



Enabling Grids for E-scienceE

gLite Application Developers Course

*Institute for Parallel Processing of the
Bulgarian Academy of Sciences (IPP-BAS),
Sofia, Bulgaria*

25 November 2008

www.eu-egee.org



Information Society





Enabling Grids for E-science

Introduction and Application Development Aspects of gLite 3.x

D.Todorov, IPP-BAS, Sofia, Bulgaria

*Materials are used from the presentations of
Mike Mineter, Training Outreach and Education, University
of Edinburgh, UK*

www.eu-egee.org



Information Society



- **Проектът EGEE (Enabling Grids for E-science)**
 - В момента обхваща около 300 сайта в 50 страни и дава на своите 10 000 потребители денонощен достъп до 80 000 процесорни ядра
 - Започва като 4-годишен проект с две фази – на 30 Април 2008 завърши втората 2-годишна фаза
 - В момента проектът продължава със трета си 2-годишна фаза (EGEE-3)
 - Финансиран е от Европейската Комисия
- **Ресурсите работещи в рамките на проекта EGEE променят начина за провеждане на научните изследвания**
- **Използването на тези ресурси зависи от потребители с приложения, които се нуждаят от:**
 - Запомняща среда с голям обем
 - Висока скорост на информационния обмен
 - Висока изчислителна мощност

1. Въвеждащите курсове в проекта

- Привличат в проекта нови участници или потребители
- Запознават участниците с:
 - Как да станат регистрирани потребители на EGEE Грид инфраструктурата
 - Първоначални познания и примери за ползването на Грид инфраструктурата
- Подготвят за участие в курс за разработчици на Грид приложения

2. Курсовете за разработчици на Грид приложения

- Предназначени са за специалисти, външни или вътрешни на проекта EGEE, имащи намерение да разработват нови приложения, които:
 - са предвидени за изпълнение в EGEE Грид инфраструктура
 - използват ресурси, участващи в проекта EGEE
 - ще бъдат на разположение на други EGEE потребители
- Обучението в такъв курс
 - предполага програмистки опит в конкретна приложна област
 - не изисква предшествващ опит в ползването на Grid или Web услуги
- Курсът за разработчици на приложения помага за
 - разбиране на оперативния контекст на платформата на EGEE
 - Запознаване с най-често ползваните функции на тази платформа и техните интерфейси към приложните програми (API)
 - Придобиване на представа за
 - *ограниченията към приложенията*
 - *ограниченията към потребителите*
 - *инструментите, които са на разположение за наблюдение и настройка на приложенията*

Introduction to Grid Computing, EGEE and Bulgarian Grid Initiatives

(Plovdiv, 4.12.2006, http://egee-2-na3.acad.bg/Events/BG_Induction-2-Plovdiv.htm)

- Introduction
 - An Introduction to Grid Computing and the EGEE Project
 - Bulgarian Grid Initiatives
- Practical - Using EGEE Middleware
 - Authentication, Authorisation and Security
 - Using a certificate and simple job submission
 - Porting applications to the GILDA grid
 - Information systems on EGEE
 - Data management
 - Next Steps
 - Closing discussion

- **Цели**

- Да задълбочи познанията по Грид на работещите в рамките на проекта EGEE, по отношение на разработката на нови приложения
- Да даде поглед от високо ниво на процеса на разработка на приложения
- Да задълбочи познанията в тази област чрез практически занимания

- **Изисквания към участниците**

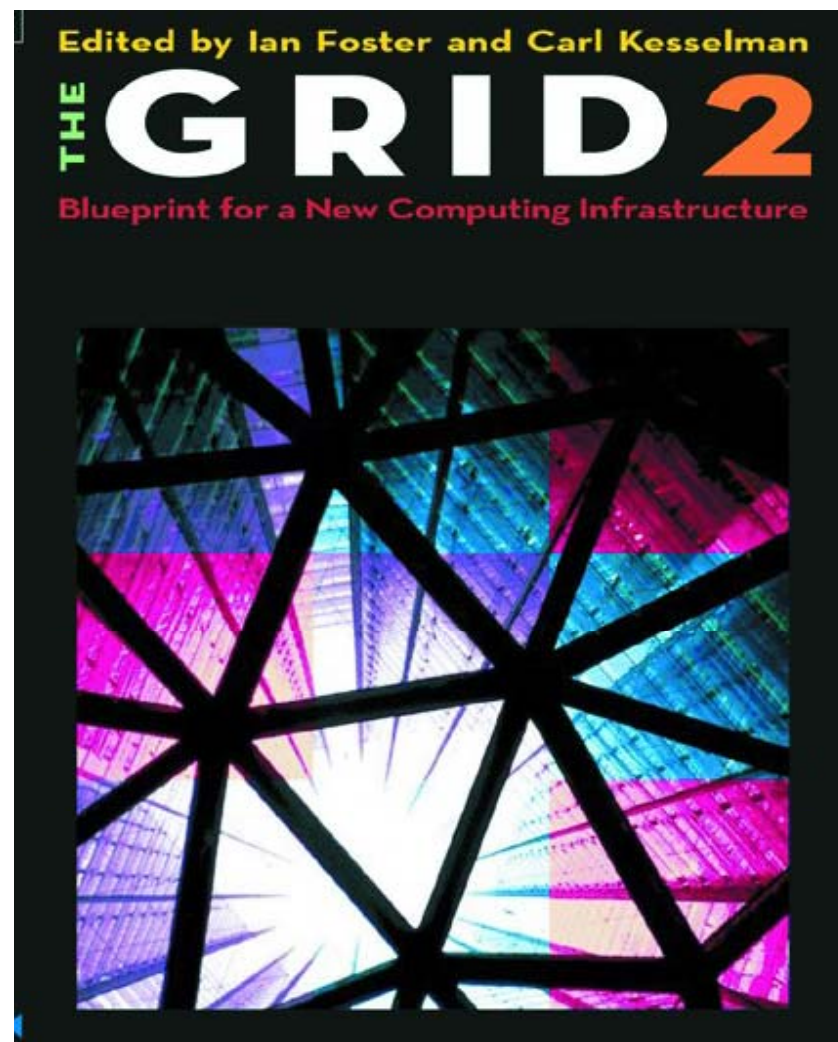
- Да имат базови познания за Грид
- Да имат интереси да разработват нови Грид приложения или да модифицират и/или ползват съществуващи такива
- Предполага се, че участниците в някой от проведените вече въвеждащи курсове отговарят на тези изисквания

- Преглед на концепциите за Грид и Грид приложения
- Видове Грид приложения
- Предизвикателства и проблеми при създаването на Грид приложения
- Повече за услугите в gLite
- Преглед на останалата част на курса

- Благодарение на споделянето на ресурси (данни, инструменти, изчисления, човешки опит ...) през интернет става възможно да се постигне сътрудничество в научната дейност, което се характеризира с
 - Размиване на границите между организациите
 - Много интензивни изчисления
 - Много големи обеми от данни
 - Много големи мащаби - многостранно и многообразно
- **e-Наука (e-Science)** – провеждането на научна дейност с висока интензивност на изчисленията, която се изпълнява в разпределена мрежова изчислителна среда и която често е свързана с огромни информационни масиви и ползване на Грид
- Отнасянето на технологиите на **e-Науката** към нови научни общности (социални науки, изкуства, хуманитарни науки и т.н.) доведе до възникване на по-универсалния термин **e-Изследвания (e-Research)**

- Грид изчисленията (Grid computing) са частен случай на разпределени изчисления, който има следните допълнителни значения:
 - Множество независими компютърни клъстери, които формират единен ресурс от възли – т.е. работят като мрежа (Грид) и не са разположени в един административен домейн (формално определение).
 - Предлагане на **online** изчисления или масова памет под формата на измерима търговска услуга (известна като **utility computing, computing on demand**, или **cloud computing**)
 - Създаване на “виртуален суперкомпютър” чрез
 - използване на свободните компютърни ресурси в рамките на дадена организация или чрез
 - използване на мрежа от географски разпръснати компютри

- Основните характеристики, на които трябва да отговарят компютърни ресурси с Грид организация, според Ian Foster са:
 - Да интегрират и координират ресурси, които не са обект на централизирано управление
 - Да използват стандартни, отворени протоколи и интерфейси с общо предназначение
 - Да предоставят услуги с гарантирано не тривиално качество



- От гледна точка на потребителите Грид технологията дава възможност за унифицирано използване на ресурсите чрез стандартизиран мидълуер
- Потребителят се абстрахира от разположението, управлението и собствеността върху ресурса, което му дава възможността да:
 - ползва ресурси, които не притежава
 - предоставя ресурси, които в даден момент не ползва
- Грид технологията е много удобна в случаите, когато става дума за осигуряване на компютърни ресурси при решаването на задачи с голям обем изчисления.

- **Грид** са компютърни ресурси, които се предоставят при поискване през Интернет посредством стандартизирани процедури и при гарантирана висока надеждност.
 - Акцентът е наличието на глобален унифициран достъп до желана от потребителя част от тези ресурси през мрежова среда.
 - Потребителят получава компютърни ресурси имащи желани от него възможности (хардуер и софтуер), които ползва за ограничено време.
- При **Cloud Computing** се акцентира върху получаването на компютърни услуги, които се предоставят при поискване през Интернет посредством стандартизирани процедури и при гарантирана висока надеждност.
 - Потребителят не се интересува от компютърния ресурс, който стои зад тези услуги.
 - В частност компютърната услуга може да е под формата на компютърна среда, която потребителят организира и ползва по желан от него начин
- Може да се приеме, че **Cloud Computing** е еволюция на концепцията за ползване на обществени компютърни услуги, в която междинен етап са **Грид**.
- От гледна точка на техническата реализация **Cloud Computing** е потребителско ниво на достъп до разпределени компютърни ресурси, което е:
 - над нивото на услугите предоставяни от компютърните ресурси;
 - фокусирано върху това какво потребителя иска да прави, а не върху начина по който то се обезпечава.

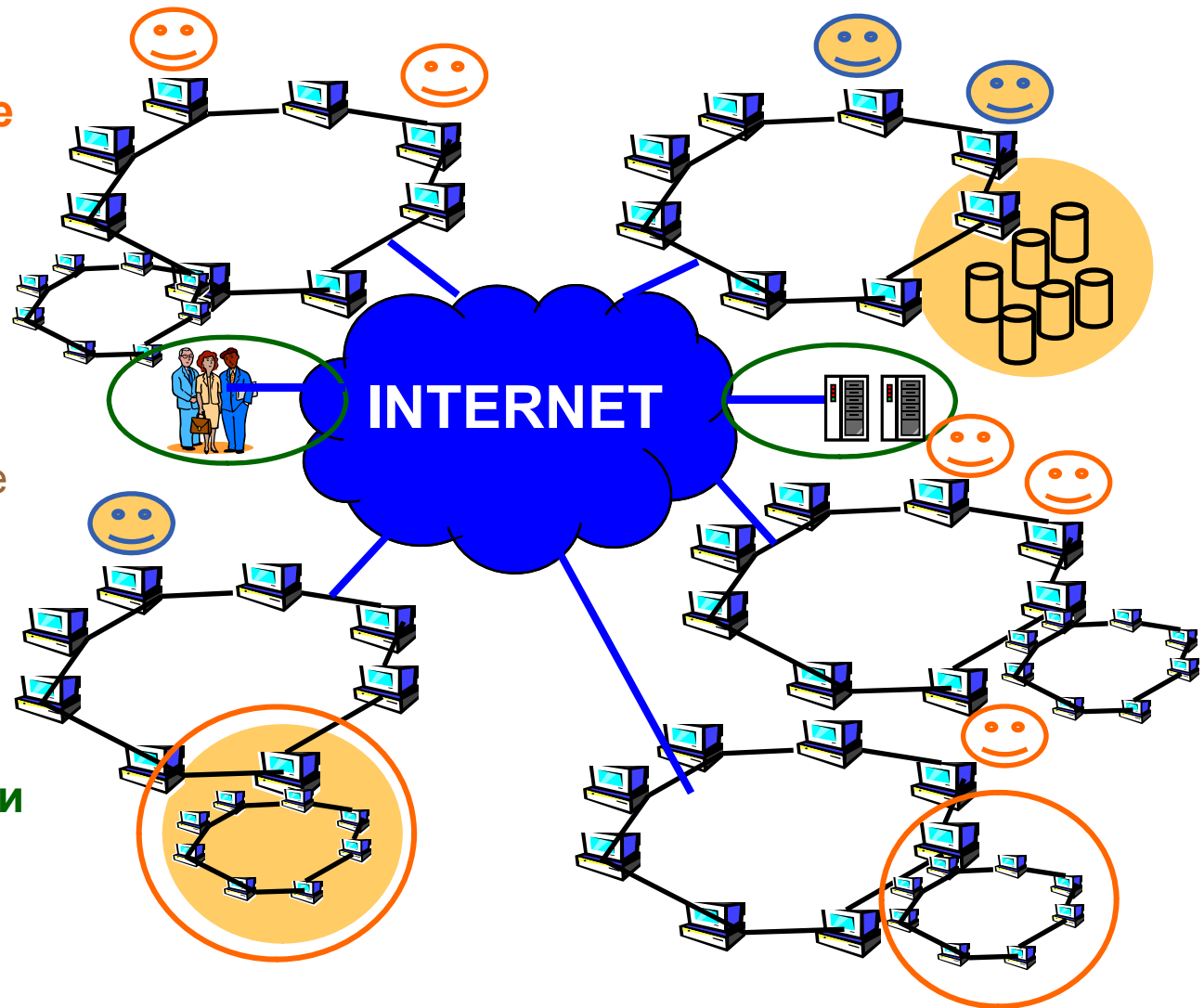
- Става възможен цялостен подход
- Съвместни изследвания, реализирани на инженерни проекти, обществени услуги



Диаграма на базата на слайд на Ян Фостер

- **Какво е Виртуална Организация (ВО)?**
 - Хора от различни организации, които желаят да си сътрудничат и да споделят своите ресурси през техните организационни граници
 - Например всяко научно сътрудничество е форма на виртуална организация
- **Грид инфраструктурата позволява на членовете на Виртуалните организации да**
 - Предоставят свои ресурси за ползване от ВО
 - Придобиват ресурси, които да се ползват от ВО
 - Организируют общ достъп до предоставените или придобитите ресурси
- **Всеки ресурс е на разположение на грида чрез абстракция, която маскира хетерогенността, например като**
 - Различни хетерогенни изчислителни платформи
 - Различни ресурси данни
- **Ресурсите в проекта EGEE се притежават от ВО**
- **ВО могат да предоставят свои ресурси или да ползват такива на друга ВО на базата на договор**

- **Виртуални организации се договарят със сайтове за достъп до ресурси**
- **На всеки общ ресурс се изпълнява Грид мидълуер, който осигурява**
 - Услуги за управление на данните
 - Изчислителни услуги
 - Управление на достъпа
- **Разпределените услуги (хората и мидълуер) правят работата на грида**





Това е нивото, където

- компютърните науки срещат приложните
- виртуалните организации разработват и използват за своите приложения
 - приложни инструменти от високо ниво (портали, виртуална среда за изследванията, организация на работата, регистри и други услуги)
 - базови услуги (core services)
- се осъществява достъпност на Грид услугите за неспециалисти

Гридовете работещи в производствен режим трябва да осигуряват тези услуги.


- **Грид приложение - Софтуер, който взаимодейства с грид услугите за да постигне специфични цели на виртуална организация или отделен потребител**
- **Грид приложенията ползват съществуващите ГРИД услуги (например тези в gLite 3.x)**
- **Грид приложенията отговарят на цели и изисквания, които ги правят специфични за даден потребител или виртуална организация**

1. Прости приложения – стартират се през WMS (Workload Management System) и се изпълняват в пакетен режим (batch mode)
2. Приложения включващи Грид услуги като
 - Запис и четене на файлове върху и от SE (Storage Element)
 - Осигуряване на информация за състоянието на Грид средата и за изпълнението на приложението (Monitoring)
 - Осигуряване на връзка с външния свят (interactive jobs)
 - Управление на мета данни
 - ...
3. Сложни приложения
 - Среда, която управлява множество приложения в зависимост от желанието на потребителите
 - Услуги от високо ниво
 - Портали и средства за организация на работата
 - Софтуер, създаден специално за потребителя или виртуалната организация
 - ...

- **Чрез UI** (User Interface е машина, която изпълнява функцията на ‘gateway’ към Грид ресурсите – чрез нея потребителя се идентифицира и получава достъп до EGEE Грид ресурсите)
 - Текстови команди или скриптове
 - APIs
 - Инструменти от високо ниво
- **Чрез Windows приложение**
 - Ползване на Грид без да разбираме за неговото съществуване
 - gLite все още не поддържа Windows
- **Чрез портали**
 - За често и лесно стартиране на задачи, които могат да бъдат:
 - Базови Грид услуги (core grid services)
 - Цели приложения или приложни инструменти
 - Достъпност от всеки browser
 - Създадени са според изискванията на даденото приложение
 - В проекта EGEE такива портали са
 - P-GRADE (<http://portal.p-grade.hu/>) и
 - GENIUS (<https://genius.ct.infn.it/>)

- **Създаване на приложение за удовлетворяване на нужда от ресурси за наше (мое) изследване**
 - Нужда от по-богата функционалност
 - MPI, обхват на входните параметри,...
 - Услуги за специфична обработка на данни и/или изчисления
- **Предоставяне на мое приложение за ползване от други специалисти**
 - Начин на ползване - как!?
 - Предварително инсталиране на изпълними кодове?
 - Средата, която ще го изпълнява?
 - Споделяне на данни
 - Ползване през портал?
- **Предоставяне на наше приложение за ползване от други специалисти**
 - Става допълнително необходимо:
 - Координиране на разработката
 - Спазване на стандарти
 - ...

**Увеличаване на сложността и
разходите за разработка**

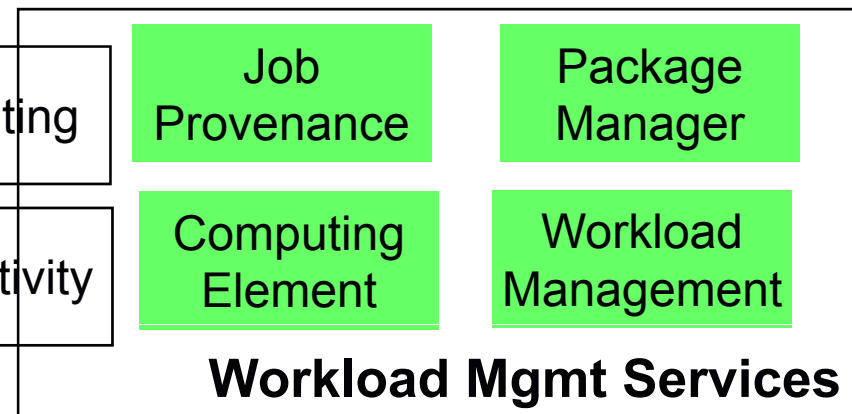
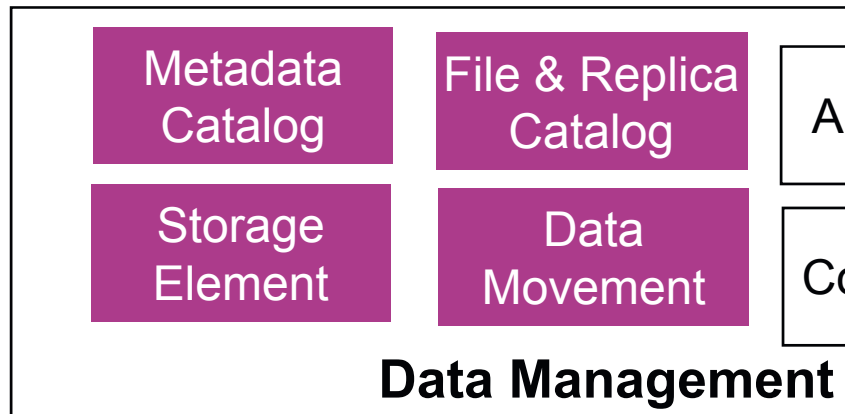
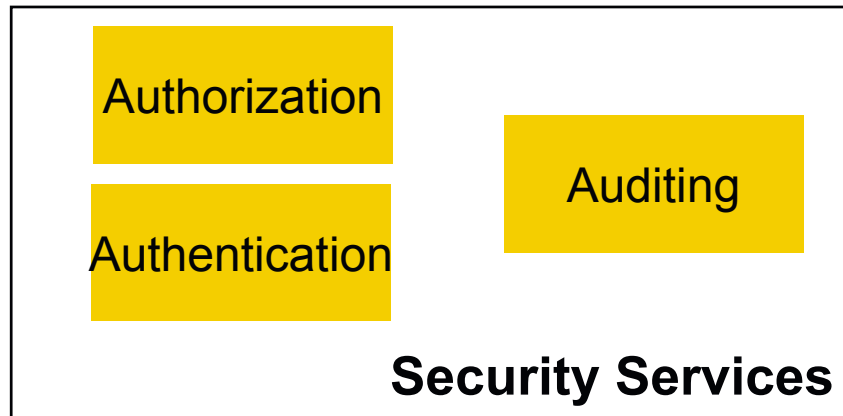
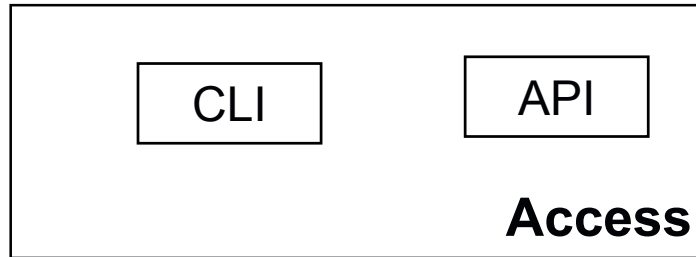


- **Изследователският софтуер е често**
 - Създаден за един потребител – самия разработчик
 - Доброто познаване го прави използваем
 - Решава близки цели: например ползва се докато се напише статията и след това се изхвърля
- **Приложенията в Грид често се използват**
 - От ВО
 - Без поддръжка от разработчика
 - В нов контекст и организация на работата
- **Разработчиците на Грид приложения са**
 - Работещи в изследователска среда
 - Техният софтуер се нуждае от още:
 - Стабилност
 - Документация
 - Възможност за ползване
 - Възможност за разширение
 - Т.е. Качество на продукцията

Необходим е опит в:

- **Софтуерен инженеринг**
- **приложната област**
- **Грид изчисленията**

- **Екипна работа!**
- **Участие в международно сътрудничество – многократно ползване, не е нужно да се прави нещо съществуващо отново, интердисциплинарен подход при намиране на решение**
- **Преходът от изследователски софтуер към работещ софтуерен продукт: около 5 пъти по-голям разход на труд**
 - 80% от времето за последните 10% от функционалността и надеждността
- **Ключовото решение е стандартизация**
 - За многократно ползване, за динамично конфигуриране на услугите, ...
 - Валидно е и за мидълуера и за специфичните приложни продукти
- **Необходимо е да се следва добре обмислен процес на разработка**
 - Изисквания за разработка, изпълнение и внедряване на приложението
 - Скорост на изработка на прототипа?
 - Поддръжка на потребителите



- **В курса фокусът ще бъде върху:**
 - Нови функционални възможности в gLite 3.x Workload Management
 - Достъп до данните върху SEs (Storage Elements)
 - Можем ли да ползваме огромни файлове, които са прекалено големи за да бъдат копирани?
 - Как да осъществяваме достъп до такива файлове?
 - Управление на метаданни
 - Можем ли да ползваме огромен брой файлове?
 - Нужда от достъп и многократно ползване, основаващо се на характеристиките на файловете, а не толкова на техните имена
 - Осигуряване на информация за начина на изпълнение на приложенията в Грид среда (Monitoring)
 - Изпълнение на множество дълги задачи
 - Какво се случва с изпълнението на задачите?!