

# MongoDB Vs Cassandra: Energy consumption Evaluation and Comparative studies towards Query Energy Efficiency

Subject proposed by Simon Pierre DEMBELE

Email: [simon.pierre.dembele@ut.ee](mailto:simon.pierre.dembele@ut.ee)

## 1. Context:

Nowadays, energy efficiency (EE) is becoming more and more critical, even a selling point in computer systems. Many efforts have been made at different levels to increase EE including as selection of energy efficient chemical components, optimal design of logic circuits, and efficient design of algorithms as well as definition of techniques to manage the trade-off between energy and performance, etc.

Nevertheless, current academic research and industrial practices on databases (DB) focus more on performance than on energy efficiency. As the consensus is to build a greener world, the design and operation of DB systems must be more sensitive to the energy dimension. Some academic work has attempted to address this objective by proposing less energy-intensive solutions. In the latter categories, mathematical models estimating the energy cost during query execution have been developed and evaluated but these do not cover NoSQL data management systems.

## 2. Scientific issues

Numerous NoSQL systems exist, they are selected and used by the giants of the computer industry such as Amazon and the industrialists on the sole basis of their performance criteria without any regard for the energy constraint. Database systems have become energy sinkholes due to the massive explosion of data, building so-called green systems is a real challenge for researchers in the sense that they must allow the continuity of services (backup and restitution of data) and meet the challenges of climate change.

## 3. Objective:

The objective of this master thesis is to overcome the limitations of existing work on the energy consumption of data management systems by covering NoSQL systems in particular by doing:

- An evaluation followed by a comparative study of the energy consumption of two NoSQL systems during data loading and querying with the objective of raising awareness and guiding the choice of the most energy sensitive system.
- Then study the functioning of these systems in order to identify and exploit the possible integration points of the energy dimension with the objective of proposing a greener system (energy sensitive).

## 4. Required

The candidate must have skills on relational and NoSQL databases, SQL language, performance analysis tools, machine learning.

**Required technologies:** MongoDB, Cassandra, R learning, Papi, Java, C++.

**Desired profile:** Master student or equivalent degree.

**Contact:** -

**Application:** -

**Gratification:** -

**Internship period:** -

**Continuity in PhD:** Yes, according to the results obtained

## 5. References:

- a. Simon Pierre Dembele, Ladjel Bellatreche, Carlos Ordonez, Amine Roukh, Think big, start small: a good initiative to design green query optimizers, Cluster Computing Journal, Springer, 2019, <https://doi.org/10.1007/s10586-019-03005-0>
- b. DEMBELE, Simon Pierre, BELLATRECHE, Ladjel, et ORDONEZ, Carlos. Towards Green Query Processing-Auditing Power Before Deploying. In: 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020. p. 2492-2501.
- c. Simon Pierre Dembele, Ladjel Bellatreche, Carlos Ordonez, al. ,Big Steps Towards Query Eco-Processing - Thinking Smart, journal ARIMA J., 34, 2020, doi 10.46298/arma.6767
- d. D. Mahajan and Z. Zong, "Energy efficiency analysis of query optimizations on MongoDB and Cassandra," 2017 Eighth International Green and Sustainable Computing Conference (IGSC), 2017, pp. 1-6, doi: 10.1109/IGCC.2017.8323581.