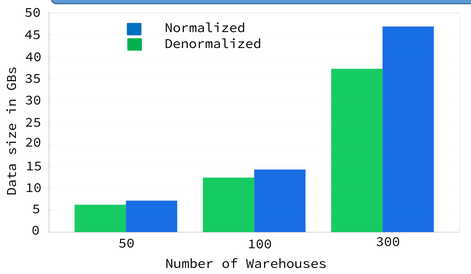
Benchmarking is HARD!

Comparing noSQL to RDBMS is extra hard: you must accept inherent differences in approaches

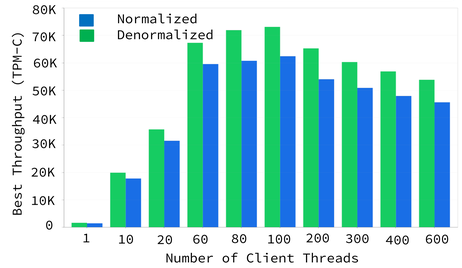
Adapting TPC-C Benchmark to Measure Performance of Multi-Document Transactions in mongoDB.

Asya Kamsky, MongoDB Inc.

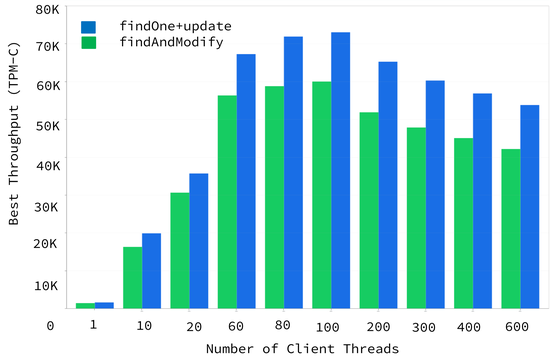
Is denormalized data larger than normalized data? It's not. Indexes are smaller too!



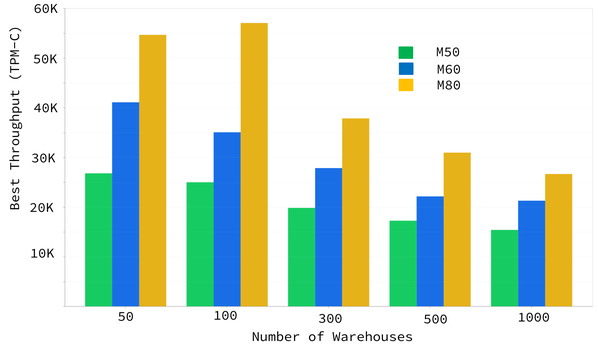
Denormalized schema performs better too



MongoDB has efficient operation to update+return same doc



Paying double for cloud resources doesn't double throughput



See the paper, the code & full results:

THE WHAT

MongoDB has more in common with traditional RDMBS systems than key-values stores it's lumped together with.

We wanted to test newly added multi-document transactions.

THE WHY

There are no meaningful accepted noSQL benchmarks that exercise the full set of capabilities that MongoDB supports.

YCSB is a simplistic key-value one-op at a time benchmark.

THE HOW

Adapted existing Python implementation of TPCC for noSQL. Made improvements and updates for best performance.

- Better denormalized document schema (no over-embedding)
- More appropriate (compound) secondary indexes
- Support MongoDB read/write preference/concern options
- Use MongoDB native operations
  i.e. `findAndModify` instead of `update+select`
- Measured & Compared everything

Conclusions

Good noSQL benchmarks should be created and discussed.

If you take a relational benchmark and run it "as is" you won't get good results. Adapting it to a different data model requires understanding how it works.

P.S. Everything ran in the cloud so you can reproduce results.

Use code VLDB2019 to get $150 credit to try MongoDB Atlas

mongoDB