The Supercomputing Grid and the European HPC initiative.

Distributed European Infrastructure for Supercomputing Applications

Towards world class global HPC infrastructures in Europe

Victor Alessandrini
IDRIS - CNRS
va@idris.fr

Issues:

Distributed European Infrastructure for Supercomputing Applications

- What is so different about HPC?
- How does HPC relate to Grid computing?
- How is DEISA enhancing HPC services in Europe?
- How is DEISA preparing the ground for new HPC initiatives?
- How should a global HPC infrastructure in Europe look like?
- How could such a global European HPC eInfrastructure be implemented?
  - New intiatives (HPCEUR)
  - Funding and operational models
  - Speculations about possible options and ways for the future

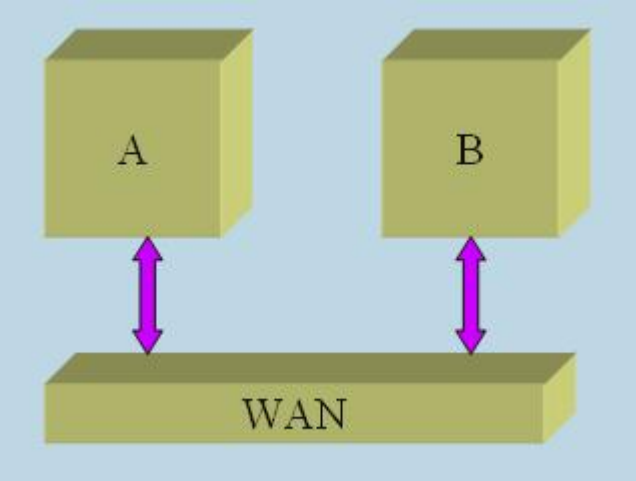
About HPC

Distributed European Infrastructure for Supercomputing Applications

- **Dealing with large complex systems requiring exceptional computational resources. For algorithmic reasons, required resources grow much, much faster than the systems size and complexity.**
- **Dealing with huge, datasets, involving large files. Typical datasets in SC centres are of the order of several PBytes. Datasets are active (not just archiving, reads are almost as frequent as writes)**
- **Little usage of commercial or public domain packages. Most applications are virtual organization corporate codes incorporating specialized know how (to make a difference in the international competition). This is why specialized user support is so important.**
- **Codes are fine tuned and targeted for a relatively small number of well identified computing platforms. They are extremely sensitive to the production environment.**
- **For any kind of data movement, transit time is $T = L + (\text{Packet size})/B$, where L is latency and B is bandwidth.**
- **The main requirement for high performance is bandwidth (from processor to memory, from processor to processor, from node to node, from system to system). Internal optimization of communications - high bandwidth and specialized tricks to hide latencies - are the main differences between supercomputers and clusters.**

HPC and Grid computing

Distributed European Infrastructure for Supercomputing Applications



Problem: the speed of light is not big enough

Finite signal propagation speed boosts message passing latencies in a WAN from a few microseconds to tens of miliseconds (if A is in Paris and B in Helsinki)

If A and B are two halves of a tightly coupled complex system, communications are frequent and the enhanced latencies will kill performance.

Grid computing works best for embarassingly parallel applications, or coupled software modules with limited communications.

Example: A is an ocean code, and B an atmospheric code. There is no bulk interaction. Systems interact only at the ocean-atmosphere boundary, with limited communications.

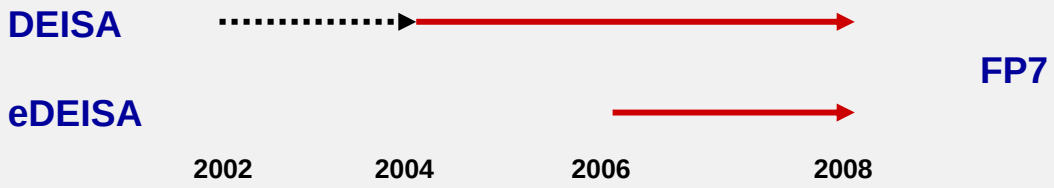
Large, tightly coupled parallel applications should be run in a single platform. This is why we still need high end supercomputers.

DEISA implements this resuirement by rerouting jobs and balancing the computational workload at a European scale.

About DEISA

Distributed European Infrastructure for Supercomputing Applications

- ***DEISA is a supercomputing Grid infrastructure, whose objective is to enhance Europe's capability computing and terascale science by the integration of Europe's most powerful supercomputing systems.***
- **DEISA is an European Supercomputing Service built on top of existing national services. This service is based on the deployment and operation of a persistent, production quality, distributed supercomputing environment with continental scope.**
- **The main objective is to add substantial value to existing HPC infrastructures, by the integration of national facilities and services, together with innovative operational models.**
- **The main focus is High Performance Computing (HPC) and Extreme Computing applications that cannot by supported by the isolated national services.**
- **Many of the services being deployed are will enable the efficient operation of future shared European petascale systems.**



How is DEISA enhancing HPC services in Europe?

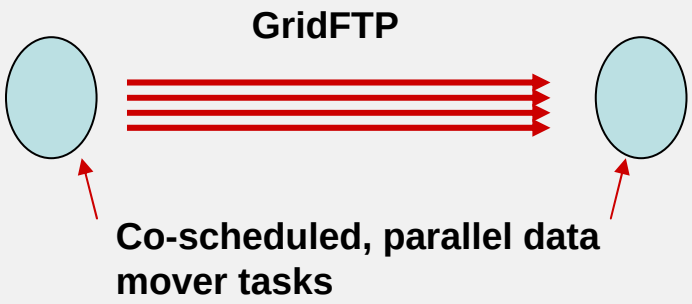
Distributed European Infrastructure for Supercomputing Applications

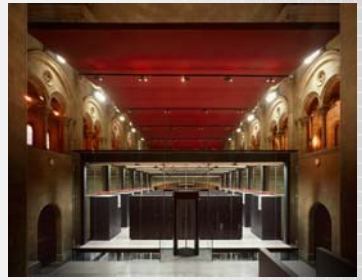
- **Running larger parallel applications in individual sites, by a cooperative reorganization of the global computational workload on the whole infrastructure.**
- **Enabling workflow applications with UNICORE (complex applicaions that are pipelined over several computing platforms)**
- **Enabling coupled multiphysics Grid applications (when it makes sense)**
- **Providing a global data management service whose primordial objectives are:**
  - **Integrating distributed data with distributed computing platforms**
  - **Enabling efficient, high performance access to remote datasets (with Global File Systems and stripped GridFTP). We believe that this service is critical for the operation of (possible) future European petascale systems**
  - **Integrating hierarchical storage management and databases in the supercomputing Grid.**
- **Deploying portals as a way to hide complex environments to new users communities, and to interoperate with another existing grid infrastructures.**

Basic DEISA data management services

Distributed European Infrastructure for Supercomputing Applications

- GPFS (Global Parallel File System) enables high performance remote IO. Remote files are seen as local files, and WANs do not spoil performance. Of course, remote data is moved to the local system, but the data movement is implicit.
- GPFS is not universal. It s an IBM product, and it does not work on all systems.
- Therefore, DEISA will deploy also explicit high performance transfers of large datasets, using GT4 striped GridFTP (which relies on parallel activation of multiple TCP streams to enhance performance)
- We said before that datasets are active. Replication of entire datasets is not practical (coherence problems). Users need to work with their traditional data repositories
- Therefore, moving large data files efficiently among remote platforms is needed to run large applications on remote systems.



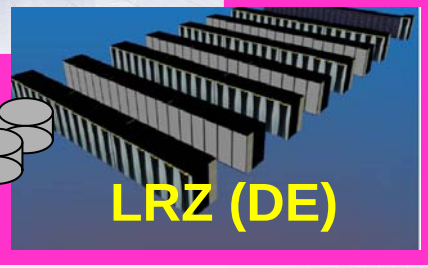


DEISA Global File System integration in 2006
(based on IBM's GPFS)

Distributed European Infrastructure for Supercomputing Applications

AIX IBM domain

Linux SGI



High Performance Common Global File System
various architectures / operating systems
High bandwidth (up to 10 Gbit/s)

LINUX Power-PC

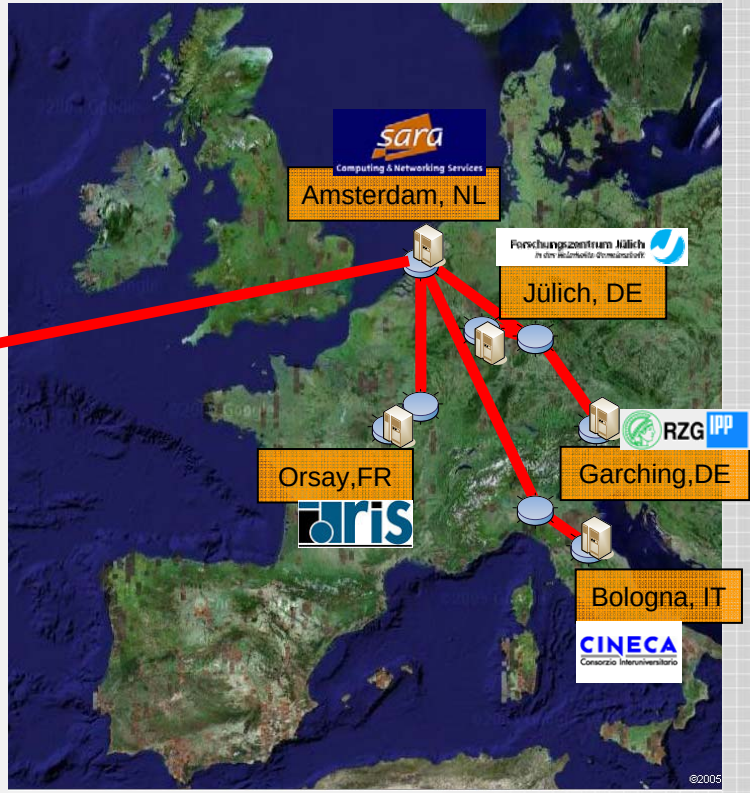
Global File System Interoperability demo during Supercomputing Conference 2005 in Seattle

Distributed European Infrastructure for Supercomputing Applications

American and European supercomputing infrastructures linked: bridging communities with scalable, wide-area global file systems

TeraGrid Sites

DEISA Sites



How should a global HPC infrastructure in Europe look like?

Distributed European Infrastructure for Supercomputing Applications

- **Emerging concensus about a global, strongly coupled eInfrastructure « à la DEISA » integrating the current national terascale systems (Tier 2) with new generation petascale systems (Tier 1) into a unique service providing environment, with a unique operational model.**
- **Ideally, T1 should provide exceptional capability computational resources, with a significant performance difference with respect to T2. We expect at any time a small number of T1 systems.**
- **Ideally, T1 systems should be shared and European.**
- **A basic requirement is cooperative operation of T1 and T2 including high performance access to existing national data repositories. This is why the data management services that DEISA is deploying are highly relevant.**
- **Ideally, exceptional resources will be concentrated in a very limited number of sites, but services, know how, competence and leadership should remain distributed.**
- **The boundaries of this « HPC island » in a global European eInfrastructure ecosystem follow from its unique operational and service provisioning models. Interoperability with the rest of the ecosystem is a major priority that is already being addressed.**

HPCEUR – STRATEGY (T1) (taken from H. Pilcher-Clayton presentation at the Brussels HPC meeting on March 21, 2006)

Distributed European Infrastructure for Supercomputing Applications

- Develop a partnership between the funding agencies of individual European countries, the European Commission and industry
- Create a sustainable model for the provision of an HPC infrastructure at European level
- Establish a new HPC centre every 2 years (3 in all over the lifetime of FP7)
- Each HPC centre to be located in a hosting country, building on that country's existing national HPC infrastructure
- Each of the 3 HPC centres to be in different hosting countries
- DEISA
  - Each HPCEUR supercomputing centre to be fully integrated into DEISA
  - Provide a % of resources to DEISA
  - Benchmarking activity for HPCEUR included as a work package within e-DEISA
- Scientific case developed by scientists from France, Germany, Spain and the UK

T1 Funding models (taken from the preliminary minutes of parallel sessions B of the Brussels HPC workshop, March 21, 2006)

Distributed European Infrastructure for Supercomputing Applications

- The German model proposes European facilities which provide supercomputing services that are open to entire Europe. They should be established in an open competition between the European member states and they are to be integrated in the future European e-Science infrastructure. The resources should be financed through (several) core investments and additional refunds from selling computer time and services.
- The HPCEUR model proposes European facilities which provide supercomputer services for a predefined set of hosting countries. They are located in the HPCEUR core countries. The facilities are integrated in a global infrastructure. The resources should be financed through investments by the partners of HPCEUR.
- In the context of the HPCEUR option of establishing three European HPC supercomputing centres, it was pointed out that the requirements of all scientific disciplines may lead to European procurements where more than one platform is requested.

Comments

- **HPCEUR: partnership between funding agencies requires further discussions to agree on the order problem for the deployment of the three HPC centres**
- **The previous observation (more that one platform per procurement) opens the way to another option where the notion of "persistent European Centre" is substituted by the notion of "European computing platform". Indeed, assuming the existence of a strongly coupled global environment "à la DEISA", two platforms of the same procurement could be installed at different sites or countries. This strategy may allow to better spreading the capital investments among several nations at a given time, and it may help to resolve some conflicting issues that arise in the funding schemes discussed above, like the order problem.**
- **However, pushing too far in this direction brings us back to the case in which individual T1 platforms will be provided by the individual nations that can afford them.**
- **There are very delicate issues to be resolved if we want to maintain the initial motivation of new HPC initiatives: pooling national funding to significantly increase the leverage of HPC in Europe**

More comments

- **The present expectations of capital investments from the EU (10-15% of the total cost of T1 European computing infrastructures) are not sufficient to catalyze a strong drive towards a unique European infrastructure for the provision of petascale supercomputing services. (I really hope I am wrong here).**
- **One open option is that individual nations (or small groups of nations) will decide to deploy next generations T1 systems, with some sort of open policy towards the EU and the remaining nations.**
- **And DEISA? DEISA will try to evolve to a consortium of national HPC funding organizations (rather that supercomputing centres) as a way to implement sustainability of the infrastructure.**
- **If T1 sites are not fully European, the only option is to integrate them into the « à la DEISA » infrastructure as the national sites are integrated today in DEISA.**
- **If, instead, the T1 service provisioning in run by a unique, European wide consortium of national funding organizations similar to the « à la DEISA » consortium, a merge of both infrastructures to provide a unique HPC infrastructure for Europe should be considered in the long run.**

Conclusions

- **The EU maintains a profound structuring action through the eInfrastructures initiatives**
- **Large concensus on the necessity of a T1-T2 global HPC infrastructure**
- **Funding and operational models for T1 have not yet been completely resolved. This is a very challenging issue for the future of HPC in Europe . The eIRG contribution to this discussion could be important.**
- **Thank you for your attention !**