[Extended Abstract] Developing a Resource Efficiency Benchmark

Norbert Schmitt
norbert.schmitt@uni-wuerzburg.de
University of Wuerzburg, Germany

Richard Vobl
vobl@retit.de
RETIT GmbH, Germany

Andreas Brunnert
brunnert@retit.de
RETIT GmbH, Germany

Samuel Kounev
samuel.kounev@uni-wuerzburg.de
University of Wuerzburg, Germany

Data centers, providing vital services to businesses and consumers, are growing in number and size. Professional performance engineers try to optimize how much of a server's resources an application can use [1] and how applications are scaled to meet demands across availabe machines [2], among other techniques. The developer of said application is responsible for designing and writing software that does not use computing resources excessively and not performance engineers and providers. Further, developers have little to no overview of available or utilization of computing resources when an application is finally deployed. Promising easy and fast scalability, giving the impression of infinite computing resources, developers also have no incentive to reduce resource wastage in their application.

Additionally, using an excess of resources also could increase the energy consumption of data centers. The New York Times sized data center power usage at about 30 billion watts, or 263 TWh of energy annually [3]. Andrea et al. also estimated that energy consumption rises to 1137 TWh until 2030 [4]. The most energy is consumed by servers when under load or in other words, when applications utilize computing resources. Hardware manufacturers counter energy usage with more efficient products and energy-saving techniques like dynamic voltage and frequency scaling (DVFS). Still, technological advances cannot fully compensate for the growth rate of data centers, and wastage of computing resources [5].

Resource-efficient hardware and runtime resource optimizations have been and are still extensively researched; the resource efficiency of the software itself is challenging. While many benchmarks exist that address resource utilization, they typically do not address the software but stress the hardware with defined workloads and load levels. This approach is, in our opinion, not suitable to determine the resource efficiency of the software. Software is a broad term and differs immensely in terms of requirements and functionality. Therefore, it is challenging to measure and describe the resource efficiency of software without the underlying hardware. Hence, we envision a new benchmark capable of measuring and describing the resource efficiency of cloud software. Such a benchmark allows rating software according to its resource efficiency, given the applications are of the same type (for example, ERP) and raise awareness among developers and operators to program and use more efficient software.

In this talk, we want to present, discuss, and gain feedback on our vision, including the following. We will present the general idea and design of a resource efficiency benchmark. This benchmark includes several key components to perform measurements and publish the results to increase comparability, foster openness, and as a result, increases awareness among developers and data center operators. But before the benchmark can be programmed, the first selection of workloads to stress important types of applications, such as Enterprise Resource Planning (ERP) software, must be made due to the immense variety of applications with different functionalities. Therefore, we plan to survey the essential application types as a starting point for our benchmark.

References

[1] Yichao Jin, Yonggang Wen, and Qinghua Chen. 2012. Energy Efficiency and Server Virtualization in Data Centers: An Empirical Investigation. In 2012 IEEE Conference on Computer Communication Workshops. 133—138.

[2] Nikolas Herbst. 2018. Methods and Benchmarks for Auto-Scaling Mechanisms in Elastic Cloud Environments. Ph.D. Dissertation. University of Würzburg. Germany. https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/16431

[3] James Glanz. 2012. Power, Pollution and the Internet. New York Times (September 23rd 2012)

[4] Anders Andrae and Tomas Edler. 2015. On Global Electricity Usage of Communication Technology: Trends to 2030. Challenges 6, 1 (Apr 2015),

117–157. https://doi.org/10.3390/challe6010117
[5] Erik Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. Recalibrating global data center energy-use estimates. Science 367, 6481 (2020), 984—986 https://doi.org/10.1126/science.aba3758