A short introduction to grid computing

Stefano Cozzini

CNR/INFM Democritos

The EU-IndiaGrid project, RI-031834, is funded by the European Commission,
DG INFSO under the Research Infrastructures Programme

European Commission
Information Society and Media

This first introduction

- What is e-science ?
- Why/What is GRID computing ?
- The elements of grid computing:
  - Middleware
  - Applications
- Comparing GRID vs HPC
- examples of GRID project/infrastructures

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

in search of E-science

- What is meant by e-Science? In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet

[from http://www.nesc.ac.uk/nesc/define.html]

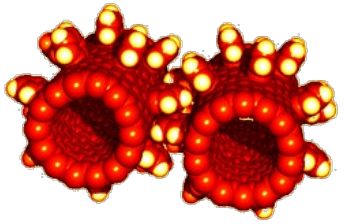
- The term e-Science (or eScience) is used to describe computationally intensive science that is carried out in highly distributed network environments

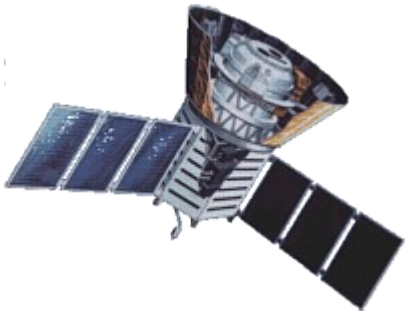
[from wikipedia]

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

e-science=computationally intensive science

- Science is becoming increasingly digital and needs to deal with increasing amounts of data and computing power
- Simulations get ever more detailed
  - Nanotechnology – design of new materials from the molecular scale
  - Modelling and predicting complex systems (weather forecasting, river floods, earthquake)
  - Decoding the human genome
- Experimental Science uses ever more sophisticated sensors to make precise measurements
  - → Need high statistics
  - → Huge amounts of data
  - → Serves user communities around the world





EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

e-science= new approach to do science

- New tools&methods
  - distribute collaborations
  - pooling of resources geographically distributed (GRID Computing)
  - powerful and modern
    - hardware
    - software
  - IT- skilled computational scientists

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Which architectures in your infrastructure ?

- Parallel computing:
  - single systems with many processors working on same problem
- Distributed computing:
  - many systems loosely coupled by a scheduler to work on related problems
- Grid Computing:
  - many systems tightly coupled by software, perhaps geographically distributed, to work together on single problems or on related problems

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Why the GRID?

- Motivation: When communication is close to free we should not be restricted to local resources when solving problems.
- A Grid Infrastructure built on the Internet and the Web to enable and exploit large scale sharing of resources
- It should provides Scalable Secure Reliable mechanisms for discovery and for remote access of resources.

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

A first definitions

- *Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations"*
  - Carl Kesselman,Ian Foster in "the anatomy of the grid" 2000

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

A GRID checklist

- a Grid is a system that:

1) coordinates resources that are not subject to centralized control ...

(Otherwise, we are dealing with a local management system.)

2)..using standard, open, general-purpose protocols and interfaces...

(Otherwise, we are dealing with an application specific system.)

3) ...to deliver nontrivial qualities of service.

(It should meet complex user demands, so that the utility of the combined system is significantly greater than that of the sum of its parts.)

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

A few concepts in GRID COMPUTING

- Resources are locally managed and controlled
- Different resources can have different policies and mechanism
  - Computing resources managed by different batch system
  - Different storage system on different node
  - Different policies granted to the same user on different resources on the GRID
- Dynamic nature:
  - Resources and users can change frequently

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

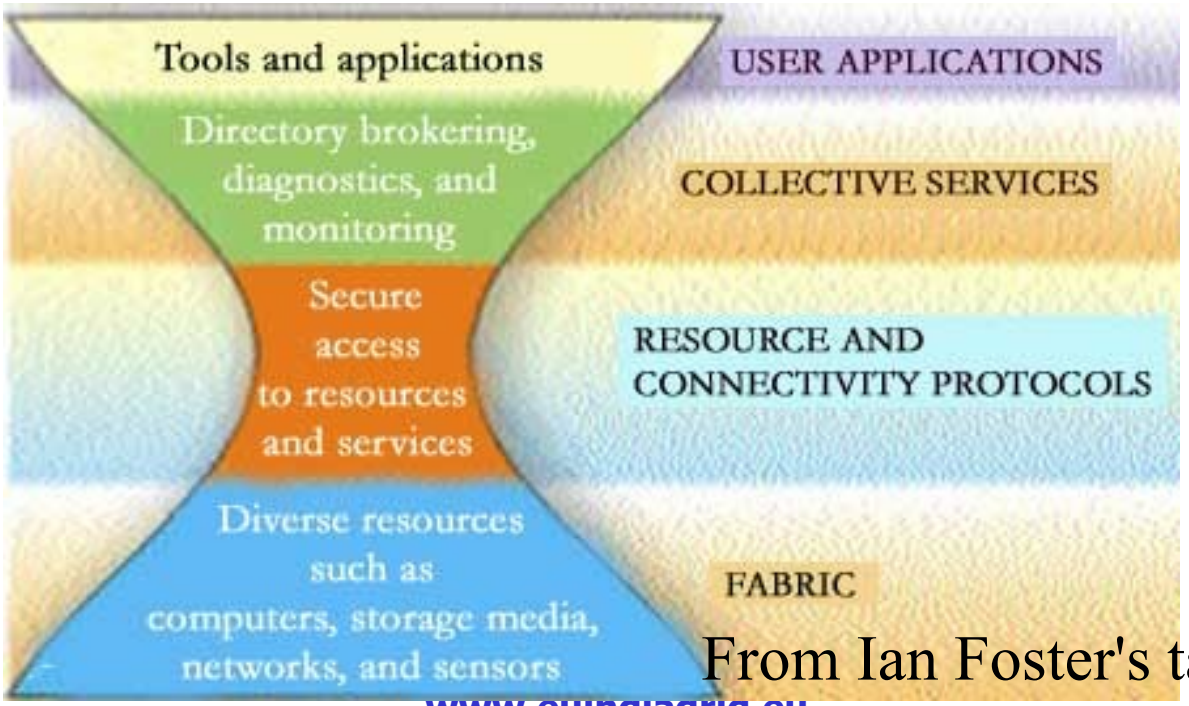
The elements of a GRID infrastructure

- Hardware/Resources
  - Made available from different sites geographically distributed
  - CPU/Storage/Instruments/DB
- Software:
  - Something that links together all these resources: the middleware
  - Some applications to use the computational power made available
- People:
  - Who maintain the Grid
  - Who use the GRID

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

GRID middleware

- Middleware is "the software layer that lies between the operating system and the applications "

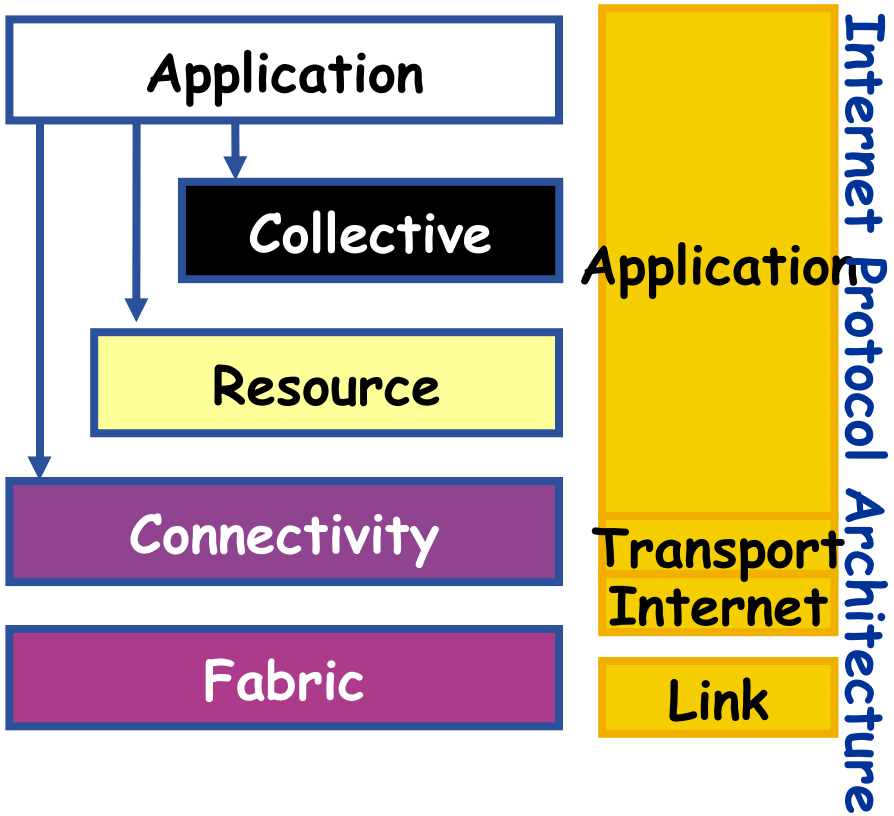


From Ian Foster's talk

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

The Grid stack

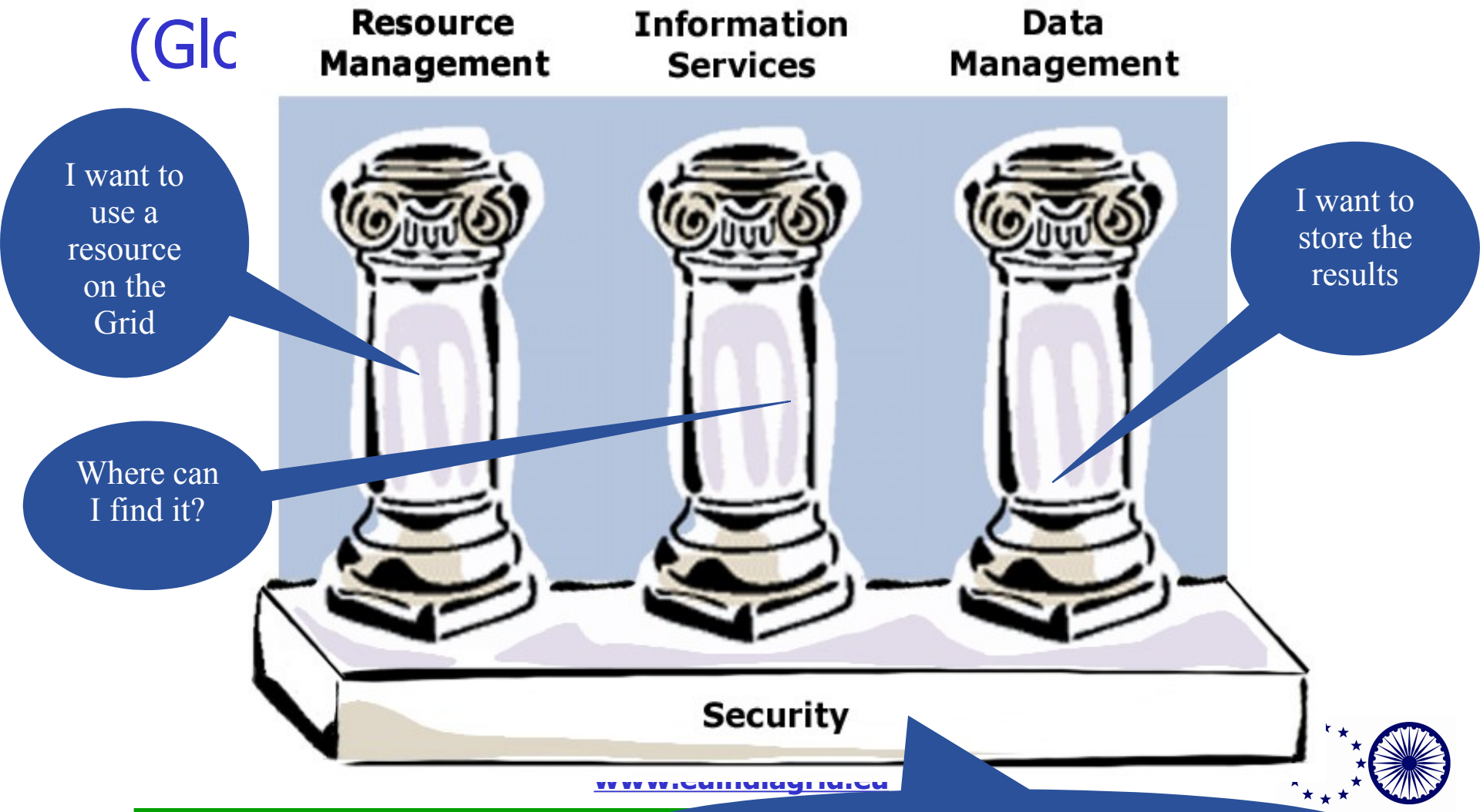
- Application layer
  - Grid programs
- Collective layer
  - Resource Co-allocation
  - Data Replica Management
- Resource layer
  - Resource Management
  - Information Services
  - Data Access
- Connectivity layer
  - Grid Security Infrastructure
  - High-performance data transfer protocols
- Fabric layer
  - the hardware: computers (parallel, clusters..), data storage servers



EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Grid foundations

- Defined by the Globus (http://globus.org) (Glc



EU-IndiaGrid (RI-031834) is funded by the Euro

Middleware available

- Globus Toolkit (Argonne+ISI)
- LCG/gLite (from EU projects)
- Gridbus (Melbourne)
- Unicore... (Germany)
- Garuda (India)
- And many other...

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

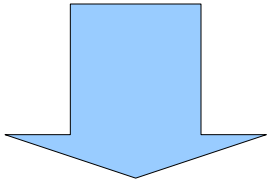
Applications for GRID computing..

- Computation intensive
  - Interactive simulation (climate modeling)
  - Large-scale simulation and analysis (atomistic simulations)
  - Engineering (parameter studies, optimization model)
- Data intensive
  - Experimental data analysis (e.g., H.E.P.)
  - Image & sensor analysis (climate)
- Distributed collaboration
  - Online instrumentation (microscopes, x-ray) Remote visualization (climate studies, biology)

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Common factor..

The size and/or complexity of the problem requires that people in several organizations collaborate and share computing resources, data, instruments



VIRTUAL ORGANIZATIONS

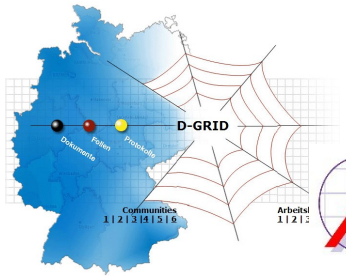
EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Virtual Organization

- Distributed resources and people
- Linked by networks, crossing admin domains
- Sharing resources, common goals
- Dynamic
- Fault Tolerant..

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

A few international GRID project.



ACCESS GRID

Open Science Grid

grangenet

National Grid
NG
SINGAPORE

GRIP
UK
Particle
Physics

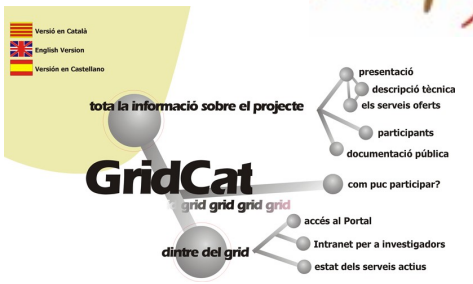
Grid3
2003

iVDgL

CERN
openlab for DataGrid applications
Developing Solutions for the Data-Intensive Science of the Large Hadron Collider

LCG

West Grid
NORDUGRID
Grid Solution for Wide Area Computing and Data Handling



INFN
GRID

grid.it
project

Grid Consortium Japan

eGee
Enabling Grids for
E-science in Europe

GridLab

NAREGI
超高速コンピュータ網形成プロジェクト
National Research Grid Initiative
Grid Applications
Grid Middleware
Networking
国立情報学研究所グリッド研究開発推進拠点 NII -The National Institute of Informatics

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Grids vs. HPC

- Not an "either/or" question
  - Each addresses different needs
  - Each are part of an integrated solution
- Grid strengths
  - Coupling necessarily distributed resources instruments, software, hardware, archives, and people
  - Eliminating time and space barriers
  - remote resource access and capacity computing
- Grids are not a cheap substitute for capability
- HPC Highest performance computing strengths
  - Supporting foundational computations
  - terascale and petascale "nation scale" problems
  - Engaging tightly coupled computations and teams

Key is easy access to resources in a transparent way

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme

Wrap-up

- Modern Science requires a large amount of computing resources
- GRID computing and HPC are now fundamental tools for scientific research
- The challenge is now to build/use the infrastructure that fits at best your computational requirements.
- HPC and GRID computing are not mutually exclusive but can be both used to address computational resources in a transparent way.
- We will focus here on the Euindiagrid infrastructure and how to use it. www.euindiagrid.eu

EU-IndiaGrid (RI-031834) is funded by the European Commission under the Research Infrastructure Programme