

Développement et déploiement d'une application « desktop » et mobile sur une mini grille de calcul à l'Université de Kinshasa pour le traitement des données GPS

P. Ntumba(1), V. Lotoy(2), S.J. Djungu(3), R. Fleury(4), M. Petitdidier(5), A. Gemünd(6), H. Schwichtenberg (7)

(1) *passyntumba@gmail.com, Univ. de Kinshasa, Département de Mathématiques et Informatique (RDC)*

(2) *vianneylotoy@gmail.com, Univ. de Kinshasa, Département de Mathématiques et Informatique (RDC)*

(3) *saintjean.djungu@unikin.ac.cd; Univ. de Kinshasa, Département de Mathématiques et Informatique (RDC)*

(4) *rolland.fleury@telecom-bretagne.eu, Institut Mines-Télécom/Télécom Bretagne, Brest, France*

(5) *monique.petitdidier@latmos.ipsl.fr, latmos, IPSL, CNRS/INSU, UVSQ, Guyancourt, France*

(6) *andre.gemuend@scai.fraunhofer.de, SCAI, Fraunhofer, Sankt Augustin, Germany*

(7) *horst.schwichtenberg@scai.fraunhofer.de, SCAI, Fraunhofer, Sankt Augustin, Germany*

Overview

All over the world, permanent networks of GNSS (Global Navigation Satellite System) receivers have been deployed and regularly-updated databases are available. The treatment of GPS data needs a large number of daily measurements to determine, locally or regionally, seasonal and inter-annual variations of the TEC (Total Electronic Content) ; or for follow-up on specific events. In equatorial-and-tropical African regions, the equatorial electrojet, located at 100km altitude, induces electric and magnetic perturbations. Those perturbations are critical for telecommunications, aviation, magnetic measurements; and need further studies; the challenge being real time information.

R. Fleury developed a software using the Matlab environment to treat GPS data to provide TEC results. He has provided his software and trained African scientists in using it. Several points have to be improved to use it efficiently. At the University of Kinshasa (RDC), P. Ntumba (2012) and V. Lotoy (2012) addressed the main points to be improved. The first stage was to convert the code to the non-commercial Octave environment. The second stage was to avoid using only one PC or several PCs separately to treat collections of GPS files. The choice of using technologies from Grid computing is justified by its capacity to dynamically integrate heterogeneous machines while providing secure and transparent access for the user. A mini computing grid (4PCs) was deployed by using *Globus toolkit 5*, *Condor* and *Simple CA* to be a Certificate Authority. As a third step, two user-friendly interfaces for the software have been implemented. One is a user interface written in Java, for use on classical desktops, while the second is aimed at Android mobile phones, while offering the same functionality.

Currently, results are obtained to validate both the code in Octave environment and the treatment via the Grid. The treatment on a local PC with the developed user interface works well and will be provided to other African users. An English interface will be done for already numerous users of this software. Additional steps for future work are to improve the mini computing grid following user feedbacks and to increase the number of compute elements in order to port other applications. Another important aspect concerns data storage and sharing, as some reduced data or files may be used by several users.

This experience shows that a Grid with a relatively low amount of resources is more efficient than the same amount of resources non-linked, especially with a collection of files to be treated. This approach helps to avoid unnecessary duplication of effort and thus decreases time that have to be invested in upscaling of applications treated.

1 Enjeux scientifiques et besoin en calcul pour le traitement de données GPS

Des réseaux de récepteurs permanents GNSS (Global Navigation Satellite System) pour des études scientifiques en géophysique ont été installés à travers le monde et de grandes bases de données, quotidiennement mises à jour, sont maintenant constituées, comme IGS (*International GPS Service*) ou EUREF (*European terrestrial reference*). En Afrique de l'Ouest et Centrale quelques unités de récepteurs GPS ont été déployés pour l'expérience AMMA (*Analyse multidisciplinaire de la Mousson Africaine*) et le projet IHY (*International Heliophysical Year*) à partir de 2005. Récemment, des initiatives nationales (*NIGNET* au Nigeria) ou internationales (*TRIGNET* en Afrique du Sud) permettent de disposer en ces régions de mesures denses et régulières. Globalement, l'exploitation des mesures GPS pour des études ionosphériques relève du calcul intensif. En effet un grand nombre de mesures sont à traiter pour déterminer les variations saisonnières et interannuelles du TEC (Contenu Total Electronique) localement ou régionalement ou faire le suivi d'événements spécifiques du TEC. Les mesures ionosphériques en zone tropicale et équatoriale sont très importantes car des perturbations électriques et magnétiques sont dues à la présence à 100km d'altitude d'un fort courant essentiellement de jour appelé électrojet équatorial. Ces perturbations gênent les télécommunications, l'aviation, les mesures magnétiques....Aussi il est important de les étudier afin de mieux connaître l'ensemble des processus en jeu avec un enjeu de fournir les informations en temps réel.

Des scientifiques Africains ont été formés au traitement de ces données GPS pour des études

ionosphériques avec le logiciel développé par R. Fleury [1]. La méthode est inspirée des travaux menés au Jet Propulsion Laboratory [2]. Le traitement des données GPS par le logiciel développé par R. Fleury et mis à disposition après formation, nécessite: (1) une licence de Matlab, langage utilisé ; (2) l'accès à des ressources de calcul et stockage suffisantes pour traiter des séries temporelles importantes, (3) plus de convivialité pour l'utilisation de ce logiciel. A l'Université de Kinshasa, P. Ntumba [3] et V. Lotoy [4] ont adressé l'ensemble de ces points afin de permettre aux étudiants et chercheurs d'exploiter les données GPS dans de bonnes conditions.

2. Déploiement d'une mini grille de calcul et d'interfaces pour l'utilisateur

La première étape, de ce travail, a été de comprendre le fonctionnement de ce logiciel puis de convertir le logiciel de traitement des mesures GPS de Matlab sous Windows en Octave sous Linux, afin de s'affranchir de licence commerciale et l'exécuter sur une plus grande variété de machines. Ensuite un questionnaire a été envoyé aux utilisateurs de ce logiciel afin de mieux cerner leurs besoins, en particulier pour l'interface utilisateur. Cette nouvelle version du logiciel et son interface utilisateur seront proposés aux autres utilisateurs de ce logiciel.

La seconde étape a consisté à définir comment utiliser au mieux des ressources de calcul limitées, la grille de calcul a été l'option choisie. Ce choix est motivé car une grille de calcul accepte des ressources hétérogènes, peut en intégrer de nouvelles et assure un accès sécurisé et transparent à l'utilisateur [5]. Dans le cas de collection de jobs, elle gère la distribution des travaux sur les processeurs qui sont ou deviennent disponibles. Elle peut être utilisée par différentes applications.

Pour cette étude de faisabilité les PCs de cette mini grille de calcul sont décrits dans le tableau 1.

Nom	CPU	Qté CPU	Vitesse (Ghz)	RAM (Mo)	HDD (Go)	SE
Master	Intel Core Duo	2	1.40	4096	120	Ubuntu 11.10
Node0	Intel Pentium 4	1	1	512	40	Ubuntu 10.10
Node1	Intel Pentium 4	1	1	512	40	Ubuntu 10.10
Node2	Intel Pentium 4	1	1	512	40	Ubuntu 10.10

Tableau 1 : Caractéristiques des PCs de la grille de calcul

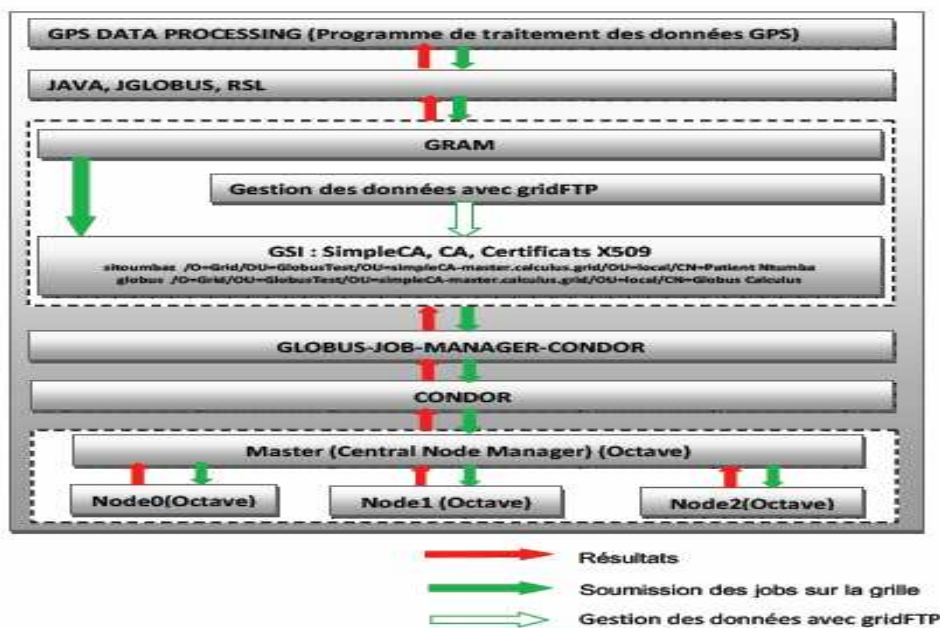


Figure 1 : Architecture de la grille de calcul

Globus Toolkit 5 est l'intergiciel choisi et Condor (v8.3), le système de gestion des ressources locales (Local Resource Manager System). Les deux systèmes ont été couplés pour permettre aux utilisateurs de lancer à distance et d'une manière sécurisée sur la grille des jobs incluant des fichiers.

- **Sur le serveur « maître »:**

- *Gram5* (Grid Resource Allocation Manager) l'utilisateur définit ce dont il a besoin pour son application et GRAM5 installe les fichiers demandés, puis soumet la tâche à l'Ordonnanceur local, Condor, et surveille l'exécution du travail;
- *Gridftp* pour la gestion et le transfert sécurisé des données, très utilisé en dehors des grilles de calcul ou de données, dans ce cas sert à transférer les fichiers à traiter sur la grille via les interfaces;
- *GSI* (*Globus Security Infrastructure*) contrôle l'accès aux ressources et *Simple CA* (simple Certificate Authority) est l'autorité de certification qui fournit à tout utilisateur, autorisé à accéder à la grille, un certificat X509. Cette solution simple n'est valable que dans le cas d'un seul site.
- *Condor* les services de Condor installés servent à la gestion de la grille. Ce sont *Condor Master* qui coordonne les différents services de ce serveur; *Condor Collector* qui collecte les informations des nœuds de travail; *Condor Negotiator* qui interroge *Condor Collector* pour connaître quel nœud de travail est disponible pour lui assigner un job; *Condor Startd* qui est chargé d'exécuter le job qui lui est assigné; *Condor Schedd* qui met dans une file d'attente les jobs qui lui sont assignés. Comme le nombre de nœuds de cette mini grille est de 3, le serveur a été aussi configuré comme nœud de travail, en activant le service *Condor Startd*.

- Sur les **3 autres nœuds de travail** (worker node) *Condor* est installé avec les services suivants : *Condor Master* qui coordonne les différents services associés au nœud; *Condor Startd* qui exécute le job qui lui est assigné; et *Condor Schedd* qui met dans une file d'attente les jobs qui lui sont assignés.
- Sur chaque PC, Octave (v3.1) est installé et configuré. Ces différents PCs ont été connectés au réseau local.

Interface utilisateur

Ce logiciel vise un public varié. Il peut s'agir d'un scientifique isolé, débutant dans le domaine ionosphérique ou d'un scientifique confirmé avec une équipe de plusieurs chercheurs. Aussi rendre ce logiciel d'une utilisation conviviale évite de faire des erreurs, de perdre du temps et donc facilite la gestion du travail.

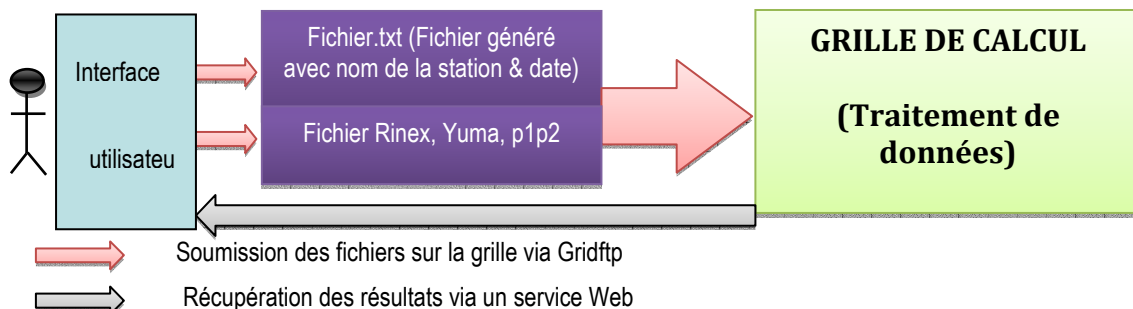


Figure 2 : Traitement par l'utilisateur de l'interface graphique jusqu'au rapatriement des résultats.

- **Interface graphique sur Desktop**

Une interface graphique destinée au traitement de données GPS a été développée en Java (cf. Fig. 3). Elle doit remplir plusieurs fonctions: (1) créer un fichier de gestion du programme, lu en début d'exécution de celui-ci, (2) vérifier dans l'espace de travail la présence des fichiers nécessaires à l'exploitation des données GPS, (3) finalement lancer le job et (4) récupérer les résultats.

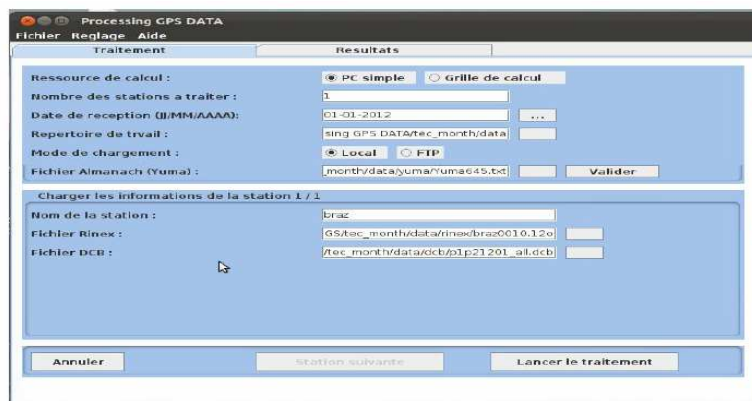


Figure 3 : interface graphique utilisateur

Le fichier de gestion du travail est créé automatiquement une fois que l'utilisateur aura indiqué s'il traite les données localement ou sur la grille, le nombre de stations à traiter et leurs noms, les dates. Au fichier GPS, il faut adjoindre les fichiers almanachs permettant de calculer l'orbite des satellites et les fichiers GIM permettant de calculer les biais récepteurs et satellites sur les mesures GPS enregistrées. Compte tenu de ces informations demandées, la présence de ces fichiers dans la zone de travail est vérifiée. S'ils ne sont pas présents, une procédure de téléchargements est lancée sur le site (ou les sites) du serveur où ils sont archivés. Ensuite le job est lancé soit localement soit sur la grille et finalement, les résultats sont rapatriés et stockés.

- **Interface sur mobile Android**

La tendance actuelle favorise la migration des technologies développées sur des ordinateurs vers les appareils mobiles. L'idée a été de développer une interface utilisateur graphique sur mobile Android qui remplisse les mêmes fonctions que l'interface Desktop. Cette interface a été développée et testée.



Figure 4 : Interface graphique en Android

En conclusion, le traitement des données peut être lancé par l'intermédiaire d'une application Desktop ou Android pour l'exploitation des données par la grille de calcul. Les 2 applications ont été testées et fonctionnent correctement.

3 Résultats et Perspectives

De nombreux tests ont été effectués pour vérifier la robustesse de l'installation. Les premiers résultats scientifiques, obtenus sur la station GPS de Brasilia (BRAZ) au Brésil, ont été validés par comparaison avec ceux obtenus avec la version originale du code en Matlab et par consistance avec ce qui est attendu pour la variation diurne du TEC. La version avec calcul en local va commencer à être utilisée et disséminée à des utilisateurs d'autres pays ; une version anglophone est prévue.

Les étapes suivantes consistent à améliorer cette grille, après des tests faits par les utilisateurs, pour la mettre en production et à augmenter le nombre de nœuds de calcul. Un autre aspect qui sera pris en compte est le stockage et partage des données. En effet plusieurs utilisateurs peuvent avoir besoin de données GPS déjà exploitées et des fichiers annexes pour différentes applications.

Cette mini grille de calcul avec un accès sécurisé par interface soit graphique soit mobile Android a montré l'intérêt de mutualiser ne serait-ce que quelques ressources de calcul, au lieu de les utiliser séparément, afin de traiter une collection de jobs, et d'éviter ainsi la duplication d'efforts.

4 Références :

- [1] Fleury R., Modélisation du TEC vertical à partir des fichiers rinex journaliers en utilisant les cartes GIM, communication privée, 2007.
- [2] Komjathy A., L. Sparks, B.D. Wilson, A.J. Mannucci, Automated daily processing of more than 1000 ground-based GPS receivers for studying intense ionospheric storms, radio science, 40,rs60006, 2005. doi:10.1029/2005RS003279
- [3] P. Ntumba, Traitement intensif des données GPS sur une grille de calcul destiné aux études ionosphériques. Licence en Sciences, option génie Informatique. Faculté des Sciences, département de Mathématiques et d'Informatique – Université de Kinshasa, RDC, 2012.
- [4] V. Lotoy, Développement d'un client léger mobile Android pour une grille mobile de traitement de données GPS. Licence en Sciences, option génie Informatique. Faculté des Sciences, département de Mathématiques et d'Informatique – Université de Kinshasa, RDC, 2012.
- [5] Foster I., and C. Kesselman (Eds.) , *The Grid:Blueprint for a New Computing Infrastructure*, 1998, 677 pp., Morgan Kaufmann, San Francisco, Calif.