

ОСОБЕННОСТИ РАЗРАБОТКИ ИНТЕРАКТИВНЫХ КОМПОЗИТНЫХ ПРИЛОЖЕНИЙ

К.В. Князьков, С.В. Ковальчук

*Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики*

E-mail: constantinvk@gmail.com

Рассматриваются принципы проектирования и разработки интерактивных композитных приложений для их исполнения в распределенной вычислительной среде. Интерактивные композитные приложения могут использоваться в качестве основы для создания систем визуализации, системы реального времени, а также для обеспечения технологии computational steering. Апробация разработанного прототипа рассматривается на примере сервиса моделирования флэшмоб-акций.

Введение

Формализм workflow широко распространен для представления сложных научных композитных приложений. Описание комплексной задачи в виде графа связанных блоков позволяет автоматически учитывать зависимости между блоками и распараллеливать исполнение на несколько распределенных ресурсов. Во многих существующих системах исполнения workflow используется модель запуска в пакетном режиме, когда сам workflow и его отдельные подзадачи после запуска переходят в режим непрерывного исполнения без возможности взаимодействия с ними [1]. Такой подход прост в реализации и оправдан для решения задач, не связанных с необходимостью работы в реальном времени. Но существуют задачи, которые сложно адаптировать к пакетному режиму работы, либо, если это возможно, адаптация приводит к значительной потере эффективности использования ресурсов. Это задачи: а) требующие обработки данных непрерывным потоком; б) требующие постоянного взаимодействия с внешними объектами и источниками данных; в) где критически важно время получения результата; г) где требуется возможность интерактивного взаимодействия с пользователем.

Вопросы перехода систем, основанных на модели workflow, от пакетного режима функционирования к интерактивному в настоящее время активно исследуются. На текущий момент существует несколько систем исполнения композитных приложений, которые поддерживают интерактивный режим работы. Например, система WS-VLAM [2] позволяет исполнять интерактивные приложения, а также поддерживает потоковую передачу данных. Помимо этого, существует ряд исследований, не связанных напрямую с workflow, но затрагивающих интерактивный режим работы задач в распределенных средах. В том числе, применение потоков данных для связи задач в Грид рассматривается в [3]. В работе [4] рассмотрены вопросы, связанные с реализацией системы на основе Грид, позволяющей планировать запуск и исполнять интерактивные задачи. В [5] приведена модель workflow с применением конвейеризации, основанная на сетях Петри. Публикация [6] посвящена вопросу сравнения моделей пакетного и потокового исполнения для задачи обработки данных с датчиков для определения опасных погодных явлений.

Таким образом, применение модели workflow для описания комплексных вычислительных задач ограничено пакетным режимом исполнения. Данная работа посвящена

вопросу использования общей модели workflow для описания как пакетных, так и сетевых, и интерактивных приложений. В работе представлена расширенная модель workflow — модель интерактивных композитных приложений, которая ориентирована на описание информационно-управляющих систем.

Интерактивные композитные приложения

Использование классической модели workflow для представления информационно-управляющих систем осложнено из-за отсутствия возможностей: интерактивного взаимодействия с пользователями для привлечения экспертов и операторов в работу композитного приложения; непрерывного режима функционирования; работы в реальном масштабе времени; управления реальными объектами.

Для удовлетворения представленным требованиям и для преодоления ограничений вследствие пакетного исполнения вводится расширение модели workflow — *интерактивные workflow*, основанные на нескольких ключевых принципах:

1. поддержка workflow и отдельных блоков, исполняющихся долгое время;
2. поддержка механизмов управления извне поведением исполняющихся заданий и их жизненным циклом;
3. поддержка коммуникации между блоками workflow во время исполнения;
4. возможность изменения workflow во время исполнения за счет сценария workflow, а также за счет внешнего управления.

Первый принцип снимает ограничение пакетного исполнения и позволяет встраивать пакеты, работающие долгое время (например, серверы). Второй принцип позволяет управлять такими пакетами, например останавливать их или менять конфигурацию. Коммуникация необходима для объединения нескольких интерактивных приложений в одно распределенное. Четвертый принцип говорит о возможности структурного изменения композитного приложения во время исполнения, например за счет добавления блоков. Реализация перечисленных принципов позволяет строить системы взаимодействия с пользователем за счет возможностей управления и коммуникации, строить системы визуализации и управления внешними объектами за счет коммуникации, а также системы обработки данных с внешних источников данных. Наличие возможности управления блоками позволяет реализовать принцип *computational steering* для интерактивного управления процессами моделирования в реальном времени.

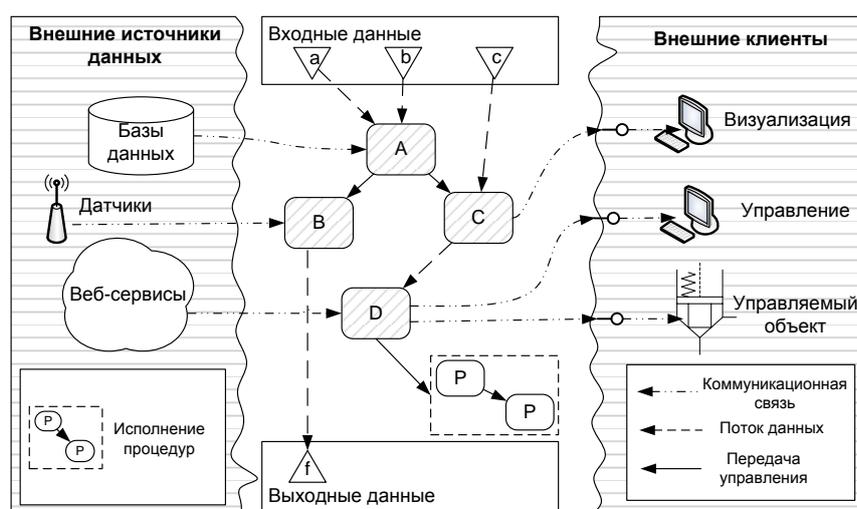


Рис. 1. Модельное интерактивное приложение

На рис. 1 представлена схема модельного интерактивного композитного приложения, на которой представлены коммуникационные связи, а также возможность запуска

пакетных процедур из интерактивных блоков (согласно принципу №4). Из схемы видно, что блоки workflow во время своей работы могут обращаться к внешним, по отношению к платформе, ресурсам для получения информации, а также являться источниками информации для внешних клиентов.

Реализация

Для воплощения принципов модели интерактивных композитных приложений в рамках платформы CLAVIRE [7] был выбран и реализован метод коммуникации узлов, метод передачи управляющих воздействий, а также реализован режим непрерывного исполнения интерактивных блоков. В качестве метода коммуникации был выбран механизм обмена сообщениями между блоками workflow по сети. Для установления коммуникационных зависимостей между блоками workflow используются порты, основной характеристикой которых является протокол, обозначающий способ передачи данных. В ходе реализации прототипа было выявлено, что эффективной абстракцией для коммуникационной связи узлов является парадигма *реактивного программирования*, ориентированная на потоки данных и распространение изменений по портам. Согласно этой парадигме, выходной порт является параметром, который может изменяться в ходе исполнения приложения, при этом присоединенные к нему входные порты других узлов будут синхронно менять свое значение. Для управления интерактивными блоками используется механизм команд, а для взаимного влияния в ходе исполнения введен механизм событий.

В рамках реализации представленной выше модели синтаксис языка описания композитных приложений EasyFlow [8] был расширен за счет новых конструкций: добавлена возможность связывания узлов по портам и объявления интерактивных узлов. Компонент интерпретации workflow модифицирован за счет добавления логики обработки коммуникационной связи, а также динамического конфигурирования запускаемых интерактивных пакетов за счет передачи им конфигурационного файла, содержащего коммуникационные настройки среды. В качестве среды коммуникации и передачи команд был выбран программный сетевой транспортный уровень — ZMQ (zero message queue), в качестве формата сериализации данных — формат BSON. При этом для встраивания пакета длительного исполнения в полученную среду необходимо подключить к нему специальную программную библиотеку (см. рис. 2).

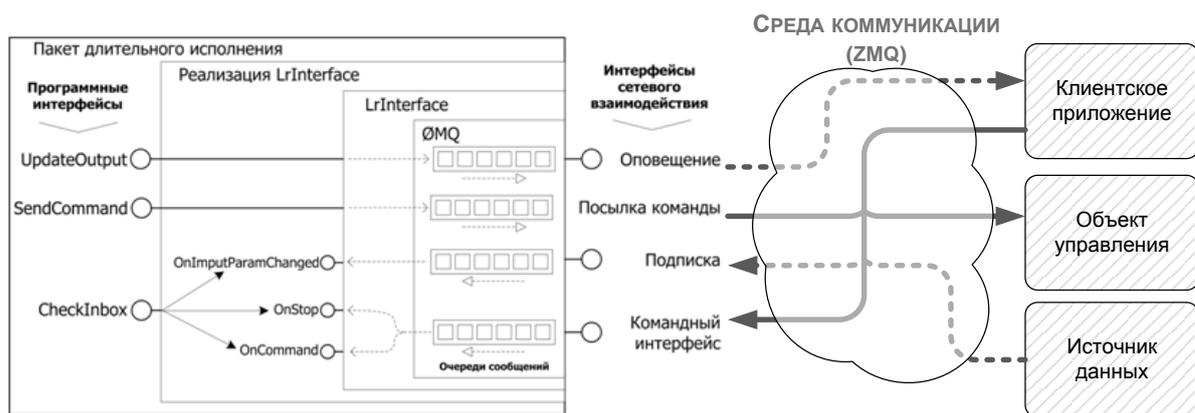


Рис. 2. Структура разработанной программной библиотеки

Пример использования

В рамках апробации предложенной модели и разработанного решения было реализовано композитное приложение — программный сервис моделирования флэшмоб-акций [9]. Композитное приложение предназначено для исследования критических ситуаций, связанных с террористическими действиями в условиях несанкционированных

флешмоб-акций, организованных с использованием социальных сетей. Композитное приложение представляет собой линейный граф из следующих последовательных шагов. По определенным характеристикам (полученным на основе краулинга) генерируется модельная сеть (шаг 1), на основе которой исследуется распространение информации (оповещение об акции). Для того чтобы оценить неопределенность прогноза распространения информации, используется *бутстрепирование* – формируется выборка на основе текущей сети с удалением части случайно выбранных узлов (шаг 2). Далее осуществляется моделирование распространения информации (шаг 3), чтобы в итоге определить интервальную оценку числа оповещенных (шаг 4). По характерным статистическим сценариям (минимальное, максимальное, среднее значение числа участников) выполняется анализ картины паники и эвакуации (шаг 5). На рисунке 3 изображен результат моделирования паники.



Рис. 3. Статическое определение зон с критическим значением плотности агентов



Рис. 4. Окно визуализации во время моделирования

Последний шаг выполняется на основе параллельных процессов моделирования и визуализации (рис. 4). Использование модели интерактивных композитных приложений позволило разделить эти процессы: в данном случае моделирование выполнялось на вычислительно мощном оборудовании, а визуализация осуществлялась в специализированном визуализационном оборудовании. При этом за счет выбранной технологии коммуникации визуализация запускалась параллельно на нескольких компьютерах. Указанные возможности позволяют более эффективно использовать различную вычислительную технику, а также сократить объемы передаваемых презентационных данных и обеспечивают возможность многопользовательского доступа к результатам расчетов.

Заключение

Результатами исследования, представленного в данной работе, являются: принципы построения интерактивных композитных приложений, программная реализация модели в рамках платформы CLAVIRE, апробация решения на сервисе моделирования флешмоб-акций. Полученные в результате апробации и оценки эффективности данные позволяют говорить о применимости модели интерактивных композитных приложений и разработанной программной инфраструктуры для создания систем визуализации, систем поддержки принятия решений, для задач обработки больших объемов данных, где хранение промежуточных файлов недопустимо из-за их большого объема, для создания интеллектуальных систем управления сложными динамическими объектами.

В то же время часть вопросов, связанных с областью интерактивных приложений, остались нерассмотренными. Вопрос взаимодействия интерактивных программных модулей, работающих в различных вычислительных средах, осложняется их неоднородностью: на текущий момент нет унифицированных способов связи приложений в рам-

ках Грид и сред облачных вычислений. Планирование интерактивных приложений не может использовать в качестве основы применяемый на данный момент в платформе CLAVIRE аппарат параметрических моделей производительности [10], т.к. задачи длительного исполнения по определению могут работать бесконечно долго. К тому же необходимо исследование вопроса взаимодействия нескольких композитных приложений в общей среде, так как в реальных задачах может возникнуть необходимость взаимодействия с уникальными реальными объектами.

Литература

1. Deelman E., Gannon D., Shields M., Taylor I. Workflows and e-science: An overview of workflow system features and capabilities // *Future Generation Computer Systems*. 2009. V. 25, № 5. P. 528–540.
2. Wibisono A., Korkhov V., Vasunin D., Guevara-Masis V., Zhao Z., Belloum A. Workshop on Workflow Systems in e-Science // *Lecture Notes in Computer Science*. 2007. Vol. 4489. P. 191–198.
3. Koulouzis S., Meij E., Marshall M.S., Belloum A. Enabling data transport between web services through alternative protocols and streaming // *Proceedings of the 2008 Fourth IEEE International Conference on eScience*. – Washington, DC, USA: IEEE Computer Society, 2008. P. 400–401.
4. Lawenda M., Okon M., Oleksiak A., Ludwiczak B., Piontek T., Pukacki J., Meyer N., Nabrzyski J., Stroinski M. Running Interactive Jobs in the Grid Environment // *Parallel Processing and Applied Mathematics. Lecture Notes in Computer Science*. 2006. Vol. 3911/2006. P. 758–765.
5. Tolosana-Calasanz R., Banares J.A., Rana O.F. Autonomic streaming pipeline for scientific workflows // *Concurrency and Computation: Practice and Experience*. 23, 16 (November 2011). P. 1868–1892.
6. Sims M., Kurose J., Lesser V. Streaming versus Batch Processing of Sensor Data in a Hazardous Weather Detection System // *Proceedings of Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005)*. 2005. P. 185–196.
7. Бухановский А.В., Васильев В.Н., Виноградов В.Н., Смирнов Д.Ю., Сухоруков С.А., Яппаров Т.Г. CLAVIRE: перспективная технология облачных вычислений второго поколения // *Известия вузов. Приборостроение*. 2011. № 10. С. 7–13.
8. Князьков К.В., Ларченко А.В. Предметно-ориентированные технологии разработки приложений в распределенных средах // *Известия вузов. Приборостроение*. 2011. № 10. С. 36–43.
9. ACLab – Демонстрационное композитное приложение моделирования флешмоб-акций – [<http://acl.ifmo.ru/index2.php?ws=142>].
10. Марьин С.В., Ковальчук С.В. Сервисно-ориентированная платформа исполнения композитных приложений в распределенной среде // *Известия высших учебных заведений. Приборостроение*. 2011. №10. С. 21–28.