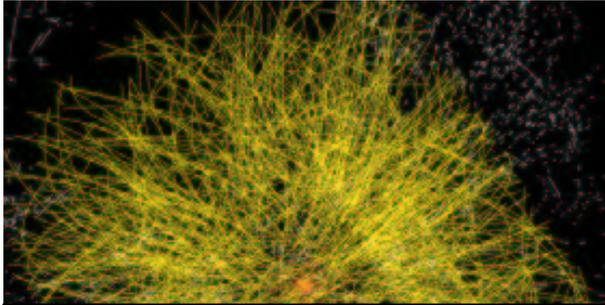


# Utilisation effective de la Grille par ATLAS

S. Jézéquel (LAPP)

# Prochaine étape

## Collecter et analyser les collisions



Traitement des données:  
100.000 PCs récents

Stockage: CD empilés sur 20 km

## LHC: Défi pour calcul scientifique

25 millions collisions /s/expt

100-200 collisions sélectionnées /s

Taille d'un événements : 1,5 MB  
(~1 photo 5 Mpixels)

Production de 100-1000 MB/s de données  
Soit 15 PBytes/an



# Solutions mises en oeuvre

Constat: CERN : 20 % des ressources nécessaires

Les pays mettent à disposition leur centre de calculs nationaux

➤ Répartition du calcul à travers la planète

Besoin: Mise en production d'outils de 'Grille' (EGEE en Europe) pour permettre

- l'échange automatique et rapide
  - de programmes informatiques
  - des données des détecteurs LHC venant du CERN
  - des résultats des calculs des différents centres
- le lancement de programmes de calculs répartis sur la planète (l'outil 'Grille' faisant le choix optimal des sites)

# Les centres de calcul accessibles par la 'Grille'



*A map of the worldwide LCG infrastructure operated by EGEE and OSG.*

**140 centres de calcul répartis dans 35 pays**

# Précisions sur le calcul ATLAS

Analyse d'événements sur collisionneur:

Les événements sont indépendants les uns des autres

- On peut traiter les événements séparément
- Utilisation des CPUs en mode scalaire

Besoin de 2 Go de mémoire par coeur (description du détecteur)

- Machine bien plus chère
  - A plus long terme, mettre assez de mémoire du carte mère
- Début de réflexion pour paralléliser le traitement des données

Exemple d'utilisation récente  
de la Grille  
par les expériences LHC

# Installation du soft ATLAS

**Soft ATLAS installé sur chaque site de manière permanente:**  
Evite de recompiler tout le soft à chaque job  
S'assure que le site a la bonne version

**Nb de sites à gérer: > 100**  
**Nb de versions de soft: > 10**

**Gérer centralement par qq personnes**  
**Installation sur site par job Grille**

**Problèmes :**  
S'assurer que le site a assez d'espace disque pour stocker les softs  
Décalage entre versions d'analyse et de simulation

# Méthode de soumission des jobs: Ressource Broker

RB envoie jobs contenant le descriptif des applications sur les sites  
(push)

**Important:**

**Vérifier régulièrement la disponibilité des composants  
nécessaires à votre application  
(Faire des SAM tests dépendant de l'expérience)**

**Sinon: utilisateur doit découvrir lui-même les sites problématiques !!!!**



# Méthode de soumission des jobs: Pilot job



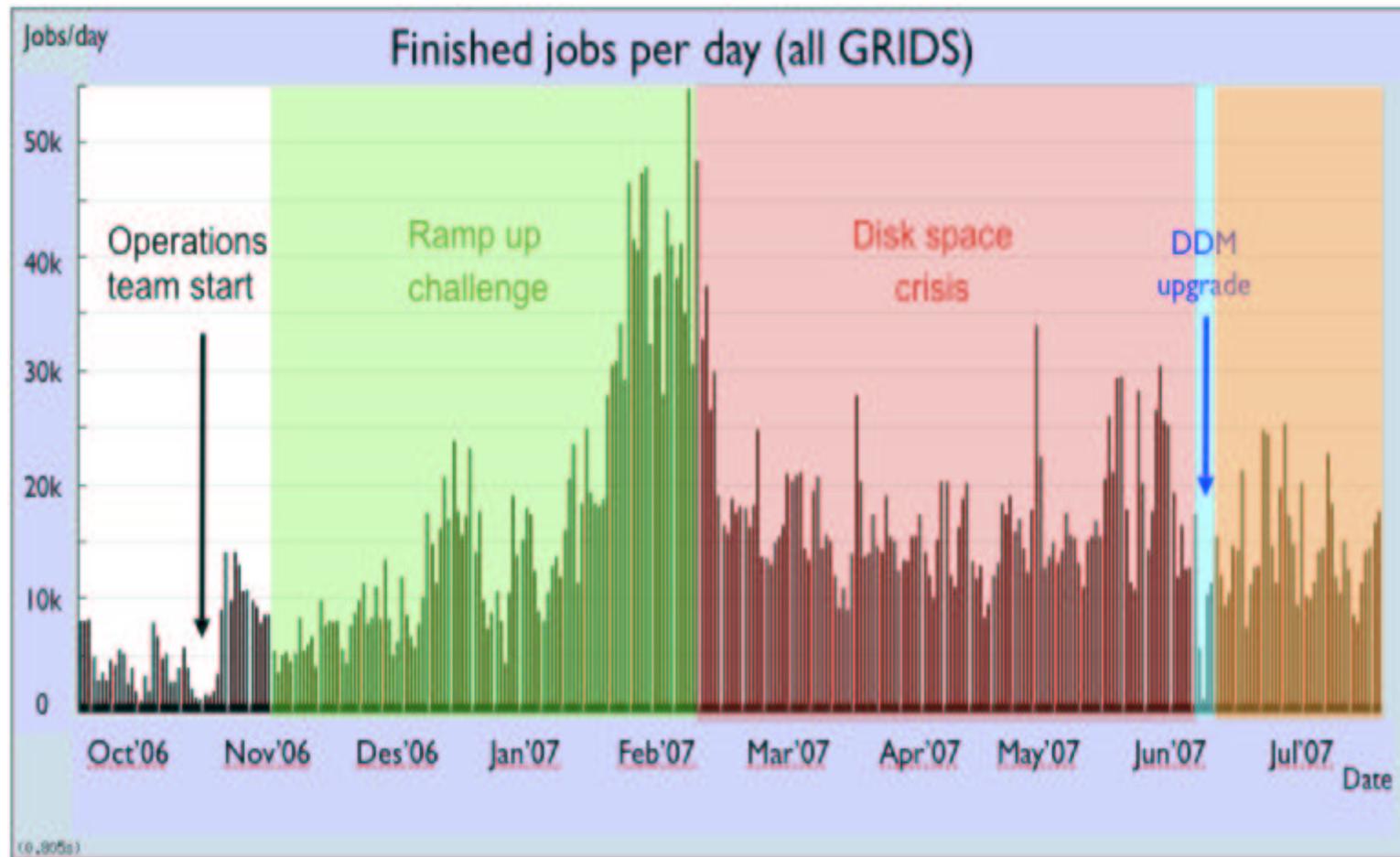
**Pros (utilisateur):** Le système découvre lui-même et en ligne les ressources  
OK

**Cons (site):** Ce système génère de l'activité sur le site même si il n'y a pas d'activités

Problème de partage des ressources sur sites multi-VO  
Problème de sécurité pour le site (savoir qui fait quoi)

Effort actuellement pour répondre aux problèmes de site

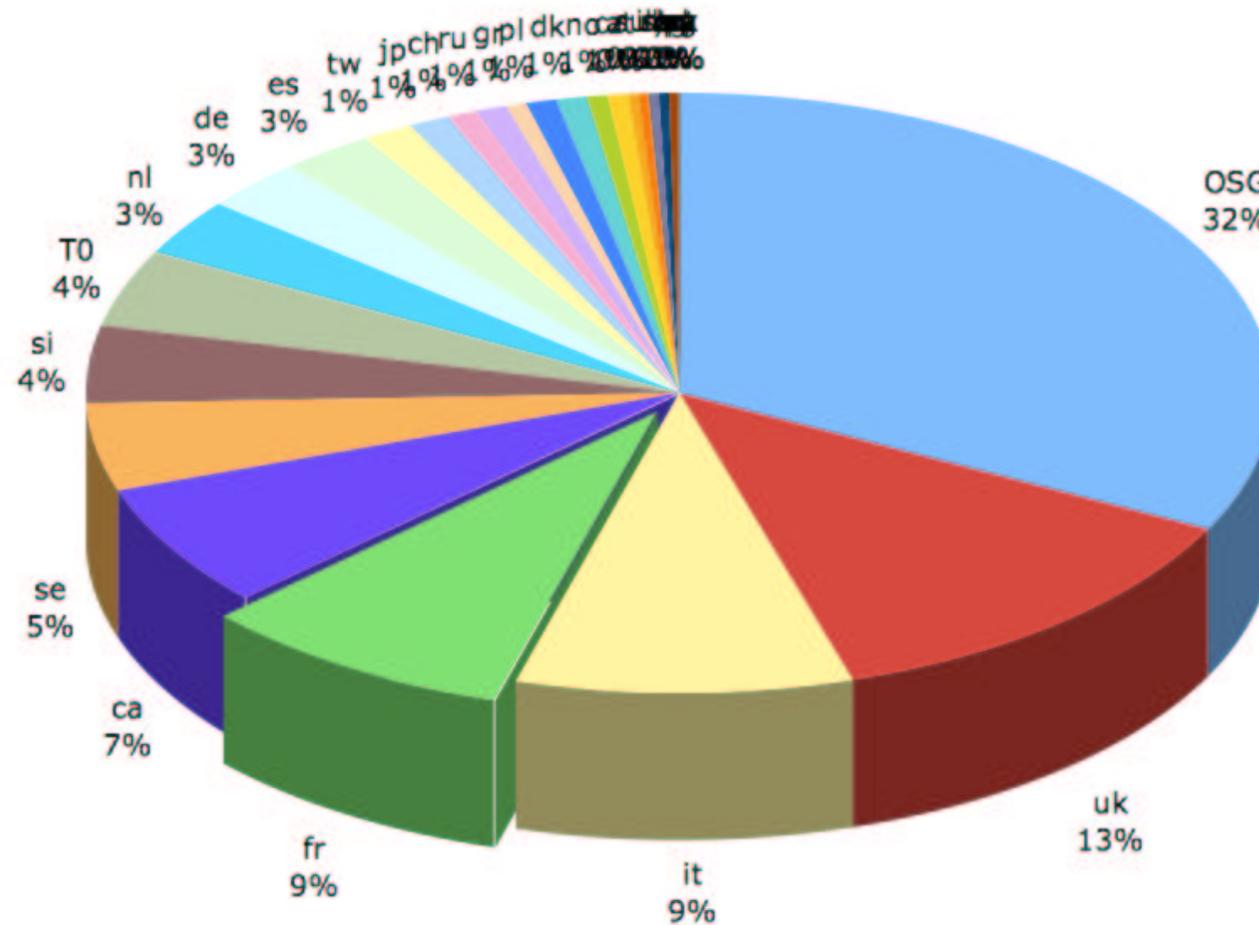
# Soumission centralisée de production



**But : 1000k jobs par jour**

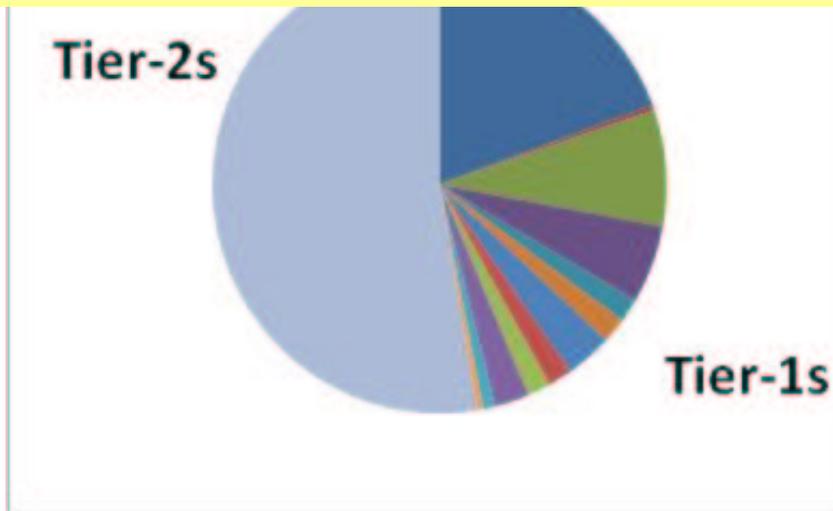
# Répartition par pays (ATLAS)

2006 : Wall Time

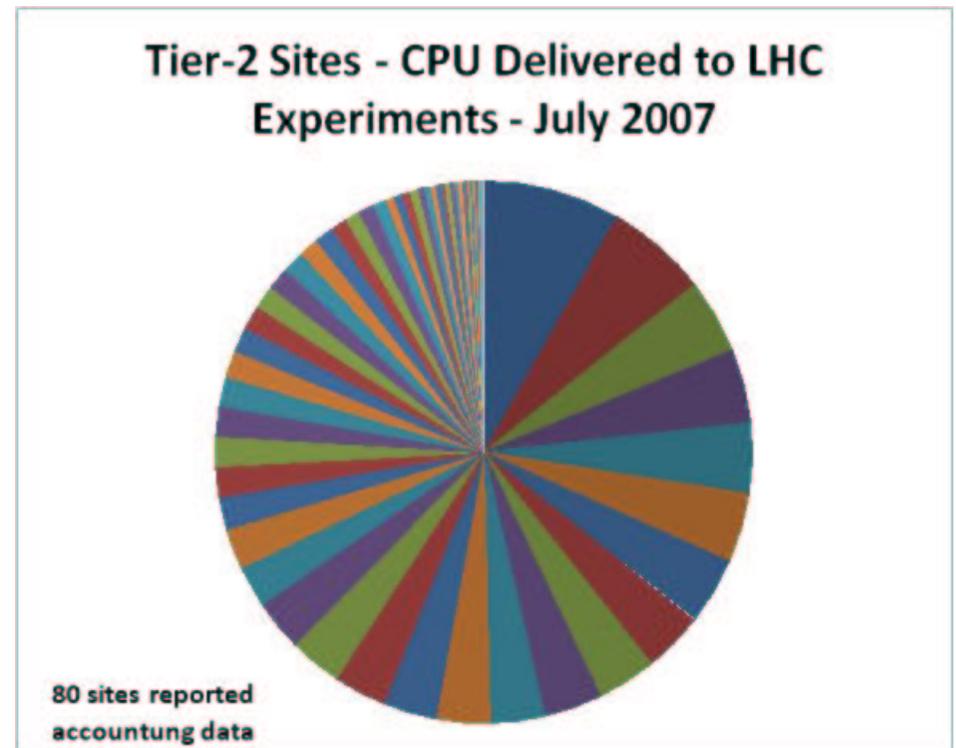


# Répartition du calcul LHC entre sites (Juillet 2007)

Ne pas négliger de pouvoir accéder à un maximum de 'petits' sites



<b>CERN</b>	<b>20%</b>
<b>11 CC nationaux</b>	<b>30%</b>
<b>80 CC locaux</b>	<b>50%</b>



# Organisation des fichiers en dataset

**Fichier: Contient une liste d'événements indépendants**

**Plusieurs milliers de fichiers par canal de physique**

**Travailler au niveau d'un fichier n'a de sens que pour 'debugger':  
A ne pas faire sur la Grille**

**Notion de dataset: Liste de fichiers ayant les mêmes paramètres physiques  
en input**

**Dataset peut contenir de 100 à 10k fichiers (attention limitation technique)**

**Question : Que contient un dataset ?**

**Job/une série de jobs Grille = Traite un dataset**

**Principe: Un job Grille tourne là où sont les données (efficacité)**

**Question: Où sont les datasets (fichiers produits dans le monde entier)?**

# Que contient un dataset ?

**Catalogue (DDM) de datasets (spécifique à l'expérience) :**

**pour chaque dataset:**

liste de fichiers caractérisés par leur guid

status (fini ou en cours de remplissage)

Liste des sites où se trouvent les datasets

Metadata pour expliquer l'histoire du dataset

**pour caractériser chaque fichier, on peut ajouter:**

**Taille (redondant avec LFC): Intégrité des données**

**Checksum: Intégrité des données (gourmand en CPU)**

**Un nom (aller chercher dans arborescence LFC)**

**Pas d'adresse physique**

**Catalogue localisé au CERN (réplication en read-only à envisager)**

**Contenu sur l'intérêt de physique du dataset:**

**Autre catalogue externe (AMI spécifique à ATLAS)**

# Où est le dataset et ses fichiers?

**Introduction: Les fichiers sont produits n'importe où dans le monde:**

**Catalogue DDM contient**

**la liste des sites contenant tout ou une partie du dataset :**

**pour connaître la liste des fichiers sur chaque site:**

**Consulte catalogue LFC associé au site:**

**Input : liste des guid du site**

**Middle : pour chaque guid, la liste des replicas**

**Output: la liste des guids sur le site**

**Prime : Adresse physique des fichiers (utiliser par les jobs)**

**Attention :**

**rien n'assure la consistance entre la liste LFC et les fichiers physiques**

**→ besoin de vérifier régulièrement la consistance**

**Opération longue et prenant des ressources**

**Mettre les données sur des sites sérieux**

# Quel catalogue LFC ?

DDM a la correspondance SE-catalogue LFC

Un catalogue LFC pour LHC a besoin d'une base de donnée robuste:  
Choix de Oracle (restreint à des gros centres = T1 ou T0)

**2 possibilités:**

**1 catalogue LFC central (option LHCb) en mode write**

**+6 replicas en mode read-only**

**Pour : facile à gérer**

**Contre : LFC central supporte une gross charge et ne doit pas s'arreter**

**N catalogues LFC (option ATLAS) en mode w/r**

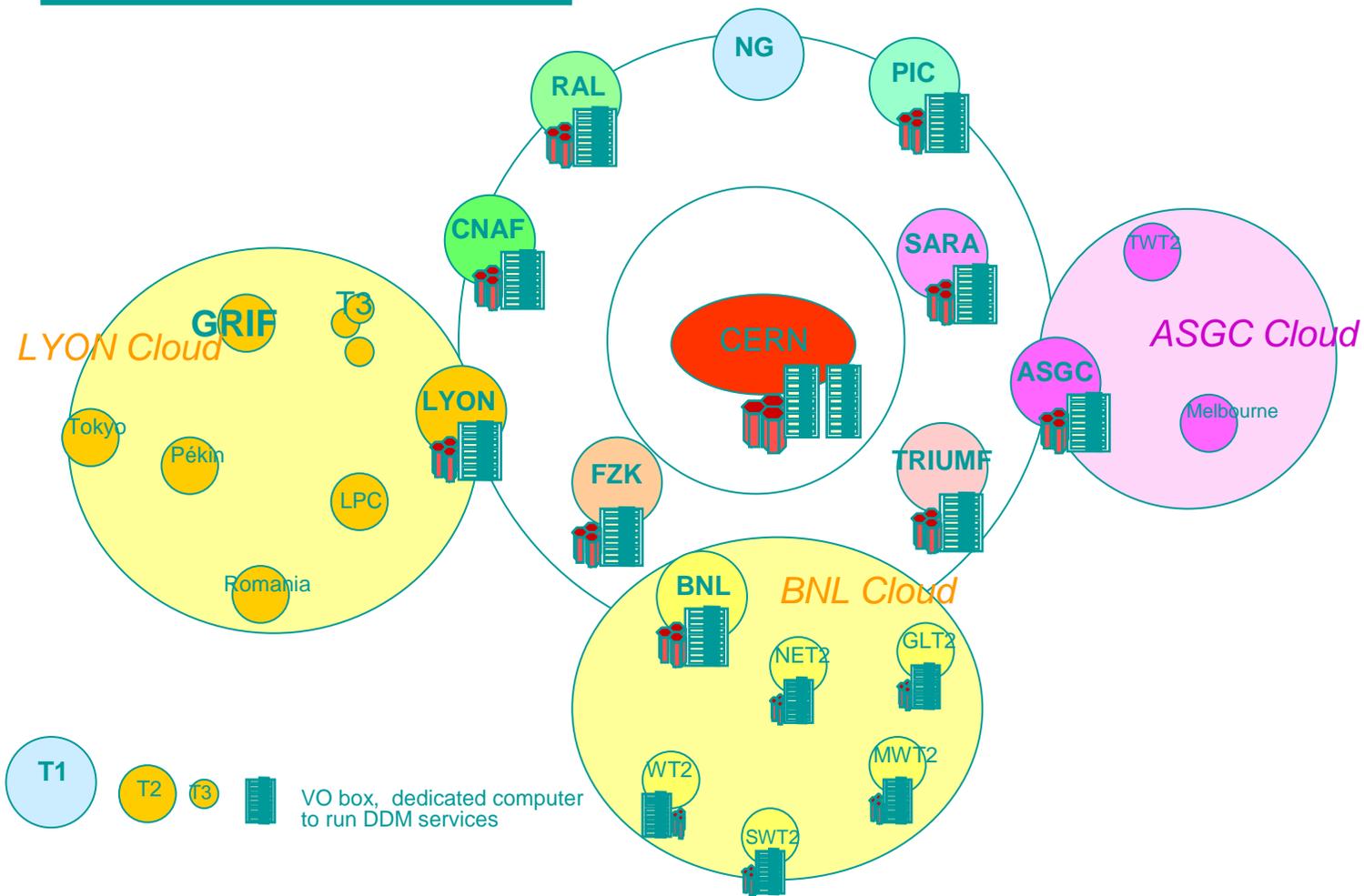
**chaque catalogue contient l'info d'un groupe de sites**

**Pour : Moins de charge sur chaque catalogue = scalabilité**

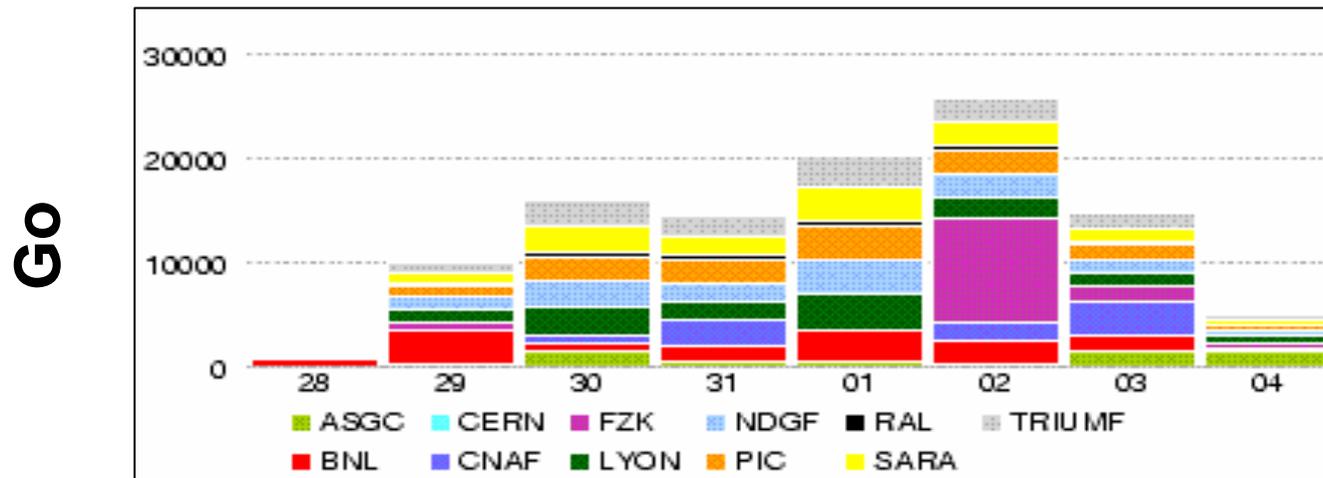
**Contre : Plus grande dépendance des sites**

# Organisation par nuage

“Tier Cloud Model”  
Unit : 1 T1 + n T2



# Transfert de datasets



## Recette ATLAS (outil Distributed Data Management)

- **DDM** fournit liste des replicas.
- **Utilisateurs peut choisir le/les sites de départ**
- **DDM** scanne les catalogues LFC des sites choisis et choisit les fichiers physiques a transferer
- **DDM declenche des transferts FTS**
- Une fois les transferts terminés, **DDM** enregistre les nouveaux fichiers physiques dans LFC
- **DDM** publie une nouvelle liste de replicas

# Rôles des sites (ATLAS)

- **Tier 0 (=CERN)**
  - Lieu de fonctionnement des détecteurs (= collecte des données)
  - Premier traitement des données réelles pour avoir premier bilan de la prise de données
  - Contient certains catalogues centraux
- **Tier 1 (Centres de calcul nationaux = GROS SITES)**
  - Réception des données en ligne du CERN
  - Reprocessing des données réelles
  - Réception des données simulées dans Tier2
  - Fournit services Grille au T2 (LFC,FTS)
- **Tier 2 (Centres de calcul locaux)**
  - Production de données simulées
  - Analyse finale des groupes de physique

# Transferts: Contraintes de mise en oeuvre

Transfert: Echange entre 2 sites (= 2 sites opérationnels)

**1 seul site doit gérer le transfert sinon conflit (peut être un tiers)**

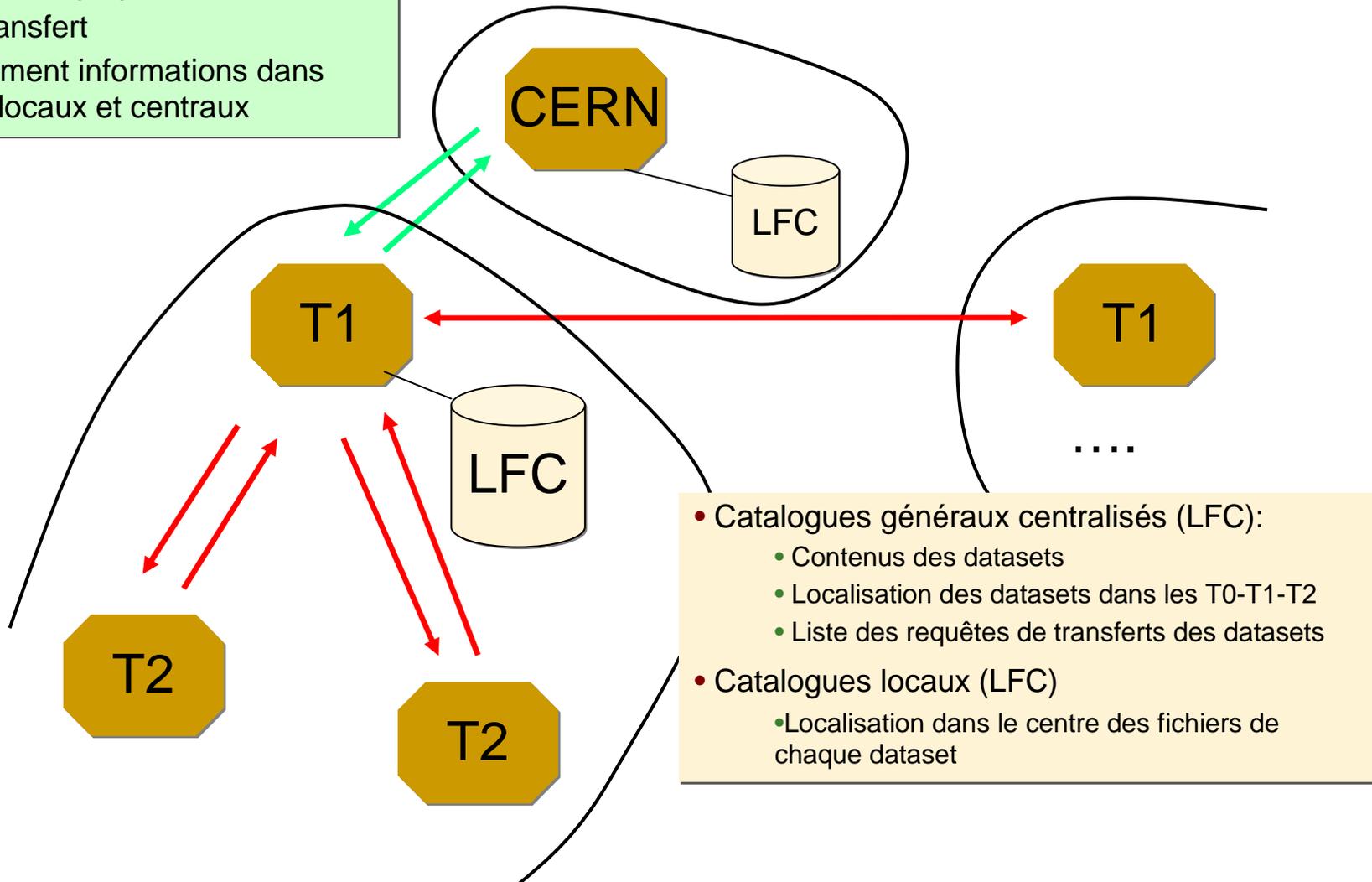
**Si site A a N interlocuteurs n ne peut pas passer son temps à gérer les problèmes des N sites**

**2 solutions :**

- Exclut les sites problématiques (CMS)
  - Minimise le nombre de sites N par site A(ATLAS)
- 
- **Ligne à haut débit entre Tier0-Tier1 et Tier1-Tier1**
  - **Ligne de débit plus faible vers/depuis les Tier2**

# Modèle de transfert des données

- Demande au catalogue central : liste des datasets à répliquer
- Gestion transfert
- Enregistrement informations dans catalogues locaux et centraux



# Conclusion

**La Grille a permis à ATLAS d'entendre les possibilités de calcul**

**ATLAS travaille toujours à la limite des possibilités concrètes de la Grille  
= cycle échec/développement/amélioration**

**ATLAS : Application pour voir les limites de la Grille localement  
Grille: Opportunité de collaboration entre sites distants**