

С.И. Соболев, К.С. Стефанов

СЕРВИС ВЫПОЛНЕНИЯ ПАКЕТОВ ОДНОПРОЦЕССОРНЫХ ЗАДАЧ НА СУПЕКОМПЬЮТЕРЕ СКИФ МГУ «ЧЕБЫШЕВ»

*Научно-исследовательский вычислительный центр МГУ
имени М.В.Ломоносова, Москва,
sergeys@parallel.ru, cstef@parallel.ru*

Опыт использования вычислительных кластеров коллективного пользования, установленных в НИВЦ МГУ [1], показывает, что часть задач, запускаемых пользователями, требует для работы одно процессорное ядро. Узлы современных кластеров часто имеют 4-8 ядер, а системы управления заданиями настраиваются так, что на одном вычислительном узле в один момент времени исполняется только одна задача. Поэтому при выполнении однопроцессорных задач ресурсы используются крайне неэффективно. Кроме того, однопроцессорные задачи часто запускаются логически связанной серией, однако системы управления заданиями оперируют с отдельными задачами, и пользователю трудно контролировать прохождение всей серии. Описываемый в данной работе сервис выполнения пакетов однопроцессорных задач предназначен как для более эффективного использования ресурсов, так и для логического объединения серии запусков для упрощения контроля.

Создаваемый сервис реализован на основе системы метакомпьютинга X-Com [2, 3]. Запуск серии однопроцессорных задач осуществляется одним запуском специальной задачи X-Com. Каждая порция данных этой задачи X-Com – это один запуск однопроцессорной задачи пользователя. В порции данных передается информация о текущем каталоге, переменных окружения и командной строке запуска однопроцессорной задачи. Кроме того, в описании задачи, передаваемом сервером X-Com клиенту, указывается имя пользователя, с чьими правами должна исполняться эта однопроцессорная задача.

Схема работы сервиса показана на рис. 1.

На каждом вычислительном узле специальной подсистемой сервиса запускается несколько клиентов X-Com, что позволяет эффективно использовать вычислительные ресурсы. Сами клиенты X-Com запускаются с помощью штатной системы управления заданиями Cleo, то есть сервис работает в нормальном режиме функционирования узлов кластера, без выделения подраздела специально для работы сервиса.

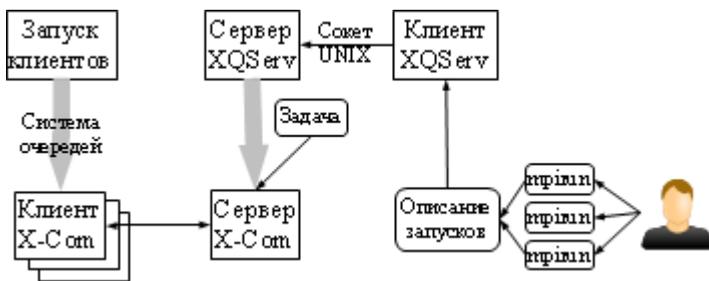


Рис. 1. Схема работы сервиса

Для реализации сервиса потребовалось внести ряд изменений во все компоненты системы X-Com, вошедшие в основную версию.

Изменения клиентской части X-Com связаны с запуском клиентской части прикладной задачи от имени пользователя, отличного от того, который запустил клиент X-Com. Для этого был реализован новый программный интерфейс для взаимодействия вычислительного модуля прикладной задачи с клиентской частью X-Com. Новый интерфейс не требует программирования клиентской части задачи на Perl в файле gtask, вместо это сервер передает клиенту командные строки, вызываемые для инициализации клиентской части приложения и обработки очередной вычислительной порции, а также имена файлов для записи входных данных порции и считывания результатов ее обработки. Кроме того, сервер X-Com передает клиенту имя пользователя, от имени которого задача была поставлена в очередь. Клиент обеспечивает выполнение программ от имени нужного пользователя при помощи программы sudo.

В серверной части системы X-Com также была добавлена поддержка нового программного интерфейса. Имя пользователя, от имени которого должна выполняться прикладная задача, сервер X-Com получает от сервера управления заданиями XQSERV, который в свою очередь получает его при помощи механизма UNIX-сокетов (см. ниже).

Еще одно изменение серверной части X-Com коснулось логики раздачи клиентам вычислительных порций. В обычных случаях сервер X-Com после выдачи узлу последней порций начинает заново раздавать те порции, результаты обработки которых на тот момент не были получены. При проведении расчетов в неоднородных динамических средах такой механизм гарантирует, что при прекращении работы части узлов результат все равно будет получен полностью. Также он позволяет снизить общее время расчета задачи в тех случаях, когда изначально порция попадает на сравнительно маломощный вычислительный узел. При повторной выдаче она может попасть на

более мощный узел, который обработает ее быстрее. Однако для функционирования сервиса распределения однопроцессорных задач на кластере такой механизм неприемлем – предполагается, что вычислительная среда однородна и надежна, а ее ресурсы должны использоваться максимально экономно, т.е. дублирование вычислений недопустимо. Поэтому в настройки сервера X-Com был добавлен ряд опций, запрещающих повторную раздачу вычислительных порций и определяющих максимально время ожидания результатов от клиентов.

В подсистеме управления заданиями XQSERV был реализован режим обмена данными между клиентом и сервером посредством UNIX-сокетов. Ограничения такого режима – необходимость работы клиента и сервера в рамках одной физической машины под управлением ОС UNIX/Linux. Однако функциональность реализуемого сервиса подразумевает работу именно в таких условиях. Обмен данными через UNIX-сокеты позволяет серверу подсистемы управления заданиями автоматически получать информацию о пользователе, вызвавшем клиента для постановки задания в очередь. Имеющийся ранее механизм обмена данными между клиентом и сервером поверх протокола TCP не мог гарантировать корректности передачи такой информации.

Для облегчения перехода пользователей на новый сервис создана утилита, которая преобразует серию постановок однопроцессорных задач в очередь в файл, описывающий задание для сервиса однопроцессорных задач. При запуске эта утилита подменяет команду постановки задачи в очередь (программа `trigger`). Информация о каждом вызове фиксируется, и после окончания постановки всей серии в очередь формируется окончательный выходной файл.

1. Жуматий С.А., Князев А.С. Исследование характеристик потока задач на вычислительном кластере // Научный сервис в сети Интернет: многоядерный мир. 15 лет РФФИ: Труды Всероссийской научной конференции. 2007. М.:Изд-во МГУ. Стр. 139.
2. Вл.В. Воеводин, Ю.А. Жолудев, С.И. Соболев, К.С. Стефанов. Эволюция системы метакомпьютинга X-Com. Параллельные вычислительные технологии (ПаВТ'2009): Труды международной научной конференции (Нижний Новгород, 30 марта - 3 апреля 2009 г.). - Челябинск: Изд. ЮУрГУ, 2009. С. 82-91.
3. <http://x-com.parallel.ru/>