La grille de calcul WLCG

Christophe DIARRA
IPNO/IN2P3-CNRS
diarra@ipno.in2p3.fr

Avertissement

- Je fais cette présentation en tant qu'Administrateur Système d'un des sites de WLCG
- Cette présentation est une synthèse personnelle et non pas officielle de WLCG
- Elle est forcement incomplète et comporte sûrement des valeurs obsolètes
- Je me suis basé en partie sur d'autres présentations faites dans le cadre d'EGI et WLCG

Quelques définitions

- **WLCG** : Worldwide LHC Computing Grid (WLCG)
  - WLCG est la grille de calcul utilisée pour stocker et analyser les données du LHC du CERN à Génève
  - WLCG est basée en partie sur la grille EGI
- **EGI** : European Grid Infrastructure
  - Infrastructure pérenne de grille en Europe
- **CERN** : Organisation européenne pour la recherche nucléaire
- **LHC** : Large Hadron Collider (Grand collisionneur de hadrons)
  - C'est un accélérateur de particules

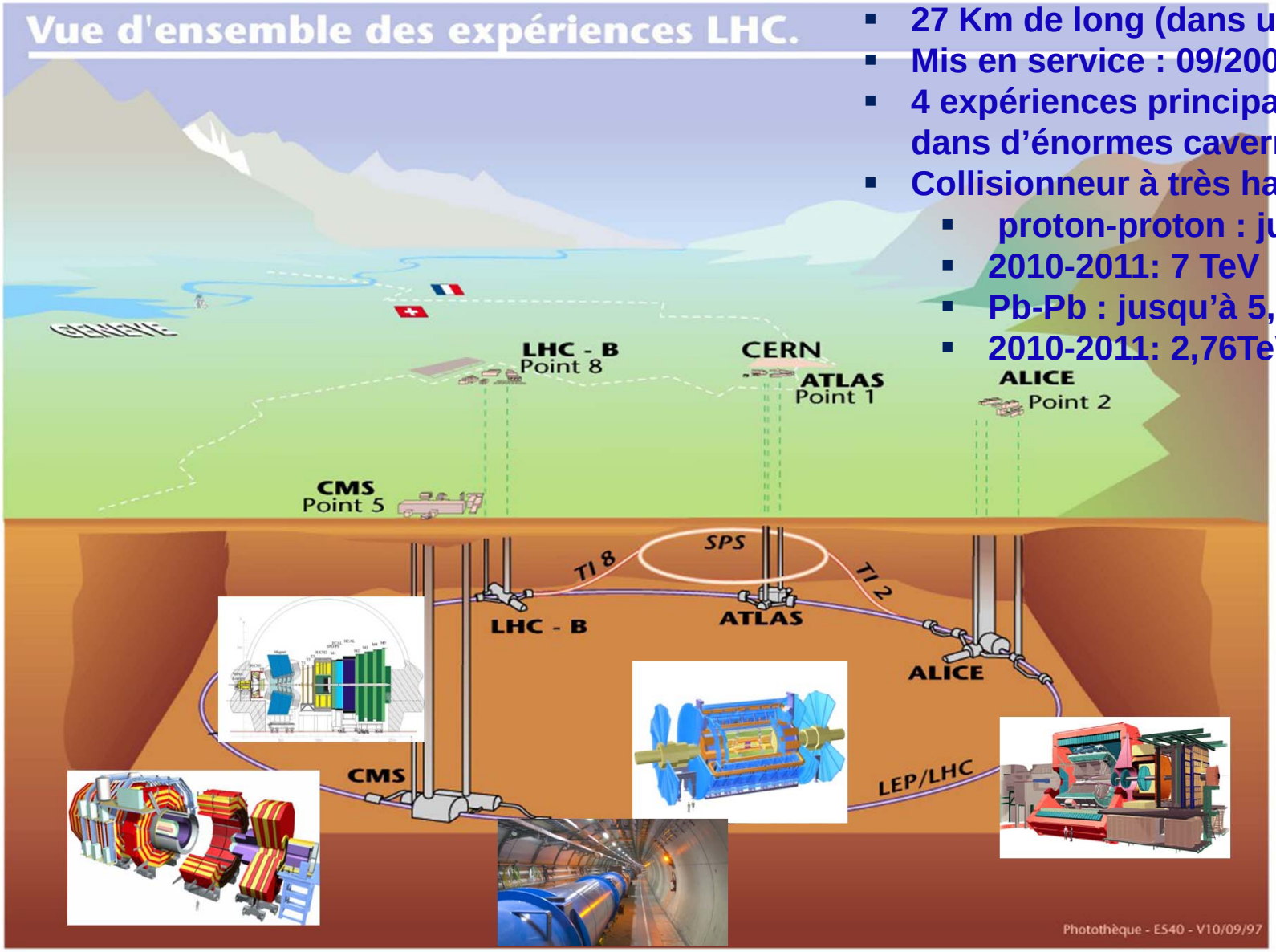
WLCG : faire face au déluge de données du LHC

- Le LHC produit 15 Péta-Octets (PO) de données par an (1 PO = 1000 000 GB)
- Où stocker ces données de façon pérenne ?
- Comment les traiter/retraiter et les rendre accessibles à 8000 physiciens à travers le monde pendant 15 ans (durée de vie estimée du LHC)
- Un supercalculateur aussi puissant soit-il ne pourrait convenir
- Solution : Interconnecter les centres informatiques participants aux projet pour créer une grille de calcul

  ➔ Un **supercalculateur virtuel** composé de ressources géographiquement distribuées

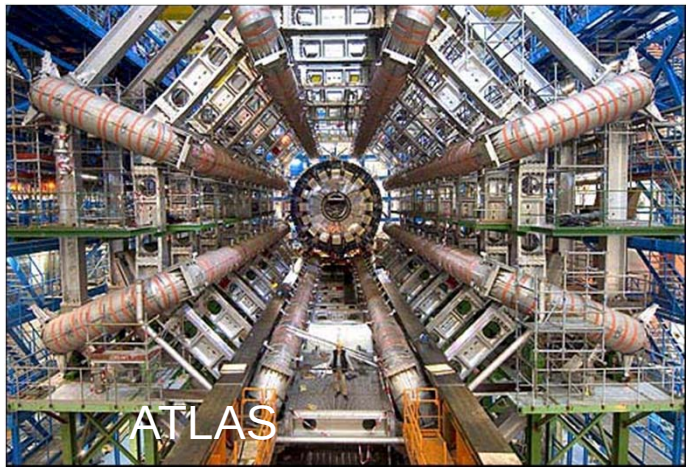
Il faut une pile de 20Km de CDs pour contenir un an de données du LHC

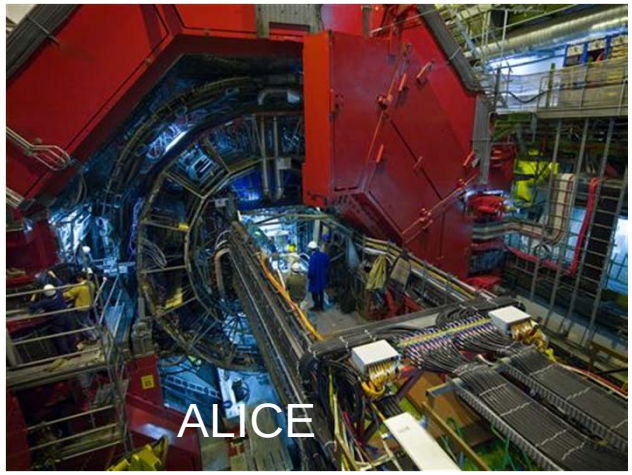
LE LHC : le plus grand accélérateur de particules au monde



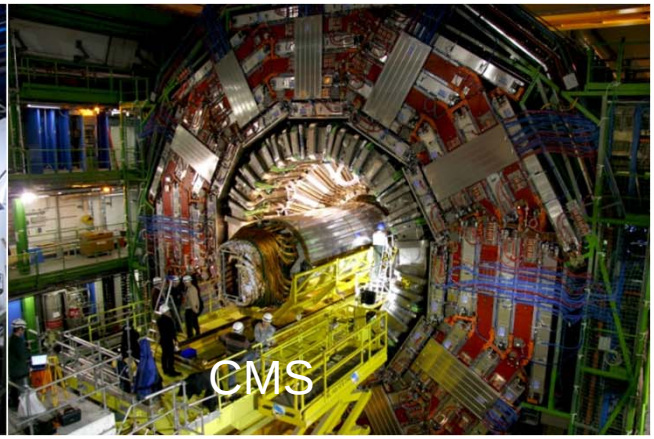
- 27 Km de long (dans un tunnel)
- Mis en service : 09/2008
- 4 expériences principales installées dans d'énormes cavernes à -100m
- Collisionneur à très haute énergie :
  - proton-proton : jusqu'à 14TeV
  - 2010-2011: 7 TeV
  - Pb-Pb : jusqu'à 5,5 TeV
  - 2010-2011: 2,76TeV

LHC : des détecteurs gigantesques



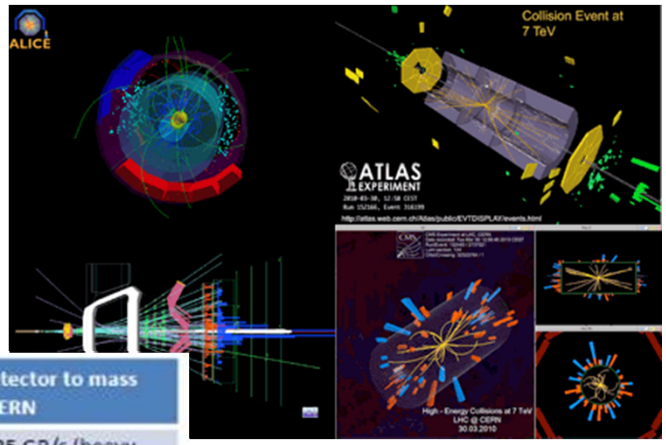


Longueur: 21 à 46m
Largeur: 13 à 25m
Hauteur: 10 à 25m
Poids: 5,6 à 10 000T



LHC: des collisions aux données

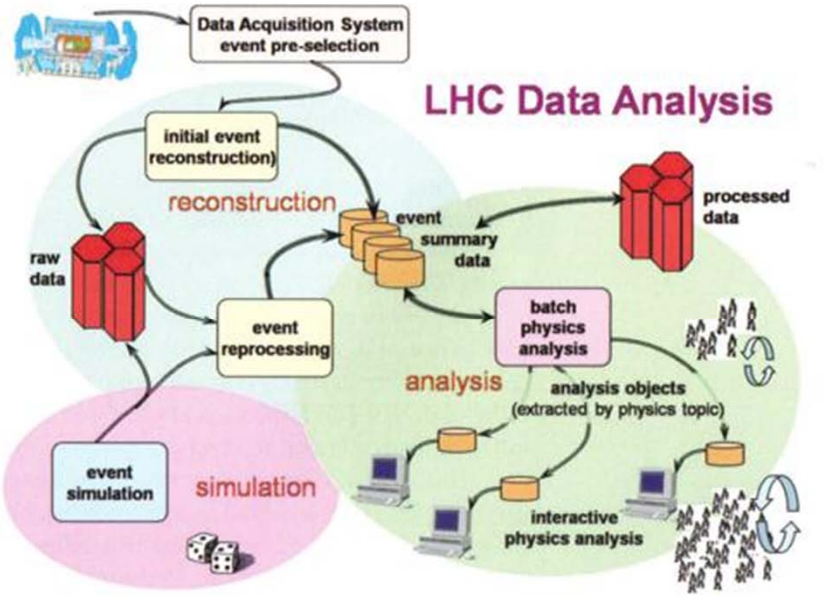
- Une collision de « packets » de protons toutes les 25ns
  - Chaque collision génère un évènement de ~1,5MB
  - 1,5MB/25ns = 60TB/s !
- L'électronique du détecteur et une ferme de calcul dédiée filtre les évènements intéressants
  - Ceci donne un taux d'évènement de ~300Hz ➔ 500MB/s
- Ces évènements (raw) sont transférés et traités au CERN pour faire la reconstruction des propriétés des particules issues de la collision
  - Les évènements reconstruits font quelques centaines de KB par évènement

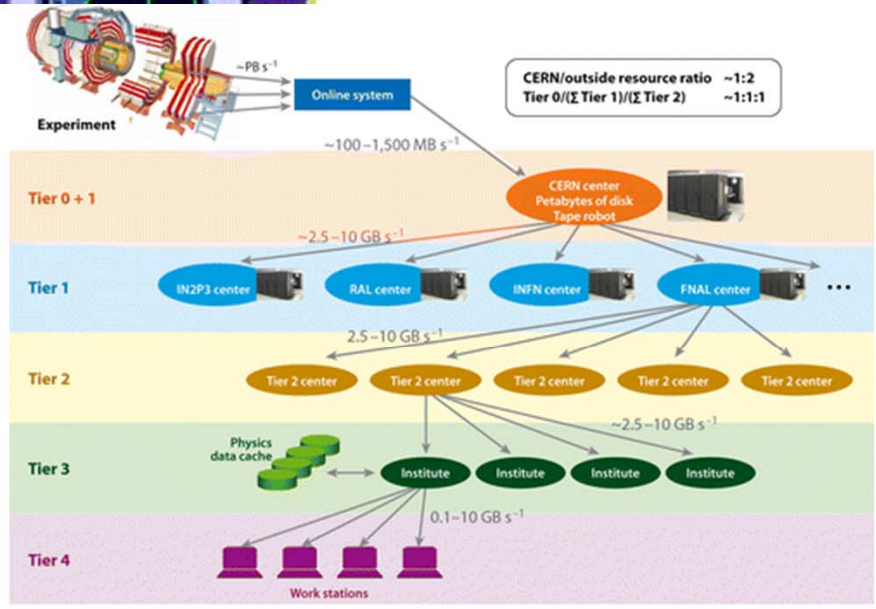


Data rates from the detector to mass storage at CERN	
ALICE	100 MB (pp), 1.25 GB/s (heavy ions)
ATLAS	320 MB/s
CMS	450 MB/s
LHCb	60 MB/s

LHC: le traitement des données

- Toutes les données stockées sont très rapidement disponibles sur la grille





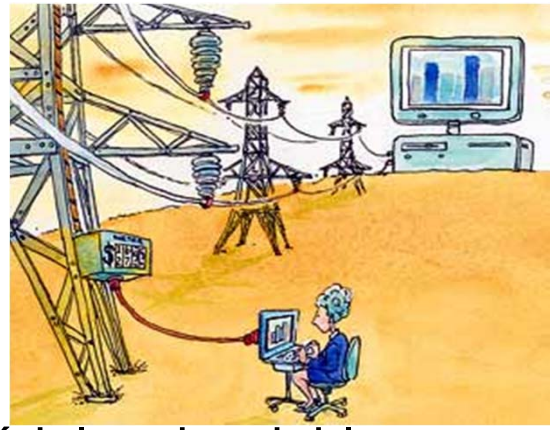
Bird I. 2011.
Annu Rev. Nucl. Part. Sci. 61:99–118

Qu’est-ce qu’une grille de calcul ?

- C’est un ensemble de ressources informatiques distribuées géographiquement , reliées à Internet et interagissant via un middleware (intergiciel)
  - Les ressources sont : les processeurs, les systèmes de stockage, les logiciels
  - Le matériel informatique est hétérogène et géré de façon décentralisée
  - Grace au middleware les utilisateurs disposent d’un supercalculateur virtuel

Avantage d'une grille de calcul

- Vue homogène de ressources hétérogènes
  - Chaque site choisit son matériel et sa configuration
  - Le Middleware cache cette hétérogénéité
- Collaboration facilité entre scientifiques
  - Partage des données et des logiciels de façon transparente à des milliers de km
- Mutualisation des ressources
  - Evite la sous-utilisation et le surdimensionnement des ressources
- Coût modéré
  - Ressources souvent basées sur du matériel « abordable »



Avantage d'une grille de calcul (suite)

- Reconfiguration et découverte dynamique
  - Les nouvelles ressources sont automatiquement disponibles pour les utilisateurs
  - Les ressources indisponibles sont automatiquement ignorées
- La grille est tolérante aux pannes et offre une redondance des données
  - La grille peut vivre avec certaines ressources manquantes
  - Les données peuvent être répliquées sur des sites différents
- Support et surveillance 24h/24
  - Grace aux décalage horaire, il y a toujours une équipe pour surveiller la grille et faire du support

Différents type de grille

- Grille des supercalculateurs : exemple DEISA
  - Chères, procédure d'allocation « lourde », bien adaptée aux applications fortement parallèles
- Grilles institutionnelles : exemple EGI
  - Moins chères, allocation des ressources simplifiée, peut accueillir des applications très variées
- Grilles de PC « desktop grid » : BOINC (cf SETI@Home), Xtremweb, EDGeS
  - Beaucoup moins chères, ressources offertes bénévolement (« cycle sharing »), applications exigeant peu de volumes de données

EGI : European Grid Infrastructure

- EGI.eu (Amsterdam) est une fondation créée en 2010
- EGI.eu est l'organisation qui coordonne et gère la grille EGI en Europe
  - EGI fédère les infrastructures de grilles nationales
  - EGI.eu repose sur les efforts de ses membres : CERN, NGIs, EMBL
  - Chaque NGI (National Grid Initiative) est responsable du bon fonctionnement des ressources de grille qu'elle offre
- EGI.eu a pour but de pérenniser l'infrastructure de grille existante en l'ouvrant à toutes les disciplines scientifiques tout en intégrant les innovations sur le calcul distribué

EGI : Historique

- 2001-2010 : projets européens sur les grilles
  - European DataGrid (EDG)
  - Enabling Grids for E-SciencE (EGEE)

➔ Buts : Créer une grille européenne pour la communauté scientifique

- Avril 2010 : fin EGEE début EGI
  - S'appuie sur les NGI (National Grid Initiative)
  - En France la NGI (GIS France Grilles) est pilotée par l'Institut des Grilles et du Cloud du CNRS

➔ Buts : pérenniser l'infrastructure de grille en Europe

- EGI comporte 37 NGI + CERN

EGI : le middleware

- Le middleware est une couche de logicielles qui se situe entre le système d'exploitation et les applications scientifiques
- C'est une série de programmes et de protocoles qui permettent aux utilisateurs d'accéder aux ressources de la grille
- Il permet d'agréger différents services et ressources hétérogènes (processeurs, stockage, …)
- Le middleware développé par EGEE s'appelle gLite
- La transition est en cours de gLite vers EMI (European Middleware Initiative)

EGI : les composants du middleware

L’architecture du middleware comprend :

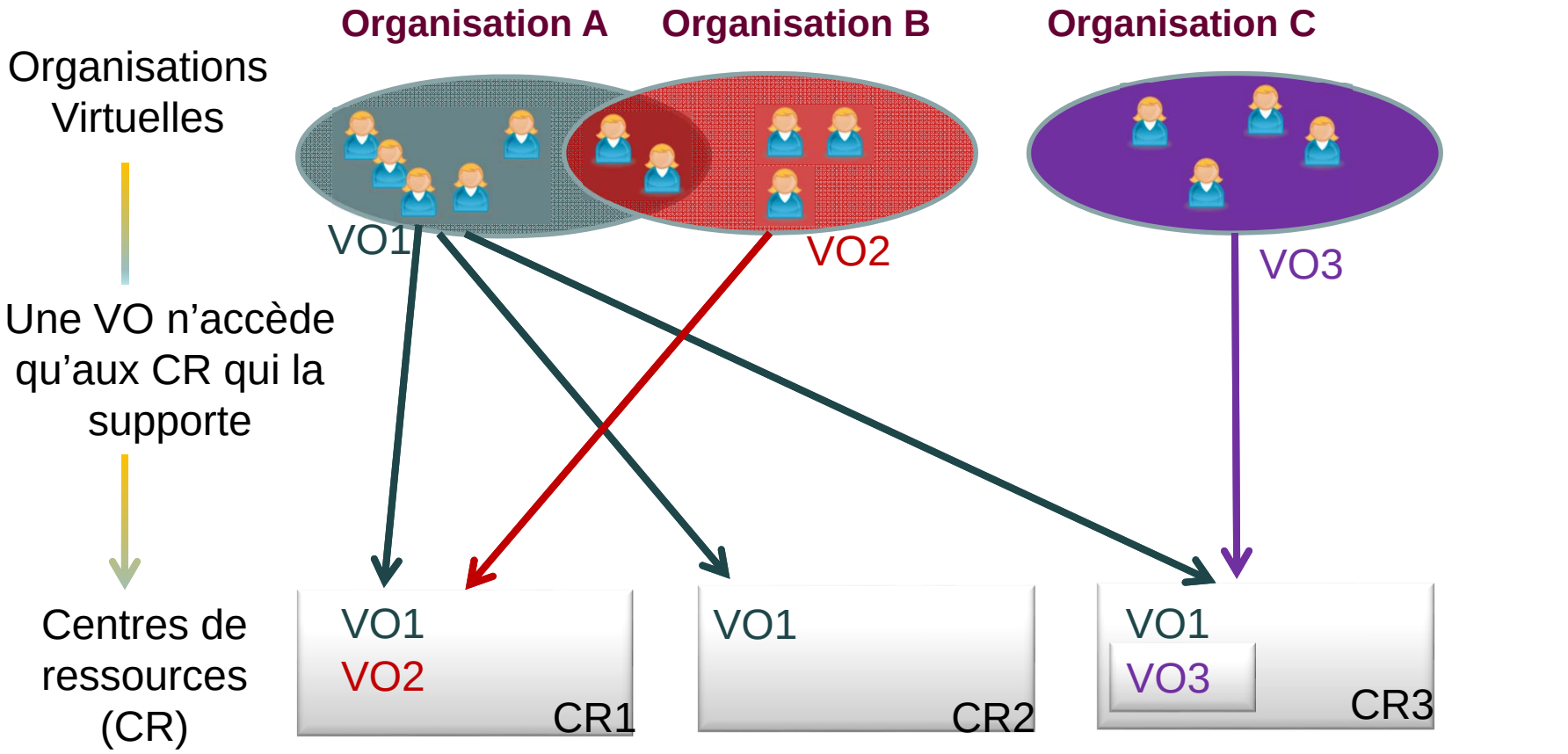
- Un système d’autorisation et d’authentification
- Un système de gestion de jobs
- Un système de gestion des données
- Un système d’information
- Un système d’accounting (comptabilité)
- Un système de monitoring (supervision)

EGI : Comment ça s'utilise ?

Pour utiliser la grille EGI :

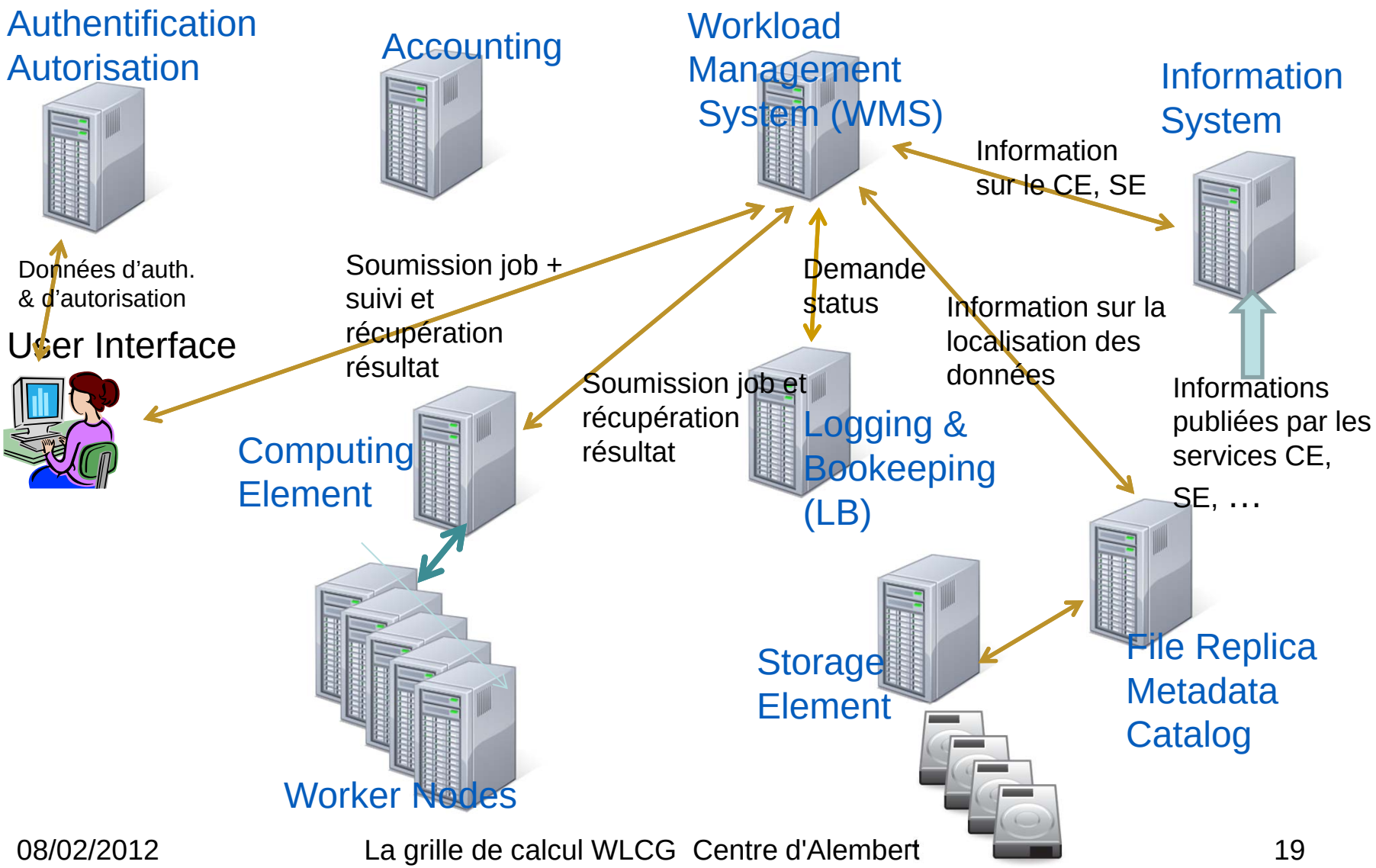
- Il faut posséder un certificat électronique
- S'inscrire dans une Organisation Virtuelle
  - Une VO (Virtual Organization) est une communauté d'utilisateurs ayant des intérêts communs et qui utilisent la grille pour collaborer et partager des ressources
- Se servir des utilitaires ou outils (fournis) pour transférer des données
- Soumettre un job (traitement par lot) pour pour faire tourner un code de calcul

EGI : les Organisations Virtuelles

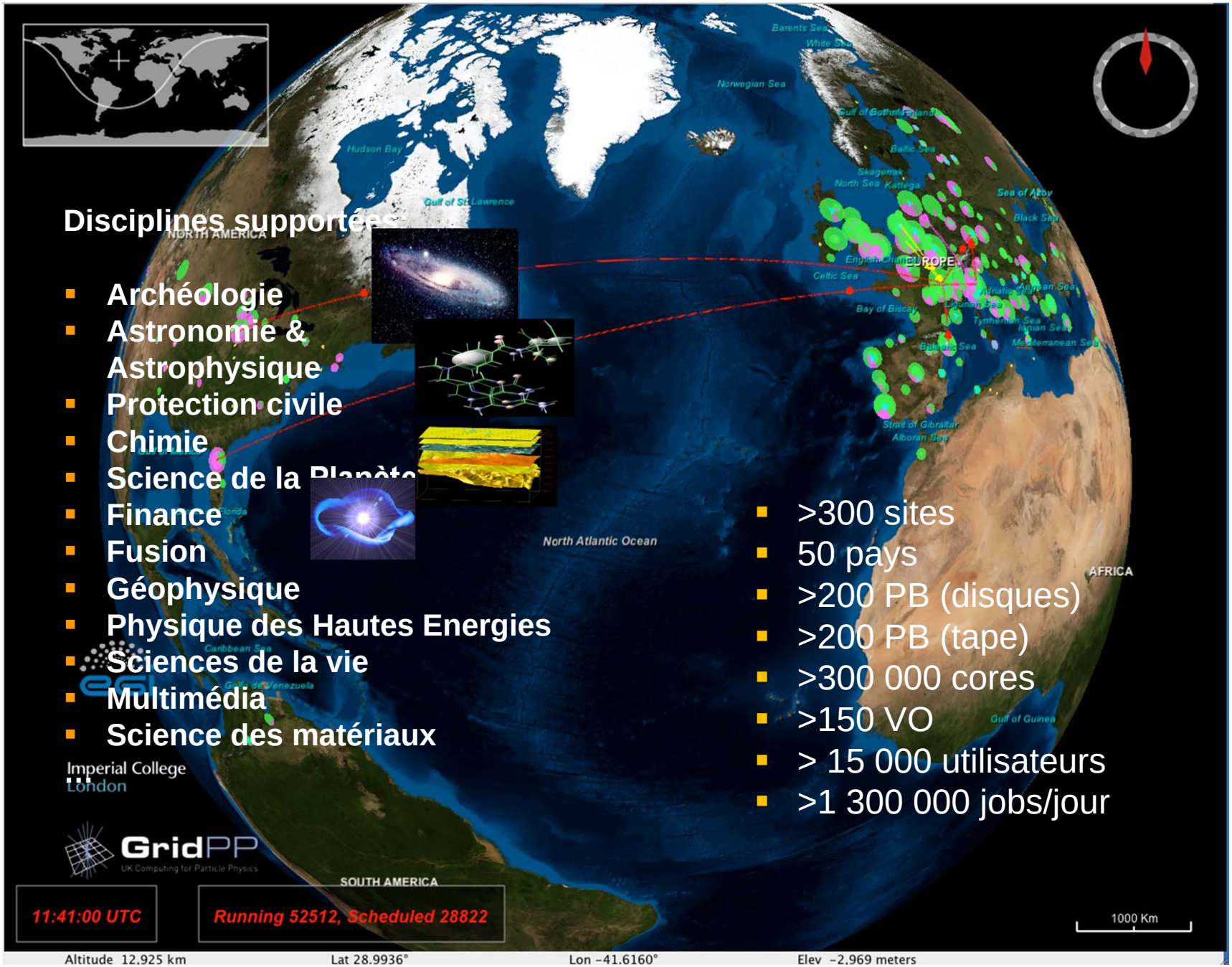


Un Centre de Ressources (ou site) est un ensemble de ressources sur lesquelles est déployé le middleware, les rendant accessibles via la grille.

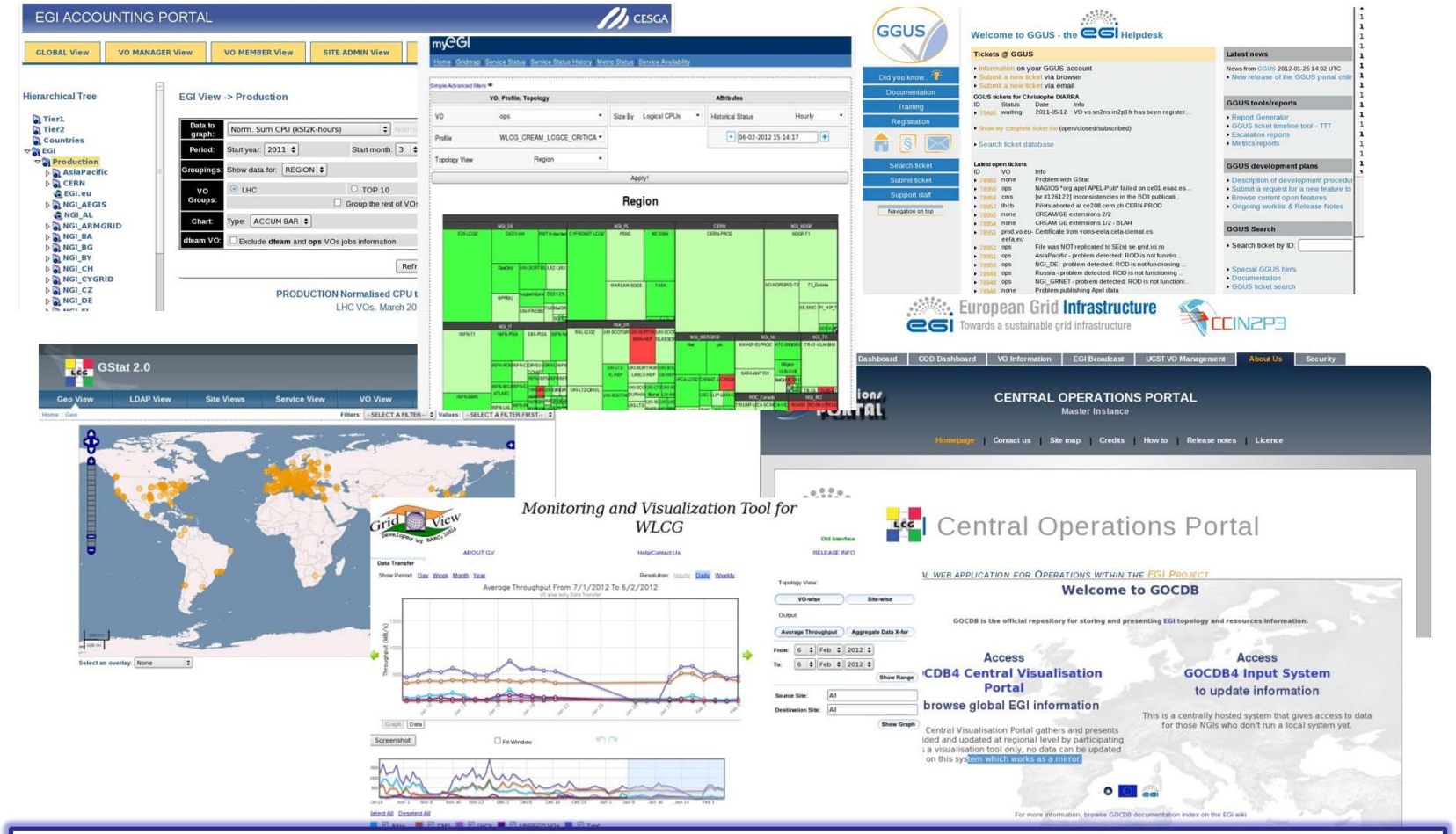
EGI : déroulement d'un job



EGI : en quelques chiffres



EGI : infrastructure opérationnelle

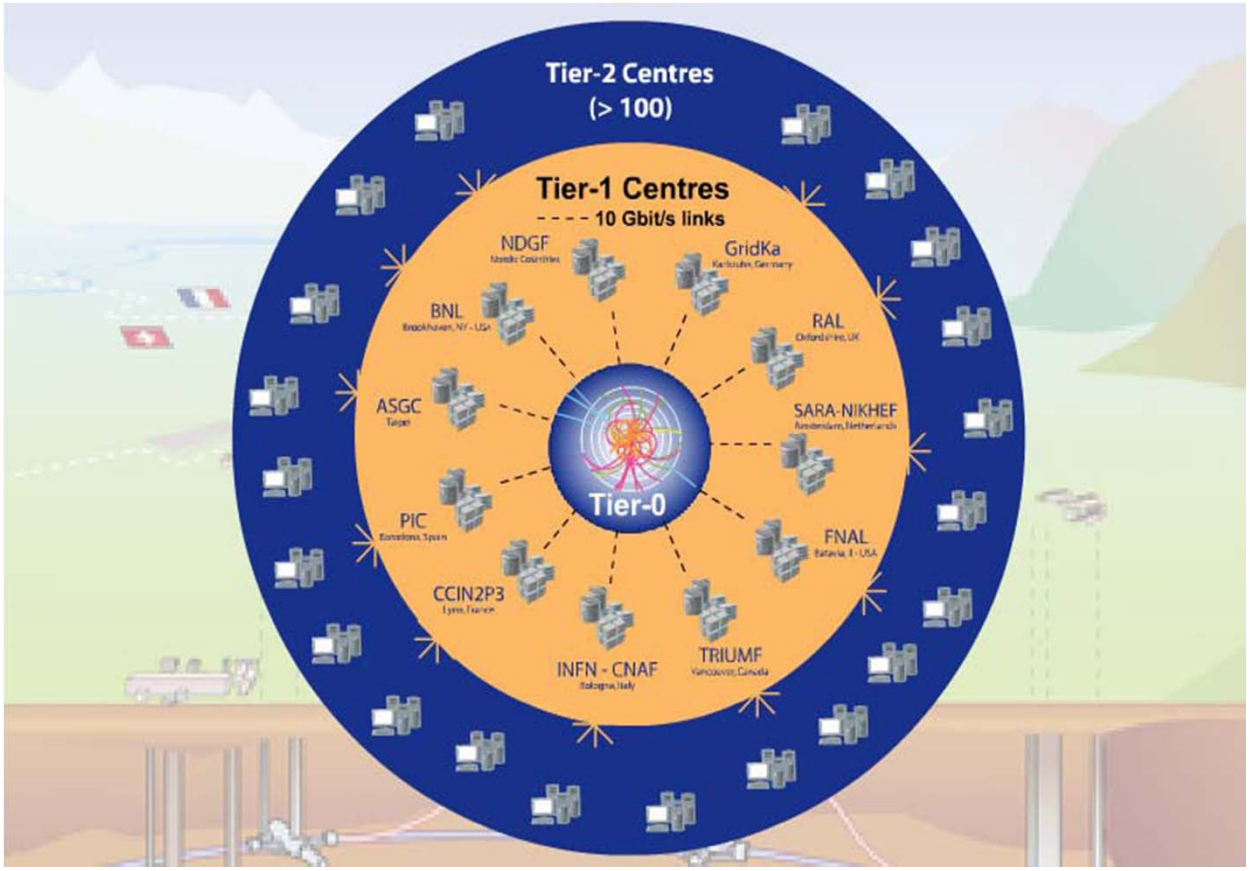


Plusieurs outils pour veiller au bon fonctionnement de la grille

La grille WLCG

- C’est la grille utilisée pour stocker et analyser les données produites par le LHC
- Elle est en production depuis 2005
- Elle repose sur les sites de EGI supportant les VOs du LHC, le CERN et la grille OSG (Open Science Grid) aux USA
- Les sites grille de WLCG sont classés en « Tier » selon une hiérarchie dépendant de leurs tailles et du niveau de qualité de service à respecter
  - Tier-0 (CERN)
  - Tier-1 (11 grands centres nationaux)
  - Tier-2 : les autres sites
  - NB.: il existe des Tier-3 qui offrent des ressources uniquement pour des usages « privés » des sites eux-mêmes

WLCG : hiérarchie des services



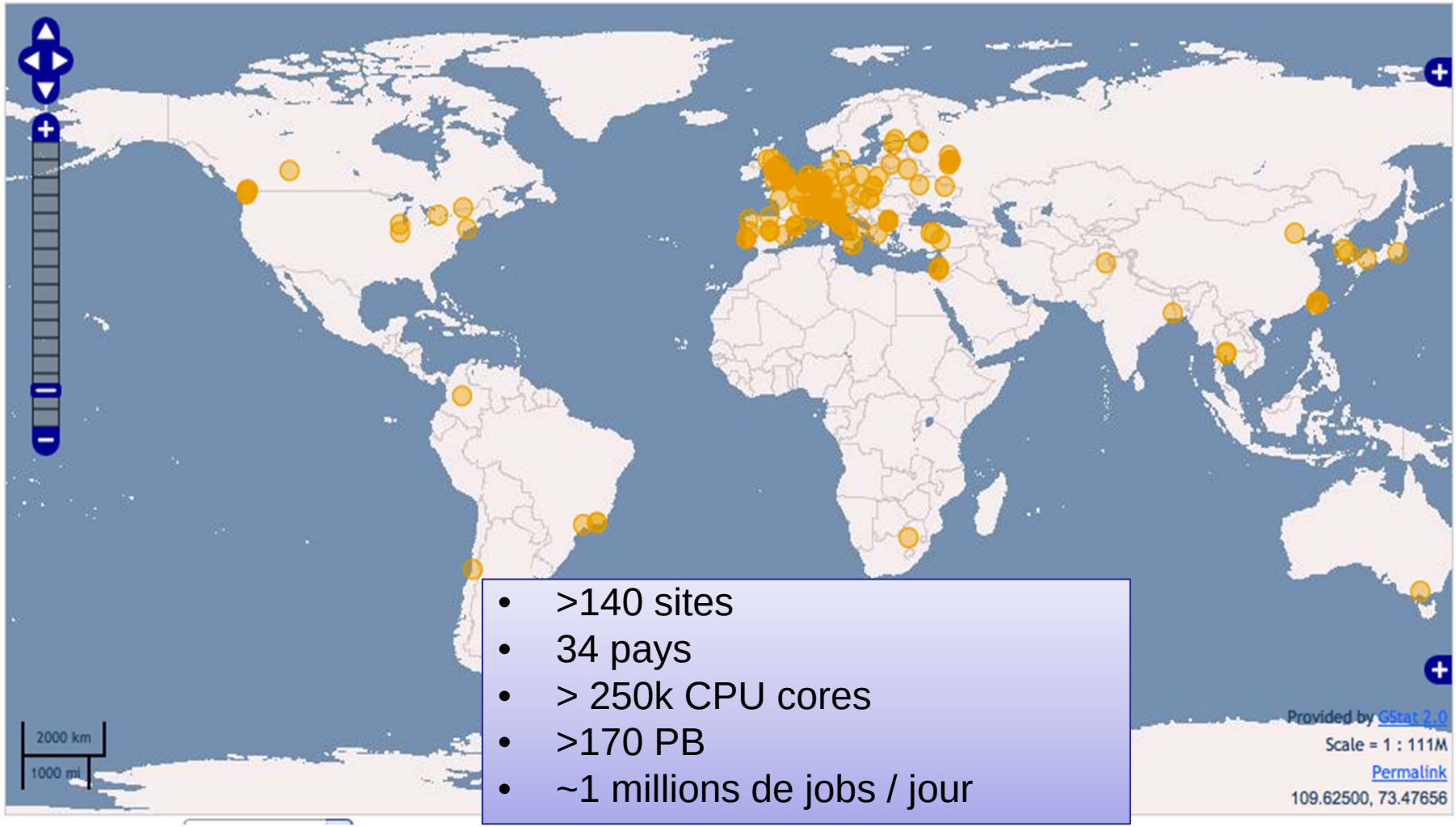
Tier-0 (CERN)
- Stockage des données
- Reconstruction
- Distribution des données

Tier-1 (11 centres)
- Stockage permanent
- Re-processing
- Analyse

Tier-2 (>140 centres)
- Simulation
- Analyse par les utilisateurs finaux

Chaque Tier1 reçoit une fraction des données du T0
Chaque fraction de ces données est repliquée sur au moins un autre T1

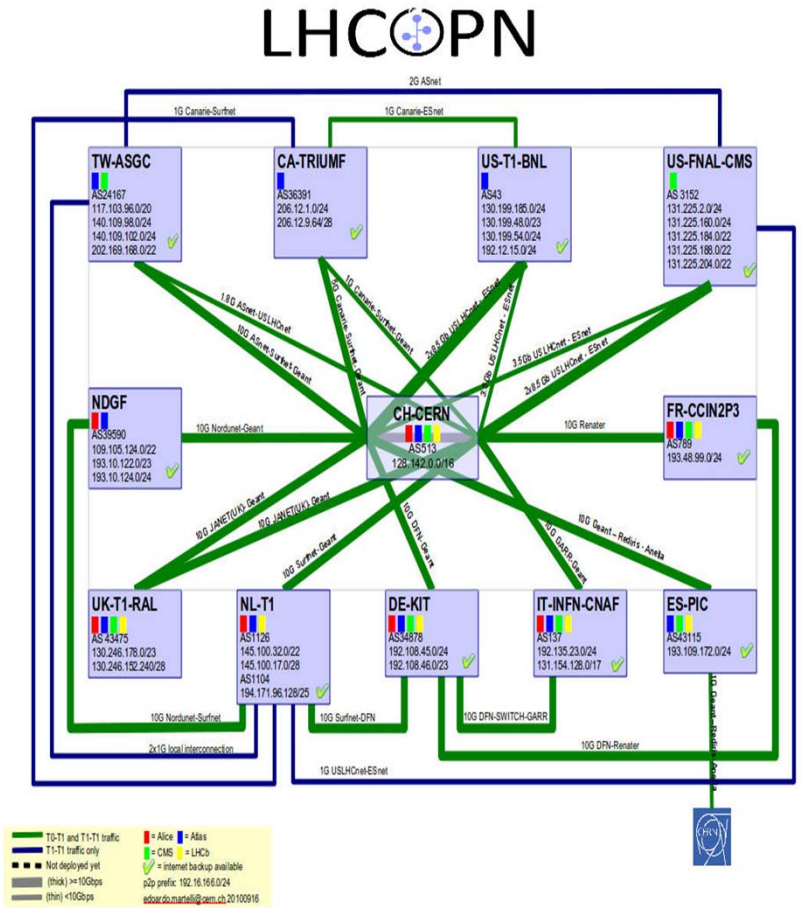
WLCG : une collaboration mondiale



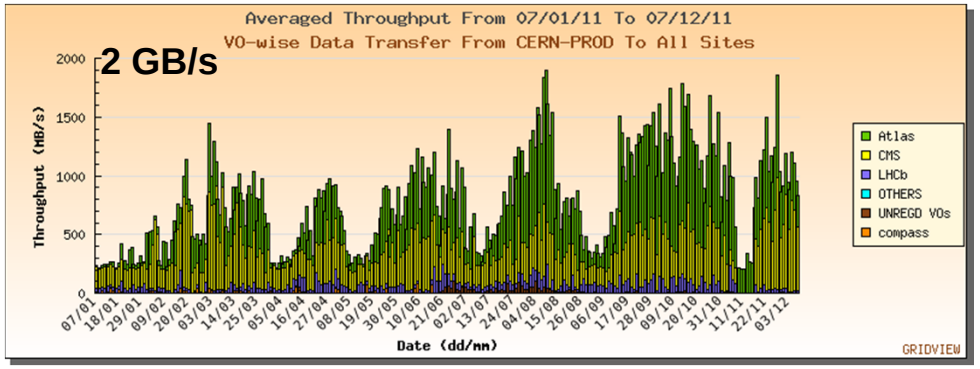
- >140 sites
- 34 pays
- > 250k CPU cores
- >170 PB
- ~1 millions de jobs / jour

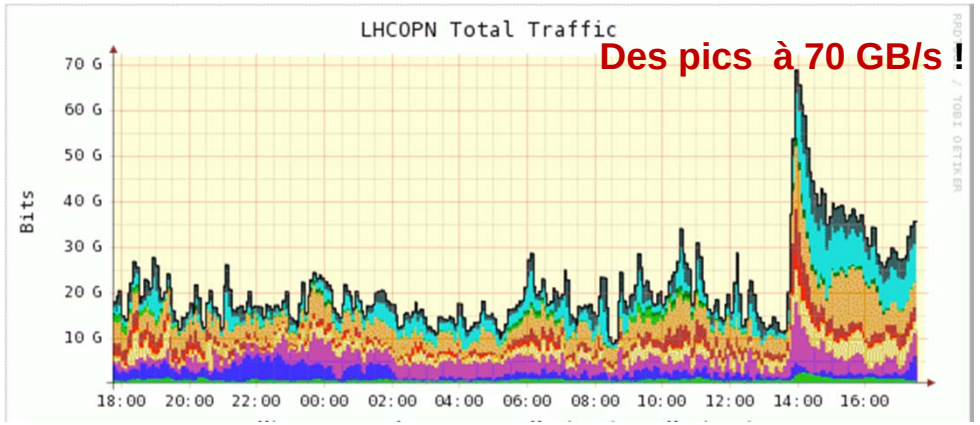
WLCG : les transferts de données

- **Les volumes de données transférées nécessitent un réseau fiable et avec un très haut débit.**
- **WLCG repose sur des réseaux privés optiques (LHCOPN, LHCONE),**
- **US-LHCNet, GEANT, les NRENs**



Des taux de transfert moyens très élevés et soutenus

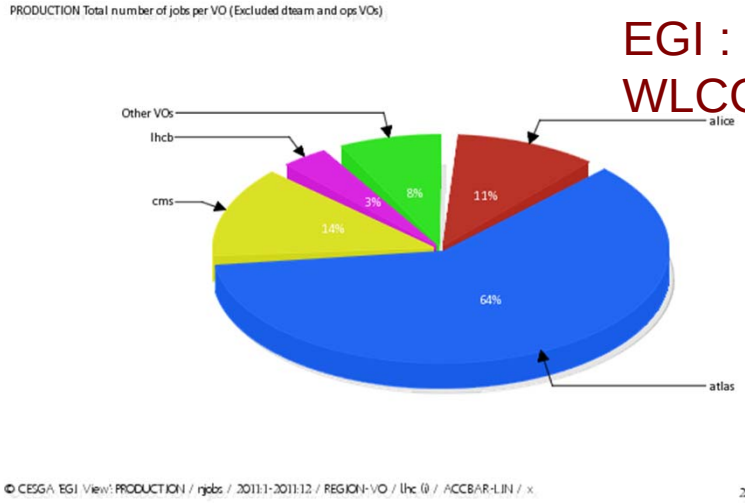


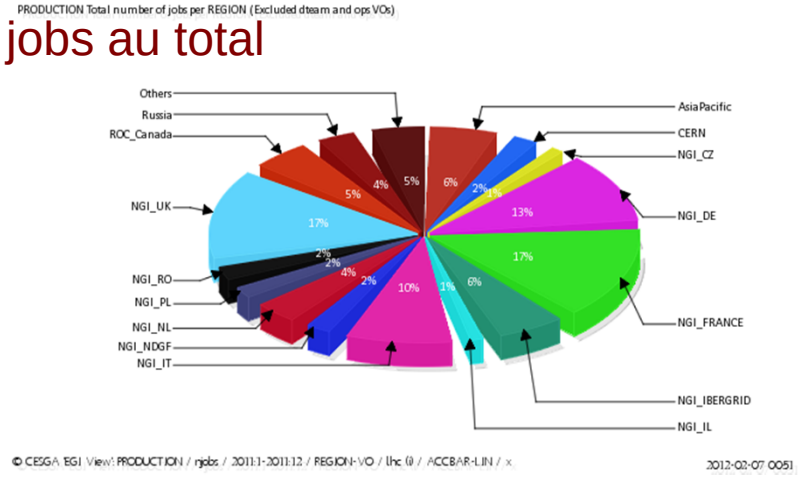


WLCG : taux d'utilisation (2011)

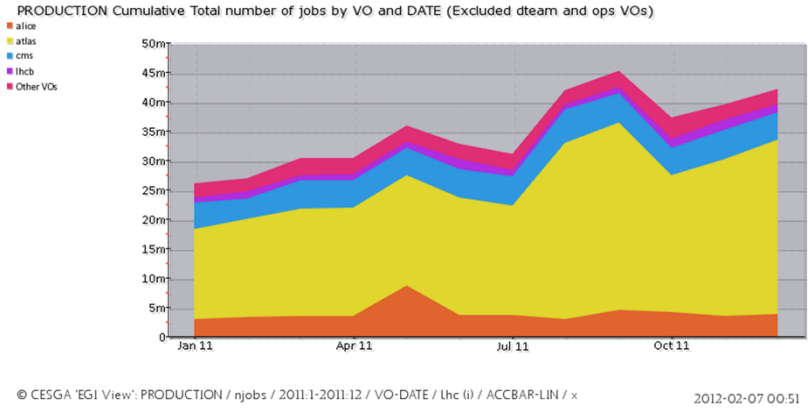
EGI : >416 000 000 jobs au total

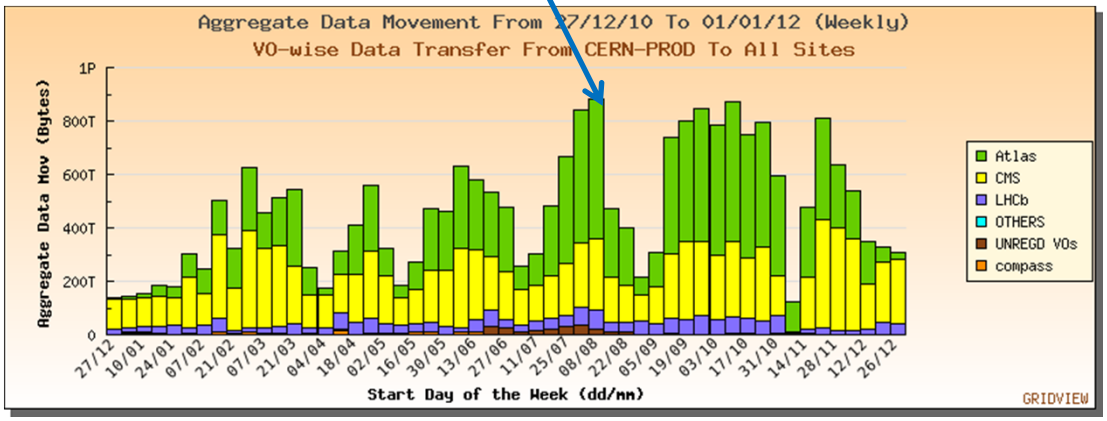
WLCG → 94%



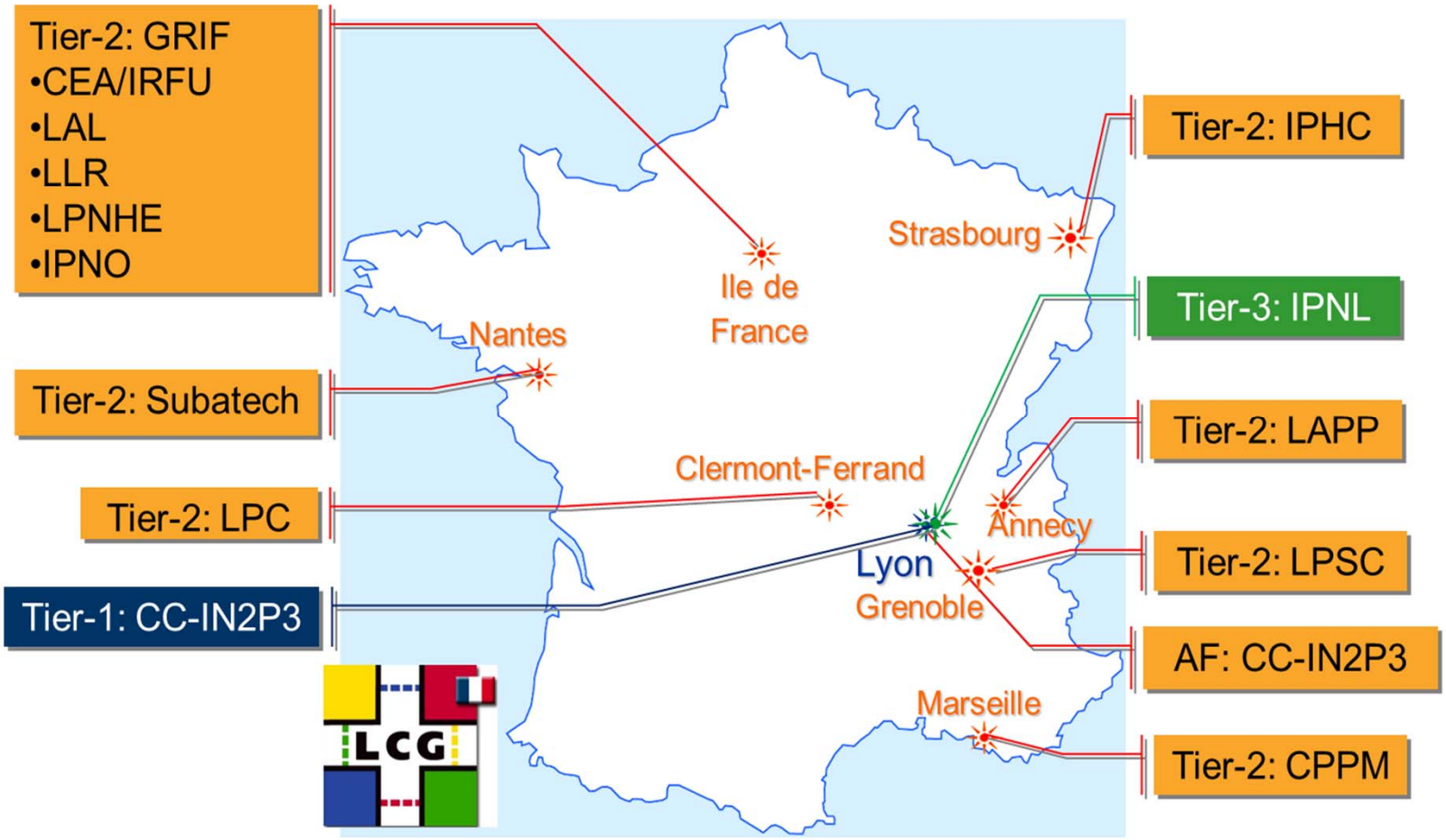


Pic de ~1PB de données transférées du CERN/semaine





WLCG : LCG-France

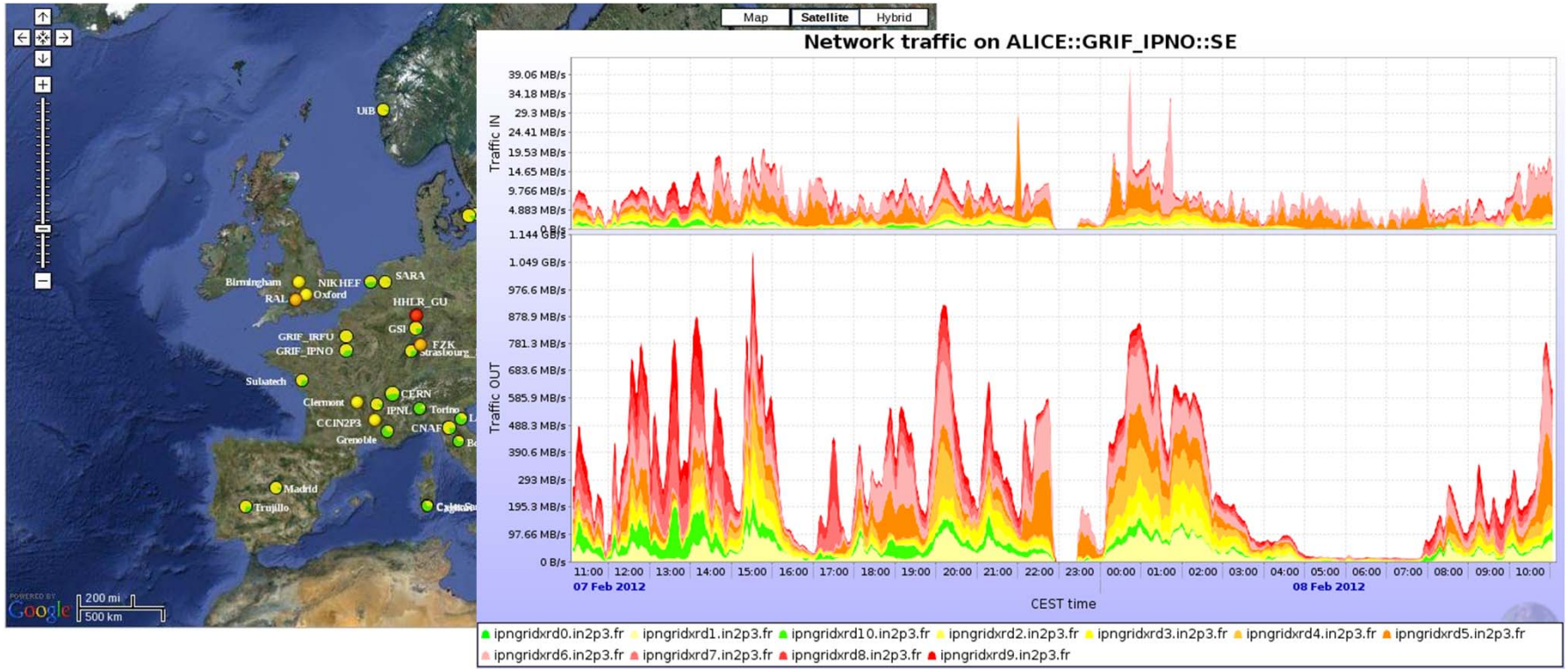


Tier-2 GRIF : Grille au service de la Recherche en Ile-de-France

- GRIF : plusieurs sites fédérés mais vus comme un site grille unique
- 12 clusters dont 3 dédiés au calcul parallèle
- ~8500 CPU cores
- >3PB de stockage disque
- Plus de 40 VO supportés
  - 4 VO LHC
  - Autres VO : psud, biomed, ILC, babar, dzero, fusion, grif, ipno, …
- Etroite collaboration entre les équipes techniques
- GRIF est relié à LHCONE

GRIF à l'IPNO

- GRIF a une vocation multi-disciplinaire
- L'IPNO supporte les VO IPNO, ALICE (LHC), AGATA, PANDA, PSUD, MURE, GRIF, …
- Plusieurs utilisations hors LHC : radio-chimie (calcul parallèle : CPMD, VASP, CP2K, GAUSSIAN, …), physique théorique, AUGER, …

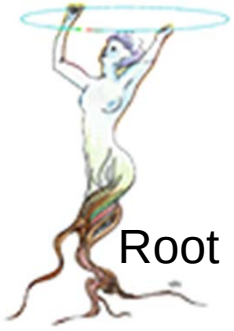


WLCG : outils des expériences

Pionnières sur la grille, les expériences du LHC on développé des outils dont certains sont devenus incontournables et disponibles pour la communauté scientifique

Ces outils permettent le traitement des données et simplifient grandement l'utilisation de la grille

PANDA



Root

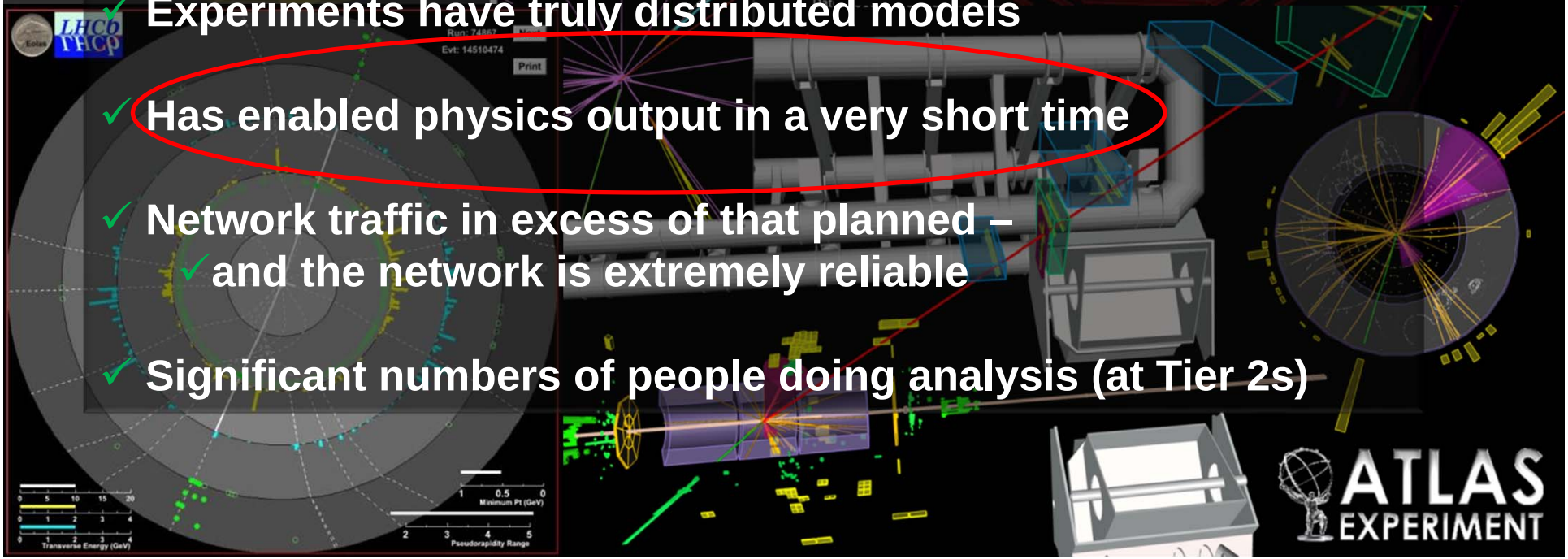
Conclusion : WLCG marche !

Extrait d'un transparent de Ian BIRD du CERN, Chef de Projet WLCG
Lors du WLCG Collaboration Workshop, DESY, 11th July 2011:

"WLCG: The 1st 18 months with data & looking to the future"

Successes:

- ✓ We have a working grid infrastructure
- ✓ Experiments have truly distributed models
- ✓ Has enabled physics output in a very short time
- ✓ Network traffic in excess of that planned –
  - ✓ and the network is extremely reliable
- ✓ Significant numbers of people doing analysis (at Tier 2s)



Liens utiles ?

- http://www.egi.eu
- http://lcg.web.cern.ch/lcg/
- http://public.web.cern.ch/public/
- http://grif.fr
- http://www.gridcafe.org/index.html