DEISA Middleware Strategies

Sophia Antipolis, October 13, 2005

V. Alessandrini
www.deisa.org

Table of contents

The DEISA Consortium

Distributed European Infrastructure for Supercomputing Applications

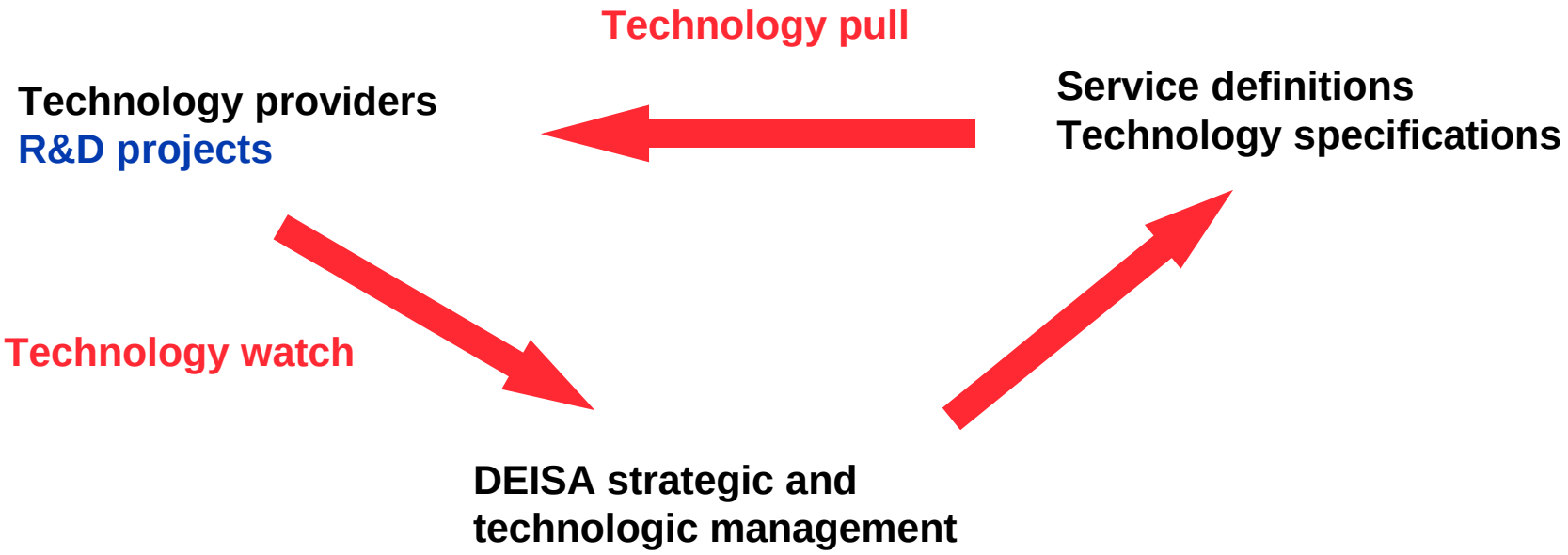
DEISA objectives

- *To enable Europe's terascale science by the integration of Europe's most powerful supercomputing systems.*
- *Enabling scientific discovery across a broad spectrum of science and technology is the only criterion for success*
- **DEISA is an European Supercomputing Service built on top of existing national services. The integration of national facilities and services using Grid technologies is expected to add substantial value to the existing infrastructures.**
- **DEISA deploys and operates a persistent, production quality, distributed supercomputing environment with continental scope.**
- **Main focus is High Performance Computing (HPC).**

Basic requirements and strategies for the DEISA research Infrastructure

- **Fast deployment of a persistent, production quality, grid empowered supercomputing infrastructure with continental scope.**
- **The European supercomputing service built on top of existing national services requires reliability and non disruptive behavior.**
- **User transparency : users should not be disturbed by technology overhead. Grid technologies should work seamlessly in the background.**
- **Application transparency : applications are part of the corporate wealth of virtual organizations, and they should be portable and independent (as much as possible) of the underlying Grid technologies.**
- **Top-bottom approach: technology choices follow from the business and operational models of our virtual organization. DEISA technology choices are pragmatic and fully open. There is practically no « DEISA specific middleware ».**

The technology cycle



Ongoing actions: WAN GPFS (IBM), Multi-cluster batch processing (IBM)
Planned actions: GPFS for non-IBM systems (IBM), Co-allocation (Platform)

Running the infrastructure

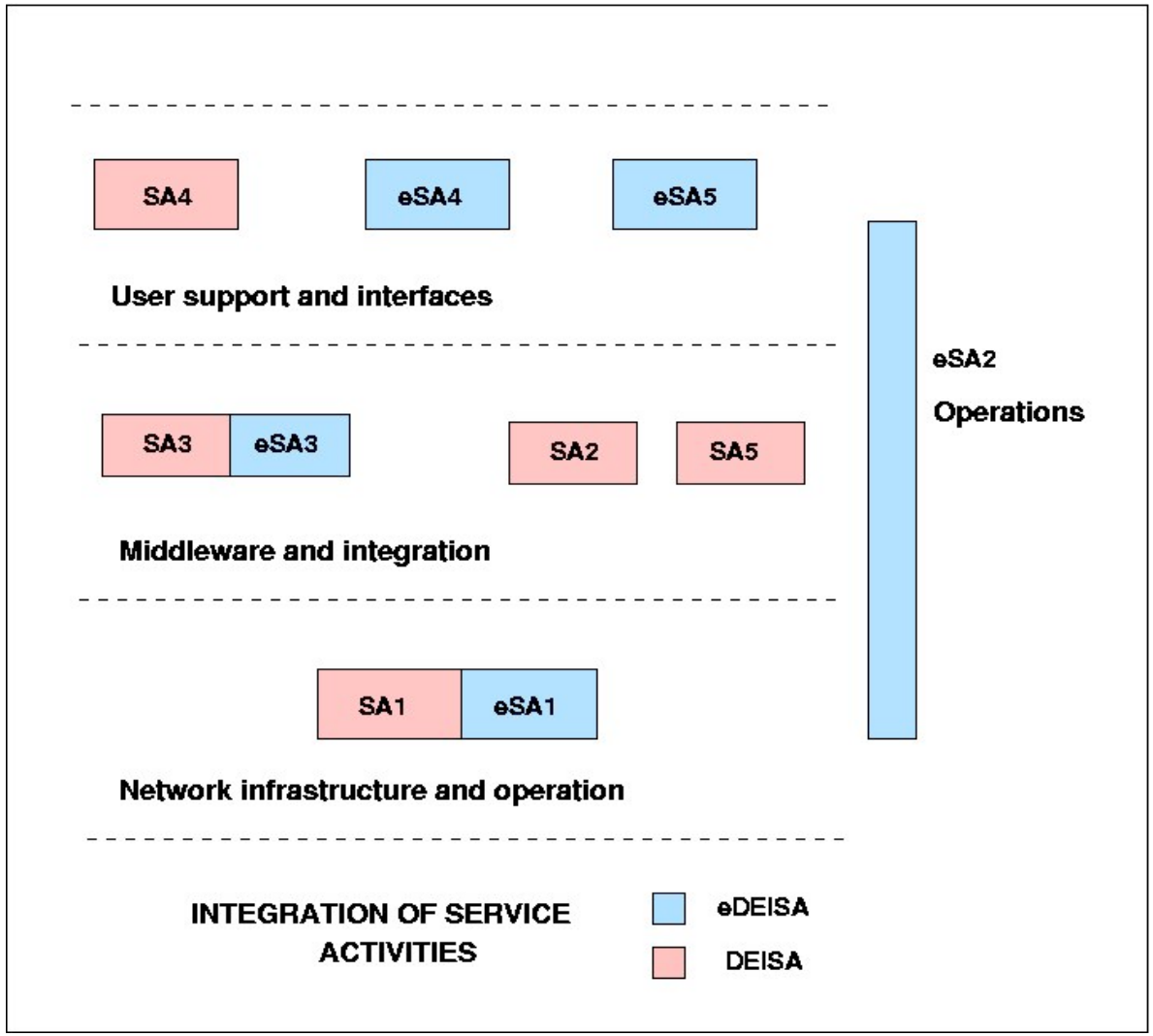
- ***System Integration, deployment and operation of the infrastructure*:**
- **SA in Networking**
- **SA in Global File Systems**
- **SA in Middleware and Ressource Management**
- **SA in User Support**
- **SA in Security**
- **JRA in Heterogeneous resource management (JRA7)**

- ***Applications***

- **Scientific JRAs (JRA1 to JRA6)**
- **The Applications Task Force (new)**
- **The DEISA Extreme Computing Initiative (new)**

DEISA Service Activities

Distributed European Infrastructure for Supercomputing Applications



DEISA

SA1: Networking
SA2: Global File Systems
SA3: Middleware
SA4: User Support
SA5: Security

eDEISA

eSA2: Operations
eSA4: Applications Enabling
eSA5: Visualization and Portals

The DEISA supercomputing Grid: a layered infrastructure

- **Inner layer: a distrubuted super-cluster resulting from the deep integration of similar IBM AIX platforms at IDRIS, FZ-Julich, RZG-Garching and CINECA (phase 1) then CSC and ECMWF (phase 2). It looks to external users as a single supercomputing platform.**
- **Outer layer: a heterogeneous supercomputing Grid:**
  - **IBM AIX super-cluster (IDRIS, FZJ, RZG, CINECA, CSC) close to 24 Tf**
  - **BSC, IBM PowerPC Linux system, 40 Tf**
  - **LRZ , Linux cluster (2.7 Tf) moving to SGI ALTIX system (33 Tf in 2006, 70 Tf in 2007)**
  - **SARA, SGI ALTIX Linux cluster, 2.2 Tf**
  - **ECMWF, IBM AIX system, 32 Tf**
  - **HLRS, NEC SX8 vector system, close to 10 Tf**

THE DEISA SUPERCOMPUTING GRID

Distributed European Infrastructure for Supercomputing Applications

AIX distributed super-cluster

Vector systems (NEC, …)

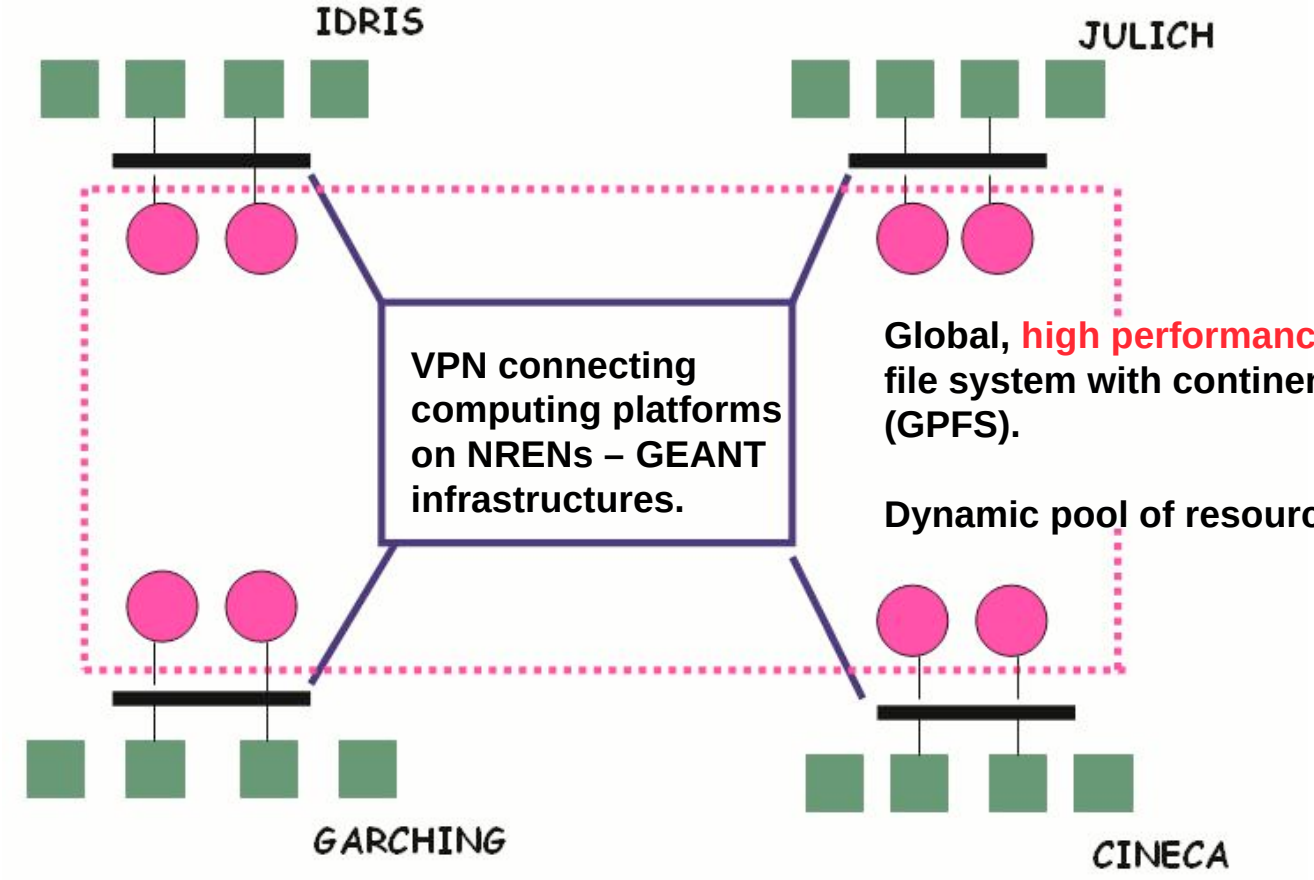
Linux systems (SGI, IBM, …)

HPC and Grid Computing

- **Grid computing is not always HPC…**
- **MPI latencies are boosted in WANs from a few microseconds to milliseconds…**
- **… *because the speed of light is not big enough !***
- **Deploying tightly coupled parallel aplications in large scale Grids is not HPC**
- **Direct Grid computing works best for (almost) embarassingly parallel applications, or coupled software modules with limited real time communications**
- **It is better to run large, tightly coupled applications in a single platform.**
- **DEISA implements this requirement by rerouting jobs and by balancing the computational workload at a Eurooean scale (a bandwidth issue, not a latency issue).**

A - The DEISA super-cluster (phase 1)

Distributed European Infrastructure for Supercomputing Applications



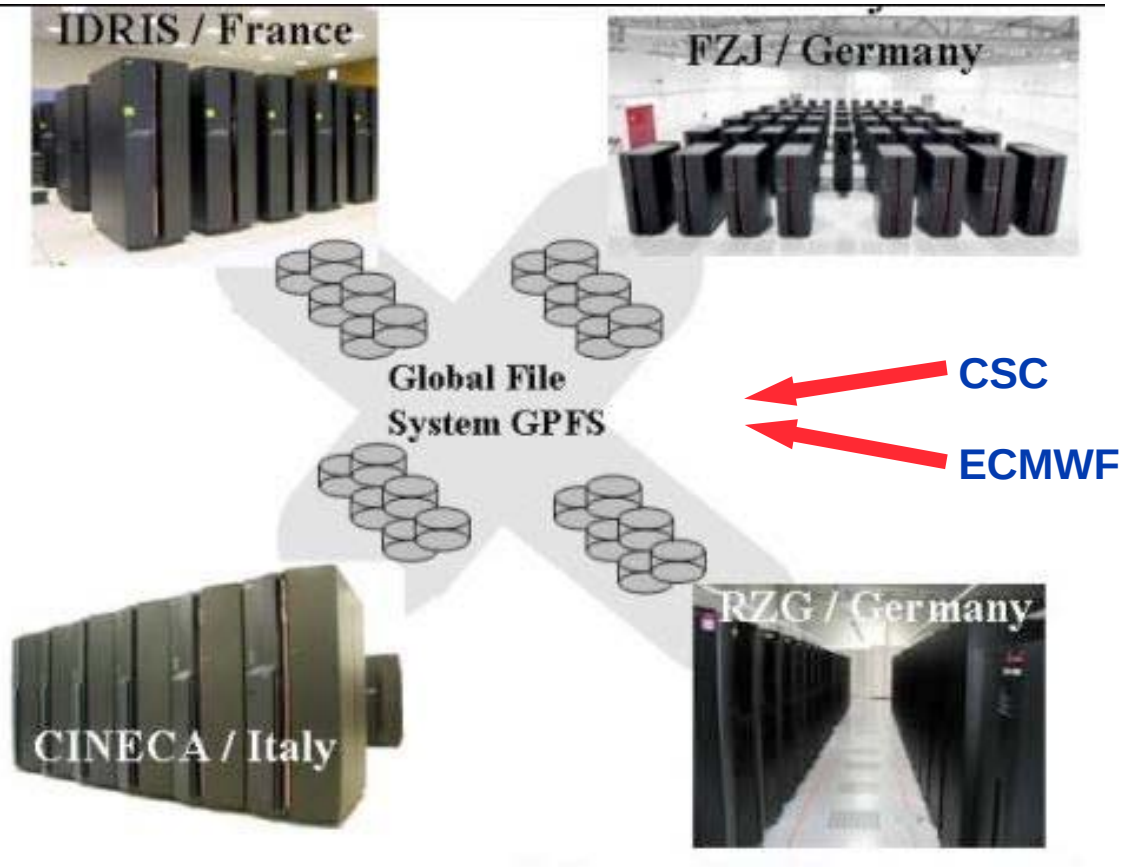
Global, high performance, distributed file system with continental scope (GPFS).

Dynamic pool of resources.

A - Operational model

- **DEISA enables job migration across sites (also transparent to end users). Exceptional resources for very demanding applications are made available by the operation of the global resource pool. We are load balancing computational workload at a European scale.**
- **Huge, demanding applications can be run "as such".**
- **For absolute portability, a Common Production Environment has been deployed.**
- **With this operational model, the DEISA super-cluster is not very different from a "true" monolithic European supercomputer (which must be partitioned in any case for fault tolerance and QoS).**
- ***The main difference comes from the coexistence of several independent administration domains*. This requires, as in TeraGrid, coordinated production environments.**

AIX SUPER-CLUSTER, September 2005



Distributed European Infrastructure for Supercomputing Applications

Services:

High performance datagrid via GPFS
Access to remote files use the full available network bandwidth

Job migration across sites
Used to load balance the global workflow when a huge partition is allocated to a DEISA project in one site

Common Production Environment

Full production status of dedicated 1 Gb/s network
GPFS : Full production IDRIS and RZG, FZJ and CINECA to follow immediately
JOB MIGRATION: test status in all sites

DEISA Heterogeneous Grid Services

- **Workflow management: based on UNICORE plus further extensions and services coming from DEISA's JRA7 and other projects (UniGrids, …)**
- **Global data management: a well defined architecture implementing extended global file systems on heterogeneous systems, fast data transfers across sites, and hierarchical data management at a continental scale.**
- **Co-scheduling, needed to support Grid applications running on the heterogeneous environment.**
- **Science Gateways and portals: specific Internet interfaces to hide complex supercomputing environments from end users, and facilitate the access of new, non traditional, scientific communities.**

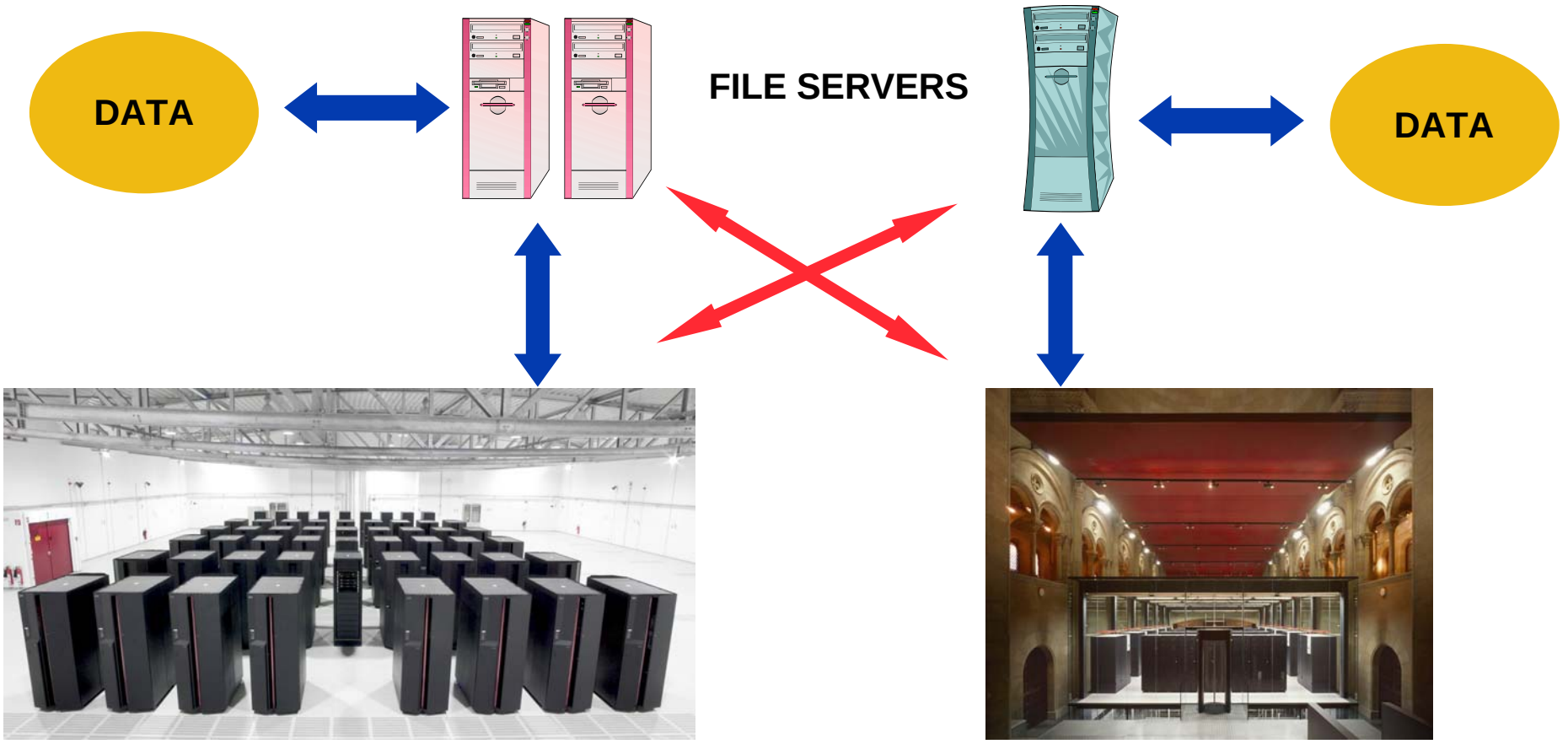
Workflow applications (UNICORE)

Distributed European Infrastructure for Supercomputing Applications

USER

Global data management

Distributed European Infrastructure for Supercomputing Applications



Co-scheduling for coupled applications

Distributed European Infrastructure for Supercomputing Applications

USER

Portals (Science Gateways)

- Similar concepts as TeraGrid's Science Gateways
- Needed to enhance the outreach of supercomputing infrastructures
- Hiding complex supercomputing environments from end users, providing discipline specific tools and support, and moving in some cases towards community allocations for anonymous users.
- There is already work done by DEISA on Genomics and Material Sciences portals
- Intense brainstorming on the desing of a global strategy, if possible interoperable with TeraGrid's Science Gateways

C. Enabling science

Distributed European Infrastructure for Supercomputing Applications

- Initial, « early users » program: a number of Joint Research Activities integrated in the project from the start.
- Moving towards « exceptional users » : the DEISA Extreme Computing Initiative

Distributed European Infrastructure for Supercomputing Applications

Activity	Scientific program	Partners	Leader
JRA1	Enabling Material Sciences CPMD codes, portals	RZG	Hermann Lederer, RZG
JRA2	Computational environment for applications in Cosmology	EPCC	Gavin Pringle, EPCC
JRA3	Enabling the TORB Plasma Physics code	RZG	HermannLederer, RZG
JRA4	Life sciences: genomic and eHealth Applications	IDRIS, (BSC)	VA , IDRIS -> BSC
JRA5	CFD in the automobile industry	CINECA, CRI	Roberto Tregnago, CRI
JRA6	Coupled applications: Astrophysics, Combustion, Environment	IDRIS (HLRS)	Gilles Grasseau, IDRIS

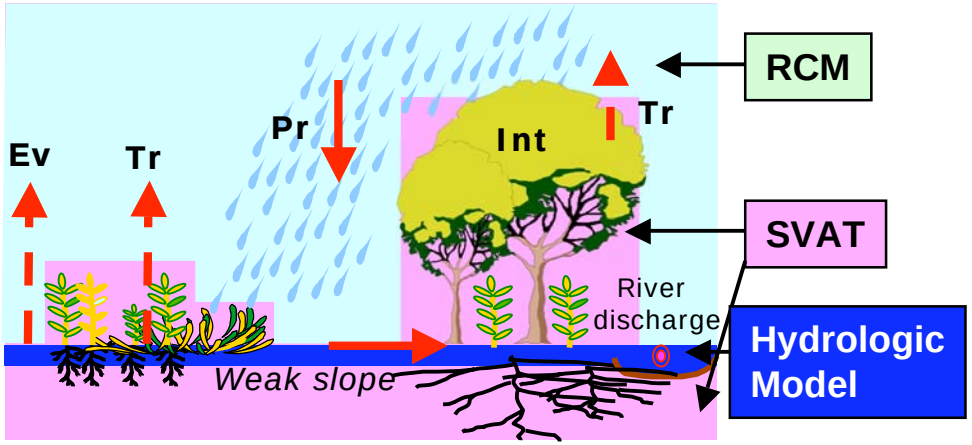
EARLY USERS PROGRAM

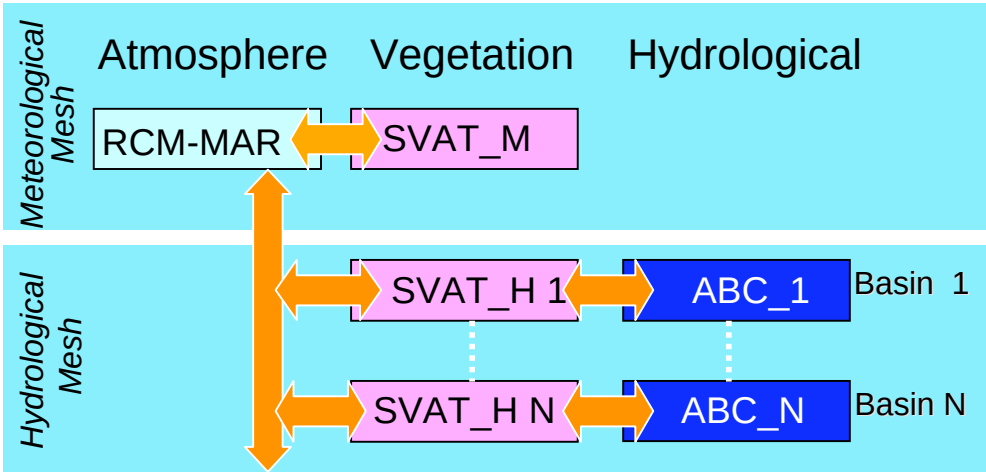
Coupled applications : Environment

Distributed European Infrastructure for Supercomputing Applications

- **Leaders: Michel Vauclin and Christophe Messager (LTHE)**
- **Evaluate the importance of the water cycles between:**
  - **the atmosphere (RCM),**
  - **the soil / vegetation (SVAT),**
  - **the hydrological basins (hydrologic model)**

  over the West Africa.
- **Add easily new basins — the coupling architecture is modular and extensible.**
- **Collaboration with the international AMMA project (Africa Monsoon Multidisciplinary Analysis) by adding a new basin (Ouémé) and a new hydrologic model (dedicated to humid catchments).**

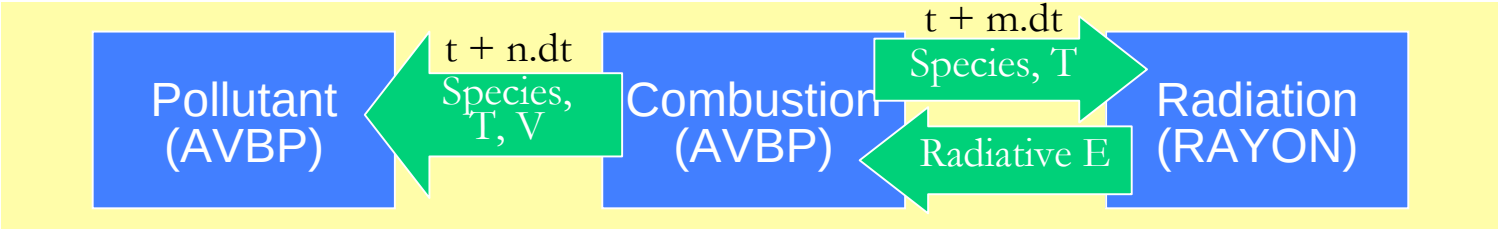




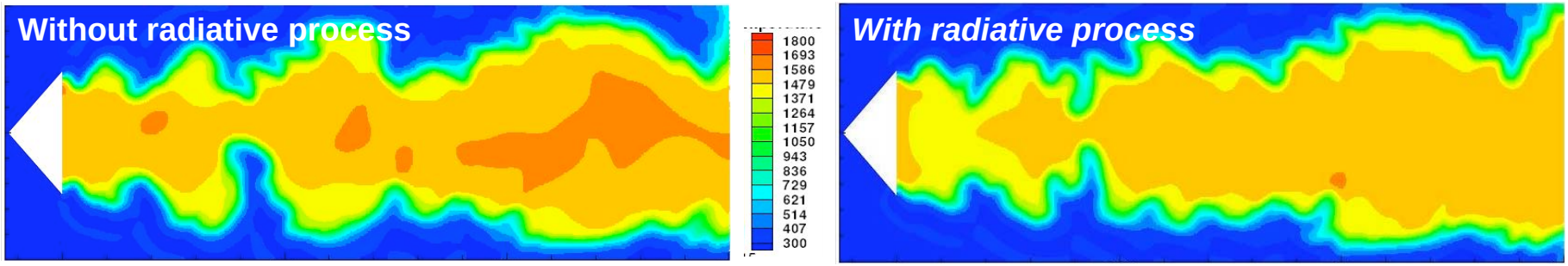
Coupled applications: combustion

Distributed European Infrastructure for Supercomputing Applications

- **Leader: Denis Veynante (EM2C)**
- **Develop and optimize the efficiency of the combustion and reduce pollutant emissions in industrial systems (engines, energy production, industrial furnaces, …)**
- **Take account of the radiative process in the combustion (rarely considered in previous works)**
- **Coupling description (3 physical phenomena → 3 coupled codes):**



- **First simulations about the impact of radiative process on the flame behaviour**

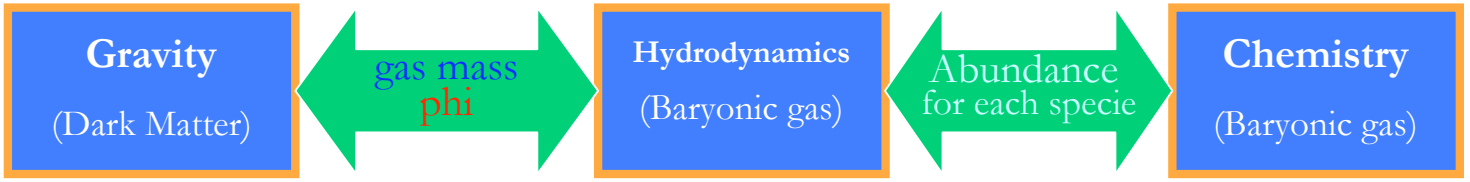


The temperature field is largely modified: the temperature decreases and the field is more homogeneous when the radiative process runs

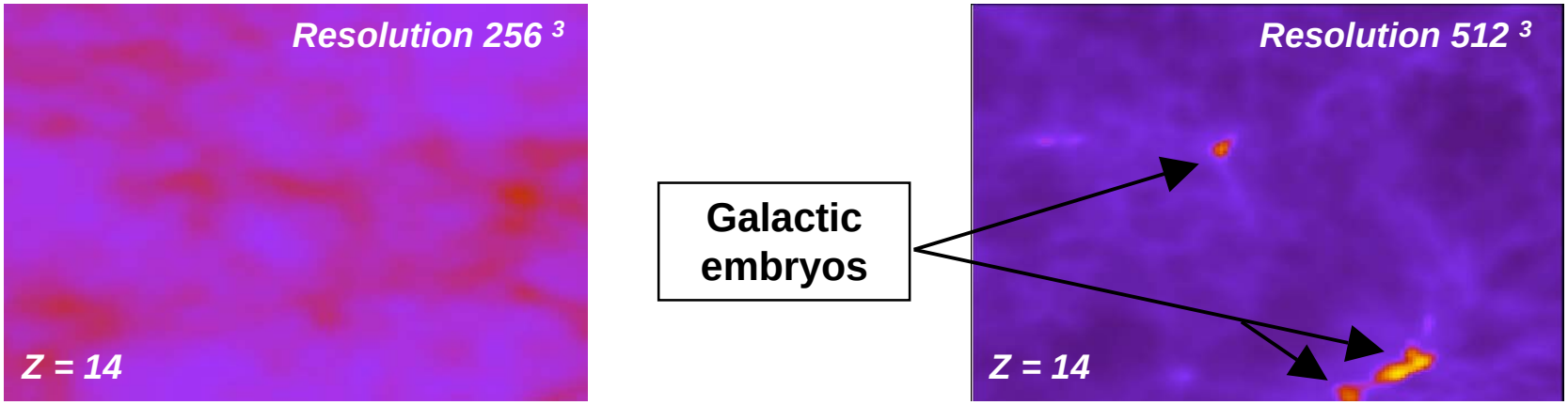
Coupled applications: astrophysics

Distributed European Infrastructure for Supercomputing Applications

- **Leader: Jean-Michel Alimi (LUTH)**
- **Modeling of the galaxy formation requires to take account of many physical processes.**
- **3 main physical phenomena are currently considered → 3 coupled codes:**



- **Impact of the resolution in the process of galaxy formation**
  Gas temperature at 240 millions of years after the Big Bang (z = 14):



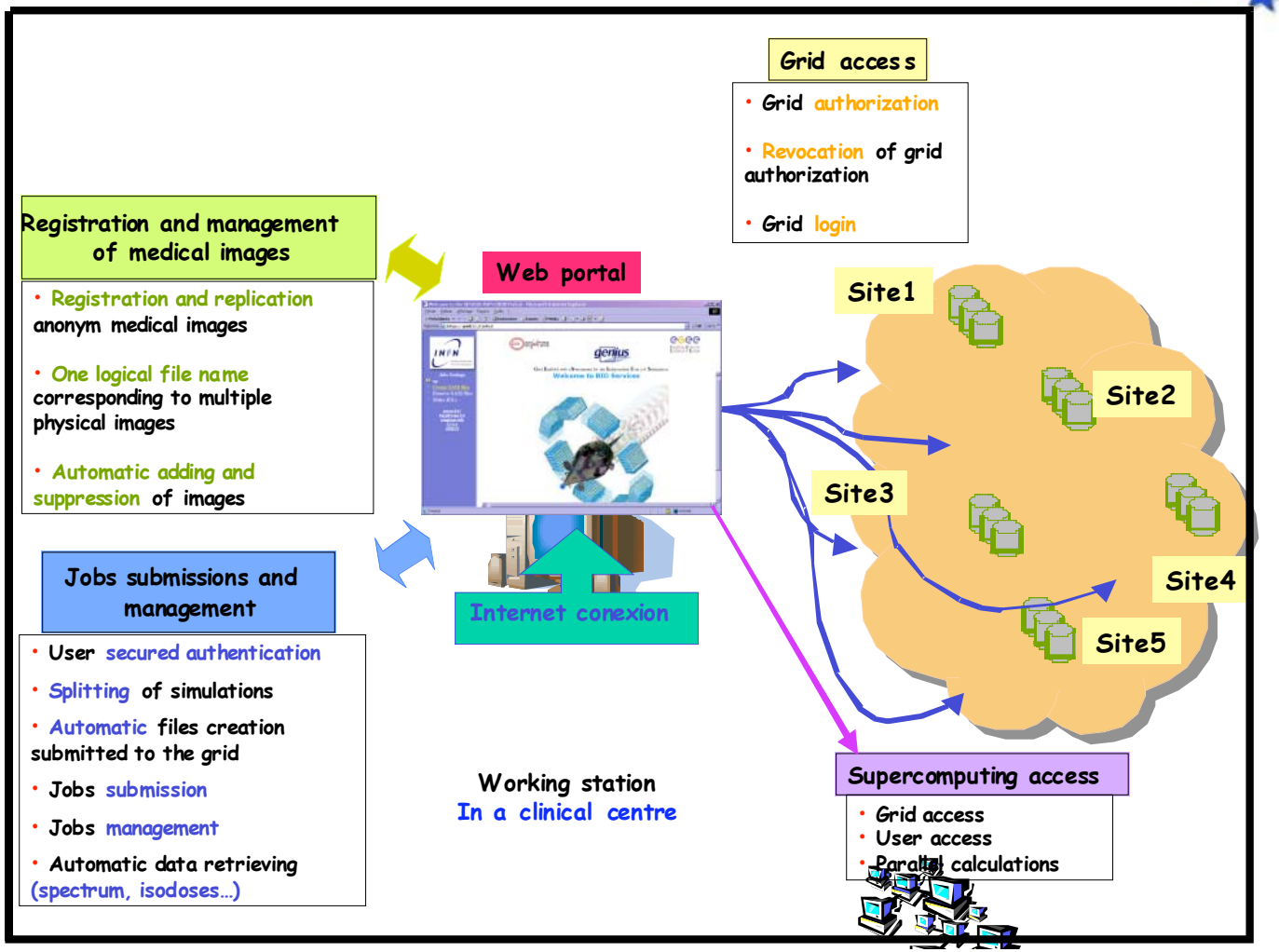
The high temperature zones (galactic embryos) appear later (470 millions of years after the Big Bang — Z=10) in the 256^3 resolution.

Life Sciences

- Genomics computational environment on AIX super-cluster: databases, fine tunning of multithreaded BLAST.
- InfoBioGen portal: transparent rerouting of jobs to the DEISA platform.
- Leading applications
  - **INSERM: *Identification of new human mithocondrial proteins* (ready for production)**
  - **INRA : *Large scale microbial genome reanotation***
  - **BSC: *Prediction of protein interactions***
- Deployment, with BSC, of a high performance *heterogeneous* platform for bioinformatics applications

Radiation Therapy planning (joint application with EGEE)

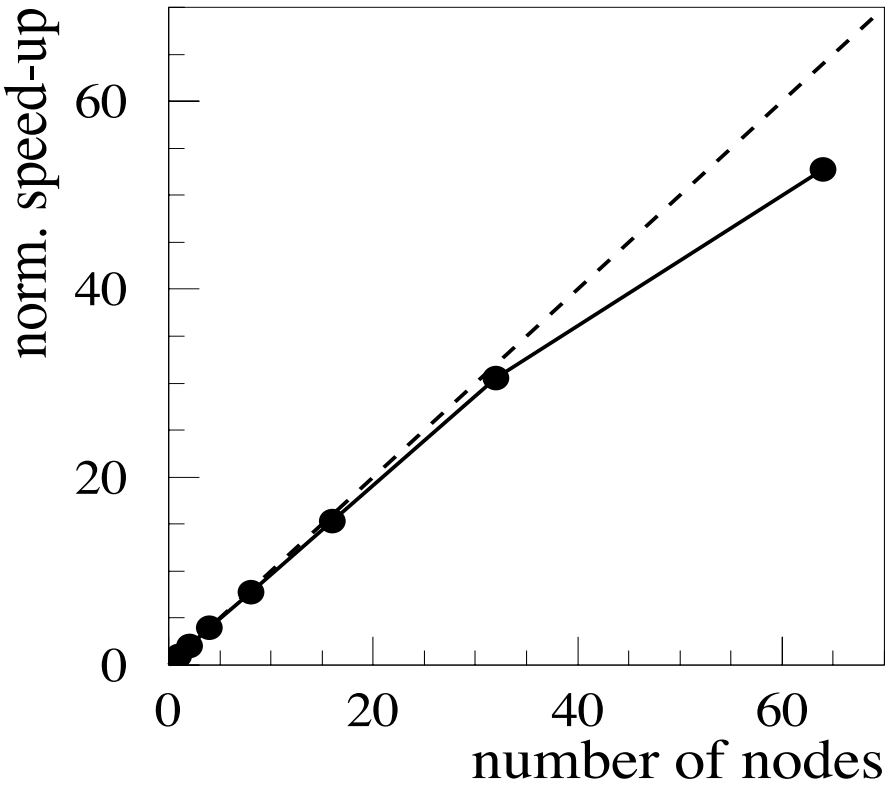
Distributed European Infrastructure for Supercomputing Applications



Distributed European Infrastructure for Supercomputing Applications

Proposals from Plasma Physics

Extreme Gyrokinetic Turbulence Simulations



64 nodes = 2048 processors

The nonlinear particle-in-cell code TORB uses a Monte Carlo particle approach to simulate the time evolution of turbulent field structures in fusion plasmas
(J. Nuehrenberg, IPP, Greifswald & L. Villard, CRPP, Lausanne)

Within DEISA, TORB has been improved for extreme scalability at IBM system at ECMWF:

On 2048 procs:

Speedup = 1680 Parallel efficiency = 82%
Sustained performance = 1.3 TF

C. The Extreme Computing Initiative

- Identification, deployment and operation of a number of « flagship » applications in selected areas of science and technology
- Applications must rely on the DEISA Supercomputing Grid services (application profiles have been clearly defined). They will benefit from exceptional resources from the DEISA pool.
- Applications are selected on the basis of scientific excellence, innovation potential, and relevance criteria.
- European call for proposals: April 1st -> May 30, 2005
- Evaluation: Juin -> September 2005.

C. Enabling new applications: ATASKF

- Creation, in April 2005, of the Aplications Task Force (ATASKF), to support the Extreme Computing initiative.
- The ATASKF carries out a prospective action with the European scientific community, to support the design on new, leading applications.
- The ATASKF provides guidance to find the best fit between the user requirements and the distributed supercomputing environment
- The ATASKF determines the actions to be taken to move applications into the « estreme computing » domain.
- We have 53 Extreme Computing proposals.

Extreme Computing proposals

- Bioinformatics 4
- Biophysics 3
- Astrophysics 11
- Fluid Dynamics 5
- Materials Sciences 11
- Cosmology 3
- Climate, Environment 5
- Quantum Chemistry 5
- Plasma Physics 2
- QCD, Quantum computing 3

Evaluation and allocation of DEISA resources

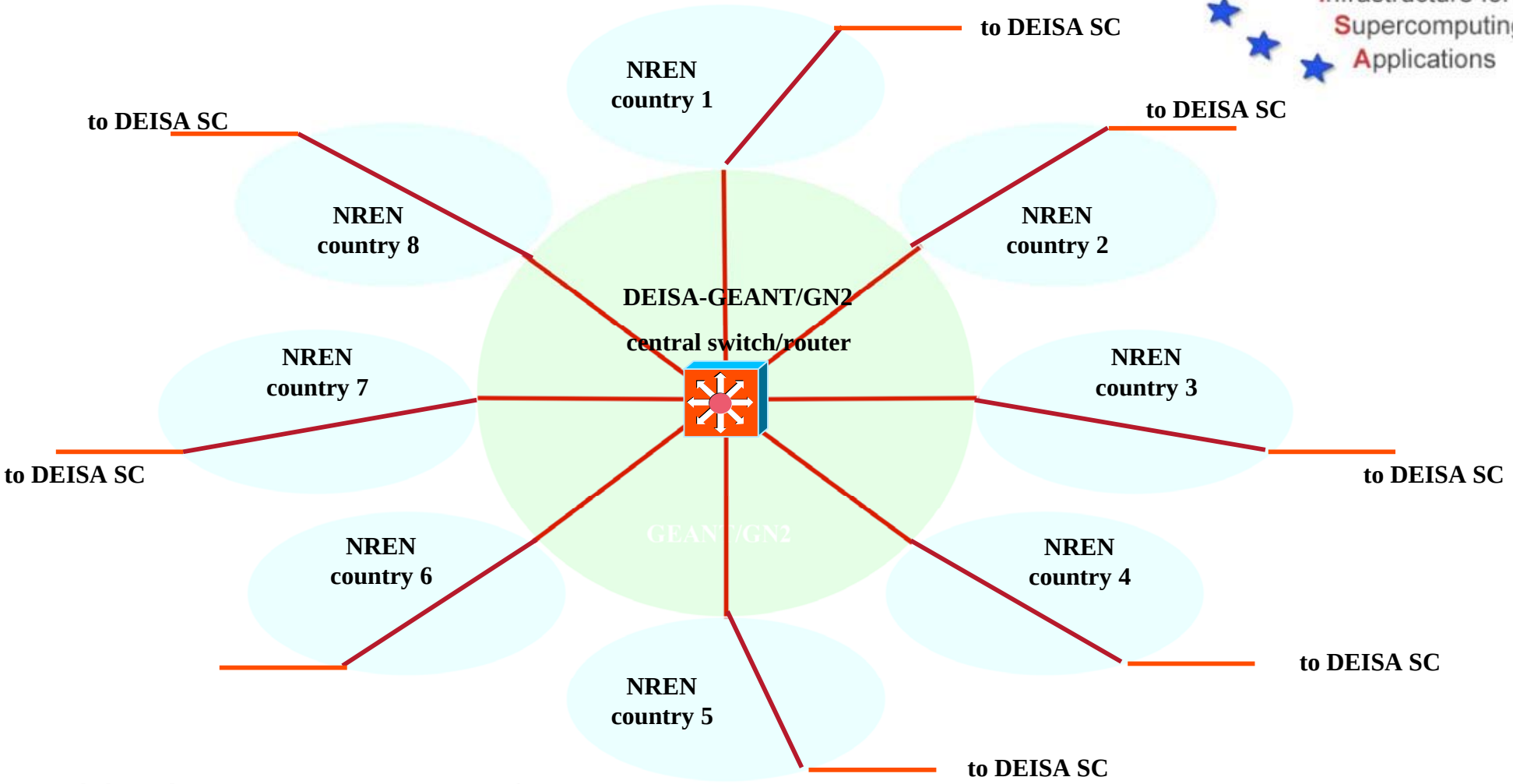
- National evaluation committees evaluate the proposals and determine priorities.
- On the basis of this information, the DEISA consortium examines how the applications map to the resources available in the DEISA pool, and negotiates internally the way the resources will be allocated and the final priorities for projects.
- Exceptional DEISA resources are allocated – as in large scientific instruments – at well defined time windows (negotiated with the users).
- DECI proposals have been evaluated by national committees. Final selection made in September 2005.
- 27 « grand challenge » projects have been retained for operation in 2005-2006 (out of 53 proposals)

DEISA Network, phase 2

- **10 Gb/s (or more) dedicated bandwidth among nine (or more) supercomputers in the hetergeneous supercomputing Grid.**
- **High performance access to remote data repositories, fast data transfers among sites.**
- **Required by distributed aplications, workflow simulations, distributed visualization grids, etc.**
- **Ongoing discussions with GN2 and the NRENs. DEISA-GN2 joint meeting held on July 4, 2005.**

Logical view of DEISA network

Distributed European Infrastructure for Supercomputing Applications



dedicated DEISA supercomputer connection

Conclusions

- **DEISA adopts Grid technologies to integrate national supercomputing infrastructures, and to provide an European Supercomputing Service.**
- **Service activities are supported by the coordinated action of the national center's staffs. DEISA operates as a virtual European supercomputing centre.**
- ***The big challenge we are facing is enabling new, first class computational science.***
- ***Integrating leading supercomputing platforms with Grid technologies may enable a new research dimension in Europe.***