

Grid

GRID

СИСТЕМЫ



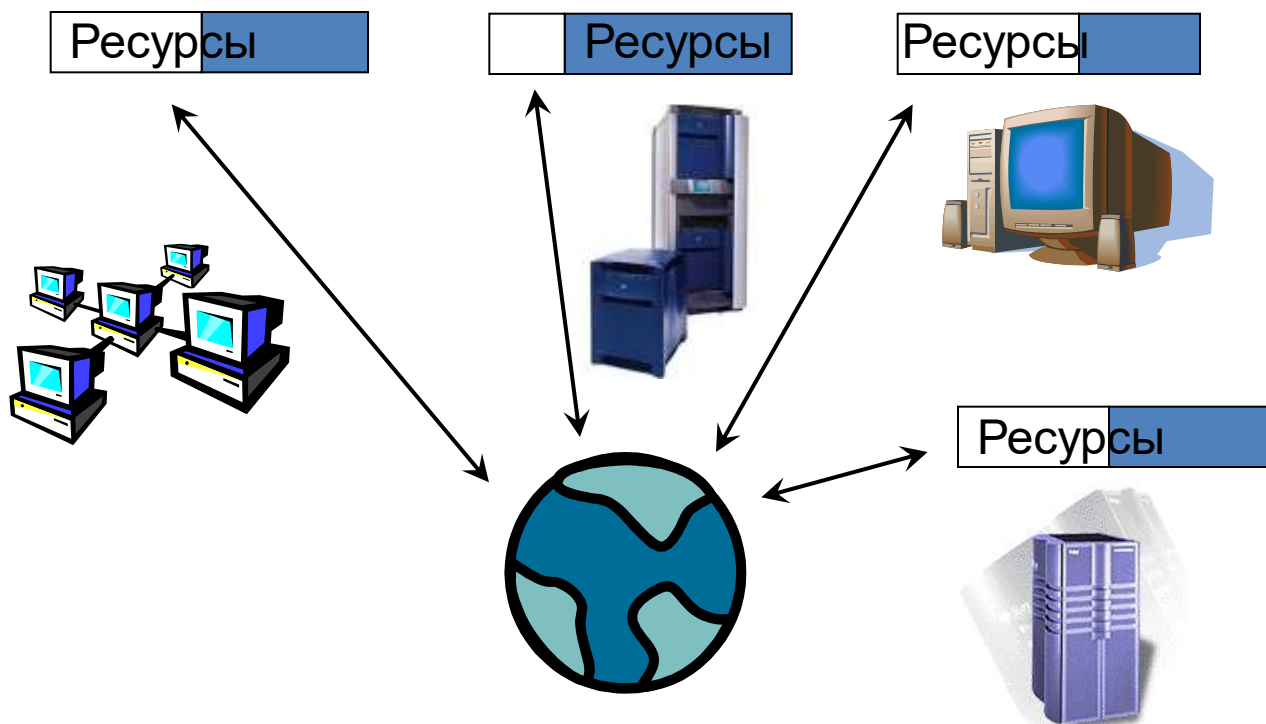
*Выполнил
студент ВМ 42
Шалыгин А. О.*

Содержание:

1. Определение и назначение.
2. Свойства.
3. История.
4. Архитектура.
5. Основные определения.
6. Сервисы.
7. Безопасность.
8. Области применения.
9. Литература.

Грид (англ. grid — решетка, сеть) —

согласованная, открытая и стандартизованная компьютерная среда, которая обеспечивает гибкое, безопасное, скоординированное разделение вычислительных ресурсов и ресурсов хранения информации, которые являются частью этой среды, в рамках одной виртуальной организации.



К прикладным задачам GRID относятся:

- сложное моделирование;
- совместная визуализация очень больших наборов научных данных;
- распределенная обработка в целях анализа данных;
- связывание научного инструментария с удаленными компьютерами и архивами данных.

Грид вычисления —

это форма распределённых вычислений, в которой «виртуальный суперкомпьютер» представлен в виде кластера соединённых с помощью сети, слабосвязанных компьютеров, работающих вместе для выполнения огромного количества заданий (операций, работ).

НАЗНАЧЕНИЕ GRID



**вычислительные
системы
(computational GRID)**



**хранение больших
массивов информации
(data GRID)**

GRID также называют распределенной информационно-вычислительной средой **РИВС**

Характерные свойства РИВС

- масштабы вычислительного ресурса
- гетерогенность среды
- пространственное (географическое) распределение информационно-вычислительного ресурса
- объединение ресурсов, которые не могут управляться централизованно (в случае, если они не принадлежат одной организации);
- использование стандартных, открытых, общедоступных протоколов и интерфейсов;
- обеспечение информационной безопасности.

История

1985–1995 - Программа Национального Научного Фонда США «National Science Foundation (NSF) Supercomputer Centers».

Октябрь 1997 - NSF инициировал новую программу развития информационных технологий – Partnerships for Advanced Computational Infrastructure (PACI).

1998 - Создан (и успешно развивается) инструментальный пакет Globus Toolkit.

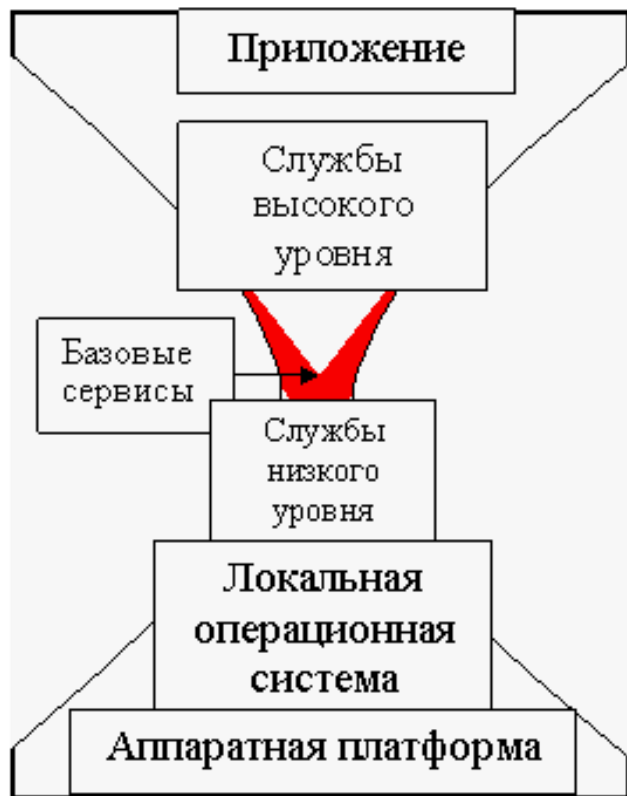
1999 - Сформировалось (и активно действует) международное научное Грид-сообщество – Global Grid Forum (GGF)

2002- GGF и IBM была представлена новая системная разработка – Open Grid Service Architecture (OGSA).

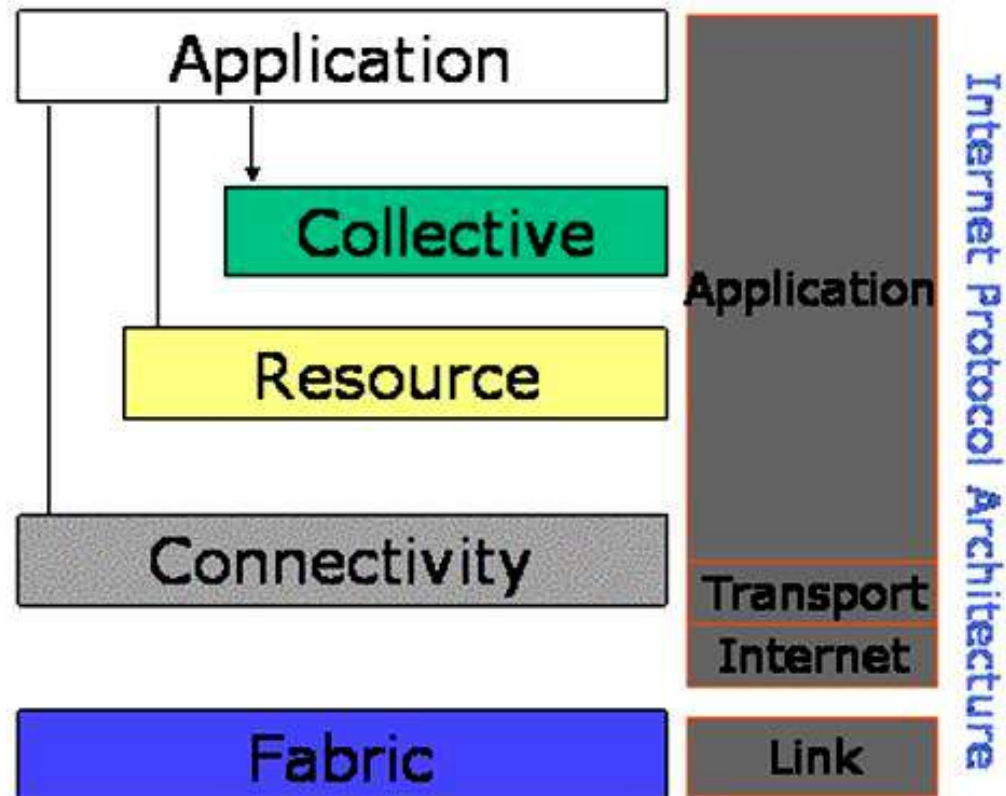
2003 - Создано объединение Enterprise Grid Alliance (EGA)

2006 - GGF и EGA объявили о слиянии и образовании Open Grid Forum (OGF)

Архитектура GRID

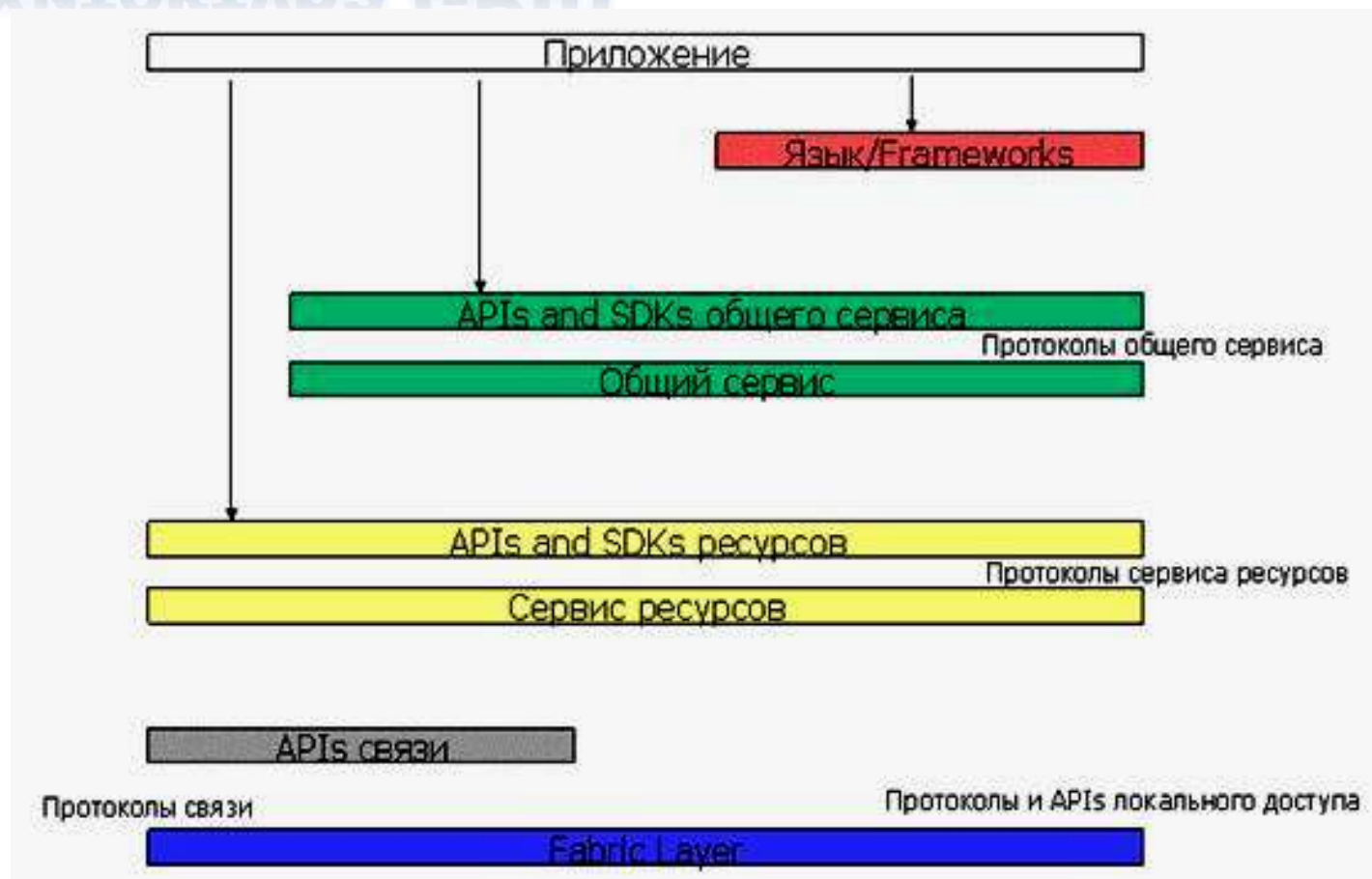


Модель песочных часов



Многоуровневая архитектура GRID

Архитектура GRID



Взаимодействие прикладной программы с различными уровнями GRID-системы

Основные определения

РЕСУРС - вычислительные ресурсы, системы хранения, каталоги, сетевые ресурсы.

Ресурсы могут быть разделены на физические и логические. К физическим относятся: оперативная память, память на долговременных носителях, количество и производительность процессоров и т.д. Примерами логических ресурсов являются распределенная файловая система, компьютерный кластер, распределенный пул компьютеров.

СЕТЕВОЙ ПРОТОКОЛ - это формальное описание форматов сообщений и набор правил, определяющий обмен сообщениями. Различают сетевые протоколы нижнего (Ethernet и др.), среднего (IP, TCP и др.) и высокого уровня (FTP, HTTP и др.).

Основные определения

СИНТАКСИС - это правила, определяющие порядок и форму записи информации в сообщении.

СЕРВИС — это сущность, которая предоставляет специфическую функциональность.

ИНТЕРФЕЙС ПРИКЛАДНЫХ ПРОГРАММ (API) представляет собой набор функций (сервисов), позволяющих прикладной программе осуществлять доступ к ресурсам через обслуживающую операционную систему.

Сервисы GRID

- идентификация выполняемой программы
- авторизация пользователя
- поиск ресурсов
- описание ресурсов
- резервирование ресурсов
- выполнение распределённых алгоритмов
- доступ к удалённым данным
- распределение ресурсов
- обнаружение неполадок

Безопасность



специализированное программное обеспечение - middleware

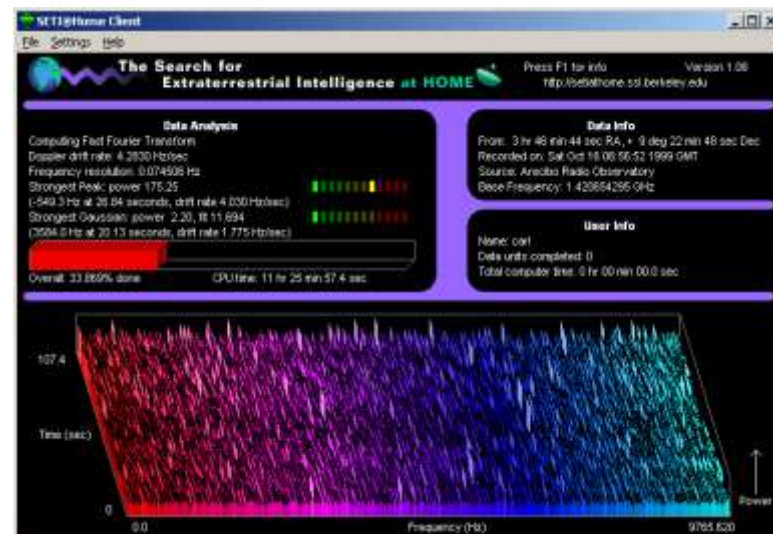
Протоколы защиты

- **Kerberos. Kerberos** – IETF стандарт, который поддерживает безопасность системы через установление подлинности.
- **TLS. TLS** (Transport Layer Security) (исходно известный как SSL) – IETF стандарт для установления подлинности, целостности и конфиденциальности сообщения
- **CMS. CMS** (Cryptographic Message syntax) - стандарт IETF определяет синтаксис, который позволяет в цифровой форме подписать, подтвердить подлинность, или зашифровать произвольные сообщение
- **GSS-API. GSS-API** (Generic Security Service API) - стандарт IETF, который определяет интерфейс прикладных программ

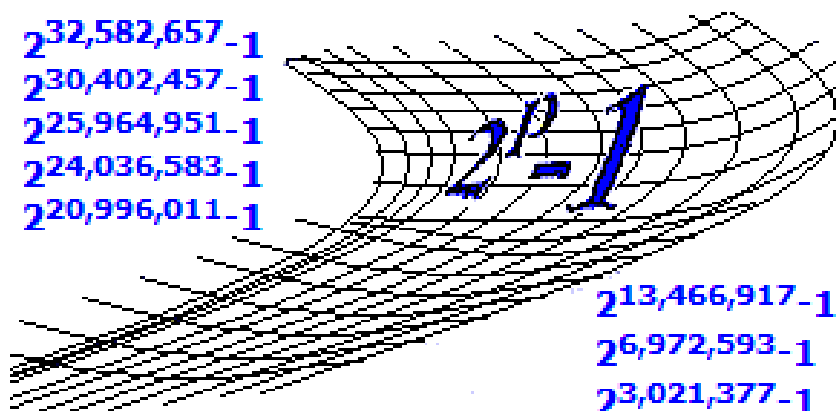
Области применения

- распределенные высокопроизводительные вычисления, решение очень крупных задач, требующих максимальных процессорных ресурсов, памяти и т.д.;
- «высокопоточные» вычисления, позволяющие организовать эффективное использование ресурсов для небольших задач, утилизируя временно простаивающие компьютерные ресурсы;
- проведение крупных разовых расчетов;
- вычисления с привлечением больших объемов распределенных данных, например, в метеорологии, астрономии, физике высоких энергий;
- коллективные вычисления: одновременная работа нескольких взаимодействующих задач разных пользователей.

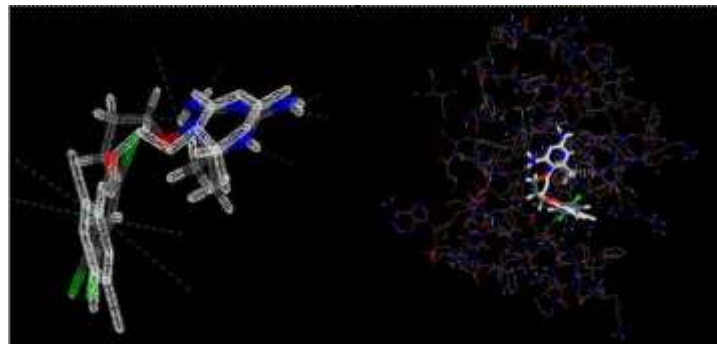
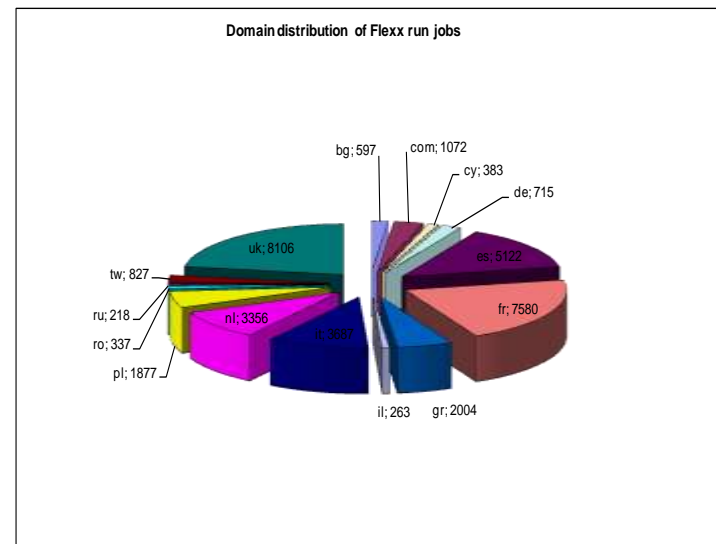
- Поиск следов внеземных цивилизаций
- Обработка данных, полученных радиотелескопом Аресибо
- Около 5 млн. участников
- 1200 CPU лет в день
- Постоянная вычислительная мощность ~34 TF (примерно такая, какая достигнута в Симуляторе Земли в Японии)
- Высокая степень гетерогенности ресурсов - >77 различных типов процессоров



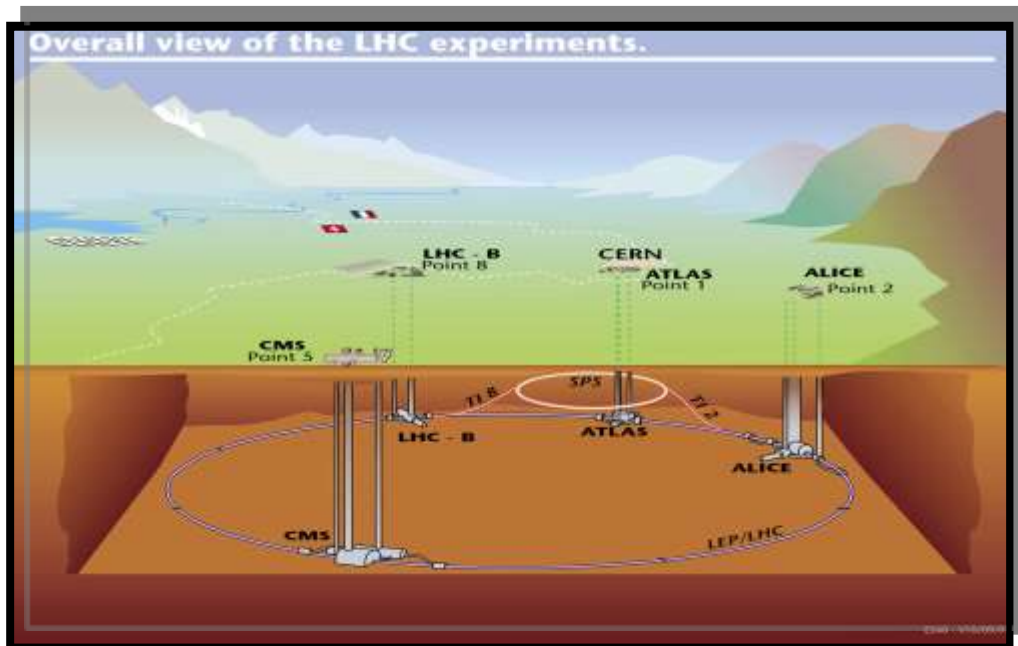
- Поиск простых чисел Мерсенна. Числа Мерсенна имеют вид $M_p = 2^p - 1$, где p -простое
- Самое большое известное на данный момент простое число $M_{43112609} = 2^{43112609} - 1$ было найдено в рамках проекта GIMPS в августе 2008 года. Оно состоит из 12,978,189 цифр!!
- Ресурсы (на 10.06.2009):
 - команд – 216
 - участников – 20888
 - CPUs – 115601
 - мощность - 38.497 терафлоп



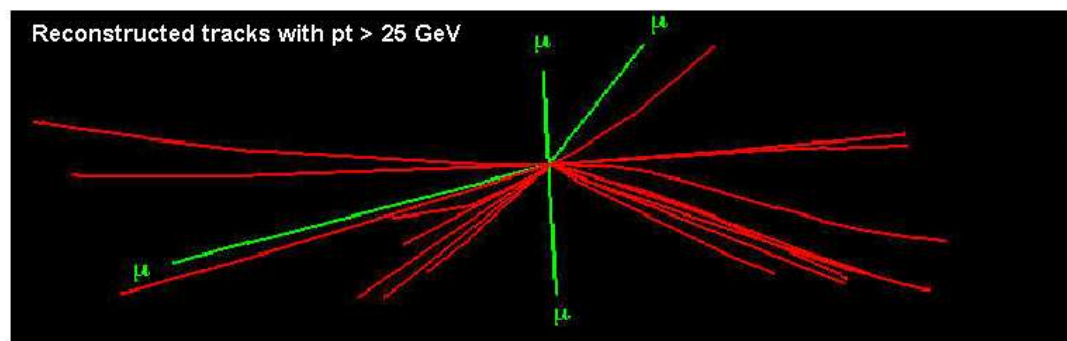
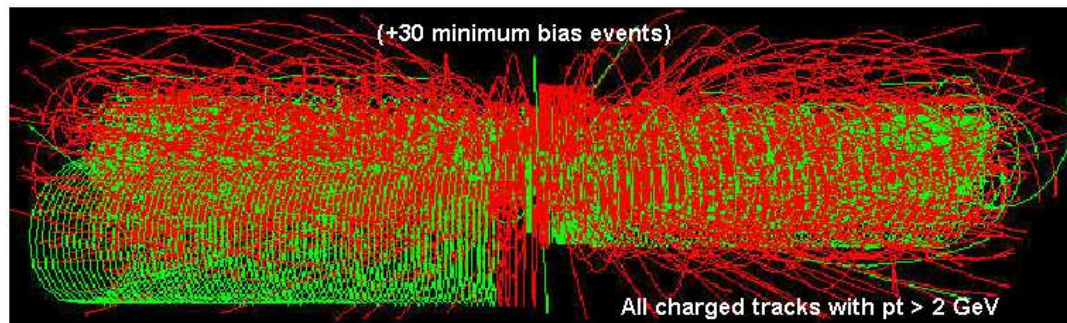
- Приложение Drug Discovery, позволяющее вычислять вероятность прямого контакта между потенциальным лекарством и белком-мишенью
- Первый в истории биомедицины сеанс массовой обработки данных (малярия)
 - Исследовано 46 миллионов посадочных лиганд
 - Получено более 1 Тб данных
 - Использованы ~1000 компьютеров из 15 стран, что составляет ~ 80 машино/лет
 - Средний фактор ускорения – 600
- Второй сеанс (птичий грипп)
 - Использованы ~5000 компьютеров из 27 стран, что составляет ~ 420 машино/лет
 - Получено более 2 Тб данных
 - Средний фактор ускорения – 2000



Обработка данных БАК
производится с
помощью GRID
вычислений.



Селективность (степень отбора) ~ 1 из 10^{13}



Литература:

- <http://ru.wikipedia.org/wiki/GRID>
- <http://jre.cplire.ru/win/dec03/4/text.html>
- <http://www.gridclub.ru>
- <http://egee.pnpi.nw.ru/> - ПИЯФ РАН