

ПРОБЛЕМЫ СОЗДАНИЯ ПОРТАЛОВ ПОДДЕРЖКИ КЛАСТЕРОВ ДЛЯ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

А.В.Богданов¹, М.И.Павлова², Е.Н.Станкова²

1) АНО Институт высокопроизводительных вычислений и информационных систем,

2) Санкт-Петербургский государственный политехнический университет, Учебно-научный центр по вычислительной механике ИТЭС РАН (СПбГПУ УНЦ ВМ)

Тел.: (812) 247-58-79, e-mail: bogdanov@csa.ru, lena@csa.ru

В настоящее время активно развивается технология высокопроизводительных вычислений для решения различных научных и прикладных задач. Появление вычислительных кластеров дало новый толчок развитию этой технологии, поскольку кластеры являются наиболее дешевыми средствами, предназначенными для высокопроизводительных вычислений, и для их работы используется, как правило, свободно распространяемое программное обеспечение. Количество кластерных вычислительных систем стремительно растёт. В первую очередь, кластерными центрами оснащаются высшие учебные заведения.

Тем не менее, для обычного пользователя работа на высокопроизводительном кластере вырастает в проблему. Вследствие этого, пользователь часто предпочитает потратить больше времени на расчет на персональном компьютере, чем осваивать новые программные продукты и технологии распараллеливания программ. В университетах в программу обучения уже введены курсы высокопроизводительных вычислений. Переподготовка же специалистов старшего возраста практически не ведется.

Основным источником сведений о технологии высокопроизводительных вычислений в настоящее время являются Интернет-ресурсы. Попытка собрать и систематизировать сведения о высокопроизводительных вычислениях и российских центрах, ими занимающихся, осуществлена на сайте, созданном на базе Научно-исследовательского вычислительного центра Московского государственного университета им. М.В.Ломоносова. Задачи создания этого www-ресурса освещены в [2]. Самостоятельно создать обучающие и информационные материалы, естественно, центр не в состоянии. В начале создания он послужил хорошей службой, но в настоящее время уже не может охватить всех проблем, по-видимому, из-за малочисленности своего персонала. Присутствует информация в основном о ресурсах Москвы и Санкт-Петербурга. Ожидалось, что основные материалы будут присылаться сторонними авторами, но, как показывает практика, этого не происходит.

На сайте [3], в создании которого принимали участие авторы, также реализован доступ к широкому кругу образовательных ресурсов, который включает описание высокопроизводительных систем различной архитектуры, электронные пособия по работе с различными операционными системами, компиляторами, системами визуализации, а также описание высокопроизводительных алгоритмов и данные мониторинга тенденций развития суперкомпьютеров и ведущих центров.

Однако, наряду с несомненными достоинствами, информационные ресурсы такого рода обладают рядом недостатков. С одной стороны, они не могут описать все обучающие и информационные материалы, и все имеющиеся высокопроизводительные вычислительные ресурсы, даже на территории России. Так в Интернете присутствует информация в основном о ресурсах Москвы и Санкт-Петербурга. Центры азиатской части практически не представлены. С другой стороны, для пользователя, желающего получить сведения о работе на конкретной вычислительной системе, часть информации является излишней, и в ней трудно ориентироваться.

Выходом из этой ситуации может служить создание специализированного портала, предназначенного для быстрого приобретения навыка работы на конкретном кластере.

В качестве примера зарубежного ресурса такого рода можно привести Интернет-ресурс Корнуэльского университета [4]. На этом сайте приведена таблица, в которой указаны все установленные программные продукты, их версия, инсталляционный статус на данном компьютере, производитель, краткое описание, ссылка на Интернет-страницу программного продукта, сведения о пароле для пользования им. Так же приведен e-mail для вопросов.

Обсудим, какие материалы должен содержать специализированный портал конкретного вычислительного ресурса. Очевидно, минимальным требованием пользователя является наличие следующей информации:

- описание технических характеристик вычислительных ресурсов;
- способы связи с системным администратором (регистрация, получение пароля и т.д.);
- список программного обеспечения, установленного на кластере;
- новости (сообщения об инсталляции новых программных продуктов, профилактических работах и т.д.).

Однако практика показывает, что при наличии только этой информации, пользователь воспринимает такой вычислительный ресурс как «недружественный». «Спаси» пользователя могут только дополнительные материалы сайта. Это, во-первых, описание работы со всеми продуктами, установленными на кластере, а, во-вторых, наличие следующих электронных комплектов пособий и учебников:

- по работе со стандартным программным обеспечением (ПО) для распараллеливания (MPI, PVM, профилировщики и т.д.);
- по методам вычислений, пригодным для работы конкретно на данном кластере;
- по работе с прикладным ПО, установленным на данном кластере (многочисленные библиотеки);
- форум взаимопомощи пользователей с участием куратора-администратора.

Следует отметить, что материалы, указанные в первом пункте, уже в какой-то мере созданы и размещение их, или ссылок на них, не представляет особых сложностей.

На сайте [3], в создании которого принимали участие авторы, также реализован доступ к широкому кругу образовательных ресурсов по высокопроизводительным технологиям.

Размещение материалов, указанных во втором пункте, на первый взгляд является не обязательным, но, как показала практика, часть пользователей со стажем еще не вполне осведомлена о преимуществах высокопроизводительных вычислительных систем и ориентирована на работу на одном процессоре и без использования прикладных библиотек. Эти материалы (или ссылки на них) могут быть предоставлены на сайте вычислительного ресурса самими пользователями (в особенности вузовскими преподавателями). Практика показала, что кроме описания методов вычислений необходимы и материалы по теории параллельных вычислений. Авторы принимали участие в создании нескольких электронных пособий, соответствующих данному разделу: "Многосеточные методы" [5], "Высокопроизводительные алгоритмы" [6].

Хуже дело обстоит с описанием прикладных библиотек. Платные библиотеки обычно имеют описание и даже русифицированы, но вычислительные кластеры ориентированы, прежде всего, на свободно распространяемое ПО, описание которого, как правило, отсутствует. Авторы имели опыт в создании русской документации [7] по пакету Scilab [8]. На это ушло два года, и работа далека от завершения, но она приостановлена в связи с прекращением финансирования. В настоящее время авторы ведут работу по привлечению новых пользователей к работе с данным пакетом, что, возможно, сможет инициировать процесс выделения средств для продолжения этой работы.

Важность создания описаний прикладных программ изначально недооценивается, результатом чего является отсутствие финансирования таких работ (как и вообще работ, связанных с консультационной поддержкой пользователей). Кроме того, такие работы не престижны. В большинстве вычислительных центров работают только системные программисты. Выход в отсутствие финансирования видится в предоставлении сторонним пользователям права бесплатной работы на кластере в обмен на создание программной документации, а также создание банка обмена программной документацией, который мог бы находиться, например, на ресурсе [1]. Кроме того, на каждом кластере должны быть созданы форумы пользователей и ссылки на соответствующие форумы производителей ПО.

Необходимость создания и поддержки специализированных информационных ресурсов при каждом вычислительном центре представляется особенно актуальной в свете широкого распространения GRID-технологий [9], изначально предполагающих наличие гетерогенных вычислительных узлов с нестандартным программным обеспечением. Для вычислений с использованием GRID-технологий особенно важно развивать входные порталы с дружественным интерфейсом и средой программирования. С их помощью пользователь сможет гораздо быстрее создавать, отлаживать и оптимизировать программы, а также обрабатывать полученные данные.

Литература

1. <http://parallel.ru>.
2. <http://www.csa.ru/ihpcdb/tutor.htm#6>.
3. А.Н.Андреев, А.С.Антонов, Вл.В.Воеводин, С.А.Жуматий, М.П.Филамофитский, П.А.Церетели, «Информационно-аналитический центр по параллельным вычислениям в сети Интернет», Сборник трудов конференции «Научный сервис в сети Интернет», 1999.
4. Cornell Theory Center (CTC) <http://www.tc.cornell.edu/Services/Software/index.asp>.
5. Многосеточные методы <http://www.csa.ru/analitik/multigrid/index.htm>.
6. Высокопроизводительные алгоритмы http://www.csa.ru/analitik/distant/q_start.html.
7. Руководство по работе с пакетом SCILAB http://www.csa.ru/~zebra/my_scilab/.
8. Пакет Scilab (производитель Inria) <http://www.scilab.org/>.
9. <http://www.gridclub.ru/>.