BRIDGING THE GAP BETWEEN TECHNICAL AND BUSINESS BENCHMARKING

Research-in-Progress

Track N° 13

Francalanci, Chiara, Politecnico di Milano, Milano, Italia, chiara.francalanci@polimi.it

Geronazzo, Angela, Politecnico di Milano, Milano, Italia, angela.geronazzo@polimi.it

Giacomazzi, Paolo, Politecnico di Milano, Milano, Italia, paolo.giacomazzi@polimi.it

Pernici, Barbara, Politecnico di Milano, Milano, Italia, barbara.pernici@polimi.it

Polidori, Lucia, Politecnico di Milano, Milano, Italia, lucia.polidori@polimi.it

Abstract

This paper proposes a methodology and a model for a better insight into the relationship between business and technical benchmarking in BDT (Big Data Technology) use cases. Technical benchmarks are aimed to help IT managers in making technical decisions by measuring key performance metrics of the underlying BD infrastructure. Business benchmarks are aimed at associating BDT use cases with measurable business benefits. In principle, making the right technical choices is key to deliver business benefits. However, the relationship between technical and business benchmarking is rarely addressed in previous research, which mostly focuses on either side of benchmarking. This methodology takes a first step towards bridging the gap between technical and business benchmarking. The paper illustrates the methodology and provides preliminary evidence gained from a desk analysis.

Keywords: Business KPI, technical benchmark, performance metric, Big Data, Big Data Technology, BDT.

1 Introduction

In the last decade, several initiatives in the industry and in academia have developed technical benchmarks able to assess the performance of Big Data Technologies (BDT) (Gao *et al.*, 2018). Although these two areas are recognized to be related (Yin and Kaynak, 2015; Marjani *et al.*, 2017), a limited effort has been devoted to understanding the relation between BDT choices and business performance. The paper (Yin and Kaynak, 2015) summarized previous contributions on Big Data potential benefits and speculates that they differ from sector to sector and in different areas, with a particular emphasis on customer-centric ones. Moreover, the paper stresses the challenges introduced by the complexity of the data used by Big Data systems. The paper speculates on the importance to address technical challenges but does not offer any guidelines on how to address these challenges. The paper (Marjani *et al.*, 2017) analyses Big Data application and technical solutions in Internet of Things (IoT) applications, with a specific focus on analytics. The paper focuses on the technical aspects but overlooks business benefits.

This work aims to link technical benchmarking and business Key Performance Indicators (KPIs) by presenting a new methodology useful to i) extract business process characteristics and Big Data features, ii) guide the design of different BDT by considering requirements and challenges, and iii) select the most suitable technical benchmark to assess the performance of these configurations. The methodology has been developed as part of the H2020 DataBench[1] project that has created the context for the BDT and business benchmarking communities to cooperate towards developing an integrated framework able to support the design and implementation of use cases. In this perspective, a use case is "a discretely funded effort designed to accomplish a particular business goal or objective through the application of big data technology to particular business processes and/or application domains, employing line-of-business and BDT resources" (Pernici, Francalanci, Geronazzo, Polidori, Ivanov, *et al.*, 2018). The methodology has been designed by extending a preliminary work (Pernici, Francalanci, Geronazzo, Polidori, Cattaneo, *et al.*, 2018; Pernici, Francalanci, Geronazzo, Polidori, Ivanov, *et al.*, 2018; Francalanci *et al.*, 2019) focused on analysing the relationships between business and technical benchmarking. In particular, the methodology spawned from the definition of three macro-areas: i) the business features macro-area, which collects information regarding business goals and features of the organization, ii) the Big Data features macro-area, which concerns the technical information about the Big Data involved in the use case, and iii) the Technical Benchmark macro-area, which refers to the features of the technical benchmarks useful to assess the performance of different BDT alternatives. The desk analysis identifies a set of key features, called dimensions, useful to describe each macro-area by distilling its relevant characteristics. Each dimension is in turn described by a set of relevant values.

The paper is structured as follows: Section 2 introduces relevant literature, Section 3 presents the methodology, Section 4 summarizes the desk analysis performed to validate the methodology. Section 5 drafts conclusions and outlines future work.

2 State of the art

Several contributions have highlighted the economic impact of BD, in industrial (James *et al.*, 2011; IBM Institute for Business Value, 2013) and academic reviews (Gandomi and Haider, 2015; Yin and Kaynak, 2015). As an example, in (James *et al.*, 2011; IBM Institute for Business Value, 2013) is explained how analyzing the exploding amount of data in our world can allow organizations to be more competitive, more innovative and can help managers in order to increase the productivity growth and

[1] https://www.databench.eu/

consumer surplus. According to research by MGI and McKinsey's Business Technology Office, that has been performed in five domains – healthcare in the United States, the public sector in Europe, retail in the United States, and manufacturing and personal-location data globally, big data can generate value in each industry among those considered. Moreover, in (Gandomi and Haider, 2015; Yin and Kaynak, 2015) authors highlight how technological advances in storage and computations have enabled cost-effective capture of the informational value of big data in a timely manner. However, limited effort has been devoted to design new methodologies to formally assess BD impact and to provide evidence of their economic benefits. Among them, the authors in (Sivarajah *et al.*, 2017) proposed a comprehensive review of BD analytics and conceptually classifies BD challenges in data, process and management. According to them, the most critical BD challenges for outdated data processing applications are the large datasets (in terms of size and complexity) and the ability to process vast amount of data. A further contribution is the methodology presented in (Supakkul, Zhao and Chung, 2016), the work aimed to bridge the gap between business goals and DB analytics by using a goal-oriented approach. Moreover, authors in (Chen, Chiang and Storey, 2018) outlined key characteristics and capabilities of business intelligence and analytics with an industry specific focus and analyze current research in BI&A. Moreover, challenges and opportunities associated with BI&A research and education are identified. This work is intended to serve, in part, as a platform and conversation guide for examining how the IS discipline can help business decision makers in understanding deeply the emerging BI&A technologies, in exploiting the new availability of big data and in managing the shortage of figures able to handle this large amount of data. A significant effort was made to develop models and indicators useful to evaluate business performance (Neely, 2002; Marr and Schiuma, 2003; Franco-Santos *et al.*, 2007). The authors in (Neely, 2002) proposed the Performance Prism framework to address the complexity of an organization's relationships with its stakeholders in its operating environment. The framework directs management attention to long-term success and helps organizations to design, build and operate their performance measurement systems in a way that is relevant to the specific conditions of their operating environment. The authors in (Franco-Santos et al., 2007) reviewed the different definitions of a business performance measurement system provided by the literature. Moreover, the research presented in (Marr and Schiuma, 2003) simplified the balanced scorecard (Kaplan and Norton, 1992) to be applied to SMEs. However, these papers overlook technology aspects and do not provide metrics useful for the assessment of processes either from a business or from a technology perspective.

An extended review of open source BD benchmarks is presented in (Han, John and Zhan, 2018), the paper organized benchmarks by type of system and discriminates among Hadoop-related systems, data stores and specialized systems and considers the three important aspects of benchmarking – workload generation techniques, workload input data generation techniques, and metric used to assess systems. It worth mentioning, among the benchmarks described, BigBench (Ghazal *et al.*, 2013, 2017; Baru *et al.*, 2015) an overarching suite to benchmark analytical capabilities of a BD platform, and Yahoo Cloud Serving Benchmark (Cooper *et al.*, 2010; Patil *et al.*, 2011) a comprehensive toolkit to benchmark OLTP solutions.

Some literature contributions reviewed and classified case studies with a specific focus on BDT and their impact (Fosso Wamba *et al.*, 2015; Sivarajah *et al.*, 2017; Gunasekaran *et al.*, 2018; Urbinati *et al.*, 2018). As an example, McKinsey (James *et al.*, 2011) analyzed the transformative potential of BD in healthcare, public sector administration, retail, manufacturing and personal location data, whereas IBM (IBM Institute for Business Value, 2013) investigated analytics capabilities, types of data and adopted infrastructures in the banking and financial sector. These contributions were analyzed in the preliminary desk analysis.

In (Pappas *et al.*, 2018) the considerable value of digital transformations that emerges from the analysis of big data is highlighted. According to the authors in the big data analytics ecosystem it is necessary to understand first the different actors, the data they generate, and how they interact, and second the capabilities of the system to be developed to harness this potential. There are different elements that are important in order to develop a data-driven culture within organizations: investing in appropriate technology, fostering technical and managerial skills, and promoting a climate of organizational learning are the most critical factors in realizing business value.

3 Methodology

The goal of our methodology is to model the relationship between business and technical performance of a specific use case. Ultimately, the goal is the selection of the technical benchmarks that support the technical choices that are, in turn, key to achieve business results. The selection of the most appropriate technical benchmark is challenging due to several factors, including: the large number of available approaches and tools, the focus of benchmarks on specific technologies, instead of aiding high-level technical decisions and the large number of technical performance metrics that can be important in BDT-related technical choices.

Figure 1. Steps of the methodology: business and technical branches.

The methodology depicted in Figure 1 has two branches, business and technical, respectively. The business branch involves the following steps: i) definition of the business features relevant for a given use case (e.g., product quality), ii) definition of critical business requirements (e.g., real-time constraints), and iii) definition of technical requirements and related metrics, associated with critical business requirements (e.g., read throughput).

The technical branch involves the following steps: i) definition of Big Data features (e.g., streaming data generated by IoT devices), ii) definition of key technical choices (e.g., data storage and acquisition), and iii) identification of alternative technologies (e.g., Apache Accumulo).

At the end of the two branches, a technical benchmarking suite can be selected to compare alternative technologies based on the technical metrics that are associated with critical business requirements. Technical decisions can then be made based on the results of technical benchmarking.

The first step of each branches, namely business and Big Data features, is aimed to describe the corresponding macro-area, business or technical. To support this description process, for each macro-area we defined a set of key features, called dimensions, useful to describe the use cases from the business and the technical perspectives.

The result of the first step along each branch in Figure 1 is a description of the data involved by the BDT use case and of the business features that managers will use to assess the benefits delivered by the BDT. The next step along the business branch identifies the critical business requirements, which are then related to corresponding technical requirements. The second and third steps along the business branch are difficult to classify based on a desk analysis, as they require context-dependent knowledge on business processes that is seldom available from public sources. Similarly, the identification of key technical choices and related alternative technologies is difficult to perform based on a desk analysis, as this type of information is most often missing.

As mentioned, BDT use cases can be described with reference to business and Big Data features, in the following we will describe these two branches. A fundamental dimension of the business features macro-area is the business KPI. In particular, a business KPI is a measurable value that demonstrates how effectively a company is achieving key business goals. The outlined values were selected as a preliminary general-purpose set of high-level benefits.

To effectively describe a business process, it is necessary to introduce other dimensions. The industry and analytics application area are dimensions useful to describe the context of the use case. The industry refers to the main sector of activity of the company, provided values group the most relevant industrial sectors, while the analytics application area specifies the main focus of the BDT use case. A further dimension, relevant to elicit technical requirements, is the level of BD integration that refers to the integration of BD and analytics in the business process. A low level of integration applies when Big Data are processed in a batch environment and made available the following day by providing reports/dashboards, while a high level of integration applies when a real-time acquisition and processing of the data supports the business process, e.g., fault detection and recovery. The dimensions, useful to describe a business feature macro-area, are summarized in Table 1.

Dimensions	Values
Business KPI	Cost reduction, time efficiency, product/service quality, revenue growth, customer satisfaction, innovation
Industry	Agriculture, banking/insurance/financial services, business professional services/IT services, healthcare, manufacturing, retail trade/wholesale trade, telco/media, transport/logistics, utilities/energy
Analytics application Area	Sales, customer service & support, IT and data operation, governance risk and compliance, product management, marketing, maintenance & logistics, product innovation, HR & legal, R&D, finance
Level of BD integration with business processes	Low, medium, high

Table 1. Dimensions and values of the business features.

The second branch shown in Figure 1 refers to the key features of the Big Data used in the BDT use case. Information about data characteristics are collected and summarized according to the Big Data features macro-area summarized in Table 2. The investigation of this macro-area is useful to link business features to technical benchmarks by means of the characteristics of the data involved in the use case. These dimensions include data variety, volume and velocity that characterize the data required/involved in the use case. The data variety dimension refers to types of data involved in the use case. The data volume dimension specifies the size of data expressed as order of magnitude and the data velocity refers to the timing of data acquisition and processing. These three dimensions raise specific challenges to the design of the BD infrastructure, as they create requirements for its critical layers. Moreover, the data source dimension refers to the type of data source and discriminates between centralized and distributed sources.

Dimensions	Values
Data Variety	Tables and structured data, graph and linked data, geospatial and temporal data (including time series and IoT data), media (image, audio or video), text and semi-structured data (XML, genomic data, etc.)
Data Volume	Gigabytes, Terabytes, Petabytes, Exabytes
Data Velocity	Batch (not in real-time), streaming (real-time), interactive/(near) real time, iterative/in-memory
Data source	Distributed, centralized
Analysis approach	Descriptive, diagnostic, predictive, prescriptive
Performance Metric	Cost, throughput, end-to-end execution time, accuracy/quality/data quality/veracity, availability

Table 2. Dimensions and values of the Big Data features.

The analysis approach dimension in Table 2 outlines a common categorization of analytic tasks and it is useful to tie key technical choices with the level of integration of the BDT use case business dimension in Table 1. In general, as the analysis approach moves from descriptive to prescriptive, the level of BD integration goes from low to high.

The performance metric dimension represents a key measurable metric useful to elicit the most appropriate technical benchmark to support the design of the BD infrastructure. Similarly, to the business branch, the technical branch requires context-dependent knowledge to perform the second and third steps. Further work in this area is ongoing and is analysing in details characteristics of different type of systems and their benchmarking opportunities.

Dimensions	Values
Output Metric	Response time, throughput, reliability, availability, architecture metrics, price-performance metrics, energy consumption metrics
BD Architecture Layer	Data visualization, data analysis, data processing, data management, data storage, communication and connectivity
System	Hadoop-related systems, data stores, specialized systems

Table 3. Dimensions and values of the technical benchmark.

In order to understand the contribution that can be provided by technical benchmarks, we extended our desk analysis to the technical benchmarks and classified their relevant dimensions and values. The technical benchmark macro-area is described in Table 3 and it is the output of a review of the available literature on technical benchmarks (Ghazal *et al.*, 2013; Wang *et al.*, 2014; Han, John and Zhan, 2018; Ivanov and Singhal, 2018). The analysis emphasized the following dimensions and associated values useful to characterize technical benchmarks. The main dimension is the benchmark output metric, that specifies its outcome, this dimension is useful to identify the final goal of the benchmarking process. However, it should be stressed that a benchmarking process requires a thorough understanding of the BD infrastructure and of the benchmark characteristics. Indeed, multiple factors, some not obvious, contribute to the performance measured by a given benchmark. These factors are summarized by three further dimensions: BD architecture layer, system type and system. The BD architecture layer identifies the different layers of an BD infrastructure, its values are drawn from the BDV Reference Model (European Big Data Value Association, 2015) and extended to aid the benchmark selection. The values of this dimension can be further categorized, e.g., the data processing value can be distinguished in streaming processing, interactive processing and batch processing. The system dimension identifies three broad families of systems that can be benchmarked using the benchmark, whereas the system type dimension further distinguishes among these systems by referencing to the main types of data storage and processing systems.
The most appropriate benchmark is then chosen based on many factors, however, the benchmark output metric needs to be able to assess the performance of the critical layers involved in the use case. The benchmark might not test the entire infrastructure, but focus only on a subset of layers considered relevant/critical to the improvement of some technical requirements that in turn are related to some business KPIs.

4 Experimental evaluation

The identified dimensions and values were validated by means of an extensive desk analysis. The complete list of use cases of the extensive desk analysis together with their mapping on the business and technical dimensions and values presented in Section 3 can be found at *http://131.175.120.100/itais2019/desk_analysis.xlsx*. The extensive desk analysis investigates each of the identified macro-areas based on use cases reported by the academic and industry literature. The desk analysis selects the group of use cases developed by the H2020 ICT 14-15 projects and the use cases presented by the literature (Fosso Wamba *et al.*, 2015; Sivarajah *et al.*, 2017; Urbinati *et al.*, 2018), including ICT vendors" white papers. The Technical Benchmark macro-area is analysed by referring to the available literature. The analysis confirmed the choice of dimensions and values extracted during the preliminary phase. However, it confirmed that some of the dimensions and values could be further specified to gain a better understanding of specific use cases.

A quantitative analysis of the classification reported at (Francalanci *et al.*, 2019) showed that some dimensions/values are cross-industry, while other dimensions/values are specific to a small subset of industries. As an example, from a business perspective the desk analysis highlighted customer satisfaction among the top relevant indicators in most industries, with a particular emphasis in industries that provide products/services to consumers, e.g., telco/media, healthcare, banking/insurance/financial services, retail trade/wholesale trade. Conversely, other KPIs appear more strictly related to a specific industry. As an example, cost reduction is the most relevant indicator in banking/insurance/financial services and utilities/energy, whereas revenue growth is the pivotal KPI in retail trade/wholesale trade, and transport/logistics and healthcare appear to be focused on product/service quality. Moreover, some industries are more concerned with innovation, e.g., utility/energy and agriculture.

From a technical perspective, the desk analysis indicated that tables and structured data tend to be present in all industries, although they are predominant in selected industries, such as banking/insurance/financial services. On the contrary, selected industries, including manufacturing, transport/logistics, utilities/energy have specific use cases addressing geospatial and temporal data created by IoT devices in monitoring and automation processes. Other types of data, such as graph and linked data, are present in all the industries that perform social media analysis.

Overall, from a data analysis perspective it emerged the need to process a growing amount of data by exploiting predictive/prescriptive methods with real-time constraints, thus making evident the quest for a structured approach able to tackle technical challenges and to support technical choices pivotal to enable business benefits. Moreover, these preliminary findings suggested the relevance of providing blueprints by industry to further investigate and structure the presented methodology.

5 Conclusions

The main goal of the presented methodology is to help organizations to understand the link between their business processes and the technical choices required to enable these processes. The methodology makes a step forward in this direction by using technical benchmarking to take more informed technical decisions with reference to the requirements imposed by business goals and to the challenges introduced by the usage of Big Data. Future work will extend the desk analysis to investigate the proposed dimensions and values. Also, essential, relevant values will be further categorized.

References

Baru, C. *et al.* (2015) "Discussion of BigBench: A Proposed Industry Standard Performance Benchmark for Big Data", in. Springer, Cham, pp. 44–63. doi: 10.1007/978-3-319-15350-6_4.

Chen, H., Chiang, R. and Storey, V. (2018) "Business intelligence and analytics: From big data to big impact. MIS Q 36 (4): 1165–1188".

Cooper, B. F. *et al.* (2010) "Benchmarking cloud serving systems with YCSB", in *Proceedings of the 1st ACM symposium on Cloud computing - SoCC "10*. New York, New York, USA: ACM Press, p. 143. doi: 10.1145/1807128.1807152.

European Big Data Value Association (2015) *European Big Data Value Strategic Research and Innovation*, *Big Data Value Europe*. Available at: http://www.nessi-europe.eu/Files/Private/EuropeanBigDataValuePartnership_SRIA__v099 v4.pdf.

Fosso Wamba, S. *et al.* (2015) "How "big data" can make big impact: Findings from a systematic review and a longitudinal case study", *International Journal of Production Economics*. Elsevier, 165, pp. 234–246. doi: 10.1016/j.ijpe.2014.12.031.

Francalanci, C. *et al.* (2019) *Data collection results*. Available at: https://www.databench.eu/wp-content/uploads/2018/10/databench-d4.1-ver-1.0.pdf.

Franco-Santos, M. *et al.* (2007) "Towards a definition of a business performance measurement system", *International Journal of Operations & Production Management*. Edited by M. Bourne. Emerald Group Publishing Limited, 27(8), pp. 784–801. doi: 10.1108/01443570710763778.

Gandomi, A. and Haider, M. (2015) "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*. Elsevier Ltd, 35(2), pp. 137–144. doi: 10.1016/j.ijinfomgt.2014.10.007.

Gao, W. *et al.* (2018) "BigDataBench: A Scalable and Unified Big Data and AI Benchmark Suite". Available at: https://arxiv.org/abs/1802.08254.

Ghazal, A. *et al.* (2013) "BigBench", in *Proceedings of the 2013 international conference on Management of data - SIGMOD "13*. New York, New York, USA: ACM Press, p. 1197. doi: 10.1145/2463676.2463712.

Ghazal, A. *et al.* (2017) "BigBench V2: The New and Improved BigBench", in *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE, pp. 1225–1236. doi: 10.1109/ICDE.2017.167.

Gunasekaran, A. *et al.* (2018) "Agile manufacturing practices: the role of big data and business analytics with multiple case studies", *International Journal of Production Research*. Taylor & Francis, 56(1–2), pp. 385–397. doi: 10.1080/00207543.2017.1395488.

Han, R., John, L. K. and Zhan, J. (2018) "Benchmarking Big Data Systems: A Review", *IEEE Transactions on Services Computing*. IEEE Computer Society, 11(3), pp. 580–597. doi: 10.1109/TSC.2017.2730882.

IBM Institute for Business Value (2013) *Analytics : The real-world use of big data in financial services - How innovative banking and financial markets organizations extract value from uncertain data*. Available at: www.sbs.ox.ac.uk.

Ivanov, T. and Singhal, R. (2018) "ABench", in *Companion of the 2018 ACM/SPEC International Conference on Performance Engineering - ICPE "18*. New York, New York, USA: ACM Press, pp. 13–16. doi: 10.1145/3185768.3186300.

James, M. *et al.* (2011) "Big data: The next frontier for innovation, competition, and productivity", *McKinsey Global Institute*, (June), p. 156. doi: 10.1080/01443610903114527.

Kaplan, R. and Norton, D. (1992) "The balanced scorecard: measures that drive performance", *Harvard Business School Pub.* Available at: http://home.bi.no/fgl99011/bok2302/MB92.pdf.

Marjani, M. *et al.* (2017) "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges", *IEEE Access*, 5, pp. 5247–5261. doi: 10.1109/ACCESS.2017.2689040.

Marr, B. and Schiuma, G. (2003) "Business performance measurement – past, present and future", *Management Decision*. MCB UP Ltd, 41(8), pp. 680–687. doi: 10.1108/00251740310496198.

Neely, A. (2002) *The Performance Prism: The Scorecard for Measuring and Managing Business Success The emergence of the redistributed manufacturing (RDM) concept: The role of big data in the consumer goods industry View project EC-HVEN View project*. Available at: https://www.researchgate.net/publication/265453886.

Pappas, I. O. *et al.* (2018) "Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies", *Information Systems and e-Business Management*. Springer Berlin Heidelberg, 16(3), pp. 479–491. doi: 10.1007/s10257-018-0377-z.

Patil, S. *et al.* (2011) "YCSB++", in *Proceedings of the 2nd ACM Symposium on Cloud Computing - SOCC "11*. New York, New York, USA: ACM Press, pp. 1–14. doi: 10.1145/2038916.2038925.

Pernici, B., Francalanci, C., Geronazzo, A., Polidori, L., Cattaneo, G., *et al.* (2018) *Industry Requirements with benchmark metrics and KPIs*. Available at: https://www.databench.eu/wp-content/uploads/2019/01/databench-d1.1-ver.1.0.pdf.

Pernici, B., Francalanci, C., Geronazzo, A., Polidori, L., Ivanov, T., *et al.* (2018) "Relating Big Data Business and Technical Performance Indicators", in *Conference of the Italian Chapter of AIS*, pp. 1–12.

Sivarajah, U. *et al.* (2017) "Critical analysis of Big Data challenges and analytical methods", *Journal of Business Research*. The Authors, 70, pp. 263–286. doi: 10.1016/j.jbusres.2016.08.001.

Supakkul, S., Zhao, L. and Chung, L. (2016) "GOMA: Supporting big data analytics with a goal-oriented approach", *Proceedings - 2016 IEEE International Congress on Big Data, BigData Congress 2016*, pp. 149–156. doi: 10.1109/BigDataCongress.2016.26.

Urbinati, A. *et al.* (2018) "Creating and capturing value from Big Data: A multiple-case study analysis of provider companies", *Technovation*. Elsevier Ltd, (May), pp. 1–16. doi: 10.1016/j.technovation.2018.07.004.

Wang, L. *et al.* (2014) "BigDataBench: A big data benchmark suite from internet services", in *2014 IEEE 20th International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, pp. 488–499. doi: 10.1109/HPCA.2014.6835958.

Yin, S. and Kaynak, O. (2015) "Big Data for Modern Industry: Challenges and Trends", *Proceedings of the IEEE*. IEEE, 103(2), pp. 143–146. doi: 10.1109/JPROC.2015.2388958.