

ТЕХНОЛОГИИ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ: ТЕНДЕНЦИИ И ПЕРСПЕКТИВЫ

Технологии распределенных вычислений активно развиваются с середины 90-х годов прошлого века. Предпосылками к такому развитию стал, с одной стороны, бурный рост числа компьютерных систем самого разного уровня производительности, а также повышение качества и проникновения каналов связи, а с другой стороны – потребность решать все более сложные задачи, для которых мощности доступных вычислительных установок оказывалось недостаточно. Собственно технологии распределенных вычислений подразумевают использование для проведения расчетов уже имеющихся компьютерных ресурсов, соединенных между собой каналами связи общего назначения. Распределенные вычисления стали отчасти дополнением традиционных суперкомпьютерных технологий, позволяя объединять несколько суперкомпьютеров для решения задач высокой вычислительной сложности, а отчасти и альтернативой им, позволяя для решения ряда задач обойтись и вовсе без суперкомпьютеров, подключив к расчетам только доступные офисные и домашние компьютеры, учебные классы и т.д.

Интересно, что распределенные вычисления в широком смысле можно рассматривать и как технологию параллельного программирования. Задача не распараллеливается традиционными средствами, такими как MPI или OpenMP, вместо этого создается «обвязка» для выбранной программной платформы распределенных вычислений, которая отвечает за разбиение задачи на независимые порции и компоновку результатов. При этом распределение заданий и все транспортные функции берет на себя выбранная платформа.

Прежде чем перейти к обзору состояний технологий распределенных вычислений, а также тенденций и перспектив их развития, стоит отметить еще два тесно связанных с ними комплекса технологий: грид-технологии и cloud computing («облач-

ные вычисления»). В настоящее время эти понятия зачастую смешивают, заменяя одно другим, однако, с нашей точки зрения, такое использование некорректно. Рассмотрим вкратце каждое из них.

Грид, согласно определению Я.Фостера и К. Кессельмана – согласованная, открытая и стандартизованная среда, которая обеспечивает гибкое, безопасное, скоординированное разделение ресурсов в рамках виртуальной организации [1]. Грид-технологии включают в себя все множество вопросов, решаемых при создании масштабных проектов по распределенной обработке и хранению данных – обеспечение безопасности доступа, поиск, учет и планирование использования ресурсов и т.д. Вследствие своей многофункциональности программное обеспечение для грид-сред (де-факто это Globus Toolkit [2] и gLite [3]) достаточно трудно в установке, настройке и использовании. Его применение не очень оправданно в тех случаях, когда необходимо достаточно оперативно развернуть распределенную среду и провести расчет на доступных ресурсах. Более того, в базовом варианте запуск задачи в грид-среду, по сути, означает ее запуск на одном из входящих в нее ресурсов; в случае когда необходимо выполнить серию таких запусков, т.е. фактически реализовать распределенный расчет, придется использовать дополнительные средства. Стоит отметить, что такую функциональность предоставляет ПО промежуточного уровня, например, Unicore [4].

Cloud computing – технология обработки данных, обеспечивающая абстракцию между компьютерным ресурсом и нижележащей технической архитектурой, предоставляющая удаленный доступ к разделяемому массиву конфигурируемых компьютерных ресурсов, которые по запросу могут быть оперативно зарезервированы и задействованы с минимальными затратами на организацию и взаимодействие с непосредственно владельцами ресурсов. Если в грид-технологиях акцент делается на скоординированном использовании распределенных ресурсов, то в облачных вычислениях цель другая – предоставление компьютерных ресурсов в аренду. Облачные сервисы эффективно используются для обработки больших массивов данных или организации удаленного доступа к определенному программному обеспечению, однако их нельзя рас-

считать как готовые платформы для поддержки распределенных вычислений. Лимитирующими факторами, как правило, являются жестко стандартизированное программное обеспечение виртуальных платформ, не очень высокая пропускная способность каналов связи внутри облаков, а также принципиально платный характер использования сервисов.

Что же остается для распределенных вычислений в их классическом понимании, в каких направлениях будут развиваться эти технологии? Попробуем выделить наиболее важные, интересные и перспективные направления.

Инфраструктура больших проектов. Это те самые грид-среды, построенные для решения задач конкретной предметной области. В качестве одного из ведущих проектов такого класса можно назвать EGEE – Enabling Grids for E-sciencE [5] (ныне проект продолжается как EGI – European Grid Infrastructure [6]). С помощью этого проекта, в частности, реализуется среда для обработки и хранения данных, получаемых с Большого адронного коллайдера.

Организация масштабных расчетов путем привлечения максимально доступного числа вычислительных ресурсов различного плана – суперкомпьютеров, мощных серверов, рабочих станций. Такие вычислительные эксперименты могут строиться на основе инфраструктуры, подобной созданной в рамках программы СКИФ-ГРИД [7], объединяющей ряд крупнейших российских суперкомпьютерных центров и использующей ПО промежуточного уровня Unicore, либо на базе системы метакомпьютинга X-Com [8], позволяющей подключать к расчетам практически любые вычислительные ресурсы.

Volunteer computing – вычисления с использованием ресурсов компьютеров добровольных участников. Самый известный проект такого рода – SETI@home [9], цель которого – распределенная обработка сигналов, получаемых с радиотелескопа. В ходе развития этого проекта была разработана программная платформа BOINC [10], позволяющая на своей основе строить аналогичные проекты для различных научных областей. В настоящее время насчитывается более сотни подобных проектов, большинство из которых построены на платформе BOINC [11].

Распределенные вычисления в рамках веб-браузера. В отличие от предыдущего направления, где для участия в распределенном проекте на компьютере требуется установить и настроить специализированное ПО, в данном случае участнику будет достаточно зайти на специализированный веб-сайт для того, чтобы подключиться к расчету. Расчет будет проходить в контейнере веб-браузера, клиентская вычислительная часть может быть написана на Java/JavaScript, Microsoft Silverlight или даже Adobe Flash. Преимущества такого подхода исключительно весомые – крайне широкая аудитория охвата, стандартизированные платформы для написания приложений, автоматическое решение существенной части вопросов, связанной с безопасностью и изоляцией работы приложения на клиентской машине. Безусловно, есть и недостатки, наиболее существенный из которых – слабая пригодность упомянутых технологий программирования для создания именно вычислительных задач. Интерес к подобному направлению возник буквально несколько лет назад, однако уже сейчас существуют компании, продающие подобные вычислительные ресурсы [12].

Вычисления на мобильных устройствах. Достаточно простая и красивая идея – на руках у пользователей появляется все больше мобильных сетевых устройств – телефонов, коммуникаторов, интернет-планшетов и др., процессоры которых уже могут сравниться по мощности с процессорами настольных компьютеров, скажем, 10-летней давности. Процессоры этих устройств в основном находятся в состоянии простоя, следовательно, их вполне можно было бы использовать для проведения расчетов в те моменты времени, когда они не заняты своей прямой работой. Безусловно, на этом пути вопросов пока больше, чем ответов. Как разработать переносимое приложение с учетом разнообразия имеющихся мобильных платформ? Как обеспечить надежное и безопасное сетевое соединение? Как быть с повышенным расходом аккумуляторной батареи мобильного устройства из-за постоянной загрузки процессора? Эти проблемы еще предстоит решить, и тем не менее публикации на эту тему появляются уже сейчас.

В заключение еще раз хотелось бы подчеркнуть, что концепция распределенных вычислений, появившаяся свыше 20 лет назад, не теряет актуальности и сейчас, с увеличением мощности и доступности традиционных вычислительных систем. Наряду с изначально возникшими направлениями (объединение отдельных вычислительных систем в единое целое, утилизация простаивающих компьютерных ресурсов) в этой области появляются и новые технологии, а следовательно, прикладным специалистам становятся доступны новые методы решения вычислительно сложных задач.

Список литературы

1. GRIDCLUB.RU: Интернет-портал по грид-технологиям. – URL: <http://www.gridclub.ru>.
2. The Globus Alliance. – URL: <http://www.globus.org>.
3. gLite: Lightweight Middleware for Grid Computing. – URL: <http://glite.web.cern.ch/glite>.
4. UNICORE: Distributed computing and data resources. – URL: <http://www.unicore.eu>.
5. EGEE Portal: Enabling Grids for E-sciencE. – URL: <http://www.eu-egee.org>.
6. EGI – European Grid Infrastructure. – URL: <http://www.egi.eu>.
7. СКИФ-ГРИД: научно-техническая программа Союзного государства. – URL: <http://skif-grid.botik.ru>.
8. Система метакомпьютинга X-Com. – URL: <http://x-com.parallel.ru>.
9. SETI@home. – URL: <http://setiathome.berkeley.edu>.
10. BOINC. – URL: <http://boinc.berkeley.edu>.
11. Distributed Computing. – URL: <http://distributedcomputing.info>.
12. Plura Processing. – URL: <http://www.pluraprocessing.com>.