

Preface

Formed in 2012, the Big Data Benchmarking Community (BDBC) represents 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, Meikel Poess, Milind Bhandarkar, and Nambiar Raghunath, the BDBC has successfully conducted five international Workshops on Big Data Benchmarking (WBDB). One strength of the WBDB is that it brings together practitioners and researchers, which leads to a balance between industrial and academic contributions. It provides the right environment in which to discuss the challenges and potential approaches to benchmark big data systems. The results of the WBDB have a high impact on the ongoing research in the domain and WBDB itself is established as the leading event focusing on big data benchmarking. Previous editions in 2012 and 2013 were organized in the USA, India, and China. In August 2014, the fifth WBDB finally came to Europe.

WBDB 2014 took place in Potsdam hosted by the Hasso Plattner Institute. It was only possible thanks to the generous support of SAP SE and HP/Suse as gold sponsors and Intel, Mellanox, and Pivotal as silver sponsors. The number of high-quality submissions allowed us to define an interesting program covering many different aspects of big data benchmarking. Each paper was reviewed by at least three reviewers. In addition to the technical program, the local arrangements team under Matthias Uacker did a great job and made it an unforgettable event. The program of the workshop was enriched with two exciting keynote talks:

- “An Approach to Benchmarking Industrial Big Data Application” by Umseh Dayal
- “A TU Delft Perspective on Benchmarking Big Data in the Data Center” by Alexander Iosup.

Beside all the exciting talks, we had many inspiring discussions on the different presentations, e.g., on how to build standard benchmarks that can be used by industry as well as academia. In this book, we collected the revised version of the submissions and grouped them into the following three sections: “Benchmark Specifications and Proposals”—submissions focusing on concrete benchmarks and/or proposals on how to implement such benchmarks, “Hadoop and MarReduce”—papers discussing Hadoop and MapReduce in the different context such as virtualization and cloud, and “In-Memory, Data Generation Graphs.”

Overall, 2014 was an exciting year for the BDBC. An important step in the direction of an industry standard benchmarks for big data benchmarking was the decision of the BDBC to join the Standard Performance Evaluation Corporation (SPEC). The BDBC formed a new working group within the SPEC Research Group (RG). The SPEC RG is a group within SPEC established to serve as a platform for collaborative research efforts in the area of quantitative system evaluation and analysis, fostering the interaction between industry and academia in the field. The collaboration between the

BDBC and SPEC was announced at WBDB 2014. Since 2015, the BDBC community has been represented in the SPEC RG Steering Committee by Tilmann Rabl, who is also heading the Big Data Benchmarking working group together with Chaitanya Baru. The working group is currently working on a proposal for an industry standard benchmark based on the BigBench proposal. Further details about BigBench are discussed in this book.

We thank the sponsors, members of the Program Committee, authors, local arrangement team, and participants for their contributions to the workshop.

March 2015

Chaitanya Baru
Tilmann Rabl
Kai Sachs

Big Data Benchmarking

5th International Workshop, WBDB 2014, Potsdam,
Germany, August 5-6- 2014, Revised Selected Papers
Rabl, T.; Sachs, K.; Poess, M.; Baru, C.; Hans-Arno, J.
(Eds.)

2015, X, 157 p. 54 illus., Softcover

ISBN: 978-3-319-20232-7