Benchmarking Cassandra Using Yahoo's Cloud Serving Benchmark (YCSB).

December 17th, 2021

- Shivali Dalmia

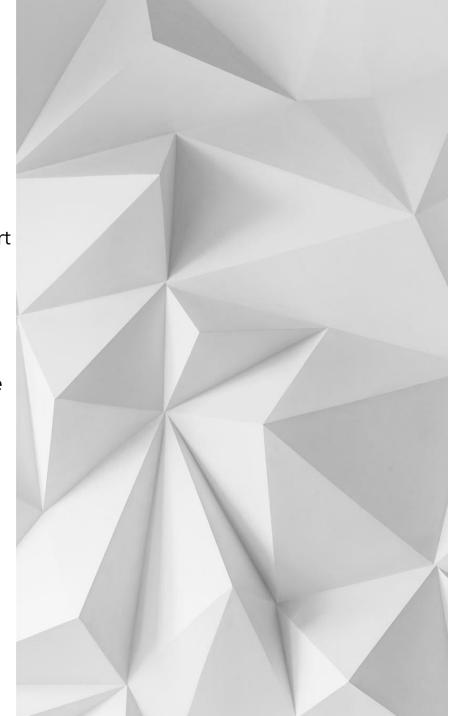# PROJECT WORKITEMS

➢ Difference between NoSQL and traditional SQL databases.

➢ Architecture of Cassandra NoSQL database.

➢ Analyzing performance of Cassandra database for different types of database operations using YCSB workloads.

➢ Experimental setup.

  – YCSB installation

  – Cassandra installation

  – Workload execution

➢ Comparison of throughput and latency for different types of workloads.

INSPIRATION

➢ To understand –
  - How distributed and non-distributed databases perform for read, insert and update operations.
  - How to select a database for different business use cases.

➢ To learn a new NoSQL database such as Cassandra.

➢ Key reference paper : Distributed NoSQL Data Stores: Performance Analysis and a Case Study by Abdeltawab Hendawi, Jayant Gupta, Jiayi Liu. [Link]
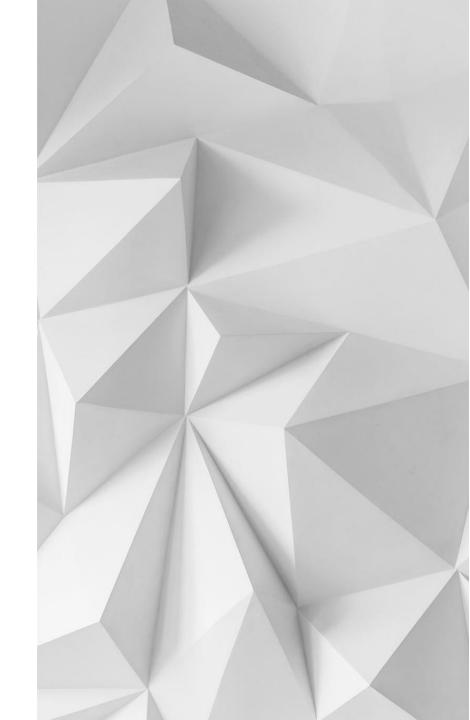
TOPICS FROM CLASS

➢ Difference between traditional SQL and NoSql databases.

➢ How CAP theorem applies to distributed and non distributed databases.

➢ Learned a new NoSql database Cassandra (e.g., architecture).

➢ How distributed databases perform on different database operations such as read, update, modify, write, delete.

➢ Role of performance metrics such as throughput and latency in analyzing different databases.

CHALLENGES

➢ Running a multi node cluster on a single windows machine.

➢ Setting up various multi node clusters with different configurations.

➢ Identifying appropriate sharding and replication factors.

FUTURE WORK

- ➢ Analyze the performance of Cassandra on:
    - – Real world datasets.
    - – Multi node cluster.
    - – Nested queries.
- ➢ Compare and contrast Cassandra with different NoSQL databases.

ACKNOWLEDGEMENT

- Thank you, Prof. Abe Kazemzadeh, for your guidance.

- Kindly reach me out for any questions and concerns.

- Please refer to the final draft for more details.