

Piattaforme Abilitanti Distribuite - PAD -

Distributed Enabling Platforms

Nicola Tonellotto
(ISTI, CNR)
nicola.tonellotto@isti.cnr.it

Today

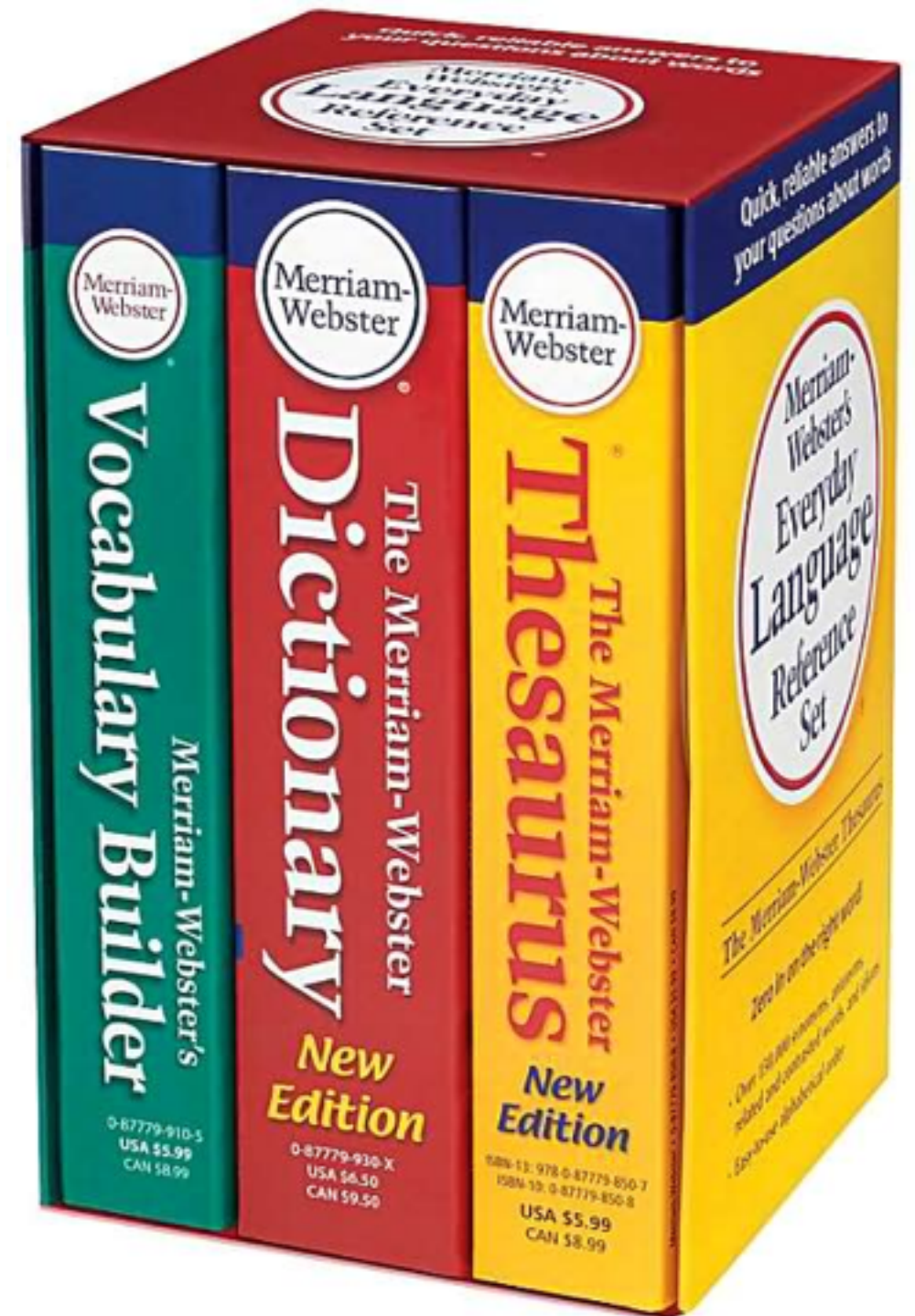


What?



What is the meaning of words?

- 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



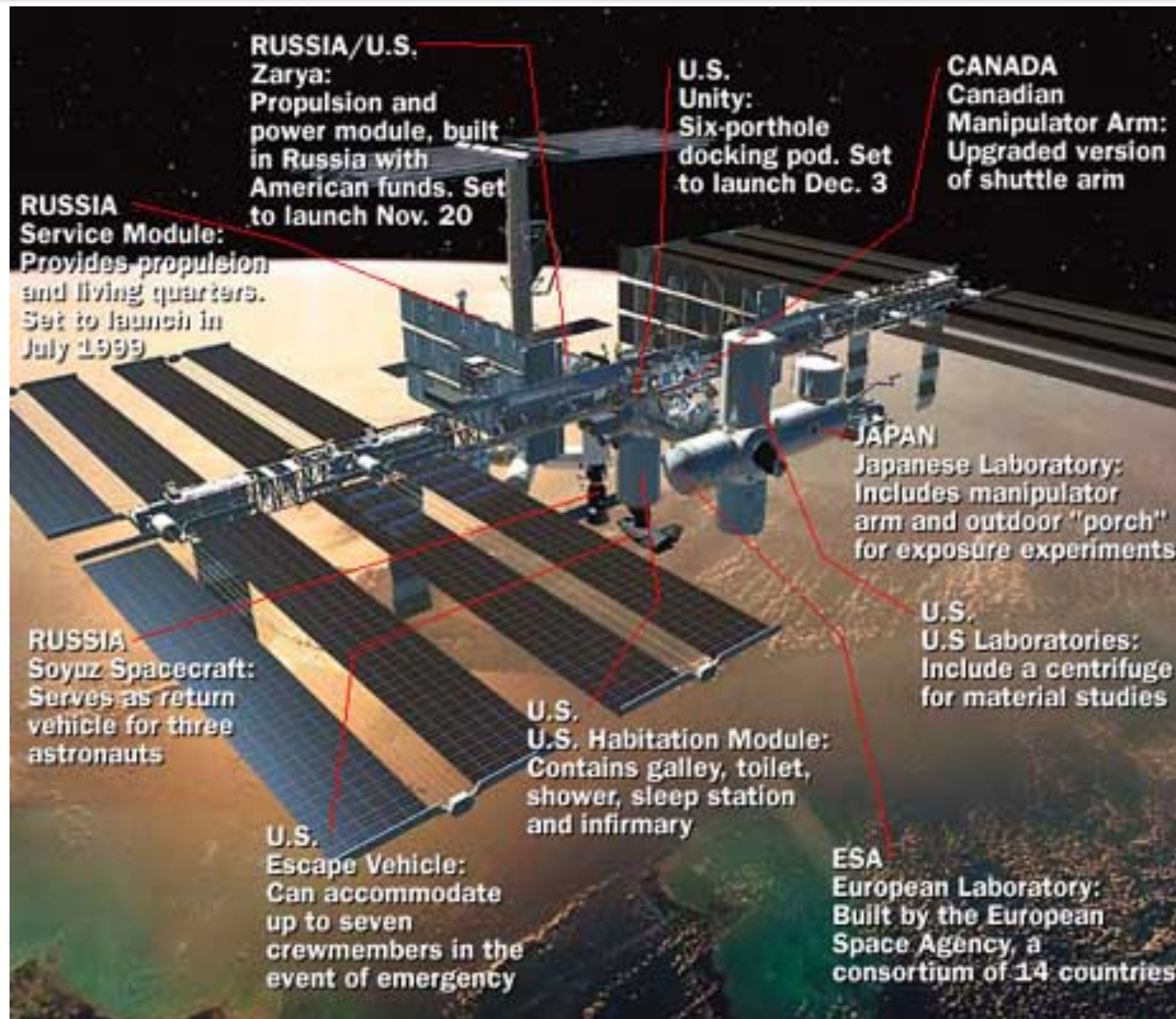
To do what?



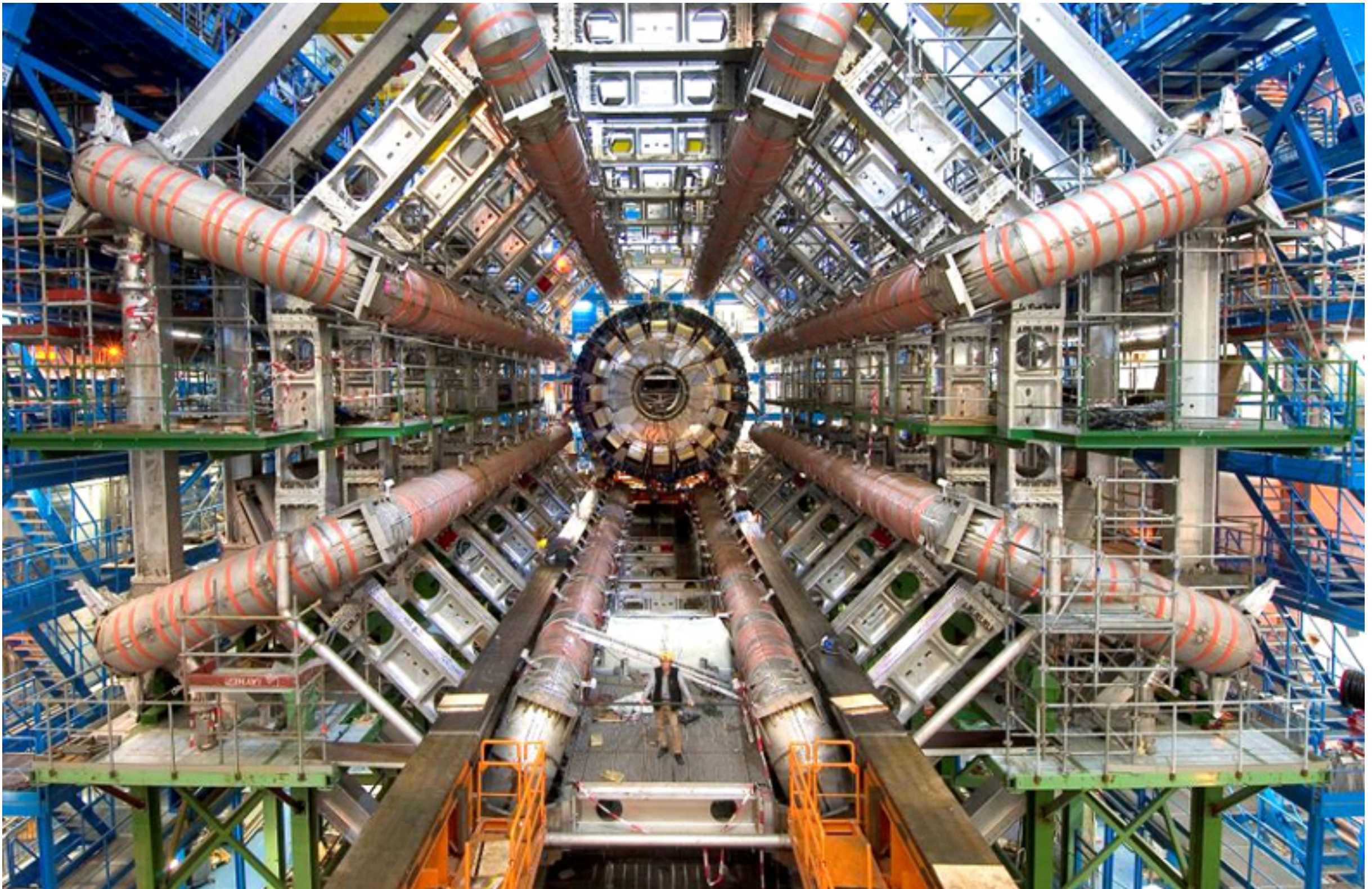
Solve large scale problems!

- 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 business
 - Crawling, indexing, searching the Web
 - Web 2.0 applications
 - Mining information
 - Highly interactive applications
 - Online analysis of gazillion-bytes of usage data

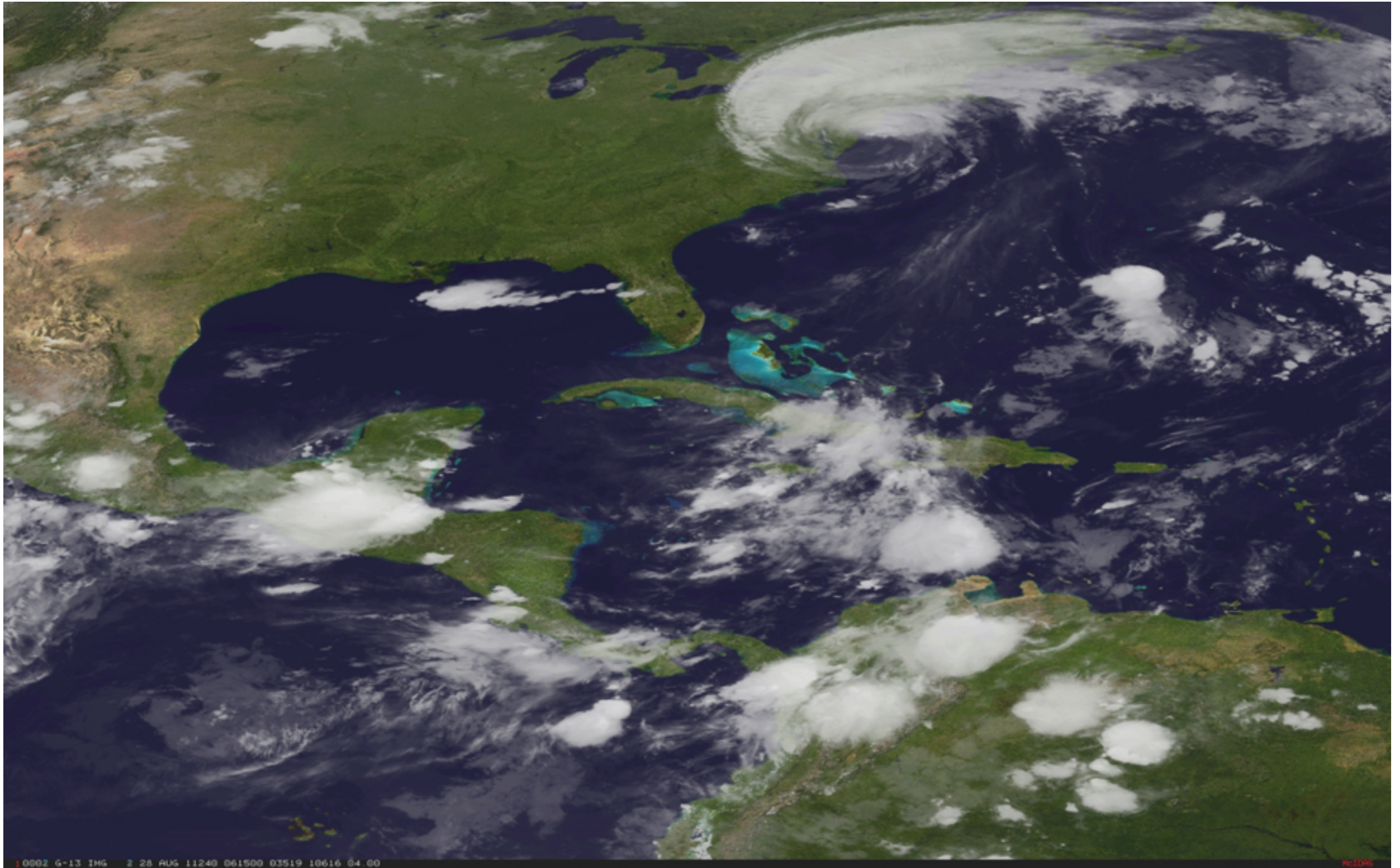
World-wide Collaborations



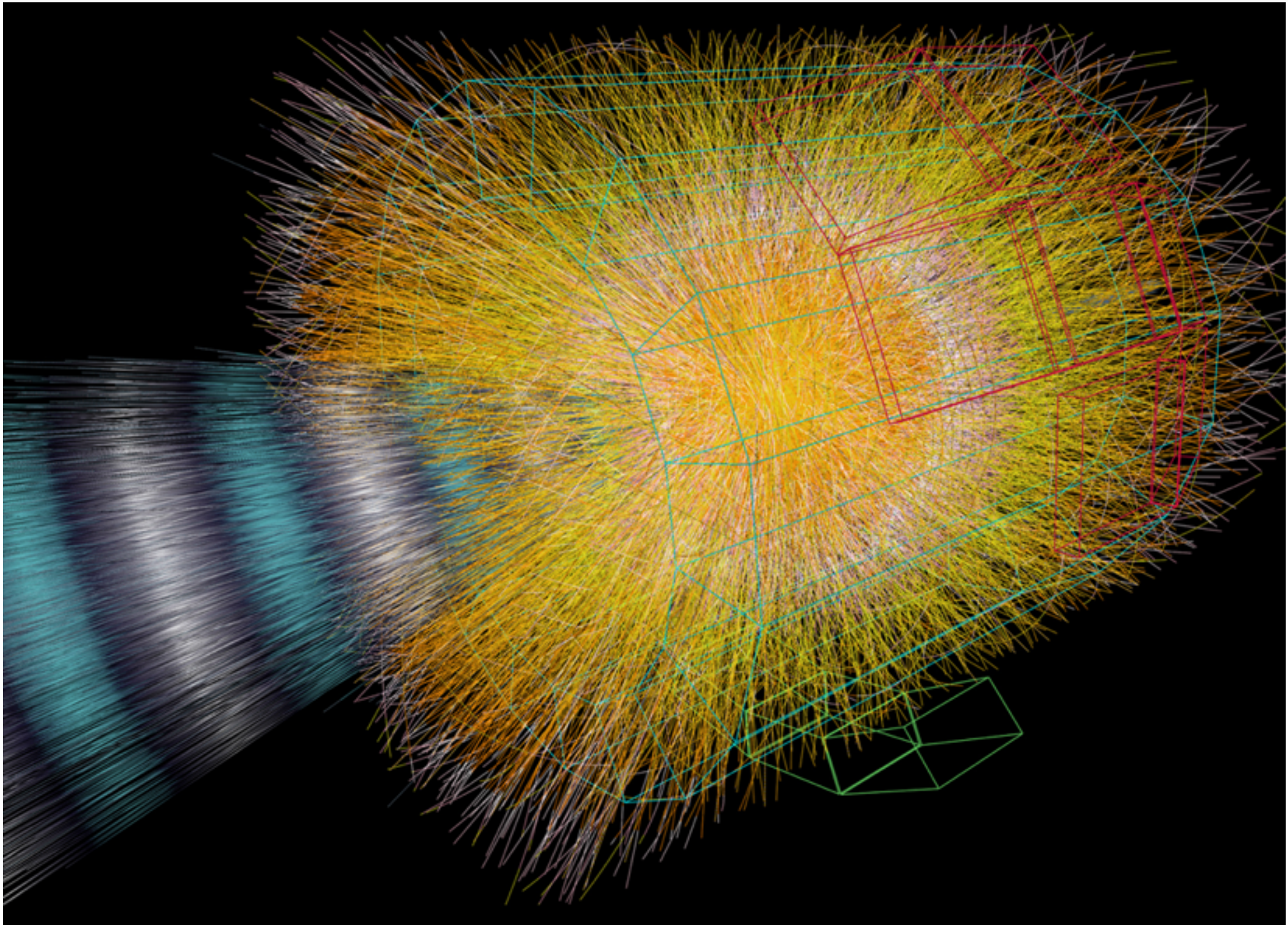
Expensive Scientific Instruments



World-scale Simulations



Batch analysis of huge data



Managing the Web



Web [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [more »](#)

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2005 Google - Searching 8,058,044,651 web pages

Web 2.0



licensed under Attribution-NonCommercial-ShareAlike 2.0 Germany | Ludwig Gatzke | <http://flickr.com/photos/stabilo-boss/>

Online analysis of huge data



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, ...

Big Enough?

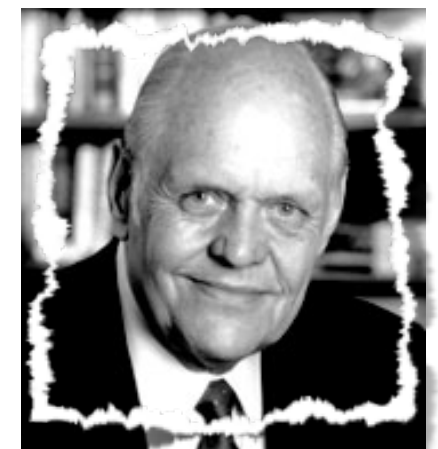
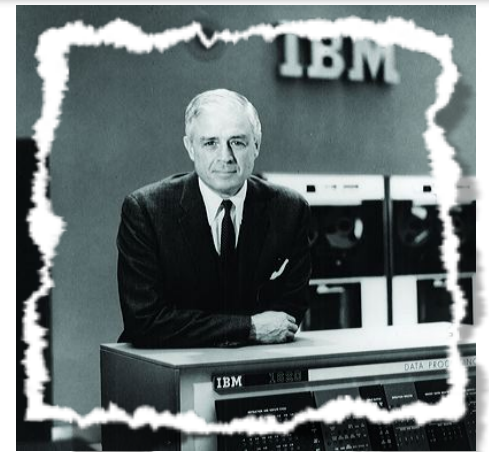
- Large Hadron Collider:
 - 10 EB/year generated
 - 1 ZB/year forecasted
 - 103 scientists
 - 102 institutions
- Large Synoptic Survey Telescope (2016)
 - 15 TB/night
 - 6.8 PB/year
- Google (2010)
 - 24 PB/day processed
 - 8 EB/day processed
 - 0.1 sec query latency
- Facebook (2009)
 - 15 TB/day user data
- eBay (2009)
 - 50 TB/day user data
- Walmart
 - 6000 stores, 267 M items/day

Data everywhere!



taken from: <http://now.sprint.com/nownetwork/>

Famous(?) predictions (I)

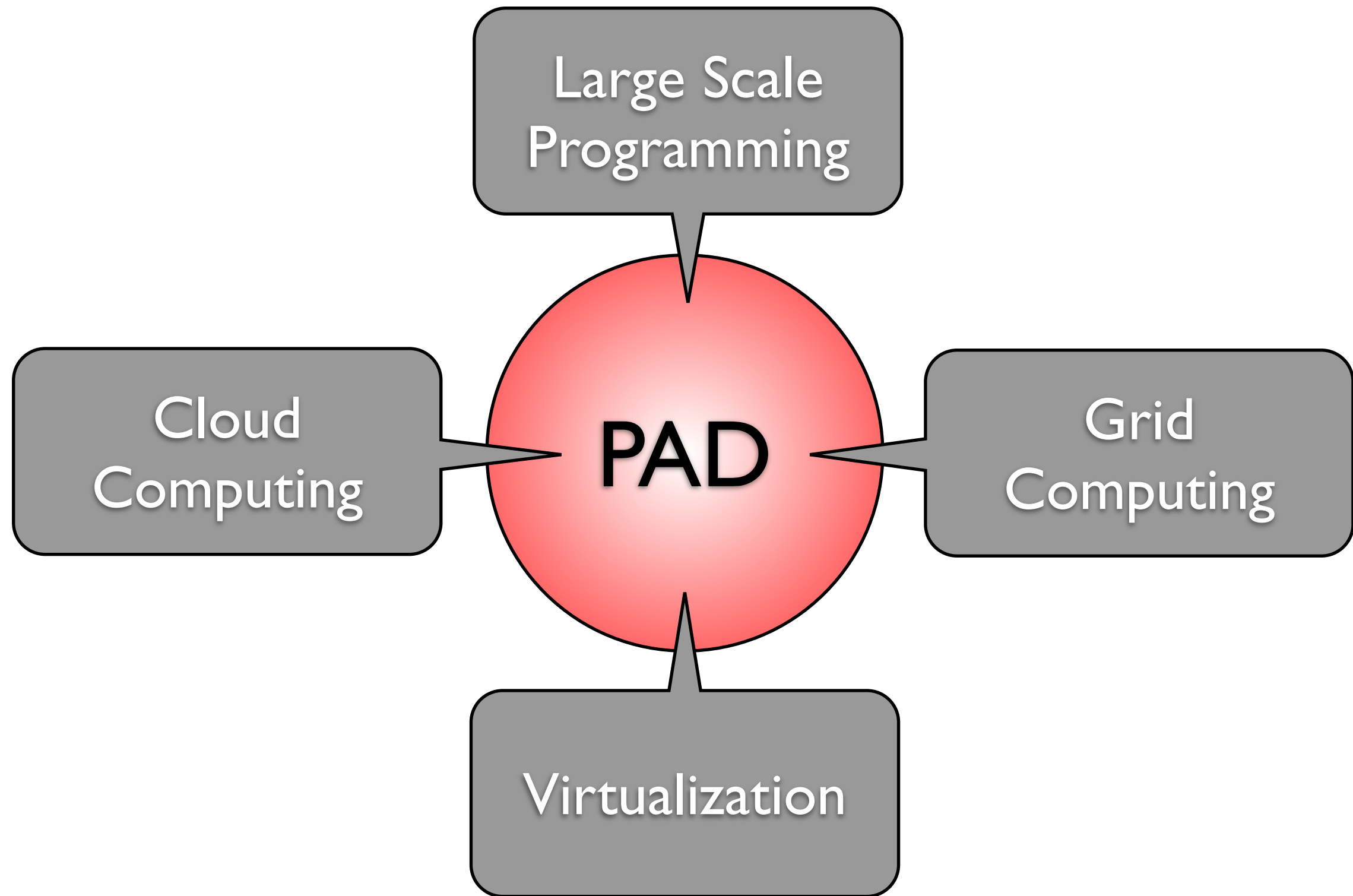


- "I think there is a world market for maybe five computers."
- Thomas Watson, chairman of IBM, 1943
- "I have travelled the length and breadth of this country and talked with the best people, and I can assure you that data processings is a fad that won't last out the year."
- The ed in charge of biz books for Prentice-Hall, 1957
- "There is no reason anyone would want a computer in their home."
- Ken Olson, president, chairman and founder of DEC, 1977

How?



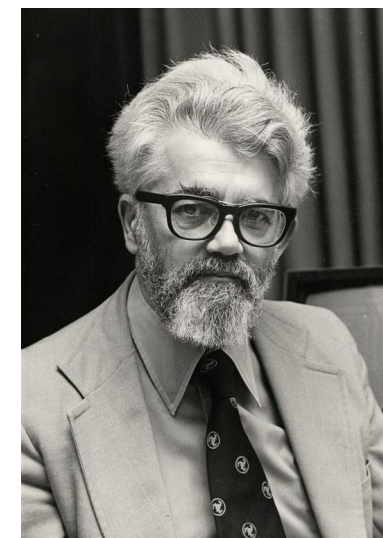
(not so?) Hot Technologies



Famous(?) predictions (II)

1961

[...] computing may someday be organized as a **public utility** just as telephone system is a public utility [...] the computer utility could become the basis of a new and important industry [...]



John McCarthy (1927-2011)
Turing Award (1971)
Artificial Intelligence

1969

As of now, computer networks are still in their infancy, but as they group up and become sophisticated, we will probably see the spread of **computer utilities** which, like present electric and telephone utilities, will service individual homes and offices across the country.



Leonard Kleinrock (1934)
Turing Award (1971)
Queueing Theory

The 5th Utility



Computing is being transformed to a model consisting of services that are commoditized and delivered in a manner similar to traditional utilities



Demand for more computing power

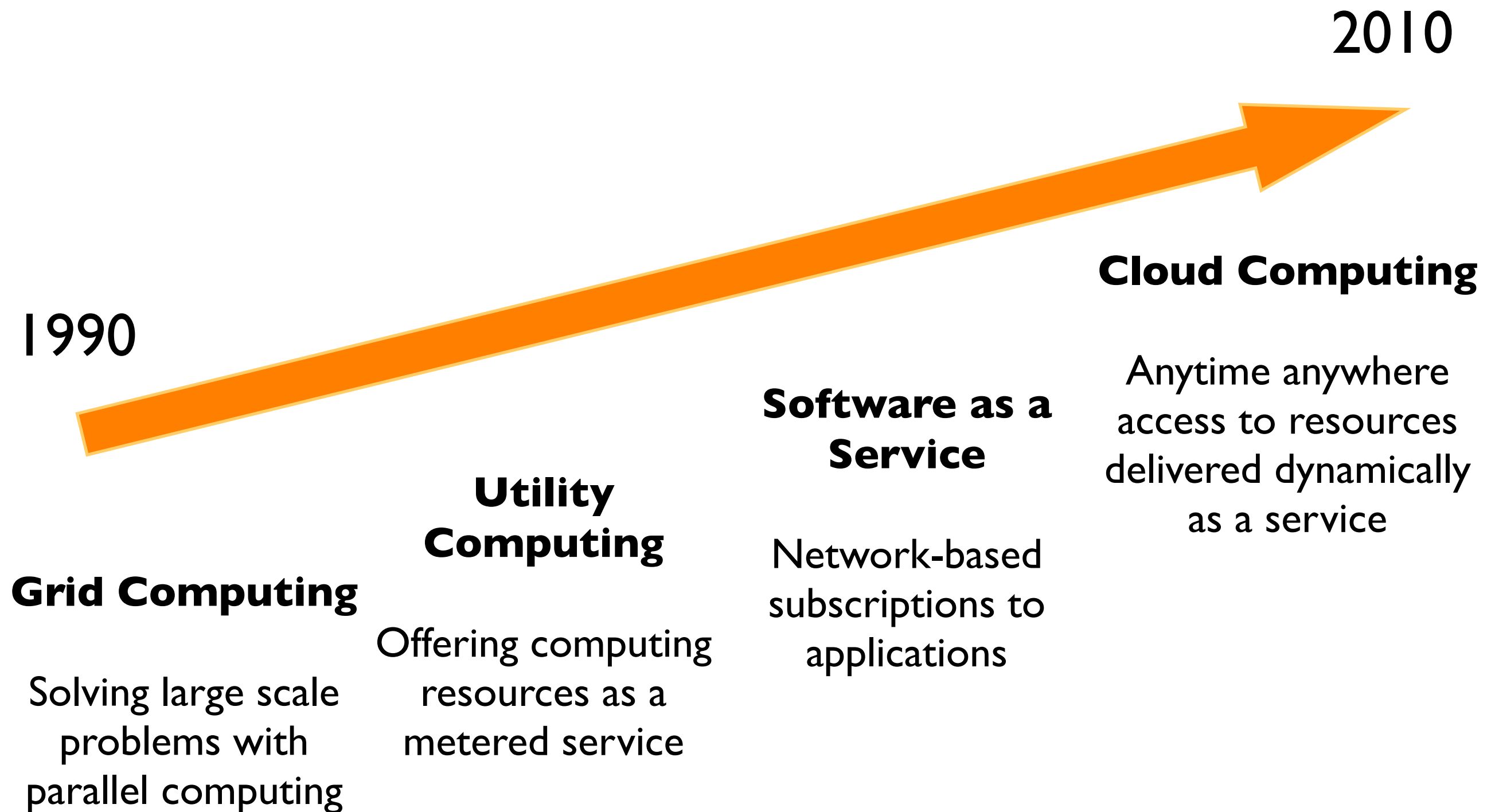
- There are three ways to improve performance:
 - Work smarter
 - Work harder
 - Get help
- In computing:
 - Using optimized algorithms and techniques
 - Using faster hardware
 - Using multiple computers

Cluster Computing

- A cluster is a type of parallel and distributed system, which consists of a collection of **inter-connected stand-alone computers** working together as a **single integrated computing resource**.
- Basic element is the **node**, a single or multiprocessor system with memory, I/O and OS
- Generally two or more nodes connected together
- In a single **rack**, or physically separated and connected via a LAN
- Appears as a single system to users and applications
- Specialized access, management and programming



Utility Computing History



Grid Computing

- Problem:

Scientific instruments and experiments provide huge amount of data

- Goal:

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

- Solution:

Networked data processing centers and "middleware" software as the "glue" of resources.

Once upon a time...



Microcomputer



Minicomputer



Cluster



Mainframe

...up to the Grid



Why not just distributed?

- Distributed applications already exist!
 - But they tend to be specialised system
 - Single purpose
 - Single User Group
- Grids go further!
 - Different kinds of resources
 - Different kinds of interactions
 - Dynamic nature
 - Multiple institutions

Key Concept

ability to negotiate resource-sharing arrangements among a set of participating parties (providers and consumers) and then to use the resulting resource pool for some purpose

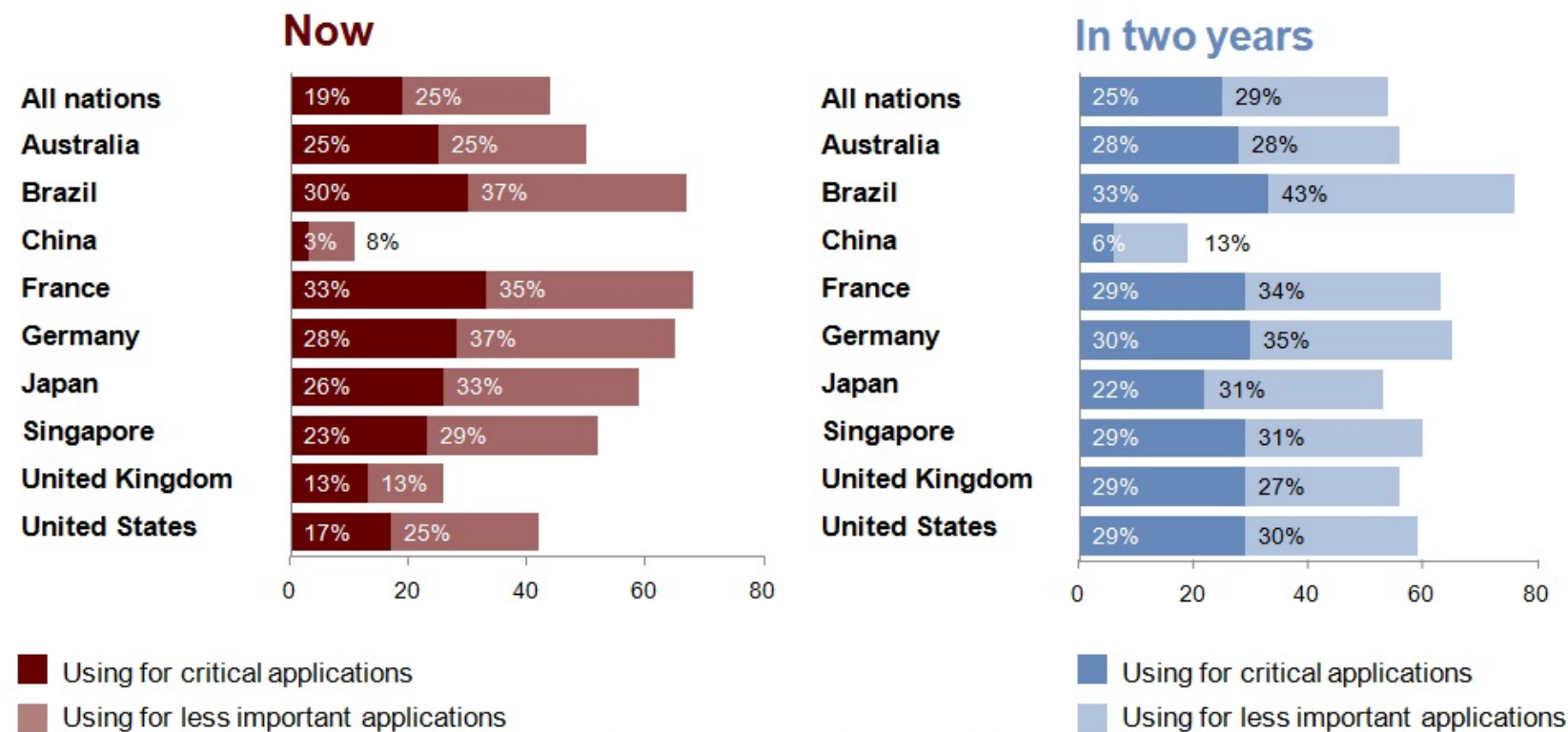
Grids in action

- High Energy Physics
 - European Data Grid
 - LHC Computing Grid
- Earth Observation
 - ESA EO Grid
 - Global Earth Observation Grid
- Bioinformatics
 - Genome Grid
- Mathematics
 - Zetagrid
- Geology
 - Earthquake Engineering Simulation
- Astronomy
 - SETI@home



Cloud Computing

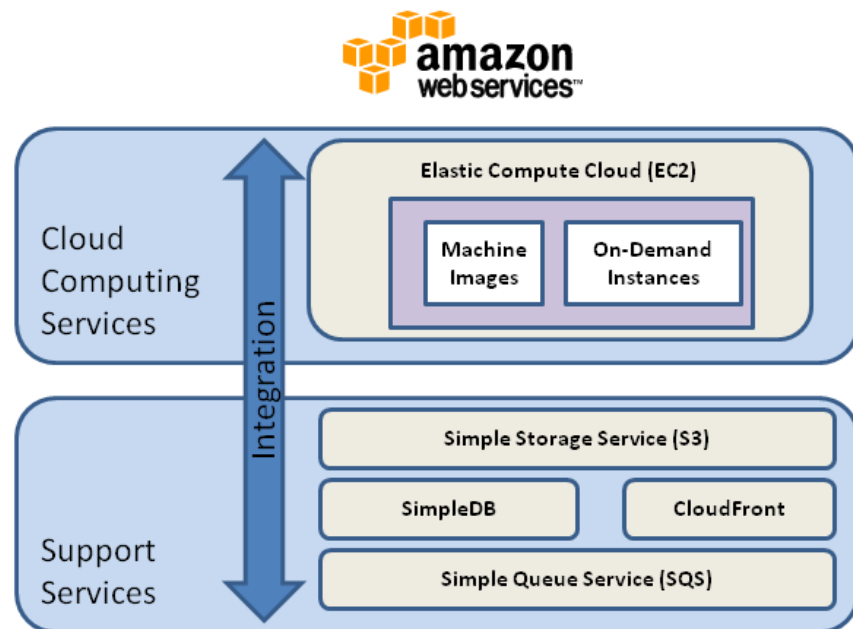
- “Cloud computing” is a very fuzzy term (to be kind)
- Depending on who you talk to:
 - a revolutionary idea that is rapidly changing the face of computing
 - an old idea whose time has come
 - just hype
 - evil
- In any case, it is changing economics behind computing in important ways



Copyright © 2010 Accenture All Rights Reserved.

Source: Jeanne G. Harris and Allan E. Alter, "Cloudrise: Reward and Risks at the Dawn of Cloud Computing," Accenture Institute for High Performance, November 2010.

Everything as a Service



Infrastructure



Software

Google App Engine

[Home](#) [Docs](#) [FAQ](#) [Articles](#)

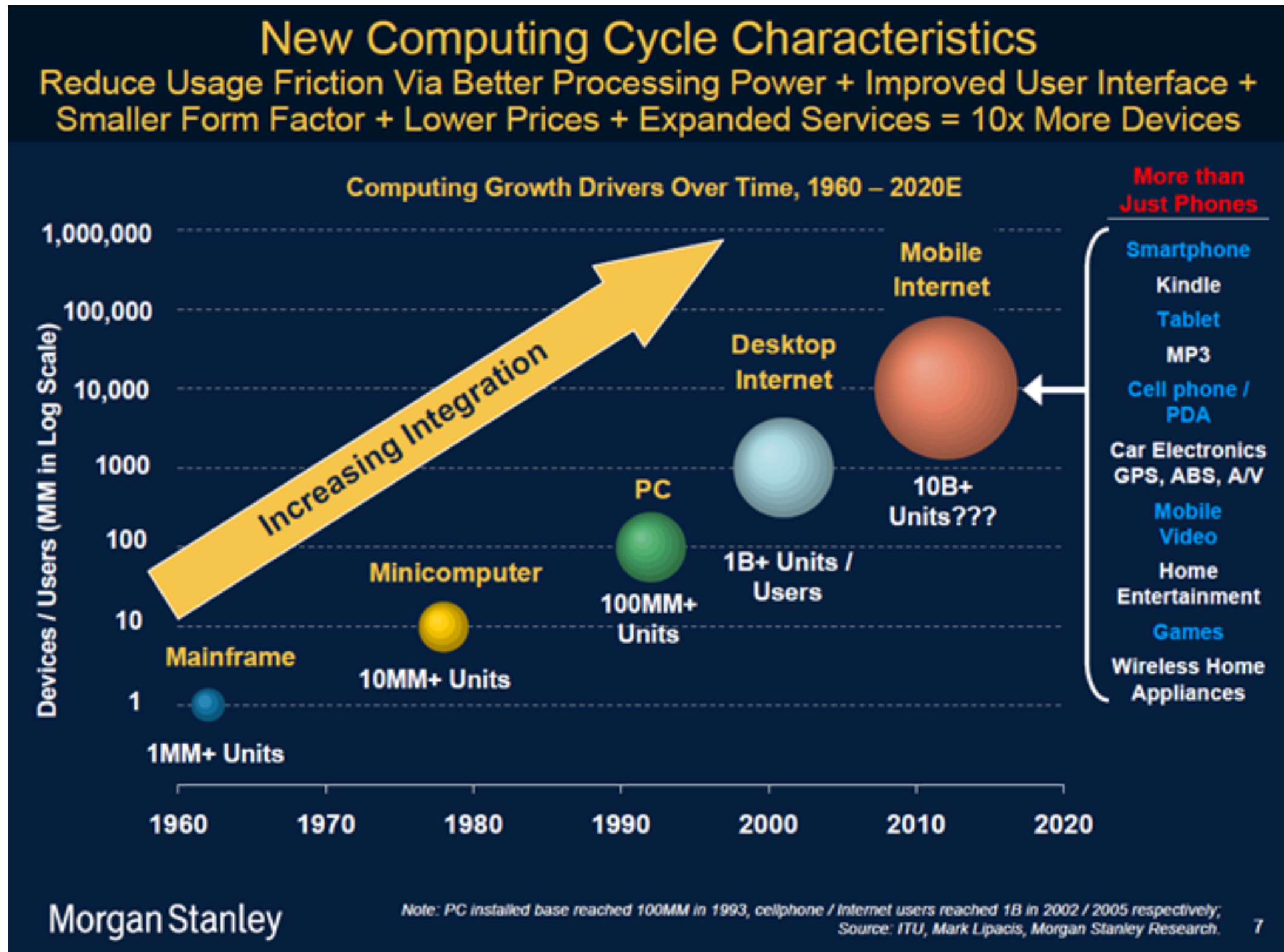
Platform



Run your web apps on Google's infrastructure
Easy to build, easy to maintain, easy to scale

Google App Engine enables you to build and host web apps on the same systems that power Google applications. App Engine offers fast development and deployment; simple administration, with no need to worry about hardware, patches or backups; and effortless scalability.

The World is going Mobile



Large Scale Programming



HDFS

Google File System

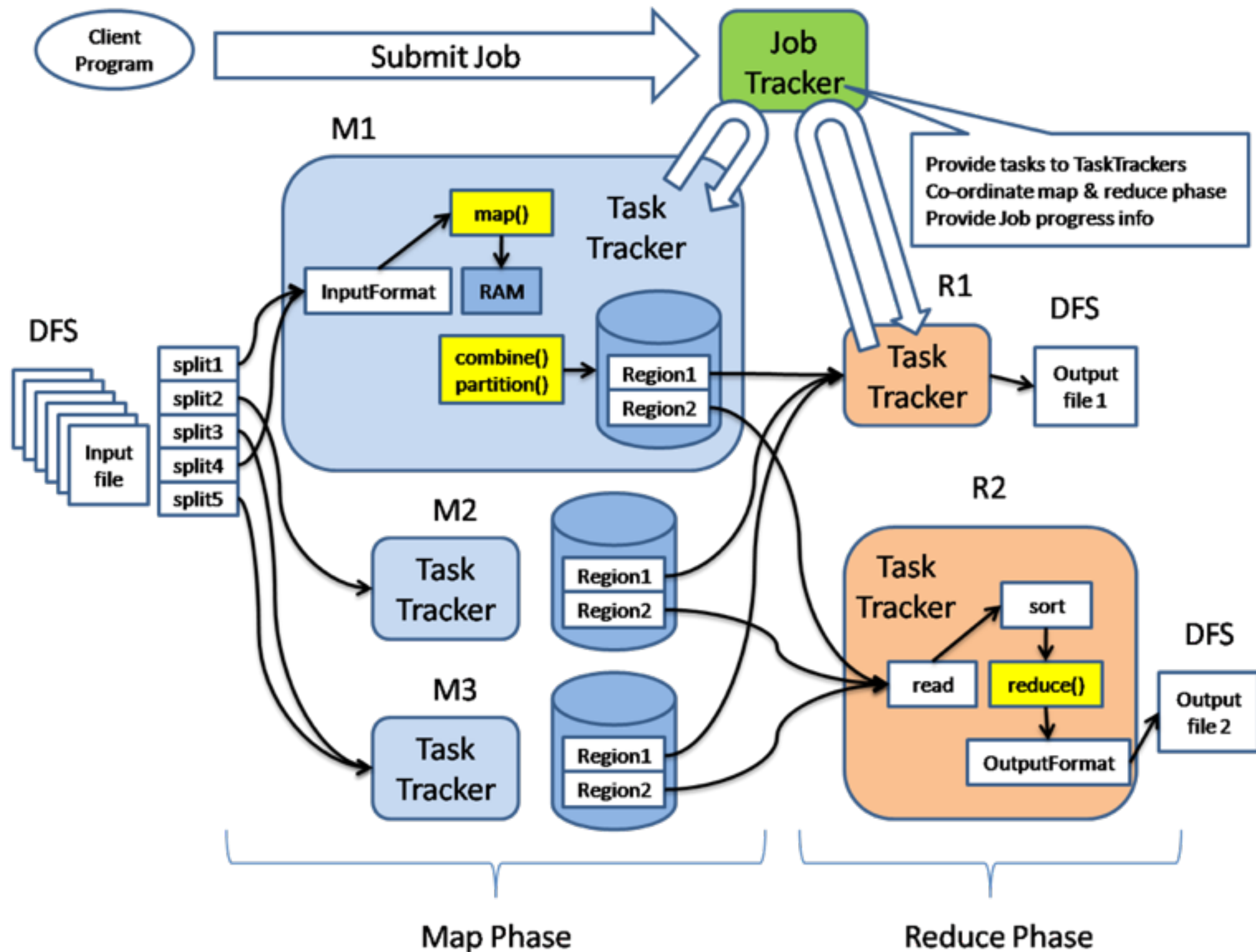
Hadoop

Google MapReduce

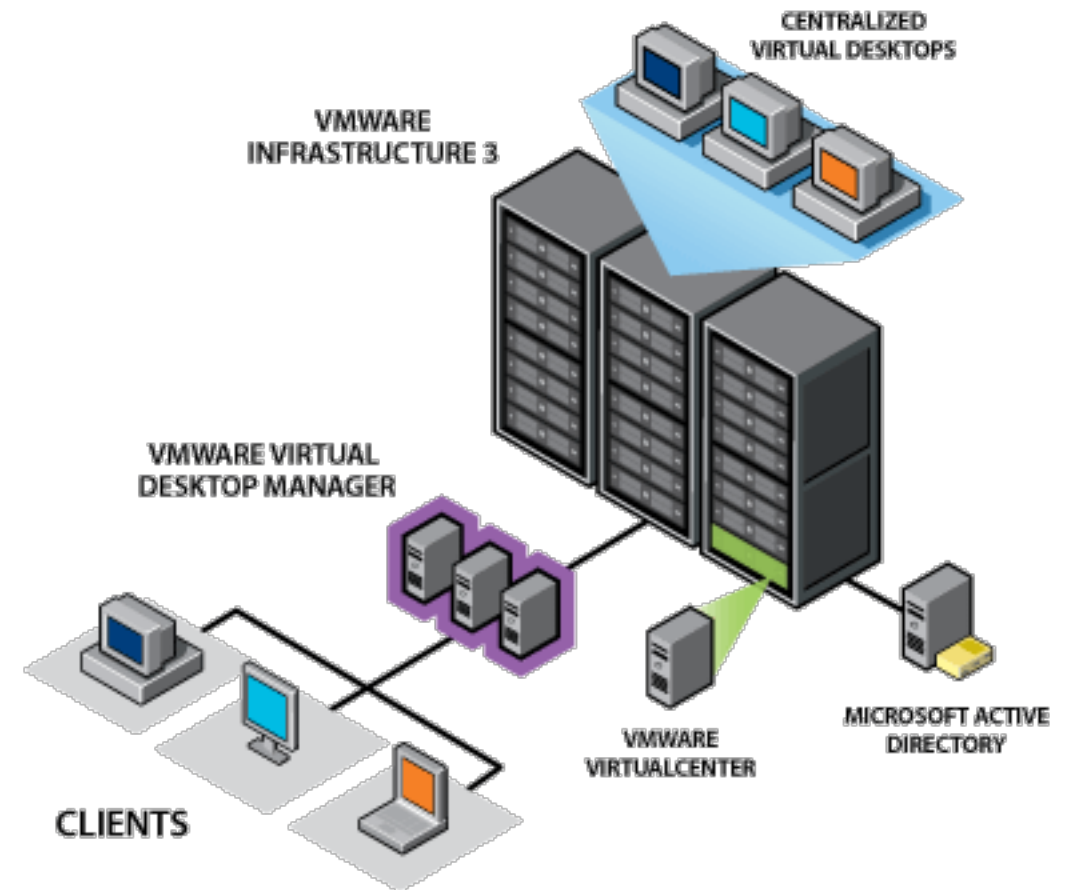
HBase

Google BigTable

Map Reduce



Virtualization



Who?



- **Nicola Tonellotto**

- Laurea degree in Computer Engineering
- PhD in Information Engineering @ UNIPI (Italy)
- PhD in Computer Engineering @ UNIDO (Germany)

- Researcher @ ISTI-CNR since 2002
 - Grid Computing
 - Scheduling
 - Information Retrieval

- TA @ UNIPI since 2002
 - Parallel and Distributed Applications
 - Fundamentals of Computer Science
 - C/C++ Programming
 - Java Programming
 - Distributed Enabling Platforms



Where? & When?



Course Organization

- 48 hours: ~32 lessons, ~16 laboratory
- TAs
 - Alberto Gotta
 - Salvatore Trani
- Agreement on room and timetable
 - Currently: Mon 11-13 (room N1), Thu 11-13 (room N1)
 - Depending on availability
- Highly interactive lectures
- Laboratory
 - Java programming skills required
- Notes and references available online
 - Updated in real time on the course wiki
- Final examination: project + oral session
 - To be agreed with teacher