

ОРГАНИЗАЦИЯ УПРАВЛЕНИЯ РАСПРЕДЕЛЁННЫМИ КЛАСТЕРНЫМИ СИСТЕМАМИ НА ПРИМЕРЕ GLOBUS TOOLKIT

Астриков Д.Ю.

Научный руководитель – к.т.н., доцент Кузьмин Д. А.

Сибирский федеральный университет

В настоящее время в области высокопроизводительных вычислений одной из актуальных проблем является проблема грамотного распределения нагрузки на вычислительные узлы комплекса, особенно в случаях с кластерами неоднородной структуры. Зачастую возникает ситуация, когда пользовательское приложение «простаивает», так как ждет завершения процесса, распределенного на наиболее медленный или загруженный вычислительный узел. Интеграция отдельных комплексов высокопроизводительных вычислений в сообщества GRID позволяет заметно ускорить процесс решения пользовательских задач. Управление комплексами в таком случае осуществляется с помощью специального программного обеспечения.

На базе института космических и информационных технологий в Сибирском федеральном университете развернут комплекс высокопроизводительных вычислений суммарной производительностью 12 ТФлопс, который активно используется для решения задач с большими объемами входных данных профессорско-преподавательским составом СФУ и других вузов. Он включает в себя такие ресурсы, как суперкомпьютер, производительность которого составляет 9.6 ТФлопс, а также малые кластеры, производительностью 1.2 ТФлопс каждый.

Для обеспечения эффективной работы этого комплекса и удобного взаимодействия пользователей с распределенными вычислительными ресурсами необходим мощный инструментарий, который включал бы в себя систему безопасности, систему управления данными, базы данных, менеджеров ресурсов, планировщиков задач, удобный пользовательский интерфейс и т.д. В качестве такого инструментария можно рассматривать такие программные пакеты, как Globus Toolkit, Condor, Unicore.

Эти пакеты имеют схожую структуру и предоставляют системному администратору примерно одинаковые наборы инструментов для настройки и управления комплексами высокопроизводительных вычислений.

Таким образом, необходимо выяснить, какой из вышеперечисленных пакетов будет наиболее оптимально использован в качестве инструментария для управления распределенными вычислительными ресурсами СФУ. Основные критерии, которыми следует руководствоваться при совершении выбора в пользу того или иного решения:

- Способ распространения. Крайне желательно, чтобы пакет распространялся свободно и имел открытый код;
- Поддержка операционной системы Linux. Комплекс высокопроизводительных вычислений ИКИТ СФУ находится под управлением ОС SLES (Suse Linux Enterprise Server);
- относительная простота установки и настройки пакета;
- полнота сопроводительной документации;
- возможность включения комплекса в уже существующие GRID сообщества;
- возможность эффективного взаимодействия инструментария со сторонними работками и решениями, используемыми для управления комплексом.

Более всего перечисленным выше критериям удовлетворяет Globus Toolkit (GT), разработанный компанией Globus Alliance и поддерживаемый IBM, Oracle и SAP, так

как именно GT на сегодняшний день является стандартом де-факто для российского GRID-сообщества РДИГ (Russian Data Intensive GRID, RDIG).

В пользу GT можно привести следующие аргументы:

- GT – инструментарий с открытым кодом, распространяется свободно;
- GT включает в себя множество мощных инструментов, позволяющих произвести гибкую настройку практически всех систем комплекса;
- проекту оказывается мощная поддержка со стороны крупных компаний разработчиков, таких как IBM, Oracle, SAP;
- GT обеспечивает возможность последующей интеграции комплекса в российское GRID-сообщество РДИГ;
- GT включает в себя удобные интерфейсы взаимодействия пользователей с распределенными вычислительными ресурсами на базе web-сервисов;
- GT включает в себя все необходимые пакеты для обеспечения эффективного выполнения пользовательских задач, такие как планировщики задач, менеджеры ресурсов, хранилища и базы данных, FTP-серверы и другие;
- GT включает в себя мощную систему обеспечения безопасности GSI (Grid Security Infrastructure);
- на сегодняшний день доступен большой объем сопроводительной документации по вопросам установки, настройки и использования инструментария.

Таким образом, использование Globus Toolkit в качестве основного инструментария для обеспечения взаимодействия пользователей и вычислительных ресурсов комплекса высокопроизводительных вычислений СФУ позволит:

- наиболее эффективно использовать распределенные вычислительные ресурсы СФУ;
- обеспечить пользователям удобный и, главное, безопасный доступ к распределенным вычислительным ресурсам СФУ;
- в будущем успешно взаимодействовать с российским GRID консорциумом РДИГ;
- позволит студентам Сибирского федерального университета ознакомиться и получить опыт работы с распределенными вычислительными системами, организованными по принципу GRID архитектуры.

Для реализации этого проекта необходимо:

- предоставить необходимое для работы инструментария аппаратное обеспечение;
- провести установку необходимого для работы инструментария программного обеспечения:
 - OpenLDAP (Lightweight Directory Access Protocol);
 - OpenSSL (Secure Sockets Layer);
 - SASL (Simple Authentication and Security Layer);
- провести установку инструментария Globus Toolkit;
- провести настройку установленного программного обеспечения и его тестирование.

В качестве интерфейса взаимодействия пользователей с ресурсами комплекса наиболее приемлемо будет использовать входящий в комплект поставки GT web-интерфейс WS-GRAM (Web-Service Grid Resource Allocation and Management), так как объективно, на сегодняшний день наиболее удобным представляется взаимодействие посредством сети Интернет.

Таким образом, предоставляя администратору довольно большой объем программных средств с обширными возможностями настройки, Globus Toolkit является гибким и мощным инструментом для управления большими кластерными системами, построенными по принципу GRID архитектуры.