

Грид-проект EGEE – новейшие технологии развития Е-науки

Б. Джонс¹, Адылова Ф.Т.²

¹ *Директор проекта EGEE, Женева, Швейцария*

² *Институт математики и информационных технологий АН Узбекистана, Ташкент, Узбекистан*

Е-наука определяется как: "наука, выполняемая через глобальное сотрудничество на базе Интернета с использованием больших объемов данных, вычислительных ресурсов и высокоэффективной визуализации". е-наука жизненно необходима для успешной эксплуатации следующего поколения мощных средств научных исследований. Синхротроны, спутники, телескопы и лазеры,- все вместе производят терабайты данных каждый день. Их пользователи требуют эффективного доступа к географически распределенной системе хранения данных, вычислительных и сетевых ресурсов, чтобы управлять и анализировать эти данные своевременно и эффективно.

Это по существу означает, что многие области науки, в настоящее время использующие вычислительные ресурсы как часть исследования, должны иметь возможность и способность применять более мощные вычислительные ресурсы в рамках новой инфраструктуры, называемой Грид. Ученые будут иметь доступ к очень большим объемам данных и выполнять на них эксперименты в режиме реального времени, что позволит начать исследования в областях, которые до настоящего времени считались недоступными.

Проект Enabling Grids for E-science (EGEE) сделал мечту о совместном использовании вычислительных ресурсов в международном масштабе действительностью. Профинансированный Европейской Комиссией, EGEE был запущен, чтобы создать инфраструктуру для поддержки европейской науки (European Research Area). Сначала проект создал Грид, -инфраструктуру с более, чем 200 сайтами в более, чем 40 странах, сделав самую большую в мире инфраструктуру для научных исследований, которая работала в режиме 24/7. Эта инфраструктура обрабатывает до 50 000 заданий в день в восьми научных областях,- от биомедицины до термоядерных исследований. Чтобы организовать и объединить несоизмеримые вычислительные средства в одну сеть Грид, и сделать их легко доступными для пользователя, проект разработал собственное промежуточное (middleware) программное обеспечение,- gLite. Он создавался из лучших middleware компонентов, чтобы обеспечить полный диапазон основных сервисов, таких, как поддержка базовых и прикладных исследований высокого уровня.

Первый двухлетний этап проекта, успешно закончившийся 31 марта 2006, достиг цели. Сегодня проект вышел за рамки Европы, объединяя группы ученых в США и Азии, и поэтому был запущен новый проект EGEE-2, который начался 1 апреля 2006г. Проект EGEE, построенный на общеевропейской сети GÉANT, значительно расширил и объединил грид инфраструктуру Европы, связывая национальные, региональные и тематические аспекты Грид, а также обеспечивая взаимодействие с другими Грид сообществами мира. Получаемая в результате большая емкость вычислительных ресурсов и международная коммуникационная инфраструктура сильно превосходит возможности локальных кластеров и является уникальным инструментом "е-науки".

EGEE сегодня,- это более чем 100 партнеров в 32 странах, организованных в 13 федераций, которые распространяют знание о Гриде, его выгодах в среде исследователей, студентов, работающих в физике высоких энергий, науках о земле, биологии, астрофизике, вычислительной химии, термояде и других областях. Грид привлекает интерес провайдеров ИКТ и приложений в бизнесе. EGEE поддерживает связи с другими большими грид-проектами во всем мире, - Open Science Grid and TeraGrid (США),

NAREGI (Япония) и DEISA (Европа), часть из которых имеют общие точки с EGEE. Работа с этими проектами включает области защиты сети и улучшение программного обеспечения. EGEE также сотрудничает с этими проектами по стандартизации через Открытый Грид Форум.

Кроме того, EGEE сотрудничает с проектами, которые обеспечивают поддержку в области образования, и безопасности. За прошлые годы Европа вложила большие средства в программы развития е-науки, как на национальном, так и на европейском уровне, и технология грид признана одним из фундаментальных блоков прогресса этой новой области знания. Исследования не остаются в национальных рамках, и в то время как национальные инфраструктуры являются базовыми в обеспечении связью и ресурсами исследователей, они должны быть общими на международном уровне, обеспечивая глобальное научное сотрудничество.

Европейская инфраструктура грид, основанная на аналогии с инфраструктурой единой европейской энергосистемы и связанная с подобными инфраструктурами вне Европы, является подходом к будущему развитию грид. Участвуя активно в этой инициативе, EGEE помогает прокладывать путь к долгосрочной, жизнеспособной грид инфраструктуре, гарантируя, что ресурсы и знание, полученные в проекте, будут доступны исследователям в последующие годы.

Рассмотрим приложения Грид в разных областях науки.

Физика высоких энергий (High Energy Physics, HEP)

HEP-сообщество является главным пользователем EGEE инфраструктуры, и мощным двигателем новых сервисов EGEE, которые гарантируют выполнение новых приложений для новых пользователей. EGEE-сообщество вокруг HEP было сформировано на экспериментах Большого Адронного Коллайдера (Large Hadron Collider, LHC), - мощного ускорителя частиц; LHC в настоящее время строится в Европейском Центре Ядерных Исследований около Женевы, в Швейцарии. Когда коллайдер запустят в 2007г., четыре главных эксперимента, ALICE, ATLAS, CMS and LHCb будут выдавать приблизительно 15 петабайт (petabytes) информации в год, что эквивалентно стеку компакт-дисков высотой более чем 20 км. Грид - реальное и единственное решение крупномасштабного управления и обработки данных, и это понимание сегодня уже есть.

Другие международные эксперименты HEP также базируются на грид-сети проекта EGEE: это ВаВаg, CDF и эксперименты DØ, использующие ускорители частиц в США, а также ZEUS и эксперименты H1, использующие коллайдер в Германии.

Биомедицина

Биомедицина - другой главный пользователь EGEE инфраструктуры; она подразделяется на три различных области:

- Обработка медицинских изображений объединяет работу ряда связанных систем, многие из которых относятся к вычислительно емкой области регистрации и записи изображения. Это методы типа «виртуальных биопсий» в онкологии, которые позволяют избегать хирургических процедур. Понятно, что исключение дорогих хирургических процедур улучшает качество жизни пациентов, и уменьшает стоимость лечения.
- EGEE также поддерживает биоинформатику, затрагивающую широкий диапазон проблем. Они включают приложения от удаления отклонений в изображении микроскопа до доступа к грид для пользователей в областях типа анализа белковых последовательностей. Биомедицинское сообщество работает также в области обработки и управления большими потоками биологических данных.
- Создатели новых лекарств используют EGEE инфраструктуру, чтобы ускорить поиск молекул препарата-кандидата против болезней. Используя грид и программное обеспечение стыковки молекул (docking), исследователи могут быстро вывести на экран много тысяч кандидатов препарата, выбрать наиболее

перспективные составы и, таким образом, сократить большие затраты на разработку препарата.

Проект WISDOM был успешно развернут против болезней типа малярии и птичьего гриппа, проверяя миллионы потенциальных лекарств в короткие сроки. В одном испытании за несколько недель были проверены почти 200 миллионов возможных конфигураций между составами препарата и искомыми белками составов паразита малярии; использование же одного компьютера потребовало бы для этого теста сотен лет. Подобный процесс поиска лекарства для лечения потенциально смертельного H5N1 птичьего вируса гриппа, требует анализа 300 000 возможных компонентов препарата. Занимаясь проблемами малярии и птичьего гриппа, проект способен продемонстрировать потенциал грида, как базы решения современных научных проблем.

Проект GEMSS (Grid-Enabled Medical Simulation Services) показал, что Грид может быть использован для обеспечения доступа исследователям и практикам медицины к современным средствам моделирования и обработки изображений для усовершенствования до-оперативного планирования и поддержки операций в режиме реального времени.

Проект MAMMOGRID имеет следующие основные цели:

- Оценить Грид-технологии и определить требования организации на этой основе европейской базы данных по маммографии .
- Внедрить базу данных, используя новейшие Грид технологии в области хранения данных, которые обеспечат доступ к распределенным данным, и позволят быстро развернуть соответствующее программное обеспечение для обработки.
- Эксплуатировать базу данных MammoGrid, алгоритмы для целей начального контроля качества данных, и, в конечном счете, обеспечить проведение сопоставительных исследований.

Астрофизика

Две главных Виртуальных Организации в этой области, Planck and MAGIC, совместно работают над проблемами получения данных, моделирования, их хранения и поиска. Спутник Planck Европейского Космического Агентства будет запущен в 2008 г. с целью ультракоротковолнового картирования неба с беспрецедентной комбинацией неба и охвата частоты, точности, стабильности и чувствительности. MAGIC моделирует поведение атмосферных ливней, порожденных первичными космическими лучами. Это моделирование необходимо, чтобы анализировать данные телескопа, расположенного на Канарских островах, и изучать происхождение и свойства гамма лучей высокой энергии.

Приложения в науках о земле (Earth Science Research)

Эти приложения охватывают большой диапазон проблем, связанных с землей, атмосферой, океаном, а также с проблемами на границах между атмосферой планеты и ее ядром. Недавно члены Виртуальной Организации ESR провели быстрый анализ землетрясения, помогая научному сообществу лучше понять эти разрушительные естественные бедствия.

Приложения в геофизике

Геокластер (Geocluster),-промышленное решение обработки сейсмических данных, является первым приложением, успешно выполняющимся на EGEE. Доступ к инфраструктуре EGEE осуществляется через виртуальную организацию EGEODE (Expanding GEOsciences on DEMand), который предназначен для поддержки геофизических исследований, как частных, так и академических лабораторий. Это дает возможность исследователям обрабатывать сейсмические данные и исследовать структуру Земли.

Приложения в термоядерных реакциях

Несколько приложений из этой области уже выполняются на EGEE инфраструктуре: трассировка плотного луча, оценка траектории микроволнового луча в плазме; кинетический транспорт и оптимизация специальных магнитных устройств ограждения. ITER,-это объединенный международный научно-исследовательский проект, который имеет целью продемонстрировать научно-техническую выполнимость термояда; он может потенциально произвести 500 МВт энергии к 2016г. Инфраструктура EGEE в настоящее время используется, чтобы развить приложения в области магнитной оболочки термояда; эти результаты могут представлять интерес для ITER .

Приложения в вычислительной химии

Главный пользователь в области вычислительной химии -проект GEMSS (Grid-Enabled Medical Simulation Services) или молекулярный тренажер. Несколько грид-приложений были выполнены на промышленной основе: вычисление наблюдаемых величин химических реакций, моделирование молекулярной динамики сложных систем, вычисление электронной структуры молекул, молекулярных совокупностей, жидкостей и тел.

Стандартные блоки Грид

Главная цель проекта EGEE состоит в том, чтобы строить, управлять, и использовать качественную международную инфраструктуру Грид, которая используется разными сообществами пользователей e-науки. Новое сотрудничество и новые услуги возникают непрерывно, и EGEE инфраструктура приспосабливается к ним. Компьютеры и ресурсы памяти, работающие в проекте, образуются за счет большого и растущего числа ресурсных центров во всех регионах, которые координируются так называемыми Региональными Операционными Центрами(Regional Operations Centers,ROC).

gLite связывает эти ресурсы в единую инфраструктуру, которая обеспечивает доступ для пользовательских сообществ проекта. Пользователи, таким образом, имеют доступ к ресурсам, которые не были бы доступны без грид, и без которых были бы невозможны некоторые научные исследования. Такая инфраструктура облегчает сотрудничество географически рассеянных сообществ, которые могут также совместно использовать данные и ресурсы. В дополнение к ресурсам и услугам, обеспечиваемыми EGEE, существенные дополнительные ресурсы, доступные исследовательскому сообществу, были получены через соответствующие инфраструктуры и взаимодействие с другими грид инфраструктурами по всему миру. Множество исследовательских и деловых групп также вкладывает свои ресурсы с тем, чтобы получить опыт как провайдеры инфраструктуры. Инфраструктура грид использует более чем 200 центров ресурсов, обеспечивающих включение в грид 35 000 процессоров, с потенциальной загрузкой до 50 000 заданий в день. Все федерации, участвующие в деятельности проекта, имеют ресурсы в наличии и связаны с EGEE. Кроме того, есть приблизительно 4 000 процессоров, связанных через не-EGEE сайты в Китае, США, Канаде, Пакистане и Индии. Последовательный доступ к ним поддерживается через согласованную политику управления доступа на уровне сервисов.

Организация и управление пользовательской поддержкой в глобальной научной вычислительной инфраструктуре типа EGEE является одним из достижений и перспектив Грид. Учитывая распределенный характер организации проекта, и распространение опыта установки, конфигурирования, управления и поиска неисправностей в грид middleware , проект разработал Глобальную Поддержку Пользователя Грид (GGUS),которая является системой с распределенной инфраструктурой и центральной координацией. Первая группа линии поддержки обеспечивает поддержку проблемам создания грид, в то время как специализированные модули отвечают за middleware, развертывание, сеть, инфраструктуры грид и проблемы

виртуальных организаций. В настоящее время, система GGUS обрабатывает до 1200 запросов в месяц.

Работа с Бизнесом и Промышленностью

Европейская Комиссия сделала большие инвестиции в проекты грид типа EGEE. С быстро развивающейся технологией грид возникает ключевой вопрос - как передать технологии от исследовательских проектов в область коммерческих услуг и вернуть вложенные Комиссией средства. Есть много признаков, что рынок грид уже есть. Однако есть несколько барьеров - технический и нетехнический, которые надо преодолеть. Проект EGEE твердо привержен налаживанию прочных отношений с промышленностью и предлагает разнообразные каналы для вовлечения бизнеса,- от совместно используемой информации, и делового анализа требований, к полному техническому сотрудничеству

Деловые Партнеры

EGEE обеспечивает механизм для разных видов коммерческой деятельности, консолидируя интерес промышленности к проекту через аналитические обзоры рынков, стратегии эксплуатации и технологического развития.

GridwiseTech, - консалтинговая фирма Грид с отделениями в Польше и Соединенных Штатах. Фирма работала с EGEE в течение некоторого времени, обеспечивая разработку порталов EGEE, инфраструктуры и объединяя EGEE с инфраструктурами других крупных фирм. Через программу ЕВА(European Business Activities) они будут работать с проектом над продвижением его в коммерческий сектор, особенно в части внедрения EGEE технологий.

NICE, другой долговременный партнер EGEE, будет работать с проектом, гарантируя, что портал GENIUS Grid и EnginFrame FrameWork будут совместимы с gLite .

Форум Промышленности

На EGEE Форум Промышленности объединяет приблизительно 250 членов, из 100 институтов 20 стран, гарантируя эффективную двухстороннюю связь экспертов проекта и промышленности и действенные обсуждения заинтересованными сторонами проблем внедрения Грид в бизнес. Это продвигает Грид в промышленность, определяя препятствия и предлагая средства их преодоления. Активный двухсторонний форум в руководящем комитете, состоящем из промышленных партнеров проекта и представителей бизнеса, возглавляет промышленность. Целевая группа по промышленности (Industry Task Force) сводит представителей промышленных приложений на EGEE инфраструктуре, включая проекты, нацеленные на промышленность, с партнерами из промышленности. На уровне научных приложений в Технической Группе Координации, ITF улучшает понимание требований промышленности и влияет на технические направления проекта, с учетом потребностей бизнеса.

Geocluster- первое индустриальное приложение, выполняющееся на EGEE

Геокластер (Geocluster) является индустриальным приложением, успешно выполняющимся на EGEE в рамках расширяющейся виртуальной организации Expanding GEOsciences on DEMand (EGEODE) по геофизическим исследованиям. Организация EGEODE проводит геофизические исследования для государственного и частного сектора, в том числе, и для академических лабораторий. Геокластер,- это есть программное обеспечение, которое разработано и выставлено на продажу компанией Compagnie Générale de Géophysique (CGG), Франция; оно включает несколько инструментальных средств для обработки сигналов, моделирования и инверсии. Это дает возможность исследователям нескольких стран в Европе, Азии и Америке обрабатывать сейсмические данные и исследовать состав земных слоев.

Перспективные направления работы

Европа вложила большие средства в программы е-науки за прошлые годы на национальных и европейском уровнях, получив внушительные результаты. Благодаря сильной поддержке европейской Комиссии, Европа утвердилась как мировой лидер в области Грид технологии, которая признана как фундаментальный компонент е-науки. Быстрый рост числа Грид-сообществ говорит о том, что многие уже полагаются на эту технологию в своей работе. Поскольку Виртуальные Организации, установленные научными сообществами в EGEE, прогрессируют от испытания Грид-приложений к рутинному использованию, становится актуальным гарантировать, что надежные и адаптивные инфраструктуры Грид будут поддержаны, независимо от циклов финансирования проекта. Без такой гарантии, научные сообщества будут полагаться на эту новую вычислительную парадигму с большой осторожностью, потенциально подвергая опасности большие инвестиции, которые уже были сделаны. Управляемая потребностями и требованиями исследовательского сообщества, предложенная европейская инициатива Грид, как ожидается, создаст базу новому прорыву в исследовательских инфраструктурах, поддерживая этим совместные научные открытия в европейской науке. EGEE обеспечивает эффективный подход к развитию политики продвижения Грид, основанный на непосредственном опыте работы с вычислительной инфраструктурой. Грид технологии могут удовлетворить потребности многих сообществ, будь то бизнес или академия. Поэтому работа EGEE могла бы быть подхвачена самым широким сообществом в последующие годы.

Развитие Грид в Узбекистане

Благодаря двум грантам Академии Наук республики, и гранту НАТО, Институт математики и информационных технологий начал работу по созданию грид узла. Работа ведется совместно с Объединенным Институтом Ядерных Исследований (Дубна, Россия), и кураторстве проекта EGEE.

Мы ожидаем получить знания и навыки в Грид, чтобы стать профессиональным ресурсным центром Грид-технологий в Узбекистане и поддержать развитие ИКТ в региональном академическом, медицинском и образовательном сообществе. Кроме того, мы планируем подключиться к проекту EGEE и в будущем к другим международным сообществам в области Грид, способствуя международному сотрудничеству.