УДК 681.3.06

ОЦЕНКА НАДЕЖНОСТИ ФУНКЦИОНИРОВАНИЯ ИНТЕГРИРОВАННОЙ КЛАСТЕРНОЙ СИСТЕМЫ С МЕТАПЛАНИРОВЩИКОМ GRIDWAY

А.С. Корсуков, к.т.н., научный сотрудник
(Институт динамики систем и теории управления СО РАН,
ул. Лермонтова, 134, г. Иркутск, 664033, Россия, alexask@icc.ru)

Надежность и эффективность работы распределенной вычислительной среды во многом зависят от планирования поступающих в систему потоков заданий. На сегодняшний день наиболее используемым метапланировщиком является Gridway. В статье рассматривается имитационное моделирование данного метапланировщика в интегрированной кластерной системе, описываются имитационные алгоритмы работы метапланировщика Gridway. Имитационное моделирование позволяет оценить степень эффективности и надежности интегрированной кластерной системы при использовании этих алгоритмов. В качестве среды имитационного моделирования использована система GPSS World. Построение модели алгоритмов функционирования метапланировщика Gridway выполнено путем описания вычислительной среды на формальном специализированном языке с последующей генерацией текста моделирующей программы на языке GPSS. В качестве узлов моделируемой интегрированной кластерной системы рассматривались вычислительные кластеры, отличающиеся по производительности и количеству ядер. Интегрированные кластерные системы имеют довольно сложную программно-аппаратную структуру. В связи с этим достаточно трудно подобрать оптимальные соотношения значений конфигурационных параметров для используемых средств планирования вычислений. Рассмотренные в статье средства имитационного моделирования позволяют частично решить данную проблему и получить в отдельных случаях повышение степени надежности и эффективности работы интегрированной кластерной системы.

Ключевые слова: *кластерная Grid, потоки заданий, планирование, инструментальные средства, моделирование.*

THE EVALUATION OF RELIABILITY OF THE INTEGRATED CLUSTER SYSTEM WITH THE META-SCHEDULER GRIDWAY

***Korsukov A.S.**, Ph.D., Research Associate*
(Institute of System Dynamics and Control Theory SB of RAS,
134, Lermontova st., Irkutsk, 664033, Russia alexask@icc.ru)

Abstract. The reliability and efficiency of a distributed computing environment mostly depend on the workflow scheduling. Nowadays, the most widely used meta-scheduler is Gridway. In the paper the simulation of this meta-scheduler in the integrated cluster system is considered. The simulation algorithms of the meta-scheduler Gridway are described. The simulation modeling allows to evaluate the efficiency and reliability of the integrated cluster system. The system GPSS World is used as a simulation environment. The model of the meta-scheduler Gridway was described on the special formal language. Next, the text of modeling program on GPSS-language was generated automatically. The computing clusters were considered as nodes of the integrated cluster system which was simulated. The clusters are characterized by different performance and cores count. The integrated cluster systems have complex software and hardware structure. As a result, it is so difficult to find the optimal values of configuration parameters for the used computing schedulers. The simulation modeling tools considered in this paper allow to solve this problem partially and obtain in some cases increase in reliability and efficiency of the integrated cluster system.

Keywords: cluster Grid, job flows, scheduling, toolkit, simulation modeling.

В последнее время большое значение придается организации интегрированных кластерных систем. Зачастую надежность и эффективность функционирования подобных систем напрямую зависят от работы планировщика, распределяющего потоки заданий по кластерам. Как правило, планировщик такого рода имеет целый ряд конфигурационных параметров для настройки процесса функционирования. Настроенный должным образом планировщик позволяет повысить надежность и эффективность использования ресурсов вычислительной среды. Таким образом, актуализируются построение и исследование модели данного планировщика с целью оценки надежности функционирования интегрированной кластерной системы, а также разработка методов и инструментальных средств повышения показателей этой надежности путем параметрической настройки алгоритмов управления распределенными и параллельными вычислениями в кластерной системе на основе полученных оценок.

Рассмотрим имитационное моделирование ряда алгоритмов работы метапланировщика Gridway в интегрированной кластерной системе.

Постановка задачи. Пусть в интегрированной кластерной системе имеется множество вычислительных кластеров $C=\{c_1, c_2, \ldots, c_k\}$. В составе i-го кластера c_i имеется r_i узлов. Через входные шлюзы в систему поступает множество потоков заданий $J=\{J_1, J_2, \ldots\}$. Отдельный поток может включать задания различных типов: стандартные последовательные и параллельные задания, многовариантные, взаимосвязанные и локальные задания [1]. Принадлежность j-го задания к тому или иному

типу определяется на основе вектора параметров $P_j=\{p_1, p_2, \ldots\}$, задающего требования, необходимые для выполнения данного задания. Кластерная система является системой массового обслуживания с очередями (Q – множество очередей) и имеет следующие характеристики: пуассоновское распределение моментов времени поступления заданий в систему и времени их выполнения; параллельная работа вычислительных узлов системы; дисциплина очереди, работающая по принципу «первым пришел – первым обслуживается» в разных вариациях; конечная емкость (количество вычислительных узлов) системы; бесконечная емкость источника, генерирующего задания.

В каждом шлюзе функционирует метапланировщик Gridway, распределяющий задания по кластерам в соответствии с его конфигурационными настройками. Работа метапланировщика заключается в подборе кластера c_i для выполнения очередного j-го задания, поступившего в систему, с учетом вектора параметров задания p_i. Если задание распределяется на кластер c_i, свободных ресурсов которого недостаточно для немедленного запуска данного задания, то оно ставится в очередь $q_i \in Q$, соответствующую кластеру c_i.

Метапланировщик имеет описание всех кластеров: число узлов, производительность отдельного узла, количество оперативной и дисковой памяти, система управления прохождением заданий. С помощью подсистемы мониторинга метапланировщик получает дополнительную информацию о числе свободных и занятых узлов, а также расписание выполнения заданий. Для каждого кластера задан коэффициент производительности. Различие коэффициентов производительности при выполнении одного и того же задания обусловливается различной конфигурацией кластеров. Построение модели функционирования метапланировщика Gridway осуществляется путем ее описания на формальном специализированном языке [2] с последующей генерацией текста моделирующей программы на языке GPSS.

Требуется исследовать алгоритмы работы метапланировщика Gridway в интегрируемой кластерной системе, использующие различные конфигурационные параметры метапланировщика. Сравнение эффективности работы исследуемых алгоритмов будем проводить на основе следующих показателей: число заданий с нулевым временем ожидания $t_0=\varphi(J, Q)$; среднее время ожидания задания в очереди $t_q=\chi(J, Q)$; среднее число заданий в очереди $n_q=\beta(J, Q)$; коэффициент полезного использования ресурсов интегрированной кластерной системы $k=\eta(C)$; коэффициент решаемости заданий $s=\gamma(C)$. Последний показатель особенно важен с точки зрения оценки надежности функционирования системы. Перечисленные выше функции $\varphi(J, Q)$, $\chi(J, Q)$, $\beta(J, Q)$, $\eta(C)$ и $\gamma(C)$ позволяют определить соответствующие им показатели на основе характеристик (числовых атрибутов) множеств J, Q, C.

Потоки заданий. Задание пользователя представляет собой спецификацию процесса решения задачи, содержащую информацию о необходимых вычислительных ресурсах, исполняемых прикладных программах, входных/выходных данных и т.п. Множество заданий пользователя, поступающих в интегрированную кластерную систему с его машины, составляет поток заданий, который в дальнейшем может сливаться с потоками заданий других пользователей, образуя новые потоки.

Описание потоков заданий, формируемых в разработанной модели, базируется на стандартизированном формате workload [3], который включает две основные группы параметров. Первая группа содержит набор характеристик потока в целом, например общее количество заданий в потоке, максимально разрешенное время выполнения задания и др., вторая – непосредственно характеристики каждого задания потока, например уникальный номер задания, время запуска, требуемое число процессоров, необходимые объемы оперативной и дисковой памяти, запрашиваемое процессорное время и др.

Характеристики потока заданий, генерируемого в разработанной модели, соответствуют характеристикам потока заданий в формате workload и их очередности. Благодаря этому в модели, помимо встроенного генератора потока заданий, можно использовать внешние файлы-описания потоков заданий, записанные в том же формате.

Конфигурирование метапланировщика Gridway. Средства конфигурирования данного метапланировщика включают две основные группы настроек: конфигурационные параметры процесса планирования и конфигурационные параметры встроенных политик планирования. Конфигурационные параметры первой группы предназначены для задания количественных показателей вычислительной среды, а также информации о ресурсах, пользователях и характеристиках процесса планирования. Конфигурационные параметры второй группы позволяют выбрать политики, определяющие приоритет задания и ресурсов. Каждая политика специфицируется своим набором параметров: вес политики, приоритеты и квоты пользователей или ресурсов в рамках политики, а также различные временные параметры. К общим конфигурационным параметрам встроенных политик планирования заданий относятся такие параметры, как максимальное количество заданий, отправляемых за итерацию планирования, максимальное число одновременно запущенных заданий от одного пользователя и максимальное количество заданий, которые планировщик может отправить на какой-либо ресурс.

Приведенный выше список конфигурационных параметров позволяет сделать вывод о том, что

задача рационального подбора параметров для определенной интегрированной кластерной системы, в которую поступают различные соотношения типов пользовательских заданий, является довольно сложной и требует проведения предварительных экспериментов для получения оценки показателей эффективности функционирования системы при определенных условиях.

Результаты моделирования. Вычислительный эксперимент проводился в моделируемой интегрированной кластерной системе, состоящей из 50 кластеров и включающей 10 000 вычислительных узлов с различными показателями производительности и надежности. Для оценки эффективности работы алгоритмов планирования в разных условиях проведен ряд экспериментов для каждого алгоритма. Варьировались количественное соотношение типов заданий во входном потоке и значения конфигурационных параметров метапланировщика Gridway.

Исследованы три алгоритма работы метапланировщика Gridway.

*Базовый алгоритм A*0 является простейшим алгоритмом метапланировщика Gridway, суть которого заключается в том, что поступающие задания обрабатываются в порядке поступления в систему и распределяются на первые же найденные подходящие к требованиям задания ресурсы. В данном алгоритме применяется только политика фиксированных (статичных) приоритетов заданий и ресурсов.

*Расширенный алгоритм A*1 основывается на политике времени ожидания, суть которой в увеличении приоритета заданий линейно с течением времени, чтобы предотвратить застои в очередях. В отличие от предыдущего алгоритма очередь заданий работает по дисциплине с динамическими приоритетами и ее можно представить как многоуровневую очередь, на каждом уровне которой находятся задания с одинаковым приоритетом. Все новые задания попадают на нижний уровень очереди и имеют минимальный приоритет. Обработка заданий происходит с самого верхнего уровня очереди, где задания имеют наивысший приоритет. Когда на этом уровне не остается заданий, они берутся из следующего уровня очереди и т.д. В данном алгоритме используется политика ранжирования ресурсов, заключающаяся в подборе максимально подходящих вычислительных ресурсов для выполнения определенного задания. Таким образом, используя данную политику, метапланировщик сможет принять более эффективное решение о распределении заданий. На основании данных о состоянии системы, полученных от информационных сервисов интегрированной кластерной системы, Gridway к каждому заданию выстраивает очередь наиболее подходящих ресурсов.

*Расширенный алгоритм A*2 работает аналогично *A*1, но с учетом дополнительной политики отказа ресурсов для повышения надежности выполнения заданий в интегрированной кластерной среде. Ресурсы, на которых часто возникают сбои, блокируются.

Сравнение показателей работы данных алгоритмов представлено в таблице.

Алгоритм	Показатели				
	t_0, сек.	t_q, сек.	n_q, ед.	k	s
*A*0	4687	733	65	0,91	0,87
*A*1	6603	563	43	0,94	0,92
*A*2	10711	185	18	0,99	0,95

Описанные выше алгоритмы применялись для обработки ряда потоков заданий с варьируемым соотношением заданий различных типов. Результаты вычислительного эксперимента позволяют сделать вывод о том, что более оптимальные значения показателей, перечисленных выше, получены при применении алгоритма *A*2. В ходе вычислительного эксперимента путем изменения параметров политики отказов ресурсов удалось повысить значение *s* для алгоритма *A*2 с 0,95 до 0,99.

Интегрированные кластерные системы характеризуются следующими свойствами: организационно-функциональная разнородность, динамичность и неполнота описания интегрируемых в них ресурсов; разнообразие спектра задач, решаемых с помощью этих ресурсов; наличие различных категорий пользователей, преследующих свои цели и задачи эксплуатации вычислительной системы. Для средств планирования вычислений, функционирующих в таких системах, достаточно сложно подобрать оптимальные соотношения значений их конфигурационных параметров. Рассмотренные в статье средства моделирования позволяют частично решать подобные задачи путем имитационного моделирования и в отдельных случаях получать повышение степени эффективности работы интегрированных кластерных систем в целом.

Литература

1. Корсуков А.С. Инструментальные средства полунатурного моделирования распределенных вычислительных систем // Современные технологии, системный анализ, моделирование. 2011. № 3 (31). С. 105.

2. Опарин Г.А., Феоктистов А.Г., Вартанян Э.К. Язык описания объектной модели Grid-системы // Программные продукты и системы. 2012. № 1. С. 3–6.

3. Chapin S., Cirne W., Feitelson D., Jones J., Leutenegger S., Schwiegelshohn U., Smith W., Talby D., Job Scheduling Strategies for Parallel Processing, Springer–Verlag, 1999, Vol. 1659, pp. 66–89.

References

1. Korsukov A.S., *Sovremennye tekhnologii, sistemny analiz, modelirovanie*, 2011, no. 3 (31), p. 105.

2. Oparin G.A., Feoktistov A.G., Vartanyan E.K., *Programmnye produkty i sistemy*, 2012, no. 1, pp. 3–6.

3. Chapin S., Cirne W., Feitelson D., Jones J., Leutenegger S., Schwiegelshohn U., Smith W., Talby D., *Job Scheduling Strategies for Parallel Proc.*, Springer-Verlag, 1999, Vol. 1659, pp. 66–89.