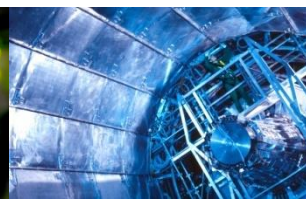
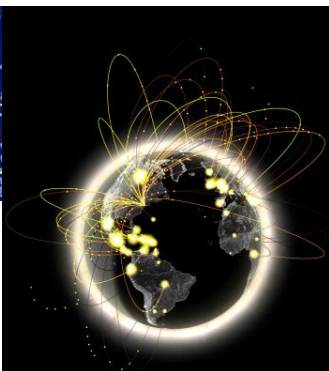




# Информационни технологии в ЦЕРН от WWW до GRID



**др. Иван Глушков**  
**Българска инженерна учителска програма**  
**ЦЕРН, октомври 2016**



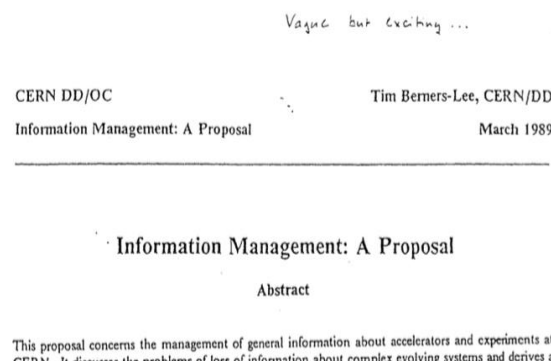
# WWW



- WWW (World Wide Web), the Web (мрежата) – Система в която ресурсите са идентифицирани еднозначно чрез URL (Uniform Resource Locator) и са свързани чрез hyperlinks.
- Създадена в ЦЕРН от Tim Berners-Lee през 1989 г. с цел споделяне на данни между учени от различни институции и университети по света.
- Публикуван през 1994 като: Безплатен и свободен
- **Днес:** Да запазим мрежата такава – безплатна и свободна.
- Връзки:
  - [Бъдещето на www](#)
  - [Първата уеб страница в света](#)
  - [Първата уеб страница през първия уеб браузър \(line-mode browser\)](#)



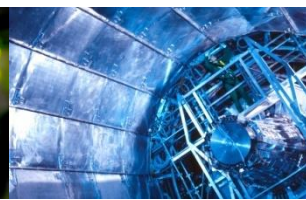
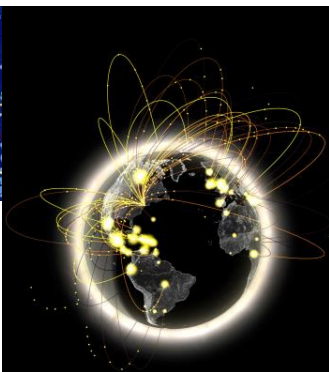
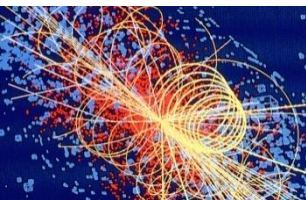
Работната NeXT станция на която е бил инсталиран първия web server.



Оригиналното предложение за "Information Management". С коментара на ръководителя на Тим – "Vague but exciting.."



# GRID ТЕХНОЛОГИИ



**др. Иван Глушков**  
**Българска инженерна учителска програма**  
**ЦЕРН, октомври 2016**



# 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?
  - 1 (1.5) MB / събитие в CMS (ATLAS)
  - 40 MHz честота на сблъскване (40 000 000 събития / секунда)
  - 40 TB/ секунда!! ( $10^6$  PB / година)
- Колко можем да запишем?
  - Максимум 1 000 събития / секунда. (Пробвахме и 2000..)
- Колко данни записваме общо? (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) =  $10^3$

М (мега) =  $10^6$

Г (гига) =  $10^9$

Т (тера) =  $10^{12}$

Р (пета) =  $10^{15}$



# Що е то тригер?

## Или какво разбираме под “онлайн” в CMS



**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 подбираме събитията в **CMS**? – На две стъпки:
  - Стъпка 1: “Тригерно ниво 1”:
    - Решението се взима въз основа само на 192 бита.
    - Данните идват от най-бързите системи в CMS: Калориметрите и мюонните камери (RPC, българско участие)
    - Имплементиран на “поръчкова” електроника - ASIC (Application Specific Integrated Circuit) и FPGA (Field-Programmable Gate Array) чипове.
    - Решението се взима в рамките на 1  $\mu$ сек.
    - Намалява потока от данни от 40 милиона (40 MHz) на 100 хиляди събития в секунда (100 kHz)
    - Ограничението идва от електрониката, която чете събитията.



# Тригер от високо ниво



- Как подбираме събитията например - в **CMS**?
  - Стъпка 2: “Тригер от високо ниво”:
    - Прочитат се данните от всички детектори (16 милиона канала!)
    - Технически детайли:
      - Ресурси: 16 000 Xeon ядра, 2 GB RAM / ядро
      - Скорост на входящите данни: 100 хиляди събития на секунда
      - Време за реконструкция на едно събитие : 160 ms.
      - Софтуеър: Набор от алгоритми, имплементирани с оптимизирана версия на стандартния пакет за обработка на данните на CMS (CMSSW). Състои се от ~ 2500 модули (C++) , изпълнени в различна последователност и с различна конфигурация (python).
    - Правят се много по-пълна оценка на данните.
    - Намалява потока на данни до 1000 събития в секунда



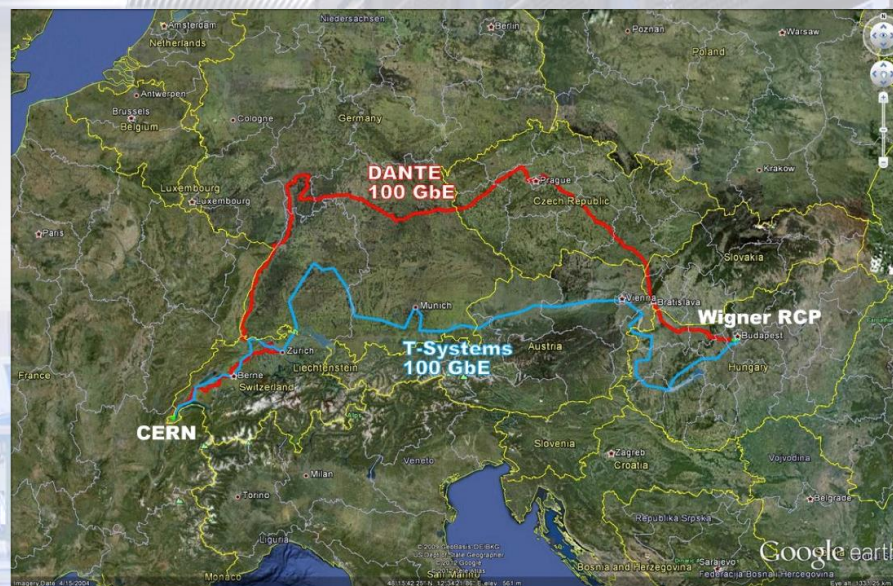
# Информационни технологии в ЦЕРН



## IT Услуги

(непълнен 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 безопасност 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.
    - Гарантирани параметри, отчетност и сигурност 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



# WLCG



- Основни елементи:
  - Мрежови компоненти
  - Хардуер
  - GRID програмно обезпечение (middleware)
  - Програми за обработване на данни от експериментите
- Статистика:
  - 40 държави
  - 170 изчислителни центъра
  - 2 милиона изчислителни задачи на ден
  - 10 ГБ/сек глобална скорост
  - Над 10 000 потребителя



# Структура на WLCG



## Ниво-0 (ЦЕРН и Вигнер): ~ 20%

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

## Ниво-1 (13 центъра): ~ 40%

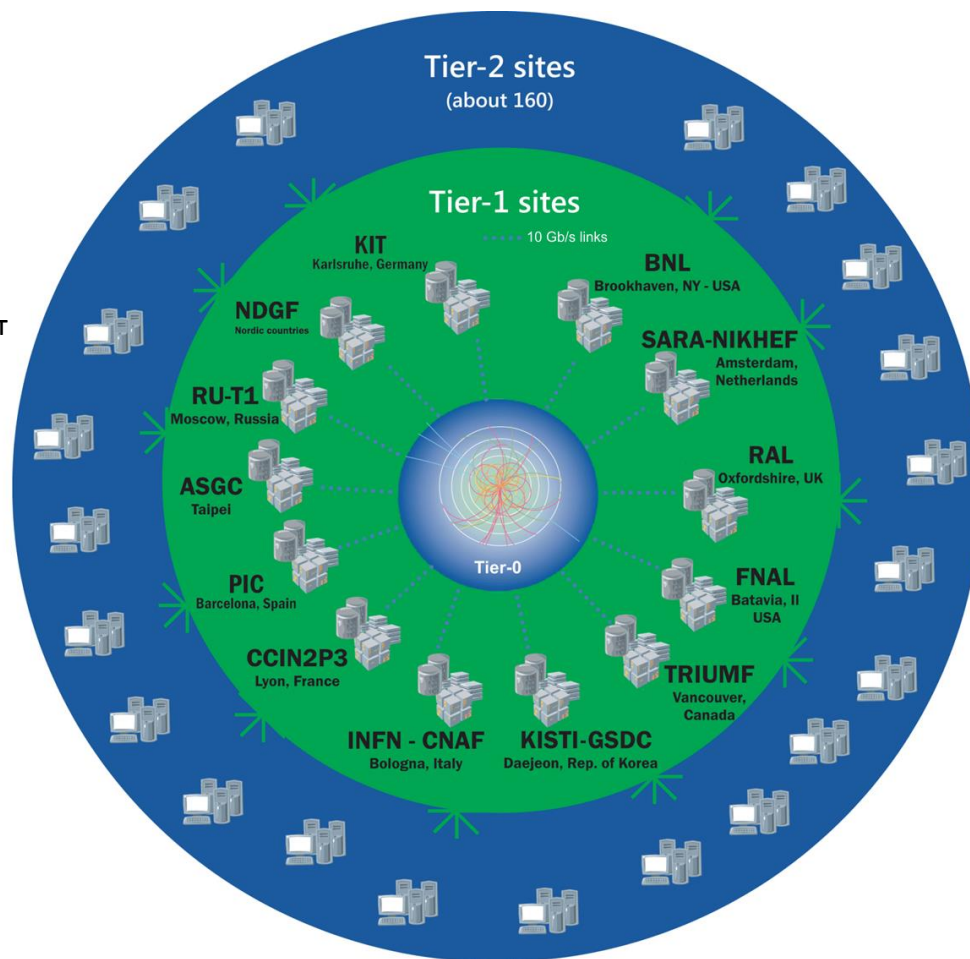
- Изисквания:
  - Лентови носители
  - Дисково пространство
  - Изчислителни ресурси
  - Свързан с LHCOPN (мрежа използвана само от WLCG) 10 GB/s
  - 24/7 поддръжка
- Функционалност:
  - Второ копие на първичните данни
  - Реконструкция на данните
  - Симулация
  - Съхранение на данните от реконструкцията

## Ниво-2 (160 центъра): ~ 40%

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

## Ниво-3:

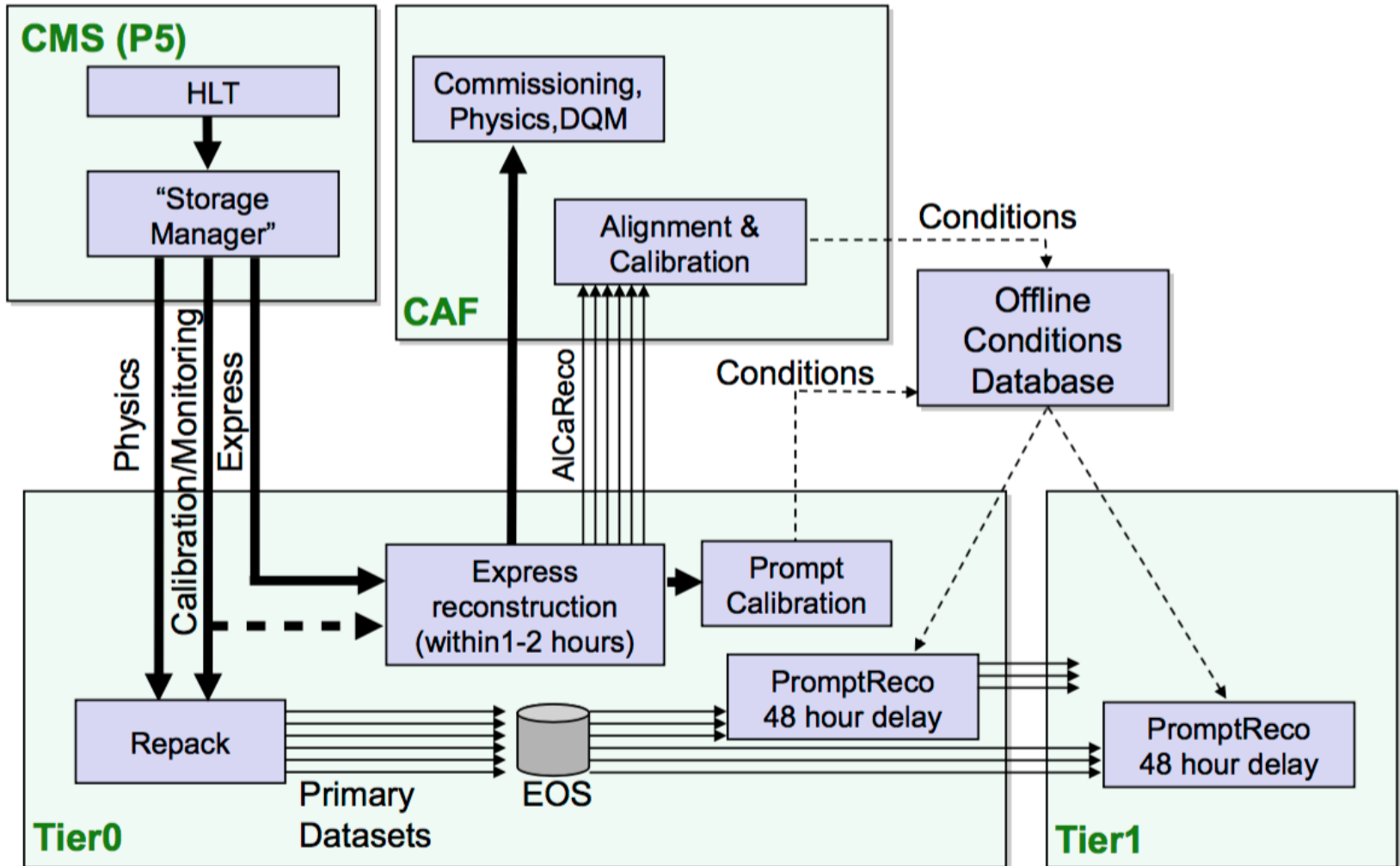
- Без фиксирани изисквания



[Обобщение](#)



# CMS Ниво-0





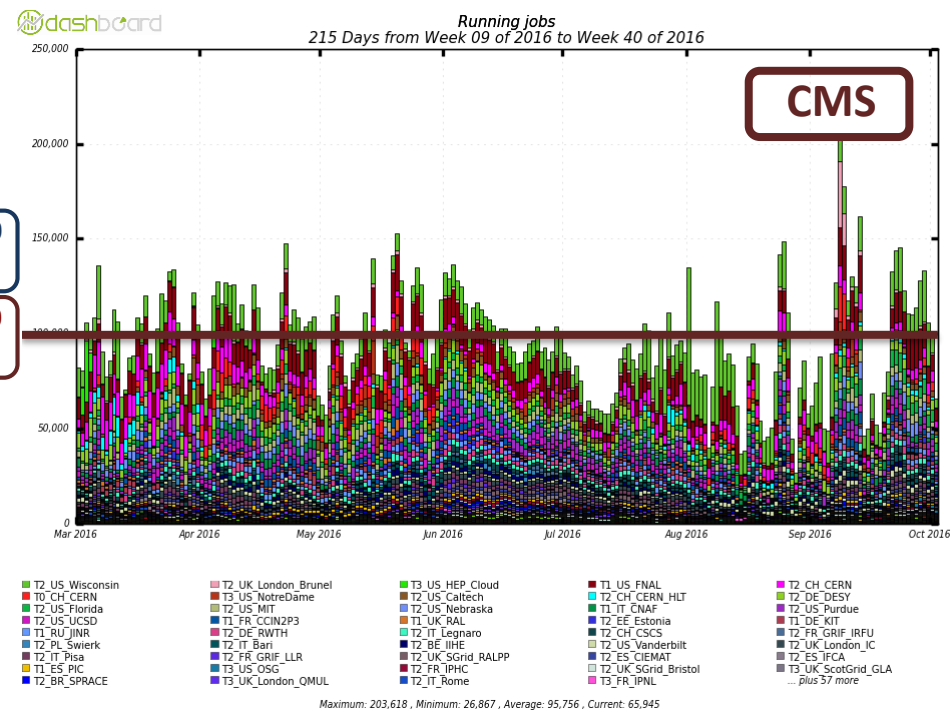
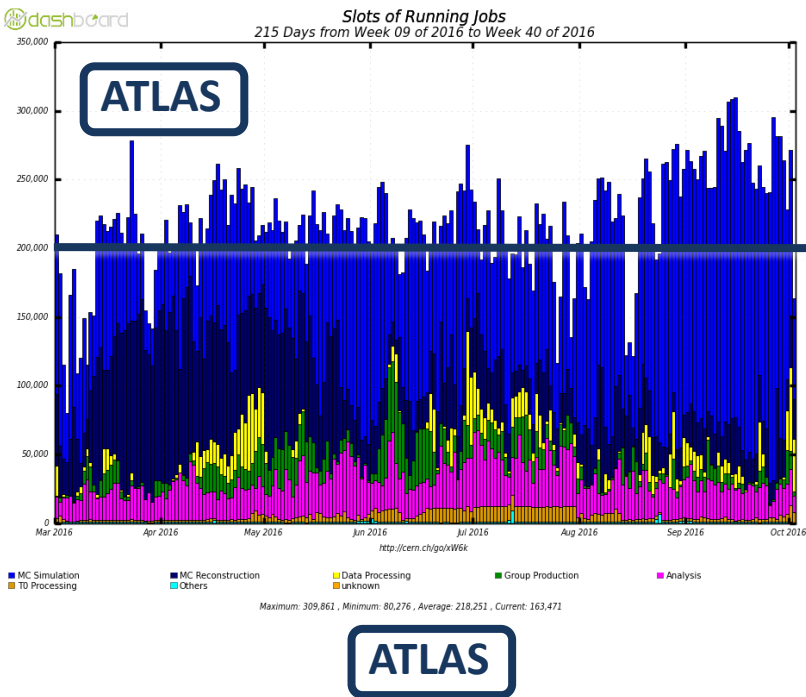
# Инфраструктура за разпределение на задачи GlideInWMS



- Проблем:
  - Имаме: Много потребители с много различни задачи с различни приоритети за смятане и ресурси от много и различни сайтове
  - Как: да свършим възможно най-много работа за възможно най-кратко време така че да е най-справедливо?
- Решение:
  - Т.нар. “пилотен подход” – разпращаме на всички **необходими** ресурси задачи (наречени “пилоти” или glidein-и), които образуват стандартна компютърна ферма разпределена върху всички сайтове. За целта използваме [GlideInWMS](#) (The Glidein-base Workload Management System)
    - Фабрика за glidein-и
    - Frontend – кой казва на фабриките какво да правят
  - За разпределение на задачите използваме стандартен софтуер за компютърни ферми – [HTCondor](#)
  - HTCondor централизирано напасва висящи задачи (“idle jobs”) към свободни ресурси (“unclaimed slots”) в зависимост от **текущите** преоритети.



# Колко ядра ползваме в ATLAS и CMS?



## Разпределение по видове задачи

- **Симулация на детектора**
- **Реконструкция на частици**
- **Работещи физици**
- **Ниво 0**
- **и т. н.**

## CMS

- Разпределение по сайтове
- 100 сайта



# Пренос на данни в CMS

## (дизайн)



### PhEDEx

### (Physics Experiment Data Export)

система за пренос на файлове базирана на множество независими агенти с определни състояния и централизирана Oracle базирана база данни за отчетност (TMDB - Transfer management DB).

#### Съставни елементи

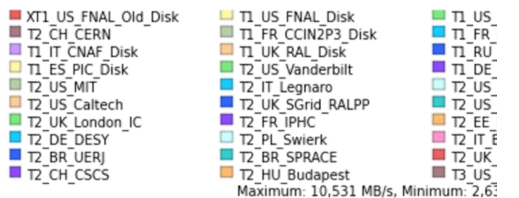
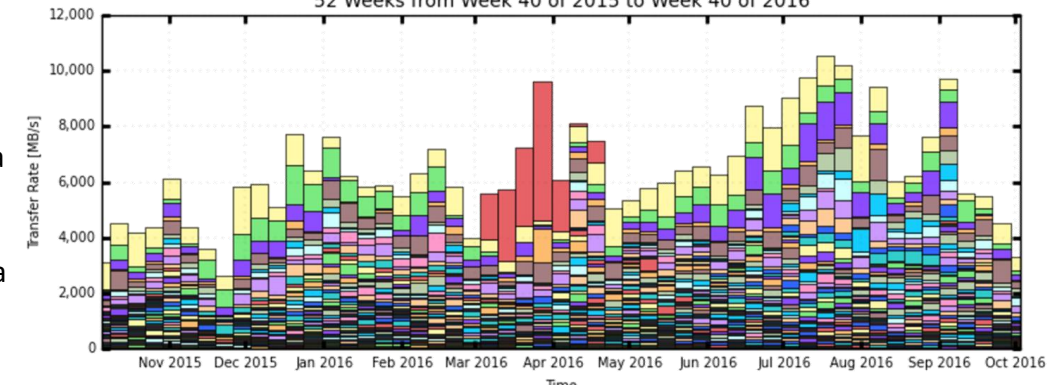
- На всяко географско място (сайт) където искаме да пренасяме данни съответства по една група агенти
- Комуникацията между отделните агенти е забранена по дизайн
- Динамични рутиращи алгоритми за оптимален трансфер на данни
- Web инструменти за наблюдение на системата

#### Предимства:

- TMDB - улеснява намирането и решаването на проблеми
- Независимостта на агентите дава възможност за имплементирането им на произволен програмен език

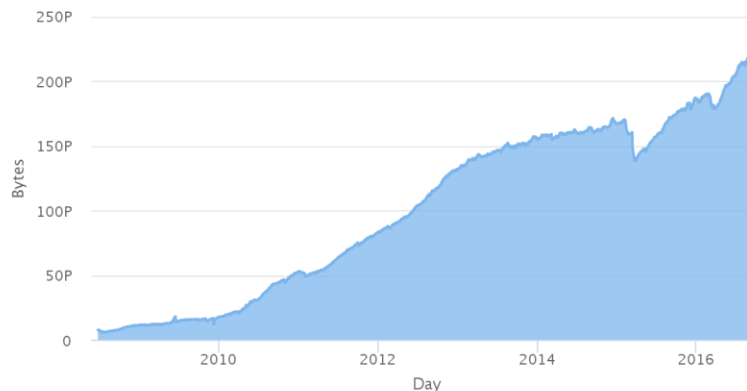
### CMS PhEDEx - Transfer Rate

52 Weeks from Week 40 of 2015 to Week 40 of 2016



### ATLAS Data Overview

Worldwide







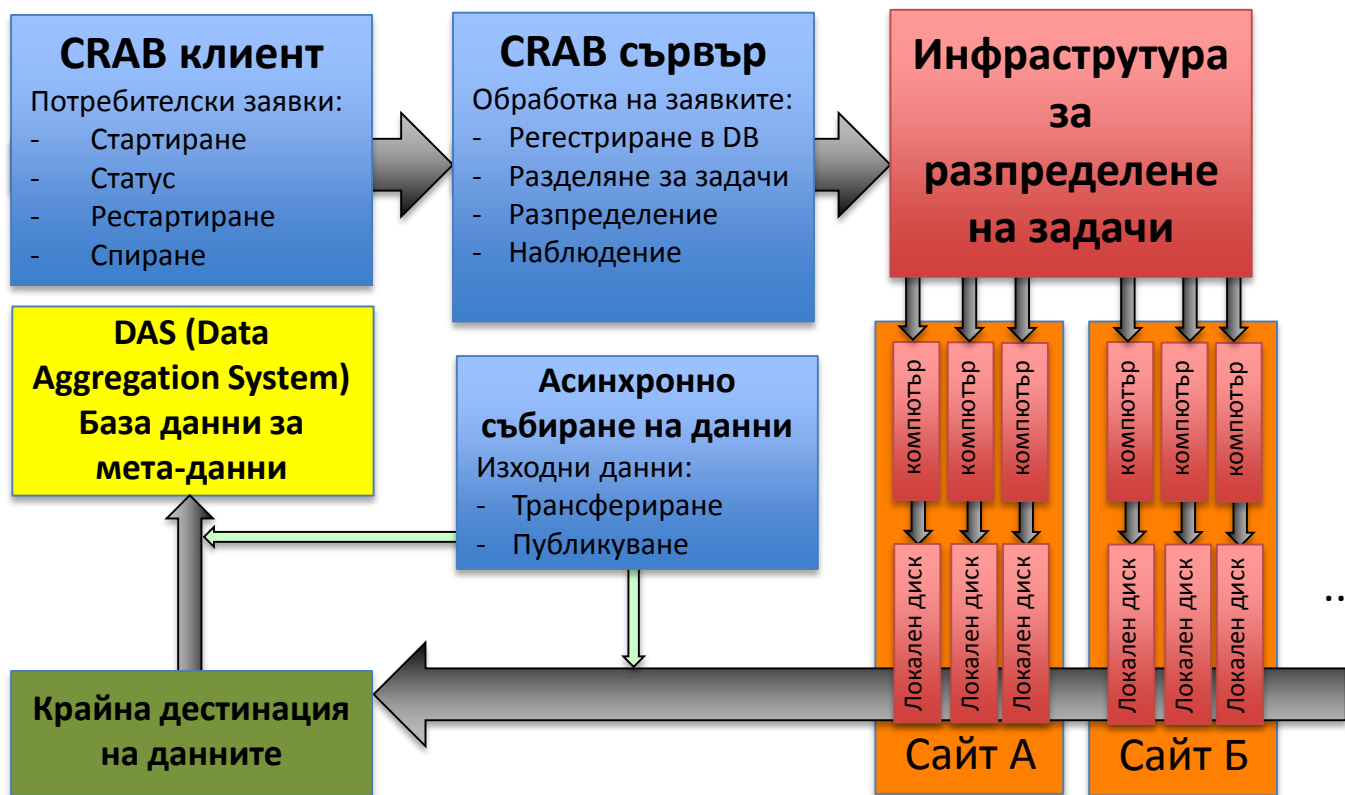
# Анализ на данни



CRAB

(CMS Remote Analysis Builder)

python базирана система за отдалечена обработка на данни.





# Какво пропускам..



- Какво пък толкова има за смятане в тези частици че използваме толкова ресурси
- Видове данни, степени на редуцирането им и групирането им според анализа за който са предназначени
- [GlideInWMS](#) – Инфраструктура за разпределение на задачи.
- [CRAB](#) (CMS Remote Analysis Builder) – Как физиците използват WLCG в CMS за анализиране на данни
- [GOCDDB](#) – 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.
- [SiteDB](#) – Каталогизиране на участващите в CMS GRID сайтовете и хората асоциирани с тях.
- 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
- Облаци (т.е. Cloud инфраструктури), виртуални машини, конфигурирането и използването им.
- Финансирането и планирането на развиртието на GRID година за година.
- [HEP-SPEC06](#) - Как да сте сигурни кой колко изчислителна мощност ви предоставя
- Над какво работим:
  - Опортюнистични ресурси или какво да правим когато някои ни заеме 20000 ядра за 2 седмици.
  - CMS@Home - Как може всеки да помогне за обработването на данните от CMS.
  - AWS (Amazon Web Services) или какво ще правим ако дори и GRID ни е тесен.



# Технологии в ЦЕРН



indico

Google Docs

collaborate with TWiki

Doodle

Vidyo

amazon web services

Open Science Grid

esi

NORDUGRID  
Grid Solution For Wide Area Computing and Data Handling

Edison

CernVM File system

CASTOR  
CERN Advanced STORage manager

EOS



CERNBox

Jenkins

puppet

ceph openstack

Nagios Hammercloud

dashboard

WLCG

Worldwide LHC Computing Grid

git

C++

DIRAC  
THE INTERWARE

python

#!/bin/bash

ROOT  
Data Analysis Framework

SQL  
Database Language

APACHE Spark  
elasticsearch  
hadoop

logstash kibana

Grafana

ORACLE

CentOS  
Scientific Linux

GNU

HTCondor  
High Throughput Computing

LSF  
PanDA

intel



Linux OS X

Windows ubuntu

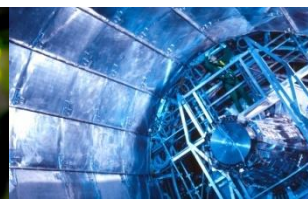
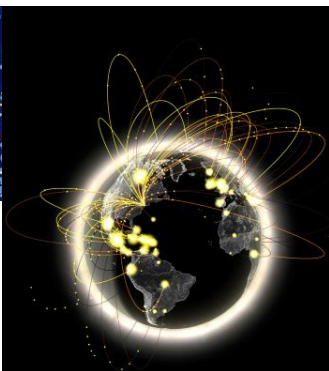
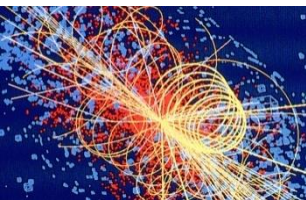
AFS

servicenow

lustre



# Благодаря за вниманието..



... Въпроси?



# Източници



- Използвани за тази презентация
  - Интернет:
    - [Страницата на WLCG](#)
    - [Център за съхранение на данни в ЦЕРН](#)
    - [Хардуерна инвентаризация на ЦЕРН](#)
    - [Тригера от виско ниво на CMS](#) (The CMS High Level Trigger), Andrea Bocci, CHEP2015, Okinawa, Japan
  - Презентации:
    - “Въведение в GRID” – 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](#)
- Последните новости (CHEP2016 ще се проведе следващата седмица):
  - [Презентации от конференцията CHEP2015 \(Computing in High Energy Physics\)](#)
  - [Конференция 2015 на колаборацията WLCG](#)