

DBCOMP

A Database Performance Comparison

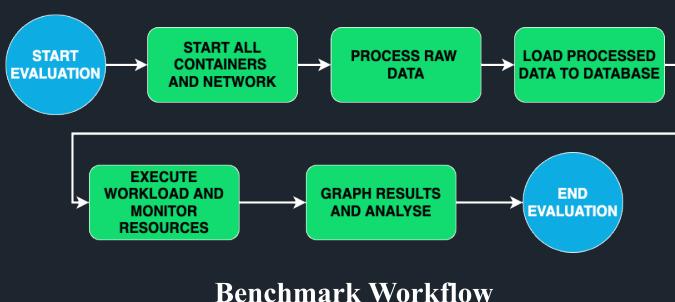


Aims

To compare the extent of latency and throughput performance of **MongoDB (document)** and **Cassandra (column)** against **MySQL (relational)** when using modest computing resources.

Context

The big data era has challenged the relational database model. NoSQL databases were developed to address these challenges and designed for use-case specific speed, in contrast to the relational model's focus on ACID properties. While NoSQL's applicability to big data is well noted, there is a lack of research comparing how these databases perform with modest resources.



System Design

Database benchmarks typically use CRUD operations to simplify performance comparisons over varying data and request sizes. This utilises the same data model for each database which is not suitable for our comparison.

A custom benchmarking tool was built, extending YCSB to utilise its metric tracking. Data transformation, loading and visualisations were done in Python, while Docker provided benchmark hygiene and a portable framework for reproducible results.

IoT - Cassandra

IoT data for a building's environmental and electrical sensors provides an environment for analytical work well-suited towards Cassandra. Benchmark queries were designed to target strengths and weaknesses of Cassandra and MySQL, which the results displayed.

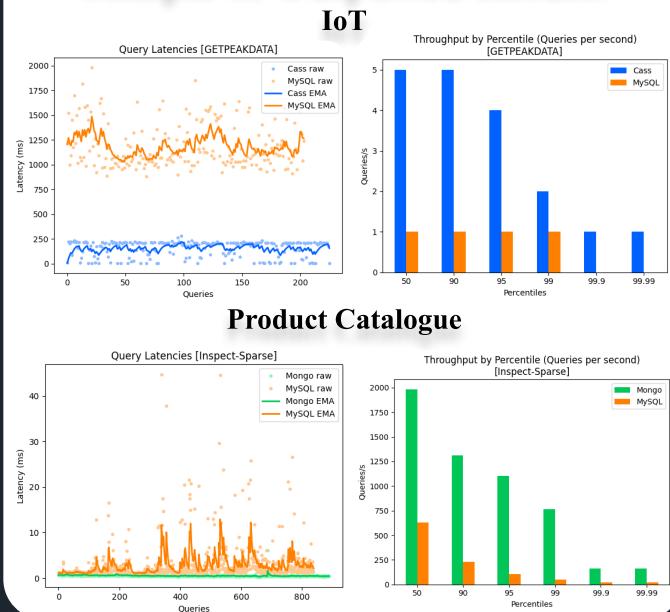
Cassandra was up to **79X** more performant than MySQL for queries targeting the column model.

Product Catalogue - MongoDB

MongoDB's flexible schema is well-suited to storing and retrieving data in this context. Queries were designed around two key use cases in online shopping, while product and review data scraped from Walmart.com formed the basis of the dataset used.

MongoDB was up to **25X** more performant than MySQL for queries targeting the document model.

Sample of Outputted Results



Conclusions

In most instances, the NoSQL databases achieved significantly better performance compared to the SQL implementations and consistently managed resource utilisation better. The less significant performance differences when SQL outperformed NoSQL, coupled with their poor resource utilisation, highlights the applicability of NoSQL to modestly resourced contexts. Similar extensions to YCSB can expand this research to more contexts and databases.



David Court
crtdav015@myuct.ac.za
Bradley Culligan
cllbra005@myuct.ac.za

Supervisor
(A/Prof) Sonia Berman
 National Research Foundation

Department of
Computer Science
University of
Cape Town

