

Опыт применения GRID-технологий в Приморском сегменте сети ДВО РАН*

Г.В. Тарасов, Д.И. Харитонов, Д.С. Шиян

В статье авторы описывают опыт построения GRID-сети на базе информационно-вычислительных ресурсов Приморского научного центра ДВО РАН (ПримНЦ ДВО РАН). Приводится информация об имеющихся вычислительных ресурсах и имеющейся сетевой инфраструктуре. На базе открытого продукта Globus Toolkit описывается архитектура GRID-сети и организация выполнения прикладных вычислительных задач. В заключении авторы анализируют достоинства и недостатки представленной схемы и определяют дальнейшие направления развития.

1. Введение

В современном научном обществе постоянно изменяются границы возможного. Непрерывный рост производительности процессоров, построение многопроцессорных систем и, наконец, GRID-технологии открывают возможности проведения крупномасштабных вычислительных экспериментов, ранее считавшихся невозможными. В настоящее время GRID применяется в таких областях, как ядерная физика, защита окружающей среды, предсказание погоды, моделирование климата, численное моделирование в автомобиле- и авиастроении, биологическое моделирование, фармацевтика, астрофизика и многое другое.

В чём отличие GRID от традиционных вычислительных технологий? В большинстве аспектов – это продолжение того, что мы имеем уже сегодня. Традиционным вычислительным центрам, по сути, не хватает уровня программного обеспечения, позволяющего системным образом дать в распоряжение приложений или сервисов безграничное множество ресурсов аппаратной платформы. Это программное обеспечение можно рассматривать как метаоперационную систему. Подобно традиционной операционной системе она отображает рабочую нагрузку на ресурсы в соответствии с определенной политикой. Однако ресурсами теперь являются не только процессоры, память и устройства ввода/вывода, но и целые серверы, кластеры, операционные системы, виртуальные машины, дисковые массивы, балансировщики нагрузки, переключатели, базы данных, и т.д. Понятие рабочей нагрузки сместилось от сравнительно простых исполняемых файлов в сторону распределенных приложений и сложных виртуальных систем. Политики, используемые в GRID, не ограничиваются планировочными приоритетами, но также предназначены для разделения доступа вычислительных задач и сервисов во времени и по ресурсам GRID сети, управляемой метаоперационной системой. В настоящее время метаоперационные системы GRID реализуются различными фрагментами программного обеспечения, количество, функциональное назначение и алгоритмы которых регулируются исключительно на уровне договорённостей активно развивающего GRID технологии сообщества учёных-разработчиков. И хотя пока что рано говорить о законченности идей, интерфейсов, протоколов и компонент, очертания GRID метаоперационной системы в целом видны довольно ясно.

2. Сетевая инфраструктура ПримНЦ ДВО РАН

GRID-сеть Приморского научного центра ДВО РАН строится на базе общей научно-образовательной сети (НОС) Владивостока. В настоящее время НОС объединяет все ведущие научные институты центра и ведущие высшие учебные заведения. Ряд институтов и университетов обладают собственными кластерами. Перечень имеющихся вычислительных ресурсов перечислен в таблице 1. Характерной особенностью распределения ресурсов является

* Работа выполнена при финансовой поддержке грантов ДВО (раздел 1) и программы Президиума РАН №15

их сосредоточение в крупном узле НОС – Академгородок. Емкость каналов связи насчитывает 100 Мбит/с и, таким образом, образует достаточно благоприятные условия для создания распределенной вычислительной сети для нужд научных и образовательных учреждений. Наиболее крупным вычислительным узлом сети Академгородка является суперкомпьютерный центр института автоматики и процессов управления (СЦ ИАПУ ДВО РАН), на базе которого строится вся инфраструктура GRID-сети Приморского научного центра. Внутренняя сеть суперкомпьютерного центра построена на базе гигабитной сети и имеет прямой доступ в НОС и Интернет (RBNет).

Таблица 1. Суперкомпьютерные ресурсы Приморского научного центра ДВО РАН.

Научно-образовательное учреждение	Характеристики ресурсов
Институт автоматики и процессов управления ДВО РАН (ИАПУ ДВО РАН)	5 кластеров. Суммарная пиковая производительность составляет 850 GFLOPS.
Институт химии ДВО РАН (ИХ ДВО РАН)	1 кластер. Пиковая производительность составляет 220 GFLOPS.
Тихоокеанский океанологический институт ДВО РАН (ТОИ ДВО РАН)	1 кластер. Пиковая производительность составляет 96 GFLOPS.
Дальневосточный государственный университет (ДВГУ)	1 кластер. Пиковая производительность составляет 10 GFLOPS.

Ближайшим участником GRID сети в России для Приморского сегмента является Хабаровск, также планируется размещение кластеров в Петропавловске-Камчатском и Благовещенске. Канал связи RBNет, используемый всеми институтами Приморского сегмента, имеет пропускную способность 12 мегабит в секунду и используется в основном для доступа в Интернет. Таким образом, Приморский сегмент Дальневосточного Отделения РАН в настоящее время является относительно изолированным, оснащённым собственной суперкомпьютерной техникой и объединённым корпоративной сетью. Всё это определяет политику развития в ней GRID-технологий, как автономного GRID сегмента, подключение которого к GRID инфраструктуре России связано с развитием Интернет каналов взаимодействия.

3. GRID инфраструктура ПримНЦ ДВО РАН

Перечень терминов, используемых в GRID-сообществе, включает следующие понятия: виртуальная организация, ресурсный центр, вычислительный элемент, элемент хранения данных и Web-сервис. В основном эти понятия сформировались в рамках реализации двух крупных проектов: Enabling Grids for E-science (EGEE) [1,2] и Globus Alliance [3]. Развитие этих проектов продолжается и в настоящее время. Само понятие GRID предполагает коллективный разделяемый режим доступа к ресурсам и к связанным с ними услугам в рамках глобально распределённых виртуальных организаций, объединяющих НИИ, вычислительные центры, коммерческие организации и отдельных специалистов. С каждой виртуальной организацией ассоциируются собственная политика доступа и правила использования ресурсов GRID. Основой объединения GRID ресурсов служит ориентированная на сервисы архитектура (Service Oriented Architecture, SOA), которая представляет собой распределённый набор программных компонент, действующих в качестве посредников для программных и аппаратных ресурсов вычислительного элемента или элемента хранения данных. Такая парадигма даёт возможность пользователям этих ресурсов иметь дело исключительно с операционным описанием сервисов. В определении SOA подчеркивается, что все сервисы имеют адресуемые сетевые интерфейсы и взаимодействуют через передачу сообщений на основе стандартных протоколов и форматов данных. Реализация архитектуры SOA в Интернет привела к развитию WSA (Web Services Architecture), которая регламентирует взаимодействие сервисов на базе связки протоколов HTTP/SOAP (Simple Object Access Protocol) и языков WSDL (Web-Service Definition Language) и XML (Extensible Markup Language). XML

используется как стандартный, гибкий и расширяемый формат представления данных и сообщений внутри Web-сервисов. SOAP обеспечивает модульную основу для компоновки и передачи XML сообщений через HTTP каналы между поставщиком и потребителем Web-сервиса. WSDL – это основанный на XML язык для описания интерфейса Web-сервиса. На основе описания сервиса на автоматизированной основе может быть скомпилирован код, конвертирующий XML сообщения в язык разработки приложений, реализующих функциональность Web-сервиса.

Следуя представленной технологии, инфраструктура GRID-сети Приморского центра организована как самостоятельная виртуальная организация, включающая распределенный набор вычислительных и информационных ресурсов. Виртуальная организация объединяет четыре ресурсных центра, обладающих собственными вычислительными мощностями, ее общая структура показана на рисунке 1.

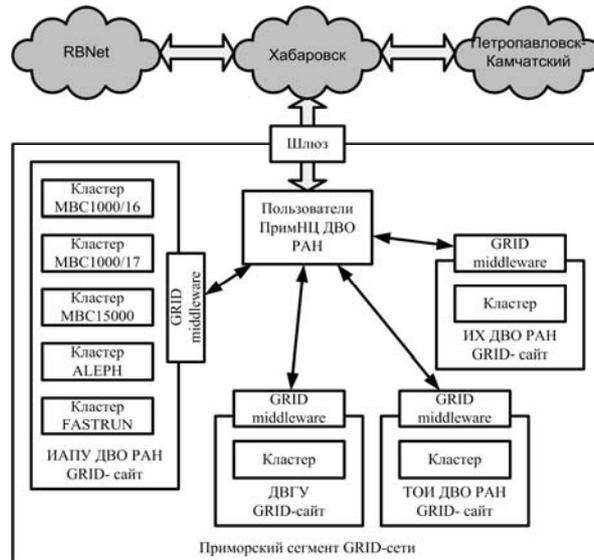


Рис. 1. Виртуальная организация ПримНЦ ДВО РАН.

Практически стандартом для построения GRID инфраструктуры стал программный продукт Globus Toolkit, разрабатываемый сообществом университетов и лабораторий Global Alliance. Данный продукт использовался для развертывания базовой GRID-инфраструктуры в Приморском сегменте сети ДВО РАН, внедрение которой происходило поэтапно. Основным этапом было определение достаточного набора Web-сервисов, позволяющих предоставить минимальный набор услуг пользователям. Такими сервисами стали: запуск и диспетчеризация выполнения прикладных задач, организация пересылки данных и исполняемых модулей между ресурсом и пользователем. На рисунке 2 изображена типовая схема конфигурации базового набора сервисов и включает следующие компоненты:

1. WS GRAM – сервис, обеспечивающий удалённый запуск и управление состоянием прикладных задач. Представляет собой вычислительный элемент (Computing Element – CE) в инфраструктуре GRID.
2. Сервис WS Delegation отвечает за проверку безопасности выполнения тех или иных взаимодействий между различными службами и пользователем на одном узле. Модель безопасности в Globus Toolkit построена на понятии сертификата и соответствует международному стандарту X.509. Права пользователя на доступ к тому или иному сервису по протоколу SOAP подтверждаются его сертификатом.
3. WS RFT (Reliable File Transfer) – это Web-сервис, обеспечивающий надёжный интерфейс передачи и управления файлами. RFT посредством SOAP сообщений через HTTP получает запросы на передачу файлов от пользователя и перенаправляет их на службу GridFTP. RFT - транзакционная служба, она использует базу данных для сохранения списка файлов, их состояний и операций над файлами. RFT способна восстанавливать корректное состояние в случае сбоев и перерывов при передаче данных.

данные являются результатом других вычислений. Вторая стадия – цикл вычислений. Вычисления бывают независимыми по данным и тогда могут выполняться параллельно, либо зависимыми по данным (вычисления на больших сетках) и тогда могут выполняться распределено. Третья стадия – сохранение результатов. Вычисленные данные, также как и исходные, обычно сохраняются в файлы или базу данных. Представленная схема с разной степенью эффективности может быть реализована как на суперкомпьютерах, так и в распределенной GRID-сети. Существенная разница в производительности кластеров единой GRID-сети, а также в производительности внутренней коммуникационной сети кластеров и сети, связывающей GRID-сайты, ограничивает взаимодействие между вычислительными элементами. При использовании вычислительных узлов нескольких кластеров для решения одной задачи может получиться большая несбалансированность вычислительной мощности выделенного ресурса, что существенно сказывается на скорости получения результата. В связи с этим было принято решения в Приморском сегменте GRID-сети ДВО РАН вычисления на кластерах организовать как локально-распределённые или локально-параллельные, использующие централизованный способ разделения начальных данных и сохранения конечных результатов.

В представленной модели работа пользователя выглядит следующим образом. Пользователь заходит на узел доступа к GRID-ресурсам и закачивает в домашнюю директорию необходимые программы и данные для выполнения задачи. На узле доступа к GRID-ресурсам пользователь готовит XML-описание задания в соответствии с требованиями клиента WS GRAM. Описываются основные этапы прохождения задачи: стадия закачивания необходимых исполняемых модулей и файлов данных из узла доступа на выбранный кластер, стадия запуска в системе пакетной обработки OpenPBS соответствующего кластера, стадия копирования результатов из кластера на узел доступа. После этого пользователь инициирует процесс запуска задачи с указанием файла с XML-описанием. Окончание процесса выполнения и просмотр текущего состояния задачи пользователь может узнать с узла доступа, вызвав соответствующие утилиты клиента сервиса WS GRAM.

Описанная модель вычислений и способ запуска в настоящее время проходят апробацию на нескольких реальных пользовательских задачах. Следует выделить такие области интересов, как анализ спутниковых данных, квантовомеханическое моделирование кристаллических наноструктур, моделирование газо- и гидродинамических процессов. Заметим, что данные задачи уже имеют реализацию в последовательном или параллельном исполнении, но никак не адаптированы для использования в GRID-сети. Предварительный анализ структуры каждой из задач показал, что не все они могут быть эффективно решены с применением GRID-технологий. Основная проблема в GRID заключается в том, что существенно больше времени тратится на своевременную подготовку пакета исходных или промежуточных расчетных данных конкретному узлу в процессе вычислений. Так, указанные задачи, связанные с моделированием динамических процессов и расчетом на сетках, требуют существенной переработки в той их части, которая касается получения, использования и сохранения данных для расчетов, а это составляет львиную долю всего алгоритма решения. В тоже время задача анализа спутниковых данных обладает хорошей степенью независимости входных данных, и ее решение может быть эффективно реализовано в GRID-сети.

5. Выводы

Опыт построения территориальной GRID сети показал, что применение GRID-технологий можно условно разделить на 2 категории: метакомпьютинг и распределённое предоставление ресурсов. Метакомпьютинг даёт возможность объединения вычислительных ресурсов с целью получения производительности превышающей производительность любого отдельного суперкомпьютера. Однако, использование парадигмы метакомпьютинга на практике не всегда оправдывается имеющейся сетевой инфраструктурой, когда участники GRID-сети связаны между собой медленными каналами связи. Более привлекательным выглядит использование второй парадигмы, позволяющей организовать распределенную среду предоставления вычислительных ресурсов.

Таким образом, в Приморском центре GRID-сеть имеет четко выраженную иерархическую структуру и представлена набором GRID-узлов, организующих доступ к локальным кластерам. Каждый кластер рассматривается как самостоятельный вычислительный ресурс, управление пользовательскими задачами внутри которого осуществляется локальной системой управления заданиями (OpenPBS). Конечно, при такой структуре прослеживается явный недостаток. Вопреки распределенной идеологии построения GRID-сети, мы внесли явно выделенные самостоятельные узлы. Однако, на начальных этапах внедрения GRID-технологий в большой региональной сети, такая структура дает ряд преимуществ. Во-первых, позволяет организационному участнику гибко настроить управление своими локальными вычислительными ресурсами для обеспечения потребностей локальных и удаленных пользователей. Во-вторых, без существенной реорганизации в будущем проводить интеграцию своих вычислительных ресурсов с вычислительными ресурсами соседних организаций.

Несмотря на полученный положительный опыт внедрения GRID-технологий, дальнейшее развитие сети требует решения многих вопросов. В частности необходимо отметить следующие задачи, которые могут быть решены в ближайшем будущем. **Тестирование, разработки и внедрение Web-интерфейса** для предоставления универсального доступа к ресурсам GRID-сети. Реализация единого Web-интерфейса сделает GRID более дружественной системой для конечного пользователя. **Ориентация на исходные тексты программ.** Вследствие гетерогенности, программное обеспечение на GRID системах должно обладать свойством переносимости с платформы на платформу. В большинстве случаев переносимость программного обеспечения обеспечивается открытыми исходными текстами, приспособленными для компиляции на UNIX системах. GRID инфраструктура не приспособлена для работы непосредственно с исходными текстами программ. Однако если разработать адаптер к системе очередей, применяющейся GRID инфраструктуре, то текст программы можно рассматривать как исходные данные, которые необходимо обработать для получения конечного результата. При этом исходный текст программы можно получать, например, из системы контроля версий (CVS), что позволяет также вести разработку программного обеспечения. **Библиотека GRID для коммуникаций.** Основное отличие выполнения GRID задач от задач, выполняющихся на кластерах – медленные коммуникации, заставляющие планировать разделение начальных данных, взаимодействие процессов и интегрирование конечных результатов заранее. Разработка шаблонных подходов обеспечивающих для GRID приложений сервис по получению начальных данных сохранению конечных результатов и обмену промежуточными данными упростит адаптацию существующего программного обеспечения для работы в GRID.

В заключение можно отметить, что потенциал технологий GRID очень велик, и в близкой перспективе GRID вслед за компьютером и Интернетом должен стать основным вычислительным инструментарием для развития высоких технологий в различных сферах человеческой деятельности. Чтобы в полной мере воспользоваться всеми преимуществами GRID, необходимо не только следить за развитием GRID, но и внедрять в GRID технологии, методы и средства, максимально удобные для его применения в конкретных прикладных областях.

Литература

1. EGEE – Enabling Grids for E-sciencE. Интернет-ресурс: <http://www.eu-egee.org/>.
2. Jacob B., Brown M., Fukui K., Trivedi N. Introducing to Grid Computing //
3. The Globus Alliance. Интернет-ресурс: <http://www.globus.org/>.
4. Bayucan A., Henderson L., Jasinskij L., Lesiak C., Mann B., Proet T., Tweten D. Portable Batch System. Administrator Guide // NASA Ames Research Center. 1998. P. 54.