Minimizing Transaction Latency in Geo-replicated Data Stores

Divyakant Agrawal[1]

[1]Department of Computer Science
University of California at Santa Barbara (UCSB)
Santa Barbara – CA – USA

agrawal@cs.ucsb.edu

Web service providers, in general, use NoSQL data stores to ensure scalability and availability of distributed data at the cost of sacrificing transactional guarantees. Recently, major web service providers such as Google have moved towards building storage systems to ensure ACID transactional guarantees for globally distributed data. Spanner from Google uses two-phase locking with Paxos based transaction commitment to ensure consistency of data replicated over multiple datacenters and incurs a large number of cross-datacenter messages. We demonstrate that an alternative approach of replicating commit processing over multiple datacenters can reduce cross-datacenter communication substantially. A common characteristic of both Spanner and the Replicated Commit protocol is that the transaction latencies are subject to round-trip time (RTT) for message communication between the datacenters. We next develop a technique based on log propagation among the datacenters and show that the RTT barrier can be broken subject to certain constraints (e.g., by decoupling consistency and fault-tolerance). A natural question then arises if there is an inherent limit to reducing the commit latency? We answer this question by deriving a lower-bound on commit latency. The sum of the commit latency of any two datacenters is at least the RTT between them. We use the insights and lessons learned while deriving the lower-bound to develop a commit protocol, called Helios, that achieves low commit latencies. Helios actively exchanges transaction logs (history) among the datacenters. The received logs are used to determine whether a transaction can commit or not. The earliest point in the received logs that is needed to commit a transaction is decided by Helios to ensure low commit latency. Also, in a real-world deployment on five datacenters, Helios achieves a commit latency that is close to the optimal.

About the Autor

Divyakant Agrawal is a Professor of Computer Science at the University of California at Santa Barbara. His research expertise is in the areas of database systems, distributed computing, data warehousing, and large-scale information systems. Divy Agrawal has had visiting appointments at IBM Research, NEC Research, ASK.com, Qatar Computing Research Institute, National University of Singapore, and at Google Inc. He has published more than 350 research manuscripts in various forums (journals, conferences, symposia, and workshops) on wide range of topics related to data management and distributed systems and has advised more than 35 Doctoral students during his academic career. He received the 2011 Outstanding Graduate Mentor Award from the Academic Senate at UC Santa Barbara. In 2015, Divy Agrawal and his co-authors received the International Conference on Database Theory Test-of-Time award. His current interests

are in the area of scalable data management and data analysis in Cloud Computing environments, security and privacy of data in the cloud, and scalable analytics over big data. Divy Agrawal is an ACM Distinguished Scientist (2010), an ACM Fellow (2012), and an IEEE Fellow (2012).