

Перспективы Грид-технологий в промышленности и бизнесе

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. По существу, на основе этого явного противоречия и возникла простая идея: обеспечить желающим постоянный доступ к свободным компьютерным ресурсам. Немного 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, явля-

ются 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.

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 (последствие закона Мура). С тех пор и мир информационных технологий переживает бурное и широкомасштабное развитие. Грид – это мощная географически распределенная компьютерная инфраструктура совершенно нового типа, которая обеспечивает глобальную интеграцию информационных и вычислительных 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. 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 СА 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.

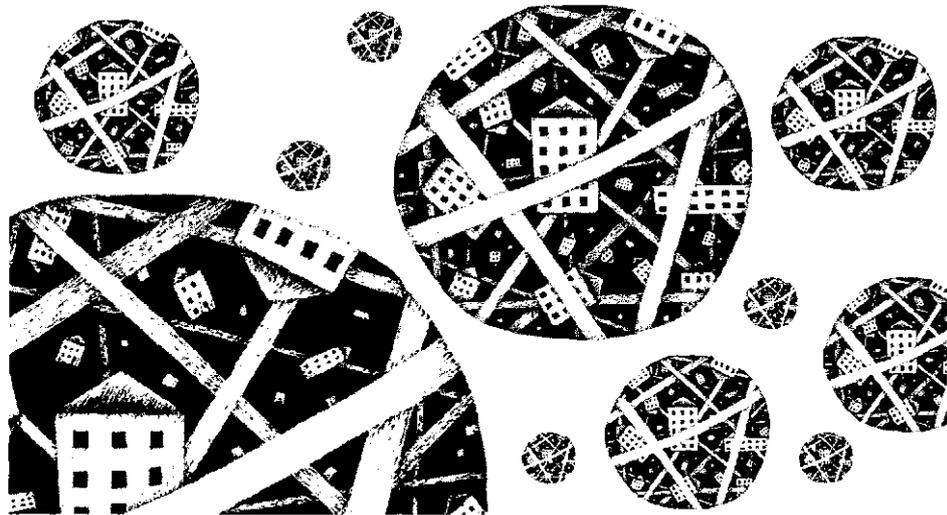
Первоначально технологии Грид использовались для научных и инженерных приложений. Однако теперь

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 основные – это моделирование сложных процессов и систем, совместная (корпоративная, как любят сегодня говорить) визуализация 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 – компьютерные вычисления 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.

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 (Европейская организация ядерных исследований в Женеве) 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 ALICE, ATLAS, CMS и LHCb. 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) производят 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 будут располагать 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 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, так и национальные, и региональные (например, в Италии и в России).

Биомедицинские науки представляют сегодня другое основное сообщество пользователей Грид-технологий. Эта сфера – одна из важнейших областей приложений, она достаточно четко делится на три области: обработка медицинских графических данных, биомедицинские дисциплины и разработка лекарств.

Первая из них – не что иное, как компьютерный анализ цифровых медицинских изображений. Он опирается на интегрированные мировые медицинские данные и алгоритмы, требующие значительных компьютерных ресурсов, на обработку больших объемов данных и статистические исследования на больших выборках населения.

Построение медицинских изображений включает работу многих связанных друг с другом систем. Некоторые из них требуют значительных вычислительных ресурсов и выполняют параллельную реконструкцию изображений. В результате появляются такие технологии, как «виртуальные биопсии» для диагностики рака без хирургического вмешательства, что по-

вышает качество жизни пациентов и помогает снизить стоимость лечения. Компьютерное моделирование дает возможность планировать сеансы радиотерапии на основе результатов графических данных обследования пациента, а Грид-инфраструктура используется для уменьшения времени, необходимого для получения более достоверных результатов. Клинические системы используют классификацию изображений для того, чтобы помочь врачам при принятии решений о выборе методов лечения. Отдельные приложения изучают распространение контрастного вещества в печени и других органах, используя последовательность изображений, полученных методом магнитного резонанса. Движения пациента во время этой процедуры делают невозможным непосредственное сравнение изображений. Однако специальная корректирующая обработка изображений (на основе нахождения подобных фрагментов) позволяет конструктивно анализировать такую последовательность изображений за разумное время.

Интерактивное формирование трехмерных изображений всего объема сложных органов можно осуществить с помощью пакета специализированных программ. Еще один пакет программного обеспечения применяется в неврологических исследованиях для ранней диагностики болезни Альцгеймера. В его основе лежит сравнение данных пациента с большим набором данных от людей, не имеющих этой патологии. Объем данных, которыми при этом приходится манипулировать, превышает возможности одиночных стандартных компьютеров, но приложение может легко быть распределено по ресурсам Грид и выполнено за приемлемое время.

Вторая область Грид-технологий в биомедицине – биоинформатика. Здесь также ведется работа со сложными алгоритмами оптимизации обработки биологических данных. Сфера интересов биоинформатики включает в себя геномику, протеомику и филогению.

Так, например, выполняется трехмерный структурный анализ больших макромолекулярных комплексов. В процессе восстановления их структуры используется множество разных изображений исследуемого образца. Эти изображения, однако, часто подвержены различным фоновым воздействиям, что заметно усложняет анализ. Поэтому для составления наиболее адекватной экспериментальным данным модели необходимо сделать много последовательных приближений, что требует значительных Грид-ресурсов. С помощью другого приложения возможно проводить моделирование эволюции генома человека. Оно позволяет восстановить исторический процесс расселения человека по Земле в географически правдоподобных ландшафтах и моделировать молекулярное разнообразие разных человеческих популяций.

И третье. Грид-инфраструктура используется для ускорения поиска молекул с целью создания препаратов от различных болезней, новых лекарств посредством компьютерного моделирования структуры и динамики белков. Используя основанные на Грид-технологиях программы расчета пристыковки молекул (молекулярного докинга), исследователи быстро просматривают многие тысячи вариантов и отбирают наиболее перспективные соединения, что снижает стоимость разработки препаратов.

Цель расчетов для поиска – определить, насколько эффективно конкретные лекарства присоединяются к определенным участкам вируса-мишени. Успешными оказались приложения для поиска средств от малярии и птичьего гриппа. Планируется поиск лекарств и от других вирусов.

В этом направлении в год выполняется около 15 тысяч различных исследований. Так, например, в ходе разработки одного лекарства за месяц был проведен анализ молекулярного докинга, который потребовал бы почти 100 лет работы обычного компьютера.

В области наук о Земле Грид-технологии сформировали ряд специа-

лизированных приложений в гидрологии, в сфере наблюдений за поверхностью Земли со спутников, в климатологии (например, прогнозирование наводнений), физике суши (твердой Земли), в области мониторинга атмосферы и океана и так далее. В рамках специализированной виртуальной организации ESR (Earth Science Research – «Исследования в области наук о Земле») развернуто несколько исследовательских приложений, где Грид-инфраструктура на базе спутниковой информации помогает вести анализ характеристик озонового слоя, а также поиск и обнаружение утечек нефти. В работе ESR важное место занимают приложения в области физики суши, которые сосредоточены на оперативном анализе механизмов землетрясений, а также на численном моделировании условий землетрясений в сложных трехмерных геологических моделях. Очевидно, что результаты этой работы помогут научному сообществу лучше понять природу этих разрушительных природных явлений. В области гидрологии с помощью Грид-технологий изучается характер проникновения морской воды в прибрежный водоносный слой в бассейне Средиземного моря, обусловленный хозяйственной деятельностью человека.

Один из способов предсказания извержения вулканов основан на преобразовании геофизической информации о поведении вулканов в звуковые волны. Картина звуковых волн анализируется, и это позволяет предугадывать поведение вулканов в ближайшем будущем, в том числе предсказывать извержения.

Грид-технологии также активно используются в таких областях, как термоядерный синтез, астрофизика (международные проекты MAGIC, Planck, ANTARES, NEMO), молекулярное моделирование (проекты вычислительной химии CHARON, CompChem), нанотехнологии, археология и в других направлениях. Везде глобальная Грид-инфраструктура упрощает сотрудничество между гео-

графически распределенными сообществами и позволяет им совместно пользоваться компьютерными ресурсами и данными. Именно поэтому число Грид-проектов быстро растет.

На основе Грид-технологий можно создать бизнес-среду для организации распределенных вычислений в таких коммерческих приложениях, как электронный бизнес, распределенное производство, коллективное проектирование сложных объектов, системы обработки с высокой пропускной способностью, распределенный супермаркетинг и тому подобное. Задача любой коммерческой структуры – получение прибыли при минимизации издержек. Поэтому любая компания, нуждающаяся в ресурсоемких вычислениях, должна быть заинтересована в применении технологии, позволяющей получить более высокопроизводительные мощности с меньшими затратами.

Учитывая радикальное различие между наукой и бизнесом, можно предположить, что Грид-среда для бизнеса будет отличаться от научно-вычислительной. У Грид-решений для бизнеса видится как минимум одно важное свойство – это экономически более эффективная модель получения результата. Все предыдущие «вычислительные идеи» в этой области сводились, главным образом, к построению инфраструктуры информационных технологий за счет «повторного» использования вычислительных ресурсов при изменении бизнес-процессов и упрощения системы администрирования.

Владелец или управляющий предприятием обычно далеко не компетентен в информационных технологиях и лишь в самых общих чертах может понять, почему нельзя черпать вычислительные ресурсы из розетки, как электроэнергию, или платить только за использованные ресурсы, как, например, за телефон или воду. Почему мощный сервер ему нужен в момент составления годового отчета, а платить за этот сервер он должен в течение всего года?

Как и всякая новая концепция, Грид для бизнеса часто воспринимается упрощенно. Многие технические специалисты полагают, что достаточно соединить несколько компьютеров в решетку, и они получат Грид-систему. Это неверно как для бизнес-Грида, так и для вычислительных Грид-систем. Фактически бизнес-Грид – это распределенная система обработки данных с единым центром управления, в которой вычисления выполняются на группе связанных между собой компьютеров. При этом нагрузка на серверы автоматически распределяется таким образом, чтобы соответствовать текущему потреблению вычислительных ресурсов.

В настоящий момент различные компании, включая такие гиганты индустрии информационных технологий, как Sun, Oracle, HP, IBM, предлагают свои Грид-решения, основанные на их понимании Грид-технологий. Главная проблема, однако, заключается в том, что, в отличие от научной сферы, в этой области пока еще не все стандартизовано и каждая компания реализует Грид-концепцию в бизнесе по своему усмотрению. Основная задача состоит в том, чтобы такие системы стали универсальными и могли адаптироваться к любым изменениям бизнес-процессов компании.

В Европе и Америке уже накоплен заметный опыт применения Грид-технологий для решения достаточно широкого класса задач, в том числе бизнес-приложений. Один из ярких примеров – это европейский проект BEinGRID (Business Experiments in Grid – «Эксперименты по применению Грид-технологий в бизнесе»), главная цель которого – содействие распространению Грид-технологий следующего поколения.

Для этого планируется создать специальный язык верхнего уровня для управления программным обеспечением. В рамках проекта будет проведен ряд целевых пробных бизнес-экспериментов для внедрения и распространения Грид-технологий в самых разных секторах европейского

бизнеса (финансы, экономика, текстильная и химическая промышленность, розничная торговля, развлечения, игры и др.). Чтобы свести к минимуму разработку новых Грид-компонентов, планируется разворачивать инновационные решения в области Грид-технологий, используя существующие Грид-компоненты Европейского союза и других регионов.

Однако в настоящее время в России потенциал такого рода практически нигде не используется, его не видно в промышленности, бизнесе, его нет и в государственном управлении.

Чтобы активизировать деятельность по бизнес-Гриду, в России следует создать Грид-портал для реализации заказов по адаптации компьютерных приложений к работе в среде Грид, в первую очередь – бизнес-приложений (методика, инструментальные средства, обучение, поддержка в рабочем состоянии и т.п.). Эту общую деятельность можно условно разделить на три главных направления. Во-первых, это исследование и тестирование различных Грид-систем, во-вторых, разработка и внедрение Грид-решений для клиентов на основе существующих продуктов, и, в-третьих, адаптация бизнес-приложений для Грид-среды.

По первому поводу напомним, что различные компании предлагают свои Грид-продукты. И каждый из них, безусловно, рекламируется как «самое замечательное решение для вашего бизнеса», что почти всегда неверно. Необходимо знать реальные возможности той или иной системы и, конечно же, потребности конкретного заказчика. Для этого следует проводить качественное тестирование и исследование возможностей предлагаемых на рынок Грид-систем.

Второе направление деятельности, очевидно, связано с требованиями клиентов к системе и с учетом результатов исследования и тестирования различных Грид-систем.

Наиболее важна, пожалуй, в настоящее время деятельность по адаптации бизнес-приложений для Грид-

среды. Сам по себе Грид как таковой не представляет особого коммерческого интереса, но исключительно важны преимущества, которые он дает по сравнению с другими вычислительными технологиями. Для компаний чрезвычайно ценны выигранное время при использовании подобного рода решений и уменьшение затрат на информационно-технологическую инфраструктуру. Грид же можно считать некоторой «ускоряющей подложкой» под эти бизнес-приложения, которые вряд ли изначально разрабатывались под Грид-платформу. И основная задача в этом направлении деятельности – как раз переработка бизнес-приложений, или создание дополнительного слоя программного обеспечения, позволяющих запускать конкретные приложения заказчиков, но уже в Грид-среде, в полной мере используя все неоспоримые преимущества Грид-технологий.

Итак, развитие и совершенствование Грид-технологий, создание Грид-инфраструктур различного уровня (национальных, региональных, проблемно-ориентированных, корпоративных и т.п.) – это задачи, стратегически важные как для развития промышленного потенциала страны, укрепления ее обороноспособности и безопасности, так и для обеспечения участия России в решении глобальных задач планетарного масштаба: мониторинг, анализ и прогноз развития политических, экономических, географических, климатических и социальных процессов на нашей планете.

