

А. В. ДУХАНОВ, Е. В. БОЛГОВА, Л. Р. ГЕРВИЧ, В. Г. КОЛПАКОВ,
Е. Н. КРАВЧЕНКО, И. И. КУРОЧКИН, Е. Д. МАСЛЕННИКОВ,
И. В. ОФЁРКИН, А. О. РУБЦОВ, С. А. СМИРНОВ, О. Б. ШТЕЙНБЕРГ, М. В. ЮРУШКИН

ПРЕДМЕТНО-ОРИЕНТИРОВАННЫЕ ТЕХНОЛОГИИ СОЗДАНИЯ ВИРТУАЛЬНЫХ РАБОЧИХ ПРОСТРАНСТВ В СРЕДЕ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ CLAVIRE

Рассматривается технология автоматизации создания виртуальных рабочих пространств для поддержки распределенных научных исследований в рамках парадигмы eScience. Представлено описание инструментального средства автоматизации, разработанного на основе среды облачных вычислений CLAVIRE, и продемонстрирована возможность его применения для создания облачных сервисов для различных предметных приложений.

Ключевые слова: композитное приложение, облачные вычисления, предметно-ориентированный язык, интеллектуальный редактор, облачный сервис.

Введение. Современные технологии облачных вычислений в рамках моделей SaaS (Software as a Service — программное обеспечение как сервис) и AaaS (Application as a Service — приложение как сервис) ориентированы на четкое разделение функций пользователей и провайдеров [1]. Как правило, все программное обеспечение (ПО) в „облаке“ устанавливается и настраивается провайдером, а работа с ним пользователя осуществляется в рамках сценариев, допускаемых облачной средой (например, ограничений на форму представления данных, интерактивность и пр.). Такая модель оправдана для большинства бизнес-приложений, обладающих высокой степенью унификации требований, однако ее перенос на задачи eScience (связанных с поддержкой научных исследований) непродуктивен [2]. Это обусловлено как разнообразием предметных областей и программного обеспечения, так и исследовательским (разовым) его применением в ходе научного поиска, что делает поддержку всего набора потенциально востребованного ПО экономически неэффективной для провайдера. Таким образом, это приводит к необходимости поиска новых коллаборативных форм модели SaaS, адаптированных для научных исследований; при этом пользователь имеет возможность не только работать с прикладными пакетами, но и *самостоятельно* встраивать их в облачную среду, обеспечивая их доступность для других пользователей.

В настоящей статье представлены результаты экспериментальных исследований по созданию автоматизированной технологии встраивания прикладного ПО в публичные среды облачных вычислений в рамках проекта по организации межвузовской мобильности молодых ученых ФЦП „Научные и научно-педагогические кадры инновационной России“ на 2009—2013 гг. [3]. Работы выполнены на базе Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (СПбНИУ ИТМО)

силами 10 привлеченных молодых ученых и специалистов из четырех сторонних организаций с использованием коллаборативной технологии создания публичных облачных сервисов [4].

Принципы построения виртуальных рабочих пространств. Виртуальное рабочее пространство (ВРП) в облачной среде обеспечивает реализацию средства доступа к инструментам распределенных вычислений для выполнения научных исследований, что позволяет консолидировать вычислительные и программные ресурсы в различных предметных областях. В общем случае к ВРП предъявляются следующие требования [5].

1. Поддержка процессов конструирования и исполнения композитных приложений (КП), состоящих из отдельных вычислительных сервисов распределенной среды, объединенных в целях решения общей задачи.

2. Ориентация на технологии удаленного взаимодействия с пользователем через веб-браузер, без использования специализированных клиентских приложений, в рамках модели SaaS или AaaS облачных вычислений.

3. Обеспечение интеллектуальной поддержки пользователей при работе с КП.

4. Обеспечение информационной безопасности ВРП в рамках модели системы с открытым контуром (для которой характеристики и местонахождение потенциальных пользователей не определены). При этом важно обеспечить защиту от легитимного пользователя, который имеет право доступа к ресурсам ВРП, но использует их не по назначению (например, для коммерческих расчетов или хранения данных из сомнительного источника).

5. Возможность эффективного совокупного использования ресурсов существующих суперкомпьютеров, сред облачных вычислений и грид-сред, поскольку для функционирования современных пакетов для научных расчетов в ряде случаев требуется специализированная вычислительная среда.

6. Поддержка „горячего“ подключения новых прикладных пакетов и вычислительных ресурсов при минимальных усилиях со стороны их разработчиков или владельцев. Данная мера обеспечивает открытость, переносимость и масштабируемость ВРП. Также сама технология является гибкой для создания ВРП силами специалистов-предметников (а не разработчиков-программистов).

Участие пользователей в создании ВРП, т.е. формировании содержания „облака“, может быть реализовано разными способами. Наиболее распространенной является агрегаторская схема, когда пользователь сам передает исходные коды и (или) исполняемые модули провайдеру, который устанавливает приложение на своих ресурсах и обеспечивает доступ к нему через веб-интерфейс. Однако такая схема подходит только для приложений массового пользования, поскольку требует выделения ресурсов провайдера и не может быть выполнена on-line, что противоречит самой парадигме облачных вычислений (выделение ресурсов по требованию, без участия провайдера). Паллиативным решением этой проблемы является использование модели облачных вычислений PaaS (Platform as a Service — платформа как сервис), в рамках которой пользователь сам должен устанавливать и поддерживать свое программное обеспечение на платформе провайдера. Однако в этом случае требуется наличие квалификации, которой многие специалисты-предметники не обладают. Кроме того, такой подход не позволяет обеспечить совместную работу пользователей „облака“; установленное ПО доступно только на виртуальной машине самого пользователя.

В настоящее время существуют три подхода в встраиванию прикладного ПО в ВРП с применением облачных вычислений в рамках модели SaaS:

— низкоуровневая разработка уникальных для каждого конкретного пакета проблемно-ориентированных интерфейсов для встраивания ПО в распределенную или облачную инфраструктуру [6];

— использование описаний пакетов и приложений на предметно-ориентированных языках, относящихся к научному направлению, которому соответствует ВРП (например, язык описания задач оптимизации — AMPL [7]);

— использование предметно-ориентированных языков, например EasyPackage [8], для описания способов применения прикладных пакетов в облачной среде.

Последний из указанных подходов представляется наиболее прогрессивным: в частности, он реализуется в многопрофильной инструментально-технологической платформе CLAVIRE (CLOUD Applications VIRTUAL Environment — виртуальная среда облачных вычислений) [5], обеспечивающей создание сред облачных вычислений в рамках моделей SaaS и AaaS; кроме того, этот подход хорошо зарекомендовал себя при исследованиях [4].

Во всех перечисленных случаях процедура встраивания пакетов в „облако“ требует навыков написания программ (или описаний) на соответствующем языке. Как следствие, для привлечения широкого круга потенциальных разработчиков ВРП провайдер облачной среды должен обеспечить средства автоматизации процесса встраивания (включая средства формирования предметно-ориентированных интерфейсов), не требующие специальных навыков в области программирования и знаний о функционировании вычислительных ресурсов.

Рассмотрим вопросы разработки и применения такого инструмента для среды облачных вычислений CLAVIRE.

Технология автоматизации построения виртуальных рабочих пространств. В общем случае процесс встраивания вычислительного пакета в платформу CLAVIRE состоит из следующих этапов:

- 1) установка прикладного пакета на зарегистрированный ресурс (ресурсы), который соответствует программно-аппаратным требованиям к устанавливаемому пакету;
- 2) определение ресурсов, на которых планируется запуск пакета;
- 3) регистрация в базе пакетов и описание на специальном унифицированном языке EasyPackage.

Если первые два этапа осуществляются администратором платформы CLAVIRE и, в общем случае, выполняются только один раз, то описание пакета может создать сам пользователь. Причем процедура описания пакета может быть выполнена неоднократно: каждый раз, добавляя, удаляя или изменяя состав используемых параметров, пользователь, тем самым, расширяет или сужает функциональность своего пакета.

Для практической реализации данной технологии разработан интеллектуальный редактор CLAVIRE/PackageManager, который предназначен для автоматизации процесса описания пакета и упрощения процедуры его встраивания. Собственно „интеллектуальность“ выражается в том, что пользователь освобожден от необходимости разрабатывать описание приложения на языке EasyPackage; он работает в диалоге с системой, отвечая на вопросы в терминах предметной области, в ходе чего автоматически формируется описание. Для обратной операции — процесса редактирования пакета — требуется лексический и синтаксический разбор файла описания пакета.

Интеллектуальный редактор PackageManager [5] реализует следующие функциональные возможности:

— просмотр доступных пользователю пакетов; доступность пакетов определяется в соответствии с назначенными пользователю правами;

— просмотр описания пакета как в виде текста на языке EasyPackage, так и в виде конструктора; описание пакета в виде конструктора представляет собой совокупность полей, каждое из которых описывает соответствующий атрибут;

— редактирование пакета; данная функция сводится к редактированию информации об атрибутах, описывающих входные и выходные данные о пакете;

— создание нового пакета; при создании пакета пользователю предоставляется полный спектр полей и атрибутов для заполнения, а также автоматически назначаются права доступа к создаваемому пакету;

— удаление пакета; данная функция доступна только администратору базы пакетов;

— запуск проблемно-ориентированного интерфейса в контексте текущего пользователя; данная функция предназначена для тестирования запуска пакета в среде CLAVIRE; компонент данного интерфейса в зависимости от описания пакета в базе пакетов динамически строит web-интерфейс, который представляет собой набор полей для заполнения, необходимых для успешного выполнения программного пакета.

Интеллектуальный редактор CLAVIRE/PackageManager реализован на языках JavaScript и HTML. Он может быть использован как web-приложение через стандартный Интернет-браузер.

Разработка предметно-ориентированных сервисов для организации виртуальных рабочих пространств. На основе интеллектуального редактора CLAVIRE/PackageManager привлеченными специалистами в рамках проекта разработаны предметно-ориентированные сервисы CLAVIRE для различных предметных областей. В частности, ими выполнены следующие работы.

Создание элементов ВРП для решения задач математической физики с помощью прикладного пакета решения задачи Дирихле для оператора Лапласа. Пакет `dirichlet_solver` позволяет осуществлять численный расчет решения задачи Дирихле для оператора Лапласа с помощью метода Якоби. Особенностью пакета является нестандартное размещение данных в системах с распределенной памятью (размещение данных с перекрытием). Такое распределение позволяет существенно увеличить производительность расчетов за счет сокращения количества пересылок данных. На вход пакета принимаются текстовые файлы с описанием граничных условий задачи, значений размерности сетки по всем трем осям, количества итераций метода Якоби и размера перекрытия. На выходе пакета формируется группа файлов, каждый из которых содержит результирующую сетку для заданного процессора.

Создание элементов ВРП для решения задач математической физики с помощью прикладного пакета `ipmeh-mhd-sim`. Пакет `ipmeh-mhd-sim` предназначен для исследования ударных магнитогидродинамических волн и влияния на них возмущений набегающего потока. Исследование основывается на решении уравнений магнитной газовой динамики. При разработке пакета использовались следующие численные методы решения: метод Рунге — Кутты, модифицированная система уравнений магнитной гидродинамики для исключения численного магнитного заряда. Исходными данными пакета служит текстовое описание вычислительного задания, а выходными данными являются распределения физических величин в расчетной области.

Создание элементов ВРП для решения задач в области генерации параллельного кода для вычислительной архитектуры с распределенной памятью. Пакет `OpsMpiParallelizer` является средством автоматической генерации параллельного MPI-кода на базе „Диалогового высокоуровневого оптимизирующего распараллеливателя“. Основу работы пакета составляют методы блочно-аффинного размещения массивов в распределенной памяти. На вход пакета принимаются программы на языке C, соответствующие стандарту C99. По ходу выполнения пакета осуществляются вывод диагностической информации в стандартный выходной поток и запись в выходной файл измененной входной программы.

Создание элементов ВРП для решения задач моделирования процесса функционирования телекоммуникационных сетей с помощью прикладного пакета `NetMax`. Пакет `NetMax` предназначен для решения задачи рационального использования ресурсов телекоммуникационных сетей. При разработке пакета реализован подход к заполнению телекоммуникационных сетей, основанный на использовании их новых характеристик при выборе каналов связи

и применении аппарата теории потоков. В этом контексте является обоснованным применение процедур имитационного моделирования, когда прокладываются все маршруты, и в каждый момент времени можно абсолютно точно определить состояние любого канала в сети.

Создание элементов ВПП для решения задач персонифицированной медицины с помощью прикладного пакета ANN. Пакет ANN (AUC for Native Network — AUC для нативной сети) предназначен для решения задачи построения ROC-кривой и вычисления площади под ней (Area under the Curve — AUC) для байесовской сети (БС) с нативной топологией. Решение задачи основывается на использовании наиболее тривиальной — нативной — топологии БС. При разработке пакета применялись алгоритм прямого безытерационного обучения нативной БС и стандартный алгоритм построения ROC-кривых с возможностью регулирования числа разбиений. Исходными данными пакета являются файлы с описанием нативной БС, база данных для обучения и параметры ее разбиения, информация о необходимости вывода ROC-кривых и префикса выходных файлов. Результатом работы пакета ANN является набор файлов, содержащих ROC-кривые, таблицы для их построения и значения AUC.

Создание элементов ВПП для решения задач молекулярного моделирования с помощью прикладного пакета OPTM. Пакет предназначен для решения задачи локальной энергетической оптимизации молекулярных структур в рамках силового поля MMFF94 (Merck Molecular Force Field). Данный вид оптимизации позволяет устранять энергетически напряженные области в молекуле, приводя молекулу к экспериментально более правдоподобным конформациям. В качестве метода оптимизации используется метод локальной оптимизации дифференцируемой функции многих переменных — LBFGS (Broyden—Fletcher—Goldfarb—Shanno). Входными данным к пакету являются файл параметров оптимизации и файл с начальной геометрией оптимизируемой молекулы. В ходе выполнения пакета в стандартный поток вывода предоставляются текущие сведения о состоянии процесса оптимизации и осуществляется запись в выходной файл конечной геометрии оптимизированной молекулы.

Создание элементов ВПП для решения задач молекулярной биологии с помощью прикладного пакета Autodock. Пакет Autodock предназначен для решения задачи молекулярного докинга — расчета и предсказания поведения биомолекулярных комплексов (соединения из двух органических молекул, одна из которых является „мишенью“ или макромолекулой, другая, которая значительно меньше, — лигандом). Под расчетом поведения комплекса понимается определение положения, которое занимает лиганд в поле макромолекулы, энергии их взаимодействия и других физических и химических параметров комплекса. Исходными данными пакета являются описание строения молекул в формате PDBQT, параметры построения сеточной модели энергетического поля молекулы, а также параметры алгоритма поиска. Выходными данными пакета является отчет, содержащий полную информацию о проведенной процедуре докинга и возможные варианты решения этой задачи.

Создание элементов ВПП для безошибочного решения задач вычислительной математики с помощью прикладного пакета Maxima. Система компьютерной алгебры и символьных вычислений Maxima предназначена для безошибочного решения задач вычислительной математики. Данный пакет обеспечивает символьное дифференцирование, интегрирование и разложение в ряд, позволяет решать обыкновенные дифференциальные уравнения и системы линейных уравнений, а также предоставляет широкий набор средств работы с математическими объектами. На вход пакета подаются строка с командой интерпретатора Maxima, текстовый файл с подпрограммами на языке Maxima Script и файл с данными в Lisp-формате. Выходными данными пакета являются строка с кратким результатом и файл в Lisp-формате с результатами решения указанной в команде задачи.

Создание элементов ВПП для решения задач автоматического распараллеливания кода с помощью прикладного пакета „разрезания“ цикла. Пакет loop_distribution позволяет решить задачу „разрезания“ тесного „гнезда“ цикла с использованием вспомогательных

эквивалентных преобразований („растягивание скаляров“ и „введение временных массивов“). Этот пакет может быть использован для „разрезания“ нераспараллеливаемых циклов, так как результирующие циклы могут оказаться пригодными для параллельного выполнения или для других распараллеливающих преобразований. Пакет был реализован в рамках „Открытой распараллеливающей системы“, а затем модифицирован для проекта „Диалоговый высокоуровневый преобразователь“. На вход пакета принимаются программы на языке С, соответствующие стандарту С99. На выходе пакета выводится файл с результирующей программой на языке С, в которой к числу переменных могут быть добавлены вспомогательные одномерные массивы, необходимые для преобразований „растягивание скаляров“ и „введение временных массивов“.

Создание элементов ВРП для решения задач вычислительной математики с помощью прикладного пакета „Оптимизированный алгоритм распределенного блочного перемножения матриц“. Пакет с рабочим названием “block_matrix_mult” разработан с применением метода блочного распределения матриц в общей памяти, который позволяет увеличить эффективность использования кэш-данных и буфера ассоциативной трансляции (Translation Lookaside Buffer — TLB). Также в пакет заложены методы оптимизации программного кода: векторизация вычислений, тайлинг (оптимизирующее преобразование кода к блочному виду) с учетом кэш-памяти нескольких уровней, выравнивание данных и кода, устранение ветвлений, устранение зависимостей между инструкциями, распараллеливание по данным с помощью библиотеки pthreads. Исходными данными пакета является файл, содержащий информацию о размере перемножаемых матриц и значения их элементов; выходными данными — выходной файл с результирующей матрицей.

Таким образом, разработки элементов ВРП, проведенные с использованием разработанной выше технологии, показывают ее применимость для различных предметных областей.

Заключение. Апробация представленной технологии для разработки элементов ВРП путем автоматизации встраивания прикладных пакетов пользователей в облачную среду на основе платформы CLAVIRE проведена в рамках проекта по обеспечению межвузовской мобильности молодых ученых, входящего в Федеральную целевую программу „Научные и научно-педагогические кадры инновационной России“ на 2009—2013 гг. Это позволило не только проверить применимость разработанной технологии для задач различных предметных областей, но и выявить ряд технических и методологических недостатков инструментального средства CLAVIRE/PackageManager в целях его дальнейшей доработки и развития.

Статья подготовлена по результатам работы, выполненной в рамках Федеральной целевой программы „Научные и научно-педагогические кадры инновационной России“ на 2009—2013 гг., соглашение № 14.В37.21.2085 от 14.11.2012 г.

СПИСОК ЛИТЕРАТУРЫ

1. Foster I., Zhao Y., Raicu I., Lu S. Cloud Computing and Grid Computing 360-Degree Compared // eprint arXiv:0901.0131, 2008 [Электронный ресурс]: <<http://arxiv.org/ftp/arxiv/papers/0901/0901.0131.pdf>>.
2. Бухановский А. В., Васильев В. Н. Современные программные комплексы компьютерного моделирования e-Science // Изв. вузов. Приборостроение. 2010. Т. 53, № 3. С. 60—64.
3. Федеральная целевая программа „Научные и научно-педагогические кадры инновационной России“ на 2009—2013 гг. [Электронный ресурс]: <<http://www.fcpr.ru/Default.aspx>>.
4. Духанов А. В. и др. Облачные технологии в задачах интерактивной 3D-визуализации: опыт организации межвузовской мобильности молодых ученых // Информационно-измерительные и управляющие системы. 2012. № 11(10). С. 25—33.
5. Болгова Е. В. и др. Автоматизация процесса разработки виртуальных лабораторных практикумов на основе облачных вычислений // Там же. 2012. № 11(10). С. 71—80.

6. UniHUB — Университетский кластер [Электронный ресурс]: <www.unihub.ru>.
7. Смирнов С. А. Применение алгебраических языков моделирования в системе распределенного решения задач глобальной оптимизации // Тр. 53-й науч. конф. МФТИ „Современные проблемы фундаментальных и прикладных наук“. М.: Изд-во МФТИ им. Н. Э. Баумана, 2010. Т. 9. С. 55—56.
8. Ковальчук С. В. и др. Организация человеко-компьютерного взаимодействия в средах компьютерного моделирования на базе облачной инфраструктуры // Прикладная информатика. 2012. № 5 (41). С. 89—102.

Сведения об авторах

- Алексей Валентинович Духанов** — канд. техн. наук, доцент; НИИ наукоемких компьютерных технологий Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики;
E-mail: dukhanov@niuitmo.ru
- Екатерина Владимировна Болгова** — НИИ наукоемких компьютерных технологий Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики; мл. науч. сотрудник;
E-mail: bolgova@gmail.com
- Лев Романович Гервич** — аспирант; Южный федеральный университет, кафедра алгебры и дискретной математики, Ростов-на-Дону;
E-mail: lgervith@gmail.com
- Василий Григорьевич Колпаков** — аспирант; Институт проблем передачи информации им. А. А. Харкевича РАН, отдел распределенных вычислений, Москва;
E-mail: vasilykolpakov@gmail.com
- Евгений Николаевич Кравченко** — аспирант; Южный федеральный университет, кафедра алгебры и дискретной математики, Ростов-на-Дону;
E-mail: peon_sxe@gmail.com
- Илья Ильич Курочкин** — Институт системного анализа РАН, отдел распределенных вычислений, Москва; ст. науч. сотрудник; E-mail: qurochkin@gmail.com
- Евгений Дмитриевич Масленников** — Московский государственный университет им. М. В. Ломоносова, научно-исследовательский вычислительный центр; техник;
E-mail: ixidiciti@gmail.com
- Игорь Владимирович Офёркин** — Московский государственный университет им. М. В. Ломоносова, научно-исследовательский вычислительный центр; техник;
E-mail: oferkin@physics.msu.ru
- Антон Олегович Рубцов** — Институт системного анализа РАН, отдел распределенных вычислений, Москва; ст. науч. сотрудник; E-mail: rubtsov493@gmail.com
- Сергей Андреевич Смирнов** — Институт системного анализа РАН, отдел распределенных вычислений, Москва; инженер; E-mail: sasmir@gmail.com
- Олег Борисович Штейнберг** — Южный федеральный университет, кафедра алгебры и дискретной математики, Ростов-на-Дону; ассистент;
E-mail: olegsteinb@gmail.com
- Михаил Викторович Юрушкин** — аспирант; Южный федеральный университет, кафедра алгебры и дискретной математики, Ростов-на-Дону;
E-mail: m.yurushkin@gmail.com

Рекомендована НИИ
наукоемких компьютерных
технологий

Поступила в редакцию
12.02.13 г.