

ОРГАНИЗАЦИЯ ПРЕДМЕТНО-ОРИЕНТИРОВАННЫХ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ ЦЕНТРОВ В РАМКАХ ПРОГРАММЫ «УНИВЕРСИТЕТСКИЙ КЛАСТЕР».

А.И. Аветисян, С.С. Гайсарян, О.И. Самоваров, Э.В. Хашба

Введение

Практически все современные вычислительные системы являются распределенными и допускают параллельное выполнение на всех уровнях: инструкций, потоков и процессов. Кроме этого на развитие технологий распределенных и параллельных вычислений оказало сильное влияние свободное ПО: Linux, GCC, Java и т.п., а рост популярности концепции облаков, позволяющих предоставлять масштабируемые практически без ограничения ресурсы в виде сервисов дополнили список факторов, определяющих сегодня спрос на параллельные вычисления. Основной причиной развития технологий облаков стало, с одной стороны, развитие сетевой инфраструктуры и создание крупномасштабных центров обработки данных, а с другой, растущие в геометрической прогрессии потребности крупного бизнеса и частных пользователей в масштабировании своих информационных систем.

Несмотря на усилия, прилагаемые со стороны государства и бизнеса в стране нет таких специалистов в нужном количестве. В результате, многие научные учреждения и промышленные предприятия используют имеющееся оборудование неэффективно, что приводит к снижению спроса на высокопроизводительные параллельные и распределенные вычисления, простою дорогостоящей аппаратуры либо к ее использованию для решения случайных задач. Все это снижает уровень научных исследований и качество промышленных разработок.

Как известно, подготовка квалифицированных кадров невозможна в отрыве от производства и науки — для этого необходимы инновационные подходы к интеграции науки, образования и промышленности. В 2008 году родилась совместная инициатива научного сообщества и индустрии, которая впоследствии переросла в программу «Университетский кластер» (www.unicluster.ru).

Программа «Университетский кластер»

Программа была учреждена Российской академией наук (ИСП РАН и МСЦ РАН), компаниями НР, «Синтерра» и направлена на повышение уровня использования технологий параллельных и распределенных вычислений в высшем образовании и научно-исследовательской деятельности, а также на их ускоренное внедрение в промышленность России. Цель Программы в создании экосистемы поддержки параллельных и распределенных вычислений, а также в создании сообщества пользователей и разработчиков. Такая инфраструктура будет служить основой для организации научных исследований, промышленных разработок и обучения студентов и аспирантов, а также для развертывания масштабируемых облачных сервисов и предметно-ориентированных виртуальных лабораторий, доступных широкому кругу пользователей.

Одним из направлений использования инфраструктуры Программы является исследование новых концепций и парадигм программирования, в частности, создание испытательного стенда для поддержки исследований по технологиям предоставления сервисов и управления ими в масштабе нескольких центров обработки данных. Эффективное использование инфраструктуры невозможно без разработки соответствующих учебных планов, программ и средств поддержки курсов.

По состоянию на май 2010 года к Программе присоединилось 70 участников, а партнерами стали компания Microsoft, Intel, Ассоциация «Отечественный Софт», ФГУП «Информика», ФГБУ «Российская государственная библиотека», издательство «Открытые системы». Центр компетенции Программы создан на базе ИСП РАН, обеспечивающем техническую поддержку и координацию мероприятий по привлечению технической поддержки сервисов со стороны партнеров программы.

Вычислительная инфраструктура Программы

Высокое качество сервисов и испытательных стендов, развернутых на базе инфраструктуры «Университетский кластер», обеспечивается за счет виртуальной частной сети и вычислительных ресурсов и хранилищ.

Виртуальная частная сеть «Университетский кластер» развернута на базе коммуникационных каналов национального оператора связи компании Синтерра, к региональным ЦОД которой подключаются участники Программы. Сеть имеет топологию «каждый с каждым», а каналы сети способны обеспечить пропускную способность до 10 Гбит/с.

На территории организации-участника организуется шлюз, через который обеспечивается доступ из его корпоративной сети к испытательным стендам и сервисам, развернутым на базе вычислительной инфраструктуры Программы.

В сети представлены общие сетевые службы, развернутые и поддерживаемые в ИСП РАН: система доменных имён (DNS), справочник ресурсов (испытательных стендов и сервисов) сети, сервер централизованной авторизации.

Учредители, партнеры и участники программы предоставляют свои вычислительные ресурсы и хранилища файлов -- сейчас это кластеры Межведомственного суперкомпьютерного центра РАН, Российского научного центра «Курчатовский институт», Технологического института Южного Федерального Университета, [Научно-исследовательского института радио и ряда](#) других партнеров и участников Программы. Общая мощность вычислительной инфраструктуры около 200 TFLOPS.

0. Испытательный стенд OpenCirrus

Проект OpenCirrus [1] основан компаниями HP, Intel и Yahoo, при участии Национального правительственного научного фонда США, Университета штата Иллинойс в Урбана Шампейн, Технологического института Карлсруэ и Министерства развития компьютерных коммуникаций Сингапура. Целью проекта является создание открытого испытательного стенда на базе распределенных центров обработки данных для поддержки разработчиков прикладных и системных программных средств для облаков. Российская Академия наук, в составе ИСП РАН, МСЦ РАН и РИЦ «Курчатовский институт» -- первые в Восточной Европе организации, присоединившиеся к программе OpenCirrus, в рамках которой доступны следующие сервисы:

- *Physical Resource Set* -- предоставление пользователю (средствами открытой системы Tусооn [2]) доступа на уровне системного администратора к затребованной им аппаратуре;
- *Elastic Compute* -- организация (средствами открытой среды Tashi [3]) доступа пользователя к виртуальным средам;
- *набор средств работы с большими распределенными массивами данных* (на базе открытой среды *Hadoop Core* [4]);
- система параллельного программирования на основе языка высокого уровня *Pig*.

Кроме сервисов, реализованных в составе испытательного стенда проекта OpenCirrus, в рамках Программы реализованы следующие сервисы:

- Сервис *Parallel Compute* обеспечивает выполнение программ, на высокопроизводительных вычислительных системах с распределенной или общей памятью.
- Сервис *Workspace* позволяет предоставить пользователям необходимое им рабочее окружение.
- Сервис *Virtual Classroom* обеспечивает проведение конференций, лекций, семинаров и лабораторных работ в режиме «он-лайн».

Предметно-ориентированный научно-исследовательский центр

Одной из задач Программы является организация интерактивного взаимодействия с сервисами для развертывания предметно-ориентированных Web-лабораторий -- хабов, позволяющих организовывать взаимодействие конструкторских организаций с прикладной и фундаментальной наукой

Аппаратура хаба образуется из Web-серверов, обработчика интерактивных запросов, хранилища данных, grid, высокопроизводительных систем и суперкомпьютеров, а также серверов, обеспечивающих визуализацию результатов.

Программное обеспечение хаба имеет многоуровневую структуру. На первом уровне обеспечивается доступ к множеству приложений через единую точку входа -- Web-портал. Здесь решаются задачи представления и организации информации, первоначальной авторизации пользователей, а также автоматизированной генерации интерфейсов, обеспечивающих интерактивное взаимодействие с приложением. Для выполнения интерактивных приложений используются распределенные вычислительные ресурсы, за доступ к которым отвечает второй уровень. Здесь решаются вопросы объединения разнородных ресурсов, принадлежащих различным организациям, разграничения доступа и безопасности. Применение открытых универсальных интерфейсов позволяет наращивать вычислительную инфраструктуру, подключая к ней новые вычислительные средства.

Важной проблемой, решаемой на втором уровне является обеспечение однородности вычислительных ресурсов для запускаемых приложений. Каждое приложение предъявляет свои требования к версии операционной системы, составу предустановленных библиотек, аппаратному обеспечению и трудно гарантировать, что все вычислительные ресурсы будут удовлетворять определенным требованиям. Здесь обычно применяется технология виртуализации, однако необходимо обеспечить управление множеством образов виртуальных машин, учитывая, что объем каждого образа -- 3-10 Гбайт.

На третьем уровне обеспечивается доступ к отдельным вычислительным ресурсам. Здесь размещается интерфейс управления кластерами и решаются задачи обеспечения эффективного использования вычислительных ресурсов, безопасности, надежности передачи данных.

Ключевой возможностью хаба является предоставление через обычный браузер доступа к интерактивным инструментам моделирования -- можно изменить масштаб графика, повернуть молекулу, исследовать изоповерхности трехмерного тела и т. п.

Сегодня в рамках программы «Университетский кластер» ведется создание новых испытательных стендов и расширение набора сервисов в рамках совместных работ с партнерами программы; создаются новые предметно-ориентированные научно-исследовательские лаборатории в области нанотехнологий совместно с Международным инновационным центром нанотехнологий стран СНГ (МИЦНТ СНГ) созданным на базе Объединенного института ядерных исследований (ОИЯИ); по договоренности с локальными операторами связи, ведется подключение к программе, образовательных и научно-исследовательских организаций Республик Беларусь и Казахстан. В ближайшее время планируется интегрировать программу «Университетский кластер» в реализацию проектов Рамочных программ ЕС

ЛИТЕРАТУРА:

1. Проект OpenCirrus. <https://opencirrus.org/>
2. Проект Tycoon. <http://www.hpl.hp.com/research/tycoon/>
3. Проект Tashi. <http://incubator.apache.org/tashi/>
4. Проект Hadoop. <http://hadoop.apache.org/>