CS554 Project Ideas

ZHT:OHT – Hierarchical Distributed Hash Tables

Overview

ZHT is a zero-hop distributed hash table, which has been tuned for the requirements of high-end computing systems. ZHT aims to be a building block for future distributed systems, such as parallel and distributed file systems, distributed job management systems, and parallel programming systems. The goals of ZHT are delivering high availability, good fault tolerance, high throughput, and low latencies, at extreme scales of millions of nodes. At this moment ZHT have achieved most of these goals. But it still has limitation. With n-to-n communication, all nodes have to maintain a large number of socket connections which is quite expensive when the system scales into extreme, such 1 million nodes. In some environment which users don't have root permission so are not able to change system parameters such as ulimit and maximum number of open file descriptors, concurrent connection will be limited to around 1024, and so would scale be limited.

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

Methodology

A practical approach to address this issue is hierarchical architecture, in which a layer of I/O nodes are added. Each I/O node serve a fixed set of compute node, and forward the request to other I/O node if needed. In this manner 1 million node scale can be easily achieved by using 1000 I/O nodes and each serves 1000 compute nodes.

Preferred/Required Skills

Required: Linux, C/C++ (no OOP skill needed)

Preferred: Shell scripting.

Evaluation and Metrics

Latency, throughput, scalability

Project Mentor

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