Performance Breakdown for In-memory OLTP Databases under Skewed Workload

Group Info: Huanchen Zhang (huanche1@andrew.cmu.edu),
Zhuo Chen (zhuoc@cs.cmu.edu)

Web Page:
www.cs.cmu.edu/~zhuoc/15740project.html

Project Description
OLTP (online transaction processing) database systems are essential building blocks for many popular web services (e.g. Amazon) [2]. Due to the slow improvement of the bandwidth between disk and memory, and the fact that main memory has become sufficiently cheap, multicore main-memory OLTP databases have emerged and have been gaining more and more attention because of their much higher throughput (usually an order of magnitude improvement over disk-based systems). However, recent studies [1] [2] [3] have shown that the performance of a multicore main-memory OLTP database suffers from skewed (contended) workloads. In this work, we will use measurement tools to conduct a detailed performance breakdown analysis on some representative systems, such as H-Store, under realistic skewed workloads, to identify the performance bottlenecks first and then propose improvement suggestions based on the evaluation.

We highly suspect that a large portion of the transaction execution critical path is dedicated to concurrency control, especially under skewed workloads. For example, when data contention and data dependencies are high, protocols such as OCC (Optimistic Concurrency Control) will abort frequently and thus wasting computing resources. If this is the case, we will propose ideas on how to smartly (but safely) bypass unnecessary concurrency controls to achieve higher throughput. If time permits (125% goal), we will implement one of our ideas on a real system (likely H-Store).

Project Goal:
100% goal: Performance breakdown + optimization ideas
75% goal: If performance breakdown does NOT expose clear bottlenecks
125% goal: Performance breakdown + optimization ideas + one implementation on a real system

Week by week plan (labor separation):
10/26 - 11/1: read related works; select benchmarks (both)
11/2 - 11/8: get familiar with using the target database(s) (both)
11/9 - 11/15: measure memory performance (Huanchen); measure CPU performance (Zhuo)
11/16 - 11/22: continue the work from last week
11/23 - 11/29: closely look at concurrency control overhead (both), prepare milestone
11/30 - 12/6: identify bottlenecks and propose ideas for improvement

Milestone (11/25)
A performance breakdown graph of a main-memory OLTP database (likely H-Store) under different skewed workloads.

Resource Needed:
Open-source and well-documented main-memory OLTP database(s).
Easy to use standard benchmarks or benchmarks from real-world applications.

Work that have been done & questions:
We have read some related work. We have found H-Store to be one potential choice to analyse. However, H-store uses shared-nothing data partitioning design to improve multicore performance. To make our analysis more general, we are still looking for an open-source well-documented main-memory OLTP database that does not do data partitioning.

Literature Search:
[1] Narula, N., Cutler, C., Kohler, E., & Morris, R. Phase Reconciliation for Contended In-Memory Transactions.

[2] Mu, S., Cui, Y., Zhang, Y., Lloyd, W., & Li, J. (2014). Extracting More Concurrency from Distribted Transactions. Technical Report TR2014-970, New York University, Courant Institute of Mathematical Sciences.

[3] A. Pavlo, E. P. C. Jones, and S. Zdonik, "On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems," Proc. VLDB Endow., vol. 5, pp. 85-96, 2011.

[4] S. Harizopoulos, D. J. Abadi, S. Madden, and M. Stonebraker, "OLTP through the looking glass, and what we found there," in SIGMOD '08: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 2008, pp. 981-992.

[5] M. Stonebraker, S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland, "The end of an Architectural Era: (It's Time for a Complete Rewrite)," in VLDB '07: Proceedings of the 33rd international conference on Very large data bases, 2007, pp. 1150-1160.

[6] R. Kallman, H. Kimura, J. Natkins, A. Pavlo, A. Rasin, S. Zdonik, E. P. C. Jones, S. Madden, M. Stonebraker, Y. Zhang, J. Hugg, and D. J. Abadi, "H-Store: a High-Performance, Distributed Main Memory Transaction Processing System,"Proc. VLDB Endow., vol. 1, iss. 2, pp. 1496-1499, 2008.

[7] Difallah, D. E., Pavlo, A., Curino, C., & Cudre-Mauroux, P. (2013). OLTP-Bench: An extensible testbed for benchmarking relational databases.Proceedings of the VLDB Endowment, 7(4).

[8] Thomson, A., Diamond, T., Weng, S. C., Ren, K., Shao, P., & Abadi, D. J. (2012, May). Calvin: fast distributed transactions for partitioned database systems. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 1-12). ACM.