

# CS554 Project Ideas

---

## ZHT:DMHDFS - Distributed Metadata Management for the Hadoop File System

### Overview

The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In other words, all metadata requests are served by the NameNode. This centralized metadata structure will be a significant bottleneck when facing intensive metadata operations such as small files or frequent access. An intuitive approach is to distribute metadata across many servers. Your work is to explore the HDFS metadata service and figure out a solution to replace it with ZHT, a zero-hop distributed hash table.

### Relevant Systems and Reading Material

HDFS: [http://hadoop.apache.org/docs/stable/hdfs\\_design.html#Introduction](http://hadoop.apache.org/docs/stable/hdfs_design.html#Introduction)

[1] Shvachko, Konstantin and Kuang, Hairong and Radia, Sanjay and Chansler, Robert. The Hadoop Distributed File System, *IEEE 26th Symposium on Mass Storage Systems and Technologies*, 2000. Available online: <http://dl.acm.org/citation.cfm?id=1914427>

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>

### Preferred/Required Skills

Required: Linux, C/C++ (no OOP skill needed), Java, Hadoop and HDFS

Preferred: Shell scripting (for experiments)

### Evaluation and Metrics

Application running time, latency, throughput, scalability

### Project Mentor

Tonglin Li, [tli13@hawk.iit.edu](mailto:tli13@hawk.iit.edu), <https://sites.google.com/site/tonglinlihome/>

**Note: This project need solid background knowledge and strong programming skills, it will need 2 to 3 students contribute enough time.**