

Preface

Big data processing has evolved from a buzz-word to a pervasive technology shift in only a few years. The plethora of systems developed or extended for big data processing makes big data benchmarking more important than ever.

Formed in 2012, the Big Data Benchmarking Community (BDBC) represented a major step in facilitating the development of benchmarks for objective comparisons of hardware and software systems dealing with emerging big data applications. Led by Chaitanya Baru, Tilmann Rabl, Milind Bhandarkar, Raghunath Nambiar, and Meikel Poess, the BDBC has successfully conducted seven international Workshops on Big Data Benchmarking (WBDB) in only 4 years bringing industry experts and researchers together to present and debate challenges, ideas, and methodologies to benchmark big data systems. In 2015, the BBDC joined the SPEC Research Group to further organize and structure the benchmarking efforts collected under the BBDC umbrella. The Big Data Working Group¹ meets regularly as a platform for discussion and development of big data benchmarks.

While being a platform of discussion, the BBDC and SPEC Research Group also demonstrated practical impact in the field of big data benchmarking. In 2015, Big-Bench, one of the benchmark efforts started at WBDB2012.us, was adopted by the Transaction Processing Performance Council to become the first industry standard end-to-end benchmark for big data analysis. The benchmark was officially released under the name TPCx-BB in January 2016².

This book contains the joint proceedings of the 6th and 7th Workshop on Big Data Benchmarking. The 6th WBDB was held in Toronto, Canada, during June 16–17, 2015, hosted by the University of Toronto. The 7th WBDB was held in New Delhi, India, during December 14–15, 2015, at the India Habitat Centre. Both workshops were well attended by industry and academia alike and featured SPEC Research Big Data Working Group meetings. Keynote speakers in WBDB2015.ca were Tamer Özsu (University of Waterloo) and Anil Goel (SAP). Tamer Özsu presented benchmarks for graph management and Anil Goel gave an overview of the SAP big data infrastructure. In WBDB2015.in Michael Franklin (UC Berkeley) presented emerging trends in big data software, Geoffrey Fox (Indiana University Bloomington) discussed the convergence of big data and high-performance computing (HPC), and Mukund Desphande (Persistent Systems) introduced the use of Web technologies in database architecture.

In this book, we have collected recent trends in big data and HPC convergence, new proposals for big data benchmarking, as well as tooling and performance results. In the first part of the book, an overview of challenges and opportunities of the convergence of HPC and big data is given. The second part presents two novel benchmarks, one based on the Indian Aadhaar database that contains biometric information of hundreds

¹ See <https://research.spec.org/working-groups/big-data-working-group>.

² See <http://www.tpc.org/tpcx-bb/default.asp>.

of millions of people and one for array database systems. The third part contains experiences from a benchmarking framework for Hadoop setups. In the final part, several performance results from extensive experiments on Cassandra, Spark, and Hadoop are presented.

The seven papers in this book were selected out of a total of 39 presentations in WBDB2015.ca and WBDB2015.in. All papers were reviewed in two rounds. We thank the sponsors, members of the Program Committee, authors, and participants for their contributions to these workshops. The hard work and close cooperation of a number of people have been critical to the success of the WBDB workshop series.

August 2016

Tilman Rabl

Big Data Benchmarking

6th International Workshop, WBDB 2015, Toronto, ON,
Canada, June 16-17, 2015 and 7th International
Workshop, WBDB 2015, New Delhi, India, December
14-15, 2015, Revised Selected Papers

Rabl, T.; Nambiar, R.; Baru, C.; Bhandarkar, M.; Poess,
M.; Pyne, S. (Eds.)

2016, IX, 129 p. 60 illus., Softcover

ISBN: 978-3-319-49747-1