

# **La Grille :**

# **Une mutualisation du calcul**

# **et du stockage**

Th.Mouthuy -CPPM 20/1/2009

# Plan

- ◆ C'est quoi une grille ?
- ◆ Qui l'utilise et pourquoi ?
  - ★ Physique des particules
  - ★ Bio-informatique
  - ★ Sciences de la terre
- ◆ Institut des grilles du CNRS
- ◆ Fonctionnement et mise en place d'un noeud de grille

# C'est quoi une grille ?

Partage **transparent** de l'utilisation de **ressources massivement distribuées** par des utilisateurs de différentes disciplines

*« A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high computation capabilities »*

(The Grid, I.Foster and C.Kesselman 1998)

# Caractéristiques importantes

- ◆ **Fédérer des ressources distribuées**
  - ★ CPU, stockage, algorithmes, expertises
- ◆ **Accès sécurisé entre des domaines différents**
  - ★ Confiance entre les administrateurs de système, développeurs des applications, utilisateurs,...
- ◆ **Accès transparent à « toutes » les ressources**
  - ★ On se connecte à la grille une seule fois
  - ★ Les aspects de calcul distribué sont cachés.

# Types de grille

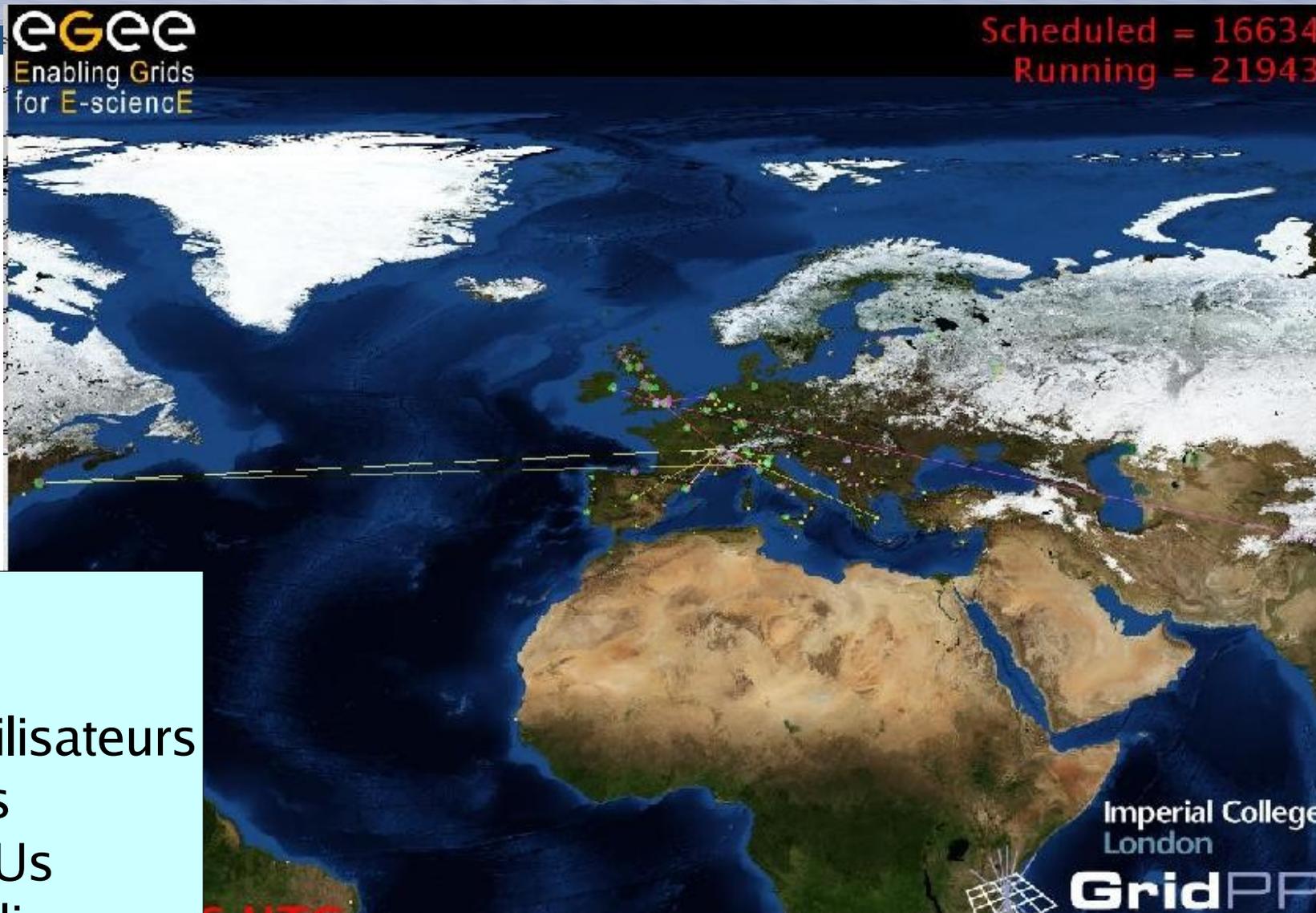
- ◆ Grille des supercalculateurs
  - ★ Deisa
  - ★ Cher, procédure d'allocation lourde, bien adapté aux applications fortement parallèles
- ◆ Grille institutionnelle
  - ★ EGEE (<http://www.eu-egee.org/>)
  - ★ Moins cher, allocations faites par les propriétaires, accueille des applications variées
- ◆ Grille des « desktops »
  - ★ BOINC (<http://boinc.berkeley.edu/>)
  - ★ Beaucoup moins cher, ressources des bénévoles, transfert de données minimum

# Le projet EGEE

- ◆ Une série de projets européens (2001- présent)
  - ★ European DataGrid (EDG)
  - ★ Enabling Grids for E-science (EGEE)
  - ★ EGEE-II (~36 M€)
  - ★ EGEE-III (~32 M€)
- ◆ Buts:
  - ★ Créer une grille européennes pour la communauté scientifique
  - ★ Aider les gens qui veulent intégrer leurs ressources dans la grille
  - ★ Aider les gens qui veulent utiliser la grille

# L'infrastructure EGEE

Scheduled = 16634  
Running = 21943



- 259 sites
- 52 pays
- > 7500 utilisateurs
- > 200 Vos
- ~ 72k CPUs
- ~ 20 PO disque
- > 200k tâches/jour

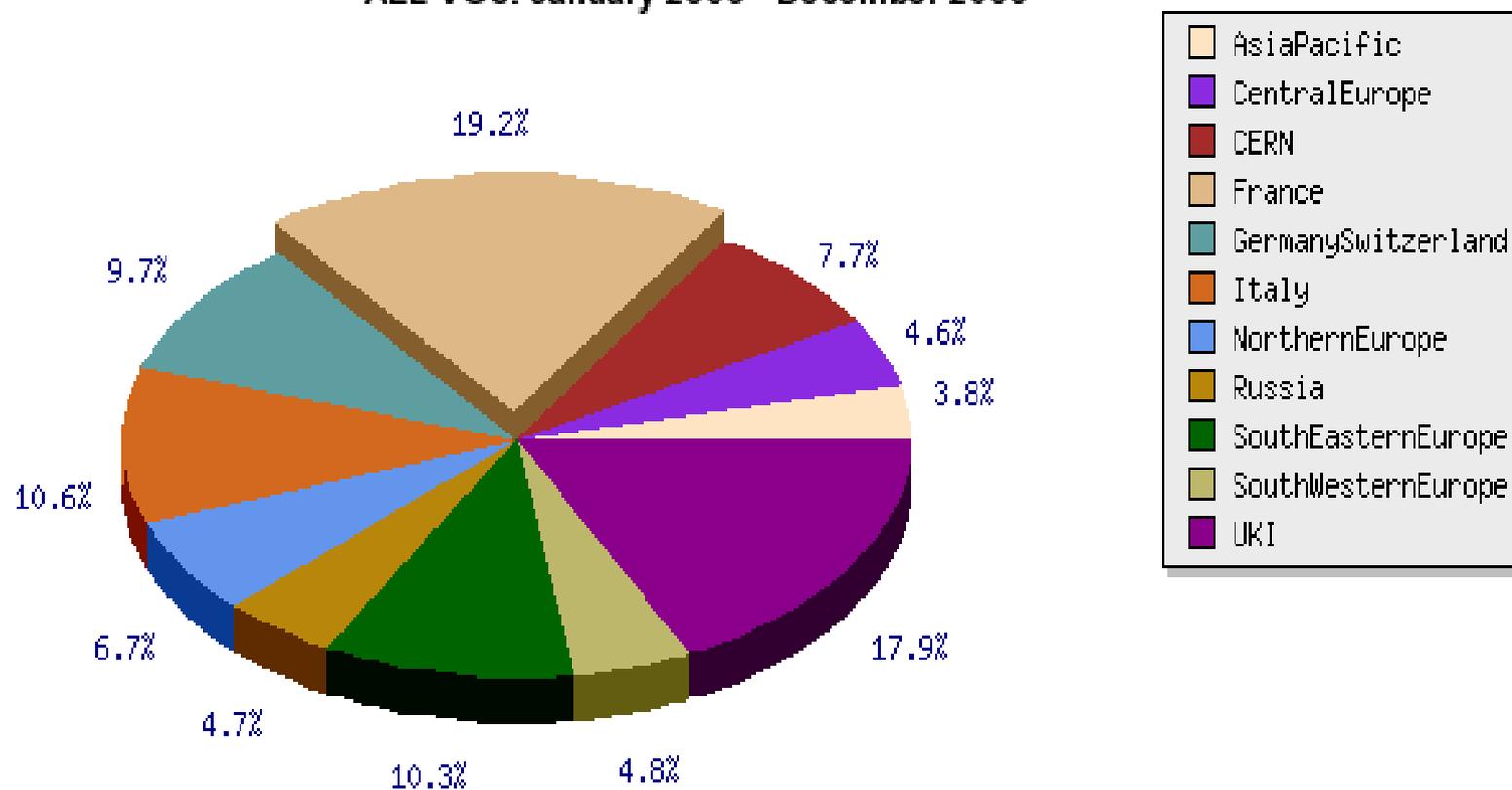
Real Time Monitor

<http://gridportal.hep.ph.ic.ac.uk/rtm>

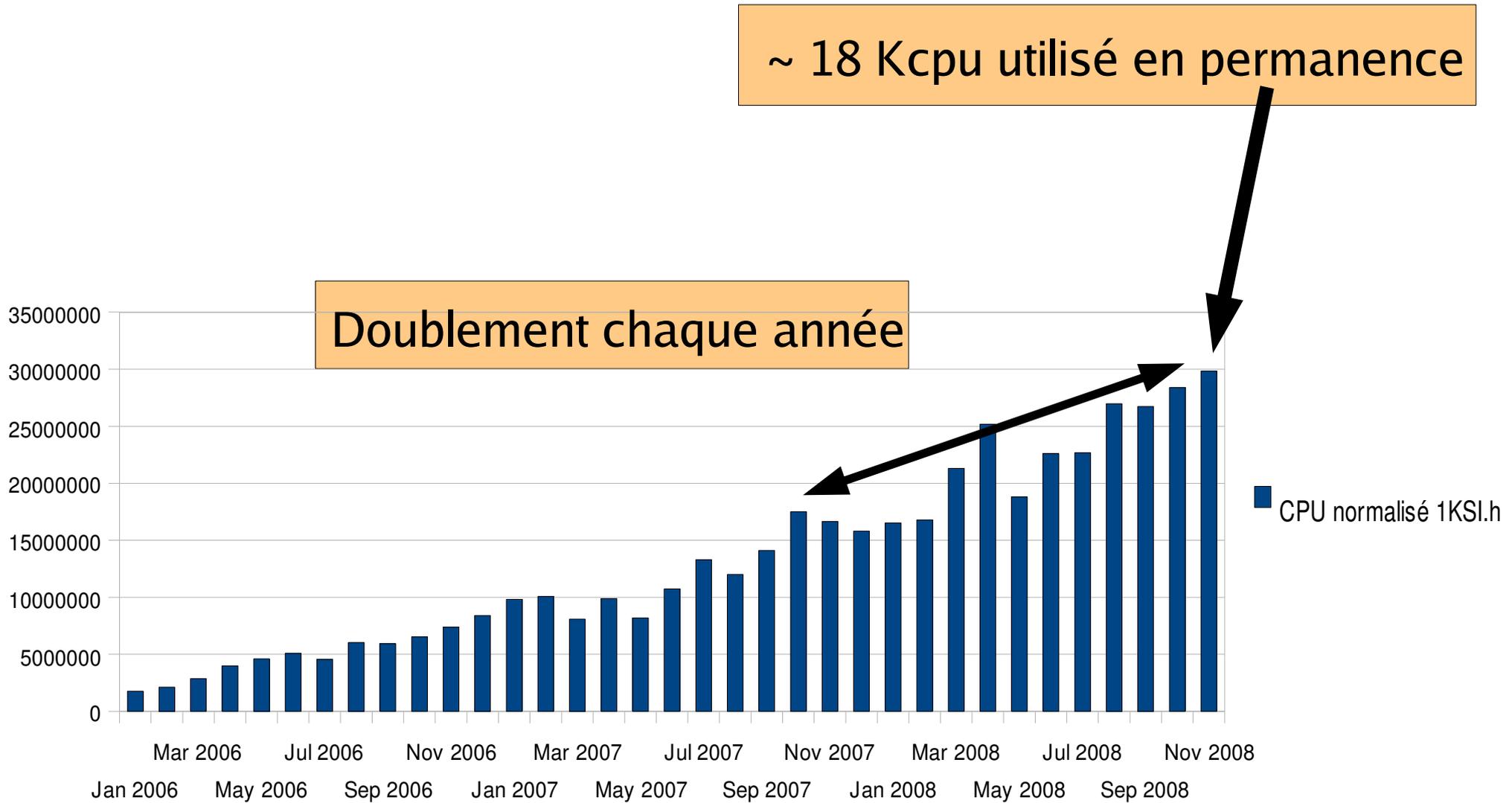
# Répartition des ressources

## PRODUCTION Normalised CPU time per REGION

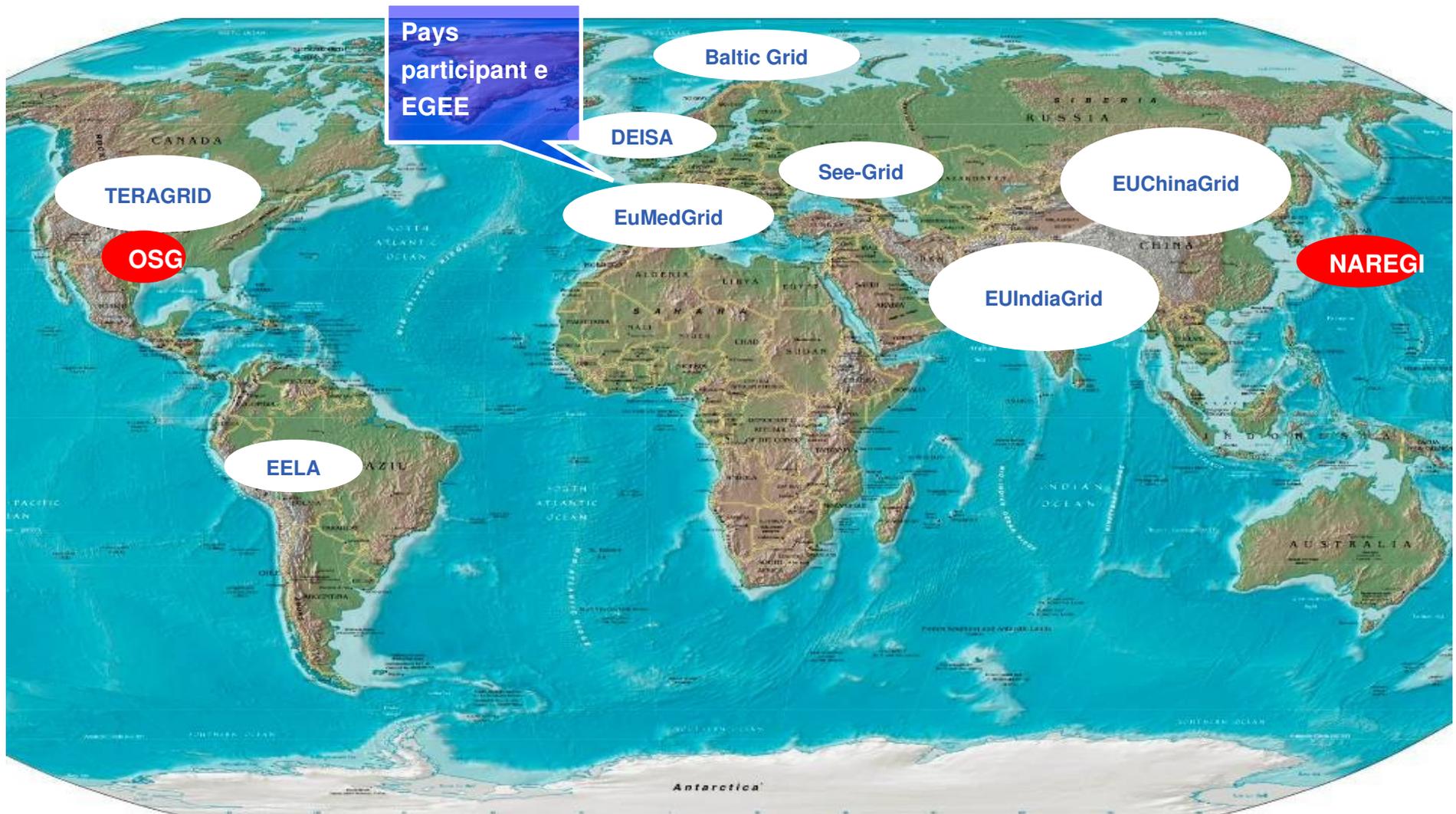
ALL VO's. January 2006 - December 2008



# Utilisation en forte croissance



# Interconnexion des grilles



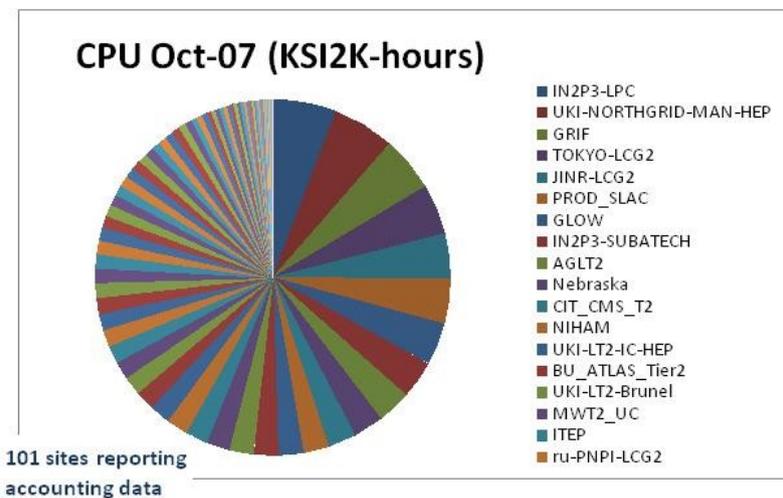
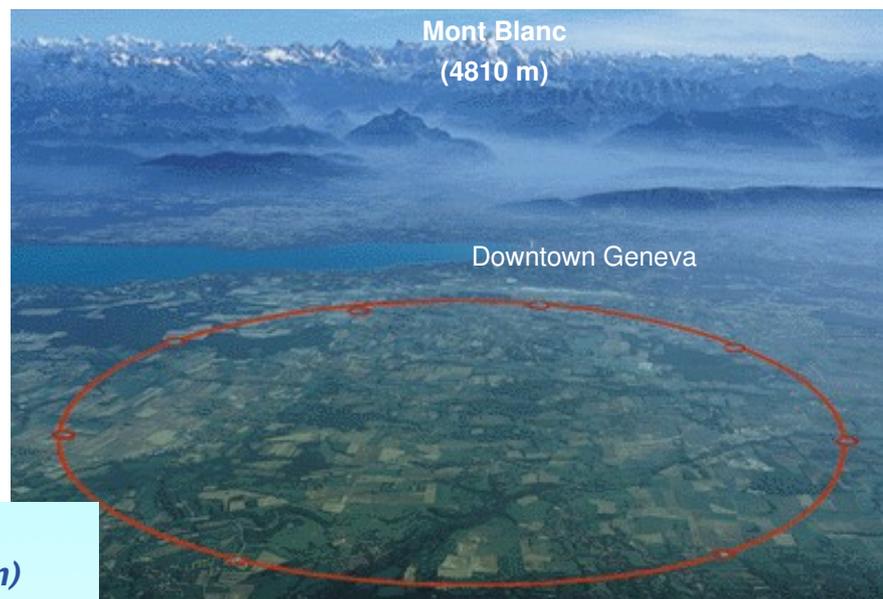
# Quelques exemples d'applications

- ◆ WLCG - La grille en physique des particules
- ◆ BioMed - La grille en sciences de la vie
- ◆ ESR - Utilisation en Sciences de la Terre

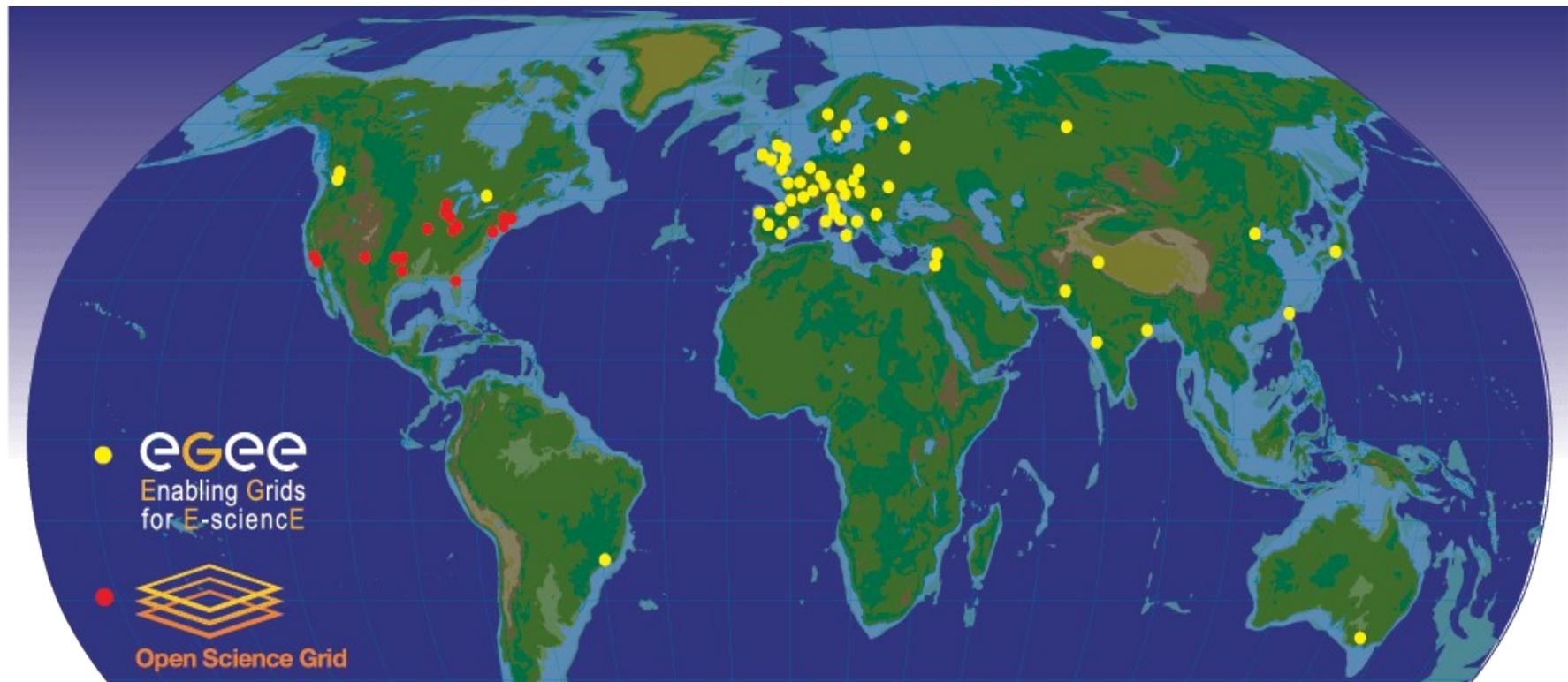


# WLCG - LHC Computing Grid

- ◆ **LHC** : l'accélérateur le plus puissant du monde au CERN
- ◆ Plusieurs expériences:
  - ★ Atlas
  - ★ CMS
  - ★ LHCb
  - ★ Alice
- ◆ Enormément de données
  - ★ 15 PO par an !
- ◆ Beaucoup de calcul !
- ◆ La grille est indispensable



WLCG = OSG + EGEE + ...



# La grille en biologie –un exemple

- ◆ WISDOM (<http://wisdom.healthgrid.org/>)
  - ★ Développer de nouveaux médicaments «*in silico*»
  - ★ Coûts réduits
  - ★ Développement accéléré  
p.ex. Malaria, grippe aviaire...



- ◆ Trois « data challenge »
  - ★ Wisdom-I : 1M molécules, 1 TO
  - ★ Grippe aviaire : 300 k molécules, 750 GO
  - ★ Wisdom-II : 125 M molécules, 2 TO

# Les sciences de la terre

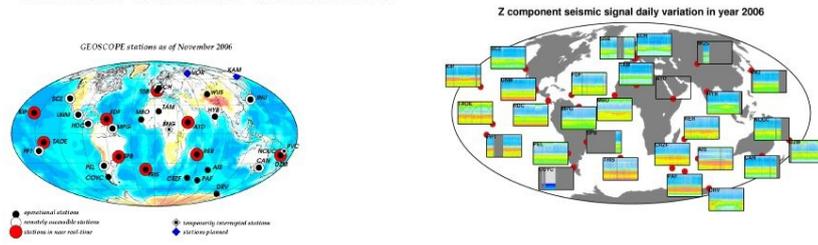
- ◆ Plusieurs domaines :
  - ★ Beaucoup de CPU, rapidement
  - ★ Calcul à long terme
  - ★ Partage de données
  - ★ Partage d'algorithmes

Etude du bruit sismique dans un large spectre de fréquences, et pour différentes durées (jour, horaire...)

IPGP: E. Stitzmann, G. Patau, IPSL/EGEE: D. Weissenbach

Geoscope: (<http://geoscope.ipgp.jussieu.fr>) IPGP-France

28 stations sismiques  
20- 25 années de données  
Calcul sur l'ensemble des données

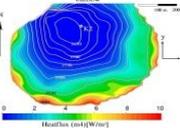
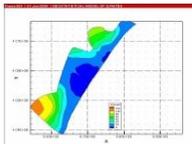
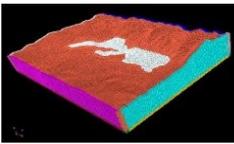


• **GEOSCOPE** (<http://geoscope.ipgp.jussieu.fr/>)

– Analyses rapides des tremblements de terre

January 6th 2008 – Southern Greece, M=5.1, origin time 05h14m17s

date = 6 Jan. 2008  
time = 5:14:17  
depth = 50.9 km  
magnitude = 6.1  
latitude = 37.150°

GEOCLUSTER	ELMER	CODESA-3D	3DSEM_UNSTRUCT
CGG-Veritas, France	CSC - Finland	CRS4 - Italy	IPGP - France
			
Logiciel Plateforme sismique (400 modules, serveur de licence)	Modèles physiques – mécanique des fluides, transfert de chaleur, acoustique...	Modélisation Intrusion d'eau de mer dans les aquifères côtiers	Modélisation Propagation 3D des ondes sismiques dans un milieu géologique complexe
VO- EGEODE	VO- ESR	EUMEDGRID	EELA

# Et en France ?



- ◆ Un institut des grilles : IDG
- ◆ Les sites EGEE

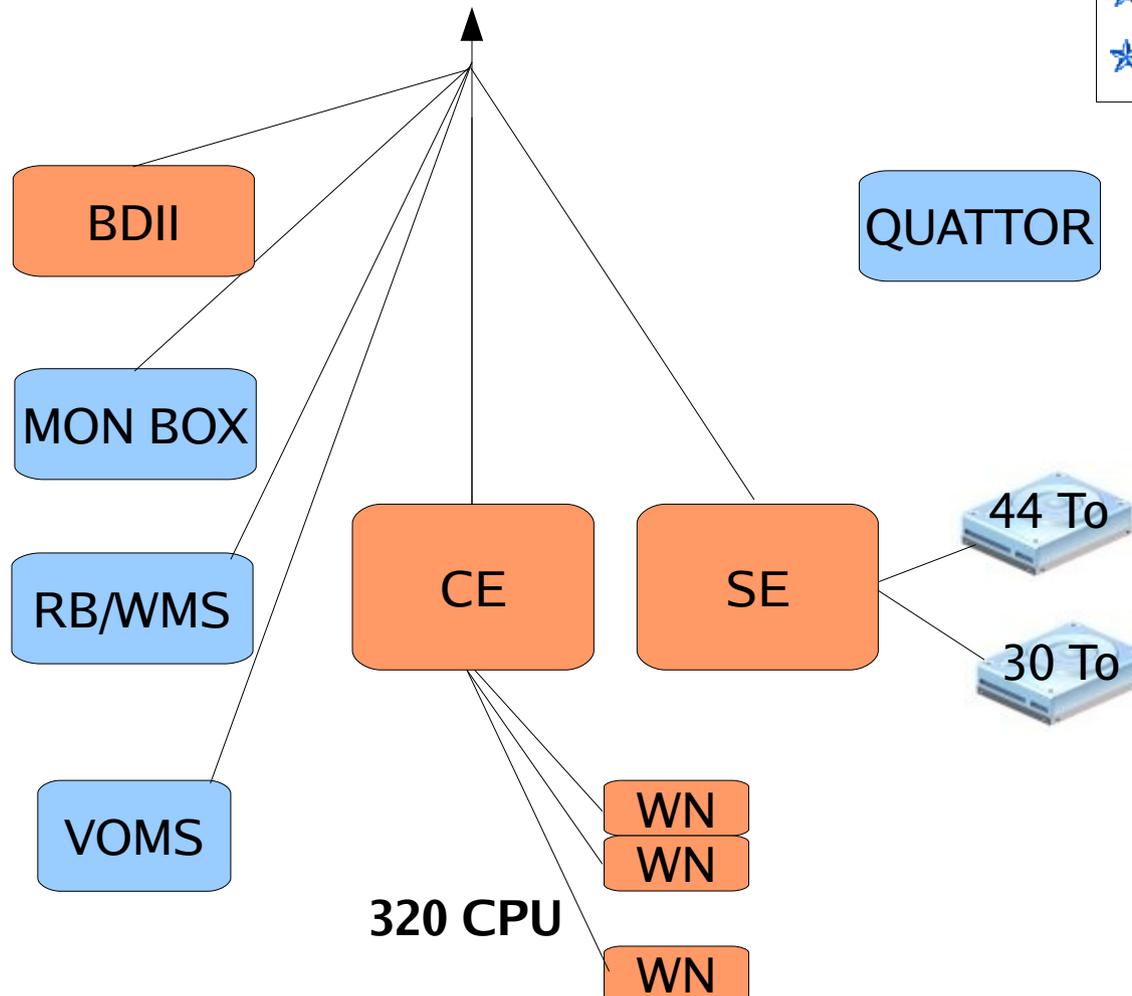


Paris	: CEA/IRFU - LAL -LLR -LPNHE -IPNO CGG/VERITAS -IPSL
Nantes	: Subatech
Clermont	: LPC -Auvergrid
Lyon	: CCIN2P3 - IPNL -IBCP
Strasbourg	: IPHC
Annecy	: LAPP
Grenoble	: LPSC
Marseille	: CPPM
<b><i>Futur</i></b>	
Grenoble	: ESRF
Montpellier	: MSFG

# Exemple de mise en œuvre

TopBDII

1Gb/s vers internet/grille



**Minimum:**

- ★ Disponibilité > 75 %
- ★ Supporter des VOs >= 1

UI

# Fonctionnalités de base

- ★ Authentification par certificat
- ★ Organisation virtuelle : VO
- ★ Gestion de tâches (jobs) –sous Linux (SL)
- ★ Gestion des données –copie, catalogues, ...
- ★ Monitoring des sites et des services

Le middleware « glite » fournit cela

