

Часть 1.
**От Веб до Грид:
к виртуальным
организациям**

В современном мире существует уже множество самых разных компьютеров. Причем далеко не все они работают в полную силу круглые сутки. Одни считают с утра до вечера, другие трудятся вполсилы, а третьи, вообще, большую часть времени простаивают, часами ожидая включения или хотя бы случайного прикосновения к клавиатуре. В тех странах, где сейчас ночь, мало кто работает, не работают и компьютеры, а ведь они могли бы в это время решать какие-нибудь полезные задачи – например, предсказывать изменения климата, изучать движение звезд по небу, искать лекарство против СПИДа и тому подобное. В то же время, на другой стороне Земли, где сейчас день, порой катастрофически не хватает вычислительных мощностей – скажем, метеорологи не могут точно и в срок предсказать погоду, нефтяники не справляются с расчетом контура месторождения и так далее... По существу, на основе этого явного противоречия и возникла простая идея: обеспечить желающим постоянный доступ к свободным компьютерным ресурсам. Несколько упрощая ситуацию, можно сказать, что именно это обстоятельство и легло в основу концепции Грид (Grid).

Действительно, идея совместного использования вычислительных возможностей соединенных друг с другом компьютеров появилась, видимо, одновременно с самим понятием «компьютерная сеть». Сначала были развиты методы и способы для информационного обмена, то есть простейший обмен файлами, электронная почта и так далее. Затем появились Web-сервисы, которые для многих пользователей, по существу, являются синонимом понятия компьютерной сети. До настоящего времени этот тип взаимодействия компьютеров остается преобладающим.

Возможность распределения собственно вычислений (то есть не только разнообразных данных, но и исполняемого компьютерного кода-программы) появилась позднее. Поддержка такого рода взаимодействия компьютеров требует довольно больших усилий даже для компьютерных кластеров с одинаковой процессорной архитектурой, не говоря уже о так называемых гетерогенных системах, когда в сеть объединены процессоры совершенно разных архитектур и производителей.

Владимир Кореньков, Вадим Бедняков
**Перспективы Грид-технологий
в промышленности и бизнесе**

Простейшим вариантом распределенных вычислений является, например, обычная пакетная система, в рамках которой пользователь, «поставив» тем или иным способом свою задачу в «очередь» на счет, не должен беспокоиться о том, как, когда и на какой конкретно машине будет выполнено его задание. Как правило, такие системы предполагают не только идентичность аппаратной и программной среды, синхронность обновления пользовательских данных на всех машинах кластера, но и совместимость средств авторизации пользователя, то есть прав доступа пользователя к тем или иным компьютерно-вычислительным ресурсам. Все это может быть достигнуто достаточно простыми средствами на компьютерных кластерах масштаба одного «предприятия», при наличии единого и централизованного администрирования. Однако более крупные масштабы, например, объединение счетных ресурсов нескольких организаций, требуют существенного расширения механизмов авторизации пользователя в рамках такого «гиперкластера».

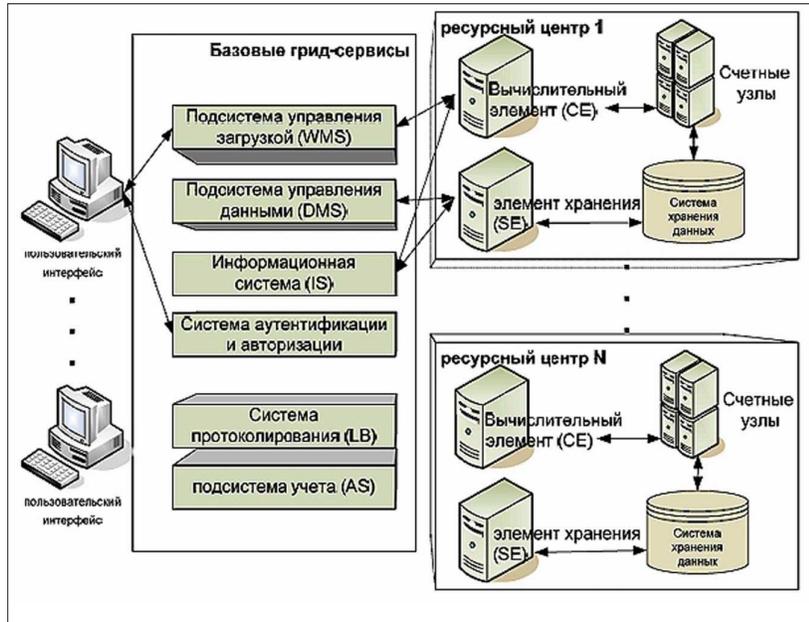
Систематическая разработка способа распределенных вычислений, получивших впоследствии название Грид, началась примерно в середине 90-х годов XX века. Сами же глобальные распределенные Грид-инфраструктуры появились в конце 90-х годов, когда скорость передачи данных вышла на гигабитный уровень, что позволило объединять суперкомпьютеры, кластеры на глобальном уровне (следствие закона Мура)¹. С тех пор и мир информационных технологий переживает бурное и широкомасштабное развитие. Грид – это мощная компьютерная инфраструктура совершенно нового типа, которая обеспечивает глобальную интеграцию информационных и вычислительных ресурсов. В настоящее время Грид предоставляет пользователям развитую технологию доступа к общим ресурсам и службам в рамках так называемых виртуальных организаций.

Виртуальная организация (Virtual Organization) – это совокупность организаций, объединенных для решения проблем в режиме скоординированного распределения своих ресурсов. Принадлежность к ней пользователь подтверждает наличием у него так называемого CA сертификата (Certificate Authority), признаваемого всеми участниками объединения на всех кластерах, объединенных в Грид-структуру. Сертификат дает пользователю право выполнять задания на конкретном кластере, входящем в виртуальную организацию пользователя. Современное понятие Грид весьма широко – это географически распределенная инфраструктура, объединяющая множество ресурсов разных типов (процессоры, долговременная и оперативная память, хранилища и базы данных, сети), доступ к которым пользователь может получить из любой точки мира, независимо от места ее расположения. Появление и быстрое внедрение Грид стало необходимым благодаря тому, что сильно возросло число сложных научно-прикладных задач, для решения которых требовались огромные вычислительные ресурсы, и такие ресурсы стали, в принципе, доступны. Помимо этого появилась возможность проводить так называемые параллельные вычисления, а скорость и надежность коммуникационных каналов связи оказались достаточными для эффективной передачи больших объемов информации.

Первоначально технологии Грид использовались для научных и инженерных приложений. Однако теперь они становятся основой для координированного совместного использования ресурсов в динамических, охватывающих многие предприятия виртуальных организациях в промышленности и бизнесе. Таким образом, сегодня Грид становится все более универсальной и вполне эффективной инфраструктурой для высокопроизводительных распределенных вычислений и обработки больших совокупностей данных.

Области наиболее эффективного применения современных Грид-технологий можно разделить на три основных – это моделирование сложных процессов и систем, совместная (корпоративная, как любят сегодня говорить) визуализация

¹ По закону Мура, каждое следующее поколение компьютеров работает в 2,5 раза быстрее, а каждая последующая версия ОС Microsoft Windows – в 1,5 раза медленнее. – Прим. ред.



ция очень больших наборов данных и распределенная обработка, хранение и анализ огромных массивов данных.

В практическом плане это, например, максимально эффективное использование временно простаивающих компьютеров для выполнения относительно небольших по объему задач. С другой стороны, Грид осуществляет распределенные, так сказать, супервычисления для решения очень крупных вычислительных (управленческих) задач, требующих как огромного времени для проведения собственно вычислений (процессорных ресурсов), так и столь же внушительных ресурсов памяти, дискового пространства и так далее. Следующая область применения Грид-технологий касается компьютерных вычислений с привлечением больших объемов географически распределенных данных и ресурсов, например, в метеорологии, астрономии, физике высоких энергий. Наконец, нельзя обойтись без Грид в том случае, когда в реальном времени проводятся коллективные вычисления, в которых одновременно принимают участие пользователи из различных организаций.

Часть 2.

Для хранения и обработки экспериментальных данных

Сообщество ученых, занимающихся физикой высоких энергий – исторически самый первый и, пожалуй, пока самый главный пользователь-потребитель Грид-технологий. Впервые Грид-подход был принят в ЦЕРН для хранения и обработки данных экспериментов на

Большом адронном коллайдере (LHC).

Действительно, Грид – это единственно эффективное средство работы с огромными объемами и потоками данных (тысячи Гигабайт ежедневно), которые распределяются и обрабатываются практически во всем мире участвующими в экспериментах международными научными коллективами.

На Большом адронном коллайдере уже запущены четыре основных эксперимента с уникальными детекторами. Это эксперимент по физике тяжелых ионов ALICE (A Large Ion Collider Experiment), два многоцелевых протон-протонных эксперимента ATLAS (A Toroidal LHC Apparatus) и CMS (Compact Muon Solenoid) и специализированного эксперимента LHCb (The Large Hadron Collider beauty experiment), задача которого – тщательное исследование физики b-кварков. Каждый из этих экспериментов проводится силами беспрецедентных по численности международных коллабораций. Так, например, коллаборация ATLAS насчитывает более 2900 человек из 172 различных физических институтов 37 стран Европы, Азии, Северной и Южной Америки, Африки и Австралии. Понятно, что только в рамках Грид-концепции возможна эффективная работа такого сообщества физиков в реальном времени.

В результате столкновений протонов образуется огромное число вторичных частиц, для регистрации и исследования которых служат упомянутые детекторы. По существу, это целые детектирующие комплексы, уникальные по сложности и размерам (так, детектор ATLAS по высоте превышает пятиэтажный

дом). Их задача – конвертировать результаты взаимодействия вторичных частиц с веществом детекторов в специальные сигналы, с которыми уже могут работать компьютеры. Эти сигналы изощренным образом обрабатываются хитроумными компьютерными программами, в результате чего физики получают возможность понять, сколько, каких именно и с какими характеристиками образуется вторичных частиц в результате каждого столкновения протонов сверхвысоких энергий на LHC. Именно эти специализированные сигналы (которые называются «сырыми» данными) производит на свет каждый из детектирующих комплексов. Объем «сырых» данных огромен, поэтому в режиме реального времени (он-лайн) отбираются лишь самые интересные конфигурации вторичных частиц (события) для их сохранения и последующего подробного и тщательного физического анализа. Далее записанные события специальным образом обрабатываются в офф-лайн режиме. В результате из первоначальных специфических для каждого детектора сигналов реконструируются физические объекты – адронные струи, фотоны и лептоны. Именно эти объекты (плюс еще дисбаланс энергии) являются теми уникальными источниками информации, которыми располагают физики в своей работе по поиску бозона Хиггса и других явлений на Большом адронном коллайдере.

На этой ключевой стадии обработки информации – реконструкции реальных событий – невозможно обойтись без Грид-вычислений. Только они позволяют параллельно использовать компьютерные сети университетов и лабораторий во всем мире для крайне ресурсоемкой (в смысле процессорного времени и дискового пространства) задачи реконструкции огромных потоков исходных данных. По существу, Грид был задуман именно для такой работы.

Комплекс программ, обеспечивающих работу на десятках тысяч машин в сотнях организаций, входящих в коллаборации экспериментов на LHC, получил название LCG (LHC Computing Grid). Ресурсы LCG доступны также целому ряду других Грид-комплексов, среди которых есть как международные (EGEE – Enabling Grids for E-Science, OSG – Open Science Grid и пр.), так и национальные и региональные, например, INFN Grid в Италии и RDIG (Russian Data Intensive Grid) в России.

*(Окончание
в следующем номере.)*