CS554 Project Ideas

ZHT:Bench - Benchmarking mainstream NoSQL databases

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. In this project, you will benchmark some mainstream key-value stores (such as Riak) and document databases (such as MongoDB, Couchbase), and compare to ZHT (a distributed key-value store project developed in the Datasys Lab at IIT). You will evaluate these NoSQL databases on Amazon EC2 at scales up to 1K nodes.

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

Methodology

You will evaluate the performance of these key-value stores on different test beds, cluster, supercomputer and cloud.

Preferred/Required Skills

Required: Linux, Shell scripting

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

Evaluation and Metrics

Operation latency and its distribution, throughput, scalability, CDF graph

Project Mentor

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