GRID science cases and projects in Grenoble:

a pragmatic multidisciplinary view.

P. Valiron

(Most slides taken from Grid workshop, Strasbourg, 7 juin 2006)

- Olivier Richard,
- Nicolas Capit,
- Françoise Roch and all other engineers on CIMENT
- Philippe Augerat (ICATIS)
- Plus the authors of the science cases… ASTROMOL, LPG

CIMENT

http://ciment.ujf-grenoble.fr

« Calcul Intensif, Modélisation, Expérimentation Numérique et Technologique »

« Intensive Computing, Modelling, Numerical and Technological Experimentation »

- An interdisciplinary HPC network involving researchers, engineers and students
- Platforms for computer science, medecine, physics, chemistry, climate, universe…

CIMENT platforms: diversity and power of hardware

- SCCI (Universe sciences, Observatory)
  - IBM Power3
  - **30 Sun v40z (november 2005, plus 12 grid'5000 add nodes, plus infiniband in 2006)**

~ 0.8 Tflop

- MIRAGE (Modelling, Environment, Climate)
  - 2 x Altix 350 SGI (16 itanium2 per SMP node)
- CECIC (Chemistry)
  - IBM Power3 + 14 bi-Xeon (Gigabit cluster)
- BioIMAGe (Biology, Medecine)
- PhyNum (Numerical Physics)
  - 40 bi-Athlons (Gigabit cluster)
  - Medetphy: 2 x Altix 350 SGI (16 itanium2 per SMP node)
- IMAG+INRIA clusters
  - Icluster 1 (225 Pentium III, now stopped)
  - **Icluster 2 (104 bi-Itanium2, myrinet, on GRID'5000)**
  - IDPOT (48 bi-Xeon cluster)
  - Student computer rooms at UFR IMA operated by ICATIS software (« dynamic » cluster, 60 Pentium 4 available on the grid when unused)

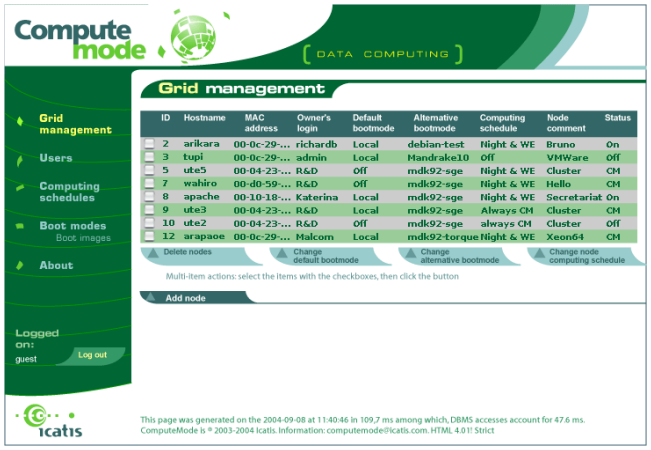
~ 0.7 Tflop

- Gigabit backbone, 10 Giga in deployment…

ICATIS / Compute Mode: enrolls non-intrusively dormant machines

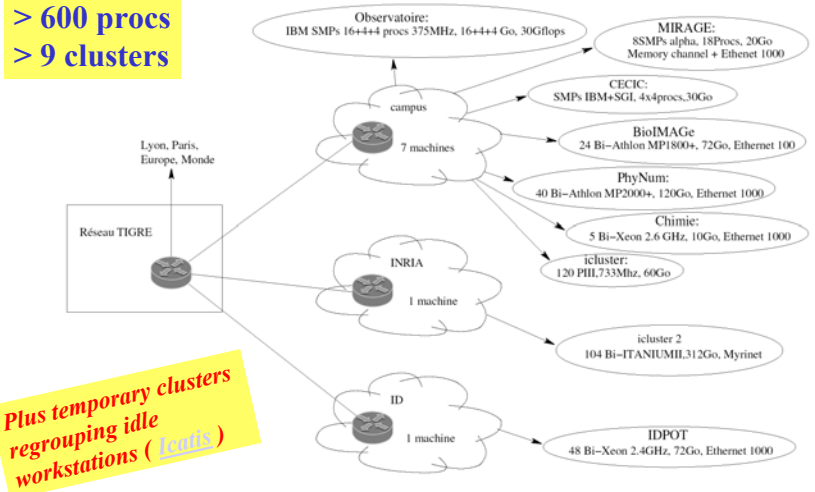
ComputeMode™ is an Icatis product that extends an organization's Computational Grid through the aggregation of unused computing resources.
For instance, **ComputeMode™** lets a virtual cluster be built using employees' PCs when they are available (at night, on weekends, during vacations, ...)

Icatis is an offspring startup from the laboratory ID (Informatics and Distribution) in Grenoble



« CIMENT » A zoo of clusters… A community of engineers and users.

> 600 procs
> 9 clusters



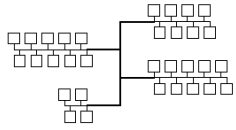
Plus temporary clusters regrouping idle workstations (Icotis)

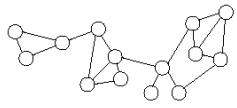
Federating the diversity ?

Evolution :

- lightweight grids : a hierarchical interconnection of clusters
- P2P networks : large number of nodes loosely organized
- general grids : a mix of both ?

...at least one order larger





Gnutella, Napster, BitTorrent…
Or SETI@home !

Need for simple and flexible strategies…

Our approach…

- Innocuity to entitled users ➔ « best effort » strategy
  - Kill grid jobs and resubmit when needed
- Security and accounting issues ➔ ssh, not Globus!
- Generality or simplicity ? ➔ lightweight, not Globus!
- Grid bottlenecks
  - Latency bottlenecks
    - Few microseconds for SMP and dedicated networking
    - 10 to 100 times more for local or departemental ethernet
    - Einstein's relativity constraints beyond. Remember your last satellite phone call !
  - Throughput bisection bottlenecks
    - Scalable for expensive dedicated local switches
    - Limited by the slowest WAN or national interconnect

  ➔ Large scale parallelism hopeless in general

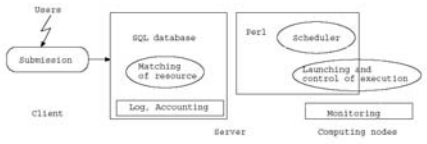
  ➔ Focus on multi-parametric applications
- Scalability for managing jobs and data ➔ be simple !
- Support for local parallelization ➔ Yes, « to do » list….

CiGri lightweight components

OAR : a different batch scheduler

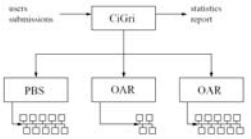
Built on top of a database engine (MySql) storing internal state

- easy debugging
- powerfull data extraction
- experiments made simple
- some support for data integrity



CiCgri : new grid specific features

- hierachization
  - meta scheduler using sub-schedulers for resources allocation
- mass submission of independant tasks
  - allow parametrized submission of a large set of similar tasks
- automatization of errors handling
  - by using rules describing actions to take in each case



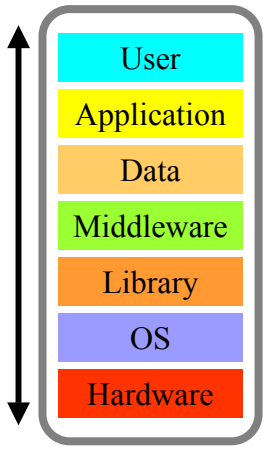
ACI GRID

Action Concertée Incitative [ACI] Globalisation des Ressources Informatiques et des Données [GRID]

- Modular architecture
- Pluggable schedulers
- Lightweight to develop
- Lightweight to use

Importance of error handling

- From initial experiments (P. Valiron, F. Roch)
  - Automatic error handling is mandatory
  - Importance of harware failures (disk errors)
- Error concept is complex
  - Harware failures
  - Application/hardware mismatch
  - Application bugs, wrong parameters…
- Actions
  - Spot the origin
  - Isolate
  - Restart
  - Other (prevention, periodic checking…)
- Needs
  - Specification of actions
  - Heuristics (remove a node if repeated errors)
  - In development…



User action

Presequisites

Get an account on several clusters

Install application on each

Validate application on each

Write JDL file (Job Data Language)

DEFAULT{

name = campaign1 ;

paramFile = param.tmp;

}

idpot.imag.fr{

execFile = /users/home/capitn/test.sh ;

}

tomte.ujf-grenoble.fr{

execFile = /users/nis/capitn/test.sh ;

}g.fr CIGRI - Une grille l´eg`ere

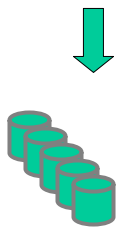
Campaign1: one line per job

Model1 100 200

Model2 300 400

Model3 3.14

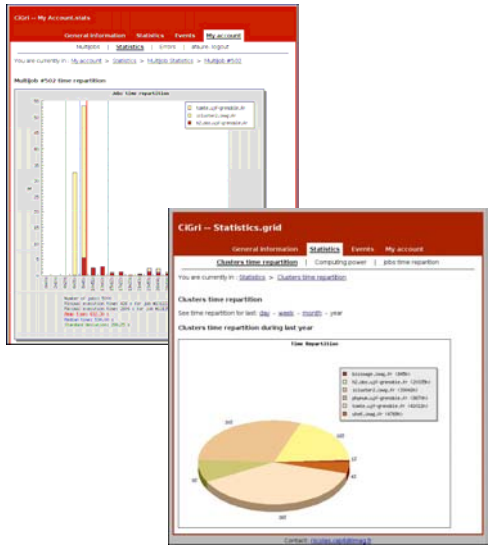
10000 more lines…g.fr CIGRI - Une grille l´eg`ere



Results automatically moved back on CiGri server

Comprehensive web interface and statistics

- Superjobs
- Individual jobs
- Error codes, timings
- Usage statistics
- And more…



CiGri: harnessing unused CPU cycles for many independant calculations

- Independant multi-parametric or Monte Carlo calculations…
- Best suited for few minutes to few hours jobs
  - Larger overhead for too short jobs.
  - Due to best effort strategy, longer jobs get a higher probability to get killed before termination.
- Selected science cases:
  - Data processing for independant samples (MARSIS)
  - Multi-D PES sampling
  - Monte Carlo trajectory calculations
  - Model fitting (e.g. radiative transfer)…
- Perspective: fulfill heavy VO requests (involving model fitting for a bunch of objects) ➔ define standard procedures…

In production on a lightweight grid (CIMENT)

- groupment of 9 clusters
- Over 600 nodes and 1.5 Tflop

Pluridisciplinary actual use

- medical imaging
- physics, astrophysics
- computational chemistry
- applied mathematics
- computer science

Easy to deploy and operate
Fun for both engineers and end users
Harnessing more science on existing equipments…

Partnerships for CIMENT and CiGri

UNIVERSITE JOSEPH FOURIER

Institut National Polytechnique de Grenoble INPG

ACI GRID

Liberté • Égalité • Fraternité RÉPUBLIQUE FRANÇAISE
MINISTERE DE LA RECHERCHE

ministère jeunesse éducation recherche — ministère délégué recherche et nouvelles technologies

INRIA

IMAG

Grid'5000

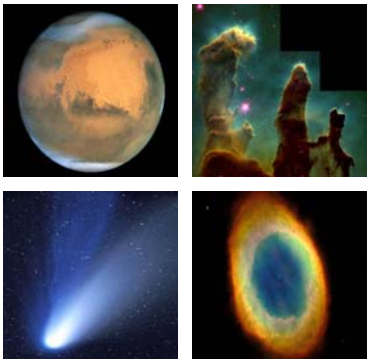
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

GRENOBLE-ALPES METROPOLE

VILLE DE GRENOBLE

RHÔNE-ALPES LA RÉGION

Water in the Universe



- H_2O is ubiquitous in Universe in either ice or vapour form.
- H_2O play a crucial role in:
  - Interstellar chemistry,
  - Stellar formation.

Herschel/HIFI 2007
• Unprecented window on cold U.
• Need detailed predictions on microscopic processes (inelastic collisions…)

A computational challenge

9-D Monte Carlo importance sampling ➔

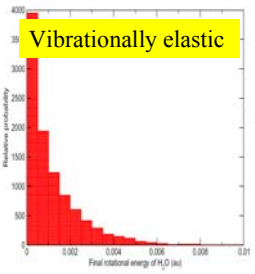
- ~ 375 000 geometries, 1 125 000 CCSD(T) runs
- ~ 200 000 CPU hours on our experimental PC grid
- Produced design specifications for CiGri

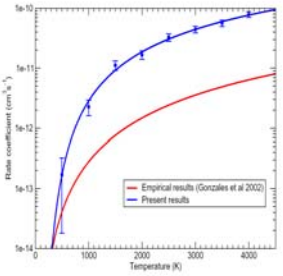
Action Concertée Incitative
[ACI]
Globalisation des Ressources
Informatiques et des Données
[GRID]

Monte-Carlo H_2O-H_2 trajectories: a 12-D problem

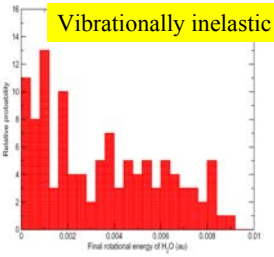
- *Objective* : **compute vibrational quenching for water bending.**
- *Méthod* : statistical analysis of trajectories campaigns with importance sampling of initial conditions.
- *CPU timing* : 5 min per trajectory in average.
- *Grid benefit* : permit to sample unfrequent collisional events, using 10,000 trajectories per temperature

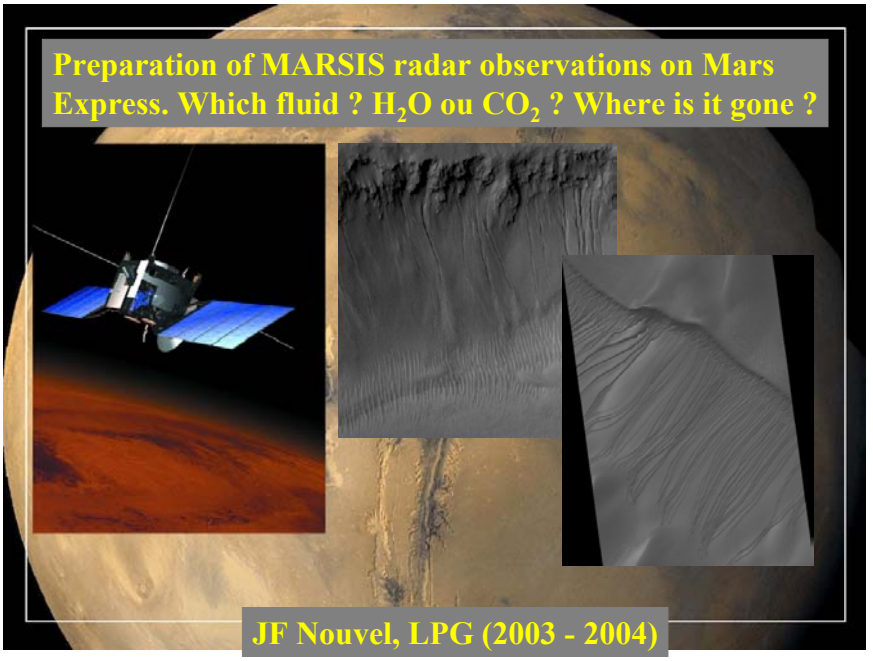
Vibrationally elastic



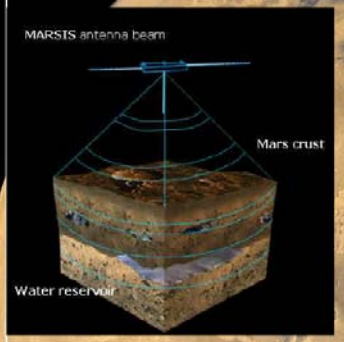


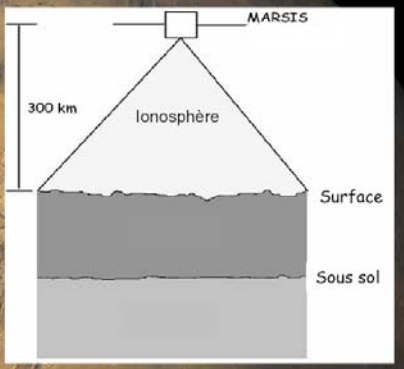
Vibrationally inelastic





L'intérêt de la simulation



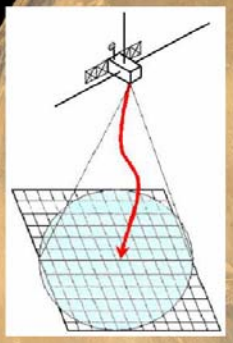


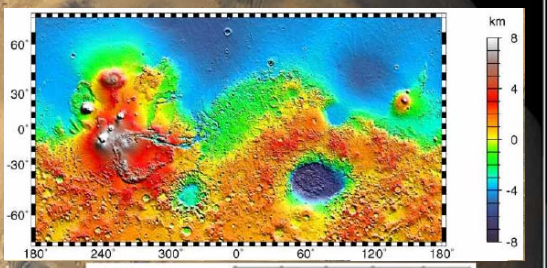
Planification: ⇒ Optimisation de la fréquence utilisée
⇒ Sélection des orbites pertinentes

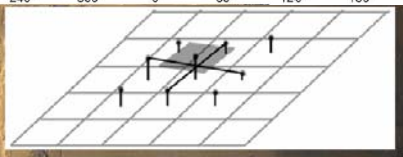
Résultats: ⇒ Préparation réduction donnée
⇒ Essai des algorithmes de traitement

La Méthode des Facettes

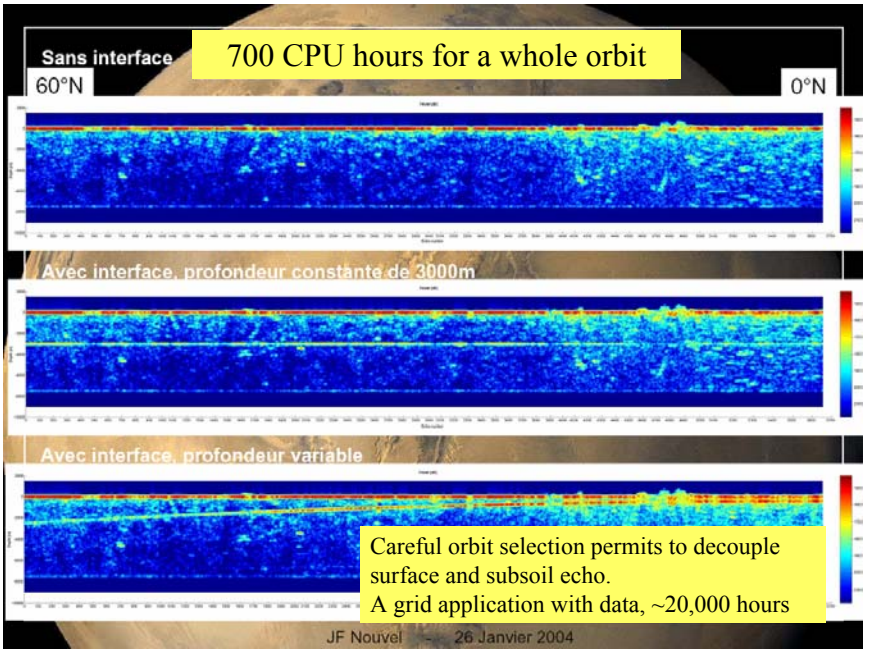
Procédé : Modéliser une surface par une série de plans tangents







- Facettes carrées de **500m de côté** => Expression analytique du lobe
- Elévation des points à la surface donnée par MOLA



Perspectives for CiGri…

- OAR and CiGri development
  - Support for parallel jobs within a cluster
  - More versatile schedulers
  - Better killing strategies (save longer or parallel jobs…)
- Interoperability with national GRID'5000
  - GRID'5000 is presently devoted to computer science experiments
  - Deploy science cases on the national grid
  - Teach general users to use grid resources
- Interoperability with other (european) grids
  - Collaboration in project with Thibaut Lery in Ireland
  - Support for Globus ➔ enter EGEE

Next step: specify interface between CiGri and VO ?

- User request
  - Request data for a class of objects (planetary disks, AGBs, …)
  - Specify model fitting tool (LVG, …)
- Get best fit parameters for all objects and publish !

Grid resources

Demande ANR « GLEMA »
Grilles légères multi-applicatives : mouvement de données et passage à l'échelle des calculs

- Passage à l'échelle du concept de « grille légère » pour calcul + stockage
- IMAG + IN2P3 + ICATIS
- Utilisateurs: OSUG + CECIC

Application OV: verrous de performance et interfaçage grille

- généraliser les requêtes à l'Observatoire Virtuel pour inclure la spécification de traitement sur les données, permettant ainsi d'appliquer un modèle à une famille d'objets et non plus au cas par cas.

- coupler la grille de données OV à une grille de calcul pour transférer les données et les traitements associés sur une grille de calcul dans un paradigme multi-paramétrique, en s'appuyant sur les progrès attendus dans le domaine du traitement des mouvements de données.

- renvoyer la synthèse des résultats à l'utilisateur et à une archive dans l'OV.

Cet objectif nécessite des développements au niveau de l'Observatoire virtuel mais également requiert des évolutions au niveau de l'intergiciel CiGri pour le traitement efficace du mouvement d'un très grand nombre de données.