A System for Dynamic Load Balancing in Large-scale Graph Processing

Panos Kalnis
Associate professor in the Division of Computer Electrical and Mathematical Sciences and Engg
The King Abdullah University of Science and Technology (KAUST)

Abstract:

Pregel was recently introduced as a scalable graph mining system that can provide significant performance improvements over traditional MapReduce implementations. Pregel follows the Bulk Synchronous Parallel (BSP) programming model; it uses asynchronous message passing to better balance computation and communication. However, achieving an optimal balance is not straightforward. Existing implementations rely primarily on graph partitioning at a preprocessing step. However, graph partitioning alone is insufficient for minimizing execution time, especially where data is very large or the runtime behavior of the algorithm is unknown. To this end, we introduce Mizan, a Pregel clone that achieves efficient load balancing to better adapt to changes in computing needs. Unlike existing approaches, Mizan does not assume any a-priori knowledge of the structure of the graph or the behavior of the algorithm. Instead, it monitors the runtime characteristics of the system and performs efficient fine-grain vertex migration to balance computation and communication. Mizan outperforms Giraph (a popular Pregel implementation) by over 200% across a wide array of algorithms and datasets. Especially for highly-dynamic workloads, Mizan provides up to 84% improvement over techniques leveraging static graph pre-partitioning.

About the Speaker:

Panos Kalnis is associate professor in the Division of Computer, Electrical and Mathematical Sciences and Engineering in the King Abdullah University of Science and Technology (KAUST). In 2009 he was visiting assistant professor in the Dept. of Computer Science, Stanford University. Before that, he was assistant professor in the Dept. of Computer Science, National University of Singapore (NUS). In the past he was involved in the designing and testing of VLSI chips in the Computer Technology Institute, Greece. He also worked in several companies on database designing, e-commerce projects and web applications. He received his Diploma in Computer Engineering from the Computer Engineering and Informatics Dept., University of Patras, Greece in 1998 and his PhD from the Computer Science Dept., Hong Kong University of Science and Technology (HKUST) in 2002. His research interests include Databases, Cloud Computing, Distributed Systems, Large Graphs and Data Privacy. Further information can be found at: http://www.kaust.edu.sa/panoskalnis

All are welcome!
For enquiries, please call 2859 2180 or email enquiry@cs.hku.hk
Department of Computer Science
The University of Hong Kong