INFO4990
Benchmark on Analytical Workload for NoSQL Systems

Presented by

[redacted]

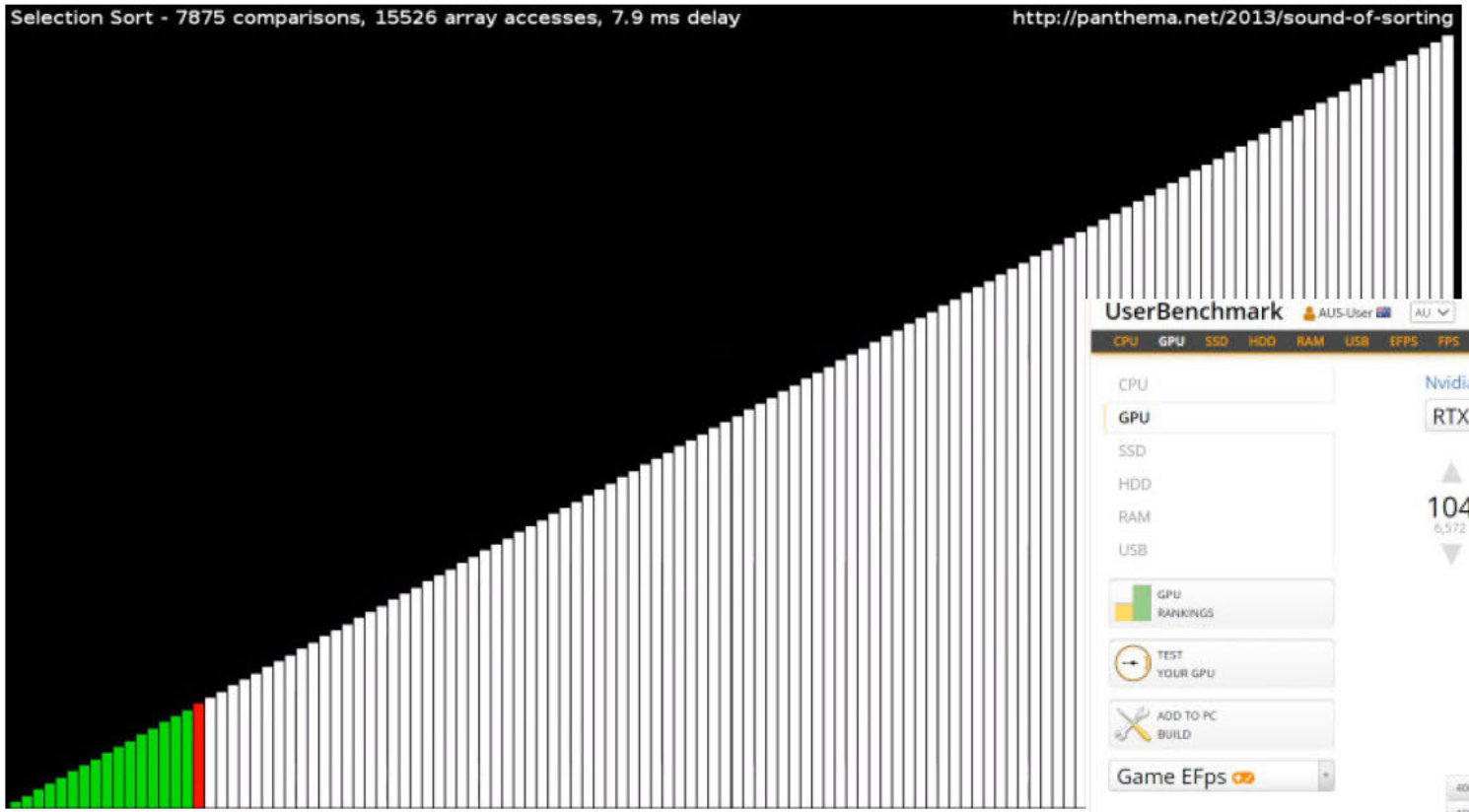
Supervised by

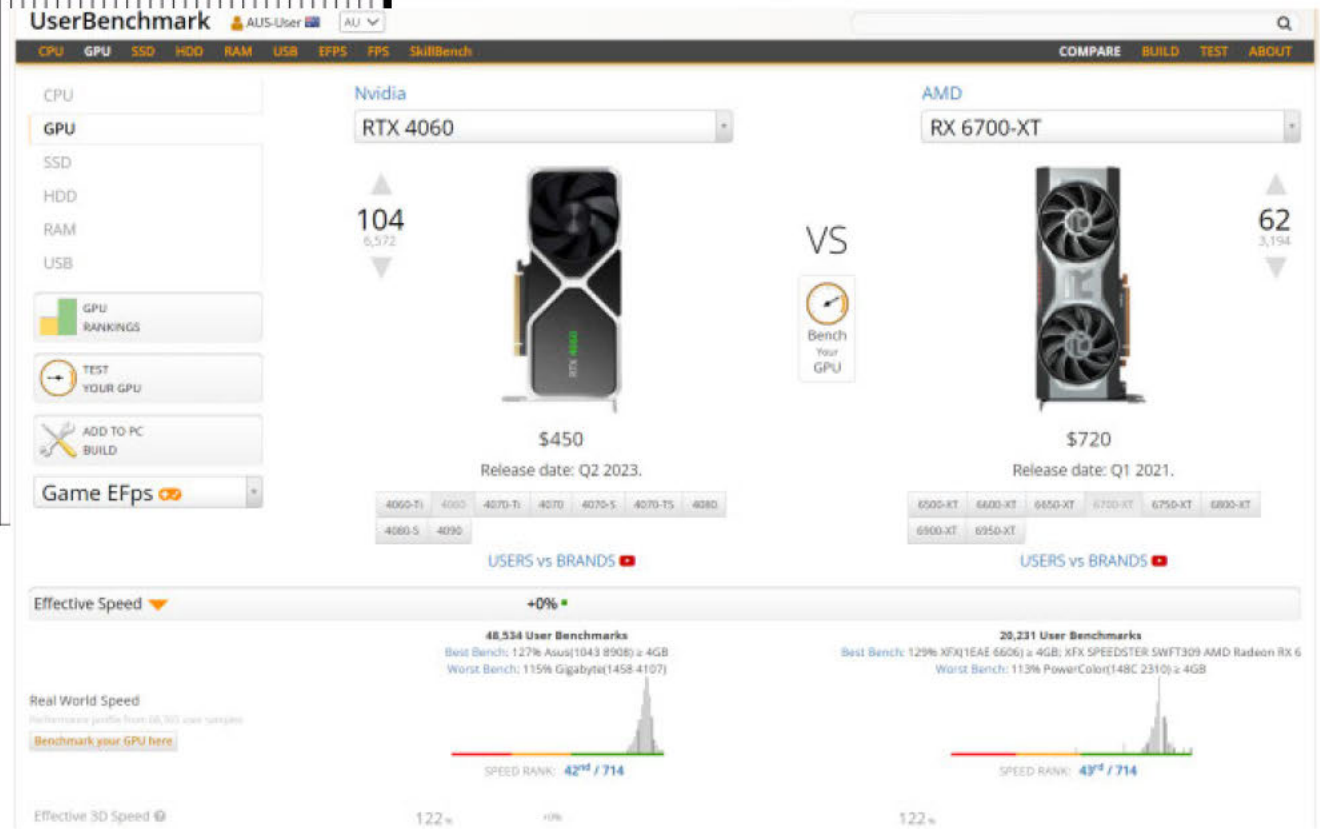
[redacted]

[redacted]

THE UNIVERSITY OF SYDNEY

What are Benchmarks?





NoSQL and Analytical Workload

NoSQL

- High performance, availability and scalability

Analytical Workload

- Complex queries running on large data sets
- Identify trends, relationships and insights
- Common workload in NoSQL as it requires high throughput and concurrency

Motivation

Why not a lot of NoSQL benchmarks?

- Long history of RDBMS vs 20 years of NoSQL existence
- Structured Schema vs NoSQL's diverse data models
- Hard to stay relevant

Objectives

Research Question:

How to design a Benchmark on Analytical Workload for NoSQL

- Objective
  - Develop Benchmark on Analytical Workloads for NoSQL
    - READ operations of Joins (Links), Filtering and Aggregation
    - Updates
  - Needs to at least provide:
    - data model (schema) – Collections of documents
    - data generator
    - query generator

Benchmark Evaluation Aspect Comparisons

Benchmark Name	Throughput	Scalability	Freshness	Query Complexity	Workload Customisation
TPC-C [1]	□	X	X	Low	X
TPC-H [2]	□	X	X	Medium	X
CH2 [9]	□	□	X	Medium	□
HTAPBench [10]	□	□	□	Medium	□
BigDataBench [11]	□	□	X	High	□
DeepBench [12]	□	□	X	High	□
CNSSB [13]	X	□	X	Low	X

Benchmark Requirements

- Common Use Cases for NoSQL – real-time analytics & continuous monitoring

Our Benchmark:

- Provides Data model, Data & Query Generator
- Tests READ and UPDATE operations
  - READ: complex analytical queries involving links, aggregations and filters
- Tests:
  - Latency: Query Response Time, 95th percentile latency
  - Scalability: Concurrency
  - Resource Utilisation: CPU, memory, I/O utilisation
  - Data Freshness: Time taken from UPDATE to query-able
  - Different Data Size: Load Time, QueryResponseTime/DataSize

System Research Method (Design > Implement > Evaluate)

Consulted professionals,
Identify
- Workloads
- Datasets
- Queries

Via Open-Source Projects

Run Benchmark on systems
- Test different query
- Test optimisation plans
- Test indexing technique
- Evaluate Query Advisor

Literature Review
- Identify NoSQL systems' common use cases
- Identify current benchmark limitations

Implementation (Build Benchmark)
- JSON doc store
- And then expand

Finish Research

References

[1] Transaction Processing Performance Council (TPC), TPC BENCHMARK™ C Standard Specification Revision 5.11, 2010

[2] Transaction Processing Performance Council (TPC), TPC BENCHMARK™ H (Decision Support) Standard Specification Revision 3.0.1, 2022

[3] DeWitt, Dave et al. Standard Benchmarks for Database Systems, Industrial Session 5, SIGMOD 1997, 1997

[4] Kamsky, Asya et al. Adapting TPC-C Benchmark to Measure Performance of Multi-Document Transactions in MongoDB, PVLDB 12(12): 2254-2262, 2019.

[5] apavlo. https://github.com/apavlo/py-tpcc/wiki.

[6] Serlin, Omri ”The History of Debitcredit and the TPC”. The benchmark handbook for database and transaction processing systems. M. Kaufmann Publishers. pp. 19–38, 1991

[7] Rutishauser, Nico TPC-H applied to MongoDB: How a NoSQL database performs, University Zurich, 2012

[8] Dagstuhl “Robust Query Processing” Breakout Group “Workload Management”, The mixed workload CH-benCHmark, DBTest ’11: Proceedings of the Fourth International Workshop on Testing Database Systems, Article No 8, pp.1-6, 2011

[9] Michael Carey, et al., CH2: A Hybrid Operational/Analytical Processing Benchmark for NoSQL, TPCTC 2021, pp. 62-80, 2022

[10] Coelho, F´abio et al. HTAPBench: Hybrid Transactional and Analytical Processing Benchmark, ACM ICPE’17, 2017.

[11] Jianfeng Zhan, et al. BigDataBench: A Scalable and Unified Big Data and AI Benchmark Suite, 2018

[12] Belloni, Stefano et al. DeepBench - Benchmarking JSON Document Stores, 9th International Workshop of Testing Database Systems (DBTest’22), 2022

[13] Dehdouh, Khaled et al. Columnar NoSQL Star Schema Benchmark, G.A. (eds) Model and Data Engineering. MEDI 2014. Lecture Notes in Computer Science, vol 8748. Springer, Cham, 2014

[14] O’Neil, P. et al. The Star Schema Benchmark (SSB), 2009

[15] 15 Sorting Algorithms in 6 Minutes, Retrieved on 22nd April 2024 from https://www.youtube.com/watch?v=kPRA0W1kECg&ab_channel=TimoBingmann

[16] UserBenchmark, Nvidia RTX4060 vs AMD RX6700-XT, Retrieved on 22nd April 2024 from https://gpu.userbenchmark.com/Compare/Nvidia-RTX-4060-vs-AMD-RX-6700-XT/4150vs4109

Intended Contributions

Develop benchmark so that

- Provide more **NoSQL benchmarking Options**
- Evaluate Optimisation Techniques
- Evaluate Query Plans
  - E.g. Query Advisor
- Improve the state-of-the-art of NoSQL systems