Historically, the Grid can be viewed as the latest step toward distributed computing solutions; the significance of the Grid comes from its ambition to extend distributed computing and storage to a global scale. In fact, every technological revolution conceals a great deal of evolution!

What is Grid

The Grid is a service for sharing computer power and data storage capacity in a geographically distributed way. Like the web, the Grid is a service which runs on Internet. Unlike the web, there is no uniform standard or protocol for the Grid today: there are many Grids for applications with different kinds of computing requirements.

Some Grids link resources at a local or national scale. Some Scientific Grids link together dozens of major computing centers on several countries, even continents. Projects such as the EU co-funded EGEE (Enabling Grids for E-sciencE) and DEISA (Distributed European Infrastructure for Supercomputing Applications) are establishing general-purpose Grid infrastructure for science and industrial research applications. DEISA links super-computers while EGEE is devoted to access and analyze large amount of data by thousands of scientists for the larger part from the High Energy Physics field.

How does the grid work?

The essential component is the middleware, which is a software that allows the user to access remote data and processing power in a simple, reliable and efficient way. The physical infrastructure consists of clusters of PCs, supercomputers, tapes and disk storage systems, as well as the network that links them together.

Data in EGEE

- **Data files**: stored on Storage Elements on EGEE. Listed on the EGEE catalogs to locate them when required, or attached to the job in the input sandbox.
- **(Meta)Databases**: accessed through Amga if inside the Grid infrastructure, accessed with OGSA-DAI authentication if outside.
- **Volume**: millions of files, only limited by availability of storage space and network rates.
- **Formats**: all accepted.
- **Data sharing and exchange**: a given set of data belonging to a VO is accessible only to the members of this VO (or a subset of it). The data can be encrypted, if needed.

Schema of EGEE - Layers

http://www.eu-egee.org around 32,000 CPUs and 13 PB storage distributed on 200 European sites.

Why EGEE for Earth Science?

- **Computer resources**: Linux machines (ia32/ia64/x64_86 architectures).
- **Distributed computing**: functionalities to manage simultaneous jobs.
- **MPI jobs**: usually less than 16 CPUs available on around 10 sites.
- **Languages**: C, C++, Python on all sites;
  - Fortran 77-90, IDL, Octave, Matlab ... not yet everywhere.
  - Implementation left to the site administrator discretion.
- **Workflows**.
- **Grid is very well adapted for Earth Science applications**:
  - for statistical approach, intensive computation and/or storage,
  - for rapid solution in case of many independent jobs,
  - for sharing and/or processing of large sets of data within the scope of large-scale European projects.

Conclusion: EGEE for Earth Science

Grid will open new fields of investigation in Earth Science.

- An open platform for handling computing resources and data, with tools not limited to high-performing computing. A partner can use a lot more resources than the ones he (she) brings in.
- Impressive number of shared resources. EGEEII is around 32,000 CPUs distributed and 13 PB storage in 200 sites.
- Secure and restricted access to resources, data, tools... Same data and software policy as outside Grid.
- A collaborative platform among teams and/or countries. Interactive collaboration to avoid effort duplication (software, tools, data...).
  - Innovative multidisciplinary applications.

⇒ more effort and time left for research
Some Earth Science applications on EGEE
August 2007

Two publications in Journal of Quantitative Spectroscopy and Radiative Transfer.

Gome/ERS2: Ozone-7 years
ESA, KNMI (NL), IPSL (FR), UTV (IT)
lugi.fusco@esa.int
sdecerf@knmi.nl
sophie.goldin@aero.jussieu.fr
Production of ozone profiles through the means of a Neural Network algorithm (2 versions) and a physical inversion algorithm.

Validation with ground-based lidar.

Grid characteristics:
- large number of files (~80000),
- metadata base with geospatial queries for collocation satellite-lidar,
- OGSA-DAI server for the database,
- complex algorithms,
- sharing raw and elaborated data.

Earthquake: CMT
IPGP (FR)
clevede@ipgp.jussieu.fr
patau@ipgp.jussieu.fr
Data: seismograms from Geoscope network.

Computation: Green function for each 3D-Grid point around the approximate earthquake location and for different times.

Results: Location of the space-time barycenter of the rupture, the seismic energy released, the source duration, and the global mechanism of the source.

Grid characteristics:
- many simultaneous independent jobs (50-100),
- greatly improved response time (<6 hours vs 1 week).


GeoCluster
CGG-Veritas (FR)
gerald.vetois@cggveritas.com
Seismic platform software (~400 modules) developed by CGG for Academy and Industry.

Available in EGEE for Academic organizations.
More resources available: possibility to use it on a larger scale.

Last version of the software, no need to implement it.

Grid characteristics:
- license server,
- large storage space,
- large network bandwidth (between sites and internal network),
- processing intensive: batch, MPI,
- fault tolerance system for central server (LFC, GGS, VOMS),
- interactive access.

Codesa - 3D
CRS4 (IT), Unime (CH), INAT (TU)
giuditta@crgs4.it
Intrusion of seawater into coastal aquifers.

Results: probabilistic maps of seawater intrusion in coastal aquifers of the Mediterranean basin, using Monte Carlo simulations, and according to different scenarios for sustainable water resources management.

Grid characteristics:
- large dataset volume,
- MPI,
- long running jobs,
- virtual data model,
- monitoring needs,
- collaborative environment,
- use of licensed software (not yet on the grid),
- EUMedGrid.

Geomorphology
IPGP (FR)
narteau@ipgp.jussieu.fr
rozier@ipgp.jussieu.fr
Formation and evolution of landscapes using a discrete model of transport.

Model: a 3D cellular automaton (CA) in which different sets of next-neighbor interactions allow to distinguish between different types of physical processes (e.g. erosion, deposition, transport). Such an innovation is necessary to implement retroaction mechanisms between a topography and a flow.

Grid characteristics:
- parametrical studies,
- CPU and memory intensive.

3DSEM_UNSTRUCT
IPGP (FR)
delavaud@ipgp.jussieu.fr
moquiliy@ipgp.jussieu.fr
This numerical tool uses the spectral element method to model 3D seismic waves propagation in complex geological media, on a local scale (sedimentary basin, topography). Its ability to handle unstructured meshes allows to take into account complicated geometries (surfaces topography, interfaces) and to adapt wavelengths resolution in heterogeneous media conditions.

Used to study the Caracas basin.

Grid characteristics:
- intensive processing with MPI/90,
- to be ported on EELA.

ELMER
http://www.csc.fi/elman

Flood Forecasting of a Danube river
UISAV (SK)
hluchy.ul@savba.sk
Data: Meteorological boundary conditions, river network, numerical field map.

Models: Meteorology: ALADIN (MPI-parallel), MM5 (MPI-parallel), Hydrology: HSPF (sequential-parametric), NLC (sequential-parametric).

Hydraulic: DaveF (MPI-parallel), EFSWMS (MPI-parallel).

Output: weather forecasting, precipitation forecasting, hydrograph, water level and velocity of flooding area.

Grid characteristics:
- complex workflow,
- MPI,
- knowledge management.

Application to be adapted to French and Ukrainian rivers.

Other Applications
- Mars Atmosphere simulation (F. Cipriani; CETP, FR): very intensive Monte-Carlo procedures on density vs altitude profiles.
- The Climate applications (DKRZ, MPI, DE): Climate model output analysis on EGEE. For this purpose a metadata and data management structure has been developed and deployed to make existing data searchable, accessible and processable on EGEE (bierring@dkrz.de).
- Stratospheric polar ozone (IPSL, FR): 10-years worth of climatologic data processed through meteorological and chemical models (sophie.goldin@aero.jussieu.fr).
- Large-scale air pollution model (BAS, BG; NERL, DK): Implementation on EGEE of the sequential version of the Danish Eulerian Model (DEM) with medium grid spatial discretization. (Tvzetan Ostromky: ceco@parallel.bas.bg).
- Interface with the Geoscope server (IPGP, FR): re-analysis of 25 years worth of seismological data (stutz@ipgp.jussieu.fr).
- Seismology (Geophysical Lab., Univ. Thessaloniki, GR): numerical simulations in the broader area of Thessaloniki, Greece. (Andreas Skarlatoudis: askarlat@geo.auth.gr).
- Meteorology (GCRAS, RU): data mining on large sets of meteorological outputs (Mikhail Zhizhin: jjn@wdec.ru).

Contacts:
Monique Pettididier
monique.pettididier@cetp.ipsl.fr
David Weissnacht
weissens@crgs4.it