



Enabling Grids for E-science

## Применение технологии Грид для БАК

*Котляр В.В.  
Институт Физики Высоких Энергий  
г.Протвино*

[www.eu-egee.org](http://www.eu-egee.org)

- **1985–1995** Программа Национального Научного Фонда США «National Science Foundation (NSF) Supercomputer Centers». Результат – создание развитой коммуникационной инфраструктуры и нескольких суперкомпьютерных центров для поддержки академических работ и исследований.
- **Октябрь 1997** NSF инициировал новую программу развития информационных технологий – Partnerships for Advanced Computational Infrastructure (PACI). Возникла концепция Грид-инфраструктуры, как некая аналогия электроэнергетической инфраструктуры. Предполагалось, что Grid будет использоваться в основном для того, чтобы задействовать **простаивающие** вычислительные компьютерные ресурсы.
- **1998** Создан (и успешно развивается) инструментальный пакет **Globus Toolkit**. Является технологической базой создания Грид-инфраструктуры (**стандарт de facto**)
- **1999** Сформировалось (и активно действует) **международное научное Грид-сообщество** – Global Grid Forum (GGF)
- **2002** GGF и IBM была представлена новая системная разработка – Open Grid Service Architecture (OGSA), инкорпорирующая в грид понятия и стандарты веб-сервисов
- **2003** Создано объединение **Enterprise Grid Alliance (EGA)** – адаптация академических разработок по Грид-технологиям к потребностям производителей (Fujitsu, Siemens, HP, Intel, NEC, Oracle, Sun, EMC)
- **2006** GGF и EGA объявили о слиянии и образовании **Open Grid Forum (OGF)**

В упрощенном виде первоначальная идея заключалась в том, что вычислительными ресурсами можно было бы воспользоваться так же, как обычные люди пользуются электросетью: как и бытовой прибор, вы включаете свой компьютер в розетку, запускаете свое задание, оно куда-то уходит, где-то выполняется, а вам только остается ждать ответ. Где выполняется задание, какие вычислительные ресурсы используются, вас, как пользователя вообще не должно волновать.

- ростом требований к вычислительным мощностям при решении сложных научных, инженерных, производственных и бизнес-задач;
- широким распространением высокоскоростных сетей передачи данных, таких как Internet, и выделенных сетей для связи научных организаций;
- наличием локальных вычислительных ресурсов во многих организациях (кластеров персональных компьютеров или специализированных суперкомпьютеров);
- необходимостью объединения ресурсов различных организаций, работающих над одной задачей.

## Типы грид-систем с позиций решаемых задач

- **Вычислительный Грид для максимальной скорости вычислений за счет распределения задач между различными центрами (Computational Grid)**
  - Высокопроизводительные вычисления (HPC)
    - сокращение времени выполнения задания, допускающего параллельные вычисления
    - требование: **поддержка параллельных вычислений**
  - Высокоэффективные вычисления (HTC)
    - выполнение как можно большего количества схожих заданий за определённый период
    - требование: **использование “свободных” ЦПУ полностью или частично**
- **Грид для интенсивной обработки огромных массивов данных (Data Grid)**
  - Как правило, с использованием физически распределённых баз данных
- **Семантический Грид для оперирования данными из различных баз данных (Semantic Grid)**
  - Интеграция пользователей из различных областей знаний для решения сложных задач

- Термин **ГРИД** был выбран по аналогии с энергетическими сетями (**electric power grid**)

Грид – это согласованная, открытая и стандартизованная среда, обеспечивающая гибкое, безопасное, скоординированное, раздельное использование разнообразных компьютерных ресурсов виртуальными организациями.



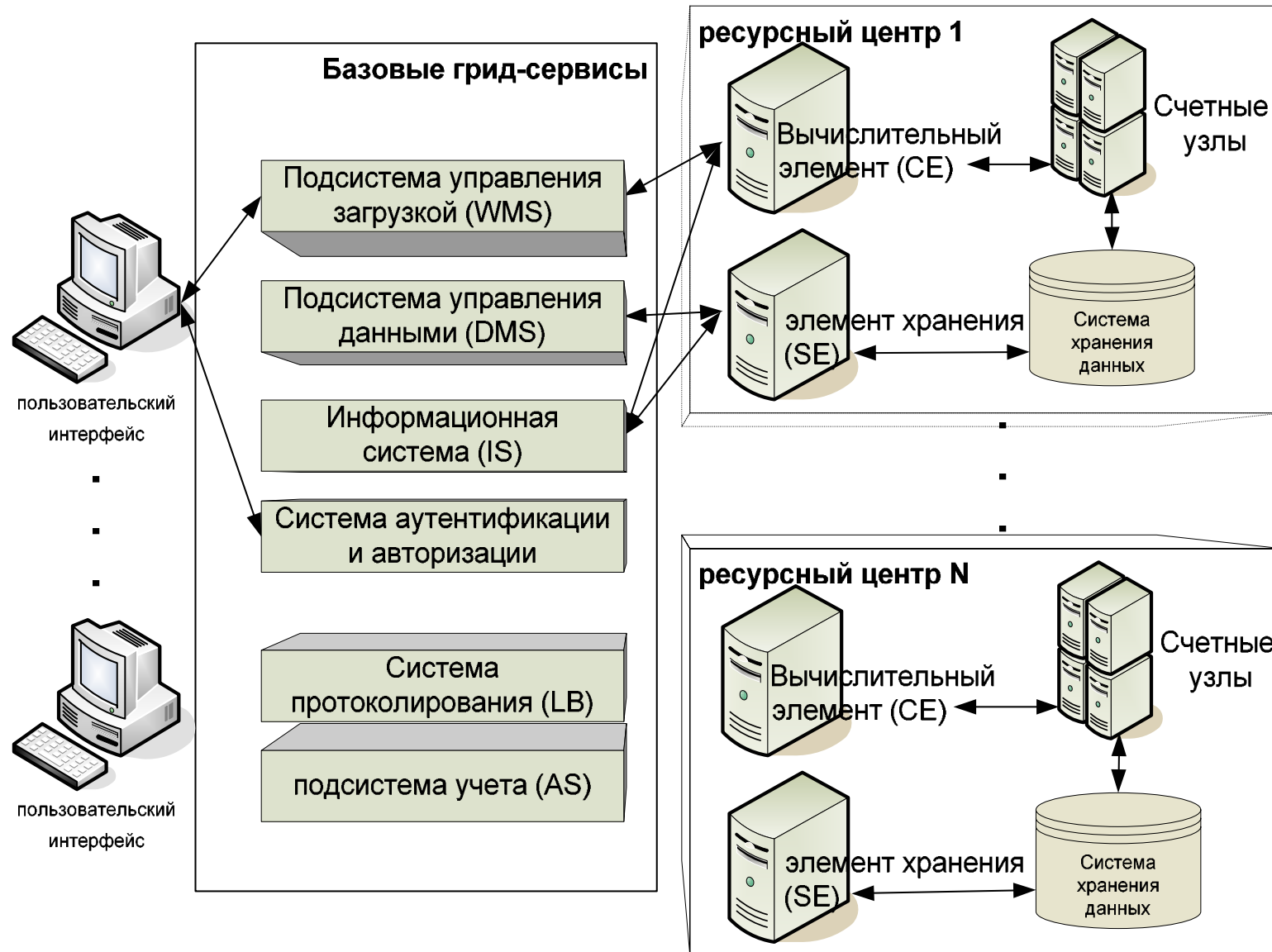
- Сотрудничество – это аспект Грид, подразумевающий справедливое разделение ресурсов методами распределенного компьютеринга.
- Объединение, когда каждый Грид больше суммы всех своих частей: Грид объединяет в одно целое много ресурсов, и поэтому дает нам объединение мощностей индивидуальных ресурсов в виртуальный ресурс с мощностями более высокого уровня, при этом возможности индивидуальных ресурсов сохраняются.
- Грид предоставляет службы, следуя принципам ориентированной на службы архитектуры.
- Грид состоит из «гетерогенных вычислительных ресурсов», т.е. имеется многообразие различных компонент аппаратных устройств и программного обеспечения с различными характеристиками латентности и эффективности.
- Ресурсы в Грид находятся в собственности разных владельцев, и одним из требований Грид является использование механизмов распределенного (децентрализованного) управления.

- Грид должен разрешать обращаться к вычислительной инфраструктуре и тем пользователям, которые не обладают глубокими знаниями базовой архитектуры или сетевой топологии.
- Грид должен быть «динамически реконфигурируемым» и легко масштабируемым.
- У авторизованных пользователей и приложений должен быть ограниченный набор операций, которые они могут выполнять с помощью служб. Безопасность Грид – один из первостепенных моментов, с которым реальные пользователи имеют дело.

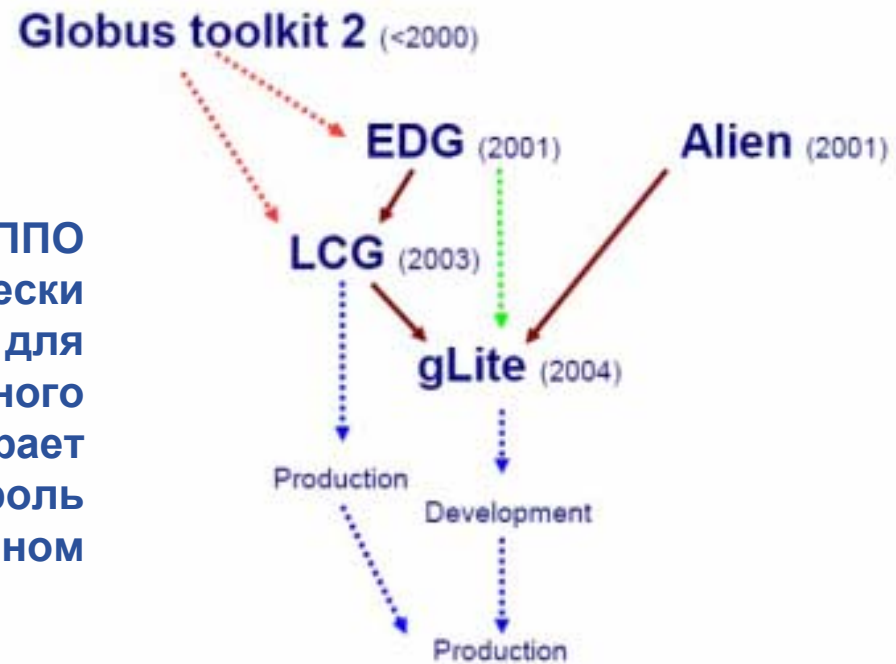


- вычислительные ресурсы (отдельные компьютеры, кластеры);
- ресурсы хранения данных (диски и дисковые массивы, ленты, системы массового хранения данных);
- сетевыми ресурсами;
- программное обеспечение (какое-либо специализированное ПО).

технология Грид не является технологией параллельных вычислений, и в ее задачи входит лишь координация использования ресурсов.



По причине использования одного ППО распределенное географически множество ресурсов представляется для пользователей в качестве единого ресурса. Промежуточное ПО играет базовую роль в Грид, например, как роль операционной системы на персональном компьютере.



**gLite 3.1** (2007)



## Виртуальная организация

Виртуальная организация (ВО) представляет собой динамическое сообщество людей, которое использует ресурсы Грид в соответствии с правилами, согласованными между ним и собственниками ресурсных центров.

Обычно в виртуальные организации объединяются научные сотрудники из различных институтов. ВО помогает организовывать временные межинститутские и международные коллаборации исследователей для решения крупных проблем. При этом, благодаря технологиям Грид, они получают доступ к общим ресурсам и могут работать с единым программным обеспечением, что обеспечивает достоверность и сравнимость результатов, полученных разными группами исследователей.

Средства аутентификации должны обладать следующими свойствами:

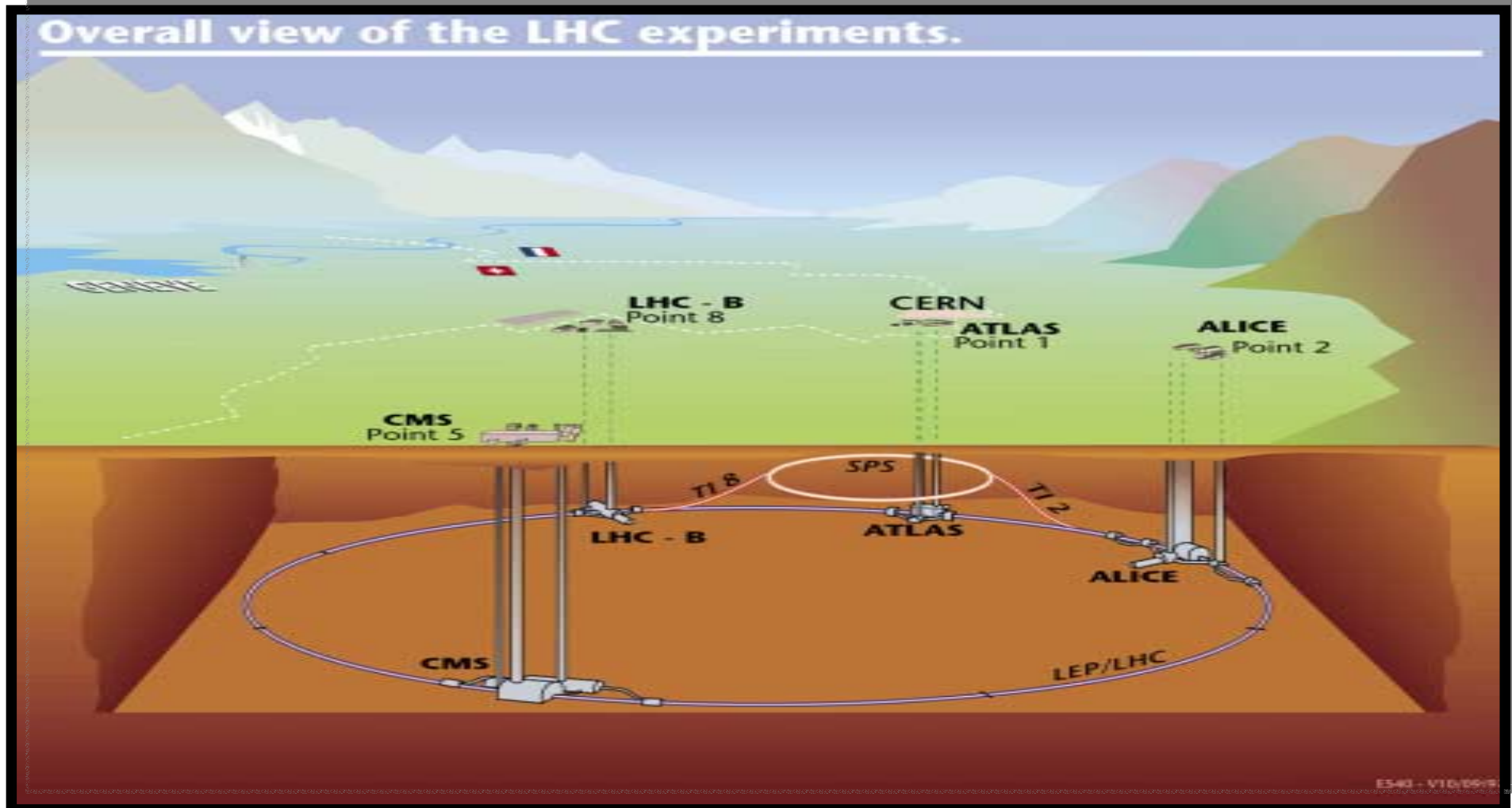
- ❖ Системой единого входа. Пользователь аутентифицируется и регистрируется только один раз в начале работы, получая тем самым доступ ко всем разрешенным ресурсам базового уровня системы Грид.
- ❖ Делегированием прав. Пользователь должен иметь возможность запуска задач от своего имени, и таким образом получать доступ ко всем ресурсам, на которых авторизован пользователь. В свою очередь запущенные задачи могут делегировать часть своих прав другим программам.
- ❖ Системой доверительных отношений к пользователю. Пользователь должен иметь возможность работы с ресурсами нескольких поставщиков, и при конфигурации защищенной среды пользователя система безопасности не должна требовать дополнительного взаимодействия поставщиков ресурсов друг с другом.

Наличие цифрового сертификата является обязательным шагом для получения доступа к системе Грид.

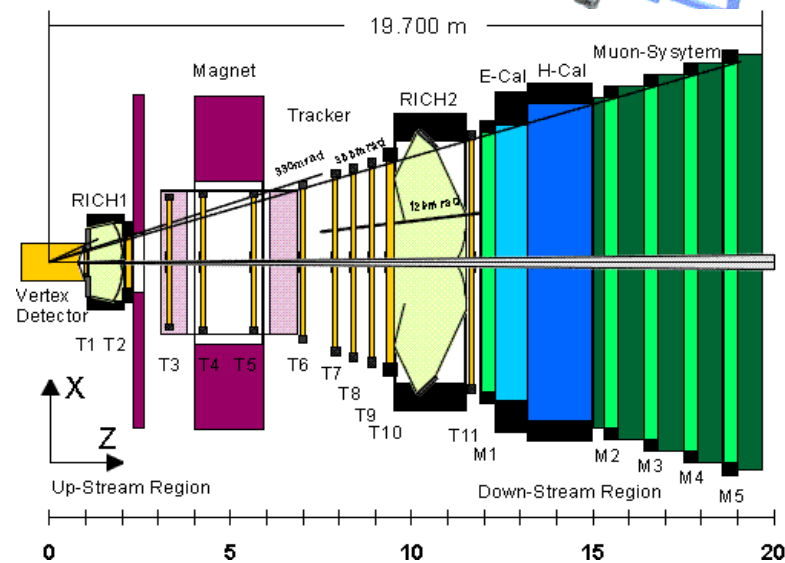
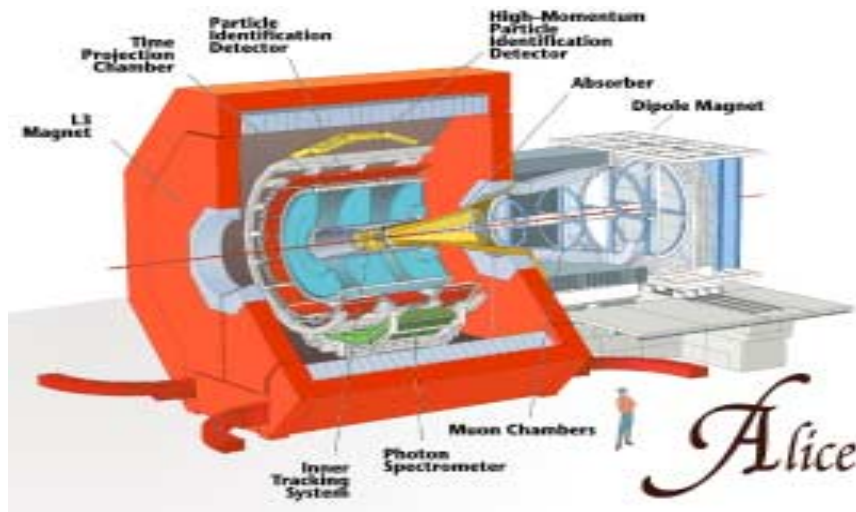
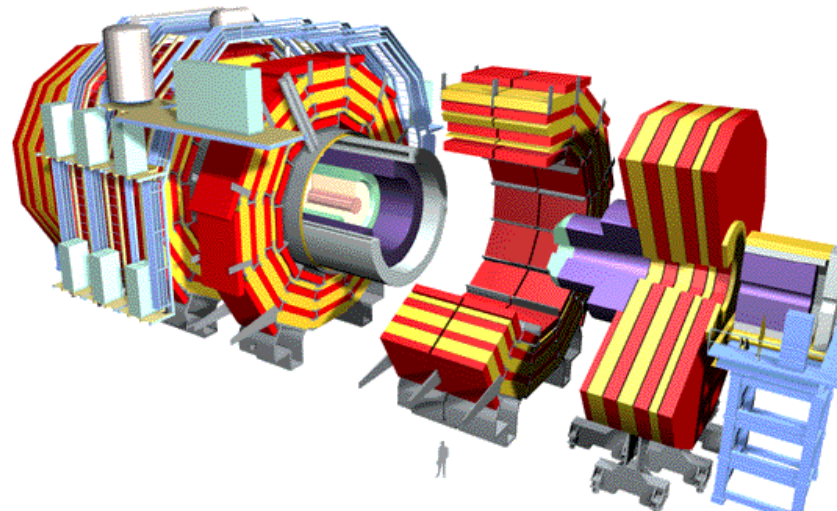
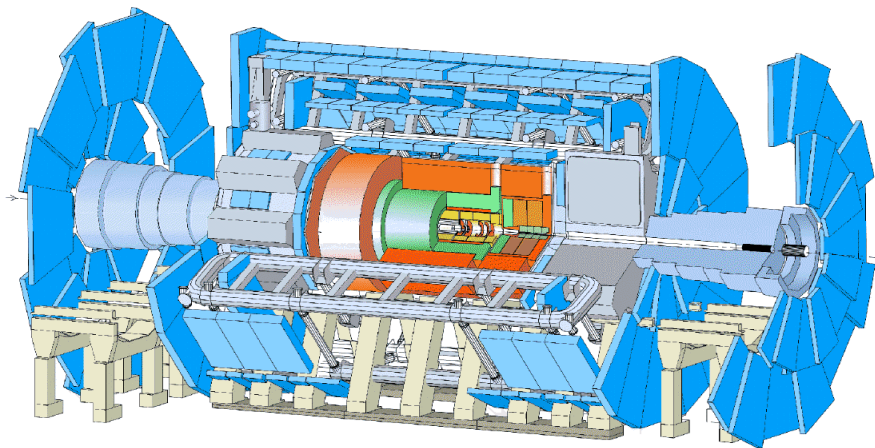


## Как стать пользователем Грид EGEE

- **Необходимо иметь account на UI локального кластера**
  - обращаться к администратору кластера
- **Получить цифровой сертификат. Для RDIG – Центр Сертификации (Certification Authority, CA) -**  
<http://ca.grid.kiae.ru/RDIG/>
  - потребуется контакт с лицом, ответственным за регистрацию в вашем институте – Registration Authority
- **Вступить в виртуальную организацию**
  - виртуальные организации LCG –  
[http://lcg-registrar.cern.ch/virtual\\_organization.html](http://lcg-registrar.cern.ch/virtual_organization.html)
  - виртуальные организации RDIG –  
<http://grid.sinp.msu.ru/grid/roc/voinrdig>



- Длина окружности ускорителя – 27 км.
- Запуск – в 2008 году







## ■ Требования по данным

- Колоссальные объемы данных (десятки и сотни Петабайт)
- Данные типа WORM (писать единожды, читать многократно)
- Структуризация данных с последующим извлечением информации из данных (data mining)
- Продолжительное время хранения данных, а также необходимость создания копий данных в разных странах мира

## ■ Требования к обработке данных

- Обработка данных подразделяется на 2 типа – регулярное производство данных и «нерегулярный» анализ данных
  - Производство (моделирование) данных происходит систематически; при этом производятся наборы данных порядка  $\sim 10^{**9}$  физических событий.
  - Анализ физических данных (на наборах данных порядка  $10^{**7}$  событий) проводится произвольным образом и в индивидуальном порядке многими сотнями отдельных пользователей
- Высокий уровень параллелизма обработки на уровне событий, который может быть описан ориентированным графом с указанием последовательности обработки
- Поскольку интерактивная работа очень важна при анализе данных, необходимо предусмотреть возможность спасения сессий с сохранением информации об источнике данных («проверяемость», provenance)
- Необходимость глобального доступа к базам данных экспериментов для получения значений констант, условий работы и т.д.



## Требования к компьютерингу для БАК

Необходимо обеспечить прозрачный доступ к данным и вычислительным ресурсам для ~5000 ученых в ~500 институтах, расположенных по всему миру

- Надежное и безопасное хранение данных (ежегодно будет производиться ~15 Петабайт данных)
- Скоростная сеть с малыми задержками и высокой пропускной способностью
- Управление разделением ресурсов между экспериментами, анализом и производством данных, различными группами анализа и индивидуальными пользователями, т.е. необходимость выработки общих правил (common policies)
- Поддержка и обучение пользователей

Проект **LCG** – the LHC Computing Grid Project (<http://lcg.web.cern.ch/LCG/>) – был организован для создания компьютерной инфраструктуры, необходимой для моделирования, обработки и анализа данных строящихся на LHC экспериментов.

Проект был принят ЦЕРН в 2001 году и включает в себя 2 этапа:  
1-й этап – 2002-05 (разработка общего прототипа мат.обеспечения и запуск пилотного вычислительного сервиса для LHC). 2-й этап – 2006-08 (оснащение и ввод в эксплуатацию вычислительного сервиса для LHC). В проект вовлечены эксперименты LHC, GRID-проекты в Европе и США, региональные и национальные компьютерные центры.

LCG-инфраструктура реализована и успешно функционирует на базе инфраструктур, обеспечиваемых проектами:

**EGEE** - Enabling Grids for E-Science

**OSG** - US Open Science Grid

**GridPP** - in the UK

**INFN Grid** - in Italy

**NorduGrid** - the Nordic region



*A map of the worldwide LCG infrastructure operated by EGEE and OSG.*

EGEE – это крупнейшая инфраструктура грид в Европе:

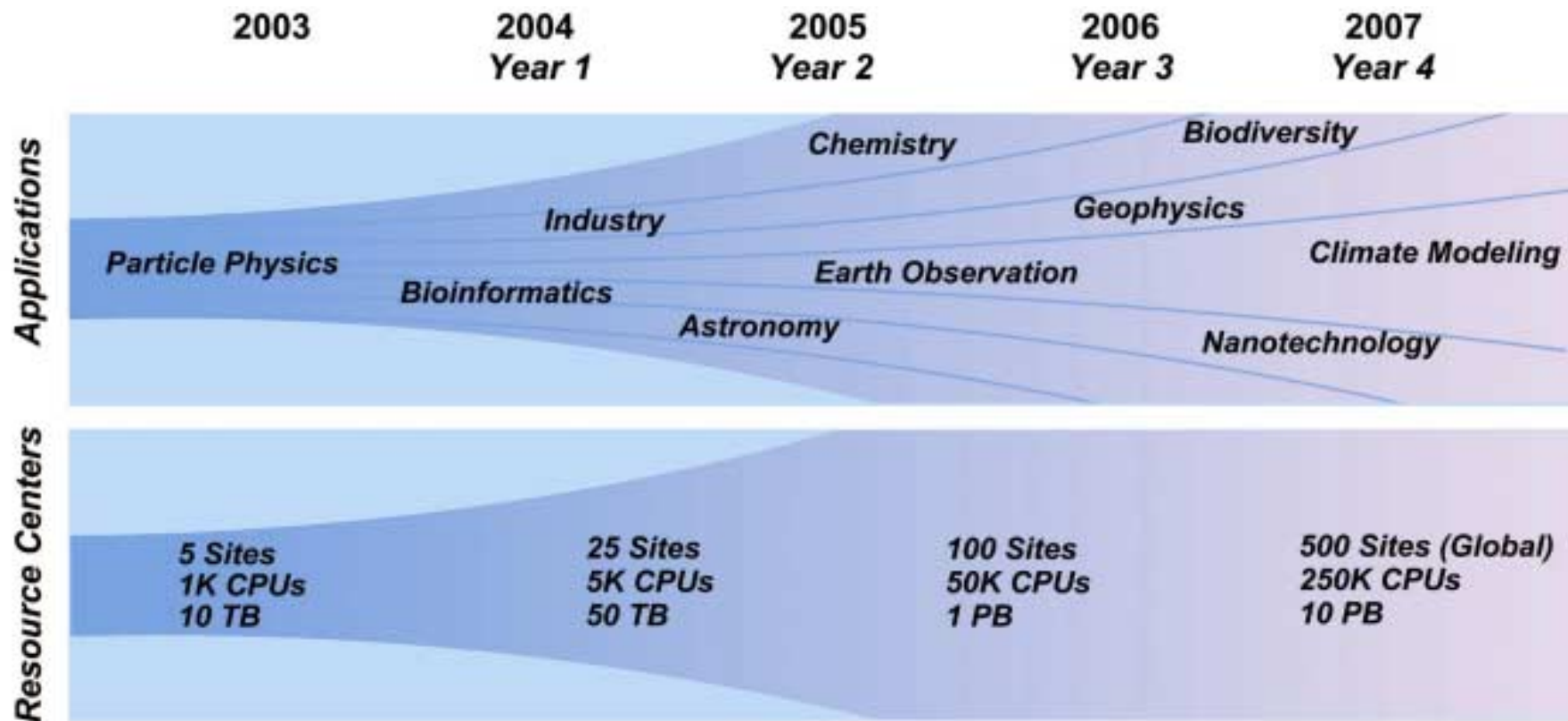
- 240 ведущих научно-исследовательских институтов 45 странах, объединенные в 12 региональных федераций грид
- расширяет национальные и региональные работы по грид
- Активизирует международное научное сотрудничество

Цель EGEE: создать глобальную инфраструктуру грид-сервисов, круглосуточно доступную для ученых и специалистов



LCG и EGEE – различные проекты  
Но именно тесное сотрудничество обеспечивает разделение, а не дублирование работ

Апрель 2004: проект стартовал

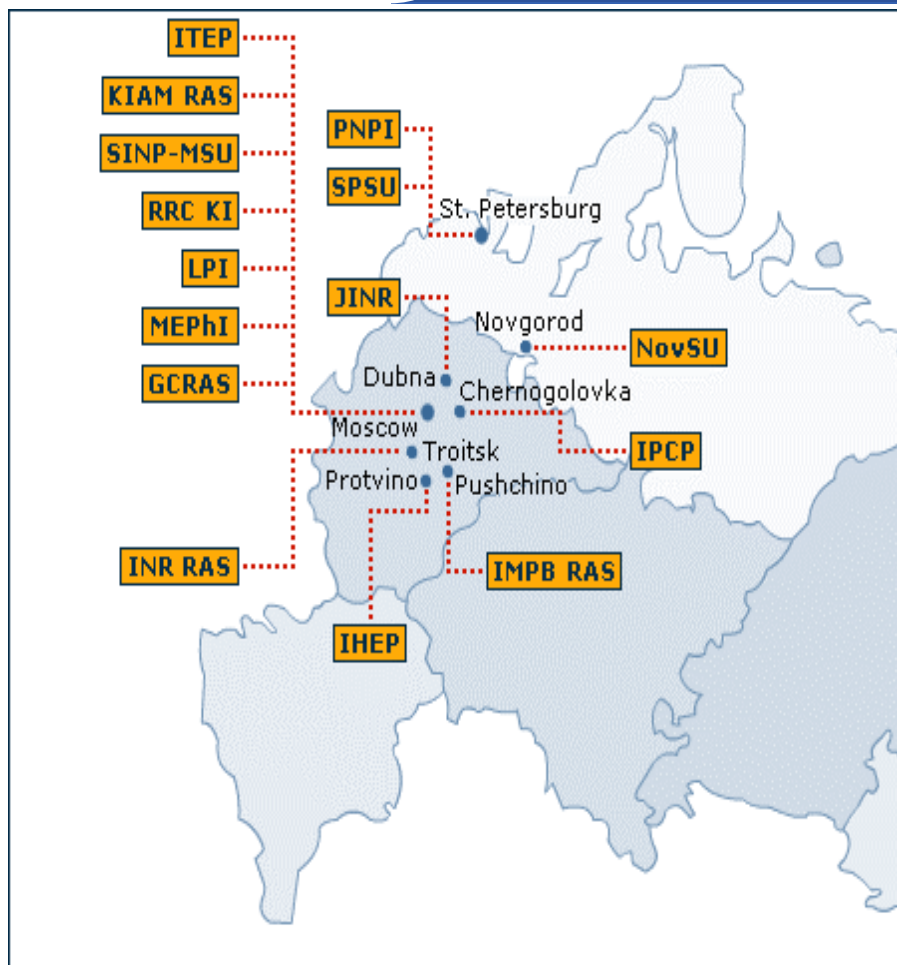


- для экспериментов на ускорителе LHC (ATLAS, CMS, ALICE, LHCb) - **LCG**
- биомедицина и биоинформатика
  - ✓ GATE - медицинская томография и планирование процедур радиотерапии
  - ✓ gPTM3D - Грид-ориентированные методы сбора, обработки и анализа трёхмерных радиологических изображений человеческих органов,
  - ✓ Mammogrid - проект, посвящённый созданию всеевропейской базы данных маммограмм
  - ✓ GPS@ : портал по биоинформатике, посвящённый различным средствам для анализа белков
  - ✓ CDSS - клиническая система поддержки принятия решений
  - ✓ моделирование взаимодействия потенциальных лекарств с белками-мишенями (проект WISDOM )



- **Науки о Земле:**
  - ✓ наблюдения за поверхностью Земли со спутников,
  - ✓ гидрология – проникновение морской воды в прибрежный водоносный слой,
  - ✓ климатология – прогнозирование наводнений,
  - ✓ физика Земли - численное моделирование землетрясений.
- Геофизика (промышленное приложение Geocluster)
- Астрофизика (проекты MAGIC, Planck, ANTARES, NEMO)
- Термоядерный синтез (проект ITER)
- Вычислительная химия (проекты CHARON, CompChem)
- Археология
- Финансовые приложения (проект EGRID)

(Всего свыше 200 виртуальных организаций)

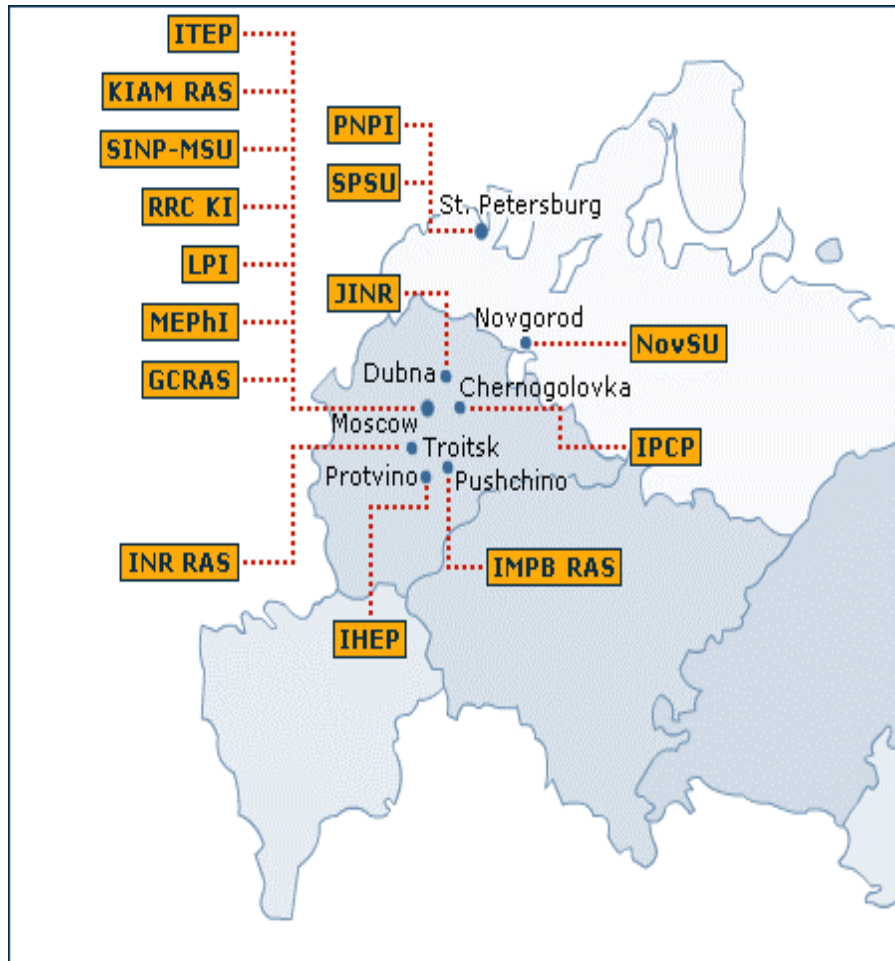


В целях обеспечения полномасштабного участия России в этом проекте был образован консорциум РДИГ (Российский ГРИД для интенсивных операций с данными – Russian Data Intensive GRID, RDIG) для эффективного выполнения работ по проекту и развитию в России инфраструктуры EGEE, с вовлечением на следующих этапах проекта других организаций из различных областей науки, образования и промышленности.

Консорциум РДИГ, согласно принятой в проекте EGEE структуре, входит в проект в качестве региональной федерации “Россия” (“Russia”).

Сейчас в РДИГ (<http://www.egee-rdig.ru>) входят 15 институтов (в 2004 году - 8)

Мониторинг: [http://rosmon.jinr.ru:8080/genimage?page=map\\_image](http://rosmon.jinr.ru:8080/genimage?page=map_image)



**Институт физики высоких энергий (Протвино)**

**Институт математических проблем биологии (Пушино)**

**Объединенный институт ядерных исследований (Дубна)**

**НИИ ядерной физики МГУ (Москва)**

**Институт теоретической и экспериментальной физики (Москва)**

**Институт прикладной математики им. М.В. Келдыша (Москва)**

**Петербургский институт ядерной физики (Санкт Петербург)**

**РНЦ «Курчатовский институт» (Москва)**

- 40 000 000 столкновений в секунду
- После фильтрации, остаётся только 100 событий в секунду
- Объём оцифрованных данных для одного события ~ 1Мб
- За год необходимо записать  $10^{10}$  событий = 15 Петабайт данных в год

*1 Мегабайт (1MB)*

*Цифровая фотография*

*1 Гигабайт (1GB) = 1000MB*

*DVD фильм*

*1 Терабайт (1TB) = 1000GB*

*Объём всех книг, изданных за год в мире*

*1 Петабайт (1PB) = 1000TB*

*Производит за год один эксперимент БАК*

*1 Экзобайт (1EB) = 1000 PB*

*Объём информации, которую производит за год всё человечество*

Объём получаемых данных БАК соответствует 20 миллионам записанных CD дисков в год.

Анализ данных LHC потребует вычислительных мощностей, эквивалентных мощности 100000 самых современных процессоров.

- **RAW** Сырые данные, как они записаны первоначально
- **SIM** Данные моделирования, аналогичные RAW
- **ESD** (Event Summary Data) Результат полной реконструкции сырых данных
- **AOD** (Analysis Object Data) Формат, содержащий характеристики восстановленных физических объектов и используемый для физического анализа
- **TAG** Краткие суммарные характеристики событий, используемые для их отбора
- **DPD** Упрощённые форматы, создаваемые исследовательскими группами для специфических выборок событий, используемых в анализах данных групп

Параметр	Единица измерения	Величина на одно событие
<b>RAW объём данных</b>	МБ	1.6
<b>ESD объём данных</b>	МБ	0.5
<b>AOD объём данных</b>	КБ	100
<b>TAG объём данных</b>	КБ	1
<b>SIM RAW объём данных</b>	МБ	2.0
<b>SIM ESD объём данных</b>	МБ	0.5
<b>Реконструкция (CPU)</b>	kSI2k*sec	15
<b>Моделирование (CPU)</b>	kSI2k*sec	100
<b>Анализ (CPU)</b>	kSI2k*sec	0.5

1 kSI2k = Pentium IV 2.5 GHz

Параметры приведены для данных, полученных при светимостях 2008, 2009 гг. При проектной светимости объём данных (RAW, ESD) возрастает на 50%, а процессорное время на реконструкцию (анализ) – на 75% из-за наложения дополнительных фоновых событий.

- **Tier-0** CERN.
- **Tier-1** – большой региональный компьютерный центр национального масштаба, имеющий все виды сервиса.
- **Tier-2** – менее крупный вычислительный центр.
- **Tier-3** – вычислительный ресурс институтского (университетского) масштаба (спутник для Tier-2/Tier-1).
- **Tier-4** – индивидуальный компьютер.





- **CERN** European Organization for Nuclear Research
- **TRIUMF** Canada's National Laboratory for Particle and Nuclear Physics
- **FNAL** Fermi National Accelerator Laboratory
- **BNL** Brookhaven National Laboratory
- **INFN** National Institute for Nuclear Physic
- **RAL** Rutherford Appleton Laboratory
- **NIKHEF/ SARA** National institute for subatomic physics
- **DESY** Deutsches Elektronen Synchrotron
- **NDGF** Nordic Data Grid Facility



## Функции Tier0 Tier1 центров ,

### Основные задачи Tier-0

- Хранение одной копии RAW данных на лентах
- Копирование в режиме реального времени всех сырых данных в Tier-1's
- Хранение на дисках калибровочных данных
- Быстрое первоначальное определение параметров калибровок/выравнивания и первоначальная реконструкция событий (создание ESD)
- Распределение ESD во внешние Tier-1's

### Основные задачи Tier-1

- АТЛАС планирует использовать 10 Tier-1's
- Хранение 1/10 RAW данных на лентах
- Обеспечение повторной реконструкции своей доли данных и доступа к реконструированным ESD, AOD, TAG
- Хранение 20% последних версий ESD и AOD на дисках и 10% предыдущей версии на лентах
- Обеспечение организованного доступа групп физического анализа к ESD
- Осуществление поддержки 3-4 прикрепленных Tier-2's
- Хранение SIM данных, произведенных в прикрепленных Tier-2's
- Повторная реконструкция осуществляется примерно 2 раза в год

### Основные задачи Tier-2

- АТЛАС планирует использовать 30 Tier-2's (по 20 пользователей)
- Возможно использование распределённого Tier-2
- Хранение 1/4 AOD и всех TAG данных на дисках
- Хранение на дисках небольшой выборки ESD
- Обеспечение возможности (CPU, дисковое пространство) физического анализа и MC симуляции данных для отдельных пользователей (20/Tier-2)
- Обеспечение централизованной MC симуляции данных
- Распределённый анализ физических данных также осуществляется с использованием Grid middleware
- Возможна обработка калибровочных данных

- Институт теоретической и экспериментальной физики (Москва) (отв.по **LHCb**)
- Институт физики высоких энергий (Протвино) (отв.по **ATLAS**)
- Курчатовский институт (Москва)
- Научно-исследовательский институт ядерной физики (Москва) (отв. по **CMS**)
- С.-Петербургский институт ядерной физики (Гатчина)
- Объединенный институт ядерных исследований (Дубна) (отв. по **ALICE** и **CMS**)



## Системы хранения и передачи данных в LCG

Castor ([www.castor.org](http://www.castor.org)), Tier0-Tier1

dCache ([www.dcache.org](http://www.dcache.org)) Tier1-Tier2

DPM ([www.gridpp.ac.uk/wiki/Disk\\_Pool\\_Manager](http://www.gridpp.ac.uk/wiki/Disk_Pool_Manager)) Tier2-Tier3

Для взаимодействия с которыми разработан специальный сервис SRM (Storage Resource Manager) ([www.gridpp.ac.uk/wiki/SRM](http://www.gridpp.ac.uk/wiki/SRM))

Для обеспечения необходимой надежности, производительности и организации взаимодействия между остальными сервисами управления данными был создан сервис передачи данных — FTS (File Transfer Service), основные обязанности которого: обеспечение надежных и удобных механизмов передачи файлов типа «точка-точка», контроль и мониторинг передач, распределение ресурсов сайта между различными организациями, управление запросами пользователей.

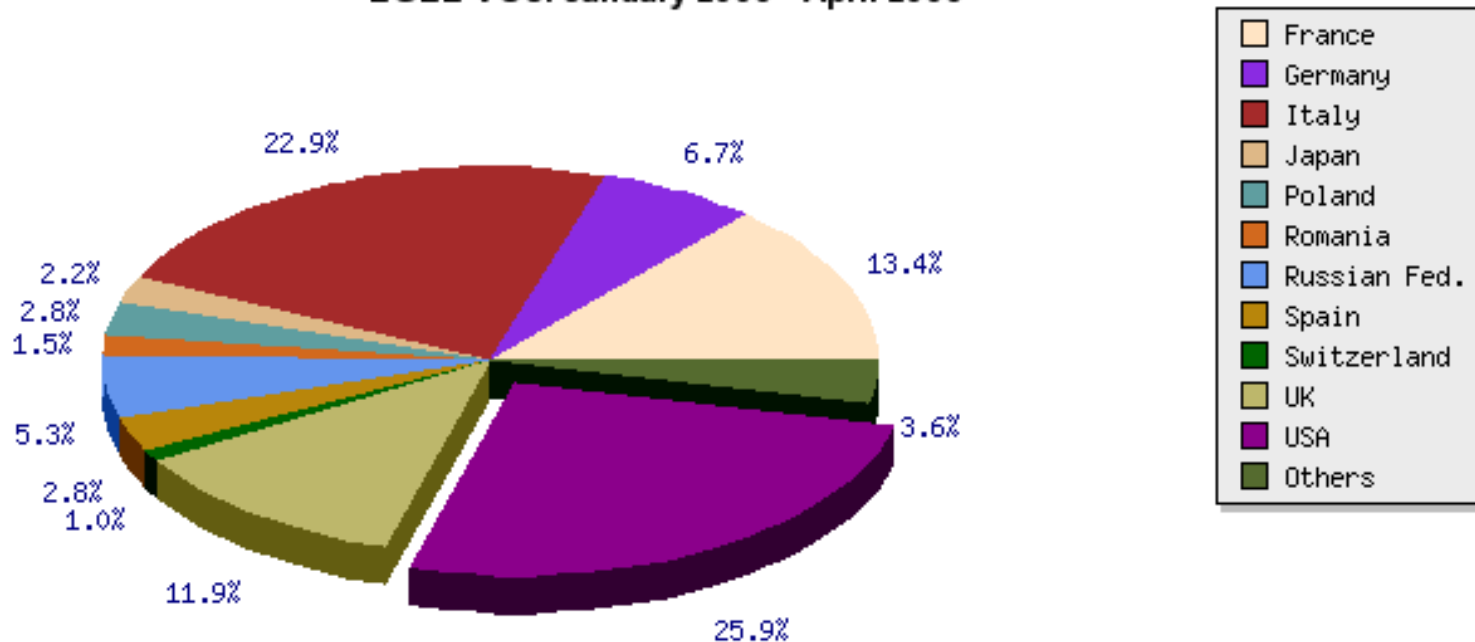
PhEDEx – Physics Experimental Data Export Service (CMS)

Xrootd - AliEn ALICE Environment

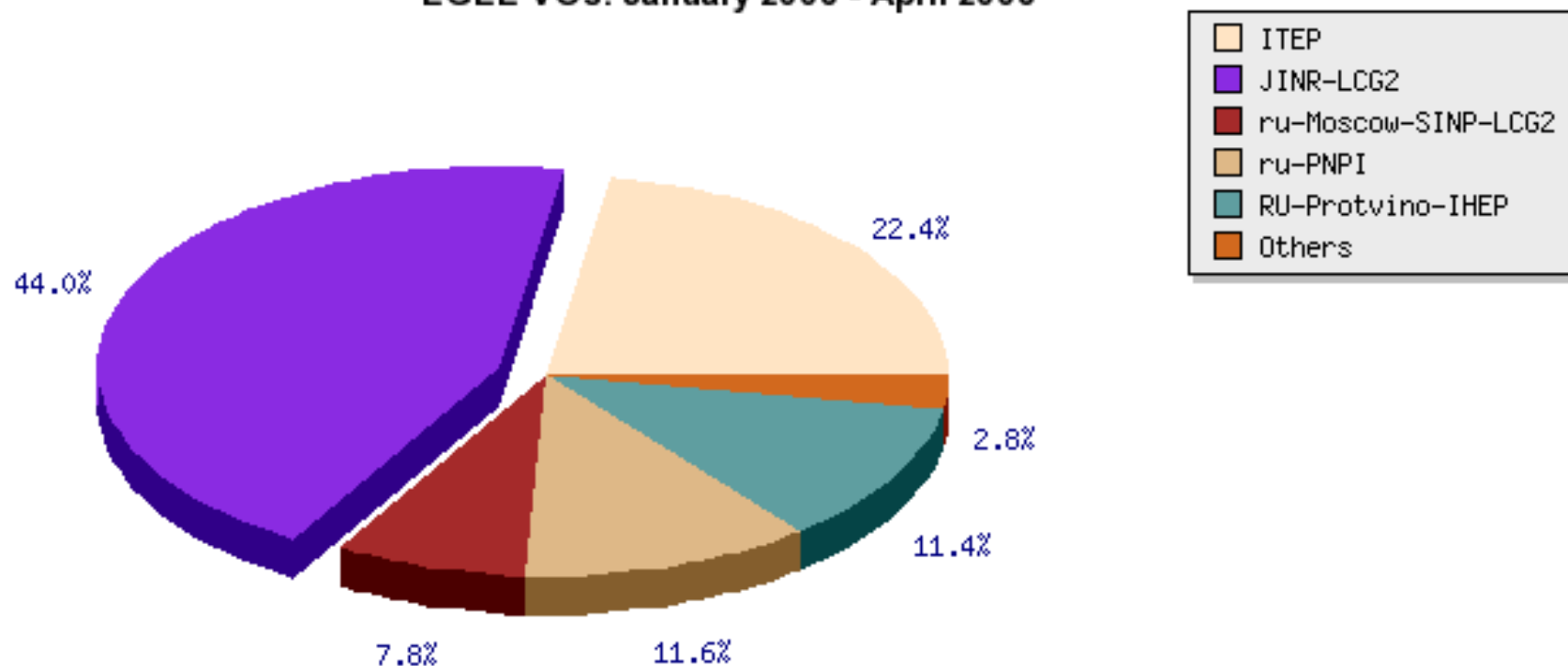
- ATLAS: The PanDA production and distributed analysis system
- CMS: CRAB -CMS Remote Analysis Builder
- LHCb: DIRAC - Distributed Infrastructure with Remote Agent Control
- ALICE: AliEn - ALICE Environment

- 1.Упрощает пользователям запуск заданий пользователей, а также production job's;
- 2.Скрывает механизмы GRID от пользователя;
- 3.Обеспечивает специфические нужды экспериментов;
- 4.Обеспечивает возможность организации гибких политик запуска заданий внутри самого эксперимента;
- 5.Имеет свою систему мониторинга и аккаунтинга;
- 6.Интеграция в грид инфраструктуру происходит на основе pilot заданий

COUNTRY\_T2 Normalised CPU time per COUNTRY\_T2  
EGEE VOs. January 2008 - April 2008



**Russia Normalised CPU time per SITE**  
EGEE VOs. January 2008 - April 2008





- Вычислительные мощности:**

	CPU type	Number of nodes	Core per node	RAM per node	performace (kSI2k)
1	Pentium 4 2,4Ghz	38	2	2GB	40
2	Pentium 4 2,8Ghz	8	2	2GB	10
3	HP ProLiant DL140G3, 2x Dual-Core Xeon 5150, 2.66GHz	20	4	8GB	200
<b>Total:</b>		<b>64</b>	<b>172CPU</b>		<b>250</b>

- Системы хранения данных:**

	Storage system	Number of systems	Size per system, TB	Total size, TB
1	Gladius storage server	10	4	40
2	EonStor RAID system	1	2,6	2,6
3	UltraTrack SX-8000 RAID system	3	0,8	2,4
<b>Total:</b>				<b>45</b>



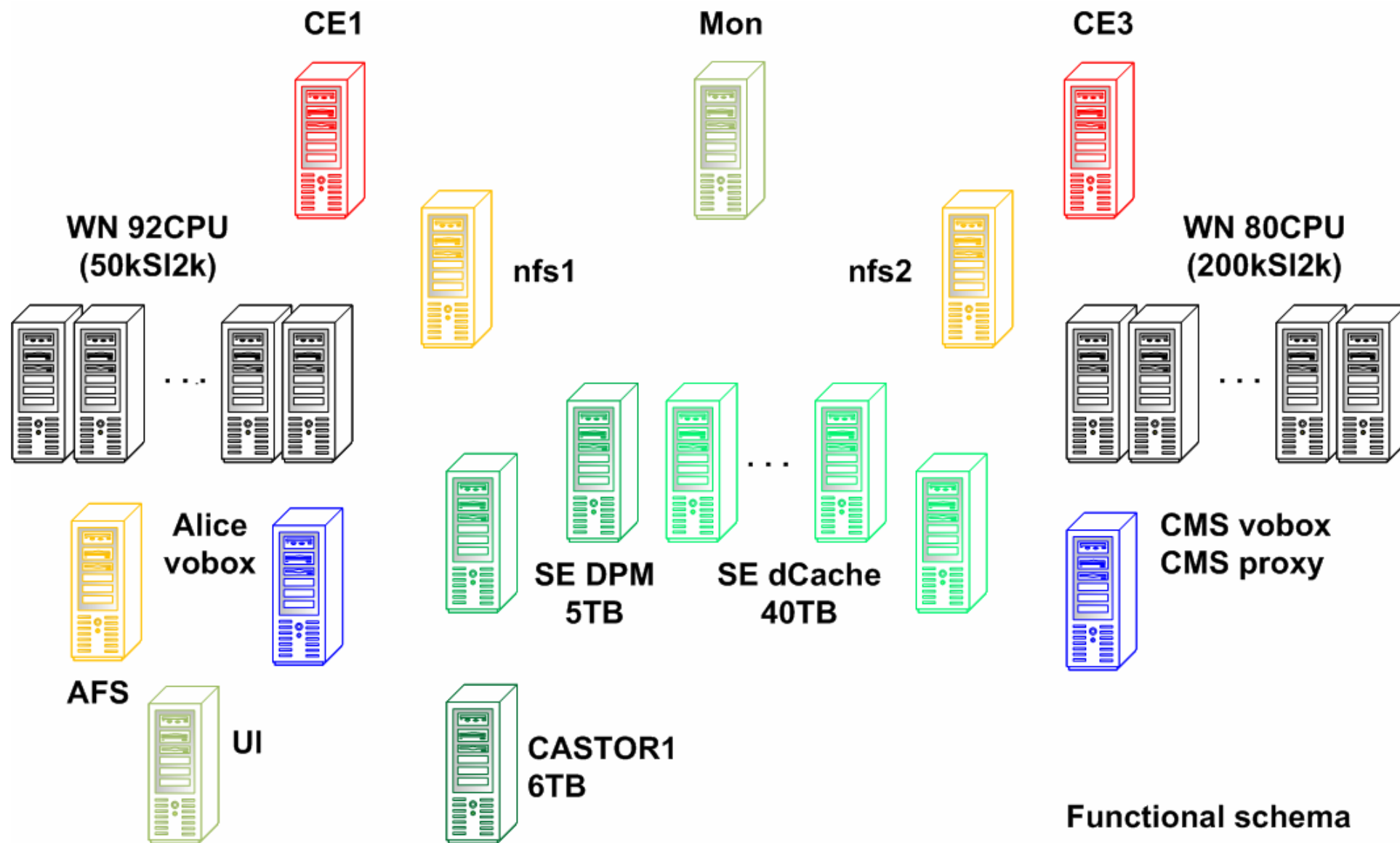
## ИФВЭ в инфраструктуре Грид





# ИФВЭ в инфраструктуре Грид



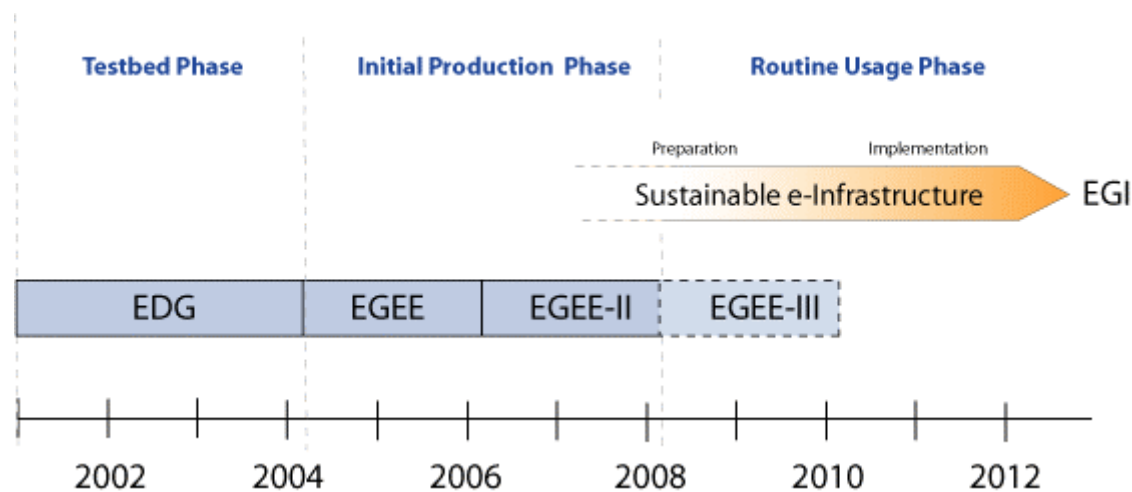


Functional schema

# eGEE Национальные Грид инфраструктуры

- Англия UK National Grid Service <http://www.ngs.ac.uk>
- Австрия Austrian Grid Initiative <http://www.austriangrid.at>
- Бельгия BEgrid <http://www.begrid.be/>
- Болгария Bulgarian Grid Consortium, BGGC, <http://www.grid.bas.bg/consortium.htm>
- Хорватия CRO-GRID <http://www.srce.hr/crogrid/>
- Чехия METACentrum <http://meta.cesnet.cz>
- Эстония Estonian Grid, working under Estonian Educational and Research Network (EENet) <http://grid.eenet.ee>
- Финляндия CSC – Scientific Computing Ltd., <http://www.csc.fi>
- Франция EGEE-FranceGermanyD-Grid, <https://www.d-grid.de/index.php?id=1&L=1>
- Греция HellasGrid, HG, <http://www.hellasgrid.gr>
- Ирландия Grid-Ireland <http://www.grid.ie/>
- Израиль Israel Academic Grid (IAG) <http://iag.iucc.ac.il>
- Италия Italian Grid Infrastructure, IGI, <http://www.italiangrid.org>
- Литва Lithuanian Grid LitGrid <http://www.litgrid.lt>
- Голландия Netherlands National Science Grid, internally known as BIGGRID project
- Норвегия Norwegian Grid (NorGrid), <http://www.norgrid.no>
- Польша PL-Grid, <http://plgrid.cyfronet.pl> (public part in preparation)
- Португалия Iniciativa Nacional Grid – INGRID <http://www.gridcomputing.pt/>
- Румыния RoGrid URL: <http://www.rogrid.ro> (old version)
- **Россия Russian Data Intensive Grid (RDIG): <http://www.egee-rdig.ru>**
- Сербия Serbia Academic and Educational Grid Initiative of Serbia AEGIS <http://aegis.phy.bg.ac.yu>
- Украина Ukrainian Grid, UGrid <http://grid.ntu-kpi.kiev.ua/>
- Швеция Swedish Grid, SweGrid, <http://www.swegrid.se/> and <http://www.snic.vr.se/>
- Турция TR-Grid National Grid Initiative, TR-Grid NGI <http://www.grid.org.tr>
- Германия(D-initiative)

- Необходимость подготовки постоянной, общей Грид инфраструктуры
- Поддержка долговременной Европейской e-инфраструктуры, независимо от выполняемых краткосрочных (1-2 года) проектов
- Координация интеграции и взаимодействия между Национальными Грид Инфраструктурами (NGI)
- Управление Европейским уровнем производственной Грид инфраструктуры для широкого круга научных дисциплин совместно с NGIs.





## Полезные ссылки

<http://www.eu-egee.org> - проект EGEE-II

<http://egee.lib.ed.ac.uk/> - репозиторий материалов (EGEE Digital Library)

<http://www.egee-rdig.ru/> - RDIG консорциум

<http://roston.jinr.ru:8080/accounting/> - мониторинг RDIG

<http://www.gridclub.ru/> - GRIDCLUB.RU :: Интернет-портал по грид-технологиям

<http://egee.sinp.msu.ru> - НИИЯФ МГУ

<http://gridcafe.web.cern.ch>

<http://EnterTheGrid.com>



*Спасибо за внимание*

