



**GRID-технологии
и
мета-компьютинг**

Введение.

Термин GRID (переводится как решетка или вычислительная сеть) только недавно начал входить в лексикон специалистов по информационным технологиям. Однако аналитики уже сейчас прогнозируют, что идея GRID может радикально изменить мир информационных технологий, точно так же, как когда-то это сделал интернет. Если интернет-ресурсы можно охарактеризовать как глобальный доступ к текстовой и графической информации, то GRID - это возможность вычислений на глобальных компьютерных ресурсах.



Введение в GRID-технологии.

Итак, GRID – это географически распределенная инфраструктура, объединяющая множество ресурсов разных типов (процессоры, долговременная и оперативная память, хранилища и базы данных, сети). При этом место их расположения роли не играет. GRID предполагает коллективный разделяемый режим доступа к ресурсам и связанным с ними услугам в рамках глобально распределенных виртуальных организаций, состоящих из предприятий и отдельных специалистов, совместно использующих общие ресурсы.

Основной идеей GRID-технологии является решение проблем создания виртуальных организаций, централизованное и скоординированное распределение ресурсов. Под распределением ресурсов понимается не только обмен файлами, но и прямой доступ к вычислительным мощностям, программному обеспечению, данным, периферийному оборудованию.

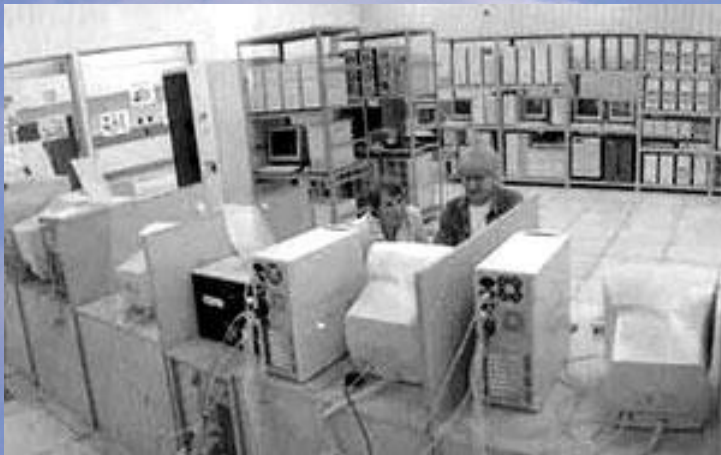
История развития

Концепция GRID появилась еще в конце 60-х годов. Ее рождение и развитие многие годы было связано с научными исследованиями.

Идея: использование для вычислений свободных ресурсов компьютеров. Эта проблема было особенно актуальна 25-35 лет назад, когда для построения серьезных математических моделей были нужны мощности суперкомпьютерных центров, которых тогда в мире было совсем не много.

До сколь-нибудь широкого применения в сфере бизнеса в те времена технология GRID не дошла. Так как:

- ✓ организовать динамическое перераспределение ресурсов для огромного числа разнородных задач — это очень непросто;
- ✓ актуальность самой проблемы отошла на второй план, так как именно в 70-е годы начался небывало быстрый рост мощностей автономных компьютеров.



Использование свободного времени процессоров стало популярным в конце 1990-х. Сам же термин "GRID" (*GRID - решётка*) возник в середине 90-х годов, он был впервые упомянут в сборнике под редакцией Яна Фостера и Карла Кессельмана. В 1997 г. и в 1999 г. появились первые GRID-проекты.

Технологию GRID можно назвать третьим поколением Интернета. Его второе поколение мы наблюдаем сегодня. GRID позволяет выйти за рамки простого обмена данными между компьютерами и в конце концов превратить их глобальную сеть в своего рода гигантский виртуальный компьютер, доступный в режиме удаленного доступа из любой точки независимо от места расположения пользователя.

Классы решаемых задач.

Изначально 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 включает в себя промышленность и бизнес, крупные предприятия создают GRID для решения собственных производственных задач. Таким образом, GRID претендует на роль универсальной инфраструктуры для обработки данных, в которой функционирует множество служб (GRID Services). Эти службы позволяют решать не только конкретные прикладные задачи, но и предлагают сервисные услуги: поиск необходимых ресурсов, сбор информации о состоянии ресурсов, хранение и обмен данными.

GRID можно применять для решения следующих классов задач:

- 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.

GRID с точки зрения разработчика.

- Software - программное обеспечение для решения самых различных задач на конкретном компьютере.
- Hardware - собственно компьютеры.
- Middleware - новый тип программного обеспечения, необходимый для управления работой географически распределенными GRID-системами.

Сегодня, говоря о GRID, обычно имеют в виду промежуточное ПО (middleware), программный инструментарий и прикладные программы, способные работать в географически распределенной и неоднородной вычислительной среде.

С точки зрения разработчика поддержка GRID реализуется в программных средствах нескольких уровней: на уровне системного ПО, middleware и конечных приложений.

Мета-компьютинг.

Понятие мета-компьютера можно определить как метафору виртуального компьютера, динамически организующегося из географически распределенных ресурсов, соединенных высокоскоростными сетями передачи данных. Отдельные установки являются составными частями мета-компьютера и в то же время служат точками подключения пользователей.

Сам термин возник в начале 90-х годов, когда начала развиваться высокоскоростная сетевая инфраструктура, и изначально относился к объединению нескольких разнородных вычислительных ресурсов в локальной сети предприятия, организованного для решения одной задачи. Основная цель построения мета-компьютера в то время заключалась в оптимальном распределении частей работы по вычислительным системам различной архитектуры и различной мощности.

В дальнейшем исследования в области технологий мета-компьютинга были развиты в сторону однородного доступа к вычислительным ресурсам большого числа (вплоть до нескольких тысяч) компьютеров в локальной или глобальной сети. Компонентами мета-компьютера могут быть как простейшие ПК, так и мощные системы.



мета-компьютер может не иметь постоянной конфигурации - отдельные компоненты могут включаться в его конфигурацию или отключаться от нее; при этом технологии мета-компьютинга обеспечивают непрерывное функционирование системы в целом.

Наилучшим образом для решения на мета-компьютерах подходят задачи переборного и поискового типа, где вычислительные узлы практически не взаимодействуют друг с другом и основную часть работы производят в автономном режиме.

Сложность применения GRID-технологий.

В связи с тем, что при использовании GRID-технологий открывается широкий доступ к информации, возникает вопрос о безопасности. Фундаментом технологии является требование обеспечения надежности услуг.

Технологические требования, предъявляемые к GRID, определены следующим образом:

1. Чёткий контроль над использованием ресурсов.
2. Многоуровневый контроль прав доступа, локальные и глобальные политики доступа.
3. Поддержка распределения различных ресурсов - программ, данных, устройств, вычислительных мощностей.
4. Поддержка различных моделей пользования - многопользовательской, однопользовательской, различных режимов.
5. Контроль над качеством предоставляемых услуг, планирование, резервное предоставление услуг.

Плюсы и минусы

Плюсы:

1. Широкий доступ к информации.
2. Высокая надежность. Все ресурсы и все пользователи в GRID имеют свои сертификаты - зашифрованные ключи. Расшифровать их практически невозможно, тем более что сертификаты периодически изменяются.
3. Постоянный доступ к отдельным ресурсам сети.

Минусы:

4. Никогда не известно заранее, сколько времени потребуется для решения конкретной задачи, насколько загружены мощности системы. Существует мнение, что есть целый класс задач, которые в принципе не могут решаться с помощью GRID.
5. Современные технологии распределённых вычислений не отвечают всему спектру предъявляемых требований.

Программные средства GRID

К настоящему времени для GRID уже создано довольно много программных средств и их количество постоянно растет.

Базовые средства - инструментальные пакеты, содержащие средства разработки систем для GRID и минимальный набор служб, поддерживающих дистанционные операции (Globus Toolkit, Legion).

Отдельные программные средства для GRID - различные разработки для GRID, выполняющие отдельные функции (Condor-G, The Network Weather Service).

Пакетированные комплексы средств - прошедшие проверку практикой собрания программных средств, как правило, от разных разработчиков. Пакеты достаточно полны функционально, хорошо документированы и на их основе можно построить реально действующий GRID (GRIDConfig, GRIDsolve).

Платформы - взаимосогласованные наборы программных средств. Различают открытые платформы (EU DataGRID, ARC – NorduGRID), коммерческие платформы, которые в свою очередь делятся на вычислительные платформы (DCGRID, LiveCluster, Frontier) и платформы для работы с данными (Avaki Data GRID).

Проекты по мета-компьютингу и распределенным вычислениям в Интернет

- *GIMPS - Great Internet Mersenne Prime Search* - поиск простых чисел Мерсенна (т.е. простых чисел вида 2^p-1).
- *SETI@home* - поиск внеземных цивилизаций с помощью распределенной обработки данных, поступающих с радиотелескопа.
- *Condor* - распределяет независимые подзадачи по существующей в организации сети рабочих станций, заставляя компьютеры работать в свободное время.
- *DataGRID* – цель: создание сетевой компьютерной инфраструктуры нового поколения для обработки распределенных тера- и петабайтных баз данных, полученных в результате научных исследований, преимущественно в области физики.
- *EGSO - European GRID of Solar Observations* – цель: создание "виртуальной солнечной обсерватории", данные в которой будут собираться из наземных и космических наблюдательных пунктов.
- *Climateprediction.net* – цель: выяснение точности современных климатических моделей и установка поправок, которые должны быть в них внесены.

и др.

Проекты GRID в России

Проект Дубна-GRID

Программа Дубна-GRID, который проводит Объединенный институт ядерных исследований (ОИЯИ) в Дубне, была создана с целью более эффективного использования мощностей компьютерного парка университета и компьютерных классов школ, объединенных городской информационно-образовательной сетью. На базе уже существующей городской инфраструктуры можно создать систему для исследования в области GRID-технологий, чтобы передавать этот опыт в другие регионы, обучать GRID-специалистов.

Первый масштабный проект, в котором участвовали многие российские центры - EU DataGRID - стартовал в 2001 году и объединял организации 14 европейских стран. Цель: создание глобальной инфраструктуры нового поколения для обработки огромных массивов информации в области физики высоких энергий, биоинформатики и системы наблюдения за Землей. Именно участие в этом проекте обеспечило российским ученым знакомство и опыт работы с новейшим программным обеспечением типа GRID.

Проект EGEE

1 апреля 2004 стартовал европейский проект создания глобальной компьютерной инфраструктуры GRID - EGEE («Развертывание GRID для электронной науки»). Сейчас в этом проекте работают специалисты из 32 стран. Его ресурсами пользуются более 200 виртуальных организаций для исследований в астрономии, астрофизике, физике высоких энергий, физике конденсированных сред, вычислительной физике, задачах термоядерного синтеза и др. В фармакологии сегодня работают 20 приложений проекта, очень много приложений интенсивно используются в вычислительной химии.

Этот проект участвовал в международном проекте термоядерного синтеза ITER, где именно использование EGEE позволяло обрабатывать и хранить большие массивы данных, принимать решения в режиме реального времени, оптимизировать параметры и так далее.

Также ресурсы EGEE могут быть полезны в правительственных структурах, в сфере образования, также и в бизнесе: возможен GRID одного предприятия, причем, сегодня это уже реализованное стандартное решение, дальше возможно развитие до GRID корпорации, а дальше - мечта всего человечества - глобальный GRID.

Актуальность GRID-технологий.

Сейчас IT компании начали активно говорить о GRID. Это обусловлено следующими факторами:

1. Для того, чтобы связать множество компьютеров в единую GRID нужны хорошие сети передачи данных. Конечно, пропускная способность и скорость сетей еще недостаточны, но они очень быстро развиваются.
2. В мире появляется все более сложные задачи и накапливаются все большие объемы данных. И для решения этих суперзадач и обработки этих огромных массивов данных уже не годятся обычные компьютеры. Нужны суперкомпьютеры с очень высокой мощностью и таких компьютеров требуется все больше. Стоимость этих суперкомпьютеров очень высока, но их мощности очень быстро перестает хватать.
3. Концепция GRID сегодня позволяет получить результат быстрее и дешевле. Затраты на администрирование уменьшаются. В качестве элементов GRID можно использовать дешевые компьютеры и операционные системы.

4. Растет количество людей, знакомых с терминами GRID, виртуализация, вычисления как коммунальная услуга и т.д. Увеличилось число публикаций, концепция достаточно понятна и уже сами заказчики и отделы IT требуют внедрения GRID технологий.
5. Уже разработаны стандарты GRID. Многие крупные фирмы – производители компьютеров и программного обеспечения участвуют в Global GRID Forum – некоммерческой организации, разрабатывающей стандарты построения GRID. Причем разрабатываются не только стандарты, но и инструментарий для реализации этих стандартов.
6. Появился опыт реализации концепции GRID и реальных проектов, построенных на основе этой концепции.
7. Основные производители компьютеров и программного обеспечения начали промышленную реализацию и продажу продуктов, позволяющих строить GRID.

Заключение

- GRID - это качественно новый этап развития системы распределенных вычислений, позволяющий получать доступ к вычислительным мощностям, данным и вообще говоря ресурсам многих машин, расположенных по всему миру.
- Потенциал технологий GRID оценивается очень высоко: в близкой перспективе GRID должен стать вычислительным инструментарием для развития высоких технологий в различных сферах человеческой деятельности. И хотя изначально GRID задумывался как подспорье для научных исследований, сейчас всё больше говорят о возможности использования этих технологий в предпринимательской деятельности. Технологии GRID применяются во многих сферах и постепенно круг этих сфер увеличивается.
- Распространение GRID не влечёт за собой кризиса профессии программиста, так как для поддержания сети и разработки качественно новых к ней приложений будут нужны высококвалифицированные специалисты.
- Несмотря на то, что это сравнительно недавнее направление, уже запущены многие проекты, связанные с GRID.