

3. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2008.

4. Breiman L. Bagging predictors // Machine Learning. – 1996. – Vol. 26, No. 2. – P. 123–140.

5. Freund Y., Schapire R. Experiments with a New Boosting Algorithm // Machine Learning: Proceedings of the Thirteenth International Conference, 1996.

6. Breiman L. Random Forests // Mach. Learn. – 2001. – Vol. 45, No. 1. – P. 5–32.

7. Breiman L., Friedman J., Olshen R., Stone C. Classification and Regression Trees. – Wadsworth, 1983.

8. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1979. – 448 с.

О.В. Джосан

Московский государственный университет имени М.В. Ломоносова

О ПЕРСПЕКТИВАХ И ПРОБЛЕМАХ ЭКЗАФЛОПСНЫХ ВЫЧИСЛЕНИЙ В РОССИИ

Осенью 2009 года, во время проведения конференции «Научный сервис в сети Интернет–2009» [1] группой энтузиастов было зарегистрировано доменное имя exascale.ru – российский аналог домена exascale.org международного сообщества International Exascale Software Project [2]. Сделано это было в рамках вечной проблемы «научных отцов и детей» – молодежь требует революции, а старшее поколение – соблюдения традиций. Молодым хотелось доказать, что те проблемы, которые обсуждают «отцы», устарели и потеряли актуальность, необходимо выходить на новый уровень и думать о препятствиях, которые нам обещают к 2018 году. Иначе шансов быть на достойном месте в мировом суперкомпьютерном сообществе у России немного. Однако широкой поддержки тогда эта тема не получила и была в целом признана несвоевременной.

За прошедший год в России многое изменилось: для широкой общественности стал доступен «полпетафлопса» [3]. В обсуждение ситуации с «пето» и «экза» вовлекалось все больше и больше ученых из разных организаций и сфер научной деятельности. На российских суперкомпьютерных конференциях появились доклады о масштабируемости на тысячи ядер, моделировании задач, которые раньше считались нерешаемыми, и т.д. Перспектива «экза» стала уже не столь далекой и абстрактной. Нужно ли «экза» в России к 2018 году? Есть ли перспективы его достичь? Какие ресурсы существуют для этого в стране? Появятся ли новые парадигмы программирования или принципиальные подходы к построению вычислительных устройств? Вопросов очень много. На большинство из них пока нет ответов, но уже пришло время формулировать проблемы, которые требуют решения.

В рамках международной конференции «Научный сервис в сети Интернет–2010» прошло первое заседание Российского сообщества экзаккомпьютерных технологий [4], посвященное обсуждению актуальных проблем экзавычислений в России и мире. Данная статья подводит итоги обсуждения и рассматривает следующие вопросы:

- проблемы подготовки кадров для работы с экзавычислителями;
- актуальные задачи, требующие расчетов на сверхбольших компьютерах;
- проблемы создания программного обеспечения для экзаккомпьютеров.

Приведенный список проблем никак не может считаться исчерпывающим и формулирующим основные проблемы перехода к «экзаскейлу», однако представляет практический интерес и позволяет сформулировать «руководство к действию» на ближайшую перспективу для российских исследователей.

Автор выражает благодарность всем участникам первого заседания сообщества: Лацису Алексею (ИПМ РАН), Дбар Светлане (ИПМ РАН), Орлову Антону (ИПС РАН), Андрею Адинцу (НИВЦ МГУ), Шворину Артему (ИПС РАН), Климову Юрию (ИПМ РАН), Сальникову Алексею (ВМК МГУ), Андрее-

ву Дмитрию (ВМК МГУ), Серебряйскому Артему (ВМК МГУ), Коробкову Сергею (ВМК МГУ), Позднееву Александру (ВМК МГУ), Коржу Антону (Т-платформы), Давыдову Александру (ИПМ РАН), Андриюшину Дмитрию (НИЦЭВТ).

Проблема подготовки кадров для работы с экзавычислителями. Проблема образования в России в последнее время обсуждается с разных сторон: почему в школах стали хуже учить, почему все ПТУ стали вузами, почему в вузах учат тому же, что 20 лет назад? Вопрос обучения на разных стадиях весьма многогранный и острый для нашей страны. В области суперкомпьютерного обучения мы безнадежно отстаем от развитых стран. К сожалению, точной статистики нет, но слово «суперкомпьютер» знают далеко не все россияне. Даже в Московском университете, где установлен тринадцатый по мощности в мире суперкомпьютер[3].

Суперкомпьютерное образование активно развивается в рамках Суперкомпьютерного консорциума университетов России [5]. Основная деятельность в области суперкомпьютерного образования связана с обучением в вузах. Однако экзасистемы требуют не только высококвалифицированных специалистов, которые их разрабатывают, и ученых, которые их используют для решения прикладных задач, но и обслуживающего персонала.

К сожалению, далеко не все процессы в настоящее время удается автоматизировать, поэтому появляется потребность в таких участниках суперкомпьютерного процесса, как, например, операторы, которые следят за показанием датчиков системы и вовремя реагируют на сбои. Также для обслуживания компьютера необходим персонал, который будет оперативно менять вышедшие из строя компоненты. На системах пета- и тем более экза-уровня регулярно выходящие из строя компоненты – это обыденная ситуация. Говорят, что в ближайшем будущем каждому суперкомпьютеру нужен будет водопроводчик или сантехник. Сейчас это воспринимается скорее как шутка, однако с повсеместным внедрением водяного охлаждения разработчикам суперкомпьютеров вплотную приходится общаться со специалистами в этой области.

Для обслуживания любой информационной системы традиционно требуется человек под кодовым названием «системный администратор» или «эникейщик» – человек, который пытается наладить взаимодействие пользователей с системой. В больших суперкомпьютерных системах функциональность системного администратора подразделяется на несколько специальностей. Например, установка различных прикладных программных пакетов на системы петафлопсного уровня уже вызывает огромное количество проблем. Зачастую прикладной исследователь не обладает компетенцией, чтобы такие проблемы решать, а системному администратору не до этого: потратить пару недель на «допиливание пакета» – это непозволительная роскошь. В системах экзафлопсного уровня такие задачи придется решать регулярно. Поэтому необходимо готовить соответствующих специалистов.

В настоящее время в России не существует многоуровневой системы образования. Считается, что нужно либо получить высшее образование, либо обходиться без образования совсем. Перспектива «экза» может послужить фактором развития образования в суперкомпьютерной области по схеме многоуровневости: когда в отрасли требуются не только специалисты с высшим образованием, но и люди, которым достаточно среднеспециального образования или краткосрочных курсов повышения квалификации.

Актуальные задачи для экзавычислителей. Перспектива появления вычислительных систем экзауровня ставит суперкомпьютерное сообщество перед проблемой определения круга задач, которые можно будет решить с появлением систем такого уровня. Многие специалисты задаются вопросом, а нужен ли будет в 2018 году тот самый экзакомпьютер или это создание «сферического коня в вакууме»? На сегодняшний день человечество смогло придумать всего несколько реальных задач, которые для своего решения требуют экзавычислителя и не могут быть решены на сегодняшних суперкомпьютерах.

Одной из основных целей сообщества International Exascale Software Project была формулировка списка задач, которые

требуют экзавычислений для своего решения. Интересна проекция этого списка на российское суперкомпьютерное сообщество. Какие перспективные направления исследований в России могут потребовать эксафлопсных вычислений?

Одним из таких направлений, безусловно, станет более точное предсказание погоды. В России в настоящее время ведутся подобные исследования. Еще одной важной задачей, которая решается в нашей стране и требует значительных вычислений, является моделирование лекарственных препаратов. Также важной отраслью для нашей страны является нефтегазовая отрасль, в которой суперкомпьютерное моделирование применяется для существенного сокращения стоимости разработки.

Антон Корж предположил, что экзакomпьютер нужен для того, чтобы стать последним компьютером на старых принципах: основная задача, которую решит экзакomпьютер – это создание компьютера на новых принципах, возможно квантового.

Проблема формулировки задач для суперкомпьютера экзакomпьютерного уровня сводится к анализу междисциплинарных исследований, которые проводятся с использованием суперкомпьютеров. В настоящее время в мире идет активное формирование междисциплинарных исследовательских команд, в том числе включающих специалистов по суперкомпьютерному моделированию. Такое объединение усилий позволяет решать задачи, которые раньше казались неразрешимыми. Для этой цели создаются специализированные междисциплинарные исследовательские центры, где совместно работают исследователи из разных областей. Для России эта тенденция не характерна. Междисциплинарные работы проводятся узкими группами, часто для иностранных заказчиков, поэтому решаемые задачи не отличаются масштабностью. Для широкого распространения междисциплинарных исследований прежде всего необходимо обучение прикладных пользователей основам суперкомпьютерных вычислений и развитие коммуникаций специалистов в суперкомпьютерной области и прикладных пользователей.

Проблема создания программного обеспечения для экзакomпьютеров. Сложно предсказать, какая парадигма про-

граммирования будет актуальна через 8 лет. Однако можно попытаться сформулировать проблемы, которые эта парадигма будет решать. Вероятно, что основными проблемами программистов, пишущих программы для экзакомпьютеров, станет ввод-вывод данных и масштабируемость программ.

Проблемы ввода-вывода возникают и в большинстве приложений, которые сейчас выполняются на петафлопсных суперкомпьютерах. Неравномерность коммуникаций приводит к появлению простоев вычислительных узлов и замедлению общего процесса вычислений. Возможны два подхода к решению проблемы ввода-вывода: использование специализированных библиотек для оптимизации ввода-вывода в программе или использование более совершенных аппаратных конфигураций, позволяющих оптимизировать ввод-вывод. Оба подхода обладают классическими недостатками: первый слишком сложен для прикладного пользователя, второй существенно увеличивает стоимость установки. Возможным решением проблемы ввода-вывода в больших системах может стать новая концепция программирования, в рамках которой накладывается строгое ограничение, требующее разделения параллельной программы на части «ввод-вывод» и «счет». Имея структуру программы с разделением частей программы на эти две категории, можно существенно оптимизировать и балансировать загруженность вычислительных узлов.

Масштабируемость задач на системе с миллионом вычислительных ядер – это проблема, которую уже приходится решать программистам. К сожалению, большинство прикладных пакетов для суперкомпьютерного моделирования имеют существенные ограничения на масштабируемость, т.е. на максимальное количество процессоров, для которых будет получаться заметное ускорение. Одним из решений этой проблемы может стать переход к программированию с использованием односторонних коммуникаций. Ярko выраженная проблема парадигмы односторонних коммуникаций – сложность в понимании и применении на практике для специалистов, которые ранее исполь-

зовали MPI для написания программ. В этом аспекте перспективной представляется разработка системы автоматизированного перевода программ из модели двухсторонних коммуникаций в модель односторонних, т.е. системы, которая скроет от пользователя трудности, связанные с переходом на односторонние коммуникации. При этом такая система заметно увеличит масштабируемость программы.

Перспектива появления вычислительных систем экзафлопсного уровня формулирует большое количество задач и проблем, которые придется решать. В частности, это проблемы создания программного и аппаратного обеспечения для достижения экзафлопсной производительности, проблемы суперкомпьютерного образования, проблемы постановки реальных задач, требующих таких больших вычислений. Отрасль суперкомпьютерных вычислений в России стремительно развивается в последние годы. В данной статье показана актуальность перечисленных выше проблем экзакомпьютерных вычислений для России. Рассмотрены некоторые пути их решения.

Список литературы

1. Сайт конференции «Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность–2009». – URL: <http://agora.guru.ru/display.php?conf=abrau2009>.
2. International Exascale Software Project. – URL: <http://exascale.org>.
3. Суперкомпьютер «Ломоносов». – URL: <http://www.t-platforms.ru/ru/clusters/clusters/unique/lomonosov.html>.
4. Российское сообщество экзакомпьютерных технологий. – URL: <http://exascale.ru>.
5. Суперкомпьютерный консорциум университетов России. – URL: <http://hpc-russia.ru>.