

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

GRID-ВЫЧИСЛЕНИЯ

РЕФЕРАТ

студента 1 курса 171 группы
направления информатика и вычислительная техника
факультета компьютерных наук и информационных технологий
Шарова Александра Вадимовича

Проверил

д.ф.-м.н.,

Абросимов М.Б.

Саратов 2018

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 История	5
2 Общие задачи грида	6
3 Типы грид-систем с точки зрения решаемых задач	8
4 Применение грид систем	10
5 Проект World Community Grid	11
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	13

ВВЕДЕНИЕ

Grid-вычисления — это быстро развивающаяся технология, объединяющая ресурсы тысяч и даже миллионов отдельных компьютеров в гигантскую «виртуальную» систему с огромной вычислительной мощностью. Под Grid обычно понимают географически распределенную инфраструктуру, охватывающую ресурсы разных типов (процессоры, долговременная и оперативная память, хранилища и базы данных, сети), доступ к которым можно получить из любой точки, независимо от того, где они реально расположены. Обычно Grid предполагает коллективный режим разделяемого доступа к ресурсам и к связанным с ними услугам в рамках глобально распределенных виртуальных организаций, состоящих из предприятий и отдельных специалистов, совместно использующих общие ресурсы. В каждой виртуальной организации имеется своя собственная политика поведения ее участников, которые должны соблюдать установленные правила. Виртуальная организация может образовываться динамически и существовать ограниченное время. С точки зрения сетевой организации Грид представляет собой согласованную, открытую и стандартизованную среду, которая обеспечивает гибкое, безопасное, скоординированное разделение вычислительных ресурсов и ресурсов хранения информации, которые являются частью этой среды, в рамках одной виртуальной организации. Потенциал технологий Grid уже сейчас оценивается очень высоко: эксперты полагают, что он имеет стратегический характер, и в близкой перспективе Grid может стать вычислительным инструментарием для развития технологий в самых разных сферах человеческой деятельности. Это можно объяснить способностью Grid на основе безопасного и надежного удаленного доступа к ресурсам глобально распределенной инфраструктуры решить две задачи:

1. Создание распределенных вычислительных систем сверхвысокой пропускной способности из серийно выпускаемого оборудования при одновременном повышении эффективности имеющегося парка вычислительной техники за счет предоставления в Grid временно простаивающих ресурсов
2. Создание широкомасштабных систем мониторинга, управления, комплексного анализа и обслуживания с глобально распределенными

источниками данных, способных поддерживать жизнедеятельность государственных структур, организаций и корпораций.

Распределённые, или грид-вычисления, в целом являются разновидностью параллельных вычислений, которое основывается на обычных компьютерах (со стандартными процессорами, устройствами хранения данных, блоками питания и т. д.), подключенных к сети (локальной или глобальной) при помощи обычных протоколов, например Ethernet. В то время как обычный суперкомпьютер содержит множество процессоров, подключенных к локальной высокоскоростной шине. Основным преимуществом распределённых вычислений является то, что отдельная ячейка вычислительной системы может быть приобретена как обычный неспециализированный компьютер. Таким образом можно получить практически те же вычислительные мощности, что и на обычных суперкомпьютерах, но с гораздо меньшей стоимостью.

1 История

Термин «грид-вычисления» появился в начале 1990-х годов, как метафора, демонстрирующая возможность простого доступа к вычислительным ресурсам как и к электрической сети в сборнике под редакцией Яна Фостера и Карла Кессельмана “The Grid: Blueprint for a new computing infrastructure”[1]. Использование свободного времени процессоров и добровольного компьютеринга стало популярным в конце 1990-х годов после запуска проектов добровольных вычислений GIMPS[2] в 1996 году, distributed.net[3], в 1997 году и SETI@home[4] в 1999 году. Эти первые проекты добровольного компьютеринга использовали мощности подсоединённых к сети компьютеров обычных пользователей для решения исследовательских задач, требующих больших вычислительных мощностей. Идеи грид-системы были собраны и объединены Иэном Фостером, Карлом Кессельманом и Стивом Тики, которых часто называют отцами грид-технологии. Они начали создание набора инструментов для грид-компьютинга Globus Toolkit[5], который включает не только инструменты менеджмента вычислений, но и инструменты управления ресурсами хранения данных, обеспечения безопасности доступа к данным и к самому гриду, мониторинга использования и передвижения данных, а также инструментарий для разработки дополнительных грид-сервисов. В настоящее время этот набор инструментария является де факто стандартом для построения инфраструктуры на базе технологии грид, хотя на рынке существует множество других инструментов для грид-систем как в масштабе предприятия, так и в глобальном.

2 Общие задачи грида

Концепция грида появилась не как абстрактная идея, а как ответ на появляющиеся потребности в крупных информационно-вычислительных ресурсах, динамически выделяемых для решения громоздких задач, в научной, индустриальной, административной и коммерческой областях деятельности. Создание грид-среды подразумевает распределение вычислительных ресурсов по территориально разделенным сайтам, на которых установлено специализированное программное обеспечение для того, чтобы распределять задания по сайтам 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 на основе сайта 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[6]. 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.

Основными общими задачами грида являются:

1. создание из серийно выпускаемого оборудования широкомасштабных распределенных вычислительных систем и систем обработки, комплексного анализа и мониторинга данных, источники которых также могут быть (глобально) распределены
2. повышение эффективности вычислительной техники путем предоставления в грид временно простаивающих ресурсов. Приоритет той или иной общей задачи, которая решается с помощью грида, определяется типом грида и характером прикладных областей, в которых он используется

3 Типы грид-систем с точки зрения решаемых задач

Анализируя существующие проекты по построению грид-систем можно сделать вывод о трех направлениях развития грид-технологии:

1. вычислительный грид (Computational Grid)
2. грид для интенсивной обработки данных(Data Grid)
3. семантический Грид для оперирования данными из различных баз данных (Semantic Grid)

Целью первого направления является достижение максимальной скорости вычислений за счет глобального распределения этих вычислений между тысячами компьютеров, а также, возможно, серверами и суперкомпьютерами. Целью второго направления является обработка огромных объемов данных относительно несложными программами. Поэтому вычислительные ресурсы грид-инфраструктуры в этом случае зачастую представляют собой кластеры персональных компьютеров. А вот доставка данных для обработки и пересылка результатов в этом случае представляют собой достаточно сложную задачу. Одним из крупнейших проектов, целью которого является создание грид-системы для обработки научных данных, является проект EGEE (Enabling Grids for E-sciencE)[7].

Грид-системы третьего направления - семантические - предоставляют инфраструктуру для выполнения вычислительных задач на основе распределенного мета-информационного окружения, позволяющего оперировать данными из разнотипных баз, различных форматов, представляя результат в формате, определяемом приложением.

Не все проблемы лучше всего решать, используя распределенные кластеры на основе грид-технологий. Суперкомпьютеры незаменимы для некоторых научных проблем, типа составления прогноза погоды, когда множество процессоров должны часто общаться друг с другом. Очевидно, что такое частое общение невозможно обеспечить для географически распределенных и, возможно, аппаратно-неоднородных ресурсов в грид-среде. Другими словами, грид не слишком подходит для параллельных вычислений с интенсивным межпроцессорным обменом. Поясним чуть более подробно почему не следует смешивать

грид-технологии с технологией параллельных вычислений. Основными препятствиями для осуществления нетривиальных параллельных вычислений в грид-среде является нестабильность, плохая предсказуемость времени отклика на запрос. Причем это связано не только с тем, что в компьютерных сетях информационные пакеты проходят через множество сетевых устройств, но и с различиями в протоколах связи используемых во "внешних" компьютерных сетях и для межпроцессорного обмена внутри суперкомпьютеров. Это не позволяет эффективно организовать параллельные вычисления с интенсивным обменом информацией между процессорами, выполняющими отдельные подзадачи, в грид-среде. Грид-технология не является технологией параллельных вычислений, она предназначена для удаленного запуска отдельных задач на территориально распределенные ресурсы. Поэтому если громоздкая задача, которую необходимо решить, может быть разбита на большое количество маленьких, независимых (не обменивающихся никакими данными) частей, - грид-технология оказывается особенно эффективным и относительно дешевым решением. Напротив, суперкомпьютеры оказываются для таких вычислений неоправданно дорогим и неэффективным решением. В англоязычной литературе такие прикладные задачи иногда называют «bag-of-tasks» - сумка/мешок задач: вычисления для каждой выполняются независимо, а в конце пользователь или программное обеспечение просто должны соединить результаты индивидуальных вычислений. Типичными примерами таких задач являются: массовая обработка потоков экспериментальных данных большого объема (зачастую изучаемое явление можно разделить на отдельные независимые события и экспериментальные результаты по каждому событию обрабатывать независимо от других); визуализация больших наборов данных (отдельные области визуального представления обрабатываются независимо, а потом «склеиваются»); сложные бизнес-приложения с большими объемами вычислений (разбиение на части зависит от конкретного характера задачи).

4 Применение грид систем

Исходно Grid-технологии предназначались для решения сложных научных, производственных и инженерных задач, которые невозможно решить в разумные сроки на отдельных вычислительных установках. Однако теперь область применения Grid не ограничивается только этим. По мере развития Grid-технологии проникают в промышленность и бизнес, крупные предприятия создают Grid-системы для решения собственных производственных задач. Таким образом, Grid претендует на роль универсальной инфраструктуры для обработки данных, в которой функционирует множество служб (Grid Services), не только решающих конкретные прикладные задачи, но и предлагающих сервисные услуги: поиск необходимых ресурсов, сбор информации о состоянии ресурсов, хранение и доставка данных.

Применение Grid может дать новое качество решения таких классов задач, как массовая обработка потоков данных большого объема; многопараметрический анализ данных; моделирование на удаленных суперкомпьютерах; реалистичная визуализация больших наборов данных; сложные бизнес-приложения с большими объемами вычислений и т. д. Сегодня Grid-технологии уже активно применяются как государственными организациями в сфере управления, обороны, коммунальных услуг, так и частными компаниями, например, финансовыми и энергетическими. Область применения Grid охватывает ядерную физику, защиту окружающей среды, предсказание погоды и моделирование климатических изменений, численное моделирование в машиностроении и авиастроении, биологическое моделирование, фармацевтику.

5 Проект World Community Grid

Научный центр Рюбена Флита (Reuben H. Fleet Science Center), один из самых популярных в Сан-Диего (шт. Калифорния) культурных центров, который посещает 550 тыс. человек в год, стал первым учреждением культуры, участвующим в крупнейшем технологическом гуманитарном проекте IBM World Community Grid[8]. Теперь административные компьютеры Научного центра Флита включены в глобальную распределенную сеть World Community Grid и предоставляют свободные вычислительные ресурсы для выполнения масштабных исследований, направленных на решение ряда сложнейших проблем. Кроме того, Центр рассказывает сотням тысяч своих посетителей о World Community Grid, привлекая их к участию в этом проекте. Проект World Community Grid использует незадействованные вычислительные ресурсы компьютеров по всему миру, направляя их на решение задач, представляющих глобальную общественную ценность. В настоящее время в мире работает более 650 млн ПК, и проект World Community Grid ставит перед собой цель создать крупнейшую в мире Grid-инфраструктуру, выполняющую функции виртуального суперкомпьютера для решения исключительно гуманитарных задач. За девять месяцев вычислительная сеть World Community Grid предоставила для выполнения научных исследований около 15 тыс. лет компьютерного времени, задействовав ресурсы более чем 130 тыс. домашних и офисных компьютеров. По мнению исполнительного директора Научного центра Флита, IBM настолько упростила процедуру подключения к Grid, что к новому проекту планируется привлечь 100 тыс. учащихся и преподавателей, которые ежегодно принимают участие в программах Центра. Стоит отметить, что ранее IBM подарила Центру систему TryScience Kiosk[9], предоставляющую учащимся мгновенный доступ к экспонатам из свыше 600 лучших научных музеев мира и к интерактивным экспериментам. Первый же проект World Community Grid был связан с изучением пространственной структуры белков человеческого организма (Human Proteome Folding Project), которое ведется Институтом системной биологии (Institute for Systems Biology, ISB). Считается, что выполнение этого проекта с использованием вычислительных ресурсов института ISB, без привлечения ресурсов World Community Grid, заняло бы примерно 100 лет. Данный проект очень

важен для исследовательского сообщества, поскольку база данных белковых структур поможет ученым сделать следующий шаг к пониманию развития заболеваний, в которых участвуют эти белки, и в конечном счете найти методы лечения рака, малярии и других заболеваний. Свободный доступ к результатам проекта Human Proteome Folding Project дает исследователям возможность использовать эту информацию в своей работе. В настоящее время ресурсы World Community Grid позволяют ежегодно вести пять-шесть проектов для общественных и некоммерческих организаций. Результаты исследований будут открыты для мирового исследовательского сообщества. World Community Grid планирует проекты в таких областях, как:

1. медицина — исследования генома, протеомика, эпидемиология и изучение биологических систем
2. экология — климатология, загрязнение и защита природных ресурсов
3. базовые исследования — работы в области здравоохранения и социального обеспечения

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Foster, I. The grid: blueprint for a new computing infrastructure / I. Foster, C. Kesselman. LA : Univ. of Southern California, 1999. 748 p.
- 2 Great Internet Mersenne Prime Search [Electronic resource]. URL: <https://www.mersenne.org> (online; accessed: 28.11.2018).
- 3 DistributedNet [Electronic resource]. URL: <http://www.distributed.net> (online; accessed: 28.11.2018).
- 4 SETI@home [Electronic resource]. URL: <https://setiathome.berkeley.edu> (online; accessed: 28.11.2018).
- 5 Globus Toolkit [Electronic resource]. URL: <http://toolkit.globus.org/toolkit> (online; accessed: 28.11.2018).
- 6 Berman, F. Grid computing. Making the Global Infrastructure a Reality / [F. Berman, G. Fox, T. Hey] : Wiley, 2003.
- 7 Enabling Grids for E-sciencE [Electronic resource]. URL: <http://eu-egee-org.web.cern.ch/eu-egee-org/index.html> (online; accessed: 28.11.2018).
- 8 World Community Grid [Electronic resource]. URL: <https://www.worldcommunitygrid.org> (online; accessed: 1.12.2018).
- 9 TryScience Kiosk [Electronic resource]. URL: <http://www.empireexhibits.com/tryscience-interactive-kiosk> (online; accessed: 29.11.2018).