

ВЫСОКОПРОИЗВОДИТЕЛЬНЫЙ ПРОГРАММНЫЙ КОМПЛЕКС МОДЕЛИРОВАНИЯ ЭКСТРЕМАЛЬНЫХ ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ЯВЛЕНИЙ: ЭФФЕКТИВНОЕ УПРАВЛЕНИЕ ПАРАЛЛЕЛЬНЫМИ ВЫЧИСЛЕНИЯМИ И ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ РАСЧЕТОВ

С.В. Ковальчук, С.В. Иванов, Л.И. Лопатухин, А.В. Бухановский

Успех разработки высокопроизводительного программного комплекса моделирования экстремальных гидрометеорологических явлений, ориентированного на вычислительные системы терафлопной производительности [1], в первую очередь, упирается в выбор квалифицированного подхода к его проектированию. Традиционно высокоуровневое проектирование ставит своей целью формализацию и обоснование внутренней структуры программной системы, отвечающей заданным функциональным требованиям с обеспечением в необходимой степени работоспособности, безопасности, безотказности, защищенности [2]. Большинство существующих на сегодняшний день подходов к проектированию программного обеспечения ориентировано на сферу типовых приложений бизнес-информатики, для которых вопросы производительности вычислений в большинстве случаев не являются ключевыми. Однако данный аспект приобретает большую актуальность при разработке наукоемкого программного обеспечения. Необходимо учитывать, что для вычислительных приложений приоритеты проектных требований, их структура, процесс проектирования, создания, тестирования, внедрения и поддержки обладают рядом характерных особенностей.

Одной из наиболее актуальных задач, связанных с разработкой наукоемкого программного обеспечения, является моделирование сложных систем [3] (к которым относится система «океан-атмосфера» с точки зрения описания экстремальных явлений). Это обусловлено ростом вычислительной мощности современных компьютеров, позволяющих решать все более ресурсоемкие задачи, а также востребованностью решения таких задач. Ключевыми особенностями сложных систем в данном случае можно считать большое количество составных элементов, дальние связи между элементами и многомасштабную изменчивость. При разработке программного обеспечения для моделирования таких систем приходится сталкиваться с рядом трудностей, таких как: необходимость использования готовых программных компонентов, организации доступа к разнообразным источникам данных, учет сложной структуры параллельной среды выполнения и пр.

Главной проблемой создания эффективного (с точки зрения параллельных вычислений) программного обеспечения для моделирования сложных систем является то, что вычислительные компоненты, соответствующие разным подсистемам, построены на различных принципах, по-разному распараллеливаются и, зачастую, используют различные технологии параллельного программирования. Учитывая, что компоненты связаны (в общем случае) в нелинейную вычислительную цепочку, возникает необходимость построения иерархии параллельных вычислительных процессов, которая обеспечивала бы наибольшую производительность на конкретной вычислительной системе. Задача осложняется тем, что в общем случае параллельная эффективность вычислительных компонентов зависит от исходных данных и характеристик вычислительной системы. Это делает маловероятной возможность проектирования эффективного параллельного программного обеспечения для моделирования сложных систем без совокупного учета особенностей вычислительных алгоритмов, специфики данных и архитектуры вычислительной системы.

В докладе обсуждается подход к проектированию адаптивного параллельного программного обеспечения для моделирования сложных систем, основанный на динамическом механизме организации иерархии параллельных вычислительных процессов, учитывающем конкретные входные данные и характеристики вычислительной системы (в том числе, их изменчивость во времени). Для построения начального плана вычислительных процессов (принципов распараллеливания и способов декомпозиции данных для каждого вычислительного компонента) механизм использует априорные знания о производительности, представленные в форме параметрических моделей параллельного ускорения. Параметры моделей идентифицируются на основе прямых измерений характеристик вычислительной системы при запуске приложения и уточняются в процессе его выполнения. Это позволяет организовать процесс динамической балансировки вычислительной нагрузки; его необходимость определяется мерой расхождения априорных оценок производительности с результатами текущих измерений. В том случае, когда представление знаний о производительности связано с большой степенью неопределенности (например, архитектура конкретной системы сильно отличается от «идеальной» архитектуры, использованной при построении модели), применяются непараметрические способы получения знаний на основе прямого моделирования параллельных вычислительных процессов с помощью специального компонента-симулятора.

Для реализации описанного механизма предлагается использование сервисно-ориентированной архитектуры [4]. С учетом возможностей динамического построения и реорганизации структуры параллельных вычислительных процессов, в работе производится попытка построения онтологии сервисов, используемых при разработке типовой архитектуры высокопроизводительного программного комплекса моделирования сложных систем. Главную роль в этой онтологии занимает семейство управляющих сервисов, которые реализуют логический вывод с целью определения оптимальной схемы параллельных вычислений. Логический вывод выполняется на основе знаний о производительности, использующих данные об аппаратной архитектуре вычислительного комплекса, текущем состоянии среды исполнения (включая приоритеты работающих задач), способах распараллеливания используемых программных модулей, отвечающих за моделирование составных частей рассматриваемой сложной системы, и схеме их взаимодействия. При этом сама база знаний носит децентрализованный характер в силу того, что она распределяется по различным вычислительным сервисам. Для каждого конкретного сервиса она позволяет определить: а) способ запуска рассматриваемого вычислительного компонента; б) принцип декомпозиции при организации параллельных вычислений; в) априорное ускорение, предназначенное для оценки оптимальности схемы распараллеливания до начала процесса вычислений. Представление базы знаний в форме набора сервисов позволяет организовать унифицированную интеграцию новых вычислительных модулей в созданную схему распараллеливания.

Особенности проектирования и реализации эффективного программного обеспечения для моделирования сложных систем в данной работе иллюстрируются на примере высокопроизводительного программного комплекса моделирования экстремальных явлений в атмосфере и океане, разрабатываемого в рамках проекта ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы». Комплекс содержит четыре вычислительных модуля: модуль усвоения метеорологической информации, модуль гидродинамического моделирования волнения, течений и уровня моря по полям ветра и давления, модуль стохастического анализа и моделирования, модуль расчета рисков от воздействий на морские объекты и сооружения. Каждый модуль включает в себя от 1 до 7 сервисов, которые могут предоставляться независимо от остальных. Предложенный способ организации управляющего сервиса (который осуществляет удаленный запуск и контроль вычислений) позволяет гибко работать с произвольными проблемно-ориентированными модулями, обеспечивая возможность балансировки нагрузки и использования базы знаний о специфике работы различных вычислительных компонентов с целью выбора оптимальной по производительности стратегии распараллеливания. Поддерживаемая сервисом работа с данными в рамках концепции REST позволяет гибко настраивать форматы данных для передачи между различными вычислительными компонентами. Исследования параллельной масштабируемости программного комплекса, проведенные на кластерной системе СПбГУ ИТМО, показали достаточно высокие значения параллельной эффективности (для 128 вычислителей – свыше 90%). Это подтверждает правомерность и целесообразность использования подхода, предлагаемого в докладе.

Специфической чертой программного комплекса моделирования экстремальных явлений в атмосфере и океане является необходимость оптимального представления полученных в результате работы данных ввиду их чрезвычайно большого объема. Для данного примера рассматривается разработка представления результирующей информации в форме интерактивного атласа-справочника. При создании такого атласа необходимо учитывать, что, с одной стороны, наглядное представление данных в виде графиков и таблиц существенно облегчает работу пользователя с системой, а с другой — зачастую пользователь предпочитает получить возможность интеграции с разрабатываемой им самим программной системой. В то же время, расчетные данные не являются статическими, что объясняется как изменением или уточнением условий моделирования, так и пополнением за счет выполнения новых запросов на моделирование. В настоящее время, в условиях повсеместного распространения сети Интернет, она становится идеальной средой для распространения информации, в том числе и в сфере научного сотрудничества. Представление информации в форме интерактивного справочника позволяет пользователю в наглядной форме получить представление о результатах работы программного комплекса. В то же время технологии уровня Web-сервисов позволяют с легкостью произвести интеграцию с существующими и разрабатываемыми программными комплексами. При этом интеграция производится на основании утвержденных стандартов, лежащих вне терминологии какого-либо конкретного языка программирования или операционной системы, что позволяет снять ограничение на платформу исполнения клиентских приложений. Современные Web-технологии и подходы (AJAX, Flash, Silverlight и пр.) позволяют произвести интеграцию в справочник потоковых данных с видео- и аудиоматериалами, разработать полностью интерактивные интерфейсы, оптимизирующие время загрузки информации и улучшающие удобство работы с ней, использовать наиболее удобные схемы пользовательских интерфейсов без ограничения стандартным набором компонентов. Таким образом, построение интерактивного справочника является завершающим этапом и позволяет оптимальным образом представить информацию, полученную в результате работы программного комплекса. В докладе будут приведены примеры представления информации о совместных экстремальных характеристиках ветра, волнения, течений и уровня Каспийского моря.

ЛИТЕРАТУРА:

1. А.В. Бухановский, О.И. Зильберштейн, С.В. Иванов, С.В. Ковальчук, Л.И. Лопатухин, С.К. Попов, М.М. Чумаков Моделирование экстремальных явлений в атмосфере и океане как задача высокопроизводительных вычислений // Вычислительные методы и программирование. Том 9. — М.: Издательством МГУ, 2008 — с. 141-153.
2. И. Соммервилл Инженерия программного обеспечения, 6-е изд. / Пер. с англ. — М.: Издательский дом «Вильямс», 2002. — 624 с.
3. Vossara N. Modeling Complex Systems. — Springer New York, 2004. — 397 p.
4. Lublinsky B. Defining SOA as an architectural style. 9 January 2007.
<http://www.ibm.com/developerworks/architecture/library/ar-soastyle/>.