

# Вычислительная инфраструктура исследовательских ядерно-физических лабораторий

(введение)

- **Современные запросы на вычисления**
  - наука/техника/образование
  - Бизнес, военные приложения
  - государственные структуры
- **История вычислителей.**
- **Организация крупномасштабных вычислений**
  - оборудование, программы
  - Вычисления в сетях Grid и «облачных» сетях
  - Вычисления в крупных физических экспериментах
- **Особенности вычислений в ФВЭ**
  - Большие объёмы данных
  - Коллективная работа

# Роль вычислений в обществе (не только в ядерной физике)

- Развитие человечества как природного феномена невозможно без экономического роста.
- Экономический рост в настоящее время и будущем может основываться лишь на использовании инноваций в массовом порядке.
- Любые инновации возможны лишь с усложняющейся переработкой больших объёмов информации (проще - вычислений).

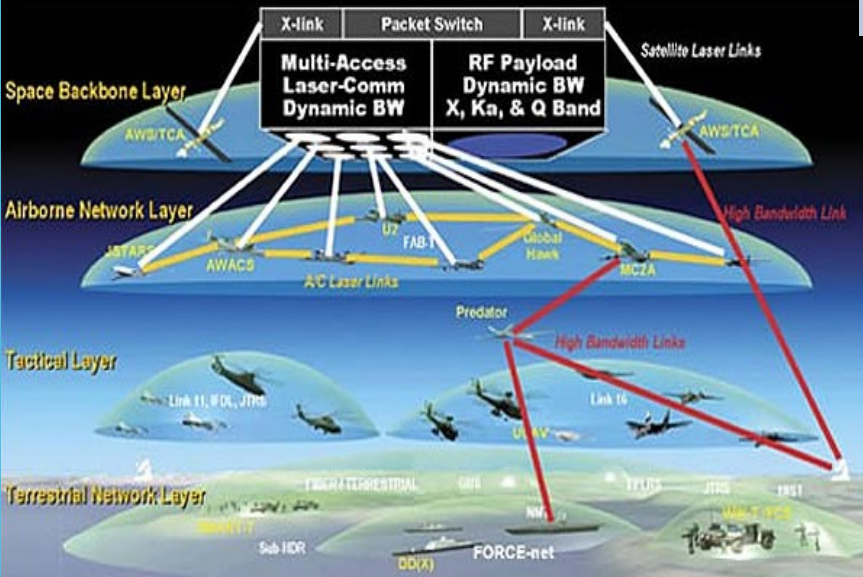
# Наука/техника/образование

- **Расчёты/моделирование в**
  - сейсмологии (землетрясения, проч.);
  - гидрологии (плотины, проч.);
  - органической и неорганической химии;
  - **ядерной физике (физика высоких энергий, реакторы);**
  - прогнозах (атмосфера, океаны, цунами, штормы, проч.);
  - биологии/медицине (томограммы);
  - поиске полезных ископаемых;
  - мировой экономике (по отраслям, глобально, по регионам, поведение биржи, кризисы);
  - космических исследованиях;
  - в коммуникациях (технических и человеческих);
  - инженерных расчётах (транспорт, двигатели, горючее, механизмы, сооружения, проч.).

# Финансовые и родственные приложения

- Выполнение традиционных банковских операций.
- «Электронные деньги»
- Кредитные карты, банкоматы, кассовые аппараты.
- Учёт движения товаров.
- Автоматизированные складские помещения.

# Военные приложения



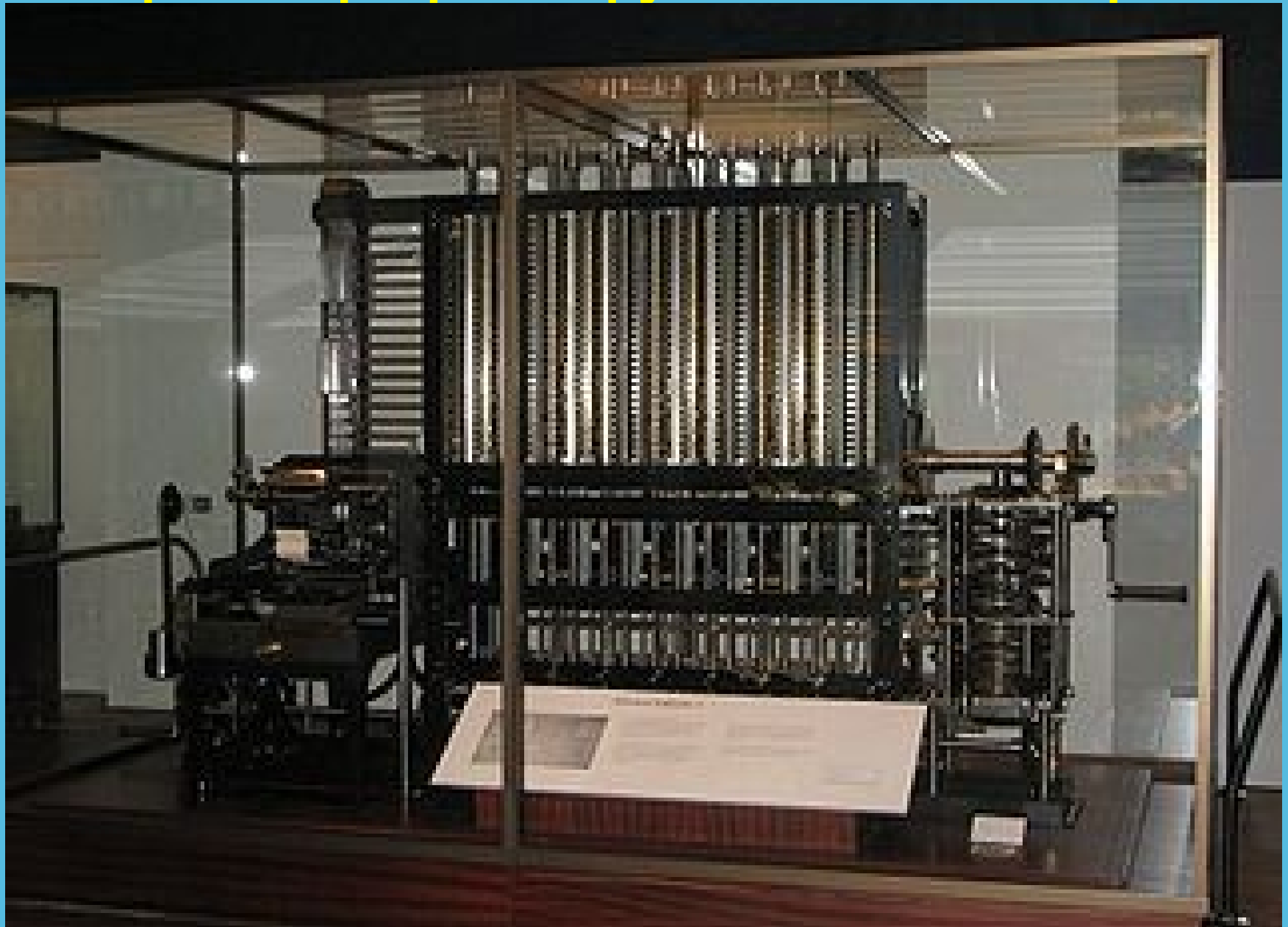
# Программы «электронное правительство»

- Москва - <http://www.mos.ru/>
- С. Петербург - <http://www.gov.spb.ru/>
- Ленинградская область - <http://www.lenobl.ru/>
- Государственная власть России - <http://www.gov.ru/>
- Государственные услуги <http://www.gosuslugi.ru/>

# Можно ли использовать один большой компьютер для любых (всех) отраслей?

- Различные требования к вычислениям и терминология в разных отраслях деятельности
- Разный характер вычислений и вычислителей
- Отличающиеся условия использования (окружающая среда)
- Неодинаковые прикладные программы
- *Вывод – каждая область человеческой деятельности имеет специфические потребности в вычислениях*

# Первый программируемый компьютер





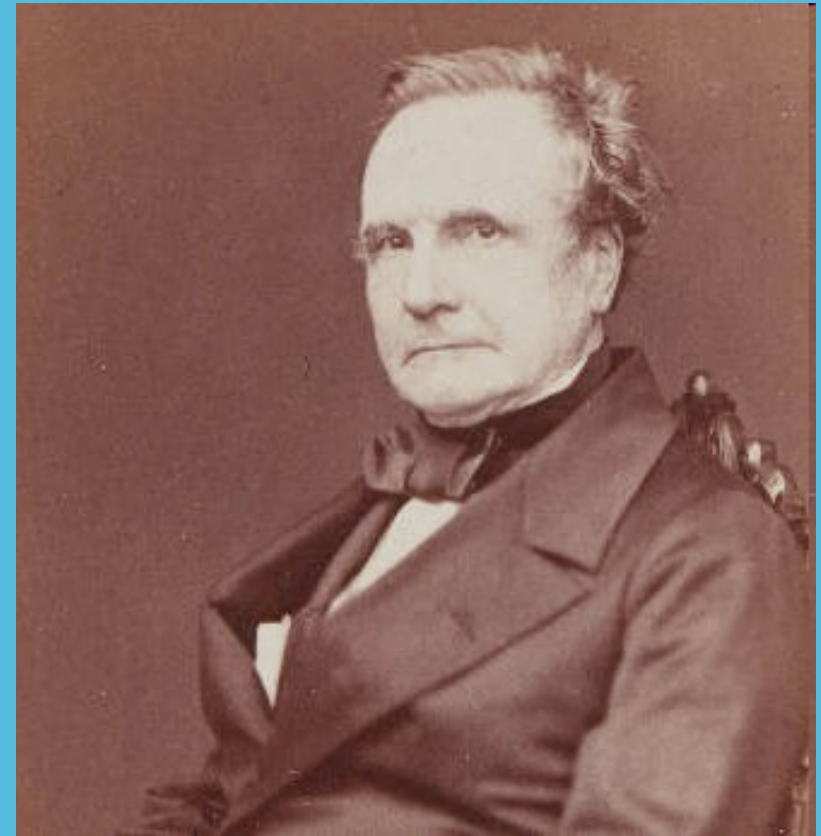
Charles Babbage (1791 - 1871) –  
Изобретатель первого  
Программируемого вычислителя



Ada Lovelace (Byron)

1815 - 1852

первый программист мира





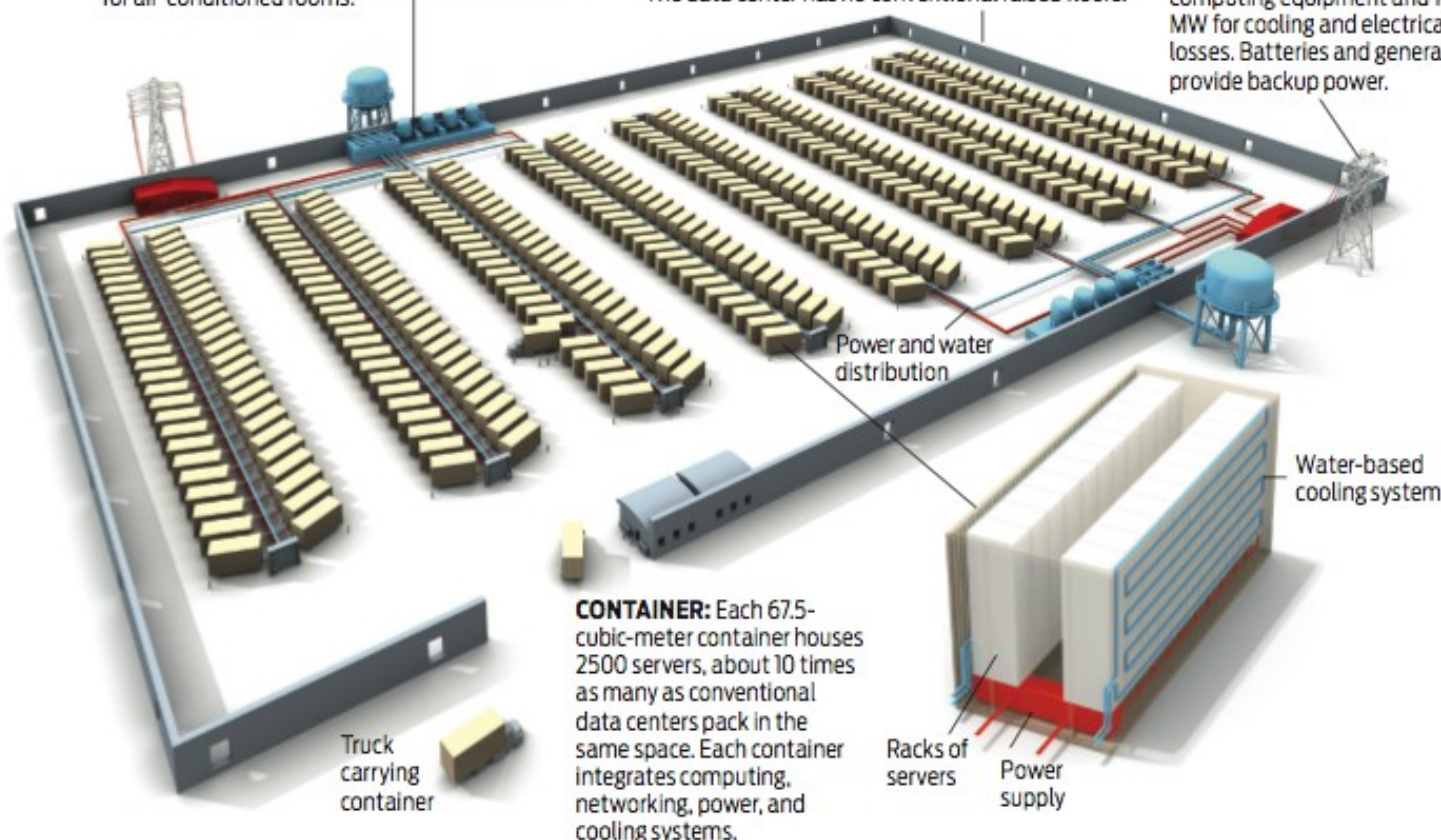
She speculated that the Engine 'might act upon other things besides number... the Engine might compose elaborate and scientific pieces of music of any degree of complexity or extent'.

# План большого центра данных (кластера)

**COOLING:** High-efficiency water-based cooling systems—less energy-intensive than traditional chillers—circulate cold water through the containers to remove heat, eliminating the need for air-conditioned rooms.

**STRUCTURE:** A 24 000-square-meter facility houses 400 containers. Delivered by trucks, the containers attach to a spine infrastructure that feeds network connectivity, power, and water. The data center has no conventional raised floors.

**POWER:** Two power substations feed a total of 300 megawatts to the data center, with 200 MW used for computing equipment and 100 MW for cooling and electrical losses. Batteries and generators provide backup power.



# Большие центры обработки данных



**WAREHOUSE-SIZE COMPUTERS:** Google has built a sprawling data center on the banks of the Columbia River, in The Dalles, Ore. The site, with two server-packed buildings and space for a third, houses tens of thousands of computers—the exact number is a closely guarded secret. Microsoft, Yahoo, and Amazon are also building data centers in the region, enticed by its readily available fiber-optic connectivity and cheap electricity. PHOTO: MELANIE CONNER

# Программное обеспечение

- **Операционные системы**
  - Unix-like: Linux, Solaris, Android, etc
  - MS Windows
- **Языки программирования**
  - C, C++ и клоны
  - Fortran и другие процедурно-ориентированные ЯЗЫКИ
  - Java, JavaScript, Ruby, Perl, Python, PHP, оболочки (bash, csh, etc) и т.д.
- **Прикладное программное обеспечение**
  - СУБД, например, MySql, PostGreSql, etc

# Наиболее обсуждаемые компьютерные темы - 2013

- Использование облачных вычислений.
- Многоядерные процессоры и использование специальных средств типа GPGPU, FPGA.
- Big Data — Данные большого объёма (~100ТВ и более в 2013)
- Увеличение объёма вычислительной мощности на ватт потребляемой мощности.
- **Безопасность в сетях** (предотвращение недружественных вторжений, вирусы, спам) **не только в кластерах.**
- Дополнительно см. конференцию [sc13.supercomputing.org](http://sc13.supercomputing.org) — Super Computing

# Крупнейшие компьютерные кластеры мира

- **В мире**

- <http://www.top500.org> (список крупнейших кластеров мира – обновляется дважды в год)

- **В России**

- <http://www.supercomputers.ru/> (список крупнейших кластеров в России и СНГ)

# Наиболее обсуждаемые темы в программном обеспечении 2013

- Использование Свободно распространяемого Программного Обеспечения (СПО) [некоммерческого]
  - Операционные системы Линукс (различные дистрибутивы); как использовать многоядерность
- Виртуализация
  - настольных компьютеров;
  - серверов;
  - компьютерной инфраструктуры.
- Новые языки программирования (например, C++11)
- Как обрабатывать большие объёмы данных.



# Перспективные вычислительные структуры 2013

- **Грид (grid) системы (много проектов).**  
**Информации о таких системах на сайтах:**
  - <https://sites.google.com/site/clustergateorg>
  - <http://www.globus.org/>
  - <http://parallel.ru/>
- **Облачные вычисления (cloud computing):**
  - [Amazon.com/ec2](http://Amazon.com/ec2), [rackspace.com](http://rackspace.com) и сотни других  
Системы организации облаков: [OpenStack.org](http://OpenStack.org)  
и десятки других
- **Гетерогенные вычислительные установки с использованием GPGPU, FPGA**

# Использование Grid для науки EGI

- **European Grid Initiative site –**
- *<http://www.egi.eu/>*
  - Основные области использования
    - Биология
    - Исследование земли и окружающего пространства
    - Физика Высоких Энергий

# Вид тоннеля ускорителя БАК (LHC)

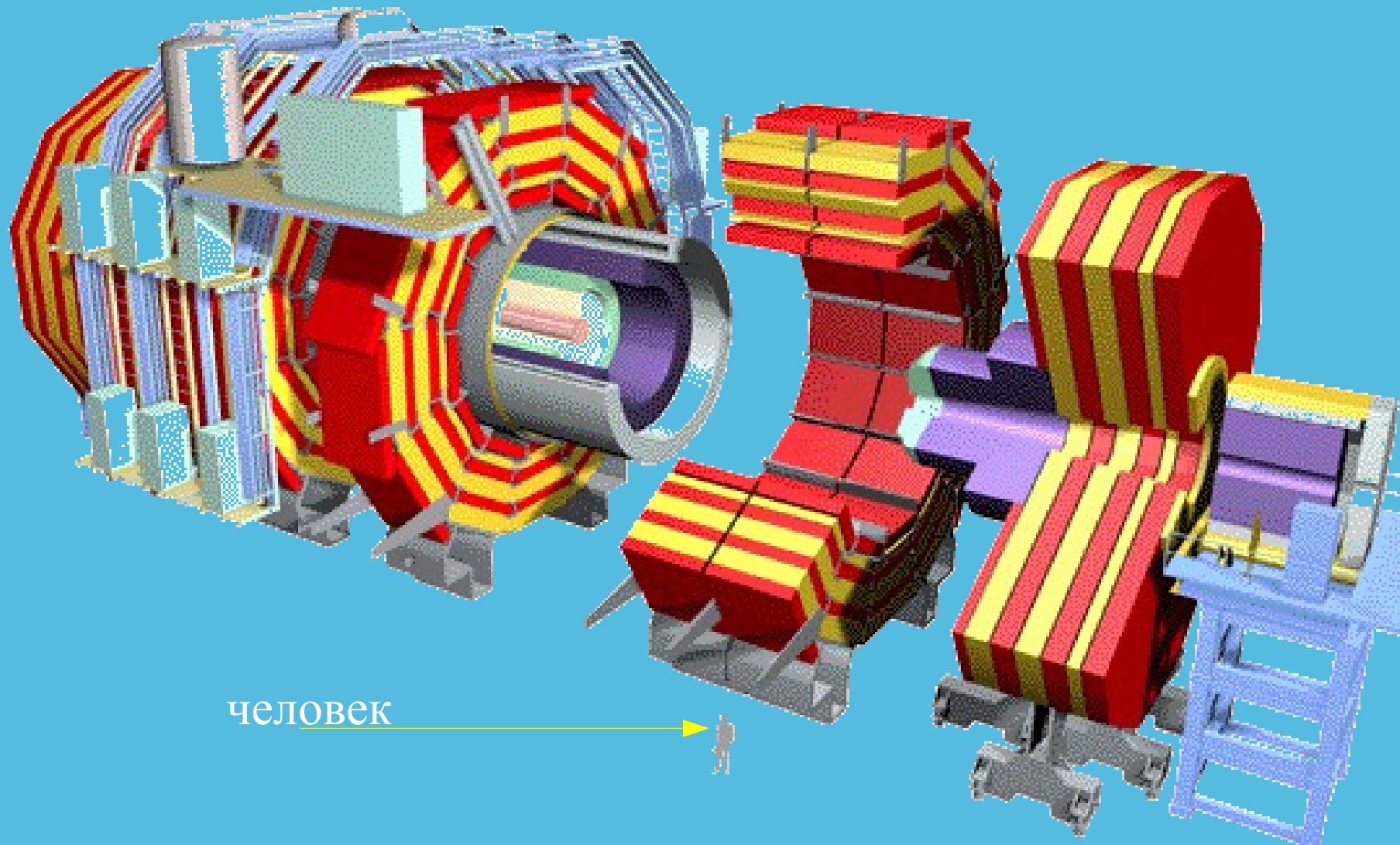
600 Million Proton Collisions/second

SC Magnets cooled to  $-193.2^{\circ}\text{C}$  (80 K),  
 $-268.7^{\circ}\text{C}$  (4.5K),  $-271.3^{\circ}\text{C}$  (1.9k)  
Using 10'000 tons of liquid  
nitrogen and 120 tonnes of liquid helium

26659m in Circumference (but varies with the moon!)

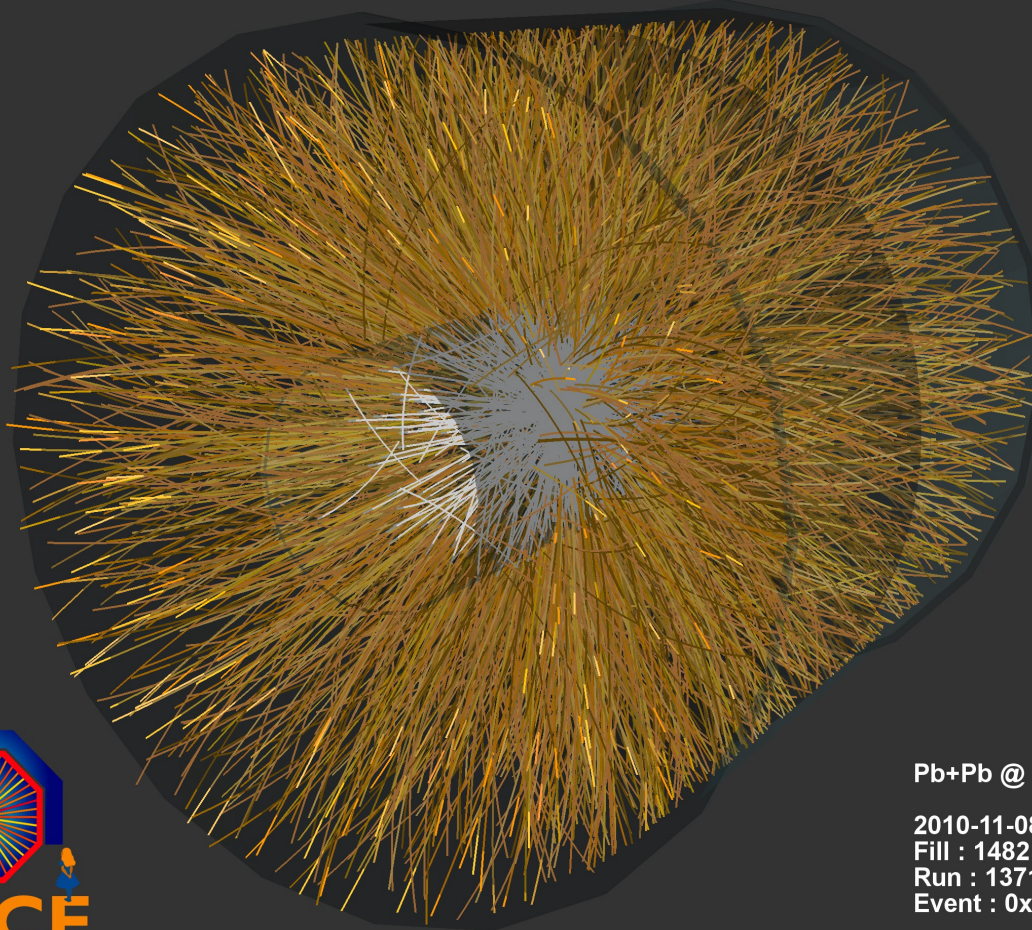


# Пример размеров детектора на LHC



ЧЕЛОВЕК

# Столкновения частиц



Pb+Pb @  $\sqrt{s} = 2.76$  ATeV

2010-11-08 11:29:52

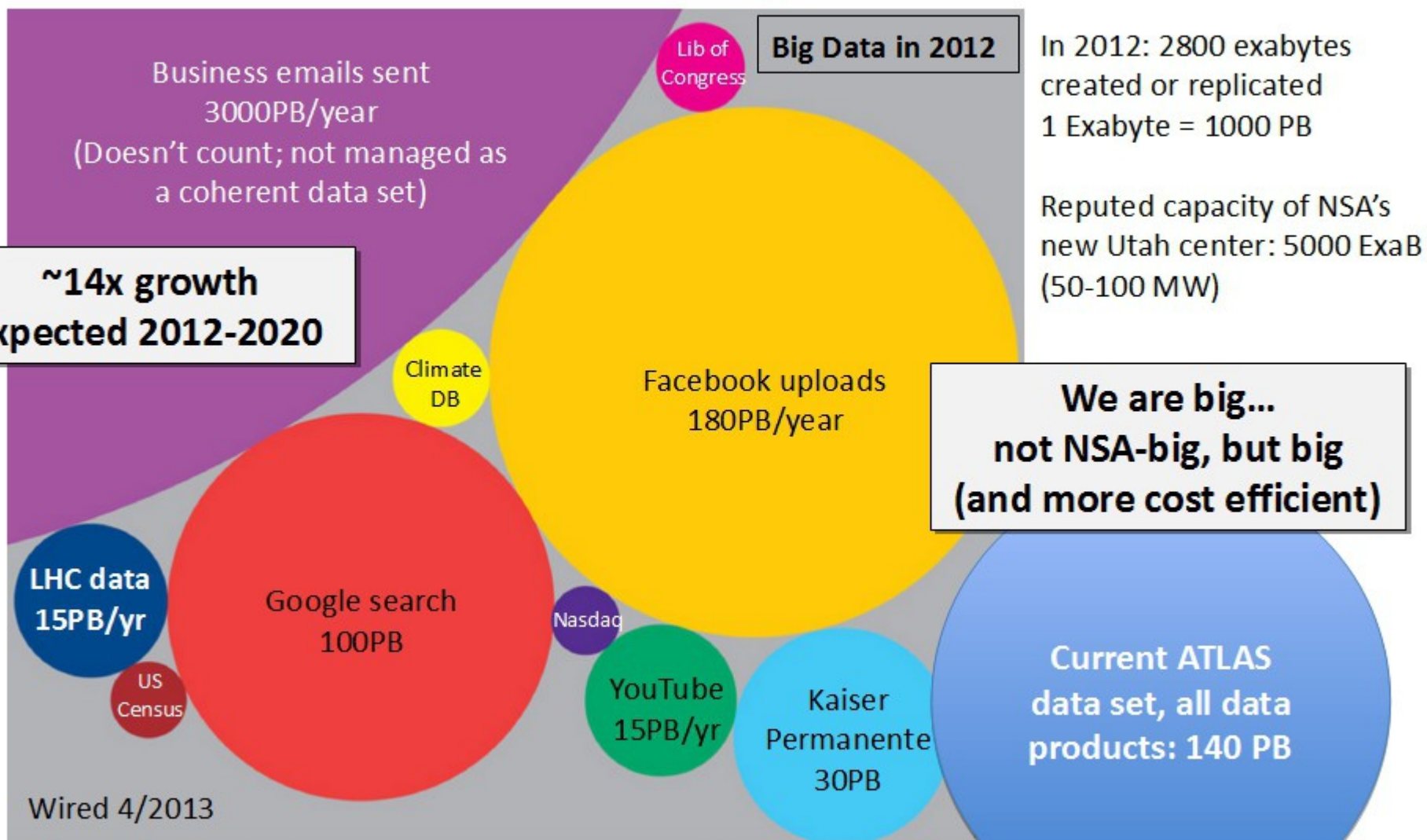
Fill : 1482

Run : 137124

Event : 0x0000000042B1B693

# Data Management

## Where is LHC in Big Data Terms?



<http://www.wired.com/magazine/2013/04/bigdata/>

# Основные типы и вычислений в ФВЭ

- **Online** («в линию») вычисления (определяются условиями измерений в эксперименте)
  - должны быть выполнены максимально быстро (основной критерий - продуктивность для физики). Коллективная работа: программисты, электронщики, физики, студенты.
- **Offline** («не на линии») вычисления (определяются логикой обработки экспериментальных измерений) – это обработка большого объёма данных:
  - часто выполняется с использованием распределённых по миру компьютерных систем;
  - большой объём данных, состоящих из статистически независимых фрагментов данных (**событий**), описывающих отдельные физические события в измерительной установке.
  - подготовкой обработки и обсуждением результатов обработки данных занимается большое количество лиц (десятки или сотни).
    - Результаты компьютерной обработки данных эксперимента – это как правило **коллективный труд, а не достижения отдельных персон.**
- **Сопутствующие вычисления:** всё остальное, что выполняется на компьютерах в исследовательских учреждениях.



# Заключение - 1

- Компьютеры используются в любых отраслях человеческой деятельности, но в каждой имеются особенности использования.
- Современные особенности компьютерных архитектур (*всё используется в ядерной физике*)
  - гетерогенные компьютерные кластеры
  - свободное программное обеспечение
  - виртуализация
  - облачные и Грид архитектуры

# Заключение - 2

- **Компьютерная инфраструктура экспериментальной ядерной физики/физики высоких энергий характеризуется:**
  - Большим объёмом данных ( часто измеряется в ТВ -  $10^{12}$  байтов и даже в PB –  $10^{15}$  байтов)
  - Высокой скоростью поступления измерительных данных (много MB/sec)
  - Большими запросами на вычислительные мощности (много тысяч компьютеров)
- **Два вида компьютерной инфраструктуры**
  - On-line – вычисления во время проведения физических измерений, где вычисления являются неотъемлемой частью физических экспериментов
  - Off-line – вычисления связанные с обработкой измерительных данных и анализом результатов обработки. Такие вычисления могут выполняться независимо от проведения физических экспериментов.

# Заключение - 3

- Данные состоят из отдельных фрагментов (могут иметь размер измеряемый в МВ) , которые содержат информацию о физических событиях. Эти фрагменты данных именуются “события”.
- В большинстве экспериментов события являются статистически независимыми, что даёт возможность обрабатывать каждое событие независимо от других. Например, параллельно, или в одно и то же время.

# Заключение - 4

- **Обработка данных полученных в эксперименте может длиться продолжительное время (до нескольких месяцев или лет). Может производиться удалённо от места хранения данных.**
- **Большое распространение получает удалённое наблюдения за проведением эксперимента.**

**Конец лекции**

# Можно ли принять участие в моделировании событий на LHC ?

- Ответ - “Да” для любого человека с компьютером, см.

<http://lhcatome.web.cern.ch/LHCathome/>