IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 26, 2020, accepted June 16, 2020, date of publication June 29, 2020, date of current version July 8, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3005461*

# Critical Success Factors for Big Data: A Systematic Literature Review

**ZAHER ALI AL-SAI, ROSNI ABDULLAH, AND MOHD HEIKAL HUSIN, (Member, IEEE)**
School of Computer Science, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia
Corresponding author: Zaher Ali Al-Sai (z_alsai@yahoo.com)

**ABSTRACT** During the last few decades, many organizations have started recognizing the benefits of Big Data (BD) to drive their digital transformation and to gain faster insights from faster data. Making smart data-driven decisions will help the organizations to ride the waves toward invaluable investments. The successful implementation of Big Data projects depends on their alignment with the current organizational, technological, and analytical aspects. Identifying the Critical Success Factors (CSFs) for Big Data is fundamental to overcome the challenges surrounding Big Data Analytics (BDA) and implementation. In recent years, the investigations related to identifying the CSFs of Big Data and Big Data Analytics expanded on a large scale trying to address the limitations in existing publications and contribute to the body of knowledge. This paper aims to provide more understanding about the existing CSFs for Big Data Analytics and implementation and contributes to the body of knowledge by answering three research questions: 1) How many studies have investigated on Big Data CSFs for analytics and implementation?, 2) What are the existing CSFs for Big Data Analytics, and 3) What are the categories of Big Data Analytics CSFs?. By conducting a Systematic Literature Review (SLR) for the available studies related to Big Data CSFs in the last twelve years (2007-2019), a final list of sixteen (16) related articles was extracted and analyzed to identify the Big Data Analytics CSFs and their categories. Based on the descriptive qualitative content analysis method for the selected literature, this SLR paper identifies 74 CSFs for Big Data and proposes a classification schema and framework in terms of 5 categories, namely Organization, Technology, People, Data Management, and Governance. The findings of this paper could be used as a referential framework for a successful strategy and implementation of Big Data by formulating more effective data-driven decisions. Future work will investigate the priority of the Big Data CSFs and their categories toward developing a conceptual framework for assessing the success of Big Data projects.

**INDEX TERMS** Big data (BD), big data analytics (BDA), big data implementation, critical success factors (CSFs), big data challenges, information system success, readiness, maturity, systematic literature review (SLR).

## I. INTRODUCTION

In recent years, Big Data (BD) has become the main tech talk between academia and practitioners in the digital competitive play yard. BD is an important asset that attracts the attention of many CEOs in different organizations to gain faster insights and high revenue [1]. The journey of BD started when many organizations recognized that the large volume of their data exceeds the capabilities of their organizations, process, capacity, structure, technology infrastructure, and governance. They struggled to deal with the requirements for analyzing the high volume of various data [2]–[4]. In 2012, Gartner formulated the ''Big Data'' term to describe the high volume of various data from several sources. Subsequently, in 2013, they defined Big Data as ''a high-volume, high-velocity and high-variety information assets that demand cost effective,

innovative forms of information processing for enhanced insight and decision making'' [5], [6]. Industrial Development Corporation (IDC) and EMC Corporation reported that the amount of created data in 2020 will be exceeding 40 zettabytes (ZB) which is greater than the amount of data in 2009 with a scale of 44 times [5]–[7].

Researchers have identified various Big Data related capabilities and resources as a potential solid base to improve organizational performance. Most of the current research in

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Hua Chen.

the Big Data Analytics (BDA) domain covers the technology dimensions, talent, and management that affect organizational performance [75]. The organization's capability to gain the benefits from various forms of massive data is essential and the readiness for investments in Big Data is now at the center of attention [55], [56].

Despite the benefits of BD and BDA, the Big Data projects often represent one of the largest risky and costly investments that need serious planning [8]. Big Data projects differ from other technology projects, where the implementation of Big Data projects require more preparation [9]. The implementation of Big Data is a complex commitment that needs a new technological and organizational approach [9]. Organizations are at the crossway where they can handle the challenges related to Big Data implementation and become Big Data winners and reap benefits from Big Data Analytics or to struggle with Big Data characteristics and challenges to become a loser in Big Data market [3], [10]. On the other hand, organizations need to improve their ability to manage the challenges and the risks related to Big Data investment, if they want to survive with a high competitive advantage [3], [10].

Many organizations often adopt several methods for identifying and assessing the business value of their Big Data investments. For example, [11] adopted the approach of ''Benefit-Risk Analysis'' whilst [12] used the Return On Investment (ROI) approach to show the business value of Big Data investment. The Critical Success Factors (CSFs) approach has been adopted by many authors [3], [8], [13]–[16]. The CSFs are considered as a powerful and applicable method used to address many of the challenges related to implementing Information Technology (IT) in several domains [17]. The origin of the ''Success Factors'' concept was introduced by D. Ronald Daniel in the early 1961s [3], [17]–[19]. Daniel highlighted the importance of the factors that determine organizational success in addressing the risk of data overload. In 1978, [20] expanded Daniel's research to develop the CSFs approach as a way to systematically identify and define the requirements for implementation [20]. Reference [20] introduced the CSFs as elements or characteristics that should be present in a particular context and if satisfactory, will lead to a competitive advantage in the organization's performance [3], [18]. The CSFs are considered as a key area where everything will be done in the right way through the business process to achieve the organization's objectives [21]. Also, the CSFs are related to what the organization could do to improve delivering a successful project [22]. CSFs analysis guides the organization in the identification of potential critical factors that can be used to successfully gain the value of BD investments. The identification of CSFs is considered a business management approach used to support the Information Systems (IS) strategic planning and readiness assessment [3]. Examples of the major areas that utilize the CSFs approach are: project management with a percentage of (63.49 %), Information Systems (IS) implementation with (49.21 %), and requirements with (47.62 %). Consultants and Information System departments used the CSF's approach to support and develop the IT strategic planning [17], [25].

The CSFs method is used for identifying the critical elements of success which is originally developed to align the technology planning with the strategic planning of an organization [23], [18] [3], [17]. CSFs as best practices are frequently identified after the successful achievement of certain activities. As a consequence, these success factors are close to real-life experiences [24]. On another hand, the use of CSFs can have a major impact on the readiness, maturity, design, development, and implementation of IS [17]. The CIOs could use the success factors to identify the information that is most important to make critical decisions about their organization [17], [21]. The decisions in this manner should be more effective and accurate when made based on data that is related to the organization's success factors. Most decisions-makers should be aware of the factors they must address to be successful, otherwise only when problems appear the origin causes will be identified and these critical factors will be clear [17], [21].

In the context of developing a comprehensive IS success model for the Big Data domain, there is a need to investigate on determining the factors influencing the success of the Big Data implementation [23], [27]. The activities regarding determining the critical factors could transform the organization towards a successful implementation of Big Data new technology [28], [29]. Since Big Data systems are related to Data Warehouse and Business Intelligence systems, most of the CSFs investigated in Big Data system were adopted from these technologies [9], [23], [27].

Many organizations adopt the CSFs methods to assess the risk of investing in Big Data project before starting the implementation [3]. For example, CSFs could be used to assess and identify the business value of Big Data projects [3]. Also, understanding the CSFs of Big Data could lead the organizations to be successful in the Big Data journey and to achieve their objectives from Big Data investments [3], [10]. Most of the existing CSFs for Big Data have been adopted according to the requirements which originated from the Big Data challenges [3], [31]. The CSFs related to readiness phase, the availability of adequate resources, and appropriate people have positive effects on the project's outcome [32]. At the implementation phase especially with the absence of data scientists, most of the experts in Big Data are career changers from a scientific field, who know the statistical and analytics tools; but lack the management experience and mind-set to make business strategic decisions [24]. The key to the successful implementation of Big Data in the organization is not only based on having sufficient technical capabilities to analyze the data. Data management also plays the most important role in the phase of implementation [9], [33]. Also, the availability of an effective assessment tool with a clear layout for the outputs can be used as a base that can enable the managers to make their right decisions toward a successful implementation [33].

There is limited research conducted on the CSFs of the BD domain, where most of the existing literature adopted the CSFs from other related technology like Business Intelligent (BI) and data mining [27]–[31] [10]. It becomes necessary to explore and understand the CSFs for the Big Data projects because Big Data adoption is relatively new and academic literature did not present sufficient systematized knowledge on the topic [9], [22], [36]. There is a dire need to investigate the factors that determine the successful implementation of Big Data [9], [37]. Based on conducting Systematic Literature Review (SLR) by [38], [39], this paper attempts to contribute to the body of knowledge by answering three research questions as:

RQ1: ''How many studies have investigated on Big Data CSFs for analytics and implementation?''

RQ2: ''What are the existing Critical Success Factors for Big Data Analytics?''

RQ3: ''What are the categories of Big Data Analytics CSFs?''.

By reviewing the existing studies that focus on the Big Data Critical Success Factors (CSFs) and their categories, this paper provides a grounded reference with critical insights for the managers to formulate and execute the Big Data strategies and decision making based on measured categories. These categories can be used by the organizations to handle the issues related to identifying the key aspects in Big Data and focusing on the key components of Big Data strategy. Future work will investigate on the priority of the critical factors of Big Data and their categories toward creating a referential platform for implementing overall Big Data projects.

The structure of this paper will be mapped as follows: after the introduction section (I), the review methodology section (II) will highlight the method that will be used to conduct this SLR paper. Then, the findings from the reporting stage will be discussed in section (III) to identify the existing literature, the CSFs for Big Data, and their categories. Next, the proposed classification framework for CSFs of Big Data Analytics will be provided in section (IV). Finally, a conclusion and future work will point out some suggestions for future direction.

## II. REVIEW METHODOLOGY

This paper performed a Systematic Literature Review (SLR) method to identify the existing literature relevant to Big Data CSFs and their categories. The SLR is a suitable, reliable and accurate technique and method that can be used to identify and evaluate all existing research related to a particular research question, topic domain, or phenomenon of interest [38]. While the individual studies that contribute to a systematic review are called primary studies, the systematic review considers as a form of secondary study that can be used in this research paper to answer the research questions [22], [38], [39].

The SLR strategy in this paper was conducted based on the guidelines for performing Systematic Literature Review in Software Engineering by [38], [39]. The SLR consists of three main stages: Planning the Review, Conducting the Review, and Reporting the Review as shown in Figure 1.

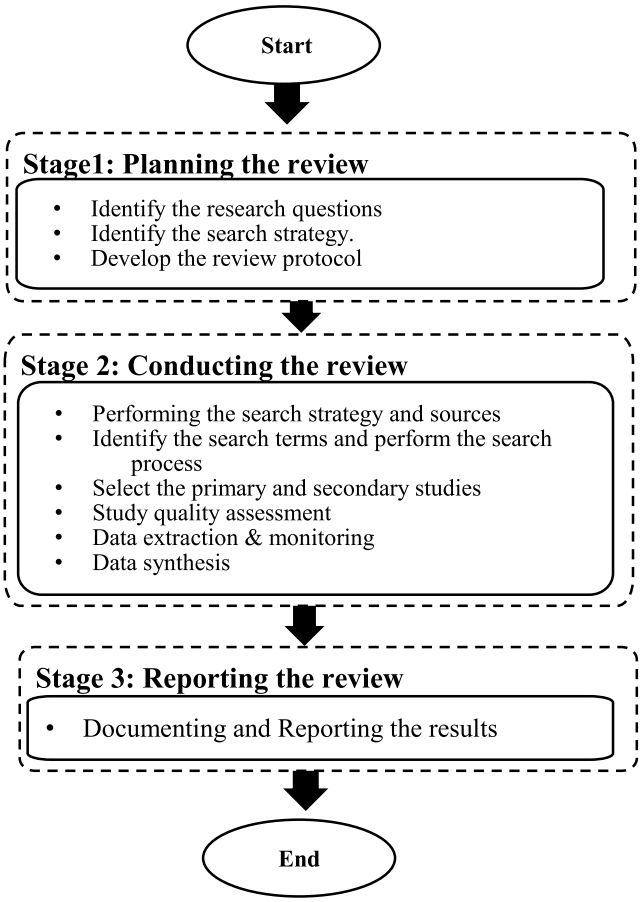



**FIGURE 1.** Systematic Literature Review Phases by [38].

### A. STAGE 1: PLANNING THE REVIEW

The activities associated with the stage of planning include identifying the need for the review, formulating the research questions, identifying the search strategy, and develop the review protocol that will be used to formulate the research question which will be addressed and the review methodology which will be used to perform the review [38]. This SLR will answer three research questions: RQ1: ''How many studies have investigated on Big Data CSFs for analytics and implementation?'', RQ2: ''What are the CSFs of Big Data Analytics?'', and RQ3: ''What are the categories of Big Data Analytics CSFs?''.

The research strategy is the initial mapping study that helps to determine the appropriate research strategy. Systematic Literature Reviews are performed based on a pre-defined search strategy that aims to identify the literature relevant to the SLR research question. The strategy aims to identify the primary studies including search key terms and resources to be used in this SLR as shown in section (II-B-2). The most common approach to identify the term is to break down the research question into individual terms. Then draw up a list of synonyms, abbreviations, and alternative spellings. More

terms can be obtained by considering subject headings used in databases and journals. Advanced search strings can then be constructed using Boolean AND's and OR's [38] as shown in section (II-B-2). This SLR paper used the database search method in the first search strategy as recommended by [26]. Resources include electronic search engines, journals, conference proceedings, digital libraries, and grey literature [38].

According to [30], the most common and key validity threats is by missing of related primary studies' which can result from threats as incomplete search terms, incorrect search method, restricted time span, incorrect or in automatic search, identification error of primary studies in the searching process, bias, and database /paper inaccessible. After identifying the related papers using the database search method, a snowballing backward method was used based on the reference lists in the papers found from the database search. The purpose of using both search methods was to avoid the bias [30] at the search stage as recommended by [26]. The conducting stage highlights the literature sources and search keywords used in this SLR.

### B. STAGE 2: CONDUCTING THE REVIEW

The activities associated with conducting the review include performing the search strategy and sources, the selection of primary studies, study quality assessment, data extraction & monitoring, and data synthesis [38]. In this research, the SLR method was used to evaluate all the available research related to the pre-defined research questions in the planning stage. The following sub-sections will provide more explanation about conducting the review.

### C. SOURCE

A total of nine databases sources were used as final main sources to identify the available literature relevant to the research questions, namely Research Gate, Google, Google Scholar, EBSCOhost, Science Direct, Scopus, Springer, USM Library Repository, and IEEE. Based on this SLR, this paper selected these nine databases as the main sources for the search because the results of the initial research have shown that these databases have a good repository that could be used in this SLR to return all of the available numbers of literature. Also, the search has shown that the other databases referred to these databases as a main source for the existing literature. These main sources were used to locate published journals, papers, conference proceedings, IEEE bulletins, book chapters, symposiums, gray literature, and workshops.

### D. SELECTION CRITERIA (INCLUSION AND EXCLUSION CRITERIA)

The selection of the relevant Literature is governed by the inclusion and exclusion criteria in the context of Big Data CSFs and their categories. Using various forms of search terms derived from the pre-identified research question, this paper identified several keywords were: ''Big Data'', ''Big Data Analytics'', ''Big Data projects'', ''success'', ''Critical Success Factors'', ''CSFs'', ''issues'', ''challenges'', ''barriers'', ''requirements'', ''capabilities'', ''Big Data Implementation'', ''Information System Success'', and ''successful implementation''. A general approach to determine the search strategy is to break down the research question into individual terms. Advanced search strings can then be executed using Boolean ''ORs'' and ''ANDs'' [38], [39] as follows: (Big Data OR Big Data Analysis, OR Big Data implementation) AND (Success OR Maturity OR Readiness) AND (critical success factors OR empirical OR success model OR factors OR barrier ).

**TABLE 1. Study quality assessment criteria.**

| No. | Item | Answer |
|---|---|---|
| Q1 | Is there a clear description of the objectives of the investigation? | Yes/No/ Partially |
| Q2 | Did the paper adequately explained the research method? | Yes/No/ Partially |
| Q3 | Is the paper supported by primary data? | Yes/No/ Partially |
| Q4 | Did the paper clarified and detailed the model constructs and structure? | Yes/No/ Partially |
| Q5 | Did the paper identify the source of the factors? | Yes/No/ Partially |

This research process included publications in the context of Big Data or Big Data Analytics, written in ''English'', and were published between the years of 2007 and 2019. Due to the novel concept of Big Data, there are no academic publications about the Big Data and CSFs of Big Data before the year 2007 as assumed by [3]. An additional filter time was applied to include the literature published between 2007 to 2019.

Each identified paper must focus on the critical factors influencing the success of Big Data, Big Data projects, or Big Data Analytics. Any papers which did not respond to the inclusion criteria will be excluded. However, due to the limited number of related papers, further criteria were specified to choose relevant paper for inclusion in this review which is as follows: the paper should focus either on (A) the critical factors that influence the success of Big Data in a case study or (B) implementation factors influencing the success of BD implementation, or (C) critical factors influencing the success of Big Data Analytics. In addition, this SLR reviewed the various types of papers including (SLR, Content analysis, Survey, Case study, Empirical Case study, Meta-Analysis).

### E. QUALITY ASSESSMENT

The ''quality assessment'' is considered a critical step to assess the quality of the selected literature. Quality assessment includes questions aimed to assess the scope to which reviewed articles have addressed bias and internal and external validity [39]. Table 1 shows the five questions (Q1-Q5) of quality assessment, where each answer has only three options: Yes=1; Partially=0.5; and No=0.

### F. STAGE 3: REPORTING THE REVIEW

The stage of ''Reporting'' is a single-stage phase [38]. In the reporting stage, the findings were reported in the following discussion section.

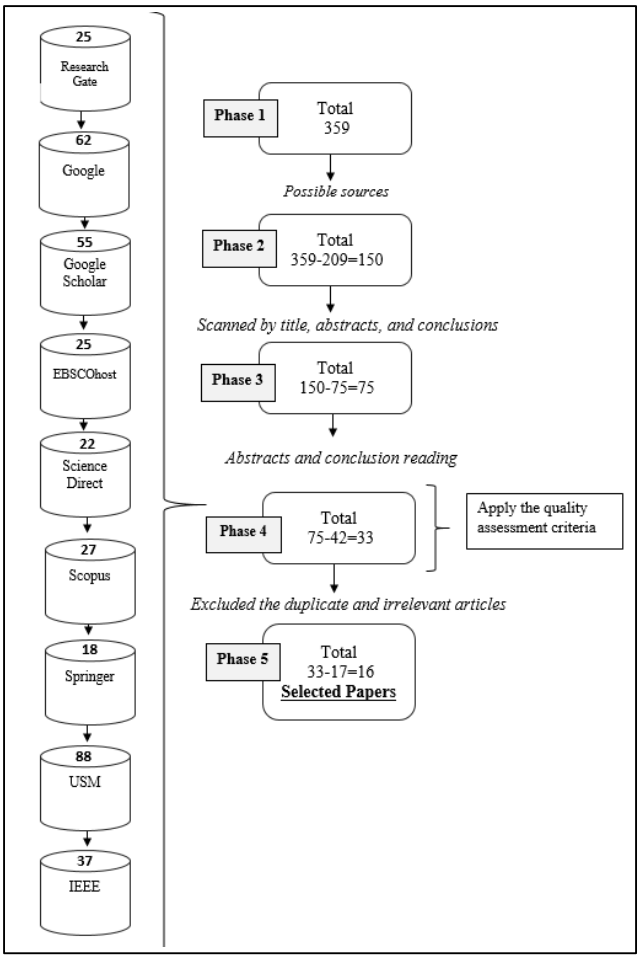



**FIGURE 2.** The Search Process and Results of Paper Selection.

## III. FINDINGS AND DISCUSSION

The findings of this paper aim to answer our specific research questions that guided this SLR. The study selection that applied in this SLR consisted of 5 phases. Starting with (1) extracting a total of 359 relevant papers from the digital search as possible sources (n=359). (2) 150 studies were scanned by title, abstracts, and conclusions (n=150). (3) Followed with abstracts and the conclusion reading for 75 papers resulted in related studies before the quality assessment (n=75). (4) Followed with a detailed view of abstracts and the full text of the returned articles where 33 papers resulted in the quality assessment(n=33). Finally, (5) after excluding the duplicate and irrelevant articles using exclusion criteria and filtering the results of the quality assessment stage for all papers, only 16 articles were accepted and identified as final sources for the data synthesis(n=16). Figure 2 shows the search process, results, and the process of paper selection.

The number of selected papers published per year has been quite limited. The average quality score appeared to be increased in the year of 2016 but decreased again in 2017 and 2018. Table 2 shows the number of studies and the percentage of publication per publication date.

**TABLE 2.** Distribution of publications \ per Year.

| Year | Number of Studies | Percentage (%) | References |
|---|---|---|---|
| 2014 | 2 | 12% | [8] [16], |
| 2015 | 3 | 19% | [31], [40], [35] |
| 2016 | 6 | 37% | [22], [9], [14], [41], [15] |
| 2017 | 3 | 19% | [3], [27], [42] |
| 2018 | 2 | 13% | [36], [43] |

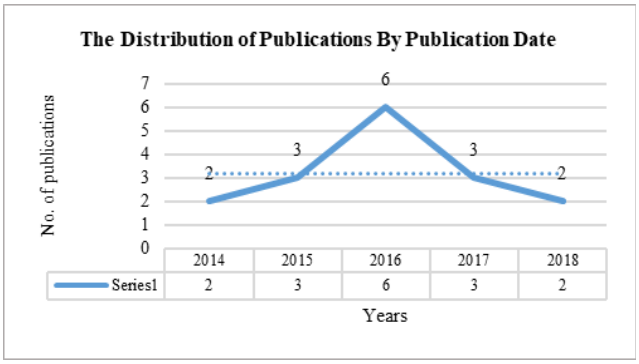



**FIGURE 3.** The Distribution of Publications by Publication Date.

**TABLE 3.** Quality assessment score.

| Quality Scale | Very Poor (<1) | Poor (1 - <2) | Good (2 - <3) | Very Good (3-4) | Total |
|---|---|---|---|---|---|
| Number of papers | 0 | 0 | 5 | 11 | 16 |
| Percentage (%) | 0 | 0 | 31 | 69 | 100 |

Figure 3 shows how the distribution of publications changes per year.

Sixteen (16) articles were selected as high-quality articles that could be used to answer the research questions in this SLR. As shown in Table 3, eleven (11) articles were rated (69%) as very good quality; five (5) articles were rated (31%) as good, and any other poor quality articles were excluded from the results. Table 3 shows the final results of the quality assessment score.

Based on the quality assessment questions (Q1-Q5) that are pre-defined in the review methodology, Table 4 shows the summary of the quality assessment for the (16) papers (A1-A16) selected for review in this SLR.

In the following sub-sections, we discuss the answers to this SLR three research questions: A. (RQ1): How many studies have investigated on Big Data CSFs for analytics and implementation? B. (RQ2): What are the CSFs of Big Data Analytics? C. (RQ3): What are the categories of Big Data Analytics CSFs?

### 1) (RQ1): HOW MANY STUDIES HAVE INVESTIGATED ON BIG DATA CSFs FOR ANALYTICS AND IMPLEMENTATION?

In recent years, many researchers conducted different investigations in order to understand and identify the various CSFs

**TABLE 4.** Quality assessment results.

| ID | Q1 | Q2 | Q3 | Q4 | Q5 | Total |
|---|---|---|---|---|---|---|
| A1 | 1 | 1 | 1 | 0.5 | 1 | 4.5 |
| A2 | 1 | 0.5 | 1 | 0.5 | 1 | 4 |
| A3 | 1 | 0.5 | 1 | 0.5 | 1 | 4 |
| A4 | 0.5 | 1 | 0.5 | 2.5 | 0.5 | 2.5 |
| A5 | 2.5 | 0.5 | 0.5 | 0.5 | 0.5 | 2.5 |
| A6 | 1 | 1 | 1 | 1 | 1 | 5 |
| A7 | 1 | 1 | 1 | 0.5 | 0.5 | 4 |
| A8 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 2.5 |
| A9 | 1 | 0.5 | 0.5 | 0.5 | 1 | 3.5 |
| A10 | 1 | 0.5 | 1 | 1 | 1 | 4.5 |
| A11 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 2.5 |
| A12 | 1 | 1 | 1 | 1 | 1 | 5 |
| A13 | 1 | 1 | 1 | 1 | 1 | 5 |
| A14 | 1 | 1 | 0.5 | 1 | 1 | 4.5 |
| A15 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 2.5 |
| A16 | 1 | 1 | 0.5 | 1 | 0.5 | 4 |

and their categories for BD. In this section, we identified the results related to the first research questions in the planning stage ''RQ1: How many studies have investigated on Big Data CSFs for analytics and implementation?''. Based on conducting a content analysis method and the quality assessment approach for the selected literature, sixteen (16) articles were selected as high-quality literature that could be used to answer the specific research questions in this SLR.

The studies by [3], [8], [9], [14]–[16], [22], [27], [31], [36], [37], [40]–[44], have investigated the factors that influence the successful implementation of Big Data. In 2014, a study by Evers investigated the CSFs of BDA systems and Business Intelligence [8]. Evers conducted a prospective analysis of multiple case studies that were obtained by structured interviews with experts in the area of Business Intelligence and Big Data Analysis to identify their CSFs. Likewise, [8] adopted the qualitative research method to link the critical success factors of BI & BDA from literature, with the perceived CSF's in the practical field of Business Intelligence, Big Data Analytics, and Data warehousing. Based on the challenges of Big Data, [8] classified the CSFs in two main categories of organizational and technology. A proposed model was included in [8] study to represent the organizational and technological factors of BI and BDA projects. While [8] did not identify a category of ''People'' to classify any of the CSFs that were identified in his study, [12] suggested to include ''People'' category in [8] study where the ''Organization'' category contains the same kind of CSFs that are related to the ''People'' category like the balanced team composition, team member skills, and the talent management [12].

Further to the point, both [16] and [31], adopted a qualitative research approach using a literature review to identifying six preliminary CSFs for Big Data projects as following (Information Strategy for Big Data; Identifiable Business Value; Top Management Support; Skills for Big Data projects; Information Quality, Security and Integrity; Technological Capability). More specifically, [16] categorized the Big Data success factors from the perspective of '' People'', ''Process'', and ''Technology'' [12], [16]. Building on the research that was carried out by [16], [31] identified the Critical Success Factors (CSFs) based on the requirements that emerge from the Big Data challenges (People, Technology, Process). Reference [31] identified and categorized the critical success factors regarding these challenges and distributed them over the lifecycle of the project.

From another perspective, the literature review by [40] exposed some of the Big Data challenges which would be highlighted as indicators that represent the success factors if selected properly. In addition, they would be considered as barriers or obstacles that prevent Big Data implementation. [40] identified six critical success factors which included the (Top management support, Organizational change, Data availability and Quality, Infrastructure, The required skillset, Privacy, and Security) but they did not identify any categories or details about these CSFs [40].

Reference [37] identified five dimensions for Big Data that enabled business value issues. A study by [3] considered these dimensions as critical factors were effecting the successful implementation for BD are as follows: (i) Organizational change and talent; (ii) technology and techniques; (iii) Data policies; (iv) Industry structure; and (v) Access to data. Reference [37] focused on change management involvement for Big Data and on the need for data policies and procedures as essential elements of the proposed CSFs. Reference [37] did not offer any further explanation of the meaning of their dimensions. The ''People'' category refers to the analytical and statistical skills required for talent working on Big Data projects which referred to the first dimension ''Organizational change and talent'' [37]. The dimensions of ''Technology and Techniques'', ''Data policies'' and ''Access to data'' that are related to the technological elements could be considered under the category of ''Technology''. The dimension of ''Industry structure'' and ''Organization change'' could be referring to the ''Organization'' category. Based on the meaning of these categories from existing literature, the study by [37] enveloped three categories and dimensions for Big Data CSFs as ''Technology'', ''Organization'', and ''People''.

Likewise, a study conducted by [22], highlighted the necessity for the CSFs with consideration to the categorization of (People, Process and Technology) in the context of Critical Success Factors for Big Data project implementation. Reference [22] proposed a project methodology adequate for the implementation of Big Data. The proposed methodology included the CSFs that follows a repeated process model that classifies the project according to complexity and clarity. Moreover, [22] assumed six categories as success factors for BD projects which included Data (Security, Ownership, Access to data), Process (Change and Project management), Governance (Performance management and Culture), Teams (Team building and Skill level), Tools (Technological side such as Infrastructure), and Objectives (Clear goals and

strategy). Reference [22] didn't include the organization's elements as critical success factors but highlighted the importance of the organization dimension and suggested that the methodology should include organizational capacity assessment and capacity building as a future work [22].

Based on studies carried out by [9], the critical factors that may impact the successful implementation of Big Data system were grouped into eight (8) main categories namely: Enterprise Data Management, Legal, Management, Organization, Corporate Culture, Scoping, Skills, Technology. The factors were identified based on an inductive coding procedure to the research material through content analysis for the related publications by software vendors and consulting organizations and academic scholars [9].

From another perspective, [14] examined the capabilities of Big Data analytics (BDA) of organizations with considerations to the effective role of Data management in the case study of the commercial vehicle industry. More specifically, [14] proposed a model with seven (7) CSFs for analytical capabilities. The seven CSFs were not identified in Nieder literature but the dimensions as ''Technological capabilities'', ''Strategic/organizational'', and ''Human'' are identified. Based on a review conducted by [12], the ''capabilities'' categories for the CSFs identified by [3] as (Strategic or organizational capabilities, Human capabilities, and Technological capabilities) could be considered similar to the three categories of People, Process and Technology identified by [16] using different labels for the categories [3].

A study conducted by [44], revealed five strategies for being successful with Big Data Analytics in the specific area of the healthcare as the following: Implementing governance for Big Data, Developing a culture for information sharing, Training key people to use Big Data analytics, Combine cloud computing into the organization's Big Data analytics projects, and Generating new business ideas from Big Data analytics. [44], categorized and linked the CSFs in their research to the three key dimensions: IT, People, Process as a foundation to harvest the IT business value from Big Data [41], [44] studied the barriers to Big Data development in a case study of Colombia and proposed a reality gap model to assess the dimensions of Big Data development. The model presented a structured framework that displays a set of factors affecting the development of Big Data. Where the CSFs could be derived from the challenges, the study by [41] used a gap analysis method to identify the challenges to Big Data adoption. Reference [41] considered these challenges as a CSFs of Big Data which included: Information, Technology, Process, Objectives and Values, Skills and Knowledge, Management System and structure, Other Resources (such as time and budget) [3], [41].

Furthermore, [15] identified seven (7) Critical Success Factors (CSFs) in Big Data innovation as following: Effective value discovery process, Direct involvement of the CEO, Service-orientation mindset, Not blindly following IT fashion, The architectural foundation for integration and growth, Confident outsourcing & Vendor management, and talent planning. These factors help organizations to innovate with Big Data and to gain faster business value [3], [15] identified their CSFs without any classification. Reference [3], proposed categories for the CSFs that were identified by [15] as generic categories as following: Organization, which includes Process-related factors (effective value discovery process; direct involvement of top management; customer-centric focus; the strong business need for Big Data implementation); Technology (technology infrastructure to handle data growth and data integration; confident outsourcing and vendor contract management); People (skills and talent planning) [3], [15].

The study by [3] focused on the identification of the categories of Big Data Critical Success Factors (CSFs) published in both journals and conference proceedings. Eybers adopted a qualitative approach using a systematic review for grouping the factors into categories and proposed a preliminary model of CSFs categorization. The model assumed that CSFs can be classified into three main categories as Organization, People, and Technology [3].

Likewise, [27] conducted a qualitative SLR technique to identify ten factors influencing the implementation success of BDA. These factors considered as CSFs for Big Data as following; Organization Capability, Technology Capability, Analytics Capability, Human Capability, Analytics Culture, Environment, Data Management, Data and Information Quality, System Quality, Perceived Benefit [27].

Another study by [42] investigated the factors for promoting the implementation and usage of Big Data in a case study of healthcare in Korea using an analytical hierarchy process (AHP) [42] proposed a research framework that considered 15 factors and four dimensions as follows Organizational analytics capability, Data privacy and security, Data quality, Data standardization, Data experts, and Organizational application capability.

Reference [36], conducted an exploratory investigation to identify the factors relevant to the adoption of Big Data by a case study in virtual retailers as one of the main users of Big Data. Reference [36] identified factors related to Strategy, Processes and leadership, Human resources, Implementation management, and Ethics and privacy as potentially relevant factors for the adoption of Big Data technologies namely, Strategic alignment, Top management involvement, Compatibility with organizational processes, Information sharing, Cultural change for data orientation and Experimentation, Overcoming the talent gap, Structure and functions for implementation Communication, Expectations management, Agile methods, and Privacy management [36]. These factors were not considered before in the previous studies as a critical success factor for Big Data. The data regarding the critical success factors was collected through the content analysis technique and using a semi-structured interview in a case study [36].

Another study by [43] highlighted some of the most important factors related to BDA. Reference [43] followed the Technology, Organization, and Environment (TOE)

**TABLE 5.** Related works on BD CSFs.

| No. | Reference | Source | Type | Database | Domain | Focus Area | Source of CSFs | Type of study | Data Collection Method | Analysis Method |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 | [8] | Master Thesis- Tilburg University | Thesis | Academia | Big Data and Business Intellie (BD and BI) | CSFs of Big Data | Challenges | Empirical Case study | Structured interviews | Content analysis of multiple cases |
| A2 | [16] | Proceedings - Pacific Asia Conference On Information Systems, PACIS 2014 | Conference | Google-Research Gate | Big Data projects success | CSFs of Big Data | Requirements and Challenges | Meta- Analysis | Content Analysis | ABC Analysis |
| A3 | [31] | Twenty-First Americas Conference On Information Systems | Conference | Semantic Scholar | Big Data Analytics projects | CSFs of Big Data | Requirements and Challenges | Content Analysis | Content Analysis | ABC Analysis |
| A4 | [40] | Journal of International Technology and Information Management | Journal | USM Library Repository + EBSCOhost+ Research Gate+ Semantic Scholar+ Google Scholar | Successful Implementation of Big Data in Organizations | Big Data implementation | Challenges or barriers of BD | Conceptual Proposed Model | Survey | NA |
| A5 | [35] | International Journal Production Economics | Journal | Science Direct | Business value for Big Data | Big Data capability | Value creation from 'big data' | Empirical longitudinal case study | Semi- structured interviews, | In-depth analysis of the longitudinal case study |
| A6 | [22] | 2016 IEEE International Conference On Big Data | Conference | IEEE+ Research Gate+ Google Scholar | Big Data Project Success | Key success factors (KSFs) of Big Data Projects | Big Data Challenges and Requirements | Content Analysis | Systematic Literature Review | SLR Based on Wixom and Watson model |
| A7 | [9] | 2015 11th International Conference On Innovations In Information Technology | Conference | IEEE+ Research Gate | Implementation Factors affecting the success of BD Systems | Big Data Systems | Big Data Project Success Factors | Empirical, Content Analysis | Systematic Literature Review | Content Analysis |
| A8 | [14] | MSc Dissertation | Dissertation | Google Scholar | Development Of Critical Success Factors and an Analysis Of Firms' Capabilities In The Automotive Industry | CSFs of Big Data Management | Big Data Capabilities | Case Study (Qualitative Study) | Interviews Survey | NA |
| A9 | [44] | Technological Forecasting And Social Change – Elsevier | Journal | Science Direct | Understanding the Big data analytics capabilities and potential benefits for healthcare organizations | Big data analytics capabilities | Big Data Analytics capabilities -The strategies for success with big data analytics -The Dimensions of Big Data CSFs | Case Study (quantitative approach) - Empirical. - Content Analysis (comprehensive literature review) | Comprehensive literature review Case Study | Content Analysis |
| A10 | [41] | Development Informatics - Working Paper | Working Paper | | Measuring the Barriers to Big Data for Development: Design-Reality Gap Analysis in Colombia's Public Sector | Barriers of Big Data Development | Challenges of BD | Systematic Analysis+ Content Analysis | Systematic analysis for Literature | Gap Analysis and Content Analysis |
| A11 | [15] | Proceedings Of The Annual Hawaii International Conference On System Sciences 2016 | Conference | Academia | Innovating with Big Data at Lufthansa (Empirical multiple case study of 25 European enterprises) | CSFs of Big Data | BD Challenges | Empirical exploratory research + Case Study | Case Study+ Semi-structured and documents analysis (slides, internal technical reports, use cases, etc.) grounded theory method | Case Analysis |
| A12 | [3] | 2017 Ist-Africa Week Conference (IST-Africa) | Conference | Research Gate | Critical success factor categories for big data: A preliminary analysis of the current academic landscape | CSFs of Big Data | Challenges | Case Study + Conceptual Model | Systematic Literature Review | A meta-analysis approach |
| A13 | [27] | 2017 International Conference On Research And Innovation In Information Systems (ICRIIS) | Conference | Google Scholar | Factors influencing the implementation success of BDA | Big Data implementation factors | Factors Influencing the successful implementation of Big Data | Empirical study +Content Analysis | Systematic literature + Review (SLR) guidelines by (Kitchenham, 2007) and (Okoli,2010) | Content Analysis |
| A14 | [42] | Information Development- Sage | Journal | Research Gate\ Google Scholar | Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare | CSFs of Big Data | Challenges and Barriers | Content Analysis+ Expert opinion + Conceptual model+ Case Study (healthcare Korea)+ Questionnaire | Expert opinion + Case Study (healthcare Korea)+ Questionnaire | Analytical hierarchy process (AHP) and Excel Template |
| A15 | [36] | Review Of Business Management | Journal | Research Gate | CSFs for Big Data adoption in Virtual retail | CSFs of Big Data | Challenges | Exploratory Study+ Empirical Case Study + Content Analysis | Literature Review+ Case Study through Semi-Structure Interview | Content Analysis |
| A16 | [43] | International Conference On Enterprise Information Systems (ICEIS 2018) | Conference | Google Scholar | Interdependencies of contextual factors | Big Data Analytics | TOE Framework | Empirical study | Online survey+ pilot study ( 1-5 Likert scales) | Confirmatory factor analysis |

** *Source: SLR and Compilation by Author*

**TABLE 6.** Related works ON BD CSFs and its categories.

| No. | Authors | CSFs category | CSFs |
|---|---|---|---|
| A1 | [8] | Organizational | • Vision and Business Case<br>• Top Management<br>• Championship and Support<br>• Change Management<br>• Project Management and Methodology<br>• Team Skills and User Participation<br>• Resource Relate Factors. |
| | | Technology | • System<br>• Infrastructure and Application<br>• Data Source<br>• Team Skills Related Facets for Success |
| | | Performance | • Technology Performance<br>• Quality of The System<br>• Information Quality<br>• Accurate Query Time.<br>• Organizational and Process Performance<br>• Systems Usages/User Acceptance<br>• Operational Costs and Time Effort |
| A2 | [46] | People | • Team Skills |
| | | Process | • Strategy<br>• Management Support<br>• Documentation<br>• Data Quality<br>• Project Management Related (For Example Project Milestones) |
| | | Technology | • Analytical tools<br>• Documentation<br>• Infrastructure<br>• Integration<br>• System Performance |
| A3 | [31] | People | • Aptitude of Team Members |
| | | Technology | • Technological Tools<br>• Documentation<br>• System Performance. |
| | | Process | • Organizational Strategy<br>• Management Support<br>• Documentation<br>• Project Management Related Items (For Example Project Milestones)<br>• Data Quality. |
| A4 | [40] | N\A | • Top Management Support<br>• Organizational Change<br>• Data Availability and Quality<br>• Infrastructure<br>• The Required Skillset<br>• Privacy, and Security |
| A5 | [35] | Data policies<br>Technology and techniques | • Policies and Procedures<br>• Ability to Gather, Disseminate, and Handle Data |
| | | Organizational change and talent | • Change Management Process<br>• Human Resources Skills |
| | | Access to data | • Availability and Accessibility of Data |
| | | Industry structure | • No Further Explanation Offered by The Authors |

**TABLE 6.** *(Continued.)* **Related works ON BD CSFs and its categories.**

| No. | Authors | CSFs category | CSFs |
|---|---|---|---|
| A6 | [22] | Data | • Data & Data Quality Management<br>• Ownership<br>• Data Integration & Security<br>• Unstructured/Structured Data (Variety, Access.)<br>• Representativeness of Data<br>• Document Collection/Access to Sources |
| | | Governance | • Management Priority / Sponsorship / Support<br>• Big Data Strategy Alignment (With Organization's Vision)<br>• Project Management Process Defined<br>• Well Defined Organizational Structure<br>• Performance Management<br>• Culture of Being Data-Driven<br>• Data Protection and Privacy by Design (Governance) |
| | | Process | • Close Collaboration Between IT and Business<br>• Communication About the Data and Initiatives<br>• Flexibility and Agility<br>• With Freedom for Experimentation<br>• Focus On Change Management<br>• Project Difficulty Explored and Communicated<br>• Clarity of Project Deliverables (Clear or Ambiguous) |
| | | Objectives | • Focus On Small Projects and Known Questions<br>• Specified Business Case<br>• Feasibility Study<br>• Skill Gap Analysis<br>• Well Defined Scope that Understood by The Team<br>• Measurable Project Outcome |
| | | Teams | • Development of Skills / Training<br>• People Skills & Ability to Self-Organize When Needed<br>• Data Science<br>• Technology<br>• Business & Management Skills<br>• Multidisciplinary Team (i.e Across Different Departments)<br>• Stakeholder Coordination / Shared Understanding |
| | | Tools | • Investment in IT Infrastructure, Technology & tools<br>• Investment in Data Sources & Data Storage<br>• Reporting and Visualization Technology<br>• Discovery Technology |
| | | Organizational capacity | Suggested for future work |
| | | Capacity building | Suggested for future work |
| A7 | [9] | Enterprise Data Management | • Master Data Management<br>• Data Quality Management<br>• Integration Solution<br>• Data Security |
| | | Legal | • Data Protection and Privacy by Design |
| | | Management | • Management Priority<br>• Sponsorship<br>• Big Data Strategy Alignment |
| | | Organization | • Close Collaboration between IT & Business<br>• Organizational Structure |
| | | Corporate Culture | • Data-driven Mind-set<br>• Change Management Program<br>• Freedom for Experimentation |
| | | Scoping | • Focus on Small Projects and Known Questions<br>• Specified Business Case<br>• Feasibility Study |

**TABLE 6.** *(Continued.)* **Related works ON BD CSFs and its categories.**

| No. | Authors | CSFs category | CSFs |
|---|---|---|---|
| | | Skills | • Conduction of Skill Gap Analysis<br>• Development of Skills/ Training<br>• Data Science Skills |
| | | Technology | • Investments in Technology<br>• Adequate Selection of Technology and Algorithms |
| A8 | [14] | Technological capabilities | • Capabilities of tools |
| | | Strategic/organizational Capabilities | • Organizational mission and vision |
| | | Human capabilities | • Project team member knowledge and skills |
| A9 | [44] | IT (Technology) | • Implementing (Big) Data Governance Data<br>• Developing an Information Sharing Culture<br>• Training Key Personnel to use Big Data Analytics<br>• Incorporating(Combine) Cloud Computing into the Organization's Big Data Analytics<br>• Generating New Business Ideas From Big Data Analytic |
| | | People | N\A |
| | | Process | N\A |
| A10 | [41] | Information | • Availability and Quality Data is Available and Accessible. |
| | | Technology | • Infrastructure |
| | | Process | • Process Supporting The "Information Value Chain" From The Gathering of Data<br>• Utilization of Information<br>• Decisions Based On the Information<br>• Action Taken Based on the Result and Effective of Final Result. |
| | | Objectives and Values | • Organizational Policies and Procedures |
| | | Skills and Knowledge | • Data Science Capabilities |
| | | Management System and structure | • Structure to Execute BD Strategy |
| | | Other Resources (such as time and budget) | • Time and Budget |
| A11 | [15] | N\A | • Effective Value Discovery Process, (For Technology and Organization)<br>• Direct Involvement of the CEO.<br>• Confident Outsourcing & Vendor Management<br>• The Architectural Foundation for Integration and Growth<br>• Service-Orientation Mind-Set<br>• Not Blindly Following IT Fashion<br>• Talent Planning |
| A12 | [3] | Organization | • Mission, Vision, and Strategy<br>• People<br>• Technology<br>• Aligned to The Overall Objectives of the Organization,<br>• Management Support and Managerial Skills<br>• The Maturity and Readiness of the Organization, Process, and Performance |
| | | People | • Knowledge and Skills of Team Members<br>• Balanced Team<br>• System Users Participation |

**TABLE 6.** ***(Continued.)*** **Related works ON BD CSFs and its categories.**

| No. | Authors | CSFs category | CSFs |
|---|---|---|---|
| | | Technology | • Infrastructure (Hardware and Networking Capability)<br>• Software Applications (such as Security, Flexibility, and Scalability to Name a Few)<br>• Data<br>• Performance |
| A13 | [27] | N/A | • Organization Capability<br>• Analytics Culture<br>• Environment<br>• Perceived Benefit<br>• Technology Capability<br>• Analytics Capability<br>• Human Capability<br>• Data Management<br>• Data and Information Quality<br>• System Quality |
| A14 | [42] | Data | • Establishing Data Standardization<br>• Opening Data Between Inter-and Intra-Organizations<br>• Securing Data Quality<br>• Enhancing Data Privacy And Security |
| | | Organization | • Expanding Organizational Investment<br>• Enhancing Organizational Analytics Capability<br>• Enhancing Organizational Application Capability |
| | | Technology | • Developing Technologies for Data Aggregation and Processing<br>• Developing Technologies for Data Storage<br>• Developing Technologies for Analytics Platform<br>• Developing Technologies for Application |
| | | Support | • Adjusting The Laws and Regulations<br>• Establishing Government Policies<br>• Strengthening Cooperation within the Ecosystem<br>• Fostering Data Experts |
| A15 | [36] | Factors related to strategy, processes, and leadership | • Strategic Alignment<br>• Top Management Involvement<br>• Compatibility with Organizational Processes<br>• Cultural Change for Data Orientation and Experimentation |
| | | Factors related to human resources | • Overcoming The Talent Gap |
| | | Relevant factors related to implementation management | •Communication<br>• Expectations Management<br>• Structure and Functions for Implementation<br>• Agile Methods |
| | | Factors related to ethics and privacy | • Privacy Management<br>• Information Sharing |
| A16 | [43] | Technology | • Usage Experience with Data from External Sources<br>• Usage Experience with Data from Internal Sources<br>• Experience with Big-Data-Related Technology<br>• Privacy and Security |
| | | Organization | • BDA Skills<br>• Management Support |
| | | Environment | • Market Pressure<br>• Big Data Pressure<br>• Market Performance |

** *Source: SLR and Compilation by Author*

framework to categorize the influencing factors. Reference [43] extended the TOE framework from the perspective of dynamic capability which requires a mixture of technology, management, and skills as a requirement to the successful adoption of Big Data Analytics. Reference [43] focused on the importance of dynamic capabilities and Big Data capabilities to enable the organization to evolve according to the requirements of a changing environment [43].

Table 5. shows the existing literature, source, domain, focus area, source of CSFs, type of study, data collection method, and data analysis method.

### 2) (RQ2): WHAT ARE THE EXISTING CRITICAL SUCCESS FACTORS (CSFs) FOR BIG DATA ANALYTICS?

Big Data implementation is covered by critical factors that organizations should identify to guarantee successful implementation [40]. The increasing investments in Big Data projects request a focused investigation on identifying the CSFs and understanding their effect on the implementation process [10], [45]. Although various studies focused on understanding the CSFs in the adoption of Information Systems (IS), limited academic research has been investigated on identifying the critical factors influencing the Big Data success [9].

This paper presents and identifies the Big Data CSFs for the various categories by reviewing the existing literature from academia which were published in the past ten years (2009-2019). According to this SLR, the number and categories of CSFs for Big Data have been identified by previous studies different from one to another. This paper started with retrieving the previous literature related to this paper. Then an initial number of 191 CSFs were identified from previous studies as shown in Table 6.

After conducting a descriptive qualitative content analysis method for the existing literature and integrated the factors with the same meaning, the 191 CSFs extracted from the literature reduced to 74 CSFs. Useful sources for identifying the Big Data CSFs were found in the existing literature. Based on the descriptive qualitative content analysis of the selected studies, this SLR identified 74 CSFs for Big Data implementation are shown in Table 7.

### 3) (RQ3): WHAT ARE THE CATEGORIES OF BIG DATA ANALYTICS CRITICAL SUCCESS FACTORS (CSFs)?

After conducting a descriptive qualitative content analysis for the existing literature and integrated the factors with the same meaning, the 191 CSFs extracted from the literature reduced to 74 CSFs. The factors could be categorized using both apparent and potential content [47]. The former is the specific, clear, surface contents that are easily categorized whereas the latter refers to the underlying meaning contained in a reference text [48]. Based on the meaning of the most common categories, this study identified and categorized the final list of the CSFs to propose five (5) main categories for Big Data successful as (1) Organization; (2) People; (3) Technology; (4) Data Management; and (5) Governance. Table 7 shows the final list of the Big Data CSFs and their proposed categories based on literature.

These BD CSFs and their categories would be used in future work as a construct to develop a conceptual framework for CSFs of Big Data Analytics.

**TABLE 7. Big data critical success factors and their proposed categories.**

| Proposed Category | Critical Success Factors | Authors |
|---|---|---|
| Organization | 1. Vision, mission, objectives, and values | [3], [8], [14] |
| | 2. Top management support and Direct involvement of the CEO | [3], [8], [15], [16], [22], [31], [36], [40], [41], [43] |
| | 3. Organizational change, change management, and Change Management Process | [8], [9], [22], [35], [40] |
| | 4. Project management | [8], [16], [31], [36] |
| | 5. Resource relate factors(Time and budget) and Organizational investments | [8], [41], [42] |
| | 6. Organizational and Process Performance | [3], [8], [22] |
| | 7. Strategy and strategic alignment | [3], [9], [16], [22], [31], [36] |
| | 8. Communication between IT and business | [9], [22], [36], [42] |
| | 9. Flexibility and agility with Freedom for experimentation | [9], [22] |
| | 10. Organizational structure | [9], [22], [36], [41] |
| | 11. Sponsorship and outsourcing | [9], [23] [15] |
| | 12. Culture (Culture of data-driven) | [22], [27], [36], [44] |
| | 13. Organizational Process | [44] [41] [3] [36] |
| | 14. The maturity and readiness | [3] |
| | 15. Organization Capability | [27], [42] |
| | 16. Environment | [27], [43] |
| | 17. Perceived Benefit | [27] |
| | 18. Industry structure | [35] |
| | 19. Effective value discovery process | [15] |
| People | 1. Team skills | [3], [8], [14], [16], [31] |
| | 2. User participation | [8] [3] |
| | 3. Skill gap analysis | [9], [22], [36] |
| | 4. Development of skills, training, and Capacity building | [9], [22], [44] |
| | 5. Stakeholder coordination | [23] |
| | 6. Talent planning | [15] |
| | 7. Human Capability | [22], [27], [41], [43] |
| | 8. Data-driven Mind set | [9] |
| | 9. Managerial skills | [3], [44] |
| | 10. User Acceptance | [8] |
| | 11. The required skillset and Data Experts | [9], [22], [35], [40], [42], [43] |
| | 12. Multidisciplinary team | [3], [22] |
| Technology | 1. Documentation | [16], [31] |
| | 2. Data sources & data storage | [8], [22] |
| | 3. Technology Infrastructure and applications | [3], [8], [16], [31], [40], [41] |
| | 4. Analytics Capability | [27], [39] |
| | 5. Technology performance | [3], [8], [16], [31] |
| | 6. Analytical and technological tools | [16], [31] |
| | 7. Capabilities of tools | [14] |
| | 8. Developing technologies for data aggregation and processing, data storage, for analytics platform, and application | [42] |

**TABLE 7.** ***(Continued.)*** **Big data critical success factors and their proposed categories.**

| Proposed Category | Critical Success Factors | Authors |
|---|---|---|
| | 9. Data Documentation | [16], [31] |
| | 10. Investment in IT infrastructure | [9], [22] |
| | 11. Integration architecture | [15] |
| | 12. Discovery technology | [22] |
| | 13. Reporting and visualization technology | [22] |
| | 14. Technology Capability | [27] |
| | 15. Adequate Selection of Technology and Algorithms | [9] |
| | 16. Combine Cloud Computing Into The Organization's Big Data Analytics | [44] |
| | 17. Service-Orientation Mind-Set | [15] |
| | 18. Not Blindly Following IT Fashion | [15], [44] |
| | 19. Flexibility And Scalability Of Software Application | [3] |
| | 20. Technology Readiness and Maturity | [3] |
| Data Management | 1. Document collection/access to sources | [16], [22], [31] |
| | 2. Master Data Management | [9] |
| | 3. Data Management | [27] |
| | 4. Data and Data Quality Management | [9], [22] |
| | 5. Unstructured/structured data (variety, access) | [22] |
| | 6. Representativeness of data | [22] |
| | 7. Integration Solutions | [9] |
| | 8. Data | [3] |
| | 9. Opening data between inter- and intra-organizations | [42] |
| | 10. Privacy management | [36] |
| | 11. Data Sharing and Data Integration | [22], [36] |
| Governance | 1. Data Privacy and Security | [9], [22], [40], [42], [45] |
| | 2. Data protection and privacy by design | [9], [22] |
| | 3. Data Availability and accessibility | [35], [40], [41] |
| | 4. Data governance | [44] |
| | 5. Laws and Regulations | [42] |
| | 6. Government Policies | [42] |
| | 7. Policies and Procedures | [35], [41] |
| | 8. Data and information Quality | [8], [16], [27], [31], [40]–[42] |
| | 9. Systems Quality | [8], [27] |
| | 10. Data Integrity | [16], [22], [31] |
| | 11. Data ownership | [22] |
| | 12. data standardization | [27] |

** *Source: SLR and Compilation by Author*

## IV. TOWARDS A CLASSIFICATION FRAMEWORK FOR CSFs OF BIG DATA ANALYTICS

To identify the categories of Big Data CSFs, individual CSFs were classified according to the categories in the existing literature and their meaning. For example, IT Infrastructure CSF was classified as a part of the Technology category and team skills were categorized as part of the People category [17]. Based on conducting the descriptive qualitative content analysis approach of the existing literature, five main categories were proposed and identified, namely (1) Organization; (2) Technology; (3) People; (4) Data Management; and (5) Governance as shown in Table 7. Figure 4 illustrates the classification for Big Data Analytics CSFs toward developing a classification framework for the CSFs of Big Data Analytics.

A Top-Down technique [3], [10] could be applied to define the main five (5) categories for the final list of the Big Data CSFs. The following section describes the five main categories and classification for Big Data CSFs that were identified in this SLR.

### A. ORGANIZATION

The category of ''Organization'' has the most CSFs in literature [3]. The category considers the mission, vision, and strategy of the organizations. For successful implementation for the Big Data projects, the organization should be aware of what they have, what they aim to achieve, and how to align the BD with the organization's objectives [3], [10]. The organization's structure affects the maturity and the successful implementation of BD within the organization including other related factors as close collaboration between Business and IT as well as the structure of the organization, and the characteristics of the organization [9].

### B. PEOPLE

The category of ''People'' covers the human side of the Information System. This category is critical to determine the maturity of Big Data projects toward a successful implementation. According to [32], the availability of qualified people on the Big Data projects has positive impacts on the successful implementation. This category is referred to the human capabilities, analytical skills, and team skills critical to success in Big Data projects [3], [49].

### C. TECHNOLOGY

Technology covers the tools that are provided to meet the Big Data needs and requirements. The category of ''Technology'' is related to data collection, storage, processing, analysis, and applications [42]. This category concentrates on the technological tools data visualization tools such as system performance, real-time availability, data quality, and integration with the existing analytical tools. At the same time, the applications and system infrastructure refer to factors including security, scalability, and system flexibility. Data sources refer to the CSFs relevant to data and influence the trustworthiness of data, like metadata management, data integrity, and data quality [3].

### D. DATA MANAGEMENT

Data Management looks at how the organization manage their Big Data issues and analyses its data. The category of ''Data Management'' is concerned with the administrative process that includes capturing, processing, validating, storing, and protecting the required data to guarantee the secure accessibility, reliability, and timeliness of the data [50]. Data management is a widespread term that covers a broad domain of data applications. This category may refer to the basic

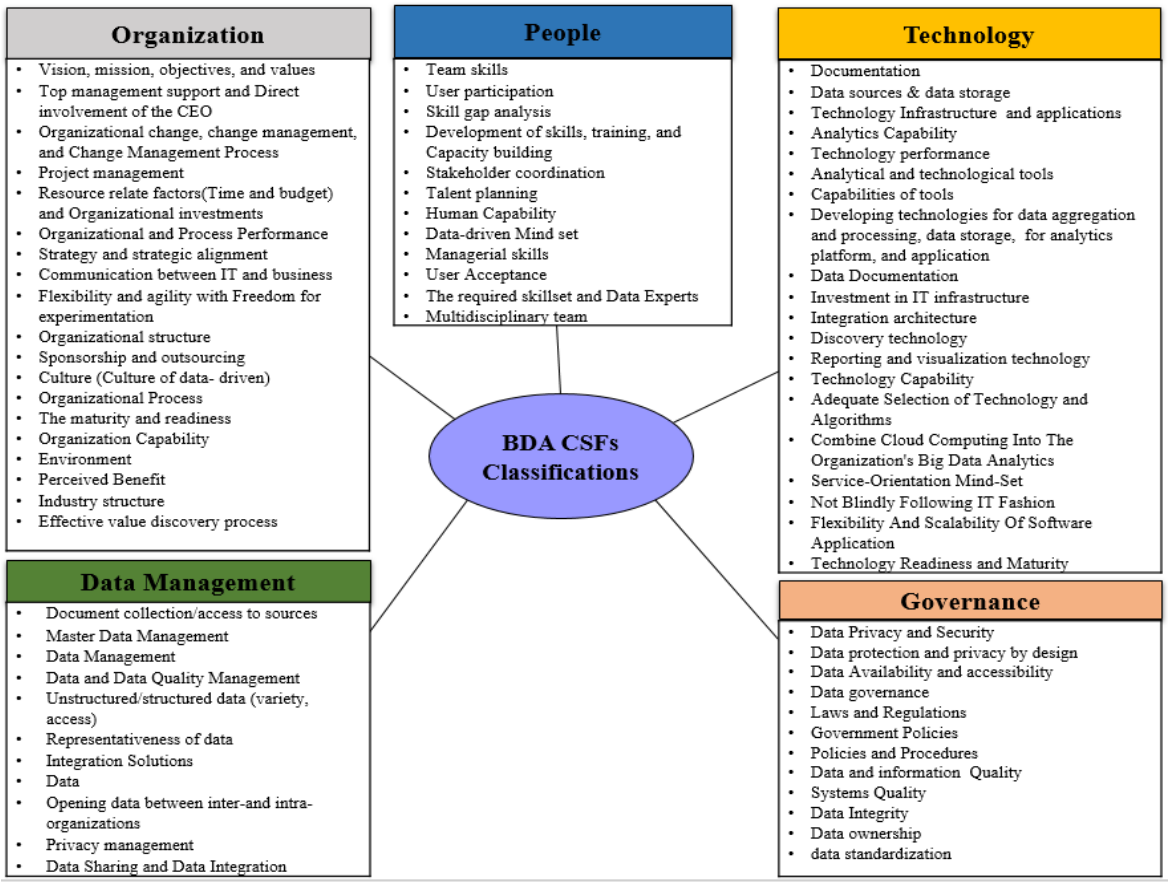



**FIGURE 4.** Classification Framework for CSFs of Big Data Analytics.

concepts of data management or specific technology as Big Data. Also, it refers to some applications of Big Data management as data storage, data security, and data design. Some applications of Big Data management include data design, data storage, and data security [51]. Without Big Data management, it is impossible to continuously manage the other important aspects of the Big Data framework.

### E. GOVERNANCE

Governance and policy address whether the BD were analyzed appropriately and whether it applies legal framework and security for the data. The category of ''Governance'' covers the social activities, processes, practices, and policies undertaken by either organizations or other stakeholders [52]. This category is usually related to the act of ''governing'' where various actors are involved in the Big Data projects [10], [30], [53]. The required Data often has to cross various organizational, departments, and even public-private boundaries. Governance is a critical category as it contributes to the effectiveness and efficiency of the processes across those different borders [10], [52].

## V. CONCLUSION AND FUTURE WORK

Organizations still struggle to succeed with Big Data projects. Identifying the Critical Success Factors (CSFs) for Big Data could help the organizations to implement Big Data projects successfully with clear and thoughtful steps. There is limited research conducted on the CSFs of Big Data domain, where most of the existing literature adopted the CSFs from other related technologies like Business Intelligent (BI) and data mining. Based on the SLR guidelines in Software Engineering by [38], [39], this paper attempts to provide more understanding about the existing CSFs for Big Data Analytics and implementation by answering three research questions RQ1: ''How many studies have investigated on Big Data CSFs for analytics and implementation?'', RQ2: ''What are the existing Critical Success Factors for Big Data Analytics'', and RQ3: ''What are the categories of Big Data Analytics CSFs?''. This paper contributes to the body of knowledge by identifying the final set of 16 papers published in ''English'' during the past twelve years (2007-2019) in the context of Big Data and Big Data Analytics. Also, this paper identified a list of 74 CSFs in the focus domain of Big Data. This list will be important as a referenced list of all existing Big Data CSFs and categories which are required by decision-makers to be successful in Big Data implementation. Based on the descriptive qualitative content analysis method of the existing Big Data CSFs and their categories and after grouping the extracted CSFs with the same meaning, this paper identified and proposed a classification schema and framework in terms

of 5 main categories of Big Data CSFs, namely Organization, Technology, People, Data Management, and Governance.

One of the limitations of this research is the limited number of literature returned from the search that could be used to answer the SLR research questions. In addition to limitations, related to some inaccessible database /paper. This SLR paper could be used as a referential framework for the successful strategy and implementation of Big Data by formulating more effective data-driven decisions. Finally, future work will investigate the priority of the CSFs of Big Data and their categories toward developing a conceptual framework for CSFs of Big Data Analytics and measuring how these factors influence Big Data Analytics' success. Moreover, future work will evaluate the extracted CSFs for Big Data by interviews or quantitatively by questionnaire toward developing a conceptual framework for assessing the success of Big Data projects.

## REFERENCES

[1] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Res.*, vol. 2, no. 2, pp. 59–64, Jun. 2015.

[2] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: A survey," *J. Big Data*, vol. 2, no. 1, p. 21, 2015.

[3] S. Eybers and M. J. Hattingh, "Critical success factor categories for big data: A preliminary analysis of the current academic landscape," in *Proc. IST–Africa Week Conf. (IST-Africa)*, May 2017, pp. 1–11, doi: 10.23919/ISTAFRICA.2017.8102327.

[4] A. Labrinidis and H. V. Jagadish, "Challenges and opportunities with big data," *Proc. VLDB Endowment*, vol. 52, no. 12, pp. 2032–2033, 2012.

[5] Z. A. Al-Sai, R. Abdullah, and M. H. Husin, "Big data impacts and challenges: A review," in *Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT)*, Apr. 2019, pp. 150–155.

[6] Z. A. Al-Sai and L. M. Abualigah, "Big data and E-government: A review," in *Proc. 8th Int. Conf. Inf. Technol.*, May 2017, pp. 580–587.

[7] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data?: Survey, technologies, opportunities, and challenges," *Sci. World J.*, vol. 2014, Jul. 2014, Art. no. 712826.

[8] J. M. Evers, "Critical success factors of business intelligence and big data analysis," M.S. thesis, Dept. Inf. Manage., Tilburg Univ., Tilburg, The Netherlands, 2014.

[9] P. Cato, P. Golzer, and W. Demmelhuber, "An investigation into the implementation factors affecting the success of big data systems," in *Proc. 11th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2015, pp. 134–139, doi: 10.1109/INNOVATIONS.2015.7381528.

[10] Z. A. Al-Sai, R. Abdullah, M. H. Husin, and S. M. Syed-Mohamad, "A preliminary systematic literature review on critical success factors categories for big data," in *Proc. AiIc*, 2019, pp. 2320–2324.

[11] J. Polonetsky, O. Tene, and J. Jerome, "Benefit-risk analysis for big data projects," in *Future Of Privacy Forum*. 2014. [Online]. Available: https://fpf.org/wp-content/uploads/2014/09/FPF_DataBenefitAnalysis_FINAL1.pdf

[12] B. P. Russom, "Big data analytics," *TDWI Best Practices Rep.*, vol. 19, no. 4, pp. 1–34, 2011.

[13] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

[14] N. Nieder, "Effective big data management: A development of critical success factors and an analysis of firms' capabilities in the automotive industry," Ph.D. dissertation, 2016.

[15] H.-M. Chen, R. Schutz, R. Kazman, and F. Matthes, "Amazon in the air: Innovating with big data at lufthansa," in *Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2016, pp. 5096–5105, doi: 10.1109/HICSS.2016.631.

[16] A. Koronios, J. Gao, and S. Selle, "Big data project success—A meta analysis," in *Proc. Pacis*, 2014, p. 376.

[17] A. Caralli, F. Stevens, J. Willke, and R. Wilson, "The critical success factor method?: Establishing a foundation for enterprise security management," Carnegie-Mellon Univ. Pittsburgh Pa Softw. Eng. Inst., Pittsburgh, PA, USA, Tech. Rep. CMU/SEI-2004-TR-010, 2004.

[18] M. Amberg, "Background of critical success factor research," *Evolution (N. Y)*, vol. 17, no. 2, p. 12, 2005.

[19] F. Aa, M. M, H. H, and R. Almamlook, "Overview success criteria and critical success factors in project management," *Ind. Eng. Manage.*, vol. 7, no. 1, pp. 1–6, 2018.

[20] J. F. Rockart, "A new approach to defining the chief executive's information needs," School Manage., Center Inf. Syst. Res., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., 1978, vol. 02139.

[21] N. Kaur and G. Singh, "Critical success factors in agile software development projects?: A review," *Int. J. Emerg. Technol.*, vol. 7, no. 1, pp. 1–4, 2016.

[22] J. S. Saltz and I. Shamshurin, "Big data team process methodologies: A literature review and the identification of key factors for a project's success," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2872–2879.

[23] A. Paidi and D. Iliopoulou, "Critical success factors in the implementation of integrated healthcare information systems," in *Concepts and Trends in Healthcare Information Systems* (Annals of Information Systems), vol. 16, 2014, pp. 103–119.

[24] T. Al-Rousan and B. Al-Shargabi, "A new maturity model for the implementation of software process improvement in web-based Projects," *J. Digit. Inf. Manag.*, vol. 15, no. 2, pp. 66–75, 2017.

[25] J. M. Esteves, "Definition and analysis of critical success factors for ERP implementation projects," Dr. thesis, Univ. Politècnica Catalunya, Barcelona, Spain, 2004.

[26] S. Jalali and C. Wohlin, "Systematic literature studies: Database searches vs. Backward snowballing," in *Proc. ACM-IEEE Int. Symp. Empirical Softw. Eng. Meas. ESEM*, 2012, pp. 29–38, doi: 10.1145/2372251.2372257.

[27] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Factors influencing to the implementation success of big data analytics: A systematic literature review," in *Proc. Int. Conf. Res. Innov. Inf. Syst. (ICRIIS)*, Jul. 2017, pp. 1–6, doi: 10.1109/ICRIIS.2017.8002536.

[28] W. H. DeLone and E. R. McLean, "Information systems success measurement," *Found. Trends Inf. Syst.*, vol. 2, no. 1, pp. 1–116, 2016.

[29] H. Macke. (2014). *Measuring Your Big Data Maturity*. Kenny & Company. [Online]. Available: https://michaelskenny.com/points-of-view/measuring-your-big-data-maturity/

[30] X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang, "A map of threats to validity of systematic literature reviews in software engineering," in *Proc. 23rd Asia–Pacific Softw. Eng. Conf. (APSEC)*, 2016, pp. 153–160, doi: 10.1109/APSEC.2016.031.

[31] J. Gao, A. Koronios, and S. Selle, "Towards a process view on critical success factors in big data analytics projects," in *Proc. 21st Amer. Conf. Inf. Syst.*, 2015, pp. 1–14.

[32] R. L. Hayen, C. D. Rutashobya, and D. E. Vetter, "An investigation of the factors affecting data warehousing success," *Int. Assoc. Comput. Inf. Ation Syst.*, vol. 8, no. 2, pp. 547–553, 2007.

[33] K. Blasiak, "Big Data; A management revolution," M.S. thesis, Dept. Appl. Sci., Bachelor Bus. Admin., Helsinki Metropolia Univ., Vantaa, Finland, 2014.

[34] B. M. Félix, E. Tavares, and N. W. F. Cavalcante, "Fatores críticos de sucesso para adoção de big data no varejo virtual: Estudo de caso do magazine Luiza," *Rev. Bras. Gest. Negocios*, vol. 20, no. 1, pp. 112–126, 2018.

[35] S. Akter, S. F. Wamba, A. Gunasekaran, R. Dubey, and S. J. Childe, "How to improve firm performance using big data analytics capability and business strategy alignment?" *Int. J. Prod. Econ.*, vol. 182. pp. 113–131, 2016.

[36] B. Muniz Félix, E. Tavares, and N. W. Freitas Cavalcante, "Critical success factors for big data adoption in the virtual retail: Magazine luiza case study," *Rev. Bus. Manage.*, vol. 20, no. 1, pp. 112–126, Jan. 2018.

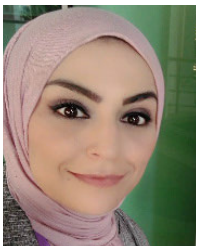
[37] S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *Int. J. Prod. Econ.*, vol. 165, pp. 234–246, Jul. 2015.

[38] B. Kitchenham, *Procedures for Performing Systematic Reviews*, no. 33. Keele, U.K.: Keele Univ., 2004, pp. 1–26.

[39] B. Kitchenham, "Guidelines for performing systematic literature reviews in software engineering," EBSE, Goyang-si, South Korea, Tech. Rep. Ver. 2.3, 2007, vol. 5.

[40] M. Halaweh and A. El Massry, "Conceptual model for successful implementation of big data in organizations," *J. Int. Technol. Inf. Manag.*, vol. 24, no. 2, pp. 21–29, 2015.

[41] L. F. Gómez and R. Heeks, "Measuring the barriers to big data for development: Design-reality gap analysis in Colombia's public sector," Develop. Inform. Working Paper 62, 2016.
[42] M.-K. Kim and J.-H. Park, "Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare," *Inf. Develop.*, vol. 33, no. 3, pp. 257–269, Jun. 2017.
[43] A. Schüll and N. Maslan, "On the adoption of big data analytics: Interdependencies of contextual factors," in *Proc. 20th Int. Conf. Enterprise Inf. Syst.*, 2018, pp. 425–431.
[44] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technol. Forecasting Social Change*, vol. 126, pp. 3–13, Jan. 2018.
[45] P. Mikalef, I. Pappas, M. Giannakos, and J. Krogstie, "Big data and strategy?: A research framework," in *Proc. MCIS*, Sep. 2016, p. 50.
[46] M. Petäjäniemi, "Big data in finish companies what has been done and the results achieved," M.S. thesis, Dept. Inf. Service Economy, Aalto Univ., Espoo, Finland, 2015.
[47] M. Kamaruddin, R. Razali, and A. Deraman, "Critical success factors of executive information systems development for education management—A preliminary investigation," in *Proc. Int. Conf. Electr. Eng. Informat.*, Jul. 2011, pp. 1–6.
[48] K. Mahamsiatus and R. Rozilawati, "A framework of successful executive information system development for education domain mahamsiatus kamaruddin and rozilawati razali centre of software technology and management, faculty of information science and technology," *Amer. J. Appl. Sci.*, vol. 8, no. 10, pp. 997–1003, 2011.
[49] Z. You and C. Wu, "A framework for data-driven informatization of the construction company," *Adv. Eng. Informat.*, vol. 39, pp. 269–277, Jan. 2019.
[50] M. Galetto, "What is data management?" NGDATA, Tech. Article, Mar. 2016. [Online]. Available: https://www.ngdata.com/what-is-data-management
[51] TechTerms. (2012). *Data Management Definition*. [Online]. Available: http://www.techterms.com/definition/data_management
[52] W. Veeneman, H. van der Voort, F. Hirschhorn, B. Steenhuisen, and B. Klievink, "PETRA: Governance as a key success factor for big data solutions in mobility," *Res. Transp. Econ.*, vol. 69, pp. 420–429, Sep. 2018.
[53] Z. A. Al-Sai, R. Abdullah, M. H. Heikal, and S. M. Syed-Mohamad, "A preliminary systematic literature review on critical success factors categories for big data analytics," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2320–2324, 2019.
[54] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Conceptual model development of big data analytics implementation assessment effect on decision-making," *IJIMAI*, vol. 5, no. 1, pp. 101–106, 2018.
[55] B. M. Kalema and M. Mokgadi, "Developing countries organizations' readiness for big data analytics," *Problems Perspect. Manage.*, vol. 15, no. 1, pp. 260–270, May 2017.
[56] M. Gupta and J. F. George, "Toward the development of a big data analytics capability," *Inf. Manage.*, vol. 53, no. 8, pp. 1049–1064, Dec. 2016.



**ZAHER ALI AL-SAI** received the B.S. degree in computer science from the Zaytoonah University of Jordan and the M.Sc. degree (Hons.) in e-government project management and the M.B.A. degree from BAU Jordan. She is currently pursuing the Ph.D. degree in software engineering and big data with the School of Computer Science, Universiti Sains Malaysia (USM). She is also an Active Researcher in empirical software engineering and big data. She has an interest in big data, big data analytics, software process improvement, big data engineering, big data development, big data readiness and maturity assessment, and evidence-based software engineering. She is an Active Ambassador and a Facilitator with the Library of USM. She is also a Presenter of many workshops on performing the Systematic Literature Review and research support topics. She is also the Founder and the CEO of the non-profit organization Big Data Minds.

**ROSNI ABDULLAH** received the bachelor's and master's degree in computer science from Western Michigan University, Kalamazoo, MI, USA, and the Ph.D. degree in parallel algorithms from Loughborough University, U.K., in April 1997. She was the first lady Dean of the School of Computer Sciences, Universiti Sains Malaysia (USM), from 2004 to 2012, where she is currently a Professor in parallel computing. She is one of the national pioneers in this field. She has been the Head of the Parallel and Distributed Processing Research Group, School of Computer Sciences, since its inception in 1994. Her research areas include parallel and distributed computing, parallel numerical algorithms, and computational biology. She has graduated 20 Ph.D. students and led 13 research grants, including two European Union grants and three Intel grants. She also drives the Makers movement at USM. She is also back as the Dean of the School of Computer Sciences, USM, as well as the Director of the National Advanced IPv6 Center (Nav6), a Center of Research Excellence, USM, which focus on Cybersecurity and Internet of Things (IoT).

**MOHD HEIKAL HUSIN** (Member, IEEE) received the bachelor's degree in multimedia computing from INTI International University (formerly known as the INTI College Malaysia) and the master's degree in e-commerce and the Ph.D. degree in IT from the School of Computer and Information Science, University of South Australia. He is currently an Information System Lecturer with the School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia. He has published in local conferences as well as international conferences on the topic of social media usage and effective policy development within organizations as well as effective technology adoption approaches. His current research interests include government 2.0, enterprise 2.0, social networking, social network analysis, knowledge management, and more recently ICT education and data mining.