



ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

С.В. Востокин

ОБЗОР ПРОЕКТА АВТОМАТИЗАЦИИ ПАРАЛЛЕЛЬНЫХ И РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ TEMPLET

(Самарский университет)

Введение. В работе представлены назначение, цели и задачи, основные результаты исследований по проекту автоматизации параллельных и распределенных вычислений Templet (templet.ssau.ru), который разрабатывается на кафедре программных систем и кафедре информационных систем и технологий Самарского университета.

В проекте Templet изучаются методы разработки программ, способных эффективно исполняться на произвольной совокупности вычислительных ресурсов: вычислительных и графических ядрах процессора; отдельных компьютерах, кластерах и облаках, объединенных сетью Интернет.

Актуальность разработки таких программ, с одной стороны, обусловлена тем, что они используются в мега-проектах для решения задач, требующих сверхбольшой вычислительной мощности. Здесь важен объем вычислений, который сложно или невозможно выполнить даже на суперкомпьютере. Широко известный пример – это обработка данных эксперимента на Большом адронном коллайдере по обнаружению бозона Хигса.

С другой стороны, даже в небольших проектах часто требуется значительная вычислительная мощность. В этом случае решающим фактором является себестоимость вычислительных ресурсов. Получить недорогие вычислительные ресурсы можно за счет использования «не специализированных» ресурсов в период их временного простоя по основному назначению. Примером является зонтичный проект BOINC, который объединяет исследователей из разных отраслей знаний и обычных добровольцев, предоставляющих исследователям свои компьютеры для вычислений посредством сети Интернет.

Помимо пользователей персональных компьютеров и мобильных устройств, простаивающие ресурсы во временное использование могут предоставить облачные провайдеры в виде бесплатных или недорогих виртуальных машин. Другой пример – кластеры и суперкомпьютеры. Обычно это высокозагруженные системы, но в силу их большой производительности объем простаивающих ресурсов также значителен и может быть использован. У предприятий, где много компьютеризированных рабочих мест, также велик объем простаива-



вающих ресурсов. Например, в ночные часы компьютеры предприятия потенциально доступны для решения производственных вычислительных задач.

Проект Templet ставит целью популяризацию перечисленных выше подходов использования вычислительных ресурсов через сеть Интернет. Для этого разрабатывается оригинальная технология программирования в парадигме акторов и задач, упрощающая организацию вычислений на произвольной совокупности ресурсов, в первую очередь в образовательных целях и для автоматизации научных исследований, а в перспективе – для решения широкого круга прикладных задач.

Направления исследований. Проблемы организация вычислений на произвольной совокупности вычислительных ресурсов относятся к области параллельных и распределенных вычислений, включая грид-вычисления, гетерогенные вычисления, многозадачные вычисления, языки и системы управления потоками работ и другие родственные направления компьютерных наук. В проекте Templet разрабатывается синтетический подход, опирающийся на результаты в перечисленных направлениях, особенностью которого является представленная ниже методология организации вычислений с точки зрения модели вычислений, библиотеки времени исполнения, программирования и генерации кода, программной инфраструктуры.

Модель вычислений. В основе проекта лежит модель вычислений, позволяющая сформировать математически точное и самодостаточное описание вычислительного процесса, чтобы на его основе можно было построить программу, исполняющуюся на произвольной совокупности вычислительных ресурсов.

Точное и привычное для программистов описание последовательного вычислительного процесса – это алгоритм. Последовательные алгоритмы самодостаточны в том смысле, что не привязаны к конкретному механизму выполнения. Для описания вычислительного процесса мы используем последовательные недетерминированные алгоритмы специальной структуры, распараллеливание которых позволяет получить искомые программы.

Таким образом, программирование алгоритма для исполнения на произвольной совокупности вычислительных ресурсов понимается как эквивалентное преобразование исходного алгоритма (последовательного либо параллельного) к новому алгоритму специальной структуры. Этот подход близок к модели акторов Хьюитта, алгоритмическим скелетам Коула, вычислениям на основе задач и отражает классическое представление о содержании теории программирования как задачи отображения алгоритма на заданное вычислительное окружение. Подход можно рассматривать как альтернативный по отношению к языкам описания потоков работ (CWL и др.); промежуточному программному обеспечению (HTCondor и др.); набору высокоуровневых алгоритмических скелетов с индивидуальными моделями исполнения (map-reduce и др.).

Библиотека времени исполнения. Особенностью библиотеки времени исполнения является ориентация на синхронизацию параллельных задач на многоядерных/многopроцессорных компьютерах с общей/разделяемой памя-



тью. Для запуска и исполнения задач используются сторонние системы и библиотеки, например, OpenMP, стандартная библиотека языка C++, MPI, платформа Everest (разработка ИППИ РАН). Эта особенность обеспечивает расширяемость для использования с разными системами управления задач, компактность кода, а также позволяет эффективно синхронизировать очень большое количество задач малой длительности счета. Текущая реализация библиотеки времени исполнения выполнена на языке C++.

В области библиотеки времени исполнения в проекте изучаются: методы оптимизация кода библиотеки с целью повышения производительности; производительность на разных платформах в алгоритмах мелкогранулярных вычислений; методы разработки, оптимизации и адаптации под библиотеку времени исполнения интерфейсных модулей к сторонним системам управления задачами.

Программирование и генерация кода. Наш подход позволяет применять любой алгоритмический язык программирования или псевдокод для передачи семантики модели вычислений, но программисту требуется поддерживать определенную структуру алгоритма при написании кода. В результате получающийся код может оказаться довольно объемным. Для преодоления этих сложностей мы применяем автоматизацию программирования, относящуюся к метапрограммированию на основе препроцессора. В проекте разрабатывается набор инструментов, включающий препроцессор кода и кодогенератор, управляемый компактным языком. Язык кодогенератора позволяет лаконично описать свойства типов программы: сообщения, принимаемые и отправляемые акторами; задачи, запускаемые в их контексте. На этой основе автоматически формируется структура программы.

Программная инфраструктура. Исследования в области программной инфраструктуры направлены на практическое применение проекта в образовательных и исследовательских целях. Цель оптимизации инфраструктуры – удобство работы с программами в соответствии с современными требованиями развертывания на произвольных виртуальных машинах в облаке и работы через веб-интерфейс в браузере. Для поддержки этих возможностей в проекте ведутся исследования в области применения технологий системной интеграции на основе: контейнеризации приложений Docker; интерактивной разработки приложений для математического моделирования, анализа данных и искусственного интеллекта проекта Jupyter; платформы Everest для управления запуском приложений на произвольных ресурсах; протокола IPFS для организации распределенного хранения данных в сети Интернет.

Практические результаты. Результаты исследований по проекту отражены в статьях, свидетельствах программ и баз данных, представленных в докторской и трех кандидатских диссертациях по темам проекта [1-4]. В настоящее время используются в образовательном процессе Самарского университета: система управления задачами кластера «Сергей Королёв» (<http://hpc.ssau.ru/node/3130>), которая предназначена для проведения лабораторных занятий, исследовательских работ и других занятий на кластере с исполь-



зованием веб-браузера на простаивающих узлах кластера; комплект для разработки программного обеспечения Templet SDK (<https://github.com/the-templet-project/templet>), использующий модель акторов для синхронизации разнотипных задач с поддержкой платформы Everest, развертыванием на платформе Binder, возможностью работы из веб-браузера.

Научные результаты были получены и использовались в проекте «Разработка комплекса технологий использования ресурсов суперкомпьютера «Сергей Королёв» в целях развития инновационной и научно-образовательной среды университета» (программа повышения конкурентоспособности Самарского университета среди ведущих мировых научно-образовательных центров на 2013-2020 годы, «Проект 5-100»); в проекте «Разработка фундаментальных основ аналитического синтеза регулярных и хаотических процессов в динамике космических аппаратов» (Госзадание № 9.1616.2017/4.6).

Будущие направления исследований в рамках проекта Templet связаны с расширением библиотеки примеров распределенных процессов; совершенствованием системы времени выполнения и метапроцессора; новыми методами анализа кода на корректность и эффективность; совершенствованием программной инфраструктуры; разработкой приложений в области автоматизации учебного процесса и решением других актуальных задач.

Литература

1. Востокин, С.В. Графическая объектная модель параллельных процессов и ее применение в программных комплексах численного моделирования : диссертация ... доктора технических наук : 05.13.18 / Востокин Сергей Владимирович; [Место защиты: Сам. гос. аэрокосм. ун-т им. С.П. Королева]. - Самара, 2007.

2. Литвинов, В.Г. Разработка и применение типовых решений для распараллеливания алгоритмов численного моделирования : диссертация ... кандидата технических наук : 05.13.18 / Литвинов Владимир Геннадьевич; [Место защиты: Сам. гос. аэрокосм. ун-т им. С.П. Королева]. - Самара, 2015.

3. Артамонов, Ю.С. Мониторинг и прогнозирование загрузки кластерной системы на основе адаптивной смеси моделей : диссертация ... кандидата технических наук : 05.13.18 / Артамонов Юрий Сергеевич; [Место защиты: Сам. нац. исслед. ун-т им. акад. С.П. Королева]. - Самара, 2017.

4. Бобылева, И.В. Метод попарной обработки элементов информационных массивов для многозадачных вычислений в гибридном облаке: дис. ... канд. техн. наук : 2.3.5 / Бобылева, Ирина Владимировна; - Самара, 2021.