Jennifer Lam General Exam Spring 2022

Thermopylae: Exploiting Skew in High Throughput Distributed Databases

Abstract

Distributed databases scale out by spreading data across many machines, which allows them to support large-scale applications whose data is too large to fit on a single-machine database. This scalability introduces a dilemma, however, as the throughput of distributed databases is much lower than that of modern single-machine databases for skewed workloads. This paper eliminates this dilemma as much as possible. We present Thermopylae, the first distributed database that is able to both support large-scale applications and match its throughput to single-machine databases under skewed workloads. Central to Thermopylae is a novel hybrid architecture that embeds a high-performance single-machine database into a highly scalable distributed database. Thermopylae applies a specialized concurrency control protocol designed for its hybrid architecture. Our evaluation shows that Thermopyale achieves orders of magnitude better performance than a state-of-the-art distributed database and closely matches its throughput to a modern single-machine database under skewed workloads.

Reading List

1. Spanner.
    a. Corbett, James C., et al. "Spanner: Google's globally distributed database." ACM Transactions on Computer Systems (TOCS) 31.3 (2013): 1-22.
2. Linearizability.
    a. Herlihy, Maurice P., and Jeannette M. Wing. "Linearizability: A correctness condition for concurrent objects." *ACM Transactions on Programming Languages and Systems (TOPLAS)* 12.3 (1990): 463-492.
3. Hybrid Logical Clocks (HLC).
    a. Kulkarni, Sandeep S., et al. "Logical physical clocks." *International Conference on Principles of Distributed Systems*. Springer, Cham, 2014.
4. The Part-time Parliament.
    a. Lamport, Leslie. "The part-time parliament." *Concurrency: the Works of Leslie Lamport*. 2019. 277-317.
5. Cicada.

    a. Lim, Hyeontaek, Michael Kaminsky, and David G. Andersen. "Cicada: Dependably fast multi-core in-memory transactions." Proceedings of the 2017 ACM International Conference on Management of Data. 2017.
6. COST.
    a. McSherry, Frank, Michael Isard, and Derek G. Murray. "Scalability! but at what {COST}?." *15th Workshop on Hot Topics in Operating Systems (HotOS XV)*. 2015.
7. Serializability.
    a. Papadimitriou, Christos H. "The serializability of concurrent database updates." *Journal of the ACM (JACM)* 26.4 (1979): 631-653.
8. CockroachDB.
    a. Taft, Rebecca, et al. "Cockroachdb: The resilient geo-distributed sql database." Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 2020.
9. Chain Replication.
    a. Van Renesse, Robbert, and Fred B. Schneider. "Chain Replication for Supporting High Throughput and Availability." *OSDI*. Vol. 4. No. 91–104. 2004.
10. Chiller.
    a. Zamanian, Erfan, et al. "Chiller: Contention-centric transaction execution and data partitioning for modern networks." *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2020.

Textbooks:

- Van Steen, Maarten, and Andrew S. Tanenbaum. *Distributed systems*. Leiden, The Netherlands: Maarten van Steen, 2017.
    - Tanenbaum textbook, 3rd edition