La grille en Physique des particules : l'expérience **ATLAS**

Karim Bernardet
bernardet@cppm.in2p3.fr
4 octobre 2006

Contributions de Stéphane Jezequel (LAPP)

De quoi allons nous parler ?

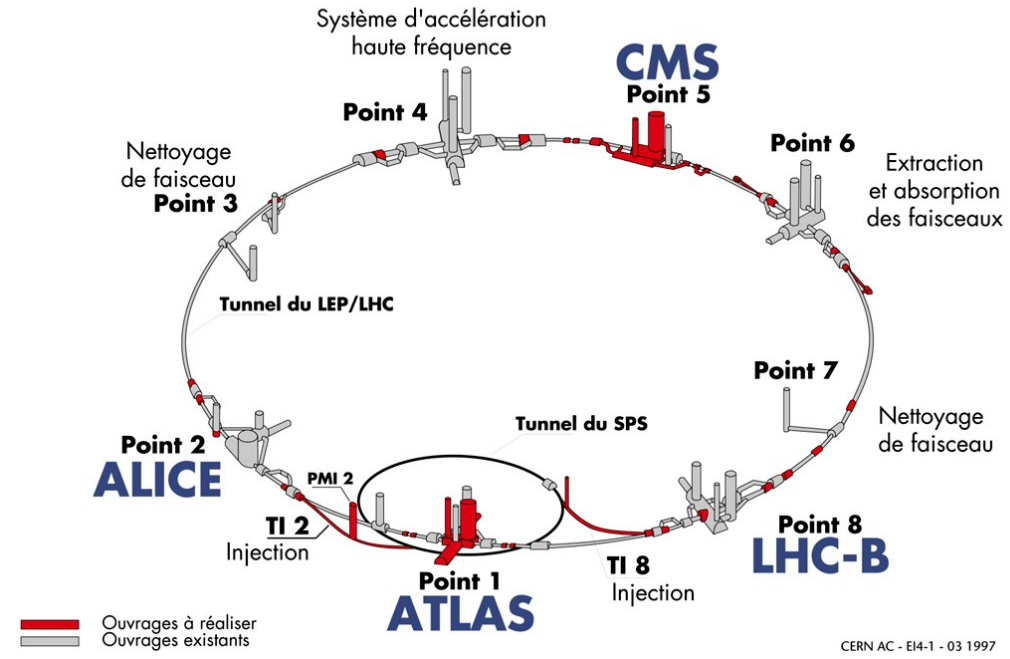
- LHC, LCG, ATLAS
- Utilisations de la grille par ATLAS

LHC, LCG, ATLAS

Large Hadron Collider

Circonférence (km)	26.66
Faisceau	p-p
Énergie dans le centre de masse (TeV)	14

Vue d'ensemble des ouvrages souterrains du LHC



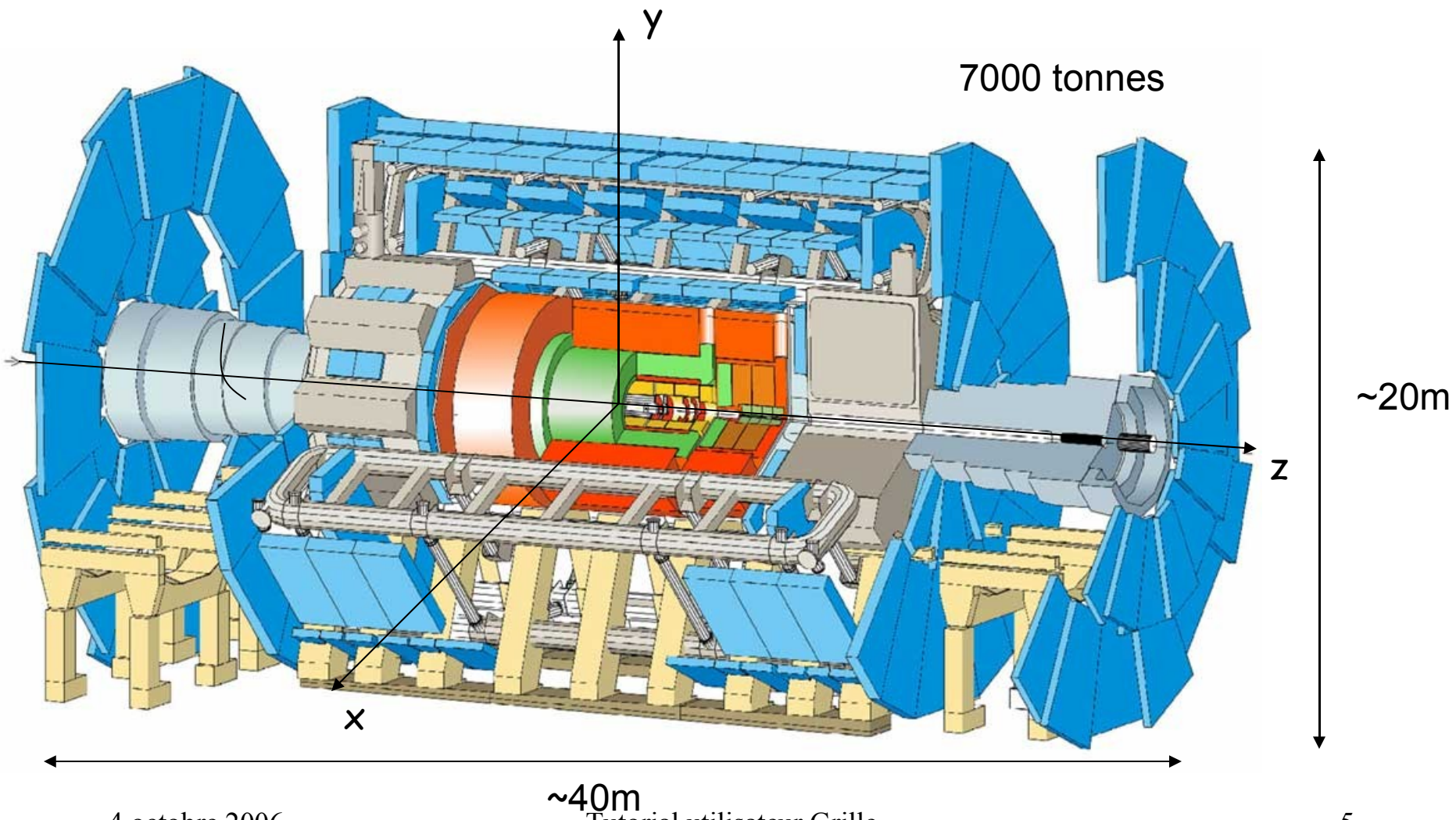
Démarrage en 2007

Démarrage de la Physique en 2008

Le détecteur ATLAS

A la recherche du boson de Higgs

A Toroidal Lhc ApparatuS



Le défi informatique

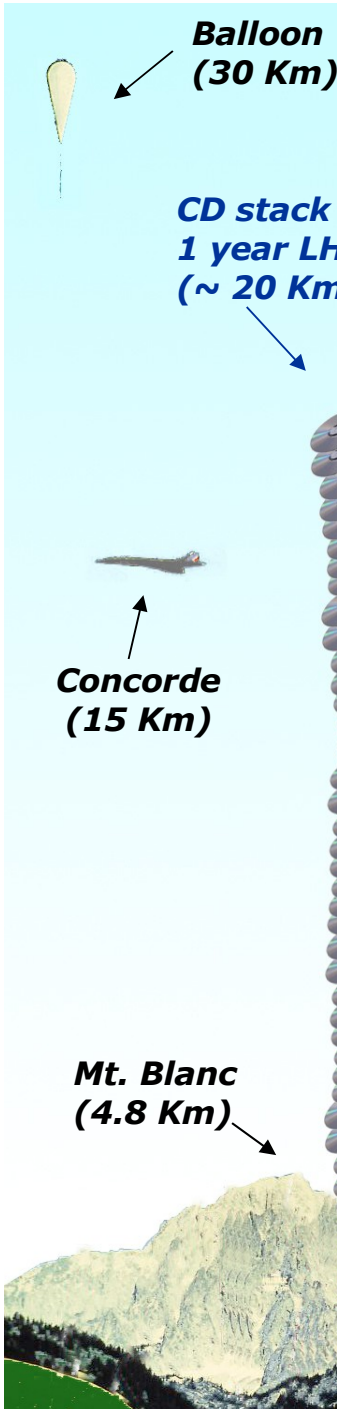
- 40 millions de collisions par seconde pour chacune des 4 expériences
- Après sélection et filtrage, il reste 100-1000 Mo/seconde
- **~15 PetaOctets** de données seront enregistrées chaque année
- Ces données doivent être reconstruites et analysées par des utilisateurs
- Production de données de Monte Carlo

Besoins énormes en calcul et stockage

15 PétaOctets/an

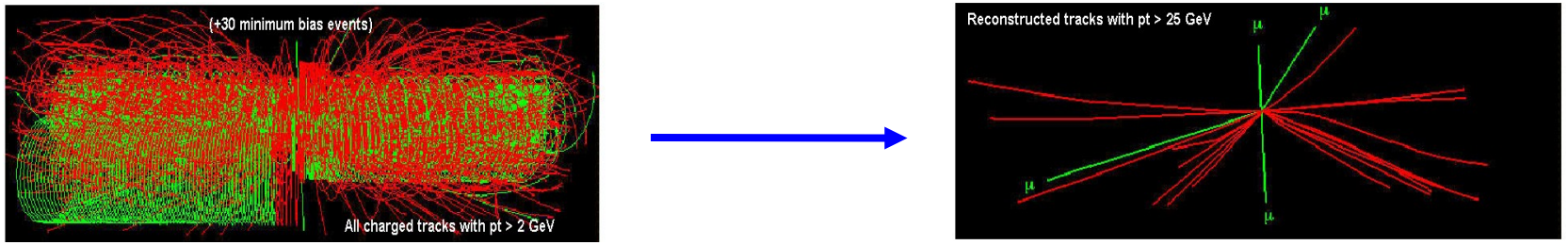
- Besoin de 100 Million SI2K
- Un Pentium IV 3 Ghz ~ 1000 SI2K
- O(100k) cpus

En attendant 2007 : les simulations



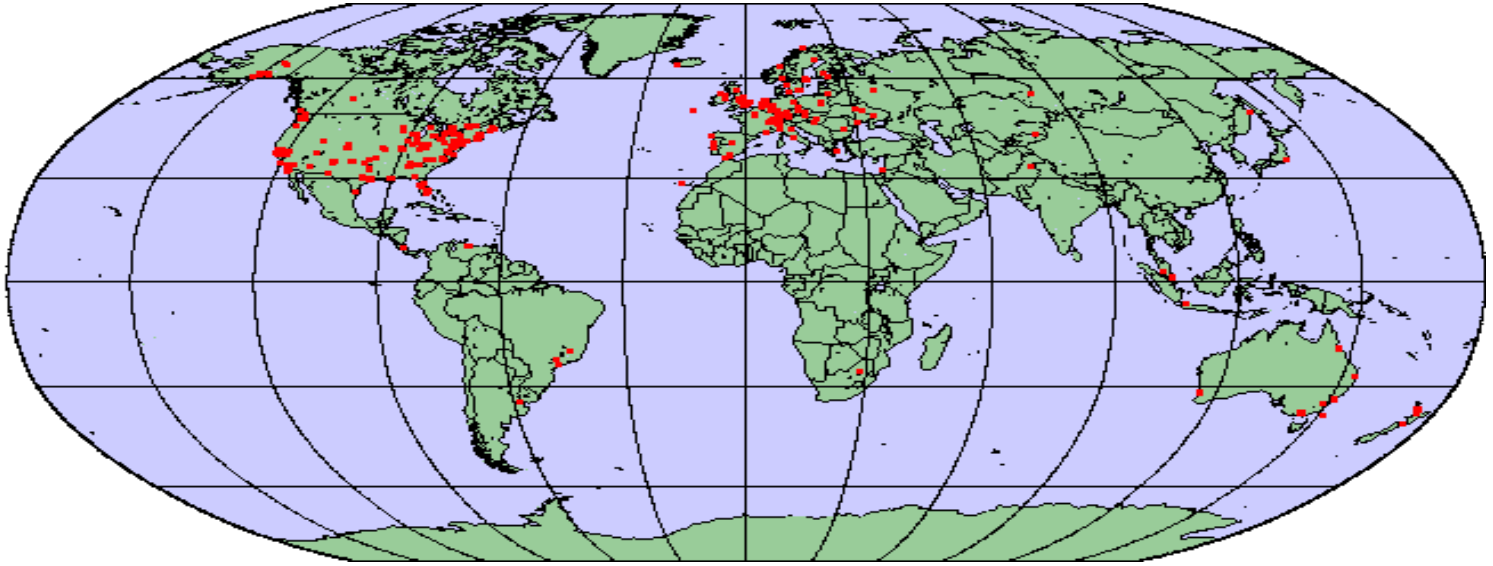
Le flot de travail ou workflow

- « **Event generation** » : Les particules qui émergent des collisions sont « générées » en utilisant des logiciels basés sur les théories de Physique et de phénoménologie,
- « **Simulation** » : Les particules sont « transportées » dans le détecteur en utilisant les lois de la Physique que gouvernent le passage des particules à travers la matière,
- « **Digitization** » : Les interactions avec les éléments sensibles du détecteur sont simulées, reproduisant ce que devraient être les données réelles,
- « **Reconstruction** » : Les données sont ensuite reconstruites,
- « **Physics Analysis** » : Puis analysées par les physiciens.



Une communauté dispersée

- Les collaborateurs du CERN
  - 6000 utilisateurs de 450 instituts
  - Personne n'a la puissance CPU requise
  - Tous ont des ressources de calcul



Solution : connecter toutes ces ressources dans une grille

Une grille

- Une grille en 4 grandes idées :
  - Mutualisation des ressources à l'échelle mondiale
  - Accès sécurisé
  - Utilisation équilibrée des ressources
  - Abolition de la distance

Les projets de grille

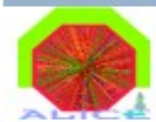
- Grilles de production :
  - EGEE : Projet européen pour la mise en place d'une infrastructure de grille pluridisciplinaire. Son infrastructure et ses développements sont utilisés par :
    - **L**HC **C**omputing **G**rid : projet international dédié aux expériences du LHC du CERN
    - INFN-GRID, ...
  - Nordugrid
  - Grid3, ...
- Grilles « légères » :
  - DIRAC
  - ...

Sites dans LCG : 18 sept 2006

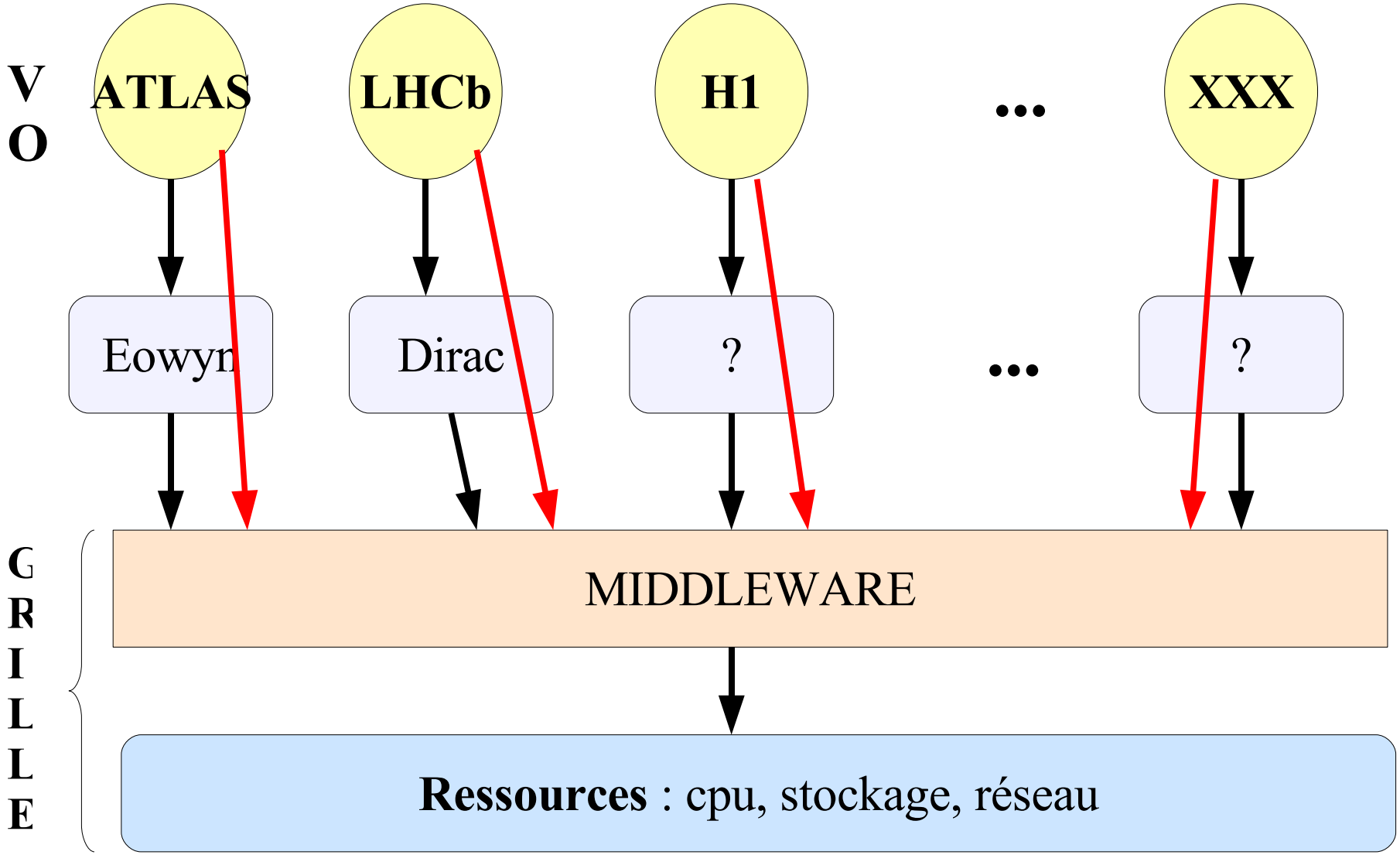
map.

Grid Projects Collaborating in LHC Computing Grid

Open Science Grid



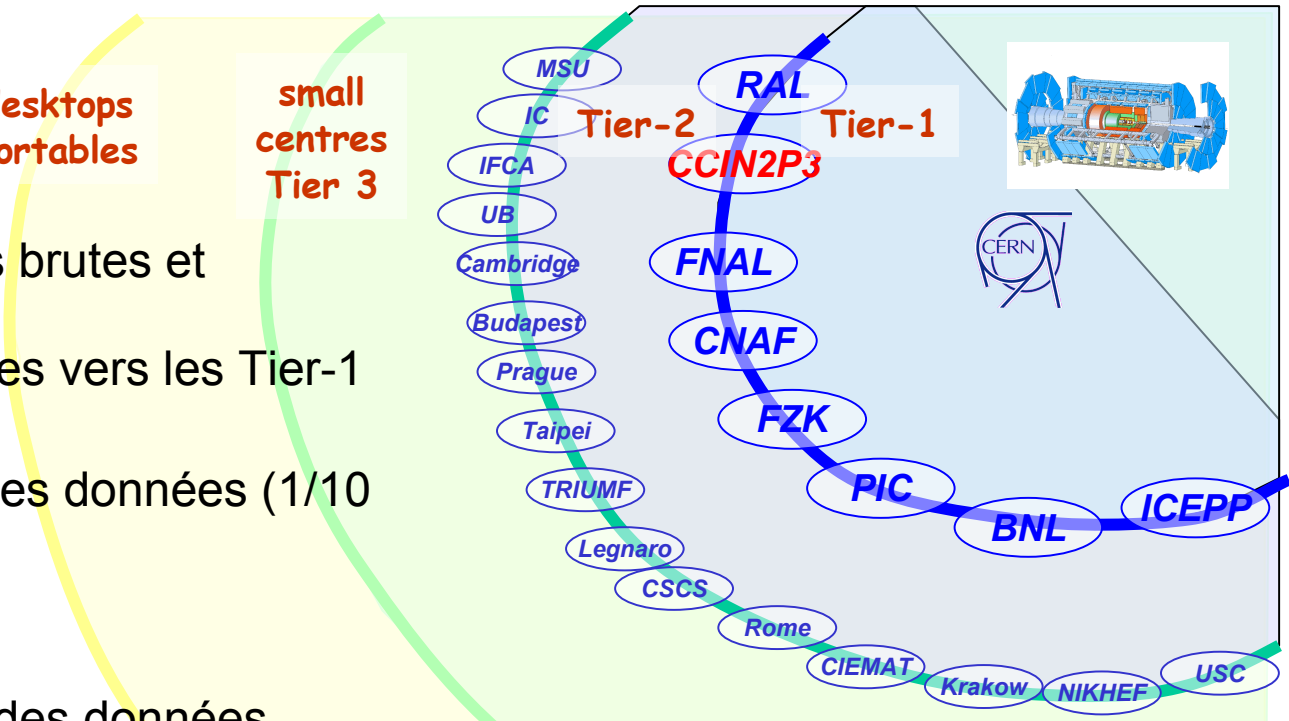
EGEE Operations Information	
Active Sites	177
Available CPU	30653
Available Storage (TB)	41063


V
O
ATLAS
LHCb
H1
...
XXX
Eowyn
Dirac
?
...
?
G
R
I
L
L
E
MIDDLEWARE
Ressources : cpu, stockage, réseau

Modèle de calcul

Pour ATLAS :

- Tier-0 (CERN)
  - Stockage des données brutes et reconstruites
  - Distribution des données vers les Tier-1
- Tier-1 (~ 10)
  - Stockage permanent des données (1/10 brutes, reconstruites)
  - Reprocessing
- Tier-2 (~ 4 T-2/T-1)
  - Stockage d'une partie des données reconstruites
  - Production de Monte Carlo
  - Analyses
- Tier-3
  - Analyses locales et production de MC



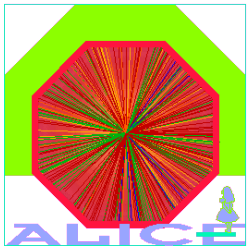
Première version: Publiée mi-2005 (TDR)

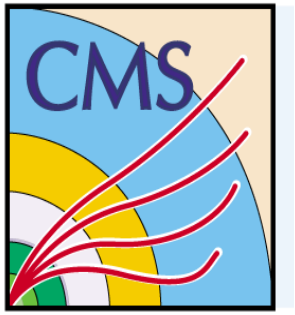
Premiers correctifs: courant 2006

Devra encore s'adapter aux réalités

Data challenge

- Production de données à grande échelle faites par les expériences du LHC
  - Pour tester et valider les modèles informatiques
  - Pour produire des données simulées
  - Pour tester les logiciels des expériences
  - Pour tester le middleware de la grille
  - Pour tester les services fournis par LCG
- Toutes les expériences du LHC ont utilisé LCG pour leur production



Données réelles: Tier-0

- **Après filtrage en ligne,**
  - **évenements bruts (RAW) : 200 Hz**
  - **Taille : 1,6 Mo/evt**

1 Po/an (1 an=4.10^6s) (2008) 3 Po/an (1 an=10^7s) (2010)

- **Première reconstruction des données au CERN:**
  - **ESD (0,5 Mo/evt)**

Réduction des données → AOD (0,1 Mo/evt)

(utilisées par les physiciens pour leur analyse)

Total : 2-5 Po/an

- **Transferts des données dans les T1**
- **Pas de traitement ultérieur au CERN**

mais archivage/backup long terme

From S.J.

Données réelles: Tier-1

T1 recoit les données reconstruite en ligne (~13%)
(~80 Mo/s ATLAS/LYON)

Rôle :

Reconstructions ultérieures (~3/an)
à partir de sa copie de données RAW
Production de nouvelles ESD/AOD

Stockage de toutes les AODs (échange d'AOD entre T1)
Distribution des AODs dans les T2

Redondance du stockage des ESDs
entre paires de T1 (CCIN2P3-BNL)

Etat de la validation

- **Transfert ~OK en phase de test**
- **Reprocessing pas encore testé**

From S.J.

Données réelles: Tier2

T2 recoit une copie des AODs (N*20 Mo/s T1→ tous les T2) (Centre de Calcul Locaux)

Rôle :

Mise à disposition des AODs pour l'analyse des données (= Ferme d'analyse)

Validation:

- **Transfert commence à marcher sur courte période**
- **Activité d'analyse pas possible sur SE DPM avec les outils Grille (accès uniquement par copie sur le disque local)**

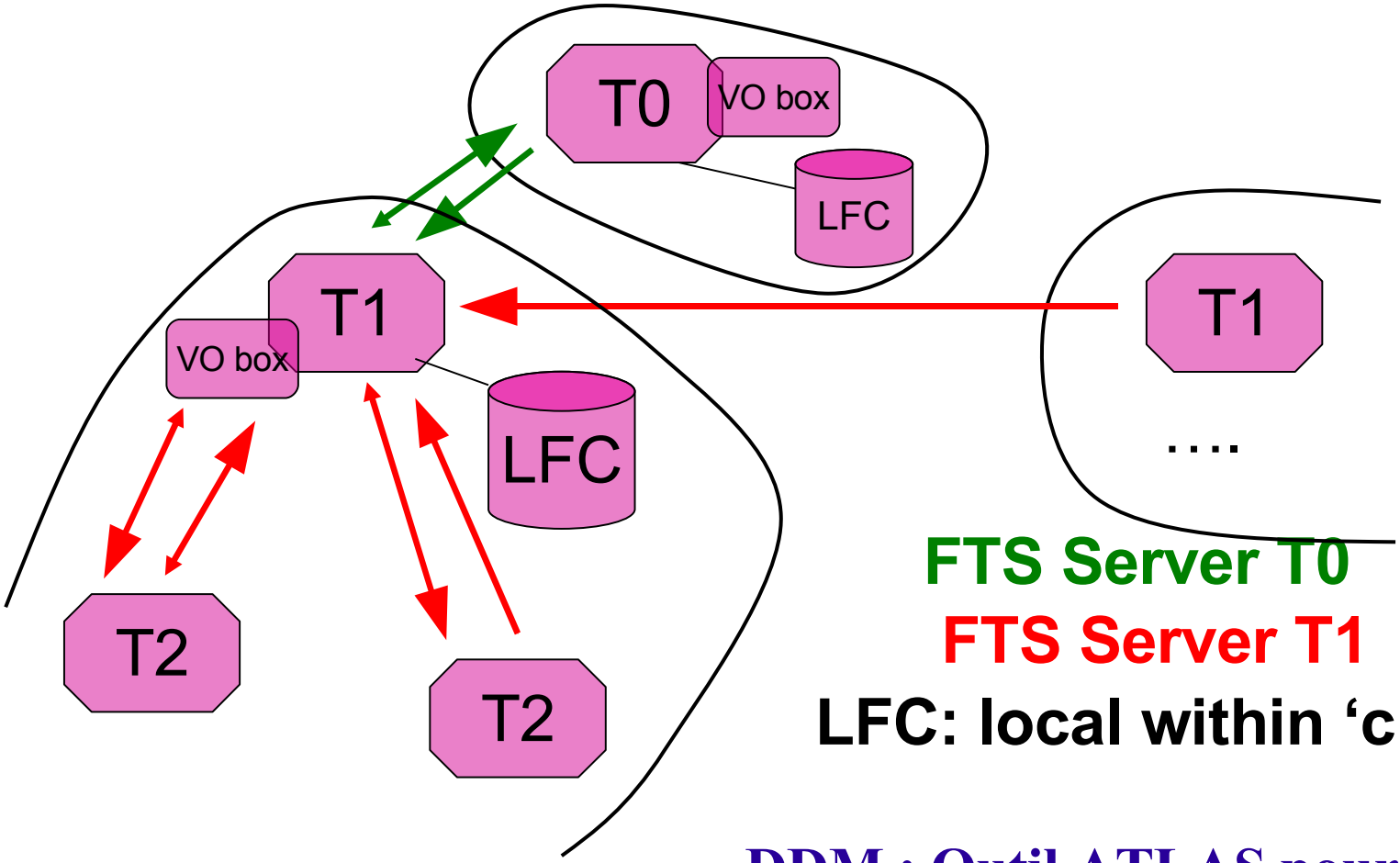
From S.J.

Production de données simulées

- **Seule activité opérationnelle et possible à ce jour**
- **Produites dans les T2/T3 (et T1 actuellement)**
- **Centralisées dans le T1 (lieu de stockage de masse)**
- **Réplication des AODs vers les autres T1 (3 To)**

From S.J.

Réplication/enregistrement data



FTS Server T0

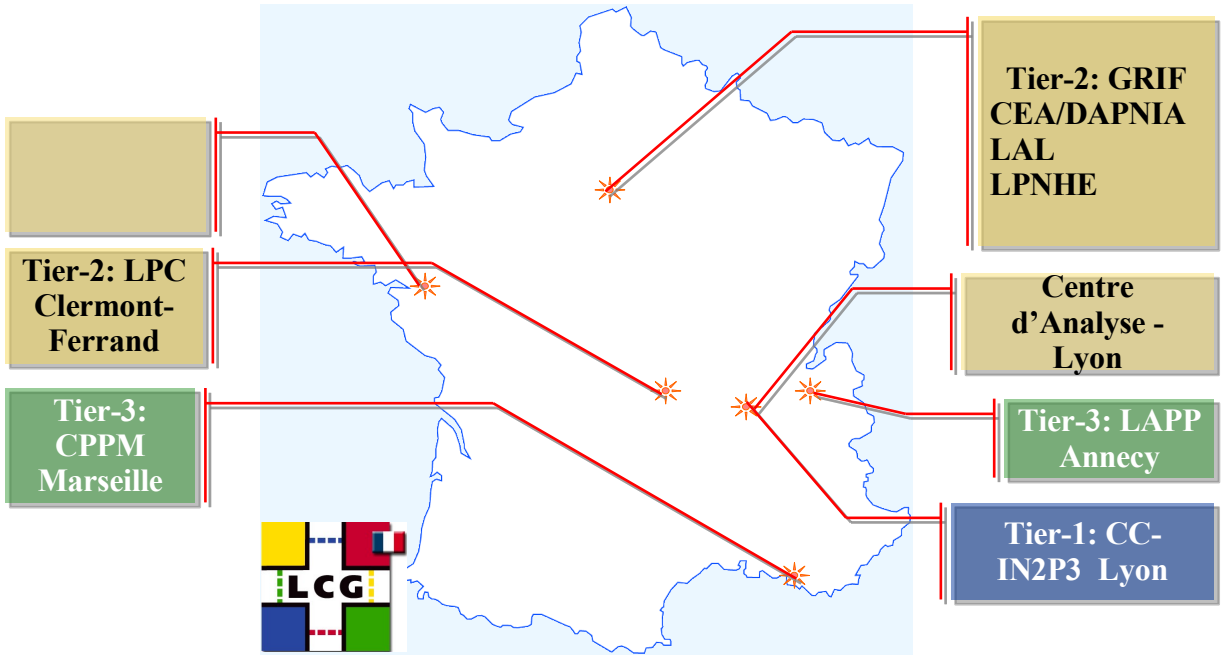
FTS Server T1

LFC: local within 'cloud'

From S.J.

DDM : Outil ATLAS pour organiser les transferts

Nuage FR :T1/T2/T3



- **BEIJING**
- **TOKYO**

From S.J.

Utilisation de la grille par ATLAS

Distribution du software ATLAS

- ~100 sites pour ATLAS
- Installation automatisée/centralisée
  - Marche dans 50% des cas (sites stables)
  - Origines d'erreurs
    - Sites en évolution (changement de SE, d'OS,...)
    - Zones disques affectées au stockage de softs pleine

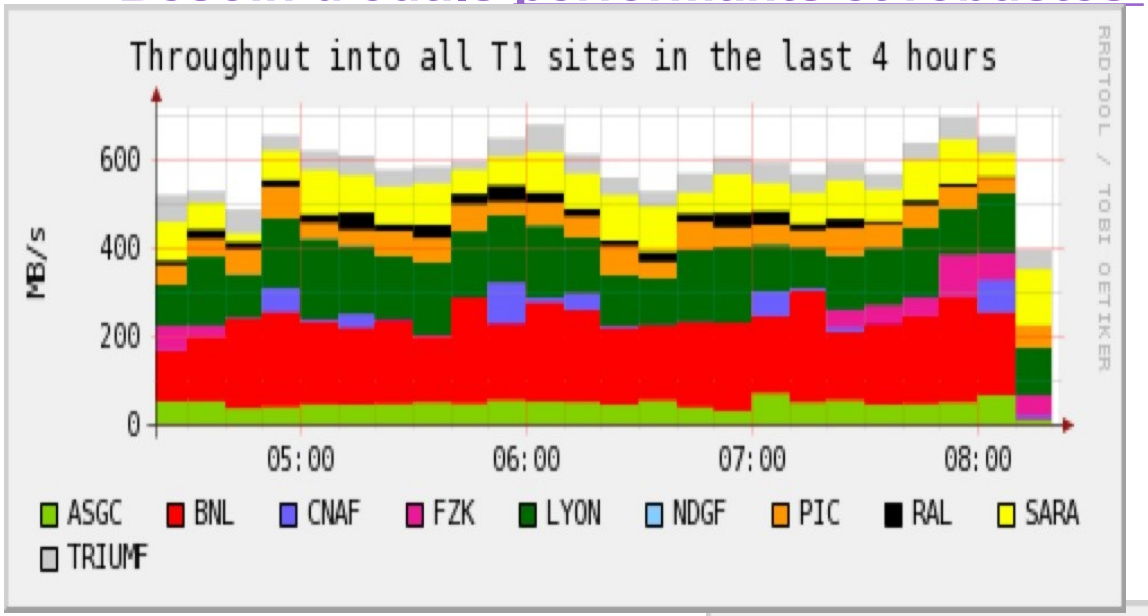
Besoin de contact direct

entre le responsable du site et ATLAS

From S.J.

Réplication/Publication des données

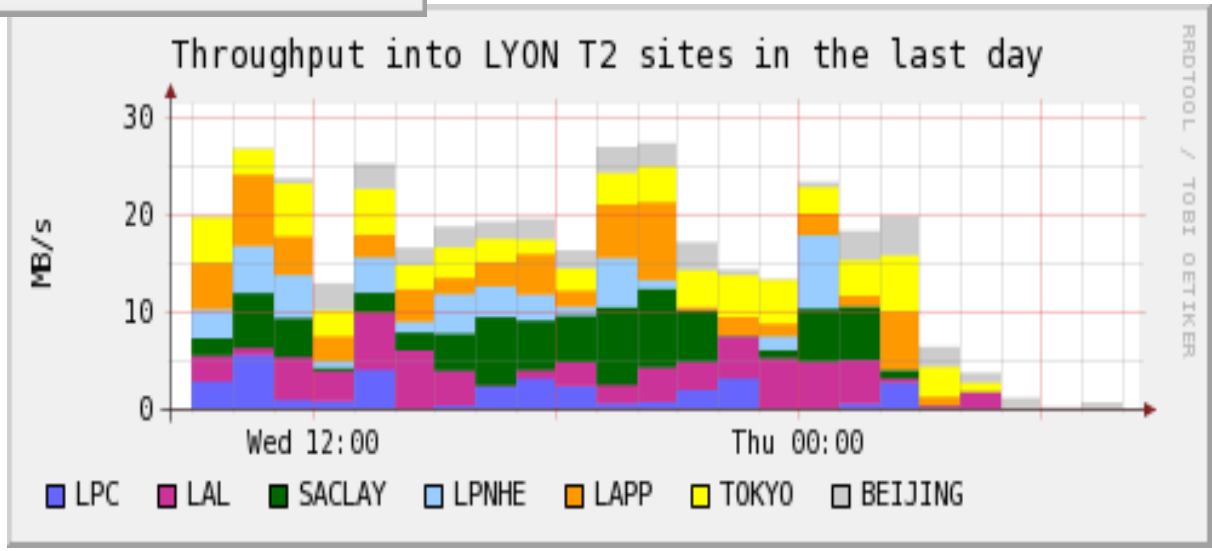
- **Passage obligé**
- **Besoin d'outils performants et robustes**



OK dans cadre restreint

FR : Premier 'nuage' opérationnel de la grille EGEE

From S.J.



4 octobre 2006

Réplication/Publication des données

- **Dans un environnement non dédié :**
  - **Surcharge du catalogue LFC**
  - **10-20% d'erreur sur les transferts FTS (transfert multi-site simultané)**
  - **Non optimisation de DDM**

Travail en cours
mais
déjà en phase de production dans ATLAS

From S.J.

Un job LCG ATLAS

- Le soft ATLAS est le framework ATHENA

- Ecrire le script qui sera exécuté sur un WN
  - Récupération de fichiers d'input (éventuellement)
  - Récupération d'une archive contenant ce qui manque (éventuellement)
  - Setup d'Athena
  - Décompression de l'archive (éventuellement)
  - Compilation de packages (éventuellement)
  - Exécution d'Athena
  - Copie des fichiers produits vers un SE et enregistrement dans un catalogue LCG (voire dans DDM)

Simulation

From S.J.

Partie de la chaine d'analyse tirant le meilleur profit de la Grille avec un minimum de développement

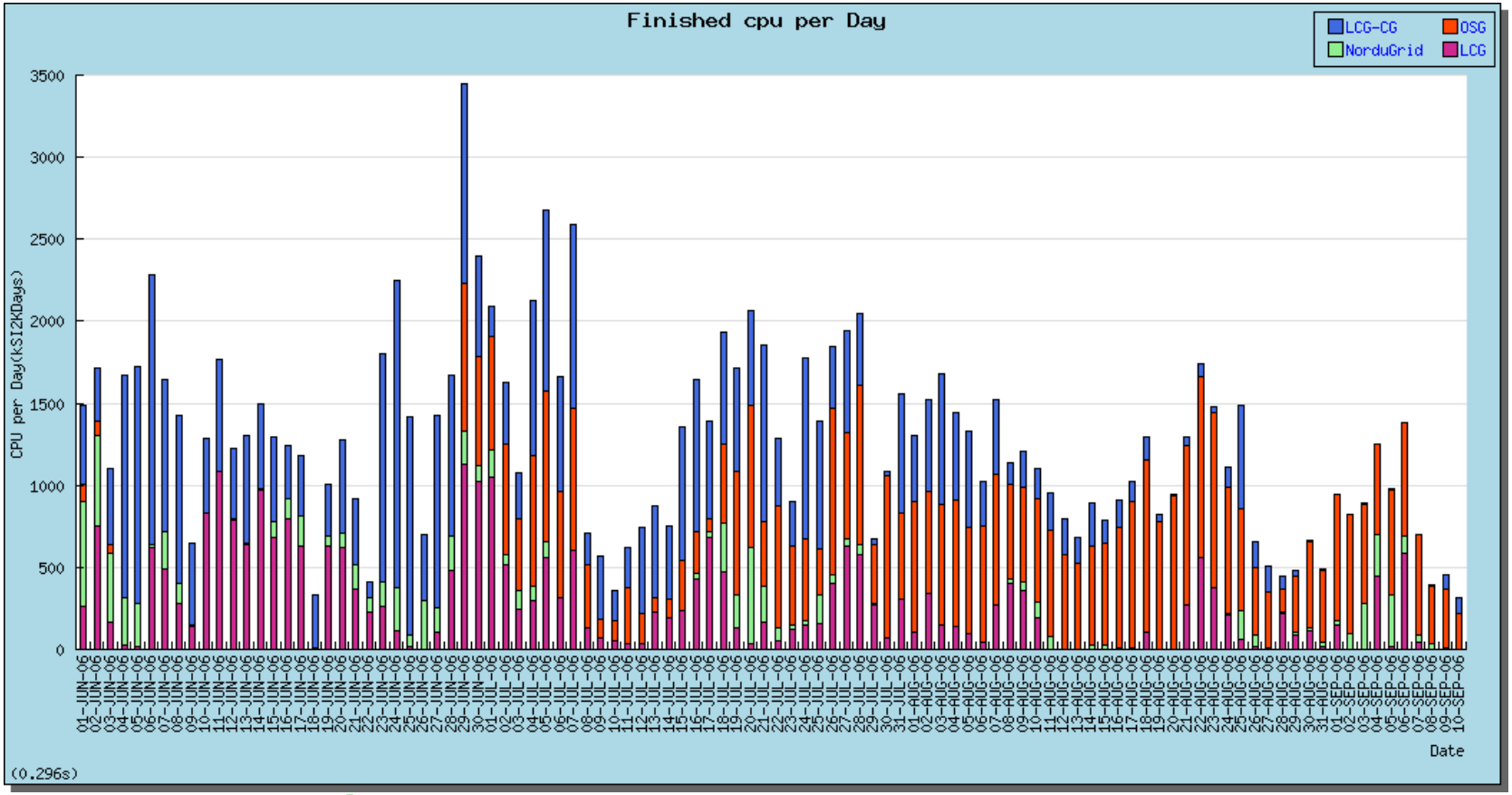
- **Besoin d'avoir un CPU disponible**
- **Un seul fichier en entrée**
- **Peu d'accès en lecture/écriture**
- **Temps de soumission << Temps d'éxecution**
- **Ne nécessite que qq experts pour lancer les jobs**

Point critique :
Accès aux données au cours de la simulation ou après

Vitesse de production

From S.J.

- Encore fortement instable
- En deçà des attentes d'ATLAS (facteur 3)



Prochaines évolutions:

- Etiqueter un job de production centralisée par un rôle VOMS
- Donner la priorité dans les queues des sites à ces jobs

Analyse des données

En collaboration avec C. Bourdarios

- **Travail de tous les physiciens**

  (ca doit marcher comme sur la doc)

Commandes grilles de bas niveau inconnues

et/ou limitations grille/infrastructure

⇨ **Bonne fiabilité et stabilité avant diffusion**

From S.J.

Analyse des données:Standard

- **En 2006:**
  - **Travail sur des données simulées avec**
    l'équivalent de qq jours de données
  - **Utilisent les réplications des données locales**

(commencent à se former aux outils de réplications)

  - **Tourne les jobs analyses en interactif**
    ou avec qq batchs BQS/LSF

From S.J.

Analyse des données:Prospective

Analyse avec infrastructure Grille :
Utile lorsque beaucoup de données seront disponibles

En phase d'évaluation et debugging des outils

Préalable: nécessite de rassembler sur un/plusieurs sites tous les fichiers d'un même job

2 interfaces utilisateurs sur le marché:

- **Panda (made in BNL)**
  Ne tournait qu'à BNL jusqu'à récemment mais avec succès
  Volonté de le porter sur LCG (1 job a déjà tourné à Lyon)

- **Ganga (made in Europe: LHCb/ATLAS)**
  Commence juste à être opérationnel dans ATLAS
  CCIN2P3 est un des sites de validation

From S.J.

Exemple : générer des évènements

1. {ccali13}>grid-proxy-init
   Your identity: /C=FR/O=CNRS/OU=CPPM/CN=Karim Bernardet/Email=bernardet@cppm.in2p3.fr
   Enter GRID pass phrase for this identity:
   Creating proxy ..... Done
   Your proxy is valid until: Mon Nov 1 06:36:36 2004
2. {ccali13}> edg-job-submit --vo atlas -o jobId job.jdl

job.jdl

Executable=«generation.sh";

Arguments=«NTUPLEFN=test7 PHYSSHORT=test7 EVGENFN=test7 DATASET=001234 TOTAL=10
00 OUTPUTDIR=Production/evgen/test NO=1";

InputSandbox="generation.sh";

Requirements=(Member("VO-atlas-release
11.0.41",other.GlueHostApplicationSoftwareRunTimeEnvironment)) ;

RetryCount = 0;

OutputSandbox="job.log";

StdOutput="job.log";

StdError="job.log";

Que fait generation.sh ?

En général, les fichiers produits sont des fichiers POOL au format ROOT. A ces fichiers sont associés un fichier XML qui contient notamment l'ID du fichier POOL.

- Extraction des paramètres à partir de la liste des arguments et initialisation de certaines variables d'environnement

```
export LFC_CATALOG_TYPE=lfc
export LFC_HOST=lfc-atlas-test.cern.ch
```

- Setup d'Athena
- Récupération d'une archive contenant des packages manquants depuis un SE

```
lcg-cp --vo atlas lfn:/grid/atlas/users/bernardet/tools/packageGeneration.tar.gz
    file:`pwd`/packageGeneration.tar.gz
```

- Décompression de l'archive packageGeneration.tar.gz
- Compilation des packages
- Exécution d'Athena
- Copie et enregistrement des fichiers produits vers un SE (ccsrm au CCIN2P3)

```
GUID=`./guid_gateway.sh -p PoolFileCatalog.xml ${OUTPUTFILE}`
lcg-cr --vo atlas  -t 7200 -d
    srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/tape/temp/${OUTPUTDIR}/${OUTPUTFIL
    E}  -g guid:${GUID} -l /grid/atlas/users/bernardet/${OUTPUTDIR}/${OUTFILE}
    file://$PWD/${OUTPUTFILE}
```

{ccali13}…> edg-job-submit --vo atlas -o jobId job.jdl

Selected Virtual Organisation name (from --vo option): atlas

Connecting to host lxb0728.cern.ch, port 7772

Logging to host lxn1186.cern.ch, port 9002

==========edg-job-submit Success ==========

The job has been successfully submitted to the Network Server.

Use edg-job-status command to check job current status. Your job identifier (edg_jobId) is:

- https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

The edg_jobId has been saved in the following file:

/sps/atlas/k/kbernard/atlprod-client-LCG2/jobs/TESTSLCG/jobEVGEN/jobId

==========

{ccali13}…>edg-job-status -i jobId

****

BOOKKEEPING INFORMATION:

Status info for the Job : https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

Current Status: **Ready**

Status Reason: unavailable

Destination: dgce0.icepp.jp:2119/jobmanager-lcgpbs-infinite

reached on: Wed Nov 3 14:57:10 2004

****

{ccali13}…>edg-job-status -i jobId

****

BOOKKEEPING INFORMATION:

Status info for the Job :
https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

Current Status: **Scheduled**

Status Reason: Job successfully submitted to Globus

Destination: dgce0.icepp.jp:2119/jobmanager-lcgpbs-infinite

reached on: Wed Nov 3 14:57:41 2004

****

{ccali13}…>edg-job-status -i jobId

****

BOOKKEEPING INFORMATION:

Status info for the Job :
https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

Current Status: **Running**

Status Reason: Job successfully submitted to Globus

Destination: dgce0.icepp.jp:2119/jobmanager-lcgpbs-infinite

reached on: Wed Nov 3 15:02:00 2004

****

{ccali13}…>edg-job-status -i jobId

****

BOOKKEEPING INFORMATION:

Status info for the Job : https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

Current Status: **Done (Success)**

Exit code: 0

Status Reason: Job terminated successfully

Destination: dgce0.icepp.jp:2119/jobmanager-lcgpbs-infinite

reached on: Wed Nov 3 15:15:26 2004

****

{ccali13}…>edg-job-get-output -dir . -i jobId

Retrieving files from host: lxb0728.cern.ch (for https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA)

****

JOB GET OUTPUT OUTCOME

Output sandbox files for the job:

- https://lxb0728.cern.ch:9000/sRjOworKtRxX5ngv0himDA

have been successfully retrieved and stored in the directory:

/sps/atlas/k/kbernard/atlprod-client-LCG2/jobs/TESTSLCG/jobEVGEN/kbernard_sRjOworKtRxX5ngv0himDA

Conclusions

- **La validation concrète du modèle de calcul a commencé**
- **Mise en place en cours des outils Grille indispensables**
- **Point de passage obligé : outil de réplication performant et robuste (ATLAS/Grille)**
- **Simulation sur la grille en cours de stabilisation**
- **Démarrage de la validation de l'analyse sur la grille (job Grille/ accès aux données sur un SE)**