

Review

Big Data Analytic Framework for Organizational Leverage

Sanjay Mathrani *  and Xusheng Lai

Operations and Engineering Innovation, School of Food and Advanced Technology, Massey University, Auckland 0632, New Zealand; gdhatake@hotmail.com

* Correspondence: s.mathrani@massey.ac.nz

Abstract: Web data have grown exponentially to reach zettabyte scales. Mountains of data come from several online applications, such as e-commerce, social media, web and sensor-based devices, business web sites, and other information types posted by users. Big data analytics (BDA) can help to derive new insights from this huge and fast-growing data source. The core advantage of BDA technology is in its ability to mine these data and provide information on underlying trends. BDA, however, faces innate difficulty in optimizing the process and capabilities that require merging of diverse data assets to generate viable information. This paper explores the BDA process and capabilities in leveraging data via three case studies who are prime users of BDA tools. Findings emphasize four key components of the BDA process framework: system coordination, data sourcing, big data application service, and end users. Further building blocks are data security, privacy, and management that represent services for providing functionality to the four components of the BDA process across information and technology value chains.

Keywords: big data; big data analytics; big data process; big data analytic framework



Citation: Mathrani, S.; Lai, X. Big Data Analytic Framework for Organizational Leverage. *Appl. Sci.* **2021**, *11*, 2340. <https://doi.org/10.3390/app11052340>

Academic Editor: Federico Divina

Received: 18 February 2021

Accepted: 3 March 2021

Published: 6 March 2021

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



Copyright: © 2021 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 have seen exponential growth over the past few decades, changing the information availability landscape and everything it interacts with. The big data concept was primarily introduced in the 1990s [1] but it was not until the early 21st century when it had its revolutionary breakthrough, evolving from decision support and business intelligence (BI) systems [2]. Now, big data refers to huge datasets comprising structured, semi-structured, and unstructured data with challenges in storage, analysis, and visualization for more processing [3].

Data are being generated at phenomenal rates. Recent research reports that only 5 exabytes of data were generated by humans when the concept of big data first gained emphasis [4]. Currently, this many data can be created in a day or two. In 2010, the world generated over 1 zettabytes of data, which was expanded to 2.7 zettabytes in 2012. This number was predicted to double every two years. Being as humongous as it is, big data have overwhelmed traditional data mining tools and now requires a different approach, which has led to the development of big data analytics (BDA). This is the process of sieving through massive and complex data streams using novel technologies such as business intelligence, cloud computing, and machine learning to disclose hidden correlations and patterns and presenting this information to business users in real time for effective decision-making [3]. The data are captured from different online platforms, such as from e-commerce activities, social media interactions, organizational and individual web sites, or from day-to-day information exchanges made by users. These data streams can also be generated from sensor-based data exchanges, such as Internet-of-Things (IoT) devices, radio-frequency identification (RFID) readers, microphones, cameras, mobile devices, and similar web-enabled equipment [5].

In recent decades, the emphasis on BDA has been increasing in organizations. It is considered to have changed the way in which organizations develop and grow [6].

The recent literature indicates that organizations that leverage BDA have been able to realize a significant return on investment by rendering customer focused services [7]. A recent case study by Teradata [8] reported that a British multinational telecommunications conglomerate trading firm, Vodafone New Zealand, was able to develop more finely targeted campaigns that had a higher chance of success by using big data to accurately predict the traits of their customers. In another renowned case of Continental Airlines [9], use of data warehousing and real-time BI helped to improve several performance metrics, resulting in its turnaround, moving it from 10th place to become the first ranked airline.

Although BDA pays off in some organizations, it has achieved little impact for others due to its innate difficulty in process development. As the literature has revealed [3,10,11], the application of BDA faces multiple challenges. First, it is difficult to set up a robust physical infrastructure which is based on a distributed model that can physically store the big data. A sophisticated architecture with thousands of nodes and multiple processors needs to be created (e.g., Hadoop framework), which comes at a substantial cost. This hardware infrastructure investment issue greatly increases the barrier to entry for organizations with less margin of investment to implement BDA [10]. Aside from this, organizations also face the challenge of hiring personnel who are trained in business analytics and implementation methods and know how to extract information that will have potential to improve business decision-making [12]. These challenges are especially apparent in developing countries as the vast majority of infrastructures and trained personnel usually reside in developed countries [10]. Even developed countries face a workforce shortage in this area. For example, the US, a country well-known for its technical advancements, is estimated to have had a shortage of 160,000 big data professionals by 2018 [13]. Therefore, organizations will have to train their own workforce to possess the necessary IT administration and application skill sets, which will further increase the cost of implementing BDA. Those who are experienced with BDA also face challenges, the main one being data communication. As a network-based application, BDA has an even higher communication cost in comparison to the processing cost. Therefore, a key challenge of BDA application is in minimizing the communication cost, keeping bandwidth and latency at a satisfactory level [14]. Another challenge that big data workforces typically face is security, as cyber attackers may tamper with the data under exchange or shut down servers with operating system attacks [11]. In big data management, the sheer amount of data entering BDA each minute becomes unaccountable. The server being shut down for even the slightest period can be detrimental to organizations.

Driven by these aspects, this paper examines the role of BDA in different organizations to develop an in-depth understanding of its capabilities and potential benefits. The existing research largely looks at the importance of big data analytics, its challenges, and the process of implementation [4,10,11,15]. Thus, this paper is motivated to explore the capabilities of BDA and its potential benefits by answering the following research question: how can organizations evaluate big data to achieve their strategic business objectives?

To address this research question, two key objectives of this research are (1) to identify the capabilities of big data analytics, and (2) to propose a theoretical framework that organizations can potentially implement to leverage big data. Case studies were conducted to examine how existing organizations leverage benefits through the BDA process. Primary data were collected using semi-structured interviews from six data analysts or associated technologists of three large companies (two participants from each company) who engage in frequent use of big data analytic tools. Questions were asked to understand the process of BDA implementation based on their experiences and how organizations source, process, analyze, and visualize big data to achieve their strategic business goals. Key findings identified four main components of BDA—system coordination, data sourcing, big data application service, and end users—each having a major significance in the BDA process to produce business insights. Findings from this research can assist in developing an architectural BDA framework of significance to both academic researchers and business professionals in identifying the major process activities and outcomes. The

study insights inform researchers and practitioners about the effectiveness of BDA and the implementation process.

The structure of this paper is as follows. First, we provide a background of big data analytics, its challenges, and the research intent of this study. Then, we review the published literature detailing present and future expectations. Next, we discuss the importance of BDA in different industry settings and the development of its capabilities and characteristics in the present context. The research methodology is described next, which explains the three cases and the theoretical framework used in this study. This is followed by the content analysis of the data collected from the different organizations and discussion of the findings, which lead to further development of the theoretical framework. Finally, the conclusions, limitations, and future research directions are presented.

2. Related Works

2.1. *Big Data Analytics—Its Past, Present, and Future*

Data generation is not new. The “big data” term was first used by Michael Cox and David Ellsworth in 1997 during their presentation at an IEEE conference [1], in which they explained data visualization features and associated issues [16]. Progression in information technology and applications with novel data sources is new (e.g., e-commerce) and has led to novel business opportunities [7]. In the past few decades, this concept has gained an explosion of interest and has generated tremendous attention worldwide. Searching for “big data” on Google yields more than 3 billion results. According to Akter et al. [6], the big data concept skyrocketed, and businesses became swamped with humungous amounts of data. These data could either be structured (data from fixed fields such as customer profiles and transactions); semi-structured (data not from fixed fields but using markers or tags to capture data elements, e.g., RFID tags); or unstructured (data not residing in any fixed fields, such as manuscript text). According to an IDC study, the size of the digital data is further expected to grow 44-fold per year to 35 ZB by 2020 [4]. This implies that data analytics will quickly become mainstream, with novel growth in business opportunities.

The massive amounts of observational data stored in databases have grown exponentially, making them extremely hard to access, process, analyze, visualize, share, and manage through standard software application techniques. In 2012, a global project known as “The Human Face of Big Data” was conducted to gather and derive in real time huge volumes of social data [3]. According to this media project, there were over 30 billion pieces of content being posted daily on Facebook, 571 new web sites being developed every minute, and around 4 billion views undertaken on YouTube. This gigantic quantum of information was predicted to grow 50 times larger by the next decade. This tremendous growth rate overwhelms any traditional data processing tools and drives the need for high-performance, fault-tolerant, and hyper-scalable data processing infrastructures. Current tools and technologies are not equipped to continuously listen to and extract value from this voluminous data for adding to business insight. Therefore, development of tools that can keep pace with data growth so that meaningful information can be leveraged from big data is vital, which will provide much advantage to organizations over their competitors [15]. Furthermore, due to the premise of higher profitability and productivity, many firms have led to development (or investment) of BDA in recent years [6].

The investments made in big data continue to grow as organizations continue their search for sustained competitive advantage [2]. Fifteen percent of US businesses invested USD 100 million in 2012 and fifty percent of these invested around USD 500 million each. This number was expected to be 75 percent higher by 2015 [13]. In another report, investments in BDA were around USD 1 trillion in 2013 [17] and expected to be USD 3.8 trillion by 2014 [6]. Yet another study conducted by IDC reports that the expenditure in storage and server devices was USD 12.4 billion in 2010, for high-performance computing capabilities, and by 2014, this would increase to USD 17.2 billion [4]. From these statistics, it is apparent that most organizations believe that BDA would be redefining their competitiveness and growth prospects. Those industry segments which are unable to deploy a BDA strategy in

the next few years could potentially lose market share and business growth [18]. These aspects make this research even more relevant in enabling more and more businesses who seek value from BDA.

2.2. Big Data Characteristics

Big data are categorized by their diverse characteristics, as indicated by early research [19]. The first three main characteristics are volume, velocity, and variety (3Vs of big data) [2]. Volume refers to adoption and growth in the size of data that is produced by various sources. The ever-growing data volume is now in zettabytes (a unit of information equal to one sextillion). Due to the sheer scale of big data, they surpass traditionally applied storage and analytical tools [3]. The velocity refers to the rate at which data are generated and the speed required to process them. The Her Majesty's Inspectorate of Constabulary (HMIC) report [20] illustrates how quickly social media information can spread. Further, the variety is from big data collected via various sources that are very varied in nature, such as text and images [15]. Some sources, such as e-transactions generated from e-commerce sites, deal with structured data [6], while server log files may comprise semi-structured data, and social media sites provide unstructured data [15]. Variety stresses diverse sources of data in different formats, data types, and frequencies. It emphasizes the data function for processing the type of data. For example, functions for image processing for image analysis and recognition are different from analyzing human interactions for human data sources.

A further five characteristics have been put forth as equally important for big data [15,21–23]. These include validity, veracity, virtue, value, and variability. Validity is sometimes referred to as data assurance or trustworthiness of data [23]. This characteristic is based on the accuracy or the tidiness of the processed data. Owing to the sheer big data variety, velocity, and volume, achieving error-free and credible data outcomes is still just a goal, not (yet) a reality [21]. Hence, combinations of precise and imprecise data may exist. For the purpose of providing decision-making data, development of a mechanism that deals with imprecise data is essential when developing a big data analytic framework. A recent study [22] has suggested that social media must augment established sources to showcase a longitudinal extension that informs the designs of existing research instead of acting as a surrogate for them.

Functional dependency as a property articulates the vital relationship between two attributes in relational database theory, highlighting constraints between two sets of attributes. With the emergence of approximations in data validity, the property of relaxed functional dependency can identify constraints beneficial to different aspects—for example, capture of data discrepancies or semantically related data patterns [24]. Veracity is another significant big data characteristic that explains the worthiness of data sources in defining their eligibility and value for providing data in important focal areas [23]. It highlights how accurate (or uncertain) the data are from the sources. Some researchers have related veracity to validity or a part of it. Validity in relation to application and context must be reviewed to increase the data value in decision-making. In the comparison of veracity with validity, veracity necessarily includes validity, but the converse may not be true.

Virtue, as listed in the same report by Williams et al. [22], relates to the ethics behind using big data in social research. The authors argue that very few social media users were concerned about their information being used by third parties. This aspect is often overlooked due to the informal and less structured characteristic of big data [4], and, while this is considered one of its strengths, it can also pose a data security and privacy problem. In the case of sensitive data being processed, whether for regulatory, security, or a privacy condition, their use may present a serious security breach issue. Thus, safety measure is another factor to consider in the development phase of big data analytics. Organizations that include any sensitive data in their BDA must incorporate some forms of safeguard to protect themselves. Finally, value (or viability) showcases another characteristic that refers to the meaning of data within a certain context [15]. Data meanings change in different contexts, so tools that can correctly interpret the connotation are important. Together, this

leads to the goal of BDA as it links the preceding Vs and focuses on the collaboration, extraction, and analysis of big data [22].

Value relates to the level of data importance based on which business decisions can be established. The value element forms a vital characteristic of BDA in formulating security needs and using techniques and methods that provide rigor and security to the process. Variability relates to the ongoing behavioral changes in data sources which provide the data on study objects. This refers to the variabilities of results, possible due to the inconsistencies in the behavior of data sources due to reliability and availability. For example, inconsistencies in data flow rates may develop complexities in query management due to the desynchronization in data retrieval from different sources. Thus, correlating data from varying sources could be an issue [23]. The outcome of the analytics allows organizations to rely on reasonably accurate data, derive meaningful patterns, and make informed decisions to gain a competitive edge in the marketplace.

2.3. Capabilities of Big Data Analytics

The prior literature in big data analytics has identified five main categories of BDA in terms of process capabilities, which include analytical capability, unstructured data analytical capability, decision-making support capability, predictive and preventive capability, and traceability.

Analytical capability refers to the use of tools and techniques in a BDA system that allow big data to be processed and rationally analyzed [25]. For example, analytical capability enables healthcare firms to uncover associations from the enormous healthcare records by identifying emerging patterns among patients [16]. Healthcare organizations can capture big data in real time for processing patients' medical data and potentially identifying previously unnoticed symptoms in patients. In the telecom industry worldwide, analytical capability enables communication service providers (CSPs) to reduce churn and properly harness large data volumes to increase their revenue intake. Furthermore, it allows them to analyze customer patterns or behaviors to identify potential market trends and put forth attractive personalized offers. By adding real-time semantic insights to business decisions, they can offer targeted marketing to improve their market brand.

Between legacy (traditional) data mining methods and big data analytics, the biggest difference is the unique ability of the latter in analyzing unstructured and semi-structured data [16]. Whether it is dealing with doctors' handwritten notes or a social media interaction, the data analytical capability of BDA entails dealing with unstructured data. According to studies, 27 percent of the digital data today are unstructured and 21 percent are semi-structured, which sums up to 48 percent of the total modern data [2,13]. Wang et al. [16] have stressed that the analysis of unstructured data leads to successful targeted marketing, revenue growth, and allows the building of customer segmentation. These data are complex in nature and must be pre-processed before useful insights can be extracted.

BDA supports strong decision-making capability. By producing informative reporting statistics, a big data analytics system can enable managers to understand customers and/or patients, which in turn helps to further improve their decisions and subsequent actions. Furthermore, the reports allow managers to identify business opportunities that traditional business intelligence would be incapable of recognizing. This is evident in the case study of the Continental Airlines, in which the company could turnaround its performance to become a top-ranked airline [9].

Analyzing big data enables organizations to build a more dependable business model with the insights collected. This allows organizations to accurately predict events that are likely to happen and accordingly plan a preventive strategy in advance. In healthcare, predictive capability allows organizations to achieve better cost-saving as well as ensure the implementation of broad-scale disease profiling. In the telecom industry, predictive and preventive capability allows for congestion control as well as churn identification and prevention.

Traceability is the ability for organizations to trace back to the origin from which the data were collected. This is especially important in the healthcare industry, where individual patients can be monitored closely even when they are away from healthcare facilities such as hospitals. This means providing personalized care and home care, which, in a way, increases the healthcare facilities' capacity, allowing for better medical supply utilization rates as well as reducing delays in patient queues. Fraud detection is also possible with real-time data analysis in both the healthcare and telecom industries.

2.4. Impact of Big Data Analytics

The growth of big data is driving the development of its analytics. The investments, however, vary widely over different industry segments. Some industries are represented with novel sources of data whilst others might still be in the development phase of a solution for data already generated. In the telecom and healthcare industries, there has been a phenomenal spurt in the growth of big data over the past decade. Many studies have been conducted on how these industries can leverage big data analytics to their business advantage [3,7,22]. This section provides an overview of these studies and examines the impact of big data analytics in modern industries, as well as the methodology involved. This section focuses on the telecom and healthcare industries, which are known to utilize multiformat (structured, semi-structured, and unstructured) data and therefore need sophisticated data analytic tools.

Today, the data generated in the telecom industry are unprecedented in human history, comprising call, text, and user data. As more and more people are connected online, environmental scanners, wearable life-logging sensors, mobile computing, e-business transactions, networked appliances, and social computing can generate tremendous amounts of data in the blink of an eyelid [7]. With the rapid rate of development and even greater rate of data growth in the telecom industry, it is vital for communication service providers (CSPs) to be equipped with the right tools to harness the ever-growing big data in order to maximize their revenue potential [26]. These data can be used to create transparency and discover customer needs. Companies that can leverage the right analytic tools will be able to identify and grasp business opportunities faster than their competitors.

With the assistance of big data analytics, telecom businesses can capture and analyze big data in real time, provide insights, and enhance decision-making. Furthermore, conducting an analytics program enables telecom firms to boost the efficiency of their networks by segmenting their customer behavior to identify the more active area or time period (e.g., if a concert is held in an originally less populated area, more-than-usual data traffic is expected), re-assigning their Internet resources to these areas in advance to provide congestion control and ensure customer satisfaction [27]. Another key benefit of BDA in the telecom industry includes fraud detection via real-time call data monitoring [26]. Other benefits include reduction of operational expenses by 10–15 percent by increasing performance, such as cell-site optimization by augmenting networks with contextual information, as well as by providing preemptive customer care services.

The major benefit of BDA in the telecom industry is identified as improving revenue via improving customer satisfaction [27,28]. Digital data are growing with the expansion in online consumption; therefore, communication service providers must develop an effective social strategy that analyzes data streams for improving consumer experiences and supporting innovation in future services [7]. Unstructured data (e.g., text, pictures, audio, video) generated via social media are complex and require new techniques for their processing. Therefore, the way in which BDA is implemented is crucial. Leveraging incoming customer data to improve existing services or tailor attractive sales offers will help to minimize churn. A study has shown that strategic use of BDA enabled CSPs to improve their campaign management and their budget by 15 to 25 percent, as well as reduce churn by 8 to 12 percent [26].

The main objective of providing customers with more tailored marketing campaigns is to increase customer satisfaction and prevent churning [27]. For example, upon receiving

a customer complaint, a customized offer is triggered to this customer, reducing their churning tendency [7]. Furthermore, BDA enables CSPs to shift their business intelligence strategy. They no longer look back at historical records, rather looking forward at current data with a preventive and predictive focus, to evaluate actions that are likely to cause and prevent churn events. Other churn-precluding actions used in real time include precise capacity planning, multi-SIM prediction, rotational churn identification, churn location, and leveraging social networking analysis [7].

The healthcare industry is quickly moving from paper to digitized format as they maintain their patient data and related health records. They have collected large amounts of unstructured clinical data which cannot be handled by conventional tools. Big data in healthcare refers to the digitized datasets which are hard (or near impossible) to manage by traditional mining techniques and methods due to their sheer size and complexity [21]. Thus, the implementation of BDA in healthcare is a promising breakthrough [16]. BDA allows healthcare organizations to analyze their voluminous data across a wide range of unstructured data formats (e.g., transcribed notes) as well as structured and semi-structured data (e.g., medical records, home monitoring, clinical trial data). These BDA tools enable processing of medical data to deliver clinical insights, such as realize disease trends, identify patients' prognosis, and improve quality of care as well as the financial performance of the facility [29].

While profit is not (and should not be) the primary goal of implementing BDA in the healthcare industry, it is still vital to achieve growth and expand. The healthcare industry requires tools and technologies for processing big data for the purpose of insightful patient treatments and overall cost savings [30]. It is estimated that savings of USD 300 billion per year can occur in the US alone if BDA is leveraged properly [21]. Of this, two thirds would accrue from reduction of around 8 percent in the country's healthcare expense. Another study indicates that investment wastes are estimated to be USD 165 billion for clinical operations and USD 108 billion for research and development [13]. Therefore, BDA opens up the following possibilities in the healthcare industry: (1) advancing clinical operations—by analyzing patient behavior patterns, BDA makes it more efficient to identify and administer clinically related treatment methods, and (2) research and development opportunities—BDA can enable a leaner and more targeted production process for new drugs and services. Clinical trial data can be monitored and shared globally, improving trial design, patient recruitment, and outcomes. This process can further assist in inspecting any adverse effects on patients globally, thus reducing failure rates and accelerating the overall trial process.

Raghupathi and Raghupathi's [21] study has indicated the benefits of applying segmentation and utilizing predictive modeling tools for patients' data. The study suggests improvements in (i) clinical monitoring and preventative measures; (ii) supporting prevention initiatives by broad-scale disease profiling; (iii) collecting and publishing clinical cases for patients to better understand and determine the care protocol; (iv) preventing fraud by analyzing large volumes of claim requests; and (v) tool development for enabling healthcare organizations to monitor in-hospital and in-house devices in real time. The growing use of mobile devices, remote biomedical equipment, and sensors is a major aspect that supports an increase in home healthcare services [16]. Implementation of these devices further means that a continuous listening approach is in place, therefore ensuring that safety and adverse event prediction are being leveraged to improve the quality of healthcare services through more accurate predictive capabilities. Raghupathi and Raghupathi's study proposed a four-layer conceptual architecture of BDA in healthcare, which represents a traditional health analytic or informatic project structure. In their proposed framework, big data collected from external (e.g., health maintenance organizations, insurance companies, pharmacies, laboratories, government sources) and internal sources (clinical decision support systems, electronic health records, etc.) are pooled as part of the first layer in big data sources. Then, raw data are transformed (cleansed and processed) and stored in data warehouses. The next component involves the selection and application of appropriate

tools for building analytics models. Finally, the data are presented in various formats, readable for decision-making.

In studies related to analytics capability at Twitter, phenomenal growth has been seen in terms of variety of use cases, number of users, complexity, and size of “big data” mining in the last decade. The evolution of Twitter’s capability development and creation of infrastructure for big data mining has required much preparatory work that has preceded the application of data mining algorithms and robust solutions. Schemas have played a vital role in assisting their data analysts in understanding petabyte-scale data stores, although they have been insufficient in generating overall “big picture” insights [31]. A major challenge for the company has been in building a data analytic platform stemming from the heterogeneousness of the different elements that must be integrated into production workflows. Short message-based sentiment mining using a keyword-based classifier has helped Twitter to outline a cataloging process with the potential to extend and further add sentiment-based dimensions, providing a deeper understanding about user preferences [32].

Sentiment analysis allows detection of trends and patterns emerging from data effectively. In another Twitter-based study looking at sentiment analytics, authors reviewed the correlation between the approval rating of (President of the United States) POTUS and his tweets [33]. The mined and cleaned Twitter feed of POTUS was quantified on content basis, or “sentiment score”. Evaluation of tweets in different stages of Mr. Trump’s election campaign showed that his sentiment score increased by 60% over time, revealing a causative correlation between approval rating and POTUS’ Twitter activity.

Wang et al. [16] extended an earlier study (conducted in 2014) to propose a five-layer architectural framework for BDA. These are (1) data layer comprising raw data, including structured data (from traditional digital clinic records), semi-structured data (as a result of monitoring patient logs), and unstructured data (or doctor notes); (2) data aggregation layer that handles data from varied sources of data. The data are read from communication channels, transformed after cleansing and sorting, and loaded into a target database for subsequent analysis; (3) analytics layer responsible for applying appropriate analytics. Depending on the purpose of the analysis, three main analytic tools—in-database analytics, stream computing, and Hadoop map/reduce—may be implemented; (4) information exploration layer that holds the final outputs of the BDA. This mainly includes readable reports for decision-makers to analyze; however, Wang et al. point out that the most important output in healthcare is information monitoring in real time from proactive notifications and alerts. Such information can be analyzed from real-time monitoring devices such as personal medical devices and mobile phones; and (5) data governance layer that emphasizes master data management (MDM) to ensure the data’s availability, accuracy, completeness, and immediacy. MDM helps to manage business information throughout its lifecycle, prevents security breaches, and assists organizations in protecting patient privacy.

3. Research Methodology

A qualitative approach has been adopted for this research study. Multiple case content analyses have been conducted to examine how existing organizations leverage benefits from big data analytics. Through a series of semi-structured interviews, primary data were gathered from six key participants of three large companies (two participants from each company). Two companies were China-based, and one was from New Zealand. The three companies were selected based on two predetermined criteria: (1) the cases should belong to the large enterprise sector (USD 100 million revenue and above) with at least 100 BDA users in each case; (2) the cases should have been engaging in use of BDA tools for data mining and analytic decision-making for at least three years. The firms were purposefully selected, as they are currently global market leaders in their business segments and make much use of BDA as a process tool for decision-making. Moreover, these firms were agreeable to participate in this study. The six participants were either data analysts in

these organizations or associated technologists who were involved in the use of big data analytics and understood the implications. The interviews were held between April and October 2019 to explore the big data analytic practices employed in these companies.

3.1. Data Collection

The participants were contacted via phone or email. A brief introduction was given to explain the purpose of this study, and further appointment was sought from these participants. The researchers' availability dates in China and New Zealand were conveyed to the participants. Once the appointments were confirmed, an information sheet along with a set of questions was sent to the participants. One face-to-face interview was conducted with each of the six participants in their office. Interviews lasted anywhere between 60 and 90 min and were recorded with the participant's permission. Questions were asked to understand the process of BDA implementation (how they sourced, processed, analyzed, and visualized big data), its architecture and infrastructure, as well as how the emerging insights assisted the users in realizing their business objectives. The participants shared their implementation of big data analytics with their perspectives and overall experiences. The confidentiality of the organizations and participants has been maintained with use of pseudonyms. A brief overview of the three case studies is given next.

1. Geevon is a mobile communications and consumer electronics manufacturing company based in Guangdong, China. The company is known for its supply of electronic devices such as Blu-ray players and smartphones. The company is one of the top smartphone manufacturers globally.
2. Yeevon is China's leading manufacturer of dairy products. The company manufactures milk products including fresh, organic, and sterilized milk, milk tea powder, and ice cream and is listed on the Shanghai stock exchange as an "A" category company.
3. Meevon is a New Zealand-based electricity generation and retailing company. All the company's electricity generation is renewable. In 2017, Meevon had a 19% share of the New Zealand retail electricity market.

3.2. Analysis and Evaluation

After each interview, the interview recordings were transcribed. The transcripts were analyzed to evaluate the process of BDA execution as well as understand how benefits were realized by these organizations. The data analysis was conducted using a qualitative software analytical tool, Nvivo 11, by applying the condensation method. This approach categorizes information summaries into clusters based on predefined constructs. The conceptual architectural frameworks on big data analytics proposed by Raghupathi and Raghupathi [21] and Wang et al. [16] (2018) provided a methodological guide for this study. The following section presents each organization's perspective.

4. Findings

The positions of the participants and size of the three companies were very varied; accordingly, the findings gathered from each case have been described separately in the following three subsections.

4.1. Geevon

Geevon is a large smartphone manufacturer, maintaining their market position due to the stylish design of their products. Geevon's BDA allows them to observe user behaviors and establish their preferred interfacing on the phones. In their big data system, the data access activities typically include collecting data, transforming and cleaning sensitive information, and creating metadata. Geevon has identified sourcing policies for accessing data with access method information and has implemented controls for push or pull data access. The data are transformed into an analytical format with the ability to implement any programmable interface through software if required. The company uses a data analyst to source and provide access of relevant data to the big data analytic system. The data

are captured from different sources, such as network operators, Web file transfer protocol (FTP) or similar applications, search engines, scientists and researchers, public agencies, business enterprises, and end users.

According to the participant, Geevon follows a three-layer BDA architectural framework. The first is the foundation layer, which is a basic system. It uses components in the Hadoop ecosystem, including real-time and offline processing, as well as data collection and online analytical processing (OLAP) analysis. These components are currently mainly open-source for applications, and some are self-developed custom systems. On top of these basic systems, Geevon have developed several autonomous services. The platform for developers includes data access and task scheduling. The application-oriented platform is mainly for their internal operators and product staff. This assists them in undertaking multi-dimensional evaluation, report analysis, and user portraits.

To simplify their architectural framework, Geevon incorporates an open-source component called NiFi (from Apache) that allows Geevon to streamline their entire architecture. The advantage of NiFi is its visual interface. One participant displayed the NiFi interface during the interview to show that each processor in the BDA is visualized as a square box. There is data flow between the boxes, and at the same time, the mode of cluster is supported. After the data are processed, they could be automatically allocated to the cluster for data access. Another benefit of NiFi is that it can support rich data conversion (including format conversion, encryption, and decryption operations), underpinning data captured on Hadoop Distributed File System (HDFS) for storage and Kafka for processing and distributed streaming. There is a message queue between each box, and each queue can implement a buffer to keep the architecture stable. Geevon has developed many extension libraries based on NiFi for managing customized data conversions, achieving end-to-end performance testing and optimization and indicator acquisition monitoring.

The impact of their system management was emphasized by the participant. That is, Geevon have developed platform portraits based on their core business values, looking at their BDA from three perspectives. The first is progress of the mission. Some tasks have a lot of pre-dependent tasks. For example, Geevon can have a task with more than 20 pre-tasks. Without task monitoring, the delay could become larger and larger. Their BDA could show progress in the history of a task. The progress is calculated by a formula: $\text{task progress} = \frac{\text{sum}(\text{completed task} \times \text{historical time consuming})}{\text{historical total time}}$. Furthermore, the system manager in Geevon is also in charge of visualizing and finalizing the process result to communicate with other departments or upper management. The system manager coordinates configuration of the big data architectural components in executing the defined workload tasks. At the back end, these workloads are provisioned to individual virtual or physical nodes in the network. A graphical user interface supports the linking of multiple components and applications at the front end. For example, as explained by the participant, users are provided a multi-view service with more than a thousand personalized face recommendations to increase their product interest. They are informed about new services, keeping them actively involved. These strategies help in improving user retention through key indicators, such as increasing the users' click rates, which in turn strengthens the overall retention rates.

4.2. Yeevon

Yeevon is a leader in the fast-moving consumer goods business segment. Yeevon cooperates with authoritative organizations such as Nielsen USA and Mintel UK and has built over 430 data sources with effective data levels to reach the entire network based on the vast amount of data from Internet consumers. Following a scanning sequence, both online and offline, Yeevon's reach is more than 90% of relevant consumer data using their big data radar platform. Through the application of big data technology, Yeevon accurately understands their consumer market. The participant explained that their IT manager monitors the execution of BDA tasks and system governance through management roles to confirm that each task met specific quality of service requirements. The participant

advised that all the data generated in production were collected and analyzed. A data provider typically created an abstract data source for communicating with various data sources (i.e., raw data or data pre-converted by other systems) to provide discovery and data accessibility from different interfaces. The interfaces included an archive that enabled big data applications to find data sources, determine which data to include, determine data access methods, locate data sources, understand the types of analysis supported, recognize the types of access allowed, and identify the data. Consequently, the interfaces provided the ability to identify datasets in the registry, query the registry, and register the data sources.

Through their internal data platform, Yeevon sorts out the needs of consumers, their future research and development direction, and their key requirements of quality and cost control. This guarantees the maintenance of the brand image of Yeevon and provides a safe and healthy product for everyone without a hefty price tag. Yeevon uses an open cloud-based analytics platform from SAP to collect the datasets and Amazon Web Services Datahub to support its storage and mining. When big data application provider services are used, as explained by the participant, they perform a series of operations as established by the IT manager, to meet system requirements, including privacy and security throughout the data lifecycle. The application providers construct specific big data applications by combining general resources with service capabilities in the big data framework to encapsulate commercial logic and functionality into architectural components. The activities conducted by the big data application service include data collection, pre-processing, analysis, visualization, and access, assisted by application experts, platform experts, and consultants. Extraction of information from data usually requires specific data processing algorithms to derive useful insights. Analytical tasks include lower-level coding of commercial reasoning in the of big data systems as well as higher-level coding of the business process logic. These codes usually involve software that implements analysis logic on a batch or stream processing component, leveraging the processing framework of the big data application service to implement the logic of these associations.

Using the message and communication framework of the big data application service, analytical activities can also transfer functions for data and control in the application logic. The visualization activities present the results of the analysis to the data consumers in a way that is most conducive to communication and understanding of knowledge. Usually, visualization features include text-based reports or a graphical rendering of the analytical results. These results could be static and stored in the big data service framework for accessing later. In more cases, visualizing activities often interact with data consumers, big data analytics activities, and big data provider processing frameworks to produce outputs based on data access parameters set by data consumers.

At the other end of the innovation process, BDA has also enabled Yeevon to test-and-learn about their products without going through the cost of full consumer data assimilation and refinement. This has hugely decreased the cost of product development for the company. Yeevon has been able to identify potential consumer demands and analyze trends, taking full advantage of their big data strategy. Yeevon has consolidated big data assimilated from over 1 billion consumers and many partners, and around 5 million points of sale. Their BDA processes have proven beneficial to gather valuable insights into their customer requirements and aspects influencing consumer behavior. For example, Yeevon set up their country's first "ecosystem of maternity and infancy" strategy. This program aimed at personalized services and utilized the "Internet Plus" BDA model to capture and analyze babies' and mothers' data to gain insight into their key nutritional needs. Additionally, Yeevon has implemented specialized chipsets on supermarket checkout PIN machines to gather and analyze regional consumer sales data related to preferences, values, and purchasing power to improve their demand planning, inventory management, and product development processes and strategize against competition. Furthermore, BDA also helps Yeevon to achieve an effective connection with consumers, enhancing Yeevon's corporate brand image. According to the Kantar [34] consulting group's 2016 Global Brand

Footprint report, Yeevon's consumers have increased to more than 1.1 billion globally in the past year and their products have become the most popular choice for Chinese consumers.

4.3. Meevon NZ Limited

Unlike Geevon and Yeevon, Meevon mainly uses BDA to ensure that its data collected is 100% accurate to ensure electricity generation and retailing compliance. As a renewable energy company, Meevon is under obligation to provide energy usage for its customers, to view, track, and see how much they have spent on power and compare with similar homes in nearby area. Based on this information, Meevon also provides tips to help electricity consumers to save their power usage and reduce their bill payments. These services allow Meevon to attract more customers. The software that Meevon has implemented to handle its data management, data cleansing, and data crunching for their BDA is called SAS business intelligence. The system manager at Meevon controlled the BDA activities, including allocation of IT resources. They could flexibly allocate and provide additional physical or virtual resources to accommodate changes/surge in workload necessities due to data or user transaction volume. The data analysts identified the data sources to access relevant data from different interfaces to meet their business analytical requirements. The analysts also ensured security and anonymity in controlling the identification of confidential data and other relevant information. The participant explained that the interface of SAS was hierarchical in nature. It helped the data consumers to collect data by geographical groups for generating visualizations. The data consumers could typically search/retrieve, download, analyze, generate reports, and visualize the graphical outputs. Data consumers accessed the information they were interested in using interfaces or services provided by big data application providers, including data reporting, data retrieval, data rendering, and more. The data consumers could also interact with big data application providers through data access activities to perform data analysis and create visualizations. Interactions could be demand-based, including interactive visualization, creating reports, or using a BI tool provided by a big data application service to drill-down data. The interaction function could also be based on a streaming-based or push-based mechanism, in which case the consumer only needed to subscribe to the output of the big data application system.

In case of Meevon, the end report produced presented a clear image of energy usage across different areas, which helped Meevon to understand their market position. This worked towards preparing a more tailored sales campaign and prevented energy churning rates for their customers. This has led to an overall customer satisfaction improvement for the company. Aside from mining data from its customers, Meevon also uses BDA to effectively reduce manual labor in data input and avoid potential manual errors. Each year, Meevon sets a business target of using SAS to reduce manual labor hours and has reported that, on average, 30 hours of work is saved each week for employees to conduct more Business-As-Usual (BAU) work, effectively reducing company costs. Overall, the company has been very successful in their BDA endeavors to achieve strong growth and customer satisfaction in recent years.

5. Discussion

This paper investigates the capabilities of big data analytics and its potential benefits by examining how organizations evaluate big data to achieve their strategic business objectives. Findings from this research highlight four main components of a successful BDA process, as identified by the study participants, that enhance the BDA capabilities in realizing the strategic goals of business firms. These are discussed below.

5.1. System Coordination

An overall system coordination is essential to meet the BDA requirements, including policies, architecture, resources, and governance, as well as auditing and monitoring actions ensuring that the system is robust to achieve the strategic objectives. The role of the system coordinator includes business leaders, software architects, privacy and security architects,

information architects, data scientists, network designers, and consultants. The relevant data application activities are defined and integrated by the system coordinator into the running vertical system based on the organization's strategic intent. System coordinators often involve one or more role players, with more specific roles, for managing and coordinating the operation of big data systems. These role players can be people, or software, or a combination of both. The system coordination function configures and manages the various big data architecture components to implement the required workloads. These workloads, managed by the system coordinator, can be assigned or provisioned to the lower-level virtual or physical individual node. At the higher level, a graphical user interface can be provided to support the connection of multiple applications and components. System coordination can also achieve monitoring of workloads and system governance through management roles to confirm that each workload meets specific quality of service requirements. To accommodate changes/surge in workload requirements due to data or user transaction volume, system coordination may also flexibly allocate and provide supplementary virtual or physical resources.

5.2. Data Sourcing

Having access to relevant data based on the goals of an organization is a vital activity, so that pertinent data are available to the BDA system. The process encapsulates data capture from network operators, Web/FTP and other applications, search engines, scientists and researchers, public agencies, business enterprises, end users, and more. In a big data system, data sourcing activities typically include collecting data, transforming and cleaning sensitive information, creating metadata, accessing policies for sourcing data and accessing method information, controlling for push or pull data streams, publishing data availability, and implementing a programmable interface through software.

Data access is typically created by identifying an abstract data source for communicating with various data sources (i.e., raw data or data pre-converted by other systems) to provide discovery and data accessibility from different interfaces. The interfaces included an archive that enabled big data applications to find data sources, determine which data to include, determine data access methods, locate data sources, understand the types of analysis supported, and recognize the types of access allowed with identification of the data. The data sourcing process also ensures the security and anonymity requirements in controlling the identification of confidential data and other relevant information. Therefore, the interfaces provided the ability to identify standard datasets in the registry, query the registry, and register the data sources.

5.3. Big Data Application Service

A series of operations are performed by the big data application providers to meet system requirements, including privacy and security, as established by the system coordinator throughout the data lifecycle. Big data application services construct specific applications by combining general resources with capabilities in the big data framework to encapsulate commercial logic and functionality into architectural components. The role of the big data application service includes application experts, platform experts, consultants, and so on. The activities conducted by the big data application service include data collection, pre-processing, analysis, visualization, and access.

The task of the analysis activity is to achieve extraction of knowledge from data. This usually requires specific data processing algorithms to process the data in order to derive useful insights. Analytical tasks include coding of commercial reasoning at the lower level into big data systems by the service providers to allow systemic evaluations. Higher-level coding of the business process logic is performed by system coordinators and data analysts for specific analytics. These codes leverage the processing framework of the big data application service to implement the logic of these associations, usually involving software that implements analytical sense on a batch or stream processing component. Using the

message and communication framework of the big data application service, analytical activities can also transfer functions for data and control in the application logic.

The task of the visualization activity is to present the results of the analysis activities to the data consumer in a way that is most conducive to communication and understanding of knowledge. Visualization features include generating text-based reports or graphically rendering analytical results. These results could be static and stored in the big data service framework for accessing later. In more cases, visualizing activities often interact with data consumers, big data analytics activities, and big data processing frameworks and platforms. This requires interactive visualization based on data sourcing parameters set by data consumers. Visualization activities can be implemented entirely by the application or by using a specialized visual processing framework provided by the big data service provider.

5.4. Data Consumption by End Users

The data clients or end users receive the big data system output focused on their decision-making requirements. Like data sourcing, data output can be consumed by end users or other application systems in the form of business insights. Activities performed by data clients/end users typically include search/retrieval, download, local analysis, report generation, visualization, and, finally, decision-making based on predetermined goals. End users source the information that they are interested in using services or interfaces provided by big data service providers, including data rendering, data retrieval, and data reporting.

End users also interact with big data service providers through data sourcing activities to perform the data analysis and visualization capabilities available. Interactions can be demand-based, including interactive visualization, creating reports, or using a BI tool provided by the big data application service to drill-down data. The interaction function can also be based on a streaming-based or push-based mechanism, in which case the client only needs to subscribe to the output of the big data application system.

The above four main components identified by the participants align with the research framework proposed by Raghupathi and Raghupathi [21] and Wang et al. [16]. As suggested by the participants in the study, three governing characteristics of a big data analytic framework—data security, data privacy, and data management—form essential building blocks that represent the services to provide functionality across the information and technology value chains. These are in concurrence with the four main components of the BDA process, namely system coordination, data sourcing, big data application service, and end users. Although these characteristics are not categorized as components in big data analytics, they are deemed as vital layers that a big data analytic framework should implement. Figure 1 presents a proposed framework with these layers implemented.

The holistic framework comprises two dimensions that represent the big data value chain—the information value chain (horizontal axis) and the IT value chain (vertical axis). In the information value chain dimension, the value of big data is achieved by accessing relevant data streams through the BDA application in four stages—data collection, pre-processing, analysis, and visualization—leading to business insights for end users. In the IT value chain dimension, big data value is achieved by providing networks, infrastructure, platforms, application tools, and other IT services that store and run big data for big data applications. The system coordination with the BDA application service is at the intersection of both dimensions. In the BDA framework, privacy and security components are considered as part of the management role, which also means that the privacy and security roles are related to all the activities and functions within the framework. According to participants, the role of management in BDA is often disregarded, since most organizations see the analytic as a technical challenge.

Although this is not wrong, as the core component of BDA is technical, it is vital for organizations to have management governance in place to ensure the functioning of an efficient system. This feedback aligns with the literature [3,10,11] that emphasizes management of a BDA as a critical challenge. As advised by Suthaharan [11] in his study, the cardinality parameter pertaining to the number of records at any point in an ever-growing dataset demands

the requirement of an effective dispersed database system for capturing data, storage, and analysis of network traffic to predict intrusion. Then, the parameters of continuity (in terms of growth of dataset) and complexity (due to high variety and speed of data processing) add extra difficulties to the task of managing the big data. Therefore, to manage BDA optimally and in a cost-effective manner, the network topology design must be robust and efficient [11]. Within the privacy and security management module, through the different technical means and safety measures, it is vital to develop a comprehensive security protection system for big data analytics by providing a reasonable disaster recovery framework to mitigate risks and realize real-time data remotely. In another study, a first of its kind BDA-based visual speech enhancement framework, VISION corpus, has been developed for evaluating audio-visual datasets captured from a variety of speakers in noisy environments [35]. Having achieved a significant improvement in performance compared to previous approaches, the speech enhancement framework aligns with the proposed architecture in this study comprising data collection (extraction and fusion), pre-processing, analysis, and visualization (validation) stages.

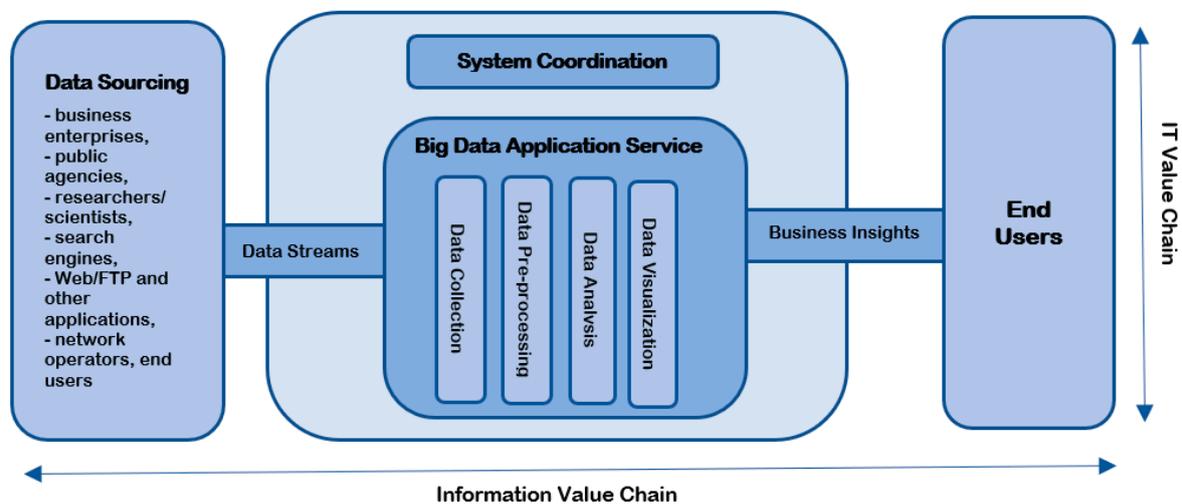


Figure 1. A proposed architectural big data analytic framework.

A summary of the effectiveness of the proposed BDA framework is presented in Table 1, which highlights requirements, activities, and outcomes of the four components of the BDA process.

Table 1. A summary of potential effectiveness of the proposed BDA framework.

Components of a BDA Process	Process Requirements	Role of Process Owner	Process Activities and Outcomes
System coordination	To meet business policies, architecture, resources, governance, auditing, and monitoring actions to ensure system is robust	Includes business leaders, software architects, privacy and security architects, data scientists, network designers, and consultants	<ul style="list-style-type: none"> - Configures and manages big data architecture components to implement the required workloads - With the help of graphical user interface, supports connection of multiple applications and components - Monitors workloads and system governance - Meets specific quality of service requirements - Allocates and provides supplementary virtual or physical resources

Table 1. Cont.

Components of a BDA Process	Process Requirements	Role of Process Owner	Process Activities and Outcomes
Data sourcing	Encapsulates data capture from network operators, Web/FTP applications, search engines, scientists and researchers, public agencies, business enterprises, and end users	Information architects, data analysts, application experts, system engineers, platform experts, network designers, and consultants	<ul style="list-style-type: none"> - Collects data, transforms and cleans sensitive information - Creates metadata - Follows access policies for sourcing data and accessing method information - Implements control for push or pull data streams - Publishes data availability - Implements a programmable interface through software for data extraction
Big data application service	To meet system requirements including privacy and security and construct specific applications by combining general resources with capabilities to encapsulate commercial logic and functionality into architectural components	Includes system coordinators, data analysts, system engineers, application experts, platform experts, and consultants	<ul style="list-style-type: none"> - Data collection, pre-processing, analysis, visualization, and access - Analysis activity extracts knowledge from data - Runs specific data processing algorithms to derive insights - Analytical tasks include coding of commercial reasoning into big data system - Coding of business process logic to implement analytic sense on batch or streaming components - Visualization activity presents results of the analyses to data consumer - Visualization includes generating textual reports or graphically rendering analytical results - Visualization activities are implemented by the application or by using a specialized visual processing framework provided by big data service provider
Data consumption by end users	Data clients, end users, business analysts, functional experts, product developers, process designers, system architects	Receiving system data output focused on decision-making requirements or other application systems in the form of business insights	<ul style="list-style-type: none"> - Search/retrieval, download, local analysis, report generation, visualization, and decision making based on predetermined goals - End users source information using data service provider interfaces, including data rendering, data retrieval, and data reporting - End users also interact with big data service providers through data sourcing activities to perform data analysis and interactive visualization - The interaction can be based on a streaming-based or push-based mechanism by subscribing to the output of big data application - Or create reports using a BI tool by drilling down data

6. Conclusions

The big data analytic framework can be considered as a generic big data system model. It represents the logical functional components of a common, technology-independent big data system with interoperability interfaces between components, which can be used as a general technical reference framework for developing various specific types of big data

application system architectures. The goal is to create an open big data technology reference architecture that enables senior decision-makers, data architects, software developers, data scientists, and system engineers to create a solution in an interoperable big data ecosystem. Using multiple methods, various big data features are congregated into a common big data application system framework that supports different environmental settings, including loosely coupled vertical industries and tightly integrated enterprise systems. This framework helps us to understand how big data systems complement or differentiate from the current intelligent analytics and traditional data application systems such as database management.

The overall layout of the big data reference framework representing the big data value chain has a two-dimensional architecture across the IT and information value chains (shown in Figure 1). In the IT value chain dimension, value of big data is achieved through the availability of required services, networks, and infrastructure to store and run the big data applications. In the information value chain dimension, the big data value is achieved through collection of data, pre-processing, analyses, and visualization to achieve strategic business objectives. The coordination and big data application service is at the intersection of the two dimensions, demonstrating that the implementation of big data analytics provides value to stakeholders of big data in both IT and information value chains.

Four main components of big data analytics are identified from this research: system coordination, data sourcing, big data application service, and end users. Each of these diverse components has a vital functional significance in the big data analytic framework to produce business insights. Furthermore, three very important aspects of security, privacy, and management are identified, which are the governing layers of a BDA framework. These layers represent the building blocks that provide functionality and services to the four major components of the big data analytics framework. The functionality of these key layers is extremely important and should therefore be integrated into the big data analytics framework. These findings, identifying the key activities and outcomes of the BDA process and the big data reference framework, are of significance to academic researchers as well as business professionals for use in both research and practice. The study insights inform researchers and practitioners about the advantages of BDA, its implementation processes, and the effectiveness of the proposed framework. Although this study is limited to three cases in specific industry sectors, the findings are applicable to a variety of organizational settings in different industry segments. Future research is suggested in more diverse industry segments using the BDA framework developed to evaluate organizational leverage of BDA and compare with the findings of this study.

Author Contributions: Conceptualization, S.M. and X.L.; Methodology, S.M. and X.L.; Software, S.M. and X.L.; Validation, S.M. and X.L.; Formal Analysis, S.M. and X.L.; Investigation, X.L.; Resources, S.M. and X.L.; Data Curation, X.L.; Writing—Original Draft Preparation, X.L.; Writing—Review and Editing, S.M.; Visualization, S.M. and X.L.; Supervision, S.M.; Project Administration, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to privacy issues.

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

References

1. Cox, M.; Ellsworth, D. Application-controlled demand paging for out-of-core visualization. In Proceedings of the IEEE Conference on Visualization'97 (Cat. No. 97CB36155), Phoenix, AZ, USA, 24 October 1997; pp. 235–244.
2. Nedelcu, B. About big data and its challenges and benefits in manufacturing. *Database Syst. J.* **2013**, *4*, 10–19.
3. Sagiroglu, S.; Sinanc, D. Big data: A review. In Proceedings of the IEEE International Conference on Collaboration Technologies and Systems (CTS), San Deigo, CA, USA, 20–24 May 2013; pp. 42–47.

4. Villars, R.L.; Olofson, C.W.; Eastwood, M. *Big Data: What It Is and Why You Should Care*; White Paper; IDC: Framingham, MA, USA, 2011; pp. 1–14.
5. O'Donovan, P.; Gallagher, C.; Leahy, K.; O'Sullivan, D.T.J. A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications. *Comput. Ind.* **2019**, *110*, 12–35. [[CrossRef](#)]
6. Akter, S.; Wamba, S.F. Big data analytics in e-commerce: A systematic review and agenda for future research. *Electron. Mark.* **2016**, *26*, 173–194. [[CrossRef](#)]
7. Nwanga, M.; Onwuka, E.; Aibinu, A.; Ubadike, O. Impact of big data analytics to Nigerian mobile phone industry. In Proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM), IEEE, Dubai, UAE, 3–5 March 2015.
8. Bloch, D. Increasing the Relevance of Special Offers to Customers: The Vodafone New Zealand Story. Available online: <https://assets.teradata.com/resourceCenter/downloads/CaseStudies/Vodafone%20Case%20Study%2003.14%20V3.pdf> (accessed on 4 August 2019).
9. Anderson-Lehman, R.; Watson, H.J.; Wixom, B.H.; Hoffer, J.A. *Continental Airlines Takes off with Real-Time Business Intelligence*; Technical Report; Teradata University Network: San Diego, CA, USA, 2005.
10. Luna, D.; Mayan, J.; García, M.; Almerares, A.; Househ, M. Challenges and potential solutions for big data implementations in developing countries. *Yearb. Med Inform.* **2014**, *23*, 36–41.
11. Suthaharan, S. Big data classification: Problems and challenges in network intrusion prediction with machine learning. *Perform. Eval. Rev.* **2014**, *41*, 70–73. [[CrossRef](#)]
12. Hilbert, M. Big Data for Development: From Information-to Knowledge Societies. Available online: <https://ssrn.com/abstract=2205145> (accessed on 7 August 2019).
13. Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A. Big Data: The Next Frontier for Innovation, Competition, and Productivity. Available online: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation> (accessed on 7 August 2019).
14. Carlin, S.; Curran, K. Cloud computing technologies. *Int. J. Cloud Comput. Serv. Sci.* **2012**, *1*, 59. [[CrossRef](#)]
15. Chandarana, P.; Vijayalakshmi, M. Big data analytics frameworks. In Proceedings of the 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), IEEE, Mumbai, India, 4–5 April 2014; pp. 430–434.
16. Wang, Y.; Kung, L.; Byrd, T.A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Chang.* **2018**, *126*, 3–13. [[CrossRef](#)]
17. Lunden, I. *Forrester: \$2.1 Trillion will Go into It Spend in 2013*; TechCrunch: Bay Area, San Francisco, CA, USA, 2013.
18. Columbus, L. 84 Percent of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes in the Next Year. Available online: <https://www.forbes.com/sites/louiscolombus/2014/10/19/84-of-enterprises-see-big-data-analytics-changing-their-industries-competitive-landscapes-in-the-next-year/#1fdf206817de> (accessed on 4 August 2019).
19. Gerhardt, B.; Griffin, K.; Klemann, R. *Unlocking Value in the Fragmented World of Big Data Analytics*; Cisco Internet Business Solutions Group: San Jose, CA, USA, 2012; p. 11.
20. *Her Majesty's Inspectorate of Constabulary. Policing Public Order: An Overview and Review of Progress against the Recommendations of Adapting to Protest and Nurturing the British Model of Policing*; Government Report; HMIC: London, UK, 2011.
21. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 1–10. [[CrossRef](#)] [[PubMed](#)]
22. Williams, M.L.; Burnap, P.; Sloan, L. Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *Br. J. Criminol.* **2017**, *57*, 320–340. [[CrossRef](#)]
23. Alsaig, A.; Alagar, V.; Ormandjieva, O. A critical analysis of the V-model of big data. In Proceedings of the 12th IEEE International Conference On Big Data Science And Engineering, New York, NY, USA, 1–3 August 2018; pp. 1809–1813.
24. Caruccio, L.; Deufemia, V.; Polese, G. Mining relaxed functional dependencies from data. *Data Min. Knowl. Discov.* **2020**, *34*, 443–477. [[CrossRef](#)]
25. Chen, H.; Chiang, R.H.; Storey, V.C. Business intelligence and analytics: From big data to big impact. *MIS Q.* **2012**, *36*, 1165–1188. [[CrossRef](#)]
26. Banerjee, A. *Big Data and Advanced Analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity*; Senior Analyst; Heavy Reading: New York, NY, USA, 2013.
27. Acker, O.; Blockus, A.; Pötscher, F. *Benefiting from Big Data: A New Approach for the Telecom Industry*; Strategy&, Analysis Report; PricewaterhouseCoopers (formerly Booz and Company): New York, NY, USA, 2013.
28. Fox, B.; Dam, R.; Shockley, R. Analytics: Real-World Use of Big Data in Telecommunications, how Innovative Communications Service Providers Are Extracting Value from Uncertain Data. Available online: <http://www-935.ibm.com/services/multimedia/Analytics.pdf> (accessed on 29 September 2016).
29. Jiang, P.; Winkley, J.; Zhao, C.; Munnoch, R.; Min, G.; Yang, L.T. An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. *IEEE Syst. J.* **2016**, *10*, 1147–1159. [[CrossRef](#)]
30. LaValle, S.; Lesser, E.; Shockley, R.; Hopkins, M.S.; Kruschwitz, N. Big data analytics and the path from insights to value. *MIT Sloan Manag. Rev.* **2011**, *52*, 21.
31. Lin, J.; Ryaboy, D. Scaling big data mining infrastructure: The Twitter experience. *ACM SIGKDD Explor. Newsl.* **2013**, *14*, 6–19. [[CrossRef](#)]

32. Baumgarten, M.; Mulvenna, M.D.; Rooney, N.; Reid, J. Keyword-based sentiment mining using twitter. *Int. J. Ambient Comput. Intell.* **2013**, *5*, 56–69. [[CrossRef](#)]
33. Sahu, K.; Bai, Y.; Choi, Y. Supervised sentiment analysis of twitter handle of president trump with data visualization technique. In Proceedings of the 10th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, Las Vegas, NV, USA, 6–8 January 2020.
34. Kantar. The New Fmcg Ranking Is Out. Available online: <https://www.kantarworldpanel.com/global/news/Brand-%20Footprint-report,-the-new-FMCG-ranking-is-out> (accessed on 4 August 2019).
35. Gogate, M.; Dashtipour, K.; Hussain, A. Visual Speech In Real Noisy Environments (VISION): A Novel Benchmark Dataset and Deep Learning-based Baseline System. *Interspeech 2020* **2020**, 4521–4525. [[CrossRef](#)]