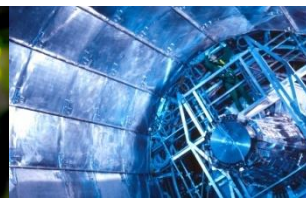
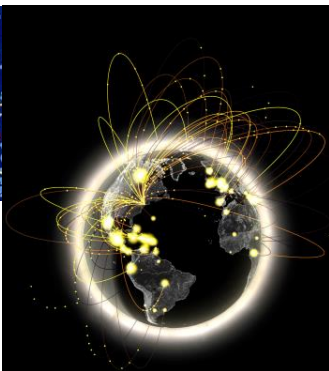
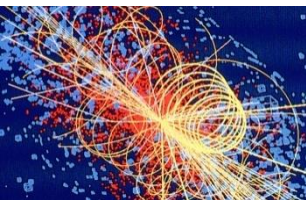




GRID

ТЕХНОЛОГИИ

И ТЯХНОТО ИЗПОЛЗВАНЕ В ЦЕРН



Иван Глушков
СУ "Климент Охридски"
Програма за обучение на български инженерни учители
ЦЕРН, октомври 2014



Накратко..



- Данните от LHC
 - Колко точно са данните от LHC?
 - Що е то тригер?
- Информационни технологии в ЦЕРН
- Какво се разбира под GRID?
- GRID за LHC или какво е WLCG
- Малък бонус..



Данните



.. и колко от тях можем да запишем

- Колко данни създаваме? (само CMS)
 - 1 MB / събитие
 - 40 MHz честота на сблъскване
 - 40 TB/ секунда!! (10^6 PB / година)
- Колко данни записваме? (LHC)
 - 30 PB/година
 - Аналог :
 - 9 милиона HD филма
 - Записани на CD-та образуват кула с височина 20 км
 - Повече от половината от всичко написано от човешвото на всички езици

Пищов

к (кило) = 10^4
М (мега) = 10^6
Г (гига) = 10^9
Т (тера) = 10^{12}
P (пета) = 10^{15}



Що е то тригер?



Trigger (анг.) – спусък, пусково устройство

- **Дефиниция:** Система която приема за по-нататъчна обработка или отхвърля данните от дадено събитие въз основа на характеристиките на събитието.
- Как го правим в **CMS**?
 - Тригерно ниво 1:
 - Търсят се интересни характеристики използвайки минимални данни и бърза електроника (FPGA чипове).
 - Решението се взема в рамките на 1 μ сек.
 - Намалява потока от данни от 40 милиона на 50 хиляди събития в секунда
 - Тригер от високо ниво:
 - Изчислителна мощ с размерите на CMS фермата за “Ниво 0”.
 - Правят се много по-пълна оценка на данните вземайки в предвид всичката налична информация.
 - Намалява потока на данни до 100 събития в секунда



Информационни технологии в ЦЕРН



IT Услуги

(непълн лист)

- Компютърни ресурси за данните от експериментите на LHC.
- Инфраструктура за експериментите
 - Ресурси за локални изчисления и съхранение на данни
- Поддръжка и развитие на програмната и хардуерната обеспеченост на работата на лабораторията



Статистика (Mayrin)

11207 сървъра
115 279 ядра
131 796 ТБ хард дискове
90 000 ТБ лентови носители
3.5 MW електрически капацитет
Резервно захранване с дизелови генератори

Статистика (Wigner)

11207 сървъра
115 279 ядра
131 796 ТБ хард дискове
90 000 ТБ лентови носители
8 MW електрически капацитет

Недостатъчно за
обработката на
всички данни -
**Необходими са 5
пъти повече
ресурси**



Какво се разбира под GRID



- Дефиниция (според Wikipedia):
 - [Electrical grid](#) (англ.) – електроснабдителна мрежа – взаимно свързана система за доставка на електроенергия от производителя до потребителите.
 - [Grid computing](#) (англ.) – компютърни ресурси намиращи се на различни места работещи за постигане на обща задача.
- Аналогията:
 - Включване към електрическата мрежа - ползване на ток
 - Включване към GRID – ползване на компютърни ресурси
 - Предимствата:
 - Няма значение къде се намират физически ресурсите (интересува ли ви точно от коя електроцентрала ви пристига тока?)
 - Няма значение как достигат до вас
 - Гарантирано параметри и отчетност на услугата



GRID за LHC

или какво е WLCG



WLCG

Worldwide LHC Computing GRID
Световен GRID за обработка на данни от LHC

- Какво е WLCG?
Глобална изчислителна инфраструктура целяща да снабдява с ресурси за съхраняване, разпределение и анализиране на данните генерирани от експериментите на LHC като подsigури достъпа до тях за всички партньори независимо от географското им местоположение.
- Състои се от следните локални GRID системи:
 - [EGI](#) (European Grid Infrastructure) – Европейска GRID инфраструктура (Европа)
 - [OSG](#) (Open Science Grid) – GRID за общодостъпна наука (САЩ)
 - [NDGF](#) (Nordic Data Grid Facility) – Скандинавски GRID
- Основни елементи:
 - Мрежови компоненти
 - Хардуер
 - GRID програмно обезпечение (middleware)
 - Програми за обработване на данни от експериментите
- Координира се от ЦЕРН
- Статистика:
 - 40 държави
 - 170 изчислителни центъра
 - 2 милиона изчислителни задачи на ден
 - 10 ГБ/сек глобална скорост
 - Над 10000 потребителя

Ниво-0 (ЦЕРН и Вигнер): ~ 20%

- Записване на първичните данни
- Първа реконструкция
- Разпределяне на данните към “Ниво-1”

Ниво-1 (13 центъра): ~ 40%

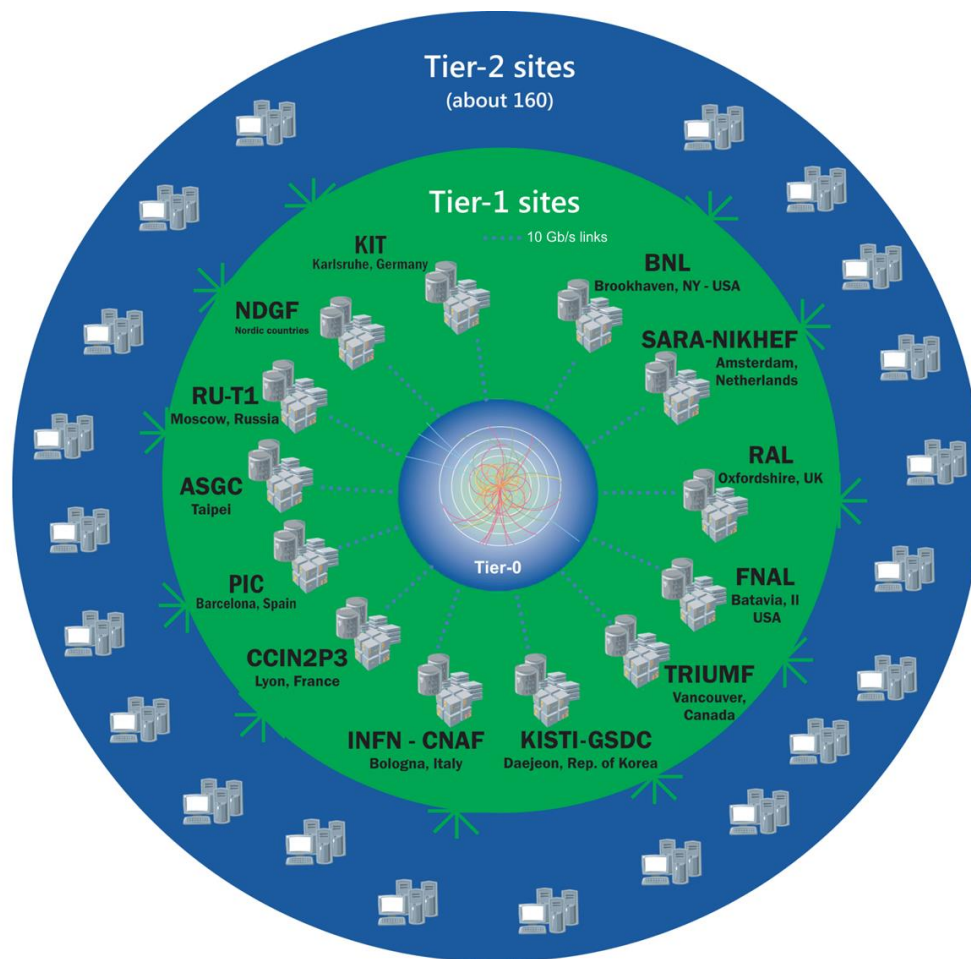
- Изисквания:
 - Лентови носители
 - Дисково пространство
 - Изчислителни ресурси
 - Свързан с LHCOPN (мрежа използвана само от WLCG) 10 GB/s
 - 24/7 поддръжка
- Функционалност:
 - Второ копие на първичните данни
 - Реконструкция на данните
 - Симулация
 - Съхранение на данните от реконструкцията

Ниво-2 (160 центъра): ~ 40%

- Изисквания..
- Функционалност:
 - Потребителски анализ на данни
 - Симулация и съхраняване на данни за потребителите

Ниво-3:

- Без фиксирани изисквания





Бонуси



- [HEP-SPEC06](#) - как да сте сигурни кой колко изчислителна мощност ви предоставя
- [CRAB](#) (CMS Remote Analysis Builder) – как физиците използват WLCG в CMS
- [GOCD](#) – как знаем кой сайт работи в момента
- [PhEDEx](#) – как контролираме и разпределяме данните в CMS по света
- [VOMS](#) (Virtual Organization Membership Service) – как в среда от множество колаборации от организации които участват до различна степен в различни операции оспяваме да знаем кой кой е какво има право да прави.
- Как наблюдаваме и контролираме качеството на данните:
 - [Dashboard](#) – визуализация на състоянието на системата
 - [SAM](#) (Service Availability Monitoring) / Hammercloud тестове – как проверяваме дали наистина всичко работи
- Видове данни, степени на редуцирането им и групирането им според анализа за който са предназначени и още, и още..
- Облаци (т.е. Cloud инфраструктури), виртуални машини, конфигурирането и използването им.



Изтоницици



- Използвани за тази презентация
 - Интернет:
 - [Страницата на WLCG](#)
 - [Център за съхранение на данни в ЦЕРН](#)
 - [Хардуерна инвентаризация на ЦЕРН](#)
 - Презентации:
 - “Въведение в ГРИД” – Преслав Константинов, Българска учителска програма, ЦЕРН, 2013
- За повече информация:
 - Книги:
 - [“The Grid 2: Blueprint for a New Computing Infrastructure”](#) – I. Foster, C. Kasseman, ISBN: 978-1-55860-933-4
 - Интернет:
 - [Натоварване на мрежата “Ниво-1” в реално време](#)
 - [Кой с колко участва в WLCG](#)
 - Видео:
 - [Обработка на данните от LHC](#)
 - [Анимиран GRID](#)
 - [Отдела на ЦЕРН за информационни технологии в 8 минути](#)
 - [Как изглежда един лентов робот на работа](#)
 - [Центъра за данни във Вигнер, Унгария](#)