Thesis abstract

As demand for big data analytics grows every day, companies have become aware of the critical role of real-time data-driven decision making to gain a competitive edge. This creates a challenge for companies needing to accelerate (fine-grained) access to massively distributed data, in particular, those dealing with online services. NoSQL systems, in particular distributed in-memory key-value stores, are vital in accelerating memory-centric and disk-centric distributed systems.

A *distributed key-value* store offers a flexible data model with more performant but weaker consistency to partition data across many nodes on a computer network. Starting in mid-2000, numerous commercial key-value stores have emerged, each with its own unique characteristics, such as Google Bigtable, Amazon Dynamo, and Facebook Cassandra to enable managing massively distributed data at unseen scale, which simply was not feasible with traditional relational database systems running on commodity hardware. These systems have become critical for large-scale applications, such as social networks, real-time processing, and recommendation engines to achieve higher performance.

Distributed systems are commonly built under the assumption that the network is the primary bottleneck, however this assumption no longer holds by emerging high-performance protocols in datacenters. Designing distributed applications over such protocols requires a fundamental rethinking in communication components in comparison with traditional protocols (i.e., TCP/IP). Much research has been carried out in order to improve the communication performance either by optimizing the existing protocol or inventing new communication standards.

A great deal of work on high-performance communication has led to modern high-speed networks including InfiniBand, RoCE, and iWARP, which support Remote Direct Memory Access (RDMA). RDMA blurs the boundary of each machine by creating a virtual distributed, shared memory among connected nodes, i.e., substantially reducing communication and processing on the host machine. Through RDMAs, clients can now directly access remote memory without the need to invoke the NoSQL's traditional client-server model. This motivates the NoSQL community to invest in developing purely in-memory key-value stores with RDMA capability, such as HydraDB, Herd, Pilaf, DrTM, FaRM. RDMA capable protocol (i.e., InfiniBand) supports legacy socket applications through *IP over InfiniBand* (IPOIB); however, running existing in-memory systems on top of it can not efficiently exploit the benefits in the infrastructure. So ex-

isting in-memory key-value stores strive to reduce latency and achieve higher performance by exploiting RDMA operations. In this thesis, commonly used underlying structure and data concurrency in RDMA-enabled in-memory key-value store are discussed. Furthermore, performance challenges of the RDMA operations have been investigated. State of the art are reviewed and evaluated based on the achieved knowledge on the RDMA operation. Finally, a novel in-memory key-value store is presented and evaluated in comparison with the state of the art.