

Rapport de la rencontre entre les représentants de l'équipe Offline ALICE et le CCIN2P3

01 janvier 2012

ALICE : Yves Schutz, Federico Carminati, Peter Hristov, Andrei Gheata et Mihai Niculescu.

CCIN2P3 : Loïc Tortay, Bernard Chambon, Frédéric Suter, Pierre Girard, Rachid Lemrani, Renaud Vernet, Xavier Canehan, Yannick Perret, Fabien Wernli.

ALICE recherche un partenariat avec le CCIN2P3 en tant que fournisseur de ressources informatiques matérielles. L'expérience souhaite tester les nouvelles implémentations orientées *manycore* de son infrastructure logicielle, qui ont d'ores-et-déjà débuté.

ALICE a exposé au cours de cette réunion son intention d'adapter son environnement logiciel afin de bénéficier de technologies multi-cœurs : *multi-core*, *manycore (MIC ou GPU)*, *cluster*. Le niveau de parallélisme et les performances offerts par ces matériels devront permettre à l'expérience d'augmenter considérablement sa puissance de calcul et d'optimiser l'utilisation de la mémoire, et de réellement profiter des opportunités offertes par ces technologies, ce dont bénéficiera en conséquence la physique.

Compte-tenu de la relative nouveauté de ces technologies, de leur manque de standards disponibles au niveau du développement logiciel et de la non-expertise de la plupart des développeurs de l'expérience, ALICE souhaite poursuivre l'utilisation du C++ en s'appuyant sur les *pragmas* pour s'adapter aux différents matériels déployés par les sites. Ceci assurera à long terme l'indépendance totale du code par rapport aux matériels hétérogènes disponibles sur chaque site.

Une ré-écriture complète de son environnement logiciel est envisagée. Le travail de R&D devrait s'achever fin 2012. La réelle phase d'implémentation commencera et s'étendra vraisemblablement jusqu'à 2014. En effet, le développement tirera parti des 2 arrêts planifiés du LHC pour paralléliser son code au sein d'un serveur pendant le premier, puis au niveau d'un cluster pendant le second.

L'expérience souhaite pouvoir commencer ses tests sur ce type de matériel dès la mi-2012, sur des machines dotées de systèmes d'exploitation et noyaux récents.

Le CCIN2P3 examine plusieurs pistes d'évolution pour son infrastructure de calcul :

1. Machines de calcul *whole-node* : la totalité d'une machine standard du batch dédiée à un job.
2. Virtualisation de la ferme de calcul : d'une part, sur le moyen terme, étude d'une infrastructure fédérant des Cloud IAAS, orientée HTC *High Throughput Computing* (ex-Capri). D'autre part, immédiatement, des tests comparatifs de machines virtuelles dédiées au calcul et pouvant aller jusqu'au *whole-node*.
3. Intégration de matériels récents : multi-cœur (dont les derniers Nehalem), MIC, GPU. Les demandes des utilisateurs permettront d'orienter les acquisitions éventuelles.

ALICE propose donc de profiter de ces investigations pour bénéficier des tests d'infrastructure. Les besoins d'ALICE s'intègrent aux études en cours, que le CCIN2P3 pourra infléchir. Le CCIN2P3 bénéficierait aussi des retours d'expérience, en particulier pour dimensionner ses futurs systèmes (CPU/MIC/GPU, changement de pattern IO...).

Trois axes principaux ont été suggérés par le CCIN2P3 :

1. À court terme, participation au test de serveurs *whole-node*. En parallèle, des tests comparatifs peuvent avoir lieu sur des machines virtuelles dont l'hyperviseur utilise un système d'exploitation récent. ALICE réclame une version du kernel plus récente que SL5, notamment pour utiliser KSM-KVM. La distribution Linux Ubuntu 11.10 est à l'heure actuelle souhaitée.

Dans ce cas ALICE demanderait l'accès à des outils de monitoring système. La façon de répondre à cette demande reste à évaluer au CCIN2P3, en particulier pour les récents tests de virtualisation.

2. Sous quelques mois, le test de matériels neufs, équipés de dispositifs *manycore* ou de processeurs récents. La phase de qualification des matériels Dell débutant, ALICE est encouragée à présenter des besoins factuels, permettant de dimensionner au mieux ces matériels.
3. Frédéric Suter propose l'utilisation de l'infrastructure française Grid'5000 qui propose une vaste gamme de matériels adéquats. Une demande spécifique devra être rédigée et présentée F. Suter, qui orientera les échanges.

Ces solutions permettraient à ALICE de faire face à plusieurs types de matériels, ce qui est souhaité par l'expérience. Des tests de comparaison de performances sur plusieurs combinaisons de types de machines physiques et virtuelles, ainsi qu'entre machines physique et virtuelle, sont envisagés par l'expérience.

Dans tous les cas, l'utilisation de machines virtuelles est une direction envisagée. Ceci pourra notamment pallier les éventuelles disparité et vétusté des systèmes d'exploitation au niveau des sites. Sa mise en œuvre pourra être effectuée à la fois grâce aux cadres fournis par les projets ex-Capri et Grid'5000.

Le projet d'upgrade de l'environnement logiciel d'ALICE devra être présenté aux autorités de tutelle, ce qui peut influencer sur le programme.

Le gain d'efficacité attendu par ALICE permettra une meilleure utilisation des ressources au CCIN2P3. Cependant, la production d'ALICE restera à terme contrainte par la quantité des futures ressources allouées pour le calcul au LHC.