CS554 Project Ideas

ZHT:YCSB - Yahoo Cloud Serving Benchmark on ZHT

Overview

A NoSQL database provides a mechanism for storage and retrieval of data that uses looser consistency models than traditional relational databases. Motivations for this approach include simplicity of design, horizontal scaling and finer control over availability. NoSQL databases are often highly optimized key–value stores intended for simple retrieval and appending operations, with the goal being significant performance benefits in terms of latency and throughput. NoSQL databases are finding significant and growing industry use in big data and real-time web applications. NoSQL systems are also referred to as "Not only SQL" to emphasize that they do in fact allow SQL-like query languages to be used. As an important building block for distributed systems, distributed key-value stores are widely used in many places. But from performance perspective, they do not equally perform. One of the actual industrial standard benchmark for NoSQL databases is YCSB (Yahoo Cloud Serving Benchmark). It now supports HBase, Cassandra, DynamoDB, HyperTable and many others, but not ZHT. In this project, you need to build an adaptor between YCSB and ZHT so that YCSB can be used to benchmark ZHT. This requires you modify ZHT API or YCSB or both. Note that YCSB is in Java and ZHT is in C++, so you will also have to handle the Java to C++ language barrier, which might be achieved by some third party cross-language tools like Apache Thrift.

Relevant Systems and Reading Material

- ZHT paper: http://datasys.cs.iit.edu/projects/ZHT/ZHT-CRC-PID2666213-Final.pdf
- Project URL: http://datasys.cs.iit.edu/projects/ZHT/index.html
- NoSQL on Wikipedia: http://en.wikipedia.org/wiki/NoSQL
- YCSB source code: https://github.com/brianfrankcooper/YCSB/
- YCSB project page: http://labs.yahoo.com/news/yahoo-cloud-serving-benchmark/
- Apache Thrift : https://thrift.apache.org/

Methodology

Design and implement APIs between ZHT and YCSB so to benchmark ZHT with YCSB.

Preferred/Required Skills

Required: Linux, C++ and Java

Preferred: Any scripting language that you are comfortable with to allow you conduct the experiments, such as Shell.

Evaluation and Metrics

Functionality: run YCSB benchmark for ZHT.

Benchmark results: Operation latency and its distribution, throughput, scalability, CDF graph, for ZHT, HBase, Cassandra, DynamoDB, and HyperTable on Amazon EC2, on up to 128 VMs.

Project Mentor

Tonglin Li, tli13@hawk.iit.edu, https://sites.google.com/site/tonglinlihome/