

УДК 004.75

Д.В. Тягунов, Р.К. Кудерметов, М.Ю. Тягунова
Запорожский национальный технический университет, г. Запорожье
кафедра компьютерных систем и сетей
E-mail: dmitry@tyagunov.org.ua, krk@zntu.edu.ua, mary@yurich.org.ua

ДИСПЕТЧЕР ЗАДАНИЙ РАСПРЕДЕЛЕННОЙ КОМПЬЮТЕРНОЙ СИСТЕМЫ

Аннотация

Тягунов Д.В., Кудерметов Р.К., Тягунова М.Ю. Диспетчер заданий распределенной компьютерной системы. Разработана структура диспетчера заданий распределенной компьютерной системы. Описаны особенности программной реализации диспетчера заданий распределенной компьютерной системы. При этом файлы оценки заданий предложено формировать в формате XML, задания представляют собой программы, разработанные на языке Java.

Ключевые слова: распределенная компьютерная система, GRID, диспетчеризация заданий, ресурсный запрос, XML.

Общая постановка проблемы.

Проблема управления ресурсами вычислительных систем решается уже долгое время, при этом постоянно находятся все новые и новые способы повышения эффективности управления. Одной из наиболее известных и имеющих острую потребность нахождения такого решения является GRID-система — система, состоящая из множества гетерогенных узлов, в качестве которых могут выступать, как локальные сети, кластеры, так и отдельные компьютеры [1,2].

Управление ресурсами такой огромной системы, какой является GRID-система, производится на так называемом «глобальном» и «локальном» уровне. Остановимся на последнем варианте. Для этого случая очень часто более эффективным является разработка и использование собственного диспетчера управления заданиями – планировщика распределения вычислительной нагрузки внутри компьютерной сети.

Существует множество различных планировщиков заданий, которые обладают как достоинствами, так и недостатками. Поэтому есть смысл в разработке собственной планировщика (диспетчера) заданий.

Все поступающие задачи для выполнения в GRID условно можно разделить на такие три подгруппы:

- научные;
- инженерные;
- коммерческие.

На данном этапе развития GRID в Украине преимущественным большинством задач, для решения которых используется GRID, являются задачи именно научного характера. Поэтому при решении вопроса распределения заданий между ресурсами GRID экономическую составляющую можно не учитывать, а в дальнейшем, при возникновении в этом такой необходимости, ввести ее, к примеру, в виде определенного коэффициента, влияющего на место задачи в очереди заданий на выполнение.

Целью работы стала разработка программного комплекса, позволяющего оценить эффективность работы предложенного диспетчера распределения заданий в системе GRID по сравнению с существующим и используемым на данном этапе развития GRID программным обеспечением.

При этом, естественно, разработанный программный продукт не сравнится по полноте и функциональности, к примеру, с gLite. Однако его применение в некоторых случаях является вполне оправданным и имеет свои преимущества. Так, к примеру, разработанная система не ограничивает в выборе операционной системы, которая должна быть установлена на выполняющем задачу компьютере, поскольку язык Java, на котором реализован диспетчер, является кроссплатформенным. Кроме того, существующий на данном этапе набор задач для решения в GRID-системе в Украине, имеющие научный либо же академический характер, вполне устраивает то количество параметров оценки, которое предусмотрено в разработанном диспетчере. И такого рода «сокращения» также имеют свое положительное влияние на производительность и эффективность работы диспетчера распределения заданий.

Структура диспетчера заданий.

Проектирование структуры диспетчера заданий необходимо для его реализации. Разработанная структура системы диспетчеризации заданий учитывает предложенные ранее модели, методы и алгоритмы решения задачи планирования нагрузки в распределенных компьютерных системах [3] и включает следующие модули:

1. Модуль сбора статистической информации. Предоставляет сведения о типе процессора, его тактовой частоте, доступной оперативной памяти компьютера, свободном (или доступном, т.е. которым разрешено воспользоваться) объеме памяти на жестком диске.

2. Модуль предоставления информации о заданиях. Представляет собой файл оценки заданий или ресурсный запрос. Предоставляет информацию о ресурсах, необходимых для выполнения задания. Все параметры устанавливаются минимально требуемыми.

3. Модуль группировки заданий и компьютеров. Группирует компьютеры и задания, которым требуются предоставляемые ресурсы этих компьютеров, в одну группу. Группировка происходит, начиная с заданий, требующих больших ресурсов. Дальнейшая диспетчеризация будет проходить именно в пределах группы.

4. Основной модуль диспетчеризации заданий. Данный модуль отвечает непосредственно за этап диспетчеризации заданий, основываясь на оценочных данных о компьютерах, на которые возможно распределение заданий для выполнения, и заданиях, поступающих для распределения, на сервер.

5. Клиентский модуль диспетчеризации заданий.

Рассмотрим более подробно функции клиентского модуля диспетчеризации заданий. В функции клиентского модуля диспетчеризации входят:

Запуск принятого задания на выполнение;

1. Пересылка полученных результатов на серверную часть диспетчера по окончании вычислений.

2. Сообщение, в случае необходимости, об ошибке вычислений или других исключительных ситуациях, к примеру, сбой в системе.

Основной отличительной особенностью структуры системы диспетчеризации заданий является наличие модуля группировки заданий и компьютеров.

На рис. 1 приведена структура системы диспетчеризации заданий.

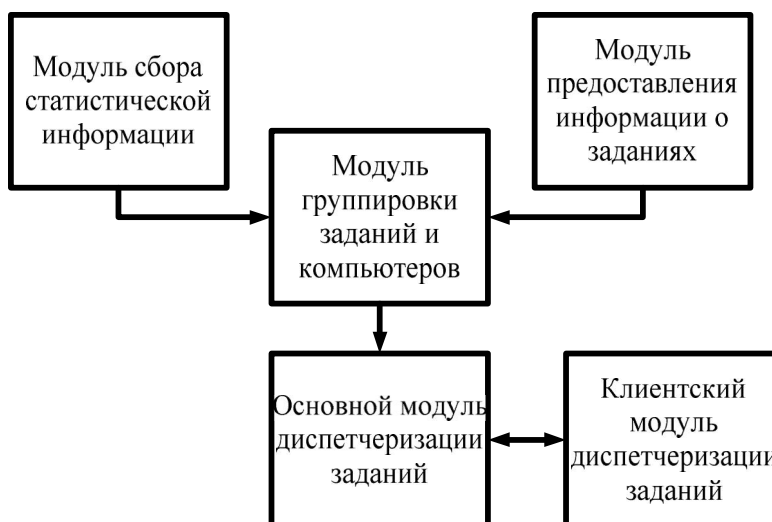


Рисунок 1 — Структура системы диспетчеризации заданий

Программная реализация диспетчера распределения заданий.

Для программной реализации диспетчера распределения заданий в системе выделим следующие модули:

1. Клиентский модуль - включает описательный модуль для человека, желающего поместить для выполнения в GRID-системе свою задачу. (Модуль, связывающий заказчика с самим диспетчером задач).
2. Непосредственно диспетчер: серверная часть – устанавливается на сервере сети.
3. Непосредственно диспетчер: клиентская часть – устанавливается на рабочих станциях (компьютерах) сети.

Клиентский модуль. Пользователю необходимо сообщить следующие данные:

1. Сервер, на который будет послана задача для распределения между компьютерами.
2. Файл, который отсылается в систему для выполнения – программа, разработанная на языке Java;
3. Информация оценки задания.
4. Максимальное время хранения результатов вычислений диспетчером (ограничено диспетчером).

Функции диспетчера (серверная часть):

1. Принимает исполняемый файл задачи (программу) и данные к задаче (файлы, содержащие необходимые для работы программы данные);
2. Принимает информацию об оценке поступающих для решения задач (файл оценки заданий);
3. При помощи реализованного алгоритма распределения определяет, какое задание, какому компьютеру назначить (для этого используется математическая модель, описанная в [4]).
4. Отсылает результаты заказчику или хранит их определенное время, в случае невозможности соединения с компьютером заказчика.

Функции диспетчера (клиентская часть) полностью совпадают с описанными ранее функциями клиентского модуля диспетчеризации.

Остановимся подробнее на том, в каком формате должны приходить файлы для выполнения в данной среде.

Исполняемый файл приходит в формате .jar - формат файлов архива Java, дающий возможность собирать многие файлы в один архивный файл. JAR-файл содержит файлы

классов и дополнительные ресурсы, связанные с апплетами и приложениями. Применение такого формата обусловлено следующими преимуществами [5]:

1. Сжатие. Формат JAR позволяет сжать файлы для экономии памяти;
2. Запечатывание пакета. Пакеты, записанные в JAR-файлы, могут выборочно быть запечатаны, чтобы усилить проверку целостности версии в пакете. Запечатывание пакета в JAR-файл означает, что все классы, определенные в пакете, должны находиться в этом же JAR-файле;
3. Управление версиями в пакете. JAR-файл может хранить данные о файлах, содержащихся в нем, такие как информацию о производителе и версии.
4. Переносимость. Механизм работы с JAR-файлами является стандартной частью базового API платформы Java.

Данные присылаются отдельным файлом с прочей информацией о поступающей задаче.

Для того, чтобы решить свою задачу в системе GRID, пользователю необходимо сделать следующее.

Для начала необходимо зайти на сайт и зарегистрироваться. При этом выполняется стандартная процедура регистрации, результатом которой является получение пользователем логина и пароля в случае успешного ее прохождения.

Следующим шагом является вход в систему. После чего пользователь попадает на интернет-страницу со следующими функциональными возможностями:

1. Добавить задачу.
2. Посмотреть результаты выполнения.
3. Настроить свой профиль.
4. Посмотреть справочную информацию.
5. Выйти из профиля.

Данная страница в будущем в зависимости от возникающих потребностей может быть дополнена и другими функционалами.

При добавлении задачи необходимо указать следующее:

- путь к исполняемому файлу – в формате .jar;
- путь к файлу описания и данных – в формате XML.

Описание требований заданий.

Вместо способа описания требований заданий в виде ресурсного запроса предлагается использовать так называемые «описательные» файлы, в которых будет содержаться информация, необходимая для проведения оптимизации распределения заданий между компьютерами вычислительной системы. Назовем эти файлы, как файл оценки задания (ФОЗ) и файл оценки компьютера (ФОК), в которых будет содержаться информация о приходящем для решения в системе задании и описании самого компьютера (вычислительного узла) системы соответственно [6].

В качестве альтернативы ресурсному предлагаем использовать ФОЗ и ФОК со следующими параметрами. Для ФОЗ:

- объем необходимой памяти для решения задания;
- длительность выполнения задания;
- тип данных, используемых в задании;
- число требуемых процессоров.

Для ФОК параметры следующие:

- объем имеющейся памяти компьютера;
- вычислительная мощность компьютера;
- тип данных, с которыми эффективнее работает данный компьютер;
- число процессоров.

Рассмотрим некоторые из этих параметров. Под объемом необходимой памяти для решения задания понимаем минимальный объем памяти, требуемый для решения задания. Указание такого параметра, как тип данных, является важным, поскольку производительность процессора компьютера отличается при работе с разными типами данных [10].

Для описания данных к задаче и описания оценочных данных задачи был выбран формат в виде XML-файла по следующим причинам. XML облегчает компьютеру задачу создания и чтения данных, обеспечивая при этом однозначность их структуры. XML позволяет избежать распространенных ошибок проектирования языков: он расширяем, независим от платформы, включает поддержку интернационализации и локализации. XML полностью совместим с Unicode и свободен от лицензирования. XML не накладывает никаких ограничений на семантику данных, а текстовая природа XML обеспечивает высокую степень переносимости данных. К тому же существует возможность создавать и редактировать XML-документы в любом стандартном текстовом редакторе [7, 8].

Так называемая «заготовка» xml-файла изначально будет предоставлена пользователю. Ему необходимо будет только заполнить своими данными содержимое соответствующих тегов (которые так же, как и функциональные возможности под профилем пользователя, могут быть в дальнейшем расширены).

В тегах xml-файла содержатся следующие данные:

1. Оценочные параметры задачи – данные ФОЗ.
2. Значения данных, необходимых для вычисления всего, что передано в поступившем jar-файле.

Основываясь на оценочных параметрах файлов и компьютеров, находящихся под управлением выбранного сервера, диспетчер и определяет, на какой компьютер, какая задача будет распределена с учетом оптимизации времени выполнения заданий.

При просмотре результатов выполнения, пользователь увидит следующие поля формы отчета:

- номер задачи в общем списке задач;
- название задачи;
- автор задачи;
- профиль, под которым она была загружена;
- сервер, который за нее «отвечает»;
- ресурсы, на которых она запущена (или уже выполнена);
- итог работы (выполнена или нет, на каком этапе выполнения находится и т.п.);
- время решения задачи.

Наличие возможности просмотра справочной информации не только поможет пользователю корректно воспользоваться предоставленными возможностями, но и даст более широкую информацию о GRID в целом.

Выводы.

1. Разработана структура системы диспетчеризации заданий, которая использует разработанные методы диспетчеризации в распределенных компьютерных системах, основной особенностью которой является наличие модуля группировки заданий и компьютеров.

2. Описаны особенности программной реализации диспетчера заданий распределенной компьютерной системы.

3. Предложено формировать файлы оценки заданий в формате XML, задания – в виде программ, разработанных на языке Java.

Литература

1. Состав и функции системы диспетчеризации заданий в гриде с некластеризованными ресурсами [Электронный ресурс] / П.С. Березовский, В.Н. Коваленко - Режим доступа: http://www.keldysh.ru/papers/2007/prep67/prep2007_67.html
2. Информационный центр GRID-вычислений [Электронный ресурс] - Режим доступа: <http://www.GRIDcomputing.com/>
3. Юрич М.Ю. Гибридный алгоритм распределения задач в вычислительной системе / М.Ю. Юрич // Сборник научных трудов Донецкого национального технического университета. Серия: Информатика, кибернетика и вычислительная техника. – 2010. – С. 66-72.
4. Юрич М.Ю. Подход к оптимальному распределению заданий в вычислительной системе / М.Ю. Юрич // Комп'ютинг. - 2008. - Том 7. – Вип. 1. - С. 27-34.
5. Саммерер А. Электронная библиотека НТУ «ХПИ» / А. Саммерер // Файлы JAR. [Электронный ресурс] Режим доступа: http://khpriip.mipk.kharkiv.edu/litbrary/extent/prog/index.html#pgm_java
6. Юрич М.Ю. Основные параметры файлов оценки в GRID / М.Ю. Юрич // Збірник тез доповідей щорічної науково-технічної конференції серед студентів, викладачів, науковців, молодих учених і аспірантів «Гиждень науки», 13–17 квітня 2009 р., Запоріжжя. – 2009. – Т.1. – С. 293-294.
7. Информационный портал IT-компания CompuTel. [Электронный ресурс] - Режим доступа: <http://raieigh.ru/pub/2001/XML-in-10-points.html#>.
8. Официальный сайт поддержки проекта cms Kate для платформы Asp.Net 2.0 [Электронный ресурс] - Режим доступа: <http://cms.xd24.ru/>.

Надійшла до редакції:
01.02.2011

Рекомендовано до друку:
д-р техн. наук, проф. Скобцов Ю.О.

Abstract

Tyagunov D.V., Kudermetov R.K., Tyagunova M.Yu. Task Manager of the distributed computer system. Developed the structure of the Task Manager of the distributed computer system. Described the features of the software implementation of the Task Manager of the distributed computer system. The file assessment tasks proposed to form in XML, the task is a program developed with Java.

Keywords: distributed computing system, GRID, scheduling tasks, resource query, XML.

Анотація

Тягунов Д.В., Кудерметов Р.К., Тягунова М.Ю. Диспетчер завдань розподіленої комп'ютерної системи. Розроблено структуру диспетчера завдань розподіленої комп'ютерної системи. Описано особливості програмної реалізації диспетчера завдань розподіленої комп'ютерної системи. При цьому файли оцінки завдань запропоновано формувати у форматі XML, завданнями є програми, розроблені на мові Java.
Ключові слова: розподілена комп'ютерна система, GRID, диспетчеризація завдань, ресурсний запит, XML.

© Тягунов Д.В., Кудерметов Р.К., Тягунова М.Ю., 2011