Overview on ZHT

Introduction to NoSQL databases and CS554 projects based on ZHT

Outlines

- General terms
- Overview to NoSQL dabases and key-value stores
- Introduction to ZHT
- CS554 projects

Databases/datastores

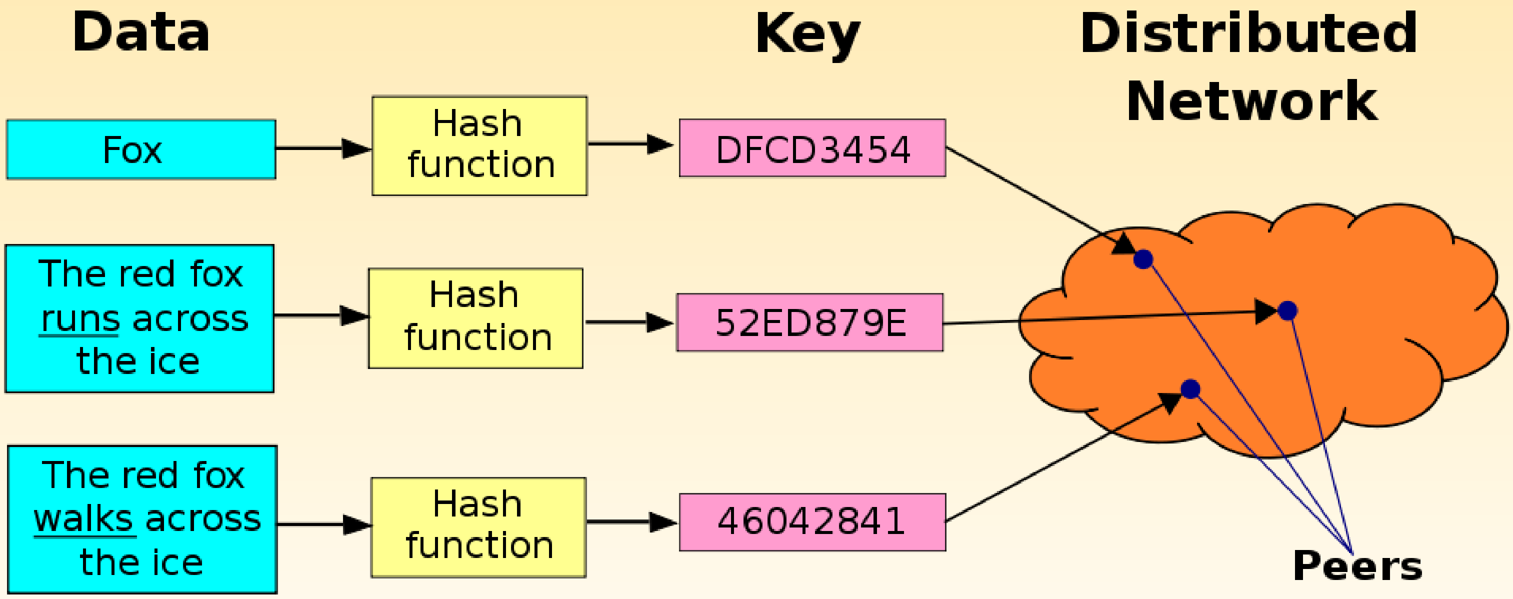
- Relational databases
  - Query with SQL
  - DB2, MySQL, Oracle, SQL Server
  - CS 425, 525
- NoSQL databses
  - Loose consistency model
  - Simpler design
  - High performance
  - Distributed design

Categories in NoSQL

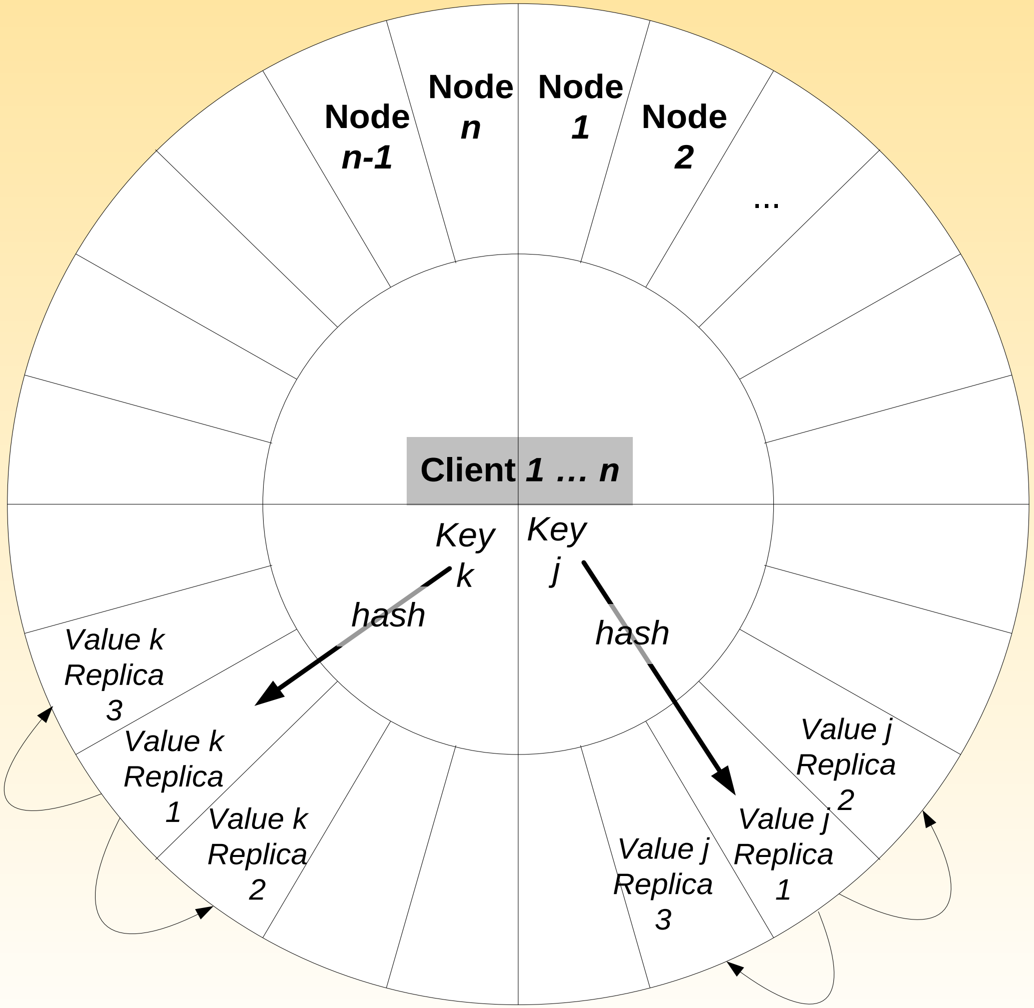
- Key-Value store
  - ZHT, Dynamo, Memcached, Cassandra, Chord
- Document Oriented Databases
  - MongoDB, Couchbase
- Graph databases
  - Neo4J, Allegro, Virtuoso

Key-value Stores

- Another name for Distributed Hash Table

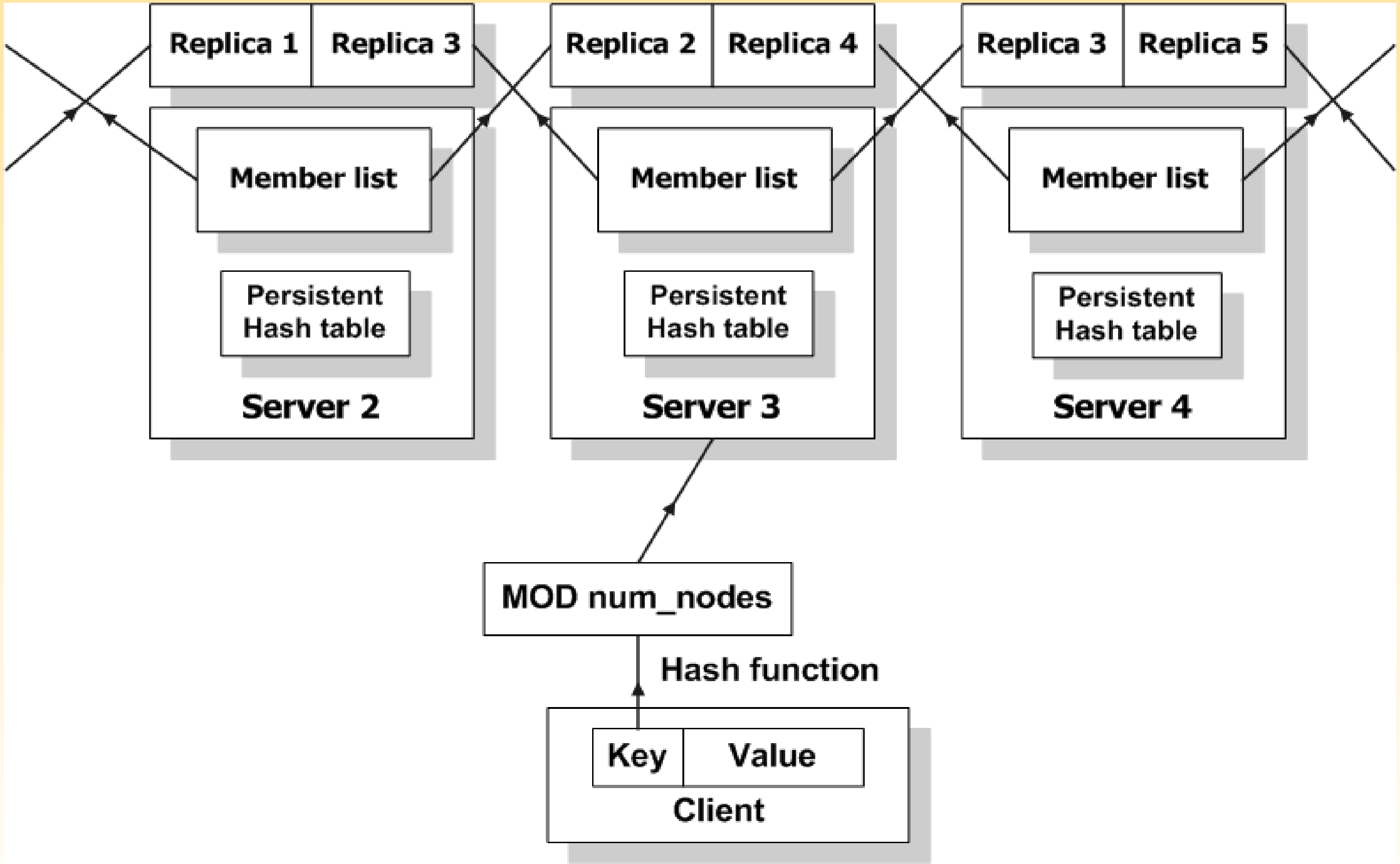


Zero-hop hash mapping



ILLINOIS INSTITUTE OF TECHNOLOGY

2-layer hashing



ILLINOIS INSTITUTE OF TECHNOLOGY

Consistency

- Updating membership tables
  - Planed nodes join and leave: strong consistency
  - Nodes fail: eventual consistency
- Updating replicas
  - Configurable
  - Strong consistency: consistent, reliable
  - Eventual consistency: fast, availability

Related work: Distributed Hash Tables

- Many DHTs: Chord, Kademlia, Pastry, Cassandra, C-MPI, Memcached, Dynamo ...
- Why another?

Name	Impl.	Routing Time	Persistence	Dynamic membership	Append Operation
Cassandra	Java	Log(N)	Yes	Yes	No
C-MPI	C	Log(N)	No	No	No
Dynamo	Java	0 to Log(N)	Yes	Yes	No
Memcached	C	0	No	No	No
ZHT	C++	0 to 2	Yes	Yes	Yes

ILLINOIS INSTITUTE OF TECHNOLOGY

Related projects

- ZHT Bench: Benchmarking mainstream NoSQL databases
- ZHT Cons: Eventual consistency support for ZHT
- ZHT DMHDFS: Distributed Metadata Management for the Hadoop File System
- ZHT Graph: Design and implement a graph database on ZHT
- ZHT OHT: Hierarchical Distributed Hash Tables
- ZHT ZST: Enhance ZHT through Range Queries and Iterators

Evaluation: test beds

- IBM Blue Gene/P supercomputer
  - Up to 8192 nodes
  - 32768 instance deployed
- Commodity Cluster
  - Up to 64 node
- Amazon EC2
  - M1.medium and Cc2.8xlarge
  - 96 VMs, 768 ZHT instances deployed

Genera requirements

- Familiar with Linux and it’s command line
- Shell scripting language (eg. Bash, zsh…)
- Programming skills in C++/C (except benchmark)
- GCC compiler
- No object oriented skill needed

ZHT Bench: Benchmarking mainstream NoSQL databases

- Goal: Extensively benchmarking NoSQL databases and analysis performance data.
- ZHT, MongoDB, Cassandra
- Neo4J (experiment for Graph)
- And others...
- Metrics
  - Latency and its distribution , throughput
- Parameters
  - Message size
  - Scales
  - Key Distributions

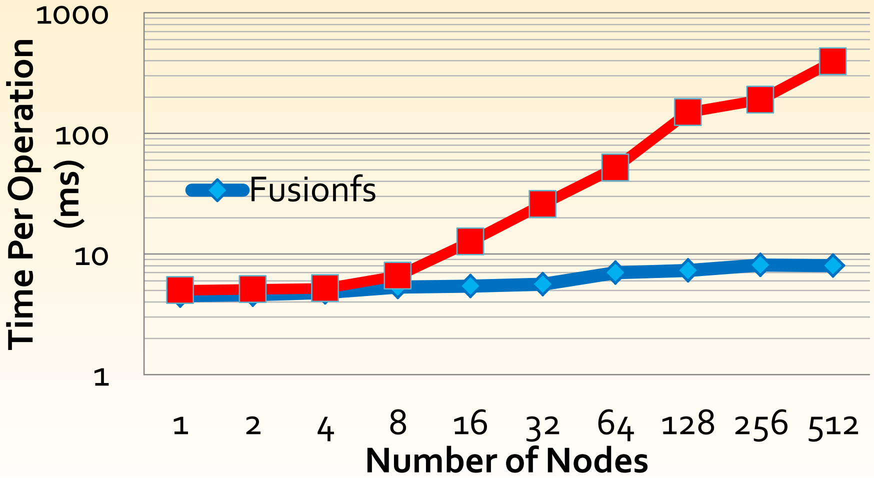
ZHT Cons: Eventual consistency support for ZHT

- Goal 1: allow replicas serve read operation
- Goal 2: maintain eventual consistency between replicas
- Goal 3: make it scale (pretty hard!)

- Optional goal: allow replicas serve write requests and maintain consistency (applying Paxos protocol, even harder)

ZHT DMHDFS: Distributed Metadata Management for the Hadoop File System

- What is metadata?
- Goal: improve HDFS performance by adding distributed metadata service
- Requirement: experience with Hadoop and HDFS; strong programming skill in both Java and C++



ILLINOIS INSTITUTE OF TECHNOLOGY

ZHT Graph: Design and implement a graph database on ZHT

- Goal: build a graph databases on top of ZHT
- How: construct a mapping from key-value store interface to graph interface

ZHT OHT: Hierarchical Distributed Hash Tables

- Goal: adding a proxy level to ZHT architecture so to reduce concurrency stress to each server
- Easy: make it work and scale
- Hard: handle failures

ZHT ZST: Enhance ZHT through Range Queries and Iterators

- Goal: design and implement new interface methods to ZHT
  - Iterator: next/previous operation
  - Range get/put: given a range of key, return a series of results in one request loop
- How?
  - Sorted map
  - B+ tree (bold!)

What do I expect?

- Communication: come and talk to me (by appointment)
- Make good use of Google
- Fail quick, fail early, fail cheap.
- Fast iteration: very small but frequent progress

- Why bother? 80% points from projects!

Welcome abroad and enjoy!

Tonglin Li

tli13@hawk.iit.edu

http://datasys.cs.iit.edu/projects/ZHT/

ILLINOIS INSTITUTE
OF TECHNOLOGY