DataBench: Evidence Based Big Data Benchmarking to Improve Business Performance*

Richard Stevens, Gabriella Cattaneo, Helena Schwenk, Cristina Pepato, Christopher Ostberg Hansen
IDC Italy

Nuria De Lama Sanchez, Tomas Pariente Lobo
ATOS SPAIN SA, Spain

Chiara Francalanci, Barbara Pernici, Angela Geronazzo, Paolo Giacomazzi, Lucia Polidori
POLITECNICO DI MILANO, Italy

Arne Jørgen Berre, Volker Hoffman
SINTEF AS, Norway

Marko Grobelnik, James Hodson
INSTITUT JOZEF STEFAN, Slovenia

Todor Ivanov, Roberto V. Zicari
FRANKFURT BIG DATA LAB, GOETHE UNIVERSITY, Germany

ABSTRACT

Organisations rely on evidence from the Benchmarking domain to provide answers on how their processes are performing. There is extensive information on how and why to perform technical benchmarks for the specific management and analytics processes, but there is a lack of objective, evidence-based methods to measure the correlation between Big Data Technology (BDT) benchmarks and an organisation's business benchmarks and demonstrate return on investment (ROI). The DataBench project addresses this significant gap in the current benchmarking community's activities, by providing certifiable benchmarks and evaluation schemes of BDT performance of high business impact and industrial significance.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*;

KEYWORDS

DataBench, Big Data Benchmarking

ACM Reference Format:
Richard Stevens, Gabriella Cattaneo, Helena Schwenk, Cristina Pepato, Christopher Ostberg Hansen, Nuria De Lama Sanchez, Tomas Pariente Lobo, Chiara Francalanci, Barbara Pernici, Angela Geronazzo, Paolo Giacomazzi, Lucia Polidori, Arne Jørgen Berre, Volker Hoffman, Marko Grobelnik, James Hodson, and Todor Ivanov, Roberto V. Zicari . 2018. DataBench: Evidence Based Big Data Benchmarking to Improve Business Performance. In *Proceedings of The 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD'18).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.475/123_4

*Contact Email: databench@databench.eu

KDD'18, August 19-23, London, United Kingdom

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

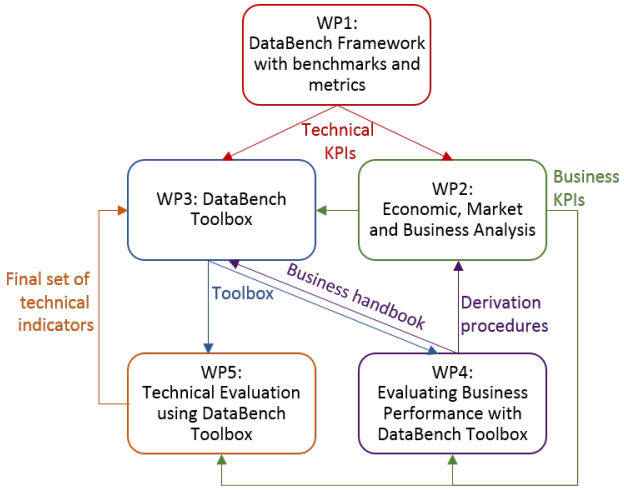

Figure 1: DataBench Workflow

1 DESCRIPTION

The approach followed by DataBench starts with a performing comparative analysis of existing benchmarking initiatives and technologies. In fact, the goal of DataBench is not to create another benchmark, but to support an approach for efficient usage, evolution, extensions and synergy of the available Big Data benchmarks from the international Big Data benchmarking community related to industrial requirements. Based on that, the project will proceed to establish metrics for the different technical priorities identified and develop a methodology, an economic and market analysis to assess the European and industrial significance of the BDT

to be benchmarked. Industrial significance will be assessed through the investigation of the main Big Data use cases, which will allow the correlation of Big Data technical performance with business processes. Based on all these inputs, the project will build the DataBench Toolbox, a tool which will connect and evaluate external benchmarking initiatives. Using the DataBench Toolbox and the methodology and metrics previously defined, evaluation and benchmarking will be carried out considering both business relevance and technical aspects. Figure 1 represents how the different elements from the DataBench ecosystem will interact in order to achieve the project goals described in the next section.

2 PROJECT GOALS

At the heart of DataBench is the goal to design a benchmarking process helping European organizations developing BDT to reach for excellence and constantly improve their performance, by measuring their technology development activity against parameters of high business relevance. DataBench will investigate existing Big Data benchmarking tools and projects, identify the main gaps and provide a robust set of metrics to compare technical results coming from those tools. To achieve this goal, the project will pursue the following objectives:

(1) Provide the BDT stakeholder communities with a comprehensive framework to integrate business and technical benchmarking approaches for BDT.
(2) Perform economic and market analysis to assess the "European economic significance" of benchmarking tools and performance parameters.
(3) Evaluate the business impacts of BDT benchmarks of performance parameters of industrial significance.
(4) Develop a tool applying methodologies to determine optimal BDT benchmarking approaches.
(5) Evaluation of the DataBench Framework and Toolbox in representative industries, data experimentation/ integration initiatives (ICT-14) and Large-Scale Pilot (ICT-15).
(6) Liaise closely with the BDVA, ICT-14 and 15 projects to build consensus and to reach out to key industrial communities, to ensure that benchmarking responds to real needs and problems.
(7) Bring together Research, Academia and industry establishing the Big Data Benchmarking Community.

3 OUTCOMES

Figure 1 shows expected major work packages (WP) of the DataBench ecosystem together with their input and output results that can be described as:

- **The DataBench Framework (WP1)** - Including a complete set of metrics for BDT assessment. Figure 2 shows a preliminary star diagram, where different dimensions represent characteristics of case studies across industries. These cases have to be assessed in order to support the high-level design of the technology architecture and the selection of corresponding technical benchmarks.
- **Economic, Market and Business Analysis (WP2)** - Assessing the European and industrial significance of the BDT examined by the project. Several of the project objectives listed in Section 2 ((1), (2), (3) and (5)) will be addressed here.
- **The DataBench Toolbox (WP3)** - Being the core technical component of the DataBench Framework, it will take as input most of the work done in the rest of the project's work packages. The Toolbox will be the entry point for users that would like to perform Big Data benchmarking and will ultimately deliver recommendations and business insights. It will include features to reuse existing Big Data benchmarks, and will help users to search, select, download, execute and get a set of homogenized results. The **AI/ML Framework (WP5)** will serve to enable recommendations for the benchmarking community based on past experiences and will be integrated with the Toolbox.
- **The DataBench Handbook (WP4)** - Providing guidelines to the use of the project's results, associating the DataBench Framework and the DataBench Toolbox, describing metrics implementation and benchmarks.

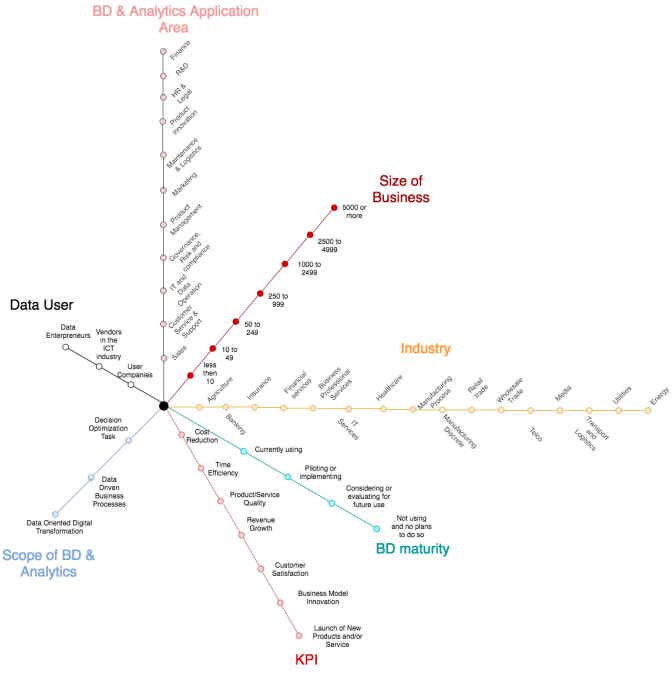

Figure 2: Business Dimensions