Curriculum Vitae

Dixin Tang

totemtang@uchicago.edu

The University of Chicago, Department of Computer Science

RESEARCH INTERESTS

Adaptive Database systems

EDUCATION

The University of Chicago, Department of Computer Science Sep. 2015 – Now
Ph.D. Candidate

Institute of Computing Technology, University of Chinese Academy of Sciences Sep. 2011 – July 2014
M.S in Computer Software and Theory

Huazhong University of Science & Technology Sep. 2007 – June 2011
B.S. in Software Engineering

EMPLOYMENT

Teaching Assistant Mar. 2016 – Now
The University of Chicago, Distributed Systems (MPCS 52040)

Research Assistant Jan. 2016 – Mar. 2016
The University of Chicago

Teaching Assistant Sep. 2015 – Dec. 2015
The University of Chicago, C programming (MPCS 51040)

Research Assistant Sep. 2012 – July 2015
Institute of Computing Technology, University of Chinese Academy of Sciences

PUBLICATION

D. Tang, T. Liu, R. Lee, H. Liu and W. Li, "A Case Study of Optimizing Big Data Analytical Stacks Using Structured Data Shuffling," BigData Congress, 2016 (To appear).

D. Tang, T. Liu, R. Lee, H. Liu and W. Li, "A Case Study of Optimizing Big Data Analytical Stacks Using Structured Data Shuffling," Cluster Computing (CLUSTER), 2015 IEEE International Conference on, Chicago, IL, 2015, pp. 70-73.

D. Tang, T. Liu, H. Liu and W. Li, "RHJoin: A fast and space-efficient join method for log processing in MapReduce," Parallel and Distributed Systems (ICPADS), 2014 20th IEEE International Conference on, Hsinchu, 2014, pp. 975-980.

PROJECTS

Adaptive Concurrency Control for Main-Memory Multicore Databases Sep. 2015 – Now

- **Motivation**
  Current concurrency control algorithms are typically optimized for some types of workloads and thus fail to address the challenges of workloads variations
- **Personal Achievements**
  (1) Develop a prototype from scratch that supports adaptive switch among three representative algorithms according to workloads characteristics from major in-memory multicore databases. Our system examines a partition-based concurrency control (PCC) based on H-Store, an optimistic concurrency control (OCC) based on Silo, and a no-wait two-phase locking method (2PL)

(2) Experiments of two benchmarks show that our method can improve throughput by as much as 25x, 1.38x and 1.26x over PCC, OCC and 2PL respectively

Optimizing Big Data Analytical Stacks using Structured Data Shuffling Sep. 2013 – Oct. 2014

- **Motivation**

  Current general purpose data shuffling mechanically transfers data but not understands them. Therefore, when it is used to processes SQL queries, it could not apply efficient compression algorithms according to data characteristics and filter unnecessary data in advance according to query conditions.

- **Personal Achievements**

  (1) Developed a novel MapReduce-style data shuffling for SQL query processing, which leverages structure and query information in structured data to reduce overall shuffled data volume.

  (2) Experiments show performance of reduce phase and query performance is improved by 35.8x and 2.4x respectively.

A Fast and Space-efficient Join Method for Log Processing in MapReduce Nov. 2012 – July 2013

- **Motivation**

  Existing join methods in MapReduce for log processing cannot gain high query speed with affordable storage consumption at the same time

- **Personal Achievements**

  (1) Design and implement a join method called RHJoin which achieves high query performance with a small extra storage cost for log processing. It shuffles the log table to avoid huge storage consumption and optimizes the shuffle procedure to achieve high query performance.

  (2) Experiments show that RHJoin runs faster than standard join method by 46% and slower than the fastest join method by 5%. In addition, RHJoin reduces 59 times of extra storage space compared to the fastest join method.