
Modèles de calcul hiérarchique, distribué = Projets de grille ~ DataGRID

Cargèse, 24 juillet 2001

Fabio HERNANDEZ + Nadia (+DL)
fabio@in2p3.fr

Vue d'ensemble

- Le projet DataGRID
- Etat d'avancement
- Le rôle de l'IN2P3
- Etat d'avancement au Centre de Calcul
- Etapes suivantes

Ceci n'est pas une revue exhaustive du projet DataGRID.

Le projet DataGRID

- Research & Technological Development for an International Data Grid
- Objectif
 - *"To enable next generation scientific exploration which requires intensive computation and analysis of shared large-scale databases, from hundreds of terabytes to petabytes, across widely distributed scientific communities."*
 - autrement dit, construire un prototype du modèle de calcul pour la période LHC (2007 = 200000 "PC-2000", 10 PB disque, 30 PB/an K7)
- Partiellement financé par l'Union Européenne (9,8 M€/trois ans)
- Projet multidisciplinaire
 - Physique des hautes énergies (rôle leader)
 - Biologie
 - Observation de la terre

Le projet DataGRID (suite)

- o Partenaires principaux



- o + Partenaires associés

- En France

- Le CEA

- La Compagnie des Signaux (communications & systems)

- D'autres instituts de recherche/sociétés en Italie, Finlande, Espagne, Allemagne, Suède, Hongrie, Grande Bretagne, Hollande

- o La communauté de physique de hautes énergies est chargée de la coordination du projet

- CERN au niveau européen

- IN2P3 au niveau français

pluridisciplinaire+entreprises
pour assurer la dissémination.

Le projet DataGRID (suite)

- Work packages

- *Middleware*

- WP1: Workload management (Francesco Prelz - INFN)
- WP2: Data management (Ben Segal - CERN)
- WP3: Monitoring services (Robin Middleton – RAL)
- WP4: Fabric management (Olof Barring – CERN)
- WP5: Mass storage management (John Gordon – RAL)

- *Infrastructure*

- WP6: Integration testbed (François Etienne – CNRS/IN2P3) ■ ■
- WP7: Network services (Pascale Primet – CNRS/ENS Lyon) ■ ■

- *Applications*

- WP8: High-Energy physics applications (Federico Carminati – ALICE-CERN)
- WP9: Earth observation science application (Luigi Fusco – ESA)
- WP10: Biology science application (Christian Michau – CNRS) ■ ■

- *Management*

- WP11: Dissemination (Maurizio Lancia – INFN)
- WP12: Project Management (Fabrizio Gagliardi – CERN)

Le projet DataGRID (suite)

- Site web du projet
 - <http://www.eu-datagrid.org>

Etat d'avancement

- Travail prospectif débuté au premier trimestre 2000 (lancement par le HEPCCC en novembre 1999)
- Travaux officiellement commencés en **janvier 2001** (date du contrat européen de 3 ans)
- Où en sommes-nous ?
 - Première itération des tests pendant le premier semestre 2001
 - Phase initiale de démarrage plutôt difficile
 - Actuellement, préparation de la deuxième itération en intégrant le travail des différents work packages d'infrastructure
 - Début de l'intégration prévue pour **octobre 2001**
 - Tests pendant le quatrième trimestre 2001

Rôle de l'IN2P3

- Coordination du work package 6 *Integration and Testbed* (F. Etienne)
- Rôle prépondérant pour les applications 10 biologie (V. Breton) et 8 physique (Y. Schutz)
- Participation du CC au WP7 réseaux
- Plusieurs laboratoires impliqués dans la phase initiale du testbed
 - Marseille
 - Clermont
 - Grenoble
 - LAL Orsay (observateur)
- Étroite collaboration avec CEA Saclay

Etat d'avancement

CNRS

- Sites de testbed :CPPM, LPC, CC-I N2P3, SACLAY, LAL.
- Interfaçage de BQS avec Globus au CC I N2P3
- Tests
 - En cours : Tests de production avec Alice (Y.Schutz)
Démarrage de tests avec le CERN (JJ.Blaising)
Début des tests LHCb (E. van Herwijnen)
 - En vue : Démarrage de tests avec LPC Clermont (Y.Legré)
Début de tests CMS (C. Charlot)
 - Autres : Tests autour de l'application BLAST= Décodage de génomes
(Clermont+Marseille/Lyon)

Bilan

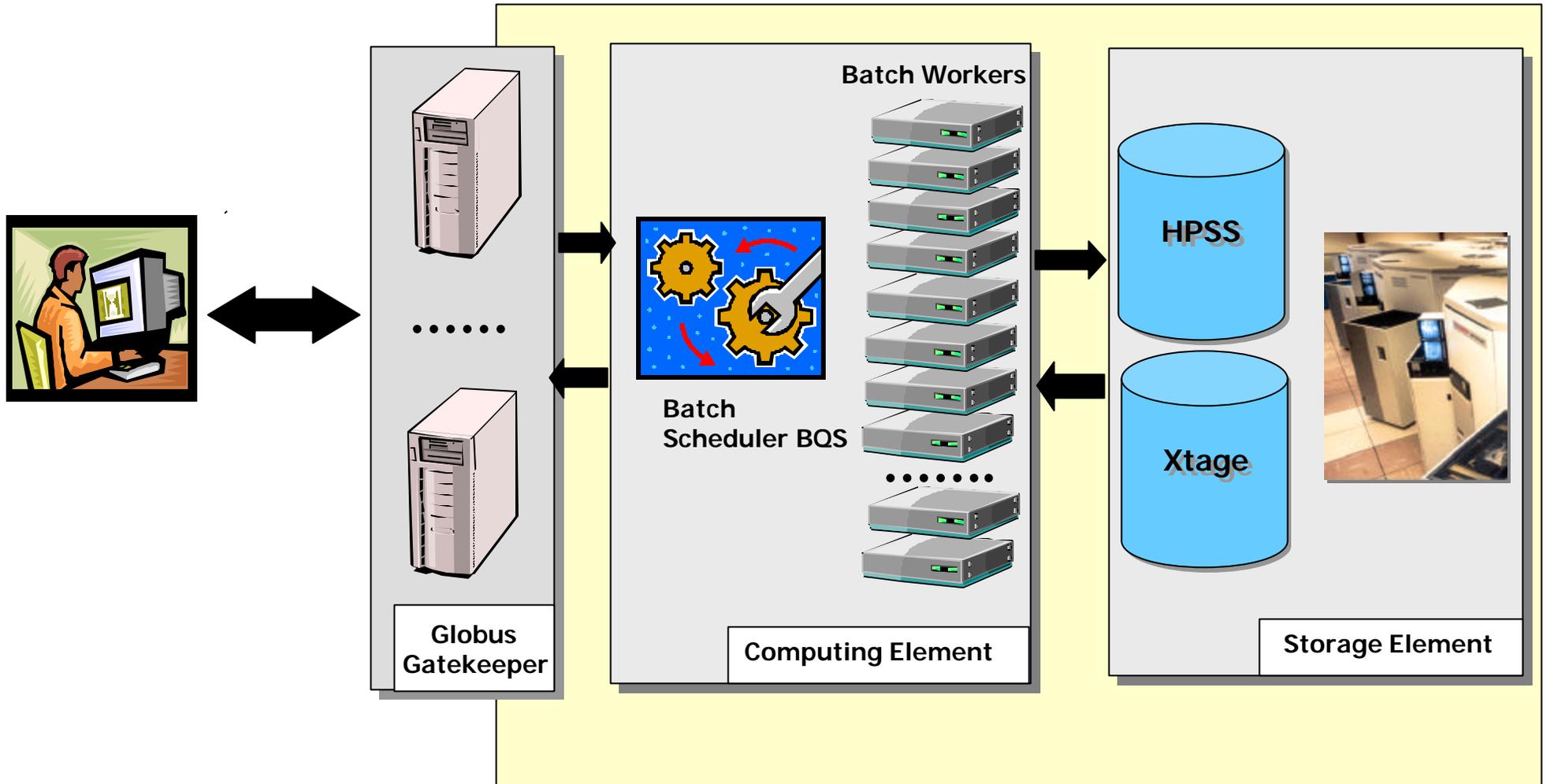
Rédaction du document définissant la stratégie d'intégration (dates et organisation des différentes étapes)

Contacts avec les autres WPS afin de tenir compte des besoins, des interdépendances et exigences

Globus2 alpha est en cours d'évaluation

Définition d'une Infrastructure de site web, d'un outil de bug tracking, d'un répertoire CVS central / miroir

Etat d'avancement au CCIN2P3



Resources Disponibles

CC-in2p3, seul site à ouvrir toutes ses ressources

- *Computing element*

- *BQS - Batch Queueing System* : ordonnanceur de batch interfacé avec Globus
- Travaux soumis à BQS via Globus tournent dans l'environnement habituel de l'utilisateur (AFS, ...) et ont accès à toute l'infrastructure de stockage
- Actuellement ~200 machines (~400 CPUs) sous Linux RedHat 6.1 accessibles aux utilisateurs de la grille à travers Globus

- *Storage element*

- 60+ machines sous AIX, Solaris, Tru64
- Capacité disque de 35 TB
- HPSS: 50TB sur cartouche / 1 TB disque cache

- Distribution du logiciel produit dans le cadre du projet sous la responsabilité du centre de calcul (repository avec l'outil CVS)

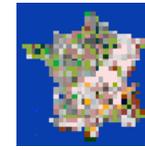
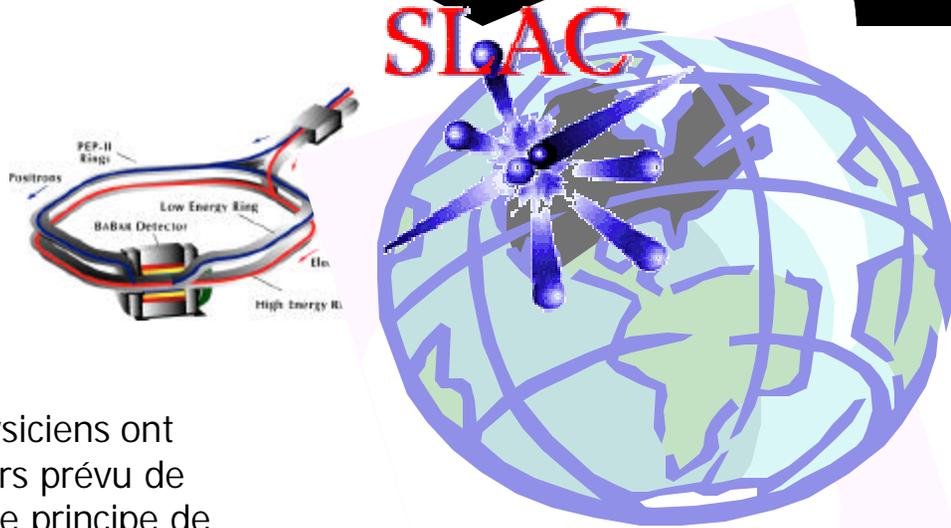
Etapes suivantes

- Deuxième campagne de test de production pour Alice pendant cet été
- LHCb a commencé à mettre en place son environnement pour des test initiaux
- Tests des applications de biologie pendant cet été impliquant plusieurs sites français
 - LPC Clermont (coordinateur)
 - Marseille
 - Lyon
- Mise en place de l'infrastructure de distribution du logiciel pour tout le projet (en cours)
- Participation au développement d'outils divers pour la gestion de la grille.

Tier A, collaboration avec les grilles US



BABAR



les physiciens ont d'ailleurs prévu de tester le principe de la Grille en commençant par les centres de calcul de Lyon et de Stanford (Californie).
Un premier pas vers l'ubiquité de la Grille.

Pierre Le Hir
Le Monde
12 avril 2001

Deux centres qui testent de fonctionner comme un seul.
- Données non dupliquées
- internationalisation

