A New Course on Systems Benchmarking - For Scientists and Engineers

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

CCS CONCEPTS

• **General and reference** → **Performance**; *Measurement*; *Evaluation*; • **Software and its engineering** → *Software performance*;

KEYWORDS

Benchmarking, Metrics, Performance Evaluation, Measurement Techniques, Experimental Design, Quantitative Analysis

A benchmark is a tool coupled with a methodology for the evaluation and comparison of systems or components with respect to specific characteristics, such as performance, reliability, or security [1]. Benchmarks enable educated purchasing decisions and play a great role as evaluation tools during system design, development, and maintenance. In research, benchmarks play an integral part in evaluation and validation of new approaches and methodologies.

Traditional benchmarks have been focused on evaluating performance, typically understood as the amount of useful work accomplished by a system (or component) compared to the time and resources used. Ranging from simple benchmarks, targeting specific hardware or software components, to large and complex benchmarks focusing on entire systems (e.g., information systems, storage systems, cloud platforms), performance benchmarks have contributed significantly to improve successive generations of systems. Beyond traditional performance benchmarking, research on dependability benchmarking has increased in the past two decades. Due to the increasing relevance of security issues, security benchmarking has also become an important research field. Finally, resilience benchmarking faces challenges related to the integration of performance, dependability, and security benchmarking as well as to the adaptive characteristics of the systems under consideration.

Each benchmark is characterized by three key aspects: metrics, workloads, and measurement methodology. The metrics determine what values should be derived based on measurements to produce the benchmark results. The workloads determine under which usage scenarios and conditions (e.g., executed programs, induced system load, injected failures / security attacks) measurements should be performed to derive the metrics. Finally, the measurement methodology defines the end-to-end process to execute the benchmark, collect measurements, and produce the benchmark results. The increasing size and complexity of modern systems make the engineering of benchmarks a challenging task. Thus, we see the need for a better education on the theoretical and practical foundations necessary for gaining a deep understanding of benchmarking and the benchmark engineering process.

In this talk, we present an overview of a new course focused on systems benchmarking, based on our book[1] *Systems Benchmarking - For Scientists and Engineers* [1]. The course captures our experiences that have been gained over the past 15 years in teaching a regular graduate course on performance engineering of computing systems. The latter was taught at four different European universities since 2006, including University of Cambridge, Technical University of Catalonia, Karlsruhe Institute of Technology, and University of Würzburg.

The conception, design, and development of benchmarks requires a thorough understanding of the benchmarking fundamentals beyond understanding of the system under test, including statistics, measurement methodologies, metrics, and relevant workload characteristics. The course addresses these issues in depth; it covers how to determine relevant system characteristics to measure, how to measure these characteristics, and how to aggregate the measurement results in a metric. Further, the aggregation of metrics into scoring systems, as well as the design of workloads, including workload characterization and modeling, are additional challenging topics that are covered. Finally, modern benchmarks and their application in industry and research are studied. We cover a broad range of different application areas for benchmarking, presenting contributions in specific fields of benchmark development. These contributions address the unique challenges that arise in the conception and development of benchmarks for specific systems or subsystems. They also demonstrate how the foundations and concepts of the first part of the course are being used in existing benchmarks.

REFERENCES

[1] Samuel Kounev, Klaus-Dieter Lange, and Jóakim von Kistowski. 2020. *Systems Benchmarking - For Scientists and Engineers* (1 ed.). Springer International Publishing. https://doi.org/10.1007/978-3-030-41705-5

[1]http://benchmarking-book.com/