

К ПОПУЛЯРИЗАЦИИ НАУЧНЫХ ПРИЛОЖЕНИЙ НА ИНФРАСТРУКТУРАХ EGEE И ГРИДННС

В.А. Вознесенский , В. Неверов

Резюме

В докладе описывается технология предоставления широкому кругу исследователей доступа к научным приложениям, требующим высокопроизводительных или распределённых вычислительных ресурсов, на примере программы моделирования рассеяния жёсткого излучения углеродными наноструктурами[1]. Доступ предоставляется через веб-портал ГИФ, разработанный в Курчатовском Институте. Доступ к вычислительным ресурсам для расчётов пользователей предоставлен гридовскими инфраструктурами EGEE-RDIG[2] и ГридННС[3].

– Введение

Появление в последние годы ряда высокопроизводительных вычислительных ресурсов в России[4] привело к двум явлениям. Во-первых, у разработчиков численных моделей появилась возможность без участия в зарубежных проектах выйти за рамки настольных ЭВМ и создавать программы, предназначенные для этих ресурсов. Во-вторых, обострился вопрос о предоставлении возможности расчётов с использованием этих программ для большого количества прикладных исследователей, не являющихся специалистами в информационных технологиях.

Вторая проблема имеет два аспекта. Во-первых, отрасли высокопроизводительных вычислений требуется набор средств для быстрого создания "дружественных" интерфейсов к разработанным программам. Во-вторых, массовый приход пользователей может привести к регулярным перегрузкам то одних, то других ресурсов, что потребует стандартных механизмов балансировки нагрузки, то есть распределения вычислительных задач на свободные ресурсы через стандартные сетевые интерфейсы этих ресурсов.

Такие механизмы распределения задач существуют в ряде грид-проектов, например ГридННС[3] и EGEE/RDIG[2]. В данном докладе представляется система ГИФ, предназначенная для создания веб-интерфейсов пользователей к приложениям, использующим эти гриды, на примере создания интерфейса для массового расчёта рассеяния жёсткого излучения на различных углеродных наноструктурах.

- Расчёты рассеяния излучения

Базовая программа[1] моделирует дифракционную картину, которая имеет вид двумерной матрицы распределения интенсивности рассеянного излучения (или частиц - нейтронов или электронов) по сферическим углам. Рассчитывается дифракция на различных углеродных наноструктурах (фуллеренах, нанотрубках и др.).

В настоящее время результаты расчетов используются для интерпретации результатов просвечивания плёнок, образующихся на стенках токамака Т-10, на Источнике Синхротронного Излучения Курчатовского Института. Эти плёнки содержат углеродные наноструктуры. Авторы программы пытаются решить обратную задачу определения функции распределения рассеивателей в исследуемом объекте по типам структур внутри заданного класса наноструктур.

С точки зрения численного моделирования задача хорошо распараллеливается, так как каждый элемент матрицы интенсивности считается независимо от остальных. В текущей реализации распараллеливание производится либо на отдельные процессоры MPI-ресурсов ГридННС, либо на отдельные задачи инфраструктуры EGEE-RDIG. Входные параметры программы описываются на языке XML определённого формата.

Развитие данной технологии предполагает создание научно-исследовательского конвейера, одной из стадий которого должно стать решение задачи определения структурного состава объекта (точнее, задачи дифрактометрической диагностики упорядоченной наноструктурной компоненты в исследуемом объекте) самим экспериментатором, не имеющим навыков работы с гридом или высокопроизводительными ресурсами.

– Функциональность ГИФ

Система ГИФ (Гридовский ИнтерФейс) уже представлялась на этой конференции в 2008 году [5].

ГИФ-приложение представляет собой набор управляющих сценариев на языке Python, а так же набор константных файлов, необходимых в работе этих сценариев. Сценарии при запуске конкретного приложения должны реализовывать следующие функции:

1. Начальная инициализация хранимых переменных
2. Декларация содержимого веб-страницы для вывода в его веб-браузере конечного пользователя на основании текущего состояния хранимых переменных
3. Обработка результатов передачи результатов пользовательского ввода через веб-браузер
4. Обработка результатов успешно завершившейся гридовской задачи

5. Обработка сбоя при запуске гридовской задачи

Продекларированные переменные хранятся в объектной базе данных. Сценарии исполняются в интерпретаторе, защищённом от доступа в операционную систему сервера. Для запуска задач используется метапланировщик Gridway[6], предоставляющий возможности запуска задач как через интерфейс CREAM[7], использующийся в инфраструктуре EGEE, так и через интерфейс WS-GRAM[8], использующийся в инфраструктуре ГридННС.

Редактирование перечня сценариев и константных файлов приложения осуществляется разработчиком приложения через специальный веб-интерфейс. Авторизация пользователей системой осуществляется исключительно с использованием сертификатов стандарта X.509[9].

Разработчик может предоставить доступ к своему приложению как отдельным пользователям системы, так и членам виртуальных организаций, поддерживаемых конфигурацией данной установки ГИФ, а так же членам подгрупп этих организаций.

Для запуска задач пользователь обязан загрузить в ГИФ свой прокси-сертификат с VOMS-расширением[10]. На данный момент это осуществляется на пользовательской ЭВМ из командной строки, что требует дальнейших усилий по созданию "дружественного", но остающегося при этом безопасным, интерфейса пользователя.

– Реализация приложения в ГИФ

На данный момент речь идёт о так называемом вариативном счёте, то есть запуске базовой программы для большого количества вариантов из многомерной области изменения параметров задачи.

Приложение имеет два основных режима, следующих друг за другом: режим задания параметров и режим расчёта. В первом режиме пользователь выбирает значения фиксированных входных параметров базовой программы, а так же последовательно выбирает параметры, подлежащие вариации. Варьируемые параметры задаются интервалами и шагами их варьирования в виде констант или арифметических выражений, зависящих от текущих значений уже выбранных параметров.

После задания параметров пользователь переходит к режиму расчёта. При этом, приложение проверяет корректность описания стационарных и варьируемых параметров, генерирует соответствующий этому описанию перечень XML-файлов с конкретными входными параметрами базовой программы, после чего декларирует для каждого XML-файла свою гридовскую задачу.

По мере завершения расчётов задач на гридовских ресурсах, пользователи получают возможность скачивать полученные файлы результатов, а так же просматривать результирующие графики через веб-интерфейс и оценивать близость смоделированных кривых к экспериментальным.

Ещё раз заметим, что вся описанная выше логика веб-приложения описана через веб-интерфейс ГИФ относительно простыми сценариями на языке Python.

Дальнейшая работа будет идти по трём направлениям. Во-первых, необходимо отработать создаваемый программный интерфейс на данных от значительного количества экспериментов и сделать его действительно удобным для пользователей. Во-вторых, возможно добавление в приложение функции автоматического подбора структурного состава исследуемых объектов на основании результатов физических экспериментов, то есть функции решения оптимизационной задачи. В-третьих, возможно создание универсального интерфейса решения вариационных и оптимизационных задач определённого класса на основании стандартизованного описания входных и выходных объектов при вычислении значений целевой функции.

● Выводы

В докладе продемонстрирован способ предоставления прикладным исследователям, не имеющим навыков в информационных технологиях, интерфейса к вычислительным приложениям, требующим значительных ресурсов, на примере решения реальной научно-прикладной задачи. Используемый способ гридификации приложений требует от разработчика значительных навыков программирования на языке Python, однако даёт возможность предоставления сложного "дружественного" веб-интерфейса конечному пользователю в сжатые сроки.

Авторы благодарят участников проектов ГридННС и EGEE-RDIG за предоставленный доступ к вычислительным ресурсам.

ЛИТЕРАТУРА:

1. A.B. Kukushkin, N.L. Marusov, V.S. Neverov, I.B.Semenov. Modeling of X-ray Diffraction by Carbon Nanotubes and Interpretation of Diffractometry of the Films Deposited in Tokamak T-10 // Будет опубликовано на конференции 36th EPS Conference on Plasma Physics, June 29 - July 3, 2009, Sofia, Bulgaria
2. EGEE-RDIG // Сайт консорциума Российского Грида для Интенсивных Операций с Данными: URL: <http://rus.egee-rdig.ru/> (2009, 31 мая)
3. Грид ННС // Сайт проекта Грида Национальной Нанотехнологической Сети: URL: <http://ngrid.ru> (2009, 31 мая)

4. TOP50 | Суперкомпьютеры // Страница совместного проекта НИВЦ МГУ и МСЦ РАН по формированию списка 50 наиболее мощных компьютеров СНГ: <http://supercomputers.ru/?page=rating> (2009, 31 мая)
5. В.А. Вознесенский. ГИФ - ВЕБ-ПОРТАЛ ДЛЯ РАЗРАБОТКИ И ИСПОЛЬЗОВАНИЯ ГРИДОВСКИХ ПРИЛОЖЕНИЙ // Научный сервис в сети Интернет: решение больших задач: Труды Всероссийской научной конференции (22-27 сентября 2008 г., г. Новороссийск). - М.: Изд-во МГУ, 2008. - 468 с. ISBN 978-5-211-05616-9 URL: <http://agora.guru.ru/abrau2008/pdf/082.pdf>
6. GridWay Metascheduler: Metascheduling Technologies for the Grid :: // Сайт проекта GridWay: URL: <http://www.gridway.org> (2009, 31 мая)
7. Cream - Main - Homepage // Сайт проекта CREAM CE: URL: <http://grid.pd.infn.it/cream/> (2009, 31 мая)
8. GT 4.0 WS_GRAM // Документация на компонент WS-GRAM программного комплекса Globus Toolkit: URL: <http://www.globus.org/toolkit/docs/4.0/execution/wsgram/> (2009, 31 мая)
9. S. Tuecke, V. Welch, D. Engert, L. Pearlman, M. Thompson. Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile // IETF RFC 3820. June 2004. URL: <http://www.ietf.org/rfc/rfc3820.txt> (2008. 31 мая)
10. WP6 Authorisation Working Group // Сервер итальянского Национального Института Ядерной Физики: URL: <http://grid-auth.infn.it/> (2008. 31 мая)