Complementi di Piattaforme Abilitanti Distribuite

Distributed Enabling Platforms | |

Topics

- State-of-the-art technologies to dealing with large scale problems
  - Frontier research in many different fields today requires world-wide collaborations
  - Batch analysis of gazillion-bytes of experimental data
  - Crawling, indexing, searching the Web
  - Web 2.0 applications
  - Online analysis of gazillion-bytes of usage data
- Grid and Cloud Platforms
  - Resource Management
  - Information Management
  - Data Management
  - System Virtualization
  - Cost Analysis
  - Data Analysis
  - Programming

Course Organization

- ~~48 hours: ~32 lessons, ~16 laboratory~~
- 36 hours: ~24 lessons, ~12 laboratory
- Timetable
  - Monday 14:00-16:00 Room 10B
  - Wednesday 17:00-19:00 Room 10B
- Highly interactive lectures
- Laboratory
  - Java programming skills required
- Notes and references available online
  - Updated in real time on the course wiki
- Grading
  - notes (20%)
  - project (50%)
    - To be agreed with teacher
  - oral session (30%)

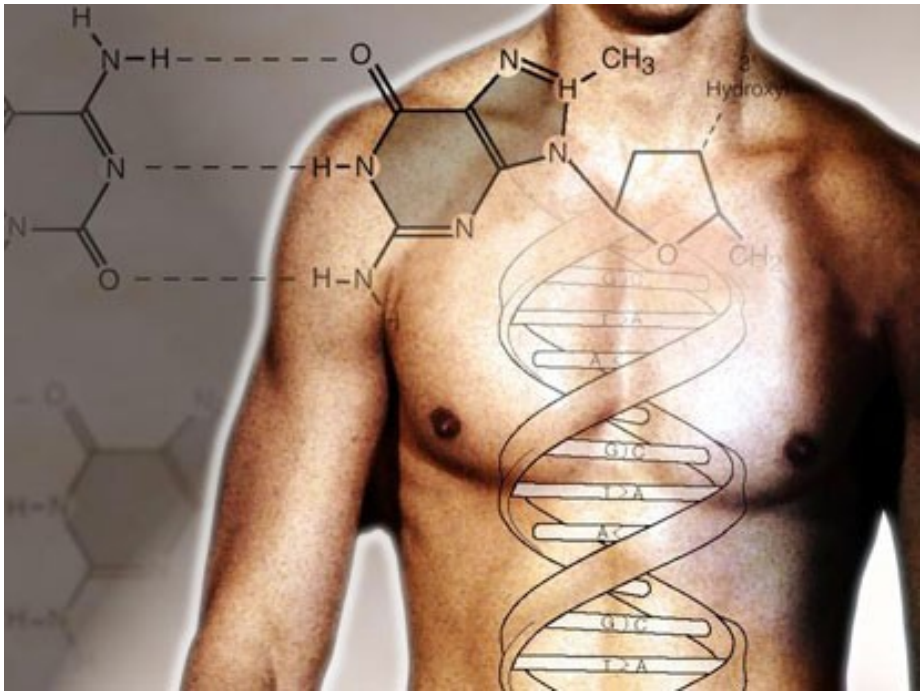
- Distributed…
  - relating to a computer **network** in which at least some of the **processing** is done by the individual computers and **information** is **shared** by and often **stored** at the computers
- Enabling…
  - to make **possible**, **practical**, or **easy**
- Platforms…
  - the computer **architecture** and **equipment** used for a **particular purpose**

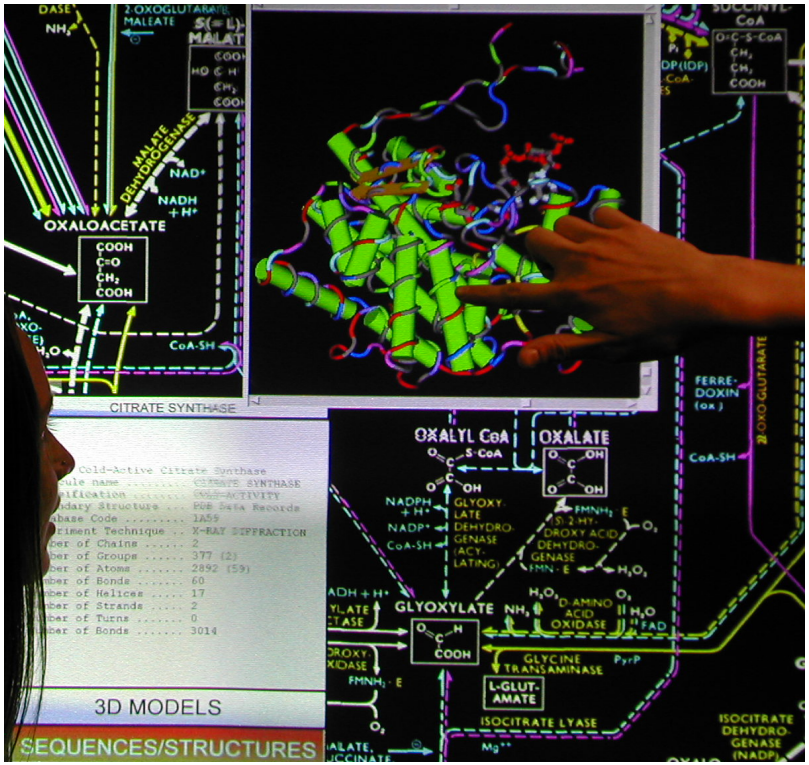
TO DO WHAT?

Large Scale Problems

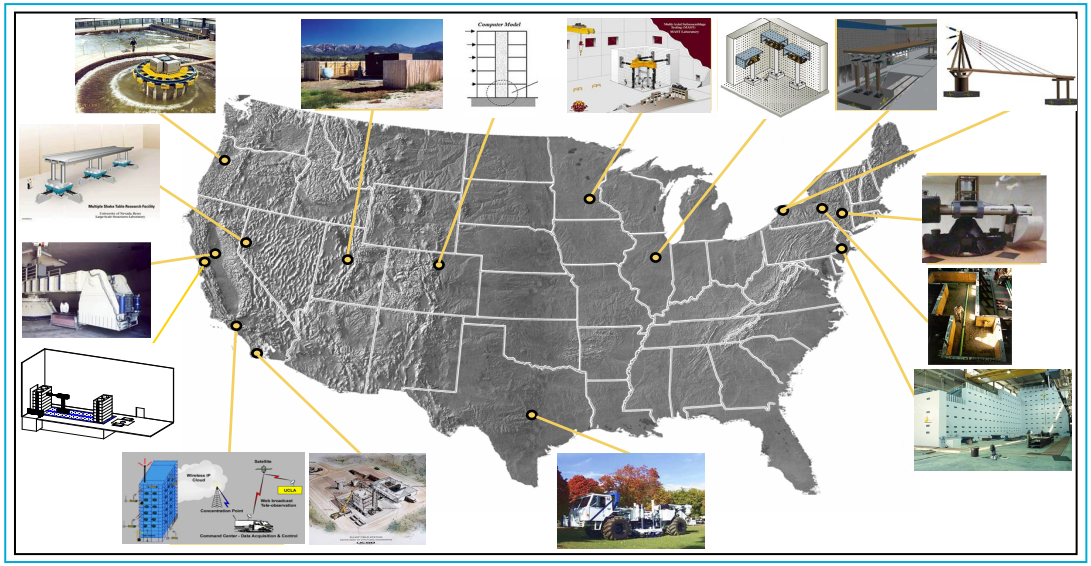
- In research
  - Frontier research in many different fields today requires world-wide collaborations
  - Online access to expensive scientific instrumentation
  - Scientists and engineers will be able to perform their work without regard to physical location
  - Simulations of world-scale mathematical models
  - Batch analysis of gazillion-bytes of experimental data
- In production
  - Crawling, indexing, searching the Web
  - Web 2.0 applications
  - Mining information
  - Highly interactive applications
  - Online analysis of gazillion-bytes of usage data

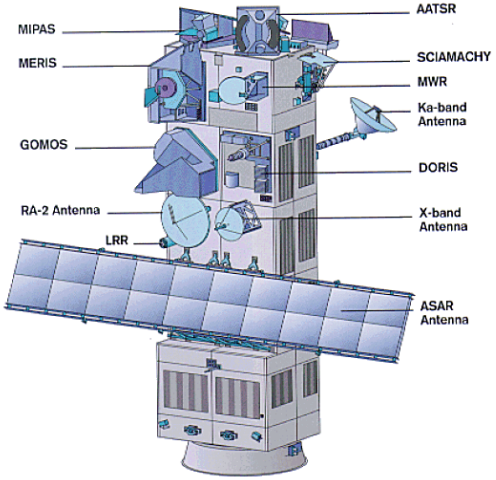
Biology

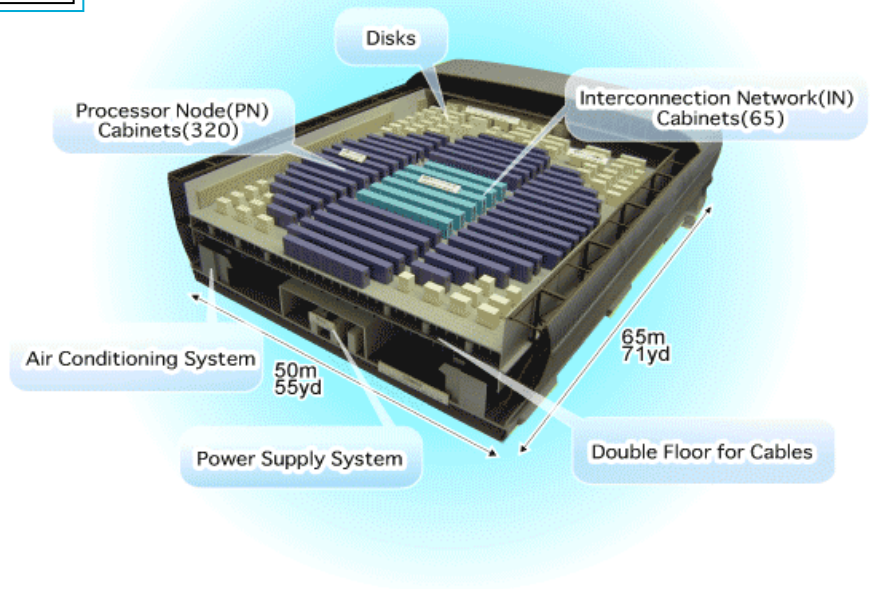



OXALOACETATE
CITRATE SYNTHASE
3D MODELS
SEQUENCES/STRUCTURES
OXALYL CoA
OXALATE
GLYOXYLATE
L-GLUT-AMATE

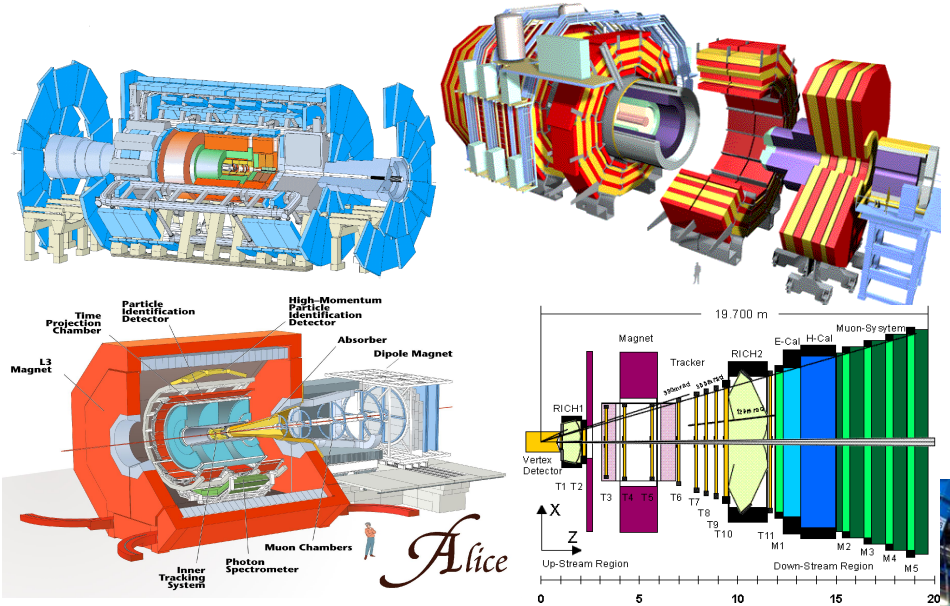
Earth Science

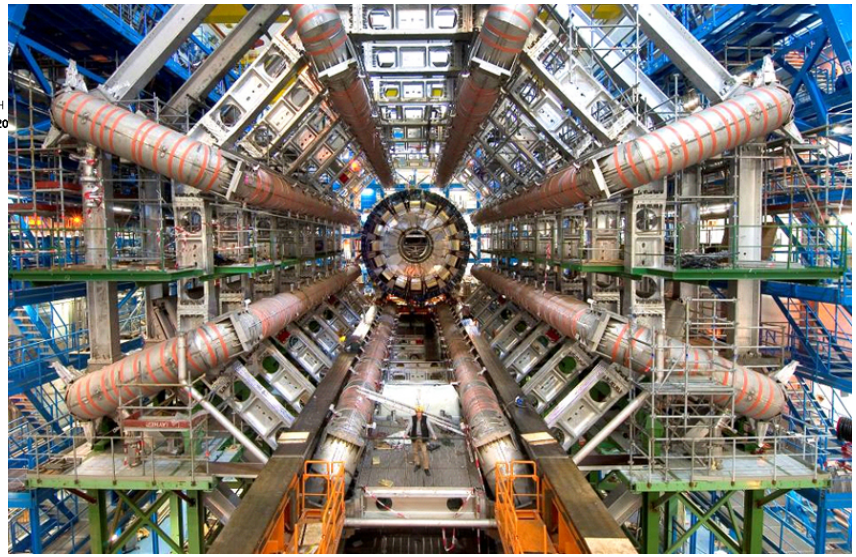




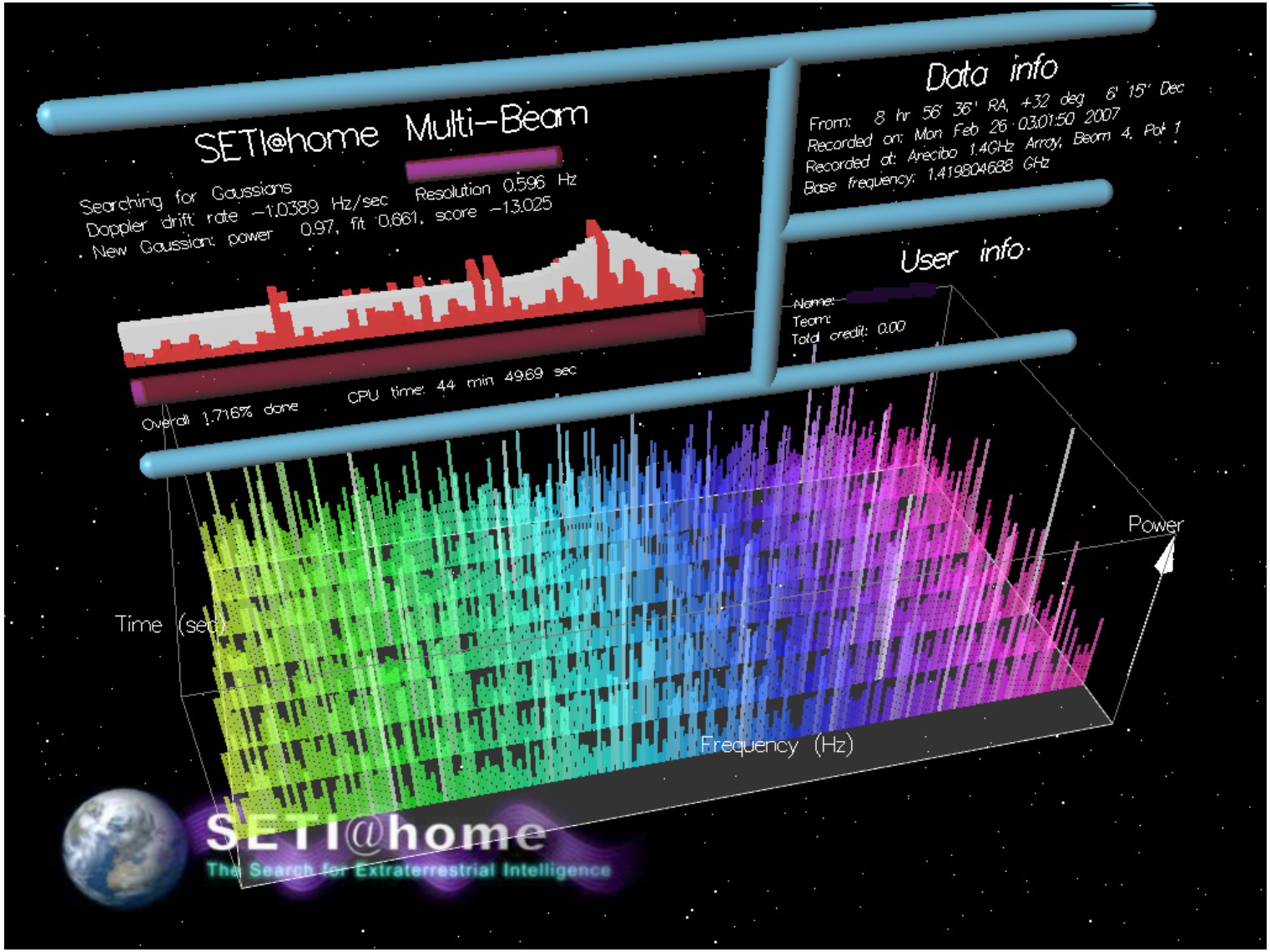


Physics

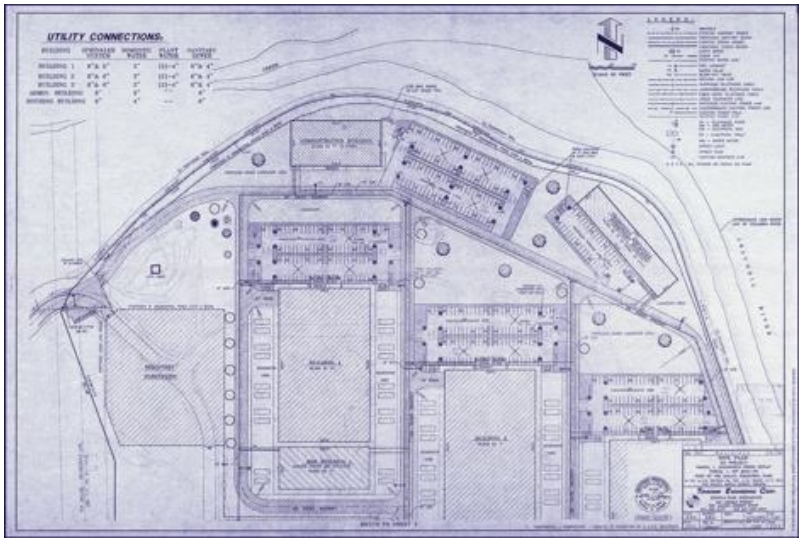

Time Projection Chamber
Particle Identification Detector
High-Momentum Particle Identification Detector
Absorber
Dipole Magnet
L3 Magnet
Muon Chambers
Photon Spectrometer
Inner Tracking System
Alice
19.700 m
Magnet
Tracker
RICH1
RICH2
E-Cal
H-Cal
Muon-System
Vertex Detector
T1 T2
T3
T4 T5
T6
T7
T8
T9
T10
T11
M1
M2
M3
M4
M5
X
Z
Up-Stream Region
Down-Stream Region
0
5
10
15
20

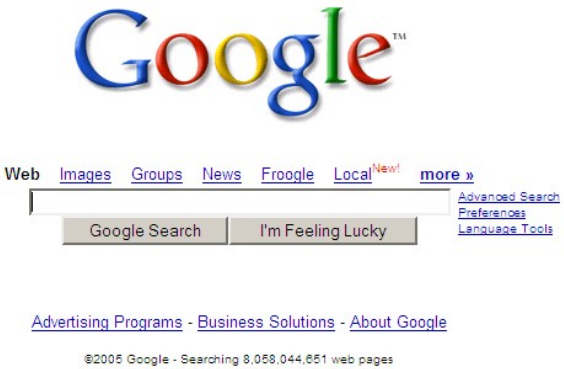


Astronomy


SETI@home Multi-Beam
Searching for Gaussians
Doppler drift rate −1.0389 Hz/sec Resolution 0.596 Hz
New Gaussian: power 0.97, fit 0.661, score −13.025
Overall 1.716% done
CPU time: 44 min 49.69 sec
Data info
From: 8 hr 56' 36" RA, +32 deg 6' 15" Dec
Recorded on: Mon Feb 26 03:01:50 2007
Recorded at: Arecibo 1.4GHz Array, Beam 4, Pol 1
Base frequency: 1.419804688 GHz
User info
Name:
Team:
Total credit: 0.00
Power
Time (sec)
Frequency (Hz)
SETI@home
The Search for Extraterrestrial Intelligence

Google


UTILITY CONNECTIONS.


Google
Web Images Groups News Froogle Local New! more »
Advanced Search
Preferences
Language Tools
Google Search
I'm Feeling Lucky
Advertising Programs - Business Solutions - About Google
©2005 Google - Searching 8,058,044,651 web pages

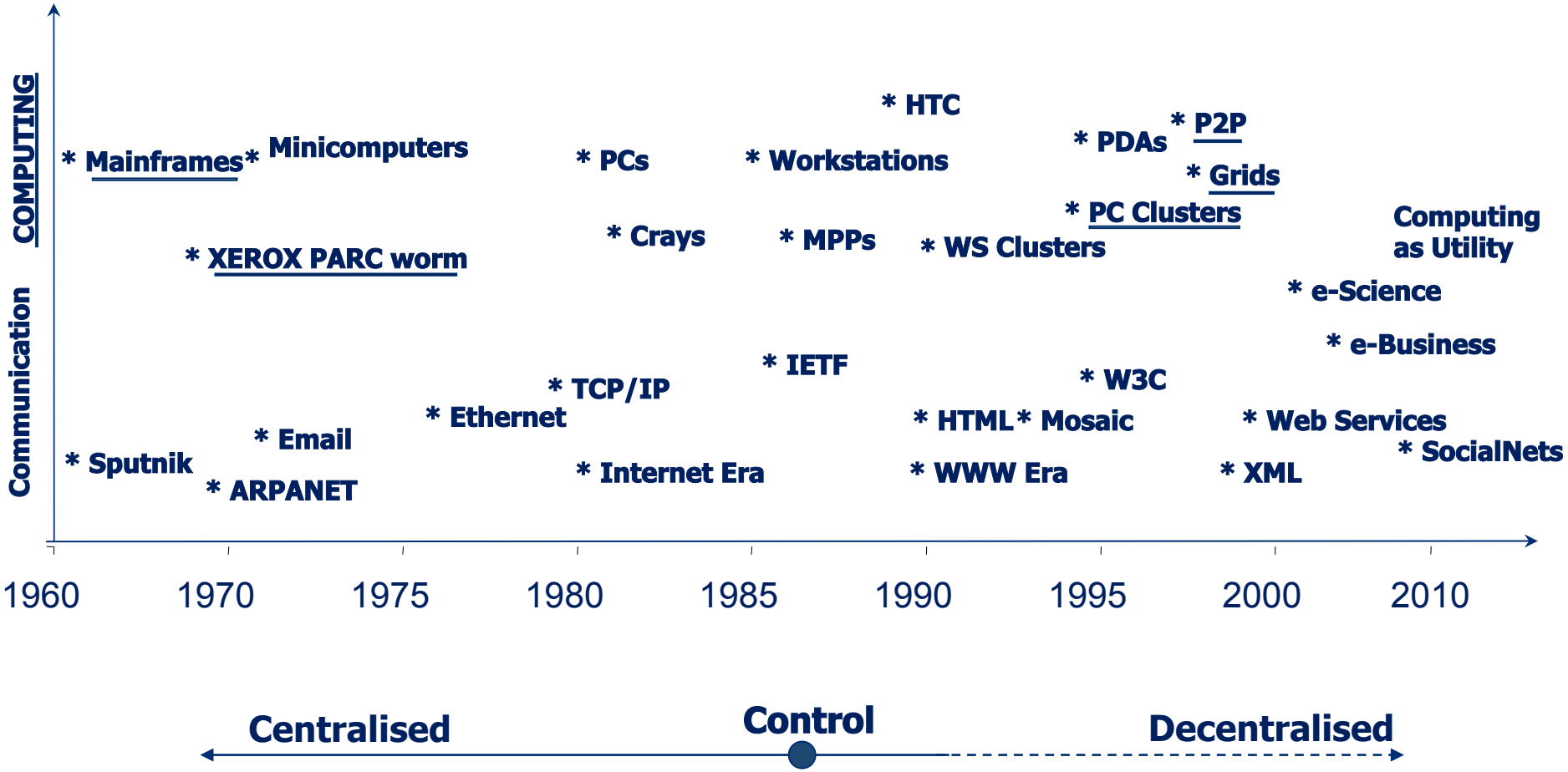
Big enough?

- Large Hadron Collider:
  - 10^{19} bytes/year generated
  - 10^{21} bytes/year forecasted
  - 10^{3} scientists
  - 10^{2} institutions
- Large Synoptic Survey Telescope (2016)
  - 15 TB/night
  - 6.8 PB/year
- Google
  - 10^{19} byte/day processed
  - 0.1 sec query latency
- Walmart
  - 6000 stores, 267 M items/day

Our Data Driven World

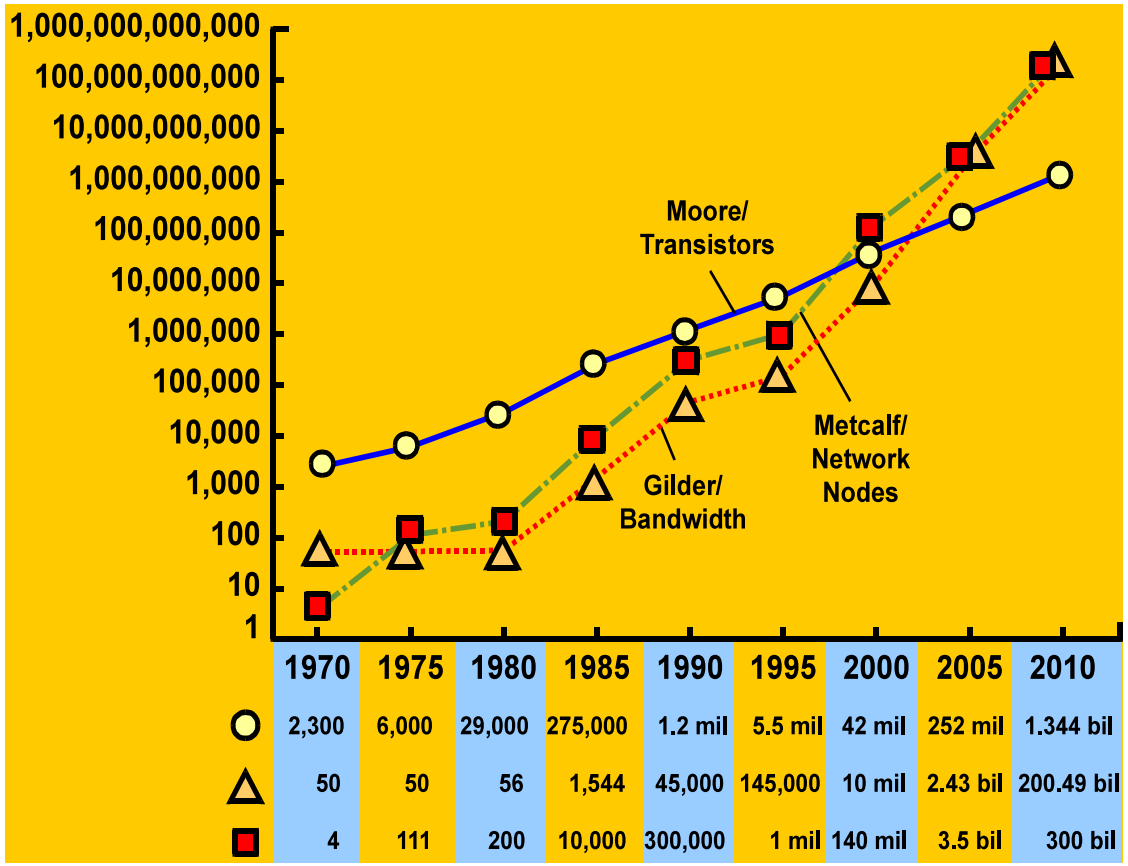
- Science
  - Databases for astronomy, genomics, natural languages, seismic modeling, …
- Humanities
  - Scanned books, historic documents, …
- Commerce
  - Corporate sales, stock market transactions, census, airline traffic, …
- Entertainment
  - Hollywood movies, Internet images, MP3 music, …
- Medicine
  - Patient records, drugs composition, …

Computing and Communication Technologies Evolution: 1960-2010!



Performance, Capability, Value of ICT as defined by the three Laws of Computing

- **Moore's Law.**
  - Transistors on a single chip doubles ~ every 18 months.
- **Gilder's Law.**
  - Aggregate bandwidth triples ~ every year.
- **Metcalfe's Law.**
  - The value of a network may grow exponentially with the number of participants.



	1970	1975	1980	1985	1990	1995	2000	2005	2010
○	2,300	6,000	29,000	275,000	1.2 mil	5.5 mil	42 mil	252 mil	1.344 bil
△	50	50	56	1,544	45,000	145,000	10 mil	2.43 bil	200.49 bil
□	4	111	200	10,000	300,000	1 mil	140 mil	3.5 bil	300 bil

Source: Cambridge Energy Resource Associates

Experiment

- You must put together your computers to calculate 10^{20} prime numbers. How do you proceed?
  - You agree to collaborate
  - You put your computers in a network
  - You install the programs
  - You run the programs
  - You wait for results
  - You publish your results on the Web
- Is really that simple?

What if…

- I do not trust someone else's computer?
- I do not trust the application?
- I want to use my laptop during lectures?
- The application wants more computers?
- I forget the IP address of some computers?
- My disk disintegrates losing the data?
- Someone pays and we must share money?
- We are still waiting the results after the class?

NOT SO SIMPLE!

Some issues

- Security
- Resource sharing
- Dynamicity
- Lack of information
- Lack of global state
- Fault tolerance
- Accounting
- …

How to solve a problem?

- Manual Computing
- Personal Computing
- Mobile Computing
- Ubiquitous Computing
- Pervasive Computing
- Parallel Computing
- Distributed Computing
- High Performance Computing
- …
- **Grid Computing**
- **Cloud Computing**