Challenges in Measuring Energy Consumption of a Data Lake with Power Joular and Scaphandre

Hernan Humberto ALVAREZ-VALERA[1][1], Alexandre MAURICE[1][2], Franck RAVAT[2][3], Jiefu SONG[2][4], Philippe ROOSE[3][5], and Nathalie VALLES-PARLANGEAU[3][6]

[1] Domolandes, Saint Geours de Maremne, France
{humberto.valera,alexandre.maurice}@domolandes.fr
[2] IRIT-UT Capitole, Toulouse, France {franck.favat,jiefu.song}@irit.fr
[3] LIUPPA, Anglet, France
{philippe.roose,nathalie.valles-parlangeau}@univ-pau.fr

1 Context

This document outlines the challenges faced in using established energy consumption measurement tools with DLBench+ processes. Given their ease of use and proven effectiveness, we have chosen Power Joular [1] and Scaphandre[4] and we compared them with the system we developed in some key aspects.

As a first step, we executed and evaluated the DLBench+ processes using both tools on a server with the following characteristics:

HARDWARE SETUP		
Hardware	**Spec.**	**Power Params.**
CPU	Intel Core I7-850H 2.20GHZ (12 vcores)	TDP: 45W
RAM	samsung M471A2K43CB1-CRC	Values in section XX
NIC	Cannon Lake PCH CNVi WiFi	Dowload Power: 0.55W Upload Power: 1.029 W
SD	Samsung MZVLW512HMJP	Write Power: 6.1 W Read Power: 5.1 W
SOFTWARE SETUP		
Software	**Spec.**	
O.S.	Ubuntu Linux - Kernel V. 6.2.0-26	
Frameworks	Neo4j 4.1.12, Elasticsearch 7.17.10, MongoDB 6.0.5, SQLite 3.37.2	

Table 1. Server's software and hardware configuration

2 Challenges encountered

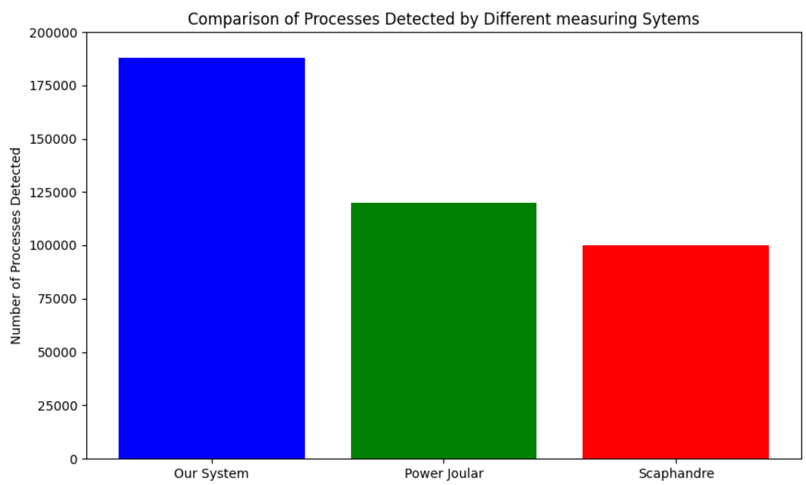
2.1 Processes Overlooked

After running the entire DLBench+ ingestion cycle twice with the three tools, we observed that there are short operating system processes (generally less than

[4] https://github.com/hubblo-org/scaphandre

two seconds) that neither PowerJoular nor Scaphandre account for. That is, they either fail to detect the PID of these processes or they report a result with zero energy consumption.

During each run, as Figure 2.1 depicts, DLBench+ generated approximately 200,000 operating system processes of varying durations. Our system successfully detected the energy consumption of 94% of these processes, while Power Joular detected only 60%, and Scaphandre detected 50%.



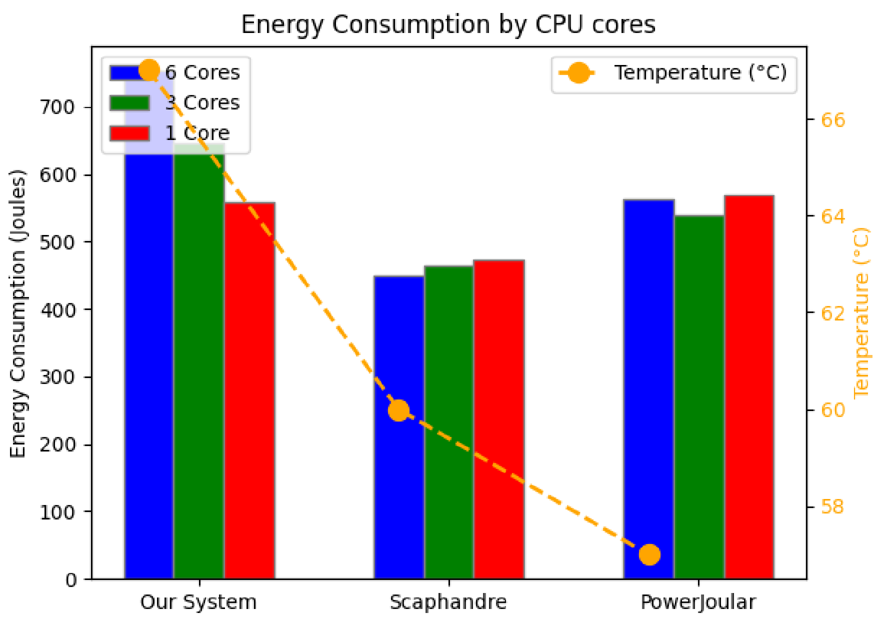
Key Insight: Measuring short processes poses a challenge for tools like Power Joular and Scaphandre. This limitation makes them less suitable for dynamic environments such as Data Lakes, where processes continuously appear and disappear.

2.2 Energy Consumption Comprehensiveness:

As mentioned in the main article, existing energy consumption measurement tools consider the CPU as the sole component. Our system, in its current version, takes into account the CPU, RAM, network card, and storage device for energy measuring.

2.3 Correlation Between Core's Load and Energy Consumption:

Another aspect we aimed to explore is the energy behavior of processes in relation to the number of cores they utilize, as this provides insight into energy consumption based on parallelism. For this, we repeated a simple experiment five times, which involved running the Sieve of Eratosthenes for two minutes with 6, 3, and 1 thread, ensuring the use of different cores. As shown in Figure 2.3, both Scaphandre and Power Joular exhibit a linear consumption regardless of the thread differences, whereas our system maintains the correlation of fewer cores, and less energy for the same workload. We also measured the average CPU temperature (lm-sensors[5]), confirming this correlation.



Key Insight: Despite the proven and undeniable effectiveness of both tools, in certain configurations like the ones used for our experiments, energy consumption is not related to the level of parallelism of operating system processes.

References

1. Noureddine, A.: Powerjoular and joularjx: Multi-platform software power monitoring tools. In: 2022 18th International Conference on Intelligent Environments (IE) (2022). https://doi.org/10.1109/IE54923.2022.9826760

[5]https://linux.die.net/man/1/sensors