**BC/NW 2017 № 1 (30):3.1**

**ПРОБЛЕМА БАЛАНСИРОВКИ НАГРУЗКИ И РЕСУРСОВ ГРИД-ТЕХНОЛОГИЙ**

Абросимов Л.И, Арутюнян Ш.Ш.

В современном обществе существует необходимость в повышении каче­ства и скорости обработки в первую очередь «больших данных». В связи с этим возрастает значение грид-систем, как средства решения этой проблемы. Особенностью такой системы является географически распределенные гетерогенные компьютеры, соединенные с помощью сети[1].



Грид-технологии позволяют создать географически распределенные вычислительные инфраструктуры, которые объединяют разнородные ресурсы и реализуют возможность коллективного доступа к этим ресурсам [2].  Инфраструктура грид состоит из аппаратных средств и служб (на основе людских и программных ресурсов), которые должны быть организованы, и постоянно поддерживаться для того, чтобы ресурсы могли совместно использоваться. Наконец, стандарты определяют формат и протоколы обмена сообщениями, как между службами, так и между службами и пользователями, а также правила работы грида.

Таким образом, в основе грид-систем лежит обеспечение стабильной работы набора служб на основе общепринятых открытых стандартов и управляющего программного обеспечения (промежуточного программного обеспечения (ППО)) для обеспечения надежного, унифицированного доступа к географически распределенным информационным и вычислительным ресурсам, включающим отдельные компьютеры, кластеры и суперкомпьютерные центры, хранилища информации и т.д. Создание таких систем стало возможным благодаря впечатляющим успехам, прежде всего, в четырех направлениях:

* повышению производительности микропроцессоров массового производства – современный персональный компьютер сравним по производительности с суперкомпьютерами десятилетней давности;
* появлению быстрых линий связи – в настоящее время осуществляется перевод основных магистралей на уровень нескольких Гигабит/сек;
* глобализации обмена информацией (Интернет/Веб);
* развитию методов метакомпьютинга - научной дисциплины по организации массовых и распределенных вычислительных процессов.

Создание грид-системы подразумевает распределение вычислительных ресурсов по территориально разделенным областям, на которых установлено специализированное программное обеспечение для того, чтобы распределять задания по узлам и принимать их там, возвращать результаты пользователю, контролировать права пользователей на доступ к тем или иным ресурсам, осуществлять мониторинг ресурсов, и так далее. Общедоступные ресурсы, распределённые на большом участке территории, могут включать вычислительные узлы и/или узлы хранения и передачи данных, собственно данные, прикладное программное обеспечение.



 **Вычислительные ресурсы** предоставляют пользователю грид-системы (точнее говоря, задаче пользователя) процессорные мощности. Вычислительными ресурсами могут быть как кластеры, так и отдельные рабочие станции. При всем разнообразии архитектур любая вычислительная система может рассматриваться как потенциальный вычислительный ресурс грид-системы. Необходимым условием для этого является наличие ППО, реализующего стандартный внешний интерфейс с ресурсом и позволяющего сделать ресурс доступным для грид-системы. Основной характеристикой вычислительного ресурса является производительность.

 Ресурсы хранения также используют ППО, реализующее унифицированный интерфейс управления и передачи данных. Как и в случае вычислительных ресурсов, физическая архитектура ресурса памяти не принципиальна для грид-системы, будь то жесткий диск на рабочей станции или система массового хранения данных на сотни терабайт. Основной характеристикой ресурсов хранения данных является их объем. В настоящее время характерный объем ресурсов хранения измеряется в Терабайтах (Тб).

**Информационные ресурсы** и каталоги являются особым видом ресурсов хранения данных. Они служат для хранения и предоставления метаданных и информации о других ресурсах грид-системы. Информационные ресурсы позволяют структурировано хранить огромный объем информации о текущем состоянии грид-системы и эффективно выполнять задачи поиска ресурсов.

**Сетевой ресурс** является связующим звеном между распределенными ресурсами грид-системы. Основной характеристикой сетевого ресурса является скорость передачи данных.

Ресурсы принадлежат различным организациям, имеющим свои правила управления ресурсами, их использования и определения их стоимости для различных пользователей в различное время. Доступность и загруженность ресурсов также может динамически изменяться во времени.

В грид владельцы и потребители ресурсов имеют различные цели, используют различные стратегии и экономические схемы регулирования спроса и предложения[3]. Таким образом, актуальной задачей является разработка систем управления ресурсами грид, нацеленных на оптимизацию отношений между владельцами ресурсов и пользователями в соответствии с выбранными ими стратегиями.

Из вышесказанного можно выделить основную проблему современных грид систем – проблема балансировки нагрузки, которая учитывала бы требования, как пользователей, так и владельцев ресурсов:

* требование пользователей к сокращению времени выполнения заданий;
* требование пользователей к сокращению денежных затрат на использование ресурсов;
* требование владельцев на минимизацию простоя вычислительных ресурсов и повышению их рентабельности.

В грид системе пользователь конкурирует с другими пользователями и владелец ресурса с другими владельцами ресурсов. Следовательно, важным становится динамическое управления ресурсами грид, определяя выделяемые ресурсы, учитывая их цену, производительность, загрузку и распределяя задачи на этих ресурсах так, чтобы удовлетворить потребности пользователей и владельцев.

Таким образом, возникает потребность в разработке и использовании таких измерителей, которые оперируют показателями, затрагивающими потребности и пользователей и владельцев, а так же позволяют перейти к количественным оценкам показателей эффективности системы. К ним можно отнести: интенсивность обслуживания заданий, коэффициент загрузки узла, интенсивность входного потока заданий, математическое ожидание времени обслуживания и т.д. Используя данные показатели как характеристику эффективности использования ресурсов грид системы[4], можно решить проблему балансировки нагрузки ресурсов, а также управлять данными параметрами, которые не зависят от архитектуры узлов и программного обеспечения системы.

Одной из основных характеристик эффективности функционирования распределенных системы является производительность, которая должна оцениваться общим количеством выполненных информационно-вычислительных работ всеми узлами, входящими в грид[5]. Обработка запроса в компьютере требует выделение вычислительного ресурса на время выполнения вычислительных операций. Для этого обрабатывающие модули, временно сосредотачивают ресурсы узла для обработки поступающих запросов. Соответственно производительность узлов грид-системы оценивается набором интервалов времени, которые затрачивает каждый элемент каждого обслуживающего узла при обработке запроса соответствующего типа.

Таким образом, основной задачей измерения необходимой производительности ресурсов грид-систем, затрачиваемых на решение типовой задачи, то есть задачи по обработке сетевых запросов, является экспериментальное определение вероятностно-временных характеристик ЭВМ, выполняющей функции узла грид-системы.

Основные затраты при обработке запросов на узлах грид-систем, в соответствии с моделью ЭМВОС, происходят не только за счет прикладной программы, но и на транспортном и сетевом уровне[6]. Протоколы сетевого и транспортного уровня реализованы в виде программных функций ядра ОС. Учитывая вышесказанное, c помощью программных и аппаратных средств ОС можно достичь поставленную цель, для этого следует решить следующие задачи:

* определение функций участвующих в обработке заявок;
* определение адресов вызовов выделенных функций;
* установка динамических зондов на адреса вызовов подпрограмм использую средство SystemTap;
* измерение интервалов времени выполнения подпрограмм в тиках;
* расчет вышеуказанных показателей эффективности на основе полученных интервалов;

Используя вышеописанные измерительные средства (ИС), были получены оценки эффективности узла ЭВМ в клиент-серверной модели, в зависимости от различных характеристик входного потока, такие как: математическое ожидание времени обработки заявок, загрузку системы, характер распределения выходного потока заявок.

Таким образом, запустив ИС на каждом узле грид-системы и реализовав сбор полученных показателей, можно реализовать распределенную систему, оперирующая количественными оценками, учитывающими потребности, как пользователей, так и владельцев системы.

Учитывая вышесказанное, выделим основные требования к грид-системам, решающим проблему балансировки нагрузки:

* Динамическое управление ресурсами
* Оперирование количественными показателями эффективности
* Централизованный сбор статистики

**Список литературы**

1. Статья грид-технологии. http://grid.jinr.ru/?page\_id=39
2. Chervenak, A., Foster, I., Kesselman, C., Salisbury, C. and Tuecke, S. The Data GRID: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets. J. Network and Computer Applications, 2011
3. Realizing the Information Future: The Internet and Beyond. National Academy Press, 1994. <http://www.nap.edu/readingroom/books/rtif/>.
4. Czajkowski, K., Fitzgerald, S., Foster, I. and Kesselman, C. GRID Information Services for Distributed Resource Sharing, 2001.
5. Абросимов Л.И. Основные положения теории производительности вычислительных сетей. Вестник МЭИ 2001, №4, с. 70-75.
6. Абросимов Л. И., Крамаренко М. Д., Гончаренко О. С. Лабораторная работа по исследованию вероятностно-временных характеристик прокси-сервера BC/NW №2 2015 (27):15.1