

Nouvelle technique pour le traitement et le mapping des requêtes SPARQL complexes sur Hive basé sur l'ingénierie des modèles

¹Mouad Banane ^{1*} Allae Erraissi² Abdessamad Belangour ²

*lead presenter

¹mouadbanane@gmail.com, Hassan 2 University, Morocco

² Hassan 2 University, Morocco

La croissance des données Web a présenté de nouveaux défis en ce qui concerne la capacité d'interroger efficacement les données RDF. Les systèmes de bases de données relationnelles traditionnels adaptent et interrogent efficacement les données distribuées. Avec le développement de Hadoop, son implémentation du Framework MapReduce avec Hive, un entrepôt de données, la sémantique du traitement et de l'interrogation des données a changé.

Nous présentons dans cet article, SPARQL2Hive un système de traitement des requêtes SPARQL compétitif sur le Framework MapReduce qui permet un traitement de requête SPARQL ad hoc sur de grands graphes RDF. Au lieu d'un mapping direct, SPARQL2Hive utilise le langage de requête de Hive(HiveQL), Hive est un système d'entrepôt de données qui permet d'interroger des systèmes utilisant HDFS, située au-dessus de Hadoop MapReduce, en tant que couche intermédiaire entre SPARQL et MapReduce. Ce niveau d'abstraction supplémentaire rend notre approche indépendante de la version actuelle de Hadoop et garantit ainsi la compatibilité avec les futures modifications du framework Hadoop car elles seront couvertes par la couche Hive sous-jacente. Notre approche consiste à utiliser les deux méta-modèles de SPARQL et Hive, et de proposer une transformation/mapping entre ces deux méta-modèles en utilisant le langage de transformation Atlas(ATL). Nous comparons SPARQL2Hive avec les implémentations SPARQL basées sur MapReduce, qui proposent le traitement des requêtes SPARQL dans un environnement basé sur Hadoop.

Références

1. Banane, M., Belangour, A., & Houssine, L. El. (2019). Lecture Notes in Real-Time Intelligent Systems (Vol. 756). Springer International Publishing.
<http://doi.org/10.1007/978-3-319-91337-7>.

2. Banane, M., Belangour, A., & Labriji, E. H. (2018). RDF Data Management Systems Based on NoSQL Databases : A Comparative Study, 58(2), 98–102..
3. Guo, Y., Pan, Z., & Heflin, J. (2005). LUBM: A benchmark for OWL knowledge base systems. *Web Semantics*, 3(2–3), 158–182. <http://doi.org/10.1016/j.websem.2005.06.005>.
4. Punnoose, R., Crainiceanu, A., & Rapp, D. (2012). Rya: a scalable RDF triple store for the clouds. *Proceedings of the 1st International Workshop on Cloud Intelligence*, 4. <http://doi.org/10.1145/2347673.2347677>.
5. Gu, R., Hu, W., & Huang, Y. (2015). Rainbow: A distributed and hierarchical RDF triple store with dynamic scalability. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 561–566. <http://doi.org/10.1109/BigData.2014.7004274>.
6. Erraissi, A., & Belangour, A. (2018). Data sources and ingestion big data layers: meta-modeling of key concepts and features. *International Journal of Engineering & Technology*, 7(4), 3607–3612. <https://doi.org/10.14419/ijet.v7i4.21742>
7. Allae Erraissi, Abdessamad Belangour, Abderrahim Tragha. “A Big Data Hadoop Building Blocks Comparative Study.” *International Journal of Computer Trends and Technology*. Accessed June 18, 2017. <http://www.ijcttjournal.org/archives/ijctt-v48p109>