

Sujet de master recherche « Architectures logicielles distribuées » 2005-2006**Traitement des streams dans un système P2P****Encadrant principal** : Patrick.Valduriez@inria.fr**Co-encadrant** : Esther.Pacitti@univ-nantes.fr**Prérequis** : Java, XML, Linux**Mots-clés** : bases de données, traitement des streams, pair-à-pair**Objectif du stage**

Récemment, de nombreux travaux de recherche d'envergure se sont intéressés au traitement des streams de données [1][2][3]. Des exemples d'applications des streams sont la surveillance du Web, le traitement d'images satellites, la surveillance des appels téléphoniques, le contrôle de fabrication, etc. Dans la plupart de ces applications, les données sont produites en continu par des sources distribuées, puis envoyées à des noeuds où des requêtes sont continûment évaluées. En général, les streams sont en très grande quantité et consomment beaucoup de ressources informatiques pour être traitées.

Une solution nouvelle pour traiter ces streams de données est d'utiliser un système P2P [4][5]. Le paradigme du P2P a émergé comme une manière populaire de partager de très grandes quantités de ressources [7] [8]. A l'origine, les systèmes P2P ont été exploités pour partager des fichiers multimédias, mais ils sont devenus intéressants dans d'autres applications, par exemple, le partage des ressources de stockage, de calcul, etc. Une raison majeure du succès du P2P sont l'aspect décentralisé qui permet une grande tolérance aux fautes et le passage à l'échelle.

Nous avons l'intention de réaliser un service de traitement des streams dans le contexte d'APPA. APPA (*Atlas Peer-to-Peer Architecture*) [6] est un nouveau système de gestion de données P2P que nous construisons au LINA. Les objectifs d'APPA sont le passage à l'échelle, la disponibilité des données et les performances pour des applications avancées. APPA a une architecture indépendante du réseau, en termes de services de base et de service avancés, afin de pouvoir fonctionner sur différents réseaux P2P (non structuré, DHT, super-peer). L'un des services avancés est le traitement de requête. APPA est en cours de réalisation avec le framework JXTA [8].

L'objectif du stage est de concevoir un service efficace pour le traitement des streams sur APPA et réaliser ce service en utilisant les services de base fournis par APPA. Le langage de programmation est Java et le langage pour représenter les structures de données communes est XML.

Travail à réaliser

1. Faire un état de l'art sur le traitement des streams.
2. Concevoir un service efficace pour le traitement des streams sur APPA.
3. Développement d'un prototype en Java sur JXTA.
4. Etude de performances par expérimentation sur un cluster avec 64 noeuds.

Références

- [1] B. Babcock et al. Models and Issues in Data Stream Systems. *Proc. of 21st ACM Symposium on Principles of Database Systems (PODS)*, 2002.

- [2] S. Babu, U. Srivastava, and J. Widom. Exploiting k-Constraints to Reduce Memory Overhead in Continuous Queries over Data Streams. *J. ACM Transactions on Database Systems*, Sept 2004.
- [3] S. Babu et al. Adaptive Ordering of Pipelined Stream Filters. *Proc. of the ACM Intl Conf. on Management of Data (SIGMOD)*, 2004.
- [4] R. Kuntschke et al. StreamGlobe: Processing and Sharing Data Streams in Grid-Based P2P Infrastructures. *Proc. of the 31st Int. Conf. on Very Large Data Bases (VLDB)*, 2005.
- [5] L. Guo et al. A Scalable and Reliable P2P Assisted Proxy Streaming System. *24th Int. Conf. on Distributed Computing Systems (ICDCS)*, 2004.
- [6] R. Akbarinia, V. Martins, E. Pacitti, P. Valduriez. Replication and Query Processing in the APPA Data Management System. *Distributed Data & Structures 6 (WDAS): Records of the 6th International Meeting (Lausanne, Switzerland)*, Waterloo: Carleton Scientific, 2004.
- [7] Gnutella. <http://www.gnutelliums.com/>.
- [8] Kazaa. <http://www.kazaa.com/>.
- [9] JXTA. <http://www.jxta.org/>.
- [10] T. Özsu, P. Valduriez: *Principles of Distributed Database Systems*. Prentice Hall, 2^{ème} edition, 666 pages, 1999.
- [11] T. Özsu, P. Valduriez: Distributed and Parallel Database Systems. *Handbook of Computer Science and Engineering*, 2nd Edition, A. Tucker (ed.), CRC Press, 2004.