

NEW CLASS – SPRING ‘10

CS395T The Semantic Web, Ontologies and Cloud Databases

Prof. Daniel P. Miranker

Mondays 3:30 – 6:30

The evolution of computing as a utility provided by a distributed network of computers brings with it new problems and opportunities in data management and data integration. This topics course will cover two contrasting developments. Cloud Databases (attribute/value stores), are nascent commercial services created and offered by some of the most famous companies. These systems provide low-level interfaces. The Semantic Web is a related collection of technologies ratified by the W3C that includes, in principle, the association of metadata encoded as an ontology with each and every web site. The promise is that that will enable, high-level interfaces that prove to be the basis for much improved document search, and automatic integration of distributed data. Although aspects of the Semantic Web are now gaining commercial traction, the larger promise is still largely the province of research projects. Curiously, these developments are intersecting in a movement called NOSQL, “a database movement which began in early to mid 2009 and which promotes non-relational data stores that do not need a fixed schema”, and hence are both harbingers of a new era of databases beginning from a fresh start (i.e. no SQL).

Organization:

Professor Miranker will present a small number of formal introductory lectures on the basic technologies. Subsequently students will present papers in round-robin fashion. This will comprise one or two presentations depending on the size of the class. A partial, preliminary reading list is below. Grading will be based on presentations, class participation and a term project. A list of term projects is available upon request (Miranker@cs). The list is not posted publicly as these areas are new and moving quickly, and a number of projects on the list are novel ideas that could easily lead to publication. Students may also organize their own projects and are encouraged to nominate papers.

Cloud Databases:

Chang, F., Dean, J., Ghemawat, S., Hsieh, W., Wallach, D., Burrows, M., Chandra, T., Fikes, A., and Gruber, R. 2008. Bigtable: A Distributed Storage System for Structured Data. *ACM Trans. Comput. Syst.* 26, 2 (Jun. 2008)

Dean, J. and Ghemawat, S. 2008. MapReduce: simplified data processing on large clusters. *Commun. ACM*

DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Voshall, P., and Vogels, W. 2007. Dynamo: amazon's highly

available key-value store. SIGOPS Oper. Syst. Rev. 41, 6 (Oct. 2007), 205-220.

Lakshman, A., and Malik, P. 2009 Cassandra – A Decentralized Structured Storage System . In Proceedings of 2009 ACM SIGOPS International Workshop on Large Scale Distributed Systems and Middleware.

Yang, H., Dasdan, A., Hsiao, R., and Parker, D. S. 2007. Map-reduce-merge: simplified relational data processing on large clusters. In Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data (Beijing, China, June 11 - 14, 2007).

Zhou Wei, Guillaume Pierre and Chi-Hung Chi. Scalable Transactions for Web Applications in the Cloud. In proc. of the Euro-Par Conference, January 2009

The Semantic Web:

Christian Bizer, Tom Heath and Tim Berners-Lee (2009) Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, Vol. 5(3)

Pérez, J., Arenas, M., and Gutierrez, C. 2009. Semantics and complexity of SPARQL. ACM Trans. Database Syst. 34, 3 (Aug. 2009)

Tirmizi, S. H., Sequeda, J., and Miranker, D. 2008. Translating SQL Applications to the Semantic Web. In Proceedings of the 19th international Conference on Database and Expert Systems Applications (Turin, Italy, September 01 - 05, 2008).

Abadi, D. J., Marcus, A., Madden, S. R., and Hollenbach, K. 2007. Scalable semantic web data management using vertical partitioning. In Proceedings of the 33rd international Conference on Very Large Data Bases (Vienna, Austria, September 23 - 27, 2007).

Abadi, D. J., Madden, S. R., and Hachem, N. 2008. Column-stores vs. row-stores: how different are they really?. In Proceedings of the 2008 ACM SIGMOD international Conference on Management of Data (Vancouver, Canada, June 09 - 12, 2008). SIGMOD '08.

Weiss, C., Karras, P., and Bernstein, A. 2008. Hexastore: sextuple indexing for semantic web data management. Proc. VLDB Endow. 1, 1 (Aug. 2008),

Ma, L., Wang, C., Lu, J., Cao, F., Pan, Y., and Yu, Y. 2008. Effective and efficient semantic web data management over DB2. In Proceedings of the 2008 ACM SIGMOD international Conference on Management of Data (Vancouver, Canada, June 09 - 12, 2008). SIGMOD '08.