

## ГРИД-ТЕХНОЛОГИИ: СТАТУС И ПЕРСПЕКТИВЫ

В. В. Кореньков

Объединенный институт ядерных исследований, Дубна

### GRID-Technologies: Status and Prospectives

V. V. Korenkov

Joint Institute for Nuclear Research, Dubna

В статье представлены базовые понятия концепции грид и перспективы использования современных грид-технологий для различных отраслей науки, промышленности и бизнеса. Дано краткое описание проектов по созданию грид-инфраструктур. Особое внимание уделено участию России в глобальных проектах EGEE и WLCG и созданию российского грид-консорциума РДИГ.

The conception of grid and perspectives of the usage of modern grid technologies for various fields of science, commerce and business are presented. Current projects on creation of grid-infrastructures are shortly described. A special attention is paid to the participation of Russia in the EGEE and WLCG global projects and the creation of Russian grid-consortium RDIG.

#### Концепция грид

Развитие научных исследований в физике высоких энергий, астрофизике, биологии, науках о Земле, химии, а также в медицине, нанотехнологиях, промышленности, бизнесе и других направлениях человеческой деятельности требуют совместной работы многих организаций по обработке большого объема данных в относительно короткие сроки. Для этого необходимы географически распределенные вычислительные системы, способные передавать, обрабатывать и хранить огромные массивы данных.

В настоящее время в мире информационных технологий интенсивно развивается грид (название по аналогии с электрическими сетями — electric power grid) — это компьютерная инфраструктура нового типа, обеспечивающая глобальную интеграцию информационных и вычислительных ресурсов [1, 2]. Суть инициативы грид состоит в создании набора стандартизированных служб для обеспечения надежного, совместимого, дешевого и безопасного доступа к географически распределенным высокотехнологичным информационным и вычислительным ресурсам — отдельным компьютерам, кластерам и суперкомпьютерным центрам, хранилищам информации, сетям, научному инструментарию и т. д.

Миссия Интернет состояла в глобализации обмена информацией, а всемирная паутина WWW стандартизовала поиск и доставку документов. Грид стал следующим этапом в этой цепочке революционных преобразований — стандартизации и глобализации использования всех видов компьютерных ресурсов.

Грид — это система, которая координирует использование ресурсов при отсутствии централизованно-

го управления этими ресурсами; использует стандартные, открытые, универсальные протоколы и интерфейсы; обеспечивает высококачественное обслуживание.

Грид предлагает технологию доступа к общим ресурсам и службам в рамках виртуальных организаций. Виртуальная организация (VO) — является сообществом пользователей грид-системы, объединённых для решения проблем в режиме скоординированного распределения ресурсов. В каждой виртуальной организации имеется своя собственная политика поведения ее участников, которые должны соблюдать установленные правила. Виртуальная организация может образовываться динамически и иметь ограниченное время существования.

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 государственном управлении, в медицине, в промышленности, в бизнесе. Появились и активно используются новые термины e-Science, e-Health, e-Commerce, которые подчеркивают теснейшую связь в развитии науки, медицины, бизнеса с современными информационными технологиями, в первую очередь, с грид-технологиями (см. рис. 1). Таким образом, грид служит универсальной эффективной инфраструктурой для высокопроизводительных распределенных вычислений и обработки данных. Общая схема взаимодействия пользователей, программной среды и ресурсов показана на рис. 2

К приложениям грид относятся:

- сложное моделирование на удаленных суперкомпьютерах;

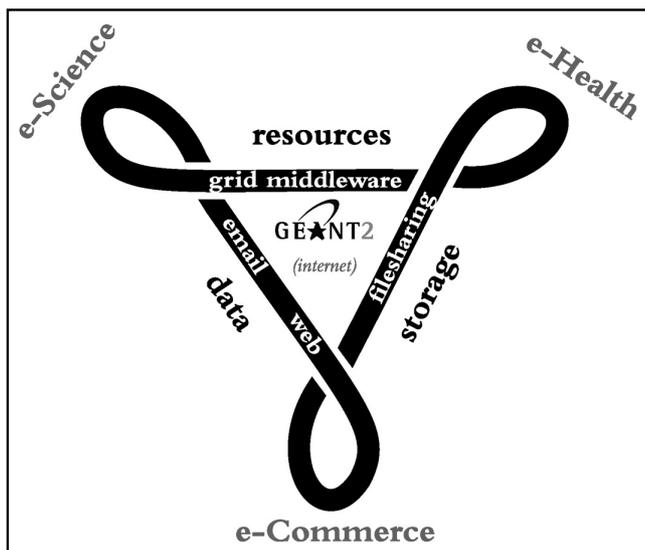


Рис. 1. Грид: наука, медицина и коммерция, использующие ресурсы глобальной информационно-вычислительной инфраструктуры на основе современных грид-технологий.

- совместная визуализация очень больших наборов научных данных;
- распределенная обработка в целях анализа данных;
- соединение научного инструментария с удаленными компьютерами и архивами данных.

Среди основных направлений использования грид на данный момент можно выделить:

- организация эффективного использования ресурсов для небольших задач, с утилизацией временно простаивающих компьютерных ресурсов;
- распределенные супервычисления, решение очень крупных задач, требующих огромных процессорных ресурсов, памяти и т. д.;

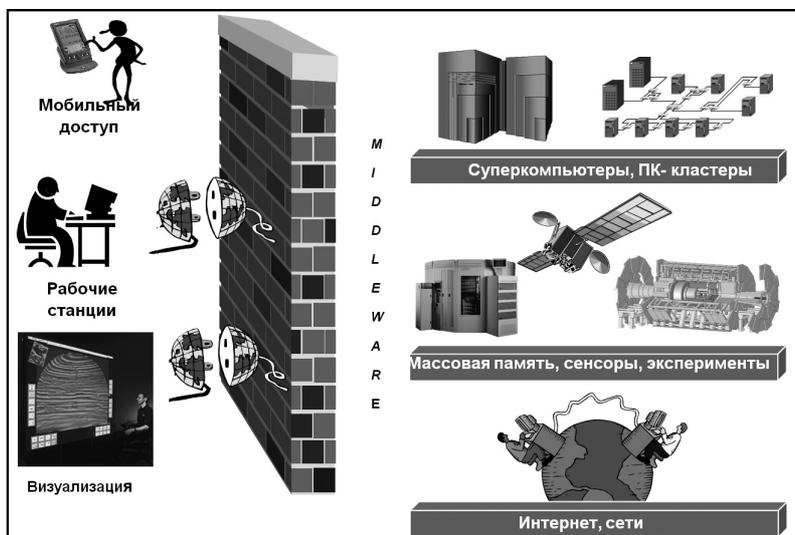


Рис. 2. Общая схема взаимодействия пользователей, программной среды и ресурсов в грид-инфраструктуре.

- вычисления с привлечением больших объемов географически распределенных данных, например, в метеорологии, астрономии, физике высоких энергий, медицине, науках о земле;
- коллективные вычисления, в которых одновременно принимают участие пользователи из различных организаций.

Таким образом, грид — это соединение технологии, инфраструктуры и стандартов. Здесь технология — это специальное программное обеспечение, которое позволяет предоставлять ресурсы (компьютеры, хранилища данных, сети и другие) в общее пользование, а потребителям — использовать их, когда необходимо. Инфраструктура состоит из аппаратных средств и служб (на основе человеческих и программных ресурсов), которые должны быть организованы и должны постоянно поддерживаться для того, чтобы ресурсы могли совместно использоваться. Наконец, стандарты должны определять формат и протоколы обмена сообщениями как между службами, так и между службами и пользователями, а также правила работы грида [3].

### Проекты развития грид-технологий

В мире накоплен большой опыт создания программной среды (middleware) для реализации распределенной грид-инфраструктуры. Одним из первых и наиболее популярных программных решений, которое стало стандартом де-факто на реализацию грид-систем — это Globus Toolkit [4]. Данный инструментарий реализует механизмы сервисов, которые охватывают вопросы защиты, обнаружения информации, управления данными и ресурсами, коммуникации, обнаружения ошибок и т. д. В настоящее время этот инструментарий применяется во многих проектах по всему миру. Кроме Globus Toolkit существуют много других программных сред (middleware) для реализации грид-систем: Legion, Condor, Unicore, ARC, gLite [5] и т. д.

В этом веке было создано большое количество масштабных грид-инфраструктур, среди которых — DEISA (Европа), NAREGI (Япония), Open Science Grid («Открытый научный грид») и TeraGrid в США, GridPP (Великобритания), INFN GRID (Италия), NorduGRID (Скандинавия).

Особое место среди этих проектов занимают EGEE (Enabling Grids for E-scienceE)[6] и WLCG (Worldwide LHC Computing GRID) [7].

Проект EGEE («Развёртывание грид-систем для развития е-науки») воплотил в действительность замысел превратить мировые компьютерные ресурсы в единую однородную среду, где ими можно пользоваться совместно в мировом масштабе.

Проект финансируется Европейским Сообществом и странами-участниками. В

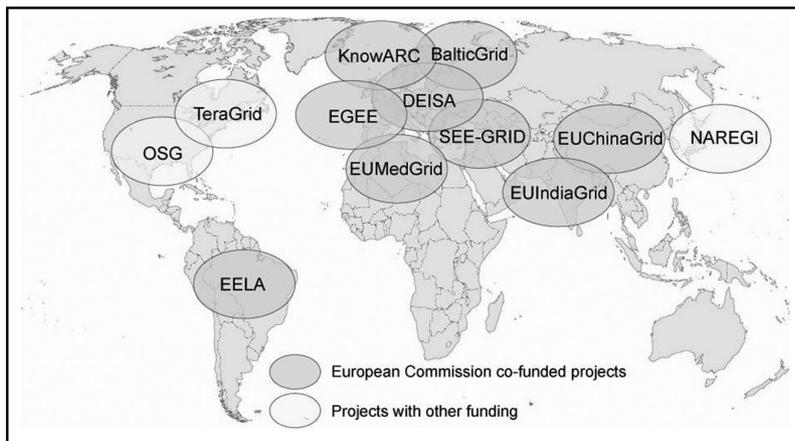


Рис. 3. Общая инфраструктура проекта EGEE и родственных проектов.

результате появилась высокопроизводительная всемирная инфраструктура, намного превосходящая по своим возможностям локальные кластеры и отдельные центры.

В консорциум EGEE входят свыше 140 организаций из более чем 50 стран, которые объединены в 13 федераций, одна из которых «Russia», и представляют почти все основные европейские международные и национальные грид-проекты, а также проекты в США и Азии. Кроме того, множество родственных проектов (более 50 ассоциированных участников) распространяют грид-инфраструктуру на Средиземноморье, Балтику, Латинскую Америку, Индию и Китай (рис. 3).

Грид-инфраструктура EGEE уже стала повседневным рабочим средством для целого ряда больших и малых исследовательских сообществ. В ней работают приложения для физики высоких энергий, биологических наук и смежных дисциплин, наук о Земле, астрофизики, вычислительной химии, термоядерной энергетики и других. Число пользователей инфраструктуры EGEE более 14000, и они объединены в более чем 200 виртуальных организациях. В настоящее время в день выполняется более 400 тысяч заданий (более 12 миллионов в месяц), и с каждым месяцем эти показатели растут.

Проект WLCG (Worldwide LHC Computing GRID) принят в 2001 году в ЦЕРНе с целью создания глобальной информационно-вычислительной инфраструктуры для обработки, хранения и анализа данных, полученных во время экспериментов, проводимых на Большом адронном коллайдере. Для реализации этой грандиозной задачи построена масштабная глобальная грид-инфраструктура на основе региональных центров различного уровня, обеспечивающая моделирование, хранение, передачу данных с Большого адронного коллайдера.

### Пользователи и виртуальные организации EGEE

Сообщество ученых, занимающихся физикой высоких энергий (ФВЭ) — главный пользователь инфраструктуры EGEE.

Биомедицинские науки представляют другое основное сообщество пользователей инфраструктуры EGEE. Эта сфера четко делится на три области. Прежде всего, это разработка систем регистрации, надежного хранения и оперативного анализа 2D и 3D медицинских изображений (различные томограммы, «маммограммы», «виртуальные биопсии», кардиограммы и т. д.). Далее следуют различные направления биоинформатики — от анализа белковых цепочек до оптимизации обработки биологических данных. Наконец, Грид-инфраструктура используется в разработке лекарств для ускорения расчёта пристыковки молекул (молекулярного докинга): исследователи быстро просматривают многие тысячи вариантов и отбирают наиболее перспективные соединения, что снижает высокую стоимость разработки препаратов.

Исследователи наук о Земле решают широкий круг проблем в рамках виртуальной организации ESR (Earth Science Research, «Исследования в области наук о Земле»).

В инфраструктуре EGEE работают несколько приложений для исследований в области термоядерного синтеза. Для демонстрации принципиальной и технической возможности термоядерного синтеза учреждён ITER — международный проект соответствующих исследований и разработок.

Главным пользователем грид-инфраструктуры в области вычислительной химии является GEMS — приложение для молекулярного моделирования. В грид-инфраструктуре размещены и используются на промышленном уровне несколько приложений для расчёта химических реакций, моделирования молекулярной динамики сложных систем и расчёта электронной структуры молекул.

### Российский грид-консорциум

Чтобы обеспечить полномасштабное участие России в осуществлении проектов EGEE и WLCG в 2003 году был образован консорциум РДИГ (Российский ГРИД для интенсивных операций с данными, Russian Data Intensive GRID, RDIG) [8,9]. Меморандум о создании консорциума был подписан руководителями восьми крупных институтов: Института физики высоких энергий (Протвино), Института математических проблем биологии (Пушино), Института теоретической и экспериментальной физики (Москва), Объединенного института ядерных исследований (Дубна), Института прикладной математики им. М.В. Келдыша (Москва), НИИ ядерной физики МГУ (Москва), Петербургского института ядерной физики (Санкт-Петербург) и РИЦ «Курчатовский институт» (Москва), а с 2008 года — Геофизического центра РАН (Москва). Схематическая карта состава РДИГ представлена на рис. 4.



Рис. 4. Консорциум РДИГ.

Консорциум РДИГ входит в структуру EGEE в качестве региональной федерации «Россия», и его целью является создание действующей грид-инфраструктуры в России, что включает в себя:

- наращивание вычислительных ресурсов и ресурсов хранения данных российского сегмента грид-среды;
- обеспечение надежной сетевой инфраструктуры;
- обеспечение работы базовых грид-сервисов в российском сегменте;
- создание Регионального операционного центра (Regional Operations Center — ROC);
- помощь ресурсным центрам в установке ППО и поддержка его функционирования;
- поддержка пользователей грида;
- участие в предоставлении ресурсов для исследований в важных прикладных областях (в области биомедицины, термоядерного синтеза, физики высоких энергий, космофизики);
- управление функционированием инфраструктуры: регистрация пользователей, региональных виртуальных организаций и мониторинг;
- популяризация технологий и вовлечение новых пользователей из научных и производственных кругов, а также обучение пользователей и администраторов.

В 2009 году была создана распределенная инфраструктура для обучения грид-технологиям [10], в которую на данный момент включены грид-сайты Москов-

ской области (Дубна и Протвино) и грид-сайты в Узбекистане (Ташкент), Болгарии (София) и Украине (Киев). Эта инфраструктура не только успешно используется для обучающих целей, но и способствует развитию международного сотрудничества РДИГ.

В настоящее время в рамках РДИГ работают более 100 ученых и специалистов в области компьютерных технологий, подключено 17 ресурсных центров, предоставляющих ресурсы для грид-среды, с общим числом процессорных узлов более 3000 и с общим объемом хранилища данных около 2 Петабайт.

Российские ученые уже сейчас активно используют грид-инфраструктуру для своих исследований, но не все направления науки включились в процесс освоения этой новой инновационной технологии.

## Перспективы развития грид-технологий

Грид-технологии вступают в пору зрелости — происходит переход от пилотных приложений к постоянной устойчивой работе по обслуживанию самых разнообразных прикладных областей науки и производства. Масштаб внедрения грид-технологий в значительной степени будет зависеть от способности новых и существующих приложений быть развернутыми в этой среде. Поэтому важно обеспечить простые (графические, основанные на технологии веб-порталов) интерфейсы к системам запуска и мониторинга заданий и управления данными в гриде.

Создание удобных интерфейсов для конкретных приложений является важным для того, чтобы сделать грид-технологии доступными широкому кругу пользователей в области промышленного производства и других бизнес-приложений.

Что касается проекта EGEE, то планируется, что с 2010 года начнет функционировать Европейская грид инфраструктура (EGI — European Grid Infrastructure). В основе этой инициативы лежит сотрудничество между национальными грид инфраструктурами (National Grid Initiatives, NGIs) и координирующей организацией (the EGI Organisation, EGI.eu). Это сотрудничество должно обеспечить дальнейшее развитие устойчивой и постоянно действующей глобальной грид-инфраструктуры, обеспечивающей оптимальное использование вычислительных ресурсов и ресурсов хранения данных.

### Литература

1. Foster I., Kesselman C. (eds), «The Grid, Blueprint for a New computing Infrastructure», Morgan Kaufmann Publishers, 1998.
2. Foster I., Kesselman C. (eds), «The Grid 2: Blueprint for a New Computing Infrastructure», Morgan Kaufmann Publishers, 2004.
3. Демичев А. П., Ильин В. А., Крюков А. П., Шамардин А. В. Грид-технологии на службе компьютеринга для Большого адронного коллайдера // Информатизация образования и науку — М., 2009. 4: 158. [http://www.informika.ru/about/informatization\\_pub/publications/2009/4/4-p158.pdf](http://www.informika.ru/about/informatization_pub/publications/2009/4/4-p158.pdf)
4. <http://www.globus.org>
5. Промежуточное программное обеспечение gLite: <http://glite.web.cern.ch/glite>
6. Европейский проект EGEE: <http://www.eu-egee.org>
7. Проект WLCG: <http://lcg.web.cern.ch/LCG>
8. Проект РДИГ: <http://egee-rdig.ru>.
9. Ильин В., Кореньков В., Солдатов А. Российский сегмент глобальной инфраструктуры LCG // Открытые системы, 2003. 1.
10. Кореньков В., Кутковский Н. Инфраструктура обучения грид-технологиям // Открытые системы, 2009. 11: 48—51.