ECOFLOC: A Multi-component Energy Measurement Tool

1st Humberto VALERA
Technopôle Domolandes, Laboratoire recherche
Saint-Geours-de-Maremne, France
humberto.valera@domolandes.fr

2nd Franck RAVAT
Université de Toulouse Capitole, IRIT (CNRS 5505)
Toulouse, France
frank.ravat@irit.fr

3rd Philippe ROOSE
Université de Pau et des Pays de l'Adour, LIUPPA
Anglet, France
philippe.roose@univ-pau.fr

4th Jiefu SONG
Université de Toulouse Capitole, IRIT (CNRS 5505)
Toulouse, France
jiefu.song@irit.fr

5th Nathalie VALLES-PARLANGEAU
Université de Pau et des Pays de l'Adour, LIUPPA
Anglet, France
nathalie.valles-parlangeau@univ-pau.fr

***Index Terms*—comprehensive energy consumption, BDA, software energy measurement tools**

I. EXTENDED ABSTRACT

Nowadays, Big Data Analytics (BDA) requires complex processing that relies on hardware resources that are often energy-intensive, whether for storage, data transfer, or computational tasks. This demand arises at a time when energy savings are crucial: to reach net-zero emissions by 2050, the digital sector's power consumption must be reduced by 55%.

Addressing this issue begins with accurately measuring energy consumption. Although some research focuses on the energy consumption of software applications, few studies provide measurement methods that account for complex architectures, like Data Lakes, which involve a full range of applications and hardware resources [1]. Current measurement approaches include physical sensors that assess device power consumption but do not distinguish between different applications or hardware components [2]. Another approach involves software tools that evaluate energy use at the process or code fragment level. However, these tools often overlook components essential for BDA, such as RAM and I/O devices [3]–[5]. Other software tools consider a range of hardware components, yet limit their analysis to specific workloads over set periods and may lack compatibility with new operating system versions [1]. Meanwhile, some approaches track the full journey of client requests in distributed environments but are unsuitable for long-term analysis of complex systems like BDA [6].

We introduce ECOFLOC, a BDA-compatible tool for measuring applications' energy consumption. ECOFLOC collects power values from hardware components and application loads —including processes, subprocesses, and threads— via GNU/Linux interfaces (with Windows support forthcoming) and converts them into energy data through specific mathematical computations. For example, CPU energy is derived from capacitance, voltage, and frequency; RAM from read/write operations; NIC and storage from data transfer loads; and GPU from real-time power data provided by the driver.

ECOFLOC allows users to specify not only the applications and hardware components to be analyzed, but also the analysis frequency, total analysis duration, and the option to monitor intermittent applications (those that may start and stop repeatedly).

We evaluated ECOFLOC during the ingestion and metadata generation phases of a Data Lake, using a benchmark specifically designed to assess Big Data operations within such an environment [7]. The ingestion phase manages the transfer and storage of data, including PDF files and structured data. The metadata generation phase enhances data governance within the Data Lake. This process involves several stages: transforming source data, optimizing data, generating metadata, and performing similarity analyses. These operations rely on a complex technical architecture comprising Elasticsearch, MongoDB, Neo4j, and SQLite instances.

Figure 1 shows the energy consumption of data ingestion and metadata generation operations in the benchmark. For data ingestion, the storage device was the most energy-intensive component, consuming 197.46 J (39.8% of the total), followed by RAM at 111.95 J (22.6%), and the network card at 86.39 J (17.4%), reflecting the I/O-intensive nature of these operations. In contrast, metadata processing showed higher energy consumption in the CPU (637.27 J, 46.7%) and RAM (581.23 J, 42.6%) due to the nature of the tasks involved.

With ECOFLOC, we can detail the energy consumption of each hardware component for specific operations performed

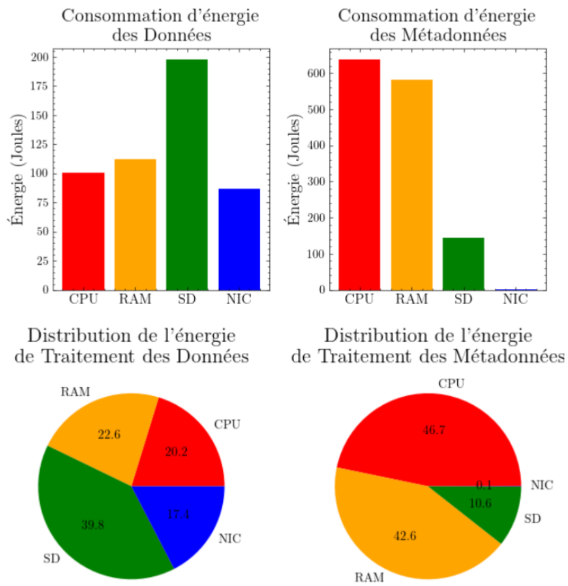

Fig. 1. Energy Consumption for Data Ingestion and Metadata Generation

by an application or complex system. This insight supports energy-saving strategies across various areas, such as load balancing, code optimization, and data processing.

REFERENCES

[1] H. Valera, A. Maurice, F. Ravat, J. Song, P. Roose, and N. Valles-Parlangeau, "Energy measurement system fordata lake: An initial approach," in *Intelligent Information and Database Systems*. Springer Nature Singapore, 2024.

[2] M. Jay, V. Ostapenco, L. Lefevre, D. Trystram, A.-C. Orgerie, and B. Fichel, "An experimental comparison of software-based power meters: focus on cpu and gpu," in *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing*, 2023.

[3] A. Noureddine, "Powerjoular and joularjx: Multi-platform software power monitoring tools," in *2022 18th International Conference on Intelligent Environments (IE)*, 2022.

[4] B. Petit, "scaphandre," 2023. [Online]. Available: https://github.com/hubblo-org/scaphandre

[5] M. Hähnel, B. Döbel, M. Völp, and H. Härtig, "Measuring energy consumption for short code paths using rapl," *SIGMETRICS Perform. Eval. Rev.*, 2012.

[6] V. Anand, Z. Xie, M. Stolet, R. De Viti, T. Davidson, R. Karimipour, S. Alzayat, and J. Mace, "The odd one out: Energy is not like other metrics," vol. 3, 2023.

[7] P. N. Sawadogo and J. Darmont, "Dlbench+: A benchmark for quantitative and qualitative data lake assessment," *Data & Knowledge Engineering*, 2023.