

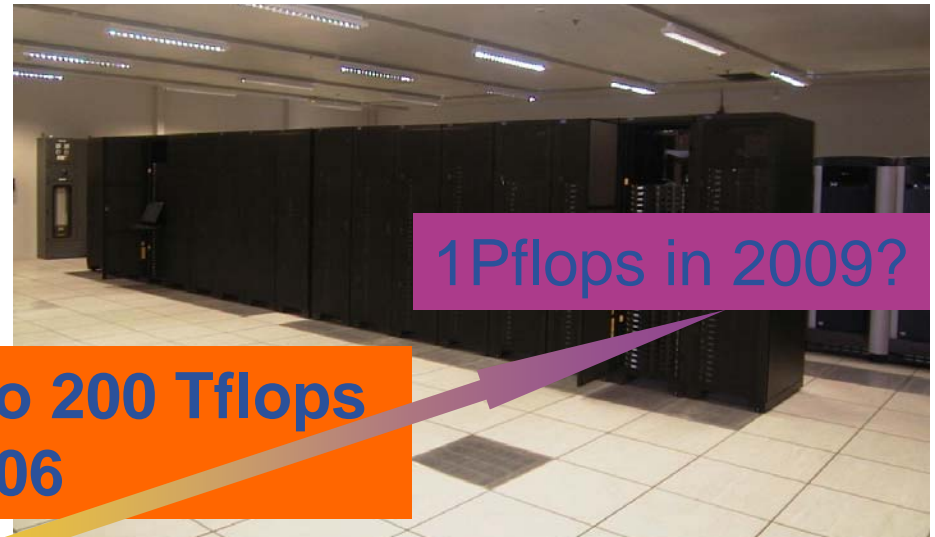


Enabling Grids for E-science

***EGEODE VO « Expanding GEosciences On DEMand »
Geocluster©: Generic Seismic Processing Platform
Gaël Youinou
Compagnie Générale de Géophysique (CGG, France)***

www.eu-egee.org

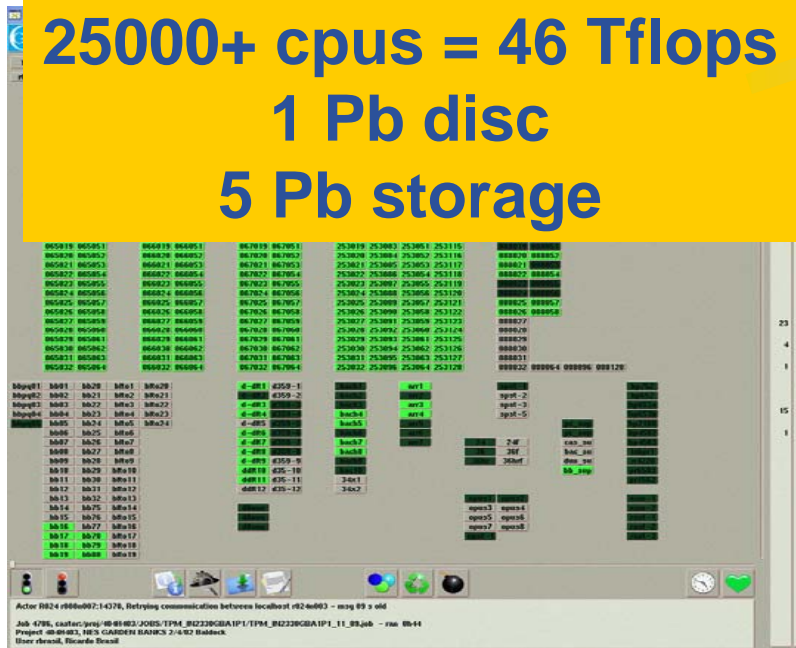




1 Pflops in 2009?

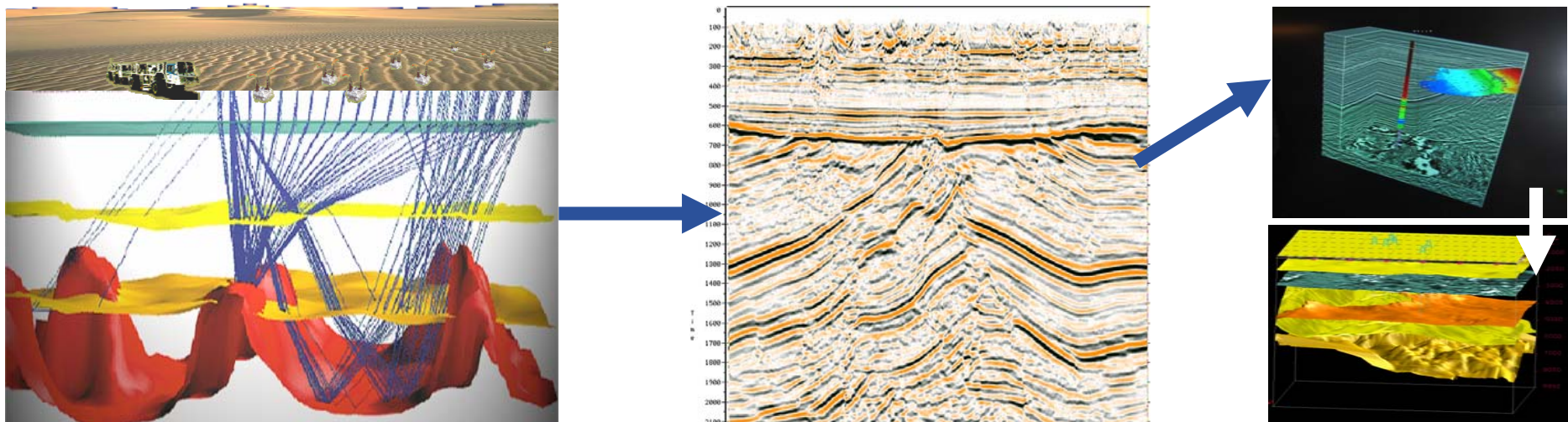
100 to 200 Tflops in 2006

Mid 2005
25000+ cpus = 46 Tflops
1 Pb disc
5 Pb storage



- **More Computing & Storage resources to solve complex problems**
 - Capability to solve complex problems and to validate innovative algorithms on real size data sets
 - Close the gap between Research and Industrial environment
 - Framework for Industry/Research collaborations
- **To optimise IT infrastructure**
 - Lower the total cost of IT by sharing available resources
 - Load balancing between CGG Processing Centres
 - Smoothing peaks of production
 - Service continuity – Business Continuity Plan
 - Better fault tolerant system and applications
- **To share and acquire knowledge**
 - Best practices and programming models
 - Enable cross-organizational teamwork and partnership

- **Seismic processing Generic Platform for research and education**
 - Based on Geocluster©, an industrial application, used in production at CGG
 - Include several standard tools for signal processing, simulation and inversion (model optimisation)
 - Being ported to EGEE for **Industry** and Academia
 - The main focus of EGEODE Virtual Organization



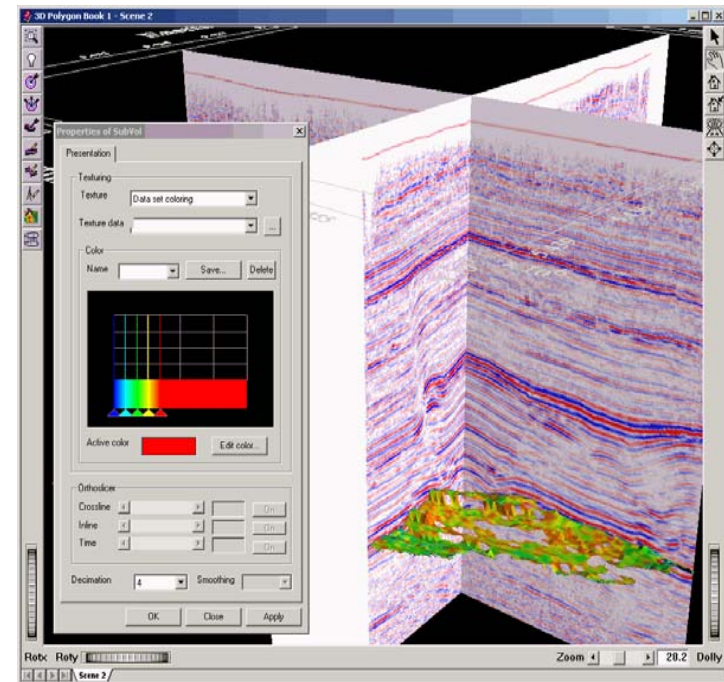
- « *Expanding Geosciences On Demand* » : **EGEODE VO**

Virtual Organization to share IT resources and best-practices.
 Opened to all Research centers in environmental geophysics from
 both **Industrial (public-private)** and **Academic** world

- Open: any user can write new algorithms in new modules
- Free access for academic research
- Controlled by license keys

- **Community challenges**

- Geophysics is a key technology for earth sciences, it shares and complements requirements of actual EGEE applications



- **Who are the targeted users**
 - Researchers in seismic processing algorithms
 - Researchers in geosciences *using* seismic processing
 - Few hundred, very scattered
- **Characteristics of geophysical data processing**
 - Huge volume of data (seismic) and complex structures (models, parameters)
 - Petabytes to store; hundred Gigabytes to transfer per job
 - Very long computation with IO-memory-cpu bounds
 - PSD/TM: one iteration = weeks on hundreds of nodes
 - Comprehensive and successful CGG experience on Very Large Clusters

- **IT side**
 - Create a node of the European grid EGEE
 - Learn how to install/use/administrate and improve the middleware
 - Evaluate the cost of Managing grid infrastructure
 - Create an Internal Technology grid
- **Software side**
 - Connect an application on top of the middleware
 - License management and compilation server
 - Understand and validate new programming models in real situation
 - Robustness, portability, performances, difficulty to develop, ...
 - Explore new ways for end user (web services,...)
- **V.O. side**
 - *EGEODE*: Expanding GEosciences On DEmand
 - Dedicated to Research in Geosciences

- **What's missing to go to full production:**
 - Disc and network bandwidth
 - An economical/usage model (and accounting tool)
 - To support a good balance between users and providers of resources
 - To include cost of network transfer
 - A mean to manage projects inside a VO: authorization for a user to access a project, accounting management at a project or user level.
 - Middleware to implement local policies about dynamic priorities and resources allocation to VOs or Projects