расширение области внедрения высокопроизводительных вычислений на производстве. Сегодня наиболее передовые возможности современных производств и технологий уже всегда основываются на результатах ресурсоемких вычислений.

Сегодня первоочередной задачей суперкомпьютерных технологий в производстве является исполнение вспомогательной функции – осуществление упрощенного внедрения новых, в т. ч. аддитивных технологий, которые все чаще рассматриваются как перспективные. Даже на существующем уровне развития аддитивных технологий без инженерных расчетов и ускоряющих их параллельных вычислений обойтись невозможно. Если на двух- или трехосевом токарном

станке умелый мастер может выточить деталь, читая бумажный чертеж, то при использовании многофункционального четырех- и пятикоординатного станка, способного фрезеровать сложные криволинейные поверхности с высокой точностью и скоростью, без цифровой модели уже не обойтись.
Для аддитивного производства наличие виртуальных электронных моделей изделия является безусловной необходимостью. Например, модель изделия может быть построена конструктором с нуля в CAD-пакете. Однако все более важным становится подход, связанный с реверсинжинирингом, локальной оптимизацией геометрии или свойств изделия. Для оперативного и эффективного применения таких подходов исходный физический прототип изделия оцифровывается по технологии 3D-сканирования и, при необходимости, — фотограмметрии. В результате этого формируется специфическая электронная модель изделия — так называемое «облако точек», представляющее собой множество точек с конкретными пространственными координатами. Чем выше разрешение сканирования, тем точнее возможно передать геометрическую форму изделия, но тем больше оперативной памяти и вычислительных ядер требуется. Актуальные требования к оцифровке изделия уже не позволяют осуществить этот процесс на обычных мощных ПК.
«Облако точек» может быть импортировано в CAD/CAM/CAE-системы, и на его основе может быть построена «привычная» цифровая трехмерная модель для ее последующей оптимизации, инженерной доработки и взаимной адаптации модели и технологических процессов аддитивного и традиционного производства. Эти процедуры связаны с цепочкой расчетов, сложность которых быстро возрастает при увеличении размеров сканируемой детали, повышении сложности ее формы, а также увеличении разрешения сканирования, что в качестве инструмента расчета требует суперкомпьютер.
Сами по себе задачи инженерной оптимизации геометрии изделия также требуют существенных вычислительных ресурсов, особенно когда речь идет об объектах больших размеров или сложной формы. Также это верно для деталей, работающих в сложных условиях со значительными напряжениями, градиентами температур и т.п. — в этом случае в расчетах приходится применять подробные расчетные сетки, что приводит к быстрому росту необходимой для расчетов оперативной памяти и вычислительных мощностей.
Накопляемая со временем база знаний позволит перейти к разработке нового аддитивного оборудования с более высокими показателями точности, для чего необходимо решить ряд вычислительных научно-исследовательских задач, представляющих значительную сложность.
Во-первых, необходимо разработать методы получения еще более мелкозернистых порошков различных пластиков и металлических сплавов, при этом интерес представляют различные варианты формы зерна. Необходимо учитывать, что однородность формы и состава гранул порошка крайне важна для обеспечения качества конечного изделия. Более того, для достижения заявленного производителем уровня качества конкретное аддитивное оборудование настроено на строго определенные параметры исходного порошкового материала, и в большинстве случаев применение сырья с другими параметрами приводит к значительным искажениям производимых деталей или полной их непригодности к использованию.
Разработка более совершенных методов получения мелкодисперсных порошков различного состава, зерен различных форм сопряжена с обстоятельными исследованиями и численными экспериментами в области материаловедения. Однако только опытным путем получения изделий можно определить требования к формам и размеру зерен порошка материала.
Кроме развития методов промышленного производства сырья актуальными являются вопросы более совершенного управления свойствами производимых изделий, особенное значение они имеют для послойного производства композитных и градиентных материалов. Эти задачи требуют применения более точных и ресурсоемких физико-математических моделей, которые описывали бы процессы, происходящие при спекании в единое целое гранул порошков, в том числе представляющих собой многокомпонентные смеси.
Эти исследования тоже невозможно представить себе без значительных численных экспериментов.
В частности, в 2014 году Министерство энергетики США выделило грант компании AltaSim на разработку инженерного пакета для обеспечения аддитивного производства, который должен работать с применением высокопроизводительных вычислений.
В итоге в рамках современного производства присутствует необходимость использования вычислительных ресурсов на всех стадиях процесса. Вычислительные ресурсы требуются и для построения исходной цифровой модели по облаку точек, полученных при трехмерном сканировании исходного объекта.
Требуются они и для задач инженерной оптимизации, и для анализа соответствия между измерениями конечного изделия и его оптимизированной цифровой модели. Фактически для аддитивного производства необходимо внедрение и использование комплексного интегрированного решения, включающего целый ряд программно-аппаратных средств, интегрированных с поддерживающим аддитивные технологии промышленным оборудованием.

Рис. 1. Оценка эффективности применения HPC и аддитивных технологий

Реалии и перспективы взаимодействия производственных и вычислительных технологий в России

Аддитивные технологии и соответствующее оборудование развиваются в мире уже около 25 лет. В России только в последние годы серьезно обратили внимание на применение этих технологий.
Для достижения максимальной синергии вычислительных технологий и методов аддитивного производства необходима тесная интеграция в единый комплекс аппаратных вычислительных ресурсов, профильного программного обеспечения, производственных мощностей и ряда околопроизводственных компетенций. Для обозначения концепции подобных интегрированных структур в мире применяется понятие «цифровое производство».
Цифровое производство обладает рядом значительных преимуществ. Применение технологий, методов и инструментов численного моделирования, находящихся в едином комплексе с автоматизированным оборудованием аддитивного производства, облегчает оптимизацию технологических процессов и ускоряет выпуск готовых изделий (рис. 1).
Важным достоинством цифрового производства является возможность оказания полного спектра производственных и сопутствующих услуг, сопровождая процесс разработки и изготовления изделий от начала до конца. В этот спектр входит: трехмерное сканирование образцов; численное моделирование и инженерная многокритериальная оптимизация на суперкомпьютере; трехмерная печать песчаных литейных форм; изготовление конечных изделий любой сложности из пластика, металлических сплавов и композитных материалов с помощью аддитивных технологий; производство отливок и заготовок с помощью высокоточного литья; неразрушающий контроль качества геометрии и внутренней структуры продукции с помощью трехмерных сканеров и томографов; разработка сопроводительной технической документации на продукцию и сертификация изделий в соответствии с требованиями нормативных документов (рис. 2).

Рис. 2. Пример сокращенной технологической цепочки промышленного производства изделия по аддитивным технологиям

Ожидания от применения описанных комплексных подходов в мире довольно высоки. Доступные к настоящему моменту сведения выглядят весьма многообещающими. Например, результаты проведенного компанией CIMdata исследования говорят о том, что компаниям, перестроившим работу в соответствии с принципами цифрового производства, удалось снизить время вывода продукта на рынок на 30%, повысить производительность на 15%, снизить себестоимость продукции на 13%, сократив затраты на оборудование на 40%. При этом дается оценка, что срок возврата инвестиций составил в среднем несколько месяцев, и выгоды от перехода на новые производственные технологии в годовом исчислении превысили сделанные инвестиции в пять—десять раз.

Первые точки роста и распространения цифровых производственных технологий в России уже начинают появляться. В Уфе функционирует центр, который в том числе занимается аддитивным производством непосредственно промышленных деталей, применяя в качестве технологии 3D-печать из металлических порошков. В настоящее время ООО «Проектно-инжиниринговая компания» создает Центр цифровых технологий в Казани. Проект реализуется при активном содействии представителей Министерства экономики и развития РФ, Министерства экономики Республики Татарстан, Министерства промышленности Республики Татарстан. С первого квартала 2015 года запланировано оказание части спектра услуг цифрового производства, а на последующих этапах развития Центра планируется постепенный выход на оказание полного спектра данных услуг. Известно, что на начальных этапах развития Центра основными потребителями создаваемой конечной продукции станут предприятия таких отраслей промышленности, как авиастроение, вертолетостроение, судостроение, космическая отрасль, автомобилестроение, двигателестроение и приборостроение. Однако, учитывая мировой опыт гораздо более широкого отраслевого распространения аддитивных производственных технологий, в будущем можно предполагать активное участие Центра в распространении инноваций и в других отраслях отечественной промышленности. Бытует мнение, что сочетание аддитивных и вычислительных технологий всерьез и надолго займет существенную нишу в отечественной промышленности и послужит существенным рычагом дальнейшей трансформации и интенсивного развития российской промышленности.

Решение связанной задачи моделирования взрыва бытового газа в жилом кирпичном здании и оценки его несущей способности с использованием программных комплексов ANSYS и FLOW VISION

Текст Г. Г. Кашеварова, А. А. Пепеляев

Аварии зданий, вызванные взрывами бытового газа, происходят регулярно. В основном проблема затрагивает газифицированные здания. Основной причиной аварий является человеческий фактор, исключить влияние которого практически невозможно. Такие ситуации, как несанкционированное подключение к системе газоснабжения, халатность при пользовании газом и газовыми приборами в бытовых нуждах, не представляется возможным контролировать или регулировать их предотвращение, т.е. исходить нужно именно из этого.

Впервые проблема защиты зданий от внутренних взрывов была рассмотрена после аварии в здании Ронан Пойнт в Лондоне в 1968 году. Данную проблему необходимо рассматривать совместно с проблемой прогрессирующего разрушения зданий или с проблемой обеспечения их механической безопасности. Термин «прогрессирующее» (лавинообразное) разрушение определяется как последовательное разрушение несущих строительных конструкций и основания, приводящее к обрушению всего сооружения или его частей при локальном повреждении. На сегодняшний день проблема расчета на прогрессирующее разрушение сформулирована следующим образом: конструктивная схема здания должна обеспечивать его прочность и устойчивость в случае локального разрушения несущих конструкций как минимум на время, необходимое для эвакуации людей.

Действующий закон РФ требует обеспечение механической безопасности зданий и сооружений и предписывает учитывать возможные аварийные воздействия, такие как взрывы или пожары, при проектировании зданий для предотвращения их прогрессирующего разрушения. Также даны предписания непосредственно по виду расчетной схемы сооружений: «Для расчета зданий против прогрессирующего обрушения следует использовать пространственную расчетную модель, которая может учитывать элементы, являющиеся при обычных эксплуатационных условиях ненесущими, а при наличии локальных воздействий активно участвуют в перераспределении нагрузки».

На сегодняшний день на российском рынке и в других странах существует большое количество нормативной методической и научной литературы, в которой содержатся рекомендации и методики, позволяющие производить расчеты взрывов опасных веществ и оценивать воздействие взрыва бытового газа на здания и сооружения. Все они различны не только в расчетных показателях избыточного давления, но и в оценке воздействия поражающих факторов на конструкции и здания, т. е. данные вопросы требуют дальнейшего серьезного изучения.

Этими вопросами занимались и занимаются многие исследователи и ученые (А. В. Мишуев, Б. С. Расторгуев, В. М. Ройтман, В. О. Алмазов, Я. Б. Зельдович, Ю. И. Стекольников, В. И. Травуш, А. А. Комаров, В. А. Котляревский и др.).

При авариях бытового газа внутри помещений зданий возникает дефлаграционный взрыв – быстрое горение газовоздушной смеси, в которой концентрация горючего находится между нижним и верхним концентрационными пределами воспламенения (5–15% для метано-воздушной смеси). Реакция протекает при дозвуковых скоростях. Дефлаграцию газовоздушной смеси часто путают с детонационным взрывом. В действительности, что доказано экспериментально и теоретически, дефлаграция значительно отличается от процесса детонации как по скорости протекания реакций, так и по величине избыточного давления (при детона-

ции выше на 2 и более порядков). Процесс дефлаграции тем не менее при определенных условиях, необходимых для интенсификации турбулизации смеси, может переходить в стадию детонации. Геометрические характеристики помещения, а именно соотношение длины, ширины и высоты 10:1 и больше могут оказывать значительное влияние на процессы турбулизации горения при взрыве, а следовательно, на величину избыточного давления. Наличие смежных помещений, наличие преград на пути распространения фронта пламени также оказывают влияние на протекание реакции и на возможность формирования мощных воздушных потоков в межквартирных и межкомнатных проходах, коридорах и т.д. Именно эти потоки (а не ударные волны, как это часто трактуется, особенно в прессе) приводят к выбросу фрагментов строительных конструкций и предметов из аварийного помещения. Следует иметь в виду, что разрушение конструкций происходит под действием избыточного давления, а последующий их выброс происходит под действием скоростного напора.

Данная проблема обычно рассматривается только в свете нового проектирования, тогда как это представляется наиболее актуальным для зданий, уже находящихся в эксплуатации, а именно:

- *при текущем определении технического состояния строительных конструкций газифицированных объектов, в том числе и после аварий;*
- *при прогнозировании ущерба от последствий внутренних взрывов;*
- *при восстановлении поврежденных или разрушенных зданий, а также их отдельных конструкций.*

А для этого требуются более точные расчетные модели и методы их расчета, а именно численное моделирование и использование современных мощных программных комплексов.

Рис. 1. Граничные условия модели во FlowVision

Целью данной работы является комплексное исследование проблемы механической безопасности кирпичных жилых зданий на действие дефлаграционного взрыва бытового газа, включающее разработку, программную реализацию и верификацию методик расчета взрывной нагрузки и воздействия этой нагрузки на конструкции здания на основе численного решения трехмерных задач гидрогазодинамики и механики деформируемого твердого тела.

Для решения этой проблемы нами была разработана и апробирована вычислительная технология, представляющая собой последовательность решения информационно связанных задач, которая включает 3 этапа:

1. *Моделирование помещения, ограниченного внешними и внутренними стенами, а также перекрытиями, в котором происходит взрыв газа, для чего используется программный комплекс ANSYS (или SolidWorks).*

2. *Экспорт модели помещения в программный комплекс FlowVision и проведение газодинамического расчета. Определяется величина избыточного давления на стенки модели (конструкции здания) во времени.*

3. *Нагрузки с модели FV экспортируются в конечно-элементный комплекс ANSYS, строится КЭ модель здания, выполняется расчет напряженно-деформированного состояния и характера повреждений конструкций здания.*

1. Газодинамический расчет интенсивности взрывной нагрузки в программном комплексе FlowVision

Тенденцией развития ведущих программных комплексов является реализация в каждом из них набора математических моделей (ММ), позволяющих как можно более полно моделировать все встречающиеся на практике физические эффекты. Пользователь подключает нужные модели на стадии постановки задачи, задавая затем соответствующие начальные и граничные условия и требуемые исходные данные для каждой конкретной задачи.

Процесс взрыва бытового газа моделируется в программном комплексе FlowVision (FV), предназначенном для моделирования трехмерных течений жидкости и газа в технических и природных объектах и визуализации этих течений методами компьютерной графики. Также FV позволяет решать гидрогазодинамические задачи и сопряженные задачи взаимодействия потока с деформируемым телом совместно с конечно-элементными программами.

```
ELEM,6517,SIDE,1,VALS,15385,76
ELEM,6519,SIDE,6,VALS,0
ELEM,6520,SIDE,1,VALS,15386,03
ELEM,6522,SIDE,6,VALS,0
ELEM,6523,SIDE,1,VALS,15382,05
ELEM,6525,SIDE,6,VALS,0
ELEM,6526,SIDE,1,VALS,15382,19
ELEM,6528,SIDE,6,VALS,0
ELEM,6529,SIDE,1,VALS,15382,38
ELEM,6531,SIDE,6,VALS,0
ELEM,6532,SIDE,1,VALS,15382,53
ELEM,6534,SIDE,6,VALS,0
ELEM,6535,SIDE,1,VALS,15382,78
```

Рис. 2. Файл, содержащий информацию о нагрузках на фасетках модели в FV

```
SFE,6517,1,PRES,15385,76
SFE,6519,6,PRES,0
SFE,6520,1,PRES,15386,03
SFE,6522,6,PRES,0
SFE,6523,1,PRES,15382,05
SFE,6525,6,PRES,0
SFE,6526,1,PRES,15382,19
SFE,6528,6,PRES,0
SFE,6529,1,PRES,15382,38
SFE,6531,6,PRES,0
SFE,6532,1,PRES,15382,53
SFE,6534,6,PRES,0
SFE,6535,1,PRES,15382,78
```

Рис. 3. Файл, содержащий информацию о нагрузках с командами SFE программы

Математическая модель дефлаграционного взрыва базируется на модели горения Зельдовича, описывающей процессы горения газовых смесей при дозвуковых числах Маха и на уравнениях модели слабосжимаемой жидкости. Базовые уравнения гидрогазодинамики: Навье-Стокса, энергии, состояния, уравнения для скалярных величин, описывающие концентрацию топлива, окислителя, продуктов сгорания, нейтрального газа, оксидов азота и маркера. Эти нелинейные уравнения замыкаются уравнениями стандартной модели турбулентности.

В качестве начальных условий задаются плотности газов, начальная температура, пульсация, стехиометрические коэффициенты при горении бытового газа в воздухе, кинетические константы, определяющие скорость реакции для горения.

Граничные условия схематично изображены на рис. 1.

Для внутренних точек расчетной области задаются уравнения, описывающие модель горения Зельдовича – простейшая модель горения, в которой постулируется бесконечная скорость брутто-реакции (интенсивность горения). Это означает, что топливо и окислитель не могут сосуществовать в одной точке (ячейке).

В общем случае нелинейные уравнения, описывающие гидрогазодинамическую задачу, не имеют аналитического решения, решать их приходится численно, находя вместо непрерывного решения дискретный набор значений в определенных точках пространства и для определенных моментов времени. Для численной реализации задачи используется метод конечных объемов (МКО), в котором используется подход Эйлера, т.е. рассматривается течение в выделенной области пространства, например, в помещении кухни. Согласно МКО при дискретизации пространства расчетной области расчетная сетка может быть любой (структурированной или неструктурированной).

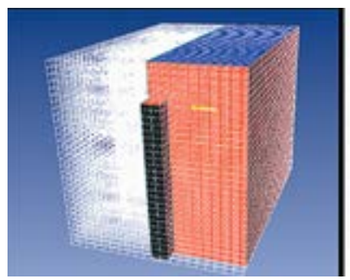


Рис. 4. Модель кухни здания во FlowVision

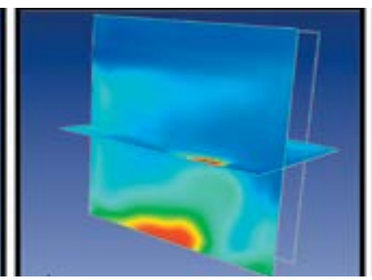


Рис. 5. Расчет холодного течения в пространстве кухни (отображается концентрация газов)

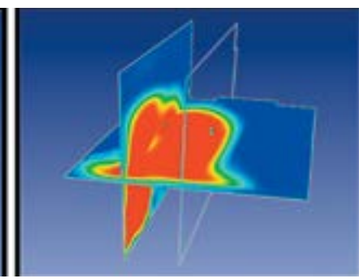


Рис. 6. Горение смеси газов (отображается распределение давления по объему)

Дискретное решение задачи может быть получено как в узлах расчетной сетки, так и в ячейках расчетной сетки. Для определенности рассмотрим типовое многоэтажное газифицированное кирпичное здание, какими застроены многие микрорайоны российских городов, в помещении кухни которого гипотетически может произойти взрыв бытового газа. Помещение кухни в типовых зданиях советского периода постройки имеет сравнительно небольшой объем (в среднем от 15 до 20 м^3) и сообщается со смежным помещением (коридором) через дверной проем. Осредненные габариты внутреннего пространства кухни можно представить в виде параллелепипеда размерами в плане 2х3 м и высотой 2,8 м. Помещение изначально заполнено воздухом.

Исходные параметры газовоздушной среды можно принять: стехеометрический коэффициент при горении смеси метан+воздух 17.24, смеси пропан-бутан+воздух 15.67; температура воспламенения: метан+воздух 923°К, пропан-бутан+воздух: 750°К; плотности газов: метан 0.71 кг/м^3, пропан 0.585 кг/м^3, воздух комнатный 1.225 кг/м^3; теплопроводность: метан 0.026 (Вт/м°С), воздух 0.0024-0,03 (Вт/м°С).

Воздухообмен помещения осуществляется посредством вытяжной вентиляции (вытяжка составляет 24 м^3/ч), а также обеспечивается воздухопроницаемостью заполнения дверных и оконных проемов. Для окон с деревянными переплетами воздухопроницаемость составляет в среднем 6 кг/ч•м^2, для деревянных внутриквартирных дверей – 15 кг/ч•м^2. Подразумевается, что образование газового облака происходит через неисправные конфорки кухонной плиты. В качестве бытового газа выбран метан плотностью 0.71 кг/м^3. Известно, что расход через конфорки при неплотностях в системе газоснабжения составляет ~ 0.19 м^3/ч. Формирование взрывоопасного облака при действующих 4 конфорках происходит в среднем за 2,5 часа (при расходе газа 0,19 ^{3}/ч).

Для решения задачи строится расчетная область – трехмерная геометрическая модель помещения, в которой происходит взрыв. Для построения модели можно использовать программу Solid Works или ANSYS. Мы будем решать связанные задачи и используем для этой цели программный комплекс ANSYS. Формат, в котором данные экспортируются из ANSYS – «*.cdb», несет в себе всю необходимую информацию о структуре расчетной конечно-элементной сетки.

На полученной модели решается задача газодинамики – моделируется взрыв (горение метана/пропана в воздухе). Отработана процедура обмена данными между программами ANSYS и FlowVision. Величины давления сохраняются на фасетках модели помещения (рис. 2) и экспортируются в ANSYS в виде текстового файла, содержащего команды ANSYS –SFE (рис. 3). Запись в файл происходит построчно, для каждой конечно-элементной ячейки, имеющей фасетки, выходящие вовнутрь модели. В каждой строке содержится информация об элементе, стороне, на которой приложена нагрузка, и величине самой нагрузки. Дальнейший импорт нагрузок в ANSYS из приведенного файла происходит после дополнения строк в файле необходимой служебной информацией. Команды читаются в ANSYS, и соответствующие нагрузки прикладываются к твердотельной модели помещения для проведения конечно-элементного расчета. Принцип подобной связки программных комплексов описан в работе И. Э. Лукьяновой и В. В. Шмелева. В отличие от указанных работ нами рассмотрены модели с разными типами жесткости и для более сложной модели горения, реализованной в программе FlowVision.

Изменение независимых физических параметров с течением времени в выделенной области пространства кухни (рис. 4) определяется физическими потоками (конвективными и диффузионными), проходящими внутрь этой области через ее поверхность, а также источниками (объемными и поверхностными), находящимися внутри этой области. Расчет делится на два этапа:

1. Расчет холодного течения (смешение газов без горения). В заполненную воздухом расчетную область поступает газ, происходит процесс смешивания газов (рис. 5); Для расчета холодного течения используется модель слабосжимаемой жидкости.

2. При достижении определенной концентрации газа в помещении производится инициация горения – поджиг (от какого-нибудь электрического прибора, например, реле холодильника) и горение смеси газов (рис. 6).

В результате расчета определяется величина максимального избыточного давления на стенках модели кухни.

При дефлаграционном взрыве реализуется принцип квазистатичности избыточного давления, который заключается в независимости взрывной нагрузки от пространственной координаты. Другими словами, давление, действующее в рассматриваемый момент времени на любой конструктивный элемент ограждения (стены, потолок, пол, окна, двери и т.д.), одинаково во всех точках помещения.

1.1. Верификация численной методики расчета интенсивности взрывной нагрузки

Применяя теорию, заложенную в современных программных комплексах, надо быть уверенным, что она корректно описывает исследуемый физический процесс. Физические аспекты дефлаграционного взрыва газа экспериментально исследованы и описаны в работах А.А. Комарова. На рис. 7 приведены

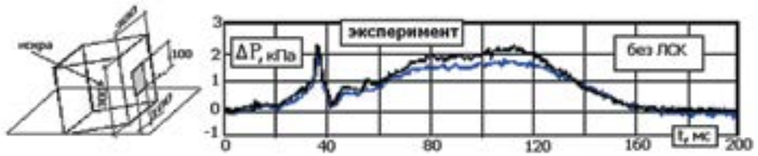


Рис. 7. Натурные эксперименты А.А. Комарова

Рис. 8. Вычислительные эксперименты с моделью FlowVision

графики изменения избыточного давления по времени при дефлаграционном взрыве в замкнутом объеме, имеющем выход в виде окна, полученные им экспериментально.

Для верификации нашей методики мы использовали этот натурный эксперимент и получили адекватные результаты (рис. 9), которые позволили нам далее решать задачи по исследованию влияния взрыва бытового газа на механическую прочность конструкций или оценку возможного ущерба для конкретного здания.

1.2. Исследование влияния различных факторов на величину избыточного давления в помещении

Представленная методика газодинамического расчета позволяет исследовать влияние различных факторов на величину избыточного давления в помещении при взрыве.

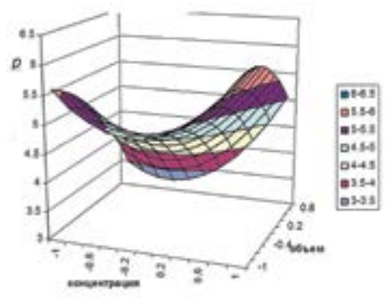


Рис. 9. Зависимость избыточного давления от объема помещения и концентрации газа

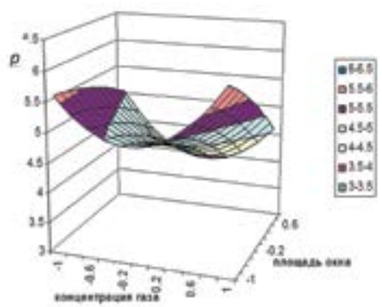


Рис. 10. Зависимость избыточного давления от площади окна и концентрации газа

Дефлаграционные взрывы отличаются многообразием проявлений, поскольку различны планировки зданий, размеры, количество и расположение проемов, через которые происходит истечение сначала исходной смеси, а затем и продуктов взрыва, качество остекления или наличие легко сбрасываемых конструкций.

Используя метод планирования многофакторного эксперимента, мы получили зависимости избыточного давления p при взрыве бытового газа (метановоздушной смеси) от факторов: объема помещения (x_1), площади оконных (дверных) проемов (x_2), концентрации газа в смеси (x_3).

В основу планирования вычислительного эксперимента положено ортогональное планирование на трех уровнях по каждому из факторов (план Хартли). На рис. 9 и 10 приведены графики зависимости избыточного давления (функции отклика) от факторов.

Анализируя полученные результаты, можно отметить следующее. Наиболее значимым фактором, влияющим на величину избыточного давления, является концентрация газа в газовоздушной смеси. Чем она больше (в пределах воспламеняемости смеси), тем выше максимум давления. В меньшей степени на величину давления влияют объем помещения и площадь сбросного проема. Чем больше объем помещения, тем больше давление. Увеличение размера окна, наоборот, снижает давление, но не оказывает влияния на первый максимум давления до момента разрушения окна.

Величина избыточного давления для любого момента времени определяется темпом роста давления, вызванного выделением продуктов сгорания на фронте пламени и темпом снижения давления вследствие истечения газа через открытый проем. Если сбросной проем остеклен, то он в процессе взрывного горения вскрывается.

В этот момент возникает локальный по времени максимум давления, затем наблюдается спад, после чего давление начинает расти, пока не выгорит вся газовоздушная смесь.
Следует отметить тот факт, что величина максимального давления в зданиях с глухим остеклением зависит от давления начала разрушения остекления, которое, в свою очередь, зависит от размеров стекла и его толщины.
В замкнутом объеме избыточное давление при внутреннем дефлаграционном взрыве может достигнуть 700…900 кПа, но благодаря наличию в зданиях предохранительных конструкций (ПК), таких как остекленные оконные проемы или другие легкосбрасываемые конструкции, уровень давления удается в некоторых случаях снизить до безопасного уровня (2 – 5 кПа). Это давление характеризует нетравмоопасное повреждение человека.
Полученные результаты могут быть использованы при прогнозировании ущерба от последствий возможных внутренних взрывов проектируемых и существующих жилых зданий, что в конечном итоге позволит разрабатывать мероприятия для исключения наиболее опасного аварийного сценария и уменьшения вероятности возникновения взрыва.

Рис. 11. Взрыв в жилом кирпичном доме с указанием расположения помещения кухни в г. Губаха Пермского края

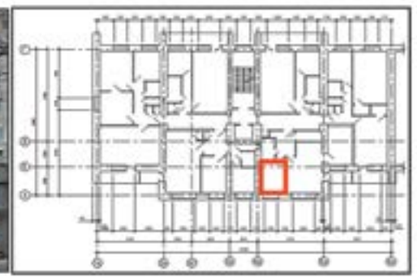


Рис. 12 План третьего этажа

2. Решение связанной задачи – ретроспективный анализ реальной аварийной ситуации взрыва бытового газа в жилом здании

Разработанная вычислительная технология была опробована при моделировании реальной аварии, произошедшей в 2006 году в г. Губаха Пермского края в квартире на 3-м этаже жилого 9-этажного здания с несущими кирпичными стенами (рис. 11, 12). Кирпичные стены в большей степени подвержены разрушению горизонтальными динамическими нагрузками, чем панели или каркасные здания.
Исходные данные для расчета.
Материалы: стены– керамический кирпич М100 на растворе М50; модуль упругости E = 75E7 Па, коэффициент Пауссона v = 0,25, плотность ρ = 1800 кг/м^3; перекрытия– плита ж/б сборная, класс бетона В15, модуль упругости E = 2.72E10 Па, коэффициент Пауссона v = 0,2, плотность ρ = 2500 кг/м^3. Нагрузки статические: на плиту: 30 кН/м^2 + собственный вес плиты; на стену 1: 6000 кН/м^2; на стену 2: 9000 кН/м^2; на стену 3: 1000 кН/м^2.
Исследование механической безопасности кирпичного здания при взрыве бытового газа началось с рассмотрения фрагмента здания – помещения кухни, где этот взрыв вероятнее всего и мог произойти. Чтобы правильно учесть конструктивное решение здания, помещение кухни моделировалось совместно со смежным помещением, имеющим с кухней общую плиту перекрытия. Влияние остальной части здания учитывалось закреплениями и приложением постоянной нагрузки от вышерасположенных конструкций (рассматривался 3-й этаж).
На рис. 13 показаны конечно-элементная модель кухни, построенная в ANSYS, и модель для газодинамического расчета в программном комплексе FV. В результате газодинамического расчета определена величина максимального давления на стенках модели кухни, которое в данном случае получилось равным 6 кПа.
Для решения комплексной задачи гидрогазодинамики и прочностного анализа в программных комплексах FlowVision и ANSYS был разработан алгоритм и исследовательский программный модуль связки ANSYS – FlowVision.

2.1. Расчет напряженно-деформированного состояния и разрушения конструктивных элементов здания в ПК ANSYS

Математическая модель расчета напряженно-деформированного состояния конструктивных элементов и здания в целом представляет собой краевую задачу механики деформируемого твердого тела, включающую: уравнения движения, геометрические уравнения Коши, физические соотношения, устанавливающие связь между тензорами напряжений и деформаций, конкретный вид которых зависит от физико-механических свойств материалов конструкций здания.
Для описания нелинейных эффектов материалов в определяющих

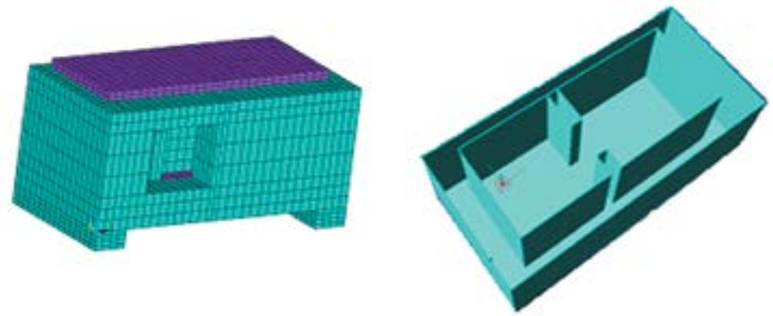


Рис. 13. Расчетные модели помещения кухни в ANSYS и FlowVision

соотношениях использована модель разрушения бетона Вильямса и Ранке и обобщенная модель упруго-хрупкого разрушения ортотропного материала кирпичной кладки, разработанная Г. Г. Кашеваровой.
Граничные условия зависят от условий закрепления и нагружения конкретной модели. Они могут быть смешанного типа: на части границы тела по некоторым направлениям могут задаваться поверхностные нагрузки P, а на некоторых частях – перемещения U. При решении задач динамики кроме статических граничных условий вводятся и динамические граничные условия. При этом динамическая нагрузка определяется в результате расчета во FlowVision или может быть представлена в виде импульсного сигнала. Кроме того, исследовалась возможность замены динамического воздействия на эквивалентную статическую нагрузку, рекомендуемую нормативными документами.
Для решения краевой задачи использовался метод конечных элементов (МКЭ), и программный комплекс ANSYS, позволяющий выполнять полноценный статический и динамический анализ широкого круга технических задач. Решение на динамическое действие нагрузки получено с использованием неявной схемы интегрирования разрешающего уравнения движения. Для оценки времени воздействия динамической нагрузки на конструкцию выполнялся модальный анализ расчетной модели.
Оценка механической безопасности конструкций кирпичного здания включала реализацию следующей последовательности задач:
1. Построение грубой конечно-элементной модели здания (рис. 14) и расчет НДС на действие статических ветровых и распределенных нагрузок, определенных в техническом задании. Для построения КЭ модели использовался трехмерный 20-узловой элемент SOLID95 с нелинейной квадратичной аппроксимацией.
2. Проблема большой размерности сформированных конечно-элементных моделей решалась с помощью метода субмоделирования. Вблизи от места взрыва выделялась часть здания, на границах которой при обследовании следов разрушения не обнаружено, наносилась более мелкая сетка конечных элементов (рис. 15) и выполнялся расчет на действие статических ветровых и распределенных нагрузок с учетом физической нелинейности материалов несущих конструкций. При этом использовались кинематические граничные условия, полученные при расчете грубой модели. Нелинейный расчет поводился МКЭ с использованием пошаговой процедуры. На каждом шаге решения для получения сходимости выполнялись равновесные итерации с помощью метода Ньютона–Рафсона.
3. Расчет фрагмента здания, где произошел взрыв бытового газа, с использованием уточненной модели на еще более мелкой сетке (рис. 15). Расчет

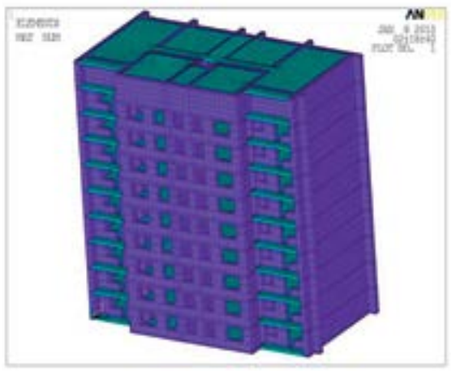


Рис. 14. Полная модель здания

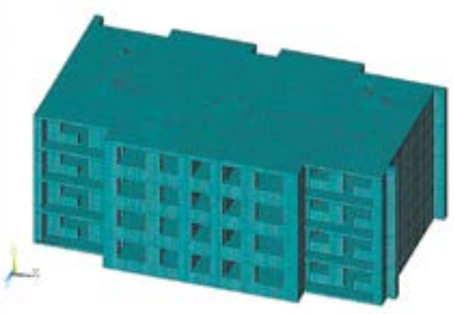


Рис. 15. Выделенная часть здания

Рис. 16. Фрагмент здания

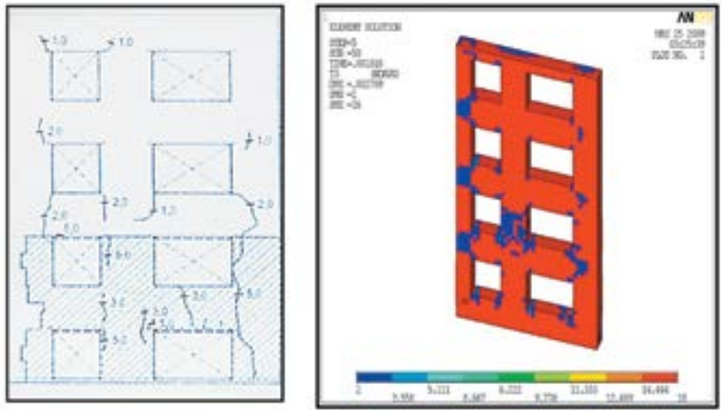


Рис. 17. Сравнение расчетной картины трещин с фактическими повреждениями при взрыве

проводился для двух вариантов нагрузки: эквивалентной статической с учетом нормативного коэффициента динамичности 1,8 и динамической. В первом случае выполнялся нелинейный статический анализ при независимом от времени поведении материала. Во втором – запускался нелинейный анализ переходных процессов, активизировались эффекты интегрирования по времени («время» представляет фактическую хронологию). Кроме того, процесс разрушения конструктивных элементов здания рассматривался с использованием разных конструктивных схем помещений кухни (наличие открытой/закрытой двери в смежное помещение, жесткая заделка или опирание перекрытий).

В результате решения определено напряженно-деформированное состояние несущих конструкций здания с учетом структурного разрушения, которое показало, что трещины появились только в конструкциях помещения кухни, где произошел взрыв, и в помещениях, непосредственно примыкающих к нему.

Схемы распространения трещин в несущих стенах здания качественно повторяют фактические, снятые на месте аварии (рис. 16). Это позволяет говорить о достоверности математических моделей и предложенной вычислительной технологии, которую в дальнейшем можно использовать для экспериментов с различными параметрами, влияющими на силу и место взрыва.

Кроме того, выполнена количественная оценка степени повреждения конструкции стены, которая определялась как отношение объема разрушенных конечных элементов к начальному объему. При этом исследовалось влияние прочностных характеристик материала и избыточного давления взрывной нагрузки на степень разрушения с использованием математической теории планирования эксперимента.

На рис. 18 изображена поверхность функции отклика, позволяющая оценить влияние прочности материала стены здания (z_1) и избыточного давления при взрыве (z_2) на степень разрушения конструкции.

Анализируя полученные результаты, можно отметить, что зависимость степени разрушения конструкции от уровня взрывной нагрузки имеет вид выпуклой унимодальной функции, и при использовании более прочного материала кирпичной кладки процент разрушения конструкции снижается.

Просуммируем результаты нашего исследования.

Во-первых, была разработана и верифицирована вычислительная технология проведения комплексного анализа действия дефлаграционного взрыва газа на механическую безопасность здания. Для решения связанной задачи гидрогазодинамики и прочностного анализа в программных комплексах FlowVision и ANSYS разработан алгоритм и исследовательский программный модуль связки ANSYS – FlowVision, позволяющий автоматизировать процесс обмена информацией.

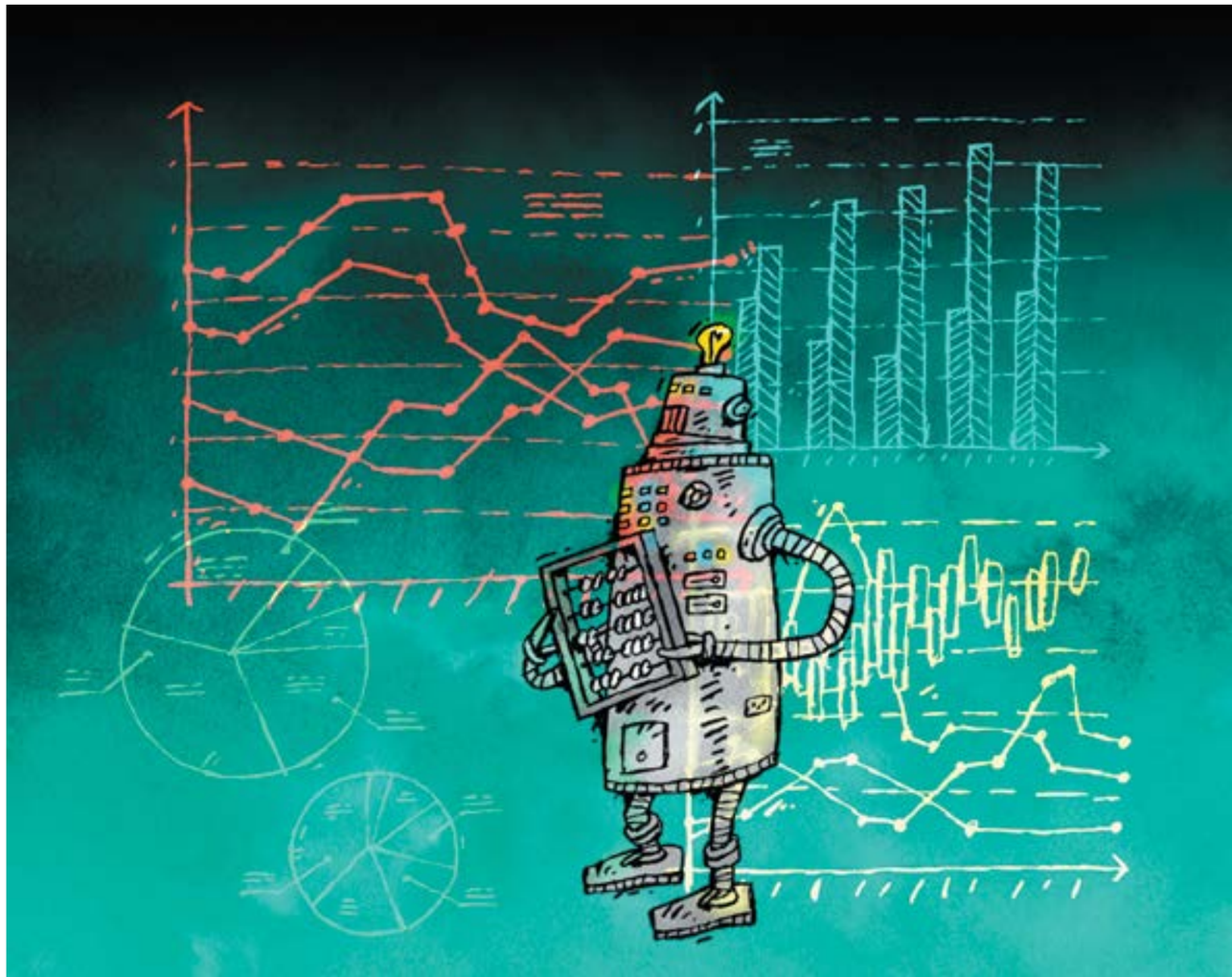
Во-вторых, с помощью разработанной методики расчета интенсивности взрывной нагрузки и теории планирования многофакторного эксперимента установлены зависимости избыточного давления при взрыве бытового газа от объема помещения, концентрации газа в смеси, размеров оконных (дверных) проемов. Наиболее значимым фактором, влияющим на рост давления, является концентрация газа в смеси. Величина максимального избыточного давления при взрыве в основном зависит от давления начала разрушения остекления и характерных размеров помещения.

В-третьих, при изучении воздействия дефлаграционного взрыва на несущие конструкции здания выявлены параметры, оказывающие наибольшее влияние на механическую безопасность здания. Наличие оконных и дверных проемов позволяет снизить напряжения в конструкциях, причем чем больше общая площадь проемов, тем это снижение значительнее. Жесткая заделка перекрытий также уменьшает напряжения в элементах конструкций (~ в 2–5 раз).

И в-четвертых, расчеты на динамическое воздействие и на эквивалентную статическую нагрузку дают качественно и количественно отличающиеся результаты. При динамическом анализе процесс изменения напряжений развивается во времени и превышает значения напряжений в статике в среднем в 2 раза (при открытой двери) и ~ на 28% – при закрытой двери помещения, где произошел взрыв.

Квантовая информатика для экономики и финансов

Текст Александра Климова
Иллюстрация Владимир Камаев



Квантовая информатика – молодая область науки, возникшая на стыке физики, информатики и математики, – сегодня стремительно развивается и широко проникает в самые разные сферы человеческой деятельности. Специалисты международной научной лаборатории «Современные коммуникационные технологии и их приложения в экономике и финансах – Fin Q» Университета ИТМО работают над актуальным проектом, цель которого – реализовать огромный потенциал квантовой информатики в решении экономических проблем, и прежде всего – в эффективном расчете сложных многоуровневых экономических и финансовых моделей, создание которых не под силу даже современным суперкомпьютерам. Лаборатория собрала вместе ученых мирового уровня из разных стран, крупнейших специалистов в фундаментальных и прикладных науках, совокупные усилия которых направлены на решение общих задач нового исследовательского объединения.

Квантовая информатика как новая наука

Колоссальный рост объема информации – определяющая черта современной эпохи и ее главный вызов научной и технологической мысли. Квантовая информатика находится на переднем крае поиска новых методов обработки, использования и хранения информации. Эта молодая наука, опирающаяся на фундаментальные свойства квантовых систем, оперирует квантовыми объектами информации, единица которых, по аналогии с понятием «бит», получила название «кубит». Кубиты обладают рядом замечательных свойств, которые обеспечивают им огромные преимущества по сравнению с битами и позволяют на их основе конструировать вычислительные устройства с высоким быстродействием.

Разрабатываемые сегодня квантовые компьютеры будут способны многократно превзойти скорость обработки данных даже самых современных суперкомпьютеров. Таким образом, открываются принципиально новые возможности для компьютерного моделирования сверхсложных процессов в природе и социальной жизни, что ранее представлялось маловероятным ввиду громоздкости объема вычислений и колоссального количества параметров.

Одной из первых заявила о создании квантового компьютера канадская компания D-Wave Systems. В 2011 году разработанный ее коллективом компьютер на 128-кубитном процессоре приобрела крупнейшая американская корпорация Lockheed Martin Corporation, специализирующаяся в области авиастроения, авиакосмической техники, судостроения, автоматизации почтовых служб и аэропортовой логистики.

Сегодня разработки в области квантовой информатики ведутся крупнейшими корпорациями и научными центрами, такими как IBM, Mitsubishi, Toshiba, University of California Berkeley, California Institute of Technology и др.

В России также проводятся исследования квантовых систем, например, в Санкт-Петербургском государственном университете на физическом и математико-механическом факультетах, а квантовая информатика как наука активно входит в учебные планы университетов.

Квантовая информатика в экономике и финансах

Экономика и финансы выступают как одна из наиболее перспективных и органичных сфер, где методы квантовой информатики могут применяться. Объясняется это тем, что сегодня практические задачи в экономической жизни как никогда ранее требуют использования развитых математических методов и продвинутых компьютерных технологий. В то же время растет конкуренция на рынке эффективных финансовых инструментов и стратегий торгов, и квантовая информатика с ее огромным потенциалом обработки сложнейших комплексов данных становится все более востребованной.

Так, методы квантовой информатики могут найти применение в так называемом HFT-трейдинге (High Frequency Trading), суть которого состоит в проведении торговых операций с минимальными задержками. Причем счет

Рис. 1

в ходе подобных операций идет не на секунды, а на мили- и даже микросекунды. В этих условиях крылатое выражение «время – деньги» приобретает буквальное значение. Например, компания Hibernia Atlantic, провайдер трансатлантических широкополосных услуг, несколько лет назад вложила около $300 млн в прокладку нового кабеля через Атлантический океан для соединения Лондона и Нью-Йорка. Для максимального сокращения длины кабеля его маршрут строился с учетом рельефа морского дна. Считается, что благодаря новому соединению сделки на фондовой бирже можно будет совершать на 6 миллисекунд быстрее, что сулит существенные преимущества трейдерам, подключенным к этой услуге. И это – не предел, ведь с появлением более мощного и производительного оборудования, работающего на принципах квантовой информатики, время осуществления трансакции может сократиться еще значительнее.

Другим привлекательным направлением для квантовой информатики служит квантовая криптография, которая способна многократно повысить надежность работы финансовых институтов, внеся решающий вклад в обеспечение краеугольного условия протекания финансовой жизни – безопасности совершения банковских операций. Это особенно актуально для современного интернет-банкинга – динамично развивающегося ключевого направления предоставления финансовых услуг. Интернет-банкинг – главная форма совершения крупных финансовых транзакций, но не только. Сегодня подобная технология все шире внедряется в практику и для более массовых операций, таких как оплата услуг ЖКХ, подача заявок на кредиты, контроль средств на счете и т.д., что в совокупности приводит в движение через Интернет колоссальные денежные потоки. Это означает, что безопасность финансовых операций в интернет-банкинге становится фундаментальной проблемой, которая и по сей день остается серьезной причиной беспокойства всех финансовых организаций и частных потребителей финансовых услуг. В настоящее время многие банки и другие финансовые институты чаще всего используют для защиты своих данных так называемые RSA-алгоритмы (алгоритмы шифрования с открытым ключом). Одной из возможных альтернатив классическим методам защиты и шифрования как раз и может стать квантовая криптография. Среди лидеров нового направления – швейцарская компания IdQuantique. Использование законов квантовой физики позволило создать такую систему распределения и смены ключей кодирования (самая уязвимая часть любого процесса шифрования, зависящая от «человеческого фактора»), которая функционирует автоматически, без участия оператора-шифровальщика. Механизм смены ключей в квантовых системах обеспечивается физическим процессом запуска одного фотонного импульса по сетям телекоммуникации. Женевские физики, объединившиеся в 2001 году в компанию IdQuantique, уже в 2002 году смогли активизировать систему запуска одиночных фотонов по кабелю, идущему по дну Женевского озера. Опираясь на подобную же технологию, американская компания MagiQ Technologies (США) создала в 2003 году свой первый продукт – систему Navajo, которая была способна не только передавать ключ по телекоммуникациям, но также обновлять его каждые 10 секунд, обеспечивая защиту данных. За последнее десятилетие совместными усилиями десятков проектно-исследовательских организаций во всем мире квантовая криптография уже показала свою эффективность и вышла на уровень коммерческой реализации. Более того, сегодня не остается сомнений в том, что именно квантовой информатике принадлежит будущее лидерство всего глобального проекта создания действенной системы защиты информации.

Таким образом, становится все более очевидным, что квантовая информатика выступает как одно из прорывных направлений развития современной науки. Это подтверждается не только взрывом исследовательского интереса к ней во всем мире, но и существенным ростом инвестиций, вкладываемых в проекты по квантовой информатике всеми развитыми государствами и многочисленными венчурными фондами. И это неудивительно, ведь сегодня как никогда актуальна старая истина: кто владеет информацией, тот владеет миром.

Формирование Санкт-Петербургского кластера

В апреле 2014 года в ведущем информационном вузе страны – Санкт-Петербургском национальном исследовательском университете информационных технологий, механики

и оптики (Университет ИТМО) – была создана новая международная лаборатория «Современные инфокоммуникационные технологии и их приложения в экономике и финансах – Fin Q». Это подразделение выступает как звено в реализации программы повышения конкурентоспособности университета среди ведущих научно-образовательных центров. Проблемы квантовой информатики и ранее были объектом исследований такого научного подразделения вуза, как Международный институт фотоники и оптоинформатики. В частности, в лаборатории квантовой информатики ведутся работы по созданию эффективной системы передачи криптографического ключа по оптическому волокну, ведущее к оптимизации защиты данных. В вузе создается первая квантовая сеть. Новая лаборатория Fin Q, возникшая на базе факультета инфокоммуникационных технологий в университете ИТМО, определила в качестве главного направления своей деятельности применение квантовых теорий и технологий в разработке новейших решений в сфере экономики и финансов.

«Благодаря высокой эффективности и быстродействию новые алгоритмы, основанные на квантовых вычислениях, должны изменить лицо вычислительной экономики и финансов», – уверен Андрей Рыбин (PhD, MBA), заведующий международной лабораторией Fin Q. Среди первоочередных задач деятельности нового научного подразделения – теоретические разработки, позволяющие развить быстродействие вычислительных устройств, необходимых для проведения операций на финансовых рынках. Исследования будут охватывать три уровня решения поставленных задач: фундаментальные квантовые вычисления, разработка их приложений в экономике и финансах и, наконец, внедрение результатов исследований, включая коммерческие приложения (рис. 1).

Претворение в жизнь этих амбициозных планов требует от Университета ИТМО решения серьезной задачи – объединить усилия лаборатории Fin Q и ведущих научных институтов Санкт-Петербурга и создать единый образовательный и научный кластер для совместной подготовки квалифицированных кадров и проведения передовых научных исследований в области экономических приложений квантовой информатики (рис. 2).

Рис. 2

«Создание единого образовательного и научного кластера в сфере квантовой информатики, безусловно, важная задача. Я убежден, что общими усилиями мы сможем создать благоприятные условия для проведения передовых исследований в данной кросс-дисциплинарной области», – считает Владимир Васильев, ректор Университета ИТМО, председатель Совета ректоров вузов Санкт-Петербурга.

Помимо ученых из научных организаций Санкт-Петербурга, лаборатория привлекает к участию также крупнейших зарубежных специалистов в сфере экономики и финансов и квантовых вычислений. В качестве главного научного сотрудника лаборатории выступает известный ученый Николай Решетихин, профессор математической физики University of California Berkeley, США, внесший фундаментальный вклад в теорию квантовых и классических интегрируемых систем, основатель нескольких новых направлений математической физики. В работе лаборатории также участвуют профессора факультета математики Brunel University London, Великобритания, Лейн Хьюстон и Дорж Броди, уже долгое время занимающиеся вопросами формирования стоимости активов и управления рисками. Будучи крупнейшими специалистами в области финансов и экономики, ученые из университета Brunel в сотрудничестве с экспертами в сфере квантовой информатики в рамках лаборатории Fin Q ищут решения актуальных экономических моделей и задач, которые не могут быть предложены на основе существующих традиционных подходов.

Прекрасная возможность обсудить все грани текущих и предстоящих исследований, наметить стратегию будущих решений и междисциплинарного партнерства была у ученых в начале ноября этого года, когда они встретились на представительном международном семинаре «Quantum Informatics and Applications in Economics and Finance», организованном в университете ИТМО. Тематика семинара затрагивала ключевые аспекты деятельности лаборатории Fin Q и отражала ее кроссдисциплинарный характер: обсуждались вопросы квантовой информатики, приложений квантовой физики в экономике и финансах, финансовой математики и топологических квантовых вычислений. Такая проблематика, в которой исследовательские задачи вплотную приближены к актуальным запросам обществ, и привлекла на семинар не только квалифицированных специалистов по информатике и экономике, но и студентов ИТМО и СПбГУ.

Успешный НСКФ

Текст Игорь Лёвшин

У Национального Суперкомпьютерного Форума (НСКФ) в этом, 2014 году круглая дата. Сам форум проходит 3-й раз, но зато главному организатору – ИПС имени А. К. Айламазяна РАН – исполнилось 30. Генеральным партнером НСКФ в этом году была Российская Венчурная Компания, а платиновым спонсором – РФФИ. Как и в прошлые годы, форум проходил в Переславле-Залесском, на окраине которого расположен институт. На этот раз было около 300 участников, более 152 организаций из 34 городов.

Самое заметное новшество – фактически самостоятельная конференция «Посткремниевые вычисления», которая привлекла большое внимание.

Традиционно к форуму была приурочена презентация компании «ИММЕРС». На ней генеральный директор компании Леонид Клюев представил новое, 4-е поколение систем – младшую модель IMMERS 660 и старшую IMMERS 880, с помощью которой рекомендуется создавать системы начиная от 70 ТФлопс и выше. Также он сообщил о запуске Академической программы ИММЕРС, в рамках которой будет в том числе организован удаленный доступ к суперкомпьютерам для тестирования пользовательских приложений.

Третья новость компании – ею получен статус резидента «Сколково», чтобы получить дополнительное финансирование исследовательских проектов.

Другим ньюсмейкером была компания «ТЕСИС». У компании свой юбилей: 20-летие работы на рынке. Ее генеральный директор Сергей Курсаков рассказал о перспективах компании, ее ключевого продукта FlowVision и представил концепцию платформы с открытым интерфейсом программирования. Цель ее создания – быстрое расширение функционала. «ТЕСИС» расчитывает на добавление новых моделей и методов вычислений.

Конференцию венчал бурный и продолжительный круглый стол. Темы, обсуждаемые на нем, практически все достойны отдельного разговора.

Николай Непейвода о «Посткремниевых вычислениях»:

Традиционная электроника подошла к своему физическому и логическому пределу, в который упирается на передовых рубежах уже сейчас. Это – скорость света, предел Ландауэра, ограничивающий снизу выделение тепла, стена памяти, усиливающая стократно действие предела Ландауэра, и предел Чейтина, ограничивающий логическую сложность программ и конструкций, после чего они становятся принципиально непонимаемыми. Приходится искать совершенно новое, что будет использовано в перспективе.

В этом русле на конференции рассматривались мемристоры (элементы, где простейшее вычисление соединено с аналоговой памятью, уже не двоичной; в России над ними работает блестящая школа А. И. Галушкина, начинавшего еще с персептронов), кристаллы (ИПС РАН и УдГУ; здесь уже не числа, а преобразования) и квантовые устройства (несть им числа, но все в России уже 15 лет по-прежнему на уровне принципиальных разработок; пытаются за счет ненадежности обхитрить границы при помощи эффектов, проявляющихся для сверхмалых количеств вещества: с некоторой вероятностью кванты могут сделать то, что считается невычислимым или исключительно трудным для вычислений).

Возникают и были представлены такие безумные идеи, как вычисления на вакууме (порождающем частицы под воздействием энергетических полей) и на сине-зеленых водорослях. Живые системы, в частности, реализуют пятилучевую симметрию, которой нет в кристаллах, и пятеричную систему. Этот доклад выпускника УдГУ К. Смирнова вызвал самое бурное обсуждение и был охарактеризован как шаг к возможному принципиальному открытию.

Экзафлопс задерживается, прагматика побеждает

Текст Леонид Черняк

Далеко не случайно девизом прошедшей в Новом Орлеане в середине ноября 26-й Ежегодной конференции Supercomputing Conference (SC14) стали слова «HPC Matters», тем самым был отмечен весьма актуальный в нынешних условиях тренд – активная переориентация HPC на практическое применение.

Не секрет, что на протяжении предшествующих десятилетий это направление компьютинга служило инструментом для ограниченного круга исследований, главным образом относящихся к физике высоких энергий и некоторым другим фундаментальным дисциплинам, а главным стимулом к его развитию служили разработка ядерного оружия и контроль за ним. Но все меняется – в 2014 году количество и состав участников и экспозиция, являющаяся важнейшей составной частью конференции, указывают на то, что HPC становятся критически важными в совершенно иных многочисленных и абсолютно мирных областях человеческой деятельности, нет сомнений, наступает новая эпоха так называемого всепроникающего суперкомпьютинга (age of pervasive supercomputing). Теперь наиболее актуальными сферами применения HPC становятся промышленное производство, финансы, медицина, общественная безопасность, развитие окружающей среды и многое другое. Показательно, что последним словом стало использование HPC в среднем и малом бизнесе, особенно при проектировании высокотехнологичных изделий. Самый образный пример задачи такого рода – отработка посадки на гоночном велосипеде. На SC14 были продемонстрированы десятки, если не сотни подобных приложений, требующих высокой, но отнюдь не рекордной производительности. О массовости HPC можно судить и по числу университетских стендов на выставке – их было не менее двух десятков из США, по несколько из Западной Европы, Японии, Южной Кореи, Россия была представлена студенческим стендом МГУ имени Ломоносова. Плюс к тому двенадцать студенческих команд (США, Германия, Австралия, Китай, Тайвань и Сингапур) приняли участие в соревновании по сборке кластеров из рыночных компонентов с ограничением по цене и потребляемой энергии. Собранные кластеры тестировались на известных пакетах ADCIRC, NAMD, MATLAB и одном «секретном», то есть не известном участникам в этом году астрофизическом коде. SC14 подтвердила, что по части процессоров лидером HPC была и остается Intel, подобная позиция более чем аргументирована, почти 90% компьютеров из нынешней редакции TOP500 собраны на процессорах этой компании, однако в своем выступлении вице-президент и руководитель отделения высокопроизводительных вычислений Intel Радж Хазра согласился с тем, что сегмент HPC переживает период серьезной трансформации. Грядущее обновление он связывает прежде всего с гетерогенным компьютингом и со следующим поколением процессоров Intel Xeon Phi, известным под кодовым именем Knights Landing и новым интеконнектом Omni Scale Fabric, который позволит эффективно объединить CPU с сопроцессорами. Его коллега Чарльз Вусчард, руководитель направления рабочих станций и HPC в Intel, подчеркнул, что Knights Landing станет первым процессором класса many-core, который будет соответствовать современным требованиям в части работы с памятью и операциями ввода-вывода. Предполагается, что с 60 ядрами архитектуры Silvermont, усовершенствованной для целей HPC, процессор Knights Landing будет иметь производительность 3 Тфлопса на операциях с двойной точностью. По словам Вусчард, целый ряд партнеров Intel готовится к выпуску компьютеров с производительностью порядка 100 Пфлопс в ближайшие несколько лет.

Отмеченный сдвиг парадигмы HPC непосредственным образом отражается в заметном изменении отношения и к культовому списку TOP500, и к заветному экзафлопному рубежу. Для большинства участников рынка HPC – а сейчас уже можно говорить о рынке, а не об отдельных проектах, – спонсируемых государствами, рекорды имеют такое же значение, как события в гонке «Формула-1» для владельца обычного автомобиля. О повышенном интересе к суперкомпьютерам средней производительности свидетельствует усилившееся присутствие в TOP500 специализирующееся в этом сегменте компании Cray – за несколько лет оно возросло с 37 до 62. Снижение интереса к TOP500 к тому же усугубляется очевидным застоем, нынешняя редакция TOP500 не сильно отличается от предыдущей, на первых позициях остаются те же, а куда более интересные события происходят в пелатоне экзафлопсной гонки. Среди них следует отметить очевидные достижения Группы компаний РСК – с суммарной пиковой производительностью поставленных ей машин более 2 Пфлопс она на равных вошла во второй эшелон, то есть в следующую за лидерами группу, насчитывающую 5–6 производителей, чья доля больше 1% от суммарной производительности.

РСК в третий раз участвовала в этом самом престижном суперкомпьютерном форуме – на ее стенде были кластерные решения «РСК Торнадо» на базе нового процессора Intel Xeon E5-2600 v3, отличающиеся высокой плотностью и энергетической эффективностью, пакеты расширений RSC Tornado Expansion Pack, служащие для решения специализированных задач заказчиков, и система массивно-параллельной архитектуры RSC PetaStream. RSC PetaStream построена на базе 60-ядерных сопроцесоров Intel Xeon Phi, процессоров семейства Intel Xeon E5-2600 v2 и твердотельных накопителей Intel SSD DC S3700. Решающим фактором отмеченного успеха РСК стала кластерная система «Политехник РСК Торнадо», созданная компанией для Санкт-Петербургского государственного политехнического университета и занявшая 81-е место с производительностью 658 Тфлопс на тесте Linpack, а система «Политехник RSC PetaStream» вышла на 390-е место в TOP500. Суммарная пиковая производительность нового суперкомпьютерного центра СПБПУ равна 1.1 Пфлопса. С его вводом среди 9 российских суперкомпьютеров, попадающих в TOP500, оказывается четыре, созданные в РСК.

В самой верхней части TOP500 первых подвижек следует ожидать через год-два, когда будут материализованы несколько основных американских преэкзамасштабных проекта (производительность – 100–200 Пфлопс). На них Министерство энергетики США выделило 325 млн долларов и еще более 100 млн ожидается от Министерства обороны в рамках проекта FastForward 2. Эти две машины будут создаваться в рамках инициативы CORAL (Collaboration of Oak Ridge, Argonne and Livermore), в которой участвуют все три крупнейшие физические национальные лаборатории. Ими будут заменены Titan в первой, Mira во второй и Sequoia в третьей. В 2017 году более мощная система Summit будет установлена в Oak Ridge, а немного уступающая ей Sierra – в Livermore. Новые компьютеры станут самым заметным результатом активности со стороны образованной под патронажем IBM ассоциации OpenPower, на первых ролях выступят, естественно, сама IBM, а также NVIDIA и Mellanox. Sequoia и Sierra (S&S) не просто в несколько раз производительнее, важнее другое: они будут заметно отличаться от машин, занимающих верхние позиции в текущем TOP500, поскольку нарушают сложившуюся традицию лобового наращивания производительности. S&S станут первыми представителями компьютинга, получившего название Multi-GPU Computing, базирующегося на двух стопах – на гетерогенной вычислительной модели (Heterogeneous Computing Model, HCM) и на высокоскоростном интерконнекте между CPU и GPU, поддерживающем HCM, в данном случае это NVIDIA NVLink. Гетерогенность HCM существенна тем, что она естественным образом отражает гетерогенную природу реальных данных – при обработке данных случаются разные периоды,

при некоторых условиях нагрузку удается с успехом распределить между тысячами узлов, но иногда возникают задержки из-за необходимости длительной последовательной работы. Ограничения по закону Амдала непреодолимы, и с их существованием следует мириться. Поэтому требуется такая архитектура, которая была бы в равной степени приспособлена к обоим типам нагрузки: и к производительной параллельной, и к вызывающей задержки последовательной. Переводя сказанное на язык процессорных технологий, можно сказать, что необходимо обеспечить на одном узле эффективное взаимодействие набора из нескольких GPU с одним CPU.

На физическом уровне обмен данными между GPU и CPU в S&S осуществляется посредством интерконнекта NVLink, который как минимум в несколько раз энергоэффективнее и на порядок быстрее, чем l PCIe Gen3 x16. Отличительная особенность NVLink – в обеспечении унифицированного доступа CPU и GPU к памяти. NVLink будет впервые доступен на процессорах NVIDIA Pascal GPU, где обеспечит связь между одним CPU и несколькими GPU. Возможно, в S&S будет использована следующая версия NVLink 2.0, поддерживающая когерентность кэша, что еще больше повышает производительность. Mellanox EDR InfiniBand послужит в качестве системного интерконнекта между узлами.

В отличие от крупных коммерческих конференций, где главное оказывается сказанным на пленарных заседаниях, содержание SC14 и подобных акций распределено по сотням сессий и рабочих групп. Одним из наиболее интересных из них стала The 2014 International Workshop on Data-Intensive Scalable Computing Systems (DISCS-2014), которая была посвящена Data-Intensive Computing (DIC), то есть инфраструктурам обработки больших объемов данных. В задачу DISCS-2014 входило объединение двух направлений – HPC и Big Data. Здесь обсуждались следующие вопросы: архитектуры HPC для систем, предназначенных для DIP (Data Intensive Applications); программные модели для DIP; рабочие среды и инструменты для поддержки DIP, а также другие. По результатам DISCS-2014 будет издан специальный номер «Special Issue on Data-Intensive High Performance Computing» журнала International Journal of High Performance Computing Applications (IJHPCA).

	Summit	Titan	Sierra	Sequoia
CPU	IBM Power9	AMD Opteron	IBM Power9	PowerPC A2
GPU	NVIDIA Volta	NVIDIA Kepler	NVIDIA Volta	N/A
Пиковая производительность, Петафлопс	150-300	27	100+	20
Потребление, МВатт	10	9	N/A	8
Число узлов	3400	18 688	N/A	N/A

Количество представленных в рамках SC14 DIC-решений было невелико, но они чрезвычайно интересны. Прежде всего, стоило обратить внимание на специализированную машину для аналитики в памяти NumaQ, плод совместного труда норвежской компании Numascale и тайваньской 1degreenorth. NumaQ напоминает кластер тем, что собирается из стандартных серверов, широко масштабируется, начиная от 128 ядер и 1 Тбайта памяти и работает под управлением Red Hat Enterprise Linux, но в отличие от наследников Беовульфа серверы объединены посредством специализированного интерконнекта NumaConnect, который можно представить как трехмерный тор, обеспечивающий всем ядрам равный доступ к общей памяти. Очевидно, что это решение на порядки дешевле многомерных торов, толстых деревьев и других подобных решений. В данный момент на NumaQ устанавливается статистический пакет R, поддерживаемый средствами ПО NumaManager. Внешне для пользователя все выглядит так, как если бы он работал на обычном ПК.

На SC14 Numascale выставила систему из 108 серверов Supermicro 1U, каждый из которых имеет на борту 48 ядер в трех процессорах AMD Opteron 6386 и 192 Гбайт памяти, вместе они образуют единое адресное пространство 20.7 Тбайт, к которому имеют доступ 5184 ядра. Самое удивительное в том, что история Numascale уходит корнями в стандарт Scalable Coherent Interconnect (SCI), разработанный более 20 лет назад и с тех пор практически забытый. Прогнозы аналитиков о наступлении в 2018 году экзафлопной эры, сделанные исходя из предположений о справедливости закона Мура, оказались несостоятельными – это событие переносится в лучшем случае на пять лет, когда в США, возможно, случится следующий очередной серьезный скачок по части суперкомпьютеров, заметим, если этому ничто не помешает. А в Европе и Японии планы скромнее – там предполагают к 2020 году подойти к рубежу, равному только половине экзафлопса.