Introduction to Grid Computing

## *Balázs Kónya*

(balazs.konya@quark.lu.se)

What is the Grid?

- Grid is a technology to share and seamlessly access resources of the world:

  - computing cycles
  - datasets,
  - software,
  - special instruments

- the Holy Grail of distributed computing

- Middleware: a bag of software which implements Grid Standards & protocols

- World Wide Web access to information

- World Wide Grid access to computing capacity and ...

What is the Grid?

- The future infrastructure of computing and data management

- The Computational Power Grid

- a very ambitious attempt to create a new utility, next to the already existing water, heating, electricity, ...

- the present hype in IT
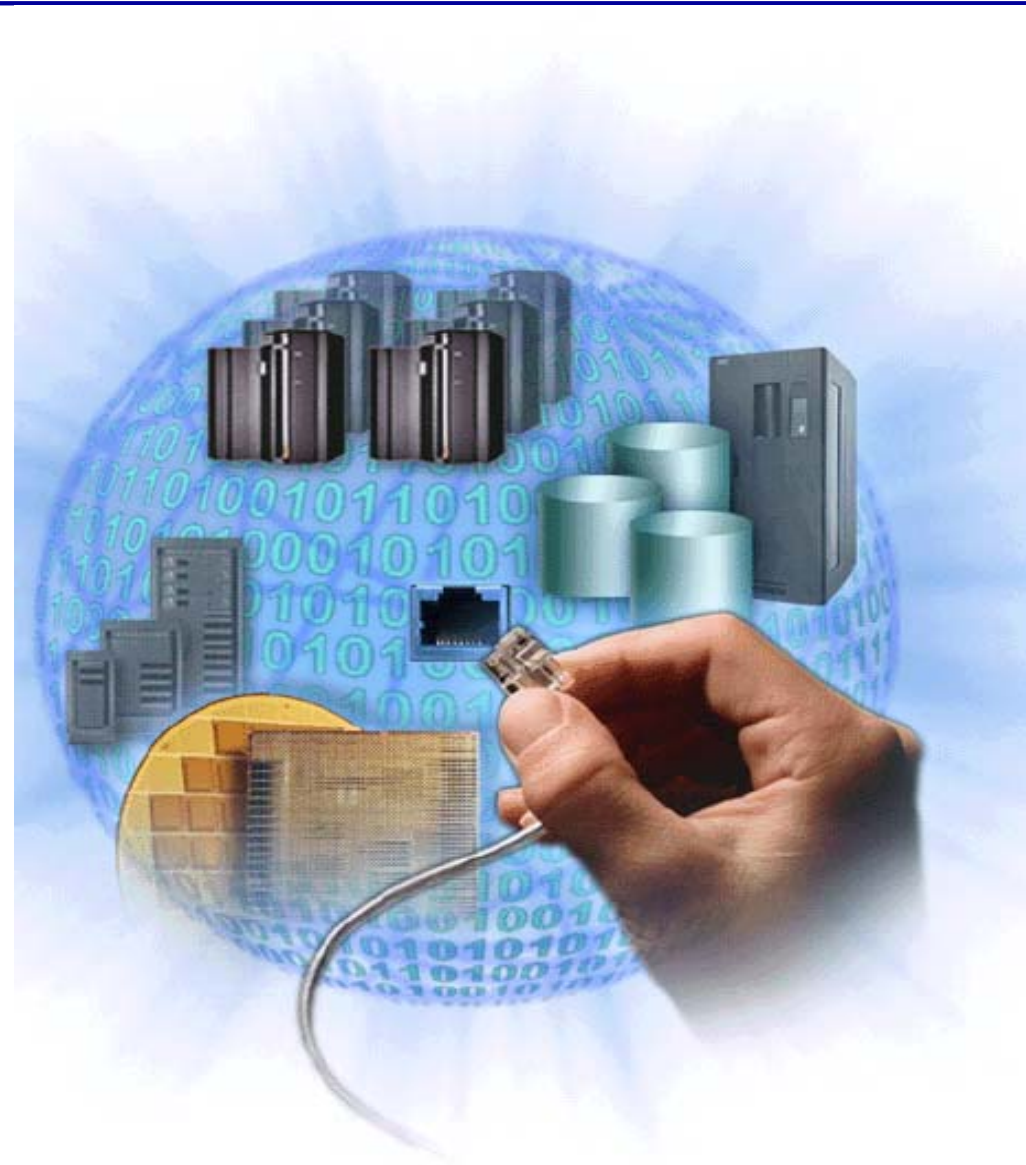
source: IBM

History

- Grand Scientific Challenges of the 80's
  - parallel computation
  - high performance & high throughput computing
- Early "Testbeds" in the USA connected supercomputing centers at the late 90's
- Ian Foster, Carl Kesselman, July, 1998:

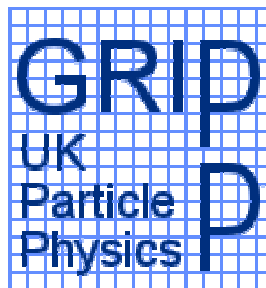 *Blueprint for a new Computing Infrastructure*
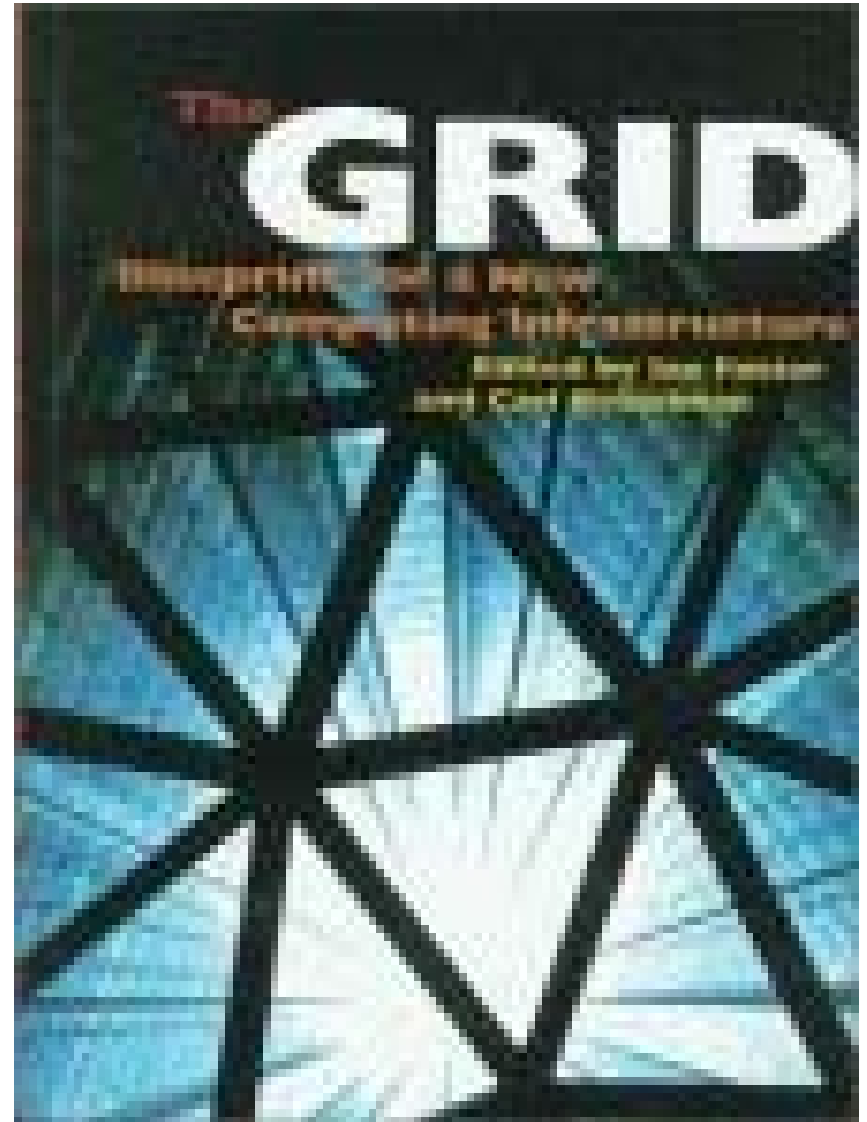
- The Computational Grid <-> Power Grid analogy was suggested

- The birth of the "ancient" middleware solutions
  - Globus, Legion, Condor, NWS, SRB, NetSolve, AppLes,Unicore
  - "demonstration quality Testbeds: Gusto" no real users
  - loose of interest in Grids

- 2000+: The Grid revives and gets "Global"
  - The High Energy Physics community picks up the nearly forgotten Grid idea
  - The appearance of the Global Grid Forum
  - de facto standard middleware: Globus

- the "Grid phenomena" or hype is started
  - Grid Projects are launched everywhere, governments & research agencies rush to support Grid project
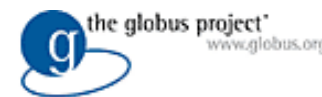
History cont.

- Huge commercial interest: startup companies & the Big Names try to sell the Grid
  - IBM wants to Grid-enable the company's entire product portfolio
  - commercial Grid software (IBM, Platform Computing, SUN)
    - commercial support, consulting, training
  - seriuos research projects (mainly biology) among the customers
- last Global Grid Forum meeting in Edinburgh July, 2002:
  - over 850 participants
  - Key speakers involved: IBM, Nec, Hewlet Packard, Microsoft, SUN
- Daily Grid magazines: www.thegridreport.com, www.gridtoday.com, www.gridcomputingplanet.com

GRID today   The Grid Report
                    the *state* of grid computing

- Everything is Grid, the word "Grid" is inflated to a marketing term
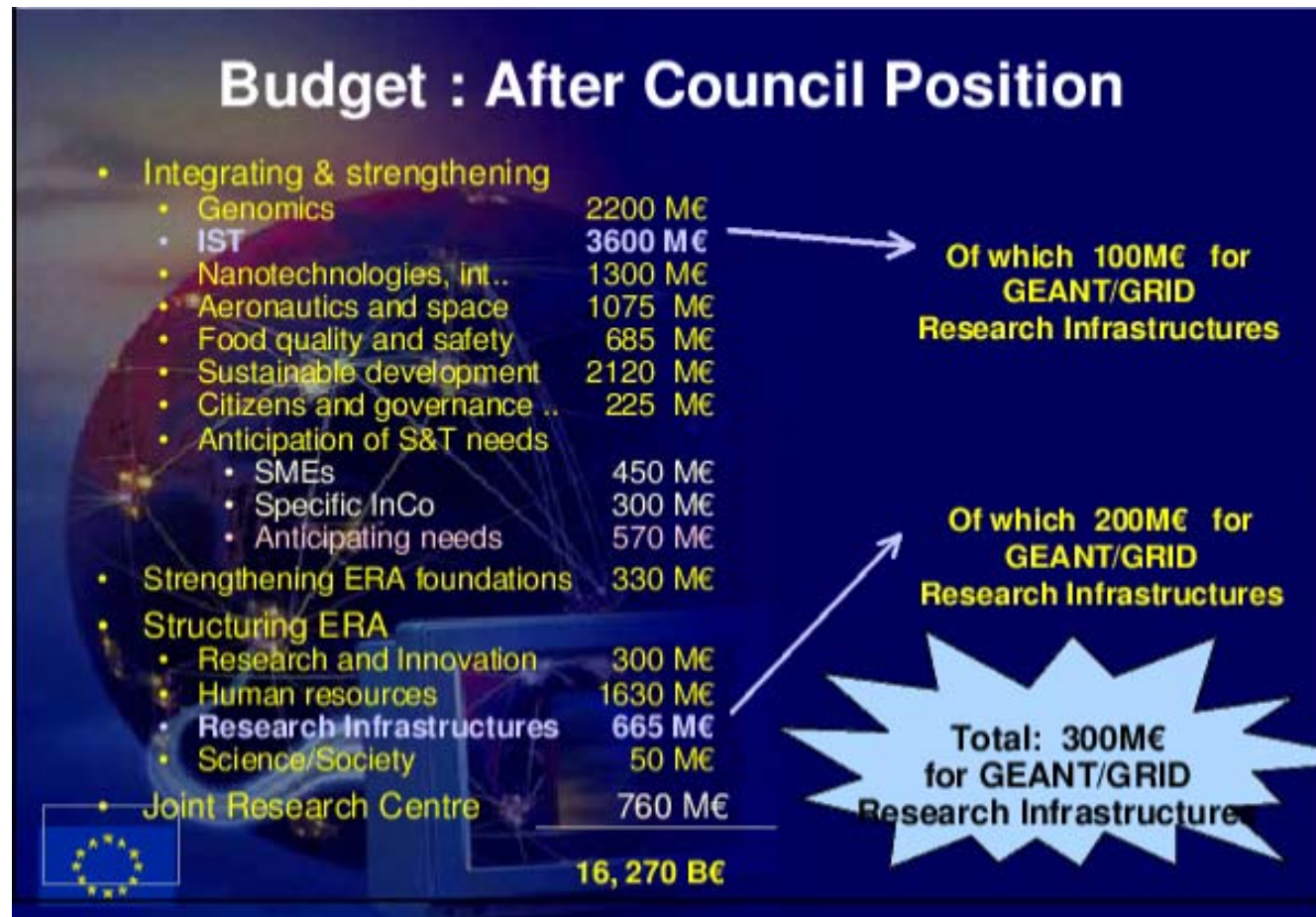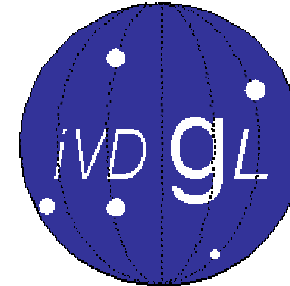- the divergence of Grid Toolkits and solutions

European projects

GRIDSTART

AVO

EGSO

GRIA

EUROGRID

DATAGRID

GRIP

GRIDLAB

CROSSGRID

DAMIEN

DATATAG

FLOWGRID
OPENMOLGRID
GRACE
COG
MOSES
GEMSS
BIOGRID
SELENE
MAMMOGRID

EU FP6

Budget : After Council Position

- Integrating & strengthening
  - Genomics — 2200 M€
  - IST — 3600 M€ → Of which 100M€ for GEANT/GRID Research Infrastructures
  - Nanotechnologies, int.. — 1300 M€
  - Aeronautics and space — 1075 M€
  - Food quality and safety — 685 M€
  - Sustainable development — 2120 M€
  - Citizens and governance .. — 225 M€
  - Anticipation of S&T needs
    - SMEs — 450 M€
    - Specific InCo — 300 M€
    - Anticipating needs — 570 M€
- Strengthening ERA foundations — 330 M€
- Structuring ERA
  - Research and Innovation — 300 M€
  - Human resources — 1630 M€
  - Research Infrastructures — 665 M€
  - Science/Society — 50 M€
- Joint Research Centre — 760 M€

16, 270 B€

Of which 200M€ for GEANT/GRID Research Infrastructures

Total: 300M€ for GEANT/GRID Research Infrastructure

Nordic Testbed for Wide Area Computing and Data Handling

the globus project
www.globus.org

NEESgrid

iVDGL

DISCOM

SinRG

APGrid

IPG …

GriPhyN
Data Intensive Science

United States virtual observatory

SDSC/UCSD • NCSA/UIUC • Caltech • ANL

TERAGRID

NSF PACI

SDSC
S R B
STORAGE RESOURCE BROKER

NASA

GrADS
Grid Application Development Software Project

NETWORK WEATHER SERVICE

LEGION Worldwide Virtual Computer
e pluribus unum: one out of many

Condor
High Throughput Computing

TeraGrid

- 53 million from the NSF

- 13.6 teraflops of Linux clusters

- 450 terabytes of data storage

- 4 sites

- 40 Gbits/sec (later 50-80) network connections

- Globus based Grid toolkits

- Visualisation environment
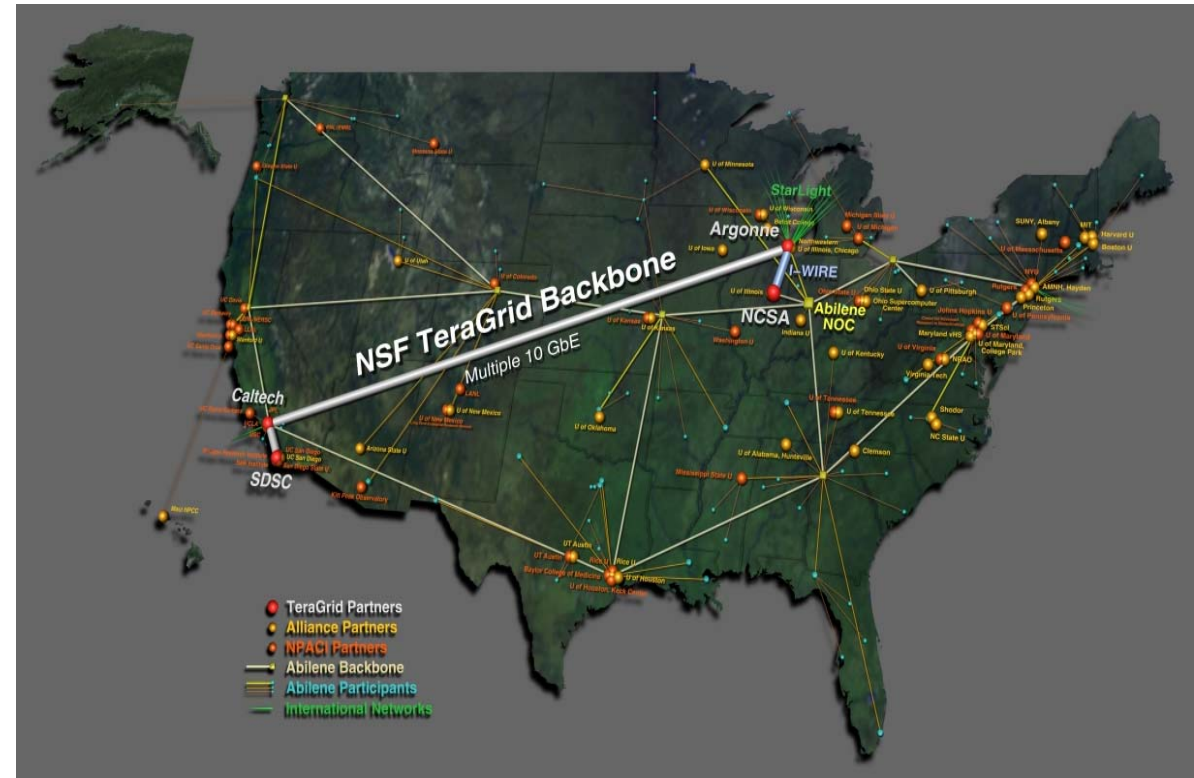
TeraGrid
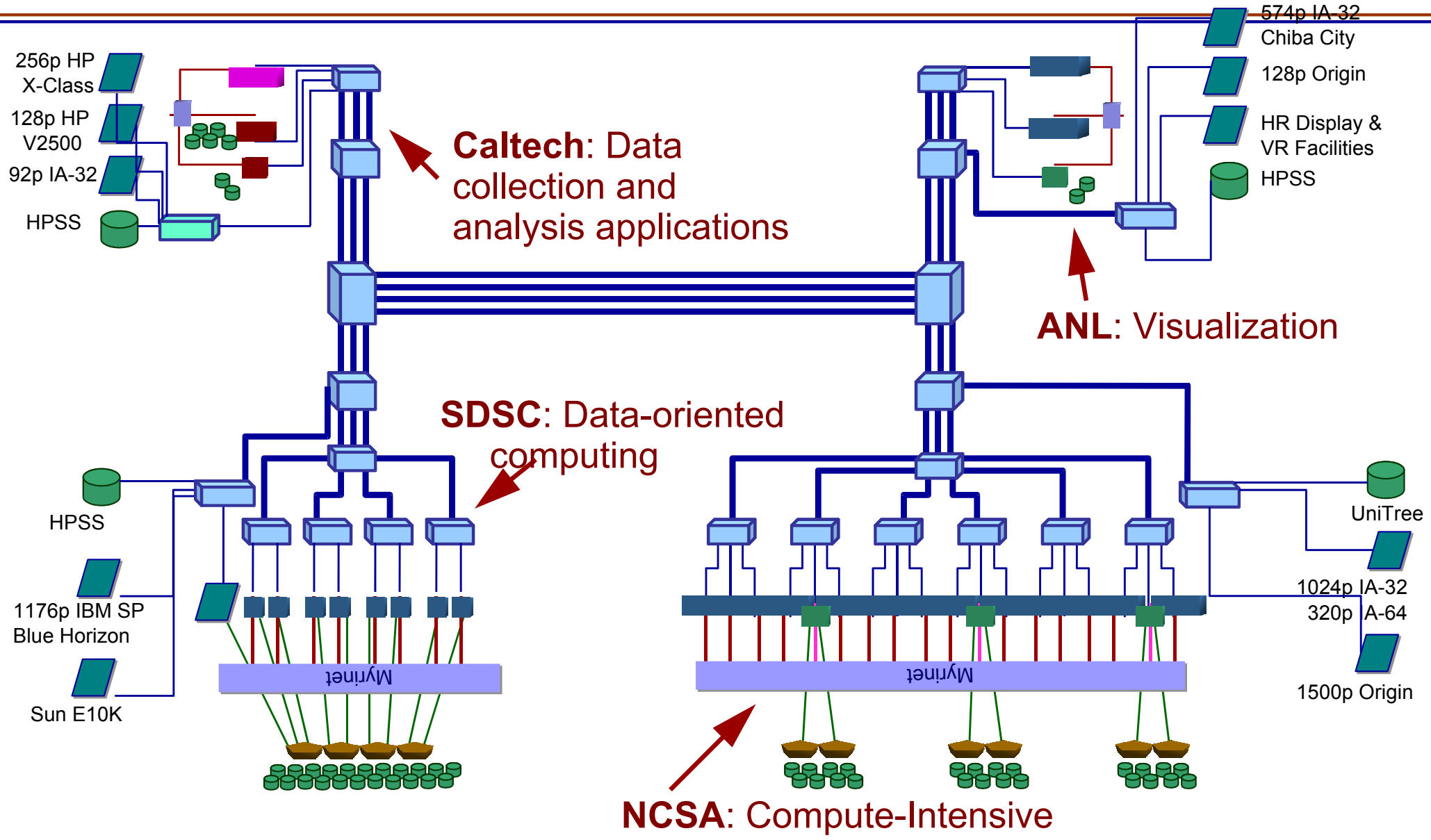
Japan: AIST GTRC
China: SDG
Korean Grid
Thailand:ThaiGrid
Australia: GRIDSLab

Grid in the NEWS

BusinessWeek online

| BW MAGAZINE | DAILY BRIEFING | INVESTING | GLOBAL BUSINESS | TECHNOLOGY | SMALL BUSINESS | B-SCHOOL |

JUNE 3, 2002

INFORMATION TECHNOLOGY

▶ BW MAGAZINE
U.S. EDITION
Full Table of Contents
Cover Story
Up Front
Readers Report

Who Needs Supercomputers?
Grid software lets companies tap other machines in their network

GRID COMPUTING PLANET.COM — **EARTHWEB**

| ...anagement | Networking & Communications | Web Development | Hardware & Systems | Software Development |

ernet.com™

...ridcomputingplanet.com : News

FREE TECH NEWSLETTERS   YOUR EMAIL ➘   Sign Up
Grid Computing Planet Text

PHYSICS TODAY.org

Gradschool

Sh... ...rch Engine ...ATEGIES 2002 ...ERENCE & EXPO

Researchers Achieve Production Grid Breakthrough
By Paul Shread

Search   ▶   ● Welcome!   logoff

| FEATURE ARTICLES | PHYSICS UPDATE | LETTERS | SEARCH & DISCOVERY | ISSUES & EVENTS | BOOKS | WE HEAR THAT | OBITUARIES |

TABLE OF CONTENTS
PAST CONTENTS
Feature Article

LINKS TO PHYSICS TODAY ADVERTISERS
PLACE AN AD
BUYER'S GUIDE
ABOUT US
CONTACT US

The Grid: A New Infrastructure for 21st Century Science
As computer networ...
of science and engi...

Ian Foster

| White Papers | IT Jobs | Events | Research | Premium Services | Training & Certification | Media Kit |

GRID COMPUTING PLANET.COM — **EARTHWEB**

| IT Management | Networking & Communications | Web Development | Hardware & Systems | Software Development |

internet.com™

www.gridcomputingplanet.com : News

FREE TECH NEWSLETTERS   YOUR EMAIL ➘   Sign Up
Grid Computing Planet Text

AMD

News Archives

TeraGrid Receives $35 Million From National Science Foundation
By Paul Shread

Grid in the NEWS

Vision...

- **Cohen Communication Group:**
  - Grid computing will be the driving force behind the 150% annual internet traffic expansion in 2005
  - in contrast to the 60% predicted growth rate accounted mainly for video streaming and video file transfer forecasted by McKinsey - JP Morgan

- **IBM:**
  - Grid is the next utility in the line of the water, drainage, gas and electricity systems
 - people will pay their "computing bills"

Oversold?

- The promise of the Grid has been not oversold but the difficulty of developing the necessary Grid infrastructure has been underestimated

Ian Foster:

- People used to overestimate the short term impact of change but underestimate the long-term effect

what is behind?

- Powerful PCs are everywhere

- Clusters are commodity

- Network & Storage & Computing exponentials:

  - Networking speed grows faster than computing power

  - Even data storage outperforms the CPUs
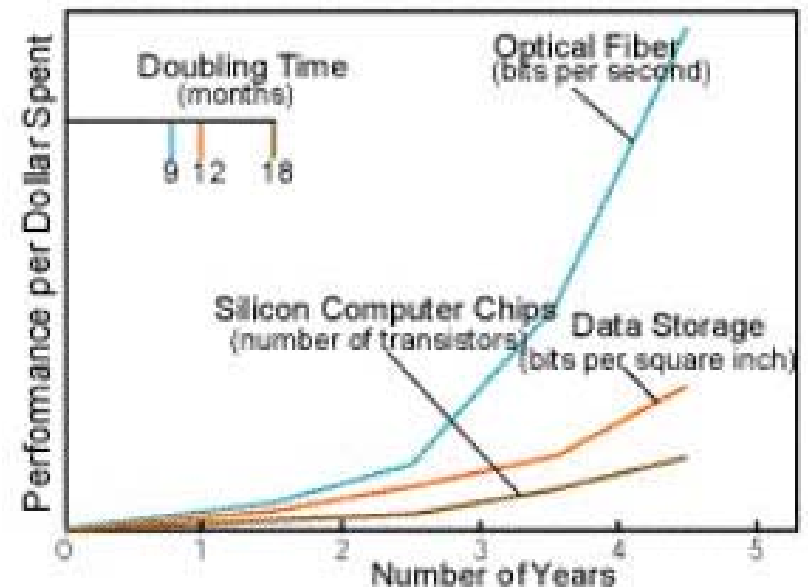
source: Scientific American, Jan 2001

The LHC Detectors

ATLAS

CMS

3.5 PetaBytes / year
~10^8 events/year

Alice
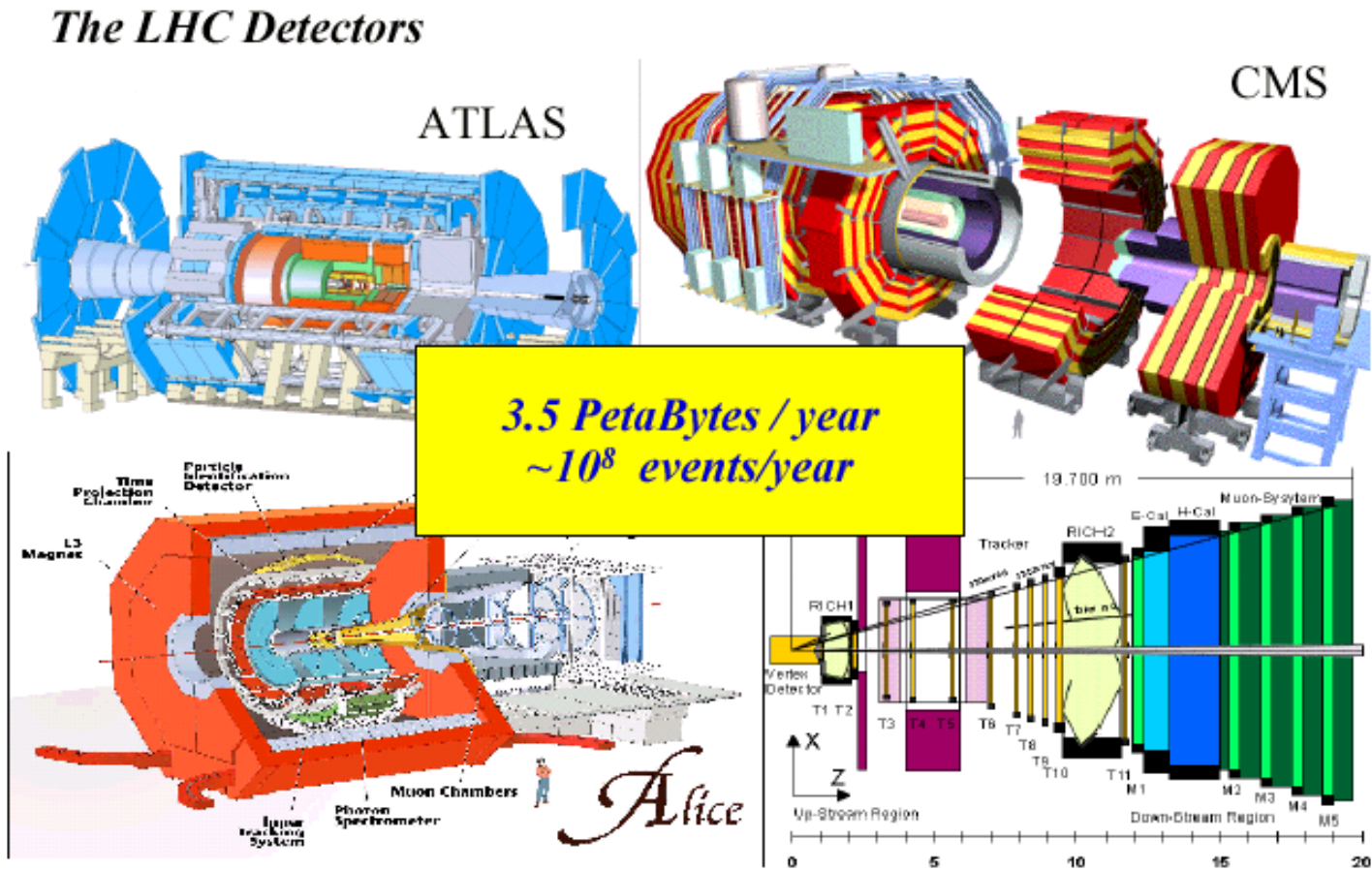
Walmart Inventory Control

→ Satellite technology used to track every item

→ Bar code information sent to remote data

→ Inventory adjusted in real time to avoid shortages and predict demand

→ Data management, prediction, real-time, wide-area synchronization

SETI@HOME

→ 3.8M users in 226 countries

→ 1200 CPU years/day

→ 1.7 ZETAflop over last 3 years (10^{21})

→ 38 TF sustained performance

(Japanese Earth Simulator is 40 TF peak)

→ Highly heterogeneous: >77 *different* processor types

Copyright ©2001 Wal-Mart Stores, Inc.

... distributed applications

Everquest

➡ 45 communal "world servers" (26 high-end PCs per server) supporting 430,000 players

➡ Real-time interaction, individualized database management, back channel communication between players

Napster, Gnutella, Kazaa, etc...

➡ file sharing

➡ ask the music industry :)

Google

➡ database, search engine

➡ more than 150 million searches per day, 2 billion indexed pages, more than 10.000 linux servers

there should be a Grid ...

- Existing real world examples demonstrate that it is technically, commercially, and economically viable to deploy robust, large-scale distributed applications

- The Grid

  - will extend those distributed applications

  - should accelerate the progress of distributed applications

  - will use common interfaces

  - will be based upon well-defined protocols & standards

  - will offer scalable Grid services for applications

where we are now?

- lots of theoretical papers
    - *The anatomy of the Grid:Enabling Scalable virtual organizations*, I.Foster et.al.
    - *The Physiology of the Grid: An Open Grid Services Architecture for Distributed System Integration*, I.Foster,C.Kesselman, et. al.
    - *The patology of the Grids*, ???
- non-existing (very few) TestBeds:
    - they are incompatible
    - Difficult to get access to them
    - very expensive to maintain
- non existing standards (GGF has not produced anything yet)
- "de facto standard" middleware is rather limited in functionality
- diverging solutions, huge amount of (overlapping) projects
- we are living in the Grid hype era

not even (hardly) addressed:

- political issues

- heterogeneity

- Grid-based authorization

- Grid schedulling

- Program development environments

- Debugging, compiling, performance tuning

- Fault tolerance

- Modeling of dynamic, unpredictable environments

- Grid market economy (allocation, accounting, cost models)

Definition

- Ian Foster *(www.gridtoday.com/02/0722/100136.html):*

  - coordinates resources that are not subject to a centralized control

  - using standard, open, general-purpose protocols and interfaces

  - delivers nontrivial qualities of service

- Rajkumar Buyya:

  - A type of parallel and distributed system that enables the sharing, selection, & aggregation of resources distributed in administrative domains depending on their availability, capability, performance, cost, and users quality of service requirements.

Data Management

Resource & Job Management

Information System

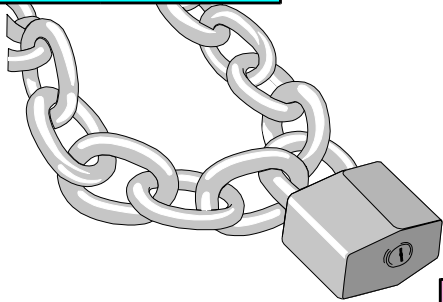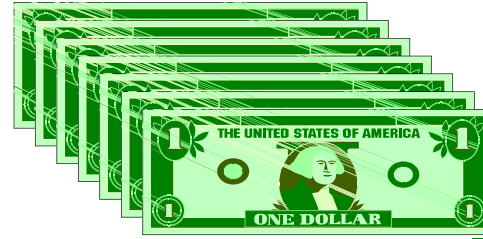+ security

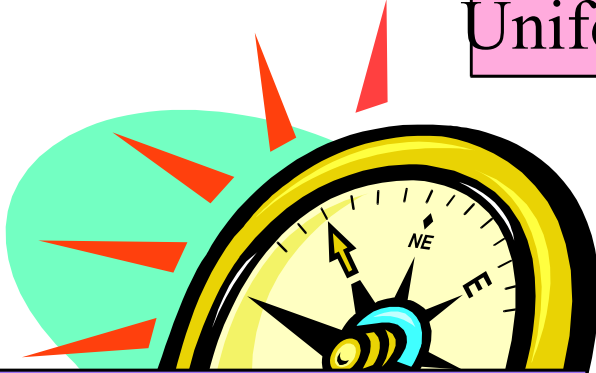# another model (basic elements)

Security

Computational Economy

System Management

Uniform Access

OCTOBER

Resource Allocation & Scheduling

Data locality

Resource Discovery
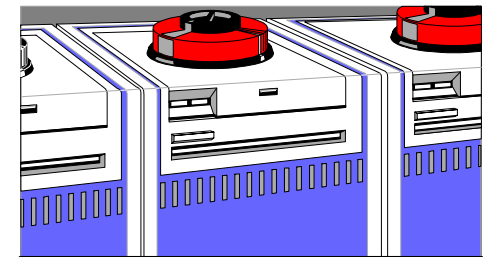
Application Development Tools
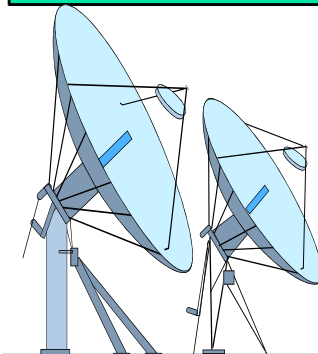
source:Rajkumar Buyya

Network Management

The layers of the Grid:

Grid Applications

science,engineering,commercial apps, web portals

Grid Programming environment

languages,interfaces,libraries,compilers, griddifying tools

User level Middleware

resource management and scheduling services

Low level Middleware

job submission, storage access, info service,accounting

Security Infrastructure

single log-on, authentication, authorization, secure communication

Grid Fabric

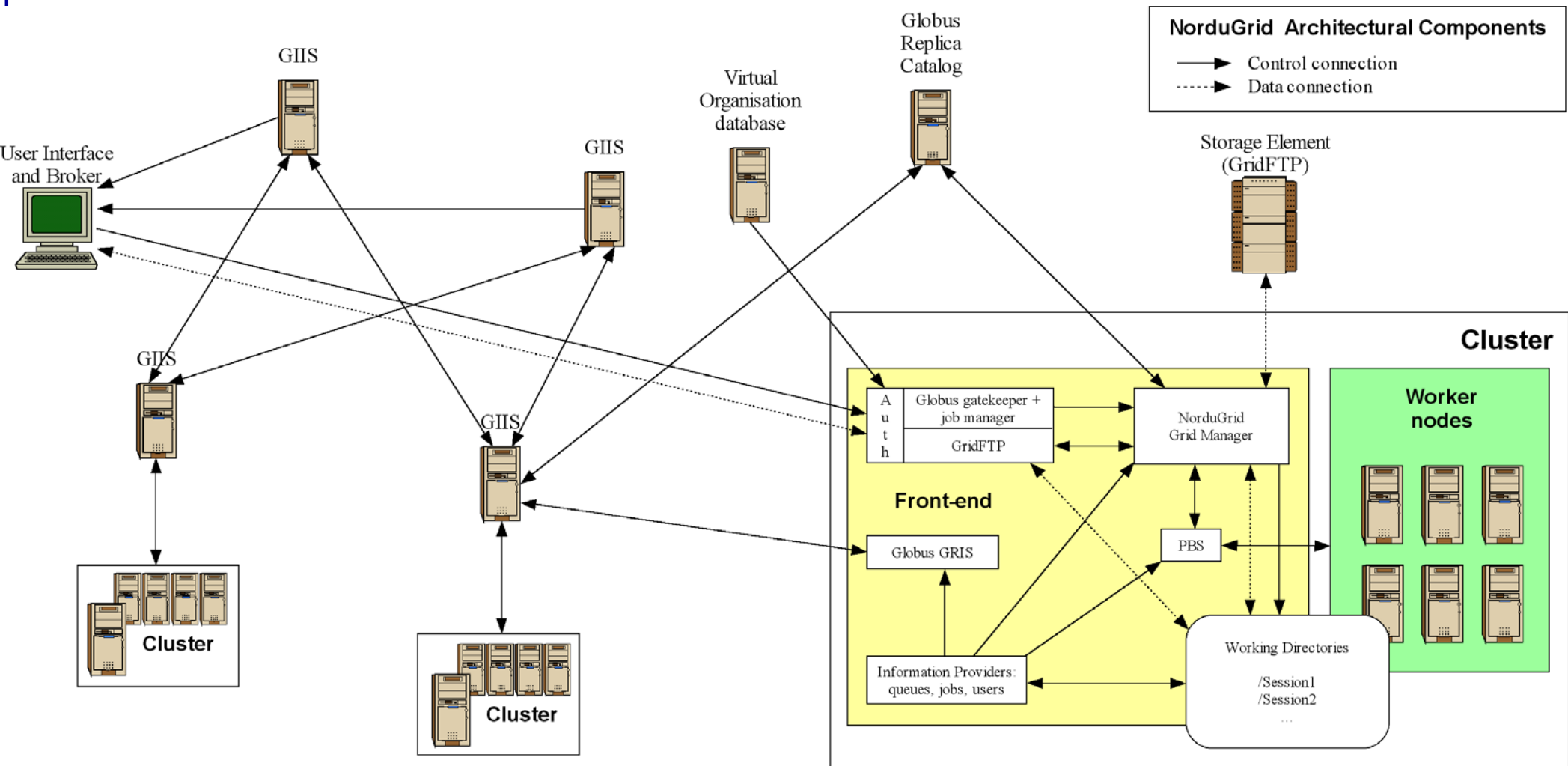clusters, networks, batch systems, devices, databases

- Linux Operating Environment
- Basic and Core Globus Services
  - GSI (Grid Security Infrastructure)
  - GSI-enabled SSH and GSIFTP
  - GRAM (Grid Resource Allocation & Management)
  - GridFTP
  - Information Service
  - Distributed accounting
  - MPICH-G2
  - Science Portals
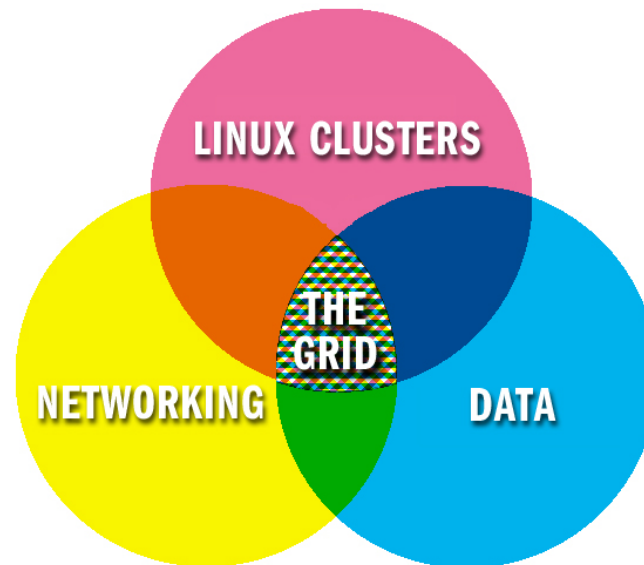

- Advanced and Data Services
  - Replica Management Tools
  - GRAM-2 (GRAM extensions)
  - CAS (Community Authorization Service)
  - Condor-G (as brokering "super scheduler")
  - SDSC SRB (Storage Resource Broker)
  - APST user middleware, etc.

NorduGrid Architectural Components

- → Control connection
- --→ Data connection

GIIS

GIIS

Virtual Organisation database

Globus Replica Catalog

Storage Element (GridFTP)

User Interface and Broker

GIIS

GIIS

Cluster

Worker nodes

Auth

Globus gatekeeper + job manager

GridFTP

NorduGrid Grid Manager

Front-end

Globus GRIS

PBS

Cluster

Cluster

Information Providers: queues, jobs, users

Working Directories

/Session1
/Session2
...

October 2002

Grid & Supercomputers

- The present day Supercomputers are the PC clusters

- Grid will provide a uniform access to all the resources

- The Supercomputing centers will be the power plants of the Grid

Cray T3E

PC cluster

Sun E10000

The Grid
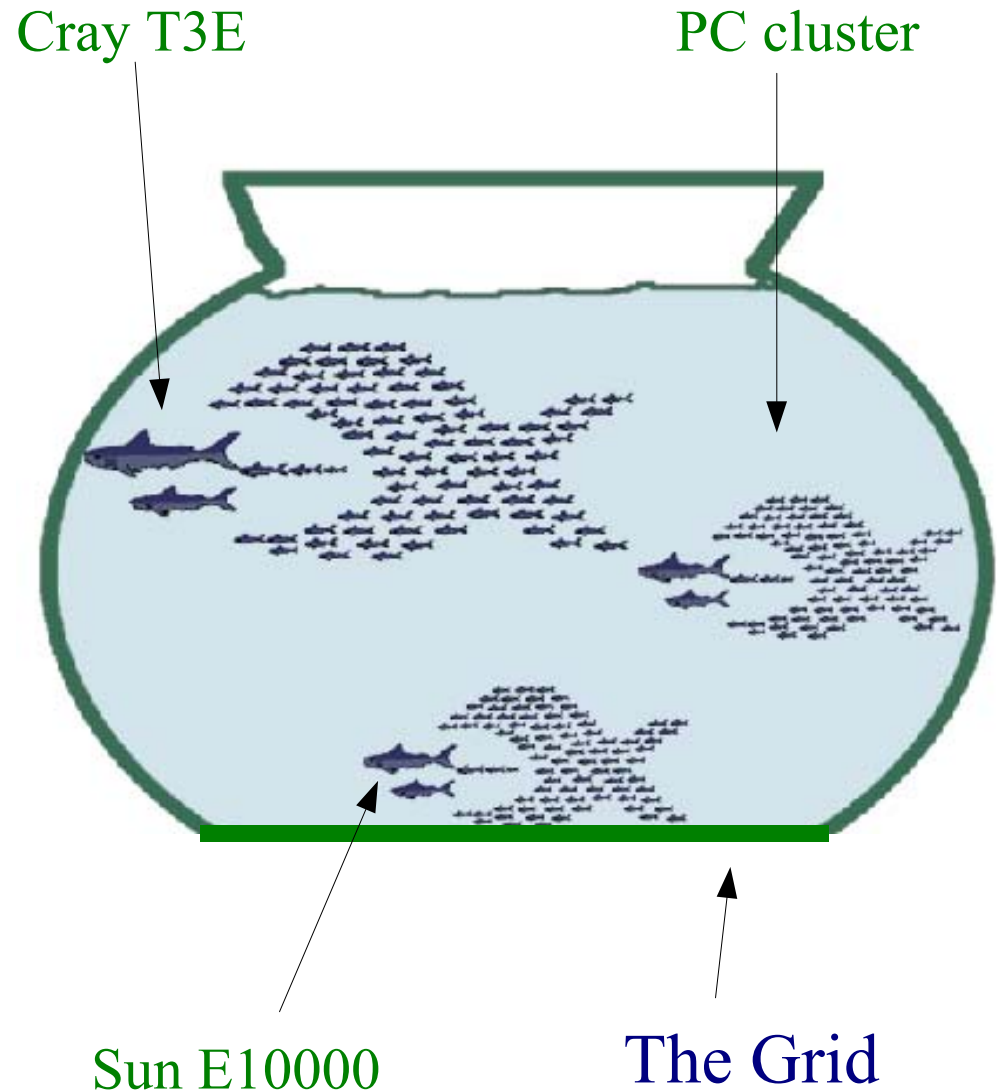
clusters,P2P,Grid

- Cluster:
  - single administrative domain
  - centralized resource management, full controll over resources
  - suitable for strongly-coupled applications
  - limited capacity
- Grid:
  - a layers on top of clusters, bunch of services on top of clusters
  - "borrows" resources from clusters, capacity will be able to be reserved
  - multiple administrative domains
- Peer-to-Peer:
  - millions of uncoordinated, unorganized desktops (screensavers)
  - parallel application pools
  - capacity varies and mostly unpredictable

Applications

- Applications are key to the success of Grid
- Application developers will only pick up the Grid IF:
    - Grid services will have a well-defined interface
    - Grid infrastructure some day be as natural part of the picture as the OS
- We are still very far from "throwing any application onto the Grid"
- Considerable porting effort in "Griddifications" of problems

targeted application areas

Genetics

Computer Graphics

VLSI Design

Bioinformatics

Data Mining

Network Simulation

Molecular Dynamics

BioInformatics

Fluid Dynamics

Astrophysics

Drug Design

Protein Structure

Meteorology

Earth Observation

Civil Engineering

Financial Risk Analysis

"best" applications for the Grid

Decoupled applications (minimal communication)

  → embarrassingly parallel apps, parameter sweeps

Staged/linked applications

  → (complete part A then do part B)

  → Includes remote instrument applications
    (get input from instrument at site A, compute/analyze
    data at site B)

Access to resources (mainly data)

  → get "something" from/do "something" at site A

  → dataGrids

  → data & controlled/shared acces to date is the critical
    issue of the future

"Data is the killer app"

there are many simillar but incompatible solutions :

SRB-G	DataCutter	I2T
Matrix	SAM	Magda
D2K	Spitfire	*Your Tool/Service Here*
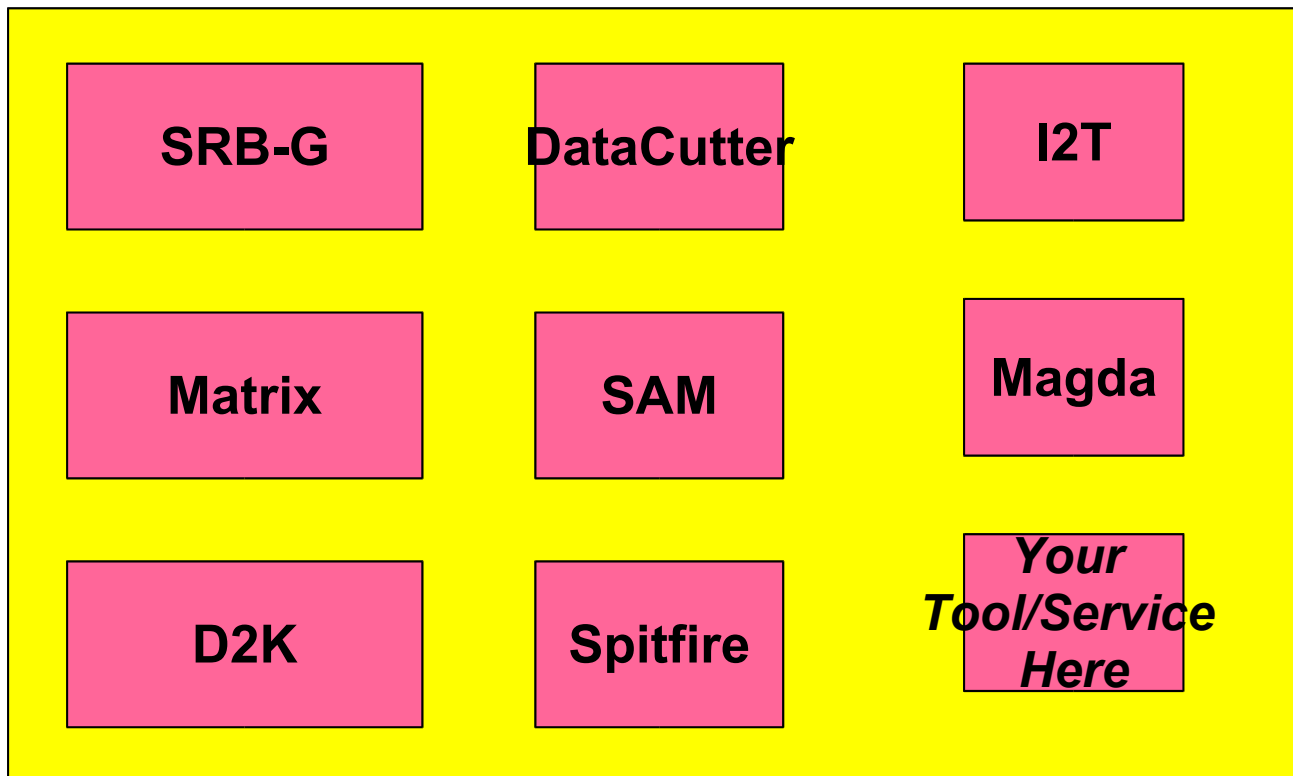
Alessandro Volta in Paris in 1801 inside French National Institute shows the battery while in the presence of Napoleon I

Fresco by N. Cianfanelli (1841)

(Zoological Section "La Specula" of National History Museum of Florence University)

source: Rajkumar Buyya

source: Rajkumar Buyya

NorduGrid Project

www.nordugrid.org

- Create a Grid infrastructure in Nordic countries

- Operate a production quality Testbed

- Expose the infrastructure to end-users of different scientific communities

- Survey current Grid technologies

- Pursue basic research on Grid Computing

- Develop Middleware Solutions

brochure:www.nordugrid.org/documents/brochure.pdf

Lund, 24/01/2003

Participants
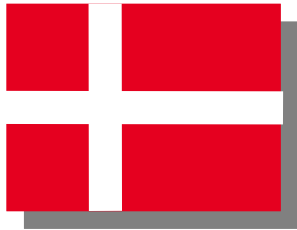
Copenhagen University: Niels Bohr Institute, Research Center COM, DIKU
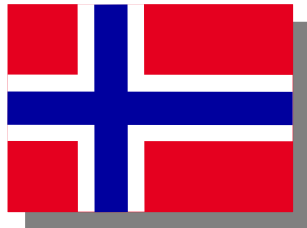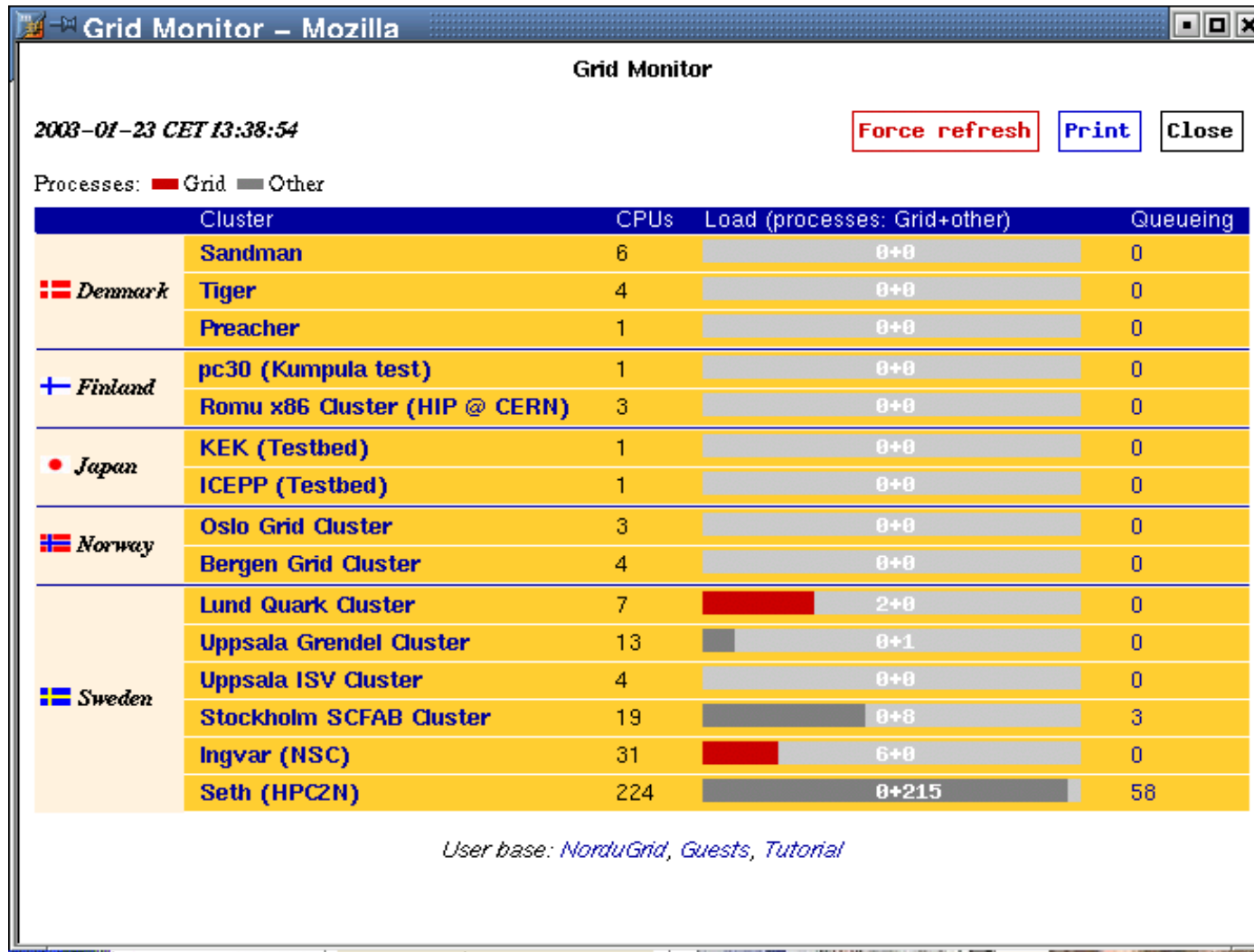
Oslo University, Bergen University

Lund University, Uppsala University, Stockholm University, KTH

Helsinki Institute of Physics

resources:

www.nordugrid.org, and click on the GridMonitor

contact us

Dept. of Elementary Particle Physics, Physics Building A427,

balazs.konya@quark.lu.se

- The Grid needs you & your ideas :)
  - computing & physics diploma works within the NorduGrid
  - PhD topics
  - try to run your application on the Grid
  - or just look at the www.nordugrid.org

Nordic Testbed for Wide Area Computing and Data Handling

→ while I was preparing for this introductory talk I "borrowed" slides, ideas, pictures from general Grid-talks. I would like to thank all the authors of those talks. Especially to Rajkumar Buyya & Fran Berman*

*GGF5 Plenary Keynote: TeraGrid "State of the Grid 2002"