



HAL
open science

Collaboration des grilles de production grenobloises CIMENT et LPSC - Développement, ouverture à de nouvelles communautés d'utilisateurs et impacts scientifiques

C. Biscarat, Bruno Bzeznik

► To cite this version:

C. Biscarat, Bruno Bzeznik. Collaboration des grilles de production grenobloises CIMENT et LPSC - Développement, ouverture à de nouvelles communautés d'utilisateurs et impacts scientifiques. Rencontres Scientifiques France Grilles 2011, Sep 2011, Lyon, France. hal-00653019

HAL Id: hal-00653019

<https://hal.science/hal-00653019v1>

Submitted on 16 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collaboration des grilles de production grenobloises CIMENT et LPSC Développement, ouverture à de nouvelles communautés d'utilisateurs et impacts scientifiques

Auteurs : Grilles CIMENT et LPSC

Interlocuteurs principaux :

Catherine Biscarat, LPSC (1)

Bruno Bzeznik, CIMENT (2)

(1) biscarat@lpsc.in2p3.fr LPSC, CNRS-IN2P3, UJF, INP

(2) Bruno.Bzeznik@imag.fr CIMENT, UJF

Overview:

In Grenoble, two well established production grid infrastructures started a new collaboration in 2010 with the support of France-Grilles. The partners are the LPSC, a grid site contributing to the worldwide LHC computing grid, and CIMENT, a mesocentre hosting a light-weighted middleware CIGRI [1] aimed at the exploitation of resources in a best effort way. These two grid sites have different and complementary technical implementations and address the needs of different scientific user communities. The aim of the collaboration is to develop a link between these two infrastructures allowing the sharing of resources among different user communities. Indeed, our new collaboration opens new computing synergies in terms of storage and CPU usage to existing scientific users and this should naturally attract new communities.

This contribution will focus on the technical developments made to strengthen the collaboration (particularly iRODS [2] storage deployment). It cites the first scientific results obtained thanks to this work, and it highlights the first successful porting of an application from a high-profile High Energy Physics analysis at LPSC to the CIMENT infrastructure.

Enjeux scientifiques, besoin de la grille :

Une large communauté d'utilisateurs de thématiques différentes utilisent les ressources de ces deux grilles de production. Par exemple, les thématiques qui ont bénéficiées de la mise en place de la collaboration CIMENT-LPSC, sont :

- Modélisation du climat au LGGE (Hubert Gallée)
- Simulation de la réponse des espèces et de la biodiversité aux changements globaux.
- Hydrologie et changement climatique
- Astronomie : Projet CONSERT
- Simulation de nouveaux concepts de réacteurs nucléaires : projet MURE
- Physique des particules : expériences auprès du LHC et du Tevatron
- Neurologie (TIDRA)
- Biomédical

La collaboration CIMENT-LPSC est naissante et ces premiers résultats ont utilisé le stockage iRODS grâce au financement obtenu par France-Grille. Toutefois, l'étape suivante de cette collaboration est de porter les utilisateurs d'une communauté, c'est-à-dire leurs applications, d'une grille à l'autre. Nos premiers résultats sont encourageants. De là nous gagnerons en expérience sur les différentes grilles et nous pourrons mieux adapter les spécificités de chacune dans le but de construire une nouvelle synergie.

Dans le cadre de la collaboration D0 [3], ce travail a déjà débuté. D0 est une expérience internationale de Physique des Hautes Energies auprès du collisionneur Tevatron, au

Laboratoire Fermi aux Etats-Unis (Fermilab), dans laquelle l'IN2P3/CNRS est très engagé. Un des thèmes principaux de D0 est l'une des questions les plus fondamentales de notre époque : la compréhension de l'origine de la masse et l'existence, ou non, du boson de Higgs, la particule manquante du Modèle Standard de la Physique des Particules. La recherche directe du boson de Higgs est actuellement engagée au LHC (le grand collisionneur de hadrons au CERN à Genève). En même temps, la recherche indirecte du boson de Higgs, par la mesure précise de la masse du bosons W, est actuellement poursuivie au Tevatron. Ces deux approches sont complémentaires et nous aideront chacune à mieux appréhender les mystères encore sans réponse de notre univers.

La méthode indirecte par la mesure précise de la masse du boson W menée à D0 est une initiative d'une équipe du LPSC. Cette méthode repose sur la relation mathématique qui existe entre la masse du boson W, celle du quark top (très précisément mesurée au Tevatron) et la masse du boson de Higgs. Cette dernière est donc contrainte (jusqu'à une possible exclusion) par les mesures précises du boson W et, en moindre importance, du quark top.

Le résultat actuel (qui est la combinaison de plusieurs expériences dont D0) conduit à une incertitude de 23 MeV [4] sur la masse du boson W. Le résultat de D0 dans cette combinaison est le plus précis [5] et il est élaboré avec 10% des données de l'expérience. En effet, il s'agit d'un travail commencé en 2003 et qui a porté ses premiers fruits en 2009. L'analyse de l'échantillon complet des données est en cours. On attend un résultat intermédiaire en 2011 et un résultat final en 2013. Une erreur de 15 MeV pour D0 seule est attendue. Cette valeur suffirait à exclure définitivement le boson de Higgs du modèle Standard [6].

L'équipe qui travaille sur cette analyse a élaboré un programme de simulation rapide du détecteur D0, PMCS. Des simulations dans diverses configurations du détecteur et de la masse du boson W doivent être faites pour parvenir à un résultat précis. Toutefois, le laboratoire Fermi consacre ses ressources informatiques à la reconstruction des données et peu à l'analyse. Trouver des ressources indépendantes est donc important pour mener cette analyse à terme et en temps utile.

Les ressources dont dispose CIMENT ne sont pas employées 100% du temps. D0, dont les tâches de calcul peuvent être soumises en nombreux ensembles de tâches courtes, représente un client idéal pour une utilisation opportuniste des ressources non utilisées de CIMENT. Pour cela, nous avons réalisé le portage de l'application PMCS de D0 sur CIMENT. Nous avons validé les résultats obtenus sur CIMENT à l'aide d'une comparaison détaillée avec une référence provenant de la ferme de calcul centralisée de D0 à Fermilab. Ces premiers résultats sont positifs et ouvrent des possibilités pour augmenter les ressources en CPU pour cette analyse. Nous continuons à accumuler de l'expérience opérationnelle sur CIMENT, et nos premiers résultats sont d'ores et déjà reconnus et appréciés de la communauté D0 au LPSC et dans l'expérience internationale D0 dans son ensemble.

Développements, déploiement sur la grille :

Les récents développements dans le cadre de la collaboration CIMENT-LPSC sont les suivants :

- Déploiement de serveurs iRODS pour le stockage sur CIMENT, et au LPSC pour TIDRA
- Création de l'organisation virtuelle MURE pour la simulation des réacteurs nucléaires
- Développement de l'intergiciel CIGRI de CIMENT (aussi destiné à être déployé sur

- GRID 5000), définition des besoins de CIGRI
- Portage sur CIMENT d'une application spécifique de D0, PMCS, originalement implémentée pour être exécutée sur une ferme de calcul dédiée à Fermilab; le « package » portable ainsi obtenu s'exécute sans difficulté sur CIMENT et il est d'ailleurs maintenant aussi exécuté sur d'autres sites, notamment l'université de l'état de New-York.

Outils, difficultés rencontrées :

Un outil majeur de la collaboration CIMENT-LPSC est la mise en place d'un stockage spécifique. Pour cela, nous avons choisi la technologie iRODS. Celle-ci est particulièrement bien adaptée à un environnement hétérogène et elle est facile à utiliser (et à administrer). Ainsi, des utilisateurs n'ayant pas d'expérience sur la grille n'ont aucune difficulté à gérer leurs données et peuvent aisément passer de leur ferme de calcul dédiée à une grille.

iRODS permet en effet d'intégrer très facilement des espaces de stockage hétérogènes sur plusieurs sites en offrant ainsi une vue globale aux utilisateurs. Un catalogue de méta-données gère les informations de localisation des fichiers et permet aux machines clientes d'entrer en relation directe avec les nœuds de stockage via des commandes spécifiques. Le déploiement sur une grille existante est très simple puisqu'il suffit d'installer un serveur principal (icat) sur une machine centrale et des serveurs secondaires sur chaque nœud de stockage qui héberge des partitions standards. Dans CIMENT, nous avons déployé plusieurs nœuds de stockage sur trois sites différents. Il est ensuite possible de créer des règles qui permettent l'optimisation des accès aux fichiers : par exemple lorsqu'un fichier est répliqué sur plusieurs sites, l'accès peut se faire sur le nœud le plus proche ou le plus rapide. Diverses interfaces sont disponibles pour faciliter l'accès aux données depuis les calculateurs ou les postes de travail des utilisateurs : clients en ligne de commande (icommands), interface web, api, webdav, montage fuse.

Pour l'utilisation avec CIGRI, les données d'entrée sont rapatriées par l'utilisateur en début de job sur le nœud de calcul et les données de sortie en fin de job sont aussi déposées sur iRODS, ceci de la même manière quel que soit le site de calcul. L'utilisateur peut ajouter des méta-données aux fichiers (par exemple l'identifiant unique de la tâche qui a servi à calculer le résultat). CIGRI a été modifié pour qu'il soit possible d'exécuter un script prologue sur chaque frontale de calculateur avant l'exécution d'un ensemble de tâches. L'application et les données communes qui sont préalablement stockées dans iRODS sont ainsi déployées à la volée au démarrage d'un ensemble de tâches.

Grâce à cette première mise en œuvre de iRODS sur la grille de CIMENT, nous avons pu identifier de nouvelles fonctionnalités à inclure dans la prochaine version de CIGRI actuellement en cours d'écriture. Par exemple, la prise en charge du dépôt des résultats en fin de job par CIGRI directement, ceci afin d'exclure le processus de la partie « best-effort » du job car il est délicat.

En ce qui concerne les difficultés rencontrées, elles sont de deux types : les difficultés liées à la technique et celles liées aux aspects humains.

Difficultés techniques :

- MPI sur grille EGI (mauvaise efficacité des jobs MURE)
- Limitations de l'utilisation MPI sur CIGRI

Difficultés côté utilisateurs :

- difficultés à accepter de mutualiser les ressources provenant de financements propres
- difficultés d'adaptation à la complexité de la grille EGI

Résultats scientifiques :

La collaboration étant récente, les thématiques scientifiques qui utilisent les ressources travaillent encore à l'exploitation de leurs données. Cependant certaines ont déjà obtenu et publié des résultats dont un échantillon est donné en référence ([7],[8]).

Perspectives :

Le développement de la collaboration entre les deux sites a pour but, à moyen terme, de :

- Mettre en place un stockage commun iRODS entre le LPSC et CIMENT
- Intégrer le site du LPSC dans la grille CIMENT
- Faciliter les échanges de communautés sur ces différentes grilles

Le but ultime est de mettre en valeur les ressources qui sont proposées mais non utilisées par ces grilles. La mutualisation des ressources en est un élément clé et l'utilisation opportuniste de D0 de la grille CIMENT est une première manifestation de cette nouvelle synergie.

Références :

- [1] The CiGri grid middleware: <https://ciment.ujf-grenoble.fr/wiki/index.php/CiGri>
- [2] IRODS:Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems: <https://www.irods.org>
- [3] Collaboration D0, V. Abazov et al., « The upgraded D0 detector », Nuclear Instruments and Methods in Physics Research, Vol. 565, 2 463-537 (2006).
- [4] The Tevatron Electroweak Working Group for the CDF and D0 Collaborations, « Updated Combination of CDF and D0 Results for the Mass of the W Boson », TEVEWWG/WZ 2009/01, arXiv:0908.1374v1 [hep-ex].
- [5] Collaboration D0, V. Abazov et al., « Measurement of the W Boson Mass », Phys. Rev. Lett. 103, 141801 (2009).
- [6] P. Renton, [34th International Conference on High Energy Physics](#) (la plus grande conférence internationale de la Physique des Particules), « Global Electroweak Fits and the Higgs Boson », Philadelphia (2008).
- [7] M. Lafaysse, B. Hingray, P. Etchevers, E. Martin et C. Obled, « Influence of spatial discretization, underground water storage and glacier melt on a physically based hydrological model of the Upper Durance River basin ». Journal of Hydrology, 2011, 403, 116129 (2011); doi:10.1016/j.jhydrol.2011.03.046.
- [8] W. Thuiller, S. Lavergne, C. Roquet, I. Boulangeat, B. Lafourcade et M. B. Araújo, « Consequences of climate change on the Tree of Life in Europe ». Nature, 470, 531-534 (2010) <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature09705.html>