Earth Sciences: EGEODE

« Expanding Geosciences On Demand »

EGEE 1st EU Review – 9th to 11th February 2005
CERN
Dominique Thomas; Gaël Youinou
Compagnie Générale de Géophysique (CGG, France)
R&D

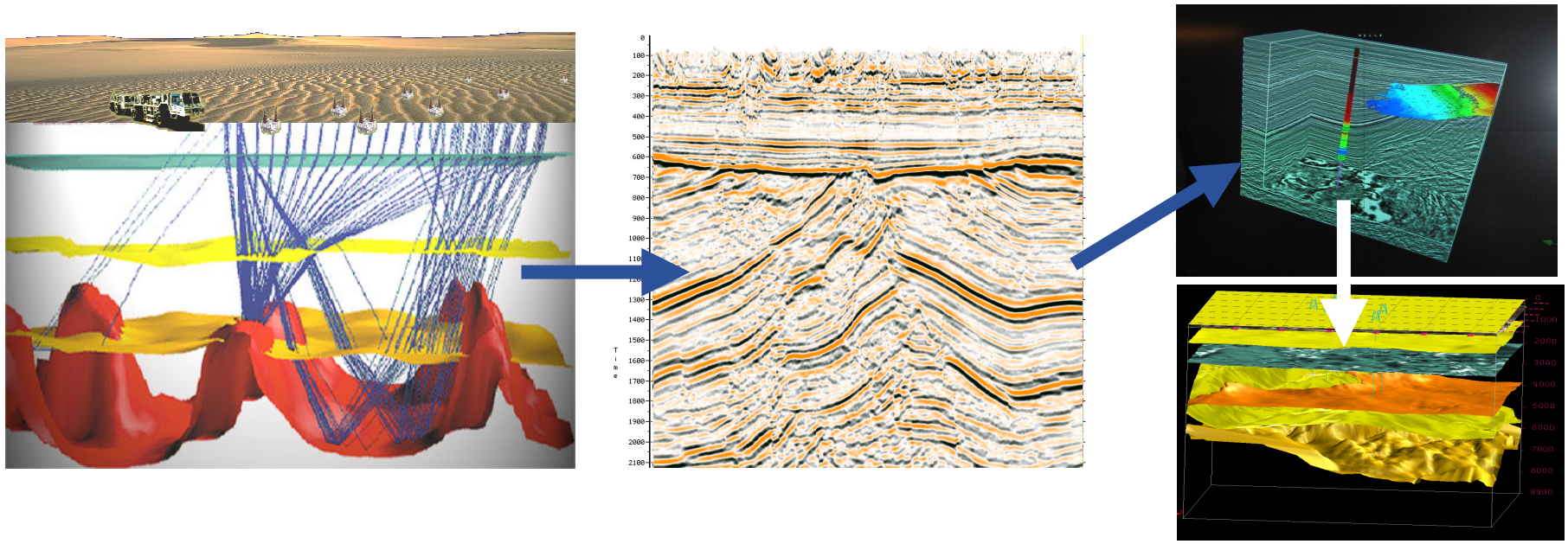
www.eu-egee.org

Information Society

Overview of the application

- **Geocluster, the *seismic processing* generic platform from CGG**
  - Being ported to EGEE for ***Industry*** and Academia
  - The main focus of EGEODE Virtual Organization

- **Close collaboration with ESR « Earth Sciences Research » VO.**
  (Earth Observation, Climate, Hydrology, Solid Earth Geophysics)



EGEODE application

Seismic processing Generic Platform for research and education:

- Based on **Geocluster**, an industrial application, used in production

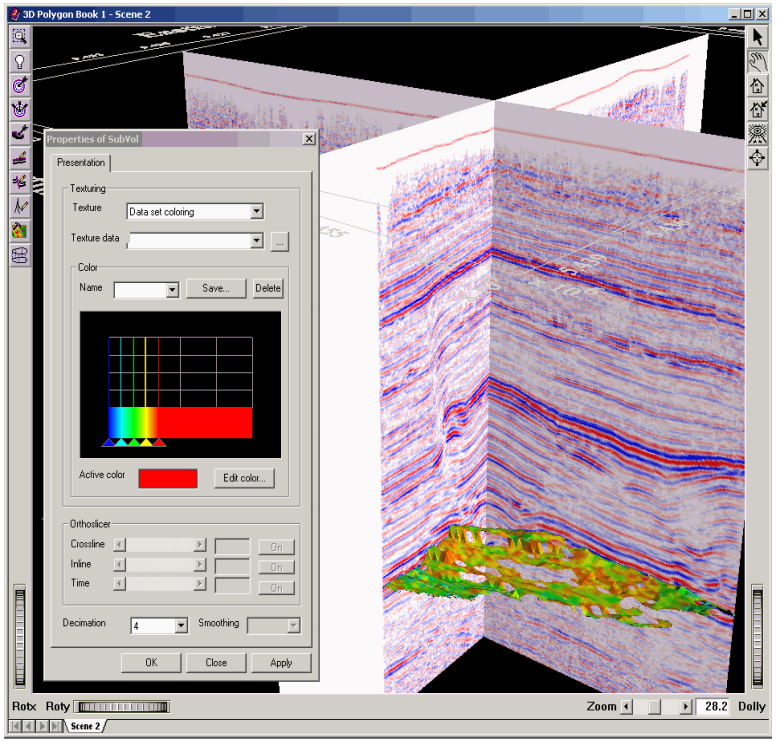
- Include several standard tools for signal processing, simulation and inversion (model optimization).

- Open: any user can write new algorithms in new modules (shared or not)

- Free access for academic research

- Controlled by license keys (opportunity to explore license issue at a grid level)

- Initial partners F, CH, NL, Russia



Global workflow

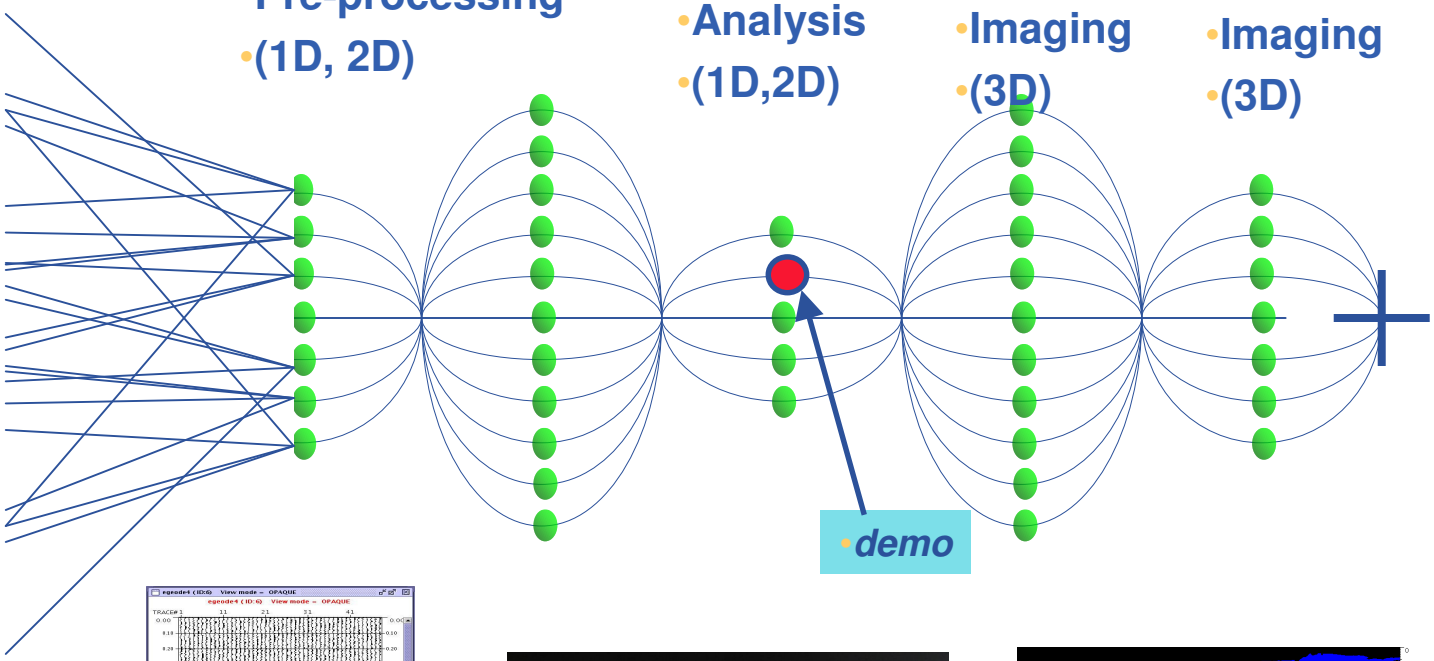
- Pre-processing
- (1D, 2D)

- Velocity
- Analysis
- (1D,2D)

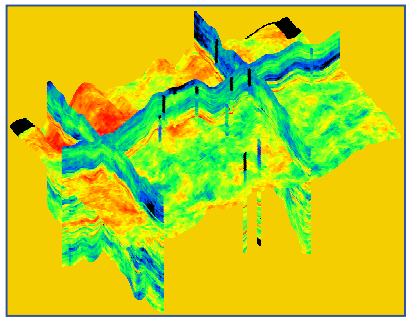
- Time
- Imaging
- (3D)

- Depth
- Imaging
- (3D)

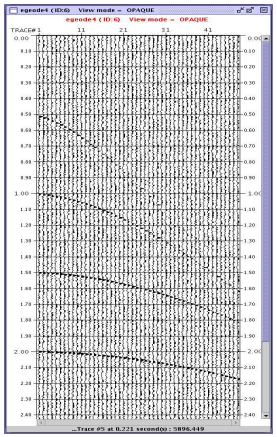
- Reservoir
- Geophysics

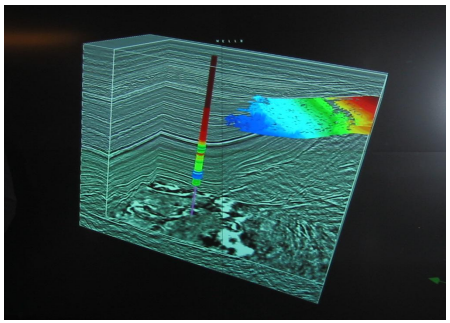


- *demo*

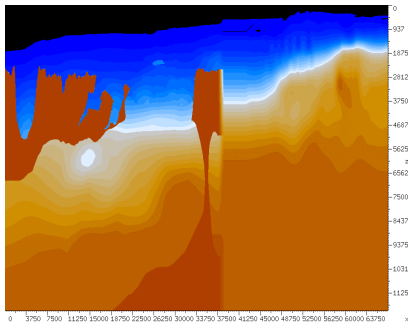


- Rock properties

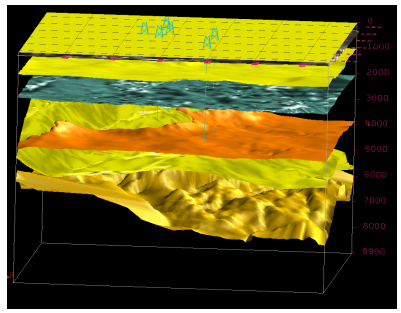




- 3D seismic



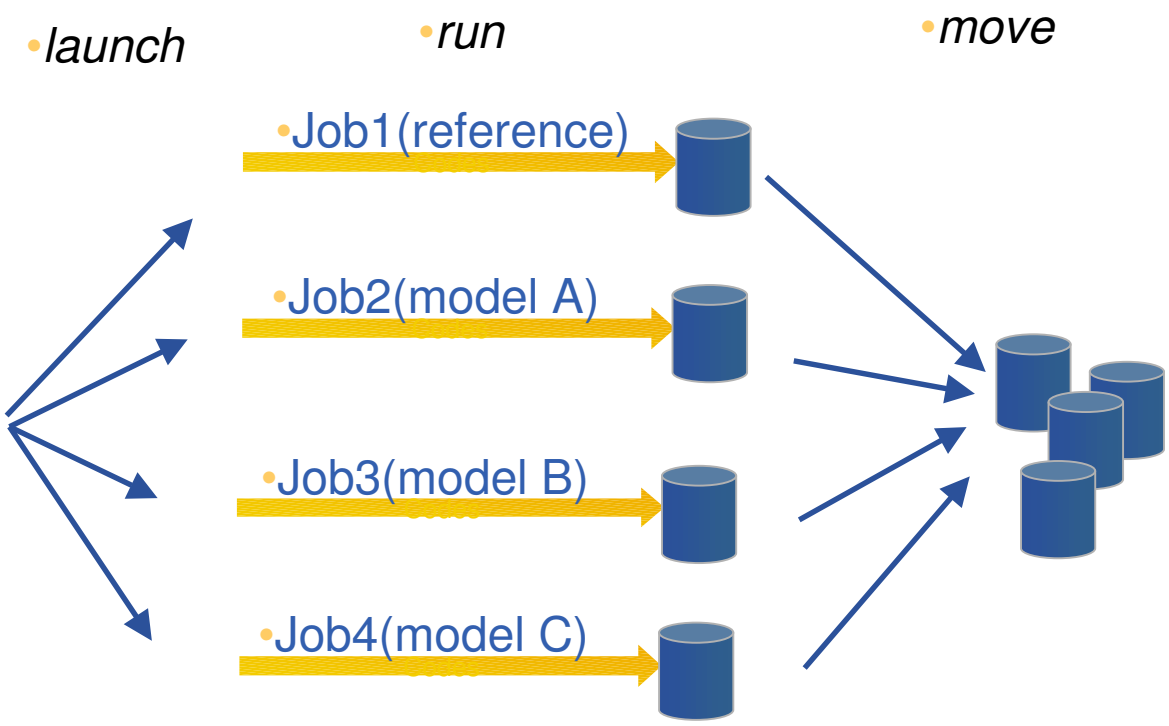
- Velocity model

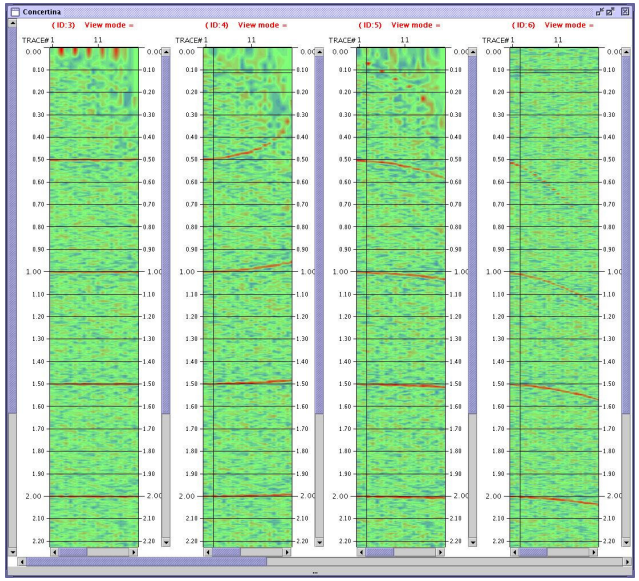


- Structural model

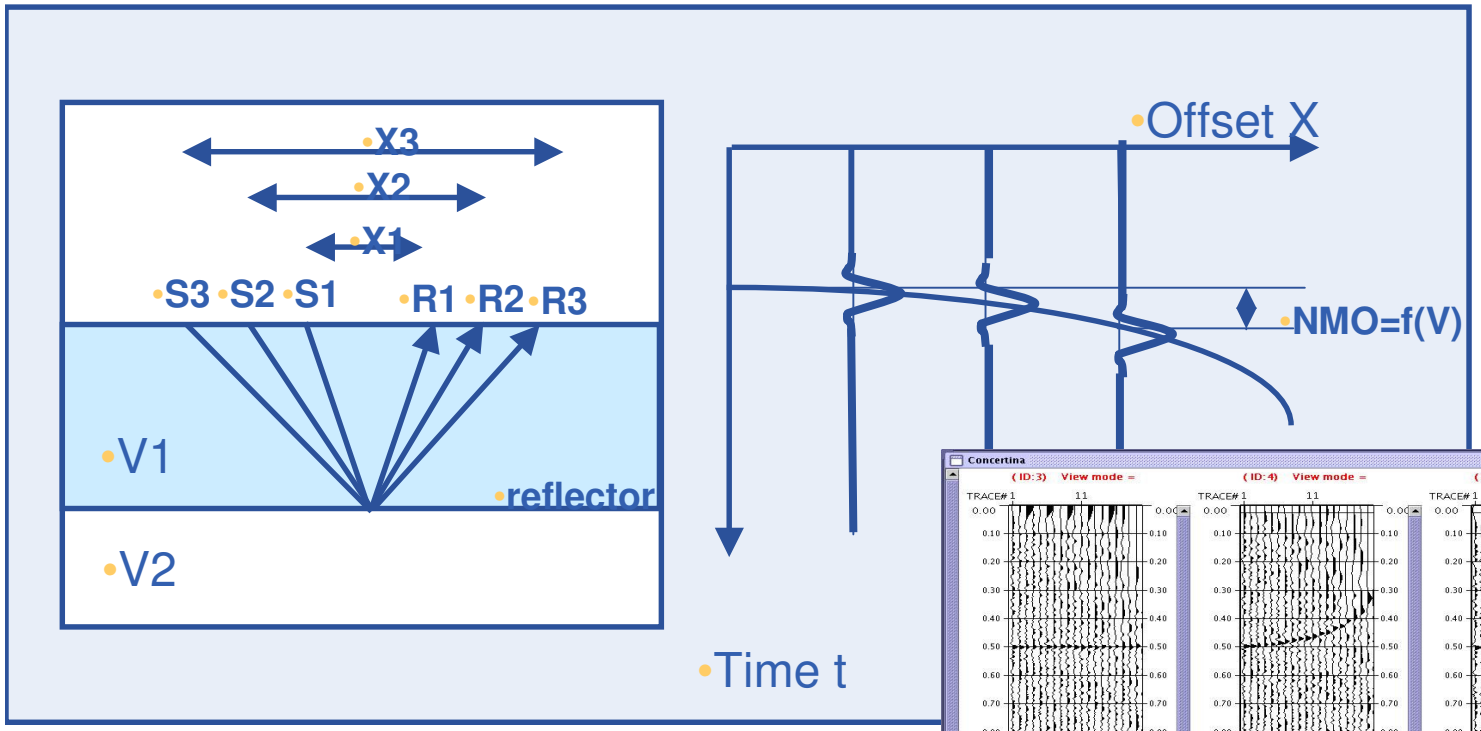
Demonstration outline

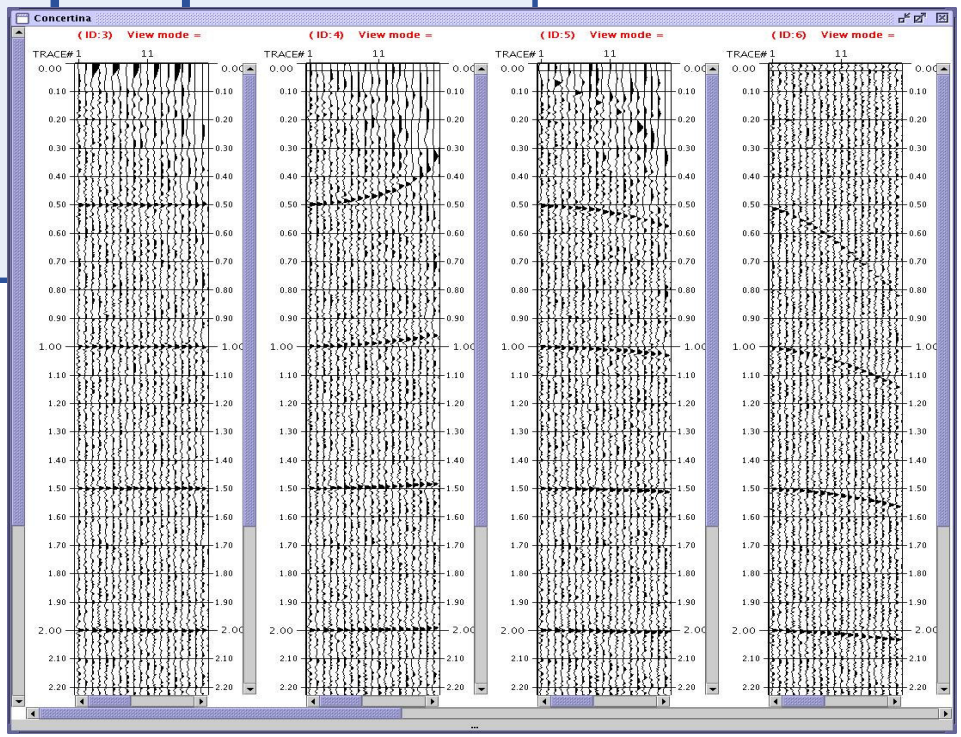
- **GeoCluster run on standard EGEE production grid and on GILDA for induction of new users**
  - Demonstration of a simple parametric study





Demonstration: NMO correction





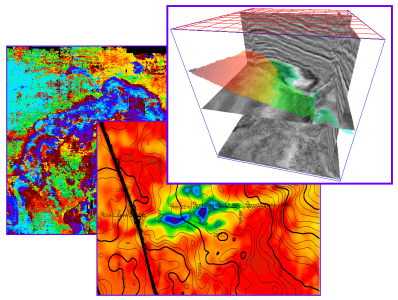
- Normal Move Out: a way to compute velocity field

CGG

Acquisition 1-10 TB/proj (IBM 34/3590)

Processing 5-50 TB/proj (SAN)

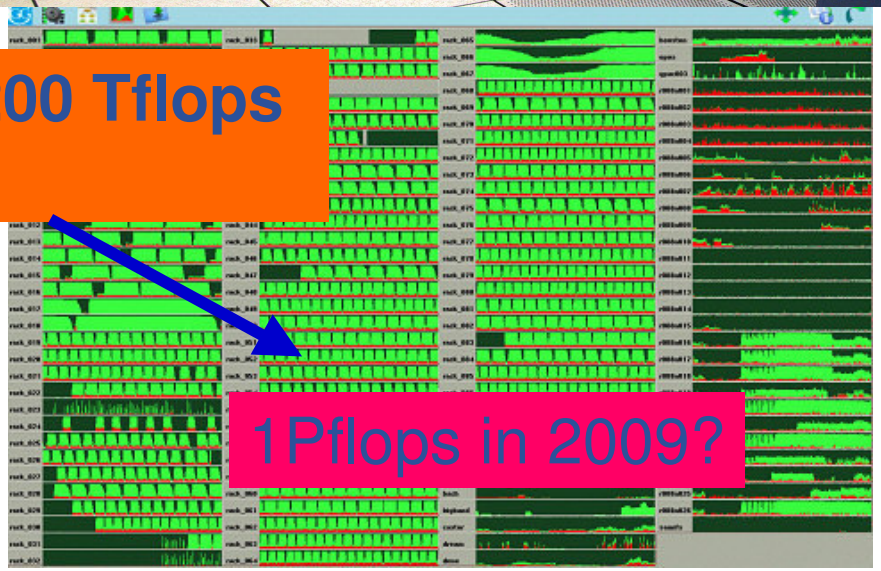
Interpretation << 1 TB/proj



CGG - Storage and Computing

End 2004
20000+ cpus = 45 Tflops
1,5 Pb disc
5 Pb storage

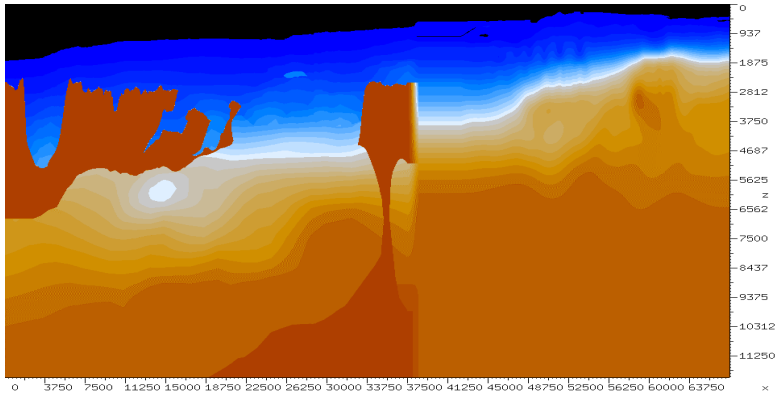
100 to 200 Tflops in 2006



1Pflops in 2009?

Application users

- **Who are the targeted users:**
  - Researchers in seismic processing algorithms
  - Researchers in geosciences *using* seismic processing
- **Few hundred, very scattered**



- **Imaging benchmark example:**
- **CGG**
  - R&D team has access to large production facilities
  - Use state of the art technology in the full sequence
- **University**
  - Process a small part of data, hiding potential quality of results
  - Use basic processing except for imaging
  - Has a very innovative algorithm for imaging

Advantages of gridifying application

- **The general benefits of grid computing:**
  - *Access to computing resources without investing in large IT infrastructure*
  - Lower the total cost of IT by sharing available resources

- **And the specific benefits for Research community:**
  - *Access to a comprehensive, industrial software*
  - Free the researcher from the additional burden of managing IT hardware and software complexity and limitations.
  - Have a framework to share data and project resources with other teams across Europe and the rest of the world,
  - Share best practices, support and training more easily.
  - Enable cross-organizational teamwork and partnership.

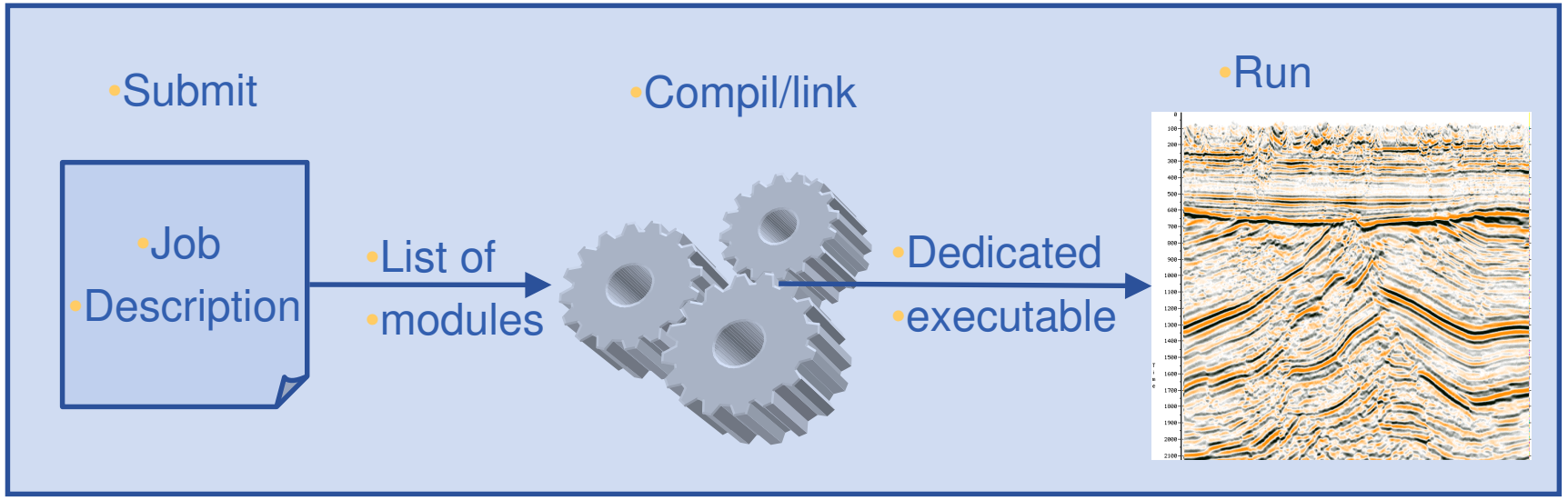
Deployment layout

- **Just started: the application is not yet fully available on the Grid**

- **3 Sites on production grid +GILDA**
  - CGG (CE/WNs; SE; UI), IPGP(CE/WN;UI), SCAI (CE/WN;SE)
  - GILDA(« get started »: easy access via portal for new grid user)
- **Key services**
  - Based on standard LCG-2
  - 2 RB (LAPP, LAL, CGG in Q1)
  - VOMS expected (managing sub-projects?)
  - MPI expected (for Imaging)
  - Licence server not yet implemented
  - Compilation/Link server implemented (not part of LCG-2)

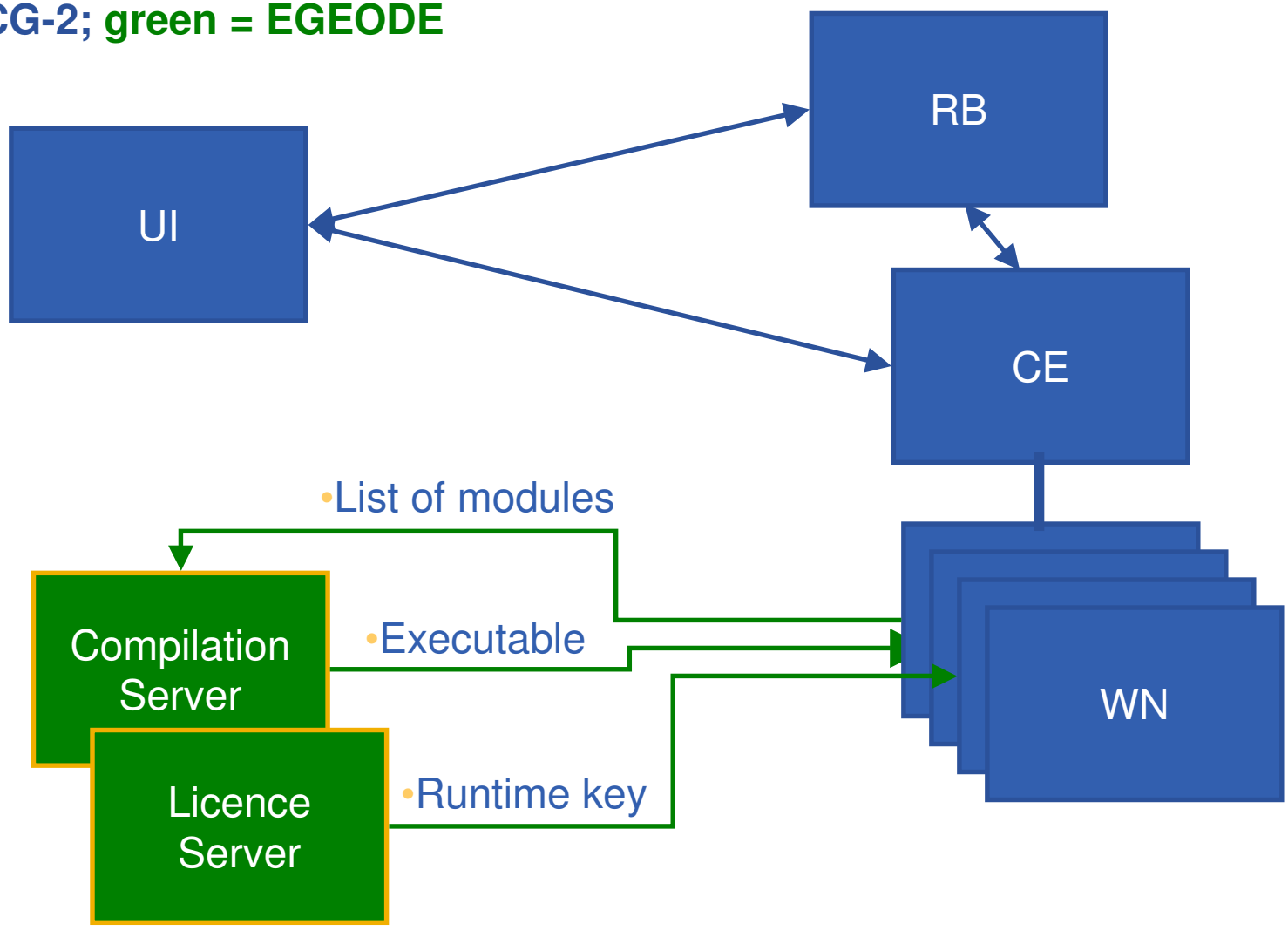
Limitations found and workaround

- **A running application**
  - Gridifying: depedencies on run-time environment (WN!)
    - (new version of Geocluster: full availability: mid-2005)
  - Commercial software
    - (licence management must be implemented)
  - Size of the binary executable (>2Gb by default)
    - (two phases: compil/link + run => compilation server)



Usage of LCG-2 middleware services

Blue=LCG-2; green = EGEODE



Cost comparison with using only local resources

- **No figure available**
  - Accounting system not yet used,
  - What are the relevant indicators (->TCO) ?

- **Significant cost saving expected from centralized software management**

Future work and direction

- **Lessons learned**
  - It works
  - The learning curve is significant

- **Next activities (in collaboration with ESR and EGEE)**
- **1 – Application**
  - Finalize the porting of the application to the grid
  - Implement a licence management system
  - Implement an accounting system
  - Work on the data management issues (data servers, SE,..)

- **2 – Users/deployment**
  - Train a support team aware of both application and grid topics
  - Create a web based, support and information system (as any other VOs)
  - Consolidate and Formalize membership

Challenges towards full production

- **what's missing to go to full production:**
  - The validated application
  - 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 and licence management at a project or user level.
  - Middleware to implement local policies about dynamic priorities and resources allocation to VOs or Projects
- **Expected deployment of EGEODE (with SA1 and NA4 support)**
  - 5 sites, 100 nodes, 10 users Q2-2005
  - Up to 25 sites, 500 nodes, 100 users Q1-2006