

Миллиард вершин

Текст: Александр Фролов, Александр Семенов

5 марта в МГУ прошел второй научно-технический семинар GraphHPC-2015 (www.dislab.org/GraphHPC-2015), посвященный проблемам параллельной обработки больших графов с использованием суперкомпьютеров. Организаторами семинара выступили компания НИЦЭВТ (лаборатория DISLab) и НИВЦ МГУ.

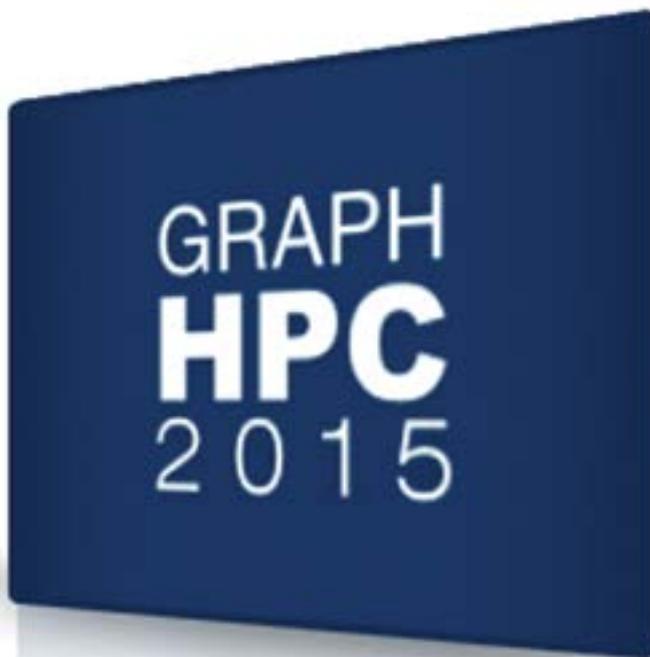
Как и год назад, тематика семинара была посвящена четырем «столпам» суперкомпьютерной обработки больших графов: приложениям, алгоритмам, программным технологиям и архитектуре систем. Приложений, анализирующих большие графы (более миллиарда вершин), становится все больше. Кроме социальных сетей и сети Интернет, «поставщиком» больших графов является биоинформатика, и тут позиции отечественной науки по меньшей мере соответствуют мировому уровню. В частности, в ИППИ РАН работает коллектив, возглавляемый В.А. Любецким,

который занимается получением новых биологических знаний при помощи собственных алгоритмов анализа самых разнообразных графов, возникающих при суперкомпьютерном моделировании процессов эволюции. В этом году Василий Александрович сделал доклад про алгоритм приближенного поиска подграфов в графе, число вершин в котором может достигать 10¹⁹. Для таких задач требуются вычислительные системы по принципу «чем больше узлов, тем лучше». Тематика социальных сетей была представлена на семинаре двумя докладами. Алексей Зиновьев из

омского университета (ОмГУ) рассказал о модели обнаружения спам-аккаунтов с использованием методов распределенной обработки графов на примере данных, полученных из социальной сети «Одноклассники». Кирилл Чихрадзе из ИСП РАН сделал доклад о параллельной генерации графов, по своим характеристикам соответствующим реальным социальным сетям, что очень важно для дальнейшей апробации алгоритмов анализа социальных сетей.

Для анализа в данных работах использовались фреймворки Giraph и Spark соответственно, что позволило выполнять параллельную обработку графов на системах типа Amazon EC2.

Анна Пирова (ННГУ) рассказала о нижегородской разработке – библиотеке MORSy переупорядочивания разреженных матриц на основе многоуровневых методов разбиения графов. В докладе были продемонстрированы результаты работы одного из параллельных алгоритмов для системы с общей памятью в сравнении с известной библиотекой ParMETIS. Неожиданное графовое приложение возникло при создании расчетных сеток со сложной геометрией для задач численного моделирования. В данном случае под сложной сеткой понимается неструктурированная многогранная расчетная сетка, состоящая из произвольных многоугольников (в том числе невыпуклых и неплоских). О подготовке такой сетки и графах, возникающих при решении этой задачи, говорилось в докладе Александра Семенова из ОАО «НИЦЭВТ» (лаборатории DISLab). Сотрудники лаборатории



DISLab выполнили реализацию разработанного алгоритма для Intel Xeon Phi, показав преимущество ускорителя по производительности по сравнению с процессором Intel Xeon E5-2660.

Оптимизациям базовых алгоритмов обработки графов было посвящено несколько докладов. Первым был доклад об алгоритме параллельного поиска сообществ внутри графа, разработанном известным коллективом авторов из IBM Research во главе с Фабрицио Петрини. Работу для участников GraphHPC-2015 представил сотрудник IBM Александр Позднеев, продемонстрировавший хорошее масштабирование алгоритма на суперкомпьютере IBM Blue Gene/Q на графах, включающих до сотен миллиардов ребер. Также можно отметить и оригинальные приемы оптимизации, предназначенные для решения задачи поиска вширь (BFS) для распределенных систем, представленные Александром Дарьиным, руководителем исследовательской лаборатории в компании «Т-Платформы». Подходы по реализации поиска вширь на кластерной системе, оснащенной графическими ускорителями, представил

Александр Колганов из ИПМ РАН. Оптимизации представления SELLPACK разреженных матриц для обработки графическими ускорителями был посвящен доклад Александра Монакова из ИСП РАН. Необходимо отметить, что в программе GraphHPC-2015 впервые присутствовал доклад об обработке динамических графов: Михаил Черноскутов разработал в IBM Research алгоритм приближенного расчета характеристики центральности подграфа. Новой оригинальной технологии обработки графов, основанной на

специализированном проблемно-ориентированном языке Scala, разрабатываемом в российском RnD-подразделении компании Huawei, был посвящен доклад руководителя лаборатории Александра Слесаренко. В докладе были представлены особенности высокоуровневого языка Scala, разнообразные структуры данных, предназначенные для хранения разреженных матриц (т. е. информации о топологии графов), а также дано сравнение эффективности этих представлений для задачи поиска минимального остовного дерева в графе.

Программу докладов GraphHPC-2015 закрывала студенческая секция, представленная двумя докладами из Новосибирского университета (НГУ) по оптимизации алгоритмов поиска минимального остовного дерева и связанных компонент в графе. Научный руководитель студентов – Константин Калгин.

Конкурс

Ярким событием GraphHPC-2015 стал конкурс (contest.dislab.org), проводившийся с 20 января по



4 марта. Лаборатория DISLab разработала систему проведения конкурса HPCJudge, которая автоматически тестировала решения участников на корректность и производительность и отображала на сайте таблицу с результатами. В рамках конкурса нужно было получить наибольшую производительность на задаче построения минимального остовного дерева в неориентированном графе с весами (Minimum Spanning Tree, MST). Для тестирования производительности решений использовались два вида синтетических графов: RMAТ-графы и SSCA2-графы. В этом году была объявлена отдельная номинация для студентов, которые не имеют профессионального опыта программирования высокопроизводительных систем. Соревнование профессионалов, обладающих таким опытом, со студентами было бы нечестным. Также для поддержки студентов организаторы конкурса провели в МГУ мастер-класс с подробным рассказом о задаче, алгоритмах ее

решения и приемах оптимизации на суперкомпьютерах. С самого начала конкурса возникла интрига: удастся ли студентам обогнать «профессионалов», а также какая архитектура вычислительного узла позволит получить лучший результат – гибридная или однородная. Чем ближе становилось завершение конкурса, тем драматичнее развивались события. Таблица результатов была заморожена на сайте за несколько дней до окончания конкурса, поэтому участники вплоть до объявления результатов на семинаре не знали, кто победит. К концу конкурса участники отправляли решения все чаще, пик активности пришелся на последнюю ночь, отправки решений выполнялись даже за несколько минут до завершения конкурса. Иногда студенты обгоняли «профессионалов», однако в этом году «профессионалы» отбились. Победителем конкурса стал сотрудник компании «Т-Платформы» Александр Дарьин с гибридной

реализацией на узле с двумя центральными процессорами Intel Xeon E5-2690 и ускорителем NVIDIA Tesla K20x с результатом 3080 MTEPS (миллионов пройденных дуг в секунду). Александр выяснил, что ускорители NVIDIA за счет возможности эффективного выполнения редукции над большим объемом данных имеют преимущество над центральными процессорами на задаче MST. Победителем в студенческой номинации стал студент факультета ВМК МГУ Артем Мазеев с реализацией, использующей на том же узле только x86-процессоры, и результатом 1651 MTEPS. Результаты других призеров и участников конкурса оказались очень близко расположены друг к другу. Победители и призеры конкурса в общей и студенческой номинациях получили ценные призы от организаторов и партнеров семинара, компаний NVIDIA, IBM, Intel, «Т-Платформы», Sapphire, AMD, а также памятные сертификаты. ■■■