CS554: DATA INTENSIVE COMPUTING

Benchmark: NoSQL Databases

MongoDB vs Riak vs ZHT

Abhilash Bhurse (A20404893)
Ashish Wagh (A20405623)
Suyog Kharage (A20402686)

INTRODUCTION

- With most scientific applications becoming data-centric, the effective management and analysis of large scale data has become a matter of prime importance.
- NoSQL systems have come up as possible solutions in this regard.
- MongoDB and Riak are two popular NoSQL systems, each offering a variety of different features.
- ZHT has emerged as new system that promises to be the foundational block for future distributed systems.

BACKGROUND AND MOTIVATION

- Traditional Relational Databases prove inefficient for dealing with data (structured or unstructured) of huge scale.
- This is where NoSQL databases come into picture.
- A NoSQL database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.
- These databases are widely used in Big-Data applications.
- NoSQL might also support SQL-like query languages as well.

PROBLEM STATEMENT

- There have been a variety of NoSQL systems that have emerged in recent years that employ different techniques for storage and analysis of largely unstructured data and each one claims to be better at some or the other aspect.
- MongoDB is Document Store system that supports field, range queries and regular expression searches, and uses Grid File System for File Storage.
- Riak is a key value store system that provides support for basic PUT, GET, POST and DELETE functions.
- A new system ZHT has emerged recently, which is a zero-hop distributed key value store system, and promises to be a building block for future distributed systems.
- This calls for a comparative study of ZHT with other popular systems such as MongoDB and Riak to establish the better system or mechanism in terms of efficiency, throughput, scalability and latency.

PROPOSED SOLUTION

- Through this project we aim to carry out an empirical performance evaluation of ZHT in comparison to MongoDB and Riak and present the results based on metrics such as latency, throughput, efficiency and scalability.

- We also aim to study and compare similar features of these systems and propose the better system for specific real world applications such as **Hospital Record Maintenance System**.

- We would mainly evaluate these systems based on the basic operations supported by each system, such as insert, lookup and remove in ZHT, Create, Read and Delete in MongoDB and PUT, GET, DELETE in Riak and compare the results.

- Datasets would be roughly in the range of 10K to 150K records.

CHALLENGES

- Setting up the all the database on the chameleon cloud environment
- Getting ZHT up and running.

MONGODB DATABASE

Classified as a NoSQL database program

MongoDB is a distributed database at its core.

A document database which stores data in JSON-like documents where fields can vary and structure can be changed at any time.

Provides features like Ad hoc queries, indexing, and real time aggregation to analyze the data.

MongoDB can be used as a file system with load balancing and data replication features over multiple machines for storing files.

RIAK DATABASE

Distributed NoSQL key-value data store

Databases is built on a set of core services providing a highly reliable, scalable distributed systems framework

Written in Erlang programming language

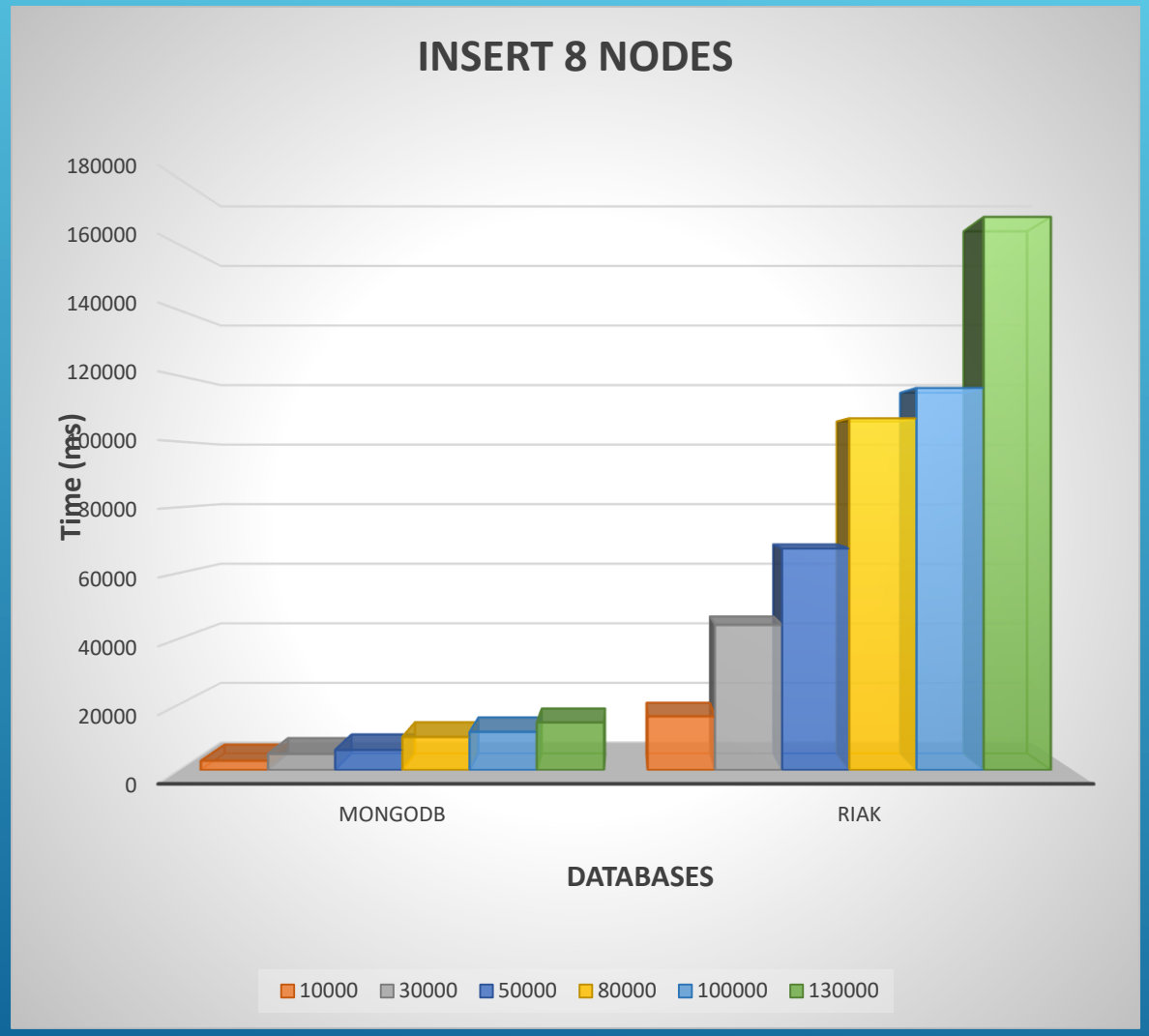
implements the principles from Amazon's Dynamo

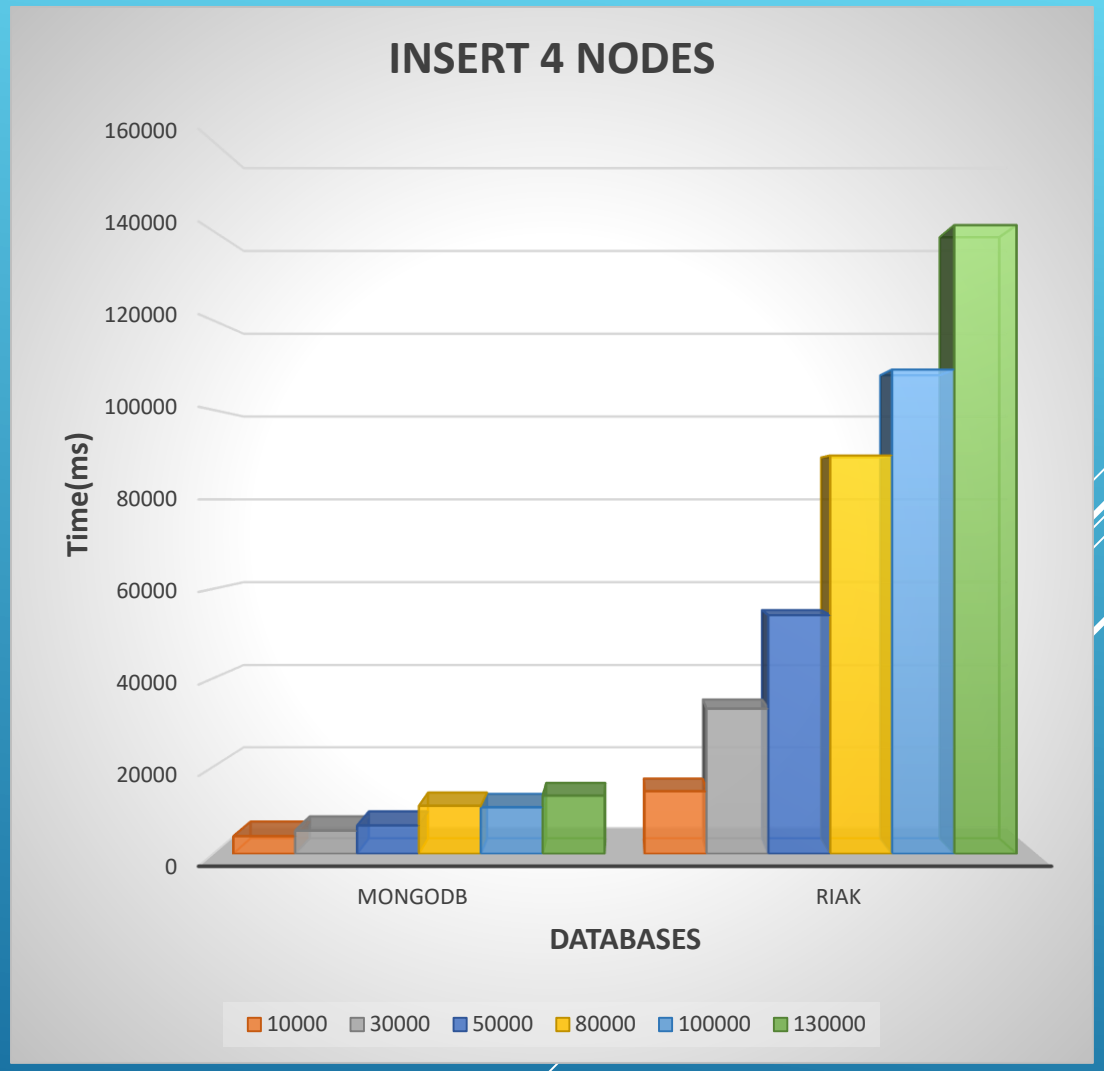
Provides features like Fault-tolerant availability, Predictable latency, Storage options and Multi-datacenter replication

OPERATIONS FOR EVALUATION

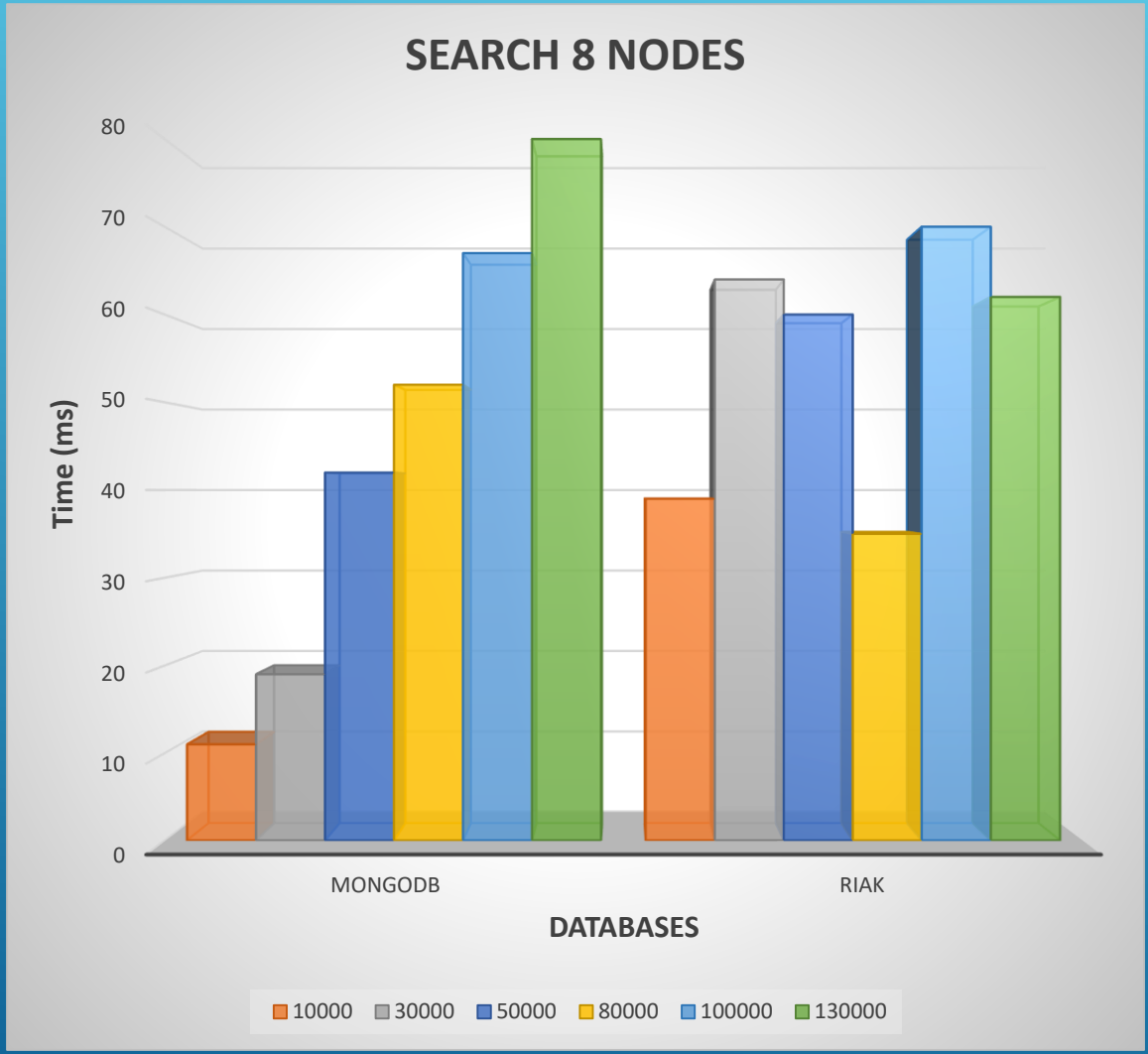
- Insert operation
- Search operation
- Update operation
- Delete operation

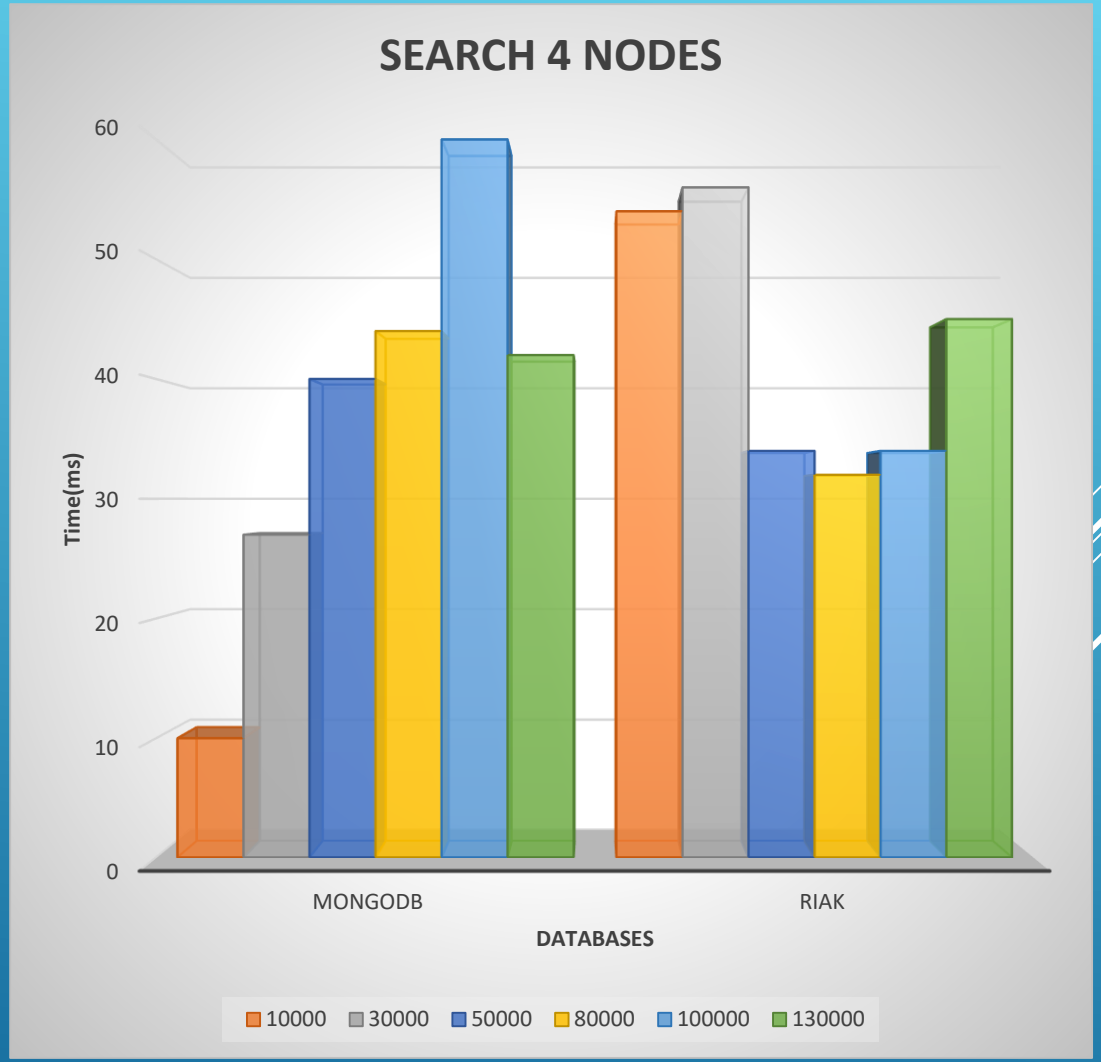
INSERT EVALUATIONS



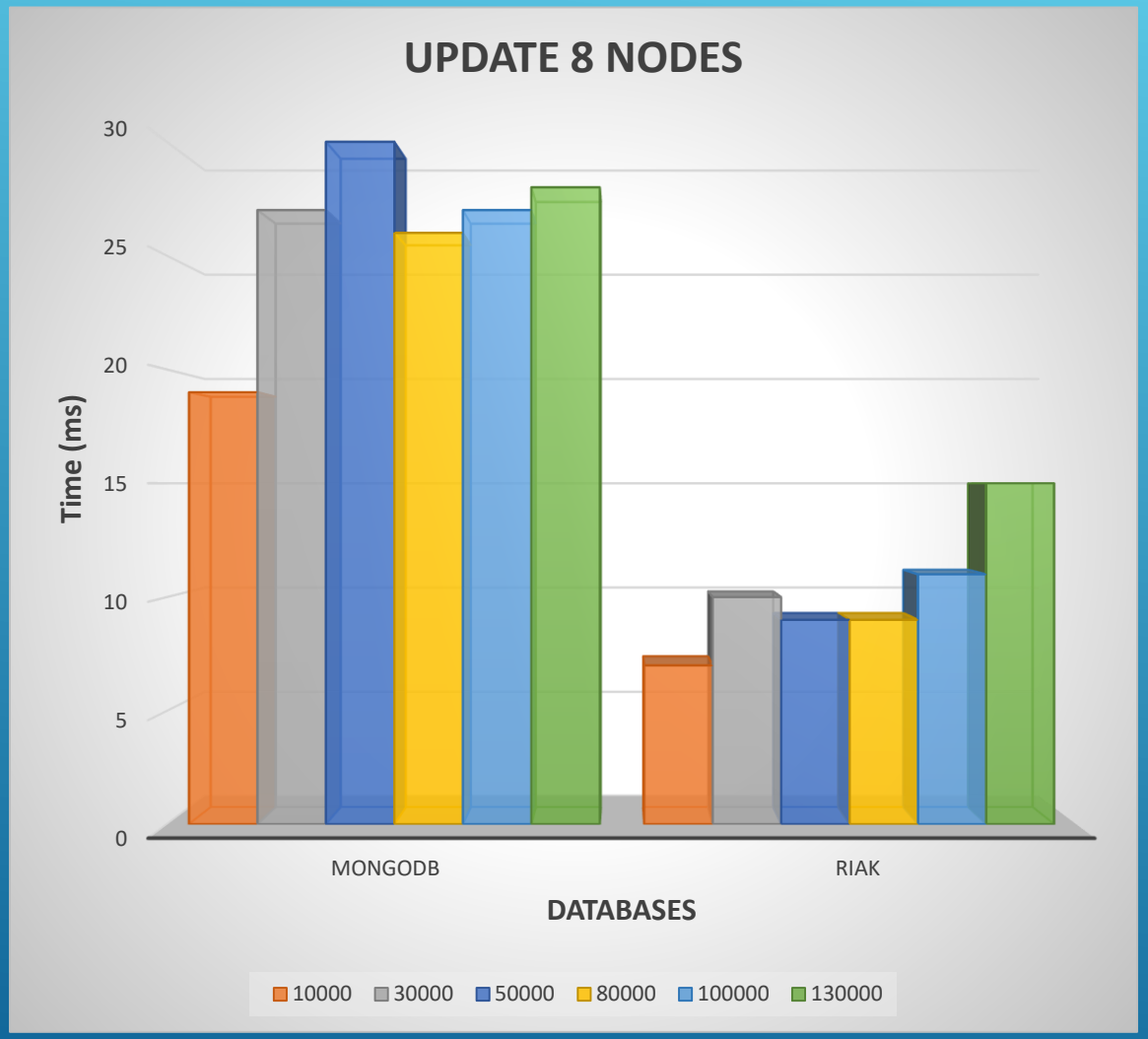


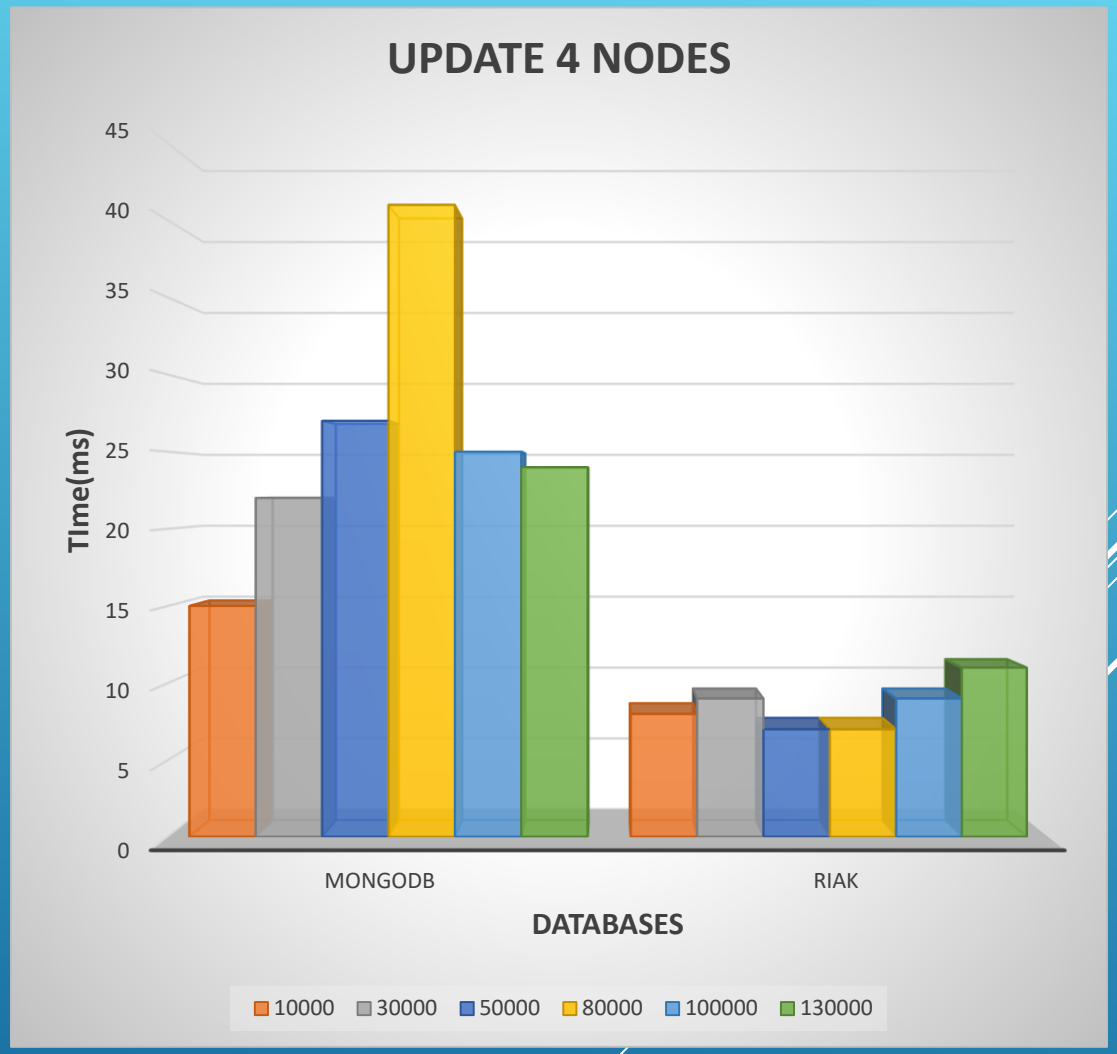
SEARCH EVALUATIONS



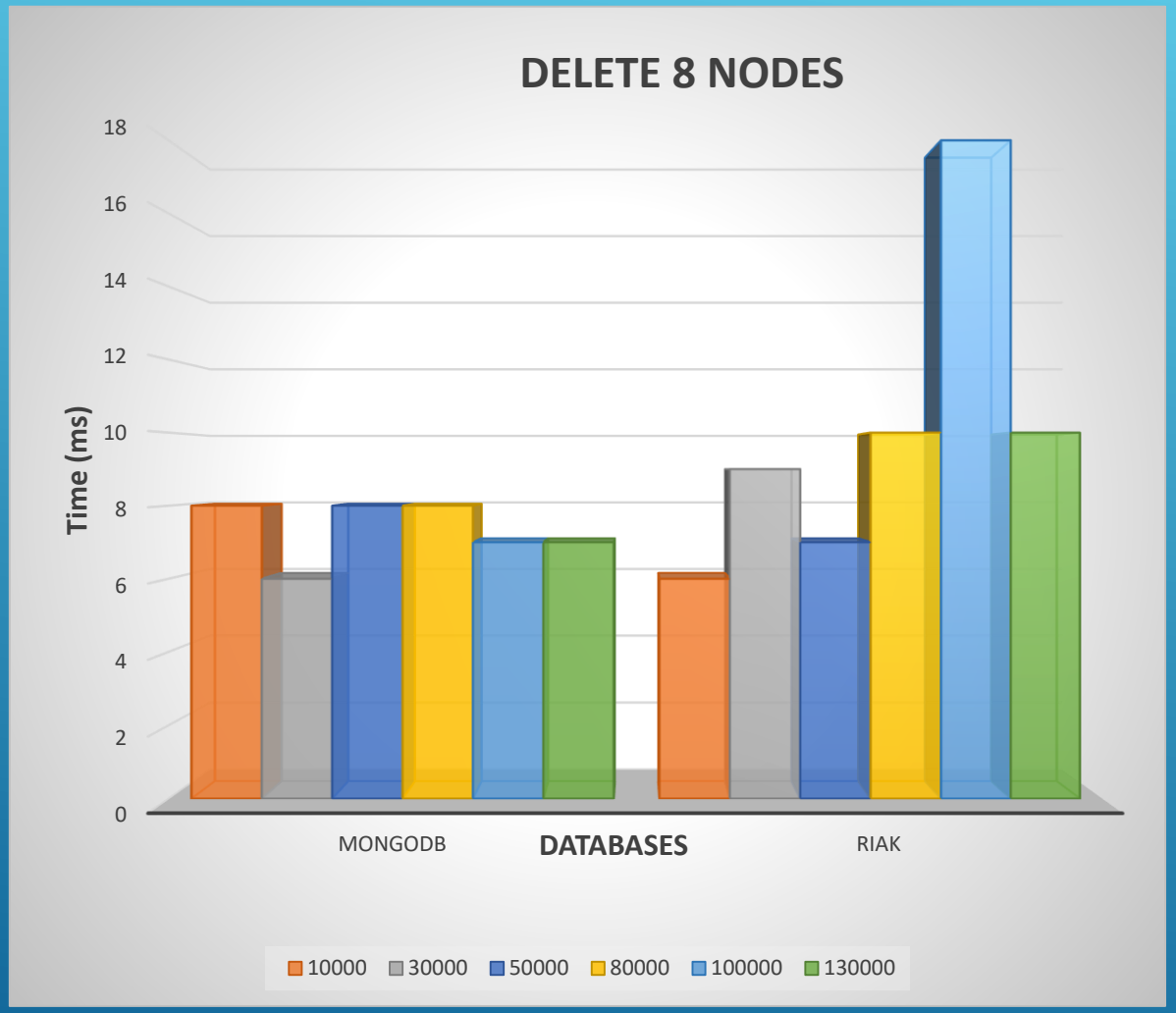


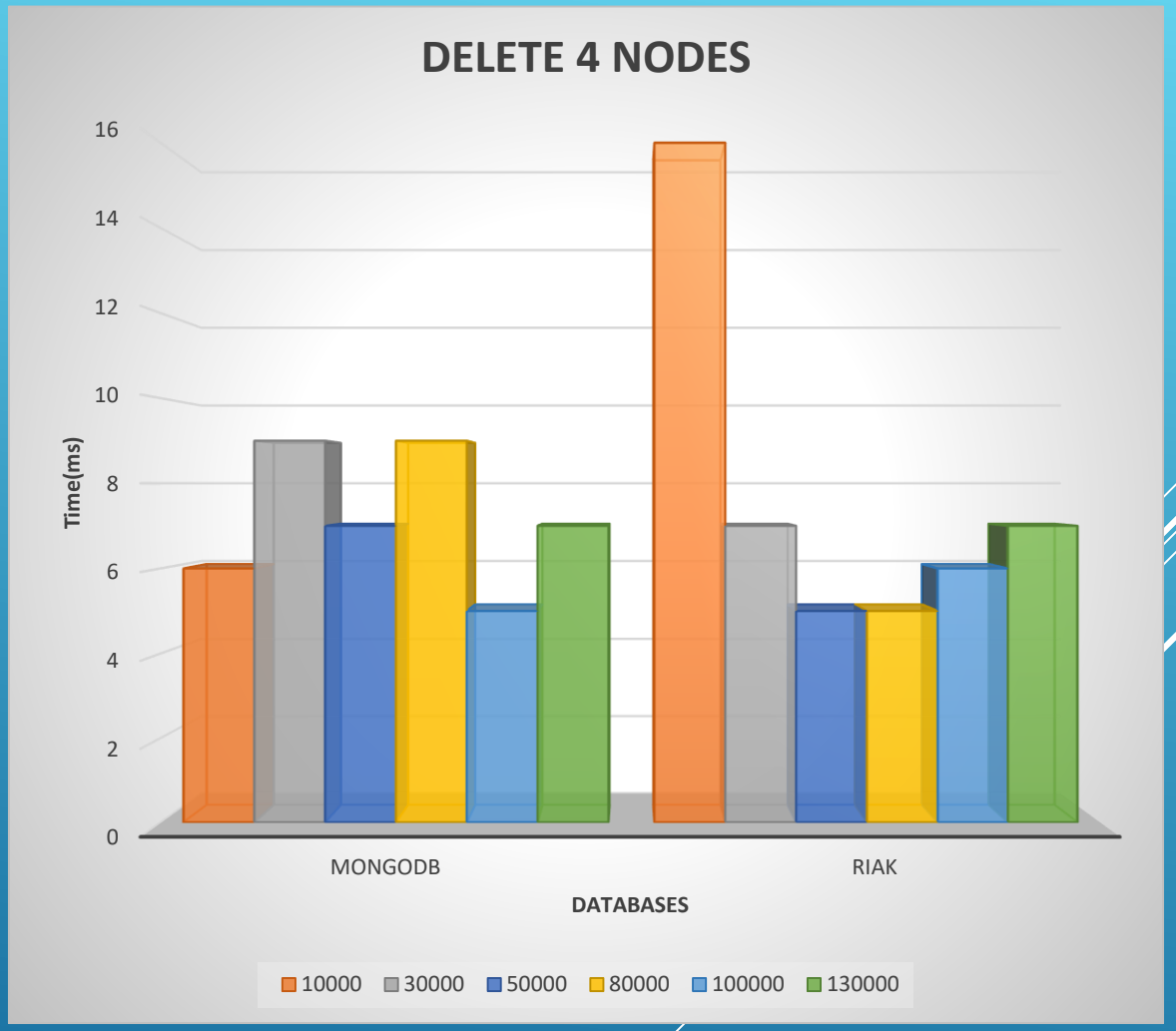
UPDATE EVALUATION



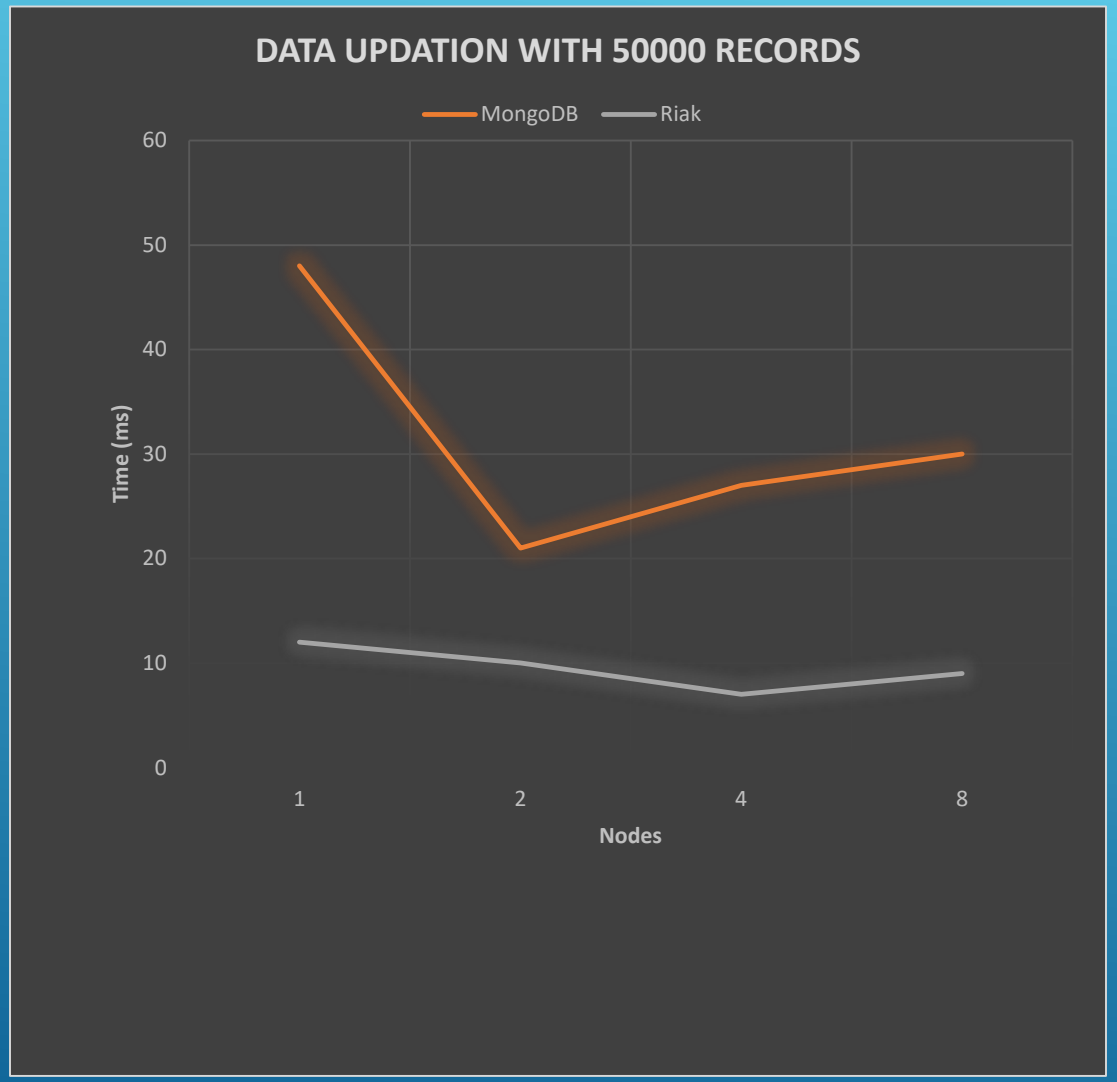


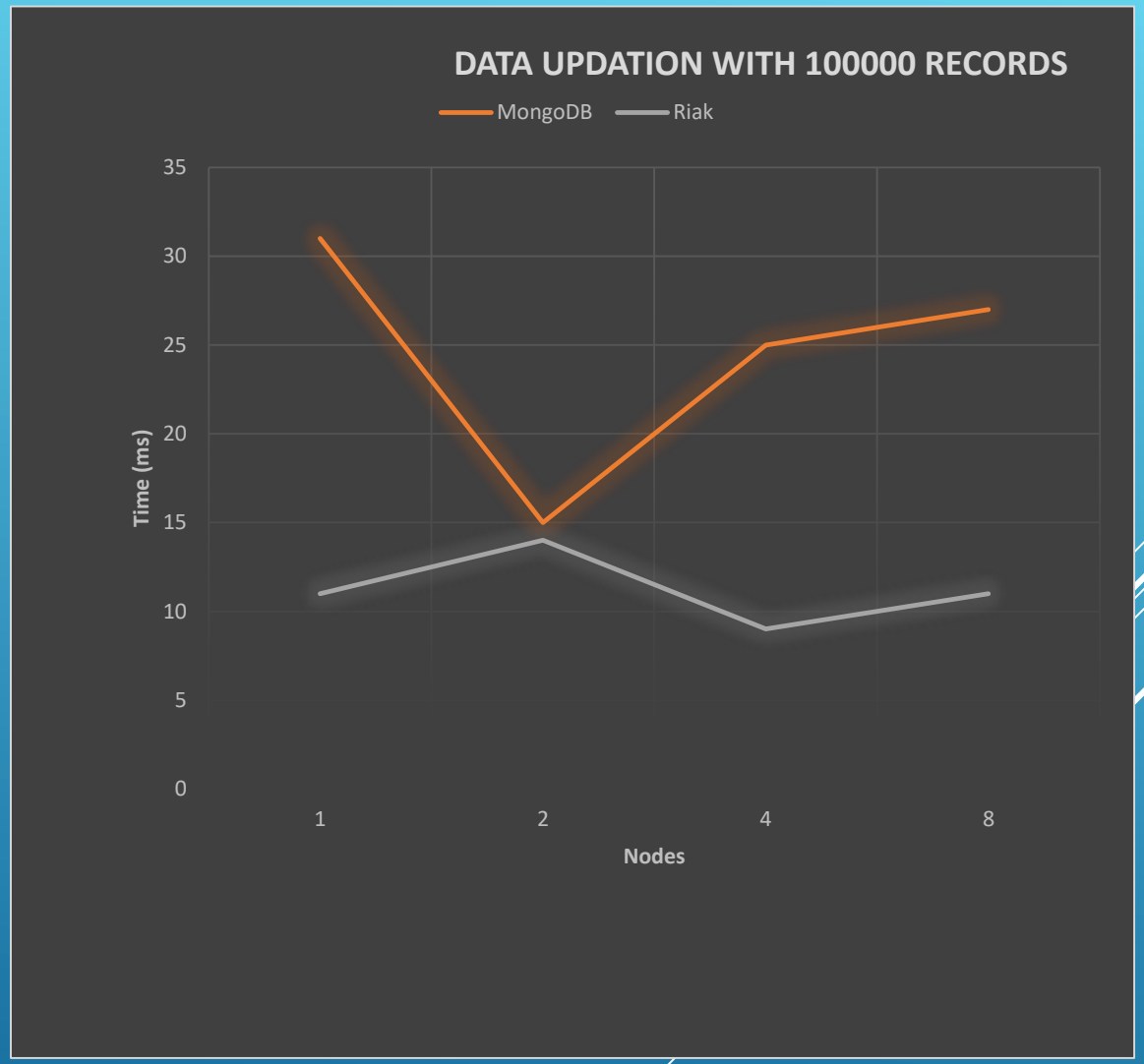
DELETE EVALUATIONS



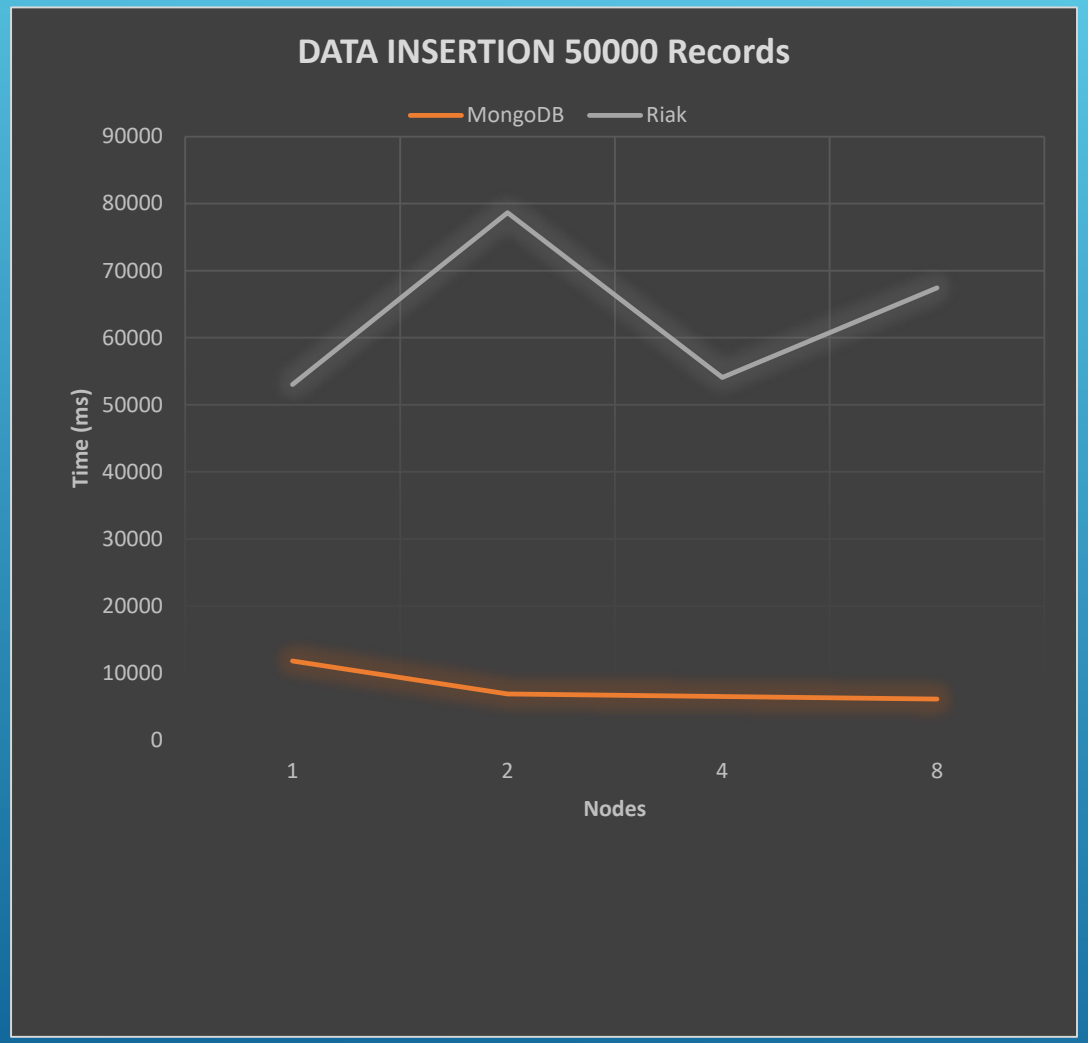


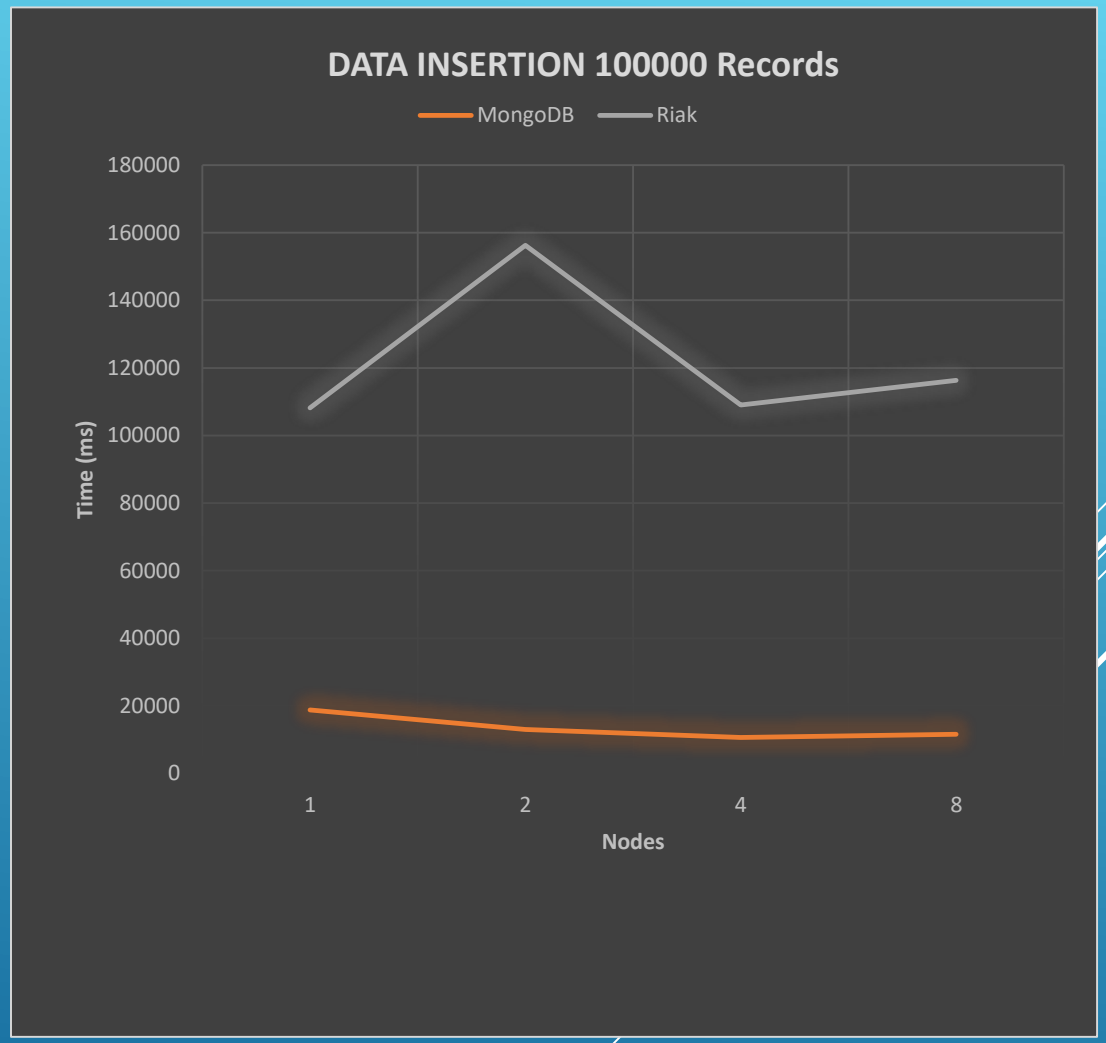
SCALABILITY (UPDATE)



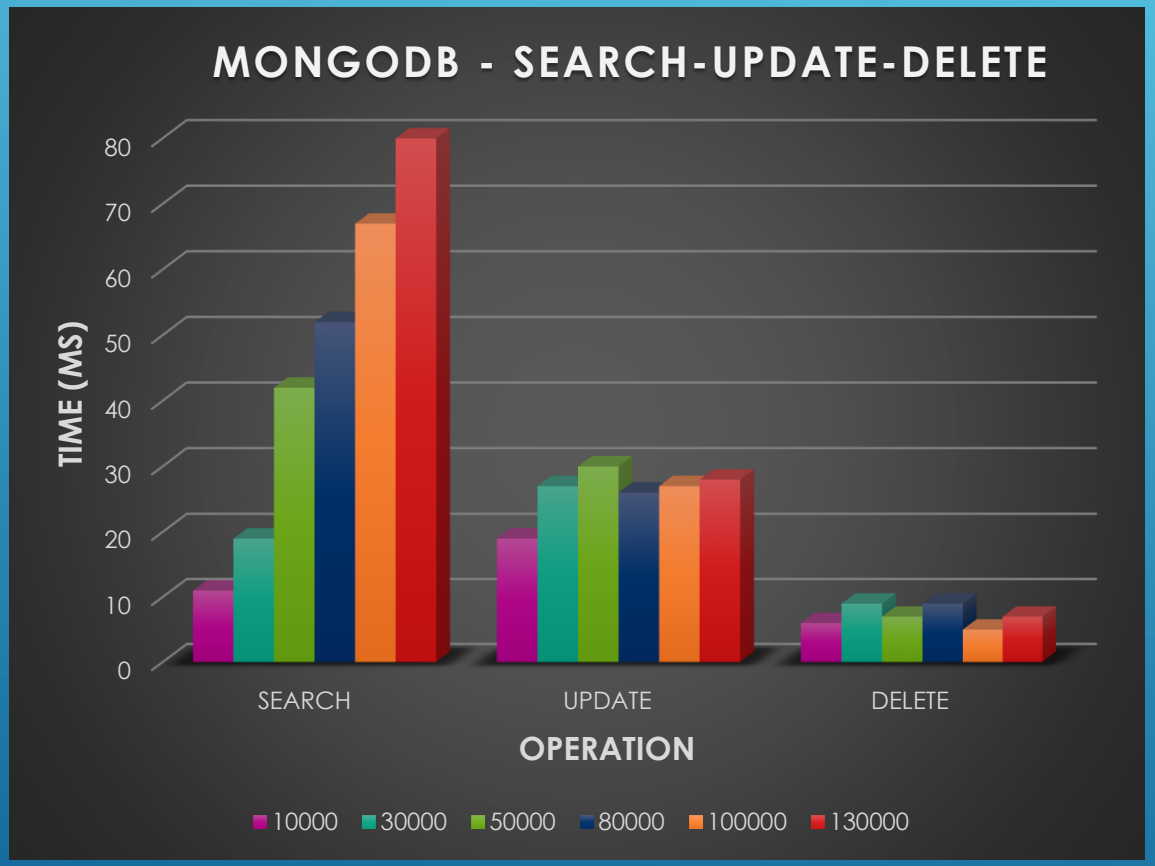


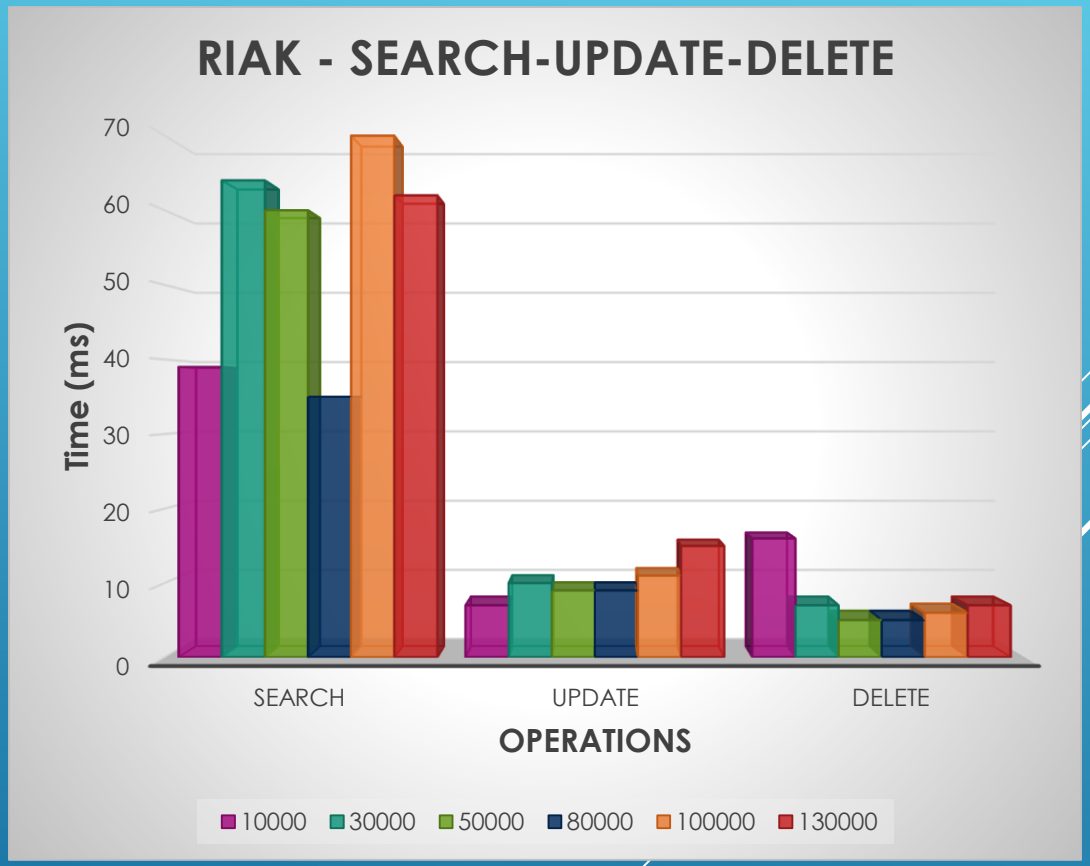
SCALABILITY (INSERT)





SERACH UPDATE DELETE COMAPRISON





CONCLUSION

Performance of MongoDB for Data Insertion is clearly much better than that of Riak. (appox 10x faster)

Although Riak performs better than MongoDB for Update and Search (appox 1.5x faster), the difference in performance is not too huge.

In future, if Hospital Systems (or other similar systems) decide to migrate to NoSQL Databases, MongoDB will prove to be a better bet as compared to Riak.

THANK YOU