electronics

MDPI

Article

Framework for Structuring Big Data Projects

Gustavo Grander [1,*], Luciano Ferreira Da Silva [1], Ernesto Del Rosário Santibañez Gonzalez [2] and Renato Penha [1]

[1] Programa de Pós-Graduação em Gestão de Projetos (PPGP), Universidade Nove de Julho, São Paulo 01156-000, Brazil
[2] Department of Industrial Engineering, Faculty of Engineering, University of Talca, Curico 3340000, Chile
* Correspondence: grandergustavo@gmail.com

Abstract: This article aims to present a framework for structuring Big Data projects. The methodological procedures were divided into two phases: in-depth interview and focus group. The first phase embraces 12 in-depth individual interviews. In the second phase, three sessions of interviews with focus groups were applied. Both phases had as research subjects professionals with experience in Big Data projects. The analysis process was based on categorization through theory-driven and data-driven codes. Based on our analysis, it was possible to present a definition of a Big Data project and explore the beginning and its phases. We also identified 17 different critical factors in Big Data projects and proposed a discussion on technical and behavioral skills and decision-making issues in Big Data projects. We also developed and validated a framework to help structure a Big Data project. As the main theoretical contribution, our study aligns with a growing body of researchers who are proposing to debate Big Data projects, and with that increasing maturity on the topic. As a practical contribution, we present a framework that we hope will contribute to professionals working in the area, helping to conduct a Big Data project through a systemic view.

Keywords: project; big data; framework; project management; big data projects

check for updates

Citation: Grander, G.; Da Silva, L.F.; Gonzalez, E.D.R.S.; Penha, R. Framework for Structuring Big Data Projects. *Electronics* **2022**, *11*, 3540. https://doi.org/10.3390/electronics11213540

Academic Editors: Ting Wang, Lu Wang and Subrota Kumar Mondal

Received: 2 September 2022
Accepted: 26 October 2022
Published: 30 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Big Data has been widely used in optimizing logistical routes, improving traffic, predicting diseases in healthcare, optimizing business, and targeting marketing campaigns [1,2]. Big Data affects many domains of daily life and has had a profound impact on the whole of society, which leads to the reflection that the practice of Big Data becomes increasingly relevant in various activities and sectors [3]. Although the importance of Big Data is recognized, there is a paradox around its concept, as at the same time as it can be used as a powerful tool to deal with various societal ills, offering the potential for new insights in areas such as research medical care, combating terrorism, and climate change, it can also allow for the invasion of the privacy of people and institutions, diminishing civil liberties and increasing state and corporate control [4].

In the context of Big Data, there is a great challenge in managing projects that include Big Data due to the increasing and cheap volume of data collection and storage [5,6]. Tang et al. [2] highlight that there are also different ways in which this data is being generated, whether it is user-generated data (in social media, texts, photos, etc.), data monitored by devices (such as weather monitors, meters, GPS, etc.), and data logging activities (web search/visit, online/offline marketing, clinical treatments, laboratory experiments, etc.). A relevant aspect in the use of Big Data is the culture of data-driven decision-making, which is essential for the success of a Big Data project [7]. However, many Big Data projects do not generate the expected result or are not even completed, generating high costs and losses for companies [8]. Therefore, monitoring and managing high-load data projects requires new approaches to infrastructure, risk management, and data-driven decision support [9].

Wu, Zhu, Wu, and Ding [10] indicate that to explore Big Data, there are challenges at the data, models, and systems levels. According to the same authors, at the data

level, autonomous information sources and the variety of data collection environments often result in data with complicated conditions, such as missing or uncertain values. At the model level, the main challenge is to generate global models by combining patterns discovered locally to form a unifying vision. At the systems level, the essential challenge is that a Big Data mining framework needs to consider complex relationships between samples, models, and data sources, along with their evolving changes over time and other possible factors.

Based on the aforementioned information, we can say that despite the promising opportunity that the moment offers us to improve Big Data projects, there is a disconnect between the potential that digital initiatives can add value and the successful implementations of these projects. Even though some studies report alarming failure rates in Big Data projects, it is very difficult to make such a measurement as there are no records of Big Data projects that can be followed to determine which projects fail over time [11]. Big Data projects are characterized by their large size, high complexity, and innovative technology, and these characteristics aggravate any deficiencies in management practices, skills of the people involved, or even in the processes [12].

The differences between Big Data projects and, for example, software development projects are mainly related to desired organizational processes [13]. Big data projects often have highly uncertain inputs (e.g., what data might be relevant), as well as highly uncertain outcomes, i.e., will there be insight derived from analyzing the data [14]. There is still the challenge of determining whether the data to be analyzed does not have time mismatch issues or what is the acceptable level of data quality [15]. These characteristics are unparalleled in a software development project [14].

Chiheb, Boumahdi, and Bouarfa [16] highlight that one of the reasons that Big Data projects fail to produce the desired results is in the way these projects were developed, since organizations usually focus on their data and Data Analytics, without establishing a focus on decision making that should be the real objective of the Big Data project. Given the complexity, the phenomenon of the failure tendency of Big Data projects refers to a few factors, including human factors related to skills and perception, technical factors related to the technologies used to implement Big Data, the nature of the data itself, legal factors related to who owns the data and how to use it, organizational factors related to data strategies, alignment with organizational objectives, and availability of funds [17]. As a result, there are serious consequences for organizations due to the high cost of Big Data projects, the loss of time of people assigned to work on these projects, and the competitive disadvantage of not using Big Data analytics while competitors develop such projects [18].

To obtain positive results in Big Data projects, professionals and researchers should not only focus on the data and data analysis aspect when studying and developing a Big Data project but also go beyond that and study the decision aspect of this project [16]. There are also results indicating the importance of choosing the approach in the management of Big Data projects, which is conditioned not only by the size and criticality of the project but also by the dynamics of the environment [19]. For Jin, Wah, Cheng, and Wang [20], the necessary conditions for a Big Data project to be successful are related to having very clear technical, social, and economic requirements, having data structures small enough to characterize the behavior to be analyzed, and a top-down management model should be sought to promote comprehensive resolutions through an integrated approach.

In this sense, it has become obvious that the complexity of the data, tools, techniques, and resources involved require an adequate level of project management for organizations to fully take advantage of Big Data [21]. Thus, organizations increasingly need to continually realign their work practices, organizational models, and stakeholder interests to reap the benefits of Big Data projects [22]. However, despite the significant progress made in the area of Big Data in recent years, Lara, De Sojo, Aljawarneh, Schumaker, and Al-Shargabi [23] point out that practitioners highlight important research opportunities mainly related to guides, models, and methodologies. In addition, Barham and Daim [24]

point out that effective models are also opportune to take advantage of lessons learned from previous projects.

Given the importance of support structures for decision making in Big Data projects, some studies propose models to fill this gap. Mousanif, Sabah, Douiji, and Sayad [25] and Mousanif, Sabah, Douiji, and Sayad [26] present a holistic approach to encompass the planning, implementation and post-implementation phases of a Big Data project. Dutta and Bose [7] propose a framework for implementing Big Data projects based on the phases of the strategic base, data analysis, and implementation. The framework proposed by the authors was applied in the manufacturing industry in India, highlighting the relevance of a clear understanding of the business problem, as well as very specific steps for the execution of the project. Muntean and Militaru [27] proposed a methodological framework based on design science research for the design and development of data and information artifacts in data analysis projects, particularly managerial performance analysis.

This need for a global model or roadmap to assist Information Technology (IT) departments should not only serve the implementation of a Big Data project but also the best possible use to meet business objectives [26]. While advances have been made in the algorithms and technologies used to perform the analyses, little has been done to determine how the team should work together to carry out a Big Data project [14].

Based on the presented context, this study aims to present a framework for structuring Big Data projects. To this end, we conducted qualitative research composed of two phases. The first included the application of 12 individual in-depth interviews totaling 9.52 h of interviews. The second phase included three sessions of interviews with focus groups using nine professionals, totaling 3.38 h of discussion. As a result, we present a definition of a Big Data project, we investigate how a Big Data project is born, what are its phases, critical factors, and problems related to decision-making. In addition, we present a management framework for structuring Big Data projects.

This study is justified as it aims to contribute to the discussion about Big Data projects. In terms of the models, we can say that they are influenced by critical factors for good project delivery. In this study, we tried to map the antecedents and consequences of Big Data projects based on the literature and reality of the interviewees. Corroborating the expectations of this research, Saltz and Shamshurin [28] identified 33 critical factors for the execution of Big Data efforts, which are grouped by six identified characteristics of a mature Big Data organization which are: Data, Governance, Processes, Objectives, Team, and Tools. Additionally, in an investigation of critical factors in Big Data projects, Gao, Koronios, and Selle [29] presented 27 factors divided into six phases, namely: Business, Data, Analysis, Implementation, Measurement, and General. Thus, based on the process of understanding the reality studied, the researchers aimed to contribute to the body of knowledge on Big Data projects.

After this introduction section, we structure the article as follows: a materials and methods section, a results section, a framework proposition and validation section, a discussion section, and finally a final remarks section.

2. Materials and Methods

We present in this section the methodological procedures carried out in this study through the description of the sample, data collection procedures, data analysis procedures, and analysis categories, providing transparency and the possibility of replicating the processes used that made it possible to legitimize the findings of this research [30]. In our research, we adopted the qualitative method to understand and explore aspects related to the phenomenon of Big Data projects. This research followed in two phases, the first being based on in-depth individual interviews that allowed access to the reality of the researched social actor. After the application of the interviews and data analysis for the construction of the proposed framework, the second phase included three focus group sessions in order to seek the validation of the proposal by the practitioners.

The questionnaire prepared before the interviews was based on the literature on the subject in order to seek an understanding of the variables related to the focus of this study. The first question, which was not linked to a category of analysis, sought to explore the interviewee's professional experience in the area of Big Data, therefore, the interviewee was initially asked to comment on their experience in the area. The interview process followed the prescriptions of an in-depth interview [31]; this type of interview is based on a guided conversation guided by a set of items. Table 1 presents the initial categories raised in the literature review and the respective questions that guided the interviews.

Table 1. Initial categories of analysis.

Categories	Description	Questions	Sources
Definition of a Big Data project	The definition of a Big Data project was sought according to the interviewees.	How do you define a Big Data project?	[12–14]
Phases of a Big Data Project	It was proposed to identify the phases that contemplate Big Data projects according to the perception of the interviewees.	According to your experience, which phases (or stages) are included in a Big Data project?	[10,25,26]
Critical factors in a Big Data project	It was proposed to identify the critical factors in a Big Data project.	In your experience, what would be the critical factors in a Big Data project?	[17,28,29]
Decision-making issues in a Big Data project	It was proposed to identify problems related to decision making in a Big Data project.	During the life cycle of a Big Data project, what issues related to decision making would you highlight?	[7,16]
Methodology/Guide/Roadmap of a Big Data project	It was proposed to verify if there are any execution standards or good practices used for the execution of a Big Data project.	Is there any methodology/guide/script that you use in the Big Data projects you participate in? If not, are you aware of any?	[19,23]

To validate and calibrate the data collection instrument, for the first three interviews we sought to identify the adherence of the answers with our research intention, as well as observing opportunities for new questions to be inserted in the research protocol. With this, guiding questions were constructed from the constituted items (Table 1), and new questions emerged during the conversation so that the researched phenomenon could be better understood from the perspective of the interviewees.

The questions included were: (i) "How is a Big Data project born?"; (ii) "About skills known as Hard skills (technical) and Soft skills (behavioral), are there any that stand out in Big Data projects?". It is noted that in-depth interviews are understood as a guided conversation in which the interviewer assumes the role of facilitator of explanations about what one wants to understand [32]. The interview focused on making the interviewee speak as much as possible about the research topics, so that in-depth questions were used: "how...", "why?...", "what is your opinion?...", "Tell me more about . . . ".

Based on this information about the interview process, the choice of qualitative research is justified, as it made it possible to identify more in-depth issues in the interviewees. How the interviews were conducted required the interviewer to be more prepared, that is, provided with information to conduct the interview openly, creating a promising environment and thus obtaining the greatest amount of information possible [33]. Dilley [34] describes that there must be some key elements for the in-depth interview to be robust, which are: information background, interview analysis, and creation of a protocol. Thus, the researcher must provide information about the cultural context of the interviewee and other information that is relevant to the beginning of the interview.

The initial sample consisted of 12 professionals with an average of 7.3 years of experience with Big Data projects, who were invited to participate in the interview and share knowledge in the area acquired throughout their careers. These professionals were mapped according to the networking of the authors of the article and who, after profile analysis, were invited to participate in the study. All invited professionals agreed to participate of their own free will. For the analysis of the duration of the interviews, we considered the counting of time from the beginning of the first question and ending at the end of the last answer, thus eliminating the initial presentation time and thanks at the end of the interviews.

The 12 interviews totaled 9.52 h (average 47.5 min per interview) that were individually transcribed without the aid of transcription software, and totaling 58 h of transcription in a text document (on average of 290 min to transcribe each interview) (Table 2).

Table 2. Description of respondents.

ID	Background	Highest Degree	Interview Duration (min)	Profile
E1	Tecnologia da Informação	Ph.D.	60	Technician
E2	Business Administration	Ph.D.	85	Technician
E3	Information Technology	Specialist	35	Technician
E4	Computer Science	Specialist	47	Management
E5	Engineering/Business Administration	Ph.D.	25	Management
E6	Mathematic	Ph.D.	36	Technician
E7	Information Technology	Specialist	47	Technician
E8	Engineering	Master's degree	40	Management
E9	Economy	Master's degree	35	Management
E10	Computer Science	Specialist	41	Management
E11	Business Administration	Master's degree	68	Management
E12	Business Administration	Master's degree	52	Technician

The sample included professionals with training in the areas of Business Administration, Computer Science, Economics, Engineering, Mathematics, and Information Technology, being 4 Ph.Ds., 4 masters, and 4 specialists. The interviews were conducted by guided conversations, to exhaust the discussions regarding the items raised here.

For data collection, the interviews were scheduled according to the availability of the interviewees, and this took place between the period from 16 July 2020 to 15 September 2020, the dates of the first and last interviews respectively. The interviews took place with the help of digital communication tools, as they took place during the COVID-19 pandemic period and therefore were not conducted face-to-face. At the beginning of each interview, permission was requested for the interview to be recorded and, with the authorization, the protocol began with the presentation of the interviewee, the objective of the research, and informing that the interviewee could ask for the interview to be interrupted at any time, as well as how, after the interview, the interviewee could have access to the recording for review.

Regarding the analysis process, the search for patterns of behavior that allowed analytical generalizations, or even the construction of theories, was carried out through an inductive process based on the development of categories or themes presented by the interviewees [30]. The process of categorizing the interviewees' speeches also followed a deductive process. Thus, following the recommendation of Charmaz [35], the categories were constituted from the interviews carried out, the perspective of data (data-driven), and constructed from the theory (theory-driven).

We emphasize that the data analysis and discussion phase mainly followed a data-driven perspective, and this type of perspective allows starting from the interview data to build analysis categories—codes [35,36]. Despite this, it is worth noting that the construction of the protocol was guided by a previous survey of the literature, among other research. Thus, the analysis process was carried out based on the coding technique, which Creswell [30] describes as a process of organizing the material and assigning labels to the highlighted categories with a term (code), based on the participant's actual language. For data analysis, we used ATLAS.ti software (version 7.5.4) and followed the analysis process

according to Charmaz [35] guidelines, a process known as an analysis technique based on the Grounded Theory methodology, following coding cycles whose objective is to obtain meaning from a corpus of analysis.

According to Charmaz [35] and Saldaña [36], the coding process follows three phases. First, open coding was performed, which aims to recognize incidents, which are represented by passages, words, and images, among others, and which help in responding to the research problem. These first-order codes were grouped to establish relationships through axial coding. Based on the general structure built from the incidents and relating them hierarchically with other second-order codes, more abstract categories were obtained following the third cycle of analysis called selective.

The analysis was also performed using the constant comparison technique, in which the coding and analysis activities are carried out concomitantly in a circular mode, which allows back and forth to the research corpus for validation of the analysis categories [35,36]. This process follows a flow towards the achievement of the research objectives that are not linearly linked to the process. Thus, as the theoretical sampling is increased (number of incidents found), the main elements are constructed to support more abstract categories and their respective properties [37]. Qualitative studies with a high degree of excellence go beyond description and identification and reach complex thematic connections [30]. It is worth mentioning that the initial categories of analysis start from a theoretical perspective. These categories represented the management styles and were built from the search and analysis of the theoretical framework.

The stage of processing the collected data allowed us to gain confidence in the analysis and expand the perception regarding what was said and not said in the interviews, while the protocol traces the path that was taken to inquire the interviewee. A data perspective oriented towards understanding new elements of the researched phenomenon was also contemplated [35]. Thus, at the end of the individual interviews, we asked the interviewees to suggest any research opportunity, or any that could be developed to assist the management of a Big Data project, and based on these responses, we proposed a future research agenda with opportunities for complementary studies to this one. Therefore, after conducting individual interviews, the data were analyzed, and we highlighted the need to develop a framework for structuring Big Data projects for practitioners. After developing the framework, we understood that it would be essential to validate the model with professionals working in the area of Big Data projects in the second phase of the research and, consequently, a new round of interviews was planned and conducted.

We started a new cycle of interviews now with three focus groups to validate the results obtained from the first part of the study. The focus group is a type of in-depth interview carried out in a group whose meetings have characteristics in terms of proposal, size, composition, and conduction procedures [38]. The use of this procedure was due to the need to seek discussions between practitioners to legitimize the results found.

The empirical research was conducted in an online format, which made it possible to conduct interviews with specialized professionals who would be difficult to physically gather in a single location. In addition, this strategy was adopted due to the pandemic period. Research indicates that participants prefer the convenience of online focus groups due to the flexibility in scheduling and the ability to participate from home or the office [39], and the distinction between the "real" and the "virtual" environment is becoming increasingly difficult [40].

The sample consisted of nine professionals with experience in Big Data projects (Table 3), and the sessions are managed by an interviewer and a facilitator. Regarding the discussion about the number of participants, some care with the number of participants is necessary so as not to lose control of the discussion [41]. The adoption of an online focus group determines that the participants must be extremely qualified for such an application. Despite the suggestion of a number between 6 to 8 participants for face-to-face sessions, this number can be reduced to ensure the achievement of the objectives of the online discussion.

It is worth mentioning that the technological means represent a challenge to be overcome because in an online session mediation is more difficult.

Table 3. Data from the interviews in the focus groups.

Session	Data	Respondents' Background	Highest Degree	Session Duration (min)
01	27 May 2021	#1—Information Technology #2—Business Administration #3—Information Technology	Ph.D. Ph.D. Specialist	81
02	8 June 2021	#4—Physics #5—Information Technology #6—Information Technology #7—Computer Science	Master's degree Specialist Master's degree Specialist	65
03	12 July 2021	#8—Information systems #9—Mathematic	Specialist Specialist	57

Session 01 was composed of two professors from stricto sensu programs and a master's student as interviewees. Session 02 was composed of a team of four professionals who are part of the Data Science department of a private car rental company. Session 03 was composed of two professionals from a private bank. Each focus group session started with a brief introduction by the interviewer and facilitator, and then the model was presented in its entirety. As a data collection procedure, a non-structured questionnaire was used to give the interviewees freedom to comment and discuss the most relevant points of the model and, with that, the interviewees were able to present their respective points of view. All sessions were recorded with the consent of the interviewees, and the 3 interviews totaled 3.38 h.

As a data analysis procedure, after each focus group, the recording was watched to review the suggestions and questions made by the interviewees. All notes were listed and discussed between the interviewer and the facilitator, to reach a consensus on whether to accept the suggestion or not. Then, the accepted suggestions were incorporated into the model, and the next focus group was carried out with the revised model. It is noteworthy that through interviews with focus groups, it was possible to validate the framework with professionals who work with Big Data projects and to present a management framework for structuring Big Data projects.

3. Results

3.1. Background of a Big Data Project

The first relevant aspect of structuring a Big Data project is the understanding of what this type of project is. In this sense, the interviewees' understanding of the definition of a Big Data project showed convergence in several aspects. For E1, in the past, Big Data was just a large report; E7 adds that the change from more traditional Analytics to Big Data occurred when it was necessary to start having an interaction almost in real-time with the data. For E8, the Big Data project is as if it were the natural evolution of working with large volumes of data because knowing the data structure and needing a shorter processing time, there is a need to be more aggressive and then we choose a Big Data project.

According to E10, a Big Data project starts by challenging the *status quo* as it has the characteristic of being a transformation project that goes beyond the perspective of volume, velocity, and variety. For E6, a Big Data project is a project essentially oriented to data that can be structured or not, collected in real-time or asynchronously, but that involves a sufficiently large volume of data about some phenomenon, which aims to understand it in a context that cannot occur from pre-assumed premises. This phenomenon, as E2 is also called, breaks the paradigm that Big Data projects are essentially quantitative. The interviewee highlights the importance of also having maturity in qualitative analyses that allow the interpretation of correlations to help decision-making in Big Data.

Interviewee E5 also highlights the characteristic of distributed computing involved in the Big Data project, as there is a need for several computers to solve a problem. E4 understands that behind a Big Data project there is an ecosystem of systems through various data sets that make it possible to identify correlations that were not perceived when the data sets were analyzed in isolation. Therefore, E7 understands that we have moved from the age of data and information to the age of intelligence when it comes to Big Data projects.

In conclusion, E11 argues that the big difference between a Big Data project concerning traditional data analytical projects is the problem to be solved. For the interviewee, there is a great tendency to look to the future when it comes to a Big Data project, while in analytical projects only look at events that have already happened. With this, through Big Data projects it is possible to predict events or guide future actions as a way to help decision making, always respecting the nature of the data.

When we asked how a Big Data project was born, some interviewees highlighted a point that caught our attention. Despite the evolution of Big Data projects, in the interviews, the intention of companies to seek Big Data projects to be in the hype, that is, invest in this type of project for a fad, even without knowing what a project will serve Big Data, was highlighted by respondents E3 and E5. These types of projects, according to E4, are projects that are born by chance and since companies do not know what the project will be for, they may often not have a very objective focus.

On the other hand, there is a convergence on the understanding that initially there must be a problem, and that the Big Data project can be a means to solve this problem (E5, E7, E11, and E12) or even to leverage a business (E8). In this case, the project is born when a company has a perception that it needs to make more assertive and faster decisions based on a large volume of data from different sources (E2). This type of situation, according to E4, involves targeted projects. There are also visionary projects, and as E4 argues; these are projects that may involve feeling, that is, intuition, as they are often oriented to what is believed, and this must be pursued until data niches are structured to the point of extracting some positive insight. For E10, this type of project often occurs with a work team that seeks to innovate and that does not necessarily have a business problem such as demand.

Another point highlighted by E6 concerns the paradigm that the project will be involved in. According to the interviewee, when assumptions about the problem to be solved are assumed, the project is inserted in a model paradigm and when there are few assumptions for the problem, the project is inserted in the data paradigm, and in this case the greater the volume of data the better. In a complementary way, the Big Data project comes from the need of each sector, each professional, and the existence of tools that allow working with data (E8).

3.2. Phases of a Big Data Project

The description of the phases of a Big Data project was perhaps the question with the least convergence in the answers of the interviewees. For E5, a Big Data project does not change much about another technology project, except for the technology implemented. For E8, the phases of a Big Data project depend on the maturity of the client or the company. E7 understands that in general, it starts with the definition of a clear problem, and with that a highly adaptable architecture is structured.

Interviewee E7 still brings a contextualization with the deliverables of a Big Data project, and according to the interviewee, in the first sprint the entire data map is delivered; in the second sprint, the infrastructure is prepared and then it begins to extract the data and, in parallel, the analysis by objective is carried out. In the third Sprint, construction takes place and, at this moment, there is a check regarding the quality of the data, and with that there is a pattern concerning whether a product or customer is missing. According to E7, this is the first delivery; after that, the next deliveries are based on reports, analysis, and activation of solutions.

Complementary to the understanding of E7, interviewee E10 argues that from the information generated with reports and analyses, there are insights that are perceptions of a moment about time. This makes it possible, for example, to structure a Data Warehouse by collecting data aimed at having this type of information as a process that is part of the business; that is, obtaining value in the Big Data project occurs when the model that was generated to respond a problem becomes part of the business process.

For E6, the phases of a Big Data project include the definition of the problem, modeling of the structure of the data that will be used, data collection, cleaning or pre-treatment of the data, definition and implementation of the methods that will be used, validation of the pre-results, extrapolation of the pre-tests to a base that is possibly larger and closer to the real base, and finally the presentation of the results. There are also references to verify the availability of data and availability of tools (E9, E11) and also compliance for information security (E3). Based on what was evidenced in the speeches and the previous research, it was possible to build a flow of procedures that make up the phases of a Big Data project that is presented in the discussion section of this article.

3.3. Critical Factors to Enable a Big Data Project

By tabulating the critical factors cited by the interviewees, it was possible to identify 17 factors, with emphasis on the critical factors of Soft Skills, Technology, Hard Skills, and Planning, which were cited respectively 8, 7, 6, and 6 times by the interviewees (Table 4).

Table 4. Critical factors in Big Data projects.

Critical Factor/Interviewees	E5	E10	E11	E1	E2	E4	E12	E9	E3	E7	E8	E6	Total
Soft Skills		x		x		x	x	x	x	x	x		8
Technologies	x			x	x	x		x			x	x	7
Hard Skills				x			x	x	x	x	x		6
Plan			x		x	x			x	x		x	6
Data availability							x	x	x	x		x	5
Partnership		x		x			x		x			x	5
Holistic vision	x								x	x	x	x	5
Risk management				x		x		x	x			x	5
Hype	x		x		x	x			x				5
Bias			x		x	x					x	x	5
Data quality			x				x			x	x		4
Business		x					x				x		3
Expectancy								x				x	2
Politic					x		x						2
Modeling										x	x		2
MVP-Minimum Viable Product										x		x	2
Analysis techniques								x					1
Total	3	3	4	5	5	6	7	7	8	8	8	9	

According to E1, Big Data projects are highly technical projects, so it is essential to have a technical team with experience. For E9, what makes a professional work with Big Data in the first place is some hard skills. According to the interviewee, this opens many doors and encourages the professional and the company to position themselves as working with Big Data, but after this position, soft skills become essential. E12 also highlighted the role of the project manager, insofar as a Big Data project is not a simple project, it is a project that involves many people and invariably a lot of need for negotiation, persuasion, and resilience on the part of the manager.

On Technology and Planning, E2 mentions that "we joke that the Information Technology area grinds people because the manager does not plan information technology". The use of inappropriate technologies was highlighted as a highly critical factor. E4 also men-

tioned situations where projects started simply because decision-makers wanted to work with new technology. Interviewee E8 also adds that if the infrastructure was developed that cannot process data at the speed that decision-makers need, the effort was in vain.

Regarding the availability of data, E3 argues that the larger the company, the more dispersed knowledge becomes, and this often makes access to data difficult. Interviewee E9 deals with the contradiction involving the availability of data, because if the objective is a very specific answer, it is common that the data available easily will not give the desired answer and this implies replanning to collect more specific data sets that often involve additional costs. Regarding the E12, when arguing that access to data sources serves to make the project technically viable, the following questions should be taken into account: How do I treat this data? How do I store this data? How do I transform this data? How do I delete this data? Are there any implications in the General Data Protection Law (GDPL) that I have to consider?

Seeking partnerships as a good practice was also considered a critical factor by the interviewees, for E10, bringing partners with know-how in the area enables acceleration in the learning curve and promotes a culture of collaborative approaches, which is so important in the execution of a Big Data project. E12 argues about the importance of supporting project stakeholders such as sponsors, area directors, and even CEOs (Chief Executive Officer) to ensure that the necessary actions are taken since digital transformation projects involve a lot of politics and many conflicting interests.

When referring to the holistic view in a Big Data project, the interviewees highlighted the importance of a business-oriented view. E7 mentions that in projects that he participated in, and that were considered successful, there were always professionals with business expertise who sometimes did not even know about Big Data technologies. E8 adds that if there is no clear direction to where you want to go, there is a great chance that the project will fail because what E3 presents as a delivery that the client does not need occurs.

Risk management also proved to be important within a Big Data project. E1 highlights risks involved with process modeling, data storage, and distribution. Interviewee E4 mentions the "escape from the topic" resulting from a poor requirement gathering in the initial phase of the project, resulting in problems of cost and time deviation, or even making the project unfeasible. As E9 commented, "*despite the great availability of data and solutions that have been created, not every project is possible to be carried out, so this is the biggest challenge and it happens much more than we can estimate*".

Regarding the hype, there was a convergence in the understanding that many companies want to use Big Data to follow a trend. For E3, "*everyone wants it, but few people understand what a Big Data project is really for*". This adoption of Big Data alone brings as a consequence the difficulty of in many cases proving value to the business (E5), and as E4 comments, decisions driven by hype result in a project without a very defined objective. E11 adds that hype generates as its main consequence the moment of perception in which the technology adopted does not make strategic sense for the organization.

Interviewee E11 argues about bias that "a major flaw in these projects that guide decision-making is related to people believing in the data without respecting the fact that the analysis was created not to work in the world, but to work in the world that you analyzed" and this, according to the interviewee, requires a change in the mindset of the people involved with Big Data projects. For E8, today we talk about data projects, but with data that we do not know, and with dubious quality. E2 mentions that having intelligent systems is useless if you do not have intelligent people capable of interpreting the data and this creates problems, for example, bias generated by the amount of data analyzed.

Another critical factor highlighted concerns the expectations generated by Big Data projects. For E6, people create sometimes overestimated expectations of a project, and underestimate or neglect the ability to obtain data. As already mentioned by E9, today not every Big Data project is possible. In this aspect, the proper gathering of requirements, modeling, and management of expectations converge as good practice in conducting a project with these characteristics.

Regarding the modeling of a Big Data project, E7 highlights that the error in the modeling makes the entire Big Data project unfeasible. E8 also reinforces that when talking about a modeling process, one of the most important steps is the revisiting of the model because revisiting it allows the calibration according to the perception throughout the project. This revisiting even helps the possibility of providing deliverables throughout the project, which helps in the perception of project value (E6).

3.4. Competencies of the People Involved in a Big Data Project

The skills of the people involved in Big Data projects are crucial for their effectiveness to be achieved. Thus, in this research, the specificities related to technical skills (hard skills) or behavioral skills (soft skills) that stand out in Big Data projects were observed. For E1, Big Data projects are highly technical projects and therefore hard skills must stand out. E4 understands that personal skills, related to soft skills, are more decisive in Big Data projects. E11 adds that hard skills make projects viable, but soft skills differentiate their solution.

For E10, as Big Data works well with agile methodology, soft skills end up being of great importance because they include collaborative work processes in tune between team members. E7 argues that increasingly horizontal organizations provide conditions for people to give their opinion and have the same opportunity to speak out as their peers, but more important than the skills themselves is the ability to learn quickly, whether soft or hard skills.

From another point of view, E6 understands that the type of project will influence the profile required for a professional involved with a Big Data project. For the interviewee, in more self-contained, specific, and closed-scope projects, hard skills stand out and the more open the problem to be solved, there is a balance between hard and soft skills. E8 highlights how difficult it is to become a professional with this combination of skills. For E11, the great soft skill that will impact the success of the project is to understand that those involved with the project are human and that, on the other hand, those who will use the project delivery are also human.

3.5. Decision-Making Issues in a Big Data Project

Concerning decision-making problems in a Big Data project, E2 highlights that in many projects, decisions are not made based on technical criteria, but on political criteria and, as a consequence, investment in technologies that do not always meet the needs that eventually emerge. E4 complements that inadequate technological directions hinder the good conduct of a Big Data project, with financial impacts or even loss of engagement of the project team for discrediting the possibility of obtaining an adequate result at the end of the project.

Respondents E3, E5, and E11 comment that the hype, that is, the decision to join a Big Data project just to follow a trend, can generate as a consequence the bias in wanting some result even without having adequate modeling, or well-defined metrics for the robustness of the project. E9 argues as a challenge, the difficulty in converting the project into a process so that the information extracted helps users to make a decision, the interviewee also adds that the decision on the part of the customers to change the objectives of the projects are serious problems that can compromise the entire project feasibility.

E12 cites as the main decision-making in a Big Data project the choice of the manager who will lead the project, as it will trigger subsequent decisions throughout the project by the manager, which can be described in "*how to do the projects by the characteristics of the projects*". According to the interviewee, this is reflected in the choice, for example, of the management methodology that is also influenced by the use of methods disseminated in the market as saviors, but which do not meet the real need for management.

3.6. Big Data Design Frameworks

When asked about knowing or using any methodology, guide, or roadmap in conducting Big Data projects, there were different perceptions about this question. E2 and E11 see

the PMBoK (Project Management Body of Knowledge) guide as suitable for conducting their projects, where they highlight the project as being seen in a "traditional" way. E1 and E4 also state that managing a Big Data project is the same as managing a software project.

For E10, "being data-driven works very well with the modern software development model through agile methodology", for the interviewee, it is very important to have the "agile mindset applied to Big Data so that the project does not become a waterfall". E6 highlights the use of the scientific method adapted to the context of the project, as there is a definition of a problem, raising hypotheses, obtaining data, and validating the hypotheses or not. The interviewee also highlights the importance of fragmented deliveries to enhance the perception of value in deliveries and also mitigate errors or scope deviations.

Interviewee E12 highlights the importance of taking into account the characteristics of each project and then looking for a form of management that applies to each situation. For E7, two pillars should influence Big Data projects, which are: (i) Governance, based on responsibilities, data ownership definitions, and access permissions; (ii) Modeling referring to data architecture. Because of these aspects, it is noteworthy that frameworks based on the characteristics of each project and available resources are well-applied, and not necessarily on a methodology or guide.

3.7. Opportunities in Big Data Projects

With the last question of our questionnaire, we sought to identify research opportunities from the perspective of the interviewees and to hear suggestions of what could be developed to assist the management of Big Data projects. Within the context of the hype problem involved in a Big Data project, E2 suggests the development of a planning tool that guides decisions based on technical criteria. The perception that a framework would have value was shared by E3, E5, E7, E8, and E11. E8 also highlights that Big Data can be applied in any company, or any business; the difficulty is to have qualified professionals who can properly structure the project life cycle to obtain results. By arguing that the academy needs to undergo an empirical transformation, E11 suggests the elaboration of an ethical framework in the area of Big Data. According to the interviewee, there is little discussion about the ethical aspect of Big Data projects when compared to the discussion about quality or safety, and a phase about ethics as a process within Design Thinking could even be developed.

In the opinion of E4, there is a lack of a "*methodology that is not too rigid and not too fast*", that is, something not so bureaucratic, but that provides control of the processes and that would be an intersection of the waterfall and agile methodologies, and which he called "*a methodology agile documentable*". E12 suggests the development of a master line for conducting Big Data projects considering the specifics of each project, and E8 proposed the development of management metrics associated with the methodology and with that to measure efficiency and performance throughout the management.

E6 suggests the development of material for professionals who will perform the role of manager of a Big Data project, but who are not in the project management area. With the increasing interest in the Big Data area, professionals end up being inserted in this environment, but they do not always master management techniques and concepts, and this ends up being a barrier in the conduct of these types of projects, so the interviewee understands that the development of something that would support this kind of situation would be of great value.

4. Framework Proposition and Validation

As highlighted earlier, despite the significant progress made in the area of Big Data in recent years, practitioners highlight important research opportunities mainly related to guides, models, or even methodologies [23]. Based on previous information about background, competencies, models, etc., the researchers moved on to a phase of compiling information and building propositions. In this way, we understand that an opportunity to contribute to the community of professionals who work with Big Data projects would

be to synthesize the structure of a Big Data project, and with that, the framework presented in Figure 1 was developed based on the information collected and analyzed of the 12 individual interviews.

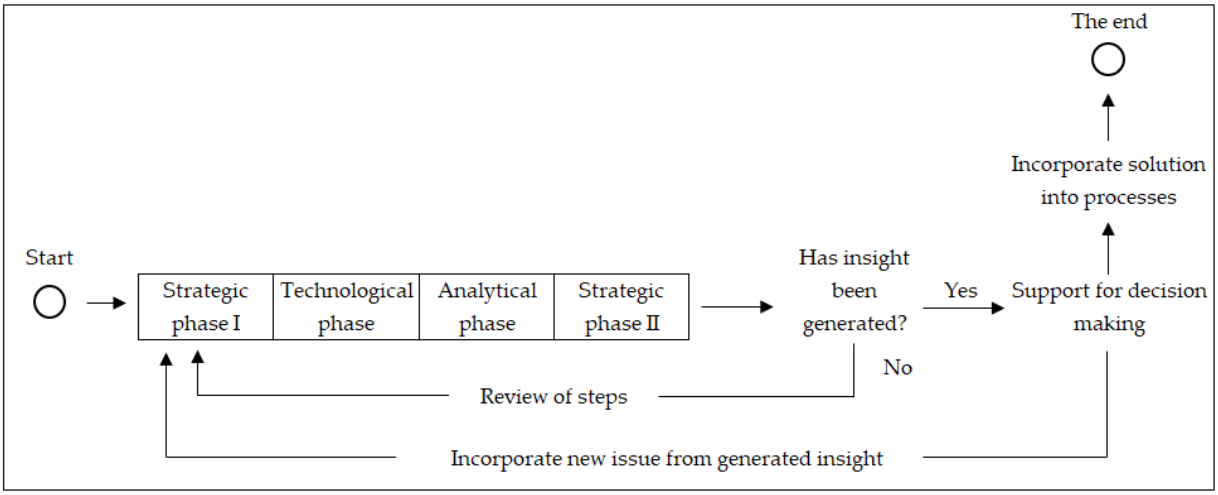

Figure 1. Framework for structuring a Big Data project.

The model presented has a managerial aspect that aims to provide support in structuring and macro decisions. Given the complexity of Big Data projects, we understand that the higher the level of detail, the model would be more restricted to specific projects. That is why we kept a managerial aspect to contemplate a greater number of projects.

From our analysis, we can conclude that a Big Data project has a first strategic phase that starts with an external trigger and that guides the type of problem to be analyzed and solved. Problems can be pulled (when it is a latent and clear problem) or pushed (when it is a problem with a bias of some actor involved in the environment). Then, the management risks must be mapped to anticipate adverse events that can compromise the progress of the project, the management of expectations must be carried out with the project's stakeholders for an alignment of the expected results and possible situations of non-achievement of the objectives. In this sense, the manager of a Big Data project should also carry out a feasibility analysis of the project together with experts in the field, and consider lessons learned from previous projects to mitigate the chance of repeating past mistakes. With this, the hypothesis or objective to be achieved is defined, and one must ask what type of analysis the problem is conducting: descriptive analysis for objectives, and predictive or prescriptive analysis for hypotheses.

The second phase of a Big Data project is the technological phase, a phase in which any ethical or legal impact on what you want to analyze must be analyzed, as well as the definition of data access policies. There is an increasing demand for transparency and visibility in the use of collected data, so every Big Data project must comply with current legislation. These definitions will influence the architecture to be modeled for the Big Data project. After modeling, there is the collection and storage of data as applicable.

The third phase of a Big Data project is the analytical phase that begins with the preprocessing of the data and with that, it is possible to have a data treatment when necessary. So, based on the type of analysis defined in the strategic phase, we have supervised, semi-supervised, or unsupervised learning algorithms in cases of prescriptive or predictive analysis, and Big Data Analytics in cases of descriptive analysis. In both situations, it is advised that experts legitimize the results generated as a way of giving more credibility to the findings, mitigating the possibility of some result that is out of touch with reality because some variable was not considered in the analysis.

Following the framework flow, there is the second strategic phase, which starts from the result generated in the analytical phase and must answer the problem (pulled or pushed)

that gave rise to the Big Data project. With this, it must be questioned whether any insight was generated with the result of the previous phase; if not, the phases must undergo revisions and, if so, the insight conditions the support for decision-making, enabling new organizational processes. We strongly recommend recording lessons learned throughout the project to serve as an output in new Big Data projects. With the insight legitimized, a recommended analysis is to discuss with the expert group whether the insight generated a new trigger to start a new Big Data project. If so, the flow must be resumed at the beginning of the framework with the new questioning and definition of the type of problem.

The main benefit of this framework is the sequential integration and structuring of the phases that comprise a Big Data project so that no phase is neglected by the project team. This was verified during the sessions with the focus groups, and according to the interviewees, it is something valuable, because, in situations experienced by them, the lack of a holistic view of the project meant that many steps were neglected, generating rework and losses. The process of validating the management framework took place with the full presentation and, later, the interviewees made comments indicating points of attention, asking questions, and suggesting improvements. Every suggestion for improvement was evaluated after the session. It should be noted that not all suggestions were included in the framework, as they were evaluated regarding the feasibility and focus of the proposal. It was also common for the interviewees to discuss themselves to legitimize some statements made with an example from their daily lives.

Accordingly, after the implementation of the focus group sessions, there were more suggestions for improvements for the first strategic phase, six in all (Figure 2). The concern with the first phase of the framework was evident to direct the subsequent phases.

Session	Strategic phase I	Technological phase	Analytical phase	Strategic phase II	TOTAL
1	3		2	1	6
2	2	1	2		5
3	1	1		1	3
TOTAL	6	2	4	2	

Figure 2. Suggestions for improvements made throughout the discussion.

In the technological phase, with two suggestions, there was an increase in the process of adapting the project according to legislation or rules, where the definition of data access policies was also considered. In the analytical phase, with four suggestions, a process of legitimation by specialists was incorporated, in which a multidisciplinary team with the ability to perceive the context in which the problem is inserted must be composed. These experts can mitigate the risk of a variable not being considered, or even sharing tacit knowledge, which is often difficult to incorporate into processes.

An outstanding point was that at each session, fewer suggestions were given to indicate an increasing acceptance of the framework, since, with fewer suggestions, the legitimation and agreement of the processes by the interviewees were noticed. At the end of the three sessions, as a way of seeking acceptance of the framework and considering it appropriate for use, the following question was asked to the focus groups: "*Do you understand that the framework can be useful when used as a guide for structuring Big data projects*?". In the three sessions, the answers were positive, and in one of the sessions, one of the interviewees even went back to a situation he had experienced and, according to his speech, if he had this framework at the time, he would have avoided a problem that he ended up experiencing. With the validations evidenced in each session, the framework was considered validated by experts who work professionally with Big Data projects.

5. Discussion

A Big Data project, as defined by Becker [21], is a data-intensive project with volume, variety, and velocity issues. In addition to the analysis process, we note the emphasis placed

on the need to have a business problem be solved as a relevant aspect of Big Data projects. In this sense, we understand that a predecessor phase to the project itself, but which should not be neglected, is the perception of which problems need to be solved. As suggested by Mousanif et al. [25], we can question whether these problems can be solved with software and techniques for conventional data analysis. Due to this, we can infer that a Big Data project can be defined as a way of solving a complex problem, through distributed computing technologies and data analysis techniques that have volume, variety, and velocity as characteristics, which results in information capable of assisting decision making.

To obtain the most benefits from Big Data, practitioners and researchers should not only focus on the data and data analysis aspect when studying and developing a Big Data project to support decision making, but also go beyond that and study the decision-making process aspect related to this type of project [16]. In our study, we noticed a maturity on the part of the interviewees for not neglecting the aspect of decision making; however, we noticed that external factors outside the scope of managers can impact the quality of decisions, such as, for example, the bias of the executive level in desiring some type of result, or even influencing the acquisition of a technology that is not the most appropriate.

When we analyze the phases of a Big Data project according to the perception of some authors (Table 5), we find that there is no consensus on the stages or phases that cover the project. For Mousanif et al. [26] the flow of a Big Data project is linear, starting with the elaboration of the global strategy as a form of planning, followed by the implementation stages and, finally, with the post-implementation stage. When presenting a holistic map for the implementation of Big Data projects, Dutta and Bose [7] presented a framework with the phases of strategic basis, data analysis, and implementation.

Table 5. Big Data Project Flow.

Mousanif et al. [26]	**Dutta and Bose [7]**	**Dietrich and Reiner [42]**
(1) Elaboration of global strategy	(1) Business Problems	(1) Explore the business context and frame the hypothesis
(2) Data collection	(2) Research	(2) Discover and ingest data sources
(3) Data processing	(3) Cross-functional team building	(3) Integrate and prepare data
(4) Intelligent data analysis	(4) Project roadmap	(4) Plan model, methods, and key variables
(5) Representation and visualization	(5) Data collection and examination	(5) Develop test, dataset, and build model
(6) Extracting actionable and timely insights	(6) Data modeling and analysis	(6) Applying the model to the analysis and decision scenarios
(7) Assessment	(7) Data visualization	(7) Engage and operationalize the model
	(8) Insights generation	
	(9) IT system integration	
	(10) Training people	
Linear	Cyclic	Cyclic

In line with this discussion, Dietrich and Reiner [42] registered an invention, which is a computer program to provide a computing system according to an analytical data lifecycle. According to the inventors, the process meets emerging needs related to Big Data projects and is effective in obtaining operational benefits from Big Data sources. In a survey of how teams work together to execute Big Data projects, Saltz and Shamshurin [28] infer that there is no agreed standard for executing these projects, but there is a growing focus on research in this area and that a process methodology enhanced would be helpful. The authors also identified 33 critical factors for the execution of Big Data efforts, which are grouped by six identified characteristics of a mature Big Data organization which are: Data, Governance, Processes, Objectives, Team, and Tools. Additionally, in an investigation of critical factors in Big Data projects, Gao et al. [29] presented 27 factors divided into six phases, namely: Business, Data, Analysis, Implementation, Measurement, and General.

In our research, we identified 17 critical factors, with the highest incidence for soft skills, technology, hard skills, and plan. In this sense, we conclude the great criticality by the perception of the interviewees are given with the skills of the people involved with the projects, also with the acquisition and use of appropriate technologies, and still with the adequate planning for the execution of Big Data projects. These critical factors converge with decision-making problems identified in our survey responses, and as stated

by Chiheb et al. [16], many Big Data projects deliver disappointing results that do not meet the needs of decision makers for reasons such as, for example, neglecting to study the decision-making aspect of projects.

When investigating which approach should be applied in the management of Big Data projects, Franková et al. [19] found that the agile approach is preferable due to the frequent change of requirements arising from new questions that arise throughout the project. In our analysis, we noticed great adherence to the use of agile methods as managers seek to deliver fragmented results with small use cases, accepting failures and promoting successive revisions to follow an iterative orientation.

As also evidenced by Franková et al. [19], we found that the choice of approach in managing Big Data projects is conditioned not only by the size and criticality of the project but also by the dynamics of the environment, by the skills of those responsible for the work of the project, and the organizational culture. For Jin et al. [19], the necessary conditions for a Big Data project to be successful are related to having very clear technical, social, and economic requirements, having data structures small enough to characterize the behavior be analyzed, and a top-down management model should be sought to promote comprehensive resolutions through an integrated approach. We identify relationships in several aspects, mainly in the relationship between the environment in which Big Data projects are inserted and the team's expertise and organizational culture.

6. Final Remarks

Through a qualitative study, we seek to investigate and contribute to the discussion about Big Data projects. We set out to present some definitions and answers that we believe are appropriate for the moment we are going through, where there is an increasing movement towards investments in technology projects and, consequently, in Big Data projects.

As the main theoretical contribution, our study aligns with a growing body of researchers who are proposing to discuss and contribute to the topic of Big Data projects, and with that increasing maturity on this topic. As a practical contribution, we present a management framework, and with this we hope to contribute with professionals who work in the area, helping to conduct a Big Data project through a holistic view. It starts from the premise that the framework, from the managerial aspect, can be applied to projects in different areas and sectors, a characteristic assimilated by the interviewees during the framework validations. The framework aims to support professionals to structure a Big Data project and make better decisions regarding the project structure, reducing the chance of failure of these projects.

As a limitation, we can consider that the number of respondents could be higher to make the study even more robust. The pandemic period during which the study was carried out ended up making access difficult for some professionals, who, between rejection to participate in the research and requests without answers, totaled 12 invitations.

In one of the focus group interview sessions, one of the interviewees expressed his desire to have a more technical detailing of the framework, especially for the technological phase. This finding has become an important opportunity for future research as new studies can delve into the model with complementary technical levels or studies aimed at a greater level of detail in a specific phase. The possibility of applying the model in a Big Data project is also highlighted since the validation was carried out by the perception of the interviewees.

Author Contributions: Conceptualization, L.F.D.S. and G.G.; Formal analysis, G.G. and R.P.; Investigation, G.G. and L.F.D.S.; Methodology, L.F.D.S.; Validation, E.D.R.S.G.; Writing—original draft, G.G.; Writing—review and editing, L.F.D.S., R.P. and E.D.R.S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported in Brazil by CAPES—Coordination of Personnel Improvement for Higher Education: Code 001 and CNPQ—National Council for Scientific and Technological Development; and funded by FONDECYT No. 1190559.

Acknowledgments: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES); and funded by FONDECYT No. 1190559.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grander, G.; da Silva, L.F.; Gonzalez, E.D.R.S. Big data as a value generator in decision support systems: A literature review. *Rev. Gestão.* **2021**, *28*, 205–222. [CrossRef]
2. Tang, L.; Li, J.; Du, H.; Li, L.; Wu, J.; Wang, S. Big Data in Forecasting Research: A Literature Review. *Big Data Res.* **2022**, *27*, 100289. [CrossRef]
3. Chen, W.; Quan-Haase, A. Big Data Ethics and Politics: Toward New Understandings. *Soc. Sci. Comput. Rev.* **2018**, *38*, 3–9. [CrossRef]
4. Boyd, D.; Crawford, K. Critical questions for big data Provocations for a cultural, technological, and scholarly phenomenon. *Inf. Tarsad.* **2012**, *15*, 7–23. [CrossRef]
5. Demirkan, H.; Delen, D. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decis. Support Syst.* **2013**, *55*, 412–421. [CrossRef]
6. Grander, G.; da Silva, L.F.; Gonzalez, E.D.R.S. A Patent Analysis on Big Data Projects. *Int. J. Bus. Anal.* **2022**, *9*, 14. [CrossRef]
7. Dutta, D.; Bose, I. Managing a big data project: The case of Ramco cements limited. *Int. J. Prod. Econ.* **2015**, *165*, 293–306. [CrossRef]
8. Barham, H.; Daim, T. Identifying critical issues in smart city big data project implementation. In Proceedings of the 1st ACM/EIGSCC Symposium On Smart Cities and Communities, SCC 2018, Portland, OR, USA, 20–22 June 2018. [CrossRef]
9. Fedushko, S.; Ustyianovych, T.; Gregus, M. Real-Time High-Load Infrastructure Transaction Status Output Prediction Using Operational Intelligence and Big Data Technologies. *Electronics* **2020**, *9*, 668. [CrossRef]
10. Wu, X.; Zhu, X.; Wu, G.W. Ding, Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 97–107. [CrossRef]
11. Berman, J.J. Big Data Failures and How to Avoid (Some of) Them. *Princ. Pract. Big Data* **2018**, 321–349. [CrossRef]
12. Kappelman, L.A.; McKeeman, R.; Zhang, L. Early warning signs of it project failure: The dominant Dozen. *Edpacs* **2007**, *35*, 1–10. [CrossRef]
13. Saltz, J. The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015.
14. Saltz, J.; Shamshurin, I.; Connors, C. A Framework for Describing Big Data Projects. *Int. Conf. Bus. Inf. Syst.* **2017**, *263*, 183–195. [CrossRef]
15. Kaisler, S.; Armour, F.; Espinosa, J.A.; Money, W. Big data: Issues and challenges moving forward. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 7–10 January 2013.
16. Chiheb, F.; Boumahdi, F.; Bouarfa, H. A conceptual model for describing the integration of decision aspect into big data. *Int. J. Inf. Syst. Model. Des.* **2019**, *10*, 23. [CrossRef]
17. Lavalle, S.; Lesser, E.; Shockley, R.; Hopkins, M.S.; Kruschwitz, N. Big Data, analytics and the path from insights to value. *MIT Sloan Man. Rev.* **2011**, *52*, 20–31.
18. Angrave, D.; Charlwood, A.; Kirkpatrick, I.; Lawrence, M.; Stuart, M. HR and analytics: Why HR is set to fail the big data challenge. *Hum. Resour. Manag. J.* **2016**, *26*, 1–11. [CrossRef]
19. Franková, P.; Drahošová, M.; Balco, P. Agile Project Management Approach and its Use in Big Data Management. *Procedia Comput. Sci.* **2016**, *83*, 576–583. [CrossRef]
20. Jin, X.; Wah, B.W.; Cheng, X.; Wang, Y. Significance and Challenges of Big Data Research. *Big Data Res.* **2015**, *2*, 59–64. [CrossRef]
21. Becker, D.K. Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD). In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 2320–2330. [CrossRef]
22. Günther, W.A.; Mehrizi, M.H.R.; Huysman, M.; Feldberg, F. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* **2017**, *26*, 191–209. [CrossRef]
23. Lara, J.A.; De Sojo, A.A.; Aljawarneh, S.; Schumaker, R.P.; Al-Shargabi, B. Developing big data projects in open university engineering courses: Lessons learned. *IEEE Access* **2020**, *8*, 22988–23001. [CrossRef]
24. Barham, H.; Daim, T. The use of readiness assessment for big data projects. *Sustain. Cities Soc.* **2020**, *60*, 102233. [CrossRef]
25. Mousannif, H.; Sabah, H.; Douiji, Y.; Sayad, Y.O. From big data to big projects: A step-by-step roadmap. In Proceedings of the 2014 International Conference on Future Internet of Things and Cloud, Barcelona, Spain, 27–29 August 2014; pp. 373–378. [CrossRef]
26. Mousannif, H.; Sabah, H.; Douiji, Y.; Sayad, Y.O. Big data projects: Just jump right in! *Int. J. Pervasive Comput. Commun.* **2016**, *12*, 260–288. [CrossRef]
27. Muntean, M.; Militaru, F.D. Design Science Research Framework for Performance Analysis Using Machine Learning Techniques. *Electronics* **2022**, *11*, 2504. [CrossRef]

28. Saltz, J.S.; Shamshurin, I. Big data team process methodologies: A literature review and the identification of key factors for a project's success. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 2872–2879. [CrossRef]
29. Gao, J.; Koronios, A.; Selle, S. Towards a process view on critical success factors in Big Data analytics projects. *AMCIS* **2015**, *2015*, 1–14.
30. Creswell, J.W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; Sage Publications: London, UK, 2017.
31. Guion, L.A.; Diehl, D.C.; McDonald, D. *Conducting an in-Depth Interview*; University of Florida: Gainesville, FL 32611, USA, 2011.
32. Boyce, C.; Neale, P. *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*; Pathfinder International Tool Series: Watertown, MA, USA, 2006.
33. Turner, D.W. Qualitative Interview Design: A Practical Guide for Novice Investigators. *Qual. Rep.* **2010**, *15*, 754–760. [CrossRef]
34. Dilley, P. Conducting successful interviews: Tips for intrepid research. *Theory Pract.* **2000**, *39*, 131–137. [CrossRef]
35. Charmaz, K. *Constructing Grounded Theory: A Practical Guide through Qualitative Research*; Sage: London, UK, 2006.
36. Saldaña, J. *The Coding Manual for Qualitative Researchers*; Sage Publications: London, UK, 2012.
37. Silva, A.B.; Godoi, C.K.; Melo, R.B. *Pesquisa Qualitativa em Estudos Organizacionais*; Saraiva: São Paulo, Brazil, 2010.
38. Oliveira, M.; Freitas, H. Focus group, pesquisa qualitativa: Resgatando a teoria, instrumentalizando o seu planejamento. *Rev. Adm. Da Univ. São Paulo.* **1998**, *33*, 83–91.
39. Zwaanswijk, M.; Van Dulmen, S. Advantages of asynchronous online focus groups and face-to-face focus groups as perceived by child, adolescent and adult participants: A survey study. *BMC Res. Notes* **2014**, *7*, 756. [CrossRef]
40. Stewart, D.W.; Shamdasani, P. Online Focus Groups. *J. Advert.* **2017**, *46*, 48–60. [CrossRef]
41. Krueger, R.A. *Focus Groups: A Practical Guide for Applied Research*; Sage Publications: London, UK, 2014.
42. Dietrich, D.I.; Reiner, D.S. Holistic Methodology for Big Data Analytics; US9798788B1; 2017. Available online: https://worldwide.espacenet.com/patent/search/family/060082219/publication/US9798788B1?q=US9798788B1 (accessed on 18 June 2021).